Authored by Dhruva Raman

Cheatsheet

Contents

Linear Algebra

How to learn

Remember this is a humble *cheatsheet*. It's only definitions. It's not useful to memorise definitions that can be found online. *You* need to develop intuition and understanding of all these definitions through the lectures and the 3blue1brown videos here. And most importantly, **by doing exercises!!**. Ideally, the exercises on the 3blue1brown transcripts, the notebook, and the questions in this cheatsheet.

For the 3blue1brown link, you can skip the videos/transcripts on Cramer's rule and cross products

Linear algeba is *lots of work*. It's also foundational material which pops up again and again. Each time you learn it, you get better. Beginners won't learn as much as people who've seen it before, and that's OK.

- The lectures are a *minimum* amount of material you need.
- The cheatsheets add *more* useful material.
- Ideally, you'd have gone through everything on the 3blue1brown course. Even if you've done Linear Algebra before, you'll get more insight by going over the geometric approach he uses carefully.
- If you're experienced and can do that, go through Chapters 1 and 2 here, instead of doing the notebooks.

For week 3 and 4 (linear algebra), my lecture *won't* last two hours. Instead I encourage you to come to the lecture with questions from the 3blue1brown course or the cheatsheet or the notebooks. This is called *flipped learning*.

You will get a week off in the middle of term, where I don't present any slides in the lecture. I encourage you to spend that week learning Linear Algebra, and giving me questions you have.

Note that after linear algebra, we will do aspects of probability that don't require linear algebra. This gives you breathing room to keep working on it without falling behind on the module.

1. Key operations and laws

1.1. Taxonomy of ordered collections

We introduce by example (below). Note that scalars are elements, not collections.

Example	Belongs to which set	Tensor-dimension	Shape	Elements
$x = 3.1$	$\in \mathbb{R}$	0-tensor (scalar)	NA	NA
$v = \begin{bmatrix} 7 \\ 3 \\ 4 \\ 5 \end{bmatrix}$	$\in \mathbb{R}^4$	1-tensor (vector)	4	$v_3 = 4$
$A = \begin{bmatrix} 1 & 2 & \dots & 10 \\ 2 & 3 & \dots & 11 \\ \vdots & \vdots & \ddots & \vdots \\ 8 & 9 & \dots & 18 \end{bmatrix}$	$\in \mathbb{R}^{8 \times 10}$	2-tensor (matrix)	(8,10)	$A_{8,2} = 9$
$T =$ see here	$\in \mathbb{R}^{2 \times 5 \times 3}$	3-tensor (tensor)	(2,5,3)	$T_{2,3,1} = 7$

- It's *really important* to note that mathematicians will often refer to v above as a four-dimensional vector. They are referring to its *vector dimension*: the dimension of the vector space it lives in. What is this? You'll find out lower down!

- This is different to the *tensor dimension* in the table above, which refers to the number of indices you need to specify an element in it.

1.2. Operations on vectors

Notation will assume that $\underline{v}, \underline{w} \in \mathbb{R}^n$

Operation	Example	Notation	Outcome is in
Cross product	Not on this course!	$\underline{v} \times \underline{w}$	
Addition	$\begin{bmatrix}7\\3\\5\end{bmatrix} + \begin{bmatrix}4\\4\\3\end{bmatrix} = \begin{bmatrix}11\\7\\8\end{bmatrix}$	$\underline{v} + \underline{w}$	$\mathbb{R}^n$
Scaling	$3\begin{bmatrix}7\\3\\5\end{bmatrix} = \begin{bmatrix}21\\9\\15\end{bmatrix}$	$k\underline{v}, k \in \mathbb{R}$	$\mathbb{R}^n$
Dot (/inner) product	$\begin{bmatrix}7\\3\\5\end{bmatrix} \bullet \begin{bmatrix}4\\4\\3\end{bmatrix} = \begin{matrix}7\times4+\\3\times4+\\5\times3\end{matrix}$	$\underline{v}^T\underline{w}$ $\underline{v} \bullet \underline{w}$ $\langle \underline{v}, \underline{w} \rangle$	$\mathbb{R}$

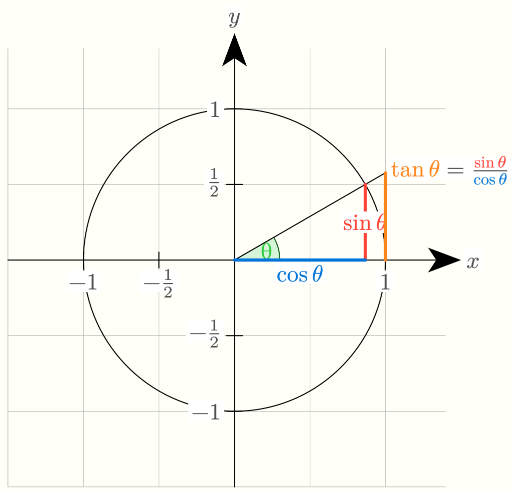
- The dot product has a *geometric interpretation* through the identity

$$\cos(\theta) = \frac{\langle \underline{v}, \underline{w} \rangle}{\| \underline{v} \|_2 \| \underline{w} \|_2} \tag{1}$$

- The double vertical lines denote the *norm* of the vector. See later section.

$\langle \underline{v}, \underline{w} \rangle = 1$	vectors must be parallel
$\langle \underline{v}, \underline{w} \rangle = 0$	vectors must be orthogonal (right angles)
$\langle \underline{v}, \underline{w} \rangle = -1$	vectors must be parallel (but opposite directions)

You should be able to figure these requirements out from the plot on the side



1.3. Operations on matrices

Assumptions for following table	$A \in \mathbb{R}^{m\times n}$	$B \in \mathbb{R}^{n\times p}$
	$\underline{v} \in \mathbb{R}^n$	$\underline{w} \in \mathbb{R}^p$

Operation	Example	Notation	Outcome is in
Transposition	$\begin{bmatrix} 3 & 5 & 4 \\ 2 & 3 & 1 \end{bmatrix}^T = \begin{bmatrix} 3 & 2 \\ 5 & 3 \\ 4 & 1 \end{bmatrix}$	A^T	$A^T \in \mathbb{R}^{n \times m}$
Matrix multiplication	$\begin{bmatrix} 3 & 5 & 4 \\ 2 & 3 & 1 \end{bmatrix} \times \begin{bmatrix} 2 & 4 \\ 3 & 1 \end{bmatrix}$	AB	$AB \in \mathbb{R}^{m \times p}$
Matrix multiplication (with vector)	$\begin{bmatrix} 3 & 5 & 4 \\ 2 & 3 & 1 \end{bmatrix} \times \begin{bmatrix} 2 \\ 3 \end{bmatrix}$	$A\underline{v}$	$AB \in \mathbb{R}^{m \times p}$
Inversion	$m = n$ (invertible square matrices only)	A^{-1}	$\mathbb{R}^{n \times n}$

1.4. Special square matrices

- A square matrix has the same number of rows and columns.

Name	Example	Notes
Identity matrix	$I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$A\mathbb{I}_n = \mathbb{I}_m A = A$ $\forall A \in \mathbb{R}^{m \times n}$
Diagonal matrix	$D = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 3 \end{bmatrix}$	
(Real) Symmetric matrix	$S = \begin{bmatrix} 4 & 2 & 1 \\ 2 & 3 & 4 \\ 1 & 4 & 3 \end{bmatrix}$	$A^T = A$
Zero matrix	$\underline{0}_3 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$	Additive inverse
Inverse matrix	$A^{-1}A = AA^{-1} = I_n$	More on inverses later

- The inverse is *only* allowable on square matrices of the form $A \in \mathbb{R}^{n \times n}$
- Not all square matrices are invertible. Invertible matrices are called *nonsingular*

1.5. Laws of vector / matrix algebra

Operation	Example	law	Notes
Law of transposition	$\left(\begin{bmatrix} 3 & 5 & 4 \\ 2 & 3 & 1 \end{bmatrix} \begin{bmatrix} 3 & 2 \\ 5 & 3 \\ 4 & 1 \end{bmatrix}\right)^T$	$(AB)^T = B^T A^T$	$(AB)^T \in \mathbb{R}^{p \times m}$
Law of transposition (with vector)	$\left(\begin{bmatrix} 3 & 5 & 4 \\ 2 & 3 & 1 \end{bmatrix} \begin{bmatrix} 3 \\ 5 \\ 3 \end{bmatrix}\right)^T$	$(A\underline{v})^T = \underline{v}^T A^T$	$(A\underline{v})^T \in \mathbb{R}^{1 \times m}$
Non-commutativity (multiplication)	$\begin{bmatrix} 3 & 5 & 4 \\ 2 & 3 & 1 \end{bmatrix} \times \begin{bmatrix} 2 & 4 \\ 3 & 1 \end{bmatrix}$	$AB \neq BA$	Multiplication with symmetric matrices *is* commutative

2. Vector spaces

I encourage you to go through these videos/transcripts before reading this section

A vector space is a *group* of mathematical objects that interact with each other in a particular way. Many groups form a vector space. Spaces of mathematical functions, matrices, images, videos, …

The '*vector*' in vector space could be many things. Not just traditional vectors like $\underline{v} = [1, 3, 2]$ (i.e. one-tensors)

If we use maths to find properties of abstract vector spaces, then these properties extend to all the actual examples of vector spaces, for free!. This is the power of *generalisation* and *abstraction*, a key part of maths.

Understanding vectors spaces doesn't feel like an immediately useful endeavour. But actually, it's key to properly understanding matrices and lots of aspects of foundational machine learning and statistics.

Just like locks need a key, vector spaces need a number field K to be defined. This field is $K = \mathbb{R}$ almost all the time, and occasionally $K = \mathbb{C}$. It's rarely anything else outside of really abstract maths. Elements of the field are referred to as *scalars*.

A simple example of a vector space is $\mathbb{R}^2$. This has a nice geometric intuition as vectors in $\mathbb{R}^2$ correspond to arrows on a Cartesian plane (see figure in Section 2.3). $\mathbb{R}^n$ is a vector space for any $n \in \mathbb{N}$, and these vector spaces pop up a lot.

2.1. Axioms of a number field

see week 1 notebook for more on axioms

- An axiom is a statement that we take to be true without questioning.
- All of mathematics is built on axioms.
- Some, like the axiom of choice, are controversial.

- A field is a type of set satisfying particular properties.
- In real life the only fields you see are $\mathbb{R}$ and $\mathbb{C}$ (real and complex numbers). Mostly $\mathbb{R}$.

Let S be a field. Then the following axioms hold

Axioms of a number field K

	ADDITION	AXIOM
Commutativity	$a+b=b+a$ $\forall a,b \in K$	A1
Associativity	$a+(b+c)=(a+b)+c$ $\forall a,b,c \in K$	A2
Zero	$\exists 0 \in K : a+0=a, \forall a \in K$	A3
Additive inverse	$\forall a \in K, \exists {-a} \in K : a+-a=$ $-a+a=0$	A4
	MULTIPLICATION	
Commutativity	$a \times b = b \times a$ $\forall a,b \in K$	M1
Transitivity	$(a \times b) \times c = a \times (b \times c)$ $\forall a,b,c \in K$	M2
One	$\exists 1 \in K : a \times 1 = 1 \times a = a$ $\forall a \in K$	M3
Multiplicative inverse	$\forall a \neq 0 \in K :$ $\exists a^{\{-1\}} \in K : a \times a^{\{-1\}} = 1$	M4
	ADDITION AND MULTIPLICATION	
Expanding brackets	$(a+b) \times c = a \times b + c \times$ $b \forall a,b,c \in K$	D1

- You use these axioms implicitly even when doing simple arithmetic. They are simple.
- This is why it is great practice to translate their complicated-looking mathematical representations into English.

Question

Identify why the sets $\mathbb{N}$ and $\mathbb{Z}$, but not $\mathbb{Q}$, fail to satisfy the requirements of a field

2.2. Axioms of a vector space

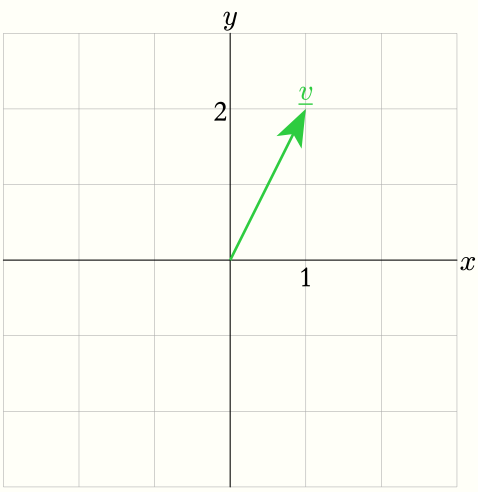
Axioms of a vector space V on a field K

Additive axioms		Axiom
Commutativity	$a+b=b+a$ $\forall a,b \in V$	A1
Associativity	$a+(b+c)=(a+b)+c$ $\forall a,b,c \in V$	A2
Zero	$\exists 0 \in V : a+0=a, \forall a \in V$	A3
Additive inverse	$\forall a \in V, \exists {-a} \in V : a+-a=-a+a=0$	A4
Distributive axioms		
Expanding Brackets 1	$a(u+v)=au+av$ $\forall a \in K, v \in V$	D1
Expanding Brackets 2	$(a+b)v=av+bv$ $\forall a,b \in K, v \in V$	D2
Multiplicative axioms		
One	$\exists 1 \in V : a \times 1 = 1 \times a = a$ $\forall a \in V$	M1
Zero	$\exists 0 \in V : a \times 0 = 0 \times a = 0$ $\forall a \in V$	M2

Repeated warning: vector confusion

- At the beginning of the cheatsheet, vectors were defined as one-tensors, e.g. $v = [1, 4, 5]$. This is common in programming.
- Now, vectors are elements of a vector space. Which might note be one-tensors. For instance, matrices are often elements of a vector space.
- You have to infer which definition is being used from context. Sorry!

2.3. Representing standard vectors (one-tensors)



The vector $\underline{v}$ depicted is usually represented as: $\underline{v} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$. Why? The short story is that this represents its x and y components. The long story is useful to know:

Let's take $\underline{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\underline{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$. Then v = 1$\underline{e}_1$ + 2$\underline{e}_2$.

The set $\{\underline{e}_1, \underline{e}_2\}$ forms a *basis* for $\mathbb{R}^2$. We will learn the definition of basis soon. How you represent a vector depends on the basis you choose.

For instance, let's take a basis $\underline{f}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\underline{f}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$. Then $v = -1\underline{f}_1 + 2\underline{f}_2$. So the representation of $\underline{v}$ in this basis is $\underline{v} = \begin{bmatrix} -1 \\ 2 \end{bmatrix}$

Why would we want to represent $\underline{v}$ differently?

- Vectors in the vector space $\mathbb{R}^2$ are easy to represent. Vectors in more abstract spaces are not. So we need the general way of doing things.
- Lots of vector *transformations* can be viewed as a change of basis. Such as the mapping from a neural network's raw inputs to a filtered representation.

Definition 2.3.1 : Linear combinations of vectors

Consider a vector space V over a field K. A linear combination of vectors $\underline{v}_1, \underline{v}_2, ...\underline{v}_n$ in V is an expression of the form $\sum_{i=1}^{n} c_i \underline{v}_i = \underline{0}$, where $c_1, c_2, ..., c_n$ are scalars in K

Example (see prev page):

$-1\underline{f}_1 + 2\underline{f}_2$ is a linear combination of the vectors $\underline{f}_1, \underline{f}_2 \in \mathbb{R}^2$, which is a vector space over the field $\mathbb{R}$.

Definition 2.3.2 : Linear dependence

Consider a vector space V over a field K. A set of vectors $\underline{v}_1, \underline{v}_2, ...\underline{v}_n$ in V are *linearly dependent* if there exists scalars $c_1, c_2, ..., c_n$ such that

$$\sum_{i=1}^{n} c_i \underline{v}_i = \underline{0} \qquad (2)$$

Otherwise, they are *linearly independent*

Another way of thinking about linear dependence is that one vector can be expressed as a linear combination of the other vectors if they are collectively linearly dependent. For instance $\underline{v}_1 = -\frac{1}{c_1} \sum_{i=2}^{n} c_i \underline{v}_i$

Definition 2.3.3: Span of a set of vectors

As before, consider a set S of vectors $\underline{v}_1, \underline{v}_2, ...\underline{v}_n$ in a vector space V over a field K. The span of S is the set of linear combinations of the elements of S. Mathematically,

$$\text{span}(S) = \left\{ \underline{v} \in V : \underline{v} = \sum_{i=1}^{n} c_i \underline{v}_i \text{ for some } c_1, c_2 ... c_n \in K \right\}$$

So $\text{span}(S)$ is the set of vectors you can *reach* by adding linear combinations of the vectors $\underline{v}_1, \underline{v}_2, ...\underline{v}_n$

- What's the span of the vector $\underline{v} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \in \mathbb{R}^2$? Can you sketch it?
- What's the span of the vectors $\left\{ \underline{v}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \underline{v}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \in \mathbb{R}^2 \right\}$?

Definition 2.3.4: Basis of a vector space

A set $S = \{\underline{v}_i\}_1^n$ of vectors in a vector space V forms a *basis* of V if

1. $\{\underline{v}_i\}_1^n$ are linearly independent
2. $V = \text{span}(S)$

- Verify for yourself that $\left\{ \underline{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \underline{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$ forms a basis of $\mathbb{R}^2$
- Verify for yourself that $\left\{ \underline{f}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \underline{f}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}$ forms a basis of $\mathbb{R}^2$

As stated at the start of this chapter, we usually use the basis $\{\underline{e}_1, \underline{e}_2\}$ to represent vectors in $\mathbb{R}^2$. It is the *canonical* basis.

Theorem 2.3.1 : (The basis theorem)

Suppose that $\underline{v}_1, \ldots, \underline{v}_m$ and $\underline{w}_1, \ldots, \underline{w}_n$ are both bases of the vector space V. Then $m = n$. In other words, all finite bases of V contain the same number of vectors.

The basis theorem means we can unambiguously define the *dimension* of a vector space as the number of elements in any basis it has.

Proving the basis theorem is difficult but possible if you want a challenge. There are various proofs online so I won't send you a particular link.

Infinite-dimensional bases

Some vector spaces are infinite dimensional. For instance, the set of square-integrable functions is a common, important one. They have a basis with (countably) infinite elements. We won't consider these in the course but it's good to know they exist.

Definition 2.3.5 : Vector subspace

Consider a vector space V on a field K. A subspace W is any set $W \subseteq V$ satisfying:

Closure under addition	$\underline{w}_1 + \underline{w}_2 \in W$ $\forall \underline{w}_1, \underline{w}_2 \in W$
Closure under scalar multiplication	$c\underline{w} \in W$ $\forall \underline{w} \in W c \in K$

In other words, if you add any two vectors in the subspace, and/or multiply them by a scalar, you stay in the subspace: it's *closed*.

- Note that the span of $\texttt{mat}[2; 4; 1] \in \mathbb{R}^3$ is a line in $\mathbb{R}^3$: it's just the set of scalar multiples of this vector. Verify it's a subspace for yourself.
- More generally the span of any set of vectors (such as a subset of a basis) is a vector subspace.As a challenge, prove this (e.g. by contradiction)!

2.4. Norms: notions of size in vector spaces

In this section we concentrate on vector space of the form $V = \mathbb{R}^n$, where $n \in \mathbb{N}$.

You can think of the representation of a vector in its basis as a position (e.g. see the Figure below Section 2.3). How *big*? is a vector $\underline{v} \in V$? Mathematically, we will denote the size, or **norm** of a vector, as $\|\underline{v}\ \|$

Any notion of size must satisfy the following properties:

PROPERTIES OF A NORM

Non-negativity	$\|\underline{v}\| \geq 0 \quad \forall \underline{v} \in V$	*Can't have a negative size*
Linear scaling	$c\ \|\underline{v}\| = \|c\underline{v}\|$ $\forall c \in K, \underline{v} \in V$	*Twice the vector means twice the size (for e.g. $c = 2$)*
Triangle inequality	$\|\underline{v} + \underline{w}\| \leq \|\underline{v}\| + \|\underline{w}\|$ $\forall \underline{v}, \underline{w} \in V$	*Going to a destination 'via' somewhere is always slower*

- Make sure you understand the description of the triangle inequality, which I've left a bit vague on purpose.

We can and do make many different notions of size that satisfy the requirements of the table above. Some common ones, which are assumed knowledge, are listed below. Their formulae assume a vector $v \in \mathbb{R}^n$:

STANDARD NORMS

L_1 norm	$\|v\|_1 = \sum_{i=1}^{n} \mid v_i \mid$
L_2 norm	$\|v\|_2 = \sqrt{\sum_{i=1}^{n} v_i^2}$
L_∞ norm	$\|v\|_\infty = \max_{i=1}^{n} \mid v_i \mid \mid$

Example: $v = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$	$\|v\|_1 = \|4\| + \|2\| = 6$ $\|v\|_2 = \sqrt{4^2 + 2^2} = \sqrt{20}$ $\|v\|_\infty = \max(4, 2) = 4$

2.5. Metrics: distance in vector spaces

We can think of the distance $d(v, w)$ between vectors $\underline{v}, \underline{w} \in V$ as the norm (size) of the vector $\underline{v} - \underline{w}$:

$$d(v, w) = \|v - w\| \tag{3}$$

More generally, any distance function $d : V \times V \to R^+$ is called a *metric* if it satisfies the following properties:

See if you can infer them from the norm properties!

PROPERTIES OF A METRIC

Positivity	$x \neq y \Rightarrow d(\underline{v}, \underline{w}) > 0$	*If they aren't the same, then they have non-negative distance between them*
Self distance	$d(\underline{v}, \underline{v}) = 0$	
Symmetry	$d(\underline{v}, \underline{w}) = d(\underline{w}, \underline{v})$	
Triangle inequality	$d(\underline{v}, \underline{u}) \leq d(\underline{v}, \underline{w}) + d(\underline{w}, \underline{u})$	*Going to a destination 'via' somewhere is always slower*

3. Matrices as transformations

Mathematics is often about transforming objects *between* vector spaces. Here are some examples

Image editing and 3d graphics	(Rotation, scaling, etc, see notebook)
Regression	Critical part of machine learning. A model predicts output features by transforming an input (e.g. predicting lifespan from medical characteristics)
Dimensionality reduction	Standard techniques like PCA and TSNE to summarise data. Or analogously compression algorithms for images/files (e.g. JPEG)
Fourier transform	Used in basically everything, but e.g. converting audio into a basis where effects can be easily applied (noise reduction, adding instruments, etc)

Definition 3.1 : Linear mapping

Consider two vector spaces V and W (e.g. $\mathbb{R}^2$ and $\mathbb{R}^3$) over the same field K (e.g. $\mathbb{R}$). A linear mapping $T : V \to W$ is a function that satisfies

$$T(\underline{v}_1) + T(\underline{v}_2) = T(\underline{v}_1 + \underline{v}_2) \forall \underline{v}_1, \underline{v}_2 \in V \tag{4}$$

$$cT(\underline{v}) = T(c\underline{v}) \forall c \in K, \underline{v} \in V \tag{5}$$

We can combine the two requirements above into one:

$$cT(\underline{v}_1) + dT(\underline{v}_2) = T(c\underline{v}_1 + d\underline{v}_2) \forall c, d \in K, \underline{v}_1, \underline{v}_2 \in V$$

- Verify for yourself that $f(x) = 2x + 1$ is a linear mapping between $\mathbb{R}$ and $\mathbb{R}$.
- When you start to view matrices at *linear transformations between vector spaces*, rather than tables of numbers, your life improves.
- Here is a great resource that visualises this.

Linear mappings are really important and useful because we can exploit linearity properties to predict a lot about how they transform unseen vectors. For finite-dimensional spaces, *every* linear mapping can be *represented* as a matrix. Just like vectors in $\mathbb{R}^2$ can be represented by an arrow on a graph. This representation makes understanding and manipulating the transformations easier.

Common notation

Domain of a map/ function	Set of possible inputs to the map
Range/image of a map / function	Set of possible outputs of the map

For a map $T : V \to W$:
- The domain of T is V.
- The range is a subset of W but might not be the entire space… what if T maps every input to $\underline{0} \in W$?

Questions

1. Consider a linear transformation $T : V \to W$ where
   - V is three dimensional and has a basis $\{\underline{e}_1, \underline{e}_2, \underline{e}_3\}$
   - W is two dimensional and has a basis $\left\{\underline{f}_1, \underline{f}_2\right\}$

Now suppose

$T(\underline{e}_1) = \underline{f}_1 + \underline{f}_2$ $\qquad$ $T(\underline{e}_2) = \underline{f}_1 - 3\underline{f}_2$ $\qquad$ $T(\underline{e}_3) = \underline{f}_1 - 2\underline{f}_2$

- Can you use linearity to express $T(4\underline{e}_1 + 5\underline{e}_2)$?
- Can you use linearity to express $T(c\underline{e}_1 + d\underline{e}_2)$ for any $c, d \in K = \mathbb{R}$?

The above question should convince you that knowing how a linear map $T : V \to W$ transforms the basis vectors of V tells you how it transforms *any* vector $v \in V$. Because v after all must be a linear combination of basis elements.

We can answer the question above more simply! (Only read once you've done that question). Consider the matrix:

$M = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -3 & -2 \end{bmatrix}$

- What is the relationship to the linear transformation in the question? Ask yourself before reading on.

M is a representation of T. Specifically let $x = \begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix}$ represent the vector $3\underline{e}_1 + 1\underline{e}_2 + 2\underline{e}_3 \in V$.

1. Use matrix multiplication to find Mx
2. Find the vector $T(x) \in W$, represented as a linear sum of the basis $\{\underline{f}_1, \underline{f}_2\}$.

… they should be the same!

This becomes clearer if you think of matrix multiplication as

$$\begin{aligned} \begin{bmatrix} 1 & 1 & 1 \\ 1 & -3 & -2 \end{bmatrix} \begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix} &= 3\begin{bmatrix} 1 \\ 1 \end{bmatrix} + 1\begin{bmatrix} 1 \\ -3 \end{bmatrix} + 2\begin{bmatrix} 1 \\ -2 \end{bmatrix} \\ &= 3T(\underline{e}_1) + 1T(\underline{e}_2) + 2T(\underline{e}_3) \end{aligned}$$

Definition 3.2: **Range (/image) of a transformation**

This is the set of vectors you can *reach* through the transformation:

$$\{\underline{w} \in W : \underline{w} = T(\underline{v}), \text{ for some } \underline{v} \in V\} \quad (6)$$

Theorem 3.1: The range of a linear map is a vector space

- As a challenge, prove this for yourself!

3.1. Matrix equations

Definition 3.1.1 : Matrix equation

A matrix equation is an equation of the form

$$A\underline{v} = \underline{w} \tag{7}$$

where

$A \in \mathbb{R}^{m\times n}$ is a *matrix*

$\underline{v} \in \mathbb{R}^{n}$ is an *unknown vector*

$\underline{w} \in \mathbb{R}^{m}$ is a *known vector*

Example

$$\begin{bmatrix} 1 & 3 \\ 5 & 2 \\ 6 & 4 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} \tag{8}$$

This can equivalently be written as a *simultaneous equation*:

$$\begin{aligned} v_1 + 3v_2 &= 4 \\ 5v_1 + 2v_2 &= 5 \\ 6v_1 + 4v_2 &= 6 \end{aligned} \tag{9}$$

- Notice there are *three constraints* (each equals sign is an extra constraint)
- Notice there are *two free variables* (v_1 and v_2)

A few observations:

- The subspace of possible solutions on the left-hand-side (LHS) of the equation can be written as

$$\left\{ v_1 \begin{bmatrix} 1 \\ 5 \\ 6 \end{bmatrix} + v_2 \begin{bmatrix} 3 \\ 2 \\ 4 \end{bmatrix} \text{ where } v_1, v_2 \in \mathbb{R} \right\} \tag{10}$$

- This is the set of linear combinations of *two* vectors. So it can only form, at most, a vector subspace with dimension *two*.
- But the result lives in $\mathbb{R}^3$, which is a *three* dimensional space.
- Hence we know that there might not be a solution.

Example

$$\begin{bmatrix} 1 & 0 & 6 \\ 0 & 1 & 4 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} 4 \\ 5 \end{bmatrix} \tag{11}$$

- Note that the set of possible solutions on the LHS is now

$$\left\{ v_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + v_2 \begin{bmatrix} 0 \\ 1 \end{bmatrix} + v_3 \begin{bmatrix} 6 \\ 4 \end{bmatrix} \right\} \tag{12}$$

- The first two vectors in this solution set form a basis of $\mathbb{R}^2$, so we can reach anything on the RHS. Hence there must be a solution to this equation.
- There are two constraints but *three* free variables. So there might actually be multiple (infinite!) solutions.

Question

- For the example above, can you find a vector $\underline{u} \in \mathbb{R}^3$ such that $A\underline{u} = \underline{0}$?
- If $A\underline{u} = 0$, then what is $A(c\underline{u})$ for any $c \in \mathbb{R}$?
- Suppose you found a single solution $\underline{v}$ satisfying $A\underline{v} = \begin{bmatrix} 4 \\ 5 \end{bmatrix}$. Why must there be infinite solutions?

Example

$$\begin{bmatrix} 1 & 3 & 2 \\ 2 & 6 & 4 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} 4 \\ 5 \end{bmatrix} \tag{13}$$

- Note that the set of possible solutions on the LHS is now

$$\left\{ v_1 \begin{bmatrix} 1 \\ 2 \end{bmatrix} + v_2 \begin{bmatrix} 3 \\ 6 \end{bmatrix} + v_3 \begin{bmatrix} 2 \\ 4 \end{bmatrix} \right\} \tag{14}$$

- Initially, one might think this is thet set of linear combinations of *three* vectors, so it could form a three-dimensional vector subspace
- However, each vector is in $\mathbb{R}^2$. You can't have more than two linearly independent vectors in $\mathbb{R}^2$ by the basis theorem…it's a two-dimensional space!
- In fact *all* of these vectors are parallel, they are scalar multiples of each other! So the solution set is actually a *one* dimensional subspace, a line drawn on $\mathbb{R}^2$.
- As such, there will only be a solution if the RHS $\begin{bmatrix} 4 \\ 5 \end{bmatrix}$ lies on the line. But it doesn't so there is no solution.

Definition 3.1.2: Kernel of a matrix / linear map

Consider a linear map $T : V \to W$, where V, W are vector spaces on a field K. The kernel of T is

$$\ker(T) = \{v \in V : T(v) = \underline{0} \in W\} \tag{15}$$

Now suppose $V = \mathbb{R}^n$, $W = \mathbb{R}^m$, and $K = \mathbb{R}$. Consider a matrix $A \in \mathbb{R}^{m \times n}$ representing the linear transformation T Equivalently, the kernel of A is

$$\ker(A) = \{v \in \mathbb{R}^N : Av = \underline{0} \in \mathbb{R}^m\} \tag{16}$$

Note that the kerneel is a vector subspace: it is closed under addition and scalar multiplication.

Definition 3.1.3: Image of a matrix / linear map

Consider a linear map $T : V \to W$, where V, W are vector spaces on a field K. The image of T is

$$\mathrm{Im}(T) = \{w \in W : w = Tv \quad \text{for some } v \in V\} \tag{17}$$

For example, consider the matrix $A = \begin{bmatrix} 1 & 3 \\ 5 & 5 \\ 2 & 4 \end{bmatrix}$ representing a linear map $T : \mathbb{R}^2 \to \mathbb{R}^3$

The image of A is

$$\mathrm{Im}(A) = \left\{ v_1 \begin{bmatrix} 1 \\ 5 \\ 2 \end{bmatrix} + v_2 \begin{bmatrix} 3 \\ 5 \\ 4 \end{bmatrix}, \text{ where } \ v_1, v_2 \in \mathbb{R} \right\}$$

Just like the Kernel, you can prove to yourself that the image of a matrix is a vector subspace.

Common notation	
Rank of a matrix	Dimension of the vector subspace constituting the *image*
Nullity of a matrix	Dimension of the vector subspace constituting the *kernel*

Note that $\mathrm{Ker}(A)$ might consist of a single element: $\underline{0}$. In this case the nullity is zero.

Example

$$A = \begin{bmatrix} 3 & 0 \\ 0 & 4 \end{bmatrix} \qquad A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \qquad A = \begin{bmatrix} 1 & 2 & 4 \\ 2 & 4 & 3 \end{bmatrix}$$

What are $\mathrm{Ker}(A)$ and $\mathrm{Im}(A)$ for these examples? (Last one is harder)

Theorem 3.1.1 : Rank-nullity theorem

The number of columns in a matrix A is equal to the rank of the matrix plus the nullity of the matrix.

A short explanation of why is that you can divide the columns of a matrix into a linearly independent set, and a set that are linearly dependent,

Example

$$A = \begin{bmatrix} 1 & 1 & 2 \\ 0 & 1 & 0 \end{bmatrix} \tag{18}$$

The first two columns are linearly independent. The third column is dependent (it's a scalar multiple of the first column).

- The linearly independent columns form a basis for $\mathrm{Im}(A)$:

$$\mathrm{Im}(A) = \left\{ v_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + v_2 \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \forall v_1, v_2 \in \mathbb{R} \right\} \tag{19}$$

This means $\mathrm{Im}(A)$ encompasses all of $\mathbb{R}^2$ as it is a two-dimensional vector subspace. In particular, this means we can solve

$$v_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + v_2 \begin{bmatrix} 1 \\ 1 \end{bmatrix} = -v_3 \begin{bmatrix} 2 \\ 0 \end{bmatrix} \tag{20}$$

for any given value of v_3. Any solution (v_1, v_2, v_3) is in the Kernel. Adding any two solutions, or multiplying them by a scalar, won't change the equality, as $0 + 0 = 0$ and $0 \times c = 0$.

Therefore Equation 20 defines a one-dimensional subspace forming the kernel of A.

The rank nullity theorem says 2+1 = 3 in this case.

Common notation

Nonsingular matrix	A square matrix whose rank is equal to the number of columns
Invertible matrix	Another name for a nonsingular matrix: all nonsingular matrices have an inverse!

Note that $\mathrm{Ker}(A)$ might consist of a single element: $\underline{0}$. In this case the nullity is zero. We say that A has a *trivial* Kernel.

Question

Can you prove why a square matrix with a nontrivial Kernel can't have an inverse? Try a proof by contradiction: suppose it does have an inverse and see if that assumption leads to an absurdity.

3.2. Inverses

Definition 3.2.1 : Inverse of a linear map

Consider a linear map $T : V \to W$, where V, W are vector spaces on a field K.

A linear map $I : W \to V$ is a *right inverse* if

$$T(I(\underline{w})) = \underline{w} \quad \forall \underline{w} \in V \tag{21}$$

A linear map $I : W \to V$ is a *left inverse* if

$$I(T(\underline{v})) = \underline{v} \quad \forall \underline{v} \in W \tag{22}$$

If a linear map I is both a right inverse and a left inverse of T, then we call it the inverse of T, and denote it T^{-1}

- A matrix has a right inverse if its rank equals its number of rows
- A matrix has a left inverse if its rank equals its number of columns
- So only a square matrix has a full inverse. Can you think of geometric intuition as to why?

3.3. Determinants

- The determinant of a square matrix is a number associated with the matrix. EG you could have $\det(A) = 4$
- The formula is complicated and unimportant. Just go `np.linalg.det(A)`!

What's the intuition behind it? It measures how much the linear transformation associated with a matrix shrinks/expands the basis vectors of the domain vector space. Rather than writing lots, I'd suggest you go to the transcript here.

3.4. Eigenvectors and eigenvalues

Definition 3.4.1 : Eigenvectors/values of a matrix

Consider a linear map $T : V \to W$, where V, W are vector spaces on a field K. A vector $v \in V$ is an *eigenvector* of T if

$$Tv = \lambda v, \text{ where } \lambda \in K. \tag{23}$$

In this case, λ is known as the associated *eigenvalue*.
In other words, T *scales* v, but doesn't change its direction. The degree of scaling is λ

Questions

1. Why is every element of $\mathrm{Ker}(A)$ an eigenvector of K? What is the corresponding eigenvalue?
2. Consider the matrix

$$R = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \tag{24}$$

- What are it's eigenvectors and eigenvalues? (Hint, they might be complex numbers)

Again, instead of overloading the cheatsheet, I encourage you to go through the 3blue1brown transcript on eigenvectors