Linearity

Adam Layne

2025-01-14

Table of contents

What are these notes?

These are notes for a first course in Linear Algebra.

The PDF version of these notes can be fount at https://fnbu.pw/linearity-book/Linearity.pdf.

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.

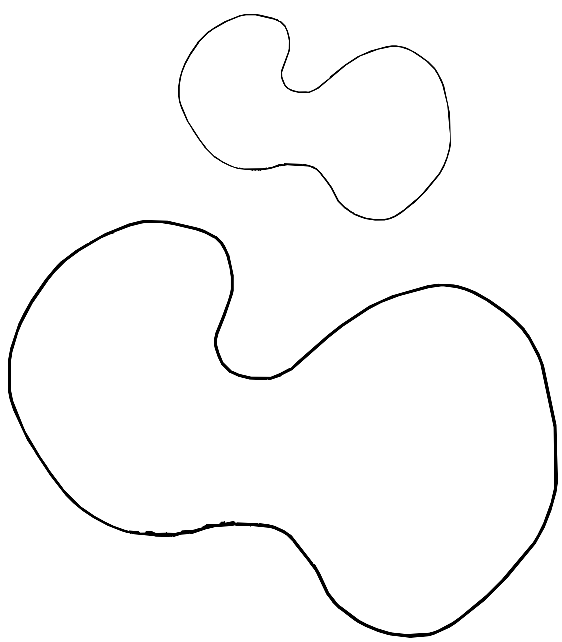

Figure 1.1: Scaling a curve by a factor of 2

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

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

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 $b = 0$.

References

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