Math 146 Notes - Linear Algebra

Laurent W. Marcoux

April 8, 2024

Preface

The following is a set of class notes for the Math 146 course I am currently teaching at the University of Waterloo in January, 2024. They are a work in progress, and – this being the "first edition" – they are replete with typos. A student should approach these notes with the same caution he or she would approach buzz saws; *they can be very useful, but you should be alert and thinking the whole time you have them in your hands.*

Just one short comment about the Exercises at the end of each chapter. These are of varying degrees of difficulty. Some are much easier than Assignment Questions, and some are of a comparable level of difficulty. Those few Supplementary Examples and Exercises that are marked by an asterisk are definitely worth doing, as they are crucial to understanding the underlying concepts. The marked exercises are also of varying levels of difficulty, but it is better for the reader to discover some things on his/her own, since the reader will then understand and retain those things better. The only way to learn mathematics is to do mathematics.

In our humble opinion, an excellent approach to reading these notes is as follows.

- One first gathers the examples of vector spaces from the second and third chapters. One then reads the statements of the theorems, propositions, corollaries, etc., and interprets those results for each of those examples. The purpose of the theory is to understand and unify the examples.
- To learn the proofs, we recommend that one read the statement of a given theorem or proposition, and tries to prove the result oneself. If one gets stuck at a certain point in the proof, one reads the proof until one gets past that point, and then one resumes the process of proving the result oneself.

Also, one should keep in mind that *if one doesn't know where to start, one can always start with the definition,* which means that one always knows where to start. Just saying.

I *strongly* recommend that the reader consult other textbooks as well as these notes. As ridiculous as this may sound, there are other people who can write as well as, if not better than, your humble author and it is important that the reader find the source which best suits the reader. Moreover, by consulting multiple sources, the reader will discover results not covered in any single reference. I shall only mention three references, namely the book of Friedberg, Insel and Spence [**FIS97**], the book

of Hoffman and Kunze [**HK71**], and the book of Strang [**Str88**]. The library will have other books which you may prefer to these.

I would like to thank (I didn't get the first name) Bell, J. Broden, J. Huang, S.L. Kaur, S. Li, V. Satish and L. Zhou for bring some typos to my attention. Any remaining typos and mistakes are the fault of my colleagues. You know which ones.

April 8, 2024

The reviews are in!

He is a writer for the ages, the ages of four to eight.

Dorothy Parker

This paperback is very interesting, but I find it will never replace a hardcover book - it makes a very poor doorstop.

Alfred Hitchcock

It was a book to kill time for those who like it better dead.

Rose Macaulay

That's not writing, that's typing.

Truman Capote

Only the mediocre are always at their best.

Jean Giraudoux

Contents

CHAPTER 1

The Axiom of Choice, posets and Zorn's Lemma

Sleep is my favourite thing in the world. It's the reason I get up in the morning.

Ross Smith

1. The Axiom of Choice

1.1. In this Chapter we shall acquire some *very* basic knowledge about the Axiom of Choice. A number of important results arising in different areas of modern mathematics are known to be equivalent to the Axiom of Choice, including one result which will play a crucial role for us, namely: Zorn's Lemma. Further equivalent formulations of the Axiom of Choice include the *Hahn-Banach Theorem* in Functional Analysis, *Krull's Theorem* in Ring Theory, *Tychonoff's Theorem* in topology, and – closer to our hearts in relation to this course – the fact that every vector space admits a Hamel basis.

While the Axiom of Choice is relatively easy to understand, and while it may in fact appear to be a self-evident truth, it implies things which on the surface appear to be either false, or impenetrable. To quote Jerry Bona:

> *The Axiom of Choice is obviously true, the well-ordering principle obviously false, and who can tell about Zorn's lemma?*

1.2. Let's begin with an analogy. Suppose that one is given a collection of boxes, and that in each box there is a pair of identical socks. A seemingly innocent question is whether one can pull a sock out of each box. Of course, if there are only finitely many boxes to begin with, then there is no problem in doing so. One pulls a sock out of the first box, then a second sock out of the second box, then a third sock from the third box, and so on, until one reaches the last box and completes the task in a most satisfactory if unfulfilling way. But what if the collection of boxes we are given is infinite? Since one can never stop the procedure, how do we know that we can actually obtain a sock from each and every box? Still, isn't it obvious that we can? What if we have uncountably many boxes – say, one box for each real number? Would that make a difference? We couldn't just order the boxes as we do the natural numbers. Still, what's the problem? After all, each box has a pair of

socks, and we know we can pull a sock out of any given box. And why do we need so many socks anyway?

1.3. To better understand the underlying issue from a mathematical viewpoint, let's begin by replacing "boxes" with sets, and "socks" with elements of those sets. We will also need to know what we mean by "picking an element from each set". To that end, we first define unions, intersections, and choice functions.

1.4. Definition. *Let $\varnothing \neq \Lambda, X$ be sets and suppose that $\{X_\lambda\}_{\lambda\in\Lambda}$ is a set of subsets of X. Then we define the* ***union*** *of the X_λ's to be*

$$\cup_{\lambda\in\Lambda} X_\lambda = \{x \in X : x \in X_\lambda \text{ for some } \lambda \in \Lambda\},$$

and the ***intersection*** *of the X_λ's to be*

$$\cap_{\lambda\in\Lambda} X_\lambda = \{x \in X : x \in X_\lambda \text{ for all } \lambda \in \Lambda\}.$$

Exercise: What should $\cap_{\lambda\in\varnothing} X_\lambda$ and $\cup_{\lambda\in\varnothing} X_\lambda$ mean?

1.5. Definition. *Let $\Lambda \neq \varnothing$ and let $\{X_\lambda\}_{\lambda\in\Lambda}$ be a set of subsets of a set X. We define the* ***product*** *of the sets X_λ to be:*

$$\prod_{\lambda\in\Lambda} X_\lambda = \{f : \Lambda \to \cup_{\lambda\in\Lambda} X_\lambda : f(\lambda) \in X_\lambda \text{ for all } \lambda \in \Lambda\}.$$

If such a function f exists, it is called a ***choice function****.*

Note: If $X_{\lambda_0} = \varnothing$ for some $\lambda_0 \in \Lambda$, then $f(\lambda_0) \in X_{\lambda_0}$ is false, and so $\prod_{\lambda\in\Lambda} X_\lambda = \varnothing$.

Given non-empty sets X and Y, we define

$$X^Y = \{f : Y \to X : f \text{ is a function}\} = \prod_{y\in Y} X_y,$$

where $X_y = X$ for all $y \in Y$.

1.6. Do choice functions always exist?

(a) Suppose that Λ is a finite, non-empty set and that for all $\lambda \in \Lambda$, $\varnothing \neq X_\lambda$ is a set. Then the answer is "yes". This follows from the basic axioms of *Zermelo-Fraenkel set theory.*

In particular, the basic axioms of Zermelo-Fraenkel set theory say that if you have a finite number of boxes, each of which contains a pair of socks, then you can take a sock out of each box.

The reader is referred to the excellent text of R. André [**And14**] for more detail about Set Theory, although I make no promises that André will mention socks anywhere in the text.

(b) Let Λ be an arbitrary non-empty set. For each $\lambda \in \Lambda$, suppose that $\varnothing \neq X_\lambda \subseteq \mathbb{N}$. Given $\lambda \in \Lambda$, define $f(\lambda)$ to be the least element of X_λ. Then $f \in \prod_{\lambda\in\Lambda} X_\lambda$ is a choice function. Note: here we have a *rule* for picking an element from each X_λ, and that rule applies to every X_λ *simultaneously*!

(c) Let Λ be an arbitrary non-empty set. Suppose that for each $\lambda \in \Lambda$, P_λ consists of a pair $\{L_\lambda, R_\lambda\}$ of shoes (where L_λ is the left shoe, and R_λ is the right shoe). Given $\lambda \in \Lambda$, set $g(\lambda) = L_\lambda$. Then $g \in \prod_{\lambda\in\Lambda} P_\lambda$ is a choice function.

Note: Once again we have a *rule* for picking an element from each X_λ, and that rule applies to every X_λ *simultaneously*!

(d) For each $n \geq 1$, let B_n denote a pair of identical socks. How do we specify a choice function $f \in \prod_{n\in\mathbb{N}} B_n$? Here, what rule do we have for selecting one sock from the pair? By Zermelo-Fraenkel set theory, we can do this for one pair at a time, but how do we do it simultaneously for all pairs?

1.7. The above question prompted the following quote from the mathematician (and philosopher) Bertrand Russel(1872-1970):

> To choose one sock from each of infinitely many pairs of socks requires the Axiom of Choice, but for shoes the Axiom is not needed.

As the reader will undoubtedly come to appreciate over their undergraduate career, it is twentieth century's obsession with socks which drove most of the mathematics discovered over the last 116 years.

1.8. One way to circumvent the question of how do we know that we can choose one sock from amongst each pair in an infinite collection of pairs of socks is to *assume* we can.

The Axiom of Choice [AC]. If $\Lambda \neq \varnothing$ is a set and for each $\lambda \in \Lambda$, X_λ is a non-empty subset of a set X, then $\prod_{\lambda\in\Lambda} X_\lambda \neq \varnothing$.

Exercise: Prove that the Axiom of Choice is equivalent to the following:

The Axiom of Choice - disjoint set version [ACD]. Suppose that $\Lambda \neq \varnothing$ is a set and that

(i) for all $\lambda \in \Lambda$, X_λ is a non-empty subset of a universe X, and
(ii) $X_\lambda \cap X_\beta = \varnothing$ if $\lambda \neq \beta \in \Lambda$.

Then $\prod_{\lambda\in\Lambda} X_\lambda \neq \varnothing$.

Exercise: Prove that the Axiom of Choice is equivalent to the following statement.

Given a non-empty set X there exists a function $f : \mathcal{P}(X) \setminus \{\varnothing\} \to X$ so that $f(A) \in A$ for all $A \in \mathcal{P}(X) \setminus \{\varnothing\}$.

1.9. At first glance, it would seem madness to even try to imagine that the Axiom of Choice is not true. As it turns out, we can appeal to the *Principle of you're damned if you do and you're damned if you don't* to begin to appreciate the can of worms we have just opened.

It can be shown (in fact it *has been* shown) that the Axiom of Choice implies the following: it is possible to "carve up" the unit ball in $\mathbb{R}^3$ into finitely many pieces and, using only rotations and translations, to reassemble those pieces into two balls each having the same volume as the original unit ball. This is known as the *Banach-Tarski Paradox*. As one might imagine, this result is non-constructive. It does not tell you *how* to cut the unit ball. It would be unwise yet strangely thirst-quenching to test this out on a bag of oranges using a typical kitchen knife.

On the other hand, the negation of (AC) implies the existence of two sets A and B so that neither of these can be mapped injectively into the other. It is unclear that this is a world in which we would like to live.

Our next goal is to obtain a couple of equivalent formulations of the Axiom of Choice which will prove useful both in analysis and in algebra. Before describing these equivalent formulations, we shall pause to develop some notation and definitions.

2. Partially ordered sets

2.1. We mentioned the Axiom of Choice in part because it is equivalent to Zorn's Lemma – which is the axiom we are *really* interested in in this course – and because it is much easier to interpret than Zorn's Lemma. To understand the statement of Zorn's Lemma, we first need to examine the concept of a *partially ordered set*, also known as *posets*. Understanding posets will also help us to understand what is meant by a *maximal* linearly independent set, or what is meant by the *smallest subspace* which contains a given set $\mathcal{S}$ of vectors in a vector space. But we are getting ahead of ourselves.

2.2. Definition. *A* ***relation*** R *on a set* X *is a subset of the Cartesian product* $X \times X = \{(x, y) : x, y \in X\}$. *We write* xRy *if* $(x, y) \in R$.

A relation $\leq$ *is called a* ***partial order*** *on* X *if it satisfies*

(i) $x \leq x$ *for all* $x \in X$ *(reflexivity);*
(ii) $x \leq y$ *and* $y \leq z$ *implies that* $x \leq z$ *(transitivity);*
(iii) $x \leq y$ *and* $y \leq x$ *implies that* $x = y$ *(anti-symmetry).*

The ordered pair $(X, \leq)$ *is called a* ***partially ordered set****, or simply a* ***poset****. Informally, it is also customary to refer to* X *as the poset with partial order* $\leq$.

A ***chain*** $\mathcal{C}$ *in* X *is a subset of* X *such that for any* $x, y \in \mathcal{C}$, *either* $x \leq y$ *or* $y \leq x$. *Alternatively, these are called* ***totally ordered sets*** *or* ***linearly ordered sets****.*

2.3. Example.

(a) $(\mathbb{R}, \leq)$ is a totally ordered (and hence a partially ordered) set using the usual order on $\mathbb{R}$. Similarly, $(\mathbb{Q}, \leq)$ is a totally ordered set using the same partial order.

(b) The list of words in the dictionary forms a totally ordered set with the usual lexicographic ordering.

2.4. Example. Let $X \neq \varnothing$ be a set. Consider the power set $\mathcal{P}(X)$. For $A, B \in \mathcal{P}(X)$, define $A \leq B$ to mean $A \subseteq B$. We say that $\mathcal{P}(X)$ is **(partially) ordered by inclusion**. Then $(\mathcal{P}(X), \leq)$ is a poset. If X has more than one element, then $(\mathcal{P}(X), \leq)$ is not a chain.

Suppose $X = \{1, 2, 3, 4, 5\}$ and that $\mathcal{P}(X)$ is ordered by inclusion. Then

$$\mathcal{C} = \{\{2\}, \{2, 5\}, \{2, 3, 5\}\}$$

is a chain in $\mathcal{P}(X)$. The set $\mathcal{D} = \{\{2\}, \{2, 5\}, \{1, 3, 5\}\}$ is not a chain.

2.5. Example. Let $X \neq \varnothing$ be a set. Consider the power set $\mathcal{P}(X)$. For $A, B \in \mathcal{P}(X)$, define $A \leq B$ to mean $A \supseteq B$. We say that $\mathcal{P}(X)$ is **ordered by containment**. Then $(\mathcal{P}(X), \leq)$ is a poset. If X has more than one element, then $(\mathcal{P}(X), \leq)$ is not a chain.

2.6. Example. Let

$$X = \mathcal{C}([0, 1], \mathbb{R}) := \{f : [0, 1] \to \mathbb{R} : f \text{ is continuous}\}.$$

For $f, g \in X$, define $f \leq g$ if $f(x) \leq g(x)$ for all $x \in [0, 1]$. Then $(X, \leq)$ is a partially ordered set.

2.7. Example. Consider $X = \mathbb{N}$, the set of positive integers, and for $m, n \in \mathbb{N}$, define $m \leq n$ if m divides n, written $m|n$. Then for $k, m, n \in \mathbb{N}$,

- $k|k$, so $k \leq k$.
- If $k \leq m$ and $m \leq k$, then k divides m and m divides k, so $k = m$.
- If $k \leq m$ and $m \leq n$, then $k|m$ – say $m = km_0$, and $m|n$ – say $n = mn_0$, and thus $n = km_0n_0$. Hence $k|n$, i.e. $k \leq n$.

Thus $(\mathbb{N}, \leq)$ is a partially ordered set. In this case, $2 \leq 4$, but $2 \nleq 3$.

2.8. Definition. *Let $(X, \leq)$ be a poset. We say that $x \in X$ is **maximal** in X if $y \in X$ and $x \leq y$ implies $x = y$. We say that $m \in X$ is a **maximum element** in X if $m \geq y$ for all $y \in X$.*

*We say that $z \in X$ is **minimal** in X if $y \in X$ and $y \leq z$ implies $y = z$. The element $n \in X$ is a **minimum element** in X if $y \in X$ implies that $n \leq y$.*

The distinction between a maximal element and a maximum element is that a *maximum* element must be comparable to (and at least as big as) *every* element of the poset $(X, \leq)$. A *maximal* element need only be as big as those elements in X

to which it is actually comparable. A analogous statement holds for *minimum* and *minimal* elements.

2.9. Example.

(a) Let $X = \{1, 2, 3, 4, 5, 6\}$, and denote by $\mathcal{P}_0(X)$ the collection of *proper* subsets of X, partially ordered by inclusion. (Recall that a subset $A \subseteq X$ is *proper* if $A \neq X$.) Then $N_1 = \{1, 2, 3, 4, 5\}$ and $N_2 = \{1, 3, 4, 5, 6\}$ are two distinct maximal elements of $\mathcal{P}_0(X)$. Neither of these is a maximum element; for example, $Y = \{6\} \in \mathcal{P}_0(X)$, but $Y \not\leq N_1$. In fact, $\mathcal{P}_0(X)$ does not have a maximum element at all.

(b) Let $X = (0, 1)$, equipped with the usual order inherited from $(\mathbb{R}, \leq)$. Again, X does not have a maximum element. In this case, it also does not have a maximal element. Moreover, $(\mathbb{R}, \leq)$ itself does not have any maximal elements.

(c) **Exercise:** Every finite poset has a maximal element. (It is also a worthwhile exercise to describe all 3 element posets to get a feeling for what is going on.)

2.10. Definition. *Let $(X, \leq)$ be a poset and $A \subseteq X$. We say that $y \in X$ is an **upper bound** for A if $a \leq y$ for all $a \in A$. We say that $x \in X$ is a **lower bound** for A if $x \leq a$ for all $a \in A$.*

*We say that $\beta \in X$ is the **least upper bound (lub)**, or **supremum (sup)** for A if*

- *β is an upper bound for A, and*
- *if y is any upper bound for A, then $\beta \leq y$.*

*Similarly, we say that $\alpha \in X$ is the **greatest lower bound (glb)** or **infimum (inf)** for A if*

- *α is a lower bound for A, and*
- *if x is any lower bound for A, then $x \leq \alpha$.*

2.11. Example.

(a) $(\mathbb{R}, \leq)$ has the **least upper bound property**, where $\leq$ is the usual ordering on $\mathbb{R}$. If $\varnothing \neq A \subseteq \mathbb{R}$ is bounded above, then A has a least upper bound β.

If $A = \varnothing$, then for any $b \in \mathbb{R}$, b is an upper bound for A. Indeed, if b were not an upper bound for $A = \varnothing$, then there would exist an element $a \in A$ such that $b \not\leq a$, which is false. Since b is an upper bound for $\varnothing$ for all $b \in \mathbb{R}$, we say that the *least upper bound* of $\varnothing$ is $-\infty$.

Here, $-\infty$ is *not* a number! The statement $\sup \varnothing = -\infty$ is to be interpreted as saying that any $b \in \mathbb{R}$ is an upper bound for $\varnothing$.

Note that using the same logic, we write $\infty = \inf \varnothing$, as every $b \in \mathbb{R}$ is also a lower bound for $\varnothing$.

(b) Let X be a non-empty set and let $\mathcal{P}(X)$ denote its power set, partially ordered by inclusion. If $\{X_\lambda\}_{\lambda\in\Lambda} \subseteq \mathcal{P}(X)$, then $\cup_{\lambda\in\Lambda} X_\lambda$ is the l.u.b. of $\{X_\lambda\}_{\lambda\in\Lambda}$, and $\cap_{\lambda\in\Lambda} X_\lambda$ is the g.l.b. of $\{X_\lambda\}_{\lambda\in\Lambda}$.

(c) Consider $(\mathbb{Q}, \leq)$ where $\leq$ denotes the usual total order inherited from $\mathbb{R}$. The set $A = \{x \in \mathbb{Q} : x^2 < 2\}$ is bounded above, but there is no least upper bound for A *in* $\mathbb{Q}$. Indeed, if $b \in \mathbb{Q}$ and $b > \sqrt{2}$, then there exists another rational number $d \in (\sqrt{2}, b)$, and thus d is an upper bound for A and $d < b$, so b is not the supremum of A. If $b \in \mathbb{Q}$ and $b < \sqrt{2}$, then clearly b is not even an upper bound for A, so it is not a supremum for A.

The Axiom of Choice was introduced by Zermelo in order to prove his Well-ordering Principle. To explain this, we first need a couple of definitions.

2.12. Definition. *A non-empty poset $(X, \leq)$ is said to be* ***well-ordered*** *if every non-empty subset $A \subseteq X$ has a mimimum element.*

It immediately follows that every well-ordered set is totally ordered, since if $x, y \in X$ and $(X, \leq)$ is well-ordered, then either $x = \min(x, y)$ and so $x \leq y$, or $y = \min(x, y)$ and so $y \leq x$. In other words, any two elements in X can be compared.

2.13. Example.

(a) The set $\mathbb{N}$ is well-ordered with the usual ordering, whereas $\mathbb{R}$ is not.

(b) Let $\omega + 7 = \{1, 2, 3, ..., \omega, \omega + 1, \omega + 2, ..., \omega + 6\}$. Define a partial order on $\omega + 7$ by setting $n \leq \omega + k$ for all $n \geq 1$, $0 \leq k \leq 6$ and $\omega + i \leq \omega + j$ if $0 \leq i \leq j \leq 6$. The ordering on $\mathbb{N} \subseteq \omega + 7$ is the usual ordering on $\mathbb{N}$. Then $\omega + 7$ is well-ordered.

2.14. Theorem. *The following are equivalent:*

(i) *The Axiom of Choice (AC): given a non-empty collection $\{X_\lambda\}_{\lambda\in\Lambda}$ of non-empty sets, $\prod_{\lambda\in\Lambda} X_\lambda \neq \varnothing$.*

(ii) *Zorn's Lemma (ZL): Let $(Y, \leq)$ be a poset. Suppose that every chain $\mathcal{C} \subseteq Y$ has an upper bound. Then Y has a maximal element.*

(iii) *The Well-Ordering Principle (WO): Every non-empty set Z admits a well-ordering.*

Proof. This result has been moved to PM433. You may consult the appendix to this Chapter for a proof.

□

2.15. Remark. A number of other results are known to be equivalent to the Axiom of Choice. We mention only two:

- If X and Y are non-empty, disjoint sets and X is infinite, then there exists a bijection between $X \times Y := \{(x, y) : x \in X, y \in Y\}$, and $X \cup Y$.
- If X is an infinite set, then there exists a bijection between X and $X \times X$.

Appendix

A1.1. In this Appendix we shall provide a proof of the equivalence of the Axiom of Choice, Zorn's Lemma and the Well-Ordering Principle. We begin with the definition of an **initial segment**, which will be required in the proof.

A1.2. *Let $(X, \leq)$ be a poset, $C \subseteq X$ be a chain in X and $d \in C$. We define*

$$P(C, d) = \{c \in C : c < d\}.$$

*An **initial segment** of C is a subset of the form $P(C, d)$ for some $d \in C$.*

A1.3.

(a) For each $r \in \mathbb{R}$, $(-\infty, r)$ is an initial segment of $(\mathbb{R}, \leq)$.
(b) For each $n \in \mathbb{N}$, $\{1, 2, ..., n\}$ is an initial segment of $\mathbb{N}$.

A1.4. Theorem. *The following are equivalent:*

(i) *The Axiom of Choice (AC): given a non-empty collection $\{X_\lambda\}_{\lambda \in \Lambda}$ of non-empty sets, $\prod_{\lambda \in \Lambda} X_\lambda \neq \varnothing$.*
(ii) *Zorn's Lemma (ZL): Let $(Y, \leq)$ be a poset. Suppose that every chain $\mathcal{C} \subseteq Y$ has an upper bound. Then Y has a maximal element.*
(iii) *The Well-Ordering Principle (WO): Every non-empty set Z admits a well-ordering.*

Proof.

(i) implies (ii): This is the most delicate of the three implications. We shall argue by contradiction.

Suppose that $(X, \leq)$ is a poset such that every chain in X is bounded above, but that X no maximal elements. Given a chain $C \subseteq X$, we can find an upper bound u_C for C. Since u_C is not a maximal element, we can find $v_C \in X$ with $u_C < v_C$. We shall refer to such an element v_C as a **strict upper bound** for C.

By the Axiom of Choice, for each chain C in X, we can choose a strict upper bound $f(C)$. If $C = \varnothing$, we arbitrarily select $x_0 \in X$ and set $f(\varnothing) = x_0$.

We shall say that a subset $A \subseteq X$ satisfies **property L** if

(I) The partial order $\leq$ on X when restricted to A is a well-ordering of A, and
(II) for all $x \in A$, $x = f(P(A, x))$.

• **Claim 1:** *if $A, B \subseteq X$ satisfy property L and $A \neq B$, then either A is an initial segment of B, or B is an initial segment of A.*

Without loss of generality, we may assume that $A \setminus B \neq \varnothing$. Let

$$x = \min \{a \in A : a \notin B\}.$$

Note that x exists because A is well-ordered. Then $P(A,x) \subseteq B$. We shall argue that $B = P(A,x)$. If not, then $B \setminus P(A,x) \neq \varnothing$, and using the well-orderedness of B,

$$y = \min\{b \in B : b \notin P(A,x)\}$$

exists. Thus $P(B,y) \subseteq P(A,x)$.

Let $z = \min(A \setminus P(B,y))$. Then $z \leq x = \min(A \setminus B)$.

- **Subclaim 1:** $P(A,z) = P(B,y)$.
  By definition, $P(A,z) \subseteq P(B,y)$.
  To obtain the reverse inclusion, we first argue that if $t \in P(B,y) = A \cap P(B,y)$, then $P(A,t) \cup \{t\} \subseteq P(B,y)$. By hypothesis, $t \in P(B,y)$, so suppose that $u \in P(A,t)$. Now $t \in P(B,y) \subseteq P(A,x)$, so $u < t < x$ implies that $u \in P(A,x)$. In other words, $P(A,t) \subseteq P(A,x) \subseteq B$. But then $u \in B$ and $u < t < y$ implies that $u \in P(B,y)$.
  We now have that if $s \in P(B,y)$, then $P(A,s) \cup \{s\} \subseteq P(B,y) \subseteq P(A,x) \subseteq A$. This forces $s < z := \min(A \setminus P(B,y))$, so that $s \in P(A,z)$. Together, we find that $P(B,y) \subseteq P(A,z) \subseteq P(B,y)$, which proves the subclaim.

Returning to the proof of the claim, we now have that $z = f(P(A,z)) = f(P(B,y)) = y$. But $y \in B$, so $y \neq x$. Hence $z < x$. Thus $y = z \in P(A,x)$, contradicting the definition of y. We deduce that $P(A,x) = B$, and hence that B is an initial segment of A, thereby proving our claim.

Suppose that $A \subseteq X$ has property L, and let $x \in A$. It follows from the above argument that given $y < x$, either $y \in A$ or y does not belong to any set B with property L.

Let $V = \cup\{A \subseteq X : A \text{ has property } L\}$.

- **Claim 2:** *We claim that if $w = f(V)$, then $V \cup \{w\}$ has property L.*

Suppose that we can show this. Then $V \cup \{w\} \subseteq V$, so $w \in V$, a contradiction. This will complete the proof.

• **Subclaim 2a:** *First we show that V itself has property L.* We must show that V is well-ordered, and that for all $x \in V$, $x = f(P(V,x))$.

(a) V is well-ordered.
  Let $\varnothing \neq B \subseteq V$. Then there exists $A_0 \subseteq X$ so that A_0 has property L and $B \cap A_0 \neq \varnothing$. Since A_0 is well-ordered and $\varnothing \neq B \cap A_0 \subseteq A_0$, $m := \min(B \cap A_0)$ exists. We claim that $m = \min(B)$.
  Suppose that $y \in B$. Then there exists $A_1 \subseteq X$ so that A_1 has property L and $y \in A_1$. Now, both A_0 and A_1 have property L:
  - ⋄ if $A_0 = A_1$, then $m = \min(B \cap A_1)$, so $m \leq y$.
  - ⋄ if $A_0 \neq A_1$, then either

- A_0 is an initial segment of A_1, so $A_0 = P(A_1, d)$ for some $d \in A_1$. Then

  $$m = \min(B \cap A_0) = \min(B \cap A_1),$$

  since $r \in A_1 \setminus A_0$ implies that $m < d \leq r$. Hence $m \leq y$;, or
- A_1 is an initial segment of A_0, say $A_1 = P(A_0, d) \subseteq A_0$ for some $d \in A_0$. Then

  $$m = \min(B \cap A_0) \leq \min(B \cap A_1).$$

  Hence $m \leq y$.

In both cases we see that $m \leq y$. Since $y \in B$ was arbitrary, $m = \min(B)$.
Thus, any non-empty subset B of V has a minimum element, and so V is well-ordered.

(b) Let $x \in V$. Then there exists $A_2 \subseteq X$ with property L so that $x \in A_2$. Then $x = P(A_2, x)$. Suppose that $y \in V$ and $y < x$. Then there exists $A_3 \subseteq X$ with property L so that $y \in A_3$. Since A_2 and A_3 both have property L, either
- $A_2 = A_3$, and so $y \in A_2$; or
- $A_2 = P(A_3, d)$ for some $d \in A_3$. Since $x \in A_2$, $P(A_2, x) = P(A_3, x)$ and therefore $y \in A_2$; or
- $A_3 = P(A_2, d)$ for some $d \in A_2$. Then $y \in A_3$ implies that $y \in A_2$.

In any of these three cases, $y \in A_2$. Hence $P(V, x) \subseteq P(A_2, x)$. Since $A_2 \subseteq V$, we have that $P(A_2, x) \subseteq P(V, x)$, whence $P(A_2, x) = P(V, x)$. But then

$$x = f(P(A_2, x)) = f(P(V, x)).$$

By (a) and (b), V has property L.

We now return to the proof of Claim 2. That is, we prove that if $w = f(V)$, then $V \cup \{w\}$ has property L.

(I) $V \cup \{w\}$ is well-ordered.
We know that V is well-ordered by part (a) above. Suppose that $\varnothing \neq B \subseteq V \cup \{w\}$. If $B \cap V \neq \varnothing$, then by (a) above, $m := \min(B \cap V)$ exists. Clearly $m \in V$ implies $m \leq f(V) = w$, so $m = \min(B \cap (V \cup \{w\}))$. If $\varnothing \neq B \subseteq V \cup \{w\}$ and $B \cap V = \varnothing$, then $B = \{w\}$, and so $w = \min(B)$ exists.
Hence $V \cup \{w\}$ is well-ordered.

(II) Let $x = V \cup \{w\}$. If $x \in V$, then $x = f(P(V, x))$ by part (a). If $x = w$, then

$$P(V \cup \{w\}, x) = P(V \cup \{w\}, w) = V,$$

so $x = w = f(V) = f(P(V \cup \{w\}, x))$.

By (I) and (II), $V \cup \{w\}$ has property L. As we saw in the statement following Claim 2, this completes the proof that the Axiom of Choice implies Zorn's Lemma. Now let us never speak of this again.

(ii) implies (iii): Let $X \neq \varnothing$ be a set. It is clear that every finite subset $F \subseteq X$ can be well-ordered. Let $\mathcal{A}$ denote the collection of pairs $(Y, \leq_Y)$, where $Y \subseteq X$ and $\leq_Y$ is a well-ordering of Y. For $(A, \leq_A)$, $(B, \leq_B) \in \mathcal{A}$, observe that A is an **initial segment** of B if the following two conditions are met:

- $A \subseteq B$ and $a_1 \leq_A a_2$ implies that $a_1 \leq_B a_2$;
- if $b \in B \setminus A$, then $a \leq_B b$ for all $a \in A$.

Let us partially order $\mathcal{A}$ by setting $(A, \leq_A) \leq (B, \leq_B)$ if A is an initial segment of B. Let $\mathcal{C} = \{C_\lambda\}_{\lambda \in \Lambda}$ be a chain in $\mathcal{A}$.

Then (**exercise**): $\cup_{\lambda \in \Lambda} C_\lambda$ is an upper bound for $\mathcal{C}$.

By Zorn's Lemma, $\mathcal{A}$ admits a maximal element, say $(M, \leq_M)$. We claim that $M = X$. Suppose otherwise. Then we can choose $x_0 \in X \setminus M$ and set $M_0 = M \cup \{x_0\}$. define a partial order on M_0 via: $x \leq_{M_0} y$ if either (a) $x, y \in M$ and $x \leq_M y$, or (b) x is arbitrary and $y = x_0$. Then $(M_0, \leq_{M_0})$ is a well-ordered set and $(M, \leq_M) < (M_0, \leq_{M_0})$, a contradiction of the maximality of $(M, \leq_M)$. Thus $M = X$ and $\leq_M$ is a well- ordering of X.

(iii) implies (i): Suppose that $\{X_\lambda\}_{\lambda \in \Lambda}$ is a non-empty collection of non-empty sets. Let $X = \cup_{\lambda \in \Lambda} X_\lambda$. By hypothesis, X admits a well-ordering $\leq_X$. Since each $\varnothing \neq X_\lambda \subseteq X$, it has a minimum element relative to the ordering on X. Define a choice function f by setting $f(\lambda)$ to be this minimum element of X_λ for each $\lambda \in \Lambda$.

□

We include a proof of an exercise mentioned earlier in the notes:

A1.5. Proposition. *The following are equivalent:*

(a) *The Axiom of choice: if $\Lambda \neq \varnothing$ and for each $\lambda \in \Lambda$ there exists a non-empty set X_λ, then*

$$\Pi_{\lambda \in \Lambda} X_\lambda \neq \varnothing.$$

(b) *If $\varnothing \neq \varnothing$, then there exists a function*

$$g : \mathcal{P}(X) \setminus \{\varnothing\} \to X$$

such that $g(Y) \in Y$ for all $Y \subseteq X$.

Proof.

(a) implies (b).

Suppose (a) holds. Let $\varnothing \neq X$ be a set and set $\Lambda = \mathcal{P}(X) \setminus \{\varnothing\}$. For each $Y \in \Lambda$, set $Z_Y = Y \neq \varnothing$.

By the Axiom of Choice, there exists a choice function

$$f \in \Pi_{Y \in \Lambda} Z_Y = \Pi_{Y \in \Lambda} Y.$$

But then $f(Y) \in Z_Y = Y$ for each $Y \in \Lambda = \mathcal{P}(X) \setminus \{\varnothing\}$.

That is, (b) holds.

(b) implies (a).

Suppose that (b) holds.

Let $\varnothing \neq \Lambda$ be a set and suppose that X_λ is a non-empty set for each $\lambda \in \Lambda$. Let $Y = \cup_{\lambda\in\Lambda} X_\lambda$.

By hypothesis, there exists a function $g : \mathcal{P}(Y) \setminus \{\varnothing\} \to Y$ so that $g(W) \in W$ for all $W \in \mathcal{P}(Y) \setminus \{\varnothing\}$. In particular, each $X_\lambda \in \mathcal{P}(Y) \setminus \{\varnothing\}$, and so $g(X_\lambda) \in X_\lambda$ for all $\lambda \in \Lambda$.

Define $f(\lambda) = g(X_\lambda)$, $\lambda \in \Lambda$. Then f is a choice function, so (a) holds.

□

A1.6. Culture.

(a) The basic axioms of set theory are referred to as the **Zermelo-Fraenkel Axioms**, or (ZF).

Gödel proved that the Axiom of Choice is consistent with (ZF), but that (ZF) does not by itself imply the Axiom of Choice. Cohen then developed the theory of "forcing" to prove that (ZF) plus the *negation* of the Axiom of Choice is also consistent.

(b) It is known that the Riemann hypothesis is true in (ZF) if and only if it is true in (ZFC), namely (ZF) plus the Axiom of Choice.

(c) The generalized Continuum hypothesis (GCH) is known to be independent of (ZFC), however (ZF) plus (GCH) together imply the Axiom of Choice (AC).

(d) Tarski tried to publish the result which says that the Axiom of Choice (AC) is equivalent to the assertion that $|A| = |A \times A|$ whenever A is infinite in the *Comptes Rendus*. It was not accepted. Fréchet said that the equivalence of two true statements is not something new, while Lebesgue said that any implication between two false propositions is of no interest.

A1.7. In our definition of unions and intersections of sets (Definition 1.4), we first specified a "*universe*" X and then required each of the sets X_λ to be a subset of that universe X. There is a good reason for this, namely: there is no universe of "all sets".

Suppose to the contrary that V is the "set of all sets". We can then use the basic axioms of set theory (in particular the so-called **Axiom of Subsets** to define the set

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

Thus $y \in A$ if and only if $y \notin y$. Then A is a set, and since B is the "set of all sets", $A \in B$. This raises the question: is $A \in A$?

- If $A \in A$, then $A \in B$ and $A \in A$, so by definition of A, $A \notin A$, a contradiction.
- We conclude that $A \notin A$. But then $A \in B$ and $A \notin A$, so $A \in A$, a contradiction.

The problem is that we supposed that there exists a "universal set" V containing all sets. This doesn't happen. Thus, when dealing with problems in set theory, we should first define which collection of sets we wish to consider – i.e. we should define the "universe" under discussion.

Exercises for Chapter 1.

Exercise 1.1.
Let X be a set with $0 \leq n < \infty$ elements. Prove that the **power set**

$$\mathcal{P}(X) := \{Y : Y \subseteq X\}$$

of X has 2^n elements.

Exercise 1.2.
Given a set $\varnothing \neq X$ and a collection of subsets X_α, $\alpha \in \Lambda$ of X, what does $\cup_{\lambda \in \varnothing} X_\lambda$ mean? What does $\cap_{\lambda \in \varnothing} X_\lambda$ mean?

Exercise 1.3.
Prove that the Axiom of Choice is equivalent to each of the following:

(a) **The Axiom of Choice - disjoint set version [ACD].** Suppose that $\Lambda \neq \varnothing$ and that
   (i) for all $\lambda \in \Lambda$, X_λ is a non-empty set, and
   (ii) $X_\lambda \cap X_\beta = \varnothing$ if $\lambda \neq \beta \in \Lambda$.
   Then $\prod_{\lambda \in \Lambda} X_\lambda \neq \varnothing$.

(b) Given a non-empty set X there exists a function $f : \mathcal{P}(X) \setminus \varnothing \to X$ such that $f(A) \in A$ for all $A \in \mathcal{P}(X) \setminus \varnothing$.

Exercise 1.4.
Prove that every finite poset $(X, \leq)$ has a maximal element and a minimal element. Give examples of a finite poset $(X, \leq)$ where X has a maximum element but no minimum element, and examples where it has both a maximum and a minimum element.

CHAPTER 2

Vector spaces and subspaces

Somewhere on this globe, every ten seconds, there is a woman giving birth to a child. She must be found and stopped.

Sam Levenson

1. Vector spaces: examples, definitions and very basic facts

1.1. Pure Mathematics is the study of mathematical objects and of the relationships between them. When we find enough interesting examples of a given phenomenon, we establish a definition to describe that property. The extent to which the definition can be used to predict and explain new phenomena determines its value. Few definitions are more useful and pervasive than that of a vector space.

Let us begin by examining a few seemingly disparate examples of mathematical objects, and let us try to find a commonality amongst them.

Throughout these notes, we will be dealing with ordered pairs $(\mathcal{V}, \mathbb{F})$, where $\mathcal{V}$ is a certain non-empty set which we will soon call a **vector space**, and $\mathbb{F}$ is a field. Amongst the most important examples of vector spaces are those for which $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$. Certainly results hold equally well in both of these cases with essentially identical proofs. In such cases, rather that repeating the same proof twice, we simply write $\mathbb{K}$ to mean either of $\mathbb{R}$ or $\mathbb{C}$.

We begin with a definition which will be used throughout the course, and far, far beyond.

1.2. Definition. *Let $\mathbb{F}$ be a field and $m, n \in \mathbb{N}$. An $m \times n$ **matrix over** $\mathbb{F}$ is a function*

$$\begin{array}{rccc} a: & \{1,2,\ldots,m\} \times \{1,2,\ldots,n\} & \to & \mathbb{F} \\ & (i,j) & \mapsto & a_{ij}. \end{array}$$

We typically write this function in the form

$$a = [a_{ij}]_{m\times n} \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}.$$

(If m and n are understood, we sometimes abbreviate this to $a = [a_{ij}]$.

The collection of all $m \times n$ matrices over $\mathbb{F}$ is denoted by $\mathbb{M}_{m\times n}(\mathbb{F})$, and when $m = n$, we abbreviate this to $\mathbb{M}_n(\mathbb{F})$.

*Given $a = [a_{ij}] \in \mathbb{M}_{m\times n}(\mathbb{F})$, we define the **transpose** of a to be*

$$a^t := [a_{ji}] \in \mathbb{M}_{n\times m}(\mathbb{F}).$$

If $a \in \mathbb{M}_{m\times n}(\mathbb{F})$, it is clear that $(a^t)^t = a$.

We allow ourselves a mild abuse of notation by conflating $a = (a_1, a_2, \ldots, a_n)$ with $a = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix}$. Thus we have that

$$(a_1, a_2, \ldots, a_n)^{\mathrm{t}} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}.$$

An element of $\mathbb{M}_{1\times n}(\mathbb{F})$ is often referred to as a **row vector**, while an element of $\mathbb{M}_{m\times 1}(\mathbb{F})$ is often referred to as a **column vector**. Clearly there is a bijection between $\mathbb{M}_{1\times n}(\mathbb{F})$ and $\mathbb{M}_{n\times 1}(\mathbb{F})$ induced by the transpose map $a \mapsto a^{\mathrm{t}}$.

We next consider a number of familiar mathematical sets, which we observe to have a common structure, namely: in each case, we can add members of the set, as well as multiply elements of the set by scalars.

1.3. Example. Let $\mathbb{F}$ be a field, $m \geq 1$ be an integer, and consider the set $\mathcal{V} := \mathbb{F}^m := \mathbb{F} \times \mathbb{F} \times \cdots \times \mathbb{F}$ (m times). That is,

$$\mathcal{V} = \{(x_1, x_2, \ldots, x_m) : x_j \in \mathbb{F}, 1 \leq j \leq m\}.$$

We define two operations $+ : \mathcal{V} \times \mathcal{V} \to \mathcal{V}$ and $\cdot : \mathbb{F} \times \mathcal{V} \to \mathcal{V}$ as follows. For $x = (x_1, x_2, \ldots, x_m), y = (y_1, y_2, \ldots, y_m) \in \mathcal{V}$ and $\kappa \in \mathbb{F}$, we set

$$x + y := (x_1 + y_1, x_2 + y_2, \ldots, x_m + y_m), \qquad \text{and}$$

$$\kappa \cdot x := (\kappa x_1, \kappa x_2, \ldots, \kappa x_m).$$

Thus $\mathbb{F}^m = \mathbb{M}_{1\times m}(\mathbb{F})$ is the set of all row vectors of length m with entries in $\mathbb{F}$. In many instances (we shall come across such instances when studying linear maps between finite-dimensional vector spaces and their representations as matrices), there is a good reason to denote the elements of $\mathcal{V} = \mathbb{F}^m$ as column vectors, that is, as elements of $\mathbb{M}_{m\times 1}(\mathbb{F})$. Of course, when representing elements of $\mathbb{F}^m$ as column vectors, the operations become: given $x, y \in \mathcal{V} = \mathbb{F}^m$,

$$x + y := \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_m + y_m \end{bmatrix} \qquad \text{and} \qquad \kappa \cdot x := \begin{bmatrix} \kappa x_1 \\ \kappa x_2 \\ \vdots \\ \kappa x_m \end{bmatrix}.$$

Note that amongst other properties, $x + y = y + x$ and $\kappa(x + y) = \kappa x + \kappa y$ for all $x, y \in \mathcal{V}$ and $\kappa \in \mathbb{F}$. If $\mathbf{0} := (0, 0, \ldots, 0) \in \mathcal{V} = \mathbb{F}^m$, then $\mathbf{0} + x = x$ for all $x \in \mathbb{F}^m$.

We emphasise that the idea of expressing elements of $\mathbb{F}^m$ either as row vectors or as column vectors is not simply a fetish of the current author, but rather a *common practice*. It should not cause problems. Furthermore, it is also a common practice (which we shall often adopt) to simply write κx to mean $\kappa \cdot x$ when $\kappa \in \mathbb{F}$ and $x \in \mathcal{V}$.

1.4. Example. Let

$$\mathcal{C}([0,1],\mathbb{K}) := \{f : [0,1] \to \mathbb{K} : f \text{ is continuous}\}.$$

As we always do with functions, given $f, g \in \mathcal{C}([0,1],\mathbb{K})$ and $\kappa \in \mathbb{K}$, we define (for all $x \in [0,1]$)

$$(f+g)(x) = f(x) + g(x),$$

and

$$(\kappa f)(x) = \kappa(f(x)).$$

That is, we define addition and scalar multiplication *pointwise*.

From Calculus, we know that $f+g$ and $\kappa f \in \mathcal{C}([0,1],\mathbb{K})$, and that $f+g = g+f$ and $\kappa(f+g) = \kappa f + \kappa g$ when $\kappa \in \mathbb{K}$. If $z(x) = 0$ for all $x \in [0,1]$, then $z \in \mathcal{C}([0,1],\mathbb{K})$, and $f + z = f = z + f$ for all $f \in \mathcal{C}([0,1],\mathbb{K})$.

1.5. Example. Let $\mathbb{F}$ be a field and consider

$$\mathbb{T}_2(\mathbb{F}) := \left\{ \begin{bmatrix} a & b \\ 0 & d \end{bmatrix} : a, b, d \in \mathbb{F} \right\}.$$

Given $A = \begin{bmatrix} a_1 & b_1 \\ 0 & d_1 \end{bmatrix}$ and $B = \begin{bmatrix} a_2 & b_2 \\ 0 & d_2 \end{bmatrix} \in \mathbb{T}_2(\mathbb{F})$, define

$$A + B := \begin{bmatrix} a_1 + a_2 & b_1 + b_2 \\ 0 & d_1 + d_2 \end{bmatrix},$$

and for $\kappa \in \mathbb{F}$,

$$\kappa A = \begin{bmatrix} \kappa a_1 & \kappa b_1 \\ 0 & \kappa d_1 \end{bmatrix}.$$

Note that $A + B, \kappa A \in \mathbb{T}_2(\mathbb{F})$, $A + B = B + A$ and $\kappa(A+B) = \kappa A + \kappa B$. Moreover, if $Z := \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$, then $Z \in \mathbb{T}_2(\mathbb{K})$ and $A + Z = A = Z + A$ for all $A \in \mathbb{T}_2(\mathbb{F})$.

1.6. Notation. Given non-empty sets A and B, we write B^A to denote the set of all functions from A to B. That is,

$$B^A := \{f : A \to B : f \text{ is a function}\}.$$

This raises the question: doesn't this contradict our previous notation of $\mathbb{F}^m$?

The answer may surprise you. Given $m \in \mathbb{N}$, if we identify m with the set $\Lambda_m := \{1, 2, \ldots, m\}$, then

$$\mathbb{F}^{\Lambda_m} := \{f : \Lambda_m \to \mathbb{F} : f \text{ is a function}\} = \{f : \{1, 2, \ldots, m\} \to \mathbb{F} : f \text{ is a function}\}.$$

But a function is determined by its value at every point in the domain (i.e. if two functions $f, g : A \to B$ satisfy $f(x) = g(x)$ for all $x \in A$, then $f = g$), and so we can alternatively define

$$\mathbb{F}^{\Lambda_m} := \{(f(1), f(2), \ldots, f(m)) : f \text{ a function from } \Lambda_m \text{ to } \mathbb{F}\}.$$

(That is, if we know the function $f : \{1, 2, \ldots, m\} \to \mathbb{F}$, then we know $f(j) \in \mathbb{F}$ for all $1 \le j \le m$, and conversely, if we know the values of $f(j)$, $1 \le j \le m$, then we know exactly which function f is.) Since $x_k := f(k) \in \mathbb{F}$ can be arbitrary, we see that

$$\mathbb{F}^{\Lambda_m} = \{(x_1, x_2, \ldots, x_m) : x_j \in \mathbb{F}, 1 \le j \le m\}$$

is precisely what we call $\mathbb{F}^m$. In other words, $\mathbb{F}^m$ is just an abbreviation for $\mathbb{F}^{\Lambda_m}$!

1.7. Example. The next example is not so familiar, but demonstrates how we may construct new, abstract, sets which behave in much the same way as the familiar sets we have listed above.

Consider $\mathcal{V} := \{\alpha\,\text{PIG} + \beta\,\text{DOG} : \alpha, \beta \in \mathbb{F}\}$, with the understanding that

$$\alpha_1\,\text{PIG} + \beta_1\,\text{DOG} = \alpha_2\,\text{PIG} + \beta_2\,\text{DOG}$$

if and only if $\alpha_1 = \alpha_2$ and $\beta_1 = \beta_2$. Given $x := \alpha_1\text{PIG} + \beta_1\text{DOG}$ and $y := \alpha_2\text{PIG} + \beta_2\text{DOG}$, and given $\kappa \in \mathbb{F}$, define

$$x + y := (\alpha_1 + \alpha_2)\text{PIG} + (\beta_1 + \beta_2)\text{DOG}$$

and

$$\kappa x := (\kappa\alpha_1)\text{PIG} + (\kappa\beta_1)\text{DOG}.$$

Again, $x + y$ and $\kappa x \in \mathcal{V}$.

In fact, if $z = \alpha_3\text{PIG} + \beta_3\text{DOG}$, then $x + y = y + x$, $(x + y) + z = x + (y + z)$, $\kappa_1(\kappa_2 x) = (\kappa_1\kappa_2)x$, and more. Exactly how much more we shall now see. For example, note that $e := 0\text{PIG} + 0\text{DOG}$ has the property that $x + e = x = e + x$ for all $x \in \mathcal{V}$, and that if $x = \alpha\text{PIG} + \beta\text{DOG}$, then $y := (-\alpha)\text{PIG} + (-\beta)\text{DOG}$ has the property that $x + y = e = y + x$.

The commonality in the above examples leads us to invent the following definition.

1.8. Definition. *A **vector space** (or **linear space**) **over a field** $\mathbb{F}$ consists of a non-empty set $\mathcal{V}$ equipped with two binary operations*

- ***addition**: $+ : \mathcal{V} \times \mathcal{V} \to \mathcal{V}$, and*
- ***scalar multiplication**: $\cdot : \mathbb{F} \times \mathcal{V} \to \mathcal{V}$*

which satisfy:

(VS 1) *$\mathcal{V}$ is **closed** under addition; that is, if $x, y \in \mathcal{V}$, then $x + y \in \mathcal{V}$.*
(VS 2) *For all $x, y \in \mathcal{V}$, $x + y = y + x$.*
(VS 3) *For all $x, y, z \in \mathcal{V}$, $(x + y) + z = x + (y + z)$.*
(VS 4) *There exists an element $\mathbf{0} \in \mathcal{V}$ such that $x + \mathbf{0} = x$ for all $x \in \mathcal{V}$.*
(VS 5) *For all $x \in \mathcal{V}$ there exists an element $y_x \in \mathcal{V}$ such that $x + y_x = \mathbf{0}$.*

(VS 6) *$\mathcal{V}$ is **closed** under scalar multiplication; that is, if $x \in \mathcal{V}$ and $\kappa \in \mathbb{F}$, then $\kappa x \in \mathcal{V}$.*
(VS 7) *For all $x \in \mathcal{V}$, $\mathbf{1} \cdot x = x$.*
(VS 8) *For all $x \in \mathcal{V}$ and $\alpha, \beta \in \mathbb{F}$, $\alpha(\beta x) = (\alpha\beta)x$.*
(VS 9) *For all $\alpha \in \mathbb{F}$ and $x, y \in \mathcal{V}$, $\alpha(x + y) = \alpha x + \alpha y$.*
(VS 10) *For all $\alpha, \beta \in \mathbb{F}$ and $x \in \mathcal{V}$, $(\alpha + \beta)x = \alpha x + \beta x$.*

*Elements of $\mathcal{V}$ are called **vectors**, while elements of $\mathbb{F}$ are called **scalars**.*

1.9. Example. Let $\mathbb{F}$ be a field and $m, n \in \mathbb{N}$. The set of all $m \times n$ **matrices over** $\mathbb{F}$ forms a vector space over $\mathbb{F}$. That is,

$$\mathbb{M}_{m\times n}(\mathbb{F}) := \{a = [a_{ij}] : a_{ij} \in \mathbb{F}, 1 \le i \le m, 1 \le j \le n\}.$$

Given $a = [a_{ij}], b = [b_{ij}] \in \mathbb{M}_{m\times n}(\mathbb{F})$ and $\kappa \in \mathbb{F}$, we define

$$a + b := [a_{ij} + b_{ij}],$$

and

$$\kappa a := [\kappa a_{ij}].$$

In particular, $\mathbb{F}^n$ is a vector space over $\mathbb{F}$, whether we view elements of $\mathbb{F}^n$ as row vectors or as column vectors.

1.10. Example. Let $m \in \mathbb{N}$. Then $\mathbb{C}^m$ is a vector space over $\mathbb{C}$. It is also a vector space over $\mathbb{R}$, and it is even a vector space over $\mathbb{Q}$.

More generally, if $\mathcal{V}$ is a vector space over a field $\mathbb{F}$, and if $\mathbb{G}$ is a subfield of $\mathbb{F}$, then $\mathcal{V}$ is a vector space over $\mathbb{G}$.

1.11. Example. Let $S \neq \varnothing$ be a set and $\mathbb{F}$ be a field. We define

$$\mathbb{F}^S := \{f : S \to \mathbb{F} : f \text{ a function}\},$$

and

$$\mathbb{F}^{(S)} := \{f \in \mathbb{F}^S : f(x) = 0 \text{ except for finitely many values of } x \in S\}.$$

Both $\mathbb{F}^S$ and $\mathbb{F}^{(S)}$ are vector spaces over $\mathbb{F}$. Note that $\mathbb{F}^S = \mathbb{F}^{(S)}$ if and only if S is finite.

1.12. Example. The set $\mathcal{C}([0,1], \mathbb{K})$ from Example 1.4 is a vector space over $\mathbb{K}$, as is the set $\mathbb{T}_2(\mathbb{K})$ from Example 1.5.

1.13. Example. Let $\mathbb{F}$ be a field. Define the set

$$\mathbb{F}[x] := \{p := p_0 + p_1 x + \cdots + p_m x^m : m \in \mathbb{N}, p_k \in \mathbb{F}, 0 \le k \le m\}.$$

The element $p_j \in \mathbb{F}$ is referred to as the j^{th} **coefficient** of p, and *if* $p \neq 0$, then the **degree** of p is defined to be

$$\deg(p) := \max\{j \in \mathbb{N} : p_j \neq 0\}.$$

We do not define the degree of the zero polynomial $z(x) = 0$**.**

$\mathbb{F}[x]$ is the set of **polynomials with coefficients in** $\mathbb{F}$, and it is a vector space over $\mathbb{F}$.

If $m \in \mathbb{N}$ is fixed, then the set $\mathbb{F}_m[x] := \{p := p_0 + p_1 x + \cdots + p_m x^m : p_k \in \mathbb{F}, 0 \le k \le m\}$ of all polynomials of degree at most m is also a vector space over $\mathbb{F}$.

1.14. Example. The following sets are all vector spaces over $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$.

(a) $\ell_\infty^{\mathbb{N}}(\mathbb{K}) := \{x = (x_n)_n \in \mathbb{K}^{\mathbb{N}} : \sup_n |x_n| < \infty\}$.
(b) $\ell_1^{\mathbb{N}}(\mathbb{K}) := \{x = (x_n)_n \in \mathbb{K}^{\mathbb{N}} : \sum_n |x_n| < \infty\}$.
(c) $c(\mathbb{K}) := \{x = (x_n)_n \in \mathbb{K}^{\mathbb{N}} : \lim_n x_n \text{ exists}\}$.
(d) $c_0(\mathbb{K}) := \{x = (x_n)_n \in \mathbb{K}^{\mathbb{N}} : \lim_n x_n = 0\}$.

1.15. Example. The set $\mathcal{V} := \{\alpha \,\text{PIG} + \beta \,\text{DOG} : \alpha, \beta \in \mathbb{F}\}$ from Example 1.7 is a vector space over $\mathbb{F}$.

1.16. Example. Consider the set $\mathcal{W} := \{f \in \mathbb{R}^{[0,1]} : f(\frac{1}{2}) = 1\}$. Let

$$g(x) = \begin{cases} 0 & x \neq \frac{1}{2} \\ 1 & x = \frac{1}{2} \end{cases} \qquad \text{and} \qquad h(x) = 1,\ x \in [0,1],$$

so that $g, h \in \mathcal{W}$.

Then $(g+h)(\frac{1}{2}) = g(\frac{1}{2}) + h(\frac{1}{2}) = 1 + 1 = 2$, which implies that $g + h \notin \mathcal{W}$. In particular, $\mathcal{W}$ is **not** a vector space over $\mathbb{R}$.

Having defined a vector space, let us now establish a couple of basic facts about them that depend only upon the definition, and not upon the specificities of any given example.

1.17. Proposition. CANCELLATION
Let $\mathcal{V}$ be a vector space over a field $\mathbb{F}$. Let $x, y, z \in \mathcal{V}$ and suppose that

$$x + y = x + z.$$

Then $y = z$.

Proof. Suppose that $x + y = x + z$. Choose $v_x \in \mathcal{V}$ such that $v_x + x = \mathbf{0}$. Then

$$\begin{aligned} y = \mathbf{0} + y = (v_x + x) + y &= v_x + (x + y) \\ &= v_x + (x + z) = (v_x + x) + z = \mathbf{0} + z = z. \end{aligned}$$

□

1.18. Proposition. *Let $\mathcal{V}$ be a vector space over a field $\mathbb{F}$.*

(a) *The element $\mathbf{0}$ from (VS 4) satisfying $x + \mathbf{0} = x$ for all $x \in \mathcal{V}$ is unique.*
(b) *Given $x \in \mathcal{V}$, the element $y_x \in \mathcal{V}$ for which $x + y_x = \mathbf{0}$ is unique. We denote it by $-x$.*
(c) *For any $x \in \mathcal{V}$, $0 \cdot x = \mathbf{0}$.*
(d) *For all $\kappa \in \mathbb{F}$, $x \in \mathcal{V}$, we have*

$$(-\kappa x) = -(\kappa x).$$

(e) *For all $\kappa \in \mathbb{F}$, $\kappa\mathbf{0} = \mathbf{0}$.*
(f) *For all $\kappa \in \mathbb{F}$ and $x \in \mathcal{V}$ we have*

$$-(\kappa x) = \kappa(-x).$$

Proof.

(a) Suppose that $z_1, z_2 \in \mathcal{V}$ and that $z_i + x = x = x + z_i$, $i = 1, 2$ for all $x \in \mathcal{V}$. Then

$$x + z_1 = x = x + z_2.$$

By Proposition 1.17, $z_1 = z_2$. Thus the neutral element under addition is unique, and we denote it by $\mathbf{0}$.

(b) Let $x \in \mathcal{V}$, and suppose that $y_1, y_2 \in \mathcal{V}$ and $x + y_1 = \mathbf{0} = x + y_2$. By Proposition 1.18, $y_1 = y_2$. Since the additive inverse of x is unique, we denote it by $-x$.

(c) Let $x \in \mathcal{V}$. Then

$$0 \cdot x + \mathbf{0} = 0 \cdot x = (0 + 0) \cdot x = 0 \cdot x + 0 \cdot x.$$

By Proposition 1.17, $0 \cdot x = \mathbf{0}$.

(d) Let $\kappa \in \mathbb{F}$ and $x \in \mathcal{V}$. Then

$$\mathbf{0} = 0 \cdot x = (-\kappa + \kappa) \cdot x = (-\kappa) \cdot x + \kappa \cdot x.$$

By uniqueness of additive inverses,

$$-(\kappa x) = (-\kappa)x.$$

(e) Let $\kappa \in \mathbb{F}$. Then

$$\kappa\mathbf{0} + \mathbf{0} = \kappa\mathbf{0} = \kappa(\mathbf{0} + \mathbf{0}) = \kappa\mathbf{0} + \kappa\mathbf{0}.$$

By Proposition 1.17, $\kappa\mathbf{0} = \mathbf{0}$.

(f) Let $\kappa \in \mathbb{F}$ and $x \in \mathcal{V}$. Then

$$\mathbf{0} = \kappa\mathbf{0} = \kappa(x + (-x)) = \kappa x + \kappa(-x).$$

By the uniqueness of additive inverses (i.e. (b)), $\kappa(-x) = -(\kappa x) = (-\kappa)x$.

□

2. Subspaces

2.1. In many areas of mathematics, we are interested in algebraic objects which are subsets of larger algebraic objects and which inherit a similar structure using the operations defined on the larger object. For example, you will have seen in a previous course that $\mathbb{C}$ is a field, and that $\mathbb{R} \subseteq \mathbb{C}$ is a field *using the same operations of addition and multiplication* inherited from $\mathbb{C}$. We say that $\mathbb{R}$ is a **subfield** of $\mathbb{C}$. Note that $\mathbb{Q}$ is a subfield of both $\mathbb{R}$ and of $\mathbb{C}$.

In this course, we are studying vector spaces, and so – unsurprisingly – the substructures we are considering will be called *subspaces*.

2.2. Definition. *Let $\mathcal{V}$ be a vector space over a field $\mathbb{F}$. A subset $\mathcal{W} \subseteq \mathcal{V}$ of $\mathcal{V}$ is called a **subspace** of $\mathcal{V}$ if $\mathcal{W}$ is a vector space over $\mathbb{F}$ with respect to the addition and scalar operation it inherits from $\mathcal{V}$.*

Keeping in mind that vector spaces are non-empty, we see that in order to be a subspace of $\mathcal{V}$, $\mathcal{W}$ must be non-empty. Let $0 \neq w \in \mathcal{W}$. Since $\mathcal{W}$ is a vector space, $\kappa w \in \mathcal{W}$ for all $\kappa \in \mathbb{F}$. In particular, if $\mathcal{W}$ is a subspace of $\mathcal{V}$, then $0 \cdot w = \mathbf{0} \in \mathcal{W}$.

2.3. Example. With $\mathcal{V}$ as above, the sets $\mathcal{W}_1 := \mathcal{V}$ and $\mathcal{W}_2 := \{\mathbf{0}\}$ are always subspaces of $\mathcal{V}$. They are called **trivial subspaces**.

Many properties of vector spaces automatically hold for subsets. In fact, we have the following:

2.4. Theorem. (The subspace test) *Let $\mathcal{V}$ be a vector space over $\mathbb{F}$ and $\mathcal{W} \subseteq \mathcal{V}$. The following are equivalent:*

(a) *$\mathcal{W}$ is a subspace of $\mathcal{V}$.*
(b) *$\mathcal{W} \neq \varnothing$, and for all $\kappa \in \mathbb{F}$, $w_1, w_2 \in \mathcal{W}$ we have $\{\kappa w_1, w_1 + w_2\} \subseteq \mathcal{W}$.*
(c) *$\mathcal{W} \neq \varnothing$, and for all $\kappa \in \mathbb{F}$, $w_1, w_2 \in \mathcal{W}$ we have $\kappa w_1 + w_2 \in \mathcal{W}$.*

Proof.

(a) implies (b). If $\mathcal{W}$ is a subspace of $\mathcal{V}$, then $\mathcal{W}$ is itself a vector space, so by definition, $\mathcal{W} \neq \varnothing$. Also by definition, if $\kappa \in \mathbb{F}$, $w_1, w_2 \in \mathcal{W}$, then κw_1 and $w_1 + w_2 \in \mathcal{W}$.

(b) implies (c). Suppose that (b) holds. Then $\mathcal{W} \neq \varnothing$ by hypothesis. Let $\kappa \in \mathbb{F}$, $w_1, w_2 \in \mathcal{W}$. Then $\kappa w_1 \in \mathcal{W}$ by (b), and so $(\kappa w_1) + w_2 \in \mathcal{W}$, again by (b). Thus (c) holds.

(c) implies (a). Suppose that (c) holds. Let $w_1, w_2, w_3 \in \mathcal{W}$ and $\kappa \in \mathbb{F}$.

- By (c), $1 \cdot w_1 + w_2 = w_1 + w_2 \in \mathcal{W}$. (Note that $1 \cdot w_1 = w_1$ because $w_1 \in \mathcal{W} \subseteq \mathcal{V}$, and $1 \cdot x = x$ for all $x \in \mathcal{V}$.)
- $w_1 + w_2 = w_2 + w_1$ because this holds in $\mathcal{V}$, and $\mathcal{W} \subseteq \mathcal{V}$.
- $(w_1 + w_2) + w_3 = w_1 + (w_2 + w_3)$ because this holds in $\mathcal{V}$, and $\mathcal{W} \subseteq \mathcal{V}$.
- $\mathbf{0} = -1 \cdot w_1 + w_1 \in \mathcal{W}$.
- $-w_1 = (-1) \cdot w_1 + \mathbf{0} \in \mathcal{W}$.
- $\kappa w_1 = \kappa w_1 + \mathbf{0} \in \mathcal{W}$.

- $\mathbf{1} \cdot w_1 = w_1$ because this holds in $\mathcal{V}$, and $\mathcal{W} \subseteq \mathcal{V}$.
- if $\alpha, \beta \in \mathbb{F}$, then $\alpha(\beta w_1) = (\alpha\beta)w_1$ because this holds in $\mathcal{V}$, and $\mathcal{W} \subseteq \mathcal{V}$.
- $\kappa(w_1 + w_2) = \kappa w_1 + \kappa w_2$ because this holds in $\mathcal{V}$, and $\mathcal{W} \subseteq \mathcal{V}$.
- If $\alpha, \beta \in \mathbb{F}$, then $(\alpha + \beta)w_1 = \alpha w_1 + \beta w_1$ because this holds in $\mathcal{V}$, and $\mathcal{W} \subseteq \mathcal{V}$.

By definition, $\mathcal{W}$ is a vector space, and thus $\mathcal{W}$ is a subspace of $\mathcal{V}$.

□

2.5. Exercise. In parts (b) and (c) of the above Theorem, we may replace the condition that $\mathcal{W} \neq \varnothing$ by the condition that $\mathbf{0} \in \mathcal{W}$.

2.6. Example.
Let $m, n \in \mathbb{N}$ and let $\mathbb{F}$ be a field. Recall that given $T = [t_{ij}] \in \mathbb{M}_{m\times n}(\mathbb{F})$, we define the **transpose** of T to be

$$T^{\mathrm{t}} := [t_{ji}] \in \mathbb{M}_{n\times m}(\mathbb{F}).$$

If $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{K}$, we define the **adjoint** of T to be $T^* := [\overline{t_{ji}}]$, where $\overline{z}$ denotes the complex conjugate of $z \in \mathbb{K}$. In the case where $\mathbb{K} = \mathbb{R}$, $T^* = T^{\mathrm{t}}$.

Then $(\mathbb{M}_n(\mathbb{F}))_{sym} := \{T \in \mathbb{M}_n(\mathbb{F}) : T = T^{\mathrm{t}}\}$ is a subspace of $\mathbb{M}_n(\mathbb{F})$.

Observe that the set $(\mathbb{M}_n(\mathbb{C}))_{sa} := \{T \in \mathbb{M}_n(\mathbb{C}) : T = T^*\}$ is a vector space over $\mathbb{R}$, but it is *not* a vector space over $\mathbb{C}$! (Why not?)

We say that $T \in \mathbb{M}_n(\mathbb{C})$ is **self-adjoint** or **hermitian** if $T = T^*$. If $\mathbb{F}$ is a field and $S \in \mathbb{M}_n(\mathbb{F})$ satisfies $S = S^{\mathrm{t}}$, we say that S is **symmetric**.

2.7. Definition. *Let $n \in \mathbb{N}$ and $T = [t_{ij}] \in \mathbb{M}_n(\mathbb{F})$. We define the **trace** of T to be*

$$tr(T) := t_{11} + t_{22} + \cdots + t_{nn} = \sum_{k=1}^{n} t_{kk}.$$

We then define

$$\mathfrak{sl}_n(\mathbb{F}) := \{T \in \mathbb{M}_n(\mathbb{F}) : tr(T) = 0\}.$$

2.8. Example.
Given a field $\mathbb{F}$ and $n \in \mathbb{N}$, $\mathfrak{sl}_n(\mathbb{F})$ is a subspace of $\mathbb{M}_n(\mathbb{F})$.

2.9. Example.
Let $F \subseteq [0, 1]$ be a closed set. Define

$$\Delta_F := \{f \in \mathcal{C}([0, 1], \mathbb{K}) : f(x) = 0 \text{ for all } x \in F\}.$$

Then Δ_F is a subspace of $\mathcal{C}([0, 1], \mathbb{K})$.

2.10. Example.

(a) Consider $\mathcal{V} := \mathbb{R}$, as a vector space over $\mathbb{R}$. Then $\{\mathbf{0}, \mathbb{R}\}$ are the trivial subspaces of $\mathcal{V}$ over $\mathbb{R}$. Are there any others?

Suppose $\{\mathbf{0}\} \neq \mathcal{W}$ is a subspace of $\mathbb{R}$, and let $\mathbf{0} \neq w \in \mathcal{W}$. For all $\kappa \in \mathbb{R}$, we must have $\kappa w \in \mathcal{W}$. But if $\mathbf{0} \neq w \in \mathcal{W} \subseteq \mathbb{R}$, then $\{\kappa w : \kappa \in \mathbb{R}\} = \mathbb{R}$, and so $\mathcal{W} = \mathbb{R}$.

That is, the only subspaces of $\mathbb{R}$ (as a vector space over $\mathbb{R}$) are the trivial subspaces.

(b) Consider $\mathbb{R}^2$ as a vector space over $\mathbb{R}$. We leave it as an exercise for the reader to show that if $\mathcal{W}$ is a subspace of $\mathbb{R}^2$, then $\mathcal{W}$ is one of
- $\{\mathbf{0}\}$ or $\mathbb{R}^2$ – the trivial subspaces; or
- a line passing through the origin.

(c) Similarly, $\mathbb{R}^3$ is a vector space over $\mathbb{R}$, and the subspaces of $\mathbb{R}^3$ are of the form
- $\{\mathbf{0}\}$ or $\mathbb{R}^3$ – the trivial subspaces;
- a line passing through the origin; or
- a plane passing through the origin.

2.11. Theorem. *Let $\mathcal{V}$ be a vector space and $(\mathcal{W}_\lambda)_{\lambda\in\Lambda}$ be a family of subspaces of $\mathcal{V}$. Then*

$$\mathcal{W} := \cap_{\lambda\in\Lambda}\mathcal{W}_\lambda$$

is a subspace of $\mathcal{V}$.

Proof. Since each $\mathcal{W}_\lambda$ is a subspace of $\mathcal{V}$, we see that $\mathbf{0} \in \mathcal{W}_\lambda$ for all $\lambda \in \Lambda$. Thus

$$\mathbf{0} \in \cap_{\lambda\in\Lambda}\mathcal{W}_\lambda \neq \varnothing.$$

Also, if $x, y \in \mathcal{W}$ and $\kappa \in \mathbb{F}$, then for any $\lambda \in \Lambda$, $x, y \in \mathcal{W}_\lambda$. Since $\mathcal{W}_\lambda$ is a subspace of $\mathcal{V}$, it follows that $\kappa x + y \in \mathcal{W}_\lambda$. Since λ was arbitrary,

$$\kappa x + y \in \mathcal{W} = \cap_{\lambda\in\Lambda}\mathcal{W}_\lambda,$$

and so by the Subspace Test, $\mathcal{W}$ is a subspace of $\mathcal{V}$.

□

2.12. Notation. Given a vector space $\mathcal{V}$ over a field $\mathbb{F}$ and two non-empty subsets $A, B \subseteq \mathcal{V}$, we define

$$A + B := \{a + b : a \in A, b \in B\}.$$

If $A = \{a\}$, we usually write $a + B$ instead of $\{a\} + B$. Yes, we really are wild like that.

2.13. Definition. *Let $\mathcal{V}$ be a vector space over a field $\mathbb{F}$ and let $\mathcal{W} \subseteq \mathcal{V}$ be a subspace of $\mathcal{V}$. For $v \in \mathcal{V}$, the set*

$$\{v\} + \mathcal{W} := \{v + w : w \in \mathcal{W}\}$$

*is called the **coset of** $\mathcal{W}$ **containing** v. We normally write $v+\mathcal{W}$ instead of $\{v\}+\mathcal{W}$, and we refer to v as a **representative** of the coset $v + \mathcal{W}$. .*

2.14. Example. Thus, if $v \in \mathbb{R}^2$, then the coset of $\mathcal{W}$ passing containing v is one of the following.

- If $\mathcal{W} = \{\mathbf{0}\}$, then $v + \mathcal{W} = \{\mathbf{0}\} = \{v\}$, just the set containing v.
- If $\mathcal{W} = \mathbb{R}^2$, then $v + \mathbb{R}^2 = \mathbb{R}^2$.
- If $\mathcal{W}$ is a line passing through the origin, then $v + \mathcal{W}$ is a line parallel to $\mathcal{W}$ but passing through v.

2.15. Example. Thus, if $v \in \mathbb{R}^3$ and $\mathcal{W}$ is a subspace of $\mathbb{R}^3$, then the coset of $\mathcal{W}$ passing containing v is one of the following.

- If $\mathcal{W} = \{\mathbf{0}\}$, then $v + \{\mathbf{0}\} = \{v\}$; just the set containing v, and if $\mathcal{W} = \mathbb{R}^3$, then $v + \mathbb{R}^3 = \mathbb{R}^3$.
- If $\mathcal{W}$ is a line in $\mathbb{R}^3$ passing through the origin, then $v + \mathcal{W}$ is a line parallel to the line $\mathcal{W}$ but passing through v.
- If $\mathcal{W}$ is a plane in $\mathbb{R}^3$ which contains the origin, then $v + \mathcal{W}$ is a plane parallel to the plane $\mathcal{W}$ but passing through v.

2.16. Exercise. We note that in general, the representative of a coset is *not unique*. If $v \in \mathcal{V}$ and $\mathcal{W}$ is a subspace of $\mathcal{V}$, then for any $w \in \mathcal{W}$, $v + w$ is another representative of $v + \mathcal{W}$. As we shall soon see, every representative of $v + \mathcal{W}$ is of this form for some $w \in \mathcal{W}$.

With this in mind – when might the representative of $v + \mathcal{W}$ be unique?

2.17. Proposition. *Let $\mathcal{V}$ be a vector space over a field $\mathbb{F}$ and let $\mathcal{W}$ be a subspace of $\mathcal{V}$. Then:*

(a) *$v + \mathcal{W}$ is a subspace of $\mathcal{V}$ if and only if $v \in \mathcal{W}$.*
(b) *$x + \mathcal{W} = y + \mathcal{W}$ if and only if $x - y \in \mathcal{W}$.*

Proof.

(a) Suppose first that $v + \mathcal{W}$ is a subspace of $\mathcal{V}$. Then $\mathbf{0} \in v + \mathcal{W}$, and therefore $\mathbf{0} = v + w$ for some $w \in \mathcal{W}$. But additive inverses are unique in $\mathcal{V}$, and thus $-v \in \mathcal{W}$. Since $\mathcal{W}$ is a subspace of $\mathcal{V}$, $v = (-1)(-v) \in \mathcal{W}$.

Next, suppose that $v \in \mathcal{W}$. Since $\mathcal{W}$ is a subspace of $\mathcal{V}$, $v + \mathcal{W} = \{v + w : w \in \mathcal{W}\} \subseteq \mathcal{W}$. If $w_0 \in \mathcal{W}$, then $w_0 - v \in \mathcal{W}$ as $\mathcal{W}$ is a subspace. But then $w_0 = v + (w_0 - v) \in v + \mathcal{W}$. This shows that $\mathcal{W} \subseteq v + \mathcal{W}$, which in turn implies that $\mathcal{W} = v + \mathcal{W}$. in particular, $v + \mathcal{W} = \mathcal{W}$ is a subspace of $\mathcal{V}$.

(b) If $x + \mathcal{W} = y + \mathcal{W}$, then $x = x + 0 \in y + \mathcal{W}$ and so $x = y + w$ for some $w \in \mathcal{W}$, i.e., $x - y = w \in \mathcal{W}$.

Conversely, suppose that $x - y \in \mathcal{W}$. Then $(x - y) + \mathcal{W} = \mathcal{W}$, by part (a). Hence

$$\begin{aligned} y + \mathcal{W} &= y + ((x - y) + \mathcal{W}) \\ &= \{y + (x - y) + w : w \in \mathcal{W}\} \\ &= \{x + w : w \in \mathcal{W}\} \\ &= x + \mathcal{W}. \end{aligned}$$

□

2.18. Proposition. *Let $\mathcal{V}$ be a vector space over a field $\mathbb{F}$, and let $\mathcal{W}$ be a subspace of $\mathcal{V}$. Let*

$$\mathcal{Q} := \{x + \mathcal{W} : x \in \mathcal{V}\}$$

denote the collection of all cosets of $\mathcal{W}$ in $\mathcal{V}$. We define two operations on $\mathcal{Q}$ as follows. Given $x + \mathcal{W}$ and $y + \mathcal{W} \in \mathcal{Q}$ and $\kappa \in \mathbb{F}$, we set

$$(x + \mathcal{W}) + (y + \mathcal{W}) := (x + y) + \mathcal{W}, \quad \text{and}$$
$$\kappa(x + \mathcal{W}) := (\kappa x) + \mathcal{W}.$$

Then

(a) *These operations are well-defined. That is, if $x_1 + \mathcal{W} = x_2 + \mathcal{W}$ and $y_1 + \mathcal{W} = y_2 + \mathcal{W}$, then $(x_1 + y_1) + \mathcal{W} = (x_2 + y_2) + \mathcal{W}$, and $(\kappa x_1) + \mathcal{W} = (\kappa x_2) + \mathcal{W}$.*

(b) *Furthermore, $\mathcal{Q}$ is a vector space with these operations.*

The space $\mathcal{Q}$ is called the **quotient space** of $\mathcal{V}$ **modulo** $\mathcal{W}$, and is usually denoted by $\mathcal{V}/\mathcal{W}$.

Proof.

(a) Suppose that $x_1 + \mathcal{W} = x_2 + \mathcal{W}$ and $y_1 + \mathcal{W} = y_2 + \mathcal{W}$. Then $(x_1 - x_2), (y_1 - y_2) \in \mathcal{W}$. Since $\mathcal{W}$ is a subspace of $\mathcal{V}$,

$$(x_1 + y_1) - (x_2 + y_2) = x_1 - x_2 + y_1 - y_2 \in \mathcal{W},$$

and so by Proposition 2.17,

$$(x_1 + y_1) + \mathcal{W} = (x_2 + y_2) \in \mathcal{W}.$$

Similarly, if $x_1 + \mathcal{W} = x_2 + \mathcal{W}$, then $(x_1 - x_2) \in \mathcal{W}$. Since $\mathcal{W}$ is a subspace of $\mathcal{V}$,

$$\kappa x_1 - \kappa x_2 = \kappa(x_1 - x_2) \in \mathcal{W},$$

and so by Proposition 2.17,

$$(\kappa x_1) + \mathcal{W} = (\kappa x_2) + \mathcal{W}.$$

(b) We shall prove half of this result, and leave the remaining half as an exercise.

- If $x+\mathcal{W}$ and $y+\mathcal{W} \in \mathcal{Q}$, then $(x+\mathcal{W})+(y+\mathcal{W}) = (x+y)+\mathcal{W} in \mathcal{Q}$, so $\mathcal{Q}$ is closed under addition.
- For $x+\mathcal{W}, y+\mathcal{W}, z+\mathcal{W} \in \mathcal{Q}$,

$$\begin{aligned}((x+\mathcal{W})+(y+\mathcal{W}))+(z+\mathcal{W}) &= ((x+y)+\mathcal{W})+(z+\mathcal{W}) \\ &= ((x+y)+z)+\mathcal{W} \\ &= (x+(y+z))+\mathcal{W} \\ &= (x+\mathcal{W})+((y+z)+\mathcal{W}) \\ &= (x+\mathcal{W})+((y+\mathcal{W})+(z+\mathcal{W})).\end{aligned}$$

  Thus addition is associative in $\mathcal{Q}$.
- For all $x+\mathcal{W}, y+\mathcal{W} \in \mathcal{Q}$,

$$(x+\mathcal{W})+(y+\mathcal{W}) = (x+y)+\mathcal{W} = (y+x)+\mathcal{W} = (y+\mathcal{W})+(x+\mathcal{W}).$$

- For all $x+\mathcal{W} \in \mathcal{Q}$,

$$(\mathbf{0}+\mathcal{W})+(x+\mathcal{W}) = (\mathbf{0}+x)+\mathcal{W} = x+\mathcal{W} = (x+\mathcal{W})+(\mathbf{0}+\mathcal{W}).$$

  Thus $\mathbf{0}+\mathcal{W}$ is the additive neutral element of $\mathcal{Q}$.
- If $x+\mathcal{W} \in \mathcal{Q}$, then $(x+\mathcal{W})+((-x)+\mathcal{W}) = (x+(-x))+\mathcal{W} = \mathbf{0}+\mathcal{W}$, so $(-x)+\mathcal{W} = -(x+\mathcal{W})$.

The remaining parts are similar.

□

2.19. Example. Let $\mathcal{V} = \mathbb{R}^3$, viewed as a vector space over $\mathbb{R}$, and let $\mathcal{W} = \{(x,y,z) \in \mathbb{R}^3 : x+y+z = 0\}$. Clearly $\mathcal{W}$ is a subspace of $\mathcal{V}$.

Then

$$\mathcal{V}/\mathcal{W} = \{(x_0,y_0,z_0)+\mathcal{W} : (x_0,y_0,z_0) \in \mathbb{R}^3\}.$$

If $(x_0,y_0,z_0) \notin \mathcal{W}$, (i.e. if $(x_0,y_0,z_0)+\mathcal{W} \neq (0,0,0)+\mathcal{W}$), then for any $(x_1,y_1,z_1) \in \mathbb{R}^3$¡ we have that

$$(x_1,y_1,z_1)+\mathcal{W} = \alpha((x_0,y_0,z_0)+\mathcal{W}),$$

where $\alpha := \dfrac{x_1+y_1+z_1}{x_0+y_0+z_0}$. (Note that the denominator is not zero!).

This shows that $\mathcal{V}/\mathcal{W}$ "looks one-dimensional" - it just consists of multiples of one element, namely of $(x_0,y_0,z_0)+\mathcal{W}$.

Supplementary Examples

S2.1. Example. The set $\mathcal{W} := \{f \in \mathbb{K}^{[0,1]} : f(\frac{1}{2}) \neq 0\}$ is **not** a vector space over $\mathbb{K}$.

S2.2. Example. The set $\mathcal{W} := \{f \in \mathbb{K}^{[0,1]} : f(\frac{1}{3}) = 0 = f(\frac{2}{3})\}$ is a vector space over $\mathbb{K}$.

S2.3. Example. The set $\mathcal{W} := \{T = [t_{ij}] \in \mathbb{M}_n(\mathbb{R}) : t_{ij} \geq 0 \text{ for all } 1 \leq i, j \leq n\}$ is **not** a vector space over $\mathbb{R}$.

S2.4. Example. Whereas $\mathbb{C}$ is a vector space over $\mathbb{Q}$, it is **not** the case that $\mathbb{Q}$ is a vector space over $\mathbb{C}$.

S2.5. Example. The set $\mathcal{D} := \{f : (0,1) \to \mathbb{R} : f \text{ is differentiable on } (0,1)\}$ is a vector space over $\mathbb{R}$. This is proven in Math 147.

S2.6. Example. The set

$$\mathcal{R}([0,1],\mathbb{R}) := \{f : [0,1] \to \mathbb{R} : f \text{ is Riemann integrable on } [0,1]\}$$

is a vector space over $\mathbb{R}$. This is proven in Math 148.

S2.7. Example. The set $\mathcal{P} := \{(x, x^2) : x \in \mathbb{R}\}$ is *not* a subspace of $\mathbb{R}^2$.

S2.8. Example. Let $m, n \in \mathbb{N}$ and suppose that $\mathbb{F}$ is a field. Let $\mathcal{V}$ be a vector space over $\mathbb{F}$. Then

$$\mathbb{M}_{m\times n}(\mathcal{V}) := \{T = [t_{ij}] : t_{ij} \in \mathcal{V}, 1 \leq i \leq m, 1 \leq j \leq n\}$$

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

In particular, $\mathcal{V} := \mathbb{M}_{23\times 7}(\mathcal{C}([0,1],\mathbb{C}))$ is a vector space over $\mathbb{C}$. A typical element of $\mathcal{V}$ is a 23×7 matrix $[f_{ij}]$, where each f_{ij} is a continuous function from $[0,1]$ into $\mathbb{C}$.

S2.9. Example. *
Let $\mathbb{F}$ be a field and $\mathcal{V}$, $\mathcal{W}$ be vector spaces over $\mathbb{F}$. Then

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

is a vector space over $\mathbb{F}$, where we set $(v_1, w_1) + (v_2, w_2) := (v_1 + v_2, w_1 + w_2)$ and $\kappa(v, w) := (\kappa v, \kappa w)$ for all $(v_1, w_1), (v_2, w_2), (v, w) \in \mathcal{V} \oplus \mathcal{W}$ and $\kappa \in \mathbb{F}$. This is often called the **external direct sum** of $\mathcal{V}$ and $\mathcal{W}$. Of course, we can extend this definition using $n \in \mathbb{N}$ spaces instead of just two. In fact, we can extend it to an infinite direct sum. We'll return to this later.

S2.10. Example. *

Let $\mathbb{F}$ be a field, and suppose that $\mathcal{Y}$ is a vector space over $\mathbb{F}$. Suppose that $\mathcal{V}$ and $\mathcal{W}$ are two subspaces of $\mathcal{Y}$ satisfying

(i) $\mathcal{V} \cap \mathcal{W} = \{0\}$, and
(ii) $\mathcal{Y} = \mathcal{V} + \mathcal{W} := \{v + w : v \in \mathcal{V}, w \in \mathcal{W}\}$.

Then we say that $\mathcal{Y}$ is the **internal direct sum** of $\mathcal{V}$ and $\mathcal{W}$, and we write $\mathcal{Y} = \mathcal{V} \dot{+} \mathcal{W}$.

We leave it as an exercise for the reader to show that in this case, there exists a bijective map $T : \mathcal{Y} \to \mathcal{V} \oplus \mathcal{W}$ which satisfies:

$$T(\kappa y_1 + y_2) = \kappa T y_1 + T y_2 \text{ for all } \kappa \in \mathbb{F},\ y_1, y_2 \in \mathcal{Y}.$$

This last condition is the assertion that the map T is "*linear*", and when T is bijective, we refer to it as a **linear isomorphism** between $\mathcal{Y}$ and $\mathcal{V} \oplus \mathcal{W}$. We shall have much more to say about these later in the course.

S2.11. Example. *

Consider a **homogeneous system of linear equations** over $\mathbb{F}$. That is, suppose that $m, n \in \mathbb{N}$ and that $a_{ij} \in \mathbb{F}$ for all $1 \le i \le m, 1 \le j \le n$. Consider the system of equations:

$$\begin{array}{rcrcccrcl} a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n & = & 0 \\ a_{21}x_1 & + & a_{22}x_2 & + & \cdots & + & a_{2n}x_n & = & 0 \\ \vdots & & & & & & \vdots & = & 0 \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \cdots & + & a_{mn}x_n & = & 0 \end{array}$$

The **solution set** of this system, namely the set $\mathcal{S}$ of all n-tuples $x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{F}^n$ which satisfy this set of equations is a vector space over $\mathbb{F}$. (Why is $\mathcal{S} \neq \varnothing$?)

S2.12. Example.

Let $n \in \mathbb{N}$ and let $\mathbb{F}$ be a field. Let $T = [t_{ij}] \in \mathbb{M}_n(\mathbb{F})$ and suppose that $\sum_{i=1}^n t_{ii} \neq 0$. Then

$$\mathbb{M}_n(\mathbb{F})/\mathfrak{sl}_n(\mathbb{F}) = \{\kappa T + \mathfrak{sl}_n(\mathbb{F}) : \kappa \in \mathbb{F}\}.$$

Appendix

A2.1. It was Descartes who first determined that points in our usual three-dimensional Euclidean space could be described by ordered triples, and similarly points in two-dimensional space could be described by order pairs. The current definition of a vector space appears to be due to Giuseppe Peano. He'll be the one you want to blame if the going gets tough, not me.

Incidentally, the emergence of vector spaces of functions is due to a number of people, including Lebesgue, Banach and Hilbert.

A2.2. In Definition 2.2.2, we required that a subspace $\mathcal{W}$ of a vector space $\mathcal{V}$ over a field $\mathbb{F}$ must carry *the addition and scalar multiplication operations it inherits from* $\mathcal{V}$. To be very precise, the addition operator $+$ is $\mathcal{V}$ is really the map $+ : \mathcal{V} \times \mathcal{V} \to \mathcal{V}$ that sends an ordered pair (x, y) to $x + y$. When talking about a subspace $\mathcal{W}$, we really mean $+|_{\mathcal{W}} : \mathcal{W} \times \mathcal{W} \to \mathcal{W}$, which means that we are restricting both the domain of $+$ to $\mathcal{W} \times \mathcal{W}$, as well as the codomain of $+$ to $\mathcal{W}$. For $\mathcal{W}$ to be a subspace, we must know that $\mathcal{W}$ is closed under the addition operation in $\mathcal{V}$. This is a consequence of the Subspace Test. Similarly, the scalar multiplication of $\mathcal{W}$ is really the scalar multiplication of $\mathcal{V}$ restricted to the domain $\mathbb{F} \times \mathcal{W}$, with codomain $\mathcal{W}$. Again, for $\mathcal{W}$ to be a subspace, it is necessary that $\mathcal{W}$ be closed under the scalar multiplication operation on $\mathcal{V}$.

Exercises for Chapter 2

Exercise 2.1.
Let $\mathcal{V}$ be a vector space over a field $\mathbb{F}$ and let $\mathcal{W}$ and $\mathcal{Y}$ be subspaces of $\mathcal{V}$. Give necessary and sufficient conditions for $\mathcal{W} \cup \mathcal{Y}$ to be a subspace of $\mathcal{V}$.

Exercise 2.2.
Note that $\mathbb{Q}$ is a subfield of $\mathbb{R}$. If $\mathcal{V}$ is a vector space over $\mathbb{R}$, prove that $\mathcal{V}$ is a vector space over $\mathbb{Q}$. Is the converse true?

Exercise 2.3.*
Let $\Lambda \neq \varnothing$ be a set, and for each $\lambda \in \Lambda$, let $\mathcal{V}_\lambda$ be a vector space. Let $\mathcal{V} := \cup_{\lambda \in \Lambda} \mathcal{V}_\lambda$, and define

$$\prod_{\lambda \in \Lambda} \mathcal{V}_\lambda := \{f : \Lambda \to \mathcal{V} : f(\lambda) \in \mathcal{V}_\lambda \text{ for all } \lambda \in \Lambda\}.$$

We refer to this as the **direct product** of the spaces $\mathcal{V}_\lambda$.

We normally write

$$(x_\lambda)_{\lambda \in \Lambda}$$

to denote the function $f : \Lambda \in \mathcal{V}$ for which $x_\lambda := f(\lambda)$ for all λ. In fact, when Λ is understood, we even abbreviate this to $(x_\lambda)_\lambda$.

Given $(x_\lambda)_\lambda$ and $(y_\lambda)_\lambda \in \prod_\lambda \mathcal{V}_\lambda$, and given $\kappa \in \mathbb{F}$, define

$$(x_\lambda)_\lambda + (y_\lambda)_\lambda := (x_\lambda + y_\lambda)_\lambda$$
$$\kappa(x_\lambda)_\lambda := (\kappa x_\lambda)_\lambda.$$

Prove that $\prod_\lambda \mathcal{V}_\lambda$ is a vector space over $\mathbb{F}$ with these operations.

Note: this is not the first time you have seen this sort of thing. A **sequence** $(x_n)_n$ of real numbers is really just a function $f : \mathbb{N} \to \mathbb{R}$, where $x_n := f(n)$ for all $n \geq 1$. All we have done is to change the index set from the natural numbers to an arbitrary index set, which amounts to changing the domain of the function from $\mathbb{N}$ to Λ.

If Λ is finite, this becomes a more familiar notion. By relabelling, we assume that $\Lambda = \{1, 2, \ldots, n\}$ for the appropriate $n \in \mathbb{N}$ (namely n is the cardinality of Λ, i.e. the number of elements in Λ), and then

$$\prod_{\lambda \in \Lambda} \mathcal{V}_\lambda = \mathcal{V}_1 \times \mathcal{V}_2 \times \cdots \times \mathcal{V}_n = \{(x_1, x_2, \ldots, x_n) : x_j \in \mathcal{V}_j, 1 \leq j \leq n\}.$$

Finally, if $\mathcal{V}_i = \mathcal{V}_j, 1 \leq i, j \leq n$, then

$$\prod_{\lambda \in \Lambda} \mathcal{V}_\lambda = \mathcal{V}_1^n,$$

the set of n-tuples with entries in $\mathcal{V}_1$. N

When Λ is infinite, the notion of **direct sum** also exists, and is defined as

$$\oplus_{\lambda \in \Lambda} \mathcal{V}_\lambda := \{(x_\lambda)_{\lambda \in \Lambda} : x_\lambda = 0 \text{ for all but finitely many} \lambda \in \Lambda\}.$$

The Subspace Test may be applied to the direct sum to prove that it is a subspace of the direct product. (When Λ is finite, the direct sum and the direct product coincide.)

Exercise 2.4.
Is $\mathbb{Q}$ a subspace of $\mathbb{R}$? (Think about this one!)

Exercise 2.5.
Recall the vector space $\mathcal{V} := \{\alpha\,\text{PIG} + \beta\,\text{DOG} : \alpha, \beta \in \mathbb{K}\}$ from Example 1.7. Let $\mathcal{W} := \{\alpha\,\text{PIG} + 7\alpha\,\text{DOG} : \alpha \in \mathbb{K}\}$. Determine whether or not $\mathcal{W}$ is a subspace of $\mathcal{V}$.

Exercise 2.6.
Let $\mathbb{F}$ be a field, $n \in \mathbb{N}$, and define

$$\mathbb{T}_n(\mathbb{F}) := \{T = [t_{ij}] \in \mathbb{M}_n(\mathbb{F}) : t_{ij} = 0 \text{ if } 1 \le j < i \le n\}.$$

Prove that $\mathbb{T}_n(\mathbb{F})$ is a subspace of $\mathbb{M}_n(\mathbb{F})$.

Exercise 2.7.
Let $\mathcal{V}$ be a vector space over a field $\mathbb{F}$ and let $\{x_\lambda : \lambda \in \Lambda\}$ be a non-empty collection of vectors in $\mathcal{V}$. Define

$$\mathcal{W} = \{\sum_{j=1}^{n} \kappa_j x_{\lambda_j} : n \in \mathbb{N}, \kappa_j \in \mathbb{F}, \lambda_j \in \Lambda, 1 \le j \le n\}.$$

Prove or disprove that $\mathcal{W}$ is a subspace of $\mathcal{V}$.

Exercise 2.8.*
Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces over the same field $\mathbb{F}$. A map $T : \mathcal{V} \to \mathcal{W}$ is said to be **linear** if

$$T(\kappa x + y) = \kappa T x + T y \quad \text{for all } \kappa \in \mathbb{F}, x, y \in \mathcal{V}.$$

Let $\mathcal{L}(\mathcal{V}, \mathcal{W}) := \{T : \mathcal{V} \to \mathcal{W} : T \text{ is linear}\}$. Prove that $\mathcal{L}(\mathcal{V}, \mathcal{W})$ is a vector space over $\mathbb{F}$.

Exercise 2.9.
Consider $\mathcal{V} = \mathbb{R}^n$ as a vector space over $\mathbb{R}$, and for $1 \le k \le n$, define $e_k := (0, 0, \ldots, 0, 1, 0, \ldots 0)$, where the unique "1" appears in the k^{th} position. Let $\mathcal{W}$ be a vector space over $\mathbb{R}$, and let $w_1, w_2, \ldots, w_n \in \mathcal{W}$.

Prove that there exists exactly one linear map $T : \mathcal{V} \to \mathcal{W}$ such that $Te_k = w_k$, $1 \le k \le n$.

Exercise 2.10.*
Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces over the same field, and let $T \in \mathcal{L}(\mathcal{V}, \mathcal{W})$ be a fixed linear map. The **kernel** of T is the set $\ker T := \{x \in \mathcal{V} : Tx = 0\}$, while the **range** of T is the set $\operatorname{ran} T := \{Tx : x \in \mathcal{V}\}$.

Prove that $\ker T$ is a subspace of $\mathcal{V}$ and that $\operatorname{ran} T$ is a subspace of $\mathcal{W}$.

Exercise 2.11.

Let $\mathbb{F}$ be a field. Recall from Exercise 1.6 above that $\mathbb{T}_n(\mathbb{F})$ is a vector space over $\mathbb{F}$. Let $\mathcal{W}_n := \{T \in \mathbb{T}_n(\mathbb{F}) : \operatorname{tr}(T) = 0\}$.

(a) Prove that $\mathcal{W}_n$ is a subspace of $\mathbb{T}_n(\mathbb{F})$.

(b) Find a matrix $A \in \mathbb{T}_2(\mathbb{F})$ such that

$$\mathbb{T}_2(\mathbb{F})/\mathcal{W}_2 = \{\alpha(A + \mathcal{W}) : \alpha \in \mathbb{F}\}.$$

(c) Let $\mathcal{Y} := \{W = \begin{bmatrix} w_{11} & 0 \\ 0 & w_{22} \end{bmatrix} : w_{11}, w_{22} \in \mathbb{F}, w_{11} + w_{22} = 0\}$. Find two matrices $A, B \in \mathbb{T}_2(\mathbb{F})$ such that

$$\mathbb{T}_2(\mathbb{F})/\mathcal{Y} := \{\alpha(A + \mathcal{Y}) + \beta(B + \mathcal{Y}) : \alpha, \beta \in \mathbb{F}\}.$$

Can you extend your construction to the case where $n \geq 3$?

CHAPTER 3

Linear spans and linear independence

I can speak Esperanto like a native.

Spike Milligan

1. Linear spans

1.1. The two most important concepts in vector space theory are those of **the linear span** and of **linear independence (or the dual notion, linear dependence)** of a given set $\mathcal{S}$ of vectors in a vector space $\mathcal{V}$ over a field $\mathbb{F}$.

We begin by adopting the practice of simply writing "*a vector space*" instead of "*a vector space over a field* $\mathbb{F}$" when there is only one field being considered. When multiple fields come into play at once and we wish to emphasise this – we might be looking at the "*dimension*" of $\mathbb{C}$ as a vector space over $\mathbb{C}$, over $\mathbb{R}$, or even over $\mathbb{Q}$ – we will have to explicitly mention which field we are dealing with at any given moment.

1.2. With the above convention, we let $\mathcal{V}$ be a vector space and $\mathcal{S} \subseteq \mathcal{V}$ be a non-empty set of vectors. The set $\mathcal{S}$ may be finite or infinite. To understand linear spans, I would like to make an analogy which will hopefully be of use to you.

Suppose that $\mathcal{V} = \mathbb{R}^2$ as a vector space over the field $\mathbb{Q}$, and that $\varnothing \neq \mathcal{S} \subseteq \mathcal{V}$. Imagine that you have a robot stationed at the origin $(0,0) \in \mathbb{R}^2$ that can move around in $\mathbb{R}^2$. The issue is that it is only programmed to take very specific (multiples of, or *scaled*) "*steps*". The steps in question are precisely those determined by elements of $\mathcal{S}$; and the "*multiple*" in question means that we can multiply that step in $\mathcal{S}$ by an element $\kappa \in \mathbb{F}$ - in this case $\mathbb{F} = \mathbb{Q}$. If $0 \neq s \in \mathbb{S}$ and $\kappa \in \mathbb{Q}$ is negative, then the step κs means that the step is taken in the exact opposite direction of that of s, and the length of the step is $|\kappa|$ times the length of the step s. No other kinds of steps are permitted. We shall give a precise definition of *linear combinations* in just a moment, but for the time being, let's agree that a linear combination of elements of $\mathcal{S}$ means that

- we allow the robot to concatenate these *scaled steps* a **finite number of times** – each time taking a "*scaled step*" as above. So - we are adding a finite number of "*extended permissible steps*" together.

The notion of "*the linear span of* $\mathcal{S}$" is the answer to the question: starting at the origin $(0,0)$, to which points in $\mathbb{R}^2$ can our robot walk?

This is the intuition, and not a mean ice-breaker at the kind of parties to which we are likely to be invited. Let us now formalise it, that is, let us make it precise so that we can do mathematics. Both at those parties and elsewhere.

1.3. Definition. *Let $\mathcal{V}$ be a vector space over a field $\mathbb{F}$ and $\varnothing \neq \mathcal{S} \subseteq \mathcal{V}$. A vector $x \in \mathcal{V}$ is said to be a* ***linear combination of vectors in*** $\mathcal{S}$ *if there exists a finite number of vectors $u_1, u_2, \ldots, u_n \in \mathcal{S}$ and scalars $\kappa_1, \kappa_2, \ldots, \kappa_n \in \mathbb{F}$ such that*

$$x = \kappa_1 u_1 + \kappa_2 u_2 + \cdots + \kappa_n u_n.$$

We refer to the κ_j's as the ***coefficients*** *of x relative to $\{u_1, u_2, \ldots, u_n\}$.*

We write span $\mathcal{S}$ to denote the set of all possible linear combinations of elements of $\mathcal{S}$, and we call this the ***linear span*** *of $\mathcal{S}$.*

Note: *By convention, we define span $\varnothing := \{0\}$.*

At this point, it might – no, make that it *would* – be a good idea for the reader to look at Exercise 2.7 to see if anything looks familiar.

1.4. Remark. With this level of generality, the coefficients of a vector $x \in \mathcal{V}$ relative to $\{u_1, u_2, \ldots, u_n\}$ are not necessarily uniquely determined. For example, if $\mathcal{V} = \mathbb{R}^2$ as a vector space over $\mathbb{R}$, and $u_1 = u_2 \in \mathcal{V}$, then $2u_1 + 5u_2 = 7u_2 = 3u_1 + 4u_1$. In fact, $2u_1 + 5u_2 = \kappa_1 u_1 + \kappa_2 u_2$ whenever $\kappa_1 + \kappa_2 = 7$. In this case, the coefficients are non-unique because of (essentially) trivial reasons, and we can get around this particular "non-uniqueness issue" by requiring all of the u_j's to be distinct. Another trivial reason why the linear combination is not "unique" is that if $x = \kappa_1 u_1 + \kappa_2 u_2$ and if $u_3 \notin \{u_1, u_2\}$, then $x = \kappa_1 u_1 + \kappa_2 u_2 + 0u_3$. Again, we can get around this "non-uniqueness issue" by asking that all of the coefficients of the vectors u_j should be non-zero.

This still does not solve the problem of "non-uniqueness", however. With $\mathcal{V} = \mathbb{R}^2$ as above, if we let $\mathcal{S} = \{u_1 = (1,0), u_2 = (0,1), u_3 = (1,1)\}$, and if we set $x = (2,1)$, then $x = 2u_1 + 1u_2 = u_1 + u_3$. This lack of uniqueness is in no way trivial, and we shall examine it more closely below. In fact, this last example of "non-uniqueness" represents the notion of "linear dependence", which is one of the most crucial concepts we shall deal with in this course.

1.5. Example.

(a) Let $\mathcal{V} = \mathbb{R}^3$, $\mathbb{F} = \mathbb{R}$ and $\mathcal{S} = \{(1,0,0), (0,1,0), (0,0,1)\}$. Then

$$\operatorname{span}_{\mathbb{R}} \mathcal{S} = \mathbb{R}^3.$$

Indeed, if $p := (x_0, y_0, z_0) \in \mathbb{R}^3$, then

$$p = x_0(1,0,0) + y_0(0,1,0) + z_0(0,0,1).$$

(b) If $\mathcal{V} = \mathbb{R}^3$, $\mathbb{F} = \mathbb{Q}$ and $\mathcal{S} = \{(1,0,0),(0,1,0),(0,0,1)\}$. Then

$$\text{span}_{\mathbb{Q}}\mathcal{S} = \mathbb{Q}^3 := \{(q_1,q_2,q_3) : q_j \in \mathbb{Q}, 1 \le j \le 3\}.$$

Indeed, if $p := (q_1,q_2,q_3) \in \mathbb{Q}^3$, then

$$p = q_1(1,0,0) + q_2(0,1,0) + q_3(0,0,1).$$

Thus $\text{span}_{\mathbb{Q}}\mathcal{S} \supseteq \mathbb{Q}^3$. Conversely, if $\alpha,\beta,\gamma \in \mathbb{Q}$, then

$$\alpha(1,0,0) + \beta(0,1,0) + \gamma(0,0,1) = (\alpha,\beta,\gamma) \in \mathbb{Q}^3,$$

so that $\text{span}_{\mathbb{Q}}\mathcal{S} \subseteq \mathbb{Q}^3$. Hence equality holds.

(c) Let $\mathcal{V} = \mathbb{C}^3$, $\mathbb{F} = \mathbb{C}$ and $\mathcal{S} = \{(q,r,0) : q,r \in \mathbb{Q}\}$. We claim that

$$\text{span}_{\mathbb{C}}\,\mathcal{S} = \Omega := \{(w,z,0) : w,z \in \mathbb{C}\}.$$

This time $\mathcal{S}$ contains infinitely many vectors. However, an arbitrary linear combination of elements of $\mathcal{S}$ is of the form

$$p := \kappa_1(q_1,r_1,0) + \kappa_2(q_2,r_2,0) + \cdots + \kappa_n(q_n,r_n,0),$$

where $\kappa_j \in \mathbb{C}$ and $q_j, r_j \in \mathbb{Q}$, $1 \le j \le n$, so that

$$p = (\sum_{j=1}^{n} \kappa_j q_j, \sum_{j=1}^{n} \kappa_j r_j, 0).$$

In particular, the third coordinate is zero, so $\text{span}_{\mathbb{C}}\,\mathcal{S} \subseteq \Omega$.

Conversely, given $(w,z,0) \in \Omega$, we have

$$(w,z,0) = w(1,0,0) + z(0,1,0) \in \text{span}_{\mathbb{C}}\,\mathcal{S},$$

so $\text{span}_{\mathbb{C}}\,\mathcal{S} = \Omega$.

(d) Consider the vector space $\mathcal{C}([0,1],\mathbb{R})$ over $\mathbb{R}$ and the set $\mathcal{S} = \{\mathbf{1}, x, x^2, x^3, \ldots\} \subseteq \mathcal{C}([0,1],\mathbb{R})$. In Math 147, one learns that

$$\sin x = \sum_{n=0}^{\infty} \frac{(-1)^{n+1}}{n!} x^{2n+1}.$$

Nevertheless, $\sin x \notin \text{span}_{\mathbb{R}}\,\mathcal{S}$.

The key is that a linear combination can only involve *finitely many* elements of $\mathcal{S}$ at a time. So how do we know that there isn't *some other* finite linear combination of vectors in $\mathcal{S}$, say $p(x) := p_0 + p_1 x + p_2 x^2 + \cdots + p_m x^m$, satisfying $p(x) := \sin x$? Hint: consider derivatives!

Given a vector space $\mathcal{V}$, let $\text{SUBSP}(\mathcal{V}) := \{\mathcal{W} : \mathcal{W} \text{ is a subspace of } \mathcal{V}\}$. As seen in Chapter One, we may partially order $\text{SUBSP}(\mathcal{V})$ by inclusion, so that for subspaces $\mathcal{Y}$ and $\mathcal{Z}$ of $\mathcal{V}$, $\mathcal{Y} \le \mathcal{Z}$ if $\mathcal{Y} \subseteq \mathcal{Z}$. If $\mathcal{S} \subseteq \mathcal{V}$, then the **smallest** subspace that contains $\mathcal{S}$ is taken to mean the **minimum** element $\mathcal{M}$ of the partially ordered set $\text{SUBSP}_{\mathcal{S}}(\mathcal{V}) = \{\mathcal{W} \in \text{SUBSP}(\mathcal{V}) : \mathcal{S} \subseteq \mathcal{W}\}$. We leave it as an exercise for the reader to show that

$$\mathcal{M} = \cap\{\mathcal{W} \in \text{SUBSP}(\mathcal{V}) : \mathcal{S} \subseteq \mathcal{W}\}.$$

1.6. Theorem. *Let $\mathcal{V}$ be a vector space and $\mathcal{S} \subseteq \mathcal{V}$. Then $\operatorname{span}\mathcal{S}$ is the smallest subspace of $\mathcal{V}$ that contains $\mathcal{S}$.*

That is, if $\mathcal{W} \subseteq \mathcal{V}$ is a subspace of $\mathcal{V}$ and $\mathcal{S} \subseteq \mathcal{W}$, then $\operatorname{span}\mathcal{S} \subseteq \mathcal{W}$.

Proof. We claim that

$$\operatorname{span}\mathcal{S} = \cap\{\mathcal{W} \in \textsc{Subsp}(\mathcal{V}) : \mathcal{S} \subseteq \mathcal{W}\}.$$

If we can show this, then clearly $\operatorname{span}\mathcal{S} \subseteq \mathcal{W}$ whenever $\mathcal{W}$ is a subspace of $\mathcal{V}$ that contains $\mathcal{S}$.

First observe that if $\mathcal{S} = \varnothing$, then $\operatorname{span}\mathcal{S} = \{\mathbf{0}\}$ by convention, while if $\mathcal{S} \neq \varnothing$, then with $s \in \mathcal{S}$, we have that $\mathbf{0} = 0 \cdot s \in \operatorname{span}\mathcal{S}$. Thus, in either case, $\mathbf{0} \in \operatorname{span}\mathcal{S} \neq \varnothing$.

Next note that if $x, y \in \operatorname{span}\mathcal{S}$, then there exist $s_1, s_2, \ldots, s_n, t_1, t_2, \ldots, t_m \in \mathcal{S}$, and $\alpha_1, \alpha_2, \ldots, \alpha_n, \beta_1, \beta_2, \ldots, \beta_m \in \mathbb{F}$ such that $x = \sum_{j=1}^{n} \alpha_j s_j$ and $y = \sum_{k=1}^{m} \beta_k t_k$. Suppose that $\gamma \in \mathbb{F}$. Then

$$\gamma x + y = \gamma \sum_{j=1}^{n} \alpha_j s_j + \sum_{k=1}^{m} \beta_k t_k = \sum_{j=1}^{n} \gamma\alpha_j s_j + \sum_{k=1}^{m} \beta_k t_k \in \operatorname{span}\mathcal{S}.$$

By the Subspace Test, $\operatorname{span}\mathcal{S}$ is a subspace of $\mathcal{V}$. Moreover, $s = 1 \cdot s \in \operatorname{span}\mathcal{S}$ for all $s \in \mathcal{S}$, so that $\operatorname{span}\mathcal{S}$ contains $\mathcal{S}$. Writing $\mathcal{W}_0 \coloneqq \operatorname{span}\mathcal{S}$, we see that

$$\mathcal{W}_0 \in \{\mathcal{W} \in \textsc{Subsp}(\mathcal{V}) : S \subseteq \mathcal{W}\}.$$

In particular, therefore,

$$\mathcal{W}_0 \supseteq \{\mathcal{W} \in \textsc{Subsp}(\mathcal{V}) : S \subseteq \mathcal{W}\}.$$

If $\mathcal{W} \subseteq \mathcal{V}$ is a subspace of $\mathcal{V}$ and $\mathcal{W}$ contains $\mathcal{S}$, then for any $\{s_1, s_2, \ldots, s_n\} \subseteq \mathcal{S} \subseteq \mathcal{W}$ and $\kappa_1, \kappa_2, \ldots, \kappa_n \in \mathbb{F}$, have that (as $\mathcal{W}$ is a subspace of $\mathcal{V}$),

$$\sum_{j=1}^{n} \kappa_j s_j \in \mathcal{W}.$$

Thus

$$\mathcal{W}_0 = \operatorname{span}\mathcal{S} \subseteq \cap\{\mathcal{W} \in \textsc{Subsp}(\mathcal{V}) : \mathcal{S} \subseteq \mathcal{W}\}.$$

Combining these two containments, we find that

$$\operatorname{span}\mathcal{S} = \{\mathcal{W} \in \textsc{Subsp}(\mathcal{V}) : S \subseteq \mathcal{W}\},$$

which completes the proof.

□

1.7. Definition. *A subset $\mathcal{S}$ of a vector space $\mathcal{V}$ is said to **span** $\mathcal{V}$ (or to **generate** $\mathcal{V}$) if*

$$\textit{span}\ \mathcal{S} = \mathcal{V}.$$

1.8. Exercise. Let $\mathcal{V} = \mathbb{C}_2[x] := \{p_0 + p_1x + p_2x^2 : p_0, p_1, p_2 \in \mathbb{C}\}$, viewed as a vector space over $\mathbb{C}$.

(a) The set $\mathcal{S}_1 := \{1, x, x^2\}$ is a generating set for $\mathcal{V}$.
(b) The set $\mathcal{S}_2 := \{1, 1+x, 1+x+x^2, x^2, x+x^2, 3+2ix^2, -x+x^2\}$ is also a generating set for $\mathcal{V}$.
(c) The set $\mathcal{V}$ is itself a generating set for $\mathcal{V}$. This example has nothing to do with the specific form of $\mathcal{V}$ set out above. If $\mathcal{W}$ is *any* vector space over a field $\mathbb{F}$, then $\mathcal{W}$ is a generating set for $\mathcal{W}$.

1.9. Example. Let $\mathbb{F}$ be a field and $m, n \in \mathbb{N}$. For $1 \leq i \leq m, 1 \leq j \leq n$, define the matrix $E_{i,j} \in \mathbb{M}_{m\times n}(\mathbb{F})$ to be the matrix all of whose entries are zero except for the (i, j)-entry which is $1 \in \mathbb{F}$.

Given $A = [a_{ij}] \in \mathbb{M}_{m\times n}(\mathbb{F})$, we have that

$$A = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} E_{ij},$$

and so $\{E_{i,j} : 1 \leq i \leq m, 1 \leq j \leq n\}$ is a generating set for $\mathbb{M}_{m\times n}(\mathbb{F})$.

We shall refer to the $E_{i,j}$'s as the **standard matrix units** in $\mathbb{M}_{m\times n}(\mathbb{F})$.

1.10. Example. Let $\mathcal{V} = \mathcal{C}([0, 1], \mathbb{R})$, viewed as a vector space over $\mathbb{R}$. Let $\mathcal{S} = \{1, x, x^2, x^3, \ldots\}$. We shall define

$$\mathfrak{P} := \{p_0 + p_1x + \cdots + p_nx^n : n \geq 1, p_j \in \mathbb{R}, 0 \leq j \leq n\}$$

denotes the set of all polynomials with real coefficients, thought of as continuous functions restricted to the interval $[0, 1]$ (so that $\mathfrak{P} \subseteq \mathcal{C}([0, 1], \mathbb{R})$). (The only difference between $\mathfrak{P}$ and $\mathbb{R}[x]$ is that we are thinking of elements of $\mathbb{R}[x]$ as "abstract polynomials" in an indeterminate x, whereas we are specifically thinking of elements of $\mathfrak{P}$ as real-valued functions on $[0, 1]$. Later, we shall learn how to deal with this kind of phenomenon and we will be able to "identify" $\mathbb{R}[x]$ with $\mathfrak{P}$.)

Note that $\mathfrak{P}$ is then a subspace of $\mathcal{V}$ and $\mathcal{S}$ is a generating set for $\mathfrak{P}$ (but not for $\mathcal{V}$ as we have seen above).

Culture: although it is beyond the scope of this course, it can be shown that any generating set for $\mathcal{C}([0, 1], \mathbb{R})$ must be *uncountable*.

1.11. Example. Recall that if p_1, p_2 and $p_3 \in \mathbb{R}^3$ are three non-collinear points, and if we set $x := p_2 - p_1$, $y := p_3 - p_1$, then the (uniquely determined) plane Π containing these three points is given by

$$\{p_1 + \alpha x + \beta y : \alpha, \beta \in \mathbb{R}\}.$$

Thus, letting $\mathcal{W} := \{\alpha x + \beta y : \alpha, \beta \in \mathbb{R}\} = \operatorname{span}\{x, y\}$, we see that $\mathcal{W}$ is a subspace of $\mathbb{R}^3$, and $\Pi = p_1 + \mathcal{W}$ is a coset of that subspace.

1.12. Exercise. Let $\mathcal{V} = \mathbb{R}^3$. Is $(-2, 0, 3)$ a linear combination of $u := (1, 3, 0)$ and $v := (2, 4, -1)$? How can one decide?

1.13. Exercise. Let $\mathcal{V}$ be a vector space and $\mathcal{S}_1 \subseteq \mathcal{S}_2$ be subsets of $\mathcal{V}$. Then

$$\operatorname{span}\mathcal{S}_1 \subseteq \operatorname{span}\mathcal{S}_2.$$

In particular, if $\mathcal{S}_1$ generates $\mathcal{V}$, then so must any set $\mathcal{S}_2$ that contains $\mathcal{S}_1$.

2. Linear independence

2.1. We argued in Section 1 above that the linear span of a set $\mathcal{S}$ of vectors in a vector space $\mathcal{V}$ is trying to tell us *which points in the vector space you can reach* if you start at the origin and you are only allowed to move in (multiples of) the directions prescribed by the vectors in $\mathcal{S}$.

Of course, since $x + y = y + x$ whenever x and y lie in $\mathcal{V}$, it is always possible to change the order of the steps that you take to get from the origin $\mathbf{0}$ to a point $p \in \operatorname{span}\mathcal{S}$. The question of "*linear independence*" asks whether or not there is anything *other than this* that you can do to get back to the origin? In other words, does $\mathcal{S}$ contain "*redundant*" vectors that you could just as well have done without and still obtained the same linear span? If so, we shall say that $\mathcal{S}$ is *linearly dependent*.

The above paragraph is, of course, rather vague. We shall need a precise mathematical definition of *redundancy*, which we now provide.

2.2. Definition. *Let $\mathcal{V}$ be a vector space and $\mathcal{S} \subseteq \mathcal{V}$. The set $\mathcal{S}$ is said to be* ***linearly dependent*** *if there exist finitely many distinct vectors $y_1, y_2, \ldots, y_n \in \mathcal{S}$ and (not necessarily distinct) scalars $\kappa_1, \kappa_2, \ldots, \kappa_n \in \mathbb{F}$, at least one of which is non-zero, such that*

$$\kappa_1 y_1 + \kappa_2 y_2 + \cdots + \kappa_n y_n = \mathbf{0}.$$

If no such finite subset of $\mathcal{S}$ exists, we say that $\mathcal{S}$ is ***linearly independent****.*

2.3. It follows immediately from the definition that $\mathcal{S}$ is linearly *independent* if and only if for all choices of *distinct* vectors $y_1, y_2, \ldots, y_n \in \mathcal{S}$ and scalars $\kappa_1, \kappa_2, \ldots, \kappa_n$, the equation $\kappa_1 y_1 + \kappa_2 y_2 + \cdots \kappa_n y_n = \mathbf{0}$ implies that $\kappa_j = 0$, $1 \leq j \leq n$.

Secondly, in our definition of linear dependence, we require that at least one, but not necessarily all, of the $\kappa_j' s \in \mathbb{F}$ should be non-zero. By simply removing any terms of the linear combination for which the coefficient is zero, we see that we could just as well have defined linear dependence by requiring that *all* of the coefficients κ_j should be non-zero, and yet $\sum_{j=1}^{n} \kappa_j y_j = \mathbf{0}$.

Given distinct vectors $u_1, u_2, \ldots, u_m$ in a vector space $\mathcal{V}$, it is clear that if $\alpha_j := 0$, $1 \leq j \leq m$, then $\alpha_1 u_1 + \alpha_2 u_2 + \cdots + \alpha_n u_n = \mathbf{0}$. It will be useful to refer to this as the **trivial representation** of $\mathbf{0}$ as a linear combination of $u_1, u_2, \ldots, u_m$.

2.4. Example. Let $\mathcal{V}$ be a vector space and $\mathcal{S} \subseteq \mathcal{V}$. If $\mathbf{0} \in \mathcal{S}$, then $\mathcal{S}$ is linearly dependent.

To see this, note that $\mathbf{0} \in \mathcal{S}$, $0 \neq 1 \in \mathbb{F}$, and $1 \cdot \mathbf{0} = \mathbf{0}$, meaning that a non-trivial linear combination of vectors in $\mathcal{S}$ (in this case just the one vector $\mathbf{0}$ itself) yields $\mathbf{0}$.

2.5. Example.

(a) In $\mathbb{R}^3$, the set $\mathcal{S} = \{(2,1,3),(-1,0,1),(0,1,5)\}$ is linearly dependent.
(b) In $\mathbb{R}^3$, the set $\mathcal{S} = \{(2,1,3),(-1,0,1)\}$ is linearly independent.
(c) Let $\mathcal{V} = \mathbb{C}^2$ as a vector space over $\mathbb{C}$. Then $\mathcal{S} = \{(1,1),(1+i,1+i)\}$ is linearly dependent.
(d) Let $\mathcal{V} = \mathbb{C}^2$ as a vector space over $\mathbb{R}$. Then $\mathcal{S} = \{(1,1),(1+i,1+i)\}$ is linearly independent.

A comparison of (c) and (d) shows that the choice of the underlying field is an integral part of the definition of linear independence/dependence.

2.6. Example. Consider the real vector space

$$\mathbb{R}_5[x] = \{p_0 + p_1 x + p_2 x^2 + p_3 x^3 + p_4 x^4 + p_5 x^5 : p_j \in \mathbb{R}, 0 \le j \le 5\}.$$

For each $0 \le j \le 5$, define the polynomial

$$q_j(x) = 1 + x + \cdots + x^j.$$

Then $\mathcal{S} := \{q_0, q_1, q_2, q_3, q_4, q_5\}$ is linearly independent.

2.7. Theorem. *Let $\mathcal{V}$ be a vector space and $\mathcal{S} \subseteq \mathcal{V}$ be a linearly independent set. Suppose that $x \in \mathcal{V} \setminus \mathcal{S}$. The following statements are equivalent:*

(a) *$\mathcal{S} \cup \{x\}$ is linearly dependent.*
(b) *$x \in \operatorname{span}\mathcal{S}$.*

Proof.

(a) implies (b). Suppose that $S \cup \{x\}$ is linearly independent. Then we can find $s_1, s_2, \ldots, s_n \in \mathcal{S}$ and $\kappa_1, \kappa_2, \ldots, \kappa_n, \kappa_{n+1} \in \mathbb{F}$, and *not all equal to* 0 such that

$$\kappa_1 s_1 + \kappa_2 s_2 + \cdots + \kappa_n s_n + \kappa_{n+1} x = \mathbf{0}.$$

If $\kappa_{n+1} = 0$, then at least one of the other κ_j's is not zero, proving that $\{s_1, s_2, \ldots, s_n\}$ is linearly *dependent*, and thus S is linearly dependent, a contradiction. Thus $\kappa_{n+1} \neq 0$.

But then

$$x = (-\kappa_{n+1})^{-1}\kappa_{n+1} x = (-\kappa_{n+1})^{-1} \sum_{j=1}^{n} \kappa_j s_j = \sum_{j=1}^{n} ((-\kappa_{n+1})^{-1}\kappa_j) s_j \in \operatorname{span}\mathcal{S}.$$

(b) implies (a). If $x \in \operatorname{span}\mathcal{S}$, then there exist $s_1, s_2, \ldots, s_n \in \mathcal{S}$ and $\alpha_1, \alpha_2, \ldots, \alpha_n \in \mathbb{F}$ such that

$$x = \sum_{j=1}^{n} \alpha_j s_j.$$

But then

$$\mathbf{0} = (-1) \cdot x + \alpha_1 s_1 + \alpha_2 s_2 + \cdots + \alpha_n s_n.$$

Since at least one of the coefficients in this linear combination is non-zero (namely the coefficient of x, which is -1), $\mathcal{S} \cup \{x\}$ is linearly dependent.

□

2.8. Example. Let $\mathcal{V}$ be a vector space and $x, y \in \mathcal{V}$ with $x \neq y$. Then $\{x, y\}$ is linearly dependent if and only if one of x and y is a multiple of the other.

(Why didn't we just say "if and only if y is a multiple of x"?)

2.9. Example. Let $n \in \mathbb{N}$ and $\mathbb{F}$ be a field. The set $\mathcal{S} := \{E_{i,j} : 1 \leq i, j \leq n\}$ of all matrix units in $\mathbb{M}_n(\mathbb{F})$ is both linearly independent and a generating set for $\mathbb{M}_n(\mathbb{F})$.

Supplementary Examples

S3.1. Example. Let $\mathbb{F}$ be a field and $n \in \mathbb{N}$. Let $E_{i,j}$, $1 \le i, j \le n$ denote the standard matrix units of $\mathbb{M}_n(\mathbb{F})$. Then $\mathcal{S} := \{E_{i,j} : 1 \le i \le j \le n\}$ is a spanning set for the space $\mathbb{T}_n(\mathbb{F})$ of upper-triangular $n \times n$ matrices over $\mathbb{F}$.

S3.2. Example. Let $\mathcal{V} = \mathbb{M}_2(\mathbb{R})$, and set

$$\mathcal{S} = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \right\}.$$

Then $\mathcal{S}$ is a generating set for $\mathfrak{sl}_2(\mathbb{R}) := \{T = [t_{i,j}] \in \mathbb{M}_2(\mathbb{R}) : \mathrm{tr}(T) = 0\}$.

S3.3. Example. Recall from Example 2.1.7 the vector space $\mathcal{V} := \{\alpha\,\mathrm{PIG} + \beta\,\mathrm{DOG} : \alpha, \beta \in \mathbb{K}\}$, with the understanding that

$$\alpha_1\,\mathrm{PIG} + \beta_1\,\mathrm{DOG} = \alpha_2\,\mathrm{PIG} + \beta_2\,\mathrm{DOG}$$

if and only if $\alpha_1 = \alpha_2$ and $\beta_1 = \beta_2$.

Then $\mathcal{S} := \{\mathrm{PIG}, \mathrm{DOG}\}$ is a spanning set for $\mathcal{V}$, and the understanding we listed above ensures that PIG and DOG are linearly independent!

Note that $\mathcal{S}_2 := \{\mathrm{PIG} + \mathrm{DOG}, \mathrm{PIG} - \mathrm{DOG}\}$ is another spanning set for $\mathcal{V}$. Are $y_1 := \mathrm{PIG} + \mathrm{DOG}$ and $y_2 := \mathrm{PIG} - \mathrm{DOG}$ linearly independent?

S3.4. Example. Let $p \in \mathbb{N}$ be a prime number and recall that $\mathbb{F} := \mathbb{Z}_p$ is a field. As in Example 1.**??**, for any $m \ge 1$, $\mathbb{F}^m$ may be thought of as a vector space over $\mathbb{F}$ – either as row vectors, or as column vectors.

How many vectors are there in $\mathcal{V} := (\mathbb{Z}_3)^4$? An arbitrary element of $(\mathbb{Z}_3)^4$ looks like $x = (x_1, x_2, x_3, x_4)$, where $x_j \in \mathbb{Z}_3$ for each $1 \le j \le 4$. This yields $81 = 3^4$ possible vectors. A spanning set for $\mathcal{V}$ is $\mathcal{V}$ itself, but more interestingly, if we set $e_1 = (1,0,0,0)$, $e_2 = (0,1,0,0)$, $e_3 = (0,0,1,0)$ and $e_4 = (0,0,0,1)$, then $\mathcal{E} := \{e_1, e_2, e_3, e_4\}$ is a spanning set for $\mathcal{V}$ which is also linearly independent.

S3.5. Example. Let $\mathcal{V}$ be a vector space over a field $\mathbb{F}$, and suppose that $\{\mathcal{W}_\lambda\}_{\lambda \in \Lambda}$ is a collection of subspaces of $\mathcal{V}$. Then

$$\mathcal{W} := \cap_{\lambda \in \Lambda} \mathcal{W}_\lambda$$

is a subspace of $\mathcal{V}$.

Indeed, since each $\mathcal{W}_\lambda$ is a subspace of $\mathcal{V}$, we have that $\mathbf{0} \in \mathcal{W}_\lambda$ for all $\lambda \in \Lambda$. But then $\mathbf{0} \in \mathcal{W}$, so $\mathcal{W} \neq \varnothing$.

If $x, y \in \mathcal{W}$ and $\kappa \in \mathbb{F}$, then for each $\lambda \in \Lambda$, $x, y \in \mathcal{W}_\lambda$ and since $\mathcal{W}_\lambda$ is a subspace of $\mathcal{V}$,

$$\kappa x + y \in \mathcal{W}_\lambda \quad \text{for all} \lambda \in \Lambda.$$

Hence $\kappa x + y \in \mathcal{W}$, so $\mathcal{W}$ is a subspace of $\mathcal{V}$.

S3.6. Example. Suppose that $x = (3,7)$, $y = (-\pi, e)$ and $z = (-4,1) \in \mathbb{R}^2$, the latter viewed as a vector space over $\mathbb{R}$. We claim that $\mathcal{S} = \{x, y, z\}$ is linearly dependent.

To see this, we try to solve the equation:

$$\alpha_1 x + \alpha_2 y + \alpha_3 z = \mathbf{0},$$

which translates to the system of equations

$$3\alpha_1 - \pi\alpha_2 - 4\alpha_3 = 0$$
$$7\alpha_1 + e\alpha_2 + 1\alpha_3 = 0.$$

This, like every homogeneous system of m equations in n variables (where $m < n$), has infinitely many solutions.

In this particular instance, we can set $\alpha_2 := t$, and then we must solve

$$3\alpha_1 - 4\alpha_3 = \pi t$$
$$7\alpha_1 + 1\alpha_3 = -et.$$

Adding four times the second row to the first yields:

$$31\alpha_1 = (\pi - 4e)t,$$

or equivalently,

$$\alpha_1 = \frac{\pi - 4e}{31} t,$$

and plugging this into the second equation yields

$$\alpha_3 = -et - \frac{7(\pi - 4e)}{31} t.$$

The point is that each choice of $t \in \mathbb{R}$ yields a choice of $\alpha_1, \alpha_2, \alpha_3 \in \mathbb{R}$ which implements the linear dependence of x, y and z.

For example, we can take $t = 1$, $\alpha_1 = \frac{\pi - 4e}{31}$, $\alpha_2 = 1$ and $\alpha_3 = -e - \frac{7(\pi - 4e)}{31}$.

This technique of solving the problem leads one to ask: can we find three vectors $x, y, z \in \mathbb{R}^2$ such that $\mathcal{S} = \{x, y, z\}$ is linearly independent?

S3.7. Example. We know that $\mathbb{R}$ is a vector space over $\mathbb{R}$. Let $\mathcal{S} = \{1, \pi\}$. Then $\mathcal{S}$ is linearly dependent. Indeed, we must solve the equation

$$\kappa_1 \cdot 1 + \kappa_2 \pi = 0.$$

This admits an easy solution: take $\kappa_1 = -\pi$ and $\kappa_2 = 1$. (This solution is not unique.)

However, $\mathbb{R}$ is also a vector space over $\mathbb{Q}$. Viewed this way, $\mathcal{S}$ is linearly *independent*. Indeed, consider $\alpha_1, \alpha_2 \in \mathbb{Q}$ – not both equal to zero – satisfying

$$\alpha_1 \cdot 1 + \alpha_2 \cdot \pi = 0.$$

It is easy to see that if $\alpha_1 \neq 0$, then we can't have $\alpha_2 = 0$, while if $\alpha_2 \neq 0$, then $\alpha_1 \neq 0$. Hence both α_1, α_2 are non-zero.

But then

$$\pi = -\frac{\alpha_1}{\alpha_2} \in \mathbb{Q},$$

which is known to be false! (If we don't know that this is false – replace π by $\sqrt{2} \notin \mathbb{Q}$. *This* you should be able to prove!)

S3.8. Example. Recall that in Example 1.**??** we defined $c_0(\mathbb{R}) := \{x = (x_n)_n \in \mathbb{R}^{\mathbb{N}} : \lim_n x_n = 0\}$.

For each $n \geq 1$, define $e_n := (0, 0, \ldots, 0, 1, 0, 0, \ldots)$, where the unique "1" occurs in the n^{th} coordinate. Let $\mathcal{S} := \{e_1, e_2, e_3, \ldots\}$. Then

$$\operatorname{span} \mathcal{S} = c_{00}(\mathbb{R}) := \{x = (x_1, x_2, x_3, \ldots) : x_j = 0 \text{ except for finitely many } j \in \mathbb{N}\}.$$

As such, the vector $y := (1, \frac{1}{2}, \frac{1}{3}, \frac{1}{3}, \ldots) \in c_0(\mathbb{R})$, but $y \notin \operatorname{span} \mathcal{S}$.

S3.9. Example. Let $\mathcal{V}$ be a vector space over $\mathbb{K}$ and suppose that $\mathcal{S} := \{x, y\} \subseteq \mathcal{V}$ is linearly independent. We claim that the set $\mathcal{S}_0 := \{x - y, x + y\}$ is also linearly independent. Indeed, if $\alpha, \beta \in \mathbb{K}$ are not both zero and $\alpha(x - y) + \beta(x + y) = \mathbf{0}$, then

$$(\alpha + \beta)x + (-\alpha + \beta)y = \mathbf{0}.$$

But $\mathcal{S}$ is linearly independent by hypothesis, and so $\alpha + \beta = 0 = -\alpha + \beta$, from which we deduce that $\alpha = 0 = \beta$, and thus $\mathcal{S}_0$ is linearly independent.

Something interesting happens, however, when the characteristic of the field is two. (In case you are not familiar with the *characteristic* of a field, we shall consider the special case where $\mathbb{F} = \mathbb{Z}_2$ (which just so happens to have characteristic two).

Thus, let $\mathcal{V} \neq \{\mathbf{0}\}$ be a vector space over the field $\mathbb{Z}_2$, and suppose that $\mathcal{S} := \{x, y\} \subseteq \mathcal{V}$ is linearly independent. Then $\mathcal{S}_0 := \{x - y, x + y\}$ is *never* linearly independent. The reason is that in this case, $-y = y$ (because $-1 = 1$ in $\mathbb{Z}_2$), and thus $x - y = x + y$, meaning that

$$1 \cdot (x - y) + (-1) \cdot (x + y) = 1 \cdot (x - y) + 1 \cdot (x + y) = \mathbf{0}.$$

Fields of characteristic two play a special role in vector space and matrix theory, and a great many theorems include a phrase such as: "let $\mathcal{V}$ be a vector space over a field of characteristic not equal to two". Just saying.

S3.10. Example. Let $\mathcal{V} := \mathbb{R}$, viewed as a vector space over $\mathbb{Q}$. Let p and $q \in \mathbb{N}$ be distinct prime numbers. Then $\mathcal{S} := \{\sqrt{p}, \sqrt{q}\}$ is linearly independent.

Indeed, suppose that $\alpha, \beta \in \mathbb{Q}$ and $\alpha\sqrt{p} + \beta\sqrt{q} = 0$. Note that by multiplying by a sufficiently large integer (namely the product of the denominators of α and β), we may assume without loss of generality that α and β are integers!

Hence

$$\begin{aligned} 0 &= (\alpha\sqrt{p} + \beta\sqrt{q})^2 \\ &= \alpha^2 p + 2\alpha\beta\sqrt{pq} + \beta^2 q, \end{aligned}$$

whence

$$\sqrt{pq} = -\frac{\alpha^2 p + \beta^2 q}{2\alpha\beta} \in \mathbb{Q}.$$

Write $\sqrt{pq} = \frac{a}{b}$, where $a, b \in \mathbb{Z}$ and a and b have no common prime factors. Then

$$pq = \frac{a^2}{b^2},$$

so p must divide a^2 and hence it must divide a. But then p does not divide b and so p^2 divides a^2 and thus $\frac{a^2}{b^2} = pq$, a contradiction.

Thus $\mathcal{S}$ is linearly independent.

If p, q, and $r \in \mathbb{N}$ are three distinct prime numbers, is $\mathcal{S}_3 := \{\sqrt{p}, \sqrt{q}, \sqrt{r}\}$ linearly independent (over $\mathbb{Q}$)? What if we consider

$$\mathcal{S}_\infty := \{\sqrt{2}, \sqrt{3}, \sqrt{5}, \sqrt{7}, \ldots\},$$

the set of square roots of *every* prime number. Is $\mathcal{S}_\infty$ linearly independent in $\mathbb{R}$ over $\mathbb{Q}$?

Appendix

A3.1. Linear algebra pervades many areas of mathematics. To wit.

If $\alpha \in \mathbb{R}$, then we say that α is **algebraic** if there exist finitely many integers $\kappa_0, \kappa_1, \ldots, \kappa_n$ – not all equal to zero!!! – such that

$$\kappa_0\alpha^0 + \kappa_1\alpha^1 + \kappa_2\alpha^2 + \cdots\kappa_n\alpha^n = 0.$$

Otherwise, we say that α is **transcendental**.

Note that this condition is equivalent to the existence of *rational numbers* $\gamma_0, \gamma_1, \ldots, \gamma_n$ (again, not all equal to 0) such that

$$\gamma_0\alpha^0 + \gamma_1\alpha^1 + \gamma_2\alpha^2 + \cdots\gamma_n\alpha^n = 0.$$

(You should convince yourselves that if the γ_j's as above exist, then so do the κ_j's.)

As such, the question of whether or not $\alpha \in \mathbb{R}$ is algebraic or transcendental becomes a question of whether the set

$$\mathcal{S} := \{1, \alpha, \alpha^2, \alpha^3, \ldots\}$$

is "*linearly dependent*" or "*linearly independent*" in $\mathbb{R}$ over the field $\mathbb{Q}$.

In case you are not aware of this fact: both e and π are transcendental. The proofs are not entirely trivial.

Note that $\alpha := \sqrt{2}$ is algebraic, as can be seen by taking $\kappa_0 = 2$, $\kappa_1 = 0$ and $\kappa_2 = -1$ to get

$$\kappa_0\alpha^0 + \kappa_1\alpha^1 + \kappa_2\alpha^2 = 2 \cdot 1 + 0 \cdot \sqrt{2} + (-1) \cdot \left(\sqrt{2}\right)^2 = 2 - 2 = 0.$$

Culture. Is the set Λ of all algebraic numbers countable or uncountable?

Exercises for Chapter 3

Exercise 3.1.
Find a finite generating set for the set $(\mathbb{M}_n(\mathbb{R}))_{sym}$ of symmetric matrices over $\mathbb{R}$.

Exercise 3.2.
Note that $\mathbb{Q}$ is a subfield of $\mathbb{R}$. If $\mathcal{V}$ is a vector space over $\mathbb{R}$, prove that $\mathcal{V}$ is a vector space over $\mathbb{Q}$. Is the converse true?

Exercise 3.3.
Consider the set $\mathcal{S} := \{\sin x, \cos x, \sin(2x), \cos(2x), \sin(3x), \cos(3x)\}$ in the real vector space $\mathcal{C}([0,1],\mathbb{R})$. Is $\mathcal{S}$ linearly dependent, or linearly independent?

Exercise 3.4.*
Let $\mathcal{V}$ be a vector space over a field $\mathbb{F}$, and suppose that $\mathcal{S}_1 \subseteq \mathcal{S}_2 \subseteq \mathcal{V}$. Prove that if $\mathcal{S}_1$ is linearly dependent, then so is $\mathcal{S}_2$. Thus, if $\mathcal{S}_2$ is linearly independent, then so is $\mathcal{S}_1$.

That is, any subset of a linearly independent set is linearly independent. This should satisfy our "intuition". If a set has no built-in redundancy, we shouldn't expect any of its subsets to have any redundancy. Of course, intuition is not proof – you are required to produce the latter!

Exercise 3.5.*
Let $\mathcal{V}$ be a vector space over a field $\mathbb{F}$, and suppose that $\mathcal{S} \subseteq \mathcal{V}$. Show that $\mathcal{S}$ is linearly dependent if and only if there exists a *proper* subset $\mathcal{S}_0 \subseteq \mathcal{S}$ of $\mathcal{S}$ (**proper** here means that $\mathcal{S}_0 \neq \mathcal{S}$) such that

$$\operatorname{span} \mathcal{S}_0 = \operatorname{span} \mathcal{S}.$$

Exercise 3.6.
Let $\mathcal{V}$ be a vector space over a field $\mathbb{F}$, and suppose that $\mathcal{S}$ is a spanning set for $\mathcal{V}$. If $\mathcal{W}$ is a subspace of $\mathcal{V}$, then

$$\mathcal{S}_{\mathcal{Q}} := \{y + \mathcal{W} : y \in \mathcal{S}\}$$

is a spanning set for the quotient space $\mathcal{V}/\mathcal{W}$.

Exercise 3.7.
Suppose that $\mathcal{V}$ is a vector space over a field $\mathbb{F}$ and that $\mathcal{W}$ is a subspace of $\mathcal{V}$. Suppose that $\mathcal{L} \subseteq \mathcal{V}$ is a linearly independent set.

(a) Does $\mathcal{L}_{\mathcal{Q}} := \{y + \mathcal{W} : y \in \mathcal{L}\}$ have to be linearly independent in $\mathcal{V}/\mathcal{W}$? Prove it is true or provide a counterexample to show that it can be false.
(b) Can $\mathcal{L}_{\mathcal{Q}}$ ever be linearly independent in $\mathcal{V}/\mathcal{W}$? Give an example to show that it can be, or prove that it can never be.
(c) Give a necessary and sufficient condition for $\mathcal{L}_{\mathcal{Q}}$ to be linearly independent in $\mathcal{V}/\mathcal{W}$.

Exercise 3.8.

Let $\mathcal{V} = \{\mathbf{0}\}$ as a vector space over a field $\mathbb{F}$. Is $\mathcal{S}_1 := \varnothing$ linearly independent or linearly dependent in $\mathcal{V}$? Is $\mathcal{S}_2 := \{\mathbf{0}\}$ linearly independent or linearly dependent in $\mathcal{V}$?

Exercise 3.9.

Let $\mathcal{V}$ be a vector space over $\mathbb{Q}$, and suppose that $\mathcal{L} := \{y_1, y_2, y_3\}$ is linearly independent in $\mathcal{V}$. Let

- $w_1 := 3y_1 - 2y_2 + 7y_3$;
- $w_2 := 3y_1 + 2y_2 - 6y_3$; and
- $w_3 := y_1 + y_2 - 3y_3$.

Determine whether or not $\mathcal{L}_2 := \{w_1, w_2, w_3\}$ is linearly independent in $\mathcal{V}$.

Exercise 3.10.*

Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces over a common field $\mathbb{F}$. Recall from Exercise 2.8 that a map $T : \mathcal{V} \to \mathcal{W}$ is said to be **linear** if

$$T(\kappa x + y) = \kappa Tx + Ty \qquad \text{for all } \kappa \in \mathbb{F},\ x, y \in \mathcal{V}.$$

Suppose that $\mathcal{B} = \{b_\lambda\}_{\lambda \in \Lambda} \subseteq \mathcal{V}$ is a linearly independent subset of $\mathcal{V}$ which also spans $\mathcal{V}$. For each $\lambda \in \Lambda$, let $w_\lambda \in \mathcal{W}$ be arbitrary.

Prove that there exists a *unique* linear map $T : \mathcal{V} \to \mathcal{W}$ such that $Tb_\lambda = w_\lambda$ for all $\lambda \in \Lambda$.

Compare this with the result from Exercise 2.9. Think about what might be underlying this phenomenon.

CHAPTER 4

Bases and dimension

If you want to know what God thinks of money, just look at the people he gave it to.

Dorothy Parker

1. Hamel bases

1.1. The word "basis" appears in more than one area of mathematics, and it can mean different things to different people. In dealing with vector spaces, one usually interprets "basis" to mean what is technically known as a "*Hamel basis*". When dealing with so-called *Banach* or *Hilbert* spaces (which are vector spaces with extra properties), there are other types of "bases" that may be more natural and useful to consider, e.g. *Hilbert space bases*. Those are, however, *not* vector space bases when the underlying space is infinite-dimensional. In this course, we are only interested in "vector space" bases, i.e. *Hamel bases*, and it is standard to simply refer to these as "bases".

1.2. Definition. *A **(Hamel) basis** for a vector space $\mathcal{V}$ is a maximal linearly independent set $\mathcal{B} \subseteq \mathcal{V}$. That is, $\mathcal{B}$ is linearly independent, and if $\mathcal{B} \subseteq \mathcal{D} \subseteq \mathcal{V}$ where $\mathcal{D}$ is linearly independent, then $\mathcal{D} = \mathcal{B}$.*

1.3. Proposition. *Let $\mathcal{V}$ be a vector space over a field $\mathbb{F}$ and let $\mathcal{B} \subseteq \mathcal{V}$. The following are equivalent.*

(a) *$\mathcal{B}$ is a basis for $\mathcal{V}$.*
(b) *$\mathcal{B}$ is linearly independent and generates $\mathcal{V}$.*

Proof. This is an immediate consequence of Theorem 3.2.7.

□

1.4. Example. The set $\mathcal{B} := \varnothing$ is a basis for $\mathcal{V} = \{\mathbf{0}\}$.

1.5. Example.

(a) Let $\mathcal{V} = \mathbb{F}^n$ be a vector space over the field $\mathbb{F}$. The **standard basis** for $\mathbb{F}^n$ is $\mathcal{B} := \{e_k\}_{k=1}^n$, where

$$e_k = (0, 0, \ldots, 0, 1, 0, \ldots, 0),$$

with the unique "1" appearing at the k^{th} coordinate, $1 \leq k \leq n$.

(b) Let $m, n \in \mathbb{N}$ and consider $\mathcal{V} = \mathbb{M}_{m\times n}(\mathbb{F})$ as a vector space over $\mathbb{F}$. The **standard basis** for $\mathbb{M}_{m\times n}(\mathbb{F})$ is the set

$$\mathcal{B} := \{E_{i,j}\}_{1\leq i\leq m, 1\leq j\leq n},$$

the set of matrix units.

(c) The **standard basis** for $\mathbb{F}[x]$ is $\mathcal{B} = \{1, x, x^2, x^3, \ldots\}$.

(d) If $n \in \mathbb{N}$, the **standard basis** for $\mathbb{F}_n[x]$ is $\mathcal{B} = \{1, x, x^2, \ldots, x^n\}$.

(e) Consider $\mathcal{C}([0, 1], \mathbb{R})$ as a vector space over $\mathbb{R}$. Does this vector space have a basis? If so, what is it?

1.6. Theorem. *Let $\mathcal{V}$ be a vector space over a field $\mathbb{F}$, and let $\mathcal{B} \subseteq \mathcal{V}$. The following are equivalent:*

(a) *$\mathcal{B}$ is a basis for $\mathcal{V}$.*

(b) *Given $\mathbf{0} \neq x \in \mathcal{V}$, there exists a unique choice of **non-zero** scalars $\kappa_1, \kappa_2, \ldots, \kappa_n$ and **distinct** vectors $b_1, b_2, \ldots, b_n \in \mathcal{B}$ such that*

$$x = \sum_{j=1}^{n} \kappa_j b_j.$$

Note: the order of the terms here doesn't matter - since addition is commutative in a vector space, we can always permute the terms. Also, no uniqueness is possible if we allow zero coefficients, or if we allow repetition of vectors – e.g. if $b_1 = b_2$ – and so to have the definition make sense, we are obliged to add the conditions that we did.

Proof.

(a) implies (b). Let $\mathbf{0} \neq x \in \mathcal{V} = \operatorname{span}\mathcal{B}$. Thus we can find $b_1, b_2, \ldots, b_m \in \mathcal{B}$ and $\alpha_1, \alpha_2, \ldots, \alpha_m \in \mathbb{F} \setminus \{0\}$ such that

$$x = \alpha_1 b_1 + \alpha_2 b_2 + \cdots + \alpha_m b_m.$$

Suppose that we can also find $d_1, d_2, \ldots, d_n \in \mathcal{B}$ and $\beta_1, \beta_2, \ldots, b_n \in \mathbb{F} \setminus \{0\}$ such that

$$x = \beta_1 d_1 + \beta_2 d_2 + \cdots + \beta_n d_n.$$

Let $p := |\{b_1, b_2, \ldots, b_m\} \cap \{d_1, d_2, \ldots, d_n\}|$. By reindexing the b_i's and the d_k's, we may assume without loss of generality that $b_i = d_i$, $1 \leq i \leq p$. Thus

$$\{b_1, b_2, \ldots, b_m\} \cup \{d_1, d_2, \ldots, d_n\} = \{b_1, b_2, \ldots, b_p, b_{p+1}, b_{p+2}, \ldots, b_m, d_{p+1}, d_{p+2}, \ldots, d_n\}.$$

Moreover,

$$\begin{aligned}\mathbf{0} &= x - x \\ &= \sum_{i=1}^{p}(\alpha_i - \beta_i)b_i + \sum_{j=p+1}^{n} \alpha_j b_j + \sum_{k=p+1}^{n} -\beta_k d_k.\end{aligned}$$

Since $\mathcal{B}$ is linearly independent, we see that $\alpha_j = 0$, $p+1 \le j \le m$ and $\beta_k = 0$, $p+1 \le k \le n$. Since α_j, β_k were non-zero, we conclude that $p = m = n$, and then we note that $\alpha_i = \beta_i$, $1 \le i \le p = m = n$, which means that the linear combination was indeed unique.

(b) implies (a). Suppose that (b) holds. Then every $\mathbf{0} \neq x \in \mathcal{V}$ lies in $\operatorname{span}\mathcal{B}$, and of course, $\mathbf{0} \in \operatorname{span}\mathcal{B}$, so that $\operatorname{span}\mathcal{B} = \mathcal{V}$. Suppose that $\mathcal{B}$ is not linearly independent. Then we can find $b_1, b_2, \ldots, b_m \in \mathcal{B}$ and $\kappa_1, \kappa_2, \ldots, \kappa_m \in \mathbb{F} \setminus \{0\}$ such that

$$\mathbf{0} = \sum_{j=1}^{m} \kappa_j b_j.$$

Since $\mathcal{V} \neq \{\mathbf{0}\}$ and $\operatorname{span}\mathcal{B} = \mathcal{V}$, there exists $\mathbf{0} \neq d \in \mathcal{B}$. Then

$$\begin{aligned} d &= 1d \\ &= 1d + \sum_{j=1}^{m} \kappa_j b_j \end{aligned}$$

are two distinct linear combinations giving rise to d. By eliminating the terms of the second expansion with 0 coefficients, we obtain a contradiction of (b). That is, (b) implies that $\mathcal{B}$ is linearly independent, and is thus a basis.

□

1.7. Theorem. *Let $\mathcal{V}$ be a vector space and suppose that $\mathcal{S} \subseteq \mathcal{V}$ be a finite spanning set for $\mathcal{V}$. Then $\mathcal{S}$ contains a basis $\mathcal{B}$ for $\mathcal{V}$.*

Proof. Let $\mathcal{J} := \{L \subseteq \mathcal{S} : L \text{ is linearly independent}\}$. Every $L \in \mathcal{J}$ is finite with $|L| \le |\mathcal{S}|$, and so we can choose $L_0 \in \mathcal{J}$ such that $|L_0| \ge |L|$ for all $L \in \mathcal{J}$.

If $s \in \mathcal{S} \setminus \mathcal{L}_0$, then $L_0 \cup \{s\}$ is linearly independent, so $s \in \operatorname{span} L_0$. Thus $\mathcal{S} \subseteq \operatorname{span} L_0$, whence

$$\mathcal{V} = \operatorname{span}\mathcal{S} \subseteq \operatorname{span} L_0.$$

Thus L_0 is a linearly independent set which spans $\mathcal{V}$, and so it is a basis for $\mathcal{V}$ which is contained in $\mathcal{S}$.

□

1.8. Example. Consider $\mathbb{Q}_2[x] := \{p = p_0 + p_1 x + p_2 x^2 \mid p_0, p_1, p_2 \in \mathbb{Q}\}$ as a vector space over $\mathbb{Q}$. Let

$$\mathcal{S} = \{1, 1+x, 2+x, 4+2x, x+x^2\},$$

and observe that $\operatorname{span}\mathcal{S} = \mathbb{Q}_2[x]$.

Note that $\mathcal{B}_1 := \{1, 1+x, x+x^2\}$ and $\mathcal{B}_2 := \{1+x, 2+x, x+x^2\}$ are both bases for $\mathbb{Q}_2[x]$, and $\mathcal{B}_j \subseteq \mathcal{S}$, $j = 1, 2$.

It is interesting to note that $\mathcal{B}_1$ and $\mathcal{B}_2$ have the same number of elements. Is this just a "fluke", or is there something behind this? Before answering this question, we require a Lemma, and an interesting result of Steinitz.

1.9. Lemma. *Let $\mathcal{V}$ be a vector space over the field $\mathbb{F}$, and let $\mathcal{S} := \{s_1, s_2, \ldots, s_m\}$, $\mathcal{T} := \{y_1, y_2, \ldots, y_n\} \subseteq \mathcal{V}$.*

(a) *If $\mathcal{T} \subseteq \operatorname{span}\mathcal{S}$ and $\mathcal{S} \subseteq \operatorname{span}\mathcal{T}$, then $\operatorname{span}\mathcal{S} = \operatorname{span}\mathcal{T}$.*

(b) *If $y \in \operatorname{span}\mathcal{S}$ but $y \notin \operatorname{span}\{s_1, s_2, \ldots, s_{m-1}\}$, then*

$$s_m \in \operatorname{span}\{s_1, s_2, \ldots, s_{m-1}, y\},$$

and so

$$\operatorname{span}\mathcal{S} = \operatorname{span}\{s_1, s_2, \ldots, s_{m-1}, y\}.$$

(c) *If $\mathcal{C} \subseteq \mathcal{V}$, then $\operatorname{span}(\mathcal{S} \cup \mathcal{C}) = \operatorname{span}(\{s_1, s_2, \ldots, s_{m-1}, y\} \cup \mathcal{C})$.*

Proof.

(a) As we have seen, if $\mathcal{X} \subseteq \mathcal{V}$ is any set, then $\operatorname{span}\mathcal{X}$ is the smallest subspace of $\mathcal{V}$ which contains $\mathcal{X}$. In other words,

$$\operatorname{span}\mathcal{X} = \cap\{\mathcal{W} : \mathcal{W} \text{ is a subspace of } \mathcal{V} \text{ and } \mathcal{X} \subseteq \mathcal{W}\}.$$

It follows that if $\mathcal{T} \subseteq \operatorname{span}\mathcal{S}$, then $\operatorname{span}\mathcal{S}$ is *a* subspace of $\mathcal{V}$ that contains $\mathcal{T}$, and thus $\operatorname{span}\mathcal{S}$ is just one of the spaces $\mathcal{W}$ occurring in the intersection above. Hence

$$\operatorname{span}\mathcal{T} \subseteq \operatorname{span}\mathcal{S}.$$

By symmetry, if $\mathcal{S} \subseteq \mathcal{T}$, then $\operatorname{span}\mathcal{S} \subseteq \operatorname{span}\mathcal{T}$, and so equality follows.

(b) Let $\mathcal{T} := \{s_1, s_2, \ldots, s_{m-1}, y\}$.

Since $y \in \operatorname{span}\mathcal{S}$, we can find $\kappa_1, \kappa_2, \ldots, \kappa_m \in \mathbb{F}$ such that

$$y = \kappa_1 s_1 + \kappa_2 s_2 + \cdots + \kappa_m s_m.$$

If $\kappa_m = 0$, then $y \in \operatorname{span}\{s_1, s_2, \ldots, s_{m-1}\}$, a contradiction. Thus $\kappa_m \neq 0$.

From this it follows that

$$\kappa_m s_m = y - \kappa_1 s_1 - \kappa_2 s_2 - \cdots - \kappa_{m-1} s_{m-1},$$

or equivalently,

$$s_m = (\kappa_m)^{-1} y - (\kappa_m)^{-1}\kappa_1 s_1 - (\kappa_m)^{-1}\kappa_2 s_2 - \cdots - (\kappa_m)^{-1}\kappa_{m-1} s_{m-1}.$$

Thus $s_m \in \operatorname{span}\mathcal{T}$.

Since $s_j \in \operatorname{span}\mathcal{T}$ for all $1 \leq j \leq m-1$, we conclude that $\mathcal{S} \subseteq \operatorname{span}\mathcal{T}$. But $\mathcal{T} \subseteq \operatorname{span}\mathcal{S}$ as well, and so by part (a), $\operatorname{span}\mathcal{S} = \operatorname{span}\mathcal{T}$, as claimed.

(c) Again, we set $\mathcal{T} = \{s_1, s_2, \ldots, s_{m-1}, y\}$. If $\mathcal{C} \subseteq \mathcal{V}$, then

$$\begin{aligned}\operatorname{span}(\mathcal{S} \cup \mathcal{C}) &= \operatorname{span}(\operatorname{span}\mathcal{S} \cup \operatorname{span}\mathcal{C}) \\ &= \operatorname{span}(\operatorname{span}\mathcal{T} \cup \operatorname{span}\mathcal{C}) \\ &= \operatorname{span}(\mathcal{T} \cup \mathcal{C}).\end{aligned}$$

□

1.10. Theorem. (Steinitz's Replacement Theorem)
Let $\mathcal{V}$ be a vector space, $n \in \mathbb{N}$ and suppose that $\mathcal{S} \subseteq \mathcal{V}$ is a set with n elements. Suppose that $\mathcal{L} := \{y_1, y_2, \ldots, y_m\} \subseteq \operatorname{span}\mathcal{S}$ is linearly independent. Then

(a) *there exists $\mathcal{H} \subseteq \mathcal{S}$ such that $\mathcal{H}$ has $n - m$ elements, and*

$$\operatorname{span}\mathcal{S} = \operatorname{span}(\mathcal{L} \cup \mathcal{H}).$$

(b) *In particular, $m \leq n$.*

Proof. We shall argue by induction on m. If $m = 0$, there is nothing to prove.

Now suppose that $m \geq 1$, and that the result holds whenever $\mathcal{L}$ has fewer than m elements. We shall prove that the result holds when $\mathcal{L}$ has m elements, completing the induction step and thereby the proof as well. Note that the induction step implies that $n \geq (m-1)$.

Let $\mathcal{L} = \{y_1, y_2, \ldots, y_m\} \subseteq \mathcal{V}$ be a linearly independent set. Then $\{y_1, y_2, \ldots, y_{m-1}\}$ is a linearly independent set with $m - 1 < m$ elements, and so our induction hypothesis implies that we can find a subset $\mathcal{H}_0 \subseteq \mathcal{S}$ with $n - (m-1)$ elements, say $\mathcal{H}_0 = \{t_m, t_{m+1}, t_{m+2}, \ldots, t_n\} \subseteq \mathcal{S}$ such that

$$\operatorname{span}\mathcal{S} = \operatorname{span}(\{y_1, y_2, \ldots, y_{m-1}\} \cup \mathcal{H}_0).$$

Recall that $y_m \in \operatorname{span}\mathcal{S}$, but $y_m \notin \operatorname{span}\{y_1, y_2, \ldots, y_{m-1}\}$, as $\mathcal{L}$ is linearly independent. Therefore we can choose $m \leq q \leq n$ such that

$$y_m \in \operatorname{span}\{y_1, y_2, \ldots, y_{m-1}, t_m, t_{m+1}, \ldots, t_q\},$$

but

$$y_m \notin \operatorname{span}\{y_1, y_2, \ldots, y_{m-1}, t_m, t_{m+1}, \ldots, t_{q-1}\}.$$

In particular, we must have $n \geq m$ (otherwise $y_m \notin \operatorname{span}\mathcal{S}$). By Lemma 1.9,

$$\begin{aligned}\operatorname{span}\{y_1, y_2, \ldots, y_{m-1}, t_m, t_{m+1}, \ldots, t_{q-1}, t_q\} = \\ \operatorname{span}\{y_1, y_2, \ldots, y_{m-1}, t_m, t_{m+1}, \ldots, t_{q-1}, y_m\},\end{aligned}$$

and so with $\mathcal{H} = \{t_m, t_{m+1}, \ldots, t_{q-1}, t_{q+1}, \ldots, q_n\}$, again by Lemma 1.9, we have

$$\operatorname{span}\mathcal{L} \cup \mathcal{H} = \operatorname{span}(\{y_1, y_2, \ldots, y_{m-1}\} \cup \mathcal{H}_0) = \operatorname{span}\mathcal{S}.$$

□

1.11. Corollary. *Let $\{\mathbf{0}\} \neq \mathcal{V}$ be a vector space over a field $\mathbb{F}$ with a finite basis containing $d \in \mathbb{N}$ elements. Then every basis for $\mathcal{V}$ has exactly d elements.*

Proof. Let $\mathcal{B}$ and $\mathcal{D}$ be bases for $\mathcal{V}$, and suppose that $\mathcal{B}$ has d elements. We claim that $\mathcal{D}$ has at most d elements. Indeed, if $\mathcal{D}$ has more than d elements, then we can choose a linearly independent subset $\mathcal{L} = \{y_1, y_2, \ldots, y_{d+1}\} \subseteq \mathcal{D}$.

But then $\mathcal{L} \subseteq \mathcal{V} = \operatorname{span}\mathcal{B}$, and so by Steinitz's Theorem above, $d + 1 < d$, an obvious contradiction. Thus $\mathcal{D}$ has at most d elements. We can therefore say that $\mathcal{D}$ has $m \leq d$ elements.

Interchanging the roles of $\mathcal{B}$ and $\mathcal{D}$, we conclude that $d \leq m$, whence $d = m$.

□

Of course, the only basis for $\mathcal{V} = \{\mathbf{0}\}$ is $\mathcal{B} := \varnothing$, so the above result also applies in this case.

In light of Corollary 1.11, the notion of *dimension* is well-defined.

1.12. Definition. *A vector space $\mathcal{V}$ over a field $\mathbb{F}$ is said to be **finite-dimensional** if it admits a basis $\mathcal{B}$ consisting of a finite number of elements. In this case, by Corollary 1.11, any two bases for $\mathcal{V}$ over $\mathbb{F}$ contain the same number of elements – say n – and we say that the **dimension** of $\mathcal{V}$ **over** $\mathbb{F}$ is n. We write $\dim_{\mathbb{F}} \mathcal{V} = n$, or, if $\mathbb{F}$ is understood, $\dim \mathcal{V} = n$.*

*If $\mathcal{V}$ does not admit a finite basis over $\mathbb{F}$, then we say that $\mathcal{V}$ in **infinite-dimensional**, and we write $\dim_{\mathbb{F}} \mathcal{V} = \infty$.*

1.13. Example.

(a) If $\mathcal{V} = \{\mathbf{0}\}$, then $\mathcal{B} := \varnothing$ is a basis for $\mathcal{V}$, so $\dim \mathcal{V} = 0$.
(b) If $\mathcal{V} = \mathbb{F}^n$ for some $n \geq 1$, then $\dim \mathcal{V} = n$.
(c) The space $\mathbb{F}_n[x]$ is $n+1$-dimensional.
(d) Let $m, n \in \mathbb{N}$. The space $\mathbb{M}_{m\times n}(\mathbb{F})$ is mn-dimensional over $\mathbb{F}$.
(e) Note: *the underlying field is crucial to the notion of dimension!* For example, $\dim_{\mathbb{C}} \mathbb{C} = 1$, while $\dim_{\mathbb{R}} \mathbb{C} = 2$, and $\dim_{\mathbb{Q}} \mathbb{C} = \infty$.

1.14. Corollary. *Let $n \in \mathbb{N}$, and let $\mathcal{V}$ be an n-dimensional vector space over a field $\mathbb{F}$.*

(a) *If $\mathcal{S} \subseteq \mathcal{V}$ is a generating set for $\mathcal{V}$, then $|\mathcal{S}| \geq n$. If $\mathcal{S} = n$, then $\mathcal{S}$ is in fact a basis for $\mathcal{V}$.*
(b) *Is $\mathcal{L} \subseteq \mathcal{V}$ is linearly independent, then $|\mathcal{L}| \leq n$, and if $|\mathcal{L}| = n$, then $\mathcal{L}$ is a basis for $\mathcal{V}$.*
(c) *Every linearly independent subset $\mathcal{L}$ of $\mathcal{V}$ can be extended to a basis for $\mathcal{V}$.*

Proof.

(a) Let $\mathcal{S}$ be a generating set for $\mathcal{V}$, so that $\operatorname{span}\mathcal{S} = \mathcal{V}$. Since $\dim \mathcal{V} = n$, we can find a basis $\mathcal{B} = \{b_1, b_2, \ldots, b_n\}$ for $\mathcal{V}$. By definition, $\mathcal{B}$ is linearly independent, and obviously $\mathcal{B} \subseteq \mathcal{V} = \operatorname{span}\mathcal{S}$.

By Steinitz's Theorem, $\mathcal{S}$ has at least n elements.

Suppose that $|\mathcal{S}| = n$. By Theorem 1.7, there exists a subset $\mathcal{T} \subseteq \mathcal{S}$ which is a basis for $\mathcal{V}$. But then $\mathcal{T}$ must have n elements (as every basis for $\mathcal{V}$ has n elements), which implies that $\mathcal{S} = \mathcal{T}$ is a basis for $\mathcal{V}$.

(b) Let $\mathcal{L} \subseteq \mathcal{V}$ be a linearly independent set, and let $\mathcal{B} = \{b_1, b_2, \ldots, b_n\}$ be a basis for $\mathcal{V}$. Then $\mathcal{B}$ generates $\mathcal{V}$, and so $\mathcal{L} \subseteq \operatorname{span}\mathcal{B}$. By Steinitz's Theorem, $\mathcal{L}$ can have at most n elements.

Furthermore, if $|\mathcal{L}| = n$, then Steinitz's Theorem implies that we can find a subset $\mathcal{H} \subseteq \mathcal{B}$ with $n - n = 0$ elements such that

$$\mathcal{V} = \operatorname{span}\mathcal{B} = \operatorname{span}(\mathcal{L} \cup \mathcal{H}) = \operatorname{span}\mathcal{L}.$$

Thus $\mathcal{L}$ spans $\mathcal{V}$ and $\mathcal{L}$ is linearly independent; i.e. $\mathcal{L}$ is a basis for $\mathcal{V}$.

(c) Let $\mathcal{L}$ be a linearly independent set in $\mathcal{V}$, and let $\mathcal{B} = \{b_1, b_2, \ldots, b_n\}$ be a basis for $\mathcal{V}$. In particular, $\mathcal{B}$ generates $\mathcal{V}$, and so $\mathcal{L} \subseteq \operatorname{span}\mathcal{B}$. By (b), $m := |\mathcal{L}| < n$. By Steinitz's Theorem, we can find a subset $\mathcal{H} \subseteq \mathcal{B}$ such that $\mathcal{H}$ has $n - m$ elements, and

$$\operatorname{span}(\mathcal{L} \cup \mathcal{H}) = \operatorname{span}\mathcal{B} = \mathcal{V}.$$

Thus $\mathcal{L} \cup \mathcal{H}$ is a spanning set for $\mathcal{V}$ and $\mathcal{L} \cup \mathcal{H}$ has exactly n elements. By part (a), $\mathcal{L} \cup \mathcal{H}$ is a basis for $\mathcal{V}$ (which obviously extends $\mathcal{L}$).

□

1.15. Example.

(a) Since $\mathcal{L} := \{(2,1),(1,7)\} \subseteq \mathbb{R}^2$ is linearly independent over $\mathbb{R}$, it is a basis for $\mathbb{R}^2$.

(b) The set $\mathcal{T} := \{(1,6,9,1),(2,1,3,1),(8,8,8,1)\} \subseteq \mathbb{R}^4$ cannot possibly generate $\mathbb{R}^4$, since it contains only three vectors. Since $\mathcal{T}$ is a linearly independent set, it generates a three-dimensional subspace of $\mathbb{R}^4$.

1.16. Theorem. *Let $\mathcal{W}$ be a subspace of a finite-dimensional vector space $\mathcal{V}$ over a field $\mathbb{F}$. Then $\mathcal{W}$ is finite-dimensional, and* $\dim_{\mathbb{F}}\mathcal{W} \leq \dim_{\mathbb{F}}\mathcal{V}$.

Moreover, if $\dim_{\mathbb{F}}\mathcal{W} = \dim_{\mathbb{F}}\mathcal{V}$, *then* $\mathcal{W} = \mathcal{V}$.

Proof. Let $n := \dim_{\mathbb{F}}\mathcal{V}$. If $\mathcal{W} = \{\mathbf{0}\}$, then $\varnothing$ is a basis for $\mathcal{W}$ and so $\dim\mathcal{W} = 0 \leq n$.

Otherwise, we may choose an element $\mathbf{0} \neq w_1 \in \mathcal{W}$. Clearly $\{w_1\}$ is linearly independent. If $\mathcal{W} = \operatorname{span}\{w_1\}$.

- If $\operatorname{span}\{w_1\} = \mathcal{W}$, we stop. Otherwise there exists $w_2 \in \mathcal{W} \setminus \operatorname{span}\{w_1\}$, which implies that $\{w_1, w_2\}$ is linearly independent.
- If $\operatorname{span}\{w_1, w_2\} = \mathcal{W}$, we stop. Otherwise there exists $w_3 \in \mathcal{W} \setminus \operatorname{span}\{w_1, w_2\}$, which implies that $\{w_1, w_2, w_3\}$ is linearly independent.
- More generally, having chosen linearly independent vectors $w_1, w_2, \ldots, w_k \in \mathcal{W}$, if $\operatorname{span}\{w_1, w_2, \ldots, w_k\} = \mathcal{W}$, we stop. Otherwise there exists $w_{k+1} \in \mathcal{W} \setminus \operatorname{span}\{w_1, w_2, \ldots, w_k\}$, which implies that $\{w_1, w_2, \ldots, w_k, w_{k+1}\}$ is linearly independent.

This process must stop after a finite number – say m – of steps, since any linearly independent subset of $\mathcal{W}$ is a linearly independent subset of $\mathcal{V}$, and thus by Corollary 1.14, it can have at most n elements. In particular, $m \leq n$, and $\{w_1, w_2, \ldots, w_m\}$ must span $\mathcal{W}$ (that being the reason why we stopped), implying that $\dim_{\mathbb{F}} \mathcal{W} = m \leq n = \dim_{\mathbb{F}} \mathcal{V}$.

If $m = n$, then $\{w_1, w_2, \ldots, w_n\}$ is a linearly independent set in $\mathcal{W} \subseteq \mathcal{V}$, and hence it must be a basis for $\mathcal{V}$, implying that $\mathcal{W} = \operatorname{span}\{w_1, w_2, \ldots, w_n\} = \mathcal{V}$.

□

1.17. Example.
Let $\mathcal{W} := \{p = p_0 + p_1 x + p_2 x^2 \in \mathbb{R}_2[x] : p_0 + 2p_1 + 3p_2 = 0\}$. Then $\mathcal{W}$ is a subspace of $\mathbb{R}_2[x]$, and

$$\mathcal{B} := \{1 + 4x - 3x^2, 3x - 2x^2\}$$

is a basis for $\mathcal{W}$.

1.18. Example. The set $\mathcal{B} := \{E_{ij}\}_{1 \leq i \leq j \leq n}$ of matrix units in $\mathbb{M}_n(\mathbb{F})$ is a basis for the subspace $\mathbb{T}_n(\mathbb{F})$.

1.19. Proposition. *Let $\mathcal{V}$ be a vector space over a field $\mathbb{F}$, and suppose that $\mathcal{Y}, \mathcal{Z}$ are finite-dimensional subspaces $\mathcal{V}$. Then:*

(a) *$\mathcal{Y} + \mathcal{Z} := \{y + z : y \in \mathcal{Y}, z \in \mathcal{Z}\}$ is a finite-dimensional subspace of $\mathcal{V}$.*
(b) *$\dim(\mathcal{Y} + \mathcal{Z}) = \dim \mathcal{Y} + \dim \mathcal{Z} - \dim(\mathcal{Y} \cap \mathcal{Z})$.*
(c) *Suppose that $\mathcal{V} = \mathcal{Y} + \mathcal{Z}$. Then $\mathcal{V} = \mathcal{Y} \dot{+} \mathcal{Z}$ if and only if $\dim \mathcal{V} = \dim \mathcal{Y} + \dim \mathcal{Z}$.*

Hint. Let $\{u_1, u_2, \ldots, u_k\}$ be a basis for $\mathcal{Y} \cap \mathcal{Z}$, and extend it to bases for $\mathcal{Y}$ and $\mathcal{Z}$ respectively.

Proof.

(a) That $\mathcal{Y} + \mathcal{Z}$ is a subspace is left as a routine exercise.

Let $\mathcal{B}_{\mathcal{Y}} := \{y_1, y_2, \ldots, y_m\}$ be a basis for $\mathcal{Y}$, and $\mathcal{B}_{\mathcal{Z}} := \{z_1, z_2, \ldots, z_n\}$ be a basis for $\mathcal{Z}$. Then

$$\mathcal{Y} + \mathcal{Z} = \operatorname{span}\{y_1, y_2, \ldots, y_m, z_1, z_2, \ldots, z_n\}.$$

That is, $\mathcal{B}_{\mathcal{Y}} \cup \mathcal{B}_{\mathcal{Z}}$ is a finite generating set for $\mathcal{Y} + \mathcal{Z}$. By Theorem 1.7, $\mathcal{B}_{\mathcal{Y}} \cup \mathcal{B}_{\mathcal{Z}}$ contains a basis for $\mathcal{Y} + \mathcal{Z}$. Thus $\dim(\mathcal{Y} + \mathcal{Z}) \leq m + n$.

(b) Now $\mathcal{Y} \cap \mathcal{Z}$ is a subspace of the finite-dimensional subspace $\mathcal{Y} + \mathcal{Z}$, and as such, $\mathcal{Y} \cap \mathcal{Z}$ is itself finite-dimensional. Set $k := \dim(\mathcal{Y} \cap \mathcal{Z})$, and let $\{u_1, u_2, \ldots, u_k\}$ be a basis for $\mathcal{Y} \cap \mathcal{Z}$.

By Corollary 1.14, we may extend $\{u_1, u_2, \ldots, u_k\}$ to a basis

$$\Lambda_{\mathcal{Y}} := \{u_1, u_2, \ldots, u_k, b_1, b_2, \ldots, b_{m-k}\}$$

for $\mathcal{Y}$, and also to a basis

$$\Lambda_{\mathcal{Z}} := \{u_1, u_2, \ldots, u_k, d_1, d_2, \ldots, d_{n-k}\}$$

for $\mathcal{Z}$.

We claim that $\Omega := \Lambda_{\mathcal{Y}} \cup \Lambda_{\mathcal{Z}}$ is a basis $\mathcal{Y} + \mathcal{Z}$. Of course, $\Omega \subseteq \mathcal{Y} + \mathcal{Z}$, and so $\operatorname{span} \Omega \subseteq \mathcal{Y} + \mathcal{Z}$. If $v = y + z \in \mathcal{Y} + \mathcal{Z}$, then $y \in \operatorname{span} \Lambda_{\mathcal{Y}} \subseteq \operatorname{span} (\Lambda_{\mathcal{Y}} \cup \Lambda_{\mathcal{Z}})$ and $z \in \operatorname{span} \Lambda_{\mathcal{Z}} \subseteq \operatorname{span} (\Lambda_{\mathcal{Y}} \cup \Lambda_{\mathcal{Z}})$, whence

$$y + z \in \operatorname{span} (\Lambda_{\mathcal{Y}} \cup \Lambda_{\mathcal{Z}}) = \operatorname{span} \Omega.$$

That is, Ω generates $\mathcal{Y} + \mathcal{Z}$. To see that Ω is linearly independent, suppose that we can find scalars $\alpha_1, \alpha_2, \ldots, \alpha_k$, $\beta_1, \beta_2, \ldots, \beta_{m-k}$ and $\gamma_1, \gamma_2, \ldots, \gamma_{n-k}$ such that

$$\mathbf{0} = \sum_{i=1}^{k} \alpha_i u_i + \sum_{j=1}^{m-k} \beta_j b_j + \sum_{p=1}^{n-k} \gamma_p d_p.$$

Then

$$\sum_{i=1}^{k} \alpha_i u_i + \sum_{j=1}^{m-k} \beta_j b_j = -\left(\sum_{p=1}^{n-k} \gamma_p d_p \right) \in \mathcal{Y} \cap \mathcal{Z},$$

so that

$$\sum_{i=1}^{k} \alpha_i u_i + \sum_{j=1}^{m-k} \beta_j b_j = \delta_1 u_1 + \delta_2 u_2 + \cdots + \delta_k u_k$$

for a unique choice of δ_i, $1 \leq i \leq k$. In particular, $\alpha_i = \delta_i$, $1 \leq i \leq k$, and $\beta_j = 0$, $1 \leq j \leq m - k$.

Hence

$$\mathbf{0} = \sum_{i=1}^{k} \alpha_i u_i + \sum_{p=1}^{n-k} \gamma_p d_p.$$

But $\Lambda_{\mathcal{Z}}$ is a basis for $\mathcal{Z}$, and so $\alpha_i = 0$, $1 \leq i \leq k$ and $\gamma_p = 0$, $1 \leq p \leq n - k$.

Since all of the coefficients are zero, Ω is linearly independent. Since it also spans $\mathcal{Y} + \mathcal{Z}$, it is a basis for $\mathcal{Y} + \mathcal{Z}$. Thus

$$\begin{aligned} \dim (\mathcal{Y} + \mathcal{Z}) &= |\Omega| \\ &= k + (m - k) + (n - k) \\ &= m + n - k \\ &= \dim \mathcal{Y} + \dim \mathcal{Z} - \dim(\mathcal{Y} \cap \mathcal{Z}). \end{aligned}$$

(c) Now $\mathcal{V} = \mathcal{Y} + \mathcal{Z}$ and $\dim \mathcal{V} = \dim \mathcal{Y} + \dim \mathcal{Z} - \dim (\mathcal{Y} \cap \mathcal{Z})$.

- If $\mathcal{V} = \mathcal{Y} \dot{+} \mathcal{Z}$, then $\mathcal{Y} \cap \mathcal{Z} = \{0\}$, so $\dim (\mathcal{Y} \cap \mathcal{Z}) = 0$ and

$$\dim \mathcal{V} = \dim \mathcal{Y} + \dim \mathcal{Z}.$$

- If $\dim \mathcal{V} = \dim \mathcal{Y} + \dim \mathcal{Z}$, then we must have $\dim (\mathcal{Y} \cap \mathcal{Z}) = 0$, so that $\mathcal{Y} \cap \mathcal{Z} = \{0\}$, and therefore $\mathcal{V} = \mathcal{Y} \dot{+} \mathcal{Z}$.

□

1.20. Proposition. *Let $\mathcal{V}$ be a finite-dimensional vector space. If $\mathcal{Y}$ is a subspace of $\mathcal{V}$, then there exists a subspace $\mathcal{Z}$ of $\mathcal{V}$ such that*

$$\mathcal{V} = \mathcal{Y} \dot{+} \mathcal{Z}.$$

Proof. Let $n = \dim \mathcal{V} < \infty$. Since $\mathcal{Y}$ is a subspace of a finite-dimensional space, we see that $\mathcal{Y}$ is finite-dimensional by Theorem 1.16 and $m := \dim \mathcal{Y} \leq n$. Let $\mathcal{B}_\mathcal{Y} := \{b_1, b_2, \ldots, b_m\}$ be a basis for $\mathcal{Y}$. By Corollary 1.14, we can extend $\mathcal{B}_\mathcal{Y}$ to a basis $\mathcal{B}_\mathcal{V} = \{b_1, b_2, \ldots, b_m, z_1, z_2, \ldots, z_{n-m}\}$ for $\mathcal{V}$.

Define $\mathcal{Z} = \operatorname{span}\{z_1, z_2, \ldots, z_{n-m}\}$. Then $\mathcal{V} = \mathcal{Y} + \mathcal{Z}$ and $\dim \mathcal{V} = \dim \mathcal{Y} + \dim \mathcal{Z}$. By Proposition 1.19 above, $\mathcal{V} = \mathcal{Y} \dot{+} \mathcal{Z}$.

□

We say that $\mathcal{Z}$ is a **complement** of $\mathcal{Y}$, and this Proposition says that every subspace of a (finite-dimensional) vector space is complemented. We mention (for the sake of culture) that the same is every subspace of $\mathcal{V}$ is complemented even if $\mathcal{V}$ is infinite-dimensional. Note that the complement need not be unique!!

Continuing with our cultural interlude: in later courses you will discover vector spaces equipped with a **topology** – that is, a notion of "*open*" and "*closed*" sets. At that point, the question might arise as to whether a *closed subspace* of $\mathcal{V}$ is *topologically complemented*, meaning that it admits a complement in the above sense, but that that complement is also "*closed*". This often fails, and makes life both difficult and interesting to those studying Functional Analysis. This marks the end of our cultural interlude. Back to "*reality*".

1.21. Example.

Let $\mathcal{V} = \mathbb{R}^3$ and $\mathcal{Y} := \{x = (x_1, x_2, x_3) \in \mathbb{R}^3 : 2x_1 - x_2 = 0\}$. Then $\mathcal{Y} = \operatorname{span}\{(1,2,0), (1,2,1)\}$, and one complement of $\mathcal{Y}$ is

$$\mathcal{Z}_1 = \operatorname{span}\{(0,1,0)\},$$

while another complement of $\mathcal{Y}$ is

$$\mathcal{Z}_2 = \operatorname{span}\{(1,1,0)\}.$$

In fact, $\mathcal{Y}$ admits infinitely many complements.

1.22. Theorem. *Let $\mathcal{W}$ be a subspace of a finite-dimensional vector space $\mathcal{V}$. Let $\{u_1, u_2, \ldots, u_k\}$ be a basis for $\mathcal{W}$, and extend this to a basis*

$$\mathcal{B} = \{u_1, u_2, \ldots, u_k, u_{k+1}, u_{k+2}, \ldots, u_n\}$$

for $\mathcal{V}$. Then

(a) *$\{u_{k+1} + \mathcal{W}, u_{k+2} + \mathcal{W}, \ldots, u_n + \mathcal{W}\}$ is a basis for $\mathcal{V}/\mathcal{W}$, and*
(b) $\dim \mathcal{V} = \dim \mathcal{W} + \dim (\mathcal{V}/\mathcal{W})$.

Proof.

(a) Let $v \in \mathcal{V}$. Since $\mathcal{B}$ is a basis for $\mathcal{V}$, we can find scalars $\alpha_1, \alpha_2, \ldots, \alpha_n \in \mathbb{F}$ such that

$$v = \alpha_1 u_1 + \alpha_2 u_2 + \cdots + \alpha_n u_n.$$

Thus

$$v + \mathcal{W} = \sum_{j=1}^{n} \alpha_j (u_j + \mathcal{W}).$$

For $1 \leq j \leq k$, $u_j + \mathcal{W} = 0 + \mathcal{W}$, as $u_j \in \mathcal{W}$. Hence

$$v + \mathcal{W} = \sum_{j=k+1}^{n} \alpha_j (u_j + \mathcal{W}).$$

That is, $\{u_{k+1} + \mathcal{W}, u_{k+2} + \mathcal{W}, \ldots, u_n + \mathcal{W}\}$ generates $\mathcal{V}/\mathcal{W}$.

Next, suppose that $\alpha_j \in \mathbb{F}$, $k+1 \leq j \leq n$ and that

$$(\sum_{j=k+1}^{n} \alpha_j u_j) + \mathcal{W} = \sum_{j=k+1}^{n} \alpha_j (u_j + \mathcal{W}) = \mathbf{0} + \mathcal{W}.$$

Then

$$(\sum_{j=k+1}^{n} \alpha_j u_j) \in \mathcal{W} = \operatorname{span}\{u_1, u_2, \ldots, u_k\},$$

say

$$(\sum_{j=k+1}^{n} \alpha_j u_j) = \sum_{i=1}^{k} \beta_i u_i$$

for an appropriate choice of scalars $\beta_1, \beta_2, \ldots, \beta_k$. Since $\mathcal{B}$ is linearly independent, we conclude that $\alpha_j = 0 = \beta_i$, $1 \leq i \leq k$, $k+1 \leq j \leq n$.

Thus $\{u_{k+1} + \mathcal{W}, u_{k+2} + \mathcal{W}, \ldots, u_n + \mathcal{W}\}$ is linearly independent, and so it is a basis for $\mathcal{V}/\mathcal{W}$.

(b) By part (a),

$$\dim \mathcal{V} = n = k + (n-k) = \dim \mathcal{W} + \dim (\mathcal{V}/\mathcal{W}).$$

□

2. Infinite-dimensional vector spaces

2.1. A number of the results we have obtained regarding linear dependence and independence did not require our vector space to be finite-dimensional. For example, we saw in Theorem 3.2.7 that if $\mathcal{V}$ is a vector space, $\mathcal{L} \subseteq \mathcal{V}$ is linear independent and $x \in \mathcal{V} \setminus \mathcal{L}$, then $\mathcal{L} \cup \{x\}$ is linearly dependent if and only if $x \in \operatorname{span} \mathcal{L}$.

So far we have shown that every finite-dimensional vector space $\mathcal{V}$ over a field $\mathbb{F}$ admits a (Hamel) basis, and that any two such bases have the same **cardinality** (i.e. in this setting, we simply mean "*the same number of elements*". The proof that an arbitrary infinite-dimensional vector space admits a basis is much, much deeper. In fact, the proof requires a new axiom, independent of the usual **Zermelo-Fraenkel Axioms** of set theory. The concept we have in mind is **Zorn's Lemma**, which we developed in Chapter 1. Let's (finally) put it to good use.

2.2. Theorem. *Let $\mathcal{V}$ be a vector space over a field $\mathbb{F}$. Then $\mathcal{V}$ admits a basis.*

Proof. Let $\Omega = \{L \subseteq \mathcal{V} : L \text{ is linearly independent}\}$. Since $\varnothing \in \Omega$, $\Omega \neq \varnothing$.

Let $\mathcal{C} = \{L_\lambda\}_{\lambda\in\Lambda}$ be a chain in Ω, where Ω is ordered by inclusion.

Let

$$L := \cup_{\lambda\in\Lambda} L_\lambda.$$

Clearly, if L is linearly independent, then L will be an upper bound for $\mathcal{C}$ in Ω. Thus we must prove that L is linearly independent. We argue by contradiction.

Suppose otherwise. Then there exist vectors $x_1, x_2, ..., x_n \in L$ and scalars $\kappa_1, \kappa_2, ..., \kappa_n \in \mathbb{F}$ (not all equal to zero) such that $\sum_{j=1}^{n} k_j x_j = 0$. Now, each $x_j \in L_{\lambda_j}$ for some $\lambda_j \in \Lambda$. Since $\mathcal{C}$ is a chain, there exists $m \in \{1, 2, ..., n\}$ so that $x_k \in L_{\lambda_k} \subseteq L_{\lambda_m}$ for all $1 \leq k \leq n$. But then $\{x_1, x_2, ..., x_n\} \subseteq L_{\lambda_m}$ and so L_{λ_m} is linearly dependent, a contradiction.

Thus L is linearly independent, and so $L \in \Omega$ is an upper bound for $\mathcal{C}$.

By Zorn's Lemma, there exists a maximal element $M \in \mathcal{F}$. Suppose that $\operatorname{span} M \neq \mathcal{V}$. Then there exists $0 \neq y \in \mathcal{V}$ but $y \notin \operatorname{span} M$. But then $M \cup \{y\}$ is linearly independent, and $M < M \cup \{y\}$, contradicting the maximality of M.

Hence $\operatorname{span} M = \mathcal{V}$, and $M \in \Omega$ implies M is linearly independent, so M is a basis for $\mathcal{V}$.

□

2.3. Remark. In fact, it can be shown that if $\mathcal{B}_1$ and $\mathcal{B}_2$ are two bases for a given vector space $\mathcal{V}$ over a field $\mathbb{F}$, then there exists a bijection between $\mathcal{B}_1$ and $\mathcal{B}_2$. We say that $\mathcal{B}_1$ and $\mathcal{B}_2$ have the same **cardinality**. This allows us to define the notion of dimension of an infinite-dimensional space as the cardinality of any one of its bases.

2.4. Remark. A much deeper result than Theorem 2.2 above is that the Axiom of Choice is *equivalent* (over the Zermelo-Fraenkel Axioms) to the statement that every vector space admits a basis. This is due to Blass [**Bla84**], and the proof is beyond the scope of this course.

The following result is also known to be equivalent to the Axiom of Choice.

- Let $\mathcal{V}$ be a vector space over the field $\mathbb{F}$ and suppose that $\mathcal{L} \subseteq \mathcal{S} \subseteq \mathcal{V}$, where $\mathcal{L}$ is a linearly independent subset of $\mathcal{V}$, and $\operatorname{span} \mathcal{S} = \mathcal{V}$. Then there exists a basis $\mathcal{B}$ for $\mathcal{V}$ satisfying $\mathcal{J} \subseteq \mathcal{B} \subseteq \mathcal{S}$.

Supplementary Examples

S4.1. Example. The Lagrange Interpolation Formula. From an abstract point of view, all bases for a given vector space were created equal. On the other hand, when dealing with a specific vector space, there may exist some basis or bases which are, as the saying goes, "*more equal than others*". By this we mean that a certain basis may reduce the number or the complexity of computations required to resolve certain problems. This is a vague statement which is worth demystifying with an example.

Consider the case of the vector space $\mathbb{F}_n[x]$, where $n \in \mathbb{N}$ and $\mathbb{F}$ is infinite.

Let $\{\theta_0, \theta_1, \theta_2, \ldots, \theta_n\}$ be distinct scalars in an infinite field $\mathbb{F}$. Define the polynomials:

$$p_i(x) \coloneqq \prod_{\substack{k=0 \\ k \neq i}}^{n} (x - \theta_k)(\theta_i - \theta_k)^{-1}.$$

Observe that for $1 \leq j \leq n$,

$$f_i(\theta_j) \coloneqq \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if} i \neq j. \end{cases}$$

The elements of $\mathcal{L} \coloneqq \{p_0, p_1, p_2, \ldots, p_n\}$ are referred to as **Lagrange polynomials**, and it is not hard to see that they lie in $\mathbb{F}_n[x]$.

We claim that $\mathcal{L}$ is linearly independent. If we can show this, then the fact that $|\mathcal{L}| = n + 1 = \dim \mathbb{F}_n[x]$ implies that $\mathcal{L}$ must be a basis for $\mathbb{F}_n[x]$.

To that end, suppose that $\alpha_0, \alpha_1, \ldots, \alpha_n \in \mathbb{F}$ and that

$$\sum_{i=0}^{n} \alpha_i p_i = 0.$$

It then follows that for each $0 \leq j \leq n$,

$$\alpha_j = \sum_{i=0}^{n} \alpha_i p_i(\theta_j) = 0,$$

from which our claim immediately follows.

Now suppose that $g \in \mathbb{F}_n[x]$ is an arbitrary element. We would like to express g as a linear combination of our basis $\mathcal{L}$. Normally, this can be an involved process. However, thanks to our judicious choice of basis, we find that if $g = \sum_{i=0}^{n} \beta_i p_i$, then for each $0 \leq j \leq n$,

$$\beta_j = \sum_{i=0}^{n} \beta_i p_i(\theta_j) = g(\theta_j),$$

and thus

$$g = \sum_{i=0}^{n} g(\theta_i) p_i.$$

This is called the **Lagrange Interpolation Formula**, because it tells us that g is the unique function in $\mathbb{F}_n[x]$ which takes on the values $g(\theta_j)$ at θ_j, $0 \le j \le n$.

S4.2. Example. Let us find the unique polynomial $g \in \mathbb{R}_2[x]$ satisfying $g(0) = 1$, $g(1) = 2$ and $g(2) = 5$.

Following the analysis above, we consider $\theta_0 = 0$, $\theta_1 = 1$ and $\theta_2 = 2$. We then define

$$\begin{aligned} p_0(x) &= (x-1)(0-1)^{-1} \cdot (x-2)(0-2)^{-1} = \frac{1}{2}(x-1)(x-2); \\ p_1(x) &= (x-0)(1-0)^{-1} \cdot (x-2)(1-2)^{-1} = -(x)(x-2) \\ p_2(x) &= (x-0)(2-0)^{-1} \cdot (x-1)(2-1)^{-1} = \frac{1}{2}(x)(x-1). \end{aligned}$$

Remark. On an exam, it would be worth checking that $p_0(\theta_j) = \delta_{i,j}$, where $\delta_{i,j}$ denotes the **Kronecker delta** function, i.e. $\delta_{i,j} := \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j. \end{cases}$

Arguing as above,

$$\begin{aligned} g &= \sum_{i=0}^{2} g(\theta_i) p_i \\ &= g(\theta_0)p_0 + g(\theta_1)p_1 + g(\theta_2)p_2 \\ &= g(0)p_0 + g(1)p_1 + g(2)p_2 \\ &= 1p_0 + 2p_1 + 5p_2 \\ &= \frac{1}{2}(x^2 - 3x + 2) + 2(-x^2 + 2x) + 5(\frac{1}{2}x^2 - \frac{1}{2}x). \end{aligned}$$

S4.3. Example. Note that if $\mathbb{F}$ is an infinite field and if $\theta_0, \theta_1, \theta_2, \ldots, \theta_n$ are $(n+1)$ distinct elements of $\mathbb{F}$, then by the Lagrange Interpolation Formula, the only polynomial q of degree at most n which satisfies $q(\theta_j) = 0$ for all $0 \le j \le n$ is the zero polynomial.

The (easy) computation is left to the reader.

Of course - there exists a non-zero polynomial $r(x) = r_0 + r_1 x + \cdots + r_{n+1}x^{n+1}$ satisfying $r(\theta_j) = 0$, $0 \le j \le n$. (Why can we say "*of course*"?)

S4.4. Example. Recall the vector space $\mathcal{V} = \{\alpha \,\text{PIG} + \beta \,\text{DOG} : \alpha, \beta \in \mathbb{K}\}$ from Example 2.1.7. Then $\{\text{PIG}, \text{DOG}\}$ is a basis for $\mathcal{V}$ over $\mathbb{K}$!

In the case where $\mathbb{K} = \mathbb{C}$, note that $\mathcal{B}_1 := \{\text{PIG}, i\text{PIG}, \text{DOG}, i\text{DOG}\}$ is a basis for $\mathcal{V}$ as a vector space over $\mathbb{R}$, as is $\mathcal{B}_2 := \{\text{PIG}, (1+i)\text{PIG}, -\text{DOG}, (3+ei)\text{DOG}\}$.

S4.5. Example. Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces over a field $\mathbb{F}$ and suppose that $\mathcal{B}_1$ (resp. $\mathcal{B}_2$) is a basis for $\mathcal{V}$ (resp. $\mathcal{W}$). Let

$$\mathcal{B} := \{(x,0),(0,y) : x \in \mathcal{B}_1, y \in \mathcal{B}_2\}.$$

We claim that $\mathcal{B}$ is a basis for the vector space $\mathcal{V}\times\mathcal{W}$. To see this, we must show that $\mathcal{B}$ spans $\mathcal{V}\times\mathcal{W}$ and that $\mathcal{B}$ is linearly independent.

- Let $(v,w) \in \mathcal{V}\times\mathcal{W}$. Choose $\alpha_i \in \mathbb{F}$, $x_i \in \mathcal{B}_1$, $1 \le i \le m$ such that $v = \sum_{i=1}^m \alpha_i x_i$. Similarly, choose $\beta_j \in \mathbb{F}, y_j \in \mathcal{B}_2$, $1 \le j \le n$ such that $w = \sum_{j=1}^n \beta_j y_j$. Then

$$\begin{aligned}(v,w) &= (v,0) + (0,w)\\ &= (\sum_{i=1}^m \alpha_i x_i, 0) + (0, \sum_{j=1}^n \beta_j y_j)\\ &= \sum_{i=1}^m \alpha_i (x_i,0) + \sum_{j=1}^n \beta_j (0,y_j) \in \operatorname{span}\mathcal{B}.\end{aligned}$$

- To see that $\mathcal{B}$ is linearly independent, suppose that $(x_i,0)$, $(0,y_j) \in \mathcal{B}$, $1 \le i \le m$, $1 \le j \le n$ are distinct vectors and that there exist $\alpha_i, \beta_j \in \mathbb{F}$, $1 \le i \le m$, $1 \le j \le n$ such that

$$\sum_{i=1}^m \alpha_i (x_i,0) + \sum_{j=1}^n \beta_j (0,y_j) = (\sum_{i=1}^m \alpha_i x_i, \sum_{j=1}^n \beta_j y_j) = (0,0).$$

  Then $\sum_{i=1}^m \alpha_i x_i = 0 = \sum_{j=1}^n \beta_j y_j$. Since each of the sets $\{x_1, x_2, \ldots, x_m\}$ and $\{y_1, y_2, \ldots, y_n\}$ is linearly independent, this implies that $\alpha_i = \beta_j = 0$ for all i and j, proving that $\mathcal{B}$ is linearly independent.

Thus $\mathcal{B}$ is a basis for $\mathcal{V}\times\mathcal{W}$.

Question. Is $\mathcal{C} := \{(b_1,b_2) : b_1 \in \mathcal{B}_1, b_2 \in \mathcal{B}_2\}$ also a basis for $\mathcal{V}\times\mathcal{W}$?

S4.6. Example. Let $\mathcal{V}$ be a vector space over a field $\mathbb{F}$ and let $x, y \in \mathcal{V}$ be linearly independent vectors. Let $\mathcal{J} := \{x+y, x-y\}$. It is interesting to ask whether or not $\mathcal{J}$ is linearly independent.

Suppose that $\alpha, \beta \in \mathbb{F}$ and $\alpha(x+y) + \beta(x-y) = 0$. Then $(\alpha+\beta)x + (\alpha-\beta)y = 0$, and since $\{x,y\}$ is linearly independent,

$$\alpha + \beta = 0 = \alpha - \beta.$$

From this we see that $2\beta = 0$, and we might be tempted to conclude that $\beta = 0$. This, however, fails miserably if $\operatorname{char}(\mathbb{F}) = 2$ - e.g. if $\mathbb{F} = \mathbb{Z}_2$. In fact, if $\operatorname{char}(\mathbb{F}) = 2$, then $1 = -1$ in $\mathbb{F}$, and so $x + y = x - y$, and therefore $\{x+y, x-y\}$ is linearly *dependent*.

If $\operatorname{char}(\mathbb{F}) \neq 2$, then $2\beta = 0$ implies that $\beta = 0$, whence $\alpha = 0$, and then $\{x+y, x-y\}$ is linearly *independent*.

This highlights the important fact that in dealing with vector spaces, the case where the characteristic of the underlying field is 2 must often be treated separately.

S4.7. Example. Let $(\mathbb{M}_2(\mathbb{C}))_{sa} := \{T = [t_{ij}] \in \mathbb{M}_2(\mathbb{C}) : t_{ij} = \overline{t_{ji}}, 1 \leq i, j \leq 2\}$. Recall that an element of $(\mathbb{M}_2(\mathbb{C}))_{sa}$ is said to be a **self-adjoint matrix**, also known as an **hermitian matrix**.

We claim that $(\mathbb{M}_2(\mathbb{C}))_{sa}$ is a *vector space over* $\mathbb{R}$.

Note that if $T = [t_{ij}]$, $R = [r_{ij}] \in (\mathbb{M}_2(\mathbb{C}))_{sa}$ and $\kappa \in \mathbb{R}$, then $\kappa T + R = [\kappa t_{ij} + r_{ij}]$, and

$$\overline{\kappa t_{ji} + r_{ji}} = \overline{\kappa}\overline{t_{ji}} + \overline{r_{ji}} = \kappa t_{ij} + r_{ij}.$$

Since the zero matrix clearly lies in $(\mathbb{M}_2(\mathbb{C}))_{sa}$, it is non-empty, and hence is a subspace of $\mathbb{M}_2(\mathbb{C})$, viewed as a vector space over $\mathbb{R}$.

We invite the reader to verify that

- $\dim_{\mathbb{R}}(\mathbb{M}_2(\mathbb{C})) = 8$;
- $\{E_{11}, E_{22}, E_{11} + E_{22}, iE_{12} - iE_{21}\}$ is a basis for $(\mathbb{M}_2(\mathbb{C}))_{sa}$ over $\mathbb{R}$, and so
- $\dim_{\mathbb{R}}(\mathbb{M}_2(\mathbb{C}))_{sa} = 4$.

Question. Is $(\mathbb{M}_2(\mathbb{C}))_{sa}$ a vector space over $\mathbb{C}$?

S4.8. Example. Let $\mathcal{V}$ be a vector space over a field $\mathbb{F}$ and suppose that $\mathcal{Y}, \mathcal{Z}$ are subspaces of $\mathcal{V}$. We claim that that the following are equivalent:

(a) $\mathcal{Y} \cap \mathcal{Z} = \{0\}$.
(b) For all $0 \neq y \in \mathcal{Y}$, $0 \neq z \in \mathcal{Z}$, the set $\{y, z\}$ is linearly independent in $\mathcal{V}$.

Proof.

(a) implies (b). We argue by contradiction. Suppose that there exist $0 \neq y \in \mathcal{Y}$ and $0 \neq z \in \mathcal{Z}$ such that $\{y, z\}$ is linearly *dependent*. Choose $\alpha, \beta \in \mathbb{F}$, not both equal to zero, such that $\alpha y + \beta z = 0$.

If $\alpha \neq 0$, then $\alpha y \neq 0$, whence $\beta z \neq 0$, and thus $\beta \neq 0$. Thus $y = \alpha^{-1}\beta z \in \mathcal{Y} \cap \mathcal{Z}$, contradicting the assumption that $\mathcal{Y} \cap \mathcal{Z} = \{0\}$. This completes this step.

(b) implies (a). Again, we argue by contradiction. Suppose that there exists a non-zero vector w which lies in $\mathcal{Y} \cap \mathcal{Z}$. Set $y = w = z$. Then $1y + (-1)z = 0$, so $\{y, z\}$ is linearly dependent.

□

S4.9. Example. Let $\{p, q\} \subseteq \mathbb{F}[x]$ be linearly independent polynomials. Let us prove that if $\min(\deg p, \deg q) \geq 1$, then $\{p, q, pq\}$ is also linearly independent.

Write $p(x) = p_0 + p_1 x + \cdots + p_n x^n$ where $p_n \neq 0$ (note that $n \geq 1$ by hypothesis), and similarly, write $q(x) = q_0 + q_1 x + \cdots q_m x^m$ where $q_m \neq 0$ (and again, $m \geq 1$). Then, for an appropriate choice of $r_j \in \mathbb{F}$, $0 \leq j \leq m + n$, we have that

$$pq(x) = r_0 + r_1 x + \cdots + r_{m+n} x^{m+n},$$

and $r_{m+n} = p_n q_m \neq 0$. Let $\alpha, \beta, \gamma \in \mathbb{F}$ and suppose that

$$\alpha p + \beta q + \gamma(pq) = 0.$$

By considering the coefficient of x^{m+n} (and keeping in mind that $m+n > \max(m,n)$ as $m, n \geq 1$), we see that $\alpha 0 + \beta 0 + \gamma r_{m+n} = 0$. But $r_{m+n} \neq 0$ then implies that $\gamma = 0$. Hence

$$\alpha p + \beta q = 0.$$

Since $\{p, q\}$ is linearly independent by hypothesis, this implies that $\alpha = 0 = \beta$, which in turn shows that $\{p, q, pq\}$ is linearly independent.

What happens if $\min(\deg p, \deg q) < 1$?

S4.10. Example. Let $\mathcal{V}$ be a vector space over a field $\mathbb{F}$ and suppose that $\mathcal{L} = \{x_1, x_2, x_3\} \subseteq \mathcal{V}$ is linearly independent. We claim that $\mathcal{M} := \{x_1, x_1+x_2, x_1+x_2+x_3\}$ is also linearly independent.

Indeed, if $\alpha, \beta, \gamma \in \mathbb{F}$ and $\alpha x_1 + \beta(x_1 + x_2) + \gamma(x_1 + x_2 + x_3) = 0$, then

$$(\alpha + \beta + \gamma)x_1 + (\beta + \gamma)x_2 + \gamma x_3 = 0.$$

Since $\{x_1, x_2, x_3\}$ is linearly independent (by hypothesis), this force $(\alpha + \beta + \gamma) = (\beta + \gamma) = \gamma = 0$. But then $\alpha = \beta = \gamma = 0$, prove that $\mathcal{M}$ is linearly independent.

Such an argument can clearly be extended to more than three vectors.

Appendix

A4.1. The notion of linear dependence of a set $\mathcal{L}$ of vectors in a vector space $\mathcal{V}$ indicates *redundancy*. If $\mathcal{L}$ is linearly dependent, then one can remove at least one vector – say x – from $\mathcal{L}$ to obtain a set $\mathcal{M} = \mathcal{L} \setminus \{x\}$ such that $\operatorname{span}\mathcal{L} = \operatorname{span}\mathcal{M}$. Of course, this doesn't mean that the choice of x is arbitrary, just that such an x exists in $\mathcal{L}$. Of course, $\mathcal{L}$ might or might not generate the whole space - that is not the issue.

The idea that a set $\mathcal{S}$ in $\mathcal{V}$ generates or spans $\mathcal{V}$ means that one can achieve any vector in $\mathcal{V}$ using a *finite* linear combination of vectors from $\mathcal{S}$. If $\mathcal{S}$ spans $\mathcal{V}$, it might have built-in redundancy, or it might not.

When dealing with bases, one asks that the set $\mathcal{B}$ be large enough to span the entire space, yet small enough to not include any redundancy. As Goldilocks would say, $\mathcal{B}$ has to be *just right*.

A4.2. It is *crucial* to keep in mind that we only consider *finite* linear combinations of vectors in a vector space. What would an infinite linear combination even mean? For those of you who have seen series in Calculus, you might be thinking that one considers sums of the form $\sum_{n=1}^{\infty} x_n$, where $x_n \in \mathbb{R}$, so why can't one do this here?

The answer is that in mathematics, one *never* considers infinite sums. The notation $\sum_{n=1}^{\infty} x_n$ is misleading in that sense. Recall that we say that a sequence $(x_n)_n \in \mathbb{R}^{\mathbb{N}}$ is said to be **summable** if

$$\alpha := \lim_{N\to\infty} \sum_{n=1}^{N} x_n$$

exists in $\mathbb{R}$, in which case we denote α by $\sum_{n=1}^{\infty} x_n$. As such, we are *not* adding infinitely many terms of the sequence, but rather *adding finitely many terms*, and *then taking limits* of the sequence of partial sums we obtain in that manner.

In order to be able to do this in a vector space context, one would need to have a notion of convergence. The technical term for this is to say that one needs to define a **topology** on the vector space. Many, many, many interesting vector spaces admit interesting topologies, and so in those spaces, we can consider limits of finite sums of vectors. But the fact remains the same – we never consider sums of infinitely many vectors, just limits of sums of finitely many vectors at a time.

A similar comment applies to verifying whether a subset $\mathcal{S}$ of a vector space $\mathcal{V}$ is linearly independent. To verify this, one must verify that every *finite* subset of $\mathcal{S}$ is linearly independent – *one can never take infinite linear combinations of vectors in* $\mathcal{S}$, even if $\mathcal{S}$ is infinite.

A4.3. In future courses, you may come across vector spaces imbued with a topology, and the word "basis" may have a different meaning there. For example, a *Hilbert space* $\mathcal{H}$ is a complete inner product space, and a **Hilbert space basis** for $\mathcal{H}$ is a *maximal orthonormal set* in $\mathcal{H}$. (Hilbert spaces will be dealt with in future

courses.) When $\mathcal{H}$ is infinite-dimensional, a Hilbert space basis is never a vector space basis (i.e. a Hamel basis). Having said that, Hilbert space bases tend to prove more useful than Hamel bases to study Hilbert spaces and the operators upon them.

But to repeat what was said in the main text: since the only type of bases we deal with *in this course* are Hamel bases, we drop the adjective "Hamel" to improve the readability of the text.

Exercises for Chapter 4

Exercise 4.1. Let $\mathcal{V}$ be a vector space over a field $\mathbb{F}$ and let $\mathcal{B} \subseteq \mathcal{V}$. Prove that the following are equivalent:

(a) $\mathcal{B}$ is a basis for $\mathcal{V}$ – i.e. it is a maximal linearly independent subset of $\mathcal{V}$.
(b) $\mathcal{B}$ is linearly independent, and $\operatorname{span} \mathcal{B} = \mathcal{V}$.

Exercise 4.2. Let $\mathcal{V} = \mathcal{C}([0,1], \mathbb{R})$. Determine which sets of functions in $\mathcal{V}$ form a linearly independent set.

(a) $f_1(x) = 3x$; $f_2(x) = x + 5$; $f_3(x) = 2x^2$; $f_4(x) = (x+1)^2$.
(b) $f_1(x) = (x+1)^2$; $f_2(x) = x^2 - 1$; $f_3(x) = 2x^2 + 2x - 3$.
(c) $f_1(x) = 1$; $f_2(x) = e^x$; $f_3(x) = e^{-x}$.
(d) $f_1(x) = 1 - x$; $f_2(x) = x(1-x)$; $f_3(x) = 1 - x^2$.

Exercise 4.3. Let $\mathcal{V}$ be a vector space over $\mathbb{C}$, and suppose that $x, y, z \in \mathcal{V}$ are linearly independent. Prove that $(x+y), (y+z)$, and $(x+z)$ are also linearly independent.

Exercise 4.4. Let $\mathbb{F}$ be a field and $\mathcal{V} = \mathbb{M}_2(\mathbb{F})$. Let $\mathcal{Y}$ denote the set of matrices in $\mathcal{V}$ of the form

$$\begin{bmatrix} x & -x \\ y & z \end{bmatrix},$$

and let $\mathcal{Z}$ denote the set of matrices in $\mathcal{V}$ of the form

$$\begin{bmatrix} a & b \\ -a & c \end{bmatrix}.$$

(a) Prove that $\mathcal{Y}$ and $\mathcal{Z}$ are subspaces of $\mathcal{V}$.
(b) Find $\dim \mathcal{Y}$, $\dim \mathcal{Z}$, $\dim(\mathcal{Y} + \mathcal{Z})$ and $\dim(\mathcal{Y} \cap \mathcal{Z})$.

Exercise 4.5. Let $\mathcal{V}$ be a vector space and $\mathcal{Y}, \mathcal{Z}$ be subspaces of $\mathcal{V}$.

(a) If $\mathcal{S}_1$ (resp. $\mathcal{S}_2$) is a spanning set for $\mathcal{Y}$ (resp. for $\mathcal{Z}$), is $\mathcal{S}_1 \cap \mathcal{S}_2$ a spanning set for $\mathcal{Y} \cap \mathcal{Z}$?
(b) If $\mathcal{B}_1$ (resp. $\mathcal{B}_2$) is a basis for $\mathcal{Y}$ (resp. for $\mathcal{Z}$), is $\mathcal{B}_1 \cup \mathcal{S}_2$ a basis for $\mathcal{Y} + \mathcal{Z}$?

Exercise 4.6. Let $\mathcal{V}$ be a vector space and $\mathcal{Y}, \mathcal{Z}$ be subspaces of $\mathcal{V}$. Suppose that $\mathcal{Y} \cap \mathcal{Z} = \{0\}$ and that $\mathcal{V} = \mathcal{Y} + \mathcal{Z} := \{y + z : y \in \mathcal{Y}, z \in \mathcal{Z}\}$.

Prove that for all $x \in \mathcal{V}$, there exist *unique* vectors $y \in \mathcal{Y}$, $z \in \mathcal{Z}$ such that $x = y + z$.

Exercise 4.7. Let $\mathcal{V} = \mathcal{C}([0,1], \mathbb{R})$. Let p_0 denote the constant function $p_0(x) = 1$, $x \in [0,1]$, and for $n \in \mathbb{N}$, let $p_n(x) = x^n$, $x \in [0,1]$.

Prove that for all $N \in \mathbb{N}$, the set $\mathcal{L}_N := \{p_0, p_1, \ldots, p_N\}$ is linearly independent.

Exercise 4.8. (Culture) Prove that $\dim_{\mathbb{Q}} \mathbb{R} = \infty$.

Exercise 4.9. Let $n \in \mathbb{N}$, and recall that $\mathbb{T}_n(\mathbb{F})$ denotes the set of $n \times n$ upper-triangular matrices over $\mathbb{F}$. Find $\dim \mathbb{T}_n(\mathbb{F})$.

Exercise 4.10. Find three vectors $x, y, z \in \mathbb{R}^3$ such that $\{x, y, z\}$ is linearly dependent, but each of $\{x, y\}$, $\{y, z\}$ and $\{x, z\}$ is linearly *independent*.

CHAPTER 5

Linear transformations and matrices

A cowboy asked me if I could help him round up 18 cows. I said, "Yes, of course. That's 20 cows."

Jake Lambert

1. Linear maps

1.1. Mathematics is the study of mathematical objects and the relationships between them. The relationships between them are determined by maps from each object to the other. When two mathematical objects share a common structure – for example, when they are both groups, rings, topological spaces or, as is of interest to us presently, vector spaces over the same field – it is both natural and useful to consider maps that respect that structure.

In the case of vector spaces $\mathcal{V}$ and $\mathcal{W}$ over a common field $\mathbb{F}$, we know that the vector space structure implies that both $\mathcal{V}$ and $\mathcal{W}$ admit a binary operation called "addition", as well as a binary operation on $\mathbb{F}\times\mathcal{V}$ (resp. $\mathbb{F}\times\mathcal{W}$) called "scalar multiplication". For this reason, the most important maps between them are maps which respect addition and scalar multiplication. We shall call such maps *linear*.

1.2. Definition. *Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces over the field $\mathbb{F}$. A function $T:\mathcal{V}\to\mathcal{W}$ is said to be* ***linear*** *if for all $\kappa\in\mathbb{F}$ and $x,y\in\mathcal{V}$,*

$$T(\kappa x+y)=\kappa Tx+Ty.$$

We denote the set of linear maps from $\mathcal{V}$ to $\mathcal{W}$ by $\mathcal{L}(\mathcal{V},\mathcal{W})$.

1.3. Remark. We leave it to the reader to prove that the condition above defining linearity is equivalent to either of the following two conditions:

(a) $T(\kappa x)=\kappa T$ and $T(x+y)=Tx+Ty$ for all $\kappa\in\mathbb{F}$, $x,y\in\mathcal{V}$.
(b) For all $n\geq 1$, $x_1,x_2,\ldots,x_n\in\mathcal{V}$ and $\kappa_1,\kappa_2,\ldots,\kappa_n\in\mathbb{F}$ we have

$$T(\sum_{j=1}^{n}\kappa_j x_j)=\sum_{j=1}^{n}\kappa_j Tx_j.$$

In particular,

$$T(\mathbf{0}_{\mathcal{V}}) = T(0 \cdot \mathbf{0}_{\mathcal{V}}) = 0(T\,\mathbf{0}_{\mathcal{V}}) = \mathbf{0}_{\mathcal{W}}$$

whenever T is linear.

1.4. Example.

(a) The map

$$\begin{array}{rccc} T_1: & \mathbb{R}^2 & \to & \mathbb{R}^2 \\ & (x,y) & \mapsto & (2x+3y, y) \end{array}$$

is linear.

(b) The map

$$\begin{array}{rccc} T_2: & \mathbb{R}^2 & \to & \mathbb{R}^2 \\ & (x,y) & \mapsto & (x^2+y, 0) \end{array}$$

is *not* linear.

(c) Let $\mathcal{V} = \mathcal{C}([0,1], \mathbb{R})$ and $\mathcal{W} = \mathbb{R}$. Define

$$\begin{array}{rccc} T_3: & \mathcal{V} & \to & \mathcal{W} \\ & f & \mapsto & \int_0^1 f(t)dt. \end{array}$$

Then T_3 is linear, as is the map

$$\begin{array}{rccc} T_4: & \mathcal{V} & \to & \mathcal{V} \\ & [T_4 f](x) & := & \int_0^x f(t)dt, \quad x \in [0,1]. \end{array}$$

(d) Let $\mathcal{V} = \mathbb{R}[x]$, the set of polynomials in x, considered as a subspace of $\mathcal{C}([0,1], \mathbb{R})$. The map

$$\begin{array}{rccc} T_5: & \mathcal{V} & \to & \mathcal{V} \\ & f & \mapsto & f' \end{array}$$

is linear. That is, differentiation is linear on the space of polynomials.

(e) Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces over the field $\mathbb{F}$. The map $Zx = \mathbf{0}_{\mathcal{W}}$ for all $x \in \mathcal{V}$ is linear, ie. $Z \in \mathcal{L}(\mathcal{V}, \mathcal{W})$. We refer to this as the **zero map**. .

Similarly, the map $I : \mathcal{V} \to \mathcal{V}$ defined by $Ix = x$ for all $x \in \mathcal{V}$ is linear. We refer to this as the **identity map**.

If $\mathcal{W} \subseteq \mathcal{V}$ is a subspace of $\mathcal{V}$, we refer to the map $\iota : \mathcal{W} \to \mathcal{V}$ defined by $\iota w = w$ for all $w \in \mathcal{W}$ as the **inclusion map**. Note that $\iota = I$ if and only if $\mathcal{W} = \mathcal{V}$.

1.5. Example. Let $\mathcal{V} = \mathbb{C}[x]$. We define two operators D and $M_x \in \mathcal{L}(\mathcal{V})$ as follows:

$$D(p_0 + p_1 x + p_2 x^2 + \cdots + p_n x^n) := p_1 + 2p_2 x + 3p_3 x^2 + \cdots + n p_n x^{n-1},$$

(this just looks like differentiation), and

$$M_x(p_0 + p_1 x + p_2 x^2 + \cdots + p_n x^n) := p_0 x + p_1 x^2 + \cdots + p_n x^{n+1},$$

so that M_x is *multiplication* by x).

We leave it to the reader to verify that these are indeed linear maps.

Observe that

$$\begin{aligned}(DM_x - M_xD)(p_0 + p_1x + p_2x^2 + \cdots + p_nx^n) &= D(p_0x + p_1x^2 + \cdots + p_nx^{n+1}) - M_x(p_1 + 2p_2x + 3p_3x^2 + \cdots + np_nx^{n-1}) \\ &= (p_0 + 2p_1x + 3p_2x^2 + \cdots + (n+1)p_nx^n) - (p_1x + 2p_2x^2 + \cdots + np_nx^n) \\ &= p_0 + p_1x + p_2x^2 + \cdots + p_nx^n \\ &= I(p_0 + p_1x + p_2x^2 + \cdots + p_nx^n).\end{aligned}$$

Since $p = p_0 + p_1x + p_2x^2 + \cdots + p_nx^n \in \mathbb{C}[x]$ was arbitrary, $DM_x - M_xD = I$.

Culture. This example is not as innocuous as it might look. A formulation of **Heisenberg's Uncertainty Principle** states that one cannot simultaneously measure with perfect accuracy both the position and momentum of a particle. A mathematical formulation of this Principle casts momentum as M_x and position as D; the fact that the difference can be a non-zero multiple of the identity is interpreted as saying that one cannot find common *eigenvalues* for M_x and D, where *eigenvalues* correspond to *observable states* of the system. We point out, however, that in the world of Physics, the underlying vector space is not $\mathbb{C}[x]$.

1.6. Example.

(a) **Rotations in $\mathbb{R}^2$.** Let $\theta \in \mathbb{R}$. Given $(x, y) \in \mathbb{R}^2$, we may write $x = r\cos\alpha, y = r\sin\alpha$ for some $\alpha \in \mathbb{R}$, where $r = \sqrt{x^2 + y^2}$. Define

$$\begin{aligned}R_\theta(x, y) &= R_\theta(r\cos\alpha, r\sin\alpha) \\ &= (r\cos(\alpha + \theta), r\sin(\alpha + \theta)) \\ &= (x\cos\theta - y\sin\theta, x\sin\theta + y\cos\theta).\end{aligned}$$

Then R_θ is *rotation by an angle of θ radians* around the origin in the *counter*clockwise direction. It is linear.

(b) **Reflection about the x-axis.** The map

$$\begin{array}{rccc} F_1: & \mathbb{R}^2 & \to & \mathbb{R}^2 \\ & (x, y) & \mapsto & (x, -y) \end{array}$$

represents the reflection of $\mathbb{R}^2$ about the x-axis. It is linear. (How would one define F_2, the reflection about the y-axis. Is it linear?)

(c) **The projection onto the y-axis.** The map

$$\begin{array}{rccc} P_2: & \mathbb{R}^2 & \to & \mathbb{R}^2 \\ & (x, y) & \mapsto & (0, y) \end{array}$$

represents the projection of $\mathbb{R}^2$ onto the y-axis. It is linear. (How would one define P_1, the reflection onto the x-axis. Is it linear?)

1.7. Theorem. *Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces over a field $\mathbb{F}$. Given $R, T \in \mathcal{L}(\mathcal{V}, \mathcal{W})$ and $\kappa \in \mathbb{F}$, we define*

$$(R+T)x := Rx + Tx, \quad x \in \mathcal{V}$$

and

$$(\kappa T)x := \kappa(Tx), \quad x \in \mathcal{V}.$$

With these operations, $\mathcal{L}(\mathcal{V}, \mathcal{W})$ is a vector space over $\mathbb{F}$.
Proof. We leave the proof of this as an exercise* for the reader.

□

The next definition actually appears in the Exercises at the end of Chapter 1.

1.8. Definition. *Let $\mathcal{V}, \mathcal{W}$ be vector spaces over a field $\mathbb{F}$ and suppose that $T \in \mathcal{L}(\mathcal{V}, \mathcal{W})$. Then the **kernel** of T is*

$$\ker T := \{x \in \mathcal{V} : Tx = \mathbf{0}_{\mathcal{W}}\},$$

*and the **range** of T is*

$$\operatorname{ran} T := \{Tx : x \in \mathcal{V}\}.$$

1.9. Remark. Note that $\mathbf{0}_{\mathcal{V}} \in \ker T$, so that $\ker T \neq \varnothing$, and that if $x, y \in \ker T$ and $\kappa \in \mathbb{F}$, then

$$T(\kappa x + y) = \kappa Tx + Ty = \kappa\, \mathbf{0}_{\mathcal{W}} + \mathbf{0}_{\mathcal{W}},$$

so that $\kappa x + y \in \ker T$ and the latter is a subspace of $\mathcal{V}$.

As for the range of T, again, $T\mathbf{0}_{\mathcal{V}} = \mathbf{0}_{\mathcal{W}} \in \operatorname{ran} T$, so that $\operatorname{ran} T \neq \varnothing$, and if $w, z \in \operatorname{ran} T$ – say $w = Tx$ and $z = Ty$ for some $x, y \in \mathcal{V}$ – and if $\kappa \in \mathbb{F}$, then

$$\kappa w + z = \kappa Tx + Ty = T(\kappa x + y) \in \operatorname{ran} T,$$

so that $\operatorname{ran} T$ is a subspace of $\mathcal{W}$.

1.10. Definition. *Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces over the field $\mathbb{F}$ and let $T \in \mathcal{L}(\mathcal{V}, \mathcal{W})$. The **nullity** of T is*

$$\operatorname{nul} T := \dim \ker T,$$

*and the **rank** of T is*

$$\operatorname{rank} T := \dim \operatorname{ran} T.$$

1.11. Example. Let

$$\begin{array}{rccc} T: & \mathbb{R}^2 & \to & \mathbb{R}^3 \\ & (x,y) & \mapsto & (2x+y, 6x, x-y). \end{array}$$

If $(x,y) \in \ker T$, then $T(x,y) = (0,0,0)$, whence $2x + y = 0$, $6x = 0$ and $x - y = 0$. Solving this system of equations yield $x = 0 = y$. Thus

$$\ker T = \{\mathbf{0}_{\mathbb{R}^2}\} = \{(0,0)\}.$$

Thus $\operatorname{nul} T = \dim\{(0,0)\} = 0$.

As for the range of T, $T(1,0) = (2,6,1)$ and $T(0,1) = (1,0,-1)$, and so for $x, y \in \mathbb{R}$,

$$T(x,y) = T(x,0) + T(0,y) = xT(1,0) + yT(0,1) = x(2,6,1) + y(1,0,-1).$$

That is,

$$\operatorname{ran} T = \operatorname{span}\{(2,6,1),(1,0,-1)\}.$$

Since $\{(2,6,1),(1,0,-1)\}$ is linearly independent, $\operatorname{rank} T = \dim \operatorname{ran} T = 2$.

1.12. Proposition. *Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces and $T : \mathcal{V} \to \mathcal{W}$ be a linear map. Let $\mathcal{S} := \{x_\alpha\}_{\alpha\in\Lambda}$ be a spanning set for $\mathcal{V}$. Then*

$$\operatorname{ran} T = \operatorname{span}\{Tx_\alpha\}_{\alpha\in\Lambda}.$$

Proof. Let $w \in \operatorname{ran} T$. Then $w = Tx$ for some $x \in \mathcal{V}$. Since $\mathcal{S}$ is a spanning set for $\mathcal{V}$, there exist $x_{\alpha_1}, x_{\alpha_2}, \ldots, x_{\alpha_n} \in \mathcal{S}$ and $\kappa_1, \kappa_2, \ldots, \kappa_n \in \mathbb{F}$ such that $x = \sum_{j=1}^n \kappa_j x_{\alpha_j}$. But then

$$w = Tx = T\Big(\sum_{j=1}^n \kappa_j x_{\alpha_j}\Big) = \sum_{j=1}^n \kappa_j T x_{\alpha_j} \in \operatorname{span}\{Tx_\alpha\}_{\alpha\in\Lambda}.$$

□

1.13. Example. Suppose that $T : \mathbb{R}^2 \to \mathbb{R}^3$ is linear, that $T(1,1) = (1,0,2)$ and $T(2,3) = (1,-1,4)$. Since $\{(1,1),(2,3)\}$ is a basis for $\mathbb{R}^2$, it generates $\mathbb{R}^2$, and so from Proposition 1.12,

$$\operatorname{ran} T = \operatorname{span}\{(1,0,2),(1,-1,4)\}.$$

For example,

$$\begin{aligned} T(8,11) &= T[2(1,1) + 3(2,3)] \\ &= 2\cdot T(1,1) + 3\cdot T(2,3) \\ &= 2(1,0,2) + 3(1,-1,4) \\ &= (5,-3,16). \end{aligned}$$

Observe that we can completely determine T just by knowing what it does to this basis. Indeed, $(1,0) = 3(1,1) - 1(2,3)$, and $(0,1) = -2(1,1) + 1(2,3)$, and thus if $(x,y) \in \mathbb{R}^2$, then

$$\begin{aligned}(x,y) &= x(1,0) + y(0,1)\\ &= x(3(1,1) - 1(2,3)) + y(-2(1,1) + 1(2,3))\\ &= (3x-2y)(1,1) + (-x+y)(2,3).\end{aligned}$$

It follows that

$$\begin{aligned}T(x,y) &= T((3x-2y)(1,1) + (-x+y)(2,3))\\ &= (3x-2y)T(1,1) + (-x+y)T(2,3)\\ &= (3x-2y)(1,0,2) + (-x+y)(1,-1,4)\\ &= (2x-y, x-y, 2x).\end{aligned}$$

1.14. Example.

(a) Let $\mathcal{V} = \ell^\infty(\mathbb{N}) := \{(x_n)_n \in \mathbb{R}^{\mathbb{N}} : \sup_n |x_n| < \infty\}$. Define the linear map

$$\begin{array}{rccc} S: & \mathcal{V} & \to & \mathcal{V} \\ & (x_n)_n & \mapsto & (0, x_1, x_2, x_3, \ldots). \end{array}$$

(You should verify that S is indeed linear.) It is called the **unilateral forward shift** operator. Then S is injective - i.e. $\ker S = \{0\}$, and $\operatorname{nul} S = 0$. We leave it as an exercise for the reader to show that $\operatorname{rank} S = \infty$.

We may also define the linear map

$$\begin{array}{rccc} T: & \mathcal{V} & \to & \mathcal{V} \\ & (x_n)_n & \mapsto & (x_2, x_3, x_4, \ldots). \end{array}$$

Again, $\operatorname{rank} T = \infty$. This time, $\ker T = \operatorname{span}\{e_1\}$, where $e_1 = (1,0,0,0,\ldots)$.

(b) Let $n \in \mathbb{N}$ and $\mathbb{F}$ be a field. Define the map $J_n \in \mathcal{L}(\mathbb{F}^n)$ via

$$J((x_1, x_2, \ldots, x_n)) = (x_2, x_3, \ldots, x_n, 0).$$

This maps is known as the $n \times n$ **nilpotent Jordan cell**.

Then $\ker J_n = \operatorname{span} e_1$, so $\operatorname{nul} J_n = 1$. Also, $\operatorname{rank} J_n = n - 1$, as $\operatorname{ran} J_n = \mathbb{F}^{n-1} \oplus \{0\}$.

Jordan cells will be extremely important in the second linear algebra course.

The next Theorem is sometimes referred to as **the Dimension Theorem**, although you should not be alarmed if you ask someone what the "Dimension Theorem" is and they are unable to tell you (even though they know the result very well).

1.15. Theorem. [The Dimension Theorem] *Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces over a field $\mathbb{F}$ and $T \in \mathcal{L}(\mathcal{V}, \mathcal{W})$. If $\mathcal{V}$ is finite-dimensional, then*

$$\operatorname{nul} T + \operatorname{rank} T = \dim \mathcal{V}.$$

Proof. Now $\ker T$ is a subspace of the finite-dimensional space $\mathcal{V}$, so $\dim \ker T \leq \dim \mathcal{V} < \infty$. Let $m := \dim \ker T$ and $n := \dim \mathcal{V}$. Let $\{u_1, u_2, \ldots, u_m\}$ be a basis for $\ker T$. We may extend this to a basis $\mathcal{B} := \{u_1, u_2, \ldots, u_m, v_1, v_2, \ldots, v_{n-m}\}$ for $\mathcal{V}$.

As we have seen earlier, if we set $\mathcal{Z} := \operatorname{span}\{v_1, v_2, \ldots, v_{n-m}\}$, then $\mathcal{Z}$ is a complement to $\ker T$; i.e. $\ker T + \mathcal{Z} = \mathcal{V}$, and $\mathcal{Z} \cap \ker T = \{\mathbf{0}\}$.

Now

$$\begin{aligned} \operatorname{ran} T &= \operatorname{span}\{Tu_1, Tu_2, \ldots, Tu_m, Tv_1, Tv_2, \ldots, Tv_m\} \\ &= \operatorname{span}\{Tv_1, Tv_2, \ldots, Tv_{n-m}\}, \end{aligned}$$

as $Tu_j = \mathbf{0}$ for all j. We claim that $\{Tv_1, Tv_2, \ldots, Tv_{n-m}\}$ is linearly independent.

Indeed, if $\kappa_j \in \mathbb{F}$, $1 \leq j \leq n-m$, and if $\sum_{j=1}^{n-m} \kappa_j Tv_j = \mathbf{0}$, then by linearity of T,

$$T\Big(\sum_{j=1}^{n-m} \kappa_j v_j\Big) = \mathbf{0},$$

so that $\sum_{j=1}^{n-m} \kappa_j v_j \in \mathcal{Z} \cap \ker T = \{\mathbf{0}\}$. But $\{v_1, v_2, \ldots, v_{n-m}\}$ is linearly independent, being a subset of a linearly independent set, and thus $\kappa_j = 0$ for all $1 \leq j \leq n-m$. That is,

$$\mathcal{D} := \{Tv_1, Tv_2, \ldots, Tv_{n-m}\}$$

is linearly independent. Thus $\mathcal{D}$ is a basis for $\operatorname{ran} T$, and so $\operatorname{rank} T = \dim \operatorname{ran} T = n - m = \dim \mathcal{V} - \operatorname{nul} T$, completing the proof.

□

Few results are as "standard" as the next one.

1.16. Proposition. *Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces over a field $\mathbb{F}$ and $T \in \mathcal{L}(\mathcal{V}, \mathcal{W})$. The map T is injective if and only if* $\ker T = \{\mathbf{0}\}$.

Proof. Suppose first that T is injective. Since $T\mathbf{0}_{\mathcal{V}} = \mathbf{0}_{\mathcal{W}}$ (this holds for any linear map from $\mathcal{V}$ to $\mathcal{W}$), we have that $\ker T = \{\mathbf{0}_{\mathcal{V}}\}$.

Conversely, suppose that $\ker T = \{\mathbf{0}_{\mathcal{V}}\}$. If $x, y \in \mathcal{V}$ and $Tx = Ty$, then $T(x - y) = Tx - Ty = \mathbf{0}_{\mathcal{W}}$, and so $x - y \in \ker T = \mathbf{0}_{\mathcal{V}}$; i.e. $x = y$. Hence T is injective.

□

As a consequence of the Dimension Theorem, we can show the following.

1.17. Theorem. *Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces over a field $\mathbb{F}$ and suppose that* $\dim \mathcal{V} = \dim \mathcal{W} < \infty$. *Let $T \in \mathcal{L}(\mathcal{V}, \mathcal{W})$. The following statements are equivalent.*

(a) *T is injective.*
(b) *T is surjective.*
(c) $\operatorname{rank} T = \dim \mathcal{V}$.

Proof.

(a) Suppose that T is injective. By the Dimension Theorem,

$$\dim \operatorname{ran} T = \operatorname{rank} T = \dim \mathcal{V} - \operatorname{nul} T = \dim \mathcal{V} - 0 = \dim \mathcal{V} = \dim \mathcal{W}.$$

Thus $\operatorname{ran} T = \mathcal{W}$, and so T is surjective.

(b) Suppose that T is surjective. Then

$$\dim \mathcal{V} = \dim \mathcal{W} = \dim \operatorname{ran} T = \operatorname{rank} T.$$

(c) Suppose that $\operatorname{rank} T = \dim \mathcal{V}$. Then

$$\dim \mathcal{V} = \operatorname{rank} T = \dim \mathcal{V} - \operatorname{nul} T$$

implies that $\operatorname{nul} T = 0$, i.e. $\ker T = \{\mathbf{0}\}$. By Proposition 1.16, T is injective.

□

1.18. Example.

(a) The above theorem fails spectacularly when the dimensions of $\mathcal{V}$ and $\mathcal{W}$ are infinite. For example, let

$$\mathcal{V} = \mathcal{W} = \ell^\infty(\mathbb{N}) = \{x = (x_n)_n \in \mathbb{R}^{\mathbb{N}} : \sup_n |x_n| < \infty\}.$$

Define

$$\begin{array}{rccc} S: & \ell^\infty(\mathbb{N}) & \to & \ell^\infty(\mathbb{N}) \\ & (x_n)_n & \mapsto & (0, x_1, x_2, x_3, \ldots). \end{array}$$

We refer to S as the **unilateral forward shift operator** on $\ell^\infty(\mathbb{N})$.

Then $\ker S = \{\mathbf{0}\}$, but S is not onto.

(b) Let

$$\begin{array}{rccc} T: & \mathbb{R}^3 & \to & \mathbb{R}^2 \\ & (x, y, z) & \mapsto & (x + y, y + z). \end{array}$$

Then T is surjective, but not injective.

(c) Let $\theta \in \mathbb{R}$ and consider

$$\begin{array}{rccc} R_\theta: & \mathbb{R}^2 & \to & \mathbb{R}^2 \\ & (x, y) & \mapsto & (x\cos\theta + y\sin\theta, -x\sin\theta + y\cos\theta). \end{array}$$

The R_θ is injective, hence surjective.

The next theorem is incredibly useful.

1.19. Theorem. *Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces over the field $\mathbb{F}$ and suppose that $\mathcal{B} := \{b_\alpha\}_{\alpha\in\Lambda}$ is a basis for $\mathcal{V}$. Given arbitrary vectors $w_\alpha \in \mathcal{W}$, $\alpha \in \Lambda$, there exists one and only one linear map $T \in \mathcal{L}(\mathcal{V}, \mathcal{W})$ such that*

$$Tx_\alpha = w_\alpha, \quad \alpha \in \Lambda.$$

Proof. This will appear as an assignment Question.

□

1.20. Example. Recall that $\{1, x, x^2, x^3, \ldots\}$ is a basis for $\mathbb{C}[x]$ over $\mathbb{C}$. By Theorem 1.19, there exists a unique linear map $T : \mathbb{C}[x] \to \mathbb{C}[x]$ such that $T1 = 0$ and $Tx^n = nx^{n-1}$, $n \geq 1$. Since the differentiation operator $Dp = p'$ satisfies this property, we must have $T = D$.

2. From linear maps to matrices

2.1. Matrices do not exist purely because they do. Their «raison d'être» is that they are an amazing computational tool that allows one to better understand linear maps, which is what we are "really" interested in. In fact, it is our need to understand the composition of linear maps that will give rise to the "unusual" multiplication of matrices that you may have seen before.

2.2. Definition. *Let $n \in \mathbb{N}$ and let $\mathcal{V}$ be an n-dimensional vector space over a field $\mathbb{F}$. An **ordered basis** for $\mathcal{V}$ is an element $\mathcal{B}_\bullet := (b_1, b_2, \ldots, b_n) \in \mathcal{V}^n$, where $\mathcal{B} := \{b_1, b_2, \ldots, b_n\}$ is a basis for $\mathcal{V}$.*

2.3. Remarks. The difference between an *ordered basis* and a *basis* is that with an ordered basis, we are keeping track of which basis element is written first, which is written second, etc.. A basis is just a set, so that

$$\mathcal{B} := \{b_1, b_2, \ldots, b_n\} = \{b_n, b_{n-1}, \ldots, b_2, b_1\},$$

whereas $\mathcal{B}_\bullet := (b_1, b_2, \ldots, b_n) \neq (b_n, b_{n-1}, \ldots, b_2, b_1) =: \mathcal{C}_\bullet$.

Thus, $\mathcal{B}_\bullet := ((1,0,0),(0,1,0),(0,0,1))$ is an ordered basis for $\mathbb{F}^3$, while $\mathcal{D}_\bullet := ((0,1,0),(1,0,0),(0,0,1))$ is a *distinct* ordered basis for $\mathbb{F}^3$.

In most of the cases below, we shall only be using an ordered basis, and to simplify the notation, we just refer to the ordered basis $\mathcal{B} = (b_1, b_2, \ldots, b_n)$.

The connection between linear maps and matrices arises through the identification of a vector in an n-dimensional vector space $\mathcal{V}$ over a field $\mathbb{F}$ with the n-tuple in $\mathbb{F}^n$ consisting of the coordinates of that vector with respect to a specified ordered basis. As we shall see – linear maps act upon *vectors*, while matrices act upon *coordinates*.

2.4. Definition. *Let $\mathcal{V}$ be an n-dimensional vector space over a field $\mathbb{F}$ and let $\mathcal{B} := (b_1, b_2, \ldots, b_n)$ be an ordered basis for $\mathcal{V}$. Let $x \in \mathcal{V}$, and choose $\kappa_1, \kappa_2, \ldots, \kappa_n \in \mathbb{F}$ so that*

$$x = \sum_{j=1}^{n} \kappa_j b_j.$$

(Recall that the fact that $\mathcal{B}$ is a basis implies that the κ_j's are uniquely determined.)

The ***coordinate vector of*** *x relative to $\mathcal{B}$ is:*

$$[x]_{\mathcal{B}} := \begin{bmatrix} \kappa_1 \\ \kappa_2 \\ \vdots \\ \kappa_n \end{bmatrix} \in \mathbb{F}^n.$$

2.5. Example. Let $\mathcal{V} = \mathbb{R}_2[x] := \{p_0 + p_1 x + p_2 x^2 : p_j \in \mathbb{R}, 0 \le j \le 2\}$. Consider two ordered bases for $\mathcal{V}$, namely $\mathcal{B} = (1+x, 1+x^2, 1+x+x^2)$ and $\mathcal{C} = (1+x+x^2, 1+x^2, 1+x)$.

Let $p = 4 + 5x + 2x^2 = 2(1+x) - 1(1+x^2) + 3(1+x+x^2)$.

Then

$$[p]_{\mathcal{B}} = \begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix} \text{ and } [p]_{\mathcal{C}} = \begin{bmatrix} 3 \\ -1 \\ 2 \end{bmatrix}.$$

The vector p has not changed – its coordinates with respect to two different coordinate systems (i.e. ordered bases) *has*.

2.6. Example. A very special and useful case is the following. Let $n \in \mathbb{N}$ and suppose that $\mathcal{V} = \mathbb{F}^n$. Let $\mathcal{B} := (e_1, e_2, \ldots, e_n) \in \mathcal{V}^n$ denote the **standard ordered basis** for $\mathbb{F}^n$, so that $e_k = (0, 0, \ldots, 0, 1, 0, \ldots, 0)$, with the unique "1" occurring at the k^{th} coordinate, $1 \le k \le n$. Then

$$[e_k]_{\mathcal{B}} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

More generally, if $x = (x_1, x_2, \ldots, x_n) \in \mathbb{F}^n$, then

$$[x]_{\mathcal{B}} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$

In other words, x looks strikingly similar to its coordinate vector $[x]_{\mathcal{B}}$! Again, this is a *very special* setting! You shouldn't expect this kind of thing in general.

We are now in a position to associate a matrix to each linear map $T \in \mathcal{L}(\mathcal{V}, \mathcal{W})$, when $\mathcal{V}$ and $\mathcal{W}$ are finite-dimensional.

2.7. Definition. *Let $\mathcal{V}$ and $\mathcal{W}$ be finite-dimensional vector spaces over a field $\mathbb{F}$ and let $T \in \mathcal{L}(\mathcal{V}, \mathcal{W})$. Let $\mathcal{D} = (v_1, v_2, \ldots, v_n)$ be an ordered basis for $\mathcal{V}$, and $\mathcal{C} = (w_1, w_2, \ldots, w_m)$ be an ordered basis for $\mathcal{W}$. (We are choosing the letter $\mathcal{D}$ to indicate that it is a basis for the domain, while $\mathcal{C}$ indicates that it is an ordered basis for the codomain.)*

For each $1 \leq j \leq n$, we write

$$Tv_j = \sum_{i=1}^{m} \alpha_{ij} w_i.$$

(The choice of the α_{ij}'s is unique!)

*The **matrix** of T **relative to** $\mathcal{D}$ and $\mathcal{C}$ is:*

$$[T]_{\mathcal{D}}^{\mathcal{C}} := [\alpha_{ij}] \in \mathbb{M}_{m \times n}(\mathbb{F}).$$

In the case where $\mathcal{W} = \mathcal{V}$ and $\mathcal{D} = \mathcal{C}$, we also write

$$[T]_{\mathcal{D}} = [\alpha_{ij}] \in \mathbb{M}_n(\mathbb{F}).$$

2.8. Example. Let $\mathcal{D}$ and $\mathcal{C}$ denote the standard ordered bases for $\mathbb{R}^3$ and $\mathbb{R}^1$ respectively, so that $\mathcal{D} = (e_1 = (1,0,0), e_2 = (0,1,0), e_3 = (0,0,1))$ and $\mathcal{C} = (f = 1)$. Define

$$\begin{array}{rccc} T: & \mathbb{R}^3 & \to & \mathbb{R}^1 \\ & (x,y,z) & \mapsto & 2x + y - 3z. \end{array}$$

Then

$$\begin{aligned} Te_1 &= T(1,0,0) = 2 = 2f \\ Te_2 &= T(0,1,0) = 1 = 1f \\ Te_3 &= T(0,0,1) = -3 = -3f, \end{aligned}$$

and so

$$[T]_{\mathcal{D}}^{\mathcal{C}} = \begin{bmatrix} 2 & 1 & -3 \end{bmatrix} \in \mathbb{M}_{1\times 3}(\mathbb{R}).$$

2.9. Example. Let

$$\begin{array}{rccc} T: & \mathbb{M}_2(\mathbb{F}) & \to & \mathbb{M}_2(\mathbb{F}) \\ & A & \mapsto & A^{\mathrm{t}}. \end{array}$$

Let $\mathcal{B} = (E_{11}, E_{12}, E_{21}, E_{22})$ be the standard ordered basis for $\mathbb{M}_2(\mathbb{F})$.

Then

$$[T]_{\mathcal{B}} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

2.10. Theorem. *Let $\mathcal{V}$ and $\mathcal{W}$ be finite-dimensional vector spaces with ordered bases $\mathcal{D} = (v_1, v_2, \ldots, v_n)$ and $\mathcal{C} = (w_1, w_2, \ldots, w_m)$ respectively. Let $R, T \in \mathcal{L}(\mathcal{V}, \mathcal{W})$ and $\kappa \in \mathbb{F}$. Then*

(a) $[R+T]_{\mathcal{D}}^{\mathcal{C}} = [R]_{\mathcal{D}}^{\mathcal{C}} + [T]_{\mathcal{D}}^{\mathcal{C}}$; *and*
(b) $[\kappa T]_{\mathcal{D}}^{\mathcal{C}} = \kappa\,[T]_{\mathcal{D}}^{\mathcal{C}}$.

Thus the map

$$\begin{array}{rccc} \Phi: & \mathcal{L}(\mathcal{V},\mathcal{W}) & \to & \mathbb{M}_{m\times n}(\mathbb{F}) \\ & T & \mapsto & [T]_{\mathcal{D}}^{\mathcal{C}} \end{array}$$

is a linear bijection.

Proof.

(a) Let $R, T \in \mathcal{L}(\mathcal{V}, \mathcal{W})$. Write $[R]_{\mathcal{D}}^{\mathcal{C}} = [r_{ij}]$ and $[T]_{\mathcal{D}}^{\mathcal{D}} = [t_{ij}]$. For each $1 \le j \le n$,

$$(R+T)(v_j) = Rv_j + Tv_j = \sum_{i=1}^{m} r_{ij}w_i + \sum_{i=1}^{m} t_{ij}w_i = \left(\sum_{i=1}^{m}(r_{ij}+t_{ij})\right)w_i.$$

It follows that

$$[R+T]_{\mathcal{D}}^{\mathcal{C}} = [r_{ij}+t_{ij}] = [r_{ij}] + [t_{ij}] = [R]_{\mathcal{D}}^{\mathcal{C}} + [T]_{\mathcal{D}}^{\mathcal{C}}.$$

(b) The proof is similar to that of (a) above. If $\kappa \in \mathbb{F}$, then for all $1 \le j \le n$,

$$(\kappa T)(v_j) = \kappa(Tv_j) = \kappa \sum_{i=1}^{m} t_{ij}w_i = \sum_{i=1}^{m}(\kappa t_{ij})w_i.$$

Thus

$$[\kappa T]_{\mathcal{D}}^{\mathcal{C}} = [\kappa t_{ij}] = \kappa[t_{ij}] = \kappa[T]_{\mathcal{D}}^{\mathcal{C}}.$$

□

2.11. Remark. It is important to note that Φ depends not only on $\mathcal{V}$ and $\mathcal{W}$, but also upon $\mathcal{D}$ and $\mathcal{C}$; in other words, $\Phi = \Phi_{\mathcal{D},\mathcal{C}}$. We drop the subscripts to improve the readability.

2.12. Example. Let $\mathcal{V} = \mathbb{R}^3$ with standard ordered basis $\mathcal{D} = (e_1, e_2, e_3)$ and $\mathcal{W} = \mathbb{R}^2$ with standard ordered basis $\mathcal{C} = (f_1, f_2)$. Suppose that $R : \mathcal{V} \to \mathcal{W}$ is the map $R(x, y, z) = (2x + y, z)$. Then

$$[R]_{\mathcal{D}}^{\mathcal{C}} = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

2.13. Example. Suppose that $\mathcal{V}$ is a vector space with ordered basis $\mathcal{D} = (v_1, v_2, \ldots, v_n)$ and $\mathcal{W}$ is a vector space over the same field with ordered basis $\mathcal{C} = (w_1, w_2, \ldots, w_m)$. By Theorem 2.10, given a matrix $A = [\alpha_{ij}] \in \mathbb{M}_{m\times n}(\mathbb{F})$, we may associate to it a linear map $T_A : \mathcal{V} \to \mathcal{W}$ via:

$$T_A v_j = \sum_{i=1}^{m} \alpha_{ij} w_i, \qquad 1 \le j \le n.$$

We may then extend the definition of T_A to all of $\mathcal{V}$ by linearity - see Theorem 1.19. Observe that $T_A \in \mathcal{L}(\mathcal{V}, \mathcal{W})$ and that $[T_A]_{\mathcal{D}}^{\mathcal{C}} = A$.

For example, suppose that

$$\mathcal{V} = \operatorname{span}\{\sin x, \sin 2x, \cos x\},$$

and that

$$\mathcal{W} = \operatorname{span}\{e^x, \log(1+x)\},$$

thought of as subspaces of $\mathcal{C}([0, 2\pi], \mathbb{R})$. Set $\mathcal{D} = (\sin x, \sin 2x, \cos x)$ and $\mathcal{C} = (e^x, \log(1+x))$ as ordered bases for $\mathcal{V}$ and $\mathcal{W}$ respectively.

If $A = \begin{bmatrix} 8 & -2 & \sqrt{3} \\ 1 & -1 & 0 \end{bmatrix} \in \mathbb{M}_{2\times 3}(\mathbb{R})$, then the associated map satisfies:

$$\begin{aligned} T_A \sin x &= 8e^x + 1\log(1+x) \\ T_A \sin 2x &= -2e^x - 1\log(1+x) \\ T_A \cos x &= \sqrt{3}e^x + 0\log(1+x). \end{aligned}$$

It follows that

$$T_A(\kappa_1 \sin x + \kappa_2 \sin 2x + \kappa_3 \cos x) = (8\kappa_1 - 2\kappa_2 + \sqrt{3}\kappa_3)e^x + (1\kappa_1 - 1\kappa_2 + 0\kappa_3)\log(1+x).$$

3. Composition of functions

3.1. There is a property that functions have that go beyond simply adding them and scaling them, namely: given functions f and g, we can compose functions, provided that the range of g is a subset of the domain of f.

3.2. Theorem. *Let $\mathcal{V}, \mathcal{W}, \mathcal{Y}$ and $\mathcal{Z}$ be vector spaces over the field $\mathbb{F}$. Suppose that $R, R_1, R_2 \in \mathcal{L}(\mathcal{V}, \mathcal{W})$, $S, S_1, S_2 \in \mathcal{L}(\mathcal{W}, \mathcal{Y})$, and $T \in \mathcal{L}(\mathcal{Y}, \mathcal{Z})$. Then $S \circ R$ is linear. Furthermore,*

(a) $S \circ (R_1 + R_2) = S \circ R_1 + S \circ R_2$;
(b) $(S_1 + S_2) \circ R = S_1 \circ R + S_2 \circ R$;
(c) $T \circ (S \circ R) = (T \circ S) \circ R$;
(d) $I_{\mathcal{W}} \circ R = R = R \circ I_{\mathcal{V}}$;
(e) $\kappa(S \circ R) = (\kappa S) \circ R = S \circ (\kappa R)$.

Proof. Most of the proof is routine and is left to the reader. We will only prove that $S \circ R$ is linear. Indeed, if $v_1, v_2 \in \mathcal{V}$ and $\kappa \in \mathbb{F}$, then

$$\begin{aligned}(S \circ R)(\kappa v_1 + v_2) &= S(R(\kappa v_1 + v_2)) \\ &= S(\kappa R v_1 + R v_2) \\ &= \kappa S(R v_1) + S(R v_2) \\ &= \kappa (S \circ R)(v_1) + (S \circ R)(v_2).\end{aligned}$$

Thus $S \circ R \in \mathcal{L}(\mathcal{V}, \mathcal{Y})$.

□

3.3. With the notation from Theorem 3.2, suppose that $\mathcal{D} = (v_1, v_2, \ldots, v_n)$, $\mathcal{M} = (w_1, w_2, \ldots, w_m)$ and $\mathcal{C} = (y_1, y_2, \ldots, y_p)$ are ordered bases for $\mathcal{V}$, $\mathcal{W}$ and $\mathcal{Y}$ respectively.

Write $[R]_{\mathcal{D}}^{\mathcal{M}} = [r_{ij}]$, $[S]_{\mathcal{M}}^{\mathcal{C}} = [s_{ij}]$. Let us determine $[S \circ R]_{\mathcal{D}}^{\mathcal{C}}$.

Now for $1 \le j \le n$,

$$\begin{aligned}(S \circ R) v_j &= S(R v_j) \\ &= S\Big(\sum_{k=1}^{m} r_{kj} w_k\Big) \\ &= \sum_{k=1}^{m} r_{kj} (S w_k) \\ &= \sum_{k=1}^{m} r_{kj} \left(\sum_{i=1}^{p} s_{ik} y_i\right) \\ &= \sum_{i=1}^{p} \left(\sum_{k=1}^{m} s_{ik} r_{kj}\right) y_i\end{aligned}$$

Letting $q_{ij} := \sum_{k=1}^{m} s_{ik} r_{kj}$, $1 \le i \le p$, $1 \le j \le n$, we find that

$$[S \circ R]_{\mathcal{D}}^{\mathcal{C}} = [q_{ij}].$$

In light of this, we **define** the product of matrices in such a way that

$$[S \circ R]_{\mathcal{D}}^{\mathcal{C}} = [S]_{\mathcal{M}}^{\mathcal{C}} \cdot [R]_{\mathcal{D}}^{\mathcal{M}},$$

namely: given $A = [a_{ij}] \in \mathbb{M}_{m \times n}(\mathbb{F})$ and $B = [b_{ij}] \in \mathbb{M}_{p \times m}(\mathbb{F})$, we set

$$B \cdot A := [d_{ij}] \in \mathbb{M}_{p \times n}(\mathbb{F}),$$

where

$$d_{ij} := \sum_{k=1}^{m} b_{ik} a_{kj}, \quad 1 \le i \le p, 1 \le j \le n.$$

In other words, the *entire reason for this form of multiplying matrices* is because we wish the product of matrices to represent the composition of linear maps.

We remark that there does exist a "naïve" version of matrix multiplication, whereby – if $A = [a_{ij}], B = [b_{ij}] \in \mathbb{M}_{m\times n}(\mathbb{F})$ – we set $B * A := [b_{ij}a_{ij}]$, and this is referred to as **Schur multiplication** or **Hadamard multiplication**. Because it does not represent the composition of B with A, we shall not consider it here, although it is relevant when one studies *tensor products* of vector spaces. In particular, it arises in quantum information theory.

3.4. Example.

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

(b) $\begin{bmatrix}4 & 9\\ 1 & 1\end{bmatrix}\begin{bmatrix}-1 & 0\\ 2 & 1\end{bmatrix} = \begin{bmatrix}14 & 9\\ 1 & 1\end{bmatrix}$.

As an immediate consequence of the definition of the product of matrices we obtain:

3.5. Corollary. *If $\mathcal{V}$ is a finite-dimensional vector space with ordered basis $\mathcal{D}$, and if $R, S \in \mathcal{L}(\mathcal{V})$, then*

$$[SR]_{\mathcal{D}} = [S]_{\mathcal{D}}\,[R]_{\mathcal{D}}.$$

3.6. Example. Consider the maps

$$\begin{array}{rccc} R: & \mathbb{R}_4[x] & \to & \mathbb{R}_3[x] \\ & p & \mapsto & p' \end{array}$$

and

$$\begin{array}{rccc} S: & \mathbb{R}_3[x] & \to & \mathbb{R}_4[x] \\ & q_0 + q_1x + q_2x^2 + q_3x^3 & \mapsto & q_0x + \frac{1}{2}q_1x^2 + \frac{1}{3}q_2x^3 + \frac{1}{4}q_3x^4. \end{array}$$

Let $\mathcal{D} = (1, x, x^2, x^3, x^4)$ be the standard ordered basis for $\mathbb{R}_4[x]$, and $\mathcal{C} = (1, x, x^2, x^3)$ be the standard ordered basis for $\mathbb{R}_3[x]$. Then

$$[R]_{\mathcal{D}}^{\mathcal{C}} = \begin{bmatrix}0 & 1 & 0 & 0 & 0\\ 0 & 0 & 2 & 0 & 0\\ 0 & 0 & 0 & 3 & 0\\ 0 & 0 & 0 & 0 & 4\end{bmatrix},$$

and

$$[S]_{\mathcal{C}}^{\mathcal{D}} = \begin{bmatrix}0 & 0 & 0 & 0\\ 1 & 0 & 0 & 0\\ 0 & \frac{1}{2} & 0 & 0\\ 0 & 0 & \frac{1}{3} & 0\\ 0 & 0 & 0 & \frac{1}{4}\end{bmatrix}.$$

Thus

$$[SR]_{\mathcal{D}} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix},$$

while

$$[RS]_{\mathcal{C}} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Note that $[RS]_{\mathcal{C}} = [I_{\mathbb{R}_3[x]}]_{\mathcal{C}}$. Is this surprising?

3.7. Notation. Let $n \in \mathbb{N}$. We write

$$I_n := \begin{bmatrix} 1 & 0 & & & 0 \\ 0 & 1 & & & \\ 0 & 0 & \ddots & & \\ 0 & 0 & & 1 & 0 \\ 0 & & & & 1 \end{bmatrix} \in \mathbb{M}_n(\mathbb{F}),$$

while

$$O_n := [0_{ij}].$$

3.8. Definition. *Given $A = [\alpha_{ij}] \in \mathbb{M}_{m\times n}(\mathbb{F})$, we define the **left multiplication operator** $L_A \in \mathcal{L}(\mathbb{F}^n, \mathbb{F}^m)$ via*

$$L_A x = Ax, \quad x \in \mathbb{F}^n.$$

Here, Ax denotes the matrix product, where x and thus Ax are written as column vectors.

*In higher-level analysis courses, the map $A \mapsto L_A$ is often referred to as the **left regular representation**. Now you know.*

3.9. Example. If $A = \begin{bmatrix} 1 & 0 \\ -1 & 1 \\ 0 & 2 \end{bmatrix} \in \mathbb{M}_{3\times 2}(\mathbb{R})$, then $L_A : \mathbb{R}^2 \to \mathbb{R}^3$ is the map

$$L_A \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -1 & 1 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 \\ -x_1 + x_2 \\ 2x_3 \end{bmatrix}.$$

The following exercise is **very important**.

3.10. Exercise. The left regular representation

$$\begin{array}{rccc} \Psi: & \mathbb{M}_{m\times n}(\mathbb{F}) & \to & \mathcal{L}(\mathbb{F}^n,\mathbb{F}^m) \\ & A & \mapsto & L_A \end{array}$$

is a linear bijection of vector spaces, and if $\mathcal{D}$ and $\mathcal{C}$ are the **standard** ordered bases for $\mathbb{F}^n$ and $\mathbb{F}^m$ respectively, then

$$[L_A]_{\mathcal{D}}^{\mathcal{C}} = A.$$

Our next result is an immediate consequence of Theorem 3.2, based upon our definition for the multiplication of matrices. We shall prove the first part of (a), and leave the remainder of the proof as an exercise.

3.11. Theorem. *Let $m,n,p,q \in \mathbb{N}$ and $\mathbb{F}$ be a field. Let $A, A_1, A_2 \in \mathbb{M}_{m\times n}(\mathbb{F})$, $B, B_1, B_2 \in \mathbb{M}_{p\times m}(\mathbb{F})$. Then*

(a) *$B(A_1 + A_2) = BA_1 + BA_2$ and $(B_1 + B_2)A = B_1A + B_2A$;*
(b) *For all $\kappa \in \mathbb{F}$, $\kappa(BA) = (\kappa B)A = B(\kappa A)$;*
(c) *$I_mA = A = AI_n$;*
(d) *If $\mathcal{V}$ is an n-dimensional vector space over $\mathbb{F}$ and $\mathcal{D}$ is **any** ordered basis for $\mathcal{V}$, then $[I_{\mathcal{V}}]_{\mathcal{D}} = I_n$.*

Proof. We prove that $B(A_1 + A_2) = BA_1 + BA_2$ via a direct computation. If $B = [b_{ij}], A_1 = [x_{ij}]$ and $A_2 = [y_{ij}]$, then $A_1 + A_2 = [x_{ij} + y_{ij}]$ and so

$$B(A_1 + A_2) = [q_{ij}],$$

where $q_{ij} = \sum_{k=1}^m b_{ik}(x_{kj} + y_{kj}) = \sum_{k=1}^m b_{ik}x_{kj} + \sum_{k=1}^m b_{ik}y_{kj}$.

Letting $r_{ij} = \sum_{k=1}^m b_{ik}x_{kj}$ and $s_{ij} = \sum_{k=1}^m b_{ik}y_{kj}$, $1 \le i \le p$, $1 \le j \le n$, we find that $BA_1 = [r_{ij}]$, $BA_2 = [s_{ij}]$ and $q_{ij} = r_{ij} + s_{ij}$ for all i,j, so that

$$B(A_1 + A_2) = [q_{ij}] = [r_{ij} + s_{ij}] = [r_{ij}] + [s_{ij}] = BA_1 + BA_2.$$

□

3.12. Theorem. *Let $\mathcal{V}$ and $\mathcal{W}$ be finite-dimensional vector spaces over the field $\mathbb{F}$ with ordered bases $\mathcal{D}$ and $\mathcal{C}$ respectively. If $T \in \mathcal{L}(\mathcal{V},\mathcal{W})$ and $x \in \mathcal{V}$, then*

$$[Tx]_{\mathcal{C}} = [T]_{\mathcal{D}}^{\mathcal{C}}\,[x]_{\mathcal{D}}.$$

Proof. As usual, we write $\mathcal{D} = (v_1, v_2, \ldots, v_n)$ and $\mathcal{C} = (w_1, w_2, \ldots, w_m)$. Given $x \in \mathcal{V}$, we find $\kappa_1, \kappa_2, \ldots, \kappa_n \in \mathbb{F}$ such that $x = \sum_{j=1}^n \kappa_j v_j$. Thus

$$[x]_{\mathcal{D}} = \begin{bmatrix} \kappa_1 \\ \kappa_2 \\ \vdots \\ \kappa_n \end{bmatrix}.$$

If $[T]_{\mathcal{D}}^{\mathcal{C}} = [t_{ij}]$, then by definition of $[T]_{\mathcal{D}}^{\mathcal{C}}$, we have that for all $1 \le j \le n$,

$$Tv_j = \sum_{i=1}^{m} t_{ij} w_i.$$

Thus

$$Tx = T(\sum_{j=1}^{n} \kappa_j v_j) = \sum_{j=1}^{n} \kappa_j T v_j = \sum_{j=1}^{n} \kappa_j (\sum_{i=1}^{m} t_{ij} w_i) = \sum_{i=1}^{m} (\sum_{j=1}^{n} t_{ij} \kappa_j) w_i,$$

so that

$$[Tx]_{\mathcal{C}} = \begin{bmatrix} \sum_{j=1}^{n} t_{1j}\kappa_j \\ \sum_{j=1}^{n} t_{2j}\kappa_j \\ \vdots \\ \sum_{j=1}^{n} t_{mj}\kappa_j \end{bmatrix} = [t_{ij}] \begin{bmatrix} \kappa_1 \\ \kappa_2 \\ \vdots \\ \kappa_n \end{bmatrix} = [T]_{\mathcal{D}}^{\mathcal{C}} [x]_{\mathcal{D}}.$$

□

3.13. Example. Consider $T : \mathbb{R}^3 \to \mathbb{R}^2$ defined by $T(x, y, z) = (x + 2y + 3z, 4x + 5y + 6z)$ as in Example 3.4. Let $e_1 = (1, 0, 0)$, $e_2 = (0, 1, 0)$ and $e_3 = (0, 0, 1) \in \mathbb{R}^3$; $f_1 = (1, 0)$ and $f_2 = (0, 1) \in \mathbb{R}^2$. Let $\mathcal{D} = (e_2, e_3, e_1)$ be an ordered basis for $\mathbb{R}^3$, $\mathcal{C} = (f_2, f_1)$ be an ordered basis for $\mathbb{R}^2$.

If $x = (1, 2, -1) \in \mathbb{R}^3$, then $[x]_{\mathcal{D}} = \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix}$, and

$$[Tx]_{\mathcal{C}} = [T]_{\mathcal{D}}^{\mathcal{C}} [x]_{\mathcal{D}} = \begin{bmatrix} 5 & 6 & 4 \\ 2 & 3 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 8 \\ 2 \end{bmatrix},$$

so that

$$T(1, 2, -1) = 8f_2 + 2f_1 = (2, 8) \in \mathbb{R}^2,$$

which agrees with the original formula for $T(1, 2, -1) = (1 + 4 - 3, 4 + 10 - 6)$!

(Note: this last statement is useful to keep in mind. If one is required to perform such a calculation on an assignment or a test – then one can check one's answer before submitting it!!!)

4. Invertibility

4.1. Recall that if X and Y are sets and $f : X \to Y$ is a function, then f is **invertible** if there exists a function $g : Y \to X$ such that $f \circ g = \mathrm{id}_Y$ and $g \circ f = \mathrm{id}_X$, where $\mathrm{id}_X : X \to X$ is the function $\mathrm{id}_X(x) = x$ for all $x \in X$ and $\mathrm{id}_Y : Y \to Y$ is the function $\mathrm{id}_Y(y) = y$ for all $y \in Y$.

Recall also that a necessary and sufficient condition for a function f to be invertible (as a function) is that f be a **bijection**, i.e. that it is both injective and surjective.

4.2. Definition. *Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces over a field $\mathbb{F}$ and suppose that $T : \mathcal{V} \to \mathcal{W}$ is a linear map. We say that T is **invertible** (as a **linear** map) if there exists $R \in \mathcal{L}(\mathcal{W}, \mathcal{V})$ such that $R \circ T = I_{\mathcal{V}}$ and $T \circ R = I_{\mathcal{W}}$.*

On the face of it, it seems harder for a *linear* map to be invertible than for a general function to be invertible, since we are requiring that the inverse function should also be linear. Having said this, we have the following theorem.

4.3. Theorem. *Suppose that $\mathcal{V}$ and $\mathcal{W}$ are vector spaces over a field $\mathbb{F}$ and that $T \in \mathcal{L}(\mathcal{V}, \mathcal{W})$. If T is invertible as a function (i.e. if T is a bijection), and if $R : \mathcal{W} \to \mathcal{V}$ is the unique function satisfying $R \circ T = I_{\mathcal{V}}$ and $T \circ R = I_{\mathcal{W}}$, then R is linear, and so T is invertible as a linear map.*

Proof. Let $w_1, w_2 \in \mathcal{W}$ and $\kappa \in \mathbb{F}$. Since T is bijective, we can find *unique* vectors $v_1, v_2 \in \mathcal{V}$ such that $Tv_1 = w_1$ and $Tv_2 = w_2$. Consider

$$\begin{aligned} R(\kappa w_1 + w_2) &= R(\kappa T v_1 + T v_2) \\ &= R(T(\kappa v_1 + v_2)) \qquad \text{as } T \text{ is linear} \\ &= I_{\mathcal{V}}(\kappa v_1 + v_2) \\ &= \kappa v_1 + v_2 \\ &= \kappa R w_1 + R w_2. \end{aligned}$$

Thus R is itself linear.

□

4.4. Remarks.

- Suppose that $T \in \mathcal{L}(\mathcal{V}, \mathcal{W})$ is invertible. Since its inverse is unique, we denote it by $T^{-1} \in \mathcal{L}(\mathcal{W}, \mathcal{V})$.
- If $\mathcal{V}, \mathcal{W}$ and $\mathcal{Y}$ are vector spaces over a field $\mathbb{F}$ and $R \in \mathcal{L}(\mathcal{V}, \mathcal{W})$ and $T \in \mathcal{L}(\mathcal{W}, \mathcal{Y})$ are invertible, then so is $T \circ R \in \mathcal{L}(\mathcal{V}, \mathcal{Y})$, and in this case $(T \circ R)^{-1} = R^{-1} \circ T^{-1} \in \mathcal{L}(\mathcal{Y}, \mathcal{V})$.
- If $T \in \mathcal{L}(\mathcal{V}, \mathcal{W})$ is invertible, then $(T^{-1})^{-1} = T$.

4.5. Example.

(a) Suppose that $\mathcal{V}$ is a vector space over $\mathbb{R}$, that $\dim \mathcal{V} = 2$, and that $\mathcal{B} = \{v_1, v_2\}$ is a basis for a vector space $\mathcal{V}$. Define $T \in \mathcal{L}(\mathcal{V})$ via

$$T(\kappa_1 v_1 + \kappa_2 v_2) = (\kappa_1 + 5\kappa_2)v_1 + (2\kappa_2)v_2.$$

Then T is invertible, and T^{-1} is the map

$$T^{-1}(\alpha_1 v_1 + \alpha_2 v_2) = (1\alpha_1 - \frac{5}{2}\alpha_2)v_1 + (\frac{1}{2}\alpha_2)v_2.$$

Indeed, consider the map $R(\alpha_1 v_1 + \alpha_2 v_2) = (1\alpha_1 - \frac{5}{2}\alpha_2)v_1 + (\frac{1}{2}\alpha_2)v_2$. Then

$$\begin{aligned}(T \circ R)v_1 &= T(Rv_1)\\ &= Tv_1\\ &= v_1\\ &= (I_{\mathcal{V}})v_1,\end{aligned}$$

while

$$\begin{aligned}(T \circ R)v_2 &= T(Rv_2)\\ &= T(-\frac{5}{2})v_1 + (\frac{1}{2})v_2\\ &= -\frac{5}{2}Tv_1 + \frac{1}{2}Tv_2\\ &= -\frac{5}{2}v_1 + \frac{1}{2}(5v_1 + 2v_2)\\ &= v_2\\ &= (I_{\mathcal{V}})v_2.\end{aligned}$$

Since $T \circ R$ and $I_{\mathcal{V}}$ agree on a basis $\mathcal{B}$ for $\mathcal{V}$, we have that $T \circ R = I_{\mathcal{V}}$. A similar calculation shows that $R \circ T = I_{\mathcal{V}}$, and thus $R = T^{-1}$.
This leave us with two questions:

- Where did our candidate for T^{-1} come from? We seem to have pulled it out of a hat.
- Is there a simpler/better way to do this?

(b) Consider the **unilateral forward shift** operator $S \in \mathcal{L}(\ell^\infty(\mathbb{N}))$:

$$\begin{array}{rccc} S: & \ell^\infty(\mathbb{N}) & \to & \ell^\infty(\mathbb{N}) \\ & (x_n)_n & \mapsto & (0, x_1, x_2, \ldots). \end{array}$$

Define

$$\begin{array}{rccc} T: & \ell^\infty(\mathbb{N}) & \to & \ell^\infty(\mathbb{N}) \\ & (x_n)_n & \mapsto & (x_2, x_3, x_4, \ldots). \end{array}$$

Then $T \circ S = \mathrm{id}_{\ell^\infty(\mathbb{N})}$, so that S is **left-invertible**. Is S **invertible**? Why or why not? The operator T is typically referred to as the **unilateral backward shift**.

4.6. Proposition. *Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces over a field $\mathbb{F}$ and suppose that $T \in \mathcal{L}(\mathcal{V}, \mathcal{W})$ is invertible. Then $\mathcal{V}$ is finite-dimensional if and only if $\mathcal{W}$ is, in which case* $\dim \mathcal{V} = \dim \mathcal{W}$.

Proof. Suppose first that $\dim \mathcal{V} = n < \infty$. Let $\{v_1, v_2, \ldots, v_n\}$ be a basis for $\mathcal{V}$. Then T is surjective, and so $\mathcal{W} = \operatorname{span}\{Tv_1, Tv_2, \ldots, Tv_n\}$, implying that $\mathcal{W}$ is finite-dimensional and $\dim \mathcal{W} \leq n$. But T is injective and $\{v_1, v_2, \ldots, v_n\}$ is linearly independent, so that $\{Tv_1, Tv_2, \ldots, Tv_n\}$ is linearly independent in $\mathcal{W}$, whence $\dim \mathcal{W} \geq n$. Together, these imply that $\dim \mathcal{V} = \dim \mathcal{W}$.

Conversely, if $\dim \mathcal{W} = n < \infty$, then we apply the above argument using the inverse map $R = T^{-1} : \mathcal{W} \to \mathcal{V}$, since R is again invertible and linear. We conclude that $\dim \mathcal{V} = \dim \mathcal{W} < \infty$.

□

4.7. Theorem. *Suppose that $\mathcal{V}$ and $\mathcal{W}$ are finite-dimensional vector spaces over a field $\mathbb{F}$ and that $\dim \mathcal{V} = \dim \mathcal{W}$. Let $T \in \mathcal{L}(\mathcal{V}, \mathcal{W})$. The following are equivalent:*

(a) *T is injective.*
(b) *T is invertible.*
(c) *T is surjective.*

Proof. Recall by the Dimension Theorem that $\operatorname{nul} T + \operatorname{rank} T = \dim \mathcal{V}$.

(a) implies (b). Suppose that T is injective. Then $\operatorname{nul} T = 0$ and so $\operatorname{rank} T = \dim \mathcal{V} = \dim \mathcal{W} < \infty$. But then $\operatorname{ran} T = \mathcal{W}$, and so T is surjective as well.
(b) implies (c). This is a triviality.
(c) implies (a). Once again, from the Dimension Theorem, it easily follows that $\operatorname{nul} T = 0$, and thus $\ker T = \{\mathbf{0}\}$, or equivalently, T is injective.

□

4.8. Theorem. *Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces over a field $\mathbb{F}$, and let $T \in \mathcal{L}(\mathcal{V}, \mathcal{W})$.*

(a) *T is injective if and only if T is left-invertible.*
(b) *T is surjective if and only if T is right-invertible.*

Proof.

(a) Suppose that T is injective. Let $\mathcal{B}_\mathcal{V}$ be a basis for $\mathcal{V}$ and consider the set $\mathcal{D} := \{Tb : b \in \mathcal{B}_\mathcal{V}\} \subseteq \mathcal{W}$. Since $\mathcal{B}_\mathcal{V}$ is linearly independent and T is injective, we find that $\mathcal{D}$ is linearly independent. As such, we can extend $\mathcal{D}$ to a basis $\mathcal{B}_\mathcal{W} := \mathcal{D} \cup \{w_\alpha\}_{\alpha \in \Lambda}\}$ for $\mathcal{W}$.

We may then define a linear map $R \in \mathcal{L}(\mathcal{W}, \mathcal{V})$ by setting $R(Tb) := b$ for all $b \in \mathcal{B}_\mathcal{V}$, and $R(w_\alpha) = 0$ for all $\alpha \in \Lambda$, and extending this definition by linearity to all of $\mathcal{W}$.

Note that $R \circ T \in \mathcal{L}(\mathcal{V})$, and that for each $b \in \mathcal{B}_\mathcal{V}$,

$$(R \circ T)(b) = R(Tb) = b = I_\mathcal{V} b.$$

Since linear maps are completely determined by their actions on a basis for their domains, $R \circ T = I_\mathcal{V}$. Thus T is left-invertible.

Now suppose that T is left-invertible and choose $R \in \mathcal{L}(\mathcal{W}, \mathcal{V})$ such that $R \circ T = I_\mathcal{V}$. If $x \in \ker T$, then $x \in \ker R \circ T = \ker I_\mathcal{V} = \{\mathbf{0}\}$, so $x = \mathbf{0}$. Thus T is injective.

(b) Suppose that $T : \mathcal{V} \to \mathcal{W}$ is surjective. Let $\mathcal{B}_\mathcal{W}$ be a basis for $\mathcal{W}$, and for each $b \in \mathcal{B}_\mathcal{W}$, choose $v_b \in \mathcal{V}$ such that $Tv_b = b$. (This is possible because T is surjective.). Once again, we define $R \in \mathcal{L}(\mathcal{W}, \mathcal{V})$ by specifying its action

on the basis $\mathcal{B}_\mathcal{W}$ for $\mathcal{W}$ and extending by linearity to all of $\mathcal{W}$. In this case, we do this by setting $Rb = v_b$ for each $b \in \mathcal{B}_\mathcal{W}$.

Then for each $b \in \mathcal{B}_\mathcal{W}$, we have

$$(T \circ R)b = T(Rb) = Tv_b = b = I_\mathcal{W} b.$$

As before, since linear maps are completely determined by their actions on a basis for their domains, $T \circ R = I_\mathcal{W}$. Thus T is right-invertible.

Now suppose that T is right-invertible and choose $R \in \mathcal{L}(\mathcal{W}, \mathcal{V})$ such that $T \circ R = I_\mathcal{W}$. Since $I_\mathcal{W}$ is surjective, so is T.

□

By combining the previous two results, we easily obtain the following.

4.9. Corollary. *Suppose that $\mathcal{V}$ and $\mathcal{W}$ are finite-dimensional vector spaces over a field $\mathbb{F}$ and that* $\dim \mathcal{V} = \dim \mathcal{W}$. *Let $T \in \mathcal{L}(\mathcal{V}, \mathcal{W})$. The following are equivalent:*

(a) *T is left-invertible.*
(b) *T is invertible.*
(c) *T is right-invertible.*

4.10. Theorem. *Let $\mathcal{V}, \mathcal{W}, \mathcal{Y}$ and $\mathcal{Z}$ be finite-dimensional vector spaces over $\mathbb{F}$ and suppose that $R \in \mathcal{L}(\mathcal{V}, \mathcal{W})$, $T \in \mathcal{L}(\mathcal{W}, \mathcal{Y})$ and $S \in \mathcal{L}(\mathcal{Y}, \mathcal{Z})$. Then*

$$\mathit{rank}\,(STR) \leq \mathit{rank}\,(T).$$

If S and R are invertible, then

$$\mathit{rank}\,(STR) = \mathit{rank}\,(T).$$

Proof. It is clear that $\operatorname{ran}(TR) \subseteq \operatorname{ran}(T)$, and thus

$$\operatorname{ran}(STR) = \{Sy : y \in \operatorname{ran}(TR)\} \subseteq \{Sy : y \in \operatorname{ran}(T)\} = \operatorname{ran}(ST).$$

Thus $\operatorname{rank}(STR) \leq \operatorname{rank}(ST)$. If $\mathcal{L} := \{y_1, y_2, \ldots, y_p\}$ is a basis for $\operatorname{ran} T$ (so that $\operatorname{rank} T = p$, then $\{Sy_1, Sy_2, \ldots, Sy_p\}$ is a spanning set for $\operatorname{ran}(ST)$, and thus $\operatorname{rank}(ST) \leq p = \operatorname{rank} T$. Thus

$$\operatorname{rank}(STR) \leq \operatorname{rank}(ST) \leq \operatorname{rank}(T).$$

Now suppose that S, R are invertible. Although we won't appeal to it, observe that we then have $\dim \mathcal{V} = \dim \mathcal{W}$ and $\dim \mathcal{Y} = \dim \mathcal{Z}$. Observe that R is surjective, and thus

$$\operatorname{ran} T = \{Tw : w \in \mathcal{W}\} = \{TRv : v \in \mathcal{V}\} = \operatorname{ran}(TR).$$

In particular, $\operatorname{rank}(TR) = \operatorname{rank}(T)$. If $\mathcal{L} = \{y_1, y_2, \ldots, y_p\}$ is a basis for $\operatorname{ran}(T) = \operatorname{ran}(TR)$, then as before, $\{Sy_1, Sy_2, \ldots, Sy_p\}$ is a spanning set for $\operatorname{ran}(ST) = \operatorname{ran}(STR)$. But S is injective and $\mathcal{L}$ is linearly independent, implying that the set $\{Sy_1, Sy_2, \ldots, Sy_p\}$ is linearly independent, and hence a basis for $\operatorname{ran}(STR)$. Thus

$$\operatorname{rank}(STR) = \operatorname{rank}(ST) = p = \operatorname{rank}(T).$$

□

4.11. Remark. With the notation of the above theorem, we find that

$$\operatorname{rank}(TR) = \operatorname{rank}(I_{\mathcal{Y}}\ TR) \leq \operatorname{rank}(T).$$

On the other hand, Theorem 4.10 also implies that

$$\operatorname{rank}(TR) = \operatorname{rank}(TR\, I_{\mathcal{V}}) \leq \operatorname{rank}(R).$$

Thus, as a general rule, we have that

$$\operatorname{rank}(TR) \leq \min(\operatorname{rank}(T), \operatorname{rank}(R)).$$

In order to answer the questions from part (a) of Example 4.5, we turn our attention to matrices.

4.12. Definition. *Let $n \in \mathbb{N}$ and let $\mathbb{F}$ be a field. A matrix $T \in \mathbb{M}_n(\mathbb{F})$ is said to be* ***invertible*** *if there exists $R \in \mathbb{M}_n(\mathbb{F})$ such that $TR = I_n = RT$.*

4.13. Remark. Recall from Definition 3.8 that for each $A \in \mathbb{M}_n(\mathbb{F})$, we may define the **left multiplication operator** $L_A \in \mathcal{L}(\mathbb{F}^n)$ via $L_A x = Ax$, $x \in \mathbb{F}^n$.

If $A \in \mathbb{M}_n(\mathbb{F})$ is invertible and $R = A^{-1} \in \mathbb{M}_n(\mathbb{F})$, then

$$L_R L_A = L_{RA} = L_{I_n} = I_{\mathbb{F}^n} = L_{I_n} = L_{AR} = L_A L_R.$$

Thus $A \in \mathbb{M}_n(\mathbb{F})$ invertible implies that $L_A \in \mathcal{L}(\mathbb{F}^n)$ is invertible with inverse $L_{A^{-1}}$.

Conversely, suppose that $L_A \in \mathcal{L}(\mathbb{F}^n)$ is invertible and that $T \in \mathcal{L}(\mathbb{F}^n) = (L_A)^{-1}$. Let $\mathcal{D}$ denote the standard ordered basis for $\mathbb{F}^n$, and recall that $[L_A]_{\mathcal{D}} = A$. Thus

$$\begin{aligned}[T]_{\mathcal{D}} A = [T]_{\mathcal{D}}\ [L_A]_{\mathcal{D}} = [TL_A]_{\mathcal{D}} = [I_{\mathcal{V}}]_{\mathcal{D}} = I_n \\ = [I_{\mathcal{V}}]_{\mathcal{D}} = [L_A T]_{\mathcal{D}} \\ = [L_A]_{\mathcal{D}}\ [T]_{\mathcal{D}} = A\ [T]_{\mathcal{D}}.\end{aligned}$$

Thus $L_A \in \mathcal{L}(\mathbb{F}^n)$ invertible implies that $A \in \mathbb{M}_n(\mathbb{F})$ is invertible with inverse $[(L_A)^{-1}]_{\mathcal{D}}$, where $\mathcal{D}$ is the standard ordered basis for $\mathbb{F}^n$.

4.14. Example. The matrix $A = \begin{bmatrix} 3 & 1 \\ 1 & 4 \end{bmatrix} \in \mathbb{M}_2(\mathbb{Q})$ is invertible with inverse

$$A^{-1} = \frac{1}{11}\begin{bmatrix} 4 & -1 \\ -1 & 3 \end{bmatrix}.$$

We now extend the above analysis to obtain our key result regarding invertibility of linear maps on *finite-dimensional spaces*. It says that to determine if a linear map T between finite-dimensional spaces is invertible, one need only verify whether its matrix with respect to (any) pair of ordered bases is invertible, and if it is, then one can use matrix inversion to determine the inverse of T.

4.15. Theorem. *Suppose that $\mathcal{V}$ and $\mathcal{W}$ are finite-dimensional vector spaces over a field $\mathbb{F}$. Let $\mathcal{D}$ and $\mathcal{C}$ be ordered bases for $\mathcal{V}$ and $\mathcal{W}$ respectively, and suppose that $T \in \mathcal{L}(\mathcal{V}, \mathcal{W})$.*

Then T is invertible if and only if $[T]_{\mathcal{D}}^{\mathcal{C}}$ is invertible, in which case

$$[T^{-1}]_{\mathcal{C}}^{\mathcal{D}} = ([T]_{\mathcal{D}}^{\mathcal{C}})^{-1}.$$

Proof. Suppose first that $T \in \mathcal{L}(\mathcal{V}, \mathcal{W})$ is invertible. As we have just seen, this implies that $n := \dim \mathcal{V} = \dim \mathcal{W}$. Now

$$I_n = [I_{\mathcal{V}}]_{\mathcal{D}} = [T^{-1}T]_{\mathcal{D}} = [T^{-1}]_{\mathcal{C}}^{\mathcal{D}}\ [T]_{\mathcal{D}}^{\mathcal{C}},$$

and

$$I_n = [I_{\mathcal{W}}]_{\mathcal{C}} = [TT^{-1}]_{\mathcal{C}} = [T]_{\mathcal{D}}^{\mathcal{C}}\ [T^{-1}]_{\mathcal{C}}^{\mathcal{D}}.$$

Thus

$$([T]_{\mathcal{D}}^{\mathcal{C}})^{-1} = [T^{-1}]_{\mathcal{C}}^{\mathcal{D}}.$$

Now suppose that $[T]_{\mathcal{D}}^{\mathcal{C}}$ is invertible in $\mathbb{M}_n(\mathbb{F})$ with inverse $[r_{ij}] \in \mathbb{M}_n(\mathbb{F})$. Recall that given a basis (e.g. $\mathcal{C} = \{w_1, w_2, \ldots, w_n\}$) for a vector space $\mathcal{W}$, we may define a linear map on that space by specifying what it does to that basis and extending by linearity.

In our case, we set $Rw_j := \sum_{i=1}^{n} r_{ij} v_i$, $1 \le j \le n$. Then $[R]_{\mathcal{C}}^{\mathcal{D}} = [r_{ij}]$, and so

$$[RT]_{\mathcal{D}} = [R]_{\mathcal{C}}^{\mathcal{D}}\ [T]_{\mathcal{D}}^{\mathcal{C}} = [r_{ij}]\,[T]_{\mathcal{D}}^{\mathcal{C}} = I_n = [I_{\mathcal{V}}]_{\mathcal{D}},$$

so that $RT = I_{\mathcal{V}}$. Similarly,

$$[TR]_{\mathcal{C}} = [T]_{\mathcal{D}}^{\mathcal{C}}\ [R]_{\mathcal{C}}^{\mathcal{D}} = [T]_{\mathcal{D}}^{\mathcal{C}}\ [r_{ij}] = I_n = [I_{\mathcal{W}}]_{\mathcal{C}},$$

and so $TR = I_{\mathcal{W}}$. That is, T is invertible with inverse R.

□

4.16. Example. Let

$$\begin{array}{rccc} T: & \mathbb{R}^2 & \to & \mathbb{R}^2 \\ & (x,y) & \mapsto & (3x+y, x+4y). \end{array}$$

Let $\mathcal{D} = \mathcal{C}$ denote the standard ordered basis for $\mathbb{R}^2$. Then $[T]_{\mathcal{D}} = \begin{bmatrix} 3 & 1 \\ 1 & 4 \end{bmatrix}$. By Example 4.14,

$$[T^{-1}]_{\mathcal{D}} = \frac{1}{11}\begin{bmatrix} 4 & -1 \\ -1 & 3 \end{bmatrix},$$

and thus

$$T^{-1}(x,y) = \left(\frac{4}{11}x - \frac{1}{11}y, -\frac{1}{11}x + \frac{3}{11}y\right).$$

The moral of the story is that if we can compute inverses of matrices, then we have a very effective way of computing inverses of linear maps. We are interested in linear maps – and matrices will be the tool we use to understand these maps!

If X and Y are non-empty sets and $f : X \to Y$ is a bijection, then we can think of Y as being a *relabelled version* of X. Of course, sets have no algebraic structure in general, so we shouldn't ask that our bijective maps be anything other than exactly that – bijective maps. When dealing with vector spaces, we place more stringent conditions on our bijections.

4.17. Definition. *Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces over the field $\mathbb{F}$. We say that $\mathcal{V}$ is **isomorphic** to $\mathcal{W}$, and we write $\mathcal{V} \simeq \mathcal{W}$, if there exists an invertible linear map $T : \mathcal{V} \to \mathcal{W}$. We say that T is an **isomorphism** from $\mathcal{V}$ to $\mathcal{W}$.*

4.18. Remark. The relation $\simeq$ is an **equivalence relation** on the class of all vector spaces. That is,

(a) The identity map $I_{\mathcal{V}}$ implements the isomorphism of $\mathcal{V}$ with $\mathcal{V}$, so $\mathcal{V} \simeq \mathcal{V}$ whenever $\mathcal{V}$ is a vector space.
(b) If $\mathcal{V} \simeq \mathcal{W}$ via $T : \mathcal{V} \to \mathcal{W}$, then $\mathcal{W} \simeq \mathcal{V}$ via $T^{-1} : \mathcal{W} \to \mathcal{V}$. Thus $\mathcal{V} \simeq \mathcal{W}$ implies $\mathcal{W} \simeq \mathcal{V}$.
(c) If $\mathcal{V} \simeq \mathcal{W}$ via $T : \mathcal{V} \to \mathcal{W}$ and $\mathcal{W} \simeq \mathcal{Y}$ via $R : \mathcal{W} \to \mathcal{Y}$, then $\mathcal{W} \simeq \mathcal{Y}$ via $R \circ T : \mathcal{V} \to \mathcal{Y}$. Thus $\mathcal{V} \simeq \mathcal{Y}$ and $\mathcal{Y} \simeq \mathcal{Y}$ implies that $\mathcal{V} \simeq \mathcal{Y}$.

Vector spaces are truly wondrous objects. Up to isomorphism, they are completely determined by their dimension! The technical way to say this is: the cardinality of a vector space basis is a complete invariant for vector spaces up to isomorphism. We are now officially cool.

4.19. Theorem. *Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces over a field $\mathbb{F}$. The following conditions are equivalent.*

(a) $\mathcal{V} \simeq \mathcal{W}$.
(b) *There exist bases $\mathcal{B}_{\mathcal{V}}$ for $\mathcal{V}$ and $\mathcal{B}_{\mathcal{W}}$ for $\mathcal{W}$ and a bijection $\Phi : \mathcal{B}_{\mathcal{V}} \to \mathcal{B}_{\mathcal{W}}$.*

Consequently, if $\dim \mathcal{V} = n \in \mathbb{N}$, *then* $\mathcal{V} \simeq \mathbb{F}^n$.

Remark. Of course, when $\mathcal{V}$ and $\mathcal{W}$ are finite-dimensional, condition (b) is merely the statement that $\dim \mathcal{V} = \dim \mathcal{W}$, and then we can choose *any* bases $\mathcal{B}_{\mathcal{V}}$ for $\mathcal{V}$ and $\mathcal{B}_{\mathcal{W}}$ for $\mathcal{W}$.

In fact, it can be shown that if $\mathcal{Z}$ is *any* vector space over $\mathbb{F}$ and if $\mathcal{B}_1$ and $\mathcal{B}_2$ are bases for $\mathcal{Z}$, then there exists a bijection $f : \mathcal{B}_1 \to \mathcal{B}_2$. This allows us to define the *dimension* of an infinite-dimensional vector space as the *cardinality* of any of its bases (which one may think of as the single equivalence class under bijective correspondence of all bases for $\mathcal{Z}$). Using this notion, we find that once again, (b) becomes the statement that $\dim \mathcal{V} = \dim \mathcal{W}$.

Proof.

(a) implies (b). Suppose first that $\mathcal{V} \simeq \mathcal{W}$. Let $T : \mathcal{V} \to \mathcal{W}$ be an invertible map, and let $\mathcal{B}_{\mathcal{V}} = \{v_\lambda\}_{\lambda \in \Lambda}$. Define $\mathcal{B}_{\mathcal{W}} := \{Tv_\lambda\}_{\lambda \in \Lambda}$. Clearly T is a bijection between $\mathcal{B}_{\mathcal{V}}$ and $\mathcal{B}_{\mathcal{W}}$, so there remains only to show that $\mathcal{B}_{\mathcal{W}}$ is a basis for $\mathcal{W}$.

Now T is surjective and $\mathcal{B}_\mathcal{V}$ generates $\mathcal{V}$, so $\mathcal{B}_\mathcal{W}$ generates $\mathcal{W}$. Also, T is injective and $\mathcal{B}_\mathcal{V}$ is linearly independent, so $\mathcal{B}_\mathcal{W}$ is linearly independent. It follows that $\mathcal{B}_\mathcal{W}$ is indeed a basis for $\mathcal{W}$, completing the proof of this half of the Theorem.

(b) implies (a). Now suppose that $\mathcal{B}_\mathcal{V}$ and $\mathcal{B}_\mathcal{W}$ are bases for $\mathcal{V}$ and $\mathcal{W}$ respectively, and that $f : \mathcal{B}_\mathcal{V} \to \mathcal{B}_\mathcal{W}$ is a bijection. We define a linear map $T : \mathcal{V} \to \mathcal{W}$ by setting $Tb := f(b)$ for all $b \in \mathcal{B}_\mathcal{V}$, and extending T to all of $\mathcal{V}$ by linearity.

Since f is surjective,

$$\operatorname{ran} T = \operatorname{span}\{Tb : b \in \mathcal{B}_\mathcal{V}\} = \operatorname{span}\{d : d \in \mathcal{B}_\mathcal{W}\} = \mathcal{W}.$$

Thus T is surjective. If $x \in \ker T$, then we may write $x = \sum_{j=1}^{n} \kappa_j b_j$ for some $b_1, b_2, \ldots, b_n \in \mathcal{B}_\mathcal{V}$ and $\kappa_1, \kappa_2, \ldots, \kappa_n \in \mathbb{F}$. Then

$$\mathbf{0}_\mathcal{W} = T(\sum_{j=1}^{n} \kappa_j b_j) = \sum_{j=1}^{n} \kappa_j T b_j = \sum_{j=1}^{n} \kappa_j f(b_j).$$

But $f(b_j) \in \mathcal{B}_\mathcal{W}$ for all $1 \le j \le n$, and f is injective. Thus

$$\{f(b_1), f(b_2), \ldots, f(b_n)\} \subseteq \mathcal{B}_\mathcal{W}$$

is linearly independent, which implies that $\kappa_j = 0$, $1 \le j \le n$, and therefore $x = \mathbf{0}_\mathcal{V}$. Hence T is injective, so it is an isomorphism of $\mathcal{V}$ onto $\mathcal{W}$.

□

4.20. Definition. *Let $\mathcal{V}$ be an n-dimensional vector space over $\mathbb{F}$ and let $\mathcal{D}$ be an ordered basis for $\mathcal{V}$. The* ***standard representation*** *of $\mathcal{V}$* ***with respect to*** *$\mathcal{D}$ is the (soon to be proven to be linear) map:*

$$\begin{array}{rccc} \varrho_\mathcal{D} : & \mathcal{V} & \to & \mathbb{F}^n \\ & x & \mapsto & [x]_\mathcal{D}. \end{array}$$

4.21. Theorem. *Let $\mathcal{V}$ be an n-dimensional vector space over $\mathbb{F}$ and let $\mathcal{D}$ be an ordered basis for $\mathcal{V}$. The standard representation $\varrho_\mathcal{D} : \mathcal{V} \to \mathbb{F}^n$ of $\mathcal{V}$ with respect to $\mathcal{D}$ is an isomorphism.*

Proof. Write $\mathcal{D} = (d_1, d_2, \ldots, d_n)$ and let $(e_1, e_2, \ldots, e_n)$ be the standard basis for $\mathbb{F}^n$. Let $x = \sum_{j=1}^n \alpha_j d_j$ and $y = \sum_{j=1}^n \beta_j d_j \in \mathcal{V}$, and let $\kappa \in \mathbb{F}$. Then

$$\begin{aligned}
\varrho_{\mathcal{D}}(\kappa x + y) &= \varrho_{\mathcal{D}}\Big(\sum_{j=1}^n (\kappa\alpha_j + \beta_j) d_j\Big) \\
&= \begin{bmatrix} \kappa\alpha_1 + \beta_1 \\ \kappa\alpha_2 + \beta_2 \\ \vdots \\ \kappa\alpha_n + \beta_n \end{bmatrix} \\
&= \kappa \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} + \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix} \\
&= \kappa \varrho_{\mathcal{D}}(x) + \varrho_{\mathcal{D}}(y).
\end{aligned}$$

Thus $\varrho_{\mathcal{D}}$ is linear.

If $x \in \ker \varrho_{\mathcal{D}}$, then $\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \varrho_{\mathcal{D}}(x)$, so $x = \sum_{j=1}^n 0\, d_j = \mathbf{0}$. Thus $\ker \varrho_{\mathcal{D}} = \{\mathbf{0}\}$, and so $\varrho_{\mathcal{D}}$ is injective. But $\dim \mathcal{V} = n = \dim \mathbb{F}^n$, so that $\varrho_{\mathcal{D}}$ is also surjective, and thus $\varrho_{\mathcal{D}}$ is an isomorphism.

□

4.22. Example. Consider $\mathcal{V} = \mathbb{R}_2[x]$, and $\mathcal{D} = (1, x, x^2)$. Then $\varrho_{\mathcal{D}} : \mathbb{R}_2[x] \to \mathbb{R}^3$ is the map

$$\varrho_{\mathcal{D}}(p_0 + p_1 x + p_2 x^2) = \begin{bmatrix} p_0 \\ p_1 \\ p_2 \end{bmatrix}.$$

We may now reinterpret Theorem 2.10 in terms of isomorphisms.

4.23. Theorem. *Let $\mathcal{V}$ and $\mathcal{W}$ be finite-dimensional vector spaces of dimensions n and m respectively. For each choice of ordered bases $\mathcal{D}$ and $\mathcal{C}$ for $\mathcal{V}$ and $\mathcal{W}$ respectively, the map*

$$\begin{array}{rccc} \Phi : & \mathcal{L}(\mathcal{V}, \mathcal{W}) & \to & \mathbb{M}_{m\times n}(\mathbb{F}) \\ & T & \mapsto & [T]_{\mathcal{D}}^{\mathcal{C}} \end{array}$$

is an isomorphism of vectors spaces.

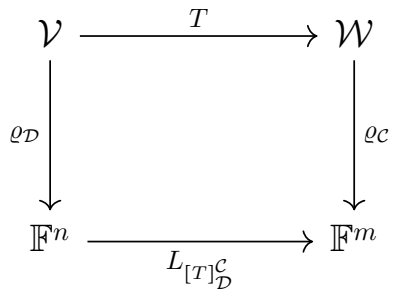
In particular, $\dim \mathcal{L}(\mathcal{V}, \mathcal{W}) = \dim \mathbb{M}_{m\times n}(\mathbb{F}) = mn$.

4.24. Remark. When $\mathcal{V} = \mathcal{W}$ and $\mathcal{D} = \mathcal{C}$, the map Φ satisfies an extra property, namely: for all $R, T \in \mathcal{L}(\mathcal{V})$,

$$\Phi(TR) = [TR]_{\mathcal{D}} = [T]_{\mathcal{D}}\,[R]_{\mathcal{D}} = \Phi(T)\,\Phi(R).$$

We say that Φ is **multiplicative**.

4.25. We may also reinterpret Theorem 3.12 as saying that **the following diagram commutes**: that is, $\varrho_{\mathcal{C}} \circ T = L_{[T]_{\mathcal{D}}^{\mathcal{C}}} \circ \varrho_{\mathcal{D}}$.

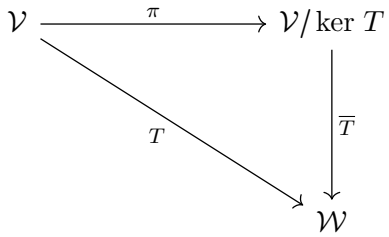


4.26. Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces over a field $\mathbb{F}$, and suppose that $T : \mathcal{V} \to \mathcal{W}$ is a **surjective** linear map. Define

$$\begin{array}{rccc} \overline{T}: & \mathcal{V}/\ker T & \to & \mathcal{W} \\ & x + \ker T & \mapsto & Tx. \end{array}$$

4.27. Proposition. *Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces over a field $\mathbb{F}$, and suppose that $T : \mathcal{V} \to \mathcal{W}$ is a **surjective** linear map. Let $\pi : \mathcal{V} \to \mathcal{V}/\ker T$ denote the canonical quotient map $\pi(x) = x + \ker T$ for all $x \in \mathcal{V}$. Then:*

(a) *$\overline{T}$ is well-defined.*
(b) *$\overline{T}$ is linear.*
(c) *$\overline{T}$ is an isomorphism.*
(d) *The diagram*



commutes. That is, $T = \overline{T} \circ \pi$.

Proof.

(a) Suppose that $x_1 + \ker\ T = x_2 + \ker T$. Then $x_2 - x_1 \in \ker T$, so $Tx_1 - Tx_2 = T(x_1 - x_2) = \mathbf{0}$. That is,

$$\overline{T}(x_1 + \ker, T) = Tx_1 = Tx_2 = \overline{T}(x_2 + \ker T),$$

and so $\overline{T}$ is well-defined.

(b) Let $x + \ker T, y + \ker T \in \mathcal{V}/\ker T$, and let $\kappa \in \mathbb{F}$. Then

$$\begin{aligned}\overline{T}(\kappa(x + \ker T) + (y + \ker T)) &= \overline{T}((\kappa x + y) + \ker T)\\ &= T(\kappa x + y)\\ &= \kappa Tx + Ty\\ &= \kappa\overline{T}(x + \ker T) + \overline{T}(y + \ker T).\end{aligned}$$

Thus $\overline{T}$ is linear.

(c) Suppose that $x + \ker T \in \ker \overline{T}$. Then $Tx = \overline{T}(x + \ker T) = \mathbf{0}$, so that $x \in \ker T$. But then $x + \ker T = \mathbf{0} + \ker T$. Hence $\overline{T}$ is injective.

For any $w \in \mathcal{W}$, we have that $w = Tx$ for some $x \in \mathcal{V}$ as T is surjective. But then $w = Tx = \overline{T}(x + \ker T)$, and so $\overline{T}$ is also surjective.

(d) For any $x \in \mathcal{V}$,

$$Tx = \overline{T}(x + \ker T) = \overline{T} \circ \pi(x),$$

whence $T = \overline{T} \circ \pi$.

□

5. Change of basis

5.1. On an abstract level, all bases for a given vector space are created equal. Having said that – depending on the situation, and to coin a phrase – some bases might be more equal than others. For example, while $\{1, x, x^2, \ldots, x^n\}$ might seem like the most natural basis for $\mathbb{F}_n[x]$, we have seen that Lagrange polynomials can be extremely useful on a computational level.

A natural question becomes: how do we translate coordinates from one basis to another?

5.2. Theorem. *Let $\mathcal{D}$ and $\mathcal{C}$ be two ordered bases for a finite-dimensional vector space $\mathcal{V}$. Define the matrix $Q := [I_\mathcal{V}]^\mathcal{C}_\mathcal{D}$. Then*

(a) *Q is invertible, with $Q^{-1} = [I_\mathcal{V}]^\mathcal{D}_\mathcal{C}$, and*
(b) *for $v \in \mathcal{V}$, $[v]_\mathcal{C} = Q[v]_\mathcal{D}$.*

Proof.

(a) This is an immediate consequence of Theorem 4.15, as the identity map $I_\mathcal{V}$ is obviously invertible in $\mathcal{L}(\mathcal{V})$ – it is its own inverse!

(b) This is Theorem 3.12 applied to the map $I_\mathcal{V}$.

□

5.3. The matrix $Q = [I_{\mathcal{V}}]_{\mathcal{D}}^{\mathcal{C}}$ is called the **change of coordinate matrix**, and it converts $\mathcal{D}$-coordinates into $\mathcal{C}$-coordinates. If $Q = [q_{ij}]$, $\mathcal{D} = (v_1, v_2, \ldots, v_n)$ and $\mathcal{C} = (w_1, w_2, \ldots, w_n)$, then for each $1 \le j \le n$,

$$v_j = \sum_{i=1}^{n} q_{ij} w_i.$$

5.4. Example. Let $\mathcal{V} = \mathbb{R}^2$, $D := ((1,1),(1,-1))$ and $\mathcal{C} = ((1,2),(2,1))$ as ordered bases for $\mathcal{V}$. Then

$$\begin{aligned}[I_{\mathcal{V}}]_{\mathcal{D}}^{\mathcal{C}} &= [[v_1]_{\mathcal{C}} : [v_2]_{\mathcal{C}}] \\ &= [[(1,1)]_{\mathcal{C}} : [(1,-1)]_{\mathcal{C}}] \\ &= \begin{bmatrix} 1/3 & -1 \\ 1/3 & 1 \end{bmatrix}.\end{aligned}$$

Hence

$$[(1,1)]_{\mathcal{C}} = Q[(1,1)]_{\mathcal{D}} = Q\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1/3 \\ 1/3 \end{bmatrix}.$$

Once again – we don't have to accept this on faith – we can check! The vector $\begin{bmatrix} 1/3 \\ 1/3 \end{bmatrix}$ relative to $\mathcal{C}$-coordinates represents

$$\frac{1}{3}w_1 + \frac{1}{3}w_2 = \frac{1}{3}(1,2) + \frac{1}{3}(2,1) = (1,1).$$

Having established a relationship between coordinates of a vector in a vector space with respect to two ordered bases, we now consider the relationship between the matrix of a linear map $T \in \mathcal{L}(\mathcal{V})$ with respect to two ordered bases. The following is an immediate result of how we *defined* the product of two (and hence of finitely many) matrices.

5.5. Theorem. *Let $\mathcal{V}$ be a finite-dimensional vector space and $T \in \mathcal{L}(\mathcal{V})$. Let $\mathcal{D}$ and $\mathcal{V}$ be ordered bases for $\mathcal{V}$, and set $Q := [I_{\mathcal{V}}]_{\mathcal{D}}^{\mathcal{C}}$. Then*

$$[T]_{\mathcal{D}} = [I_{\mathcal{V}}]_{\mathcal{C}}^{\mathcal{D}}\ [T]_{\mathcal{C}}\ [I_{\mathcal{V}}]_{\mathcal{D}}^{\mathcal{C}} = Q^{-1}\ [T]_{\mathcal{C}}\ Q.$$

Proof. Simply note that $T = I_{\mathcal{V}} \circ T \circ I_{\mathcal{V}}$, and thus

$$[T]_{\mathcal{D}} = [I_{\mathcal{V}} \circ T \circ I_{\mathcal{V}}]_{\mathcal{D}} = [I_{\mathcal{V}} \circ T]_{\mathcal{C}}^{\mathcal{D}}\ [I_{\mathcal{V}}]_{\mathcal{D}}^{\mathcal{C}} = [I_{\mathcal{V}}]_{\mathcal{C}}^{\mathcal{D}}\ [T]_{\mathcal{C}}\ [I_{\mathcal{V}}]_{\mathcal{D}}^{\mathcal{C}} = Q^{-1}\ [T]_{\mathcal{C}}\ Q.$$

□

5.6. Definition. *Two matrices $A, B \in \mathbb{M}_n(\mathbb{F})$ are said to be* ***similar*** *if there exists $R \in \mathbb{M}_n(\mathbb{F})$ invertible such that $B = S^{-1}AS$. We write $A \sim B$ if A is similar to B.*

If $\mathcal{V}$ is a vector space and $S, T \in \mathcal{L}(\mathcal{V})$, then S and T are said to be ***similar*** *if there exists an invertible linear map $Y \in \mathcal{L}(\mathcal{V})$ such that $T = Y^{-1}SY$. We also write $S \sim T$ if S is similar to T.*

5.7. Remarks.

(a) Similarity of matrices is an equivalence relation on $\mathbb{M}_n(\mathbb{F})$ and similarity of linear maps is an equivalence relation on $\mathcal{L}(\mathcal{V})$. Thus for all $A_1, A_2, A_3 \in \mathbb{M}_n(\mathbb{F})$, we have
- $A_1 \sim A_1$;
- if $A_1 \sim A_2$, then $A_2 \sim A_1$; and
- if $A_1 \sim A_2$ and $A_2 \sim A_3$, then $A_1 \sim A_3$.

The proof of these three conditions is left as an exercise. The same analysis applies to similarity of linear maps in $\mathcal{L}(\mathcal{V})$.

(b) Theorem 5.5 is the statement that if $\mathcal{V}$ is a finite-dimensional vector space, $T \in \mathcal{L}(\mathcal{V})$, and $\mathcal{D}$ and $\mathcal{C}$ are ordered bases for $\mathcal{V}$, then $[T]_{\mathcal{D}}$ and $[T]_{\mathcal{C}}$ are similar.

(c) If $\mathcal{V}$ is n-dimensional with ordered basis $\mathcal{D}$, and if $T, R \in \mathcal{L}(\mathcal{V})$, then T is similar to R if and only if $[T]_{\mathcal{D}}$ is similar to $[R]_{\mathcal{D}}$.

5.8. Example. Define the map

$$\begin{array}{rccc} T: & \mathbb{R}^2 & \to & \mathbb{R}^2 \\ & (x,y) & \mapsto & (\frac{1}{2}x + \frac{1}{2}y, \frac{1}{2}x + \frac{1}{2}y). \end{array}$$

Let $\mathcal{D} = ((1,0),(0,1))$ be the standard ordered basis for $\mathbb{F}^2$, and let $\mathcal{C} = ((1,1),(1,-1))$ be a second ordered basis for $\mathbb{R}^2$.

Then $[T]_{\mathcal{D}} = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$. Moreover

$$\begin{aligned} Q &= [I_{\mathcal{V}}]_{\mathcal{D}}^{\mathcal{C}} \\ &= [[(1,0)]_{\mathcal{C}} : [(0,1)]_{\mathcal{C}}] \\ &= \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}. \end{aligned}$$

Meanwhile,

$$\begin{aligned} Q^{-1} = [I_{\mathcal{V}}]_{\mathcal{C}}^{\mathcal{D}} &= [[(1,1)]_{\mathcal{D}} : [(1,-1)]_{\mathcal{D}}] \\ &= \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}. \end{aligned}$$

Thus

$$[T]_{\mathcal{C}} = Q^{-1}\,[T]_{\mathcal{D}}\,Q = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}\frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}.$$

5.9. Example. Let $A \in \mathbb{M}_n(\mathbb{F})$ and $\mathcal{D} = (v_1, v_2, \ldots, v_n)$ be an ordered basis for $\mathbb{F}^n$. Let $\mathcal{B} = (e_1, e_2, \ldots, e_n)$ denote the standard ordered basis for $\mathbb{F}^n$, and consider the left regular representation $L_A : \mathbb{F}^n \to \mathbb{F}^n$ defined by $L_A x = Ax$, $x \in \mathbb{F}^n$.

As we have seen in Exercise 3.10, $[L_A]_{\mathcal{B}} = A$. It now follows that

$$[L_A]_{\mathcal{D}} = [I_{\mathbb{F}^n}]_{\mathcal{B}}^{\mathcal{D}} \, [L_A]_{\mathcal{B}} \, [I_{\mathbb{F}^n}]_{\mathcal{D}}^{\mathcal{B}} = Q^{-1} A Q,$$

where $Q = [I_{\mathbb{F}^n}]_{\mathcal{D}}^{\mathcal{B}}$.

But $Q = [I_{\mathbb{F}^n}]_{\mathcal{D}}^{\mathcal{B}}$ is the matrix whose j^{th} column consist of the entries of v_j, $1 \le j \le n$.

Let us consider a concrete example. Suppose that

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

Let $\mathcal{D} = ((1,1,1),(0,1,1),(0,0,1))$. Then

$$\begin{aligned} [L_A]_{\mathcal{D}} &= \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ -1 & -2 & -3 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ -1 & -2 & -3 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 6 & 5 & 3 \\ 9 & 6 & 3 \\ -21 & -16 & -9 \end{bmatrix}. \end{aligned}$$

Supplementary Examples

S5.1. Example. Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces over a common field $\mathbb{F}$, and let $\mathcal{F} := \{f : \mathcal{V} \to \mathcal{W} \mid f \text{ is a function}\}$. For $f, g \in \mathcal{F}$ and $\kappa \in \mathbb{F}$, define $(f+g)(x) = f(x) + g(x)$ and $(\kappa f)(x) := \kappa(f(x))$ for all $x \in \mathcal{V}$.

It is tedious but simple to verify that $\mathcal{F}$ becomes a vector space over $\mathbb{F}$. If one does this, then one may then apply the Subspace Test to $\mathcal{L}(\mathcal{V}, \mathcal{W})$ to obtain that it too is a vector space over $\mathbb{F}$.

S5.2. Example. Let us recall the *rotation operator* R_θ from Example 1.6. That is, given $\theta \in \mathbb{R}$, we define

$$\begin{array}{rccc} R_\theta : & \mathbb{R}^2 & \to & \mathbb{R}^2 \\ & (x,y) & \mapsto & (x\cos\theta - y\sin\theta, x\sin\theta + y\cos\theta). \end{array}$$

Relative to the standard ordered basis $\mathcal{D} = \mathcal{C} = (e_1, e_2)$ for $\mathbb{R}^2$, we see that

$$R_\theta(e_1) = R_\theta \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix},$$

while

$$R_\theta(e_2) = R_\theta \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} -\sin\theta \\ \cos\theta \end{bmatrix}.$$

Since the coordinates of a vector in $\mathbb{R}^m$ relative to the standard basis look like the vector itself, we see that

$$[R_\theta]_{\mathcal{D}}^{\mathcal{C}} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}.$$

S5.3. Example. Let us next consider the projection operator P_y from Example 1.6 (c), namely:

$$\begin{array}{rccc} P_2 : & \mathbb{R}^2 & \to & \mathbb{R}^2 \\ & (x,y) & \mapsto & (0,y). \end{array}$$

Again, if we let $\mathcal{D} = \mathcal{C} = (e_1, e_2)$ be the standard ordered basis for $\mathbb{R}^2$, then

$$P_2(e_1) = P_y(1,0) = (0,0),$$

and

$$P_2(e_2) = P_y(0,1) = (0,1).$$

Arguing as in the above example,

$$[P_y]_{\mathcal{D}}^{\mathcal{C}} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

We leave it to the reader to show that if $P_1(x,y) := (x,0)$ denotes the projection of $\mathbb{R}^2$ onto the x-axis, then relative to $\mathcal{D} = \mathcal{C} = (e_1, e_2)$,

$$[P_x]_{\mathcal{D}}^{\mathcal{D}} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}.$$

S5.4. Example. Let $\mathcal{D} = ((1,2),(1,3))$ and $\mathcal{C} = ((3,23),(1,-5))$ be ordered bases for $\mathbb{R}^2$. Suppose that

$$T(x,y) = (4x - y, -x + 5y), \qquad (x,y) \in \mathbb{R}^2.$$

Then

$$T(1,2) = (2,9) = \frac{1}{2}(3,23) + \frac{1}{2}(1,-5)$$
$$T(1,3) = (1,14) = \frac{1}{2}(3,23) - \frac{1}{2}(1,-5).$$

Thus

$$[T]_{\mathcal{D}}^{\mathcal{C}} = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & -1/2 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}.$$

S5.5. Example. Let us find the inverse of the matrix $A := \begin{bmatrix} 1 & 1 & -1 \\ 2 & 1 & -1 \\ 1 & 1 & 1 \end{bmatrix}$.

We must solve

$$\begin{bmatrix} 1 & 1 & -1 \\ 2 & 1 & -1 \\ 1 & 1 & 1 \end{bmatrix}\begin{bmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ z_1 & z_2 & z_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

This leads us to solve three systems of three equations in three variables. We shall solve for x_1, y_1, z_1 and leave the other cases as an exercise.

By calculating the first column of the product, we have

S5.6. Example. Let $2 \le n \in \mathbb{N}$ and $\mathcal{V} = \mathbb{T}_n(\mathbb{F})$. Define $\mathcal{W} := \{w = [w_{ij}] \in \mathbb{T}_n(\mathbb{F}) : w_{kk} = 0, 1 \le k \le n\}$.

Consider the map

$$\begin{array}{rccc} T: & \mathcal{V} & \to & \mathbb{F}^n \\ & [x_{ij}] & \mapsto & (x_{11}, x_{22}, \ldots, x_{nn}). \end{array}$$

If $a = [a_{ij}, b = [b_{ij}] \in \mathbb{T}_n(\mathbb{F})$ and $\kappa \in \mathbb{F}$, then

$$\begin{aligned} T(\kappa a + b) &= T(\kappa[a_{ij}] + [b_{ij}]) \\ &= T([\kappa a_{ij} + b_{ij}]) \\ &= (\kappa a_{11} + b_{11}, \kappa a_{22} + b_{22}, \ldots, \kappa a_{nn} + b_{nn}) \\ &= \kappa(a_{11}, a_{22}, \ldots, a_{nn}) + (b_{11}, b_{22}, \ldots, b_{nn}) \\ &= \kappa\, Ta + Tb, \end{aligned}$$

proving that T is linear.

Now $a \in \ker T$ if and only if $(a_{11}, a_{22}, \ldots, a_{nn}) = Ta = (0, 0, \ldots, 0)$; that is, if and only if $a \in \mathcal{W}$. Thus $\ker T = \mathcal{W}$. Also, if $x = (x_{11}, x_{22}, \ldots, x_{nn}) \in \mathbb{F}^n$ and $d := [d_{ij}]$, where $d_{kk} = x_{kk}$, $1 \le k \le n$ and $d_{ij} = 0$ if $i \ne j$, then $Td = x$, so that T is surjective.

It follows from the First Isomorphism Theorem for vector spaces that

$$\mathbb{T}_n(\mathbb{F})/\mathcal{W} = \mathbb{T}_n(\mathbb{F})/\ker T \simeq \operatorname{ran} T = \mathbb{F}^n.$$

S5.7. Example. Let $\mathcal{V} := \mathcal{C}^1((0,1),\mathbb{R}) := \{f : (0,1) \to \mathbb{R} : f' \text{ is continuous on } (0,1)\}$. Let

$$\begin{array}{rccc} D: & \mathcal{V} & \to & \mathcal{C}((0,1),\mathbb{R}) \\ & f & \mapsto & f'. \end{array}$$

Clearly D is linear.

Given $g \in \mathcal{C}((0,1),\mathbb{R})$, we may define $h(x) = \int_0^x g(t)dt$, $x \in (0,1)$. Then h is continuous on $(0,1)$, and in fact, by the Fundamental Theorem of Calculus, $h : (0,1) \to \mathbb{R}$ is differentiable on $(0,1)$ and $h'(x) = g(x)$, $x \in (0,1)$. Thus $h \in \mathcal{C}^1((0,1),\mathbb{R})$ and $Dh = g$, proving that D is surjective.

Note that if $f \in \ker D$, then $f' = 0$, so f must be a constant function: that is, there exists $\alpha \in \mathbb{R}$ such that $f(x) = \alpha$, $x \in (0,1)$. Thus $\ker D \simeq \mathbb{R}$. (Consider the map $\Theta : \ker D \to \mathbb{R}$ that sends $f(x) = \alpha$ for all $x \in (0,1)$ to $\alpha \in \mathbb{R}$.)

By the First Isomorphism Theorem for vector spaces,

$$\mathcal{C}^1((0,1),\mathbb{R})/\mathbb{R} = \mathcal{V}/\ker D \simeq \operatorname{ran} D = \mathcal{C}((0,1),\mathbb{R}).$$

S5.8. Example. Let $A = \begin{bmatrix} 1 & 3 \\ 1 & 1 \end{bmatrix} \in \mathbb{M}_2(\mathbb{Q})$. Let $\mathcal{D} = (d_1 := \begin{bmatrix} 1 \\ 1 \end{bmatrix}, d_2 := \begin{bmatrix} 1 \\ 2 \end{bmatrix})$ be an ordered basis for $\mathbb{Q}^2$. Let $\mathcal{B} = (e_1, e_2)$ be the standard ordered basis for $\mathbb{A}^2$. Then

$$[L_A]_{\mathcal{D}} = [[L_A d_1]_{\mathcal{D}} \ [L_A d_2]_{\mathcal{D}}] = \begin{bmatrix} [Ad_1]_{\mathcal{D}} & [Ad_2]_{\mathcal{D}} \end{bmatrix}.$$

Now $Ad_1 = \begin{bmatrix} 1 & 3 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$, while $Ad_2 = \begin{bmatrix} 1 & 3 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 7 \\ 3 \end{bmatrix}$.

Consider $Ad_1 = \begin{bmatrix} 4 \\ 2 \end{bmatrix} = q_{11}d_1 + q_{21}d_2 = q_{11}\begin{bmatrix} 1 \\ 1 \end{bmatrix} + q_{21}\begin{bmatrix} 1 \\ 2 \end{bmatrix}$. Solving for q_{11}, q_{21} yields

$$\begin{bmatrix} q_{11} \\ q_{21} \end{bmatrix} = \begin{bmatrix} 6 \\ -2 \end{bmatrix}.$$

Similarly, solving for $Ad_2 = \begin{bmatrix} 7 \\ 3 \end{bmatrix} = q_{12}d_1 + q_{22}d_2$ yields

$$\begin{bmatrix} q_{12} \\ q_{22} \end{bmatrix} = \begin{bmatrix} 11 \\ -4 \end{bmatrix}.$$

Thus $[L_A]_{\mathcal{D}} = \begin{bmatrix} 6 & 11 \\ -2 & -4 \end{bmatrix}$, and

$$\begin{bmatrix} 6 & 11 \\ -2 & -4 \end{bmatrix} = [L_A]_{\mathcal{D}} = [I_{\mathbb{Q}^2}]^{\mathcal{D}}_{\mathcal{B}} \ [L_A]_{\mathcal{B}} \ [I_{\mathbb{Q}^2}]^{\mathcal{B}}_{\mathcal{D}} = Q^{-1}AQ,$$

where $Q = [I_{\mathbb{Q}^2}]^{\mathcal{B}}_{\mathcal{D}}$.

Now $[I_{\mathbb{R}^2}]_{\mathcal{D}}^{\mathcal{B}} = \begin{bmatrix}[I_{\mathbb{Q}^2}d_1]_{\mathcal{B}} & [I_{\mathbb{Q}^2}d_2]_{\mathcal{B}}\end{bmatrix} = \begin{bmatrix}1 & 1\\ 1 & 2\end{bmatrix}$.. Thus we are claiming that

$$\begin{bmatrix}6 & 11\\ -2 & -4\end{bmatrix} = \begin{bmatrix}1 & 1\\ 1 & 2\end{bmatrix}^{-1}\begin{bmatrix}1 & 3\\ 1 & 1\end{bmatrix}\begin{bmatrix}1 & 1\\ 1 & 2\end{bmatrix}.$$

We can verify this without calculating $\begin{bmatrix}1 & 1\\ 1 & 2\end{bmatrix}^{-1}$. Indeed, $[L_A]_{\mathcal{D}} = Q^{-1}AQ$ is equivalent to the equation

$$Q[L_A]_{\mathcal{D}} = AQ,$$

and so we only need check that

$$\begin{bmatrix}1 & 1\\ 1 & 2\end{bmatrix}\begin{bmatrix}6 & 11\\ -2 & -4\end{bmatrix} = \begin{bmatrix}1 & 3\\ 1 & 1\end{bmatrix}\begin{bmatrix}1 & 1\\ 1 & 2\end{bmatrix},$$

which is easily verified, as both are equal to $\begin{bmatrix}4 & 7\\ 2 & 3\end{bmatrix}$.

S5.9. Example. Let $\mathcal{V}$ and $\mathcal{W}$ be finite-dimensional vector spaces over a field $\mathbb{F}$, and let $\mathcal{D}_1, \mathcal{D}_2$ be two ordered bases for $\mathcal{V}$ and $\mathcal{C}_1, \mathcal{C}_2$ be two ordered bases for $\mathcal{W}$. Then

$$[T]_{\mathcal{D}_2}^{\mathcal{C}^2} = [I_{\mathcal{W}}]_{\mathcal{C}_1}^{\mathcal{C}_2}\,[T]_{\mathcal{D}_1}^{\mathcal{D}_2}\,[I_{\mathcal{V}}]_{\mathcal{D}_2}^{\mathcal{D}_1}.$$

For example, let $\mathcal{D}_1 = (e_1, e_2, e_3)$ be the standard ordered basis for $\mathcal{V} = \mathbb{R}^3$ and $\mathcal{D}_2 = ((1,1,2),(1,2,1),(2,1,0))$. Let $\mathcal{C}_1 = (f_1, f_2)$ be the standard ordered basis for $\mathbb{R}^2$, and $\mathcal{C}_2 = ((1,1),(2,1))$.

Let $T : \mathbb{R}^3 \to \mathbb{R}^2$ be the linear map defined by $T(x,y,z) = (2x+3y, 4y+z)$.

Then

$$\begin{aligned}[T]_{\mathcal{D}_1}^{\mathcal{C}_1} &= \begin{bmatrix}[Te_1]_{\mathcal{C}_1} & [Te_2]_{\mathcal{C}_1} & [Te_3]_{\mathcal{C}_1}\end{bmatrix}\\ &= \begin{bmatrix}\begin{bmatrix}2\\0\end{bmatrix}_{\mathcal{C}_1} & \begin{bmatrix}3\\4\end{bmatrix}_{\mathcal{C}_1} & \begin{bmatrix}0\\2\end{bmatrix}_{\mathcal{C}_1}\end{bmatrix}\\ &= \begin{bmatrix}2 & 3 & 0\\ 0 & 4 & 1\end{bmatrix}.\end{aligned}$$

Next,

$$[I_{\mathbb{R}^3}]_{\mathcal{D}_2}^{\mathcal{D}_1} = \begin{bmatrix}[(1,1,2)]_{\mathcal{D}_1} & [(1,2,1)]_{\mathcal{D}_1} & [(2,1,0)]_{\mathcal{D}_1}\end{bmatrix} = \begin{bmatrix}1 & 1 & 2\\ 1 & 2 & 1\\ 2 & 1 & 0\end{bmatrix}.$$

Also,

$$[I_{\mathbb{R}^2}]_{\mathcal{C}_1}^{\mathcal{C}_2} = \begin{bmatrix}[f_1]_{\mathcal{C}_2} & [f_2]_{\mathcal{C}_2}\end{bmatrix} = \begin{bmatrix}-1 & 2\\ 1 & -1\end{bmatrix}.$$

Thus

$$[T]_{\mathcal{D}_2}^{\mathcal{C}_2} = \begin{bmatrix} -1 & 2 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 2 & 3 & 0 \\ 0 & 4 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 2 \\ 1 & 2 & 1 \\ 2 & 1 & 0 \end{bmatrix}$$
$$= \begin{bmatrix} 7 & 10 & 1 \\ -1 & -1 & 3 \end{bmatrix}.$$

Once again, we can check that this is right!!!

$$T(1,1,2) = (5,6) = 7(1,1) + (-1)(2,1)$$
$$T(1,2,1) = (8,9) = 10(1,1) + (-1)(2,1)$$
$$T(2,1,0) = (7,4) = 1(1,1) + 3(2,1).$$

Maybe not all of life is good, but this part is.

S5.10. Example. Let $2 \leq n \in \mathbb{N}$ and $T \in \mathbb{M}_n(\mathbb{F})$. We define the **similarity orbit** of T to be

$$\mathcal{S}(T) := \{S^{-1}TS : S \in \mathbb{M}_n(\mathbb{F}) \text{ invertible}\}.$$

Recall that for $X = [x_{ij}] \in \mathbb{M}_n(\mathbb{F})$, we defined the **trace** of X to be

$$\operatorname{tr}(X) := \sum_{k=1}^{n} x_{kk}.$$

Recall also from Assignment 6 that if $X, Y \in \mathbb{M}_n(\mathbb{F})$, then $\operatorname{tr}(XY) = \operatorname{tr}(YX)$. It follows that $\operatorname{tr}(\cdot)$ is constant on similarity orbits: that is, if $T \in \mathbb{M}_n(\mathbb{F})$ and $Y \in \mathcal{S}(T)$, then $\operatorname{tr}(Y) = \operatorname{tr}(T)$. Indeed, choose $S \in \mathbb{M}_n(\mathbb{F})$ invertible such that $Y = S^{-1}TS$. Then

$$\operatorname{tr}(Y) = \operatorname{tr}(S^{-1}TS) = \operatorname{tr}(S(S^{-1}T)) = \operatorname{tr}(T).$$

Observe also that if $Y = S^{-1}TS$, then $Y^2 = (S^{-1}TS)^2 = S^{-1}T^2S$, and more generally, by a routine induction argument, for $k \geq 1$, $Y^k = S^{-1}T^kS$. If $p(x) = p_0 + p_1x + \cdots + p_nx^n \in \mathbb{F}[x]$, then

$$p(S^{-1}TS) = S^{-1}p(T)S.$$

This simple fact will prove very useful in your next linear algebra course.

Appendix - dual spaces

A5.1. Given two vector spaces $\mathcal{V}$ and $\mathcal{W}$ over a field $\mathbb{F}$, we have seen that $\mathcal{L}(\mathcal{V},\mathcal{W})$ is a vector space. One particularly important instance of this phenomenon is when $\mathcal{W}=\mathbb{F}$.

A5.2. Definition. Let $\mathcal{V}$ be a vector space. The vector space $\mathcal{L}(\mathcal{V},\mathbb{F})$ is called the **(algebraic) dual space** of $\mathcal{V}$, and is denoted by $\mathcal{V}^{\#}$. Elements of $\mathcal{V}^{\#}$ are referred to as **linear functionals**.

A5.3. Example. Let $\mathcal{V}=\mathcal{C}([0,2\pi],\mathbb{R})$. Fix a function $g\in\mathcal{C}([0,2\pi],\mathbb{R})$ and define

$$\begin{array}{rccc}\mu_g: & \mathcal{C}([0,2\pi],\mathbb{R}) & \to & \mathbb{R}\\ & f & \mapsto & \int_0^{2\pi} f(x)g(x)dx.\end{array}$$

Then μ_g is a linear functional on $\mathcal{V}$.
Exercise. The map

$$\begin{array}{rccc}\Phi: & \mathcal{V} & \to & \mathcal{V}^{\#}\\ & g & \mapsto & \mu_g\end{array}$$

is itself linear. It can be shown to be injective as well, but this is harder.

The maps of the form μ_g, $g\in\mathcal{V}$ are not all of the linear functionals on $\mathcal{V}$. For example, for each $x_0\in[0,2\pi]$, we may define the **evaluation functional**

$$\begin{array}{rccc}\delta_{x_0}: & \mathcal{C}([0,2\pi],\mathbb{R}) & \to & \mathbb{R}\\ & f & \mapsto & f(x_0).\end{array}$$

Unlike the linear functionals μ_g above, the evaluation functionals have an extra property, namely they are **multiplicative**. That is, for all $f,h\in\mathcal{C}([0,2\pi],\mathbb{R})$,

$$\delta_{x_0}(fh)=(fh)(x_0)=f(x_0)\,h(x_0)=\delta_{x_0}(f)\,\delta_{x_0}(h).$$

Interestingly, the evaluation functionals are the only multiplicative linear functionals on $\mathcal{C}([0,2\pi],\mathbb{R})$. The standard proof of this is not easy.

A5.4. Example. Consider the map

$$\begin{array}{rccc}\mathrm{tr}: & \mathbb{M}_n(\mathbb{F}) & \to & \mathbb{F}\\ & T=[t_{ij}] & \mapsto & \sum_{j=1}^n t_{jj}.\end{array}$$

Then tr is a linear functional, referred to as the **trace** on $\mathbb{M}_n(\mathbb{F})$. When $\mathbb{F}=\mathbb{R}$ or $\mathbb{F}=\mathbb{C}$, it can be shown that tr is the unique linear functional φ satisfying $\varphi(RT)=\varphi(TR)$ for all $R,T\in\mathbb{M}_n(\mathbb{F})$.

A5.5. Example. For $1 \le j \le n$, the maps

$$\begin{array}{rccc} \gamma_j : & \mathbb{F}^n & \to & \mathbb{F} \\ & x = (x_1, x_2, \ldots, x_n) & \mapsto & x_j \end{array}$$

are linear functionals, called **coordinate functionals**.

If $\dim \mathcal{V} = m$, we have seen that for any order basis $\mathcal{D} = (v_1, v_2, \ldots, v_m)$ for $\mathcal{V}$, the map

$$\begin{array}{rccc} \Phi : & \mathcal{V} & \to & \mathbb{F}^m \\ & v & \mapsto & [v]_{\mathcal{D}} \end{array}$$

is an isomorphism. Thus, for $1 \le j \le m$, the maps

$$\begin{array}{rccc} \gamma_j \circ \Phi : & \mathcal{V} & \to & \mathbb{F} \\ & \sum_{i=1}^m \kappa_i v_i & \mapsto & \kappa_j \end{array}$$

is a linear functional.

A5.6. Remark. If $\dim \mathcal{V} = n < \infty$, then

$$\dim \mathcal{V}^\# = \dim \mathcal{L}(\mathcal{V}, \mathbb{F}) = \dim \mathcal{V} \cdot \dim \mathbb{F} = \dim \mathcal{V}.$$

Thus $\mathcal{V} \simeq \mathcal{V}^\#$ as vector spaces.

A5.7. Proposition. Suppose that $n \in \mathbb{N}$ and that $\mathcal{V}$ is an n-dimensional vector space over the field $\mathbb{F}$. Suppose that $\mathcal{D} = (v_1, v_2, \ldots, v_n)$ is an ordered basis for $\mathcal{V}$, and let $\mathcal{D}^\# := (\gamma_1, \gamma_2, \ldots, \gamma_n)$ denote the n-tuple of coordinate functionals defined in Example A5.5.

Let us show that $\mathcal{D}^\#$ is an ordered basis for $\mathcal{V}^\#$, called the **dual basis** to $\mathcal{D}$. First note that

$$\gamma_j(v_i) = \delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \ne j. \end{cases}$$

(The function δ_{ij} is known as the **Kronecker delta function**.) Thus, if $\kappa_1, \kappa_2, \ldots, \kappa_n \in \mathbb{F}$ and $\sum_{j=1}^n \kappa_j \gamma_j = 0$, then for each $1 \le i \le n$,

$$0 = (\sum_{j=1}^n \kappa_j \gamma_j)(v_i) = \kappa_i.$$

Thus $\mathcal{D}^\#$ is linearly independent. Since $|\mathcal{D}^\#| = n = \dim \mathcal{V}^\#$, $\mathcal{D}^\#$ is a basis for $\mathcal{V}^\#$.

A5.8. Example. If $\mathcal{V} = \mathbb{F}^n$ and $\mathcal{D} = \{e_1, e_2, \ldots, e_n\}$ is the standard ordered basis for $\mathcal{V}$, then $\mathcal{D}^\# = \{\gamma_1, \gamma_2, \ldots, \gamma_n\}$ where $\gamma_j(e_i) = \delta_{ij}$ is the dual basis to $\mathcal{D}$. In this special instance, we often write $e_j^\#$ instead of γ_j.

A5.9. Let $\mathcal{V}$ be a vector space. Consider the **double dual** $\mathcal{V}^{\#\#} = (\mathcal{V}^{\#})^{\#} = \mathcal{L}(\mathcal{V}^{\#}, \mathbb{F})$ of $\mathcal{V}$. We leave it as an exercise for you to show that the map

$$\begin{array}{rccc} \Gamma: & \mathcal{V} & \to & \mathcal{V}^{\#\#} \\ & v & \mapsto & \widehat{v}, \end{array}$$

where $\widehat{v}(\varphi) = \varphi(v)$ for all $\varphi \in \mathcal{V}^{\#}$ is linear. When $\dim \mathcal{V} < \infty$, Γ is an isomorphism, and $\widehat{\mathcal{D}} := \{\widehat{v_1}, \widehat{v_2}, \ldots, \widehat{v_n}\}$ is the dual basis to $\mathcal{D}^{\#}$!

A5.10. Example. Let $\mathcal{V} = \mathbb{R}_3[x]$ and $\mathcal{D} = \{1, x, x^2, x^3\}$ be an ordered basis for $\mathcal{V}$. The $\mathcal{D}^{\#} = \{\gamma_0, \gamma_1, \gamma_2, \gamma_3\}$, where

$$\gamma_j(x^i) = \delta_{ij}.$$

Thus for $0 \le j \le 3$,

$$\gamma_j(p_0 + p_1 x + p_2 x^2 + p_3 x^3) = p_j.$$

If $\varphi \in \mathcal{V}^{\#}$, then there exist $\alpha_0, \alpha_1, \alpha_2, \alpha_3 \in \mathbb{R}$ such that

$$\varphi(p_0 + p_1 x + p_2 x^2 + p_3 x^3) = \sum_{i=0}^{3} \alpha_i p_i.$$

Also,

$$\widehat{x^i}(\gamma_j) = \gamma_j(x^i) = \delta_{ij}, \quad 0 \le i, j \le 3.$$

A5.11. Definition. Suppose that $\mathcal{V}, \mathcal{W}$ are vector spaces over the field $\mathbb{F}$, and let $T \in \mathcal{L}(\mathcal{V}, \mathcal{W})$. We define the **adjoint** of T to be the map

$$\begin{array}{rccc} T^{\#}: & \mathcal{W}^{\#} & \to & \mathcal{V}^{\#} \\ & \varphi & \mapsto & T^{\#}\varphi, \end{array}$$

where $(T^{\#}\varphi)(v) = \varphi(Tv)$, $v \in \mathcal{V}$.

A5.12. Example. Let $\mathcal{V} = \mathbb{R}_3[x]$ and $\mathcal{W} = \mathbb{R}_2[x]$. Let $\mathcal{D} = (1, x, x^2, x^3)$ and $\mathcal{C} = (1, x, x^2)$ denote the standard ordered bases for $\mathcal{V}$ and $\mathcal{W}$ respectively. Then $\mathcal{D}^{\#} = (\gamma_0, \gamma_1, \gamma_2, \gamma_3)$ and $\mathcal{C}^{\#} = (\theta_0, \theta_1, \theta_2)$, where

$$\gamma_j(x^i) = \delta_{ij}, \quad 0 \le i, j \le 3,$$

and

$$\theta_j(x^i) = \delta_{ij}, \quad 0 \le i, j \le 2.$$

Let $D : \mathbb{R}_3[x] \to \mathbb{R}_2[x]$ denote the usual derivative map

$$D(p_0 + p_1 x + p_2 x^2 + p_3 x^3) = p_1 + 2p_2 x + 3p_3 x^2.$$

Then $D^{\#} : (\mathbb{R}_2[x])^{\#} \to (\mathbb{R}_3[x])^{\#}$ is the map which satisfies (for all $\kappa_0, \kappa_1, \kappa_2 \in \mathbb{R}$):

$$\begin{aligned}
&[D^{\#}(\kappa_0\theta_0 + \kappa_1\theta_1 + \kappa_2\theta_2)](p_0 + p_1x + p_2x^2 + p_3x^3) \\
&\qquad = [\kappa_0\theta_0 + \kappa_1\theta_1 + \kappa_2\theta_2](D(p_0 + p_1x + p_2x^2 + p_3x^3)) \\
&\qquad = [\kappa_0\theta_0 + \kappa_1\theta_1 + \kappa_2\theta_2](p_1 + 2p_2x + 3p_3x^2) \\
&\qquad = \kappa_0 \cdot p_1 + \kappa_1 \cdot (2p_2) + \kappa_2 \cdot (3p_3) \\
&\qquad = \kappa_0 \cdot p_1 + (2\kappa_1) \cdot p_2 + (3\kappa_2) \cdot p_3.
\end{aligned}$$

A5.13. Theorem. Suppose that $\mathcal{V}, \mathcal{W}$ are finite-dimensional vector spaces over the field $\mathbb{F}$, and let $T \in \mathcal{L}(\mathcal{V}, \mathcal{W})$. The adjoint $T^{\#}$ of T is linear. If $\mathcal{D}$ (resp. $\mathcal{C}$) is an ordered basis for $\mathcal{V}$ (resp. $\mathcal{W}$), and if $\mathcal{D}^{\#}$ (resp. $\mathcal{C}^{\#}$) is the dual basis to $\mathcal{D}$ (resp. $\mathcal{C}$), then

$$[T^{\#}]^{\mathcal{D}^{\#}}_{\mathcal{C}^{\#}} = ([T]^{\mathcal{C}}_{\mathcal{D}})^{\mathrm{t}}.$$

Indeed, we leave it to the reader to verify that in Example A5.12 above,

$$[D]^{\mathcal{C}}_{\mathcal{D}} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix},$$

while

$$[D^{\#}]^{\mathcal{D}^{\#}}_{\mathcal{C}^{\#}} = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}.$$

Exercises for Chapter 5

Exercise 5.1.

Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces, and let $\mathcal{B} = \{b_\alpha\}_{\alpha\in\Lambda}$ be a basis for $\mathcal{V}$. Suppose that $S, T \in \mathcal{L}(\mathcal{V}, \mathcal{W})$ and that $Sb_\alpha = Tb_\alpha$ for all $\alpha \in \Lambda$.

Prove that $S = T$.

Exercise 5.2.

Let $m, n \in \mathbb{N}$ and $\mathbb{F}$ be a field. Suppose that $A \in \mathbb{M}_{m\times n}(\mathbb{F})$, and define

$$\begin{array}{rccc} L_A: & \mathbb{F}^n & \to & \mathbb{F}^m \\ & x & \mapsto & Ax. \end{array}$$

Let $\mathcal{D} = \{e_1, e_2, \ldots, e_n\}$ denote the standard basis for $\mathbb{F}^n$, and $\mathcal{C} = \{f_1, f_2, \ldots, f_m\}$ denote the standard basis for $\mathbb{F}^m$.

Prove that $[L_A]_{\mathcal{D}}^{\mathcal{C}} = A$.

Exercise 5.3.

This question is a bit difficult, but it is interesting, important, and within reach. Let $\mathcal{V} = \mathcal{C}([0,1], \mathbb{R})$, and consider the linear map $V \in \mathcal{L}(\mathcal{C}([0,1], \mathbb{R}))$ defined by

$$[Vf](x) := \int_0^x f(t)dt, \qquad x \in [0,1].$$

Prove that if $\ker V = \{0\}$.

Exercise 5.4.

Observe that the matrix $S = \begin{bmatrix} 1 & 2 & -1 & 2 & 0 \\ 2 & -1 & 0 & 2 & 1 \\ 1 & 1 & 1 & 1 & -1 \\ 0 & 3 & 1 & 0 & -1 \\ 1 & 2 & 1 & 1 & 0 \end{bmatrix}$ is invertible. Suppose that $T \in \mathbb{M}_5(\mathbb{C})$ and that $S^{-1}TS = J_5 := \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$.

(a) Recall from Math 147/148 that $\exp(z) = \sum_{n=0}^{\infty} \frac{1}{n!} z^n$. Using this, find $\exp(T)$.

(b) Recall from Math 147/148 that $\sin(z) = \sum_{n=1}^{\infty} (-1)^{n+1} \frac{1}{(2n-1)!} z^n$. Using this, find $\sin(T)$.

Exercise 5.5.

Let $S \in \mathbb{M}_5(\mathbb{C})$ be the matrix from Exercise 5.4 above. Suppose that $X \in \mathbb{M}_5(\mathbb{C})$ and that $S^{-1}XS = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 0 & 5 \end{bmatrix}$.

(a) Recall from Math 147/148 that $\exp(z) = \sum_{n=0}^{\infty} \frac{1}{n!} z^n$. Using this, find $\exp(X)$.

(b) Recall from Math 147/148 that $\sin(z) = \sum_{n=1}^{\infty} (-1)^{n+1} \frac{1}{(2n-1)!} z^n$. Using this, find $\sin(X)$.

Exercise 5.6.

Let $n \in \mathbb{N}$ and suppose that $E \in \mathcal{L}(\mathbb{C}^n)$ satisfies $E = E^2$. Show that there exists a basis $\mathcal{B} = \{b_1, b_2, \ldots, b_n\}$ for $\mathbb{C}^n$ and $1 \leq k \leq n$ such that

$$[E]_{\mathcal{B}} = \begin{bmatrix} I_k & E_2 \\ 0 & 0_{n-k} \end{bmatrix}$$

for some matrix $E_2 \in \mathbb{M}_{k \times (n-k)}(\mathbb{C})$.

Exercise 5.7.

Let $n \in \mathbb{N}$ and suppose that $E = E^2, F = F^2$ are two idempotent operators in $\mathcal{L}(\mathbb{C}^n)$. Prove that E is similar to F if and only if $\operatorname{rank} E = \operatorname{rank} F$.

Hint. Let $P \in \mathcal{L}(\mathbb{C}^n)$ be the linear operator whose matrix relative to the standard basis $\mathcal{E} := \{e_1, e_2, \ldots, e_n\}$ for $\mathbb{C}^n$ is

$$[P]_{\mathcal{E}} = \begin{bmatrix} I_k & 0 \\ 0 & 0_{n-k} \end{bmatrix}.$$

Prove that any idempotent of rank equal to k is similar to P.

Exercise 5.8.

Let $\mathcal{B}$ and $\mathcal{D}$ be two bases for a vector space $\mathcal{V}$ over the a field $\mathbb{F}$. Suppose that $R, T \in \mathcal{L}(\mathbb{C}^n)$ and that

$$[R]_{\mathcal{B}} = [T]_{\mathcal{D}}.$$

Prove that R is similar to T.

Exercise 5.9.

Let $A = \begin{bmatrix} 1 & 2 \\ 0 & 4 \end{bmatrix}$ and $B = \begin{bmatrix} 3 & 5 \\ 0 & 7 \end{bmatrix} \in \mathbb{T}_2(\mathbb{C})$. Consider the maps

$$\begin{array}{rccc} L_A: & \mathbb{T}_2(\mathbb{C}) & \to & \mathbb{T}_2(\mathbb{C}) \\ & T & \mapsto & AT \end{array} \qquad \text{and} \qquad \begin{array}{rccc} R_B: & \mathbb{T}_2(\mathbb{C}) & \to & \mathbb{T}_2(\mathbb{C}) \\ & T & \mapsto & TB. \end{array}$$

(a) Show that $L_A R_B = R_B L_A \in \mathcal{L}(\mathbb{T}_2(\mathbb{C}))$.

(b) For which $\alpha \in \mathbb{C}$ does there exist $0 \neq T_\alpha \in \mathbb{T}_2(\mathbb{C})$ such that

$$(L_A - R_B)(T_\alpha) = \alpha T_\alpha?$$

Exercise 5.10.

Let $A \in \mathbb{M}_2(\mathbb{C})$ and consider the map

$$\begin{array}{rccc} \delta_A : & \mathbb{M}_2(\mathbb{C}) & \to & \mathbb{M}_2(\mathbb{C}) \\ & T & \mapsto & AT - TA. \end{array}$$

(a) Prove that $\delta_A(XY) = X\delta_A(Y) + \delta_A(X)Y$ for all $X, Y \in \mathbb{M}_2(\mathbb{C})$.
(b) Prove that δ_A is not invertible.

Culture. The map δ_A is referred to as an **inner derivation** on $\mathbb{M}_2(\mathbb{C})$.

CHAPTER 6

Matrix operations and systems of linear equations

My dad has suggested that I register for a donor card. He's a man after my own heart.

Masai Graham

1. Elementary matrix operations

1.1. In this Chapter we shall learn how to use matrices to solve linear equations. The familiar techniques used to solve such systems correspond to matrix operations that preserve the rank of the matrix. These techniques involve:

(a) interchanging two equations;
(b) multiplying an equation by a non-zero constant; and
(c) adding a multiple of one equation to another to help eliminate variables from an equation.

To a system $\mathbb{S}$ of the form:

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

we shall associate the matrix

$$A := \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & & \cdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{bmatrix}.$$

In light of this, the following definitions are motivated.

1.2. Definition. *Let $A \in \mathbb{M}_{m\times n}(\mathbb{F})$. An **elementary row operation** on A is one of the following:*

(a) *interchanging two rows of A;*
(b) *multiplying a row by a non-zero scalar from $\mathbb{F}$; and*
(c) *adding a scalar multiple of one row to another.*

By changing the word "row" for the word "column" in (a), (b), (c), we may analogously define ***elementary column operations****.*

If $A = [a_{ij}] \in \mathbb{M}_{m\times n}(\mathbb{F})$, we denote the i^{th} row of A by $\mathrm{R}_i(A)$, and the j^{th} column of A by $\mathrm{C}_j(A)$. When no confusion may arise, we abbreviate this notation to R_i and C_j, $1 \le i \le m, 1 \le j \le n$.

1.3. Example. Let $A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \end{bmatrix} \in \mathbb{M}_{2\times 4}(\mathbb{R})$.

(a) Interchanging C_1 and C_2 of A yields:

$$B = \begin{bmatrix} 2 & 1 & 3 & 4 \\ 6 & 5 & 7 & 8 \end{bmatrix}.$$

(b) Adding $(-5)\mathrm{R}_1$ of B to R_2 yields:

$$B_1 := \begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & -4 & -8 & -12 \end{bmatrix}.$$

(c) multiplying R_2 of B_1 by $-\frac{1}{4}$ yields:

$$B_2 := \begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 1 & 2 & 3 \end{bmatrix}.$$

(d) Adding $(-2)\mathrm{R}_2$ of B_2 to R_1 yields:

$$B_3 := \begin{bmatrix} 1 & 0 & -1 & -2 \\ 0 & 1 & 2 & 3 \end{bmatrix}.$$

1.4. Definition. *An* ***elementary matrix*** *$E \in \mathbb{M}_n(\mathbb{F})$ is one obtained from I_n by a single elementary operation on I_n.*

We leave it as an exercise for the reader to show that if $E \in \mathbb{M}_n(\mathbb{F})$ is an elementary matrix, then one can obtain E from I_n by an elementary row operation, and one can also obtain E from I_n by an elementary column operation.

1.5. Example. Let $E = \begin{bmatrix} 1 & 0 & 4 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \in \mathbb{M}_3(\mathbb{Q})$.

Then E is obtained from I_3 either

(a) by adding $4 \cdot \mathrm{R}_3$ of I_3 to R_1; or
(b) by adding $4 \cdot \mathrm{C}_1$ of I_3 to C_3.

Alternatively, if $F = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \in \mathbb{M}_3(\mathbb{Q})$, then F can be obtained from I_3 either by

(a) multiplying R_1 of I_3 by 2; or
(b) by multiplying C_1 of I_3 by 2.

1.6. Theorem. *Let $A \in \mathbb{M}_{m\times n}(\mathbb{F})$. Suppose that B is obtained from A by an elementary row operation. Let E be the elementary matrix obtained from I_m by the* ***same*** *elementary operation. Then $B = EA$.*

The converse also holds. That is to say: if $E \in \mathbb{M}_m(\mathbb{F})$ is obtained from I_m by an elementary row operation, the $B := EA$ is the matrix obtained from A by the same elementary row operation.

Proof. The proof is a routine calculation and is left as an exercise.

□

We remark that there is a corresponding result for column operations. An matrix C is obtained from A by an elementary operation if and only if there exists an elementary operation $E \in \mathbb{M}_n(\mathbb{F})$ obtained from I_n through the same operation such that $C = AE$. Note that in dealing with column operations, multiplication by E is now on the right, which means that we must have $E \in \mathbb{M}_n(\mathbb{F})$.

1.7. Example. Let $E = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$. Then E is obtained from I_3 by interchanging R_1 and R_3. If $A = [a_{ij}] \in \mathbb{M}_3(\mathbb{F})$, then

$$EA = \begin{bmatrix} a_{31} & a_{32} & a_{33} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

is obtained from A by the same elementary row operation.

Of course, E is also obtained from I_3 by interchanging C_1 and C_3. If $A = [a_{ij}] \in \mathbb{M}_3(\mathbb{F})$, then

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

is obtained from A by the same elementary column operation.

1.8. Theorem. *Elementary matrices are invertible, and their inverses are of the same type.*

Proof.

(a) If E is obtained from I_n by interchanging R_i and R_j, then $E = E^{-1}$.

(b) If E is obtained from I_n by scaling R_i by $0 \neq \kappa$, then E^{-1} is obtained from I_n by scaling R_i by κ^{-1}.

(c) If E is obtained from I_n by adding $\kappa \cdot \mathrm{R}_i$ to R_j, then E^{-1} is obtained from I_n by adding $-\kappa \cdot \mathrm{R}_i$ to R_j.

□

2. Rank and matrix inversion

2.1. Our goal in this section is to devise an algorithm to compute the inverse of a matrix, and hence of a linear map between finite-dimensional spaces. Elementary row operations are the building blocks of this algorithim.

2.2. Definition. *Let $A \in \mathbb{M}_{m\times n}(\mathbb{F})$. The **rank** of A is defined to be* $\operatorname{rank} L_A$*, where L_A is the left-regular representation of A; that is,*

$$\begin{array}{rccc} L_A: & \mathbb{F}^n & \to & \mathbb{F}^m \\ & x & \mapsto & Ax. \end{array}$$

Recall that $A \in \mathbb{M}_n(\mathbb{F})$ is invertible if and only if $L_A \in \mathcal{L}(\mathbb{F}^n)$ is invertible, and that if $\dim \mathcal{V} = n < \infty$ and $T \in \mathcal{L}(\mathcal{V})$, then T is invertible if and only if T is surjective, i.e. if and only if $\operatorname{rank} T = n$. Thus $A \in \mathbb{M}_n(\mathbb{F})$ is invertible if and only if $\operatorname{rank} A = n$.

2.3. Theorem. *Let $\mathcal{V}, \mathcal{W}$ be finite-dimensional vector spaces with ordered bases $\mathcal{D}$ and $\mathcal{C}$ respectively. Let $T \in \mathcal{L}(\mathcal{V}, \mathcal{W})$. Then* $\operatorname{rank} T = \operatorname{rank} [T]_{\mathcal{D}}^{\mathcal{C}}$.

Proof. Let $n := \dim \mathcal{V}$ and $m := \dim \mathcal{W}$. Recall that the map

$$\begin{array}{rccc} \Gamma_{\mathcal{W}}: & \mathcal{W} & \to & \mathbb{F}^m \\ & w & \mapsto & [w]_{\mathcal{C}} \end{array}$$

is an isomorphism of $\mathcal{W}$ onto $\mathbb{F}^m$. By Theorem 5.4.10,

$$\begin{aligned} \operatorname{rank} T &= \operatorname{rank} (\Gamma_{\mathcal{W}} \circ T) = \dim \operatorname{ran} (\Gamma_{\mathcal{W}} \circ T) \\ &= \dim \{[Tv]_{\mathcal{C}} : v \in \mathcal{V}\} = \dim \{[T]_{\mathcal{D}}^{\mathcal{C}} [v]_{\mathcal{D}} : v \in \mathcal{V}\} \\ &= \operatorname{rank} L_{[T]_{\mathcal{D}}^{\mathcal{C}}} = \operatorname{rank} [T]_{\mathcal{D}}^{\mathcal{C}}. \end{aligned}$$

□

The next result is an immediate consequence of the definition of the rank of a matrix combined with Theorem 5.4.10.

2.4. Theorem. *Let $B \in \mathbb{M}_{m\times n}(\mathbb{F})$ and $A \in \mathbb{M}_{p\times m}(\mathbb{F})$. Then*

$$\operatorname{rank} (AB) \le \min(\operatorname{rank} (A), \operatorname{rank} (B)).$$

When one of the terms is invertible, we can do better.

2.5. Theorem. *Let $A \in \mathbb{M}_{m\times n}(\mathbb{F})$. Suppose that $R \in \mathbb{M}_m(\mathbb{F})$ and $S \in \mathbb{M}_n(\mathbb{F})$ are invertible matrices. Then*

$$\operatorname{rank} A = \operatorname{rank} RA = \operatorname{rank} AS = \operatorname{rank} RAS.$$

Proof. Note that $L_R \in \mathcal{L}(\mathbb{F}^m)$ and $L_S \in \mathcal{L}(\mathbb{F}^n)$ are invertible. By Theorem 5.4.10,

$$\operatorname{rank} A = \operatorname{rank} L_A = \operatorname{rank}(L_R L_A) = \operatorname{rank}(L_{RA}) = \operatorname{rank}(RA),$$

and

$$\operatorname{rank} A = \operatorname{rank} L_A = \operatorname{rank}(L_A L_S) = \operatorname{rank}(L_{AS}) = \operatorname{rank}(AS).$$

Finally, combining these two equalities:

$$\operatorname{rank} A = \operatorname{rank}(AS) = \operatorname{rank}(R(AS)).$$

□

2.6. Corollary. *Elementary row and column operations preserve the rank of a matrix.*

2.7. Notation. We define the **column space** of the matrix $A \in \mathbb{M}_{m\times n}(\mathbb{F})$ to be

$$\textsc{Col}(A) := \operatorname{span}\{\textsc{c}_1, \textsc{c}_2, \ldots, \textsc{c}_n\} \subseteq \mathbb{F}^m,$$

where $\textsc{c}_j$ denotes the j^{th} column of A, $1 \le j \le n$. Analogously, we define the **row space** of A to be

$$\textsc{Row}(A) := \operatorname{span}\{\textsc{r}_1, \textsc{r}_2, \ldots, \textsc{r}_m\} \subseteq \mathbb{F}^n,$$

where $\textsc{r}_i$ denotes the i^{th} row of A, $1 \le i \le m$.

2.8. Theorem. *Let $A \in \mathbb{M}_{m\times n}(\mathbb{F})$. Then* $\operatorname{rank} A = \dim \textsc{Col}(A)$ *is the number of linearly independent columns of A.*

Proof. It suffices to note that if $\mathcal{D} = (e_1, e_2, \ldots, e_n)$ is the standard ordered basis for $\mathbb{F}^n$, then

$$\operatorname{ran} L_A = \operatorname{span}\{Ae_j : 1 \le j \le n\} = \operatorname{span}\{\textsc{c}_j : 1 \le j \le n\} = \textsc{Col}(A).$$

From this, the result immediately follows.

□

2.9. Example. Let $A = \begin{bmatrix} 0 & -2 & 4 \\ 1 & 1 & -1 \\ 3 & 4 & -5 \end{bmatrix}$. Now,

$$1\mathrm{C}_1 + (-2)\mathrm{C}_2 - 1\mathrm{C}_3 = \mathbf{0},$$

so rank $A \leq 2$. Since the first two columns of A are linearly independent, rank $A \geq 2$, whence rank $A = 2$.

Alternatively, we may apply elementary operations to determine the rank of A.

$$A = \begin{bmatrix} 0 & -2 & 4 \\ 1 & 1 & -1 \\ 3 & 4 & -5 \end{bmatrix} \xrightarrow{\mathrm{R}_1 \leftrightarrow \mathrm{R}_2} \begin{bmatrix} 1 & 1 & -1 \\ 0 & -2 & 4 \\ 3 & 4 & -5 \end{bmatrix} \xrightarrow{\mathrm{R}_3 + (-3)\mathrm{R}_1} \begin{bmatrix} 1 & 1 & -1 \\ 0 & -2 & 4 \\ 0 & 1 & -2 \end{bmatrix} \xrightarrow{\mathrm{R}_2 \leftrightarrow \mathrm{R}_3} \begin{bmatrix} 1 & 1 & -1 \\ 0 & 1 & -2 \\ 0 & -2 & 4 \end{bmatrix} \xrightarrow{\mathrm{R}_3 + (2)\mathrm{R}_2} \begin{bmatrix} 1 & 1 & -1 \\ 0 & 1 & -2 \\ 0 & 0 & 0 \end{bmatrix} \xrightarrow{\mathrm{R}_1 + (-1)\mathrm{R}_2} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & -2 \\ 0 & 0 & 0 \end{bmatrix}.$$

Clearly rank $A = 2$.

2.10. Theorem. *Let $A \in \mathbb{M}_{m\times n}(\mathbb{F})$, and suppose that* rank $A = r$. *Then $r \leq \min(m, n)$, and by a finite number of elementary row and column operations, A can be transformed into the matrix*

$$D = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}_{m\times n}.$$

The proof of this result has been (unfairly, perhaps?) relegated to the Appendix of this Chapter. As for the first statement, it is clear that

$$\text{rank } A = \text{rank } L_A = \dim \text{ran } L_A \leq \dim \mathbb{F}^m = m.$$

But as we saw above, rank A is the number of linearly independent columns of A, of which there can be no more than the total number of columns of A, namely n.

Let us nevertheless illustrate the process from this theorem with an example:

2.11. Example. Let $A = \begin{bmatrix} 1 & 2 & 1 & 0 \\ 2 & 5 & 5 & 1 \\ -2 & -3 & 0 & 3 \\ 3 & 4 & -2 & -3 \end{bmatrix}$. Then $D = I_4$.

$$A = \begin{bmatrix} 1 & 2 & 1 & 0 \\ 2 & 5 & 5 & 1 \\ -2 & -3 & 0 & 3 \\ 3 & 4 & -2 & -3 \end{bmatrix} \xrightarrow{R_2+(-2)R_1;\ R_3+(2)R_1;\ R_4+(-3)R_1} \begin{bmatrix} 1 & 2 & 1 & 0 \\ 0 & 1 & 3 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & -2 & -5 & -3 \end{bmatrix}$$

$$\xrightarrow{R_1+(-2)R_2;\ R_3+(-1)R_2;\ R_4+(2)R_2} \begin{bmatrix} 1 & 0 & -5 & -2 \\ 0 & 1 & 3 & 1 \\ 0 & 0 & -1 & 2 \\ 0 & 0 & 1 & -1 \end{bmatrix}$$

$$\xrightarrow{(-1)R_3} \begin{bmatrix} 1 & 0 & -5 & -2 \\ 0 & 1 & 3 & 1 \\ 0 & 0 & 1 & -2 \\ 0 & 0 & 1 & -1 \end{bmatrix}$$

$$\xrightarrow{R_1+(5)R_3;\ R_2+(-3)R_3;\ R_4+(-1)R_3} \begin{bmatrix} 1 & 0 & 0 & -12 \\ 0 & 1 & 0 & 7 \\ 0 & 0 & 1 & -2 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\xrightarrow{R_1+(12)R_4;\ R_2+(-7)R_4;\ R_3+(2)R_4} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

2.12. Corollary. *Let* $A = \mathbb{M}_{m\times n}(\mathbb{F})$ *and suppose that* rank $A = r$. *Then there exist invertible matrices* $B \in \mathbb{M}_m(\mathbb{F})$ *and* $C \in \mathbb{M}_n(\mathbb{F})$ *such that*

$$D = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} = BAC.$$

Furthermore, each of B *and* C *is a product of elementary matrices.*

Proof. Each row operation in Theorem 2.10 corresponds to multiplying by an elementary row matrix on the left, and each elementary column operation corresponds to multiplying by an elementary column matrix on the right. Since these are invertible, so are their products.

□

2.13. Example. Let $A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 5 & 7 & 9 \\ 3 & 7 & 10 & 13 \end{bmatrix}$.

$$A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 5 & 7 & 9 \\ 3 & 7 & 10 & 13 \end{bmatrix} \xrightarrow{R_2+(-2)R_1;\ R_3+(-3)R_1} \begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \end{bmatrix}$$

$$\xrightarrow{R_1+(-2)R_2;\ R_3+(-1)R_2} \begin{bmatrix} 1 & 0 & 1 & 2 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\xrightarrow{C_3+(-1)C_1;\ C_4+(-2)C_1} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\xrightarrow{C_3+(-1)C_2;\ C_4+(-1)C_2} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Now, to construct B, we consider:

$$I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow{R_2+(-2)R_1;\ R_3+(-3)R_1} \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ -3 & 0 & 1 \end{bmatrix}$$

$$\xrightarrow{R_1+(-2)R_2;\ R_3+(-1)R_2} \begin{bmatrix} 5 & -2 & 0 \\ -2 & 1 & 0 \\ -1 & -1 & 1 \end{bmatrix}.$$

Thus $B = \begin{bmatrix} 5 & -2 & 0 \\ -2 & 1 & 0 \\ -1 & -1 & 1 \end{bmatrix}$. Note that

$$BA = \begin{bmatrix} 5 & -2 & 0 \\ -2 & 1 & 0 \\ -1 & -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 5 & 7 & 9 \\ 3 & 7 & 10 & 13 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 & 2 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

To construct C, we consider:

$$I_4 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \xrightarrow{C_3+(-1)C_1;\ C_4+(-2)C_1} \begin{bmatrix} 1 & 0 & -1 & -2 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\xrightarrow{C_3+(-1)C_2;\ C_4+(-1)C_2} \begin{bmatrix} 1 & 0 & -1 & -2 \\ 0 & 1 & -1 & -1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Thus $C = \begin{bmatrix} 1 & 0 & -1 & -2 \\ 0 & 1 & -1 & -1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$. It follows that

$$BAC = (BA)C = \begin{bmatrix} 1 & 0 & 1 & 2 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & -1 & -2 \\ 0 & 1 & -1 & -1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

2.14. Corollary. *Let $A \in \mathbb{M}_{m\times n}(\mathbb{F})$. Then*

(a) $\operatorname{rank} A^{\mathrm{t}} = \operatorname{rank} A$.
(b) *The rank of A is equal to the dimension of* $\operatorname{Row}(A)$. *Thus,* $\operatorname{rank} A$ *represents the number of linearly independent rows of A.*

Proof. By Theorem 2.10, we can find invertible matrices B and C such that

$$BAC = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}.$$

Note that $r = \operatorname{rank}(BAC) = \operatorname{rank}(A)$ as B and C are invertible.

But it is also clear that $r = \operatorname{rank}(BAC)^{\mathrm{t}} = \operatorname{rank}(C^{\mathrm{t}}A^{\mathrm{t}}B^{\mathrm{t}})$ and that C^{t}, B^{t} are also invertible. Thus

$$r = \operatorname{rank}(A^{\mathrm{t}}) = \operatorname{rank}(A).$$

The second statement follows immediately from this.

□

2.15. Corollary. *Let $A \in \mathbb{M}_n(\mathbb{F})$. The following are equivalent.*

(a) *A is invertible.*
(b) *A is the product of elementary matrices.*

Proof.

(a) implies (b). By Corollary 2.12 above applied to the case where $m = n = r$, we can find $B, C \in \mathbb{M}_n(\mathbb{F})$ such that $I_n = BAC$. Thus

$$B^{-1}C^{-1} = B^{-1}I_nC^{-1} = B^{-1}(BAC)C^{-1} = A.$$

But if $B = E_jE_{j-1}E_{j-2}\cdots E_2E_1$, then $B^{-1} = E_1^{-1}E_2^{-1}\cdots E_{j-1}^{-1}E_j^{-1}$, and each E_i^{-1} is again an elementary matrix. Thus B^{-1}, and similarly C^{-1}, is a product of elementary matrices, meaning that A is also a product of elementary matrices.

(b) implies (a). Since each elementary matrix is invertible, any finite product of elementary matrices must also be invertible.

□

2.16. Example. Let

$$\begin{array}{rccc} D: & \mathbb{R}_3[x] & \to & \mathbb{R}_3[x] \\ & p & \mapsto & p'. \end{array}$$

With respect to the standard ordered basis $\mathcal{D} = (1, x, x^2, x^3)$ for $\mathbb{R}_3[x]$,

$$D = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

has rank equal to 3. Thus D is not invertible.

2.17. Definition. *Let $A \in \mathbb{M}_{m\times n}(\mathbb{F})$, $B \in \mathbb{M}_{m\times p}(\mathbb{F})$. The* ***augmented matrix*** $[\ A \mid B\ ]$ *is the matrix whose first n columns are those of A, and whose last p columns are those of B.*

2.18. Example. If $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ and $B = \begin{bmatrix} 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \end{bmatrix}$, then

$$[\ A \mid B\ ] = \left[\begin{array}{cc|cccc} 1 & 2 & 5 & 6 & 7 & 8 \\ 3 & 4 & 9 & 10 & 11 & 12 \end{array}\right].$$

2.19. Exercise. Let $\mathbb{F}$ be a field and $m, n, p, q \in \mathbb{N}$. Given $A \in \mathbb{M}_{m\times n}(\mathbb{F})$, $B \in \mathbb{M}_{m\times p}(\mathbb{F})$ and $T \in \mathbb{M}_{q\times m}(\mathbb{F})$,

$$T[A \mid B] = [TA \mid TB].$$

2.20. Inverting matrices. Suppose that $A \in \mathbb{M}_n(\mathbb{F})$ is invertible. Let $G := [\ A \mid I_n\ ] \in \mathbb{M}_{n\times 2n}(\mathbb{F})$.

Since A is invertible, we know by Corollary 2.15 that we may write $A = E_1E_2\cdots E_p$, where E_j is an elementary matrix, $1 \le j \le p$. Since elementary matrices are invertible and their inverses are also elementary matrices,

$$A^{-1} = E_r^{-1}E_{r-1}^{-1}\cdots E_2^{-1}E_1^{-1}.$$

Combining this with the above exercise, we see that

$$A^{-1}G = A^{-1}[\ A \mid I_n\ ] = [\ I_n \mid A^{-1}\ ].$$

In other words, if $A \in \mathbb{M}_n(\mathbb{F})$ and we apply a sequence of row operations to $[\ A \mid I_n\ ]$ to obtain $[\ I_n \mid B\ ]$, then $B = A^{-1}$.

2.21. Example. Consider $A = \begin{bmatrix} 1 & 2 & 1 \\ -1 & 1 & 2 \\ 1 & 0 & 1 \end{bmatrix} \in \mathbb{M}_3(\mathbb{R})$. We leave it as an exercise for the reader to show that

$$[\ A \mid I_n\ ] = \left[\begin{array}{ccc|ccc} 1 & 2 & 1 & 1 & 0 & 0 \\ -1 & 1 & 2 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \end{array}\right] \longrightarrow \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & \frac{1}{6} & -\frac{1}{3} & \frac{1}{2} \\ 0 & 1 & 0 & \frac{1}{2} & 0 & -\frac{1}{2} \\ 0 & 0 & 1 & -\frac{1}{6} & \frac{1}{3} & \frac{1}{2} \end{array}\right].$$

Thus

$$A^{-1} = \begin{bmatrix} \frac{1}{6} & -\frac{1}{3} & \frac{1}{2} \\ \frac{1}{2} & 0 & -\frac{1}{2} \\ -\frac{1}{6} & \frac{1}{3} & \frac{1}{2} \end{bmatrix}.$$

2.22. Example. Let $A = \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix}$. Then

$$[\ A \mid I_n\ ] = \left[\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 1 & 1 & 0 & 1 \end{array}\right] \longrightarrow \left[\begin{array}{cc|cc} 1 & 0 & -1 & 2 \\ 0 & 1 & 1 & -1 \end{array}\right],$$

and therefore

$$A^{-1} = \begin{bmatrix} -1 & 2 \\ 1 & -1 \end{bmatrix}.$$

3. Systems of linear equations

3.1. In this system we use what we have learnt about matrices to help us solve systems of linear equations. Given a system $\mathbb{S}$ of m equations in n unknowns,

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

the matrix $A \coloneqq [a_{ij}] \in \mathbb{M}_{m\times n}(\mathbb{F})$ is called the **coefficient matrix** of $\mathbb{S}$. . By writing

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{F}^n, \qquad b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix} \in \mathbb{F}^m,$$

the system $\mathbb{S}$ may be written as the single equation $Ax = b$, or as $L_A x = b$, where $L_A \in \mathcal{L}(\mathbb{F}^n, \mathbb{F}^m)$.

A **solution** to the system $\mathbb{S}$ is a vector $s = \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_n \end{bmatrix} \in \mathbb{F}^n$ such that $As = b$. The set

$$\mathrm{Sol}(\mathbb{S}) := \{s \in \mathbb{F}^n : As = b\}$$

is called the **solution set** of the system.

The system $\mathbb{S}$ is said to be **consistent** if $\mathrm{Sol}(\mathbb{S}) \neq \varnothing$, otherwise we say that $\mathbb{S}$ is **inconsistent**.

Furthermore, the system $\mathbb{S}$ is said to be **homogeneous** if $b = \mathbf{0}_m \in \mathbb{F}^m$. Then $\mathbf{0}_n \in \mathrm{Sol}(\mathbb{S})$, and so the system is automatically consistent. In fact, by the Dimension Theorem, we have the following.

3.2. Theorem. *If $\mathbb{S}$ is a homogeneous system $Ax = \mathbf{0}$ of m linear equations in n unknowns, then* $\mathrm{Sol}(\mathbb{S}) = \ker L_A$. *Thus $\mathbb{S}$ is a subspace of $\mathbb{F}^n$, and*

$$\dim \mathrm{Sol}(\mathbb{S}) = n - \mathrm{rank}\ L_A = n - \mathrm{rank}\ A.$$

3.3. Example. Consider the homogeneous system $\mathbb{S}$ of linear equations:

$$\begin{array}{rcrcrcl} x_1 & + & 2x_2 & - & x_3 & = & 0 \\ 2x_1 & + & x_2 & + & x_3 & = & 0 \end{array}$$

As above, we may write this as a single matrix equation:

$$\begin{bmatrix} 1 & 2 & -1 \\ 2 & 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

The solution set to the system $\mathbb{S}$ is $\mathrm{Sol}(\mathbb{S}) = \ker L_A$, where $A = \begin{bmatrix} 1 & 2 & -1 \\ 2 & 1 & 1 \end{bmatrix}$. Note that

$$\dim \mathrm{Sol}(\mathbb{S}) = n - \mathrm{rank}\ L_A = 3 - 2 = 1,$$

so that $\mathrm{Sol}(\mathbb{S})$ is a 1-dimensional space. Since $\begin{bmatrix} 1 \\ -1 \\ -1 \end{bmatrix} \in \ker L_A$, we see that

$$\mathrm{Sol}(\mathbb{S}) = \ker L_A = \mathrm{span}\left\{\begin{bmatrix} 1 \\ -1 \\ -1 \end{bmatrix}\right\} = \left\{\begin{bmatrix} a \\ -a \\ -a \end{bmatrix} : a \in \mathbb{F}\right\}.$$

3.4. Corollary. *If* $\mathrm{Sol}(\mathbb{S})$ *is a homogeneous system* $Ax = \mathbf{0}$ *of* m *equations in* n *unknowns, where* $1 \leq m < n$*, then the system has a non-zero solution.*
Proof. This follows immediately from Theorem 3.2, since if A is the matrix corresponding to the linear system, then rank $A \leq m < n$, so

$$\dim \mathrm{Sol}(\mathbb{S}) = n - \mathrm{rank}\ A > 0.$$

□

Suppose that $Ax = b$ represents a system of m linear equations in n unknowns. Then the equation

$$Ax = \mathbf{0}$$

is called the **homogeneous system corresponding** to $Ax = b$. We shall refer to the associated homogeneous system as $\mathbb{S}_{\mathrm{hom}}$.

3.5. Theorem. *Let* $\mathbb{S}$ *be a consistent system* $Ax = b$ *of* m *equations in* n *unknowns. Let* $s_0 \in \mathrm{Sol}(\mathbb{S})$*. Then*

$$\mathrm{Sol}(\mathbb{S}) = s_0 + \mathrm{Sol}(\mathbb{S}_{\mathrm{hom}}) = \{s_0 + k : k \in \mathrm{Sol}(\mathbb{S}_{\mathrm{hom}})\}.$$

In other words, $\mathrm{Sol}(\mathbb{S})$ *is a coset of* $\mathrm{Sol}(\mathbb{S}_{\mathrm{hom}})$ *in* $\mathbb{F}^n/\mathrm{Sol}(\mathbb{S}_{\mathrm{hom}})$*, and it has any particular solution* s_0 *of the system* $\mathbb{S}$ *as its representative.*
Proof. Let $s_0 \in \mathrm{Sol}(\mathbb{S})$.

Let $k \in \mathrm{Sol}(\mathbb{S}_{\mathrm{hom}})$. Then $A(s_0 + k) = As_0 + Ak = b + \mathbf{0} = b$, so $s_0 + k \in \mathrm{Sol}(\mathbb{S})$. That is,

$$s_0 + \mathrm{Sol}(\mathbb{S}_{\mathrm{hom}}) \subseteq \mathrm{Sol}(\mathbb{S}).$$

Next, let $s_1 \in \mathrm{Sol}(\mathbb{S})$. Then $As_1 = b = As_0$, so $k := s_1 - s_0 \in \mathrm{Sol}(\mathbb{S}_{\mathrm{hom}})$. That is,

$$\mathrm{Sol}(\mathbb{S}) \subseteq s_0 + \mathrm{Sol}(\mathbb{S}_{\mathrm{hom}}).$$

This completes the proof.

□

3.6. Example. Consider the system $\mathbb{S}$ given by:

$$\begin{array}{rcrcrcl} x_1 & + & 2x_2 & - & x_3 & = & 3 \\ 2x_1 & + & x_2 & + & x_3 & = & 3 \end{array}$$

The homogeneous system $\mathbb{S}_{\text{hom}}$ associated to $\mathbb{S}$ is the system from Example 3.3 above. Note that $\begin{bmatrix}1\\1\\0\end{bmatrix}$ is a particular solution to this system. By Theorem 3.5,

$$\begin{aligned}\text{Sol}(\mathbb{S}) &= \begin{bmatrix}1\\1\\0\end{bmatrix} + \text{Sol}(\mathbb{S}_{\text{hom}})\\ &= \begin{bmatrix}1\\1\\0\end{bmatrix} + \left\{\begin{bmatrix}a\\-a\\-a\end{bmatrix} : a \in \mathbb{F}\right\}\\ &= \left\{\begin{bmatrix}1+a\\1-a\\-a\end{bmatrix} : a \in \mathbb{F}\right\}.\end{aligned}$$

3.7. Example. Let us consider the system $\mathbb{S}$ given by:

$$\begin{array}{rcrcrcrcl} x_1 & + & 2x_2 & + & x_3 & + & x_4 & = & 1\\ & & x_2 & - & x_3 & + & x_4 & = & 1 \end{array}$$

Let $A = \begin{bmatrix}1 & 2 & 1 & 1\\ 0 & 1 & -1 & 1\end{bmatrix}$ denote the coefficient matrix of the system. We must find $\text{Sol}(\mathbb{S}_{\text{hom}}) = \ker L_A$. Let E be an elementary matrix in $\mathbb{M}_2(\mathbb{F})$, and observe that $Ax = \mathbf{0}$ if and only if $EAx = E\mathbf{0} = \mathbf{0}$. In other words, if B is the matrix obtained from A by performing elementary row operations on A, then $\ker L_B = \ker L_A$.Now

$$\begin{bmatrix}1 & 2 & 1 & 1\\ 0 & 1 & -1 & 1\end{bmatrix} \longrightarrow \begin{bmatrix}1 & 0 & 3 & -1\\ 0 & 1 & -1 & 1\end{bmatrix}.$$

Setting $x_4 = a, x_3 = b \in \mathbb{F}$, we solve for $x_2 = -a + b$ and $x_1 = a - 3b$. Thus

$$\text{Sol}(\mathbb{S}_{\text{hom}}) = \ker L_A = \{(a-3b, -a+b, b, a) : a, b \in \mathbb{F}\} = \text{span}\{(1,-1,0,1), (-3,1,1,0)\}.$$

In particular, $\dim \text{Sol}(\mathbb{S}_{\text{hom}}) = 2$. (We can check this by noting that $\text{rank } A = 2$, so $\text{nul } A = 4 - \text{rank } A = 4 - 2 = 2$.)

Observe that $(0,0,0,1) \in \mathrm{Sol}(\mathbb{S})$, and so by the above Theorem,

$$\begin{aligned}\mathrm{Sol}(\mathbb{S}) &= \begin{bmatrix}0\\0\\0\\1\end{bmatrix} + \mathrm{Sol}(\mathbb{S}_{\mathrm{hom}}) \\ &= \left\{ \begin{bmatrix}0\\0\\0\\1\end{bmatrix} + a\begin{bmatrix}1\\-1\\0\\1\end{bmatrix} + b\begin{bmatrix}-3\\1\\1\\0\end{bmatrix} : a,b \in \mathbb{F} \right\} \\ &= \left\{ \begin{bmatrix}a-3b\\-a+b\\b\\1+a\end{bmatrix} : a,b \in \mathbb{F} \right\}.\end{aligned}$$

When the coefficient matrix A of a system $\mathbb{S}$ is invertible, we obtain the following.

3.8. Theorem. *Let $\mathbb{S}$ be the system $Ax = b$ consisting of n equations in n unknowns. If the corresponding coefficient matrix $A \in \mathbb{M}_n(\mathbb{F})$ is invertible, then the system is consistent and admits the unique solution* $\mathrm{Sol}(\mathbb{S}) = \{A^{-1}b\}$.

Conversely, if the system $\mathbb{S}$ admits a unique solution, then A must be invertible.

Proof. Suppose that $A \in \mathbb{M}_n(\mathbb{F})$ is invertible. Clearly $A(A^{-1}b) = b$, and thus $A^{-1}b \in \mathrm{Sol}(\mathbb{S})$. Also, if $s \in \mathrm{Sol}(\mathbb{S})$, then $As = b$ implies that $s = A^{-1}(As) = A^{-1}b$, so $\mathrm{Sol}(\mathbb{S}) = \{A^{-1}b\}$.

Conversely, suppose that $Ax = b$ admits a unique solution. By Theorem 3.5,

$$\{A^{-1}b\} = \mathrm{Sol}(\mathbb{S}) = A^{-1}b + \mathrm{Sol}(\mathbb{S}_{\mathrm{hom}}).$$

Thus $\mathrm{Sol}(\mathbb{S}_{\mathrm{hom}}) = \{\mathbf{0}\}$.

By Theorem 3.2,

$$0 = \dim(\mathrm{Sol}(\mathbb{S}_{\mathrm{hom}})) = n - \mathrm{rank}\ A,$$

implying that $\mathrm{rank}\ A = n$, and thus that A is invertible.

□

3.9. Example. Consider the system $\mathbb{S}$ given by

$$\begin{aligned} x_1 &+ 3x_2 &&= 4 \\ 2x_1 &+ 5x_2 &&= 3 \end{aligned}$$

The corresponding coefficient matrix is $A = \begin{bmatrix}1 & 3\\ 2 & 5\end{bmatrix}$, which is invertible with inverse

$$A^{-1} = \begin{bmatrix}-5 & 3\\ 2 & -1\end{bmatrix}.$$

Thus

$$\mathrm{Sol}(\mathbb{S}) = \{A^{-1}b\} = \left\{\begin{bmatrix} -5 & 3 \\ 2 & -1 \end{bmatrix}\begin{bmatrix} 4 \\ 3 \end{bmatrix}\right\} = \left\{\begin{bmatrix} -11 \\ 5 \end{bmatrix}\right\}.$$

3.10. Definition. *Given a system $\mathbb{S}$ of m linear equations in n unknowns, say $Ax = b$, the matrix $[\ A \mid b\ ]$ is called the* ***augmented matrix*** *of the system.*

3.11. Remark. It is worth noting that any vector Ax is a linear combination of the columns of A, and if $y = \sum_{j=1}^{n} \kappa_j \mathrm{c}_j(A)$, then

$$y = A\begin{bmatrix} \kappa_1 \\ \kappa_2 \\ \vdots \\ \kappa_n \end{bmatrix} \in \operatorname{ran} L_A.$$

In other words, $\operatorname{ran} L_A = \textsc{Col}(A)$.

3.12. Theorem. *The system $\mathbb{S}$ defined by $Ax = b$ is consistent if and only if* $\operatorname{rank} A = \operatorname{rank} [\ A \mid b\ ]$.

Proof. As just noted, $\operatorname{ran} L_A = \textsc{Col}(A)$, and so $Ax = b$ is consistent if and only if $b \in \textsc{Col}(A)$; i.e. if and only if

$$\operatorname{rank} A = \dim \textsc{Col}(A) = \dim \textsc{Col}([\ A \mid b\ ]) = \operatorname{rank}([\ A \mid b\ ]).$$

□

3.13. Example. Consider the system $\mathbb{S}$ given by:

$$\begin{bmatrix} 1 & 2 & -1 \\ 2 & 1 & 2 \\ 1 & -4 & 7 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 4 \end{bmatrix}.$$

Then $\operatorname{rank} A = 2$, while $\operatorname{rank} [\ A \mid b\ ] = 3$. By Theorem 3.12, the system $\mathbb{S}$ is inconsistent, i.e. it does not have a solution.

3.14. Solving linear equations. We now turn our attention to the problem of solving a system of linear equations. We begin with a definition.

3.15. Definition. *Two systems $\mathbb{S}_1$ and $\mathbb{S}_2$ of linear equations are said to be* ***equivalent*** *if they admit the same solution set.*

3.16. Theorem. *Let $Ax = b$ represent a system of m linear equations in n unknowns over a field $\mathbb{F}$. Let $C \in \mathbb{M}_m(\mathbb{F})$ be an invertible matrix. Then $Ax = b$ is equivalent to the system $CAx = Cb$.*

Proof. We leave this as an exercise for the reader.

□

3.17. Corollary. *Applying a finite number of elementary row operations to a system of linear equations results in an equivalent system.*

3.18. Given a system $\mathbb{S}$ of m linear equations in n unknowns represented as $Ax = b$, we wish to apply elementary row operation to the augmented matrix $\left[\begin{array}{c|c} A & b \end{array}\right]$ of the system so as to yield:

- an upper triangular matrix in which the first non-zero entry of each row is 1, and this entry occurs to the right of the first non-zero entry of any preceding row.

3.19. Example. Consider the system $\mathbb{S}$ given by:

$$\begin{array}{rcrcrcrcr} x_1 & - & 4x_2 & - & x_3 & + & x_4 & = & 3 \\ 2x_1 & - & 8x_2 & + & x_3 & - & 4x_4 & = & 9 \\ -x_1 & + & 4x_2 & - & 2x_3 & + & 5x_4 & = & -6 \end{array}$$

The augmented matrix corresponding to $\mathbb{S}$ is:

$$\begin{bmatrix} 1 & -4 & -1 & 1 & 3 \\ 2 & -8 & 1 & -4 & 9 \\ -1 & 4 & -2 & 5 & -6 \end{bmatrix} \longrightarrow \cdots \longrightarrow \begin{bmatrix} 1 & -4 & -1 & 1 & 3 \\ 0 & 0 & 1 & -2 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

3.20. Example. Consider the system $\mathbb{S}$ given by:

$$\begin{array}{rcrcrcrcr} x_1 & + & 2x_2 & - & x_3 & + & x_4 & = & 2 \\ 2x_1 & + & x_2 & + & x_3 & - & x_4 & = & 3 \\ x_1 & + & 2x_2 & - & 3x_3 & + & 2x_4 & = & 2 \end{array}$$

The augmented matrix corresponding to $\mathbb{S}$ is:

$$\begin{bmatrix} 1 & 2 & -1 & 1 & 2 \\ 2 & 1 & 1 & -1 & 3 \\ 1 & 2 & -3 & 2 & 2 \end{bmatrix} \longrightarrow \cdots \longrightarrow \begin{bmatrix} 1 & 2 & -1 & 1 & 2 \\ 0 & 1 & -1 & 1 & \frac{1}{3} \\ 0 & 0 & 1 & -\frac{1}{2} & 0 \end{bmatrix}.$$

3.21. Definition. *A matrix $A \in \mathbb{M}_{m\times n}(\mathbb{F})$ is said to be in* ***reduced row echelon form*** *if*

(a) *any row containing a non-zero entry precedes a zero row;*
(b) *the first non-zero entry in any row is the only non-zero entry in its column; and*
(c) *the first non-zero entry of each row is* 1*, and this entry occurs to the right of the first non-zero entry of any preceding row.*

3.22. Example. Which of the following are in reduced row echelon form?

(a) $A_1 = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$.

(b) $A_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$.

(c) $A_3 = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$.

(d) $A_4 = \begin{bmatrix} 1 & 0 & 4 & 0 & 3 \\ 0 & 1 & 2 & 0 & 9 \\ 0 & 0 & 0 & 1 & 4 \end{bmatrix}$.

(e) $A_5 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$.

The procedure below for reducing a matrix to its reduced row echelon form is called **Gaussian elimination**.

3.23. Example. Solve the system $\mathbb{S}$ given by:

$$\begin{array}{rcrcrcrcrcr} 3x_1 & - & x_2 & + & x_3 & - & x_4 & + & 2x_5 & = & 5 \\ x_1 & - & x_2 & - & x_3 & - & 2x_4 & - & x_5 & = & 2 \\ 5x_1 & - & 2x_2 & + & x_3 & - & 3x_4 & + & 3x_5 & = & 10 \\ 2x_1 & - & x_2 & & & - & 2x_4 & + & x_5 & = & 5 \end{array}$$

The augmented matrix corresponding to $\mathbb{S}_1$ is:

$$\begin{bmatrix} 3 & -1 & 1 & -1 & 2 & 5 \\ 1 & -1 & -1 & -2 & -1 & 2 \\ 5 & -2 & 1 & -3 & 3 & 10 \\ 2 & -1 & 0 & -2 & 1 & 5 \end{bmatrix} \longrightarrow \cdots \longrightarrow \begin{bmatrix} 1 & 0 & 1 & 0 & 2 & 3 \\ 0 & 1 & 2 & 0 & 5 & 7 \\ 0 & 0 & 0 & 1 & -1 & -3 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

This corresponds to the equivalent system $\mathbb{S}_2$ given by:

$$\begin{array}{rcrcrcrcr} x_1 & & + & x_3 & & + & 2x_5 & = & 3 \\ & x_2 & + & 2x_3 & & + & 5x_5 & = & 7 \\ & & & & x_4 & - & x_5 & = & -3 \end{array}$$

To solve this system, we assign parameters to the **non-leading variables** (i.e. the variables which *do not* correspond to a leading one in the non-augmented matrix of the system), and we use these to solve for the leading variables:

In our case, the non-leading variables are x_5 and x_3; we set $x_5 = t \in \mathbb{F}$, $x_3 = s \in \mathbb{F}$ and obtain:

$$\begin{aligned} x_5 &= t \\ x_4 &= -3 + t \\ x_3 &= s \\ x_2 &= 7 - 5t + 2s \\ x_1 &= 3 - 2t + s. \end{aligned}$$

The solution set is therefore

$$\mathrm{Sol}(\mathbb{S}_1) = \mathrm{Sol}(\mathbb{S}_2) = \left\{ \begin{bmatrix} 3 - 2t + s \\ 7 - 5t + 2s \\ s \\ -3 + t \\ t \end{bmatrix} : s, t \in \mathbb{F} \right\} = \begin{bmatrix} 3 \\ 7 \\ 0 \\ -3 \\ 0 \end{bmatrix} + \mathrm{span} \left\{ \begin{bmatrix} 1 \\ 2 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -2 \\ -5 \\ 0 \\ 1 \\ 1 \end{bmatrix} \right\}.$$

3.24. Theorem. *Let $\mathbb{S}$ be the system $Ax = b$ of m non-zero equations in n unknowns over a field $\mathbb{F}$. Suppose that* $\operatorname{rank} A = \operatorname{rank}[\ A \mid b\ ]$ *and that* $[\ A \mid b\ ]$ *is in reduced row echelon form. Then*

(a) $\operatorname{rank} A = m$.

(b) *if the general solution to the system obtained through Gaussian elimination is*

$$s = s_0 + \kappa_1 u_1 + \kappa_2 u_2 + \cdots + \kappa_{n-m} u_{n-m},$$

where $s_0, u_j \in \mathbb{F}^n$ are fixed and $\kappa_j \in \mathbb{F}$, $1 \le j \le n - m$ are arbitrary, then

$$\mathcal{D} := \{u_1, u_2, \ldots, u_{n-m}\}$$

is a basis for

$$\ker A := \ker L_A = \mathrm{Sol}(\mathbb{S}_{\mathrm{hom}}).$$

Moreover, s_0 may be replaced by any other solution t_0 to the original system $Ax = b$.

Proof.

(a) Since all of the rows of $[A \mid b]$ are non-zero (by hypothesis), and since $[A \mid b]$ is assumed to be in reduced row-echelon form, each row must have a leading 1 that is the only non-zero entry in its column. This means that there must be m linearly independent columns, and so $\operatorname{rank}[A \mid b] = m$. Since $\operatorname{rank} A = \operatorname{rank}[A \mid b]$ by hypothesis, it follows that $\operatorname{rank} A = m$.

(b) Choose arbitrary $\kappa_i \in \mathbb{F}$, $1 \le i \le n - m$, and let $s_1 = s_0 + \sum_{j=1}^{n-m} \kappa_j u_j$. Note that $s_0 = s_0 + \sum_{j=1}^{n-m} 0\, u_j$ is given as another solution to the system $Ax = b$. Thus $As_1 = b = As_0$, and so

$$s_1 - s_0 = \sum_{j=1}^{n-m} \kappa_j u_j \in \ker L_A = \mathrm{Sol}(\mathbb{S}_{\mathrm{hom}}).$$

Since the κ_j's were arbitrary,

$$\operatorname{span}\{u_1, u_2, \ldots, u_{n-m}\} \subseteq \operatorname{Sol}(\mathbb{S}_{\text{hom}}) = \ker L_A.$$

Conversely, if $y \in \operatorname{Sol}(\mathbb{S}_{\text{hom}})$, then $A(s_0 + y) = As_0 + Ay = As_0 = b$, so $s_0 + y \in \{s_0 + \sum_{j=1}^{n-m} \alpha_j u_j : \alpha_j \in \mathbb{F}, 1 \leq j \leq n-m\}$. In particular, $y_0 \in \operatorname{span}\{u_1, u_2, \ldots, u_{n-m}\}$, and thus

$$\ker L_A = \operatorname{Sol}(\mathbb{S}_{\text{hom}}) \subseteq \operatorname{span}\{u_1, u_2, \ldots, u_{n-m}\}.$$

Thus

$$\ker L_A = \operatorname{Sol}(\mathbb{S}_{\text{hom}}) = \operatorname{span}\{u_1, u_2, \ldots, u_{n-m}\}.$$

But $\operatorname{rank} L_A = \operatorname{rank} A = m$, whence $\operatorname{nul} L_A = n - m$. Since $\{u_j : 1 \leq j \leq n-m\}$ spans the $n-m$ dimensional space $\ker L_A$, it must be a basis for that space.

If $t_0 \in \operatorname{Sol}(\mathbb{S})$ is arbitrary, then $t_0 + \ker L_A = s_0 + \ker L_A$ since $At_0 = b = As_0$ implies that $t_0 - s_0 \in \ker L_A$, and thus t_0 and s_0 are representatives of the same coset in $\mathbb{F}^n / \ker L_A$.

□

3.25. Theorem. *Suppose that $A \in \mathbb{M}_{m\times n}(\mathbb{F})$, and that B is obtained from A through a finite number of elementary row operations. Let $\mathrm{C}_1(A), \mathrm{C}_2(A), \ldots, \mathrm{C}_n(A)$ denote the columns of A, and $\mathrm{C}_1(B), \mathrm{C}_2(B), \ldots, \mathrm{C}_n(B)$ denote the columns of B. Given $\alpha_1, \alpha_2, \ldots, \alpha_n \in \mathbb{F}$, we have that*

$$\sum_{j=1}^{n} \alpha_j \mathrm{C}_j(A) = 0 \qquad \textit{if and only if} \qquad \sum_{j=1}^{n} \alpha_j \mathrm{C}_j(B) = 0.$$

Proof. Let M be the matrix which implements the sequence of elementary row operations so that $B = MA$. Then M is invertible, $\mathrm{C}_j(B) = M\mathrm{C}_j(A)$, and so $\mathrm{C}_j(A) = M^{-1}\mathrm{C}_j(B)$. From this the result easily follows.

□

3.26. Example. Suppose that the reduced row echelon form of a matrix A is given by

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

Determine A if the first, second and fourth columns are given by

$$\begin{bmatrix} 1 \\ -1 \\ 3 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 1 \\ -2 \\ 0 \end{bmatrix}.$$

Note that $B = MA$, where M is invertible (being the product of the elementary row operations that induce the elementary operations on A). Thus Theorem 3.25 applies.

Since $\mathrm{C}_3(B) = 2\mathrm{C}_1(B) - 5\mathrm{C}_2(B)$, it follows that

$$\mathrm{C}_3(A) = 2\mathrm{C}_1(A) - 5\mathrm{C}_2(A) = 2\begin{bmatrix}1\\-1\\3\end{bmatrix} - 5\begin{bmatrix}0\\-1\\1\end{bmatrix} = \begin{bmatrix}2\\3\\1\end{bmatrix}.$$

Since $\mathrm{C}_5(B) = -2\mathrm{C}_1(B) - 3\mathrm{C}_2(B) + 6\mathrm{C}_4(B)$, it follows that

$$\mathrm{C}_5(A) = -2\mathrm{C}_1(A) - 3\mathrm{C}_2(A) + 6\mathrm{C}_4(A) = (-2)\begin{bmatrix}1\\-1\\3\end{bmatrix} + (-3)\begin{bmatrix}0\\-1\\1\end{bmatrix} + 6\begin{bmatrix}1\\-2\\0\end{bmatrix} = \begin{bmatrix}4\\-7\\-9\end{bmatrix}.$$

Thus

$$A = \begin{bmatrix}1 & 0 & 2 & 1 & 4\\ -1 & -1 & 3 & -2 & -7\\ 3 & 1 & 1 & 0 & -9\end{bmatrix}.$$

3.27. Example. Let

$$A = \begin{bmatrix}1 & 2 & 1 & -1 & 2\\ 1 & 1 & 1 & 0 & 3\\ 3 & 2 & 3 & -2 & 1\end{bmatrix} \longrightarrow \cdots \longrightarrow B = \begin{bmatrix}1 & 0 & 1 & 0 & 1\\ 0 & 1 & 0 & 0 & 2\\ 0 & 0 & 0 & 1 & 3\end{bmatrix}.$$

Since $\mathrm{C}_1(B), \mathrm{C}_2(B), \mathrm{C}_4(B)$ are linearly independent, so are $\mathrm{C}_1(A), \mathrm{C}_2(A)$, and $\mathrm{C}_4(A)$. Also,

$$\mathrm{C}_3(B) = \mathrm{C}_1(B) \qquad \text{and} \qquad \mathrm{C}_5(B) = \mathrm{C}_1(B) + 2\mathrm{C}_2(B) + 3\mathrm{C}_4(B),$$

whence $\mathrm{C}_3(A) = \mathrm{C}_1(A)$ and

$$\mathrm{C}_5(A) = \mathrm{C}_1(A) + 2\mathrm{C}_2(A) + 3\mathrm{C}_4(A).$$

Keep in mind yet again that on an exam or assignment, you can check your answer!

3.28. Example. Let $\mathcal{L} := \left\{L_1 := \begin{bmatrix}1 & 2\\ 3 & 4\end{bmatrix}, L_2 := \begin{bmatrix}0 & 1\\ 2 & 3\end{bmatrix}\right\}$, and observe that $\mathcal{L}$ is linearly independent in $\mathbb{M}_2(\mathbb{C})$. Let us extend $\mathcal{L}$ to a basis for $\mathbb{M}_2(\mathbb{C})$.

The standard basis for $\mathbb{M}_2(\mathbb{C})$ is $\mathcal{B} := \{E_{11}, E_{12}, E_{21}, E_{22}\}$.

Recall that the map $\Phi : \mathbb{M}_2(\mathbb{C}) \to \mathbb{C}^4$ defined by $\Phi(A) = [A]_{\mathcal{B}}$ is an isomorphism, and that isomorphisms preserve linearly independent sets. We therefore consider the matrix $A \in \mathbb{M}_{4\times 6}(\mathbb{C})$ whose columns are $[L_1]_{\mathcal{B}}, [L_2]_{\mathcal{B}}, [E_{11}]_{\mathcal{B}}, \ldots, [E_{22}]_{\mathcal{B}}$.

That is, consider

$$A := \begin{bmatrix}1 & 0 & 1 & 0 & 0 & 0\\ 2 & 1 & 0 & 1 & 0 & 0\\ 3 & 2 & 0 & 0 & 1 & 0\\ 4 & 3 & 0 & 0 & 0 & 1\end{bmatrix} \overset{\text{ROW}}{\longrightarrow} \cdots \overset{\text{REDUCE}}{\longrightarrow} B := \begin{bmatrix}1 & 0 & 0 & 0 & 3 & -2\\ 0 & 1 & 0 & 0 & -4 & 3\\ 0 & 0 & 1 & 0 & -3 & 2\\ 0 & 0 & 0 & 1 & -2 & 1\end{bmatrix}.$$

Since $\{\mathrm{C}_1(B), \mathrm{C}_2(B), \mathrm{C}_3(B), \mathrm{C}_4(B)\}$ is linearly independent, so is the set $\{\mathrm{C}_1(A), \mathrm{C}_2(A), \mathrm{C}_3(A), \mathrm{C}_4(A)\}$. Since Φ is an isomorphism, this means that $\mathcal{L}$ extends to the linearly independent set

$$\mathcal{M} := \left\{ \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 2 & 3 \end{bmatrix}, E_{11}, E_{12} \right\}.$$

But $\dim \mathbb{M}_2(\mathbb{C}) = 4$, so $\mathcal{M}$ must be a basis, as it is linearly independent in $\mathbb{M}_2(\mathbb{C})$ and it has four elements.

3.29. Example. The set

$$\mathcal{S} := \{s_1 = (1,2,3,4), s_2 = (1,1,0,0), s_3 = (6,7,3,4), s_4 = (2,1,0,0)\}$$

spans a subspace $\mathcal{W}$ of $\mathbb{R}^4$. Let us find a basis $\mathcal{B}$ for $\mathcal{W}$ such that $\mathcal{B} \subseteq \mathcal{S}$.

Let $\mathcal{D} = (e_1, e_2, e_3, e_4)$ be the standard ordered basis for $\mathbb{R}^4$, and again note that the map $\Phi : \mathbb{R}^4 \to \mathbb{R}^4$ defined by $\Phi(x) = [x]_{\mathcal{D}}$ is an isomorphism. (In fact, it looks like the identity map because we chose the standard basis! The only difference is that we are writing the coordinate vectors as columns instead of rows. Having said the "*only*" difference, it is an important one!)

Writing $[s_1]_{\mathcal{D}}, [s_2]_{\mathcal{D}}, [s_3]_{\mathcal{D}}, [s_4]_{\mathcal{D}}$ as columns of a matrix A, we find that

$$A := \begin{bmatrix} 1 & 1 & 6 & 2 \\ 2 & 1 & 7 & 1 \\ 3 & 0 & 3 & 0 \\ 4 & 0 & 4 & 0 \end{bmatrix} \longrightarrow \cdots \longrightarrow B := \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 5 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Since $\mathrm{C}_1(B), \mathrm{C}_2(B), \mathrm{C}_4(B)$ are linearly independent, so are $\mathrm{C}_1(A), \mathrm{C}_2(A), \mathrm{C}_4(A)$. Since $\mathrm{C}_3(B) = \mathrm{C}_1(B) + 5\mathrm{C}_2(B)$, we have that

$$\mathrm{C}_3(A) = \mathrm{C}_1(A) + 5\mathrm{C}_2(A) \in \operatorname{span}\{\mathrm{C}_1(A), \mathrm{C}_2(A), \mathrm{C}_4(A)\}.$$

Thus $s_3 \in \operatorname{span}\{s_1, s_2, s_4\}$ and we may take

$$\mathcal{B} = \{s_1, s_2, s_4\} = \{(1,2,3,4), (1,1,0,0), (2,1,0,0)\}.$$

Supplementary Examples

S6.1. Example. Let $\mathbb{F}$ be a field and $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathbb{M}_2(\mathbb{F})$. Suppose that $\Delta := ad - bc \neq 0$, and let

$$Y := \Delta^{-1} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

An easy calculation then shows that

$$AY = I_2 = YA,$$

and so $Y = A^{-1}$.

If $\Delta = 0$, then $\operatorname{rank} A < 2$ (check!), and so A is not invertible.

In the next Chapter, we shall refer to Δ as the **determinant** of A, and we shall generalise the notion of determinants to all matrices in $\mathbb{M}_n(\mathbb{F})$, showing that a matrix $B \in \mathbb{M}_n(\mathbb{F})$ is invertible if and only if $\operatorname{Det}(B) \neq 0$. Unfortunately, the computation of B^{-1} when B is invertible is not going to be as nice as the the one above for 2×2 matrices over $\mathbb{F}$.

S6.2. Example.

S6.3. Example.

S6.4. Example.

S6.5. Example.

S6.6. Example.

S6.7. Example.

S6.8. Example.

S6.9. Example.

S6.10. Example.

Appendix

Let's prove Theorem 6.2.10. We begin with a Lemma.

A6.1 Lemma. *Let $m, n \in \mathbb{N}$ and $\mathbb{F}$ be a field. Suppose that $0 \neq A \in \mathbb{M}_{m\times n}(\mathbb{F})$. Then there exist invertible matrices $B \in \mathbb{M}_m(\mathbb{F})$ and $C \in \mathbb{M}_n(\mathbb{C})$ such that if $T = [t_{ij}] := BAC$, then*

(i) $t_{11} = 1$;
(ii) $t_{1j} = 0$ *if* $2 \leq j \leq n$; *and*
(iii) $t_{i1} = 0$ *if* $2 \leq i \leq m$.

Proof. Recall that applying elementary row operations to A is equivalent to multiplying A on the left by an (invertible) elementary matrix in $\mathbb{M}_m(\mathbb{F})$, while applying elementary column operations to A is equivalent to multiplying A on the right by an (invertible) elementary matrix in $\mathbb{M}_n(\mathbb{F})$. Thus if we can obtain the matrix T from A through a finite sequence of elementary operations, this completes the proof.

The hypothesis that $0 \neq A$ implies that there exist $1 \leq i_0 \leq m$, $1 \leq j_0 \leq n$ such that $a_{i_0 j_0} \neq 0$. Consider the sequence of elementary operations:

$$A \overset{\mathrm{R}_{i_0}\leftrightarrow\mathrm{R}_1}{\longrightarrow} A_1 \overset{\mathrm{C}_{j_0}\leftrightarrow\mathrm{C}_1}{\longrightarrow} A_2 \overset{a_{i_0 j_0}^{-1}\mathrm{R}_1}{\longrightarrow} A_3.$$

This has the effect of first moving the (i_0, j_0) entry of A to the $(1, j_0)$ spot, then to the $(1, 1)$ spot, and then changing it from a_{i_0,j_0} to 1. In other words, if we set $X = A_3$ and write $X = [x_{ij}]$, then $x_{11} = 1$. (Note that if $i_0 = 1$, then the operation of switching rows i_0 and 1 is just doing nothing, which is represented by multiplication on the left by the identity matrix. A similar statement holds if $j_0 = 1$, or if $a_{i_0 j_0} = 1$ to begin with.)

Next, consider the chain of elementary operations

$$X \overset{\mathrm{C}_2+(-x_{12})\mathrm{C}_1}{\longrightarrow} X_2 \overset{\mathrm{C}_3+(-x_{13})\mathrm{C}_1}{\longrightarrow} X_3 \overset{\mathrm{C}_4+(-x_{14})\mathrm{C}_1}{\longrightarrow} \cdots\cdots \overset{\mathrm{C}_n+(-x_{1n})\mathrm{C}_1}{\longrightarrow} X_n.$$

Letting $Y = X_n$ and writing $Y = [y_{ij}]$, we find that $y_{11} = 1$ and $y_{1j} = 0$ for all $2 \leq j \leq n$.

We then apply the following chain of elementary operations.

$$Y \overset{\mathrm{R}_2+(-y_{21})\mathrm{R}_1}{\longrightarrow} Y_2 \overset{\mathrm{R}_3+(-y_{31})\mathrm{R}_1}{\longrightarrow} Y_3 \overset{\mathrm{R}_4+(-y_{41})\mathrm{R}_1}{\longrightarrow} \cdots\cdots \overset{\mathrm{R}_m+(-y_{m1})\mathrm{R}_1}{\longrightarrow} Y_m.$$

Let $T := Y_m$, and observe that T was obtained from A by at most $3 + (n-1) + (m-1) < \infty$ elementary operations, so that $T = BAC$ for some invertible matrices B and C (namely the products of the elementary matrices applied above). Write $T = [t_{ij}]$.

Now, since none of these operations in the last chain affected the first row of Y, we find that $t_{11} = 1$ and $t_{1j} = 0$, $2 \leq j \leq n$. The purpose of the last chain of elementary operations was to eliminate any non-zero entries in the first column of the last $(m-1)$ rows, and thus we have that $t_{i1} = 0$ for all $2 \leq i \leq m$.

This is exactly what we wanted.

□

A6.2 Theorem. *Let $A \in \mathbb{M}_{m\times n}(\mathbb{F})$, and suppose that* $\operatorname{rank} A = r$. *Then $r \le \min(m,n)$, and by a finite number of elementary row and column operations, A can be transformed into the matrix*

$$D = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}_{m\times n}.$$

Proof. That $r \le \min(m,n)$ was proven in the main body of the notes.

As for the second statement, we shall argue by induction on m, the number of rows of A.

Case One. $m = 1$. If $A = 0$, there is nothing to do, and the result holds. If $A \neq 0$, then by Lemma A6.1, we can find $B \in \mathbb{M}_1(\mathbb{F})$ and $C \in \mathbb{M}_n(\mathbb{F})$ such that

$$T = BAC = [1\ 0\ 0 \cdots 0\ 0].$$

Since $\operatorname{rank} A \ge 1$ (as $A \neq 0$) and since $\operatorname{rank} A \le 1 = m$, we have completed the proof in this case.

Case Two. $m > 1$. Let $M \ge 2$ and suppose that the result holds for all matrices in $\mathbb{M}_{m\times n}(\mathbb{F})$ where $m < M$. We must prove that it holds if $A \in \mathbb{M}_{M\times n}(\mathbb{F})$.

Again, if $A = 0$, there is nothing to do and the result holds. Thus we may suppose that $A \neq 0$.

By Lemma A6.1, we can find invertible matrices (i.e. products of elementary matrices) $B_0 \in \mathbb{M}_M(\mathbb{F})$ and $C_0 \in \mathbb{M}_n(\mathbb{F})$ such that if $T = B_0AC_0 = [t_{ij}]$, then

(i) $t_{11} = 1$;
(ii) $t_{1j} = 0$ if $2 \le j \le n$; and
(iii) $t_{i1} = 0$ if $2 \le i \le m$.

Thus

$$T = \begin{bmatrix} 1 & 0 \\ 0 & T_4 \end{bmatrix},$$

where $T_4 \in \mathbb{M}_{(M-1)\times(n-1)}(\mathbb{F})$. By our induction hypothesis, we can apply finitely many elementary row operations *involving only rows* 2 *through* M and finitely many elementary column operations *involving only columns* 2 *through* n so that the resulting matrix D is of the form

$$D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & I_s & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

where $s = \operatorname{rank}(T_4)$. Since D was obtained from T using finitely many elementary operations, and since T was obtained from A through finitely many elementary operations, we conclude that D was obtained from A through finitely many elementary operations, and thus

$$D = BAC$$

for some invertible matrices $B \in \mathbb{M}_M(\mathbb{F})$ and $C \in \mathbb{M}_n(\mathbb{F})$.

That $r := s+1 = \operatorname{rank} A$ is now clear, since this is $\dim \textsc{Col}(D) = \operatorname{rank} D = \operatorname{rank} A$, as elementary operations do not affect the rank of a matrix.

□

Exercises for Chapter 6

Exercise 6.1.

Recall from Exercise 2.19 that if $\mathbb{F}$ is a field and $m, n, p, q \in \mathbb{N}$, then, given $A \in \mathbb{M}_{m\times n}(\mathbb{F})$, $B \in \mathbb{M}_{m\times p}(\mathbb{F})$ and $T \in \mathbb{M}_{q\times m}(\mathbb{F})$,

$$T[A\,|B] = [TA \mid TB].$$

Of course, if there is any justice in the world (and let's face it, do we really believe that there is?), then there should be a corresponding result for multiplication on the right. Well, despite the appeal to justice in the world, there actually is a dual result.

Given $C \in \mathbb{M}_{n\times q}(\mathbb{F})$ and $D \in \mathbb{M}_{p\times q}(\mathbb{F})$, we define

$$\begin{bmatrix} C \\ D \end{bmatrix} \in \mathbb{M}_{(n+p)\times q}(\mathbb{F})$$

as the matrix whose first n rows are the rows of C and whose last p rows are the rows of D. Your new mission, should you choose to accept it, is to verify that

$$[A \mid B]\begin{bmatrix} C \\ D \end{bmatrix} = AC + BD \in \mathbb{M}_{m\times q}(\mathbb{F}).$$

Exercise 6.2.

There's no stopping us now. Let $\mathbb{F}$ be a field, and $n_1, n_2, \ldots, n_r \in \mathbb{N}$. Set $n = \sum_{j=1}^{r} n_j$. For each $1 \le i, j \le r$, suppose that $A_{ij}, B_{ij} \in \mathbb{M}_{n_i\times n_j}(\mathbb{F})$.

Let $A = [A_{ij}], B = [B_{ij}]$, which we may think of as elements of $\mathbb{M}_n(\mathbb{F})$ (if we disregard** the "["'s and "]"'s around each of the matrices A_{ij} and B_{ij}). Show that

$$AB = [C_{ij}],$$

where $C_{ij} = \sum_{k=1}^{r} A_{ik}B_{kj}$.

For example**, if $n_1 = 2, n_2 = 3$ and $n = n_1 + n_2 = 5$, we conflate

$$\begin{bmatrix} \begin{bmatrix} 1 & 2 \\ 6 & 7 \end{bmatrix} & \begin{bmatrix} 3 & 4 & 5 \\ 8 & 9 & 10 \end{bmatrix} \\ \begin{bmatrix} 11 & 12 \\ 16 & 17 \\ 21 & 22 \end{bmatrix} & \begin{bmatrix} 13 & 14 & 15 \\ 18 & 19 & 20 \\ 23 & 24 & 25 \end{bmatrix} \end{bmatrix}$$

and

$$\begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 6 & 7 & 8 & 9 & 10 \\ 11 & 12 & 13 & 14 & 15 \\ 16 & 17 & 18 & 19 & 20 \\ 21 & 22 & 23 & 24 & 25 \end{bmatrix}.$$

Exercise 6.3.

Prove that in Example 3.22 of this Chapter, the answers are: YES for (d), and NO for (a), (b), (c) and (e). That is, in each case you must justify the answer.

Exercise 6.4.

Exercise 6.5.

Exercise 6.6.

Exercise 6.7.

Exercise 6.8.

Exercise 6.9.

Exercise 6.10.

CHAPTER 7

Determinants

Whenever someone says "I don't believe in coincidences", I say: "Oh my God, me neither!"

Alasdair Beckett-King

1. The basics

1.1. In this section we shall associate to every square matrix $A \in \mathbb{M}_n(\mathbb{F})$ an element $\alpha \in \mathbb{F}$ called the **determinant** of A, denoted by $\det(A)$. The importance of determinants *can* be overstated. For example, if one were to say that determinants are more important than the Fundamental Theorem of Calculus, I think that most working mathematicians, and even the lazy ones, would disagree. But they are useful.

For example, we shall see below that the matrix A is invertible if and only if $\det(A) \neq 0$, and so a very nice application of determinants is to determine the invertibility of a matrix A, and hence the consistency and uniqueness of solutions to systems of n linear equations in n unknowns. Determinants also have geometrical interpretations which time (or rather a lack of time) does not allow us to explore.

Theorem 7.1.15 below essentially says that the determinant function yields a *group homomorphism* from the (generally non-abelian) group $\mathrm{GL}_n(\mathbb{F}) := \{T \in \mathbb{M}_n(\mathbb{F}) : T \text{ is invertible}\}$ to the abelian group $\mathbb{F} \setminus \{0\}$, which can't be all bad. Indeed, this is nothing to sneeze at unless you are allergic to groups and group homomorphisms.

In this section we develop the basic properties of the determinant function. Alas, it is not as nice as the trace functional on $\mathbb{M}_n(\mathbb{F})$, but we shouldn't judge it just based upon that.

1.2. Definition. *Let $\mathbb{F}$ be a field. If $A := [a] \in \mathbb{M}_1(\mathbb{F})$, then we define the **determinant** of A to be*

$$\mathrm{Det}(A) := a \in \mathbb{F}.$$

*If $B := \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \in \mathbb{M}_2(\mathbb{F})$, we define the **determinant** of B to be*

$$\mathrm{Det}(B) := b_{11}b_{22} - b_{21}b_{12}.$$

1.3. Remarks. Suppose that $k \in \{1,2\}$, and let $A, B \in \mathbb{M}_k(\mathbb{F})$. We invite the reader to verify the following.

(a) A is invertible if and only if $\det A \neq 0$;
(b) $\det(AB) = \det(A) \cdot \det(B)$; and
(b) if $A \in \mathbb{M}_k(\mathbb{F})$ is invertible, then $\det(A^{-1}) = (\det(A))^{-1}$.

1.4. Definition. *Let $n \in \mathbb{N}$, $\mathbb{F}$ be a field and $A = [a_{ij}] \in \mathbb{M}_n(\mathbb{F})$. Given $1 \leq i, j \in n$, we define the (i,j)-submatrix $\widehat{A}_{ij}$ of A to be the $(n-1) \times (n-1)$ matrix obtained from A by deleting the i^{th} row and the j^{th} column.*

Having defined the determinants of 1×1 and 2×2 matrices over $\mathbb{F}$, given $3 \leq n \in \mathbb{N}$, we now recursively defined the ***determinant*** *of a matrix $A = [a_{ij}] \in \mathbb{M}_n(\mathbb{F})$ as follows:*

$$\det(A) := \sum_{i=1}^{n} (-1)^{i+1}\ a_{i1}\ \det(\widehat{A}_{i1}).$$

For $1 \leq i, j \leq n$, the element

$$\det(\widehat{A}_{ij}) \in \mathbb{F}$$

is referred to as the (i,j) ***minor*** *of A, and*

$$(-1)^{i+j}\ \det(\widehat{A}_{ij}) \in \mathbb{F}$$

is called the (i,j) ***cofactor*** *of A.*

The above definition of $\det(A)$ is then called the ***expansion by cofactors on the first column*** *of A.*

1.5. Example.

(a) If $B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \in \mathbb{M}_2(\mathbb{F})$, then $\widehat{B}_{11} = [b_{22}] \in \mathbb{M}_1(\mathbb{F})$, while $B_{21} = [b_{12}] \in \mathbb{M}_1(\mathbb{F})$. If we expand the determinant of B by cofactors along the first column, we obtain:

$$\begin{aligned} \det(B) &= (-1)^{1+1} b_{11} \det(\widehat{B}_{11}) + (-1)^{2+1} b_{21} \det(\widehat{B}_{21}) \\ &= b_{11}(b_{22}) + (-1) b_{21}(b_{12}). \end{aligned}$$

This agrees with our original definition of $\det(B)$ - which is good news indeed.

(b) Let $T = [t_{ij}] \in \mathbb{M}_3(\mathbb{F})$. Then

$$\begin{aligned} \det(T) = (-1)^{1+1} t_{11}\ \det \begin{bmatrix} t_{22} & t_{23} \\ t_{32} & t_{33} \end{bmatrix} + (-1)^{2+1} t_{21}\ \det \begin{bmatrix} t_{12} & t_{13} \\ t_{32} & t_{33} \end{bmatrix} \\ (-1)^{3+1} t_{31}\ \det \begin{bmatrix} t_{12} & t_{13} \\ t_{22} & t_{23} \end{bmatrix} \\ = t_{11}(t_{22}t_{33} - t_{23}t_{32}) - t_{21}(t_{12}t_{33} - t_{13}t_{32}) + t_{31}(t_{12}t_{23} - t_{13}t_{22}) \\ = t_{11}t_{22}t_{33} + t_{13}t_{32}t_{21} + t_{31}t_{12}t_{23} - t_{11}t_{23}t_{32} - t_{21}t_{12}t_{33} - t_{31}t_{13}t_{22}. \end{aligned}$$

1.6. Exercise. Suppose that $T = [t_{ij}] \in \mathbb{T}_n(\mathbb{F})$. Then a routine induction argument shows that

$$\det(T) = t_{11}\ t_{22}\ t_{33} \cdots t_{nn}.$$

In particular, $\det(I_n) = 1$ for all $n \geq 1$.

1.7. Theorem. *Let $n \in \mathbb{N}$ and $A = [a_{ij}], B = [b_{ij}]$ and $C = [c_{ij}] \in \mathbb{M}_n(\mathbb{F})$. Suppose that there exists $1 \leq p \leq n$ such that*

- $\mathrm{R}_i(A) = \mathrm{R}_i(B) = \mathrm{R}_i(C)$ *if* $1 \leq i \neq p \leq n$; *and*
- $\mathrm{R}_p(A) = \mathrm{R}_p(B) + \mathrm{R}_p(C)$.

Then

$$\det(A) = \det(B) + \det(C).$$

Proof. We shall argue by induction on n.

The base case is $n = 1$. Here $A = [b_{11} + c_{11}]$, $B = [b_{11}]$ and $C = [c_{11}]$. Thus

$$\det A = b_{11} + c_{11} = \det B + \det C.$$

Now suppose that $N > 1$ and that the result holds if $n < N$. Let $A = [a_{ij}]$, $B = [b_{ij}]$, $C = [c_{ij}]$, and suppose that the conditions of the Lemma hold, namely

- $\mathrm{R}_i(A) = \mathrm{R}_i(B) = \mathrm{R}_i(C)$ if $1 \leq i \neq p \leq n$; and
- $\mathrm{R}_p(A) = \mathrm{R}_p(B) + \mathrm{R}_p(C)$.

By definition,

$$\det A = \sum_{i=1}^{N} (-1)^{i+1} a_{i1}\ \det \widehat{A}_{i1}.$$

Now $a_{p1} = b_{p1} + c_{p1}$, and $\widehat{A}_{p1} = \widehat{B}_{p1} = \widehat{C}_{p1}$, so

$$\begin{aligned} a_{p1} \det \widehat{A}_{p1} &= (b_{p1} + c_{p1}) \det \widehat{A}_{p1} \\ &= b_{p1} \det \widehat{A}_{p1} + c_{p1} \det \widehat{A}_{p1} \\ &= b_{p1} \det \widehat{B}_{p1} + c_{p1} \det \widehat{C}_{p1} \end{aligned}$$

If $1 \leq i \neq p \leq N$, then $\widehat{A}_{i1}$, $\widehat{B}_{i1}$ and $\widehat{C}_{i1}$ are identical, except that for some q_i, the q_i^{th} row of A satisfies $a_{q_i,j} = b_{q_i,j} + c_{q_i,j}$, $2 \leq j \leq n$. (One can explicitly calculate what q_i is, but we just need to know that the induction hypothesis works no matter which row that is.)

Thus by the induction hypothesis, $\det \widehat{A}_{i1} = \det \widehat{B}_{i1} + \det \widehat{C}_{i1}$, and since $a_{i1} = b_{i1} = c_{i1}$, we get

$$\begin{aligned} a_{i1} \det \widehat{A}_{i1} &= a_{i1} \det \widehat{B}_{i1} + a_{i1} \det \widehat{C}_{i1} \\ &= b_{i1} \det \widehat{B}_{i1} + c_{i1} \det \widehat{C}_{i1}. \end{aligned}$$

Thus

$$\begin{aligned}\det A &= \sum_{i=1}^{N}(-1)^{i+1}(b_{i1}\det \widehat{B}_{i1} + c_{i1}\det \widehat{C}_{i1}) \\ &= \sum_{i=1}^{N}(-1)^{i+1}b_{i1}\det \widehat{B}_{i1} + \sum_{i=1}^{N}(-1)^{i+1}c_{i1}\det \widehat{C}_{i1} \\ &= \det B + \det C.\end{aligned}$$

This concludes the induction step and the proof.

□

1.8. Theorem. *Let $n \in \mathbb{N}$, $\mathbb{F}$ be a field, $\kappa \in \mathbb{F}$ and $A \in \mathbb{M}_n(\mathbb{F})$.*

(a) *If $1 \le p \le n$ and $B \in \mathbb{M}_n(\mathbb{F})$ is obtained from A by multiplying $\mathrm{R}_p(A)$ by κ, then*

$$\det(B) = \kappa \det(A).$$

(b) *If $\mathrm{R}_p(A) = \mathbf{0} \in \mathbb{F}^n$ for some $1 \le p \le n$, then*

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

Proof.

(a) We argue by induction on n.

If $n = 1$, the result obviously holds. Let $N \in \mathbb{N}$ and suppose that the result holds when $n < N$. Let $B = [b_{ij}] \in \mathbb{M}_N(\mathbb{F})$, and suppose that $\mathrm{R}_i(B) = \mathrm{R}_i(A)$ if $1 \le i \ne p \le N$; $\mathrm{R}_p(B) = \kappa\,\mathrm{R}_p(A)$.

Then $\widehat{A}_{p1} = \widehat{B}_{p1}$, so

$$b_{p1}\det(\widehat{B}_{p1}) = \kappa a_{p1}\det(\widehat{A}_{p1}).$$

For $1 \le i \ne p \le N$, one row of $\widehat{B}_{i1}$ is κ times the same row of $\widehat{A}_{i1}$, and so by our induction hypothesis,

$$\det(\widehat{B}_{i1}) = \kappa\det(\widehat{A}_{i1}).$$

Since $b_{i1} = a_{i1}$ when $i \ne p$,

$$b_{i1}\det(\widehat{B}_{i1}) = \kappa a_{i1}\det(\widehat{A}_{i1}).$$

Thus

$$\begin{aligned}\det(B) &= \sum_{i=1}^{N}(-1)^{i+1}b_{i1}\det(\widehat{B}_{i1}) \\ &= \sum_{i=1}^{N}(-1)^{i+1}\kappa a_{i1}\det(\widehat{B}_{i1}) \\ &= \kappa\sum_{i=1}^{N}(-1)^{i+1}a_{i1}\det(\widehat{B}_{i1}) \\ &= \kappa\det(A).\end{aligned}$$

(b) Let B be the matrix obtained from A by multiplying $\mathrm{R}_p(A)$ by $\kappa = 0$. By part (a),

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

On the other hand, $B = A$, from which the result follows.

□

By combining Theorem 1.7 and Theorem 1.8, we see that $\det(\cdot)$ is linear in any row. See the Appendix at the end of the Chapter for more details.

1.9. Corollary. *Let $n \in \mathbb{N}$ and $A = [a_{ij}] \in \mathbb{M}_n(\mathbb{F})$. If $B = -A$, then*

$$\det(B) = (-1)^n \det(A).$$

Proof. Let $A_0 = A$ and for $1 \le k \le n$, let A_k be the matrix obtained from A_{k-1} by multiplying the k^{th} row of A_{k-1} by -1. Note that $B = A_n$ and that $\det(A_k) = (-1)\det(A_{k-1})$ for each $1 \le k \le n$, so that

$$\begin{aligned}\det(B) = \det(A_n) &= (-1)\det(A_{n-1}) \\ &= (-1)^2\det(A_{n-2}) \\ &= (-1)^3\det(A_{n-3}) \\ &= \cdots \\ &= (-1)^n\det(A_0) = (-1)^n\det(A).\end{aligned}$$

□

1.10. Theorem. *Let $n \in \mathbb{N}$, $\mathbb{F}$ be a field, $\kappa \in \mathbb{F}$ and $A \in \mathbb{M}_n(\mathbb{F})$. If $1 \le p \ne q \le n$ and B is obtained from A by interchanging $\mathrm{R}_p(A)$ and $\mathrm{R}_q(A)$, then*

$$\det(B) = -\det(A).$$

Proof.

We argue by induction on n.

STEP 1. The case where $n = 2$ is a simple calculation: if $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$, then

$$B = \begin{bmatrix} a_{21} & a_{22} \\ a_{11} & a_{12} \end{bmatrix},$$

and so

$$\det(B) = (a_{21}a_{12} - a_{22}a_{11}) = -(a_{11}a_{22} - a_{21}a_{22}) = \det(A).$$

STEP 2. Let $3 \le N < \infty$ and suppose that the result holds whenever $n < N$, $1 \le p \ne q \le n$. Let $A \in \mathbb{M}_N(\mathbb{F})$, and let B be obtained from A by interchanging the p^{th} and the q^{th} rows of A. We first observe that it suffices to prove that $\det(T) = -\det(A)$ in the case where T is obtained from A by interchanging the first row and any other row of A. Indeed, suppose that the conclusion holds in this case.

Let $T_0 = A$ and consider the sequence of operations:

$$A = T_0 \overset{\mathrm{R}_1 \leftrightarrow \mathrm{R}_p}{\longrightarrow} T_1 \overset{\mathrm{R}_1 \leftrightarrow \mathrm{R}_q}{\longrightarrow} T_2 \overset{\mathrm{R}_1 \leftrightarrow \mathrm{R}_p}{\longrightarrow} T_3 = B.$$

Since each T_k is obtained from T_{k-1} by interchanging the first row and another row of T_{k-1}, $1 \leq k \leq 3$, we find that

$$\det(B) = \det(T_3) = -\det(T_2) = \det(T_1) = -\det(T_0) = -\det(A).$$

Thus we shall prove the result when $n = N$, $p = 1$ and $2 \leq q \leq N$, completing the induction step and the proof.

Writing $A = [a_{ij}], B = [b_{ij}] \in \mathbb{M}_N(\mathbb{F})$, we have that

$$\det(A) = \sum_{i=1}^{n} (-1)^{i+1} a_{i1} \det(\widehat{A}_{i1}),$$

and

$$\det(B) = \sum_{i=1}^{n} (-1)^{i+1} b_{i1} \det(\widehat{B}_{i1}).$$

If $1 \leq i \leq N$ and $i \notin \{1, q\}$, then $a_{i1} = b_{i1}$ and $\widehat{B}_{i1}$ is obtained from $\widehat{A}_{i1}$ by interchanging two of the rows of the latter matrix. Since $\widehat{B}_{i1}, \widehat{A}_{i1} \in \mathbb{M}_{N-1}(\mathbb{F})$, our induction hypothesis implies that

$$\det(\widehat{B}_{i1}) = -\det(\widehat{A}_{i1}).$$

Thus

$$(-1)^{i+1} b_{i1} \det(\widehat{B}_{i1}) = -((-1)^{i+1} a_{i1} \det(\widehat{A}_{i1})).$$

There remains to consider the cases where $i = 1$ and $i = q$. We claim that

$$(-1)^{q+1} a_{q1} \det(\widehat{A}_{q1}) = -(-1)^{1+1} b_{11} \det(\widehat{B}_{11})$$

and

$$(-1)^{1+1} a_{11} \det(\widehat{A}_{11}) = -(-1)^{q+1} b_{q1} \det(\widehat{B}_{q1}).$$

If this holds, then from the above formulae for $\det(A)$ and $\det(B)$, we find that

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

and we are done. We prove the first of these equalities - the proof of the second being similar.

Observe first that $a_{q1} = b_{11}$. Next, note that $\widehat{B}_{11}$ is obtained from $\widehat{A}_{q1}$ by reordering the rows $\mathrm{R}_1, \mathrm{R}_2, \ldots, \mathrm{R}_{q-1}$ of $\widehat{A}_{q1}$ into the order $\mathrm{R}_2, \mathrm{R}_3, \ldots, \mathrm{R}_{q-1}, \mathrm{R}_1$. But this can be accomplished by a finite sequence of operations, each consisting of switching only two rows of the previous $(N-1) \times (N-1)$ matrix. Indeed, let $X_0 \coloneqq \widehat{A}_{q1}$ and consider the sequence of operations:

$$\widehat{A}_{q1} = X_0 \overset{\mathrm{R}_1 \leftrightarrow \mathrm{R}_2}{\longrightarrow} X_1 \overset{\mathrm{R}_2 \leftrightarrow \mathrm{R}_3}{\longrightarrow} X_2 \overset{\mathrm{R}_3 \leftrightarrow \mathrm{R}_3}{\longrightarrow} X_3 \cdots \longrightarrow \cdots \overset{\mathrm{R}_{q-2} \leftrightarrow \mathrm{R}_{q-1}}{\longrightarrow} X_{q-2} = \widehat{B}_{11}.$$

By the induction hypothesis (keeping in mind that $N - 1 < N$),

$$\det(X_j) = -\det(X_{j-1}), \quad 1 \leq j \leq q-2,$$

and so

$$\det(\widehat{A}_{q1}) = (-1)^{q-2}\det(\widehat{B}_{11}).$$

Thus

$$(-1)^{q+1}(a_{q1}\det(\widehat{A}_{q1})) = (-1)^{q+1}(b_{11}(-1)^{q+2}\det(\widehat{B}_{11})) = -b_{11}(-1)^{1+1}\det(\widehat{B}_{11}),$$

as claimed. The second identity holds by symmetry. This completes the proof.

□

1.11. Theorem. *Let $n \in \mathbb{N}$, $\mathbb{F}$ be a field, $\kappa \in \mathbb{F}$ and $A \in \mathbb{M}_n(\mathbb{F})$.*

(a) *If A has two identical rows, then*

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

(b) *If $1 \le p \ne q \le n$ and B is obtained from A be adding $\kappa \mathrm{R}_p(A)$ to $\mathrm{R}_q(A)$, then*

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

Proof.

(a) We shall argue by induction on n. When $n = 2$, this is a routine calculation which we leave to the reader. Suppose that $3 \le N \in \mathbb{N}$ and that $\det(X) = 0$ whenever $2 \le n < N$ and $X \in \mathbb{M}_n(\mathbb{F})$ has two identical rows.

Let T be the matrix obtained from A by interchanging $\mathrm{R}_p(A)$ and $\mathrm{R}_1(A)$. (If $p = 1$, we set $T = A$.)

Let B be the matrix obtained from T by interchanging $\mathrm{R}_q(T)$ and $\mathrm{R}_2(T)$. (If $q = 2$, we set $B = T$.)

As a consequence of Theorem 1.10, we easily find that $\det(B) = \pm\det(A)$.

Note that $\mathrm{R}_1(B) = \mathrm{R}_2(B)$ and that

$$\det(B) = \sum_{i=1}^{N}(-1)^{i+1}b_{i1}\det(\widehat{B}_{i1}).$$

Now, if $3 \le i \le n$, then $\widehat{B}_{i1} \in \mathbb{M}_{N-1}(\mathbb{F})$ has two identical rows, and so $\det(\widehat{B}_{i1}) = 0$ by the induction hypothesis. Thus

$$\det(B) = (-1)^{1+1}b_{11}\det(\widehat{B}_{11}) + (-1)^{2+1}b_{21}\det(\widehat{B}_{21}).$$

But $b_{11} = b_{21}$ and $\widehat{B}_{11} = \widehat{B}_{21}$, so that

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

Since $\det(A) = \pm\det(B)$, we see that $\det(A) = 0$ as well.

(b) Let C be the matrix obtained from A by replacing the q^{th} row of A by the p^{th} row of A; that is,

$$\mathrm{R}_i(C) = \mathrm{R}_i(A), \quad 1 \le i \ne q \le n; \text{ and}$$
$$\mathrm{R}_q(C) = \mathrm{R}_p(A).$$

Let D be the matrix obtained from C by multiplying the q^{th} row of C by κ, so that

$$\begin{aligned} &\mathrm{R}_i(D) = \mathrm{R}_i(C) = \mathrm{R}_i(A), \quad 1 \le i \ne q \le n; \text{ and} \\ &\mathrm{R}_q(C) = \kappa \mathrm{R}_q(C) = \kappa \mathrm{R}_p(A). \end{aligned}$$

By Theorem 1.8, $\det(D) = \kappa \det(C)$. But C has two identical rows, and so from part (a) above, $\det(C) = 0$. Thus $\det(D) = 0$.

Note that

$$\begin{aligned} &\mathrm{R}_i(B) = \mathrm{R}_i(A) = \mathrm{R}_i(D), \quad 1 \le i \ne q \le n; \text{ and} \\ &\mathrm{R}_q(B) = \mathrm{R}_q(A) + \mathrm{R}_q(D). \end{aligned}$$

By Theorem 1.7,

$$\det(B) = \det(A) + \det(D) = \det(A) + 0 = \det(A).$$

□

1.12. Remark. In the case where the characteristic of the field $\mathbb{F}$ is not equal to 2, part (a) of Theorem 1.11 admits a much simpler proof, namely: if A has two identical rows R_p and R_q with $1 \le p \ne q \le n$, we let B be the matrix obtained from A by interchanging these two rows.

On the one hand, $\det(B) = -\det(A)$ by Theorem 1.10, but on the other hand, $B = A$ and so $\det(B) = \det(A)$. This shows that

$$\det(A) = -\det(A),$$

and if $\mathrm{CHAR}(\mathbb{F}) \ne 2$, then $\det(A) = 0$.

A number of references appear to have overlooked the issue of the characteristic of $\mathbb{F}$ and falsely claimed this as the proof for all fields. Caveat emptor.

Recall that if B is a matrix obtained from the matrix $A \in \mathbb{M}_n(\mathbb{F})$ by performing an elementary row operation, and if E is the elementary matrix obtained from I_n by performing the same elementary row operation, then $B = EA$. Combining that with the previous Theorems immediately yields the following:

1.13. Proposition. *Let $n \in \mathbb{N}$, $\kappa \in \mathbb{F}$ and $A \in \mathbb{M}_n(\mathbb{F})$.*

(a) *If E is the elementary matrix obtained from I_n by multiplying one of its rows by $\kappa \in \mathbb{F}$, then ,* $\det(EA) = \kappa \det(A)$. *In particular,* $\det(E) = \det(EI_n) = \kappa \det(I_n) = \kappa$.

(b) *If E is the elementary matrix obtained from I_n by interchanging two of its rows, then* $\det(EA) = -\det A$. *In particular,* $\det(E) = \det(E) = \det(EI_n) = -1 \det(I_n) = -1$.

(c) *If E is the elementary matrix obtained from I_n by adding $\kappa \mathrm{R}_p(I_n)$ to $\mathrm{R}_q(I_n)$, then* $\det(EA) = \det(A)$. *In particular,* $\det(E) = \det(EI_n) = \det(I_n) = 1$.

1.14. Remark. Observe that in each of the above three cases, we find that

- FOR ANY $A \in \mathbb{M}_n(\mathbb{C})$ AND ANY ELEMENTARY MATRIX E, WE HAVE

$$\det(EA) = \det(E)\,\det(A).$$

By an easy induction argument, if $X = E_1E_2\cdots E_q$, where E_j is an elementary matrix, $1 \le j \le q$, then

$$\det(XA) = \left(\prod_{j=1}^{q} \det(E_j)\right)\det(A).$$

In particular, by setting $A = I_n$, we see that $\det(X) = \prod_{j=1}^{n} \det(E_j)$.

Recall also that if $\operatorname{rank}(A) = r$, then there exist invertible operators $B, C \in \mathbb{M}_n(\mathbb{C})$ such that

$$BAC = \begin{bmatrix} I_r & 0 \\ 0 & 0_{n-r} \end{bmatrix}.$$

1.15. Theorem. *Let $A, B \in \mathbb{M}_n(\mathbb{F})$. Then*

(a) $\det(AB) = \det(A)\det(B)$.
(b) *If $S \in \mathbb{M}_n(\mathbb{F})$ is invertible,* $\det(S^{-1}) = (\det(S))^{-1}$.
(c) *If $A \sim B$, then* $\det(A) = \det(B)$.

Proof.

(a) We consider two cases.

CASE ONE. If A is invertible, then by Corollary 6.**??**, A invertible implies that we may write as a finite product of elementary matrices, say $A = E_1E_2\cdots E_q$. The result now follows from Remark 1.14.

CASE TWO.

Now suppose that A is not invertible, so that $r := \operatorname{rank}(A) < n$. It follows that $\dim(\text{ROW}(A)) = r < n$, and so there exists a row $\text{R}_q(A)$ which is a linear combination of the remaining rows of A, say

$$\text{R}_q(A) = \sum_{1 \le j \ne q \le n} \kappa_j \text{R}_j(A).$$

Let $T_0 = A$ and for $1 \le j \ne q \le n$, let T_j be the matrix obtained from A by adding $-\kappa_j \text{R}_j(T_{j-1})$ to $\text{R}_j(T_{j-1})$. Then $\text{R}_q(T_n) = \mathbf{0}$, and so $\det(T_n) = 0$. But by Theorem 1.11,

$$\begin{aligned}\det(T_n) &= \det(T_{n-1}) = \cdots = \det(T_{q+1}) \\ &= \det(T_{q-1}) = \cdots = \det(T_1) = \det(T_0) = \det(A).\end{aligned}$$

Thus $\det(A) = 0$.

Applying the same argument to AB which also has rank at most A (and thus less than n), we see that

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

(b) If $S \in \mathbb{M}_n(\mathbb{F})$ is invertible, then

$$1 = \det(I_n) = \det(SS^{-1}) = \det(S)\,\det(S^{-1}),$$

from which the result immediately follows.

(c) If $B \sim A$, then there exists $S \in \mathbb{M}_n(\mathbb{C})$ such that $B = S^{-1}AS$. It follows that

$$\det(B) = \det(S^{-1}AS) = \det(S^{-1})\,\det(A)\,\det(S) = \det(A).$$

□

1.16. Exercise. Let $n \in \mathbb{N}$ and suppose that $E \in \mathbb{M}_n(\mathbb{F})$ is an elementary matrix. Then $\det(E^{\mathrm{t}}) = \det(E)$.

1.17. Corollary. *If $A \in \mathbb{M}_n(\mathbb{F})$, then $\det(A) = \det(A^t)$.*

Proof. If A is not invertible, then the row space of A has dimension less than n, and so the column space of A^{t} has dimension less than n, implying that A^{t} is not invertible either. Thus if A is not invertible, then

$$\det(A) = 0 = \det(A^{\mathrm{t}}).$$

If A is invertible, then we can find elementary matrices E_j, $1 \le j \le r$ such that $A = E_1E_2\cdots E_r$. By the above Exercise, $\det E = \det(E^t)$ for any elementary matrix, and so

$$\begin{aligned}
\det(A^{\mathrm{t}}) &= \det(E_r^{\mathrm{t}}E_{r-1}^{\mathrm{t}}\cdots E_1^{\mathrm{t}}) \\
&= \prod_{j=1}^{r} \det(E_j^{\mathrm{t}}) \\
&= \prod_{j=1}^{r} \det(E_j) \\
&= \det(E_1E_2\cdots E_r) \\
&= \det(A).
\end{aligned}$$

□

1.18. Corollary. *Let $A \in \mathbb{M}_n(\mathbb{F})$ and suppose that B is the matrix obtained from A by interchanging two columns of A. Then*

$$\det B = -\det(A).$$

Moreover, the determinant function is linear in each column (in the sense of Exercise 7.1).

We have used the next Proposition earlier without stating it in its gory detail. We are now ready to do so.

1.19. Proposition. *Let $A \in \mathbb{M}_n(\mathbb{F})$ and enumerate the rows of A as*

$$(\mathrm{R}_1, \mathrm{R}_2, \ldots, \mathrm{R}_n).$$

Let $1 \leq q \leq n$ and suppose that B is the matrix whose rows are

$$(\mathrm{R}_2, \mathrm{R}_3, \ldots, \mathrm{R}_q, \mathrm{R}_1, \mathrm{R}_{q+1}, \mathrm{R}_{q+2}, \ldots, \mathrm{R}_n).$$

*We shall say that B is obtained from A by **cyclically permuting the first q rows of A**.*

Then

$$\det(B) = (-1)^{q-1}\det(A).$$

Similarly, if B is obtained from A by cyclically permuting the first q columns of A, then

$$\det(B) = (-1)^{q-1}\det(A).$$

Proof. Clearly the second statement follows from the first by simply considering transposes.

As for the first statement, let $T_0 = A$ and for each $1 \leq j \leq q-1$, let T_j be the matrix obtained from T_{j-1} by interchanging $\mathrm{R}_j(T_{j-1})$ and $\mathrm{R}_{j+1}(T_{j-1})$. A simple calculation shows that $T_{q-1} = B$, whence

$$\begin{aligned}
\det(B) &= \det(T_{q-1}) \\
&= (-1)\det(T_{q-2}) \\
&= (-1)^2\det(T_{q-3}) \\
&= \cdots \\
&= (-1)^{q-1}\det(T_0) \\
&= (-1)^{q-1}\det(A).
\end{aligned}$$

□

1.20. Lemma. *Let $A \in \mathbb{M}_n(\mathbb{F})$ and suppose that there exists $1 \leq p, q \leq n$ such that $\mathrm{R}_p = e_q$, where $e_q = (0, 0, \ldots, 0, 1, 0, \ldots, 0)$ with the non-zero entry occurring in the q^{th} position.. Then*

$$\det(A) = (-1)^{p+q}\det(\widehat{A}_{pq}).$$

Proof. Case 1. $q = 1$. Then

$$A = \begin{bmatrix}
a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\
a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\
\vdots & & & & \vdots \\
1 & 0 & 0 & \cdots & 0 \\
\vdots & & & & \vdots \\
a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn}
\end{bmatrix}.$$

Now $\det(A) = \sum_{i=1}^{n}(-1)^{i+1}a_{i1}\det(\widehat{A}_{i1})$.

But if $1 \le i \ne p \le n$, then $\widehat{A}_{i1}$ admits a row of 0's, and thus $\det(\widehat{A}_{i1}) = 0$. Moreover, $a_{p1} = 1$, and hence

$$\det(A) = (-1)^{p+1} a_{p1} \det(\widehat{A}_{p1}) = (-1)^{p+1} \det(\widehat{A}_{p1}),$$

as required.

CASE 2. $2 \le q \le n$.

Let B be the matrix obtained from A by cyclically permuting the first q columns of A. By Proposition 1.19,

$$\det(B) = (-1)^{q-1} \det(A).$$

Since B is of the form required for CASE 1,

$$\det(B) = (-1)^{p+1} \det(\widehat{B}_{p1}).$$

Finally, we leave it to the reader to verify that $\widehat{B}_{p1} = \widehat{A}_{pq}$, and thus

$$\det(A) = (-1)^{q-1}(-1)^{p+1} \det(\widehat{B}_{p1}) = (-1)^{p+q} \det(\widehat{A}_{pq}).$$

□

1.21. Theorem. *Let $A \in \mathbb{M}_n(\mathbb{F})$. Then*

(a) *for any $1 \le p \le n$,*

$$\det(A) = \sum_{j=1}^{n} (-1)^{p+j} a_{pj} \det(\widehat{A}_{pj}).$$

(b) *For any $1 \le q \le n$,*

$$\det(A) = \sum_{i=1}^{n} (-1)^{i+q} a_{iq} \det(\widehat{A}_{iq}).$$

In other words, we may expand the determinant along any row or any column.

Proof. The proof of (b) follows from that of (a) by considering transposes. Thus, we restrict our attention to (a).

Let $A = [a_{ij}] \in \mathbb{M}_n(\mathbb{F})$ and note that $r_p(A) = (a_{p1}, a_{p2}, a_{p3}, \ldots, a_{pn}) = \sum_{j=1}^{n} a_{pj} e_j$. By the linearity of the determinant in any row (see Exercise 1), and by Lemma 1.20,

$$\det(A) = \sum_{j=1}^{n} a_{pj} \det \begin{bmatrix} \mathrm{R}_1(A) \\ \mathrm{R}_2(A) \\ \vdots \\ \mathrm{R}_{p-1}(A) \\ e_j \\ \mathrm{R}_{p+1}(A) \\ \vdots \\ \mathrm{R}_n(A) \end{bmatrix}$$

$$= \sum_{j=1}^{n} a_{pj} (-1)^{p+j} \det(\widehat{A}_{pj}).$$

□

It is time we considered an example.

1.22. Example. Let

$$A = \begin{bmatrix} 2 & 0 & -1 & 3 \\ 3 & -2 & 4 & 0 \\ 6 & 1 & 3 & 0 \\ 7 & 2 & 0 & 5 \end{bmatrix}.$$

We wish to calculate $\det(A)$.

By Theorem 1.21, we may calculate it by expansion along any row or any column. Since the fourth column has two zero entries, this will reduce the number of calculations we must make.

Thus

$$\det(A) = (-1)^{1+4}3 \det \begin{bmatrix} 3 & -2 & 4 \\ 6 & 1 & 3 \\ 7 & 2 & 0 \end{bmatrix} + 0(*) + 0(*) + (-1)^{4+4}5 \det \begin{bmatrix} 2 & 0 & -1 \\ 3 & -2 & 4 \\ 6 & 1 & 3 \end{bmatrix}.$$

Let $T_1 = \begin{bmatrix} 3 & -2 & 4 \\ 6 & 1 & 3 \\ 7 & 2 & 0 \end{bmatrix}$ and $T_2 = \begin{bmatrix} 2 & 0 & -1 \\ 3 & -2 & 4 \\ 6 & 1 & 3 \end{bmatrix}$.

To calculate $\det(T_1)$, we might expand along the third row, since it has a zero entry.

$$\begin{aligned} \det(T_1) &= (-1)^{3+1}\ 7 \det \begin{bmatrix} -2 & 4 \\ 1 & 3 \end{bmatrix} + (-1)^{3+2}\ 2 \det \begin{bmatrix} 3 & 4 \\ 6 & 3 \end{bmatrix} + 0(*) \\ &= 7(-6-4) - 2(9-24) \\ &= -40. \end{aligned}$$

To calculate $\det(T_2)$, we might expand along the first row, since it has a zero entry.

$$\begin{aligned} \det(T_2) &= (-1)^{1+1}\ 2 \det \begin{bmatrix} -2 & 4 \\ 1 & 3 \end{bmatrix} + 0(*) + (-1)^{1+3}\ (-1) \det \begin{bmatrix} 3 & -2 \\ 6 & 1 \end{bmatrix} + 0(*) \\ &= 2(-6-4) + (-1)(3+12) \\ &= -35. \end{aligned}$$

From above, we see that

$$\det(A) = (-3)\det(T_1) + (5)\det(T_2) = (-3)(-40) + (5)(-35) = -55.$$

Supplementary Examples

S7.1. Example. Suppose that we wish to determine whether or not the vectors $x_1 = (3,2,1)$, $x_2 = (31,2,-2)$ and $x_3 = (9,-6,5) \in \mathbb{R}^3$ are linearly independent. Consider

$$A = \begin{bmatrix} 3 & 2 & 1 \\ 31 & 2 & -2 \\ 9 & -6 & 5 \end{bmatrix},$$

the matrix whose i^{th} row is x_i, $1 \leq i \leq 3$.

Then $\det(A) = -556 \neq 0$. It follows that A is invertible, and so the rows of A are linearly independent; i.e. $\{x_1, x_2, x_3\}$ is linearly independent.

Note that had $\det(A)$ been equal to zero, then A would not have been invertible, and so the rows of A would have been linearly *dependent*. In other words, this test works in both cases.

S7.2. Example. The following example is based upon **Cramer's Rule**, which is proved in **Theorem A7.6**.

Consider the system $\mathbb{S}$ of two equations in two unknowns given by

$$\begin{array}{rcrcl} 3x_1 & + & 5x_2 & = & 4 \\ 2x_1 & - & 4x_2 & = & 3 \end{array}$$

Observe that the coefficient matrix $A = \begin{bmatrix} 3 & 5 \\ 2 & -4 \end{bmatrix}$ is invertible, as (for example) $\det(A) = -22 \neq 0$. As such, the system is consistent, and the solution is given by

$$x = A^{-1} \begin{bmatrix} 4 \\ 3 \end{bmatrix}.$$

Applying Cramer's Rule, we see that with $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, we have

$$x_1 = \det(A)^{-1} \cdot \det(\begin{bmatrix} 4 & 5 \\ 3 & -4 \end{bmatrix}) = -\frac{1}{22}(-31) = \frac{31}{22},$$

and

$$x_2 = \det(A)^{-1} \cdot \det(\begin{bmatrix} 3 & 4 \\ 2 & 3 \end{bmatrix}) = -\frac{1}{22}(1) = -\frac{1}{22}.$$

S7.3. Example. Using the same system $\mathbb{S}$ as in Example S7.3, we can use **Theorem A7.5** to determine A^{-1}:

We find that $\text{COF}(A) = \begin{bmatrix} -4 & -2 \\ -5 & 3 \end{bmatrix}$, and thus $\text{ADJ}(A) = \begin{bmatrix} -4 & -5 \\ -2 & 3 \end{bmatrix}$, so that

$$A^{-1} = -\frac{1}{22} \begin{bmatrix} -4 & -5 \\ -2 & 3 \end{bmatrix},$$

and therefore

$$x = -\frac{1}{22}\begin{bmatrix} -4 & -5 \\ -2 & 3 \end{bmatrix}\begin{bmatrix} 4 \\ 3 \end{bmatrix} = -\frac{1}{22}\begin{bmatrix} -31 \\ 1 \end{bmatrix} = \begin{bmatrix} 31/22 \\ -1/22 \end{bmatrix}.$$

This answers agrees with that found above, which is a positive sign about the overall consistency of mathematics.

S7.4. Example. Using the same system $\mathbb{S}$ as in Example S7.3, we can also compute the inverse of A directly as follows:

$$\begin{bmatrix} 3 & 5 & 1 & 0 \\ 2 & -4 & 0 & 1 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 5/3 & 1/3 & 0 \\ 2 & -4 & 0 & 1 \end{bmatrix}$$
$$\longrightarrow \begin{bmatrix} 1 & 5/3 & 1/3 & 0 \\ 0 & -22/3 & -2/3 & 1 \end{bmatrix}$$
$$\longrightarrow \begin{bmatrix} 1 & 5/3 & 1/3 & 0 \\ 0 & 1 & 1/11 & -3/22 \end{bmatrix}$$
$$\longrightarrow \begin{bmatrix} 1 & 0 & 2/11 & 5/22 \\ 0 & 1 & 1/11 & -3/22 \end{bmatrix}$$

so that

$$A^{-1} = \begin{bmatrix} 2/11 & 5/22 \\ 1/11 & -3/22 \end{bmatrix},$$

after which we find that $x = A^{-1}\begin{bmatrix} 4 \\ 3 \end{bmatrix} = \begin{bmatrix} 31/22 \\ -1/22 \end{bmatrix}$ is the unique solution to the system.

S7.5. Example. We could also have solved the system by applying Gaussian elimination, which is far less painful than its name might suggest.

$$\begin{array}{rcrcl} 3x_1 & + & 5x_2 & = & 4 \\ 2x_1 & - & 4x_2 & = & 3 \end{array}$$

We apply elementary row operations to the augmented system $[A\ b]$ to put it in RREF.

$$[A|b] = \begin{bmatrix} 3 & 5 & 4 \\ 2 & -4 & 3 \end{bmatrix}$$
$$\longrightarrow \begin{bmatrix} 1 & 5/3 & 4/3 \\ 2 & -4 & 3 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 5/3 & 4/3 \\ 0 & -22/3 & 1/3 \end{bmatrix}$$
$$\longrightarrow \begin{bmatrix} 1 & 5/3 & 4/3 \\ 0 & 1 & -1/22 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 0 & 31/22 \\ 0 & 1 & -1/22 \end{bmatrix}$$

Thus $\mathrm{Sol}(\mathbb{S}) = \{\begin{bmatrix} 31/22 \\ -1/22 \end{bmatrix}\}$.

The last four examples shows that, even without alluding to taxidermal tendencies, there are often multiple ways to accomplish the same task.

S7.6. Example. Let $A = \begin{bmatrix} 1 & 0 & 0 \\ 3 & 6 & 0 \\ -2 & 3 & 19 \end{bmatrix} \in \mathbb{M}_3(\mathbb{Q})$.

Since $\det(A) = \det(A^{\mathrm{t}})$, and since $A^{\mathrm{t}} = \begin{bmatrix} 1 & 3 & -2 \\ 0 & 6 & 3 \\ 0 & 0 & 19 \end{bmatrix}$ is upper-triangular, we find that

$$\det(A) = 1 \cdot 6 \cdot 19 = 105.$$

Alternatively, we could have expanded $\det(A)$ directly by cofactors along the first row to find:

$$\det(A) = (-1)^{1+1}(1)\det\begin{bmatrix} 6 & 0 \\ 3 & 19 \end{bmatrix} = 1 \cdot (6(19) - 0(3)) = 105.$$

S7.7. Example. Suppose that $p_1 = (x_1, y_1)$, $p_2 = (x_2, y_2)$ and $p_3 = (x_3, y_3)$ are three points in $\mathbb{R}^2$. We may view $\mathbb{R}^2$ as the xy-plane of $\mathbb{R}^3$, and thereby identify p_1, p_2, p_3 with the points $q_1 = (x_1, y_1, 0)$, $q_2 = (x_2, y_2, 0)$ and $q_3 = (x_3, y_3, 0)$ sitting in $\mathbb{R}^3$. Of course p_1, p_2, p_3 are collinear if and only if q_1, q_2, and q_3 are collinear.

In fact, we given q_1, q_2, q_3 in the xy-plane of $\mathbb{R}^3$, we can translate the points upwards by one unit (i.e. in the direction of the positive z-axis) to obtain points

$$w_k := q_k + (0, 0, 1), \quad 1 \leq k \leq 3,$$

and the points w_1, w_2, w_3 will be collinear if and only if q_1, q_2, q_3 are, and thus if and only if p_1, p_2, p_3 are.

But $\{w_1, w_2, w_3\}$ are collinear if and only if there exists $\alpha \in \mathbb{R}$ such that $w_3 = w_1 + \alpha(w_2 - w_1) = (1 - \alpha)w_1 + \alpha w_2 \in \mathrm{span}\{w_1, w_2\}$, and this in turn happens if and only if $\det(A) = 0$, where $\mathrm{R}_k(A) = w_k$, $1 \leq k \leq 3$.

Putting this all together, p_1, p_2, p_3 are collinear if and only if

$$\det\begin{bmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{bmatrix} = 0.$$

For example, if $p_1 = (1, 3)$, $p_2 = (5, 3)$ and $p_3 = (2, 2)$, then

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

so the points p_1, p_2, p_3 are not collinear.

S7.8. Example. Let's put Example S7.7 to better use. Suppose we are asked to find the equation of the line in $\mathbb{R}^2$ passing through $p_1 := (1,7)$ and $p_2 := (4,6)$. Any point (x,y) on the line must be collinear with p_1 and p_2, so from the above analysis,

$$\det\begin{bmatrix} x & y & 1 \\ 1 & 7 & 1 \\ 4 & 6 & 1 \end{bmatrix} = 0.$$

That is,

$$x + 3y - 22 = 0,$$

or equivalently,

$$y = -\frac{1}{3}x + \frac{22}{3}.$$

More generally, if $p_1 = (x_1, y_1)$ and $p_2 = (x_2, y_2)$ are two points in $\mathbb{R}^2$, then the equation of the linear through these points is given by the equation

$$\det\begin{bmatrix} x & y & 1 \\ x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \end{bmatrix} = 0.$$

S7.9. Example. We leave it to the reader to convince themselves (by applying similar arguments to those that were so profitably used in Example S7.7) that the equation of the plane in $\mathbb{R}^3$ passing through the (non-collinear) points $p_1 = (x_1, y_1, z_1)$, $p_2 = (x_2, y_2, z_2)$ and $p_3 = (x_3, y_3, z_3)$ is given by the equation

$$\det\begin{bmatrix} x & y & z & 1 \\ x_1 & y_1 & y_3 & 1 \\ x_2 & y_2 & z_2 & 1 \\ x_3 & y_3 & z_3 & 1 \end{bmatrix} = 0.$$

For example, if $p_1 = (3,2,1)$, $p_2 = (5,3,7)$ and $p_3 = (2,-1,1)$, then the equation of the plane π passing through these three points is:

$$0 = \det\begin{bmatrix} x & y & z & 1 \\ 3 & 2 & 1 & 1 \\ 5 & 3 & 7 & 1 \\ 2 & -1 & 1 & 1 \end{bmatrix} = 18x - 6y - 5z - 37.$$

Note: It is important that the three points be non-collinear, otherwise the calculation of the determinant of the matrix will be zero regardless of the choice of $(x,y,z) \in \mathbb{R}^3$, which stands to reason, geometrically speaking, as three collinear points and any other point will always lie in the same plane.

S7.10. Example. The next Example is based upon Exercise 7.3 at the end of this Chapter. It states that if $x = (x_1, x_2, x_3)$ and $y = (y_1, y_2, y_3) \in \mathbb{R}^3$, then x is perpendicular to y if and only if

$$\langle x, y \rangle := x_1y_1 + x_2y_2 + x_3y_3 = 0.$$

Now suppose that $v = (v_1, v_2, v_3)$ and $w = (w_1, w_2, w_3) \in \mathbb{R}^3$ are linearly independent. Consider

$$z := \det \begin{bmatrix} e_1 & e_2 & e_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{bmatrix} = (v_2w_3 - w_2v_3)e_1 - (v_1w_3 - w_1v_3)e_2 + (v_1w_2 - w_1v_2)e_3.$$

Now we could do the unthinkable and we treat e_1, e_2 and e_3 as the standard basis vectors for $\mathbb{R}^3$, and use this to compute z. This, however, makes no sense as we can't have e_1 be both a vector in $\mathbb{R}^3$, and a single entry in the matrix whose determinant we are computing. In order to avoid doing something that is basically so against the laws of nature, we just define z to be the vector in $\mathbb{R}^3$ whose coordinates are the coefficients of e_1, e_2 and e_3 above. There – no harm done, our consciences are clean, and we obtain the same result.

With this convention, we find that $z = (v_2w_3 - w_2v_3, -v_1w_3 + w_1v_3, v_1w_2 - w_1v_2)$. Then

(a) $\langle z, v \rangle = v_1(v_2w_3 - w_2v_3) + v_2(-v_1w_3 + w_1v_3) + v_3(v_1w_2 - w_1v_2) = 0$, and
(b) $\langle z, w \rangle = w_1(v_2w_3 - w_2v_3) + w_2(-v_1w_3 + w_1v_3) + w_3(v_1w_2 - w_1v_2) = 0$.

Thus z is perpendicular to both v and w. In fact, a finer analysis can be done so that one can determine the "*orientation*" of the system v, w, z, but we shall not discuss this here.

Appendix

A7.1. We made a bold claim just after Theorem 7.1.8 that "$\det(\cdot)$ is linear in any row".

Of course, by considering transposes, we see that $\det(\cdot)$ is also "linear" in any column.

Let's prove this. The first thing that we shall do will be to use induction to extend Theorem 7.1.7 as follows:

Theorem. *Let $m, n \in \mathbb{N}$ with $m \geq 2$, and suppose that $A, B_1, B_2, \ldots, B_m \in \mathbb{M}_n(\mathbb{F})$. Suppose that there exists $1 \leq p \leq n$ such that*

- $\mathrm{R}_i(A) = \mathrm{R}_i(B_j)$ *for all* $1 \leq j \leq m$ *and* $1 \leq i \leq p \leq n$*; and*
- $\mathrm{R}_p(A) = \mathrm{R}_p(B_1) + \mathrm{R}_p(B_2) + \cdots \mathrm{R}_p(B_m)$.

Then

$$\det(A) = \det(B_1) + \det(B_2) + \cdots + \det(B_m).$$

Proof. If $m = 2$, then this is exactly Theorem 7.1.7.

Now suppose that $M \geq 3$ and that the result holds whenever we have $m < M$. Let $B_1, B_2, \ldots, B_M \in \mathbb{M}_n(\mathbb{F})$ with

- $\mathrm{R}_i(A) = \mathrm{R}_i(B_j)$ for all $1 \leq j \leq M$ and $1 \leq i \neq p \leq n$; and
- $\mathrm{R}_p(A) = \mathrm{R}_p(B_1) + \mathrm{R}_p(B_2) + \cdots \mathrm{R}_p(B_M)$.

Let $C \in \mathbb{M}_n(\mathbb{F})$ be the matrix satisfying

- $\mathrm{R}_i(C) = \mathrm{R}_i(A)$ for all $1 \leq i \neq p \leq n$; and
- $\mathrm{R}_p(C) = \mathrm{R}_p(B_1) + \mathrm{R}_p(B_2) + \cdots \mathrm{R}_p(B_{M-1})$.

Note that

- $\mathrm{R}_i(A) = \mathrm{R}_i(C) = \mathrm{R}_i(B_M)$ for all $1 \leq i \neq p \leq n$; and
- $\mathrm{R}_p(A) = \mathrm{R}_p(C) + \mathrm{R}_p(B_M)$.

Since our result holds for $m = 2$, we find that

$$\det(A) = \det(C) + \det(B_M).$$

By the induction step, however,

$$\det(C) = \det(B_1) + \det(B_2) + \cdots + \det(B_{M-1}).$$

Thus

$$\det(A) = \det(B_1) + \det(B_2) + \cdots + \det(B_M),$$

completing the induction step and the proof.

□

A7.2. Here is the promised version of linearity we claimed above.

Corollary. *Let $n \in \mathbb{N}$ and $A \in \mathbb{M}_n(\mathbb{F})$. Fix $1 \le p \le n$ and write $\mathrm{R}_p(A) = \sum_{j=1}^n a_{pj} e_j$, where $\mathcal{D} = (e_1, e_2, \ldots, e_n)$ is the standard ordered basis for $\mathbb{F}^n$.*

For $1 \le j \le n$, let $B_j \in \mathbb{M}_n(\mathbb{F})$ be the matrix satisfying

- $\mathrm{R}_i(B_j) = \mathrm{R}_i(A)$ *for all $1 \le i \ne p \le n$; and*
- $\mathrm{R}_p(B_j) = e_j$.

Then

$$\det(A) = \sum_{j=1}^n a_{pj} \det(B_j).$$

Proof. For $1 \le j \le n$, let $D_j \in \mathbb{M}_n(\mathbb{F})$ be the matrix obtained from B_j by multiplying the p^{th} row of B_j by a_{pj}. Then

- $\mathrm{R}_i(A) = \mathrm{R}_i(D_j)$ for all $1 \le i \ne p \le n$; and
- $\mathrm{R}_p(A) = \mathrm{R}_p(D_1) + \mathrm{R}_p(D_2) + \cdots + \mathrm{R}_p(D_n)$.

By the above Theorem,

$$\det(A) = \det(D_1) + \det(D_2) + \cdots + \det(D_n).$$

On the other hand, by Theorem 7.1.8(a),

$$\det(D_j) = a_{pj} \det(B_j), \quad 1 \le j \le n.$$

Thus

$$\det(A) = \sum_{j=1}^n a_{pj} \det(B_j).$$

□

A7.3 An application of determinants. Let $n \in \mathbb{N}$, $\mathbb{F}$ be a field, and $A = [a_{ij}] \in \mathbb{M}_n(\mathbb{F})$. We define the **cofactor matrix** of A to be the matrix

$$\mathrm{COF}(A) := [(-1)^{i+j} \det(\widehat{A}_{ij})].$$

In other words, the (i,j)-entry of $\mathrm{COF}(A)$ is the (i,j)-cofactor of A.

We next define the **adjugate** of A to be $\mathrm{ADJ}(A) := (\mathrm{COF}(A))^{\mathrm{t}}$.

A7.4 Examples. Let $A = \begin{bmatrix} 1 & 3 & 1 \\ 3 & 0 & 1 \\ -2 & 2 & 1 \end{bmatrix}$.

Then

$$\mathrm{COF}(A) = \begin{bmatrix} -2 & -5 & 6 \\ -1 & 3 & -8 \\ 3 & 2 & -9 \end{bmatrix},$$

while

$$\mathrm{ADJ}(A) = \begin{bmatrix} -2 & -1 & 3 \\ -5 & 3 & 2 \\ 6 & -8 & -9 \end{bmatrix}.$$

Observe furthermore that $\det(A) = -11$, implying that A is invertible.

Here is where something interesting happens: consider

$$\begin{aligned} A \cdot \frac{1}{\det(A)}\text{ADJ}(A) &= -\frac{1}{11}\begin{bmatrix} 1 & 3 & 1 \\ 3 & 0 & 1 \\ -2 & 2 & 1 \end{bmatrix} \cdot \begin{bmatrix} -2 & -1 & 3 \\ -5 & 3 & 2 \\ 6 & -8 & -9 \end{bmatrix} \\ &= -\frac{1}{11}\begin{bmatrix} -11 & 0 & 0 \\ 0 & -11 & 0 \\ 0 & 0 & -11 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \end{aligned}$$

Let us prove that this is not just a coincidence.

A7.5 Theorem. *Let $n \in \mathbb{N}$, $\mathbb{F}$ be a field, and $A = [a_{ij}] \in \mathbb{M}_n(\mathbb{F})$. Then*

$$A \cdot \text{ADJ}(A) = \det(A) I_n.$$

In particular, if A is invertible, then

$$A^{-1} = \det(A)^{-1}\text{ADJ}(A).$$

Proof. The proof relies heavily upon the following elementary and useful observation:

if $X = [x_{ij}] \in \mathbb{M}_{m\times q}(\mathbb{F})$ and $Y = [y_{ij}] \in \mathbb{M}_{q\times n}(\mathbb{F})$, then

$$X \cdot Y = [z_{ij}] \in \mathbb{M}_{m\times n}(\mathbb{F}),$$

where $z_{ij} = \text{R}_i(X) \cdot \text{C}_j(Y)$.

Returning to the proof, let $1 \le i \ne j \le n$, and let B be the matrix obtained from A by replacing the j^{th} row of A by the i^{th} row of A, and leaving all other rows of A unchanged.

Then $\text{R}_j(B) = \text{R}_i(A) = \text{R}_i(B)$, and so $\det(B) = 0$, as it has two identical rows. In particular, if we expand the determinant of B by cofactors along the j^{th} row, then we obtain

$$0 = \det(B) = \sum_{k=1}^{n} (-1)^{j+k} b_{jk} \det(\widehat{B}_{jk}).$$

Note, however, that $b_{jk} = a_{ik}$ and $\widehat{B}_{jk} = \widehat{A}_{jk}$ for $1 \le k \le n$, and therefore

$$0 = \sum_{k=1}^{n} (-1)^{j+k} a_{ik} \det(\widehat{A}_{jk}) = \sum_{k=1}^{n} a_{ik}\left((-1)^{j+k}\det(\widehat{A}_{jk})\right) = \text{R}_i(A) \cdot \text{C}_j(\text{ADJ}(A)).$$

That is, if $i \ne j$, then $\text{R}_i(A) \cdot \text{C}_j(\text{ADJ}(A)) = 0$.

On the other hand, for any $1 \leq i \leq n$,

$$\det(A) = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \det(\widehat{A}_{ij}) = \sum_{j=1}^{n} a_{ij} \left((-1)^{i+j} \det(\widehat{A}_{ij}) \right) = \mathrm{R}_i(A) \cdot \mathrm{C}_i(\mathrm{ADJ}(A)).$$

Thus for all $1 \leq i \leq n$, $\mathrm{R}_i(A) \cdot \mathrm{C}_i(\mathrm{ADJ}(A)) = \det(A)$.

By applying these results to the observation at the start of the proof, we find that

$$A \cdot \mathrm{ADJ}(A) = [\mathrm{R}_i(A) \cdot \mathrm{C}_j(\mathrm{ADJ}(A))] = \det(A) I_n.$$

The second statement follows trivially from this.

□

A7.6. The following technique for solving linear equations is known as **Cramer's Rule**.

Theorem. *Let $n \in \mathbb{N}$, $\mathbb{F}$ be a field and $\mathbb{S}$ be a system $Ax = b$ of n linear equations in n unknowns which admits a unique solution over $\mathbb{F}$. (Equivalently, suppose that $A \in \mathbb{M}_n(\mathbb{F})$ is invertible.) For $1 \leq j \leq n$, let A_j be the matrix obtained from A by replacing the j^{th} column of A by b. Then*

$$\mathrm{Sol}(\mathbb{S}) = \{x\},$$

where $x = (x_1, x_2, \ldots, x_n)^t$ is given by $x_j = \det(A_j) \cdot (\det(A))^{-1}$.

Proof. The hypothesis that $A \in \mathbb{M}_n(\mathbb{F})$ is invertible implies that $\det(A) \neq 0$. Moreover, we have seen that $\mathrm{Sol}(\mathbb{S}) = \{x\}$, where

$$x = A^{-1} b.$$

But from above, $A^{-1} = \det(A)^{-1} \mathrm{ADJ}(A)$, and so

$$x = \det(A)^{-1} \mathrm{ADJ}(A) b.$$

Using the observation from the beginning of the previous proof, we obtain that for each $1 \leq i \leq n$,

$$\begin{aligned} x_i &= \det(A)^{-1} \ \mathrm{R}_i(\mathrm{ADJ}(A)) \cdot \mathrm{C}_1(b) \\ &= \det(A)^{-1} \ \mathrm{R}_i(\mathrm{ADJ}(A)) \cdot b \\ &= \det(A)^{-1} \ \sum_{j=1}^{n} b_j \left((-1)^{i+j} \det(\widehat{A}_{ij}) \right) \\ &= \det(A)^{-1} \cdot \det(A_j). \end{aligned}$$

□

Exercises for Chapter 7

Exercise 7.1.
Let $\alpha_0, \alpha_1, \alpha_2, \ldots, \alpha_n \in \mathbb{F}$ and define the matrix

$$V = \begin{bmatrix} 1 & \alpha_0 & \alpha_0^2 & \cdots & \alpha_0^n \\ 1 & \alpha_1 & \alpha_1^2 & \cdots & \alpha_1^n \\ 1 & \alpha_2 & \alpha_2^2 & \cdots & \alpha_2^n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \alpha_n & \alpha_n^2 & \cdots & \alpha_n^n \end{bmatrix}.$$

Find $\det(V)$.

Exercise 7.2.
Let $x = (x_1, x_2)$ and $y = (y_1, y_2) \in \mathbb{R}^2$. Prove that x is perpendicular to y if and only if

$$\langle x, y \rangle := x_1 y_1 + x_2 y_2 = 0.$$

Exercise 7.3.
Let $x = (x_1, x_2, x_3)$ and $y = (y_1, y_2, y_3) \in \mathbb{R}^3$. Prove that x is perpendicular to y if and only if

$$\langle x, y \rangle := x_1 y_1 + x_2 y_2 + x_3 y_3 = 0.$$

Exercise 7.4.
Let $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$, $n \in \mathbb{N}$, and let $y = (y_j)_{j=1}^n, z = (z_j)_{j=1}^n \in \mathbb{K}^n$. We define the **inner product** of y and z to be:

$$\langle y, z \rangle := \sum_{j=1}^{n} y_j \overline{z_j}.$$

Of course, when $\mathbb{K} = \mathbb{R}$, the complex conjugate is superfluous, and we obtain the formulae indicated in the two previous exercises for $n = 2$ and $n = 3$ respectively.

Prove that

(a) $\langle x, x \rangle \geq 0$, and $\langle x, x \rangle = 0$ if and only if $x = \mathbf{0}$.
(b) $\langle \kappa x, y \rangle = \kappa \langle x, y \rangle$ for all vectors x, y and $\kappa \in \mathbb{K}$.
(c) $\langle x, y \rangle = \overline{\langle y, x \rangle}$ for all vectors x, y, and
(d) $\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$ for all vectors x, y, z.

Exercise 7.5.
More generally, let $\mathcal{V}$ be a vector space over $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$. Any function

$$\begin{array}{rccc} \langle \cdot, \cdot \rangle : & \mathcal{V} \times \mathcal{V} & \to & \mathbb{K} \\ & (x, y) & \mapsto & \langle x, y \rangle \end{array}$$

satisfying the four properties (a) - (d) of Exercise 7.4 above is called an **inner product** on $\mathcal{V}$.

Let $\mathcal{V} = \mathcal{C}([0,1],\mathbb{C})$ as a vector space over $\mathbb{C}$. For $f, g \in \mathcal{V}$, define

$$\langle f, g\rangle := \int_0^1 f(x)\overline{g(x)}\,dx.$$

Prove that this defines an inner product on $\mathcal{C}([0,1],\mathbb{C})$.

Exercise 7.6.

Let $\mathcal{V} = \mathcal{C}([0,1],\mathbb{C})$ as a vector space over $\mathbb{C}$, and let $r \in \mathcal{C}([0,1],\mathbb{C})$ be a function satisfying $r(x) > 0$ for all $x \in [0,1]$. For $f, g \in \mathcal{V}$, define

$$\langle f, g\rangle := \int_0^1 r(x)f(x)\overline{g(x)}\,dx.$$

Prove that this defines an inner product on $\mathcal{C}([0,1],\mathbb{C})$.

Would this still be the case if there were an interval $[a,b] \subseteq [0,1]$ where $r(x) = 0$ for all $x \in [a,b]$?

Exercise 7.7.

Let $A = \begin{bmatrix} 1 & 3 & 7 \\ 0 & 18 & -245 \\ 0 & 0 & 2 \end{bmatrix}$ and $B = \begin{bmatrix} 12 & 0 & 0 \\ -4 & 2 & 0 \\ 15 & 24 & 8 \end{bmatrix}$. Find $\det(AB)$.

Exercise 7.8.

Let $A \in \mathbb{M}_3(\mathbb{C})$. Prove or disprove the following statement: there exists a matrix $D = \begin{bmatrix} \alpha & 0 & 0 \\ 0 & \beta & 0 \\ 0 & 0 & \gamma \end{bmatrix} \in \mathbb{M}_3(\mathbb{C})$ such that $\operatorname{tr}(DA) \in \mathbb{R}$ and $\det(DA) \in \mathbb{R}$.

Exercise 7.9.

Let $n \in \mathbb{N}$ and $N \in \mathbb{M}_n(\mathbb{F})$. We say that N is **nilpotent of order** k if $N^k = 0 \neq N^{k-1}$. For example,

$$J_3 := \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

is nilpotent of order 3 in $\mathbb{M}_3(\mathbb{F})$.

Suppose that N is nilpotent of order k in $\mathbb{M}_n(\mathbb{F})$. Prove that $\det(N) = 0$.

Exercise 7.10.

Let $n \in \mathbb{N}$ and $E \in \mathbb{M}_n(\mathbb{F})$. We say that E is **idempotent** $E^2 = E$. For example,

$$E := \begin{bmatrix} 1 & 1023451 \\ 0 & 0 \end{bmatrix}$$

is idempotent in $\mathbb{M}_2(\mathbb{F})$.

Suppose that E is idempotent in $\mathbb{M}_n(\mathbb{F})$. Prove that $\det(E) = \{1, -1\}$.

CHAPTER 8

An introduction to eigenvalues and eigenvectors

Hedgehogs - why can't they just share the hedge?

Dan Antopolski

1. Eigenvalues, eigenvectors and eigenspaces

1.1. The last chapter of these notes will deal with eigenvalues, eigenvectors, and diagonalisation of matrices. Given more time, we would cover these in much greater depth. A number of issues regarding eigenvalues and eigenvalues will be dealt with in Math 245.

1.2. Definition. *Let $\mathcal{V}$ be a vector space over a field $\mathbb{F}$ and let $T \in \mathcal{L}(\mathcal{V})$. An element $\alpha \in \mathbb{F}$ is said to be an **eigenvalue** for T if there exists $0 \neq x \in \mathcal{V}$ such that $Tx = \alpha x$. We then say that x is an **eigenvector for** T **corresponding to** α.*

*We denote by $\sigma_p(T)$ the set of all eigenvalues of T. (The notation reflects the fact that the set of eigenvalues of T is also known as the **point spectrum** of T.)*

*Note that $x \in \mathcal{V}$ is an eigenvector for $T \in \mathcal{L}(\mathcal{V})$ if and only if $x \neq 0$ and $x \in \ker(T - \alpha I)$. More generally, therefore, given a vector space $\mathcal{V}$, $T \in \mathcal{L}(\mathcal{V})$ and $\beta \in \mathbb{F}$ as above, we define the **eigenspace** $E(T;\alpha)$ of T **corresponding to** α as*

$$E(T;\alpha) := \ker(T - \alpha I).$$

1.3. As always, we extend such concepts from linear maps (acting on finite-dimensional vector spaces) to matrices via the left-regular representation, so that if $n \in \mathbb{N}$, $\mathbb{F}$ is a field and $A \in \mathbb{M}_n(\mathbb{F})$, we say that α is an **eigenvalue** of A with corresponding **eigenvector** $0 \neq x \in \mathbb{F}^n$ provided that α is an eigenvalue of $L_A \in \mathcal{L}(\mathbb{F}^n)$ with corresponding eigenvector x. We leave it to the reader's overactive imagination to consider what an *eigenspace* for A corresponding to α might be.

1.4. Examples.

(a) Let $T \in \mathcal{L}(\mathbb{R}^3)$ be the operator $T(x,y,z) = (7x-2y, -2x+6y-2z, -2y+5z)$. Consider $v_1 = (1,2,2), v_2 = (-2,-1,2), v_3 = (2,-2,1)$. Then

$$T(v_1) = (3,6,6) = 3v_1$$
$$T(v_2) = (-12,-6,12) = 6v_2$$
$$T(v_3) = (18,-18,9) = 8v_3$$

Thus v_1 is an eigenvector for T corresponding to the eigenvalue $\alpha_1 = 3$; v_2 is an eigenvector for T corresponding to the eigenvalue $\alpha_2 = 6$, and v_3 is an eigenvalue for T corresponding to the eigenvalue $\alpha_3 = 9$.

(b) Let $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \in \mathbb{M}_2(\mathbb{R})$. If $\alpha \in \mathbb{R}$, $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \in \mathbb{R}^2$ and $L_A x = \alpha x$, then

$$\begin{bmatrix} \alpha x_1 \\ \alpha x_2 \end{bmatrix} = \alpha x = L_A x = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -x_2 \\ x_1 \end{bmatrix}.$$

Thus $\alpha x_1 = -x_2$ and $\alpha x_2 = x_1$.

If $x_2 \neq 0$, then $-x_2 = \alpha x_1 = \alpha(\alpha x_2) = \alpha^2 x_2$, a contradiction as $\alpha^2 = -1$ has no solution with $\alpha \in \mathbb{R}$. Thus $x_2 = 0$.

But then $x_1 = \alpha x_2 = \alpha 0 = 0$, so that $x = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ is not an eigenvector of A. In other words, A has no eigenvectors, and hence no eigenvalues either.

(c) Let $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \in \mathbb{M}_2(\mathbb{C})$. Let $x = \begin{bmatrix} 1 \\ -i \end{bmatrix}$. Then

$$L_A x = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ -i \end{bmatrix} = \begin{bmatrix} i \\ 1 \end{bmatrix} = i \begin{bmatrix} 1 \\ -i \end{bmatrix},$$

so that $x = \begin{bmatrix} 1 \\ -i \end{bmatrix}$ is an eigenvector for A corresponding to the eigenvalue $\alpha = i$.

We leave it to the reader to verify that $-i$ is also an eigenvalue for A, and to find an eigenvector for A corresponding to $-i$.

1.5. Theorem. *Let $n \in \mathbb{N}$, $\mathbb{F}$ be a field, and $A \in \mathbb{M}_n(\mathbb{F})$. Then $\alpha \in \mathbb{F}$ is an eigenvalue for A (that is, $\alpha \in \sigma_p(A)$ if and only if $\det(A - \alpha I_n) = 0$, or equivalently if and only if $(A - \alpha I_n)$ is not invertible.*

Proof. Consider the following sequence of statements, where each of the last five statements is equivalent to the one above it (and thus to all of the others). The equivalence that proves our Theorem is that of the first and fifth statements.

- $\alpha \in \sigma_p(L_A)$;
- there exists $0 \neq x \in \ker(L_A - \alpha I_{\mathbb{F}^n})$;
- $(L_A - \alpha I_{\mathbb{F}^n}) = L_{A-\alpha I_n}$ is not invertible;
- $(A - \alpha I_n)$ is not invertible;
- $\det(A - \alpha I_n) = 0$.

- $\operatorname{rank}(A - \alpha I_n) < n$.

□

1.6. Remark. Let $n \in \mathbb{N}$ and let $\mathbb{F}$ be a field. So far we have defined the determinant of an element $A = [a_{ij}] \in \mathbb{M}_n(\mathbb{F})$ recursively by setting $\det([a]) = a$ whenever $[a] \in \mathbb{M}_1(\mathbb{F})$ and then using the formula

$$\det(A) = \sum_{i=1}^{n} (-1)^{i+1} (a_{i1}) \det(\widehat{A}_{ij}),$$

where $\widehat{A}_{ij} \in \mathbb{M}_{n-1}(\mathbb{F})$ is the matrix obtained from A by deleting the i^{th} row and the first column.

Using elementary row and column operations, we were able to obtain a good deal of information about the determinant function, including the fact that

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

for all $A, B \in \mathbb{M}_n(\mathbb{F})$.

In what follows we wish to define $\det(A)$ when $A \in \mathbb{M}_n(\mathbb{F}[x])$. That is, suppose that $A = [a_{ij}]$, where each a_{ij} is a polynomial in x with coefficients in $\mathbb{F}$ for $1 \leq i, j \leq n$. Clearly we may set $\det([a_{11}]) := a_{11}$, and for $n \geq 1$, we can apply the same recursive formula as above, namely:

$$\det(A) = \sum_{i=1}^{n} (-1)^{i+1} (a_{i1}) \det(\widehat{A}_{ij}),$$

where $\widehat{A}_{ij} \in \mathbb{M}_{n-1}(\mathbb{F})$ is the matrix obtained from A by deleting the i^{th} row and the first column.

The question of whether or not $\det(AB) = \det(A) \cdot \det(B)$ when $A, B \in \mathbb{M}_n(\mathbb{F}[x])$ is less clear. For example, if $A = \begin{bmatrix} 3 + x^2 & 2 + x + x^3 \\ \pi + ex - 19\sqrt{2}x^5 & -17 + x^9 \end{bmatrix} \in \mathbb{M}_2(\mathbb{R}[x])$, is it obvious that we can perform elementary row operations on A to put it in reduced row echelon form? (This strategy was behind our proof of the multiplicativity of the determinant function.)

While not obvious, it happens to be true. Strangely enough, our inspiration comes from looking at matrices with integer entries. Suppose that $A, B \in \mathbb{M}_n(\mathbb{Z})$. Then $A, B \in \mathbb{M}_n(\mathbb{Q})$, and so $\det(AB) = \det(A) \cdot \det(B)$ when thinking of A and B as matrices with rational entries. But given $T \in \mathbb{M}_n(\mathbb{Z})$, the formula for $\det(T)$ is the same whether we view T as having integer entries or rational entries, implying that

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

when thinking of A and B as matrices with integer entries as well.

What does this have to do with matrices with entries in $\mathbb{F}[x]$? To answer this, we first ask ourselves "how were we able to find a field, in this case $\mathbb{Q}$, that contained $\mathbb{Z}$?"

We shall give a very abridged version of the construction, leaving the verification of the details to the interested reader.

The integers $\mathbb{Z}$ are an example of what is known as an **integral domain**. An integral domain is a non-empty set D with two operations $+$ and $\cdot$ (referred to as **addition** and **multiplication**) satisfying:

(D1) $x + y \in D$ for all $x, y \in D$;
(D2) $(x + y) + z = x + (y + z)$ for all $x, y, z \in D$;
(D3) there exists $0 \in D$ such that $x + 0 = x = 0 + x$;
(D4) for each $x \in D$ there exists an element $y \in D$ such that $x + y = 0 = y + x$;
(D5) $x + y = y + x$ for all $x, y \in D$.
(D6) $x \cdot y \in D$ for all $x, y \in D$;
(D7) $x \cdot (y \cdot z) = (x \cdot y) \cdot z$ for all $x, y, z \in D$;
(D8) $x \cdot (y + z) = x \cdot y + x \cdot z$ for all $x, y, z \in D$;
(D9) $(x + y) \cdot z = x \cdot z + y \cdot z$ for all $x, y, z \in D$;
(D10) there exists an element $1 \neq 0$ such that $1 \cdot x = x = x \cdot 1$ for all $x \in D$.
(D11) $x \cdot y = y \cdot x$ for all $x, y \in D$;
(D11) If $x, y \in D$ and $x \cdot y = 0$, then either $x = 0$ or $y = 0$.

The first five conditions should be reminiscent of the definition of a vector space. Indeed, the first five conditions define what is known as an **abelian group** under addition, and both vector spaces and integral domains are abelian groups under addition.

As a first exercise, we leave it to the reader to verify that $\mathbb{F}[x]$ is also an integral domain whenever $\mathbb{F}$ is a field.

Of course, we think of rational numbers as numbers of the form $\dfrac{a}{b}$, where $a, b \in \mathbb{Z}$ and $b \neq 0$. It is not hard to see that we could just as easily write this as an ordered pair $(a, b) \in \mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$. There is a slight complication, however, in that we know that (for example) $\dfrac{3}{15} = \dfrac{-9}{-45}$. This is verified by noting that $3(-45) = 15(-9)$. Given an integral domain D, we define an equivalence relation on $D \times (D \setminus \{0\})$ by setting $(a_1, b_1) \equiv (a_2, b_2)$ if $a_1 \cdot b_2 = a_2 \cdot b_1$.

We then define two operations $+$ and $\cdot$ on $\mathbb{E} := \{[(a, b)] : a, b \in D, b \neq 0\}$ as follows:

$$[(a, b)] + [(c, d)] = [(ad + bc, bd)],$$

and

$$[(a, b)] \cdot [(c, d)] := [(ac, bd)].$$

(We should think of $[(a, b)]$ as the equivalent of $\dfrac{a}{b}$, and our usual addition and multiplication of rational numbers then explains the two operations above.) Of course, since we are dealing with *equivalence classes* of ordered pairs, it behooves us to verify that these operations are well-defined! (In other words, is $\frac{3}{4} + \frac{1}{2} = \frac{6}{8} + \frac{-23}{-46}$?) The well-definedness of the two operations is the second exercise left to the reader.

The third exercise left to the reader is to show that $(\mathbb{E}, +, \cdot)$ is an integral domain, and that if $[(a,b)] \neq [(0,1)] \in \mathbb{E}$ (note that $[(0,1)]$ is the zero-element of $\mathbb{E}$ as per condition D3 above – think of it as $\frac{0}{1}$), then $[(a,b)]$ is invertible under multiplication with inverse $[(b,a)]$. (After all, you would expect the inverse of $\frac{a}{b}$ to be $\frac{b}{a}$!)

Together, these conditions imply that $\mathbb{E}$ is a field. In the same way that the map $\varphi : \mathbb{Z} \to \mathbb{Q}$ defined by $\varphi(n) = \frac{n}{1}$ for all $n \in \mathbb{Z}$ embeds a copy of $\mathbb{Z}$ into $\mathbb{Q}$, the map

$$\begin{array}{rccc} \varrho: & D & \to & \mathbb{E} \\ & d & \mapsto & [(d,1)] \end{array}$$

embeds a copy of D into $\mathbb{E}$. We may therefore conflate, or identify, D with the subset $\varrho(D) \subseteq \mathbb{E}$ and think of D as sitting "*inside*" its **field of quotients** $\mathbb{E}$. When doing this, we suppress the "ϱ" altogether and just write $D \subseteq \mathbb{E}$.

Finally, in our case, $\mathbb{E} := \{[(f(x), g(x)] : f(x), g(x) \in \mathbb{F}[x], g(x) \neq 0\}$ is a field which contains a copy of $\mathbb{F}[x]$, namely $\varrho(\mathbb{F}[x]) = \{[(f(x), 1] : f(x) \in \mathbb{F}[x]\}$. Again, we tend to think of $\mathbb{F}[x]$ as a subset of $\mathbb{E}$, and some people with a funny disposition might even write $\frac{f(x)}{g(x)}$ to mean $[f(x), g(x)] \in \mathbb{E}$.

Thus, when given $A = [a_{ij}] \in \mathbb{M}_n(\mathbb{F}[x])$, we may view A as an element of $\mathbb{M}_n(\mathbb{E})$, and all of the familiar properties of determinants hold because $\mathbb{E}$ is a field. In particular, if $A, B \in \mathbb{M}_n(\mathbb{F}) \subseteq \mathbb{M}_n(\mathbb{E})$, then

$$\det(AB) = \det(A) \cdot \det(B) \in \mathbb{E}.$$

We needed all of this to be able to define the characteristic polynomial of a matrix $A \in \mathbb{M}_n(\mathbb{F})$, which we now do.

1.7. Definition. *Let $n \in \mathbb{N}$ and $\mathbb{F}$ be a field. Given $A \in \mathbb{M}_n(\mathbb{F})$, the **characteristic polynomial** of T is the polynomial*

$$p_A(x) := \det(A - xI_n) \in \mathbb{F}[x].$$

1.8. Example. Let $A = \begin{bmatrix} -3 & 2 \\ 7 & -11 \end{bmatrix} \in \mathbb{M}_2(\mathbb{C})$. Then

$$\begin{aligned} p_A(x) &= \det(A - xI_n) \\ &= \det \begin{bmatrix} -3-x & 2 \\ 7 & -11-x \end{bmatrix} \\ &= (-3-x)(-11-x) - 2(7) \\ &= x^2 + 14x + 19. \end{aligned}$$

1.9. Proposition. *Let $n \in \mathbb{N}$, $\mathbb{F}$ be a field, and $A \in \mathbb{M}_n(\mathbb{F}_1[x])$. Then* $\det(A) \in \mathbb{F}_n[x]$.

Proof. We shall argue by induction on n.

If $n = 1$, then $A = [a_{11}]$ where $a_{11} = p_{11}+q_{11}x$. By definition, $\det(A) = a_{11} \in \mathbb{F}_1[x]$.

Now let $N \geq 2$ and suppose that the result holds for all matrices $A \in \mathbb{M}_n(\mathbb{F}_1[x])$, $1 \leq n < N$. Let $B = [b_{ij}] \in \mathbb{M}_N(\mathbb{F}_1[x]) \subseteq \mathbb{M}_N(\mathbb{F}[x])$. By definition,

$$\det(B) = \sum_{i=1}^{N}(-1)^{i+1}b_{ij}\det(\widehat{B}_{i1}).$$

Now $\widehat{B}_{i1} \in \mathbb{M}_{N-1}(\mathbb{F}_1[x])$, and so by our induction hypothesis,

$$\det(\widehat{B}_{i1}) \in \mathbb{F}_{N-1}[x], \quad 1 \leq i \leq N.$$

Since $b_{i1} \in \mathbb{F}_1[x]$, we then have that

$$b_{i1}\det(\widehat{B}_{i1}) \in \mathbb{F}_N[x], \quad 1 \leq i \leq N,$$

and thus

$$\det(B) \in \mathbb{F}_N[x],$$

completing the induction step.

□

1.10. Theorem. *Let $n \in \mathbb{N}$, $\mathbb{F}$ be a field and $A = [a_{ij}] \in \mathbb{F}$. Then $p_A(x)$ is a polynomial of degree n in x.*

Proof. Since $A - xI_n \in \mathbb{M}_n(\mathbb{F}_1[x])$, from the above Proposition we find that $p_A(x) = \det(A - xI_n) \in \mathbb{F}_n[x]$. To see that $\deg(p_A(x)) = n$, we once again argue by induction on n.

If $n = 1$, then $A = [a_{11}]$ and so $\det(A - xI_1) = \det([a_{11} - x]) = a_{11} - x \in \mathbb{F}_1[x]$, as required.

Suppose that $N \geq 2$ and that the result holds for all $1 \leq n < N$. Let $A = [a_{ij}] \in \mathbb{M}_N(\mathbb{F})$. Let $B := A - xI_N = [b_{ij}]$, where $b_{ij} \in \mathbb{F}_1[x]$ for all i, j. Then

$$\det(A - xI_N) = \sum_{i=1}^{N}(-1)^{i+1}b_{i1}\det(\widehat{B}_{i1}).$$

For $2 \leq i \leq N$, we have that $\widehat{B}_{i1} \in \mathbb{M}_{N-1}(\mathbb{F}_1[x])$, and so from Proposition 1.9, $\det(\widehat{B}_{i1}) \in \mathbb{F}_{N-1}[x]$. Since $b_{i1} \in \mathbb{F}$, $2 \leq i \leq N$, we see that

$$\sum_{i=2}^{N}(-1)^{i+1}b_{i1}\det(\widehat{B}_{i1}) \in \mathbb{F}_{N-1}[x].$$

On the other hand, $b_{11} = a_{11} - x$ has degree 1, and by our induction hypothesis, $\det(\widehat{B}_{11})$ has degree $N - 1$, as $\widehat{B}_{11} = \widehat{A}_{11} - xI_{N-1}$. Thus

$$(-1)^{1+1}b_{11}\det(\widehat{B}_{11})$$

has degree N, and therefore

$$\det(A - xI_N) = \det(B) = \sum_{i=1}^{N} (-1)^{i+1} b_{i1} \det(\widehat{B}_{i1})$$

has degree N. This completes the induction step, and hence the proof.

□

1.11. Corollary. *Let $n \in \mathbb{N}$, $\mathbb{F}$ be a field, and $A \in \mathbb{M}_n(\mathbb{F})$. Then A has at most n distinct eigenvalues.*

Proof. The eigenvalues of A are the roots of $p_A(x)$ which lie in $\mathbb{F}$. Since $p_A(x)$ has degree n, it can have at most n roots in $\mathbb{F}$.

□

1.12. Example. Let $n \in \mathbb{N}$, $\mathbb{F}$ be a field, and $A \in \mathbb{M}_n(\mathbb{F})$. Suppose that

$$p_A(x) = x^4 + x^3 + x^2 + x + 1.$$

(a) If $\mathbb{F} = \mathbb{Q}$, then $p_A(x)$ does not factor. Thus A has no eigenvalues.

(b) If $\mathbb{F} = \mathbb{R}$, then

$$p_A(x) = \left(x^2 + \frac{\sqrt{5}+1}{2}x + 1\right) \cdot \left(x^2 - \frac{\sqrt{5}+1}{2}x + 1\right).$$

Again, A has no eigenvalues.

(c) If $\mathbb{F} = \mathbb{C}$, then

$$p_A(x) = (x - \alpha_1)(x - \alpha_2)(x - \alpha_3)(x - \alpha_4),$$

where $\alpha_k = \cos(2k\pi/5) + i\sin(2k\pi/5) \in \mathbb{C}$, $1 \le k \le 4$. Thus A has four eigenvalues.

(d) If $\mathbb{F} = \mathbb{Z}_5$, then

$$p_A(x) = (x-1)^4 = (x-1)(x-1)(x-1)(x-1).$$

Here A has one eigenvalue, namely 1. We shall soon define the multiplicity of this eigenvalue to be 4.

1.13. Proposition. *Let $n \in \mathbb{N}$, $\mathbb{F}$ be a field, and $A, B \in \mathbb{M}_n(\mathbb{F})$. If A and B are similar, then $p_A(x) = p_B(x)$.*

Proof. Choose $R \in \mathbb{M}_n(\mathbb{F})$ invertible such that $B = R^{-1}AR$. Then $B - xI_n = R^{-1}AR - xI_n = R^{-1}(A - xI_n)R$. Thus

$$p_B(x) = \det(B - xI_n) = \det(R^{-1}(A - xI_n)R) = (\det R^{-1})\ (p_A(x))\ (\det R) = p_A(x).$$

□

1.14. Corollary. *Suppose that $\mathcal{V}$ is an n-dimensional vector space over $\mathbb{F}$, $T \in \mathcal{L}(\mathcal{V})$, and that $\mathcal{D}$ and $\mathcal{C}$ are two ordered bases for $\mathcal{V}$. Then*

$$\det([T]_{\mathcal{D}} - xI_n) = \det([T]_{\mathcal{C}} - xI_n).$$

Proof. This is an immediate consequence of the above Proposition combined with the fact that $[T]_{\mathcal{D}}$ is similar to $[T]_{\mathcal{C}}$ via the matrix $[I_{\mathcal{V}}]_{\mathcal{D}}^{\mathcal{C}}$.

□

Because the characteristic polynomial for the matrix for T doesn't depend upon the ordered basis that we choose to represent T, the following definition makes sense.

1.15. Definition. *Let $n \in \mathbb{N}$ and $\mathcal{V}$ be an n-dimensional vector space over a field $\mathbb{F}$. Suppose that $T \in \mathcal{L}(\mathcal{V})$. Then the* ***characteristic polynomial*** *of T is the degree-n polynomial*

$$p_T(x) = \det([T]_{\mathcal{D}} - xI_n),$$

where $\mathcal{D}$ is any ordered basis for $\mathcal{V}$.

1.16. Example. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the map given by $T(x, y) = (3x + 4y, 9x - 7y)$. Let $\mathcal{D} = ((1, 0), (0, 1))$ denote the standard ordered basis for $\mathbb{R}^2$. Then

$$[T]_{\mathcal{D}} = [[T(1,0)]_{\mathcal{D}}\ [T(0,1)]_{\mathcal{D}}] = [[(3,9)]_{\mathcal{D}}\ [(4,-7)]_{\mathcal{D}}] = \begin{bmatrix} 3 & 4 \\ 9 & -7 \end{bmatrix}.$$

Thus

$$\begin{aligned} p_T(x) &= \det([T]_{\mathcal{D}} - xI_2) \\ &= \det \begin{bmatrix} 3-x & 4 \\ 9 & -7-x \end{bmatrix} \\ &= (3-x)(-7-x) - (9)(4) \\ &= x^2 + 4x - 57. \end{aligned}$$

2. Multiplicities of eigenvalues and diagonalisability

2.1. If $\mathcal{V}$ is a vector space over a field $\mathbb{F}$ and $T \in \mathcal{L}(\mathcal{V})$, we may define the **spectrum** of T to be the set

$$\sigma(T) = \{\alpha \in \mathbb{F} : (T - \alpha I_{\mathcal{V}}) \text{ is not invertible.}\}$$

As we saw in Theorem 1.5, when $\dim \mathcal{V} < \infty$, $\sigma(T) = \sigma_p(T)$.

When $\dim \mathcal{V} = \infty$, this need not be the case. For example, let $\mathcal{V} = \mathcal{C}([0, 1], \mathbb{R})$ be a vector space over $\mathbb{R}$, and define the linear map

$$\begin{array}{rccc} M_x : & \mathcal{C}([0,1],\mathbb{R}) & \to & \mathcal{C}([0,1],\mathbb{R}) \\ & f(x) & \mapsto & xf(x). \end{array}$$

If $f \in \ker M_x$, then $[M_x f](x) = xf(x) = 0$ for all $x \in [0,1]$, and thus $f(x) = 0$ for all $x \in (0,1]$. But f is continuous on $[0,1]$, and therefore $f(0) = \lim_{x\to 0^+} f(x) = 0$. That is, $f \equiv 0$. Thus $0 \notin \sigma_p(M_x)$.

On the other hand, $M_x = M_x - 0I_{\mathcal{V}}$ is not invertible, since $\operatorname{ran} M_x \neq \mathcal{C}([0,1],\mathbb{R})$. (If $g \in \operatorname{ran} M_x$ – say $g = M_x f$ for some $f \in \mathcal{C}([0,1],\mathbb{R})$, then $g(0) = [M_x f](0) = 0f(0) = 0$.) Thus $0 \in \sigma(M_x)$.

A great deal of information about a linear map can be gleaned from its spectrum and its point spectrum (and there exist many, many other types of spectrum as well). In this introductory course, we shall limit ourselves to finite-dimensional vector spaces where $\sigma(T) = \sigma_p(T)$ is better behaved.

2.2. Definition. *Let $n \in \mathbb{N}$, and $\mathcal{V}$ be an n-dimensional vector space over a field $\mathbb{F}$. Suppose that $\alpha \in \mathbb{F}$ is an eigenvalue for $T \in \mathcal{L}(\mathcal{V})$. We define the* ***geometric multiplicity*** *of α to be*

$$\gamma_T(\alpha) := \dim \ker (T - \alpha I).$$

Since $\alpha \in \sigma_p(T)$, we know that $(z-\alpha)$ divides $p_T(z)$, and this allows us to define the ***algebraic multiplicity*** *of α to be*

$$\mu_T(\alpha) := \max\{k \geq 1 : (z-\alpha)^k \text{ divides } p_T(z).\}$$

As always, if $A \in \mathbb{M}_n(\mathbb{F})$, then we define the geometric (resp. algebraic) multiplicity of $\alpha \in \sigma_p(A)$ as the geometric (resp. algebraic) multiplicity of $L_A \in \mathcal{L}(\mathbb{F}^n)$.

2.3. Example.

(a) Let $n \geq 1$, and let $\mathcal{D} = (e_1, e_2, \ldots, e_n)$ be the standard ordered basis for $\mathbb{F}^n$. Let $J_n \in \mathbb{M}_n(\mathbb{F})$ be the $n \times n$-**Jordan cell**, defined by setting $J_n := [a_{ij}^{(n)}]$, where $a_{ij}^{(n)} = 0$ if $j \neq i+1$, and $a_{i\,i+1}^{(n)} = 1$, $1 \leq i \leq n-1$. Thus

$$J_n = \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ \vdots & & \ddots & \ddots & & \vdots \\ \vdots & & & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & 0 & \cdots & 0 & 0 \end{bmatrix}.$$

A routine inductive argument shows that $p_{J_n}(x) = (0-x)^n = (-1)^n x^n$, so that 0 is an eigenvalue of algebraic multiplicity n for J_n. That is, $\mu(0) = n$.

Note, however, that $\operatorname{nul} J_n = \dim \ker J_n = 1$, as $\ker J_n = \mathbb{F}e_1$. Thus

$$\gamma_{J_n}(0) = 1 \leq n = \mu_{J_n}(0),$$

with equality holding if and only if $n = 1$ in which case $J_1 = [0] \in \mathbb{M}_1(\mathbb{F})$.

If you search the literature, you should not be surprised to see J_n refer to both the linear map L_{J_n} and the matrix $J_n = [L_{J_n}]_{\mathcal{D}}$, where $\mathcal{D}$ is the standard ordered basis for $\mathbb{F}^n$. In fact, this holds general $A \in \mathbb{M}_n(\mathbb{F})$. We

have been careful to distinguish between the two concepts as this is the first time you are seeing the material. Most authors simply conflate $[L_A]_{\mathcal{D}}$ and $A \in \mathbb{M}_n(\mathbb{F})$. Just saying.

(b) Given $m, n \in \mathbb{N}$ and matrices $A \in \mathbb{M}_m(\mathbb{F})$ and $B \in \mathbb{M}_n(\mathbb{F})$, we define the **direct sum** of A and B to be the matrix

$$A \oplus B = \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix} \in \mathbb{M}_{m+n}(\mathbb{F}).$$

For example, if $A = -4I_3 + J_3$ and $B = 7I_2 + J_2$, then with $T := A \oplus B$,

$$T := A \oplus B = \begin{bmatrix} -4 & 1 & 0 & 0 & 0 \\ 0 & -4 & 1 & 0 & 0 \\ 0 & 0 & -4 & 0 & 0 \\ 0 & 0 & 0 & 7 & 1 \\ 0 & 0 & 0 & 0 & 7 \end{bmatrix}.$$

Again, because this is upper triangular, as is $T - xI_5$, it is easy to calculate $p_T(x)$:

$$p_T(x) = \det(T - xI_5) = (-4 - x)^3(7 - x)^2.$$

Thus $\sigma_p(T) = \{-4, 7\}$, $\mu_T(-4) = 3$ and $\mu_T(7) = 2$, while $\gamma_T(-4) = 1 = \gamma_T(7)$.

(c) Let $T = \begin{bmatrix} 19 & -9 & -6 \\ 25 & -11 & -9 \\ 17 & -9 & -4 \end{bmatrix} \in \mathbb{M}_3(\mathbb{R})$. Then

$$p_T(x) = \det(T - xI_3) = -x^3 + 4x^2 - 5x + 2.$$

By inspection, we see that $\alpha = 1$ is a root of this polynomial, so that

$$p_T(x) = (1 - x)(x^2 - 3x + 2) = (1 - x)(x - 1)(x - 2) = -(x - 1)^2(x - 2).$$

It follows that $\sigma_p(T) = \{1, 2\}$, that $\mu_T(1) = 2$ and $\mu_T(2) = 1$.

As for the geometric multiplicities of these eigenvalues, note that (by solving homogeneous systems of three equations in three variables)

$$\ker(T - 1I_3) = \ker \begin{bmatrix} 18 & -9 & -6 \\ 25 & -12 & -9 \\ 17 & -9 & -5 \end{bmatrix} = \operatorname{span}\left\{ \begin{bmatrix} 3 \\ 4 \\ 3 \end{bmatrix} \right\},$$

while

$$\ker(T - 2I_3) = \ker \begin{bmatrix} 17 & -9 & -6 \\ 25 & -13 & -9 \\ 17 & -9 & -6 \end{bmatrix} = \operatorname{span}\left\{ \begin{bmatrix} 3 \\ 3 \\ 4 \end{bmatrix} \right\}.$$

Thus $\gamma_T(1) = 1 = \gamma_T(2)$.

Observe that in each of the above examples, whenever $T \in \mathbb{M}_n(\mathbb{F})$ and $\alpha \in \sigma_p(T)$, we have that $\gamma_T(\alpha) \leq \mu_T(\alpha)$. This is not a coincidence.

2.4. Proposition. *Let $n \in \mathbb{N}$, $\mathcal{V}$ be an n-dimensional vector space over a field $\mathbb{F}$, and $T \in \mathcal{L}(\mathcal{V})$. If $\alpha \in \mathbb{F}$ is an eigenvalue of T, then*

$$\gamma_T(\alpha) \leq \mu_T(\alpha).$$

That is, the geometric multiplicity of α is at most equal to the algebraic multiplicity of α as an eigenvalue of T.

Proof. Let $\mathcal{M} = \ker(T - \alpha I_n)$, and define $d := \gamma_T(\alpha) = \dim \mathcal{M}$. Let $(v_1, v_2, \ldots, v_d)$ be a basis for $\mathcal{M}$, and extend this to an ordered basis

$$\mathcal{D} = (v_1, v_2, \ldots, v_d, x_1, x_2, \ldots, x_{n-d})$$

for $\mathcal{V}$. Since $Tv_j = \alpha v_j$, $1 \leq j \leq d$, it follows that

$$[T]_{\mathcal{D}} = \begin{bmatrix} \alpha I_d & B \\ 0 & D \end{bmatrix}$$

for some $B \in \mathbb{M}_{d\times(n-d)}(\mathbb{F})$ and $D \in \mathbb{M}_{n-d}(\mathbb{F})$. By Exercise 8.3,

$$p_T(x) = p_{\alpha I_d}(x)\; p_D(x) = (\alpha - x)^d\; p_D(x).$$

Thus $(x - \alpha)^d$ divides $p_T(x)$, from which we deduce that $\mu_T(\alpha) \geq d = \gamma_T(\alpha)$, as required.

□

2.5. Theorem. *Let $\mathcal{V}$ be a vector space over a field $\mathbb{F}$ and $T \in \mathcal{L}(\mathcal{V})$. Suppose that $\alpha \neq \beta$ are distinct eigenvalues of T. If x (resp. y) is an eigenvector for T corresponding to α (resp. corresponding to β), then x and y are linearly independent.*

Proof. Suppose that $\kappa_1, \kappa_2 \in \mathbb{F}$ and that $\kappa_1 x + \kappa_2 y = \mathbf{0}$. Then

$$\begin{aligned} \mathbf{0} &= (T - \alpha I_{\mathcal{V}})\mathbf{0} \\ &= (T - \alpha I_{\mathcal{V}})(\kappa_1 x + \kappa_2 y) \\ &= \kappa_1(T - \alpha I_{\mathcal{V}})x + \kappa_2(T - \alpha I_{\mathcal{V}})y \\ &= \kappa_1\mathbf{0} + \kappa_2(\beta - \alpha)y \\ &= \kappa_2(\beta - \alpha)y. \end{aligned}$$

Since y is an eigenvector for T corresponding to β, we know that $y \neq \mathbf{0}$, and since $\beta \neq \alpha$, this forces $\kappa_2 = 0$. But then $\kappa_1 x = \mathbf{0}$ and $x \neq \mathbf{0}$, whence $\kappa_1 = 0$ and so x and y are linearly independent.

□

2.6. Definition. *Let $n \in \mathbb{N}$ and $\mathbb{F}$ be a field. We say that $D \in \mathbb{M}_n(\mathbb{F})$ is a* ***diagonal matrix*** *if*

$$D = \begin{bmatrix} d_1 & 0 & 0 & \cdots & 0 \\ 0 & d_2 & 0 & \cdots & 0 \\ 0 & 0 & \ddots & \cdots & 0 \\ \vdots & & & & \vdots \\ 0 & 0 & 0 & \cdots & d_n \end{bmatrix}$$

for some choice of $d_1, d_2, \ldots, d_n \in \mathbb{F}$. We also denote this by $D = \text{DIAG}(d_1, d_2, \ldots, d_n)$.

*We say that $B \in \mathbb{M}_n(\mathbb{F})$ is **diagonalisable** if B is similar to a diagonal matrix.*

*Finally, if $\mathcal{V}$ is an n-dimensional vector space over $\mathbb{F}$ and $T \in \mathcal{L}(\mathcal{V})$, then we say that T is **diagonalisable** if there exists an ordered basis $\mathcal{D}$ for $\mathcal{V}$ relative to which $[T]_{\mathcal{D}}$ is a diagonal matrix.*

2.7. Remark. Note that if $D = \text{DIAG}(d_1, d_2, \ldots, d_n) \in \mathbb{M}_n(\mathbb{F})$, then each of the standard basis vectors e_j is an eigenvector for D corresponding to the eigenvalue d_j, $1 \leq j \leq n$.

Suppose now that $T \in \mathbb{M}_n(\mathbb{F})$ is diagonalisable, and choose $R \in \mathbb{M}_n(\mathbb{F})$ invertible such that $D := R^{-1}TR$ is diagonal, say $D = \text{DIAG}(\alpha_1, \alpha_2, \ldots, \alpha_n)$. Set $b_j := c_j(R)$, $1 \leq j \leq n$, and observe that the fact that R is invertible implies that

$$\mathcal{B} := (b_1, b_2, \ldots, b_n)$$

is an ordered basis for $\mathbb{F}^n$. Moreover, $R = [I_{\mathbb{F}^n}]_{\mathcal{B}}^{\mathcal{D}}$, and so $R^{-1} = [I_{\mathbb{F}^n}]_{\mathcal{D}}^{\mathcal{B}}$. From this we find that

$$D = R^{-1}TR = [I_{\mathbb{F}^n}]_{\mathcal{D}}^{\mathcal{B}}[L_T]_{\mathcal{D}}[I_{\mathbb{F}^n}]_{\mathcal{B}}^{\mathcal{D}} = [L_T]_{\mathcal{B}}.$$

Using the observation at the start of this Remark, we see that

$$[L_T b_j]_{\mathcal{B}} = [L_T]_{\mathcal{B}}[b_j]_{\mathcal{B}} = De_j = \alpha_j e_j = [\alpha_j b_j]_{\mathcal{B}},$$

whence $Tb_j = L_T b_j = \alpha_j b_j$, $1 \leq j \leq n$.

In other words, the j^{th} column of R is precisely an eigenvector of T corresponding to the eigenvalue α_j, $1 \leq j \leq n$. We shall return to this in Proposition 2.9 below.

2.8. Example. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the map $T(x, y) = (-14x + 12y, -20x + 17y)$. Relative to the standard ordered basis $\mathcal{D} = (e_1, e_2)$ for $\mathbb{R}$, we see that

$$[T]_{\mathcal{D}} = \begin{bmatrix} -14 & 12 \\ -20 & 17 \end{bmatrix}.$$

We leave it to the reader to verify that $p_T(x) = (x-2)(x-1)$, and thus $\sigma_p(T) = \{1, 2\}$. Moreover, $\ker(T - 1I_2) = \text{span}\left\{\begin{bmatrix} 4 \\ 5 \end{bmatrix}\right\}$, while $\ker(T - 2I_2) = \text{span}\left\{\begin{bmatrix} 3 \\ 4 \end{bmatrix}\right\}$.

Let $\mathcal{C} := ((4, 5), (3, 4))$, and note that this is an ordered basis for $\mathbb{R}^2$ since eigenvectors corresponding to distinct eigenvalues are automatically linearly independent.

If we define $R := [I_{\mathbb{R}^2}]_{\mathcal{D}}^{\mathcal{C}}$, then $R^{-1} = [I_{\mathbb{R}^2}]_{\mathcal{C}}^{\mathcal{D}} = \begin{bmatrix} 4 & 3 \\ 5 & 4 \end{bmatrix}$. A routine calculation (using elementary operations) shows that $R = \begin{bmatrix} 4 & -3 \\ -5 & 4 \end{bmatrix}$.

Finally,

$$[T]_{\mathcal{C}} = [I_{\mathbb{R}^2}]_{\mathcal{C}}^{\mathcal{D}}[T]_{\mathcal{D}}[I_{\mathbb{R}^2}]_{\mathcal{C}}^{\mathcal{D}} = \begin{bmatrix} 4 & -3 \\ -5 & 4 \end{bmatrix}\begin{bmatrix} -14 & 12 \\ -20 & 17 \end{bmatrix}\begin{bmatrix} 4 & 3 \\ 5 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}.$$

2.9. Proposition. *Let $n \in \mathbb{N}$, $\mathbb{F}$ be a field and $\mathcal{V}$ be an n-dimensional vector space over $\mathbb{F}$. A linear map $T \in \mathcal{L}(\mathcal{V})$ is diagonalisable if and only if there exists an ordered basis for $\mathcal{V}$ consisting of eigenvectors of T.*

Proof. Suppose that T is diagonalisable, and let $\mathcal{D} = (v_1, v_2, \ldots, v_n)$ be an ordered basis for $\mathcal{V}$ for which $[T]_{\mathcal{D}}$ is diagonal, say $[T]_{\mathcal{D}} = \text{DIAG}(d_1, d_2, \ldots, d_n)$. If $\mathcal{B} := (e_1, e_2, \ldots, e_n)$ is the standard ordered basis for $\mathbb{F}^n$, then

$$[Tv_j]_{\mathcal{D}} = [T]_{\mathcal{D}}[v_j]_{\mathcal{D}} = [T]_{\mathcal{D}}e_j = d_j e_j = [d_j v_j]_{\mathcal{D}},$$

implying that $Tv_j = d_j v_j$, $1 \leq j \leq n$. Thus $\mathcal{D}$ consists of eigenvectors for T.

Conversely, if $\mathcal{B} := (b_1, b_2, \ldots, b_n)$ is an ordered basis for $\mathcal{V}$ consisting of eigenvectors for T – say $Tb_j = \beta_j b_j$, $1 \leq j \leq n$ – then

$$[T]_{\mathcal{B}} = \text{DIAG}(\beta_1, \beta_2, \ldots, \beta_n),$$

and so T is diagonalisable.

□

As a corollary to this result, we obtain:

2.10. Theorem. *Let $n \in \mathbb{N}$, $\mathbb{F}$ be a field and $\mathcal{V}$ be an n-dimensional vector space over $\mathbb{F}$. The following conditions are equivalent:*

(a) *The map $T \in \mathcal{L}(\mathcal{V})$ is diagonalisable.*
(b) *The characteristic polynomial $p_T(z)$ splits over $\mathbb{F}$ and $\gamma_T(\alpha) = \mu_T(\alpha)$ for all $\alpha \in \sigma_p(T)$.*

Proof.

(a) implies (b). Suppose that the map $T \in \mathcal{L}(\mathcal{V})$ is diagonalisable. Let $\mathcal{D}$ be an ordered basis for $\mathcal{V}$ relative to which $[T]_{\mathcal{D}} = \text{DIAG}(\alpha_1, \alpha_2, \ldots, \alpha_n)$.

Clearly $p_T(x) = \det(T - xI_n) = (\alpha_1 - x)(\alpha_2 - x)\cdots(\alpha_n - x)$, and so $p_T(x)$ splits over $\mathbb{F}$.

Also, for $1 \leq j \leq n$,

$$\begin{aligned}\ker([T]_{\mathcal{D}} - \alpha_j I_n) &= \ker \text{DIAG}(\alpha_1 - \alpha_j, \alpha_2 - \alpha_j, \ldots, \alpha_n - \alpha_j) \\ &= \text{span}\{e_i : e_i = e_j\}.\end{aligned}$$

Thus $\gamma_T(\alpha_j) = |\{e_i : e_i = e_j\}| = \mu_T(\alpha_j)$.

(b) implies (a). Conversely, suppose that the characteristic polynomial $p_T(z)$ splits over $\mathbb{F}$ and $\gamma_T(\alpha) = \mu_T(\alpha)$ for all $\alpha \in \sigma_p(T)$. Let

$$p_T(x) = (\alpha_1 - x)^{\mu_T(\alpha_1)}(\alpha_2 - x)^{\mu_T(\alpha_2)}\cdots(\alpha_r - x)^{\mu_T(\alpha_r)},$$

where $\sum_{i=1}^{r} \mu_T(\alpha_i) = n$. Since $\gamma_T(\alpha_i) = \mu_T(\alpha_i)$ for each $1 \leq i \leq r$, we can find an ordered basis $\mathcal{B}_i = (b_{i1}, b_{i2}, \ldots, b_{i\mu_T(\alpha_i)})$ for $\ker(T - \alpha_i I_n)$.

By Assignment 7, $\mathcal{B} = (\mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_r)$ is a linearly independent set containing $\sum_{i=1}^{r} \mu_T(\alpha_i) = n$ elements, it must be a basis for $\mathcal{V}$. By Proposition 2.9, T is diagonalisable.

□

2.11. Example. Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be the linear map $T(x, y, z) = (8x + 3y - 4z, -3x + y + 3z, 4x + 3y)$. Relative to the standard basis $\mathcal{D} = (e_1, e_2, e_3)$ for $\mathbb{R}^3$, we have

$$[T]_{\mathcal{D}} = \begin{bmatrix} 8 & 3 & -4 \\ -3 & 1 & 3 \\ 4 & 3 & 0 \end{bmatrix}.$$

Thus

$$p_T(x) = \det(T - xI_3) = -(x-4)^2(x-1).$$

Thus $\sigma_p(T) = \{1, 4\}$.

Now $\ker(T - 4I_3) = \operatorname{span}\{\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}\}$, and so $\gamma_T(4) = 1 < 2 = \mu_T(4)$. By Theorem 2.10, T is not diagonalisable.

2.12. Theorem. *Let $n \in \mathbb{N}$, $\mathbb{F}$ be a field and $\mathcal{V}$ be an n-dimensional vector space over $\mathbb{F}$. If $T \in \mathcal{L}(\mathcal{V})$ has n distinct eigenvalues in $\mathbb{F}$, then T is diagonalisable.*

Proof. Suppose that T has n distinct eigenvalues, say $\sigma_p(T) = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ with $\alpha_i \in \mathbb{F}$ for all i and $\alpha_i \neq \alpha_j$ if $i \neq j$. Now

$$p_T(x) = (\alpha_1 - x)(\alpha_2 - x)\cdots(\alpha_n - x),$$

and $\mu_T(\alpha_j) = 1$ for all $1 \leq j \leq n$, since all α_i's are distinct. Since $\gamma_T(\alpha) \geq 1$ when $\alpha \in \sigma_p(T)$, and since $\gamma_T(\alpha) \leq \mu_T(\alpha)$ for all $\alpha \in \sigma_p(T)$, we conclude that $\gamma_T(\alpha_i) = 1 = \mu_T(\alpha_i)$, $1 \leq i \leq n$. By Theorem 2.10, T is diagonalisable.

□

2.13. Example.

(a) The matrix $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ is not diagonalisable over $\mathbb{R}$, as it has no real eigenvalues. On the other hand, over $\mathbb{F} = \mathbb{C}$, it has two eigenvalues i and $-i$, and as such it is diagonalisable and with respect to the ordered basis $\mathcal{D} = ((1, -i), (1, i))$ consisting of the eigenvectors of A, we see that

$$[A]_{\mathcal{D}} = \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix}.$$

(b) If $T \in \mathbb{M}_4(\mathbb{C})$ and $\sigma_p(T) = \{1, 3 + i, 3 - i, 14 + 5i\}$, then we conclude that there exists a basis $\mathcal{D}$ for $\mathbb{C}^4$ relative to which

$$T = \text{DIAG}(1, 3 + i, 3 - i, 14 + 5i).$$

2.14. We illustrate the power of diagonalisability through the following example. Let $n \in \mathbb{N}$, $\mathbb{F}$ be a field and $\mathcal{V}$ be an n-dimensional vector space over $\mathbb{F}$. Suppose that $T \in \mathcal{L}(\mathcal{V})$ and that $q(z) = 24z - 14z^{147}$. Let us find $q(T)$.

If we can find an ordered basis $\mathcal{D}$ for $\mathcal{V}$ relative to which

$$[T]_{\mathcal{D}} = \text{DIAG}(d_1, d_2, \ldots, d_n),$$

then an easy induction argument shows that for all $k \geq 1$,

$$[T^k]_{\mathcal{D}} = ([T]_{\mathcal{D}})^k = (\text{DIAG}(d_1, d_2, \ldots, d_n))^k = \text{DIAG}(d_1^k, d_2^k, \ldots, d_n^k).$$

From this we see that

$$[q(T)]_{\mathcal{D}} = \text{DIAG}(q(d_1), q(d_2), \ldots, q(d_n)),$$

and once we know $[q(T)]_{\mathcal{D}}$, we know $q(T)$.

Let us finish the course by illustrating this with a concrete example.

2.15. Example. Let $T = \begin{bmatrix} -1 & 7 & -1 \\ 0 & 1 & 0 \\ 0 & 15 & -2 \end{bmatrix} \in \mathbb{M}_3(\mathbb{R})$. Then a routine calculation shows that

$$p_T(x) = -(x-1)(x+1)(x+2).$$

Thus $\sigma_p(T) = \{1, -1, -2\}$ has three distinct elements. Since $T \in \mathbb{M}_3(\mathbb{R})$, we conclude that T is diagonalisable.

By solving systems of linear equations, we find that

- $\ker(T - 1I_3) = \text{span}\{\begin{bmatrix} 1 \\ 1 \\ 5 \end{bmatrix}\}$.
- $\ker(T - (-1)I_3) = \text{span}\{\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}\}$.
- $\ker(T - (-2)I_3) = \text{span}\{\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}\}$.

Letting $\mathcal{D} = (e_1, e_2, e_3)$ be the standard ordered basis for $\mathbb{R}^3$ and $\mathcal{C} = ((1,1,5), (1,0,0), (1,0,1))$, we let

$$R := [I_{\mathbb{R}^3}]_{\mathcal{C}}^{\mathcal{D}} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 5 & 0 & 1 \end{bmatrix},$$

and we calculate

$$R^{-1} = [I_{\mathbb{R}^3}]_{\mathcal{D}}^{\mathcal{C}} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 4 & -1 \\ 0 & -5 & 1 \end{bmatrix}.$$

Thus

$$Y := [T]_{\mathcal{C}} = R^{-1}[L_T]_{\mathcal{D}}R = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 4 & -1 \\ 0 & -5 & 1 \end{bmatrix} \begin{bmatrix} -1 & 7 & -1 \\ 0 & 1 & 0 \\ 0 & 15 & -2 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 5 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 2 \end{bmatrix}.$$

Suppose that $q(x) = 13 + 24x^3 - ex^{417}$. Then $q(T) = Rq(Y)R^{-1}$.
But

$$\begin{aligned} q(Y) &= 13I_3 + 24Y^3 - eY^{417} \\ &= \begin{bmatrix} q(1) & 0 & 0 \\ 0 & q(-1) & 0 \\ 0 & 0 & q(2) \end{bmatrix} \\ &= \begin{bmatrix} 37 - e & 0 & 0 \\ 0 & -11 + e & 0 \\ 0 & 0 & 61 - 2^{417}e \end{bmatrix} \end{aligned}$$

Hence

$$\begin{aligned} q(T) = Rq(Y)R^{-1} &= \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 5 & 0 & 1 \end{bmatrix} \begin{bmatrix} 37 - e & 0 & 0 \\ 0 & -11 + e & 0 \\ 0 & 0 & 61 - 2^{417}e \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 1 & 4 & -1 \\ 0 & -5 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 37 - e & -11 + e & 61 - 2^{417}e \\ 37 - e & 0 & 0 \\ 185 - 5e & 0 & 61 - 2^{417}e \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 1 & 4 & -1 \\ 0 & -5 & 1 \end{bmatrix} \\ &= \begin{bmatrix} -11 + e & -312 + (e + 5 \cdot 2^{417})e & 72 - (1 + 2^{417})e \\ 0 & 37 - e & 0 \\ 0 & -120 + (5 \cdot 2^{417} - 5)e & 61 - 2^{417}e \end{bmatrix} \end{aligned}$$

Supplementary Examples

S8.1. Example. Let $\mathcal{W} = \mathcal{C}([0,1],\mathbb{R})$, and consider the map $V \in \mathcal{L}(\mathcal{W})$ defined by:

$$[Vf](x) := \int_0^x f(t)dt, \qquad x \in [0,1].$$

This map is referred to as the **Volterra operator** on $\mathcal{C}([0,1],\mathbb{R})$.

Observe that $[Vf](0) = 0$ for all $f \in \mathcal{C}([0,1],\mathbb{R})$, and thus there does not exist $f \in \mathcal{W}$ such that $Vf = \eta$, where $\eta : [0,1] \to \mathbb{R}$ is the constant function $\eta(x) = 1$ for all $x \in [0,1]$. In particular, V is not surjective, and so

$$0 \in \sigma(V) := \{\alpha \in \mathbb{R} : (V - \alpha I_{\mathcal{W}}) \text{ is not invertible}\}.$$

Nevertheless, $0 \notin \sigma_p(V)$. Indeed, suppose that $0 \neq f \in \mathcal{W}$ and that $Vf = 0$. Thus $[Vf](x) = 0$ for all $x \in [0,1]$. Since $f \neq 0$ and f is continuous, there exists $x_0 \in (0,1)$ such that $f(x_0) \neq 0$. By replacing f by $-f$ if necessary, we may assume that $f(x_0) > 0$. (Observe that if $Vf = 0$, then $V(-f) = -(Vf) = 0$, so this really is "without loss of generality".)

Let $\varepsilon := f(x_0)/2 > 0$. By continuity of f at x_0, there exists $\delta > 0$ such that $|x - x_0| < \delta$ implies that $|f(x) - f(x_0)| < \varepsilon$, and thus $f(x) > f(x_0) - \varepsilon = \varepsilon$. Hence

$$\begin{aligned}
[Vf](x_0) &= \int_0^{x_0} f(t)dt \\
&= \int_0^{x_0-\delta} f(t)dt + \int_{x_0-\delta}^{x_0} f(t)dt \\
&= [Vf](x_0-\delta) + \int_{x_0-\delta}^{x_0} f(t)dt \\
&\geq 0 + \int_{x_0-\delta}^{x_0} \varepsilon dt \\
&= 0 + \varepsilon(\delta) > 0,
\end{aligned}$$

a contradiction.

Thus V does not have 0 as an eigenvalue, despite the fact that $V - 0I_{\mathcal{W}}$ is not invertible.

S8.2. Example. Let $n \in \mathbb{N}$, $\mathcal{V}$ be a vector space of dimension n over a field $\mathbb{F}$, and $T \in \mathcal{L}(\mathcal{V})$. Suppose furthermore that $\alpha \in \mathbb{F}$ is an eigenvalue of T, and choose $0 \neq x \in \mathcal{V}$ such that $Tx = \alpha x$. Then $\{x\}$ is linearly independent, and so we can extend $\{x\}$ to a basis $\mathcal{B} := \{x, b_2, b_2, \ldots, b_n\}$ for $\mathcal{V}$.

Write $[T]_{\mathcal{B}} = [t_{ij}] \in \mathbb{M}_n(\mathbb{F})$, and observe that $Tx = \alpha x$ implies that $[Tx]_{\mathcal{B}} := \begin{bmatrix} \alpha \\ 0 \\ \vdots \\ 0 \end{bmatrix}$.

But $[Tx]_{\mathcal{B}} = \begin{bmatrix} t_{11} \\ t_{21} \\ \vdots \\ t_{n1} \end{bmatrix}$! Thus

$$[T]_{\mathcal{B}} = \begin{bmatrix} \alpha & t_{12} & t_{13} & \cdots & t_{1n} \\ 0 & t_{22} & t_{23} & \cdots & t_{2n} \\ 0 & t_{32} & t_{33} & \cdots & t_{3n} \\ \vdots & & & & \vdots \\ 0 & t_{n2} & t_{n3} & \cdots & t_{nn} \end{bmatrix}.$$

If the matrix $\begin{bmatrix} t_{22} & t_{23} & \cdots & t_{2n} \\ t_{32} & t_{33} & \cdots & t_{3n} \\ \vdots & & & \vdots \\ t_{n2} & t_{n3} & \cdots & t_{nn} \end{bmatrix}$ also has an eigenvalue, we can repeat this argument on the space $\mathcal{W} := \operatorname{span}\{b_2, b_3, \ldots, b_n\}$. This observation, combined with the fact that $\mathbb{C}$ is algebraically closed, is the key to Exercise 8.1 below.

S8.3. Example. Let $T = \begin{bmatrix} -1 & 2 & 0 \\ -4 & 2 & 3 \\ -4 & -2 & 7 \end{bmatrix}$. Then (after some calculation) we find that

$$p_T(x) = \det(T - xI_3) = (1-x)(3-x)(4-x).$$

Moreover,

- $\ker(T - 1I_3) = \operatorname{span}\{\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}\}$;
- $\ker(T - 3I_3) = \operatorname{span}\{\begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}\}$; and
- $\ker(T - 4I_3) = \operatorname{span}\{\begin{bmatrix} 2 & 5 & 6 \end{bmatrix}\}$.

Thus, if we let $R = \begin{bmatrix} 1 & 1 & 2 \\ 1 & 2 & 5 \\ 1 & 2 & 6 \end{bmatrix}$, then more calculation shows that $R^{-1} = \begin{bmatrix} 2 & -2 & 1 \\ -1 & 4 & -3 \\ 0 & -1 & 1 \end{bmatrix}$, and

$$R^{-1}TR = \operatorname{DIAG}(1,3,4) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 4 \end{bmatrix}.$$

S8.4. Example. With $T = \begin{bmatrix} -1 & 2 & 0 \\ -4 & 2 & 3 \\ -4 & -2 & 7 \end{bmatrix}$ as in Example S8.3, suppose we wished to find

$$Y := 17T^6 + 5T^3 - 6T + 3I_3.$$

Letting $q(z) = 17z^6 + 5z^3 - 6z + 3$, we find that $Y = q(T)$.

Now $D = \text{DIAG}(1, 3, 4) = R^{-1}TR$, and so $T = RDR^{-1}$. As previously noted, by induction and linearity, we find that

$$p(T) = p(RDR^{-1}) = Rp(D)R^{-1}$$

for any polynomial $p(z)$, and thus $q(T) = R(q(D))R^{-1}$.

Again, the whole point of this exercise is that $q(D)$ is easy to evaluate when D is diagonal. Indeed,

$$\begin{aligned} q(D) &= q(\text{DIAG}(1, 3, 4)) \\ &= \text{DIAG}(q(1), q(3), q(4)) \\ &= \text{DIAG}(17 + 5 - 6 + 3, 17(3^6) + 5(3^3) - 6(3) + 3, 17(4^6) + 5(4^3) - 6(4) + 3) \\ &= \text{DIAG}(19, 12513, 69931). \end{aligned}$$

Thus

$$q(T) = R \begin{bmatrix} 19 & 0 & 0 \\ 0 & 12513 & 0 \\ 0 & 0 & 69931 \end{bmatrix} R^{-1} = \begin{bmatrix} -12457 & -52309 & 102342 \\ -24988 & -174511 & 274596 \\ -24988 & -244442 & 344527 \end{bmatrix}.$$

You might want to check by computing $17T^6 + 5T^3 - 6T + 3I_3$ directly. Then again, you might not want to do so.

S8.5. Example. Let $n \in \mathbb{N}$ and $\mathbb{F}$ be a field. Then $n \times n$ **Jordan cell** is the matrix

$$J_n := \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \ddots & \cdots & 0 \\ \vdots & & & \ddots & & 0 \\ \vdots & & & & 0 & 1 \\ 0 & 0 & 0 & 0 & \cdots & 0 \end{bmatrix} \in \mathbb{M}_n(\mathbb{F}).$$

That is, if $J_n = [x_{ij}^{(n)}]$, then $x_{i\,i+1}^{(n)} = 1$ for all $1 \le i \le n-1$, while $x_{ij}^{(n)} = 0$ if $j \ne i+1$.

Also, given $T_j \in \mathbb{M}_{n_j}(\mathbb{F})$, $1 \le j \le m$, we define the **direct sum** of the matrices T_j to be

$$T := T_1 \oplus T_2 \oplus \cdots \oplus T_m := \begin{bmatrix} T_1 & 0 & 0 & \cdots & 0 \\ 0 & T_2 & 0 & \cdots & 0 \\ 0 & 0 & \ddots & 0 & \vdots \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & 0 & \cdots & T_m \end{bmatrix} \in \mathbb{M}_n(\mathbb{F}),$$

where $n = \sum_{j=1}^{m} n_j$.

If we set

$$T = (-3I_4 + J_4) \oplus (-3I_4 + J_4) \oplus (-3I_3 + J_3) \oplus (6I_3 + J_3) \oplus (2I_2 + J_2) \in \mathbb{M}_{16}(\mathbb{F}),$$

then $T \in \mathbb{T}_{16}(\mathbb{F})$ is upper-triangular, and so the eigenvalues of T appear on its diagonal, namely:

$$\sigma_p(T) = \{-3, 6, 2\}.$$

Moreover, since $T = [t_{ij}]$ is upper-triangular,

$$p_T(x) = \prod_{j=1}^{16}(t_{jj} - x) = (x - (-3))^{11}(x - 6)^3(x - 2)^2.$$

Thus $\mu_T(-3) = 11$, $\mu_T(6) = 3$, and $\mu_T(2) = 2$.

As for the geometric multiplicity of these eigenvalues, we invite the reader to verify that if $\mathcal{D} = (e_1, e_2, \ldots, e_{16})$ is the standard ordered basis for $\mathbb{F}^{16}$, then

- $\ker(T + 3I_{16}) = \ker(T - (-3)I_{16}) = \operatorname{span}\{e_4, e_8, e_{11}\}$, and so $\gamma_T(-3) = 3$;
- $\ker(T - 6I_{16}) = \operatorname{span}\{e_{14}\}$, and so $\gamma_T(6) = 1$; while
- $\ker(T - 2I_{16}) = \operatorname{span}\{e_{16}\}$, and so $\gamma_T(2) = 1$.

By Theorem 8.2.10, T is not diagonalisable.

We leave it as an exercise for the reader to determine whether or not there exists an example of a matrix $R \in \mathbb{M}_{16}(\mathbb{F})$ satisfying

(i) $\sigma_p(R) = \{-3, 6, 2\}$,
(ii) $\mu_R(-3) = 11$, $\mu_R(6) = 3$, $\mu_R(2) = 2$, and
(iii) $\gamma_R(-3) = 5$, $\gamma_R(6) = 2$ and $\gamma_R(2) = 2$.

S8.6. Example. Let $T = \begin{bmatrix} 5 & 8 \\ -4 & 5 \end{bmatrix} \in \mathbb{M}_2(\mathbb{R})$. In order for T to be diagonalisable, a minimum necessary condition is that $\sigma_p(T) \subseteq \mathbb{R}$. Note, however, that

$$p_T(x) = (5 - x)(5 - x) - (-4)(8) = (5 - x)^2 + 32 \geq 32 > 0 \text{ for all } x \in \mathbb{R}.$$

Thus $\sigma_p(T) = \varnothing$, and so T has no hope of being diagonalisable.

More generally, if $T = \begin{bmatrix} \alpha & \beta \\ \gamma & \alpha \end{bmatrix} \in \mathbb{M}_2(\mathbb{R})$, then

$$p_T(x) = (\alpha - x)^2 - \gamma\beta.$$

Thus if $\gamma\beta < 0$, $p_T(x) \geq |\beta\gamma| > 0$, and $\sigma_p(T) = \varnothing$.

Suddenly the idea of computing $q(T) = T^{2024} + 54T^{102} + 3I_2$ seems very daunting.

S8.7. Example. Let $T = \begin{bmatrix} 5 & 8 \\ -4 & 5 \end{bmatrix} \in \mathbb{M}_2(\mathbb{R})$ as above. We can take advantage of the fact that $\mathbb{R} \subseteq \mathbb{C}$ to think of T as a 2×2 complex matrix. Suddenly, when thinking of $T \in \mathbb{M}_2(\mathbb{C})$, we see that

$$p_T(x) = (5-x)^2 + 32 = x^2 - 10x + 57.$$

The *complex* roots of this polynomial are $\alpha_1 := 5 - 4\sqrt{2}i$ and $\alpha_2 = 5 + 4\sqrt{2}i$. Hence

$$\sigma_p(T) = \{\alpha_1, \alpha_2\} = \{5 - 4\sqrt{2}i, 5 + 4\sqrt{2}i\}.$$

Since $T \in \mathbb{M}_2(\mathbb{C})$ has two distinct eigenvalues, it is diagonalisable, and so we may use the techniques of Example S8.4 to compute $q(T) = T^{2024} + 54T^{102} + 3I_2$.

Since $T \in \mathbb{M}_2(\mathbb{R})$, and since $q(x) = x^{2024} + 54x^{102} + 3$ has real coefficients, it should be clear that $q(T) \in \mathbb{M}_2(\mathbb{R})$, even though the diagonal form of T in $\mathbb{M}_2(\mathbb{C})$ has complex entries. (It's applying the similarities that will get rid of these for us.)

S8.8. Example. Let $n \in \mathbb{N}$, $\mathbb{F}$ be a field, and $T \in \mathbb{M}_n(\mathbb{F})$. Suppose that $\alpha \in \sigma_p(T)$, and let $d := \mu_T(\alpha)$. Choose $0 \neq y_1 \in \mathbb{F}^n$ such that $Ty_1 = \alpha y_1$, and extend the linearly independent set $\{y_1\}$ to a basis $\mathcal{B}_1 := (y_1, b_2, b_3, \ldots, b_n)$ for $\mathbb{F}^n$. Relative to this basis, we find that

$$[T]_{\mathcal{B}} := \begin{bmatrix} \alpha & t_{12} & t_{13} & \cdots & t_{1n} \\ 0 & t_{22} & t_{23} & \cdots & t_{2n} \\ \vdots & \vdots & & \cdots & \vdots \\ 0 & t_{n2} & t_{n3} & \cdots & t_{nn} \end{bmatrix} = \begin{bmatrix} \alpha & T_2 \\ 0 & T_4 \end{bmatrix},$$

where $T_2 = \begin{bmatrix} t_{12} & t_{13} & \cdots & t_{1n} \end{bmatrix} \in \mathbb{M}_{1\times(n-1)}(\mathbb{F})$ and $T_4 = [t_{ij}]_{2\leq i,j\leq n} \in \mathbb{M}_{n-1}(\mathbb{F})$. By computing $p_T(x) = \det(T - xI_n)$ by cofactors along the first column, we deduce that

$$p_T(x) = (\alpha - x)\det(T_4 - xI_{n-1}) = (\alpha - x)p_{T_4}(x).$$

If $d = \mu_T(\alpha) > 1$, then this means that $(x-\alpha)^2 | p_T(x)$, and therefore $(x-\alpha)|p_{T_4}(x)$. In other words, $\alpha \in \sigma_p(T_4)$.

We may therefore repeat this argument with T_4 to find a new basis

$$\mathcal{C} = \{y_2, c_3, c_4, \ldots, c_n\}$$

for $\operatorname{span}\{b_2, b_3, \ldots, b_n\}$ where

$$[T_4]_{\mathcal{C}} = \begin{bmatrix} \alpha & X_2 \\ 0 & X_4 \end{bmatrix},$$

where $X_2 \in \mathbb{M}_{1\times(n-2)}(\mathbb{F})$ and $X_4 \in \mathbb{M}_{n-2}(\mathbb{F})$. Then with $\mathcal{B}_2 := (y_1, y_2, c_3, c_4, \ldots, c_n)$, we find that

$$[T]_{\mathcal{C}} = \begin{bmatrix} Z_1 & Z_2 \\ 0 & Z_4 \end{bmatrix},$$

where $Z_1 = \begin{bmatrix} \alpha & * \\ 0 & \alpha \end{bmatrix}$.

Again, if $d \geq 3$, then the same logic shows that $\alpha \in \sigma_p(Z_4)$ and we can repeat the argument yet again. More generally, we can repeat the argument d times to find a basis $\mathcal{D}$ relative to which

$$[T]_{\mathcal{D}} = \begin{bmatrix} Q_1 & Q_2 \\ 0 & Q_4 \end{bmatrix},$$

where $Q_1 = \begin{bmatrix} \alpha & r_{12} & r_{13} & \cdots & r_{1d} \\ 0 & \alpha & r_{23} & \cdots & r_{2d} \\ & \ddots & \ddots & & \vdots \\ & & & & \vdots \\ 0 & 0 & \cdots & 0 & \alpha \end{bmatrix} \in \mathbb{M}_d(\mathbb{F})$.

Once again, we find that $p_T(x) = p_{Q_1}(x)p_{Q_4}(x) = (\alpha - x)^d p_{Q_4}(x)$. Since $\mu_T(\alpha) = d$, this implies that $(x - \alpha)$ does not divide $p_{Q_4}(x)$, and so α is not an eigenvalue of Q_4.

The moral of the story is: you can get a *lot* of information about T from $\sigma_p(T)$, and the algebraic and geometric multiplicities of α for each $\alpha \in \sigma_p(T)$.

This is the content of the next linear algebra course.

Appendix

A8.1. In a small number of texts, one refers to eigenvalues and eigenvectors as **characteristic values** and **characteristic vectors** respectively. The adjective "*eigen*" is of German origin, and can be used to mean "intrinsic to" or "inherent" or even "characteristic" – to wit:

> *Sie, mit allem ihr eigenen Charme, hob das Pferd auf und warf es auf den Fußballplatz.*

Exercises for Chapter 8

Exercise 8.1.*

This problem is challenging, but definitely worthwhile. The reader might first want to consult Example S8.2 above.

Let $n \in \mathbb{N}$ and $T = [t_{ij}] \in \mathbb{M}_n(\mathbb{C})$. Prove that there exists a basis $\mathcal{B} = \{b_1, b_2, \ldots, b_n\}$ relative to which the matrix $[T]_{\mathcal{B}}$ of T is upper-triangular.

Exercise 8.2.

Let $m, n \in \mathbb{N}$, $\mathbb{F}$ be a field, $A \in \mathbb{M}_m(\mathbb{F})$ and $B \in \mathbb{M}_n(\mathbb{F})$. Recall that

$$A \oplus B := \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix} \in \mathbb{M}_{m+n}(\mathbb{F}).$$

Prove that $p_{A\oplus B}(x) = p_A(x)p_B(x)$.

Exercise 8.3.

Let $m, n \in \mathbb{N}$, $\mathbb{F}$ be a field, $A \in \mathbb{M}_m(\mathbb{F})$, $B \in \mathbb{M}_{m+n}(\mathbb{F})$, and $D \in \mathbb{M}_n(\mathbb{F})$. Let

$$T := \begin{bmatrix} A & B \\ 0 & D \end{bmatrix} \in \mathbb{M}_{m+n}(\mathbb{F}).$$

Prove that $p_T(x) = p_A(x)p_D(x)$, and that $\sigma_p(T) = \sigma_p(A) \cup \sigma_p(D)$.

Exercise 8.4.

Let $n \in \mathbb{N}$, $\mathbb{F}$ be a field, $T \in \mathbb{M}_n(\mathbb{F})$, and $D = [d_{ij}] \in \mathbb{M}_n(\mathbb{F})$ be a diagonal matrix – i.e. $d_{ij} = 0$ if $1 \leq i \neq j \leq n$. Let $\{T\}' := \{X \in \mathbb{M}_n(\mathbb{F}) : TX = XT\}$, the **commutant** of T. More generally, if $\mathcal{F} \subseteq \mathbb{M}_n(\mathbb{F})$, then

$$\mathcal{F}' := \{X \in \mathbb{M}_n(\mathbb{F}) : XF = FX \text{ for all } F \in \mathcal{F}\}.$$

(a) Prove that $\{T\}'$ is a vector space over $\mathbb{F}$, and that if $X, Y \in \{T\}'$, then $XY \in \{T\}'$. (In fact, $\{T\}'$ is an **algebra** over $\mathbb{F}$. That is, it is both a vector space over $\mathbb{F}$ and a **ring**.)

(b) Describe $\{D\}'$.

(c) Show that there exists $T_0 \in \mathbb{M}_n(\mathbb{F})$ such that $\{D, T_0\}' = \mathbb{F}I_n$.

Exercise 8.5.

Let $\mathcal{V}$ be an n-dimensional vector space over $\mathbb{C}$, and let $T \in \mathcal{L}(\mathcal{V})$. Recall that a subspace $\mathcal{M} \subseteq \mathcal{V}$ is said to be **invariant** for T if $Tm \in \mathcal{M}$ for all $m \in \mathcal{M}$. In this case, we may decompose $\mathcal{V} = \mathcal{M} \dot{+} \mathcal{N}$ and relative to this decomposition,

$$T = \begin{bmatrix} T_1 & T_2 \\ 0 & T_4 \end{bmatrix}.$$

Let $\varnothing \neq \Omega \subsetneq \sigma_p(T)$ be a subset of the eigenvalues of T. Prove that there exists an invariant subspace $\mathcal{M}$ for T such that with the above notation, $\sigma_p(T_1) = \Omega$ and $\sigma_p(T_4) = \sigma_p(T) \setminus \Omega$.

Exercise 8.6.

Let $n \in \mathbb{N}$, $\mathbb{F}$ be a field, $\alpha \in \mathbb{F}$ and $A \in \mathbb{M}_n(\mathbb{F})$. Prove that for $j \geq 1$, $\ker(A - \alpha I_n)^j$ is invariant for all $T \in \{A\}'$.

We say that a subspace $\mathcal{M}$ of $\mathcal{V}$ is **hyperinvariant** for $T \in \mathcal{L}(\mathcal{V})$ if $X\mathcal{M} \subseteq \mathcal{M}$ for all $X \in \{T\}'$. Thus $\ker(A - \alpha I_n)^j$ is hyperinvariant for A.

Exercise 8.7.

Let $\mathcal{V}$ be a finite-dimensional vector space over a filed $\mathbb{F}$. Given an example of an operator $T \in \mathcal{L}(\mathcal{V})$, and of a subspace $\mathcal{M} \subseteq \mathcal{V}$ which is invariant for T, but not hyperinvariant for T.

Exercise 8.8.

Let $n \in \mathbb{N}$, $\mathbb{F}$ be a field and $A = [a_{ij}] \in \mathbb{M}_n(\mathbb{F})$. Suppose that $\sigma_p(A) = \{0\}$. Prove that $A^n = 0$.

Hint. First find a basis $\mathcal{D}$ for $\mathbb{F}^n$ relative to which $[L_A]_{\mathcal{D}}$ is **strictly upper-triangular**, that is: $[L_A]_{\mathcal{D}} \in \mathbb{T}_n(\mathbb{F})$, and all of the diagonal values of $[L_A]_{\mathcal{D}}$ are also equal to zero. It might help to consider $\mathcal{M}_k := \ker A^k$, $1 \leq k \leq n$.

Exercise 8.9.*

Let $n \in \mathbb{N}$, $\mathbb{F}$ be a field of characteristic zero and $A, B \in \mathbb{M}_n(\mathbb{F})$. Let $\lambda_A : \mathbb{M}_n(\mathbb{F}) \to \mathbb{M}_n(\mathbb{F})$ be the map $\lambda_A(X) = AX$, and $\varrho_B : \mathbb{M}_n(\mathbb{F}) \to \mathbb{M}_n(\mathbb{F})$ be the map $\varrho_B(X) = XB$.

(a) Show that λ_A and $\varrho_B \in \mathcal{L}(\mathbb{M}_n(\mathbb{F}))$.
(b) Find $\sigma_p(\lambda_A)$ and $\sigma_p(\varrho_B)$.
(c) Let $\tau := \lambda_A - \varrho_B$. Show that $\sigma_p(\tau) \subseteq \sigma_p(A) - \sigma_p(B) := \{\alpha - \beta : \alpha \in \sigma_p(A), \beta \in \sigma_p(B)\}$.
(d) Let $\delta_A := \lambda_A - \varrho_A$. Show that δ_A is not invertible, and find an element $X \in \mathbb{M}_n(\mathbb{F})$ such that $X \notin \operatorname{ran} \delta_A$.
(e) Prove that $\delta_A(XY) = X\delta_A(Y) + \delta_A(X)Y$. We refer to such a linear map as a **derivation**.

Exercise 8.10.

Let $n \in \mathbb{N}$, $T \in \mathbb{T}_n(\mathbb{C})$. Prove that there exists a sequence $(R_m)_m$ of invertible matrices such that if $X := [x_{ij}^{(m)}] := R_m^{-1} T R_m$, then $\lim_m x_{ij}^{(m)} = 0$ for all $1 \leq i \neq j \leq n$, while $x_{ii}^{(m)} = t_{ii}$ for all $1 \leq i \leq n$ and $m \geq 1$.

Bibliography

[And14] R. André. *Axioms and Set Theory; a first course in set theory*. R. André, 2014.

[Bla84] A. Blass. *Existence of bases implies the axiom of choice*, volume 31 of *Contemp. Math.*, pages 31–33. Amer. Math. Soc., Providence, RI, 1984.

[FIS97] S. Friedberg, A. Insel, and L. Spence. *Linear algebra*. Prentice-Hall Inc., Upper Saddle River, 3rd edition edition, 1997.

[HK71] K.M. Hoffman and R. Kunze. *Linear algebra*. Prentice-Hall Inc., Upper Saddle River, 2nd edition edition, 1971.

[Str88] G. Strang. *Introduction to linear algebra*. Wellesley-Cambridge Press, Wellesley, MA, 6th edition edition, 1988.

Index