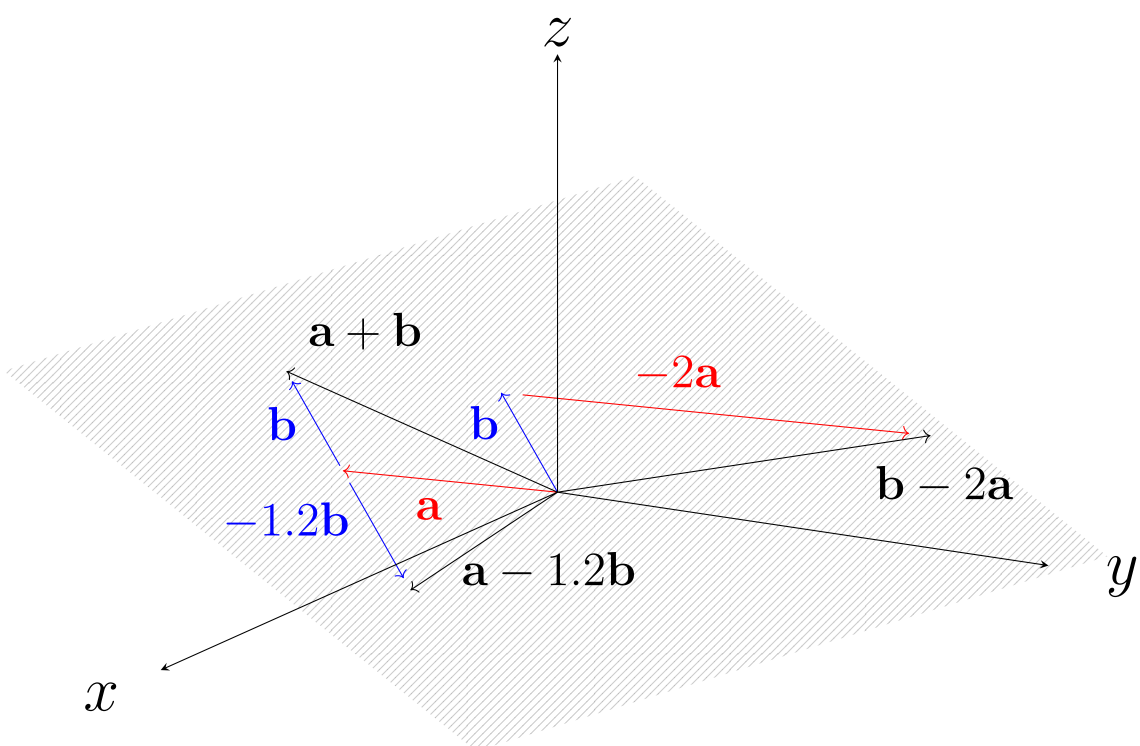
Incomplete Notes On Linear Algebra

Andrew Lehmann

Last update: October 26, 2023



Typeset in LaTeX 2_ε.

Contents

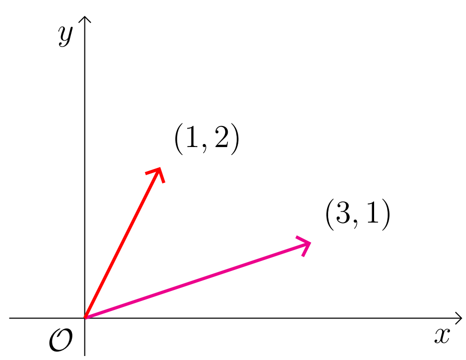
Chapter 1

Euclidean Vectors

1.1 Basic definitions

Definition 1.1: Euclidean vector or tuple.
A Euclidean vector is a list of n real numbers, also called an n-tuple. We write this list in parentheses, for example $(1, 3, -2, \dots, 0)$, and we say that this object belongs to $\mathbb{R}^n$. An arbitrary tuple can be written $\mathbf{v} = (v_1, v_2, \cdots, v_n)$ where the *components* $v_i \in \mathbb{R}$ for any index i.

It will be particularly useful to build intuition on vectors in $\mathbb{R}^2$. These are pairs of numbers like $(1, 2)$, $(-1, \pi)$ etc. We represent these vectors on a cartesian plane by associating an arrow pointing from the origin, $\mathcal{O} = (0, 0)$, to the given pair of numbers (x, y) as below



Arrow representation of two euclidean vectors.

The components of euclidean vectors represented this way are generally called the x and y components (there will be a z component for triples), and so we often write $\mathbf{v} = (v_x, v_y)$. For obvious reasons we also call these 2 dimensional vectors.

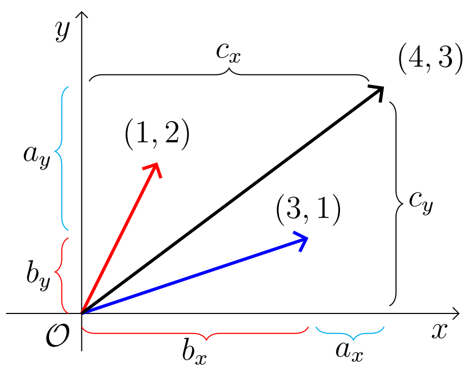
Now we would like to create a way to add vectors together. When we add two vectors together the result should also be a vector.

Definition 1.2: Tuple addition.
Euclidean vectors are added to each other component by component. In symbols

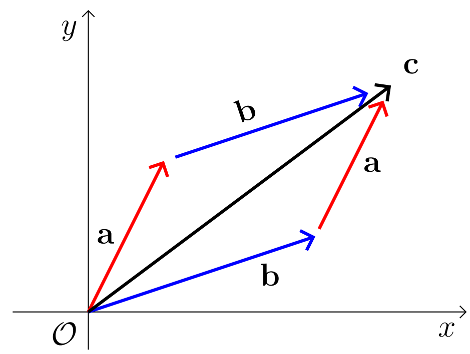
$$(a_1, a_2, \ldots, a_n) + (b_1, b_2, \ldots, b_n) = (a_1 + b_1, a_2 + b_2, \ldots, a_n + b_n).$$

Note: this means you can only add two tuples together *of the same size*. It makes no sense to add a 3-tuple to a 5-tuple.

Let's understand this tuple addition geometrically. For the two vectors we represented above, $\mathbf{a} = (1, 2)$, $\mathbf{b} = (3, 1)$ and so their addition is the vector $\mathbf{c} = \mathbf{a} + \mathbf{b} = (1 + 3, 2 + 1) = (4, 3)$. This component by component addition is represented below left. We can also understand this addition as though we move the vector $\mathbf{b}$ so that its tail is at the tip of $\mathbf{a}$. This tip-to-tail method is shown below to the right. It equally works by moving $\mathbf{a}$ to the tip of $\mathbf{b}$ meaning that $\mathbf{a} + \mathbf{b} = \mathbf{b} + \mathbf{a}$.



Component by component addition.

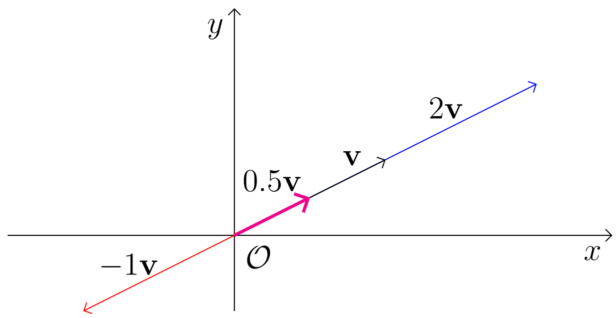


Tip-to-tail representation.

Now lets think about multiplcation. We can naturally think of doubling or tripling a vector, which should be adding a vector to itself 2 or 3 times, and geometrically it just becomes longer. So we want $2\mathbf{v} = \mathbf{v} + \mathbf{v}$. In components this gives $2(v_x, v_y) = (v_x, v_y) + (v_x, v_y) = (2v_x, 2v_y)$. So we see that multiplying a tuple by a number multiplies each of its components by that number.

Definition 1.3: Scalar multiplication.
Let $c \in \mathbb{R}$, called a scalar quantity, and $\mathbf{v} \in \mathbb{R}^n$ with components v_i. Then the *scalar multiplication* $c\mathbf{v}$ gives a vector $\mathbf{w}$ with components $w_i = cv_i$ for every index i. In tuple form

$$c(v_1, v_2, \ldots, v_n) = (cv_1, cv_2, \ldots, cv_n).$$



Arrow representation of scalar multiplication.

We can use scalar multiplication and some manipulation to form a second way to represent Euclidean vectors:

$$\begin{aligned}(v_1, v_2, \ldots, v_n) &= (v_1, 0, \ldots, 0) + (0, v_2, \ldots, 0) + \cdots + (0, 0, \ldots, v_n)\\ &= v_1(1, 0, \ldots, 0) + v_2(0, 1, \ldots, 0) + \cdots + v_n(0, 0, \ldots, 1)\\ &= v_1\mathbf{e}_1 + v_2\mathbf{e}_2 + \cdots + v_n\mathbf{e}_n\end{aligned}$$

where we have introduced the canonical Euclidean vectors.

Definition 1.4: Canonical Euclidean unit vectors.

The canonical Euclidean vectors in $\mathbb{R}^n$ are the n vectors of the form

$$\begin{aligned}\mathbf{e}_1 &= (1, 0, \ldots, 0)\\ \mathbf{e}_2 &= (0, 1, \ldots, 0)\\ &\vdots\\ \mathbf{e}_n &= (0, 0, \ldots, 1).\end{aligned}$$

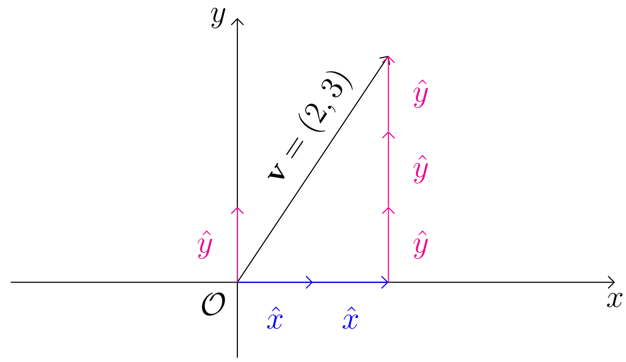
More compactly

$$\mathbf{e}_k = (\alpha_1, \alpha_2, \ldots, \alpha_n) \quad \text{where} \quad \alpha_j = \begin{cases} 1 & \text{for } j = k,\\ 0 & \text{for } j \neq k.\end{cases}$$

In two dimensions we have several notations for these unit vectors

$$\mathbf{v} = v_x\hat{x} + v_y\hat{y} = v_x\mathbf{e}_x + v_y\mathbf{e}_y = v_x\hat{\imath} + v_y\hat{\jmath}$$

but we will stick to the hat notation, $\hat{x}$, in this text when referring to the geometric representations in 2 or 3d ($\hat{z} = \mathbf{e}_z = \hat{k}$ being used for the third direction). We can understand these unit vectors as the building blocks of the arrow:



Unit vectors as building blocks of a Euclidean vector.

1.2 Dot product

Notice that with scalar multiplication we multiplied two different kinds of objects together, a scalar by a vector resulting in a vector. How about multiplying two vectors together? You might be tempted to simply form a new vector with component by component multiplication, as we did for addition. This turns out not to be so useful. One very useful way to define vector multiplication is as follows.

Definition 1.5: Dot product.

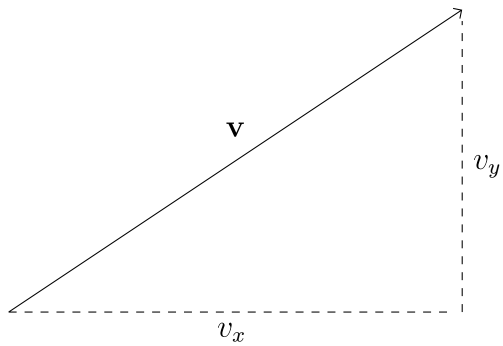
For two n-tuples $\mathbf{a}$ and $\mathbf{b}$, their *dot product*, also called *scalar* product and *Euclidean inner* product, is the real number given by the addition of component by component multiplication

$$\mathbf{a} \cdot \mathbf{b} = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n = \sum_{i=1}^{n} a_i b_i .$$

Notice that the result of this product is a real number, not another tuple. Let's first try to understand the dot product by considering a 2d vector "dotted" with itself:

$$\mathbf{v} \cdot \mathbf{v} = (v_x, v_y) \cdot (v_x, v_y) = v_x^2 + v_y^2 .$$

This dot product gives the square of the length of the right triangle made from the x and y components of $\mathbf{v}$:



Components of the vector $\mathbf{v}$

So we can identify the dot product $\mathbf{v} \cdot \mathbf{v}$ with the square of the length of $\mathbf{v}$. We generalise this concept of vector length in the operation of the norm.

Definition 1.6: Norm.
The norm of an n-tuple $\mathbf{v}$, denoted $\|\mathbf{v}\|$, is given by

$$\|\mathbf{v}\| = \sqrt{\mathbf{v} \cdot \mathbf{v}} = \sqrt{v_1^2 + v_2^2 + \cdots + v_2^2}.$$

So far we have defined a vector by its components. Through the norm, we can have a method of *calculating* the components from 2 other pieces of information: its length and its direction. Its direction will mean its angle with respect to the x-axis. Denote that angle θ and the norm as the non-bold font version of the vector $v = \|\mathbf{v}\|$. From the above figure, we see that $\cos\theta = v_x/v$ and $\sin\theta = v_y/v$. So, given a vector length and its direction we have

$$\mathbf{v} = (v\cos\theta, v\sin\theta) = v(\cos\theta, \sin\theta).$$

This gives us the following result: for any 2d Euclidean vector $\mathbf{v}$ its components are given by

$$v_x = \|\mathbf{v}\| \cos\theta$$
$$v_y = \|\mathbf{v}\| \sin\theta$$

where θ is the angle that $\mathbf{v}$ makes with the x axis.

What happens if we dot two different vectors with this representation. Let $\mathbf{a} = a(\cos\alpha, \sin\alpha)$ and

$\mathbf{b} = b(\cos\beta, \sin\beta)$. Then

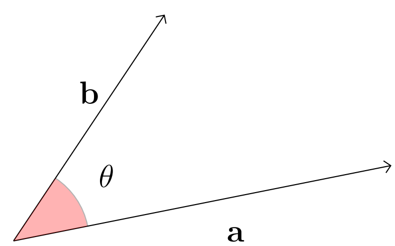
$$\begin{aligned}\mathbf{a}\cdot\mathbf{b} &= [a(\cos\alpha, \sin\alpha)]\cdot[b(\cos\beta,\sin\beta)]\\ &= ab(\cos\alpha,\sin\alpha)\cdot(\cos\beta,\sin\beta)\\ &= ab(\cos\alpha\cos\beta + \sin\alpha\sin\beta)\\ &= ab\cos(\alpha-\beta)\end{aligned}$$

where we have used a trigonometric identity in the last line and a little trickery in the first line that you should verify: $(c_1\mathbf{v}_1)\cdot(c_2\mathbf{v}_2) = (c_1c_2)\mathbf{v}_1\cdot\mathbf{v}_2$. In the end we have shown the following theorem.

Theorem 1.1: Dot product with angle.

The Euclidean dot product of two n-tuples $\mathbf{a}$ and $\mathbf{b}$ can be calculated solely from their norms and the angle between them, θ, given by the formula

$$\mathbf{a}\cdot\mathbf{b} = \|\mathbf{a}\|\|\mathbf{b}\|\cos\theta$$



The angle between two vectors.

From this expression of the dot product we see immediately what happens in the special case of perpendicular vectors. When two vectors are perpendicular the angle between them is $\pi/2$ and the dot product is then zero (recall that $\cos(\pi/2) = 0$). We can extend this result to define orthogonality for higher dimensional Euclidean vectors.

Definition 1.7: Orthogonal Euclidean vectors.

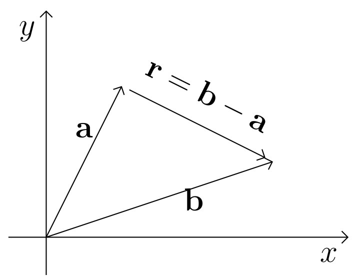
Two vectors in $\mathbb{R}^n$ are orthogonal if and only if their dot product equals zero.

1.3 Applications

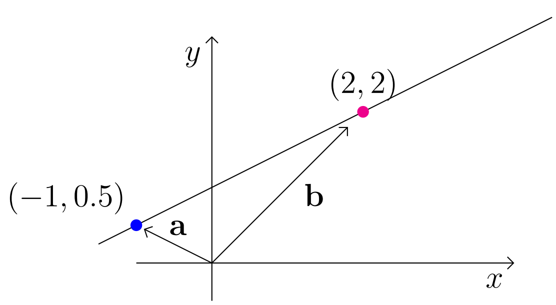
So far we have been considering Euclidean vectors as arrows pointing from the origin to the given tuple. It can be useful to consider arrows pointing from an arbitrary point in space to another.

Definition 1.8: Displacement vector.

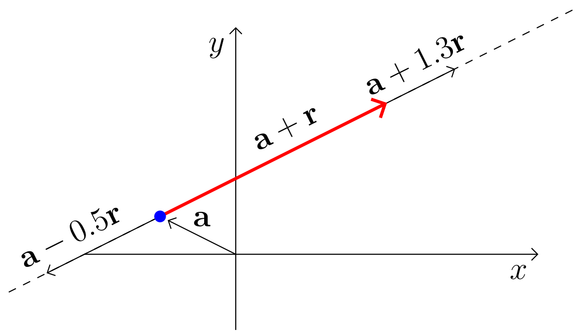
Given two Euclidean vectors $\mathbf{a}$ and $\mathbf{b}$, the displacement vector pointing from $\mathbf{a}$ to $\mathbf{b}$ is given by $\mathbf{r} = \mathbf{b} - \mathbf{a}$ as pictured below. Of course we can also create the displacement vector in the other direction, from $\mathbf{b}$ to $\mathbf{a}$, given by $\mathbf{a} - \mathbf{b}$.



Why is this useful? Well we can, for example, represent straight lines in a new way. Recall the equation for a straight line $y = mx + b$ where m gives the slope of the line and b gives the y-intercept (or vertical offset). Let's consider the line $y = (0.5)x + 1$. Now that means we can choose any two x values, say $x = -1$ and $x = 2$, and compute corresponding y values, in this case $y = 0.5$ and $y = 2$. So the Euclidean vectors $\mathbf{a} = (-1, 0.5)$ and $\mathbf{b} = (2, 2)$ belong to the line, as pictured.



The displacement vector from $\mathbf{a}$ to $\mathbf{b}$ is given by $\mathbf{r} = \mathbf{b} - \mathbf{a} = (2, 2) - (-1, 0.5) = (3, 1.5)$. This displacement vector can be used to represent any point on the line. We just have to start at $\mathbf{a}$ and add any multiple of $\mathbf{r}$:



So this gives as the vector representation of a line.

Definition 1.9: Vector form of a straight line.
The set of vectors in $\mathbb{R}^n$ of the form $\mathbf{v} = \mathbf{a} + t\mathbf{r}$ for a parameter $t \in \mathbb{R}$ represents a straight line through the space $\mathbb{R}^n$. That is,

$$\{(x, y) \,|\, \forall x \in \mathbb{R} \,\text{and}\, y = mx + b\} = \{\mathbf{a} + t\mathbf{r} \,|\, \forall t \in \mathbb{R}\}$$

where $\mathbf{a}$ is an arbitrary pair $(x, mx + b)$ and $\mathbf{r}$ is a displacement vector between any two distinct pairs $(x_1, mx_1 + b)$ and $(x_2, mx_2 + b)$.

To do

Unit vector

unit vector pointing in the direction of a given vector

Dot product as projection into another direction

1.4 Summary of properties of Euclidean vectors

Let's collect a list of key properties of operations on tuples that we will see, next chapter, are not unique to tuples. The following is true for any tuples $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c} \in \mathbb{R}^n$ and any real numbers k and $l \in \mathbb{R}$.

addition of tuples gives another tuple	$\mathbf{a}+\mathbf{b} \in \mathbb{R}^n$
tuple addition is associative	$(\mathbf{a}+\mathbf{b})+\mathbf{c}=\mathbf{a}+(\mathbf{b}+\mathbf{c})$
the zero tuple is a tuple of zeros	$\mathbf{a}+\mathbf{0}=\mathbf{0}+\mathbf{a}=\mathbf{a}$
the negative of a tuple is the additive inverse	$\mathbf{a}+(-\mathbf{a})=\mathbf{0}$
the order of tuple addition doesn't matter	$\mathbf{a}+\mathbf{b}=\mathbf{b}+\mathbf{a}$
scalar multiplication of a tuple gives another tuple	$k\mathbf{a} \in \mathbb{R}^n$
scalar multiplication distributes of tuple addition	$k(\mathbf{a}+\mathbf{b})=k\mathbf{a}+k\mathbf{b}$
scalar addition distributes over tuples	$(k+l)\mathbf{a}=k\mathbf{a}+l\mathbf{a}$
scalar multiplication order doesn't matter	$k(l\mathbf{a})=(kl)\mathbf{a}$
scalar multiplication by 1 is an identity operation	$1\mathbf{a}=\mathbf{a}$

Chapter 2

Vector Spaces

2.1 Set of polynomials of degree up to n

Recall that the degree of a polynomial with variable x is the power of the highest power of x amongst all of the terms with non-zero coefficient. Let's consider all of the polynomials with degree up to and including n with only real coefficients, denoted $P_n(\mathbb{R})$. An arbitrary member of this set can be written

$$\mathbf{p} = p_0 + p_1 x + p_2 x^2 + \cdots + p_n x^n = \sum_{i=0}^{n} p_i x^i$$

where the coefficients $p_i \in \mathbb{R}$. Let's see what happens when we add another polynomial in the same set, say $\mathbf{q} \in P_n(\mathbb{R})$:

$$\begin{aligned} \mathbf{p} + \mathbf{q} &= \sum_{i=0}^{n} p_i x^i + \sum_{i=0}^{n} q_i x^i \\ &= \sum_{i=0}^{n} (p_i + q_i) x^i \\ &= \mathbf{r}. \end{aligned}$$

We see the result is another polynomial of degree up to n (its degree with be the maximum of the degrees of $\mathbf{p}$ and $\mathbf{q}$). That means $\mathbf{r} \in P_n(\mathbb{R})$ and its coefficients are given by $r_i = p_i + q_i$.

Now what happens if we multiply a polynomial $\mathbf{p}$ by a real number $c \in \mathbb{R}$.

$$\begin{aligned} c\mathbf{p} &= c \sum_{i=0}^{n} p_i x^i \\ &= \sum_{i=0}^{n} (cp_i) x^i \\ &= \mathbf{r}. \end{aligned}$$

We see the result is another polynomial of degree up to n. That means $\mathbf{r} \in P_n(\mathbb{R})$ and its coefficients are given by $r_i = cp_i$. So just like Euclidean vectors we have addition of two objects resulting in an object of the same type, and scalar multiplication resulting in an object of the same type.

Let's prove that these polynomials also satisfy one of the other properties in the summary list of Section 1.4, for example the distributivity of scalar multiplication. For polynomials $\mathbf{p}$ and $\mathbf{q} \in P_n(\mathbb{R})$ and a real number $c \in \mathbb{R}$ we have

$$\begin{aligned} c(\mathbf{p} + \mathbf{q}) &= c \left(\sum_{i=0}^{n} p_i x^i + \sum_{i=0}^{n} q_i x^i \right) \\ &= c \sum_{i=0}^{n} p_i x^i + c \sum_{i=0}^{n} q_i x^i \\ &= c\mathbf{p} + c\mathbf{q}. \end{aligned}$$

So we see that polynomials also satisfy this distributivity property. It would be a good idea to convince yourself that they also satisfy all of the other properties in the list of Section 1.4.

2.2 Set of functions continuous on an interval

Let's recall the definition of continuity of a function.

Definition 2.1: Continuity of a function.

A function $f : A \to B$ is continuous at a point $c \in A$ if it satisfies the following limit

$$\lim_{x \to c} f(x) = f(c).$$

Then the function is continuous on an interval $[a, b]$ if it is continuous at all points in the interval. That is

$$\forall c \in [a, b] \quad \lim_{x \to c} f(x) = f(c).$$

Now let's consider the set of all functions continuous on the interval $[0,1]$, denoted $\mathcal{C}([0,1])$. Given two functions f and $g \in \mathcal{C}([0,1])$, lets define their addition as a third function h such that

$$\forall x \in [0,1] \quad h(x) = f(x) + g(x).$$

Is this function also continuous on $[0,1]$? Well let's see, we have

$$\forall c \in [0,1] \quad \lim_{x\to c} f(x) = f(c) \quad \text{and} \quad \lim_{x\to c} g(x) = g(c).$$

Now consider the same limit for h

$$\begin{aligned}\forall c \in [0,1] \quad \lim_{x\to c} h(x) &= \lim_{x\to c} (f(x) + g(x)) \\ &= \lim_{x\to c} f(x) + \lim_{x\to c} g(x) \\ &= f(c) + g(c) \\ &= h(c).\end{aligned}$$

So $h \in \mathcal{C}([0,1])$. What about scalar multiplication? For any $k \in \mathbb{R}$ we have

$$\lim_{x\to c} kf(x) = k \lim_{x\to c} f(x) = kf(c)$$

so that $kf \in \mathcal{C}([0,1])$.

Just like Euclidean vectors and polynomials, we have that addition and scalar multiplication remains within the set of objects. Let's prove that these continuous functions also satisfy one of the other properties in the summary list of Section 1.4, for example that there exists an additive inverse.

Let $f \in \mathcal{C}([0,1])$. Define h as the function

$$\forall x \in [0,1] \quad h(x) = -f(x).$$

This obviously satisfies the definition of an additive inverse: $f + h = 0$. Now let's show that h is continuous on the interval.

$$\begin{aligned}\forall c \in [0,1] \quad \lim_{x\to c} h(x) &= \lim_{x\to c} (-f(x)) \\ &= -\lim_{x\to c} f(x) \\ &= -f(c) \\ &= h(c).\end{aligned}$$

Hence $h \in \mathcal{C}([0,1])$. So every function continuous on $[0,1]$ has an additive inverse function which is also continuous on $[0, 1]$. It would be a good idea to convince yourself that they also satisfy all of the other properties in the list of Section 1.4.

2.3 Vector space axioms and properties

So now you should have had enough examples of sets of objects that seem to satisfy the same properties. We abstract away from these particular sets, Euclidean vectors, polynomials, functions, to talk about the properties themselves and the relations between any mathematical objects that satisfy these properties. Then if we look at a new set of objects and we recognise these properties, all of our results will automatically apply. The name for this abstracted algebraic structure is the vector space. Though the power of linear algebra is in this abstraction, we will often try to concretely understand a result by referring back to a particular vector space. For the most part I will use 2d Euclidean vectors to illustrate results.

Definition 2.2: Vector space.

A *vector space over a field* $\mathbb{F}$ is a set, call it V, with elements called vectors supplied with definitions of two operations, *vector addition* (VA) and *scalar multiplication* (SM), that satisfy the following *vector space axioms*:

$$\forall\, \mathbf{u}, \mathbf{v}, \mathbf{w} \in V \quad \text{and} \quad \forall\, k, l \in \mathbb{F}$$

(VA1)	$\mathbf{u} + \mathbf{v} \in V$	(closure under vector addition)
(VA2)	$(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$	(associativity of vector addition)
(VA3)	$\exists\, \mathbf{0} \in V$, such that $\mathbf{u} + \mathbf{0} = \mathbf{0} + \mathbf{u} = \mathbf{u}$	(additive identity)
(VA4)	$\exists\, -\mathbf{u} \in V$ such that $\mathbf{u} + (-\mathbf{u}) = \mathbf{0}$	(additive inverse)
(VA5)	$\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$	(commutativity of vector addition)
(SM1)	$k\mathbf{u} \in V$	(closure under scalar multiplication)
(SM2)	$k(\mathbf{u} + \mathbf{v}) = k\mathbf{u} + k\mathbf{v}$	(distributivity over vector addition)
(SM3)	$(k + l)\mathbf{u} = k\mathbf{u} + l\mathbf{u}$	(distributivity over field addition)
(SM4)	$k(l\mathbf{u}) = (kl)\mathbf{u}$	(compatibility of scalar and field multiplication)
(SM5)	$1\mathbf{u} = \mathbf{u}$	(multiplicative identity)

For now we will restrict ourselves to *real* vector spaces by assuming the field is the real numbers, $\mathbb{F} = \mathbb{R}$.

Theorem 2.1: Zero vector.
If we take the zero from the field, $0 \in \mathbb{R}$, and multiply it by any vector, $\mathbf{u} \in V$, then we get the zero vector $\mathbf{0}_V \in V$.

Proof. Let's use axiom SM3 by choosing $k = 0$ and keeping the other terms arbitrary. This then says

$$(0 + l)\mathbf{u} = 0\mathbf{u} + l\mathbf{u}.$$

But we know for real numbers that $0 + l = l$. Hence we have the equation

$$l\mathbf{u} = 0\mathbf{u} + l\mathbf{u}$$

which is exactly the form of axiom VA3 which defines the zero vector. Hence we have

$$0\mathbf{u} = \mathbf{0}.$$

Often we distinguish between the zero *vector* and zero number (in the field) by using a subscript: $\mathbf{0}_V$ for the zero vector in the vector space V.

Example 2.1: Euclidean line vector space.
The set of all real tuples (x, y) satisfying $y = 3x$ forms a vector space, call it V. For example, if $\mathbf{u} = (u_x, u_y)$ and $\mathbf{v} = (v_x, v_y)$ are two vectors of V, then $u_y = 3u_x$, $v_y = 3v_x$ and their addition $\mathbf{w} = \mathbf{u} + \mathbf{v}$ is a tuple

$$\begin{aligned}(w_x, w_y) &= (u_x, u_y) + (v_x, v_y) \\ &= (u_x + v_x, 3u_x + 3v_y) \\ &= (u_x + v_x, 3(u_x + v_y)).\end{aligned}$$

Thus the vector $\mathbf{w}$ has a y component that is 3 times its x component, i.e. $w_y = 3w_x$, and so it is also a vector in V. That proves the vector space axiom (VA1), the closer under vector addition. The other 9 axioms also hold and it is a good exercise to prove that. We can write this vector space in the form of a set

$$V = \{(x, y) \in \mathbb{R}^2 \,|\, 3x - y = 0\}.$$

This example showed that even a subset of a vector space (V was a subset of $\mathbb{R}^2$) can also satisfy the vector space axioms.

Definition 2.3: Vector subspace.
Suppose that V is a vector space and W is a subset of V. We call W a *vector subspace* if it satisfies the vector space axioms for the same definition of vector addition and scalar multiplication defined for V.

In practice it can be tedious to show all 10 axioms hold for the subset. However, many of the properties are automatically inherited from the known vector space. For example any subset of vectors will obviously satisfy commutativity and associativity. In the end it suffices to prove just 3 properties for the candidate subspace.

Theorem 2.2: Demonstration of a vector subspace.
Let W be a subset of a vector space V. W is a vector subspace if and only if

1. W is a non-empty set,
2. W is closed under vector addition: $\mathbf{u}, \mathbf{v} \in W \implies \mathbf{u} + \mathbf{v} \in W$,
3. W is closed under scalar multiplication: $\mathbf{u} \in W$ and $k \in \mathbb{R} \implies k\mathbf{u} \in W$.

In fact we can do even better, the closure under vector addition and scalar multiplication can be proven in just 1 step:

$$\forall \mathbf{u}, \mathbf{v} \in W \text{ and } \forall k \in \mathbb{R}$$
$$k\mathbf{u} + \mathbf{v} \in W.$$

Example 2.2: Plane as a vector subspace.
Consider a flat 2 dimensional plane as a set of triples within 3d space, (x, y, z), satisfying a plane equation $2x - y + z = 0$. The set of these points are a subset, call it W, of the Euclidean vector space $\mathbb{R}^3$. In set notation we can write

$$W = \{(x, y, z) \in \mathbb{R}^3 \,|\, 2x - y + z = 0\}.$$

Let's show that W is a vector subspace of $\mathbb{R}^3$. Firstly, the vector $(0, 0, 0) \in W$ because its components satisfy the plane equation. Hence W has at least one vector, that is, it is a non-empty set. Let $\mathbf{u} = (u_x, u_y, u_z)$ and $\mathbf{v} = (v_x, v_y, v_z)$ be two arbitrary vectors in W. That means their components satisfy the plane equation, i.e. $2u_x - u_y + u_z = 0$ and $2v_x - v_y + v_z = 0$. For any real

k we therefore have

$$k\mathbf{u} + \mathbf{v} = (ku_x + v_x, ku_y + v_y, ku_z + v_z) = (w_x, w_y, w_z).$$

We need to check whether the components of this resultant vector satisfies the plane equation governing W.

$$\begin{aligned} 2w_x - w_y + w_z &= 2(ku_x + v_x) - (ku_y + v_y) + ku_z + v_z \\ &= k(\underbrace{2u_x - u_y + u_z}_{=0}) + \underbrace{2v_x - v_y + v_z}_{=0}. \end{aligned}$$

Since the vector $k\mathbf{u}+\mathbf{v}$ has components satisfying the plane equation, we conclude that $k\mathbf{u}+\mathbf{v} \in W$. So W is a non-empty subset of a vector space and W is closed under vector addition and scalar multiplication. Thus W is a vector subspace.

In the previous example we *verified* that a space satisfied the vector space axioms. Now let's generate a vector space out of some given vectors. First we consider the most general expression of creating a new vector from some given vectors.

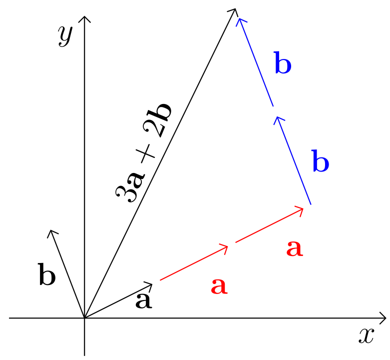
Definition 2.4: Linear Combination.

Let $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be a set of vectors in a vector space V. A linear combination of these vectors is a new vector, $\mathbf{w} \in V$, of the form

$$\mathbf{w} = \alpha_1\mathbf{v}_1 + \cdots + \alpha_n\mathbf{v}_n$$

where the α_k are real numbers.

In 2d space, taking linear combinations of two vectors $\mathbf{a}$ and $\mathbf{b}$ is like choosing a pair of directions as reference directions in pirate map explorations. Normally you would say "3 steps east, 2 steps north", but you could equally say "3 steps in direction $\mathbf{a}$, 2 steps in direction $\mathbf{b}$" as pictured below.



Now we can use this idea of linear combination to generate a whole set of vectors.

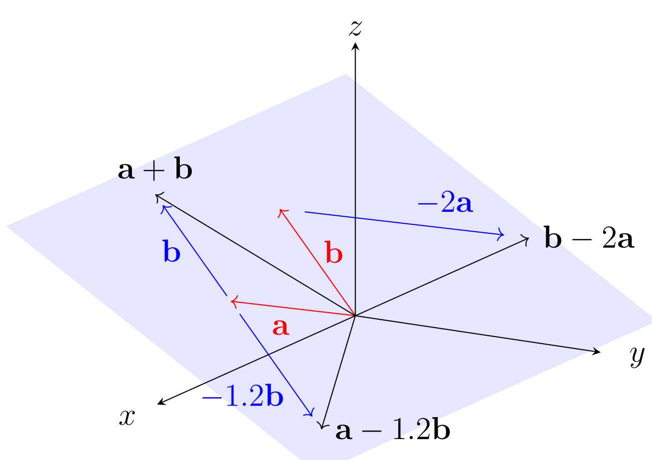
Definition 2.5: Span.

Let $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be a set of vectors from a vector space V. The span of these vectors is the set of all linear combinations of those vectors:

$$\mathrm{SPAN}(\mathcal{B}) = \mathrm{SPAN}(\mathbf{v}_1, \ldots, \mathbf{v}_n) = \{\alpha_1 \mathbf{v}_1 + \cdots + \alpha_n \mathbf{v}_n \,|\, \alpha_1, \ldots, \alpha_n \in \mathbb{R}^n\}\,.$$

This set forms a vector subspace of V. It is obviously non-empty because it at least contains the vectors of $\mathcal{B}$. It is also automatically closed under vector addition and scalar multiplication because those are exactly the operations we used to create all the vectors in the span! Therefore $\mathrm{SPAN}(\mathcal{B})$ is a vector subspace of V.

In 3d space, the span of any two vectors pointing in different directions will form a plane. In the picture below we see some linear combinations of $\mathbf{a}$ and $\mathbf{b}$. Hopefully you can convince yourself that no such linear combination could leave the blue plane.



Some selected linear combinations, the black arrows, of $\mathbf{a}$ and $\mathbf{b}$.

Example 2.3: Span of two 3d vectors.

Let $\mathbf{u} = (1, 1, 2)$ and $\mathbf{v} = (0, 3, 1)$ be two Euclidean vectors. Let $\mathrm{SPAN}(\mathbf{u}, \mathbf{v}) = V$. Does the vector $\mathbf{w} = (1, -5, 0)$ belong to V? Let's answer this question twice. First we'll find the equation of the plane that governs this span. By definition

$$V = \{\alpha\mathbf{u} + \beta\mathbf{v} \,|\, \forall \alpha, \beta \in \mathbb{R}\}.$$

So a generic vector (x, y, z) in V must satisfy

$$(x, y, z) = \alpha(1, 1, 2) + \beta(0, 3, 1).$$

Let's look for a single equation relating the x, y and z:

$$(x, y, z) = (\alpha, \alpha + 3\beta, 2\alpha + \beta) \implies \begin{cases} x = \alpha \\ y = \alpha + 3\beta \\ z = 2\alpha + \beta \end{cases} \implies \begin{cases} x = \alpha \\ y = x + 3\beta \\ z = 2x + \beta \end{cases} \implies \begin{cases} x = \alpha \\ y = x + 3\beta \\ y - 3z = -5x \end{cases}$$

This last line gives us the equation we seek, and hence V can be written

$$V = \{(x, y, z) \in \mathbb{R}^3 \,|\, 5x + y - 3z = 0\}.$$

With this equation we can easily check whether $\mathbf{w} = (1, -5, 0)$ belongs to V or not:

$$5w_x + w_y - 3w_z = 5(1) + (-5) - 3(0) = 0.$$

The equation is satisfied and so $\mathbf{w} \in V$.

The second method to check this is to find whether $\mathbf{w}$ really is a linear combination of $\mathbf{u}$ and $\mathbf{v}$ or not. Assume that it is:

$$\mathbf{w} = \alpha\mathbf{u} + \beta\mathbf{v}$$

for some α and β that we can find or else we will find a contradiction. We have assumed

$$\begin{aligned} (1, -5, 0) &= \alpha(1, 1, 2) + \beta(0, 3, 1) \\ &= (\alpha, \alpha + 3\beta, 2\alpha + \beta) \end{aligned}$$

giving the 3 equations

$$\alpha = 1, \quad \alpha + 3\beta = -5, \quad 2\alpha + \beta = 0$$

Putting the value of α into either the second or third gives the same result: $\beta = -2$. Hence $\mathbf{w}$ is a linear combination of $\mathbf{u}$ and $\mathbf{v}$

$$\mathbf{w} = \mathbf{u} - 2\mathbf{v}$$

and therefore $\mathbf{w} \in V$.

When we considered the span of two 3d vectors giving a plane I casually inserted an important qualification, that the two vectors *are not in the same direction*. If we have two vectors $\mathbf{u}$ and $\mathbf{v}$ in the same direction then, for example, $\mathbf{v} = k\mathbf{u}$ for some real k. Then any linear combination of these two vectors gives a third vector

$$\begin{aligned} \mathbf{w} &= \alpha\mathbf{u} + \beta\mathbf{v} \\ &= \alpha\mathbf{u} + \beta k\mathbf{u} \\ &= (\alpha + \beta k)\,\mathbf{u} \end{aligned}$$

that is, $\mathbf{w}$ must also be in the same direction as $\mathbf{u}$ and $\mathbf{v}$. The span of these 2 vectors gives only vectors in their single direction. In this case we say that $\mathbf{u}$ and $\mathbf{v}$ are not independent vectors. One can be represented in terms of the other. Let's formalise this notion of vector dependence.

Definition 2.6: Linear dependence.
A set of vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ from a vector space V is said to be *linearly dependent* if there exists a set of constants $\{\alpha_1, \ldots, \alpha_n\}$ *not all zero* such that

$$\alpha_1\mathbf{v}_1 + \cdots + \alpha_n\mathbf{v}_n = \mathbf{0}_V.$$

Note: the right hand side of the equation is the *zero vector*, not the real number 0.

How does this definition relate to what we understood earlier, that two vectors are dependent on each other if one can be expressed in terms of the other? Well, consider three vectors $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w} \in \mathbb{R}^3$. If there exists some constants α, β and γ, *not all zero*, such that

$$\alpha\mathbf{u} + \beta\mathbf{v} + \gamma\mathbf{w} = \mathbf{0}_V$$

then we can write the vector with the non-zero constant in terms of the other. For example suppose $\alpha \neq 0$, then

$$\mathbf{u} = -\frac{\beta}{\alpha}\mathbf{v} - \frac{\gamma}{\alpha}\mathbf{w}.$$

In this way $\mathbf{u}$ depends on the other two, or we could say $\mathbf{u} \in \text{SPAN}(\mathbf{v}, \mathbf{w})$. This also means that the span of the three vectors $\text{SPAN}(\mathbf{u}, \mathbf{v}, \mathbf{w}) = \text{SPAN}(\mathbf{v}, \mathbf{w})$. The vector $\mathbf{u}$ doesn't give anything new. In this particular case, we have a plane and adding the new vector gives linear combinations that remain in the plane. What would it mean for $\mathbf{u}$ to be independent of $\mathbf{v}$ and $\mathbf{w}$? The formal definition is simple:

Definition 2.7: Linear independence.

A set of vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ from V is said to be *linearly independent* if they are not linearly dependent. That is, the equation

$$\alpha_1 \mathbf{v}_1 + \cdots + \alpha_n \mathbf{v}_n = \mathbf{0}_V.$$

implies that the constants $\alpha_1, \ldots, \alpha_n$ *are all zero.*

This definition is a little cheeky. Independent means not dependent. Let's understand it in the sense we were thinking earlier, where we think a vector $\mathbf{u}$ being independent of two other vectors $\mathbf{v}$ and $\mathbf{w}$ should mean that we cannot express $\mathbf{u}$ as a linear combination of $\mathbf{v}$ and $\mathbf{w}$. This follows from the definition. If we could write such a linear combination, then there are constants α and β such that

$$\begin{aligned} \mathbf{u} &= \alpha\mathbf{u} + \beta\mathbf{v} \\ \implies \mathbf{u} - \alpha\mathbf{u} - \beta\mathbf{v} &= \mathbf{0}_V, \end{aligned}$$

but this equation is impossible as $\alpha_1\mathbf{u} + \alpha_2\mathbf{v} + \alpha_3\mathbf{w} = \mathbf{0}_V \implies \alpha_1 = \alpha_2 = \alpha_3 = 0$.

Theorem 2.3: The span of a dependent set of vectors can be reduced.

Let $\mathcal{B} = \{\mathbf{v}_1,\ \mathbf{v}_2,\ \ldots,\ \mathbf{v}_n\}$ be a set of vectors of V. If $\mathcal{B}$ is a set of linearly dependent vectors, then we can always remove one of the vectors to form a new set, $\mathcal{B}' = \mathcal{B}\backslash\{\mathbf{v}_k\}$ for some k, without changing the span: $\text{SPAN}(\mathcal{B}') = \text{SPAN}(\mathcal{B})$.

Proof. Due to the linear dependence of the vectors of $\mathcal{B}$, there exists a set of constants $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$ not all zero such that

$$\alpha_1 \mathbf{v}_1 + \cdots + \alpha_n \mathbf{v}_n = \mathbf{0}_V.$$

Suppose $\alpha_k \neq 0$. Then we can write

$$\mathbf{v}_k = -\frac{\alpha_1}{\alpha_k}\mathbf{v}_1 - \cdots - \frac{\alpha_{k-1}}{\alpha_k}\mathbf{v}_{k-1} - \frac{\alpha_{k+1}}{\alpha_k}\mathbf{v}_{k+1} - \cdots - \frac{\alpha_n}{\alpha_k}\mathbf{v}_n.$$

which is to say $\mathbf{v}_k \in \text{SPAN}(\mathbf{v}_1, \cdots, \mathbf{v}_{k-1}, \mathbf{v}_{k+1}, \mathbf{v}_n)$.... (need to prove $\mathbf{v} \in \text{SPAN}(\mathcal{B}) \implies \mathbf{v} \in \text{SPAN}(\mathcal{B}')$ and $\mathbf{v} \in \text{SPAN}(\mathcal{B}') \implies \mathbf{v} \in \text{SPAN}(\mathcal{B})$)

Now linear independence is a very important concept because it allows us to...

Definition 2.8: Basis.

A *basis of a vector space* V is a minimal set of vectors which spans the vector space. Formally, the set of vectors $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ in a vector space V is a basis of V if it is a set of linearly independent vectors and $\text{SPAN}(\mathbf{v}_1, \ldots, \mathbf{v}_n) = V$. *Note*: bases are not unique, but they always contain the same number of vectors.

Definition 2.9: Dimension.

The *dimension of a vector space* is the number of elements in a basis for that vector space.

Definition 2.10: Canonical basis of $\mathbb{R}^n$.

The *canonical basis* of the vector space of real n-tuples, $\mathbb{R}^n$, is the ordered set of n n-tuples with k^{th} element, $\mathbf{c}_k = (\alpha_1, \ldots, \alpha_n)$ such that

$$\alpha_j = \begin{cases} 1 & \text{for } j = k, \\ 0 & \text{for } j \neq k. \end{cases}$$

That is, as a set the canonical basis is

$$\mathcal{C}_n = \{(1,0,\ldots,0),\ (0,1,\ldots,0),\ \ldots,\ \underbrace{(0,0,\ldots,0,\ \overbrace{1}^{k^{th}\text{ place}}\ ,0,\ldots,0)}_{k^{th}\text{ tuple}},\ \ldots,\ (0,0,\ldots,1)\}.$$

Definition 2.11: Canonical basis of $\mathcal{P}_n$.

The *canonical basis* of the vector space of polynomials with degree up to n, $\mathcal{P}_n$, is the ordered set of n polynomials with k^{th} element, $\mathbf{c}_k = x^k$. That is, as a set the canonical basis is

$$\mathcal{C}_n = \{1,\ x,\ x^2,\ \ldots,\ x^n\}.$$

Definition 2.12: Coordinates of a vector.

Let $\mathbf{v}$ be a vector in a vector space V. The coordinates of $\mathbf{v}$ *with respect to a given basis* $\mathcal{B}$, denoted $[\mathbf{v}]_{\mathcal{B}}$, is a column of the unique set of coefficients in the linear combination of $\mathbf{v}$ in terms of the basis vectors.

Example 2.4: Coordinates of a Euclidean vector.

For the Euclidean vector

$$\mathbf{v} = (2,\, -1,\, 8) \in \mathbb{R}^3$$

(implicitly written in the canonical basis) and basis of $\mathbb{R}^3$

$$\mathcal{B} = \{\mathbf{b}_1,\, \mathbf{b}_2,\, \mathbf{b}_3\} = \{(2,\, 0,\, 0),\, (0,\, 1,\, -2),\, (0,\, 0,\, 2)\}$$

the vector $\mathbf{v}$ can be written in terms of the basis vectors as $\mathbf{v} = 1\mathbf{b}_1 - 2\mathbf{b}_2 + 2\mathbf{b}_3$ and hence its coordinates with respect to $\mathcal{B}$ are

$$[\mathbf{v}]_{\mathcal{B}} = \begin{pmatrix} 1 \\ -2 \\ 2 \end{pmatrix}$$

Example 2.5: Coordinates of a polynomial vector.

For a polynomial vector

$$\mathbf{v} = 3 - x + 2x^2 \in \mathcal{P}_2[\mathbb{R}]$$

and basis of $\mathcal{P}_2[\mathbb{R}]$

$$\mathcal{B} = \{\mathbf{b}_1,\, \mathbf{b}_2,\, \mathbf{b}_3\} = \left\{1 - x,\, 1 + x,\, x - x^2\right\}$$

we express the vector $\mathbf{v}$ as a linear combination of the basis vectors $\mathbf{v} = \alpha\mathbf{b}_1 + \beta\mathbf{b}_2 + \gamma\mathbf{b}_3$ and our goal is to determine the constants α, β and γ. Develop the linear combination

$$\begin{aligned} \mathbf{v} &= \alpha(1 - x) + \beta(1 + x) + \gamma(x - x^2) \\ 3 - x + 2x^2 &= (\alpha + \beta)1 + (-\alpha + \beta + \gamma)x + (-\gamma)x^2. \end{aligned}$$

By equating polynomial terms we get the system

$$\begin{cases} \alpha + \beta = 3 \\ -\alpha + \beta + \gamma = -1 \\ -\gamma = 2 \end{cases} \implies \begin{cases} \alpha + \beta = 3 \\ -\alpha + \beta = 1 \\ \gamma = -2 \end{cases} \implies \begin{cases} 2\alpha = 2 \\ 2\beta = 4 \\ \gamma = -2 \end{cases}$$

Hence $3 - x + 2x^2 = \mathbf{b}_1 + 2\mathbf{b}_2 - 2\mathbf{b}_3$ so that the coordinates of the vector are

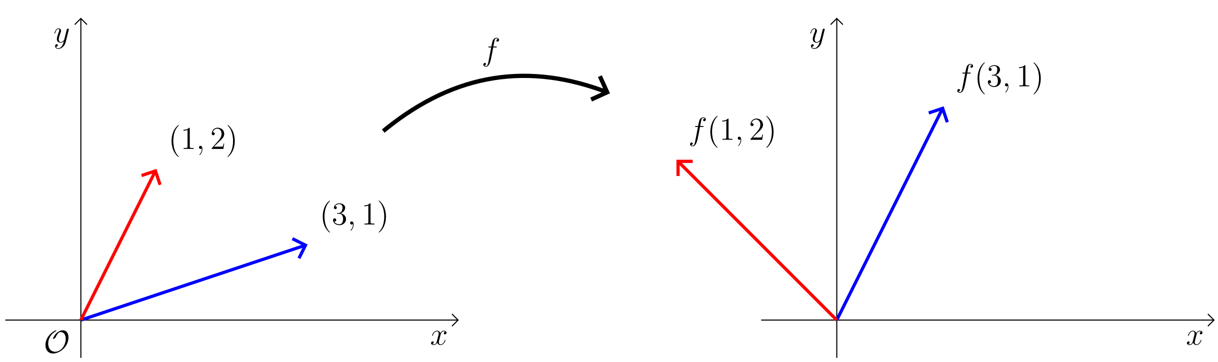
$$[\mathbf{v}]_{\mathcal{B}} = \begin{pmatrix} 1 \\ 2 \\ -2 \end{pmatrix}$$

Chapter 3

Linear Maps

3.1 Basic properties

Let's build the concept of a function, call it f, taking vectors as inputs and giving vectors as outputs. We therefore write $f : V \to W$ for vector spaces V and W. For example, imagine a function that takes any Euclidean vector in $\mathbb{R}^2$ and rotates it by 45° anti-clockwise, keeping the length of the vector fixed. We could write $f(\mathbf{v}) = \mathbf{w}$ for $\mathbf{v}, \mathbf{w} \in \mathbb{R}^2$. Below we sketch a couple of examples.



For an arbitrary vector $\mathbf{v} = (x, y)$, f maps $\mathbf{v}$ to the coordinates $(x/\sqrt{2} - y/\sqrt{2},\, x/\sqrt{2} + y/\sqrt{2})$. We then ask the question, if we have two arbitrary vectors $\mathbf{v} = (x, y)$ and $\mathbf{u} = (s, t)$ which add to the third vector $\mathbf{w} = \mathbf{v} + \mathbf{u}$, where does f map the addition $\mathbf{w}$ to? Do we get the same answer as if we first rotate $\mathbf{v}$ and $\mathbf{u}$ and then add the rotated vectors together? Let's see

$$\begin{aligned} f(\mathbf{v}+\mathbf{u}) &= f(x+s,\, y+t) \\ &= \frac{1}{\sqrt{2}}(x+s-y-t,\, x+s+y+t) \\ &= \frac{1}{\sqrt{2}}(x-y,\, x+y) + \frac{1}{\sqrt{2}}(s-t,\, s+t) \\ &= f(x,y) + f(s,t) \\ &= f(\mathbf{v}) + f(\mathbf{u}) \end{aligned}$$

We have indeed that we can rotate $\mathbf{v}$ and $\mathbf{u}$ and then add up the result, or we can add $\mathbf{v}$ and $\mathbf{u}$ and then rotate the result to get the same outcome.

This lack of importance in the order of the application of the function is not necessarily true for any function we could think of. Consider a function, g, that takes a vector in $\mathbb{R}^2$ and gives another vector in the same direction with length equal to the square of the original vector's length. This would be represented by

$$g(x,y) = \sqrt{x^2+y^2}(x,y).$$

Now, for example, take two vectors $\mathbf{v} = (2,0)$ and $\mathbf{u} = (0,2)$. We have $g(\mathbf{v}) = (4,0)$, $g(\mathbf{u}) = (0,4)$ and $g(\mathbf{v}+\mathbf{u}) = g(2,2) = (4\sqrt{2}, 4\sqrt{2})$. So for this function, we have $g(\mathbf{v}+\mathbf{u}) \neq g(\mathbf{v}) + g(\mathbf{u})$. The order matters. In linear algebra we study functions of the first type and not the second. These functions are called *linear maps*, defined below:

Definition 3.1: Linear map.

A mapping, f, from a vector space V to a vector space W, denoted $f : V \to W$, is called a *linear map* if it satisfies the following property:

$$\begin{aligned} &\forall \mathbf{u},\, \mathbf{v} \in V,\, \forall \alpha, \beta \in \mathbb{R} \\ &f(\alpha\mathbf{u} + \beta\mathbf{v}) = \alpha f(\mathbf{u}) + \beta f(\mathbf{v}). \end{aligned}$$

We say that a linear map *preserves linear combinations.*

Example 3.6: Differentiation as a linear map.

Let's define a mapping $f : \mathcal{P}_n \to \mathcal{P}_{n-1}$ that takes a polynomial of degree up to n (a member of the vector space of polynomials of degree up to n) and differentiates it. For example

$$
\begin{aligned}
f(1+3x^2) &= 6x \\
f(3) &= 0 \\
f(2x+x^2) &= 2+2x
\end{aligned}
$$

and you get the idea. Consider two arbitrary vectors

$$
\begin{aligned}
\mathbf{u} &= \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \cdots + \alpha_n x^n \\
\mathbf{v} &= \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_n x^n.
\end{aligned}
$$

We consider, separately, the action of f on these vectors and on their addition

$$
\begin{aligned}
f(\mathbf{u}) &= \alpha_1 + 2\alpha_2 x + \cdots + n\alpha_n x^{n-1} \\
f(\mathbf{v}) &= \beta_1 + 2\beta_2 x + \cdots + n\beta_n x^{n-1} \\
f(\mathbf{u}+\mathbf{v}) &= f\left((\alpha_0+\beta_0) + (\alpha_1+\beta_1)x + (\alpha_2+\beta_2)x^2 + \cdots + (\alpha_n+\beta_n)x^n\right) \\
&= (\alpha_1+\beta_1) + 2(\alpha_2+\beta_2)x + \cdots + n(\alpha_n+\beta_n)x^{n-1}
\end{aligned}
$$

This last expression can be split by collecting alphas and betas

$$
(\alpha_1 + 2\alpha_2 x + \cdots + n\alpha_n x^{n-1}) + (\beta_1 + 2\beta_2 x + \cdots + n\beta_n x^{n-1}) = f(\mathbf{u}) + f(\mathbf{v})
$$

showing that the derivative of the addition is the addition of the derivatives. Hence we can consider differentiation of polynomials as a linear map.

Theorem 3.1: A linear map conserves the zero vector.

For any linear map, $f : V \to W$, we have

$$
f(\mathbf{0}_V) = \mathbf{0}_W
$$

where $\mathbf{0}_V$ is the zero vector of V and $\mathbf{0}_W$ is the zero vector of W.

Proof. By the definition of a linear map, we can choose $\alpha = \beta = 0 \in \mathbb{R}$ and we must have for any $\mathbf{u}$, $\mathbf{v} \in V$ the following

$$
f(0\mathbf{u} + 0\mathbf{v}) = 0f(\mathbf{u}) + 0f(\mathbf{v}).
$$

We proved that the number 0 multiplied by any vector gives the zero vector in that space. Hence

$$0\mathbf{u} + 0\mathbf{v} = \mathbf{0}_V + \mathbf{0}_V = \mathbf{0}_V \quad \text{and} \quad 0f(\mathbf{u}) + 0f(\mathbf{v}) = \mathbf{0}_W + \mathbf{0}_W = \mathbf{0}_W$$

and so we have proved

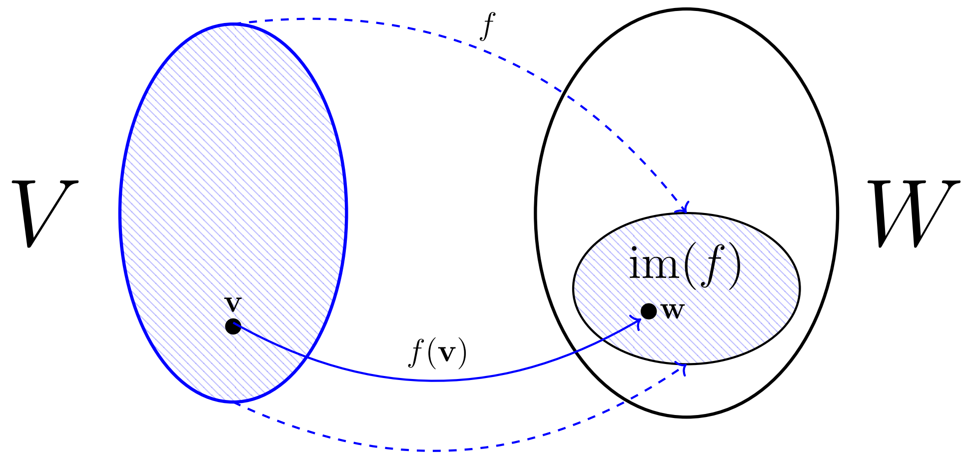
$$f(\mathbf{0}_V) = \mathbf{0}_W.$$

Definition 3.2: Image.

The *image of a linear map* $f : V \to W$, denoted $\operatorname{im}(f)$, is the set of all possible "output" vectors of the map:

$$\operatorname{im}(f) = \{\mathbf{w} \in W \mid \exists \mathbf{v} \in V \; f(\mathbf{v}) = \mathbf{w}\} \subseteq W.$$

This can be understood pictorially as so:



The image is a vector subspace of W. Let's prove this. Firstly, we previously showed that any linear map takes the zero vector of V to the zero vector of W. So $\mathbf{0}_W \in \operatorname{im}(f)$ and hence it is not an empty set. Let $\mathbf{u}$, $\mathbf{v} \in \operatorname{im}(f)$ and α, $\beta \in \mathbb{R}$. There must exist corresponding vectors in V, $\mathbf{u}'$ and $\mathbf{v}'$ such that $f(\mathbf{u}') = \mathbf{u}$ and $f(\mathbf{v}') = \mathbf{v}$. Hence

$$\alpha\mathbf{u} + \beta\mathbf{v} = \alpha f(\mathbf{u}') + \beta f(\mathbf{v}') = f(\alpha\mathbf{u}' + \beta\mathbf{v}')$$

where we have used the definition of a linear map in the last step. This shows that for any linear combination of vectors in the image of f, we can find a corresponding vector in V. This means the $\alpha\mathbf{u} + \beta\mathbf{v} \in \operatorname{im}(f)$. Hence the image is a non-empty subset of W that is closed under linear combinations, that is, it is a vector subspace of W.

Theorem 3.2: Generator of the image.
The image of any linear map, $f : V \to W$, has a generator set that can be found by the action of f on any basis of V. That is, let V be an n dimensional vector space and $\mathcal{B}_V = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ be a basis of V. Then the set $\mathcal{C} = \{f(\mathbf{b}_1), \dots, f(\mathbf{b}_n)\} \subset W$ generates $\operatorname{im}(f)$, that is,

$$\operatorname{im}(f) = \operatorname{SPAN}\left(f(\mathbf{b}_1), \dots, f(\mathbf{b}_n)\right)$$

Proof. Let $\mathbf{w} \in \operatorname{im}(f)$. This means there must exist a $\mathbf{v} \in V$ such that $f(\mathbf{v}) = \mathbf{w}$. As $\mathcal{B}_V$ is a basis of V, the vector $\mathbf{v}$ can be expressed as a linear combination of the basis vectors

$$\mathbf{v} = \alpha_1 \mathbf{b}_1 + \cdots + \alpha_n \mathbf{b}_n$$

hence

$$\mathbf{w} = f(\mathbf{v}) = f(\alpha_1 \mathbf{b}_1 + \cdots + \alpha_n \mathbf{b}_n) = \alpha_1 f(\mathbf{b}_1) + \cdots + \alpha_n f(\mathbf{b}_n).$$

This shows that any vector of the image can be expressed as a linear combination of vectors in the set $\mathcal{C} = \{f(\mathbf{b}_1), \dots, f(\mathbf{b}_n)\}$, i.e.

$$\operatorname{im}(f) \subset \operatorname{SPAN}\left(f(\mathbf{b}_1), \dots, f(\mathbf{b}_n)\right).$$

The other direction is trivial, but worth the practice. Let $\mathbf{w} \in \operatorname{SPAN}\left(f(\mathbf{b}_1), \dots, f(\mathbf{b}_n)\right)$. This means it can be written as a linear combination of these vectors

$$\begin{aligned} \mathbf{w} &= \alpha_1 f(\mathbf{b}_1) + \cdots + \alpha_n f(\mathbf{b}_n) \\ &= f(\alpha_1 \mathbf{b}_1 + \cdots + \alpha_n \mathbf{b}_n) \end{aligned}$$

which means we have found a corresponing vector $\mathbf{v} = \alpha_1 \mathbf{b}_1 + \cdots + \alpha_n \mathbf{b}_n \in V$ such that $f(\mathbf{v}) = \mathbf{w}$. Hence $\mathbf{w} \in \operatorname{im}(f)$. This proves

$$\operatorname{SPAN}\left(f(\mathbf{b}_1), \dots, f(\mathbf{b}_n)\right) \subset \operatorname{im}(f)$$

and so we have the equality of these two sets.

Definition 3.3: Rank.
The *rank of a linear map* is the dimension of its image: $\operatorname{rank}(f) = \dim(\operatorname{im}(f))$.

Example 3.7: Rank of a linear map from $\mathbb{R}^3$ to $\mathbb{R}^3$.

Let's find the dimension of a given linear map. Let $f : \mathbb{R}^3 \to \mathbb{R}^3$ be the linear map defined by

$$f(x, y, z) = (x + z, z - y, y - x - 2z).$$

We take the canonical basis of $\mathbb{R}^3$, $\mathcal{C} = \{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\} = \{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}$. From theorem ... we can form the generator set

$$\begin{aligned}\mathcal{B} &= \{f(\mathbf{e}_1),\ f(\mathbf{e}_2),\ f(\mathbf{e}_3)\}\\ &= \{(1, 0, -1),\ (0, -1, 1),\ (1, 1, -2)\}.\end{aligned}$$

Now this set is not a basis, because the third vector is the subtraction of the second from the first (and thus the set $\mathcal{B}$ is not a set of linearly independent vectors). So we can drop the third vector to find

$$\operatorname{im}(f) = \mathrm{SPAN}((1, 0, -1),\ (0, -1, 1),\ (1, 1, -2)) = \mathrm{SPAN}((1, 0, -1),\ (0, -1, 1))$$

and thus we have the basis of the image

$$\mathcal{C} = \{(1, 0, -1),\ (0, -1, 1)\}$$

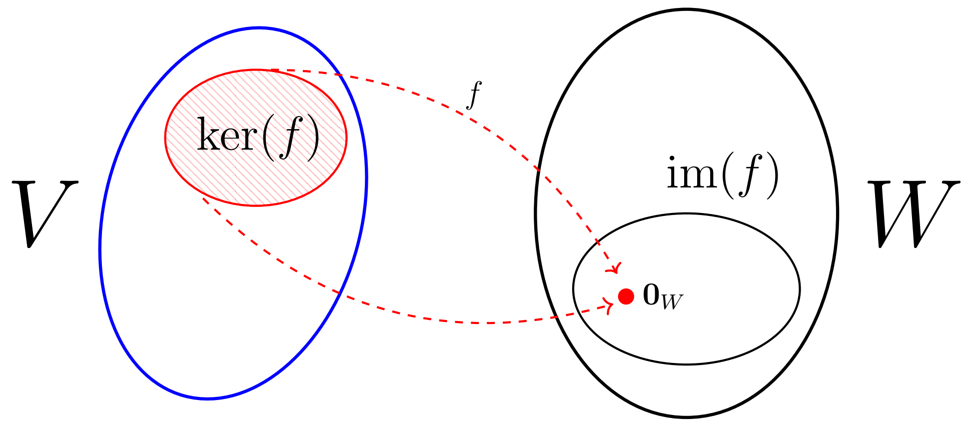
which means the dimension is 2 and hence $\operatorname{rank}(f) = 2$.

Definition 3.4: Kernel.

The *kernel of a linear map* $f : V \to W$, denoted $\ker(f)$, is the set of vectors that f maps to the zero vector, $\mathbf{0}_W$, of W. That is,

$$\ker(f) = \{\mathbf{v} \in V \mid f(\mathbf{v}) = \mathbf{0}_W\}.$$

This can be understood pictorially as so:



Example 3.8: Kernel of a linear map from $\mathbb{R}^3$ to $\mathbb{R}^3$.

Let's find the kernel of the linear map from the previous example, the map $f : \mathbb{R}^3 \to \mathbb{R}^3$ defined by

$$f(x, y, z) = (x + z, z - y, y - x - 2z).$$

We want to find all the triples that f takes to the zero vector of $\mathbb{R}^3$, which is $(0, 0, 0)$. Hence we are looking to solve

$$f(x, y, z) = (0, 0, 0) \quad \implies \quad (x + z, z - y, y - x - 2z) = (0, 0, 0).$$

Equating components gives us the linear system

$$\begin{cases} x + z = 0 \\ z - y = 0 \\ y - x - 2z = 0 \end{cases} \quad \implies \quad \begin{cases} x = -z \\ y = z \\ y - x - 2z = 0 \end{cases}.$$

So x and y can each be expressed in terms of z and the 3rd equation gives no extra constraint. z is therefore a free variable, denote it $z = t \in \mathbb{R}$, and we have

$$f(x, y, z) = (0, 0, 0) \quad \implies \quad (x, y, z) = (-t, t, t) = (-1, 1, 1)t.$$

So we can write the kernel as the set

$$\ker(f) = \{(-1, 1, 1)t \mid t \in \mathbb{R}\}.$$

The kernel is a vector subspace of W. It is non-empty because the zero vector necessarily maps to the zero, $f(\mathbf{0}_V) = \mathbf{0}_W$, and so $\mathbf{0}_V \in \ker(f)$. Let $\mathbf{u}, \mathbf{v} \in \ker(f)$ and $\alpha, \beta \in \mathbb{R}$. Then f maps the linear combination as

$$f(\alpha\mathbf{u} + \beta\mathbf{v}) = \alpha f(\mathbf{u}) + \beta f(\mathbf{v}) = \alpha\mathbf{0}_W + \beta\mathbf{0}_W = \mathbf{0}_W$$

and so $\alpha\mathbf{u} + \beta\mathbf{v}$ is also in the kernel of f. Thus $\ker(f)$ is closed under linear combinations and is a non-empty subset of V. Therefore it is a vector subspace of V.

Definition 3.5: Nullity.

The *nullity of a linear map* is the dimension of its kernel: $\text{nullity}(f) = \dim(\ker(f))$.

Example 3.9: Nullity of a linear map from $\mathbb{R}^3$ to $\mathbb{R}^3$.

Let's retake the linear map from the previous example, $f : \mathbb{R}^3 \to \mathbb{R}^3$ defined by

$$f(x, y, z) = (x + z, z - y, y - x - 2z).$$

We found the kernel as the set

$$\ker(f) = \{(-1, 1, 1)t \mid t \in \mathbb{R}\} = \text{SPAN}(\,(-1, 1, 1)\,).$$

This means that the set $\mathcal{B} = \{(-1, 1, 1)\}$ is a basis for the kernel. Hence the dimension of kernel is 1, and so

$$\text{nullity}(f) = \dim(\ker(f)) = 1.$$

Theorem 3.3: Rank-Nullity.

For any linear map $f : V \to W$ we have

$$\text{rank}(f) + \text{nullity}(f) = \dim(V)$$

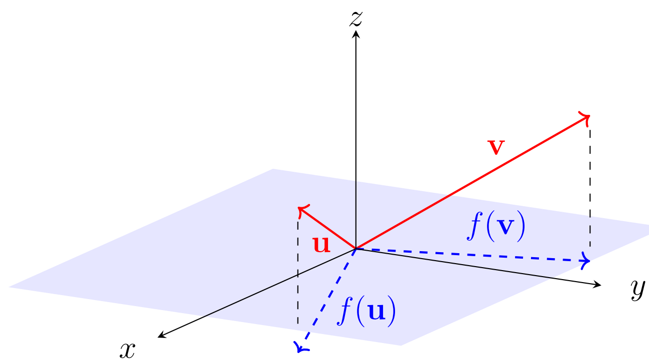
or

$$\dim(\text{im}(f)) + \dim(\ker(f)) = \dim(V).$$

Example 3.10: Projection map onto the x-y plane.
Consider the linear map $f : \mathbb{R}^3 \to \mathbb{R}^3$ defined by

$$f(x, y, z) = (x, y, 0).$$

This map takes any vector in 3d space and gives you the component of that vector in the x-y plane. Here's a sketch of the action of f on a couple of example vectors



Now the image of this function is clearly all of the x-y plane, but let's show that mathematically. Take the canonical basis of $\mathbb{R}^3$: $\mathcal{C} = \{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\} = \{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}$. We have the generator set

$$\begin{aligned}\mathcal{B} &= \{f(\mathbf{e}_1),\ f(\mathbf{e}_2),\ f(\mathbf{e}_3)\} \\ &= \{(1, 0, 0),\ (0, 1, 0),\ (0, 0, 0)\}.\end{aligned}$$

The span of this set is the same as if we drop the zero vector, so

$$\mathcal{C} = \{(1, 0, 0),\ (0, 1, 0)\}$$

is a set of linearly independent vectors such that $\text{im}(f) = \text{SPAN}(\mathcal{C})$, which means $\mathcal{C}$ is a basis for the image. With two vectors in this basis, we have the dimension of the image, and therefore the rank of f, is 2.

Now for the kernel we must solve

$$f(x, y, z) = (0, 0, 0).$$

This gives us $(x, y, 0) = (0, 0, 0)$ so that $x = 0$ and $y = 0$. There is no restriction on z, so that the kernel can be written as the set

$$\ker(f) = \{(0,0,1)t \,|\, t \in \mathbb{R}\} = \mathrm{SPAN}((0,0,1)).$$

So we can form the obvious basis $\mathcal{D} = \{(0,0,1)\}$, showing the dimension of the kernel, and hence the nullity of f, is 1. We thus verify that for this linear map, we have $\mathrm{rank}(f) + \mathrm{nullity}(f) = 2 + 1 = 3 = \dim(\mathbb{R}^3)$.

Definition 3.6: Injectivity.

Let $f : V \to W$ be a linear map. We say f is injective if no two vectors of V are mapped to the same vector of W. In symbols we have two equivalent expressions

$$\forall\, \mathbf{x}, \mathbf{y} \in V, \quad (f(\mathbf{x}) = f(\mathbf{y}) \implies \mathbf{x} = \mathbf{y})$$

or

$$\forall\, \mathbf{x}, \mathbf{y} \in V, \quad (\mathbf{x} \neq \mathbf{y} \implies f(\mathbf{x}) \neq f(\mathbf{y}))$$

Definition 3.7: Surjectivity.

Let $f : V \to W$ be a linear map. We say that f is surjective if every vector in the output space has a corresponding input vector. In symbols

$$\forall\, \mathbf{w} \in W \quad \exists \mathbf{v} \in V \text{ such that } f(\mathbf{v}) = \mathbf{w}.$$

Definition 3.8: Categories of linear maps.

Let $f : V \to W$ be a linear map.

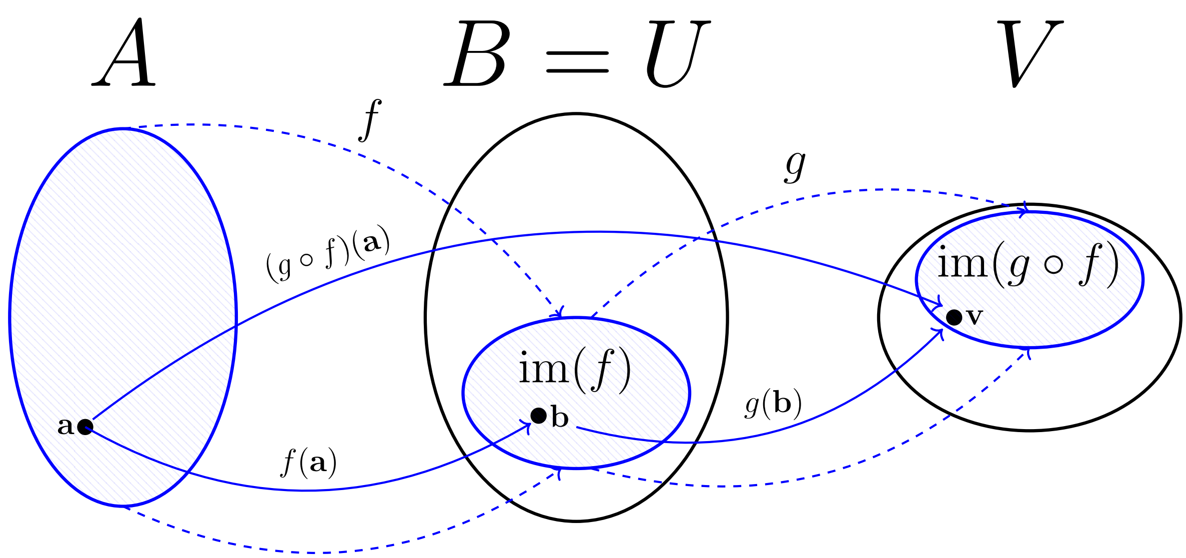
- If $W = V$ we call f an *endomorphism.*
- If f is both injective and surjective then we say it is bijective and we call it an *isomorphism.*
- If f is both an isomorphism and an endomorphism we call it an *automorphism.*

Definition 3.9: Composition of linear maps.

Composition of linear maps works exactly as you would expect if you remember the composition of regular functions. We must have a coherence between the output of one linear map and the input of another. So, two linear maps $f : U \to V$ and $g : A \to B$ can be composed as a well defined linear map $g \circ f$ ("g of f") if and only if the output space of f is the input space of g: $B = U$. For any $\mathbf{u} \in A$ the composition is written

$$g \circ f : A \to V \quad \text{and} \quad (g \circ f)(\mathbf{u}) = g(f(\mathbf{u})).$$

The composition can be represented pictorially as so:



Example 3.11: Composition of linear maps.

Let's come up with a couple of linear maps and then take their composition, why not? Consider $f : \mathbb{R}^2 \to \mathcal{P}_3[\mathbb{R}]$ and $g : \mathcal{P}_3[\mathbb{R}] \to \mathbb{R}^3$ defined by

$$f(\alpha, \beta) = \alpha + (\beta - \alpha)x + (\alpha + 2\beta)x^3$$
$$g(a_0 + a_1x + a_2x^2 + a_3x^3) = (a_3,\ a_2 - a_1,\ a_0 + a_3).$$

Now the composition of these linear maps will skip the polynomial space (the output of f and input of g):

$$
\begin{aligned}
g \circ f : & \mathbb{R}^2 \to \mathbb{R}^3 \\
(g \circ f)(\alpha, \beta) &= g\left(f(\alpha, \beta)\right) \\
&= g\left(\alpha + (\beta - \alpha)x + (\alpha + 2\beta)x^3\right) \\
&= (\alpha + 2\beta,\ \alpha - \beta,\ 2\alpha + 2\beta)
\end{aligned}
$$

So let's see for example what happens to the vector $\mathbf{v} = (1, 1)$. If we consider the step-by-step process, first act with f to obtain a polynomial, then hit that polynomial with g, we find

$$
\begin{aligned}
f(\mathbf{v}) &= f(1, 1) = 1 + 3x^3 \\
\implies g(f(\mathbf{v})) &= g(1 + 3x^3) = (3,\ 0,\ 4)\,.
\end{aligned}
$$

If we want to avoid this 2-step process we can use the composition rule as we found above

$$
(g \circ f)(1, 1) = (1 + 2,\ 1 - 1,\ 2 + 2) = (3,\ 0,\ 4).
$$

Of course we get the same answer, but the lesson here is that we never had to think about polynomials this way.

Theorem 3.4: Inverse linear map.

Let $f : V \to W$ be a linear map. If f is bijective then there exists a linear map $g : W \to V$ such that

$$
\begin{gathered}
\forall\, \mathbf{v} \in V \text{ we have } (g \circ f)(\mathbf{v}) = \mathbf{v} \\
\text{and} \\
\forall\, \mathbf{w} \in W \text{ we have } (f \circ g)(\mathbf{w}) = \mathbf{w}
\end{gathered}
$$

g is called the inverse of f and is denoted f^{-1}.

3.2 The vector space of linear maps

Theorem 3.5: A set of linear maps as a vector space.
Consider the set of all possible linear maps from the vector space V to the vector space W, denoted $\mathcal{L}(V, W)$. This set satisfies all of the vector space axioms if we define vector addition and scalar multiplication as follows:

$$\forall f, g \in \mathcal{L}(V, W), \quad f + g = h \quad \text{such that} \quad h(\mathbf{v}) = f(\mathbf{v}) + g(\mathbf{v})$$
$$\text{and}$$
$$\forall \alpha \in \mathbb{R} \quad \alpha f = f_\alpha \quad \text{such that} \quad f_\alpha(\mathbf{v}) = \alpha\left(f(\mathbf{v})\right).$$

Proof. Let $f, g, h \in \mathcal{L}(V, W)$ and $\alpha, \beta \in \mathbb{R}$. Let's go through the vector space axioms in order:
VA1 - closure under vector addition: $f + g \in \mathcal{L}(V, W)$
Let the addition be denoted $h = f + g$. We have to show that h is a linear map from V to W. For every $\mathbf{v}, \mathbf{u} \in V$ and $\alpha, \beta \in \mathbb{R}$ we have

$$\begin{aligned} h(\alpha\mathbf{v} + \beta\mathbf{u}) &= f(\alpha\mathbf{v} + \beta\mathbf{u}) + g(\alpha\mathbf{v} + \beta\mathbf{u}) \\ &= \alpha f(\mathbf{v}) + \beta f(\mathbf{u}) + \alpha g(\mathbf{v}) + \beta g(\mathbf{u}) \quad \text{(because } f \text{ and } g \text{ are linear maps)} \\ &= \alpha\left(f(\mathbf{v}) + g(\mathbf{v})\right) + \beta\left(f(\mathbf{u}) + g(\mathbf{u})\right) \\ &= \alpha h(\mathbf{v}) + \beta h(\mathbf{u}). \end{aligned}$$

So $f + g$ preserves linear combinations and is therefore a linear map. Hence $\mathcal{L}(V, W)$ is closed under vector addition.

VA2 - associativity of vector addition: $f + (g + h) = (f + g) + h \in \mathcal{L}(V, W)$

For every $\mathbf{v} \in V$ we have

$$\begin{aligned} (f + (g + h))(\mathbf{v}) &= f(\mathbf{v}) + (g + h)(\mathbf{v}) && \text{(definition of vector addition of linear maps)} \\ &= f(\mathbf{v}) + (g(\mathbf{v}) + h(\mathbf{v})) && \text{(definition of vector addition of linear maps)} \\ &= (f(\mathbf{v}) + g(\mathbf{v})) + h(\mathbf{v}) && \text{(associativity of vectors in } W\text{)} \\ &= (f + g)(\mathbf{v}) + h(\mathbf{v}) && \text{(definition of vector addition of linear maps)} \\ &= ((f + g) + h)(\mathbf{v}) && \text{(definition of vector addition of linear maps).} \end{aligned}$$

This shows that $f + (g + h)$ is the same linear map as $(f + g) + h$. Hence the addition of vectors in $\mathcal{L}(V, W)$ is associative.

VA3 - additive identity: $\exists\, f_0 \in \mathcal{L}(V,W)$, such that $f + f_0 = f_0 + f = f$

Define the zero map $f_0 : V \to W$ by

$$f_0(\mathbf{v}) = \mathbf{0}_W$$

for every $\mathbf{v} \in V$. Firstly, is this a linear map? Let's see if it preserves linear combinations

$$\begin{aligned}
f_0(\alpha\mathbf{v} + \beta\mathbf{u}) &= \mathbf{0}_W && \text{(definition of the zero map)}\\
&= \mathbf{0}_W + \mathbf{0}_W && \text{(definition of the zero vector of } W\text{)}\\
&= \alpha\mathbf{0}_W + \beta\mathbf{0}_W && \text{(a number multiplied by the zero vector is the zero vector)}\\
&= \alpha f_0(\mathbf{v}) + \beta f_0(\mathbf{u}) && \text{(reverse definition of the zero map).}
\end{aligned}$$

Indeed, this zero map is a linear map from V to W, that is $f_0 \in \mathcal{L}(V,W)$. Now we must show that this map acts as the zero vector of $\mathcal{L}(V,W)$. For every $\mathbf{v} \in V$ we have

$$\begin{aligned}
(f_0 + f)(\mathbf{v}) &= f_0(\mathbf{v}) + f(\mathbf{v}) && \text{(definition of vector addition of linear maps)}\\
&= \mathbf{0}_W + f(\mathbf{v}) && \text{(definition of zero map)}\\
&= f(\mathbf{v}) + \mathbf{0}_W && \text{(commutativity of vector addition in } W\text{)}\\
&= f(\mathbf{v}) && \text{(definition of zero vector of } W\text{)}
\end{aligned}$$

Hence this zero map is a member of $\mathcal{L}(V,W)$ and satisifes $f + f_0 = f_0 + f = f$. This proves the zero map is the zero *vector* of $\mathcal{L}(V,W)$.

VA4 - additive inverse: $\exists\, f_- \in \mathcal{L}(V,W)$ such that $f + f_- = f_0$

This can easily become symbolically ambiguous, so I will try to be pedantically careful here. Define the additive inverse of any map f, denote it f_-, as the scalar multiplication of that map (as defined) by the real number -1. So that the map $f_- : V \to W$ is given by

$$f_-(\mathbf{v}) = -1 \times f(\mathbf{v}).$$

First let's show this map belongs to $\mathcal{L}(V, W)$ by checking the preservation of linear combinations

$$\begin{aligned}
f_-(\alpha\mathbf{v} + \beta\mathbf{u}) &= -1 \times f(\alpha\mathbf{v} + \beta\mathbf{u}) && \text{(definition of this map)} \\
&= -1 \times (\alpha f(\mathbf{v}) + \beta f(\mathbf{u})) && \text{(linearity of the map } f\text{)} \\
&= -1 \times \alpha f(\mathbf{v}) - 1 \times \beta f(\mathbf{u}) && \text{(distributivity of the reals)} \\
&= \alpha \times -1 \times f(\mathbf{v}) + \beta \times -1 \times f(\mathbf{u}) && \text{(commutativity of real multiplication)} \\
&= \alpha f_-(\mathbf{v}) + \beta f_-(\mathbf{u}) && \text{(definition of this map).}
\end{aligned}$$

Indeed, this map is a linear map from V to W, that is $f_- \in \mathcal{L}(V, W)$. Now we show that it acts as an additive inverse. For every $\mathbf{v} \in V$ we have

$$\begin{aligned}
(f_- + f)(\mathbf{v}) &= f_-(\mathbf{v}) + f(\mathbf{v}) && \text{(definition of vector addition of linear maps)} \\
&= -1 \times f(\mathbf{v}) + f(\mathbf{v}) && \text{(definition of this map)} \\
&= -f(\mathbf{v}) + f(\mathbf{v}) && \text{(multiplication by 1 for vectors in } W\text{)} \\
&= \mathbf{0}_W && \text{(additive inverse of vectors in } W\text{)}
\end{aligned}$$

With commutativity of vectors in W you can also show $f_- + f = f + f_-$. Hence this additive inverse map exists in $\mathcal{L}(V, W)$.

VA5 - commutativity of vector addition: $f + g = g + f$

This one is pretty simple. For every $\mathbf{v} \in V$ we have

$$\begin{aligned}
(f + g)(\mathbf{v}) &= f(\mathbf{v}) + g(\mathbf{v}) && \text{(definition of vector addition of linear maps)} \\
&= g(\mathbf{v}) + f(\mathbf{v}) && \text{(commutativity of vector addition in } W\text{)} \\
&= (g + f)(\mathbf{v}) && \text{(definition of vector addition of linear maps).}
\end{aligned}$$

Hence the addition of vectors in $\mathcal{L}(V, W)$ is commutative.

SM1 - closure under scalar multiplication: $\alpha f \in \mathcal{L}(V, W)$

Recall the definition of scalar multiplication of the linear map f by a real number k gives a new map, denote it f_k, such that

$$f_k(\mathbf{v}) = kf(\mathbf{v}).$$

Let's show this map belongs to $\mathcal{L}(V, W)$ by checking the preservation of linear combinations

$$\begin{aligned}
f_k(\alpha\mathbf{v} + \beta\mathbf{u}) &= kf(\alpha\mathbf{v} + \beta\mathbf{u}) && \text{(definition of this map)}\\
&= k\left(\alpha f(\mathbf{v}) + \beta f(\mathbf{u})\right) && \text{(linearity of the map } f\text{)}\\
&= k\alpha f(\mathbf{v}) + k\beta f(\mathbf{u}) && \text{(distributivity of the reals)}\\
&= \alpha \times kf(\mathbf{v}) + \beta \times kf(\mathbf{u}) && \text{(commutativity of real multiplication)}\\
&= \alpha f_k(\mathbf{v}) + \beta f_k(\mathbf{u}) && \text{(definition of this map).}
\end{aligned}$$

Indeed, this map is a linear map from V to W, and so $\mathcal{L}(V, W)$ is closed under scalar multiplication.

SM2 - distributivity over vector addition: $\alpha(f + g) = \alpha f + \alpha g$

For every $\mathbf{v} \in V$ we have

$$\begin{aligned}
(\alpha(f + g))(\mathbf{v}) &= \alpha((f + g)(\mathbf{v})) && \text{(definition of scalar multiplication for linear maps)}\\
&= \alpha(f(\mathbf{v}) + g(\mathbf{v})) && \text{(definition of vector addition for linear maps)}\\
&= \alpha\left(f(\mathbf{v})\right) + \alpha\left(g(\mathbf{v})\right) && \text{(distributivity over vector addition in } W\text{)}\\
&= (\alpha f)(\mathbf{v}) + (\alpha g)(\mathbf{v}) && \text{(definition of scalar multiplication for linear maps)}\\
&= (\alpha f + \alpha g)(\mathbf{v}) && \text{(definition of addition for linear maps).}
\end{aligned}$$

This shows that $\alpha(f + g)$ is the same linear map as $\alpha f + \alpha g$. Hence scalar multiplication for linear maps distributes over vector addition of linear maps.

SM3 - distributivity with field addition: $(\alpha + \beta)f = \alpha f + \beta g$

For every $\mathbf{v} \in V$ we have

$$\begin{aligned}
((\alpha + \beta)f)(\mathbf{v}) &= (\alpha + \beta)\left(f(\mathbf{v})\right) && \text{(definition of scalar multiplication for linear maps)}\\
&= \alpha f(\mathbf{v}) + \beta f(\mathbf{v}) && \text{(distributivity over field addition for } W\text{)}\\
&= (\alpha f)(\mathbf{v}) + (\beta f)(\mathbf{v}) && \text{(definition of scalar multiplication for linear maps)}\\
&= (\alpha f + \beta f)(\mathbf{v}) && \text{(definition of vector addition for linear maps).}
\end{aligned}$$

This shows that $(\alpha + \beta)f$ is the same linear map as $\alpha f + \beta f$. Hence field addition distributes over linear maps.

SM4 - compatibility of scalar and field multiplication: $\alpha(\beta f) = (\alpha\beta)f$

For every $\mathbf{v} \in V$ we have

$$\begin{aligned}
(\alpha(\beta f))(\mathbf{v}) &= \alpha\left((\beta f)(\mathbf{v})\right) && \text{(definition of scalar multiplication for linear maps)} \\
&= \alpha\left(\beta(f(\mathbf{v}))\right) && \text{(definition of scalar multiplication for linear maps)} \\
&= (\alpha\beta)\left(f(\mathbf{v})\right) && \text{(compatibility of scalar and field multiplication for } W\text{)} \\
&= \left((\alpha\beta) f\right)(\mathbf{v}) && \text{(definition of scalar multiplication for linear maps)}
\end{aligned}$$

This shows that $\alpha(\beta f)$ is the same linear map as $(\alpha\beta)f$. Hence scalar multiplication is compatible with field multiplication for linear maps.

SM5 - multiplicative identity: $1f = f$ For every $\mathbf{v} \in V$ we have

$$\begin{aligned}
(1f)(\mathbf{v}) &= 1\left(f(\mathbf{v})\right) && \text{(definition of scalar multiplication for linear maps)} \\
&= f(\mathbf{v}) && \text{(multiplicative identity for } W\text{)}
\end{aligned}$$

This shows that $1f$ is the same linear map as f. Hence the real number 1 is the scalar multiplicative identity for linear maps.

It was long and perhaps tedious but we have now completed the proof that with this definition of linear map vector addition and scalar multiplication, the set $\mathcal{L}(V, W)$ satisfies all 10 of the vector space axioms.

3.3 Matrices

Let's recall the definition of coordinates of a vector. Given any vector $\mathbf{v}$ in some n-dimensional vector space V, if we have a basis $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ then the coordinates of $\mathbf{v}$ *in this basis*, denoted $[\mathbf{v}]_{\mathcal{B}}$, is a column of n numbers which are the coefficients of the linear combination of $\mathbf{v}$ in the basis vectors

$$\mathbf{v} = \alpha_1 \mathbf{b}_1 + \cdots + \alpha_n \mathbf{b}_n \quad \Longrightarrow \quad [\mathbf{v}]_{\mathcal{B}} = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix}.$$

I hope to show in this section something quite powerful about linear algebra, that any n-dimensional vector space can be mapped to $\mathbb{R}^n$, via the coordinates, and that linear maps take on a particularly simple to use form when mapping between coordinates. We'll start by an example.

Example 3.12: Rotation linear map.

Let's retake the first example of this chapter, the linear map, $f : \mathbb{R}^2 \to \mathbb{R}^2$, that rotates 2d vectors by 45° anti-clockwise, defined in detail by

$$f(x,y) = \frac{1}{\sqrt{2}}\left(x-y, x+y\right).$$

Let $\mathcal{C} = \{\mathbf{e}_x,\, \mathbf{e}_y\} = \{(1,0),\, (0,1)\}$. We will want to represent f by some operation, call it F, that takes the coordinates of any vector to the coordinates of the rotated vector. Let's write this desired property as

$$F[\mathbf{v}]_{\mathcal{C}} = [f(\mathbf{v})]_{\mathcal{C}}.$$

Now let $\mathbf{v} = (x, y)$ be some arbitrary vector, so that

$$\begin{aligned}
[f(\mathbf{v})]_{\mathcal{C}} &= [f(x\mathbf{e}_x + y\mathbf{e}_y)]_{\mathcal{C}} \\
&= [\frac{1}{\sqrt{2}}\left(x-y, x+y\right)]_{\mathcal{C}} \\
&= [\left(\frac{x-y}{\sqrt{2}}\right)\mathbf{e}_x + \left(\frac{x+y}{\sqrt{2}}\right)\mathbf{e}_y]_{\mathcal{C}} \\
&= \begin{pmatrix} \frac{x-y}{\sqrt{2}} \\ \frac{x+y}{\sqrt{2}} \end{pmatrix}.
\end{aligned}$$

These coordinates can be rearranged to focus on the x and y

$$\begin{pmatrix} \frac{x-y}{\sqrt{2}} \\ \frac{x+y}{\sqrt{2}} \end{pmatrix} = \begin{pmatrix} \frac{x}{\sqrt{2}} \\ \frac{x}{\sqrt{2}} \end{pmatrix} + \begin{pmatrix} \frac{-y}{\sqrt{2}} \\ \frac{y}{\sqrt{2}} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} x + \begin{pmatrix} \frac{-1}{\sqrt{2}} \\ \frac{}{\sqrt{2}} \end{pmatrix} y.$$

So at this point we have

$$F[\mathbf{v}]_{\mathcal{C}} = [f(\mathbf{v})]_{\mathcal{C}} \qquad \Longrightarrow \qquad F\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} x + \begin{pmatrix} \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} y.$$

We will define a certain rearrangement of the right hand side

$$\begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} x + \begin{pmatrix} \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} y = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

so that the equation

$$F\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}$$

let's us identify the operation of F with this array of 4 numbers

$$F = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}$$

This kind of array of numbers is called a *matrix*. In this case we say that F is a matrix representation of the linear map f. It's job is to represent the same mapping as f but on the *coordinates* of vectors in the domain and codomain of f.

Let's take another example to show that the matrix representation is not always an array of 4 numbers.

Example 3.13: Non-square matrix.

Consider a linear map $f : \mathbb{R}^2 \to \mathcal{P}_2[\mathbb{R}]$ defined by

$$f(\alpha, \beta) = \alpha + \beta x + (\alpha - \beta)x^2.$$

As in the previous example, we desire that the matrix F maps the coordinates of an arbitrary vector $\mathbf{v} = (\alpha, \beta)$ to the coordinates of $f(\mathbf{v})$. The coordinates depend on the choices of bases for the domain and codomain vector spaces. Let's take the canonical bases

$$\mathcal{A} = \{\mathbf{e}_x,\, \mathbf{e}_y\} = \{(1,0),\, (0,1)\}$$
$$\mathcal{B} = \{\mathbf{p}_0,\, \mathbf{p}_1,\, \mathbf{p}_2\} = \{1,\, x,\, x^2\}$$

so that

$$F[(\alpha,\beta)]_{\mathcal{A}} = [f(\alpha,\beta)]_{\mathcal{B}} \quad\implies\quad \begin{aligned} F\begin{pmatrix}\alpha\\ \beta\end{pmatrix} &= [\alpha+\beta x+(\alpha-\beta)x^2]_{\mathcal{B}} \\ &= \begin{pmatrix}\alpha\\ \beta\\ \alpha-\beta\end{pmatrix} \\ &= \begin{pmatrix}1\\0\\1\end{pmatrix}\alpha + \begin{pmatrix}0\\1\\-1\end{pmatrix}\beta. \end{aligned}$$

We then use this last term to define a matrix

$$\begin{pmatrix}1\\0\\1\end{pmatrix}\alpha + \begin{pmatrix}0\\1\\-1\end{pmatrix}\beta. = \begin{pmatrix}1 & 0\\0 & 1\\1 & -1\end{pmatrix}\begin{pmatrix}\alpha\\ \beta\end{pmatrix}.$$

So we see that we have a matrix with 3 rows and 2 columns

$$F\begin{pmatrix}\alpha\\ \beta\end{pmatrix} = \begin{pmatrix}1 & 0\\0 & 1\\1 & -1\end{pmatrix}\begin{pmatrix}\alpha\\ \beta\end{pmatrix} \quad\implies\quad F = \begin{pmatrix}1 & 0\\0 & 1\\1 & -1\end{pmatrix}.$$

So now that we perhaps can see that matrices can take any shape we may as well define them clearly.

Definition 3.10: Matrix.

A matrix, denoted A, is an array of numbers organised into rows and columns.

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1j} & \cdots & a_{1n} \\ \vdots & \cdots & \vdots & \cdots & \vdots \\ a_{i1} & \cdots & a_{ij} & \cdots & a_{in} \\ \vdots & \cdots & \vdots & \cdots & \vdots \\ a_{m1} & \cdots & a_{mj} & \cdots & a_{mn} \end{pmatrix}$$

We say A is an $m \times n$ matrix if it has m rows and n columns. The numbers in the array are called coefficients or elements of A, and are often denoted by their indices

$$a_{ij} = (A)_{ij} = a_{i,j}.$$

In the previous examples, there was a key step in the creation of the matrices, when we claimed

$$\begin{pmatrix} \alpha \\ \gamma \end{pmatrix} x + \begin{pmatrix} \beta \\ \delta \end{pmatrix} y = \begin{pmatrix} \alpha & \beta \\ \delta & \gamma \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} \alpha \\ \gamma \\ \epsilon \end{pmatrix} x + \begin{pmatrix} \beta \\ \delta \\ \zeta \end{pmatrix} y = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \\ \epsilon & \zeta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

We will simply reverse the direction of the equality and take this as a defining statement of how to multiply a matrix by a column:

Definition 3.11: Matrix multiplication by a column.

The multiplication of an $m \times n$ matrix, A, by a column, X, is defined only if the column contains as many elements as the columns of A. It is given by

$$AX = \begin{pmatrix} a_{11} & \cdots & a_{1j} & \cdots & a_{1n} \\ \vdots & \cdots & \vdots & \cdots & \vdots \\ a_{i1} & \cdots & a_{ij} & \cdots & a_{in} \\ \vdots & \cdots & \vdots & \cdots & \vdots \\ a_{m1} & \cdots & a_{mj} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} a_{11} \\ \vdots \\ a_{j1} \\ \vdots \\ a_{m1} \end{pmatrix} x_1 + \cdots + \begin{pmatrix} a_{1n} \\ \vdots \\ a_{jn} \\ \vdots \\ a_{mn} \end{pmatrix} x_n$$

$$= \begin{pmatrix} a_{11}x_1 + \cdots + a_{1j}x_j + \cdots + a_{1n}x_n \\ \vdots \\ a_{i1}x_1 + \cdots + a_{ij}x_j + \cdots + a_{in}x_n \\ \vdots \\ a_{m1}x_1 + \cdots + a_{mj}x_j + \cdots + a_{mn}x_n \end{pmatrix}$$

This gives an expression for the i^{th} element of the resulting column that you may have seen in other textbooks

$$(AX)_i = \sum_{j=1}^{n} a_{ij}x_j.$$

Now let's be very general and consider an undefined linear map. Let $f \in \mathcal{L}(V, W)$, $\mathcal{A} = \{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$ be a basis for the n-dimensional vector space V and $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_m\}$ be a basis for the m-dimensional vector space W. Then let F be the operation of f that takes the coordinates of an arbitrary vector $\mathbf{v} \in V$ to the coordinates of $f(\mathbf{v}) \in W$

$$F[\mathbf{v}]_{\mathcal{A}} = [f(\mathbf{v})]_{\mathcal{B}}.$$

As $\mathcal{A}$ is basis of V, the vector $\mathbf{v}$ can be expressed as a linear combination of these vectors and the coefficients give its coordinates in this basis

$$\exists \alpha_1, \ldots, \alpha_n \in \mathbb{R} \quad \text{such that} \quad \mathbf{v} = \alpha_1\mathbf{a}_1 + \cdots + \alpha_n\mathbf{a}_n \quad \Longrightarrow \quad [\mathbf{v}]_{\mathcal{A}} = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix}.$$

Hence we have

$$
\begin{aligned}
F\begin{pmatrix}\alpha_1\\ \vdots\\ \alpha_n\end{pmatrix} &= [f(\alpha_1\mathbf{a}_1+\alpha_2\mathbf{a}_2+\cdots+\alpha_n\mathbf{a}_n)]_{\mathcal{B}}\\
&= [\alpha_1 f(\mathbf{a}_1)+\alpha_2 f(\mathbf{a}_2)+\cdots+\alpha_n f(\mathbf{a}_n)]_{\mathcal{B}}\\
&= \alpha_1[f(\mathbf{a}_1)]_{\mathcal{B}}+\alpha_2[f(\mathbf{a}_2)]_{\mathcal{B}}+\cdots+\alpha_n[f(\mathbf{a}_n)]_{\mathcal{B}}.
\end{aligned}
$$

If we remember that the $[f(\mathbf{a}_k)]_{\mathcal{B}}$ are columns, then this last line gives us the form of a matrix multiplied by a column

$$
F\begin{pmatrix}\alpha_1\\ \vdots\\ \alpha_n\end{pmatrix} = \alpha_1[f(\mathbf{a}_1)]_{\mathcal{B}}+\cdots+\alpha_n[f(\mathbf{a}_n)]_{\mathcal{B}} = \begin{pmatrix} | & | & & | \\ [f(\mathbf{a}_1)]_{\mathcal{B}} & [f(\mathbf{a}_2)]_{\mathcal{B}} & \ldots & [f(\mathbf{a}_n)]_{\mathcal{B}} \\ | & | & & | \end{pmatrix}\begin{pmatrix}\alpha_1\\ \vdots\\ \alpha_n\end{pmatrix}
$$

This gives us a general method of finding the matrix representation of any linear map. In this case we've proved the following theorem before expressing it:

Theorem 3.6: Matrix representation of a linear map.

Let $f: V \to W$ be a linear map, $\mathcal{A} = \{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$ be a basis of V, $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_m\}$ be a basis of W. Let $\mathbf{v}$ be any vector in V and $\mathbf{w} = f(\mathbf{v}) \in W$. Then the *matrix representation* of f in bases $\mathcal{A}$ and $\mathcal{B}$, defined by the unique $m \times n$ matrix, denoted $\mathcal{M}_{\mathcal{A}\to\mathcal{B}}(f)$, which takes the coordinates of $\mathbf{v}$ to the coordinates of $\mathbf{w}$ in their respective bases:

$$
\mathcal{M}_{\mathcal{A}\to\mathcal{B}}(f)[\mathbf{v}]_{\mathcal{A}} = [\mathbf{w}]_{\mathcal{B}}.
$$

can be calculated by expressing the coordinates of the linear map acting on the basis vectors of the input space

$$
\mathcal{M}_{\mathcal{A}\to\mathcal{B}}(f) = \begin{pmatrix} | & & | \\ [f(\mathbf{a}_1)]_{\mathcal{B}} & \ldots & [f(\mathbf{a}_n)]_{\mathcal{B}} \\ | & & | \end{pmatrix}
$$

where the vertical lines are reminders that the coordinates of the $f(\mathbf{a}_k)$ vectors are columns. We often shorten "matrix representation of f" to just "matrix of f". If the input and output vector spaces are the same, i.e. if f is an endomorphism, we can use the same basis for both spaces and we may shorten the notation $\mathcal{M}_{\mathcal{A}\to\mathcal{A}}(f) = \mathcal{M}_{\mathcal{A}}(f)$.

Theorem 3.7: Identity map and identity matrix.

The identity map is the endomorphism $f_I : V \to V$ for any n-dimensional vector space V, defined by

$$f_I(\mathbf{v}) = \mathbf{v} \quad \forall\, \mathbf{v} \in V.$$

Let $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ be some basis of V. Then the matrix of f_I in this basis for both the input and output space is given by

$$\mathcal{M}_{\mathcal{B}}(f_I) = \begin{pmatrix} | & & | \\ [f_I(\mathbf{b}_1)]_{\mathcal{B}} & \dots & [f_I(\mathbf{b}_n)]_{\mathcal{B}} \\ | & & | \end{pmatrix} = \begin{pmatrix} | & & | \\ [\mathbf{b}_1]_{\mathcal{B}} & \dots & [\mathbf{b}_n]_{\mathcal{B}} \\ | & & | \end{pmatrix} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

This matrix is called the identity matrix of size n and is denoted I_n or just I when the context is clear.

Theorem 3.8: Map composition and matrix multiplication.

$$F[\mathbf{v}]_{\mathcal{A}} = [f(\mathbf{v})]_{\mathcal{B}}, \quad G[\mathbf{w}]_{\mathcal{B}} = [g(\mathbf{w})]_{\mathcal{C}}, \quad M[\mathbf{v}]_{\mathcal{A}} = [g \circ f(\mathbf{v})]_{\mathcal{C}} \quad \implies \quad M = GF$$

Theorem 3.9: Inverse linear map and matrix inversion.

$$F[\mathbf{v}]_{\mathcal{A}} = [\mathbf{w}]_{\mathcal{B}}, \quad G[\mathbf{w}]_{\mathcal{B}} = [\mathbf{v}]_{\mathcal{A}}, \quad \implies \quad G = F^{-1}$$

Definition 3.12: Transition matrix (change-of-basis matrix).

The *transition matrix* changes the representation of the coordinates of a vector from one basis into another. Let $\mathcal{A}$ and $\mathcal{B}$ be two bases of the same vector space, V, and let $\mathbf{v} \in V$. The transition matrix from $\mathcal{A}$ to $\mathcal{B}$, denoted $P_{\mathcal{A}\to\mathcal{B}}$ or $P_{\mathcal{A},\mathcal{B}}$, satisfies

$$P_{\mathcal{A},\mathcal{B}}[\mathbf{v}]_{\mathcal{A}} = [\mathbf{v}]_{\mathcal{B}}.$$

If we let the bases $\mathcal{A} = \{\mathbf{a}_1, \dots, \mathbf{a}_n\}$ and $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_m\}$ then the transition matrix can be calculated by

$$P_{\mathcal{A}\to\mathcal{B}} = P_{\mathcal{A},\mathcal{B}} = \begin{pmatrix} | & | & & | \\ [\mathbf{a}_1]_{\mathcal{B}} & [\mathbf{a}_2]_{\mathcal{B}} & \dots & [\mathbf{a}_n]_{\mathcal{B}} \\ | & | & & | \end{pmatrix}$$

where the vertical lines are reminders that the coordinates of the A basis vectors are columns. You can think of the transition matrix as the matrix representation of the identity linear map $I : V \to V$, $I(\mathbf{v}) = \mathbf{v}$ for every $\mathbf{v} \in V$, in these bases: $P_{\mathcal{A}\to\mathcal{B}} = \mathcal{M}_{\mathcal{A}\to\mathcal{B}}(I)$.

Chapter 4

Matrix Algebra and Linear Systems

Definition 4.1: Matrix.

A matrix, denoted A, is an array of numbers organised into rows and columns.

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1j} & \cdots & a_{1n} \\ \vdots & \cdots & \vdots & \cdots & \vdots \\ a_{i1} & \cdots & a_{ij} & \cdots & a_{in} \\ \vdots & \cdots & \vdots & \cdots & \vdots \\ a_{m1} & \cdots & a_{mj} & \cdots & a_{mn} \end{pmatrix}$$

We say A is an $m \times n$ matrix if it has m rows and n columns. The numbers in the array are called coefficients or elements of A, and are often denoted by their indices

$$a_{ij} = (A)_{ij} = a_{i,j}.$$

Definition 4.2: Matrix addition.

Matrix addition is done entry by entry, that is, for two matrices A and B we define $(A+B)_{ij} = (A)_{ij} + (B)_{ij}$. This means the addition is only well defined if the matrices have the same size. For example

$$\begin{pmatrix} 1 & 2 & -2 \\ 3 & -1 & 0 \end{pmatrix} + \begin{pmatrix} 0 & -1 & 3 \\ 2 & 0 & 3 \end{pmatrix} = \begin{pmatrix} 1+0 & 2-1 & -2+3 \\ 3+2 & -1+0 & 0+3 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 \\ 5 & -1 & 3 \end{pmatrix}$$

Definition 4.3: Transpose of a matrix.

The transpose of an $m \times n$ matrix, A, is an $n \times m$ matrix, denoted A^T, with rows equal to the columns of A. That is, $\left(A^T\right)_{ij} = (A)_{ji}$ for all combinations of i and j. For example

$$\begin{pmatrix} 1 & -2 \\ 0 & -1 \\ -1 & 0 \end{pmatrix}^T = \begin{pmatrix} 1 & 0 & -1 \\ -2 & -1 & 0 \end{pmatrix}$$

Definition 4.4: Identity matrix.

The n-dimensional identity matrix I is a square matrix of size $n \times n$ with 1s along the diagonal and 0s elsewhere, that is,

$$(I)_{ij} = \begin{cases} 1 & \text{whenever } i = j, \\ 0 & \text{whenever } i \neq j. \end{cases}$$

Definition 4.5: Diagonal matrix.

A square matrix A is said to be diagonal if all its non-diagonal elements are zero, e.g. $(A)_{ij} = 0$ whenever $i \neq j$.

Definition 4.6: Symmetric matrix.

A matrix A is symmetric if it is equal to its transpose, $A = A^T$.

Definition 4.7: Invertible matrix.

A matrix A is invertible if there exists a matrix B such that

$$AB = BA = I$$

This matrix B is called the inverse of A and is denoted A^{-1}.

Definition 4.8: Determinant of a 2×2 matrix.
For a 2×2 matrix

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

its determinant, denoted $\det(A)$ or $|A|$, is given by

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$$

Definition 4.9: Submatrix.
From a matrix A we generate the *submatrix* A_{ij} by deleting the ith row and jth column:

$$\text{For } A = \begin{pmatrix} a_{1,1} & \cdots & a_{1,j-1} & a_{1,j} & a_{1,j+1} & \cdots & a_{1,n} \\ \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ a_{i-1,1} & \cdots & a_{i-1,j-1} & a_{i-1,j} & a_{i-1,j+1} & \cdots & a_{i-1,n} \\ a_{i,1} & \cdots & a_{i,j-1} & a_{i,j} & a_{i,j+1} & \cdots & a_{i,n} \\ a_{i+1,1} & \cdots & a_{i+1,j-1} & a_{i+1,j} & a_{i+1,j+1} & \cdots & a_{i+1,n} \\ \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ a_{m,1} & \cdots & a_{m,j-1} & a_{m,j} & a_{m,j+1} & \cdots & a_{m,n} \end{pmatrix}$$

$$\text{The submatrix } A_{ij} = \begin{pmatrix} a_{1,1} & \cdots & a_{1,j-1} & a_{1,j+1} & \cdots & a_{1,n} \\ \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ a_{i-1,1} & \cdots & a_{i-1,j-1} & a_{i-1,j+1} & \cdots & a_{i-1,n} \\ a_{i+1,1} & \cdots & a_{i+1,j-1} & a_{i+1,j+1} & \cdots & a_{i+1,n} \\ \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ a_{m,1} & \cdots & a_{m,j-1} & a_{m,j+1} & \cdots & a_{m,n} \end{pmatrix}$$

Note: we generally have to specify in words that we create a submatrix. The notation A_{ij} is a little ambiguous without being explicit.

Definition 4.10: Cofactor matrix.

From a matrix A we generate its cofactor matrix C_A which has entries given by determinants of submatrices of A with the same plus/minus pattern as in a determinant calculation. That is, the entries of C_A are $c_{ij} = (-1)^{i+j}\det(A_{ij})$:

$$C_A = \begin{pmatrix} |A_{11}| & -|A_{12}| & |A_{13}| & \cdots \\ -|A_{21}| & |A_{22}| & -|A_{23}| & \cdots \\ |A_{31}| & -|A_{32}| & |A_{33}| & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

Theorem 4.1: Inverse matrix (cofactor method).

The inverse of a matrix A can be computed from its cofactor matrix:

$$A^{-1} = \frac{1}{\det(A)} C_A^T$$

Theorem 4.2: Inverse of a 2×2 matrix.

With the cofactor method, the inverse of a 2×2 matrix is given by

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -c \\ -b & a \end{pmatrix}$$