Vectors and Matrices: outline notes

John Magorrian
john.magorrian@physics.ox.ac.uk

11 November 2021

Week 1

Introduction

1.1 Vectors

When you think of a "vector" you might think of one or both of the following ideas:

- a physical quantity that has both length and direction
  - e.g., displacement, velocity, acceleration, force, electric field,
  - angular momentum, torque, ...
- a column of numbers (coordinate vector)

The mathematical definition of a vector generalises these familiar notions by redefining vectors in terms of their relationship to other vectors: vectors are objects that live (with other like-minded vectors) in a so-called **vector space**; anything that inhabits a vector space is then a vector. A vector space is a set of vectors, equipped with operations of addition of two vectors and multiplication of vectors by scalars (i.e., scaling) that must obey certain rules. These distil the fundamental essence of "vector" into a small set of rules that are easily generalisable to, say, spaces of functions. From this small set of rules we can introduce concepts like basis vectors, coordinates and the dimension of the vector space. We'll meet vector spaces in week 3, after limbering up by recalling some of the more familiar geometric properties of 2d and 3d vectors next week (week 2).

In these notes I use a calligraphic font to denote vector spaces (e.g., $\mathcal{V}$, $\mathcal{W}$), and a bold font for vectors (e.g., $\mathbf{a}$, $\mathbf{b}$).

1.2 Matrices

You have probably already used matrices to represent

- geometrical transformations, such as rotations or reflections, or
- system of simultaneous linear equations.

But, just as we'll generalise the notion of vector, we'll also generalise that of matrix: a matrix represents a **linear map** from one vector space to another. A map $f : \mathcal{V} \to \mathcal{W}$ from one vector space $\mathcal{V}$ to another space $\mathcal{W}$ is linear if it satisfies the pair of conditions

$$\begin{aligned} f(\mathbf{a}+\mathbf{b}) &= f(\mathbf{a}) + f(\mathbf{b}), \\ f(\alpha\mathbf{a}) &= \alpha f(\mathbf{a}), \end{aligned} \tag{1.1}$$

for any vectors $\mathbf{a}, \mathbf{b} \in \mathcal{V}$ and for any scalar value α. An alternative way of writing these **linearity conditions** is that

$$f(\alpha\mathbf{a} + \beta\mathbf{b}) = \alpha f(\mathbf{a}) + \beta f(\mathbf{b}), \tag{1.2}$$

for any pair of vectors $\mathbf{a}, \mathbf{b} \in \mathcal{V}$ and any pair of scalars α, β. Examples of linear maps include rotations, reflections, shears and projections. Linear maps are fundamental to undergraduate physics: the first step in understanding a complex system is nearly always to "linearize" it: that is, to approximate its equations of motion by some form of linear map, $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x})$. We'll spend most of this course studying the basic properties of these maps and learning different ways of characterising them.

Even though our intuition about vectors and linear maps comes directly from 2d and 3d real geometry, in this course we'll mostly focus on their *algebraic* properties: that is why this subject is usually called "linear algebra".

1.3 How to use these notes

These notes serve only as a broad overview of what I cover in lectures and to give you the opportunity to read ahead. In parallel with the lectures I expect you to read Andre Lukas' excellent lecture notes and will refer to them throughout the course. For alternative treatments see your college library: it will be bursting with books on this topic, from the overly formal to the most practical "how-to" manuals. *Mathematical methods for physics and engineering* by Riley, Hobson & Bence is a very good source of examples beyond what I can cover in lectures.

Week 2

Back to school: vectors and matrices

2.1 Coordinates and bases

Let's start with familiar examples of vectors, such as the displacement $\mathbf{r}$ of a particle from a reference point O, its velocity $\mathbf{v} = \dot{\mathbf{r}}$, or maybe the electric field strength $\mathbf{E}$ at its location. These vectors "exist" independent of how we think of them. For calculations, however, it is usually convenient to express them in terms of their **coordinates**. That is, we choose a set of **basis vectors**, say $(\mathbf{i}, \mathbf{j}, \mathbf{k})$, or $(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$, that span 3d space and expand AL p.7

$$\mathbf{r} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}, \tag{2.1}$$

or

$$\mathbf{r} = r_1\mathbf{e}_1 + r_2\mathbf{e}_2 + r_3\mathbf{e}_3, \tag{2.2}$$

and then use these coordinates (x, y, z) or (r_1, r_2, r_3) for doing calculations. The numerical values of these coordinates depend on our choice of basis vectors.

Now let us identify the basis vectors with 3d column vectors, like this:

$$\mathbf{e}_1 \leftrightarrow \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \mathbf{e}_2 \leftrightarrow \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad \mathbf{e}_3 \leftrightarrow \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}. \tag{2.3}$$

Then the vector $\mathbf{r}$ in (2.2) corresponds to the column vector

$$\mathbf{r} \leftrightarrow \begin{pmatrix} r_1 \\ r_2 \\ r_3 \end{pmatrix} = r_1 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + r_2 \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + r_3 \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \tag{2.4}$$

This $(r_1, r_2, r_3)^{\mathrm{T}}$ is known as the **coordinate vector** for $\mathbf{r}$. Once we've chosen a set of basis vectors, every physical vector $\mathbf{r}$, $\mathbf{v}$, $\mathbf{E}$ has a corresponding coordinate vector and vice versa, and this correspondence respects the rules of vector arithmetic: if, say,

$$\mathbf{r}(r) = \mathbf{r}_0 + \mathbf{v}t, \tag{2.5}$$

the the corresponding coordinate vectors satisfy the same relation. For that reason we often don't distinguish between "physical" vectors and their corresponding coordinate vectors, but in this course we'll be careful to make that distinction.

Throughout the following I'll use the convention that scalar indices a_1, a_2, a_3 etc refer to coordinates of the vector $\mathbf{a}$.

2.2 Products

Apart from adding vectors and scaling them, there are a number of other ways of operating on pairs or triplets of vectors, which are naturally defined using by appealing to familiar geometrical ideas:[1]

- The scalar product is AL p.21
$$\mathbf{a}\cdot\mathbf{b} \equiv |\mathbf{a}|\,|\mathbf{b}|\,\cos\theta, \tag{2.6}$$
where $|\mathbf{a}|$ and $|\mathbf{b}|$ are the *lengths* or magnitudes of $\mathbf{a}$ and $\mathbf{b}$ and θ is the angle between them. The result is a scalar, hence the name. Geometrically, it measures the projection of the vector $\mathbf{b}$ along the direction of $\mathbf{a}$.

- The vector product of two vectors is another vector, AL p.24
$$\mathbf{a}\times\mathbf{b} \equiv |\mathbf{a}|\,|\mathbf{b}|\,\sin\theta\,\hat{\mathbf{n}}, \tag{2.7}$$
where θ again is the angle between $\mathbf{a}$ and $\mathbf{b}$ and now $\hat{\mathbf{n}}$ is a unit vector perpendicular to both $\mathbf{a}$ and $\mathbf{b}$, oriented such that $(\mathbf{a},\mathbf{b},\hat{\mathbf{n}})$ are a right-handed triple. Geometrically, the magnitude AL p.30 of the vector product is equal to the area spanned by the vectors $\mathbf{a}$ and $\mathbf{b}$.

- The triple product of the vectors $\mathbf{a},\mathbf{b},\mathbf{c}$ is simply AL p.28
$$\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c}). \tag{2.8}$$
This is the (oriented) volume of the parallelepiped spanned by $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$.

The scalar product is clearly commutative: $\mathbf{b}\cdot\mathbf{a} = \mathbf{a}\cdot\mathbf{b}$.[2] Notice that the (square of the) length of a vector is given by the scalar product of the vector with itself:
$$|\mathbf{a}|^2 = \mathbf{a}\cdot\mathbf{a}, \tag{2.9}$$
because $\cos\theta = 1$ in this case. Two non-zero vectors $\mathbf{a}$ and $\mathbf{b}$ are perpendicular ($\theta = \pm\pi/2$) if and only if $\mathbf{a}\cdot\mathbf{b} = 0$. The vector product is *anti*-commutative: $\mathbf{b}\times\mathbf{a} = -\mathbf{a}\times\mathbf{b}$.

The scalar and vector products are both linear:[3]
$$\begin{aligned} \mathbf{c}\cdot(\alpha\mathbf{a}+\beta\mathbf{b}) &= \alpha\mathbf{a}\cdot\mathbf{b}+\beta\mathbf{a}\cdot\mathbf{b},\\ \mathbf{c}\times(\alpha\mathbf{a}+\beta\mathbf{b}) &= \alpha\mathbf{a}\times\mathbf{b}+\beta\mathbf{a}\times\mathbf{b}. \end{aligned} \tag{2.10}$$

Because the triple product is made up of a scalar and a vector product, each of which is linear in each of its arguments, it follows that the triple product itself is also linear in each of $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$.

Later we'll see that the scalar product can be defined for vectors of any dimension. In constrast, the vector and triple products apply only to three-dimensional vectors.[4]

[1]The corresponding algebraic definitions of these will follow later in the course.

[2]They commute if $\mathbf{a}$ and $\mathbf{b}$ are *real* vectors; for complex vectors the relation is slightly more complex (week NN).

[3]Again, true for real vectors; only half true for complex vectors.

[4]After your first pass through the course have a think about how they might be generalised to spaces of arbitrary dimension

2.3 Turning geometry into algebra

Lines Referred to some origin O, the displacement vector of points on a line passing through the point with displacement vector $\mathbf{p}$ and running parallel to $\mathbf{q}$ satisfies joining them is AL p.33

$$\mathbf{r}(\lambda) = \mathbf{p} + \lambda\mathbf{q}, \tag{2.11}$$

the position along the line being controlled by the parameter $\lambda \in \mathbb{R}$. Expressing this in coordinates and rearranging, we have

$$\lambda = \frac{r_1 - p_1}{q_1} = \frac{r_2 - p_2}{q_2} = \frac{r_3 - p_3}{q_3}. \tag{2.12}$$

Planes Similarly, AL p.36

$$\mathbf{r}(\lambda_1, \lambda_2) = \mathbf{p} + \lambda_1\mathbf{q}_1 + \lambda_2\mathbf{q}_2 \tag{2.13}$$

is the parametric equation of the plane spanned by $\mathbf{q}_1$ and $\mathbf{q}_2$ that passes through $\mathbf{p}$. A neater alternative way of representing a plane is

$$\mathbf{r} \cdot \hat{\mathbf{n}} = d. \tag{2.14}$$

where $\hat{\mathbf{n}}$ is the unit vector normal to the plane and d the shortest distance between the plane and the origin $\mathbf{r} = \mathbf{0}$.

Spheres Points on the sphere of radius R centred on the point displaced $\mathbf{a}$ from O satisfy AL p.40

$$|\mathbf{r} - \mathbf{a}|^2 = R^2. \tag{2.15}$$

Exercises

1. Show that the shortest distance d_{min} between the point $\mathbf{p}_0$ and the line $\mathbf{r}(\lambda) = \mathbf{p} + \lambda\mathbf{q}$ is given by

$$d_{\text{min}} = \frac{|(\mathbf{p} - \mathbf{p}_0) \times \mathbf{q}|}{|\mathbf{q}|}. \tag{2.16}$$

What is the minimum distance of the point $(1, 1, 1)$ from the line AL p.36

$$\frac{x-2}{3} = -\frac{y+1}{5} = \frac{z-4}{2}? \tag{2.17}$$

2. Show that the shortest distance between the pair of lines $\mathbf{r}_1(\lambda_1) = \mathbf{p}_1 + \lambda_1\mathbf{q}_1$ and $\mathbf{r}_2(\lambda_2) = \mathbf{p}_2 + \lambda_2\mathbf{q}_2$ is given by

$$d_{\text{min}} = |(\mathbf{p}_2 - \mathbf{p}_1) \cdot \hat{\mathbf{n}}|, \tag{2.18}$$

where $\hat{\mathbf{n}} = (\mathbf{q}_1 \times \mathbf{q}_2)/|\mathbf{q}_1 \times \mathbf{q}_2|$.

3. Show that the line $\mathbf{r}(\lambda) = \mathbf{P} + \Lambda\mathbf{Q}$ intersects the plane $\mathbf{r}(\lambda_1, \lambda_2) = \mathbf{p} + \lambda_1\mathbf{q}_1 + \lambda_2\mathbf{q}_2$ at the point $\mathbf{r} = \mathbf{P} + \Lambda_0\mathbf{Q}$, where the parameter

$$\Lambda_0 = \frac{(\mathbf{p} - \mathbf{P}) \cdot (\mathbf{q}_1 \times \mathbf{q}_2)}{\mathbf{Q} \cdot (\mathbf{q}_1 \times \mathbf{q}_2)}. \tag{2.19}$$

This is well defined unless the denominator $\mathbf{Q} \cdot (\mathbf{q}_1 \times \mathbf{q}_2) = 0$. Explain geometrically what happens in this case.

4. What is the minimum distance of the point $(1, 1, 1)$ from the plane

$$\mathbf{r} = \begin{pmatrix} 2 \\ -1 \\ 4 \end{pmatrix} + \lambda_1 \begin{pmatrix} 3 \\ -5 \\ 2 \end{pmatrix} + \lambda_2 \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}? \tag{2.20}$$

2.4 Orthonormal bases

Any $(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$ with $\mathbf{e}_1 \cdot (\mathbf{e}_2 \times \mathbf{e}_3) \neq 0$ is a basis for 3d space, but it nearly always makes sense to choose an **orthonormal basis**, which is one that satisfies

$$\mathbf{e}_i \cdot \mathbf{e}_j = \delta_{ij} \tag{2.21}$$

for any choice of i, j, where AL p.25

$$\delta_{ij} \equiv \begin{cases} 1, & i = j, \\ 0, & i \neq j, \end{cases} \tag{2.22}$$

is known as the **Kronecker delta**. Then, using the linearity of the scalar product, we have that

$$\begin{aligned} \mathbf{a} \cdot \mathbf{b} &= (a_1\mathbf{e}_1 + a_2\mathbf{e}_2 + a_3\mathbf{e}_3) \cdot (b_1\mathbf{e}_1 + b_2\mathbf{e}_2 + b_3\mathbf{e}_3) = \cdots \\ &= a_1b_1 + a_2b_2 + a_3b_3. \end{aligned} \tag{2.23}$$

Similarly, to find, say, the coordinate a_1, just do

$$\mathbf{e}_1 \cdot \mathbf{a} = \mathbf{e}_1 \cdot (a_1\mathbf{e}_1 + a_2\mathbf{e}_2 + a_3\mathbf{e}_3) = a_1 \underbrace{\mathbf{e}_1 \cdot \mathbf{e}_1}_{\delta_{11}=1} + a_2 \underbrace{\mathbf{e}_1 \cdot \mathbf{e}_2}_{\delta_{12}=0} + a_3 \underbrace{\mathbf{e}_1 \cdot \mathbf{e}_3}_{\delta_{13}=0} = a_1. \tag{2.24}$$

More generally, $a_i = \mathbf{e}_i \cdot \mathbf{a}$ for an orthonormal basis.

2.5 The Levi-Civita (alternating) symbol

Let $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ be orthonormal. From the definition (2.21) this means that

$$\mathbf{e}_i \times \mathbf{e}_j = \pm \mathbf{e}_k \tag{2.25}$$

when $i \neq j \neq k \neq i$. It's conventional to choose the sign according to the **right-handed rule**:

$$\begin{aligned} \mathbf{e}_1 \times \mathbf{e}_2 &= \mathbf{e}_3, \\ \mathbf{e}_2 \times \mathbf{e}_3 &= \mathbf{e}_1, \\ \mathbf{e}_3 \times \mathbf{e}_1 &= \mathbf{e}_2. \end{aligned} \tag{2.26}$$

Then (expand it out!)

$$\mathbf{a} \times \mathbf{b} = (a_2b_3 - a_3b_2)\mathbf{e}_1 + (a_3b_1 - a_1b_3)\mathbf{e}_2 + (a_1b_2 - a_2b_1)\mathbf{e}_3. \tag{2.27}$$

Now we can go algebraic on the vector product. Let $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ be a right-handed, orthonormal basis. Then we have that AL p.25

$$\mathbf{e}_i \times \mathbf{e}_j = \sum_{k=1}^{3} \epsilon_{ijk} \mathbf{e}_k, \tag{2.28}$$

where the **alternating** or **Levi-Civita** symbol is defined to be

$$\epsilon_{ijk} \equiv \begin{cases} +1, & \text{if } (i,j,k) \text{ is even permutation of } (1,2,3), \\ -1, & \text{if } (i,j,k) \text{ is odd permutation of } (1,2,3), \\ 0, & \text{otherwise.} \end{cases} \tag{2.29}$$

That is, $\epsilon_{123} = \epsilon_{231} = \epsilon_{312} = 1$, $\epsilon_{213} = \epsilon_{132} = \epsilon_{321} = -1$ and $\epsilon_{ijk} = 0$ for all other 21 possible choices of (i, j, k) .

Now we can easily express vector equations in terms of their coordinates:

$$\begin{aligned}
\mathbf{r} = \mathbf{p} + \lambda\mathbf{q} \quad &\rightarrow \quad r_i = p_i + \lambda q_i, \\
d = \mathbf{a}\cdot\mathbf{b} \quad &\rightarrow \quad d = \sum_{i=1}^{3} a_i b_i, \\
\mathbf{c} = \mathbf{a}\times\mathbf{b} \quad &\rightarrow \quad c_i = \sum_{j=1}^{3}\sum_{k=1}^{3} \epsilon_{ijk} a_j b_k, \\
\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c}) \quad &\rightarrow \quad \sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{k=1}^{3} \epsilon_{ijk} a_i b_j c_k.
\end{aligned}$$

Exercises

1. Show that $\mathbf{a}\cdot(\mathbf{a}\times\mathbf{c}) = 0$.

2. Show that the triple product $\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c})$ is unchanged under cyclic interchange of $\mathbf{a}, \mathbf{b}, \mathbf{c}$. Corrected from $\mathbf{a}\times(\mathbf{b}\times\mathbf{c})$!

3. Use the (important) identity AL p.25

$$\sum_{i=1}^{3} \epsilon_{ijk}\epsilon_{ilm} = \delta_{jl}\delta_{km} - \delta_{jm}\delta_{kl}, \tag{2.30}$$

to show that

$$\begin{aligned}
\sum_{i=1}^{3}\sum_{j=1}^{3} \epsilon_{ijk}\epsilon_{ijm} &= 2\delta_{km}, \\
\sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{k=1}^{3} \epsilon_{ijk}\epsilon_{ijk} &= 6,
\end{aligned} \tag{2.31}$$

and to prove the following relations involving the vector product: AL p.26

$$\begin{aligned}
\mathbf{a}\times(\mathbf{b}\times\mathbf{c}) &= (\mathbf{a}\cdot\mathbf{c})\mathbf{b} - (\mathbf{a}\cdot\mathbf{b})\mathbf{c}, \\
(\mathbf{a}\times\mathbf{b})\cdot(\mathbf{c}\times\mathbf{d}) &= (\mathbf{a}\cdot\mathbf{c})(\mathbf{b}\cdot\mathbf{d}) - (\mathbf{a}\cdot\mathbf{d})(\mathbf{b}\cdot\mathbf{c}), \\
\mathbf{a}\cdot(\mathbf{a}\times\mathbf{b}) &= 0.
\end{aligned} \tag{2.32}$$

2.6 Matrices

If you're not familiar with the rules of matrix arithmetic see §3.2.1 of Lukas' notes or §8.4 of Riley, Hobson & Bence. Here is a recap: AL p.51

- An $n \times m$ matrix has n rows and m columns.
- An n-dimensional column vector is also an $n \times 1$ matrix.
- Write A_{ij} for the **matrix element** at row i, column j of the matrix A.
- If A is an $n \times m$ matrix and B is an $m \times p$ one then their product AB is an $n \times p$ matrix. Written out in terms of matrix elements, the equation $C = AB$ becomes AL p.57

$$C_{ik} = \sum_{k=1}^{m} A_{ij} B_{jk}. \tag{2.33}$$

The important message of this section is that every $n \times m$ matrix is a linear map from n-dimensional column vectors to m-dimensional ones. Conversely, any linear map from n-dimensional column vectors to m-dimensional ones can be represented by some $n \times m$ matrix, as we'll now show.

2.6.1 Any linear map can be represented by a matrix

Let B be a beast that eats m-dimensional vectors and emits n-dimensional ones.We want it to satisfy the linearity condition AL p.56

$$B(\alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2) = \alpha_1 B(\mathbf{v}_1) + \alpha_2 B(\mathbf{v}_2). \tag{2.34}$$

What's the most general such B?

Choose *any* basis $\mathbf{e}_1^m, ..., \mathbf{e}_m^m$ for B's mouth and $\mathbf{e}_1^n, ..., \mathbf{e}_n^n$ for the nuggets it emits. Examine these nuggets when fed some $\mathbf{v}$:

$$B(\mathbf{v}) = B\left(\sum_{j=1}^{m} v_j \mathbf{e}_j^m\right) = \sum_{j=1}^{m} v_j B(\mathbf{e}_j^m). \tag{2.35}$$

Each emission $B(\mathbf{e}_j^m)$ is an n-dimensional nugget and so can be expanded as

$$B(\mathbf{e}_j^m) = \sum_{i=1}^{n} B_{ij} \mathbf{e}_i^n, \tag{2.36}$$

for some set of coefficients B_{ij}. Writing $\mathbf{w} = B(\mathbf{v})$ we have

$$\sum_{i=1}^{n} w_i \mathbf{e}_i^n = \sum_{j=1}^{m} v_j B(\mathbf{e}_j^m) = \sum_{i=1}^{n} \sum_{j=1}^{m} v_j B_{ij} \mathbf{e}_i^n. \tag{2.37}$$

That is, the coordinates w_i of the output are given by $w_i = \sum_{j=1}^{m} B_{ij} v_j$. But the RHS of this expression is simply (pre)multiplication of the coordinate vector $(v_1, ...v_m)^T$ by the $n \times m$ matrix having elements B_{ij}.

2.6.2 Terminology

- **Square** matrices have as many rows as columns.
- The **(main) diagonal** of a matrix A consists of the elements $A_{11}, A_{22}, ...$ (i.e., top left to bottom right). A **diagonal matrix** is one for which $A_{ij} = 0$ when $i \neq j$; all elements off the main diagonal are zero.
- The n-dimensional **identity matrix** is the $n \times n$ diagonal matrix with elements $I_{ij} = \delta_{ij}$.

- Given an $n \times m$ matrix A, its **transpose** A^T is the $m \times n$ matrix with elements $[A^T]_{ij} = A_{ji}$. Its **Hermitian conjugate** $A^\dagger$ has elements $[A^\dagger]_{ij} = A^\star_{ji}$.
- If $A^T = A$ then A is **symmetric**. If $A^\dagger = A$ it is **Hermitian**.
- If $AA^T = A^T A = I$ then A is **orthogonal**. If $AA^\dagger = A^\dagger A = I$ then A is **unitary**.
- Suppose A and B are both $n \times n$ matrices. Then in general $AB \neq BA$. The difference

$$[A, B] \equiv AB - BA \tag{2.38}$$

is called the **commutator** of A and B. It's another $n \times n$ matrix. If it is equal to the $n \times n$ zero matrix then A and B are said to **commute**. AL p.59

Exercises

1. A 2×2 matrix is a linear map from the plane (2d column vectors) onto itself. It is completely defined by 4 numbers. Identify the different types of transformation that can be constructed from these 4 numbers.

2. A 3×3 matrix maps 3d space onto itself. It has 9 free parameters. Identify the geometrical transformations that it can effect.

3. Given a $n \times n$ matrix with complex coefficients, how to characterise the map it represents?

4. Show that $(AB)^T = B^T A^T$ and $(AB)^\dagger = B^\dagger A^\dagger$. Hence show that $AA^T = I$ implies $A^T A = I$.

5. Show that the map $B(\mathbf{x}) = \mathbf{n} \times \mathbf{x}$ is linear. Find the find the matrix that represents it. AL p.57

Week 3

Vector spaces and linear maps

3.1 Vector spaces

A vector space consists of AL p.10

- a set $\mathcal{V}$ of **vectors**,
- a field $\mathcal{F}$ of **scalars**
- a rule for adding two vectors to produce another vector,
- a rule for multiplying vectors by scalars,

that together satisfy the following conditions: AL p.11

1. The set $\mathcal{V}$ of vectors is closed under addition, i.e.,
$$\mathbf{a} + \mathbf{b} \in \mathcal{V} \quad \text{for all } \mathbf{a}, \mathbf{b} \in \mathcal{V}; \tag{3.1}$$
2. $\mathcal{V}$ is also closed under multiplication by scalars, i.e.,
$$\alpha\mathbf{a} \in \mathcal{V} \quad \text{for all } \mathbf{a} \in \mathcal{V} \text{ and } \alpha \in \mathcal{F}; \tag{3.2}$$
3. $\mathcal{V}$ contains a special zero vector, $\mathbf{0} \in \mathcal{V}$, for which
$$\mathbf{a} + \mathbf{0} = \mathbf{a} \quad \text{for all } \mathbf{a} \in \mathcal{V}; \tag{3.3}$$
4. Every vector has an additive inverse: for all $\mathbf{a} \in \mathcal{V}$ there is some $\mathbf{a}' \in \mathcal{V}$ for which
$$\mathbf{a} + \mathbf{a}' = \mathbf{0}; \tag{3.4}$$
5. The vector addition operator must be commutative and associative:
$$\begin{aligned} \mathbf{a} + \mathbf{b} &= \mathbf{b} + \mathbf{a}, \\ (\mathbf{a} + \mathbf{b}) + \mathbf{c} &= \mathbf{a} + (\mathbf{b} + \mathbf{c}). \end{aligned} \tag{3.5}$$
6. The multiplication-by-scalar operation must be distributive with respect to vector and scalar addition, consistent with the operation of multiplying two scalars and must satisfy the multiplicative identity:
$$\begin{aligned} \alpha(\mathbf{a} + \mathbf{b}) &= \alpha\mathbf{a} + \alpha\mathbf{b}; \\ (\alpha + \beta)\mathbf{a} &= \alpha\mathbf{a} + \beta\mathbf{a}; \\ \alpha(\beta\mathbf{a}) &= (\alpha\beta)\mathbf{a}; \\ 1\mathbf{a} &= \mathbf{a}. \end{aligned} \tag{3.6}$$

For our purposes the scalars $\mathcal{F}$ will usually be either the set $\mathbb{R}$ of all real numbers (in which case we have a **real vector space**) or the set $\mathbb{C}$ of all complex numbers (giving a **complex vector space**). Note that the "type" of vector space refers to the type of *scalars*!

Examples abound: see lectures.

A subset $\mathcal{W} \subseteq \mathcal{V}$ is a **subspace** of $\mathcal{V}$ if it satisfies the first 4 conditions above. That is: it must be closed under addition of vectors and multiplication by scalars; it must contain the zero vector; the additive inverse of each element must be included. The other conditions are automatically satisfied because they depend only on the definition of the addition and multiplication operations. AL p.12

Notice that the conditions for defining a vector space involve only **linear combinations** of vectors: new vectors constructed simply by scaling and adding other vectors. There is not (yet) any notion of length or angle: these require the introduction of a scalar product (§5.1). The following important ideas are associated with linear combinations of vectors: AL p.14

- The **span** of a list of vectors $(\mathbf{v}_1, ..., \mathbf{v}_m)$ is the set of all possible linear combinations of them:

$$\operatorname{span}(\mathbf{v}_1, ..., \mathbf{v}_m) \equiv \left\{ \sum_{i=1}^{m} \alpha_i \mathbf{v}_i \,|\, \alpha_i \in \mathcal{F} \right\}. \tag{3.7}$$

  The span is a vector subspace of $\mathcal{V}$.

- A list of vectors $\mathbf{v}_1, ..., \mathbf{v}_m$ is **linearly independent** (or LI) if the only solution to AL p.15

$$\sum_{i=1}^{m} \alpha_i \mathbf{v}_i = \mathbf{0} \tag{3.8}$$

  is if all the scalars $\alpha_i = 0$. Otherwise the list is **linearly dependent** (LD).

- A **basis** for $\mathcal{V}$ is a list of vectors $\mathbf{e}_1, \mathbf{e}_2, ...$ drawn from $\mathcal{V}$ that is LI and spans $\mathcal{V}$. AL p.17

- Given a basis $\mathbf{e}_1, ..., \mathbf{e}_n$ for $\mathcal{V}$ we can expand any vector $\mathbf{v} \in \mathcal{V}$ as $\mathbf{v} = \sum_{i=1}^{n} v_i \mathbf{e}_i$, where the $v_i \in \mathcal{F}$ are the **coordinates** of $\mathbf{v}$.

In lectures we establish the following basic facts:

- given a basis these coordinates are unique;

- if a list of vectors is LD then at least one of them can be expressed as a linear combination of the others;

- if $\mathbf{v}_1, ..., \mathbf{v}_n$ are LI, but $\mathbf{v}_1, ..., \mathbf{v}_n, \mathbf{w}$ are LD then $\mathbf{w}$ is a linear combination of $\mathbf{v}_1, ...\mathbf{v}_n$.

- **(Exchange lemma)** If $\mathbf{v}_1, ..., \mathbf{v}_m$ is a basis for $\mathcal{V}$ then any subset $\{\mathbf{w}_1, ..., \mathbf{w}_n\}$ of elements of $\mathcal{V}$ with $n > m$ is LD.

A corollary of the exchange lemma is that all bases of $\mathcal{V}$ have the same number of elements. This number – the maximum number of LI vectors – is called the **dimension** of $\mathcal{V}$. AL p.20

3.2 Linear maps

Recall that a function is a mapping $f : X \to Y$ from a set X (the **domain** of f) to another set Y AL p.42
(the **codomain** of f). The **image** of f is defined as

$$\operatorname{Im} f \equiv \{f(x) | x \in X\} \subseteq Y. \tag{3.9}$$

The map f is

- **one-to-one (injective)** if each $y \in Y$ is mapped to by *at most* one element of X;
- **onto (surjective)** if each $y \in Y$ is mapped to by *at least* one element of X;
- **bijective** if each $y \in Y$ is mapped to by *precisely* one element of X (i.e., f is both injective and surjective).

A trivial but important map is the identity map,

$$\mathrm{id}_X : X \to X,$$

defined by $\mathrm{id}_X(x) = x$. Given $f : X \to Y$ and $g : Y \to Z$ we define their **composition** $g \circ f : X \to Z$ as the new map, AL p.43

$$(g \circ f)(x) = g(f(x)),$$

obtained by applying f first, then g to the result. A map $g : Y \to X$ is the **inverse** of $f : X \to Y$ if $f \circ g = \mathrm{id}_Y$ and $g \circ f = \mathrm{id}_X$. If it exists this inverse map is usually written as f^{-1}. It is a few lines' work to prove the following fundamental properties of inverse maps:

- f is invertible $\Leftrightarrow$ f is bijective. The inverse function is unique. AL p.44
- if $f : X \to Y$ and $h : Y \to Z$ are both invertible then $(h \circ f)^{-1} = f^{-1} \circ h^{-1}$;
- $(f^{-1})^{-1} = f$.

For the rest of the course we'll focus on maps $f : \mathcal{V} \to \mathcal{W}$, whose domain $\mathcal{V}$ and codomain $\mathcal{W}$ are vector spaces, possibly of different dimensions, but over the same field $\mathcal{F}$. Such an $f : \mathcal{V} \to \mathcal{W}$ is AL p.45
linear if for all vectors $\mathbf{v}_1, \mathbf{v}_2 \in \mathcal{V}$ and all scalars $\alpha \in \mathcal{F}$,

$$\begin{aligned} f(\mathbf{v}_1 + \mathbf{v}_2) &= f(\mathbf{v}_1) + f(\mathbf{v}_2), \\ f(\alpha \mathbf{v}) &= \alpha f(\mathbf{v}). \end{aligned} \tag{3.10}$$

It's easy to see from that that if $f : \mathcal{V} \to \mathcal{W}$ and $g : \mathcal{W} \to \mathcal{U}$ are both linear maps then their composition $g \circ f : \mathcal{V} \to \mathcal{U}$ is also linear. And if a linear $f : \mathcal{V} \to \mathcal{V}$ is bijective then its inverse f^{-1} is also a linear map.

3.2.1 Properties of linear maps

The **kernel** of f is the set of all elements $\mathbf{v} \in \mathcal{V}$ for which $f(\mathbf{v}) = \mathbf{0}$. The **rank** of f is the dimension AL p.45
of its image.

$$\begin{aligned} \operatorname{Ker} f &\equiv \{\mathbf{v} \in \mathcal{V} \,|\, f(\mathbf{v}) = \mathbf{0}\}, \\ \operatorname{rank} f &\equiv \dim \operatorname{Im} f. \end{aligned} \tag{3.11}$$

Directly from these definitions we have that AL p.46

- $f(\mathbf{0}) = \mathbf{0}$. Therefore $\mathbf{0} \in \operatorname{Ker} f$.
- $\operatorname{Ker} f$ is a vector subspace of $\mathcal{V}$.
- $\operatorname{Im} f$ is a vector subspace of $\mathcal{W}$.
- f surjective $\Leftrightarrow \operatorname{Im} f = W \Leftrightarrow \dim \operatorname{Im} f = \dim \mathcal{W}$.

- f injective $\Leftrightarrow \operatorname{Ker} f = \{\mathbf{0}\} \Leftrightarrow \dim \operatorname{Ker} f = 0$.

But the main result in this section is that any linear map $f : \mathcal{V} \to \mathcal{W}$ satisfies the **dimension theorem** AL p.47

$$\dim \operatorname{Ker} f + \dim \operatorname{Im} f = \dim \mathcal{V}. \tag{3.12}$$

We'll be invoking this theorem over the next few lectures. Here are some immediate consequences of it:

- if f has an inverse (bijective) then $\dim \mathcal{V} = \dim \mathcal{W}$; AL p.48
- if $\dim \mathcal{V} = \dim \mathcal{W}$ then the following statements are equivalent (they are all either true or false): f is bijective $\Leftrightarrow \dim \operatorname{Ker} f = 0 \Leftrightarrow \operatorname{rank} f = \dim \mathcal{W}$;

3.2.2 Examples of linear maps

- **Matrices** Any $n \times m$ matrix is a linear map from $\mathcal{F}^m$ to $\mathcal{F}^n$: for any $\mathbf{u}, \mathbf{v} \in \mathcal{F}^m$ we have that $A(\mathbf{u}+\mathbf{v}) = A(\mathbf{u}) + A(\mathbf{v})$ and $A(\alpha\mathbf{v}) = \alpha(A\mathbf{v})$. Composition is multiplication. AL p.49
- **Coordinate maps** Given a vector space $\mathcal{V}$ over a field $\mathcal{F}$ with basis $\mathbf{e}_1, ..., \mathbf{e}_n$, we have that any vector in $\mathcal{V}$ can be expressed as

$$\mathbf{v} = \sum_{i=1}^{n} v_i \mathbf{e}_i,$$

where the v_i are the coordinates of the vector wrt the $\mathbf{e}_i$ basis. Introduce a mapping $\varphi : \mathcal{F}^n \to \mathcal{V}$ defined by

$$\varphi \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} = \sum_{i=1}^{n} v_i \mathbf{e}_i. \tag{3.13}$$

This is called the **coordinate map** (for the basis $\mathbf{e}_1, ..., \mathbf{e}_n$). It is clearly linear. Because $\dim \mathcal{F}^n = \dim \mathcal{V}$ and $\operatorname{Im} \varphi = \mathcal{V}$ it follows as a consequence of the dimension theorem that φ is bijective and therefore has an inverse mapping $\varphi^{-1} : \mathcal{V} \to \mathcal{F}^n$. The coordinates $(v_1, ..., v_n)$ of the vector $\mathbf{v}$ are simply $\varphi^{-1}(\mathbf{v})$, the i^{th} element of which is v_i. typo removed

- **Linear differential operators** Consider the vector space $\mathcal{C}^\infty(\mathbb{R})$ of smooth functions $f : \mathbb{R} \to \mathbb{C}$. Then the differential operator

$$\begin{aligned} L &: \mathcal{C}^\infty(\mathbb{R}) \to \mathcal{C}^\infty(\mathbb{R}) \\ L\bullet &= \sum_{i=0}^{n} p_i(x) \frac{\mathrm{d}^i}{\mathrm{d}x^i} \bullet \end{aligned} \tag{3.14}$$

is a linear map.

3.3 Change of basis of the matrix representing a linear map

Let $f : \mathcal{V} \to \mathcal{W}$ be a linear map. Suppose that $\mathbf{v}_1, ..., \mathbf{v}_m$ is a basis for $\mathcal{V}$ with coordinate map $\phi : \mathcal{F}^m \to \mathcal{V}$, and $\mathbf{w}_1, ..., \mathbf{w}_n$ is a basis for $\mathcal{W}$ with coordinate map $\psi : \mathcal{F}^n \to \mathcal{W}$. The matrix A that represents $f : \mathcal{V} \to \mathcal{W}$ in these bases is simply AL p.71

$$A = \psi^{-1} \circ f \circ \phi. \tag{3.15}$$

Now consider another pair of bases for $\mathcal{V}$ and $\mathcal{W}$ with coordinate maps

$$\begin{aligned} \varphi' : \mathcal{F}^m \to \mathcal{V}, &\quad \text{basis } \mathbf{v}'_1, ..., \mathbf{v}'_m \in \mathcal{V}; \\ \psi' : \mathcal{F}^n \to \mathcal{W}, &\quad \text{basis } \mathbf{w}'_1, ..., \mathbf{w}'_n \in \mathcal{W}. \end{aligned}$$

The matrix A' representing $f : \mathcal{V} \to \mathcal{W}$ in the new primed bases is given by

$$\begin{aligned} A' &= (\psi')^{-1} \circ f \circ \varphi' \\ &= \underbrace{(\psi')^{-1} \circ \psi}_{Q} \circ \underbrace{\psi^{-1} \circ f \circ \varphi}_{A} \circ \underbrace{\varphi^{-1} \circ \varphi'}_{P^{-1}}. \end{aligned} \tag{3.16}$$

Here P^{-1} is an $m \times m$ matrix that transforms coordinate vectors wrt the $\mathbf{v}'_i$ basis to the $\mathbf{v}_i$ one. Inverting, P transforms unprimed coordinates to primed ones. Similarly, Q transforms from the $\mathbf{w}_i$ basis to the $\mathbf{w}'_i$ one.

The most important case is when $\mathcal{V} = \mathcal{W}$, $\mathbf{v}_i = \mathbf{w}_i$, $\mathbf{v}'_i = \mathbf{w}'_i$. Then A and A' are both square and are related through

$$A' = PAP^{-1}. \tag{3.17}$$

Here's how this works: P^{-1} transforms coordinate vectors from the primed to unprimed basis. Then A acts on that before we transform the resulting coordinates back to the primed basis. Think P does the priming of the *coordinates!*

Exercises

1. We have just seen that coordinate vectors $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_m)^T$ and $\boldsymbol{\alpha}' = (\alpha'_1, ..., \alpha'_m)^T$ with respect to these two bases are related via $\boldsymbol{\alpha}' = P\boldsymbol{\alpha}$, or, in index form, $\alpha'_j = \sum_k P_{jk}\alpha_k$. Show that the *basis vectors* transform in the opposite way: $\mathbf{v}_j = \sum_k (P^{-1})_{jk}\mathbf{v}'_k$.

2. Consider the reflection matrix $A = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$. What is the corresponding matrix in a new basis rotated counterclockwise by α: $\mathbf{v}'_1 = \begin{pmatrix} \cos\alpha \\ \sin\alpha \end{pmatrix}$, $\mathbf{v}'_2 = \begin{pmatrix} -\sin\alpha \\ \cos\alpha \end{pmatrix}$?

Week 4

More on matrices and maps

4.1 Systems of linear equations

AL p.73

Suppose we have n simultaneous equations in m unknowns:

$$\begin{aligned} A_{11}x_1 + \cdots A_{1m}x_m &= b_1, \\ \vdots \qquad & \quad \vdots \\ A_{n1}x_1 + \cdots A_{nm}x_m &= b_n. \end{aligned} \tag{4.1}$$

In matrix form this is $A\mathbf{x} = \mathbf{b}$, where A is an $n \times m$ matrix, $\mathbf{b}$ is an n-dimensional column vector and $\mathbf{x}$ is the m-dimensional column vector we're trying to find. If $\mathbf{b} = \mathbf{0}$ the system of equations is called **homogeneous**.

Exercises

1. Show that the full space of solutions to $A\mathbf{x} = \mathbf{b}$ is the set $\{\mathbf{x}_0 + \mathbf{x}_1 \,|\, \mathbf{x}_0 \in \ker A\}$, where $\mathbf{x}_1$ is *any* vector for which $A\mathbf{x}_1 = \mathbf{b}$.

4.1.1 Rows versus columns

AL p.54

AL p.74

In §3.2.1 (eq. 3.11) we defined the rank of a linear map as being equal to the dimension of its image. Here we distinguish between the row rank and the column rank of a matrix and show that they are in fact equal.

We'll need the following result. Let $\mathcal{W}$ be a vector subspace of $\mathcal{F}^m$. The **orthogonal complement** to $\mathcal{W}$ is

$$\mathcal{W}^\perp \equiv \{\mathbf{v} \,|\, \mathbf{v}\cdot\mathbf{w} = 0, \text{ for all } \mathbf{w} \in \mathcal{W}, \mathbf{v} \in \mathcal{F}^m\}, \tag{4.2}$$

in which $\mathbf{v}\cdot\mathbf{w} = \sum_{i=1}^m v_i w_i$.[1] The **orthogonal complement dimension theorem** states that AL p.100

$$\dim \mathcal{W} + \dim \mathcal{W}^\perp = m. \tag{4.3}$$

The proof is similar to that for the original dimension theorem (3.12).

Now let's apply this to matrices. We can view the $n \times m$ matrix A as either

1. m column vectors, $\mathbf{A}^1, ..., \mathbf{A}^m$ (each having n entries), or
2. n row vectors $\mathbf{A}_1, ...\mathbf{A}_n$ (each having m entries).

These lead to two different ways of thinking about solutions to the inhomogeneous equation $A\mathbf{x} = \mathbf{0}$:

[1] In §6.4 we'll recycle this idea, but replacing the $\mathbf{v}\cdot\mathbf{w}$ in the definition of $\mathcal{W}^\dagger$ by the scalar product $\langle \mathbf{v}, \mathbf{w}\rangle$. Equation (4.3) still holds for this modified definition of $\mathcal{W}^\perp$.

1. The solution space is simply $\ker A$, i.e., everything that doesn't map to $\operatorname{Im} A \backslash \{\mathbf{0}\}$. This unmapped space is $\operatorname{Im} A = \operatorname{span}(\mathbf{A}^1, ..., \mathbf{A}^m)$, the span of the *columns* of A. By the (original) dimension theorem (3.12) we have

$$\underbrace{\text{column rank}}_{\dim \operatorname{Im} A} + \underbrace{\text{dim space of solns}}_{\dim \operatorname{Ker} A} = m. \tag{4.4}$$

2. Consider instead the n rows of A. The vector equation $A\mathbf{x} = \mathbf{0}$ is equivalent to the n scalar equations $\mathbf{A}_i \cdot \mathbf{x} = 0$, where $\cdot$ is matrix multiplication as used in the definition (4.2) of $\mathcal{W}^\perp$ above. That is, solutions live in the space "orthogonal" to that spanned by the rows. Let $\mathcal{W}$ be the vector subspace of $\mathcal{F}^m$ spanned by the row vectors $\mathbf{A}_1, ..., \mathbf{A}_n$. Its dimension is the **row rank** of A. $\mathcal{W}^\perp$ is then the space of solutions to $\mathbf{A}_i \cdot \mathbf{v} = 0$ $(i = 1, ..., n)$. From the orthogonal complement dimension theorem (4.3),

$$\underbrace{\text{row rank}}_{\dim \mathcal{W}} + \underbrace{\text{dim space of solns}}_{\dim \mathcal{W}^\perp} = m. \tag{4.5}$$

Comparing (4.4) and (4.5) shows that the row rank is equal to the (column) rank. [Lukas p.55]

4.1.2 How to calculate the rank of a matrix

AL p.63

That's wonderful, but given a matrix how do we actually calculate its rank? To do this observe that the vector subspace spanned by a list of vectors $\mathbf{v}_1, ..., \mathbf{v}_n \in \mathcal{V}$ is unchanged if we perform any sequence of the following operations:

- Swap any pair $\mathbf{v}_i$ and $\mathbf{v}_j$;
- Multiply any $\mathbf{v}_i$ by a nonzero scalar;
- Replace $\mathbf{v}_i$ by $\mathbf{v}_i + \alpha \mathbf{v}_j$.

We can apply these operations to the rows $\mathbf{A}_1, \mathbf{A}_2, ...$ of A to reduce it to a form from which we can read off the row rank by inspection. For example, if A is a 5×5 matrix and we reduce it to

$$\begin{pmatrix} (\mathbf{A}_1)_1 & (\mathbf{A}_1)_2 & (\mathbf{A}_1)_3 & (\mathbf{A}_1)_4 & (\mathbf{A}_1)_5 \\ 0 & (\mathbf{A}_2)_2 & (\mathbf{A}_2)_3 & (\mathbf{A}_2)_4 & (\mathbf{A}_2)_5 \\ 0 & 0 & (\mathbf{A}_3)_3 & (\mathbf{A}_3)_4 & (\mathbf{A}_3)_5 \\ 0 & 0 & 0 & 0 & (\mathbf{A}_4)_5 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

then it's clear that the (row) rank of A is 4. This procedure is called **row reduction**: the idea is to reduce the matrix to so-called **row echelon form** in which the first nonzero element of row i is rightwards of the first nonzero element of the preceding row $i-1$. [2]

4.1.3 How to solve systems of linear equations: Gaussian elimination

AL p.63

Row reduction is based on the invariance of rank A under the following **elementary row operations**:

- Swap two rows $\mathbf{A}_i$, $\mathbf{A}_j$;
- Multiply any $\mathbf{A}_i$ by a nonzero value;
- Add $\alpha \mathbf{A}_j$ to $\mathbf{A}_i$.

Each of these operations has a corresponding **elementary matrix**. For example, to add $\alpha \mathbf{A}_1$ to $\mathbf{A}_2$ we can premultiply $\mathbf{A}$ by

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ \alpha & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{4.6}$$

[2] Because row rank=column rank we could just as well apply the same idea to A's columns, but it's conventional to use rows.

To solve $A\mathbf{x} = \mathbf{b}$ we can premultiply both sides by a sequence of elementary matrices E_1, E_2, ..., E_k, reducing it to

$$(E_k \cdots E_2 E_1 A)\mathbf{x} = (E_k \cdots E_2 E_1)\mathbf{b}, \tag{4.7}$$

in which the matrix $(E_k \cdots E_2 E_1 A)$ on the LHS is reduced to row echelon form. That is,

1. Use row reduction to bring A into echelon form; AL p.77
2. apply the same sequence of row operations to $\mathbf{b}$;
3. solve for $\mathbf{x}$ by **backsubstitution**.

The advantage of this procedure over other methods (e.g., using the inverse) is that we can construct $\ker A$ and deal with the case when the solution is not unique.

4.1.4 How to find the inverse of a matrix

AL p.66

We can use the same idea to construct the inverse A^{-1} of A (if it exists, or, if doesn't, to diagnose why not). The steps are as follows:

0. Stop! Do you really need to know A^{-1}? For example, it is better to use row reduction to solve $A\mathbf{x} = \mathbf{b}$ than to try to apply A^{-1} to both sides.
1. Apply row-reduction operations to eliminate all elements of A below the main diagonal.
2. Apply row-reduction operations to eliminate all elements of A above the main diagonal.
3. Apply the same sequence of operations to the identity matrix I. The result will be A^{-1}.

How does this work? In reducing A to the identity matrix we are implicitly finding a sequence of elementary matrices $E_1, ..., E_k$ for which

$$E_k \cdots E_2 E_1 A = I. \tag{4.8}$$

From this it follows that the inverse is given by

$$A^{-1} = E_k \cdots E_2 E_1 I. \tag{4.9}$$

Exercises

1. Use row reduction operations to obtain the rank and kernel of the matrix

$$\begin{pmatrix} 0 & 1 & -1 \\ 2 & 3 & -2 \\ 2 & 1 & 0 \end{pmatrix}. \tag{4.10}$$

2. Solve the system of linear equations

$$\begin{aligned} y - z &= 1, \\ 2x + 3y - 2z &= 1, \\ 2x + y \qquad &= b. \end{aligned} \tag{4.11}$$

3. Find the inverse of the matrix

$$\begin{pmatrix} 1 & 1 & 2 \\ 2 & -1 & 10 \\ 1 & -2 & 3 \end{pmatrix} \tag{4.12}$$

4. Given a matrix A explain how to find $\ker A$. Find $\ker A$ for

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} \tag{4.13}$$

Another, less efficient, way of calculating A^{-1} is by using the Laplace expansion of the determinant, equation (4.2.3) below.

4.2 Determinant

Suppose $\mathcal{V}_1, ..., \mathcal{V}_k$ are vector spaces over a common field of scalars $\mathcal{F}$. A map $f : \mathcal{V}_1 \times \cdots \times \mathcal{V}_k \to \mathcal{F}$ is **multilinear**, specifically k-linear, if it is linear in each argument separately: AL p.82

$$f(\mathbf{v}_1, ..., \alpha\mathbf{v}_i + \alpha'\mathbf{v}'_i, ..., \mathbf{v}_k) = \alpha f(\mathbf{v}_1, ..., \mathbf{v}_i, ..., \mathbf{v}_k) + \alpha' f(\mathbf{v}_1, ..., \mathbf{v}'_i, ..., \mathbf{v}_k). \tag{4.14}$$

For the special case $k = 2$ the map is called bilinear. A multilinear map is **alternating** if it returns zero whenever two of its arguments are equal:

$$f(\mathbf{v}_1, ..., \mathbf{v}_i, ..., \mathbf{v}_i, ..., \mathbf{v}_k) = 0. \tag{4.15}$$

This means that swapping any pair of arguments to a alternating multilinear map flips the sign of its output.

4.2.1 Definition of the determinant

The **determinant** is the (unique) mapping from $n \times n$ matrices to scalars that is n-linear alternating in the columns, and takes the value 1 for the identity matrix. It directly follows that:

1. If two columns of A are identical then $\det A = 0$.
2. Swapping two columns of A changes the sign of $\det A$.
3. If B is obtained from A by multiplying a single column of A by a factor c then $\det B = c \det A$.
4. If one column of A consists entirely of zeros then $\det A = 0$.
5. Adding a multiple of one column to another does not change $\det A$.

4.2.2 The Leibniz expansion of the determinant

We can express each column $\mathbf{A}^j$ of the matrix A as a linear combination $\mathbf{A}^j = \sum_i A_{ij}\mathbf{e}_i$ of the column basis vectors $\mathbf{e}_1 = (1, 0, 0, ..)^T, ..., \mathbf{e}_n = (0, .., 0, 1)^T$. For any k-linear map δ we have that, by definition,

$$\delta(\mathbf{A}_1, ..., \mathbf{A}_k) = \delta\left(\sum_{i_1=1}^{n} A_{i_1,1}\mathbf{e}_{i_1}, ..., \sum_{i_k=1}^{n} A_{i_k,k}\mathbf{e}_{i_k}\right) = \sum_{i_1=1}^{n} \cdots \sum_{i_k=1}^{n} A_{i_1,1} \cdots A_{i_k,k}\delta(\mathbf{e}_{i_1}, ..., \mathbf{e}_{i_k}),$$

showing that the map is completely determined by the n^k possible results of applying it to the basis vectors. Imposing the condition that δ be alternating means that that $\delta(\mathbf{e}_{i_1}, ..., \mathbf{e}_{i_n})$ vanishes if two or more of the i_k are equal. Therefore we need consider only those $(i_1, ..., i_n)$ that are **permutations** P of the list $(1, ..., n)$. The change of sign under pairwise exchanges implied by the alternating condition means that

$$\delta(\mathbf{e}_{P(1)}, ..., \mathbf{e}_{P(n)}) = \operatorname{sgn}(P)\delta(\mathbf{e}_1, ..., \mathbf{e}_n),$$

where $\operatorname{sgn}(P) = \pm 1$ is the sign of the permutation P ($+1$ for even, -1 for odd). Finally the condition that $\det I = 1$ sets $\delta(\mathbf{e}_1, ..., \mathbf{e}_n) = 1$, completely determining δ. The result is that AL p.84

$$\begin{aligned}\det A &= \sum_P \operatorname{sgn}(P) A_{P(1),1} A_{P(2),2} \cdots A_{P(n),n} \\ &= \sum_{i_1=1}^{n} \cdots \sum_{i_n=1}^{n} \epsilon_{i_1,...,i_n} A_{i_1,1} \cdots A_{i_n,n},\end{aligned} \tag{4.16}$$

where $\epsilon_{i_1,...,i_n}$ is the multidimensional generalisation of the Levi-Civita or alternating symbol (4.15).

There are two important results that follow hot on the heels of the Leibniz expansion: AL p.86

- $\det(A^{\mathrm{T}}) = \det A$
  - This implies that the determinant is also multilinear and alternating in the *rows* of A.

- $\det(AB) = (\det A)(\det B)$
  - From this it follows that A is invertible iff $\det A \neq 0$. Moreover, $\det(A^{-1}) = 1/\det A$.
  - The determinant of the matrix that represents a linear map is independent of basis used: in another basis the matrix representing the map becomes $A' = PAP^{-1}$ (equ. 3.17) and so $\det A' = \det A$. As every linear map is represented by some matrix, this means that we can sensibly extend our definition of determinant to include *any* linear map $f : \mathcal{V} \to \mathcal{V}$.

4.2.3 Laplace's expansion of the determinant

AL p.88

Although the Leibniz expansion (4.16) of $\det A$ is probably unfamiliar to you, you are likely to have encountered its Laplace expansion, which we now summarise. Given an $n \times n$ matrix A, let $A_{(i,j)}$ be the $(n-1) \times (n-1)$ matrix obtained by omitting the i^{th} row and j^{th} column of A. Define the **cofactor matrix** as

$$c_{ij} = (-1)^{i+j} \det(A_{(i,j)}). \tag{4.17}$$

Its transpose is known as the **adjugate** matrix or **classical adjoint** of A:

$$(\text{adj}A)_{ji} = (-1)^{i+j} \det(A_{(i,j)}). \tag{4.18}$$

Then with some work we can show from the Leibniz expansion that

$$\delta_{ij} \det A = \sum_{k=1}^{n} A_{ik} c_{jk} = \sum_{k=1}^{n} c_{ki} A_{kj}, \tag{4.19}$$

or, in matrix notation,

$$(\det A)I = A\,\text{adj}A = (\text{adj}A)A, \tag{4.20}$$

where I is the $n \times n$ identity matrix. This is known as the Laplace expansion of the determinant. An immediate corollary is an explicit expression for the inverse of a matrix: AL p.90

$$A^{-1} = \frac{1}{\det A}\text{adj}A.$$

4.2.4 Geometrical meaning of the determinant

$\det A$ is the (oriented) n-dimensional volume of the n-dimensional parallelpiped spanned by the columns (or rows) of A. See, e.g., XX,§4 of Lang's *Undergraduate analysis* for the details.

4.2.5 A standard application of determinants: Cramer's rule

Consider the set of simultaneous equation $A\mathbf{x} = \mathbf{b}$. For each $i = 1, ..., n$ introduce a new matrix AL p.91

$$B_{(i)} = (\mathbf{A}^1, ..., \mathbf{A}^{(i-1)}, \mathbf{b}, \mathbf{A}^{(i+1)}, ...\mathbf{A}^n),$$

obtained by replacing column i in A with $\mathbf{b}$. Note that because $A\mathbf{x} = \mathbf{b}$ we have $\mathbf{b} = \sum_j x_j \mathbf{A}^j$. So, using the multilinearity property of the determinant,

$$\begin{aligned}
\det(B_{(i)}) &= \det(\mathbf{A}^1, ..., \mathbf{A}^{(i-1)}, \mathbf{b}, \mathbf{A}^{(i+1)}, ...\mathbf{A}^n) \\
&= \det\left(\mathbf{A}^1, ..., \mathbf{A}^{(i-1)}, \sum x_j \mathbf{A}^j, \mathbf{A}^{(i+1)}, ...\mathbf{A}^n\right) \\
&= \sum_j x_j \det\left(\mathbf{A}^1, ..., \mathbf{A}^{(i-1)}, \mathbf{A}^j, \mathbf{A}^{(i+1)}, ...\mathbf{A}^n\right) \\
&= \sum_j x_j \delta_{ij} \det A = x_i \det A,
\end{aligned} \tag{4.21}$$

the last line following because the determinants vanish unless $i = j$. This shows that we may solve $A\mathbf{x} = \mathbf{b}$ in a cute but inefficient matter by calculating

$$x_i = \det(B_{(i)}) / \det A \tag{4.22}$$

for each $i = 1, ..., n$.

4.3 Trace

The trace of an $n \times n$ matrix A is defined to be the sum of its diagonal elements: AL p.110

$$\operatorname{tr} A \equiv \sum_{i=1}^{n} A_{ii}. \tag{4.23}$$

Exercises

1. Show that

$$\begin{aligned} \operatorname{tr}(AB) &= \operatorname{tr}(BA), \\ \operatorname{tr}(ABC) &= \operatorname{tr}(CAB). \end{aligned} \tag{4.24}$$

2. Let A and A' be matrices that represent the same linear map in two different bases. Show that $\operatorname{tr} A' = \operatorname{tr} A$.

3. Show that the trace of any matrix that represents a rotation by an angle θ in 3d space is equal to $1 + 2\cos\theta$.

Week 5

Scalar product, dual vectors and adjoint maps

5.1 Scalar product

Our definition (§3.1) of vector space relies only on the most fundamental idea of taking linear combinations of vectors. By introducing a scalar product (also known as an inner product) between pairs of vectors we can extend our algebraic manipulations to include lengths and angles.

Formally, a scalar product $\langle \bullet, \bullet \rangle$ on a vector space $\mathcal{V}$ is a mapping $\mathcal{V} \times \mathcal{V} \to \mathcal{F}$ between pairs of vectors and scalars that satisfies the following conditions: AL p.93

$$\langle \mathbf{c}, \alpha\mathbf{a} + \beta\mathbf{b} \rangle = \alpha\langle \mathbf{c}, \mathbf{a} \rangle + \beta\langle \mathbf{c}, \mathbf{b} \rangle \qquad \text{(linear in } \textit{second} \text{ argument)} \tag{5.1}$$

$$\langle \mathbf{b}, \mathbf{a} \rangle = \langle \mathbf{a}, \mathbf{b} \rangle^\star \qquad \text{(swapping conjugates)} \tag{5.2}$$

$$\begin{aligned} \langle \mathbf{a}, \mathbf{a} \rangle &= 0, \quad \text{iff } \mathbf{a} = \mathbf{0}, \\ &> 0, \quad \text{otherwise.} \end{aligned} \qquad \text{(positive definite)} \tag{5.3}$$

Comments:

- if $\mathcal{V}$ is a real vector space then the scalar product is real.
- if $\mathcal{V}$ is a complex vector space then the scalar product is complex.
- Condition (5.2) guarantees that $\langle \mathbf{a}, \mathbf{a} \rangle \in \mathbb{R}$. Together with (5.3) this means that we can define the length (or **magnitude** or norm) $|\mathbf{a}|$ of the vector $\mathbf{a}$ through

$$|\mathbf{a}|^2 = \langle \mathbf{a}, \mathbf{a} \rangle. \tag{5.4}$$

  Similarly in real vector spaces we define the **angle** θ between nonzero vectors $\mathbf{a}$ and $\mathbf{b}$ via

$$\langle \mathbf{a}, \mathbf{b} \rangle = |\mathbf{a}|\,|\mathbf{b}|\,\cos\theta, \tag{5.5}$$

  and say that they are **orthogonal** if

$$\langle \mathbf{a}, \mathbf{b} \rangle = 0. \tag{5.6}$$

- Using conditions (5.1) and (5.2) together we have that

$$\langle \alpha\mathbf{a} + \beta\mathbf{b}, \mathbf{c} \rangle = \alpha^\star\langle \mathbf{a}, \mathbf{c} \rangle + \beta^\star\langle \mathbf{b}, \mathbf{c} \rangle. \tag{5.7}$$

  That is, $\langle \bullet, \bullet \rangle$ is linear in the first argument only for *real* vector spaces!

Some examples of scalar products: AL p.94

- For $\mathbb{C}^n$, the space of n-dimensional vectors with complex coefficients the natural scalar product is

$$\langle \mathbf{a}, \mathbf{b} \rangle = \mathbf{a}^\dagger \cdot \mathbf{b} = \sum_{i=1}^{n} a_i^\star b_i. \tag{5.8}$$

- For the space $L^2(a,b)$ of square-integrable functions $f:[a,b]\to\mathbb{C}$ the natural choice is

$$\langle f,g\rangle = \int_a^b f^\star(x)g(x)\,\mathrm{d}x. \tag{5.9}$$

- There is one important exception to these rules in undergraduate physics. In relativity the **metric** defines a scalar product between pairs of tangent vectors in four-dimensional spacetime in which the positive-definiteness condition (5.3) is relaxed. Spacetime is a real vector space, which means that only the linearity condition (5.1) matters. In this course, however, we'll demand positive definiteness.

Exercises

1. Consider a generalisation of (5.8) to

$$\langle \mathbf{a},\mathbf{b}\rangle = \mathbf{a}^\dagger M\mathbf{b} = \sum_{i,j=1}^{n} a_i^\star M_{ij} b_j. \tag{5.10}$$

where M is an $n\times n$ matrix. This gives us linearity in $\mathbf{b}$ (5.1) for free. What constraints do the other two conditions, (5.2) and (5.3), impose on M?

2. [For your second pass through the notes:] In the definition (5.5) of the angle θ between two vectors how do we know that $|\cos\theta|\le 1$?

3. If the nonzero $\mathbf{v}_1,\ldots,\mathbf{v}_n$ are pairwise orthogonal (that is $\langle\mathbf{v}_i,\mathbf{v}_j\rangle=0$ for $i\ne j$) show that they must be LI. AL p.95

4. Show that if $\langle\mathbf{v},\mathbf{w}\rangle=0$ for all $\mathbf{v}\in\mathcal{V}$ then $\mathbf{w}=\mathbf{0}$. AL p.99

5. Suppose that the maps $f:\mathcal{V}\to\mathcal{V}$ and $g:\mathcal{V}\to\mathcal{V}$ satisfy

$$\langle\mathbf{v},f(\mathbf{w})\rangle = \langle\mathbf{v},g(\mathbf{w})\rangle \tag{5.11}$$

for all $\mathbf{v},\mathbf{w}\in\mathcal{V}$. Show that $f=g$. (Do not assume that f and g are linear maps.)

5.1.1 Important relations involving the scalar product

$$\begin{aligned}
|\mathbf{a}+\mathbf{b}|^2 &= |\mathbf{a}|^2+|\mathbf{b}|^2 \text{ if } \langle\mathbf{a},\mathbf{b}\rangle=0 && \text{(Pythagoras)} && (5.12)\\
|\mathbf{a}+\mathbf{b}|^2+|\mathbf{a}-\mathbf{b}|^2 &= 2\left(|\mathbf{a}|^2+|\mathbf{b}|^2\right) && \text{(Parallelogram Law)} && (5.13)\\
|\langle\mathbf{a},\mathbf{b}\rangle|^2 &\le |\mathbf{a}|^2|\mathbf{b}|^2 && \text{(Cauchy–Schwarz inequality)} && (5.14)\\
|\mathbf{a}+\mathbf{b}| &\le |\mathbf{a}|+|\mathbf{b}| && \text{(Triangle inequality).} && (5.15)
\end{aligned}$$

AL p.22

5.1.2 Orthonormal bases

An **orthonormal basis** for $\mathcal{V}$ is a set of basis vectors $\mathbf{e}_1,\ldots,\mathbf{e}_n$ that satisfy AL p.95

$$\langle\mathbf{e}_i,\mathbf{e}_j\rangle = \delta_{ij}. \tag{5.16}$$

Referred to such an orthonormal basis,

- the coordinates of $\mathbf{a}$ are given by
$$a_j = \langle\mathbf{e}_j,\mathbf{a}\rangle; \tag{5.17}$$

- in terms of these coordinates the scalar product becomes

$$\langle\mathbf{a},\mathbf{b}\rangle = \sum_{i=1}^{n} a_i^\star b_i; \tag{5.18}$$

- identifying $\mathbf{e}_1, \mathbf{e}_2, ...$ with the column vectors $(1,0,0,..)^T$, $(0,1,0,0,...)^T$ etc, the matrix M that represents a linear map $f : \mathcal{V} \to \mathcal{V}$ has **matrix elements**

$$M_{ij} = \langle \mathbf{e}_i, f(\mathbf{e}_j) \rangle. \tag{5.19}$$

Given a list $\mathbf{v}_1, ..., \mathbf{v}_m$ of vectors we can construct an orthonormal basis for their span using the following, known as the **Gram–Schmidt algorithm**: AL p.97

1. Choose one of the vectors $\mathbf{v}_1$. Let $\mathbf{e}'_1 = \mathbf{v}_1$. Our first basis vector is $\mathbf{e}_1 = \mathbf{e}'_1/|\mathbf{e}'_1|$.
2. Take the second vector $\mathbf{v}_2$ and let $\mathbf{e}'_2 = \mathbf{v}_2 - \langle \mathbf{e}_1, \mathbf{e}_2 \rangle \mathbf{e}_1$: that is, we subtract off any component that is parallel to $\mathbf{e}_1$. If the result is nonzero then our second normalized basis vector is $\mathbf{e}_2 = \mathbf{e}'_2/|\mathbf{e}'_2|$.
3. ...

k. By the k^{th} step we'll have constructed $\mathbf{e}_1, ...\mathbf{e}_j$ for some $j \leq k$. Subtract off any component of $\mathbf{v}_k$ that is parallel to any of these: $\mathbf{e}'_k = \mathbf{v}_k - \sum_{i=1}^{j-1} \langle \mathbf{e}_i, \mathbf{v}_k \rangle \mathbf{e}_i$. If $\mathbf{e}'_k \neq \mathbf{0}$ then our next basis vector is $\mathbf{e}_{j+1} = \mathbf{e}'_k/|ve'_k|$.

Exercises

1. Construct an orthnormal basis for the space spanned by the vectors $\mathbf{v}_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$, $\mathbf{v}_2 = \begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix}$, $\mathbf{v}_3 = \begin{pmatrix} 1 \\ -2 \\ -2 \end{pmatrix}$.

5.2 Adjoint map

Let $f : \mathcal{V} \to \mathcal{V}$ be a linear map. Its **adjoint** $f^\dagger : \mathcal{V} \to \mathcal{V}$ is defined as the map that satisfies AL p.100

$$\langle f^\dagger(\mathbf{u}), \mathbf{v}\rangle = \langle \mathbf{u}, f(\mathbf{v})\rangle. \tag{5.20}$$

Directly from this definition we have that

- $f^\dagger$ **is unique**. Suppose that both $f^\dagger = f_1$ and $f^\dagger = f_2$ satisfy (5.20). Then

  $$\langle f_1(\mathbf{u}), \mathbf{v}\rangle = \langle \mathbf{u}, f(\mathbf{v})\rangle = \langle f_2(\mathbf{u}), \mathbf{v}\rangle. \tag{5.21}$$

  Then using (5.2) together with the result of exercise 5 we have that $f_1 = f_2$.

- $f^\dagger$ **is a linear map**. Setting $\mathbf{u} = \alpha_1\mathbf{u}_1 + \alpha_2\mathbf{u}_2$ in (5.20) and taking the complex conjugate gives

  $$\begin{aligned}\big\langle \mathbf{v}, f^\dagger(\alpha_1\mathbf{u}_1 + \alpha_2\mathbf{u}_2)\big\rangle &= \langle f(\mathbf{v}), \alpha_1\mathbf{u}_1 + \alpha_2\mathbf{u}_2\rangle\\ &= \alpha_1\langle f(\mathbf{v}), \mathbf{u}_1\rangle + \alpha_2\langle f(\mathbf{v}), \mathbf{u}_2\rangle\\ &= \alpha_1\big\langle \mathbf{v}, f^\dagger(\mathbf{u}_1)\big\rangle + \alpha_2\big\langle \mathbf{v}, f^\dagger(\mathbf{u}_2)\big\rangle\\ \big\langle \mathbf{v}, f^\dagger(\alpha_1\mathbf{u}_1 + \alpha_2\mathbf{u}_2) - \alpha_1 f^\dagger(\mathbf{u}_1) - \alpha_2 f^\dagger(\mathbf{u}_2)\big\rangle &= 0.\end{aligned} \tag{5.22}$$

  Then by the result of exercise 4 it follows that

  $$f^\dagger(\alpha_1\mathbf{u}_1 + \alpha_2\mathbf{u}_2) = \alpha_1 f^\dagger(\mathbf{u}_1) + \alpha_2 f^\dagger(\mathbf{u}_2). \tag{5.23}$$

- $f^\dagger$ has the following properties:

  $$(f^\dagger)^\dagger = f \tag{5.24}$$
  $$(f+g)^\dagger = f^\dagger + g^\dagger \tag{5.25}$$
  $$(\alpha f)^\dagger = \alpha^\star f \tag{5.26}$$
  $$(f \circ g)^\dagger = g^\dagger \circ f^\dagger \tag{5.27}$$
  $$(f^{-1})^\dagger = (f^\dagger)^{-1} \quad \text{(assuming } f \text{ invertible)} \tag{5.28}$$

Suppose that A is the matrix that represents f in some orthonormal basis $\mathbf{e}_1, ..., \mathbf{e}_n$. Then by (5.19), (5.20) and (5.2) the matrix $A^\dagger$ that represents $f^\dagger$ in the same basis has elements

$$(A^\dagger)_{ij} = \big\langle \mathbf{e}_i, f^\dagger(\mathbf{e}_j)\big\rangle = A^\star_{ji}. \tag{5.29}$$

That is, $A^\dagger$ is the Hermitian conjugate of A.

5.3 Hermitian, unitary and normal maps

A **Hermitian** operator $f : \mathcal{V} \to \mathcal{V}$ is one that is self-adjoint: AL p.101

$$f^\dagger = f. \tag{5.30}$$

The corresponding matrices satisfy $A = A^\dagger$, or, more explicitly,

$$A_{ij} = A^\star_{ji}. \tag{5.31}$$

If $\mathcal{V}$ is a real vector space then the matrix A is **symmetric**.

A **unitary** operator $f : \mathcal{V} \to \mathcal{V}$ preserves scalar products: AL p.102

$$\langle f(\mathbf{u}), f(\mathbf{v})\rangle = \langle \mathbf{u}, \mathbf{v}\rangle. \tag{5.32}$$

[If $\mathcal{V}$ is a real vector space then **orthogonal** is more often used.] Using (5.20) the condition (5.32) can also be written as $\big\langle \mathbf{u}, f^\dagger(f(\mathbf{v}))\big\rangle = \langle \mathbf{u}, \mathbf{v}\rangle$. As this holds for any $\mathbf{u}, \mathbf{v}$ an equivalent condition for unitary is that

$$f \circ f^\dagger = f^\dagger \circ f = \mathrm{id}_\mathcal{V}\,. \tag{5.33}$$

That is, unitary operators satisfy $f^\dagger = f^{-1}$. Taking the determinant of (5.33) we have

$$|\det f|^2 = 1. \tag{5.34}$$

If f is an orthogonal map then this becomes $\det f = \pm 1$. If $\det f = +1$ then the map f is a pure rotation. If $\det f = -1$ then f is a rotation plus a reflection.

Let U be the matrix that represents a unitary map in some orthonormal basis. Then (5.33) becomes

$$\sum_{j=1}^n U_{ij} U^\star_{kj} = \sum_{j-1}^n U^\star_{ji} U_{jk} = \delta_{ik}. \tag{5.35}$$

The first sum shows that the rows of U are orthonormal, the second its columns.

Hermitian and unitary maps are examples of so-called **normal** maps, which are those for which AL p.112

$$[f, f^\dagger] \equiv f \circ f^\dagger - f^\dagger \circ f = 0. \tag{5.36}$$

Exercises

1. Let R be an orthogonal 3×3 matrix with $\det R = +1$. How is $\operatorname{tr} R$ related to the rotation angle?

2. Let U_1 and U_2 be unitary. Show that $U_2 \circ U_1$ is also unitary. Does the commutator $[U_2, U_1] \equiv U_2 \circ U_1 - U_1 \circ U_2$ vanish?

3. Suppose that H_1 and H_2 are Hermitian. Is $H_2 \circ H_1$ also Hermitian? If not, are they any conditions under which it becomes Hermitian?

5.4 Dual vector space

Let $\mathcal{V}$ be a vector space over the scalars $\mathcal{F}$. The **dual vector space** $\mathcal{V}^\star$ is the set of all linear maps $\varphi : \mathcal{V} \to \mathcal{F}$: AL p.108

$$\varphi(\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2) = \alpha_1\varphi(\mathbf{v}_1) + \alpha_2\varphi(\mathbf{v}_2). \tag{5.37}$$

It is easy to confirm that the set of all such maps satisfies the vector space axioms (§3.1).

Some examples of dual vectors spaces:

- If $\mathcal{V} = \mathbb{R}^n$, the space of real, n-dimensional column vectors, then elements of $\mathcal{V}^\star$ are real, n-dimensional *row* vectors.
- Take any vector space $\mathcal{V}$ armed with a scalar product $\langle \bullet, \bullet \rangle$. Then for any $\mathbf{v} \in \mathcal{V}$ the mapping $\varphi_{\mathbf{v}}(\bullet) = \langle \mathbf{v}, \bullet \rangle$ is a linear map from vectors $\bullet \in \mathcal{V}$ to scalars. The dual space $\mathcal{V}^\star$ consists of all such maps (i.e., for all possible choices of $\mathbf{v}$).

In these examples $\dim \mathcal{V}^\star = \dim \mathcal{V}$. This is true more generally: just as we used the magic of coordinate maps to show that any n-dimensional vector space $\mathcal{V}$ over the scalars $\mathcal{F}$ is isomorphic[1] to n-dimensional column vectors $\mathcal{F}^n$, we can show that the dual space $\mathcal{V}^\star$ is isomorphic to the space of n-dimensional row vectors. AL p.109

When playing with dual vectors it is conventional to write elements of $\mathcal{V}$ and $\mathcal{V}^\star$ like so,

$$\begin{aligned}\mathbf{v} &= \sum_{i=1}^{n} v^i \mathbf{e}_i \in \mathcal{V},\\ \varphi &= \sum_{i=1}^{n} \varphi_i \mathbf{e}_\star^i \in \mathcal{V}^\star,\end{aligned} \tag{5.38}$$

with upstairs indices for the coordinates v^i of vectors in $\mathcal{V}$, balanced by downstairs indices for the corresponding basis vectors, and the other way round for $\mathcal{V}^\star$.

[1] I don't define what an isomorphism is in these notes, but do in the lectures. So look it up if you zoned out.

Week 6

Eigenthings

An eigenvector of a linear map $f : \mathcal{V} \to \mathcal{V}$ is any nonzero $\mathbf{v} \in \mathcal{V}$ that satisfies the **eigenvalue equation**

$$f(\mathbf{v}) = \lambda \mathbf{v}. \tag{6.1}$$

The constant λ is the **eigenvalue** associated with $\mathbf{v}$. AL p.113

Some examples:

- Let R be a rotationg about some axis $\mathbf{z}$. Clearly R leaves $\mathbf{z}$ unchanged. Therefore $R\mathbf{z} = \mathbf{z}$. That is, $\mathbf{z}$ is an eigenvector of R with eigenvalue 1.
- The 2×2 matrix
$$M = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \tag{6.2}$$
projects the (x, y) plane onto the line $y = x$. The vector $\left(\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right)$ is a eigenvector with eigenvalue 1. Another eigenvector is $\left(\begin{smallmatrix} -1 \\ 1 \end{smallmatrix}\right)$ with corresponding eigenvalue 0.

We can rewrtite the eigenvalue equation (6.1) as

$$(f - \lambda \,\mathrm{id}_{\mathcal{V}})(\mathbf{v}) = \mathbf{0}. \tag{6.3}$$

It has nontrivial solutions if $\operatorname{rank}(f - \mathrm{id}\,\mathcal{V}) < \dim \mathcal{V}$, which is equivalent to AL p.114

$$\det(f - \lambda \,\mathrm{id}_{\mathcal{V}}) = 0, \tag{6.4}$$

known as the **characteristic equation** for f. We define the **eigenspace** associated with the eigenvalue λ to be

$$\mathrm{Eig}_f(\lambda) \equiv \ker(f - \lambda \,\mathrm{id}_{\mathcal{V}}). \tag{6.5}$$

6.1 How to find eigenvalues and eigenvectors

AL p.114

Choose a basis for $\mathcal{V}$ and let A be the $n \times n$ matrix that represents f in this basis. Then the characteristic equation (6.4) becomes

$$0 = \det(A - \lambda I) = \sum_Q \operatorname{sgn}(Q)(A - \lambda I)_{Q(1),1)} \cdots (A - \lambda I)_{Q(n),n}, \tag{6.6}$$

which is an n^{th}-order polynomial equation for λ. Notice that

- This will have n roots, $\lambda_1, ..., \lambda_n$, although some might be repeated.
- These roots (including repetitions) are the eigenvalues of A (and therefore of f).
- The eigenvalues are in general complex, even when $\mathcal{V}$ is a real vector space!

Having found $\lambda_1, ..., \lambda_n$ we can find each of the corresponding eigenvectors by constructing the eigenspace $\text{Eig}_\lambda(A) = \ker(A - \lambda_i I)$. If for some λ the dimension $m = \dim\ker(A - \lambda I) > 1$ then the eigenvalue λ is said to be **(m-fold) degenerate** or to have a **degeneracy** of m.

Exercises

1. Find the eigenvalues and eigenvectors of the following matrices:

$$\frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \quad \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix}, \quad \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}, \quad \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}. \tag{6.7}$$

2. Show that $\det A = \prod_{i=1}^n \lambda_i$ and $\operatorname{tr} A = \sum_{i=1}^n \lambda_i$.

3. Our procedure for finding the eigenvalues of $f : \mathcal{V} \to \mathcal{V}$ relies on choosing a basis for $\mathcal{V}$ and then representing f as a matrix A. Show that the eigenvalues this returns are independent of basis used to construct the matrix.

6.2 Eigenproperties of Hermitian and unitary maps

The following statements are easy to prove. If f is a Hermitian map ($f = f^\dagger$) then AL p.119

- the eigenvalues of f are real, and
- the eigenvectors corresponding to distinct eigenvalues of f are orthogonal.

On the other hand, if $f : \mathcal{V} \to \mathcal{V}$ is a unitary map ($f^\dagger = f^{-1}$) then AL p.123

- the eigenvalues of f are complex numbers with unit modulus, and
- the eigenvectors corresponding to distinct eigenvalues of f are orthogonal.

So, if the eigenvalues $\lambda_1, ..., \lambda_n$ of an $n \times n$ unitary or Hermitian matrix are distinct, then its n (normalized) eigenvectors are automatically an orthonormal basis.

6.3 Diagonalisation

AL p.116

A map $f : \mathcal{V} \to \mathcal{V}$ is **diagonalisable** if there is a basis in which the matrix that represents f is diagonal. An important lemma is that

$$f \text{ is diagonalisable} \quad \Leftrightarrow \quad \mathcal{V} \text{ has a basis of eigenvectors of } f. \tag{6.8}$$

Exercises

1. Now suppose that f is diagonalisable and let A be the $n \times n$ matrix that represents f in some basis. According to the preceding lemma, this A has eigenvectors $\mathbf{v}_1, ..\mathbf{v}_n$ with $A\mathbf{v}_i = \lambda_i \mathbf{v}_i$ that form an eigenbasis for $\mathbb{C}^n$. Let $P = (\mathbf{v}_1, ..., \mathbf{v}_n)$ whose columns are these eigenvectors. Show that

$$P^{-1}AP = \mathrm{diag}(\lambda_1, ...\lambda_n). \tag{6.9}$$

Diagonal matrices have such appealing properties (e.g., they commute, involve only n numbers instead of n^2) that we'd like to understand under what conditions a matrix representing a map can be diagonalised. The results of §6.2 imply that both Hermitian and unitary matrices are diagonalisable *provided the eigenvalues are distinct.* But we can do better than this. Next we'll show that:

1. a map f is diagonalisable iff it is normal (5.36);
2. if we have two normal maps f_1 and f_2 then there is a basis in which the corresponding matrices are simultaneously diagonal iff $[f_1, f_2] = 0$.

6.4 Eigenproperties of normal maps

Let $f : \mathcal{V} \to \mathcal{V}$ be a normal map ($f \circ f^\dagger = f^\dagger \circ f$). Then

- $f(\mathbf{v}) = \lambda \mathbf{v} \Rightarrow f^\dagger(\mathbf{v}) = \lambda^\star \mathbf{v}$; AL p.122
- $\mathcal{V}$ has an orthogonal basis of eigenvectors of f. AL p.123

It immediately follows that $\mathcal{V}$ has an *orthonormal* basis of eigenvectors of f.

Here is how this works in practice. If the eigenvalues of f are distinct then the eigenvectors are automatically orthogonal. If there are degenerate eigenvalues (e.g., $\lambda_1 = \lambda_2$ then $\mathbf{v}_1$ and $\mathbf{v}_2$ are LI and span $\mathrm{Eig}_f(\lambda)$, but aren't necessarily orthogonal; we can use Gram–Schmidt to make them so.

In the $\mathbf{v}_1, ..., \mathbf{v}_n$ eigenbasis the matrix D that represents f has elements

$$D_{ij} = \langle \mathbf{v}_i, f(\mathbf{v}_j) \rangle = \lambda_i \delta_{ij}. \tag{6.10}$$

Now suppose that A is the matrix that represents f in our "standard" $\mathbf{e}_1 = (1, 0, 0..)^T$, $\mathbf{e}_2 = (0, 1, 0, ...)^T$ etc basis. We can expand the eigenvectors $\mathbf{v}_i = \sum_j Q_{ji} \mathbf{e}_j$: that is,

$$Q = (\mathbf{v}_1, ..., \mathbf{v}_n) \tag{6.11}$$

is a matrix whose i^{th} column is $\mathbf{v}_i$ expressed in the standard $\mathbf{e}_1, ..., \mathbf{e}_n$ basis. Then

$$Q^{-1} A Q = \mathrm{diag}(\lambda_1, ..., \lambda_n). \tag{6.12}$$

Exercises

1. Diagonalise the matrix

$$A = \frac{1}{2} \begin{pmatrix} 3 & -1 \\ -1 & 3 \end{pmatrix}. \tag{6.13}$$

2. Diagonalize AL p.121

$$\frac{1}{4} \begin{pmatrix} 2 & 3\sqrt{2} & 3\sqrt{2} \\ 3\sqrt{2} & -1 & 3 \\ 3\sqrt{2} & 3 & -1 \end{pmatrix} \tag{6.14}$$

3. Show that if f has an orthonormal basis of eigenvectors then it is normal. [Hint: choose a basis and show that $(A^\dagger)_{ij} = A^\star_{ji}$.]

6.5 Simultaneous diagonalisation

Let A and B be two diagonalisable matrices. Then there is a basis in which both A and B are diagonalisable iff $[A, B] = 0$. AL p.125

6.6 Applications

- Exponentiating a matrix: AL p.129

$$\exp[\alpha G] = \lim_{m\to\infty}\left[I + \frac{1}{m}\alpha G\right]^m = I_n + \alpha G + \frac{1}{2}\alpha^2 G^2 + \cdots = \sum_{j=0}^{\infty}\frac{1}{j!}(\alpha G)^j. \qquad (6.15)$$

  If G is diagonalisable then $G = P\,\mathrm{diag}(\lambda_1, ..., \lambda_n)P^{-1}$ and (6.15) becomes

$$\exp[\alpha G] = P\mathrm{diag}(\mathrm{e}^{\alpha\lambda_1}, ..., \mathrm{e}^{\alpha\lambda_n})P^{-1}. \qquad (6.16)$$

- Quadratic forms. A quadratic form is an expression such as AL p.131

$$q(x, y, z) = ax^2 + by^2 + cz^2 + dxy + eyz + fzx \qquad (6.17)$$

  that is a sum of terms quadratic in the variables x, y, z. We can express any real quadratic form as

$$q(x_1, ..., x_n) = \sum_{i=1}^{n} Q_{ii}^2 x_i^2 + 2\sum_{i=1}^{n}\sum_{j=1}^{i-1} Q_{ij}x_i x_j = \mathbf{x}^T Q\mathbf{x}, \qquad (6.18)$$

  where $\mathbf{x} = (x_1, ..., x_n)^{rmT}$ and Q is a symmetric matrix, $Q_{ij} = Q_{ji}$. Symmetric implies diagonalisability and so we can express $Q = P\,\mathrm{diag}(\lambda_1, ..., \lambda_n)P^T$. In new coordinates $\mathbf{x}' = P^T\mathbf{x}$ the quadratic form becomes

$$q(\mathbf{x}) = q'(\mathbf{x}') = \mathbf{x}'^T\,\mathrm{diag}(\lambda_1, ..., \lambda_n)\mathbf{x}'.\lambda_1(x_1')^2 + \lambda_2(x_2')^2 + \cdots + \lambda_n(x_n')^2, \qquad (6.19)$$

  so that the level surfaces of $q(\mathbf{x})$ (i.e., the solutions to $q(\mathbf{x}) = c$ for some constant level c) become easy to interpret geometrically.

Exercises

1. Calculate $\exp(\alpha G)$ for $G = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$.

2. Show that

$$\det\exp A = \exp(\mathrm{tr}\,A). \qquad (6.20)$$

3. Sketch the solution to the equation $3x^2 + 2xy + 2y^2 = 1$.