A First Course in Linear Algebra

I. Bokor

Draft, May, 2014

To err is human.
To forgive is divine.
To undertsand is everything.

Preface

What Is Linear Algebra?

1. **What is the subject matter of linear algebra?**
2. **Why is linear algebra called "linear algebra"?**
3. **Why study linear algebra?**

These are natural questions deserving at least provisional answers.

1. Linear algebra is the study of vector spaces and linear transformations between them.

2. The name "linear algebra" may seem peculiar, given the answer to the first question. The reasons for linear algebra's being called "linear algebra" are historical. It is useful to think of it as the algebra of mappings of co-ordinate systems which map lines to lines.

3. Linear algebra is probably the most widely applied part of mathematics. It is applied in statistics, physics, economics, computer science, ... It will become apparent that the calculus you have leant so far is, in essence, linear algebra, being the study of certain vector spaces of real valued functions, and the taking of limits, differentiation and integration are linear transformations.

There are two principal aspects to mathematics in general, and to linear algebra in particular: theoretical and computational. The theory not only organises and explains the relevant concepts, but also provides the foundations for the calculations, providing explicit formulæ and sometimes even algortihms. A major part of mastering mathematics subjecting of learning how these two aspects are related and how to move from one to the other.

One difficulty faced when learning linear algebra is that many calculations are very similar, and can therefore cause confusing without a grasp of their theoretical context and significance. It can be tempting to draw false conclusions.

On the other hand, while many statements are easier to express elegantly and to understand from a purely theoretical point of view, you will need to "get your hands dirty" to apply them to concrete problems.

Mathematics often has different formulations of the same concept or theorem, when it is not obvious that they are equivalent. It is common for one formulation to express general aspects with clarity, while being all but useless for practical calculation, and another formulation to be computationally convenient, but give no insight into the underlying structure or concepts.

Both are indispensable. Without the conceptual formulation, there can be no understanding of the underlying structures, and so no theoretical development, no expansion of the range of applicability, no solid foundation for the techniques of calculation. On the other hand, concrete applications require calculation.

We develop the theory first, starting with a handful of examples familiar from your prior studies, on which the development can be tested, and then show how this leads concrete calculations. One significant advantage of this approach is that important definitions, which are usually presented *ad hoc* without any explanation, and seem obscure and esoteric, become natural and obvious.

Where You Have Already Met Linear Algebra

You have already met aspects of linear algebra already in your study of mathematics, although your attention may not have been drawn to this fact at the time. Here are some of the occasions you have met linear algebra — or seen its application.

1. In the algebra section of MATH101, matrices and determinants are studied, including eigenvalues and eigenvectors, algebraic operations on matrices and determinants.

2. The calculus section of MATH101 studies an example of a real vector space (even though it is not called one) and show explicitly and in detail that it is, in fact, a vector space. Certain important subspaces are also explicitly studied. It is also shown in detail in MATH101 that taking limits and differentiation are linear transformations.

3. MATH102 continues the study of these vector spaces, showing explicitly and in detail that integration (the definite integral at least) is a linear transformation.

4. The differential equations section of MATH102 studies, at some length and in some detail, ordinary linear differential equations with constant coefficients. The techniques for solving them are, historically, amongthe first applications of linear algebra, and illustrate the power and importance of *characteristic equations*, *characteristic values* and *characteristic vectors*. (The last two terms are synonymous with eigenvalues and eigenvectors)

5. You have met and used matrices in AMTH140, learning some of their algebraic properties and see them applied to computing such things as the number of different paths between any two vertices in a graph.

6. In MATH140 you have studied at some length and in some detail, *linear difference equations with constant coefficients*. The techniques for solving them are identical to those used in the differential equations section of MATH102 — in fact the two correspond perfectly if "λ^n" and "n" in the case of difference equations are replaced, respectively by"$e^{\lambda x}$" "x^n" in the case of differential equations — and illustrate, once again, the power and importance of characteristic equations, characteristic values, and characteristic vectors.

7. The first part of PMTH212 deals with two and three dimensional real vector spaces, including the notion of "inner product". Matrix products arise as the "Chain Rule" for differentiation, and determinants enter when integrating by substitution. The axioms defining a vector space are given explicitly and the study of quadrics treats an application of the theory of bilinear forms.

An Overview

An analysis of the features common to the examples listed above leads to the notions of *vector space* and *linear transformation*. Linear algebra is their study.

A *vector space* is a mixed object. It has two components, *vectors* and *scalars*.

Scalars behave like the rational numbers in that they can be added, subtracted and multiplied. Division by any non-zero scalar is also possible. In other words, the scalars form a *field*.

Vectors can be added and subtracted but not, in general, multiplied by each other. The vectors form an *abelian group*.

The interaction between vectors and scalars consists of "multiplying" a vector by a scalar. The field of scalars *acts* on the abelian group of vectors.

To pass one vector space to another, to *transform* vector spaces, or to compare them, we have *linear transformations*. These are functions between vector spaces (with common field of scalars) respecting the vector space structure. We consider two vector spaces with common scalars to be essentially the same if the only if the difference between them is purely notational: what the elements are called, or how they are designated. This intuition is formulated mathematically by the notion of *isomorphism*: an isomorphism of vector spaces is a linear transformation which has an inverse linear transformation.

We commence with a handful of basic examples of vector spaces and show how to construct other vector spaces from these. In particular, we construct the *direct sum* of vector spaces with common scalars and determine when a subset of a vector space forms a *vector sub-space*. We also show that the set of all linear transformations between two vector spaces is again a vector space. Because of the importance of this last vector space, we examine it in detail, showing that algebraic operations can be defined on it, allowing computations with linear transformations which parallel algebraic computations with integers, with the exception that the "multipication" is not commutative.

It is natural to ask whether there are vector spaces essentially different from those we have constructed. This leads to one of the central tasks of linear algebra, namely, the *classification* of vector spaces with common scalars into *isomorphism classes*.

It is an amazing fact this can be achieved by calculating a single numerical invariant, the *dimension* of a vector space over a given field, for two vector spaces with common scalars are isomorphic if and only if they have the same dimension.

We prove this only for *finitely generated* vector spaces, by showing that each such vector space has a *basis* and that two vector spaces with common scalars are isomorphic if and only if any basis for one has the same number of elements as any basis for the other. This number is the dimension.

The proof of this comprises showing that every vector space can be written as the direct sum of non-trivial subspaces, none of which can be further decomposed in this manner, the number of direct summands being precisely the dimension of the vector space: Each indecomposable summand has dimension 1.

Moreover, choosing a basis is the same as choosing such a decomposition.

Choosing a basis for each of our vector spaces also allows us to represent each linear transformation by a *matrix*, whose coefficients are scalars, as long as each vector space is finitely generated. It is this use of matrices which makes many things computable.

We define an *addition* for matrices to represent the addition of linear transformations and A *multiplication* for matrices to represent the composition of linear transformations. The requirement that the algebraic operations on matrices represent the corresponding algebraic operations on linear transformations forces the familiar definitions and restrictions. Their previous apparent

arbitrariness vanishes, and their properties follow, without any computation from the properties of the operations on linear transformations.

An analysis of the finer structure of matrices not only helps to facilitate calculations, but it also leads to the definition of the *determinant* for "square" matrices. This, in turn, determines whether a matrix represents an isomorphism, for a matrix has non-zero *determinant* if and only if it represents an isomorphism, and vice-versa. We even derive an algorithm for find the inverse.

We can also form *directs sum of linear transformations*. But whereas every vector space is a direct sum of as many one-dimensional subspaces as its dimension, it is not true in general that every linear transformation of a vector space to itself can be expressed as the direct sum of that many linear transformations between such subspaces!. Indeed, this is possible if and only if the vector space has a basis consisting of *eigenvectors*. Then the matrix (in the finite-dimensional case) of the linear transformation, with respect to such a basis, has all of its non-diagonal coefficients 0, and the diagonal entries are precisely the *eigenvalues*.

The above applies to any vector space, with restriction, at most, to finite dimensionality.

Some vector spaces admit additional structure. One such structure is an *inner product*. This allows us to study the vector space in question geometrically. For the inner product allows us to measure angles and speak of distances in the vector space concerned. This richer structure allows for a deeper study and finds wide application both within mathematics, statistics and in other sciences and technology.

A familiar example is digital recording of sound and pictures. Other important applications include quantum mechanics and relativity theory.

Contents

It is not of the essence of mathematics to be occupied with the ideas of number and quantity.

George Boole

Chapter 1

Notation; Sets and Functions

We revise concepts used in these notes, expressing them in the form they are used, and use the occasion to fix notation and conventions.

As the Greek alphabet is commonly used in mathematics, but may not be familiar to the reader, we include it first.

1.1 The Greek Alphabet

alpha	α	A
beta	β	B
gamma	γ	Γ
delta	δ	Δ
epsilon	ϵ, ε	E
zeta	ζ	Z
eta	η	H
theta	θ, ϑ	Θ
iota	ι	I
kappa	κ	K
lambda	λ	Λ
mu	μ	M
nu	ν	N
xi	ξ	Ξ
omicron	o	O
pi	π, ϖ	Π
rho	ρ, ϱ	P
sigma	σ, ς	Σ
tau	τ	T
upsilon	υ	Υ
phi	ϕ, φ	Φ
chi	χ	X
psi	ψ	Ψ
omega	ω	Ω

1.2 Logical Notation

It is sometimes convenient to use logical notation.

We list the notation we use.

$P \Longrightarrow Q$	for	"if P, then Q", or "Q whenever P", or "P only if Q";
$P \Longleftrightarrow Q$	for	"P if and only if Q", that is to say P and Q are logically equivalent;
$P :\Longleftrightarrow Q$	for	"P is defined to be equivalent to Q";
$\forall$	for	"For every ...";
$\exists$	for	"There is at least one ... ";
$\exists!$	for	"There is a unique ...", or "There is one and only one ...".

1.3 Sets

The mathematics we study in this course can be expressed entirely in terms of *sets* and *functions* between sets.

While the notion of sets and functions are presumed to be familiar, we present a summary of the set-theoretical concepts and definitions used in this course and use the occasion to summarise notational conventions we use.

A *set* is almost any reasonable collection of things. We shall not attempt a more formal definition in this course. The things in the collection are called the *elements* of the set in question. We write

$$x \in A$$

to denote that x is an element of the set A and

$$x \notin A$$

to denote that x is not an element of the set A.

We do not exclude the possibility that x be a set in its own right, except that x cannot be A:

We explicitly exclude $A \in A$.

Two sets are considered to be the same when they comprise precisely the same elements, in other words, when every element of the first set is also an element of the second and vice versa.

When two sets are not necessarily the same, elements of one could still be elements of the other.

Definition 1.1. Given two sets A and B, $A = B$ if and only if x is an element of A when and only when x is an element of B.

The set A is a *subset* of B if and only if $x \in B$ whenever $x \in A$. This is denoted

$$A \subseteq B$$

B is called a *proper subset* of A if and only if B is a subset of A, but $B \neq A$. This is denoted

$$B \subset A$$

Using our notational conventions, given two sets A and B,

$$A = B :\Longleftrightarrow \Big((x \in A) \Leftrightarrow (x \in B)\Big).$$

$$A \subseteq B \colon \iff \Big((x \in A) \Rightarrow (x \in B) \Big).$$

We see that $A = B$ if and only if $A \subseteq B$ and $B \subseteq A$.

When we wish to describe a set, we can do so by *listing* all of its elements. Thus, if the set A has precisely a, b and c as its elements, then we write

$$A = \{a, b, c\}.$$

Example 1.2. By Definition 1.1 on the facing page, $\{a, b\}$, $\{a, b, b, b\}$ and $\{a, a, a, a, a, a, b\}$ are all the same set.

Another way of describing a set is by prescribing a number of *conditions* for membership of the set. In this case we write

$$A = \{x \mid P(x), Q(x), \ldots\}$$

to denote that the set in question consists of all those x for which $P(x), Q(x), \ldots$ all hold.

There are important operations on sets.

Definition 1.3. The *union* of the sets A and B is again a set. It is the set of all those objects which are in one, or other (or both). It is denoted by

$$A \cup B.$$

Using the notation above,

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

[Here, := has been used to signify that the expression on the left hand side is *defined* to be equal to the expression on the right hand side.]

Definition 1.4. The *intersection* of the sets A and B is again a set. It is the of all those objects which are elements of both. It is denoted by

$$A \cap B.$$

In other words,

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

Those elements of A that are not also elements of B form a set in their own right.

Definition 1.5. The *relative complement of B in A* comprises those elements of A that are not also elements of B. It is again a set in their own right. It is denoted by

$$A \setminus B,$$

so that

$$A \setminus B := \{x \in A \mid x \notin B\}.$$

Definition 1.6. Given sets A, B, their *(Cartesian) product* is again a set. It is the set of all ordered pairs, with the first member of each pair an element of A, and the second an element of B. It is denoted by

$$A \times B,$$

so that

$$A \times B \colon = \{(x, y) \mid x \in A, y \in B\}.$$

Forming unions, intersections and cartesian products can be extend to larger collections of sets than just two.

Definition 1.7. An *indexed family* of sets, with *indexing set* Λ consists of a collection of sets, containing one set, A_λ, for each element λ of the indexing set Λ. This is written as

$$\{A_\lambda \mid \lambda \in \Lambda\}.$$

Definition 1.8. Given the indexed family of sets $\{A_\lambda \mid \lambda \in \Lambda\}$, their *union*, *intersection* and *Cartesian) product* are the sets defined, respectively ,by

$$\bigcup_{\lambda\in\Lambda} A_\lambda := \{x \mid x \in A_\lambda \text{ for at least one } \lambda \in \Lambda\}$$
$$\bigcap_{\lambda\in\Lambda} A_\lambda := \{x \mid x \in A_\lambda \text{ for every } \lambda \in \Lambda\}$$
$$\prod_{\lambda\in\Lambda} A_\lambda := \{(x_\lambda)_{\lambda\in\Lambda} \mid x_\lambda \in A_\lambda \text{ for all } \lambda \in \Lambda\}.$$

Here $(x_\lambda)_{\lambda\in\Lambda}$ denotes a *generalised sequence*, namely, an ordered choice of elements x_λ, one for each $\lambda \in \Lambda$. Ordered pairs arise when $\Lambda = \{1, 2\}$ and sequences when $\Lambda = \mathbb{N}$.

A number of sets occur with such frequency that special notation has been introduced for them. These include the sets $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$ consisting respectively of all *natural numbers*, all *integers*, all *rational numbers*, all *real numbers* and all *complex numbers*.

Explicitly,

$$\mathbb{N} := \{0, 1, 2, 3, \ldots\}$$
$$\mathbb{Z} := \{\ldots -3, -2 - 1, 0, 1, 2, 3, \ldots\}$$
$$\mathbb{Q} := \{x \in \mathbb{R} \mid x = \frac{p}{q} \text{ for some } p, q \in \mathbb{Z}, \text{ with } q \neq 0\}$$
$$= \{x \in \mathbb{R} \mid x = \frac{p}{q} \text{ with } p \in \mathbb{Z} \text{ and } q \in \mathbb{N} \setminus \{0\}\}$$

Observe that

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

We write $\emptyset$ for the *empty set*, which is the (unique!) set with no elements. Note that it is a subset of every set, that is, if X is any set, then $\emptyset \subseteq X$.

1.4 Functions

To compare sets, we have the notion of a *function* or *map* or *mapping*.

Definition 1.9. A *function*, *map*, or *mapping* consists of three separate data,

(i) a *domain* that is, a set on which the function is defined,

(ii) a *co-domain*, that is, a set in which the function takes its values, and

(iii) the assignment to each element of the domain of definition of a uniquely determined element from the set in which the function takes it values.

This is conveniently depicted diagrammatically by

$$f\colon X \longrightarrow Y,$$

or

$$X \xrightarrow{f} Y$$

Here X is the domain of definition, Y is theco-domain and f is the name of the function.

We write $X = \mathrm{dom}(f)$ and $Y = \mathrm{codom}(f)$ when X is the domain and Y the co-domain of f.

It is common to denote the function by f alone. We shall only do so when there is no danger of confusion. If we wish to express explicitly that the function, $f\colon X \longrightarrow Y$, assigns the element $y \in Y$ to the element $x \in X$, then we write $f\colon x \longmapsto y$ or, equivalently, $y = f(x)$, a form undoubtedly familiar to the reader.

Sometimes the two parts are combined as

$$f\colon X \longrightarrow Y, \quad x \longmapsto y$$

or as

$$\begin{aligned} f\colon X &\longrightarrow Y \\ x &\longmapsto y. \end{aligned}$$

Observation 1.10. A function should not be thought of just in terms of mathematical formalæ, even if most functions the reader will deal with are of this form.

One reason is that not every function can be expressed in terms of a mathematical formula.

Example 1.11. Let X be the set of all human beings and Y the set of all male human beings.

The function

$$f\colon X \longrightarrow Y, \quad x \longmapsto \text{ the biological father of } x$$

cannot be expressed in terms of a mathematical formula.

Other reasons why functions should not be thought of just in terms of mathematical formalæ, and more examples, will arise shortly.

Definition 1.12. If f assigns $y \in Y$ to $x \in X$, then we say that y is the *image of x under f* or just the *image of x*.

Two functions f and g are *equal*, that is $f = g$ if and only if

(i) $\mathrm{dom}(f) = \mathrm{dom}(g)$

(ii) $\mathrm{codom}(f) = \mathrm{codom}(g)$

(iii) $f(x) = g(x)$ for every $x \in \mathrm{dom}(f)$.

In other words, to be the same, two functions must share both domain and co-domain as well as agreeing everywhere.

Observation 1.13. This provides another three reasons why functions should not be thought of purely in terms of formalæ.

In the first place, when a function can be defined in terms of a mathematical formula, it can be defined in terms of other formalæ. It can be a significant theorem to show that two different formulæ define the same function

Example 1.14. An example familiar from trigonometry is that the function

$$\mathbb{R} \longrightarrow \mathbb{R}, \quad x \longmapsto 1,$$

expressed by the formula $f(x) = 1$ is the same function as

$$\mathbb{R} \longrightarrow \mathbb{R}, \quad x \longmapsto \cos^2(x) + \sin^2(x)$$

expressed by the formula $f(x) = \cos^2(x) + \sin^2(x)$.

This is *Pythagoras' Theorem*, one of the oldest and most frequently used theorems in the history of mathematics.

In the second place, there are distinct functions whose domains agree, which agree at every point (and therefore have the same range). Thus the only difference between them is that they have different co-domains: *They only differ in the values they do* **not** *take!*

Example 1.15. The two functions

$$f\colon \mathbb{N} \longrightarrow \mathbb{N}, \quad x \longmapsto x$$
$$g\colon \mathbb{N} \longrightarrow \mathbb{Z}, \quad x \longmapsto x$$

are given by the same formula, but are different functions.

At this stage, it may not be clear to the reader that they are, in fact, different functions and it may seem peculiarly pedantic to distinguish these two functions. However there are important similar examples in algebraic and geometric setting, where the difference is crucial.

In the third place, we can define functions *piecewise*, so that its values are determined differently in different parts of its domain.

Example 1.16. Let $X = \mathbb{R} \setminus \{0\}$ and $Y = \mathbb{R}$.

Then

$$f\colon X \longrightarrow Y, \quad x \longmapsto \begin{cases} x & \text{for } x < 0 \\ x^3 & \text{for } x > 0 \end{cases}$$

is a well-defined function which cannot be expressed in terms of a single mathematical formula.

The next lemma shows when a single function can be defined by defining it, possibly differently, on different parts of its domain. Such functions are *defined piece-wise*.

Lemma 1.17. *Given functions $g\colon A \longrightarrow Y$ and $h\colon B \longrightarrow Y$, with $g(x) = h(x)$ whenever $x \in A \cap B$, there is a unique function $f\colon A \cup B \longrightarrow Y$ such that $f(a) = g(a)$ for all $a \in A$ and $f(b) = h(b)$ for all $b \in B$.*

Proof. Put $X := A \cup B$ and define f by

$$f\colon X \longrightarrow Y, \quad x \longmapsto \begin{cases} g(x) & \text{if } x \in A \\ h(x) & \text{if } x \in B \end{cases} \qquad (*)$$

This definition is forced by the requirement that $f(a) = g(a)$ for $a \in A$ and $f(b) = h(b)$ for $b \in B$. This means that $(*)$ is the only possible definition of f. In other words, there cannot be more than one function meeting our requirements.

The only question remaining is whether f is, in fact, a function.

(i) Since $X = A \cup B$ is the union of two sets, it is, itself, a set.

(ii) Y is, by hypothesis, also a set.

(iii) If $x \in X = A \cup B$, then either $x \in A$ or $x \in B$ (or possibly both).

To $x \in A$, f assigns $g(x) \in Y$, which is uniquely determined, since $g\colon A \longrightarrow Y$ is a function.

To $x \in B$, f assigns $h(x) \in Y$, which is uniquely determined, since $h\colon B \longrightarrow Y$ is a function.

Hence, $f\colon X \longrightarrow Y$ is a function unless it happens to assign two different elements of Y to some element of X, which could only occur if $x \in A \cap B$, for then f assigns both $g(x)$ and $h(x)$ to x. As, by assumption, $g(x) = h(x)$ for all $x \in A \cap B$, $f\colon X \longrightarrow Y$ is, indeed, a function.

□

Observation 1.18. In Lemma 1.17 on the facing page, the fact that $X = A \cup B$ ensures that there cannot be more than one function meeting our requirements, and the fact that g and h agree on $A \cap B$ ensure that there must be at least one such function.

Example 1.19. Consider the definition

$$|\ |\colon \mathbb{R} \longrightarrow \mathbb{R}, \quad x \longmapsto \begin{cases} -x & \text{for } x \leq 0 \\ x & \text{for } x \geq 0 \end{cases}$$

To see that $|\ |$ is a function, we define $\mathbb{R}_0^- := \{x \in \mathbb{R} \mid x \leq 0\}$ and $\mathbb{R}_0^+ := \{x \in \mathbb{R} \mid x \geq 0\}$. Then

(i) $g\colon \mathbb{R}_0^- \longrightarrow \mathbb{R},\ x \longmapsto -x$ and $h\colon \mathbb{R}_0^+ \longrightarrow \mathbb{R},\ x \longmapsto x$ are functions;

(ii) $\mathbb{R} = \mathbb{R}_0^- \cup \mathbb{R}_0^+$;

(iii) $\mathbb{R} = \mathbb{R}_0^- \cap \mathbb{R}_0^+ = \{0\}$ and $g(0) = -0 = 0 = h(0)$.

Hence, by Lemma 1.17 on the preceding page, $|\ |$ is a function.

We shall continue the practice of specifying functions in formally correct manner, in order that it become matter of course for the reader to do so as well.

A function, $f\colon X \longrightarrow Y$, can also be represented by means of its *graph*.

Definition 1.20. The *graph*, $\mathrm{Gr}(f)$, of the function $f\colon X \longrightarrow Y$ is

$$\mathrm{Gr}(f) := \{(x, y) \in X \times Y \mid y = f(x)\}.$$

This representation should be familiar from calculus.

Definition 1.21. The *range* or *image* of the function $f\colon X \longrightarrow Y$ is the subset $\mathrm{im}(f)$ of Y defined by

$$\begin{aligned} \mathrm{im}(f) &:= \{y \in Y \mid y = f(x) \text{ for some } x \in X\} \\ &= \{f(x) \mid x \in X\}. \end{aligned}$$

Notice that $\mathrm{im}(f) \subseteq \mathrm{codom}(f)$ always holds, with equality holding only sometimes.

Example 1.22. For the function

$$f\colon \mathbb{R} \longrightarrow \mathbb{R}, \quad x \longmapsto 1,$$

$\mathrm{im}(f) = \{1\} \neq \mathbb{R} = \mathrm{codom}(f)$.

Definition 1.23. Given a function $f\colon X \longrightarrow Y$ and subsets A of X and B of Y, the *image of A under f*, denoted by $f(A)$, and the *inverse image of B under f*, or the *pre-image of B under f*, denoted by $f^{-1}(B)$, are defined by

$$\begin{aligned} f(A) &:= \{y \in Y \mid y = f(x) \text{ for some } x \in A\} \\ &= \{f(x) \mid x \in A\} \\ f^{-1}(B) &:= \{x \in X \mid f(x) \in B\}. \end{aligned}$$

Definition 1.24. The *identity function*, on the set X, denoted id_X, is the function

$$id_X\colon X \longrightarrow X, \quad x \longmapsto x$$

Notice that both the domain and co-domain must be precisely X for the identity function.

Definition 1.25. If X is a subset of Y, then the *inclusion map,* is

$$i_X^Y\colon X \longrightarrow Y,\ x \longmapsto x.$$

We sometimes denote this simply by i when the context makes the domain and co-domain clear.

Observation 1.26. Let X be a subset of Y. Then the functions

$$\begin{aligned} id_X\colon X &\longrightarrow X, \quad x \longmapsto x \\ i_X^Y\colon X &\longmapsto Y, \quad x \longmapsto x \end{aligned}$$

are both given by the same mathematical formula: "$f(x) = x$".

But they are different functions, unless $Y = X$.

For in the definition of i_X^Y, the x on the left of the equality sign is viewed as an element of the set X, whereas on the right hand side it is viewed as an element of the set Y. By contrast, in the definition of id_X, both occurrences of x are as elements of X.

Sometimes we are only interested in the behaviour of a function on a subset of its domain.

Definition 1.27. Given a function $f\colon X \longrightarrow Y$ and a subset A of X, the restriction of f to A, $f\,|_A$, is the function

$$f\,|_A\colon A \longrightarrow Y, \quad a \longmapsto f(a)$$

Note that unless $A = X$, this is *not* the same function as f, even though the two functions agree everywhere they are both defined.

Functions can sometimes be *composed*.

Definition 1.28. Given functions $f\colon X \longrightarrow Y$ and $g\colon Y \longrightarrow Z$ their *composition*, $g \circ f$, is the function

$$g \circ f\colon X \longrightarrow Z,\ x \longmapsto g\big(f(x)\big),$$

In other words, $g \circ f$ is the function defined by

$$\begin{aligned} \operatorname{dom}(g \circ f) &= \operatorname{dom}(f) \\ \operatorname{codom}(g \circ f) &= \operatorname{codom}(g) \\ (g \circ f)(x) &= g\big(f(x)\big) \qquad \text{for all } x \in X \end{aligned}$$

We read $g \circ f$ as "g following f".

We depict this using diagrams by

$$X \xrightarrow{f} Y \xrightarrow{g} Z$$

or

$$\begin{array}{ccc} X & \xrightarrow{f} & Y \\ & & \downarrow{\scriptstyle g} \\ & & Z \end{array}$$

Observation 1.29. The functions g and f can be composed if and only if $\mathrm{dom}(g) = \mathrm{codom}(f)$.

Observation 1.30. It is immediate that $\mathrm{im}(g \circ f) \subseteq \mathrm{im}(g)$.

Equality need not hold in the last of these statements.

Example 1.31. The functions $f\colon \mathbb{R} \longrightarrow \mathbb{R},\ x \longmapsto 1$ and $g\colon \mathbb{R} \longrightarrow \mathbb{R},\ y \longmapsto y$ can be composed, and, clearly, $\mathrm{im}(g \circ f) = \{1\} \neq \mathbb{R} = \mathrm{im}(g)$.

Observation 1.32. The reader has almost certainly made use of the composition of functions, even if (s)he is not aware of it.

Example 1.33. When we evaluate the value of the function

$$h\colon \mathbb{Z} \longrightarrow \mathbb{Z}, \quad x \longmapsto x^2 + 1$$

we normally first square x then add 1 to the result.

This is an application of the composition of functions. For we have used the composition $h = g \circ f$, with

$$\begin{aligned} f\colon \mathbb{Z} &\longrightarrow \mathbb{Z}, \quad x \longmapsto x^2 \\ g\colon \mathbb{Z} &\longrightarrow \mathbb{Z}, \quad y \longmapsto y + 1, \end{aligned}$$

as given any real number, x, we have $h(x) = g\big(f(x)\big)$.

Example 1.34. Another application of the composition of functions provides the restriction of a function to a subset of its domain. For given $A \subseteq X$ and a function $f\colon X \longrightarrow Y$, the restriction $f\,|_A\colon A \longrightarrow Y$ is, in fact, the composition of f and the inclusion of A into X:

$$f\,|_A = f \circ i_A^X\colon A \longrightarrow Y$$

To see this, observe that

(i) $\mathrm{dom}(f\,|_A) = A = \mathrm{dom}(i_A^X) = \mathrm{dom}(f \circ i_A^X)$

(ii) $\mathrm{codom}(f\,|_A) = Y = \mathrm{codom}(f) = \mathrm{codom}(f \circ i_A^X)$

(iii) Given $a \in A$,

$$\begin{aligned} (f \circ i_A^X)(a) &:= f\big(i_A^X(a)\big) \\ &= f(a) \\ &=: f\,|_A\,(a) \end{aligned}$$

Composition of functions being central in mathematics, we investigate some of its properties.

Lemma 1.35. *The composition of functions is* associative*:*

Given functions $g\colon W \longrightarrow X, f\colon X \longrightarrow Y$ and $e\colon Y \longrightarrow Z$, the compositions $(e \circ f) \circ g\colon W \longrightarrow Z$ and $e \circ (f \circ g)\colon W \longrightarrow Z$ are the same function.

Proof. $\operatorname{dom}\big((e \circ f) \circ g\big) = \operatorname{dom}\big(g\big) = \operatorname{dom}\big(f \circ g\big) = \operatorname{dom}\big(e \circ (f \circ g)\big)$.

Similarly, $\operatorname{codom}\big((e \circ f) \circ g\big) = \operatorname{codom}(e \circ f) = \operatorname{codom}(e) = \operatorname{codom}\big(e \circ (f \circ g)\big)$.

It only remains to show that the two functions agree on their common domain, W.

Given $w \in W$,

$$\begin{aligned}\Big((f \circ g) \circ h\Big)(w) &:= (f \circ g)\big(h(w)\big)\\ &:= f\Big(g\big(h(w)\big)\Big)\\ &=: f\Big((g \circ h)(w)\Big)\\ &=: \Big(f \circ (g \circ h)\Big)(w)\end{aligned}$$

□

Lemma 1.36. *Let $f : X \longrightarrow Y$ be a function, then $id_Y \circ f = f$ and $f \circ id_X = f$.*

Proof. Plainly, $\operatorname{dom}(f \circ id_x) = \operatorname{dom}(f)$, $\operatorname{codom}(f \circ id_x) = \operatorname{codom}(f)$.

Similarly, $\operatorname{dom}(fid_Y \circ f) = \operatorname{dom}(f)$, $\operatorname{codom}(id_Y \circ f) = \operatorname{codom}(f)$.

Take $x \in X$. Then

$$\begin{aligned}(id_Y \circ f)(x) &:= id_Y\big(f(x)\big) := f(x)\\ (\circ id_X)(x) &:= f\big(id_X(x)\big) := f(x)\end{aligned}$$

□

We say that the identity functions act as *neutral elements* with respect to composition.

Sometimes the effect of one function can be "undone" by another: If the first assigns y to x, the second allows us to determine x from knowing y.

Composition of functions and the identity functions allow us to formulate this precisely.

Definition 1.37. The function $f\colon X \longrightarrow Y$ is *invertible* if there is a function $g\colon Y \longrightarrow X$ such that

(i) $f \circ g = id_Y$, and

(ii) $g \circ f = id_X$

In other words, f has inverse g if and only if

(a) $f\big(g(y)\big) = y$ for every $y \in Y$, and

(b) $g\big(f(x)\big) = x$ for every $x \in X$.

In such a case, g is said to be the *inverse* of f.

Example 1.38. Let $\mathbb{R}^+ := \{r \in \mathbb{R} \mid r > 0\}$ be the set of all positive real numbers.
The function

$$f\colon \mathbb{R} \longrightarrow \mathbb{R}^+, \quad x \longmapsto e^x$$

has inverse

$$g\colon \mathbb{R}^+ \longrightarrow \mathbb{R}, \quad y \longmapsto \ln y$$

Example 1.39. The function

$$f\colon \mathbb{R} \longrightarrow \mathbb{R}, \quad x \longmapsto x^2$$

has no inverse.
For if $g\colon \mathbb{R} \longrightarrow \mathbb{R}$ is any function,

$$(f \circ g)(-1) = f\Big(g(-1)\Big) > 0$$

Since $(f \circ g)(-1) \neq -1$, $f \circ g \neq id_{\mathbb{R}}$

Observation 1.40. In our definition of invertibility of the function $f\colon X \longrightarrow Y$, the function $g\colon Y \longrightarrow X$ needed to satisfy two conditions. We consider these separately, and introduce terminology tailored to this.

Definition 1.41. The function $g\colon Y \longrightarrow X$ is a *left inverse* of $f\colon X \longrightarrow Y$ if and only if $g \circ f = id_X$ and the function $h\colon Y \longrightarrow X$ is a *right inverse* of $f\colon X \longrightarrow Y$ if and only if $f \circ h = id_Y$.

Theorem 1.42. *If $f\colon X \longrightarrow Y$ has both a left and a right inverse, then these must be the same, and hence f is invertible with a uniquely determined inverse.*

Proof. If $e\colon Y \longrightarrow X$ is left inverse to $f\colon X \longrightarrow Y$ and $g\colon Y \longrightarrow X$ is right inverse, then

$$\begin{aligned}
e &= e \circ id_Y && \text{by Lemma 1.36} \\
&= e \circ (f \circ g) && \text{as } g \text{ is right inverse to } f \\
&= (e \circ f) \circ g && \text{by Lemma 1.35} \\
&= id_X \circ g && \text{as } e \text{ is left inverse to } f \\
&= g && \text{by Lemma 1.36}
\end{aligned}$$

□

The fact that a function, $f\colon X \longrightarrow Y$ cannot have more than one inverse justifies the notation f^{-1} usually used to denote the function $Y \longrightarrow X$ inverse to f, for it is uniquely determined by f whenever f is invertible.

To decide whether the function $f\colon X \longrightarrow Y$ has an inverse does not seem to be an easy task at first glance. If we blindly follow our definition, we would need to try all possible functions from Y to X and see which, if any, satisfy the conditions in the definition. It would be preferable to be able to determine from *intrinsic* properties of f — that is, properties of f alone, without reference to other functions — whether it admits an inverse. We show that such an intrinsic criterion is available. To do so, we introduce some important properties of functions.

Definition 1.43. The function $f\colon X \longrightarrow Y$ is

(i) *1–1* or *injective* or *mono* if and only if it follows from $f(x) = f(u)$ that $x = u$;

(ii) *onto* or *surjective* or *epi* if and only if given any $y \in Y$ there is an $x \in X$ with $f(x) = y$ — in other words $\mathrm{im}(f) = \mathrm{codom}(f)$;

(iii) *1–1 and onto* or *bijective* or *iso* if and only if it is both 1–1 and onto.

Thus a function is injective if and only if it distinguishes different elements of its domain: different elements of its domain are mapped to different elements of its co-domain.

Similarly, a function is surjective if and only if its image coincides with its co-domain.

Example 1.44. We write $\mathbb{R}_0^+$ for $\{x \in \mathbb{R} \mid x \geq 0\}$.

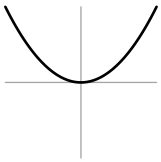
(i) $f\colon \mathbb{R} \longrightarrow \mathbb{R},\ x \longmapsto x^2$ is neither injective nor surjective, as $f(1) = f(-1)$ and there is no $x \in \mathbb{R}$ with $f(x) = -4$.

(ii) $g\colon \mathbb{R} \longrightarrow \mathbb{R}_0^+,\ x \longmapsto x^2$ is not injective, but it is surjective, as $f(1) = f(-1)$ and every non-negative real number can be written as the square of a real number.

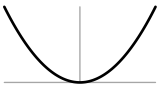
(iii) $h\colon \mathbb{R}_0^+ \longrightarrow \mathbb{R},\ x \longmapsto x^2$ is injective, but it not surjective, as $f(x) = f(u)$ if and only if $x^2 = u^2$ if and only if $u = \pm x$ if and only if $u = x$ as, by definition, $x, u \geq 0$. On the other hand, there is no $x \in \mathbb{R}_0^+$ with $f(x) = -4$.

(iv) $k\colon \mathbb{R}_0^+ \longrightarrow \mathbb{R}_0^+,\ x \longmapsto x^2$ is both injective, and surjective, as should be clear from parts (ii) and (iii).

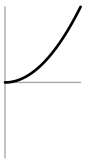
The differences between these functions is illustrated by their respective graphs



Graph of f



Graph of g



Graph of h

Graph of k

Observation 1.45. The notions of injectivity, surjectivity and bijectivity can also be expressed in terms of equations.

Take sets X and Y, and suppose we have a relation between elements of X and elements of Y, which we express by writing

$$y = f(x)$$

whenever $y \in Y$ is related to $x \in X$.

Then f is a function if and only if for each $x \in X$, the equation $y = f(x)$ has one and only one solution $y \in Y$.

If we restrict attention to relations which are functions, then f is injective if and only if for each $y \in Y$, the equation $y = f(x)$ has *at most one* solution $x \in X$, and it is surjective if and only if for each $y \in Y$, the equation $y = f(x)$ has *at least one* solution $x \in X$.

The formulation in terms of equations suggests that a function has an inverse if and only if it is bijective (1–1 and onto).

We next prove that this is, indeed, the case.

Theorem 1.46. *Given a non-empty set X, a function $f\colon X \longrightarrow Y$ has*

(i) a left inverse if and only if it is injective (or 1–1),

(ii) a right inverse if and only if it is surjective (or onto) and

(iii) an inverse if and only if it is bijective.

Proof. (i) Suppose that $f\colon X \longrightarrow Y$ has a left inverse $g\colon Y \longrightarrow X$.

To see that f must then be injective (that is 1–1), suppose that $f(x) = f(u)$. Then

$$\begin{aligned} x &= id_X(x) \\ &= (g \circ f)(x) \\ &= g\big(f(x)\big) \\ &= g\big(f(u)\big) \\ &= (g \circ f)(u) \\ &= id_X(u) \\ &= u \end{aligned}$$

For the converse, suppose that $f\colon X \longrightarrow Y$ is injective.

Choose $x_0 \in X$ and consider

$$g\colon Y \longrightarrow X, \quad y \longmapsto \begin{cases} x & \text{if } \ y = f(x) \\ x_0 & \text{otherwise} \end{cases}$$

That $g \circ f = id_X$ is immediate from the definition of g.

It only remains to show that g so defined is, in fact a function.

For this, g must assign to each $y \in Y$ a uniquely determined $x \in X$.

It is immediate from the definition of g that the only possible obstruction is that g might assign more than one element of X to some element y of Y.

By the definition of g, this could only happen when $y \in \mathrm{im}(f)$, that is, when $y = f(x) = f(u)$.

But then $x = u$, since f is injective.

Hence, g is, indeed, a function.

(ii) Suppose that $f\colon X \longrightarrow Y$ has a right inverse $g\colon Y \longrightarrow X$.

To see that f must be surjective, take $y \in Y$ and put $x := g(y)$. Then

$$\begin{aligned} f(x) &= f\big(g(y)\big) \\ &= (f \circ g)(y) \\ &= id_Y(y) \\ &= y \end{aligned}$$

For the converse, suppose that $f\colon X \longrightarrow Y$ is surjective.

Define $g\colon Y \longrightarrow X$ by choosing for each $y \in Y$ a specific element, x_y, of X with $f(x_{y)} = x$.

There is always at least one such an element of X is because f is surjective.[1]

This g is a function, because we have chosen for each $y \in Y$ a single corresponding element of X.

Since, by the definition of g, $(f \circ g)(y) = f\big(g(y)\big) = f(x_y) = y$, for each $y \in Y$, $f \circ g = id_Y$.

(iii) This follows from Theorem 1.42 on page 11 and parts (i) and (ii) here. □

[1] This requires the *Axiom of Choice* in the general case. In fact, it is equivalent to the Axiom of Choice, but we do not pursue such matters here.

Example 1.47. Theorem 1.46 on page 12 illustrates one of the ways in which two functions can have the same domain and agree everywhere without being the same function.

Let X be a non-empty proper subset of the set Y, so that $X \subset Y$. Then the two functions

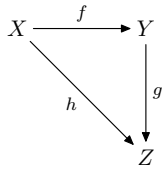
$$id_X \colon X \longrightarrow X, \quad x \longmapsto x$$
$$i_X^Y \colon X \longrightarrow Y \quad x \longmapsto x$$

share a common domain and agree at every point, so that they have the same range: X. But they cannot be the same function. For whereas id_X is invertible — it is its own inverse — Theorem 1.46 on page 12 tells us that i_X^Y cannot be invertible, since it fails to be surjective, whence it has no right inverse.

We have seen how to represent functions using a diagram. We extend this to represent several functions simultaneously.

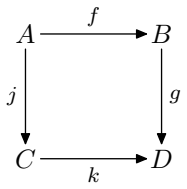
Definition 1.48. A diagram *commutes* whenever any two paths from one fixed vertex to any other fixed vertex traced by following consecutive arrows in their given directions represent the same function.

Example 1.49. The diagram



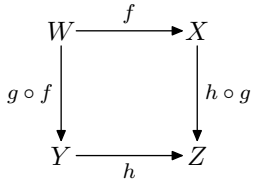
commutes whenever $h = g \circ f$, in other words, when the composition $g \circ f$ coincides with h.

Similarly, the diagram

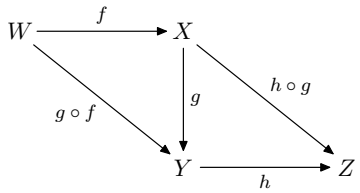


commutes when $k \circ j = g \circ f$, in other words, when the compositions $g \circ f$ and $k \circ j$ coincide.

Example 1.50. That the composition of functions is associative — that is, $(h \circ g) \circ f = h \circ (g \circ h)$ for functions $f \colon W \longrightarrow X$, $g \colon X \longrightarrow Y$ and $h \colon Y \longrightarrow Z$ — is expressed by the commutativity of



or, equivalently, of



1.5 Equivalence Relations and Partitions

Sometimest two distinct objects are indistinguishable, or that their differences are irrelevant , for some purpose: they are equivalent for that purpose. We define this notion formally.

A *relation* between the elements of the set X and those of Y can be represented by the subset of $X \times Y$ comprising those pairs (x, y) $(x \in X, y \in Y)$ such that x stands in the relation R to y. We often write xRy to denote this.

Example 1.51. An example is provided by the telephone book. Here we regard X as the set of all subscribers, and Y as all telephone numbers.

If Y happens to coincide with X, we speak of a *binary relation on* X.

Definition 1.52. An *equivalence relation* on X, $\sim$, is a binary relation on X, which is reflexive, symmetric and transitive. That it to say, for all $x, y, z \in X$, we have

Reflexiveness $x \sim x$

Symmetry $x \sim y$ if and only if $y \sim x$.

Transitivity If $x \sim y$ and $y \sim z$, then $x \sim z$.

Given an equivalence relation $\sim$ on X, and $x \in X$, we define

$$[x] := \{t \in X \mid x \sim t\},$$

and call it the *equivalence class* of x. We call any element z of $[x]$ a *representative* of $[x]$.
Finally, we let $X/\!\sim$ denote the set of all such equivalence classes, so that

$$X/\!\sim := \{[x] \mid x \in X\}.$$

We then have a function, the *natural map* or the *quotient map*

$$\eta\colon X \longrightarrow X/\!\sim, \quad x \longmapsto [x].$$

Example 1.53. Let X be the set of all Australian citizens registered to vote in federal elections.

$$x \sim y \quad \text{if and only if } x \text{ and } y \text{ are enrolled in the same federal electorate}$$

defines an equivalence relation on X, and the equivalence classes are the individual electorates.

The above construction enjoys a *universal property*:

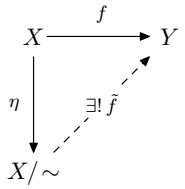
Theorem 1.54. *Let $\sim$ be an equivalence relation on the set X.*
Given any set Y and any function $f\colon X \longrightarrow Y$ with the property that $f(x) = f(u)$ whenever $x \sim u$, there is a unique function

$$\tilde{f}\colon X/\!\sim \longrightarrow Y$$

such that $f = \tilde{f} \circ \eta$, that is, $\hat{f}([x]) = f(x)$ for all $[x] \in X/\!\sim$.

Proof. The proof is left as an exercise. □

This theorem can be summarised by the commutative diagram



Example 1.55. Given the function $f \circ X \longrightarrow Y$ define the relation $\sim$ on X by

$$x \sim u \qquad \text{if and only if} \qquad f(x) = f(u)$$

It is easy to verify directly that $\sim$ is an equivalence relation.

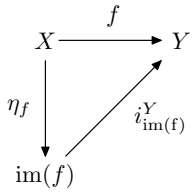
We may identify each equivalence class $[x]$ with the element $f(x)$ of Y, for these uniquely determine each other. This has the effect of identifying ${}^{X}/_{\sim}$ with $\{y \in Y \mid y = f(x) \text{ for some } x \in X\}$, that is, the range of f, $\mathrm{im}(f)$. The natural projection $\eta\colon X \longrightarrow {}^{X}/_{\sim}$ then induces the function

$$\eta_f\colon X \longrightarrow \mathrm{im}(f), \quad x \longmapsto f(x)$$

If we apply the universal property of the quotient construction to the function $f\colon X \longrightarrow Y$, we obtain a uniquely determined function $\tilde{f}\colon {}^{X}/_{\sim} \longrightarrow Y$ with $f = \tilde{f} \circ \eta$.

Using the identifications introduced, this becomes a uniquely determined function $\tilde{f}^{\sharp}\colon \mathrm{im}(f) \longrightarrow Y$ with $f = \tilde{f}^{\sharp} \circ \eta_f$.

As the inclusion function $i^{Y}_{\mathrm{im}(f)}\colon \mathrm{im}(f) \longrightarrow Y$, shares this property, we have $\tilde{f}^{\sharp} = i^{Y}_{\mathrm{im}(f)}$, and we obtain the commutative diagram



Plainly, η_f is surjective (epi) and $i^{Y}_{\mathrm{im}(f)}$ is injective (mono).

What we have shown is that every function $f\colon X \longrightarrow Y$ can be expressed as a mono (injective function) following an epi (surjective function). We summarise this in our next theorem, whose statement requires a definition.

Definition 1.56. A *mono-epi factorisation* of the function $f\colon X \longrightarrow Y$ consists of a mono (injective function), $m\colon W \longrightarrow Y$, and an epi (surjective function), $e\colon W \longrightarrow W$ with $f = m \circ e$

Theorem 1.57. *Every function has a mon-epi factorisation.*

Another important notion is that of a *partition* of a set.

Definition 1.58. A *partition* of the set X is a collection of disjoint non-empty subsets of X, $\{X_\lambda \mid \lambda \in \Lambda\}$, whose union is X. Thus $\{X_\lambda \mid \lambda \in \Lambda\}$ is a partition of X if and only if

1. $\emptyset \subset X_\lambda \subseteq X$ for each $\lambda \in \Lambda$
2. $X_\lambda \cap X_\mu = \emptyset$ whenever $\lambda \neq \mu$

3. $X = \bigcup_{\lambda \in \Lambda} X_\lambda$

The notions of an equivalence relation on a set and of a partition of a set may appear to be unrelated, but that is not the case. Rather, they are two sides of the same coin, as the next theorem shows.

Theorem 1.59. *Every equivalence relation on the set X determines a unique partition of X, and conversely.*

Proof. We outline a proof, leaving the details as an exercise for the reader.

Given the equivalence relation $\sim$ on X, the equivalence classes form a partition of X, that is every $x \in X$ belongs to some equivalence class, and if $[x] \cap [u] \neq \emptyset$, then $[x] = [u]$.

If $\{X_\lambda \mid \lambda \in \Lambda\}$ is a partition of X, then

$$x \sim u \text{ if and only if } x, u \in X_\lambda \text{ for some } \lambda \in \Lambda$$

defines an equivalence relation on X

Now show that if we start with an equivalence relation, construct the associated partition, then the associated equivalence relation is the original one.

Finally, show that if we start with a partition, define the associated equivalence relation, then the associated partition is the original one. □

1.6 Exercises

Exercise 1.1. Given the function $f\colon X \longrightarrow Y$ and subsets A of X and B of Y, prove the following statements.

(i) $A \subseteq f^{-1}(f(A))$.

(ii) $f\left(f^{-1}(B)\right) \subseteq B$.

(iii) In general, equality need not hold in either (i) or (ii).

(iv) $G = f^{-1}(f(G))$ for every subset G of X if and only if f is injective (1–1).

(v) $f\left(f^{-1}(H)\right) = H$ for every subset H of Y if and only if f is surjective (onto).

Exercise 1.2. Take functions $f\colon X \longrightarrow Y$ and $g\colon \longrightarrow Z$. Prove the following statements.

(a) If f and g are both injective, then so is $g \circ f$.

(b) If f and g are both surjective, then so is $g \circ f$.

(c) If $g \circ f$ is injective, then so is f, but not necessarily g.

(d) If $g \circ f$ is surjective, then so is g, but not necessarily f.

(e) If f and g are bijective, so is $g \circ f$.

(f) If $g \circ f$ is bijective, then neither f nor g need be bijective.

Exercise 1.3. Let A, B, C and D be sets. Determine the relationships between

(i) $(A \times C) \cap (B \times D)$ and $(A \cap B) \times (C \cap D)$;

(ii) $(A \times C) \cup (B \times D)$ and $(A \cup B) \times (C \cup D)$.

Exercise 1.4. Given a function $f\colon X \longrightarrow Y$ and subsets G, H of Y, prove the following statements.

(i) $f^{-1}(G \cap H) = f^{-1}(G) \cap f^{-1}(H)$.

(ii) $f^{-1}(G \cup H) = f^{-1}(G) \cup f^{-1}(H)$.

(iii) $f^{-1}(G \setminus H) = f^{-1}(G) \setminus f^{-1}(H)$.

(iv) $f^{-1}(Y \setminus G) = X \setminus f^{-1}(G)$.

Exercise 1.5. Given a function $f\colon X \longrightarrow Y$ and subsets A, B of X, find the relationship between the following pairs of subsets of Y.

(i) $f(A \cap B)$ and $f(A) \cap f(B)$.

(ii) $f(A \cup B)$ and $f(A) \cup f(B)$.

(iii) $f(A \setminus B)$ and $f(A) \setminus f(B)$.

(iv) $f(X \setminus A)$ and $Y \setminus f(A)$.

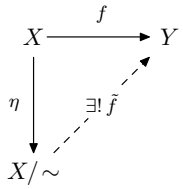
Exercise 1.6. Let $\sim$ be an equivalence relation on the set X.

Prove that if Y is any set and if $f\colon X \longrightarrow Y$ is any function with the property that $f(x) = f(u)$ whenever $x \sim u$, then there is a unique function

$$\hat{f}\colon {}^{X}\!/_{\sim} \longrightarrow Y$$

such that $f = \hat{f} \circ \eta$, that is, $\hat{f}([x]) = f(x)$ for all $[x] \in {}^{X}\!/_{\sim}$.

This is equivalent to the statement that the following diagram commutes.



In mathematics the art of proposing a question must be held of higher value than solving it.

Georg Cantor

Chapter 2

Introductory Examples

We consider several problems familiar to the reader from earlier courses. The rest of these notes comprises a thorough mathematical analysis and systematic solution of these problems, developing the theory which is needed to explain the methods and solutions. The reader should bear these examples in mind while working through the rest of these notes.

2.1 Solving Systems of Linear Equations

We begin with systems of simultaneous linear equations.

We first investigate the equation

$$ax = b \tag{2.1}$$

where a, b are given real numbers, and we wish to find all real numbers, x, satisfying Equation 2.1.

There are two cases to consider: (i) $a = 0$ and (ii) $a \neq 0$.

(i) $a = 0$**:** When $a = 0$, $ax = 0x = 0$ for every real number x.

Consequently, (2.1) has no solution unless $b = 0$.

On the other hand, *every* real number x is a solution when $b = 0$.

(ii) $a \neq 0$**:** When $a \neq 0$, we know from arithmetic that $x = \dfrac{b}{a}$ is the one and only solution.

Summarising, Equation (2.1) has

- a unique solution, when $a \neq 0$, one for each b;
- no solution, when $a = 0$ and $b \neq 0$;
- infinitely many solutions, when $a = 0$ and $b = 0$.

We next consider the system of equations

$$ax + by = e \tag{2.2a}$$
$$cx + dy = f \tag{2.2b}$$

A solution of (2.2) consists of a pair of real numbers (x, y) such that both (2.2a) and (2.2b) are satisfied.

Our approach to finding all solutions of (2.2) is to try to replace (2.2a) and (2.2b) by equations of the form (2.1) which, taken together, have the same solutions as (2.2).

If we multiply (2.2a) by d and subtract b times (2.2b), as well as subtracting c times (2.2b) from a times (2.2a) we obtain

$$(ad - bc)x = (ed - bf) \tag{2.3a}$$
$$(ad - bc)y = (af - ec) \tag{2.3b}$$

Each of these equations is of the same form as (2.1).

From our analysis of (2.1), if $ad - bc \neq 0$, we obtain the unique solution

$$\left(\frac{ed - bf}{ad - bc}, \frac{af - ec}{ad - bc} \right) \tag{2.4}$$

Direct substitution verifies that (2.4) solves our system of equations.

If, on the other hand, $ad - bc = 0$, then there is no solution whatsoever if either $ed - bf \neq 0$ or $af - ec \neq 0$, and every pair of real numbers (x, y) is a solution if both $ed - bf = 0$ and $af - ec = 0$.

It follows from our derivation of (2.3) from (2.2) that any solution of (2.2) is also a solution of (2.3). Thus, if $ad - bc = 0$ and either $ed - bf \neq 0$ or $af - ec \neq 0$, then (2.2) has no solution, for then (2.3) has none.

The situation is more delicate when $ad - bc = ed - bf = af - ec = 0$, for it is then possible that some solutions of (2.3) are not solutions of (2.2), as the next example shows.

Example 2.1. For $a = 1, b = -1, c = d = e = f = 0$, (2.2) becomes

$$x - y = 0 \tag{2.5a}$$
$$0x + 0y = 0 \tag{2.5b}$$

and, plainly, the complete set of solutions is the set of all pairs of real numbers of the form (x, x). But (2.3) becomes

$$0x = 0 \tag{2.6a}$$
$$0y = 0 \tag{2.6b}$$

which is solved by any pair of real numbers (x, y).

In particular, $(1, 0)$ solves (2.6) without solving (2.5).

Observation 2.2. Whether (2.2) has a unique solution is determined by $ad - bc$. For this reason $ad - bc$ is known as the *determinant* of the system of equations (2.2).

We continue our investigation of (2.2).

We claim to have found all possible solutions.

But, how can we be sure?

We address this question.

Suppose that (x_1, y_1) and (x_2, y_2) both solve (2.2).

Then

$$a(x_1 - x_2) + b(y_1 - y_2) = (ax_1 + by_1) - (ax_2 + by_2) = e - e = 0 \tag{2.7a}$$
$$c(x_1 - x_2) + d(y_1 - y_2) = (cx_1 + dy_1) - (cx_2 + dy_2) = f - f = 0 \tag{2.7b}$$

Thus, any two solutions of (2.2) differ by a solution of

$$ax + by = 0 \tag{2.8a}$$
$$cx + dy = 0 \tag{2.8b}$$

This means that once we have found one solution, (x_s, y_s) of (2.2), we can find all other solutions by adding the various solutions of (2.8).

In other words, the general solution of (2.2) can be found by adding to (x_h, y_h), the general solution of (2.8), any one solution, (x_s, y_s), of (2.2).

In particular, (2.2) cannot have a unique solution unless (2.8) has a unique solution.

We therefore investigate (2.8).

Unlike (2.2), the system of equations (2.8) always has at least one solution, namely the *trivial* solution $x = 0, y = 0$.

Suppose that (x_1, y_1) and (x_2, y_2) are solutions of (2.8), and that λ, μ are real numbers. Then

$$a(\lambda x_1 + \mu x_2) + b(\lambda y_1 + \mu y_2) = \lambda(ax_1 + by_1) + \mu(ax_2 + by_2) = \lambda 0 + \mu 0 = 0$$
$$c(\lambda x_1 + \mu x_2) + d(\lambda y_1 + \mu y_2) = \lambda(cx_1 + dy_1) + \mu(cx_2 + dy_2) = \lambda 0 + \mu 0 = 0,$$

so that $(\lambda x_1 + \mu x_2, \lambda y_1 + \mu y_2)$ is also a solution of (2.8).

If we define an "addition" of solutions of (2.8) by

$$(x_1, y_1) \boxplus (x_2, y_2) := (x_1 + x_2, y_1 + y_2),$$

and a "multiplication" by real numbers of solutions of (2.8) by

$$\lambda \boxdot (x, y) := (\lambda x, \lambda y)$$

then "adding" any two solutions of (2.8) yields a solution of (2.8), and "multiplying" a solution of (2.8) by a real number yields a solution of (2.8).

Systems of equations like (2.8) are called *homogeneous*. They are, of course, just the special case of (2.2) where $e = f = 0$. In particular, if in (2.2) either $e \neq 0$ or $f \neq 0$, then we call the corresponding system (2.8) the *associated homogeneous system*.

This analysis can be extended to larger systems of simultaneous linear equations.

A system of m linear equations in the n variables $x_1, \ldots, x_n$ is a set of m equations

$$\begin{array}{ccccccc} a_{11}x_1 & + & \cdots & + & a_{1n}x_n & = & b_1 \\ \vdots & & & & \vdots & & \vdots \\ a_{m1}x_1 & + & \cdots & + & a_{mn}x_n & = & b_m \end{array} \tag{$*$}$$

where $a_{11}, \ldots, a_{mn}, b_1, \ldots, b_m$ are fixed.

A solution is an n-tuple $(r_1, \ldots, r_n)$ such that each of the m equations holds when each x_j is replaced by r_j $(1 \leq j \leq n)$. The general solution is again given by any one specific solution plus the general solution of the associated homogeneous system of equations.

The example of systems of simultaneous linear equations, and the features just highlighted, provides one of the motivations for these notes: Elementary linear algebra was first the systematic study of such systems of equations, their solutions and the transformations these admit.

Matrices were introduced in the course of the systematic study of such equations, allowing simpler, more efficient notation. Matrix algebra then systematised the computations. For example, the system of equations $(*)$ has matrix representation[1]

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

The solutions of homogeneous systems of linear equations are elementary examples of *vector spaces*. Vector spaces are the objects of study in linear algebra. We present a rigorous definition of the notion of a vector space in Chapter 3 on page 31.

While its application to the systematic study of simultaneous linear equations would be sufficient, by itself, to justify the study of vector spaces, it is an astounding fact that there are numerous other examples and applications.

- The vectors of physics, such as force, also provide examples, as the language suggests: the "sum" of two forces acting simultaneously is their resultant force, and "multiplication" of a force by a real number corresponds to scaling the force.
- Binary computer code is another example of a vector space, a point of view which finds application in theoretical computer science.
- Solutions to specific systems of differential equations also form vector spaces.
- Vector spaces also appear in number theory in several places, including the study of field extensions, and form the basis from which the important algebraic notion of *module* has been abstracted.
- Finally, vector spaces, particularly inner product spaces, are central to the study of statistics and geometry.

It is not immediately apparent that the examples listed have much in common. This explains and justifies why we need to develop a general theory of vector spaces: we need to account for the common features of our diverse examples, without being distracted by the special features of any specific example.

Before launching into the formal study of linear algebra, we illustrate how linear algebra can be applied to solving *linear difference equations with constant coefficients*, by writing such difference equations in terms of *matrices*. This not only provides an application of linear algebra and its techniques, but also provides motivation for deeper investigation.

2.2 Linear Difference Equations with Constant Coefficients

We begin by recalling the definition of a linear difference equation over $\mathbb{R}$, the set of all real numbers.

[1] The later chapters contain a thorough explanation of this.

Definition 2.3. Let $(x_n)_{x\in\mathbb{N}}$ be a sequence of real numbers. A *linear difference equation of degree k with constant coefficients* is an equation of the form

$$x_{n+k} + a_{k-1}x_{n+k-1} + \cdots + a_1x_{n+1} + a_0x_n = g(n), \tag{2.10}$$

where each $a_j \in \mathbb{R}$ and $g(n)$ is a function of $n \in \mathbb{N}$.

Example 2.4. Consider the difference equation

$$x_{n+2} - 4x_{n+1} + 3x_n = 0. \tag{2.11}$$

This can be rewritten as

$$x_{n+2} = 4x_{n+1} - 3x_n$$

which has precisely the same solutions as the system of simultaneous equations

$$x_{n+2} = 4x_{n+1} - 3x_n \tag{2.12a}$$
$$x_{n+1} = x_{n+1} \tag{2.12b}$$

Using ($\diamondsuit$), this system of equations is represented by the matrix equation

$$\begin{bmatrix} x_{n+2} \\ x_{n+1} \end{bmatrix} = \begin{bmatrix} 4 & -3 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_{n+1} \\ x_n \end{bmatrix}$$

As this holds for all $n \in \mathbb{N}$, we see that for $n = 0$

$$\begin{bmatrix} x_2 \\ x_1 \end{bmatrix} = \begin{bmatrix} 4 & -3 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_0 \end{bmatrix}$$

For $n = 1$, we have

$$\begin{aligned} \begin{bmatrix} x_3 \\ x_2 \end{bmatrix} &= \begin{bmatrix} 4 & -3 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_2 \\ x_1 \end{bmatrix} \\ &= \begin{bmatrix} 4 & -3 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 4 & -3 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_0 \end{bmatrix} \\ &= \begin{bmatrix} 4 & -3 \\ 1 & 0 \end{bmatrix}^2 \begin{bmatrix} x_1 \\ x_0 \end{bmatrix} \end{aligned}$$

We see, by induction, that for all $n \in \mathbb{N}$,

$$\begin{bmatrix} x_{n+1} \\ x_n \end{bmatrix} = \begin{bmatrix} 4 & -3 \\ 1 & 0 \end{bmatrix}^n \begin{bmatrix} x_1 \\ x_0 \end{bmatrix} \tag{2.13}$$

Example 2.5. Consider the difference equation

$$x_{n+2} - 4x_{n+1} + 4x_n = 0, \tag{2.14}$$

This can be rewritten as

$$x_{n+2} = 4x_{n+1} - 4x_n$$

which has precisely the same solutions as the system of simultaneous equations

$$x_{n+2} = 4x_{n+1} - 4x_n \tag{2.15a}$$
$$x_{n+1} = x_{n+1} \tag{2.15b}$$

This system of equations is represented by the matrix equation

$$\begin{bmatrix} x_{n+2} \\ x_{n+1} \end{bmatrix} = \begin{bmatrix} 4 & -4 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_{n+1} \\ x_n \end{bmatrix},$$

from which we deduce, by induction, that

$$\begin{bmatrix} x_{n+1} \\ x_n \end{bmatrix} = \begin{bmatrix} 4 & -4 \\ 1 & 0 \end{bmatrix}^n \begin{bmatrix} x_1 \\ x_0 \end{bmatrix}. \tag{2.16}$$

Example 2.6. Consider the difference equation

$$x_{n+2} - 4x_{n+1} + 5x_n = 0, \tag{2.17}$$

This can be rewritten as

$$x_{n+2} = 4x_{n+1} - 5x_n$$

which has precisely the same solutions as the system of simultaneous equations

$$x_{n+2} = 4x_{n+1} - 5x_n \tag{2.18a}$$
$$x_{n+1} = x_{n+1} \tag{2.18b}$$

This system of equations is represented by the matrix equation

$$\begin{bmatrix} x_{n+2} \\ x_{n+1} \end{bmatrix} = \begin{bmatrix} 4 & -5 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_{n+1} \\ x_n \end{bmatrix},$$

from which we deduce, by induction, that

$$\begin{bmatrix} x_{n+1} \\ x_n \end{bmatrix} = \begin{bmatrix} 4 & -5 \\ 1 & 0 \end{bmatrix}^n \begin{bmatrix} x_1 \\ x_0 \end{bmatrix}. \tag{2.19}$$

Thus, solving these difference equations has been reduced to "merely" computing, respectively

$$\begin{bmatrix} 4 & -3 \\ 1 & 0 \end{bmatrix}^n, \quad \begin{bmatrix} 4 & -4 \\ 1 & 0 \end{bmatrix}^n \quad \text{and} \quad \begin{bmatrix} 4 & -5 \\ 1 & 0 \end{bmatrix}^n$$

Observation 2.7. These examples can be generalised to solve Equation (2.10) whenever $g(n) = 0$ for all $n \in \mathbb{N}$. We do not pursue this greater generality here, since our aim here is merely to present some concrete examples to assist the reader and to motivate the theory we develop. As the general case is an application of this theory, nothing will be lost.

The reader has met in his her earlier studies, the following explicit formula for computing the product of two matrices:

The product of the $m \times n$ matrix $\underline{\mathbf{A}} = [a_{ij}]_{m\times n}$ and the $n \times p$ matrix $\underline{\mathbf{B}} = [b_{jh}]_{n\times p}$ is the $m \times p$ matrix $\underline{\mathbf{A}}\,\underline{\mathbf{B}} = [c_{ih}]_{m\times p}$, where

$$c_{ih} := \sum_{j=1}^{n} a_{ij} b_{jh}$$

This provides an inductive formula for $\underline{\mathbf{A}}^n$ $(n \in \mathbb{N})$:
Writing $\underline{\mathbf{A}}^n := [a_{ij}^{(n)}]_{k\times k}$,

1. $a_{ij}^{(0)} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$ as, by convention, $\underline{\mathbf{A}}^0$ is the $k \times k$ identity matrix.

2. $a_{ij}^{(n+1)} = \sum_{h=1}^{k} a_{ih} a_{hj}^{(n)}$

While it is comforting to have a recursive formula and so be able to use a programmable calculator or computer for the actual calculation, it is easy to see that this is neither an efficient nor an insightful way to proceed.

Example 2.8. Try to compute the onethousandth power of $\begin{bmatrix} 1 & 2 \\ 2 & 5 \end{bmatrix}$ using this procedure.

Plainly all the coefficients are positive integers and $a_{ij}^{(n+1)} > a_{ij}^{(n)}$. But little more can be said!

Even when you have used the inductive formula to complete such a calculation, you are unlikely to guess any formula for calculating the $a_{ij}^{(n)}$'s directly for $n > 2$. On the other hand, using the theory developed during this course, you will be able to see that for the matrices above, we have the explicit formulæ below.

Example 2.4 Continued. Since

$$\begin{bmatrix} 4 & -3 \\ 1 & 0 \end{bmatrix}^n = \frac{1}{2}\begin{bmatrix} 3 & 1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}^n\begin{bmatrix} 1 & -1 \\ -1 & 3 \end{bmatrix}$$
$$= \frac{1}{2}\begin{bmatrix} 3 & 1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 3^n & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & -1 \\ -1 & 3 \end{bmatrix},$$

$$a_{11}^{(n)} = \frac{1}{2}(3^{n+1} - 1) \qquad a_{12}^{(n)} = \frac{1}{2}(3 - 3^{n+1})$$
$$a_{21}^{(n)} = \frac{1}{2}(3^n - 1) \qquad a_{22}^{(n)} = \frac{1}{2}(3 - 3^n).$$

Example 2.5 Continued. Since

$$\begin{bmatrix} 4 & -4 \\ 1 & 0 \end{bmatrix}^n = \begin{bmatrix} 2 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}^n\begin{bmatrix} 0 & 1 \\ 1 & -2 \end{bmatrix}$$
$$= \begin{bmatrix} 2 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 2^n & n2^{n-1} \\ 0 & 2^n \end{bmatrix}\begin{bmatrix} 0 & 1 \\ 1 & -2 \end{bmatrix},$$

$$a_{11}^{(n)} = (n+1)2^n \qquad a_{12}^{(n)} = -n2^{n+1}$$
$$a_{21}^{(n)} = n2^{n-1} \qquad a_{22}^{(n)} = (1-n)2^n.$$

Example 2.6 Continued. Since

$$\begin{bmatrix} 4 & -5 \\ 1 & 0 \end{bmatrix}^n = \frac{-i}{2}\begin{bmatrix} 2+i & 2-i \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 2+i & 0 \\ 0 & 2-i \end{bmatrix}^n\begin{bmatrix} 1 & -2+i \\ -1 & 2+i \end{bmatrix}$$

$$= \frac{-i}{2}\begin{bmatrix} 2+i & 2-i \\ 1 & 1 \end{bmatrix}\begin{bmatrix} (2+i)^n & 0 \\ 0 & (2-i)^n \end{bmatrix}\begin{bmatrix} 1 & -2+i \\ -1 & 2+i \end{bmatrix}$$

where $i^2 = -1$

$$a_{11}^{(n)} = \frac{-i}{2}\left((2+i)^{n+1} - (2-i)^{n+1}\right) \qquad a_{12}^{(n)} = \frac{5i}{2}\left((2+i)^n - (2-i)^n\right)$$

$$a_{21}^{(n)} = \frac{-i}{2}\left((2+i)^n - (2-i)^n\right) \qquad a_{22}^{(n)} = \frac{5i}{2}\left((2+i)^{n-1} - (2-i)^{n-1}\right)$$

Example 2.8 Continued. Since

$$\begin{bmatrix} 1 & 2 \\ 2 & 5 \end{bmatrix}^n = \frac{\sqrt{2}-1}{2\sqrt{2}}\begin{bmatrix} \sqrt{2}+1 & 1 \\ -1 & \sqrt{2}+1 \end{bmatrix}\begin{bmatrix} \left(3-2\sqrt{2}\right)^n & 0 \\ 0 & \left(3+2\sqrt{2}\right)^n \end{bmatrix}\begin{bmatrix} \sqrt{2}+1 & -1 \\ 1 & \sqrt{2}+1 \end{bmatrix},$$

$$a_{11}^{(n)} = \frac{1}{2\sqrt{2}}\left[\left(\sqrt{2}+1\right)\left(3-2\sqrt{2}\right)^n + \left(\sqrt{2}-1\right)\left(3+2\sqrt{2}\right)^n\right]$$

$$a_{12}^{(n)} = \frac{1}{2\sqrt{2}}\left[\left(3+2\sqrt{2}\right)^n - \left(3-2\sqrt{2}\right)^n\right]$$

$$a_{21}^{(n)} = \frac{1}{2\sqrt{2}}\left[\left(3+2\sqrt{2}\right)^n - \left(3-2\sqrt{2}\right)^n\right]$$

$$a_{22}^{(n)} = \frac{1}{2\sqrt{2}}\left[\left(\sqrt{2}-1\right)\left(3-2\sqrt{2}\right)^n + \left(\sqrt{2}+1\right)\left(3+2\sqrt{2}\right)^n\right]$$

Observation 2.9. It is difficult to envisage how anyone could have guessed any of the four sets of formulæ we have just presented.

On the other hand, it is just a matter of simple direct calculation to verify the formulæ.

This is common in mathematics. Given a prospective solution to a problem, testing it is often straightforward. Finding a reasonable candidate to test is frequently much more difficult, often requiring a more theoretical approach.

In the case here, the reader is likely to be perplexed by how the various matrices were "pulled out of the air" to find the explicit formulaæ. For example, how could anyone come up with the matrices

$$\begin{bmatrix} 2+i & 2-i \\ 1 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & -2+i \\ -1 & 2+i \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 2+i & 0 \\ 0 & 2-i \end{bmatrix}$$

in connection with the matrix $\begin{bmatrix} 4 & -5 \\ 1 & 0 \end{bmatrix}$?

The reader may usefully regard the rest of these notes as explaining how the various matrices above can be found. It will also become apparent that the matrices we provided are not the only ones which provide the explicit formulaæ, and how the theory needed for such explanation has much broader application.

Observation 2.10. Each of the four matrices, $\underline{\mathbf{A}}$, introduced in Examples 2.4 on page 23 to 2.8 on the preceding page had integer coefficients. This, and the definition of matrix multiplication, make it obvious that every coefficients of each of the matrices $\underline{\mathbf{A}}^n$ is an integer.

Furthermore, it is obvious that in Example 2.8 on the previous page, where

$$\underline{\mathbf{A}} = \begin{bmatrix} 1 & 2 \\ 2 & 5 \end{bmatrix}$$

each sequence, $\left(a_{ij}^{(n)}\right)_{n\in\mathbb{N}}$, is a monotonically increasing sequence of positive integers.

Yet, when we turn to the explicit formulæ for the coefficients of the matrices $\underline{\mathbf{A}}^n$, it is only in Example 2.5 on page 23 that it is immediately clear that they must all be integers.

In Example 2.4 on page 23, the observation that the sum of two odd integers must be even is enough to show from the explicit formulæ that all the coefficients must, indeed, be integers.

In Example 2.6 on page 24, it is hard to see from the explicit formulæ that all the coefficients are real numbers.

In Examples 2.8 on page 25, the explicit formulæ for $a_{ij}^{(n)}$ as function of n contain negative numbers, fractions and irrational numbers, making it hard to see that the coefficients must be even rational and/or that they must all be positive.

This illustrates a recurring theme in mathematics: In order to solve problems which are simple to express, it is frequently necessary to go beyond the terms in which the problem is expressed, to a deeper or more abstract level, in order to find a solution. We seem to have made the problems more complicated. But this has made them easier to solve!

Perhaps the most striking recent example of this is Andrew Wiles' proof in 1995 of Fermat's Last Theorem:

> **The integer equation $\mathbf{x^n + y^n = z^n}$, with $\mathbf{x, y, z \neq 0}$, has no solution if $\mathbf{n > 2}$.**

This was enunciated in 1657, but no proof was known until Andrew Wiles' work in the 1990s! While the problem is simple to express and understand — a student in the first year of high school can begin to work on it — its proof by Andrew Wiles depends upon results drawn from algebraic topology, algebraic geometry and other fields of mathematics.

Observation 2.11. The matrices in Examples 2.4 on page 23 to Examples 2.6 on page 24 differ only in a single coefficient — the 3 in Example 2.4 on page 23 is replaced by a 4 in Example 2.5 on page 23 and a 5 in Example 2.6 on page 24.

It is easy to jump to the conclusion that the explicit formulæ for the coefficients of the n^{th} powers of the matrices must also be similar.

We have shown that this is far from the case, providing another important reason for rigorous theoretical analysis: Superficially similar problems can have radically different solutions.

The theory developed in these notes provides a uniform analysis of these examples, explaining why and when such similar problems have different solutions.

Observation 2.12. The reader is unlikely to have guessed any of the explicit formulaæ we provided. Indeed, the reader is probably perplexed at how anyone could have come up with the matrices we introduced. By carefully working through these notes, the reader will see and understand how these matrices arise naturally from the initial ones.

2.3 Exercises

These exercises revise material from pre-requisite courses. Additional thought may be needed.

Exercise 2.1. Solve the following system of equations, where the solutions are to be real numbers.

$$\begin{array}{rcrcrcr} x & + & 7y & + & 4z & = & 21 \\ 3x & - & 6y & + & 5z & = & 2 \\ 5x & + & y & - & 3z & = & 14 \end{array}$$

Exercise 2.2. Let $f, g\colon \mathbb{R} \longrightarrow \mathbb{R}$ be real valued functions of the real variable x such that both $y = f(x)$ and $y = g(x)$ satisfy the differential equation

$$\frac{d^2y}{dx^2} - 2\frac{dy}{dx} - 3y = 0$$

Let λ, μ be real numbers.

Show that the function $h\colon \mathbb{R} \longrightarrow \mathbb{R}$ defined by

$$h(x) := \lambda f(x) + \mu g(x)$$

for all $x \in \mathbb{R}$, also satisfies the given differential equation.

Exercise 2.3. Solve the system of differential equations

$$\begin{array}{lcll} x'(t) & - & 2y'(t) & = & x(t) \\ x'(t) & + & y'(t) & = & y(t) + x(t) \end{array}$$

where $x(t)$ and $y(t)$ denote real valued functions of the real variable t, and $'$ stands for the derivative.

Exercise 2.4. Find all integral matrices $\underline{\mathbf{A}} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ satisfying $\underline{\mathbf{A}}^2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$.

Exercise 2.5. Let $\underline{\mathbf{A}} = \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix}$.

Prove that for $n \in \mathbb{N},\ n \geq 1$,

$$\underline{\mathbf{A}}^n = \begin{bmatrix} a^n & 0 \\ 0 & b^n \end{bmatrix}$$

Exercise 2.6. (i) Let $\underline{\mathbf{A}} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$.

Prove that for $n \in \mathbb{N}$

$$\underline{\mathbf{A}}^n = \begin{bmatrix} \cos n\theta & -\sin n\theta \\ \sin n\theta & \cos n\theta \end{bmatrix}$$

(ii) For $\underline{\mathbf{A}} = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$, with $a, b \in \mathbb{R}$, find $\underline{\mathbf{A}}^n$.

Exercise 2.7. For each of the following matrices $\underline{\mathbf{A}}$, find $\underline{\mathbf{A}}^n$.

(i) $\underline{\mathbf{A}} = \begin{bmatrix} a & 1 \\ 0 & a \end{bmatrix}$.

(ii) $\underline{\mathbf{A}} = \begin{bmatrix} a & 1 & 0 \\ 0 & a & 1 \\ 0 & 0 & a \end{bmatrix}$.

(iii) $\underline{\mathbf{A}} = \begin{bmatrix} a & 1 & 0 & 0 \\ 0 & a & 1 & 0 \\ 0 & 0 & a & 1 \\ 0 & 0 & 0 & a \end{bmatrix}$.

(iv) $\underline{\mathbf{A}} = a\underline{\mathbf{1}}_k + \underline{\mathbf{N}}_k$, where $\underline{\mathbf{N}}_k := [x_{ij}]_{k\times k}$ is given by

$$x_{ij} = \begin{cases} 1 & \text{if } j = i+1 \\ 0 & \text{otherwise.} \end{cases}$$

Mathematics is so widely applicable because of, and not despite, its being abstract.

Peter Hilton

Chapter 3

Vector Spaces

Linear algebra is the theory of *vector spaces*, and this chapter begins their formal study. The previous chapter, looked in detail at examples of a vector space, beginning with solutions to homogeneous systems of linear equations with real coefficients. The exercises looked at further examples. The definitions below express the essential features of our examples without reference to their special features.

In our first example, there were actually two distinct sets: the solutions to the homogeneous systems of equations on the one hand, and the real numbers on the other. We had operations defined on each of these sets and a way of combining them.

The other examples in Chapter 2 were similar: a vector space is a set with additional structure.

The additional structure is *algebraic* in nature. It allows us to compute and solve numerous problems explicitly with relative ease. This ease of computation and broad range of application comes at the price of requiring a relatively large number of axioms to describe the structure.

The structure is a mixed structure, for, as our examples illustrated, a vector space is actually a set-with-structure upon which another set-with-structure acts. The latter set is a *field* and we begin with the axioms for a field.

3.1 Fields

The axioms for a field capture and formulate the structure common to the set of all rational numbers, $\mathbb{Q}$, the set of all real numbers, $\mathbb{R}$ and the set of all complex numbers, $\mathbb{C}$, which underlies arithmetic.

Definition 3.1. A *field* comprises a set, $\mathbb{F}$, together with two binary operations, *addition* and *multiplication*,

$$+_{\mathbb{F}} \colon \mathbb{F} \times \mathbb{F} \longrightarrow \mathbb{F}, \quad (x, y) \longmapsto x +_{\mathbb{F}} y$$
$$\times_{\mathbb{F}} \colon \mathbb{F} \times \mathbb{F} \longrightarrow \mathbb{F}, \quad (x, y) \longmapsto x \times_{\mathbb{F}} y$$

together with distinguished elements $0_{\mathbb{F}} \neq 1_{\mathbb{F}}$ satisfying the following axioms.

Given $x, y, z \in \mathbb{F}$,

A1 (**Associativity of Addition**)
$x +_{\mathbb{F}} (y +_{\mathbb{F}} z) = (x +_{\mathbb{F}} y) +_{\mathbb{F}} z$

A2 (**Existence of a Neutral Element for Addition**)
$x +_{\mathbb{F}} 0_{\mathbb{F}} = x = 0_{\mathbb{F}} +_{\mathbb{F}} x$

A3 (**Existence of Additive Inverses**)
There is a $-x \in \mathbb{F}$ with
$x +_{\mathbb{F}} (-x) = 0_{\mathbb{F}} = (-x) +_{\mathbb{F}} x$

A4 (**Commutativity of Addition**)
$y +_{\mathbb{F}} x = x +_{\mathbb{F}} y$

M1 (**Associativity of Multiplication**)
$x \times_{\mathbb{F}} (y \times_{\mathbb{F}} z) = (x \times_{\mathbb{F}} y) \times_{\mathbb{F}} z$

M2 (**Existence of a Neutral Element for Multiplication**)
$x \times_{\mathbb{F}} 1_{\mathbb{F}} = x = 1_{\mathbb{F}} \times_{\mathbb{F}} x$

M3 (**Existence of Mutiplicative Inverses**)
If $x \neq 0_{\mathbb{F}}$, there is a $x^{-1} \in \mathbb{F}$ with
$x \times_{\mathbb{F}} x^{-1} = 1_{\mathbb{F}} = x^{-1} \times_{\mathbb{F}} x$

M4 (**Commutativity of Multiplication**)
$y \times_{\mathbb{F}} x = x + \times_{\mathbb{F}} y$

D (**Distributivity of Multiplication over Addition**)
$x \times_{\mathbb{F}} (y +_{\mathbb{F}} z) = (x \times_{\mathbb{F}} y) + (x \times_{\mathbb{F}} z)$
$(x +_{\mathbb{F}} y) \times_{\mathbb{F}} z = (x \times_{\mathbb{F}} z) + (y \times_{\mathbb{F}} z)$

Observation 3.2. We write $+_{\mathbb{F}}$ and $\times_{\mathbb{F}}$ for the two operations in the definition of the field structure on the set $\mathbb{F}$ to emphasise that they need not actually be addition and multiplication as the reader is used to, and depend on the particular field in question. Example 3.9 on the next page illustrates this dramatically.

Observation 3.3. Axioms A1, A2 and A3 assert that $\mathbb{F}$ is a *group* with respect to addition.

Axioms M1, M2 and M3 assert that $\mathbb{F} \setminus \{0_{\mathbb{F}}\}$ is a group with respect to multiplication.

Axioms A4 and M4 assert that these two group structures commutative (or *abelian*).

Finally, Axiom D describes how these two group structures interact.

Observation 3.4. Axioms A1, A2, A3, A4, M1, M2 and D assert that $\mathbb{F}$ is a *(unital) ring* with respect to addition and multiplication.

Axiom M4 asserts that this ring structure is *commutative*.

Example 3.5. The rational numbers, $\mathbb{Q}$, the real numbers, $\mathbb{R}$, and the complex numbers, $\mathbb{C}$, all form fields with respect to their usual addition and multiplication.

Example 3.6. The set of integers, $\mathbb{Z}$, forms a commutative ring, but not a field, with respect to their usual addition and multiplication. $\mathbb{Z}$ fails to be a field because Axiom M3 does not hold: there is no integer, x, with $3x = 1$.

Example 3.7. The set of all natural numbers, $\mathbb{N}$, does not form a group with respect to its addition, since Axiom A3 does not hold: there is no $x \in \mathbb{N}$ with $1 + x = 0$.

Example 3.8. Take $\mathbb{F} = \{a, b\}$, with $a \neq b$.

Define binary operations, $+_{\mathbb{F}}$ and $\times_{\mathbb{F}}$ on $\mathbb{F}$ by means of their *Cayley tables*: The value of the operation at (x, y) is the entry in the row labelled by x and column labelled by y.

$+_{\mathbb{F}}$	a	b
a	a	b
b	b	a

and

$\times_{\mathbb{F}}$	a	b
a	a	a
b	a	b

One way to verify that this, or any other finite set, is a field is to list all the possible combinations and check that all the required equalities hold. This can be done systematically by listing all the combinations in a table, wit hone column for each variable, and a column for each of the operations performed. The equalities are verified if the column representing the left side of the equality agrees with the column representing the right side of the equality.

We illustrate this by drawing up the table to show that for all $x, y, z \in \mathbb{F}$,

$$x \times_{\mathbb{F}} (y +_{\mathbb{F}} z) = (x \times_{\mathbb{F}} y) +_{\mathbb{F}} (x \times_{\mathbb{F}} z)$$

and completing two of the rows.

x	y	z	$y +_{\mathbb{F}} z$	$x \times_{\mathbb{F}} (y +_{\mathbb{F}} z)$	$x \times_{\mathbb{F}} y$	$x \times_{\mathbb{F}} z$	$(x \times_{\mathbb{F}} y) +_{\mathbb{F}} (x \times_{\mathbb{F}} z)$
a	a	a					
a	a	b					
a	b	a					
a	b	b					
b	a	a	a	a	a	a	a
b	a	b					
b	b	a	b	b	b	a	b
b	b	b					

The entries in the column labelled $x \times_{\mathbb{F}} (y +_{\mathbb{F}} z)$ agree with the entries in the column labelled $(x \times_{\mathbb{F}} y) +_{\mathbb{F}} (x \times_{\mathbb{F}} z)$, at least in the rows we have completed.

It is left to the reader to complete this table and to do the same for the other axioms.

This field is often denoted by $\mathbb{F}_2$ and sometimes by $\mathbb{Z}/2\mathbb{Z}$.

Example 3.9. Let $\mathbb{F} := \{x \in \mathbb{R} \mid x > 0\}$ and define

$$+_{\mathbb{F}} \colon \mathbb{F} \times \mathbb{F} \longrightarrow \mathbb{F}, \quad (x, y) \longmapsto xy$$
$$\times_{\mathbb{F}} \colon \mathbb{F} \times \mathbb{F} \longrightarrow \mathbb{F}, \quad (x, y) \longmapsto x^{\ln y}.$$

These operations render $\mathbb{F}$ a field. The verification, including the identification of $0_{\mathbb{F}}$ and $1_{\mathbb{F}}$, is left to the reader as an exercise.

Example 3.10. Let $\mathbb{F}$ be the set of all matrices of the form $\begin{bmatrix} x & -y \\ y & x \end{bmatrix}$ with $x, y \in \mathbb{R}$.

Taking $+_{\mathbb{F}}$, $\times_{\mathbb{F}}$ to be the usual addition and multiplication of matrices, $\mathbb{F}$ becomes a field, as the reader can verify through direct computation.

Example 3.11. Let $\mathbb{R}[t]$ denote the set of all polynomials in the indeterminate t, with real coefficients, so that

$$\mathbb{R}[t] := \{a_0 + a_1 t + \cdots + a_n t^n \mid a_j \in \mathbb{R} \ (j = 1, \ldots, n) \ \text{ and } \ a_n \neq 0 \text{ if } n \neq 0\}$$

Define $\sim$ on $\mathbb{R}[t]$ by

$$p(t) \sim q(t) \ \text{ if and only if } \ t^2 + 1 \ \text{ divides } \ p(t) - q(t).$$

Direct verification shows that $\sim$ is an equivalence relation on $\mathbb{R}[t]$.

Let $\mathbb{F}$ denote the set of all $\sim$-equivalence classes, and denote the equivalence class of $p(t)$ by $[p(t)]$. $p(t)$ is called a *representative* of $[p(t)]$.

Each equivalence class is represented by a polynomial of the form $a + bt$, with $a, b \in \mathbb{R}$, so that

$$\mathbb{F} = \{\, [a + bt] \mid a, b \in \mathbb{R} \,\}.$$

Then $\mathbb{F}$ is a field with respect to addition and multiplication defined by

$$+_{\mathbb{F}} : \mathbb{F} \times \mathbb{F} \longrightarrow \mathbb{F}, \quad ([a + bt], [c + dt]) \longmapsto [(a + c) + (b + d)t]$$
$$\times_{\mathbb{F}} : \mathbb{F} \times \mathbb{F} \longrightarrow \mathbb{F}, \quad ([a + bt], [c + dt]) \longmapsto [(ac - bd) + (ad + bc)t]$$

3.2 Vector Spaces

Vectors in physics motivate the mathematical notion of a vector spaces. Vectors, such as forces, can be added: if two forces act upon a given object, there is a net resultant vector. Vectors can also be scaled, that is multiplied by a *scalar*.

The next definition formulates the essential features of vectors in physics and our other examples.

Definition 3.12. A *vector space over the field* $\mathbb{F}$ — or an $\mathbb{F}$*-vector space* — is a set, V, with a distinguished element $\mathbf{0}_V$, together with a binary operation on V, the *addition of two vectors*

$$\boxplus_V : V \times V \longrightarrow V, \quad (\mathbf{u}, \mathbf{v}) \longmapsto \mathbf{u} \boxplus_V \mathbf{v},$$

and an operation of the field $\mathbb{F}$ on V, the *multiplication of a vector by a scalar*

$$\boxdot_V : \mathbb{F} \times V \longrightarrow V, \quad (\lambda, \mathbf{v}) \longmapsto \lambda \boxdot_V \mathbf{v}$$

satisfying the axioms listed below.

Given $\mathbf{x}, \mathbf{y}, \mathbf{z} \in V$ and $\lambda, \mu \in \mathbb{F}$,

VS1 $\mathbf{x} \boxplus_V (\mathbf{y} \boxplus_V \mathbf{z}) = (\mathbf{x} \boxplus_V \mathbf{y}) \boxplus_V \mathbf{z}$

VS2 $\mathbf{x} \boxplus_V \mathbf{0}_V = \mathbf{x} = \mathbf{0}_V \boxplus_V \mathbf{x}$

VS3 There is a $-\mathbf{x}$ such that $\mathbf{x} \boxplus_V (-\mathbf{x}) = \mathbf{0}_V = (-\mathbf{x}) \boxplus_V \mathbf{x}$

VS4 $\mathbf{y} \boxplus_V \mathbf{x} = \mathbf{x} \boxplus_V \mathbf{y}$

VS5 $1_{\mathbb{F}} \boxdot_V \mathbf{x} = \mathbf{x}$

VS6 $\lambda \boxdot_V (\mathbf{x} \boxplus_V \mathbf{y}) = (\lambda \boxdot_V \mathbf{x}) \boxplus_V (\lambda \boxdot_V \mathbf{y})$

VS7 $(\lambda +_{\mathbb{F}} \mu) \boxdot_V \mathbf{x} = (\lambda \boxdot_V \mathbf{x}) \boxplus_V (\mu \boxdot_V \mathbf{x})$

VS8 $(\lambda \times_{\mathbb{F}} \mu) \boxdot_V \mathbf{x} = \lambda \boxdot_V (\mu \boxdot_V \mathbf{x})$

When V is a vector space over $\mathbb{F}$, the elements of V are the *vectors* and those of $\mathbb{F}$ the *scalars*.

Thus a vector space V over the field $\mathbb{F}$ comprises an abelian group, V (Axioms VS1 to VS4) together with an *action* of the field $\mathbb{F}$ on V (Axioms VS5 to VS8).

Example 3.13. Let $\mathbb{F}$ be a field and put $\mathbb{F}^n := \{(x_1, \ldots, x_n) \mid x_1, \ldots, x_n \in \mathbb{F}\}$.

Define addition and multiplication by scalars by

$$(x_1, \ldots, x_n) \boxplus_V (y_1, \ldots, y_n) := (x_1 +_{\mathbb{F}} y_1, \ldots, x_n +_{\mathbb{F}} y_n)$$
$$\lambda \boxdot_V (x_1, \ldots, x_n) := (\lambda \times_{\mathbb{F}} x_1, \ldots, \lambda \times_{\mathbb{F}} x_n)$$

Then $\mathbb{F}^n$ is a vector space over $\mathbb{F}$. When we refer to $\mathbb{F}^n$ as a vector space over $\mathbb{F}$, we shall always mean the vector space structure just defined.

A familiar case is $\mathbb{R}^2 := \{(x_1, x_2) \mid x_1, x_2 \in \mathbb{R}\}$, the set of all ordered pairs of real numbers. [Here $\mathbb{F} := \mathbb{R}$ and $n = 2$.]

Observation 3.14. When $n = 1$ in Example 3.13, the addition of vectors, $\boxplus_V$, coincides with the addition, $+_{\mathbb{F}}$, in the field $\mathbb{F}$ and multiplication of a vector by a scalar, $\boxdot_V$, is just the multiplication, $\times_{\mathbb{F}}$, in the field. Thus, a field is always a vector space over itself and any property of an arbitrary vector space over $\mathbb{F}$ is also a property of $\mathbb{F}$ itself. In other words,

The notion of a vector space is a generalisation of the notion of a field.

Example 3.15. Let V denote the set of all solutions of the homogeneous system of real linear equations

$$\begin{array}{rcrcrcl} x & + & 2y & + & 4z & = & 0 \\ 2x & + & 5y & + & 11z & = & 0 \end{array}$$

so that

$$V = \{(x, y, z) \in \mathbb{R}^3 \mid x + 2y + 4z = 2x + 5y + 11z = 0\}$$

If we use the addition and multiplication on elements of $\mathbb{R}^3$ defined in Example 3.13, then V becomes a vectors space over $\mathbb{R}$.

Notice that we do not need to solve the system of equations to verify that the set of all solutions forms a vector space.

Example 3.16. Though the reader has already met matrices elsewhere, we revise the formal definition and show that the set of all matrices of a fixed size forms a vector space.

Definition 3.17. An $m \times n$ matrix over $\mathbb{F}$ is an array of mn elements of $\mathbb{F}$ arranged into m rows and n columns. We write $[a_{ij}]$ to denote the $m \times n$ matrix over $\mathbb{F}$ with a_{ij} the ij-th *coefficient* or *entry*. Here, the first subscript specifies the row and the second specifies the column. Thus,

$$\underline{\mathbf{A}} = [a_{ij}] = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}$$

$\mathbf{M}(m \times n; \mathbb{F})$ denotes the set of all $m \times n$ matrices over $\mathbb{F}$, simplified to $\mathbf{M}(n; \mathbb{F})$ when $m = n$.

Given $\underline{\mathbf{A}} = [a_{ij}]_{m \times n}$ and $\underline{\mathbf{B}} = [b_{ij}]_{m \times n}$ as well as $\lambda \in \mathbb{F}$ we define

$$[a_{ij}]_{m \times n} \boxplus_V [b_{ij}]_{m \times n} := [a_{ij} +_{\mathbb{F}} b_{ij}]_{m \times n}$$
$$\lambda \boxdot_V [a_{ij}]_{m \times n} := [\lambda \times_{\mathbb{F}} a_{ij}]_{m \times n}.$$

This renders $\mathbf{M}(m \times n; \mathbb{F})$ an $\mathbb{F}$-vector space.

Example 3.18. Let $\mathbb{F}$ be a field and X a non-empty set.

Let V be the set of all $\mathbb{F}$-valued functions defined on X , so that

$$V = \mathcal{F}(X) := \{f \colon X \longrightarrow \mathbb{F} \mid f \text{ is a function } \}$$

Define

$$\boxplus_V \colon V \times V \longrightarrow V, \quad (f, g) \longmapsto f \boxplus_V g$$

where, for all $x \in X$

$$(f \boxplus_V g)(x) := f(x) +_{\mathbb{F}} g(x)$$

and

$$\boxdot_V \colon \mathbb{F} \times V \longrightarrow V, \quad (\lambda, f) \longmapsto \lambda \boxdot_V f$$

where, for all $x \in X$

$$(\lambda \boxdot_V f)(x) := \lambda \times_{\mathbb{F}} f(x)$$

This defines an $\mathbb{F}$-vector space structure on V.

Two closely related vector spaces, which we meet again later, are introduced in Exercises 3.11 on page 43 and 3.12 on page 43.

Example 3.19. Take $V = \{f \colon \mathbb{R} \longrightarrow \mathbb{R} \mid xf'(x) - f(x) = 0 \text{ for all } x \in \mathbb{R}\}$.

Direct verification shows that V is a vector space over $\mathbb{R}$ with respect to the operations defined in Example 3.18.

We note that there is no need to solve the differential equation in order to see that the set of all solutions forms a vector space.

Example 3.20. Let V be the set of all sequences of elements of the field, $\mathbb{F}$, so that

$$V = \{(x_n)_{n\in\mathbb{N}} \mid x_n \in \mathbb{F} \text{ for every } n \in \mathbb{N}\}$$

Define

$$\boxplus_V \colon V \times V \longrightarrow V, \quad \big((x_n)_{n\in\mathbb{N}}, (y_n)_{n\in\mathbb{N}}\big) \longmapsto (x_n +_{\mathbb{F}} y_n)_{n\in\mathbb{N}}$$
$$\boxdot_V \colon \mathbb{F} \times V \longrightarrow V, \quad \big(\lambda, (x_n)_{n\in\mathbb{N}}\big) \longmapsto (\lambda \times_{\mathbb{F}} x_n)_{n\in\mathbb{N}}$$

This defines an $\mathbb{F}$-vector space structure on V.

Example 3.21. Let $V = \mathbb{R}[t]$, the set of all polynomials in the indeterminate t with real coefficients:

$$\begin{aligned}\mathbb{R}[t] :&= \{a_0 + a_1 t + \cdots + a_n t^n \mid n \in \mathbb{N},\ a_j \in \mathbb{R} \text{ for } 0 \le j \le n, \text{ and } a_n \neq 0 \text{ if } n > 0\} \\ &= \Big\{ \sum_{j=0}^{n} a_j t^j \mid n \in \mathbb{N},\ a_j \in \mathbb{R}, \text{ and } a_n \neq 0 \text{ if } n > 0 \Big\}\end{aligned}$$

where we have adopted the convention that $a_0 t^0 = a_0$.

Define

$$\boxplus_V \colon V \times V \longrightarrow V, \quad \Big(\sum_{i=0}^{m} a_i t^i,\ \sum_{j=0}^{n} b_j t^j \Big) \longmapsto \sum_{k=0}^{\max\{m,n\}} c_k t^k$$

$$\text{where } c_k = \begin{cases} a_k + b_k & \text{if } k \leq \min\{m,n\} \\ b_k & \text{if } m < n \text{ and } m < k \leq n \\ a_k & \text{if } n < m \text{ and } n < k \leq m \end{cases}$$

$$\boxdot_V : \mathbb{R} \times V \longrightarrow V, \quad \Big(\lambda, \sum_{j=0}^{n} a_j t^j\Big) \longmapsto \sum_{j=0}^{n} (\lambda a_j) t^j$$

$\mathbb{R}[t]$ is a real vector space with respect to these operations. This example should be familiar from calculus.

The reader should note that the only feature of $\mathbb{R}$ needed here is that it is a field. This allows us to generalise this example:

Given a field $\mathbb{F}$, the set of all polynomials in the indeterminate t with coefficients in $\mathbb{F}$,

$$\mathbb{F}[t] := \Big\{ \sum_{j=0}^{n} a_j t^j \mid n \in \mathbb{N},\ a_j \in \mathbb{F},\ \text{and } a_n \neq 0 \ \text{ if } n > 0 \Big\}$$

is a vector space over $\mathbb{F}$.

Example 3.22. Our final example here is also taken from calculus.

Let $V = \mathbb{R}[[t]]$, the set of all power series in t with real coefficients:

$$\mathbb{R}[[t]] : = \Big\{ \sum_{j=0}^{\infty} a_j t^j \mid a_j \in \mathbb{R} \text{ for all } j \in \mathbb{R} \Big\}$$

where we have adopted the convention that $a_0 t^0 = a_0$.

Define

$$\boxplus_V : V \times V \longrightarrow V, \quad \Big(\sum_{i=0}^{\infty} a_i t^i,\ \sum_{j=0}^{\infty} b_j t^j \Big) \longmapsto \sum_{k=0}^{\infty} (a_k + b_k) t^k$$

$$\boxdot_V : \mathbb{R} \times V \longrightarrow V, \quad \Big(\lambda, \sum_{j=0}^{\infty} a_j t^j\Big) \longmapsto \sum_{j=0}^{\infty} (\lambda a_j) t^j$$

$\mathbb{R}[t]$ is a real vector space with respect to these operations.

The reader should note that, once again, the only feature of $\mathbb{R}$ needed here is that it is a field. This allows us to generalise this example:

Given a field $\mathbb{F}$, the set of all power series in t with coefficients in $\mathbb{F}$,

$$\mathbb{F}[[t]] := \Big\{ \sum_{j=0}^{\infty} a_j t^j \mid a_j \in \mathbb{F},\ \text{for all } j \in \mathbb{N} \Big\}$$

is a vector space over $\mathbb{F}$.

Our less than exhaustive list of examples of vector spaces is far from exhaustive is enough to shows the diversity of vector spaces. It is difficult to see, at first glance, that these examples have anything in common.

It is for this reason for developing the theory of vector spaces. Rather than working in each specific class of vector spaces separately, and re-inventing the wheel on each occasion, a single mathematical theory was developed, with its techniques and tools, having a very broad range of applications. In other word, linear algebra unifies significant aspects of a broad range of mathematics and its applications.

The notation we introduced — $+_{\mathbb{F}}, \times_{\mathbb{F}}, \boxplus_V, \boxtimes_V$ — is initially important for separating the different algebraic operations, and for reminding the reader that the "addition" and "multiplication" in question may have nothing to do with the addition and multiplication familiar from arithmetic. But it is cumbersome, and so we simplify our notation in order to make the text easier to read.

Notational Convention. Except when there is danger of confusion, or for emphasis, we shall avail ourselves of *systematic ambiguity*:

1. We write 0 and 1 for $0_{\mathbb{F}}$ and $1_{\mathbb{F}}$ respectively.
2. We write $\lambda + \mu$ and $\lambda\mu$ for $\lambda +_{\mathbb{F}} \mu$ and $\lambda \times_{\mathbb{F}} \mu$ respectively.
3. We write $\mathbf{x} + \mathbf{y}$ for $\mathbf{x} \boxplus_V \mathbf{y}$ and $\lambda\mathbf{x}$ for $\lambda \boxdot_V \mathbf{x}$.
4. We write $\mathbf{0}$ for $\mathbf{0}_V$.

There is little danger of confusion, for it is usually clear from the context whether two vectors or two scalars are being added, and whether two scalars are being multiplies, or a vector by a scalar.

We leave it to the good sense of the reader to recognise from the context which operations are intended and turn to establishing a number of elementary properties of vector space, which we are indispensable for computing applications and are needed when developing theory.

Theorem 3.23. *Let V be a vector space over the field $\mathbb{F}$. Take $\lambda \in \mathbb{F}$ and $\mathbf{x}, \mathbf{y}, \mathbf{z} \in V$. Then*

(a) $\mathbf{x} + \mathbf{y} = \mathbf{x} + \mathbf{z}$ *if only if* $\mathbf{y} = \mathbf{z}$. *(In particular,* $\mathbf{x} + \mathbf{y} = \mathbf{0}_V$ *if and only if* $\mathbf{y} = -\mathbf{x}$.*)*

(b) $-\mathbf{0}_V = \mathbf{0}_V$

(c) $\lambda\mathbf{0}_V = \mathbf{0}_V$

(d) $0\mathbf{x} = \mathbf{0}_V$

(e) $\lambda\mathbf{x} = \mathbf{0}_V$ *if and only if either* $\lambda = 0$ *or* $\mathbf{x} = \mathbf{0}_V$

(f) $(-1)\mathbf{x} = -\mathbf{x}$

Proof. Take $\mathbf{x}, \mathbf{y}, \mathbf{z} \in V$ and $\lambda \in \mathbb{F}$.

(a) Suppose $\mathbf{x} + \mathbf{y} = \mathbf{x} + \mathbf{z}$.

$$\begin{aligned} -\mathbf{x} + (\mathbf{x} + \mathbf{y}) &= -\mathbf{x} + (\mathbf{x} + \mathbf{z}) && \text{by VS3} \\ (-\mathbf{x} + \mathbf{x}) + \mathbf{y} &= (-\mathbf{x} + \mathbf{x}) + \mathbf{z} && \text{by VS1} \\ \mathbf{0}_V + \mathbf{y} &= \mathbf{0}_V + \mathbf{z} && \text{by VS3} \end{aligned}$$

Hence, by VS2

$$\mathbf{y} = \mathbf{z}$$

(b)

$$\begin{aligned} -\mathbf{0}_V &= -\mathbf{0}_V + \mathbf{0}_V && \text{by VS2} \\ &= \mathbf{0}_V && \text{by VS3} \end{aligned}$$

(c)

$$\lambda\mathbf{x} + \lambda\mathbf{0}_V = \lambda(\mathbf{x} + \mathbf{0}_V) \qquad \text{by VS6}$$

$$\begin{aligned} &= \lambda\mathbf{x} && \text{by VS2} \\ &= \lambda\mathbf{x} + \mathbf{0}_V && \text{by VS2.} \end{aligned}$$

Hence, by (a),

$$\lambda\mathbf{0}_V = \mathbf{0}_V$$

(d)

$$\begin{aligned} \mathbf{x} + 0\mathbf{x} &= 1\mathbf{x} + 0\mathbf{x} && \text{by VS5} \\ &= (1+0)\mathbf{x} && \text{by VS7} \\ &= 1\mathbf{x} && \text{by properties of fields} \\ &= \mathbf{x} && \text{by VS5} \\ &= \mathbf{x} + \mathbf{0}_V && \text{by VS2} \end{aligned}$$

Hence, by (a),

$$0\mathbf{x} = \mathbf{0}_V$$

(e) Suppose that $\lambda\mathbf{x} = \mathbf{0}_V$ and $\lambda \neq 0$.

$$\begin{aligned} \mathbf{x} &= 1\mathbf{x} && \text{by VS5} \\ &= (\frac{1}{\lambda}\lambda)\mathbf{x} && \text{by properties of fields} \\ &= \frac{1}{\lambda}(\lambda\mathbf{x}) && \text{by VS8} \\ &= \frac{1}{\lambda}\mathbf{0}_V && \text{by hypothesis} \\ &= \mathbf{0}_V && \text{by (c).} \end{aligned}$$

(f)

$$\begin{aligned} \mathbf{x} + (-1)\mathbf{x} &= 1\mathbf{x} + (-1)\mathbf{x} && \text{by VS5} \\ &= (1 + (-1))\mathbf{x} && \text{by VS7} \\ &= 0\mathbf{x} && \text{by properties of fields} \\ &= \mathbf{0}_V && \text{by (d)} \\ &= \mathbf{x} + (-\mathbf{x}) && \text{by VS3} \end{aligned}$$

Hence, by (a), $(-1)\mathbf{x} = -\mathbf{x}$ □

Corollary 3.24. *Let $\mathbb{F}$ be a field and take $x \in \mathbb{F}$. Then $-(-x) = x$.*

Proof. As $(-x) + (-(-x)) = 0$ and $(-x) + x = 0$, the conclusion follows by Theorem 3.23 on the preceding page(a). □

Observation 3.25. Theorem 3.23 on the facing page applied to $\mathbb{Q}$, $\mathbb{R}$ or $\mathbb{C}$ proves the usual "Laws of Arithmetic", which are taught at school, usually without proper explanation.

Observation 3.26. A set, V, may be a vector space over the same field in more than one way, even when the vector addition is the same in both cases.

Example 3.27. As an example, let V be a vector space over $\mathbb{C}$, with operations $\boxplus_V$ and $\boxdot_V$. We define a new vector space structure on V by defining

$$\overline{\boxplus}_V : V \times V \longrightarrow V, \quad (\mathbf{u}, \mathbf{v}) \longmapsto \mathbf{u} \boxplus_V \mathbf{v}$$
$$\overline{\boxdot}_V : \mathbb{C} \times V \longrightarrow V, \quad (\alpha, \mathbf{v}) \longmapsto \overline{\alpha} \boxdot_V \mathbf{v}.$$

In other words, we "twist" multiplication by a scalar: instead of multiplying vectors by a given complex number, we multiply them by its complex conjugate.

To see that this is a genuinely different vector space, observe that for any vector, $\mathbf{v} \in V$,

$$\begin{aligned} i \boxdot_V \mathbf{v} = i \overline{\boxdot}_V \mathbf{v} \quad & \text{if and only if} \quad i \boxdot_V \mathbf{v} = -i \boxdot_V \mathbf{v} \\ & \text{if and only if} \quad 2i \boxdot_V \mathbf{v} = \mathbf{0}_V \\ & \text{if and only if} \quad \mathbf{v} = \mathbf{0}_V \qquad \text{by Theorem 3.23 on page 38(e)} \end{aligned}$$

3.3 Exercises

Exercise 3.1. Show that $\mathbb{F} = \{a, b\}$, with $a \neq b$ is a field with respect to the operations + and . defined by:

+	a	b
a	a	b
b	b	a

and

.	a	b
a	a	a
b	a	b

Here we have defined the two binary operations by means of their *Cayley tables*: The value of the operation at (x, y) is the entry in the row labelled by x and column labelled by y.

This field is often denoted by $\mathbb{F}_2$, and sometimes by $\mathbb{Z}/2\mathbb{Z}$.

Exercise 3.2. Show that $\mathbb{F} := \{a, b, c\}$, with all elements distinct, is a field with respect to the operations + and . defined by:

+	a	b	c
a	a	b	c
b	b	c	a
c	c	a	b

and

.	a	b	c
a	a	a	a
b	a	b	c
c	a	c	b

This field is often denoted by $\mathbb{F}_3$, and sometimes by $\mathbb{Z}/3\mathbb{Z}$.

Exercise 3.3. Show that $\mathbb{F} := \{a, b, c, d\}$, with all elements distinct, is a field with respect to the operations + and . defined by:

+	a	b	c	d
a	a	b	c	d
b	b	a	d	c
c	c	d	a	b
d	d	c	b	a

and

.	a	b	c	d
a	a	a	a	a
b	a	b	c	d
c	a	c	d	b
d	a	d	b	c

This field is usually denoted by $\mathbb{F}_4$, or $\mathbb{F}_{2^2}$.

Exercise 3.4. Show that the usual addition and multiplication render

(a) $\mathbb{C}$ a vector space over $\mathbb{R}$;

(b) $\mathbb{R}$ a vector space over $\mathbb{Q}$;

(c) $\mathbb{C}$ a vector space over $\mathbb{Q}$.

These examples illustrate a general result.

If the field $\mathbb{F}$ is a subfield of the field $\mathbb{K}$, then $\mathbb{K}$ is a vector space over $\mathbb{F}$.

The reader is invited to adapt his/her solution to this exercise to prove of the general result.

Exercise 3.5. Let V be the set of all real solutions of the system of homogeneous linear equations

$$\begin{array}{rcrcrcl} 3x & + & 2y & - & z & = & 0 \\ x & - & 5y & + & 7z & = & 0 \end{array}$$

so that

$$V = \{(x, y, z) \in \mathbb{R}^3 \mid 3x + 2y - z = 0 \text{ and } x - 5y + 7z = 0\}$$

Given $(x, y, z), (u, v, w) \in V$ and $\alpha \in \mathbb{R}$, define

$$\begin{aligned} (x, y, z) \boxplus_V (u, v, w) &:= (x + u, y + v, z + w) \\ \alpha \boxdot_V (x, y, z) &:= (\alpha x, \alpha y, \alpha z) \end{aligned}$$

Prove that these render V a vector space over $\mathbb{R}$.

Exercise 3.6. Decide whether the following are vector spaces.

(a) Take $\mathbb{F} := \mathbb{C}$ and $V := \mathbb{C}$.
Define $\boxplus_V$ to be the usual addition of complex numbers, and $\boxdot_V$ by

$$\alpha \boxdot_V z := \alpha^2 z \qquad (\alpha, z \in \mathbb{C})$$

(b) Let $\mathbb{F}$ be any field and $V := \mathbb{F}^2$.
Define $\boxplus_V$ to be the usual (component-wise) addition of ordered pairs, and $\boxdot_V$ by

$$\alpha \boxdot_V (\beta, \gamma) := (\alpha, \beta) \qquad (\alpha, \beta, \gamma \in \mathbb{F})$$

(c) Take $\mathbb{F} := \mathbb{F}_{2^2}$ and $V := \mathbb{F}^2$.
Define $\boxplus_V$ to be the usual addition of complex numbers, and $\boxdot_V$ by

$$\alpha \boxdot_V \mathbf{v} := \begin{cases} (\alpha\beta, \alpha\gamma) & \text{if } \gamma \neq 0 \\ (\alpha^2\beta, 0) & \text{if } \gamma = 0 \end{cases}$$

(d) Take $\mathbb{F} := \mathbb{C}$ and $V := \mathbb{C}$.
Define $\boxplus_V$ to be the usual addition of complex numbers, and $\boxdot_V$ by

$$\alpha \boxdot_V z := \Re(\alpha) z \qquad (\alpha, z \in \mathbb{C}),$$

where $\Re(\alpha)$ denotes the real part of the complex number α.

(e) Take $\mathbb{F} := \mathbb{R}$ and $V := \mathbb{R}^+ = \{r \in \mathbb{R} \mid r > 0\}$.

Define $\boxplus_V$ and $\boxdot_V$ by

$$x \boxplus_V y := xy \qquad (x, y \in \mathbb{R}^+)$$
$$\alpha \boxdot_V x := x^\alpha \qquad (\alpha \in \mathbb{R}, x \in \mathbb{R}^+)$$

Exercise 3.7. Let $\mathbb{F} := \{x \in \mathbb{R} \mid x > 0\}$ and define

$$+_{\mathbb{F}} : \mathbb{F} \times \mathbb{F} \longrightarrow \mathbb{F}, \quad (x, y) \longmapsto xy$$
$$\times_{\mathbb{F}} : \mathbb{F} \times \mathbb{F} \longrightarrow \mathbb{F}, \quad (x, y) \longmapsto x^{\ln y}.$$

Prove that these operations render $\mathbb{F}$ a field.

Exercise 3.8. Put $\mathbb{F} := \left\{ \begin{bmatrix} x & -y \\ y & x \end{bmatrix} \mid x, y \in \mathbb{R} \right\}$.

Let $+_{\mathbb{F}}$, $\times_{\mathbb{F}}$ be the usual addition and multiplication of matrices, so that

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} +_{\mathbb{F}} \begin{bmatrix} r & s \\ t & u \end{bmatrix} := \begin{bmatrix} a+r & b+s \\ c+t & d+u \end{bmatrix}$$

and

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \times_{\mathbb{F}} \begin{bmatrix} r & s \\ t & u \end{bmatrix} := \begin{bmatrix} ar+bt & as+bu \\ cr+dt & cs+du \end{bmatrix}$$

Prove that $\mathbb{F}$ a field with respect to these operations.

Exercise 3.9. Let $\mathbb{R}[t]$ denote the set of all polynomials in the indeterminate t with real coefficients, so that

$$\mathbb{R}[t] := \{a_0 + a_1 t + \cdots + a_n t^n \mid a_j \in \mathbb{R} \ (j = 1, \ldots, n) \ \text{ and } \ a_n \neq 0 \text{ if } n \neq 0\}$$

(a) Define a relation, $\sim$, on $\mathbb{R}[t]$ by

$$p(t) \sim q(t) \ \text{ if and only if } \ t^2 + 1 \ \text{ divides } \ p(t) - q(t).$$

Prove that $\sim$ is an equivalence relation on $\mathbb{R}[t]$.

(b) Let $\mathbb{F}$ be the set of all $\sim$-equivalence classes, and $[p(t)]$ the $\sim$-equivalence class of $p(t)$.

Prove that every equivalence class contains a polynomial of the form $a + bt$, with $a, b \in \mathbb{R}$, so that

$$\mathbb{F} = \{[a + bt] \mid a, b \in \mathbb{R}\}.$$

(c) Define

$$+_{\mathbb{F}} : \mathbb{F} \times \mathbb{F} \longrightarrow \mathbb{F}, \quad ([a + bt], [c + dt]) \longmapsto [(a + c) + (b + d)t]$$
$$\times_{\mathbb{F}} : \mathbb{F} \times \mathbb{F} \longrightarrow \mathbb{F}, \quad ([a + bt], [c + dt]) \longmapsto [(ac - bd) + (ad + bc)t]$$

Prove that these definitions render $\mathbb{F}$ a field.

Exercise 3.10. Let $\mathbb{F}$ be any field and X a non-empty set and V the set all $\mathbb{F}$-valued functions defined on X, so that

$$V := \{f \colon X \longrightarrow \mathbb{F} \mid f \text{ is a function}\}.$$

Define

$$\boxplus_V \colon V \times V \longrightarrow V, \quad (f, g) \longmapsto f \boxplus_V g$$

where, for all $x \in X$,

$$(f \boxplus_V g)(x) := f(x) +_{\mathbb{F}} g(x)$$
$$\boxdot_V \colon \mathbb{F} \times V \longrightarrow V, \quad (\lambda, f) \longmapsto \lambda \boxdot_V f$$

where, for all $x \in X$,

$$(\lambda \boxdot_V f)(x) := \lambda \times_{\mathbb{F}} f(x).$$

Prove that these definitions render V a vector space over $\mathbb{F}$.

Exercise 3.11. Let W be a vector space over the field $\mathbb{F}$ and X a non-empty set.
Let V be $\mathcal{F}(X, W)$ be the set all W-valued functions defined on X, so that

$$V = \mathcal{F}(X, W) := \{f \colon X \longrightarrow W \mid f \text{ is a function}\}.$$

Define

$$\boxplus_V \colon V \times V \longrightarrow V, \quad (f, g) \longmapsto f \boxplus_V g$$

where, for all $x \in X$,

$$(f \boxplus_V g)(x) := f(x) \boxplus_W g(x)$$
$$\boxdot_V \colon \mathbb{F} \times V \longrightarrow V, \quad (\lambda, f) \longmapsto \lambda \boxdot_V f$$

where, for all $x \in X$

$$(\lambda \boxdot_V f)(x) := \lambda \boxdot_W f(x).$$

Prove that these definitions render V a vector space over $\mathbb{F}$.
Since every field is a vector space over itself, this exercise generalises Exercise 3.10.

Exercise 3.12. Let X be a non-empty set.
Let V be the set all $\mathbb{F}$-valued functions defined on X, such that $f(x) = 0$ for all but finitely many $x \in X$, so that

$$V = \{f \colon X \longrightarrow | f(x) \neq 0 \text{ has only finitely many solutions } x \in X\}.$$

Define

$$\boxplus_V \colon V \times V \longrightarrow V, \quad (f, g) \longmapsto f \boxplus_V$$

where, for all $x \in X$,

$$(f \boxplus_V g)(x) := f(x) \boxplus_W g(x)$$

$$\boxdot_V : \mathbb{F} \times V \longrightarrow V, \quad (\lambda, f) \longmapsto \lambda \boxdot_V f$$

where, for all $x \in X$

$$(\lambda \boxdot_V f)(x) := \lambda \boxdot_W f(x).$$

Prove that these definitions render V a vector space over $\mathbb{F}$.

Note that when X is a finite set, the vector space V in this exercise is the same as the vector space V in Example 3.18 on page 36.

Exercise 3.13. Let V be the set of all solutions of the differential equation

$$x\frac{dy}{dx} = y,$$

so that

$$V = \{f : \mathbb{R} \longrightarrow \mathbb{R} \mid xf'(x) - f(x) = 0\}$$

Prove that V is a vector space over $\mathbb{R}$ with respect to the operations defined in Example 3.18 on page 36.

Exercise 3.14. Let V be $\mathbf{M}(m \times n; \mathbb{F})$, the set of all $m \times n$ matrices over $\mathbb{F}$.

Given $\underline{\mathbf{A}} = [a_{ij}]_{m\times n}$ and $\underline{\mathbf{B}} = [b_{ij}]_{m\times n}$ as well as $\lambda \in \mathbb{F}$ we define

$$[a_{ij}]_{m\times n} \boxplus_V [b_{ij}]_{m\times n} := [a_{ij} +_{\mathbb{F}} b_{ij}]_{m\times n}$$
$$\lambda \boxdot_V [a_{ij}]_{m\times n} := [\lambda \times_{\mathbb{F}} a_{ij}]_{m\times n}.$$

Prove that $\mathbf{M}(m \times n; \mathbb{F})$ an $\mathbb{F}$-vector space with respect to these operations.

Exercise 3.15. Let V be a vector space over $\mathbb{C}$, with respect to operations $\boxplus_V$ and $\boxdot_V$.

Prove that

$$\overline{\boxplus}_V : V \times V \longrightarrow V, \quad (\mathbf{u}, \mathbf{v}) \longmapsto \mathbf{u} \boxplus_V \mathbf{v}$$
$$\overline{\boxdot}_V : \mathbb{C} \times V \longrightarrow V, \quad (\alpha, \mathbf{v}) \longmapsto \overline{\alpha} \boxdot_V \mathbf{v}$$

also defines a $\mathbb{C}$-vector space structure on V.

Science is a way of thinking much more than it is a body of knowledge.

Carl Sagan

Chapter 4

Geometric Interpretation

Our definitions are abstract because this level of abstractness has several distinct advantages.

(i) It allows us to isolate the essence of the matter.

(ii) It broadens the applicability of the theory we develop.

(iii) It makes proofs of general results simpler, more elegant and more transparent, even if this might not be apparent upon first encounter. For the abstractness forces us to use only general concepts rather than special tricks tailored to specific examples.

Nevertheless, this abstractness can be daunting, if one is unaccustomed to abstract methods.

It is therefore important to have a few standard examples, or common applications, both as a guide and as a warning: These examples offer concrete illustrations of the ideas investigated, show some of the difficulties which can arise, and display what can "go wrong".

The oldest applications of linear algebra are to geometry and to physical situations, as the term *vector space* attests. More recent applications include number theory, statistics, economics and computer science.

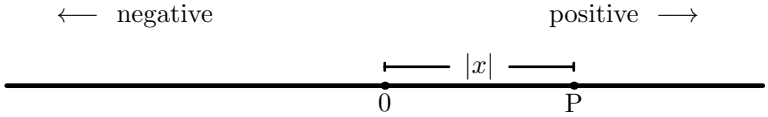
To master the concepts in this course, it is important to bear these examples in mind, with the caveat that examples may exhibit special features not shared by all examples.

Chapter 2 contained such examples. This chapter provides geometric examples.

Descartes introduced what we now call *Cartesian co-ordinates* to study geometry. (These are typically introduced in secondary school mathematics.)

To study the geometry of the line, we draw a line, ℓ, and choose a fixed point $\mathbf{0}$ on it. We choose one side of the line (one direction) from $\mathbf{0}$ and call it *positive*. The other side (the opposite direction) is called *negative*. Finally, the point P on the line ℓ is assigned the real number, x, as its *co-ordinate* when the distance from $\mathbf{0}$ to P is $|x|$, with x positive or negative according to whether P is on the positive or negative side of $\mathbf{0}$.

Assigning each point P of the line ℓ its co-ordinate x defines a bijection between ℓ and $\mathbb{R}$.



For plane geometry and spatial geometry, we take two (resp. three) mutually perpendicular, concurrent lines in the plane (resp. in space). These are the x_1- *and* x_2*–axes* (resp. x_1-, x_2- *and* x_3-*axes*).

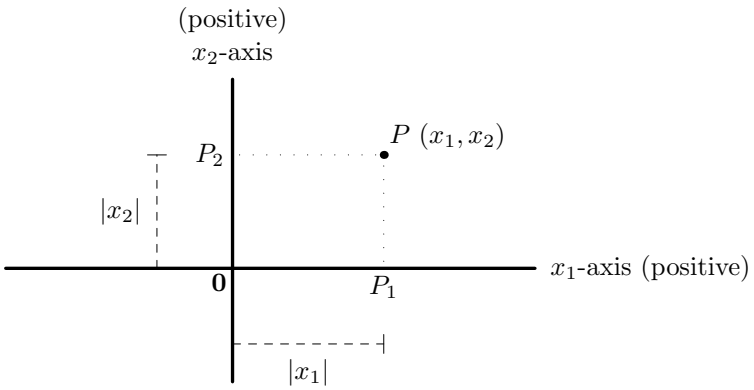
Their point of intersection is called the *origin*, which we denote by $\mathbf{0}$.

We choose *positive* and *negative* directions for each co-ordinate axis.

To each point, P, in the plane (resp. in space) we assign an *ordered pair* (resp. *ordered triple*) of real numbers, (x_1, x_2) (resp. (x_1, x_2, x_3)), called the *co-ordinates* of P. The i-th co-ordinate, x_i is obtained by taking the line through P perpendicular to the i-th co-ordinate axis and finding P_i, the point of intersection with the i-th co-ordinate axis. Then the distance of P_i form $\mathbf{0}$ is $|x_i|$, with x_i positive or negative according to whether $P_i$is on the positive or negative side of the i-th co-ordinate axis.

Clearly $\mathbf{0}$ has co-ordinates $(0, 0)$ (resp. $(0, 0, 0)$.

We illustrate the case of the plane.



Assigning each point P in the plane its co-ordinate pair (x_1, x_2) defines a bijection between the plane and $\mathbb{R}^2 = \mathbb{R}\times\mathbb{R}$, and assigning each point P in space its co-ordinate triple (x_1, x_2, x_3) defines a bijection between space and $\mathbb{R}^3 = \mathbb{R} \times \mathbb{R} \times \mathbb{R}$.

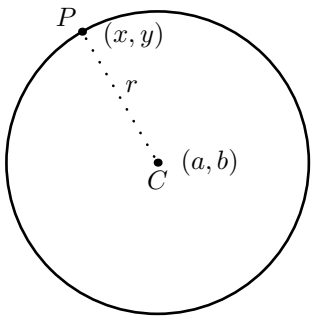
Observation 4.1. We often write (x, y) instead of (x_1, x_2) and (x, y, z) in place of (x_1, x_2, x_3) when dealing with plane geometry or spatial geometry.

The introduction of co-ordinates means that relations between points can be translated into relations between their co-ordinates.

Example 4.2. If the points in space P and Q have co-ordinates (x_1, x_2, x_3) and (y_1, y_2, y_3) respectively then the distance, d, between P and Q is given by

$$d = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + (x_3 - y_3)^2} = \sqrt{\sum_{j=1}^{3}(x_j - y_j)^2}.$$

If C is the point in the plane with co-ordinates (a, b), then the circle of radius r with centre C comprises all points P whose co-ordinates (x, y) satisfy the equation $(x - a)^2 + (y - b)^2 = r^2$.



Conversely, this also means that we can give equations in two variables a geometric interpretation: Given any three real numbers a, b, c with either a or b non-zero, the set of all points P in the plane whose co-ordinates (x, y) satisfy the equation

$$ax + by = c$$

comprise *a* line, ℓ, in the plane.

Indeed, every line in the plane is obtained in this manner, with different lines corresponding to *essentially different* equations, where we consider two such equations to be *essentially the same* if one can be obtained form the other by multiplying through by a non-zero constant.

This indicates why equations of the above form are called *linear equations*.

Continuing in this vein, we take fixed real numbers a, b, c, d, e and f, and consider the *system of linear equations*

$$ax + by = e$$
$$cx + dy = f$$

where we assume that either $a \neq 0$ or $b \neq 0$ and that either $c \neq 0$ or $d \neq 0$.

If we take a solution to be the co-ordinates of a point P in the plane, then the set of all solutions has a geometric interpretation.

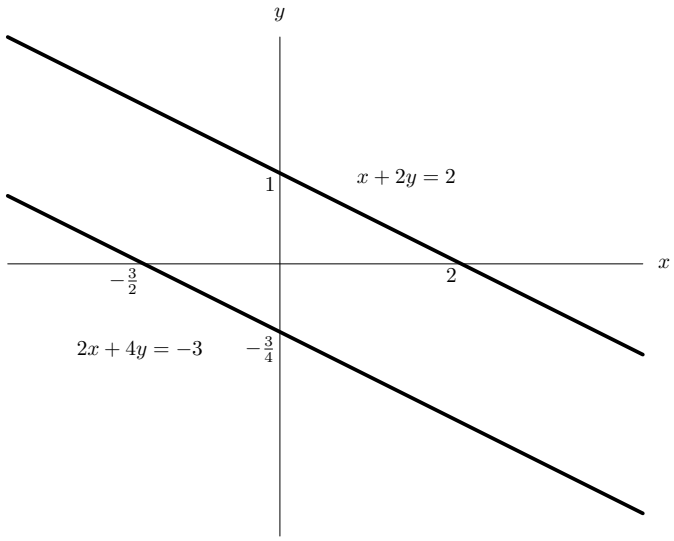
If the solutions to the first equation comprise the co-ordinates of the points of the line ℓ_1 and the solutions to the second equation are the co-ordinates of the point of the ℓ_2, then there are several possibilities.

(i) ℓ_1 and ℓ_2 represent the same line. In this case there are infinitely many solutions to the system of equations.

(ii) ℓ_1 and ℓ_2 represent parallel but distinct lines. In this case there is no solution of the system.

(iii) ℓ_1 and ℓ_2 are not parallel. In this case they have a unique point of intersection, P, whose co-ordinates are the unique solution to our system of linear equations.

We note that cases (i) and (ii) correspond to the relation $ad - bc = 0$.



The situation is similar, but slightly more involved, when we move to spatial geometry.

If we take real numbers a, b, c, and j with at least one of a, b, c non-zero, then points whose co-ordinates, (x, y, z), comprise all solutions of the linear equation

$$ax + by + cz = j$$

form a *plane*, and every plane arises this way.

A given line is represented by a system of two linear equations with the property that the planes they represent intersect in the given line.

A new possibility arises for two lines in space, say ℓ_1 and ℓ_2, in addition to the three listed above: ℓ_1 and$_2$ could be *skew* — they do not meet despite not being parallel. Since each line in space is determined by two equations, two lines require, in general, four equations. Thus the system of equations we obtain comprises four linear equations in three unknowns – an *over-determined system* – and it is possible that any three have a common solution without the four having any solution, as the next example shows

Example 4.3.

$$\begin{array}{rcrcrcl} \mathrm{x} & & & & & = & 0 \\ & & \mathrm{y} & & & = & 0 \\ & & & & \mathrm{z} & = & 0 \\ \mathrm{x} & + & \mathrm{y} & + & \mathrm{z} & = & 1 \end{array}$$

The reader is invited to try to represent what we do geometrically, bearing in mind that our geometric representation is both illuminating and misleading. It is simple to avoid many pitfalls by remembering that the geometric representation depends intimately on properties of the real numbers and that in other situations those features which depend upon properties specific to $\mathbb{R}$ are not available.

A frequently useful heuristic guide is to regard a field as corresponding to a "generalised line", the set of all ordered pairs of elements of the field as corresponding to a "generalised plane", the set of all ordered triples f elements of the field as corresponding to a "generalised (3-)space", and so on. But in doing so, the reader must be mindful that not all properties generalise from the case when the field in question is $\mathbb{R}$.

4.1 Exercises

The purpose of these exercises is to provide practice in moving between equational, parametric and vectorial representations of lines, planes, etc. The questions are formulated in their general form.The reader who has difficulty with such generality should first attempt a numerical example, by taking, say, $a = 3, b = 4$ and $c = 5$.

Exercise 4.1. Choose a co-ordinate system for the plane, with origin $\mathbf{0}$.

Let P have co-ordinates (x, y) and A have co-ordinates (a, b).

Let θ be the acute angle between the line through $\mathbf{0}$ and A and the line through $\mathbf{0}$ and P.

Prove that

$$\cos\theta = \frac{ax + by}{\sqrt{a^2 + b^2}\sqrt{x^2 + y^2}}.$$

Exercise 4.2. Choose a co-ordinate system for the plane.

Let ℓ be the line comprising all points P in the plane whose co-ordinates (x, y) satisfy the equation $ax + by = c$, where either $a \neq 0$ or $b \neq 0$, or, equivalently, $a^2 + b^2 \neq 0$.

Find the co-ordinates of the point P_ℓ on ℓ which is closest to the origin, $\mathbf{0}$.

Exercise 4.3. A *parametric representation* of the line ℓ is a function

$$\varphi\colon \mathbb{R} \longrightarrow \mathbb{R} \times \mathbb{R},\ t \longmapsto (x(t), y(t))$$

whose image is ℓ.

Find a parametric representation of the line ℓ in Exercise 4.2.

Exercise 4.4. Chose a co-ordinate system for the plane.

Assign to each point P a vector, its *co-ordinate vector*: If P has co-ordinates (x, y), then its co-ordinate vector is

$$\mathbf{x} := \begin{bmatrix} x \\ y \end{bmatrix}$$

This allows us to represent geometric objects and relations using vectorial equations and to interpret vector equations geometrically.

For example, taking $\mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix}$, $\mathbf{u} = \begin{bmatrix} u \\ v \end{bmatrix}$, $\mathbf{r} = \begin{bmatrix} r \\ s \end{bmatrix}$, the vector equation

$$\mathbf{x} = \mathbf{u} + \lambda\mathbf{r} \qquad (\lambda \in \mathbb{R}$$

or, equivalently,

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} u \\ v \end{bmatrix} + \lambda \begin{bmatrix} r \\ s \end{bmatrix} \qquad (\lambda \in \mathbb{R}),$$

with either $r \neq 0$ o $s \neq 0$ represents a line in the plane.

(a) Find an equation for this line.

(b) Find a vectorial representation of the line ℓ in Exercise 4.2.

The longer mathematics lives the more abstract — and therefore, possibly also the more practical — it becomes.

E. T. Bell

Chapter 5

Linear Transformations and Isomorphism

One of the key insights of 20th Century mathematics is the most effective way to study (classes of) objects of interest is to study the transformations between them. In the case of sets, we study functions between sets.

5.1 Linear Transformations

Since vector spaces are sets with additional structure, and since functions are what allows us to compare sets, it is natural to use functions which are compatible with this additional structure to compare vector spaces: They must respect addition of vectors and the multiplcation of vectors by scalars. Such functions are precisely the *linear transformations*. Formally,

Definition 5.1. Let V and W be vector spaces over the field $\mathbb{F}$.

A *linear transformation from V to W* is a function

$$T\colon V \longrightarrow W$$

such that for all $\mathbf{x}, \mathbf{y} \in V$ and $\lambda, \mu \in \mathbb{F}$

$$T(\lambda\mathbf{x} + \mu\mathbf{y}) \;=\; \lambda T(\mathbf{x}) + \mu T(\mathbf{y}).$$

Example 5.2. Take $V = \mathbb{R}^3$, $W = \mathbb{R}^2$ and

$$T\colon V \longrightarrow W, \quad (x, y, z) \longmapsto (2x + y + z, x - y)$$

Then T is a linear transformation.

Example 5.3. Let V be a real vector space over $\mathbb{R}$. Take $\mathbf{u}, \mathbf{v} \in V$, with $\mathbf{v} \neq \mathbf{0}_V$.

We saw in Exercise 4.3 on page 49 that the set

$$\ell = \{\mathbf{u} + \lambda\mathbf{v} \mid \lambda \in \mathbb{R}\}$$

comprises a line in V.

If $T\colon V \longrightarrow W$ is a linear transformation, then, for each $\lambda \in \mathbb{R}$,

$$\begin{aligned} T(\mathbf{u}+\lambda\mathbf{v}) &= T(1\mathbf{u}+\lambda\mathbf{v}) && \text{by VS5} \\ &= 1T(\mathbf{u})+\lambda T(\mathbf{v}) && \text{as } T \text{ is a linear transformation} \\ &= T(\mathbf{u})+\lambda T(\mathbf{v}) && \text{by VS5} \end{aligned}$$

Hence the image of ℓ under T is

$$T(\ell) = \{T(\mathbf{u})+\lambda T(\mathbf{v}) \mid \lambda \in \mathbb{R}\},$$

which comprises a line in W.

We have just shown that a linear transformation is a function which maps lines to lines.

Example 5.4. Put

$$\mathcal{D}(\mathbb{R}) := \{f : \mathbb{R} \longrightarrow \mathbb{R} \mid f \text{ is differentiable}\}.$$

It is an elementary result from differential calculus that $\mathcal{D}(\mathbb{R})$ is a real vector space and that the derivative defines a linear transformation

$$D\colon \mathcal{D}(\mathbb{R}) \longrightarrow \mathcal{F}(\mathbb{R}), \quad f \longmapsto f'$$

where $\mathcal{F}(\mathbb{R})$ is as defined Example 3.18 on page 36, and we have written f' for the derivative of f.

That D is a linear transformation is simply a restatement of the familiar rule from calculus that given $\alpha, \beta \in \mathbb{R}$ and $f, g \in \mathcal{D}(\mathbb{R})$,

$$\frac{d}{dx}(\alpha f+\beta g) = \alpha\frac{df}{dx}+\beta\frac{dg}{dx}$$

Example 5.5. Put

$$\mathcal{I}(\mathbb{R}) := \{f\colon \mathbb{R} \longrightarrow \mathbb{R} \mid f \text{ is integrable}\}.$$

It is an elementary result from integral calculus that $\mathcal{I}(\mathbb{R})$ is a real vector space and that for each $a \in \mathbb{R}$ integration defines a linear transformation

$$I_a\colon \mathcal{I}(\mathbb{R}) \longrightarrow \mathbb{R}, \quad f \longmapsto \int_a^x f(t)dt$$

That I_a is a linear transformation is simply a restatement of the familiar rule from calculus that given $\alpha, \beta \in \mathbb{R}$ and $f, g \in \mathcal{I}(\mathbb{R})$,

$$\int_a^x \Big(\alpha f(t)+\beta g(t)\Big)dt = \alpha\int_a^x f(t)dt+\beta\int_a^x g(t)dt$$

Example 5.6. Let V be the real vector space of all sequences of real numbers, so that

$$V = \{(x_n)_{n\in\mathbb{R}} \mid x_n \in \mathbb{R} \text{ for every } n \in \mathbb{N}\}$$

Then

$$T\colon V \longrightarrow V, \quad (x_n)_{n\in\mathbb{N}} \longmapsto (x_{n+2}-4x_{n+1}+3x_n)_{n\in\mathbb{N}}$$

is a linear transformation.

Example 5.7. Given any field $\mathbb{F}$, the linear transformation

$$T\colon \mathbb{F}[t] \longrightarrow \mathbb{F}[[t]], \quad \sum_{j=0}^{n} a_j t^j \longmapsto \sum_{j=0}^{\infty} c_j t^j$$

where

$$c_j = \begin{cases} a_j & \text{for } j \leq n \\ 0 & \text{for } j > n \end{cases}$$

explains why a polynomial may be thought of as a "finite" or "truncated" power series.

It is natural to try to determine all linear transformations between given vector spaces.

If V is a vector space over the field $\mathbb{F}$, we always have at least one linear transformation from V to itself, namely the identity map, and if W is any vector space over $\mathbb{F}$, we always have at least one linear transformation from V to W, namely, the *zero map*. Moreover, the composition of any two linear transformations is again a linear transformation.

Lemma 5.8. *Let U, V and W be vector spaces over the field $\mathbb{F}$.*

(a) The zero map

$$\mathbf{0} : V \longrightarrow W, \quad \mathbf{v} \longmapsto \mathbf{0}_W$$

is a linear transformation.

(b) The identity map

$$id_V\colon V \longrightarrow V, \quad \mathbf{v} \longmapsto \mathbf{v}$$

is a linear transformation.

(c) If $S\colon U \longrightarrow V$ and $T\colon V \longrightarrow W$ are linear transformations, so is their composition

$$T \circ S \ : U \longrightarrow W, \quad \mathbf{u} \longmapsto T(S(\mathbf{u}))$$

Proof. (a) is an immediate consequence of the fact that $\lambda\mathbf{0}_W + \mu\mathbf{0}_W = \mathbf{0}_W$ for all $\lambda, \mu \in \mathbb{F}$.

(b) is immediate from definition.

(c) Take $\mathbf{u}, \mathbf{v} \in U$ and $\lambda, \mu \in \mathbb{F}$. Then

$$\begin{aligned} (T \circ S)(\lambda\mathbf{u} + \mu\mathbf{v}) &:= T\Big(S(\lambda\mathbf{u} + \mu\mathbf{v})\Big) \\ &= T\Big(\lambda S(\mathbf{u}) + \mu S(\mathbf{v})\Big) && \text{as } S \text{ is linear} \\ &= \lambda T\big(S(\mathbf{u})\big) + \mu T\big(S(\mathbf{v})\big) && \text{as } T \text{ is linear} \\ &=: \lambda(T \circ S)(\mathbf{u}) + \mu(T \circ S)(\mathbf{v}) \end{aligned}$$

□

Lemma 5.8 shows that we always have linear transformations, but does not provide more than one or two, and certainly provides no guide to finding them all. There is a good reason for this lack: there is no method for finding all linear transformations between two arbitrary vector spaces over a given field.

However, we can classify all linear transformations between vector spaces of a specific form, namely those of the form $\mathbb{F}^n$ for $\mathbb{F}$ a field and n a counting number.

Theorem 5.9. *The function $T\colon \mathbb{F}^n \longrightarrow \mathbb{F}^m$ is a linear transformation if and only if there are $a_{ij} \in \mathbb{F}$ $(i = 1, \ldots, m,\ j = 1, \ldots, n)$ such that for all $(x_1, \ldots, x_n) \in \mathbb{F}^n$,*

$$T(x_1, \ldots, x_n) = (\sum_{j=1}^{n} a_{1j}x_j, \ldots, \sum_{j=1}^{n} a_{mj}x_j).$$

In particular $T\colon \mathbb{F} \longrightarrow \mathbb{F}$ is linear if and only if there is an $a \in \mathbb{F}$ such that $T(x) = ax$ for all $x \in \mathbb{F}$.

Proof. Suppose that there $a_{ij} \in \mathbb{F}$ $(i = 1, \ldots, m,\ j = 1, \ldots, n)$ such that for all $(x_1, \ldots, x_n) \in \mathbb{F}^N$,

$$T(x_1, \ldots, x_n) = (\sum_{j=1}^{n} a_{1j}x_j, \ldots, \sum_{j=1}^{n} a_{mj}x_j).$$

Take $(x_1, \ldots, x_n), (y_1, \ldots, y_n) \in \mathbb{F}^n$ and $\lambda, \mu \in \mathbb{F}$. Then

$$\begin{aligned}
&T\Big(\lambda(x_1, \ldots, x_n) + \mu(y_1, \ldots, y_n)\Big) \\
&\qquad = T\Big((\lambda x_1 + \mu y_1), \ldots, (\lambda x_n + \mu y_n)\Big) \\
&\qquad = \Big(\sum_{j=1}^{n} a_{1j}(\lambda x_j + \mu y_j), \ldots, \sum_{j=1}^{n} a_{mj}(\lambda x_j + \mu y_j)\Big) \\
&\qquad = \Big(\lambda \sum_{j=1}^{n} a_{1j}x_j + \mu \sum_{j=1}^{n} y_j), \ldots, \lambda \sum_{j=1}^{n} a_{mj}x_j + \mu \sum_{j=1}^{n} y_j)\Big) \\
&\qquad = \lambda \left(\sum_{j=1}^{n} a_{1j}x_j, \ldots, \sum_{j=1}^{n} a_{mj}x_j\right) + \mu \left(\sum_{j=1}^{n} a_{mj}y_j, \ldots, , \sum_{j=1}^{n} a_{mj}y_j\right) \\
&\qquad = \lambda T(x_1, \ldots, x_n) + \mu T(y_1, \ldots, y_n)
\end{aligned}$$

For the converse, suppose $T\colon \mathbb{F}^n \longrightarrow \mathbb{F}^m$ is a linear transformation.

Take $(x_1, \ldots, x_n) \in \mathbb{F}^n$. Then

$$(x_1, \ldots, x_n) = x_1(1, 0, \ldots, 0) + \cdots + x_n(0, \ldots, 0, 1),$$

so that

$$T(x_1, \ldots, x_n) = x_1T(1, 0, \ldots, 0) + \cdots + x_nT(0, \ldots, 0, 1).$$

Since $T(1, 0, \ldots, 0), \ldots, T(0, \ldots, 0, 1) \in \mathbb{F}^m$, there are $a_{11}, \ldots, a_{m1}, \ldots, a_{1n}, \ldots, a_{mn} \in \mathbb{F}$ such that

$$\begin{aligned}
T(1, 0, \ldots, 0) &= (a_{11}, \ldots, a_{m1}) \\
&\vdots \\
T(0, 0, \ldots, 1) &= (a_{1n}, \ldots, a_{mn}),
\end{aligned}$$

whence

$$\begin{aligned}
T(x_1, \ldots, x_n) &= x_1(a_{11}, \ldots, a_{m1}) + \cdots + x_n(a_{1n}, \ldots, a_{mn}) \\
&= (a_{11}x_1, \ldots, a_{m1}x_1) + \cdots + (a_{1n}x_n, \ldots, a_{mn}x_n) \\
&= \Big(\sum_{j=1}^{n} a_{1j}x_j, \ldots, \sum_{j=1}^{n} a_{mj}x_j\Big)
\end{aligned}$$

□

While we have only been able to classify all linear transformations between vector spaces of this one special form, we shall see that this is enough to classify all linear transformations between a much larger class of vector spaces.

While we cannot show this without further study of linear transformations and vector spaces, we can already begin to see how the theory we are developing applies to the examples in Chapter 2.

For by Theorem 5.9, studying the system of linear equations

$$\begin{array}{ccccccc} a_{11}x_1 & + & \cdots & + & a_{1n}x_n & = & y_1 \\ \vdots & & & & \vdots & & \vdots \\ a_{m1}x_1 & + & \cdots & + & a_{mn}x_n & = & y_m \end{array}$$

is the same as studying the linear transformation

$$T\colon \mathbb{F}^n \longrightarrow \mathbb{F}^m, \quad (x_1, \ldots, x_n) \longmapsto (y_1, \ldots, y_m)$$

where, for $1 \leq i \leq m$

$$y_i = \sum_{j=1}^{n} a_{ij}x_j$$

We next present a reformulation of the definition of linear transformation, which is sometimes more convenient to apply.

Lemma 5.10. *Let V and W be vector spaces over the field $\mathbb{F}$.*

The function $T\colon V \longrightarrow W$ is a linear transformation if and only for all $\mathbf{u}, \mathbf{v} \in V$ and $\lambda \in \mathbb{F}$,

(a) $T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$

(b) $T(\lambda\mathbf{v}) = \lambda T(\mathbf{v})$

Proof. Let $T\colon V \longrightarrow W$ be a function.

Take $\mathbf{u}, \mathbf{v} \in V$ and $\lambda, \mu \in \mathbb{F}$.

Suppose that T is a linear transformation. Then

(a)

$$\begin{array}{lll} T(\mathbf{u} + \mathbf{v}) = T(1\mathbf{u} + 1\mathbf{v}) & & \text{by VS5} \\ \quad = 1T(\mathbf{u}) + 1T(\mathbf{v}) & & \text{as } T \text{ is a linear transformation} \\ \quad = T(\mathbf{u}) + T(\mathbf{v}) & & \text{by VS5} \end{array}$$

(b)

$$\begin{array}{lll} T(\lambda\mathbf{v}) = T(\mathbf{0}_V + \lambda\mathbf{v}) & & \text{by VS2} \\ \quad = T(0\mathbf{u} + \lambda\mathbf{v}) & & \text{by Theorem 3.23} \\ \quad = 0T(\mathbf{u}) + \lambda T(\mathbf{v}) & & \text{as } T \text{ is a linear transformation} \\ \quad = \mathbf{0}_W + \lambda T(\mathbf{v}) & & \text{by Theorem 3.23} \\ \quad = \lambda T(\mathbf{v}) & & \text{by VS2} \end{array}$$

Conversely, suppose that T satisfies (a) and (b). Then

$$\begin{array}{lll} T(\lambda\mathbf{u} + \mu\mathbf{v}) = T(\lambda\mathbf{u}) + T(\mu\mathbf{v}) & & \text{by (a)} \\ \quad = \lambda T(\mathbf{u}) + \mu T(\mathbf{v}) & & \text{by (b)} \end{array}$$

□

Definition 5.11. Let V and W be vector spaces over the field $\mathbb{F}$.

The function $T\colon V \longrightarrow W$ is *additive* if and only if for all $\mathbf{u}, \mathbf{v} \in V$

$$T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$$

T is *homogeneous (of degree* 1*)* if and only if given $\lambda \in \mathbb{F}$ and $\mathbf{v} \in V$

$$T(\lambda\mathbf{v}) = \lambda T(\mathbf{v})$$

Thus, a linear transformation is a homogeneous, additive function. It is sometimes more convenient to verify the two conditions in Lemma 5.10 on the previous page separately, and at other times it is more convenient to use Definition 5.1 on page 51 directly. Lemma 5.10 on the previous page allows us to ouse whatever seems more convenient.

We continue our analysis of linear transformations with the notion of *kernel*.

Definition 5.12. The *kernel*, denoted $\ker(T)$, of the linear transformation $T\colon V \longrightarrow W$ is the subset of V comprising those vectors that are mapped to $\mathbf{0}_W$ by T.

$$\ker(T) := \{\mathbf{x} \in V \mid \ T(\mathbf{x}) = \mathbf{0}_W\}$$

The kernel of a linear transformation contains important information about it.

Before illustrating this, we recall the notions of invectiveness, surjectiveness and bijectiveness of functions and apply it to the special case of linear transformations.

Definition 5.13. The linear transformation $T : V \longrightarrow W$ is

(a) *1–1, injective* or a *monomorphism* if and only if $\mathbf{u} = \mathbf{v}$ whenever $T(\mathbf{u}) = T(\mathbf{v})$;

(b) *onto, surjective* or an *epimorphism* if and only if for each $\mathbf{w} \in W$ there is a $\mathbf{v} \in V$ with $\mathbf{w} = T(\mathbf{v})$;

(c) *1–1 and onto* or *bijective* if and only it is both 1–1 and onto;

(d) an *endomorphism* if and only if $W = V$.

Lemma 5.14. *Let $T\colon V \longrightarrow W$ be a linear transformation. Then*

(i) T is injective if and only if $\ker(T) = \{\mathbf{0}_V\}$*;*

(ii) T is surjective if and only if $\mathrm{im}(T) = W$*.*

Proof. (i). $T(\mathbf{u}) = T(\mathbf{v})$ if and only if $T(\mathbf{u} - \mathbf{v}) = \mathbf{0}_W$.

Since T is a linear transformation, this is the case if and only if $\mathbf{u} - \mathbf{v} \in \ker(T)$.

If $\ker(T) = \{\mathbf{0}_V\}$, then this occurs only if $\mathbf{u} - \mathbf{v} = \mathbf{0}_V$, that is $\mathbf{u} = \mathbf{v}$.

Conversely, suppose that T is injective and $\mathbf{v} \in \ker(T)$, so that $T(\mathbf{v}) = \mathbf{0}_W$.

Since T is a linear transformation, $T(\mathbf{0}_V) = \mathbf{0}_W$.

Since T is injective, $\mathbf{v} = \mathbf{0}_V$.

(ii) This is just a restatement of the definition. □

Observation 5.15. Determining the kernel of the linear transformation $T\colon V \longrightarrow W$ is "simply" a matter of solving the equation

$$T(\mathbf{x}) = \mathbf{0}_W$$

However, doing so can be a subtle and/or difficult problem, sometimes requiring the application or development of theory from other parts of mathematics.

Example 5.16. Let $V = \mathcal{C}^\infty(\mathbb{R})$ be the real vector space of all infinitely differentiable real-valued functions of a real variable.

Then

$$T\colon \mathcal{C}^\infty(\mathbb{R}) \longrightarrow \mathcal{C}^\infty(\mathbb{R}), \quad f \longmapsto \frac{d^2 f}{dx^2} - 4\frac{df}{dx} + 4f$$

is a linear transformations, and determining its kernel comprises finding all solutions to the differential equation

$$\frac{d^2 y}{dx^2} - 4\frac{dy}{dx} + 4y = 0$$

Example 5.17. To find the kernel of the linear transformation in Example 5.6 on page 52 comprises finding all solutions of the difference eqiuation

$$x_{n+2} - 4x_{n+1} + 3x_n = 0$$

in Example 2.4 on page 23.

5.2 Isomorphism

We consider two vector spaces over $\mathbb{F}$ to be essentially the same when the only differences between them are purely notational. We formulate rigorously using the language of linear transformations, including the notion of *isomorphism*.

Definition 5.18. The linear transformation $T\colon V \longrightarrow W$ is an *isomorphism* if and only if there is a linear transformation $S\colon W \longrightarrow V$ such that

$$S \circ T = id_V \qquad \text{and} \qquad T \circ S = id_W.$$

In such a case, S is the *inverse linear transformation to* T.

The vector spaces V and W over the field $\mathbb{F}$ are *isomorphic* if and only if there is an isomorphism $T\colon V \longrightarrow W$.

We write $V \cong W$ when V is isomorphic to W.

An *automorphism* is an endomorphism $T\colon V \longrightarrow V$ which is also an isomorphism.

Since the linear function S in Definition 5.18 is, in particular, a function, we se that if there is such a linear transformation, if must be the inverse function to T. In other words, a necessary condition for T to be an isomorphism is that it be an invertible function.

It follows that the only possible inverse linear transformation to the linear transformation $T\colon V \longrightarrow W$ is the inverse function to T, which, by Theorem 1.46 on page 12, exists if and only if T is bijective.

Theorem 5.19. *A linear transformation $T\colon V \longrightarrow W$ is an isomorphism if and only if it is bijective.*

Proof. Since a function has an inverse if and only if it is bijective (Theorem 1.46 on page 12), it is sufficient to show that if the linear transformation $T\colon V \longrightarrow W$ has an inverse *function*, $S\colon W \longrightarrow V$, then S must also be a linear transformation.

Let $S\colon W \longrightarrow V$ be the inverse function to the linear transformation $T\colon V \longrightarrow W$. Take $\mathbf{u}, \mathbf{w} \in W$ and $\lambda, \mu \in \mathbb{F}$. Then

$$S(\lambda\mathbf{u} + \mu\mathbf{w}) = S\big(\lambda(T \circ S)(\mathbf{u}) + \mu(T \circ S)(\mathbf{w})\big) \qquad \text{as } T \circ S = id_W$$

$$
\begin{aligned}
&= S\big(\lambda T(S(\mathbf{u})) + \mu T(S(\mathbf{w}))\big) \\
&= S\big(T(\lambda S(\mathbf{u}) + \mu S(\mathbf{w}))\big) && \text{as } T \text{ is linear} \\
&= (S \circ T)(\lambda S(\mathbf{u}) + \mu S(\mathbf{w})) \\
&= \lambda S(\mathbf{u}) + \mu S(\mathbf{w}) && \text{as } S \circ T = id_V
\end{aligned}
$$

□

While this theorem provides a satisfactory intrinsic criterion for deciding whether a given linear transformation is an isomorphism, it does not do the same for deciding whether two given vector spaces are isomorphic. After all, we still face the unwieldy and, in principle, infinite task of finding an isomorphism between the vector spaces in question.

It need not be obvious that two given vector spaces are isomorphic.

Example 5.20. The set of all solutions, $f\colon \mathbb{R} \longrightarrow \mathbb{R}$ which solve the differential equation

$$\frac{d^2 f}{dx^2} + f = 0$$

is a real vector space which is isomorphic with the real vector space $\mathbb{C}$.

We shall see (Theorem 8.3 on page 87) that we can decide whether two vector spaces are isomorphic by means of a single numerical invariant, the *dimension* of a vector space.

Example 5.21. Let V be the real vector space of all sequences of real numbers, so that

$$V = \{(x_n)_{n\in\mathbb{N}} \mid x_n \in \mathbb{R} \text{ for all} n \in \mathbb{N}\}$$

and $\mathcal{F}(\mathbb{N})$ the real vector space of all real-valued functions defined on $\mathbb{N}$, the set of all natural numbers.

The function

$$T\colon \mathcal{F}(\mathbb{N}) \longrightarrow V, \quad f \longmapsto \big(f(n)\big)_{n\in\mathbb{N}}$$

is an isomorphism of real vector spaces.

5.3 Exercises

Exercise 5.1. Show that

$$f\colon \mathbb{R}^2 \longrightarrow \mathbb{R}^3, \quad (x, y) \mapsto (2x + y, 4x + 17y, 3x - 4y)$$

is a linear transformation of vector spaces over $\mathbb{R}$.

Exercise 5.2. Let $V := \mathcal{C}^\infty(\mathbb{R})$ be the real vector space of all infinitely differentiable functions $f\colon \mathbb{R} \longrightarrow \mathbb{R}$, so that

$$V := \mathcal{C}^\infty(\mathbb{R}) := \{f\colon \mathbb{R} \longrightarrow \mathbb{R} \mid \frac{d^n f}{dx^n} \text{ is continuous for every } n \in \mathbb{N}\}$$

with the vector space operations defined point-wise, as in Example 3.18 on page 36.

Prove that

$$D\colon V \longrightarrow V, \quad f \longmapsto \frac{df}{dx}$$

is a linear transformation.

Exercise 5.3. Prove that if the function

$$f\colon \mathbb{R}^2 \longrightarrow \mathbb{R}^2$$

is a linear transformation of real vector spaces, then there are uniquely determined real numbers a, b, c, d such that for all $(x, y) \in \mathbb{R}^2$,

$$f(x, y) = (ax + by, cx + dy)$$

Exercise 5.4. Let $\mathcal{I}(\mathbb{R})$ be the real vector space of all integrable functions $f\colon \mathbb{R} \longrightarrow \mathbb{R}$, so that

$$\mathcal{I}(\mathbb{R}) = \{f\colon \mathbb{R} \longrightarrow \mathbb{R} \mid f \text{ is integrable}\}.$$

Prove that for each $a \in \mathbb{R}$

$$I_a\colon \mathcal{I}(\mathbb{R}) \longrightarrow \mathbb{R}, \quad f \longmapsto \int_a^x f(t)dt$$

is a linear transformation.

Exercise 5.5. We use the fact from the elementary theory of ordinary differential equations (as presented in MATH102) that the function $f\colon \mathbb{R} \longrightarrow \mathbb{R}$ solves the differential equation

$$\frac{d^2 f}{dx^2} + f = 0$$

if and only if there a $A, B \in \mathbb{R}$ such that for all $x \in \mathbb{R}$

$$f(x) = A\cos x + B\sin x$$

Let V be the real vector space of all solutions to this ordinary differential equation.

Prove that

$$T\colon V \longrightarrow \mathbb{C}, \quad A\cos x + B\sin x \longrightarrow A + iB,$$

where we have written $A\cos x + B\sin x$ for the fnction

$$f\colon \mathbb{R} \longrightarrow \mathbb{R}, \quad x \longmapsto A\cos x + B\sin x,$$

is an isomorphism of real vector spaces.

Exercise 5.6. Let V be the real vector space of all sequences of real numbers, so that

$$V = \{(x_n)_{n\in\mathbb{N}} \mid x_n \in \mathbb{R} \text{ for all } n \in \mathbb{N}\}$$

Let $\mathcal{F}(\mathbb{N})$ be the real vector space of all real-valued functions defined on $\mathbb{N}$.

Prove that the function

$$T\colon \mathcal{F}(\mathbb{N}) \longrightarrow V, \quad f \longmapsto \big(f(n)\big)_{n\in\mathbb{N}}$$

is an isomorphism of real vector spaces.

Beauty is the first test: there is no permanent place in the world for ugly mathematics.

G. H. Hardy

Chapter 6

Deriving Vector Spaces from Given Ones

Now that we know what a vector space is and have met enough examples to show that the concept is not an empty one, we investigate the problem of deriving vector spaces from given ones.

6.1 Vector Subspaces

One possibility is to start with a vector space, say V, and to consider subsets of V. When is a subset U of V form a vector space in its own right?

It is tempting to call a subset U of V that is a vector space in its own right a *vector subspace of* V. The defects of such a definition are readily illustrated by concrete examples, which ultimately suggest a more useful definition

Example 6.1. Take the set of real numbers $\mathbb{R}$ with its usual structure as $\mathbb{R}$-vector space. Then the set of rational numbers, $\mathbb{Q}$, clearly forms a subset.

Being a field, $\mathbb{Q}$ is a vector space in its own right, but only over $\mathbb{Q}$. It is **not** a vector space over $\mathbb{R}$ with respect to the usual addition and multiplication, for multiplying together a rational number by a real number need not result in a rational number: $\sqrt{2}.2$ is not rational

In fact, there is no way whatsoever to make $\mathbb{Q}$ a vector space over $\mathbb{R}$, because, as we shall see later, a vector space over $\mathbb{R}$ has either precisely one element, or has at least as many elements as $\mathbb{R}$. But it is a basic result from set theory, that this is not true of $\mathbb{Q}$.

The salient feature this example was that the two vector space structures in question have different fields of scalars. So we need to insist that the two vector spaces have a common field of scalars.

But even this is not enough, as we now show.

Example 6.2. We take $\mathbb{Q}$, the field of rational numbers, as the field of scalars.

Example 3.13 on page 35 showed that

$$\mathbb{Q} \times \mathbb{Q} = \{(x, y) \mid x, y \in \mathbb{Q}\}$$

is a vector space over $\mathbb{Q}$

We construct V, a vector space over $\mathbb{Q}$, which is a subset of $\mathbb{Q}\times\mathbb{Q}$, but whose vector space structure is not related to that of $\mathbb{Q} \times \mathbb{Q}$.

Our example requires the following elementary facts from basic number theory.

1. Any two integers, say x and y, which are not both 0, have a *greatest common divisor*[1] denoted by $\gcd(x, y)$. This is defined to be the (uniquely determined) positive integer d such
   (a) d divides both x and y and
   (b) if the integer c divides both x and y, then c divides d.
2. The integers x and y are said to be *relatively prime* if and only if their greatest common divisor is 1.
3. Given any integers, x and y, not both 0, with $\gcd(x, y) = d$, there are (uniquely determined) relatively prime integers u and v with $x = du$ and $y = dv$.

Put $V := \{(x, y) \in \mathbb{Z}^2 \mid y > 0 \text{ and } \gcd(x, y) = 1\}$.

Since $\mathbb{Z} \subset \mathbb{Q}$,

$$V \subset \mathbb{Z} \times \mathbb{Z} \subset \mathbb{Q} \times \mathbb{Q}$$

Define

$$\boxplus\colon V \times V \longrightarrow V$$
$$\boxdot\colon \mathbb{Q} \times V \longrightarrow V$$

by

$$(u, v) \boxplus (x, y) := (r, s) \qquad \text{where } rvy = s(uy + vx) \text{ with } \gcd(r, s) = 1$$

$$\frac{p}{q} \boxdot (x, y) := (r, s) \qquad \text{where } rqy = spx \text{ with } \gcd(r, s) = 1.$$

One way to see that V is a vector space over $\mathbb{Q}$ is to verify the vector space axioms by direct computation. This is to be recommended to those who are still wary of abstract methods and feel more comfortable with brute-force computation.

Alternatively, observe that if we rewrite $(u, v) \in V$ as $\frac{u}{v}$, then we see that V essentially consists of the set of all rational numbers, written in *reduced form*, and that $\boxplus$ and $\boxdot$ are just the usual addition and multiplication of rational numbers re-written in the form appropriate to V.

Thus the statement: "V is a vector space over $\mathbb{Q}$" is just a restatement of the fact that every field is a vector space over itself (cf. Example 3.13 on page 35).[2]

But the two vector space structures, have nothing to do with each other, beyond having a common field of scalars.

For a meaningful and useful notion of a vector subspace we require not merely that the subset form a vector space in its own right over the same field, but that this vector space structure be precisely that derived from the ambient vector space.

One way of ensuring this is to insist that the inclusion of the subspace be a linear transformation.

[1] It is also called their *highest common factor* and denoted by $\mathrm{hcf}(x, y)$.

[2] Strictly speaking, we have shown that

$$T\colon V \longrightarrow \mathbb{Q}, \quad (u, v) \longmapsto \frac{u}{v}$$

is an isomorphism, so that V is isomorphic to $\mathbb{Q}$ as vector space over $\mathbb{Q}$.

Definition 6.3. The subset U of the vector space V over the field $\mathbb{F}$ is a *vector subspace of* V if and only if the inclusion

$$i_U^V : U \longrightarrow V, \quad \mathbf{u} \longmapsto \mathbf{u}$$

is a linear transformation.

We write $U \leq V$ to denote that U is a vector subspace of V.

Theorem 6.4. *Let U be a subset of the $\mathbb{F}$-vector space V. Then the following are equivalent.*

(i) U is a vector subspace of V.

(ii) Given $\mathbf{u}, \mathbf{u}' \in U$ and $\lambda, \mu \in \mathbb{F}$, $\lambda\mathbf{u} + \mu\mathbf{u}' \in U$.

(iii) (a) Given $\mathbf{u}, \mathbf{u}' \in U$, $\mathbf{u} + \mathbf{u}' \in U$.
(b) Given $\mathbf{u} \in U$ and $\lambda \in \mathbb{F}$, $\lambda\mathbf{u} \in U$.

Proof. Since the inclusion function

$$i_U^V : U \longrightarrow V$$

is defined by

$$i_U^V(\mathbf{x}) = \mathbf{x}$$

for all $\mathbf{x} \in U$, the equivalence of (i) and (ii) is just the definition of what it means for i_U^V to be a linear transformation (cf. Definition 5.1 on page 51).

Similarly, the equivalence of (ii) and (iii) is just the restatement of Lemma 5.10 on page 55 for i_U^V. □

The advantage of Theorem 6.4 is that in order to determine whether the subset U of the vector space V is, in fact, a vector subspace, it is not necessary to first establish that U is, itself, a vector space with the same field of scalars as V. It is enough t show that if we apply the vector space operations in V to elements of U, the resulting vectors are again elements of U. This is sometimes expressed by saying that U *is closed under the vector space operations on* V.

Example 6.5. If we consider $\mathbb{C}$ and $\mathbb{R}$ as vector spaces over $\mathbb{Q}$ in the usual manner, then

$$\mathbb{Q} \leq \mathbb{R} \leq \mathbb{C}$$

Example 6.6. The real vector space of all solutions, $f : \mathbb{R} \longrightarrow \mathbb{R}$, of the differential equation

$$\frac{d^2y}{dx^2} - 4\frac{dy}{dx} + 4y = 0$$

is a vector subspace of $\mathcal{C}^\infty(\mathbb{R})$, the real vector space of all infinitely differentiable real-valued functions of a real variable, which is, in turn, a vector subspace of $\mathcal{F}(\mathbb{R})$, the real vector space of all real-valued functions of a era variable.

Example 6.7. Possibly the most significant single example of a vector subspace of the vector space V requires the following definition.

Given $\mathbf{v} \in V$,

$$\mathbb{F}\mathbf{v} \;:=\; \{\lambda\mathbf{v} \mid \lambda \in \mathbb{F}\}.$$

For each $\mathbf{v} \in V$, $\mathbb{F}\mathbf{v}$ is a vector subspace of V, as proven in the next lemma.

Lemma 6.8. *Let V be a vector space over $\mathbb{F}$.*

For each $\mathbf{v} \in V$, $\mathbb{F}\mathbf{v}$ is a vector subspace of V.

Proof. Take $\mathbf{u}, \mathbf{u}' \in \mathbb{F}\mathbf{v}$ and $\lambda, \mu \in \mathbb{F}$.

Then there are $\alpha, \beta \in \mathbb{F}$ with $\mathbf{u} = \alpha\mathbf{v}$ and $\mathbf{u}' = \beta\mathbf{v}$.

Thus

$$\lambda\mathbf{u} + \beta\mathbf{u}' = \lambda(\alpha\mathbf{v}) + \mu(\beta\mathbf{v}) = (\lambda\alpha)\mathbf{v} + (\mu\beta)\mathbf{v} = \nu\mathbf{v},$$

where $\nu := \lambda\alpha + \mu\beta \in \mathbb{F}$. □

The vector subspace $\mathbb{F}\mathbf{v}$ of the vector space V generalises the notion of a line through the origin in $\mathbb{R}^n$, for such a line is determined uniquely by a point on it, other than the origin.

If this point has co-ordinates, $(a_1, \ldots, a_n)$, then the line is the set

$$\begin{aligned}\mathbb{R}((a_1, \ldots, a_n)) &= \{(\lambda a_1, \ldots, \lambda a_n) \mid \lambda \in \mathbb{R}\} \\ &= \{(x_1, \ldots, x_n) \in \mathbb{R}^n \mid x_\imath = \lambda a_\imath (\imath = 1, \ldots n) \text{ for some } \lambda \in \mathbb{R}\}\end{aligned}$$

and this is just the familiar *parametric representation* of the given line.

Example 6.9. Given a non-empty set X and a field $\mathbb{F}$, the vector space of all functions $f\colon X \longrightarrow \mathbb{F}$ such that $f(x) = 0$ for all but finitely many $x \in X$ (cf. Exercise 3.12 on page 43) is a vector subspace of the vector space of all functions $f\colon X \longrightarrow \mathbb{F}$ (cf. Example 3.18 on page 36).

Example 6.10. Let V, W be vector spaces over the field $\mathbb{F}$ and $T\colon V \longrightarrow W$ a linear transformation.

Then $\ker(T)$ is a vector subspace of V and $\mathrm{im}(T)$ is a vector subspace of W.

Given a family of vector subspaces of a fixed vector space V, their intersection is again a vector subspace of V.

Theorem 6.11. *Let V be a vector space.*

If for each $\gamma \in \Gamma$, W_γ is a vector subspace of V, then

$$W := \bigcap_{\gamma \in \Gamma} W_\gamma$$

is a vector subspace of V.

Proof. Take $\mathbf{u}, \mathbf{v} \in W$ and $\lambda \in \mathbb{F}$.

Then $\mathbf{u}, \mathbf{v} \in W_\gamma$ for each $\gamma \in \Gamma$.

Since each W_γ is a vector subspace of V, $\mathbf{u} + \mathbf{v} \in W_\gamma$ and $\lambda\mathbf{u} \in W_\gamma$ for each $\gamma \in \Gamma$.

It follows that $\mathbf{u} + \mathbf{v}, \lambda\mathbf{u} \in \bigcap_{\gamma \in \Gamma} W_\gamma = W$. □

While the intersection of vector subspaces of a given vector space is again a vector subspace, the same is not true of the union of vector subspaces, as the next example shows.

Example 6.12. Consider $\mathbb{R}^2$ as vector space over $\mathbb{R}$.

Then $U := \{(x, 0) \mid x \in \mathbb{R}\}$ and $V := \{(0, y) \mid y \in \mathbb{R}\}$ are both vector subspaces of $\mathbb{R}^2$, and

$$U \cup V = \{(x, y) \in \mathbb{R}^2 \mid x = 0 \text{ or } \ y = 0\}$$

which is not a vector subspace of $\mathbb{R}^2$.

For while $(1, 0), (0, 1) \in U \cup V$, their sum, $(1, 0) + (0, 1) = (1, 1)$, is not an element of $U \cup V$.

However, given vector subspaces U, W of the vector space V, the subset of V comprising those vectors in V that can be written as the sum of a vector from U and one from W is, in fact, a vector subspace of V.

Definition 6.13. Let U and W be subsets of the vector subspace V.

Their *sum*, $U + W$, is defined by

$$U + W \;:=\; \{\mathbf{u} + \mathbf{w} \mid \mathbf{u} \in U, \mathbf{w} \in W\}$$

Example 6.14. Take $\mathbb{R}^3$ with its usual structure as real vector space.

Put

$$U := \{(x, 0, 1) \mid x > 0\} \qquad \text{and} \qquad W := \{(0, y, 0) \mid y \in \mathbb{R}\}$$

Then

$$U + W = \{(x, y, 1) \mid x, y \in \mathbb{R},\ x > 0\}$$

Lemma 6.15. *Let U and W be vector subspaces of the vector subspace V.*

Then $U + W$ is a vector subspace of V.

Proof. Take $\mathbf{u}_1 + \mathbf{w}_1, \mathbf{u}_2 + \mathbf{w}_2 \in U + W$ and $\lambda \in \mathbb{F}$. Then

$$(\mathbf{u}_1 + \mathbf{w}_1) + (\mathbf{u}_2 + \mathbf{w}_2) = \mathbf{u}_1 + \mathbf{u}_2 + \mathbf{w}_1 + \mathbf{w}_2 = \mathbf{u} + \mathbf{w},$$

where $\mathbf{u} := \mathbf{u}_1 + \mathbf{u}_2 \in U$ and $\mathbf{w} := \mathbf{w}_1 + \mathbf{w}_2 \in W$, and

$$\lambda(\mathbf{u}_1 + \mathbf{w}_1) = \lambda\mathbf{u}_1 + \lambda\mathbf{w}_1 = \mathbf{u}' + \mathbf{w}',$$

where $\mathbf{u}' := \lambda\mathbf{u}_1 \in U$ and $\mathbf{w}' = \lambda\mathbf{w}_1 \in W$. □

This construction can be generalised; every subset S of the vector space V *generates* a unique vector subspace of V.

Definition 6.16. The *vector subspace of the vector space V generated by S*, S of V, denoted $\langle S \rangle$, is the *smallest* vector subspace of V containing S. In other words,

(i) $\langle S \rangle \leq V$

(ii) Given $W \leq V$, if $S \subseteq W$, then $\langle S \rangle \subseteq W$.

The elements of S are called *generators*, and S a *generating set* for $\langle S \rangle$.

We also write $\langle \mathbf{v}_1, \ldots \rangle$ when $S := \{\mathbf{v}_1, \ldots\}$.

Theorem 6.11 on the facing page is the key to proving that there is such a vector subspace of V, and that it is unique.

Theorem 6.17. *Let S be a subset of the vector space V.*

Then the vector subspace of V generated by S is the intersection of all vector subspaces U of V with $S \subseteq U$.

Proof. Put $\mathfrak{A} := \{W \leq V \mid S \subseteq W\}$.

Then $\mathfrak{A} \neq \emptyset$, as $V \in \mathfrak{A}$.

Put $U := \bigcap_{W \in \mathfrak{A}} W$.

By Theorem 6.11 on page 64, U is a vector subspace of V, and clearly $S \subseteq U$.

Take $W \leq V$ with $S \subseteq W$.

Then, by definition, $W \in \mathfrak{A}$, so that $U = \bigcap_{X \in \mathfrak{A}} X \subseteq W$.

It follows that our U is the smallest vector subspace of V containing all the elements of S. In other words,

$$\langle S \rangle = \bigcap \{W \leq V \mid S \subseteq W\}$$

□

Corollary 6.18. *U is a vector subspace of V if and only if $\langle U \rangle = U$.*

Example 6.19. Take $\mathbb{R}^3$ with its usual structure as real vector space.

Then $S := \{(1,0,0),(0,0,3)\} \subset \mathbb{R}^3$ and

$$\langle (1,0,0),\ (0,0,3) \rangle = \{(x,0,z) \mid x,z \in \mathbb{R}\}$$

Observation 6.20. If U and W are vector subspaces of V then $U + W = \langle U \cup W \rangle$.

A linear transformation $T\colon V \longrightarrow W$ naturally determines both a vector subspace of V and a vector subspace of W.

Theorem 6.21. *Let $T : V \longrightarrow W$ be a linear transformation. Then*

(i) $\ker(T)$ *is a vector subspace of V, and*

(ii) $\mathrm{im}(T)$ *is a vector subspace of W.*

Proof. Exercise. □

When working with vectors spaces, it is always useful and often important to find convenient generating sets. For example, the values of a linear transformation on a generating set completely determines al its values, as we shall later.

Definition 6.22. The vector space V is *finitely generated* if and only if there is a finite subset, S, of V with $\langle S \rangle = V$.

Otherwise V is *infinitely generated*.

When $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$, we often write

$$V = \langle \mathbf{v}_1, \ldots, \mathbf{v}_n \rangle$$

Example 6.23. $\mathbb{R}^3$ is finitely generated as real vector space, for it is generated, as real vector space by

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

Observation 6.24. Definition 6.22 on the facing page, in effect, divides the class of all vectors spaces over the field $\mathbb{F}$ into two subclasses, the finitely generated ones and the infinitely generated ones.

This distinction is actually quite significant, for in the case of finitely generated vector spaces, we can carry out computations involving linear transformations using the algebra of matrices. This algebra is developed in these notes as an application of general properties of linear transformations between vector spaces in the special case of finitely generated spaces.

We have chosen this approach for two main reasons.

In the first place, the main results about linear transformations between vector spaces are no more difficult to prove in general than in the special case of finitely generated vector spaces. This means that they can be applied equally to a broader range of situations.

Secondly, when matrices are introduced in the usual *ad hoc* manner, the definitions and the restrictions needed for them are mysterious and usually unexplained. Our approach makes them clear and natural, and many properties which are usually proven with extended calculation, follow without any calculation.

6.2 The Direct Sum of Vector Spaces

Let V and W be vector spaces over the field $\mathbb{F}$. We construct an $\mathbb{F}$-vector space structure on their Cartesian product, $V \times W$, which reflects the vector space structures on V and W.

Definition 6.25. The *direct sum*, $V \oplus W$, of the vector spaces V and W over the field $\mathbb{F}$, is the set $V \times W = \{(\mathbf{v}, \mathbf{w}), \mid \mathbf{v} \in V, \mathbf{w} \in W\}$, together with the operations defined by

$$\begin{aligned} (\mathbf{v}_1, \mathbf{w}_1) \boxplus_{V\oplus W} (\mathbf{v}_2, \mathbf{w}_2) &:= (\mathbf{v}_1 \boxplus_V \mathbf{v}_2, \mathbf{w}_1 \boxplus_W \mathbf{w}_2) && (\mathbf{v}_1, \mathbf{v}_2 \in V,\ \mathbf{w}_1, \mathbf{w}_2 \in W) \\ \lambda \boxtimes_{V\oplus W} (\mathbf{v}, \mathbf{w}) &:= (\lambda\mathbf{v}, \lambda\mathbf{w}) && (\lambda \in \mathbb{F},\ \mathbf{v} \in V,\ \mathbf{w} \in W) \end{aligned}$$

Theorem 6.26. *Given vector spaces V and W over $\mathbb{F}$, $V \oplus W$ is a vector space over $\mathbb{F}$ with respect to the operations in Definition 6.25 defined, with $\mathbf{0}_{V\oplus W} = (\mathbf{0}_V, \mathbf{0}_W)$ and $-(\mathbf{v}, \mathbf{w}) = (-\mathbf{v}, -\mathbf{w})$.*

Proof. The proof is routine verification.

We illustrate this by verifying VS2.

Take $(\mathbf{v}, \mathbf{w}) \in V \oplus W$. Then

$$\begin{aligned} (\mathbf{v}, \mathbf{w}) \boxplus_{V\oplus W} \mathbf{0}_{V\oplus W} &= (\mathbf{v}, \mathbf{w}) \boxplus_{V\oplus W} (\mathbf{0}_V, \mathbf{0}_W) \\ &= (\mathbf{v} \boxplus_V \mathbf{0}_V,\ \mathbf{w} \boxplus_W \mathbf{0}_W) \\ &= (\mathbf{v}, \mathbf{w}) \\ &= (\mathbf{0}_V \boxplus_V \mathbf{v},\ \mathbf{0}_W \boxplus_W \mathbf{w}) \\ &= (\mathbf{0}_V, \mathbf{0}_W) \boxplus_{V\oplus W} (\mathbf{v}, \mathbf{w}) \end{aligned}$$

The rest is left as an exercise. □

Example 6.27. Let $V = W = \mathbb{F}$ with its usual structure as $\mathbb{F}$-vector space.

Then $V \oplus W = \{(x, y) \mid x, y \in \mathbb{F}\}$ with the vector space structure in Definition **??** on page ?? yields precisely $\mathbb{F}^2$ with its usual structure as $\mathbb{F}$-vector space.

The direct sum of given vector spaces over the same field is a new vector space constructed from two given ones. This raises the question: Is a given vector space "in effect" non-trivial direct sum of vector spaces?

Definition 6.28. The vector space V is a *(non-trivial) direct sum* if and only if there are non-zero vector spaces U and W such that V and $U \oplus W$ are isomorphic.

Example 6.29. Take $V := \mathbb{R}^3$, $U := \mathbb{R}^2$ and $W := \mathbb{R}$ with their usual real vector space structures. Then

$$U \oplus W = \{\big((x, y), z\big) \mid (x, y) \in \mathbb{R}^2, z \in \mathbb{R}\}$$

and

$$T\colon V \longrightarrow U \oplus W, \quad (u, v, w) \longmapsto \big((u, v), w\big)$$

is an isomorphism of real vector spaces.

More generally, let $\mathbb{F}$ be a field ad take $m, n \in \mathbb{N}$. Then

$$\mathbb{F}^{m+n} \cong \mathbb{F}^m \oplus \mathbb{F}^n$$

Observation 6.30. A subtle, but important, point is illustrated by Examples 6.27 on the previous page and 6.29.

In the first case, we have an equality, $\mathbb{R}^2 = \mathbb{R} \oplus \mathbb{R}$.

In the latter, we have only an isomorphism $\mathbb{R}^3 \cong \mathbb{R}^2 \oplus \mathbb{R}$.

While isomorphism is enough for many applications, difficulties can arise when isomorphic vector spaces are treated as if they were equal, for many calculations depend on the particular isomorphism chosen. This is a source of complications when using matrices, as we shall see later.

6.2.1 Internal Direct Sum

Of particular importance is the case when U and W in Definition 6.28 may be chosen to be subspaces of V.

We first need to introduce convenient notation.

Definition 6.31. Let A, B be subsets of the vector space V. Then

$$A + B := \{\mathbf{x} + \mathbf{y} \mid \mathbf{x} \in A, \mathbf{y} \in B\}$$

When A is a singleton set, say $A = \{\mathbf{v}\}$, it is customary to write $A + B$ as

$$\mathbf{v} + B$$

Theorem 6.32. *Let U, W be non-trivial vector subspaces of the vector space V.*
If $U + W = V$ and $U \cap W = \{\mathbf{0}_V\}$, then V is isomorphic with $U \oplus W$.

Proof. Suppose that $V = U + W$ and that $U \cap W = \{\mathbf{0}_V\}$.

Then each $\mathbf{v} \in V$ can be written uniquely as $\mathbf{u} + \mathbf{w} \in U + W$.

For if $\mathbf{u} + \mathbf{w} = \mathbf{u}' + \mathbf{w}'$ with $\mathbf{u}, \mathbf{u}' \in U$ and $\mathbf{w}, \mathbf{w}' \in W$, then

$$\mathbf{u} - \mathbf{u}' = \mathbf{w}' - \mathbf{w}$$

As $\mathbf{u} - \mathbf{u}' \in U$ and $\mathbf{w}' - \mathbf{w} \in W$,

$$\mathbf{u} - \mathbf{u}' = \mathbf{w}' - \mathbf{w} \in U \cap W$$

As $U \cap W = \{\mathbf{0}_V\}$, it follows that $\mathbf{u} = \mathbf{u}'$ and $\mathbf{w}' = \mathbf{w}$.

Hence

$$T\colon U \oplus W \longrightarrow V, \quad (\mathbf{u}, \mathbf{w}) \longmapsto \mathbf{u} + \mathbf{w}$$

is a bijection.

As it is plainly a linear transformation, it is, in fact, an isomorphism. □

Definition 6.33. The vector space V is the *internal direct sum of the subspaces* U, W if and only if $V = U + W$ and $U \cap W = \{\mathbf{0}_V\}$. We then write $V = U \oplus W$.

The importance of this notion is difficult to overstate. For one thing it is one of the keys to our programme of classifying (up to isomorphism) vector spaces over a given field. We shall see that a non-trivial vector space, V, can be *decomposed* as the internal direct sum of vector subspaces, V_j, each of which is isomorphic with $\mathbb{F}$.

Corollary 6.34. *The vector space V is the internal direct sum of the subspaces U and V if and only if each $\mathbf{v} \in V$ can be written uniquely as*

$$\mathbf{v} = \mathbf{u} + \mathbf{w}$$

with $\mathbf{u} \in U$ and $\mathbf{w} \in W$.

Proof. Exercise. □

Example 6.35. We saw in Example 6.27 on page 67 that $\mathbb{R} = \mathbb{R} \oplus \mathbb{R}$.

As $\mathbb{R}$ is not a subspace of $\mathbb{R}^2$, this expresses $\mathbb{R}^2$ as an (external) direct sum, but not as an internal direct sum.

On the other hand, putting

$$\begin{aligned} U &:= \{(x, 0) \mid x \in \mathbb{R}\} \\ W &:= \{(0, y) \mid y \in \mathbb{R}\} \end{aligned}$$

we have $U \cap W = (0, 0) = \mathbf{0}_{\mathbb{R}^2}$ and $\mathbb{R}^2 = U + W$, showing that $\mathbb{R}^2$ is the internal direct sum of $\{(x, 0) \mid x \in \mathbb{R}\}$ and $\{(0, y) \mid y \in \mathbb{R}\}$.

Notice that $U \oplus W \neq \mathbb{R}^2$, for

$$\begin{aligned} U \oplus W &= \{(\mathbf{u}, \mathbf{w}) \mid \mathbf{u} \in U, \mathbf{w} \in W\} \\ &= \left\{ \Big((x, 0), (0, y) \Big) \mid x, y \in \mathbb{R} \right\} \\ &\subset \mathbb{R}^2 \oplus \mathbb{R}^2 \end{aligned}$$

Observation 6.36. The subspaces U and W in Example 6.35 are not uniquely determined, for we could, instead, have chosen

$$\begin{aligned} U &:= \{(x, x) \mid x \in \mathbb{R}\} \\ W &:= \{(2y, y) \mid y \in \mathbb{R}\} \end{aligned}$$

Example 6.35 illustrates a general phenomenon.

Theorem 6.37. *Let V and W be vector spaces over the field $\mathbb{F}$. Then*

$$\begin{aligned} in_V &\colon V \longrightarrow V \oplus W, \quad \mathbf{v} \longmapsto (\mathbf{v}, \mathbf{0}_W) \\ in_W &\colon V \longrightarrow V \oplus W, \quad \mathbf{w} \longmapsto (\mathbf{0}_V, \mathbf{w}) \end{aligned}$$

are injective linear transformations,

$$\begin{aligned}\mathrm{im}(in_V) &= \{(\mathbf{x}, \mathbf{0}_W) \mid \mathbf{x} \in V\}\\ \mathrm{im}(in_W) &= \{(\mathbf{0}_V, \mathbf{y}) \mid \mathbf{y} \in W\}\end{aligned}$$

and $V \oplus W$ *is the internal direct sum of* $\mathrm{im}(in_V)$ *and* $\mathrm{im}(in_W)$

Proof. Exercise. □

Definition 6.38. Given vector spaces V and W over the field $\mathbb{F}$, the linear transformation

$$\begin{aligned}in_V &: V \longrightarrow V \oplus W, \quad \mathbf{v} \longmapsto (\mathbf{v}, \mathbf{0}_W)\\ in_W &: V \longrightarrow V \oplus W, \quad \mathbf{w} \longmapsto (\mathbf{0}_V, \mathbf{w})\end{aligned}$$

are the *natural inclusions of the direct sum.*

Observation 6.39. The natural inclusions of a direct sum capture a familiar geometric idea. When we draw horizontal and vertical axes — usually call the "x-axis" and the "y-axis" in the Cartesian plane, we are drawing the images of the two natural inclusions

$$\begin{aligned}in_1 &: \mathbb{R} \longrightarrow \mathbb{R} \oplus \mathbb{R} = \mathbb{R}^2, \quad x \longmapsto (x, 0)\\ in_2 &: \mathbb{R} \longrightarrow \mathbb{R} \oplus \mathbb{R} = \mathbb{R}^2, \quad y \longmapsto (0, y)\end{aligned}$$

This illustrates another way in which linear algebra captures and formulates certain geometric concepts, and allows them to be used in a broader context.

We can have direct sums of more than two spaces. We consider the case of finitely many spaces.

Definition 6.40. Given vector spaces, $V_1, \ldots V_n$ over the field $\mathbb{F}$, their *direct sum*, $V = \bigoplus_{j=1}^{n} V_j$, consists of the set

$$V = V_1 \times \cdots V_n = \{(\mathbf{v}_1, \ldots, \mathbf{v}_n) \mid \mathbf{v}_j \in V_j,\ j = 1, \ldots, n\}$$

with vector space operations given by

$$\begin{aligned}\boxplus_V &: V \times V \longrightarrow V, \quad \Big((\mathbf{v}_1, \ldots, \mathbf{v}_n), (\mathbf{w}_1, \ldots, \mathbf{w}_n)\Big) \longmapsto (\mathbf{v}_1 \boxplus_{V_1} \mathbf{w}_1, \ldots, \mathbf{v}_n \boxplus_{V_n} \mathbf{w}_n)\\ \boxtimes_V &: \mathbb{F} \times V \longrightarrow V, \quad \Big(\lambda, (\mathbf{v}_1, \ldots, \mathbf{v}_n)\Big) \longmapsto (\lambda \boxtimes_{V_1} \mathbf{v}_1, \ldots, \lambda \boxtimes_{V_n} \mathbf{v}_n\Big)\end{aligned}$$

Example 6.41. Regarding the field, $\mathbb{F}$, as vector space itself, for any counting number, n,

$$\mathbb{F}^n = \bigoplus_{j=1}^{n} \mathbb{F}$$

Definition 6.42. The vector space, V, is the *internal direct sum of its subspaces* $W_1, \ldots, , W_n$ if and only if each $\mathbf{v} \in V$ can be expressed uniquely as

$$\mathbf{v} = \sum_{j=1}^{n} \mathbf{w}_j$$

with $\mathbf{w}_j \in W_j \ \ (j = 1, \ldots, n)$.

Observation 6.43. We can also define the direct sum of infinitely many vector spaces. However, a number of difficulties arise, whose detailed analysis we leave to more advanced courses.

6.3 Quotient Spaces

Another important construction is that of a *quotient vector space*. While its true significance will not be apparent until you have have studied more mathematics, we introduce its construction here to demonstrate that new constructions are possible even with the limited theory we have already developed, and to illustrate some of the interrelationships between the concepts introduced.

Let U be a vector subspace of the vector space V over the field $\mathbb{F}$.

Define a relation, $\sim_U$, on V by

$$\mathbf{v} \sim_U \mathbf{v}' \quad \text{if and only if} \quad \mathbf{v}' - \mathbf{v} \in U$$

It is easy to see that $\sim_U$ defines an equivalence relation on V. We establish reflexivity, leaving symmetry and transitivity to the reader as an exercise.

Take $\mathbf{v} \in V$.

By definition, $\mathbf{v} \sim_U \mathbf{v}$ if and only if $\mathbf{v} - \mathbf{v} \in U$.

But $\mathbf{v} - \mathbf{v} = \mathbf{0}_V$, and $\mathbf{0}_V \in U$ as $U \leq V$.

Thus, $\mathbf{v} \sim_U \mathbf{v}$.

Let $[\mathbf{v}]$ denote the $\sim_U$-equivalence class containing $\mathbf{v}$, and put

$$V/U := \{[\mathbf{v}] \mid \mathbf{v} \in V\}.$$

We have the natural function

$$\eta \colon V \longrightarrow V/U, \quad \mathbf{v} \longmapsto [\mathbf{v}].$$

Define vector space operations on V/U by

$$\begin{aligned} [\mathbf{v}] + [\mathbf{v}'] &:= [\mathbf{v} + \mathbf{v}'] \\ \lambda.[\mathbf{v}] &:= [\lambda \mathbf{v}] \end{aligned}$$

for all $[\mathbf{v}], [\mathbf{v}'] \in V/U$ and $\lambda \in \mathbb{F}$.

It must first be established that these operations are well defined, for they are defined by choosing representatives of the equivalence classes, operating on these, and then forming new equivalence classes, and not directly from the equivalence classes themselves. We must show that the result does not depend on the particular choices made.

We show this to be the case for the addition of vectors, and leave it to the reader as an exercise to show that the multiplication of a vector by a scalar is also well defined.

We need to show that if $[\mathbf{u}] = [\mathbf{u}']$ and $[\mathbf{v}] = [\mathbf{v}']$, then $[\mathbf{u} + \mathbf{v}] = [\mathbf{u}' + \mathbf{v}']$.

Now, if $[\mathbf{u}] = [\mathbf{u}']$ and $[\mathbf{v}] = [\mathbf{v}']$, then $\mathbf{u} \sim_U \mathbf{u}'$ and $\mathbf{v} \sim_U \mathbf{v}'$, or, equivalently,

$$\mathbf{u} - \mathbf{u}', \mathbf{v} - \mathbf{v}' \in U$$

Then

$$\begin{aligned} (\mathbf{u} + \mathbf{v}) - (\mathbf{u}' + \mathbf{v}') &= (\mathbf{u} - \mathbf{u}') + (\mathbf{v} - \mathbf{v}') \\ &\in U \qquad \text{as } \mathbf{u} - \mathbf{u}', \mathbf{v} - \mathbf{v}' \in U \text{ and } U \leq V \end{aligned}$$

Hence, $\mathbf{u} + \mathbf{v} \sim_U \mathbf{u}' + \mathbf{v}'$, or, equivalently,

$$[\mathbf{u} + \mathbf{v}] = [\mathbf{u}' + \mathbf{v}']$$

It is now easy to verify that these definitions do, indeed, render V/U a vector space over $\mathbb{F}$.

We illustrate this by establishing VS1 (the associativity of the addition of vector), leaving the rest to the reader as an exercise.

Take $[\mathbf{u}], [\mathbf{v}], [\mathbf{w}] \in V/U$. Then

$$\begin{aligned}\big([\mathbf{u}] + [\mathbf{v}]\big) + [\mathbf{w}] &= [\mathbf{u} + \mathbf{v}] + [\mathbf{w}] \\ &= [(\mathbf{u} + \mathbf{v}) + \mathbf{w}] \\ &= [\mathbf{u} + (\mathbf{v} + \mathbf{w})] && \text{by VS1 for } V \\ &= [\mathbf{u}] + [\mathbf{v} + \mathbf{w}] \\ &= [\mathbf{u}] + \big([\mathbf{v}] + [\mathbf{w}]\big)\end{aligned}$$

It is also easy to see that $\eta\colon V \longrightarrow V/U$ is a linear transformation, whose kernel is precisely U.

We illustrate this by establishing that η is homogeneous, leaving the rest to the reader as an exercise.

For $\lambda \in \mathbb{F}$ and $\mathbf{v} \in V$,

$$\begin{aligned}\eta(\lambda\mathbf{v}) &= [\lambda\mathbf{v}] \\ &=: \lambda[\mathbf{v}] \\ &= \lambda\big(\eta(\mathbf{v})\big)\end{aligned}$$

Observation 6.44. In fact, this is the only way of defining a vector space structure on V/U with respect to which η is a linear transformation.

Definition 6.45. If U is a vector subspace of V, then the *quotient space of V modulo U* is V/U with the vector space operations just defined.

Example 6.46. Take $V = \mathbb{R}^2$ with its usual structure as real vector space.

Then $U := \{(x, y) \mid y = 2x\}$ is a vector subspace of V.

Thus $(x, y) \sim_U (x', y')$ if and only if $(x - x', y - y') \in U$, which is the case if and only if $y - y' = 2(xx')$. It follows that

$$[(x, y)] = \{(x + r, y + 2r) \mid r \in \mathbb{R}\}$$

In particular, each $[(x, y)] = [(0, y - 2x)] = [(x - \frac{y}{2}, 0)]$.

In other words, each equivalence class contains a unique representative of the form $(a, 0)$, as well as a unique one of the form $(0, b)$. It follows that

$$T\colon \mathbb{R} \longrightarrow V/U, \quad x \longmapsto [(x, 0)]$$

is an isomorphism of real vector spaces, as is also

$$S\colon \mathbb{R} \longrightarrow V/U, \quad y \longmapsto [(0, y)]$$

Observation 6.47. The vector subspace, U in Example 6.46 can be expressed in the form $\mathbb{F}\mathbf{v}$. Since $(x, y) \in U$ if and only if $y = 2x$, every $(x, y) \in U$ must be of the form $(x, 2x) = x(1, 2)$, with x being any real number whatsoever, whence

$$U = \mathbb{R}(1, 2)$$

and so

$$[(x, y)] = (x, y) + \mathbb{R}(1, 2)$$

This has a convenient geometric interpretation.

U is the line, ℓ, in the Cartesian plane through the origin and the point with co-ordinates $(1,2)$, and the equivalence class $[(x,y)]$ comprises the (unique) line ℓ' in the Cartesian plane parallel to ℓ and passing through the point with co-ordinates (x,y).

Recall from Chapter 1 on page 1 that an equivalence relation on a set is "essentially the same as" partitioning the set in question, with the equivalence classes being the partitioning subsets. In Example 6.46 on the preceding page we have partitioned the Cartesian plane, which we identify with $\mathbb{R}^2$, into the family of all lines parallel to ℓ.

This illustrates that our algebraic considerations encapsulate geometry.

6.4 $\mathrm{Hom}_{\mathbb{F}}(V,W)$ and the Dual of a Vector Space

Another way of constructing a new vector space from two given vector spaces is to consider the set of all linear transformations between then.

Definition 6.48. Let V and W be vector spaces over the field $\mathbb{F}$.

We write $\mathrm{Hom}_{\mathbb{F}}(V,W)$ for the set of all linear transformations from V to W, so that

$$\mathrm{Hom}_{\mathbb{F}}(V,W) := \{T\colon V \to W \mid T \text{ is a linear transformation}\}$$

Define

$$\begin{aligned} \boxplus\colon \mathrm{Hom}_{\mathbb{F}}(V,W) \times \mathrm{Hom}_{\mathbb{F}}(V,W) &\longrightarrow \mathrm{Hom}_{\mathbb{F}}(V,W), \quad (S,T) \longmapsto \big(S \boxplus T\colon V \to W\big) \\ \boxdot\colon \mathbb{F} \times \mathrm{Hom}_{\mathbb{F}}(V,W) &\longrightarrow \mathrm{Hom}_{\mathbb{F}}(V,W), \quad (\lambda,T) \longmapsto \big(\lambda \boxdot T\colon V \to W\big) \end{aligned}$$

by

$$\begin{aligned} &S \boxplus T\colon \mathbf{v} \longmapsto S(\mathbf{v}) + T(\mathbf{v}) \\ &\lambda \boxdot T\colon \mathbf{v} \longmapsto \lambda\big(T(\mathbf{v})\big), \end{aligned}$$

Theorem 6.49. *If V and W are vector spaces over the field $\mathbb{F}$, then $\mathrm{Hom}_{\mathbb{F}}(V,W)$, as defined in Definition 6.48, is a vector space over $\mathbb{F}$, with zero vector*

$$\mathbf{0}_{\mathrm{Hom}_{\mathbb{F}}(V,W)}\colon V \longrightarrow W, \quad \mathbf{v} \longmapsto \mathbf{0}_W$$

One way to prove the theorem is to verify each of the eight axioms VS1 – VS8. Any reader, who does not trust theoretical methods, is encouraged to do so, as the verifications are routine.

We provide, instead, and alternative, utilising the theory we have already developed.

Proof. Recall from Exercise 3.11 on page 43 (which was a generalisation of Example 3.18 on page 36) that $\mathcal{F}(V,W)$, the set of all functions $f\colon V \longrightarrow W$, is a vector space over $\mathbb{F}$ with respect to the operations we have just defined.

Since $\mathrm{Hom}_{\mathbb{F}}(V,W)$ is a subset of $\mathcal{F}(V,W)$, it follows by Theorem 6.4 on page 63 that $\mathrm{Hom}_{\mathbb{F}}(V,W)$ is a vector subspace of $\mathcal{F}(V,W)$ — and therefore a vector space in its own right — if and only if $\mathrm{Hom}_{\mathbb{F}}(V,W)$ is closed under the operations we have just defined.

In other words, all we need to do is to show that for all $S,T \in \mathrm{Hom}_{\mathbb{F}}(V,W)$ and $\lambda \in \mathbb{F}$, both $S \boxplus T$ and $\lambda \boxdot T$ are again in $\mathrm{Hom}_{\mathbb{F}}(V,W)$, which is to say, that they are, in fact, linear transformations.

Take $\mathbf{u},\mathbf{v} \in V$ and $\alpha,\beta \in \mathbb{F}$. Then

$$(S \boxplus T)(\alpha\mathbf{u} + \beta\mathbf{v}) := S(\alpha\mathbf{u} + \beta\mathbf{v}) + T(\alpha\mathbf{u} + \beta\mathbf{v})$$

$$
\begin{aligned}
&= (\alpha S(\mathbf{u}) + \beta S(\mathbf{v})) + (\alpha T(\mathbf{u}) + \beta T(\mathbf{v})) && \text{as } S, T \text{ are linear}\\
&= \alpha(S(\mathbf{u}) + T(\mathbf{u})) + \beta(S(\mathbf{v}) + T(\mathbf{v})) && \text{as } W \text{ is an } \mathbb{F}\text{-vector space}\\
&=: \alpha\big((S \boxplus T)\big)(\mathbf{u}) + \beta\big((S \boxplus T)\big)(\mathbf{v})\\
(\lambda \boxdot T)(\alpha\mathbf{u} + \beta\mathbf{v}) &:= \lambda T(\alpha\mathbf{u} + \beta\mathbf{v})\\
&= \lambda(\alpha T(\mathbf{u}) + \beta T(\mathbf{v})) && \text{as } T \text{ is linear}\\
&= (\alpha(\lambda T(\mathbf{u})) + \beta(\lambda T(\mathbf{v})) && \text{as } W \text{ is an } \mathbb{F}\text{-vector space}\\
&= \alpha\big((\lambda \boxdot T)\big)(\mathbf{u}) + \beta\big((\lambda \boxdot T)\big)(\mathbf{v})
\end{aligned}
$$

□

$\mathcal{L}(V, W)$ and $\mathrm{Hom}(V, W)$ are two common notations for $\mathrm{Hom}_{\mathbb{F}}(V, W)$ when there is no ambiguity about the field in question.

In the special case that $W = \mathbb{F}$, $\mathrm{Hom}_{\mathbb{F}}(V, W)$ is called the *dual* space of V.

Definition 6.50. The *dual space* of the vector V over the field $\mathbb{F}$ is $\mathrm{Hom}_{\mathbb{F}}(V, \mathbb{F})$.

The elements of $\mathrm{Hom}_{\mathbb{F}}(V, \mathbb{F})$ are called *linear forms (on V)*.

We sometimes write V^* for $\mathrm{Hom}_{\mathbb{F}}(V, \mathbb{F})$.

Composition with the linear transformation $T\colon V \longrightarrow W$ defines functions

$$
\begin{aligned}
T^*\colon \mathrm{Hom}(W, X) &\longrightarrow \mathrm{Hom}(V, X), \quad S \longmapsto S \circ T\\
T_*\colon \mathrm{Hom}(U, V) &\longrightarrow \mathrm{Hom}(U, W), \quad R \longmapsto T \circ R
\end{aligned}
$$

We investigate properties of these two functions.

Theorem 6.51. *Let U, V, W, X be vector spaces over the field $\mathbb{F}$. Let $R, R'\colon U \longrightarrow V$, $T\colon V \longrightarrow W$ and $S, S'\colon W \longrightarrow X$ be linear transformations. Take $\alpha \in \mathbb{F}$.*

(i) $(S \boxplus S') \circ T = (S \circ T) \boxplus (S' \circ T)\colon V \longrightarrow X$

(ii) $T \circ (R \boxplus R') = (T \circ S) \boxplus (T \circ S')\colon U \longrightarrow W$

(iii) $(\alpha \boxdot S) \circ T = \alpha \boxdot (S \circ T) = S \circ (\alpha \boxdot T)\colon V \longrightarrow X$

Proof. Since, in each case, the linear transformations in question have the same domain and the same co-domain as each other, it is sufficient to show that they assign the same vector in the co-domain to a given vector in the domain.

For this, take $\mathbf{u} \in U$ and $\mathbf{v} \in V$.

(i)

$$
\begin{aligned}
\big((S \boxplus S') \circ T\big)(\mathbf{v}) &:= (S \boxplus S')(T(\mathbf{v})) && \text{by the definition of composition}\\
&:= S(T(\mathbf{v})) + S'(T(\mathbf{v})) && \text{by the definition of } \boxplus\\
&=: (S \circ T)(\mathbf{v}) + (S' \circ T)(\mathbf{v}) && \text{by the definition of composition}\\
&=: \big((S \circ T) \boxplus (S' \circ T)\big)(\mathbf{v}) && \text{by the definition of } \boxplus
\end{aligned}
$$

(ii)

$$
\begin{aligned}
\big(T \circ (R \boxplus R')\big)(\mathbf{u}) &:= T\big((R \boxplus R')(\mathbf{u})\big) && \text{by the definition of composition}\\
&:= T\big((R(\mathbf{u}) + R'(\mathbf{u})\big) && \text{by the definition of } \boxplus
\end{aligned}
$$

$$
\begin{aligned}
&= T\big((R(\mathbf{u})\big) + T\big(R'(\mathbf{u})\big) && \text{as } T \text{ is a linear transformation}\\
&=: (T \circ R)(\mathbf{u}) + (T \circ R')(\mathbf{u}) && \text{by the definition of composition}\\
&=: \big((T \circ R) \boxplus (T \circ R')\big)(\mathbf{u}) && \text{by the definition of } \boxplus
\end{aligned}
$$

(iii)

$$
\begin{aligned}
\big((\alpha \boxdot S) \circ T\big)(\mathbf{v}) &:= (\alpha \boxdot S)(T\mathbf{v}))) && \text{by the definition of composition}\\
&:= \alpha S(T(\mathbf{v})) && \text{by the definition of } \boxdot\\
&=: \alpha(S \circ T)(\mathbf{v}) && \text{by the definition of composition}\\
&=: \big(\alpha \boxdot (S \circ T)\big)(\mathbf{v}) && \text{by the definition of } \boxdot\\
\big(S \circ (\alpha \boxdot T)\big)(\mathbf{v}) &:= S\big((\alpha \boxdot T)(\mathbf{v})\big) && \text{by the definition of composition}\\
&:= S(\alpha T(\mathbf{v})) && \text{by the definition of } \boxdot\\
&= \alpha S(T(\mathbf{v})) && \text{as } S \text{ is a linear transformation}\\
&=: \alpha(S \circ T)(\mathbf{v}) && \text{by the definition of composition}\\
&=: \big(\alpha \boxdot (S \circ T)\big)(\mathbf{v}) && \text{by the definition of } \boxdot
\end{aligned}
$$

□

Corollary 6.52. *The linear transformation $T\colon V \longrightarrow W$ induces linear transformations*

$$
\begin{aligned}
T_*\colon \operatorname{Hom}_{\mathbb{F}}(V,W) &\longrightarrow \operatorname{Hom}_{\mathbb{F}}(V,X), \qquad R \longmapsto T \circ R\\
T^*\colon \operatorname{Hom}_{\mathbb{F}}(V,W) &\longrightarrow \operatorname{Hom}_{\mathbb{F}}(U,W), \qquad S \longmapsto S \circ T
\end{aligned}
$$

Proof. Theorem 6.51 on the facing page (i) shows that T^* is additive.

Theorem 6.51 on the preceding page (ii) shows that T_* is additive.

Theorem 6.51 on the facing page (iii) shows that both T^* and T_* are homogeneous. □

Observation 6.53. Theorem 6.51 on the preceding page and Corollary 6.52, when applied to finitely generated vector spaces, provide the basis for computation with matrices, explaining the conditions imposed on matrices for them to be added and multiplied, as well proving, without further argument, the properties of these algebraic operations on matrices.

6.5 Exercises

Exercise 6.1. Let $\mathcal{C}(\mathbb{R}) := \{f : \mathbb{R} \to \mathbb{R}\}$ be the set of all real valued functions defined on $\mathbb{R}$.

$\mathcal{C}(\mathbb{R})$ is a real vector space with respect to point-wise operations (cf. Example 3.18 on page 36).

Decide which of the following subsets of $\mathcal{C}(\mathbb{R})$, are, in fact, vector subspaces.

(a) $\mathcal{C}^0(\mathbb{R}) := \{f : \mathbb{R} \to \mathbb{R} \mid f \text{ is continuous}\}$

(b) $\mathcal{C}^r(\mathbb{R}) := \{f : \mathbb{R} \to \mathbb{R} \mid \dfrac{d^r f}{dx^r} \text{ is continuous}\}$ $\quad (r \in \mathbb{N} \setminus \{0\})$.

(c) $\big(\mathcal{F}(\mathbb{R})\big)_{(x_0)} := \{f : \mathbb{R} \to \mathbb{R} \mid f(x_0) = 0\}$, where x_0 is a fixed real number.

(d) $\big(\mathcal{F}(\mathbb{R})\big)_{0,1} := \{f : \mathbb{R} \to \mathbb{R} \mid f(0)f(1) = 0\}$.

Exercise 6.2. Show $\mathbb{R}^3$ is generated as real vector space by

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

Exercise 6.3. Determine the vector subspace of $\mathbb{R}^3$ generated by

(i) $\{(0,1,2),(1,2,3)\}$

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

(iii) $\{(0,1,2),(1,3,5)\}$

Exercise 6.4. Given vector spaces V and W over the field $\mathbb{F}$, let

$$\mathcal{F}(V,W) := \{f\colon V \longrightarrow W\}$$

be the set of all functions from V to W.

Prove that $\mathcal{F}(V,W)$ forms a vector space over $\mathbb{F}$ with respect to point-wise operations and that

$$\mathrm{Hom}_{\mathbb{F}}(V,W) := \{f\colon V \to W \mid f \text{ is an } \mathbb{F}\text{-linear transformation}\}$$

is a vector subspace of $\mathcal{F}(V,W)$.

Exercise 6.5. Find all vector subspaces of $\mathbb{C}^2$, when

(a) $\mathbb{C}^2$ is taken as a vector space over $\mathbb{C}$.

(b) $\mathbb{C}^2$ is taken as a vector space over $\mathbb{R}$ in the usual manner.

Exercise 6.6. Prove Theorem 6.26 on page 67: If V and W are vector spaces over $\mathbb{F}$, then $V \oplus W$ is a vector space over $\mathbb{F}$ with respect to the operations defined by

$$(\mathbf{v}_1,\mathbf{w}_1)+(\mathbf{v}_2,\mathbf{w}_2) := (\mathbf{v}_1+\mathbf{v}_2,\mathbf{w}_1+\mathbf{w}_2)$$
$$\lambda(\mathbf{v},\mathbf{w}) := (\lambda\mathbf{v},\lambda\mathbf{w})$$

for all $\lambda \in \mathbb{F}$, $\mathbf{v},\mathbf{v}_1,\mathbf{v}_2 \in \mathbb{F}$ and $\mathbf{w},\mathbf{w}_1,\mathbf{w}_2 \in W$.

Exercise 6.7. Prove that U and W are vector subspaces of V then

$$U + W = \langle U \cup W \rangle$$

Exercise 6.8. Let U, W be vector subspaces of V, with $U \cap W = \{\mathbf{0}_V\}$ and $V = U + W$.

Prove that

$$T\colon U \oplus W \longrightarrow V, \quad (\mathbf{u},\mathbf{w}) \longmapsto \mathbf{u}+\mathbf{w}$$

is an isomorphism.

Exercise 6.9. Prove the $\mathbb{R}^2$ is the internal direct sum of the subspaces

$$U := \{(x,x) \mid x \in \mathbb{R}\}$$
$$W := \{(2y,y) \mid y \in \mathbb{R}\}$$

Exercise 6.10. Prove that the vector space V is the internal direct sum of the subspaces U and V if and only if each $\mathbf{v} \in V$ can be written uniquely as

$$\mathbf{v} = \mathbf{u} + \mathbf{w}$$

with $\mathbf{u} \in U$ and $\mathbf{w} \in W$.

Exercise 6.11. Let U be a vector subspace of the vector space V over the field $\mathbb{F}$.

Define a relation $\sim_U$ on V by

$$\mathbf{v} \sim_U \mathbf{v}' \quad \text{if and only if} \quad \mathbf{v}' - \mathbf{v} \in U$$

a. Prove that $\sim_U$ defines an equivalence relation on V.

Let $[\mathbf{v}]$ denote the $\sim_U$-equivalence class containing $\mathbf{v}$, and put

$$V/U := \{[\mathbf{v}] \mid \mathbf{v} \in V\}.$$

We have the natural function

$$\eta : V \longrightarrow V/U, \quad \mathbf{v} \longmapsto [\mathbf{v}].$$

Define

$$\boxplus : V/U \times V/U \longrightarrow V/U, \quad ([\mathbf{v}], [\mathbf{v}']) \longmapsto [\mathbf{v} + \mathbf{v}']$$

$$\boxtimes : \mathbb{F} \times V/U \longrightarrow V/U, \quad (\lambda, [\mathbf{v}]) \longmapsto [\lambda \mathbf{v}].$$

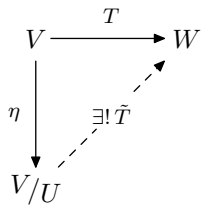
b. Prove the following statements.

(i) These definitions render V/U a vector space over $\mathbb{F}$.

(ii) $\eta\colon V \longrightarrow V/U$ is a linear transformation.

(iii) $\ker(\eta) = U$.

c. Prove that if W is any vector space over $\mathbb{F}$ and $T\colon V \longrightarrow W$ is any linear transformation withker$(T) \subseteq U$, then there is a uniquely determined linear transformation $\tilde{T} : V/U \longrightarrow W$ such that $T = \tilde{T} \circ \eta$.In diagrammatic form



[This is an example of a *universal property*. You will meet universal properties if you pursue further studies in mathematics, especially in *category theory*.]

Exercise 6.12. Prove Theorem 6.49 on page 73 by verifying directly that the axioms hold.

Exercise 6.13. Show that

$$U = \{(x, y, z) \in \mathbb{R}^3 \mid x + y + z = 0\}$$

is a vector subspace of $\mathbb{R}^3$, and that

$$\mathbb{R}^3/U \cong \mathbb{R}$$

Exercise 6.14. Take vector spaces V, W and a linear transformation $T\colon V \longrightarrow W$.

Prove that

$$V/\ker(T) \cong \mathrm{im}(T)$$

[This central result is *the Noether Isomorphism Theorem*, sometimes called *Noether's First Isomorphism Theorem*. The Noether in question being Emmy Noether (1882-1935), often referred to as *the father of modern algebra*.]

The art of doing mathematics consists in finding that special case which contains all the germs of generality.

David Hilbert

Chapter 7

Linear Dependence and Bases

We saw in Chapter 6 that for each subset, S, of the vector space, V, there is a unique "smallest" vector subspace, $\langle S \rangle$ of V containing all the elements of S. Theorem 6.17 on page 65 showed that $\langle S \rangle$ is the intersection of all those vector subspaces of V, that contain all the element of S.

While this establishes both the existence and uniqueness of the vector subspace $\langle S \rangle$, it provides no indication of how to find $\langle S \rangle$ directly from S.

We show how to do so in this chapter. The concepts and techniques we introduce have broader application. For example, they provide the key to classifying vector spaces up to isomorphism.

All vector spaces are understood to be over a fixed field, $\mathbb{F}$. We only mention the specific field in concrete examples.

Definition 7.1. The vector, $\mathbf{x}$, is a *linear combination of the vectors in* S if and only if there are $n \in \mathbb{N}$, $\mathbf{v}_1, \ldots, \mathbf{v}_n \in S$ and $\lambda_1, \ldots, \lambda_n \in \mathbb{F}$ with

$$\mathbf{x} = \lambda_1 \mathbf{v}_1 + \cdots + \lambda_n \mathbf{v}_n = \sum_{i=1}^{n} \lambda_i \mathbf{v}_i.$$

The vectors in S are *linearly independent (over* $\mathbb{F}$*)* if and only if for all $n \in \mathbb{N}$ and all $\mathbf{v}_1, \ldots, \mathbf{v}_n \in S$, the equation

$$\sum_{i=1}^{n} \lambda_i \mathbf{v}_i = \lambda_1 \mathbf{v}_1 + \cdots + \lambda_n \mathbf{v}_n = \mathbf{0}_V \quad (\lambda_1, \ldots, \lambda_n \in \mathbb{F})$$

has only the trivial solution $\lambda_1 = \cdots \lambda_n = 0$.

Otherwise the vectors in S are *linearly dependent*.

Example 7.2. Let $V = \mathbb{R}^2$, with its usual vector space structure over $\mathbb{R}$.

Then $(3, 2)$ is a linear combination of $(1, 1)$ and $(5, 4)$, because

$$(3, 2) = -2(1, 1) + (5, 4)$$

The vectors $(1, 1), (5, 4)$ and $(3, 2)$ are linearly dependent, because

$$2.(1, 1) + (-1).(5, 4) + 1.(3, 2) = (0, 0)$$

The vectors $(1, 1)$ and $(3, 2)$ are linearly independent, because

$$\lambda(1, 1) + \mu(3, 2) = (0, 0)$$

if and only if

$$\begin{array}{rcrcl} \lambda & + & 3\mu & = & 0 \\ \lambda & + & 2\mu & = & 0 \end{array}$$

which, clearly, is the case if and only if $\lambda = \mu = 0$.

Lemma 7.3. *The vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are linearly dependent if and only if at least one of them can be expressed as a linear combination of the others.*

Proof. Since $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are linearly dependent, there are $\lambda_1, \ldots, \lambda_n \in \mathbb{F}$, not all 0, with

$$\sum_{j=1}^{n} \lambda_j \mathbf{v}_j = \mathbf{0}_V.$$

Suppose that $\lambda_i \neq 0$. Then

$$\lambda_i \mathbf{v}_i = -\sum_{j \neq i} \lambda_j \mathbf{v}_j$$

whence

$$\mathbf{v}_i = \sum_{j \neq i} \mu_j \mathbf{v}_j,$$

with $\mu_j := \dfrac{-\lambda_j}{\lambda_i} \in \mathbb{F},\ (j = 1, \ldots, n,\ j \neq i)$. □

Theorem 7.4. *If each $\mathbf{w}_i \ \ (i = 1, \ldots, m)$ is a linear combination of $\mathbf{v}_1, \ldots, \mathbf{v}_n$, then any linear combination of $\mathbf{w}_1, \ldots, \mathbf{w}_m$ is a linear combination of $\mathbf{v}_1, \ldots, \mathbf{v}_n$.*

Proof. Suppose that for each $i = 1, \ldots, m$

$$\mathbf{w}_i = \sum_{j=1}^{n} a_{ij} \mathbf{v}_j \qquad \big(= a_{i1}\mathbf{v}_1 + \cdots a_{in}\mathbf{v}_n \big)$$

If $\mathbf{u}$ is a linear combination of $\mathbf{w}_1, \ldots, \mathbf{w}_m$, there are $\lambda_1, \ldots, \lambda_m \in \mathbb{F}$ with

$$\begin{aligned} \mathbf{u} &= \sum_{i=1}^{m} \lambda_i \mathbf{w}_i \\ &= \sum_{i=1}^{m} \lambda_i \Big(\sum_{j=1}^{n} a_{ij} \mathbf{v}_j \Big) \\ &= \sum_{i=1}^{m} \Big(\sum_{j=1}^{n} \lambda_i a_{ij} \mathbf{v}_j \Big) \\ &= \sum_{j=1}^{n} \Big(\sum_{i=1}^{m} \lambda_i a_{ij} \mathbf{v}_j \Big) && \text{as the summations are independent of each other} \\ &= \sum_{j=1}^{n} \Big(\sum_{i=1}^{m} \lambda_i a_{ij} \Big) \mathbf{v}_j \\ &= \sum_{j=1}^{n} \mu_j \mathbf{v}_j \end{aligned}$$

where $\mu_j = \displaystyle\sum_{i=1}^{m} \lambda_i a_{ij}$ □

Recall that $\langle S \rangle$, the vector subspace of V generated by the subset, S, of V, is the smallest vector subspace of V containing S. It is the intersection of all those vector subspaces of V which have S as a subset.

We can now provide an alternative and intrinsic description of it.

Corollary 7.5. *Given $S \subseteq V$, $\langle S \rangle$ comprises all linear combinations of elements of S.*

Proof. Let $\mathcal{LC}(S)$ be the set of all linear combinations of elements of S.

$$\mathcal{LC}(S) := \{ \sum_{j=1}^{n} \lambda_j \mathbf{v}_j \mid \lambda_j \in \mathbb{F}, \mathbf{v}_j \in S \text{ for all } j \leq n, n \in \mathbb{N} \}$$

By Theorem 7.4 on the preceding page, $\mathcal{LC}(S)$ is closed under the addition of vectors and the multiplication of vectors by scalalrs.

Hence, by Theorem 6.4 on page 63 $\mathcal{LC}(S)$ is a vector subspace of V.

By definition, $S \subseteq \mathcal{LC}(S)$.

Hence, $\langle S \rangle \leq \mathcal{LC}(S)$.

For the reverse inclusion, note that by Theorem 6.4 on page 63, every vector space is closed under forming linear combinations of its elements.

Hence, if $U \leq V$ and $S \subseteq U$, then $\mathcal{LC}(S) \leq U$.

Taking $U = \langle S \rangle$ completes the proof. □

We have shown that $\langle S \rangle$ is obtained by taking all linear combinations of elements of S and we have the following reformulation of the condition for the subset U of the vector space V to be a vector subspace.

$U \subseteq V$ **is a vector subspace of** V **if and only if** U **is closed under linear combinations.**

Lemma 7.6. *Any non-zero vector is linearly independent.*

Proof. If $\mathbf{v} \neq \mathbf{0}_V$, then, by (e) in Theorem 3.23 on page 38, $\lambda \mathbf{v} = \mathbf{0}_V$ if and only if $\lambda = 0$. □

Theorem 7.7. *Let $\mathbf{v}_1, \ldots, \mathbf{v}_n$ be linearly independent.*

If $\mathbf{v}_1, \ldots, \mathbf{v}_n, \mathbf{w}$ are linearly dependent, then $\mathbf{w}$ is a linear combination of $\mathbf{v}_1, \ldots, \mathbf{v}_n$.

Proof. Since $\mathbf{v}_1, \ldots, \mathbf{v}_n, \mathbf{w}$ are linearly dependent, there are $\lambda_1, \ldots, \lambda_{n+1} \in \mathbb{F}$ such that

$$\lambda_1 \mathbf{v}_1 + \cdots + \lambda_n \mathbf{v}_n + \lambda_{n+1} \mathbf{w} = \mathbf{0}_V,$$

with not all $\lambda_j = 0$.

If $\lambda_{n+1} = 0$, then our equation reduces to $\lambda_1 \mathbf{v}_1 + \cdots + \lambda_n \mathbf{v}_n = \mathbf{0}_V$.

Since $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are linearly independent, this has only the trivial solution $\lambda_1 = \cdots = \lambda_n = 0$.

Thus, $\lambda_{n+1} \neq 0$, whence

$$\mathbf{w} = \sum_{j=1}^{n} \mu_j \mathbf{v}_j$$

with $\mu_j = -\dfrac{\lambda_j}{\lambda_{n+1}}$. □

Example 7.8. We saw in Example 7.2 on page 79 that $(1,1)$ and $(3,2)$ are linearly independent, but $(1,1),(3,2)$ and $(5,4)$ are linearly dependent in $\mathbb{R}^2$. Plainly

$$(5,4) = 2.(1,1) + 1.(3,2)$$

Theorem 7.9. *The vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are linearly independent if and only if any vector which can be written as a linear combination of them can be written in this form in precisely one way.*

Proof. As

$$\mathbf{x} = \lambda_1\mathbf{v}_1 + \cdots + \lambda_n\mathbf{v}_n = \mu_1\mathbf{v}_1 + \cdots + \mu_n\mathbf{v}_n$$

if and only if

$$(\lambda_1 - \mu_1)\mathbf{v}_1 + \cdots + (\lambda_n - \mu_n)\mathbf{v}_n = \mathbf{0}_V,$$

the expression of each vector $\mathbf{x}$ as a linear combination of $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is unique if and only if

$$\gamma_1\mathbf{v}_1 + \cdots + \gamma_n\mathbf{v}_n = \mathbf{0}_V$$

has only the trivial solution $\gamma_1 = \cdots = \gamma_n = 0$. □

Linear independence and the property of generating a vector space are closely linked to significant properties of linear transformations, as the next theorem shows.

Theorem 7.10. *Let $T\colon V \longrightarrow W$ be a linear transformation. Then*

(i) T is injective if and only if the vectors $T(\mathbf{v}_1), \ldots, T(\mathbf{v}_n)$ are linearly independent in W whenever $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are linearly independent in V.

(ii) T is surjective if and only if $T(S)$ generates W, whenever S generates V.

Proof. (i) Suppose that T is injective and that $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are linearly independent in V. Then

$$\begin{aligned}
\sum_{j=1}^{n} \lambda_j T(\mathbf{v}_j) = \mathbf{0}_W \quad &\text{if and only if} \quad T(\sum_{j=1}^{n} \lambda_j \mathbf{v}_j) = \mathbf{0}_W, && \text{as } T \text{ is a linear transformation.} \\
&\text{if and only if} \quad \sum_{j=1}^{n} \lambda_j \mathbf{v}_j \mathbf{v}_n \in \ker(T) \\
&\text{if and only if} \quad \sum_{j=1}^{n} \lambda_j \mathbf{v}_j = \mathbf{0}_V, && \text{as } T \text{ is injective} \\
&\text{if and only if} \quad \text{each } \lambda_j = 0, && \text{as } \mathbf{v}_1, \ldots \mathbf{v}_n \text{ are linearly independent,}
\end{aligned}$$

showing that $T(\mathbf{v}_1), \ldots T(\mathbf{v}_n)$ are linearly independent in W.

Suppose, now, that $T(\mathbf{v}_1), \ldots, T(\mathbf{v}_n)$ are linearly independent in W whenever $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are linearly independent in V.

Take $\mathbf{v} \in V$, $\mathbf{v} \neq \mathbf{0}_V \in V$.

Then $\mathbf{v}$ is linearly independent in V.

By hypothesis, $T(\mathbf{v})$ is then linearly independent in W.

Thus, $T(\mathbf{v}) \neq \mathbf{0}_W$, whence $\mathbf{v} \notin \ker(T)$.

Hence $\ker(T) = \{\mathbf{0}_V\}$.

By Lemma 5.14 on page 56, T is injective.

(ii) Suppose that T is surjective and that S generates V.

Take $\mathbf{w} \in W$.

Since T is surjective, there is a $\mathbf{v} \in V$ with $T(\mathbf{v}) = \mathbf{w}$.

Since S generates V, there are $\mathbf{v}_1, \ldots, \mathbf{v}_n \in S$ and $\lambda_1, \ldots, \lambda_n \in \mathbb{F}$ with

$$\mathbf{v} = \lambda_1 \mathbf{v}_1 + \cdots + \lambda_n \mathbf{v}_n.$$

Since T is a linear transformation,

$$\mathbf{w} = T(\mathbf{v}) = T(\lambda_1 \mathbf{v}_1 + \cdots + \lambda_n \mathbf{v}_n) = \lambda_1 T(\mathbf{v}_1) + \cdots + \lambda_n T(\mathbf{v}_n),$$

showing that $T(S)$ generate W.

Conversely, suppose that $T(S)$ generate W whenever S generates V.

Since clearly V generates V, $T(V)$ must generate W, that is, $\langle\, T(V) \,\rangle = W$.

Since $T(V) = \operatorname{im}(T)$ is a vector subspace of W, it follows by Corollary 6.18 on page 66 that $\langle\, \operatorname{im}(T) \,\rangle = \operatorname{im}(T)$.

Hence $W = T(V)$, showing that T is surjective. □

Sets of vectors that are both linearly independent and generate a given vector space lie at the heart of working with vector spaces, since their behaviour completely determines the behaviour of the entire vector space. Such a set of vectors comprises a *basis* for the vector space in question.

Definition 7.11. Let V be a vector space over the field $\mathbb{F}$. The vectors $\{\mathbf{e}_\lambda \mid \lambda \in \Lambda\}$ form a *basis for* V if and only if they are linearly independent and generate V.

Example 7.12. By the definition of the standard real vector space structure on $\mathbb{R}^2$, $\{(1,0),(0,1)\}$ is a basis for $\mathbb{R}^2$. This is the *standard basis* for $\mathbb{R}^2$.

It is left as an exercise for the reader to verify that $\{(0,-1),(\frac{1}{\sqrt{2}},-\frac{1}{\sqrt{2}})\}$ is also a basis for $\mathbb{R}^2$.

Theorem 7.13. *The subset B of the vector space V is a basis for V if and only if every vector in V can be expressed uniquely as a linear combination of the elements of B.*

Proof. By definition, B is a basis for V if and only if B generates V and the laments of B are linearly independent.

By Corollary 7.5 on page 81 B generates V if and only if every vector in V can be written as a linear combination of the elements of B.

By Theorem 7.9 on the preceding page, the elements of B are linearly independent if and only if no vector can be written as a linear combination of the elements of B in more than one way. □

Theorem 7.13 is useful for determining whether a given set of vectors forms a basis.

Example 7.14. The set of all solutions of the real differential equation

$$\frac{d^2y}{dx^2} = -y$$

forms a real vector space, $V = \{f \colon \mathbb{R} \to \mathbb{R} \mid \frac{d^2f}{dx^2} + f = 0\}$.

From the theory of linear differential equations with constant coefficients (cf. MATH102), each $f \in V$ can be expressed uniquely as

$$f\colon \mathbb{R} \longrightarrow \mathbb{R}, \quad x \longmapsto A\cos x + B\sin x$$

Thus, the functions

$$\cos\colon \mathbb{R} \longrightarrow \mathbb{R}, \quad x \longmapsto \cos x$$
$$\sin\colon \mathbb{R} \longrightarrow \mathbb{R}, \quad x \longmapsto \sin x$$

form a basis for V.

It is left as an exercise for the reader to verify that the functions

$$\mathbf{e}_1\colon \mathbb{R} \longrightarrow \mathbb{R}, \quad x \longmapsto \cos(x + \tfrac{\pi}{4})$$
$$\mathbf{e}_2\colon \mathbb{R} \longrightarrow \mathbb{R}, \quad x \longmapsto \cos(x + \tfrac{\pi}{2})$$

also form a basis for V.

The significance of bases is indicated by the facts, to be proved in Chapter 8 for the case of finitely generated vector spaces, that any two bases for a given vector space must have the same number of elements and that two vector spaces over a given field are isomorphic if and only if a basis for one can be found with the same number of elements as a basis for the other.

It is therefore crucial to know when a vector space admits a basis. The answer is provided by the next theorem.

Theorem 7.15. *Every finitely generated vector space admits a basis.*

Proof. Let $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be a finite generating set for V.

As $\lambda\mathbf{0}_V = \mathbf{0}_V$ for all $\lambda \in \mathbb{F}$, we may assume, without loss of generality, that $\mathbf{v}_j \neq \mathbf{0}_V \ \ (j = 1, \ldots n)$.

We construct inductively a basis from S by omitting successively those elements of S which are linearly dependent on their predecessors in the ordering induced by their subscripts.

Since $\mathbf{v}_1 \neq \mathbf{0}_V$, it follows from Lemma 7.6 on page 81 that $\mathbf{v}_1$ is linearly independent.

Put $\mathbf{e}_1 := \mathbf{v}_1$.

Clearly, $\langle \mathbf{e}_1 \rangle = \langle \mathbf{v}_1 \rangle$.

Now suppose that for $j \geq 1$ we have chosen $\mathbf{e}_1, \ldots \mathbf{e}_j$ from S in such a manner that

(i) $\mathbf{e}_1, \ldots \mathbf{e}_j$ are linearly independent and

(ii) $\langle \mathbf{e}_1, \ldots \mathbf{e}_j \rangle = \langle \mathbf{v}_1, \ldots \mathbf{v}_{n_j} \rangle$.

If $\langle \mathbf{e}_1, \ldots \mathbf{e}_j \rangle = V$, we are finished.

Otherwise, let n_{j+1} be the least integer such that $\mathbf{v}_{n_{j+1}} \notin \langle \mathbf{e}_1, \ldots \mathbf{e}_j \rangle$, or equivalently, such that $\mathbf{e}_1, \ldots \mathbf{e}_j, \mathbf{v}_{n_{j+1}}$ are linearly independent.

Put $\mathbf{e}_{j+1} := \mathbf{v}_{n_{j+1}}$.

Then $\mathbf{e}_1, \ldots \mathbf{e}_{j+1}$ are obviously linearly independent and $\langle \mathbf{e}_1, \ldots \mathbf{e}_{j+1} \rangle = \langle \mathbf{v}_1, \ldots \mathbf{v}_{n_{j+1}} \rangle$.

Since S is finite, this procedure must terminate after at most n steps. □

Example 7.16. The vectors $(1,0), (1,1), (0,1)$ generate $\mathbb{R}^2$ as real vector space.

Since $\lambda(1,0) + \mu(1,1) = (\lambda + \mu, \mu) = (0,0)$ if and only if $\mu, \lambda = 0$, $(1,0)$ and $(1,1)$ are linearly independent.

As $(0,1) = -(1,0) + (1,1)$, $(0,1)$ is a linear combination of $(1,0)$ and $(1,1)$.

Hence our procedure produces the basis $\{(1,0),(1,1)\}$ for $\mathbb{R}^2$ as real vector space.

It is left to the reader to verify that $\{(1,0),(1,1)\}$ is, indeed, a basis for $\mathbb{R}^2$.

Observation 7.17. The statement of Theorem 7.15 on the facing page is still true without the restriction to finitely generated vector spaces, but, of course, the above proof would not suffice then. The more general statement requires the *Axiom of Choice* or some equivalent of it. This would take us into the realm of formal set theory, which is not within the scope of this course.

Set theory lies at the basis of most of mathematics and it was through axiomatic set theory that the theory of recursive functions, the theory of computability and the theory of Turing machines arose. Thus, in addition to its centrality in the development of modern mathematics, set theory is the historical, conceptual and theoretical parent of modern computing and modern computers.

7.1 Exercises

Exercise 7.1. Let $T : V \to W$ be a linear transformation of $\mathbb{F}$ vector spaces.

Show that if $\mathbf{v}_1, \ldots, \mathbf{v}_k$ are linearly dependent, then so are $T(\mathbf{v}_1), \ldots, T(\mathbf{v}_k)$.

Find an example with $\mathbf{v}_1, \ldots, \mathbf{v}_k$ linearly independent and $T(\mathbf{v}_1), \ldots, T(\mathbf{v}_k)$ linearly dependent.

Exercise 7.2. Find a basis for the vector subspace of $\mathbb{R}^4$ generated by

$$\{(1,1,0,0),(0,0,0,0),(1,0,0,1),(0,1,0,1),(1,-1,0,3)\}$$

Exercise 7.3. Let $\mathcal{B} = \{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ be a basis for the vector space V over the field $\mathbb{F}$.

Show that V is isomorphic with $\mathcal{F}(\mathcal{B})$, the vector space of all functions $f \colon \mathcal{B} \to \mathbb{F}$.

Exercise 7.4. Show that $\{(0,-1),(\frac{1}{\sqrt{2}},-\frac{1}{\sqrt{2}})\}$ is a basis for $\mathbb{R}^2$ with its standard vector space structure.

Exercise 7.5. Show that the functions

$$\mathbf{e}_1 \colon \mathbb{R} \longrightarrow \mathbb{R}, \quad x \longmapsto \cos(x + \tfrac{\pi}{4})$$
$$\mathbf{e}_2 \colon \mathbb{R} \longrightarrow \mathbb{R}, \quad x \longmapsto \cos(x + \tfrac{\pi}{2})$$

form a basis for $V = \{f \colon \mathbb{R} \to \mathbb{R} \mid \frac{d^2 f}{dx^2} + f = 0\}$.

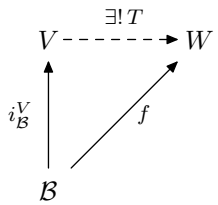
Exercise 7.6. Show that $\{(1,0),(1,1)\}$ is a basis for $\mathbb{R}^2$ as real vector space.

Exercise 7.7. Let $\mathcal{B} = \{\mathbf{e}_1, \ldots, \mathbf{e}_m\}$ be a basis for the vector space V over the field $\mathbb{F}$.

Prove that for every vector space, W, over $\mathbb{F}$ and every function, $f : \mathcal{B} \to W$, there is a unique linear transformation $T \colon V \longrightarrow W$ such that for all $j \in \{1, \ldots, m\}$,

$$T(\mathbf{e}_j) = \varphi(\mathbf{e}_j)$$

This is the *universal property* of a basis, conveniently expressed by the commutative diagram



Put differently, good general theory does not search for the maximum generality, but for the right generality.

Saunders MacLane

Chapter 8

Classification of Finitely Generated Vector Spaces

Recall that two vector space over a given field are equivalent (as vector spaces) if and only if they are isomorphic. This raises the classification problem for vector spaces:

Given vector spaces $\mathbf{V}$ and $\mathbf{W}$ over the field $\mathbb{F}$, decide whether $\mathbf{V} \cong \mathbf{W}$.

This problem has an elegant solution. We can assign to each vector space, V, over the field $\mathbb{F}$ a numerical invariant, its *dimension*, $\dim_{\mathbb{F}}(V)$, which solves the classification problem completely.

Main Theorem (Classification Theorem). *Given vector spaces V, W over the field $\mathbb{F}$, $V \cong W$ if and only if $\dim_{\mathbb{F}}(V) = \dim_{\mathbb{F}}(W)$.*

This chapter is devoted to introducing the necessary concepts for and a proof of the Classification Theorem for finitely generated vector spaces.[1]

The first concept we need is that of the *dimension* of a vector space.

Definition 8.1. Let V be a finitely generated vector space over $\mathbb{F}$. The *dimension of V over $\mathbb{F}$*, $\dim_{\mathbb{F}} V$, is the number of vectors in a basis for V.

Observation 8.2. Since the dimension of a vector space is defined in terms of the number of elements in a basis for the vector space, it is not immediately clear that the dimension of a vector space depends only on the vector space itself, and not on the choice of a particular basis.

Theorem 7.15 on page 84 solved part of the problem, at least for finitely generated vector spaces, by proving that every finitely generated vector space does, indeed, have a basis.

To justify Definition 8.1, it then remains to show that any two bases for the same vector space must have the same number of elements.

We formulate this in our next theorem.

Theorem 8.3. *If $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ and $\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ are bases for the vector space V, then $m = n$.*

We do not prove Theorem 8.3 immediately. Rather, it is a corollary to another theorem, which we illustrate with an explicit example before formulating and proving it.

[1] The theorem actually holds for all vector spaces. Since the general case uses the Axiom of Choice, we omit it.

Example 8.4. Let $U = \langle \mathbf{u}_{1,}, \mathbf{u}_2 \rangle$ be a vector subspace of V. Consider $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3 \in U$, where

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

We add suitable multiples of $\mathbf{v}_1$ to $\mathbf{v}_2$ and $\mathbf{v}_3$ to eliminate $\mathbf{u}_1$, obtaining

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

It follows that

$$\mathbf{v}_1 + \mathbf{v}_2 = 2(\mathbf{v}_1 - \mathbf{v}_3),$$

or

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

Thus , we see that $\mathbf{v}_1, \mathbf{v}_2$ and $\mathbf{v}_3$ are linearly dependent.

Our next theorem shows that Example 8.4 is generic, and our proof follows the above calculation.

Theorem 8.5. *Let U be a vector subspace of the vector space V.*

If U can be generated by a set of n vectors, then any set of more than n vectors from U is linearly dependent.

Observation 8.6. This apparently innocuous technical result, whose proof is just an extension of the calculation in Example 8.4 is actually the key to many important results in the theory of finitely generated vector spaces and its applications.

Proof of Theorem 8.5. We use induction on n.

$\mathbf{n = 1}$: In this case $U = \langle \mathbf{u} \rangle$ for some $\mathbf{u} \in V$.

Take $\mathbf{v}_1, \ldots, \mathbf{v}_m \in U$ for some $m > 1$.

Then there are $\lambda_1, \ldots, \lambda_m \in \mathbb{F}$ such that for each $i \in \{1, \ldots, m\}$

$$\mathbf{v}_i = \lambda_i \mathbf{u}$$

If for some i, $\lambda_i = 0$, then $\mathbf{v}_i = \mathbf{0}_V$, whence $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are linearly dependent.

If no $\lambda_i = 0$, then

$$\begin{aligned} \lambda_2\mathbf{v}_1 - \lambda_1\mathbf{v}_2 + 0\mathbf{v}_3 + \cdots + 0\mathbf{v}_m &= \lambda_2\lambda_1\mathbf{u} - \lambda_1\lambda_2\mathbf{u} + \mathbf{0}_V + \cdots \mathbf{0}_V \\ &= \mathbf{0}_V, \end{aligned}$$

showing that $\mathbf{v}_1, \ldots, \mathbf{v}_m$ are linearly dependent.

$\mathbf{n > 1}$: We make the inductive hypothesis that if a vector subspace, S, of V can be generated by $n-1$ vectors, then every set of more than $n-1$ vectors in S must be linearly dependent.

Let $U := \langle \mathbf{u}_1, \ldots, \mathbf{u}_n \rangle$ be a vector subspace of V and put $S := \langle \mathbf{u}_2, \ldots, \mathbf{u}_n \rangle$.

If $\mathbf{v}_1, \ldots, \mathbf{v}_m \in U$, then, for each $i \in \{1, \ldots, m\}$ there are $\lambda_{ij} \in \mathbb{F} \quad 1 \leq j \leq n$ such that

$$\mathbf{v}_i = \sum_{j=1}^{n} \lambda_{ij} \mathbf{u}_i$$

Suppose $m > n$.

If $\lambda_{i1} = 0$ for every i, then $\mathbf{v}_1, \ldots, \mathbf{v}_m \in S$ and we have $m > n > n-1$ vectors in the vector subspace S of V generated by $n-1$ vectors.

By the inductive hypothesis, $\mathbf{v}_1, \ldots, \mathbf{v}_m$ must be linearly dependent.

Otherwise $\lambda_{i1} \neq 0$ for some i.

Renumbering the vectors if necessary, we may assume that $\lambda_{11} \neq 0$. Then, for each $i > 1$,

$$\begin{aligned}\lambda_{11}\mathbf{v}_i - \lambda_{i1}\mathbf{v}_1 &= \sum_{j=1}^{n}(\lambda_{11}\lambda_{ij} - \lambda_{i1}\lambda_{1j})\mathbf{u}_j \\ &= \sum_{j=2}^{n}(\lambda_{11}\lambda_{ij} - \lambda_{i1}\lambda_{1j})\mathbf{u}_j \qquad \text{as } \lambda_{11}\lambda_{i1} - \lambda_{i1}\lambda_{11} = 0\end{aligned}$$

Thus if for $i \geq 2$ we put $\mathbf{w}_i := \lambda_{11}\mathbf{v}_i - \lambda_{i1}\mathbf{v}_1$, we obtain $m-1$ vectors, $\mathbf{w}_2, \ldots, \mathbf{w}_m$, in S.

Since S is generated by $n-1$ vectors and $m-1 > n-1$, it follows from the inductive hypothesis, that $\mathbf{w}_1, \ldots, \mathbf{w}_{m-1}$ are linearly dependent.

Hence there are $\alpha_2, \ldots \alpha_m \in \mathbb{F}$, not all 0, such that

$$\alpha_2\mathbf{w}_2 + \cdots + \alpha_m\mathbf{w}_m = \mathbf{0}_V.$$

Putting $\alpha_1 := -\alpha_2\lambda_{21} - \ldots - \alpha_m\lambda_{m1}$, we have

$$\alpha_1\mathbf{v}_1 + \alpha_2\lambda_{11}\mathbf{v}_2 + \cdots + \alpha_m\lambda_{11}\mathbf{v}_m = \mathbf{0}_V$$

Since $\lambda_{11} \neq 0$, at least one $\alpha_i\lambda_{11} \neq 0$, showing that $\mathbf{v}_1, \ldots, \mathbf{v}_m$ are linearly dependent. □

Corollary 8.7 (**Theorem 8.3 on page 87**). *Let $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ and $\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ be bases for the vector space V. Then $m = n$.*

Proof. Since $\mathbf{u}_1, \ldots, \mathbf{u}_n$ form a basis for V, we have $V = \langle \mathbf{u}_1, \ldots, \mathbf{u}_n \rangle$.

Since $\mathbf{v}_1, \ldots, \mathbf{v}_m$ form a basis for V, the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_m \in V$ are linearly independent.

Hence, by Theorem 8.5 on the preceding page, $m \leq n$.

Reversing the rôles of the $\mathbf{u}$s and the $\mathbf{v}$s, it follows that $n \leq m$.

Thus $m = n$. □

Observation 8.8. By proving Theorem 8.3 on page 87 we have completed showing that, in the case of finitely generated vector spaces, the notion of the *dimension* of the vector space V, $\dim_{\mathbb{F}} V$, defined in Definition 8.1 on page 87 as the number of vectors in a basis for V, is well defined, as it depends only on the vector space in question, and not on the choice of basis.

For this reason, finitely generated vector spaces are often also called *finite dimensional* vector spaces

Before we continuing our study of vector spaces and linear transformations, we deduce further corollaries which will prove to be useful, especially in applications of linear algebra.

Corollary 8.9. *Let V be a vector space of dimension n. If $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are linearly independent, they must generate V (and hence form a basis for V).*

Proof. Take any $\mathbf{x} \in V$.

Since $\dim V = n$, V is generated by a set of n vectors.

Hence, by Theorem 8.5 on page 88, the $n+1$ vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n, \mathbf{x}$ must be linearly dependent.

Since $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are linearly dependent, it follows from Theorem 7.7 on page 81 that $\mathbf{x}$ is a linear combination of $\mathbf{v}_1, \ldots, \mathbf{v}_n$.

Thus $\langle \mathbf{v}_1, \ldots, \mathbf{v}_n \rangle = V$. □

Corollary 8.10. *Let V be a vector space of dimension n. If $\mathbf{v}_1, \ldots, \mathbf{v}_n$ generate V, they must be linearly independent (and hence form a basis for V).*

Proof. Since $\langle \mathbf{v}_1, \ldots, \mathbf{v}_n \rangle = V$, it follows by the method used in the proof of Theorem 7.15 on page 84, that some subset of $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ containing $q \leq n$ vectors must be linearly independent and still generate V, thus forming a basis for V.

But $\dim V = n$. By Theorem 8.3 on page 87, $q = n$.

Thus, $\mathbf{v}_1, \ldots, \mathbf{v}_n$ must be linearly independent. □

We combine Corollary 8.9 on the preceding page and Corollary 8.10 in the next theorem.

Theorem 8.11. *Let V be an n-dimensional vector space, that is* $\dim_{\mathbb{F}} V = n$.
Given $\mathbf{v}_1, \ldots, \mathbf{v}_n \in V$, the following are equivalent.

(i) $\mathbf{v}_1, \ldots, \mathbf{v}_n$ *are linearly independent.*

(ii) $\langle \mathbf{v}_1, \ldots, \mathbf{v}_n \rangle = V$, *that is,* $\mathbf{v}_1, \ldots, \mathbf{v}_n$ *generate V.*

(iii) $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ *is a basis for V.*

Our next result is a refinement of Theorem 7.10 on page 82.

Lemma 8.12. *Let V and W be finite dimensional vector spaces, $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ a basis for V and $T\colon V \longrightarrow W$ a linear transformation.*

(i) T is injective if and only if $T(\mathbf{v}_1), \ldots, T(\mathbf{v}_n)$ are linearly independent.

(ii) T is surjective if and only if $T(\mathbf{v}_1), \ldots, T(\mathbf{v}_n)$ generate W.

(iii) T is an isomorphism if and only if $\{T(\mathbf{v}_1), \ldots, T(\mathbf{v}_n)\}$ is a basis for W.

Proof. **(i)** $\Rightarrow$: Suppose that T is injective and that $\lambda_1 T(\mathbf{v}_1) + \cdots + \lambda_n T(\mathbf{v}_n) = \mathbf{0}_W$.

By the linearity of T, $T(\lambda_1 \mathbf{v}_1 + \cdots + \lambda_n \mathbf{v}_n) = \mathbf{0}_W$.

Thus, by the injectivity of T, $\lambda_1 \mathbf{v}_1 + \cdots + \lambda_n \mathbf{v}_n = \mathbf{0}_V$.

But $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are linearly independent.

Hence $\lambda_1 = \cdots = \lambda_n = 0$, showing that $T(\mathbf{v}_1), \ldots, T(\mathbf{v}_n)$ are linearly independent.

(i) $\Leftarrow$: Suppose that $T(\mathbf{v}_1), \ldots, T(\mathbf{v}_n)$ are linearly independent and take $\mathbf{x} \in V$.

Since $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ a basis for V, $\mathbf{x} = \sum_{j=1}^{n} x_j \mathbf{v}_j$, for uniquely determined $x_1, \ldots, x_n \in \mathbb{F}$.

Since T is a linear transformation, $T(\mathbf{x}) = \sum_{j=1}^{n} x_j T(\mathbf{v}_j)$.

Hence, $\mathbf{x} \in \ker(T)$ only if

$$\sum_{j=1}^{n} x_j T(\mathbf{v}_j) = \mathbf{0}_W$$

Since $T(\mathbf{v}_1), \ldots, T(\mathbf{v}_n)$ are linearly independent, $x_1 = \cdots = x_n = 0$.

Thus $\mathbf{x} = \mathbf{0}_V$, showing that T is injective.

(ii) $\Rightarrow$: Suppose that T is surjective and take $\mathbf{y} \in W$.

By the surjectivity of T, $\mathbf{y} = T(\mathbf{x})$ for some $\mathbf{x} \in V$.

Since $\langle \mathbf{v}_1, \ldots, \mathbf{v}_n \rangle = V$, $\mathbf{x} = \lambda_1 \mathbf{v}_1 + \cdots + \lambda_n \mathbf{v}_n$.

Since T is a linear transformation, $\mathbf{y} = T(\mathbf{x}) = \lambda_1 T(\mathbf{v}_1) + \cdots + \lambda_n T(\mathbf{v}_n)$.

Thus $T(\mathbf{v}_1), \ldots, T(\mathbf{v}_n)$ generate W.

(ii) $\Leftarrow$: Suppose that $T(\mathbf{v}_1), \ldots, T(\mathbf{v}_n)$ generate W.

Take $\mathbf{y} \in W$.

Then $\mathbf{y} = \lambda_1 T(\mathbf{v}_1) + \cdots + \lambda_m T(\mathbf{v}_m)$, since $T(\mathbf{v}_1), \ldots, T(\mathbf{v}_m)$ generate W.

Since T is a linear transformation, $\mathbf{y} = T(\mathbf{x})$ for $\mathbf{x} = \lambda_1 \mathbf{e}_1 + \cdots + \lambda_m \mathbf{e}_m$.

Thus T is surjective.

(iii) Exercise. □

Corollary 8.13. *Let $T \colon V \longrightarrow W$ be a linear transformation.*

(a) If T is injective, then $\dim(V) \leq \dim(W)$.

(b) If T is surjective, then $\dim(V) \geq \dim(W)$.

Proof. Exercise. □

Corollary 8.14. *Let $T \colon V \longrightarrow V$ be an emndomorphism of the finitely generated vector space, V. Then the following are equivalent.*

(i) T is injective.

(ii) T is surjective.

(iii) T is an isomorphism.

Proof. Let $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be a basis for V, and consider $\{T(\mathbf{v}_1), \ldots, T(\mathbf{v}_n)\}$.

By Theorem 7.10 on page 82, T is injective if and only if $T(\mathbf{v}_1), \ldots, T(\mathbf{v}_n)$ are linearly independent.

As $\dim(V) = n$, it follows from Theorem 8.12 on the facing page, that $T(\mathbf{v}_1), \ldots, T(\mathbf{v}_n)$ are linearly independent if and only if they generate V.

By Theorem 7.10 on page 82, $T(\mathbf{v}_1), \ldots, T(\mathbf{v}_n)$ generate V if and only if T is surjective.

This shows that (i) and (ii) are equivalent.

Hence, each of (i) and (ii) is equivalent to T's being a bijective linear transformation.

By Theorem 5.19 on page 57 this is equivalent to T's being an isomorphism. □

The following example show that Corollary 8.14 does not hold when V is not finitely generated.

Example 8.15. Put $V := \mathbb{F}[t]$, the vector space of all polynomials in the indeterminate t with coefficients in $\mathbb{F}$.

$$T\colon V \longrightarrow V, \quad \sum_{j=0}^{n} a_j t^j \longmapsto \sum_{j=0}^{n} a_j t^{j+1}$$

is an injective endomorphism which is not surjective.

Corollary 8.16. *$\mathbb{F}^m \cong \mathbb{F}^n$ if and only if $m = n$.*

Proof. Plainly, only the "only if" part requires proof.

Let $T\colon \mathbb{F}^m \longrightarrow \mathbb{F}^n$ be an isomorphism.

The standard basis for for $\mathbb{F}^m$, $\{(1, 0, \ldots, 0), \ldots, (0, \ldots, 0, 1)\}$, is a basis for $\mathbb{F}^m$

By Corollary 8.12 on page 90(iii), the m vectors $T(1, 0, \ldots, 0), \ldots, T(0, \ldots, 0, 1)$ comprise a basis for $\mathbb{F}^n$.

Since the standard basis for $\mathbb{F}^n$ contains n vectors, it follows from Theorem 8.3 on page 87 that $m = n$. □

The choice of a basis for the finitely generated vector space V is really the choice of an isomorphism $V \longrightarrow \mathbb{F}^n$, where $n = \dim_{\mathbb{F}} V$, as we show in the next theorem.

Theorem 8.17. *Let V be a finitely generated vector space over $\mathbb{F}$. Then $\dim_{\mathbb{F}} V = n$ if and only if V is isomorphic with $\mathbb{F}^n$.*

Proof. $\Rightarrow$: Let $\mathbf{e}_1, \ldots, \mathbf{e}_n$ be a basis for V.

By Theorem 7.13 on page 83, each $\mathbf{x} \in V$ can be written uniquely as

$$\mathbf{x} = \sum_{j=1}^{n} x_j \mathbf{e}_j$$

with $x_1, \ldots, x_n \in \mathbb{F}$.

In other words, we have a bijection

$$T\colon V \longrightarrow \mathbb{F}^n, \quad \mathbf{x} \longmapsto (x_1, \ldots, x_n)$$

We show that T is a linear transformation.

For $\mathbf{x} = x_1\mathbf{e}_1 + \cdots x_n\mathbf{e}_n$, $\mathbf{y} = y_1\mathbf{e}_1 + \cdots + y_n\mathbf{e}_n \in V$ and $\lambda, \mu \in \mathbb{F}$

$$\begin{aligned} T(\lambda\mathbf{x} + \mu\mathbf{y}) &= T(\lambda \sum_{i=1}^{n} x_i\mathbf{e}_i + \mu \sum_{i=1}^{n} y_i\mathbf{e}_i) \\ &= T(\sum_{i=1}^{n} (\lambda x_i + \mu y_i)\mathbf{e}_i) \\ &= (\lambda x_1 + \mu y_1, \ldots, \lambda x_n + \mu y_n) \\ &= \lambda(x_1, \ldots, x_n) + \mu(y_1, \ldots, y_n) \\ &= \lambda T(\mathbf{x}) + \mu T(\mathbf{y}) \end{aligned}$$

Being a bijection linear transformations, T is, by Theorem 5.19 on page 57, an isomorphism.

$\Leftarrow$: Let $S\colon \mathbb{F}^n \longrightarrow V$ be an isomorphism and $\mathbf{e}_1, \ldots, \mathbf{e}_n$ the standard basis for $\mathbb{F}^n$.

By Lemma 8.12 on page 90, $\{S(\mathbf{e}_1), \ldots, S(\mathbf{e}_n)\}$ is a basis for V, whence $\dim_{\mathbb{F}}(V) = n$. □

Corollary 8.18 (**Classification Theorem for Finitely Generated Vector Spaces**). *Two finitely generated vector spaces over the same field are isomorphic if and only if they have the same dimension.*

Proof. By Theorem 8.17 on the preceding page, two finitely generated vector spaces over the same field have the same dimension, n, if and only if they are both isomorphic with $\mathbb{F}^n$. □

Observation 8.19. The Classification Theorem for Finitely Generated Vector Spaces provides us with a complete answer to the question "When are two finitely generated vector spaces over $\mathbb{F}$ isomorphic?".

But its significance does not end there. The proof of the theorem makes it clear that if V is a finitely generated vector space over $\mathbb{F}$, then there are as many different isomorphisms $V \cong \mathbb{F}^n$ as their are choices of a basis for V.

Choosing a basis for $\mathbf{V}$ is the same as choosing an isomorphism $\mathbf{V} \cong \mathbb{F}^{\mathbf{n}}$.

This is the first step to the use of matrices for calculations involving linear transformations between finitely generated vector spaces.

Theorem 8.17 on the facing page can also be formulated in terms of direct sums, illustrating the close relationship between various conceptions and constructions introduced earlier, which, at the time, seemed to have nothing to do with each other.

Theorem 8.20. *Let V be a finitely generated vector space over $\mathbb{F}$. Then*

$$V \cong \mathbb{F} \oplus \cdots \oplus \mathbb{F},$$

where the number of copies of $\mathbb{F}$ in the direct sum is precisely the dimension of V.

Proof. By Theorem 8.17 on the preceding page, it is enough to show that the dimension of $\mathbb{F}\oplus\cdots\oplus\mathbb{F}$ is n, the number of copies of $\mathbb{F}$ in the direct sum.

By Theorem 7.13 on page 83, the n vectors $(1, 0, \ldots, 0), (0, 1, 0, \ldots, 0), \ldots, (0, \ldots, 0, 1)$ comprise a basis — called *the standard basis* — for $\mathbb{F} \oplus \cdots \oplus \mathbb{F}$. □

We provide another model for finitely generated vector spaces.

Theorem 8.21. *Every finitely generated vector space over the field $\mathbb{F}$ is isomorphic with one of the form $\mathcal{F}(X, \mathbb{F}) = \{f \colon X \longrightarrow \mathbb{F} \mid f \text{ is a function}\}$.*

Proof. We present the essential idea for a proof, leaving the details as an exercise for the reader.

Let $X = \{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ be a basis for the vector space V over $\mathbb{F}$.

Then the function

$$T \colon \{f \colon X \longrightarrow \mathbb{F} \mid f \text{ is a function}\} \longrightarrow V, \quad f \longmapsto \sum_{j=1}^{n} f(\mathbf{e}_j)\mathbf{e}_j$$

is an isomorphism of vector spaces over $\mathbb{F}$.

□

Observation 8.22. The classification of finitely generated vector spaces in this chapter shows that several of the examples of vector spaces provided in Chapter 3 are essentially the same vector spaces, but presented differently — they are isomorphic as vector spaces. For example, Theorem 8.21 asserts that Examples 3.13 on page 35 and 3.18 on page 36 define isomorphic vector spaces when and only when the set X in Example 3.18 on page 36 has precisely n elements.

Observation 8.23. By the definition of infinite direct sums, Theorem 8.20 on the previous page holds even without the restriction to finite dimensional spaces.

In the case of vector spaces which are not finitely generated, Theorem 8.21 on the preceding page requires replacing the vector space

$$\{f\colon X \longrightarrow \mathbb{F} \mid f \text{ is a function}\}$$

by its vector subspace

$$\{f\colon X \longrightarrow \mathbb{F} \mid f \text{ is a function with} f(x) \neq 0 \text{ for only finitely many } x \in X\}$$

as in Exercise **??** on page ??.

8.1 The Universal Property of a Basis

We have shown that every finitely generated vector space has a basis, and that the number of vectors in a basis for a fixed vector space is independent of the choice of basis. This enabled a complete classification of finitely generated vector spaces over a fixed field up to isomorphism in terms of a single intrinsic numerical invariant, the *dimension*, which is the number of vectors in any basis for V.

Example 3.13 on page 35 showed that for every natural number n, $\mathbb{F}^n$ admits a standard vector space structure, and we have seen in this chapter that every finitely generated vector space is isomorphic to precisely one such vector space, namely $\mathbb{F}^{\dim V}$, with the choice of a basis providing an isomorphism.

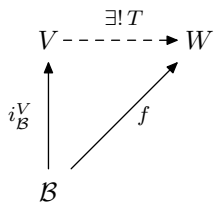
Hence, up to isomorphism, finitely generated vector spaces over a field are in bijection with $\mathbb{N}$, the set of all natural numbers.

While this is already sufficient to justify the importance of bases, they have another property with far-reaching consequences. Specifically, a basis does not only determine a vector space up to isomorphism, it also determines completely all linear transformations defined on a vector space.

Theorem 8.24 (**Universal Property of a Basis**). *Let $\mathcal{B}$ be a basis for the vector space V over the field $\mathbb{F}$.*

Given any vector space W over $\mathbb{F}$ and any function $f\colon \mathcal{B} \longrightarrow W$, there is a unique linear transformation $T\colon V \longrightarrow W$ with $T(\mathbf{e}) = f(\mathbf{e})$ for every $\mathbf{e} \in \mathcal{B}$.

This is expressed diagrammatical by



(∗)

Observation 8.25. The significance and practical importance of Theorem 8.24 cannot be overstated. For it shows that every linear transformation, T, defined on V is completely determined by the values T takes on any basis for V, and that we can assign any value to any of the vectors in such a basis — we are *free* to choose the values of T on the basis vectors in any way whatsoever.

This is particularly useful when $\mathbb{F}$ is an infinite field, for then it reduces an in principle infinite calculation to a finite one.

We expressed this property in the form of a universal property because universal properties play a central rôle in modern mathematics. As the reader will meet more examples when studying more advanced topics, we have taken this opportunity to work through an example.

Proof of Theorem 8.24. The commutativity of $(*)$ is equivalent to $f = T \circ i^V_{\mathcal{B}}$.

Given $\mathbf{e} \in \mathcal{B}$. Then

$$\begin{aligned} T(\mathbf{e}) &= T(i^V_{\mathcal{B}}(\mathbf{e})) \\ &= (T \circ i^V_{\mathcal{B}})(\mathbf{e}) \\ &= f(\mathbf{e}) \end{aligned}$$

This forces the definition of T.

For given $\mathbf{v} \in V$, there are uniquely determined $n \in \mathbb{N}$, $\mathbf{e}_1, \ldots, \mathbf{v}_n \in \mathcal{B}$ and $x_1, \ldots, x_n \in \mathbb{F}$ with

$$\mathbf{v} = \sum_{j=1}^{n} x_j \mathbf{e}_j$$

Hence, in order for T to be a linear transformation satisfying the requirements, we must have

$$\begin{aligned} T(\mathbf{v}) &= \sum_{j=1}^{n} x_j T(\mathbf{e}_j) \\ &= \sum_{j=1}^{n} x_j f(\mathbf{e}_j) \end{aligned}$$

The only possible definition of T is. therefore,

$$T\colon V \longrightarrow W, \quad \sum_{j=1}^{n} x_j \mathbf{e}_j \longmapsto \sum_{j=1}^{n} x_j f(\mathbf{e}_j)$$

It is plain from the discussion above, that T is a function.

To see that T is additive, take $\mathbf{v} = \sum_{j=1}^{n} x_j \mathbf{e}_j, \mathbf{v}' = \sum_{j=1}^{n} x'_j \mathbf{e}_j \in V$. Then

$$\begin{aligned} T(\mathbf{v} + \mathbf{v}') = \; & T\Big(\sum_{j=1}^{n} x_j \mathbf{e}_j + \sum_{j=1}^{n} x'_j \mathbf{e}_j \Big) \\ = \; & T\Big(\sum_{j=1}^{n} (x_j + x'_j) \mathbf{e}_j \Big) \\ = \; & \sum_{j=1}^{n} (x_j + x'_j) f(\mathbf{e}_j) \\ = \; & \sum_{j=1}^{n} x_j f(\mathbf{e}_j) + \sum_{j=1}^{n} x'_j f(\mathbf{e}_j) \\ = \; & T(\mathbf{v}) + T(\mathbf{v}') \end{aligned}$$

To see that T is homogenous, and hence a linear transformation, take $\alpha \in \mathbb{F}$. Then

$$T(\alpha \mathbf{v}) = \; T\Big(\alpha \sum_{j=1}^{n} x_j \mathbf{e}_j \Big)$$

$$
\begin{aligned}
&= T\Big(\sum_{j=1}^{n} \alpha\big(x_j \mathbf{e}_j\big)\Big) \\
&= T\Big(\sum_{j=1}^{n} (\alpha x_j)\mathbf{e}_j\Big) \\
&= \sum_{j=1}^{n} (\alpha x_j) f(\mathbf{e}_j) \\
&= \sum_{j=1}^{n} \alpha\big(x_j f(\mathbf{e}_j)\big) \\
&= \alpha \sum_{j=1}^{n} x_j f(\mathbf{e}_j) \\
&= \alpha T(\mathbf{v})
\end{aligned}
$$

□

8.2 Exercises

Exercise 8.1. Let $\mathcal{C}^\infty(\mathbb{R})$ be the real vector space of all smooth — that is, infinitely differentiable — real-valued functions of a real variable. Put

$$V := \{f \in \mathcal{C}^\infty(\mathbb{R}) \mid \frac{d^2 f}{dx^2} + f = 0\}.$$

Prove that V is a real vector space, and that it is isomorphic to $\mathcal{P}_1$, the real vector space of all real polynomials of degree less than two.

Exercise 8.2. Let V and W be vector spaces over the field $\mathbb{F}$, $\mathcal{B}$ be a basis for V, $\mathcal{C}$ a basis for W and $\varphi : \mathcal{B} \longrightarrow \mathcal{C}$ a function.

By Exercise 7.7 on page 85, or Theorem 8.24 on page 94, there is a unique linear transformation $T : V \longrightarrow W$ such that $T(\mathbf{v}) = \varphi(\mathbf{v})$ for all $\mathbf{v} \in \mathcal{B}$.

Prove that this T is an isomorphism if and only if φ is bijective.

Exercise 8.3. Let V be a finitely generated vector space over $\mathbb{F}$ and W a vector subspace of V. Prove that if $\dim_{\mathbb{F}}(W) = \dim_{\mathbb{F}}(V)$, then $W = V$.

Exercise 8.4. Prove that every finitely generated vector space over $\mathbb{F}$ is (isomorphic with one) of the form

$$\mathcal{F}(X, \mathbb{F}) := \{f \colon X \longrightarrow \mathbb{F} \mid f \text{ is a function}\}$$

Exercise 8.5. Put $V := \mathbb{R}[t]$, the vector space of all polynomials in the indeterminate t with coefficients in $\mathbb{R}$.

Prove that

$$I \colon V \longrightarrow V, \quad \sum_{j=0}^{n} a_j t^j \longmapsto \sum_{j=0}^{n} \frac{a_j}{j+1} t^{j+1}$$

is an injective endomorphism, which is not surjective.

Exercise 8.6. Given a field $\mathbb{F}$ and $n \in \mathbb{N}$, define

$$\mathbb{F}^{(n)} = \{ \begin{bmatrix} x_1 & \cdots x_n \end{bmatrix} \mid x_1, \ldots, x_n \in \mathbb{F} \}$$

$$\mathbb{F}_{(n)} = \left\{ \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \mid x_1, \ldots, x_n \in \mathbb{F} \right\}$$

For $\begin{bmatrix} x_1 & \ldots & x_n \end{bmatrix}, \begin{bmatrix} y_1 & \ldots & y_n \end{bmatrix}$ and $\lambda \in \mathbb{F}$, define

$$\begin{bmatrix} x_1 & \ldots & x_n \end{bmatrix} + \begin{bmatrix} y_1 & \ldots & y_n \end{bmatrix} = \begin{bmatrix} x_1 + y_1 & \ldots & x_n + y_n \end{bmatrix}$$
$$\lambda \cdot \begin{bmatrix} x_1 & \ldots & x_n \end{bmatrix} = \begin{bmatrix} \lambda x_1 & \ldots & \lambda x_n \end{bmatrix}$$

For $\begin{bmatrix} x_1 \\ \ldots \\ x_n \end{bmatrix}, \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$ and $\lambda \in \mathbb{F}$, define

$$\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_1 + y_1 \\ \vdots \\ x_n + y_n \end{bmatrix}$$
$$\lambda \cdot \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} \lambda x_1 \\ \vdots \\ \lambda x_n \end{bmatrix}$$

Show that both $\mathbb{F}^{(n)}$ and $\mathbb{F}_{(n)}$ are vector spaces over $\mathbb{F}$ with respect to these operations, and that each is isomorphic with $\mathbb{F}^n$, that is,

$$\mathbb{F}_{(n)} \cong \mathbb{F}^n \cong \mathbb{F}^{(n)}$$

If I only had an hour to solve a problem, I would use the first 55 minutes to pose the right question. For once I have formulated the right question, I would be able to solve the problem in less than five minutes.

Albert Einstein

Chapter 9

Matrix Representation of a Linear Transformation

We have developed the theory of vector spaces and linear transformations far enough to classify vector spaces over a fixed field up to isomorphism (at least for finitely generated vector spaces), thereby answering one of the central questions in linear algebra. Moreover, our answer is a particularly satisfying one, since whether or not two (finitely generated) vector spaces over the same field are isomorphic can be decided by calculating a single numerical invariant for each — its dimension — and the two vector spaces are isomorphic if and only if these two natural numbers are the same.

While this single result would be enough to justify the effort we have expended and the concepts we have introduced, our labours have actually borne more fruit, for we have made possible practical applications to many concrete situations. In this chapter we show how the theory we have developed can be used to devise and apply convenient computational methods.

9.1 Introducing Matrices

We restrict attention to finitely generated vector spaces over a fixed field $\mathbb{F}$ and exploit the fact that every finitely generated vector space admits a basis to introduce computational techniques of great power and practical importance.

Definition 9.1. Given counting numbers m and n, an $m \times n$ *matrix with coefficients in* $\mathbb{F}$, or an $m \times n$ *matrix over* $\mathbb{F}$ is an array of mn elements of $\mathbb{F}$, $\underline{\mathbf{A}}$, arranged in m *rows* and n *columns*.

We write

$$\underline{\mathbf{A}} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}$$

or, more compactly,

$$\underline{\mathbf{A}} = \Big[a_{ij}\Big]_{m\times n}$$

where $a_{ij} \in \mathbb{F}$ is the *coefficient* in the i^{th} row and j^{th} column, or $(i,j)^{th}$ *coefficient*, of $\underline{\mathbf{A}}$.

We write $\mathbf{M}(m \times n; \mathbb{F})$ for the set of all $m \times n$ matrices over $\mathbb{F}$, and $\mathbf{M}(n; \mathbb{F})$ when $m = n$.

Observation 9.2. The reader has met matrices in Example 3.16, where a vector space structure was introduced in an *ad hoc* manner, without any motivation or justification. Our discussion here will correct this.

Observation 9.3. In Exercise 8.6, the sets $\mathbb{F}_{(n)}$ and $\mathbb{F}^{(n)}$ were defined and a vector space structure was introduced for each in an *ad hoc* manner.

The elements of $\mathbb{F}_{(n)}$ are sometimes referred to as *column vectors* and those of $\mathbb{F}^{(n)}$ as *row vectors*.

Both of these vector spaces are isomorphic with $\mathbb{F}^n$. In fact the reader may have observed that the difference between the three vector spaces $\mathbb{F}^n$, $\mathbb{F}^{(n)}$ and $\mathbb{F}_{(n)}$ is, essentially, notational — we have an n-tuple of elements of $\mathbb{F}$, a row of n elements of $\mathbb{F}$ and a column of n elements of $\mathbb{F}$ respectively, and the operations are defined element-by-element. It is tempting to conclude that they are, in fact, the same vector space. But this temptation should be resisted. For while in this case we have the obvious isomorphisms

$$\mathbb{F}^n \longrightarrow \mathbb{F}^{(n)}, \quad (x_1, \ldots, x_n) \longmapsto \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix}$$

$$\mathbb{F}^n \longrightarrow \mathbb{F}_{(n)}, \quad (x_1, \ldots, x_n) \longmapsto \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

there are numerous other vector spaces isomorphic with each of these, in situations where there is no immediately obvious isomorphism and, even more importantly, there are many practical applications which require us to use different isomorphisms between $\mathbb{F}^n$ and $\mathbb{F}_{(n)}$, as we shall illustrate.

Another reason for distinguishing these three obviously isomorphic vector spaces is provided by Definition 9.1, which makes it apparent that

$$\mathbb{F}_{(n)} = \mathbf{M}(n \times 1; \mathbb{F})$$
$$\mathbb{F}^{(n)} = \mathbf{M}(1 \times n; \mathbb{F})$$

Thus, while $\mathbf{M}(m \times n; \mathbb{F})$ simultaneously generalises both $\mathbb{F}_{(m)}$ and $\mathbb{F}^{(n)}$, there is no such immediately obvious relationship between $\mathbb{F}^n$ and $\mathbf{M}(m \times n; \mathbb{F})$.

Let $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ be a basis for the vector space V and $\{\mathbf{f}_1, \ldots, \mathbf{f}_m\}$ for W. Let $T \colon V \longrightarrow W$ be a linear transformation.

Convention. When $W = V$, we assume that the same basis has been chosen in the co-domain as in the domain, unless otherwise specified.

Each $\mathbf{x} \in V$ can be expressed as a linear combination of $\mathbf{e}_1, \ldots, \mathbf{e}_n$ in precisely one way, say as,

$$\mathbf{x} = x_1\mathbf{e}_1 + \cdots + x_n\mathbf{e}_n = \sum_{j=1}^{n} x_j\mathbf{e}_j \tag{9.1}$$

with $x_j \in \mathbb{F} \quad (j = 1, \ldots, n)$.

Similarly, each $\mathbf{y} \in W$ can be written uniquely as

$$\mathbf{y} = y_1\mathbf{f}_1 + \cdots + y_m\mathbf{f}_m = \sum_{i=1}^{m} y_i\mathbf{f}_i \tag{9.2}$$

with $y_i \in \mathbb{F} \quad (i = 1, \ldots, m)$.

Put $\mathbf{y} = T(\mathbf{x})$. Then, since T is a linear transformation,

$$\begin{aligned}\mathbf{y} &= T(\mathbf{x}) \\ &= T(\sum_{j=1}^{n} x_j \mathbf{e}_j) \\ &= \sum_{j=1}^{n} x_j T(\mathbf{e}_j) \end{aligned} \tag{9.3}$$

This means T is completely determined by $T(\mathbf{e}_j) \quad (j = 1, \ldots, n)$.

Since $T(\mathbf{e}_j) \in W \quad (j = 1, \ldots, n)$,

$$T(\mathbf{e}_j) := \sum_{i=1}^{m} a_{ij} \mathbf{f}_i \tag{9.4}$$

for suitable (uniquely determined) $a_{ij} \in \mathbb{F} \quad (i = 1, \ldots, m, \ j = 1, \ldots, n)$.

It now follows from (9.3) and (9.4) that

$$\begin{aligned}\sum_{i=m}^{n} y_i \mathbf{f}_i &= \sum_{j=1}^{n} x_j \Big(\sum_{i=1}^{m} a_{ij} \mathbf{f}_i \Big) \\ &= \sum_{i=1}^{m} \Big(\sum_{j=1}^{n} x_j a_{ij} \Big) \mathbf{f}_i \end{aligned}$$

By uniqueness in (9.2), we deduce that

$$y_i = \sum_{j=1}^{n} x_j a_{ij} = \sum_{j=1}^{n} a_{ij} x_j. \tag{9.5}$$

We use matrices to rewrite this. This is our first step to developing convenient computational methods.

Since we have fixed a basis for V and a basis for W, we can use the uniqueness of the expressions in Equations (9.1), (9.2) and (9.4) to represent $\mathbf{x} \in V, \mathbf{y} \in W$ and $T \colon V \longrightarrow W$ by

$$\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \qquad \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}$$

respectively. We rewrite Equation 9.5

$$\begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \tag{9.6}$$

Observation 9.4. For the moment, this is just notation and nothing else. Think of it as merely a way of storing the data from which the linear transformation can be reconstructed. No algebraic operation has been defined here, even if the reader correctly anticipates further developments. It is important to realise that, at this stage, this is no more than a convenient notational convention.

Having introduced this notation to represent vectors and linear transformations, we turn to defining algebraic operations which reflect the operations we introduced earlier on vectors and on linear transformations.

Definition 9.5.

$$\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

is the *coordinate vector of the vector* $\mathbf{x} \in V$ *with respect to the basis* $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ *for* V and

$$\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}$$

is the *the matrix of the linear transformation* $T\colon V \longrightarrow W$ *with respect to the bases* $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ *for* V *and* $\{\mathbf{f}_1, \ldots, \mathbf{f}_m\}$ *for* W.

Observation 9.6. We see from Definition 9.5 that the number of rows of the matrix of a linear transformation is the dimension of the co-domain and the number of columns is the dimension of the domain of the linear transformation.

Observation 9.7. It is immediate form Equation 9.4 and Definition 9.5 that the j^{th} column of the matrix of T with respect to the basis $\{\mathbf{e_j} \mid 1 \leq j \leq n\}$ for V and $\{\mathbf{f}_i \mid 1 \leq i \leq m\}$ for W,

$$\begin{bmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{bmatrix},$$

is the coordinate vector with respect to the basis $\{\mathbf{f}_i \mid 1 \leq i \leq m\}$, of $T(\mathbf{e}_j)$, the image under T of $\mathbf{e}_j$, the j^{th} basis vector for V.

Theorem 9.8. *The matrix of* $id_V : V \longrightarrow V$ *with respect to any basis for* V *is*

$$\underline{\mathbf{1}}_n := [\delta_{ij}]_{n \times n}$$

where δ_{ij} *is Kronecker's "delta":*

$$\delta_{ij} := \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$

The matrix of the zero linear transformation, $0\colon V \longrightarrow W, \quad \mathbf{v} \longmapsto \mathbf{0}_W$, *with respect to the bases* $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ *for* V *and* $\{\mathbf{f}_1, \ldots, \mathbf{f}_m\}$ *for* W *is*

$$\underline{\mathbf{0}}_{m \times n} := [x_{ij}]_{m \times n}$$

with $x_{ij} = 0$ *for all* i, j.

Proof. The first statement follows from the fact that, for all j,

$$id_V(\mathbf{e}_j) = \mathbf{e}_j$$

and the second from the fact that, for all j

$$0(\mathbf{e}_j) = \mathbf{0}_W$$

□

Definition 9.9. $\underline{\mathbf{0}}_{m\times n}$ is the $(m \times n)$ *zero matrix* and $\underline{\mathbf{1}}_n$ the $(n \times n)$ *identity matrix*.

Observation 9.10. The matrix representations of the vector $\mathbf{x} \in V$ and the linear transformation $T : V \longrightarrow W$ as

$$\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}$$

depend critically upon the choice of bases for V and W, as the next examples illustrate.

Example 9.11. Take $\mathbb{F} := \mathbb{R}$, $V := \mathbb{R}^3$, $W := \mathbb{R}^2$ and $T : V \longrightarrow W,\ (x, y, z) \longmapsto (x, y)$.
Choose $\mathbf{e}_1 := (1, 0, 0)$, $\mathbf{e}_2 := (0, 1, 0)$, $\mathbf{e}_3 := (0, 0, 1) \in V$ and $\mathbf{f}_1 := (1, 0)$, $\mathbf{f}_2 := (0, 1) \in W$.

We verify that the above vectors do, in fact, forms bases for V and W respectively.
Take $\mathbf{x} = (x, y, z) \in V = \mathbb{R}^3$. Then $\mathbf{x} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + x_3\mathbf{e}_3$ if and only if

$$(x, y, z) = x_1(1, 0, 0) + x_2(0, 1, 0) + x_3(0, 0, 1) = (x_1, x_2, x_3)$$

so that $x_1 = x, x_2 = y, x_3 = z$ is the unique solution. The expression being unique, $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ does form a basis for $\mathbb{R}^3$ and the coordinate vector of $(x, y, z) \in \mathbb{R}^3$ with respect to the basis $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ is

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

An analogous calculation shows that $\{\mathbf{f}_1, \mathbf{f}_2\}$ is a basis for $W = \mathbb{R}^2$ and that the coordinate vector of $(u, v) \in \mathbb{R}^2$ with respect to the basis $\{\mathbf{f}_1, \mathbf{f}_2\}$ is

$$\begin{bmatrix} u \\ v \end{bmatrix}$$

Since

$$\begin{aligned} T(\mathbf{e}_1) &= T(1, 0, 0) = (1, 0) = \mathbf{f}_1 = 1\mathbf{f}_1 + 0\mathbf{f}_2 \\ T(\mathbf{e}_2) &= T(0, 1, 0) = (0, 1) = \mathbf{f}_2 = 0\mathbf{f}_1 + 1\mathbf{f}_2 \\ T(\mathbf{e}_3) &= T(0, 0, 1) = (0, 0) = \mathbf{0}_W = 0\mathbf{f}_1 + 0\mathbf{f}_2, \end{aligned}$$

the matrix of T with respect to the bases $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ for V and $\{\mathbf{f}_1, \mathbf{f}_2\}$ for W is

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

The matrix version of $T(x, y, z) = (x, y)$ is thus

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

Example 9.12. Take $\mathbb{F} := \mathbb{R}$, $V := \mathbb{R}^3$, $W := \mathbb{R}^2$ and $T\colon V \longrightarrow W,\ (x, y, z) \longmapsto (x, y)$.

Choose $\mathbf{e}_1 := (1, 0, 0),\ \mathbf{e}_2 := (1, 1, 0),\ \mathbf{e}_3 := (1, 1, 1) \in V$ and $\mathbf{f}_1 := (0, 1), \mathbf{f}_2 := (1, 1) \in W$.

We verify that the above vectors form bases for V and W respectively.

Take $\mathbf{x} = (x, y, z) \in V = \mathbb{R}^3$. Then $\mathbf{x} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + x_3\mathbf{e}_3$ if and only if

$$(x, y, z) = x_1(1, 0, 0) + x_2(1, 1, 0) + x_3(1, 1, 1) = (x_1 + x_2 + x_3, x_2 + x_3, x_3)$$

so that $x_3 = z, x_2 = y - z, x_1 = x - y$ is the unique solution. Since the expression is unique, $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ forms a basis for $\mathbb{R}^3$, and the coordinate vector of $(x, y, z) \in \mathbb{R}^3$ with respect to the basis $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ is

$$\begin{bmatrix} x - y \\ y - z \\ z \end{bmatrix}$$

To see that $\{\mathbf{f}_1, \mathbf{f}_2\}$ is a basis for $W = \mathbb{R}^2$, note that $(u, v) = y_1\mathbf{f}_1 + y_2\mathbf{f}_2$ if and only if

$$(u, v) = y_1(0, 1) + y_2(1, 1) = (y_2, y_1 + y_2),$$

which has the unique solution $y_2 = u, y_1 = v - u$ showing that the coordinate vector of $(u, v) \in \mathbb{R}^2$ with respect to the basis $\{\mathbf{f}_1, \mathbf{f}_2\}$ is

$$\begin{bmatrix} v - u \\ u \end{bmatrix}$$

Since

$$\begin{aligned} T(\mathbf{e}_1) &= T(1, 0, 0) = (1, 0) &&= -1\mathbf{f}_1 + 1\mathbf{f}_2 \\ T(\mathbf{e}_2) &= T(1, 1, 0) = (1, 1) &&= 0\mathbf{f}_1 + 1\mathbf{f}_2 \\ T(\mathbf{e}_3) &= T(1, 1, 1) = (1, 1) &&= 0\mathbf{f}_1 + 1\mathbf{f}_2, \end{aligned}$$

the matrix of T with respect to the bases $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ for V and $\{\mathbf{f}_1, \mathbf{f}_2\}$ for W is

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

The matrix version of $T(x, y, z) = (x, y)$ is thus

$$\begin{bmatrix} y - x \\ x \end{bmatrix} = \begin{bmatrix} -1 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} x - y \\ y - z \\ z \end{bmatrix}$$

Example 9.13. Let $\mathbb{F}[t]$ be the set of all polynomials in t with coefficients from $\mathbb{F}$. This is a real vector space with respect to the usual addition of polynomials and the usual multiplication of polynomials by constants:

Recall that as polynomials

$$\sum_{i=0}^{m} a_i t^i = \sum_{j=0}^{n} a_j t^j$$

if and only if $m = n$ and $a_i = b_i$ for each $i \in \{1, \ldots, n\}$.

Plainly, the subset, $\mathcal{P}_n$, of $\mathbb{F}[t]$ consisting of all polynomials, whose degree does not exceed n, forms a vector subspace of $\mathbb{F}[t]$.

Take $\mathbb{F} = \mathbb{R}, V = W = \mathcal{P}_2$ and

$$T\colon V \longrightarrow W, \quad p \longmapsto p' = \frac{d}{dt}(p),$$

where for $p = a_0 + \cdots + a_n t^n$

$$p' = a_1 + 2a_2 t + \cdots + n a_n t^{n-1}.$$

Choose the vectors $\mathbf{e}_1 = \mathbf{f}_1 = 1, \mathbf{e}_2 = \mathbf{f}_2 = t, \mathbf{e}_3 = \mathbf{f}_3 = t^2 \in V (= W)$.

To see that these form a basis of $\mathcal{P}_2$, observe that the definition of a polynomial means that every element of $\mathcal{P}_2$ can be written uniquely as a linear combination of $\mathbf{e}_1, \mathbf{e}_2$ and $\mathbf{e}_3$.

It follows that the coordinate vector of $a + bt + ct^2 \in \mathcal{P}_2$ with respect to the basis $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ is

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

Since

$$\begin{aligned} T(\mathbf{e}_1) &= \frac{d}{dt}(1) = 0 &&= 0\mathbf{f}_1 + 0\mathbf{f}_2 + 0\mathbf{f}_3 \\ T(\mathbf{e}_2) &= \frac{d}{dt}(t) = 1 &&= 1\mathbf{f}_1 + 0\mathbf{f}_2 + 0\mathbf{f}_3 \\ T(\mathbf{e}_3) &= \frac{d}{dt}(t^2) = 2t &&= 0\mathbf{f}_1 + 2\mathbf{f}_2 + 0\mathbf{f}_3, \end{aligned}$$

the matrix of T with respect to the bases $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ for V and $\{\mathbf{f}_1, \mathbf{f}_2, \mathbf{f}_3\}$ for W is

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

The matrix version of $T(p) = p'$ with respect to these bases is thus

$$\begin{bmatrix} b \\ 2c \\ 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix}$$

Example 9.14. We continue with the vector space V in the previous example.

Choose $\mathbf{e}_1 = t^2, \mathbf{e}_2 = t, \mathbf{e}_3 = 1 \in V$ and $\mathbf{f}_1 = 1, \mathbf{f}_2 = t + 1, \mathbf{f}_3 = t^2 + 1 \in W$.

Plainly $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ forms a basis for V, as it is just a re-ordering of our previous basis.

To see that $\{\mathbf{f}_1, \mathbf{f}_2, \mathbf{f}_3\}$ forms a basis of $\mathcal{P}_2$, note that $p(t) = a + bt + ct^2 = y_1\mathbf{f}_1 + y_2\mathbf{f}_2 + y_3\mathbf{f}_3$ if and only if

$$a + bt + ct^2 = y_1 + y_2(t + 1) + y_3(t^2 + 1) = (y_1 + y_2 + y_3) + y_2 t + y_3 t^2,$$

which is the case if and only if $y_1 = a - b - c, y_2 = b, y_3 = c$.

The coordinate vector of $a + bt + ct^2 \in V$ with respect to the basis $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ is therefore

$$\begin{bmatrix} c \\ b \\ a \end{bmatrix},$$

and that of $a + bt + ct^2 \in W$ with respect to the basis $\{\mathbf{f}_1, \mathbf{f}_2, \mathbf{f}_3\}$ is

$$\begin{bmatrix} a - b - c \\ b \\ c \end{bmatrix}$$

Since

$$\begin{aligned} T(\mathbf{e}_1) &= \frac{d}{dx}(x^2) = 2x = -2\mathbf{f}_1 + 2\mathbf{f}_2 + 0\mathbf{f}_3 \\ T(\mathbf{e}_2) &= \frac{d}{dx}(x) = 1 = 1\mathbf{f}_1 + 0\mathbf{f}_2 + 0\mathbf{f}_3 \\ T(\mathbf{e}_3) &= \frac{d}{dx}(1) = 0 = 0\mathbf{f}_1 + 0\mathbf{f}_2 + 0\mathbf{f}_3, \end{aligned}$$

the matrix of T with respect to the bases $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ for V and $\{\mathbf{f}_1, \mathbf{f}_2, \mathbf{f}_3\}$ for W is

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

The matrix version of $T(p) = p'$ with respect to these bases is thus

$$\begin{bmatrix} b - 2c \\ 2c \\ 0 \end{bmatrix} = \begin{bmatrix} -2 & 1 & 0 \\ 2 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} c \\ b \\ a \end{bmatrix}$$

9.2 The Relationship between Linear Transformations and Matrices

We have shown that choosing bases $\mathcal{B} = \{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ for the finitely generated vector space V and $\mathcal{C} = \{\mathbf{f}_1, \ldots, \mathbf{f}_m$ for W allows us to represent each linear transformation $T\colon V \longrightarrow W$ by an $m \times n$ matrix,$\underline{\mathbf{A}}$, with coefficients in the field of scalars $\mathbb{F}$.

We have seen that the matrix depends not only on the linear transformation itself, but also on the choice of the bases. In particular, the order of the vectors in a given basis matters, for the j^{th} column of $\underline{\mathbf{A}}$ is the coordinate vector of $T(\mathbf{e}_j)$ with respect to the basis $\{\mathbf{f}_1, \ldots, \mathbf{f}_m\}$ for W. We summarise the dependence of the matrix on the order of the basis vectors.

Permuting the basis vectors for the domain of T permutes the columns of $\underline{\mathbf{A}}$.

Permuting the basis vectors for the co-domain of T permutes the rows of $\underline{\mathbf{A}}$.

The fact that the choice of bases for V and W determines for each vector in V (resp. W) a unique co-ordinate vector and for linear transformation $T\colon V \longrightarrow W$ a unique matrix representing it means that choosing bases defines functions

$$\beta_{\mathcal{B}}\colon V \longrightarrow \mathbb{F}_{(n)}, \quad \mathbf{v} \longmapsto \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

$$\beta_{\mathcal{C}} \colon W \longrightarrow \mathbb{F}_{(m)}, \quad \mathbf{w} \longmapsto \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}$$

and

$$M_{\mathcal{B},\mathcal{C}} \colon \operatorname{Hom}_{\mathbb{F}}(V,W) \longrightarrow \mathbf{M}(m \times n; \mathbb{F}), \quad T \longmapsto \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}$$

as detailed above.

These functions, the key to calculations, have very pleasant features, the first of which we state in the next theorems.

Theorem 9.15. *For the basis $\mathcal{B} = \{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ for the $\mathbb{F}$-vector space V, the function*

$$\beta_{\mathcal{B}} : V \longrightarrow \mathbb{F}_{(n)}, \quad \mathbf{v} \longmapsto \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

when $\mathbf{v} = \sum_{j=1}^{n} x_j \mathbf{e}_j$, is a bijection.

Proof. That $\beta_{\mathcal{B}}$ is a bijection is as restatement of the fact that every element of V can be expressed as a linear combination of the elements of $\mathcal{B}$ in precisely one way. □

Observation 9.16. The reader may have noticed that if we take $\mathbb{F}_{(n)}$ with the vector space structure introduced in an *ad hoc* manner in Exercise 8.6, then $\beta_{\mathcal{B}}$ becomes in isomorphism.

Moreover, when $\mathbb{F}_{(n)}$ and $\mathbb{F}_{(m)}$ are taken with this vector space, making merely notational adjustments to the proof of Theorem 5.9 shows directly that $\operatorname{Hom}_{\mathbb{F}}(\mathbb{F}_{(n)}, \mathbb{F}_{(m)})$ is in bijection with $\mathbf{M}(m \times n; \mathbb{F})$.

This is true more generally.

Theorem 9.17. *Choose the basis $\mathcal{B} = \{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ for the $\mathbb{F}$-vector space V, and the basis $\mathcal{C} = \{\mathbf{f}_1, \ldots, \mathbf{f}_m\}$ for W.*

The function

$$M_{\mathcal{B},\mathcal{C}} : \operatorname{Hom}_{\mathbb{F}}(V,W) \longrightarrow \mathbf{M}(m \times n; \mathbb{F}), \quad T \longmapsto \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}$$

when $T(\mathbf{e}_j) = \sum_{i=1}^{m} a_{ij} \mathbf{f}_i \quad (1 \leq j \leq n)$ is a bijection.

Proof. That $M_{\mathcal{B},\mathcal{C}}$ is a bijection follows from the universal property of bases (Theorem 8.24). For

$$\underline{\mathbf{A}} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}$$

is the matrix of $T\colon V \longrightarrow W$ with respect to $\mathcal{B}$ and $\mathcal{C}$ if and only if

$$T(\mathbf{e}_j) = \sum_{i=1}^{m} a_{ij}\mathbf{f}_i$$

As it determines the values T takes on a basis for V, $\underline{\mathbf{A}} = M_{\mathcal{B},\mathcal{C}}(T)$ determines T uniquely. □

9.3 Algebraic Operations on Matrices

We introduce algebraic operations on matrices to reflect the algebraic operations on the linear transformations they represent, which were introduced in Chapter 6 — we can multiply a linear transformation by a scalar, we can add two linear transformations between the same vector spaces and we can compose two linear transformations when the co-domain of one agrees with the domain of the second.

9.3.1 The Matrix of the Scalar Multiple of a Linear Transformation

Let $\mathcal{B} = \{\mathbf{e}_1, \ldots \mathbf{e}_n\}$ be a basis for V and $\mathcal{C} = \{\mathbf{f}_1, \ldots, \mathbf{f}_m\}$ a basis for W. Then the matrix of the linear transformation $T\colon V \longrightarrow W$ with respect to $\mathcal{B}$ and $\mathcal{C}$ is

$$\underline{\mathbf{A}} = \Big[a_{ij}\Big]_{m\times n} \in \mathbf{M}(m \times n; \mathbb{F})$$

where

$$T(\mathbf{e}_j) = \sum_{i=1}^{m} a_{ij}\mathbf{f}_i$$

so that the coordinate vector of $T(\mathbf{e}_j)$ with respect to $\mathcal{C}$ is

$$\begin{bmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{bmatrix}$$

Take $\lambda \in \mathbb{F}$. For $1 \leq j \leq n$,

$$\begin{aligned} (\lambda \boxtimes T)(\mathbf{e}_j) &= \lambda\Big(T(\mathbf{e})_j\Big) && \text{by Definition 6.48} \\ &= \lambda\Big(\sum_{i=1}^{m} a_{ij}\mathbf{f}_i\Big) \\ &= \sum_{i=1}^{m} (\lambda a_{ij})\mathbf{f}_i \end{aligned}$$

Hence, the coordinate vector of $(\lambda \boxtimes T)(\mathbf{e}_j)$ with respect to $\mathcal{C}$ is

$$\begin{bmatrix} \lambda a_{1j} \\ \vdots \\ \lambda a_{mj} \end{bmatrix}$$

from which it follows that the matrix of $\lambda \boxtimes T$ with respect to $\mathcal{B}$ and $\mathcal{C}$ is

$$\Big[\lambda a_{ij}\Big]_{m\times n} \in \mathbf{M}(m \times n; \mathbb{F})$$

This motivates our definition of the scalar multiple of an $m \times n$ matrix.

Definition 9.18. The *multiplication by a scalar of an $m \times n$ matrix* over the field $\mathbb{F}$ is the function

$$\boxtimes : \mathbb{F} \times \mathbf{M}(m \times n; \mathbb{F}) \longrightarrow \mathbf{M}(m \times n; \mathbb{F}), \quad \big(\lambda, [a_{ij}]_{m\times n}\big) \longmapsto [\lambda a_{ij}]_{m\times n}$$

In other words,

the matrix resulting from multiplying by the scalar, λ, the matrix representing the linear transformation, T, is the matrix representing the linear transformation obtained by multiplyin T by λ.

Notational Convention. We write $\lambda\underline{\mathbf{A}}$ for $\boxtimes(\lambda, \underline{\mathbf{A}})$.

Observation 9.19. We note that there is no restriction on the matrix in question when it is to be multiply it by a scalar.

Observation 9.20. Let $\mathcal{B}$ be a basis for V and $\mathcal{C}$ a basis for W, with $\dim V = n$ and $\dim W = m$. It is immediate from Definition 9.18 that function

$$M_{\mathcal{B},\mathcal{C}} : \operatorname{Hom}(V, W) \longrightarrow \mathbf{M}(m \times n; \mathbb{F})$$

defined in Section 9.2, is homogeneous.

Example 9.21. We consider the linear transformation

$$T\colon \mathbb{R}^2 \longrightarrow \mathbb{R}^2, \quad (x, y) \longmapsto (x + 2y, 3x - y)$$

Then

$$3T\colon \mathbb{R}^2 \longrightarrow \mathbb{R}^2, \quad (x, y) \longmapsto 3(x + 2y, 3x - y) = (3x + 6y, 9x - 3y)$$

Using the basis $\{(1, 0), (0, 1)\}$ for each of the vector spaces, the matrix of T is

$$\underline{\mathbf{A}} = \begin{bmatrix} 1 & 2 \\ 3 & -1 \end{bmatrix}$$

and that of $3T$ is

$$\underline{\mathbf{B}} = \begin{bmatrix} 3 & 6 \\ 9 & -3 \end{bmatrix}$$

By Definition 9.18,

$$3 \begin{bmatrix} 1 & 2 \\ 3 & -1 \end{bmatrix} = \begin{bmatrix} 3 \times 1 & 3 \times 2 \\ 3 \times 3 & 3 \times (-1) \end{bmatrix} = \begin{bmatrix} 3 & 6 \\ 9 & -3 \end{bmatrix}$$

9.3.2 The Matrix of the Sum of Linear Transformations

Let $R\colon V \longrightarrow W$ be a linear transformation, whose matrix with respect to $\mathcal{B}$ and $\mathcal{C}$ is

$$\underline{\mathbf{B}} = \Big[b_{ij}\Big]_{m\times n} \in \mathbf{M}(m \times n; \mathbb{F})$$

so that the coordinated vector of $R(\mathbf{e}_j)$ with respect to $\mathcal{C}$ is

$$\begin{bmatrix} b_{1j} \\ \vdots \\ b_{mj} \end{bmatrix}$$

For $1 \leq j \leq n$,

$$\begin{aligned} (T \boxplus R)(\mathbf{e}_j) &= T(\mathbf{e}_j) + R(\mathbf{e}_j) && \text{by Definition 6.48} \\ &= \sum_{i=1}^{m} a_{ij}\mathbf{f}_i \ + \ \sum_{i=1}^{m} b_{ij}\mathbf{f}_i \\ &= \sum_{i=1}^{m} (a_{ij} + b_{ij})\mathbf{f}_i \end{aligned}$$

Hence, the coordinate vector of $(T \boxplus R)(\mathbf{e}_j)$ with respect to $\mathcal{C}$ is

$$\begin{bmatrix} a_{1j} + b_{1j} \\ \vdots \\ a_{mj} + bmj \end{bmatrix}$$

from which it follows that the matrix of $T \boxplus R$ with respect to $\mathcal{B}$ and $\mathcal{C}$ is

$$\Big[a_{ij} + b_{ij}\Big]_{m\times n} \in \mathbf{M}(m \times n; \mathbb{F})$$

This motivates our definition of matrix addition.

Definition 9.22. For counting numbers m and n, *addition of $m \times n$ matrices with coefficients in the field $\mathbb{F}$* is the function

$$\boxplus : \mathbf{M}(m \times n; \mathbb{F}) \times \mathbf{M}(m \times n; \mathbb{F}) \longrightarrow \mathbf{M}(m \times n; \mathbb{F}), \quad \big([a_{ij}]_{m\times n}, [b_{ij}]_{m\times n}\big) \longmapsto [a_{ij} + b_{ij}]_{m\times n}$$

In other words,

> **the sum of the matrices representing the linear transformations T and R, is the matrix representing $T \boxplus R$, is the sum of the linear transformations T and R.**

Notational Convention. We write $\underline{\mathbf{A}} + \underline{\mathbf{B}}$ for $\boxplus(\underline{\mathbf{A}}, \underline{\mathbf{B}})$.

Observation 9.23. We note that two matrices can be added if and only if they each have the same number of rows and each have the same number of columns.

This is because two linear transformations can be added if and only if they have a common domain and a common co-domain. The dimension of the domain is the number of columns of any matrix representing a linear transformation and the dimension of the co-domain is the number of rows.

Observation 9.24. Let $\mathcal{B}$ be a basis for V and $\mathcal{C}$ a basis for W, with $\dim V = n$ and $\dim W = m$.

It is immediate from Definition 9.22 that function

$$M_{\mathcal{B},\mathcal{C}} : \ \mathrm{Hom}(V, W) \longrightarrow \mathbf{M}(m \times n; \mathbb{F})$$

defined in Section 9.2, is additive.

Observation 9.25. Observation 9.20, Observation 9.24 and Theorem 9.17 show that for each choice of a basis $\mathcal{B}$ for V and a basis $\mathcal{C}$ for W, the function

$$M_{\mathcal{B},\mathcal{C}}\colon \operatorname{Hom}_{\mathbb{F}}(V,W) \longrightarrow \mathbf{M}(m\times n;\mathbb{F})$$

is, in fact, an isomorphism of vector spaces.

Example 9.26. We consider the linear transformations

$$T\colon \mathbb{R}^2 \longrightarrow \mathbb{R}^2, \quad (x,y)\longmapsto (x+2y, 3x-y)$$
$$R\colon \mathbb{R}^2 \longrightarrow \mathbb{R}^2, \quad (x,y)\longmapsto (2x+4y, -5x+7y)$$

Using the basis $\{(1,0),(0,1)\}$ for each of the vector spaces, the matrix of T is

$$\underline{\mathbf{A}} = \begin{bmatrix} 1 & 2 \\ 3 & -1 \end{bmatrix}$$

and that of R is

$$\underline{\mathbf{B}} = \begin{bmatrix} 2 & 4 \\ -5 & 7 \end{bmatrix}.$$

The sum of these linear transformations, $T+R$, is the linear transformation

$$T+R\colon \mathbb{R}^2 \longrightarrow \mathbb{R}^2, \quad (x,y)\longmapsto (x+2y,3x-y)+(2x+4y,-5x+7y) = (3x+6y,-2x+6y)$$

whose matrix is

$$\begin{bmatrix} 3 & 6 \\ -2 & 6 \end{bmatrix}$$

By Definition 9.22,

$$\begin{bmatrix} 1 & 2 \\ 3 & -1 \end{bmatrix} + \begin{bmatrix} 2 & 4 \\ -5 & 7 \end{bmatrix} = \begin{bmatrix} 1+2 & 2+4 \\ 3+(-5) & -1+7 \end{bmatrix}$$
$$= \begin{bmatrix} 3 & 6 \\ -2 & 6 \end{bmatrix}$$

9.3.3 The Matrix of a Composite Linear Transformation

The final operation on linear transformation investigated in Chapter 6 was composition. We introduce an algebraic operation on matrices to represent composition of linear transformations.

To compose two linear transformations, the domain of the second must be the co-domain of the first.

Consider finitely generated vector spaces, U, V, W, with $\dim U = p$, $\dim V = n$ and $\dim W = m$.

Take bases $\mathcal{B} = \{\mathbf{d}_1,\ldots,\mathbf{d}_p\}$ for U, $\mathcal{C} = \{\mathbf{e}_1,\ldots,\mathbf{e}_n\}$ for V and $\mathcal{D} = \{\mathbf{f}_1,\ldots,\mathbf{f}_m\}$ for W.

Take linear transformations $S\colon V \longrightarrow W$ and $T\colon U \longrightarrow V$.

Let the matrix of S with respect to $\mathcal{C}$ and $\mathcal{D}$ be

$$\underline{\mathbf{A}} = \Big[a_{ij}\Big]_{m\times n} \in \mathbf{M}(m\times n;\mathbb{F}),$$

the matrix of T with respect to $\mathcal{B}$ and $\mathcal{C}$ be

$$\underline{\mathbf{B}} = \Big[b_{jk}\Big]_{n\times p} \in \mathbf{M}(n\times p;\mathbb{F})$$

and the matrix of $S\circ T$ with respect to $\mathcal{B}$ and $\mathcal{D}$ be

$$\underline{\mathbf{C}} = \Big[c_{ik}\Big]_{m\times p} \in \mathbf{M}(m\times p;\mathbb{F})$$

By our earlier discussion this means that

$$T(\mathbf{d}_k) = \sum_{j=1}^{n} b_{jk}\mathbf{e}_j \qquad \text{for each } k \tag{9.7}$$

$$S(\mathbf{e}_j) = \sum_{i=1}^{m} a_{ij}\mathbf{f}_i \qquad \text{for each } j \tag{9.8}$$

$$(S\circ T)(\mathbf{d}_k) = \sum_{i=1}^{m} c_{ik}\mathbf{e}_j \qquad \text{for each } k \tag{9.9}$$

Then

$$\begin{aligned}
(S\circ T)(\mathbf{d}_k) &= S\big(T(\mathbf{d}_k)\big) && \text{by definition}\\
&= S\Big(\sum_{j=1}^{n} b_{jk}\mathbf{e}_j\Big) && \text{by Equation (9.7)}\\
&= \sum_{j=1}^{n} b_{jk}S(\mathbf{e}_j) && \text{as } S \text{ is a linear transformation}\\
&= \sum_{j=1}^{n} b_{jk}\Big(\sum_{i=1}^{m} a_{ij}\mathbf{f}_i\Big) && \text{by Equation (9.8)}\\
&= \sum_{j=1}^{n} \Big(\sum_{i=1}^{m} b_{jk}a_{ij}\mathbf{f}_i\Big)\\
&= \sum_{j=1}^{n} \Big(\sum_{i=1}^{m} a_{ij}b_{jk}\mathbf{f}_i\Big) && \text{as } \mathbb{F} \text{ is a field}\\
&= \sum_{i=1}^{m} \Big(\sum_{j=1}^{n} a_{ij}b_{jk}\mathbf{f}_i\Big) && \text{as the summations are independent}\\
&= \sum_{i=1}^{m} \Big(\sum_{j=1}^{n} a_{ij}b_{jk}\Big)\mathbf{f}_i && (9.10)
\end{aligned}$$

Since $\{\mathbf{f}_1,\dots\mathbf{f}_m\}$ is a basis for W, out follows from Equations (9.9) and (9.10), that for all i,k

$$c_{ik} = \sum_{j=1}^{n} a_{ij}b_{jk} \tag{9.11}$$

This motivates our definition of matrix multiplication.

Definition 9.27. For counting numbers m,n and p, *multiplication of an $m\times n$ matrix (on the right) by an $n\times p$ matrix over the field* $\mathbb{F}$ is the function

$$\boxdot : \mathbf{M}(m\times n;\mathbb{F})\times\mathbf{M}(n\times p;\mathbb{F}) \longrightarrow \mathbf{M}(m\times p;\mathbb{F}), \quad \Big(\Big[a_{ij}\Big]_{m\times n}, \Big[b_{jk}\Big]_{n\times p}\Big) \longmapsto \Big[\sum_{j=1}^{n} a_{ij}b_{jk}\Big]_{m\times p}$$

In other words,

> **the product of the matrices representing the linear transformations S and T, is the matrix representing $S \circ T$, the composition of the linear transformations S and T.**

Observation 9.28. We note that two matrices can be multiplied if and only if the number of columns of one (the one on the left) is the same as the number of rows of the other.

This is because two linear transformations can be composed if and only if the domain of the second one top be applied is the co-domain of the other.. The dimension of the domain is the number of columns of any matrix representing a linear transformation and the dimension of the co-domain is the number of rows.

Notational Convention. We write $\underline{\mathbf{A}}\,\underline{\mathbf{B}}$ for $\boxdot(\underline{\mathbf{A}}, \underline{\mathbf{B}})$.

Example 9.29. We consider the linear transformations

$$\begin{aligned} T\colon \mathbb{R}^2 &\longrightarrow \mathbb{R}^2, \quad (x, y) \longmapsto (x + 2y, 3x - y) \\ S\colon \mathbb{R}^2 &\longrightarrow \mathbb{R}^2, \quad (u, v) \longmapsto (2u + 4v, -5u + 7v) \end{aligned}$$

Using the basis $\{(1, 0), (0, 1)\}$ for each of the vector spaces, the matrix of T is

$$\underline{\mathbf{B}} = \begin{bmatrix} 1 & 2 \\ 3 & -1 \end{bmatrix}$$

and that of S is

$$\underline{\mathbf{A}} = \begin{bmatrix} 2 & 4 \\ -5 & 7 \end{bmatrix}$$

The composition of these linear transformations, $S \circ T$, is the linear transformation

$$\begin{aligned} S \circ T\colon \mathbb{R}^2 \longrightarrow \mathbb{R}^2, \quad (u, v) \longmapsto \;& S(x + 2y,\, 3x - y) \\ &= \Big(2(x + 2y) + 4(3x - y),\; -5(x + 2y) + 7(3x - y)\Big) \\ &= (14x,\, 11x - 17y) \end{aligned}$$

whose matrix is

$$\begin{bmatrix} 14 & 0 \\ 11 & -17 \end{bmatrix}$$

By Definition 9.27,

$$\begin{aligned} \begin{bmatrix} 2 & 4 \\ -5 & 7 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & -1 \end{bmatrix} &= \begin{bmatrix} 2 \times 1 + 4 \times 3 & 2 \times 2 + 4 \times (-1) \\ (-5) \times 1 + 7 \times 3 & (-5) \times 2 + 7 \times (-1) \end{bmatrix} \\ &= \begin{bmatrix} 14 & 0 \\ 11 & -17 \end{bmatrix} \end{aligned}$$

Since the matrix operations we have defined represents operations on linear transformations between finite dimensional vectors spaces, we can interpret properties of linear transformations as properties of matrices.

We illustrate this by defining left and right inverses for matrices. Recall that if $T\colon V \longrightarrow W$ is a linear transformation, then $S\colon W \longrightarrow V$ is right (resp. left) inverse to T if and only if $T \circ S = id_W$ (resp. $S \circ T = id_V$), and S is inverse to T if and only if it is both left and right inverse to T.

Since we choose a fixed basis for each vector space, the identity linear transformation on a p-dimensional vector space is represented by the $p \times p$ identity matrix $\underline{\mathbf{1}}_p$.

This motivates our next definition.

Definition 9.30. Given an $m \times n$ matrix, $\underline{\mathbf{A}}$, the $n \times m$ matrix, $\underline{\mathbf{B}}$, is

(i) *left inverse to* $\underline{\mathbf{A}}$ if and only if

$$\underline{\mathbf{B}}\,\underline{\mathbf{A}} = \underline{\mathbf{1}}_n$$

(ii) *right inverse to* $\underline{\mathbf{A}}$ if and only if

$$\underline{\mathbf{A}}\,\underline{\mathbf{B}} = \underline{\mathbf{1}}_m$$

(iii) *inverse to* $\underline{\mathbf{A}}$ if and only if

$$\underline{\mathbf{A}}\,\underline{\mathbf{B}} = \underline{\mathbf{1}}_m \qquad \text{and} \qquad \underline{\mathbf{A}}\,\underline{\mathbf{B}} = \underline{\mathbf{1}}_n$$

The matrix $\underline{\mathbf{A}}$ is *invertible* if and only if it has an inverse.

Observation 9.31. Since $\underline{\mathbf{A}}$ an $m \times n$ matrix, $\underline{\mathbf{B}}$ must be an $\ell \times m$ matrix for $\underline{\mathbf{B}}\,\underline{\mathbf{A}}$ to be defined.

Since $\underline{\mathbf{1}}_n$ is an $n \times n$ matrix, a necessary condition for the $\ell \times n$ matrix $\underline{\mathbf{B}}\,\underline{\mathbf{A}}$ to be $\underline{\mathbf{1}}_n$, is that $\ell = n$.

Lemma 9.32. *Let V and W be finitely generated vector spaces.*

The linear transformation $T\colon V \longrightarrow W$ is an isomorphism if and only if every matrix representing it is invertible.

Corollary 9.33. *If the $m \times n$ matrix, $\underline{\mathbf{A}}$, is invertible, then $m = n$.*

Proof. The result follows immediately from the Classification Theorem for Finitely Generated Vector Spaces. □

Proof. Exercise. □

9.3.4 Matrix Algebra

In Section 6.4, we saw that the set of all linear transformations between given vector spaces is again a vector space and investigated algebraic operations on this vector space.

The introduction of algebraic operations on matrices to represent the operations in Section 6.4, allows us to translate results about linear transformations into statements about matrices, as foreshadowed in Observation 6.53.

Theorem 9.34. *For $\alpha, \beta \in \mathbb{F}$, $\underline{\mathbf{A}}, \underline{\mathbf{B}}, \underline{\mathbf{C}} \in \mathbf{M}(m \times n; \mathbb{F}), \underline{\mathbf{D}}, \underline{\mathbf{E}} \in \mathbf{M}(n \times p; \mathbb{F})$ and $\underline{\mathbf{G}} \in \mathbf{M}(p \times q; \mathbb{F})$,*

(i) $(\underline{\mathbf{A}} + \underline{\mathbf{B}}) + \underline{\mathbf{C}} = \underline{\mathbf{A}} + (\underline{\mathbf{B}} + \underline{\mathbf{C}})$

(ii) $\underline{\mathbf{A}} + \underline{\mathbf{0}}_{m\times n} = \underline{\mathbf{A}} = \underline{\mathbf{0}}_{m\times n} + \underline{\mathbf{A}}$

(iii) $\underline{\mathbf{A}} + (-\underline{\mathbf{A}}) = \underline{\mathbf{0}}_{m\times n} = (-\underline{\mathbf{A}}) + \underline{\mathbf{A}}$

(iv) $\underline{\mathbf{A}} + \underline{\mathbf{B}} = \underline{\mathbf{B}} + \underline{\mathbf{A}}$

(v) $(\underline{\mathbf{A}}\,\underline{\mathbf{D}})\underline{\mathbf{G}} = \underline{\mathbf{A}}(\underline{\mathbf{D}}\,\underline{\mathbf{G}})$

(vi) $\underline{\mathbf{A}}\,\underline{\mathbf{1}}_n = \underline{\mathbf{A}} = \underline{\mathbf{1}}_m\underline{\mathbf{A}}$

(vii) $\underline{\mathbf{A}}(\underline{\mathbf{D}} + \underline{\mathbf{E}}) = \underline{\mathbf{A}}\,\underline{\mathbf{D}} + \underline{\mathbf{A}}\,\underline{\mathbf{E}}$

(vi) $(\underline{\mathbf{A}} + \underline{\mathbf{B}})\underline{\mathbf{D}} = \underline{\mathbf{A}}\,\underline{\mathbf{D}} + \underline{\mathbf{B}}\,\underline{\mathbf{D}}$

(viii) $1_{\mathbb{F}}\,\underline{\mathbf{A}} = \underline{\mathbf{A}}$

(ix) $\alpha(\underline{\mathbf{A}} + \underline{\mathbf{B}}) = \alpha\underline{\mathbf{A}} + \alpha\underline{\mathbf{B}}$

(x) $(\alpha + \beta)\underline{\mathbf{A}} = \alpha\underline{\mathbf{A}} + \beta\underline{\mathbf{A}}$

(xi) $(\alpha\beta)\underline{\mathbf{A}} = \alpha(\beta\underline{\mathbf{A}})$

(xii) $(\alpha\underline{\mathbf{A}})\underline{\mathbf{D}} = \alpha(\underline{\mathbf{A}}\,\underline{\mathbf{D}}) = \underline{\mathbf{A}}(\alpha\underline{\mathbf{D}})$

(xiii) $L_{\underline{\mathbf{A}}}\colon \mathbf{M}(n \times p; \mathbb{F}) \longrightarrow \mathbf{M}(m \times p; \mathbb{F}), \quad \underline{\mathbf{X}} \longmapsto \underline{\mathbf{A}}\,\underline{\mathbf{X}}$ *is a linear transformation.*

(xiv) $R_{\underline{\mathbf{D}}}\colon \mathbf{M}(m \times n; \mathbb{F}) \longrightarrow \mathbf{M}(m \times p; \mathbb{F}), \quad \underline{\mathbf{X}} \longmapsto \underline{\mathbf{X}}\,\underline{\mathbf{D}}$ *is a linear transformation.*

Proof. No further proof is required, because, by Definitions 9.18, 9.22 and 9.27, the statements simply reformulate, in the language of matrices, results proved in Chapter 6. □

Observation 9.35. Theorem 9.34 (i), (ii), (iii), (iv), (viii), (ix), (x) and (xi) show that $\mathbf{M}(m\times n; \mathbb{F})$ is a vector space over $\mathbb{F}$.

The following theorem now follows from immediately Observations 9.20 and 9.24.

Theorem 9.36. *If V is an n-dimensional vector space over $\mathbb{F}$ and W an m-dimensional one, then choosing a basis $\mathcal{B}$ for V and a basis $\mathcal{C}$ for W provides isomorphisms*

$$\begin{aligned} \beta_{\mathcal{B}} &: V \longrightarrow \mathbb{F}_{(n)} \\ \beta_{\mathcal{C}} &: W \longrightarrow \mathbb{F}_{(m)} \\ M_{\mathcal{B},\mathcal{C}} &: \operatorname{Hom}_{\mathbb{F}}(V, W) \longrightarrow \mathbf{M}(m \times n; \mathbb{F}) \end{aligned}$$

Observation 9.37. Since $\mathbb{F}_{(m)} = \mathbf{M}(m\times 1; \mathbb{F})$ and $\mathbb{F}^{(n)} = \mathbf{M}(1\times n; \mathbb{F})$, we obtain a "natural" vector space structure on $\mathbf{M}(m \times n; \mathbb{F})$ which simultaneously extends both the vector space structure on $\mathbb{F}_{(m)}$ and that on $\mathbb{F}^{(n)}$.

By inspection, these are precisely the vector space structures we introduced in an *ad hoc* manner in Exercise 8.6.

This allows us to classify all linear transformations between vector spaces of these forms.

Theorem 9.38. *(a) The function*

$$T\colon \mathbb{F}^{(n)} \longrightarrow \mathbb{F}^{(m)}, \quad \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix} \longmapsto \begin{bmatrix} y_1 & \cdots & y_m \end{bmatrix}$$

is a linear transformation if and only if there are $a_{ij} \in \mathbb{F}$ $(i = 1, \ldots, m,\ j = 1, \ldots, n)$ such that for all i

$$y_i = \sum_{j=1}^{n} a_{ij}x_j$$

(b) The function

$$T\colon \mathbb{F}_{(n)} \longrightarrow \mathbb{F}_{(m)}, \quad \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \longmapsto \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}$$

is a linear transformation if and only if there are $a_{ij} \in \mathbb{F}$ $(i = 1, \ldots, m,\ j = 1, \ldots, n)$ *such that for all* i

$$y_i = \sum_{j=1}^{n} a_{ij} x_j$$

Proof. A proof can be obtained by copying the proof of Theorem 5.9 on page 54, making only notational changes. □

Corollary 9.39. *(a) The function*

$$T\colon \mathbb{F}^{(n)} \longrightarrow \mathbb{F}^{(m)}$$

is a linear transformation if and only if there is an $n \times m$ *matrix,* $\underline{\mathbf{B}}$*, such that for each* $\underline{\mathbf{x}} \in \mathbb{F}_{(n)}$

$$T(\underline{\mathbf{x}}) = \underline{\mathbf{x}}\,\underline{\mathbf{B}}$$

(b) The function

$$T\colon \mathbb{F}_{(n)} \longrightarrow \mathbb{F}_{(m)}$$

is a linear transformation if and only if there is an $m \times n$ *matrix,* $\underline{\mathbf{A}}$*, such that for each* $\underline{\mathbf{x}} \in \mathbb{F}_{(n)}$

$$T(\underline{\mathbf{x}}) = \underline{\mathbf{A}}\,\underline{\mathbf{x}}$$

Proof. The corollary directly follows from Theorem 9.38 and the definition of matrix multiplication.

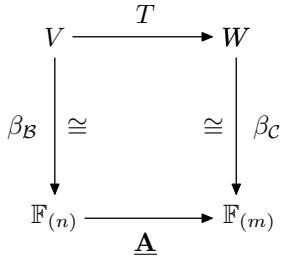
Note that the matrix $\underline{\mathbf{A}}$ is $\big[a_{ij}\big]_{m\times n}$ and the matrix $\underline{\mathbf{B}}$ is $\big[a_{ji}\big]_{n\times m}$, with a_{rs} as in Theorem 9.38. □

Observation 9.40. By Corollary 9.39, a linear transformation from $\mathbb{F}^{(n)}$ to $\mathbb{F}^{(p)}$ **is** multiplication (on the right) by and $n \times p$ matrix and a linear transformation $\mathbb{F}_{(n)}$ to $\mathbb{F}_{(m)}$ **is** multiplication (on the left) by an $m \times n$ matrix. Moreover, these matrices do not depend on choosing bases.

It is this feature which makes $\mathbb{F}_{(n)}$ our preferred "standard" vector space of dimension n over $\mathbb{F}$.

Every other vector space of dimension n over $\mathbb{F}$ is isomorphic with this, with an isomorphism being the same thing as choosing a basis. Choosing bases for V and W also assigns to each linear transformation, $T\colon V \longrightarrow W$ and $m \times n$ matrix $\underline{\mathbf{A}}$ compatible with the isomorphisms arising from the chosen bases.

In other words, we obtain a commutative diagram



We could have chosen $\mathbb{F}^{(n)}$ in lieu of $\mathbb{F}_{(n)}$, equally well. But, had we done so, the matrix representing T would need to be written on the right, and the matrix representing the composition of linear transformation would be the product of the matrices representing each linear transformation, but with the order reversed. We prefer to avoid this inconvenience.

9.4 Another Look at Matrix Multiplication

The vector space structure we have defined on $\mathbf{M}(m \times n; \mathbb{F})$, the set of all $m \times n$ matrices over $\mathbb{F}$, renders it isomorphic with the vector space of all linear transformations $V \longrightarrow W$, whenever $\dim_{\mathbb{F}}(V) = n$ and $\dim_{\mathbb{F}}(W) = m$.

We have also seen that this vector space structure simultaneously extends the vector space structure on $\mathbb{F}_{(m)}$ and that on $\mathbb{F}^{(n)}$. We examine this aspect more closely.

The key observation is an obvious one: an $m \times n$ matrix can be regarded as comprising m rows, or as n columns.

Definition 9.41. Given the matrix $\underline{\mathbf{A}} = \left[a_{ij}\right]_{m\times n} \in \mathbf{M}(m \times n; \mathbb{F})$, its i^{th} *row* is

$$\mathbf{r}_i^{\mathbf{A}} := \begin{bmatrix} a_{i1} & \cdots & a_{in} \end{bmatrix} \in \mathbb{F}^{(n)}$$

and its j^{th} *column* is

$$\mathbf{c}_j^{\mathbf{A}} := \begin{bmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{bmatrix} \in \mathbb{F}_{(m)}$$

The next theorem, which assists working with matrices, is a direct consequence of the vector space structures on $\mathbf{M}(m \times n; \mathbb{F}), \mathbb{F}^{(n)}$ and $\mathbb{F}_{(m)}$, together with the construction of the direct sum of vector spaces is the availability of two isomorphisms.

Theorem 9.42. *The functions*

$$\mathbf{M}(m \times n; \mathbb{F}) \longrightarrow \bigoplus_{i=1}^{m} \mathbb{F}^{(n)}, \qquad \underline{\mathbf{A}} \longmapsto (\mathbf{r}_1^{\mathbf{A}}, \ldots, \mathbf{r}_m^{\mathbf{A}})$$

$$\mathbf{M}(m \times n; \mathbb{F}) \longrightarrow \bigoplus_{j=1}^{n} \mathbb{F}_{(m)}, \qquad \underline{\mathbf{A}} \longmapsto (\mathbf{c}_1^{\mathbf{A}}, \ldots, \mathbf{c}_n^{\mathbf{A}})$$

are both isomorphisms.

Theorem 9.42 allows us to regard the rows of the $m \times n$ matrix $\underline{\mathbf{A}}$ as m vectors in $\mathbb{F}^{(n)}$ and its columns as n vectors in $\mathbb{F}_{(m)}$.

Definition 9.43. The *column space of* $\underline{\mathbf{A}} \in \mathbf{M}(m \times n; \mathbb{F})$ is the vector subspace of $\mathbb{F}_{(m)}$ generated by

$$\{\mathbf{c}_j^{\mathbf{A}} \mid 1 \leq j \leq n\}$$

and its *row space* is the vector subspace of $\mathbb{F}^{(m)}$ generated by

$$\{\mathbf{r}_i^{\mathbf{A}} \mid 1 \leq i \leq m\}$$

We can now investigate the relationship between the rows (resp. columns) of the product of two matrices and the rows (resp. columns) of the matrices being multiplied.

Take $\underline{\mathbf{A}} = [a_{ij}]_{m\times n}, \underline{\mathbf{B}} = [b_{jk}]_{n\times p}$ and $\underline{\mathbf{C}} = [c_{ik}]_{m\times p}$ and suppose that $\underline{\mathbf{C}} = \underline{\mathbf{A}}\,\underline{\mathbf{B}}$, that is

$$c_{ik} = \sum_{j=1}^{n} a_{ij}b_{jk} \qquad (1 \le i \le m,\ 1 \le k \le p)$$

Observation 9.44. An observation, which is useful in many applications, follows directly from our definitions:

$$c_{ij} = \mathbf{r}_i^{\underline{\mathbf{A}}}\,\mathbf{c}_j^{\underline{\mathbf{B}}}$$

The k^{th} column of $\underline{\mathbf{C}}$ is obtained by fixing k, in which case we obtain

$$\begin{aligned}
\begin{bmatrix} c_{1k} \\ \vdots \\ c_{mk} \end{bmatrix} &= \begin{bmatrix} a_{11}b_{1k} + \cdots + a_{1n}b_{nk} \\ \vdots \\ a_{m1}b_{1k} + \cdots + a_{mn}b_{nk} \end{bmatrix} \\
&= \begin{bmatrix} a_{11}b_{1k} \\ \vdots \\ a_{m1}b_{1k} \end{bmatrix} + \cdots + \begin{bmatrix} a_{1n}b_{nk} \\ \vdots \\ a_{mn}b_{nk} \end{bmatrix} \\
&= \begin{bmatrix} a_{11} \\ \vdots \\ a_{m1} \end{bmatrix} b_{1k} + \cdots + \begin{bmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{bmatrix} b_{nk}
\end{aligned}$$

Thus, for $\underline{\mathbf{C}} = \underline{\mathbf{A}}\,\underline{\mathbf{B}}$

$$\mathbf{c}_k^{\underline{\mathbf{C}}} = \sum_{j=1}^{n} \mathbf{c}_j^{\underline{\mathbf{A}}} b_{jk}$$

In other words:

> **The k^{th} column of $\underline{\mathbf{A}}\,\underline{\mathbf{B}}$ is the linear combination of the columns of $\underline{\mathbf{A}}$ given by the entries in the k^{th} column of $\underline{\mathbf{B}}$.**

We can apply this to solving systems of linear equations. We represent he system of m equations in n unknowns

$$\begin{array}{ccccccc}
a_{11}x_1 & + & \cdots & + & a_{1n}x_n & = & b_1 \\
\vdots & & & & \vdots & & \vdots \\
a_{m1}x_1 & + & \cdots & + & a_{mn}x_n & = & b_m
\end{array} \qquad (\ddagger)$$

by the matrix equation

$$\underline{\mathbf{A}}\mathbf{x} = \mathbf{b},$$

where

$$\underline{\mathbf{A}} = [a_{ij}]_{m\times n}, \qquad \mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \qquad \text{and} \qquad \mathbf{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$$

Theorem 9.45. *The system of equations,* (‡)*, has a solution if and only if* $\mathbf{b}$ *is in the column space of* $\underline{\mathbf{A}}$*.*

Now fix i instead of k.

$$\begin{aligned}\begin{bmatrix} c_{i1} & \cdots & c_{ip} \end{bmatrix} &= \begin{bmatrix} a_{i1}b_{11} + \cdots + a_{in}b_{n1} & \cdots & a_{i1}b_{1p} + \cdots + a_{in}b_{np} \end{bmatrix} \\ &= \begin{bmatrix} a_{i1}b_{11} & \cdots & a_{i1}b_{1p} \end{bmatrix} + \cdots + \begin{bmatrix} a_{in}b_{n1} & \cdots & a_{in}b_{np} \end{bmatrix} \\ &= a_{i1}\begin{bmatrix} b_{11} & \cdots & b_{1p} \end{bmatrix} + \cdots + a_{in}\begin{bmatrix} b_{n1} & \cdots & b_{np} \end{bmatrix}\end{aligned}$$

Thus, for $\underline{\mathbf{C}} = \underline{\mathbf{A}}\,\underline{\mathbf{B}}$

$$\mathbf{r}_i^{\mathbf{C}} = \sum_{i=1}^{m} a_{ij}\mathbf{r}_j^{\mathbf{B}}$$

In other words:

> **The i^{th} row of $\underline{\mathbf{A}}\,\underline{\mathbf{B}}$ is the linear combination of the rows of $\underline{\mathbf{B}}$ given by the entries in the i^{th} row of $\underline{\mathbf{A}}$.**

Observation 9.46. An important consequence is that such operations on the rows (resp. columns) of a matrix as the elementary row (resp. column) operations can be performed by multiplying the given matrix on the left (resp. right) by a suitable matrix.

In particular, Gaußian elimination for solving systems of simultaneous linear equations by reducing a matrix to *(reduced) row echelon form* can be achieved using matrix multiplication. We shall return to this later.

9.5 Matrices Representing the Same Linear Transformation

In order to represent a linear transformation between finite dimensional vector spaces by a matrix, we need to choose a basis for each of the vector spaces in question, and the resulting matrix depends not only on the linear transformation itself, but also on the particular bases chosen. It is, therefore, natural to ask:

> **What is the relationship between two matrices representing a linear transformation between finitely generated vector spaces?**

This section is devoted to answering this question.

We know that each finitely generated vector space is determined up to isomorphism by a single (numerical) invariant, its dimension, which is the number of vectors in any basis for it.

Moreover, we have also seen that if $\dim_{\mathbb{F}} V = n$ and $\dim_{\mathbb{F}} W = m$, then any matrix representing the linear transformation $T\colon V \longrightarrow W$ must be an $m \times n$ matrix over $\mathbb{F}$. One immediate, necessary condition for two matrices to represent one and the same linear transformation between two finite dimensional vector spaces is that they both be "of the same size".

The following example shows that this necessary condition is not sufficient.

Example 9.47. No linear transformation can be represented by both $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$.

One way to see why these matrices cannot represent the same linear transformation is to observe that any linear transformation represented by the former must be bijective, whereas no linear transformation represented by the latter can be either injective or surjective. (We leave it to the reader to contemplate why these statements are true. The reasons will become evident later.)

To determine when two matrices represent the same linear, consider a linear transformation $T\colon V \longrightarrow W$ from the n-dimensional vector space V to the m-dimensional vector space W.

Choose bases $\mathcal{B} = \{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ and $\mathcal{B}' = \{\mathbf{e}'_1, \ldots, \mathbf{e}'_n\}$ for V as well as bases $\mathcal{C} = \{\mathbf{f}_1, \ldots, \mathbf{f}_m\}$ and $\mathcal{C}' = \{\mathbf{f}'_1, \ldots, \mathbf{f}'_m\}$ for W.

Let $\underline{\mathbf{A}} = [a_{ij}]_{m\times n}$ be the matrix of T with respect to the bases $\mathcal{B}$ for V and $\mathcal{C}$ for W, and $\underline{\mathbf{A}}' = [a'_{ij}]_{m\times n}$ the matrix with respect to the bases $\mathcal{B}'$ for V and $\mathcal{C}'$ for W, so that

$$T(\mathbf{e}_j) = \sum_{i=1}^{m} a_{ij}\mathbf{f}_i \qquad \text{and} \qquad T(\mathbf{e}'_j) = \sum_{i=1}^{m} a_{ij}\mathbf{f}'_i$$

Finally, we express the basis vectors in $\mathcal{B}'$, (resp. $\mathcal{C}'$) as linear combinations of the basis vectors in $\mathcal{B}$, (resp. $\mathcal{C}$),

$$\mathbf{e}'_\ell = \sum_{j=1}^{n} \lambda_{j\ell}\mathbf{e}_j \qquad \text{and} \qquad \mathbf{f}'_k = \sum_{i=1}^{m} \mu_{ik}\mathbf{f}_i$$

In other words, the co-ordinate vector of $\mathbf{e}'_\ell$ with respect to the basis $\mathcal{B}$ and that of $\mathbf{f}'_k$ with respect to $\mathcal{C}$ are, respectively,

$$\begin{bmatrix} \lambda_{1\ell} \\ \vdots \\ \lambda_{n\ell} \end{bmatrix} \qquad \text{and} \qquad \begin{bmatrix} \mu_{1k} \\ \vdots \\ \mu_{mk} \end{bmatrix}$$

We form the $n \times n$ matrix $\underline{\mathbf{L}} := [\lambda_{ij}]_{n\times n}$ and the $m \times m$ matrix $\underline{\mathbf{M}} := [\mu_{ij}]_{m\times m}$.

Take $\mathbf{v} \in V$ and let $T(\mathbf{v}) = \mathbf{w} \in W$. Then

$$\begin{aligned} \mathbf{v} &= \sum_{\ell=1}^{n} x'_\ell \mathbf{e}'_\ell \\ &= \sum_{\ell=1}^{n} x'_\ell \left(\sum_{j=1}^{n} \lambda_{j\ell}\mathbf{e}_j \right) \\ &= \sum_{j=1}^{n} \left(\sum_{\ell=1}^{n} \lambda_{j\ell} x'_\ell \right) \mathbf{e}_j \end{aligned}$$

But $\mathbf{v} = \sum_{j=1}^{n} x_j e_j$, with the x_j's unique. So $x_j = \sum_{\ell=1}^{n} \lambda_{j\ell} x'_\ell$, or

$$\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} \lambda_{11} & \cdots & \lambda_{1n} \\ \vdots & & \vdots \\ \lambda_{n1} & \cdots & \lambda_{nn} \end{bmatrix} \begin{bmatrix} x'_1 \\ \vdots \\ x'_n \end{bmatrix}.$$

Similarly, if $\mathbf{w} \in W$, $\mathbf{w} = \sum_{k=1}^{m} y'_k \mathbf{e}'_k = \sum_{k=1}^{m} y_k \mathbf{e}_k$, with

$$\begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} \mu_{11} & \dots & \mu_{1m} \\ \vdots & & \vdots \\ \mu_{m1} & \dots & \mu_{mm} \end{bmatrix} \begin{bmatrix} y'_1 \\ \vdots \\ y'_m \end{bmatrix}$$

Consequently,

$$\begin{aligned} \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} &= \begin{bmatrix} \mu_{11} & \dots & \mu_{1m} \\ \vdots & & \vdots \\ \mu_{m1} & \dots & \mu_{mm} \end{bmatrix} \begin{bmatrix} y'_1 \\ \vdots \\ y'_m \end{bmatrix} \\ &= \begin{bmatrix} \mu_{11} & \dots & \mu_{1m} \\ \vdots & & \vdots \\ \mu_{m1} & \dots & \mu_{mm} \end{bmatrix} \begin{bmatrix} a'_{11} & \dots & a'_{1n} \\ \vdots & & \vdots \\ a'_{m1} & \dots & a'_{mn} \end{bmatrix} \begin{bmatrix} x'_1 \\ \vdots \\ x'_n \end{bmatrix} \end{aligned}$$

On the other hand,

$$\begin{aligned} \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} &= \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \\ &= \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix} \begin{bmatrix} \lambda_{11} & \dots & \lambda_{1n} \\ \vdots & & \vdots \\ \lambda_{n1} & \dots & \lambda_{nn} \end{bmatrix} \begin{bmatrix} x'_1 \\ \vdots \\ x'_n \end{bmatrix} \end{aligned}$$

In other words, both $\underline{\mathbf{A}}\,\underline{\mathbf{L}}$ and $\underline{\mathbf{M}}\,\underline{\mathbf{A}}'$ are the matrix of T with respect to the basis $\mathcal{B}'$ for V and $\mathcal{C}$ for W.

By the uniqueness of the matrix of a linear transformation with respect to given bases,

$$\underline{\mathbf{A}}\,\underline{\mathbf{L}} = \underline{\mathbf{M}}\,\underline{\mathbf{A}}'$$

This leads us to the next definition and theorem, summarising the above.

Definition 9.48. The matrix

$$\underline{\mathbf{L}} := \begin{bmatrix} \lambda_{11} & \dots & \lambda_{1n} \\ \vdots & & \vdots \\ \lambda_{n1} & \dots & \lambda_{nn} \end{bmatrix}$$

where $\mathbf{e}'_\ell = \sum_{j=1}^{n} \lambda_{j\ell} \mathbf{e}_j$, is the *change of basis matrix (from the basis* $\{\mathbf{e}'_i\}$ *to the basis* $\{\mathbf{e}_i\}$*).*

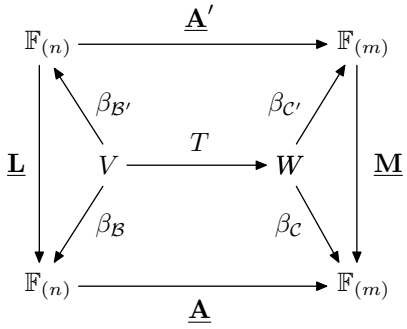
Of course, $\underline{\mathbf{M}} := \begin{bmatrix} \mu_{11} & \dots & \mu_{1m} \\ \vdots & & \vdots \\ \mu_{m1} & \dots & \mu_{mm} \end{bmatrix}$ is also a change of basis matrix.

We now summarise the above in convenient form.

Theorem 9.49. *If* $\underline{\mathbf{A}}$ *and* $\underline{\mathbf{A}}'$ *are the matrices of the linear transformation* $T\colon V \longrightarrow W$ *and if* $\underline{\mathbf{L}}$ *and* $\underline{\mathbf{M}}$ *are the change of basis matrices from the bases which give rise to* $\underline{\mathbf{A}}'$ *to the bases which give rise to* $\underline{\mathbf{A}}$*, then*

$$\underline{\mathbf{M}}\,\underline{\mathbf{A}}' = \underline{\mathbf{A}}\,\underline{\mathbf{L}}$$

This is expressed by the commutative diagram



Corollary 9.50. *If* $\underline{\mathbf{A}}$ *and* $\underline{\mathbf{B}}$ *are the matrices of the linear transformation* $T\colon V \longrightarrow V$ *and if* $\underline{\mathbf{M}}$ *is the change of basis matrix from the basis which gives rise to* $\underline{\mathbf{B}}$ *to the basis which give rise to* $\underline{\mathbf{A}}$*, then, for any counting number* n*,*

$$\underline{\mathbf{B}}^n = \underline{\mathbf{M}}^{-1}\underline{\mathbf{A}}^n\,\underline{\mathbf{M}}.$$

Proof. By Theorem 9.49, $\underline{\mathbf{M}}\,\underline{\mathbf{B}} = \underline{\mathbf{A}}\,\underline{\mathbf{M}}$, or, equivalantly, $\underline{\mathbf{B}} = \underline{\mathbf{M}}^{-1}\underline{\mathbf{A}}\,\underline{\mathbf{M}}$. Thus,

$$\begin{aligned}
\underline{\mathbf{B}}^n &= (\underline{\mathbf{M}}^{-1}\underline{\mathbf{A}}\,\underline{\mathbf{M}})^n \\
&= (\underline{\mathbf{M}}^{-1}\underline{\mathbf{A}}\,\underline{\mathbf{M}})(\underline{\mathbf{M}}^{-1}\underline{\mathbf{A}}\,\underline{\mathbf{M}})\cdots(\underline{\mathbf{M}}^{-1}\underline{\mathbf{A}}\,\underline{\mathbf{M}}) \\
&= \underline{\mathbf{M}}^{-1}\underline{\mathbf{A}}\,(\underline{\mathbf{M}}\,\underline{\mathbf{M}}^{-1})\underline{\mathbf{A}}\,(\underline{\mathbf{M}}\,\underline{\mathbf{M}}^{-1})\cdots(\underline{\mathbf{M}}\,\underline{\mathbf{M}}^{-1})\underline{\mathbf{A}}\,\underline{\mathbf{M}} && \text{by associativity} \\
&= \underline{\mathbf{M}}^{-1}\underline{\mathbf{A}}^n\underline{\mathbf{M}} && \text{since } \underline{\mathbf{M}}\,\underline{\mathbf{M}}^{-1} = \underline{\mathbf{1}}_n.
\end{aligned}$$

□

Observation 9.51. Examples 2.4, 2.5, 2.6 and 2.8 are applications of this corollary.

Example 9.52. Let $\begin{bmatrix} 4 & -3 \\ 1 & 0 \end{bmatrix}$ to be the matrix of a linear transformation $T\colon V \longrightarrow V$ with respect to the basis $\{\mathbf{e}_1, \mathbf{e}_2\}$ of the real vector space V.

Note that $\dim_{\mathbb{R}}(V) = 2$.

Put

$$\begin{aligned}
\mathbf{e}_1' &:= 3\mathbf{e}_1 + \mathbf{e}_2 \\
\mathbf{e}_2' &:= \mathbf{e}_1 + \mathbf{e}_2
\end{aligned}$$

Direct calculation shows that

$$\begin{aligned}
\mathbf{e}_1 &= \frac{1}{2}\,(\mathbf{e}_1' - \mathbf{e}_2') \\
\mathbf{e}_2 &= \frac{1}{2}\,(-\mathbf{e}_1' + 3\mathbf{e}_2')
\end{aligned}$$

Since each $\mathbf{e}_i$ is a linear combination of the $\mathbf{e}'_j$s, it follows from Theorem 7.4,

$$V = \langle \mathbf{e}_1, \mathbf{e}_2 \rangle \subseteq \langle \mathbf{e}'_1, \mathbf{e}'_2 \rangle \subseteq V$$

This shows that the two vectors $\mathbf{e}'_1$ and $\mathbf{e}'_2$ generate V.
Since $\dim V = 2$, it follows from Theorem 8.11, that $\{\mathbf{e}'_1, \mathbf{e}'_2\}$ is also a basis for V.
Let $\mathbf{v} \in V$ have co-ordinate vector $\begin{bmatrix} r \\ s \end{bmatrix}$ with respect to $\{\mathbf{e}_1, \mathbf{e}_2\}$ and $\begin{bmatrix} x \\ y \end{bmatrix}$ with respect to $\{\mathbf{e}'_1, \mathbf{e}'_2\}$.
As $x\mathbf{e}'_1 + y\mathbf{e}'_2 = (3x + y)\mathbf{e}_1 + (x + y)\mathbf{e}_2$,

$$\begin{bmatrix} r \\ s \end{bmatrix} = \begin{bmatrix} 3 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix},$$

and since $r\mathbf{e}_1 + s\mathbf{e}_2 = \frac{1}{2}\Big((r - s)\mathbf{e}'_1 + (-r + 3s)\mathbf{e}'_2\Big)$,

$$\begin{bmatrix} x \\ y \end{bmatrix} = \tfrac{1}{2} \begin{bmatrix} 1 & -1 \\ -1 & 3 \end{bmatrix} \begin{bmatrix} r \\ s \end{bmatrix},$$

It follows that the matrix of T with respect to the basis $\{\mathbf{e}'_1, \mathbf{e}'_2\}$ is

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

Observation 9.53. The reader should compare Example 9.52 with Example 2.4.

Observation 9.54. Every $n \times n$ matrix with coefficients in $\mathbb{F}$ is the matrix of a linear transformation $T : V \longrightarrow V$. It is natural to ask:

Which linear transformation is represented by the change of basis matrix $\underline{\mathbf{L}}$?

To answer this, note that when we change the basis, we do *not* change V: Each vector $\mathbf{v} \in V$ is left unchanged, only the co-ordinate vector we assign to it changes. In other words, $\underline{\mathbf{L}}$ is the matrix of the identity linear transformation $id_V : V \longrightarrow V,\ \mathbf{v} \longmapsto \mathbf{v}$.

9.6 Exercises

Exercise 9.1. Prove that if the linear transformation $T : V \longrightarrow W$ has matrix $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ with respect to some bases, then it is neither injective nor surjective.

Exercise 9.2. Prove that if linear transformation $T : V \longrightarrow W$ has matrix $\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \end{bmatrix}$ with respect to some bases, then it has a right inverse, but no left inverse.

Exercise 9.3. Let $T : \mathbb{R}^3 \longrightarrow \mathbb{R}^3$ be rotation about the y-axis through an angle of θ. Show that T is a linear transformation of $\mathbb{R}^3$ to itself and find a matrix representation of T with respect to the bases

(a) $\{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}$ for both the domain and co-domain;

(b) $\{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}$ for the domain and $\{(1, 0, 0), (1, 1, 0), (1, 1, 1)\}$ for the co-domain;

(c) $\{(1,0,0),(1,1,0),(1,1,1)\}$ for both the domain and co-domain;

(d) $\{(1,0,1),(0,1,0),(-1,0,1)\}$ for both the domain and co-domain.

Exercise 9.4. Show that the linear transformation $T : V \longrightarrow W$ is an isomorphism if and only if every matrix which represents it is invertible.

Exercise 9.5. Use the definition of the multiplication of a linear transformation by a scalar to define a multiplication of matrices by scalars.

Exercise 9.6. Prove that if the $m \times n$ matrix $\underline{\mathbf{A}}$ is invertible, then $m = n$ and its inverse is uniquely determined.

Exercise 9.7. Let $\{\mathbf{e}_1, \mathbf{e}_2\}$ be a basis for the vector space V and $T\colon V \to V$ a linear transformation.

Show that in each of the following cases $\{\mathbf{e}'_1, \mathbf{e}'_2\}$ is also a basis for V and find the matrix $\underline{\mathbf{A}}'$ of T with respect to the basis $\{\mathbf{e}'_1, \mathbf{e}'_2\}$ given that its matrix with respect to $\{\mathbf{e}_1, \mathbf{e}_2\}$ is $\underline{\mathbf{A}}$.

(a) $\mathbf{e}'_1 := 2\mathbf{e}_1 + \mathbf{e}_2$, $\mathbf{e}'_2 := \mathbf{e}_1$ and $\underline{\mathbf{A}} := \begin{bmatrix} 4 & -4 \\ 1 & 0 \end{bmatrix}$

(b) $\mathbf{e}'_1 := (\sqrt{2}+1)\mathbf{e}_1 - \mathbf{e}_2$, $\mathbf{e}'_2 := \mathbf{e}_1 + (\sqrt{2}+1)\mathbf{e}_2$ and $\underline{\mathbf{A}} := \begin{bmatrix} 1 & 2 \\ 2 & 5 \end{bmatrix}$

Exercise 9.8. Prove Theorem 9.34 by direct calculation, using only Definitions 9.9, 9.18, 9.22 and 9.27.

The traditional mathematician recognizes and appreciates mathematical elegance when he sees it. I propose to go one step further, and to consider elegance an essential ingredient of mathematics: if it is clumsy, it is not mathematics.

Edsger Dijkstra

Chapter 10

Rank and Nullity

Associated with each linear transformation $T\colon V \longrightarrow W$ are two vector subspaces:

(a) the kernel of T, $\ker T$, which is a vector subspace of the domain of T, V;

(b) the image of T, $\operatorname{im} T$, which is a vector subspace of the co-doain of T, W.

These subspaces contain crucial information about T and we investigate them and their relationship to each other.

Definition 10.1. The *rank* of T, $\operatorname{rk}(T)$, is the dimension of $\operatorname{im}(T)$ and the *nullity* of T, $\operatorname{n}(T)$, is the dimension of $\ker(T)$:

$$\operatorname{rk} T := \dim_{\mathbb{F}} (\operatorname{im} T)$$
$$\operatorname{n}(T) := \dim_{\mathbb{F}} (\ker T)$$

When the domain of the linear transformation $T\colon V \longrightarrow W$ is finitely generated, the rank and the nullity of determine each other.

Theorem 10.2. *Let $T\colon V \to W$ be a linear transformation.*

If V is finitely generated, then $\operatorname{rk}(T) + n(T) = \dim V$.

Proof. Let $\{\mathbf{e}_1, \ldots, \mathbf{e}_p\}$ be a basis for $\ker(T)$.

Extend this to a basis $\{\mathbf{e}_1, \ldots, \mathbf{e}_{p+q}\}$ of V.

We show that $\{T(\mathbf{e}_{p+1}), \ldots, T(\mathbf{e}_{p+q})\}$ is a basis for $\operatorname{im}(T)$.

Take $\mathbf{w} \in \operatorname{im}(T)$.

Then $\mathbf{w} = T(\mathbf{v})$ for some $\mathbf{v} \in V$.

Since $\{\mathbf{e}_1, \ldots, \mathbf{e}_{p+q}\}$ is a basis for V, there are $\lambda_1, \ldots, \lambda_{p+q}$ with $\mathbf{v} = \sum_{j=1}^{p+q} \lambda_j \mathbf{e}_j$. Thus,

$$\begin{aligned} \mathbf{w} = T(\mathbf{v}) &= T\left(\sum_{j=1}^{p+q} \lambda_j \mathbf{e}_j\right) \\ &= \sum_{j=1}^{p+q} \lambda_j T(\mathbf{e}_j) \end{aligned}$$

$$= \sum_{j=p+1}^{p+q} \lambda_j T(\mathbf{e}_j) \qquad \text{since } T(\mathbf{e}_j) = \mathbf{0}_W \text{ for all } j \leq p,$$

showing that $\{T(\mathbf{e}_{p+1}), \ldots, T(\mathbf{e}_{p+q})\}$ generates $\mathrm{im}(T)$.
If $\lambda_{p+1}T(\mathbf{e}_{p+1}) + \cdots + \lambda_{p+q}T(\mathbf{e}_{p+q}) = \mathbf{0}_W$, then

$$\lambda_{p+1}\mathbf{e}_{p+1} + \cdots + \lambda_{p+q}\mathbf{e}_{p+q} \in \ker(T) = \langle \mathbf{e}_1, \ldots, \mathbf{e}_p \rangle$$

Thus $\lambda_{p+1}\mathbf{e}_{p+1} + \cdots + \lambda_{p+q}\mathbf{e}_{p+q} = \mu_1\mathbf{e}_1 + \cdots + \mu_p\mathbf{e}_p$. whence,

$$\sum_{j=1}^{p+q} \lambda_j \mathbf{e}_j = \mathbf{0}_V,$$

where $\lambda_j = -\mu_j$ for $j > p$.
As $\{\mathbf{e}_1, \ldots, \mathbf{e}_{p+q}\}$ is a basis for V, $\lambda_j = 0$ for all j. In particular, $\lambda_{p+i} = 0$ for $i = 1, \ldots, q$.
Thus $T(\mathbf{e}_{p+1}), \ldots, T(\mathbf{e}_{p+q})$ are linearly independent. □

Corollary 10.3. *Let $T\colon V \longrightarrow W$ be a linear transformation. Then*

$$V \cong \ker(T) \oplus \mathrm{im}(T).$$

Proof. Exercise. □

Lemma 10.4. *Let $T\colon V \longrightarrow W$ be a linear transformation .*

(i) T is injective if and only if $n(T) = 0$.

(ii) If W is finitely generated, then T is surjective if and only if $\mathrm{rk}(T) = \dim_{\mathbb{F}} W$.

Proof. Let $T\colon V \longrightarrow W$ be a linear transformation.
(i) $\mathrm{n}(T) = 0$ if and only if $\ker(T) = \{\mathbf{0}_V\}$ if and only if T is injective.

(ii) $\Rightarrow$: Immediate from the definition.

$\Leftarrow$: Suppose that $\mathrm{rk}(T) := \dim_{\mathbb{F}}(\mathrm{im}(T)) = \dim_{\mathbb{F}}(W) = n$.
Let $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ be a basis for $\mathrm{im}(T)$.
Then $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ is a set of n linearly independent vectors in the n dimensional vector space W.
By Theorem 8.11 on page 90, $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ is a basis for W, showing that $\mathrm{im}(T) = W$. □

The hypothesis that W be finitely generated cannot be dispensed with in Lemma 10.4(ii), as our next example shows.

Example 10.5. Consider $T\colon \mathbb{R}[t] \longrightarrow \mathbb{R}[t]$ given, heuristically, by

$$T(p) := \int_0^t p(x)dx,$$

so that

$$T(a_0 + a_1t + \cdots + a_nt^4) = a_0t + \tfrac{1}{2}a_1t^2 + \cdots + \tfrac{1}{n+1}a_nt^{n+1}$$

By the definition of a polynomial, $\{1, t, t^2, \ldots\}$ is a basis for $\mathbb{R}[t]$.

Clearly, $\{t, t^2, \ldots\}$ is a basis for $\mathrm{im}(T)$.

Thus $\mathrm{im}(T) \neq \mathbb{R}[t]$, as $1 \notin \mathrm{im}(T)$; in other words, T is not surjective.

Since the function

$$\{1, t, t^2, \ldots\} \longrightarrow \{t, t^2, \ldots\}, \qquad t^j \longmapsto t^{j+1}$$

is a bijection between our basis for $\mathbb{R}[t]$ and a basis for $\mathrm{im}(T)$, we have $\dim_{\mathbb{F}}(W) = \mathrm{rk}\,(T)$.

10.1 Rank and Nullity for Matrices

When we restrict attention to finitely generated vector spaces, we may choose a finite basis for each vector space. As we saw in the previous chapter, choosing bases allowed us to introduce matrices for calculating with linear transformations, with algebraic operations on matrices defined to reflect operations on linear transformations.

This allows us to introduce the notions of rank and nullity for matrices, but first, we review the relevant parts of last chapter.

Choosing a basis, $\mathcal{B} = \{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$, for the vector space, V, and a basis $\mathcal{C} = \{\mathbf{f}_1, \ldots, \mathbf{f}_m\}$ for the vector space, W, is the choice of isomorphisms

$$V \longrightarrow \mathbb{F}_{(n)}, \qquad \mathbf{v} \longmapsto \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \qquad \text{for } \mathbf{v} = \sum_{j=1}^{n} x_j \mathbf{e}_j$$

$$W \longrightarrow \mathbb{F}_{(m)}, \qquad \mathbf{w} \longmapsto \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} \qquad \text{for } \mathbf{w} = \sum_{i=1}^{m} y_i \mathbf{f}_i$$

and an isomorphism

$$\mathrm{Hom}(V, W) \longrightarrow \mathbf{M}(m \times n; \mathbb{F}), \quad T \longmapsto \underline{\mathbf{A}} = [a_{ij}]_{m \times n}$$

where, for each j,

$$T(\mathbf{e}_j) = \sum_{i=1}^{m} a_{ij} \mathbf{f}_i$$

The algebraic operations on matrices were so defined that the coordinate vector of the image of a vector under the linear transformation is obtained by multiplying (on the left), the coordinate vector of the original vector by the matrix of the linear transformation.

In other words, when $\underline{\mathbf{x}}$ is the coordinate vector of $\mathbf{v} \in V$ with respect to $\mathcal{B}$, $\underline{\mathbf{y}}$ the coordinate vector of $\mathbf{w} \in W$ with respect to $\mathcal{C}$, and $\underline{\mathbf{A}}$ the matrix of T with respect to $\mathcal{B}$ and $\mathcal{C}$,

$$T(\mathbf{v}) = \mathbf{w} \quad \text{ if and only if } \quad \underline{\mathbf{y}} = \underline{\mathbf{A}}\,\underline{\mathbf{x}}$$

We also showed in Theorem 9.42 on page 117, that the $m \times n$ matrix, $\underline{\mathbf{A}}$, may be regard a comprising n "columns"

$$\underline{\mathbf{A}} = \begin{bmatrix} \mathbf{c}_1^{\mathbf{A}} & \cdots & \mathbf{c}_n^{\mathbf{A}} \end{bmatrix}$$

The above allows us to compute the matrix $\underline{\mathbf{A}}$:

$\mathbf{c}_j^{\mathbf{A}}$, the j^{th} column of $\underline{\mathbf{A}}$, is the coordinate vector of $T(\mathbf{e}_j)$ with respect to $\mathcal{C}$.

Further, Theorem 9.42 on page 117 allows us to regard the rows of the $m \times n$ matrix $\underline{\mathbf{A}}$ as m vectors in $\mathbb{F}^{(n)}$ and its columns as n vectors in $\mathbb{F}_{(m)}$. This lead to the introduction in Definition 9.43 on page 117 of the column space and the row space of the matrix $\underline{\mathbf{A}}$ as, respectively

$$\langle \mathbf{c}_1^{\mathbf{A}}, \ldots, \mathbf{c}_n^{\mathbf{A}} \rangle \ \leq \ \mathbf{F}_{(m)} \qquad \text{and} \qquad \langle \mathbf{r}_1^{\mathbf{A}}, \ldots, \mathbf{r}_m^{\mathbf{A}} \rangle \ \leq \ \mathbf{F}^{(n)}$$

The column rank and the row rank of the matrix $\underline{\mathbf{A}}$, are the dimensions of these spaces.

Definition 10.6. The *column rank* of $\underline{\mathbf{A}} := [a_{ij}]_{m \times n}$ is the dimension of its column space

$$\text{colrk}(\underline{\mathbf{A}}) := \dim_{\mathbb{F}} \left(\langle \mathbf{c}_1, \ldots, \mathbf{c}_n \rangle \right)$$

and its *row rank* is the dimension of its row space

$$\text{rowrk}(\underline{\mathbf{A}}) := \dim_{\mathbb{F}} \left(\langle \mathbf{r}_1, \ldots, \mathbf{r}_m \rangle \right)$$

The *null space* of $\underline{\mathbf{A}}$ is $N(\underline{\mathbf{A}}) := \{ \mathbf{x} \in \mathbb{F}_{(n)} \mid \underline{\mathbf{A}}\, \mathbf{x} = \underline{\mathbf{0}} \}$ and the *nullity* of $\underline{\mathbf{A}}$ is the dimension of its null space

$$\text{n}(\underline{\mathbf{A}}) := \dim_{\mathbb{F}}(N(\underline{\mathbf{A}}))$$

Observation 10.7. When $\underline{\mathbf{A}}$ is the matrix of the linear transformation T, then the column space of $\underline{\mathbf{A}}$ corresponds to the image of T and the null space of $\underline{\mathbf{A}}$ corresponds to the kernel of T under the isomorphisms defined by choosing bases.

Hence, we may identify the column rank of $\underline{\mathbf{A}}$ with the rank of T.

We can apply the above to solving systems of linear equations. We represent the system of m equations in n unknowns

$$\begin{array}{ccccccc} a_{11}x_1 & + & \cdots & + & a_{1n}x_n & = & b_1 \\ \vdots & & & & \vdots & & \vdots \\ a_{m1}x_1 & + & \cdots & + & a_{mn}x_n & = & b_m \end{array} \qquad (\ddagger)$$

by the matrix equation

$$\underline{\mathbf{A}}\, \underline{\mathbf{x}} = \underline{\mathbf{b}},$$

where

$$\underline{\mathbf{A}} = [a_{ij}]_{m \times n}, \qquad \underline{\mathbf{x}} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \qquad \text{and} \qquad \underline{\mathbf{b}} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}$$

Theorem 10.8. *The system of equations, $(\ddagger)$, has a solution if and only if $\underline{\mathbf{b}}$ is in the column space of $\underline{\mathbf{A}}$.*

Proof. Immediate. □

The central fact about the row rank and the column rank of a matrix is surprising: they agree.

Theorem 10.9. *The row rank and the column rank of a matrix agree.*

Proof. Let $\underline{\mathbf{A}} := [a_{ij}]_{m\times n}$ have row rank p and column rank q.

Let $\mathbf{r}_i \in \mathbb{F}_{(n)}$ be the i-th row and $\mathbf{c}_j \in \mathbb{F}^{(m)}$, so that

$$\mathbf{c}_j := \begin{bmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{bmatrix} \quad \text{and} \quad \mathbf{r}_i := \begin{bmatrix} a_{i1} & \cdots & a_{in} \end{bmatrix} \qquad (i = 1, \ldots, m, \quad j = 1, \ldots, n).$$

Since the column rank of $\underline{\mathbf{A}}$ is not altered when we permute the columns of $\underline{\mathbf{A}}$, and since the row rank is unaltered by permuting its rows, we may assume, without loss of generality, that $\{\mathbf{c}_1, \ldots, \mathbf{c}_q\}$ and $\{\mathbf{r}_1, \ldots, \mathbf{r}_p\}$ are linearly independent.

Suppose that $q < n$.

Then $\mathbf{c}_n$ is a linear combination of $\{\mathbf{c}_1, \ldots, \mathbf{c}_{n-1}\}$, say

$$\mathbf{c}_n = \alpha_1\mathbf{c}_1 + \cdots + \alpha_{n-1}\mathbf{c}_{n-1}$$

so that

$$a_{in} = \sum_{j=1}^{n-1} \alpha_j a_{ij} \qquad (i = 1, \ldots, m).$$

Define the vector subspace W of $\mathbb{F}^{(n)}$ by

$$W = \left\{ [x_1 \cdots x_n] \in \mathbb{F}^{(n)} \mid x_n = \sum_{j=1}^{n-1} \alpha_j x_j \right\}$$

Then, plainly, $\mathbf{r}_i \in W$, for each i.

Hence, the row space of $\underline{\mathbf{A}}$

$$V := \langle \mathbf{r}_1, \ldots, \mathbf{r}_m \rangle$$

is a subspace of W.

Let $\underline{\mathbf{A}}'$ be the $m \times (n-1)$ matrix obtain from $\underline{\mathbf{A}}$ by deleting its n^{th} Êcolumn $\mathbf{c}_n$ and $\mathbf{r}'_i \in \mathbb{F}^{(n-1)}$ the i^{th} row of $\underline{\mathbf{A}}'$.

The function

$$T\colon V \longrightarrow \mathbb{F}^{(n-1)}, \quad [x_1 \cdots x_{n-1}\ x_n] \longmapsto [x_1 \cdots x_{n-1}]$$

is clearly a linear transformation.

Since $T(\mathbf{r}_i) = \mathbf{r}'_i$, $\mathrm{im}(T)$ is the row space of $\underline{\mathbf{A}}'$.

Thus, the rank of T is the row rank of $\underline{\mathbf{A}}'$.

Take $[x_1 \cdots x_n] \in \ker(T)$.

Then $T([x_1 \cdots x_n]) = [x_1 \cdots x_{n-1}] = [0 \cdots 0,]$ whence $x_j = 0$ for $1 \leq j < n$.

Since $x_n = \sum_{j=1}^{n-1} \alpha_j x_j$, we have $x_n = 0$ as well.

Hence $\ker(T) = \{[0 \cdots 0,]\}$, so that $\mathrm{n}(T) = 0$. It follows that

$$\begin{aligned} \mathrm{rowrk}(\underline{\mathbf{A}}') &= \mathrm{rk}(T) \\ &= \dim(V) - \mathrm{n}(T) \qquad \text{by Theorem 10.2} \end{aligned}$$

$$\begin{aligned} &= \dim(V) && \text{as } \mathrm{n}(T) = 0 \\ &= \mathrm{rowrk}(\underline{\mathbf{A}}) \end{aligned}$$

Thus $\underline{\mathbf{A}}$ and $\underline{\mathbf{A}}'$ have the same row ranks and the same column ranks.

If $p < m$, we may eliminate the last row, $\mathbf{r}'_m$, from $\underline{\mathbf{A}}'$ to form $\underline{\mathbf{A}}''$, an $(m-1) \times (n-1)$ matrix, which, by a similar argument to the one just presented, has the same column and row rank as $\underline{\mathbf{A}}$.

We continue this process of eliminating rows and columns until we arrive at a matrix $\widehat{\underline{\mathbf{A}}}$ with the same column and row rank as $\underline{\mathbf{A}}$, but whose columns and rows are all linearly independent.

By hypothesis, $\widehat{\underline{\mathbf{A}}}$ is a $p \times q$ matrix, whose rows, $\widehat{\mathbf{r}}_i \quad (i = 1, \dots p)$, and columns, $\widehat{\mathbf{c}}_j \quad (j = 1, \dots q)$, are linearly independent.

Since $\widehat{\mathbf{r}}_i \in \mathbb{F}^{(q)}$, it follows from Theorem 8.5 on page 88, that $p \leq q$.

Similarly, since $\widehat{\mathbf{c}}_j \in \mathbb{F}_{(p)}$, $q \leq p$.

Thus $p = q$. □

Theorem 10.9 on page 128 allows us to speak unambiguously of the *rank* of a matrix.

Definition 10.10. The *rank* of the matrix $\underline{\mathbf{A}}$, $\mathrm{rk}(\underline{\mathbf{A}})$, is its row rank, or, equivalently, its column rank.

Theorem 10.11. *Let $T \colon V \longrightarrow W$ be a linear transformation of finitely generated vector spaces. Let $\underline{\mathbf{A}}$ be any matrix representing T. Then $\mathrm{rk}(\underline{\mathbf{A}}) = \mathrm{rk}(T)$ and $n(\underline{\mathbf{A}}) = n(T)$.*

Proof. Exercise. □

Lemma 10.12. *If the $m \times n$ matrix, $\underline{\mathbf{A}}$, is invertible then $m = n$.*

Proof. By Corollary 9.39 on page 116, we may identify the $m \times n$ matrix, $\underline{\mathbf{A}}$ with the linear transformation

$$L_{\underline{\mathbf{A}}} \colon \mathbb{F}_{(n)} \longrightarrow \mathbb{F}_{(m)}, \quad \mathbf{x} \longmapsto \underline{\mathbf{A}}\mathbf{x}.$$

and $\underline{\mathbf{A}}$ is invertible if and only if $L_{\underline{\mathbf{A}}}$ is an isomorphism.

In this case $L_{\underline{\mathbf{A}}}$ is surjective and injective, whence $\mathrm{rk}(L_{\underline{\mathbf{A}}}) = \dim(\mathbb{F}_{(m)}) = m$ and $\mathrm{n}(L_{\underline{\mathbf{A}}}) = 0$, respectively.

By Theorem 10.2 on page 125, $m + 0 = \dim(\mathbb{F}_{(n)}) = n$. □

Finally, we turn to the relation between the ranks of two matrices and the rank of their product.

Lemma 10.13. *Take $\underline{\mathbf{A}} \in \mathbf{M}(p \times q; \mathbb{F})$ and $\underline{\mathbf{B}} \in \mathbf{M}(q \times r; \mathbb{F})$. Then*

$$\mathrm{rk}(\underline{\mathbf{A}}\,\underline{\mathbf{B}}) \;\leq\; \min\{\mathrm{rk}(\underline{\mathbf{A}}), \mathrm{rk}(\underline{\mathbf{B}})\}$$

Proof. By Section 9.4, the rows of $\underline{\mathbf{A}}\,\underline{\mathbf{B}}$ are linear combinations of the rows of $\underline{\mathbf{B}}$. Hence, there cannot be more linearly independent rows in $\underline{\mathbf{A}}\,\underline{\mathbf{B}}$ than there are in $\underline{\mathbf{B}}$.

Thus, $\mathrm{rk}(\underline{\mathbf{A}}\,\underline{\mathbf{B}}) \leq \mathrm{rk}(\underline{\mathbf{B}})$.

By Section 9.4, the columns of $\underline{\mathbf{A}}\,\underline{\mathbf{B}}$ are linear combinations of the columns of $\underline{\mathbf{A}}$. Thus, there cannot be more linearly independent columns in $\underline{\mathbf{A}}\,\underline{\mathbf{B}}$ than there are in $\underline{\mathbf{A}}$.

Thus, $\mathrm{rk}(\underline{\mathbf{A}}\,\underline{\mathbf{B}}) \leq \mathrm{rk}(\underline{\mathbf{A}})$.

Since the row rank and the column rank of a matrix agree, $\mathrm{rk}(\underline{\mathbf{A}}\,\underline{\mathbf{B}}) \leq \min\{\mathrm{rk}(\underline{\mathbf{A}}), \mathrm{rk}(\underline{\mathbf{B}})\}$. □

We show that equality need not hold, in general.

Example 10.14. Clearly, both the matrices

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

have rank 1. But their product

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

has rank 0.

If, on the other hand, one of the matrices is invertible, then equality does hold.

Corollary 10.15. *If* $\underline{\mathbf{A}}$ *is invertible, then* $\mathrm{rk}(\underline{\mathbf{A}}\,\underline{\mathbf{B}}) = \mathrm{rk}(\underline{\mathbf{B}})$

Proof. If $\underline{\mathbf{A}} \in \mathbf{M}(q; \mathbb{F})$ is invertible, then

$$L_{\underline{\mathbf{A}}} \colon \mathbb{F}_{(m)} \longrightarrow \mathbb{F}_{(m)}, \quad \mathbf{x} \longmapsto \underline{\mathbf{A}}\mathbf{x}$$

is an isomorphism.

Let $\mathbf{c}_1, \ldots, \mathbf{c}_n \in \mathbb{F}_{(m)}$ be the columns of $\underline{\mathbf{B}}$.

Then $\underline{\mathbf{A}}\mathbf{c}_1 = L_{\underline{\mathbf{A}}}(\mathbf{c}_1), \ldots, \underline{\mathbf{A}}\mathbf{c}_n = L_{\underline{\mathbf{A}}}(\mathbf{c}_n)$ are the columns of $\underline{\mathbf{A}}\,\underline{\mathbf{B}}$.

Since $L_{\underline{\mathbf{A}}}$ is an isomorphism, it follows by Lemma 8.12 on page 90(iii) that $\{\mathbf{c}_{j_1}, \ldots, \mathbf{c}_{j_r}\}$ is a basis for the column space of $\underline{\mathbf{B}}$ if and only if $\{\underline{\mathbf{A}}\mathbf{c}_{j_1}, \ldots, \underline{\mathbf{A}}\mathbf{c}_{j_r}\}$ is a basis for the column space of $\underline{\mathbf{A}}\,\underline{\mathbf{B}}$. □

Corollary 10.16. *If* $\underline{\mathbf{B}}$ *is invertible, then* $\mathrm{rk}(\underline{\mathbf{A}}\,\underline{\mathbf{B}}) = \mathrm{rk}(\underline{\mathbf{A}})$

Proof. Exercise. □

10.2 Calculating the Column Space and the Null Space of a Matrix

Given their significance, it is important to be able to calculation of the column space and the null space of a matrix.

It may surprise the reader that the calculation can be completed using the Gauß-Jordan procedure, familiar from your earlier studies, for example, MATH101. This comprises applying the *elementary row operations* to the given matrix in question to transform it to *(reduced) row-echelon form*.

By Observation 9.46 on page 119, each elementary row operation can be performed on a matrix $\underline{\mathbf{B}}$ by mulitiplying it on the left by a suitable matrix.

The details follow.

10.2.1 Elementary Row Operations

The elementary row operations apply to matrices and we can regard each of these operations as a function $\mathbf{M}(m \times n; \mathbb{F}) \longrightarrow \mathbf{M}(m \times n; \mathbb{F})$. Recall that we write $\mathbf{M}(n; \mathbb{F})$ for $\mathbf{M}(n \times n; \mathbb{F})$.

The elementary row operations are

ER01 **M**ultiplication of the $\mathbf{i}^{\text{th}}$ row by $\lambda \in \mathbb{F} \quad (\lambda \neq 0)$.

ER02 **A**ddition of μ times the $\mathbf{j}^{\text{th}}$ row to the $\mathbf{i}^{\text{th}}$ row $\quad (\mathbf{i} \neq \mathbf{j})$.

ER03 **S**wapping the $\mathbf{i}^{\text{th}}$ row with the $\mathbf{j}^{\text{th}}$ row $\quad (\mathbf{i} \neq \mathbf{j})$.

Clearly, each the elementary row operations defines a function $\mathbf{M}(n;\mathbb{F}) \longrightarrow \mathbf{M}(n;\mathbb{F})$. Moreover, each of these functions is bijective, with obvious inverses, namely, multiplying the i^{th} row by $\frac{1}{\lambda}$, adding $-\mu$ times the j^{th} row to the i^{th} row, and, finally, swapping the i^{th} row with the j^{th} row.

We illustrate how the operations can be performed using matrix multiplication on the left, using concrete examples for $\mathbf{M}(2;\mathbb{F})$, before presenting the general form.

Example 10.17. Take $\underline{\mathbf{B}} := \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathbf{M}(2;\mathbb{F})$

ER01 If we multiply the second row of $\underline{\mathbf{B}}$ by λ, we obtain

$$\begin{bmatrix} a & b \\ \lambda c & \lambda d \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & \lambda \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

ER02 If we add μ times the second row of $\underline{\mathbf{B}}$ to the first row of $\underline{\mathbf{B}}$ we obtain

$$\begin{bmatrix} a+\mu c & b+\mu d \\ c & d \end{bmatrix} = \begin{bmatrix} 1 & \mu \\ 0 & 1 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

ER03 If we swap the first and second rows of $\underline{\mathbf{B}}$, we obtain

$$\begin{bmatrix} c & d \\ a & b \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

The general form of the matrices performing the elementary row operations is now determined, and we list them next, illustrating each using examples from $\mathbf{M}(4;\mathbb{F})$.

ERO1 $\mathbf{M}(m \times n;\mathbb{F}) \longrightarrow \mathbf{M}(m \times n;\mathbb{F}), \quad \underline{\mathbf{B}} \longmapsto \underline{\mathbf{M}}(i,\lambda)\underline{\mathbf{B}}$ where

$$\underline{\mathbf{M}}(i,\lambda) := [m_{k\ell}]_{n\times n}, \quad \text{with } m_{k\ell} = \begin{cases} \lambda & \text{if } k = \ell = i \\ 1 & \text{if } k = \ell \neq i \\ 0 & \text{if } k \neq \ell \end{cases}. \tag{ero1}$$

Example 10.18. $\underline{\mathbf{M}}(2,\lambda) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \lambda & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$

ERO2 $\mathbf{M}(m \times n;\mathbb{F}) \longrightarrow \mathbf{M}(m \times n;\mathbb{F}), \quad \underline{\mathbf{B}} \longmapsto \underline{\mathbf{A}}(i,\mu j)\underline{\mathbf{B}}$ where

$$\underline{\mathbf{A}}(i,\mu j) := [a_{k\ell}]_{n\times n}, \quad \text{where } a_{k\ell} = \begin{cases} 1 & \text{if } k = \ell \\ \mu & \text{if } k = i,\ \ell = j \\ 0 & \text{otherwise} \end{cases}. \tag{ero2}$$

Example 10.19. $\underline{\mathbf{A}}(1,\mu 3) = \begin{bmatrix} 1 & 0 & \mu & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$

ERO3 $\quad \mathbf{M}(m\times n;\mathbb{F}) \longrightarrow \mathbf{M}(m\times n;\mathbb{F}), \quad \underline{\mathbf{B}} \longmapsto \underline{\mathbf{S}}(i,j)\underline{\mathbf{B}}$ $\quad$ where

$$\underline{\mathbf{S}}(i,j) := [s_{k\ell}]_{n\times n}, \quad \text{where } s_{k\ell} = \begin{cases} 1 & \text{if } k=\ell\neq i,j \\ 1 & \text{if } k=i,\ \ell=j \\ 1 & \text{if } k=j,\ \ell=i \\ 0 & \text{otherwise} \end{cases} \tag{\textbf{ero3}}$$

Example 10.20. $\underline{\mathbf{S}}(2,4) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$

The next theorem summarises these considerations.

Theorem 10.21. *Take $\lambda,\mu\in\mathbb{F}$, with $\lambda\neq 0$, $m,n\in\mathbb{N}$ and $i,j\in\mathbb{N}$ with $1\leq i\leq m,\ 1\leq j\leq n$ and $i\neq j$. Then we have isomorphisms.*

$$\begin{aligned} M_{(i,\lambda)}&\colon \mathbf{M}(m\times n;\mathbb{F}) \longrightarrow \mathbf{M}(m\times n;\mathbb{F}), && \underline{\mathbf{B}} \longmapsto \underline{\mathbf{M}}(i,\lambda)\underline{\mathbf{B}} \\ A_{(ß,\mu j)}&\colon \mathbf{M}(m\times n;\mathbb{F}) \longrightarrow \mathbf{M}(m\times n;\mathbb{F}), && \underline{\mathbf{B}} \longmapsto \underline{\mathbf{A}}(i,\mu j)\underline{\mathbf{B}} \\ S_{(i,j)}&\colon \mathbf{M}(m\times n;\mathbb{F}) \longrightarrow \mathbf{M}(m\times n;\mathbb{F}), && \underline{\mathbf{B}} \longmapsto \underline{\mathbf{S}}(i,j)\underline{\mathbf{B}} \end{aligned}$$

Proof. That these functions are linear transformations follows directly from the definition of matrix multiplication, and we have already seen that they are bijective. □

Theorem 10.21, together with the discussion in Section 9.4, provides us with a procedure for determining the column space and the null space of a given matrix, which we state before illustrating with a concrete example.

Let the $\mathbf{c}_1,\dots\mathbf{c}_n$ be the columns of the $m\times n$ matrix. $\underline{\mathbf{B}}$. Apply elementary row operations to reduce $\underline{\mathbf{B}}$ to a matrix $\underline{\mathbf{E}}$ in echelon form.

Since each step comprises multiplication on the left by a suitable matrix of the form $\underline{\mathbf{M}}(i,\lambda), \underline{\mathbf{A}}(i,\mu j)$ or $\underline{\mathbf{S}}(i,j)$, it comprise the application of an isomorphism of the form $M_{(i,\lambda)}, A_{(i,\mu j)}$ or $S_{(i,j)}$.

Since we obtain $\underline{\mathbf{E}}$ by composing isomorphisms, we have an isomorphism

$$T\colon \mathbf{M}(m\times n;\mathbb{F}) \longrightarrow \mathbf{M}(m\times n;\mathbb{F}), \qquad \underline{\mathbf{X}} \longmapsto \underline{\mathbf{C}}\,\underline{\mathbf{X}}$$

with $\underline{\mathbf{C}}$ a product of matrices of the form $\underline{\mathbf{M}}(i,\lambda), \underline{\mathbf{A}}(i,\mu j)$ or $\underline{\mathbf{S}}(i,j)$.

Thus, $\underline{\mathbf{C}}$ is invertible and $\underline{\mathbf{E}} = \underline{\mathbf{C}}\,\underline{\mathbf{B}}$.

By the proof of Corollary 10.15 on page 131, $\{\mathbf{c}_{j_1},\dots,\mathbf{c}_{j_r}\}$ is a basis for the column space of $\underline{\mathbf{B}}$ if and only if $\{\underline{\mathbf{C}}\mathbf{c}_{j_1},\dots,\underline{\mathbf{C}}\mathbf{c}_{j_r}\}$ is a basis for the column space of $\underline{\mathbf{C}}\,\underline{\mathbf{B}} = \underline{\mathbf{E}}$.

The columns of $\underline{\mathbf{E}}$ which contain the "pivot 1"s — that is, the first non-zero element in a row — form a basis for the column space of $\underline{\mathbf{E}}$.

Hence the corresponding columns of $\underline{\mathbf{B}}$ form a basis for the column space of $\underline{\mathbf{B}}$.

Example 10.22.

$$\underline{\mathbf{B}} = \begin{bmatrix} 1 & 2 & 2 & 5 \\ 3 & 6 & 1 & 10 \\ 1 & 2 & -1 & 2 \end{bmatrix}$$

We apply elementary row operations to transform $\underline{\mathbf{B}}$ to reduced row echelon form

$$\begin{bmatrix} 1 & 2 & 2 & 5 \\ 3 & 6 & 1 & 10 \\ 1 & 2 & -1 & 2 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & 2 & 2 & 5 \\ 0 & 0 & -5 & -5 \\ 0 & 0 & -3 & -3 \end{bmatrix} \qquad \begin{matrix} R_1 \rightsquigarrow R_1 \\ R_2 \rightsquigarrow R_2 - 3R_1 \\ R_3 \rightsquigarrow R_3 - R_1 \end{matrix}$$

$$\rightsquigarrow \begin{bmatrix} 1 & 2 & 2 & 5 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & -3 & -3 \end{bmatrix} \qquad \begin{matrix} R_1 \rightsquigarrow R_1 \\ R_2 \rightsquigarrow \frac{1}{-5} R_2 \\ R_3 \rightsquigarrow R_3 \end{matrix}$$

$$\rightsquigarrow \begin{bmatrix} 1 & 2 & 2 & 5 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \qquad \begin{matrix} R_1 \rightsquigarrow R_1 \\ R_2 \rightsquigarrow R_2 \\ R_3 \rightsquigarrow R_3 + 3R_2 \end{matrix}$$

Thus $\underline{\mathbf{E}} = \begin{bmatrix} 1 & 2 & 2 & 5 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$.

Since the pivot elements are in the first and third columns of $\underline{\mathbf{E}}$, a basis for the column space of $\underline{\mathbf{B}}$ is given by the first and third columns of $\underline{\mathbf{B}}$.

In other words, $\begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix}$ comprise a basis for the column space of $\underline{\mathbf{B}}$.

We can also read off a basis for the null space of $\underline{\mathbf{B}}$ from its echelon form.

Example 10.23. Turning to $N(\underline{\mathbf{B}})$, the null space of $\underline{\mathbf{B}}$, recall that $\mathbf{x} \in \mathbb{F}_{(n)}$ is an element of $N(\underline{\mathbf{B}})$ if and only if $\underline{\mathbf{B}}\mathbf{x} = \mathbf{0} \in \mathbb{F}_{(m)}$.

Since $\underline{\mathbf{C}}$ is an invertible matrix, this is equivalent to $\underline{\mathbf{E}}\mathbf{x} = \underline{\mathbf{C}}\underline{\mathbf{B}}\mathbf{x} = \mathbf{0}$.

Putting $\mathbf{x} = \begin{bmatrix} w \\ x \\ y \\ z \end{bmatrix}$, we obtain

$$w = -2x - 2y - 5z$$
$$y = -z$$

or, equivalently

$$\mathbf{x} = \begin{bmatrix} w \\ x \\ y \\ z \end{bmatrix} = \begin{bmatrix} -2x - 3z \\ x \\ -z \\ z \end{bmatrix} = x \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + z \begin{bmatrix} -3 \\ 0 \\ -1 \\ 1 \end{bmatrix}$$

In other words, $\begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} -3 \\ 0 \\ -1 \\ 1 \end{bmatrix}$ comprise a basis for the null space of $\underline{\mathbf{B}}$.

Observation 10.24. The above allows us to find a basis for the image and kernel of a linear transformation. We illustrate this using the matrix above.

Example 10.25. Let $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3, \mathbf{e}_4\}$ be a basis for V and $\{\mathbf{f}_1, \mathbf{f}_2, \mathbf{f}_3\}$ a basis for W.

Let $T\colon V \to W$ be a linear transformation whose matrix with respect to these bases is

$$\underline{\mathbf{B}} = \begin{bmatrix} 1 & 2 & 2 & 5 \\ 3 & 6 & 1 & 10 \\ 1 & 2 & -1 & 2 \end{bmatrix}$$

It is immediate from the above that $\mathbf{f}_1 + 3\mathbf{f}_2 + \mathbf{f}_3$ and $2\mathbf{f}_1 + \mathbf{f}_2 - \mathbf{f}_3$ form a basis for $\mathrm{im}(T)$ and that $-2\mathbf{e}_1 + \mathbf{e}_2$ and $-3\mathbf{e}_1 - \mathbf{e}_3 + \mathbf{e}_4$ comprise a basis for $\ker(T)$.

10.3 Finding the Inverse of an $n \times n$ Matrix

We have shown that the $n \times n$ matrix $\underline{\mathbf{A}}$ may be identified with the linear transformation

$$L_{\underline{\mathbf{A}}}\colon \mathbb{F}_{(n)} \longrightarrow \mathbb{F}_{(n)}, \quad \underline{\mathbf{x}} \longmapsto \underline{\mathbf{A}}\,\underline{\mathbf{x}}$$

and that $\underline{\mathbf{A}}$ is invertible if and only if $L_{\underline{\mathbf{A}}}$ is an isomorphism.

Since $\dim_{\mathbb{F}}(\mathbb{F}_{(n)}) = n$, it follows from Theorem 10.2 on page 125 and Lemma 10.4 on page 126 that this is the case if and only if $\mathrm{rk}(\underline{\mathbf{A}}) = n$.

To see how this can be applied, let the $n \times n$ matrix $\underline{\mathbf{A}}$ have rank n. This means that the column space of $\underline{\mathbf{A}}$ is all of $\mathbb{F}_{(n)}$. Let $\underline{\mathbf{e}}_j$ be the j^{th} column of $\underline{\mathbf{1}}_n$, the $n \times n$ identity matrix, so that

$$\underline{\mathbf{e}}_j = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

where

$$x_i = \begin{cases} 1 & \text{when } i = j \\ 0 & \text{otherwise} \end{cases}$$

Since each $\underline{\mathbf{e}}_j$ is in the column space of $\underline{\mathbf{A}}$, there are $d_{ij} \in \mathbb{F}$ $(1 \leq i \leq n)$, such that

$$\underline{\mathbf{e}}_j = \underline{\mathbf{c}}_1^{\mathbf{A}}\, d_{1j} + \cdots + \underline{\mathbf{c}}_n^{\mathbf{A}}\, d_{nj}$$

where $\underline{\mathbf{c}}_j^{\mathbf{A}}$ is the j^{th} column of $\underline{\mathbf{A}}$. This is equivalent to

$$\underline{\mathbf{e}}_j = \underline{\mathbf{A}}\,\underline{\mathbf{d}}_j$$

where

$$\underline{\mathbf{d}}_j = \begin{bmatrix} d_{1j} \\ \vdots \\ d_{nj} \end{bmatrix}$$

It follows immediately that

$$\underline{\mathbf{A}}\,\underline{\mathbf{D}} = \underline{\mathbf{1}}_n$$

where $\underline{\mathbf{D}}$ is the $n \times n$ matrix whose j^{th} column is $\underline{\mathbf{d}}_j$. In other words, $\underline{\mathbf{D}}$ is right inverse to $\underline{\mathbf{A}}$.

Since the row rank of a matrix agrees with its column rank, an $n \times n$ has a left inverse if and only if it has a right inverse, whence $\underline{\mathbf{D}}$ must be the inverse of $\underline{\mathbf{A}}$.

The following lemma provides necessary and sufficent conditions for an $n \times n$ matrix to have rank n. The proof, which is essentially the Gauß-Jordan algorithm for transforming a matrix to reduced row echelon form, also provides a method for finding the inverse of an invertible $n \times n$ matrix.

Lemma 10.26. *The rank of $\underline{\mathbf{B}} \in \mathbf{M}(n; \mathbb{F})$ is n if and only if $\underline{\mathbf{B}}$ can be transformed into $\underline{\mathbf{1}}_n$ by means of elementary row operations.*

Proof. $\Leftarrow$: This follows from Corollary 10.15 on page 131 because $\mathrm{rk}(\underline{\mathbf{1}}_n) = n$.

$\Rightarrow$: Let $\underline{\mathbf{B}} = [b_{ij}]_{n \times n}$ have rank n.

We apply the Gauß-Jordan procedure to $\underline{\mathbf{B}}$.

Since the columns of $\underline{\mathbf{B}}$ are linearly independent, no column of $\underline{\mathbf{B}}$ can be the zero column.

In particular, there is an i such that $b_{i1} \neq 0$.

Then $\underline{\mathbf{B}}''_1 = \underline{\mathbf{M}}(i, \frac{1}{b_{1i}}1)\underline{\mathbf{S}}(1, i)\underline{\mathbf{B}}$ is of the form

$$\begin{bmatrix} 1 & * \\ * & * \end{bmatrix}.$$

Put $\underline{\mathbf{B}}_1 := \underline{\mathbf{A}}(n, -\hat{b}_{n1}1) \cdots \underline{\mathbf{A}}(2, -\hat{b}_{21}1)\underline{\mathbf{B}}''_1$, where $\hat{b}_{j1} = \begin{cases} b_{11} & \text{if } j = i \\ b_{j1} & \text{otherwise} \end{cases}$.

Then $\underline{\mathbf{B}}_1$ is of the form

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

In other words, using only elementary row operations, we have transformed $\underline{\mathbf{B}}$ into a matrix whose first column contains a 1 in the first row and all other coefficients are 0.

Now suppose that we have applied elementary row operations to transform $\underline{\mathbf{B}}$ into a matrix each of whose first k columns contains precisely one 1, and all other coefficents 0, with the only 1 of the j^{th} column sitting in the j^{th} row.

In other words, we assume that we have applied elementary row operations to transform $\underline{\mathbf{B}}$ into $\underline{\mathbf{B}}_k = [c_{ij}]_{n \times n}$ for some $k < n$, with $\underline{\mathbf{B}}_k$ of the form

$$\begin{bmatrix} \underline{\mathbf{1}}_k & * \\ \underline{\mathbf{0}} & * \end{bmatrix}.$$

Since $\mathrm{rk}(\underline{\mathbf{B}}_k) = \mathrm{rk}(\underline{\mathbf{B}}) = n$, $c_{i(k+1)} \neq 0$ for some $i > k$. For otherwise, the $(k+1)^{\text{st}}$ column would be a linear combination of the first k columns.

Swap the i^{th} row and the $(k+1)^{\text{st}}$ row to obtain the matrix

$$\underline{\mathbf{B}}'_{k+1} = \underline{\mathbf{S}}(k+1, i)\underline{\mathbf{B}}_k$$

which is of the form

$$\begin{bmatrix} \underline{\mathbf{1}}_k & * & * \\ \underline{\mathbf{0}} & c_{i(k+1)} & * \\ \underline{\mathbf{0}} & * & * \end{bmatrix}$$

Multiplying the $(k+1)^{\text{st}}$ row by $\frac{1}{c_{i(k+1)}}$ results in the matrix

$$\underline{\mathbf{B}}''_{k+1} = \underline{\mathbf{M}}(k+1, \tfrac{1}{c_{i(k+1)}})$$

which is of the form

$$\begin{bmatrix} \underline{\mathbf{1}}_k & * & * \\ \underline{\mathbf{0}} & 1 & * \\ \underline{\mathbf{0}} & * & * \end{bmatrix}$$

Put

$$\underline{\mathbf{B}}_{k+1} := \underline{\mathbf{A}}(n, -\hat{c}_{n1}1) \cdots \underline{\mathbf{A}}(2, -\hat{c}_{21}1)\underline{\mathbf{B}}''_{k+1}$$

where $\hat{c}_{j(k+1)} = \begin{cases} c_{(k+1)(k+1)} & \text{if } j = i \\ c_{j(k+1)} & \text{otherwise} \end{cases}$.

Then $\underline{\mathbf{B}}_{k+1}$ is of the form

$$\begin{bmatrix} \underline{\mathbf{1}}_{k+1} & * \\ \underline{\mathbf{0}} & * \end{bmatrix}$$

In particular $\underline{\mathbf{B}}_n = \underline{\mathbf{1}}_n$. □

Observation 10.27. Each step proof of Lemma 10.26 on the preceding page was consisted of multiplying (on the left) by an $n \times n$ matrix of the form $\underline{\mathbf{M}}(i, \lambda)$, $\underline{\mathbf{A}}(i, \mu j)$ or $\underline{\mathbf{S}}(i, j)$.

Letting the matrix $\underline{\mathbf{A}}$ be the product (in the order in which they were applied) of the matrices used to transform $\underline{\mathbf{B}}$ into $\underline{\mathbf{1}}_n$, it is immediate that $\underline{\mathbf{A}}$ is the left inverse of $\underline{\mathbf{B}}$, and hence, by our earlier observation, its inverse.

This provides a practical procedure for determining whether the $n \times n$ matrix $\underline{\mathbf{B}}$ has an inverse and, at the same time, finding the inverse when the matrix is invertible.

Step 1 Augment the $n \times n$ matrix $\underline{\mathbf{B}}$ by the $n \times n$ identity matrix $\underline{\mathbf{1}}_n$, to obtain

$$\left[\begin{array}{c|c} \underline{\mathbf{B}} & \underline{\mathbf{1}}_n \end{array}\right]$$

Step 2 Use elementary row operations to transform the augmented matrix into reduced row echelon form

$$\left[\begin{array}{c|c} \underline{\mathbf{E}} & \underline{\mathbf{A}} \end{array}\right]$$

Notice that $\underline{\mathbf{A}}$ is the product (in the order in which they were applied) of the matrices used to transform $\underline{\mathbf{B}}$ into $\underline{\mathbf{E}}$.

Step 3 $\underline{\mathbf{B}}$ is invertible if and only if $\underline{\mathbf{E}} = \underline{\mathbf{1}}_n$, in which case $\underline{\mathbf{B}}^{-1} = \underline{\mathbf{A}}$.

Example 10.28. We consider the real matrix

$$\underline{\mathbf{B}} = \begin{bmatrix} 1 & 2 & 4 \\ 1 & 3 & 9 \\ 1 & 4 & 16 \end{bmatrix}$$

Step 1 We augment $\underline{\mathbf{B}}$ by $\underline{\mathbf{1}}_3$ and obtain

$$\left[\begin{array}{ccc|ccc} 1 & 2 & 4 & 1 & 0 & 0 \\ 1 & 3 & 9 & 0 & 1 & 0 \\ 1 & 4 & 16 & 0 & 0 & 1 \end{array}\right]$$

Step 2 We apply elementary row operations to transform the augmented matrix to reduced row echelon form.

$$\begin{array}{lll}
\left[\begin{array}{ccc|ccc} 1 & 2 & 4 & 1 & 0 & 0 \\ 1 & 3 & 9 & 0 & 1 & 0 \\ 1 & 4 & 16 & 0 & 0 & 1 \end{array}\right] & \rightsquigarrow & \left[\begin{array}{ccc|ccc} 1 & 2 & 4 & 1 & 0 & 0 \\ 0 & 1 & 5 & -1 & 1 & 0 \\ 0 & 2 & 12 & -1 & 0 & 1 \end{array}\right] \quad \begin{array}{l} \\ R_2 - R_1 \\ R_3 - R_1 \end{array} \\
& \rightsquigarrow & \left[\begin{array}{ccc|ccc} 1 & 0 & -6 & 3 & -2 & 0 \\ 0 & 1 & 5 & -1 & 1 & 0 \\ 0 & 0 & 2 & 1 & -2 & 1 \end{array}\right] \quad \begin{array}{l} R_1 - 2R_2 \\ \\ R_3 - 2R_2 \end{array} \\
& \rightsquigarrow & \left[\begin{array}{ccc|ccc} 1 & 0 & -6 & 3 & -2 & 0 \\ 0 & 1 & 5 & -1 & 1 & 0 \\ 0 & 0 & 1 & \frac{1}{2} & -1 & \frac{1}{2} \end{array}\right] \quad \begin{array}{l} \\ \\ \frac{1}{2} \times R_3 \end{array} \\
& \rightsquigarrow & \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 6 & -8 & 3 \\ 0 & 1 & 0 & -\frac{7}{2} & 6 & -\frac{5}{2} \\ 0 & 0 & 1 & \frac{1}{2} & -1 & \frac{1}{2} \end{array}\right] \quad \begin{array}{l} R_1 + 6R_3 \\ R_2 - 5R_3 \\ \\ \end{array}
\end{array}$$

The augmented matrix is now in reduced row echelon form.

Step 3 Since the left hand matrix in the reduced row echelon form of the augmented matrix is $\underline{\mathbf{1}}_3$, our original matrix, $\underline{\mathbf{B}}$, is invertible and its inverse is

$$\underline{\mathbf{B}}^{-1} = \frac{1}{2}\begin{bmatrix} 12 & -16 & 6 \\ -7 & 12 & -5 \\ 1 & -2 & 1 \end{bmatrix}$$

10.4 Exercises

Exercise 10.1. Let $T\colon V \longrightarrow W$ be a linear transformation of finitely generated vector spaces.

Let $\underline{\mathbf{A}}$ be any matrix representing T.

Prove that $\mathrm{rk}(\underline{\mathbf{A}}) = \mathrm{rk}(T)$ and $\mathrm{n}(\underline{\mathbf{A}}) = \mathrm{n}(T)$.

Exercise 10.2. Prove that for the linear transformation $T\colon V \longrightarrow W$,

(i) $V \cong \ker(T) \oplus \mathrm{im}(T)$.

(ii) $\operatorname{im} T \cong {}^{V}/_{\ker(T)}$.

Exercise 10.3. Find a basis for the column space and a basis for the null space of each of the following real matrices.

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

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

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

(iv) $\begin{bmatrix} 2 & 3 & 6 & 5 \\ 1 & 0 & 8 & 7 \\ 6 & 8 & 1 & 3 \\ 7 & 14 & -11 & -8 \end{bmatrix}$

Exercise 10.4. Find a basis for the image and a basis for the kernel of each of the following linear transformations.

(a) $T\colon \mathbb{C}^3 \longrightarrow \mathbb{C}^3, \quad (u, v, w) \longmapsto (3u + 18v + 13w,\ 2u + 11v + 8w,\ u + 10v + 7w)$

(b) $T\colon \mathbb{R}^3 \longrightarrow \mathbb{R}^2, \quad (x, y, z) \longmapsto (x + y + z, 2x + 3y + 4z)$

Exercise 10.5. Prove that if $\underline{\mathbf{B}}$ is invertible, then $\operatorname{rk}(\underline{\mathbf{A}}\,\underline{\mathbf{B}}) = \operatorname{rk}(\underline{\mathbf{A}})$, whenever $\underline{\mathbf{A}}\,\underline{\mathbf{B}}$ is defined.

Exercise 10.6. Prove that an $n \times n$ matrix has a left inverse if and only if it has a right inverse.

Exercise 10.7. We work over $\mathbb{F}_3$, the field with precisely three elements (cf Exercise 3.2 on page 40).

Find, if possible, the inverse of each of the folowing matrices

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

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

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

Exercise 10.8. We work over $\mathbb{R}$, the field of all real numbers.

Find, if possible, the inverse of each of the folowing matrices

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

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

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

If I only had an hour to solve a problem, I would use the first 55 minutes to pose the right question. For once I have formulated the right question, I would be able to solve the problem in less than five minutes.

Albert Einstein

Chapter 11

The Determinant and the Trace

This chapter introduces two important functions defined on matrices: the *determinant of an $n \times n$ matrix* and the *trace* of an arbitrary matrix. A closer examination of these functions reveals that they depend only on the linear transformations represented by the matrices in question.

We begin with the determinant of an $n \times n$ matrix, showing that there is one and only one function with its characteristic properties and derive further properties.

We shall also see that both the determinant and the trace are *invariant* in the sense that for $\underline{\mathbf{A}}, \underline{\mathbf{B}} \in \mathbf{M}(n; \mathbb{F})$ with $\underline{\mathbf{B}}$ invertible, $\underline{\mathbf{B}}^{-1}\underline{\mathbf{A}}\,\underline{\mathbf{B}}$ has the same determinant and trace as $\underline{\mathbf{A}}$. This means that any two matrices representing the same linear transformation must have the same determinant and trace, and so we can define the determinant and trace of a linear transformation $T\colon V \longrightarrow V$ whenever V is finite dimensional.

11.1 The Determinant

We begin with a statement of the main theorem on determinants and make some observations before attending to the proof.

Theorem 11.1 (**Main Theorem on Determinants**). *There is a unique function*

$$D\colon \mathbf{M}(n; \mathbb{F}) \longrightarrow \mathbb{F}$$

such that

D1 *D is linear in each row;*

D2 $D(\underline{\mathbf{A}}) = 0$ *whenever* $\operatorname{rk} \underline{\mathbf{A}} < n$*;*

D3 $D(\underline{\mathbf{1}}_n) = 1$*;*

Definition 11.2. The unique function $D\colon \mathbf{M}(n; \mathbb{F}) \longrightarrow \mathbb{F}$ satisfying D1, D2 and D2 is called the *determinant* function.

We write $|\, a_{ij} \,|$ for $\det([a_{ij}])$ so that, for example, $\begin{vmatrix} a & b \\ c & d \end{vmatrix}$ is the determinant of $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$.

Observation 11.3. Strictly speaking, there is not just one determinant function, but a family of determinant functions, one for each counting number n. However, it is customary to speak as if there were only one, and, in any case, there is little danger of confusion.

Our proof of Theorem 11.1 on the previous page makes use of elementary row operations, as formulated in Section 10.2.1.

Observation 11.4. The third elementary row operation — swapping rows of matrix — is superfluous, since it can be replaced by a sequence of the other two:

Exchanging the i^{th} and j^{th} rows can be achieved by

1. adding the i^{th} row to the j^{th}, then
2. subtracting the j^{th} row from the i^{th}, then
3. adding the i^{th} row to the j^{th} and, finally,
4. multiplying the i^{th} row by -1.

We can express this with the matrices introduced in Section 10.2.1 by

$$\underline{\mathbf{S}}(i,j) = \underline{\mathbf{M}}(i,-1)\underline{\mathbf{A}}(j,i)\underline{\mathbf{A}}(i,(-1)j)\underline{\mathbf{A}}(i,j)$$

Hence, while in applications they are extremely convenient and useful, we may dispense with our third elementary row operation, and the matrices $\underline{\mathbf{S}}(i,j)$ when proving general results.

Lemma 11.5. *If the function $G\colon \mathbf{M}(n;\mathbb{F}) \longrightarrow \mathbb{F}$ satisfies D1 and D2, then, for each $\underline{\mathbf{B}} \in \mathbf{M}(n;\mathbb{F})$,*

(i) $G\big(\underline{\mathbf{M}}(i,\lambda)\underline{\mathbf{B}}\big) = \lambda G(\underline{\mathbf{B}})$

(ii) $G\big(\underline{\mathbf{A}}(i,\mu j)\underline{\mathbf{B}}\big) = G(\underline{\mathbf{B}})$

Proof. **(i)** This is immediate from D1.

(ii) Given the matrix $\underline{\mathbf{B}}$, let $\underline{\mathbf{B}}^{(i)}_{(j)}$ be obtained by replacing the i^{th} row of $\underline{\mathbf{B}}$ by its j^{th} row.

Since it has two identical rows, $\operatorname{rk}(\underline{\mathbf{B}}^{(i)}_{(j)}) < n$. Thus

$$\begin{aligned} G\big(\underline{\mathbf{A}}(i,\mu j)\underline{\mathbf{B}}\big) &= G(\underline{\mathbf{B}}) + \mu G(\underline{\mathbf{B}}^{(i)}_{(j)}) && \text{by D1} \\ &= G(\underline{\mathbf{B}}) + 0 && \text{by D2, since } \operatorname{rk}(\underline{\mathbf{B}}^{(i)}_{(j)}) < n. \end{aligned}$$

□

Proof of the Main Theorem on Determinants (Theorem 11.1). Uniqueness:

Let $F, G\colon \mathbf{M}(n;\mathbb{F}) \longrightarrow \mathbb{F}$ satisfy D1, D2 and D3.

Choose $\underline{\mathbf{B}} \in \mathbf{M}(n;\mathbb{F})$.

If $\operatorname{rk}(\underline{\mathbf{B}}) < n$, then, by D2, $F(\underline{\mathbf{B}}) = G(\underline{\mathbf{B}}) = 0$.

If $\operatorname{rk}(\underline{\mathbf{B}}) = n$, then, by Lemma 10.26 on page 136, we can transform $\underline{\mathbf{B}}$ into $\underline{\mathbf{1}}_n$ by means of elementary row operations.

In other words, there is a matrix $\underline{\mathbf{T}} \in \mathbf{M}(n;\mathbb{F})$ such that $\underline{\mathbf{T}}\,\underline{\mathbf{B}} = \underline{\mathbf{1}}_n$. In fact, $\underline{\mathbf{T}}$ is the product of finitely many matrices each of which is of the form $\underline{\mathbf{M}}(i,\lambda)$ or $\underline{\mathbf{A}}(i,\mu j)$.

By Lemma 11.5 $F(\underline{\mathbf{T}}\,\underline{\mathbf{B}}) = \Delta F(\underline{\mathbf{B}})$ and $G(\underline{\mathbf{T}}\,\underline{\mathbf{B}}) = \Delta G(\underline{\mathbf{B}})$, where Δ is the product of the λs (with repetition) which occur in the $\underline{\mathbf{M}}(i,\lambda)$s.

Then

$$\Delta F(\underline{\mathbf{B}}) = F(\underline{\mathbf{T}}\,\underline{\mathbf{B}})$$

$$
\begin{aligned}
&= F(\underline{\mathbf{1}}_n) \\
&= 1 \\
&= G(\underline{\mathbf{1}}_n) \\
&= G(\underline{\mathbf{T}}\,\underline{\mathbf{B}}) \\
&= \Delta G(\underline{\mathbf{B}})
\end{aligned}
$$

Since $\Delta \neq 0$, we conclude that $F(\underline{\mathbf{B}}) = G(\underline{\mathbf{B}})$.

Existence:

Take any of the usual definitions from a first year mathematics course, and verify D1, D2 and D3.

We verify D1, D2 and D3 for the definition of the determinant using "expansion by the j^{th} column", which we recall.

Definition 11.6. For the 1×1 matrix $\underline{\mathbf{A}} = [a]$,

$$\det(\underline{\mathbf{A}}) := a.$$

For $n \geq 1$, $\underline{\mathbf{A}} = [a_{ij}]_{(n+1)\times(n+1)}$.
Let $\underline{\mathbf{A}}_{(i)(j)}$ be the $n \times n$ matrix obtained by deleting the i^{th} row and the j^{th} column from $\underline{\mathbf{A}}$, so that

$$\underline{\mathbf{A}}_{(i)(j)} = [x_{k\ell}]_{n\times n}$$

where

$$x_{k\ell} = \begin{cases} a_{kl} & \text{if } 1 \leq k < i \text{ and } 1 \leq \ell < j \\ a_{k(\ell+1)} & \text{if } 1 \leq k < i \text{ and } j \leq \ell \leq n \\ a_{(k+1)\ell} & \text{if } i \leq k \leq n \text{ and } 1 \leq \ell < j \\ a_{(k+1)(\ell+1)} & \text{if } i \leq k \leq n \text{ and } j \leq \ell \leq n \end{cases}$$

Then

$$\det(\underline{\mathbf{A}}) := \sum_{i=1}^{n+1} (-1)^{i+j} a_{ij} \det(\underline{\mathbf{A}}_{(i)(j)}) \qquad (\circledast)$$

Example 11.7. The reader may find it useful to bear a concrete instance in mind:

$$\det\left(\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}\right) =$$
$$a_{11} \det\left(\begin{bmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{bmatrix}\right) - a_{21} \det\left(\begin{bmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{bmatrix}\right) + a_{31} \det\left(\begin{bmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{bmatrix}\right).$$

It remains to verify that D1, D2 and D3 are satisfied.

D1, D2 and D3 are clearly satisfied when $n = 1$.

Given $n > 1$, we make the inductive hypothesis that D1, D2 and D3 hold for all $r \leq n$.

D1: The summand $(-1)^{i+j} a_{ij} \det\left(\underline{\mathbf{A}}_{(i)(j)}\right)$ is linear in the i^{th} row of $\underline{\mathbf{A}}$ because a_{ij} is and $\underline{\mathbf{A}}_{(i)(j)}$ is independent of it.

It is linear in the k^{th} row of $\underline{\mathbf{A}}$ when $k \neq i$, because $\underline{\mathbf{A}}_{(i)(j)}$ is and a_{ij} is independent of it.

Being the sum of functions linear in the rows of $\underline{\mathbf{A}}$, $\det(\underline{\mathbf{A}})$ is linear in the rows of $\underline{\mathbf{A}}$.

D2: If $\mathrm{rk}(\underline{\mathbf{A}}) < n$, then there is a row, say the p^{th}, of $\underline{\mathbf{A}}$ which is a linearly combination of the others.

We can find $\lambda_i \;\; (i = 1, \ldots, n, i \neq p)$ such that for each j

$$a_{pj} = \sum_{\substack{i=1 \\ i \neq p}}^{n} \lambda_i a_{ij}.$$

Thus, since we have already established linearity in each row, the determinant of $\underline{\mathbf{A}}$ is a linear combination of determinants of matrices with two identical rows.

It is therefore sufficient to show that the determinant of an $n \times n$ matrix with two identical rows is 0. We prove this using induction on n.

This is plainly true for $n = 2$.

So take $n > 2$ and assume the inductive hypothesis that the assertion is true for all $m \times m$ matrices with $m < n$.

Assume that rows p and $p + t$ are identical ($t \geq 1$). Then, since $\underline{\mathbf{A}}_{(i)(j)}$ has two identical rows whenever $i \neq p, p + t$, the inductive hypothesis implies that

$$\begin{aligned}\det(\underline{\mathbf{A}}) &= \sum_{i=1}^{n} (-1)^{i+j} a_{ij} \det(\underline{\mathbf{A}}_{(i)(j)}) \\ &= (-1)^{p+j} a_{pj} \det(\underline{\mathbf{A}}_{(p)(j)}) + (-1)^{p+t+j} a_{(p+t)j} \det(\underline{\mathbf{A}}_{(p+t)(j)})\end{aligned}$$

As $a_{pj} = a_{(p+t)j}$ it only remains to investigate the relationship between $\det(\underline{\mathbf{A}}_{(p)(j)})$ and $\det(\underline{\mathbf{A}}_{(p+t)(j)})$ when the p^{th} and $(p + t)^{\text{h}}$ rows of $\underline{\mathbf{A}}$ are identical.

If $t = 1$, then $\underline{\mathbf{A}}_{(p)(j)}$ and $\underline{\mathbf{A}}_{(p+t)(j)}$ are identical, so their determinants agree. In this case

$$\det(\underline{\mathbf{A}}) = (-1)^{p+j} a_{pj} \det(\underline{\mathbf{A}}_{(p)(j)}) + (-1)^{p+1+j} a_{pj} \det(\underline{\mathbf{A}}_{(p)(j)}) = 0.$$

If, on the other hand, $t > 1$, then $\underline{\mathbf{A}}_{(p+t)(j)}$ can be obtained from $\underline{\mathbf{A}}_{(p)(j)}$ by interchanging rows $(t - 1)$ times. We may assume as part of our inductive hypothesis that for $m \times m$ matrices with $m < n$, each such interchange alters the sign of the determinant. Then

$$\begin{aligned}\det(\underline{\mathbf{A}}) &= (-1)^{p+j} a_{pj} \det(\underline{\mathbf{A}}_{(p)(j)}) + (-1)^{p+t+j} a_{pj} (-1)^{t-1} \det(\underline{\mathbf{A}}_{(p)(j)}) \\ &= (-1)^{p+j} a_{pj} \det(\underline{\mathbf{A}}_{(p)(j)} (1 + (-1)^{2t-1}) \\ &= 0.\end{aligned}$$

D3: If $\underline{\mathbf{A}} = \underline{\mathbf{1}}_n$, then $a_{ij} = \delta_{ij}$ and the only non-zero summand in $\displaystyle\sum_{i=1}^{n} (-1)^{i+j} a_{ij} \det(\underline{\mathbf{1}}_{n\,(p)(j)})$ is $(-1)^{2j} \delta_{jj} \det(\underline{\mathbf{1}}_{n-1})$, and, plainly, this is 1. □

Observation 11.8. It follows from the uniqueness of the determinant function that the expansion by the j^{th} column is independent of the choice of j.

Corollary 11.9. *Given* $\underline{\mathbf{B}} \in \mathbf{M}(n; \mathbb{F})$, $\mathrm{rk}(\underline{\mathbf{B}}) = n$ *if and only if* $\det(\underline{\mathbf{B}}) \neq 0$.

Proof. By Lemma 10.26 on page 136 and using the notation from the proof of Theorem 11.1 on page 141, we see that $\underline{\mathbf{B}}$ is invertible if and only if $\operatorname{rk}(\underline{\mathbf{B}}) = n$ if and only if

$$\det(\underline{\mathbf{B}}) = \frac{1}{\Delta}.$$

□

Corollary 11.10. *Given* $\underline{\mathbf{A}}, \underline{\mathbf{B}} \in \mathbf{M}(n;\mathbb{F})$, $\det(\underline{\mathbf{A}}\,\underline{\mathbf{B}}) = \det(\underline{\mathbf{A}})\det(\underline{\mathbf{B}})$.

Proof. Since, by Lemma 10.13 on page 130, $\operatorname{rk}(\underline{\mathbf{A}}\,\underline{\mathbf{B}}) \leq \min\{\operatorname{rk}(\underline{\mathbf{A}}), \operatorname{rk}(\underline{\mathbf{B}})\}$,

$$\det(\underline{\mathbf{A}}\,\underline{\mathbf{B}}) = 0 = \det(\underline{\mathbf{A}})\det(\underline{\mathbf{B}})$$

whenever $\det(\underline{\mathbf{A}}) = 0$.

Suppose now, that $\det(\underline{\mathbf{A}}) \neq 0$, or, equivalently, $\operatorname{rk}(\underline{\mathbf{A}}) = n$. Define

$$F\colon \mathbf{M}(n;\mathbb{F}) \longrightarrow \mathbb{F}, \quad \underline{\mathbf{B}} \longmapsto \det(\underline{\mathbf{A}}\,\underline{\mathbf{B}}).$$

Since $L_{\underline{\mathbf{A}}}\colon \mathbf{M}(n;\mathbb{F}) \longrightarrow \mathbf{M}(n;\mathbb{F})$, $\underline{\mathbf{B}} \longmapsto \underline{\mathbf{A}}\,\underline{\mathbf{B}}$ is a linear transformation, F satisfies D1.

Moreover, since $\operatorname{rk}(\underline{\mathbf{A}}) = n$, this transformation is an isomorphism, whence $\operatorname{rk}(\underline{\mathbf{A}}\,\underline{\mathbf{B}}) = \operatorname{rk}(\underline{\mathbf{B}})$.

Thus $F(\underline{\mathbf{B}}) = 0$ whenever $\operatorname{rk}(\underline{\mathbf{B}}) < n$, showing that F satisfies D2.

Now $F(\underline{\mathbf{1}}_n) = \det(\underline{\mathbf{A}}\,\underline{\mathbf{1}}_n) = \det(\underline{\mathbf{A}})$, which is, in general, not 1.

However, since $\underline{\mathbf{A}}$ is invertible, it follows from Corollary 11.9 that $\det(\underline{\mathbf{A}}) \neq 0$.

Thus

$$\tilde{F}\colon \mathbf{M}(n;\mathbb{F}) \longrightarrow \mathbb{F}, \quad \underline{\mathbf{B}} \longmapsto \frac{1}{\det(\underline{\mathbf{A}})}F(\underline{\mathbf{B}}) = \frac{1}{\det(\underline{\mathbf{A}})}\det(\underline{\mathbf{A}}\,\underline{\mathbf{B}})$$

satisfies D1, D2 and D3.

By Theorem 11.1 on page 141, $\tilde{F}(\underline{\mathbf{B}}) = \det(\underline{\mathbf{B}})$, that is,

$$\frac{1}{\det(\underline{\mathbf{A}})}\det(\underline{\mathbf{A}}\,\underline{\mathbf{B}}) = \det(\underline{\mathbf{B}})$$

or, equivalently, $\det(\underline{\mathbf{A}}\,\underline{\mathbf{B}}) = \det(\underline{\mathbf{A}})\det(\underline{\mathbf{B}})$. □

Corollary 11.11. *If* $\underline{\mathbf{A}}$ *is invertible, then* $\det(\underline{\mathbf{A}}^{-1}) = (\det(\underline{\mathbf{A}}))^{-1}$.

Proof. Let $\underline{\mathbf{A}}$ be an invertible $n \times n$ matrix. Then $\underline{\mathbf{A}}^{-1}\underline{\mathbf{A}} = \underline{\mathbf{1}}_n$.

Hence, by Corollary 11.10 and by D3, $\det(\underline{\mathbf{A}}^{-1})\det(\underline{\mathbf{A}}) = \det(\underline{\mathbf{1}}_n) = 1$. □

Corollary 11.12. *If the matrices* $\underline{\mathbf{A}}$ *and* $\underline{\mathbf{B}}$ *represent the same endomorphism,* $T\colon V \longrightarrow V$*, then* $\det(\underline{\mathbf{A}}) = \det(\underline{\mathbf{B}})$.

Proof. $\underline{\mathbf{A}}$ and $\underline{\mathbf{B}}$ represent the same endomorphism if and only if there is an invertible matrix, $\underline{\mathbf{C}}$, with $\underline{\mathbf{B}} = \underline{\mathbf{C}}^{-1}\underline{\mathbf{A}}\,\underline{\mathbf{C}}$. The conclusion now follows by Corollaries 11.10 and 11.11. □

Corollary 11.12 allows us to define the deteminant of an endomorphism.

Definition 11.13. Let $T\colon V \longrightarrow V$ be an endomorphism of the finitely generated vector space, V. The *determinant of* T, $\det(T)$, is the determinant of any matrix representing T.

11.2 Applications of the Determinant

Our discussion of the determinant focussed on its definition and principal properties, without regard to its applications. We now turn to two applications which require no further theory. We shall meet other applications later.

11.2.1 When Do $\mathbf{v}_1, \ldots, \mathbf{v}_n \in V$ Comprise a Basis?

Suppose given the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ in the vector space V.

By Theorem 8.3 and Definition 8.1 on page 87, a necessary (but not sufficient) condition for these vectors to comprise a basis for V is that $\dim_{\mathbb{F}}(V) = n$.

If $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ is a basis for V, then we can use the determinant function to decide whether $\mathbf{v}_1, \ldots, \mathbf{v}_n \in V$ comprise a basis for V.

We have seen that the choice of a basis for V is the choice of an isomorphism $V \longrightarrow \mathbb{F}_{(n)}$, with the vector $\mathbf{v} \in V$ being mapped to its co-ordinate vector with respect to the chosen basis.

Let $\mathbf{c}_j$ be the co-ordinate vector of $\mathbf{v}_j$ with respect to the basis $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$.

The vector subspace of V generated by $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is then mapped to the vector subspace of $\mathbb{F}_{(n)}$ generated by $\mathbf{c}_1, \ldots, \mathbf{c}_n$. But this is precisely the column space of the matrix $\underline{\mathbf{A}}$ whose j^{th} column is $\mathbf{c}_j$, and so $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ is a basis for V if and only if the column space of $\underline{\mathbf{A}}$ is $\mathbb{F}_{(n)}$.

This is the case if and only if $\mathrm{rk}(\underline{\mathbf{A}}) = n$, which is equivalent to $\det(\underline{\mathbf{A}}) \neq 0$.

The next theorem summarises the above.

Theorem 11.14. *Let $\mathcal{B} = \{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ be a basis for the vector space V. Take $\mathbf{v}_1, \ldots, \mathbf{v}_n \in V$. Let $\mathbf{c}_j$ be the co-ordinate vector of $\mathbf{v}_j$ with respect to $\mathcal{B}$ and $\underline{\mathbf{A}}$ be the matrix with $\mathbf{c}_j$ as j^{th} column. Then $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is a basis for V if and only if $\det(\underline{\mathbf{A}}) \neq 0$.*

Example 11.15. By the theory of linear differentiable functions with constant coefficients (cf. MATH102), the functions

$$\sin\colon \mathbb{R} \longrightarrow \mathbb{R}, \quad x \longmapsto \sin x$$
$$\cos\colon \mathbb{R} \longrightarrow \mathbb{R}, \quad x \longmapsto \cos x$$

comprise a basis for the real vector space

$$V := \{f\colon \mathbb{R} \longrightarrow \mathbb{R} \mid \tfrac{d^2 f}{dx^2} + f = 0\}$$

It is easy to see that

$$f_1\colon \mathbb{R} \longrightarrow \mathbb{R}, \quad x \longmapsto \cos(x + \tfrac{\pi}{4})$$
$$f_2\colon \mathbb{R} \longrightarrow \mathbb{R}, \quad x \longmapsto \cos(x + \tfrac{\pi}{3})$$

are vectors in V.

Since

$$\cos(x + \tfrac{\pi}{4}) = \cos x \cos \tfrac{\pi}{4} - \sin x \sin \tfrac{\pi}{4} = \tfrac{1}{\sqrt{2}} \cos x - \tfrac{1}{\sqrt{2}} \sin x$$
$$\cos(x + \tfrac{\pi}{3}) = \cos x \cos \tfrac{\pi}{3} - \sin x \sin \tfrac{\pi}{3} = \tfrac{\sqrt{3}}{2} \cos x - \tfrac{1}{2} \sin x$$

the co-ordinate vectors of f_1 and f_2 with respect to our chosen basis, $\{\sin, \cos\}$, are

$$\begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} \frac{\sqrt{3}}{2} \\ -\frac{1}{2} \end{bmatrix}$$

respectively. Since

$$\det\left(\begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{\sqrt{3}}{2} \\ -\frac{1}{\sqrt{2}} & -\frac{1}{2} \end{bmatrix}\right) = \frac{1}{\sqrt{2}}(-\frac{1}{2}) - (-\frac{1}{\sqrt{2}})\frac{\sqrt{3}}{2} = \frac{\sqrt{3}-1}{2\sqrt{2}} \neq 0,$$

we see that $\{f_1, f_2\}$ is also a basis for V.

Observation 11.16. The determinant to decide whether given vectors form a basis depended on having available basis for the vector space in question. When there is an obvious basis, such as in the case of $V = \mathbb{F}^n$, or when there is a well-known basis, as in our example, this is convenient.

If there is no convenient basis at hand, it is usually easier to settle the question by other means.

11.2.2 Fitting Curves to Given Points

Take distinct points $(x_1, y_1), \ldots (x_n, y_n) \in \mathbb{F}^2$.

Can we find a function $f\colon \mathbb{F} \longrightarrow \mathbb{F}$ with $f(x_j) = y_j \quad (1 \leq j \leq n)$?

This is equivalent to finding a function whose graph passes through the given points.

There is an obvious necessary condition, namely, that if $x_i = x_j$, then $y_i = y_j$, for if $f(x) = y$ and $f(x) = \tilde{y}$, Definition 1.9 on page 4 forces $\tilde{y} = y$.

So we may restrict attention to the case $x_i = x_j$ if and only if $i = j$.

We can be more specific, by seeking a function which is easily computable, such as a polynomial function. We formulate our problem accordingly.

Is there a polynomial, $p(t) = a_0 + a_1 t + \cdots + a_{n-1}t^{n-1} \in \mathbb{F}[t]$ with the function

$$f_p\colon \mathbb{F} \longrightarrow \mathbb{F}, \quad x \longmapsto a_0 + a_1 x + \cdots + a_{n-1}x^{n-1}$$

having the property that, for $1 \leq j \leq n$,

$$f_p(x_j) = y_j$$

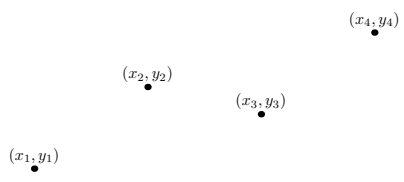
This condition is expressed by the system of linear equations

$$\begin{array}{ccccccccc} a_0x_1 & + & a_1x_1 & + & \cdots & + & a_{n-1}x_1^{n-1} & = & y_1 \\ \vdots & & & & & & \vdots & & \vdots \\ a_0x_n & + & a_1x_n & + & \cdots & + & a_{n-1}x_n^{n-1} & = & y_n \end{array}$$

As we are seeking $a_0, \ldots a_{n-1}$ which simultaneously satisfy these equations, we represent them by the matrix equation

$$\begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \cdots & x_2^{n-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^{n-1} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_{n-1} \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

Example 11.17. Take $\mathbb{F} = \mathbb{R}$ and $(-1, -15), (0, 3), (1, -3), (2, 15) \in \mathbb{R}^2$.



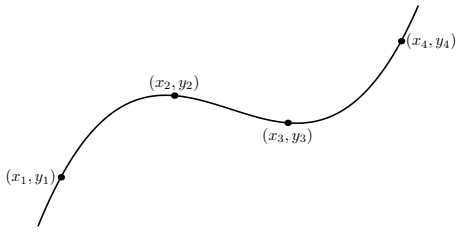
We seek $p(t) = a_0 + a_1 t + a_2 t^2 + a_3 t^3 \in \mathbb{R}[t]$ satisfying

$$\begin{bmatrix} 1 & -1 & 1 & -1 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 2 & 4 & 8 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} -15 \\ 3 \\ -3 \\ 15 \end{bmatrix}$$

It is left to the reader to verify that the polynomial we obtain is $p(t) = t^3 - 12t^2 - 2t + 3$.

Our curve is thus the graph of

$$f_p \colon \mathbb{R} \longrightarrow \mathbb{R}, \quad x \longmapsto x^3 - 12x^2 - 2x + 3$$



11.3 The Trace

The trace of a matrix complements the determinant, in a sense. For while the determinant of a product of matrices is the product of the their determinants and there is no general relationship between the determinant of a sum of matrices and the individual determinants, the opposite is true of the trace: the trace of a sum of matrices is the sum of their traces, but there is no general relationship between the trace of a product of matricesand the individual traces.

Definition 11.18. The *trace* of an $n \times n$ matrix is the sum of its diagonal coefficients.

$$\mathrm{tr} \colon \mathbf{M}(n; \mathbb{F}) \longrightarrow \mathbb{F}, \quad \underline{\mathbf{A}} = [a_{ij}]_{n \times n} \longmapsto \mathrm{tr}(\underline{\mathbf{A}}) = \sum_{j=1}^{n} a_{jj}.$$

The central properties of the trace are contained in our next theorem.

Theorem 11.19. *Take* $\underline{\mathbf{A}}, \underline{\mathbf{B}} \in \mathbf{M}(n; \mathbb{F})$. *Then*

(i) $\mathrm{tr}(\underline{\mathbf{A}} + \underline{\mathbf{B}}) = \mathrm{tr}(\underline{\mathbf{A}}) + \mathrm{tr}(\underline{\mathbf{B}})$

(ii) $\mathrm{tr}(\underline{\mathbf{A}}\,\underline{\mathbf{B}}) = \mathrm{tr}(\underline{\mathbf{B}}\,\underline{\mathbf{A}})$.

A word of warning before we prove the theorem. It is important to avoid drawing the tempting, but false, conclusion from Theorem 11.19 (ii) that there is some regular relationship between $\mathrm{tr}(\underline{\mathbf{A}}\,\underline{\mathbf{B}})$ on the one hand, and $\mathrm{tr}(\underline{\mathbf{A}})$ and $\mathrm{tr}(\underline{\mathbf{B}})$ on the other.

Example 11.20. Put

$$\underline{\mathbf{A}} = \underline{\mathbf{B}} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

Plainly $\mathrm{tr}(\underline{\mathbf{A}}) = \mathrm{tr}(\underline{\mathbf{B}}) = 0$, but $\mathrm{tr}(\underline{\mathbf{A}}\,\underline{\mathbf{B}}) = 2$.

We now prove Theorem 11.19.

Proof. Take $\underline{\mathbf{A}} = [a_{ij}], \underline{\mathbf{B}} = [b_{ij}] \in \mathbf{M}(n; \mathbb{F})$.

(i) Since $\underline{\mathbf{A}} + \underline{\mathbf{B}} := [a_{ij} + b_{ij}]$,

$$\begin{aligned}\operatorname{tr}(\underline{\mathbf{A}} + \underline{\mathbf{B}}) &= \sum_{j=1}^{n}(a_{jj} + b_{jj}) \\ &= \sum_{j=1}^{n} a_{jj} + \sum_{j=1}^{n} b_{jj} \\ &= \operatorname{tr}(\underline{\mathbf{A}}) + \operatorname{tr}(\underline{\mathbf{B}}).\end{aligned}$$

(ii) Since $\underline{\mathbf{A}}\,\underline{\mathbf{B}} = [\sum_{j=1}^{n} a_{ij}b_{jk}]$ and $\underline{\mathbf{B}}\,\underline{\mathbf{A}} = [\sum_{j=1}^{n} b_{ij}a_{jk}]$,

$$\begin{aligned}\operatorname{tr}(\underline{\mathbf{A}}\,\underline{\mathbf{B}}) &= \sum_{j=1}^{n}\sum_{k=1}^{n} a_{jk}b_{kj} \\ &= \sum_{j=1}^{n}\sum_{k=1}^{n} b_{kj}a_{jk} \\ &= \sum_{k=1}^{n}\sum_{j=1}^{n} a_{jk}b_{kj} \\ &= \operatorname{tr}(\underline{\mathbf{B}}\,\underline{\mathbf{A}}).\end{aligned}$$

□

Exercise 11.3 on page 151 extends Theorem 11.19 (ii).

Corollary 11.21. *If* $\underline{\mathbf{C}}$ *is invertible, then* $\operatorname{tr}(\underline{\mathbf{C}}^{-1}\underline{\mathbf{A}}\,\underline{\mathbf{C}}) = \operatorname{tr}\underline{\mathbf{A}}$.

Proof.

$$\begin{aligned}\operatorname{tr}(\underline{\mathbf{C}}^{-1}\underline{\mathbf{A}}\,\underline{\mathbf{C}}) &= \operatorname{tr}(\underline{\mathbf{C}}^{-1}(\underline{\mathbf{A}}\,\underline{\mathbf{C}})) \\ &= \operatorname{tr}((\underline{\mathbf{A}}\,\underline{\mathbf{C}})\underline{\mathbf{C}}^{-1}) \\ &= \operatorname{tr}(\underline{\mathbf{A}}(\underline{\mathbf{C}}\,\underline{\mathbf{C}}^{-1})) && \text{by Theorem 11.19} \\ &= \operatorname{tr}(\underline{\mathbf{A}}\,\underline{\mathbf{1}}_n) \\ &= \operatorname{tr}(\underline{\mathbf{A}}).\end{aligned}$$

□

Corollary 11.22. *If the matrices* $\underline{\mathbf{A}}$ *and* $\underline{\mathbf{B}}$ *represent the same endomorphism* $T\colon V \longrightarrow V$, *then* $\operatorname{tr}(\underline{\mathbf{A}}) = \operatorname{tr}(\underline{\mathbf{B}})$.

Corollary 11.22 allows us to define the trace of an endomorphism of finitely generated vector spaces.

Definition 11.23. Let $T\colon V \longrightarrow V$ be an endomorphism of the finitely generated vector space V. The *trace of* T, $\operatorname{tr}(T)$, is defined to be the trace of any matrix representing T.

11.4 The Transpose of a Matrix

We introduce an important operation on matrices, whose true significance will only become apparent later.

Definition 11.24. The *transpose* of the $m \times n$ matrix $\underline{\mathbf{A}}$ is the $n \times m$ matrix $\underline{\mathbf{A}}^t$ obtained by interchanging each row with the corresponding column.

This defines a function

$$(\quad)^t \colon \mathbf{M}(m \times n; \mathbb{F}) \longrightarrow \mathbf{M}(n \times m; \mathbb{F}), \quad [a_{ij}]_{m\times n} \longmapsto [x_{\nu\mu}]_{n\times m}$$

where $x_{\nu\mu} := a_{\mu\nu} \qquad (1 \leq \nu \leq n,\ 1 \leq \mu \leq n)$.

The next theorems summarises the basic properties of the transpose.

Theorem 11.25. *(i)* $(\quad)^t \colon \mathbf{M}(m \times n; \mathbb{F}) \longrightarrow \mathbf{M}(n \times m; \mathbb{F})$ *is a linear transformation.*

(ii) Given an $m \times n$ *matrix,* $\underline{\mathbf{A}}$, $(\underline{\mathbf{A}}^t)^t = \underline{\mathbf{A}}$.

(iii) Given an $m \times n$ *matrix,* $\underline{\mathbf{A}}$, *and an* $n \times p$ *matrix,* $\underline{\mathbf{B}}$, $(\underline{\mathbf{A}}\,\underline{\mathbf{B}})^t = \underline{\mathbf{B}}^t\underline{\mathbf{A}}^t$.

Proof. The assertions follow by direct calculations, which are left as exercises. □

Observation 11.26. It follows from Theorem 11.25(ii) that $(\quad)^t \colon \mathbf{M}(m\times n; \mathbb{F}) \longrightarrow \mathbf{M}(n\times m; \mathbb{F})$ is actually an isomorphism.

Theorem 11.27. *Let* $\underline{\mathbf{A}}$ *be an* $n \times n$ *matrix. Then*

(i) $\operatorname{tr}(\underline{\mathbf{A}}^t) = \operatorname{tr}(\underline{\mathbf{A}})$;

(ii) $\det(\underline{\mathbf{A}}^t) = \det(\underline{\mathbf{A}})$.

Proof. **(i)** : The assertion follows immediately from the fact that the diagonal of a matrix is unchanged by taking the transpose.

(ii): By the Main Theorem on Determinants (Theorem 11.1 on page 141), there is a unique function, $D \colon \mathbf{M}(n; \mathbb{F}) \longrightarrow \mathbb{F}$, which is linear in the rows of a matrix, which takes the value 0 on matrices whose rank is less than n, and which takes the value 1 on $\underline{\mathbf{1}}_n$.

Consider

$$F \colon \mathbf{M}(n; \mathbb{F}) \longrightarrow \mathbb{F}, \quad \underline{\mathbf{A}} \longmapsto \det\left(\underline{\mathbf{A}}^t\right).$$

Since the row and column ranks of a matrix agree, and since the identity matrix is its own transpose, D2 and D3 are clearly satisfied by F.

It remains only to establish D1 for F. Since the rows of $\underline{\mathbf{A}}^t$ are the columns of $\underline{\mathbf{A}}$, and since det is linear in each row, this is equivalent to proving that the determinant is linear in each column.

To verify the linearity of det in the j^{th} oclumn of $\underline{\mathbf{A}}$, recall from Observation 11.8 on page 144 that $\det(\underline{\mathbf{A}})$ can be calculated using the expansion by any column.

The expansion of the determinant of $\underline{\mathbf{A}}$ by the j^{th} column of $\underline{\mathbf{A}}$, $\det(\underline{\mathbf{A}})$ is

$$\det(\underline{\mathbf{A}}) = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} \det\left(\underline{\mathbf{A}}_{(i)(j)}\right),$$

which is linear in the j^{th} column of $\underline{\mathbf{A}}$, as $\det(A_{(i)(j)})$ is independent of the j^{th} column of $\underline{\mathbf{A}}$. □

11.5 Exercises

Exercise 11.1. Let $\mathcal{P}_2$ denote the real vector space of all real polynomials of degree at most 2. Let $D : \mathcal{P}_2 \longrightarrow \mathcal{P}_2$ be differentiation.

Find the determinant of D.

Exercise 11.2. Let $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ be a basis for the vector space V.

Take $\mathbf{v}_1, \ldots, \mathbf{v}_n \in V$ an let

$$\begin{bmatrix} a_{1j} \\ \vdots \\ a_{nj} \end{bmatrix}$$

be the coordinate vector of $\mathbf{v}_j$ $(j = 1, \ldots, n)$ with respect to the basis $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$.

Prove that $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is a basis for V if and only if $\det \underline{\mathbf{A}} \neq 0$, where $\underline{\mathbf{A}} := [a_{ij}]_{n \times n}$.

Exercise 11.3. Take $\underline{\mathbf{A}} \in \mathbf{M}(m \times n; \mathbb{F})$ and $\underline{\mathbf{B}} \in \mathbf{M}(n \times m; \mathbb{F})$.

Prove that $\operatorname{tr}(\underline{\mathbf{A}}\,\underline{\mathbf{B}}) = \operatorname{tr}(\underline{\mathbf{B}}\,\underline{\mathbf{A}})$.

Exercise 11.4. Suppose that $\underline{\mathbf{A}} \in \mathbf{M}(n; \mathbb{F})$ can be written in the form

$$\underline{\mathbf{A}} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \begin{bmatrix} y_1 & \cdots & y_n \end{bmatrix}$$

Show that $\underline{\mathbf{A}}^{r+1} = (\operatorname{tr} \underline{\mathbf{A}})^r \, \underline{\mathbf{A}}$ for all $r \geq 1$, and find $\det(\underline{\mathbf{A}}^r)$.

Exercise 11.5. Prove Theorem 11.25.

Exercise 11.6. Take $x_1, \ldots, x_n \in \mathbb{F}$.

Show that the determinant of the $n \times n$ matrix

$$\begin{bmatrix} 1 & x_1 & \cdots & x_1^{n-1} \\ \vdots & & & \vdots \\ 1 & x_n & \cdots & x_n^{n-1} \end{bmatrix}$$

is

$$\prod_{1 \leq i < j \leq n} (x_j - x_i)$$

Many who have had an opportunity of knowing any more about mathematics confuse it with arithmetic, and consider it an arid science. In reality, however, it is a science which requires a great amount of imagination.

Sofia Kovalevskaya

Chapter 12

Eigenvalues and Eigenvectors

The direct sum of the vector spaces $V_1, \ldots, V_n$, $V_1 \oplus \cdots \oplus V_n$ is defined by

$$\bigoplus_{j=1}^{n} V_j = \{(\mathbf{v}_1, \ldots, \mathbf{v}_n) \mid \mathbf{v}_j \in V_j,\ j = 1, \ldots, n\}$$

with the vector spaces operations are defined "componentwise":

$$\begin{aligned}(\mathbf{v}_1, \ldots, \mathbf{v}_n) + (\mathbf{v}'_1, \ldots, \mathbf{v}'_n) &:= (\mathbf{v}_1 + \mathbf{v}'_1, \ldots, \mathbf{v}_n + \mathbf{v}'_n)\\ \lambda(\mathbf{v}_1, \ldots, \mathbf{v}_n) &:= (\lambda\mathbf{v}_1, \ldots, \lambda\mathbf{v}_n),\end{aligned}$$

It follows that $\mathbf{0}_{V_1 \oplus \cdots \oplus V_n} = (\mathbf{0}_{V_1}, \ldots, \mathbf{0}_{V_n})$ and $-(\mathbf{v}_1, \ldots, \mathbf{v}_n) = (-\mathbf{v}_1, \ldots, -\mathbf{v}_n)$

Since $\mathbb{F}^n = \bigoplus_{j=1}^{n} \mathbb{F}$, we can reformulate the Classification Theorem for Finitely Generated Vector Spaces over the field $\mathbb{F}$ as stating that every such vector space is (up to isomorphism) a direct sum of copies of $\mathbb{F}$,

$$V \cong \bigoplus_{j=1}^{n} \mathbb{F},$$

where n is the dimension of the vector space in question.

We cannot decompose this further as a direct sum, because $\mathbb{F}$ itself cannot be written as a direct sum of non-trivial vector spaces over $\mathbb{F}$.

What we have achieved is a decomposition of the finitely generated vector space, V, into finitely many components, none of which can be so decomposed further.

The direct sum construction also applies to linear transformations. The direct sum of the linear transformations $T_j \colon V_j \longrightarrow W_j$ $(j = 1, \ldots, n)$ is

$$\bigoplus_{j=1}^{n} T_j \colon \bigoplus_{j=1}^{n} V_j \longrightarrow \bigoplus_{j=1}^{n} W_j$$

defined by

$$(T_1 \oplus \cdots \oplus T_n)(\mathbf{v}_1, \ldots \mathbf{v}_n) := (T_1(\mathbf{v}_1), \ldots, T_n(\mathbf{v}_n))$$

The verification that $(T_1 \oplus \cdots \oplus T_n)$ is a linear transformation is routine, and left to the reader.

Example 12.1. For

$$R\colon \mathbb{R}^2 \longrightarrow \mathbb{R}^2, \quad (x,y) \longmapsto (2x+y, 3y)$$
$$T\colon \mathbb{R}^3 \longrightarrow \mathbb{R}, \quad (u,v,w) \longmapsto u+v+w$$

$$R \oplus T\colon \mathbb{R}^2 \oplus \mathbb{R}^3 \longrightarrow \mathbb{R}^2 \oplus \mathbb{R}, \quad \big((x,y),(u,v,w)\big) \longmapsto \big((2x+y,3y),u+v+w\big)$$

We may identify $\mathbb{R}^2 \oplus \mathbb{R}^3$ with $\mathbb{R}^5$ and $\mathbb{R}^2 \oplus \mathbb{R}$ with $\mathbb{R}^3$, using the obvious isomorphisms

$$\mathbb{R}^2 \oplus \mathbb{R}^3 \longrightarrow \mathbb{R}^5, \quad \big((x,y),(u,v,w)\big) \longmapsto (x,y,u,v,w)$$
$$\mathbb{R}^2 \oplus \mathbb{R} \longrightarrow \mathbb{R}^3, \quad \big((r,s),t\big) \longmapsto (r,s,t)$$

Using these identifications, we may regard $R \oplus T$ as the linear transformation

$$\mathbb{R}^5 \longrightarrow \mathbb{R}^3, \quad (x,y,u,v,w) \longmapsto (2x+y, 3y, u+v+w)$$

The question arises:

> **Given a finitely generated vector space V over $\mathbb{F}$, is each linear transformation $T\colon V \longrightarrow V$ the direct sum of linear transformations $T_j\colon \mathbb{F} \longrightarrow \mathbb{F}$ $(j = 1, \ldots, \dim_{\mathbb{F}}(V))$?**

This is the question we pursue here.

Let $\dim(V) = m$ and $\dim(W) = n$. Take endomorphisms $R\colon V \longrightarrow V$ and $S\colon W \longrightarrow W$.

Lemma 12.2. *Let $\{\mathbf{e}_1, \ldots \mathbf{e}_m\}$ be a basis for V and $\{\mathbf{f}_1, \ldots, \mathbf{f}_n\}$ for W. Then putting*

$$\mathbf{u}_i := \begin{cases} (\mathbf{e}_i, \underline{\mathbf{0}}_W) & \text{if } i \le m \\ (\underline{\mathbf{0}}_V, \mathbf{f}_{i-m}) & \text{if } i > m \end{cases}$$

defines a basis, $\{\mathbf{u}_1, \ldots, \mathbf{u}_{m+n}\}$, for $V \oplus W$.

Proof. For $\mathbf{x} \in V \oplus W$, there are unique $\mathbf{v} \in V, \mathbf{w} \in W$ with $\mathbf{x} = (\mathbf{v}, \mathbf{w})$.
Since $\{\mathbf{e}_1, \ldots \mathbf{e}_m\}$ is a basis for V and $\{\mathbf{f}_1, \ldots, \mathbf{f}_n\}$ is a basis for W, there are unique scalars $\alpha_i, \beta_j \in \mathbb{F}$ $(1 \le i \le m,\ 1 \le j \le n)$ with $\mathbf{v} = \sum_{i=1}^{m} \alpha_i \mathbf{e}_i$ and $\mathbf{w} = \sum_{j=1}^{n} \beta_j \mathbf{f}_j$, so that

$$\begin{aligned} \mathbf{x} &= (\mathbf{v}, \mathbf{w}) \\ &= (\sum_{i=1}^{m} \alpha_i \mathbf{e}_i, \sum_{j=1}^{n} \beta_j \mathbf{f}_j) \\ &= (\sum_{i=1}^{m} \alpha_i \mathbf{e}_i, \underline{\mathbf{0}}_W) + (\underline{\mathbf{0}}_V, \sum_{j=1}^{n} \beta_j \mathbf{f}_j) \\ &= \sum_{i=1}^{m} \alpha_i (\mathbf{e}_i, \underline{\mathbf{0}}_W) + \sum_{j=1}^{n} \beta_j (\underline{\mathbf{0}}_V, \mathbf{f}_j) \\ &= \sum_{k=1}^{m+n} \gamma_k \mathbf{u}_k, \end{aligned}$$

where the coefficients $\gamma_k = \begin{cases} \alpha_k & \text{if } 1 \le k \le m \\ \beta_{k-m} & \text{if } m < k \le m+n \end{cases}$ are uniquely determined. □

Corollary 12.3. *If the matrix of R with respect to $\{\mathbf{e}_i\}$ is $\underline{\mathbf{A}}$ and that of S with respect to $\{\mathbf{f}_j\}$ is $\underline{\mathbf{B}}$, then the matrix of $R \oplus S$ with respect to $\{\mathbf{u}_k\}$ is*

$$\underline{\mathbf{A}} \oplus \underline{\mathbf{B}} := \begin{bmatrix} \underline{\mathbf{A}} & \underline{\mathbf{0}} \\ \underline{\mathbf{0}} & \underline{\mathbf{B}} \end{bmatrix}$$

Proof. Exercise. □

Convention. The meaning of $\underline{\mathbf{A}} \oplus \underline{\mathbf{B}} = \begin{bmatrix} \underline{\mathbf{A}} & \underline{\mathbf{0}} \\ \underline{\mathbf{0}} & \underline{\mathbf{B}} \end{bmatrix}$ needs clarification.

This should not be read here as a matrix of matrices, that is a 2×2 matrix, each of whose coefficients is itself a matrix, even though it is possible to do so sensibly.

Rather, what is intended is that if $\underline{\mathbf{A}}$ is an $m \times n$ matrix and $\underline{\mathbf{B}}$ is a $p \times q$ matrix, then $\begin{bmatrix} \underline{\mathbf{A}} & \underline{\mathbf{0}} \\ \underline{\mathbf{0}} & \underline{\mathbf{B}} \end{bmatrix}$ is the $(m+q) \times (n+q)$ matrix obtained by copying the coefficients of $\underline{\mathbf{A}}$ into the top left, those of $\underline{\mathbf{B}}$ into the bottom right and placing 0s everywhere else.

Explicitly, if $\underline{\mathbf{A}} = [a_{ij}]_{m \times n}$ and $\underline{\mathbf{B}} = [b_{k\ell}]_{p \times q}$, then

$$\underline{\mathbf{A}} \oplus \underline{\mathbf{B}} = [c_{rs}]_{(m+p)\times(n+q)}$$

where

$$c_{rs} = \begin{cases} a_{rs} & 1 \le r \le m \text{ and } 1 \le s \le n \\ 0 & 1 \le r \le m \text{ and } n+1 \le s \le n+q \\ 0 & m+1 \le r \le m+p \text{ and } 1 \le s \le n \\ b_{(r-m)(s-n)} & m+1 \le r \le m+p \text{ and } n+1 \le s \le n+p \end{cases}$$

Example 12.4. Take $\underline{\mathbf{A}} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ and $\underline{\mathbf{B}} = \begin{bmatrix} e & f & g \\ h & j & k \end{bmatrix}$. Then

$$\begin{bmatrix} \underline{\mathbf{A}} & \underline{\mathbf{0}} \\ \underline{\mathbf{0}} & \underline{\mathbf{B}} \end{bmatrix} = \begin{bmatrix} a & b & 0 & 0 & 0 \\ c & d & 0 & 0 & 0 \\ 0 & 0 & e & f & g \\ 0 & 0 & h & j & gk \end{bmatrix} \neq \begin{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} & \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \\ \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} & \begin{bmatrix} e & f & g \\ h & j & k \end{bmatrix} \end{bmatrix}$$

The direct summ construction can be generalised to any finite number of vector spaces, $V_1, \ldots, V_n$ and endomorphisms $T_j \colon V_j \longrightarrow V_j$ $(j = 1, \ldots, k)$.

In particular, if each V_j is 1-dimensional — equivalently, if each $V_j \cong \mathbb{F}$ — then the matrix, $[a_{ij}]_{n \times n}$, of $T = \oplus T_j \colon \oplus V_j \longrightarrow \oplus V_j$ with respect to the canonically induced basis is a diagonal matrix: $a_{ij} = 0$, whenever $i \neq j$.

In other words, the endomorphism $T \colon V \longrightarrow V$ is of the form $T_1 \oplus \cdots \oplus T_{\dim(V)}$, with each T_j a linear transformation $T_j \colon \mathbb{F} \longrightarrow \mathbb{F}$ if and only if there is a basis for V with respect to which the matrix of T is a diagonal matrix.

Thus we may reformulate our question as:

Is there a basis for V with respect to which the matrix of T is in diagonal form?

Note that if $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ is a basis for V with respect to which the matrix of $T\colon V \longrightarrow V$ is in diagonal form, then

$$T(\mathbf{e}_j) = \lambda_j \mathbf{e}_j,$$

where λ_j is the j-th diagonal entry in the matrix of T with respect to $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$.

Since scalars λ_j and vectors, $\mathbf{e}_j$ with this property play an important rôle in the study of endomorphisms and are central to many applications of linear algebra, special terminology has been introduced for them.

Definition 12.5. The scalar $\lambda \in \mathbb{F}$ is an *eigenvalue* of the endomorphism $T\colon V \longrightarrow V$ if and only if there is a non-zero vector $\mathbf{v} \in V$ such that

$$T(\mathbf{v}) = \lambda \mathbf{v}. \tag{12.1}$$

Such a vector $\mathbf{v}$ is an *eigenvector for the eigenvalue* λ. The *eigenspace* of λ, V_λ, is the set of all solutions of $T(\mathbf{v}) = \lambda \mathbf{v}$, that is

$$V_\lambda := \{\mathbf{x} \in V \mid T(\mathbf{x}) = \lambda \mathbf{x}\} \qquad (\lambda \in \mathbb{F}).$$

Observation 12.6. Eigenvalues and eigenvectors are sometimes called *characteristic values* and *characteristic vectors* respectively.

The next theorem, the main theorem of this section, summarises the preceding discussion.

Theorem 12.7 (**Main Theorem on Endomorphisms**). *If* $\dim(V) = n$*, then* $T\colon V \longrightarrow V$ *is the direct sum of endomorphisms* $T_i\colon V_i \longrightarrow V_i$*, with* $\dim(V_i) = 1$*, if and only if* V *has a basis consisting of eigenvectors of* T*.*

We discuss related results of independent interest.

Theorem 12.8. *Take an endomorphism* $T\colon V \longrightarrow V$*. Then for each* $\lambda \in \mathbb{F}$

$$V_\lambda := \{\mathbf{v} \in V \mid T(\mathbf{v}) = \lambda \mathbf{v}\}$$

is a vector subspace of V*, and* λ *is an eigenvalue for* T *if and only if* $V_\lambda \neq \{\underline{\mathbf{0}}_V\}$*.*

Proof. Take $\mathbf{u}, \mathbf{v} \in V_\lambda$ and $\alpha, \beta \in \mathbb{F}$. Then

$$\begin{aligned} T(\alpha\mathbf{u} + \beta\mathbf{v}) &= \alpha T(\mathbf{u}) + \beta T(\mathbf{v}) \\ &= \alpha\lambda\mathbf{u} + \beta\lambda\mathbf{v} \\ &= \lambda(\alpha\mathbf{u} + \beta\mathbf{v}). \end{aligned}$$

Thus, $\alpha\mathbf{u} + \beta\mathbf{v} \in V_\lambda$. □

Example 12.9. Let V be a vector space and take the identity morphism on V

$$id_V\colon V \longrightarrow V, \quad \mathbf{v} \longmapsto \mathbf{v}$$

Then $T(\mathbf{v}) = \mathbf{v}$ for every $\mathbf{v} \in V$, so that 1 is the only eigenvalue and every non-zero vector is an eigenvector for 1.

Example 12.10. Let V be a vector space and take the zero linear transformation

$$0\colon V \longrightarrow V, \quad \mathbf{v} \longmapsto \mathbf{0}_V$$

Then $T(\mathbf{v}) = \underline{\mathbf{0}}_V = 0\mathbf{v}$ for every $\mathbf{v} \in V$. Plainly, 0 is the only possible eigenvalue and every non-zero vector is an eigenvector for 0.

Observation 12.11. The eigenvalue 0 plays a distinguished role, for, plainly, $V_0 = \ker(T)$. This establishes the following lemma.

Lemma 12.12. *Let $T\colon V \longrightarrow V$ be an endomorphism of the vector space V. Then 0 is an eigenvalue if and only if T is not injective.*

Example 12.13. Let V be the Euclidean plane, regarded as $\mathbb{R}^2$.

Rotating the plane through an angle of θ (with $0 \leq \theta < 2\pi$) about the origin defines the linear transformation

$$T_\theta\colon \mathbb{R}^2 \longrightarrow \mathbb{R}^2, \quad (x, y) \longrightarrow (x\cos\theta - y\sin\theta, x\sin\theta + y\cos\theta).$$

The real number λ is then an eigenvalue for T_θ if and only if

$$(x\cos\theta - y\sin\theta, x\sin\theta + y\cos\theta) = (\lambda x, \lambda y)$$

for some $(x, y) \neq (0, 0)$.

Thus, λ is an eigenvalue for T_θ if and only if there are real $x, y \in \mathbb{R}$ with $x^2 + y^2 \neq 0$ such that

$$x\cos\theta - y\sin\theta = \lambda x$$
$$x\sin\theta + y\cos\theta = \lambda y$$

Squaring and adding these equations we see that

$$x^2 + y^2 = \lambda^2(x^2 + y^2).$$

Since $x^2 + y^2 \neq 0$, $\lambda^2 = 1$ and so the only possible eigenvalues are -1 and 1

$\lambda = \mathbf{1}$: Then

$$x\cos\theta - y\sin\theta = x$$
$$x\sin\theta + y\cos\theta = y.$$

By elementary trigonometry, $\theta = 0$, since $(x, y) \neq (0, 0)$.

Thus $T_0 = id_V$, and $V_1 = V$.

$\lambda = -\mathbf{1}$: Then

$$x\cos\theta - y\sin\theta = -x$$
$$x\sin\theta + y\cos\theta = -y.$$

By elementary trigonometry, $\theta = \pi$, since $(x, y) \neq (0, 0)$.

Thus, $T_\pi(x, y) = (-x, -y)$ for all $(x, y) \in V$, and $V_{-1} = V$.

Furthermore, if $\theta \neq 0, \pi$, then T_θ has no real eigenvalues.

Example 12.14. Let V be the Euclidean plane, regarded as $\mathbb{R}^2$. Reflecting the plane in the x-axis defines the linear transformation

$$T\colon \mathbb{R}^2 \longrightarrow \mathbb{R}^2, \quad (x, y) \longmapsto (x, -y).$$

Thus, $\lambda \in \mathbb{R}$ is an eigenvalue for T if and only if $(x, -y) = (\lambda x, \lambda y)$ for some $(x, y) \neq (0, 0)$.

In other words, λ is an eigenvalue for T if and only if there are $x, y \in \mathbb{R}$ with $x^2 + y^2 \neq 0$ and

$$x = \lambda x \qquad -y = \lambda y$$

Thus, $x^2 + y^2 = \lambda^2(x^2 + y^2)$.

Since $x^2 + y^2 \neq 0$, it follows that $\lambda^2 = 1$, so that the only possible eigenvalues are -1 and 1.

$\lambda = \mathbf{1}$: Then $x = x$ and $-y = y$, whence $y = 0$ and x is arbitrary, showing that

$$V_1 = \{(x, 0) \mid x \in \mathbb{R}\}$$

$\lambda = -\mathbf{1}$: Then $x = -x$ and $-y = -y$, whence $x = 0$ and y is arbitrary, showing that

$$V_{-1} = \{(0, y) \mid y \in \mathbb{R}\}$$

We see that $\mathbb{R}^2 = V_1 \oplus V_{-1}$, and $\{(1, 0), (0, 1)\}$ is a basis consisting of eigenvectors for T.

Example 12.15. Let $V = \mathcal{C}^\infty(\mathbb{R}) = \{f : \mathbb{R} \longrightarrow \mathbb{R} \mid f \text{ is infinitely differentiable}\}$. Take

$$T : V \longrightarrow V, \quad f \longmapsto f'',$$

where f'' denotes the second derivative of f.

Then, as is well known -1 is an eigenvalue of T. and

$$f : \mathbb{R} \longrightarrow \mathbb{R}, \quad t \longmapsto \cos t$$
$$g : \mathbb{R} \longrightarrow \mathbb{R}, \quad t \longmapsto \sin t$$

are eigenvectors for -1.

It follows from the general theory of differential equations that they form a basis for V_{-1}.

The details are left as an exercise.

Theorem 12.16. *Let $\mathbf{v}_1$ be an eigenvector of $T : V \longrightarrow V$ for the eigenvalue $\lambda_i, (1 \leq i \leq m)$. If the λ_i are pairwise distinct, then $\mathbf{v}_1, \ldots, \mathbf{v}_m$ are linearly independent.*

Proof. We prove the theorem by induction on m.

$m = 1$: Since $\mathbf{v}_i \neq \mathbf{0}_V$, it is linearly independent.

$m > 1$: Suppose that the theorem is true for m, an let $\mathbf{v}_i$ be an eigenvector for the eigenvalue λ_i of $T : V \longrightarrow V$, with $1 \leq i, j \leq m + 1$ and $\lambda_i = \lambda_j$ if and only if $i = j$.

Then $\displaystyle\sum_{i=1}^{m+1} \alpha_i \mathbf{v}_i = \underline{\mathbf{0}}_V$ if and only if $\displaystyle\sum_{i=1}^{m} \alpha_i \mathbf{v}_i = -\alpha_{m+1}\mathbf{v}_{m+1}$, and, in that case

$$\begin{aligned}
\sum_{i=1}^{m} \alpha_i \lambda_i \mathbf{v}_i &= \sum_{i=1}^{m} \alpha_i T(\mathbf{v}_i) && \text{as } \mathbf{e}_i \text{ is an eigenvector for } \lambda_i \\
&= T(\sum_{i=1}^{m} \alpha_i \mathbf{v}_i) && \text{as } T \text{ is a linear transformation} \\
&= T(-\alpha_{m+1}\mathbf{v}_{m+1}) \\
&= -\alpha_{m+1} T(\mathbf{v}_{m+1}) && \text{as } T \text{ is a linear transformation} \\
&= -\alpha_{m+1}\lambda_{m+1}\mathbf{v}_{m+1} && \text{as } \mathbf{e}_{m+1} \text{ is an eigenvector for } \lambda_{m+1} \\
&= \lambda_{m+1}(-\alpha_{m+1}\mathbf{v}_{m+1}) \\
&= \lambda_{m+1} \sum_{i=1}^{m} \alpha_i \mathbf{v}_i \\
&= \sum_{i=1}^{m} \alpha_i \lambda_{m+1} \mathbf{v}_i
\end{aligned}$$

Hence, $\sum_{i=1}^{m} \alpha_i(\lambda_i - \lambda_{m+1})\mathbf{v}_i = \mathbf{0}_V$

By the inductive hypothesis, $\mathbf{e}_1, \ldots, \mathbf{e}_m$ are linearly independent, whence $\alpha_i(\lambda_i - \lambda_{m+1}) = 0$ for $1 \leq i \leq m$.

Since $\lambda_{m+1} \neq \lambda_i$ for $i < m+1$, it follows that $\alpha_i = 0$ for $i = 1, \ldots, m$.

Then $\alpha_{m+1}\mathbf{v}_{m+1} = -\sum_{i=1}^{m} \alpha_i \mathbf{v}_i = \underline{\mathbf{0}}_V$.

Since $\mathbf{v}_{m+1}$ is an eigenvector to λ_{m+1}, $\mathbf{v}_{m+1} \neq \underline{\mathbf{0}}_V$ and so $\alpha_{m+1} = 0$ as well. □

Corollary 12.17. *$T\colon V \longrightarrow V$ has at most $\dim_{\mathbb{F}}(V)$ distinct eigenvalues.*

Corollary 12.18. *If $T\colon V \longrightarrow V$ has n distinct eigenvalues, then V has a basis consisting of eigenvectors of T.*

Given the significance of eigenvalues and eigenvectors, it would be more than merely convenient to find a practical procedure for determining the eigenvalues of a given endomorphism.

When V is finite dimensional, each endomorphism $T\colon V \longrightarrow V$ has an associated polynomial whose zeroes are precisely the eigenvalues of T, as we now show.

Recall that $\lambda \in \mathbb{F}$ is an eigenvalue for $T\colon V \longrightarrow V$ if and only if the equation

$$T(\mathbf{v}) = \lambda\,\mathbf{v} \qquad (*)$$

has a non-zero solution, $\mathbf{v}$.

Choose a basis for V. Let $\underline{\mathbf{A}}$ be the matrix of T and $\mathbf{x} \in \mathbb{F}_{(n)}$ the co-ordinate vector of $\mathbf{v}$ with respect to this basis, so that $T(\mathbf{v}) = \lambda\mathbf{v}$ if and only if $\underline{\mathbf{A}}\mathbf{x} = \lambda\mathbf{x}$.

Theorem 12.19. *λ is an eigenvalue for T if and only if* $\det(\underline{\mathbf{A}} - \lambda\underline{\mathbf{1}}_n) = 0$.

Proof.

$$\begin{aligned}
\lambda \text{ is and eigenvalue for } T \quad & \text{if and only if } \ T(\mathbf{v}) = \lambda\mathbf{v} \text{ for some } \mathbf{v} \in V,\ \mathbf{v} \neq \underline{\mathbf{0}}_V, \\
& \text{if and only if } \ \underline{\mathbf{A}}\,\mathbf{x} = \lambda\mathbf{x} \text{ for some } \mathbf{x} \in \mathbb{F}_{(n)},\ \mathbf{x} \neq \mathbf{0}, \\
& \text{if and only if } \ (\underline{\mathbf{A}} - \lambda\underline{\mathbf{1}}_n)\mathbf{x} = \mathbf{0} \text{ for some } \mathbf{x} \in \mathbb{F}_{(n)},\ \mathbf{x} \neq \mathbf{0}, \\
& \text{if and only if } \ \mathrm{rk}(\underline{\mathbf{A}} - \lambda\underline{\mathbf{1}}_n) < n, \\
& \text{if and only if } \ \det(\underline{\mathbf{A}} - \lambda\underline{\mathbf{1}}_n) = 0.
\end{aligned}$$

□

Since the determinant of $\underline{\mathbf{A}} - \lambda\underline{\mathbf{1}}_n$ is a polynomial function of λ, Theorem 12.19 provides for each endomorphism T a concrete polynomial in λ whose zeroes are precisely the eigenvalues the eigenvalues of T. This polynomial appears to be dependent on the basis chosen.

Fortunately, this is a case where appearances are deceptive. For if $\underline{\mathbf{B}}$ is the matrix of T with respect to another basis, then there is an invertible matrix $\underline{\mathbf{M}}$ such that $\underline{\mathbf{B}} = \underline{\mathbf{M}}\,\underline{\mathbf{A}}\,\underline{\mathbf{M}}^{-1}$. But then

$$\begin{aligned}
\det(\underline{\mathbf{B}} - \lambda\underline{\mathbf{1}}_n) &= \det(\underline{\mathbf{M}}\,\underline{\mathbf{A}}\,\underline{\mathbf{M}}^{-1} - \lambda\underline{\mathbf{M}}\,\underline{\mathbf{M}}^{-1}) \\
&= \det(\underline{\mathbf{M}}(\underline{\mathbf{A}} - \lambda\underline{\mathbf{1}}_n)\underline{\mathbf{M}}^{-1}) \\
&= \det(\underline{\mathbf{M}})\det(\underline{\mathbf{A}} - \lambda\underline{\mathbf{1}}_n)\det(\underline{\mathbf{M}}^{-1}) \\
&= \det(\underline{\mathbf{A}} - \lambda\underline{\mathbf{1}}_n) \qquad \text{as} \quad \det(\underline{\mathbf{M}}^{-1}) = (\det(\underline{\mathbf{M}}))^{-1}.
\end{aligned}$$

Thus the polynomial does not depend on the basis chosen.

Definition 12.20. Let T be an endomorphism of the n-dimensional vector space V. Then

$$\chi_T(t) := \det(T - t\,id_V) = \det(\underline{\mathbf{A}} - t\,\underline{\mathbf{1}}_n) =: \chi_{\underline{\mathbf{A}}}(t)$$

is the *characteristic polynomial* of T and of $\underline{\mathbf{A}}$, where $\underline{\mathbf{A}}$ is any matrix representing T.

The eigenvalues of T are the zeroes of the characteristic polynomial, or, equivalently, the solutions of the *characteristic equation*, $\chi_T(t) = 0$.

We define eigenvalues, eigenvectors and eigenspaces for $n \times n$ matrices, by regarding the $n \times n$ matrix, $\underline{\mathbf{A}}$, over $\mathbb{F}$ as the linear transformation

$$L_{\underline{\mathbf{A}}} \colon \mathbb{F}_{(n)} \longrightarrow \mathbb{F}_{(n)}, \quad \mathbf{x} \longmapsto \underline{\mathbf{A}}\,\mathbf{x}$$

Definition 12.21. The *eigenvalues, eigenvectors and eigenspaces of* $\underline{\mathbf{A}}$ are those of the linear transformation

$$L_{\underline{\mathbf{A}}} \colon \mathbb{F}_{(n)} \longrightarrow \mathbb{F}_{(n)}, \quad \mathbf{x} \longmapsto \underline{\mathbf{A}}\mathbf{x}$$

Observation 12.22. If $\underline{\mathbf{A}} \in$ has characteristic polynomial, $\chi_{\underline{\mathbf{A}}}(t) = b_0 + b_1 t + \cdots + b_n t^n$, then it follows form the definition of the characterstic function that

$$\begin{aligned} b_0 &= \det(\underline{\mathbf{A}}) \\ b_{n-1} &= (-1)^{n-1}\,\mathrm{tr}(\underline{\mathbf{A}}) \\ b_n &= (-1)^n \end{aligned}$$

Observation 12.23. The matrix of $L_{\underline{\mathbf{A}}}$ with respect to the standard basis of $\mathbb{F}_{(n)}$ is $\underline{\mathbf{A}}$ itself.

Observation 12.24. In particular, if $\underline{\mathbf{A}}$ is an $n \times n$ matrix, then the eigenspace of the eigenvalue 0 is the null space of $\underline{\mathbf{A}}$.

Thus, the null space (or kernel) of $\underline{\mathbf{A}}$ is trivial if and only if 0 is not an eigenvalue of $\underline{\mathbf{A}}$.

We list further properties of eigenvalues.

Theorem 12.25. *Let λ be an eigenvalue of the matrix $\underline{\mathbf{A}}$.*

(i) λ^n is an eigenvalue of $\underline{\mathbf{A}}^n$ for any $n \in \mathbb{N}$.

(ii) If $\underline{\mathbf{A}}$ is invertible, then λ^n is an eigenvalue of $\underline{\mathbf{A}}^n$ for any $n \in \mathbb{Z}$.

(iii) λ is an eigenvalue of $\underline{\mathbf{A}}^t$.

Proof. **(i)** We adopt here the convention that $0^0 = 1$ and proceed by induction on n.

$\mathbf{n = 0}$: Since $\underline{\mathbf{A}}^0 = \underline{\mathbf{1}}_n$ and $\lambda^0 = 1$, the statement is true for $n = 0$

$\mathbf{n > 0}$: Suppose that $\underline{\mathbf{A}}^n\,\mathbf{x} = \lambda^n \mathbf{x}$. Then

$$\begin{aligned} \underline{\mathbf{A}}^{n+1}\mathbf{x} &= \underline{\mathbf{A}}(\underline{\mathbf{A}}^n\mathbf{x}) \\ &= \underline{\mathbf{A}}\lambda^n\mathbf{x} && \text{by the inductive hypothesis} \\ &= \lambda^n \underline{\mathbf{A}}\,\mathbf{x} \\ &= \lambda^n \lambda\mathbf{x} \\ &= \lambda^{n+1}\mathbf{x} \end{aligned}$$

(ii) Since $\underline{\mathbf{A}}$ is a square matrix, it is invertible if and only if its null space is trivial.

This is equivalent to 0's not being an eigenvalue of $\underline{\mathbf{A}}$.

In such a case, $\underline{\mathbf{A}}\,\mathbf{x} = \lambda\mathbf{x}$ if and only if $\lambda^{-1}\mathbf{x} = \underline{\mathbf{A}}$.

The result now follows by applying Part (i) to $\underline{\mathbf{A}}^{-1}$

(iii)

$$\begin{aligned}\det(\underline{\mathbf{A}}^t - \lambda\underline{\mathbf{1}}_n) &= \det(\underline{\mathbf{A}}^t - \lambda\underline{\mathbf{1}}_n{}^t)\\ &= \det((\underline{\mathbf{A}} - \lambda\underline{\mathbf{1}}_n)^t)\\ &= \det(\underline{\mathbf{A}} - \lambda\underline{\mathbf{1}}_n)\end{aligned}$$

□

Corollary 12.26. *Let λ be an eigenvalue of the endomorphism $T\colon V \longrightarrow V$.*

(i) λ^n is an eigenvalue of T^n for any $n \in \mathbb{N}$, where T^n denotes the composition $T \circ \cdots \circ T$ with n terms.

(ii) If T is invertible, then λ^n is an eigenvalue of T^n for any $n \in \mathbb{Z}$

Example 12.27. We attempt to diagonalise the real matrix $\underline{\mathbf{A}} = \begin{bmatrix}1 & 6\\ 4 & 3\end{bmatrix}$

We apply elementary row operations to

$$\underline{\mathbf{A}} - \lambda\underline{\mathbf{1}}_2 = \begin{bmatrix}1-\lambda & 6\\ 4 & 3-\lambda\end{bmatrix}$$

in order to bring it to a form which makes the eigenvalues and eigenvectors evident.

$$\begin{bmatrix}1-\lambda & 6\\ 4 & 3-\lambda\end{bmatrix}$$

Adding $(\lambda - 1)$ times the second row to four four times the first, we obtain

$$\begin{bmatrix}0 & -(\lambda^2 - 4\lambda - 21)\\ 4 & 3-\lambda\end{bmatrix} \qquad (\diamondsuit)$$

Because of the first column, this matrix has rank at least 1.

So, the only way its determinant can be 0, is if the second column is a multiple of the first.

By inspection, this occurs if and only if $\lambda^2 - 4\lambda - 21 = 0$.

Since $\lambda^2 - 4\lambda - 21 = (\lambda + 3)(\lambda - 7)$, the eigenvalues of $\underline{\mathbf{A}}$ are -3 and 7.

It follows from the second row of the matrix in $(\diamondsuit)$, that

$$\begin{bmatrix}x\\ y\end{bmatrix}$$

is in the eigenspace for the eigenvalue λ if and only if $4x + (3-\lambda)y = 0$, that is $4x = (\lambda - 3)y$, or equivalently,

$$\begin{bmatrix}x\\ y\end{bmatrix} = r\begin{bmatrix}\lambda - 3\\ 4\end{bmatrix}$$

for some $r \in \mathbb{R}$.

We substitute these values successively to obtain the corresponding eigenvectors.

$\lambda = -\mathbf{3}$: Our transformed matrix is

$$\begin{bmatrix} 0 & 0 \\ 4 & 6 \end{bmatrix}$$

from which it follows that $\begin{bmatrix} x \\ y \end{bmatrix}$ is in the eigenspace of $\underline{\mathbf{A}}$ for -3 if and only if $2x + 3y = 0$.

Thus, $\begin{bmatrix} 3 \\ -2 \end{bmatrix}$ generates the eigenspace V_{-3}, and

$$\begin{bmatrix} 1 & 6 \\ 4 & 3 \end{bmatrix} \begin{bmatrix} 3 \\ -2 \end{bmatrix} = -3 \begin{bmatrix} 3 \\ -2 \end{bmatrix} = \begin{bmatrix} -9 \\ 6 \end{bmatrix} \tag{C1}$$

$\lambda = \mathbf{7}$: Our transformed matrix is

$$\begin{bmatrix} 0 & 0 \\ 4 & -4 \end{bmatrix}$$

from which it follows that $\begin{bmatrix} x \\ y \end{bmatrix}$ is in the eigenspace of $\underline{\mathbf{A}}$ for 7 if and only if $x - y = 0$.

Hence, $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ generates the eigenspace V_7, and

$$\begin{bmatrix} 1 & 6 \\ 4 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 7 \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 7 \\ 7 \end{bmatrix} \tag{C2}$$

We combine (C1) and (C2) to obtain

$$\begin{aligned} \begin{bmatrix} 1 & 6 \\ 4 & 3 \end{bmatrix} \begin{bmatrix} 3 & 1 \\ -2 & 1 \end{bmatrix} &= \begin{bmatrix} -9 & 7 \\ 6 & 7 \end{bmatrix} \\ &= \begin{bmatrix} (-3).3 & 7.1 \\ (-3).(-2) & 7.1 \end{bmatrix} \\ &= \begin{bmatrix} 3 & 1 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} -3 & 0 \\ 0 & 7 \end{bmatrix} \qquad \text{by Section 9.4} \end{aligned}$$

We may thus regard $\begin{bmatrix} 3 & 1 \\ -2 & 1 \end{bmatrix}$ as a "change-of-basis" or "transition" matrix.

Since its inverse is $\frac{1}{5}\begin{bmatrix} 1 & -1 \\ 2 & 3 \end{bmatrix}$, we obtain

$$\frac{1}{5}\begin{bmatrix} 1 & -1 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} 1 & 6 \\ 4 & 3 \end{bmatrix} \begin{bmatrix} 3 & 1 \\ -2 & 1 \end{bmatrix} = \begin{bmatrix} -3 & 0 \\ 0 & 7 \end{bmatrix},$$

which is a diagonal matrix, whose diagonal entries are precisely the eigenvaluse of $\underline{\mathbf{A}}$.

This diagonal matrix is the matrix of the linear transformation

$$L_{\underline{\mathbf{A}}} \colon \mathbb{R}_{(2)} \longrightarrow \mathbb{R}_{(2)}, \quad \underline{\mathbf{x}} \longmapsto \underline{\mathbf{A}}\,\underline{\mathbf{x}}$$

with respect to the basis

$$\left\{ \begin{bmatrix} 3 \\ -2 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}$$

for $\mathbb{R}_{(2)}$.

Emulating the above for the matrices

$$\begin{bmatrix} 4 & -3 \\ 1 & 0 \end{bmatrix}, \quad \begin{bmatrix} 4 & -4 \\ 1 & 0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 4 & -5 \\ 1 & 0 \end{bmatrix}$$

illustrates not only what can be done, but also some of the difficulties that can arise.

Example 12.28. In particular, direct computation shows that the matrix

$$\begin{bmatrix} 4 & -4 \\ 1 & 0 \end{bmatrix}$$

has only one eigenvalue, namely 2, and that every eigenvector must be of the form

$$\begin{bmatrix} 2t \\ t \end{bmatrix}$$

Hence there can be no basis for $\mathbb{R}_{(2)}$ consisting of eigenvectors of our matrix, showing that the conclusion of Corollary 12.18 is not true without some condition being imposed.

This example illustrates what can go wrong when an $n \times n$ matrix has at least one eigenvalue, but does not have n distinct ones.

Observation 12.29. While our procedure for finding eigenvalues and eigenvectors is, in principle, quite simple, significant problem do arise.

An immediate one is finding the zeroes of a polynomial, or, equvalently, expressing a polynomial as the product of linear factors (factors of the form $(t - a)$). There is no general formula for this even in the most familiar case, when the scalars are all complex numbers. In this case, the *Fundamental Theorem of Algebra* ensures that every polynomial can be factorised into linear factors, and *Cardano's formulæ*, dating from the 16$^{\text{th}}$ century, provide the factors when the polynomial in question has degree at most four. However, Lagrange, Abel and Galois proved in the 19$^{\text{th}}$ century, that no such general formula is possible for polynomials of degree at least five. This problem is studied in abstract algebra, where a proof is available using *Galois theory*.

The fact that exact solutions are only available in special cases means that in many practical situations, we are forced rely on *numerical methods* or other means to find sufficiently accurate approximations. This, in turn, leads to other interesting and important mathematical problems, such as finding efficient algorithms for the approximation and the question of the *stability* of the eigenvalues and eigenvectors when the coefficients are perturbed. Such questions are studied in courses on numerical methods and computer algebra.

12.1 The Cayley-Hamilton Theorem

If V is an n-dimensional vector space over $\mathbb{F}$ and $T\colon V \longrightarrow V$ a linear transformation, then so is T^k for any $k \in \mathbb{N}$. Now the linear transformations $V \longrightarrow V$ form a vector space $\mathrm{Hom}_{\mathbb{F}}(V,V)$ over $\mathbb{F}$ whose dimension is n^2. (To see this, recall that for a fixed basis, there is a bijection between $\mathrm{Hom}_{\mathbb{F}}(V,V)$ and $\mathbf{M}(n;\mathbb{F})$, which is actually a linear transformation, and hence an isomorphism: T corresponds to $\underline{\mathbf{A}}_T$.)

By Theorem 8.5 on page 88, $id_V, T, T^2, \ldots, T^{n^2}$ must be linearly dependent.

This means that there are $a_0, \ldots a_{n^2} \in \mathbb{F}$, not all 0, with

$$a_0 id_V + a_1 T + \cdots + a_{n^2} T^{n^2} = 0.$$

The corresponding matrix version is that for any $\underline{\mathbf{A}} \in \mathbf{M}(n;\mathbb{F})$ there are $a_0, \ldots a_{n^2} \in \mathbb{F}$, not all 0, with

$$a_0 \underline{\mathbf{1}}_n + a_1 \underline{\mathbf{A}} + \cdots + a_{n^2} \underline{\mathbf{A}}^{n^2} = \underline{\mathbf{0}}_n.$$

We can express this by saying that every endomorphism of an n-dimensional vector space over $\mathbb{F}$ is a zero of of polynomial equation of degree at most n^2 over $\mathbb{F}$, or, equivalently, every $n \times n$ matrix over $\mathbb{F}$ is a zero of a polynomial of degree at most n^2 over $\mathbb{F}$.

This immediately raises two questions:

1. **Is this the best we can do, or is there a polynomial, p, of lower degree which also has T (resp. $\underline{\mathbf{A}}$) as a zero?**
2. **Given T (or $\underline{\mathbf{A}}$), determine the polynomial p explicitly.**

If we let m_T (or $m_{\underline{\mathbf{A}}}$) be the lowest degree of any non-zero polynomial for which T (or $\underline{\mathbf{A}}$) is a zero, then what we have show is that if $\dim V = n$, then $m_T = m_{\underline{\mathbf{A}}} \leq n^2$.

The following example shows that, the best universal bound for $m$cannot be less than n. The Cayley-Hamilton Theorem (Theorem 12.34 on page 167) then shows that T (or $\underline{\mathbf{A}}$) is always the zero of a specific polynomial of degree precisely n.

Example 12.30. Choose a basis $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ of V. Take $T\colon V \longrightarrow V$ be defined by

$$T(\mathbf{e}_j) := \begin{cases} e_{j+1} & \text{if } j < n \\ e_1 & \text{if } j = n \end{cases}$$

It follows, successively, that $T(\mathbf{e}_1) = \mathbf{e}_2,\ T^2(\mathbf{e}_1), \ldots, T^{n-1}(\mathbf{e}_1) = \mathbf{e}_n$.

Let $p(t) = a_0 + a_1 t + \cdots + a_{n-1} t^{n-1}$ be a polynomial in $\mathbb{F}[t]$ for which $\mathbb{F}(T) = 0$.

This means that $p(T)(\mathbf{v}) = \mathbf{0}_V$ for every $\mathbf{v} \in V$.

In particular, take $\mathbf{v} = \mathbf{e}_1$. Then

$$p(T)(\mathbf{v}) = a_0 \mathbf{e}_1 + a_1 \mathbf{e}_2 + \cdots + a_{n-1} \mathbf{e}_n = \mathbf{0}_V.$$

Since $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ is a basis of V, the on;y possibility is that $a_0 = a_1 = \cdots = a_{n-1} = 0$.

This is the worst that can occur: If $T\colon V \longrightarrow V$ is an endomorphism of an n-dimensional vector space, then it is a zero of a polynomial of specific degree n, the *characteristic polynomial*. We prove this in terms of matrices as the Cayley-Hamilton Theorem, which asserts that any $n \times n$ matrix satisfies its own *characteristic equation*.

As we need the construction of an $(n-1)\times(n-1)$ matrix from a given $n\times n$ matrix by deleting one row and one column, we recall our earlier definition.

Let $\underline{\mathbf{A}} = [a_{ij}]_{n\times n}$ be an $n\times n$ matrix. For $1\le p,q\le n$ let $\underline{\mathbf{A}}_{(p)(q)} = [x_{ij}]_{(n-1)\times(n-1)}$ where

$$x_{ij} = \begin{cases} a_{ij} & i<p,\ j<q \\ a_{i(j+1)} & i<p,\ j\ge q \\ a_{(i+1)j} & i\ge p,\ j<q \\ a_{(i+1)(j+1)} & i\ge p,\ j\ge q \end{cases}.$$

Definition 12.31. Using the notation above, put

$$A_{ji} := (-1)^{i+j}\det\left(\underline{\mathbf{A}}_{(i)(j)}\right).$$

The *adjugate* of $\underline{\mathbf{A}}$ is the matrix

$$\operatorname{adj}\underline{\mathbf{A}} := [A_{ij}]_{n\times n}.$$

Lemma 12.32. *Given any $n\times n$ matrix $\underline{\mathbf{A}}$,*

$$(\operatorname{adj}\underline{\mathbf{A}})\underline{\mathbf{A}} = \underline{\mathbf{A}}(\operatorname{adj}\underline{\mathbf{A}}) = (\det\underline{\mathbf{A}})\underline{\mathbf{1}}_n.$$

Proof. The proof follows directly from the definition of matrix multiplication together with the definition and properties of the determinant function. □

Before proving it, we illustrate the Cayley-Hamilton Theorem with an example. Our proof of the theorem is a generalisation of this example.

Example 12.33. Take $\underline{\mathbf{A}} := \begin{bmatrix} c & b & a \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$.

Then

$$\begin{aligned}\chi_{\underline{\mathbf{A}}}(t) &= \det\left(\begin{bmatrix} c-t & b & a \\ 1 & -t & 0 \\ 0 & 1 & -t \end{bmatrix}\right) \\ &= -t^3 + ct^2 + bt + a\end{aligned}$$

As

$$\underline{\mathbf{A}}^2 = \begin{bmatrix} c^2+b & cb+a & ca \\ c & b & a \\ 1 & 0 & 0 \end{bmatrix}$$

and

$$\underline{\mathbf{A}}^3 = \begin{bmatrix} c^3+2bc+a & c^2b+b^2+ca & c^2a+ba \\ c^2+b & cb+a & ca \\ c & b & a \end{bmatrix}$$

it follows by direct substitution that

$$\begin{aligned}\chi_{\underline{\mathbf{A}}}(\underline{\mathbf{A}}) &= \begin{bmatrix} -c^3-2bc-a & -c^2b-b^2-ca & -c^2a-ba \\ -c^2-b & -cb-a & -ca \\ -c & -b & -a \end{bmatrix} \\ &\quad + \begin{bmatrix} c^3+cb & c^2b+ca & c^2a \\ c^2 & bc & ac \\ c & 0 & 0 \end{bmatrix} + \begin{bmatrix} bc & b^2 & ba \\ b & 0 & 0 \\ 0 & b & 0 \end{bmatrix} + \begin{bmatrix} a & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & a \end{bmatrix} \\ &= \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}\end{aligned}$$

Our proof of the Caylet-Hamilton Theorem makes use of the adjugate of $\underline{\mathbf{A}} - t\underline{\mathbf{1}}_n$.
We illustrate how this can be expressed as a polynomial in t with matrices as coefficients.
Put $\underline{\mathbf{B}} := \underline{\mathbf{A}} - t\underline{\mathbf{1}}_3$, so that $\chi_{\underline{\mathbf{A}}}(t) = \det(\underline{\mathbf{B}})$.
We compute $\mathrm{adj}(\underline{\mathbf{B}}) = [x_{ij}]_{3\times 3}$, where

$$x_{ij} = (-1)^{i+j}\det(\underline{\mathbf{B}}_{(j)(i)}),$$

with $\underline{\mathbf{B}}_{(ji)}$ the 2×2 matrix obtained from $\underline{\mathbf{B}}$ by deleting its j^{th} row and i^{th} column.

$$\begin{aligned} x_{11} &= (-1)^{1+1}\det\left(\begin{bmatrix} -t & 0 \\ 0 & -t \end{bmatrix}\right) \\ &= t^2 \\ x_{12} &= (-1)^{1+2}\det\left(\begin{bmatrix} b & a \\ 0 & -t \end{bmatrix}\right) \\ &= bt \\ x_{13} &= (-1)^{1+3}\det\left(\begin{bmatrix} b & a \\ -t & 0 \end{bmatrix}\right) \\ &= at \\ x_{21} &= (-1)^{2+1}\det\left(\begin{bmatrix} 1 & 0 \\ 0 & -t \end{bmatrix}\right) \\ &= t \\ x_{22} &= (-1)^{2+2}\det\left(\begin{bmatrix} c-t & a \\ 0 & -t \end{bmatrix}\right) \\ &= t^2 - ct \\ x_{23} &= (-1)^{2+3}\det\left(\begin{bmatrix} c-t & a \\ 1 & 0 \end{bmatrix}\right) \\ &= a \\ x_{31} &= (-1)^{3+1}\det\left(\begin{bmatrix} 1 & -t \\ 0 & 1 \end{bmatrix}\right) \\ &= 1 \end{aligned}$$

$$\begin{aligned} x_{32} &= (-1)^{3+2}\det\left(\begin{bmatrix} c-t & b \\ 0 & 1 \end{bmatrix}\right) \\ &= t-c \\ x_{33} &= (-1)^{3+3}\det\left(\begin{bmatrix} c-t & b \\ 1 & -t \end{bmatrix}\right) \\ &= t^2 - ct - b \end{aligned}$$

Thus,

$$\begin{aligned} \operatorname{adj}(\underline{\mathbf{B}}) &= \begin{bmatrix} t^2 & bt & at \\ t & t^2-ct & a \\ 1 & t-c & t^2-ct-b \end{bmatrix} \\ &= t^2\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} + t\begin{bmatrix} 0 & b & a \\ 1 & -c & 0 \\ 0 & 1 & -c \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & a \\ 1 & 0 & -b \end{bmatrix} \end{aligned}$$

Theorem 12.34 (**Cayley-Hamilton**). *Let $\chi_{\underline{\mathbf{A}}}(t)$ be the characteristic polynomial of the $n \times n$ matrix $\underline{\mathbf{A}}$. Then $\chi_{\underline{\mathbf{A}}}(\underline{\mathbf{A}}) = \underline{\mathbf{0}}_n$.*

Proof. Put $\underline{\mathbf{B}} := \underline{\mathbf{A}} - t\underline{\mathbf{1}}_n$.

By the definition of the determinant, there are $b_0, \ldots, b_n \in \mathbb{F}$ with

$$\chi_{\underline{\mathbf{A}}}(t) = \det \underline{\mathbf{B}} = b_0 + b_1 t + \cdots + b_n t^n = \sum_{j=0}^{n-1} b_j t^j. \tag{i}$$

Since $\operatorname{adj}(\underline{\mathbf{B}}) := [x_{ij}]_{n\times n}$, with $x_{ij} := (-1)^{i+j} \det \underline{\mathbf{B}}_{(ji)}$ and $(-1)^{i+j} \det \underline{\mathbf{B}}_{(ji)}$ is a polynomial in t of degree at most $n-1$, there are $n \times n$ matrices $\underline{\mathbf{B}}_0, \ldots, \underline{\mathbf{B}}_{n-1}$ with

$$\operatorname{adj} \underline{\mathbf{B}} \;=\; \underline{\mathbf{B}}_0 \;+\; \underline{\mathbf{B}}_1 t \;+ \cdots +\; \underline{\mathbf{B}}_{n-1} t^{n-1} = \sum_{j=0}^{n-1} t^j \underline{\mathbf{B}}^j. \tag{ii}$$

Hence,

$$\begin{aligned} (\det \underline{\mathbf{B}})\underline{\mathbf{1}}_n &= \underline{\mathbf{B}} \operatorname{adj} \underline{\mathbf{B}} \\ &= (\underline{\mathbf{A}} - t\underline{\mathbf{1}}_n) \operatorname{adj} \underline{\mathbf{B}} \\ &= \underline{\mathbf{A}} \operatorname{adj} \underline{\mathbf{B}} - t \operatorname{adj} \underline{\mathbf{B}} \end{aligned} \tag{iii}$$

and

$$\begin{aligned} \chi_{\underline{\mathbf{A}}}(t)\underline{\mathbf{1}}_n &= \sum_{j\,=\,0}^{n} b_j t^j \underline{\mathbf{1}}_n \\ &= (\det \underline{\mathbf{B}})\underline{\mathbf{1}}_n && \text{by (i)} \\ &= \underline{\mathbf{A}} \operatorname{adj} \underline{\mathbf{B}} - t \operatorname{adj} \underline{\mathbf{B}} && \text{by (iii)} \\ &= \underline{\mathbf{A}} \sum_{j=0}^{n-1} t^j \underline{\mathbf{B}}^j - t \sum_{j=0}^{n-1} t^j \underline{\mathbf{B}}^j && \text{by (ii)} \\ &= \underline{\mathbf{A}}\,\underline{\mathbf{B}}_0 + t(\underline{\mathbf{A}}\,\underline{\mathbf{B}}_1 - \underline{\mathbf{B}}_0) + \cdots + t^{n-1}(\underline{\mathbf{A}}\,\underline{\mathbf{B}}_{n-1} - \underline{\mathbf{B}}_{n-2}) - t^n \underline{\mathbf{B}}_{n-1}. \end{aligned}$$

Thus

$$\begin{aligned}\chi_{\underline{\mathbf{A}}}(\underline{\mathbf{A}}) &= \underline{\mathbf{A}}\,\underline{\mathbf{B}}_0 + \underline{\mathbf{A}}(\underline{\mathbf{A}}\,\underline{\mathbf{B}}_1 - \underline{\mathbf{B}}_0) + \cdots + \underline{\mathbf{A}}^{n-1}(\underline{\mathbf{A}}\,\underline{\mathbf{B}}_{n-1} - \underline{\mathbf{B}}_{n-2}) - \underline{\mathbf{A}}^n\underline{\mathbf{B}}_{n-1} \\ &= \underline{\mathbf{0}}\end{aligned}$$

□

By Observation 12.22 on page 160, the characteristic polynomial of $\underline{\mathbf{A}} \in \mathbf{M}(n;\mathbb{F})$ is a polynomial over $\mathbb{F}$ of the form $(-1)^n(b_0 + b_1t + \cdots + b_{n-1}t^{n-1} + t^n)$.

It is natural to ask whether every polynomial of this form is the characteristic polynomial of a matrix $\underline{\mathbf{A}} \in \mathbf{M}(n;\mathbb{F})$, or, equivalently, of an endomorphism $T\colon V \longrightarrow V$, where $\dim_{\mathbb{F}}(V) = n$.

As suggested by Example 12.33 on page 165, the answer is affirmative, as the following example shows.

Example 12.35. The $n \times n$ matrix

$$\begin{bmatrix} -b_{n-1} & -b_{n-2} & \cdots & \cdots & -b_0 \\ 1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{bmatrix}$$

has characteristic polynomial $(-1)^n(b_0 + b_1t + \cdots + b_{n-1}t^{n-1} + t^n)$.

The verification is left as an exercise.

This matrix is the *companion matrix* of the polynomial $b_0 + b_1t + \cdots + b_{n-1}t^{n-1} + t^n$.

12.2 Discussion

While every $n \times n$ matrix is a zero of its characteristic polynomial, which has degree n, some matrices are zeroes of polynomial of lower degree. For example the zero matrix is a zero of the polynomial t, and the identity matrix is a zero of the polynomial $t - 1$.

We summarise a more complete analysis without providing proofs, since these require the introduction of concepts and techniques beyond the scope of these notes. They are investigated in abstract algebra.

If we restrict attention to polynomials whose the leading coefficient is 1, then there is a unique polynomial of lowest possible degree for which the matrix $\underline{\mathbf{A}}$ is a zero. This is the *minimm polynomial* of $\underline{\mathbf{A}}$, $\mu_{\underline{\mathbf{A}}}$. It divides every polynomial for which $\underline{\mathbf{A}}$ is a zero, and its zeroes are precisely the eigenvalues of $\underline{\mathbf{A}}$, that is, the zeroes of the characteristic polynomial of $\underline{\mathbf{A}}$. The main result on the minimum polynomial is that the matrix $\underline{\mathbf{A}}$ is diagonalisable if and only if

$$\mu_{\underline{\mathbf{A}}}(t) = (t - \lambda_1)\cdots(t - \lambda_m)$$

with $\lambda_i = \lambda_j$ if and only if $i = j$.

The field $\mathbb{F}$ is *algebraically closed* if and only if every polynomial in one indeterminate over $\mathbb{F}$ can be written as a product of linear factors. In such a case, every matrix, $\underline{\mathbf{A}}$ over $\mathbb{F}$ can be brought to *block diagonal form*, or *Jordan normal form*

$$\begin{bmatrix} \underline{\mathbf{A}}_{\lambda_1} & \underline{\mathbf{0}} & \cdots & \underline{\mathbf{0}} \\ \underline{\mathbf{0}} & \underline{\mathbf{A}}_{\lambda_2} & \cdots & \\ \vdots & \vdots & \ddots & \end{bmatrix}$$

with each *Jordan block*, $\underline{\mathbf{A}}_{\lambda_j}$,of the form

$$\begin{bmatrix} \lambda_j & 1 & 0 & \cdots & 0 \\ 0 & \lambda_j & 1 & 0 & \cdots \\ \vdots & & \ddots & \ddots & \vdots \end{bmatrix}.$$

Example 12.36. The minimum polynomial of the matrix in Example 2.4 on page 23,

$$\begin{bmatrix} 4 & -3 \\ 1 & 0 \end{bmatrix},$$

is $(t-1)(t-2)$, which is of degree 2 and has two distinct zeroes. The block diagonal form comprises the two Jordan blocks

$$\underline{\mathbf{A}}_3 = \begin{bmatrix} 3 \end{bmatrix} \quad \text{and} \quad \underline{\mathbf{A}}_1 = \begin{bmatrix} 1 \end{bmatrix}$$

The minimum polynomial of the matrix in Example 2.5 on page 23,

$$\begin{bmatrix} 4 & -4 \\ 1 & 0 \end{bmatrix},$$

must divide its characteristic polynomial, $(t-2)^2$. Hence it must be either $t-2$ or $(t-2)^2$. Since

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

the minimum polynomial cannot be $t-2$. Hence it must be $(t-2)^2$, which is of degree 2.

Since this fails to have two distinct zeroes, the block diagonal form has the single Jordan block

$$\underline{\mathbf{A}}_2 = \begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}$$

The minimum polynomial of the matrix in Example 2.6 on page 24,

$$\begin{bmatrix} 4 & -5 \\ 1 & 0 \end{bmatrix},$$

is $(t-2)^2+1$, which is of degree 2, which fails to have any real zeroes, but has two distinct complex zeroes. If we now regard is as a complex matrix, the block diagonal form comprises the two Jordan blocks

$$\underline{\mathbf{A}}_{2+i} = \begin{bmatrix} 2+i \end{bmatrix} \quad \text{and} \quad \underline{\mathbf{A}}_{2-i} = \begin{bmatrix} 2-i \end{bmatrix}$$

where $i^2 = 1$.

We summarise the above.

The matrix $\underline{\mathbf{A}}$ is diagonalisable if and only if each of its Jordan blocks is 1×1.

We turn to an an alternative formulation.

Definition 12.37. Let $T\colon V \longrightarrow V$ be an endomorphism of the finitely generated vector space V (or, equivalently, take $\underline{\mathbf{A}} \in \mathbf{M}(n;\mathbb{F})$) and λ an eigenvalue of T (or $\underline{\mathbf{A}}$).

The *algebraic multiplicity* of λ is $a \in \mathbb{N}$ if and only if $(t-\lambda)^a$ divides $\chi_T(t)$ (or $\chi_{\underline{\mathbf{A}}}(t)$), but $(t-\lambda)^{a+1}$ does not.

The *geometric multiplicity* of λ is $\dim(V_\lambda)$, that is to say, the number of linearly independent eigenvectors for the eigenvalue λ.

Example 12.38. Take $\underline{\mathbf{A}} = \begin{bmatrix} 4 & -4 \\ 1 & 0 \end{bmatrix}$.

Since $\chi_{\underline{\mathbf{A}}}(t) = (t-2)^2$, the algebraic multiplicity of the eigenvalue 2 is 2.

As we saw in Example 12.28 on page 163, every eigenvector is of the form $\begin{bmatrix} 2t \\ t \end{bmatrix}$, showing that $\dim(V_2) = 1$, that is, the geometric multiplicity of 2 is 1.

We show that this is typical.

Lemma 12.39. *The geometric multiplicity of λ cannot exceed its algebraic multiplicity.*

Proof. Let λ be an eigenvalue of $T\colon V \longrightarrow V$ with geometric multiplicity g.

Choose linearly independent eigenvectors $\mathbf{e}_1, \ldots, \mathbf{e}_g$ for the eigenvalue λ. Extend this to a basis $\{\mathbf{e}_1, \ldots, \mathbf{e}_g, \ldots \mathbf{e}_n\}$ of V.

The matrix, $\underline{\mathbf{A}}$, of T with respect to this basis is of the form

$$\begin{bmatrix} \lambda & 0 & \cdots & 0 & * \\ 0 & \lambda & & \vdots & * \\ \vdots & & \ddots & 0 & * \\ 0 & \cdots & 0 & \lambda & * \\ \vdots & & & 0 & * \end{bmatrix}$$

This being the case, $(t-\lambda)^g$ must divide $\chi_{\underline{\mathbf{A}}}(t) = \chi_T(t)$. □

Theorem 12.40. *An $n \times n$ matrix, is diagonalisable if and only if each eigenvalue has the same geometric and algebraic multiplicity, and the sum of these is n.*

Proof. A little thought shows that these conditions are necessary and sufficient to ensure that there is a basis consisting of eigenvectors. □

12.3 Exercises

Exercise 12.1. Given linear transformations $R\colon V \longrightarrow V'$ and $S\colon W \longrightarrow W'$, let $\underline{\mathbf{A}}$, be the matrix of R with respect to the bases $\{\mathbf{e}_i\}$ for V and $\{\mathbf{e}_{k'}\}$ for V' and $\underline{\mathbf{B}}$ the matrix of S with respect to the bases $\{\mathbf{f}_j\}$ for W and $\{\mathbf{f}_{l'}\}$ for W'.

Show that

$$\begin{bmatrix} \underline{\mathbf{A}} & \underline{\mathbf{0}} \\ \underline{\mathbf{0}} & \underline{\mathbf{B}} \end{bmatrix}$$

is the matrix of $R \oplus S$ with respect to the bases $\{(\mathbf{e}_i, \underline{\mathbf{0}}_W), (\underline{\mathbf{0}}_V, \mathbf{f}_j)\}$ for $V \oplus W$ and $\{(\mathbf{e}'_k, \mathbf{0}_{W'}), (\mathbf{0}_{V'}, \mathbf{f}'_l)\}$ for $V' \oplus W'$.

Exercise 12.2. Find the eigenvalues and eigenvectors of the following matrices:

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

(b) $\begin{bmatrix} -1 & 0 & 0 \\ 0 & -2 & 2 \\ 12 & 1 & -3 \end{bmatrix}$

Exercise 12.3. Find the real eigenvalues and eigenvectors of the following matrices.

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

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

(c) $\begin{bmatrix} 4 & -5 \\ 1 & 0 \end{bmatrix}$

(d) $\begin{bmatrix} 1 & 2 \\ 2 & 5 \end{bmatrix}$

Exercise 12.4. Consider the real matrix

$$\underline{\mathbf{A}}_\varepsilon = \begin{bmatrix} 1+\varepsilon & 1 \\ 1 & 0 \end{bmatrix}.$$

Find its eigenvalues and eigenvectors as a function of $\varepsilon \geq 0$.

Exercise 12.5. Eigenvalues, eigenvectors and eigenspaces make sense in *any* vector space, not merely in finite dimensional vector spaces, and many problems can be formulated as eigenvalue problems. This exercise is devoted to examples of this.

Let $\mathcal{C}^\infty(\mathbb{R})$ denote the set of all smooth (that is, infinitely differentiable) real-valued functions defined on $\mathbb{R}$. Let

$$D\colon \mathcal{C}^\infty(\mathbb{R}) \longrightarrow \mathcal{C}^\infty(\mathbb{R}), \quad f \longmapsto f'$$

be differentiation.

In other words, $\big(D(f)\big)(x) = f'(x)$ for all $f \in \mathcal{C}^\infty(\mathbb{R})$ and $x \in \mathbb{R}$.

(a) Show that D is an endomorphism of the real vector space $\mathcal{C}^\infty(\mathbb{R})$, and find its eigenvalues and corresponding eigenvectors.

(b) Given $b \in \mathbb{R}$, show that

$$(D^2 + 2bD)\colon \mathcal{C}^\infty(\mathbb{R}) \longrightarrow \mathcal{C}^\infty(\mathbb{R}), \quad f \longmapsto f'' + 2bf'$$

defines an endomorphism, and find its eigenvalues and corresponding eigenvectors.

Exercise 12.6. Verify the Cayley-Hamilton Theorem for the following matrices.

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

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

Exercise 12.7. Verify that the $n \times n$ matrix

$$\begin{bmatrix} -b_{n-1} & -b_{n-2} & \cdots & \cdots & -b_0 \\ 1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{bmatrix}$$

has characteristic polynomial

$$(-1)^n(b_0 + b_1 t + \cdots + b_{n-1}t^{n-1} + t^n)$$

Neglect of mathematics work injury to all knowledge, since he who is ignorant of it cannot know the other sciences or things of this world. And what is worst, those who are thus ignorant are unable to perceive their own ignorance, and so do not seek a remedy.

Roger Bacon

Chapter 13

Inner Product Spaces

The discussion and the theory developed so far have applied to vector spaces over any field. The only restriction we occasionally made was to vector spaces which are finitely generated. Even then, not all of our results not actually depended on this hypothesis.

On the other hand, we have frequently appealed to geometry to provide motivation, illustration or graphical representation of concepts and theorems, which meant restricting attention to $\mathbb{R}^n$ ($n \in \mathbb{N}$). This is hardly surprising, since $\mathbb{R}^n$ is not only the most familiar vector space, but also the locus of analytic geometry since Descartes.

We now turn our attention to formulating such informal discussion and heuristic arguments more rigorously. Specifically, we investigate the additional structure a vector space must support in order for us to be able to "do geometry", that is, to speak of *distances* and *angles*. [Recall that we have already discussed what we mean by a "line", a "plane", and so on, in any vector space.]

Surprisingly, measuring angles also provides a way of measuring distance. However, the converse is not true. We do not enter a discussion here of why, for this and related questions are discussed in detail in courses on functional analysis, which may be fruitfully thought of as the study of infinite dimensional real and complex vector spaces, requiring, in addition, concepts from topology.

Our approach is to first briefly discuss making sense of the "length" of a vector, show how this permits us to define a notion of distance and to define continuity of functions between vector spaces. It follows that all linear transformations are continuous.

We then introduce the additional structure required to make sense of the notion of an "angle" between vectors and show how this allows us to speak of length, hence distance and hence continuity.

Our intuition is based on our experience with Euclidean space, which is a real vector space. The discussion actually applies to vector spaces over any sub-field of the field of complex numbers, although not for finite fields, or fields constructed from finite fields, for reasons beyond the scope of this course.

Since the proofs of the central results are simplest when we work over the complex numbers, with the more familiar cases being easy applications, we will work primarily with complex numbers

13.1 Normed Vector Spaces

We with the notion of *length*, or *magnitude*, of a vector.

(i) It should be clear that the length of a vector should be a non-negative real number, which is 0 for, and only for, the zero vector.

(ii) If we scale a vector, its length is multiplied by the magnitude of the scaling factor.

(iii) The length of the sum of two vectors cannot exceed the sum of the lengths of the two vectors.

We mention here, without further explanation, that it is essentially the second condition which forces us to restrict ourselves to vector spaces over sub-fields of $\mathbb{C}$. Hence, unless otherwise specified, $\mathbb{F}$ henceforth denotes a sub-field of $\mathbb{C}$. This means, in particular, that $\mathbb{F}$ contains $\mathbb{Q}$, the field of rational numbers.

The next definition formulates the properties above rigorously.

Definition 13.1. A *norm* on the vector space V over the sub-field $\mathbb{F}$ of $\mathbb{C}$ is a function

$$\| \ \| \colon V \longrightarrow \mathbb{R}_0^+$$

such that for all $\mathbf{u}, \mathbf{v} \in V$ and $\lambda \in \mathbb{F}$

N1 $\|\mathbf{u}\| = 0$ if and only if $\mathbf{u} = \mathbf{0}_V$

N2 $\|\lambda\mathbf{u}\| = |\lambda| \, \|\mathbf{u}\|$

N3 $\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$.

A *normed vector space* is a vector space, V, over the field $\mathbb{F}$ ($\subseteq \mathbb{C}$), equipped with a norm, $\| \ \|$. It is denoted by $(V, \| \ \|)$, or simply by V when the norm is understood.

The vector, $\mathbf{v}$, in the normed vector space $(V, \| \ \|)$ is *normal* or *normalised*, or a *unit vector* if and only if $\|\mathbf{v}\| = 1$.

Example 13.2. The *absolute value* or *modulus* of a complex number defines a norm on any sub-field $\mathbb{F}$ of $\mathbb{C}$. The verification is left as an exercise.

Example 13.3. Let $\mathbb{F}$ be a sub-field of $\mathbb{C}$. Then

$$\| \ \|_{\mathbb{F}^n} \colon \mathbb{F}^n \longrightarrow \mathbb{R}_0^+, \quad (x_1, \ldots, x_n) \longmapsto \big(\sum_{j=1}^{n} |x_j|^2 \big)^{\frac{1}{2}}$$

defines a norm on $\mathbb{F}^n$, called the *Euclidean norm* on $\mathbb{F}$. The verification is left as an exercise.

Example 13.2 is just the case $n = 1$, and when $\mathbb{F} \subseteq \mathbb{R}$, we may replace $|x_j|^2$ by x_j^2.

Example 13.4. Recall that

$$G \colon \mathbb{R}^2 \longrightarrow \mathbb{C}, \quad (x, y) \longmapsto z := x + iy$$

is an isomorphism of real vector spaces, and that

$$\| \ \|_{\mathbb{R}^2} = \sqrt{x^2 + y^2} = |x + iy| = \|G(x, y)\|_{\mathbb{C}^1}$$

Similarly,

$$G_n \colon \mathbb{R}^{2n} \longrightarrow \mathbb{C}^n, \quad (x_1, \ldots, x_{2n}) \longmapsto (z_1, \ldots, z_n),$$

where $z_j := x_{2j-1} + ix_{2j}$ $(j = 1, \ldots, n)$ is an isomorphism of real vector spaces, with

$$\|(x_1, \ldots, x_{2n})\|_{\mathbb{R}^{2n}} = \left(\sum_{k=1}^{2n} |x_k|^2 \right)^{\frac{1}{2}} = \left(\sum_{j=1}^{n} |z_j|^2 \right)^{\frac{1}{2}} = \|G_n(x_1, \ldots, x_n)\|_{\mathbb{C}^n}.$$

Example 13.5. Let $V = \{f\colon [0,1] \longrightarrow \mathbb{R} \mid \ f \text{ continuous}\}$ denote the real vector space of all continuous real valued functions defined on the closed unit interval. Then

$$\| \ \|_1 \colon V \longrightarrow \mathbb{R}_0^+, \quad f \longmapsto \int_0^1 |f(t)|\, dt$$

defines a norm on V, called the $\mathcal{L}^1$ *norm* on V.

Observation 13.6. If $(V, \| \ \|)$ is a non-trivial normed vector space then every $\mathbf{v} \in V$ in is a multiple of a unit vector.

For if $\mathbf{v} \neq \mathbf{0}_V$, define

$$\mathbf{v}_u := \frac{\mathbf{v}}{\|\mathbf{v}\|}.$$

Plainly, $\mathbf{v} = \|\mathbf{v}\|\mathbf{v}_u$.

If, on the other hand, $\mathbf{v} = \mathbf{0}_V$, take any $\mathbf{v} \neq \mathbf{0}_V$ and define

$$\mathbf{v}_u := \frac{\mathbf{v}}{\|\mathbf{v}\|}.$$

Then $\mathbf{0}_V = 0\mathbf{v}_u = \|\mathbf{0}\|\mathbf{v}_u$

The norm on a vector space can be used to provide a measure of distance between any two elements of V. We first characterise what we mean by the *distance* between two points in a set.

(i) The distance between two points is a non-negative real number, which is 0 if and only if the two points coincide.

(ii) The distance from one point to another is the same as the distance from the second to the first.

(iii) The distance between two points cannot exceed the sum of the distances of the first to any point plus the distance from that point to the second.

We mention here, without further explanation, that it these properties do not require any structure beyond being a set — in particular, there is no need to consider vector spaces. The study of sets equipped with a notion of distance between its points is the *theory of metric spaces*, a part of the study of topology.

We now express the properties above formally, turning them into a definition.

Definition 13.7. A *metric* (or *distance function*) on the set X is a function

$$d\colon X \times X \longrightarrow \mathbb{R}_0^+$$

such that for all $x, y, z \in X$

MS1 $d(x, y)$ if and only if $x = y$

MS2 $d(y, x) = d(x, y)$

MS3 $d(x, z) \leq d(x, y) + d(y, z)$.

A *metric space* comprises a set, X, equipped with a metric, d. We denote it by (X, d), writing only X when the metric is understood.

We now show that every normed vector space is a metric space in a natural way.

Definition 13.8. For the normed vector space, $(V, \| \ \|)$, over the field $\mathbb{F}$, define

$$d_{\| \ \|} \colon V \times V \longrightarrow \mathbb{R}_0^+, \quad (\mathbf{u}, \mathbf{v}) \longmapsto \|\mathbf{u} - \mathbf{v}\|$$

Lemma 13.9. *Let $(V, \| \ \|)$ be a normed vector space over the field $\mathbb{F}$. Then $(V, d_{\| \ \|})$ is a metric space.*

Proof. Take $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$.

Since $\|\mathbf{x}\| \geq 0$ for all $\mathbf{x} \in V$, $d(\mathbf{u}, \mathbf{v}) := \|\mathbf{u} - \mathbf{v}\| \geq 0$ for all $\mathbf{u}, \mathbf{v} \in V$, showing that $d_{\| \ \|}$ is well defined. Moreover,

$$\begin{aligned}
d(\mathbf{u}, \mathbf{v}) = 0 \ &\text{ if and only if } \ \|\mathbf{u} - \mathbf{v}\| \\
&\text{ if and only if } \ \mathbf{u} - \mathbf{v} = \mathbf{0}_V && \text{by } \mathbf{N1} \\
&\text{ if and only if } \ \mathbf{u} = \mathbf{v}, && \text{verifying } \mathbf{MS1}. \\
d(\mathbf{v}, \mathbf{u}) &:= \|\mathbf{v} - \mathbf{u}\| \\
&= |-1| \|\mathbf{u} - \mathbf{v}\| && \text{by } \mathbf{N2} \\
&= \|\mathbf{u} - \mathbf{v}\| \\
&=: d(\mathbf{u}, \mathbf{v}), && \text{verifying } \mathbf{MS2}. \\
d(\mathbf{u}, \mathbf{w}) &:= \|\mathbf{u} - \mathbf{w}\| \\
&= \|\mathbf{u} - \mathbf{v} + \mathbf{v} - \mathbf{w}\| \\
&\leq \|\mathbf{u} - \mathbf{v}\| + \|\mathbf{v} - \mathbf{w}\| && \text{by } \mathbf{N3} \\
&=: d(\mathbf{u}, \mathbf{v}) + d(\mathbf{v}, \mathbf{w}), && \text{verifying } \mathbf{MS3}.
\end{aligned}$$

□

Definition 13.10. If $(V, \| \ \|)$ is a normed vector space over the field $\mathbb{F}$, then $d_{\| \ \|}$ is the metric on V *induced by the norm* $\| \ \|$.

In particular, the Euclidean distance between points in $\mathbb{R}^n$ is the metric induced by the Euclidean norm. This allows us to reformulate the definition of continuity met in univariate and multivariate calculus in terms of metrics and so extend the notion of continuity to more general spaces.

Definition 13.11. Let (X, d) and (Y, e) be two metric spaces.

The function $f \colon X \longrightarrow Y$ is *continuous* at $a \in X$ if and only if given any $\varepsilon > 0$ there is a $\delta > 0$ such that $e(f(x), f(a)) < \varepsilon$ whenever $d(x, a) < \delta$.

We do not pursue these ideas further here, but return to our main interest, the quest for the structure required to be able to sense of "angle" between two vectors.

13.2 Inner Products

Recall that if we take two points P and Q in the Cartesian plane, neither of which is the origin, O, with co-ordinates (x, y) and (u, v) respectively, then we can compute the cosine of the angle $\angle POQ$ directly from the co-ordinates.

Suppose that the angle in question is θ. We express (x, y) in polar co-ordinates,

$$x = r\cos\alpha \quad \text{and} \quad y = r\sin\alpha$$

for uniquely determined $r > 0$ and $0 \leq \alpha < 2\pi$, so that, $r = \sqrt{x^2 + y^2}$.
Then, without loss of generality,

$$u = s\cos(\alpha + \theta) \quad \text{and} \quad v = s\sin(\alpha + \theta),$$

so that $s = \sqrt{u^2 + v^2}$.
Since $\theta = \alpha + \theta - \alpha$, it follows that

$$\begin{aligned}\cos\theta &= \cos(\alpha+\theta)\cos(\alpha) + \sin(\alpha+\theta)\sin(\alpha)\\ &= \frac{u}{s}\frac{x}{r} + \frac{v}{s}\frac{y}{r}\\ &= \frac{ux+vy}{\sqrt{u^2+v^2}\sqrt{x^2+y^2}}.\end{aligned}$$

Define

$$\langle\!\langle\ ,\ \rangle\!\rangle \colon \mathbb{R}^2 \times \mathbb{R}^2 \longrightarrow \mathbb{R}, \quad \big((u,v),(x,y)\big) \longmapsto ux + vy.$$

Then

$$\cos\theta = \frac{\langle\!\langle (u,v),(x,y)\rangle\!\rangle}{\sqrt{\langle\!\langle (u,v),(u,v)\rangle\!\rangle}\sqrt{\langle\!\langle (x,y),(x,y)\rangle\!\rangle}} \tag{13.1}$$

and we can define the angle between (u, v) and (x, y) to be the unique angle $\theta \in [0, \pi]$ satisfying

$$\sqrt{\langle\!\langle (u,v),(u,v)\rangle\!\rangle}\sqrt{\langle\!\langle (x,y),(x,y)\rangle\!\rangle}\cos\theta = \langle\!\langle (u,v),(x,y)\rangle\!\rangle.$$

Since we can define the angle purely in terms of the function $\langle\!\langle\ ,\ \rangle\!\rangle$, its characteristic properties provide a basis for the definition of a general notion allowing the use of the Equation (13.1) to define an angle between two vectors in a real vector space. We first characterise $\langle\!\langle\ ,\ \rangle\!\rangle$.

Lemma 13.12. *Take $(x, y), (u, v), (r, s) \in \mathbb{R}^2$ and $\alpha \in \mathbb{R}$. Then*

(i) $\langle\!\langle (x,y),(x,y)\rangle\!\rangle \geq 0$ *with equality if and only if* $(x, y) = (0, 0)$.

(ii) $\langle\!\langle (x,y),(u,v)\rangle\!\rangle = \langle\!\langle (u,v),(x,y)\rangle\!\rangle$

(iii) $\langle\!\langle \alpha(x,y),(u,v)\rangle\!\rangle = \alpha\langle\!\langle (x,y),(u,v)\rangle\!\rangle$

(iv) $\langle\!\langle (r,s)+(x,y),(u,v)\rangle\!\rangle = \langle\!\langle (r,s),(u,v)\rangle\!\rangle + \langle\!\langle (x,y),(u,v)\rangle\!\rangle$

Proof. The verifications are routine and left as an exercise. □

The function $\langle\!\langle\ ,\ \rangle\!\rangle$ just introduced leads naturally to

$$\|\ \|_{\langle\!\langle\ ,\ \rangle\!\rangle} \colon \mathbb{R}^2 \longrightarrow \mathbb{R}^+_O, \quad (x,y) \longmapsto \sqrt{\langle\!\langle (x,y),(x,y)\rangle\!\rangle}\,.$$

Lemma 13.13. $\|\ \|_{\langle\!\langle\ ,\ \rangle\!\rangle}$ *is a norm on* $\mathbb{R}^2$

Proof. It is routine to verify that $\|\ \|_{\langle\!\langle\ ,\ \rangle\!\rangle}$ is well defined and that **N1** and **N2** hold. On the other hand, the verification of **N3** is not quite as trivial.

We leave these as an exercise, since we prove a more general version a little later. □

We extend the above discussion to complex vector spaces, using the results of Lemma 13.12 on the previous page and Lemma 13.13 on the preceding page as a guide. One obvious generalisation, namely,

$$\langle\!\langle\ ,\ \rangle\!\rangle\colon \mathbb{C}^2 \times \mathbb{C}^2 \longrightarrow \mathbb{C}, \quad \big((u,v),(x,y)\big) \longmapsto ux + vy$$

will not do, for neither does Lemma 13.12 (i) hold, nor do we obtain a norm, since, by this definition, $\langle\!\langle (i,1),(i,1)\rangle\!\rangle = 0$, even though $(1,i) \neq (0,0)$.

If, on the other hand, we define

$$\langle\!\langle\ ,\ \rangle\!\rangle\colon \mathbb{C}^2 \times \mathbb{C}^2 \longrightarrow \mathbb{C}, \quad \big((u,v),(x,y)\big) \longmapsto u\overline{x} + v\overline{y},$$

then all our desired results hold, except for Lemma 13.12 (ii), which must be replaced by

$$\text{(ii)}' \quad \langle\!\langle (x,y),(u,v)\rangle\!\rangle = \overline{\langle\!\langle (u,v),(x,y)\rangle\!\rangle}$$

We take this as the model for our definition, and the characteristic properties serve as axioms.

Definition 13.14. Let $\mathbb{F}$ be a subfield of $\mathbb{C}$. An *inner product* on the $\mathbb{F}$-vector space, V, is a function

$$\langle\!\langle\ ,\ \rangle\!\rangle\colon V \times V \longrightarrow \mathbb{F}$$

such that for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ and $\lambda \in \mathbb{F}$

IP1 $\langle\!\langle \mathbf{u}, \mathbf{u}\rangle\!\rangle \geq 0$[1], with equality when, and only when, $\mathbf{u} = \mathbf{0}_V$;

IP2 $\langle\!\langle \mathbf{v}, \mathbf{u}\rangle\!\rangle = \overline{\langle\!\langle \mathbf{u}, \mathbf{v}\rangle\!\rangle}$;

IP3 $\langle\!\langle \lambda\mathbf{u}, \mathbf{v}\rangle\!\rangle = \lambda\langle\!\langle \mathbf{u}, \mathbf{v}\rangle\!\rangle$;

IP4 $\langle\!\langle \mathbf{u} + \mathbf{v}, \mathbf{w}\rangle\!\rangle = \langle\!\langle \mathbf{u}, \mathbf{w}\rangle\!\rangle + \langle\!\langle \mathbf{v}, \mathbf{w}\rangle\!\rangle$.

Observation 13.15. When $\mathbb{F}$ is a subfield of $\mathbb{R}$, condition **IP2** reduces to Lemma 13.12 (ii).

Example 13.16. Take $\mathbb{F} = \mathbb{C}$ and $V = \mathbb{C}^n$ $(n \in \mathbb{N} \setminus \{0\})$ Then

$$\langle\!\langle (w_1,\ldots,w_n),(z_1,\ldots,z_n)\rangle\!\rangle := \sum_{j=1}^{n} w_j\overline{z_j}$$

defines an inner product, called the *Euclidean inner product*. It (and its restriction to $\mathbb{F}^n$ for a subfield, $\mathbb{F}$, of $\mathbb{C}$) is frequently also referred to as the *standard inner product* on $\mathbb{C}^n$ or $\mathbb{F}^n$.

Example 13.17. Take $\mathbb{F} = \mathbb{C}$ and $V := \{f\colon [0,1] \longrightarrow \mathbb{C} \mid f \text{ is continuous}\}$ Then

$$\langle\!\langle f, g\rangle\!\rangle := \int_0^1 f(t)\overline{g(t)}dt$$

defines an inner product on V, giving rise to an inner product space which is closely related to the space $\mathcal{L}^2([0,1])$ studied in functional analysis as well as in measure and integration theory. The reader will also meet it and related spaces in statistics, the theory of differential equations and theoretical physics.

The verification that the inner product axioms hold requires a little of the theory of functions of complex variables, namely, that we may write $f(t)$ as $x(t) + iy(t)$ (with $i^2 = -1$) and that

$$\int_0^1 f(t)dt := \int_0^1 x(t)dt + i\int_0^1 y(t)dt.$$

[1]Here we use the convention for complex numbers that when we write $z \geq 0$, we assert that z is, in fact, a real number.

The prooerty of inner product spaces crucial for the definition of the angle between two vectors is the Cauchy-Schwarz Inequality, which we establish next.

Theorem 13.18 (**Cauchy-Schwarz Inequality**). *Let $\langle\!\langle\ ,\ \rangle\!\rangle$ be an inner product on the vector space V over the subfield $\mathbb{F}$ of $\mathbb{C}$. Then for all $\mathbf{u}, \mathbf{v} \in V$*

$$|\langle\!\langle \mathbf{u}, \mathbf{v} \rangle\!\rangle| \leq \sqrt{\langle\!\langle \mathbf{u}, \mathbf{u} \rangle\!\rangle}\,\sqrt{\langle\!\langle \mathbf{v}, \mathbf{v} \rangle\!\rangle}\,.$$

Proof. Take $\mathbf{u}, \mathbf{v} \in V$ and $\alpha, \beta \in \mathbb{F}$. Then

$$\begin{aligned} 0 &\leq \langle\!\langle \alpha\mathbf{u} - \beta\mathbf{v}, \alpha\mathbf{u} - \beta\mathbf{v} \rangle\!\rangle \\ &= \alpha\overline{\alpha}\langle\!\langle \mathbf{u}, \mathbf{u} \rangle\!\rangle - \alpha\overline{\beta}\langle\!\langle \mathbf{u}, \mathbf{v} \rangle\!\rangle - \beta\overline{\alpha}\langle\!\langle \mathbf{v}, \mathbf{u} \rangle\!\rangle + \beta\overline{\beta}\langle\!\langle \mathbf{v}, \mathbf{v} \rangle\!\rangle \end{aligned}$$

Put $\alpha := \langle\!\langle \mathbf{v}, \mathbf{v} \rangle\!\rangle$ and $\beta := \langle\!\langle u, \mathbf{v} \rangle\!\rangle$. Then, since $\langle\!\langle \mathbf{v}, \mathbf{v} \rangle\!\rangle \in \mathbb{R}$ and $z\overline{z} = |z|^2$,

$$\begin{aligned} 0 &\leq \langle\!\langle \mathbf{v}, \mathbf{v} \rangle\!\rangle\overline{\langle\!\langle \mathbf{v}, \mathbf{v} \rangle\!\rangle}\langle\!\langle \mathbf{u}, \mathbf{u} \rangle\!\rangle - \langle\!\langle \mathbf{v}, \mathbf{v} \rangle\!\rangle\overline{\langle\!\langle \mathbf{u}, \mathbf{v} \rangle\!\rangle}\langle\!\langle \mathbf{u}, \mathbf{v} \rangle\!\rangle - \langle\!\langle \mathbf{u}, \mathbf{v} \rangle\!\rangle\overline{\langle\!\langle \mathbf{v}, \mathbf{v} \rangle\!\rangle}\langle\!\langle \mathbf{v}, \mathbf{u} \rangle\!\rangle\langle\!\langle \mathbf{u}, \mathbf{v} \rangle\!\rangle\overline{\langle\!\langle \mathbf{u}, \mathbf{v} \rangle\!\rangle}\langle\!\langle \mathbf{v}, \mathbf{v} \rangle\!\rangle \\ &= \langle\!\langle \mathbf{v}, \mathbf{v} \rangle\!\rangle\left(\langle\!\langle \mathbf{u}, \mathbf{u} \rangle\!\rangle\langle\!\langle \mathbf{v}, \mathbf{v} \rangle\!\rangle - |\langle\!\langle \mathbf{u}, \mathbf{v} \rangle\!\rangle|^2\right) \end{aligned}$$

Now $\langle\!\langle \mathbf{v}, \mathbf{v} \rangle\!\rangle = 0$ if and only if $\mathbf{v} = \mathbf{0}_V$, in which case $|\langle\!\langle \mathbf{u}, \mathbf{v} \rangle\!\rangle| = 0 = \sqrt{\langle\!\langle \mathbf{u}, \mathbf{u} \rangle\!\rangle}\,\sqrt{\langle\!\langle \mathbf{v}, \mathbf{v} \rangle\!\rangle}\,$.
Otherwise, $\langle\!\langle \mathbf{v}, \mathbf{v} \rangle\!\rangle > 0$, and so, $|\langle\!\langle \mathbf{u}, \mathbf{v} \rangle\!\rangle|^2 \leq \langle\!\langle \mathbf{u}, \mathbf{u} \rangle\!\rangle\langle\!\langle \mathbf{v}, \mathbf{v} \rangle\!\rangle$. □

This allows us to define the *angle* between two vectors in an inner product space. We do this only for vector spaces when the scalars are real numbers in order to avoid questions about the meaning of "complex angles".

Definition 13.19. Let $\langle\!\langle\ ,\ \rangle\!\rangle$ be an inner product on the vector space V over the subfield $\mathbb{F}$ of $\mathbb{R}$. For $\mathbf{u}, \mathbf{v} \in V \setminus \{\mathbf{0}_V\}$, the *angle between* $\mathbf{u}$ *and* $\mathbf{v}$, $\angle\mathbf{uv}$, is the unique real number $\theta \in [0, \pi]$ with

$$\cos\theta = \frac{\langle\!\langle \mathbf{u}, \mathbf{v} \rangle\!\rangle}{\sqrt{\langle\!\langle \mathbf{u}, \mathbf{u} \rangle\!\rangle\langle\!\langle \mathbf{v}, \mathbf{v} \rangle\!\rangle}}$$

We show that each inner product space is a normed vector space in a natural way.

Definition 13.20. Let $\langle\!\langle\ ,\ \rangle\!\rangle$ be an inner product on the vector space, V, over the subfield, $\mathbb{F}$, of $\mathbb{C}$. Define

$$\|\ \|_{\langle\!\langle\ ,\ \rangle\!\rangle} : V \longrightarrow \mathbb{F}, \quad \mathbf{v} \longmapsto \sqrt{\langle\!\langle \mathbf{v}, \mathbf{v} \rangle\!\rangle}$$

Theorem 13.21. *If $\langle\!\langle\ ,\ \rangle\!\rangle$ is an inner product on the vector space, V, over the field $\mathbb{F}$, then $\|\ \|_{\langle\!\langle\ ,\ \rangle\!\rangle}$ is a norm on V.*

Proof. Take $\mathbf{u}, \mathbf{v} \in V$ and $\lambda \in \mathbb{F}$.

$$\begin{aligned} \|\mathbf{u}\|_{\langle\!\langle\ ,\ \rangle\!\rangle} = 0 \ &\text{if and only if } \ \|\mathbf{u}\|^2_{\langle\!\langle\ ,\ \rangle\!\rangle} = 0 \\ &\text{if and only if } \ \langle\!\langle \mathbf{u}, \mathbf{u} \rangle\!\rangle = 0 \\ &\text{if and only if } \ \mathbf{u} = \mathbf{0}_V, && \text{by } \mathbf{IP1}\text{, verifying } \mathbf{N1}. \end{aligned}$$

$$\begin{aligned} \|\lambda\mathbf{u}\|^2_{\langle\!\langle\ ,\ \rangle\!\rangle} &= \langle\!\langle \lambda\mathbf{u}, \lambda\mathbf{u} \rangle\!\rangle \\ &= \lambda\overline{\lambda}\langle\!\langle \mathbf{u}, \mathbf{u} \rangle\!\rangle && \text{by } \mathbf{IP2} \text{ and } \mathbf{IP3} \\ &= |\lambda|^2\|\mathbf{u}\|^2_{\langle\!\langle\ ,\ \rangle\!\rangle}, && \text{verifying } \mathbf{N2}. \end{aligned}$$

$$
\begin{aligned}
\|\mathbf{u}+\mathbf{v}\|^2_{\langle\!\langle\ ,\ \rangle\!\rangle} &:= \langle\!\langle \mathbf{u}+\mathbf{v}, \mathbf{u}+\mathbf{v}\rangle\!\rangle \\
&= \langle\!\langle \mathbf{u},\mathbf{u}\rangle\!\rangle + \langle\!\langle \mathbf{u},\mathbf{v}\rangle\!\rangle + \overline{\langle\!\langle \mathbf{u},\mathbf{v}\rangle\!\rangle} + \langle\!\langle \mathbf{v},\mathbf{v}\rangle\!\rangle && \text{by } \mathbf{IP2} \text{ and } \mathbf{IP4} \\
&= \langle\!\langle \mathbf{u},\mathbf{u}\rangle\!\rangle + 2\mathbb{R}e(\langle\!\langle \mathbf{u},\mathbf{v}\rangle\!\rangle) + \langle\!\langle \mathbf{v},\mathbf{v}\rangle\!\rangle \\
&\leq \langle\!\langle \mathbf{u},\mathbf{u}\rangle\!\rangle + 2|\langle\!\langle \mathbf{u},\mathbf{v}\rangle\!\rangle| + \langle\!\langle \mathbf{v},\mathbf{v}\rangle\!\rangle && \text{as } \mathbb{R}e(z) \leq |z| \\
&\leq \langle\!\langle xu,\mathbf{u}\rangle\!\rangle + 2\sqrt{\langle\!\langle \mathbf{u},\mathbf{u}\rangle\!\rangle\langle\!\langle \mathbf{v},\mathbf{v}\rangle\!\rangle} + \langle\!\langle \mathbf{v},\mathbf{v}\rangle\!\rangle && \text{by Theorem 13.18} \\
&= \left(\sqrt{\langle\!\langle \mathbf{u},\mathbf{u}\rangle\!\rangle} + \sqrt{\langle\!\langle \mathbf{v},\mathbf{v}\rangle\!\rangle}\right)^2 \\
&=: \left(\|\mathbf{u}\|_{\langle\!\langle\ ,\ \rangle\!\rangle} + \|\mathbf{v}\|_{\langle\!\langle\ ,\ \rangle\!\rangle}\right)^2, && \text{verifying } \mathbf{N3}.
\end{aligned}
$$

□

13.3 Exercises

Exercise 13.1. Let $\mathbb{F}$ be a sub-field of $\mathbb{C}$. Verify that the function

$$\|\ \|\colon \mathbb{F} \longrightarrow \mathbb{R}^+_0, \quad z \longmapsto |z|,$$

where $|z|$ denotes the modulus of the complex number z, defines a norm on $\mathbb{F}$.

Exercise 13.2. Let $\mathbb{F}$ be a sub-field of $\mathbb{C}$. Take $\mathbb{F}^n$ with its standard vector space structure over $\mathbb{F}$. Verify that the following function defines a norm on $\mathbb{F}^n$.

$$\|\ \|\colon \mathbb{F}^n \longrightarrow \mathbb{R}^+_0, \quad (x_1,\ldots,x_n) \longmapsto \left(\sum_{j=1}^{n} |x|_j^2\right)^{\frac{1}{2}},$$

Exercise 13.3. Verify that multiplication of real numbers defines an inner product on $\mathbb{R}$.

Exercise 13.4. (a) $\mathbf{M}(m\times n;\mathbb{R})$ is a real vector space with respect to matrix addition and multiplication of a matrix by a constant. Show that

$$\langle\!\langle\ ,\ \rangle\!\rangle_M\colon \mathbf{M}(m\times n;\mathbb{R})\times\mathbf{M}(m\times n;\mathbb{R}) \longrightarrow \mathbb{R}, \quad (\underline{\mathbf{A}},\underline{\mathbf{B}}) \longmapsto \mathrm{tr}(\underline{\mathbf{A}}^t\underline{\mathbf{B}})$$

defines a real inner product on $\mathbf{M}(m\times n;\mathbb{R})$.

(b) Show that

$$\varphi\colon \mathbf{M}(m\times n;\mathbb{R}) \longrightarrow \mathbb{R}^{mn}, \quad [a_{ij}]_{m\times n} \longmapsto (x_{11},\ldots,x_{mn}),$$

where $x_{(i-1)n+j} := a_{(i-1)n\,j}$ defines an isomorphism of real vector spaces.

(c) Show that for all $\underline{\mathbf{A}},\underline{\mathbf{B}} \in \mathbf{M}(m\times n;\mathbb{R})$,

$$\langle\!\langle \underline{\mathbf{A}},\underline{\mathbf{B}}\rangle\!\rangle_M = \langle\!\langle \varphi(\underline{\mathbf{A}}),\varphi(\underline{\mathbf{B}})\rangle\!\rangle,$$

where we have taken the Euclidean inner product on $\mathbb{R}^{mn}$. [Such an isomorphism is called a *linear isometry*.]

Exercise 13.5. Prove Lemma 13.12 on page 177.

Exercise 13.6. Prove Lemma 13.13 on page 177.

Mathematical Knowledge adds a manly Vigour to the Mind, frees it from Prejudice, Credulity, and Superstition.

John Arbuthnot

Chapter 14

Orthogonality

Definition 13.19 on page 179 introduced the angle between two vectors in an inner product space, for vector spaces all of whose scalars are real numbers.

In particular, two (non-zero) vectors, $\mathbf{u}$ and $\mathbf{v}$, are perpendicular to each other, or *orthogonal* if the angle between them is a right angle. Since $\cos\frac{\pi}{2} = 0$, this is equivalent to $\langle\langle \mathbf{u}, \mathbf{v} \rangle\rangle = 0$.

Observe that this is expressed purely in terms of the inner product, without appeal to the notion of angle, so we have no need to restrict ourselves to real scalars. Hence we can define orthogonality in any inner product space.

Definition 14.1. Let $\langle\langle\ ,\ \rangle\rangle$ be an inner product on the vector space V over the subfield $\mathbb{F}$ of $\mathbb{C}$. The vectors $\mathbf{u}, \mathbf{v} \in V$ are *orthogonal* if and only if $\langle\langle \mathbf{u}, \mathbf{v} \rangle\rangle = 0$.

Before investigating orthogonality in any detail, we provide a geometric application.

Orthogonality generalises the notion of a right angle, which is central to Pythagoras' Theorem in geometry. We prove a generalised Pythagoras' Theorem.

Theorem 14.2 (Pythagoras' Theorem). *Let $\|\quad\|$ be the norm induced on the V by the inner product $\langle\langle\ ,\ \rangle\rangle$.*

If $\mathbf{u}, \mathbf{v} \in V$ are orthogonal, then

$$\|\mathbf{u} + \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2.$$

Proof. Since $\mathbf{u}, \mathbf{v}$ are orthogonal, $\langle\langle \mathbf{u}, \mathbf{v} \rangle\rangle = 0$. Thus

$$\begin{aligned} \|\mathbf{u} + \mathbf{v}\|^2 &:= \langle\langle \mathbf{u} + \mathbf{v}, \mathbf{u} + \mathbf{v} \rangle\rangle \\ &= \langle\langle \mathbf{u}, \mathbf{u} \rangle\rangle + \langle\langle \mathbf{u}, \mathbf{v} \rangle\rangle + \langle\langle \mathbf{v}, \mathbf{u} \rangle\rangle + \langle\langle \mathbf{v}, \mathbf{v} \rangle\rangle \\ &= \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 \qquad \text{by orthogonality} \end{aligned}$$

□

The next lemma is an immediate consequence of the axioms for inner products.

Lemma 14.3. *Let $(V, \langle\langle\ ,\ \rangle\rangle)$ be an inner product space. Then $\mathbf{0}_V$ is orthogonal to every $\mathbf{v} \in V$.*

Another easy consequence is that non-zero orthogonal vectors must be linearly independent.

Theorem 14.4. *Let $(V, \langle\!\langle\ ,\ \rangle\!\rangle)$ be an inner product space. Take $\{\mathbf{v}_i \mid i \in I\} \subseteq V \setminus \{\mathbf{0}_V\}$ such that $\langle\!\langle \mathbf{v}_i, \mathbf{v}_j \rangle\!\rangle$ whenever $i \neq j$.*

Then $\{\mathbf{v}_i \mid i \in I\}$ is a set of linearly independent vectors.

Proof. Suppose $\sum \alpha_i \mathbf{v}_i = \mathbf{0}_V$ for $\alpha_i \in \mathbb{F}$ $(i \in I)$. Then, for each $j \in I$

$$\begin{aligned} 0 &= \langle\!\langle \sum_{i \in I} \alpha_i \mathbf{v}_{ii}, \mathbf{v}_j \rangle\!\rangle \\ &= \sum_{i \in I} \alpha_i \langle\!\langle \mathbf{v}_i, \mathbf{v}_j \rangle\!\rangle \\ &= \alpha_j \langle\!\langle \mathbf{v}_j, \mathbf{v}_j \rangle\!\rangle \qquad \text{as } \langle\!\langle \mathbf{v}_i, \mathbf{v}_j \rangle\!\rangle = 0 \text{ unless } i = j. \end{aligned}$$

But $\langle\!\langle \mathbf{v}_j, \mathbf{v}_j \rangle\!\rangle \neq 0$ since $\mathbf{v}_j \neq \mathbf{0}_V$.

Hence $\alpha_j = 0$. □

Unit vectors, that is, vectors whose length (norm) is 1, play a special rôle, especially when they are mutually orthogonal.

Definition 14.5. Let $\langle\!\langle\ ,\ \rangle\!\rangle$ be an inner product on the vector space V. Then the vectors $\mathbf{u}_i$ $(i \in I)$ are *orthonormal* if and only if $\langle\!\langle \mathbf{u}_i, \mathbf{u}_j \rangle\!\rangle = \delta_{ij}$, where is the *Kronecker delta*, defined by

$$\delta_{ij} := \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases}$$

The basis $\mathcal{B} = \{\mathbf{v}_i \mid i \in I\}$ is an *orthonormal basis* if and only if the vectors in $\mathcal{B}$ are orthonormal.

Example 14.6. $V :\ = \{f : [0, 2\pi] \longrightarrow \mathbb{R} \mid f \text{ is continuous}\}$ is a real vector space and

$$\langle\!\langle\ ,\ \rangle\!\rangle : V \times V \longrightarrow \mathbb{R}, \quad (f, g) \longmapsto \frac{1}{\pi} \int_0^{2\pi} f(t) g(t) dt$$

defines an inner product on V.

For $n \in \mathbb{N} \setminus \{0\}$ define

$$\begin{aligned} c_n &: [0, 2\pi] \longrightarrow \mathbb{R}, \quad x \longmapsto \cos(nx) \\ s_n &: [0, 2\pi] \longrightarrow \mathbb{R}, \quad x \longmapsto \sin(nx) \end{aligned}$$

Then $\{c_n, s_n \mid n = 1, 2, \ldots\}$ is a set of orthonormal vectors in V. (The verification is left as an exercise.)

Orthonormal bases are particularly convenient for numerous purposes. For example, the coordinates of any vector with respect to an orthonormal basis can be computed directly, using only the inner product.

Theorem 14.7. *Let $\{\mathbf{e}_i \mid i \in I\}$ be an orthonormal basis for $(V, \langle\!\langle\ ,\ \rangle\!\rangle)$. Given $\mathbf{v} \in V$,*

$$\mathbf{v} = \sum_{i \in I} \langle\!\langle \mathbf{v}, \mathbf{e}_i \rangle\!\rangle \mathbf{e}_i.$$

Proof. Take $\mathbf{v} \in V$.

Since $\{\mathbf{e}_i \mid i \in I\}$ is a basis for V, there are uniquely determined $\alpha_i \in \mathbb{F}$ $(i \in I)$ with $\mathbf{v} = \sum \alpha_i \mathbf{e}_i$.

For $j \in I$,

$$\begin{aligned}
\langle\langle \mathbf{v}, \mathbf{e}_j \rangle\rangle &= \langle\langle \sum_{i\in I} \alpha_i \mathbf{e}_i, \mathbf{e}_j \rangle\rangle \\
&= \sum_{i\in I} \alpha_i \langle\langle \mathbf{e}_i, \mathbf{e}_j \rangle\rangle \\
&= \alpha_j && \text{since } \langle\langle \mathbf{e}_i, \mathbf{e}_j \rangle\rangle = \delta_{ij}.
\end{aligned}$$

□

Corollary 14.8. *Let $\{\mathbf{e}_i \mid i \in I\}$ be an orthonormal basis for the inner product space V. Given $\mathbf{v} \in V$,*

$$\|\mathbf{v}\|^2_{\langle\langle\ ,\ \rangle\rangle} = \sum_{i\in I} |\langle\langle \mathbf{v}, \mathbf{e}_i \rangle\rangle|^2.$$

In particular, if $\mathbf{v} = \sum_{i\in I} \alpha_i \mathbf{e}_i$, then

$$\|\mathbf{v}\|^2_{\langle\langle\ ,\ \rangle\rangle} = \sum_{i\in I} |\alpha_i|^2.$$

Proof. Take $\mathbf{v} \in V$.

$$\begin{aligned}
\|\mathbf{v}\|^2_{\langle\langle\ ,\ \rangle\rangle} &= \langle\langle \mathbf{v}, \mathbf{v} \rangle\rangle \\
&= \langle\langle \sum_{i\in I} \langle\langle \mathbf{v}, \mathbf{e}_i \rangle\rangle \mathbf{e}_i, \sum_{j\in I} \langle\langle \mathbf{v}, \mathbf{e}_j \rangle\rangle \mathbf{e}_j \rangle\rangle && \text{by Theorem 14.7} \\
&= \sum_{i,j\in I} \langle\langle \mathbf{v}, \mathbf{e}_i \rangle\rangle \overline{\langle\langle \mathbf{v}, \mathbf{e}_j \rangle\rangle} \langle\langle \mathbf{e}_i, \mathbf{e}_j \rangle\rangle \\
&= \sum_{i\in I} \langle\langle \mathbf{v}, \mathbf{e}_i \rangle\rangle \overline{\langle\langle \mathbf{v}, \mathbf{e}_i \rangle\rangle} && \text{since } \langle\langle \mathbf{e}_i, \mathbf{e}_j \rangle\rangle = \delta_{ij} \\
&= \sum_{i\in I} |\langle\langle \mathbf{v}, \mathbf{e}_i \rangle\rangle|^2.
\end{aligned}$$

□

Since orthonormal bases are so useful and important, it is particularly satisfying that they can always be constructed. Given any basis whatsoever for an inner product space, there is an algorithm for constructing an orthonormal basis from it.

Theorem 14.9 (**Gram-Schmidt Orthonormalisation**)**.** *Every finitely generated inner product space admits an orthonormal basis.*[1]

Proof. Let $\langle\langle\ ,\ \rangle\rangle$ be a inner product on the finitely generated vector space V over $\mathbb{F}$. Given a basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ of V, we construct an orthonormal basis $\{\mathbf{e}_1, \ldots, \mathbf{e}_m\}$ by means of a recursive procedure (algorithm), called the *Gram-Schmidt procedure*.

Since $\{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ is a basis, $\mathbf{u}_1 \neq \mathbf{0}_V$. Put

$$\mathbf{e}_1 := \frac{\mathbf{u}_1}{\|\mathbf{u}_1\|}.$$

[1] There is an extension of this to inner product spaces which are not finitely generated. You will meet such problems in measure and integration theory, and in functional analysis, for example. We do not pursue such matters further here.

Then $\mathbf{u}_1 = \|\mathbf{u}_1\|\mathbf{e}_1$.

Hence both $\{\mathbf{u}_1\}$ and $\{\mathbf{e}_1\}$ are sets of linearly independent vectors, $\|\mathbf{e}_1\| = 1$ and $\langle \mathbf{u}_1 \rangle = \langle \mathbf{e}_1 \rangle$.

Thus, $\{\mathbf{e}_1\}$ is an orthonormal basis for the subspace of V generated by $\mathbf{v}_1$.

Suppose that orthonormal vectors $\mathbf{e}_1, \dots \mathbf{e}_j$ have been constructed for $1 \leq j < m$ such that $\langle \mathbf{e}_1, \dots, \mathbf{e}_j \rangle = \langle \mathbf{u}_1, \dots, \mathbf{u}_j \rangle$. In other words, $\{\mathbf{e}_1, \dots \mathbf{e}_j\}$ is an orthonormal basis for the subspace of V generated by $\{\mathbf{v}_1, \dots \mathbf{v}_j\}$.

Put

$$\mathbf{v}_{j+1} := \mathbf{u}_{j+1} - \sum_{i=1}^{j} \langle\langle \mathbf{u}_{j+1}, \mathbf{e}_i \rangle\rangle \mathbf{e}_i \tag{14.1}$$

Then for each $k \leq j$,

$$\begin{aligned}
\langle\langle \mathbf{v}_{j+1}, \mathbf{e}_k \rangle\rangle &= \langle\langle \mathbf{u}_{j+1} - \sum_{i=1}^{j} \langle\langle \mathbf{u}_{j+1}, \mathbf{e}_i \rangle\rangle \mathbf{e}_i, \mathbf{e}_k \rangle\rangle \\
&= \langle\langle \mathbf{u}_{j+1}, \mathbf{e}_k \rangle\rangle - \sum_{i=1}^{j} \langle\langle \mathbf{u}_{j+1}, \mathbf{e}_i \rangle\rangle \langle\langle \mathbf{e}_i, \mathbf{e}_k \rangle\rangle && \text{by linearity in the first variable} \\
&= \langle\langle \mathbf{u}_{j+1}, \mathbf{e}_k \rangle\rangle - \langle\langle \mathbf{u}_{j+1}, \mathbf{e}_k \rangle\rangle && \text{as } \langle\langle \mathbf{e}_i, \mathbf{e}_k \rangle\rangle = \delta ik \\
&= 0
\end{aligned}$$

Hence the vectors $\mathbf{e}_1, \dots, \mathbf{e}_j, \mathbf{v}_{j+1}$ are mutually orthogonal.

Moreover, $\mathbf{v}_{j+1} \neq \mathbf{0}_V$. For otherwise, by Equation 14.1, $\mathbf{u}_{j+1} = \sum_{i=1}^{j} \langle\langle \mathbf{u}_{j+1}, \mathbf{e}_{ii} \rangle\rangle \mathbf{e}_i$.

Since $\langle \mathbf{e}_1, \dots, \mathbf{e}_j \rangle = \langle \mathbf{u}_1, \dots, \mathbf{u}_j \rangle$, this would contradict the linear independence of $\mathbf{u}_1, \dots, \mathbf{u}_{j+1}$.

We may therefore put

$$\mathbf{e}_{j+1} := \frac{1}{\|\mathbf{v}_{j+1}\|} \mathbf{v}_{j+1}.$$

This clearly renders $\mathbf{e}_1, \dots \mathbf{e}_j$ orthonormal, and hence, by Theorem 14.4 on page 182, linearly independent. Thus, by Theorem 8.11 on page 90, $\langle \mathbf{e}_1, \dots, \mathbf{e}_{j+1} \rangle = \langle \mathbf{u}_1, \dots \mathbf{u}_{j+1} \rangle$. In particular, $\{\mathbf{e}_1, \dots \mathbf{e}_m\}$ is an orthonormal basis for V. □

14.1 Orthogonal Complements

Definition 14.10. Let $\langle\langle\ ,\ \rangle\rangle$ be an inner product on the vector space V. The *orthogonal complement*, $S^\perp$, of $S \subseteq V$ is the set of all vectors in V, orthogonal to every vector in S:

$$S^\perp := \{\mathbf{v} \in V \mid \langle\langle \mathbf{v}, \mathbf{x} \rangle\rangle = 0 \text{ for all } \mathbf{x} \in S\}$$

Theorem 14.11. *Let S be a subset of the inner product space $(V, \langle\langle\ ,\ \rangle\rangle)$. Then*

(i) $S^\perp$ *is a vector subspace of* V.

(ii) *If* $S \subseteq T$, *then* $T^\perp \subseteq S^\perp$

(iii) $S^\perp = \langle S \rangle^\perp$

(iv) $\langle S \rangle \leq (S^\perp)^\perp$

If, in addition, V is finitely generated,

(v) $V = \langle S\rangle \oplus \langle S\rangle^{\perp}$

(vi) $(S^{\perp})^{\perp} = \langle S\rangle$.

Proof. (i) Take $\mathbf{u}, \mathbf{v} \in S^{\perp}, \alpha, \beta \in \mathbb{F}$ and $\mathbf{x} \in S$. Then

$$\langle\langle \alpha\mathbf{u} + \beta\mathbf{v}, \mathbf{x}\rangle\rangle = \alpha\langle\langle \mathbf{u}, \mathbf{x}\rangle\rangle + \beta\langle\langle \mathbf{v}, \mathbf{x}\rangle\rangle = 0,$$

showing that $\alpha\mathbf{u} + \beta\mathbf{v} \in S^{\perp}$.

(ii) Take $\mathbf{v} \in T^{\perp}$ and $\mathbf{x} \in S$. Since $S \subseteq T$, $\mathbf{x} \in T$, and so $\langle\langle \mathbf{v}, \mathbf{x}\rangle\rangle = 0$, whence $\mathbf{v} \in S^{\perp}$.

(iii) Since $S \subseteq \langle S\rangle$, we know form (ii) that $\langle S\rangle^{\perp} \subseteq S^{\perp}$.

For the reverse inclusion, take $\mathbf{v} \in S^{\perp}$ and $\mathbf{x} \in \langle S\rangle$.

Then $\mathbf{x} = \alpha_1\mathbf{x}_1 + \cdots + \alpha_k\mathbf{x}_k$ for some $\alpha_1, \ldots, \alpha_k \in \mathbb{F}$ and $\mathbf{x}_1, \ldots, \mathbf{x}_k \in S$. Thus

$$\langle\langle \mathbf{x}, \mathbf{v}\rangle\rangle = \langle\langle \sum_{j=1}^{k} \alpha_j\mathbf{x}_j, \mathbf{v}\rangle\rangle = \sum_{j=1}^{k} \alpha_j\langle\langle \mathbf{x}_j, \mathbf{v}\rangle\rangle = \sum_{j=1}^{k} \alpha_j 0 = 0,$$

(iv) Take $\mathbf{x} \in \langle S\rangle$ and $\mathbf{v} \in S^{\perp} = \langle S\rangle^{\perp}$.

Then $\langle\langle \mathbf{x}, \mathbf{v}\rangle\rangle = \langle\langle \mathbf{v}, \mathbf{x}\rangle\rangle = 0$, whence $\mathbf{x} \in (S^{\perp})^{\perp}$.

Now suppose that V is finitely generated.

(v) First, observe that for any subspace W of V, if $\mathbf{v} \in W \cap W^{\perp}$, then $\langle\langle \mathbf{v}, \mathbf{v}\rangle\rangle = 0$, whence $\mathbf{v} = \mathbf{0}_V$. Hence $W \cap W^{\perp} = \{\mathbf{0}_V\}$.

It is therefore sufficient to show that $V = W + W^{\perp}$ and then take $W := \langle S\rangle$.

Let $\{\mathbf{e}, \ldots, \mathbf{e}_k\}$ be an orthonormal basis for W, take $\mathbf{v} \in V$.

Put $\mathbf{x} := \langle\langle \mathbf{v}, \mathbf{e}_1\rangle\rangle\mathbf{e}_1 + \cdots + \langle\langle \mathbf{v}, \mathbf{e}_k\rangle\rangle\mathbf{e}_k$ and $\mathbf{y} := \mathbf{v} - \mathbf{x}$.

Clearly $\mathbf{v} = \mathbf{x} + \mathbf{y}$, with $\mathbf{x} \in W$.

To show that $\mathbf{y} \in W^{\perp}$, note that

$$\begin{aligned}
\langle\langle \mathbf{y}, \mathbf{e}_i\rangle\rangle &= \langle\langle \mathbf{v}, \mathbf{e}_i\rangle\rangle - \langle\langle \mathbf{x}, \mathbf{e}_i\rangle\rangle \\
&= \langle\langle \mathbf{v}, \mathbf{e}_i\rangle\rangle - \langle\langle \sum_{j=1}^{k} \langle\langle \mathbf{v}, \mathbf{e}_j\rangle\rangle\mathbf{e}_j, \mathbf{e}_i\rangle\rangle \\
&= \langle\langle \mathbf{v}, \mathbf{e}_i\rangle\rangle - \sum_{j=1}^{k} \langle\langle \mathbf{v}, \mathbf{e}_j\rangle\rangle\langle\langle \mathbf{e}_j, \mathbf{e}_i\rangle\rangle && \text{by linearity in the first variable} \\
&= \langle\langle \mathbf{v}, \mathbf{e}_i\rangle\rangle - \langle\langle \mathbf{v}, \mathbf{e}_i\rangle\rangle \\
&= 0 && \text{by orthonormality}
\end{aligned}$$

(vi) Using (v) twice, $V = W \oplus W^{\perp} = W^{\perp} \oplus (W^{\perp})^{\perp}$.

Hence $W \cong (W^{\perp})^{\perp}$.

By (iv), $W \leq (W^{\perp})^{\perp}$, so since V, and so also $(W^{\perp})^{\perp}$, is finitely generated

$$W = (W^{\perp})^{\perp}$$

□

The following example shows that (iv) and (v) do not hold without some assumption, such as the vector space in question being finitelygenerated.

Example 14.12. For our real inner product space, we take $V := \mathbb{R}[t]$, the set of all real polynomials in the indeterminate t.

$\mathbb{R}[t]$ is, plainly, not finitely generated, as $\{t^n \mid n \in \mathbb{N}\}$ is an infinite set of linearly independent vectors in V.

We use the inner product $\langle\!\langle\ ,\ \rangle\!\rangle$ where

$$\langle\!\langle p, q \rangle\!\rangle := \int_0^1 p(x)q(x)\,dx$$

As subspace we take

$$W := \{p \mid p(0) = 0\}$$

Take any $h \in W^\perp$. Then

$$\begin{aligned} \|th\|^2 &= \int_0^1 \big(xh(x)\big)^2\,dx \\ &= \int_0^1 h(x)x^2h(x)\,dx \\ &= \langle\!\langle h, t^2h \rangle\!\rangle \\ &= 0 \qquad \text{since } t^2h \in W \text{ and } h \in W^\perp \end{aligned}$$

Thus th is the 0 polynomial, whence h must be the zero polynomial.

Consequently $W^\perp = \{\mathbf{0}_V\}$.

It follows that $(W^\perp)^\perp = V \neq W$ and also that $W + W^\perp = W \neq V$.

14.2 Orthogonal Transformations

When we studied vector spaces without considering any additional structure, the appropriate notion for comparing them was that of a linear transformation: linear transformations are precisely those functions between vector spaces over the same field which respect the vector space operations.

We have now specialised to subfields $\mathbb{F}$ of $\mathbb{C}$ in order to be able to introduce the notion of an inner product, which, as we have already seen, allows us to speak of angles and distances, thereby allowing us to "do geometry".

It therefore behooves us to introduce an appropriate notion to characterise those linear transformations between inner product spaces, that respect the additional structure.

Definition 14.13. Let $(V, \langle\!\langle\ ,\ \rangle\!\rangle_V)$ and $(W, \langle\!\langle\ ,\ \rangle\!\rangle_W)$ be inner product spaces over $\mathbb{F}$. Then the linear transformation $T\colon V \longrightarrow W$ is said to *preserve the inner product* if and only if for all $\mathbf{u}, \mathbf{v} \in V$

$$\langle\!\langle T(\mathbf{u}), T(\mathbf{v}) \rangle\!\rangle_W = \langle\!\langle \mathbf{u}, \mathbf{v} \rangle\!\rangle_V.$$

Theorem 14.14. *Let $T\colon V \longrightarrow W$ be an linear transformation of finite dimensional inner product spaces $(V, \langle\!\langle\ ,\ \rangle\!\rangle_V), (W, \langle\!\langle\ ,\ \rangle\!\rangle_W)$. Then the following are equivalent.*

(a) *T preserves the inner product.*

(b) $\|T(\mathbf{u})\|_W = \|\mathbf{u}\|_V$ *for all* $\mathbf{u} \in V$.

(c) *If* $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ *is an orthonormal basis for* V*, then* $\{T(\mathbf{e}_1), \ldots, T(\mathbf{e}_n)\}$ *is an orthonormal basis for* $im(T)$.

Proof. For ease of reading, we omit the subscripts $_V$ and $_W$, relying on the good sense of the reader to recognise which space is being considered.

(a) $\Rightarrow$ (b): Take $\mathbf{u} \in V$. Then

$$\begin{aligned} \|T(\mathbf{u})\|^2 &= \langle\langle T(\mathbf{u}), T(\mathbf{u})\rangle\rangle \\ &= \langle\langle \mathbf{u}, \mathbf{u}\rangle\rangle && \text{as } T \text{ preserves the inner product} \\ &= \|\mathbf{u}\|^2 \end{aligned}$$

(b) $\Rightarrow$ (a): Take $\mathbf{u}, \mathbf{v} \in V$ and $\alpha \in \mathbb{F}$. Then

$$\begin{aligned} \|\mathbf{u} + \alpha\mathbf{v}\|^2 &= \langle\langle \mathbf{u} + \alpha\mathbf{v}, \mathbf{u} + \alpha\mathbf{v}\rangle\rangle \\ &= \langle\langle \mathbf{u}, \mathbf{u}\rangle\rangle + \langle\langle \mathbf{u}, \alpha\mathbf{v}\rangle\rangle + \langle\langle \alpha\mathbf{v}, \mathbf{u}\rangle\rangle + \langle\langle \alpha\mathbf{v}, \alpha\mathbf{v}\rangle\rangle \\ &= \|\mathbf{u}\|^2 + \overline{\alpha}\langle\langle \mathbf{u}, \mathbf{v}\rangle\rangle + \alpha\overline{\langle\langle \mathbf{u}, \mathbf{v}\rangle\rangle} + |\alpha|^2\|\mathbf{v}\|^2 \end{aligned}$$

and, similarly

$$\begin{aligned} \|T(\mathbf{u} + \alpha\mathbf{v})\|^2 &= \|T(\mathbf{u}) + \alpha T(\mathbf{v})\|^2 \\ &= \|T(\mathbf{u})\|^2 + \overline{\alpha}\langle\langle T(\mathbf{u}), T(\mathbf{v})\rangle\rangle + \alpha\overline{\langle\langle T(\mathbf{u}), T(\mathbf{v})\rangle\rangle} + |\alpha|^2\|T(\mathbf{v}\|^2 \end{aligned}$$

Hence, if $\|T(\mathbf{x}\| = \|\mathbf{x}\|$ for every $\mathbf{x} \in V$, then, for all $\mathbf{u}, \mathbf{v} \in V$ and $\alpha \in \mathbb{F}$,

$$\overline{\alpha}\langle\langle \mathbf{u}, \mathbf{v}\rangle\rangle + \alpha\overline{\langle\langle \mathbf{u}, \mathbf{v}\rangle\rangle} = \overline{\alpha}\langle\langle T(\mathbf{u}), T(\mathbf{v})\rangle\rangle + \alpha\overline{\langle\langle T(\mathbf{u}), T(\mathbf{v})\rangle\rangle}$$

Choosing $\alpha = 1$ shows that

$$\Re\langle\langle \mathbf{u}, \mathbf{v}\rangle\rangle = \Re\langle\langle T(\mathbf{u}), T(\mathbf{v})\rangle\rangle$$

whereas choosing $\alpha = i$ shows that

$$\Im\langle\langle \mathbf{u}, \mathbf{v}\rangle\rangle = \Im\langle\langle T(\mathbf{u}), T(\mathbf{v})\rangle\rangle$$

Thus, since their real and imaginary parts agree,

$$\langle\langle T(\mathbf{u}), T(\mathbf{v})\rangle\rangle = \langle\langle \mathbf{u}, \mathbf{v}\rangle\rangle$$

(a) $\Rightarrow$ (c): Let $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ be an orthonormal basis for V. Then

$$\langle\langle T(\mathbf{e}_i), T(\mathbf{e}_j)\rangle\rangle = \langle\langle \mathbf{e}_i, \mathbf{e}_j\rangle\rangle = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases},$$

showing that $\{T(\mathbf{e}_1), \ldots, T(\mathbf{e}_n)\}$ is orthonormal

(c) $\Rightarrow$ (b) Let $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ be an orthonormal basis for V and take $\mathbf{v} \in V$.
Then

$$\mathbf{v} = \langle\langle \mathbf{v}, \mathbf{e}_1\rangle\rangle\mathbf{e}_1 + \cdots + \langle\langle \mathbf{v}, \mathbf{e}_n\rangle\rangle\mathbf{e}_n \qquad (*)$$

whence, since T is a linear transformation,

$$T(\mathbf{v}) = \langle\langle \mathbf{v}, \mathbf{e}_1\rangle\rangle T(\mathbf{e}_1) + \cdots + \langle\langle \mathbf{v}, \mathbf{e}_n\rangle\rangle T(\mathbf{e}_n) \qquad (\diamondsuit)$$

On the other, since $T(\mathbf{e}_1), \ldots, T(\mathbf{e}_n)$ iÃs an orthonormal basis for W

$$T(\mathbf{v}) = \langle\langle T(\mathbf{v}), T(\mathbf{e}_1)\rangle\rangle T(\mathbf{e}_1) + \cdots + \langle\langle T(\mathbf{v}), T(\mathbf{e}_n)\rangle\rangle T(\mathbf{e}_n) \qquad (\diamondsuit\diamondsuit)$$

By $(\diamondsuit\diamondsuit)$,

$$\|T(\mathbf{v})\|^2 = \sum_{j=1}^{n} |\langle\langle T(\mathbf{v}), T(\mathbf{e}_j)\rangle\rangle|^2$$

On the other hand, by $(\diamondsuit)$,

$$\|T(\mathbf{v})\|^2 = \sum_{j=1}^{n} |\langle\langle \mathbf{v}, \mathbf{e}_j\rangle\rangle|^2$$

But, by $(*)$,

$$\sum_{j=1}^{n} |\langle\langle \mathbf{v}, \mathbf{e}_j\rangle\rangle|^2 = \|\mathbf{v}\|^2$$

Thus, $\|T(\mathbf{v})\|^2 = \|\mathbf{v}\|^2$. □

Corollary 14.15. *Let $T\colon V \longrightarrow W$ be an linear transformation of the inner product spaces $(V, \langle\langle\ ,\ \rangle\rangle_V), (W, \langle\langle\ ,\ \rangle\rangle_W)$. Then T is injective.*

Proof. $T(\mathbf{v}) = \mathbf{0}_W$ if and only if $\|T(\mathbf{v})\|_W = 0$ if and only if $\|\mathbf{v}\|_V = 0$ if and only if $=\mathbf{0}_V$. □

The most important case, particularly from the point of view of applications, is when $W = V$ and $\langle\langle\ ,\ \rangle\rangle_W = \langle\langle\ ,\ \rangle\rangle_V$.

Definition 14.16. Let $(V, \langle\langle\ ,\ \rangle\rangle)$ be an inner product space. Then the linear transformation $T\colon V \longrightarrow V$ is said to be an *orthogonal transformation* with respect to $\langle\langle\ ,\ \rangle\rangle$ if and only if for all $\mathbf{u}, \mathbf{v} \in V$

$$\langle\langle T(\mathbf{u}), T(\mathbf{v})\rangle\rangle = \langle\langle \mathbf{u}, \mathbf{v}\rangle\rangle.$$

Observation 14.17. Traditionally, orthogonal endomorphisms of a complex inner product space are called *unitary*.

Lemma 14.18. *Let $(V, \langle\langle\ ,\ \rangle\rangle)$ be an inner product space. Then each orthogonal transformation $T\colon V \longrightarrow V$ is an isomorphism.*

Proof. If $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ is an orthonormal basis for V, then $T(\mathbf{e}_1), \ldots, T(\mathbf{e}_n)$ are orthonormal, hence linearly independent, and hence form a basis. □

14.3 Exercises

Exercise 14.1. For each of the following symmetric matrices, find an orthogonal matrix which diagonalises it.

(a) $\begin{bmatrix} 1 & 2 \\ 2 & 5 \end{bmatrix}$

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

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

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

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

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

The issue, then, is not, What is the best way to teach? but, What is mathematics really all about? ... Controversies about ... teaching cannot be resolved without confronting problems about the nature of mathematics.

Reuben Hersh

Chapter 15

Matrix Representation of Inner Products

Matrices enabled us to represent linear transformations and perform computations on them in the case of finitely generated vector spaces. They also be used to represent inner products and to carry out concrete computations in the case of finitely generated inner product spaces.

Inner products are special cases of sesqui-linear forms, and these can also be represented by matrices when dealing with finitely generated spaces. Most properties are simpler to state and prove at this generality, with the case of inner product spaces a simple, direct application.

Let U and V be finitely generated vector spaces over the sub-field $\mathbb{F}$ of $\mathbb{C}$ and $\{\mathbf{e}_1, \ldots, \mathbf{e}_m\}$, $\{\mathbf{f}_1, \ldots, \mathbf{f}_n\}$ bases for U and V respectively.

Let $\beta \colon U \times V \longrightarrow \mathbb{F}$ be sesqui-linear.

Given $\mathbf{u} \in U$ and $\mathbf{v} \in V$, there are unique $x_1, \ldots, x_m, y_1, \ldots y_n \in \mathbb{F}$ with

$$\mathbf{u} = \sum_{i=1}^{m} x_i \mathbf{e}_i \qquad \text{and} \qquad \mathbf{v} = \sum_{j=1}^{n} y_j \mathbf{f}_j,$$

so that

$$\begin{aligned} \beta(\mathbf{u}, \mathbf{v}) &= \beta(\sum_{i=1}^{m} x_i \mathbf{e}_i, \sum_{j=1}^{n} y_j \mathbf{f}_j) \\ &= \sum_{i=1}^{m} x_i \sum_{j=1}^{n} \overline{y_j} \beta(\mathbf{e}_i, \mathbf{f}_j) \\ &= \sum_{i=1}^{m} \sum_{j=1}^{n} x_i a_{ij} \overline{y_j}, \end{aligned}$$

where $a_{ij} := \beta(\mathbf{e}_i, \mathbf{f}_j) \qquad (1 \leq i \leq m,\ 1 \leq j \leq n)$

Let

$$\underline{\mathbf{x}} := \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix} \qquad \text{and} \qquad \underline{\mathbf{y}} \begin{bmatrix} y_1 \\ \vdots \\ \mathbf{y}_n \end{bmatrix}$$

be the co-ordinate vectors of $\mathbf{u}$ and $\mathbf{v}$ with respect to the given bases, and put $\underline{\mathbf{A}} := [a_{ij}]_{m\times n}$ with $a_{ij} := \beta(\mathbf{e}_i, \mathbf{f}_j)$.

Direct calculation shows that

$$\beta(\mathbf{u}, \mathbf{v}) = \underline{\mathbf{x}}^t \underline{\mathbf{A}}\, \underline{\overline{\mathbf{y}}}$$

Since our purposes require us to consider only the case $U = V$ and $\mathbf{f}_j = \underline{\mathbf{e}}_j$, we dispense with the greater generality for the rest of this chapter. But the importance and usefulness of the more general approach cannot be over-emphasised, for it marks the beginnings of tensor analysis, which has many applications in statistics, geometry, physics, chemistry and engineering.

Definition 15.1. Given a sesqui-linear form $\beta\colon V \times V \longrightarrow \mathbb{F}$, the *matrix of β with respect to the basis $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ of V* is the matrix

$$\underline{\mathbf{A}} := \Big[\beta(\mathbf{e}_i, \mathbf{e}_j)\Big]_{n\times n}$$

Since an inner product is a sesqui-linear form, we can already deduce an important fact.

Theorem 15.2. *Let $\langle\!\langle\ ,\ \rangle\!\rangle$ be an inner product on V.*

Let $\underline{\mathbf{A}}$ be the matrix of $\langle\!\langle\ ,\ \rangle\!\rangle$ with respect to the basis $\mathcal{B} = \{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$.

Then $\mathcal{B}$ is an orthogonal basis if and only if $\underline{\mathbf{A}}$ is a diagonal matrix, and $\mathcal{B}$ is an orthonormal basis if and only if $\underline{\mathbf{A}} = \underline{\mathbf{1}}_n$

The above discussion forms the basis of our computational techniques. The next theorem summarises it.

Theorem 15.3. *Let $\{\mathbf{e}_i, \ldots, \mathbf{e}_n\}$ be a basis for the vector space V.*

Let $\underline{\mathbf{x}}$ is the co-ordinate vector of $\mathbf{u} \in V$ and $\underline{\mathbf{y}}$ that of $\mathbf{v} \in V$.

If $\beta\colon V \times V \longrightarrow \mathbb{F}$ is a sesqui-linear form on V, then

$$\beta(\mathbf{u}, \mathbf{v}) = \underline{\mathbf{x}}^t \underline{\mathbf{A}}\, \underline{\overline{\mathbf{y}}}$$

and if β is bi-linear, then

$$\beta(\mathbf{u}, \mathbf{v}) = \underline{\mathbf{x}}^t \underline{\mathbf{A}}\, \underline{\mathbf{y}}.$$

We investigate the relationship between endomorphisms and changes of basis on the one hand, and sesqui-linear forms on the other.

Lemma 15.4. *Let $\beta\colon V \times V \longrightarrow \mathbb{F}$ be sesqui-linear and $T\colon V \longrightarrow V$ an endomorphism. Then*

$$\gamma\colon V \times V \longrightarrow \mathbb{F}, \quad (\mathbf{u}, \mathbf{v}) \longmapsto \beta(T(\mathbf{u}), T(\mathbf{v}))$$

is a sesqui-linear form.

We write $\beta \circ (T \times T)$ for γ in this theorem.

Proof. Take $\lambda, \mu \in \mathbb{F}$ and $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$. Then

$$\begin{aligned}
\gamma(\lambda\mathbf{u} + \mu\mathbf{v}, \mathbf{w}) &= \beta(T(\lambda\mathbf{u} + \mu\mathbf{v}), T(\mathbf{w})) \\
&= \beta(\lambda T(\mathbf{u}) + \mu T(\mathbf{v}), T(\mathbf{w})) \\
&= \lambda\beta(T(\mathbf{u}), T(\mathbf{w})) + \mu\beta(T(\mathbf{v}), T(\mathbf{w}))
\end{aligned}$$

$$= \lambda\gamma(\mathbf{u},\mathbf{w}) + \mu\gamma(\mathbf{v},\mathbf{w}).$$

On the other hand,

$$\begin{aligned}\gamma(\mathbf{u},\lambda{+}\mu\mathbf{w}) &= \beta(T(\mathbf{u}),T(\lambda{+}\mu\mathbf{w}))\\ &= \beta(\mathbf{u},\lambda T(\mathbf{v}) + \mu T(\mathbf{w}))\\ &= \overline{\lambda}\beta(T(\mathbf{u}),T(\mathbf{v})) + \overline{\mu}\beta(T(\mathbf{u}),T(\mathbf{w}))\\ &= \overline{\lambda}\gamma(\mathbf{u},\mathbf{v}) + \overline{\mu}\gamma(\mathbf{u},\mathbf{w}),\end{aligned}$$

which shows that γ is sesqui-linear. □

Corollary 15.5. *If $\beta\colon V\times V\longrightarrow\mathbb{F}$ is bi-linear and $T\colon V\longrightarrow V$ is linear, then $\beta\circ(T\times T)$ is bi-linear.*

Theorem 15.6. *Let $\beta,\gamma\colon \times V\longrightarrow\mathbb{F}$ be sesqui-linear forms and $T\colon V\longrightarrow V$ an endomorphism such that $\gamma=\beta\circ(T\times T)$.*

Choose a basis $\mathcal{B}=\{\mathbf{e}_1,\ldots,\mathbf{e}_n\}$ for V.

If the matrices of β,γ and T with respect to $\mathcal{B}$ are $\underline{\mathbf{B}},\underline{\mathbf{C}}$ and $\underline{\mathbf{A}}$ respectively, then

$$\underline{\mathbf{C}} = \underline{\mathbf{A}}^t\underline{\mathbf{B}}\,\overline{\underline{\mathbf{A}}}.$$

Proof. Let the co-ordinate vector of $\mathbf{u}$ with respect to the chosen basis be $\underline{\mathbf{x}}$ and that of $\mathbf{v}$ be $\underline{\mathbf{x}}$. Then the co-ordinate vectors of $T(\mathbf{u})$ and $T(\mathbf{v})$ are $\underline{\mathbf{A}}\,\underline{\mathbf{x}}$ and $\underline{\mathbf{A}}\,\underline{\mathbf{x}}$ respectively, and so

$$\begin{aligned}\underline{\mathbf{x}}^t\underline{\mathbf{C}}\,\overline{\underline{\mathbf{y}}} &= \gamma(\mathbf{u},\mathbf{v})\\ &= \beta\big(T(\mathbf{u}),T(\mathbf{v})\big)\\ &= \big(\underline{\mathbf{A}}\,\underline{\mathbf{x}}\big)^t\underline{\mathbf{B}}\,\overline{\underline{\mathbf{A}}\,\underline{\mathbf{y}}}\\ &= \underline{\mathbf{x}}^t\,\underline{\mathbf{A}}^t\underline{\mathbf{B}}\,\overline{\underline{\mathbf{A}}}\,\overline{\underline{\mathbf{y}}}\end{aligned}$$

By the uniqueness of the matrix representing γ, $\underline{\mathbf{C}} = \underline{\mathbf{A}}^t\underline{\mathbf{B}}\,\overline{\underline{\mathbf{A}}}$ □

Corollary 15.7. *Let $\beta,\gamma\colon V\times V\longrightarrow\mathbb{F}$ be bi-linear forms and $T:V\longrightarrow V$ an endomorphism such that $\gamma=\beta\circ(T\times T)$.*

Choose a basis $\mathcal{B}=\{\mathbf{e}_1,\ldots,\mathbf{e}_n\}$ for V. If the matrices of β,γ and T with respect to $\mathcal{B}$ are $\underline{\mathbf{B}},\underline{\mathbf{C}}$ and $\underline{\mathbf{A}}$ respectively, then

$$\underline{\mathbf{C}} = \underline{\mathbf{A}}^t\underline{\mathbf{B}}\,\underline{\mathbf{A}}.$$

Corollary 15.8. *Let $\beta\colon V\times V\longrightarrow\mathbb{F}$ be a sesqui-linear form.*

Choose bases $\mathcal{B}=\{\mathbf{e}_1,\ldots,\mathbf{e}_n\}$ and $\mathcal{C}=\{\mathbf{f}_1,\ldots,\mathbf{f}_n\}$ for V.

If the matrix of β with respect to $\mathcal{B}$ is $\underline{\mathbf{B}}$ and that with respect to $\mathcal{C}$ is $\underline{\mathbf{C}}$ then

$$\underline{\mathbf{C}} = \underline{\mathbf{A}}^t\underline{\mathbf{B}}\,\overline{\underline{\mathbf{A}}},$$

where $\underline{\mathbf{A}}$ is the "change of basis matrix" from the basis $\mathcal{C}$ to $\mathcal{B}$.

Proof. Recall that if $\mathbf{v}\in V$ has $\underline{\mathbf{x}}$ as co-ordinate vector with respect to $\mathcal{B}$ and $\underline{\mathbf{y}}$ with respect of $\mathcal{C}$, then $\underline{\mathbf{x}} = \underline{\mathbf{A}}\,\underline{\mathbf{y}}$. □

Corollary 15.9. *Let $\beta\colon V \times V \longrightarrow \mathbb{F}$ be a bi-linear form.*
Choose bases $\mathcal{B} = \{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ and $\mathcal{C} = \{\mathbf{f}_1, \ldots, \mathbf{f}_n\}$ for V.
If the matrix of β with respect to $\mathcal{B}$ is $\underline{\mathbf{B}}$ and that with respect to $\mathcal{C}$ is $\underline{\mathbf{C}}$ then

$$\underline{\mathbf{C}} = \underline{\mathbf{A}}^t \underline{\mathbf{B}}\, \underline{\mathbf{A}},$$

where $\underline{\mathbf{A}}$is the "change of basis matrix" from the basis $\mathcal{C}$ to $\mathcal{B}$.

In order to represent inner products on finitely generated spaces by matrices, we have only exploited the fact that an inner product on V is a sesqui-linear form on V. The other requirements impose conditions on the matrix representing our form.

Because $\langle\langle \mathbf{u}, \mathbf{v}\rangle\rangle = \overline{\langle\langle \mathbf{v}, \mathbf{u}\rangle\rangle}$ for all $\mathbf{u}, \mathbf{v} \in V$, we must have

$$a_{ji} := \langle\langle \mathbf{e}_j, \mathbf{e}_i\rangle\rangle = \overline{\langle\langle \mathbf{e}_i, \mathbf{e}_j\rangle\rangle} =: a_{ij}$$

for any basis vectors $\mathbf{e}_i$ and $\mathbf{e}_j$. Thus, if $\underline{\mathbf{A}}$ is the matrix of the inner product, we must have

$$\underline{\mathbf{A}}^t = \overline{\underline{\mathbf{A}}}$$

Definition 15.10. The complex matrix $\underline{\mathbf{A}}$ is *Hermitian* if and only if

$$\underline{\mathbf{A}}^t = \overline{\underline{\mathbf{A}}}.$$

Of course, when $\mathbb{F} \subseteq \mathbb{R}$, sesqui-linearity becomes bi-linearity and $\langle\langle \mathbf{v}, \mathbf{u}\rangle\rangle = \langle\langle \mathbf{v}, \mathbf{u}\rangle\rangle$ for all $\mathbf{u}, \mathbf{v} \in V$, so that if $\underline{\mathbf{A}}$ is the matrix of the inner product, then

$$\underline{\mathbf{A}}^t = \underline{\mathbf{A}}$$

Definition 15.11. The real matrix $\underline{\mathbf{A}}$ is *symmetric* if and only if

$$\underline{\mathbf{A}}^t = \underline{\mathbf{A}}.$$

We summarise our discussion in the next theorem.

Theorem 15.12. *Any matrix representing a complex inner product must be Hermitian, and any matrix representing a real inner product must be symmetric.*

So far, we have not exploited the positive definiteness of inner products. This also has consequences for the matrix representation of inner products. It is again more convenient to present the discussion at the level of sesqui-linear forms, rather than restricting only to inner products.

First observe that for any inner product space V, given $\lambda \in \mathbb{F}$ and $\mathbf{v} \in V$, it follows from (IP1) and (IP2) that

$$\langle\langle \lambda\mathbf{v}, \lambda\mathbf{v}\rangle\rangle = \lambda\overline{\lambda}\langle\langle \mathbf{v}, \mathbf{v}\rangle\rangle = |\lambda|^2\langle\langle \mathbf{v}, \mathbf{v}\rangle\rangle$$

In particular, given a basis $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$, we have $\mathbf{v} = \sum x_j\mathbf{e}_j$, for suitable x_j $(j = 1, \ldots, n)$, whence $\langle\langle \mathbf{v}, \mathbf{v}\rangle\rangle = \sum a_{ij}x_i\overline{x_j}$, with $a_{ij} := \langle\langle \mathbf{e}_i, \mathbf{e}_j\rangle\rangle$.

Thus we require that for all $x_1, \ldots, x_n$

$$\sum_{j=1}^{n} a_{ij}x_i\overline{x_j} \geq 0,$$

with equality if and only if each $x_j = 0$.

This last expression is a homogeneous quadratic polynomial in the co-ordinates of $\mathbf{x}$.

Given the importance of such functions, especially in the real case, we later devote a chapter to them.

We can now characterise the matrices which represent inner products.

Theorem 15.13. *The matrix $\underline{\mathbf{A}}$ represents an inner product on a finitely generated vector space over the field $\mathbb{F} \subseteq \mathbb{C}$ if and only if it is a positive definite Hermitian matrix.*

It represents an inner product on a finitely generated vector space over the field $\mathbb{F} \subseteq \mathbb{R}$ if and only if it is a positive definite symmetric matrix.

15.1 Exercises

Exercise 15.1. Let V and W be finitely generated vector spaces over the subfield $\mathbb{F}$ of $\mathbb{C}$. Let $\gamma : W \times W \longrightarrow \mathbb{F}$ be a bi-linear form on W and $T : V \longrightarrow W$ a linear transformation.

Show that

$$\beta : V \times V \longrightarrow \mathbb{F}, \quad (\mathbf{u}, \mathbf{v}) \longmapsto \gamma(T(\mathbf{u}), T(\mathbf{v}))$$

defines a bi-linear form on V. [We write $\beta = \gamma \circ (T \times T)$.]

Show that if, instead, γ is sesqui-linear, then so is β.

Choose bases $\mathcal{B} = \{\mathbf{e}_i, \ldots, \mathbf{e}_n\}$ for V and $\mathcal{C} = \{\mathbf{f}_1, \ldots, \mathbf{f}_m\}$ for W. Let the matrix of T with respect to these bases be $\underline{\mathbf{A}}$. Let the matrix of γ with respect to $\mathcal{C}$ be $\underline{\mathbf{C}}$ and that of β with respect to $\mathcal{B}$ be $\underline{\mathbf{B}}$.

Show that if γ is bi-linear, then

$$\underline{\mathbf{B}} = \underline{\mathbf{A}}^t \underline{\mathbf{C}}\, \underline{\mathbf{A}},$$

and if γ is sesqui-linear, then

$$\underline{\mathbf{B}} = \underline{\mathbf{A}}^t \underline{\mathbf{C}}\, \overline{\underline{\mathbf{A}}}.$$

Exercise 15.2. Show that if $\underline{\mathbf{A}}$ is a complex Hermitian $n \times n$ matrix, and $\underline{\mathbf{B}}$ is any other complex $n \times n$ matrix, then $\underline{\mathbf{B}}^t \underline{\mathbf{A}}\, \overline{\underline{\mathbf{B}}}$ is also Hermitian.

Show that if $\underline{\mathbf{A}}$ is a real symmetric $n \times n$ matrix, and $\underline{\mathbf{B}}$ is any other real $n \times n$ matrix, then $\underline{\mathbf{B}}^t \underline{\mathbf{A}}\, \underline{\mathbf{B}}$ is also symmetric.

The presence of an inner product has significant consequences. In particular, it enables us to map a vector space into its dual space, and gives rise to the notion of the *adjoint* of a linear transformation, which is a cornerstone of several applications of linear algebra, such as to quantum mechanics.

To discuss the adjoint we first investigate the relation between a vector space and its dual in the presence of an inner product.

Recall that if V is a vector space over the field $\mathbb{F}$, then its dual space, V^*, is $\mathrm{Hom}_{\mathbb{F}}(V, \mathbb{F})$, the $\mathbb{F}$ vector space of all $\mathbb{F}$-linear transformations $V \to \mathbb{F}$. Such a linear transformation is often called a *1-form* or a *linear form*.

Lemma 15.14. *Let $(V, \langle\!\langle\ ,\ \rangle\!\rangle)$ be an inner product space over $\mathbb{F}$. For each $\mathbf{v} \in V$*

$$\langle\!\langle\ , \mathbf{v} \rangle\!\rangle : V \longrightarrow \mathbb{F}, \quad \mathbf{x} \longmapsto \langle\!\langle \mathbf{x}, \mathbf{v} \rangle\!\rangle$$

is a linear transformation.

Proof. Take $\mathbf{x}, \mathbf{y} \in V$ and $\lambda, \mu \in \mathbb{F}$.

By the definition of inner product, $\langle\langle \lambda\mathbf{x} + \mu\mathbf{y}, \mathbf{v}\rangle\rangle = \lambda\langle\langle \mathbf{x}, \mathbf{v}\rangle\rangle + \mu\langle\langle \mathbf{y}, \mathbf{v}\rangle\rangle$ □

We use Lemma 15.14 on the preceding page to embed V in V^*.

Lemma 15.15. *Let $(V, \langle\langle\ ,\ \rangle\rangle)$ be an inner product space over $\mathbb{F}$. Then*

$$R\colon V \longrightarrow V^*, \quad \mathbf{v} \longmapsto \langle\langle\ , \mathbf{v}\rangle\rangle$$

is injective.

Proof. Take $\mathbf{u}, \mathbf{v} \in V$.

$R(\mathbf{u}) = R(\mathbf{v})$ if and only if for all $\mathbf{x} \in V$, $\big(R(\mathbf{u})\big)(\mathbf{x}) = \big(R(\mathbf{v})(\mathbf{x})$, or, equivalently, $\langle\langle \mathbf{x}, \mathbf{u}\rangle\rangle = \langle\langle \mathbf{x}, \mathbf{v}\rangle\rangle$.

By the definition of inner product, this is equivalent to $\langle\langle \mathbf{x}, \mathbf{u} - \mathbf{v}\rangle\rangle = 0$ for all $\mathbf{x} \in V$.

In particular $\langle\langle \mathbf{u} - \mathbf{v}, \mathbf{u} - \mathbf{v}\rangle\rangle = 0$, whence $\mathbf{u} = \mathbf{v}$ since $\langle\langle\ ,\ \rangle\rangle$ is an inner product. □

15.2 Riesz Representation Theorem

For general vector spaces, there are no "natural" linear forms, that is linear transformations from the given space to the field of scalars.

The preceding discussion shows that for inner product spaces, there is a rich supply of them, at least one for each vector in V. The Riesz Representation Theorem states that under some additional conditions, these are the only linear forms. We prove the Riesz Representation Theorem for finitely generated inner product spaces.

Theorem 15.16 (Riesz Representation Theorem). *Let $(V, \langle\langle\ ,\ \rangle\rangle)$ be a finitely generated inner product space over $\mathbb{F}$.*

For each linear transformation $\varphi\colon V \longrightarrow \mathbb{F}$, there is a unique $\mathbf{v}_\varphi \in V$ such that for all $\mathbf{x} \in V$

$$\varphi(\mathbf{x}) = \langle\langle \mathbf{x}, \mathbf{v}_\varphi\rangle\rangle$$

Proof. **Uniqueness:**

Suppose that $\langle\langle \mathbf{x}, \mathbf{u}\rangle\rangle = \langle\langle \mathbf{x}, \mathbf{v}\rangle\rangle$ for all $\mathbf{x} \in V$.

Since this is equivalent to $\langle\langle \mathbf{x}, \mathbf{v} - \mathbf{u}\rangle\rangle = 0$ for all $\mathbf{x} \in V$.

In particular, $\langle\langle \mathbf{v} - \mathbf{u}, \mathbf{v} - \mathbf{u}\rangle\rangle = 0$, whence $\mathbf{v} = \mathbf{u}$.

Existence:

Since $\operatorname{im}(\varphi)$ is a vector subspace of $\mathbb{F}$, either $\operatorname{im}(\varphi) = \{0\}$ or $\operatorname{im}(\varphi) = \mathbb{F}$.

In the former case, $\varphi(\mathbf{x}) = 0$ for all $\mathbf{x} \in V$, and so we may choose $\mathbf{v}_\varphi := \mathbf{0}_V$.

In the latter case, the rank of φ being 1, its nullity is $\dim V - 1$.

Choose an orthonormal basis for $\ker(\varphi)$, say $\{\mathbf{e}_2, \ldots, \mathbf{e}_n\}$, and extend to an orthonormal basis, $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ of V.

Observe that since $\mathbf{e}_1 \notin \ker(\varphi)$, $\varphi(\mathbf{e}_1) \neq 0$.

Take $\mathbf{x} \in V$. By Theorem 14.7 on page 182,

$$\mathbf{x} = \sum_{j=1}^{n} \langle\langle \mathbf{x}, \mathbf{e}_j\rangle\rangle \mathbf{e}_j,$$

so that

$$\begin{aligned}
\varphi(\mathbf{x}) &= \varphi\Big(\sum_{j=1}^{n}\langle\!\langle \mathbf{x}, \mathbf{e}_j\rangle\!\rangle \mathbf{e}_j\Big) \\
&= \sum_{j=1}^{n}\langle\!\langle \mathbf{x}, \mathbf{e}_j\rangle\!\rangle \varphi(\mathbf{e}_j) && \text{as } \varphi \text{ is linear} \\
&= \langle\!\langle \mathbf{x}, \mathbf{e}_1\rangle\!\rangle \varphi(\mathbf{e}_1) && \text{as } \varphi(\mathbf{e}_j) = 0 \text{ for } j > 1 \\
&= \langle\!\langle \mathbf{x}, \overline{\varphi(\mathbf{e}_1)}\mathbf{e}_1\rangle\!\rangle && \text{as } \varphi(\mathbf{e}_1) \in \mathbb{F}.
\end{aligned}$$

So $\mathbf{v}_\varphi := \overline{\varphi(\mathbf{e}_1)}\mathbf{e}_1$ clearly has the required property. □

Corollary 15.17. *Let $(V, \langle\!\langle\ ,\ \rangle\!\rangle)$ be a finite dimensional inner product space over $\mathbb{F}$. Then the function*

$$R\colon V \longrightarrow \mathrm{Hom}(V, \mathbb{F}), \qquad \mathbf{v} \longmapsto \langle\!\langle\ , \mathbf{v}\rangle\!\rangle$$

is an additive bijection, which is an isomorphism of vector spaces whenever $\mathbb{F} \subseteq \mathbb{R}$.

Proof. That R is well defined follows from the fact that given $\mathbf{x}, \mathbf{y}, \mathbf{v} \in V,\ \alpha, \beta \in \mathbb{F}$,

$$\langle\!\langle \alpha\mathbf{x} + \beta\mathbf{y}, \mathbf{v}\rangle\!\rangle = \alpha\langle\!\langle \mathbf{x}, \mathbf{y}\rangle\!\rangle + \beta\langle\!\langle \mathbf{y}, \mathbf{v}\rangle\!\rangle$$

That R is bijective is a restatement of the Riesz Representation Theorem.

Finally, take $\alpha, \beta \in \mathbb{F}$ and $\mathbf{u}, \mathbf{v} \in V$. Then, for any $\mathbf{x} \in V$,

$$\begin{aligned}
\big(R(\alpha\mathbf{u} + \beta\mathbf{v})\big)(\mathbf{x}) &= \langle\!\langle \mathbf{x}, \alpha\mathbf{u} + \beta\mathbf{v}\rangle\!\rangle \\
&= \overline{\alpha}\langle\!\langle \mathbf{x}, \mathbf{u}\rangle\!\rangle + \overline{\beta}\langle\!\langle \mathbf{x}, \mathbf{v}\rangle\!\rangle. \\
&= \overline{\alpha}\big(R(\mathbf{u})\big)(\mathbf{x}) + \overline{\beta}\big(R(\mathbf{v})\big)(\mathbf{x}) \\
&= \big(\overline{\alpha}R(\mathbf{u}) + \overline{\beta}R(\mathbf{v})\big)(\mathbf{x})
\end{aligned}$$

Thus $R(\alpha\mathbf{u} + \beta\mathbf{v}) = \overline{\alpha}R(\mathbf{u}) + \overline{\beta}R(\mathbf{v})$.

If, in fact, $\mathbb{F} \subseteq \mathbb{R}$, then $R(\alpha\mathbf{u} + \beta\mathbf{v}) = \alpha R(\mathbf{u}) + \beta R(\mathbf{v})$ □

Example 15.18. Our version of the Riesz Representation Theorem is not the original one.

While the restriction to finitely generated inner product spaces is not necessary, some restriction (either on the inner product space, V, or on the class of linear forms, $\varphi\colon V \longrightarrow \mathbb{F}$) is required, as we now show, recalling Example 14.12 on page 186.

Take $V := \mathbb{R}[t]$ with the inner product

$$\langle\!\langle\ ,\ \rangle\!\rangle\colon V \times V \longrightarrow \mathbb{F}, \quad (f, g) \longrightarrow \int_0^1 f(x)g(x)\,dx$$

Take the linear transformation

$$\varphi\colon V \longrightarrow \mathbb{R}, \quad f \longmapsto f(0).$$

For any $h \in V$, $f := t^2h \in \ker(\varphi)$ and

$$\langle\!\langle f, h\rangle\!\rangle = \langle\!\langle t^2h, h\rangle\!\rangle = \int_0^1 x^2(h(x))^2dx = \langle\!\langle th, th\rangle\!\rangle.$$

Thus $\varphi(f) = \langle\langle f, h\rangle\rangle$ if and only if $\varphi(f) = \|th\|^2$.

Since $\varphi(f) = 0$, this is the case if and only if $th = 0$.

As t is not the zero polynomial, this implies that h is the zero polynomial. Then $\langle\langle p, h\rangle\rangle = 0$ for all $p \in V$.

But φ is not the zero transformation, since $\varphi(1) = 1$.

Hence there is no $\mathbf{v}_\varphi \in V$ with $\varphi(\mathbf{x}) = \langle\langle \mathbf{x}, \mathbf{v}_\varphi\rangle\rangle$ for all $\mathbf{x} \in V$.

15.3 The Adjoint of a Linear Transformation

We consider the effect of a linear transformation, $T\colon V \longrightarrow W$ between inner product spaces. While our primary interest is in the case where $W = V$, the greater generality does not make the analysis more difficult and is needed for several important applications.

Let $T\colon V \longrightarrow W$ be a linear transformation from the finitely generated inner product space $(V, \langle\langle\ ,\ \rangle\rangle_V)$ to the finitely generated inner product space $(W, \langle\langle\ ,\ \rangle\rangle_W)$.

Take $\mathbf{w} \in W$. Then

$$L_T^{\mathbf{w}}\colon V \longrightarrow \mathbb{F}, \quad \mathbf{x} \longmapsto \langle\langle T(\mathbf{x}), \mathbf{w}\rangle\rangle_W$$

is a linear transformation.

By the Riesz Representation Theorem, there is a unique $\mathbf{v}_{L_T^{\mathbf{w}}} \in V$ with $L_T^{\mathbf{w}}(\mathbf{x}) = \langle\langle \mathbf{x}, \mathbf{v}_{L_T^{\mathbf{v}}}\rangle\rangle_V$ for all $\mathbf{x} \in V$. In other words, for each $\mathbf{x} \in V$,

$$\langle\langle T(\mathbf{x}), \mathbf{w}\rangle\rangle_W = \langle\langle \mathbf{x}, \mathbf{w}_{L_T^{\mathbf{v}}}\rangle\rangle_V$$

Given the linear transformation $T\colon V \longrightarrow W$ we obtain a function

$$T^*\colon W \longrightarrow V, \quad \mathbf{v} \longmapsto \mathbf{v}_{L_T^{\mathbf{w}}},$$

characterised by

$$\langle\langle T(\mathbf{x}), \mathbf{y}\rangle\rangle_W = \langle\langle \mathbf{x}, T^*(\mathbf{y})\rangle\rangle_V \qquad \text{for all } \mathbf{x} \in V, \mathbf{y} \in W$$

Lemma 15.19. *For each linear trasformation $T\colon V \longrightarrow W$,*

$$T^*\colon W \longrightarrow V, \quad \mathbf{w} \longmapsto \mathbf{v}_{L_T^{\mathbf{w}}}$$

is a linear transformation

Proof. Take $\mathbf{u}, \mathbf{v} \in W$ and $\alpha, \beta \in \mathbb{F}$. Then, for each $\mathbf{x} \in V$,

$$\begin{aligned}
\langle\langle \mathbf{x}, T^*(\alpha\mathbf{u} + \beta\mathbf{v})\rangle\rangle_V &= \langle\langle T(\mathbf{x}), \alpha\mathbf{u} + \beta\mathbf{v}\rangle\rangle_W \\
&= \overline{\alpha}\langle\langle T(\mathbf{x}), \mathbf{u}\rangle\rangle_W + \overline{\beta}\langle\langle T(\mathbf{x}), \mathbf{v}\rangle\rangle_W \\
&= \overline{\alpha}\langle\langle \mathbf{x}, T^*(\mathbf{u})\rangle\rangle_V + \overline{\beta}\langle\langle \mathbf{x}, T^*(\mathbf{v})\rangle\rangle_V \\
&= \langle\langle \mathbf{x}, \alpha T^*(\mathbf{u}) + \beta T^*(\mathbf{v})\rangle\rangle_V
\end{aligned}$$

By the uniqueness of $\mathbf{v}_{L_T^{\alpha\mathbf{u}+\beta\mathbf{v}}}$, $T^*(\alpha\mathbf{u} + \beta\mathbf{v}) = \alpha T^*(\mathbf{u}) + \beta T^*(\mathbf{v})$. □

Definition 15.20. The linear transformation $T^*\colon W \longrightarrow V$ is the *adjoint* of $T\colon V \longrightarrow W$.

Lemma 15.21. *Let $(U, \langle\langle\ ,\ \rangle\rangle_U)$, $(V, \langle\langle\ ,\ \rangle\rangle_V)$ and $(W, \langle\langle\ ,\ \rangle\rangle_W)$ be inner product spaces and $S\colon U \longrightarrow V$, $T\colon V \longrightarrow W$ linear transformations. Then*

(a) $id_V^* = id_V : V \longrightarrow V$

(b) $(S \circ T)^* = T^* \circ S^* : W \longrightarrow U$

(c) $(T^*)^* = T : V \longrightarrow W$

(d) $\ker(T^* \circ T) = \ker T$

(e) $(\operatorname{im} T^*)^\perp = \ker T$

Proof. (a) Take $\mathbf{v} \in V$. For every $\mathbf{x} \in V$

$$\begin{aligned}\langle\!\langle \mathbf{x}, id_V^*(\mathbf{v})\rangle\!\rangle &= \langle\!\langle id_V(\mathbf{x}), \mathbf{v}\rangle\!\rangle \\ &= \langle\!\langle \mathbf{x}, \mathbf{v}\rangle\!\rangle\end{aligned}$$

By uniqueness, $id_V^*(\mathbf{v}) = \mathbf{v}$ for every $\mathbf{v} \in V$.

(b) Take $\mathbf{u} \in U$ and $\mathbf{w} \in W$. Then

$$\begin{aligned}\langle\!\langle \mathbf{u}, (S \circ T)^*(\mathbf{w})\rangle\!\rangle_U &= \langle\!\langle (S \circ T)(\mathbf{u}), \mathbf{w}\rangle\!\rangle_W \\ &= \langle\!\langle S(T(\mathbf{u})), \mathbf{w}\rangle\!\rangle_W \\ &= \langle\!\langle T(\mathbf{u}), S^*(\mathbf{w})\rangle\!\rangle_V \\ &= \langle\!\langle \mathbf{u}, T^*(S^*(\mathbf{w}))\rangle\!\rangle_U \\ &= \langle\!\langle \mathbf{u}, (T^* \circ S^*)(\mathbf{w})\rangle\!\rangle_U\end{aligned}$$

By uniqueness, $(S \circ T)^*(\mathbf{w}) = (T^* \circ S^*)(\mathbf{w})$ for every $\mathbf{w} \in W$.

(c) Take $\mathbf{v} \in V$. For every $\mathbf{w} \in W$

$$\begin{aligned}\langle\!\langle \mathbf{w}, T(\mathbf{v})\rangle\!\rangle_W &= \overline{\langle\!\langle T(\mathbf{v}), \mathbf{w}\rangle\!\rangle_W} \\ &= \overline{\langle\!\langle \mathbf{v}, T^*(\mathbf{w})\rangle\!\rangle_V} \\ &= \langle\!\langle T^*(\mathbf{w}), \mathbf{v}\rangle\!\rangle_V \\ &= \langle\!\langle \mathbf{w}, (T^*)^*(\mathbf{v})\rangle\!\rangle_W\end{aligned}$$

By uniqueness, $(T^*)^*(\mathbf{v}) = T(\mathbf{v})$ for every $\mathbf{v} \in V$.

(d) By the properties of composition of linear transformations,

$$\ker T \subseteq \ker(T^* \circ T)$$

To establish the opposite inclusion, suppose that $\mathbf{v} \in \ker(T^* \circ T)$. Then

$$\begin{aligned}\langle\!\langle T(\mathbf{v}), T(\mathbf{v})\rangle\!\rangle_W &= \langle\!\langle \mathbf{v}, T^*\big(T(\mathbf{v})\big)\rangle\!\rangle_V \\ &= \langle\!\langle \mathbf{v}, \mathbf{0}_V\rangle\!\rangle_V \\ &= 0\end{aligned}$$

Since $\langle\!\langle\ ,\ \rangle\!\rangle_W$ is an inner product on W, $T(\mathbf{v}) = \mathbf{0}_W$, whence

$$\ker(T^* \circ T) \subseteq \ker T$$

(e)

$$\begin{aligned}\mathbf{v} \in (\operatorname{im} T^*)^\perp \quad &\text{if and only if} \quad \langle\!\langle T^*(\mathbf{w}), \mathbf{v}\rangle\!\rangle_V = 0 \quad \text{for all } \mathbf{w} \in W \\ &\text{if and only if} \quad \langle\!\langle \mathbf{w}, T(\mathbf{v})\rangle\!\rangle_W = 0 \quad \text{for all } \mathbf{w} \in W \\ &\text{if and only if} \quad T(\mathbf{v}) = \mathbf{0}_W \quad \text{as } \langle\!\langle\ ,\ \rangle\!\rangle_W \text{ is an inner product} \\ &\text{if and only if} \quad \mathbf{v} \in \ker T\end{aligned}$$

□

We now restrict attention to a single inner product space.

There is a useful and important relationship between subspaces invariant under an endomorphism and those invariant under the adjoint of the endomorphism.

Theorem 15.22. *Let $T\colon V \longrightarrow V$ be an endomorphism of the inner product space $(V, \langle\langle\ ,\ \rangle\rangle)$. If the subspace W of V is invariant under T, then $W^\perp$ is invariant under T^*. In other words, if $T(\mathbf{w}) \in W$ for all $\mathbf{w} \in W$, then $T^*(\mathbf{x}) \in W^\perp$ for all $\mathbf{x} \in W^\perp$.*

Proof. Take $\mathbf{w} \in W$ and $\mathbf{x} \in W^\perp$. Then

$$\begin{aligned} \langle\langle \mathbf{w}, T^*(\mathbf{x}) \rangle\rangle &= \langle\langle T(\mathbf{w}), \mathbf{x} \rangle\rangle \\ &= 0 \qquad\qquad \text{since } T(\mathbf{w}) \in W \text{ and } \mathbf{x} \in W^\perp \end{aligned}$$

Thus $T^*(\mathbf{x}) \in W^\perp$. □

15.4 Self-Adjoint Linear Transformations

An important class of endomorphisms are those which agree with their adjoints.

Definition 15.23. The endomorphism $T\colon V \longrightarrow V$ is *self-adjoint* if and only if $T^* = T$.

Self-adjoint endomorphisms are plentiful and arise naturally, as the next lemma shows.

Lemma 15.24. *Let $T\colon V \longrightarrow W$ be a linear transformation between inner product spaces. Then both $T^* \circ T\colon V \longrightarrow V$ and $T \circ T^*\colon W \longrightarrow W$ are self-adjoint endomorphisms.*

Proof.

$$\begin{aligned} (T^* \circ T)^* &= T^* \circ (T^*)^* && \text{by Lemma 15.21 (b)} \\ &= T^* \circ T && \text{by Lemma 15.21 (c)} \end{aligned}$$

The same argument applies to $T \circ T^*$. □

Being self-adjoint has significant consequences for the eigenvalues of an endomorphism.

Theorem 15.25. *The eigenvalues of a self-adjoint endomorphism are all real.*

Proof. Let $\mathbf{v} \neq \mathbf{0}_V$ be an eigenvector of the self-adjoint endomorphism T for the eigenvalue λ. Then

$$\begin{aligned} \lambda\langle\langle \mathbf{v}, \mathbf{v} \rangle\rangle &= \langle\langle \lambda\mathbf{v}, \mathbf{v} \rangle\rangle \\ &= \langle\langle T(\mathbf{v}), \mathbf{v} \rangle\rangle && \text{as } \mathbf{v} \text{ is an eigenvector for } \lambda \\ &= \langle\langle \mathbf{v}, T^*(\mathbf{v}) \rangle\rangle \\ &= \langle\langle \mathbf{v}, T(\mathbf{v}) \rangle\rangle && \text{a } T \text{ is self-adjoint} \\ &= \langle\langle \mathbf{v}, \lambda\mathbf{v} \rangle\rangle && \text{as } \mathbf{v} \text{ is an eigenvector for } \lambda \\ &= \overline{\lambda}\langle\langle \mathbf{v}, \mathbf{v} \rangle\rangle \end{aligned}$$

Since $\mathbf{v} \neq \mathbf{0}_V$, it follows that $\lambda = \overline{\lambda}$, which is the case if and only if λ is real. □

Corollary 15.26. *Every self-adjoint endomorphism has n eigenvalues (with multiplicities), whenever the ground field contains all real numbers.*

Proof. By the Fundamental Theorem of Algebra, the characteristic polynomial factors into linear factors over the complex numbers, so that

$$\chi_T(t) = \prod_{j=i}^{n}(t - \lambda_j)$$

The λ_j's are precisely the eigenvalues of T. Since these are all real, this is, in fact, a factorisation over the reals.

Hence T has n real eigenvalues (with multiplicities). □

Corollary 15.27. *Eigenvectors for distinct eigenvalues of a self-adjoint endomorphism are mutually orthogonal.*

Proof. Let $T\colon V \longrightarrow V$ be a self-adjoint endomorphism.

Let $\lambda \neq \mu$ be eigenvalues of T.

Let $\mathbf{u}$ be an eigenvector for $\lambda\mathbf{u}$ and $\mathbf{v}$ an eigenvector for μ. Then

$$\begin{aligned}
\lambda\langle\langle \mathbf{u}, \mathbf{v}\rangle\rangle &= \langle\langle \lambda\mathbf{u}, \mathbf{v}\rangle\rangle \\
&= \langle\langle T(\mathbf{u}), \mathbf{v}\rangle\rangle \\
&= \langle\langle \mathbf{u}, T(\mathbf{v})\rangle\rangle && \text{as } T \text{ is self-adjoint} \\
&= \langle\langle \mathbf{u}, \mu\mathbf{v}\rangle\rangle \\
&= \mu\langle\langle \mathbf{u}, \mathbf{v}\rangle\rangle && \text{as } \mu \in \mathbb{R}.
\end{aligned}$$

Thus, $(\lambda - \mu)\langle\langle \mathbf{u}, \mathbf{v}\rangle\rangle = 0$.

Since $\lambda \neq \mu$, $\langle\langle \mathbf{u}, \mathbf{v}\rangle\rangle = 0$. □

We come to our main result on self-adjoint endomorphisms.

Theorem 15.28. *If $T\colon V \longrightarrow V$ is a self-adjoint endomorphism of the finitely generated inner product space $(V, \langle\langle\ ,\ \rangle\rangle)$, then V has an orthonormal basis of eigenvectors of T.*

Proof. Let $T\colon V \longrightarrow V$ be self-adoint.

We use induction on $\dim(V)$.

If $\dim(V) = 1$, let $\mathbf{v}$ be any non-zero vector in V, and put

$$\mathbf{e} := \frac{\mathbf{v}}{\|\mathbf{v}\|}$$

Plainly, $\|\mathbf{e}\| = 1$, so that $\{\mathbf{e}\}$ is an orthonormal basis for V.

As $\dim(V) = 1$, $T(\mathbf{e}) = \lambda\mathbf{e}$, showing that $\mathbf{e}$ s an eigenvector of T.

Now suppose that the result holds for self-adjoint endomorphisms of inner product spaces of dimension less than n.

Suppose that $\dim(V) = n$..

By Corollary 15.26, T has an eigenvalue, say λ.

Let $\mathbf{v} \neq \mathbf{0}_V$ be an eigenvector for the eigenvalue λ.

Then $W = \mathbb{F}\mathbf{v} = \{\alpha\mathbf{v} \mid \alpha \in \mathbb{F}\}$ is a T-invariant subspace of V.

Since V is finitely generated, $V \cong W \oplus W^{\perp} = \mathbb{F}\mathbf{v} \oplus (\mathbb{F}\mathbf{v})^{\perp}$, whence $\dim W = n - 1$

By Theorem 15.22, $W^\perp$ is a T^*-invariant subspace of V.

Since T is self-adjoint, $W^\perp$ is an $(n-1)$–dimensional T-invariant subspace of V.

By the inductive hypothesis, $W^\perp$ has an orthonormal basis, $\{\mathbf{e}_2, \ldots, \mathbf{e}_n\}$, comprising eigenvectors of T.

Putting $\mathbf{e}_1 := \frac{\mathbf{v}}{\|\mathbf{v}\|}$, $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ is an orthonormal basis for V comprising eigenvectors of T,. □

Corollary 15.29. *If $\underline{\mathbf{A}}$ is a Hermitian $n \times n$ complex matrix, then there is a unitary matrix, $\underline{\mathbf{B}}$, such that $\overline{\underline{\mathbf{B}}}^t \underline{\mathbf{A}}\,\underline{\mathbf{B}}$ is a (real) diagonal matrix.*

Proof. Recall that if take $\mathbb{C}_{(n)}$ with the standard (Euclidean) inner product and regard the $n \times n$ complex matrix $\underline{\mathbf{A}}$ as the linear transformation

$$\mathbb{C}_{(n)} \longrightarrow \mathbb{C}_{(n)}, \quad \mathbf{x} \longmapsto \underline{\mathbf{A}}\mathbf{x},$$

then $\underline{\mathbf{A}}$ is self-adjoint if and only if it is Hermitian.

In that case $\mathbb{C}_{(n)}$ has an orthonormal $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ basis comprising eigenvectors of $\underline{\mathbf{A}}$.

Let $\underline{\mathbf{B}}$ be the $n \times n$ complex matrix whose j^{th} column is $\mathbf{e}_j$.

Since $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ is orthonormal, $\underline{\mathbf{B}}^t\overline{\underline{\mathbf{B}}} = \underline{\mathbf{1}}_n$.

Thus $\underline{\mathbf{B}}^{-1} = \overline{\underline{\mathbf{B}}}^t$, that is to say, $\underline{\mathbf{B}}$ is unitary.

Moreover, since for each j there is a $\lambda_j \in \mathbb{R}$ with $\underline{\mathbf{A}}\mathbf{e}_j = \lambda_j \mathbf{e}_j$, we have

$$\underline{\mathbf{A}}\,\underline{\mathbf{B}} = \underline{\mathbf{B}}\,\underline{\mathbf{diag}}(\lambda_1, \ldots, \lambda_n),$$

where $\underline{\mathbf{diag}}(\lambda_1, \ldots, \lambda_n)$ is the $n \times n$ matrix $\Big[x_{ij}\Big]_{n\times n}$ defined by

$$x_{ij} = \begin{cases} \lambda_j & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$

Thus $\overline{\underline{\mathbf{B}}}^t \underline{\mathbf{A}}\,\underline{\mathbf{B}} = \underline{\mathbf{B}}^{-1}\underline{\mathbf{A}}\,\underline{\mathbf{B}} = \underline{\mathbf{diag}}(\lambda_1, \ldots, \lambda_n)$ is a (real) diagonal matrix. □

Corollary 15.30. *If $\underline{\mathbf{A}}$ is a symmetric $n \times n$ real matrix, then there is an orthogonal matrix, $\underline{\mathbf{B}}$, such that $\underline{\mathbf{B}}^t \underline{\mathbf{A}}\,\underline{\mathbf{B}}$ is a diagonal matrix.*

Proof. The statement follows from Corollary 15.29 by recalling that a real matrix is Hermitian if and only if it is symmetric and that a real matrix is unitary if and only if it is orthogonal. □

15.5 Exercises

Exercise 15.3. Take $V := \mathcal{P}_3$, the set of all polynomials of degree at most 2 in the indeterminate t with real coefficients, with inner product $\langle\!\langle\ ,\ \rangle\!\rangle$ given by

$$\langle\!\langle p, q \rangle\!\rangle := \int_0^1 p(x)q(x)dx$$

Consider the linear form

$$\varphi : V \longrightarrow \mathbb{R}, \quad p \longmapsto p(0)$$

Find the element of $\mathcal{P}_3$ which represents φ.

Exercise 15.4. Find an orthogonal matrix which diagonalises the real matrix

$$\begin{bmatrix} 1 & 2 \\ 2 & 5 \end{bmatrix}$$

Exercise 15.5. Find a unitary matrix which diagonalises the complex matrix

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

Mathematics is the door and key to the sciences.

Roger Bacon

Chapter 16

Real Quadratic Forms

Recall from multivariate calculus that to find the extreme values of a sufficiently smooth function

$$f\colon \mathbb{R}^n \longrightarrow \mathbb{R}, \qquad (x_1, \ldots, x_n) \longmapsto f(x_1, \ldots, x_n)$$

we first look at its *gradient*

$$\nabla f\,(x_1, \ldots, x_n) := (f_{x_1}(x_1, \ldots, x_n), \ldots, f_{x_n}(x_1, \ldots, x_n)),$$

where

$$f_{x_i} := \frac{\partial f}{\partial x_i}.$$

Because of the conditions we have imposed on f, a necessary — but not sufficient —condition for f to have an extreme value at a point in $\mathbb{R}^n$ is that the gradient be the zero vector at that point.

We then examine the *Hessian* of f,

$$\begin{bmatrix} \dfrac{\partial^2 f}{\partial x_1 \partial x_1} & \cdots & \dfrac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & & \vdots \\ \dfrac{\partial^2 f}{\partial x_n \partial x_1} & \cdots & \dfrac{\partial^2 f}{\partial x_n \partial x_n} \end{bmatrix}$$

whose properties provide sufficient — but not necessary — conditions for an extremum: f has a (local) minimum whenever the Hessian is "positive definite" and a (local) maximum whenever it is "negative definite".

This Hessian is an example of a *(real) quadratic form*, to whose study this chapter is devoted.

If $\beta\colon V \times V \longrightarrow \mathbb{R}$ is a symmetric bi-linear form on V, a finitely generated real vector space, we can associate with it the real-valued function

$$q\colon V \longrightarrow \mathbb{R}, \quad \mathbf{v} \longmapsto \beta(\mathbf{v}, \mathbf{v})$$

Take a basis, $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$, for V.

Then $\mathbf{v} = x_1\mathbf{e}_1 + \cdots x_n\mathbf{e}_n$ for suitable $x_1, \ldots, x_n \in \mathbb{R}$, and it follows from the bi-linearity of β that

$$q(\mathbf{v}) = \sum_{i=1}^{n}\sum_{j=1}^{n} x_i x_j \beta(\mathbf{e}_i, \mathbf{e}_j)$$

or, putting $a_{ij} := \beta(\mathbf{e}_i, \mathbf{e}_j)$,

$$q(\mathbf{v}) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_i x_j \tag{16.1}$$

In other words, $q(\mathbf{v})$ is a homogeneous quadratic polynomial in n variables, viz. the co-ordinates of $\mathbf{v}$. In particular, given $\mathbf{v} \in V$ and all $\lambda \in \mathbb{R}$,

$$q(\lambda\mathbf{v}) = \lambda^2 q(\mathbf{v})$$

Definition 16.1. A *quadratic form* on the vector space, V, over the subfield, $\mathbb{F}$, of $\mathbb{R}$ is a function

$$q\colon V \longrightarrow \mathbb{F}$$

such that for all $\mathbf{x} \in V$ and $\lambda \in \mathbb{F}$

$$q(\lambda\mathbf{x}) = \lambda^2 q(\mathbf{x})$$

and that there is a symmetric bi-linear form

$$\beta\colon V \times V \longrightarrow \mathbb{F}$$

such that for all $\mathbf{x} \in V$

$$q(\mathbf{x}) = \beta(\mathbf{x}, \mathbf{x})$$

Given a basis $\mathcal{B} = \{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ for V, the matrix $\underline{\mathbf{A}}_q := \Big[\beta(\mathbf{e}_i, \mathbf{e}_j)\Big]_{n\times n}$ is the *matrix of the quadratic from q with respect to the basis $\mathcal{B}$.*

Observation 16.2. The matrix $\underline{\mathbf{A}}_q$ is a symmetric matrix, since it is the matrix of the symmetric bi-linear form in the definition of a quadratic form.

We defined quadratic forms in terms of bi-linear forms. In fact, real quadratic forms and real symmetric bi-linear forms completely determine each other.

Theorem 16.3. *Given a quadratic form q on the vector space V over $\mathbb{F} \subseteq \mathbb{R}$, there is a unique bi-linear form, β, on V such that*

$$q(\mathbf{v}) = \beta(\mathbf{v}, \mathbf{v})$$

for all $\in V$.

Proof. As the existence of such a bi-linear form is ensured by the definition of a quadratic form, it remains only to demonstrate its uniqueness.

But observe that

$$\begin{aligned} q(\mathbf{u}-\mathbf{v}) &= \beta(\mathbf{u}-\mathbf{v}, \mathbf{u}-\mathbf{v}) \\ &= q(\mathbf{u}) - 2\beta(\mathbf{u}, \mathbf{v}) + q(\mathbf{v}) && \text{as } \beta \text{ is bi-linear and symmetric} \\ q(\mathbf{u}+\mathbf{v}) &= \beta(\mathbf{u}+\mathbf{v}, \mathbf{u}+\mathbf{v}) \\ &= q(\mathbf{u}) + 2\beta(\mathbf{u}, \mathbf{v}) + q(\mathbf{v}) && \text{as } \beta \text{ is bi-linear and symmetric} \end{aligned}$$

Thus,

$$\beta(\mathbf{u}, \mathbf{v}) = \frac{1}{4}\left(q(\mathbf{u}+\mathbf{v}) - q(\mathbf{u}-\mathbf{v})\right).$$

□

Quadratic forms are real valued functions. So we may ask about the values they take.

Definition 16.4. The real quadratic form $q\colon V \longrightarrow \mathbb{R}$ is be *positive (negative) semi-definite* if and only if $q(\mathbf{v}) \geq 0$ (resp. $q(\mathbf{v}) \leq 0$) for all $\mathbf{v} \in V$.

It is *positive (negative) definite* if, in addition, $q(\mathbf{v}) = 0$ only for $\mathbf{v} = \mathbf{0}_V$.

Otherwise, it is *indefinite*.

Observation 16.5. The quadratic form derived from an inner product must be positive definite.

Example 16.6. We illustrate the above using quadratic forms $q\colon \mathbb{R}^3 \longrightarrow \mathbb{R}$.

(i) $q(x, y, z) := x^2 + y^2$ is, plainly, positve semi-definite. It is not positive definite, since $q(0, 0, 1) = 0$.

(ii) $q(x, y, z) := x^2 + y^2 - z^2$ is idefinite, for $q(1, 0, 0) = 1$, whereas $q(0, 0, 1) = -1$

(iii) $q(x, y, z) := -x^2 - y^3 - z^3$ is, clearly, negative definite.

We shall see (Sylvester's Theorem), that these examples are typical: every quadratic form is equivalent to one like the ones above.

We consider two quadratic forms to be equivalent if there is an automorphism of the vector space such that one form is the composite of the other with the automorphism. Formally,

Definition 16.7. The quadratic forms q and $\tilde{q}$ on the real vector space V are *equivalent* if and only if there is an isomorphism $\varphi\colon V \longrightarrow V$ such that $\tilde{q} = q \circ q$.

In the case of finitely generated real vector spaces, we can formulate this in terms of matrices.

Lemma 16.8. *Let $q\colon V \longrightarrow \mathbb{R}$ be a quadratic form on the finitely generated real vector space V and $\varphi\colon V \longrightarrow V$ a linear transformation.*

Then $q \circ \varphi\colon V \longrightarrow \mathbb{R}$ is also a quadratic form.

Moreover, if $\mathcal{B}$ is a basis for V, $\underline{\mathbf{A}}$ is the matrix q with respect to $\mathcal{B}$ and $\underline{\mathbf{B}}$ the matrix of φ with respect to $\mathcal{B}$, then the matrix of $q \circ \varphi$ with repect to $\mathcal{B}$ is

$$\underline{\mathbf{B}}^t \underline{\mathbf{A}}\, \underline{\mathbf{B}}$$

Proof. By Theorem 16.3, it is sufficient to prove the corresponding result for the bi-linear form determined by q. But that is precisely the content of Corollary 15.5 and Theorem 15.6 on page 193. □

Real quadratic forms are classified up to isomorphism by a triple of natural numbers, as the next theorem shows.

Theorem 16.9 (Sylvester's Theorem).

The real quadratic form $q\colon \mathbb{R}^n \longrightarrow \mathbb{R}$ is equivalent to one or the form

$$\sum_{i=1}^{r-s} x_i^2 - \sum_{i=r-s+1}^{r} x_j^2$$

with $0 \leq s \leq r \leq n$.

Moreover, the triple of natural numbers (n, r, s) determines q up to isomorphism.

Proof. It is sufficient to prove that there is a basis for $\mathbb{R}^n$ with respect to which the matrix of q is a diagonal matrix all of whose entries are 0 or ± 1.

By Corollary 15.30 on page 202 there is a basis, $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$, for $\mathbb{R}^n$ with respect to which the matrix of q is the diagonal matrix

$$\begin{bmatrix} d_1 & 0 & \cdots \\ 0 & d_2 & \\ \vdots & 0 & \ddots \end{bmatrix}$$

where we may assume that this basis has been so ordered that

$$\begin{aligned} d_i &> 0 \qquad \text{for } 1 \leq i \leq r-s \\ d_i &< 0 \qquad \text{for } r-s < i \leq r \\ d_i &= 0 \qquad \text{for } r < i \leq n \end{aligned}$$

Putting

$$\lambda_i := \begin{cases} \frac{1}{\sqrt{|d_i|}} & \text{for } i \leq r \\ 1 & \text{for } i > r \end{cases}$$

it follows immediately that the matrix of q with respect to the basis $\left\{\frac{1}{\sqrt{\lambda_i}}\mathbf{e}_i \mid 1 \leq i \leq n\right\}$ has the form required. □

16.1 Exercises

Exercise 16.1. Let $\beta : V \times V \longrightarrow \mathbb{R}$ be a symmetric bi-linear form on the real vector space V and $q : V \longrightarrow \mathbb{R}$ a quadratic form on V. Show that for all $\mathbf{u}, \in V$

(a) $\beta_{q_\beta}(\mathbf{u}, \mathbf{v}) = \beta(\mathbf{u}, \mathbf{v})$

(b) $q_{\beta_q}(\mathbf{u}) = q(\mathbf{u})$

Exercise 16.2. Let $\langle\!\langle\ ,\ \rangle\!\rangle$ be an inner product on the real vector space V and $\|\ \|$ the norm it induces. Decide whether

$$q : V \longrightarrow \mathbb{R}, \quad \mathbf{u} \longmapsto \|\mathbf{u}\|^2$$

defines a quadratic form on V.

Exercise 16.3. Let q be a positive definite quadratic form on the real vector space V. Prove that

$$\|\ \| : V \longrightarrow \mathbb{R}, \quad \mathbf{u} \longmapsto \sqrt{q(\mathbf{u})}$$

defines a norm on V.

Exercise 16.4. Classify each of the following bi-linear forms according to its definiteness property:

(a) $\beta : \mathbb{R}^3 \times \mathbb{R}^3 \longrightarrow \mathbb{R}, \quad ((u,v,w),(x,y,z)) \longmapsto 2ux + uy - 2uz + vx + 3vy - vz - 2wx - wy + 3wz$

(b) $\beta : \mathbb{R}^3 \times \mathbb{R}^3 \longrightarrow \mathbb{R}, \quad ((u,v,w),(x,y,z)) \longmapsto ux + 3uy + 3uz + 3vx + vy + vz + 3wx + wy + 2wz$

Index