Points, lines, planes and more

Ronald van Luijk

Make your own pictures!

1. Lines in the plane, equations and parametrizations

1.1. **Equations.** Lines in the usual (x_1, x_2)-plane can be given in different ways. For instance by *equations* such as

$$\begin{aligned} L_1 &: x_2 = 2x_1 + 1, \\ L_2 &: x_2 = -4, \\ L_3 &: x_1 = 3, \\ L_4 &: 2x_1 + 7x_2 = -5, \\ L_5 &: 4x_1 + 14x_2 = -10, \\ L_6 &: 6x_1 - 3x_2 = -3. \end{aligned}$$

Note that lines L_4 and L_5 are the same and so are lines L_1 and L_6, so a line does not have a unique equation. Each line can be given by an equation of the form

$$L : a_1x_1 + a_2x_2 = c \tag{1}$$

for some $a_1, a_2, c \in \mathbb{R}$ with a_1 and a_2 not both equal to 0. For $a_1 = 0$ the line L is parallel to the x_1-axis and for $a_2 = 0$ it is parallel to the x_2-axis. More generally, if we fix a_1 and a_2, but vary c, then all the corresponding lines are parallel to each other. The line L goes through the origin $(0, 0)$ if and only if we have $c = 0$.

1.2. **Parametrizations.** Another way to describe lines is by *parametrizations*. For instance the line L_1 can be parametrized by t as

$$L_1 : \begin{cases} x_1 &= t, \\ x_2 &= 2t + 1. \end{cases}$$

This parametrization can be found by setting $x_1 = t$ and solving for x_2. Of course this does not work for L_3, where x_1 is fixed; in that case we can take $x_2 = t$. The other lines can be parametrized by t as

$$L_2 : \begin{cases} x_1 &= t, \\ x_2 &= -4, \end{cases} \qquad L_3 : \begin{cases} x_1 &= 3, \\ x_2 &= t, \end{cases} \qquad \text{and} \qquad L_4 : \begin{cases} x_1 &= -\frac{7}{2}t - \frac{5}{2}, \\ x_2 &= t. \end{cases}$$

Note that as equations, parametrizations are not unique either. Line L_1, for instance, may also be parametrized by s as

$$L_1 : \begin{cases} x_1 &= 3s - 2, \\ x_2 &= 6s - 3, \end{cases}$$

which can be checked by substituting $x_1 = 3s - 2$ and $x_2 = 6s - 3$ in any equation for L_1. Each line in the (x_1, x_2)-plane may be parametrized by t as

$$L : \begin{cases} x_1 &= p_1t + q_1, \\ x_2 &= p_2t + q_2 \end{cases} \tag{2}$$

for some $p_1, p_2, q_1, q_2 \in \mathbb{R}$ with p_1 and p_2 not both equal to 0.

1.3. **From equations to parametrizations and back.** We will now try to understand these parametrizations better. Suppose a line L is given by equation (1) for certain a_1, a_2, c. Then the line L is parametrized by (2) for $p_1, p_2, q_1, q_2 \in \mathbb{R}$ if and only if for all $t \in \mathbb{R}$ the equation

$$(a_1p_1 + a_2p_2)t + (a_1q_1 + a_2q_2) = c$$

holds. This is equivalent to the two equations

$$a_1p_1 + a_2p_2 = 0 \qquad \text{and} \qquad a_1q_1 + a_2q_2 = c$$

Remark 1.1. *The second equation says exactly that the point $Q = (q_1, q_2)$ is on the line L. The first equation says that the point $P = (p_1, p_2)$ is on the line L' through the origin that is parallel to L. In other words, finding a parametrization for L is equivalent to finding a point on L, as well as a point on L' other than the origin.*

Example 1.2. *The point $(-2, -3)$ lies on L_1. The line L_1' through the origin that is parallel to L_1 is given by $x_2 = 2x_1$. A point on L_1' is $(3, 6)$. This gives $p_1 = 3$, $p_2 = 6$, $q_1 = -2$, and $q_2 = -3$, which corresponds to the second parametrization for L_1 given above. Of course finding a point on a given line can again be done by choosing a fixed value for one of the coordinates and solving for the other coordinate.*

Conversely, to find an equation for the line M given by, say, the parametrization

$$\begin{cases} x_1 &= 3t + 5, \\ x_2 &= 4t - 3, \end{cases} \tag{3}$$

we just eliminate the parameter t; multiply the first equation by 4, the second by 3, and subtract the resulting equations to obtain the equation $4x_1 - 3x_2 = 29$ defining M.

1.4. **Vectors and the inner product.** We now introduce some new notation, which will allow us to write equations and parametrizations more compactly. The trick is to write just one symbol for a pair (x_1, x_2). We just did this by interpreting for instance the pair (p_1, p_2) as a point and writing P for it. You can do the same thing with any other interpretation, or better still, without any interpretation.

Definition 1.3. *In this section, a* vector *$v = (v_1, v_2)$ is a pair of real numbers. The real numbers v_1 and v_2 are called the* coefficients *of v. The set of all pairs of real numbers is denoted by $\mathbb{R}^2$. The* zero vector *$(0, 0)$ is also denoted by 0.*

Remark 1.4. *Be careful! In the definition above, the symbols v_1 and v_2 denote real numbers, the coefficients of a vector. In a different context these same symbols could denote vectors themselves. Similarly, the symbol 0 can denote both a real number and a vector. When reading mathematics, it is important always to ask yourself what each symbol refers to. For instance, if we have $p = (p_1, p_2)$, then the condition that p_1 and p_2 not both be 0 is equivalent to $p \neq 0$. In this statement, 0 denotes the zero vector.*

Definition 1.5. *The* inner product *of two vectors $v = (v_1, v_2)$ and $w = (w_1, w_2)$, denoted by $\langle v, w \rangle$, is defined as*

$$\langle v, w \rangle = v_1w_1 + v_2w_2.$$

Note that the inner product of two vectors is a real number, not a vector! Some books write the inner product with a dot (as $v \cdot w$), which is why it is also often called the *dot product*. This notation can easily cause confusion, so we will not use it.

Equation (1) can now be written as

$$\langle a, x\rangle = c$$

with $a = (a_1, a_2)$ and $x = (x_1, x_2)$. We often interpret points in the plane as vectors, i.e., pairs of real numbers, so that every line L in the plane corresponds to a subset of $\mathbb{R}^2$. For each line L there is a vector $a \neq 0$ (this is the **vector** 0) and a constant c such that L consists of those vectors x that satisfy $\langle a, x\rangle = c$.

Example 1.6. *The line L_4 is given by $\langle a, x\rangle = c$ for $a = (2, 7)$ and $c = -5$.*

1.5. **Sum and scalar multiplication.** Before we can also write the parametrizations more compactly, we need to define the *sum* of two vectors, as well as the *scalar product* of a *scalar* (an element of $\mathbb{R}$) and a vector.

Definition 1.7. *The* sum *of two vectors $v = (v_1, v_2)$ and $w = (w_1, w_2)$ is*

$$v + w = (v_1 + w_1, v_2 + w_2).$$

The scalar product *of a real number λ and a vector $v = (v_1, v_2)$ is*

$$\lambda \cdot v = (\lambda v_1, \lambda v_2).$$

Example 1.8. *We have $(2, 3) + (-7, 4) = (-5, 7)$ and $3 \cdot (2, 5) = (6, 15)$.*

Remark 1.9. *Also the symbol "+" now has a double meaning: the sum of numbers and the sum of vectors. From the context it should always be clear what is meant. For instance, we easily check that $0 + v = v$ for every vector v. Here 0 is the zero vector and the sum is a sum of vectors. This shows that the vector 0 behaves with respect to the sum as you would expect from the symbol "0."*

We often leave out the dot in the scalar product. If we write $p = (p_1, p_2)$ and $q = (q_1, q_2)$, then the parametrization (2) may also be written as

$$x = tp + q,$$

where q satisfies $\langle a, q\rangle = c$ and p satisfies $\langle a, p\rangle = 0$.

Example 1.10. *Take $a = (2, -3)$ and $c = 7$ and let L be the line consisting of all vectors x satisfying $\langle a, x\rangle = c$, so that L is given by $2x_1 - 3x_2 = 7$. One point on L is $q = (2, -1)$. The point $p = (3, 2)$ satisfies $\langle a, p\rangle = 0$ (and therefore lies on the line through 0 that is parallel to L). A parametrization by t for L is therefore given by $x = tp + q$, or, written in the old way,*

$$L : \begin{cases} x_1 = 3t + 2, \\ x_2 = 2t - 1. \end{cases}$$

1.6. **Exercises.**

Exercise 1.1. *Compute the inner product of the given vectors v and w.*

- $v = (-2, 5)$ *and* $w = (7, 1)$,
- $v = 2(-3, 2)$ *and* $w = (1, 3) + (-2, 4)$,
- $v = (-3, 4)$ *and* $w = (4, 3)$,
- $v = (-3, 4)$ *and* $w = (8, 6)$,

- $v = (2, -7)$ *and* $w = (x, y)$,
- $v = w = (a, b)$.

Exercise 1.2. *Write the following equations in the form* $\langle a, x\rangle = c$, *i.e., specify a vector* a *and a constant* c *in each case.*

(1) $L_1\colon 2x_1 + 3x_2 = 0$,
(2) $L_2\colon x_2 = 3x_1 - 1$,
(3) $L_3\colon 2(x_1 + x_2) = 3$,
(4) $L_4\colon x_1 - x_2 = 2x_2 - 3$,
(5) $L_5\colon x_1 = 4 - 3x_1$,
(6) $L_6\colon x_1 - x_2 = x_1 + x_2$.
(7) $L_7\colon 6x_1 - 2x_2 = 7$

Exercise 1.3. *Paramerize the lines in Exercise 1.2.*

Exercise 1.4. *Write the following parametrizations in the form* $x = tp + q$, *i.e., specify vectors* p *and* q *in each case.*

(1)
$$M_1\colon \begin{cases} x_1 &= t + 2, \\ x_2 &= -t - 3., \end{cases}$$
(2)
$$M_2\colon \begin{cases} x_1 &= 2t + 1, \\ x_2 &= 6t - 3, \end{cases}$$
(3)
$$M_3\colon \begin{cases} x_1 &= 3t - 1, \\ x_2 &= t, \end{cases}$$
(4)
$$M_4\colon \begin{cases} x_1 &= 3t + 5, \\ x_2 &= 7, \end{cases}$$
(5)
$$M_5\colon \begin{cases} x_1 &= t + 4, \\ x_2 &= t + 4, \end{cases}$$
(6)
$$M_6\colon \begin{cases} x_1 &= 3t, \\ x_2 &= 4t. \end{cases}$$

Exercise 1.5. *Give an equation for lines in Exercise 1.4.*

Exercise 1.6. *True or False? If true, explain why. If false, give a counter example.*

(1) *If* a *and* b *are nonzero vectors and* $a \neq b$, *then the lines given by* $\langle a, x\rangle = 0$ *and* $\langle b, x\rangle = 1$ *are not parallel.*
(2) *If* a *and* b *are nonzero vectors and the lines given by* $\langle a, x\rangle = 0$ *and* $\langle b, x\rangle = 1$ *are parallel, then* $a = b$.
(3) *Two different lines may be given by the same equation.*
(4) *Two different parametrizations may parametrize the same line.*
(5) *The intersection of two lines is either empty or consists of one point.*
(6) *For each vector* $v \in \mathbb{R}^2$ *we have* $0 \cdot v = 0$. *(What do the zeros in this statement refer to?)*

Exercise 1.7. *Which of all the lines in Exercises 1.2 and 1.4 are parallel?*

Exercise 1.8. *Compute the point of intersection of the following lines from Exercises 1.2 and 1.4, if it exists.*

(1) L_1 *and* L_2,
(2) L_2 *and* L_3,
(3) L_5 *and* M_4,
(4) L_4 *and* M_2,
(5) L_2 *and* M_2,
(6) M_2 *and* M_3.

Exercise 1.9. *When you intersect two lines, would you prefer them to be given by a parametrization or by an equation?*

Exercise 1.10. *Given the vectors* $p = (p_1, p_2)$ *and* $q = (q_1, q_2)$ *with* $p \neq 0$, *we let* L *be the line parametrized by* $x = tp + q$. *Find a vector* a *and a constant* c *(in terms of* p_1, p_2, q_1, q_2*) such that* L *is defined by the equation* $\langle a, x \rangle = c$.

Exercise 1.11. *Let* a *and* b *be two vectors and* c *and* d *two constants. Let* L *be the line given by* $\langle a, x \rangle = c$ *and* M *the line given by* $\langle b, x \rangle = d$. *Show that* M *and* L *are perpendicular if and only if* $\langle a, b \rangle = 0$.

Exercise 1.12. *Give an equation and a parametrization, both written in terms of vectors, of the following lines in* $\mathbb{R}^2$.

(1) *The line through the points* $(1, 2)$ *and* $(3, -5)$.
(2) *The line through the point* $(3, 2)$ *that is parallel to the line given by* $3x_1 - x_2 = 0$.
(3) *The line through the point* $(3, -2)$ *that is perpendicular to the line given by* $\langle a, x \rangle = -1$ *with* $a = (1, -2)$.
(4) *The line through the point* $(1, 4)$ *that is parallel to the line parametrized by* $x = pt + q$ *with* $p = (1, 2)$ *and* $q = (-3, 4)$.
(5) *The line through the point* $(1, 4)$ *that is perpendicular to the line parametrized by*

$$\begin{cases} x_1 &= 3t - 1, \\ x_2 &= 4t + 1. \end{cases}$$

(6) *The perpendicular bisector of the line segment from* $(-2, -1)$ *to* $(4, 3)$.
(7) *The line obtained by rotating the line given by* $\langle (3, 5), x \rangle = 2$ *around the origin over* $\pi/2$ *(so counter clockwise).*
(8) *(*) The line obtained by rotating the line given by* $\langle (3, 5), x \rangle = 2$ *around the origin over* α.

2. Higher dimensions, angles and distances

Although a lot more can be said about lines in the plane, we immediately start with higher dimensions.

2.1. Vectors, sums, scalar multiplication.

Definition 2.1. *The set* $\mathbb{R}^n$ *consists of all* n*-tuples of real numbers. Its elements are called* vectors. *We abbreviate the zero vector* $(0, 0, \ldots, 0)$ *to* 0.

Example 2.2. *The set* $\mathbb{R}^3$ *contains the vectors* $(1, -3, 2)$ *and* $(\pi, e, \sqrt{2})$. *The set* $\mathbb{R}^n$ *contains the vector* $(1, 2, 3, \ldots, n)$ *and* $\mathbb{R}^1$ *contains the vector* (17).

Later we will see a more general definition of *vector*. We now define the sum and scalar multiplication as in the two-dimensional case, namely componentwise.

Definition 2.3. *For any scalar $\lambda \in \mathbb{R}$ and any vectors $v = (v_1, v_2, \ldots, v_n)$ and $w = (w_1, w_2, \ldots, w_n)$ in $\mathbb{R}^n$, the scalar product of λ and v is*

$$\lambda \cdot v = (\lambda v_1, \lambda v_2, \ldots, \lambda v_n)$$

and the sum of v and w is

$$v + w = (v_1 + w_1, v_2 + w_2, \ldots, v_n + w_n).$$

Sometimes we will leave out the dot in the scalar multiplication.

Note that for all $v \in \mathbb{R}^n$ we have $0 \cdot v = 0$ and $0 + v = v$. (Which 0 means what???) Note that also we have $v + (-1) \cdot v = 0$, so $-v$ is the *negative* of v and we write $-v$ for $(-1) \cdot v$. In fact, we also write $v - w$ for $v + (-w)$, so that we have

$$v - w = (v_1 - w_1, v_2 - w_2, \ldots, v_n - w_n).$$

2.2. **Interpretation of vectors.** We can interpret $\mathbb{R}^n$ in many different ways. For $n = 2$, we can identify the vectors with points in the (x_1, x_2)-plane; here the real numbers in the pair correspond with the coordinates of the point. We can similarly identify $\mathbb{R}^3$ with three-dimensional space. We will often make these identifications and talk about points as if they are vectors. By doing so, we can now add points in the plane, as well as in space!

In physics, more precisely in relativity theory, $\mathbb{R}^4$ is often interpreted as space with a fourth coordinate for time.

For $n = 2$ or $n = 3$, we may also interpret vectors as arrows in the plane or space, respectively. In the plane, the arrow from the point $p = (p_1, p_2)$ to the point $q = (q_1, q_2)$ represents the vector $v = (q_1 - p_1, q_2 - p_2) = q - p$. (A careful reader notes that here we do indeed identify points and vectors.) We say that the point p is the tail of the arrow and the point q is the head. Note the distinction we make between an arrow and a vector, the latter of which is by definition just a sequence of real numbers. Many different arrows may represent the same vector v, but all these arrows have the same direction and the same length, which together narrow down the vector. One arrow is special, namely the one with 0 as its tail; the head of this arrow is precisely the point $q - p$! Of course we can do the same for $\mathbb{R}^3$.

Example 2.4. *Take the two points $p = (3, 1, -4)$ and $q = (-1, 2, 1)$ and set $v = q - p$. Then we have $v = (-4, 1, 5)$. The arrow from p to q has the same direction and length as the arrow from 0 to the point $(-4, 1, 5)$. Both these arrows represent the vector v.*

We can now interpret negation, scalar multiples, sums, and differences of vectors geometrically, namely in terms of arrows. Make your own pictures! If a vector v corresponds to a certain arrow, then $-v$ corresponds to any arrow with the same length but opposite direction; more generally, for $\lambda \in \mathbb{R}$ the vector λv corresponds to the arrow obtained by scaling the arrow for v by a factor λ.

If v and w correspond to two arrows that have common tail p, then these two arrows are the sides of a unique parallelogram; the vector $v + w$ corresponds to a diagonal in this parallelogram, namely the arrow that also has p as tail and whose head is the opposite point in the parallelogram. Another construction for $v + w$ is to take two arrows, for which the head of the one representing v equals the tail of the one representing w; then $v + w$ corresponds to the arrow from the tail of the first to the head of the second. Compare the two constructions in a picture!

For the same v and w, still with common tail and with heads q and r respectively, the difference $v - w$ corresponds to the other diagonal in the same parallelogram, namely the arrow from r to q. Another construction for $v - w$ is to write this difference as the sum $v + (-w)$, which can be constructed as above. Make a picture again!

2.3. **Translation.** We define translations as follows.

Definition 2.5. *For every vector $v \in \mathbb{R}^n$, the translation by v, written as T_v, is the map from $\mathbb{R}^n$ to itself that adds v to each vector. In other words, we have*

$$T_v \colon \mathbb{R}^n \to \mathbb{R}^n, x \mapsto x + v.$$

Example 2.6. *Take $v = 0$, then the translation T_0 is the identity.*

Example 2.7. *For $v = (1, -3) \in \mathbb{R}^2$, the translation $T_v \colon \mathbb{R}^2 \to \mathbb{R}^2$ sends the vector (x_1, x_2) to $(x_1 + 1, x_2 - 3)$.*

2.4. **Inner product.** Similar to $\mathbb{R}^2$, we define the inner product as follows.

Definition 2.8. *The inner product of two vectors $v = (v_1, v_2, \ldots, v_n)$ and $w = (w_1, w_2, \ldots, w_n)$ in $\mathbb{R}^n$, written $\langle v, w\rangle$, equals*

$$v_1 w_1 + v_2 w_2 + \ldots + v_n w_n.$$

Note that the inner product is an element of $\mathbb{R}$.

Example 2.9. *The inner product of the vectors $v = (3, 1, 2, 0)$ and $w = (0, 2, -3, 1)$ is $3 \cdot 0 + 1 \cdot 2 + 2 \cdot (-3) + 0 \cdot 1 = -4$.*

Again, some books write the inner product with a dot (as $v \cdot w$), which is why it is also often called the *dot product*. We will not do this. The inner product satisfies the following properties.

Proposition 2.10. *Let $\lambda \in \mathbb{R}$ be a scalar and $u, v, w \in \mathbb{R}^n$ three vectors. Then*

1. $\langle v, w\rangle = \langle w, v\rangle$,
2. $\langle u + v, w\rangle = \langle u, w\rangle + \langle v, w\rangle$,
3. $\langle \lambda v, w\rangle = \lambda \cdot \langle v, w\rangle$,
4. $\langle v, v\rangle \geq 0$ *with equality if and only if* $v = 0$.

Proof. Property (1) is obvious. For property (2), write u, v, w as $u = (u_1, u_2, \ldots, u_n)$, $v = (v_1, v_2, \ldots, v_n)$, and $w = (w_1, w_2, \ldots, w_n)$. Then we have

$$u + v = (u_1 + v_1, u_2 + v_2, \ldots, u_n + v_n),$$

so

$$\begin{aligned} \langle u + v, w\rangle &= (u_1 + v_1)w_1 + (u_2 + v_2)w_2 + \ldots + (u_n + v_n)w_n \\ &= (u_1 w_1 + \ldots + u_n w_n) + (v_1 w_1 + \ldots + v_n w_n) = \langle u, w\rangle + \langle v, w\rangle \end{aligned}$$

Property (3) can be proved similarly and is left as an exercise. For property (4), note that we have

$$\langle v, v\rangle = v_1^2 + v_2^2 + \ldots + v_n^2.$$

Since every term is nonnegative, this is nonnegative; the sum equals 0 if and only if each term is 0, so if and only if $v = 0$. □

Note that from properties (1) and (2) we also conclude that $\langle u, v+w\rangle = \langle u, v\rangle + \langle u, w\rangle$, and from (1) and (3) that $\langle v, \lambda w\rangle = \lambda \cdot \langle v, w\rangle$. Properties (2) and (3), together with these two similar properties, mean that the inner product is *bilinear*.

Note that from the properties above it also follows that $\langle u, v-w\rangle = \langle u, v\rangle - \langle u, w\rangle$ for all vectors $u, v, w \in \mathbb{R}^n$.

2.5. **Length.**

Proposition 2.11. *Let v be a vector in $\mathbb{R}^2$ or $\mathbb{R}^3$. Then the length of any arrow representing v equals $\sqrt{\langle v, v\rangle}$.*

Proof. Suppose $v = (v_1, v_2, v_3) \in \mathbb{R}^3$. Without loss of generality, we may assume the arrow representing v goes from 0 to $(v_1, v_2, v_3) = v$. By Pythagoras, the square of the length of this arrow equals $v_1^2 + v_2^2 + v_3^2 = \langle v, v\rangle$. The proof for $v \in \mathbb{R}^2$ is similar. □

Inspired by Proposition 2.11, we define the length of a vector in $\mathbb{R}^n$ for any n.

Definition 2.12. *The* length *or* norm *of a vector $v \in \mathbb{R}^n$, written $\|v\|$, is $\sqrt{\langle v, v\rangle}$.*

Example 2.13. *The length of the vector $(1, -2, 2, 3)$ in $\mathbb{R}^4$ equals $\sqrt{1+4+4+9} = 3\sqrt{2}$.*

Lemma 2.14. *For all vectors $v \in \mathbb{R}^n$ and all scalars $\lambda \in \mathbb{R}$ we have $\|\lambda v\| = |\lambda| \cdot \|v\|$.*

Proof. This is an exercise. □

2.6. **Inequalities, angles, and orthogonality.** Besides the notion of length that we now have also defined for higher dimension, we want to define the angle between two vectors as well. First, we will see what angles satisfy in dimension 2 and 3, where the angle between two vectors refers to the angle between two arrows representing these vectors with a common tail.

Proposition 2.15. *Let v, w be two nonzero vectors, both in $\mathbb{R}^2$ or both in $\mathbb{R}^3$, and let θ be the angle between these vectors. Then we have*

$$\cos\theta = \frac{\langle v, w\rangle}{\|v\| \cdot \|w\|}.$$

Proof. The vectors $v, w, v-w$ (or rather some arrows representing them) form a triangle with sides of length $\|v\|$, $\|w\|$, and $\|v-w\|$. The cosine formula gives

$$\|v-w\|^2 = \|v\|^2 + \|w\|^2 - 2\|v\| \cdot \|w\| \cdot \cos\theta,$$

while by the properties of the inner product we also have

$$\|v-w\|^2 = \langle v-w, v-w\rangle = \langle v, v\rangle - \langle v, w\rangle - \langle w, v\rangle + \langle w, w\rangle = \|v\|^2 + \|w\|^2 - 2\langle v, w\rangle.$$

Comparing the two equations, we find the claimed equality. □

If we say that the zero vector is by definition orthogonal to every vector, then we find the following Corollary.

Corollary 2.16. *Two vectors v, w in $\mathbb{R}^2$ or $\mathbb{R}^3$ are orthogonal if and only if $\langle v, w\rangle = 0$.*

Proof. If one of the vectors is 0, then this is obvious, as 0 is orthogonal to everything and the inner product is automatically 0. Assume $v, w \neq 0$, and let θ be the angle between v and w. Then v and w are orthogonal if and only if $\cos\theta = 0$, so if and only if $\langle v, w\rangle = 0$. □

Inspired by Corollary 2.16, we define orthogonality in all dimensions.

Definition 2.17. *Two vectors v, w in $\mathbb{R}^n$ are said to be* orthogonal*, written $v \perp w$, when $\langle v, w\rangle = 0$. Instead of orthogonal, people often say* perpendicular*.*

With this definition, Pythagoras holds in any dimension.

Proposition 2.18. *Two vectors $v, w \in \mathbb{R}^n$ are orthogonal if and only if $\|v+w\|^2 = \|v\|^2 + \|w\|^2$.*

Proof. We have

$$\|v + w\|^2 = \langle v + w, v + w\rangle = \|v\|^2 + \|w\|^2 + 2\langle v, w\rangle,$$

so the equality $\|v + w\|^2 = \|v\|^2 + \|w\|^2$ holds if and only if $\langle v, w\rangle = 0$. □

The following proposition shows that every vector v can be decomposed as the sum of two vectors, one in the same (or opposite) direction as a given vector w, and one orthogonal to it.

Proposition 2.19. *For every vectors $v, w \in \mathbb{R}^n$ there are unique vectors $v_1, v_2 \in \mathbb{R}^n$ such that $v = v_1 + v_2$ and $v_2 \perp w$ and $v_1 = \lambda w$ for some scalar $\lambda \in \mathbb{R}$. In case $w \neq 0$, this scalar equals $\langle v, w\rangle/\|w\|^2$.*

Proof. For $w = 0$ this is obvious, as we have $v_1 = 0$ and $v_2 = v$. Suppose $w \neq 0$. Then $v_2 = v - v_1$ with $v_1 = \lambda w$ is orthogonal to w if and only if $0 = \langle v - \lambda w, w\rangle = \langle v, w\rangle - \lambda\langle w, w\rangle$, so if and only if $\lambda = \langle v, w\rangle/\|w\|^2$. □

We call v_1 the *projection* of v onto w, or onto the line of multiples of w.

Proposition 2.20 (Cauchy-Schwarz)**.** *For every two vectors $v, w \in \mathbb{R}^n$ we have $|\langle v, w\rangle| \leq \|v\| \cdot \|w\|$. Equality holds if and only if there are scalars $\mu, \nu \in \mathbb{R}$, not both 0, such that $\mu v + \nu w = 0$.*

Proof. Given the vectors $v, w \in \mathbb{R}^n$, let $v_1, v_2 = \lambda w \in \mathbb{R}^n$ be as in Proposition 2.19. Then we have

$$|\langle v, w\rangle| = |\langle v_1 + v_2, w\rangle| = |\langle v_1, w\rangle| = |\langle \lambda w, w\rangle| = |\lambda| \cdot \|w\|^2 = \|v_1\| \cdot \|w\| \leq \|v\| \cdot \|w\|,$$

where the last equality follows from Lemma 2.14 and the last inequality from $\|v\|^2 = \|v_1\|^2 + \|v_2\|^2 \geq \|v_1\|^2$. Equality holds if and only if $\|w\| = 0$ or $\|v_2\| = 0$, so if and only if $w = 0$ or $v_2 = 0$. The latter case is equivalent to $v = v_1$, so $v = \lambda w$. Clearly in both cases there are μ, ν as claimed. Conversely, if there are μ, ν, not both zero, with $\mu v + \nu w = 0$, then either $\mu = 0$, in which case $w = 0$, or we have $v = \lambda w$ with $\lambda = -\nu/\mu$. □

Note that in the proof of Proposition 2.20, we see that the condition that there exist scalars $\mu, \nu \in \mathbb{R}$, not both 0, such that $\mu v + \nu w = 0$, is equivalent to the condition that $w = 0$ or v is a scalar multiple of w. The condition used in Proposition 2.20 has the benefit of being more symmetric. It is also easier to generalize, which we will do later when we define *linear combinations*.

We can now define the angle between two nonzero vectors.

Definition 2.21. *Let v, w be two nonzero vectors in $\mathbb{R}^n$. Then the* angle *between these vectors is defined to be that $\theta \in [0, \pi]$ for which we have*

$$\cos\theta = \frac{\langle v, w\rangle}{\|v\| \cdot \|w\|}.$$

Note that the fraction in Definition 2.21 lies in the interval $[-1, 1]$ by Proposition 2.20, so that θ is well defined.

Example 2.22. *For $v = (3, 0)$ and $w = (2, 2)$ in $\mathbb{R}^2$ we have $\langle v, w\rangle = 6$, while $\|v\| = 3$ and $\|w\| = 2\sqrt{2}$. Therefore, the angle θ between v and w satisfies $\cos\theta = 6/(3 \cdot 2\sqrt{2}) = \frac{1}{2}\sqrt{2}$, so we have $\theta = \pi/4$.*

Example 2.23. *For $v = (1, 1, 1, 1)$ and $w = (1, 2, 3, 4)$ in $\mathbb{R}^4$ we have $\langle v, w\rangle = 10$, while $\|v\| = 2$ and $\|w\| = \sqrt{30}$. Therefore, the angle θ between v and w satisfies $\cos\theta = 10/(2 \cdot \sqrt{30}) = \frac{1}{6}\sqrt{30}$, so $\theta = \arccos\left(\frac{1}{6}\sqrt{30}\right)$.*

Proposition 2.24 (Triangle inequality)**.** *For any vectors $v, w \in \mathbb{R}^n$ we have*

$$\|v + w\| \leq \|v\| + \|w\|.$$

Proof. By Proposition 2.20 we have

$$\begin{aligned}\|v + w\|^2 &= \langle v + w, v + w\rangle = \langle v, v\rangle + \langle w, w\rangle + 2\langle v, w\rangle \\ &\leq \|v\|^2 + \|w\|^2 + 2\|v\| \cdot \|w\| = (\|v\| + \|w\|)^2.\end{aligned}$$

Since all lengths are nonnegative, we may take square roots and conclude the claimed inequality holds. □

2.7. **Hyperplanes and lines.** Just as we have planes in $\mathbb{R}^3$, we have hyperplanes in $\mathbb{R}^n$.

Definition 2.25. *Let $a \in \mathbb{R}^n$ be a nonzero vector and $b \in \mathbb{R}$ a constant. Then the solution set $V \subset \mathbb{R}^n$ of the equation $\langle a, x\rangle = b$ is called a* hyperplane.

Example 2.26. *Lines in $\mathbb{R}^2$ and planes in $\mathbb{R}^3$ are hyperplanes.*

Example 2.27. *The equation $3x_1 - x_2 + 2x_3 - x_4 = 3$ defines a hyperplane in $\mathbb{R}^4$, given by $\langle a, x\rangle = b$ with $a = (3, -1, 2, -1)$ and $b = 3$.*

Lemma 2.28. *The hyperplane given by $\langle a, x\rangle = b$ goes through 0 if and only if $b = 0$.*

Proof. The hyperplane goes through 0 if and only if $x = 0$ is a solution to the equation, so if and only if $b = \langle a, 0\rangle = 0$. □

Now let's see what translation (see Definition 2.5) does to hyperplanes.

Proposition 2.29. *Let $H \subset \mathbb{R}^n$ be the hyperplane given by $\langle a, x\rangle = b$ for some vector $a \in \mathbb{R}^n$ and some constant $b \in \mathbb{R}$, and let $v \in \mathbb{R}^n$ be a vector. Then the translation T_v by v sends H to the hyperplane $H' = T_v(H)$ given by $\langle a, x - v\rangle = b$, or equivalently, by $\langle a, x\rangle = b'$ with $b' = \langle a, v\rangle + b$.*

Proof. This follows from the fact that the point x_0 satisfies $\langle a, x_0\rangle = b$ if and only if the point $x_0' = x_0 + v = T_v(x_0)$ satisfies $\langle a, x_0' - v\rangle = b$. □

Note the minus-sign in the new equation $\langle a, x - v\rangle = b$ for H'! Indeed, this is consistent with the following example.

Example 2.30. *Let $L \in \mathbb{R}^2$ with coordinates x, y be given by $y = 3x$. Then the line L' obtained by translating L by $(2, 5)$, i.e., 2 to the right and 5 up, is given by $(y - 5) = 3(x - 2)$.*

Example 2.31. *Let $H \subset \mathbb{R}^n$ be the hyperplane given by $\langle a, x\rangle = b$ for some vector $a \in \mathbb{R}^n$ and some constant $b \in \mathbb{R}$. Suppose the point p is contained in H, so that we have $\langle a, p\rangle = b$. Then translation by $-p$ sends p to 0. Indeed, the hyperplane $H' = T_{-p}(H)$ goes through 0, as it is given by $\langle a, x-(-p)\rangle = b$, or equivalently, by $\langle a, x+p\rangle = b$, or equivalently, by $\langle a, x\rangle = b'$ with $b' = \langle a, -p\rangle + b = -b + b = 0$.*

Remark 2.32. *Note that translation only affects the constant term b. So without knowing exactly by which vector to translate, we can still translate any hyperplane H to a hyperplane H' that contains 0. If H is given by $\langle a, x\rangle = b$, then H' is given by $\langle a, x\rangle = 0$ by Lemma 2.28, cf. Example 2.31.*

While the constant b in an equation $\langle a, x\rangle = b$ determines which translate of the hyperplane H given by $\langle a, x\rangle = 0$ we are considering, the vector a determines the direction.

Definition 2.33. *Let $H \subset \mathbb{R}^n$ be the hyperplane given by $\langle a, x\rangle = b$ for some vector $a \in \mathbb{R}^n$ and some constant $b \in \mathbb{R}$. Then a is called a* normal *of H.*

Note that two equations $\langle a, x\rangle = b$ and $\langle a', x\rangle = b'$ determine the same hyperplane H if and only if the equations are multiples of each other. This implies that all normal vectors of a given hyperplane are scalar multiples of each other.

Remark 2.34. *Suppose a is a normal of the hyperplane H. Then for any $p, q \in H$ we have $\langle a, p\rangle = \langle a, q\rangle$, so $\langle a, q-p\rangle = 0$, so a is orthogonal to $q-p$. This means that a is orthogonal to all arrows that lie within H. We also say that H itself is orthogonal to a.* **Be careful:** *this does not mean that the elements of H are orthogonal to a! Draw a picture.*

Example 2.35. *The plane V in $\mathbb{R}^3$ given by $2x_1 - x_2 + 3x_2 = 1$ has normal $a = (2, -1, 3)$. The arrows within V are all orthogonal to a. Any point in V, however, has by definition inner product 1 with a, so it is not orthogonal to a. For instance, the points $p = (-1, 0, 1)$ and $q = (1, 1, 0)$ lie in V; the corresponding vectors are not orthogonal to a, but their difference $q - p = (2, 1, -1)$ corresponds to an arrow in V, namely $q - p$, that is orthogonal to a.*

A line in $\mathbb{R}^3$ can be given by two equations. Similarly, a line in $\mathbb{R}^n$ can be given by $n-1$ equations, or equivalently, as the intersection of $n-1$ hyperplanes (see Section 3). It is therefore much easier to give a line by a parametrization. In fact, this can be taken as the definition of a line.

Definition 2.36. *Let $p, w \in \mathbb{R}^n$ be vectors with $w \neq 0$. Then the set of all vectors of the form $p + \lambda w$ for some $\lambda \in \mathbb{R}$ is called a* line.

Example 2.37. *The line through the points p and q in $\mathbb{R}^n$, with $p \neq q$, consists of all vectors of the form $p + \lambda(q-p)$ for some $\lambda \in \mathbb{R}$. For $\lambda = 1$ we get the point q. Note how we switch between* vectors *and* points *without notification.*

Example 2.38. *The line L through the points 0 and $w \neq 0$ in $\mathbb{R}^n$ consists of all multiples λw of w. In the case $w \neq 0$, Proposition 2.19 can be phrased by stating that for every vector $v \in \mathbb{R}^n$ there are unique $v_1, v_2 \in \mathbb{R}^n$ with $v = v_1 + v_2$ and $v_1 \in L$ and $v_2 \perp L$. We also call v_1 the* projection *of v onto L.*

Proposition 2.39. *Given $p, w \in \mathbb{R}^n$. Suppose the line L consists of all vectors of the form $p + \lambda w$ for some $\lambda \in \mathbb{R}$. Then for every $v \in \mathbb{R}^n$, the translation $L' = T_v(L)$ of L by v consists of all vectors of the form $p' + \lambda w$ for some $\lambda \in \mathbb{R}$, with $p' = p + v$.*

Proof. This is clear. □

Note that translation only affects the constant vector in the parametrization. Note also that for the new parametrization of L' we do *not* get a minus-sign in $p' = p + v$!

Proposition 2.40. *Given $p, w \in \mathbb{R}^n$ with $w \neq 0$. Suppose the line L consists of all vectors of the form $p + \lambda w$ for some $\lambda \in \mathbb{R}$. Also given $a \in \mathbb{R}^n$ and $b \in \mathbb{R}$, with $a \neq 0$, and the hyperplane H given by $\langle a, x\rangle = b$. Then exactly one of the following three statements about the intersection $H \cap L$ is true.*

(1) *The line L is contained in H; in this case we have $\langle a, w\rangle = 0$ and $\langle a, p\rangle = b$ (i.e., $p \in H$).*
(2) *The line L and hyperplane H do not intersect; in this case we have $\langle a, w\rangle = 0$ and $\langle a, p\rangle \neq b$ (i.e., $p \notin H$).*
(3) *The line L and H intersect in exactly one point; in this case we have $\langle a, w\rangle \neq 0$ and the intersection point is $p + \lambda w$ with*

$$\lambda = \frac{b - \langle a, p\rangle}{\langle a, w\rangle}.$$

Proof. The intersection of the line L and the hyperplane H consists of all points of the form $p + \lambda w$ that are contained in H, i.e., that satisfy $\langle a, p + \lambda w\rangle = b$, or equivalently, $\langle a, p\rangle + \lambda\langle a, w\rangle = b$. When $\langle a, w\rangle = 0$, this holds for either all or no λ depending on whether $\langle a, p\rangle = b$; this covers the first two cases. When $\langle a, w\rangle \neq 0$, the equation only holds for λ as given. □

The following definition states when a line is orthogonal to a hyperplane. We consider the *arrows* lying within the hyperplane and arrows lying within the line. Recall that an arrow with head q and tail p corresponds with the vector $q - p$.

Definition 2.41. *We say that a hyperplane H and a line L are orthogonal when for all $p, q \in H$ and for all $x, y \in L$ we have $(q-p) \perp (y-x)$, i.e., $\langle q-p, y-x\rangle = 0$.*

We actually do not need to let all the vectors x, y, p, q vary in Definition 2.41. We would get an equivalent definition if we fixed vectors p_0 and x_0.

Proposition 2.42. *Suppose H is a hyperplane and L a line and $p_0 \in H$ and $x_0 \in L$ vectors. Then H and L are orthogonal if and only if for all $q \in H$ and all $y \in L$ we have $\langle q - p_0, y - x_0\rangle = 0$.*

Proof. It suffices to show that if we have $\langle q - p_0, y - x_0\rangle = 0$ for all $q \in H$ and all $y \in L$, then we have $\langle q - p, y - x\rangle = 0$ for all $p, q \in H$ and all $x, y \in L$. This follows from the fact that

$$\begin{aligned}\langle q - p, y - x\rangle &= \langle (q - p_0) - (p - p_0), (y - x_0) - (x - x_0)\rangle \\ &= \langle q - p_0, y - x_0\rangle - \langle q - p_0, x - x_0\rangle - \langle p - p_0, y - x_0\rangle + \langle p - p_0, x - x_0\rangle \\ &= 0 - 0 - 0 + 0 = 0.\end{aligned}$$

□

Corollary 2.43. *Suppose H is a hyperplane and L a line, both containing 0. Then H and L are orthogonal if and only if for all $q \in H$ and all $y \in L$ we have $\langle q, y\rangle = 0$.*

Proof. This follows immediately from Proposition 2.42 with $x_0 = p_0 = 0$. □

We will now define the distance from a point to a line or a hyperplane. We will assume that the line or hyperplane contains 0; the general case is easily reduced to this case by an appropriate translation.

Definition 2.44. *Let $H \subset \mathbb{R}^n$ and $L \subset \mathbb{R}^n$ be a hyperplane and a line, orthogonal to each other, and both containing 0. Given a point $v \in \mathbb{R}^n$, let $v_1 \in L$ and $v_2 \in H$ be such that $v = v_1 + v_2$ (see Proposition 2.19 and Example 2.38). Then the* distance *$d(v, L)$ from v to L is defined to be $\|v_2\|$. The* distance *$d(v, H)$ from v to H is defined to be $\|v_1\|$.*

Remark 2.45. *Let H and L be as in Definition 2.44, just as v, v_1, and v_2. Let x be any point on L. Then we have $\|v - x\| \geq \|v - v_1\| = \|v_2\|$. This follows easily from Pythagoras after writing $v - x$ as the sum $(v - v_1) + (v_1 - x)$ of two orthogonal vectors. This shows that the distance $d(v, L) = \|v_2\|$ is the minimal distance from v to any point on L. Similarly, the distance $d(v, H) = \|v_1\|$ is the minimal distance from v to any point on H. Make a picture to support all these arguments!*

Example 2.46. *Let L be the line in $\mathbb{R}^3$ consisting of all multiples of $w = (1, 2, -1)$. Take the vector $v = (3, 0, -1)$. By Proposition 2.19, we can write $v = v_1 + v_2$ with $v_2 \perp w$ and $v_1 = \lambda w$ with $\lambda = \langle v, w\rangle / \|w\|^2$. Then we have*

$$v_2 = v - v_1 = v - \frac{\langle v, w\rangle}{\|w\|^2} w.$$

From $\langle v, w\rangle = 4$ and $\|w\|^2 = 6$, we find $v_2 = (\frac{7}{3}, -\frac{4}{3}, -\frac{1}{3})$, so

$$d(v, L) = \|v_2\| = \tfrac{1}{3}\sqrt{49 + 16 + 1} = \tfrac{1}{3}\sqrt{66}.$$

Example 2.47. *Let $H \subset \mathbb{R}^3$ be the plane given by $x_1 - 2x_2 + x_3 = 3$ and take the point $p = (1, 2, -1)$. To find the distance $d(p, H)$, we first translate everything so that the new plane contains 0. Note that H contains the vector $v = (1, -1, 0)$. Set $H' = T_{-v}(H)$. Then H' is given by $x_1 - 2x_2 + x_3 = 0$. Set $p' = T_{-v}(p) = (0, 3, -1)$. The vector $a = (1, -2, 1)$ is a normal of H' (and of H). Write $p' = p_1 + p_2$ with p_1 a multiple of a and $p_2 \perp a$, i.e., $p_2 \in H'$. The by Proposition 2.19 we have $p_1 = \lambda a$ with*

$$\lambda = \frac{\langle p', a\rangle}{\|a\|^2} = \frac{7}{6},$$

so $p_1 = \frac{7}{6}(1, -2, 1)$. We find $d(p, H) = d(p', H') = \|p_1\| = \frac{7}{6}\sqrt{6}$.

2.8. **Exercises.**

Exercise 2.1. *Prove property (3) of Proposition 2.10.*

Exercise 2.2. *Check for each definition, lemma, and proposition in this section whether its proof makes use of the explicit coefficients $x_1, x_2, \ldots x_n$ that a vector $x = (x_1, x_2, \ldots x_n)$ consist of, or of more efficient rules proved before.*

Exercise 2.3. *Using the properties of Proposition 2.10, show that the following properties also hold for all $\lambda \in \mathbb{R}$ and all $u, v, w \in \mathbb{R}^n$, as was already claimed.*

1. $\langle u, v + w\rangle = \langle u, v\rangle + \langle u, w\rangle$,
2. $\langle v, \lambda w\rangle = \lambda \cdot \langle v, w\rangle$,
3. $\langle u, v - w\rangle = \langle u, v\rangle - \langle u, w\rangle$,
4. $\langle u - v, w\rangle = \langle u, w\rangle - \langle v, w\rangle$.

Exercise 2.4. *Prove Lemma 2.14.*

Exercise 2.5. *Explain why Proposition 2.24 might be called the triangle inequality, which usually refers to* $c \leq a+b$ *for the sides* a, b, c *of a triangle. Prove that for all* $v, w \in \mathbb{R}^n$ *we have* $\|v-w\| \leq \|v\| + \|w\|$.

Exercise 2.6. *Prove the cosine rule in* $\mathbb{R}^n$.

Exercise 2.7. *Show that two vectors* $v, w \in \mathbb{R}^n$ *are have the same length if and only* $v-w$ *and* $v+w$ *are orthogonal.*

Exercise 2.8. *Prove that the diagonals of a parallelogram are orthogonal to each other if and only if all sides have the same length.*

Exercise 2.9. *Compute the distance from the point* $(1,1,1,1) \in \mathbb{R}^4$ *to the line consisting of all scalar multiples of* $(1,2,3,4)$.

Exercise 2.10. *Given the vectors* $p = (1,2,3)$ *and* $w = (2,1,5)$*, let* L *be the line consisting of all points of the form* $p + \lambda w$ *for some* $\lambda \in \mathbb{R}$*. Compute the distance* $d(v, L)$ *for* $v = (2,1,3)$.

Exercise 2.11. *Let* $H \subset \mathbb{R}^4$ *be the hyperplane with normal* $a = (1,-1,1,-1)$ *going though the point* $q = (1,2,-1,-2)$*. Determine the distance from the point* $(2,1,-3,1)$ *to* H.

Exercise 2.12. *Given* $p, w \in \mathbb{R}^n$ *with* $w \neq 0$*. Suppose the line* L *consists of all vectors of the form* $p + \lambda w$ *for some* $\lambda \in \mathbb{R}$*. Show that* L *goes through* 0 *if and only if* $p = \lambda w$ *for some* $\lambda \in \mathbb{R}$.

Exercise 2.13. *The angle between two hyperplanes is defined as the angle between their normal vectors. Determine the angle between the hyperplanes in* $\mathbb{R}^4$ *given by* $x_1 - 2x_2 + x_3 - x_4 = 2$ *and* $3x_1 - x_2 + 2x_3 - 2x_4 = -1$*, respectively.*

Exercise 2.14. *Let* $H \subset \mathbb{R}^n$ *be a hyperplane containing* 0 *with normal* a*. Let* $p \in \mathbb{R}^n$ *be a point. Show that the distance from* p *to* H *equals*

$$d(p, H) = \frac{\langle p, a\rangle}{\|a\|}.$$

Exercise 2.15. *Let* $H \subset \mathbb{R}^n$ *be a hyperplane with normal* a*, going through the point* q*. Let* $p \in \mathbb{R}^n$ *be a point. Show that the distance from* p *to* H *equals*

$$d(p, H) = \frac{\langle p-q, a\rangle}{\|a\|}.$$

Exercise 2.16. *Suppose* $w \in \mathbb{R}^n$ *is nonzero. Let* $L \subset \mathbb{R}^n$ *be the line consisting of all multiples* λw *of* w*. Let* $p \in \mathbb{R}^n$ *be any point. Show that the distance* $d = d(p, L)$ *from* p *to* L *satisfies*

$$d^2 = \|p\|^2 - \frac{\langle p, w\rangle^2}{\|w\|^2}.$$

Exercise 2.17. *Suppose* $q, w \in \mathbb{R}^n$ *are vectors with* w *nonzero. Let* $L \subset \mathbb{R}^n$ *be the line consisting of all vectors of the form* $q + \lambda w$ *for some* $\lambda \in \mathbb{R}$*. Let* $p \in \mathbb{R}^n$ *be any point. Show that the distance* $d = d(p, L)$ *from* p *to* L *satisfies*

$$d^2 = \|p-q\|^2 - \frac{\langle p-q, w\rangle^2}{\|w\|^2}.$$

3. Between lines and hyperplanes, intersections

Hyperplanes in $\mathbb{R}^n$ can be given with one equation, lines with $n-1$ equations or one parameter. In this section we will look at the solution sets of any number of equations, and their parametrizations.

3.1. **Equations to parametrizations.** One way to solve the system of equations

$$\begin{cases} x_1 - x_2 &= 1 \\ x_1 + 2x_2 &= 4 \end{cases}$$

is to subtract the first equation from the second to obtain $3x_2 = 3$, solve for x_2 to get $x_2 = 1$, and substitute in the first equation to conclude $x_1 = 2$. Note that as soon as we had the new equation $3x_2 = 3$, we could forget the second of the two original equations without losing information; we could always add the first equation to the equation $3x_2 = 3$ to recover the second equation.

Similarly, we solve the system

$$\begin{cases} 3x_2 + 2x_3 &= 3 \\ x_1 + 2x_2 + x_3 &= 4 \\ 2x_1 + 3x_2 + 3x_3 &= 2 \end{cases}$$

by first eliminating the variable x_1 from all but one equation. We choose to keep the second, as well as the first, where x_1 already does not occur. We replace the last equation by the equation obtained by subtracting the second equation twice from it. The original system is then equivalent to the system

$$\begin{cases} x_1 + 2x_2 + x_3 &= 4 \\ 3x_2 + 2x_3 &= 3 \\ -x_2 + x_3 &= -6. \end{cases}$$

To see that this new system is indeed equivalent with the original, note that indeed all solutions of the first system are solutions of the second, as the last equation holds whenever the original system holds. But the other way around, the last equation of the original system is obtained by addign twice the first to the last equation from the new system; therefore the last equation of the original system holds whenever the new system holds and all solutions to the new system are solutions to the original system. We conclude that they have the same solutions.

The next variable to consider is x_2. In order to avoid denominators, we switch the last two equations of the new system; we also multiply the last one by -1 to obtain coefficient 1 for x_2. We get

$$\begin{cases} x_1 + 2x_2 + x_3 &= 4 \\ x_2 - x_3 &= 6 \\ 3x_2 + 2x_3 &= 3. \end{cases}$$

Clearly this system is equivalent to the previous. We then eliminate x_2 by replacing the last equation by the equation obtained by subtracting three times the second equation (of the latest system) from it. We now get

$$\begin{cases} x_1 + 2x_2 + x_3 &= 4 \\ x_2 - x_3 &= 6 \\ 5x_3 &= -15. \end{cases}$$

This system is again equivalent to the previous ones, because of the same arguments that explained that the first two systems were equivalent. We then solve for the variables from x_3 to x_1. The last equation gives $x_3 = -3$. The second equation then gives $x_2 = 3$ and finally, the first equation gives $x_1 = 1$. The solution therefore is $x = (x_1, x_2, x_3) = (1, 3, -3)$.

Sometimes the solution consists of more than one point. Consider the system of equations

$$\begin{cases} 2x_1 + x_2 + 3x_3 + 2x_4 &= 1 \\ x_1 + \qquad 2x_3 + x_4 &= 1 \\ x_1 - x_2 + 3x_3 + 3x_4 &= 0 \\ \qquad x_2 - x_3 + 2x_4 &= -3. \end{cases}$$

First, again to avoid denominators (which will not always be possible), we switch the first two equations to obtain

$$\begin{cases} x_1 + \qquad 2x_3 + x_4 &= 1 \\ 2x_1 + x_2 + 3x_3 + 2x_4 &= 1 \\ x_1 - x_2 + 3x_3 + 3x_4 &= 0 \\ \qquad x_2 - x_3 + 2x_4 &= -3. \end{cases}$$

Then we replace the second equation by the equation obtained from subtracting twice the first equation from it. We also subtract the first equation from the third, and obtain the equivalent system

$$\begin{cases} x_1 + \qquad 2x_3 + x_4 &= 1 \\ \qquad x_2 - x_3 &= -1 \\ \qquad -x_2 + x_3 + 2x_4 &= -1 \\ \qquad x_2 - x_3 + 2x_4 &= -3. \end{cases}$$

This time, explain for yourself why this system is indeed equivalent to the previous! As the coefficient of x_2 in the second equation of the newest system is already equal to 1, we immediately proceed by adding this second equation to the third and subtracting it from the fourth, so that the variable x_2 is eliminated from the later equations. We get

$$\begin{cases} x_1 + \qquad 2x_3 + x_4 &= 1 \\ \qquad x_2 - x_3 &= -1 \\ \qquad 2x_4 &= -2 \\ \qquad 2x_4 &= -2. \end{cases}$$

We divide the third equation by 2 and subtract it from the fourth to get

$$\begin{cases} x_1 + \qquad 2x_3 + x_4 &= 1 \\ \qquad x_2 - x_3 &= -1 \\ \qquad x_4 &= -1 \\ \qquad 0 &= 0. \end{cases} \tag{4}$$

Now we can again find solutions by solving for the variables from x_4 to x_1. The last equation is useless, but the third gives $x_4 = -1$. Whatever the value of x_3, we can solve for x_2 to make the second equation hold. Therefore, we choose a parameter s for x_3, so $x_3 = s$, and use the second equation $x_2 - x_3 = -1$ to find $x_2 = -1 + x_3 = -1 + s$. Finally, we use the first equation to get $x_1 = 1 - 2x_3 - x_4 = 1 - 2s - (-1) = 2 - 2s$. The solution is therefore the parametrization

$$x = (x_1, x_2, x_3, x_4) = (2 - 2s, -1 + s, s, -1) = (2, -1, 0, -1) + s \cdot (-2, 1, 1, 0).$$

Note that the vector $q = (2, -1, 0, -1)$ is indeed a solution to the system of equations, while $p = (-2, 1, 1, 0)$ is a solution to the associated system

$$\begin{cases} 2x_1 + x_2 + 3x_3 + 2x_4 &= 0 \\ x_1 + 2x_3 + x_4 &= 0 \\ x_1 - x_2 + 3x_3 + 3x_4 &= 0 \\ x_2 - x_3 + 2x_4 &= 0. \end{cases}$$

where we have just replaced all constants in the right-hand side of the equations by zeros.

This process of solving a system of equations by eliminating more and more variables is called *Gaussian elimination*. Although we made some human-friendly choices to avoid denominators, this process can be fully automated and is available in many mathematical software packages.

In order to write this all more efficiently, we leave out the variables, and write the coefficients as a block of numbers, with the constants in the right-hand sides next to it, each equation corresponding to a row. Such a block of numbers is called a *matrix*.

$$\left(\begin{array}{cccc|c} 2 & 1 & 3 & 2 & 1 \\ 1 & 0 & 2 & 1 & 1 \\ 1 & -1 & 3 & 3 & 0 \\ 0 & 1 & -1 & 2 & -3 \end{array}\right) \begin{array}{l} \text{s} \\ \text{s} \\ \\ \\ \end{array}$$

The letters 's' next to the first two rows indicate that in the first reduction step we switch these rows to obtain

$$\left(\begin{array}{cccc|c} 1 & 0 & 2 & 1 & 1 \\ 2 & 1 & 3 & 2 & 1 \\ 1 & -1 & 3 & 3 & 0 \\ 0 & 1 & -1 & 2 & -3 \end{array}\right), \begin{array}{l} \\ -2R_1 \\ -R_1 \\ \\ \end{array}$$

where it is indicated that in the next step we subtract twice the first row from the second and once the first row from the third. We then get

$$\left(\begin{array}{cccc|c} 1 & 0 & 2 & 1 & 1 \\ 0 & 1 & -1 & 0 & -1 \\ 0 & -1 & 1 & 2 & -1 \\ 0 & 1 & -1 & 2 & -3 \end{array}\right), \begin{array}{l} \\ \\ +R_2 \\ -R_2 \end{array}$$

which already indicates that the next matrix looks like

$$\left(\begin{array}{cccc|c} 1 & 0 & 2 & 1 & 1 \\ 0 & 1 & -1 & 0 & -1 \\ 0 & 0 & 0 & 2 & -2 \\ 0 & 0 & 0 & 2 & -2 \end{array}\right). \begin{array}{l} \\ \\ \cdot\frac{1}{2} \\ \\ \end{array}$$

We multiply the third row by $\frac{1}{2}$, as indicated, to find

$$\left(\begin{array}{cccc|c} 1 & 0 & 2 & 1 & 1 \\ 0 & 1 & -1 & 0 & -1 \\ 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 2 & -2 \end{array}\right). \begin{array}{l} \\ \\ \\ -2R_3 \end{array}$$

The last step reduces the matrix to

$$\left(\begin{array}{cccc|c} \textcircled{1} & 0 & 2 & 1 & 1 \\ 0 & \textcircled{1} & -1 & 0 & -1 \\ 0 & 0 & 0 & \textcircled{1} & -1 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right),$$

where in each row we have circled the first nonzero element, which is always a one, if it exists. These ones, or rather their positions, are called *pivots*. By repeatedly applying one of three operations, we have put the matrix in a special form. These operations, called *elementary row operations*, are (i) switch two rows, (ii) add a multiple of one row to another, and (iii) multiply a row by a nonzero scalar.

The form of the final matrix is special in the sense that any row starts with at least as many zeros as any row above it, with equality if and only if the rows contain only zeros. This special form is called the *row echelon form*.

Compare the row echelon form to the equations in (4). The first four columns correspond to the variables x_1, x_2, x_3, x_4. The columns without a pivot allow us to choose a free parameter, in this case we chose $x_3 = s$. The columns with pivots correspond to variables that we can express in terms of the parameters using the equations. For instance, the second column contains a pivot in the second row; the variable x_2 can be expressed in terms of the later variables with the equation $x_2 - x_3 = -1$ corresponding to this second row. We get $x_2 = -1 + x_3 = -1 + s$ as before.

As the number of pivots equals the number of nonzero equations, and the total number of columns corresponding to variables is n, we find the following Theorem.

Theorem 3.1. *Let $V \subset \mathbb{R}^n$ be given by equations and assume V is not empty. Then there is an integer r such that V can be given by r equations and V can also be parametrized with $n - r$ parameters.*

A good part of linear algebra is devoted to generalize and prove a more precise version of Theorem 3.1 (for instance, we have never defined very precisely what a parametrization is). We will see that $n - r$ is the *dimension* of V, while r is the *codimension* of V.

The theorem requires that V is not empty. Indeed, one might encounter inconsistent equations, leading to an empty solution set.

Example 3.2. *Consider the system of equations*

$$\begin{cases} x_1 + x_2 + x_3 &= 3 \\ x_1 + 2x_2 - x_3 &= 4 \\ 3x_1 + 5x_2 - x_3 &= 2. \end{cases}$$

We write down the associated matrix and apply Gaussian elimination.

$$\left(\begin{array}{ccc|c} 1 & 1 & 1 & 3 \\ 1 & 2 & -1 & 4 \\ 3 & 5 & -1 & 2 \end{array}\right) \begin{array}{l} \\ -R_1 \\ -3R_1 \end{array} \dashrightarrow \left(\begin{array}{ccc|c} 1 & 1 & 1 & 3 \\ 0 & 1 & -2 & 1 \\ 0 & 2 & -4 & -7 \end{array}\right) \begin{array}{l} \\ \\ -2R_2 \end{array}$$

$$\dashrightarrow \left(\begin{array}{ccc|c} 1 & 1 & 1 & 3 \\ 0 & 1 & -2 & 1 \\ 0 & 0 & 0 & -9 \end{array}\right).$$

The last row corresponds to the equation $0 = -9$*, so this system of equations has no solution.*

When all equations in a system are given by $\langle a, x\rangle = 0$ for some vector a, we call the system *homogeneous*. In that case we may leave out the right-most column of zeros.

Example 3.3. *Consider the system of equations*

$$\begin{cases} \qquad\qquad\quad x_3 + x_4 + 2x_5 &= 0 \\ x_1 + 2x_2 \qquad\quad - x_4 + x_5 &= 0 \\ 2x_1 + 4x_2 + x_3 \qquad + 3x_5 &= 0. \end{cases}$$

We write down the corresponding matrix and apply Gaussian elimination.

$$\begin{pmatrix} 0 & 0 & 1 & 1 & 2 \\ 1 & 2 & 0 & -1 & 1 \\ 2 & 4 & 1 & 0 & 3 \end{pmatrix} \begin{matrix} s \\ s \\ \end{matrix} \dashrightarrow \begin{pmatrix} 1 & 2 & 0 & -1 & 1 \\ 0 & 0 & 1 & 1 & 2 \\ 2 & 4 & 1 & 0 & 3 \end{pmatrix} \begin{matrix} \\ \\ -2R_1 \end{matrix} \dashrightarrow$$

$$\begin{pmatrix} 1 & 2 & 0 & -1 & 1 \\ 0 & 0 & 1 & 1 & 2 \\ 0 & 0 & 1 & 2 & 1 \end{pmatrix} \begin{matrix} \\ \\ -R_2 \end{matrix} \dashrightarrow \begin{pmatrix} 1 & 2 & 0 & -1 & 1 \\ 0 & 0 & 1 & 1 & 2 \\ 0 & 0 & 0 & 1 & -1 \end{pmatrix}$$

The pivots correspond to variables x_1, x_3, x_4*, so we can choose* x_2 *and* x_5 *freely, say* $x_2 = t$ *and* $x_5 = s$*. Then the last row of the matrix in row echelon form gives* $x_4 = x_5 = s$*, while the second row gives* $x_3 = -x_4 - 2x_5 = -3s$*. Finally, the first row gives* $x_1 = -2x_2 + x_4 - x_5 = -2t$*. The parametrization of the solution set of this system of equation is therefore given by*

$$x = (x_1, x_2, x_3, x_4, x_5) = (-2t, t, -3s, s, s) = s \cdot (0, 0, -3, 1, 1) + t \cdot (-2, 1, 0, 0, 0).$$

3.2. **Parametrizations to equations.** Going back from a parametrization to equations is in some sense the same as going from equations to parametrizations. A system of equations can be written as

$$\begin{cases} \langle a_1, x\rangle &= b_1 \\ \langle a_2, x\rangle &= b_2 \\ &\vdots \\ \langle a_r, x\rangle &= b_r \end{cases}$$

for some vectors $a_1, \ldots, a_r \in \mathbb{R}^n$ and constants $b_1, \ldots, b_r \in \mathbb{R}$. A parametrization of the set of solutions to this system is given by

$$x = p + \lambda_1 v_1 + \lambda_2 v_2 + \ldots + \lambda_t v_t, \tag{5}$$

with $p \in \mathbb{R}^n$ a solution to the equations and $v_1, \ldots, v_t$ independent solutions to the homogeneous system obtained by replacing all the constants b_i by 0. A precise definition of the word independent will be given soon. In other words, the vectors $v_1, v_2, \ldots, v_t$ are orthogonal to the vectors $a_1, a_2, \ldots a_r$. Note that we have $r+t \geq n$, as the equations may be redundant.

This means that the other way around, if we are given a parametrization as in (5), we can find corresponding equations by first finding enough vectors $a_1, a_2, \ldots, a_r$

that are orthogonal to $v_1, \ldots, v_t$. This means they are solutions to the system

$$\begin{cases} \langle v_1, x\rangle &= 0 \\ \langle v_2, x\rangle &= 0 \\ &\vdots \\ \langle v_t, x\rangle &= 0. \end{cases}$$

Note that by Theorem 3.1, if the t vectors $v_1, \ldots v_t$ are independent, i.e., if these equations are not redundant, then we will find $r = n - t$ independent vectors $a_1, \ldots, a_r$. Now the constants $b_1, \ldots, b_r$ that we are looking for are easy to find as for each i we have $b_i = \langle a_i, p\rangle$. This means that going from parametrizations to equations is in fact easier than the other way around.

Example 3.4. *Consider the line $L \subset \mathbb{R}^3$ parametrized by $p + \lambda w$ with $p = (1,1,1)$ and $w = (1,-1,2)$. Then for any equation $\langle a, x\rangle = b$ that the points on L satisfy, the vector a is orthogonal to w. Therefore, in order to find such vectors a, we will actually solve the equation $\langle w, x\rangle = 0$ in x first. The matrix*

$$\begin{pmatrix} 1 & -1 & 2 \end{pmatrix}$$

with exactly one row, namely the coefficients of w, is of course already in row echelon form, with the first entry being the unique pivot. We get a parametrization of its solutions by choosing freely $x_2 = s$ and $x_3 = t$ and computing $x_1 = s - 2t$, so we obtain

$$a = x = s \cdot (1,1,0) + t \cdot (-2,0,1).$$

This means we find two independent vectors $a_1 = (1,1,0)$ and $a_2 = (-2,0,1)$. Note that these could have been found more directly, by not choosing $x_2 = s$ and $x_3 = t$, but by choosing $(x_2, x_3) = (1,0)$ and $(x_2, x_3) = (0,1)$, and computing for each choice the corresponding value for x_1. With $b_1 = \langle a_1, p\rangle = 2$ and $b_2 = \langle a_2, p\rangle = -1$ we find the equations

$$\begin{cases} \langle a_1, x\rangle &= 2 \\ \langle a_2, x\rangle &= -1 \end{cases}$$

for L, or equivalently,

$$\begin{cases} x_1 + x_2 & &= 2 \\ -2x_1 & + x_3 &= -1. \end{cases}$$

Example 3.5. *Let $V \subset \mathbb{R}^3$ be the plane through the points $p_1 = (1,2,-1)$, $p_2 = (0,1,-3)$ and $p_3 = (-1,0,1)$. Determining a parametrization for V does not require much computation: the parametrization*

$$x = p_1 + s \cdot (p_2 - p_1) + t \cdot (p_3 - p_1)$$

gives a plane that contains p_1 (for $(s,t) = (0,0)$) as well as p_2 and p_3 (for $(s,t) = (1,0)$ and $(s,t) = (0,1)$), so it has to be the plane V. Putting $p_2 - p_1$ and $p_3 - p_1$ in a matrix, we apply Gaussian elimination

$$\begin{pmatrix} -1 & -1 & -2 \\ -2 & -2 & 2 \end{pmatrix} \cdot -1 \dashrightarrow \begin{pmatrix} 1 & 1 & 2 \\ -2 & -2 & 2 \end{pmatrix} +2R_1 \dashrightarrow \begin{pmatrix} 1 & 1 & 2 \\ 0 & 0 & 6 \end{pmatrix}.$$

This means that for a normal $n = (n_1, n_2, n_3)$ of V we have $6n_3 = 0$ and $n_1 + n_2 = 0$, so we may take $n = (1,-1,0)$. This gives an equation $\langle n, x\rangle = b$ for V, with $b = \langle n, p_1\rangle = -1$. We have therefore found the equation

$$x_1 - x_2 = -1.$$

Indeed, all three points satisfy this equation.

3.3. **Intersections.** Suppose $V \subset \mathbb{R}^n$ is given by the equations

$$\begin{cases} \langle a_1, x\rangle &= b_1 \\ \langle a_2, x\rangle &= b_2 \\ &\vdots \\ \langle a_r, x\rangle &= b_r \end{cases}$$

and $W \subset \mathbb{R}^n$ by the equations

$$\begin{cases} \langle c_1, x\rangle &= d_1 \\ \langle c_2, x\rangle &= d_2 \\ &\vdots \\ \langle c_t, x\rangle &= d_t. \end{cases}$$

Then the intersection $V \cap W$ is given by the union of all these $r + t$ equations. In this sense taking intersections is easy. If V and W are given by parametrizations, or if you want a parametrization for the intersection, then you can use the previous sections to go from equations to parametrizations or the other way around.

There is a situation, though, where you can reduce the computations involved. Suppose $V \subset \mathbb{R}^n$ is still given by equations as above, while $W \subset \mathbb{R}^n$ is parametrized by

$$x = p + \lambda_1 v_1 + \lambda_2 v_2 + \ldots + \lambda_t v_t$$

(different t). If we want a parametrization for the intersection $V \cap W$, then we do not need to first find equations for W, put them together with the equations for V, and then finally find a parametrization for the intersection. Instead, we can substitute the parametrization for W into the equations for V. This gives equations for the parameters. We can parametrize these new equations, thereby parametrizing the parameters. Substitution then gives a parametrization for $V \cap W$. This will become much more clear with an example.

Example 3.6. *Suppose the plane $V \subset \mathbb{R}^3$ is given by $\langle a, x\rangle = 3$ with $a = (4, -2, 3)$, while the plane $W \subset \mathbb{R}^3$ is parametrized by*

$$x = p + \lambda v + \mu w$$

with $p = (1, 1, 1)$, $v = (1, 3, -2)$, and $w = (2, -1, 0)$. The intersection consists of those points $x = p + \lambda v + \mu w$ that satisfy the equation for V, which results in the equation

$$3 = \langle a, x\rangle = \langle a, p + \lambda v + \mu w\rangle = \langle a, p\rangle + \lambda\langle a, v\rangle + \mu\langle a, w\rangle = 5 - 8\lambda + 10\mu$$

for λ and μ. The equation $8\lambda - 10\mu = 2$ can be parametrized by $(\lambda, \mu) = (-1 + 5s, -1+4s)$. Substituting this in the parametrization for W, we get the parametrization

$$x = p + (-1 + 5s)v + (-1 + 4s)w = (p - v - w) + s \cdot (5v + 4w) = p' + sq'$$

with

$$p' = p - v - w = (-2, -1, 3) \qquad \text{and} \qquad q' = 5v + 4w = (13, 11, -10)$$

for the intersection $V \cap W$. Note that in the notation we refer as much as possible to vectors, rather than all their entries separately. This keeps everything more transparent and easier to read.

3.4. **Exercises.**

Exercise 3.1. *Use Gaussian elimination to bring the following matrices into row echelon form.*

$$\begin{pmatrix} 2 & 1 & 1 \\ 3 & 3 & 2 \\ 3 & 2 & 3 \end{pmatrix} \begin{pmatrix} 2 & 4 & -3 & 3 & 4 \\ 1 & 2 & 0 & 3 & 5 \\ 3 & 6 & -3 & 7 & 8 \\ 1 & 2 & -2 & 1 & 1 \end{pmatrix}$$

$$\begin{pmatrix} -2 & 1 & 0 & 1 & 2 \\ 1 & 0 & -1 & 2 & 3 \\ -3 & 1 & 2 & 1 & -1 \end{pmatrix} \begin{pmatrix} -2 & 1 & 0 & 1 & 2 \\ 3 & -1 & -1 & 1 & 1 \\ 8 & -3 & -2 & 1 & 0 \end{pmatrix}$$

Exercise 3.2. *Give equations and parametrization of*

(1) *the plane* $V \subset \mathbb{R}^3$ *through the points* $p_1 = (1, -2, 1)$, $p_2 = (3, 1 - 2)$, *and* $p_3 = (2, -1, 3)$,
(2) *the plane* $W \subset \mathbb{R}^3$ *with normal* $n = (1, 2 - 1)$ *through the point* p_1,
(3) *the intersection* $V \cap W$.

Exercise 3.3. *Let* $V \subset \mathbb{R}^5$ *be parametrized by* $x = \lambda v_1 + \mu v_2 + \nu v_3$ *with*

$$v_1 = (-1, 1, -1, 3, -1), \quad v_2 = (-1, 1, 2, 0, 6), \quad \text{and} \quad v_3 = (1, 0, 2, -1, 4).$$

Let $W \subset \mathbb{R}^5$ *be parametrized by* $x = \lambda w_1 + \mu w_2 + \nu w_3$ *with*

$$w_1 = (1, 0, -1, 2, -3), \quad w_2 = (-2, 1, 0, 1, 2), \quad \text{and} \quad w_3 = (-3, 1, 2, 1, -1).$$

Give equations and a parametrization for the intersection $V \cap W$.

Exercise 3.4. *Consider the point* $p = (1, 1, 1) \in \mathbb{R}^3$ *and the vectors*

$$n_1 = (1, 2, 3), \quad n_2 = (0, 1, 2), \quad \text{and} \quad n_3 = (0, 0, 1).$$

Let $V_i \subset \mathbb{R}^3$ *be the plane through* p *with normal* n_i. *Give a parametrization for the intersection* $V_1 \cap V_2 \cap V_3$.

Exercise 3.5. *Give equations and parametrization of*

(1) *the plane* $V \subset \mathbb{R}^4$ *through the points* $q_1 = (1, 2, -1, 0)$, $q_2 = (1, -3, 1 - 2)$, *and* $q_3 = (0, 2, 1, 3)$,
(2) *the hyperplane* $W \subset \mathbb{R}^4$ *with normal* $n = (1, 2, -1, 1)$ *through the point* q_2,
(3) *the hyperplane* $X \subset \mathbb{R}^4$ *through the point* q_3 *that is parallel to* W,
(4) *the hyperplane* $Y \subset \mathbb{R}^4$ *with normal* $n = (-2, -1, 3, 1)$ *through the point* q_1,
(5) *the intersection* $V \cap W$,
(6) *the intersection* $V \cap X$,
(7) *the intersection* $X \cap W$,
(8) *the intersection* $W \cap Y$.

References

[S1] M. Stoll: *Linear Algebra I.* Lecture notes from Jacobs University, Bremen.
[S2] M. Stoll: *Linear Algebra II.* Lecture notes from Jacobs University, Bremen.

Index