MATH20222: Introduction to Geometry

by Simon Peacock and Ben Smith
Edited by Aram Dermenjian and David Stewart

January 23, 2023

Thanks to H. M. Khudaverdian for giving access to previous years material with which these notes are based off of.

Contents

Introduction

Welcome to Introduction to Geometry at the University of Manchester! This semester we'll be going through a ton of various stuff in order to get you better equipped with some tools in geometry. These notes are designed to better help you learn the material and guide you through the course.

How to use these notes: These notes are different than most maths class notes. They have holes and gaps in them; they are missing information! These are the notes that we will be using in our class lectures (literally) and so it should help you follow along. The purpose of these notes is so that you don't have to write everything down; you only need to fill in the gaps giving you more time to listen to lectures and better understand the material. If you need to add extra notes, there is *tons* of room in the margins, use them! These are YOUR notes. We've also added links to Wikipedia in case you want another reference to the terms and definitions we go over. If you feel using these notes would hinder your learning experience, then *please* use another system. The point is to learn; so use whatever works best for you.

Before starting a lecture we'd recommend quickly glancing over the notes (5-10 minutes) in order to get a brief idea of what we'll be covering in class. Remember to review these notes periodically so that you don't forget terms and examples.

Good luck, we believe in you!

-Aram & David

Week 1

1 Euclidean vector spaces

To begin doing geometry properly, we will need to recall quite a few notions from linear algebra. Note that we will be recalling many results from linear algebra, and therefore will not repeat proofs.

1.1 Vector spaces and basis vectors

1.1.1 Definition of a vector space

We denote the set of real numbers by $\mathbb{R}$. Informally, we think of a vector space as a set of vectors such that

- adding two vectors together gives us another vector in the vector space,
- multiplying a vector by a real number (or *scalar*) gives another vector in the vector space.

Definition 1.1 (Vector space) A *vector space* (over $\mathbb{R}$) $(V, +, \cdot)$ is a set of vectors, along with an addition operation $+$ and a multiplication operation $\cdot$ satisfying the following axioms for all $\mathbf{a}, \mathbf{b}, \mathbf{c} \in V$ and $\lambda, \mu \in \mathbb{R}$:

- (Additive closure) $\mathbf{a} + \mathbf{b} \in V$,
- (Additive commutativity) $\mathbf{a} + \mathbf{b} = \mathbf{b} + \mathbf{a}$,
- (Additive associativity) $(\mathbf{a} + \mathbf{b}) + \mathbf{c} = \mathbf{a} + (\mathbf{b} + \mathbf{c})$,
- (Zero) $\exists \mathbf{0} \in V$ such that $\forall \mathbf{a} \in V$, $\mathbf{a} + \mathbf{0} = \mathbf{a}$,
- (Additive inverses) $\exists \mathbf{a}^{-1} \in V$ such that $\mathbf{a} + \mathbf{a}^{-1} = \mathbf{0}$,
- (Multiplicative closure) $\lambda \cdot \mathbf{a} \in V$,
- (Multiplicative associativity) $(\lambda\mu) \cdot \mathbf{a} = \lambda \cdot (\mu \cdot \mathbf{a})$,
- (Distributivity) $\lambda \cdot (\mathbf{a} + \mathbf{b}) = \lambda \cdot \mathbf{a} + \lambda \cdot \mathbf{b}$,
- (Distributivity) $(\lambda + \mu) \cdot \mathbf{a} = \lambda \cdot \mathbf{a} + \mu \cdot \mathbf{a}$,

While we shall only be concerned with vector spaces over the real numbers in this course, in general vector spaces can have scalars in any field, e.g. $\mathbb{Q}, \mathbb{C}, \mathbb{F}_p$, etc.
Wikipedia: vector space

- (Unity) $1 \cdot \mathbf{a} = \mathbf{a}$.

Remark 1.2 We will denote vectors by writing them bold in the notes, or underlining them when writing. This is particularly necessary when we have to distinguish between 0, the real number, and $\mathbf{0}$, the zero element of the vector space.

Example 1.3 The most natural example of a vector space (over real numbers) is the space of ordered n-tuples of real numbers:

$$\mathbb{R}^n = \left\{(x_1, x_2, \ldots, x_n)^\mathsf{T} \mid x_1, \ldots, x_n \in \mathbb{R}\right\} \tag{1.1}$$

where $(-)^\mathsf{T}$ denotes the transpose.

Let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ be two vectors where for $\mathbf{x} = (x_1, \ldots, x_n)^\mathsf{T}$, $\mathbf{y} = (y_1, \ldots, y_n)^\mathsf{T}$, we define addition as

$$\mathbf{x} + \mathbf{y} = (x_1 + y_1, \ldots, x_n + y_n)^\mathsf{T}$$

and multiplication by scalars $\lambda \in \mathbb{R}$ as

$$\lambda \cdot \mathbf{x} = (\lambda x_1, \ldots, \lambda x_n)^\mathsf{T}$$

Remark 1.4 When defining this vector space we have a choice between row vectors and column vectors. Here, and throughout these notes, we will use the convention of column vectors; this is chosen so that matrices multiply on the left of vectors: $M\mathbf{x}$, mirroring the idea of functions being denoted on the left: $f(x)$.

1.1.2 Linear dependence

As vector spaces are closed under addition and scalar multiplication, we will often want to consider *linear combinations* of vectors:

$$\sum_{i=1}^{m} \lambda_i \mathbf{x}_i = \lambda_1 \mathbf{x}_1 + \lambda_2 \mathbf{x}_2 + \cdots + \lambda_m \mathbf{x}_m, \tag{1.2}$$

where $\lambda_i \in \mathbb{R}$ are scalars (real numbers) and $\mathbf{x}_i \in V$ are vectors from the vector space V.

Definitions 1.5 (Linear dependence and independence) The vectors $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m\}$ in vector space V are *linearly dependent* if there exist m scalars $\lambda_1, \ldots, \lambda_m \in \mathbb{R}$ (not all equal to zero) such that

$$\lambda_1 \mathbf{x}_1 + \lambda_2 \mathbf{x}_2 + \cdots + \lambda_m \mathbf{x}_m = \mathbf{0} \tag{1.3}$$

If the vectors are not linearly dependent, we say they are *linearly independent*.

Wikipedia: linearly independent

Note We have to demand not all of our scalars be equal to zero, otherwise (1.3) is true for *every* set of vectors.

There are a couple of equivalent definitions of linear independence and dependence that can be more useful in practice. A set of vectors $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m\}$ are linearly independent if and only if

$$(\mu_1\mathbf{x}_1 + \mu_2\mathbf{x}_2 + \cdots + \mu_m\mathbf{x}_m = \mathbf{0} \text{ if and only if } \mu_1 = \mu_2 = \cdots = \mu_m = 0)$$

The following proposition gives a useful definition of linear dependence.

Proposition 1.6 *The vectors $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m\}$ are linearly dependent if and only if at least one of these vectors can be expressed as a linear combination of other vectors*

$$\mathbf{x}_i = \sum_{j \neq i} \lambda_j \mathbf{x}_j, \quad \lambda_j \in \mathbb{R}$$

1.1.3 Basis and dimension of a vector space

As vector spaces are closed under addition of vectors, we would like to find a small set of vectors such that we can write any vector as a linear combination of vectors in this set. This leads to the notion of a *basis*.

Given a set of vectors $\mathbf{x}_1, \ldots, \mathbf{x}_m \in V$, we define their *span* to be the set of all vectors that can be written as a linear combination of $\mathbf{x}_1, \ldots, \mathbf{x}_m$, *i.e.*,

Wikipedia: span

$$\operatorname{span}(\mathbf{x}_1, \ldots, \mathbf{x}_m) = \left\{ \mathbf{y} \in V \;\middle|\; \mathbf{y} = \sum_{i=1}^{m} \lambda_i \mathbf{x}_i,\ \lambda_i \in \mathbb{R} \right\}. \tag{1.4}$$

We say the vectors *span* V if $\operatorname{span}(\mathbf{x}_1, \ldots, \mathbf{x}_m) = V$.

Definitions 1.7 (Basis and ordered basis) A set of vectors $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\} \subset V$ form a *basis* of V if they are linearly independent and span V. A tuple of vectors $(\mathbf{e}_1, \ldots, \mathbf{e}_n)$ form an *ordered basis* if $\mathbf{e}_1, \ldots, \mathbf{e}_n$ form a basis as a set.

Wikipedia: basis

Example 1.8 Let $V = \mathbb{R}^2$ be a vector space and suppose we have the following two vectors:

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \qquad \mathbf{e}_2 = \begin{bmatrix} 0 \\ 2 \end{bmatrix}.$$

Show that they form a basis.

To show that they form a basis we must show two things: that the vectors are linearly independent and that they span all of V.

For linear independence we must show that whenever

$$\lambda_1 \mathbf{e}_1 + \lambda_2 \mathbf{e}_2 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

then we have $\lambda_1 = 0$ and $\lambda_2 = 0$. Calculating the above we have:

$$\lambda_1 \mathbf{e}_1 + \lambda_2 \mathbf{e}_2 = \begin{bmatrix} \lambda_1 \\ \lambda_1 \end{bmatrix} + \begin{bmatrix} 0 \\ 2\lambda_2 \end{bmatrix} = \begin{bmatrix} \lambda_1 \\ \lambda_1 + 2\lambda_2 \end{bmatrix}$$

From the first equality we have $\lambda_1 = \underline{0}$ and from the second we have

$$0 = \lambda_1 + 2\lambda_2 = 2\lambda_2 = \lambda_2$$

Therefore our vectors are linearly independent.

To see if they span all of V then let $\left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right]$ be an arbitrary vector. In a similar manner we want to find λ_1 and λ_2 such that

$$\lambda_1 \mathbf{e}_1 + \lambda_2 \mathbf{e}_2 = \begin{bmatrix} x \\ y \end{bmatrix}$$

is true. From above we know

$$\begin{bmatrix} \lambda_1 \\ \lambda_1 + 2\lambda_2 \end{bmatrix} = \lambda_1 \mathbf{e}_1 + \lambda_2 \mathbf{e}_2 = \begin{bmatrix} x \\ y \end{bmatrix}$$

and therefore we have $\lambda_1 = \underline{x}$ and

$$y = \lambda_1 + 2\lambda_2 = x + 2\lambda_2 \Rightarrow \frac{y - x}{2} = \lambda_2$$

Therefore, the two vectors span all of V. This implies that $(\mathbf{e}_1, \mathbf{e}_2)$ is a basis of V.

The difference between basis and ordered basis is a subtle, but important one. In a tuple, the order in which we list the element matters, whereas order does not matter as a set. For example, the sets $\{\mathbf{e}_1, \mathbf{e}_2\}$ and $\{\mathbf{e}_2, \mathbf{e}_1\}$ are the same, whereas the tuples $(\mathbf{e}_1, \mathbf{e}_2)$ and $(\mathbf{e}_2, \mathbf{e}_1)$ are *not* the same.

The following theorem highlights why bases are hugely important objects for describing vector spaces.

Theorem 1.9 *A set of vectors $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ form a basis for V if and only if any vector $\mathbf{x} \in V$ can be expressed uniquely as a linear combination*

$$\mathbf{x} = \sum_{i=1}^{n} \lambda_i \mathbf{e}_i, \ \lambda_i \in \mathbb{R}$$

Remark 1.10 The most important word in this theorem is *uniquely*: there is precisely one expression for $\mathbf{x}$. This does not hold if our vectors

are not linearly independent. If we take a set of vectors $\mathbf{x}_1, \ldots, \mathbf{x}_m$ that span V, we can express any $\mathbf{x} \in V$ as some linear combination

$$\sum_{i=1}^{m} \lambda_i \mathbf{x}_i = \lambda_1 \mathbf{x}_1 + \cdots + \lambda_m \mathbf{x}_m = \mathbf{x}.$$

However, if $\{\mathbf{x}_1, \ldots, \mathbf{x}_m\}$ don't form a basis, they must be linearly dependent and so there exists some choice of scalars $\mu_i \in \mathbb{R}$ such that

$$\sum_{i=1}^{m} \mu_i \mathbf{x}_i = \mu_1 \mathbf{x}_1 + \cdots + \mu_m \mathbf{x}_m = \underline{\mathbf{0}}.$$

Now we can deduce $\sum_{i=1}^{m} \lambda_i \mathbf{x}_i$ is not a unique expression for $\mathbf{x}$, as we can write

$$\sum_{i=1}^{m} (\lambda_i + \mu_i)\mathbf{x}_i = \sum_{i=1}^{m} \lambda_i \mathbf{x}_i + \sum_{i=1}^{m} \mu_i \mathbf{x}_i = \mathbf{x} + \mathbf{0} = \mathbf{x}.$$

We say "a basis" of a vector space rather than "the basis" as a vector space has many different bases. However, every basis of a given vector space has the same size: if $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ and $\{\mathbf{f}_1, \ldots, \mathbf{f}_m\}$ are both a basis of V, then $n = m$. This invariance is what leads to the notion of *dimension*.

Definition 1.11 (Dimension) The *dimension* of a vector space V is the size of a basis for V.

Wikipedia: dimension

Example 1.12 (The canonical basis of $\mathbb{R}^n$) Recall from (1.1) the vector space

$$\mathbb{R}^n = \left\{ (x_1, x_2, \ldots, x_n)^\mathsf{T} \mid x_1, \ldots, x_n \in \mathbb{R} \right\}.$$

Consider the vectors $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n \in \mathbb{R}^n$:

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad \mathbf{e}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} \quad \cdots \quad \mathbf{e}_n = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} \tag{1.5}$$

One can check that these vectors span $\mathbb{R}^n$ and are linearly independent, therefore they form a basis for $\mathbb{R}^n$. For any vector $\mathbf{a} = (a_1, \ldots, a_n)^\mathsf{T} \in \mathbb{R}^n$, we can express it as a linear combination of

$\mathbf{e}_1, \ldots, \mathbf{e}_n$ as follows:

$$\mathbf{a} = \begin{bmatrix} a_1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ a_2 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \cdots + \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ a_n \end{bmatrix} = a_1 \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + a_2 \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \cdots + a_n \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$$
$$= a_1\mathbf{e}_1 + a_2\mathbf{e}_2 + \cdots + a_n\mathbf{e}_n = \textstyle\sum_{i=1}^{n} a_i\mathbf{e}_i$$

Wikipedia: canonical basis

This basis is called the *canonical basis* of $\mathbb{R}^n$: it is not the only choice of basis, but it is the most natural due to how simple it is to express vectors in it.

Remark 1.13 While all vector spaces we consider will be finite dimensional, some vector spaces have a basis that consists of infinitely many basis vectors. An example of an infinite-dimensional vector space is the space of polynomials in one variable,

$$\mathbb{R}[x] = \left\{ a_0 + a_1x + a_2x^2 + a_3x^3 + \cdots \mid a_i \in \mathbb{R} \right\}.$$

This is indeed a vector space and has the basis $\{1, x, x^2, x^3, \ldots, x^N, \ldots\}$. There are infinitely many elements in this basis, therefore $\mathbb{R}[x]$ is infinite-dimensional.

Exercise 1.14 Check $\mathbb{R}[x]$ is a vector space and $\{1, x, x^2, x^3, \ldots\}$ is a basis - do not be put off by the word infinite!

1.1.4 Change of basis and transition matrices

Vector spaces do not have a unique choice of basis, there are many choices of basis. Furthermore, if we need to change from one basis to another, we would like to do it in a controlled way. This is where *transition matrices* are useful.

Let $\mathcal{B} = (\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n)$ be an arbitrary *ordered* basis of an n-dimensional vector space V. Once this basis is fixed, we can represent any vector $\mathbf{x} \in V$ as a column vector, or an $n \times 1$ matrix, in the following way:

$$\mathbf{x} = x_1\mathbf{e}_1 + \cdots + x_n\mathbf{e}_n = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \; x_i \in \mathbb{R}.$$

Note that it is very important that the basis be ordered to do this, as the ordering is what determines which entry corresponds to which basis vector.

Suppose we have an ordered set of n vectors $\mathcal{C} = (\mathbf{f}_1, \ldots, \mathbf{f}_n)$ of V.

As $\mathcal{B}$ is a basis, we can write $\mathbf{f}_i$ as

$$\mathbf{f}_i = a_{1,i}\mathbf{e}_1 + \cdots + a_{n,i}\mathbf{e}_n = \begin{bmatrix} a_{1,i} \\ \vdots \\ a_{n,i} \end{bmatrix}, \ a_{k,i} \in \mathbb{R}.$$

We can concatenate these n column vectors together to get an $n \times n$ matrix ${}_{\mathcal{B}}T_{\mathcal{C}}$:

$${}_{\mathcal{B}}T_{\mathcal{C}} = \begin{matrix} \mathbf{f}_1 & \cdots & \mathbf{f}_i & \cdots & \mathbf{f}_n \end{matrix} \begin{bmatrix} a_{1,1} & \cdots & a_{1,i} & \cdots & a_{1,n} \\ a_{2,1} & \cdots & a_{2,i} & \cdots & a_{2,n} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{n,1} & \cdots & a_{n,i} & \cdots & a_{n,n} \end{bmatrix}$$

The matrix ${}_{\mathcal{B}}T_{\mathcal{C}}$ tells us how to represent the ordered set of vectors $\mathcal{C}$ in terms of the basis $\mathcal{B}$.

We would like to know when this new set of vectors $\mathcal{C}$ forms an ordered basis. The following proposition states we can find this out using only properties of ${}_{\mathcal{B}}T_{\mathcal{C}}$.

Proposition 1.15 *Let $\mathcal{B}$ be an ordered basis for the n-dimensional vector space V. An ordered set of n vectors $\mathcal{C}$ form an ordered basis of V if and only if (TFAE)*

- *the matrix ${}_{\mathcal{B}}T_{\mathcal{C}}$ has rank n*
- *$\det({}_{\mathcal{B}}T_{\mathcal{C}}) \neq 0$, or*
- *${}_{\mathcal{B}}T_{\mathcal{C}}$ is invertible.*

Sketch of proof. This proposition is a consequence of the rank-nullity theorem from linear algebra. Suppose the rank of ${}_{\mathcal{B}}T_{\mathcal{C}}$ is less than n; by the rank-nullity theorem this occurs if and only if the nullity of ${}_{\mathcal{B}}T_{\mathcal{C}}$ is greater than zero. This is equivalent to there being a nonzero vector $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_n)^{\mathsf{T}}$ in the kernel of ${}_{\mathcal{B}}T_{\mathcal{C}}$, *i.e.*,

$${}_{\mathcal{B}}T_{\mathcal{C}}\boldsymbol{\lambda} = \lambda_1\mathbf{f}_1 + \cdots + \lambda_n\mathbf{f}_n = \mathbf{0}.$$

This gives a linear dependence on $(\mathbf{f}_1, \ldots, \mathbf{f}_n)$, therefore they do not form a basis. □

Recall that a matrix M such that $\det(M) \neq 0$ is called *nondegenerate* or *nonsingular*.

Wikipedia: nondegenerate

Definition 1.16 (Transition matrix) The $n \times n$ nonsingular matrix ${}_{\mathcal{B}}T_{\mathcal{C}}$ that describes an ordered basis $\mathcal{C} = (\mathbf{f}_1, \ldots, \mathbf{f}_n)$ in terms of the ordered basis $\mathcal{B} = (\mathbf{e}_1, \ldots, \mathbf{e}_n)$ is called the *transition matrix* from $\mathcal{B}$ to $\mathcal{C}$.

The transition matrix is also referred to as a *change of basis matrix*.
Wikipedia: transition matrix

Example 1.17 Consider the matrix

$$T = \begin{bmatrix} 1 & 3 & 7 \\ 0 & \lambda & 5 \\ 0 & 0 & 3 \end{bmatrix}$$

where $\lambda \in \mathbb{R}$ is an arbitrary parameter. Let $(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$ be an ordered basis of V. What ordered vectors does T map this ordered basis to? Is it an ordered basis as well?

The matrix T maps this ordered basis to the ordered vectors

$$(\mathbf{e}_1,\ 3\mathbf{e}_1 + \lambda\mathbf{e}_2,\ 7\mathbf{e}_1 + 5\mathbf{e}_2 + 3\mathbf{e}_3).$$

By Proposition 1.15, this forms an ordered basis for V if and only if $0 \neq \det(T) = 3\lambda$. Therefore $(\mathbf{e}_1,\ 3\mathbf{e}_1 + \lambda\mathbf{e}_2,\ 7\mathbf{e}_1 + 5\mathbf{e}_2 + 3\mathbf{e}_3)$ forms an ordered basis for any $\lambda \in \mathbb{R} \setminus \{0\}$.

Remark 1.18 We stress the importance of considering $\mathcal{B}$ and $\mathcal{C}$ as ordered bases: without the ordering, we do not know which vectors of $\mathcal{B}$ get mapped to $\mathcal{C}$. For example, the ordered bases $\mathcal{B} = (\mathbf{e}_1, \mathbf{e}_2)$ and $\mathcal{C} = (\mathbf{e}_2, \mathbf{e}_1)$ of a 2-dimensional vector space V are equal as bases, but not as ordered bases. The transition matrix that reverses the order is

$${}_{\mathcal{B}}T_{\mathcal{C}} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

As ordering is key for working with transition matrices (and other matrices later on), we will only consider ordered bases for the remainder of the course. Therefore we shall mostly drop the prefix "ordered". We shall continue to use round brackets to denote a tuple or ordered set of basis vectors, rather than a set.

1.2 Euclidean vector spaces

1.2.1 Inner products

Much of the geometry we learn at school is centred around distances and angles. We can give vector spaces some additional structure so that we have some notion of distance and angle. The key to this is an *inner product*.

There is a more general definition of inner products for vector spaces over other fields. As we will only be interested in real vector spaces, we'll stick with this definition.
Wikipedia: inner product

Definition 1.19 (Inner product) An *inner product* (or scalar product) on a vector space V (over $\mathbb{R}$) is a function

$$\langle -, - \rangle : V \times V \to \mathbb{R}$$

that maps two vectors to a scalar satisfying the following conditions for all $\mathbf{x}, \mathbf{y}, \mathbf{z} \in V$ and for all $\lambda, \mu \in \mathbb{R}$:

- (symmetry) $\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle$,
- (linearity) $\langle \lambda\mathbf{x} + \mu\mathbf{y}, \mathbf{z} \rangle = \lambda\langle \mathbf{x}, \mathbf{z} \rangle + \mu\langle \mathbf{y}, \mathbf{z} \rangle$,
- (positive-definite) $\langle \mathbf{x}, \mathbf{x} \rangle \geq 0$ with equality if and only if $\mathbf{x} = \mathbf{0}$.

Definition 1.20 (Euclidean vector space) A *Euclidean vector space* is a vector space over $\mathbb{R}$ equipped with an inner product.

The connection between inner products and notions of distance and angle is emphasised in the following example.

Example 1.21 The vector space $\mathbb{R}^n$ can be viewed as a Euclidean vector space via the inner product

$$\langle \mathbf{x}, \mathbf{y} \rangle = x_1 y_1 + \cdots + x_n y_n \tag{1.6}$$

This is sometimes called the *canonical inner product*, or the *dot product*. Wikipedia: dot product

Exercise 1.22 Check that the canonical inner product is an inner product.

Example 1.23 Consider a 2-dimensional vector space V with basis $\{\mathbf{e}_1, \mathbf{e}_2\}$. We define the inner product $\langle -, - \rangle$ such that $\langle \mathbf{e}_1, \mathbf{e}_1 \rangle = 3$, $\langle \mathbf{e}_2, \mathbf{e}_2 \rangle = 5$ and $\langle \mathbf{e}_1, \mathbf{e}_2 \rangle = 0$. What is the inner product for any two vectors?

Using the symmetry and linearity axioms, we can determine the value of the inner product for any vectors. Explicitly, for $\mathbf{x} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2$ and $\mathbf{y} = y_1\mathbf{e}_1 + y_2\mathbf{e}_2$ with $x_i, y_i \in \mathbb{R}$, their inner product is given by

$$\begin{aligned}
\langle \mathbf{x}, \mathbf{y} \rangle &= \langle x_1\mathbf{e}_1 + x_2\mathbf{e}_2, y_1\mathbf{e}_1 + y_2\mathbf{e}_2 \rangle \\
&= x_1\langle \mathbf{e}_1, y_1\mathbf{e}_1 + y_2\mathbf{e}_2 \rangle + x_2\langle \mathbf{e}_2, y_1\mathbf{e}_1 + y_2\mathbf{e}_2 \rangle \\
&= x_1y_1\langle \mathbf{e}_1, \mathbf{e}_1 \rangle + x_1y_2\langle \mathbf{e}_1, \mathbf{e}_2 \rangle + x_2y_1\langle \mathbf{e}_2, \mathbf{e}_1 \rangle + x_2y_2\langle \mathbf{e}_2, \mathbf{e}_2 \rangle \\
&= 3x_1y_1 + 5x_2y_2.
\end{aligned}$$

Example 1.24 Consider the following nonexample of an inner product defined on V. We define $\langle -, - \rangle$ such that $\langle \mathbf{e}_1, \mathbf{e}_1 \rangle = \langle \mathbf{e}_2, \mathbf{e}_2 \rangle = 0$ and $\langle \mathbf{e}_1, \mathbf{e}_2 \rangle = $. Is this an inner product?

A similar calculation to the previous example shows

$$\langle \mathbf{x}, \mathbf{y} \rangle = x_1y_2 + x_2y_1.$$

However, if we consider $\mathbf{x} = \mathbf{e}_1 - \mathbf{e}_2$ then $x_1 = 1$, $x_2 = -1$, $y_1 = 1$, and $y_2 = -1$. We have that $\langle \mathbf{x}, \mathbf{x} \rangle = -2 < 0$ and therefore positive definiteness doesn't hold.

1.2.2 Geometry of Euclidean vector spaces

We'll now use the definition of inner product to reconstruct certain geometric concepts of Euclidean vector spaces. Throughout we let V be an n-dimensional Euclidean vector space.

Let $\mathbf{x} \in V$, we define the *length* (or *magnitude*) of $\mathbf{x}$ to be

$$\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}.$$

We remark that the formula for length does not depend on a choice of basis: we can change basis freely and the length of a vector stay the same. Note that when $\langle -, - \rangle$ is the dot product, our definition of length agrees with the standard definition of length from Euclidean geometry, *i.e.*,

$$\|\mathbf{x}\| = \sqrt{\mathbf{x} \cdot \mathbf{x}} = \sqrt{x_1^2 + \cdots + x_n^2}$$

Wikipedia: norm
Wikipedia: Euclidean norm

The map $\|\cdot\| : V \to \mathbb{R}$ is an example of a *norm*, a way of defining distance on a vector space. Length is called the *Euclidean norm* as it is the norm associated with Euclidean space. However, there are many more ways of defining distance on a space, and so there are many more norms one can define.

We can also use the inner product to define the *angle* between two vectors $\mathbf{x}, \mathbf{y} \in V$. Explicitly, the angle θ between $\mathbf{x}, \mathbf{y}$ is defined by

$$\cos(\theta) = \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{x}\|\|\mathbf{y}\|}. \tag{1.7}$$

In particular, we can use the inner product to quickly infer general behaviour about the angle:

- $\langle \mathbf{x}, \mathbf{y} \rangle > 0$ if and only if θ is acute,
- $\langle \mathbf{x}, \mathbf{y} \rangle < 0$ if and only if θ is obtuse,
- $\langle \mathbf{x}, \mathbf{y} \rangle = 0$ if and only if $\mathbf{x}, \mathbf{y}$ are *orthogonal*.

Wikipedia: orthogonal
In the case of Euclidean vector spaces, two vectors are orthogonal if and only if they are *perpendicular*: the angle between them is a right angle. Orthogonality is a generalisation of perpendicularity to any vector space.
Wikipedia: Perpendicular

This last case will be of particular importance in the following section.

Similarly to length, setting $\langle -, - \rangle$ to the dot product recovers the identity

$$\mathbf{x} \cdot \mathbf{y} = \|\mathbf{x}\|\|\mathbf{y}\| \cos(\theta)$$

and so our definition of angle agrees with the standard definition from Euclidean geometry.

Remark 1.25 As $\cos(\theta) = \cos(-\theta)$, the inner product only defines angles up to sign. The intuition behind this is there is no good notion of what is a positive or negative angle. You may *choose* for angles to be measured "clockwise", but there is no good reason why you shouldn't

choose “anti-clockwise”. This issue will be dealt with in Subsection 1.5 when we discuss the orientation of a vector space.

Week 2

1.3 Orthonormal bases

1.3.1 Orthonormal bases

Consider the canonical basis in $\mathbb{R}^2$. We learn to use the canonical basis at a young age, whether it is plotting graphs at school or playing Battleships. But why *this* basis, what makes it special? It satisfies two properties that we often take for granted, but are incredibly useful:

- each of the basis vectors has unit length,
- the basis vectors are orthogonal (or perpendicular) to each other.

We would like to restrict ourselves to using bases that also have this property. This leads to the notion of an orthonormal basis.

Definition 1.26 (Orthonormal basis) Let $(\mathbf{e}_1, \ldots, \mathbf{e}_n)$ be a set of vectors of an n-dimensional vector space V such that

$$\langle \mathbf{e}_i, \mathbf{e}_j \rangle = \delta_{ij} = \begin{cases} 1 & \text{if} \quad i = j \\ 0 & \text{if} \quad i \neq j \end{cases}, \tag{1.8}$$

where δ_{ij} is the Kronecker delta function. We call $(\mathbf{e}_1, \ldots, \mathbf{e}_n)$ an *orthonormal basis* of V.

Wikipedia: orthonormal basis

Remark 1.27 Note that Definition 1.26 does not require the vectors to form a basis. In fact, one can check that any set of n vectors satisfying condition (1.8) must form a basis .

Explicitly condition (1.8) implies the n vectors are linearly independent. Suppose that $\lambda_1 \mathbf{e}_1 + \lambda_2 \mathbf{e}_2 + \cdots + \lambda_n \mathbf{e}_n = 0$. We can multiply this relation by the vector $\mathbf{e}_i$, implying that $\lambda_i = 0$. Repeating this for every i, we see the vectors $(\mathbf{e}_1, \ldots, \mathbf{e}_n)$ are linearly dependent.

Example 1.28 Recall the Euclidean vector space with canonical basis (1.5) $(\mathbf{e}_1, \ldots, \mathbf{e}_n)$ and canonical inner product (1.6). It is quick to check that $(\mathbf{e}_1, \ldots, \mathbf{e}_n)$ satisfies condition (1.8) and therefore forms an orthonormal basis.

What nice properties do orthonormal bases have? Once we have fixed an orthonormal basis for our Euclidean vector space, the inner product on that space reduces to the canonical inner product. To see this, let

$(\mathbf{e}_1, \ldots, \mathbf{e}_n)$ be an orthonormal basis for V. Then for any two vectors $\mathbf{x} = \sum_{i=1}^n x_i \mathbf{e}_i$, $\mathbf{y} = \sum_{j=1}^n y_j \mathbf{e}_j$, their inner product becomes

$$\begin{aligned} \langle \mathbf{x}, \mathbf{y} \rangle &= \left\langle \sum_{i=1}^n x_i \mathbf{e}_i, \sum_{j=1}^n y_j \mathbf{e}_j \right\rangle \\ &= \sum_{i,j=1}^n x_i y_j \langle \mathbf{e}_i, \mathbf{e}_j \rangle \\ &= \sum_{i,j=1}^n x_i y_j \delta_{ij} \\ &= \sum_{i=1}^n x_i y_i. \end{aligned} \tag{1.9}$$

Furthermore, we will can always find an orthonormal basis for a Euclidean vector space.

Proposition 1.29 *Every (finite-dimensional) Euclidean vector space has an orthonormal basis.*

As a result, we can choose to work only with orthonormal bases from now on. We let $\mathbb{E}^n$ denote an n-dimensional Euclidean vector space with orthonormal basis $\mathcal{B} = (\mathbf{e}_1, \ldots, \mathbf{e}_n)$ (and therefore the canonical inner product under this basis).

1.3.2 Orthogonal matrices

Suppose that $\mathcal{B}$ is an orthonormal basis of $\mathbb{E}^n$. What is the condition on the transition matrix that an ordered set of vectors $\mathcal{C}$ also forms an orthonormal basis? The answer to this question is when the transition matrix is *orthogonal*.

Definition 1.30 (Orthogonal matrix) An $n \times n$ matrix M is *orthogonal matrix* if its product with its transpose is equal to the $n \times n$ identity matrix:

Wikipedia: orthogonal matrix

$$M^\mathsf{T} M = I_n. \tag{1.10}$$

Remark 1.31 If M is orthogonal then $\det(M) = \pm 1$. This holds as

$$\begin{aligned} 1 = \det(I_n) = \det(M^\mathsf{T} M) = \det(M^\mathsf{T}) \det(M) = \det(M)^2 \\ \Rightarrow \det(M) = \pm 1 \end{aligned}$$

(using some properties of the determinant we shall recall in Subsubsection 1.4.2). In particular, the determinant is always non-zero and so it is a transition matrix from one basis to another.

Proposition 1.32 *Let $\mathcal{B} = (\mathbf{e}_1, \ldots, \mathbf{e}_n)$ be an orthonormal basis of an n-dimensional Euclidean vector space V. The set of vectors $\mathcal{C} = (\mathbf{f}_1, \ldots, \mathbf{f}_n)$ form an orthonormal basis if and only if the transition matrix ${}_\mathcal{B}T_\mathcal{C}$ is an orthogonal matrix.*

Proof. Denote the (i,j)-th entry of ${}_\mathcal{B}T_\mathcal{C}$ as $({}_\mathcal{B}T_\mathcal{C})_{i,j} = a_{i,j}$. We can write

the inner product of $\mathbf{f}_i, \mathbf{f}_j$ as follows:

$$\begin{aligned}
\langle \mathbf{f}_i,\, \mathbf{f}_j \rangle &= \left\langle \sum_{k=1}^{n} a_{k,i}\mathbf{e}_k,\, \sum_{\ell=1}^{n} a_{\ell,j}\mathbf{e}_\ell \right\rangle && \text{(by defn. of transition matrices)} \\
&= \sum_{k,\ell=1}^{n} a_{k,i}a_{\ell,j}\langle \mathbf{e}_k,\, \mathbf{e}_\ell \rangle && \text{(by linearity of inner products)} \\
&= \sum_{k,\ell=1}^{n} a_{k,i}a_{\ell,j}\delta_{k\ell} && \text{(as } \mathcal{B} \text{ is orthonormal)} \\
&= \sum_{k=1}^{n} a_{k,i}a_{k,j} && \text{(nonzero only when } k=\ell) \\
&= \sum_{k=1}^{n} a^{\mathsf{T}}_{i,k}a_{k,j} \\
&= ({}_{\mathcal{B}}T_{\mathcal{C}})^{\mathsf{T}}({}_{\mathcal{B}}T_{\mathcal{C}}))_{ij}
\end{aligned}$$

The vectors $\mathcal{C}$ form an orthonormal basis if and only if $\langle \mathbf{f}_i, \mathbf{f}_j \rangle = \delta_{ij}$, therefore if and only if $(({}_{\mathcal{B}}T_{\mathcal{C}})^{\mathsf{T}}({}_{\mathcal{B}}T_{\mathcal{C}}))_{ij} = \delta_{ij}$. This condition is equivalent to $({}_{\mathcal{B}}T_{\mathcal{C}})^{\mathsf{T}}({}_{\mathcal{B}}T_{\mathcal{C}}) = I_n$, *i.e.*, the (i,j)-entry is 1 if $i = j$ and 0 otherwise. Therefore $\langle \mathbf{f}_i, \mathbf{f}_j \rangle = \delta_{ij}$ if and only if ${}_{\mathcal{B}}T_{\mathcal{C}}$ is orthogonal. □

Remark 1.33 The set of $n \times n$ transition matrices (with matrix multiplication as the operation) form a group called the *general linear group* $\mathrm{GL}(n, \mathbb{R})$. It can be checked that they form a group: in particular if $M, N \in \mathrm{GL}(n, \mathbb{R})$ then $MN \in \mathrm{GL}(n, \mathbb{R})$ as $\det(MN) = \det(M)\det(N)$ (see Proposition 1.46), both of which have non-zero determinant.

Wikipedia: general linear group

The set of $n \times n$ orthogonal matrices form a subgroup of $\mathrm{GL}(n, \mathbb{R})$ called the *orthogonal group* $O(n)$. Checking this forms a group is very similar to $\mathrm{GL}(n, \mathbb{R})$.

Wikipedia: orthogonal group

1.4 Linear operators

1.4.1 Matrix of a linear operator in a given basis

Definition 1.34 (Linear transformation/operator) Let V and W be real vector spaces. A map $P\colon V \to W$ is called a *linear transformation* if

$$P(\lambda\mathbf{x} + \mu\mathbf{y}) = \underline{\lambda P(\mathbf{x}) + \mu P(\mathbf{y})} \tag{1.11}$$

for all $\lambda, \mu \in \mathbb{R}$ and $\mathbf{x}, \mathbf{y} \in V$. This property is commonly referred to as *linearity*.

Wikipedia: linear transformation

In the special case that $V = W$ we also refer to P as a *linear operator*: a transformation that operates on a single vector space.

Let $\mathcal{B} = (\mathbf{e}_1, \ldots, \mathbf{e}_n)$ be a given (ordered) basis for the vector space V and consider the action of the operator P on the basis vectors. The image $P(\mathbf{e}_i)$ of a basis element is in V and hence there are coefficients

$p_{k,i} \in \mathbb{R}$ such that

$$P(\mathbf{e}_i) = \sum_{k=1}^{n} p_{k,i}\mathbf{e}_k. \tag{1.12}$$

Representing $(\mathbf{e}_1, \ldots, \mathbf{e}_n)$ as standard column vectors as in Subsubsection 1.1.4 this gives

$$P(\mathbf{e}_1) = \begin{bmatrix} p_{1,1} \\ p_{2,1} \\ \vdots \\ p_{n,1} \end{bmatrix} \qquad P(\mathbf{e}_2) = \begin{bmatrix} p_{1,2} \\ p_{2,2} \\ \vdots \\ p_{n,2} \end{bmatrix} \qquad \cdots \qquad P(\mathbf{e}_n) = \begin{bmatrix} p_{1,n} \\ p_{2,n} \\ \vdots \\ p_{n,n} \end{bmatrix}$$

The column vector representing $\mathbf{e}_i$ has a 1 in the i^{th} row and 0s elsewhere. The result of a matrix multiplication by this vector is simply the i^{th} column of the matrix. Thus, by concatenating these columns, we have a matrix for which multiplication is equivalent to applying the linear operator.

Definition 1.35 (Matrix of a linear operator) Let $\mathcal{B} = (\mathbf{e}_1, \ldots, \mathbf{e}_n)$ be a basis for a vector space V, and let P be a linear operator on V. The matrix

$$[P]_{\mathcal{B}} = \begin{bmatrix} p_{1,1} & p_{1,2} & \cdots & p_{1,n} \\ p_{2,1} & p_{2,2} & \cdots & p_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n,1} & p_{n,2} & \cdots & p_{n,n} \end{bmatrix}$$

determined by equation (1.12) is the *matrix of the linear transformation P in the basis $\mathcal{B}$*.

Wikipedia: matrix of the linear transformation P in the basis $\mathcal{B}$

Notice that if the linear operator is invertible then $\mathcal{C} = (P(\mathbf{e}_1), \ldots, P(\mathbf{e}_n))$ is a basis for V and the matrix of the linear operator coincides with the transition matrix from basis $\mathcal{B}$ to a basis $\mathcal{C}$.

Remark 1.36 A linear operator does not depend on a choice of basis, however the matrix of a linear operator does require a choice of basis since we are applying a linear operator to the choice of basis.

Example 1.37 Suppose that we have a vector space V with basis $\mathcal{B} = (\mathbf{e}_1, \mathbf{e}_2)$. Let P be the following linear operator:

$$P(\mathbf{e}_1) = \mathbf{e}_1$$
$$P(\mathbf{e}_2) = \mathbf{e}_1 + 2\mathbf{e}_2.$$

What is the matrix of P? Does $P(\mathcal{B})$ define a basis?

Then the matrix of P is the matrix:

$$[P]_{\mathcal{B}} = \begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix}$$

It can be checked that this matrix is invertible and this implies that $\mathcal{C} = (\mathbf{e}_1, \mathbf{e}_1 + 2\mathbf{e}_2)$ is a basis and

$$[P]_{\mathcal{B}} = {}_{\mathcal{B}}T_{\mathcal{C}}.$$

We wish to consider linear operators, and vectors in general, in different bases. As such we shall introduce the following notation:

Notation 1.38 Let $\mathcal{B} = (\mathbf{e}_1, \ldots, \mathbf{e}_n)$ be a basis for a vector space V and let $\mathbf{v}$ be a vector in V. We shall denote the column vector of $\mathbf{v}$ in the basis $\mathcal{B}$ as follows:

$$\text{if} \quad \mathbf{v} = \sum_{i=1}^{n} v_i \mathbf{e}_i \qquad \text{then} \quad [\mathbf{v}]_{\mathcal{B}} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$$

Moreover this notation is linear (property (1.11)): that is, if $\mathbf{w} = \sum_{i=1}^{n} w_i \mathbf{e}_i$ is another vector in V and $\lambda, \mu \in \mathbb{R}$ are scalars, then

$$\begin{aligned} \lambda[\mathbf{v}]_{\mathcal{B}} + \mu[\mathbf{w}]_{\mathcal{B}} &= \lambda \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} + \mu \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix} \\ &= \begin{bmatrix} \lambda v_1 + \mu w_1 \\ \vdots \\ \lambda v_n + \mu w_n \end{bmatrix} \\ &= [\lambda \mathbf{v} + \mu \mathbf{w}]_{\mathcal{B}}. \end{aligned}$$

Proposition 1.39 *Let $P\colon V \to V$ be a linear operator acting on the vector space V and let $\mathcal{B}$ be a basis for V. Then*

$$[P(\mathbf{v})]_{\mathcal{B}} = [P]_{\mathcal{B}}\,[\mathbf{v}]_{\mathcal{B}} \tag{1.13}$$

for all vectors $\mathbf{v} \in V$.

The proof of this proposition is left as an exercise.

We shall consider the effect of multiplication by a transition matrix on column vectors in the given bases.

Lemma 1.40 *Let $\mathcal{B} = (\mathbf{e}_1, \ldots, \mathbf{e}_n)$ and $\mathcal{C} = (\mathbf{f}_1, \ldots, \mathbf{f}_n)$ be two bases for a vector space V. Let $T = {}_{\mathcal{B}}T_{\mathcal{C}}$ be the transition matrix from basis $\mathcal{B}$ to basis $\mathcal{C}$ and let $\mathbf{v} \in V$. Then*

$$[\mathbf{v}]_{\mathcal{B}} = T[\mathbf{v}]_{\mathcal{C}} \tag{1.14}$$

Proof. Since $T = {}_{\mathcal{B}}T_{\mathcal{C}}$ then (by definition) the i^{th} column of T is $[\mathbf{f}_i]_{\mathcal{B}}$. Thus

$$[\mathbf{f}_i]_{\mathcal{B}} = T \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} i^{\text{th}} \text{ row} = T[\mathbf{f}_i]_{\mathcal{C}}$$

We have proven the result whenever $\mathbf{v} \in \mathcal{C}$, but we wish to prove the equality for a general $\mathbf{v} = \sum v_i \mathbf{f}_i \in V$. This follows by linearity (property (1.11)):

$$\begin{aligned} T[\mathbf{v}]_{\mathcal{C}} &= T\left(v_1[\mathbf{f}_1]_{\mathcal{C}} + \cdots + v_n[\mathbf{f}_n]_{\mathcal{C}}\right) \\ &= v_1[\mathbf{f}_1]_{\mathcal{B}} + \cdots + v_n[\mathbf{f}_n]_{\mathcal{B}} \\ &= [\mathbf{v}]_{\mathcal{B}} \end{aligned}$$

□

Lemma 1.41 *Let $\mathcal{B}, \mathcal{C}, \mathcal{D}$ be three bases of a vector space V. Then*

$${}_{\mathcal{B}}T_{\mathcal{D}} = {}_{\mathcal{B}}T_{\mathcal{C}}\ {}_{\mathcal{C}}T_{\mathcal{D}}.$$

Proof. Lemma 1.40 states that

$$\begin{aligned} [\mathbf{v}]_{\mathcal{B}} &= {}_{\mathcal{B}}T_{\mathcal{C}}\ [\mathbf{v}]_{\mathcal{C}} \\ &= {}_{\mathcal{B}}T_{\mathcal{C}}\ {}_{\mathcal{C}}T_{\mathcal{D}}\ [\mathbf{v}]_{\mathcal{D}} \end{aligned}$$

but this property applied to vectors in the basis $\mathcal{D}$ defines ${}_{\mathcal{B}}T_{\mathcal{D}}$. □

Lemma 1.42 *Let $\mathcal{B}, \mathcal{C}$ be bases of a vector space V. Then ${}_{\mathcal{C}}T_{\mathcal{B}} = ({}_{\mathcal{B}}T_{\mathcal{C}})^{-1}$.*

The proof is left as an exercise, see homework 2.

The lemmas above gives us a way to translate between column vectors in one basis to column vectors in another basis. Using these ideas we can translate from the matrix of a linear operator in one basis to the matrix of the same linear operator in another basis.

Proposition 1.43 *Let P be a linear operator acting on a vector space V. Let $\mathcal{B} = (\mathbf{e}_1, \ldots, \mathbf{e}_n)$ and $\mathcal{C} = (\mathbf{f}_1, \ldots, \mathbf{f}_n)$ be two bases for V. Let T be the transition matrix from basis $\mathcal{B}$ to basis $\mathcal{C}$.*

Then $[P]_{\mathcal{C}} = T^{-1}[P]_{\mathcal{B}}T$.

Proof. The i^{th} column of $[P]_{\mathcal{C}}$ is defined to be $[P(\mathbf{f}_i)]_{\mathcal{C}}$. We have that

$$\begin{aligned}[P(\mathbf{f}_i)]_{\mathcal{C}} &= T^{-1}[P(\mathbf{f}_i)]_{\mathcal{B}} && \text{(Lemma 1.40)}\\ &= T^{-1}[P]_{\mathcal{B}}[\mathbf{f}_i]_{\mathcal{B}}\\ &= T^{-1}[P]_{\mathcal{B}}T[\mathbf{f}_i]_{\mathcal{C}}, \quad \text{the } i^{\text{th}} \text{ column of } T^{-1}[P]_{\mathcal{B}}T. \end{aligned}$$ □

Remark 1.44 This proposition demonstrates that the matrix for a linear operator is not determined by the linear operator alone and a choice of basis must be made.

1.4.2 Determinant and Trace of linear operator

We recall the determinant of a matrix for completeness.

Definition 1.45 (Determinant of a matrix) Let M be an $n \times n$ matrix. The *determinant* of a matrix is defined as

Wikipedia: determinant

$$\det M = \sum_{\sigma \in S_n} \left(\operatorname{sgn}(\sigma) \prod_{i=1}^{n} (M)_{i,\sigma(i)} \right)$$

where S_n is the group of permutations on the set $\{1, \ldots, n\}$.

Wikipedia: Permutation group

However, we shall mostly be concerned with 2×2 and 3×3 matrices and so we also recall the specialised version of this definition to these cases:

$$\det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = ad - bc \tag{1.15}$$

$$\det \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} = a \det \begin{bmatrix} e & f \\ h & i \end{bmatrix} - b \det \begin{bmatrix} d & f \\ g & i \end{bmatrix} + c \det \begin{bmatrix} d & e \\ g & h \end{bmatrix} \tag{1.16}$$

We also recall (without proof) the following facts about the determinant:

Proposition 1.46 *For all $n \times n$ matrices M and N, the following properties hold:*

(1) M is invertible if and only if $\det(M) \neq 0$;

(2) $\det(MN) = \det(M)\det(N)$;

(3) $\det(M^{\mathsf{T}}) = \det(M)$;

(4) $\det(\lambda M) = \lambda^n \det(M)$ for all scalars $\lambda \in \mathbb{R}$.

Notice that because the determinant of the identity is 1, property (2) of the lemma above immediately implies that if M is invertible then $\det(M^{-1}) = (\det M)^{-1}$.

Next we introduce the trace of a matrix.

Definition 1.47 (Trace of a matrix) Let M be an $n \times n$ matrix. We define the *trace* of M to be the sum of the diagonal elements

Wikipedia: trace

$$\operatorname{Tr} M = \sum_{i=1}^{n} (M)_{i,i}$$

We shall state without proof several useful properties of the trace map.

Proposition 1.48 *The trace map satisfies the following properties:*

(1) $\operatorname{Tr}(M+N) = \operatorname{Tr}(M) + \operatorname{Tr}(N)$ *for all* $n \times n$*-matrices* M *and* N*;*

(2) $\operatorname{Tr}(\lambda M) = \lambda \operatorname{Tr}(M)$ *for all matrices* M *and scalars* $\lambda \in \mathbb{R}$*;*

(3) $\operatorname{Tr}(MN) = \operatorname{Tr}(NM)$ *for any* $m \times n$*-matrix* M *and any* $n \times m$*-matrix* N*.*

Remark 1.49

- Notice that together, properties (1) and (2) of the above proposition mean that trace satisfies linearity.
- As an aside we also remark that any map that takes matrices to scalars satisfying the properties in Proposition 1.48 must be a scalar multiple of the trace map.

Proposition 1.50 *Let* P *be a linear operator on the vector space* V*. Let* $\mathcal{B}$ *and* $\mathcal{C}$ *be two bases for* V*. Then*

$$\det([P]_{\mathcal{B}}) = \det([P]_{\mathcal{C}}) \text{ and } \operatorname{Tr}([P]_{\mathcal{B}}) = \operatorname{Tr}([P]_{\mathcal{C}}).$$

Proof. Let T be the transition matrix from basis $\mathcal{B}$ to basis $\mathcal{C}$.

$$\begin{aligned}\det([P]_{\mathcal{C}}) &= \det(T^{-1}[P]_{\mathcal{B}}T)\\ &= \det(T^{-1})\det([P]_{\mathcal{B}})\det T\\ &= (\det T)^{-1}\det([P]_{\mathcal{B}})\det T = \det([P]_{\mathcal{B}})\end{aligned}$$

$$\begin{aligned}\operatorname{Tr}([P]_{\mathcal{C}}) &= \operatorname{Tr}(T^{-1}[P]_{\mathcal{B}}T)\\ &= \operatorname{Tr}([P]_{\mathcal{B}}TT^{-1}) = \operatorname{Tr}[P]_{\mathcal{B}}\end{aligned}$$

□

Definition 1.51 (Determinant and Trace of linear operators) Let P be a linear operator on the vector space V. Let $\mathcal{B}$ be any basis for V.

We define the *determinant of the linear operator* P to be the determinant of the matrix $[P]_{\mathcal{B}}$

$$\det P = \det\left([P]_{\mathcal{B}}\right).$$

Similarly we define the *trace of the linear operator P* as the trace of the matrix

$$\operatorname{Tr} P = \operatorname{Tr}([P]_{\mathcal{B}}).$$

Remark 1.52 As defined the determinant and trace of a linear operator requires the choice of a basis, however in light of Proposition 1.50 we see that this choice does not actually matter.

1.4.3 Eigenvalues and eigenvectors of linear operators

We can determine some behaviour of a linear operator by considering the vectors whose direction do not change under the action of the operator, only their magnitude.

The word "eigen" comes from German and it means "self" or "own". So an eigenvector or eigenvalue can be seen as a linear operators own proper vector/value.
Wikipedia: eigenvector
Wikipedia: eigenvalue

Definition 1.53 (Eigenvalues and eigenvectors) Let P be a linear operator on the vector space V. An *eigenvector* of P is a nonzero vector $\mathbf{x} \in V$ such that

$$P(\mathbf{x}) = \lambda\mathbf{x}$$

where $\lambda \in \mathbb{R}$ is the *eigenvalue* associated to $\mathbf{x}$. That is, P scales $\mathbf{x}$ by λ.

Remark 1.54 Of particular note is the case when $\lambda = 1$, as then the associated eigenvectors of λ are completely fixed by the action of P. These vectors will be particularly useful in Subsubsection 1.6.2 when computing rotations in $\mathbb{E}^3$.

The following lemma shows that every eigenvector gives rise to a one-dimensional line of eigenvectors.

Lemma 1.55 *Let P be a linear operator on the vector space V. If $\mathbf{x}$ is an eigenvector of P with eigenvalue λ, then all vectors $\mathbf{y} \in \operatorname{span}(\mathbf{x})$ are also eigenvectors with eigenvalue λ.*

Proof. Pick some $\mathbf{y} = \mu\mathbf{x} \in \operatorname{span}(\mathbf{x})$ where $\mu \in \mathbb{R}$. Then by linearity of P

$$P(\mathbf{y}) = P(\mu\mathbf{x}) = \mu P(\mathbf{x}) = \mu\lambda\mathbf{x} = \lambda\mathbf{y}. \qquad \square$$

There is a rich theory behind eigenvalues and eigenvectors, however for our purposes this is all we will need.

Example 1.56 Let $\mathcal{B} = \{\mathbf{e}_x, \mathbf{e}_y, \mathbf{e}_z\}$ be an orthonormal basis for $\mathbb{E}^3$ and consider the linear operator P on $\mathbb{E}^3$ such that

$$P(\mathbf{e}_x) = \mathbf{e}_y, \qquad P(\mathbf{e}_y) = \mathbf{e}_x, \qquad P(\mathbf{e}_z) = -\mathbf{e}_z. \tag{1.17}$$

This operator swaps first two basis vectors and inverts the third one.

The matrix of P in the basis $\mathcal{B}$ is

$$[P]_{\mathcal{B}} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix}$$

We are told that P has an eigenvalue of $\lambda = 1$, and wish to find its associated eigenvectors. An eigenvector $\mathbf{x} = x\mathbf{e}_x + y\mathbf{e}_y + z\mathbf{e}_z$ with $\lambda = 1$ satisfies

$$\begin{aligned} P(\mathbf{x}) = \lambda\mathbf{x} = \mathbf{x} \quad &\Rightarrow \quad P(\mathbf{x}) - \mathbf{x} = 0 \\ &\Rightarrow \quad ([P]_{\mathcal{B}} - I_3)[\mathbf{x}]_{\mathcal{B}} = \mathbf{0} \\ &\Rightarrow \quad \begin{bmatrix} -1 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & -2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}. \end{aligned}$$

Therefore, we have the following three equations:

$$\begin{aligned} -x + y &= 0 \\ x - y &= 0 \\ -2z &= 0 \end{aligned}$$

Solving these three simultaneous equations, we find $x = y$, $z = 0$ and so the eigenvectors with eigenvalue 1 are of the form $\mathbf{x} = \alpha(\mathbf{e}_x + \mathbf{e}_y)$ where $\alpha \in \mathbb{R}$.

1.4.4 Orthogonal linear operators

Definition 1.57 (Orthogonal linear operator) A linear operator P acting on a Euclidean vector space V is called an *orthogonal linear operator* if P preserves inner products. That is,

Wikipedia: orthogonal linear operator

$$\langle P\mathbf{x},\, P\mathbf{y}\rangle = \langle \mathbf{x},\, \mathbf{y}\rangle \text{ for all } \mathbf{x}, \mathbf{y} \in V. \tag{1.18}$$

The following proposition relates orthogonal operators to orthogonal matrices.

Proposition 1.58 *Let $\mathcal{B} = (\mathbf{e}_1, \ldots, \mathbf{e}_n)$ be an orthonormal basis for V and P a linear operator on V. Then the following are equivalent:*

(1) P is an orthogonal operator,

(2) $P(\mathcal{B})$ is an orthonormal basis,

(3) $[P]_{\mathcal{B}}$ is an orthogonal matrix.

Proof. The equivalence between (2) and (3) is exactly Proposition 1.32 where $[P]_{\mathcal{B}}$ is the transition matrix from $\mathcal{B}$ to $P(\mathcal{B})$. We show that (1) and (2) are equivalent.

Suppose P is an orthogonal operator, then equation (1.18) gives

$$\langle P(\mathbf{e}_i),\, P(\mathbf{e}_j)\rangle = \langle \mathbf{e}_i,\, \mathbf{e}_j\rangle = \delta_{i,j}. \tag{1.19}$$

Hence, $P(\mathcal{B})$ is an orthonormal basis for V.

Conversely, let $P(\mathcal{B})$ be an orthonormal basis. Let $\mathbf{x} = \sum_{i=1}^n x_i\mathbf{e}_i$ and $\mathbf{y} = \sum_{j=1}^n y_j\mathbf{e}_j$ be two arbitrary vectors of V, then by linearity of inner products and of P we see:

$$\begin{aligned}
&\langle P(\mathbf{x}),\, P(\mathbf{y})\rangle \\
&\quad= \left\langle P\left(\sum_{i=1}^n x_i\mathbf{e}_i\right),\, P\left(\sum_{j=1}^n y_j\mathbf{e}_j\right)\right\rangle \\
&\quad= \left\langle \sum_{i=1}^n x_iP(\mathbf{e}_i),\, \sum_{j=1}^n y_jP(\mathbf{e}_j)\right\rangle && \text{(linearity of } P\text{)} \\
&\quad= \sum_{i,j} x_iy_j\langle P(\mathbf{e}_i),\, P(\mathbf{e}_j)\rangle && \text{(linearity of } \langle -,\, -\rangle\text{)} \\
&\quad= \sum_{i=1}^n x_iy_i && (P(\mathcal{B}) \text{ orthornomal}).
\end{aligned}$$

But recall from equation (1.9) that when $\mathbf{x}, \mathbf{y}$ are written in an orthonormal basis, their inner product is $\langle \mathbf{x},\, \mathbf{y}\rangle = \sum_{i=1}^n x_iy_i$. This agrees with $\langle P(\mathbf{x}),\, P(\mathbf{y})\rangle$, and so P is an orthogonal operator. □

Note as a corollary of this proposition that the determinant of an orthogonal operator is ± 1. In particular, orthogonal linear operators are invertible.

Week 3

1.5 Orientation of vector spaces

You may have heard the term *orientation* before. In particular, you may have heard phrases such as:

> The basis $(\mathbf{a}, \mathbf{b}, \mathbf{c})$ has the same orientation as the basis $(\mathbf{a}', \mathbf{b}', \mathbf{c}')$ if they both obey right hand rule or if they both obey left hand rule.

or

> A mirroring image has the opposite orientation to its source.

In this section we try to give exact meaning to these ideas.

Definition 1.59 (Same/opposite orientation) Let $\mathcal{B}$ and $\mathcal{C}$ be two bases for a vector space V and let T be the transition matrix from basis $\mathcal{B}$ to basis $\mathcal{C}$.

Wikipedia: Orientation

We say that $\mathcal{C}$ has *the same orientation* as $\mathcal{B}$ if $\det T > 0$.

We say that $\mathcal{C}$ has *an opposite orientation* to $\mathcal{B}$ if $\det T < 0$.

Recall that a transition matrix between bases is nondegenerate, hence its determinant cannot be equal to zero.

Example 1.60 The simplest example is that of a line, $\mathbb{E}$, a 1-dimensional vector space. Any non-zero element of $\mathbb{E}$ spans the space, so let us consider the following three bases, each with a single element:

$$\mathcal{B} = (2) \quad \mathcal{C} = (-8) \quad \mathcal{D} = (10)$$

Which basis have the same orientations and which have opposite orientation?

The transition matrix from $\mathcal{B}$ to $\mathcal{C}$ is the 1×1 matrix $[-4]$ and the transition matrix from $\mathcal{B}$ to $\mathcal{D}$ is $[5]$. Thus, $\mathcal{D}$ has the same orientation as $\mathcal{B}$, whilst $\mathcal{C}$ has opposite orientation.

Example 1.61 Let us now consider a 2-dimensional example. Consider the following two bases bases for $\mathbb{E}^2$

$$\mathcal{B} = \left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right) \qquad \mathcal{C} = \left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right)$$

Do they have the same or opposite orientation? The transition matrix from $\mathcal{B}$ to $\mathcal{C}$ is $T = \left[\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix} \right]$ and $\det T = -2$. Hence $\mathcal{C}$ has an opposite orientation to $\mathcal{B}$.

We shall denote "basis $\mathcal{C}$ has the same orientation as basis $\mathcal{B}$" by the notation $\mathcal{B} \sim \mathcal{C}$.

Wikipedia: Equivalence relation

Proposition 1.62 *The relation $\sim$ is an equivalence relation on the set of all bases for a given vector space. That is, $\sim$ is reflexive, symmetric and transitive.*

Proof.

Reflexivity

The transition matrix of $\mathcal{B}$ to itself is the identity matrix, which has positive determinant. Thus $\mathcal{B} \sim \mathcal{B}$.

Symmetricity

Let $\mathcal{B} \sim \mathcal{C}$ and T be the transition matrix from $\mathcal{B}$ to $\mathcal{C}$. The transition matrix from $\mathcal{C}$ to $\mathcal{B}$ is then T^{-1}. Now since $\det(T^{-1}) = (\det T)^{-1}$ and $\det T > 0$ we have that $\det(T^{-1}) > 0$ and so $\mathcal{C} \sim \mathcal{B}$.

Transitivity

Let $\mathcal{B} \sim \mathcal{C}$ and $\mathcal{C} \sim \mathcal{D}$ and let

$$\begin{aligned} {}_{\mathcal{B}}T_{\mathcal{C}} &\quad \text{be the transition matrix from } \mathcal{B} \text{ to } \mathcal{C} \\ {}_{\mathcal{C}}T_{\mathcal{D}} &\quad \text{be the transition matrix from } \mathcal{C} \text{ to } \mathcal{D} \\ {}_{\mathcal{B}}T_{\mathcal{D}} &\quad \text{be the transition matrix from } \mathcal{B} \text{ to } \mathcal{D} \end{aligned}$$

We know that $\det {}_{\mathcal{B}}T_{\mathcal{C}} > 0$ and $\det {}_{\mathcal{C}}T_{\mathcal{D}} > 0$ and we need to establish that $\det {}_{\mathcal{B}}T_{\mathcal{D}} > 0$. By Lemma 1.41, ${}_{\mathcal{B}}T_{\mathcal{D}} = {}_{\mathcal{B}}T_{\mathcal{C}}\,{}_{\mathcal{C}}T_{\mathcal{D}}$ therefore $\det({}_{\mathcal{B}}T_{\mathcal{D}}) = \det({}_{\mathcal{C}}T_{\mathcal{D}}) \det({}_{\mathcal{B}}T_{\mathcal{C}})$, a product of two positive values and so $\mathcal{B} \sim \mathcal{D}$. □

Since orientation is an equivalence relation this means that the set of all bases decomposes into a disjoint union of equivalence classes. Two bases of a vector space have the same orientation if and only if there are in the same equivalence class.

Example 1.60 shows that there are at least two orientation classes in a 1-dimensional vector space. In general, for a space with dimension at least 2, one can show that the determinant of the transition matrix from the basis $(\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n)$ to the basis $(\mathbf{e}_2, \mathbf{e}_1, \ldots, \mathbf{e}_n)$, with the first two vectors swapped, has determinant -1. This means that for all vec-

tor spaces there are always at least two equivalence classes of bases: a given orientation and its opposite orientation.

We want to show that these are the only two possibilities. That is, there are exactly two orientation classes.

Proposition 1.63 *Let $\mathcal{B}$ be a basis for a vector space V and let $\mathcal{B}'$ be a basis with an opposite orientation to $\mathcal{B}$.*

If $\mathcal{C}$ is a basis for V then either $\mathcal{C} \sim \mathcal{B}$ or $\mathcal{C} \sim \mathcal{B}'$.

Proof. Let T be the transition matrix from basis $\mathcal{B}$ to $\mathcal{C}$ and T' be the transition matrix from basis $\mathcal{B}'$ to $\mathcal{C}$. Let S be the transition matrix from $\mathcal{B}$ to $\mathcal{B}'$.

If $\det T' > 0$ then we are done, so let us assume that $\det T' < 0$. We have seen that $T = ST'$ (Lemma 1.41) and therefore $\det T = \det T' \det S$. As $\mathcal{B}$ and $\mathcal{B}'$ have opposite orientation then $\det S$ is negative and we have assumed that $\det T'$ is also negative. This shows that $\det T > 0$ and therefore that $\mathcal{B} \sim \mathcal{C}$. □

Definition 1.64 (Orientation) An *orientation* of a vector space is an equivalence class of bases under the equivalence relation $\sim$.

Wikipedia: orientation

Note that any choice of basis $\mathcal{B}$ implicitly chooses an orientation: the equivalence class of $\mathcal{B}$ under the relation $\sim$. The proposition above tells us that there are two orientations, and that every basis has either the same orientation as a fixed given basis or the opposite orientation to it.

We may pick an orientation and call it the *left orientation* and its opposite the *right orientation*, though such a choice is arbitrary. A basis with a left orientation is sometimes referred to as a *left basis* and a basis with a right orientation is sometimes referred to as a *right basis*.

Definition 1.65 (Oriented vector space) An *oriented vector space* is a vector space together with a choice of orientation.

Example 1.66 Let $(\mathbf{e}, \mathbf{f})$ be a basis of a 2-dimensional vector space. We shall say that $(\mathbf{e}, \mathbf{f})$ has a *left orientation*.

We will consider the bases $(\mathbf{e}, -\mathbf{f})$, $(\mathbf{f}, -\mathbf{e})$ and $(\mathbf{f}, \mathbf{e})$. Which have right orientations and which have left orientations?

(1) In order to transition from $(\mathbf{e}, \mathbf{f})$ to $(\mathbf{e}, -\mathbf{f})$ we use the matrix $T_{(\mathbf{e},-\mathbf{f})} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$. This is a reflection in the line spanned by $\mathbf{e}$. Its determinant is -1 and so $(\mathbf{e}, -\mathbf{f})$ has the opposite orientation to $(\mathbf{e}, \mathbf{f})$. We may say that it has a *right orientation* or that it is a *right basis*.

(2) The transition matrix from $(\mathbf{e}, \mathbf{f})$ to $(\mathbf{f}, -\mathbf{e})$ is $T_{(\mathbf{f},-\mathbf{e})} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$. This has determinant 1 and so we may say that both $(\mathbf{e}, \mathbf{f})$ and $(\mathbf{f}, -\mathbf{e})$ are *left bases*. Note that $T_{(\mathbf{f},-\mathbf{e})}$ is a rotation of the original basis by an angle of $\pi/2$.

(3) The transition matrix from $(\mathbf{e}, \mathbf{f})$ to $(\mathbf{f}, \mathbf{e})$ is $T_{(\mathbf{f},\mathbf{e})} = \left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right]$, which has determinant -1. Thus $(\mathbf{f}, \mathbf{e})$ is a right basis. Note that the transition here is a reflection in the line bisecting the angle between $\mathbf{e}$ and $\mathbf{f}$.

Of course if we had declared our initial basis to be a right basis, then all terms left and right would have to be interchanged. The choice is entirely arbitrary.

Example 1.67 Let $\{\mathbf{e}_x, \mathbf{e}_y, \mathbf{e}_z\}$ be a basis of $\mathbb{E}^3$ and let

$$T = \begin{bmatrix} a_x & b_x & c_x \\ a_y & b_y & c_y \\ a_z & b_z & c_z \end{bmatrix}$$

be any 3×3 matrix with entries in $\mathbb{R}$. Let $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$ be defined by

$$\mathbf{a} = a_x\mathbf{x} + a_y\mathbf{y} + a_z\mathbf{z}$$
$$\mathbf{b} = b_x\mathbf{x} + b_y\mathbf{y} + b_z\mathbf{z}$$
$$\mathbf{c} = c_x\mathbf{x} + c_y\mathbf{y} + c_z\mathbf{z}$$

We have three cases:

$\det T > 0$**:** In this case $(\mathbf{a}, \mathbf{b}, \mathbf{c})$ is a basis and this basis has the same orientation as $(\mathbf{e}_x, \mathbf{e}_y, \mathbf{e}_z)$.

$\det T < 0$**:** In this case $(\mathbf{a}, \mathbf{b}, \mathbf{c})$ is a basis and this basis has the opposite orientation to $(\mathbf{e}_x, \mathbf{e}_y, \mathbf{e}_z)$.

$\det T = 0$**:** In this case the set of vectors $\{\mathbf{a}, \mathbf{b}, \mathbf{c}\}$ are not linearly-independent and hence do not define a basis. As such they do not have an orientation.

Notice that since T was chosen arbitrarily, even if $(\mathbf{a}, \mathbf{b}, \mathbf{c})$ is a basis it need not be orthonormal.

Remark 1.68 The important message from this section is that there exactly two orientations of any real vector space. Given a basis as a reference point, any other basis either has the same orientation or the opposite orientation as this reference point.

If two bases $\mathcal{B}$ and $\mathcal{C}$ have the same orientation then one can transform from one basis to the other via a continuous transformation. Making this statement precise is beyond the scope of this course, however we can demonstrate the point in $\mathbb{E}^3$. If $\mathcal{B}$ and $\mathcal{C}$ are orthonormal bases of $\mathbb{E}^3$ with the same orientation then there is an axis $\mathbf{v}$ such that the transformation of $\mathcal{B}$ to $\mathcal{C}$ is given by a rotation about $\mathbf{v}$. This is Euler's Theorem and will be proved in Theorem 1.77.

1.5.1 Orientation of linear operator

Let P be a linear operator acting on a vector space V and let $\mathcal{B} = (\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n)$ be a chosen basis for V.

Consider the image of this basis $P\mathcal{B} = (P\mathbf{e}_1, P\mathbf{e}_2, \ldots, P\mathbf{e}_n)$. If P is nondegenerate then this is again a basis for V. We want to consider its orientation. As we have already seen $[P]_{\mathcal{B}}$ is the same as the transition matrix of $\mathcal{B}$ to $P\mathcal{B}$. Thus we know that if $[P]_{\mathcal{B}}$ has positive determinant then $\mathcal{B}$ and $P\mathcal{B}$ have the same orientation. If the matrix has a negative determinant then the two bases have opposite orientations .

If the determinant is zero, so P is degenerate, then $P\mathcal{B}$ is not a basis.

Since the determinant of P as a linear operator is defined to be the same as the determinant of $[P]_{\mathcal{B}}$ as a matrix we can now say the follow:

- If a linear operator P has positive determinant then the action of P preserves the orientation of a basis.
- If a linear operator P has negative determinant then the action of P swaps the orientation of a basis to the opposite orientation.

Definition 1.69 Let P be a nondegenerate (invertible) linear operator acting on a vector space V.

We say that P *preserves the orientation* of V if $\det P > 0$.

We say that P *changes the orientation* of V if $\det P < 0$.

1.6 Orthogonal operators of $\mathbb{E}^n$

Recall the notion of orthogonal operator (see Subsubsection 1.4.4). In this section, we shall consider orthogonal operators in $\mathbb{E}^2$ and $\mathbb{E}^3$. In particular, we shall try to classify them and show how they relate to geometric notions that we are familiar with: rotations and reflections.

1.6.1 Orthogonal operators in $\mathbb{E}^2$

In this section, we will show that an orthogonal operator in $\mathbb{E}^2$ induces either a rotation or a reflection of $\mathbb{E}^2$, depending on whether it preserves orientation or not.

Throughout, we let $\mathcal{B} = (\mathbf{e}_1, \mathbf{e}_2)$ be an orthonormal basis $\mathbb{E}^2$. Recall that this implies

$$\langle \mathbf{e}_1, \mathbf{e}_1 \rangle = \langle \mathbf{e}_2, \mathbf{e}_2 \rangle = 1 \quad \langle \mathbf{e}_1, \mathbf{e}_2 \rangle = 0,$$

i.e., vectors $\mathbf{e}_1, \mathbf{e}_2$ have unit length and are orthogonal to each other. Also recall that by fixing an ordering on the basis, we have fixed an orientation on the basis and therefore the vector space. We shall call this the *left orientation*.

Let P be an orthogonal operator acting on $\mathbb{E}^2$, *i.e.*,

$$\langle P(\mathbf{x}),\, P(\mathbf{y})\rangle = \langle \mathbf{x},\, \mathbf{y}\rangle \text{ for all } \mathbf{x}, \mathbf{y} \in \mathbb{E}^2.$$

Consider a new basis $\mathcal{C} = (\mathbf{f}_1, \mathbf{f}_2)$ defined by applying P to $\mathcal{B}$:

$$\mathbf{f}_1 = P(\mathbf{e}_1) = \alpha\mathbf{e}_1 + \gamma\mathbf{e}_2,\ \alpha, \gamma \in \mathbb{R}$$
$$\mathbf{f}_2 = P(\mathbf{e}_2) = \beta\mathbf{e}_1 + \delta\mathbf{e}_2,\ \beta, \delta \in \mathbb{R}$$

By Proposition 1.58, as P is an orthogonal operator, this new basis $\mathcal{C}$ is an orthonormal basis.

Because P is orthogonal, there are extra restriction on $\alpha, \beta, \gamma, \delta \in \mathbb{R}$ we have not considered yet. Consider the matrix for P in the basis $\mathcal{B}$:

$$[P]_{\mathcal{B}} = \begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix}.$$

By Proposition 1.58, P is orthogonal if and only if $[P]_{\mathcal{B}}$ is an orthogonal matrix:

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = ([P]_{\mathcal{B}})^{\mathsf{T}}[P]_{\mathcal{B}} = \begin{bmatrix} \alpha & \gamma \\ \beta & \delta \end{bmatrix}\begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix} = \begin{bmatrix} \alpha^2+\gamma^2 & \alpha\beta+\gamma\delta \\ \alpha\beta+\gamma\delta & \beta^2+\delta^2 \end{bmatrix}$$

This gives us three extra constraints on what the entries of $[P]_{\mathcal{B}}$ could be

$$\alpha^2 + \gamma^2 = 1, \quad \beta^2 + \delta^2 = 1, \quad \alpha\beta + \gamma\delta = 0. \tag{1.20}$$

Recall that $\alpha^2 + \gamma^2 = 1$ and $\beta^2 + \delta^2 = 1$ are the equations that cut out a circle of radius one. As a result, we can satisfy these equations immediately by picking angles φ, ψ and setting

$$\alpha = \cos\varphi,\ \gamma = \sin\varphi,\ \beta = \cos\psi,\ \delta = \sin\psi. \tag{1.21}$$

The final constraint $\alpha\beta + \gamma\delta = 0$ implies that

$$\cos\varphi\cos\psi + \sin\varphi\sin\psi = \cos(\psi - \varphi) = 0.$$

Observation 1.70 You may notice that we could have equally picked $\alpha = \sin\varphi, \gamma = \cos\varphi$ (similarly for β, δ) and still satisfied the constraints. However, this can be put in the form of equation (1.21) by picking a different angle:

$$\cos(\frac{\pi}{2} - \varphi) = \cos(\varphi - \frac{\pi}{2}) = \cos\varphi\cos\frac{\pi}{2} + \sin\varphi\sin\frac{\pi}{2} = \sin\varphi$$
$$\sin(\frac{\pi}{2} - \varphi) = -\sin(\varphi - \frac{\pi}{2}) = -\sin\varphi\cos\frac{\pi}{2} + \cos\varphi\sin\frac{\pi}{2} = \cos\varphi$$

Therefore without loss of generality, we can always put $\alpha, \beta, \gamma, \delta$ in the form (1.21).

Let us pause to reinforce what we have shown so far:

Lemma 1.71 *Let $\mathcal{B}$ be an orthonormal basis and P a linear operator of $\mathbb{E}^2$. Then P is an orthogonal operator if and only if its matrix $[P]_{\mathcal{B}}$ can be written in the form*

$$[P]_{\mathcal{B}} = \begin{bmatrix} \cos\varphi & \cos\psi \\ \sin\varphi & \sin\psi \end{bmatrix},$$

and satisfies the constraint $\cos(\psi - \varphi) = 0$.

The condition $\cos(\psi - \varphi) = 0$ only occurs when $\psi - \varphi = \frac{\pi}{2} + k\pi$ for some integer $k \in \mathbb{Z}$. The geometry of P varies depending on whether k is odd or even. This splits the orthogonal operators of $\mathbb{E}^2$ into two classes:

(1) when $\psi = \varphi + \frac{\pi}{2} + 2m\pi$ for some $m \in \mathbb{Z}$ ($k = 2m$ is even),

(2) when $\psi = \varphi - \frac{\pi}{2} + 2m\pi$ for some $m \in \mathbb{Z}$ ($k = 2m - 1$ is odd).

Case 1: Rotations in $\mathbb{E}^2$ We first consider the case where k is even. Let $k = 2m$ for some integer $m \in \mathbb{Z}$. Then $\psi = \varphi + \frac{\pi}{2} + 2m\pi$ and so

$$\cos\psi = \cos\left(\varphi + \frac{\pi}{2} + 2m\pi\right) = \cos\left(\varphi + \frac{\pi}{2}\right) = -\sin\varphi,$$
$$\sin\psi = \sin\left(\varphi + \frac{\pi}{2} + 2m\pi\right) = \sin\left(\varphi + \frac{\pi}{2}\right) = \cos\varphi.$$

Therefore the orthogonal operator depends on a single parameter φ. To emphasise this, we denote the operator P_φ and write its matrix as

$$[P_\varphi]_{\mathcal{B}} = \begin{bmatrix} \cos\varphi & \cos\psi \\ \sin\varphi & \sin\psi \end{bmatrix} = \begin{bmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{bmatrix}. \tag{1.22}$$

Therefore, P_φ acts on an arbitrary vector $\mathbf{x} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2$ as follows:

$$[P_\varphi]_{\mathcal{B}}[\mathbf{x}]_{\mathcal{B}} = \begin{bmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1\cos\varphi - x_2\sin\varphi \\ x_1\sin\varphi + x_2\cos\varphi \end{bmatrix}$$
$$\Rightarrow P_\varphi(\mathbf{x}) = (x_1\cos\varphi - x_2\sin\varphi)\mathbf{e}_1 + (x_1\sin\varphi + x_2\cos\varphi)\mathbf{e}_2 \tag{1.23}$$

Note also that $\det P_\varphi = \det [P_\varphi]_{\mathcal{B}} = 1$ and so P_φ *preserves orientation.*

What is the geometric behaviour of P_φ? You may recognise the matrix (1.22) as a *rotation matrix.* Explicitly, if we apply the operator P_φ to the vector $\mathbf{x}$, it *rotates* $\mathbf{x}$ by the angle φ.

Example 1.72 Fix some orthonormal basis $\mathcal{B} = (\mathbf{e}_1, \mathbf{e}_2)$ for $\mathbb{E}^2$. Consider the vector $\mathbf{x} = 2\mathbf{e}_1 + \mathbf{e}_2$ and let $\varphi = \frac{\pi}{2}$. What is the P_φ?

Then using equation (1.23), we see

$$\begin{aligned} P_{\frac{\pi}{2}}(\mathbf{x}) &= \left(2\cos\left(\frac{\pi}{2}\right) - \sin\left(\frac{\pi}{2}\right)\right)\mathbf{e}_1 + \left(2\sin\left(\frac{\pi}{2}\right) + \cos\left(\frac{\pi}{2}\right)\right)\mathbf{e}_2 \\ &= -\mathbf{e}_1 + 2\mathbf{e}_2 \end{aligned}$$

i.e., $P_{\frac{\pi}{2}}$ rotates $\mathbf{x}$ by $\frac{\pi}{2}$ radians, see Figure 1.

You may notice that we have not addressed in this example whether the rotation acts clockwise or anti-clockwise. These two notions do not make sense in an arbitrary vector space: it is completely determined by the *orientation* of the vector space and therefore the ordering of the basis. Explicitly, the rotation operator P_φ rotates the first basis vector towards the second basis vector.

Figure 1 shows Example 1.72 for two different choices of basis. In both cases, the rotation operator rotates $\mathbf{x}$ by $\frac{\pi}{2}$ radians in the direction of $\mathbf{e}_1$ to $\mathbf{e}_2$.

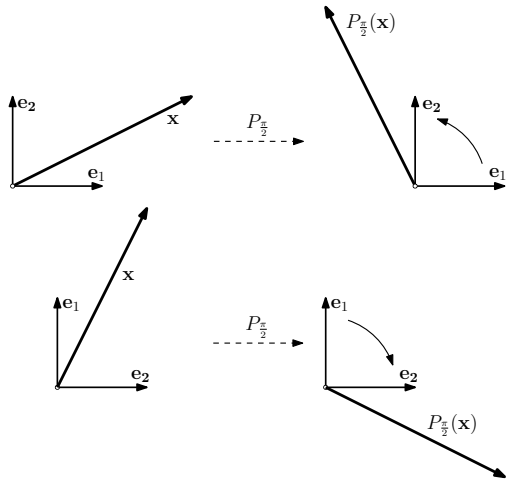


Figure 1: The rotation operator $P_{\frac{\pi}{2}}$ from Example 1.72 for two different ordered bases. The operator rotates the vector $\mathbf{x}$ by $\frac{\pi}{2}$ radians in the direction of $\mathbf{e}_1$ to $\mathbf{e}_2$.

Case 2: Reflections in $\mathbb{E}^2$ We now consider the case where k is odd. Let $k = 2m - 1$ for some integer $m \in \mathbb{Z}$. Then $\psi = \varphi - \frac{\pi}{2} + 2m\pi$ and so

$$\begin{aligned} \cos\psi &= \cos\left(\varphi - \frac{\pi}{2} + 2m\pi\right) = \cos\left(\varphi - \frac{\pi}{2}\right) = \sin\varphi \\ \sin\psi &= \sin\left(\varphi - \frac{\pi}{2} + 2m\pi\right) = \sin\left(\varphi - \frac{\pi}{2}\right) = -\cos\varphi \end{aligned}$$

Again, the orthogonal operator depends on a single variable φ. We

shall denote the operator Q_φ and write its matrix as:

$$[Q_\varphi]_{\mathcal{B}} = \begin{bmatrix} \cos\varphi & \cos\psi \\ \sin\varphi & \sin\psi \end{bmatrix} = \begin{bmatrix} \cos\varphi & \sin\varphi \\ \sin\varphi & -\cos\varphi \end{bmatrix}. \tag{1.24}$$

By a similar calculation to (1.23), Q_φ acts on an arbitrary vector $\mathbf{x} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2$ as follows:

$$[Q_\varphi]_{\mathcal{B}}[\mathbf{x}]_{\mathcal{B}} = \begin{bmatrix} \cos\varphi & \sin\varphi \\ \sin\varphi & -\cos\varphi \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1\cos\varphi + x_2\sin\varphi \\ x_1\sin\varphi - x_2\cos\varphi \end{bmatrix}$$
$$\Rightarrow Q_\varphi(\mathbf{x}) = (x_1\cos\varphi + x_2\sin\varphi)\mathbf{e}_1 + (x_1\sin\varphi - x_2\cos\varphi)\mathbf{e}_2 \tag{1.25}$$

Note that in this case, $\det Q_\varphi = \det[Q_\varphi]_{\mathcal{B}} = -1$, and so Q_φ does *not* preserve orientation.

We saw that P_φ preserves orientation whereas Q_φ doesn't. How does the geometry of Q_φ compare to P_φ? To answer this, we introduce a new linear operator R on $\mathbb{E}^2$ defined as follows:

$$R(\mathbf{e}_1) = \mathbf{e}_1\,,\ R(\mathbf{e}_2) = -\mathbf{e}_2 \quad \Rightarrow \quad [R]_{\mathcal{B}} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

R is the *reflection* that sends $\mathbf{e}_2$ to $-\mathbf{e}_2$, or alternatively is the reflection in the line spanned by $\mathbf{e}_1$. Recall from Example 1.66 that this operator does not preserve orientation. Furthermore, by comparing $[P_\varphi]_{\mathcal{B}}$ and $[Q_\varphi]_{\mathcal{B}}$ we can deduce that Q_φ is *the composition of a rotation and a reflection.*

$$[Q_\varphi]_{\mathcal{B}} = \begin{bmatrix} \cos\varphi & \sin\varphi \\ \sin\varphi & -\cos\varphi \end{bmatrix} = \begin{bmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} = [P_\varphi]_{\mathcal{B}}[R]_{\mathcal{B}}. \tag{1.26}$$

Example 1.73 Similar to the previous example, we fix some orthonormal basis $(\mathbf{e}_1, \mathbf{e}_2)$ for $\mathbb{E}^2$ and consider the vector $\mathbf{x} = 2\mathbf{e}_1 + \mathbf{e}_2$. Let $\varphi = 0$. What is Q_0?

Q_0 should simply be R, the reflection in the line spanned by $\mathbf{e}_1$, *i.e.*, that sends $\mathbf{e}_2$ to $-\mathbf{e}_2$. By equation (1.25) we get:

$$\begin{aligned} Q_0(\mathbf{x}) &= (2\cos 0 + \sin 0)\,\mathbf{e}_1 + (2\sin 0 - \cos 0)\,\mathbf{e}_2 \\ &= 2\mathbf{e}_1 - \mathbf{e}_2 \end{aligned}$$

Let us consider Q_φ for $\varphi = \frac{\pi}{2}$:

$$\begin{aligned} Q_{\frac{\pi}{2}}(\mathbf{x}) &= \left(2\cos\left(\frac{\pi}{2}\right) + \sin\left(\frac{\pi}{2}\right)\right)\mathbf{e}_1 + \left(2\sin\left(\frac{\pi}{2}\right) - \cos\left(\frac{\pi}{2}\right)\right)\mathbf{e}_2 \\ &= \mathbf{e}_1 + 2\mathbf{e}_2 \end{aligned}$$

There are two geometric interpretations of this operator. The first is that it is the reflection in the line spanned by $\mathbf{e}_1$, followed by a rotation by $\frac{\pi}{2}$ radians from $\mathbf{e}_1$ to $\mathbf{e}_2$. The second, cleaner, formulation is that it is the reflection in the line spanned by $\mathbf{e}_1 + \mathbf{e}_2$ (see Figure 2).

In fact, any two dimensional reflection can be realised by a composition of reflection operator R and a rotation.

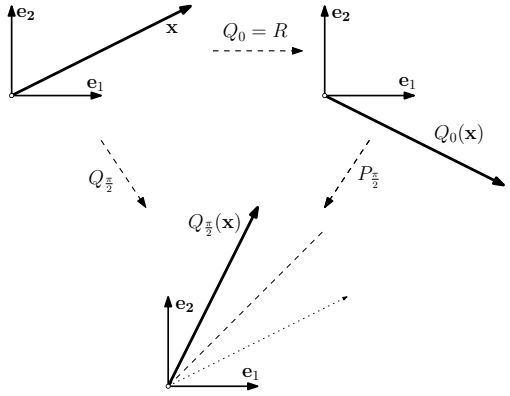


Figure 2: The reflections $Q_0 = R$ and $Q_{\frac{\pi}{2}}$ applied to the vector $\mathbf{x}$ from Example 1.73. Note that $Q_{\frac{\pi}{2}}$ can be viewed as the reflection in the line $\mathbf{e}_1 + \mathbf{e}_2$, or as the composition of R and $P_{\frac{\pi}{2}}$.

This completes a full characterisation of orthogonal operators in $\mathbb{E}^2$. We recall what we have shown in the following proposition.

Proposition 1.74 *Let P be an arbitrary orthogonal linear operator on $\mathbb{E}^2$, then $\det P = \pm 1$.*

If $\det P = 1$ then there exists an angle $\varphi \in [0, 2\pi)$ such that $P = P_\varphi$ is the operator that rotates a vector by φ.

If $\det P = -1$ then there exists an angle $\varphi \in [0, 2\pi)$ such that $P = Q_\varphi$ is the operator that rotates a vector by φ composed with a reflection.

Week 4

1.6.2 Orthogonal operators in $\mathbb{E}^3$ and rotations

In the previous section, we saw that the orthogonal operators of $\mathbb{E}^2$ that preserved orientation were rotation operators. The main result in this section will be Euler's Rotation Theorem that shows the same holds in $\mathbb{E}^3$. We will give the precise statement of Euler's Theorem at the end of this section. For now, we will just formulate a preliminary statement:

Theorem 1.75 *An orthogonal operator in $\mathbb{E}^3$ that preserves orientation is a rotation about an axis L by the angle φ.*

We shall slowly build up this statement by precisely defining rotations in $\mathbb{E}^3$ and how we can derive the axis and angle via standard linear algebra techniques.

Recall that a linear operator on $\mathbb{E}^2$ is a rotation by φ if it is of the form P_φ, see (1.22) and (1.23). The definition of a rotation is little bit more subtle in $\mathbb{E}^3$. Explicitly, a rotation in $\mathbb{E}^3$ occurs around an *axis*.

Let $\mathbf{n} \neq \mathbf{0}$ be an arbitrary non-zero vector in $\mathbb{E}^3$. Consider the line

$$L_{\mathbf{n}} = \operatorname{span}(\mathbf{n}) = \left\{ \lambda \mathbf{n} \in \mathbb{E}^3 \;\middle|\; \lambda \in \mathbb{R}_{\geq 0} \right\}$$

spanned by vector $\mathbf{n}$. We say $L_{\mathbf{n}}$ is the *axis* directed along the vector $\mathbf{n}$.

Note that $L_{\mathbf{n}}$ depends only on the direction of the vector $\mathbf{n}$, not the magnitude, *i.e.*, $L_{\mathbf{n}} = L_{\lambda \mathbf{n}}$ for all $\lambda \neq 0$. As a result, we shall often consider the *normalisation* of $\mathbf{n}$

Wikipedia: normalisation

$$\widehat{\mathbf{n}} = \frac{\mathbf{n}}{\|\mathbf{n}\|}, \tag{1.27}$$

i.e., the unit vector in the direction of $\mathbf{n}$. This will allow us to work with an orthonormal basis.

Definition 1.76 Let P be a linear operator on $\mathbb{E}^3$, and $\mathcal{B} = (\hat{\mathbf{n}}, \mathbf{f}, \mathbf{g})$ an orthonormal basis. We say P is a *rotation about the axis* $L_{\mathbf{n}}$ by the

Wikipedia: rotation about the axis

angle φ if it acts on $\mathcal{B}$ as follows:

$$P(\widehat{\mathbf{n}}) = \widehat{\mathbf{n}} \tag{1.28}$$

$$P(\mathbf{f}) = \mathbf{f}\cos\varphi + \mathbf{g}\sin\varphi \tag{1.29}$$

$$P(\mathbf{g}) = -\mathbf{f}\sin\varphi + \mathbf{g}\cos\varphi \tag{1.30}$$

i.e., the matrix representation of P in the basis $\mathcal{B}$ is

$$[P]_{\mathcal{B}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\varphi & -\sin\varphi \\ 0 & \sin\varphi & \cos\varphi \end{bmatrix}. \tag{1.31}$$

Let's break down this definition and see what is happening geometrically.

- (1.28) implies that P fixes all vectors on the axis $L_{\mathbf{n}} = L_{\widehat{\mathbf{n}}}$, as

  $$P(\lambda\widehat{\mathbf{n}}) = \lambda P(\widehat{\mathbf{n}}) = \lambda\widehat{\mathbf{n}}$$

  for any scalar $\lambda \in \mathbb{R}$. Note that in linear algebra terms, this is equivalent to $\mathbf{n}$ (and, by Lemma 1.55, any scalar multiple of $\mathbf{n}$) being an eigenvector of P with eigenvalue1.

- (1.29) and (1.30) state that if we restrict to the two dimensional plane spanned by $(\mathbf{f}, \mathbf{g})$, then P behaves exactly like a two dimensional rotation and rotates the plane by the angle φ. As the basis is orthonormal, this plane is orthogonal to $\widehat{\mathbf{n}}$, and so the plane rotates around the axis $L_{\mathbf{n}}$.

Recalling the definition of the trace of linear operator, we see from (1.31) that

$$\mathrm{Tr}(P) = 1 + 2\cos\varphi \tag{1.32}$$

where φ is angle of rotation. Recall by Proposition 1.50, that the trace of P does not depend on the choice of basis. This formula determines the cosine of the angle of rotation purely in terms of the operator rather than its matrix with respect to a basis. Furthermore, as $\cos(\varphi) = \cos(-\varphi)$ it determines the angle of rotation *up to sign*.

$[P]_{\mathcal{B}}$ is an orthogonal matrix, therefore Proposition 1.58 implies P is orthogonal. It preserves orientation as $\det P = \det [P]_{\mathcal{B}} = 1$

It is quick to check that P is orthogonal and preserves orientation. Euler's Theorem states that the converse is true: any orientation preserving orthogonal operator of $\mathbb{E}^3$ is a rotation.

Theorem 1.77 (Euler's Rotation Theorem) *Let P be an orientation preserving orthogonal operator of $\mathbb{E}^3$. Then P is a rotation around an axis L by the angle φ.*

Explicitly, the axis L is the one dimensional space of eigenvectors of

P with eigenvalue 1, and the angle φ is determined up to sign by

$$\mathrm{Tr}(P) = 1 + 2\cos\varphi.$$

Not only does this theorem completely characterise which operators define rotations, it tells us precisely what the geometry of that rotation is using only linear algebra. Before we prove this, let's consider an example that demonstrates the power of Euler's Theorem.

Example 1.78 Consider the linear operator P we saw in Example 1.56 that maps the orthonormal basis $\mathcal{B} = (\mathbf{e}_x, \mathbf{e}_y, \mathbf{e}_z)$ to

$$P(\mathbf{e}_x) = \mathbf{e}_y\,,\; P(\mathbf{e}_y) = \mathbf{e}_x\,,\; P(\mathbf{e}_z) = -\mathbf{e}_z. \tag{1.33}$$

Is P an orientation preserving orthogonal operator? If so, what is the angle of rotation?

This operator swaps the first two basis vectors and reflects the third one. Recall the matrix of operator in the basis $\mathcal{B}$ is given by

$$[P]_{\mathcal{B}} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix}$$

We can immediately check that P is an orthogonal operator as $[P]_{\mathcal{B}}$ is an orthogonal matrix. Furthermore, the determinant is $\det P = 1$, and so P preserves orientation. Hence by Euler's Theorem, it must be a rotation.

We wish to know exactly what rotation P defines, therefore we shall calculate the axis of rotation L and the angle φ. Recall from Example 1.56 that the vector $\mathbf{e}_x + \mathbf{e}_y$ is an eigenvector with eigenvalue 1. Hence the axis of rotation is $L = \mathrm{span}(\mathbf{e}_x + \mathbf{e}_y)$. To find the angle, we use (1.32),

$$\mathrm{Tr}(P) = 1 + 2\cos\varphi = -1 \;\Rightarrow\; \varphi = \pi.$$

Therefore P is the rotation around the axis L by the angle π.

Remark 1.79 The identity operator I that leaves everything fixed, *i.e.*,

$$I(\mathbf{x}) = \mathbf{x} \text{ for all } \mathbf{x} \in \mathbb{E}^3$$

is an orthogonal operator that preserves orientation. However, intuitively we can see that I doesn't rotate anything, and this is reflected by Euler's Theorem. As $I(\mathbf{x}) = \mathbf{x}$ for all vectors, every vector is an eigenvector with eigenvalue 1, and so *every* line in $\mathbb{E}^3$ could be considered an axis of rotation. Furthermore, the trace of I is $\mathrm{Tr}(I) = 3$, therefore the angle φ is zero.

The proof of the Euler's Theorem has two parts: firstly that there exists an axis that P leaves fixed, and secondly that P rotates the remaining vectors around that axis. We prove these in the following two lemmas.

Lemma 1.80 *Let P be an orientation preserving orthogonal operator of $\mathbb{E}^3$. There exists an axis L such that $P(\mathbf{x}) = \mathbf{x}$ for all $\mathbf{x} \in L$.*

Proof. If P has eigenvalue 1, then L is the one-dimensional space of eigenvectors $\mathbf{x}$ with eigenvalue 1, *i.e.*, $P(\mathbf{x}) = \mathbf{x}$. Therefore we show that P has an eigenvalue 1.

The polynomial $C_P(\lambda) = \det(\lambda I - P)$ is called the *characteristic polynomial of P*. The roots of this polynomial are exactly the eigenvalues of P.
Wikipedia: Characteristic polynomial

Recall from linear algebra that P has an eigenvalue λ if and only if $\det(\lambda I - P) = 0$, where I is the identity operator. Therefore it remains to show that $\det(I - P) = 0$. We show this using many of the determinant properties stated in Proposition 1.46.

$$\begin{aligned}
\det(I-P) &= \det(P)\det(I-P) && (P \text{ preserves orientation})\\
&= \det(P^{\mathsf{T}})\det(I-P) && (\det M = \det M^{\mathsf{T}})\\
&= \det(P^{\mathsf{T}} - P^{\mathsf{T}}P) && (\det MN = \det M \det N)\\
&= \det(P^{\mathsf{T}} - I) && (P \text{ orthogonal})\\
&= \det(P^{\mathsf{T}} - I^{\mathsf{T}})\\
&= \det((P-I)^{\mathsf{T}}) && (M^{\mathsf{T}} + N^{\mathsf{T}} = (M+N)^{\mathsf{T}})\\
&= \det(P-I)\\
&= -\det(I-P) && (\det(-M) = (-1)^n \det(M))
\end{aligned}$$

Therefore $\det(I - P) = -\det(I - P) = 0$ and so 1 is an eigenvalue of P. By Lemma 1.55, the span of a corresponding eigenvector $\mathbf{n}$ forms a fixed axis $L_{\mathbf{n}}$. □

Lemma 1.81 *Let P be an orientation preserving orthogonal operator of $\mathbb{E}^3$ that fixes an axis L. Then P is a rotation about the axis L by some angle φ.*

Proof. Pick a unit vector $\widehat{\mathbf{n}}$ in L and pick an arbitrary orthonormal basis $(\widehat{\mathbf{n}}, \mathbf{f}, \mathbf{g})$ with $\widehat{\mathbf{n}}$ as the first basis vector. By Lemma 1.80 we know that $P(\widehat{\mathbf{n}}) = \widehat{\mathbf{n}}$, we wish to show that

$$P(\mathbf{f}) = 0 \cdot \widehat{\mathbf{n}} + \alpha\mathbf{f} + \beta\mathbf{g}, \quad P(\mathbf{g}) = 0 \cdot \widehat{\mathbf{n}} + \gamma\mathbf{f} + \delta\mathbf{g}, \quad \alpha, \beta, \gamma, \delta \in \mathbb{R}.$$

i.e., they have no $\widehat{\mathbf{n}}$ component.

Suppose $P(\mathbf{f}) = \mu\widehat{\mathbf{n}} + \alpha\mathbf{f} + \gamma\mathbf{g}$ and consider the inner product

$\langle P(\mathbf{f}),\, P(\widehat{\mathbf{n}})\rangle$. By inner product manipulation, we see that

$$\begin{aligned}\langle P(\mathbf{f}),\, P(\widehat{\mathbf{n}})\rangle &= \langle \mu\widehat{\mathbf{n}} + \alpha\mathbf{f} + \gamma\mathbf{g},\, \widehat{\mathbf{n}}\rangle \\ &= \mu\langle \widehat{\mathbf{n}},\, \widehat{\mathbf{n}}\rangle + \alpha\langle \mathbf{f},\, \widehat{\mathbf{n}}\rangle + \gamma\langle \mathbf{g},\, \widehat{\mathbf{n}}\rangle \\ &= \mu\end{aligned}$$

However, as P is orthogonal, we have $\langle P(\mathbf{f}),\, P(\widehat{\mathbf{n}})\rangle = \langle \mathbf{f},\, \widehat{\mathbf{n}}\rangle = 0$. Therefore $\mu = 0$ and $P(\mathbf{f})$ has no $\widehat{\mathbf{n}}$ component. A similar calculation holds for $P(\mathbf{g})$. As a result, the matrix of P is

$$[P]_{\mathcal{B}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \alpha & \beta \\ 0 & \gamma & \delta \end{bmatrix}$$

As the matrix is orthogonal and orientation preserving, we can repeat the calculations as in (1.20) and (1.22) to show that

$$P(\mathbf{f}) = \mathbf{f}\cos\varphi + \mathbf{g}\sin\varphi, \quad P(\mathbf{g}) = -\mathbf{f}\sin\varphi + \mathbf{g}\cos\varphi.$$

Therefore P is the rotation about the axis L by φ. □

1.7 Area, volume and determinant

You may have seen before that the area of a parallelogram and the volume of a parallelepiped can be calculated in terms of the vector (cross) product, which in turn is related to determinants. In this section we will give a rigorous definition of the vector product and prove the link to determinants. These formulas will help develop a geometrical understanding of the determinant of a linear operator.

1.7.1 Vector product in oriented $\mathbb{E}^3$

We begin with a formal definition of the vector product for an oriented 3-dimensional Euclidean vector space.

Definition 1.82 (Vector product) Let $V = \mathbb{E}^3$ be a 3-dimensional oriented Euclidean vector space. A *vector product* (also known as a *cross product*) is a map,

Wikipedia: vector product

$$- \times - \colon V \times V \to V$$

satisfying the following properties:

- The vector $\mathbf{x} \times \mathbf{y} \in V$ is orthogonal to $\mathbf{x}$ and $\mathbf{y}$, that is

$$\langle \mathbf{x},\, \mathbf{x} \times \mathbf{y}\rangle = 0 \quad \text{and} \quad \langle \mathbf{y},\, \mathbf{x} \times \mathbf{y}\rangle = 0 \qquad \text{(VP-}\perp\text{)}$$

- The product is anticommutative:

$$(\mathbf{x} \times \mathbf{y}) = -(\mathbf{y} \times \mathbf{x}) \qquad \text{(VP-AC)}$$

- The product is linear:

$$(\lambda\mathbf{x} + \mu\mathbf{y}) \times \mathbf{z} = \lambda(\mathbf{x} \times \mathbf{z}) + \mu(\mathbf{y} \times \mathbf{z}) \qquad \text{(VP-Lin)}$$

- For perpendicular vectors $\mathbf{x}$ and $\mathbf{y}$, the length of the vector product, $\mathbf{x} \times \mathbf{y}$, is equal to the area of the rectangle formed by $\mathbf{x}$ and $\mathbf{y}$:

$$\langle \mathbf{x}, \mathbf{y} \rangle = 0 \quad \Rightarrow \quad \|\mathbf{x} \times \mathbf{y}\| = \|\mathbf{x}\|\|\mathbf{y}\| \qquad \text{(VP-Len)}$$

- For linearly independent vectors $\mathbf{x}$ and $\mathbf{y}$ the basis

$$(\mathbf{x}, \mathbf{y}, \mathbf{x} \times \mathbf{y}) \quad \text{has the same orientation as } V. \qquad \text{(VP-O)}$$

Remark 1.83 As with the inner product (also known as scalar product), we can use the anticommutativity of the inner product together with linearity in the first variable to show linearity in the second variable:

$$\begin{aligned} \mathbf{z} \times (\lambda\mathbf{x} + \mu\mathbf{y}) &= -(\lambda\mathbf{x} + \mu\mathbf{y}) \times \mathbf{z} \\ &= -\lambda(\mathbf{x} \times \mathbf{z}) - \mu(\mathbf{y} \times \mathbf{z}) = \lambda(\mathbf{z} \times \mathbf{x}) + \mu(\mathbf{z} \times \mathbf{y}). \end{aligned}$$

Wikipedia: bilinear

This means that the vector product is actually *bilinear*.

Remark 1.84 Anticommutativity implies that $(\mathbf{x} \times \mathbf{x}) = -(\mathbf{x} \times \mathbf{x})$ so the self product is always zero. Moreover, if $\mathbf{x}$ and $\mathbf{y}$ are not linearly independent, so that for some λ we have $\mathbf{y} = \lambda\mathbf{x}$, then the vector product $\mathbf{x} \times \mathbf{y}$ is equal to:

$$\mathbf{x} \times \mathbf{y} = \mathbf{x} \times \lambda\mathbf{x} = \lambda(\mathbf{x} \times \mathbf{x}) = 0.$$

Remark 1.85 Axiom (VP-$\perp$) means that if $\mathbf{x}$ and $\mathbf{y}$ are linearly independent then $\mathbf{x} \times \mathbf{y}$ is perpendicular to the plane spanned by $\mathbf{x}$ and $\mathbf{y}$. This implies that $(\mathbf{x}, \mathbf{y}, \mathbf{x} \times \mathbf{y})$ is a basis.

Remark 1.86 The orientation of V is important: if W is the oriented vector space that contains the same vector space as V but comes with opposite orientation, then for vectors $\mathbf{x}$ and $\mathbf{y}$, the basis $(\mathbf{x}, \mathbf{y}, \mathbf{x} \times \mathbf{y})$ has the same orientation as V by axiom (VP-O). This basis has the opposite orientation to W and hence the vector product on W would give a vector in the opposite direction, yet still perpendicular to both $\mathbf{x}$ and $\mathbf{y}$.

1.7.2 Existence and uniqueness of the vector product

At this point we have specified a list of axioms that a vector product should satisfy, it is not yet clear that such a function need necessarily exist. Our goal now is to both show that such a product exists and also that the vector product is unique.

We shall begin by showing that if a vector product exists for the oriented 3-dimensional Euclidean space V, then this product is unique. Let us fix an orthonormal basis $\mathcal{B} = (\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$ that shares the same orientation as V. We assume that a vector product $- \times -$ exists for V and may deduce the following facts:

$$\mathbf{e}_i \times \mathbf{e}_i = \underline{0} \quad \forall i \in \{1,2,3\} \qquad \text{by Remark 1.84.}$$

$$\begin{aligned} \mathbf{e}_1 \times \mathbf{e}_2 &= \lambda \mathbf{e}_3 \quad \text{for some scalar } \lambda && \text{by axiom (VP-}\perp\text{)} \\ &= \pm \mathbf{e}_3 && \text{by axiom (VP-Len)} \\ &= \mathbf{e}_3 && \text{by axiom (VP-O).} \end{aligned}$$

Similarly, since the bases $(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$, $(\mathbf{e}_2, \mathbf{e}_3, \mathbf{e}_1)$ and $(\mathbf{e}_3, \mathbf{e}_1, \mathbf{e}_2)$ all have the same orientation:

$$\mathbf{e}_2 \times \mathbf{e}_3 = \mathbf{e}_1 \quad \text{and} \quad \mathbf{e}_3 \times \mathbf{e}_1 = \mathbf{e}_2.$$

Finally by anticommutativity (axiom (VP-AC)) we know that

$$\mathbf{e}_2 \times \mathbf{e}_1 = -\mathbf{e}_3, \ \mathbf{e}_3 \times \mathbf{e}_2 = -\mathbf{e}_1, \quad \text{and} \quad \mathbf{e}_1 \times \mathbf{e}_3 = -\mathbf{e}_2.$$

The facts above mean that the axioms determine the value of the vector product for any pair of elements in an orthonormal basis. From this we can determine the value for any pair of vectors by bilinearity. Explicitly, if $\mathbf{v} = v_1\mathbf{e}_1 + v_2\mathbf{e}_2 + v_3\mathbf{e}_3$ and $\mathbf{w} = w_1\mathbf{e}_1 + w_2\mathbf{e}_2 + w_3\mathbf{e}_3$ are arbitrary vectors then bilinearity gives:

$$\begin{aligned} \mathbf{v} \times \mathbf{w} &= (v_1\mathbf{e}_1 + v_2\mathbf{e}_2 + v_3\mathbf{e}_3) \times (w_1\mathbf{e}_1 + w_2\mathbf{e}_2 + w_3\mathbf{e}_3) \\ &= v_1w_1(\mathbf{e}_1 \times \mathbf{e}_1) + v_1w_2(\mathbf{e}_1 \times \mathbf{e}_2) + v_1w_3(\mathbf{e}_1 \times \mathbf{e}_3) \\ &\quad + v_2w_1(\mathbf{e}_2 \times \mathbf{e}_1) + v_2w_2(\mathbf{e}_2 \times \mathbf{e}_2) + v_2w_3(\mathbf{e}_2 \times \mathbf{e}_3) \\ &\quad + v_3w_1(\mathbf{e}_3 \times \mathbf{e}_1) + v_3w_2(\mathbf{e}_3 \times \mathbf{e}_2) + v_3w_3(\mathbf{e}_3 \times \mathbf{e}_3) \\ &= (v_2w_3 - v_3w_2)\mathbf{e}_1 + (v_3w_1 - v_1w_3)\mathbf{e}_2 + (v_1w_2 - v_2w_1)\mathbf{e}_3 \end{aligned}$$

It is convenient to represent this formula in the following very familiar

way:

$$\mathbf{v} \times \mathbf{w} = \det \begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{bmatrix} \tag{1.34}$$

At this point we have shown that the value of the vector product for two arbitrary vectors is given by the determinant formula (1.34). It remains to show that this formula satisfies all of the required axioms (VP-$\perp$) to (VP-O). The proof of this is left as an extended exercise (see Exercise Sheet 4, Question 4).

Week 5

1.7.3 Area of a parallelogram

Axiom (VP-Len) states that the length of the vector product of two perpendicular vectors is given by the area of a rectangle bordered by the pair. The following proposition generalises this to the area of the parallelogram formed by two arbitrary vectors.

We shall represent the parallelogram formed by two vectors, $\mathbf{v}$ and $\mathbf{w}$, with the notation $\parallelogram(\mathbf{v}, \mathbf{w})$.

Proposition 1.87 *The area of the parallelogram formed by the vectors* $\mathbf{x}$ *and* $\mathbf{y}$ *is given by the length of their vector product,* $\|\mathbf{x} \times \mathbf{y}\|$.

$$\text{Area}\big(\parallelogram(\mathbf{x}, \mathbf{y})\big) = \|\mathbf{x} \times \mathbf{y}\| \tag{1.35}$$

Proof. Consider the expansion $\mathbf{y} = \mathbf{y}_{\|} + \mathbf{y}_{\perp}$, where the vector $\mathbf{y}_{\perp}$ is orthogonal to the vector $\mathbf{x}$ and the vector $\mathbf{y}_{\|}$ is parallel to vector $\mathbf{x}$. The area of $\parallelogram(\mathbf{x}, \mathbf{y})$ is equal to the product of the length of the vector $\mathbf{x}$ (the base) and the length of vector $\mathbf{y}_{\perp}$ (the height).

On the other $\mathbf{x} \times \mathbf{y} = \mathbf{x} \times (\mathbf{y}_{\|} + \mathbf{y}_{\perp}) = \mathbf{x} \times \mathbf{y}_{\|} + \mathbf{x} \times \mathbf{y}_{\perp}$. But $\mathbf{x} \times \mathbf{y}_{\|} = 0$, because these vectors are collinear. Hence $\mathbf{x} \times \mathbf{y} = \mathbf{x} \times \mathbf{y}_{\perp} = \|\mathbf{x}\|\|\mathbf{y}_{\perp}\|$ because vectors $\mathbf{x}$ and $\mathbf{y}_{\perp}$ are perpendicular. □

This proposition is very important in understanding the meaning of the vector product. Succinctly, the vector product of two vectors is a vector that is orthogonal to the plane spanned by these vectors, with magnitude equal to the area of the parallelogram formed by the vectors. The direction of the vector is defined by orientation.

It is worth recalling a formula relating the area of a parallelogram to the length of its sides and the angle between them:

$$\|\mathbf{x} \times \mathbf{y}\| = \|\mathbf{x}\|\|\mathbf{y}\||\sin\theta|. \tag{1.36}$$

Compare this to the formula we saw for inner products:

$$\langle \mathbf{x}, \mathbf{y} \rangle = \|\mathbf{x}\|\|\mathbf{y}\| \cos\theta. \tag{1.7}$$

These two formulas demonstrate a fundamental property of the two dif-

ferent products:

- The inner product is *zero* if the pair of vectors are orthogonal.
- The vector product is *zero* if the pair of vectors are collinear.

In fact equation (1.36) can be derived from equation (1.7) using the identity $\|\mathbf{v}\times\mathbf{w}\|^2+\langle\mathbf{v},\,\mathbf{w}\rangle^2=\|\mathbf{v}\|^2\|\mathbf{w}\|^2$, which we prove in Lemma 1.88 below. Using this identity we have

$$\begin{aligned}\|\mathbf{x}\times\mathbf{y}\|^2+\langle\mathbf{x},\,\mathbf{y}\rangle^2&=\|\mathbf{x}\|^2\|\mathbf{y}\|^2\\&=\|\mathbf{x}\|^2\|\mathbf{y}\|^2(\sin^2\theta+\cos^2\theta)\\&=\|\mathbf{x}\|^2\|\mathbf{y}\|^2\sin^2\theta+\langle\mathbf{x},\,\mathbf{y}\rangle^2.\end{aligned}$$

Eliminating the inner product from both sides and taking square roots gives equation (1.36). Thus if we abstractly define angles using the inner product formula, this is consistent with doing so with the vector product formula.

Lemma 1.88 *For a pair of vectors* $\mathbf{v}$ *and* $\mathbf{w}$ *in* $\mathbb{E}^3$ *the following identity holds:*

$$\|\mathbf{v}\times\mathbf{w}\|^2+\langle\mathbf{v},\,\mathbf{w}\rangle^2=\|\mathbf{v}\|^2\|\mathbf{w}\|^2.$$

Proof. Fix an orthonormal basis $\mathcal{B}$ and let

$$[\mathbf{v}]_{\mathcal{B}}=\begin{bmatrix}v_1\\v_2\\v_3\end{bmatrix}\qquad\text{and}\qquad[\mathbf{w}]_{\mathcal{B}}=\begin{bmatrix}w_1\\w_2\\w_3\end{bmatrix}.$$

We can be a little lazy with signs in the proof since each determinant is squared.

Using the determinant formula we have the following:

$$\begin{aligned}\|\mathbf{v}\times\mathbf{w}\|^2&=(\det\left[\begin{smallmatrix}v_1&v_2\\w_1&w_2\end{smallmatrix}\right])^2+(\det\left[\begin{smallmatrix}v_1&v_3\\w_1&w_3\end{smallmatrix}\right])^2+(\det\left[\begin{smallmatrix}v_2&v_3\\w_2&w_3\end{smallmatrix}\right])^2\\&=(v_1w_2-v_2w_1)^2+(v_1w_3-v_3w_1)^2+(v_2w_3-v_3w_2)^2\\&=(v_1w_2)^2+(v_1w_3)^2+(v_2w_1)^2\\&\quad+(v_2w_3)^2+(v_3w_1)^2+(v_3w_2)^2\\&\quad-2v_1w_1v_2w_2-2v_1w_1v_3w_3-2v_2w_2v_3w_3\end{aligned}$$

Calculating the square of the inner product we have:

$$\begin{aligned}\langle\mathbf{v},\,\mathbf{w}\rangle^2&=(v_1w_1+v_2w_2+v_3w_3)^2\\&=(v_1w_1)^2+(v_2w_2)^2+(v_3w_3)^2\\&\quad+2v_1w_1v_2w_2+2v_1w_1v_3w_3+2v_2w_2v_3w_3\end{aligned}$$

Finally the square product of norms gives:

$$\begin{aligned}\|\mathbf{v}\|^2\|\mathbf{w}\|^2 &= (v_1^2+v_2^2+v_3^2)(w_1^2+w_2^2+w_3^2)\\ &= (v_1w_1)^2+(v_2w_2)^2+(v_3w_3)^2\\ &\quad+(v_1w_2)^2+(v_1w_3)^2+(v_2w_1)^2\\ &\quad+(v_2w_3)^2+(v_3w_1)^2+(v_3w_2)^2\end{aligned}$$

We can now see that $\|\mathbf{v}\times\mathbf{w}\|^2+\langle\mathbf{v},\mathbf{w}\rangle^2=\|\mathbf{v}\|^2\|\mathbf{w}\|^2$. □

1.7.4 Area and determinants in $\mathbb{E}^2$

Let $\mathbf{a}$ and $\mathbf{b}$ be two linearly independent vectors in a 2-dimensional Euclidean vector space, $\mathbb{E}^2$. We can consider the 2-dimensional space as a plane in an oriented 3-dimensional Euclidean space, $\mathbb{E}^3$. Our aim is to calculate the area of the parallelogram $\mathit{/\!\!/}(\mathbf{a},\mathbf{b})$ formed by vectors $\mathbf{a}$ and $\mathbf{b}$.

Let $\mathbf{n}$ be a unit vector in $\mathbb{E}^3$ which is orthogonal to $\mathbb{E}^2$, chosen so that the basis $(\mathbf{a},\mathbf{b},\mathbf{n})$ has the same orientation as $\mathbb{E}^3$. Axiom (VP-$\perp$) means that the vector product in $\mathbf{a}\times\mathbf{b}$ is proportional to the normal vector $\mathbf{n}$:

If we had selected the other unit normal vector $-\mathbf{n}$, the vector product would still have been proportional to $\mathbf{n}$, however we would need to replace α with $-\alpha$.

$$\mathbf{a}\times\mathbf{b}=\alpha\mathbf{n},\quad \text{where } \alpha \text{ is the area of } \mathit{/\!\!/}(\mathbf{a},\mathbf{b}).$$

Let $(\mathbf{e},\mathbf{f})$ be an orthonormal basis for the plane $\mathbb{E}^2$, again chosen in the order so that the orthonormal basis $(\mathbf{e},\mathbf{f},\mathbf{n})$ has the same orientation as $\mathbb{E}^3$. This is equivalently to choosing $(\mathbf{e},\mathbf{f})$ to have the same orientation as $(\mathbf{a},\mathbf{b})$. Let $\mathbf{a}=a_1\mathbf{e}+a_2\mathbf{f}$ and $\mathbf{b}=b_1\mathbf{e}+b_2\mathbf{f}$. Then

$$\mathbf{a}\times\mathbf{b}=\det\begin{bmatrix}\mathbf{e}&\mathbf{f}&\mathbf{n}\\ a_1&a_2&0\\ b_1&b_2&0\end{bmatrix}=\mathbf{n}\det\begin{bmatrix}a_1&a_2\\ b_1&b_2\end{bmatrix}\tag{1.37}$$

Thus $\alpha=\det\left[\begin{smallmatrix}a_1&a_2\\ b_1&b_2\end{smallmatrix}\right]$. If we had instead selected a basis with the opposite orientation, for example $(\mathbf{f},\mathbf{e},\mathbf{n})$, we would instead have α equal to the negative of the determinant. Thus if $(\mathbf{e}_1,\mathbf{e}_2)$ is any orthonormal basis for a 2-dimensional Euclidean space, with $\mathbf{v}=v_1\mathbf{e}_1+v_2\mathbf{e}_2$ and $\mathbf{w}=w_1\mathbf{e}_1+w_2\mathbf{e}_2$ arbitrary vectors then

$$\text{Area}\big(\mathit{/\!\!/}(\mathbf{v},\mathbf{w})\big)=\left|\det\begin{bmatrix}v_1&v_2\\ w_1&w_2\end{bmatrix}\right|.\tag{1.38}$$

Next we want to consider the action of a linear operator on the parallelogram formed by two vectors. We shall see in the next proposition that for a linear operator acting on a 2-dimensional space, the determinant of the linear operator controls how the area scales.

Proposition 1.89 *Let $P\colon \mathbb{E}^2 \to \mathbb{E}^2$ be a linear operator, and let $\mathbf{a}$ and $\mathbf{b}$ be vectors in $\mathbb{E}^2$. Denote the images under P by $\mathbf{a}' = P(\mathbf{a})$ and $\mathbf{b}' = P(\mathbf{b})$. Then*

$$\operatorname{Area}\big(\parallelogram(\mathbf{a}', \mathbf{b}')\big) = |\det P| \operatorname{Area}\big(\parallelogram(\mathbf{a}, \mathbf{b})\big).$$

Proof. Fix a basis $\mathcal{B} = (\mathbf{e}_1, \mathbf{e}_2)$ for $\mathbb{E}^2$ and let

$$[\mathbf{a}]_{\mathcal{B}} = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \qquad [\mathbf{b}]_{\mathcal{B}} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \qquad [P]_{\mathcal{B}} = \begin{bmatrix} p_{1,1} & p_{1,2} \\ p_{2,1} & p_{2,2} \end{bmatrix}.$$

We obtain the coefficients of $\mathbf{a}'$ and $\mathbf{b}'$ by multiplying the matrix by each column vector. This together with equation (1.38) of the last section gives:

$$\begin{aligned}
\operatorname{Area}\big(\parallelogram(\mathbf{a}', \mathbf{b}')\big) &= \left| \det \begin{bmatrix} p_{1,1}a_1 + p_{1,2}a_2 & p_{2,1}a_1 + p_{2,2}a_2 \\ p_{1,1}b_1 + p_{1,2}b_2 & p_{2,1}b_1 + p_{2,2}b_2 \end{bmatrix} \right| \\
&= \left| \det \left(\begin{bmatrix} a_1 & a_2 \\ b_1 & b_2 \end{bmatrix} \begin{bmatrix} p_{1,1} & p_{2,1} \\ p_{1,2} & p_{2,2} \end{bmatrix} \right) \right| \\
&= \left|\det P^{\mathsf{T}}\right| \cdot \left| \det \begin{bmatrix} a_1 & a_2 \\ b_1 & b_2 \end{bmatrix} \right| \\
&= |\det P| \operatorname{Area}\big(\parallelogram(\mathbf{a}, \mathbf{b})\big)
\end{aligned}$$

□

1.7.5 Volume and determinants in $\mathbb{E}^3$

The vector product of a pair of vectors is related with area of the parallelogram they form. We will now consider the parallelepiped formed by three vectors.

Let $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$ be three vectors in $\mathbb{E}^3$. We shall denote the parallelepiped formed by these three vectors with the notation $\parallelepiped(\mathbf{a}, \mathbf{b}, \mathbf{c})$. We may consider the parallelogram $\parallelogram(\mathbf{b}, \mathbf{c})$ as the base of the parallelepiped. The height vector $\mathbf{h}$, is now proportional to $\mathbf{b} \times \mathbf{c}$ (as it is perpendicular to both $\mathbf{b}$ and $\mathbf{c}$) and forms some angle θ, with $\mathbf{a}$.

The volume of $\parallelepiped(\mathbf{a}, \mathbf{b}, \mathbf{c})$ is equal to the length of the height vector $\mathbf{h}$, multiplied by the area of the base, $\parallelogram(\mathbf{b}, \mathbf{c})$.

$$\begin{aligned}
\operatorname{Vol}\big(\parallelepiped(\mathbf{a}, \mathbf{b}, \mathbf{c})\big) &= \operatorname{Area}\big(\parallelogram(\mathbf{b}, \mathbf{c})\big) \|\mathbf{h}\| \\
&= \operatorname{Area}\big(\parallelogram(\mathbf{b}, \mathbf{c})\big) \|\mathbf{a}\| |\cos\theta| \\
&= \|\mathbf{b} \times \mathbf{c}\| \|\mathbf{a}\| |\cos\theta| && (\textit{Proposition } 1.87) \\
&= |\langle \mathbf{a},\ \mathbf{b} \times \mathbf{c} \rangle| && (\textit{Equation } (1.7))
\end{aligned}$$

Let us express the vectors in terms of an orthonormal basis

$\mathcal{B} = (\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$ in the usual way:

$$[\mathbf{a}]_{\mathcal{B}} = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}, \quad [\mathbf{b}]_{\mathcal{B}} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}, \quad \text{and} \quad [\mathbf{c}]_{\mathcal{B}} = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix}.$$

We can expand the inner and vector products using the chosen basis:

$$\begin{aligned} \mathrm{Vol}\big(\mathbb{Z}(\mathbf{a}, \mathbf{b}, \mathbf{c})\big) &= |\langle \mathbf{a},\, \mathbf{b} \times \mathbf{c} \rangle| \\ &= \left| \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} \bullet \begin{bmatrix} \det \left[\begin{smallmatrix} b_2 & b_3 \\ c_2 & c_3 \end{smallmatrix}\right] \\ -\det \left[\begin{smallmatrix} b_1 & b_3 \\ c_1 & c_3 \end{smallmatrix}\right] \\ \det \left[\begin{smallmatrix} b_1 & b_2 \\ c_1 & c_2 \end{smallmatrix}\right] \end{bmatrix} \right| \\ &= \left| a_1 \det \left[\begin{smallmatrix} b_2 & b_3 \\ c_2 & c_3 \end{smallmatrix}\right] - a_2 \det \left[\begin{smallmatrix} b_1 & b_3 \\ c_1 & c_3 \end{smallmatrix}\right] + a_3 \det \left[\begin{smallmatrix} b_1 & b_2 \\ c_1 & c_2 \end{smallmatrix}\right] \right| \\ &= \left| \det \begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{bmatrix} \right|. \end{aligned}$$

Putting these together we come to the beautiful formula:

$$\mathrm{Vol}\big(\mathbb{Z}(\mathbf{a}, \mathbf{b}, \mathbf{c})\big) = |\langle \mathbf{a},\, \mathbf{b} \times \mathbf{c} \rangle| = \left| \det \begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{bmatrix} \right|. \tag{1.39}$$

Remark 1.90 Just as we remarked for the area of a parallelogram, sometimes it is useful to consider the *algebraic area* of a parallelepiped as either positive or negative. In this situation, we would define the volume to be $\langle \mathbf{a}, \mathbf{b} \times \mathbf{c} \rangle$ without taking the absolute value. The sign would then depend on the orientation of the vector space.

We can now state and prove a proposition for linear operators and volumes in $\mathbb{E}^3$, analogous to Proposition 1.89 that considered areas and operators in $\mathbb{E}^2$.

Proposition 1.91 *Let $P\colon \mathbb{E}^3 \to \mathbb{E}^3$ be a linear operators and let $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$ be vectors in $\mathbb{E}^3$. Denote the images under P by $\mathbf{a}' = P(\mathbf{a})$, $\mathbf{b}' = P(\mathbf{b})$ and $\mathbf{c}' = P(\mathbf{c})$. Then*

$$\mathrm{Vol}\big(\mathbb{Z}(\mathbf{a}', \mathbf{b}', \mathbf{c}')\big) = |\det P|\, \mathrm{Vol}\big(\mathbb{Z}(\mathbf{a}, \mathbf{b}, \mathbf{c})\big). \tag{1.40}$$

Proof. The arguments in the proof of Proposition 1.89 can be applied, mutatis mutandis, to the three dimensional case.

More succinctly (using $\det M = \det M^{\mathsf{T}}$) we can see that, having

fixed an orthonormal basis $\mathcal{B}$, the volume is given by

$$\begin{aligned}\mathrm{Vol}\big(\text{▱}(\mathbf{a}', \mathbf{b}', \mathbf{c}')\big) &= \left|\det\begin{bmatrix}[\mathbf{a}']_{\mathcal{B}} & [\mathbf{b}']_{\mathcal{B}} & [\mathbf{c}']_{\mathcal{B}}\end{bmatrix}\right| \\ &= \left|\det\Big([P]_{\mathcal{B}}\begin{bmatrix}[\mathbf{a}]_{\mathcal{B}} & [\mathbf{b}]_{\mathcal{B}} & [\mathbf{c}]_{\mathcal{B}}\end{bmatrix}\Big)\right| \\ &= |\det P|\,\mathrm{Vol}\big(\text{▱}(\mathbf{a}, \mathbf{b}, \mathbf{c})\big).\end{aligned}$$

□

2 Differential geometry

2.1 Affine space

Section 1 was entirely focused on vectors and vector spaces. However, geometry often deals with spaces whose the elements are "points". Moreover, we would like a space where both points and vectors can interact with one another. This leads to the notion of *affine spaces*.

Wikipedia: Euclidean affine space

Definition 2.1 (Euclidean affine space) Let $\mathbb{E}^n$ be an n-dimensional Euclidean vector space. A *Euclidean affine space* (associated with $\mathbb{E}^n$) is a set of points $\mathbb{A}^n$, along with an addition map that allows us to add points with vectors

$$\begin{aligned}\mathbb{A}^n \times \mathbb{E}^n &\to \mathbb{A}^n \\ (\mathbf{P}, \mathbf{v}) &\mapsto \mathbf{P} + \mathbf{v}\end{aligned}$$

such that

(1) $\forall\ \mathbf{v}, \mathbf{w} \in \mathbb{E}^n, \mathbf{P} \in \mathbb{A}^n$, $\mathbf{P} + (\mathbf{v} + \mathbf{w}) = (\mathbf{P} + \mathbf{v}) + \mathbf{w}$,

(2) $\forall\ \mathbf{P} \in \mathbb{A}^n$, $\mathbf{P} + \mathbf{0} = \mathbf{P}$,

(3) $\forall\ \mathbf{P}, \mathbf{Q} \in \mathbb{A}^n$, $\exists$ a unique $\mathbf{v} \in \mathbb{E}^n$ such that $\mathbf{P} + \mathbf{v} = \mathbf{Q}$.

Note that as they behave differently, we denote points in uppercase and vectors in lower case.

This definition may seem a bit abstract and unintuitive, but in fact $\mathbb{A}^n$ behaves exactly how we expect $\mathbb{R}^n$ to behave when doing geometry. We can add a point and a vector to get to a new point, we can add two vectors to get a new vector, but we cannot add points together.

Remark 2.2 Unlike vector spaces, affine spaces do not come with a fixed "zero" or "origin". If we want to use one, we have to make this choice.

Let $\mathbb{A}^n$ be a Euclidean affine space with associated vector space $\mathbb{E}^n$ with orthonormal basis $\mathcal{B} = (\mathbf{e}_1, \ldots, \mathbf{e}_n)$. We wish to find a way to describe the points of $\mathbb{A}^n$. This leads to the notion of a *coordinate system on* $\mathbb{A}^n$.

Definition 2.3 (Coordinate system) A *coordinate system on* $\mathbb{A}^n$ is a surjective map $\mathbb{R}^n \to \mathbb{A}^n$ that assigns every n-tuple of real numbers to a point $\mathbf{P}$ in $\mathbb{A}^n$.

Wikipedia: Coordinate system
Wikipedia: Coordinates in affine space

Note that there are many different choices of coordinate system, but we shall begin with the most natural choice, *Cartesian coordinates*.

Example 2.4 (Cartesian coordinates) We first pick some arbitrary point $\mathbf{O} \in \mathbb{A}^n$ to act as an "origin" of the space, as well as an orthonormal basis $\mathcal{B} = (\mathbf{e}_1, \ldots, \mathbf{e}_n)$ for $\mathbb{E}^n$. The *Cartesian coordinate system* assigns the n-tuple of real numbers $(x_1, \ldots, x_n)$ to the point

$$\mathbf{P} = \mathbf{O} + \mathbf{v} = \mathbf{O} + x_1\mathbf{e}_1 + \cdots + x_n\mathbf{e}_n, \tag{2.1}$$

where $\mathbf{v} = \sum_i x_i\mathbf{e}_i$ is some vector in $\mathbb{E}^n$. Note that by (3), there always exists some vector $\mathbf{v}$ such that $\mathbf{P} = \mathbf{O} + \mathbf{v}$, and so every point has some n-tuple associated with it.

In the definition of a coordinate system, $\mathbb{R}^n$ is used without any vector space structure: every element is just a list of n real numbers.

Wikipedia: Cartesian coordinate system

Once $\mathbf{O}$ and $\mathcal{B}$ have been fixed, the Cartesian coordinate representation of $\mathbf{P}$ is the column vector

$$\mathbf{P} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}. \tag{2.2}$$

Unless stated otherwise, we shall work with Cartesian coordinates. Note that we will use round brackets to denote that (2.2) represents a point rather than a vector.

Remark 2.5 Example 2.4 leads to exactly the same notion of Cartesian coordinates we are familiar with on $\mathbb{R}^n$: we have a coordinate axes spanned by $\mathcal{B}$ centered at the origin $\mathbf{O}$. The only difference is we had to explicitly choose $\mathbf{O}$ and $\mathcal{B}$.

Remark 2.6 A warning about literature: many authors use $\mathbb{R}^n$ for both the affine space and corresponding vector space. However this can lead to confusion, and so we shall stick with $\mathbb{A}^n$ to emphasise the difference with the vector space $\mathbb{E}^n$.

Example 2.7 (Polar coordinates) We know that Cartesian coordinates are not the only choice of coordinates in $\mathbb{A}^n$. For example, we can define *polar coordinates* on $\mathbb{A}^2$ in the following way. Fix an origin $\mathbf{O} \in \mathbb{A}^2$ and an orthonormal basis $(\mathbf{e}_1, \mathbf{e}_2)$, the tuple (r, θ) gets mapped to the point $\mathbf{P} = \mathbf{O} + \mathbf{v}$ where $\mathbf{v}$ has magnitude r and angle θ with $\mathbf{e}_1$, *i.e.*,

Wikipedia: polar coordinates

$$r = \|\mathbf{v}\|, \quad \langle \mathbf{v}, \mathbf{e}_1 \rangle = \|\mathbf{v}\| \cos\theta.$$

However we can also describe polar coordinate by how it relates to Cartesian coordinates. In particular, if (x, y) is the Cartesian representation,

it is related to polar coordinates via

$$x = r\cos\theta \qquad r = \sqrt{x^2+y^2}$$
$$y = r\sin\theta \qquad \theta = \arctan\left(\frac{y}{x}\right)$$

There are many more coordinate systems we may pick, however the underlying geometry should stay the same regardless of our choice. Therefore we shall work with Cartesian coordinates and show our methods hold for arbitrary choices of coordinates later.

Note Unless stated otherwise, we shall fix an origin $\mathbf{O} \in \mathbb{A}^n$ and an orthonormal basis $\mathcal{B} = (\mathbf{e}_1, \ldots, \mathbf{e}_n)$ for $\mathbb{E}^n$, and use Cartesian coordinates on $\mathbb{A}^n$ with respect to $\mathbf{O}, \mathcal{B}$.

2.1.1 Tangent spaces

In vector spaces, all vectors began at the same point. This is not the case with affine space, we have to specify which point a vector is based at. This leads to the notion of tangent vectors and tangent spaces.

Definition 2.8 (Tangent vectors and tangent spaces) Let $\mathbf{P}$ be a point in $\mathbb{A}^n$. A *tangent vector* $\mathbf{v_P}$ to $\mathbb{A}^n$ is a vector $\mathbf{v} \in \mathbb{E}^n$ beginning at the point $\mathbf{P} \in \mathbb{A}^n$.

Wikipedia: tangent vector

The *tangent space of* $\mathbb{A}^n$ *at* $\mathbf{P}$ is the set $T_{\mathbf{P}}(\mathbb{A}^n)$ of all tangent vectors $\mathbf{v_P}$ to $\mathbb{A}^n$ beginning at $\mathbf{P}$.

Wikipedia: tangent space of $\mathbb{A}^n$ at $\mathbf{P}$

We first note that the tangent space $T_{\mathbf{P}}(\mathbb{A}^n)$ is a vector space, as adding and scaling vectors does not change their base point. Furthermore, as we can add any vector in $\mathbb{E}^n$ to a point $\mathbf{P}$, the vector space $T_{\mathbf{P}}(\mathbb{A}^n)$ is a copy of $\mathbb{E}^n$, *i.e.*, $T_{\mathbf{P}}(\mathbb{A}^n)$ is isomorphic to $\mathbb{E}^n$.

As $T_{\mathbf{P}}(\mathbb{A}^n)$ is a vector space, we can describe it with a basis. In theory, we may pick a different basis $\mathcal{B}_{\mathbf{P}}$ for each tangent space $T_{\mathbf{P}}(\mathbb{A}^n)$, and we shall see later that for some coordinate systems this is the correct thing to do. However, as we are working with Cartesian coordinates on $\mathbb{A}^n$, **we shall fix the same basis $\mathcal{B} = (\mathbf{e}_1, \ldots, \mathbf{e}_n)$ for every tangent space $T_{\mathbf{P}}(\mathbb{A}^n)$.** This allows us to write elements of $T_{\mathbf{P}}(\mathbb{A}^n)$ as column vectors (with square brackets) as we did in $\mathbb{E}^n$.

Remark 2.9 The name tangent vector and tangent space may seem odd here, considering they don't appear to be "tangential" to anything. This connection will become more apparent when we define tangent vectors to curves and surfaces.

Remark 2.10 When doing vector calculus in $\mathbb{R}^3$, you may have come across the notation where tangent vectors are given by $a\mathbf{i}+b\mathbf{j}+c\mathbf{k}$. This gives a vector of the tangent space: without knowing the point at which the vector begins, it does not make sense on it's own.

Finally, we make a quick note about functions on affine space. As we are doing differential geometry, we want our functions to be as differentiable as possible!

Definition 2.11 (Smooth functions) Let f be a real-valued function on $\mathbb{A}^n$

$$f\colon \mathbb{A}^n \to \mathbb{R}$$
$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \mapsto f(x_1, \ldots, x_n).$$

We say f is *smooth* if every partial derivative of f exists, *i.e.*, Wikipedia: smooth

$$\frac{\partial^a f}{\partial x_1^{a_1} \cdots \partial x_n^{a_n}} \text{ exists for all } a_i \in \mathbb{Z}_{\geq 0},\ a_1 + \cdots + a_n = a.$$

We won't need to worry too much about this definition: all this means for us is we are free to differentiate smooth functions as much as we want and not worry about awkward issues such as whether a derivative exists or not. For us, practically all the functions we will work with will be smooth.

Week 6

2.2 Vector fields

We briefly recall a key object in vector calculus.

Wikipedia: vector field

Definition 2.12 (Vector field) Let $h_1, \ldots, h_n$ be real-valued functions on $\mathbb{A}^n$. A *vector field* $\mathbf{W}$ on $\mathbb{A}^n$ is a function

$$\mathbf{W}(x_1, \ldots, x_n) = h_1(x_1, \ldots, x_n)\mathbf{e}_1 + \cdots + h_n(x_1, \ldots, x_n)\mathbf{e}_n \tag{2.3}$$

that assigns to each point $\mathbf{P} \in \mathbb{A}^n$ a tangent vector $\mathbf{W}(\mathbf{P}) \in T_{\mathbf{P}}(\mathbb{A}^n)$ from its tangent space.

Example 2.13 Let $\mathbf{W}(x, y) = x\mathbf{e}_x + y\mathbf{e}_y$ be a vector field on $\mathbb{A}^2$. The vector field given by $\mathbf{W}$ is in Figure 4. Here $h_1(x, y) = x$ and $h_2(x, y) = y$.

Wikipedia: gradient of f

Vector fields often arise in mathematics and physics in the following way. Let $f\colon \mathbb{A}^n \to \mathbb{R}$ be a smooth function, the *gradient of* f is the vector field

$$\nabla f = \frac{\partial f}{\partial x_1}\mathbf{e}_1 + \cdots + \frac{\partial f}{\partial x_n}\mathbf{e}_n.$$

Wikipedia: conservative

A vector field that arises as the gradient of some function f is called a *conservative* vector field. Geometrically, the vector field ∇f points in the direction in which f increases the most.

Example 2.14 Define

$$f(x, y) = \frac{x^2 + y^2}{2}.$$

then $\frac{\partial f}{\partial x} = \frac{2x}{2} = x$ and $\frac{\partial f}{\partial y} = \frac{2y}{y} = y$. Therefore $\nabla f = \frac{\partial f}{\partial x}\mathbf{e}_x + \frac{\partial f}{\partial y}\mathbf{e}_y = x\mathbf{e}_x + y\mathbf{e}_y$ is precisely the vector field $\mathbf{W}$ defined in Example 2.13. Therefore $\mathbf{W}$ is conservative. Note that the vector field points in the direction that f increases the most, *i.e.*, orthogonally to the level sets.

Wikipedia: level set

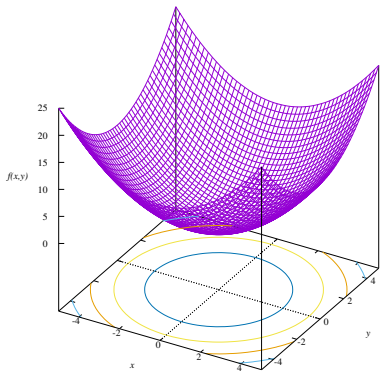


Figure 3: The graph of the function $f = \frac{x^2+y^2}{2}$ with level sets shown below.

Example 2.15 Is the vector field $\widetilde{\mathbf{W}}(x, y) = -y\mathbf{e}_x + x\mathbf{e}_y$ (shown in Figure 4) is conservative?

No. If it were conservative, there would exist some function f such that

$$\frac{\partial f}{\partial x} = -y, \quad \frac{\partial f}{\partial y} = x.$$

However as differentiation commutes, we arrive at a contradiction by considering

$$1 = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right) = \frac{\partial^2 f}{\partial x \partial y} = \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right) = -1.$$

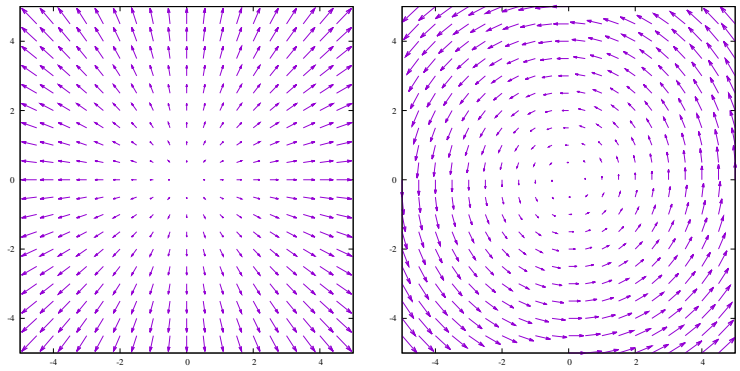


Figure 4: Two vector fields in $\mathbb{A}^2$. The left is the conservative vector field $\mathbf{W} = x\mathbf{e}_x + y\mathbf{e}_y$. The right is the non-conservative vector field $\widetilde{\mathbf{W}} = -y\mathbf{e}_x + x\mathbf{e}_y$.

We briefly recall the notion of the directional derivative of a function.

Definition 2.16 (Directional derivative) Let f be a smooth function on $\mathbb{A}^n$ and $\mathbf{v} \in T_{\mathbf{P}}(\mathbb{A}^n)$ a tangent vector to a point $\mathbf{P}$. The *directional*

Wikipedia: directional derivative of f at $\mathbf{P}$ along $\mathbf{v}$

derivative of f at $\mathbf{P}$ along $\mathbf{v}$ is

$$D_{\mathbf{v}}f(\mathbf{P}) = \nabla f(\mathbf{P}) \cdot \mathbf{v} = \frac{\partial f}{\partial x_1}(\mathbf{P})v_1 + \cdots + \frac{\partial f}{\partial x_n}(\mathbf{P})v_n. \tag{2.4}$$

More generally, if $\mathbf{W}$ is a vector field of the form (2.3) then the directional derivative of f along the vector field is

$$D_{\mathbf{W}}f = \frac{\partial f}{\partial x_1}h_1 + \cdots + \frac{\partial f}{\partial x_n}h_n \tag{2.5}$$

where evaluating at $\mathbf{P}$ gives you the directional derivative $D_{\mathbf{W}(\mathbf{P})}f(\mathbf{P})$ of f at $\mathbf{P}$ along $\mathbf{W}(\mathbf{P})$.

Example 2.17 Consider again the function $f = \frac{x^2+y^2}{2}$. Its graph is given in Figure 3. For the vector field $\mathbf{W}(x, y) = x\mathbf{e}_x + y\mathbf{e}_y$, what is the directional derivative of f along $\mathbf{W}$?

$$D_{\mathbf{W}}f = \frac{\partial f}{\partial x}h_1 + \frac{\partial f}{\partial y}h_2 = x \cdot x + y \cdot y = x^2 + y^2,$$

i.e., the slope of f increases quadratically as we get further away from the origin.

For the vector field $\widetilde{\mathbf{W}}(x, y) = -y\mathbf{e}_x + x\mathbf{e}_y$, what is the directional derivative of f along $\widetilde{\mathbf{W}}$?

$$D_{\widetilde{\mathbf{W}}}f = x \cdot (-y) + y \cdot x = 0,$$

i.e., the vector field $\widetilde{\mathbf{W}}$ matches the levels sets/contours of the function f. As one moves along the vector field $\widetilde{\mathbf{W}}$, the value of f does not change.

2.3 Differential 0-forms and 1-forms

2.3.1 Definitions

Differential forms are a very powerful, and yet subtle, approach to multivariable calculus and differential geometry. Their purpose is to unify methods of integrating over curves, surfaces and multi-dimensional objects, as well as providing an approach that is independent of choosing coordinates in the space. Unfortunately, their full power is outside the scope of this course and will have to be covered in future courses. Instead, we intend to give an introduction to 0-forms and 1-forms, with an emphasis on how to compute with them, and leave the full gory details for later.

We shall begin by defining 0-forms and 1-forms. We note that the connection between them shall not be immediately apparent, but will be covered in Subsubsection 2.3.2.

Wikipedia: differential 0-form

Definition 2.18 (0-form) A *differential 0-form* is a smooth real-valued

function $f\colon \mathbb{A}^n \to \mathbb{R}$.

Before introducing differential 1-forms, we consider the following linear maps on the tangent space

$$\mathrm{d}x_i\colon T_{\mathbf{P}}(\mathbb{A}^n) \to \mathbb{R} \tag{2.6}$$

$$\mathbf{v} = \sum_{i=1}^{n} v_i \mathbf{e}_i \mapsto \mathrm{d}x_i(\mathbf{v}) = v_i. \tag{2.7}$$

that maps a vector to its i^{th} entry. We call these *elementary forms*, and these are going to be our building blocks for differential 1-forms.

Wikipedia: elementary forms

Remark 2.19 The classical approach to calculus uses $\mathrm{d}x_i$ to denote an infinitesimal in the x_i coordinate direction. However, the modern viewpoint is that $\mathrm{d}x_i$ should not be an infinitesimal, rather an element of the tangent space in the x_i coordinate direction.

A *linear functional* f on a vector space V is a linear map that maps to $\mathbb{R}$. Note that the elementary forms $\mathrm{d}x_i$ are linear functionals on $T_{\mathbf{P}}(\mathbb{A}^n)$, as are linear combinations of them.

Wikipedia: linear functional

Definition 2.20 (1-form) Let $g_1, \ldots, g_n$ be smooth real-valued functions on $\mathbb{A}^n$. A *differential 1-form* on $\mathbb{A}^n$ is a function

$$\omega = g_1(x_1, \ldots, x_n)\mathrm{d}x_1 + \cdots + g_n(x_1, \ldots, x_n)\mathrm{d}x_n \tag{2.8}$$

that associates to each point $\mathbf{P} \in \mathbb{A}^n$ a linear functional $\omega(\mathbf{P}, -)$ on its tangent space $T_{\mathbf{P}}(\mathbb{A}^n)$

$$\omega(\mathbf{P}, -)\colon T_{\mathbf{P}}(\mathbb{A}^n) \to \mathbb{R}$$
$$\mathbf{v} \mapsto \omega(\mathbf{P}, \mathbf{v}) = g_1(\mathbf{P})\mathrm{d}x_1(\mathbf{v}) + \cdots + g_n(\mathbf{P})\mathrm{d}x_n(\mathbf{v})$$

Equivalently, we can assume each g_i is a 0-form.
Wikipedia: differential 1-form

We shall sometimes denote this linear functional as $\omega_{\mathbf{P}} := \omega(\mathbf{P}, -)$.

Example 2.21 Consider the 1-form $\omega = x\mathrm{d}x + y\mathrm{d}y$ on $\mathbb{A}^2$. Then $g_1(\mathbf{P}) = x$ and $g_2(\mathbf{P}) = y$. At the point $\mathbf{P} = (2, 1)^{\mathsf{T}}$ we get the linear functional

$$\omega_{\mathbf{P}}(\mathbf{v}) = \omega(\mathbf{P}, \mathbf{v}) = 2\mathrm{d}x(\mathbf{v}) + \mathrm{d}y(\mathbf{v}) = 2v_x + v_y$$

for $\mathbf{v} = \left[\begin{smallmatrix} v_x \\ v_y \end{smallmatrix}\right] \in T_{\mathbf{P}}\left(\mathbb{A}^2\right)$.

Remark 2.22 Just by comparing definitions, we see that 1-forms and vector fields look like very similar objects. In fact, there is a correspondence between them via

$$g_1\mathrm{d}x_1 + \cdots + g_n\mathrm{d}x_n \longleftrightarrow g_1\mathbf{e}_1 + \cdots + g_n\mathbf{e}_n.$$

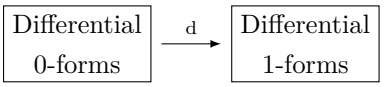
While they look very similar, we reiterate that they are different objects:

- A vector field assigns to every point a vector of $T_{\mathbf{P}}(\mathbb{A}^n)$.
- A 1-form assigns to every point a linear functional on $T_{\mathbf{P}}(\mathbb{A}^n)$.

Vector fields and 1-forms are in fact *dual* objects, similar to the relationship between row and column vectors. Duality is an abstract but important concept that shows up in all area of mathematics.
Wikipedia: Duality in mathematics

2.3.2 Differentiating 0-forms

What is the relationship between differential 0-forms and 1-forms? The answer is that we can "differentiate" 0-forms via a differential map d to get 1-forms.



Definition 2.23 (Differential of a 0-form) Let f be a 0-form on $\mathbb{A}^n$. We define the *differential of f* to be the 1-form

$$\mathrm{d}f = \frac{\partial f}{\partial x_1}\mathrm{d}x_1 + \cdots + \frac{\partial f}{\partial x_n}\mathrm{d}x_n. \tag{2.9}$$

Although we do not cover them in this course, we can define 2-forms to be what we get if we differentiate a 1-form. By extension a k-form is what we get if we differentiate a $(k-1)$-form.

Wikipedia: total derivative

You may have also come across $\mathrm{d}f$ as the *total derivative* of f. From this viewpoint, 1-forms are generalisations of the total derivative, as not all 1-forms need be of the form $\mathrm{d}f$.

Example 2.24 Consider the 0-form on $\mathbb{A}^2$

$$f(x,y) = \frac{x^2+y^2}{2}.$$

What is the differential of f? The differential of f is

$$\begin{aligned}\mathrm{d}f &= \frac{\partial f}{\partial x}\mathrm{d}x + \frac{\partial f}{\partial y}\mathrm{d}y \\ &= x\mathrm{d}x + y\mathrm{d}y.\end{aligned}$$

1-forms that can be obtained as the differential of a 0-form are particularly nice, especially when we begin integrating them.

Wikipedia: exact

Definition 2.25 (Exact form) A 1-form ω is called *exact* if there exists a 0-form f such that $\omega = \mathrm{d}f$.

As an example, note that the 1-form ω from Example 2.21, is equal to the 1-form $\mathrm{d}f$ from Example 2.24. As $\omega = \mathrm{d}f$, it is an exact 1-form.

Not all 1-forms are exact forms, just as not all vector fields are conservative.

Proposition 2.26 *Let $\omega = g_1 \mathrm{d}x_1 + \cdots + g_n \mathrm{d}x_n$ be a 1-form on $\mathbb{A}^n$. If ω is exact then*

$$\frac{\partial g_i}{\partial x_j} = \frac{\partial g_j}{\partial x_i} \quad \forall 1 \leq i, j \leq n. \tag{2.10}$$

Proof. ω is exact if and only if it is of the form $\omega = \mathrm{d}f$ for some 0-form f, *i.e.*, g_i is of the form

$$g_i = \frac{\partial f}{\partial x_i}.$$

If we differentiate g_i with respect to some x_j, we see that

$$\frac{\partial g_i}{\partial x_j} = \frac{\partial}{\partial x_j}\left(\frac{\partial f}{\partial x_i}\right) = \frac{\partial}{\partial x_i}\left(\frac{\partial f}{\partial x_j}\right) = \frac{\partial g_j}{\partial x_i}.$$

□

This proposition is useful for trying to rule out whether a 1-form is exact: if condition (2.10) does not hold then it cannot be exact.

Example 2.27 Consider the 1-form $\omega = \lambda y \mathrm{d}x + x \mathrm{d}y$. When is ω exact? If ω is exact, then

$$\frac{\partial}{\partial y}(\lambda y) = \lambda = 1 = \frac{\partial}{\partial x}(x).$$

Therefore if $\lambda \neq 1$ then ω is not exact. When $\lambda = 1$, we find ω is exact: it is equal to $\mathrm{d}f$ where $f = xy$.

Remark 2.28 Recall that the purpose of differentiating a function is to give a linear approximation of it at a point. With exact forms, we can now shed some light on the geometric intuition behind 1-forms.

Consider again the 0-form $f = \frac{x^2+y^2}{2}$ from Example 2.24. Its differential $\mathrm{d}f = x\mathrm{d}x + y\mathrm{d}y$ assigns to the point $\mathbf{P} = (1,1)^\mathsf{T}$ the linear functional

$$\begin{aligned} \mathrm{d}f(\mathbf{P}, -) \colon T_{\mathbf{P}}(\mathbb{A}^n) &\to \mathbb{R} \\ \mathbf{v} &\mapsto v_x + v_y \end{aligned}$$

on the tangent space at $\mathbf{P}$. Figure 5 shows the level sets of this linear functional on $T_{\mathbf{P}}(\mathbb{A}^n)$. However, we can also consider the level sets of f locally around $\mathbf{P}$. We see that by comparing, $\mathrm{d}f(\mathbf{P}, -)$ gives a linear approximation of f at $\mathbf{P}$. This is essentially the purpose of differentiation.

With this intuition, we can think of an exact 1-form $\mathrm{d}f$ as encoding the linear approximation of f at every point $\mathbf{P}$. For a 1-form ω is not exact, it is also encoding many "linear approximations" but there is no 0-form f that can approximated in this way.

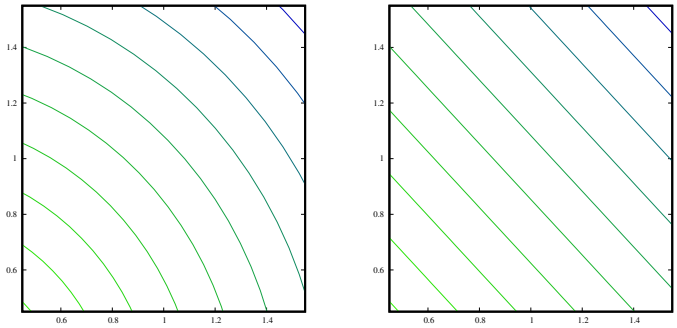


Figure 5: The level sets of the function $f = \frac{x^2+y^2}{2}$ in $\mathbb{A}^2$ on the left, and the level sets of the linear functional $v_x + v_y$ in the tangent space $T_{\mathbf{P}}\left(\mathbb{A}^2\right)$ for $\mathbf{P} = (1,1)^{\mathsf{T}}$ on the right.

2.3.3 Applying 1-forms to vector fields

We saw in the definition of 1-forms that if we fix a point $\mathbf{P}$ and treat the tangent vectors as variables, then $\omega(\mathbf{P}, -)$ is a linear functional. What happens when we fix the tangent vectors and vary the points?

To make this precise, consider a 1-form ω and a vector field $\mathbf{W}$ defined by

$$\begin{aligned}\omega &= g_1(x_1, \ldots, x_n)\mathrm{d}x_1 + \cdots + g_n(x_1, \ldots, x_n)\mathrm{d}x_n,\\ \mathbf{W} &= h_1(x_1, \ldots, x_n)\mathbf{e}_1 + \cdots + h_n(x_1, \ldots, x_n)\mathbf{e}_n.\end{aligned}$$

The vector field fixes a tangent vector $\mathbf{W}(\mathbf{P})$ for each point $\mathbf{P}$. We can "apply" ω to the vector field $\mathbf{W}$ to define a function on $\mathbb{A}^n$

$$\omega(-, \mathbf{W})\colon \mathbb{A}^n \to \mathbb{R} \tag{2.11}$$

$$\mathbf{P} \mapsto \omega(\mathbf{P}, \mathbf{W}(\mathbf{P})) \tag{2.12}$$

Using the definition of a 1-form, we can simplify the expression for $\omega(-, \mathbf{W})$ down to

$$\begin{aligned}\omega(-, \mathbf{W}) &= g_1(x_1, \ldots, x_n)\mathrm{d}x_1(\mathbf{W}) + \cdots + g_n(x_1, \ldots, x_n)\mathrm{d}x_n(\mathbf{W})\\ &= g_1(x_1, \ldots, x_n)h_1(x_1, \ldots, x_n) + \cdots\\ &\quad + g_n(x_1, \ldots, x_n)h_n(x_1, \ldots, x_n),\end{aligned} \tag{2.13}$$

as $\mathrm{d}x_i(\mathbf{W}) = h_i(x_1, \ldots, x_n)$.

Example 2.29 Consider the 1-form $\omega = x\mathrm{d}x + y\mathrm{d}y$ and the vector field $\mathbf{W} = -y\mathbf{e}_x + x\mathbf{e}_y$. Applying equation (2.13), we see that

$$\omega(-, \mathbf{W}) = x \cdot (-y) + y \cdot x = 0.$$

This seems slightly anticlimactic, but we will show there is some nice geometry going on in the background here.

Note that something particularly nice happens when ω is exact. If $\omega = \mathrm{d}f$, then evaluating $\mathrm{d}f(-, \mathbf{W})$ at $\mathbf{P}$ gives

$$\mathrm{d}f(-, \mathbf{W}) = \frac{\partial f}{\partial x_1} h_1 + \cdots + \frac{\partial f}{\partial x_n} h_n = D_{\mathbf{W}} f,$$

the directional derivative of f along the vector field $\mathbf{W}$. From this point of view, 1-forms are generalisations of directional derivatives of a function along a vector field.

Example 2.30 Let us reconsider Example 2.29. Recall that $\omega = \mathrm{d}f$ where $f = \frac{x^2+y^2}{2}$, a function whose level sets are circles in the plane. The graph of f and the vector field $\mathbf{W}$ are shown again in Figure 4.

As ω is exact, the geometric intuition is that $\omega(-, \mathbf{W})$ is the directional derivative of f along $\mathbf{W}$. Note that a point moving around the vector field $\mathbf{W}$ traces out a circle, and on these circles the value of f does not change. Therefore the directional derivative of f along $\mathbf{W}$ is 0, lining up with our value of $\omega(-, \mathbf{W})$.

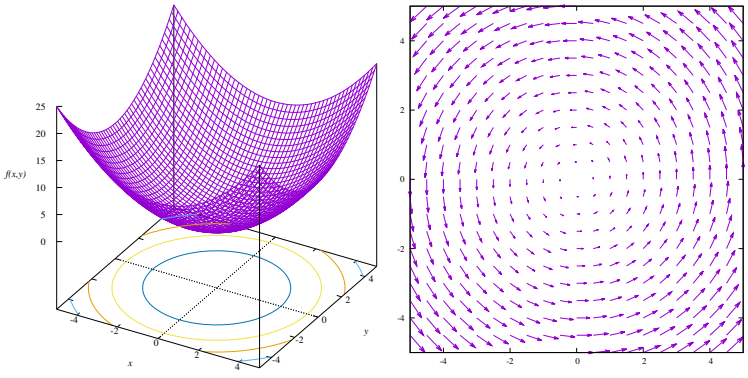


Figure 6: Recalling the 0-form $f = \frac{x^2+y^2}{2}$ and the vector field $\mathbf{W} = -y\mathbf{e}_x + x\mathbf{e}_y$.

Week 7

2.4 Curves in $\mathbb{A}^n$

2.4.1 Definitions

Aside from points, curves are the simplest geometric objects we can work with in affine space. This does not mean they are easy: even the simplest objects have some real complexity to them!

We shall define curves via parametrisation. To do this, recall the following notation for open/closed intervals:

$$(a,b) = \{t \in \mathbb{R} \mid a < t < b\}$$
$$[a,b] = \{t \in \mathbb{R} \mid a \leq t \leq b\}$$
$$[a,b) = \{t \in \mathbb{R} \mid a \leq t < b\}$$

In particular, we note that $(-\infty, +\infty) = \mathbb{R}$.

Definition 2.31 (Curve) Let $I \subseteq \mathbb{R}$ be an interval of the real numbers. A *curve* C in $\mathbb{A}^n$ is the image of a continuous map $\boldsymbol{\gamma} : I \to \mathbb{A}^n$,

Wikipedia: curve

$$C = \{\mathbf{P} \in \mathbb{A}^n \mid \exists\, t \in I \text{ such that } \mathbf{P} = \boldsymbol{\gamma}(t)\}\,. \tag{2.14}$$

The map $\boldsymbol{\gamma}$ is a *parametrisation* for C.

Wikipedia: parametrisation

When working in Cartesian coordinates, it will be convenient to write our parametrisation maps as

$$\boldsymbol{\gamma}(t) = \begin{pmatrix} \gamma_1(t) \\ \vdots \\ \gamma_n(t) \end{pmatrix}$$

where each $\gamma_i : I \to \mathbb{R}$ is a continuous map to the reals.

Example 2.32 Consider the parametrisation map

$$\begin{aligned} \boldsymbol{\gamma} : [0, 2\pi) &\to \mathbb{A}^2 \\ t &\mapsto \begin{pmatrix} a\cos t \\ a \sin t \end{pmatrix}. \end{aligned} \tag{2.15}$$

If we were to compute this parametrisation map, we'd see that the image of γ is the circle of the radius a,

$$C = \left\{ \begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{A}^2 \mid x^2 + y^2 = a^2 \right\}. \tag{2.16}$$

Remark 2.33 The following viewpoint is helpful when considering parametrisations of curves. If we consider the parameter t as "time", we can consider a parametrisation as a point moving along a path in space. The curve is the path that is traced out by this point moving.

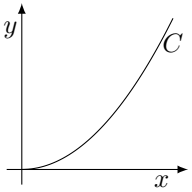
2.4.2 Reparametrisation and orientation

Curves may also be defined as a set of points that satisfy certain conditions or equations. These are sometimes known as *implicit curves*. If we want a parametrisation for these curves, we have to pick it ourselves.

Wikipedia: implicit curves

Example 2.34 Consider the following curve in $\mathbb{A}^2$

$$C = \left\{ \begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{A}^2 \mid y = x^2 \,,\, 0 < x < 1 \right\}.$$



Find a parametrisation for C.

We introduce our time parameter t and choose to set $x = t$. This implies that $y = x^2 = t^2$. Furthermore, as $0 < x < 1$, our parametrisation will be a map from $(0, 1)$. This gives the following parametrisation of C:

$$\begin{aligned} \gamma_1 : (0,1) &\to \mathbb{A}^2 \\ t &\mapsto \begin{pmatrix} t \\ t^2 \end{pmatrix}. \end{aligned}$$

However, this is far from the only parametrisation. We could instead choose to set $y = t$, implying that $x = \sqrt{t}$. As $0 < x = \sqrt{t} < 1$, we have the same interval $(0, 1)$, and so we get a different parametrisation of C:

$$\begin{aligned} \gamma_2 : (0,1) &\to \mathbb{A}^2 \\ t &\mapsto \begin{pmatrix} \sqrt{t} \\ t \end{pmatrix}. \end{aligned}$$

To verify that γ_1 and γ_2 are different parametrisations, consider the

corresponding points they map to at "time" $t = \frac{1}{2}$:

$$\boldsymbol{\gamma}_1\left(\frac{1}{2}\right) = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{4} \end{pmatrix}, \quad \boldsymbol{\gamma}_2\left(\frac{1}{2}\right) = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{2} \end{pmatrix}.$$

We would like a way to see when two parametrisation maps give rise to the same curve. Furthermore, we would like a controlled way of transforming one parametrisation into another. Both of these can be accomplished via a *reparametrisation map*.

Definition 2.35 (Reparametrisation) Consider the parametrisations $\boldsymbol{\gamma}_1 : I_1 \to \mathbb{A}^n$ and $\boldsymbol{\gamma}_2 : I_2 \to \mathbb{A}^n$. We say that $\boldsymbol{\gamma}_2$ is a *reparametrisation* of $\boldsymbol{\gamma}_1$ if there exists a smooth bijective map $\varphi\colon I_1 \to I_2$ such that $\forall\, t \in I_1$:

- $\boldsymbol{\gamma}_1(t) = \boldsymbol{\gamma}_2(\varphi(t))$,
- $\frac{\mathrm{d}\varphi}{\mathrm{d}t} \neq 0$.

We call φ a *reparametrisation map*.

The intuition we should have behind this definition is that φ "deforms" the time interval I_1 into I_2. The first condition states that this deformation makes the two parametrisations equal, implying that they must parametrise the same curve. The second condition states that this deformation cannot "stop time" in one of the intervals, and will have more implications when considering the orientation of curves.

Example 2.36 Recall the two parametrisations $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2$ from Example 2.34. These give rise to the same curve, and seem relatively well behaved, so we expect $\boldsymbol{\gamma}_2$ to be a reparametrisation of $\boldsymbol{\gamma}_1$. Show that $\boldsymbol{\gamma}_2$ is a reparametrisation of $\boldsymbol{\gamma}_1$. It turns out that they are via the reparametrisation map

$$\begin{aligned} \varphi\colon (0,1) &\to (0,1) \\ t &\mapsto t^2 \end{aligned}$$

Note that φ is smooth and bijective, and satisfies the conditions in Definition 2.35:

$$\boldsymbol{\gamma}_2(\varphi(t)) = \begin{pmatrix} \sqrt{\varphi(t)} \\ \varphi(t) \end{pmatrix} = \begin{pmatrix} \sqrt{t^2} \\ t^2 \end{pmatrix} = \begin{pmatrix} t \\ t^2 \end{pmatrix} = \boldsymbol{\gamma}_1(t)$$
$$\frac{\mathrm{d}\varphi}{\mathrm{d}t} = 2t > 0 \quad \forall t \in (0,1)$$

Note that if we extended the domain of $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2$ to $[0,1]$, then this would no longer be a parametrisation map, as $\frac{\mathrm{d}\varphi}{\mathrm{d}t}$ is equal to zero at $t = 0$.

Example 2.37 We can introduce a third parametrisation of C via the reparametrisation map

$$\begin{aligned}\psi\colon \left(0,\frac{\pi}{2}\right) &\to (0,1)\\ t &\mapsto \sin t\end{aligned}$$

We define $\boldsymbol{\gamma}_3$ by deforming the parametrisation $\boldsymbol{\gamma}_1$ via ψ:

$$\begin{aligned}\boldsymbol{\gamma}_3\colon \left(0,\frac{\pi}{2}\right) &\to \mathbb{A}^2\\ t &\mapsto \boldsymbol{\gamma}_1(\psi(t)) = \begin{pmatrix}\sin t\\ \sin^2 t\end{pmatrix}\end{aligned}$$

Note that $\boldsymbol{\gamma}_3$ is clearly a new parametrisation as its domain is a different interval to $\boldsymbol{\gamma}_1$ and $\boldsymbol{\gamma}_2$.

Remark 2.38 In Example 2.37, as the map ψ goes between I_3 to I_1, the reparametrisation process deforms $\boldsymbol{\gamma}_3$ into $\boldsymbol{\gamma}_1$. Therefore we say that $\boldsymbol{\gamma}_1$ is a reparametrisation of $\boldsymbol{\gamma}_3$. This may seem counterintuitive, as we had to define $\boldsymbol{\gamma}_3$ via $\boldsymbol{\gamma}_1$. We are actually doing the following: as we know the behaviour of $\boldsymbol{\gamma}_1$ and as ψ is bijective, we can look at its inverse to see what the behaviour of $\boldsymbol{\gamma}_3$ must be to deform into $\boldsymbol{\gamma}_1$.

If a curve has endpoints $\mathbf{P}, \mathbf{Q}$, a parametrisation can traverse the curve either from $\mathbf{P}$ to $\mathbf{Q}$ or from $\mathbf{Q}$ to $\mathbf{P}$. While these two parametrisations give rise to the same curve, it is helpful to distinguish that they traverse the curve in opposite directions. This gives rise to the notion of the *orientation* of a curve.

Definition 2.39 (Orientation of a curve) Let $\boldsymbol{\gamma}_1\colon I_1 \to \mathbb{A}^n$, $\boldsymbol{\gamma}_2\colon I_2 \to \mathbb{A}^n$ be parametrisations with a reparametrisation map $\varphi\colon I_1 \to I_2$.

- We say $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2$ have the *same orientation* if $\frac{\mathrm{d}\varphi}{\mathrm{d}t} > 0$.
- We say $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2$ have the *opposite orientation* if $\frac{\mathrm{d}\varphi}{\mathrm{d}t} < 0$.

An *orientation* of the curve C is an equivalence class of parametrisations with the same orientation.

Wikipedia: orientation

As with orientation of bases, the parametrisations of a curve with the same orientation form an equivalence class. Picking a parametrisation for a curve fixes its orientation: if we want to reparametrise and preserve orientation then we must remain in this equivalence class.

Example 2.40 The parametrisations $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2$ from Example 2.36 have the same orientation as the reparametrisation map φ has positive derivative for all $t \in (0,1)$. Intuitively this makes sense, as they both begin at the point $\left(\begin{smallmatrix}0\\0\end{smallmatrix}\right)$ and end at the point $\left(\begin{smallmatrix}1\\1\end{smallmatrix}\right)$.

Consider another parametrisation $\boldsymbol{\gamma}_4$ of C defined by the reparametrisation map ξ and deforming in $\boldsymbol{\gamma}_1$:

$$\begin{aligned} \xi\colon (0,1) &\to (0,1) & \boldsymbol{\gamma}_4\colon (0,1) &\to \mathbb{A}^2 \\ t &\mapsto 1-t & t &\mapsto \boldsymbol{\gamma}_1(\xi(t)) = \begin{pmatrix} 1-t \\ (1-t)^2 \end{pmatrix} \end{aligned}$$

Does it have the same or opposite orientation as γ_1? $\boldsymbol{\gamma}_4$ has the opposite orientation to $\boldsymbol{\gamma}_1$, as $\frac{\mathrm{d}\xi}{\mathrm{d}t} = -1$. This matches with the intuitive notion of orientation, as $\boldsymbol{\gamma}_4$ begins at the point $\left(\begin{smallmatrix}1\\1\end{smallmatrix}\right)$ and ends at the point $\left(\begin{smallmatrix}0\\0\end{smallmatrix}\right)$.

2.4.3 Differential properties of curves

To define the correct notion of a tangent vector and tangent space to a curve, we need to go via *velocity vectors*.

Definition 2.41 (Velocity vector of a parametrisation) Let $\boldsymbol{\gamma}\colon I \to \mathbb{A}^n$ be a parametrisation for a curve C. The *velocity vector of* $\boldsymbol{\gamma}$ at the point $\boldsymbol{\gamma}(t_0)$ is

$$\boldsymbol{\gamma}'(t_0) = \begin{bmatrix} \gamma_1'(t_0) \\ \vdots \\ \gamma_n'(t_0) \end{bmatrix}, \quad \gamma_i'(t_0) = \left.\frac{\mathrm{d}\gamma_i}{\mathrm{d}t}\right|_{t=t_0}. \tag{2.17}$$

Remark 2.42 The name velocity vector comes from the idea of a parametrisation $\boldsymbol{\gamma}$ being a point moving through space. The velocity vector $\boldsymbol{\gamma}'(t_0)$ is precisely the velocity of the point at time t_0.

Note that we can consider $\boldsymbol{\gamma}'(t)$ as a vector field on the curve that assigns to every point $\boldsymbol{\gamma}(t_0)$ its velocity vector $\boldsymbol{\gamma}'(t_0)$.

Example 2.43 Recall the parametrisation $\boldsymbol{\gamma}$ of the circle with radius a from Example 2.32,

$$\begin{aligned} \boldsymbol{\gamma} : [0, 2\pi) &\to \mathbb{A}^2 \\ t &\mapsto \begin{pmatrix} a\cos t \\ a\sin t \end{pmatrix}. \end{aligned} \tag{2.18}$$

What is the velocity vector of $\boldsymbol{\gamma}$ at time t?

The velocity vector of $\boldsymbol{\gamma}$ at time t is

$$\boldsymbol{\gamma}'(t) = \begin{bmatrix} \frac{\mathrm{d}}{\mathrm{d}t}(a\cos t) \\ \frac{\mathrm{d}}{\mathrm{d}t}(a\sin t) \end{bmatrix} = \begin{bmatrix} -a\sin t \\ a\cos t \end{bmatrix}.$$

As a vector field, $\boldsymbol{\gamma}'(t)$ assigns to every point a vector "tangent" to the circle with magnitude a, see Figure 7.

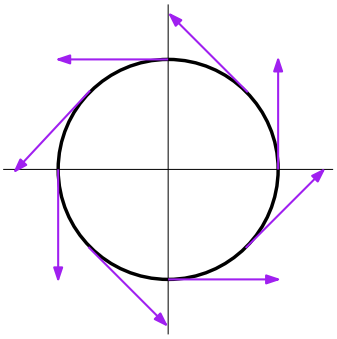


Figure 7: The velocity vectors of the parametrisation from Example 2.43.

A velocity vector is always tangent to the curve at that point. Suppose we reparametrise C, what happens to the velocity vector at a point? The magnitude of the velocity vector may change, but the direction (up to sign) will stay the same.

Definition 2.44 (Tangent vector and space to a curve) A *tangent vector* to C at $\mathbf{P}$ is a velocity vector $\boldsymbol{\gamma}'(t_0)$ such that $\boldsymbol{\gamma}$ is a parametrisation of C and $\mathbf{P} = \boldsymbol{\gamma}(t_0)$.

The *tangent space* to C at $\mathbf{P}$ is the set $T_{\mathbf{P}}(C)$ of tangent vectors to C at $\mathbf{P}$.

Note that $T_{\mathbf{P}}(C) \subset T_{\mathbf{P}}(\mathbb{A}^n)$: the tangent space to a curve is a subspace of the tangent space to $\mathbb{A}^n$. We shall always consider velocity vectors in the tangent space to the curve it parametrises.

Example 2.45 Let C be the circle of radius a and the point $\mathbf{P} = (-a, 0)^\mathsf{T}$ on it. Reconsider the parametrisation $\boldsymbol{\gamma}$ of C given in equation (2.18). Is its velocity vector a tangent vector?

We have $\mathbf{P} = \boldsymbol{\gamma}(t)$ at time $t = \pi$. Therefore, its velocity vector

$$\boldsymbol{\gamma}'(\pi) = \begin{bmatrix} -a\sin\pi \\ a\cos\pi \end{bmatrix} = \begin{bmatrix} 0 \\ -a \end{bmatrix} \in T_{\mathbf{P}}(C)$$

is a tangent vector to C at $\mathbf{P}$.

We consider a different parametrisation $\widetilde{\boldsymbol{\gamma}}$ of C via the reparametrisation map φ:

$$\varphi\colon \left[0, \frac{2\pi}{k}\right) \to [0, 2\pi) \qquad \widetilde{\boldsymbol{\gamma}}\colon \left[0, \frac{2\pi}{k}\right) \to \mathbb{A}^n$$

$$t \mapsto kt \qquad\qquad t \mapsto \boldsymbol{\gamma}(\varphi(t)) = \begin{pmatrix} a\cos(kt) \\ a\sin(kt) \end{pmatrix}$$

where $k \in \mathbb{R} \setminus 0$. Note if $k < 0$, then $\frac{2\pi}{k} < 0$ and so we rewrite the interval as $\left(\frac{2\pi}{k}, 0\right]$.

Is $\widetilde{\boldsymbol{\gamma}}$ a tangent vector? We have $\mathbf{P} = \widetilde{\boldsymbol{\gamma}}(t)$ at time $t = \frac{\pi}{k}$, and therefore we have

$$\widetilde{\boldsymbol{\gamma}}'(\pi) = \begin{bmatrix} -ka\sin\pi \\ ka\cos\pi \end{bmatrix} = \begin{bmatrix} 0 \\ -ka \end{bmatrix} \in T_{\mathbf{P}}(C)$$

is a tangent vector to C at $\mathbf{P}$. As k can be any nonzero real number, the tangent space to C at $\mathbf{P}$ is the 1-dimensional vector space span $\left(\begin{bmatrix} 0 \\ -1 \end{bmatrix}\right)$.

We can't realise the zero vector as via the reparametrisation map φ as it is not defined for $k = 0$. However we can still find a parametrisation of C which has zero velocity at $\mathbf{P}$.

While curves are some of the simplest geometric objects, they can still get quite horrible to work with if we are not careful when picking parametrisations.

Definition 2.46 (Smooth parametrisations and curves) Let $\boldsymbol{\gamma}\colon I \to \mathbb{A}^n$ be a parametrisation for a curve. We say $\boldsymbol{\gamma}$ is *smooth* if every $\gamma_i\colon I \to \mathbb{R}$ is smooth, *i.e.*, $\frac{\mathrm{d}^k\gamma_i}{\mathrm{d}t^k}$ is well defined for all positive integers k and for all $t \in I$.

A curve C is smooth if it has a smooth parametrisation.

Example 2.47 Consider the curve C from Example 2.34, but with the endpoints included. Is the parametrisation of C

$$\begin{aligned} \boldsymbol{\gamma}_1\colon [0,1] &\to \mathbb{A}^2 \\ t &\mapsto \begin{pmatrix} t \\ t^2 \end{pmatrix} \end{aligned}$$

smooth? is smooth as all derivatives of t and t^2 are well defined on $[0,1]$. Therefore C is a smooth curve.

Is there (another) parametrisation of C that is not smooth?

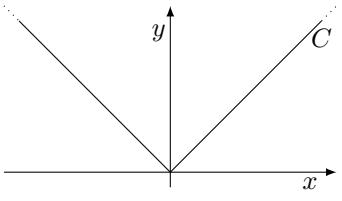
The parametrisation

$$\begin{aligned} \boldsymbol{\gamma}_2\colon [0,1] &\to \mathbb{A}^2 \\ t &\mapsto \begin{pmatrix} \sqrt{t} \\ t \end{pmatrix} \end{aligned}$$

is not smooth as $\frac{\mathrm{d}}{\mathrm{d}t}\left(\sqrt{t}\right) = \frac{1}{2\sqrt{t}}$ is not defined at $t = 0$.

Example 2.48 Consider the curve

$$C = \left\{ \begin{pmatrix} x \\ y \end{pmatrix} \mid y = |x| \right\}.$$



Is this curve smooth?

This curve is not smooth as the derivative of y with respect to x is not defined for $x = 0$:

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \begin{cases} 1 & x > 0 \\ -1 & x < 0 \\ \text{undefined} & x = 0 \end{cases}$$

Therefore for any parametrisation $\boldsymbol{\gamma} = \begin{pmatrix} \gamma_x \\ \gamma_y \end{pmatrix}$ of C, we have

$$\gamma_y'(t) = \frac{\mathrm{d}y}{\mathrm{d}t} = \frac{\mathrm{d}y}{\mathrm{d}x}\frac{\mathrm{d}x}{\mathrm{d}t}$$

is also undefined.

Definition 2.49 (Regular parametrisation) A parametrisation $\boldsymbol{\gamma}$ is *regular* if $\boldsymbol{\gamma}'(t) \neq \mathbf{0}$ for all $t \in I$.

Note that all curves we will consider will always have a regular parametrisation.

To define a curve with no regular parametrisation is difficult and all known examples are not natural!

Example 2.50 Consider the parametrisation

$$\boldsymbol{\gamma}\colon [0, \pi] \to \mathbb{A}^2, \quad t \mapsto \begin{pmatrix} \sin t \\ \sin^2 t \end{pmatrix}, \quad \boldsymbol{\gamma}'(t) = \begin{bmatrix} \cos t \\ 2 \sin t \cos t \end{bmatrix}.$$

This is a parametrisation for the curve C from Example 2.34, but with the endpoints $\left(\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right)$ and $\left(\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right)$ included. Is this parametrisation regular?

This parametrisation is not regular as the velocity vector $\boldsymbol{\gamma}'(\frac{\pi}{2}) = \mathbf{0}$ is zero at $t = \frac{\pi}{2}$. The geometric intuition is that the parametrisation traces C from $\left(\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right)$ to $\left(\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right)$, but then changes direction and traces back to $\left(\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right)$. As a result, we cannot define the orientation of a non-regular parametrisation.

Finally, we would like to know how a 0-form f changes as we follow a curve.

Definition 2.51 (Directional derivative along a curve) Let $C \subset \mathbb{A}^n$ be a curve with parametrisation $\boldsymbol{\gamma}$. Let $f : \mathbb{A}^n \to \mathbb{R}$ a 0-form, the *directional derivative of f along $\boldsymbol{\gamma}$* is

Wikipedia: directional derivative of f along $\boldsymbol{\gamma}$

$$D_{\boldsymbol{\gamma}'} f = \frac{\mathrm{d}}{\mathrm{d}t}(f(\boldsymbol{\gamma}(t))) \tag{2.19}$$

Recall that $\boldsymbol{\gamma}'$ can be considered a vector field on C that assigns to every point its velocity vector. The notation $D_{\boldsymbol{\gamma}'} f$ hints that this should be the directional derivative of f along the vector field $\boldsymbol{\gamma}'$. As $x_i = \gamma_i(t)$,

we can use the chain rule to show this is true:

$$\begin{aligned} D_{\boldsymbol{\gamma}'} f &= \frac{\mathrm{d}}{\mathrm{d}t}(f(\boldsymbol{\gamma}(t))) \\ &= \sum_{i=1}^{n} \frac{\mathrm{d}\gamma_i}{\mathrm{d}t} \frac{\partial f}{\partial x_i} \\ &= \gamma_1'(t) \frac{\partial f}{\partial x_1} + \cdots + \gamma_n'(t) \frac{\partial f}{\partial x_n} \end{aligned} \tag{2.20}$$

and so $D_{\boldsymbol{\gamma}} f$ is just the directional derivative of f in the direction of the velocity vector.

Remark 2.52 In general, the directional derivative along a curve is sensitive to which parametrisation we pick. If we pick a parametrisation that traverses the curve much "faster", the magnitude of the velocity vectors will be greater, therefore the directional derivative will be greater.

Week 8

2.5 Integration of 1-forms over curves

In Subsubsection 2.3.2, we showed how 1-forms could arise as the differential of a 0-form. Intuitively, we should also have an inverse operation that allows us to integrate a 1-form. This is precisely what we consider in this section.

We shall only consider integration of 1-forms over curves. The reason for this is a 1-form ω requires a point $\mathbf{P}$ and a tangent vector $\mathbf{v}$ to give back a real number $\omega(\mathbf{P}, \mathbf{v})$. Curves with a parametrisation have all this information, as each point has a velocity vector already associated to it.

Let $\omega = g_1 \mathrm{d}x_1 + \cdots + g_n \mathrm{d}x_n$ be a 1-form and $\boldsymbol{\gamma} : [a, b] \to \mathbb{A}^n$ a parametrisation of a curve C. We can apply ω to the parametrisation $\boldsymbol{\gamma}$ by evaluating at the point $\boldsymbol{\gamma}(t)$ with velocity vector $\boldsymbol{\gamma}'(t)$:

$$\begin{aligned}\omega(\boldsymbol{\gamma}(t), \boldsymbol{\gamma}'(t)) &= \sum_{i=1}^{n} g_i(\boldsymbol{\gamma}(t)) \mathrm{d}x_i(\boldsymbol{\gamma}'(t)) \\ &= \sum_{i=1}^{n} g_i(\boldsymbol{\gamma}(t)) \boldsymbol{\gamma}_i'(t).\end{aligned}$$

Note that if $\omega = \mathrm{d}f$ is exact, this is the directional derivative of f at $\boldsymbol{\gamma}(t)$ in the direction of the velocity vector $\boldsymbol{\gamma}'(t)$.

Definition 2.53 (Integration of 1-form over curves) Let $\omega = g_1 \mathrm{d}x_1 + \cdots + g_n \mathrm{d}x_n$ be a 1-form and $\boldsymbol{\gamma} \colon [a, b] \to \mathbb{A}^n$ a smooth, regular parametrisation of C. The *integral of ω over C* is

Wikipedia: integral of ω over C

$$\begin{aligned}\int_C \omega &= \int_a^b \omega(\boldsymbol{\gamma}(t), \boldsymbol{\gamma}'(t)) \mathrm{d}t \\ &= \int_a^b \left(\sum_{i=1}^{n} g_i(\gamma_1(t), \ldots, \gamma_n(t)) \frac{\mathrm{d}\gamma_i}{\mathrm{d}t} \right) \mathrm{d}t \end{aligned} \tag{2.21}$$

Remark 2.54 The same definition holds if our curve is defined over an open interval (a, b), as removing a finite number of points from the curve does not impact the integral.

Example 2.55 Consider the 1-form $\omega = x\mathrm{d}x + y\mathrm{d}y$ and the curve

$$C = \left\{ \begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{A}^2 \mid y = x^2\,,\ 0 < x < 1 \right\}.$$

Compute the integral $\int_C \omega$.

We have computed plenty of parametrisations of this curve, so we consider the parametrisation

$$\begin{aligned} \gamma_1 : (0,1) &\to \mathbb{A}^2 \\ t &\mapsto \begin{pmatrix} t \\ t^2 \end{pmatrix}, \quad \gamma_1'(t) = \begin{bmatrix} 1 \\ 2t \end{bmatrix}. \end{aligned}$$

We feed this into equation (2.21) to obtain

$$\int_C x\mathrm{d}x + y\mathrm{d}y = \int_0^1 t \cdot 1 + t^2 \cdot 2t\,\mathrm{d}t = \int_0^1 t + 2t^3 \mathrm{d}t = \left[\frac{t^2}{2} + \frac{t^4}{2}\right]_0^1 = 1$$

We have computed lots of parametrisations of C, so a natural question is how does this integral change if we change the parametrisation? The following theorem shows that if we fix the orientation of C, the value of the integral does not change.

Theorem 2.56 *Let C be an oriented curve. The integral $\int_C \omega$ does not depend on the parametrisation of C.*

Let C' be the curve C with the opposite orientation. Then

$$\int_{C'} \omega = -\int_C \omega$$

Example 2.57 The first part of Theorem 2.56 states the value of $\int_C \omega$ does not change if we reparametrise C while preserving orientation. We shall verify this for the curve C and the 1-form ω from Example 2.55.

We consider another parametrisation of C,

$$\begin{aligned} \gamma_3 \colon \left(0, \frac{\pi}{2}\right) &\to \mathbb{A}^2 \\ t &\mapsto \begin{pmatrix} \sin t \\ \sin^2 t \end{pmatrix} \quad , \quad \gamma_3'(t) = \begin{bmatrix} \cos t \\ 2\sin t \cos t \end{bmatrix}. \end{aligned}$$

Do we get the same value of $\int_C \omega$? We have computed previously that γ_3 has the same orientation as γ_1. By Theorem 2.56, we expect that

$\int_C \omega$ will be equal to 1 if we calculate it via γ_3.

$$\begin{aligned}\int_C x\mathrm{d}x + y\mathrm{d}y &= \int_0^{\frac{\pi}{2}} \sin t \cdot \cos t + \sin^2 t \cdot 2\sin t \cos t \,\mathrm{d}t \\ &= \int_0^{\frac{\pi}{2}} \sin t \cos t + (1-\cos^2 t)\cdot 2\sin t\cos t \,\mathrm{d}t \\ &= \int_0^{\frac{\pi}{2}} 3\sin t\cos t - 2\sin t\cos^3 t \,\mathrm{d}t \\ &= \left[\frac{3}{2}\sin^2 t + \frac{1}{2}\cos^4 t\right]_0^{\frac{\pi}{2}} = 1\end{aligned}$$

We do get the same outcome with this different parametrisation. However, the integral is a bit trickier to compute: picking the correct parametrisation can make our lives much easier!

Example 2.58 The second part of Theorem 2.56 states that if we consider the integral C with the opposite orientation, we just need to flip the sign on the value of the integral. Again, we shall verify this for the curve C and the 1-form ω from Example 2.55.

Consider another parametrisation of C:

$$\begin{aligned}\gamma_4 \colon (0,1) &\to \mathbb{A}^2 \\ t &\mapsto \begin{pmatrix} 1-t \\ (1-t)^2 \end{pmatrix}, \quad \gamma_4'(t) = \begin{bmatrix} -1 \\ 2t-2 \end{bmatrix}.\end{aligned}$$

Do we get the opposite value of $\int_C \omega$? We have previously computed that γ_4 has opposite orientation to γ_1. By Theorem 2.56, we expect that $\int_C \omega$ will be equal to -1 if we calculate it via γ_4.

$$\begin{aligned}\int_C x\mathrm{d}x + y\mathrm{d}y &= \int_0^1 (1-t)\cdot(-1) + (1-t)^2\cdot(2t-2)\,\mathrm{d}t \\ &= \int_0^1 2t^3 - 6t^2 + 7t - 3\,\mathrm{d}t \\ &= \left[\frac{t^4}{2} - 2t^3 + \frac{7t^2}{2} - 3t\right]_0^1 = -1\end{aligned}$$

The full calculation verifies this: changing the orientation of C simply changes the sign.

2.6 Integrating exact 1-forms

When we defined exact 1-forms, we noted that they would be particularly nice when it came to integrating them. The following theorem is the reason for this:

Theorem 2.59 *Let $\omega = \mathrm{d}f$ be an exact 1-form and C a curve in $\mathbb{A}^n$. The integral of ω over C depends only on the value of f at the endpoints*

of C.

Explicitly, for a parametrisation $\boldsymbol{\gamma}\colon [a,b] \to \mathbb{A}^n$ of C, the integral of ω over C

$$\int_C \omega = f(\boldsymbol{\gamma}(b)) - f(\boldsymbol{\gamma}(a)). \tag{2.22}$$

Proof. Recall that is $\omega = \mathrm{d}f$ is exact, then $\mathrm{d}f(\mathbf{P}, \mathbf{v})$ is just the directional derivative of f at $\mathbf{P}$ along $\mathbf{v}$. In particular, $\mathrm{d}f(\boldsymbol{\gamma}(t), \boldsymbol{\gamma}'(t))$ is the directional derivative along f along $\boldsymbol{\gamma}$. Therefore

$$\begin{aligned}\int_C \mathrm{d}f &= \int_a^b \mathrm{d}f(\boldsymbol{\gamma}(t), \boldsymbol{\gamma}'(t))\mathrm{d}t = \int_a^b D_{\boldsymbol{\gamma}'} f \mathrm{d}t = \int_a^b \frac{\mathrm{d}}{\mathrm{d}t}(f(\boldsymbol{\gamma}(t)))\mathrm{d}t \\ &= \int_a^b f(\boldsymbol{\gamma}(t)) = [f(\boldsymbol{\gamma}(t))]_a^b = f(\boldsymbol{\gamma}(b)) - f(\boldsymbol{\gamma}(a)).\end{aligned}$$

□

Example 2.60 If we reconsider the 1-form from Example 2.55, we have already seen that this is an exact 1-form whose corresponding 0-form is $f = \frac{x^2+y^2}{2}$. Considering the curve C from the same example, what is $\int_C \omega$? Its endpoints are $\binom{0}{0}$ and $\binom{1}{1}$, therefore by Theorem 2.59 the integral of $\omega = \mathrm{d}f$ over C is

$$\int_C \omega = f\left(\binom{1}{1}\right) - f\left(\binom{0}{0}\right) = \frac{1+1}{2} + \frac{0+0}{2} = 1.$$

Clearly this makes exact 1-forms far easier to compute with!

Remark 2.61 As the integral of exact 1-forms depends only on the endpoints, we do not care what path the curve takes between those endpoints, it will have no impact on the value of the integral.

Wikipedia: closed

We call a curve *closed* if its endpoints are the same point. For example, circles and ellipses are both closed curves, as are any deformations of them. Exact forms are even easier to integrate over closed curves.

Corollary 2.62 *If C is a closed curve and ω an exact 1-form, then $\int_C \omega = 0$.*

Proof. As C is closed, its endpoints $\boldsymbol{\gamma}(a), \boldsymbol{\gamma}(b)$ are equal and so $f(\boldsymbol{\gamma}(a)) = f(\boldsymbol{\gamma}(b))$ for any 0-form f. Using this and Theorem 2.59, we have

$$\int_C \omega = f(\boldsymbol{\gamma}(b)) - f(\boldsymbol{\gamma}(a)) = 0.$$

□

3 Conic sections

In this section we consider very important curves that you have no doubt encountered before: ellipses; hyperbolas; and parabolas. We will begin with the geometric definitions of these curves, we will follow this with the algebraic definitions in terms of both Cartesian and polar coordinates. These curves can all be realised as sections of a cone and are therefore collectively known by the name *conic sections*, see Figure 17. We will revisit conic sections again in Section 5 when we review projective geometry, giving a more modern perspective on these curves.

3.1 Geometric definitions

Geometrically we define an ellipse or hyperbola, as the locus of points that satisfy some geometric condition with respect to a pair of points called the *foci* (or sometimes *focuses*) of the curve. Similarly, we define a parabola as the locus of points that satisfy a geometric condition with respect to a single focus and a line called the *directrix*.

The following geometric definitions take place in affine space. Recall that for a point $\mathbf{P}$ in $\mathbb{A}^n$ and a vector $\mathbf{v} \in \mathbb{E}^n$ we can add the vector to the point to get a new point in affine space: $\mathbf{P} + \mathbf{v} \in \mathbb{A}^n$. In fact, for each pair of points $\mathbf{P}, \mathbf{Q} \in \mathbb{A}^n$ there is a unique vector $\mathbf{v} \in \mathbb{E}^n$ such that $\mathbf{P} + \mathbf{v} = \mathbf{Q}$. Thus, although there is no addition of points in affine space, there is a concept of subtraction. The notation $\mathbf{Q} - \mathbf{P}$, is really a shorthand meaning: "the vector $\mathbf{v}$ for which $\mathbf{P} + \mathbf{v} = \mathbf{Q}$".

The definitions that follow involve distances between points in affine space. As there is a unique vector from $\mathbf{P}$ to $\mathbf{Q}$, we use this vector to define the distance between points. Using the idea of subtraction of points above, we can denote the distance between $\mathbf{P}$ and $\mathbf{Q}$ as $\|\mathbf{Q} - \mathbf{P}\| = \|\mathbf{P} - \mathbf{Q}\|$.

Definition 3.1 (Ellipse in the affine plane, see Figure 8) Let $\mathbf{F}_1$ and $\mathbf{F}_2$ be two points, called *foci*, in the affine plane $\mathbb{A}^2$ and let $c \in \mathbb{R}$ be half the distance between the two foci: $\|\mathbf{F}_2 - \mathbf{F}_1\| = 2c$.

Wikipedia: foci

Then for each constant $a > c \geq 0$ we define the *ellipse* with foci $\mathbf{F}_1$, $\mathbf{F}_2$ and with distance $2a$ to be the set of points $\mathbf{P}$, such that the sum of the distances of $\mathbf{P}$ to each focus is $2a$:

Wikipedia: ellipse

$$\text{Ellipse} = \left\{ \mathbf{P} \in \mathbb{A}^2 \;\middle|\; \|\mathbf{F}_1 - \mathbf{P}\| + \|\mathbf{F}_2 - \mathbf{P}\| = 2a \right\}. \tag{3.1}$$

If $\mathbf{F}_1$ and $\mathbf{F}_2$ are the same point then this is simply the circle with centre $\mathbf{F}_1$ and radius a.

Remark 3.2 Notice that if we allowed $a = c$, then this definition would degenerate to the line segment from $\mathbf{F}_1$ and $\mathbf{F}_2$; whilst with $a < c$, no

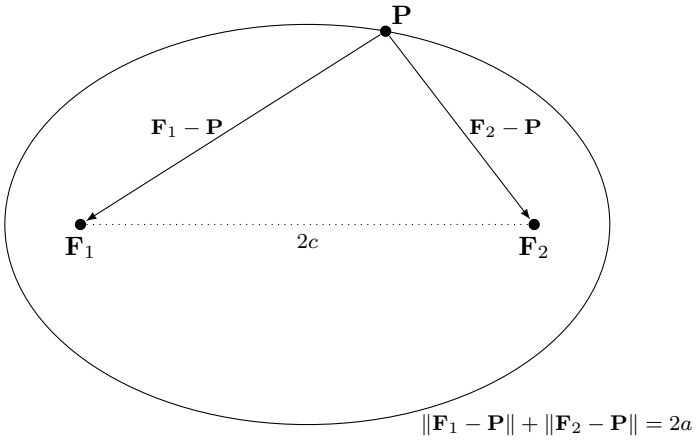


Figure 8: An ellipse with foci $\mathbf{F}_1$ and $\mathbf{F}_2$ and distance $2a$.

points would satisfy the condition.

Definition 3.3 (Hyperbola in the affine plane, see Figure 9) Let $\mathbf{F}_1$ and $\mathbf{F}_2$ be two points in the affine plane $\mathbb{A}^2$, and let $c \in \mathbb{R}$ be half the distance between the two foci.

Wikipedia: hyperbola

Then for each constant a, with $0 < a < c$, we define the *hyperbola* with foci $\mathbf{F}_1$, $\mathbf{F}_2$ and with distance $2a$ to be the set of points $\mathbf{P}$, such that the absolute difference between the distances of $\mathbf{P}$ to each focus is $2a$:

$$\text{Hyperbola} = \left\{ \mathbf{P} \in \mathbb{A}^2 \mid \|\mathbf{F}_1 - \mathbf{P}\| - \|\mathbf{F}_2 - \mathbf{P}\| = \pm 2a \right\}. \tag{3.2}$$

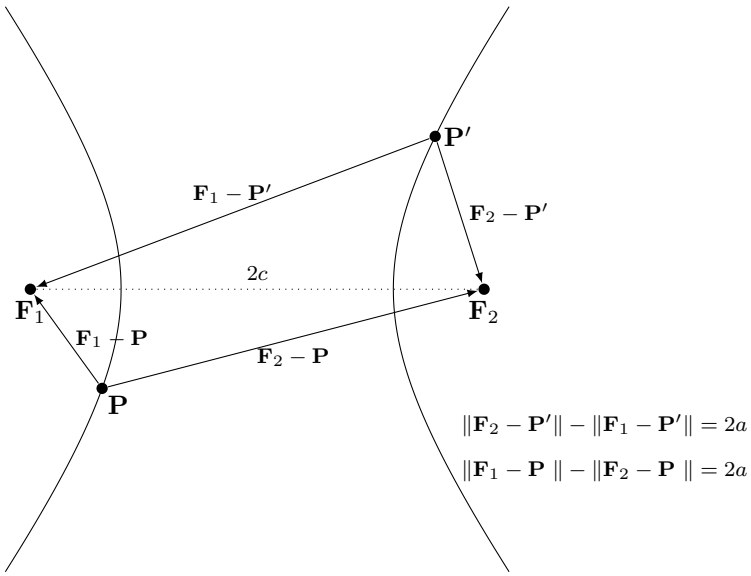


Figure 9: A hyperbola with foci $\mathbf{F}_1$ and $\mathbf{F}_2$ and distance $2a$.

Remark 3.4 Notice that if we allowed $a = c$ then this would degenerate to the union of two half-lines: the first beginning at $\mathbf{F}_1$ and extending infinitely away from $\mathbf{F}_2$; the second beginning at $\mathbf{F}_2$ and extending infinitely away from $\mathbf{F}_1$. On the other hand, if we allowed $a = 0$ then this would degenerate to the perpendicular bisector of the line from $\mathbf{F}_1$ to $\mathbf{F}_2$.

Definition 3.5 (Parabola in the affine plane, see Figure 10) Let $\mathbf{F}$ be a point, called a focus, in the affine plane $\mathbb{A}^2$, and let l be a line, called the *directrix*.

Then we define the *parabola* with focus $\mathbf{F}$ and directrix l to be set of points $\mathbf{P}$, such that the distance between $\mathbf{P}$ and $\mathbf{F}$ is equal to the distance between $\mathbf{P}$ and l. Denoting the closest point on the line l to the point $\mathbf{P}$ by $l_{\mathbf{P}}$ we have

Wikipedia: parabola

$$\text{Parabola} = \left\{ \mathbf{P} \in \mathbb{A}^2 \;\middle|\; \|\mathbf{F} - \mathbf{P}\| = \|l_{\mathbf{P}} - \mathbf{P}\| \right\}. \tag{3.3}$$

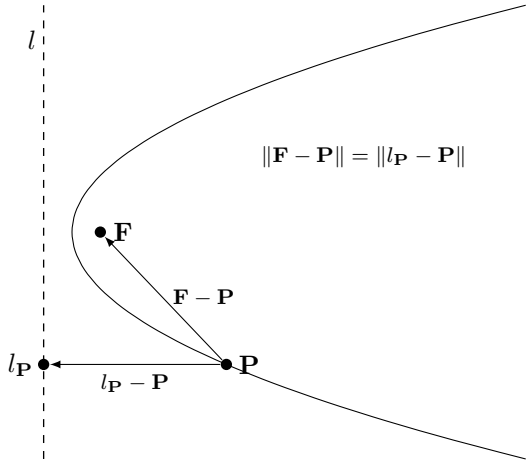


Figure 10: A parabola with focus $\mathbf{F}$ and directrix l.

3.2 Algebraic definitions

In this section we will give algebraic expressions for ellipses, hyperbolas and parabolas. In order to give these expressions we must use a coordinate system for affine space; initially we will restrict ourselves to Cartesian coordinates. Recall from Example 2.4 that we can specify a Cartesian coordinate system by selecting a point $\mathbf{O} \in \mathbb{A}^n$, called the origin, and selecting an orthonormal basis $(\mathbf{e}_1, \ldots, \mathbf{e}_n)$ for $\mathbb{E}^n$.

3.2.1 Expression for an ellipse

We shall first derive an equation for the ellipse. Let C be an ellipse with foci $\mathbf{F}_1, \mathbf{F}_2 \in \mathbb{A}^2$ and distance $2a \in \mathbb{R}$. We first need to define a Cartesian coordinate system: $\mathbf{O}, (\mathbf{e}_x, \mathbf{e}_y)$ (see Figure 11)

- Let the origin be the point half way between $\mathbf{F}_1$ and $\mathbf{F}_2$:

$$\mathbf{0} = \mathbf{F}_1 + \frac{\mathbf{F}_2 - \mathbf{F}_1}{2}.$$

- Let the first basis vector be a unit vector in the same direction as the vector from $\mathbf{F}_1$ to $\mathbf{F}_2$:

$$\mathbf{e}_x = \frac{\mathbf{F}_2 - \mathbf{F}_1}{\|\mathbf{F}_2 - \mathbf{F}_1\|}.$$

- Let $\mathbf{e}_y$ be either of the unit vectors orthogonal to $\mathbf{e}_x$.

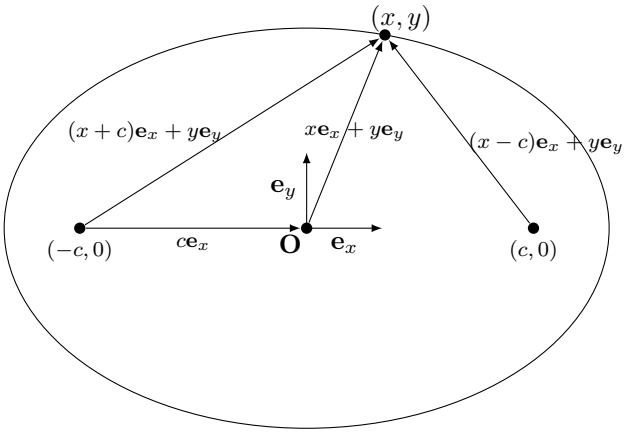


Figure 11: An ellipse shown against a Cartesian coordinate system.

Letting $2c$ be the distance between the two foci, we can express the $\mathbf{F}_1$ and $\mathbf{F}_2$ using this coordinate system:

$$\mathbf{F}_1 = (-c, 0) \qquad \mathbf{F}_2 = (c, 0).$$

Let $\mathbf{P} = (x, y)$ be a point on the ellipse. Then by definition

$$\begin{aligned} 2a &= \|\mathbf{P} - \mathbf{F}_1\| + \|\mathbf{P} - \mathbf{F}_2\| \\ &= \|(x + c)\mathbf{e}_x + y\mathbf{e}_y\| + \|x - c\mathbf{e}_x + y\mathbf{e}_y\| \\ &= \sqrt{(x + c)^2 + y^2} + \sqrt{(x - c)^2 + y^2} \end{aligned}$$

Rearranging slightly and taking squares we have

$$\begin{aligned}(x+c)^2+y^2 &= \left(2a-\sqrt{(x-c)^2+y^2}\right)^2\\ x^2+2xc+c^2+y^2 &= 4a^2-4a\sqrt{(x-c)^2+y^2}+x^2-2xc+c^2+y^2\\ 2xc &= 4a^2-4a\sqrt{(x-c)^2+y^2}-2xc\end{aligned}$$

Rearranging and taking the square a second time leads to

$$\begin{aligned}4a\sqrt{(x-c)^2+y^2} &= 4a^2-4xc\\ a\sqrt{(x-c)^2+y^2} &= a^2-xc\\ \left(a\sqrt{(x-c)^2+y^2}\right)^2 &= \left(a^2-xc\right)^2\\ a^2\left(x^2-2xc+c^2+y^2\right) &= a^4-2a^2xc+x^2c^2\\ a^2x^2-2a^2xc+a^2c^2+a^2y^2 &= a^4-2a^2xc+x^2c^2\\ a^2x^2+a^2c^2+a^2y^2 &= a^4+x^2c^2\end{aligned}$$

Letting $b^2=a^2-c^2$ (which we can do since $a>c$), we can rearrange the equation above to find

$$\begin{aligned}(a^2-c^2)x^2+a^2y^2 &= a^4-a^2c^2,\\ b^2x^2+a^2y^2 &= a^2b^2\\ \frac{x^2}{a^2}+\frac{y^2}{b^2} &= 1.\end{aligned}$$

We have proved that every point on the ellipse satisfies the equation

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1. \tag{3.4}$$

We now show the converse: that any point satisfying equation (3.4) lies on the ellipse.

Let $\mathbf{Q}$ be the point with coordinates (x,y) and assume that these values satisfy the identity of equation (3.4). Then

$$y^2=b^2\left(1-\frac{x^2}{a^2}\right).$$

We wish to calculate the distances between $\mathbf{Q}$ and the foci $\mathbf{F}_1=(-c,0)$

and $\mathbf{F}_2 = (c, 0)$. Recall that $a^2 = b^2 + c^2$.

$$\begin{aligned}
\|\mathbf{Q} - \mathbf{F}_1\| &= \sqrt{(x+c)^2 + y^2} \\
&= \sqrt{(x+c)^2 + b^2\left(1 - \frac{x^2}{a^2}\right)} \\
&= \sqrt{x^2 + 2cx + c^2 + b^2 - \frac{b^2x^2}{a^2}} \\
&= \sqrt{\underbrace{\left(1 - \frac{b^2}{a^2}\right)}_{c^2/a^2} x^2 + 2cx + \underbrace{c^2 + b^2}_{a^2}} \\
&= \sqrt{\frac{c^2x^2}{a^2} + 2cx + a^2} \\
&= \sqrt{\left(\frac{c}{a}x + a\right)^2} \\
&= \left|\frac{c}{a}x + a\right|.
\end{aligned}$$

Now since $0 \leq c < a$ and $|x| < a$ we see that

$$\|\mathbf{Q} - \mathbf{F}_1\| = \frac{c}{a}x + a. \tag{3.5}$$

A similar process for the distance to $\mathbf{F}_2$ shows that

$$\begin{aligned}
\|\mathbf{Q} - \mathbf{F}_2\| &= \sqrt{(x-c)^2 + y^2} \\
&= \sqrt{\frac{c^2x^2}{a^2} - 2cx + a^2} \\
&= \left|\frac{c}{a}x - a\right| \\
&= a - \frac{c}{a}x.
\end{aligned}$$

Putting this together we see that

$$\|\mathbf{Q} - \mathbf{F}_1\| + \|\mathbf{Q} - \mathbf{F}_2\| = \left(\frac{c}{a}x + a\right) + \left(a - \frac{c}{a}x\right) = 2a.$$

Using the analysis above we come to the algebraic definition of an ellipse.

Proposition 3.6 (Equation of an ellipse) *Let C be a curve in $\mathbb{A}^2$. The curve C is an ellipse if and only if there exists a Cartesian coordinate system (x, y) of $\mathbb{A}^2$ and real numbers $a \geq b > 0$ such that*

$$C = \left\{(x, y) \in \mathbb{A}^2 \mid \frac{x^2}{a^2} + \frac{y^2}{b^2} = 1\right\}. \tag{3.6}$$

Moreover, $a \in \mathbb{R}$ is exactly as in Definition 3.1; and if $2c$ is the distance between the foci then $b^2 = a^2 - c^2$.

Just as we derived an expression for an ellipse in terms of a Cartesian coordinate system we can do the same for hyperbolas and parabolas. We leave the required analysis as an exercise in each case and merely present the definitions.

Proposition 3.7 (Equation of a hyperbola) *Let C be a curve in $\mathbb{A}^2$. The curve C is a hyperbola if and only if there exists a Cartesian coordinate system (x, y) of $\mathbb{A}^2$ and real numbers $a > 0$, $b > 0$ such that*

$$C = \left\{ (x, y) \in \mathbb{A}^2 \mid \frac{x^2}{a^2} - \frac{y^2}{b^2} = 1 \right\}.$$

Moreover, $a \in \mathbb{R}$ is exactly as in Definition 3.3; and if $2c$ is the distance between the foci then $b^2 = c^2 - a^2$.

Proposition 3.8 (Equation of a parabola) *Let C be a curve in $\mathbb{A}^2$. The curve C is a parabola if and only if there exists a Cartesian coordinate system (x, y) of $\mathbb{A}^2$ and a real number $p > 0$ such that*

$$C = \left\{ (x, y) \in \mathbb{A}^2 \mid y^2 = 2px \right\}.$$

Moreover, $p \in \mathbb{R}$ is the distance between the directrix and the focus.

It is useful to compile all the relevant details, which we do in the following table.

	Geometric	**Algebraic**	**Parameters**
Ellipse	$\|\mathbf{F}_1 - \mathbf{P}\| + \|\mathbf{F}_2 - \mathbf{P}\| = 2a$	$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$	$2c = \|\mathbf{F}_2 - \mathbf{F}_1\|$; $b^2 = a^2 - c^2$
Hyperbola	$\|\|\mathbf{F}_1 - \mathbf{P}\| - \|\mathbf{F}_2 - \mathbf{P}\|\| = 2a$	$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$	$2c = \|\mathbf{F}_2 - \mathbf{F}_1\|$; $b^2 = c^2 - a^2$
Parabola	$\|\mathbf{F} - \mathbf{P}\| = \|l_{\mathbf{P}} - \mathbf{P}\|$	$y^2 = 2px$	$p = \|\mathbf{F} - l_{\mathbf{F}}\|$

Table 1: A comparison of the geometric and algebraic definitions of the ellipse, hyperbola and parabola.

Week 9

3.3 Polar coordinates

The algebraic expressions from the preceding section suggest a connection between the three types of curves. The first observation is that all the curves are described by quadratic expression, but looking closely we see that if we were working with complex numbers then (after substituting b with ib) ellipses and hyperbolas have the same expression.

In this section we will look at expressions for these curves in terms of focal polar coordinates, that is polar coordinates where the origin is a focus for the curve. This will highlight a deeper relationship between the curves and introduce the notion of eccentricity.

3.3.1 Ellipse or hyperbola

We begin with a treatment of the ellipse and hyperbola together. Let the origin $\mathbf{O} = \mathbf{F}_1$, be the first focus and let $\theta = 0$ indicate a direction towards $\mathbf{F}_2$. Let $\mathbf{P} = (r, \theta)$ be a point on the curve C, which is an ellipse or a hyperbola. For each θ there are two choices of r that give a point on the curve C. In the case of the ellipse we may always assume that $r > 0$, see Figure 12. In the case of the hyperbola things are slightly more complicated, here we say either: $\mathbf{P}$ is closer to $\mathbf{F}_2$ and $r > 0$; or instead $\mathbf{P}$ is closer to $\mathbf{F}_1$ and $r < 0$, see Figure 13. Therefore we have two cases for the hyperbola. Note that making the opposite choice would lead to a different, but perfectly good equation for the curve C.

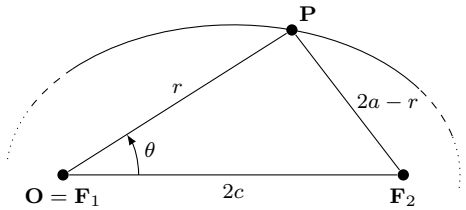


Figure 12: The radial distance r and angle θ for a point on an ellipse.

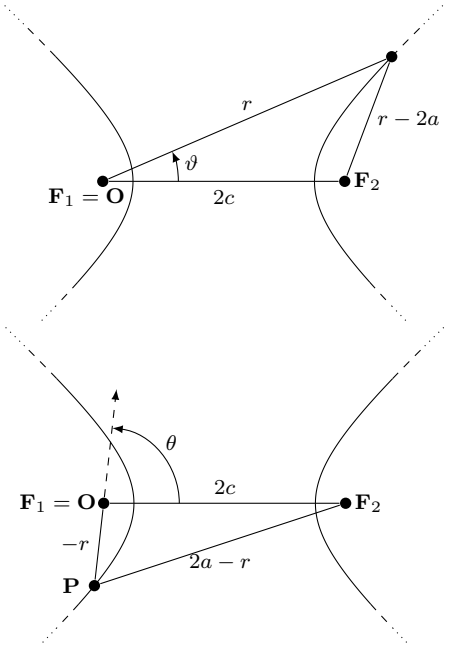


Figure 13: The radial distance r and angle θ for points on a hyperbola. In the left-hand diagram $r > 0$, whilst on the right $r < 0$.

Wikipedia: cosine rule

We can now apply the cosine rule to the triangle formed by the points $\mathbf{F}_1$, $\mathbf{F}_2$ and $\mathbf{P}$ to deduce that

$$(2a - r)^2 = r^2 + 4c^2 - 4cr\cos\theta \qquad \text{(ellipse or hyperbola case 1)}$$
$$(2a - r)^2 = r^2 + 4c^2 + 4cr\cos(\pi - \theta) \qquad \text{(hyperbola case 2).}$$

But since $\cos(\pi - \theta) = -\cos\theta$ these are the same and we have

$$\begin{aligned} 4a^2 - 4ar + r^2 &= r^2 + 4c^2 - 4cr\cos\theta \\ \Rightarrow \quad a^2 - ar &= c^2 - cr\cos\theta \\ \Rightarrow \quad r(a - c\cos\theta) &= a^2 - c^2 \\ \Rightarrow \quad r &= \frac{a^2 - c^2}{a - c\cos\theta}. \end{aligned}$$

Letting $p = \frac{a^2 - c^2}{a}$ and $e = \frac{c}{a}$ we say that the equation of the ellipse in polar coordinates is

$$r = \frac{p}{1 - e\cos\theta}. \tag{3.7}$$

The parameter e is called the *eccentricity* of the curve and notice that for an ellipse $0 \leq e < 1$, whilst for a hyperbola $e > 1$. In particular, if $e = 0$ then $c = 0$ and C is a circle.

Wikipedia: eccentricity

3.3.2 Parabola

The derivation of polar coordinates for a parabola is slightly simpler than for the other types of curve. In this situation we let the origin be the focus $\mathbf{O} = \mathbf{F}$, and say that the $\theta = 0$ direction is away from the directrix. As in the algebraic expression we say that the distance from focus $\mathbf{F}$ to directrix l is p.

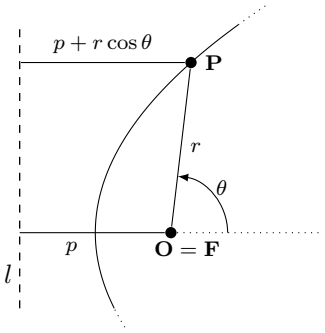


Figure 14: The radial distance r and angle θ for a point on an parabola.

Let $\mathbf{P} = (r, \theta)$ be a point on a parabola C. The distance from $\mathbf{P}$ to the directrix is $p + r\cos\theta$. The distance between $\mathbf{P}$ and $\mathbf{F}$ is simply r. Thus we have

$$r = p + r\cos\theta \quad \Rightarrow \quad r = \frac{p}{1 - \cos\theta}.$$

Notice that this matches equation (3.7) with $e = 1$, which is exactly the value of the eccentricity that we were missing.

Proposition 3.9 *Let C be an conic section. There exists a polar coordinate system (r, θ) of $\mathbb{A}^2$ with origin a focus point, and parameters $p, e \in \mathbb{R}$ with $e \geq 0$ such that*

$$C = \left\{ (r, \theta) \in \mathbb{A}^2 \mid r = \frac{p}{1 - e\cos\theta} \right\}. \tag{3.8}$$

Moreover, e is called the eccentricity (see Figure 15) and

$$\begin{aligned} e = 0 \quad &\Rightarrow \quad C \text{ is a circle;} \\ 0 < e < 1 \quad &\Rightarrow \quad C \text{ is an ellipse;} \\ e = 1 \quad &\Rightarrow \quad C \text{ is a parabola;} \\ e > 1 \quad &\Rightarrow \quad C \text{ is a hyperbola.} \end{aligned}$$

If C is a parabola then p is the distance between the focus and directrix. If C is an ellipse (including a circle) or a hyperbola then the distance between focus points is $|2c|$, and the defining distance of the curve is $|2a|$

where

$$a = \frac{p}{1-e^2} \qquad\qquad c = \frac{pe}{1-e^2}. \tag{3.9}$$

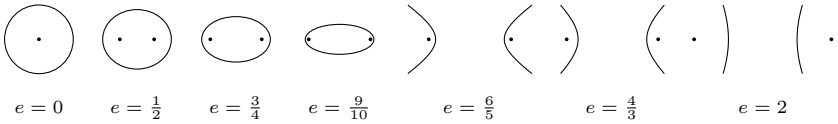


Figure 15: The conic sections shown with fixed distance $2a$, and in order of increasing eccentricity. The parabola (with $e = 1$) is not included as this has "infinite" distance.

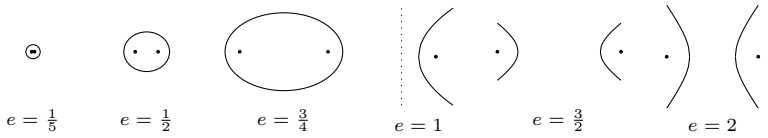


Figure 16: The conic sections shown with fixed focal parameter, p/e, and in order of increasing eccentricity. The circle (with e = 0) is not included as this has "infinite" focal parameter.

As a final remark to this section we address why ellipses, hyperbolas and parabolas are called conic sections. This is because all these curves are exactly given by the section of a cone passing through a plane. The type of curve, is dependent on the gradient of the plane with relative to the gradient of the cone. A plane with a smaller gradient intersects the cone in an ellipse; a plane with an equal gradient intersects the cone in a parabola; and a plane with a steeper gradient intersects the cone in a hyperbola. See Figure 17.

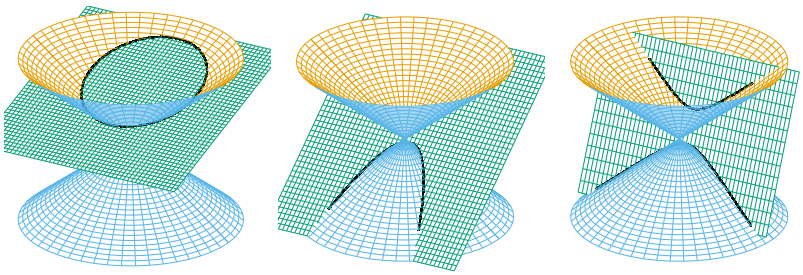


Figure 17: The curves formed by a plane intersecting a cone. On the left is a plane with a shallow gradient, giving an ellipse as its conic section. In the middle the plane has an equal gradient to the cone and intersects in a parabola. On the right the plane has a steeper gradient, producing a hyperbola.

4 Affine transformations

In this section we look at affine transformations; these are maps that act on affine spaces in a similar way to linear operators acting on vector spaces.

Wikipedia: affine transformation

Definition 4.1 (Affine transformation) Let $\mathbb{A}^n$ be an affine space. A function $f\colon \mathbb{A}^n \to \mathbb{A}^n$ is called an *affine transformation* if the following properties hold for all $\mathbf{P}_1, \mathbf{P}_2, \mathbf{Q}_1$ and $\mathbf{Q}_2$ in $\mathbb{A}^n$:

(1) the vectors $\mathbf{P}_2 - \mathbf{P}_1$ and $\mathbf{Q}_2 - \mathbf{Q}_1$ are linearly independent if and only if the vectors $f(\mathbf{P}_2) - f(\mathbf{P}_1)$ and $f(\mathbf{Q}_2) - f(\mathbf{Q}_1)$ are linearly independent;

(2) if $(\mathbf{P}_2 - \mathbf{P}_1) = \lambda(\mathbf{Q}_2 - \mathbf{Q}_1) \neq \mathbf{0}$ for some value $\lambda \in \mathbb{R}$ then

$$f(\mathbf{P}_2) - f(\mathbf{P}_1) = \lambda\big(f(\mathbf{Q}_2) - f(\mathbf{Q}_1)\big) \neq \mathbf{0}.$$

Remark 4.2 Two vectors are linearly dependent if and only if they are scalar multiples of one another, thus $P_2 - P_1$ and $Q_2 - Q_1$ are linearly dependent if and only if the line segment from P_1 to P_2 is parallel to the line segment from Q_1 to Q_2. Using this we can paraphrase the definition as:

(1) line segments are parallel if and only if their images under f are parallel;

(2) the map f scales parallel line segments by the same amount.

Recall that to each affine space $\mathbb{A}^n$, there is an associated vector space $\mathbb{E}^n$. Similarly, to each affine transformation of $\mathbb{A}^n$ there is an induced linear operator acting on $\mathbb{E}^n$.

Proposition 4.3 *Let $f\colon \mathbb{A}^n \to \mathbb{A}^n$ be an affine transformation and let $\mathbf{O}$ be a chosen origin in $\mathbb{A}^n$. The induced map $F\colon \mathbb{E}^n \to \mathbb{E}^n$ given by*

$$F(\mathbf{v}) = f(\mathbf{O} + \mathbf{v}) - f(\mathbf{O})$$

is an invertible linear operator of $\mathbb{E}^n$, and is independent of the choice of origin.

Proof. First notice that for any vectors $\mathbf{v}, \mathbf{w} \in \mathbb{E}^n$ and scalars $\lambda, \mu \in \mathbb{R}$ we have

$$(\mathbf{O} + \mu\mathbf{w}) - \mathbf{O} = \mu\mathbf{w} = (\mathbf{O} + \mu\mathbf{w} + \lambda\mathbf{v}) - (\mathbf{O} + \lambda\mathbf{v})$$

and so

$$f(\mathbf{O} + \mu\mathbf{w}) - f(\mathbf{O}) = f(\mathbf{O} + \mu\mathbf{w} + \lambda\mathbf{v}) - f(\mathbf{O} + \lambda\mathbf{v}).$$

Furthermore, by condition (2) in Definition 4.1, for any vector $\mathbf{v} \in \mathbb{E}^n$ and scalar $\lambda \in \mathbb{R}$ we have

$$\begin{aligned}(\mathbf{O} + \lambda\mathbf{v}) - \mathbf{O} &= \lambda\left((\mathbf{O} + \mathbf{v}) - \mathbf{O}\right)\\ \Rightarrow f(\mathbf{O} + \lambda\mathbf{v}) - f(\mathbf{O}) &= \lambda\left(f(\mathbf{O} + \mathbf{v}) - \mathbf{O}\right)\end{aligned}$$

We use this to show that the map is linear

$$\begin{aligned}F(\lambda\mathbf{v} + \mu\mathbf{w}) &= f(\mathbf{O} + (\lambda\mathbf{v} + \mu\mathbf{w})) - f(\mathbf{O})\\ &= \Big(f(\mathbf{O} + \lambda\mathbf{v}) - f(\mathbf{O})\Big) + \Big(f(\mathbf{O} + \lambda\mathbf{v} + \mu\mathbf{w}) - f(\mathbf{O} + \lambda\mathbf{v})\Big)\\ &= \Big(f(\mathbf{O} + \lambda\mathbf{v}) - f(\mathbf{O})\Big) + \Big(f(\mathbf{O} + \mu\mathbf{w}) - f(\mathbf{O})\Big)\\ &= \lambda\Big(f(\mathbf{O} + \mathbf{v}) - f(\mathbf{O})\Big) + \mu\Big(f(\mathbf{O} + \mathbf{w}) - f(\mathbf{O})\Big)\\ &= \lambda F(\mathbf{v}) + \mu F(\mathbf{w}).\end{aligned}$$

To see that the map is independent of the choice of origin we note that for any $\mathbf{P} \in \mathbb{A}^n$

$$(\mathbf{P} + \mathbf{v}) - \mathbf{P} = \mathbf{v} = (\mathbf{O} + \mathbf{v}) - \mathbf{O}$$

and hence

$$f(\mathbf{P} + \mathbf{v}) - f(\mathbf{P}) = f(\mathbf{O} + \mathbf{v}) - f(\mathbf{O}) = F(\mathbf{v})$$

Finally, to see that the map is invertible we need only show $F(\mathbf{v}) = \mathbf{0}$ implies $\mathbf{v} = \mathbf{0}$, and then appeal to the rank-nullity theorem. This is clear from the Definition 4.1 (2) since

$$\mathbf{v} = (\mathbf{O} + \mathbf{v}) - \mathbf{O} \neq \mathbf{0} \qquad \text{and} \qquad F(\mathbf{v}) = f(\mathbf{O} + \mathbf{v}) - f(\mathbf{O}) = \mathbf{0}$$

is forbidden. □

The only additional information required to recover the affine transformation from its induced linear operator is the image of a single point, usually the origin: $\mathbf{O} \mapsto \mathbf{O} + \mathbf{r}$.

Proposition 4.4 *Let $f\colon \mathbb{A}^n \to \mathbb{A}^n$ be an affine transformation. Let F be the induced linear operator defined in Proposition 4.3. Fix any point $\mathbf{O} \in \mathbb{A}^n$ and let $\mathbf{O}' = f(\mathbf{O})$ be its image under f. The map f can be expressed in terms of F and $\mathbf{O}'$ by*

$$f(\mathbf{P}) = \mathbf{O}' + F(\mathbf{P} - \mathbf{O}).$$

Proof. This is immediate from the definition of F since

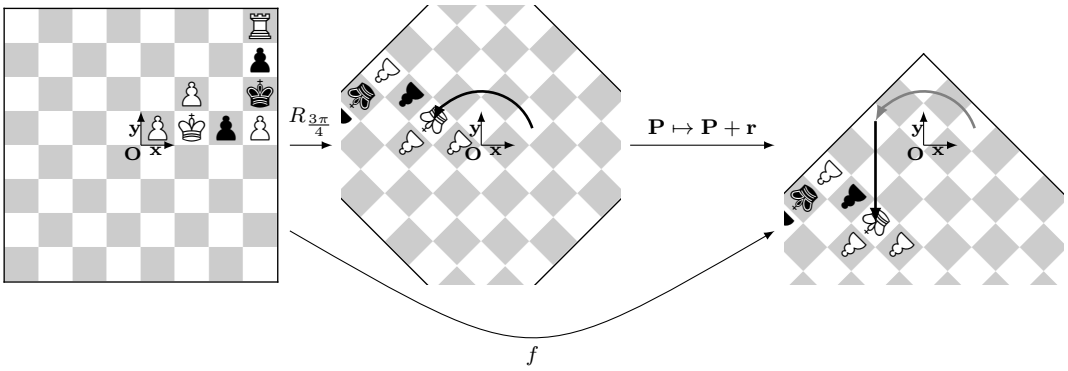
$$F(\mathbf{P}-\mathbf{O}) = f(\mathbf{P}) - f(\mathbf{O}) = f(\mathbf{P}) - \mathbf{O}'. \qquad \square$$

It is now clear that having fixed any origin $\mathbf{O}$ for affine space, an affine transformation can be expressed in terms of a linear operator together with a translation of the origin. Conversely, given any invertible linear operator and a translation we can construct an associated affine transformation. This leads to an alternative definition:

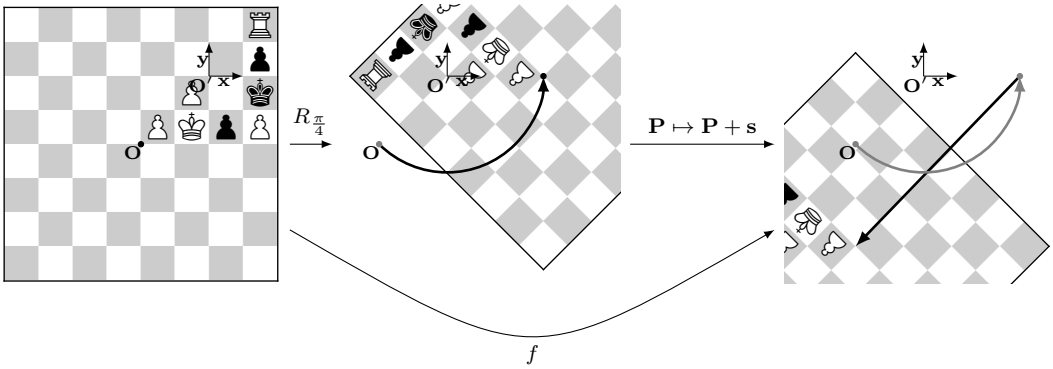
Alternative Definition 4.5 (Affine transformation) A map $f\colon \mathbb{A}^n \to \mathbb{A}^n$ is called an *affine transformation* if there is a point $\mathbf{O} \in \mathbb{A}^n$ (the origin), a translation vector $\mathbf{r}$ and an invertible linear operator $F\colon \mathbb{E}^n \to \mathbb{E}^n$ such that for any point $\mathbf{P} \in \mathbb{A}^n$

$$f(\mathbf{P}) = \mathbf{O} + F(\mathbf{P}-\mathbf{O}) + \mathbf{r}.$$

Example 4.6 The affine transformation f is made up of a rotation about $\mathbf{O}$ by $\frac{3\pi}{4}$ radians and a translation by $\mathbf{r} = -3\mathbf{y}$. The movement of the white king is indicated in the diagrams.



Example 4.7 If we express the map f from Example 4.6 with respect to a different origin, the linear map $R_{\frac{3\pi}{4}}$ remains the same but we must change the translation. For example, if $\mathbf{O}' = \mathbf{O} + 2\mathbf{x} + 2\mathbf{y}$ is the new origin then we must instead translate by the vector $\mathbf{s} = -2\sqrt{2}\mathbf{x} - 3\mathbf{y}$. The movement of the previous origin $\mathbf{O}$ is indicated in the diagrams.



Proposition 4.8 *An affine transformation $f\colon \mathbb{A}^2 \to \mathbb{A}^2$ is determined by the image of any three non-collinear points.*

Proof. Let $\mathbf{O}$, $\mathbf{P}_1$ and $\mathbf{P}_2$ be three points that do not lie on the same line. Let $\mathbf{e}_1 = \mathbf{P}_1 - \mathbf{O}$ and $\mathbf{e}_2 = \mathbf{P}_2 - \mathbf{O}$, so that $\mathcal{B} = (\mathbf{e}_1, \mathbf{e}_2)$ is a basis for $\mathbb{E}^2$. Any linear operator is determined by the image of a basis, therefore the linear operator F, induced by f is determined by $F(\mathbf{e}_1) = f(\mathbf{P}_1) - f(\mathbf{O})$ and $F(\mathbf{e}_2) = f(\mathbf{P}_2) - f(\mathbf{O})$. Now we can use the fact that f is determined by F and a translation, which we can determine using the image of any of our three points. □

4.1 Matrix of an affine transformation

Just as we have matrix representations of a linear operator, which make calculations convenient, we can define a matrix representation of an affine transformation. Instead of the matrix being defined with respect to just a basis, we also need to pick an origin for the representation. The pair of a basis and an origin is called a *frame* for the affine space.

Definition 4.9 (Frame) Let $\mathbb{A}^n$ be affine space and let $\mathbb{E}^n$ be its associated Euclidean vector space. We call any pair $\mathcal{F} = (\mathcal{B}, \mathbf{O})$, consisting of a basis $\mathcal{B}$ of $\mathbb{E}^n$, together with a point $\mathbf{O} \in \mathbb{A}^n$ a *frame* for $\mathbb{A}^n$. Wikipedia: frame

Example 4.10 A Cartesian coordinate system of $\mathbb{A}^n$ is a frame in which the basis is orthonormal.

A frame is convenient as it allows us to represent both points and vectors in column vector notation. We do this by adding an additional entry (row) to our vectors that is 0 in the case of vectors and 1 in the case of points.

Definition 4.11 (Vectors and points with respect to a frame) Let $\mathbb{A}^n$ be an affine space and $\mathbb{E}^n$ its associated vector space. Let $\mathcal{F} = \big((\mathbf{e}_1, \ldots, \mathbf{e}_n), \mathbf{O}\big)$ be a frame for $\mathbb{A}^n$. Let $\mathbf{v} = v_1\mathbf{e}_1 + \cdots + v_n\mathbf{e}_n$ be a vector in $\mathbb{E}^n$ and let $\mathbf{P} = \mathbf{O} + \mathbf{v}$ be a point of $\mathbb{A}^n$. Then we denote $\mathbf{v}$ and $\mathbf{O}$ *with respect to the frame* $\mathcal{F}$ by

$$[\mathbf{v}]_{\mathcal{F}} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \\ 0 \end{bmatrix} = \begin{bmatrix} [\mathbf{v}]_{\mathcal{B}} \\ 0 \end{bmatrix} \quad \text{and} \quad (\mathbf{P})_{\mathcal{F}} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \\ 1 \end{pmatrix} = \begin{pmatrix} [\mathbf{v}]_{\mathcal{B}} \\ 1 \end{pmatrix}. \tag{4.1}$$

Note that the choice of bracket is simply an additional aid to help distinguish points from vectors and is not of practical importance.

Definition 4.12 (Matrix of a affine transformation) Let $f\colon \mathbb{A}^n \to \mathbb{A}^n$ be an affine transformation and $\mathcal{F} = (\mathcal{B}, \mathbf{O})$ be a chosen frame for the

affine space, with $\mathcal{B} = (\mathbf{e}_1, \ldots, \mathbf{e}_n)$. Let $F\colon \mathbb{E}^n \to \mathbb{E}^n$ be the induced linear operator of f and $F(\mathbf{e}_i) = a_{1,i}\mathbf{e}_1 + \cdots + a_{n,i}\mathbf{e}_n$ for each basis vector $\mathbf{e}_i$ in $\mathcal{B}$. Let $f(\mathbf{O}) = \mathbf{O} + \mathbf{b}$, where $\mathbf{b} = b_1\mathbf{e}_1 + \cdots + b_n\mathbf{e}_n$.

Wikipedia: (augmented) matrix representation of f with respect to $\mathcal{F}$

Then the *(augmented) matrix representation of f with respect to $\mathcal{F}$* is

$$[f]_{\mathcal{F}} = \left[\begin{array}{cccc|c} a_{1,1} & a_{1,2} & \ldots & a_{1,n} & b_1 \\ a_{2,1} & a_{2,2} & \ldots & a_{2,n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{n,1} & a_{n,2} & \ldots & a_{n,n} & b_n \\ \hline 0 & 0 & \ldots & 0 & 1 \end{array}\right] = \begin{bmatrix} [F]_{\mathcal{B}} & [\mathbf{b}]_{\mathcal{B}} \\ 0 \cdots 0 & 1 \end{bmatrix}. \tag{4.2}$$

Note that the dividing lines in the matrix are not strictly necessary and are included for illustrative purposes. These lines will be dropped in most situations.

Notice that, just as with the matrix of a linear operators, matrix multiplication takes the place of applying the either of the maps f or F.

Remark 4.13 Multiplication of the matrix representation of f with a vector is equivalent to applying the linear operator F to the vector: let $\mathbf{w} = \sum w_i\mathbf{e}_i \in \mathbb{E}^n$ be a general vector. Then

$$\begin{aligned} [f]_{\mathcal{F}}\,[\mathbf{w}]_{\mathcal{F}} &= \begin{bmatrix} a_{1,1} & a_{1,2} & \ldots & a_{1,n} & b_1 \\ a_{2,1} & a_{2,2} & \ldots & a_{2,n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{n,1} & a_{n,2} & \ldots & a_{n,n} & b_n \\ 0 & 0 & \ldots & 0 & 1 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \\ 0 \end{bmatrix} \\ &= \begin{bmatrix} [F]_{\mathcal{B}}\,[\mathbf{w}]_{\mathcal{B}} \\ 0 \end{bmatrix} \\ &= \begin{bmatrix} [F(\mathbf{w})]_{\mathcal{B}} \\ 0 \end{bmatrix} \\ &= [F(\mathbf{w})]_{\mathcal{F}}. \end{aligned}$$

Remark 4.14 Multiplication of the matrix representation of f with a point is equivalent to applying the affine transformation to the point:

let $\mathbf{P} = \mathbf{O} + \mathbf{w} = \mathbf{O} + \sum w_i \mathbf{e}_i \in \mathbb{A}^n$ be a general point. Then

$$
\begin{aligned}
[f]_{\mathcal{F}} (\mathbf{P})_{\mathcal{F}} &= \begin{bmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,n} & b_1 \\ a_{2,1} & a_{2,2} & \dots & a_{2,n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{n,1} & a_{n,2} & \dots & a_{n,n} & b_n \\ 0 & 0 & \dots & 0 & 1 \end{bmatrix} \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \\ 1 \end{pmatrix} \\
&= \begin{pmatrix} [F(\mathbf{w})]_{\mathcal{B}} + [\mathbf{b}]_{\mathcal{B}} \\ 1 \end{pmatrix} \\
&= (f(\mathbf{P}))_{\mathcal{F}}.
\end{aligned}
$$

Example 4.15 The affine transformation f, of Example 4.6 was given with respect to a basis $\mathcal{B} = (\mathbf{x}, \mathbf{y})$ and an origin $\mathbf{O}$. Let $\mathcal{F} = (\mathcal{B}, \mathbf{O})$ be the frame determined by these data and recall that f was given by a rotation of $\frac{3\pi}{4}$ about $\mathbf{O}$ followed by a translation by $-3\mathbf{y}$. What is the matrix of f with respect tot he frame $\mathcal{F}$? The matrix of f with respect to the frame $\mathcal{F}$ is

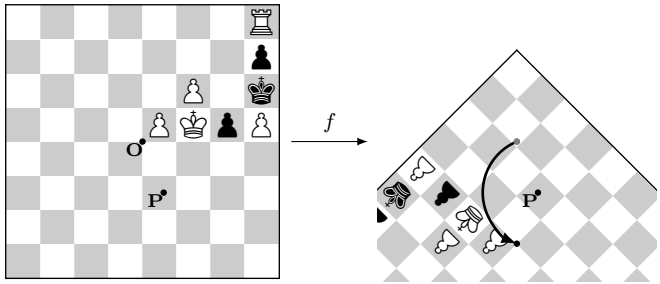
$$
[f]_{\mathcal{F}} = \left[\begin{array}{cc|c} \cos\frac{3\pi}{4} & -\sin\frac{3\pi}{4} & 0 \\ \sin\frac{3\pi}{4} & \cos\frac{3\pi}{4} & -3 \\ \hline 0 & 0 & 1 \end{array}\right] = \begin{bmatrix} \frac{-1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & -3 \\ 0 & 0 & 1 \end{bmatrix}.
$$

This matrix is far more convinient for calculations. For example, we can find a fixed point of the affine transformation by solving the equation

$$
\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{bmatrix} \frac{-1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} & -3 \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{-x-y}{\sqrt{2}} \\ \frac{x-y}{\sqrt{2}} - 3 \\ 1 \end{pmatrix}
$$

$$
\Rightarrow x = \frac{3}{2}\left(\sqrt{2} - 1\right), \qquad y = \frac{-3}{2}.
$$

The diagram below shows the affine transformation f, displayed as a rotation about the point $\mathbf{P} = \mathbf{O} + \frac{3}{2}(\sqrt{2} - 1)\mathbf{x} - \frac{3}{2}\mathbf{y}$; the path of the point $\mathbf{O}$ is highlighted.



Week 10

4.2 Affine transformations of conic sections

In this section we will consider how affine transformations act on the conic sections. Initially we will prove that any ellipse can be mapped to any other ellipse via an affine transformation. The same is true for the other types of conic section: a hyperbola can be mapped to any other hyperbola; and a parabola can be mapped to any other parabola. It is not possible however, to take one type of conic section to a different type via an affine transformation. In fact, the image of an ellipse, hyperbola or parabola under an affine transformation is always again an ellipse, hyperbola or parabola respectively.

Proposition 4.16 *Let C_1 and C_2 be ellipses in $\mathbb{A}^2$. Then there is an affine transformation f such that*

$$\mathbf{P} \in C_1 \quad \textit{if and only if} \quad f(\mathbf{P}) \in C_2.$$

Proof. By Proposition 3.6 we know that there is a pair $a, b \in \mathbb{R}$ and a frame $\mathcal{F} = (\mathcal{B}, \mathbf{O})$, with $\mathcal{B} = (\mathbf{e}_x, \mathbf{e}_y)$ orthonormal such that

$$\mathbf{P} = \mathbf{O} + x\mathbf{e}_x + y\mathbf{e}_y \in C_1 \quad \text{if and only if} \quad \frac{x^2}{a^2} + \frac{y^2}{b^2} = 1.$$

Similarly there is a pair $c, d \in \mathbb{R}$, a frame $\mathcal{F} = (\mathcal{C}, \mathbf{Q})$ with $\mathcal{C} = (\mathbf{f}_x, \mathbf{f}_y)$ orthonormal such that

$$\mathbf{P} = \mathbf{Q} + x\mathbf{f}_x + y\mathbf{f}_y \in C_2 \quad \text{if and only if} \quad \frac{x^2}{c^2} + \frac{y^2}{d^2} = 1.$$

Let $L\colon \mathbb{E}^2 \to \mathbb{E}^2$ be the linear operator given by

$$L(\mathbf{e}_1) = \frac{c}{a}\mathbf{f}_x \quad \text{and} \quad L(\mathbf{e}_2) = \frac{d}{b}\mathbf{f}_y.$$

Let f be the affine transformation defined by

$$f(\mathbf{P}) = \mathbf{Q} + L(\mathbf{P} - \mathbf{O}),$$

so

$$f(\mathbf{O} + x\mathbf{e}_x + y\mathbf{e}_y) = \mathbf{Q} + \frac{cx}{a}\mathbf{f}_x + \frac{dy}{b}\mathbf{f}_y.$$

With f as defined we have

$$\begin{aligned}
&\mathbf{P} = \mathbf{O} + x\mathbf{e}_x + y\mathbf{e}_y \in C_1 \\
&\qquad \iff \quad \frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 \\
&\qquad \iff \quad \frac{\left(\frac{cx}{a}\right)^2}{c^2} + \frac{\left(\frac{dy}{b}\right)^2}{d^2} = 1 \\
&\qquad \iff \quad f(\mathbf{P}) = \mathbf{Q} + \frac{cx}{a}\mathbf{f}_x + \frac{dy}{b}\mathbf{f}_y \in C_2. \qquad \square
\end{aligned}$$

We state the next two propositions, which concern hyperbolas and parabolas, without proofs since the arguments are very similar to those of Proposition 4.16.

Proposition 4.17 *Let C_1 and C_2 be hyperbolas in $\mathbb{A}^2$. Then there is an affine transformation f such that*

$$\mathbf{P} \in C_1 \quad \textit{if and only if} \quad f(\mathbf{P}) \in C_2.$$

Proposition 4.18 *Let C_1 and C_2 be parabolas in $\mathbb{A}^2$. Then there is an affine transformation f such that*

$$\mathbf{P} \in C_1 \quad \textit{if and only if} \quad f(\mathbf{P}) \in C_2.$$

Example 4.19 Let us consider the area of (the interior of) an ellipse, C which is given by the equation

$$\mathbf{P} = \mathbf{O} + x\mathbf{e}_x + y\mathbf{e}_y \in C \quad \text{if and only if} \quad \frac{x^2}{a^2} + \frac{y^2}{b^2} = 1.$$

with respect to a Cartesian frame $\mathcal{F} = \big((\mathbf{e}_x, \mathbf{e}_y), \mathbf{O}\big)$. Let U be the circle of radius 1 about $\mathbf{O}$: this has equation

$$\mathbf{P} = \mathbf{O} + x\mathbf{e}_x + y\mathbf{e}_y \in U \quad \text{if and only if} \quad x^2 + y^2 = 1.$$

Using Proposition 4.16 we have an affine transformation that maps U to C. Moreover, using the proof of the proposition, we see that the induced linear operator has matrix $\left[\begin{smallmatrix} a & 0 \\ 0 & b \end{smallmatrix}\right]$ and determinant ab. Since the determinant controls how areas scale, and we know the area of a unit circle is π, we see that C has area $ab\pi$. Recall that a is the length of the major axis and b is the length of the minor axis.

Theorem 4.20 *Let $C \subset \mathbb{A}^2$ be a conic section and $f\colon \mathbb{A}^2 \to \mathbb{A}^2$ be an affine transformation. Then*

- *the curve C is an ellipse if and only if $f(C)$ is an ellipse;*

- *the curve C is a hyperbola if and only if $f(C)$ is a hyperbola;*

- *the curve C is a parabola if and only if $f(C)$ is a parabola.*

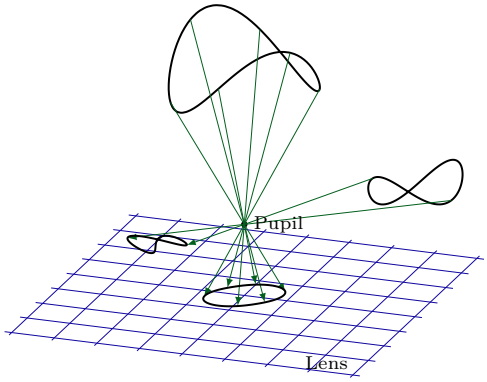
We shall not prove this theorem: the proof is most easily proved using algebraic ideas involving quadratic forms. These ideas lack much geometric intuition and hence are not particularly suited to this course. We will however make some comments on why one type of curve cannot be mapped to a different type.

Let E be an ellipse and C be a parabola or a hyperbola. The area of E is finite (see Example 4.19) and for any affine transformation, the determinant of the associated linear operator scales this area. Since the area of C is infinite the ellipse E cannot be mapped onto C, nor can C be mapped onto E.

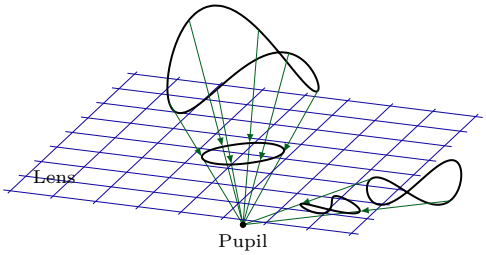
To see that a hyperbola H cannot be mapped onto a parabola P, consider the two tangent lines to H that are perpendicular to the line through the foci. These lines are parallel and each contain a single point of H: it follows that if a linear transformation mapped H to P then we would need two parallel lines, each of which containing a single point of P, but every point on a parabola has a tangent with a different gradient, so this cannot happen.

5 Projective Geometry

Projective geometry has its origins in understanding how people see the world, this is often called perspective. Consider a human eye: light enters through the pupil and makes an image on the lens at the back of the eye. The image on the lens is upside-down and reversed but is obviously interpreted by the brain correctly. As a simplified model, let us consider the pupil as a single point and the lens a plane behind the pupil. The light hitting a particular point on the lens is determined by a straight line from that point, through the pupil. In fact, if we wish the image to be correct orientated then we can instead model the lens in front of the pupil; see Figure 18. This simple model of an eye leads to the definition of projective geometry, where the *projective points* are the lines in affine space through a fixed origin.



(a) The lens is shown behind the pupil and the image reversed.



(b) Modelling the lens in front of the pupil gives a correctly orientated image.

Figure 18: Simple model of an eye: the image on the lens is determined by straight lines from objects in space through the pupil.

Throughout this section we will implicitly use Cartesian coordinates for $\mathbb{A}^n$. This means that we assume there is a chosen origin and points will be represented as n-tuples $(x_1, \ldots, x_n)$; or often (x, y) or (x, y, z) in 2-, or 3-dimensions.

5.1 Projective space

We will begin by considering the lines in $\mathbb{A}^2$ through the origin. Let us fix the affine line $y = 1$. It straightforward to see that all lines through the origin intersect this affine line in a unique point, except the line $y = 0$, which is parallel and therefore does not intersect it at all. Since the *projective points* are the lines through the origin, we see that the points in projective space are fully determined by an affine line $\mathbb{A}^1$, together with one extra point, which we call a *point at infinity*. Since this is intuitively one-dimensional, we call this space the 1-dimensional projective space; the projective line; or $\mathbb{P}^1$. Similarly, n-dimensional projective space, or $\mathbb{P}^n$, will be the lines through the origin in $\mathbb{A}^{n+1}$.

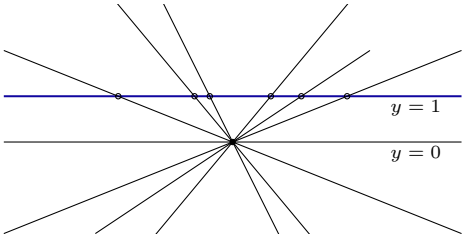


Figure 19: The projective line $\mathbb{P}^1$, is determined by points on a fixed affine line $y = 1$, together with one additional *point at infinity*.

Definition 5.1 (Projective space) We define *n-dimensional projective space*, denoted $\mathbb{P}^n$, to be the lines through the origin of the affine space

Wikipedia: n-dimensional projective space

$\mathbb{A}^{n+1}$.

Just as we have coordinates for $\mathbb{A}^{n+1}$, we wish to have a coordinate system for $\mathbb{P}^n$. Given a line through the origin, any point $(x_1, \ldots, x_{n+1})$ on the line, except the origin itself, determines the line. It is straightforward to see that two points $(a_1, \ldots, a_{n+1})$ and $(b_1, \ldots, b_{n+1})$ determine the same line if and only if there is a scalar λ such that $a_i = \lambda b_i$ for all $i \in \{1, \ldots, n+1\}$. Using this concept we define the *homogeneous coordinates* of projective space.

Wikipedia: homogeneous coordinate

Definition 5.2 (Homogeneous coordinates) Any $(n+1)$-tuple of real numbers, $(x_1, \ldots, x_{n+1})$, which are not all zero, determines a point in $\mathbb{P}^n$. This point is called a *homogeneous coordinate* of $\mathbb{P}^n$ and is denoted by $[x_1 : x_2 : \cdots : x_{n+1}]$. Note that for any $\lambda \neq 0$ we have

$$[x_1 : x_2 : \cdots : x_{n+1}] = [\lambda x_1 : \lambda x_2 : \cdots : \lambda x_{n+1}].$$

Example 5.3 Consider the projective line $\mathbb{P}^1$. Every point is given by a pair $[x : y]$, where at least one of x and y is non-zero. If $y \neq 0$ then $[x : y] = [\frac{x}{y} : 1]$ and these are exactly the points corresponding to lines that intersect $y = 1$ in Figure 19. On the other hand, if $y = 0$, then $x \neq 0$ and $[x : 0] = [1 : 0]$ is the only remaining point.

Example 5.4 Generalizing Example 5.3, consider the points in $\mathbb{P}^n$. These points are given by homogeneous coordinates $[x_1 : \cdots : x_n : z]$. If $z \neq 0$ then we see that

$$[x_1 : \cdots : x_n : z] = [\tfrac{x_1}{z} : \cdots : \tfrac{x_n}{z} : 1].$$

Since the first n values are arbitrary, we see that points of this form are in one-to-one correspondence with points in $\mathbb{A}^n$. The remaining points have $z = 0$ and have coordinates

$$[x_1 : \cdots : x_n : 0] = [\lambda x_1 : \cdots : \lambda x_n : 0]$$

for any $\lambda \in \mathbb{R}$. These points are in one-to-one correspondence with points in $\mathbb{P}^{n-1}$ and are often referred to as points at infinity.

The ideas in Example 5.3 and Example 5.4 show that we can determine most points in $\mathbb{P}^n$ by points in $\mathbb{A}^n$. These ideas lead to the concept of affine charts and associated affine coordinates for projective space.

Definition 5.5 (Affine chart) Let $\mathbb{P}^n$ be a projective space with homogeneous coordinates $[x_1 : \cdots : x_{n+1}]$. To each coordinate x_i there is an

associated *affine chart* $\mathbb{A}^n_i$ containing all projective points with $x_i \neq 0$:

$$\begin{aligned}\mathbb{A}^n_i &= \left\{[x_1 : \cdots : x_{n+1}] \in \mathbb{P}^n \;\middle|\; x_i \neq 0\right\} \\ &= \left\{[u_1 : \cdots : u_{i-1} : 1 : u_i : \cdots : u_n] \in \mathbb{P}^n\right\}.\end{aligned}$$

With respect to a given chart: the points on the chart are called *finite points* and the points outside the chart are called *points at infinity*. Wikipedia: points at infinity

In light of Definition 5.5, we see that the term *point at infinity* is a little ambiguous: a point at infinity with respect to one chart can become finite with respect to a different chart. Also notice that since a projective point cannot have all coordinates equal to zero, there always exists an affine chart containing any given projective point. That is, all projective points are finite with respect to some affine chart.

Example 5.6 Consider again the projective line $\mathbb{P}^1$ and Figure 20. The projective points that intersect the line $y = 1$ correspond to the points of the affine chart $\mathbb{A}^1_2$, where the second coordinate is non-zero. With respect to this chart the point $[1 : 0]$, corresponding to the x-axis, is a point at infinity. If instead we looked at the point of intersection with the line $x = 1$ we have the affine chart $\mathbb{A}^1_1$. The only projective point that does not intersect this line is $[0 : 1]$, corresponding to the y-axis.

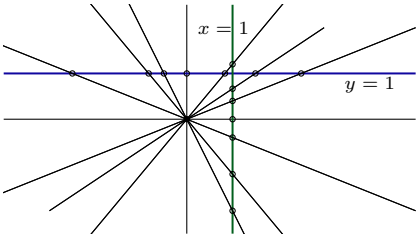


Figure 20: The projective line $\mathbb{P}^1$: the lines that intersect $y = 1$ correspond to one affine chart; the points that intersect $x = 1$ corresponds to the other.

Definition 5.7 (Affine coordinates) Let $\mathbb{A}^n_i$ be an affine chart for $\mathbb{P}^n$. Any point $\mathbf{P} = [x_1 : \cdots : x_{n+1}]$ in this chart has $x_i \neq 0$. We define the *affine coordinates* for $\mathbf{P}$ with respect to the affine chart $\mathbb{A}^n_i$ by the n-tuple $(\frac{x_1}{x_i}, \ldots, \frac{x_{n+1}}{x_i}) = (u_1, \ldots, u_n)$, with $\frac{x_i}{x_i}$ excluded. Wikipedia: affine coordinates

Example 5.8 Consider the point $\mathbf{P} = [1 : 2 : 0]$ in the projective plane, $\mathbb{P}^2$. What are its affine coordinates with respect to the first, second and third affine charts?

(1) $\mathbf{P}$ has affine coordinates $(2, 0)$ with respect to the first affine chart (first coordinate is removed).

(2) **P** has affine coordinates $(\frac{1}{2}, 0)$ with respect to the second affine chart (second coordinate is set to 1 and removed).

(3) **P** is a point at infinity with respect to the third affine chart and therefore has no associated affine coordinate.

Remark 5.9 Most of the time when we fix an affine chart, it will be the last affine chart; that is the one with non-zero last coordinate.

5.2 Projective transformations

We wish to define transformations of $\mathbb{P}^n$, and we will do so using transformations of $\mathbb{A}^{n+1}$. In order for an affine transformation of $\mathbb{A}^{n+1}$ to induce a map of projective space it must take lines through the origin to lines through the origin. Recall that affine transformations of this kind are in one-to-one correspondence with invertible linear operators.

Definition 5.10 (Projective transformation) Let $T\colon \mathbb{A}^{n+1} \to \mathbb{A}^{n+1}$ be an affine transformation that fixes the origin (equivalently T is an invertible linear operator). The map $\mathbb{P}^n \to \mathbb{P}^n$

$$[x_1 : x_2 : \cdots : x_{n+1}] \longmapsto [x'_1 : x'_2 : \cdots : x'_{n+1}],$$

where $T(x_1, \ldots, x_{n+1}) = (x'_1, \ldots, x'_{n+1})$, is called a *projective transformation*.

Wikipedia: projective transformation

Notice that scaling T by any non-zero value gives rise to the same projective transformation.

Example 5.11 Consider the projective transformation of $\mathbb{P}^1$ given by the matrix $\begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix}$. What does this map look like in terms of homogeneous and affine coordinates? In homogeneous coordinates this has the action

$$[x : y] \mapsto [\alpha x + \beta y : \gamma x + \delta y].$$

Considering affine coordinates (in the last affine chart) we see that

$$\frac{x}{y} \mapsto \frac{\alpha x + \beta y}{\gamma x + \delta y} = \frac{\alpha \frac{x}{y} + \beta}{\gamma \frac{x}{y} + \delta} \qquad \text{or simply} \qquad u \mapsto \frac{\alpha u + \beta}{\gamma u + \delta}.$$

This appears to induce a transformation on $\mathbb{A}^1$, however it is not defined for all values of u. Note that when $u = \frac{-\delta}{\gamma}$, the induced transformation is

$$\frac{-\delta}{\gamma} \mapsto \frac{-\alpha \frac{\delta}{\gamma} + \beta}{-\gamma \frac{\delta}{\gamma} + \delta} = \frac{-\alpha\delta + \beta\gamma}{0} = \infty \notin \mathbb{A}^1,$$

Considering the same transformation on $\mathbb{P}_1$, we see that this point gets mapped to

$$[-\delta : \gamma] \mapsto [-\alpha\delta + \beta\gamma : 0] = [1 : 0],$$

hence why we describe $[1 : 0]$ as a point at infinity on this affine chart.

Where does the point at infinity get mapped to via this transformation? Doing this on $\mathbb{P}^1$, we see that

$$[1 : 0] \mapsto [\alpha : \gamma].$$

However, with a massive abuse of notation we can also calculate this on the affine chart. If $u = \infty$, then the affine coordinates suggest the image

$$\infty \mapsto \frac{\alpha\infty + \beta}{\gamma\infty + \delta} = \frac{\alpha}{\gamma}$$

in affine coordinates. This matches the image given by homogeneous coordinates.

Remark 5.12 Later it will be helpful to be able to do arithmetic on the projective line. We can do arithmetic on any projective line $\mathbb{P}_1 = \mathbb{A}^1 \cup \{\infty\}$ via an abuse of notation with the following rules:

$$\frac{a}{0} = \infty, \quad \frac{a\infty}{b\infty} = \frac{a}{b}, \quad \infty + a = \infty \quad \forall a, b \in \mathbb{A}^1.$$

Formally, ∞ is just shorthand for the the point $[1 : 0]$ on the projective line, so you can't get away with using infinity like this unless you are working on a projective line!

In general, restricting a projective transformation to an affine chart does not give a well defined transformation. As Example 5.11 shows, these transformations tend to be defined almost everywhere on the affine chart. We call these *rational transformations*.

The formal definition of rational transformation is quite technical, therefore we won't define them properly.

When does a projective transformation define an affine transformation on an affine chart? The issue with Example 5.11 is that points on the affine chart were getting mapped to infinity and vice versa. The next proposition shows that if the projective transformation only maps affine points to affine points, it is an affine transformation. In particular, this shows that projective transformations generalize affine transformations.

Proposition 5.13 *Fix an affine chart $\mathbb{A}^n \subset \mathbb{P}^n$. Let $T\colon \mathbb{P}^n \to \mathbb{P}^n$ be a projective transformation such that for any $\mathbf{P} \in \mathbb{A}^n$ the image $T(\mathbf{P})$ is again in $\mathbb{A}^n$. Then T induces an affine transformation of the affine chart $\mathbb{A}^n$.*

Proof. Without loss of generality we fix $\mathbb{A}^n = \mathbb{A}^n_{n+1}$, where the last

coordinate is non-zero. Let

$$\begin{bmatrix} a_{1,1} & \dots & a_{1,n} & a_{1,n+1} \\ \vdots & \ddots & \vdots & \vdots \\ a_{n,1} & \dots & a_{n,n} & a_{n,n+1} \\ a_{n+1,1} & \dots & a_{n+1,n} & a_{n+1,n+1} \end{bmatrix}$$

be a matrix representation for T. (Equivalently, a matrix representation for a linear operator inducing T.) Consider the point $\mathbf{P} = [0 : \cdots : 0 : 1] \in \mathbb{P}^n$. This point is in the affine chart and so $T(\mathbf{P}) = [a_{1,n+1} : \cdots : a_{n+1,n+1}]$ is also in the affine chart. This means that $a_{n+1,n+1} \neq 0$ and so by rescaling the matrix we can assume the matrix is of the form

$$\begin{bmatrix} b_{1,1} & \dots & b_{1,n} & b_{1,n+1} \\ \vdots & \ddots & \vdots & \vdots \\ b_{n,1} & \dots & b_{n,n} & b_{n,n+1} \\ b_{n+1,1} & \dots & b_{n+1,n} & 1 \end{bmatrix} \qquad \text{where} \qquad b_{i,j} = \frac{a_{i,j}}{a_{n+1,n+1}}.$$

Now let us assume that $b_{n+1,1}$ is non-zero: then the affine chart contains the point $\left[\frac{-1}{b_{n+1,1}} : 0 : \cdots : 0 : 1\right]$ and this would have an image with last coordinate 0. Since this cannot happen we conclude that $b_{n+1,1} = 0$. Similarly, we can conclude that all entries in the last row are zero apart from the last and the matrix of P has the form

$$\begin{bmatrix} b_{1,1} & \dots & b_{1,n} & b_{1,n+1} \\ \vdots & \ddots & \vdots & \vdots \\ b_{n,1} & \dots & b_{n,n} & b_{n,n+1} \\ 0 & \dots & 0 & 1 \end{bmatrix}. \tag{5.1}$$

This is exactly the matrix of an affine transformation. □

Example 5.14 Let us consider how a general projective transformation acts on an affine chart. Let $T\colon \mathbb{P}^2 \to \mathbb{P}^2$ be a projective transformation given by the matrix

$$\begin{bmatrix} a_x & a_y & a_z \\ b_x & b_y & b_z \\ c_x & c_y & c_z \end{bmatrix} \tag{5.2}$$

and with action

$$\begin{aligned} T\colon \mathbb{P}^2 &\longrightarrow \mathbb{P}^2 \\ [x : y : z] &\longmapsto [a_x x + a_y y + a_z z : b_x x + b_y y + b_z z : c_x x + c_y y + c_z z] \end{aligned}$$

What is the induced rational transformation on $\mathbb{A}^2_3$? and when is this

transformation affine? On the affine chart $\mathbb{A}^2_3$ we see that the induced rational transformation is

$$\begin{aligned} \mathbb{A}^2 &\xrightarrow{T} \mathbb{A}^2 \\ (u,v) &\xmapsto{T} \left(\frac{a_x u + a_y v + a_z}{c_x u + c_y v + c_z}, \frac{b_x u + b_y v + b_z}{c_x u + c_y v + c_z} \right). \end{aligned}$$

When is this transformation affine? Notice that the map $f\colon \mathbb{A}^2 \to \mathbb{A}^2$ with the action

$$(u,v) \xmapsto{f} (\alpha u + \beta v + \gamma, \delta u + \epsilon v + \zeta)$$

is just the general form for an affine transformation. We see therefore that T induces an affine transformation if and only if $c_x = c_y = 0$ and $c_z \neq 0$. In particular, we can scale the matrix (5.2) by $\frac{1}{c_z}$ to get it in the form (5.1).

Week 11/12

5.3 Collinearity

In this section we consider what is meant by points in projective space being *collinear*. Intuitively we want to say that if points are collinear in affine coordinates then they are collinear in projective space. Let us consider the projective plane $\mathbb{P}^2$, with points $[x : y : z]$ and the affine chart with non-zero final coordinate, with affine coordinates $(\frac{x}{z}, \frac{y}{z}) = (u, v)$. The general equation of a line in $\mathbb{A}^2$ is given by $au + bv + c = 0$ for some real values a, b and c. A point $[x : y : z]$ has affine coordinates on this line if

$$a\frac{x}{z} + b\frac{y}{z} + c = 0,$$

which is true if

$$ax + by + cz = 0.$$

This second equation is the general equation for a plane in $\mathbb{A}^3$ that passes through the origin. The same ideas hold in higher dimensions: three points will be collinear in affine coordinates if the homogeneous coordinates represent points on the same plane.

Wikipedia: collinear

Definition 5.15 (Collinearity) We say that three points, $[x_1 : x_2 : \cdots : x_{n+1}]$, $[y_1 : y_2 : \cdots : y_{n+1}]$ and $[z_1 : z_2 : \cdots : z_{n+1}]$ in $\mathbb{P}^n$ are *collinear*, or that they lie on the same projective line, if there is a plane through the origin in $\mathbb{A}^{n+1}$ that contains the points $(x_1, \ldots, x_{n+1})$, $(y_1, \ldots, y_{n+1})$ and $(z_1, \ldots, z_{n+1})$.

Recall that three arbitrary vectors support points on the same plane if they are linearly dependent, this gives an alternative way to calculate collinearity in $\mathbb{P}^2$. Three points $[x_1 : y_1 : z_1]$, $[x_2 : y_2 : z_2]$ and $[x_3 : y_3 : z_3]$ are collinear if their associated vectors are linearly dependent; or equivalently when

$$\det \begin{bmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ z_1 & z_2 & z_3 \end{bmatrix} = 0.$$

Example 5.16 Consider the points

$$\mathbf{A} = [0 : -1 : 2],\ \mathbf{B} = [-1 : 0 : 3],\ \mathbf{C} = [2 : -4 : 2].$$

Are they colinear?

All these points belong to the affine chart $\mathbb{A}^2_3$ and so we can check collinearity in their affine coordinates. The affine coordinates are:

$$\mathbf{A} = (0, -1/2),\ \mathbf{B} = (-1/3, 0),\ \mathbf{C} = (1, -2).$$

These all satisfy the equation $3u + 2v + 1 = 0$ and are hence collinear.

Consider now the point $\mathbf{D} = [2 : -3 : 0]$. Are $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{D}$ colinear?

This point is not in the same affine chart; we have two options: check if this point is collinear on a different affine chart; or calculate collinearity using homogeneous coordinates. Using homogeneous coordinates we can check if the determinant of the matrix given by $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{D}$ is zero:

$$\det \begin{bmatrix} 0 & -1 & 2 \\ -1 & 0 & -3 \\ 2 & 3 & 0 \end{bmatrix} = 1\,(0+6) + 2\,(-3+0) = 6 - 6 = 0.$$

We conclude that all four points are collinear.

5.4 The cross ratio

Recall that the property preserved by affine transformations was the ratio of lengths of collinear line segments. This means that if T is an affine transformation and $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ are collinear points then

$$\frac{\mathbf{A} - \mathbf{C}}{\mathbf{B} - \mathbf{C}} = \frac{T(\mathbf{A}) - T(\mathbf{C})}{T(\mathbf{B}) - T(\mathbf{C})}. \tag{5.3}$$

When we generalize to projective transformations, it is no longer true that this ratio is preserved (see Example 5.21). The equivalent property that is preserved for projective transformations is called the *cross ratio* and involves a fourth point.

Definition 5.17 (Cross ratio) Let $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ and $\mathbf{D}$ be four collinear points in affine space $\mathbb{A}^n$. The *cross ratio* $(\mathbf{A}, \mathbf{B}; \mathbf{C}, \mathbf{D})$ is defined to be

Wikipedia: cross ratio

$$(\mathbf{A}, \mathbf{B}; \mathbf{C}, \mathbf{D}) = \frac{(\mathbf{A} - \mathbf{C})}{(\mathbf{B} - \mathbf{C})} \frac{(\mathbf{B} - \mathbf{D})}{(\mathbf{A} - \mathbf{D})} \tag{5.4}$$

By extension, the cross ratio of four collinear points in projective space is the cross ratio of the four points in any affine chart.

Remark 5.18 Note that Definition 5.17 and equation (5.3) do not make sense if the points are not collinear as we cannot divide arbitrary

vectors. This is only well defined when the vectors are parallel and therefore multiples of one another.

Example 5.19 Consider the following four points in $\mathbb{P}^2$:

$$\mathbf{A} = [1:0:0],\ \mathbf{B} = [1:1:1],\ \mathbf{C} = [1:2:2],\ \mathbf{D} = [1:3:3].$$

What is their cross-ratio? These points all lie on the first affine chart $\mathbb{A}^2_1$, given by affine coordinates:

$$\mathbf{A} = (0,0),\ \mathbf{B} = (1,1),\ \mathbf{C} = (2,2),\ \mathbf{D} = (3,3).$$

Furthermore they are collinear, as they all lie on the line $u-v=0$ in $\mathbb{A}^2_1$. Therefore the vectors between these points are in the span of $\mathbf{v} = [1,1]$. Therefore we can calculate their cross ratio in this affine chart to be

$$\begin{aligned}(\mathbf{A},\mathbf{B};\mathbf{C},\mathbf{D}) &= \frac{(\mathbf{A}-\mathbf{C})}{(\mathbf{B}-\mathbf{C})}\frac{(\mathbf{B}-\mathbf{D})}{(\mathbf{A}-\mathbf{D})}\\ &= \frac{0-2\mathbf{v}}{\mathbf{v}-2\mathbf{v}}\cdot\frac{\mathbf{v}-3\mathbf{v}}{0-3\mathbf{v}}\\ &= \frac{-2\mathbf{v}}{-\mathbf{v}}\cdot\frac{-2\mathbf{v}}{-3\mathbf{v}}\\ &= \frac{4}{3}.\end{aligned}$$

The cross ratio should be well defined for any choice of affine chart, therefore we can consider these points on the third affine chart $\mathbb{A}^2_3$:

$$\mathbf{A} = \infty,\ \mathbf{B} = (1,1),\ \mathbf{C} = \left(\frac{1}{2},1\right),\ \mathbf{D} = \left(\frac{1}{3},1\right).$$

While $\mathbf{A}$ is at point at infinity on the third affine chart, these points are collinear and so live on a copy of the projective line $\mathbb{P}^1$.

Therefore we can use the same arithmetic abuse with infinity as in Remark 5.12.

$$(\mathbf{A},\mathbf{B};\mathbf{C},\mathbf{D}) = \frac{(\mathbf{A}-\mathbf{C})}{(\mathbf{B}-\mathbf{C})}\frac{(\mathbf{B}-\mathbf{D})}{(\mathbf{A}-\mathbf{D})} = \frac{\infty\cdot{}^2\!/_3}{{}^1\!/_2\cdot\infty} = \frac{{}^2\!/_3}{{}^1\!/_2} = \frac{4}{3}.$$

While we calculated this on a different affine chart, the cross ratio remains the same.

It is not immediately clear that the cross ratio is well-defined for projective points, however the following proposition deals with this issue.

Proposition 5.20 *Let* $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ *and* $\mathbf{D}$ *be four collinear points in* $\mathbb{P}^n$ *and let* $T\colon \mathbb{P}^n \to \mathbb{P}^n$ *be a projective transformation. Then the cross ratio of the four points is preserved by* T*:*

$$(\mathbf{A},\mathbf{B};\mathbf{C},\mathbf{D}) = (T(\mathbf{A}),T(\mathbf{B});T(\mathbf{C}),T(\mathbf{D}))$$

In particular, the cross-ratio is independent of the choice of affine chart.

Proof. We prove the result for $\mathbb{P}^1$.

Let $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ and $\mathbf{D}$ be (collinear) points of $\mathbb{P}^1$ and let $T\colon \mathbb{P}^1 \to \mathbb{P}^1$ be the projective transformation:

$$[x : y] \overset{T}{\longmapsto} [\alpha x + \beta y : \gamma x + \delta y].$$

After possibly changing basis we can assume that all the points lie in the second affine chart and we have the map of affine coordinates

$$u \overset{T}{\longmapsto} \frac{\alpha u + \beta}{\gamma u + \delta}.$$

Let $u_{\mathbf{A}}$, $u_{\mathbf{B}}$, $u_{\mathbf{C}}$ and $u_{\mathbf{D}}$ be the affine coordinates of the points in this affine chart. The cross ratio of the images of the points is now

$$\begin{aligned}
&(T(\mathbf{A}), T(\mathbf{B}); T(\mathbf{C}), T(\mathbf{D})) \\
&= \frac{\left(\frac{\alpha u_{\mathbf{A}}+\beta}{\gamma u_{\mathbf{A}}+\delta} - \frac{\alpha u_{\mathbf{C}}+\beta}{\gamma u_{\mathbf{C}}+\delta}\right)\left(\frac{\alpha u_{\mathbf{B}}+\beta}{\gamma u_{\mathbf{B}}+\delta} - \frac{\alpha u_{\mathbf{D}}+\beta}{\gamma u_{\mathbf{D}}+\delta}\right)}{\left(\frac{\alpha u_{\mathbf{B}}+\beta}{\gamma u_{\mathbf{B}}+\delta} - \frac{\alpha u_{\mathbf{C}}+\beta}{\gamma u_{\mathbf{C}}+\delta}\right)\left(\frac{\alpha u_{\mathbf{A}}+\beta}{\gamma u_{\mathbf{A}}+\delta} - \frac{\alpha u_{\mathbf{D}}+\beta}{\gamma u_{\mathbf{D}}+\delta}\right)} \\
&= \frac{\left(\frac{(\alpha u_{\mathbf{A}}+\beta)(\gamma u_{\mathbf{C}}+\delta)-(\alpha u_{\mathbf{C}}+\beta)(\gamma u_{\mathbf{A}}+\delta)}{(\gamma u_{\mathbf{A}}+\delta)(\gamma u_{\mathbf{C}}+\delta)}\right)\left(\frac{(\alpha u_{\mathbf{B}}+\beta)(\gamma u_{\mathbf{D}}+\delta)-(\alpha u_{\mathbf{D}}+\beta)(\gamma u_{\mathbf{B}}+\delta)}{(\gamma u_{\mathbf{B}}+\delta)(\gamma u_{\mathbf{D}}+\delta)}\right)}{\left(\frac{(\alpha u_{\mathbf{B}}+\beta)(\gamma u_{\mathbf{C}}+\delta)-(\alpha u_{\mathbf{C}}+\beta)(\gamma u_{\mathbf{B}}+\delta)}{(\gamma u_{\mathbf{B}}+\delta)(\gamma u_{\mathbf{C}}+\delta)}\right)\left(\frac{(\alpha u_{\mathbf{A}}+\beta)(\gamma u_{\mathbf{D}}+\delta)-(\alpha u_{\mathbf{D}}+\beta)(\gamma u_{\mathbf{A}}+\delta)}{(\gamma u_{\mathbf{A}}+\delta)(\gamma u_{\mathbf{D}}+\delta)}\right)} \\
&= \frac{\left(\frac{(\alpha\delta-\beta\gamma)(u_{\mathbf{A}}-u_{\mathbf{C}})}{(\gamma u_{\mathbf{A}}+\delta)(\gamma u_{\mathbf{C}}+\delta)}\right)\left(\frac{(\alpha\delta-\beta\gamma)(u_{\mathbf{B}}-u_{\mathbf{D}})}{(\gamma u_{\mathbf{B}}+\delta)(\gamma u_{\mathbf{D}}+\delta)}\right)}{\left(\frac{(\alpha\delta-\beta\gamma)(u_{\mathbf{B}}-u_{\mathbf{C}})}{(\gamma u_{\mathbf{B}}+\delta)(\gamma u_{\mathbf{C}}+\delta)}\right)\left(\frac{(\alpha\delta-\beta\gamma)(u_{\mathbf{A}}-u_{\mathbf{D}})}{(\gamma u_{\mathbf{A}}+\delta)(\gamma u_{\mathbf{D}}+\delta)}\right)} \\
&= \frac{(u_{\mathbf{A}} - u_{\mathbf{C}})(u_{\mathbf{B}} - u_{\mathbf{D}})}{(u_{\mathbf{B}} - u_{\mathbf{C}})(u_{\mathbf{A}} - u_{\mathbf{D}})} \\
&= (\mathbf{A}, \mathbf{B}; \mathbf{C}, \mathbf{D}).
\end{aligned}$$

In higher dimensions we can use the fact that collinear points are all in 1-dimensional subspace equivalent to $\mathbb{P}^1$ and use the proof above. This makes sense geometrically, however to make this rigorous would require more technically details than we wish to cover. □

The alternative approach in higher dimensions is to rearrange some even more tedious equations, which the author (and likely the reader) would rather avoid doing.

Example 5.21 Consider the projective transformation $T\colon \mathbb{P}^2 \to \mathbb{P}^2$ given by the matrix

$$\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}.$$

Does the projective transofmration preserve ratio of distances? Does it preserve the cross ratio? This is the matrix of an affine transformation

so it is immediate that equation (5.3) is satisfied on the affine chart $\mathbb{A}^2_3$. Let us see what happens on the affine chart $\mathbb{A}^2_1$: here the projective transformation acts as

$$(u, v) \longmapsto \left(\frac{1+v}{u}, \frac{v}{u}\right).$$

Recall the collinear points from Example 5.19, all except $\mathbf{A}$ are on the first affine chart. The points $\mathbf{B} = (1, 1)$, $\mathbf{C} = (2, 2)$ and $\mathbf{D} = (3, 3)$ are collinear with images $T(\mathbf{B}) = (2, 1)$, $T(\mathbf{C}) = (3/2, 1)$ and $T(\mathbf{D}) = (4/3, 1)$. The ratio of distances are therefore

$$\frac{\mathbf{B}-\mathbf{D}}{\mathbf{C}-\mathbf{D}} = 2 \qquad \text{and} \qquad \frac{T(\mathbf{B})-T(\mathbf{D})}{T(\mathbf{C})-T(\mathbf{D})} = \frac{2-4/3}{3/2-4/3} = \frac{2/3}{1/6} = 4.$$

We can see that in general the ratio of distances preserved by affine transformations is not preserved for projective transformations.

Let us check that T does indeed preserve the cross ratio. We see that T acts on each of the points as follows:

$$T(\mathbf{A}) = [0 : 1 : 0], \qquad T(\mathbf{B}) = [1 : 2 : 1],$$
$$T(\mathbf{C}) = [2 : 3 : 2], \qquad T(\mathbf{D}) = [3 : 4 : 3].$$

All of these points live on the second affine chart with affine coordinates

$$T(\mathbf{A}) = (0, 0), \qquad T(\mathbf{B}) = \left(\frac{1}{2}, \frac{1}{2}\right),$$
$$T(\mathbf{C}) = \left(\frac{2}{3}, \frac{2}{3}\right), \qquad T(\mathbf{D}) = \left(\frac{3}{4}, \frac{3}{4}\right),$$

Therefore we can calculate the cross ratio as

$$\frac{(T(\mathbf{A})-T(\mathbf{C}))\,(T(\mathbf{B})-T(\mathbf{D}))}{(T(\mathbf{B})-T(\mathbf{C}))\,(T(\mathbf{A})-T(\mathbf{D}))} = \frac{-2/3 \cdot -1/4}{-3/4 \cdot -1/6} = \frac{4}{3}$$

and see that it is preserved by T. Note that we could equally calculate the cross ratio on any other affine chart, although would have to deal with points at infinity.

5.5 Projective transformations of conic sections

We proved in Theorem 4.20 that affine transformations map ellipses to ellipses, hyperbolas to hyperbolas and parabolas to parabolas, but that we cannot map one type of conic section onto another. Projective transformations do not have this limitation. In this section we will prove that we can map different types of conic section onto one another via projective transformations.

Consider the set of points in the projective plane

$$D = \{[x : y : z] \in \mathbb{P}^2 \mid x^2 + y^2 = z^2\} \subset \mathbb{P}^2.$$

When restricted to the third affine chart, D just becomes the unit circle:

$$D \cap \mathbb{A}^2_3 = \{(u, v) \in \mathbb{A}^2 \mid u^2 + v^2 = 1\} \subset \mathbb{A}^2.$$

The following theorem shows we can map the unit circle to any other conic section given the correct choice of projective transformation.

Theorem 5.22 *For any conic section $C \subset \mathbb{A}^2$, there exists a projective transformation $T\colon \mathbb{P}^2 \to \mathbb{P}^2$ such that C is $T(D)$ restricted to the third affine chart.*

Proof. We split this into cases for each type of conic section.

Ellipse: The case of an ellipse has already been covered. A special case of Proposition 4.16 shows that the affine transformation

$$(u, v) \mapsto (au, bv)$$

maps the unit circle to the ellipse with equation $\frac{u^2}{v^2} + \frac{y^2}{b^2} = 1$. As a projective transformation, this is

$$[x : y : z] \mapsto [ax : by : z].$$

Hyperbola: Let C be the hyperbola described by the equation $1 + \frac{v^2}{b^2} - \frac{u^2}{a^2} = 0$. Define the projective transformation

$$T\colon [x : y : z] \mapsto [az : by : x].$$

For a point $\mathbf{P} = [x : y : z] \in D$, let $T(\mathbf{P}) = [x' : y' : z'] = [az : by : x]$ where

$$x = z', \quad y = \frac{y'}{b}, \quad z = \frac{x'}{a}.$$

The point $\mathbf{P} \in D$ satisfies the equation $x^2 + y^2 - z^2 = 0$ if and only if $T(\mathbf{P})$ satisfies the equation

$$0 = z'^2 + \left(\frac{y'}{b}\right)^2 - \left(\frac{x'}{a}\right)^2.$$

Restricting to the third affine chart, we see this is the equation of the

hyperbola C:

$$\begin{aligned} 0 &= \frac{z'^2}{z'^2} + \frac{y'^2}{b^2 z'^2} - \frac{x'^2}{a^2 z'^2} \\ &= 1 + \frac{\left(\frac{y'}{z'}\right)^2}{b^2} - \frac{\left(\frac{x'}{z'}\right)^2}{a^2} \\ &= 1 + \frac{v^2}{b^2} - \frac{u^2}{a^2}. \end{aligned}$$

Parabola: Let C be the parabola described by the equation $v^2 - 2pu = 0$. Define the projective transformation

$$T\colon [x : y : z] \mapsto [z - x : \sqrt{2p}y : z + x].$$

For a point $\mathbf{P} = [x : y : z] \in D$, let $T(\mathbf{P}) = [x' : y' : z'] = [z - x : \sqrt{2p}y : z + x]$ where

$$x = \frac{z' - x'}{2}, \quad y = \frac{y'}{\sqrt{2p}}, \quad z = \frac{z' + x'}{2}.$$

The point $\mathbf{P} \in D$ satisfies the equation $x^2 + y^2 - z^2 = 0$ if and only if $T(\mathbf{P})$ satisfies the equation

$$\begin{aligned} 0 &= \left(\frac{z' - x'}{2}\right)^2 + \left(\frac{y'}{\sqrt{2p}}\right)^2 - \left(\frac{z' + x'}{2}\right)^2 \\ &= \frac{z'^2 - 2x'z' + x'^2}{4} + \frac{y'^2}{2p} - \frac{z'^2 + 2x'z' + x'^2}{4} \\ &= y'^2 - 2px'z'. \end{aligned}$$

Restricting to the third affine chart, we see this is the equation of the parabola C:

$$\begin{aligned} 0 = \frac{y'^2}{z'^2} - \frac{2px'z'}{z'^2} &= \left(\frac{y'}{z'}\right)^2 - 2p\left(\frac{x'}{z'}\right) \\ &= v^2 - 2pu. \end{aligned}$$

□

Corollary 5.23 *Any two conic sections are projectively equivalent, i.e., there exists a projective transformation from one to another.*

Proof. Projective transformations are invertible: if T is represented by the matrix M then T^{-1} is represented by the matrix M^{-1}. As a result we can map any conic section to the circle via the inverse transformation of the one constructed in Theorem 5.22. Therefore we can map any conic section to another by transforming to the circle. □

Remark 5.24 When considered as lines in $\mathbb{A}^3$, the set D is the cone shown in Figure 17. Recall that we can obtain all conic sections as intersections of this cone with a plane. It is not a coincidence that we

use this same cone in the proof of Theorem 5.22. An alternative intuition behind this proof is that we are applying a linear transformation to this cone and then intersecting with the plane $z = 1$. While it is conceptually more complex, working in projective space makes this proof much easier: in $\mathbb{A}^3$ the proof is far messier!

Thank you for taking Introduction to Geometry :)

Index