

# Before you start

This book, like the lectures, is written in the ordinary language of mathematics: sets and the symbol  $\in$ , set-builder braces such as  $\{x \in \mathbb{R} : x > 0\}$  (the colon is read “such that”), quantifiers  $\forall$  and  $\exists$ , implications, sums with  $\Sigma$ , and the words *definition*, *theorem*, *proposition*, *lemma*, *corollary* and *proof*. If any of these is unfamiliar, start with the course’s notation primer, *The Language of Mathematics*: an eight-page sheet (`notation-primer.pdf`) and a twelve-minute video in seven parts, with a one-page quick reference you can keep beside this book.

# Chapter 1

## Vectors and the geometry of $\mathbb{R}^n$

### Learning goals

After this chapter you should be able to:

- move freely between the arrow picture and the list picture of a vector;
- state the definition of  $\mathbb{R}^n$  and decide when two vectors are equal;
- add vectors, multiply them by scalars, and form linear combinations, both algebraically and geometrically;
- compute dot products, lengths and angles, and prove the Cauchy–Schwarz and triangle inequalities;
- describe lines and planes in parametric and normal form, and use the cross product in  $\mathbb{R}^3$ .

### 1.1 What linear algebra is about

#### Lecture 1

Linear algebra studies *vectors* and the operations that combine and transform them. It is one of the most widely used parts of mathematics: computer graphics, statistics, signal processing, economics, quantum mechanics and machine learning all rest on it. The reason is that linear problems are the problems we can solve completely, and many problems that are not linear can be approximated, at least locally, by problems that are.

This section introduces the vector in its two standard forms, as an arrow and as a list of numbers, fixes the notation of the course, and previews the questions the course is built to answer. Most of the ideas here are developed carefully in later sections; the aim now is a map of the territory. Symbols such as  $\in$ , set braces and  $\forall$  are explained in the notation primer *The Language of Mathematics* (see *Before you start*).

#### 1.1.1 Vectors as arrows

In the plane, a vector is an arrow: it has a *length* and a *direction*, and nothing else. In particular it has no fixed starting point. Two arrows that have the same length and point the same way represent the same vector, wherever they are drawn (Figure 1.1). This matches the quantities vectors usually

describe. A displacement of “three steps east and two steps north” means the same thing from any starting position, and the same is true of a velocity or a force.

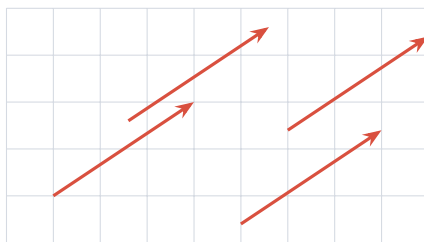


Figure 1.1: Four arrows, one vector: each arrow is three units across and two units up.

It is still convenient to agree on a standard position. Unless we say otherwise, we draw vectors with their tail at the origin. A vector drawn this way is determined by the point where its tip lands, and the coordinates of that point are called the *components* of the vector. The arrow from the origin to the point  $(3, 2)$  has first component 3 (the horizontal step) and second component 2 (the vertical step); see Figure 1.2. Conversely, any pair of numbers determines exactly one arrow from the origin. A negative component points the other way along its axis, and the pair  $(0, 0)$  gives the *zero vector*, an arrow of length zero, drawn as a point.

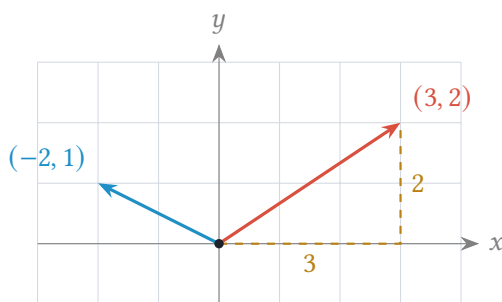


Figure 1.2: Drawn from the origin, a vector is described by its components.

### 1.1.2 Vectors as lists

Algebraically, a vector is an ordered list of numbers. We write lists vertically, as *column vectors*:

$$\mathbf{v} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}.$$

The order matters:  $\begin{bmatrix} 3 \\ 2 \end{bmatrix}$  and  $\begin{bmatrix} 2 \\ 3 \end{bmatrix}$  are different vectors. To save space in running text we also write  $(3, 2)$  for the column vector with entries 3 and 2.

**Definition 1.1.1** (The space  $\mathbb{R}^n$ ). Let  $n$  be a positive integer. The set  $\mathbb{R}^n$  consists of all ordered lists of  $n$  real numbers,

$$\mathbb{R}^n = \left\{ \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} : x_1, \dots, x_n \in \mathbb{R} \right\}.$$

Elements of  $\mathbb{R}^n$  are called *vectors*, and the numbers  $x_1, \dots, x_n$  are the *components* (or *entries*) of the vector. Real numbers, in contrast, are called *scalars*.

**Definition 1.1.2** (Equality of vectors). Two vectors  $\mathbf{x} = (x_1, \dots, x_n)$  and  $\mathbf{y} = (y_1, \dots, y_n)$  in  $\mathbb{R}^n$  are *equal*, written  $\mathbf{x} = \mathbf{y}$ , when  $x_i = y_i$  for every  $i = 1, \dots, n$ .

A single equation between vectors in  $\mathbb{R}^n$  therefore says  $n$  separate things, one for each component. Vectors with different numbers of components belong to different spaces and are never equal:  $(1, 2) \in \mathbb{R}^2$  and  $(1, 2, 0) \in \mathbb{R}^3$  are different objects.

**Example 1.1.3.** Find all real numbers  $a$  and  $b$  with  $\begin{bmatrix} a+b \\ a-2b \end{bmatrix} = \begin{bmatrix} 7 \\ 1 \end{bmatrix}$ .

*Solution.* By the definition of equality, the vector equation means  $a + b = 7$  and  $a - 2b = 1$ . Subtracting the second equation from the first gives  $3b = 6$ , so  $b = 2$  and then  $a = 5$ . Conversely  $a = 5$ ,  $b = 2$  satisfies both equations, so it is the only solution.

$\mathbb{R}^2$  is the plane and  $\mathbb{R}^3$  is ordinary three-dimensional space, where a vector needs three components and its arrow needs a third axis. For  $n \geq 4$  we can no longer draw arrows, but lists of  $n$  numbers make perfect sense, and the algebra we develop works in every dimension. This is the source of much of the power of the subject: lists of numbers appear everywhere.

**Example 1.1.4** (Vectors in the wild).

- *Colour.* A colour on a screen is a vector in  $\mathbb{R}^3$  recording the amounts of red, green and blue light, each between 0 and 1. The gold accent of the course slides is approximately  $(0.95, 0.72, 0.29)$ .
- *Recipes.* A recipe that uses 500 g flour, 200 g sugar and 3 eggs is the vector  $(500, 200, 3)$ . Doubling the recipe doubles every component.
- *Measurements.* A weather record of temperature, pressure, humidity and wind speed is a vector in  $\mathbb{R}^4$ .
- *Data.* In data analysis each survey response, image or document is recorded as a vector, often with thousands of components.

**Notation.** In this book vectors are set in bold,  $\mathbf{v}, \mathbf{w}, \mathbf{x}$ ; on the lecture slides they carry a small arrow, so  $\mathbf{v}$  in the book is  $\vec{v}$  on the slides. Scalars are plain italic letters  $a, b, c, x, y$ . The components of  $\mathbf{x}$  are  $x_1, x_2, \dots$ , and the zero vector of  $\mathbb{R}^n$  is written  $\mathbf{0}$ .

### 1.1.3 Two operations

What makes vectors more than lists is what we can do with them. Linear algebra is built on two operations.

- *Addition.* To add arrows, place the tail of  $\mathbf{w}$  at the tip of  $\mathbf{v}$ ; the sum  $\mathbf{v} + \mathbf{w}$  runs from the tail of  $\mathbf{v}$  to the tip of  $\mathbf{w}$ . For lists, add matching components:  $(2, 1) + (1, 2) = (3, 3)$ .
- *Scalar multiplication.* Multiplying by a scalar  $c$  stretches the arrow by the factor  $|c|$ , reversing it when  $c < 0$ . For lists, multiply every component:  $2(2, 1) = (4, 2)$  and  $-1(2, 1) = (-2, -1)$ .

Combining the two operations gives *linear combinations*  $a\mathbf{v} + b\mathbf{w}$ . These operations, their rules, and the geometry of linear combinations are the subject of Sections 1.2 and 1.3.

### 1.1.4 Four questions

The course is organised around four questions. Every later topic is a step towards answering one of them.

1. **What can we build?** Given vectors  $\mathbf{v}_1, \dots, \mathbf{v}_k$ , which vectors can be written as linear combinations of them? Asking whether a given vector  $\mathbf{b}$  can be built is the same as asking whether a system of linear equations has a solution (Chapter 2).
2. **What does a matrix do?** A matrix is a rectangular array of numbers, best understood as a map that sends vectors to vectors. Such maps rotate, stretch, shear, project and reflect, always keeping grid lines straight, parallel and evenly spaced (Chapters 3 and 6).
3. **Which directions are special?** Most vectors are turned by a matrix, but some are only stretched. These eigenvectors and their eigenvalues reveal the structure of the map and govern processes that repeat over time (Chapters 8 and 9).
4. **What is the best approximation?** Systems built from measured data often have no exact solution. Finding the closest possible answer is a question about lengths and right angles, answered by least squares and the singular value decomposition (Chapters 10 and 11).

### 1.1.5 One system, three pictures

**Example 1.1.5.** Solve the system

$$\begin{aligned} 2x - y &= 1, \\ x + y &= 5. \end{aligned}$$

*Row picture.* Each equation describes a line in the plane,  $y = 2x - 1$  and  $y = 5 - x$ . A solution lies on both lines, so it is their intersection. Setting  $2x - 1 = 5 - x$  gives  $x = 2$ , and then  $y = 3$ . Check:  $2 \cdot 2 - 3 = 1$  and  $2 + 3 = 5$ .

*Column picture.* Group the coefficients by unknown:

$$x \begin{bmatrix} 2 \\ 1 \end{bmatrix} + y \begin{bmatrix} -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 5 \end{bmatrix}.$$

By the definition of equality this vector equation is exactly the original system. It asks for the linear combination of the two columns that reaches the target  $(1, 5)$ . Indeed  $2(2, 1) + 3(-1, 1) = (4, 2) + (-3, 3) = (1, 5)$ .

*Matrix picture.* Place the columns side by side to form a matrix, and collect the unknowns and the right-hand side into vectors:

$$A = \begin{bmatrix} 2 & -1 \\ 1 & 1 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 1 \\ 5 \end{bmatrix}.$$

The system becomes the single equation  $A\mathbf{x} = \mathbf{b}$ , whose meaning is made precise in Section 3.1.

**Remark 1.1.6.** Two lines in the plane usually meet in exactly one point, but they may be parallel (no common point) or coincide (infinitely many common points). In Chapter 2 we prove that every system of linear equations, in any number of unknowns, has either no solution, exactly one solution, or infinitely many.

#### Check yourself

Write the system  $3x + y = 4$ ,  $x - 2y = 6$  as a single vector equation in the column picture, and check that  $(x, y) = (2, -2)$  solves it.

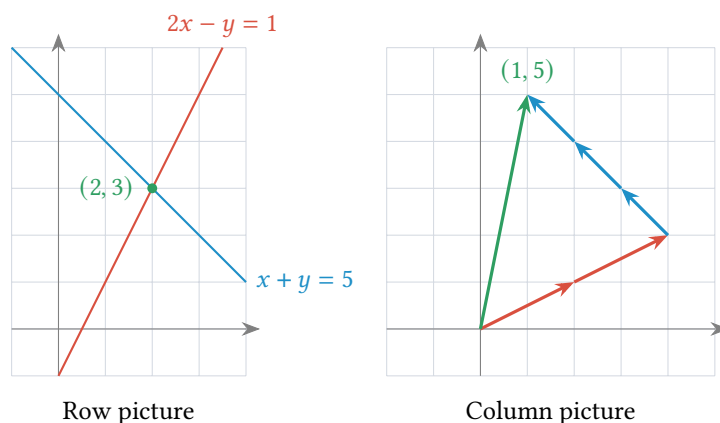


Figure 1.3: Example 1.1.5: two lines meeting at  $(2, 3)$ , and two copies of  $(2, 1)$  plus three copies of  $(-1, 1)$  reaching  $(1, 5)$ .

## How this course works

Each lecture has a matching section of this book. The book states every definition precisely and proves every theorem it states; the lectures prove the main theorems and give the intuition behind them. Each week ends with a problem sheet in four tiers (A warm-up, B core, C proof and concept, D challenge). Try the problems before reading the solutions: the effort of getting stuck and unstuck is where the learning happens.

## 1.2 Adding and scaling vectors

### Lecture 2

Section 1.1 introduced the two operations on vectors informally. This section defines them precisely in  $\mathbb{R}^n$ , explains each one geometrically, and proves the eight rules they obey. Those rules are the most important result of the section: from Chapter 4 on they become the definition of a vector space, and every fact we derive from them alone will hold far beyond lists of numbers.

Throughout, for a vector  $\mathbf{x} \in \mathbb{R}^n$  we write  $x_i$ , or  $(\mathbf{x})_i$  when the vector is given by an expression, for its  $i$ -th entry. By the definition of equality (Section 1.1), two vectors in  $\mathbb{R}^n$  are equal exactly when  $(\mathbf{x})_i = (\mathbf{y})_i$  for every  $i = 1, \dots, n$ ; almost every proof in this section uses that fact.

### 1.2.1 Vector addition

**Definition 1.2.1** (Vector addition). Let  $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ . Their *sum* is the vector in  $\mathbb{R}^n$  obtained by adding matching entries:

$$\mathbf{v} + \mathbf{w} = \begin{bmatrix} v_1 + w_1 \\ v_2 + w_2 \\ \vdots \\ v_n + w_n \end{bmatrix}, \quad \text{that is,} \quad (\mathbf{v} + \mathbf{w})_i = v_i + w_i \quad (i = 1, \dots, n).$$

The sum is defined only for two vectors with the same number of entries. A vector in  $\mathbb{R}^2$  and a vector in  $\mathbb{R}^3$  have no sum: the third entry of the second vector would have nothing to be added to.

**Example 1.2.2.** In  $\mathbb{R}^3$ ,

$$\begin{bmatrix} 1 \\ 4 \\ -2 \end{bmatrix} + \begin{bmatrix} 3 \\ -1 \\ 5 \end{bmatrix} = \begin{bmatrix} 1+3 \\ 4+(-1) \\ -2+5 \end{bmatrix} = \begin{bmatrix} 4 \\ 3 \\ 3 \end{bmatrix}.$$

**Addition as a picture.** In  $\mathbb{R}^2$ , draw  $\mathbf{v}$  and then place the tail of  $\mathbf{w}$  at the tip of  $\mathbf{v}$ . The sum  $\mathbf{v} + \mathbf{w}$  is the arrow from the tail of  $\mathbf{v}$  to the tip of  $\mathbf{w}$  (Figure 1.4, left). This *tip-to-tail rule* agrees with Definition 1.2.1: the arrow  $\mathbf{v} = (2, 1)$  is a step of 2 across and 1 up, the arrow  $\mathbf{w} = (1, 2)$  is a further step of 1 across and 2 up, so the whole trip is  $2 + 1 = 3$  across and  $1 + 2 = 3$  up, which is the arrow  $(3, 3) = \mathbf{v} + \mathbf{w}$ . The same argument works one coordinate at a time in any  $\mathbb{R}^n$ : the horizontal and vertical steps simply add.

Walking along  $\mathbf{w}$  first and then along  $\mathbf{v}$  ends at the same point. The two routes are the sides of a parallelogram with corners  $\mathbf{0}$ ,  $\mathbf{v}$ ,  $\mathbf{w}$  and  $\mathbf{v} + \mathbf{w}$ , and the sum is the diagonal from  $\mathbf{0}$  (Figure 1.4, right). This is the *parallelogram rule*, and it already shows that  $\mathbf{v} + \mathbf{w} = \mathbf{w} + \mathbf{v}$ , which we prove in Theorem 1.2.7.

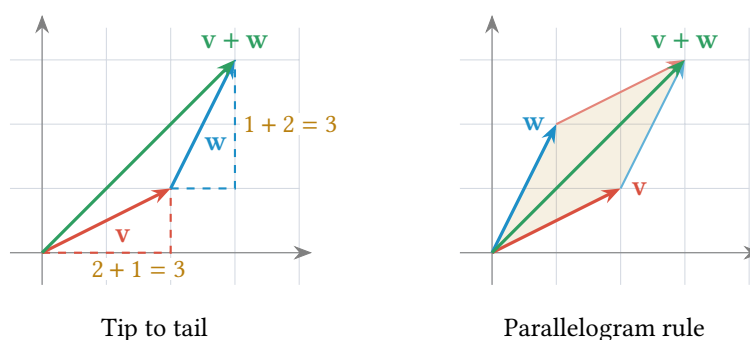


Figure 1.4: Adding  $\mathbf{v} = (2, 1)$  and  $\mathbf{w} = (1, 2)$ . Left: the steps add coordinate by coordinate. Right: both routes around the parallelogram end at  $\mathbf{v} + \mathbf{w} = (3, 3)$ .

## 1.2.2 Scalar multiplication

**Definition 1.2.3** (Scalar multiplication). Let  $c \in \mathbb{R}$  be a scalar and  $\mathbf{v} \in \mathbb{R}^n$ . The *scalar multiple*  $c\mathbf{v}$  is the vector in  $\mathbb{R}^n$  obtained by multiplying every entry of  $\mathbf{v}$  by  $c$ :

$$c\mathbf{v} = \begin{bmatrix} c v_1 \\ c v_2 \\ \vdots \\ c v_n \end{bmatrix}, \quad \text{that is,} \quad (c\mathbf{v})_i = c v_i \quad (i = 1, \dots, n).$$

**Example 1.2.4.**  $3 \begin{bmatrix} 2 \\ -1 \\ 4 \end{bmatrix} = \begin{bmatrix} 3 \cdot 2 \\ 3 \cdot (-1) \\ 3 \cdot 4 \end{bmatrix} = \begin{bmatrix} 6 \\ -3 \\ 12 \end{bmatrix}, \quad \text{and} \quad 0 \begin{bmatrix} 2 \\ -1 \\ 4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$

**Scaling as a picture.** For  $c > 0$  the arrow  $c\mathbf{v}$  points the same way as  $\mathbf{v}$  and is  $c$  times as long:  $2\mathbf{v}$  is twice as long,  $\frac{1}{2}\mathbf{v}$  half as long. For  $c < 0$  the arrow is also reversed:  $-\mathbf{v}$  has the length of  $\mathbf{v}$  and points the opposite way, and  $-2\mathbf{v}$  is twice as long and reversed. For  $c = 0$  we get the zero vector. In short, scaling by  $c$  multiplies the length by  $|c|$  and reverses the direction when  $c < 0$ ; we prove this once length has been defined, in Section 1.4. All the multiples  $c\mathbf{v}$  of a nonzero vector lie on one line through the origin, the line in the direction of  $\mathbf{v}$  (Figure 1.5).

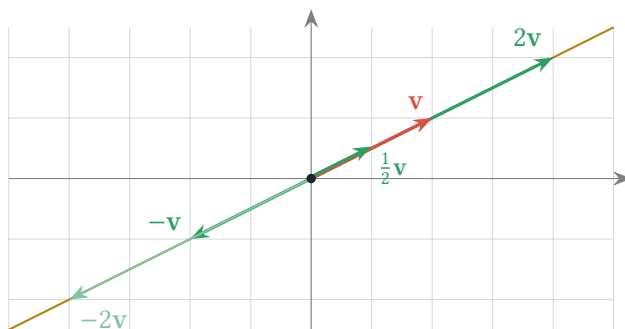


Figure 1.5: Multiples of  $\mathbf{v} = (2, 1)$ . Every multiple  $c\mathbf{v}$  lies on the line through  $\mathbf{0}$  in the direction of  $\mathbf{v}$ .

### 1.2.3 The zero vector, negatives and subtraction

- Definition 1.2.5.** (a) The *zero vector* of  $\mathbb{R}^n$  is  $\mathbf{0} = (0, 0, \dots, 0)$ , the vector whose entries are all 0.
- (b) The *negative* of  $\mathbf{v} \in \mathbb{R}^n$  is  $-\mathbf{v} = (-v_1, \dots, -v_n)$ , the vector whose entries are those of  $\mathbf{v}$  with their signs changed. Equivalently,  $-\mathbf{v} = (-1)\mathbf{v}$ .
- (c) The *difference* of  $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$  is  $\mathbf{v} - \mathbf{w} = \mathbf{v} + (-\mathbf{w})$ . Its entries are  $(\mathbf{v} - \mathbf{w})_i = v_i - w_i$ .

**Subtraction as a picture.** Draw  $\mathbf{v}$  and  $\mathbf{w}$  from the origin. Then  $\mathbf{v} - \mathbf{w}$  is the arrow from the tip of  $\mathbf{w}$  to the tip of  $\mathbf{v}$ , because starting at  $\mathbf{w}$  and adding  $\mathbf{v} - \mathbf{w}$  brings us to  $\mathbf{w} + (\mathbf{v} - \mathbf{w}) = \mathbf{v}$  (Figure 1.6). Unlike addition, subtraction is not commutative:  $\mathbf{w} - \mathbf{v} = -(\mathbf{v} - \mathbf{w})$  is the same arrow reversed.

The same idea describes movement between points. If  $P$  and  $Q$  are points with coordinate vectors  $\mathbf{p}$  and  $\mathbf{q}$ , the step from  $P$  to  $Q$  is the vector  $\mathbf{q} - \mathbf{p}$ , “the end minus the start”. In running text we also write it  $Q - P$ .

**Example 1.2.6.** For  $\mathbf{v} = (4, 1)$  and  $\mathbf{w} = (1, 3)$ ,  $\mathbf{v} - \mathbf{w} = (4 - 1, 1 - 3) = (3, -2)$ . The step from  $P = (-3, 1)$  to  $Q = (2, 3)$  is  $Q - P = (2 - (-3), 3 - 1) = (5, 2)$ : five units right and two up.

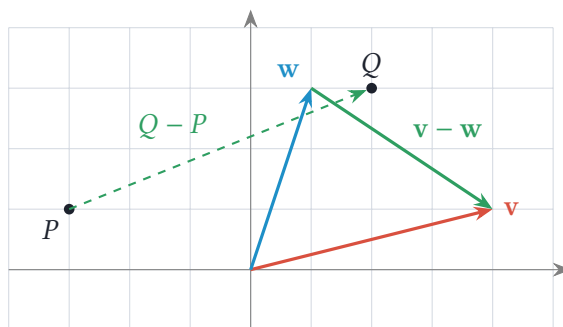


Figure 1.6: The difference  $\mathbf{v} - \mathbf{w}$  runs from the tip of  $\mathbf{w}$  to the tip of  $\mathbf{v}$ ; the step from  $P$  to  $Q$  is  $Q - P$ .

## 1.2.4 The rules of vector algebra

**Theorem 1.2.7** (Rules of vector algebra). For all vectors  $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^n$  and all scalars  $a, b \in \mathbb{R}$ :

- |                                                                                       |                                         |
|---------------------------------------------------------------------------------------|-----------------------------------------|
| (1) $\mathbf{v} + \mathbf{w} = \mathbf{w} + \mathbf{v}$                               | (addition is commutative)               |
| (2) $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$ | (addition is associative)               |
| (3) $\mathbf{v} + \mathbf{0} = \mathbf{v}$                                            | (the zero vector)                       |
| (4) $\mathbf{v} + (-\mathbf{v}) = \mathbf{0}$                                         | (negatives)                             |
| (5) $1 \mathbf{v} = \mathbf{v}$                                                       | (scaling by one)                        |
| (6) $a(b \mathbf{v}) = (ab) \mathbf{v}$                                               | (scaling twice)                         |
| (7) $a(\mathbf{v} + \mathbf{w}) = a \mathbf{v} + a \mathbf{w}$                        | (distributive law over vector addition) |
| (8) $(a + b) \mathbf{v} = a \mathbf{v} + b \mathbf{v}$                                | (distributive law over scalar addition) |

*Proof.* Each rule is an equation between two vectors of  $\mathbb{R}^n$ , so by the definition of equality it is enough to show that the two sides have the same  $i$ -th entry for every  $i = 1, \dots, n$ . Fix such an  $i$ . In each case we compute both entries from Definitions 1.2.1, 1.2.3 and 1.2.5 and compare them using a rule of arithmetic for real numbers.

- (1)  $(\mathbf{v} + \mathbf{w})_i = v_i + w_i = w_i + v_i = (\mathbf{w} + \mathbf{v})_i$ , because addition of real numbers is commutative.
- (2)  $((\mathbf{u} + \mathbf{v}) + \mathbf{w})_i = (u_i + v_i) + w_i = u_i + (v_i + w_i) = (\mathbf{u} + (\mathbf{v} + \mathbf{w}))_i$ , because addition of real numbers is associative.
- (3)  $(\mathbf{v} + \mathbf{0})_i = v_i + 0 = v_i$ .
- (4)  $(\mathbf{v} + (-\mathbf{v}))_i = v_i + (-v_i) = 0 = (\mathbf{0})_i$ .
- (5)  $(1 \mathbf{v})_i = 1 \cdot v_i = v_i$ .
- (6)  $(a(b \mathbf{v}))_i = a(b v_i) = (ab) v_i = ((ab) \mathbf{v})_i$ , because multiplication of real numbers is associative.
- (7)  $(a(\mathbf{v} + \mathbf{w}))_i = a(v_i + w_i) = a v_i + a w_i = (a \mathbf{v})_i + (a \mathbf{w})_i = (a \mathbf{v} + a \mathbf{w})_i$ , by the distributive law for real numbers.
- (8)  $((a + b) \mathbf{v})_i = (a + b) v_i = a v_i + b v_i = (a \mathbf{v})_i + (b \mathbf{v})_i = (a \mathbf{v} + b \mathbf{v})_i$ , again by the distributive law for real numbers.

In every case the two sides agree in every entry, so they are equal.  $\square$

By rule (2) we may write  $\mathbf{u} + \mathbf{v} + \mathbf{w}$  without brackets, and by rules (1) and (2) together a sum of any number of vectors can be added in any order and grouped in any way. Two of the rules have a picture worth keeping in mind (Figure 1.7). Associativity says that walking along  $\mathbf{u}$ , then  $\mathbf{v}$ , then  $\mathbf{w}$  ends at the same point however the steps are grouped. The first distributive law says that scaling a parallelogram by  $a$  scales its diagonal by  $a$ : the parallelogram on  $a\mathbf{v}$  and  $a\mathbf{w}$  is the parallelogram on  $\mathbf{v}$  and  $\mathbf{w}$  enlarged by the factor  $a$ , so its diagonal  $a\mathbf{v} + a\mathbf{w}$  is  $a(\mathbf{v} + \mathbf{w})$ .

**Looking ahead.** In Chapter 4 these eight rules are turned around. Any set of objects that can be added and multiplied by scalars, in a way that obeys rules (1) to (8), will be called a *vector space*. Polynomials, matrices and functions all form vector spaces. For that reason, from now on we try to prove facts about vectors from the eight rules alone, without looking at entries: such proofs hold in every vector space at once.

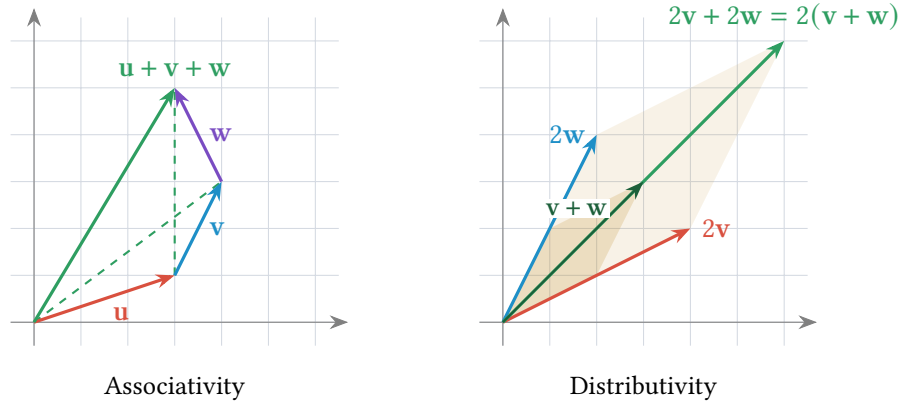


Figure 1.7: Left:  $(u + v) + w$  and  $u + (v + w)$  (dashed) reach the same point. Right: doubling the parallelogram on  $v = (2, 1)$  and  $w = (1, 2)$  doubles its diagonal.

### 1.2.5 Consequences of the rules

**Proposition 1.2.8.** Let  $u, v, w, x \in \mathbb{R}^n$  and  $a \in \mathbb{R}$ . Then:

- (a) (cancellation) if  $v + x = w + x$ , then  $v = w$ ;
- (b) (the zero vector is unique) if  $v + z = v$  for some  $z \in \mathbb{R}^n$ , then  $z = 0$ ;
- (c) (negatives are unique) if  $v + u = 0$ , then  $u = -v$ ;
- (d)  $0v = 0$ ;
- (e)  $a0 = 0$ ;
- (f)  $(-1)v = -v$ ;
- (g) if  $av = 0$ , then  $a = 0$  or  $v = 0$ .

Several of these are easy to see entry by entry in  $\mathbb{R}^n$ . The point of the proof below is that it uses only rules (1) to (8) of Theorem 1.2.7, so the same statements hold in every vector space. In particular, in (c) and (f) the symbol  $-v$  means only “the vector of rule (4)”, and (c) shows that there is just one such vector.

*Proof.* (a) Add  $-x$  to both sides of  $v + x = w + x$ :

$$\begin{aligned}
 (v + x) + (-x) &= (w + x) + (-x) \\
 v + (x + (-x)) &= w + (x + (-x)) && \text{rule (2)} \\
 v + 0 &= w + 0 && \text{rule (4)} \\
 v &= w && \text{rule (3).}
 \end{aligned}$$

(b) By rule (3),  $v + z = v = v + 0$ . By rule (1) this reads  $z + v = 0 + v$ , and cancellation (a) gives  $z = 0$ .

(c) By rule (4),  $v + u = 0 = v + (-v)$ . By rule (1) and cancellation,  $u = -v$ .

(d) Since  $0 = 0 + 0$  in  $\mathbb{R}$ , rule (8) gives

$$0v = (0 + 0)v = 0v + 0v.$$

Add  $-(0v)$  to both sides. The left side becomes  $0$  by rule (4). The right side becomes  $(0v + 0v) + (-(0v)) =$

$0\mathbf{v} + (0\mathbf{v} + (-(0\mathbf{v}))) = 0\mathbf{v} + \mathbf{0} = 0\mathbf{v}$ , by rules (2), (4) and (3). Hence  $\mathbf{0} = 0\mathbf{v}$ . (This is the proof given in Lecture 2.)

(e) In the same way, by rules (3) and (7),  $a\mathbf{0} = a(\mathbf{0} + \mathbf{0}) = a\mathbf{0} + a\mathbf{0}$ , and adding  $-(a\mathbf{0})$  to both sides gives  $\mathbf{0} = a\mathbf{0}$ .

(f) By rules (5) and (8) and part (d),

$$\mathbf{v} + (-1)\mathbf{v} = 1\mathbf{v} + (-1)\mathbf{v} = (1 + (-1))\mathbf{v} = 0\mathbf{v} = \mathbf{0},$$

so  $(-1)\mathbf{v}$  is a negative of  $\mathbf{v}$ , and by (c) it equals  $-\mathbf{v}$ .

(g) Suppose  $a\mathbf{v} = \mathbf{0}$  and  $a \neq 0$ . Then  $a$  has an inverse  $a^{-1}$  in  $\mathbb{R}$ , and by rules (5) and (6) and part (e),

$$\mathbf{v} = 1\mathbf{v} = (a^{-1}a)\mathbf{v} = a^{-1}(a\mathbf{v}) = a^{-1}\mathbf{0} = \mathbf{0}.$$

So either  $a = 0$  or  $\mathbf{v} = \mathbf{0}$ . □

**Why bother?** In  $\mathbb{R}^n$  every part of Proposition 1.2.8 can be checked in one line by looking at entries, so the proofs above may look like a lot of effort for something obvious. They are foundational all the same. In a general vector space (Chapter 4) the eight rules of Theorem 1.2.7 are all we are given: there are no entries to look at. Even a fact as simple as  $0\mathbf{v} = \mathbf{0}$  must then be earned from the rules, and much of what follows, from subspaces and linear maps to the dimension formula, rests on small steps like these. Proofs of this kind are demanding the first time you meet them; it pays to read them slowly, one rule per step.

The rules also let us solve equations in which the unknown is a vector, much as we solve equations for numbers.

**Example 1.2.9.** Find all  $\mathbf{x} \in \mathbb{R}^2$  with  $2\mathbf{x} + \begin{bmatrix} 1 \\ -3 \end{bmatrix} = \begin{bmatrix} 7 \\ 5 \end{bmatrix}$ .

*Solution.* Suppose  $\mathbf{x}$  is a solution. Adding  $-(1, -3)$  to both sides and using rules (2), (4) and (3) gives

$$2\mathbf{x} = \begin{bmatrix} 7 \\ 5 \end{bmatrix} - \begin{bmatrix} 1 \\ -3 \end{bmatrix} = \begin{bmatrix} 7-1 \\ 5-(-3) \end{bmatrix} = \begin{bmatrix} 6 \\ 8 \end{bmatrix}.$$

Multiplying both sides by  $\frac{1}{2}$  and using rules (6) and (5),  $\frac{1}{2}(2\mathbf{x}) = (\frac{1}{2} \cdot 2)\mathbf{x} = 1\mathbf{x} = \mathbf{x}$ , so

$$\mathbf{x} = \frac{1}{2} \begin{bmatrix} 6 \\ 8 \end{bmatrix} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}.$$

Conversely,  $2(3, 4) + (1, -3) = (6, 8) + (1, -3) = (7, 5)$ , so  $\mathbf{x} = (3, 4)$  is the unique solution.

## 1.2.6 Geometry with vectors

Fix an origin  $O$ . Every point  $P$  of the plane (or of space) then has a *position vector*  $\mathbf{p}$ , the arrow from  $O$  to  $P$ , whose components are the coordinates of  $P$ . As in Example 1.2.6, the step from  $P$  to  $Q$  is  $\mathbf{q} - \mathbf{p}$ .

**Proposition 1.2.10** (Midpoint). The midpoint  $M$  of the segment from  $P$  to  $Q$  has position vector

$$\mathbf{m} = \mathbf{p} + \frac{1}{2}(\mathbf{q} - \mathbf{p}) = \frac{1}{2}\mathbf{p} + \frac{1}{2}\mathbf{q}.$$

*Proof.* The midpoint is reached from  $P$  by taking half of the step from  $P$  to  $Q$ , so  $\mathbf{m} = \mathbf{p} + \frac{1}{2}(\mathbf{q} - \mathbf{p})$ . By rule (7), Definition 1.2.5 and Proposition 1.2.8(f) and rule (6),  $\frac{1}{2}(\mathbf{q} - \mathbf{p}) = \frac{1}{2}\mathbf{q} + \frac{1}{2}(-1)\mathbf{p} = \frac{1}{2}\mathbf{q} - \frac{1}{2}\mathbf{p}$ . Using rules (1) to (4), (8) and (5),

$$\mathbf{m} = \mathbf{p} - \frac{1}{2}\mathbf{p} + \frac{1}{2}\mathbf{q} = (1 - \frac{1}{2})\mathbf{p} + \frac{1}{2}\mathbf{q} = \frac{1}{2}\mathbf{p} + \frac{1}{2}\mathbf{q}. \quad \square$$

So the midpoint is the average of the two position vectors: the midpoint of  $(1, 3)$  and  $(5, 1)$  is  $(3, 2)$ .

**Theorem 1.2.11.** The diagonals of a parallelogram bisect each other: they meet at a point that is the midpoint of each.

*Proof.* Let the parallelogram have corners  $A, B, C, D$  in order, with position vectors  $\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}$ . Opposite sides  $AB$  and  $DC$  are the same arrow, so  $\mathbf{b} - \mathbf{a} = \mathbf{c} - \mathbf{d}$ . Adding  $\mathbf{a} + \mathbf{d}$  to both sides and using the rules gives  $\mathbf{b} + \mathbf{d} = \mathbf{a} + \mathbf{c}$ . By Proposition 1.2.10, the midpoint of the diagonal  $AC$  is  $\frac{1}{2}(\mathbf{a} + \mathbf{c})$ , and the midpoint of the diagonal  $BD$  is  $\frac{1}{2}(\mathbf{b} + \mathbf{d})$ . These are equal, so the two diagonals share their midpoint, which is therefore the point where they cross, and it cuts each of them in half.  $\square$

In the lecture we placed the corners at  $\mathbf{0}, \mathbf{v}, \mathbf{v} + \mathbf{w}$  and  $\mathbf{w}$ . Then the midpoint of the diagonal from  $\mathbf{0}$  to  $\mathbf{v} + \mathbf{w}$  is  $\frac{1}{2}(\mathbf{v} + \mathbf{w})$ , the midpoint of the diagonal from  $\mathbf{v}$  to  $\mathbf{w}$  is  $\frac{1}{2}\mathbf{v} + \frac{1}{2}\mathbf{w}$ , and rule (7) says these are equal (Figure 1.8). Notice that the proof uses no angles, lengths or congruent triangles, only the rules of vector algebra.

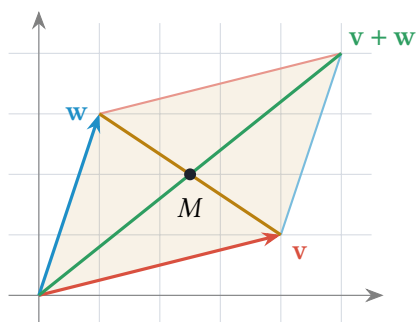


Figure 1.8: The diagonals of the parallelogram on  $\mathbf{v}$  and  $\mathbf{w}$  meet at the point  $M$  with position vector  $\frac{1}{2}(\mathbf{v} + \mathbf{w})$ , the midpoint of each.

### Check yourself

Solve  $3\mathbf{x} - \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 4 \\ -7 \\ 6 \end{bmatrix}$  for  $\mathbf{x} \in \mathbb{R}^3$ , and name the rule of Theorem 1.2.7 behind each step.

## 1.3 Linear combinations and span

### Lecture 3

Section 1.2 set up the two operations on vectors, addition and scalar multiplication, and the rules they obey. This section combines them. Scaling some vectors and adding the results produces a *linear combination*, and the set of everything that can be produced this way is the *span*. The question “which vectors can be built from given ones?” was the first of the four questions in Section 1.1; spans are the precise language for it, and they lead directly to systems of linear equations (Chapter 2), to linear independence (Section 2.6) and to subspaces (Chapter 4).

### 1.3.1 Linear combinations

**Definition 1.3.1** (Linear combination). Let  $\mathbf{v}_1, \dots, \mathbf{v}_k$  be vectors in  $\mathbb{R}^n$  and let  $c_1, \dots, c_k$  be scalars. The vector

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k$$

is called a *linear combination* of  $\mathbf{v}_1, \dots, \mathbf{v}_k$  with *weights* (or *coefficients*)  $c_1, \dots, c_k$ .

The weights may be any real numbers: positive, negative, fractional or zero. Since  $\mathbb{R}^n$  is closed under addition and scalar multiplication, a linear combination of vectors in  $\mathbb{R}^n$  is again a vector in  $\mathbb{R}^n$ . With two vectors we usually write the weights as  $a$  and  $b$ , and the combination as  $a\mathbf{v} + b\mathbf{w}$ .

**Example 1.3.2.** Let  $\mathbf{v} = (2, 1)$  and  $\mathbf{w} = (-1, 1)$ . Compute  $3\mathbf{v} + 2\mathbf{w}$ .

*Solution.* Scale first, entry by entry, then add matching entries:

$$3 \begin{bmatrix} 2 \\ 1 \end{bmatrix} + 2 \begin{bmatrix} -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 6 \\ 3 \end{bmatrix} + \begin{bmatrix} -2 \\ 2 \end{bmatrix} = \begin{bmatrix} 6 + (-2) \\ 3 + 2 \end{bmatrix} = \begin{bmatrix} 4 \\ 5 \end{bmatrix}.$$

Geometrically, walk three copies of  $\mathbf{v}$  tip to tail, then two copies of  $\mathbf{w}$ ; the walk ends at the point  $(4, 5)$  (Figure 1.9).

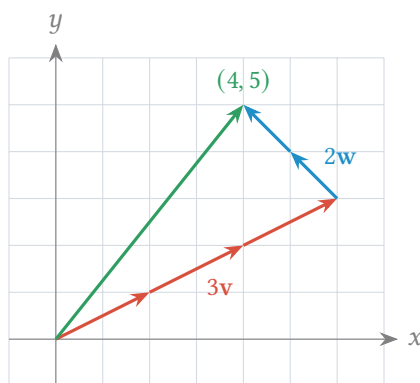


Figure 1.9: Example 1.3.2: three copies of  $\mathbf{v} = (2, 1)$  followed by two copies of  $\mathbf{w} = (-1, 1)$  reach  $3\mathbf{v} + 2\mathbf{w} = (4, 5)$ .

**Example 1.3.3.** In  $\mathbb{R}^3$ , with  $\mathbf{v}_1 = (1, 0, 1)$ ,  $\mathbf{v}_2 = (0, 1, 1)$  and  $\mathbf{v}_3 = (1, 1, 0)$ ,

$$2\mathbf{v}_1 - \mathbf{v}_2 + 3\mathbf{v}_3 = \begin{bmatrix} 2 \\ 0 \\ 2 \end{bmatrix} + \begin{bmatrix} 0 \\ -1 \\ -1 \end{bmatrix} + \begin{bmatrix} 3 \\ 3 \\ 0 \end{bmatrix} = \begin{bmatrix} 5 \\ 2 \\ 1 \end{bmatrix}.$$

**Remark 1.3.4.** Two combinations are always available. Taking every weight equal to 0 gives  $0\mathbf{v}_1 + \cdots + 0\mathbf{v}_k = \mathbf{0}$ , and taking  $c_i = 1$  and every other weight 0 gives  $\mathbf{v}_i$ . So  $\mathbf{0}$  and each of  $\mathbf{v}_1, \dots, \mathbf{v}_k$  are linear combinations of  $\mathbf{v}_1, \dots, \mathbf{v}_k$ .

### 1.3.2 The span

Fix  $\mathbf{v}$  and  $\mathbf{w}$  and let the weights vary. Holding  $b = 0$ , the combinations  $a\mathbf{v}$  fill the line through the origin in the direction of  $\mathbf{v}$ ; holding  $a = 0$ , the combinations  $b\mathbf{w}$  fill the line along  $\mathbf{w}$ . Letting both weights take whole-number values produces a slanted grid, built from  $\mathbf{v}$  and  $\mathbf{w}$  in the same way that the usual grid is built from the coordinate axes (Figure 1.10). The weights  $a$  and  $b$  act as coordinates measured along  $\mathbf{v}$  and  $\mathbf{w}$ .

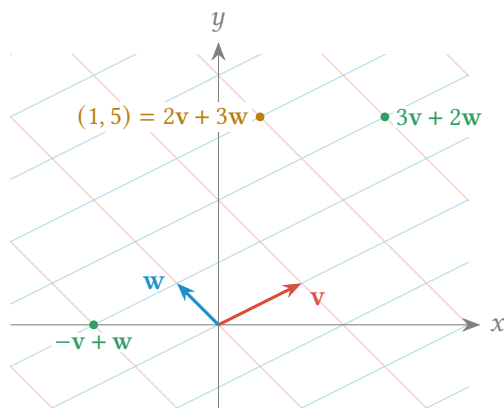


Figure 1.10: The slanted grid of  $\mathbf{v} = (2, 1)$  and  $\mathbf{w} = (-1, 1)$ : coral lines have constant  $a$ , blue lines constant  $b$ . Each crossing is a combination  $a\mathbf{v} + b\mathbf{w}$  with whole-number weights.

**Definition 1.3.5** (Span). The *span* of vectors  $\mathbf{v}_1, \dots, \mathbf{v}_k$  in  $\mathbb{R}^n$  is the set of all their linear combinations,

$$\text{Span}\{\mathbf{v}_1, \dots, \mathbf{v}_k\} = \{c_1\mathbf{v}_1 + \dots + c_k\mathbf{v}_k : c_1, \dots, c_k \in \mathbb{R}\}.$$

A vector  $\mathbf{b}$  is *in the span* if  $\mathbf{b} = c_1\mathbf{v}_1 + \dots + c_k\mathbf{v}_k$  for some scalars  $c_1, \dots, c_k$ . By convention the span of the empty set is  $\{\mathbf{0}\}$ .

A linear combination is a single vector; the span is a set, usually infinite, containing every combination at once. The convention for the empty set matches Remark 1.3.4: every span contains  $\mathbf{0}$ .

**Example 1.3.6** (The span of one vector). A combination of a single vector  $\mathbf{v}$  is a multiple  $c\mathbf{v}$ , so  $\text{Span}\{\mathbf{v}\} = \{c\mathbf{v} : c \in \mathbb{R}\}$ .

- If  $\mathbf{v} \neq \mathbf{0}$ , this is the line through the origin in the direction of  $\mathbf{v}$ : positive multiples on one side, negative multiples on the other, and  $0\mathbf{v} = \mathbf{0}$  at the origin. For  $\mathbf{v} = (2, 1)$ , the vector  $(6, 3) = 3\mathbf{v}$  lies in the span, while  $(1, 1)$  does not: every multiple  $(2c, c)$  has first entry twice its second.
- If  $\mathbf{v} = \mathbf{0}$ , every multiple is  $\mathbf{0}$ , so  $\text{Span}\{\mathbf{0}\} = \{\mathbf{0}\}$ , a single point.

Since every span contains  $\mathbf{0}$ , a line in  $\mathbb{R}^2$  that does not pass through the origin, such as  $y = x + 1$ , is not the span of any set of vectors.

### 1.3.3 Spans in the plane

**Example 1.3.7.** Let  $\mathbf{v} = (2, 1)$  and  $\mathbf{w} = (-1, 1)$ . We show that  $\text{Span}\{\mathbf{v}, \mathbf{w}\} = \mathbb{R}^2$ .

Let  $(x, y)$  be any vector in  $\mathbb{R}^2$ . We look for  $a, b$  with  $a\begin{bmatrix} 2 \\ 1 \end{bmatrix} + b\begin{bmatrix} -1 \\ 1 \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix}$ . Comparing entries gives the system

$$\begin{aligned} 2a - b &= x, \\ a + b &= y. \end{aligned}$$

Adding the equations gives  $3a = x + y$ , so  $a = \frac{x+y}{3}$ , and then  $b = y - a = \frac{2y-x}{3}$ . Check:

$$2a - b = \frac{2x + 2y - (2y - x)}{3} = x, \quad a + b = \frac{x + y + 2y - x}{3} = y.$$

So every  $(x, y)$  is a combination of  $\mathbf{v}$  and  $\mathbf{w}$ . For the target  $(1, 5)$  of Example 1.1.5 the formulas give  $a = 2$ ,  $b = 3$ , as found there.

If instead  $\mathbf{w} = (4, 2) = 2\mathbf{v}$ , then  $a\mathbf{v} + b\mathbf{w} = (a + 2b)\mathbf{v}$  is always a multiple of  $\mathbf{v}$ , and the span is only the line through  $\mathbf{v}$ . Two vectors are called *parallel* when one of them is a scalar multiple of the other. (By this definition  $\mathbf{0}$  is parallel to every vector, since  $\mathbf{0} = 0\mathbf{w}$ .)

**Theorem 1.3.8.** Let  $\mathbf{v} = (p, q)$  and  $\mathbf{w} = (r, s)$  be vectors in  $\mathbb{R}^2$ . The following are equivalent:

- (a)  $\text{Span}\{\mathbf{v}, \mathbf{w}\} = \mathbb{R}^2$ ;
- (b)  $\mathbf{v}$  and  $\mathbf{w}$  are not parallel;
- (c)  $ps - qr \neq 0$ .

If  $\mathbf{v}$  and  $\mathbf{w}$  are parallel, their span is a line through the origin, or  $\{\mathbf{0}\}$  when  $\mathbf{v} = \mathbf{w} = \mathbf{0}$ .

*Proof.* Write  $D = ps - qr$ .

(c)  $\Rightarrow$  (a). Suppose  $D \neq 0$  and let  $(x, y) \in \mathbb{R}^2$ . Put

$$a = \frac{sx - ry}{D}, \quad b = \frac{py - qx}{D}.$$

Then the two entries of  $a\mathbf{v} + b\mathbf{w}$  are

$$\frac{p(sx - ry) + r(py - qx)}{D} = \frac{(ps - qr)x}{D} = x, \quad \frac{q(sx - ry) + s(py - qx)}{D} = \frac{(ps - qr)y}{D} = y,$$

so  $(x, y) \in \text{Span}\{\mathbf{v}, \mathbf{w}\}$ .

(a)  $\Rightarrow$  (b). We prove the contrapositive. Suppose  $\mathbf{v}$  and  $\mathbf{w}$  are parallel. If both are  $\mathbf{0}$  the span is  $\{\mathbf{0}\} \neq \mathbb{R}^2$ . Otherwise one of them, call it  $\mathbf{u} = (u_1, u_2)$ , is nonzero and the other is a multiple of it, say  $\lambda\mathbf{u}$ . (If one is  $\mathbf{0}$ , it is 0 times the other. If  $\mathbf{v} = c\mathbf{w}$  with both nonzero, then  $c \neq 0$  and  $\mathbf{w} = c^{-1}\mathbf{v}$ , so either vector can play the role of  $\mathbf{u}$ .) Every combination is then a multiple of  $\mathbf{u}$ , since  $\alpha\mathbf{u} + \beta\lambda\mathbf{u} = (\alpha + \beta\lambda)\mathbf{u}$ , and every multiple of  $\mathbf{u}$  is such a combination, so the span is the line  $\text{Span}\{\mathbf{u}\}$ . This line is not all of  $\mathbb{R}^2$ : the vector  $(-u_2, u_1)$  is not on it. Indeed, if  $(-u_2, u_1) = c(u_1, u_2)$  then  $-u_2 = cu_1$  and  $u_1 = cu_2$ , so  $-u_2 = c^2u_2$  and  $u_1 = -c^2u_1$ ; since  $1 + c^2 > 0$  this forces  $u_1 = u_2 = 0$ , contradicting  $\mathbf{u} \neq \mathbf{0}$ . This also proves the last sentence of the theorem.

(b)  $\Rightarrow$  (c). Again the contrapositive: suppose  $D = 0$ ; we show the vectors are parallel. If  $\mathbf{v} = \mathbf{0}$  then  $\mathbf{v} = 0\mathbf{w}$ . Otherwise  $p \neq 0$  or  $q \neq 0$ . If  $p \neq 0$ , then  $s = qr/p$  and  $\mathbf{w} = (r, qr/p) = \frac{r}{p}(p, q) = \frac{r}{p}\mathbf{v}$ . If  $p = 0$ , then  $q \neq 0$  and  $D = 0$  gives  $qr = 0$ , so  $r = 0$  and  $\mathbf{w} = (0, s) = \frac{s}{q}\mathbf{v}$ . In every case  $\mathbf{w}$  is a multiple of  $\mathbf{v}$ .  $\square$

The number  $ps - qr$  is the *determinant* of the matrix with columns  $\mathbf{v}$  and  $\mathbf{w}$ . In Chapter 7 we show that its absolute value is the area of the parallelogram with sides  $\mathbf{v}$  and  $\mathbf{w}$ , which is zero exactly when the two vectors lie on one line.

### 1.3.4 Spans in $\mathbb{R}^3$

In  $\mathbb{R}^3$  spans come in more shapes. One nonzero vector spans a line through the origin. Two non-parallel vectors  $\mathbf{v}, \mathbf{w}$  span a *plane through the origin*: all combinations  $a\mathbf{v} + b\mathbf{w}$  lie in the flat sheet containing both arrows, and the slanted grid of  $\mathbf{v}$  and  $\mathbf{w}$  covers that sheet (Figure 1.11).

**Example 1.3.9.** Let  $\mathbf{e}_1 = (1, 0, 0)$ ,  $\mathbf{e}_2 = (0, 1, 0)$ ,  $\mathbf{e}_3 = (0, 0, 1)$ . Then

$$\text{Span}\{\mathbf{e}_1, \mathbf{e}_2\} = \{a\mathbf{e}_1 + b\mathbf{e}_2\} = \{(a, b, 0) : a, b \in \mathbb{R}\},$$

the horizontal plane of vectors with third entry 0. In particular  $\mathbf{e}_3 \notin \text{Span}\{\mathbf{e}_1, \mathbf{e}_2\}$ , because its third entry is 1. On the other hand  $(1, 1, 0) = \mathbf{e}_1 + \mathbf{e}_2$  is in the span, and (by Theorem 1.3.11 below)  $\text{Span}\{\mathbf{e}_1, \mathbf{e}_2, (1, 1, 0)\}$  is the same plane. Finally  $\text{Span}\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\} = \mathbb{R}^3$ , since  $(x, y, z) = x\mathbf{e}_1 + y\mathbf{e}_2 + z\mathbf{e}_3$ .

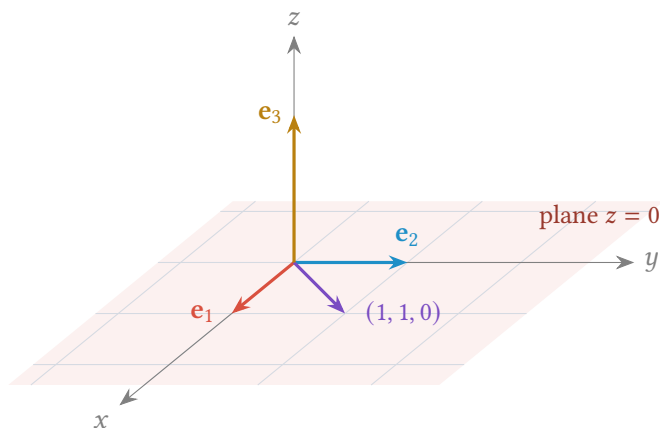


Figure 1.11: Example 1.3.9:  $\mathbf{e}_1$  and  $\mathbf{e}_2$  span the horizontal plane  $z = 0$ , which also contains  $(1, 1, 0)$ . The vector  $\mathbf{e}_3$  points out of the plane.

Adding a third vector  $\mathbf{u}$  to two non-parallel vectors  $\mathbf{v}, \mathbf{w}$  in  $\mathbb{R}^3$  either changes nothing (when  $\mathbf{u}$  lies in their plane) or gives all of  $\mathbb{R}^3$  (when it does not). In the second case one reaches a point by moving along  $\mathbf{u}$  until level with the point and then within a copy of the plane; Chapter 2 turns this picture into a proof. As a result, every span in  $\mathbb{R}^3$  is one of four sets:  $\{\mathbf{0}\}$ , a line through the origin, a plane through the origin, or  $\mathbb{R}^3$ . Which one occurs depends not on the number of vectors but on how many independent directions they contribute; making this precise is the business of linear independence (Section 2.6) and dimension (Chapter 4).

### 1.3.5 Properties of the span

**Theorem 1.3.10.** Let  $\mathbf{v}_1, \dots, \mathbf{v}_k \in \mathbb{R}^n$  and  $S = \text{Span}\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ .

- (a)  $\mathbf{0} \in S$  and  $\mathbf{v}_i \in S$  for every  $i$ .
- (b) If  $\mathbf{x}, \mathbf{y} \in S$  then  $\mathbf{x} + \mathbf{y} \in S$  ( $S$  is closed under addition).
- (c) If  $\mathbf{x} \in S$  and  $t \in \mathbb{R}$  then  $t\mathbf{x} \in S$  ( $S$  is closed under scalar multiplication).
- (d)  $S$  is the smallest such set: if  $W \subseteq \mathbb{R}^n$  contains  $\mathbf{v}_1, \dots, \mathbf{v}_k$  and is closed under addition and scalar multiplication, then  $S \subseteq W$ .

*Proof.* (a) is Remark 1.3.4.

(b) Write  $\mathbf{x} = c_1\mathbf{v}_1 + \dots + c_k\mathbf{v}_k$  and  $\mathbf{y} = d_1\mathbf{v}_1 + \dots + d_k\mathbf{v}_k$ . Using commutativity and associativity of addition and the distributive law  $c\mathbf{v} + d\mathbf{v} = (c + d)\mathbf{v}$  from Section 1.2,

$$\mathbf{x} + \mathbf{y} = (c_1 + d_1)\mathbf{v}_1 + \dots + (c_k + d_k)\mathbf{v}_k,$$

which is a linear combination of  $\mathbf{v}_1, \dots, \mathbf{v}_k$ .

(c) With  $\mathbf{x}$  as above, the rules  $t(\mathbf{a} + \mathbf{b}) = t\mathbf{a} + t\mathbf{b}$  and  $t(c\mathbf{v}) = (tc)\mathbf{v}$  give  $t\mathbf{x} = (tc_1)\mathbf{v}_1 + \dots + (tc_k)\mathbf{v}_k \in S$ .

(d) Let  $\mathbf{x} = c_1\mathbf{v}_1 + \dots + c_k\mathbf{v}_k \in S$ . Each  $\mathbf{v}_i$  lies in  $W$ , so each  $c_i\mathbf{v}_i$  lies in  $W$  by closure under scalar multiplication. Then  $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 \in W$  by closure under addition, and adding the remaining terms one at a time (formally, by induction on the number of terms) shows  $\mathbf{x} \in W$ .  $\square$

A subset of  $\mathbb{R}^n$  that contains  $\mathbf{0}$  and is closed under addition and scalar multiplication is called a *subspace*;

these are studied from Section 4.1 onwards. Theorem 1.3.10 says that a span is a subspace, and the smallest one containing the given vectors.

**Theorem 1.3.11.** Let  $\mathbf{v}_1, \dots, \mathbf{v}_k, \mathbf{u} \in \mathbb{R}^n$ . Then

$$\text{Span}\{\mathbf{v}_1, \dots, \mathbf{v}_k, \mathbf{u}\} = \text{Span}\{\mathbf{v}_1, \dots, \mathbf{v}_k\} \quad \text{if and only if} \quad \mathbf{u} \in \text{Span}\{\mathbf{v}_1, \dots, \mathbf{v}_k\}.$$

*Proof.* Write  $S = \text{Span}\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$  and  $S' = \text{Span}\{\mathbf{v}_1, \dots, \mathbf{v}_k, \mathbf{u}\}$ . Always  $S \subseteq S'$ : a combination  $c_1\mathbf{v}_1 + \dots + c_k\mathbf{v}_k$  equals  $c_1\mathbf{v}_1 + \dots + c_k\mathbf{v}_k + 0\mathbf{u}$ .

If  $S' = S$ , then  $\mathbf{u} \in S' = S$  by Theorem 1.3.10(a).

Conversely, suppose  $\mathbf{u} = d_1\mathbf{v}_1 + \dots + d_k\mathbf{v}_k$ . For any element of  $S'$ ,

$$c_1\mathbf{v}_1 + \dots + c_k\mathbf{v}_k + c\mathbf{u} = (c_1 + c d_1)\mathbf{v}_1 + \dots + (c_k + c d_k)\mathbf{v}_k \in S,$$

so  $S' \subseteq S$ , and therefore  $S' = S$ . □

**Example 1.3.12.** For any vectors  $\mathbf{v}, \mathbf{w}$  and any scalar  $\lambda$ ,  $\text{Span}\{\mathbf{v}, \mathbf{w}\} = \text{Span}\{\mathbf{v}, \mathbf{w} + \lambda\mathbf{v}\}$ .

*Proof.* Since  $\mathbf{w} + \lambda\mathbf{v} \in \text{Span}\{\mathbf{v}, \mathbf{w}\}$ , Theorem 1.3.11 gives  $\text{Span}\{\mathbf{v}, \mathbf{w}, \mathbf{w} + \lambda\mathbf{v}\} = \text{Span}\{\mathbf{v}, \mathbf{w}\}$ . Since also  $\mathbf{w} = (\mathbf{w} + \lambda\mathbf{v}) - \lambda\mathbf{v} \in \text{Span}\{\mathbf{v}, \mathbf{w} + \lambda\mathbf{v}\}$ , the same theorem gives  $\text{Span}\{\mathbf{v}, \mathbf{w} + \lambda\mathbf{v}, \mathbf{w}\} = \text{Span}\{\mathbf{v}, \mathbf{w} + \lambda\mathbf{v}\}$ . The two left-hand sides are spans of the same three vectors, so the two spans are equal. Replacing one vector by itself plus a multiple of another is exactly the move made by row operations in Gaussian elimination (Section 2.3).

### 1.3.6 Is a vector in the span?

**Example 1.3.13.** Let  $\mathbf{v} = (1, 0, 1)$  and  $\mathbf{w} = (0, 1, 1)$ . Decide whether  $\mathbf{b} = (1, 2, 3)$  and  $\mathbf{b}' = (1, 2, 4)$  lie in  $\text{Span}\{\mathbf{v}, \mathbf{w}\}$ .

*Solution.* We need  $x, y$  with  $x\mathbf{v} + y\mathbf{w} = \mathbf{b}$ , that is,

$$x \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + y \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \quad \Longleftrightarrow \quad \begin{array}{l} x = 1, \\ y = 2, \\ x + y = 3. \end{array}$$

The first two equations force  $x = 1, y = 2$ , and then  $x + y = 3$  holds, so  $\mathbf{b} = \mathbf{v} + 2\mathbf{w}$  is in the span. For  $\mathbf{b}'$  the first two equations again force  $x = 1, y = 2$ , but the third requires  $x + y = 4$ . No weights work, so  $\mathbf{b}' \notin \text{Span}\{\mathbf{v}, \mathbf{w}\}$ : it lies off the plane spanned by  $\mathbf{v}$  and  $\mathbf{w}$ .

In general,  $\mathbf{b} \in \text{Span}\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$  exactly when the vector equation  $x_1\mathbf{v}_1 + \dots + x_k\mathbf{v}_k = \mathbf{b}$  has a solution. By the definition of equality in  $\mathbb{R}^n$  this vector equation is a system of  $n$  linear equations, one for each coordinate, in the  $k$  unknowns  $x_1, \dots, x_k$ . Deciding membership in a span is therefore the same problem as deciding whether a linear system is consistent. Chapter 2 develops Gaussian elimination, which settles it for any number of vectors in any dimension, and Section 3.1 writes the system compactly as  $A\mathbf{x} = \mathbf{b}$ , where  $A$  has columns  $\mathbf{v}_1, \dots, \mathbf{v}_k$ .

#### Check yourself

(a) Is  $(2, 5)$  in  $\text{Span}\{(1, 2), (3, 6)\}$ ? (b) Is  $\text{Span}\{(1, 2), (3, 5)\} = \mathbb{R}^2$ ?

## 1.4 Dot product, length and angle

### Lecture 4

So far vectors are things we add, scale and combine. Two of the most natural questions about arrows have no answer yet: *how long is a vector*, and *what is the angle between two vectors*? This section answers both with a single new operation, the dot product. It turns two vectors into one number, and from that number we read off lengths, distances, angles and right angles. In  $\mathbb{R}^2$  and  $\mathbb{R}^3$  these agree with the lengths and angles of school geometry; in  $\mathbb{R}^n$  for larger  $n$ , where there is nothing to draw, they *define* length and angle. The step that makes this possible is the Cauchy–Schwarz inequality.

### 1.4.1 Length

In the plane, the arrow of  $\mathbf{v} = (3, 4)$  is the hypotenuse of a right triangle with legs 3 and 4 (Figure 1.12, left). By the theorem of Pythagoras its length is  $\sqrt{3^2 + 4^2} = \sqrt{25} = 5$ . In space we use Pythagoras twice. The vector  $\mathbf{v} = (1, 2, 2)$  has the “shadow”  $(1, 2, 0)$  in the floor plane, of length  $\sqrt{1^2 + 2^2} = \sqrt{5}$ . The arrow of  $\mathbf{v}$  is the hypotenuse of an upright right triangle with legs  $\sqrt{5}$  and 2, so its length is  $\sqrt{5 + 4} = 3$  (Figure 1.12, right). In both cases we square the components, add, and take the square root. In  $\mathbb{R}^n$  we turn this pattern into a definition.

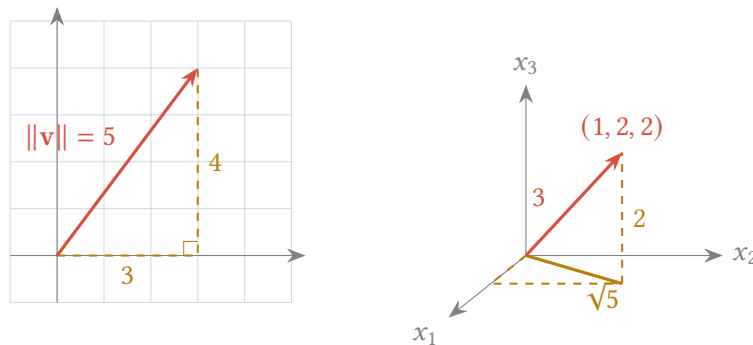


Figure 1.12: Lengths from Pythagoras:  $\|(3, 4)\| = 5$  in the plane, and  $\|(1, 2, 2)\| = \sqrt{5 + 4} = 3$  in space.

**Definition 1.4.1 (Norm).** The *norm* (or *length*) of a vector  $\mathbf{v} = (v_1, \dots, v_n) \in \mathbb{R}^n$  is the number

$$\|\mathbf{v}\| = \sqrt{v_1^2 + v_2^2 + \dots + v_n^2}.$$

Since a sum of squares is never negative,  $\|\mathbf{v}\| \geq 0$ ; and since it is zero only when every square is zero,  $\|\mathbf{v}\| = 0$  exactly when  $\mathbf{v} = \mathbf{0}$ .

**Example 1.4.2.**

$$\|(3, 4)\| = 5, \quad \|(1, 2, 2)\| = 3, \quad \|(1, 1, 1)\| = \sqrt{4} = 2, \quad \|(2, -1, 0, 2, 0)\| = \sqrt{9} = 3.$$

The norm of a vector in  $\mathbb{R}^5$  is defined by the same formula, even though there is no triangle to draw.

### 1.4.2 The dot product

The sum of squares inside the norm is the special case  $\mathbf{w} = \mathbf{v}$  of a recipe that takes two vectors.

**Definition 1.4.3** (Dot product). The *dot product* of  $\mathbf{v} = (v_1, \dots, v_n)$  and  $\mathbf{w} = (w_1, \dots, w_n)$  in  $\mathbb{R}^n$  is the number

$$\mathbf{v} \cdot \mathbf{w} = v_1 w_1 + v_2 w_2 + \cdots + v_n w_n = \sum_{i=1}^n v_i w_i.$$

We multiply matching entries and add the products. The result is a scalar, not a vector, which is why the dot product is also called the *scalar product*. It is defined only for two vectors in the same  $\mathbb{R}^n$ .

**Example 1.4.4.** For  $\mathbf{v} = (2, 1)$  and  $\mathbf{w} = (1, 3)$ , coordinate by coordinate:  $2 \cdot 1 = 2$  and  $1 \cdot 3 = 3$ , so  $\mathbf{v} \cdot \mathbf{w} = 2 + 3 = 5$ . In  $\mathbb{R}^4$ ,  $(1, -2, 0, 3) \cdot (4, 1, 7, -1) = 4 - 2 + 0 - 3 = -1$ . A dot product may be positive, negative or zero.

Taking  $\mathbf{w} = \mathbf{v}$  gives  $\mathbf{v} \cdot \mathbf{v} = v_1^2 + \cdots + v_n^2$ , so

$$\|\mathbf{v}\|^2 = \mathbf{v} \cdot \mathbf{v}, \quad \|\mathbf{v}\| = \sqrt{\mathbf{v} \cdot \mathbf{v}}. \quad (1.1)$$

For example  $(2, 1) \cdot (2, 1) = 4 + 1 = 5$ , so  $\|(2, 1)\| = \sqrt{5}$ .

**Definition 1.4.5** (Distance). The *distance* between  $\mathbf{v}$  and  $\mathbf{w}$  in  $\mathbb{R}^n$  is  $d(\mathbf{v}, \mathbf{w}) = \|\mathbf{w} - \mathbf{v}\|$ .

Drawn from the origin,  $\mathbf{w} - \mathbf{v}$  is the arrow from the tip of  $\mathbf{v}$  to the tip of  $\mathbf{w}$ , so  $d(\mathbf{v}, \mathbf{w})$  is the distance between the two tips, viewed as points. For  $\mathbf{v} = (2, 1)$  and  $\mathbf{w} = (1, 3)$ ,  $\mathbf{w} - \mathbf{v} = (-1, 2)$  and  $d(\mathbf{v}, \mathbf{w}) = \sqrt{1 + 4} = \sqrt{5}$ . Since  $\|-\mathbf{x}\| = \|\mathbf{x}\|$  (Proposition 1.4.7 below),  $d(\mathbf{v}, \mathbf{w}) = d(\mathbf{w}, \mathbf{v})$ .

### 1.4.3 The rules of the dot product

**Theorem 1.4.6** (Properties of the dot product). For all  $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^n$  and every scalar  $c$ :

- (a) *symmetry*:  $\mathbf{v} \cdot \mathbf{w} = \mathbf{w} \cdot \mathbf{v}$ ;
- (b) *additivity*:  $\mathbf{u} \cdot (\mathbf{v} + \mathbf{w}) = \mathbf{u} \cdot \mathbf{v} + \mathbf{u} \cdot \mathbf{w}$ ;
- (c) *homogeneity*:  $(c\mathbf{v}) \cdot \mathbf{w} = c(\mathbf{v} \cdot \mathbf{w})$ ;
- (d) *positivity*:  $\mathbf{v} \cdot \mathbf{v} \geq 0$ , with  $\mathbf{v} \cdot \mathbf{v} = 0$  only when  $\mathbf{v} = \mathbf{0}$ .

*Proof.* Each part follows from the definition and the rules of arithmetic for real numbers.

(a)  $\mathbf{v} \cdot \mathbf{w} = \sum_i v_i w_i = \sum_i w_i v_i = \mathbf{w} \cdot \mathbf{v}$ , because  $v_i w_i = w_i v_i$  for each  $i$ .

(b) The  $i$ -th entry of  $\mathbf{v} + \mathbf{w}$  is  $v_i + w_i$ , so

$$\mathbf{u} \cdot (\mathbf{v} + \mathbf{w}) = \sum_{i=1}^n u_i (v_i + w_i) = \sum_{i=1}^n (u_i v_i + u_i w_i) = \sum_{i=1}^n u_i v_i + \sum_{i=1}^n u_i w_i = \mathbf{u} \cdot \mathbf{v} + \mathbf{u} \cdot \mathbf{w}.$$

(c) The  $i$ -th entry of  $c\mathbf{v}$  is  $cv_i$ , so  $(c\mathbf{v}) \cdot \mathbf{w} = \sum_i (cv_i) w_i = c \sum_i v_i w_i = c(\mathbf{v} \cdot \mathbf{w})$ .

(d)  $\mathbf{v} \cdot \mathbf{v} = \sum_i v_i^2$  is a sum of squares of real numbers, hence  $\geq 0$ . If it equals 0, then every term  $v_i^2$ , being non-negative, must be 0, so every  $v_i = 0$  and  $\mathbf{v} = \mathbf{0}$ .  $\square$

Combining (a) with (b) and (c) gives the same rules in the other slot:  $(\mathbf{v} + \mathbf{w}) \cdot \mathbf{u} = \mathbf{v} \cdot \mathbf{u} + \mathbf{w} \cdot \mathbf{u}$  and  $\mathbf{v} \cdot (c\mathbf{w}) = c(\mathbf{v} \cdot \mathbf{w})$ . Taking  $c = 0$  in (c) shows  $\mathbf{0} \cdot \mathbf{w} = 0$  for every  $\mathbf{w}$ . These rules let us expand dot products just like products of numbers. The expansion we use most often is

$$\|\mathbf{v} \pm \mathbf{w}\|^2 = (\mathbf{v} \pm \mathbf{w}) \cdot (\mathbf{v} \pm \mathbf{w}) = \|\mathbf{v}\|^2 \pm 2\mathbf{v} \cdot \mathbf{w} + \|\mathbf{w}\|^2, \quad (1.2)$$

where the two middle terms  $\mathbf{v} \cdot \mathbf{w}$  and  $\mathbf{w} \cdot \mathbf{v}$  combine by symmetry.

**Proposition 1.4.7.** For every  $\mathbf{v} \in \mathbb{R}^n$  and scalar  $c$ ,  $\|c\mathbf{v}\| = |c| \|\mathbf{v}\|$ .

*Proof.* By (1.1) and homogeneity in both slots,  $\|c\mathbf{v}\|^2 = (c\mathbf{v}) \cdot (c\mathbf{v}) = c^2(\mathbf{v} \cdot \mathbf{v}) = c^2\|\mathbf{v}\|^2$ . Both  $\|c\mathbf{v}\|$  and  $|c| \|\mathbf{v}\|$  are non-negative, and their squares agree, so they are equal (using  $\sqrt{c^2} = |c|$ ).  $\square$

So  $2\mathbf{v}$  is twice as long as  $\mathbf{v}$ , and  $-\mathbf{v}$  has the same length as  $\mathbf{v}$ , as the arrow picture of Section 1.1 suggested.

**Definition 1.4.8** (Unit vector). A *unit vector* is a vector of norm 1. For  $\mathbf{v} \neq \mathbf{0}$ , the vector  $\frac{1}{\|\mathbf{v}\|} \mathbf{v}$  is called the *normalisation* of  $\mathbf{v}$ .

By Proposition 1.4.7,  $\|\frac{1}{\|\mathbf{v}\|} \mathbf{v}\| = \frac{1}{\|\mathbf{v}\|} \|\mathbf{v}\| = 1$ , so the normalisation is a unit vector. It points the same way as  $\mathbf{v}$ , since it is a positive multiple of  $\mathbf{v}$ .

**Example 1.4.9.**  $\|(3, 4)\| = 5$ , so the normalisation of  $(3, 4)$  is  $\frac{1}{5}(3, 4) = (\frac{3}{5}, \frac{4}{5})$ . Check:  $\frac{9}{25} + \frac{16}{25} = 1$ . In  $\mathbb{R}^2$  the unit vectors are exactly the points of the circle of radius 1 about the origin, and every one of them has the form  $(\cos \alpha, \sin \alpha)$ .

#### 1.4.4 The angle between two vectors

What does the number  $\mathbf{v} \cdot \mathbf{w}$  say about the arrows? Draw nonzero vectors  $\mathbf{v}$  and  $\mathbf{w}$  in the plane from the origin, and let  $\theta \in [0, \pi]$  be the angle between them. The third side of the triangle they span runs from the tip of  $\mathbf{v}$  to the tip of  $\mathbf{w}$ ; it is the arrow of  $\mathbf{w} - \mathbf{v}$  (Figure 1.13).

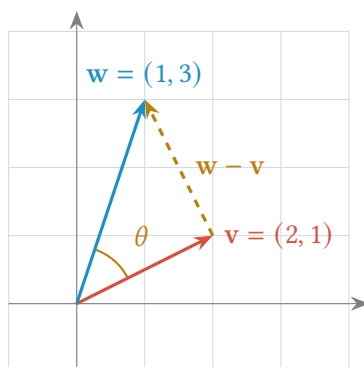


Figure 1.13: The triangle spanned by  $\mathbf{v}$  and  $\mathbf{w}$ . Here  $\theta = 45^\circ$  (Example 1.4.12).

**Proposition 1.4.10** (Dot product and angle in the plane). Let  $\mathbf{v}, \mathbf{w}$  be nonzero vectors in  $\mathbb{R}^2$  (or  $\mathbb{R}^3$ ), and let  $\theta$  be the angle between their arrows. Then

$$\mathbf{v} \cdot \mathbf{w} = \|\mathbf{v}\| \|\mathbf{w}\| \cos \theta.$$

*Proof.* By (1.2), the squared length of the third side is  $\|\mathbf{w} - \mathbf{v}\|^2 = \|\mathbf{v}\|^2 + \|\mathbf{w}\|^2 - 2\mathbf{v} \cdot \mathbf{w}$ . The law of cosines from plane geometry, applied to the triangle with sides  $\|\mathbf{v}\|$ ,  $\|\mathbf{w}\|$  and  $\|\mathbf{w} - \mathbf{v}\|$  and angle  $\theta$  between the first two, gives  $\|\mathbf{w} - \mathbf{v}\|^2 = \|\mathbf{v}\|^2 + \|\mathbf{w}\|^2 - 2\|\mathbf{v}\| \|\mathbf{w}\| \cos \theta$ . (If  $\mathbf{v}$  and  $\mathbf{w}$  lie on one line,  $\theta$  is 0 or  $\pi$ , the “triangle” is flat, and the same formula holds by direct computation.) Comparing the two

expressions gives the result. The vectors in  $\mathbb{R}^3$  lie in a common plane through the origin, so the same argument applies there.  $\square$

**Remark 1.4.11.** In  $\mathbb{R}^2$  there is also a proof by trigonometry that avoids the law of cosines. Write  $\mathbf{v} = \|\mathbf{v}\|(\cos \alpha, \sin \alpha)$  and  $\mathbf{w} = \|\mathbf{w}\|(\cos \beta, \sin \beta)$ , which is possible by Example 1.4.9. Then  $\mathbf{v} \cdot \mathbf{w} = \|\mathbf{v}\| \|\mathbf{w}\|(\cos \alpha \cos \beta + \sin \alpha \sin \beta) = \|\mathbf{v}\| \|\mathbf{w}\| \cos(\beta - \alpha)$ , and  $\cos(\beta - \alpha) = \cos \theta$  because  $\beta - \alpha$  differs from  $\pm \theta$  by a multiple of  $2\pi$ .

**Example 1.4.12.** Find the angle between  $\mathbf{v} = (2, 1)$  and  $\mathbf{w} = (1, 3)$ .

*Solution.*  $\mathbf{v} \cdot \mathbf{w} = 5$ ,  $\|\mathbf{v}\| = \sqrt{5}$  and  $\|\mathbf{w}\| = \sqrt{1+9} = \sqrt{10}$ . Hence

$$\cos \theta = \frac{\mathbf{v} \cdot \mathbf{w}}{\|\mathbf{v}\| \|\mathbf{w}\|} = \frac{5}{\sqrt{5} \sqrt{10}} = \frac{5}{\sqrt{50}} = \frac{5}{5\sqrt{2}} = \frac{1}{\sqrt{2}},$$

so  $\theta = \pi/4$ , that is  $45^\circ$ .

Since  $\|\mathbf{v}\| \|\mathbf{w}\| > 0$ , the sign of  $\mathbf{v} \cdot \mathbf{w}$  is the sign of  $\cos \theta$ . So the angle is *acute* ( $\theta < 90^\circ$ ) when  $\mathbf{v} \cdot \mathbf{w} > 0$ , *obtuse* when  $\mathbf{v} \cdot \mathbf{w} < 0$ , and a *right angle* exactly when  $\mathbf{v} \cdot \mathbf{w} = 0$ .

### 1.4.5 Orthogonality

**Definition 1.4.13** (Orthogonal vectors). Vectors  $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$  are *orthogonal*, written  $\mathbf{v} \perp \mathbf{w}$ , when  $\mathbf{v} \cdot \mathbf{w} = 0$ .

For example,  $(2, 1) \cdot (-1, 2) = -2 + 2 = 0$ , so  $(2, 1) \perp (-1, 2)$ , and their arrows meet at a right angle. In  $\mathbb{R}^3$ ,  $(1, 1, 1) \perp (1, -2, 1)$ . The zero vector is orthogonal to every vector, since  $\mathbf{0} \cdot \mathbf{w} = 0$ ; “orthogonal” means “dot product zero”, even when one of the arrows has no direction.

**Theorem 1.4.14** (Pythagoras in  $\mathbb{R}^n$ ). For  $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ ,  $\|\mathbf{v} + \mathbf{w}\|^2 = \|\mathbf{v}\|^2 + \|\mathbf{w}\|^2$  if and only if  $\mathbf{v} \perp \mathbf{w}$ .

*Proof.* By (1.2),  $\|\mathbf{v} + \mathbf{w}\|^2 = \|\mathbf{v}\|^2 + 2\mathbf{v} \cdot \mathbf{w} + \|\mathbf{w}\|^2$ . This equals  $\|\mathbf{v}\|^2 + \|\mathbf{w}\|^2$  exactly when  $2\mathbf{v} \cdot \mathbf{w} = 0$ , that is, when  $\mathbf{v} \cdot \mathbf{w} = 0$ .  $\square$

### 1.4.6 The Cauchy–Schwarz inequality

Proposition 1.4.10 was proved with a picture in the plane. In  $\mathbb{R}^4$  or  $\mathbb{R}^{10}$  there is no picture, so we turn the formula around and use it as the *definition* of the angle. A cosine lies in  $[-1, 1]$ , so this is only possible if  $\mathbf{v} \cdot \mathbf{w} / (\|\mathbf{v}\| \|\mathbf{w}\|)$  always lies in  $[-1, 1]$ . That is the content of the following theorem.

**Theorem 1.4.15** (Cauchy–Schwarz inequality). For all  $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ ,

$$|\mathbf{v} \cdot \mathbf{w}| \leq \|\mathbf{v}\| \|\mathbf{w}\|.$$

Equality holds if and only if one of the two vectors is a scalar multiple of the other.

*Proof.* If  $\mathbf{w} = \mathbf{0}$ , both sides are 0, so the inequality holds with equality, and  $\mathbf{w} = 0\mathbf{v}$  is a multiple of  $\mathbf{v}$ . Assume from now on that  $\mathbf{w} \neq \mathbf{0}$ . For a real number  $t$ , consider the vector  $\mathbf{v} - t\mathbf{w}$ . Its squared norm is never negative, and by the rules of Theorem 1.4.6,

$$0 \leq \|\mathbf{v} - t\mathbf{w}\|^2 = (\mathbf{v} - t\mathbf{w}) \cdot (\mathbf{v} - t\mathbf{w}) = \|\mathbf{w}\|^2 t^2 - 2(\mathbf{v} \cdot \mathbf{w})t + \|\mathbf{v}\|^2 \quad \text{for all } t \in \mathbb{R}.$$

The right-hand side is a quadratic polynomial  $q(t) = at^2 + bt + c$  with  $a = \|\mathbf{w}\|^2 > 0$ ,  $b = -2(\mathbf{v} \cdot \mathbf{w})$  and  $c = \|\mathbf{v}\|^2$ . A quadratic with  $a > 0$  and positive discriminant has two distinct real roots and is negative between them; since  $q(t) \geq 0$  for all  $t$ , the discriminant satisfies

$$b^2 - 4ac = 4(\mathbf{v} \cdot \mathbf{w})^2 - 4\|\mathbf{v}\|^2\|\mathbf{w}\|^2 \leq 0.$$

Dividing by 4 gives  $(\mathbf{v} \cdot \mathbf{w})^2 \leq \|\mathbf{v}\|^2\|\mathbf{w}\|^2$ , and taking non-negative square roots gives  $|\mathbf{v} \cdot \mathbf{w}| \leq \|\mathbf{v}\| \|\mathbf{w}\|$ .

*Equality.* If  $\mathbf{v} = s\mathbf{w}$  for a scalar  $s$ , then  $|\mathbf{v} \cdot \mathbf{w}| = |s| \|\mathbf{w}\|^2 = \|s\mathbf{w}\| \|\mathbf{w}\|$  by Proposition 1.4.7, so equality holds; the case  $\mathbf{w} = s\mathbf{v}$  is the same with the roles exchanged. Conversely, suppose equality holds and  $\mathbf{w} \neq \mathbf{0}$ . Then the discriminant is 0, so  $q$  has a real root  $t_0$ . Then  $\|\mathbf{v} - t_0\mathbf{w}\|^2 = q(t_0) = 0$ , so  $\mathbf{v} - t_0\mathbf{w} = \mathbf{0}$  by positivity, and  $\mathbf{v} = t_0\mathbf{w}$ .  $\square$

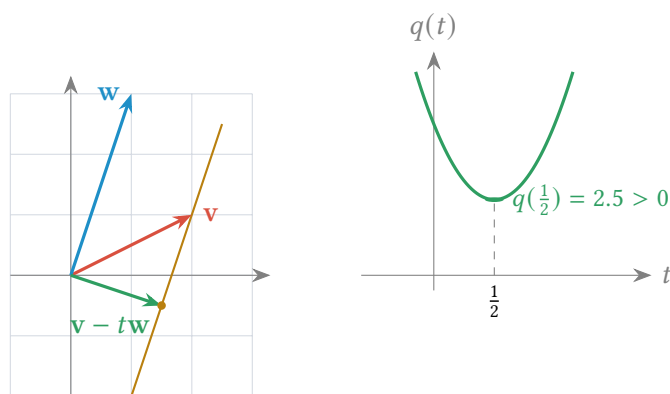


Figure 1.14: The proof of Cauchy-Schwarz for  $\mathbf{v} = (2, 1)$ ,  $\mathbf{w} = (1, 3)$ . Left: as  $t$  varies, the tip of  $\mathbf{v} - t\mathbf{w}$  moves along the gold line. Right:  $q(t) = \|\mathbf{v} - t\mathbf{w}\|^2 = 10t^2 - 10t + 5$  stays above the axis, so its discriminant  $100 - 200$  is negative.

For  $\mathbf{v} = (2, 1)$  and  $\mathbf{w} = (1, 3)$ , Cauchy-Schwarz says  $5 \leq \sqrt{5} \sqrt{10} = \sqrt{50} \approx 7.07$ . The inequality is strict because neither vector is a multiple of the other.

**Definition 1.4.16** (Angle in  $\mathbb{R}^n$ ). The *angle* between nonzero vectors  $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$  is the unique  $\theta \in [0, \pi]$  with

$$\cos \theta = \frac{\mathbf{v} \cdot \mathbf{w}}{\|\mathbf{v}\| \|\mathbf{w}\|}.$$

By Cauchy-Schwarz the fraction lies in  $[-1, 1]$ , and  $\cos$  maps  $[0, \pi]$  one-to-one onto  $[-1, 1]$ , so  $\theta$  exists and is unique. By Proposition 1.4.10 this agrees with the geometric angle in  $\mathbb{R}^2$  and  $\mathbb{R}^3$ . By the equality case,  $\theta = 0$  or  $\theta = \pi$  exactly when  $\mathbf{v}$  and  $\mathbf{w}$  are multiples of each other. Nonzero vectors are orthogonal exactly when the angle between them is  $\pi/2$ .

**Example 1.4.17.** In  $\mathbb{R}^4$ , find the angle between the diagonal  $\mathbf{d} = (1, 1, 1, 1)$  and  $\mathbf{e} = (1, 0, 0, 0)$ .

*Solution.*  $\mathbf{d} \cdot \mathbf{e} = 1 \cdot 1 + 1 \cdot 0 + 1 \cdot 0 + 1 \cdot 0 = 1$ ,  $\|\mathbf{d}\| = \sqrt{4} = 2$  and  $\|\mathbf{e}\| = 1$ . So  $\cos \theta = \frac{1}{2}$  and  $\theta = \pi/3 = 60^\circ$ . The long diagonal of the four-dimensional cube makes an angle of  $60^\circ$  with each edge at a corner. (In  $\mathbb{R}^3$  the same computation gives  $\cos \theta = 1/\sqrt{3}$ , about  $54.7^\circ$ ; in  $\mathbb{R}^n$  it gives  $1/\sqrt{n}$ , so for large  $n$  the diagonal is almost orthogonal to every edge.)

## 1.4.7 The triangle inequality

**Theorem 1.4.18** (Triangle inequality). For all  $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ ,

$$\|\mathbf{v} + \mathbf{w}\| \leq \|\mathbf{v}\| + \|\mathbf{w}\|.$$

Equality holds if and only if one of the vectors is a non-negative multiple of the other.

*Proof.* Both sides are non-negative, so it is enough to compare their squares. By (1.2) and Cauchy–Schwarz,

$$\|\mathbf{v} + \mathbf{w}\|^2 = \|\mathbf{v}\|^2 + 2\mathbf{v} \cdot \mathbf{w} + \|\mathbf{w}\|^2 \leq \|\mathbf{v}\|^2 + 2|\mathbf{v} \cdot \mathbf{w}| + \|\mathbf{w}\|^2 \leq \|\mathbf{v}\|^2 + 2\|\mathbf{v}\|\|\mathbf{w}\| + \|\mathbf{w}\|^2 = (\|\mathbf{v}\| + \|\mathbf{w}\|)^2.$$

*Equality.* Equality holds in the theorem exactly when both inequalities in this chain are equalities, that is, when  $\mathbf{v} \cdot \mathbf{w} = |\mathbf{v} \cdot \mathbf{w}| = \|\mathbf{v}\|\|\mathbf{w}\|$ . If  $\mathbf{w} = \mathbf{0}$  this holds, and  $\mathbf{w} = \mathbf{0} \cdot \mathbf{v}$ . If  $\mathbf{w} \neq \mathbf{0}$ , then by the equality case of Cauchy–Schwarz  $\mathbf{v} = s\mathbf{w}$  for some  $s$ , and  $\mathbf{v} \cdot \mathbf{w} = s\|\mathbf{w}\|^2 \geq 0$  forces  $s \geq 0$ . Conversely, if  $\mathbf{v} = s\mathbf{w}$  with  $s \geq 0$  then  $\|\mathbf{v} + \mathbf{w}\| = (1 + s)\|\mathbf{w}\| = \|\mathbf{v}\| + \|\mathbf{w}\|$ , and similarly if  $\mathbf{w} = s\mathbf{v}$  with  $s \geq 0$ .  $\square$

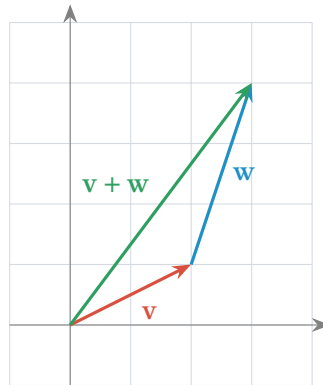


Figure 1.15: The triangle inequality:  $\|\mathbf{v} + \mathbf{w}\| = 5 \leq \sqrt{5} + \sqrt{10} \approx 5.40$  for  $\mathbf{v} = (2, 1)$ ,  $\mathbf{w} = (1, 3)$ .

**Proposition 1.4.19** (Triangle inequality for distances). For all  $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ ,  $d(\mathbf{u}, \mathbf{w}) \leq d(\mathbf{u}, \mathbf{v}) + d(\mathbf{v}, \mathbf{w})$ .

*Proof.*  $\mathbf{w} - \mathbf{u} = (\mathbf{v} - \mathbf{u}) + (\mathbf{w} - \mathbf{v})$ , so by Theorem 1.4.18,  $\|\mathbf{w} - \mathbf{u}\| \leq \|\mathbf{v} - \mathbf{u}\| + \|\mathbf{w} - \mathbf{v}\|$ .  $\square$

In words: a detour through a third point never makes the trip shorter.

#### Check yourself

Let  $\mathbf{v} = (1, 2, 2)$  and  $\mathbf{w} = (2, -2, 1)$ . Compute  $\mathbf{v} \cdot \mathbf{w}$ ,  $\|\mathbf{v}\|$  and  $\|\mathbf{w}\|$ . Are  $\mathbf{v}$  and  $\mathbf{w}$  orthogonal? Verify Theorem 1.4.14 by computing  $\|\mathbf{v} + \mathbf{w}\|^2$  directly.

#### Supplement: beyond the lecture

This material is not needed later in the course, but it rounds out the picture.

**The parallelogram law and polarisation.** Adding and subtracting the two cases of (1.2) gives

$$\|\mathbf{v} + \mathbf{w}\|^2 + \|\mathbf{v} - \mathbf{w}\|^2 = 2\|\mathbf{v}\|^2 + 2\|\mathbf{w}\|^2, \quad \mathbf{v} \cdot \mathbf{w} = \frac{1}{4}(\|\mathbf{v} + \mathbf{w}\|^2 - \|\mathbf{v} - \mathbf{w}\|^2).$$

The first identity says that the squares of the two diagonals of a parallelogram add up to the squares of its four sides. The second, the *polarisation identity*, shows that the dot product can be recovered from lengths alone: anything that preserves all lengths also preserves all dot products, and hence all angles. This fact returns when we study orthogonal matrices in Chapter 11.

**A second proof of Cauchy–Schwarz.** If  $\mathbf{v}$  and  $\mathbf{w}$  are nonzero, let  $\hat{\mathbf{v}} = \mathbf{v}/\|\mathbf{v}\|$  and  $\hat{\mathbf{w}} = \mathbf{w}/\|\mathbf{w}\|$  be their normalisations. Then  $0 \leq \|\hat{\mathbf{v}} \mp \hat{\mathbf{w}}\|^2 = 2 \mp 2\hat{\mathbf{v}} \cdot \hat{\mathbf{w}}$ , so  $-1 \leq \hat{\mathbf{v}} \cdot \hat{\mathbf{w}} \leq 1$ , and multiplying by  $\|\mathbf{v}\| \|\mathbf{w}\|$  gives  $|\mathbf{v} \cdot \mathbf{w}| \leq \|\mathbf{v}\| \|\mathbf{w}\|$ . This proof shows the inequality is really a statement about unit vectors: two unit vectors have dot product between  $-1$  and  $1$ .

**The reverse triangle inequality.** Writing  $\mathbf{v} = (\mathbf{v} - \mathbf{w}) + \mathbf{w}$  gives  $\|\mathbf{v}\| - \|\mathbf{w}\| \leq \|\mathbf{v} - \mathbf{w}\|$ , and exchanging the roles gives  $|\|\mathbf{v}\| - \|\mathbf{w}\|| \leq \|\mathbf{v} - \mathbf{w}\|$ : vectors that are close together have nearly equal lengths.

**Cosine similarity.** In text analysis a document is recorded as a vector of word counts, one component per word of a fixed vocabulary. Long documents have long vectors, so the raw distance between two documents mostly measures their lengths. The *cosine similarity*  $\mathbf{v} \cdot \mathbf{w} / (\|\mathbf{v}\| \|\mathbf{w}\|)$  ignores length and compares only direction. For instance, with the vocabulary (*matrix*, *vector*, *flour*), the counts  $(6, 4, 0)$  and  $(3, 2, 0)$  have cosine similarity  $1$  (one document is the other repeated twice), while  $(6, 4, 0)$  and  $(0, 1, 5)$  have cosine similarity  $4/(\sqrt{52}\sqrt{26}) \approx 0.11$ . Search engines and recommendation systems use this measure, or refinements of it, every day.

## 1.5 Lines, planes and the cross product

### Lecture 5

Lines and planes are the simplest shapes in geometry, and with vectors we can describe them exactly, in any number of dimensions. There are two ways to do it. A *parametric* description says how to reach every point of the shape: start at one point and move along some directions. A *normal* description says how to recognise the points of the shape: they are the solutions of an equation, written with the dot product. The first is good for producing points, the second for testing them, and moving between the two is a basic skill. In  $\mathbb{R}^3$  the bridge from a parametric plane to its equation is a new product of two vectors, the *cross product*, which is special to three dimensions. This section ends Chapter 1.

Throughout, a point  $P$  is identified with its position vector  $\mathbf{p}$  (Section 1.2), so “the point  $(1, 1)$ ” and “the vector  $(1, 1)$ ” are the same object, drawn as a dot or as an arrow from the origin.

#### 1.5.1 Lines

In Section 1.3 we saw that the multiples  $t\mathbf{d}$  of a nonzero vector  $\mathbf{d}$  fill the line through the origin in the direction of  $\mathbf{d}$ : that line is  $\text{Span}\{\mathbf{d}\}$ . To get a line through any other point  $\mathbf{p}$ , we shift this span by adding  $\mathbf{p}$  to every vector in it.

**Definition 1.5.1** (Line). Let  $\mathbf{p}, \mathbf{d} \in \mathbb{R}^n$  with  $\mathbf{d} \neq \mathbf{0}$ . The line through  $\mathbf{p}$  with direction vector  $\mathbf{d}$  is the set

$$L = \{ \mathbf{p} + t\mathbf{d} : t \in \mathbb{R} \}.$$

The equation  $\mathbf{x} = \mathbf{p} + t\mathbf{d}$  is called a *parametric equation* (or *parametric form*) of  $L$ , and  $t$  is the *parameter*.

Each value of  $t$  gives one point:  $t = 0$  gives  $\mathbf{p}$  itself,  $t = 1$  gives  $\mathbf{p} + \mathbf{d}$ ,  $t = -1$  gives  $\mathbf{p} - \mathbf{d}$ , and  $t = \frac{1}{2}$  the point halfway between  $\mathbf{p}$  and  $\mathbf{p} + \mathbf{d}$ . As  $t$  runs through all real numbers the point  $\mathbf{p} + t\mathbf{d}$  runs along the whole line (Figure 1.16). One can think of  $t$  as time: at time  $t$  a moving point is at  $\mathbf{p} + t\mathbf{d}$ , so it starts at  $\mathbf{p}$  and moves with constant velocity  $\mathbf{d}$ .

The line through two different points  $\mathbf{a}$  and  $\mathbf{b}$  is the line through  $\mathbf{a}$  with direction  $\mathbf{d} = \mathbf{b} - \mathbf{a}$ , the step from  $\mathbf{a}$  to  $\mathbf{b}$ :

$$\mathbf{x} = \mathbf{a} + t(\mathbf{b} - \mathbf{a}), \quad t \in \mathbb{R}. \quad (1.3)$$

It contains  $\mathbf{a}$  (at  $t = 0$ ) and  $\mathbf{b}$  (at  $t = 1$ ), and the points with  $0 \leq t \leq 1$  form the segment from  $\mathbf{a}$  to  $\mathbf{b}$ ;  $t = \frac{1}{2}$  gives the midpoint  $\frac{1}{2}(\mathbf{a} + \mathbf{b})$  of Proposition 1.2.10.

**Example 1.5.2.** Find a parametric equation of the line  $L$  through  $A = (1, 1)$  and  $B = (4, 3)$ , and decide whether the points  $(7, 5)$  and  $(2, 4)$  lie on it.

*Solution.* The direction is  $\mathbf{d} = \mathbf{b} - \mathbf{a} = (4 - 1, 3 - 1) = (3, 2)$ , so

$$L : \mathbf{x} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} + t \begin{bmatrix} 3 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 + 3t \\ 1 + 2t \end{bmatrix}.$$

The point  $(7, 5)$  lies on  $L$  if some single  $t$  gives both coordinates:  $1 + 3t = 7$  needs  $t = 2$ , and  $1 + 2t = 5$  also needs  $t = 2$ . So  $(7, 5) = \mathbf{a} + 2\mathbf{d}$  is on  $L$ . For  $(2, 4)$ , the first coordinate needs  $t = \frac{1}{3}$  and the second needs  $t = \frac{3}{2}$ ; no single  $t$  works, so  $(2, 4)$  is not on  $L$ .

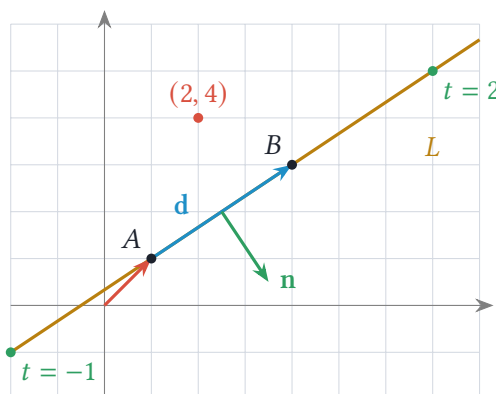


Figure 1.16: Example 1.5.2: the line  $\mathbf{x} = (1, 1) + t(3, 2)$  through  $A$  and  $B$ . The point  $(2, 4)$  is off the line. The vector  $\mathbf{n} = (2, -3)$  of Example 1.5.5, drawn from a point of  $L$ , is perpendicular to it.

A line has many parametric equations. Any of its points can serve as  $\mathbf{p}$ , and any nonzero multiple of  $\mathbf{d}$  can serve as the direction:  $\mathbf{x} = (7, 5) + s(-6, -4)$  describes the same line as in Example 1.5.2 (take  $s = -\frac{1}{2}(t - 2)$ ). The next proposition says exactly when two parametric equations describe the same line.

**Proposition 1.5.3** (When two lines are equal). Let  $L = \{\mathbf{p} + t\mathbf{d}\}$  and  $L' = \{\mathbf{q} + s\mathbf{e}\}$  be lines in  $\mathbb{R}^n$ . Then  $L = L'$  if and only if  $\mathbf{q} \in L$  and  $\mathbf{e}$  is a (nonzero) multiple of  $\mathbf{d}$ .

*Proof.* Suppose  $L = L'$ . Then  $\mathbf{q} \in L'$  (take  $s = 0$ ), so  $\mathbf{q} \in L$ . The points  $\mathbf{q}$  and  $\mathbf{q} + \mathbf{e}$  of  $L'$  lie in  $L$ , so  $\mathbf{q} = \mathbf{p} + t_0\mathbf{d}$  and  $\mathbf{q} + \mathbf{e} = \mathbf{p} + t_1\mathbf{d}$  for some  $t_0, t_1$ ; subtracting,  $\mathbf{e} = (t_1 - t_0)\mathbf{d}$ .

Conversely, suppose  $\mathbf{q} = \mathbf{p} + t_0\mathbf{d}$  and  $\mathbf{e} = k\mathbf{d}$ ; then  $k \neq 0$  because  $\mathbf{e} \neq \mathbf{0}$ . Every point of  $L'$  is  $\mathbf{q} + s\mathbf{e} = \mathbf{p} + (t_0 + ks)\mathbf{d} \in L$ . Every point of  $L$  is  $\mathbf{p} + t\mathbf{d} = \mathbf{q} + \frac{t-t_0}{k}\mathbf{e} \in L'$ . So  $L = L'$ .  $\square$

In particular, two lines with parallel direction vectors are either equal or have no point in common; such lines are called *parallel*. Parametric equations work in every  $\mathbb{R}^n$ . For example, the line in  $\mathbb{R}^3$  through  $(1, 0, 2)$  with direction  $(1, 2, -1)$  is  $\mathbf{x} = (1 + t, 2t, 2 - t)$ ; it passes through  $(2, 2, 1)$  at  $t = 1$  and meets the floor plane  $x_3 = 0$  at  $t = 2$ , in the point  $(3, 4, 0)$ .

### 1.5.2 The normal equation of a line in the plane

At school a line in the plane is described by an equation such as  $2x - 3y = -1$ . The dot product explains where such an equation comes from. Let  $L$  be the line through  $\mathbf{p}$  with direction  $\mathbf{d}$  in  $\mathbb{R}^2$ , and let  $\mathbf{n} \neq \mathbf{0}$  be a vector orthogonal to  $\mathbf{d}$ . A point  $\mathbf{x}$  lies on  $L$  exactly when the step  $\mathbf{x} - \mathbf{p}$  is a multiple of  $\mathbf{d}$ , and in the plane that happens exactly when  $\mathbf{x} - \mathbf{p}$  is orthogonal to  $\mathbf{n}$  (Figure 1.16). So

$$\mathbf{x} \in L \iff \mathbf{n} \cdot (\mathbf{x} - \mathbf{p}) = 0 \iff \mathbf{n} \cdot \mathbf{x} = \mathbf{n} \cdot \mathbf{p}. \quad (1.4)$$

Writing  $\mathbf{n} = (a, b)$ ,  $\mathbf{x} = (x, y)$  and  $c = \mathbf{n} \cdot \mathbf{p}$ , this is the familiar equation  $ax + by = c$ . The middle step of (1.4) uses a fact about  $\mathbb{R}^2$  that the proof below makes precise.

**Proposition 1.5.4** (Lines in the plane as equations). Let  $\mathbf{n} = (a, b) \neq \mathbf{0}$  and  $c \in \mathbb{R}$ . Then the set  $\{(x, y) \in \mathbb{R}^2 : ax + by = c\}$  is a line, with direction vector  $(-b, a)$ . Conversely, the line through  $\mathbf{p}$  with direction  $\mathbf{d} = (d_1, d_2)$  is the set of solutions of  $\mathbf{n} \cdot \mathbf{x} = \mathbf{n} \cdot \mathbf{p}$ , where  $\mathbf{n} = (-d_2, d_1)$ .

*Proof.* First, the vectors  $\mathbf{y} = (y_1, y_2)$  with  $\mathbf{n} \cdot \mathbf{y} = 0$  are exactly the multiples of  $\mathbf{m} = (-b, a)$ . Indeed  $\mathbf{n} \cdot (k\mathbf{m}) = k(-ab + ba) = 0$ . Conversely, let  $ay_1 + by_2 = 0$ . If  $a \neq 0$  then  $y_1 = -by_2/a$  and  $\mathbf{y} = \frac{y_2}{a}(-b, a)$ ; if  $a = 0$  then  $b \neq 0$ , so  $y_2 = 0$  and  $\mathbf{y} = (y_1, 0) = -\frac{y_1}{b}(-b, 0)$ .

Now pick one solution  $\mathbf{p}$  of  $\mathbf{n} \cdot \mathbf{x} = c$ , for example  $\mathbf{p} = \frac{c}{\|\mathbf{n}\|^2}\mathbf{n}$ . Then  $\mathbf{n} \cdot \mathbf{x} = c$  if and only if  $\mathbf{n} \cdot (\mathbf{x} - \mathbf{p}) = 0$ , if and only if  $\mathbf{x} - \mathbf{p} = t\mathbf{m}$  for some  $t$ , by the first step. So the solution set is the line through  $\mathbf{p}$  with direction  $\mathbf{m}$ , which is nonzero because  $\mathbf{n}$  is.

For the converse,  $\mathbf{n} = (-d_2, d_1)$  is nonzero and  $(-n_2, n_1) = (-d_1, -d_2) = -\mathbf{d}$ . By the first part, the solutions of  $\mathbf{n} \cdot \mathbf{x} = \mathbf{n} \cdot \mathbf{p}$  form the line through  $\mathbf{p}$  with direction  $-\mathbf{d}$ , which is the given line by Proposition 1.5.3.  $\square$

A nonzero vector orthogonal to the direction of a line is called a *normal vector* of the line, and  $\mathbf{n} \cdot \mathbf{x} = c$  a *normal equation* of it. The coefficients of  $x$  and  $y$  in  $ax + by = c$  are the entries of a normal vector, so a normal vector can be read straight off the equation.

**Example 1.5.5.** Find a normal equation of the line  $L$  of Example 1.5.2, and convert  $3x + y = 6$  to parametric form.

*Solution.* The direction is  $\mathbf{d} = (3, 2)$ , so  $\mathbf{n} = (2, -3)$  is a normal vector:  $\mathbf{n} \cdot \mathbf{d} = 6 - 6 = 0$ . With  $\mathbf{p} = (1, 1)$ ,  $\mathbf{n} \cdot \mathbf{p} = 2 - 3 = -1$ , so  $L$  is  $2x - 3y = -1$ . Check:  $B = (4, 3)$  gives  $8 - 9 = -1$ , and  $(7, 5)$  gives  $14 - 15 = -1$ . Testing a point is now a single computation:  $(2, 4)$  gives  $4 - 12 = -8 \neq -1$ , so it is not on  $L$ .

For  $3x + y = 6$ , the normal vector is  $(3, 1)$ , so a direction is  $(-1, 3)$ , and  $x = 0$  gives the point  $(0, 6)$ :  $\mathbf{x} = (0, 6) + t(-1, 3)$ .

## 1.5.3 Planes

Two non-parallel vectors  $\mathbf{u}, \mathbf{v} \in \mathbb{R}^3$  span a plane through the origin (Section 1.3). Shifting it by  $\mathbf{p}$  gives a plane through  $\mathbf{p}$ .

**Definition 1.5.6** (Plane). Let  $\mathbf{p}, \mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ , where  $\mathbf{u}$  and  $\mathbf{v}$  are not parallel. The *plane through  $\mathbf{p}$  with direction vectors  $\mathbf{u}$  and  $\mathbf{v}$*  is the set

$$P = \{ \mathbf{p} + s\mathbf{u} + t\mathbf{v} : s, t \in \mathbb{R} \},$$

and  $\mathbf{x} = \mathbf{p} + s\mathbf{u} + t\mathbf{v}$  is a *parametric equation* of  $P$ , with parameters  $s$  and  $t$ .

The plane through three points  $\mathbf{a}, \mathbf{b}, \mathbf{c}$  that do not lie on one line has the parametric equation  $\mathbf{x} = \mathbf{a} + s(\mathbf{b} - \mathbf{a}) + t(\mathbf{c} - \mathbf{a})$ ; the directions are the steps from  $\mathbf{a}$  to the other two points.

**Example 1.5.7.** Find a parametric equation of the plane  $P$  through  $A = (1, 0, 0)$ ,  $B = (0, 2, 0)$  and  $C = (0, 0, 3)$ , and decide whether  $(0, 1, \frac{3}{2})$  lies on  $P$ .

*Solution.* The directions are  $\mathbf{u} = \mathbf{b} - \mathbf{a} = (-1, 2, 0)$  and  $\mathbf{v} = \mathbf{c} - \mathbf{a} = (-1, 0, 3)$ . They are not parallel (the second entry of  $\mathbf{v}$  is 0, so a multiple of  $\mathbf{v}$  equal to  $\mathbf{u}$  would need  $2 = 0$ , and a multiple of  $\mathbf{u}$  equal to  $\mathbf{v}$  would need  $3 = 0$  in the third entry). So

$$P : \quad \mathbf{x} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + s \begin{bmatrix} -1 \\ 2 \\ 0 \end{bmatrix} + t \begin{bmatrix} -1 \\ 0 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 - s - t \\ 2s \\ 3t \end{bmatrix}.$$

For  $(0, 1, \frac{3}{2})$  we need  $2s = 1$  and  $3t = \frac{3}{2}$ , so  $s = t = \frac{1}{2}$ ; then the first entry is  $1 - \frac{1}{2} - \frac{1}{2} = 0$ , as required. So the point is on  $P$ . Testing a point against a parametric equation means solving three equations for two unknowns; Example 1.5.21 below gives a faster test.

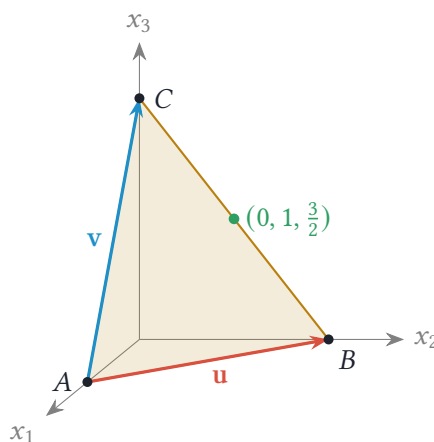


Figure 1.17: Example 1.5.7: the part of the plane through  $A, B, C$  that lies between the axes, with the direction vectors  $\mathbf{u} = \mathbf{b} - \mathbf{a}$  and  $\mathbf{v} = \mathbf{c} - \mathbf{a}$ .

**Planes as equations.** In  $\mathbb{R}^3$  a single linear equation  $ax + by + cz = d$  describes a plane, for the same reason as in the plane: it says that  $\mathbf{x} - \mathbf{p}$  is orthogonal to  $\mathbf{n} = (a, b, c)$ , and the vectors orthogonal to a nonzero vector of  $\mathbb{R}^3$  fill a plane through the origin.

**Proposition 1.5.8** (The solutions of one equation form a plane). Let  $\mathbf{n} = (a, b, c) \in \mathbb{R}^3$ ,  $\mathbf{n} \neq \mathbf{0}$ , and  $d \in \mathbb{R}$ . The set  $\{\mathbf{x} \in \mathbb{R}^3 : \mathbf{n} \cdot \mathbf{x} = d\}$  is a plane, and both of its direction vectors can be chosen orthogonal to  $\mathbf{n}$ .

*Proof.* Suppose first that  $c \neq 0$ . Put  $\mathbf{u} = (c, 0, -a)$  and  $\mathbf{v} = (0, c, -b)$ . Then  $\mathbf{n} \cdot \mathbf{u} = ac - ca = 0$  and  $\mathbf{n} \cdot \mathbf{v} = bc - cb = 0$ , and  $\mathbf{u}, \mathbf{v}$  are not parallel: their first two entries are  $(c, 0)$  and  $(0, c)$ , and no multiple of either pair is the other since  $c \neq 0$ .

We claim that the vectors  $\mathbf{y}$  with  $\mathbf{n} \cdot \mathbf{y} = 0$  are exactly the combinations  $s\mathbf{u} + t\mathbf{v}$ . Every such combination satisfies  $\mathbf{n} \cdot (s\mathbf{u} + t\mathbf{v}) = s\mathbf{n} \cdot \mathbf{u} + t\mathbf{n} \cdot \mathbf{v} = 0$ . Conversely, if  $ay_1 + by_2 + cy_3 = 0$ , then  $y_3 = -(ay_1 + by_2)/c$  and

$$\frac{y_1}{c}\mathbf{u} + \frac{y_2}{c}\mathbf{v} = \left(y_1, y_2, -\frac{ay_1 + by_2}{c}\right) = \mathbf{y}.$$

Now let  $\mathbf{p} = \frac{d}{\|\mathbf{n}\|^2}\mathbf{n}$ , so that  $\mathbf{n} \cdot \mathbf{p} = d$ . Then  $\mathbf{n} \cdot \mathbf{x} = d$  if and only if  $\mathbf{n} \cdot (\mathbf{x} - \mathbf{p}) = 0$ , if and only if  $\mathbf{x} - \mathbf{p} = s\mathbf{u} + t\mathbf{v}$  for some  $s, t$ . So the set is the plane through  $\mathbf{p}$  with directions  $\mathbf{u}, \mathbf{v}$ .

If  $c = 0$  but  $b \neq 0$ , the same argument works with the roles of the second and third coordinates exchanged:  $\mathbf{u} = (b, -a, 0)$  and  $\mathbf{v} = (0, 0, b)$ . If  $b = c = 0$ , then  $a \neq 0$  and we use  $\mathbf{u} = (0, a, 0)$ ,  $\mathbf{v} = (0, 0, a)$ . In each case the same checks go through.  $\square$

**Definition 1.5.9** (Normal vector of a plane). If a plane  $P \subseteq \mathbb{R}^3$  is the solution set of  $\mathbf{n} \cdot \mathbf{x} = d$  with  $\mathbf{n} \neq \mathbf{0}$ , then  $\mathbf{n}$  is called a *normal vector* of  $P$ , and  $\mathbf{n} \cdot \mathbf{x} = d$ , or  $ax + by + cz = d$ , a *normal equation* of  $P$ . Equivalently,  $P$  is the set of  $\mathbf{x}$  with  $\mathbf{n} \cdot (\mathbf{x} - \mathbf{p}) = 0$ , for any one point  $\mathbf{p}$  of  $P$ .

A normal vector is orthogonal to every step  $\mathbf{x} - \mathbf{y}$  between two points of the plane, since  $\mathbf{n} \cdot (\mathbf{x} - \mathbf{y}) = d - d = 0$ . It sticks straight out of the plane, like a nail hammered into a board.

**Example 1.5.10.** Find a normal equation of the plane through  $\mathbf{p} = (1, 2, 0)$  with normal vector  $\mathbf{n} = (2, 1, 2)$ , and test the points  $(0, 0, 2)$  and  $(1, 1, 1)$ .

*Solution.*  $\mathbf{n} \cdot \mathbf{p} = 2 + 2 + 0 = 4$ , so the plane is  $2x + y + 2z = 4$ . For  $(0, 0, 2)$ :  $0 + 0 + 4 = 4$ , so it lies on the plane. For  $(1, 1, 1)$ :  $2 + 1 + 2 = 5 \neq 4$ , so it does not. The plane meets the three axes at  $(2, 0, 0)$ ,  $(0, 4, 0)$  and  $(0, 0, 2)$ , found by setting two of the coordinates to zero.

**Remark 1.5.11.** In  $\mathbb{R}^3$  one equation leaves two free directions, so it describes a plane, not a line. To pin down a line in  $\mathbb{R}^3$  by equations we need two of them: a line is the intersection of two non-parallel planes. (Finding that intersection means solving a system of two equations, the subject of Section 2.2.) In  $\mathbb{R}^n$  the solution set of a single equation  $a_1x_1 + \cdots + a_nx_n = b$  with  $(a_1, \dots, a_n) \neq \mathbf{0}$  is called a *hyperplane*; the same argument as in Proposition 1.5.8 shows that it is a shifted span of  $n - 1$  vectors orthogonal to  $(a_1, \dots, a_n)$ . A hyperplane in  $\mathbb{R}^2$  is a line, and in  $\mathbb{R}^3$  a plane.

To go from a parametric equation  $\mathbf{x} = \mathbf{p} + s\mathbf{u} + t\mathbf{v}$  to a normal equation we need a nonzero vector  $\mathbf{n}$  that is orthogonal to both  $\mathbf{u}$  and  $\mathbf{v}$ . The cross product produces one.

### 1.5.4 The cross product

**Definition 1.5.12** (Cross product). The *cross product* of  $\mathbf{u} = (u_1, u_2, u_3)$  and  $\mathbf{v} = (v_1, v_2, v_3)$  in  $\mathbb{R}^3$

is the vector

$$\mathbf{u} \times \mathbf{v} = \begin{bmatrix} u_2v_3 - u_3v_2 \\ u_3v_1 - u_1v_3 \\ u_1v_2 - u_2v_1 \end{bmatrix}.$$

Unlike the dot product, the result is a vector, and it is defined only in  $\mathbb{R}^3$ . To remember the formula, note that the first entry uses only indices 2 and 3, the second only 3 and 1, the third only 1 and 2: each entry is “the next index times the one after, minus the reverse”. In the order  $1 \rightarrow 2 \rightarrow 3 \rightarrow 1$ , entry  $i$  is  $u_jv_k - u_kv_j$  where  $j$  and  $k$  are the two indices that follow  $i$ . A second way: cover the  $i$ -th row of the two columns  $\mathbf{u}, \mathbf{v}$ ; the entry is the “cross” of the two rows that remain, top-left times bottom-right minus top-right times bottom-left, with the sign flipped for the middle entry. (This is a  $2 \times 2$  determinant, the subject of Chapter 7.)

**Example 1.5.13.** For  $\mathbf{u} = (-1, 2, 0)$  and  $\mathbf{v} = (-1, 0, 3)$  from Example 1.5.7, entry by entry:

$$\text{first: } u_2v_3 - u_3v_2 = 2 \cdot 3 - 0 \cdot 0 = 6,$$

$$\text{second: } u_3v_1 - u_1v_3 = 0 \cdot (-1) - (-1) \cdot 3 = 0 + 3 = 3,$$

$$\text{third: } u_1v_2 - u_2v_1 = (-1) \cdot 0 - 2 \cdot (-1) = 0 + 2 = 2,$$

so  $\mathbf{u} \times \mathbf{v} = (6, 3, 2)$ . Check:  $(6, 3, 2) \cdot (-1, 2, 0) = -6 + 6 + 0 = 0$  and  $(6, 3, 2) \cdot (-1, 0, 3) = -6 + 0 + 6 = 0$ , so the result is orthogonal to both  $\mathbf{u}$  and  $\mathbf{v}$ . For the axis vectors  $(1, 0, 0) \times (0, 1, 0) = (0 \cdot 0 - 0 \cdot 1, 0 \cdot 0 - 1 \cdot 0, 1 \cdot 1 - 0 \cdot 0) = (0, 0, 1)$ .

**Theorem 1.5.14** (Properties of the cross product). For all  $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^3$  and every scalar  $c$ :

- (a) *anticommutativity*:  $\mathbf{v} \times \mathbf{u} = -(\mathbf{u} \times \mathbf{v})$ ; in particular  $\mathbf{u} \times \mathbf{u} = \mathbf{0}$ ;
- (b) *additivity*:  $\mathbf{u} \times (\mathbf{v} + \mathbf{w}) = \mathbf{u} \times \mathbf{v} + \mathbf{u} \times \mathbf{w}$ ;
- (c) *homogeneity*:  $(c\mathbf{u}) \times \mathbf{v} = c(\mathbf{u} \times \mathbf{v}) = \mathbf{u} \times (c\mathbf{v})$ ;
- (d) *orthogonality*:  $\mathbf{u} \cdot (\mathbf{u} \times \mathbf{v}) = 0$  and  $\mathbf{v} \cdot (\mathbf{u} \times \mathbf{v}) = 0$ .

*Proof.* (a) Exchanging  $\mathbf{u}$  and  $\mathbf{v}$  in the definition turns each entry  $u_jv_k - u_kv_j$  into  $v_ju_k - v_ku_j$ , its negative. With  $\mathbf{v} = \mathbf{u}$  this gives  $\mathbf{u} \times \mathbf{u} = -(\mathbf{u} \times \mathbf{u})$ , so  $2(\mathbf{u} \times \mathbf{u}) = \mathbf{0}$  and  $\mathbf{u} \times \mathbf{u} = \mathbf{0}$ .

(b) The first entry of  $\mathbf{u} \times (\mathbf{v} + \mathbf{w})$  is  $u_2(v_3 + w_3) - u_3(v_2 + w_2) = (u_2v_3 - u_3v_2) + (u_2w_3 - u_3w_2)$ , the sum of the first entries of  $\mathbf{u} \times \mathbf{v}$  and  $\mathbf{u} \times \mathbf{w}$ . The other entries are the same computation with the indices shifted.

(c) Each entry of  $(c\mathbf{u}) \times \mathbf{v}$  is  $(cu_j)v_k - (cu_k)v_j = c(u_jv_k - u_kv_j)$ , and similarly for  $\mathbf{u} \times (c\mathbf{v})$ .

(d) Using the definition,

$$\begin{aligned} \mathbf{u} \cdot (\mathbf{u} \times \mathbf{v}) &= u_1(u_2v_3 - u_3v_2) + u_2(u_3v_1 - u_1v_3) + u_3(u_1v_2 - u_2v_1) \\ &= u_1u_2v_3 - u_1u_3v_2 + u_2u_3v_1 - u_1u_2v_3 + u_1u_3v_2 - u_2u_3v_1 = 0, \end{aligned}$$

because the six terms cancel in pairs. For the second claim, by (a) and the first claim with the roles of  $\mathbf{u}$  and  $\mathbf{v}$  exchanged,  $\mathbf{v} \cdot (\mathbf{u} \times \mathbf{v}) = -\mathbf{v} \cdot (\mathbf{v} \times \mathbf{u}) = 0$ .  $\square$

Combining (a) with (b) and (c) gives the same rules in the first slot. So the cross product can be expanded like a product of numbers, except that the order of the factors matters: swapping them changes the sign.

**Remark 1.5.15.** The cross product is not associative. With  $\mathbf{e}_1 = (1, 0, 0)$ ,  $\mathbf{e}_2 = (0, 1, 0)$ ,  $\mathbf{e}_3 = (0, 0, 1)$  we have  $\mathbf{e}_1 \times \mathbf{e}_2 = \mathbf{e}_3$  (Example 1.5.13), and similarly  $\mathbf{e}_2 \times \mathbf{e}_3 = \mathbf{e}_1$ ,  $\mathbf{e}_3 \times \mathbf{e}_1 = \mathbf{e}_2$ . So  $(\mathbf{e}_1 \times \mathbf{e}_1) \times \mathbf{e}_2 = \mathbf{0} \times \mathbf{e}_2 = \mathbf{0}$ , while  $\mathbf{e}_1 \times (\mathbf{e}_1 \times \mathbf{e}_2) = \mathbf{e}_1 \times \mathbf{e}_3 = -\mathbf{e}_2$ .

Of the two directions orthogonal to  $\mathbf{u}$  and  $\mathbf{v}$ , the cross product picks the one given by the *right-hand rule*: if the fingers of the right hand curl from  $\mathbf{u}$  towards  $\mathbf{v}$ , the thumb points along  $\mathbf{u} \times \mathbf{v}$ . This is what  $\mathbf{e}_1 \times \mathbf{e}_2 = \mathbf{e}_3$  says when the axes are drawn the usual way. The rule is a convention about how we draw the axes, and it is made precise by the sign of the determinant (Chapter 7).

The length of the cross product has a geometric meaning, which follows from an algebraic identity.

**Theorem 1.5.16** (Length of the cross product). For all  $\mathbf{u}, \mathbf{v} \in \mathbb{R}^3$ ,

$$\|\mathbf{u} \times \mathbf{v}\|^2 = \|\mathbf{u}\|^2 \|\mathbf{v}\|^2 - (\mathbf{u} \cdot \mathbf{v})^2. \quad (1.5)$$

If  $\mathbf{u}$  and  $\mathbf{v}$  are nonzero and  $\theta$  is the angle between them, then  $\|\mathbf{u} \times \mathbf{v}\| = \|\mathbf{u}\| \|\mathbf{v}\| \sin \theta$ , which is the area of the parallelogram with sides  $\mathbf{u}$  and  $\mathbf{v}$ .

*Proof.* Identity (1.5) is checked by expanding both sides. The left side is

$$(u_2v_3 - u_3v_2)^2 + (u_3v_1 - u_1v_3)^2 + (u_1v_2 - u_2v_1)^2,$$

which expands into the six squares  $u_j^2v_k^2$  with  $j \neq k$ , minus the cross terms  $2u_2u_3v_2v_3 + 2u_1u_3v_1v_3 + 2u_1u_2v_1v_2$ . The right side is

$$(u_1^2 + u_2^2 + u_3^2)(v_1^2 + v_2^2 + v_3^2) - (u_1v_1 + u_2v_2 + u_3v_3)^2.$$

The first product consists of the nine terms  $u_j^2v_k^2$ ; the square  $(u_1v_1 + u_2v_2 + u_3v_3)^2$  consists of the three terms  $u_j^2v_j^2$  plus the same cross terms  $2u_ju_kv_jv_k$  ( $j < k$ ). After subtracting, the terms  $u_j^2v_j^2$  cancel, leaving the six squares with  $j \neq k$  minus the cross terms. The two sides agree.

For nonzero  $\mathbf{u}, \mathbf{v}$ , Definition 1.4.16 gives  $\mathbf{u} \cdot \mathbf{v} = \|\mathbf{u}\| \|\mathbf{v}\| \cos \theta$ , so (1.5) becomes  $\|\mathbf{u} \times \mathbf{v}\|^2 = \|\mathbf{u}\|^2 \|\mathbf{v}\|^2 (1 - \cos^2 \theta) = \|\mathbf{u}\|^2 \|\mathbf{v}\|^2 \sin^2 \theta$ . Since  $\theta \in [0, \pi]$ ,  $\sin \theta \geq 0$ , and taking square roots gives  $\|\mathbf{u} \times \mathbf{v}\| = \|\mathbf{u}\| \|\mathbf{v}\| \sin \theta$ . The parallelogram with sides  $\mathbf{u}$  and  $\mathbf{v}$  has base  $\|\mathbf{u}\|$  and height  $\|\mathbf{v}\| \sin \theta$  (the distance from the tip of  $\mathbf{v}$  to the line of  $\mathbf{u}$ ), so its area is  $\|\mathbf{u}\| \|\mathbf{v}\| \sin \theta$ .  $\square$

**Proposition 1.5.17** (Parallel vectors). For  $\mathbf{u}, \mathbf{v} \in \mathbb{R}^3$ ,  $\mathbf{u} \times \mathbf{v} = \mathbf{0}$  if and only if  $\mathbf{u}$  and  $\mathbf{v}$  are parallel (one is a multiple of the other).

*Proof.* By (1.5),  $\mathbf{u} \times \mathbf{v} = \mathbf{0}$  exactly when  $(\mathbf{u} \cdot \mathbf{v})^2 = \|\mathbf{u}\|^2 \|\mathbf{v}\|^2$ , that is, when  $|\mathbf{u} \cdot \mathbf{v}| = \|\mathbf{u}\| \|\mathbf{v}\|$ . By the equality case of the Cauchy–Schwarz inequality (Theorem 1.4.15) this happens exactly when one of the vectors is a multiple of the other.  $\square$

**Example 1.5.18.** The parallelogram with sides  $\mathbf{u} = (-1, 2, 0)$  and  $\mathbf{v} = (-1, 0, 3)$  has area  $\|(6, 3, 2)\| = \sqrt{36 + 9 + 4} = \sqrt{49} = 7$ . Check with (1.5):  $\|\mathbf{u}\|^2 = 5$ ,  $\|\mathbf{v}\|^2 = 10$  and  $\mathbf{u} \cdot \mathbf{v} = 1$ , so  $50 - 1 = 49$ . The triangle  $ABC$  of Example 1.5.7 is half of this parallelogram, so its area is  $\frac{7}{2}$ .

### 1.5.5 From three points to an equation; distance to a plane

We can now convert any parametric plane into a normal equation. The proof needs one more identity.

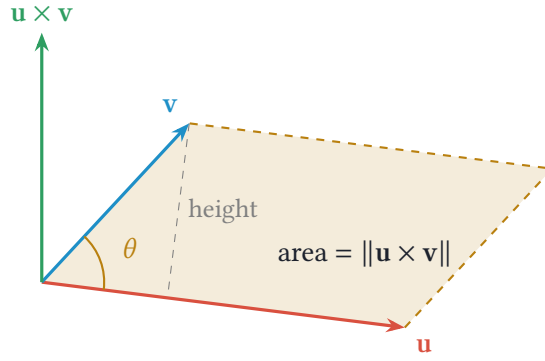


Figure 1.18: The cross product is orthogonal to  $\mathbf{u}$  and  $\mathbf{v}$ , and its length is the area  $\|\mathbf{u}\| \|\mathbf{v}\| \sin \theta$  of the parallelogram they span. The dashed drop from the tip of  $\mathbf{v}$  is the height  $\|\mathbf{v}\| \sin \theta$ .

**Proposition 1.5.19** (Triple cross product). For all  $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^3$ ,  $(\mathbf{u} \times \mathbf{v}) \times \mathbf{w} = (\mathbf{u} \cdot \mathbf{w}) \mathbf{v} - (\mathbf{v} \cdot \mathbf{w}) \mathbf{u}$ .

*Proof.* Write  $\mathbf{n} = \mathbf{u} \times \mathbf{v}$ . The first entry of  $\mathbf{n} \times \mathbf{w}$  is

$$n_2 w_3 - n_3 w_2 = (u_3 v_1 - u_1 v_3) w_3 - (u_1 v_2 - u_2 v_1) w_2 = v_1(u_2 w_2 + u_3 w_3) - u_1(v_2 w_2 + v_3 w_3).$$

Adding and subtracting  $u_1 v_1 w_1$  turns this into  $v_1(\mathbf{u} \cdot \mathbf{w}) - u_1(\mathbf{v} \cdot \mathbf{w})$ , the first entry of the right side. Shifting every index by one ( $1 \rightarrow 2 \rightarrow 3 \rightarrow 1$ ) leaves the definition of the cross product and the dot product unchanged, and turns this computation into the one for the second entry, and then the third.  $\square$

**Proposition 1.5.20** (Normal equation from a parametric plane). Let  $P$  be the plane through  $\mathbf{p}$  with directions  $\mathbf{u}, \mathbf{v} \in \mathbb{R}^3$ , and let  $\mathbf{n} = \mathbf{u} \times \mathbf{v}$ . Then  $\mathbf{n} \neq \mathbf{0}$  and

$$P = \{\mathbf{x} \in \mathbb{R}^3 : \mathbf{n} \cdot (\mathbf{x} - \mathbf{p}) = 0\}.$$

In particular every plane in  $\mathbb{R}^3$  has a normal equation.

*Proof.*  $\mathbf{n} \neq \mathbf{0}$  by Proposition 1.5.17, since  $\mathbf{u}$  and  $\mathbf{v}$  are not parallel. If  $\mathbf{x} = \mathbf{p} + s\mathbf{u} + t\mathbf{v}$ , then  $\mathbf{n} \cdot (\mathbf{x} - \mathbf{p}) = s\mathbf{n} \cdot \mathbf{u} + t\mathbf{n} \cdot \mathbf{v} = 0$  by Theorem 1.5.14(d).

Conversely, let  $\mathbf{y} = \mathbf{x} - \mathbf{p}$  satisfy  $\mathbf{n} \cdot \mathbf{y} = 0$ ; we show  $\mathbf{y} \in \text{Span}\{\mathbf{u}, \mathbf{v}\}$ . Put  $D = \|\mathbf{u}\|^2 \|\mathbf{v}\|^2 - (\mathbf{u} \cdot \mathbf{v})^2 = \|\mathbf{n}\|^2 > 0$  and

$$\alpha = \frac{(\mathbf{u} \cdot \mathbf{y}) \|\mathbf{v}\|^2 - (\mathbf{v} \cdot \mathbf{y})(\mathbf{u} \cdot \mathbf{v})}{D}, \quad \beta = \frac{(\mathbf{v} \cdot \mathbf{y}) \|\mathbf{u}\|^2 - (\mathbf{u} \cdot \mathbf{y})(\mathbf{u} \cdot \mathbf{v})}{D}.$$

A direct check shows  $\alpha \|\mathbf{u}\|^2 + \beta(\mathbf{u} \cdot \mathbf{v}) = \mathbf{u} \cdot \mathbf{y}$  and  $\alpha(\mathbf{u} \cdot \mathbf{v}) + \beta \|\mathbf{v}\|^2 = \mathbf{v} \cdot \mathbf{y}$ , so  $\mathbf{w} = \mathbf{y} - \alpha\mathbf{u} - \beta\mathbf{v}$  satisfies  $\mathbf{u} \cdot \mathbf{w} = 0$  and  $\mathbf{v} \cdot \mathbf{w} = 0$ . By Proposition 1.5.19,  $\mathbf{n} \times \mathbf{w} = (\mathbf{u} \cdot \mathbf{w})\mathbf{v} - (\mathbf{v} \cdot \mathbf{w})\mathbf{u} = \mathbf{0}$ , so  $\mathbf{w}$  is parallel to  $\mathbf{n}$  by Proposition 1.5.17; as  $\mathbf{n} \neq \mathbf{0}$ ,  $\mathbf{w} = \lambda\mathbf{n}$  for some  $\lambda$ . But also  $\mathbf{n} \cdot \mathbf{w} = \mathbf{n} \cdot \mathbf{y} - \alpha\mathbf{n} \cdot \mathbf{u} - \beta\mathbf{n} \cdot \mathbf{v} = 0$ , so  $\lambda \|\mathbf{n}\|^2 = 0$ ,  $\lambda = 0$  and  $\mathbf{w} = \mathbf{0}$ . Hence  $\mathbf{y} = \alpha\mathbf{u} + \beta\mathbf{v}$  and  $\mathbf{x} = \mathbf{p} + \alpha\mathbf{u} + \beta\mathbf{v} \in P$ .  $\square$

**Example 1.5.21.** Find a normal equation of the plane  $P$  through  $A = (1, 0, 0)$ ,  $B = (0, 2, 0)$ ,  $C = (0, 0, 3)$ .

*Solution.* By Example 1.5.13,  $\mathbf{n} = (\mathbf{b} - \mathbf{a}) \times (\mathbf{c} - \mathbf{a}) = (6, 3, 2)$ , and  $\mathbf{n} \cdot \mathbf{a} = 6$ . So  $P$  is  $6x + 3y + 2z = 6$ . Check:  $B$  gives 6 and  $C$  gives 6. The point  $(0, 1, \frac{3}{2})$  gives  $0 + 3 + 3 = 6$ , so it lies on  $P$ , as Example 1.5.7 found with more work. Dividing by 6 gives the intercept form  $\frac{x}{1} + \frac{y}{2} + \frac{z}{3} = 1$ , which shows the three points where  $P$  meets the axes.

The normal vector also gives the distance from a point to a plane: the shortest way from the point to the plane runs along the normal.

**Proposition 1.5.22** (Distance from a point to a plane). Let  $P$  be the plane  $\mathbf{n} \cdot \mathbf{x} = d$  in  $\mathbb{R}^3$  and let  $\mathbf{q} \in \mathbb{R}^3$ . The point  $\mathbf{f} = \mathbf{q} - \frac{\mathbf{n} \cdot \mathbf{q} - d}{\|\mathbf{n}\|^2} \mathbf{n}$  lies on  $P$ , and it is the point of  $P$  closest to  $\mathbf{q}$ : for every  $\mathbf{x} \in P$  with  $\mathbf{x} \neq \mathbf{f}$ ,  $\|\mathbf{q} - \mathbf{x}\| > \|\mathbf{q} - \mathbf{f}\|$ . The distance from  $\mathbf{q}$  to  $P$  is therefore

$$\|\mathbf{q} - \mathbf{f}\| = \frac{|\mathbf{n} \cdot \mathbf{q} - d|}{\|\mathbf{n}\|}.$$

*Proof.*  $\mathbf{n} \cdot \mathbf{f} = \mathbf{n} \cdot \mathbf{q} - (\mathbf{n} \cdot \mathbf{q} - d) = d$ , so  $\mathbf{f} \in P$ . Let  $\mathbf{x} \in P$ . Then  $\mathbf{q} - \mathbf{x} = (\mathbf{q} - \mathbf{f}) + (\mathbf{f} - \mathbf{x})$ , where  $\mathbf{q} - \mathbf{f}$  is a multiple of  $\mathbf{n}$  and  $\mathbf{n} \cdot (\mathbf{f} - \mathbf{x}) = d - d = 0$ . So the two pieces are orthogonal and, by Pythagoras (Theorem 1.4.14),  $\|\mathbf{q} - \mathbf{x}\|^2 = \|\mathbf{q} - \mathbf{f}\|^2 + \|\mathbf{f} - \mathbf{x}\|^2$ , which is larger than  $\|\mathbf{q} - \mathbf{f}\|^2$  unless  $\mathbf{x} = \mathbf{f}$ . Finally  $\|\mathbf{q} - \mathbf{f}\| = \frac{|\mathbf{n} \cdot \mathbf{q} - d|}{\|\mathbf{n}\|^2} \|\mathbf{n}\| = \frac{|\mathbf{n} \cdot \mathbf{q} - d|}{\|\mathbf{n}\|}$  by Proposition 1.4.7.  $\square$

The point  $\mathbf{f}$  is called the *foot of the perpendicular* from  $\mathbf{q}$  to  $P$ , and  $\mathbf{q} - \mathbf{f}$  is the *orthogonal projection* of  $\mathbf{q} - \mathbf{x}$  onto the normal direction, for any  $\mathbf{x} \in P$ . Projections return in Chapter 10. The same formula, with  $\mathbf{n} \in \mathbb{R}^2$ , gives the distance from a point to a line  $\mathbf{n} \cdot \mathbf{x} = d$  in the plane.

**Example 1.5.23.** Find the distance from the origin to the plane  $6x + 3y + 2z = 6$ , and the closest point.

*Solution.*  $\mathbf{n} = (6, 3, 2)$ ,  $\|\mathbf{n}\| = 7$  and  $\mathbf{n} \cdot \mathbf{0} - 6 = -6$ . The distance is  $\frac{|-6|}{7} = \frac{6}{7}$ , and the closest point is  $\mathbf{f} = \mathbf{0} + \frac{6}{49}(6, 3, 2) = (\frac{36}{49}, \frac{18}{49}, \frac{12}{49})$ . Check:  $6 \cdot \frac{36}{49} + 3 \cdot \frac{18}{49} + 2 \cdot \frac{12}{49} = \frac{216+54+24}{49} = \frac{294}{49} = 6$ . For the point  $(1, 1, 1)$ ,  $6 + 3 + 2 - 6 = 5$ , so its distance is  $\frac{5}{7}$ .

### Check yourself

Find a normal equation of the plane through  $A = (1, 0, 1)$ ,  $B = (2, 1, 1)$  and  $C = (1, 1, 2)$ . Check that all three points satisfy it, and find the area of the triangle  $ABC$ .

### Supplement: beyond the lecture

This material is not needed later in the course, but it rounds out the picture.

**Distance from a point to a line in  $\mathbb{R}^3$ .** Let  $L$  be the line  $\mathbf{p} + t\mathbf{d}$  and  $\mathbf{q}$  a point. The parallelogram with sides  $\mathbf{d}$  and  $\mathbf{q} - \mathbf{p}$  has area  $\|\mathbf{d} \times (\mathbf{q} - \mathbf{p})\|$  and base  $\|\mathbf{d}\|$ , so its height, which is the distance from  $\mathbf{q}$  to  $L$ , is

$$\frac{\|\mathbf{d} \times (\mathbf{q} - \mathbf{p})\|}{\|\mathbf{d}\|}.$$

For the line through  $(1, 0, 2)$  with direction  $(1, 2, -1)$  and  $\mathbf{q} = \mathbf{0}$ :  $\mathbf{q} - \mathbf{p} = (-1, 0, -2)$  and  $\mathbf{d} \times (\mathbf{q} - \mathbf{p}) = (2 \cdot (-2) - (-1) \cdot 0, (-1)(-1) - 1 \cdot (-2), 1 \cdot 0 - 2 \cdot (-1)) = (-4, 3, 2)$ , so the distance is  $\sqrt{29}/\sqrt{6} = \sqrt{29}/6 \approx 2.20$ .

**The scalar triple product.** The number  $\mathbf{w} \cdot (\mathbf{u} \times \mathbf{v})$  is called the *scalar triple product*. Its absolute value is the volume of the slanted box (parallelepiped) with edges  $\mathbf{u}, \mathbf{v}, \mathbf{w}$ : the base has area  $\|\mathbf{u} \times \mathbf{v}\|$ , and the height is the length of the component of  $\mathbf{w}$  along the normal  $\mathbf{u} \times \mathbf{v}$ , which is  $|\mathbf{w} \cdot (\mathbf{u} \times \mathbf{v})|/\|\mathbf{u} \times \mathbf{v}\|$ . The triple product is zero exactly when the three vectors lie in one plane through the origin. It is the  $3 \times 3$  determinant of Chapter 7 in disguise.

**Why only  $\mathbb{R}^3$ ?** A product of two vectors of  $\mathbb{R}^n$  that is again in  $\mathbb{R}^n$ , orthogonal to both factors, bilinear, and with length  $\|\mathbf{u}\| \|\mathbf{v}\| \sin \theta$  exists only for  $n = 3$  and  $n = 7$ , apart from the trivial case  $n = 1$ ; the seven-dimensional one is connected to the octonions. In  $\mathbb{R}^2$  the closest relative is the number  $u_1v_2 - u_2v_1$ , the signed area of the parallelogram on  $\mathbf{u}$  and  $\mathbf{v}$ , which is the third entry of  $(u_1, u_2, 0) \times (v_1, v_2, 0)$ .

## Answers to Check yourself

### Section 1.1

$x \begin{bmatrix} 3 \\ 1 \end{bmatrix} + y \begin{bmatrix} 1 \\ -2 \end{bmatrix} = \begin{bmatrix} 4 \\ 6 \end{bmatrix}$ . With  $(x, y) = (2, -2)$ :  $2 \begin{bmatrix} 3 \\ 1 \end{bmatrix} - 2 \begin{bmatrix} 1 \\ -2 \end{bmatrix} = \begin{bmatrix} 6 \\ 2 \end{bmatrix} + \begin{bmatrix} -2 \\ 4 \end{bmatrix} = \begin{bmatrix} 4 \\ 6 \end{bmatrix}$ , so it is a solution.

### Section 1.2

$\mathbf{x} = (2, -2, 2)$ . Add  $(2, 1, 0)$  to both sides (rules 2, 1, 4 and 3), so  $3\mathbf{x} = (6, -6, 6)$ ; then multiply by  $\frac{1}{3}$  (rules 6 and 5).

### Section 1.3

(a) No.  $(3, 6) = 3(1, 2)$ , so the span is the set of multiples  $(c, 2c)$ , and  $(2, 5)$  would need  $5 = 4$ . (b) Yes.  $ps - qr = 1 \cdot 5 - 2 \cdot 3 = -1 \neq 0$ , so by Theorem 1.3.8 the span is  $\mathbb{R}^2$ .

### Section 1.4

$\mathbf{v} \cdot \mathbf{w} = 2 - 4 + 2 = 0$  and  $\|\mathbf{v}\| = \|\mathbf{w}\| = 3$ , so the vectors are orthogonal.  $\mathbf{v} + \mathbf{w} = (3, 0, 3)$  and  $\|\mathbf{v} + \mathbf{w}\|^2 = 18 = 9 + 9 = \|\mathbf{v}\|^2 + \|\mathbf{w}\|^2$ .

### Section 1.5

$\mathbf{b} - \mathbf{a} = (1, 1, 0)$  and  $\mathbf{c} - \mathbf{a} = (0, 1, 1)$ , so  $\mathbf{n} = (1, 1, 0) \times (0, 1, 1) = (1, -1, 1)$  and the plane is  $x - y + z = 2$  (check:  $1 - 0 + 1 = 2$ ,  $2 - 1 + 1 = 2$ ,  $1 - 1 + 2 = 2$ ). The triangle's area is  $\frac{1}{2}\|\mathbf{n}\| = \frac{\sqrt{3}}{2}$ .