Linearity

Adam Layne

2025-01-14

Table of contents

What are these notes?

These are notes for a first course in Linear Algebra.

License

Preface

This section discusses why these notes exist. Students may skip this section.

Why publish a new set of Linear Algebra notes?

Linear Algebra, like Calculus, is one of the math subjects with the most textbooks, so it's reasonable to ask why a new set of notes is needed. Plainly, I looked at the six open-access books on the subject on the AIMath website and found that none of them were fit for my purpose (detailed below).

The perspective of these notes

These notes are constructed to vindicate the following objectives:

1. It is morally right for course materials to be free. Few existing books on this subject in English satisfy this criterion. This book, and the source used to generate it are freely available with a permissive license.

2. In practice, scientists, engineers, etc. need to be able to recognize linearity so that they may choose the correct solution techniques. They also need to understand why linear problems are preferable to non-linear ones so that they might try to massage their current problem into a linear one.

3. When we say "solution techniques" as above, 99% of the time we mean software packages. Mathematicians and physicists teach linear algebra techniques in colleges and universities, and emphasize by-hand solution techniques for historical and cultural reasons. Most working people who encounter such problems do not use such techniques, they recognize that their problem is linear and offload the problem to a software package. Mathematicians and physicists generally get a second pass at learning linear algebra in a more theory-heavy context (at the very least when learning modules), and so do not need that approach in a first course.

4. The usefulness of linear algebra techniques stems wholly from the homomorphism property of linear maps:

$$L(aV + bW) = aL(V) + bL(W)$$

> No introductory, open-access, English language books on the topic that I am aware of motivate the study of the subject with this point. They traditionally begin with coordinate geometry or solving systems of linear equations. It is a very mathematicians' way of thinking to motivate study of a topic by identifying a class of equations and asking "How do we solve them? What properties do they have?" This is not a way of thinking that is useful for people encountering linearity in the wild. Axler (2024) is an example of an open text that properly emphasizes this aspect from the beginning, but is not suitable for a class where students have not yet learned proofs.
>
> Coordinate geometry is, at least, a class of real problems where linear techniques naturally arise, but the relevance of this as an example from "the wild" has basically vanished in the last 70 years. Today's scientists and engineers are more likely to encounter linearity in optimization, data science, machine learning, or numerical PDEs.

So, for these reasons, I set out to write my own course notes.

Part I

Vector Spaces

Chapter 1

What is Linearity?

The function $C : \mathbb{R} \to \mathbb{R}$ given by

$$C(r) := 2\pi r$$

computes the circumference of a circle, given its radius.

> **Notation**
>
> First, let's talk about this notation.
> The notation $C : \mathbb{R} \to \mathbb{R}$ tells you about the inputs and outputs of the function C. When we write $f : A \to B$, we mean that f takes inputs from the set A and creates outputs in the set B. You can think of f as a machine transforming As into Bs.
> The notation $C(r) := 2\pi r$ tells you that the *definition* of the function C appears here. This is to avoid confusion like the following:
>
> $$f(x) = x^2$$
>
> If the function f hasn't appeared before, then this equation is probably a definition. But if we wrote "Given $f(x) = x - 2$, solve the equation $f(x) = x^2$", then the same equation is not a definition. To avoid this ambiguity, when we write an equals sign with a colon on one side like this $A := B$ or $B =: A$, we mean that the name on the side of the $=$ *with* the colon is defined to be the expression on the side without.

The circumference of a circle has a couple nice properties. First, the circumference of a circle of radius 14 is twice the circumference of a circle of radius 7:

$$C(7) = 2\pi(7) = 14\pi$$
$$C(14) = 2\pi(14) = 28\pi$$

This holds in general; multiplying the radius of a circle by k also changes the circumference by a factor of k:

$$C(kr) = kC(r).$$

Furthermore, adding any amount to the radius increases the circumference in a predictable way:

$$C(r_1 + r_2) = C(r_1) + C(r_2).$$

It's a bit remarkable that these two properties hold not just for circles; scaling any shape in the plane (with circumference c) by a factor of k multiplies its circumference by k (its new circumference is kc), and increasing the scale by a constant s increases its circumference by sc.

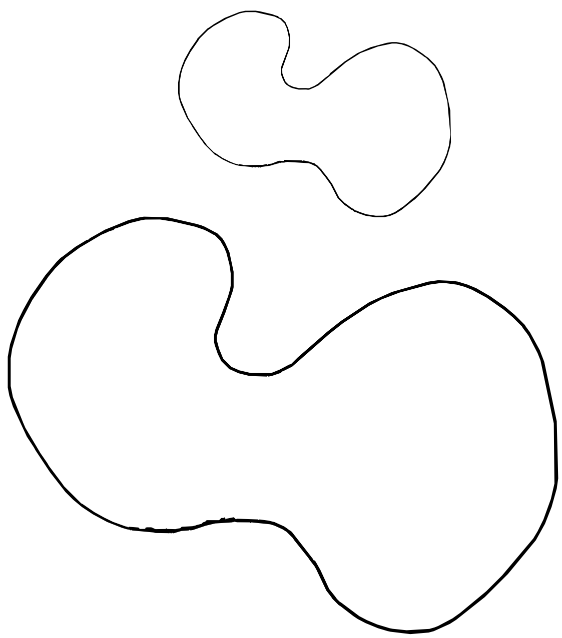

Figure 1.1: Scaling a curve by a factor of 2

1.1 Two properties

What other functions have the following two properties?

$$\begin{aligned} &\text{Property 1:} \quad f(ax) = af(x) \\ &\text{Property 2:} \quad f(x+y) = f(x) + f(y) \end{aligned}$$

Exercise 1.1. Demonstrate that the *area* of a circle as a function of its radius does not satisfy properties 1 and 2.

Exercise 1.2. Can you think of any shape in the plane whose area (as a function of scaling) satisfies properties 1 and 2? If yes, which? If no, why not?

Exercise 1.3. Can you think of any shape in the 3-dimensional space whose *volume* (as a function of scaling) satisfies properties 1 and 2? If yes, which? If no, why not?false

Classification of functions satisfying Properties 1 and 2

So far, we have seen that some functions satisfy properties 1 and 2, and others do not.

Table 1.1: Functions that do and don't satisfy properties 1 and 2

do	don't
circumference of a circle as a function of radius	area of a circle as a function of radius
circumference of any shape in the plane as a function of scale	area of any shape in the plane as a function of scale
	volume of any shape in 3-dimensions as a function of scale

Why does scaling satisfy properties 1 & 2 for any shape, not just circles?

We asserted above that property 1 is not just satisfied by circles (when you scale the radius) but is satisfied by all curves. Why is this the case?

What do we mean by scaling a figure in the plane by a factor of 2? Well, a reasonable answer is to say that a figure is a set of points and each point has an x and y coordinate. For example, the circle of radius r is the set of points with coordinates given by

$$\begin{aligned} x =& r\cos\theta \\ y =& r\sin\theta \end{aligned}$$

where $\theta \in [0, 2\pi]$.

Notation

When we want to refer to a set, we will often use notation like the following:

$$\left\{(x, y) \in \mathbb{R}^2 \mid x = r\cos\theta, y = r\sin\theta, \theta \in [0, 2\pi]\right\}.$$

In general, this notation has the form

$$B = \{f(x_1, x_2, \ldots) \in A \mid \text{ constraints involving } x_1, x_2, \ldots\}.$$

The process for constructing the set B is the following:

1. Find all the x_i that satisfy the constraints to the right of the $\mid$ symbol.
2. Plug all the x_i you found in the previous step into the function f.
3. The function f produces things in A, and so the set of all the things you produced in the last step is a collection of some (but not necessarily all) of the things in A.
   This is the set B.

Consider the point $(4, 2)$ in the plane. We can think of this point as the "sum" of its x and y coordinates:

$$(4, 2) = (4, 0) + (0, 2).$$

To scale this point by a factor of 2, it seems reasonable to multiply both coordinates by 2:

$$S_2((4, 2)) = (8, 4).$$

Notice that the function $S_2 : \mathbb{R}^2 \to \mathbb{R}^2$ "multiply coordinates by 2" has properties 1 and 2.

Exercise 1.4. Verify this.

Exercise 1.5. Check that applying S_2 to a circle of radius r produces a circle of radius $2r$.

Now, for any curve $\gamma : [a, b] \to \mathbb{R}^2$ given by

$$\gamma(t) = (x(t), y(t)),$$

its length can be computed by

$$L = \int_a^b \sqrt{x'(t)^2 + y'(t)^2} \quad dt.$$

(You may have seen this in multivariable calculus or physics.) Consider the composition

$$[a,b] \xrightarrow{\gamma} \mathbb{R}^2 \xrightarrow{S_2} \mathbb{R}^2$$

The formula is

$$(S_2 \circ \gamma)(x) = (2x(t), 2y(t))$$

and the length is given by

$$\begin{aligned} L &= \int_a^b \sqrt{\left([2x(t)]'\right)^2 + \left([2y(t)]'\right)^2} \quad dt \\ &= \int_a^b \sqrt{\left(2x'(t)\right)^2 + \left(2y'(t)\right)^2} \quad dt \\ &= \int_a^b \sqrt{4\left(x'(t)\right)^2 + 4\left(y'(t)\right)^2} \quad dt \\ &= \int_a^b 2\sqrt{\left(x'(t)\right)^2 + \left(y'(t)\right)^2} \quad dt \\ &=2\int_a^b \sqrt{\left(x'(t)\right)^2 + \left(y'(t)\right)^2} \quad dt \end{aligned}$$

Notice that what we obtain on the last line is exactly twice the length of the curve γ. Convince yourself that there is nothing special about the number 2 here; if we had replaced S_2 by S_{17}, then we would have obtained 17 times the length of γ in the last line.

More functions which break up over "sums"

Differentiation

Now consider how we compute the derivative of a function like the following:

$$\begin{aligned} \frac{d}{dx}\left[2x^2 + x\right] &= \frac{d}{dx}\left[2x^2\right] + \frac{d}{dx}\left[x\right] \\ &= 2\frac{d}{dx}\left[x^2\right] + \frac{d}{dx}\left[x\right] \\ &= 4x + 1 \end{aligned}$$

If we let D : FUNCTIONS $\to$ FUNCTIONS be the operation of taking a derivative, then in the first line we used

$$D[f_1(x) + f_2(x)] = D[f_1(x)] + D[f_2(x)]$$

and in the second line we used the fact that, when k is constant,

$$D[kf(x)] = kD[f(x)].$$

Thus, D (that is, differentiation) satisfies properties 1 and 2. (Although you should probably be uncomfortable that we wrote D : FUNCTIONS $\to$ FUNCTIONS above. What is the set FUNCTIONS? Are all functions differentiable? We will address this later. For now, it suffices to replace FUNCTIONS above with P_n, the set of all polynomials of degree n. In fact, it is the case that $D : P_n \to P_{n-1}$ if $n \geq 1$.)

Definite integration

Fix an interval $[a, b]$ and consider

$$I(f) := \int_a^b f(x) \quad dx.$$

Exercise 1.6. Check that I has properties 1 and 2.

Why are properties 1 and 2 useful

Why is it useful that D satisfies properties 1 and 2? It allows us to compute derivatives of complicated expressions like $2x^2 + x$ if we only know the computation on some simple parts of the expression. Knowing the derivative of x^2 and x is all that is needed.

Similarly, if we know the circumference of a shape in the plane at one scale, we can compute its circumference at all scales using property 1.

Not all functions are linear, but if a function is linear, it is much easier to compute with.

1.2 Linear functions

Definition 1.1 (linear function)**.** A function satisfying properties 1 and 2 is called *linear*.

> ⚠ Warning
>
> We use the term "linear" for these functions, but we also use the word "line" for graphs in the plane with formula $f(x) = mx + b$. This term is overloaded and means different things in these two contexts.

Exercise 1.7. Show that $f(x) = mx + b$ is only a linear function when $b = 0$.

Chapter 2

The notion of a Vector Space

If $L : A \to B$ is linear, what must be true about A and B?

Let's go back to the definition of a linear function. A function is linear if and only if it satisfies the following two properties:

$$\begin{aligned} &\text{Property 1:} \quad L(ax) = aL(x) \\ &\text{Property 2:} \quad L(x+y) = L(x) + L(y). \end{aligned}$$

Let's list a few things that must be true to arrive at these expressions:

- there are terms (like a in property 1) that we can factor through L
- there are terms (like x, y in properties 1 and 2) that we cannot factor through L
- there is some kind of addition on the xs and ys

There are perhaps some more properties that would be nice, and that are true about all the domains and codomains of linear functions we have seen so far:

- the as (that we can pull through L) have nice algebraic properties $(+, -, \times, \text{division})$
- the $+$ operation on the xs and ys has some nice properties, too (existence of an identity, commutativity, etc)

Eventually, mathematicians (who were working with linear functions intuitively) worked out the minimal set of facts that one needs about the domain and codomain of a linear function for everything to be coherent. Here it is

2.1 Definition

Definition 2.1. A *vector space* is a set V of *vectors* and a set F of *scalars* that satisfy the following properties

- the set of scalars F is a field (see the next section for more information on what we mean by field),
- there is a function $+\colon V \times V \to V$, called *vector addition* that
  - is associative: $(X + Y) + Z = X + (Y + Z)$,
  - is commutative: $X + Y = Y + X$,
  - has a 0 (an additive identity) which means that 0 satisfies $0 + X = X$ for all $X \in V$,
  - has negatives (additive inverses) which means that for each $X \in V$ there is an element $Y \in V$ such that $X + Y = 0$ (one can prove that there is only one inverse of X, and this is usually written $-X$),
- V is closed under vector addition and scalar multiplication, which means that
  - for every $X, Y \in V$, $X + Y$ is also in V
  - for every $a \in F$ and $X \in V$, aX is also in V
- the following distributive laws hold:
  - $(a + b)X = aX + bX, \quad \forall a, b \in F, X \in V$
  - $a(X + Y) = aX + aY, \quad \forall a \in F, X, Y \in V$

Notation

Above, we used the symbol $\forall$. This is a symbol that mathematicians use that literally just means "for all".
So, as an example $x \in P_3, \forall x \in P_2$ is read as "for all x in the set of second degree polynomials, x is in the set of third degree polynomials", which is just a wordy way of saying that all second degree polynomials are third degree polynomials.

All of the properties above hold (with the field of scalars F set to $\mathbb{R}$) for all the spaces we have used so far as the domain and codomain of a linear function. (You may want to check this for yourself quickly. Most of these facts obviously hold for polynomials, functions, points in $\mathbb{R}^2$, and $\mathbb{R}$ itself. However, one must still check all of them before using a set as the domain or codomain of a linear function.)

Not all sets of mathematical objects satisfy these properties, though. An example is the set of points on the sphere. It turns out that there is no way (although the proof is hard) to turn that set into a vector space.

The way you should think about this definition is the following:

- a set A satisfying all of the properties in Definition 2.1 **may** be used as the domain or codomain of a linear function, but

- a set A that fails to satisfy any of the properties above **can never be used** as the domain or codomain of a linear function. If you tried to use it in this way, you would eventually run into statements that make no sense.[1]

An example of the kind of problem you can run into is the following: Let $L : A \to B$ be linear and suppose A is a vector space as defined above but that B fails to have an additive identity (there is no zero vector in B). We can compute

$$L(0) = 0L(0) = 0(?) = 0$$

where the ? stands for whatever $L(0)$ maps to. Notice that in the end, it doesn't matter because we conclude that $L(0) = 0 \in B$. But this is nonsensical since we assumed there was no zero vector in B.

Exercise 2.1. Check that $\mathbb{R}$ forms a vector space (the set of scalars is $\mathbb{R}$).

Exercise 2.2. Check that $\mathbb{R}^2$ is a vector space.

2.2 Scalars

In the above definition, we said that the scalars for a vector space must come from a field. What does this mean?

For the purposes of this text, our field will always be $\mathbb{R}$ or the field of complex numbers:

$$\mathbb{C} := \{a + bi \mid a, b \in \mathbb{R}\}$$

where the symbol i has the property that $i^2 = -1$ (and so we sometimes write it as $\sqrt{-1}$). These two sets are the set of scalars for most vector spaces found in applications in the wild.

Exercise 2.3. Check that $\mathbb{C}$ can be thought of as a vector space with field of scalars $\mathbb{R}$.

Definition 2.2. Let P_n be the set (which we mentioned informally in the previous chapter)

$$\{a_n x^n + a_{n-1} x^n + \cdots + a_1 x + a_0 \mid a_n, a_{n-1}, \ldots, a_1, a_0 \in \mathbb{R}\}.$$

This is called the *set of polynomials of degree n*.

[1] It is the case that some maps between spaces that are weaker than vector spaces have defining properties 1 and 2. The spaces are a generalization of vector spaces called modules and the maps are called module homomorphisms. If you look at the definition of module homomorphism on wikipedia, you'll see the same two equations we used to define linear functions.

Exercise 2.4. Is P_n a vector space for each $n \in \mathbb{N}$?

From the previous two examples, you might notice that the definitions

$$\mathbb{C} := \{a + bi \mid a, b \in \mathbb{R}\}$$

and

$$P_n := \{a_n x^n + a_{n-1} x^n + \cdots + a_1 x + a_0 \mid a_n, a_{n-1}, \dots, a_1, a_0 \in \mathbb{R}\}$$

look similar, and they give you a hint about how to think of these as vector spaces; in both cases, they are constructed as the sum of things with coefficients in $\mathbb{R}$. We will investigate this further in the next chapter.

Abstract algebra

In case you are interested, the full definition of a field can be found on Wikipedia. A field that you already know about, but that is not $\mathbb{R}$ or $\mathbb{C}$ is the field of rational numbers:

$$\mathbb{Q} := \left\{ \frac{a}{b} \middle| a, b \in \mathbb{Z}, b \neq 0 \right\}.$$

There are esoteric examples of fields, and also algebraic structures stranger than fields. If this is interesting to you, you might try to take a course in Abstract Algebra. I learned this topic from Liz Stanhope and Dummit and Foote (2004).

In computing, real numbers (which do form a field) are often represented by 32 or 64 bit floating point numbers. It is perhaps interesting to know that floating point numbers (in any number of bits) (which are encountered **often** in the wild) do not form a field because addition of floating point numbers is not associative.

Chapter 3

Bases

Remember that the derivative function $D(f) := \frac{df}{dx}$ allows us to compute the derivative of, for example

$$f(x) = x^2 + 4\sin x$$

if we know how to compute D on only x^2 and $\sin x$. In fact, knowing just how to compute D on these two functions and knowing that D is linear allows us to compute the derivative of any function of the form

$$x \mapsto ax^2 + b\sin x, \quad a, b \in \mathbb{R}.$$

Notation

We just introduced the notation $\mapsto$. This is a mathematician's way of referring to a function without giving it a name.
As an example both of

$$x \mapsto x^2 \quad \text{and} \quad f(x) = x^2$$

are the same function. On the right, we choose to give the function a name "f". On the left, we describe the function by saying how inputs are transformed into outputs, but we don't give it a name. The function on the left is just "the function that squares its input". It doesn't have a name, so this is called an *anonymous function*.
Mathematicians use an anonymous function when they won't want to refer to a function again later, and so have no need of giving it a name. If you'll want to refer to a function later, it's useful to give it a name when you define it, so you'll write $f(x) = x^2$ (or use some other letter if f is already

used for something else).
This pattern also shows up in computer science, although they often call anonymous functions *lambdas* and instead of writing $x \mapsto x^2$ they will write `x.x^2` or `fun x => x^2` or some variation thereof.

From a very small amount of information, we actually know a lot about D. Thus, we begin this chapter straightforwardly with a question, which is the title of the next section:

3.1 What is the minimal amount of information needed to unambiguously describe a linear function?

Consider the following example:

Assume that $L : P_2 \to P_0$ is linear and that

$$L(x^2 + 3x) = 4.$$

Can we determine L on every element of P_2? That is, can we compute

$$L(ax^2 + bx + c)?$$

Let's try our best. We would like to isolate a term that looks like x^2+3x because that is something we have information about.

$$\begin{aligned} L(ax^2 + bx + c) =& L(ax^2 + bx) + L(c) && \text{by property 1} \\ =& L\left(a\left[x^2 + \frac{b}{a}x\right]\right) + L(c) && \text{assuming } a \neq 0 \\ =& aL\left(x^2 + \frac{b}{a}x\right) + L(c) && \text{by property 2} \end{aligned}$$

then we are basically stuck. Although we can force x^2 to show up, we can't *at the same time* force $3x$ to show up; forcing a coefficient of 1 on x^2 will always effect the coefficient of x. Furthermore, we have no information about how to deal with $L(c)$.

So, in this case, knowing that L is linear and knowing its value on *one* input is insufficient.

Exercise 3.1. In the example above, is it enough to know that

$$\begin{aligned} L(x^2) =& 1, \\ L(x) =& 1, \end{aligned}$$

to be able to determine L on all of P_2?

What if we know

$$\begin{aligned} L(x^2) =& 1, \\ L(x) =& 1, \quad ? \\ L(1) =& -2 \end{aligned}$$

3.2 Writing vectors relative to a set of vectors

Let $X, Y \in \mathbb{R}^2$ be given by $X := (1, 1)$ and $Y := (1, -1)$. Suppose $L : \mathbb{R}^2 \to \mathbb{R}$ is linear and also that we know

$$\begin{aligned} L(X) =& 1, \\ L(Y) =& -2 \end{aligned}.$$

Can we compute L on the vector $(2, 3)$? If the answer to this question were yes then there would have to be a way to express $(2, 3)$ in terms of X and Y (since those are the only two values we know anything about). That is, there would be $a, b \in \mathbb{R}$ such that

$$(2, 3) = aX + bY = a(1, 1) + b(1, -1) = (a, a) + (b, -b) = (a + b, a - b).$$

So, L can be computed on $(2, 3)$ if and only if the following system has a solution:

$$\begin{aligned} 2 =& a + b, \\ 3 =& a - b \end{aligned}.$$

We can solve this system (solving for, say, a in the first and substituting the value in the second). We find that there is only one solution: $a = \frac{5}{2}, b = -\frac{1}{2}$.

Using this information, how do we compute $L(2, 3)$? Well, now we know that

$$(2, 3) = \frac{5}{2}(1, 1) + \left(-\frac{1}{2}\right)(1, -1)$$

so let's just apply the function L to both sides of this equation:

$$\begin{aligned}
L(2,3) =& L\left[\frac{5}{2}(1,1) + \left(-\frac{1}{2}\right)(1,-1)\right] \\
=& L\left[\frac{5}{2}(1,1)\right] + L\left[\left(-\frac{1}{2}\right)(1,-1)\right] \\
=& \frac{5}{2}L(1,1) + \left(-\frac{1}{2}\right)L(1,-1) \\
=& \frac{5}{2}1 + \left(-\frac{1}{2}\right)(-2) \\
=& \frac{5}{2} + 1 \\
=& \frac{7}{2}
\end{aligned}$$

Exercise 3.2. For each =-symbol in the computation above, write the assumption, property, or rule that tells us we are allowed to conclude the left hand side is equal to the right hand side.

The above computation shows that the answer to our question is yes, this amount of information about L is sufficient to compute $L(2,3)$. Notice that this is the case even though we were not given the formula for L.

You may want to convince yourself that the vector $(2,3)$ above was not special; given the information we have about L, we can compute $L(k,l)$ for any $(k,l) \in \mathbb{R}^2$.

Exercise 3.3. If we know that $L : \mathbb{R}^2 \to \mathbb{R}$ is linear and that

$$L(0,1) = 1, \quad L(0,2) = -6,$$

can we compute $L(1,2)$? If yes, compute it. If no, why not?

Exercise 3.4. If we know that $L : \mathbb{R}^2 \to \mathbb{R}$ is linear and that

$$L(0,1) = 2, \quad L(1,1) = 2,$$

can we compute $L(1,2)$? If yes, compute it. If no, why not?

Linear combinations

The trick to computing $L(2,3)$ in the previous section was to rewrite $(2,3)$ in the form

$$(2,3) = a(1,1) + b(1,-1).$$

This is a good ansatz[^1] because we are able to split over the + and pull the a and b out (using the fact that L is linear). [^1]: This is a German word that, in math, means a guess with some free parameters (in this case, a and b) in it. We compute with this guess, and then later find values of the parameters that make the computation valid. This pattern is so common when working with linear functions that we have a name for it:

Definition 3.1. If $X_1, X_2, \dots, X_n$ is a collection of vectors in the vector space V, then any expression of the form

$$a_1 X_1 + a_2 X_2 + \cdots + a_n X^n, \quad a_i \in F$$

is called a *linear combination* of $X_1, X_2, \dots, X_n$.

We then have the following generalization of the computation we did in the example above:

Theorem 3.1. *Let $L : A \to B$ be linear, and suppose we know the value of L on a set of vectors $X_1, X_2, \dots, X_n \in A$. That is,*

$$\begin{aligned} L(X_1) &= Y_1 \\ L(X_2) &= Y_2 \\ &\vdots \\ L(X_n) &= Y_n \end{aligned}$$

for some $Y_i \in B$.

Then the value of L is known on any linear combination of $X_1, X_2, \dots, X_n$.

The proof of this theorem is a straightforward application of properties 1 and 2 of linear functions. You might try to generate the justification yourself.

The span of a set of vectors

[TODO]

Linear independence

[TODO]

A basis for a vector space

[TODO]

3.3 Subspaces

[TODO]

3.4 Applying linear functions using basis-representation

[TODO]

3.5 Other bases

[TODO]

3.6 Dimension

[TODO]

References

Axler, Sheldon. 2024. *Linear Algebra Done Right*. Undergraduate Texts in Mathematics. Springer, Cham. https://doi.org/10.1007/978-3-031-41026-0.

Dummit, David S., and Richard M. Foote. 2004. *Abstract Algebra*. Third. John Wiley & Sons, Inc., Hoboken, NJ.