UNIVERSITY
OF ARCHITECTURE
CIVIL ENGINEERING
AND GEODESY

INTRODUCTION TO LINEAR ALGEBRA AND ANALYTIC GEOMETRY

by

Danail S. Brezov

e-mail: danail.brezov@gmail.com

Second Edition

ISBN 978-954-724-112-1

University of Architecture, Civil Engineering and Geodesy

May 15, 2018

Contents

Preface

The present introductory course is based on a set of lectures I had the pleasure to give at the University of Architecture, Civil Engineering and Geodesy (Sofia, Bulgaria) but may be used also by students in physics, chemistry and various engineering disciplines. It may be used as a supplementary tool for students in pure mathematics as well, but one should note that the exposition is rather intuitive in some parts, relying much more on extensive examples and geometric analogies, rather than mathematical rigor. However, the main objects are carefully introduced and the most important results - given with constructive proofs. In some parts of the content, however, the reader is only being guided through in a semi-intuitive manner, the focus being much more on the applications, rather than the consistency of the theory.

The text is organized as follows: in the beginning we cover some basic college level mathematics - combinatorics, the binomial formula and the induction method, as well as complex numbers and polynomials. After that, we jump to planimetry resorting on both the high school geometric intuition and the standard constructions in $\mathbb{C}$, such as vector summation, dot and wedge products, and polar coordinates. Before we extend to the three-dimensional case, we introduce the usual matrix algebra and apply it to systems of linear equations (and matrix equations in general). Then, we consider the vector algebra in $\mathbb{R}^3$ and some of the basic problems of stereometry. Next, we go a little bit more abstract discussing properties of linear operators and in particular, projections, reflections and Euclidean motions with their explicit matrix constructions. Finally, we consider the eigenvalue problem and its application on finding the normal forms of quadratic curves and surfaces, as well as some basic properties and classification results for the classical quardics and conics. Once we have crossed the boundaries of linearity, we give in the temptation of discussing tangents of normals of curves and surfaces and their application in the problem of curved mirrors in geometric optics. This concludes the standard course. My initial idea was to include

an additional chapter (appendix) dedicated to geometric algebras and their applications in Euclidean geometry, but I managed to resist this temptation only mentioning sporadically quaternions and dual numbers instead in the comments and examples, where a little bit more advanced material somehow "slipped through the cracks" here and there - hopefully in moderation.

Sofia,
18.05.2017

Chapter 1

Combinatorics and the Induction Method

1.1 Combinations and the Binomial Formula

We start this algebra course with a brief revision of basic combinatorics. The standard procedure we follow here is to consider a set of elements $\{b_k\}$ and a set of spots $\{c_k\}$ (usually called cells) to place these elements. The simplest possible situation is to have equal number of elements and cells, say n. The combinatorial problem is then to determine the number of different ways to place the elements into the cells, provided each cell c_i contains one and only one element b_j, or equivalently, we are looking for the number of possible one-to-one correspondences between the two finite sets. This problem may be treated as follows: start filling the cells one by one, counting the number of possible choices each time. For c_1 we obviously have n possibilities, since each of the b_k's can be placed there. After c_1 is filled, we are left with $n-1$ choices for c_2, then for c_3 we have $n-2$ and so on, until we move to the last cell c_n, for which we have only one element left. Note that for each choice of element for the cell c_1 we have all the $n-1$ possibilities for c_2 and for each of them - $n-2$ to put in c_3 and so on. Thus, for the total number of possibilities we need to multiply

$$P_n = n(n-1)(n-2)\ldots 3.2.1 \tag{1.1}$$

and these different orderings are also referred to as *permutations* of n elements. The above operation is used quite often in different contexts and has

one more definition - the *factorial* sequence (or function). We write

$$n! = \prod_{k=1}^{n} k \tag{1.2}$$

and read "*n factorial*" for the product of all natural numbers from 1 to n. There is another way to define the above sequence, which is based on the idea of *recursion*. Roughly speaking, a recursive, or recurrent formula is one, which defines the n^{th} element of a sequence a_n via a_{n-1} and possibly some more terms with lower value of the index

$$a_n = f(a_{n-1}, a_{n-2}, \dots a_{n-k}) \tag{1.3}$$

is a recursive formula with step k. In order to be able to determine that way a sequence uniquely, however, one needs also the values of the first k terms. This is called *bottom* of the recursion. In our example we may construct the factorial sequence recursively as the unique function of a non-negative integer variable that obeys the conditions

$$f(n) = nf(n-1), \qquad f(0) = 1. \tag{1.4}$$

Note that without the latter each sequence of the type $a_n = an!$, $a \in \mathbb{R}$ would be a solution. Instead, we define $0! = 1$ and the recurrent formula $n! = n(n-1)!$ gives all successive terms in the sequence:

$$1! = 1.0! = 1, \quad 2! = 2.1! = 2, \quad 3! = 3.2! = 6, \quad 4! = 4.3! = 24, \dots$$

In this way $n!$ is defined as a map $f : \mathbb{N} \to \mathbb{N}$ (the zero being included). We do not have $(\sqrt{2})!$ or $\pi!$ for example - these expressions simply make no sense. Although there exists a generalization of $n!$ for non-integer arguments (the so-called *Euler gamma function*), it demands some more advanced calculus and we are not going to need it in this course.

Now, let us consider a slightly different situation, in which the number of elements n exceeds the number of cells k, so there is scarcity of places for the elements to be positioned (for example a parking lot with k spots and n cars waiting to be parked with $n > k$). We may treat this situation in the exact same way as before: for the first cell there are n possible candidates, in each case leaving $n-1$ for the second one and so on, until we reach the last one, for which this time the number of possibilities are $n-k+1$ since we count backwards from n and have passed $k-1$ steps. Then we end up with

$$V_n^k = n(n-1)(n-2)\dots(n-k+1) = \frac{n!}{(n-k)!} \tag{1.5}$$

possible arrangements of n elements into k cells ($n > k$), called *variations* of n elements of order k. Variations can be thought of as permutations of various *samples*, i.e., different choices of the k elements extracted from all n possibilities. Sometimes, however, we may ignore the order and focus only on that particular choice, e.g. we are usually interested only in the particular numbers that have won the lottery, without asking what order they were chosen in. Then we identify all the $k!$ permutations of a given sample and obtain

$$C_n^k = \frac{V_n^k}{k!} = \frac{n!}{k!(n-k)!} = \binom{n}{k} \tag{1.6}$$

called *combinations* of order k (of n elements). These numbers play important role not only in combinatorics - they appear in various branches of mathematics. For example, they give the coefficients in the *Newton binomial formula*

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k} \tag{1.7}$$

and for that reason combinations are also called *binomial coefficients*. Certainly, the origin of this formula is purely combinatorial - we may think of $(a+b)^n$ as a product of n pairs of brackets $(a+b)(a+b)\dots(a+b)$ and the coefficient in front of $a^k b^{n-k}$ equals the number of possible ways to obtain the term a^k (b always has the complementary power) by taking a from k of all the n pairs of brackets, so we may say an element (bracket) is in the sample if we pick a from it, and is left out if we pick b. The binomial coefficients C_n^k can be also obtained recursively from a construction, known as the *Pascal triangle*

$$\begin{array}{ccccccccc}
 & & & & 1 & & & & \\
 & & & 1 & & 1 & & & \\
 & & 1 & & 2 & & 1 & & \\
 & 1 & & 3 & & 3 & & 1 & \\
1 & & 4 & & 6 & & 4 & & 1
\end{array}$$

in which the n^{th} row gives the coefficients of $(a+b)^n$, starting with

$$(a+b)^0 = 1, \quad (a+b)^1 = a+b, \quad (a+b)^2 = a^2 + 2ab + b^2, \dots$$

etc., so the k^{th} column gives the factor in front of $a^k b^{n-k}$. There are two obvious properties of $C_n^k = \binom{n}{k}$ form the *Pascal* triangle:

1. Firstly, it is symmetric, i.e.,

$$\binom{n}{k} = \binom{n}{n-k}$$

2. Secondly, each coefficient equals the sum of the two numbers placed right above it[1], i.e.,

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$$

These can both be proved by the definition - for the first one we have

$$\frac{n!}{k!(n-k)!} = \frac{n!}{(n-k)!\,(n-(n-k))!}$$

and in particular

$$\binom{n}{0} = \binom{n}{n} = 1.$$

Verify the second identity and define with its help C_n^k in a recursive way! Setting $a = b = 1$ and $b = -a = 1$ in (1.7), we may prove two more identities for the binomial coefficients, namely

$$\sum_{k=0}^{n} \binom{n}{k} = 2^n \tag{1.8}$$

and respectively

$$\sum_{k=0}^{n} (-1)^k \binom{n}{k} = 0. \tag{1.9}$$

There is one more kind of variations, called *variations with repetitions*, in which we have n variants for each of the k cells, but identical elements can appear in different cells - this gives a total number of

$$\tilde{V}_n^k = k^n \tag{1.10}$$

variations with repetitions. With this new notation (1.8) can be written as

$$\sum_{k=0}^{n} C_n^k = \tilde{V}_n^2. \tag{1.11}$$

[1] for the end terms one of these numbers is missing, so we set it to be equal to zero.

We also note that (1.7) may be used as a discrete distribution (also called *binomial*) for a probabilistic process with two possible outcomes (e.g. tossing a coin) with $a = p \in (0, 1)$ denoting the probability for one of the outcomes (e.g. "heads") in each individual repetition, and $b = 1 - p$ the probability for the other one ("tails"). Then, after n experiments (tosses) we have a total of 2^n possibilities, among which C_n^k are expected to have a number of k heads and a number of $n - k$ tails, so the probability for such an event is estimated to be $p(k) = 2^{-n}C_n^k$. In the coin example we have $p = 1/2$ and thus after 4 tosses the probability for 2 heads will be $2^{-4}C_4^2 = 3/8$.

1.2 The Method of Full Induction

The method of full mathematical induction is a powerful means of proving relations (such as equalities and inequalities) involving a natural parameter n. The idea is to use recursion, starting with the lowest possible value of this parameter (sometimes also called *index*) and show that if the relation holds for n, then it should hold also for the successive value of the index, $n + 1$. Thus, the procedure consists of two steps: to show the relation $f(n) \sim g(n)$, it is sufficient to prove

1. $f(n_0) \sim g(n_0)$, where $n_0 = \min n$ (base of the induction)
2. $f(n) \sim g(n) \Rightarrow f(n + 1) \sim g(n + 1)$ (recursive step).

In other words, after ensuring the validity of the relation for the smallest possible value n_0, from the recursive step we conclude that the latter holds also for $n_0 + 1$, but then applying it once more - we obtain $n_0 + 2$ and so on, for all succeeding values of n. Let us consider a simple example: the sum all integers from 1 to n, for which the property of arithmetic sequences yields

$$S_n = \sum_{k=1}^{n} k = 1 + 2 + \ldots + n = \frac{n(n+1)}{2}.$$

To prove this result using mathematical induction we first point out that the above is obviously true for $n = 1$, so the initial part of the procedure is complete. The second step is to prove that it the equality holds for n, this implies an analogous relation for $n + 1$. In order to see this, we use the induction assumption and write

$$S_{n+1} = \sum_{k=1}^{n+1} k = \frac{n(n+1)}{2} + (n + 1) = \frac{(n+1)(n+2)}{2}$$

which is exactly the above relation with the substitution $n \to n+1$, so this completes the proof. Consider one more example, this time with an inequality:

$$\left(\frac{n}{3}\right)^n < n! < \left(\frac{n}{2}\right)^n, \qquad n \geq 6$$

To begin with, for $n = 6$ the inequalities $64 < 720 < 729$ hold, so the base of the induction is ensured. Now, for the second part of the prove one uses transitivity to show

$$\left(\frac{n+1}{2}\right)^{n+1} = \left(\frac{n}{2}\right)^n (n+1) > n!(n+1) = (n+1)!$$

where we have implemented Newton's binomial formula to see that

$$(n+1)^n = n^n + nn^{n-1} + \ldots + 1 > 2n^n.$$

We may prove both inequalities at once with the aid of the well-known result

$$2 < \left(1 + \frac{1}{n}\right)^n < 3$$

and in particular

$$\lim_{n\to\infty} \left(1 + \frac{1}{n}\right)^n = e \approx 2.718 \tag{1.12}$$

yields the estimate (for large values of n)

$$\sqrt[n]{n!} \underset{n\gg 1}{\longrightarrow} \frac{n}{e} \tag{1.13}$$

that is usually derived as a consequence of the famous Stirling formula

$$n! \underset{n\gg 1}{\longrightarrow} \sqrt{2\pi n}\left(\frac{n}{e}\right)^n.$$

Finally, let us also prove the Bernouli inequality

$$(1+x)^n \geq 1 + nx, \qquad x > -1. \tag{1.14}$$

Obviously, for $n = 0$ we have an equality, so let us proceed by writing

$$(1+x)^{n+1} = (1+x)^n(1+x) \geq (1+nx)(1+x) = 1+(n+1)x+nx^2 \geq 1+(n+1)x$$

where we use the induction assumption for the first inequality and the non-negativity of the term nx^2 for the second one, which completes the proof.

Problem Session

1. If there are twelve noble knights and a round table with twelve chairs, in how many different ways can these knights be seated on the chairs (one in each chair), considering configurations up to a rotation?

2. What is the probability to win the lottery 5 of 35 and 6 of 49? What is the probability for the latter to have won exactly three 4's after six tries?

3. How many tickets with holes minimum you need to carry with you to make sure that you have the one that fits a randomly chosen bus (tickets have between one and nine holes, in different configurations)?

4. What is the probability for a dice to show exactly three times 6 after nine throws?

5. Prove the equalities

   a) $\displaystyle\sum_{k=1}^{n} k^3 = \frac{n^2(n+1)^2}{4}$

   b) $\displaystyle\sum_{k=1}^{n} \frac{1}{k(k+1)} = \frac{n}{n+1}$

6. Prove the inequalities

   a) $\displaystyle\sum_{k=1}^{n} \frac{1}{k^2} < 2 - \frac{1}{n}$

   b) $\displaystyle\sum_{k=1}^{n} \frac{1}{k^3} < 2 - \frac{1}{n^2}$

7. Prove that

   a) $n(n+1)(n+2)(n+3)$ is divisible by 24

   b) $n(n+1)(n+2)(n+3)\dots(n+r–1)$ is divisible by $r!$

Chapter 2

Complex Numbers

The demands of nowadays rapidly evolving technologies naturally put higher and higher standards for even the most basic tools and materials in use. The situation in mathematics is quite similar, the "tools and materials" certainly being in this case associated with concepts and interconnections. Moreover, we may add the natural desire for generalization that pushed mathematicians and philosophers through the centuries to to explore farther and deeper into the realms of abstract thinking. This may be observed even in one of the most fundamental notions in mathematics: numbers, starting with the most intuitive ones - the natural numbers $\mathbb{N}$ we use for counting, to integers $\mathbb{Z}$ (so that we may introduce subtraction), then rational numbers (in order to divide). Since the latter do not include quantities as $\sqrt{2}$ and π that appear naturally in even the simplest geometric problems, we need one more extension: the field of real numbers $\mathbb{R}$ defined as the topological closure of $\mathbb{Q}$, i.e., it contains the limits of all rational numerical sequences, including algebraic numbers as $\sqrt{2}$ and transcendent ones such as π. We are so happy with this object that we often accept it as the ultimate set of numbers. Nevertheless, restricting ourselves to $\mathbb{R}$ we face another inconvenience: negative discriminants. Consider for example the simple quadratic equation $x^2 + 1 = 0$. There are obviously no real solutions, thus we define a new set - the so called *imaginary numbers* that have negative squares. In particular, we refer to $i = \sqrt{-1}$ as the *imaginary unit* and all other representatives of this set have the form $i\lambda$ with $\lambda \in \mathbb{R}$. This new construction naturally yields $x_{1,2} = \pm i$. More generally, the quadratic equation

$$ax^2 + bx + c = 0, \quad a, b, c \in \mathbb{R} \tag{2.1}$$

always has two roots: in the case $\Delta = b^2 - 4ac > 0$ these are

$$x_{1,2} = \frac{-b \pm \sqrt{\Delta}}{2a}. \tag{2.2}$$

In particular we may have $x_1 = x_2 = \dfrac{-b}{2a}$ in the case $\Delta = 0$ and

$$x_{1,2} = \frac{-b \pm i\sqrt{|\Delta|}}{2a}, \qquad \Delta < 0. \tag{2.3}$$

Note that the imaginary number contribution always comes with different signs for the two solutions. This observation is a base point in some deep theorems of Algebra. Since $i^2 = -1 \in \mathbb{R}$ the set of imaginary numbers cannot be closed under algebraic operations, so it makes more sense to consider numbers with real and imaginary parts instead

$$\mathbb{C} \cong \{\, z = x + iy \mid x, y \in \mathbb{R},\ i^2 = -1\} \tag{2.4}$$

and call such objects *complex numbers*. This completes the number sequence

$$\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R} \subset \mathbb{C}$$

and going farther along this chain one needs to violate some of the most fundamental properties regular numbers have, such as commutativity (with quaternions $\mathbb{H}$), associativity (octonions $\mathbb{O}$) etc. Furthermore, it has been shown that there are no other norm division algebras[1] apart from $\mathbb{R}$, $\mathbb{C}$, $\mathbb{H}$ and $\mathbb{O}$. Why $\mathbb{Q}$ fails to satisfy this property is pretty obvious: $\sqrt{x^2 + y^2}$ is not rational for any pair of numbers $x, y \in \mathbb{Q}$, we have a non-trivial Pythagorean type relation to confine with. Apart from that, $\mathbb{Q}$ is a *number field*, i.e., all axioms are satisfied and both algebraic operations are invertible. The latter cannot be said about $\mathbb{Z}$ - only summation is invertible there, and objects like this are called *rings* (another famous example is the ring of polynomials). Finally, the integer numbers $\mathbb{Z}$ have the structure of a *module* (over itself), i.e., we are allowed to add, subtract and multiply, but not divide. Let us give an example of an algebra, similar to $\mathbb{C}$, which is not a division algebra, consisting of numbers of the type $x + \epsilon\, y$ with $x, y \in \mathbb{R}$ and $\epsilon^2 = 1$ (but $\epsilon \neq \pm 1$) playing the role of the pseudo-imaginary unit. This object usually referred to as *split complex* (or *hyperbolic*) numbers and denoted $\mathbb{C}'$ plays an essential role in hyperbolic geometry. However, it has the problems of zero divisors, i.e., non-vanishing elements whose product is zero, e.g. defining $a_\pm = 1 \pm \epsilon$ clearly yields $a_+ a_- = 0$. Another famous example is given by the so-called *dual numbers* (used in modern mechanics) $x + \epsilon\, y$ with $\epsilon^2 = 0$, in which the pseudo-imaginary unit is itself a zero divisor.

[1] i.e., each non-zero element has a non-vanishing norm and thus, multiplicative inverse.

2.1 The Algebra of Complex Numbers

From what has been said so far about complex numbers $\mathbb{C}$ one easily concludes that as a set it is equivalent to the real plane $\mathbb{C} \cong \mathbb{R}^2$, i.e., $z = x + iy \in \mathbb{C} \quad \leftrightarrow \quad (x, y) \in \mathbb{R}^2$ where the corresponding *Cartesian* coordinates $(x, y) \in \mathbb{R}^2$ are referred to as the *real* and the *imaginary* part of z, respectively. We also write $x = \Re(z)$, $y = \Im(z)$ and call the latter *algebraic coordinates* of z or simply say that $z = x + iy$ is given in *algebraic form*. Now, it remains to construct the algebra of complex numbers in this representation postulating

1. The sum of two complex numbers $z_{1,2} = x_{1,2} + iy_{1,2}$ is given as

$$z = z_1 + z_2 = x_1 + x_2 + i(y_1 + y_2) \tag{2.5}$$

2. For any $\lambda \in \mathbb{R}$ and $z = x + iy \in \mathbb{C}$ we have $\lambda z = \lambda x + i\lambda y \in \mathbb{C}$

3. The product of two complex numbers $z_{1,2} = x_{1,2} + iy_{1,2}$ is given as

$$z_1 z_2 = x_1 x_2 - y_1 y_2 + i(x_1 y_2 + y_1 x_2) \tag{2.6}$$

$\otimes$	1	i
1	1	i
i	i	-1

where for the latter we only use the above multiplication table. Thus we see that $\mathbb{C}$ is *closed* under summation and multiplication, i.e., it contains all sums and products of its elements, and the following properties are satisfied (verify them!)

$$\begin{aligned} z_1 + z_2 &= z_2 + z_1, & z_1 z_2 &= z_2 z_1 \\ z_1 + (z_2 + z_3) &= (z_1 + z_2) + z_3, & z_1(z_2 z_3) &= (z_1 z_2) z_3 \\ (z_1 + z_2) z_3 &= z_1 z_3 + z_2 z_3, & z_1(z_2 + z_3) &= z_1 z_2 + z_1 z_3 \end{aligned}$$

so it has the structure of a *commutative*, *associative* algebra. A generic *algebra*, on the other hand, only needs the third one (distributivity) to exist (apart from all properties on the left). Moreover, all non-zero elements in $\mathbb{C}$ are invertible with respect to summation and multiplication[2] (as we shall see

[2] the additive and multiplicative neutral elements are the same as in $\mathbb{R}$, 0 and 1.

later), so it is also a number field. This is in fact the maximal extension of $\mathbb{R}$ that preserves the above three pairs of algebraic properties, as well as both additive and multiplicative invertibility of non-zero elements. Thus, we say that $\mathbb{C}$ is the maximal number field, but note that it is not an ordered field: we can compare separately the real and imaginary parts of two complex numbers, but there is no relation $z_1 > z_2$ or $z_1 < z_2$. This lack of ordering makes it irrelevant to consider $\pm\infty$ as in $\mathbb{R}$: whatever direction we choose in the complex plane, we are heading to the very same point ∞ and therefore, adding this point to $\mathbb{C}$, we get the topology of a two-dimensional sphere

$$\bar{\mathbb{C}} \cong \mathbb{C} \cup \{\infty\} \cong \mathbb{S}^2$$

referred to as *extended complex plane* (or the *Riemann sphere*).

Exercises:

1. Find all roots of $z^3 - 1 = 0$
2. Find the real and imaginary part of $z = (1 - i)(2i - 3)$
3. Write a condition for z_1 and z_2 so that their sum/product has zero real/imaginary part
4. Write explicit expressions for $f(z) = z^2$ and $f(z) = z^3$ in terms of real and imaginary parts

2.2 The Complex Conjugate and the Modulus

Looking at (2.3) we note that if $x + iy$ is a root of a quadratic equation with real coefficients, so is $x - iy$. Such symmetry is observed for all roots of polynomials, as we shall see in the next chapter, and it makes sense to define it as an operation over the complex numbers. We refer to $\bar{z} = x - iy$ as the *complex conjugate* of $z = x + iy$ (sometimes also denoted by z^*). Geometrically it is clearly a reflection about the x-axis (see Figure 2.1). Can you write similarly the corresponding expressions for a reflection about the y-axis and a point symmetry with respect to the origin? Complex conjugation is an *involution*, meaning that its square equals the identity transformation $\bar{\bar{z}} = z$. We may use this new operation to generate a bunch real numbers associated with z. For example, since $z + \bar{z} = 2x$ and $z - \bar{z} = 2iy$, we have

$$\Re(z) = \frac{z + \bar{z}}{2}, \qquad \Im(z) = \frac{z - \bar{z}}{2i}. \tag{2.7}$$

This construction, often used in mathematics, basically decomposes the ambient space into a direct sum of two subspaces - one that is preserved under given involution and another, which is reflected by it. Similarly, we may separate the even and odd part of a function, symmetric and skew-symmetric part of a matrix, parallel and orthogonal projection of a vector etc. On the other hand, it is easy to see that

$$z\bar{z} = x^2 + y^2 \tag{2.8}$$

which has a clear geometric meaning in view of the *Pythagoras theorem*: since the real and imaginary parts are nothing but the Cartesian coordinates of the point z, the above expression yields the square of the Euclidean distance between z and the origin 0. Thus, it makes sense to denote

$$|z| = \sqrt{z\bar{z}} = \sqrt{x^2 + y^2}. \tag{2.9}$$

which we refer to as the *modulus* of z and it plays central role in our further considerations. First of all, note that the modulus is a natural extension to the notion of *absolute value* for real numbers. More precisely, $|z| = R > 0$ is a set of points at equal distance from the origin, that is a circle of radius R, centered at the origin. This circle cuts from the real axis an open interval $z \in [-R, +R]$ that is the locus of the real modulus inequality $|x| \leq R$. In the same way we can write an equation of a circle of radius R and centered at the point $z_0 \in \mathbb{C}$ as $|z - z_0| = R$, and its interior respectively with inequality. We use complex conjugation for one more practical reason - to define division in $\mathbb{C}$. Note that the multiplicative inverse of a complex number z may be defined as

$$z^{-1} = \frac{1}{z} = \frac{\bar{z}}{z\bar{z}} = \frac{\bar{z}}{|z|^2}$$

and if z lies on the unit circle $\mathbb{S}^1$ we have $z^{-1} = \bar{z}$. With this in mind, we easily derive

$$\frac{z_1}{z_2} = \frac{z_1\bar{z}_2}{|z_2|^2} = \frac{x_1x_2 + y_1y_2}{x_2^2 + y_2^2} + i\frac{x_2y_1 - x_1y_2}{x^2 + y^2}, \quad z_2 \neq 0 \tag{2.10}$$

which allows for dividing complex numbers.

Example:

$$\frac{5 + 3i}{1 - i} = \frac{(5 + 3i)(1 + i)}{|1 - i|^2} = \frac{5 - 3 + i(5 + 3)}{1 + 1} = 1 + 4i.$$

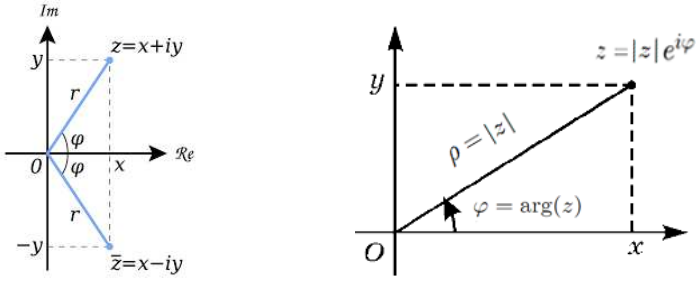

Figure 2.1: Complex conjugation as a reflection from the real axis (on the left) and the trigonometric representation as polar coordinates in the plane (on the right).

Note that on the unit circle the inversion coincides with complex conjugation. For example, for the imaginary unit we have $i^{-1} = \bar{i} = -i$, i.e., the additive and multiplicative inverse coincide. For more exciting examples see the exercises at the end of this chapter.

2.3 Trigonometric form and *Moivre*'s Formula

It is clear that there is more than one way to give a description of points in the plane - what we used so far was the standard *Cartesian* coordinates, orthogonal projections over the coordinate axes, providing the algebraic representation of a complex number. We may think of these coordinates as of some infinitely fine net, knitted by vertical ($x = const.$) and horizontal ($y = const.$) straight lines. However, there is another quite natural way to assign coordinates to a planar point, that relies on the notion of distance: firstly, foliate the plane into concentric circles and then describe each point by saying which circle it belongs to (or how far it is from the origin) and what is its angular parameter on this particular circle. The angle φ is measured counterclockwise with respect to the positive x-axis. Then we introduce an alternative orthogonal net in $\mathbb{R}^2$, consisting of concentric circles $|z| = const.$ and radial rays, starting from the origin $\varphi = const.$ These coordinates, usually referred to as *polar coordinates*, are denoted by[3]

$$\rho = \sqrt{x^2 + y^2}, \qquad \varphi = \operatorname{atan}_2 \frac{y}{x}$$

[3]the *proper quadrant inverse tangent* $\operatorname{atan}_2(x)$ accounts for the signs of $\sin x$ and $\cos x$.

and in this particular context, also called *modulus* $\rho = |z|$ and *argument* $\varphi = \arg(z)$ of the complex number z. Then, the polar change (Figure 2.1)

$$\begin{aligned} x &= \rho \cos \varphi \\ y &= \rho \sin \varphi, \quad \varphi \in [0, 2\pi) \end{aligned}$$

may be directly substituted in the algebraic representation of z to obtain

$$z = |z|(\cos \varphi + i \sin \varphi) \tag{2.11}$$

which is the *trigonometric representation* of the complex number z.

Polar coordinates are singular at the origin. To see this, we only need to try and follow one ray $\varphi = const$ through this point - the angular variable undergoes a jump from φ to $\varphi \pm \pi$ (a phase jump). In the same way each smooth curve passing through the origin has this property. This effect is well known in classical electrodynamics, optics and quantum mechanics, where complex numbers play a central role in the description of natural phenomena. On the other hand, properties of complex numbers should be independent on the choice of coordinates. In particular, summation is rather straightforward, complex conjugation may be viewed as inversion of the argument $\varphi \to -\varphi$, while multiplication (2.6) can be written as

$$z_1 z_2 = |z_1||z_2| \, [\, \cos \varphi_1 \cos \varphi_2 - \sin \varphi_1 \sin \varphi_2 + i(\sin \varphi_1 \cos \varphi_2 + \cos \varphi_1 \sin \varphi_2) \,]$$

in which we easily recognize the well-known formulae for $\cos(\varphi_1 + \varphi_2)$ and $\sin(\varphi_1 + \varphi_2)$, so we may write the latter as

$$z_1 z_2 = |z_1||z_2| \, [\, \cos(\varphi_1 + \varphi_2) + i \sin(\varphi_1 + \varphi_2) \,]. \tag{2.12}$$

From the above it becomes clear that multiplication complex numbers may be interpreted geometrically as spiral motion (Figure 2.2), i.e., rotation and dilatation (scaling), that one of the factors inflicts on the other. In particular, a complex number of unit modulus $z \in \mathbb{S}^1$ acts as a pure rotation by an angle, equal to its argument. On the other hand, if the argument is zero we end up with a real non-negative number that acts as a pure stretch, or dilatation. We focus on the former case as the latter is studied in high-school.

Formula (2.12) may be implemented repeatedly, which immediately implies

$$\prod_{k=1}^{n} z_k = \prod_{k=1}^{n} |z_k| \left(\cos \sum_{k=1}^{n} \varphi_k + i \sin \sum_{k=1}^{n} \varphi_k \right) \tag{2.13}$$

and in particular, for $z_1 = z_2 = \ldots = z_n = z$ we deduce *Moivre's formula*:

$$z^n = |z|^n \left(\cos n\varphi + i \sin n\varphi \right) \tag{2.14}$$

that is very convenient for calculating a power of a complex number.

Example 1. To calculate $(1+i)^{101}$ first we need to express $1+i$ in trigonometric form. Since $|1+i| = \sqrt{1^2+1^2} = \sqrt{2}$, we have $1+i = \sqrt{2}\left(\frac{1}{\sqrt{2}} + \frac{i}{\sqrt{2}}\right)$ and the number in the brackets lies on the unit circle, which means that its real part is cosine and the imaginary - sine of some argument φ. Since they are equal, it is not hard to find that $\varphi = \frac{\pi}{4}$ and according to (2.14) we have $(1+i)^{101} = \sqrt{2}^{101} \cos\frac{101\pi}{4} + i \sin\frac{101\pi}{4} = 2^{50}\sqrt{2} \cos\frac{5\pi}{4} + i \sin\frac{5\pi}{4}$ where make use of the fact that $\frac{101\pi}{4} = \frac{5\pi}{4} + 24\pi$ and the last term on the right can be dropped from the argument of sin and cos as they are both 2π-periodic functions. Then, finally, we obtain $(1+i)^{101} = -2^{50}(1+i)$.

The present paragraph uses some basic constructions of calculus, e.g. limits and derivatives, which may skipped by an unprepared student. For two complex numbers on the unit circle $z, w \in \mathbb{S}^1$ we have the representation

$$z = \cos\varphi + i\sin\varphi, \qquad w = \cos\psi + i\sin\psi$$

and their product is reduced, according to formula (2.12), to addition of the arguments, i.e., $zw = \cos\varphi + \psi + i\sin(\varphi + \psi)$, which resembles suspiciously the multiplication law for exponential functions $\alpha^\varphi \alpha^\psi = \alpha^{\varphi+\psi}$ and we assume there is a complex number α (real exponents cannot be periodic), such that $z = \alpha^\varphi$. It is actually possible to show that $z = e^{i\varphi}$, where the base $e \approx 2.718\ldots$ (similarly to π) is a *transcendent* number - it cannot be written as any kind of fraction, finite or infinite periodic decimal number. Instead, it is usually defined as a limit of a sequence or the sum of an infinite series

$$e = \lim_{n\to\infty} \left(1 + \frac{1}{n}\right)^n = \sum_{k=0}^{\infty} \frac{1}{k!} \tag{2.15}$$

which allows for writing

$$e^{i\varphi} = \lim_{n\to\infty} \left(1 + \frac{i\varphi}{n}\right)^n = \sum_{n=0}^{\infty} i^n \frac{\varphi^n}{n!} = \sum_{n=0}^{\infty} (-1)^n \frac{\varphi^{2n}}{(2n)!} + i \sum_{n=0}^{\infty} (-1)^n \frac{\varphi^{2n+1}}{(2n+1)!}$$

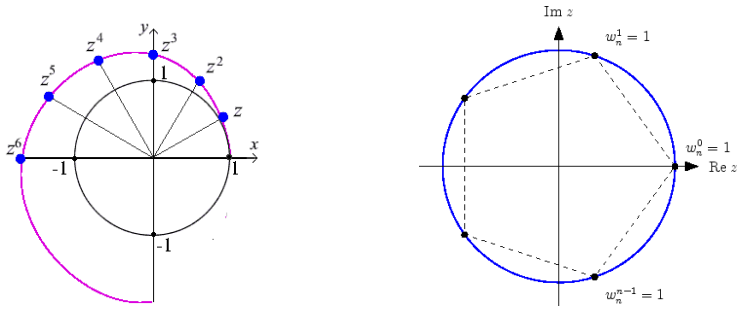

Figure 2.2: Complex powers (on the left) and complex roots of unity (on the right).

where for the last equality we use the fact that i^n equals ± 1 for even n and $\pm i$ for odd power n. The two series on the right are the well-known expansions of $\cos\varphi$ and $\sin\varphi$. However, a much simpler way to derive the above exponential expression is to consider $z(\varphi) = \cos\varphi + i\sin\varphi$ as a function of φ and differentiate, which yields $(\cos\varphi + i\sin\varphi)' = i\cos\varphi - \sin\varphi$, so we need to have $z' = iz$ with the initial condition $z(0) = \cos 0 = 1$, which has the unique solution $z = e^{i\varphi}$. Then, allowing non-unit moduli, we may write also

$$z_1 z_2 = |z_1||z_2| e^{i(\varphi_1+\varphi_2)} \tag{2.16}$$

while from the even and the odd part of $z(\varphi)$, using complex conjugation

$$e^{\pm i\varphi} = \cos\varphi \pm i\sin\varphi$$

we easily obtain the expressions (due to *Euler*)

$$\cos\varphi = \frac{e^{i\varphi} + e^{-i\varphi}}{2}, \qquad \sin\varphi = \frac{e^{i\varphi} - e^{-i\varphi}}{2i} \tag{2.17}$$

which may be considered as a definition and appear useful in various cases.

Complex Roots. We may use (2.14) in order to derive a formula for the n^{th} root of a complex number. All we have to do is assume that $z = w^n$ for some complex number $w = |w|(\cos\psi + i\sin\psi)$. Then, we obviously have $|w|^n = |z|$ and $n\psi = \varphi + 2k\pi$ as formula (2.14) suggests. In other words

$$|w| = |z|^{\frac{1}{n}}, \qquad \psi = \frac{\varphi + 2k\pi}{n}. \tag{2.18}$$

where $|z|^{\frac{1}{n}}$ denotes the (unique) positive real n-th root of $|z|$. Note that these are in fact n solutions for $k = 0 \ldots n-1$. As k becomes equal to n,

a period of 2π can be factored out and thus the "new" solutions begin to repeat the "old" ones. Then, we may conclude

$$\sqrt[n]{z} = |z|^{\frac{1}{n}} \left(\cos \frac{\varphi + 2k\pi}{n} + i \sin \frac{\varphi + 2k\pi}{n} \right), \quad k = 0 \dots n-1. \tag{2.19}$$

Example 2. We may easily find all zeroes of the polynomial $z^6 - 1 = 0$ (sixth roots of unity) - since all of them have the same (unit) modulus, we only need to take care of the angular parameter. The base root is obviously real $(\sqrt[6]{1})_0 = 1$ as we have $\varphi = 0$ and $\psi_0 = \frac{\varphi}{6} = 0$. Each successive root then is obtained from the previous one by adding a phase $\Delta\psi = \frac{2\pi}{6} = \frac{\pi}{3}$ to the previous one: $\psi_k = \frac{k\pi}{3}$, $k = 0 \dots 5$, which comes to show that the roots are vertices of a regular hexagon, inscribed in the unit circle. Substituting $z = 1$ and $n = 6$ into (2.19) it is straightforward to obtain

$$(\sqrt[6]{1})_0 = 1, \qquad (\sqrt[6]{1})_1 = \frac{1}{2} + i\frac{\sqrt{3}}{2}, \qquad (\sqrt[6]{1})_2 = -\frac{1}{2} + i\frac{\sqrt{3}}{2}$$

$$(\sqrt[6]{1})_3 = -1, \qquad (\sqrt[6]{1})_4 = -\frac{1}{2} - i\frac{\sqrt{3}}{2}, \qquad (\sqrt[6]{1})_5 = \frac{1}{2} - i\frac{\sqrt{3}}{2}.$$

Derive the same result using (2.3) and the factorization

$$z^6 - 1 = (z^3 - 1)(z^3 + 1) = (z - 1)(z^2 + z + 1)(z + 1)(z^2 - z + 1).$$

Note that the procedure of obtaining the n^{th} root of an arbitrary complex number does not differ much from the above example - first we need to define the circle all roots lie on - it is centered at the origin and has radius $R = |z|^{\frac{1}{n}}$, then we need the "base" root (the one with argument $\frac{\varphi}{n}$) obtain each successive by adding a phase of $\frac{2\pi}{n}$ (one n^{th} of the full revolution) and thus, inscribe a regular polygon in the corresponding circle with one of its vertices pointing in the direction of $\psi_0 = \frac{\varphi}{n}$ (see Figure 2.2).

Exercises:

1. What is $|z^2|$ compared to $|z|^2$ and when are these equal?
2. Find the intersection points of $|z - i| = 1$ and $|z - 1| = 1$.
3. Describe visually the set of points z such that $|z - z_0| = R$ for a given $z_0 \in \mathbb{C}$ and $R > 0$.

4. Calculate the Euclidean distance between the points $3-4i$ and $1+i$.

5. Present the numbers $\dfrac{1-i}{1+i}$, $\dfrac{2+3i}{3-i}$ and $\dfrac{i+\sqrt{2}}{1-i\pi}$ with their real and imaginary parts.

6. Make a sketch of the annulus $1 < |z-i| < 2$.

7. Draw the area $\pi < |z| < 4, \quad \pi/3 < \arg(z) < 2\pi/3, \quad \Re z > 3$.

8. Describe the image of the coordinate net $x = const$, $y = const$ under the transformation $f : z \to z^2$.

9. Find the trigonometric form of $i-1$ and $1+i\sqrt{3}$.

10. Calculate $\dfrac{(1-i)^{2013}}{(1-i\sqrt{3})^{137}}$.

11. Find all roots of

    a) $z^6 + 32\,i = 0$

    b) $z^5 + i = 1$

    c) $z^4 + 1 = i\sqrt{3}$

    d) $(z+1)^7 + (z-1)^7 = 0$

12. Express $\dfrac{z^n}{\bar{z}^k}$ and $\dfrac{(z-z_0)^n}{(z+z_0)^k}$ in trigonometric form.

13. Show that $e^{ik\pi} = ie^{-i(k+\frac{1}{2})\pi} = (-1)^k, \quad k \in \mathbb{N}$.

14. Prove that $\displaystyle\sum_{k=0}^{n-1} e^{i\frac{2k\pi}{n}} = 0, \quad n \in \mathbb{N}^+$.

15. Draw pictures of the loci in $\mathbb{C}$ described by

    a) $1 < |z-i| < \pi$

    b) $|z| = \arg z$

c) $|z-1|+|z+1|=1$

d) $z^2<1$

16. Find the complex root

$$\sqrt[3]{i-1}, \qquad \sqrt[5]{\frac{1-i}{1+i\sqrt{3}}}$$

17. Using Moivre's result, derive a formula for $\cos nx$ and $\sin nx$.

18. Find the real and imaginary part of e^z, $\sin z$ and $\cos z^2$.

Hint: Use (2.17) and the so-called *hyperbolic* sine and cosine functions

$$\cosh x = \frac{e^x + e^{-x}}{2}, \qquad \sinh x = \frac{e^x - e^{-x}}{2} \tag{2.20}$$

that constitute the *even* and the *odd* part of the real exponent. Note that the above may be defined also from the equalities

$$\cosh x = \cos ix, \quad \sinh x = i \sin ix$$

$$\cos x = \cosh ix, \quad \sin x = -i \sinh ix.$$

Moreover, we have the hyperbolic trigonometric identity

$$\cosh^2 x - \sinh^2 x = 1$$

that can be used to parameterize a hyperbola, as well as

$$\cosh' x = \sinh x, \quad \sinh' x = \cosh x$$

since $(e^x)' = e^x$ and finally

$$\sinh(x+y) = \sinh x \cosh y + \cosh x \sinh y$$

$$\cosh(x+y) = \cosh x \cosh y + \sinh x \sinh y.$$

Chapter 3

Polynomials

We refer to any sum of the type

$$P_n(z) = \sum_{k=0}^{n} a_k z^{n-k} = a_0 z^n + a_1 z^{n-1} + \ldots + a_1 z + a_n, \qquad a_0 \neq 0 \quad (3.1)$$

with $a_k, z \in \mathbb{C}$ as a *polynomial* of order (degree) n on the complex variable z with complex coefficients a_k. In particular, a_0 is called the *highest order coefficient*, while a_n is referred to as the *free term*. With a few modifications one may generalize the above definition to polynomials on several variables, matrix polynomials, or such of infinite order (*power series*). Furthermore, we shall often restrict our considerations to polynomials with real, rational or even integer coefficients, as well as such, depending on a real parameter. Non-zero constants, linear and quadratic functions are common examples. Polynomials have some distinguished properties among other functions that make them a preferred tool for many analytical and numerical considerations. For example, they form a closed set under summation and multiplication. In order to sum two polynomials, we only need to sum the coefficients in front of corresponding powers of the variable z. Similarly, the difference $P-Q$ is defined as a sum of P and $-Q$, where multiplication with a number (in this case -1) means multiplying all the coefficients by this number. Two polynomials are said to be equal if all their coefficients are equal (equivalently, $P - Q = 0$). Multiplication of polynomials, however, is a bit more complicated - the product of $P_n(z) = \sum_{k=0}^{n} a_k z^{n-k}$ and $Q_m(z) = \sum_{k=0}^{m} b_k z^{m-k}$, is a polynomial $R_{m+n}(z) = \sum_{k=0}^{n} c_k z^{m+n-k}$ with coefficients $c_k = \sum_{i+j=k} a_i b_j$.

3.1 Zeroes and Factorization

Polynomials with complex coefficients possess finite sets of isolated *zeroes* (roots), i.e., points $z_k \in \mathbb{C}$, for which $P_n(z_k) = 0$. This set is non-empty if the order (degree) is greater than zero, as stated by the following

Theorem 1 *Each polynomial $P_n(z)$ with $n > 0$ has at least one zero.*

known as the *main theorem of the algebra*. We are not going to prove this theorem here, but instead, only illustrate its validity for low values of n. Firstly, for $n = 0$, $P_n(z)$ is a constant, so unless $P_n \equiv 0$, there are no zeroes. For linear functions of the type $P_1(z) = a_0 z + a_1$, since $a_0 \neq 0$ by definition, one always has the zero $z_0 = -\dfrac{a_1}{a_0}$. For quadratic functions $P_2(z) = a_0 z^2 + a_1 z + a_2$, it is well-known that the two zeroes are given by

$$z_{1,2} = \frac{-a_1 \pm \sqrt{\Delta}}{2a_0}, \qquad \Delta = a_1^2 - 4a_0 a_2. \tag{3.2}$$

Next, we proceed with polynomial division: like with integer numbers we write

$$\frac{P_n(z)}{Q_m(z)} = R_{n-m}(z) + \frac{S_l(z)}{Q_m(z)}, \qquad m \leq n \tag{3.3}$$

and refer to P_n as to the *dividend*, Q_m the *divider* and R_{n-m} - as the *quotient* of the division. The polynomial S_l with $l < m$ is called the *remainder* of that division and we say that Q_m *divides* P_n whenever S_l vanishes. Next, we consider an example, explaining the so-called *long division* of polynomials. Let $P_n(z) = 3z^4 - z^2 + 1$ and $Q_m(z) = z^2 + 1$, then we take the highest order terms and see how many times the one in the divider fits into that of the dividend. Obviously, we need to multiply by a factor of $3z^2$, so this is the first part of the quotient:

$$\frac{3z^4 - z^2 + 1}{z^2 + 1} = 3z^2 + \dots$$

To see what follows we need to subtract

$$3z^4 - z^2 + 1 - 3z^2(z^2 + 1) = -4z^2 + 1$$

which appears as an intermediate remainder in our long division procedure. Its order, however, is not less than the order of the divider $m = 2$, so we may continue dividing it to $z^2 + 1$ in the same way. The coefficient of proportionality between the highest order terms is now -4, which appears as

the next contribution in the quotient, and to see what the next remainder is, we subtract $-4z^2+1-4(z^2+1)=-3$, which is already the final remainder, since its order (zero) is now strictly less than the one of the divider. Therefore, we may write

$$\frac{3z^4-z^2+1}{z^2+1}=3z^2-4+\frac{-3}{z^2+1}$$

so the quotient in this case is $3z^2-4$ and the remainder is -3, respectively. In many situations it is important to consider a particular case of division, in which the divider is a linear function of the type $Q_1=z-c$. For arbitrary order n one has the following very useful

Theorem 2 *The remainder of the division of $P_n(z)$ with $z-c$ equals $P_n(c)$.*

Proof. It suffices to set $z=c$ in the expression

$$P_n(z)=(z-c)R_{n-1}(z)+S_0. \qquad \square$$

One particular application of this theorem is to choose $c=z_1$ to be a zero of $P_n(z)$, that always exists for $n>0$ according to the main theorem of algebra. Then, $z-z_0$ divides P_n and the quotient is a polynomial of degree $n-1$. If this number is still greater than 0, we may find a zero, say z_2, so that $z-z_2$ divides R_{n-1}, and thus P_n, which isolates a factor of $(z-z_1)(z-z_2)$. Proceeding like this we end up with a constant remainder, given by the above theorem, which implies the following

Theorem 3 *Each polynomial of order n has exactly n zeroes $z_i\in\mathbb{C}$ (some of which may be coincident) and the may thus be factorized as*

$$P_n(z)=a_0(z-z_1)(z-z_2)\dots(z-z_n). \tag{3.4}$$

One straightforward application of the above result yields the famous *Vieta's formulae*. Let us express the polynomial (3.4) in two familiar ways

$$a_0z^n+\ldots+a_{n-1}z+a_n=a_0(z-z_1)(z-z_2)\dots(z-z_n).$$

Then, we open the brackets on the righthand side and equate the coefficients in front of the corresponding powers of z, thus obtaining the n relations

$$\sum_{i_1<i_2\ldots<i_k} z_{i_1}\dots z_{i_k}=(-1)^k\frac{a_k}{a_0} \tag{3.5}$$

and in particular $\displaystyle\sum_{k=1}^{n} z_k=-\frac{a_1}{a_0}$ and $\displaystyle\prod_{k=1}^{n} z_k=(-1)^n\frac{a_n}{a_0}$ are the first and the last of those (except for the trivial one $a_0=a_0$). In the familiar example $n=2$, one has $z_1+z_2=-\dfrac{a_1}{a_0}$ and $z_1z_2=\dfrac{a_2}{a_0}$ (derive the formulae for $n=3$).

3.2 Horner's Method and its Applications

Let us consider once more Theorem 2 and equate, just as before, the two expressions of

$$a_0 z^n + \ldots + a_{n-1} z + a_n = (z - c) R_{n-1}(z) + S_0.$$

where $R_{n-1}(z) = b_0 z^{n-1} + \ldots b_{n-2} z + b_{n-1}$. Opening the brackets on the righthand side and relating the coefficients in front of equal powers of z, we end up with

$$a_0 = b_0, \qquad a_1 = b_1 - b_0 c, \ldots \qquad a_n = S_0 - b_{n-1} c$$

which may be reversed into an iterative procedure, known as *Horner's method*, for obtaining the coefficients b_k of the quotient and the remainder S_0 (denoted for convenience as b_n) from those of the dividend a_k, namely as

$$b_0 = a_0, \qquad b_k = a_k + b_{k-1} c. \tag{3.6}$$

This is a powerful technique for polynomial division (referred to as *synthetic division*), although the divider is bound to be linear. On the other hand, one may factorize the latter as formula (3.4) shows and then perform successive division. Consider the simple example $(z^3 + 1) : (z^2 - 1)$, in which

$\{a_k\}$	1	0	0	1
$c_1 = 1$	1	1	1	2
$c_2 = -1$	1	0	1	1

$$\Rightarrow \quad \frac{z^3 + 1}{z^2 - 1} = z + \frac{1}{z - 1}$$

as can be seen also by long division. Now, to explain how it can be obtained from the above table. The first two numbers in the final row are the coefficients of the quotient, while the last two - those of the remainder. In particular, it is not hard to see that $S_1(z) = b_{n-1}(z - c_1 - c_2) + b_n$, which in this case turns out to be $z - 1 + 1 + 1 = z + 1$ and thus the above result. Note that it is more efficient to divide by $z + 1$ first as the intermediate remainder would be zero. This construction may easily be generalized to dividers of higher order, but it appears more complicated compared to the long division. Let us now consider another example, in which the dividend is $P_n(x) = 3z^4 - 2z^3 - 3z^2 - z + 1$ and the for the divider we have $z^2 - 3z + 2$. The latter obviously[1] has roots $z_1 = 1$ and $z_2 = 2$, so it may be written as $(z-1)(z-2)$. Then, instead of the usual long division, explained previously, one may divide $P_n(z)$ first by $z - 1$ and then by $z - 2$ via the Horner method

[1] the sum of the coefficient is zero, thus $z_1 = 1$ and z_2 follows from the Vieta's formulae.

$\{a_k\}$	3	-2	-3	-1	1
$c_1 = 1$	3	1	-2	-3	-2
$c_2 = 2$	3	7	12	21	40

where the first row consists of the coefficients of the dividend and in each cell below we write the sum of the number in the cell above and the product of the one on the left with the number c written in the first column, except for the first coefficient, which remains the came in every row. Thus, we have

$$3z^4 - 2z^3 - 3z^2 - x + 1 = (3z^2 + 7z + 12)(z^2 - 3z + 2) + 21(z - 3) + 40.$$

Another use of Horner's method is obtaining an expansion of a given polynomial $P_n(z)$ in terms of a shifted argument, say $z - c$. For example, let us expand $z^3 + 3z^2 - 2z + 1$ in powers of $z + 1$. This yields applying successively the procedure with $c = -1$ and taking the intermediate remainders as coefficients of the resulting polynomial, starting with the free term and building up to the highest order coefficient, as shown in the table below

$\{a_k\}$	1	2	-2	1
-1	1	2	-4	$\boxed{5}$
-1	1	1	$\boxed{-5}$	
-1	1	$\boxed{0}$		
-1	$\boxed{1}$			

which finally yields the corresponding expansion in the form

$$z^3 + 3z^2 - 2z + 1 = (z + 1)^3 - 5(z + 1) + 5.$$

Rational Zeroes of Polynomials with Integer Coefficients

There is one particularly well-known application of Horner's algorithm as a numerical procedure for obtaining the rational zeroes (if any) of polynomials with integer coefficients. Let us first note that if c is a zero of $P_n(z)$, then Theorem 2 asserts that the remainder S_0 in the division of P_n to $z - c$ has to vanish. Thus, obtaining zero in the last cell of a Horner table shows that we have come across a root. Guessing polynomial zeroes, however, is not reasonable, unless we have some initial clue. The following theorem suggests one such clue.

Theorem 4 *If $z_0 = \dfrac{p}{q}$ $(p, q \in \mathbb{Z})$ is a rational zero of $P_n(z)$, then p divides the free term a_n and q - the highest-order coefficient a_0, respectively.*

Proof. Since z_0 is a root, we obviously have

$$a_0\frac{p^n}{q^n}+\ldots+a_{n-1}\frac{p}{q}+a_n=0$$

which yields, upon multiplication by q^{n-1}, the equality

$$\frac{a_0}{q}p^n=-a_1p^{n-1}-\ldots-a_nq^{n-1}\in\mathbb{Z}.$$

Since, on the one hand, the righthand side is integer, while on the other, p and q are *co-prime*, i.e., have no common divisors except 1, it follows that $\frac{a_0}{q}\in\mathbb{Z}$ and therefore, q divides a_0. Similarly, multiplying the initial equality by $\frac{q^n}{p}$, we obtain $\frac{a_n}{p}q^n=-a_0p^{n-1}-\ldots-a_{n-1}q^{n-1}\in\mathbb{Z}$, which comes to show that p divides a_n. □

Corollary 1 *If $P_n(z)$ has integer coefficients and $a_0=1$, then all rational zeroes are also integer.*

Let us consider one example for the latter, namely $P_n(z)=z^3-2z^2-4z+3$. Obviously, rational roots can only be ±1 and ±3. We may easily test each of these possibilities and then use Horner's method to effectively reduce the order. Thus, we find the only rational root to be $z_0=3$ and

$\{a_k\}$	1	−2	−4	3
$c=3$	1	1	−1	0

so we have $P_n(z)=(z-3)(z^2+z-1)$ and the quadratic equation on the righthand side can easily be solved via formula (12.1), which yields $P_n(z)=(z-3)(z+\varphi_+)(z+\varphi_-)$, where $\varphi_\pm=\frac{\sqrt{5}\pm1}{2}$ are the two solutions of the famous *golden ratio*. Another example, which involves rational (not only integer) zeroes might be the cubic equation $2z^3-z^2-18z+9=0$. According to Theorem 4, the only possible rational zeroes are $\pm1,\pm3,\pm9,\pm\frac{1}{2},\pm\frac{3}{2},\pm\frac{9}{2}$. It is easy to see that ±1 cannot be a root, so we proceed with the remaining values, for which the Horner's method yields

	2	−1	−18	9
$c=3$	2	5	−3	0
$c=-3$	2	−1	0	
$c=\frac{1}{2}$	2	0		

so we may factorize $2z^3-z^2-18z+9=(2z-1)(z+3)(z-3)$.

3.3 Partial Fraction Expansion

In the context of polynomial division, we consider a useful technique to transform *rational functions*, i.e., fractions of polynomials, into simpler form. Here we consider only proper rational functions $R(z) = \dfrac{Q_m(z)}{P_n(z)}$ with $m < n$. Then, if P_n has only real simple zeroes z_k, the representation (3.4) yields

$$R(z) = \frac{Q_m(z)}{P_n(z)} = \frac{A_1}{z - z_1} + \frac{A_2}{z - z_2} + \ldots + \frac{A_n}{z - z_n} \tag{3.7}$$

where the constants A_k are determined by the above equality. It is easy to illustrate this with an example: take $P_n = 1 - z^2$ and $Q_m = 1$, hence

$$\frac{1}{1 - z^2} = \frac{A_1}{1 - z} + \frac{A_2}{1 + z} = \frac{A_1(1 + z) + A_2(1 - z)}{1 - z^2} = \frac{(A_1 - A_2)z + A_1 + A_2}{1 - z^2}$$

and since the denominators are equal, so must be the numerators, but equality of polynomials demands equality of all corresponding coefficients, which yields a system of equations for the constants A_i, namely $A_1 - A_2 = 0$ and $A_1 + A_2 = 1$ with the obvious solution $A_1 = A_2 = 1/2$. Then, we may write

$$\frac{1}{1 - z^2} = \frac{1}{2}\left[\frac{1}{1 - z} + \frac{1}{1 + z}\right].$$

Let us now consider the case of a multiple root z_k, which contributes to the factorization as $(z - z_k)^p$. It adds to the expansion (3.7) a sum of the type

$$\frac{B_1}{z - z_k} + \frac{B_2}{(z - z_k)^2} + \ldots + \frac{B_p}{(z - z_k)^p}. \tag{3.8}$$

To see this we use another example, in which the dividend is $Q = z^2 + 1$ and the divider $P = z^3 - z^2 - z + 1$, respectively. Obviously $z_0 = 1$ is a zero of the later, since the sum of its coefficients vanishes. Thus, it may be written also as $P = (z^2 - 1)(z - 1) = (z + 1)(z - 1)^2$ and in the expansion

$$\frac{z^2 + 1}{(z + 1)(z - 1)^2} = \frac{A}{z + 1} + \frac{B_1}{(z - 1)} + \frac{B_2}{(z - 1)^2}$$

we already have three constants to deal with. Transforming the righthand side as in the previous example, we obtain the polynomial equality

$$z^2 + 1 = A(z - 1)^2 + B_1(z^2 - 1) + B_2(z + 1)$$

in which the coefficients in front of independent powers of z yield the system

$$A + B_1 = 1, \qquad 2A - B_2 = 0, \qquad A - B_1 + B_2 = 1$$

with the obvious solution $A = B_1 = 1/2$ and $B_2 = 1$, so we may write

$$\frac{z^2+1}{z^3-z^2-z+1} = \frac{1}{2}\left[\frac{1}{z+1} + \frac{1}{z-1}\right] + \frac{1}{(z-1)^2} = \frac{x}{z^2-1} + \frac{1}{(z-1)^2}.$$

Finally, let us consider the case of complex roots with non-vanishing imaginary parts. Since the latter come in pairs of complex conjugates, as it became clear in the previous chapter, they may all be thought of as arising from quadratic factors with negative discriminants in the decomposition. Each such factors contributes to the expansion (3.7) with a term of the form

$$\frac{Cz+D}{az^2+bz+c}, \qquad b^2 - 4ac < 0. \tag{3.9}$$

Take for example the rational function $\dfrac{z^4}{z^3+1}$. If we attempt to apply the procedure directly, it would simply not work, since the power of the numerator exceeds that of the denominator. In order to obtain a "proper fraction", we first need to divide and then work with the remainder alone, leaving the polynomial quotient aside. Long division in this case yields

$$\frac{z^4}{z^3+1} = z - \frac{z}{(z+1)(z^2-z+1)} + \frac{A}{z+1} + \frac{Cz+D}{z^2-z+1}$$

thus leading to the equality

$$A(z^2-z+1) + (Cz+D)(z+1) = z$$

from which we derive the system of equations

$$A + C = 0, \qquad A - C = 1, \qquad A + D = 0$$

determining the values of the coefficients $A = 1/2$ and $C = D = -1/2$. Finally, to sum up the general result, one has

$$R(z) = \frac{Q_m(z)}{P_n(z)} = \sum_{i=1}^{k-1} \frac{A_i}{z - z_i} + \sum_{j=1}^{p} \frac{B_j}{(z-z_k)^j} + \frac{Cz+D}{az^2+bz+c} \tag{3.10}$$

where the first sum on the righthand side corresponds to the $k-1$ simple roots, the second one - to the multiple root z_k (which is assumed to be only one and of multiplicity p), while the third sum gives the contribution of the complex pair of roots, arising form the quadratic equation with negative discriminant in the denominator (also assumed to be just one).

Heaviside Expansion

Here we present an alternative method for partial fraction expansion due to Heaviside, which yields the solutions in a simpler form, but relies on some basic constructions of mathematical analysis (limits and derivatives). Consider the expansion (3.7) of the proper ($m < n$) rational function $R(z)$ assuming for now that all the roots of the denominator $P_n(z)$ are simple (not necessarily real). Multiplying both sides of the equality successively with $z - z_k$ for all the roots z_k and then taking the limit $z \to z_k$ we may easily isolate the corresponding coefficient A_k in the form

$$A_k = \lim_{z \to z_k} (z - z_k) R(z) \tag{3.11}$$

since all the remaining terms contain $z - z_k$ as a factor, and thus vanish in the limit. Note that, after the cancelation, it suffices to set $z = z_k$ in the righthand side above as the zeroes are simple and isolated. However, we keep this notation to focus on the generality of the construction[2]. On the other hand, the above expression may also be written via differentiation as

$$A_k = \frac{Q_m(z_k)}{P_n'(z_k)}. \tag{3.12}$$

Now, let us consider the case of a multiple root with the notation (3.8), in which it is not possible to isolate the coefficients B_m directly. Instead, we multiply with $(z - z_k)^p$ that yields a polynomial expression of the type

$$(z - z_k)^p R(z) = \sum_{m=1}^{p} B_m (z - z_k)^{p-m}.$$

Then, differentiating $p - m$ times, we eliminate the terms B_j for $j < m$, while taking the limit $z \to z_k$ eliminates those with $j > m$ since they contain $z - z_k$ as a factor, thus leaving only the constant term $(p - m)!\, B_m$ on the righthand side (the pre-factor is due to the successive differentiation). Finally, this proves the formula

$$B_m = \frac{1}{(p-m)!} \lim_{z \to z_k} \left[(z - z_k)^p R(z) \right]^{(p-m)}. \tag{3.13}$$

Consider the above example involving double roots

$$R(z) = \frac{z^2 + 1}{z^3 - z^2 - z + 1} = \frac{A}{z + 1} + \frac{B_1}{(z - 1)} + \frac{B_2}{(z - 1)^2}.$$

[2] which reappears in various contexts, e.g. in the so-called *residual calculus*.

Using either (3.11) or (3.12) it is straightforward to obtain

$$A = \lim_{z\to -1} \frac{z^2+1}{(z-1)^2} = \frac{1}{2}$$

while for the B-coefficients it is necessary to multiply with $(x-1)^2$ and then, differentiate in order to obtain

$$B_1 = \lim_{z\to 1} \left(\frac{z^2+1}{z+1} \right)' = \frac{z^2+2z-1}{z^2+2z+1}_{|z=1} = \frac{1}{2}, \qquad B_2 = \frac{z^2+1}{z+1}_{|z=1} = 1.$$

Note that the presence of simple roots does not disturb the procedure for the B-coefficients given by formula (3.13) as the corresponding terms are multiplied by $(z-z_k)^p$, which vanishes after the limit is taken (there is no cancelation of factors whatsoever). Moreover, it is not necessary to assume that the zeroes are real in any of the above methods - they both work perfectly well also with complex roots. However, in many practical applications (e.g. for the integration of rational functions) we prefer to work with real expressions and thus, combine the contribution of complex conjugate pairs into a real expression without further decomposing quadratic polynomials with negative discriminants. Using the Heaviside expansion (also known as *cover-up method*), it is also possible to express the fractional decomposition in terms of real rational functions only. Consider one more simple example

$$R(z) = \frac{z+1}{z^3-z^2+z-1} = \frac{z+1}{(z-1)(z^2+1)} = \frac{A_0}{z-1} + \frac{A_1}{z-i} + \frac{A_2}{z+i}$$

in which we clearly have three simple roots in the denominator, namely $z_0 = 1$ and $z_{1,2} = \pm i$. Then, formula (3.11) determines the A-coefficients as

$$A_0 = (z-1)R(z)_{|z=1} = 1$$

and respectively

$$A_{1,2} = (z \mp i)R(z)_{|z=\pm i} = \frac{z+1}{(z-1)(z\pm i)}_{|z=\pm i} = \pm\frac{i}{4}(1\pm i)^2 = -\frac{1}{2}$$

so one may finally express the fractional expansion of $R(z)$ in the form

$$\frac{z+1}{z^3-z^2+z-1} = \frac{1}{z-1} - \frac{1}{2}\left[\frac{1}{z-i} + \frac{1}{z+i}\right] = \frac{1}{z-1} - \frac{z}{z^2+1}.$$

Problem Session

1. Divide the polynomials via long or synthetic division (Horner's rule)
   a) $(x^4 + 2x^3 + x - 1) : (x^2 + 1)$
   b) $(4x^4 + x^2) : (x + 1 + i)$
   c) $(2x^6 + 3x^4 - x^2 - 5x + 1) : (1 - i)$
2. Solve the equations
   a) $2x^3 - x^2 - 18x + 9 = 0$
   b) $x^4 + 2x^3 - 4x^2 - 5x - 6 = 0$
   c) $2x^4 - 11x^3 + 18x^2 - 4x - 8 = 0$
   d) $x^5 + 3x^4 - x^3 - 11x^2 - 12x - 4 = 0$
3. Prove that the roots x_k of the polynomial $P_3(x) = 36x^3 - 12x^2 - 5x + 1$ satisfy $x_1 + x_2 = x_3$ and then find them.
4. Set the value of the parameter λ in the equation $x^3 - 7x + \lambda = 0$, so that two of the roots are in a proportion $2 : 1$.
5. Compose a polynomial of order six with zeroes
$$\alpha,\ \frac{1}{\alpha},\ 1-\alpha, \frac{1}{1-\alpha},\ 1-\frac{1}{\alpha},\ \frac{1}{1-\frac{1}{\alpha}}, \qquad \alpha \in \mathbb{R}$$
6. Find the relation between the parameters a, b and c, so that the roots of the equation $x^3 + ax^2 + bx + c = 0$ constitute a geometric progression.
7. Obtain the fractional expansion of the following rational functions
   a) $\dfrac{x^5 + 1}{x^6 + x^4}$
   b) $\dfrac{x^4 - 3x^3 + x^2 + 4x - 1}{x^3 - 3x^2 + x - 3}$
   c) $\dfrac{3x^3 + 6}{x^4 - 5x^2 + 6}$
   d) $\dfrac{x^4 - 2x^2 + 3}{(x + 1)^5}$

   Hint: in the last example expand the numerator in powers of $x + 1$.

Chapter 4

Points and Vectors in $\mathbb{R}^2$

A *real vector space* (also referred to as a linear space) is a set V, which contains all *linear combinations* of its elements with real coefficients, i.e.,

$$\forall \mathbf{v}_{1,2} \in V, \quad \alpha_{1,2} \in \mathbb{R} \quad \Rightarrow \quad \alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 \in V. \tag{4.1}$$

Complex vector spaces are defined in a similar way, letting this time $\alpha_{1,2} \in \mathbb{C}$. To see that this definition actually makes sense, let us provide a couple of examples. A trivial one is the field[1] of real numbers $\mathbb{R}$ itself, but it is not a complex vector space. Similarly, $\mathbb{Q}$ may be considered a rational vector space, i.e., such with rational coefficients, but it is neither real, nor complex one. For example, pick a coefficient $\alpha_1 = \sqrt{2}$ and $\alpha_2 = \sqrt{-1}$, that obviously destroys the rationality (and even reality) of the construction (4.1) restricted to non-zero pairs of rational numbers $\mathbf{v}_{1,2} \in \mathbb{Q}$. It is only fair to call elements of vector spaces *vectors*. Therefore, we may think of real and complex numbers as of vectors, but those are only very basic examples. We are going to gain more experience with vectors and relate them directly to plane and space geometry as we advance in the algebra. Moreover, this will allow us to enjoy and benefit some additional structures these linear spaces possess. For now, we start with the familiar complex numbers and use some of their properties to build a consistent tool for studying planar geometry. Most of the constructions below, however, have a natural generalization in higher dimensions, which we are going to introduce in a more formal manner once we become acquainted with matrices.

[1] in algebra we refer to nonzero commutative division rings as *fields*.

Complex Numbers as Vectors in the Plane

Let us recall some facts about complex numbers. Firstly, both summation and multiplication with a real number are performed component-wise, i.e., if $z_k = x_i + iy_k \in \mathbb{C}$, $k = 1, 2$, then $z_1 + z_2 = (x_1 + x_2) + i(y_1 + y_2)$ and $\lambda z_i = \lambda x_k + i\lambda y_k$ for $\lambda \in \mathbb{R}$. Those are exactly the explicit rules needed for operations with vectors in the plane. We may write the *cartesian coordinates*, which measure the orthogonal projections over the coordinate axes, as we already know, in ordered rectangular tables, called *matrices*

$$z_k = x_k + iy_k \quad \longleftrightarrow \quad \begin{pmatrix} x_k \\ y_k \end{pmatrix}$$

which is more customary when dealing with vectors. Besides, this framework generalizes naturally to arbitrary dimension without the need of complex structures. The geometric interpretation as pointing arrows with given directions lengths remains in the higher dimensional case too.

4.1 Free and Attached Vectors

It is rather tempting to interpret vectors in $\mathbb{R}^2$ as complex numbers, or equivalently, points in the plane. Note, however, that points are usually associated with *radius-vectors* depicted by an arrows connecting the origin of the coordinate system with the given point, whose cartesian coordinates coincide with the vector components. The corresponding space (of points) is called *affine* and we also speak of *attached* vectors. They may be depicted as arrows fixed to a given point in the plane (or space). The coordinates of attached vectors and radius-vectors in particular undergo an alteration whenever the origin of the reference frame is shifted. In mechanics, such vectors are used for example to model velocities or forces acting on a certain point (e.g. the center of masses) of the body. *Free* vectors, on the other hand, are not related to any particular point and can thus be moved anywhere in space. Their coordinates remain the same in a moving (without rotation) reference frame. A good mechanical example would be that of a homogenous force field. An easy way to obtain a free vector is by subtracting two radius-vectors $\mathbf{r}_{1,2}$: since each of them undergoes a transformation of the type $\mathbf{r}_i = \mathbf{r}_i - \mathbf{a}$ whenever the reference frame is shifted by a *translation vector* $\mathbf{a}$, the difference $\mathbf{v}_{12} = \mathbf{r}_2 - \mathbf{r}_1$ remains unaltered. This difference also has a simple geometric interpretation. Recall that the vector sum of $\mathbf{r}_1$ and $\mathbf{r}_2$ is depicted as the main diagonal of the parallelogram, determined by the two radius-vectors (Figure 4.2). Similarly, the difference $\mathbf{r}_2 - \mathbf{r}_1$ is associated with

the second diagonal, i.e., an arrow starting at (the point with radius-vector) $\mathbf{r}_1$ and ending at $\mathbf{r}_2$. This arrow, however can be moved anywhere in space as long as its direction and length are preserved. If the subtraction takes place in the reverse direction, namely $\mathbf{r}_1 - \mathbf{r}_2$, it points form the end of $\mathbf{r}_2$ towards the one of $\mathbf{r}_1$. More generally, any linear combination of radius-vectors, in which the sum of all coefficients vanishes, yields a free vector. Note also that summation and multiplication with a number appear quite naturally in geometry if one attempts to change the origin of the coordinate system, or the scale of measurement, respectively. For example, if the point A is represented by a radius-vector $\mathbf{r}$ in the reference frame $\{K\}$, then shifting the latter to $\{K'\}$ with a translation vector $\mathbf{a}$, the coordinates of the point A are also altered (the point itself being fixed) and it is not difficult to see that this shifted radius-vector is given as $\mathbf{r}' = \mathbf{r} - \mathbf{a}$. Similarly, if we want for some reason to change the scale of measurement, e.g. from centimeters to meters, we need to multiply all vectors, and in particular, the radius-vectors, representing point locations, by a re-scaling factor, in our case $\lambda = 10^{-2}$.

4.2 The Polar Change

As we discussed above, cartesian coordinates represent a given vector via its orthogonal projections over the coordinate axes. Similarly, the trigonometric form of complex numbers, more broadly known in mathematics by the name of *polar coordinates*, represents a vector in terms of its magnitude (or length) usually denoted by ρ and the angle φ it makes with the positive direction of the x-axis. As it can be shown with an elementary geometric argument (right triangles), the relation between the two descriptions is given by the so-called *polar change* (see Figure 4.1)

$$x = \rho\cos\varphi, \qquad y = \rho\sin\varphi. \tag{4.2}$$

Conversely, one may easily derive expressions for the polar coordinates as functions of the cartesian ones in the form

$$\rho = \sqrt{x^2 + y^2}, \qquad \varphi = \operatorname{atan}_2\Big(\frac{y}{x}\Big)\cdot \tag{4.3}$$

Polar coordinates depict most naturally the idea of vectors having magnitude and orientation. However, they have some disadvantages, such as the singularity at the origin, where the polar angle is ill-defined: passing through this point adds a phase of $\pm\pi$ to the angular variable[2]. Moreover, these coordinates are nonlinear, i.e., if we represent points (or radius-vectors) as

[2]this fact has some non-trivial consequences in optics.

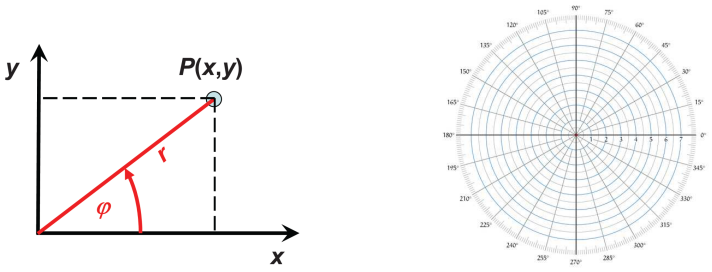

Figure 4.1: The usual introduction of the polar change (on the left) and the associated coordinate net (on the right).

$\mathbf{r}_i \to (\rho_i, \varphi_i)$, the addition law would look cumbersome and in particular

$$(\rho_1, \varphi_1) + (\rho_2, \varphi_2) \neq (\rho_1 + \rho_2,\ \varphi_1 + \varphi_2).$$

On the other hand, in many cases such representation appears much more convenient. For instance, we may use them to derive some angular relations for the basic operations in Euclidean space, already familiar from the consideration of complex numbers and in particular, the dot and wedge products.

4.3 Dot and Wedge Products

Modeling vectors in $\mathbb{R}^2$ with complex numbers provides a variety of useful constructions that may naturally be generalized to higher dimensions. For example, let us consider the modulus of a complex number

$$|z| = \sqrt{\bar{z}z} = \sqrt{x^2 + y^2}$$

which, according to the famous Pythagoras theorem, has the interpretation of length or distance. More generally, identifying radius-vectors $\mathbf{r}_i$ with points $z_i \in \mathbb{C} \cong \mathbb{R}^2$, we may define the *scalar* (or *dot*) *product* of $\mathbf{r}_{1,2}$ as

$$\mathbf{r}_1 \cdot \mathbf{r}_2 = (\mathbf{r}_1, \mathbf{r}_2) = \Re(\bar{z}_1 z_2) = x_1 x_2 + y_1 y_2. \tag{4.4}$$

Similarly, for vectors in three-dimensional space we have

$$\mathbf{r}_1 \cdot \mathbf{r}_2 = x_1 x_2 + y_1 y_2 + z_1 z_2.$$

On the other hand, complex numbers provide one more interesting type of multiplication, usually referred to as *exterior* (or *wedge*) product, namely

$$\mathbf{r}_1 \wedge \mathbf{r}_2 = z_1 \wedge z_2 = \Im(\bar{z}_1 z_2) = x_1 y_2 - x_2 y_1. \tag{4.5}$$

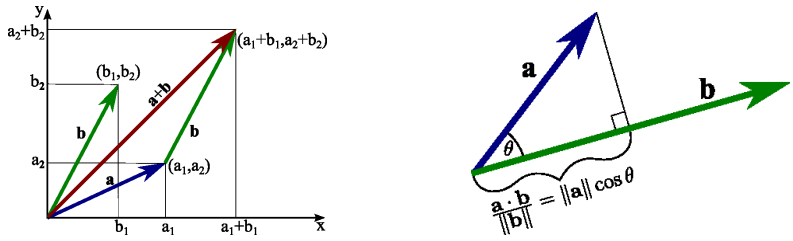

Figure 4.2: The vector sum (on the left) and and dot product (on the right).

Note that the former one is *symmetric*, i.e., $\mathbf{r}_1 \cdot \mathbf{r}_2 = \mathbf{r}_2 \cdot \mathbf{r}_1$, while the latter is *skew-symmetric*, meaning that $\mathbf{r}_1 \wedge \mathbf{r}_2 = -\mathbf{r}_2 \wedge \mathbf{r}_1$. Moreover, with the aid of the polar change (4.2) we easily see that (see Figure 4.2)

$$\mathbf{r}_1 \cdot \mathbf{r}_2 = |\mathbf{r}_1||\mathbf{r}_2| \cos \varphi_{12}$$

where the vector norm $|\mathbf{r}_i| = \rho_i$ is actually the radial component of the polar coordinates and $\varphi_{12} = \varphi_2 - \varphi_1$ is the relative angle between the two vectors. Similarly, using simple trigonometric identities, one may show that

$$\mathbf{r}_1 \wedge \mathbf{r}_2 = |\mathbf{r}_1||\mathbf{r}_2| \sin \varphi_{12}.$$

These two relations suggest a direct geometric interpretation of the dot and wedge products. Namely, the former measures the orthogonal projection of $\mathbf{r}_1$ onto the direction of $\mathbf{r}_2$, additionally multiplied with the magnitude of $\mathbf{r}_2$ (or vice versa). In particular, if both vectors are unit, i.e., $|\mathbf{r}_1| = |\mathbf{r}_2| = 1$, their dot product simply equals the cosine of the angle they spread. The wedge product, on the other hand, gives the *oriented area*[3] of the parallelogram determined by the two vectors. Moreover, these operations may be used also for obtaining the angle between two vectors, given in terms of their cartesian coordinates as

$$\varphi_{12} = \operatorname{atan}_2 \frac{\mathbf{r}_1 \wedge \mathbf{r}_2}{\mathbf{r}_1 \cdot \mathbf{r}_2} \tag{4.6}$$

as well as for criteria of orthogonality or parallelism between nonzero vectors

$$\mathbf{r}_1 \perp \mathbf{r}_2 \;\Leftrightarrow\; \mathbf{r}_1 \cdot \mathbf{r}_2 = 0, \qquad \mathbf{r}_1 \parallel \mathbf{r}_2 \;\Leftrightarrow\; \mathbf{r}_1 \wedge \mathbf{r}_2 = 0. \tag{4.7}$$

Note that all the above results hold also for free vectors or arbitrary attached vectors. The latter may be obtained via shifting a radius-vector with a

[3] i.e., with a sign depending on the orientation.

free translation vector $\tilde{\mathbf{r}} = \mathbf{r} + \mathbf{a}$, which is interpreted equivalently as a displacement of the coordinate system. For example, one obviously has

$$\sum_{i=1}^{N} \tilde{\mathbf{r}}_i = \sum_{j=1}^{N} \mathbf{r}_j + N\mathbf{a}$$

and summation demands all vectors to be attached to the same point in space. As for free vectors, no such restriction is necessary: each term can be moved freely in space so that its base point coincides with the end point of the previous one. Then, the *resultant* (i.e., the sum) is obtained by connecting the base of the first with the end of the last vector. In particular, if the vectors $\mathbf{r}_i$ form a closed polygon with consistent orientation of the sides, their sum is bound to be zero. Let us consider an example, in which the free vectors $\mathbf{v}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and $\mathbf{v}_2 = \begin{pmatrix} -1 \\ 0 \end{pmatrix}$ form a triangle with the third side given by their sum $\mathbf{v}_3 = \mathbf{v}_1 + \mathbf{v}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$. The task is to determine the angles and the lengths of the sides. The latter are simply given by the vector norms $|\mathbf{v}_1| = \sqrt{1^2 + 1^2} = \sqrt{2}$, $|\mathbf{v}_2| = |\mathbf{v}_3| = 1$. As for the angles, we do not need formula (4.6) in this case, since obviously $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$ and thus $\mathbf{v}_1 \perp \mathbf{v}_2$, which yields a right isosceles triangle with $\varphi_{31} = \dfrac{\pi}{2}$ and $\varphi_{12} = \varphi_{23} = \dfrac{\pi}{4}$.

4.4 Linear Dependence and Bases

Another issue that we barely discussed is the representation or, as they say, expansion of vectors in a given basis. To make this point more rigorous, we need the notion of *linear dependence*. A system of vectors $\{\mathbf{v}_i\}$ is called *linearly dependent* if there exists a vanishing linear combination of the type

$$\sum_{i=1}^{n} \alpha_i \mathbf{v}_i = 0$$

where at least one of the coefficients $\alpha_k \in \mathbb{R}$ (or $\mathbb{C}$) is non-zero. An equivalent definition demands one of the vectors to be expressed as a linear combination of the others. Otherwise the system is called *linearly independent*[4]. For example, a system of two vectors is linearly dependent if and only if they

[4] in short, a system $\{\mathbf{v}_k\}$ for which $\sum_{i=1}^{n} \alpha_i \mathbf{v}_i = 0$ implies $\alpha_k = 0 \; \forall \, k = 1 \ldots n$.

are proportional (parallel). Adding the *zero vector* (the neutral element of our vector space) to any system, obviously makes it linearly dependent. Note that dependent systems of more than two vectors might contain linearly independent subsystems. For example $\mathbf{v}_1$ and $\mathbf{v}_2$ might not be proportional, but their sum equals $\mathbf{v}_3$, as in the example considered above. The maximal number of linearly independent vectors within a given system $\{\mathbf{v}_i\}$ is referred to as its *rank*. In our example, the rank is 2 although we have three vectors in the system. Generally speaking, the rank equals the dimension of the space spanned by the vectors in the system and cannot exceed the dimension of the ambient space. Thus, in the plane, which may be described with only two independent directions 'forth-back' and 'left-right' used as coordinate axes, the maximal possible rank is two, i.e., each vector there can be expressed as a linear combination of two arbitrary non-parallel vectors. Similarly, in space, the additional 'up-down' degree of freedom appears, so one needs three independent directions to introduce coordinates. A system of n linearly independent vectors is referred to as a *basis* of its *linear span* (the space of its linear combinations). The choice of a basis is therefore equivalent to choosing a reference frame. Needless to say, there are infinitely many bases in each dimension and regarding some simple rules, we may go from one basis to another. One particular choice is the *standard basis* in the plane

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix},\ \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad \Rightarrow \quad \mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = v_1\,\mathbf{e}_1 + v_2\,\mathbf{e}_2, \quad v_i = (\mathbf{v}, \mathbf{e}_i).$$

Note that here the components in the expansion (parallel projections) are equal to the scalar products (orthogonal projections). This coincidence holds only for *orthonormal bases* consisting of mutually perpendicular unit vectors. Consider another such basis, namely $\mathbf{a}_1 = \dfrac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix},\ \mathbf{a}_2 = \dfrac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \end{pmatrix}$. In this basis we have different components of $\mathbf{v}$ given by $\tilde{v}_i = (\mathbf{v}, \mathbf{a}_i)$, which are easy to obtain explicitly in the form $\mathbf{v} = \dfrac{1}{\sqrt{2}}\begin{pmatrix} v_1 + v_2 \\ v_2 - v_2 \end{pmatrix}$. When the basis is arbitrary (neither normalized, nor orthogonal), one must proceed in a different way, preferably working directly with the relations between the two bases. For example, let $\mathbf{v} = \begin{pmatrix} 4 \\ 3 \end{pmatrix}$ in the standard basis and we are interested in obtaining its components in the basis $\mathbf{a}_1 = \begin{pmatrix} 2 \\ 1 \end{pmatrix},\ \mathbf{a}_2 = \begin{pmatrix} 5 \\ 3 \end{pmatrix}$. It is convenient to express first $\mathbf{a}_1 = 2\mathbf{e}_1 + \mathbf{e}_2$ and $\mathbf{a}_2 = 5\mathbf{e}_1 + 3\mathbf{e}_2$. Then, we want to obtain the combination $\mathbf{v} = 4\mathbf{e}_1 + 3\mathbf{e}_2$ from the above two, which are not difficult to invert into $\mathbf{e}_1 = 3\mathbf{a}_1 - \mathbf{a}_2$ and $\mathbf{e}_2 = 2\mathbf{a}_2 - 5\mathbf{a}_1$.

Substituting the latter into the above expression for $\mathbf{v}$, one finally obtains $\mathbf{v} = 4(3\mathbf{a}_1 - \mathbf{a}_2) + 3(2\mathbf{a}_2 - 5\mathbf{a}_1) = 2\mathbf{a}_2 - 3\mathbf{a}_1$, so in the new basis we have[5] $\mathbf{v} = \begin{pmatrix} -3 \\ 2 \end{pmatrix}$. One alternative approach is based on the notion of the so-called *dual bases*, but we leave this discussion for the three-dimensional case.

4.5 Proportions and the Geometric Center

Here we use another famous theorem in elementary geometry due to Thales, asserting that proportions are preserved under projection. One immediate consequence is that if a point $M(x_m; y_m)$ divides a line segment connecting $A(x_0; y_0)$ and $B(x_1; y_1)$ into a given proportion, say $\lambda : 1$, then the same relation holds for the x and y-components of the vectors $\vec{AB}$, $\vec{AM}$ and $\vec{MB}$:

$$\frac{|AM|}{|MB|} = \lambda \quad \Rightarrow \quad \frac{x_m - x_0}{x_1 - x_m} = \frac{y_m - y_0}{y_1 - y_m} = \lambda.$$

Ultimately, this yields for the coordinates of the point M the expressions

$$x_m = \frac{x_0 + \lambda x_1}{1 + \lambda}, \qquad y_m = \frac{y_0 + \lambda y_1}{1 + \lambda} \tag{4.8}$$

which in particular turns into an arithmetic mean if M is a midpoint ($\lambda = 1$). This is the usual way to construct medians in triangles: connecting the midpoint of a given side with the opposite vertex, say $C(x_2; y_2)$. Furthermore, as we know, the intersection point $G(x_g; y_g)$ of the three medians, which is also the geometric center of the triangle, divides each individual median in ratio $2 : 1$ from the vertex to the side, so formula (4.8) yields in this case

$$x_g = \frac{x_2 + 2x_m}{1 + 2}, \qquad y_g = \frac{y_2 + 2y_m}{1 + 2}$$

and expressing the components of M we obtain for the geometric center

$$x_g = \frac{x_1 + x_2 + x_3}{3}, \qquad y_g = \frac{y_1 + y_2 + y_3}{3}. \tag{4.9}$$

The latter generalizes for arbitrary polygons with N vertices as

$$\mathbf{r}_g = \frac{1}{N} \sum_{i=1}^{N} \mathbf{r}_i \tag{4.10}$$

[5] equality between vectors and matrices needs to be understood only in a fixed basis.

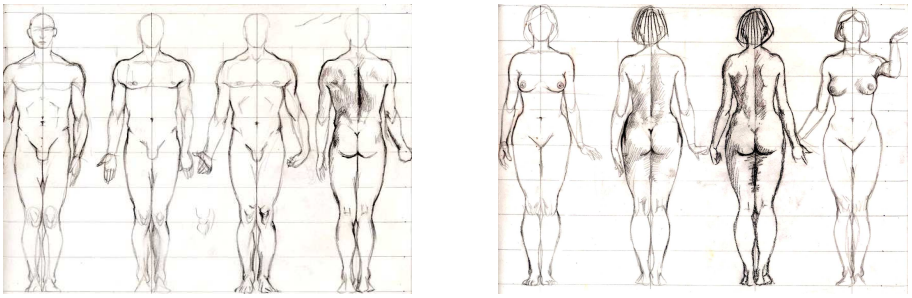

Figure 4.3: Proportions of the human body - male and female.

so the arithmetic mean property remains. This formula is also closely connected to the mechanical notion "center of masses" defined for a system of *material points*[6] described by radius vectors $\mathbf{r}_i$ and individual masses m_i as

$$\mathbf{r}_g = \frac{1}{M}\sum_{i=1}^{N} m_i \mathbf{r}_i, \qquad M = \sum_{i=1}^{N} m_i. \tag{4.11}$$

The coefficient λ in formula (4.8) may also be interpreted in this context as a ratio between the two masses that determines how much closer the center of the system is to the more massive object (this works well for binary stars).

How to bisect an angle?

We shall propose several ideas for constructing a vector, which bisects a given angle. The first one resorts on the geometric property of vector addition: since the vector sum is interpreted as a diagonal in the parallelogram formed by the two summands, it suffices to re-scale those summands in such a way that this parallelogram becomes a rhombus and has the property that its diagonals bisect the angles at the vertices they connect. Consider the following example. Let the angle we are interested in be determined by the radius-vectors of the two points $A(1;0)$ and $B(3;4)$. Then, since $|\mathbf{r}_A| = 1$ and $|\mathbf{r}_B| = 5$, it suffices to multiply the first vector by 5 and then the sum

$$\mathbf{r} = 5\mathbf{r}_A + \mathbf{r}_B = \begin{pmatrix} 5 \\ 0 \end{pmatrix} + \begin{pmatrix} 3 \\ 4 \end{pmatrix} = \begin{pmatrix} 8 \\ 4 \end{pmatrix} \sim \begin{pmatrix} 2 \\ 1 \end{pmatrix}$$

yields the desired vector (note that we may multiply by any convenient factor since only the direction is relevant). More generally, we have

$$\mathbf{r} = |\mathbf{r}_B|\,\mathbf{r}_A + |\mathbf{r}_A|\,\mathbf{r}_B. \tag{4.12}$$

[6] idealized point objects with physical properties, such as mass, electric charge etc.

Problem Session:

1. Is the affine space $\mathbb{R}^n$ also a vector space? Give rigorous arguments.
2. If the coordinate system $\{K\}$ travels with a constant speed v in the positive x-direction, what would be the coordinates of the point A_t at $t = 13$ if $A_0(3; 2)$ are the ones in the initial moment $t = 0$?
3. A TV antenna needs to be build so that it is the same distance to the towns A and B and twice closer to C. Assuming that the coordinates of A, B and C on the map form an acute triangle, can you write an exact expression for the antenna's location and is the solution unique?
4. A triangle has vertices $A(4; 0)$, $B(3; \sqrt{3})$ and the origin $O(0; 0)$.
   a) What are the angles and the area of this triangle?
   b) Calculate the distance form the geometric center of $\triangle ABO$ to the reflection of B with respect to the axis Oy.
   c) Find the vector, which bisects $\measuredangle AOB$.
5. Consider the polygon formed by the roots z_k of the equation $z^5+1 = 0$.
   a) Find the vectors bisecting each of the vertex angles. Would it have been easier or harder to consider $z^6+1=0$ instead?
   b) Find the center of masses, if the individual mass of z_k is defined by $m_k = \alpha^k$ where $\alpha > 0$ is a real parameter.
6. Given two points $A(1; 3)$ and $B(2; -1)$ find the coordinates of a point M, which divides the line segment AB into the "golden ratio", i.e.,
$$|AM| : |MB| = |MB| : |AB|.$$
7. Can you think of a way to trisect an angle ϑ between two vectors given with their components? Thy to reason your construction thoroughly and extend it to dividing the angle ϑ into any number of equal parts.
8. Which projections yield the coordinates v_i of the vector $\mathbf{v}$ in the expansion $\mathbf{v} = v_1\mathbf{a}_1 + v_2\mathbf{a}_2$ in $\{\mathbf{a}_k\}$, the parallel or the orthogonal ones?
9. If the vector $\mathbf{v}$ is represented in the standard basis $\{\mathbf{e}_k\}$ by the matrix $\begin{pmatrix} 3 \\ -1 \end{pmatrix}$, find its components in the basis $\{\mathbf{a}_k\}$ given by the vectors $\mathbf{a}_1 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ and $\mathbf{a}_2 = \begin{pmatrix} 2 \\ -1 \end{pmatrix}$ in $\{\mathbf{e}_k\}$. Hint: you may use that $\mathbf{a}_1 \perp \mathbf{a}_2$.

Chapter 5

Lines in the Plane

Let us recall that points in the plane may be represented with their cartesian coordinates, e.g. $A(x_0; y_0)$ and $B(x_1; y_1)$, or equivalently, with their radius-vectors $\mathbf{r}_0 = \begin{pmatrix} x_0 \\ y_0 \end{pmatrix}$ and $\mathbf{r}_1 = \begin{pmatrix} x_1 \\ y_1 \end{pmatrix}$. Since there is a one-to-one correspondence between these two descriptions, sometimes we write points, but use the vector formalism for convenience. Moreover, we denote $\vec{AB} = \mathbf{r}_1 - \mathbf{r}_0$ the free vector pointing from A to B and similarly $\vec{BA} = \mathbf{r}_0 - \mathbf{r}_1$ the one from B to A. Similarly, with three points A, B and C, we easily construct the *direction vectors* along the sides of the triangle $\triangle ABC$. This certainly works for any polygon, but how do we generalize to an infinite number of points? In the generic case we need some tools in the arsenal of calculus, but simpler shapes such as straight lines, ellipses, hyperbolae and parabolas are perfectly within the reach of algebra. Let us begin with the former case.

5.1 General and Parametric Equations of a Line

We may describe a straight line in the plane in several ways - one of them is to fix a point $A(x_0; y_0) \in \mathbb{R}^2$ and a direction vector $\mathbf{t}$. Then, we may introduce a real parameter $\lambda \in \mathbb{R}$ the obtain the (vector) *parametric equation* of the line g passing though the point A and oriented along $\mathbf{t}$ as

$$\mathbf{r} = \mathbf{r}_0 + \lambda \mathbf{t}, \qquad \lambda \in \mathbb{R} \tag{5.1}$$

where $\mathbf{r} = \begin{pmatrix} x \\ y \end{pmatrix}$ represents the radius-vector of a *generic point* on the line, i.e., all points on g are obtained in this way for different values of the parameter λ. One easy way to explain this in geometric terms is that the

vector difference of any pair of radius-vectors of points on the line is always proportional to the direction vector $\mathbf{t}$ (with coefficient of proportionality λ)

$$\mathbf{r} - \mathbf{r}_0 = \lambda \mathbf{t}.$$

Writing (5.1) in components we obtain the scalar parametric equations of g:

$$\begin{aligned} x &= x_0 + \lambda t_1 \\ y &= y_0 + \lambda t_2. \end{aligned} \tag{5.2}$$

Furthermore, one may isolate the parameter λ from the two equations and thus obtain an alternative representation of the form

$$\frac{x - x_0}{t_1} = \frac{y - y_0}{t_2} \tag{5.3}$$

which no longer contains a free parameter, but rather, imposes one linear constraint on the two variables x and y. Suppose that instead of the direction vector $\mathbf{t}$ we are given a pair of points $A(x_0; y_0)$ and $B(x_1; y_1)$ contained in the line. Then, the direction vector may be obtained as the vector difference $\mathbf{r}_1 - \mathbf{r}_0$, as already explained. Hence, the above equation may be written as

$$\frac{x - x_0}{x_1 - x_0} = \frac{y - y_0}{y_1 - y_0} \tag{5.4}$$

which is broadly known as the *equation of a line through a pair of points*. Another possibility is to be given a point $A(x_0; y_0) \in g$ and a *normal vector* $\mathbf{n} \perp g$. Then, instead of (5.1) one has $\mathbf{n} \cdot (\mathbf{r} - \mathbf{r}_0) = 0$, or in components

$$n_1(x - x_0) + n_2(y - y_0) = 0 \tag{5.5}$$

which may be referred to as the *general equation of a line* (see Figure 5.1). Usually in literature one finds under this name the following equation

$$ax + by + c = 0 \tag{5.6}$$

where the constants a, b and c now have perfectly transparent geometric interpretation - the first two, $a = n_1$ and $b = n_2$, are just the components of the normal vector, while the third one is given by the scalar product

$$c = -\mathbf{n} \cdot \mathbf{r}_0 = -n_1 x_0 - n_2 y_0.$$

This interpretation allows for directly identifying the normal direction form the coefficients in front of x and y in the generic equation, just as the parametric equation reveals the direction vector from the coefficients in front of

the parameter λ in formula (5.1). These three situations appear in many practical problems in geometry, physics and engineering. For example, the bisectrix has a realization as a line passing through a point and having a prescribed direction. Similarly, the median is most naturally defined as a line through a pair of points, while the height is usually determined by an incident point and a normal direction. Note that the direction and normal vectors are in a natural one-to-one correspondence (only in dimension two!). Namely, let $\mathbf{t} = \begin{pmatrix} t_1 \\ t_2 \end{pmatrix}$ be the direction vector of the line g, then its normal may be written as $\mathbf{n} = \begin{pmatrix} t_2 \\ -t_1 \end{pmatrix}$. Given a vector in $\mathbb{R}^2$, one can always obtain its normal direction by swapping its components and changing one of the signs. The proof relies on the fact that the vanishing dot product is necessary and sufficient (so it may be considered as a definition) for orthogonality.

Incidence of Points and Lines

Two distinct points in the plane $A \neq B$, or in higher-dimensional Euclidean space, determine a free vector $\vec{AB} = \mathbf{r}_1 - \mathbf{r}_0$ and by formula (5.4), a unique line $g \ni A, B$ containing them. The (shortest) distance between A and B is measured along the line g and given by the famous Pythagoras' theorem as

$$\mathrm{d}(A, B) = |\mathbf{r}_1 - \mathbf{r}_0| = \sqrt{(x_1 - x_0)^2 + (y_1 - y_0)^2}. \tag{5.7}$$

Two lines may either coincide $g_1 \equiv g_2 \Leftrightarrow g_1 \cap g_2 = g_1$ or differ $g_1 \neq g_2$ and in the latter case there are two options: they either intersect at a point $\tilde{A}(\tilde{x}; \tilde{y}) = g_1 \cap g_2$, or are parallel, i.e., with no common points $g_1 \parallel g_2 \Leftrightarrow g_1 \cap g_2 = \emptyset$. To make a clear distinction between these cases, we consider some examples. Let $g_i : \ a_i x + b_i y + c_i = 0$, $i = 1, 2$ be the general equations of the two lines. Then, if all the coefficients are proportional, the lines obviously coincide (equivalent equations determine equivalent loci). If, on the other hand, $a_1 = \lambda a_2$ and $b_1 = \lambda b_2$, but $c_1 \neq \lambda c_2$ for some $\lambda \in \mathbb{R}/\{0\}$, then the system is inconsistent and the lines are parallel[1]. Finally, if the normal vectors are not parallel, the lines intersect as the system has a unique solution - point $C = g_1 \cap g_2$. Formulate the above criteria for all three cases in terms of the parametric equation! Let us now discuss the *incidence relations* involving both points and lines. To start with, a point is said to *belong* to line $A \in g$ if its coordinates satisfy the equation of g. In this case, we also say that g *contains* or *passes through* A. Otherwise, according to

[1] the angle between two lines is equal to the angle between their normal vectors.

Euclid's axioms of geometry, there exists a unique line h, which contains A and is parallel to g. Obtaining its equation is actually quite easy: we only need the normal vector of g and the coordinates of A in formula (5.5). Similarly, one may build the unique normal from A to g by using $\mathbf{n}^{\perp}$ instead. Now, let us consider a second technique for constructing a bisectrix that is a bit less direct, but still finds applications in practice. Namely, one may choose to make use of the well-known property that the intersection point L of a bisectrix from a vertex C with the opposite side AB divides the latter into the same proportion that relates the lengths of the two adjacent sides, i.e., the coordinates of L are given by formula (4.8) with $\lambda = \dfrac{|CA|}{|CB|}\cdot$. Consider an example, in which $A(3;4)$, $B(2;1)$ and $C(-1;1)$ are the three vertices, so we easily obtain the direction vectors along the sides of $\triangle ABC$

$$\vec{c} = A\vec{B} = -\begin{pmatrix} 1 \\ 3 \end{pmatrix}, \qquad \vec{a} = B\vec{C} = -\begin{pmatrix} 3 \\ 0 \end{pmatrix}, \qquad \vec{b} = A\vec{C} = -\begin{pmatrix} 4 \\ 3 \end{pmatrix}$$

as well as their lengths $|AB| = \sqrt{10}$, $|BC| = 3$ and $|AC| = 5$, thus $\lambda = 5/3\cdot$ Then, with the aid of formula (4.8) we obtain the coordinates of $L\left(\frac{19}{8}; \frac{17}{8}\right)$ and write the equation of the line ℓ through C and L according to (5.4) as

$$\frac{x+1}{\frac{19}{8}+1} = \frac{y-1}{\frac{17}{8}-1} \quad \Rightarrow \quad \ell: \; x - 3y + 4 = 0.$$

Let us see whether we would get the same result using the method based on re-scaling and vector sum (property of the diagonals in the rhombus). Firstly, we need to make $C\vec{A}$ and $C\vec{B}$ equal in magnitude[2] so that they form a rhombus, whose diagonal (the vector sum of the re-scaled sides) determines the direction of the bisectrix. Thus, in the sum we multiply each of the vectors with the length of the other, which yields

$$\vec{\ell} \sim \; |CA|\, C\vec{B} + |CB|\, C\vec{A} = 5\begin{pmatrix} 3 \\ 0 \end{pmatrix} + 3\begin{pmatrix} 4 \\ 3 \end{pmatrix} = 9\begin{pmatrix} 3 \\ 1 \end{pmatrix}\cdot$$

Then, we may easily construct the general equation of ℓ having the normal direction $\mathbf{n} = \begin{pmatrix} 1 \\ -3 \end{pmatrix} \perp \vec{\ell}$ that gives the coefficients in front of x and y, and the coordinates of the point $C(-1;1)$, whose radius-vector's dot product with $\mathbf{n}$ yields minus the free term. Thus, we finally obtain $\ell: \; x - 3y + 4 = 0$ that is identical to the above result.

[2] as they are attached to a common vertex C we may use directly formula (4.12).

5.2 Normal Equations and Distances

As we may see from equations (5.1) and (5.5), both the directional and the normal vectors are determined only up to a non-zero factor (coefficient of proportionality). One may factor out this freedom (at least partially) via normalization. In particular, if we denote $\mathbf{n}^\circ$ the unit normal in the $\mathbf{n}$ direction, there are only two possibilities $\mathbf{n}^\circ_\pm = \pm\dfrac{\mathbf{n}}{|\mathbf{n}|}$, which differ only in their orientation. Now, let us use this normalized vector in (5.5) and write

$$g: \quad (\mathbf{r}-\mathbf{r}_0)\cdot\mathbf{n}^\circ_\pm = 0 \quad \Rightarrow \quad \pm\frac{ax+by+c}{\sqrt{a^2+b^2}} = 0 \tag{5.8}$$

which is referred to as the *normal equation* of the line g. It seems that the effort of multiplying an equation by a non-zero factor must go in vain, but actually, it turns out to be rather useful as will become apparent shortly. Consider the following application: finding the shortest distance $\mathrm{d}(B,g)$ between a point $B(x_1;y_1)$ and a line $g:\ (\mathbf{r}-\mathbf{r}_0)\cdot\mathbf{n}=0$. Firstly, it is clear that the shortest of all possible distances $\mathbf{r}_1-\mathbf{r}$, written for a generic point with radius-vector $\mathbf{r}$ ($\mathbf{r}_1$ being the radius-vector of B), is the one along the normal direction $\mathbf{n}$. Equivalently, one may project each of these vectors, and in particular $\mathbf{r}_1-\mathbf{r}_0$, since all projections are equal and as we already know, given by a dot product with the unit normal in the so chosen direction, which is exactly $\mathbf{n}^\circ$. On the other hand, this unit normal has two possible directions denoted by $\mathbf{n}^\circ_\pm$. One of them makes an acute angle with $\mathbf{r}_1-\mathbf{r}_0$ and the other - obtuse. Thus, in order to guarantee non-negative distance, we may simply take the absolute value of this dot product, finally obtaining

$$\mathrm{d}(B,g) = |(\mathbf{r}_1-\mathbf{r}_0)\cdot\mathbf{n}^\circ| \tag{5.9}$$

so the distance between the point and the line is obtained (up to a sign) by substituting the coordinates of the point into the normal equation of the line. One standard application is to obtain the equation of the bisectrix of an angle determined by the intersection of two lines. Let $g_i:\ a_ix+b_iy+c_i=0$, $i=1,2$ be the general equations of the two lines. Then, the bisectrix may be defined also as the *locus* of points at equal distance from the two lines, as we know for elementary geometry. But this condition can be written as

$$\frac{a_1x+b_1y+c}{\sqrt{a_1^2+b_1^2}} = \pm\frac{a_2x+b_2y+c}{\sqrt{a_2^2+b_2^2}} \tag{5.10}$$

where the $\pm$ sing is due to the modulus in equation (5.9) and comes to show that there are actually two bisectrices, which makes perfect sense since there are in fact two complementary angles determined by the intersection $g_1\cap g_2$.

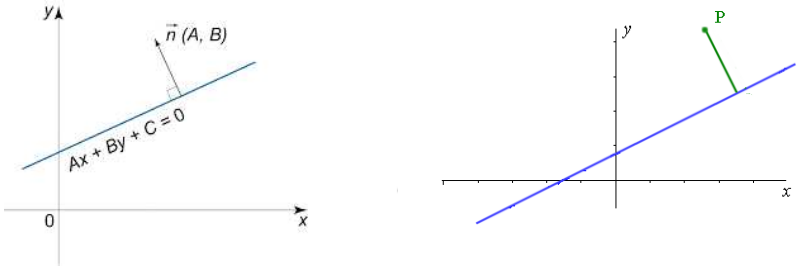

Figure 5.1: General equation of a line (on the left) and distance from a point P to a line (on the right).

5.3 The Slope-Intercept Equation

There are several more representations of straight lines in $\mathbb{R}^2$. For example, one may use the so-called *intercept-intercept equation*

$$\frac{x}{a'} + \frac{y}{b'} = 1 \tag{5.11}$$

where the constants a' and b' are interpreted as points at which g intersects the x and y-axes, respectively. This is easy to see by setting respectively $x = 0$ and $y = 0$ in (5.11). Another useful representation is given by the *slope-intercept equation*

$$y = kx + n \tag{5.12}$$

where the *slope* k can be expressed in terms of the angle α between g and the x-axis as $k = \tan\alpha$, while n is the intersection point with the latter. Note that the slope k is ill-defined for vertical lines so we have $\alpha \neq \pm\frac{\pi}{2}\cdot$ In all other cases it is convenient to introduce the so-called *relative slopes* k_{ij}. Namely, let us consider a pair of (non-vertical) lines $g_i : k_i x + n_i$, $i = 1, 2$ and denote the angle between them by $\alpha_{ij} = \alpha_j - \alpha_i$. Then, due to basic trigonometric identities, for the relative slope $k_{ij} = \tan\alpha_{ij}$ we have[3]

$$k_{ij} = \frac{k_j - k_i}{1 + k_i k_j}\cdot \tag{5.13}$$

This formula has many applications, e.g., it may be used it for an alternative construction of the bisectrix. Let g_{12} be the two intersecting lines and g_3 denote the bisectrix of the angle α_{12} between them. Then, by definition we

[3]note that $k_{ij} = -k_{ji}$, so orientation here is crucial.

have $\alpha_{13} = \alpha_{32}$, which yields for the relative slopes $k_{13} = k_{32}$ and via formula (5.13) we easily obtain the quadratic equation for the unknown slope $\kappa = k_3$

$$(k_1 + k_2)\kappa^2 + 2(1 - k_1k_2)\kappa - (k_1 + k_2) = 0. \tag{5.14}$$

The obvious solution

$$\kappa^{\pm} = \frac{k_1k_2 - 1 \pm \sqrt{(1 + k_1^2)(1 + k_2^2)}}{k_1 + k_2} \tag{5.15}$$

yields

$$\alpha^{\pm} = \arctan\left[-\cot(\alpha_1 + \alpha_2) \pm \csc(\alpha_1 + \alpha_2)\right]$$

so ultimately, we may write for the unknown angle $\alpha = \alpha_3$

$$\alpha^{+} = \frac{\alpha_1 + \alpha_2}{2}, \qquad \alpha^{-} = \frac{\pi + \alpha_1 + \alpha_2}{2}\cdot \tag{5.16}$$

Note that although this time we chose only one of the two complementary angles determined by the intersection of g_1 and g_2, the solutions to the quadratic equation automatically generate both bisectrices. It is not hard to see that α_3^+ corresponds to the so chosen angle, and α_3^- to its complement. Furthermore, the coefficients $n_3^{\pm}$ are easy to obtain from the necessity that g_1, g_2 and g_3 intersect at one and the same point $\tilde{A}(\tilde{x}, \tilde{y})$, namely as

$$n^{\pm} = \tilde{y} - \kappa^{\pm}\tilde{x}.$$

We conclude this section with a brief comment on the equation of a line through a pair of points $A(x_0; y_0)$ and $B(x_1; y_1)$. First of all, let us note that the latter may be obtained as a vanishing condition of the determinant

$$g: \quad \begin{vmatrix} x - x_0 & y - y_0 \\ x_1 - x_0 & y_1 - y_0 \end{vmatrix} = 0. \tag{5.17}$$

On the other hand, the above determinant can be written also in the form of a wedge product $(\mathbf{r} - \mathbf{r}_0) \wedge (\mathbf{r}_1 - \mathbf{r}_0) = 0$, which is equal to the area of the parallelogram spanned by the vectors $\mathbf{r} - \mathbf{r}_0$ and $\mathbf{r} - \mathbf{r}_0$, i.e., twice that of the triangle ΔABC, with $C(x; y)$. Therefore, the equation of a line can be derived from the vanishing of certain area element. This idea has a straightforward generalization to the three dimensional case as shown below.

Problem Session

1. In the right isosceles triangle ABC we are given the coordinates of the vertex $C(4;-1)$ and the equation of the hypothenuse: $3x-y+5=0$. Find the coordinates of the other two vertices.

2. In ΔABC we know the coordinates of two of the vertices $A(-4;3)$, $B(4;-1)$ and the intersections point of the heights $h_a \frown h_b = H(3;3)$. Find the vertex C, the median m_c from it and the distance $d(H, m_c)$.

3. With the point $A(3;0)$ and the line $g : y = x + 12$ build an isosceles triangle ΔABC with base $BC \in g$ for which the vertex C also lies on the y-axis. Write the equation of the circumscribed circle.

4. With the points $A(3;0), B(0;5)$ and the line $g : 2x+y = 0$ find a point $C \in g$ such that ΔABC is a right triangle with the right angle at the vertex B. Write the equation of the circumscribed circle of ΔAOB, where $O(0;0)$ is the origin.

Chapter 6

Matrices and Determinants

In this chapter we introduce the basic tool of linear algebra, matrices (as higher dimensional vectors, which allow for addition and multiplication) and thus, gradually build their algebra. The non-commutative matrix algebra has many applications in geometry, some of which we discuss below. Apart from that, it is a crucial tool in both fundamental and engineering disciplines, such as quantum mechanics, probability theory, game theory, optimal control, numerical methods and programming.

6.1 The Matrix Algebra

A rectangular table of elements of the type

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & \dots & & a_{2n} \\ \dots & & a_{ij} & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} \tag{6.1}$$

ordered in m *rows* and n *columns* is called a *matrix* with *dimensions* $m \times n$. As can be seen above, the first index denotes the row and the second one - the column, so the *matrix entry* a_{ij} fills the cell at the i-th row and j-th column. These entries may be numbers, functions or any kind of retrievable data. In most of our considerations, however, we assume that $a_{ij} \in \mathbb{C}$.

Instead of (6.1), we make use of the brief notation

$$A_{m \times n} \quad \leftrightarrow \quad \{a_{ij}\}$$

and show that matrices of the same dimension constitute a vector space with

1. Given two matrices $A_{m\times n} \leftrightarrow \{a_{ij}\}$ and $B_{m\times n} \leftrightarrow \{b_{ij}\}$, their sum is a matrix $C_{m\times n} \leftrightarrow \{c_{ij}\}$ of the same dimension with entries given by

$$c_{ij} = a_{ij} + b_{ij}$$

2. Multiplication with a scalar $\lambda \in \mathbb{C}$ also yields a matrix with entries

$$\lambda A_{m\times n} \leftrightarrow \ \{\lambda a_{ij}\}.$$

Note that this is a straightforward generalization of the linear laws for vectors introduced above. In this context, it seems reasonable to view matrices as ordered systems of vector-rows or vector-columns, respectively. Such understanding justifies also the way we define multiplication of two matrices. Among the many possibilities to do that, people have chosen to work with the so-called *row-by-column* rule. It consist of the following: the first matrix factor A is considered as a system of vector-rows and the second one B, as a system of vector-columns. Then, the entries of the product $C = AB$ are given by all the possible dot products between the vectors in the first system with those of the second one. The ordering of the rows is given by the fist factor and that of columns - from the second one, namely

$$A_{m\times n}B_{n\times k} = C_{m\times k} \ \leftrightarrow \ \{c_{ij}\}, \qquad c_{ij} = \sum_{l=1}^{n} a_{il}b_{lj}. \tag{6.2}$$

Note that these dot products demands that vectors from the two systems have equal dimensions, i.e., the number of columns of the first matrix to be equal to the number of rows of the second one. Then, the common middle index is factored out in the summation process and the dimension of the product is determined by what is left. Consider the following

Example: Let $A = \begin{pmatrix} 4 & 5 \\ 0 & 7 \end{pmatrix}$ and $B = \begin{pmatrix} 3 & 2 & 1 \\ 2 & 1 & 3 \end{pmatrix}$. Then, according to formula (6.2), the matrix product AB is given by

$$\begin{pmatrix} 4 & 5 \\ 0 & 7 \end{pmatrix}\begin{pmatrix} 3 & 2 & 1 \\ 2 & 1 & 3 \end{pmatrix} = \begin{pmatrix} 4\times3+5\times2 & 4\times2+5\times1 & 4\times1+5\times3 \\ 0\times3+7\times2 & 0\times2+7\times1 & 0\times1+7\times3 \end{pmatrix} = \begin{pmatrix} 22 & 13 & 19 \\ 14 & 7 & 21 \end{pmatrix}.$$

If we attempt to multiply these matrices in the reverse order, however, the corresponding dimensions would not agree, so unlike AB, the expression BA

does not make any sense. This property of the matrix algebra, namely that $AB \neq BA$, known as non-commutativity[1] has no analogue in the familiar number algebras. Note that the algebra of matrices thus constructed satisfies

- the commutative law of summation $A + B = B + A$
- the associative law of both summation and multiplication

$$A + (B + C) = (A + B) + C, \qquad A(BC) = (AB)C$$

- the distributive laws $A(B+C) = AB+AC$ and $(A+B)C = AC+BC$.

However, it fails to satisfy the commutative law of multiplication - in the generic case we have $AB \neq BA$ and as the above example clearly indicates, the two products are not necessarily simultaneously well-defined.

Special Types of Matrices

The whole algebra of matrices is something vast and it is convenient to consider only certain *subalgebras*, i.e., subspaces that are closed under summation and multiplication. Note that the two operations impose different restrictions on the matrix dimensions, which can be satisfied simultaneously only if the number of rows and columns are equal, i.e., if we consider matrices of the type $A_{n\times n}$, referred to as *square matrices*. Clearly, both the sums and products of square matrices of the same dimension are well-defined. Moreover, one may consider powers of the same matrix in this class, such as $A^2 = AA$, $A^3 = AAA$ and so on. Then, since AB and BA both exist, one may also define the binary operation

$$[A, B] = AB - BA \tag{6.3}$$

known as the *commutator* of two square matrices, which measures how far from commuting the latter are. This peculiar type of skew-symmetric product defines a different algebraic structure over the linear space of matrices, known as *Lie algebra*, which lies in the foundation of modern theoretical physics. There are two more subalgebras imbedded in the algebra of square matrices, namely the *upper-triangular* and *lower-triangular* matrices, defined as $A_{n\times n} \leftrightarrow \{a_{ij}\}$, $a_{ij} \equiv 0$ for $i > j$ and $B_{n\times n} \leftrightarrow \{b_{ij}\}$ with $b_{ij} \equiv 0$ for $i < j$, respectively. In other words, these are matrices with only zero entries respectively below or above the *main diagonal*, i.e., the set $\{a_{ii}\}$. The

[1] in particular, the existence of AB does not imply the existence of BA and vice versa.

intersection of these two subalgebras again determines a subalgebra - consisting of all *diagonal matrices*, for which $a_{ij} \equiv 0$ for $i \neq j$ (prove that these are all indeed subalgebras!). A specific case to consider is the class of the so-called *scalar matrices*, which are diagonal with all diagonal entries being equal $a_{ii} = \lambda \in \mathbb{C}$. They also form a subalgebra, which is commutative: this is actually the algebra of the complex numbers in a matrix representation. Moreover, they commute with all other square matrices of the same dimension, since they act as a multiplication with the *scalar* factor λ. One particular choice is $\lambda = 0$, which yields the *zero* (or *null*) *matrix* 0 with the property $A+0 = A$ for arbitrary matrix A of the same dimension. Similarly, for $\lambda = 1$ we obtain the *identity matrix* I, for which $AI = IA = A$ etc.

Transposition and Conjugation

We may introduce additional operations in the algebra of matrices. One that is particularly useful is the so-called *transposition* $^t : A \to A^t$, which turns rows to columns and vice versa. The *transposed matrix* has entries $a^t_{ij} = a_{ji}$. Here are some properties of matrix transposition:

- It is linear, meaning that

$$(A + B)^t = A^t + B^t, \qquad (\lambda A)^t = \lambda A^t \tag{6.4}$$

- It is *idempotent*, i.e.,

$$(A^t)^t = A \tag{6.5}$$

- It reverses the order of multiplication

$$(AB)^t = B^t A^t \tag{6.6}$$

The above make transposition a linear *anti-involution* of the matrix algebra. In our previous example we easily obtain $A^t = \begin{pmatrix} 4 & 0 \\ 5 & 7 \end{pmatrix}$ and $B^t = \begin{pmatrix} 3 & 2 \\ 2 & 1 \\ 1 & 3 \end{pmatrix}$, which allows for calculating also $B^t A^t = \begin{pmatrix} 3 & 2 \\ 2 & 1 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} 4 & 0 \\ 5 & 7 \end{pmatrix} = \begin{pmatrix} 22 & 14 \\ 13 & 7 \\ 19 & 21 \end{pmatrix}$ and comparing it to the previously obtained result for AB, we notice (6.6) holds, i.e., the matrix transposition changes the order of multiplication, which may be derived directly from (6.2) and holds for an arbitrary number of factors.

Another operation with similar properties is the so-called *hermitian conjugation* $^\dagger : A \to A^\dagger = \bar{A}^t$, which consists of transposition and complex conjugation. Clearly, if all entries are real $A^\dagger = A^t$. Such a generalization is useful for the definition of real scalar product for complex-valued vectors. This construction, known as *hermitian form*, is given as

$$\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u}^\dagger \mathbf{v}, \qquad \mathbf{u}, \mathbf{v} \in \mathbb{C}^n. \tag{6.7}$$

Note that we defined the dot and wedge products in $\mathbb{R}^2$ as real and imaginary parts of the hermitian form in $\mathbb{C}^1$. Can you think of an example in $\mathbb{C}^2$?

In the case of square matrices, transposition acts as a reflection about the main diagonal. Naturally, there are *symmetric matrices*, which are preserved by this reflection, i.e., have the property $A^t = A$, although they do not constitute a subalgebra under summation and multiplication[2]. Similarly, in the complex case we consider the class of *hermitian matrices* with the property $A^\dagger = A$. One may, on the other hand, define the class of *skew-symmetric matrices* satisfying the property $A^t = -A$, i.e., $a_{ij} = -a_{ji}$, which means that all diagonal elements must vanish $a_{ii} \equiv 0$. On the other hand, each *skew-hermitian* matrix, defined by the property $A^\dagger = -A$ can be decomposed into a sum of a real skew-symmetric matrix and a purely imaginary symmetric one (prove this!). Moreover, we have the following

Theorem 5 *Every square matrix A has the representation $A = A_0 + A_1$, where $A_0^\dagger = A_0$ and $A_1^\dagger = -A_1$.*

Proof. Define the hermitian part as $A_0 = \dfrac{1}{2}(A + A^\dagger)$ and and the skew-hermitian as $A_0 = \dfrac{1}{2}(A + A^\dagger)$, respectively. Restriction to the real case gives

Corollary 2 *Every square matrix may be decomposed into a sum of a symmetric and a skew-symmetric component: $A = A_0 + A_1$, $A_k^t = (-1)^k A_k$.*

Consider the example of 2×2 skew-hermitian matrices, spanned by the basis

$$\mathbf{i} = \begin{pmatrix} \mathrm{i} & 0 \\ 0 & -\mathrm{i} \end{pmatrix}, \qquad \mathbf{j} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \qquad \mathbf{k} = \begin{pmatrix} 0 & \mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix} \tag{6.8}$$

and show that $\mathbf{ij} = \mathbf{k}$, $\mathbf{jk} = \mathbf{i}$ and $\mathbf{ki} = \mathbf{j}$. Besides, these matrices anti-commute, e.g. $\mathbf{ij} = -\mathbf{ji}$, and they are all square roots minus one, i.e.,

[2] however, they do if the latter is replaced by the *anti-commutator* $\{A, B\} = AB + BA$.

$\mathbf{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 = -I$, where $I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ stands for the 2×2 identity matrix. In this sense, the above basis, extended with the identity matrix I allows for a natural generalization of complex numbers, known as *quaternions*. More precisely, a quaternion is a "number" with four components

$$q = q_0 I + q_1 \mathbf{i} + q_2 \mathbf{j} + q_3 \mathbf{k}, \qquad q_\mu \in \mathbb{R}$$

where q_0 is called a *real part* of q, while the imaginary one is the whole vector $\mathbf{q} = (q_1, q_2, q_3)^t$, which changes its sign under (hermitian) conjugation

$$q = (q_0, \mathbf{q}) \to \bar{q} = (q_0, -\mathbf{q}).$$

In particular, the algebra of complex numbers can be realized as real 2×2 matrices of the type $Z = xI + y\,\mathbf{j}$ (the multiplication rule is the same). Quaternions, usually denoted by $\mathbb{H}$, can be obtained from $\mathbb{C}$ in the same way $\mathbb{C} = \mathbb{R} \oplus \mathbf{j}\,\mathbb{R}$ is obtained as an extension of $\mathbb{R}$, namely $\mathbb{H} = \mathbb{C} \oplus \mathbf{j}\,\mathbb{C}$. They are associative but non-commutative and thus, constitute a a more general object, which is not a field (the largest number field is $\mathbb{C}$). However, they are commonly used in the context of three dimensional rotations and their complex generalization, the *bi-quaternions* (corresponding to $q_\mu \in \mathbb{R}$), in special relativity, electrodynamics and various other fields. We also have *dual quaternions*, whose coefficients are dual numbers, or *split quaternions*, for which some matrices in the above basis become hermitian. All those objects find many applications in various areas of science and technology: from classical and quantum mechanics to virtual reality and computer vision. They also present simple examples of Clifford geometric algebras, which have proven to provide a more efficient description of the physical reality compared to the classical approach.

Note also that matrices find application in practically any field which deals with data (that is any field, which allows for measurement). For example, we may use them in modeling stochastic processes with conditional probabilities: a_{ij} denotes the probability that the event i takes place under the condition j. Matrix multiplication then yields the correct overall probabilities for compound processes, e.g., repetitive experiments. Similarly, a_{ij} may be interpreted as transition coefficients and the entries of the squared matrix - as the corresponding coefficients for two-step transitions, A^3 yields the three-step transitions and so on. Then, it is no wonder that matrices are being constantly used in all kinds of simulations and optimization problems.

Problem Session

1. Obtain an expression for A^n for the square matrix

   a) $A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$, $\qquad A = \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix}$

   b) $A = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}$, $\qquad A = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$

2. Calculate all the commutators of the matrices (6.8) and show the are all square roots of the identity matrix.

3. Find all matrices that commute with $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$.

4. Show that each hermitian matrix may be decomposed into a sum of a real symmetric matrix and a purely imaginary skew-symmetric one.

5. Is it true that for each square matrix A, the products AA^t and A^tA are symmetric? Similarly, are $AA^\dagger$ and $A^\dagger A$ always hermitian? Can you construct skew-symmetric/skew-hermitian matrices in this way?

6. Show that each *strictly upper(lower)-triangular matrix*[3] is nilpotent, i.e., $A^m = 0$ for some $m \in \mathbb{N}$. Can you find the value of the order m?

7. Prove that for every matrix A the product $\langle A, A\rangle = A^tA$ is a well-defined square and symmetric matrix with the property

   $$\langle A+B, A+B\rangle = \langle A, A\rangle + \langle A, B\rangle + \langle B, A\rangle + \langle B, B\rangle.$$

   What is the geometric interpretation of its coefficients?

8. Show the transitivity of the vanishing commutator, i.e., $[A, B]=0,\ [B, C]=0 \ \Rightarrow\ [A, C] = 0$ and prove the Jacobi identity

   $$[A, [B, C]] + [B, [C, A]] + [C, [A, B]] = 0.$$

[3] i.e., such that $a_{ij} \equiv 0$ for $i \geq j$ and $i \leq j$, respectively.

6.2 Determinants

To each square matrix we may associate a homogenous polynomial of its entries called a *determinant*, which turns out to be quite useful both in algebraic considerations, e.g. for the solutions of systems of linear equations, as well as in geometry - for the calculation of volumes, areas and distances.

Determinants of 2×2 and 3×3 Matrices

We define the determinant of a 2×2 matrix to be

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc. \tag{6.9}$$

When we introduced vectors in the plane, we also defined the wedge product and showed that it is equal to the oriented area of the parallelogram, spanned by $\mathbf{u}$ and $\mathbf{v}$. This area may now be given by a determinant

$$\mathbf{u} \wedge \mathbf{v} = \begin{vmatrix} u_1 & v_1 \\ u_2 & v_2 \end{vmatrix}$$

which is generalized to arbitrary dimension n with the so-called *volume form*, given by an $n \times n$ determinant, which is an anti-symmetrized sum of $n!$ terms (all possible permutations), each consisting of n factors (one from each row and column). Moreover, this construction shares some common properties visible form the two-dimensional case:

- $\det A = \det A^t$, so properties for rows refer to columns and vice versa.
- $\det AB = \det A \det B$, but note that $\det(A + B) \neq \det A + \det B$.
- $\det(\lambda A) = \lambda^n \det A$ for $\lambda \in \mathbb{C}$ and in particular, a common factor can be taken out from each row/column in front of the determinant sign.
- Switching two successive rows/columns changes only the sign of $\det A$. Thus, moving the i-th row to the j-th position yields a factor $(-1)^{i-j}$.
- A determinant with a zero row/column, or more generally, with linearly dependent rows/columns, is always zero.
- If A is (upper or lower) triangular, then one has $\det A = a_{11} a_{22} \dots a_{nn}$.
- Similar matrices have equal determinants: $A \sim A' \Rightarrow \det A = \det A'$.

The notion of similarity here should be understood in a very specific manner. Namely, it demands that each vector-row $\mathbf{a}'_k$ of A' can be obtained if one adds to the corresponding vector-row $\mathbf{a}_k$ of A some linear combination of the remaining vector-rows $\mathbf{a}_j$ $(j \neq k)$. We use this idea below to transform higher-dimensional matrices into triangular form an thus, express their determinants simply as products of the corresponding diagonal entries.

Following a similar approach in the 3×3 case, we obtain the expression

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} \\ - a_{13}a_{22}a_{31} - a_{12}a_{21}a_{33} - a_{11}a_{23}a_{32}.$$

Note that the first indices in the three factors of each term are ordered (fixed to 1, 2 and 3), while the second ones exhaust all the even and odd permutations with positive and negative sign, respectively. This technique applies to any dimension, while the Saros and triangle rules, used to depict the way a determinant is obtained, works only for 3×3 matrices. The number of summands, however, increases rapidly with n, e.g., $4! = 24$, $5! = 120$ etc., hence we need other methods to calculate higher-dimensional determinants.

Higher-Dimensional Determinants

One standard way of calculating determinants in arbitrary dimension is based on the property that for triangular matrices the only non-zero term in the corresponding expression is the product of all diagonal elements, i.e.,

$$\text{if } A = \{a_{ij}\} \text{ and } a_{ij} \equiv 0 \text{ for } i > j \text{ (or } i < j), \text{ then } \det A = \prod_{i=1}^{n} a_{ii}.$$

We may also use the fact that each *elementary transposition* (exchange of two succeeding rows or columns) changes the sign of the determinant. Thus, it the i-th row (or column) should go to the place of the j-th one, the determinant is multiplied by a factor $(-1)^{i-j}$. Consider the following example:

$$\begin{vmatrix} 0 & 0 & 3 \\ 0 & 2 & 1 \\ 1 & 2 & 3 \end{vmatrix} = \begin{vmatrix} 0 & 2 & 1 \\ 1 & 2 & 3 \\ 0 & 0 & 3 \end{vmatrix} = - \begin{vmatrix} 1 & 2 & 3 \\ 0 & 2 & 1 \\ 0 & 0 & 3 \end{vmatrix} = -3\,! = -6$$

where the first transformation does not change the sign as it consists of two elementary transpositions, while the second one does for obvious reasons.

Now, we go a little farther and claim that the determinant is preserved under *similarity transformations*, which in this context means that we are free to add to each row linear combinations of the remaining rows. This certainly holds for columns too, but we are not allowed to manipulate both simultaneously. It is convenient to think of the determinant as representing an ordered system of vectors $\{\mathbf{a}_1, \mathbf{a}_2, \dots \mathbf{a}_n\}$, which undergoes an equivalence transformation of the form

$$\mathbf{a}_i \to \mathbf{a}'_i = \mathbf{a}_i + \sum_{j \neq i} \alpha_j \, \mathbf{a}_j$$

that preserves the value of the determinant. This technique is used to transform numerical matrices into triangular form and thus obtain their determinants. We may consider one more example to illustrate this, namely

$$\begin{vmatrix} 1 & 4 & 3 & 2 \\ 2 & 1 & 4 & 3 \\ 3 & 2 & 1 & 4 \\ 4 & 3 & 2 & 1 \end{vmatrix} = \begin{vmatrix} 1 & 4 & 3 & 2 \\ 0 & -7 & -2 & -1 \\ 0 & -3 & -6 & -1 \\ 0 & 1 & -6 & -5 \end{vmatrix} = \begin{vmatrix} 1 & 4 & 3 & 2 \\ 0 & 0 & 4 & -4 \\ 0 & 0 & -24 & -16 \\ 0 & 1 & -6 & -5 \end{vmatrix} = \begin{vmatrix} 1 & 4 & 3 & 2 \\ 0 & 1 & -6 & -5 \\ 0 & 0 & 4 & -4 \\ 0 & 0 & 0 & -40 \end{vmatrix} = -160.$$

Now, let us explain this result denoting $\mathbf{a}_k$ the k-th vector-row of the determinant. To begin with, we want to eliminate the first components of all $\mathbf{a}_k$ except $\mathbf{a}_1$. We have chosen to do so by the transformation $\mathbf{a}'_2 = \mathbf{a}_2 - 2\mathbf{a}_1$, $\mathbf{a}'_3 = \mathbf{a}_3 - \mathbf{a}_1 - \mathbf{a}_2$ and $\mathbf{a}'_4 = \mathbf{a}_4 - 2\mathbf{a}_2$. On each successive step we preserve one more of the transformed vectors and use it to eliminate the first non-zero component of the remaining ones. In this particular case it is convenient to work with $\mathbf{a}'_4$ as it begins with the number 1 (any other choice involves fractions), so the next step yields $\mathbf{a}''_2 = \mathbf{a}'_2 - \mathbf{a}'_3 + \mathbf{a}'_4$ and $\mathbf{a}''_3 = \mathbf{a}'_3 + 3\mathbf{a}'_4$. Finally, we set $\mathbf{a}'''_3 = \mathbf{a}''_3 + 6\mathbf{a}''_2$ and put the fourth row in the place of the second one (which preserves the sign) to obtain a triangular determinant.

In the attempt to extend the low-dimensional cases while satisfying the all properties postulated above, one comes to an expression for the determinant of a matrix A with entries $\{a_{ij}\}$ in the form

$$\det A = \sum_{i_1, i_2, \dots i_n} (-1)^{[i_1, i_2, \dots i_n]} a_{1 i_1} a_{2 i_2} \dots a_{n i_n} \tag{6.10}$$

where the sum is taken over all permutations of the non-repeating indices $i_1, i_2, \dots i_n$ and the sign is determined by the parity of the particular permutation, i.e., whether the number of elementary transpositions $[i_1, i_2, \dots i_n]$ needed to obtain from it the trivial permutation $1, 2, \dots n$ is even or odd.

Since repetitions are forbidden, in each term we have representatives from every row and column, but never from the same row or column. Therefore, the determinant may also be expressed as a linear combination of the elements of one fixed row or column, respectively

$$\det A = \sum_{j=1}^{n} A_{ij} a_{ij} = \sum_{i=1}^{n} A_{ij} a_{ij}. \tag{6.11}$$

The coefficients A_{ij} in these expansions are called *co-factors* of the entries a_{ij} and their explicit calculation is reduced to lower order determinants, also referred to as *minors*. In particular, all minors of order $n-1$ may be labeled by two indices, say Δ_{ij}^{n-1} denoting that this sub-determinant is obtained from $\det A$ by deleting namely the i-th row and the j-th column. The co-factor A_{ij} is then equal to this minor, up to a sign. More precisely,

$$A_{ij} = (-1)^{i+j} \Delta_{ij}^{n-1}. \tag{6.12}$$

Before we proceed with more examples, let us note that these two iterative procedures for calculating higher order determinants differ in many aspects. First of all, the one using triangular matrices is surely more efficient in terms of computational operations and thus, preferred in numerical algorithms. However, it cannot cope with symbolic expressions, as it uses numerical values explicitly. This difference persists in the context of linear matrix equations and we devote the following example to its clarification.

Example: Let us consider this time a parameter-dependent determinant

$$\Delta(x) = \begin{vmatrix} x & 0 & 1 & 1-x^2 \\ 0 & 1-x^2 & x & 1 \\ 1+x^2 & -x & 1 & 0 \\ 1 & x & 0 & -x \end{vmatrix}.$$

We may simplify a little, performing the transformations $\mathbf{a}_1' = \mathbf{a}_1 - x\mathbf{a}_4$ and $\mathbf{a}_3' = \mathbf{a}_3 - \mathbf{a}_4 - x\mathbf{a}_1$, that yield the equivalent expression

$$\Delta(x) = \begin{vmatrix} 0 & -x^2 & 1 & 1 \\ 0 & 1-x^2 & x & 1 \\ 0 & -2x & 1-x & x^3 \\ 1 & x & 0 & -x \end{vmatrix} = \begin{vmatrix} x^2 & 1 & 1 \\ x^2-1 & x & 1 \\ 2x & 1-x & x^3 \end{vmatrix}$$

where we expand over the first column since we only have one non-zero entry there, which reduces the above to a single 3×3 determinant. Furthermore,

we may eliminate more terms, for example by adding the first row, multiplied by -1 to the second one and by $-x^3$ the third one, thus obtaining

$$\Delta(x) = \begin{vmatrix} x^2 & 1 & 1 \\ -1 & x-1 & 0 \\ 2x-x^5 & 1-x-x^3 & 0 \end{vmatrix} = x^6 - x^5 + x^3 - 2x^2 + x - 1$$

expanding this time over the third column, which finally yields a 2×2 determinant. Note that we may not transform $\Delta(x)$ into a triangular form as before, since in order to do that, at some point we would have to divide by an expression depending on x, and since x is unknown, this expression might be zero.

Problem Session

1. The $n \times n$ Vandermonde determinant is defined as

$$V(x_1, x_2, \dots x_n) = \begin{vmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_n \\ x_1^2 & x_2^2 & \dots & x_n^2 \\ \dots & & & \\ x_1^{n-1} & x_2^{n-1} & \dots & x_n^{n-1} \end{vmatrix}.$$

   Use induction to prove that $V(x_1, x_2, \dots x_n) = \prod_{i>j}(x_i - x_j)$. Calculate $V(z_1, z_2, z_3, z_4)$, where z_k are the complex roots of the equation $z^4 = 1$.

2. Calculate the determinants

$$\begin{vmatrix} 1 & \cos\varphi & \cos\psi \\ \cos\varphi & 1 & \cos(\varphi+\psi) \\ \cos\psi & \cos(\varphi+\psi) & 1 \end{vmatrix}, \qquad \begin{vmatrix} 1 & x & x^2 & \dots & x^{n-1} \\ x^{n-1} & 1 & x & \dots & x^{n-2} \\ x^{n-2} & x^{n-1} & 1 & \dots & x^{n-3} \\ \dots & & & & \\ x & x^2 & x^3 & \dots & 1 \end{vmatrix}.$$

3. There are known to exist *nilpotent* matrices of order $n \in \mathbb{N}$ satisfying $A^n = 0$ with $A \neq 0$. Is there similarly a matrix A with $\det A \neq 0$ and $\det A^n = 0$ for some n?

4. Solve the equation $\begin{vmatrix} 1 & x & x^2 & x^3 \\ x^3 & x^2 & x & 1 \\ 1 & 2x & 3x^2 & 4x^3 \\ 4x^3 & 3x^2 & 2x & 1 \end{vmatrix} = 0.$

6.3 The Inverse Matrix

Up to this point we presented the algebra of square matrices as a linear space with an additional operation: matrix multiplication. For the applications it is often important to know (if possible) how to invert this operation, i.e., divide. For example, it order to solve a simple matrix equation of the form $AX = B$, we need such an inversion procedure. Since multiplication is non-commutative, we expect the same for division and it is best to view it as a multiplication with A^{-1}, referred to as the *inverse matrix* of A. In particular, we may have left and right *pseudo-inverses*, satisfying respectively $A_L^{-1}A = I$ and $AA_R^{-1} = I$, which are different in the generic case. However, for square matrices they coincide and we are going to restrict our considerations mostly to this case, in which we have for its inverse

$$AA^{-1} = A^{-1}A = I \tag{6.13}$$

and in particular, $I^{-1} = I$ (are there other matrices with this property?). It is natural to ask if all matrices are invertible, i.e., does the equation $AX = B$ always have a solution. The answer is negative as we are going to show in a while - it turns out that A^{-1} exists if and only if A is *regular*, also referred to as *non-singular*[4], meaning simply that $\det A \neq 0$. Actually, that is one of the two main reasons to consider determinants in the first place.

Let us recall the row/column expansion (6.11) and point out that if we sum the elements of a given row/column with the co-factors of another one, this corresponds to a determinant with two equal rows/columns (explain why!). This vanishing expression is called a *fake expansion* and we may include it in formula (6.11), which now adopts the more general form

$$\sum_{j=1}^{n} A_{ij}a_{kj} = \delta_{ik}\det A, \qquad \sum_{i=1}^{n} A_{ij}a_{ik} = \delta_{jk}\det A \tag{6.14}$$

where δ_{ij} is the *Kronecker symbol* defined as $\delta_{ii} = 1$ and $\delta_{ij} = 0$ for $i \neq j$. Note that δ_{ij} are exactly the components of the identity matrix $I \to \{\delta_{ij}\}$, so if we interpret (6.14) as a matrix multiplication, it may be the key to finding the inverse matrix according to the definition (6.13). The only problem is that summation should be taken over the middle indices as formula (6.2) determines. To overcome this, we may transpose the matrix consisting of all co-factors, thus obtaining what is usually referred to as the *adjoint matrix*

$$A^* : \qquad (A^*)_{ij} = A_{ji} \tag{6.15}$$

[4]likewise, matrices with a vanishing determinant are called *singular*.

which allows for writing (6.11) in the matrix form

$$AA^* = A^*A = \det(A)E$$

and dividing by $\det A$ (as long it is nonzero!), we finally obtain

$$A^{-1} = \frac{1}{\det A}\, A^*. \tag{6.16}$$

Let us apply this to the simplest non-trivial example we know

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

We have $\det A = ad - bc$ and the co-factors are easy to calculate as well, namely $A_{11} = d$, $A_{12} = -c$, $A_{21} = -b$ and $A_{22} = a$. Arranging them in a matrix an then transposing we obtain the adjoint of A as

$$A^* = \begin{pmatrix} A_{11} & A_{21} \\ A_{12} & A_{22} \end{pmatrix} = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

In other words, the diagonal elements switch their places, and the off-diagonal ones change their signs, which finally yields the inverse of an arbitrary regular 2×2 matrix in the form

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}. \tag{6.17}$$

Let us consider also a three-dimensional example, namely

$$A = \begin{pmatrix} 2 & -1 & 0 \\ 1 & 0 & 1 \\ 0 & 2 & 3 \end{pmatrix} \quad \Rightarrow \quad \det A = -1$$

in which it is not hard to obtain the co-factors

$$A_{11} = (-1)^{1+1} \begin{vmatrix} 0 & 1 \\ 2 & 3 \end{vmatrix} = -2, \quad A_{12} = (-1)^{1+2} \begin{vmatrix} 1 & 1 \\ 0 & 3 \end{vmatrix} = -3, \quad A_{13} = (-1)^{1+3} \begin{vmatrix} 1 & 0 \\ 0 & 2 \end{vmatrix} = 2$$

$$A_{21} = (-1)^{2+1} \begin{vmatrix} -1 & 0 \\ 2 & 3 \end{vmatrix} = 3, \quad A_{22} = (-1)^{2+2} \begin{vmatrix} 2 & 0 \\ 0 & 3 \end{vmatrix} = 6, \quad A_{23} = (-1)^{2+3} \begin{vmatrix} 2 & -1 \\ 0 & 2 \end{vmatrix} = -4$$

$$A_{31} = (-1)^{3+1} \begin{vmatrix} -1 & 0 \\ 0 & 1 \end{vmatrix} = -1, \quad A_{32} = (-1)^{3+2} \begin{vmatrix} 2 & 0 \\ 1 & 1 \end{vmatrix} = -2, \quad A_{33} = (-1)^{3+3} \begin{vmatrix} 2 & -1 \\ 1 & 0 \end{vmatrix} = 1$$

while for the determinant we have $\det A = -1$ and therefore

$$A^{-1} = \frac{1}{\det A}\begin{pmatrix} A_{11} & A_{21} & A_{31} \\ A_{12} & A_{22} & A_{32} \\ A_{13} & A_{23} & A_{33} \end{pmatrix} = \begin{pmatrix} 2 & -3 & 1 \\ 3 & -6 & 2 \\ -2 & 4 & -1 \end{pmatrix}. \tag{6.18}$$

Let us point out some major properties of inverse matrices (prove them!):

a) $(A^t)^{-1} = (A^{-1})^t$

b) $(\lambda A)^{-1} = \frac{1}{\lambda}A^{-1}, \qquad \lambda \neq 0$

c) $(AB)^{-1} = B^{-1}A^{-1}$

d) $\det A^{-1} = \frac{1}{\det A}$

and we have $(A \pm B)^{-1} \neq A^{-1} \pm B^{-1}$. Note that the existence of A^{-1} and B^{-1} does not necessarily imply that $(A \pm B)$ is invertible: consider for example

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad B = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

For non-square matrices one cannot define a determinant or inverse, but there is a generalized notion of a *pseudo-inverse*, which also exists only if certain conditions hold. This construction allows for defining a left and right (pseudo-)inverse of a matrix with the properties $A_L^{-1}A = I$ and $AA_R^{-1} = I$, respectively. Note that A_L^{-1} and A_R^{-1} may exist simultaneously only for square matrices (in which case they coincide), while in the most generic setting none of them does. We investigate this matter more thoroughly in the following chapter and relate it to the solution of linear matrix equations.

Problem Session:

1. Find the inverse matrix of A if

$$A = \begin{pmatrix} 0 & 0 & 0 & 4 \\ 0 & 0 & 3 & 0 \\ 0 & 2 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}, \qquad A = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & -1 \end{pmatrix}, \qquad A = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

2. Prove that for every regular matrix $(A^n)^{-1} = (A^{-1})^n = A^{-n}$.

3. Assuming that A is regular, prove the implication

$$[A, B] = 0 \quad \Rightarrow \quad [A^n, B] = 0, \qquad n \in \mathbb{Z}$$

4. Show that if $(I - A)^{-1}$ exists, then $(I - A^{N+1})(I - A)^{-1} = I + A + \ldots + A^N$.

Chapter 7

Systems of Linear Equations

Consider the system of linear equations

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 &= b_1 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 &= b_2 \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 &= b_3 \end{aligned} \tag{7.1}$$

which may be written as a matrix equation in the form

$$A\vec{x} = \vec{b} \tag{7.2}$$

where $A \leftrightarrow \{a_{ij}\}$ is referred to as the *matrix of the system*, the *righthand side* $\vec{b}$ is a matrix-column and so is the *unknown* (vector) $\vec{x}$. This construction is naturally generalized to the n-dimensional case. In the case of systems with constant coefficients it is convenient to work with the so-called *augmented matrix* $(A\,|\,\vec{b})$, which contains all the information in a concise form.

7.1 Crammer's Rule

Consider the so called *regular case* first, in which $\Delta = \det A \neq 0$ and thus, A is invertible. The system has a unique solution given by $\vec{x} = A^{-1}\,\vec{b}$ and substituting the explicit form (6.18) of A^{-1} yields for its components

$$x_k = \Delta^{-1} \sum_{j=1}^{3} A_{jk} b_j.$$

Note that the above sum may be interpreted as a k-th column expansion of the determinant consisting of the entries of A with $\vec{b}$ substituting the k-th

column. We denote the latter by Δ_k, more precisely

$$\Delta_1 = \begin{vmatrix} b_1 & a_{12} & a_{13} \\ b_2 & a_{22} & a_{23} \\ b_3 & a_{32} & a_{33} \end{vmatrix}, \qquad \Delta_2 = \begin{vmatrix} a_{11} & b_1 & a_{13} \\ a_{21} & b_2 & a_{23} \\ a_{31} & b_3 & a_{33} \end{vmatrix} \qquad \Delta_3 = \begin{vmatrix} a_{11} & a_{12} & b_1 \\ a_{21} & a_{22} & b_2 \\ a_{31} & a_{32} & b_3 \end{vmatrix}.$$

With this notation the unique solution in the regular case is given by components via the so-called *Crammer's rule*

$$x_i = \frac{\Delta_i}{\Delta}, \qquad \Delta \neq 0 \tag{7.3}$$

which is more efficient to express the entries of the solution $\vec{x} = A^{-1}\,\vec{b}$ and not as efficient as the Gauss method discussed below. Nevertheless, the above formula suggests a closed form solution that is a major advantage compared to the (purely numerical) Gauss method especially in the case of parameter-dependent systems, in which it is often the only tool for obtaining the solution. Consider for example the system with augmented matrix

$$\left(\begin{array}{ccc|c} 2 & 1 & \lambda & 6 \\ 7 & 1 & -1 & 10 \\ 1 & -2 & 2 & -5 \end{array}\right), \qquad \lambda \in \mathbb{R}$$

in which one easily obtains

$$\Delta = \begin{vmatrix} 2 & 1 & \lambda \\ 7 & 1 & -1 \\ 1 & -2 & 2 \end{vmatrix} = \begin{vmatrix} 2 & 1 & \lambda+1 \\ 7 & 1 & 0 \\ 1 & -2 & 0 \end{vmatrix} = -15(\lambda+1)$$

by first adding the second column to the third one and then expanding over the latter. Similarly, we have

$$\Delta_1 = \begin{vmatrix} 6 & 1 & \lambda \\ 10 & 1 & -1 \\ -5 & -2 & 2 \end{vmatrix} = \begin{vmatrix} 6 & 1 & \lambda+1 \\ 10 & 1 & 0 \\ -5 & -2 & 0 \end{vmatrix} = -15(\lambda+1)$$

as well as

$$\Delta_2 = \begin{vmatrix} 2 & 6 & \lambda \\ 7 & 10 & -1 \\ 1 & -5 & 2 \end{vmatrix} = \begin{vmatrix} 2 & 16 & \lambda-4 \\ 7 & 45 & -15 \\ 1 & 0 & 0 \end{vmatrix} = -15(3\lambda+4)$$

and finally

$$\Delta_3 = \begin{vmatrix} 2 & 1 & 6 \\ 7 & 1 & 10 \\ 1 & -2 & -5 \end{vmatrix} = \begin{vmatrix} 2 & 5 & 16 \\ 7 & 15 & 45 \\ 1 & 0 & 0 \end{vmatrix} = -15$$

which yields, with the aid of formula (7.3), the unique solution for the regular case $\Delta \neq 0 \Leftrightarrow \lambda \neq -1$ in the form

$$x_1 = 1, \qquad x_2 = \frac{3\lambda+4}{\lambda+1}, \qquad x_3 = \frac{1}{\lambda+1}.$$

7.2 The Gauss Method

There is another way of solving linear systems with numerical coefficients, which is far more efficient in terms of complexity, but is given as an algorithm rather than a closed formula. It is based on the property that elementary matrix transformations of the augmented matrix $(A \,|\, \vec{b})$ do preserve the system (up to equivalence) and hence, its solution. As we learned from working with determinants, such transformations allow for obtaining the matrix in a triangular form. Then the last equation is only for x_n and we may substitute its solution into the previous one (which involves x_{n-1} and x_n) and so on until we obtain all the unknowns. This is the so-called *Gauss method*. In the regular case one may proceed with the elimination from the bottom up, thus transforming the matrix into a diagonal form and then, dividing each row by its diagonal element, obtain the identity matrix on the left. Then, on the right, one ends up with the solution, i.e.,

$$(A \,|\, \vec{b}) \sim \ldots \sim (I \,|\, \vec{x}).$$

The whole procedure is usually referred to as the *Gauss-Jordan method* (elimination). Let us begin with a simple 2×2 example, namely the system

$$x - y = 1, \qquad x + 2y = 4$$

and let us transform its augmented matrix into the form

$$\left(\begin{array}{cc|c} 1 & -1 & 1 \\ 1 & 2 & 4 \end{array}\right) \sim \left(\begin{array}{cc|c} 1 & -1 & 1 \\ 0 & 3 & 3 \end{array}\right) \sim \left(\begin{array}{cc|c} 1 & -1 & 1 \\ 0 & 1 & 1 \end{array}\right) \sim \left(\begin{array}{cc|c} 1 & 0 & 2 \\ 0 & 1 & 1 \end{array}\right).$$

Then, the solution is given by $x = 2$, $y = 1$. Note that unlike in the case of determinants, here we are allowed to swap rows and multiply them by arbitrary non-zero constants. On the other hand, we cannot work with columns as this alters the equations. Consider a second example in the form

$$\left(\begin{array}{ccc|c} 1 & 2 & 3 & 1 \\ 2 & 3 & 4 & 1 \\ 3 & 4 & 5 & 1 \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & 2 & 3 & 1 \\ 0 & 1 & 2 & 1 \\ 0 & 2 & 4 & 2 \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & 2 & 3 & 1 \\ 0 & 1 & 2 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right).$$

The last row corresponds to an equation of the form $0 = 0$ that is always satisfied, so it does not provide any information. Thus, we end up with two equations for three variables, which cannot lead to an unique solution and we introduce a parameter, e.g. $x_3 = t$. Then, from the second (transformed) equation we have $x_2 = 1 - 2t$ and substituting in the first one, finally $x_1 = t - 1$. The solutions are ultimately given by the one-parameter set

$$(x_1,\ x_2,\ x_3) = (t - 1,\ 1 - 2t,\ t), \qquad t \in \mathbb{R}.$$

Such systems, in which linear dependence of the matrix-rows effectively reduces the number of equations and the solutions are therefore expressed in terms of one or more parameters, are called *dependent*. Note that such dependance is only possible in the singular case $\Delta = 0$. Now, let us return to the example considered in the previous section, in which we used Crammer's formula to retrieve the solution in the regular case. In the singular one we only need to substitute λ with -1 and apply the Gauss method to obtain

$$\left(\begin{array}{ccc|c} 2 & 1 & -1 & 6 \\ 7 & 1 & -1 & 10 \\ 1 & -2 & 2 & -5 \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & -2 & 2 & -5 \\ 2 & 1 & -1 & 6 \\ 7 & 1 & -1 & 10 \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & -2 & 2 & -5 \\ 0 & 5 & -5 & 16 \\ 0 & 15 & -15 & 45 \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & -2 & 2 & -5 \\ 0 & 5 & -5 & 16 \\ 0 & 0 & 0 & 1 \end{array}\right)$$

where the last equation in the transformed system may be written as

$$0x_1 + 0x_2 + 0x_3 = 1$$

and there are obviously no real (or complex) numbers x_i, for which it might possibly be satisfied. But then, the whole system has no solution since its equations contradict with each other. Such systems are called *inconsistent*. We devote the following section to the study of determined, dependent and inconsistent systems. Let us only note that in the 3×3 case, each linear equation has a geometric interpretation of an infinite plane. Two such planes are either parallel (in particular coincident) or intersect in a line. In the regular case the two infinite lines corresponding to the intersection of pairs of equations in the system, themselves intersect at a point, which gives the solution. These two lines, on the other hand, may also coincide (in the dependent case), or be parallel (in the inconsistent one). Moreover, two or even all three of the planes may be parallel or coincident, which corresponds to either inconsistency or a two-parameter degenerate solution, respectively. These settings are intrinsically related to the notion of *rank* discussed below.

7.3 The Kronecker-Capelli Theorem

In the previous section it become clear that in the regular case $\det A \neq 0$ the system of n linear equations for n unknowns $A\vec{x} = \vec{b}$ has a unique solution given by the Crammer's formula, that is essentially the inverse matrix construction $\vec{x} = A^{-1}\,\vec{b}$ written in components. In the singular case, on the other hand, i.e., when $\det A = 0$, one has two possibilities: the solutions are either infinitely many and the system is said to be dependent, or none, in which case it is referred to as inconsistent. It turns out that there is a simple way to determine to which of the two classes a given singular

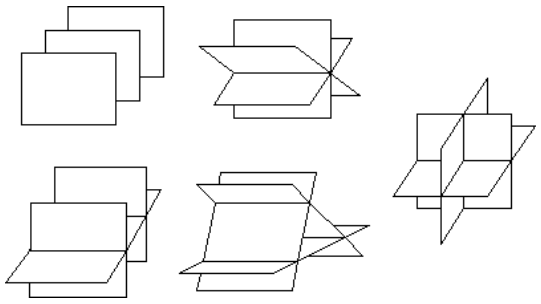

Figure 7.1: Illustration of the solutions to linear systems as intersections of planes.

system belongs without having to solve it. The following theorem makes use of the notion of rank that has various definitions in literature, some of which sound quite puzzling. We first define the *rank of a system of vectors* as the dimension of their linear span that is certainly equal to the maximal number of linearly independent vectors within the system. The *rank of a matrix* A is then naturally defined as the rank of the system of its vector-rows or vector-columns. First, let us note that even in the case of non-square matrices the two definitions are equivalent due to the following

Theorem 6 *Each system of* $n+1$ *vectors in* $\mathbb{R}^n$ *is linearly dependent.*

In other words, a system of vectors in $\mathbb{R}^n$ can never have a linear span which exceeds the ambient space. The existence of bases in $\mathbb{R}^n$ is a straightforward consequence allowing each vector to be expressed as a linear combination of a system with maximal rank. Now, let A be an $m \times n$ matrix, so it may be interpreted as an ordered system of m vectors in $\mathbb{R}^n$ or n vectors in $\mathbb{R}^m$. Then, according to the above theorem, the maximal rank it can have is $\min(n, m)$, i.e., the smaller of the two dimensions. Thus, one may always isolate a square sub-matrix, which contains all the information about the rank and square matrices have maximal rank exactly when the determinant is non-vanishing, i.e., there are no linear dependencies between rows or columns. In fact, the equivalence of the definitions for rows and columns follows from the property of determinants $\det A = \det A^t$ as the rank may also be thought of as the dimension of the maximal non-trivial minor contained in A. Each new relation (linear dependence) between vector-rows (or columns) reduces the rank by one, so we may also define the latter as the total number of vectors in the system minus the number of independent

relations between them. Consider for example the matrix

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \\ 1 & -1 & 0 \end{pmatrix}.$$

Its rank is clearly not maximal, since the second row is equal to the sum of the other two and since this is the only non-trivial relation between the vector-rows, we have $\mathrm{r}(A) = 2$. The invariance of the definition with respect to matrix transpositions demands that there is exactly one linear relation between the vector-columns too. Indeed, it is not difficult to see that the third column may be obtained as a sum of the first two. In this case each pair of vector-rows or columns spans a plane (all minors of order two are non-vanishing), but the three vectors cannot span a three-dimensional space.

With this in mind, we may formulate and proof the famous

Theorem 7 (*Kronecker-Capelli*) *The system of linear equations* $A\vec{x}=\vec{b}$ *has a solution if and only if the rank of* A *is equal to that of* $(A \,|\, \vec{b})$.

Proof. Let $\vec{a}_k$ denote the k-th vector column of A. Then, the existence of a solution $\vec{x}$ yields $x_1\vec{a}_1 + x_2\vec{a}_2 + \ldots + x_n\vec{a}_n = \vec{b}$, so the ranks of the systems $\{\vec{a}_1, \ldots \vec{a}_n\}$ and $\{\vec{a}_1, \ldots \vec{a}_n, b\}$ are equal. Conversely, if the ranks of the latter systems are equal, then $\vec{b}$ can be expressed as a linear combination of the $\vec{a}_k$'s in the form $\vec{b} = x_1\vec{a}_1 + x_2\vec{a}_2 + \ldots + x_n\vec{a}_n$, which may be written also as $A\vec{x} = \vec{b}$, so that x_i are the components of the solution. □

Note that the above examples illustrate perfectly the theorem - in the regular case $\Delta \neq 0$ the rank is already maximal and the additional vector-column $\vec{b}$ cannot increase it, so a solution always exists. As for the singular case, the equivalence between row and column based definitions yields that whenever the ranks of A and $(A \,|\, \vec{b})$ are equal, the Gauss method will generate the same number of identically zero rows in both. On the contrary, if the two ranks disagree, we are left with an equation of the form $0x_1 + \ldots + 0x_n = 1$ that is impossible to solve and thus reveals the inconsistency of the whole system. Moreover, if the conditions of the theorem are satisfied, the number of unknowns minus the number of independent equations is equal to the number of free parameters in the solution. For example, in the regular case the two numbers are equal and the solution determines a point in $\mathbb{R}^n$ (no free parameters). In the dependent case, on the other hand, we may have k independent relations between the vector-rows (or columns) of A. Then,

obviously $\mathrm{r}(A) = n - k$ and we end up with a k-parametric degenerate solution. For $k = 1$ it is a line, for $k = 2$, a plane etc. In this sense, the dimension of the resolution set equals the *codimension*[1] of $\mathrm{Span}\{\vec{a}_k\}$ and vice versa. Consider the following example

$$\left(\begin{array}{ccc|c} \lambda & \mu & 2 & 1 \\ \lambda & 2\mu-1 & 3 & 1 \\ \lambda & \mu & \mu+3 & 2\mu-1 \end{array}\right), \qquad \lambda, \mu \in \mathbb{R}$$

and let us figure out for which values of the real parameters λ and μ the system has a solution. Firstly, this is obviously so in the regular case

$$\Delta = \begin{vmatrix} \lambda & \mu & 2 \\ \lambda & 2\mu-1 & 3 \\ \lambda & \mu & \mu+3 \end{vmatrix} = \begin{vmatrix} \lambda & \mu & 2 \\ 0 & \mu-1 & 1 \\ 0 & 0 & \mu+1 \end{vmatrix} = \lambda(\mu^2 - 1) \neq 0.$$

In the singular case, for $\lambda = 0$ we have $\mathrm{r}(A) = 2$ and thus μ needs to satisfy

$$\Delta_1 = \begin{vmatrix} 1 & \mu & 2 \\ 1 & 2\mu-1 & 3 \\ 2\mu-1 & \mu & \mu+3 \end{vmatrix} = \begin{vmatrix} 1 & \mu & 2 \\ 0 & \mu-1 & 1 \\ 2(\mu-1) & 0 & \mu+1 \end{vmatrix} = (\mu - 1)(5 - \mu) = 0$$

while Δ_2 and Δ_3 are zero, because they both have a vanishing first column. Furthermore, if $\mu = 1$, then all the Δ_i's vanish and $\mu = -1$ is not an option since $\Delta_1 \neq 0$ at that point. To sum up, solutions always exist for $\mu = 1$ and in the case $\lambda = 0$, $\mu = 5$. Only then we have $\Delta = \Delta_k = 0$, which guarantees the equality $\mathrm{r}(A) = \mathrm{r}(A \,|\, \vec{b})$. Thus, apart from the regular case, in which the Crammer's formula applies (obtain the solutions in this case!), we have two more configurations that allow for a one-parameter solution, namely[2]

* $\mu = 1$, while λ is arbitrary:

$$\left(\begin{array}{ccc|c} \lambda & 1 & 2 & 1 \\ \lambda & 1 & 3 & 1 \\ \lambda & 1 & 4 & 1 \end{array}\right) \sim \left(\begin{array}{ccc|c} \lambda & 1 & 2 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \end{array}\right) \quad \Rightarrow \quad \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} t \\ 1-\lambda t \\ 0 \end{bmatrix}, \quad t \in \mathbb{R}.$$

** $\lambda = 0,\ \mu = 5$:

$$\left(\begin{array}{ccc|c} 0 & 5 & 2 & 1 \\ 0 & 9 & 3 & 1 \\ 0 & 5 & 8 & 9 \end{array}\right) \sim \left(\begin{array}{ccc|c} 0 & 5 & 2 & 1 \\ 0 & 9 & 3 & 1 \\ 0 & 0 & 3 & 4 \end{array}\right) \sim \left(\begin{array}{ccc|c} 0 & 5 & 2 & 1 \\ 0 & 3 & 0 & -1 \\ 0 & 0 & 3 & 4 \end{array}\right) \quad \Rightarrow$$

[1] i.e., the dimension of the complementary space $\mathrm{codim}\, V = \dim \mathbb{R}^n / V$, where $V \subset \mathbb{R}^n$.

[2] in this particular case, we end up with a vanishing column (instead of row).

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 3s \\ -1 \\ 4 \end{bmatrix}, \quad s \in \mathbb{R}.$$

Note that one may also consider systems of equations, in which the relations are more or less than the number of unknowns. In the former case, the system is either over-determined, or inconsistent, while in the latter, it might only be dependent or inconsistent. Fortunately, the Kronecker-Capelli allows to distinguish between those possibilities here as well. As we shall see later, each linear equation for x, y and z has the interpretation of a plane in $\mathbb{R}^3$. Then, the solutions to 3 linear systems correspond to intersections (common points, if any) of three such planes, and the Kronecker-Capelli theorem allows us to distinguish between different cases (see Figure 7.1) based on the rank. This theorem remains valid in the more general context of matrix equations considered in the following section.

7.4 Matrix Equations

In this section we consider more generic linear equations

$$AX + B, \qquad XA + B \tag{7.4}$$

in which both the righthand side and the unknown are square matrices. Note that matrix transposition allows for converting one type into another:

$$(XA)^t = A^t X^t = B^t$$

so we may consider only the first one. One way to think of it is as of a system of systems of linear equations, for which the unknown vectors are given by the columns of X and the righthand sides - by the corresponding columns of the matrix B. Such interpretation allows for applying directly most of the results obtained above. For example, Crammer's rule may be generalized in this case as

$$x_{ij} = \frac{\Delta_{ij}}{\Delta}, \qquad \Delta \neq 0 \tag{7.5}$$

where Δ_{ij} is the determinant obtained from A with the j-th column of B replacing its i-th column[3]. This is just a more efficient way of writing the unique solution in the regular case $X = A^{-1}B$. Moreover, the Gauss method

[3] it should not be confused with the minor $\Delta_{ij}^{(n-1)}$ in the definition of the co-factor A_{ij}.

applies directly too. In particular, if A is invertible, one may transform the augmented matrix in such a way that the identity matrix is on the left. Then, on the right one has the unique solution, i.e.,

$$(A\,|\,B) \sim \ldots \sim (I\,|\,X).$$

Let us consider a simple example

$$\left(\begin{array}{cc|cc} 1 & -1 & 1 & 2 \\ 1 & 2 & 4 & 5 \end{array}\right) \sim \left(\begin{array}{cc|cc} 1 & -1 & 1 & 2 \\ 0 & 3 & 3 & 3 \end{array}\right) \sim \left(\begin{array}{cc|cc} 1 & -1 & 1 & 2 \\ 0 & 1 & 1 & 1 \end{array}\right) \sim \left(\begin{array}{cc|cc} 1 & 0 & 2 & 3 \\ 0 & 1 & 1 & 1 \end{array}\right)$$

which comes to show that the solution to the matrix equation

$$\begin{pmatrix} 1 & -1 \\ 1 & 2 \end{pmatrix} X = \begin{pmatrix} 1 & 2 \\ 4 & 5 \end{pmatrix} \quad \text{is} \quad X = \begin{pmatrix} 2 & 3 \\ 1 & 1 \end{pmatrix}$$

that may easily be tested via straightforward multiplication. Next, we take an equation with the same A and B, but this time in the form $XA = B$. We cannot apply the Gauss algorithm directly when X is a left factor, so we do it for the transposed system $A^t X^t = B^t$ instead, and thus obtain

$$\left(\begin{array}{cc|cc} 1 & 1 & 1 & 4 \\ -1 & 2 & 2 & 5 \end{array}\right) \sim \left(\begin{array}{cc|cc} 1 & 1 & 1 & 4 \\ 0 & 3 & 3 & 9 \end{array}\right) \sim \left(\begin{array}{cc|cc} 1 & 1 & 1 & 4 \\ 0 & 1 & 1 & 2 \end{array}\right) \sim \left(\begin{array}{cc|cc} 1 & 0 & 0 & 2 \\ 0 & 1 & 1 & 2 \end{array}\right).$$

Since the solution on the right is for X^t, we need to transpose once more

$$X^t = \begin{pmatrix} 0 & 2 \\ 1 & 2 \end{pmatrix} \quad \Rightarrow \quad X = \begin{pmatrix} 0 & 1 \\ 2 & 2 \end{pmatrix}.$$

In these examples it is quite easy to derive the inverse matrix

$$A^{-1} = \begin{pmatrix} 1 & -1 \\ 1 & 2 \end{pmatrix}^{-1} = \frac{1}{3} \begin{pmatrix} 2 & 1 \\ -1 & 1 \end{pmatrix}$$

using formula (6.17) and then obtain the solutions to the two matrix equations (7.4) respectively as $X = A^{-1}B$ and $X = BA^{-1}$. For $n > 2$, on the other hand, the Gauss-Jordan elimination is much more efficient. In particular, one may use it to derive A^{-1} as a solution to the equation $AX = I$

$$(A\,|\,I) \sim \ldots \sim (I\,|\,A^{-1}).$$

Let us calculate as an example the inverse matrix of $A = \begin{pmatrix} 2 & 1 & -1 \\ 2 & -1 & 2 \\ 3 & 0 & 1 \end{pmatrix}$. The Gauss-Jordan algorithm yields in this case

$$\left(\begin{array}{ccc|ccc} 2 & 1 & -1 & 1 & 0 & 0 \\ 2 & -1 & 2 & 0 & 1 & 0 \\ 3 & 0 & 1 & 0 & 0 & 1 \end{array}\right) \sim \left(\begin{array}{ccc|ccc} 2 & 1 & -1 & 1 & 0 & 0 \\ 0 & -2 & 3 & -1 & 1 & 0 \\ 1 & 1 & -1 & 0 & -1 & 1 \end{array}\right) \sim \left(\begin{array}{ccc|ccc} 1 & 1 & -1 & 0 & -1 & 1 \\ 2 & 1 & -1 & 1 & 0 & 0 \\ 0 & -2 & 3 & -1 & 1 & 0 \end{array}\right) \sim$$

$$\left(\begin{array}{ccc|ccc} 1 & 1 & -1 & 0 & -1 & 1 \\ 0 & -1 & 1 & 1 & 2 & -2 \\ 0 & -2 & 3 & -1 & 1 & 0 \end{array}\right) \sim \left(\begin{array}{ccc|ccc} 1 & 1 & -1 & 0 & -1 & 1 \\ 0 & -1 & 1 & 1 & 2 & -2 \\ 0 & 0 & 1 & -3 & -3 & 4 \end{array}\right) \sim \left(\begin{array}{ccc|ccc} 1 & 1 & -1 & 0 & -1 & 1 \\ 0 & -1 & 0 & 4 & 5 & -6 \\ 0 & 0 & 1 & -3 & -3 & 4 \end{array}\right) \sim$$

$$\left(\begin{array}{ccc|ccc} 1 & 1 & 0 & -3 & -4 & 5 \\ 0 & -1 & 0 & 4 & 5 & -6 \\ 0 & 0 & 1 & -3 & -3 & 4 \end{array}\right) \sim \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 1 & -1 \\ 0 & 1 & 0 & -4 & -5 & 6 \\ 0 & 0 & 1 & -3 & -3 & 4 \end{array}\right) \Rightarrow A^{-1} = \begin{pmatrix} 1 & 1 & -1 \\ -4 & -5 & 6 \\ -3 & -3 & 4 \end{pmatrix}$$

while the definition of matrix inverse demands the computation of nine determinants for the co-factors. In the case of 3×3 matrices this does not look like such an enormous amount of work, but for a generic n this means n^2 determinants of order $n-1$ (e.g. for $n=5$ those are 25 determinants of order 4, or 100 of order 3), so the complexity of this method increases very rapidly with n, while for the Gauss-Jordan algorithm this increase is relatively slow.

Linear Homogeneous Equations

We consider a particular type of matrix equations (or systems of linear equations) with a vanishing righthand side, referred to as *homogeneous*. In this setting our work is greatly simplified, e.g. in the regular case the unique solution guaranteed by the Crammer's rule is the trivial one[4]. Moreover, for $B = 0$ one always has $\mathrm{r}\,(A \,|\, B) = \mathrm{r}\,(A)$ as no new vectors are introduced to the system and by the Kronecker-Capelli theorem a solution always exists. This proves the following

Theorem 8 *The necessary and sufficient condition for the existence of a non-trivial solution to the homogeneous equation* $AX = 0$ *is* $\det A = 0$.

One more convenience provided by the homogeneous case is that we do not need to bother about the extensions when applying the Gauss algorithm as it simply does not affect them. Consider for example the following solution

$$\left(\begin{array}{cccc|c} 3 & 4 & -5 & 7 & 0 \\ 2 & -3 & 3 & -2 & 0 \\ 4 & 11 & -13 & 16 & 0 \\ 7 & -2 & 1 & 3 & 0 \end{array}\right) \sim \left(\begin{array}{cccc|c} 1 & 7 & -8 & 9 & 0 \\ 2 & -3 & 3 & -2 & 0 \\ 0 & 17 & -19 & 20 & 0 \\ 0 & -17 & 19 & -20 & 0 \end{array}\right) \sim \left(\begin{array}{cccc|c} 1 & 7 & -8 & 9 & 0 \\ 0 & -17 & 19 & -20 & 0 \\ 0 & 17 & -19 & 20 & 0 \\ 0 & -17 & 19 & -20 & 0 \end{array}\right)$$

in which the last two rows do not contribute any new information to the system and may thus be ignored, effectively leaving two independent relations for a total of four unknowns. Thus, the solution will contain two free parameters and may be written for example as

$$x_1 = 3s - 13t, \quad x_2 = 19s - 20t, \quad x_3 = 17s, \quad x_4 = 17t \quad (s, t \in \mathbb{R}).$$

[4] all determinants Δ_{ij} in formula (7.5) contain a zero column.

Note that in the homogeneous case adding more trivial columns in the right-hand side does not demand more calculations, e.g. if the extension consists of n zero columns, the above expressions would still adequately describe the k-th column of the solution with independent parameters $s_k, t_k \in \mathbb{R}$. Another property of linear homogeneous systems is given by the following

Theorem 9 *The solutions of a linear homogeneous matrix equation* (*and in particular, system of linear equations*) *constitute a vector space.*

Proof. Let X_1 and X_2 be solutions to $AX = 0$. Then all linear combinations of the type $\lambda X_1 + \mu X_2$ with $\lambda, \mu \in \mathbb{C}$ are also solutions, since linearity yields

$$A(\lambda X_1 + \mu X_2) = \lambda A(X_1) + \mu A(X_2) = 0\lambda + 0\mu = 0. \qquad \square$$

7.5 Pseudo-Inverse Matrices

In this section we consider a generic type of matrix equations (7.4), in which the matrices A and B (and hence the unknown X) are not square. Let for example A be an $m \times n$ matrix and B - an $m \times k$ one, so if $AX = B$, X has to be of dimensions $n \times k$. Suppose also that A has maximal rank, i.e.,

$$\mathrm{r}(A) = \min(m, n).$$

Then, although one may not invert in the strict sense to obtain the solutions respectively as $X = A^{-1}B$ and $X = BA^{-1}$, there is a construction called the *generalized inverse*, or just *pseudo-inverse* of a non-square matrix with very similar properties. More precisely, we define the *left inverse* of a matrix of maximal rank A as the matrix A_L^{-1}, for which $A_L^{-1}A = I$. Similarly, the *right inverse* in this case is defined as A_R^{-1} with the property $AA_R^{-1} = I$. Note the two identity matrices belong to vector spaces of different dimension, so their equality is possible (and unavoidable) only in the case of square matrices. However, even for $m \neq n$ the products $(AA^t)_{m\times m}$ and $(A^tA)_{n\times n}$ always exist and the one with smaller dimension has a non-vanishing determinant (so it is invertible) as long as the rank of A is maximal. Thus, in the case $\mathrm{r}(A) = n < m$ we may only construct the left inverse in the form

$$A_L^{-1} = (A^tA)^{-1}A^t \tag{7.6}$$

while for $\mathrm{r}(A) = m < n$ one has a similar definition for the right inverse as

$$A_L^{-1} = A^t(AA^t)^{-1}. \tag{7.7}$$

If A is an $m\times n$ matrix and $\mathrm{r}(A) = n < m$, then A_L^{-1} is $n\times m$, so applying A_L^{-1} on the left to both sides of the equation $AX = B$, we seem to obtain the unique solution $X = A_L^{-1}B$ with the correct dimensions. Note, however, that the Kronecker-Capelli theorem still applies to this case and the rank of $(A\,|\,B)$ may exceed that of A, since we add more columns in the extension. Therefore, one needs to check first whether the two ranks agree. On the other hand, the familiar Gauss method works here with almost no modifications. Moreover, in the case $\mathrm{r}(A) = n < m$, i.e., maximal rank and more equations than variables, we may obtain the upper square block of A in a triangular and even identity matrix form. Then, the remaining rows will either be trivial, in which case the system is simply over-determined, or contain a contradiction and thus, yield inconsistency. In the case $m < n$ the number of equations is not sufficient to determine all the unknowns, so one has to introduce parameters in the solution if it exists[5]. Let us illustrate this with several examples. Consider first a matrix equation in the form

$$\left(\begin{array}{cc|cc} 1 & 1 & 1 & 4 \\ -1 & 2 & 2 & 5 \\ 2 & 3 & 3 & 11 \end{array}\right) \sim \left(\begin{array}{cc|cc} 1 & 1 & 1 & 4 \\ 0 & 3 & 3 & 9 \\ 0 & 1 & 1 & 3 \end{array}\right) \sim \left(\begin{array}{cc|cc} 1 & 1 & 1 & 4 \\ 0 & 1 & 1 & 2 \\ 0 & 1 & 1 & 3 \end{array}\right) \sim \left(\begin{array}{cc|cc} 1 & 0 & 0 & 2 \\ 0 & 1 & 1 & 2 \\ 0 & 0 & 0 & 1 \end{array}\right)$$

which clearly shows contradiction as the last row is not entirely zero. If we had chosen the last entry in the extension above (the matrix B) to be 10, instead of 11, the ranks of A and $(A\,|\,B)$ would agree and we would have ended up with a vanishing row, hence, with a solution contained above as the upper-right 2×2 block, namely $X = \begin{pmatrix} 1 & 2 \\ 1 & 2 \end{pmatrix}$, which may easily be checked. Let us also consider one more example, this time of the second type - with less equations than variables:

$$\left(\begin{array}{ccc|cc} 1 & 1 & 0 & 2 & 3 \\ 2 & 3 & 1 & 4 & 5 \end{array}\right) \sim \left(\begin{array}{ccc|cc} 1 & 1 & 0 & 2 & 3 \\ 0 & 1 & 1 & 0 & -1 \end{array}\right)$$

to which it is straightforward to apply the Gauss method, but one cannot go any further. However, we notice that the rank of A is maximal - in this case, equal to the number of rows, which cannot be increased by adding more columns, so $\mathrm{r}(A) = \mathrm{r}(A\,|\,B)$ and we have a solution, or more precisely, a two-parameter set of solutions for $X_{3\times 2}(s,t)$. It may be easily obtained by writing the above as a pair of entangled systems for the vector-columns of X, namely

$$x_{11} + x_{21} = 2, \qquad x_{21} + x_{31} = 0$$

[5] contradiction is still possible if the ranks in Theorem 7 do not agree.

and respectively

$$x_{12} + x_{22} = 3, \qquad x_{22} + x_{32} = -1.$$

The obvious substitution is $x_{21} = s$ and $x_{22} = t$, where $s, t \in \mathbb{R}$ are two arbitrary real parameters, and thus we have the solution in the form

$$X = \begin{pmatrix} 2-s & 3-t \\ s & t \\ -s & -1-t \end{pmatrix}, \qquad s, t \in \mathbb{R}.$$

Problem Session

1. Determine the rank of the matrix, depending on the real parameter λ

$$\begin{pmatrix} 1 & 0 & -1 \\ 2 & \lambda-2 & 3 \\ 1 & 0 & 4 \end{pmatrix}, \qquad \begin{pmatrix} -1 & 2 & 1 \\ 2 & \lambda & -2 \\ 3 & -6 & -3 \end{pmatrix}, \qquad \begin{pmatrix} 1 & 1 & 1 & 1-\lambda \\ 2 & 2 & 2-\lambda & 2 \\ 3 & 3-\lambda & 3 & 3 \end{pmatrix}.$$

2. Prove that

$$\mathrm{r}(AB) \leq \min[\mathrm{r}(A), \mathrm{r}(B)]$$

and in particular

$$\mathrm{r}(A^n) \leq \mathrm{r}(A).$$

Provide an example, in which the inequality does not hold.

3. Solve the linear system, represented by the augmented matrix

a) $\left(\begin{array}{cccc|c} 1 & -2 & 3 & -4 & 4 \\ 0 & 1 & -1 & 1 & -3 \\ 1 & 3 & 0 & -3 & 1 \\ 0 & -7 & 3 & 1 & -3 \end{array}\right), \qquad \left(\begin{array}{cccc|c} 1 & 0 & -3 & 4 & -5 \\ 1 & 0 & -2 & 3 & -4 \\ 3 & 2 & 0 & -5 & 12 \\ 4 & 3 & -5 & 0 & 5 \end{array}\right)$

b) $\left(\begin{array}{ccc|c} 1 & 1 & 2 & -1 \\ 2 & -1 & 2 & -4 \\ 4 & 1 & 4 & -2 \end{array}\right), \qquad \left(\begin{array}{cccc|c} 1 & -2 & 1 & 1 & 1 \\ 1 & -2 & 1 & -1 & -1 \\ 1 & -2 & 1 & 5 & 5 \end{array}\right)$

c) $\left(\begin{array}{ccc|c} \lambda & \lambda & \lambda+1 & \lambda \\ \lambda & \lambda & \lambda-1 & \lambda \\ \lambda+1 & \lambda & 2\lambda+3 & 1 \end{array}\right), \qquad \left(\begin{array}{ccc|c} \mu & 1 & 1 & 4 \\ 1 & \lambda & 1 & 3 \\ 1 & 2\lambda & 1 & 4 \end{array}\right), \qquad \lambda, \mu \in \mathbb{R}.$

4. If A and B are square matrices and the system $AX = B$ is dependent/inconsistent, then what can be told about the system $XA = B$?

5. Obtain the left inverse of the matrix

$$A = \begin{pmatrix} 1 & 2 & 0 \\ 2 & 0 & 1 \end{pmatrix}$$

and the right one of A^t. What is the explicit relation between the two?

6. Solve for X:

$$X \begin{pmatrix} 1 & 1 & -1 \\ 2 & 1 & 0 \\ 1 & -1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & -1 & 3 \\ 4 & 3 & 2 \\ 1 & -2 & 5 \end{pmatrix}$$

7. Find the solutions to the system $Ax = 0$ where

$$A = \begin{pmatrix} \lambda & 3 & 1 \\ 1 & 1+\lambda & -1 \\ 2\lambda - 1 & 2 & 4 \end{pmatrix}, \qquad \lambda \in \mathbb{R}$$

depending on the value of the parameter.

Chapter 8

Vectors in $\mathbb{R}^3$

8.1 Properties of Vectors in $\mathbb{R}^3$ and $\mathbb{R}^n$

The very definition of a vector is of course independent on the dimensionality and thus, basic properties of vectors in $\mathbb{R}^2$ and $\mathbb{R}^n$ coincide. For example, linear combinations are the most fundamental operations in any vector space and the notion of linear dependence remains identical, i.e., a system of vectors $\{\mathbf{a}_k\}$ is said to be linearly dependent if there exists a vanishing linear combination of the $\mathbf{a}_k$'s whose coefficients are not all zero (equivalently, if one of the vectors can be expressed as a linear combination of the others). Moreover, the number of independent vectors in a given system usually referred to as its *rank* cannot exceed the dimension n of the corresponding space. In the case of equality $\mathrm{r}\{\mathbf{a}_k\} = n$, we say that $\{\mathbf{a}_k\}$ constitutes a basis in $\mathbb{R}^n$ as each vector $\mathbf{v} \in \mathbb{R}^n$ can be expressed (expanded) as a linear combination in the form

$$\mathbf{v} = \sum_{k=1}^{n} v^i \mathbf{a}_i \tag{8.1}$$

where the coefficient v^i is the parallel projection of $\mathbf{v}$ in the direction determined by $\mathbf{a}_i$. In the case of orthogonal bases $\mathbf{a}_i \perp \mathbf{a}_j$ for $i \neq j$ (and only in that case) the latter coincides with the orthogonal projections given by means of the scalar product with the unit vector $\hat{\mathbf{a}}_i$ in the i-th direction:

$$v_i = \mathbf{v} \cdot \hat{\mathbf{a}}_i, \qquad \hat{\mathbf{a}}_i = \frac{\mathbf{a}_i}{|\mathbf{a}_i|}. \tag{8.2}$$

Then, $\{\hat{\mathbf{a}}_i\}$ is said to be an *orthonormal* (i.e., orthogonal normalized) basis. The entries of the *Gram matrix* $G \leftrightarrow \{g_{ij}\}$ for orthonormal bases satisfy

$$g_{ij} = \hat{\mathbf{a}}_i \cdot \hat{\mathbf{a}}_j = \delta_{ij}, \qquad G = \mathrm{Id}. \tag{8.3}$$

The relation between the vector components of $\mathbf{v}$ in two different bases, say $\{\hat{\mathbf{a}}_k\}$ and $\{\hat{\mathbf{a}}'_k\}$, is also linear as one might expect. More precisely, it is defined via the *transition matrix* $T \leftrightarrow \{\tau_i^j\}$ between $\{\hat{\mathbf{a}}_k\}$ and $\{\hat{\mathbf{a}}'_k\}$ as

$$\hat{\mathbf{a}}_i = \sum_{j=1}^{n} \tau_i^j \, \hat{\mathbf{a}}'_j. \tag{8.4}$$

In other words, the vector-columns of T are given by the coordinates of the old basis vectors $\hat{\mathbf{a}}_i$ in the new basis $\{\hat{\mathbf{a}}'_j\}$. Now, consider the expansion (8.1) and the analogous one with components v'^j in the basis $\{\mathbf{a}'_k\}$. Since the two expansion describe the same invariant geometric object $\mathbf{v}$ we have

$$\mathbf{v} = \sum_{i=1}^{n} v'^j \mathbf{a}'_j = \sum_{i=1}^{n} v^i \mathbf{a}_i = \sum_{i,j=1}^{n} v^i \tau_i^j \mathbf{a}'_j \quad \Rightarrow \quad v'^j = \sum_{i=1}^{n} \tau_i^j v^i$$

so the parallel (also referred to as *contravariant*) components of the vector $\mathbf{v}$ are transformed namely with the transition matrix T, i.e.,

$$T: \quad v^i \;\to\; v'^j. \tag{8.5}$$

The normal projections (*covariant components*) v_i, on the other hand, are being transformed by the inverse transpose $(T^t)^{-1}$ usually referred to as the *contragradient* transformation of T, as one may easily deduce from (8.2) and (8.4). This crucial difference between upper and lower indices remains hidden in the case of orthonormal bases, in which the two types of coordinates coincide and the transition matrix T is clearly equal to its contragradient. Such transformations are called *orthogonal* and they play a central role in both geometry and mechanics. More precisely, orthogonal transformations defined via the condition $T^t = T^{-1}$ are compositions of reflections and preserve the scalar product, i.e., $\mathbf{u}\cdot\mathbf{v} = T\mathbf{u}\cdot T\mathbf{u}$, hence, it also preserves lengths and angles. Moreover, we obviously have $\det T = \pm 1$ due to the properties of $\det T^t$ and $\det T^{-1}$. As it turns out, the positive sign of the determinant corresponds to an orientation-preserving transformation, which in our case reduces to a pure rotation. For $\det T = -1$, on the other hand, we end up with a composition of a rotation and a mirror reflection. At the end of the present chapter we consider the more general setting of non-orthogonal bases and show how one may still derive the vector components of $\mathbf{v} \in \mathbb{R}^3$ as normal projections using the notion of the so-called *dual basis*.

8.2 Cross and Triple Products

Recall the usual definition of the dot product given in terms of matrix multiplication that extends to an arbitrary dimension as[1]

$$\mathbf{u}\cdot\mathbf{v} = U^tV = \sum_{k=1}^{n} u^k v_k = |\mathbf{u}||\mathbf{v}|\cos\measuredangle(\mathbf{u},\mathbf{v}) \tag{8.6}$$

and is obviously invariant under orthogonal transformations of the basis:

$$U'^tV' = (TU)^t(TV) = U^tT^tTV = U^tV, \qquad T^t = T^{-1}.$$

In Chapter 4 we also introduced the wedge product of two planar vectors as

$$\mathbf{u}\wedge\mathbf{v} = \begin{vmatrix} u_1 & u_2 \\ v_1 & v_2 \end{vmatrix} = |\mathbf{u}||\mathbf{v}|\sin\measuredangle(\mathbf{u},\mathbf{v}) \tag{8.7}$$

which is equal (up to a sign) to the area of the parallelogram determined by $\mathbf{u}$ and $\mathbf{v}$. This product extends naturally to an arbitrary finite dimension as

$$(\mathbf{u}\wedge\mathbf{v})_{ij} = u_iv_j - v_iu_j, \qquad i,j = 1,2,\dots n \tag{8.8}$$

which yields a skew-symmetric tensor with two indices and $k = \dfrac{n(n-1)}{2}$ independent components (explain why!). In particular, for $n = 2$ it is a scalar ($k = 1$) and $n = 3$ yields $k = 3$ as well, so we may associate to the exterior product a vector (sometimes also referred as *pseudovector*). Namely, its components are equal to the co-factors of the first row in the formal determinant written below ($\hat{\mathbf{e}}_1$, $\hat{\mathbf{e}}_2$ and $\hat{\mathbf{e}}_3$ denote the standard basis vectors)

$$\mathbf{u}\times\mathbf{v} = \begin{vmatrix} \hat{\mathbf{e}}_1 & \hat{\mathbf{e}}_2 & \hat{\mathbf{e}}_3 \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{vmatrix} = \begin{pmatrix} u_2v_3 - u_3v_2 \\ u_3v_1 - v_3u_1 \\ u_1v_2 - u_2v_1 \end{pmatrix}. \tag{8.9}$$

This type of product is called a *cross product* of the vectors $\mathbf{u}$ and $\mathbf{v}$ and naturally generalizes the wedge product in $\mathbb{R}^2$. Namely, if $\mathbf{u}$ and $\mathbf{v}$ belong to the x,y-plane, it is a vector, pointing in the z-direction, with magnitude exactly $|\mathbf{u}\wedge\mathbf{v}|$. Actually, for arbitrary $\mathbf{u}$ and $\mathbf{v}$ it is a vector, perpendicular to the plane $\mathrm{Span}\{\mathbf{u},\mathbf{v}\}$, such that $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{u}\times\mathbf{v}$ form a right oriented triple (counterclockwise rotation or the "right hand rule" illustrated in Figure 8.1) and its magnitude is $|\mathbf{u}\times\mathbf{v}| = |\mathbf{u}||\mathbf{v}|\sin\measuredangle(\mathbf{u},\mathbf{v})$. The product is obviously skew-symmetric by construction (why?) and may be used for generating bases in $\mathbb{R}^3$. In particular, if $\mathbf{u}\perp\mathbf{v}$, the corresponding basis is orthogonal. We note that, although the wedge product extends to arbitrary dimensions, the cross product exists only in $\mathbb{R}^3$ where one has the coincidence $k = n$.

[1] the infinite-dimensional extension for *Hilbert spaces* is given by means of integration.

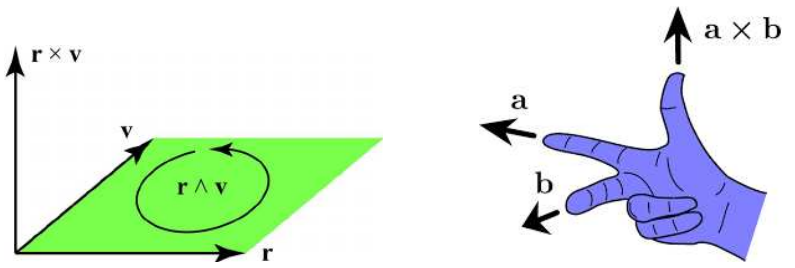

Figure 8.1: The cross product of two vectors (on the left) and the right hand rule to determine the orientation (on the right).

Triple Products and Double Cross Products

Although the wedge product is associative, i.e., $\mathbf{u} \wedge (\mathbf{v} \wedge \mathbf{w}) = (\mathbf{u} \wedge \mathbf{v}) \wedge \mathbf{w}$, surprisingly this does not hold for the cross product. Instead, we have the two different expressions below for the so-called *double cross products*

$$\begin{aligned} (\mathbf{u} \times \mathbf{v}) \times \mathbf{w} &= (\mathbf{u}, \mathbf{w})\, \mathbf{v} - (\mathbf{v}, \mathbf{w})\, \mathbf{u} \\ \mathbf{u} \times (\mathbf{v} \times \mathbf{w}) &= (\mathbf{u}, \mathbf{w})\, \mathbf{v} - (\mathbf{u}, \mathbf{v})\, \mathbf{w}. \end{aligned} \tag{8.10}$$

Note that in each case the double cross product lies in the plane determined by the two vectors in the brackets (explain why and prove the equalities!).

There is one more generalization of the wedge product in $\mathbb{R}^2$. This time we put the emphasis on the fact that it yields the oriented area determined by its two factors. Then, in $\mathbb{R}^3$ we take three, rather than two vectors, and express the oriented volume of the prism they span in a similar way, namely

$$(\mathbf{u}, \mathbf{v}, \mathbf{w}) = \begin{vmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_2 \\ w_1 & w_2 & w_3 \end{vmatrix} \tag{8.11}$$

usually referred to as the *triple product* of $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$. Note that we have

$$(\mathbf{u}, \mathbf{v}, \mathbf{w}) = (\mathbf{u} \times \mathbf{v}, \mathbf{w}) = (\mathbf{u}, \mathbf{v} \times \mathbf{w}) = \mathbf{u} \wedge \mathbf{v} \wedge \mathbf{w}$$

which follows straight from the definitions of the cross, dot and wedge products. For its interpretation as a volume, however, we first need to express

$$\mathbf{u} \times \mathbf{v} = |\mathbf{u}||\mathbf{v}| \sin \measuredangle(\mathbf{u}, \mathbf{v})\mathbf{n} = S_B \mathbf{n}$$

where $\mathbf{n}$ is a unit vector, normal to both $\mathbf{u}$ and $\mathbf{v}$, and S_B denotes the oriented area of the base, which lies in the $\mathbf{u}$, $\mathbf{v}$-plane. Then, we take into account that $\mathbf{w}\cdot\mathbf{n} = |\mathbf{w}|\cos\measuredangle(\mathbf{n},\mathbf{w})$ that is precisely the height of the prism h, so the triple product finally yields (see also Figure 8.2)

$$(\mathbf{u},\mathbf{v},\mathbf{w}) = |\mathbf{u}||\mathbf{v}||\mathbf{w}|\sin\measuredangle(\mathbf{u},\mathbf{v})\cos\measuredangle(\mathbf{n},\mathbf{w}) = S_B h = V.$$

One consequence is that coplanar vectors have vanishing triple product. Here is also a working formula for one particular type type of triple products:

$$(\mathbf{u}\times\mathbf{v},\mathbf{w}\times\mathbf{z}) = (\mathbf{u},\mathbf{w})(\mathbf{v},\mathbf{z}) - (\mathbf{u},\mathbf{z})(\mathbf{v},\mathbf{w}). \tag{8.12}$$

We leave the proof to the curious reader (see the exercises at the end).

8.3 Dual Bases

There is one peculiar application of the cross and triple product that gives a powerful technique in the geometry of non-orthogonal bases. It resorts on the fact that the parallel projections v^k of a vector $\mathbf{v}$ in the basis $\{\mathbf{a}_k\}$ may be given as orthogonal projections in another basis $\{\mathbf{a}^k\}$ we refer to as *dual* to $\{\mathbf{a}_k\}$. In order to obtain this duality let us first note that since

$$\mathbf{v} = \sum_{i=1}^{n} v^i \mathbf{a}_i, \qquad v^j = \mathbf{v}\cdot\mathbf{a}^j \tag{8.13}$$

it is quite clear that for the two bases we are going to have (see Figure 8.2)

$$\mathbf{a}^i\cdot\mathbf{a}_j = \delta^i_j. \tag{8.14}$$

This is quite easy to visualize in the plane: the vector $\mathbf{a}^1$ needs to be perpendicular to $\mathbf{a}_2$ and its dot product with $\mathbf{a}_1$ needs to be equal to one. The former is straightforward as each nonzero vector in $\mathbb{R}^2$ determines its normal direction uniquely, while for the latter we need a normalization factor of $\mathbf{a}_1\wedge\mathbf{a}_2 = \mathbf{a}_1\cdot\mathbf{a}_2^{\perp}$. With a similar argument for $\mathbf{a}^2$ we finally obtain

$$\mathbf{a}^1 = \frac{\mathbf{a}_2^{\perp}}{\mathbf{a}_1\wedge\mathbf{a}_2}, \qquad \mathbf{a}^2 = -\frac{\mathbf{a}_1^{\perp}}{\mathbf{a}_1\wedge\mathbf{a}_2} \tag{8.15}$$

where we have chosen the clockwise rotation from $\mathbf{v}$ to $\mathbf{v}^{\perp} = (v_2, -v_1)^t$ and the minus for the second equality is necessary to compensate the reverse orientation with this choice. For the case $n = 3$ the construction is quite

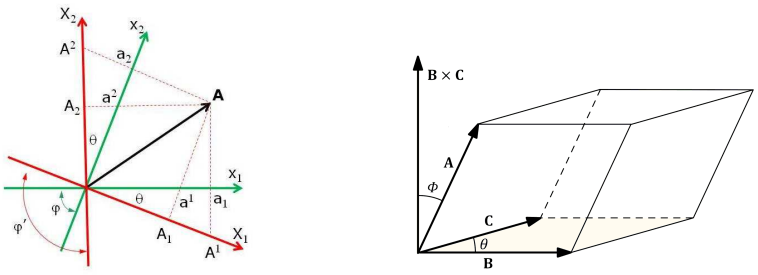

Figure 8.2: Dual bases (on the left) and the triple product (on the right).

similar, but this time formula (8.14) implies that each vector in $\{\mathbf{a}^k\}$ is normal to two of the directions determined by the initial basis $\{\mathbf{a}_k\}$. Hopefully, we can make use of the cross product to determine for example $\mathbf{a}^3 \parallel \mathbf{a}_2 \times \mathbf{a}_3$. However, we are going to need normalization factors here too, and it is not difficult to see that the triple product, i.e., the volume element determined by the basis $\{\mathbf{a}_k\}$, plays that role in $\mathbb{R}^3$. More precisely, we may define

$$\mathbf{a}^1 = \frac{\mathbf{a}_2 \times \mathbf{a}_3}{(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3)}, \qquad \mathbf{a}^2 = \frac{\mathbf{a}_3 \times \mathbf{a}_1}{(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3)}, \qquad \mathbf{a}^3 = \frac{\mathbf{a}_1 \times \mathbf{a}_2}{(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3)}. \tag{8.16}$$

Now, it is straightforward to see that with this choice the orthonormality conditions (8.14) are satisfied and thus, the parallel (contravariant) components in the basis $\{\mathbf{a}_k\}$ are precisely the normal (covariant) components in the dual basis $\{\mathbf{a}^k\}$. Note also that the concept of duality is always mutual, i.e., if $\{\mathbf{a}^k\}$ is dual to $\{\mathbf{a}_k\}$, then the inverse is also true (verify this in $\mathbb{R}^3$). The dual basis technique naturally extends to higher dimensions as well, but then one does not have the convenience of using the cross product. Try to formulate an analogous construction for $\mathbb{R}^n$ based on the exterior product.

Problem Session

1. Let $\{\hat{\mathbf{e}}_k\}$ be the standard basis and $T : \{\hat{\mathbf{e}}_k\} \to \{\mathbf{a}_k\}$ the transition matrix to the (not necessarily orthogonal) basis $\{\mathbf{a}_k\}$. Show that one has $G^{-1} = TT^t$ where G is the Gram matrix associated with $\{\mathbf{a}_k\}$.

2. Prove that the cross product satisfies the Jacobi identities

$$(\mathbf{u} \times \mathbf{v}) \times \mathbf{w} + (\mathbf{v} \times \mathbf{w}) \times \mathbf{u} + (\mathbf{w} \times \mathbf{u}) \times \mathbf{v} = 0$$

$$\mathbf{u} \times (\mathbf{v} \times \mathbf{w}) + \mathbf{v} \times (\mathbf{w} \times \mathbf{u}) + \mathbf{w} \times (\mathbf{u} \times \mathbf{v}) = 0$$

3. Prove the particular case of formula (8.12)

$$(\mathbf{u}\times\mathbf{v},\mathbf{u}\times\mathbf{v})=\mathbf{u}^2\mathbf{v}^2-(\mathbf{u},\mathbf{v})^2$$

Can you prove the generic case?
Hint: Show by the properties of determinants that

$$(\mathbf{a},\mathbf{b}\times\mathbf{c})=(\mathbf{a}\times\mathbf{b},\mathbf{c})$$

and for the generic case use the equalities for double cross products.

4. Determine $\tan\measuredangle(\mathbf{u},\mathbf{v})$ and the volume spanned by $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{u}\times\mathbf{v}$, where $\mathbf{u}=\begin{pmatrix}1\\1\\1\end{pmatrix}$ and $\mathbf{v}=\begin{pmatrix}1\\-1\\0\end{pmatrix}$.

5. Calculate the volume of the tetrahedron $ABCD$ with vertices $A(1;2;1)$, $B(-1;0;3)$, $C(2;0;0)$ and $D(3;1;0)$. Find also the coordinates of the geometric center, the total surface area and the distance from the point D to the plane determined by A, B and C. Did you have fun?

6. Given the two planar vectors $\mathbf{a}_1=(1,\,-1)^t$ and $\mathbf{a}_2=(1,\,1)^t$, obtain the contragradient matrix transformation associated with it and the dual basis $\{\mathbf{a}^k\}$. Resize $\mathbf{a}_{1,2}$ to construct an orthonormal basis $\{\hat{\mathbf{a}}_k\}$. What geometric transformation does the orthogonal map $T:\{\hat{\mathbf{e}}_k\}\to\{\hat{\mathbf{a}}_k\}$ correspond to? Can you extend this concept to obtain a one-parameter set of orhonormal bases $\{\hat{\mathbf{a}}_k\}_t$? How about a two-parameter set?

7. Given the three non-coplanar vectors $\mathbf{a}_1=(0,\,1,\,1)^t$, $\mathbf{a}_2=(1,\,0,\,1)^t$ and $\mathbf{a}_3=(1,\,1,\,0)^t$, find the matrix transformation that yields the contravariant components of an arbitrary vector $\mathbf{v}\in\mathbb{R}^3$ in the basis $\mathbf{a}_k$ expressed in terms of the ones in the standard basis $\hat{\mathbf{e}}_k$. Then, obtain the dual basis $\{\mathbf{a}^k\}$ and see that the latter coincide with the covariant components of $\mathbf{v}$ in it. Try to explain this coincidence algebraically.

Chapter 9

Lines and Planes in $\mathbb{R}^3$

Most of the constructions in the plane have natural generalization to $\mathbb{R}^3$. However, one should be careful, since there are differences too. To begin with, we may still consider two types of vectors - attached and free, typically represented by radius-vectors of points and their differences. The latter constitute the vector space $\mathbb{R}^3$, while the former, the *affine* space $\mathbb{R}^3$ as a space of points. Thus, one may think that there are two copies of $\mathbb{R}^3$ involved in our geometric considerations. When lines and planes come into play, more complicated spaces appear naturally in the model, such as Grassmannian and projective spaces, but we are not going to discuss them here explicitly.

9.1 Equations of Lines and Planes

As we already discussed, most of the constructions are not affected by increasing the dimension. For example, the vector parametric equation of a line in $\mathbb{R}^3$ is written in the exact same way as in $\mathbb{R}^2$ (See Figure 9.1), i.e.,

$$g: \ \mathbf{r} = \mathbf{r}_0 + \lambda \mathbf{t}, \qquad \lambda \in \mathbb{R} \tag{9.1}$$

but this time the scalar equations are three, since there is also a z-coordinate

$$\begin{aligned} x &= x_0 + \lambda t_1 \\ y &= y_0 + \lambda t_2 \\ z &= z_0 + \lambda t_3 \end{aligned} \tag{9.2}$$

while both the derivation of the above formula and the interpretation of $\mathbf{t}$ as a direction vector along the line remain the same. Similarly, one may write

the equation of a line passing through a pair of points with radius-vectors $\mathbf{r}_0$ and $\mathbf{r}_1$, respectively as

$$\frac{x - x_0}{x_1 - x_0} = \frac{y - y_0}{y_1 - y_0} = \frac{z - z_0}{z_1 - z_0} \tag{9.3}$$

where the difference seems to be only in the presence of a third coordinate. Note, however, that here, unlike in the two-dimensional case we have two, rather than one linear equation. The reason for this is actually quite simple: in order to obtain a one-dimensional object in three-dimensional space, such as the line, described by the free parameter λ in formula (9.1), we need to impose two independent relations for the coordinates $3 - 2 = 1$, while in the plane only one equation is sufficient since $2 - 1 = 1$. So what is the object

$$\alpha : \; ax + by + cz + d = 0 \tag{9.4}$$

described by a single linear equation for x, y and z? Well, it is obviously linear and has $3 - 1 = 2$ independent parameters, so it must be a two-dimensional plane. The above is actually referred to as the *general equation of a plane* and its geometric interpretation is quite similar to that of a general equation of a line in $\mathbb{R}^2$ (See Figure 9.1). Namely, the coefficients a, b and c can be thought of as the components of the normal vector $\mathbf{n}$ (note that each plane in $\mathbb{R}^3$ determines a single normal direction, while a line has infinitely many) and $d = -\mathbf{n} \cdot \mathbf{r}_0$, where $\mathbf{r}_0$ is the radius-vector of an arbitrary point $A \in \alpha$. This is quite easy to obtain when we take into account that the radius-vector $\mathbf{r}$ of a any point in α defines a free vector $\mathbf{r} - \mathbf{r}_0$, which is always parallel to α and thus $\mathbf{n} \cdot (\mathbf{r} - \mathbf{r}_0) = 0$. Similarly, one may consider a unit normal vector $\mathbf{n}_0 = \dfrac{\mathbf{n}}{|\mathbf{n}|}$ and write the above equation in the form

$$\mathbf{n}_0 \cdot (\mathbf{r} - \mathbf{r}_0) = 0 \tag{9.5}$$

which is usually called *normal equation of a plane* and may be used to measure distances between points and planes. Consider a point B with radius-vector $\mathbf{r}_1$ and project the difference vector $\vec{AB} = \mathbf{r}_1 - \mathbf{r}_0$ pointing from A to B along the normal direction by means of a dot product with $\mathbf{n}_0$. This clearly yields the (shortest) distance between B and α in the form

$$\mathrm{d}(B, \alpha) = |(\mathbf{r}_1 - \mathbf{r}_0) \cdot \mathbf{n}_0| \tag{9.6}$$

or, in other words, we simply substitute the coordinates of the point into the normal equation of the plane and, due to the ambiguous orientation

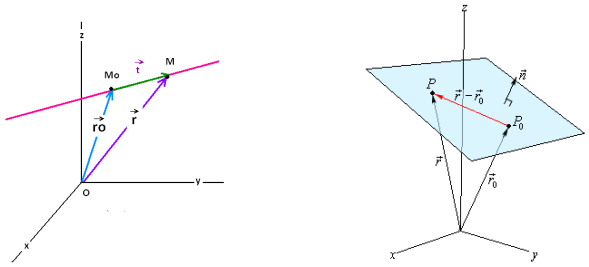

Figure 9.1: Lines and planes in three-dimensional space.

of $\mathbf{n}_0$, take the absolute value of the result. Similarly, if we generalize the intercept-intercept equation of a line (5.11) to

$$\frac{x}{a'} + \frac{y}{b'} + \frac{z}{c'} = 1 \tag{9.7}$$

the latter already represents a plane in $\mathbb{R}^3$ and the constants a', b' and c' yield the points at which this plane intersects the coordinate axes Ox, Oy and Oz, respectively. More generally, one may define a plane passing through three non-collinear points $A(x_0, y_0, z_0)$, $B(x_1, y_1, z_1)$ and $C(x_2, y_2, z_2)$ demanding for any point $F \in \alpha$ with radius-vector $\mathbf{r}$, the three vectors $\vec{AB} = \mathbf{r}_1 - \mathbf{r}_0$, $\vec{AC} = \mathbf{r}_2 - \mathbf{r}_0$ and $\vec{AF} = \mathbf{r} - \mathbf{r}_0$ to be coplanar, which may be written as

$$\alpha : \ (\vec{AX}, \vec{AB}, \vec{AC}) = \begin{vmatrix} x - x_0 & y - y_0 & z - z_0 \\ x_1 - x_0 & y_1 - y_0 & z_1 - z_0 \\ x_2 - x_0 & y_2 - y_0 & z_2 - z_0 \end{vmatrix} = 0. \tag{9.8}$$

Expanding the above determinant over the first row, we obtain the coefficients in the general equation of the plane (9.4) as co-factors[1].

9.2 Intersections and Mutual Positions

The excessive dimensions (degrees of freedom) naturally make the picture more complex, in the sense that they offer a richer variety of possible situations to consider. Things remain relatively simple as long as there are only points and vectors to consider. For example, the distance between two points $A(x_0, y_0, z_0)$ and $B(x_1, y_1, z_1)$ is given by the three-dimensional

[1] note that the same result is obtained if we construct the normal vector as $\mathbf{n} = \vec{AB} \times \vec{AC}$.

Pythagoras' theorem in the form

$$|AB| = |\mathbf{r}_1 - \mathbf{r}_0| = \sqrt{(x_1 - x_0)^2 + (y_1 - y_0)^2 + (z_1 - z_0)^2}. \tag{9.9}$$

Moreover, the formulae (9.3) and (9.8) may be used as criteria for the collinearity of three points and the coplanarity of four points, respectively. Then, we also have incidence relations, e.g. a point belongs to a line or a plane if its coordinates satisfy the corresponding equation or set of equations. As for the parametric case (9.1), the system for λ is over-determined and fixes a single value of the parameter for any particular $B \in g$. If, on the other hand, a point B does not belong to a plane α, the distance between the two is given by formula (9.9), but a similar situation with a line g involved needs further study, since the latter does not determine a unique normal direction. However, we may construct a plane β, normal to g through B, using $\mathbf{t}$ as a normal vector in the general equation. Then, the distance between B and g is equal to the distance between B and the intersection point $C = g \cap \beta$, which is easy to obtain by substituting the parametric equations (9.1) into the generic equation of β and thus determining the value of λ. Finally, $\mathrm{d}(B, g) = |BC|$ is obtained via formula (9.9).

Let us now consider the mutual position of two lines: g, determined by its direction vector $\mathbf{t}$ and a point A, and h - by the pair $\mathbf{s}$ and B, respectively. The shortest distance is measured along the common normal direction determined by the cross product $\mathbf{t} \times \mathbf{s}$ as long as $\mathbf{t}$ and $\mathbf{s}$ are not parallel. Then, each vector connecting a point from g with a point from h (and in particular, the vector $\vec{AB}$) projected over this direction yields the distance

$$\mathrm{d}(g, h) = \frac{(\vec{AB}, \mathbf{t} \times \mathbf{s})}{|\mathbf{t} \times \mathbf{s}|}, \qquad \mathbf{t} \times \mathbf{s} \neq 0. \tag{9.10}$$

If, on the other hand, the two lines are parallel and thus $\mathbf{t} \times \mathbf{s} = 0$, we may construct the normal direction by means of a double cross product. First, note that $\mathbf{t} \times \vec{AB}$ is a vector normal to the plane, determined by g and h, thus $(\mathbf{t} \times \vec{AB}) \times \mathbf{t} = \mathbf{t}^2 \vec{AB} - (\vec{AB}, \mathbf{t})\mathbf{t}$ lies in this plane and is normal to $\mathbf{t} \parallel \mathbf{s}$ by construction. Projecting the vector $\vec{AB}$ over its direction as before yields

$$\mathrm{d}(g, h) = \frac{(\mathbf{t} \times \vec{AB})^2}{|(\mathbf{t} \times \vec{AB}) \times \mathbf{t}|} = |AB| \sin \measuredangle(\vec{AB}, \mathbf{t}), \qquad \mathbf{t} \times \mathbf{s} = 0. \tag{9.11}$$

As far as planes are concerned, a line and are plane are either parallel, or intersect (the former penetrates the latter) at a point with coordinates obtained by substituting the parametric equations of the line in the general

equation of the plane and thus fixing the value of the parameter. Similarly, two non-parallel planes intersect at a line, which is the one-parameter solution to the system of two equations for three variable. If a line or a plane α is parallel to another plane β, the distance between the two objects is equal for any pair of point $A \in \alpha$ and $B \in \beta$. In particular, this distance may be zero if the two planes coincide or the line belongs to the plane, respectively.

Finally, let us consider the case of three planes in $\mathbb{R}^3$ from the perspective of systems of linear equations and the Kronecker-Capelli theorem in particular. In the non-degenerate case, in which the system of the three normals has maximal rank, the planes are bound to intersect at a single point much like the coordinate planes OXY, OYZ and OZX intersect at the origin. The other extreme setting involves parallelism of two or even three normals (corresponding to rank two and rank one, respectively). In between, however, we see something with no analogue in $\mathbb{R}^2$: neither two of the normals are parallel and yet, the system is either dependant of inconsistent, i.e., the planes may be spread like a fan with a common axis or set like the roof of a house, with three parallel intersections. Consider other possible settings.

9.3 Basic Examples from Stereometry

Let us begin as we did in the two-dimensional case: with a set of points, e.g. $A(1;2;0)$, $B(2;1;3)$, $C(3;0;1)$ and $D(1;-1;2)$. As a first task, we may obtain all direction vectors starting from the vertex A given by coordinate differences as $\vec{AB} = (1, -1, 3)^t$, $\vec{AC} = (2, -2, 1)^t$ and $\vec{AD} = (0, -3, 2)^t$. Now, the rank of the system the above vectors constitute determines whether the four points are coplanar or in a generic position. It is easy to see that

$$(\vec{AB}, \vec{AC}, \vec{AD}) = \begin{vmatrix} 1 & -1 & 3 \\ 2 & -2 & 1 \\ 0 & -3 & 2 \end{vmatrix} = -15$$

so the rank is maximal and the four points are not coplanar. Furthermore, the volume of the corresponding prism is equal to 15 measure units and hence, we may obtain the volume of the pyramid $ABCD$ as

$$V_{ABCD} = \frac{1}{6}|(\vec{AB}, \vec{AC}, \vec{AD})| = \frac{5}{2}.$$

Another typical task is to determine the distance between the point D and the plane α determined by A, B and C. We can do this in several ways: the

standard one is to express the equation of α using formula (9.8) namely as

$$\alpha: \quad x + y - 3 = 0$$

and then, dividing by the length of the normal vector $\mathbf{n}_\alpha = (1,\, 1,\, 0)^t$ and substituting the coordinates of D, obtain $d(D, \alpha) = \frac{1}{\sqrt{2}}|1 - 1 - 3| = \frac{3}{\sqrt{2}}$. Another way uses the fact that the above distance is equal to the height of the pyramid and thus may be determined by dividing its volume to one third of the area of $\triangle ABC$, which is equal to $\frac{1}{2}|\vec{AB} \times \vec{AC}| = \frac{5}{\sqrt{2}}$. Next, let us consider the bisectrix plane δ between α and $\beta \ni A, B, D$. To derive its equation, we first note that the normal $\mathbf{n}_\delta$ bisects the angle determined by the normals $\mathbf{n}_\alpha \sim \vec{AB} \times \vec{AC}$ and $\mathbf{n}_\beta \sim \vec{AB} \times \vec{AD} = (-7,\, 2,\, 3)^t$, so we may re-scale $\mathbf{n}_\alpha \to \mathbf{n}'_\alpha$ assuring $|\mathbf{n}'_\alpha| = |\mathbf{n}_\beta|$ and use vector summation to obtain

$$\mathbf{n}^\pm_\delta = \mathbf{n}_\beta \pm \mathbf{n}'_\alpha = (-7 \pm \sqrt{31},\, 2 \pm \sqrt{31},\, 3)^t$$

where the two solutions correspond to the two complementary angles determined by the intersection of the normal directions $\mathbf{n}_\alpha$ and $\mathbf{n}_\beta$. Finally, the free terms of the general equations of the two planes $\delta^\pm$ are given as $c^\pm = -\mathbf{n}^\pm_\delta \cdot \mathbf{r}_A = 3 \mp 3\sqrt{31}$, where $\mathbf{r}_A$ denotes the radius-vector of A, i.e.,

$$\delta^\pm: \qquad (-7 \pm \sqrt{31})x + (2 \pm \sqrt{31})y + 3z + 3 \mp 3\sqrt{31} = 0.$$

Another typical task in stereometry is to determine the mutual position and distance between two lines. Let us take for example the one passing through the pair of points A and B (we shall denote it with g) and another one (say, h) containing C and D. Firstly, we have the direction vectors $\vec{AB} = (1,\, -1,\, 3)^t$ and $\vec{DC} = (2,\, 1,\, -1)^t$, so using A and C for base points it is straightforward to write the parametric equations of the two lines as

$$g: \qquad x = 1 + \lambda, \quad y = 2 - \lambda, \quad z = 3\lambda, \qquad \lambda \in \mathbb{R}$$

$$h: \qquad x = 3 + 2\mu, \quad y = \mu, \quad z = 1 - \mu, \qquad \mu \in \mathbb{R}.$$

Note that equating x, y and z we obtain an inconsistent system for λ and μ as the two lines do not intersect. Since they are not parallel either, the common normal direction may be obtained as a cross product $\mathbf{n} \sim \vec{AB} \times \vec{DC}$. Then, normalization to unit length yields $\hat{\mathbf{n}} = \frac{1}{\sqrt{62}}(-2,\, 7,\, 3)^t$ and projecting the vector difference between any pair of points from the two lines (and in particular, A and C), we find the distance to be $d = \frac{15}{\sqrt{62}}$. It turns out to be useful also to obtain the equation of the unique line ℓ that intersects g and h and is perpendicular to both of them. Clearly, it is oriented along the

vector $\mathbf{n}$, but the free terms in the parametric equations are yet to be fixed. A straightforward, but rather technical way to do so is by demanding the systems $\ell \cap g$ and $\ell \cap h$ to have a solution. Another constructive approach is to obtain ℓ as an intersection of two planes: $\sigma \ni \mathbf{n}, g$ and $\psi \ni \mathbf{n}, h$. Their normals are then naturally given as double cross products

$$\mathbf{n}_\sigma \sim \vec{AB} \times (\vec{AB} \times \vec{DC}), \qquad \mathbf{n}_\psi \sim \vec{DC} \times (\vec{AB} \times \vec{DC})$$

and we readily express $\ell = \sigma \cap \psi$, where

$$\sigma : \quad 24x + 9y - 5z = 42, \qquad \psi : \quad 5x - 2y + 8z = 23.$$

Resolving the above system we finally obtain

$$\ell : \qquad x = -2\lambda, \qquad y = \frac{451}{62} + 7\lambda, \qquad z = \frac{291}{62} + 3\lambda, \qquad \lambda \in \mathbb{R}.$$

Note that we automatically obtained the direction vector $\mathbf{n} = (-2,\, 7,\, 3)^t$ as before and the distance $d(g, h) = |\vec{GH}|$ can be measured between the two intersection points $G = \ell \cap g$ and $H = \ell \cap H$. See if it yields the same result.

Problem Session

1. Let $A(-2; -1; 2)$, $B(0; 1; 2)$, $C(1; 2; 0)$ and $D(3; -1; 1)$ be four points in space. Find the coordinates of a point E, such that $ABCE$ is a parallelogram and calculate its area. Find the distance between B and the plane α determined by A, B and C.

2. Given the points $A(-4; -1; 2)$ and $B(3; 5; -16)$, find a third point C, such that the midpoint M of AC lies on the axis OY and the midpoint N of BC is in the plane XOZ.

3. Find a plane α that contains the point $A(3; 4; 0)$ and the line

$$g : \frac{x-2}{1} = \frac{y-3}{2} = \frac{z+1}{3}.$$

   Also, find the distance between A and g.

4. Find the vertices C and D of the square $ABCD$ if the other two are $A(1; 4; -3)$ and $B(1; 1; 1)$, and we know that $ABCD$ lies in a plane β, normal to $\alpha : x + 3y - 4z + 1 = 0$.

5. Given the points $A(0;-1;3)$, $B(2;1;-2)$, $C(0;3;1)$ and $D(-1;5;3)$ find the volume of the pyramid $ABCD$ and the cosine of $\measuredangle CAB$. Write the equation of a line g containing the vertex D and parallel to $\vec{AB}$.

6. Given the set of points $A(3;-1;2)$, $B(2;1;2)$, $C(0;2;0)$ and $D(1;1;1)$ find the geometric center of the pyramid $ABCD$, the area of $\triangle ABC$ and the coordinates of the orthogonal projection of the vertex D in the plane α determined by A, B and C.

7. Given the set of points $A(2;-1;2)$, $B(-2;1;2)$, $C(3;-2;3)$ and $D(1;0;1)$ find the area of $\triangle BCD$, the cosine of $\measuredangle BAD$ and the equation of the plane α containing $\vec{AB}$ and parallel to $\vec{CD}$.

Chapter 10

Linear Operators

10.1 The Rank and Nullity Theorem

In this chapter we consider linear operators $\mathcal{A} : V \to W$ from one vector space V to another W, i.e., maps satisfying the linear property

$$\mathcal{A}(\mathbf{u} + \lambda\mathbf{v}) = \mathcal{A}\mathbf{u} + \lambda\mathcal{A}\mathbf{v} \qquad \forall\, \mathbf{u}, \mathbf{v} \in V, \quad \lambda \in \mathbb{R}.$$

Note that if we have fixed bases in both V and W, then the operator $\mathcal{A}$ may be given as a $m \times n$ matrix, where $n = \dim V$ and $m = \dim W$. In the case $n = m$ and in particular, if $W \equiv V$, the corresponding matrix is square. For $m < n$, on the other hand, one clearly encounters a loss of information[1], whereas in the case $n < m$ there is not enough data in V to cover W in an invertible way, e.g. $V \subset W$. For a better understanding of the geometric properties of operators it is convenient to introduce the notions of *range* and *kernel*. More precisely, the former one, denoted here by $\mathrm{rg}\,\mathcal{A}$, is defined as the set of vectors in $\mathbf{w} \in W$, which are an image of some vector in V, i.e.,

$$\mathrm{rg}\,\mathcal{A}_{V\to W} = \{\mathbf{w} \in W |\ \exists\, \mathbf{v} \in V,\ \mathbf{w} = \mathcal{A}\mathbf{v}\}$$

and if the range of $\mathcal{A}$ covers all W we say that $\mathcal{A}$ is *onto*, or *surjective* (e.g. projection onto a lower-dimensional subspace). Similarly, the kernel is defined as the set of vectors in V mapped by $\mathcal{A}$ to the zero vector in W:

$$\ker \mathcal{A}_{V\to W} = \{\mathbf{v} \in V |\ \mathcal{A}\mathbf{v} = 0\}.$$

Obviously, due to linearity, $\ker \mathcal{A}$ always contains the zero in V and is said to be *trivial* if it does not contain anything else. In this case $\mathcal{A}$ is called

[1] projecting spacial vectors onto the XY-plane we lose track of the vertical component.

injective (for instance, the embedding $V \subset W$ is clearly injective). Now, if an operator is both injective and surjective, it is referred to as *bijective* and yields a one-to-one correspondence between V and W, which are said to be *isomorphic* (as vector spaces) that is usually denoted as $V \cong W$. One obvious isomorphism is given by the identity map $\mathcal{I}: \ \mathbf{v} \to \mathbf{v}, \ \forall\, \mathbf{v} \in V$. A non-trivial example would be a reflection or a rotation in the plane - it covers all $\mathbb{R}^2$ and the image of each non-zero vectors is also a non-zero vector. More generally, as we shall see later, each square matrix with non-vanishing determinant may be thought of as a matrix of a linear isomorphism. Before we come to this conclusion, however, it is useful to prove the following

Theorem 10 *If $\mathcal{A}: V \to W$ is linear, both $\operatorname{rg}\mathcal{A}$ and $\ker\mathcal{A}$ are vector spaces.*

Proof: Let $\mathbf{u}, \mathbf{v} \in \ker\mathcal{A}$. Then linearity yields $\mathcal{A}(\mathbf{u}+\lambda\mathbf{v}) = \mathcal{A}\mathbf{u}+\lambda\mathcal{A}\mathbf{v} = 0$ and thus, $\mathbf{u} + \lambda\mathbf{v} \in \ker\mathcal{A}$, which proves that $\ker\mathcal{A} \subset V$ is a vector space. A similar argument shows that linear combinations of images (under $\mathcal{A}$) are images of linear combinations, so $\operatorname{rg}\mathcal{A} \subset W$ is also a vector space. □

As vector spaces, $\operatorname{rg}\mathcal{A}$ and $\ker\mathcal{A}$ have their dimensions, referred to as *rank* and *nullity* of the operator $\mathcal{A}$, respectively:

$$\operatorname{rk}\mathcal{A} = \dim\left(\operatorname{rg}\mathcal{A}\right), \qquad \operatorname{nul}\mathcal{A} = \dim\left(\ker\mathcal{A}\right).$$

The rank of $\mathcal{A}$ is easily seen to be equal to the rank of the corresponding matrix (it is clearly independent on the choice of basis), while the geometric meaning of $\operatorname{nul}\mathcal{A}$ is best illustrated by the so-called rank and nullity theorem:

Theorem 11 *Let $\mathcal{A}: V \to W$ be a linear operator. Then, one has*

$$\operatorname{rk}\mathcal{A} + \operatorname{nul}\mathcal{A} = \dim V.$$

Proof: We only give a sketch of the proof here. Since each vector in V clearly either belongs to $\ker\mathcal{A}$, or is mapped into a non-zero vector in W, we have the direct sum decomposition $V = V_0 \oplus V_1$, i.e., $V_0 \cap V_1 = \{0\}$, where $V_0 = \ker\mathcal{A}$ and V_1 is in one-to one correspondence with $\operatorname{rg}\mathcal{A} \subset W$ since the restriction of $\mathcal{A}$ on V_1 has maximal rank (equal to $\dim V_1$). Thus, it yields an isomorphism between V_1 and W, which means in particular that $\dim V_1 = \dim W$, so we have $\dim V = \dim V_0 + \dim V_1 = \operatorname{rk}\mathcal{A} + \operatorname{nul}\mathcal{A}$. □

For linear systems the above result asserts that the dimension of the resolution set W equals the number of variables minus the number of independent equations. In the next section we provide more explicit geometric examples.

10.2 Projections, Reflections, Euclidean Motions

There are several classes of linear operators which are particularly interesting for the applications, e.g. the (skew-)symmetric type, the orthogonal and unitary types, the uni-modular type (satisfying $\det \mathcal{A} = \pm 1$) etc. Since this course is focused mainly on Euclidean geometry, we shall pay special attention to operators related to planar and spacial motions. There is a beautiful theorem due to Cartan–Dieudonné asserting that all such transformations are generated by reflections and the latter, as we shall see in a while, are most naturally expressed in terms of projectors, which is where we begin.

Projectors and Mirror Reflections

Recall that the projection of any vector $\mathbf{v}$ in a given direction is given by the dot product with the unit vector $\hat{\mathbf{u}}$ along this direction. Then, one may easily decompose the vector $\mathbf{v}$ into a parallel and perpendicular to $\hat{\mathbf{u}}$ components denoted respectively as $\mathbf{v}_{\|}$ and $\mathbf{v}_{\perp}$, given explicitly in the form

$$\mathbf{v}_{\|} = (\mathbf{v} \cdot \hat{\mathbf{u}})\, \hat{\mathbf{u}}, \qquad \mathbf{v}_{\perp} = \mathbf{v} - \mathbf{v}^{\|}.$$

The above are clearly linear functions of $\mathbf{v}$, so let us find their associated matrices - we only need an expression for the (parallel) projector $\mathcal{P}^{\|}_{\hat{\mathbf{u}}}$ along the direction determined by $\hat{\mathbf{u}}$. It is not hard to see that it has the form

$$\mathcal{P}^{\|}_{\hat{\mathbf{u}}} = \hat{\mathbf{u}}\hat{\mathbf{u}}^t \quad \Rightarrow \quad \mathcal{P}^{\perp}_{\hat{\mathbf{u}}} = \mathcal{I} - \hat{\mathbf{u}}\hat{\mathbf{u}}^t \tag{10.1}$$

where we make use of the so-called *tensor* or *dyadic* product of vectors given in components as $(\hat{\mathbf{u}}\hat{\mathbf{u}}^t)^i_j = \hat{\mathbf{u}}^i \hat{\mathbf{u}}_j$ and the fact that the two projectors add up to identity. Let us consider a particular example: the vector $\mathbf{u} = (3,\, 4)^t$ yields the corresponding projectors in the form[2]

$$\mathcal{P}^{\|}_{\mathbf{u}} = \frac{\mathbf{u}\mathbf{u}^t}{\mathbf{u}^2} = \frac{1}{25}\begin{pmatrix} 9 & 12 \\ 12 & 16 \end{pmatrix}, \qquad \mathcal{P}^{\perp}_{\mathbf{u}} = \mathcal{I} - \mathcal{P}^{\|}_{\mathbf{u}} = \frac{1}{25}\begin{pmatrix} 16 & -12 \\ -12 & 9 \end{pmatrix}.$$

Then, an arbitrary vector in the plane, say $\mathbf{v} = (1,\, 2)^t$, may easily be decomposed into a parallel and a normal with respect to $\mathbf{u}$ directions, namely

$$\mathbf{v} = \mathbf{v}_{\|} + \mathbf{v}_{\perp} = \mathcal{P}^{\|}_{\mathbf{u}}\mathbf{v} + \mathcal{P}^{\|}_{\mathbf{u}}\mathbf{v}$$

where it is straightforward to see by the rules of matric multiplication that

$$\mathbf{v}_{\|} = \frac{11}{25}\,(3,\, 4)^t, \qquad \mathbf{v}_{\perp} = \frac{2}{25}\,(-4,\, 3)^t.$$

[2] since $\mathbf{u}$ is not unit we divide by its squared norm.

Our next task is to construct the mirror reflection $\mathcal{M}_{\mathbf{u}}$ with respect to the direction determined by $\mathbf{u}$. In order to do this we only need to note that the parallel to $\mathbf{u}$ is preserved by $\mathcal{M}_{\mathbf{u}}$, while the one in the normal direction changes its sign, i.e.,

$$\mathcal{M}_{\mathbf{u}}\mathbf{v} = \mathbf{v}_{\parallel} - \mathbf{v}_{\perp} = \left(\mathcal{P}^{\parallel}_{\mathbf{u}} - \mathcal{P}^{\perp}_{\mathbf{u}}\right)\mathbf{v}$$

and we may express the matrix of the transformation as

$$\mathcal{M}_{\mathbf{u}} = 2\mathcal{P}^{\parallel}_{\mathbf{u}} - \mathcal{I} = \mathcal{I} - 2\mathcal{P}^{\perp}_{\mathbf{u}}. \tag{10.2}$$

In the above example this clearly yields

$$\mathcal{M}_{\mathbf{u}} = \frac{1}{25}\begin{pmatrix} -7 & 24 \\ 24 & 7 \end{pmatrix} \quad \Rightarrow \quad \mathcal{M}_{\mathbf{u}}\mathbf{v} = \frac{1}{24}\,(41,\, 38)^t.$$

To extend this construction to $\mathbb{R}^3$ one may either do it in a straightforward manner or consider reflections about planes and surfaces (see Figure 10.1). To see the difference more clearly, let us take for example the plane

$$\alpha: \quad 2x - y + 2z + 1 = 0 \quad \Rightarrow \quad \mathbf{n} = (2,\, -1,\, 2)^t.$$

One may easily construct the parallel (and hence, the normal) projector onto the direction determined by the normal vector $\mathbf{n}$ as

$$\mathcal{P}^{\parallel}_{\mathbf{n}} = \frac{1}{9}\begin{pmatrix} 4 & -2 & 4 \\ -2 & 1 & -2 \\ 4 & -2 & 4 \end{pmatrix}, \qquad \mathcal{P}^{\perp}_{\mathbf{n}} = \mathcal{I} - \mathcal{P}^{\parallel}_{\mathbf{n}} = \frac{1}{9}\begin{pmatrix} 5 & 2 & -4 \\ 2 & 8 & 2 \\ -4 & 2 & 5 \end{pmatrix}.$$

Now, the axial symmetry with respect to $\mathbf{n}$ in this case is nothing but a half-turn (rotation by an angle π) about $\mathbf{n}$ and it is given explicitly as

$$\mathcal{M}_{\mathbf{n}} = \mathcal{P}^{\parallel}_{\mathbf{n}} - \mathcal{P}^{\perp}_{\mathbf{n}} = 2\mathcal{P}^{\parallel}_{\mathbf{n}} - \mathcal{I} = \mathcal{I} - 2\mathcal{P}^{\perp}_{\mathbf{n}}.$$

On the other hand, spacial reflections are usually taken with respect to a plane, or some more general (mirror) surface. It is not difficult to realize that the parallel projector with respect to $\mathbf{n}$ is the normal one with respect to the plane α and vice versa, hence the two reflections differ only in sign:

$$\mathcal{P}^{\parallel}_{\alpha} = \mathcal{P}^{\perp}_{\mathbf{n}}, \quad \mathcal{P}^{\perp}_{\alpha} = \mathcal{P}^{\parallel}_{\mathbf{n}} \quad \Rightarrow \quad \mathcal{M}_{\alpha} = -\mathcal{M}_{\mathbf{n}} = \mathcal{I} - 2\mathcal{P}^{\parallel}_{\mathbf{n}}. \tag{10.3}$$

Using the above calculations we easily derive the reflector

$$\mathcal{M}_{\alpha} = \frac{1}{9}\begin{pmatrix} 1 & 4 & -8 \\ 4 & 7 & 4 \\ -8 & 4 & 1 \end{pmatrix}$$

and use it to obtain the mirror images of both vectors and points in $\mathbb{R}^3$. Take for instance $A(3;2;1)$ with radius-vector $\mathbf{r}$ and let us find the coordinates of the reflected point $A'(x;y;z)$. One is tempted to write $\mathbf{r}' = \mathcal{M}_\alpha \mathbf{r}$, but that would be wrong as long as α does not contain the origin. Actually, our construction can be used directly only for free vectors, not for radius-vectors (explain why) and we already know a straightforward way to obtain them in terms of coordinate differences. Take an arbitrary point on the plane, say $A_0(1;1;-1) \in \alpha$ and then reflect $\mathbf{r}-\mathbf{r}_0$ so that it is mapped to $\mathbf{r}'-\mathbf{r}_0$. Thus, the actual formula for the coordinates of the mirror image A' reads

$$\mathbf{r}' = \mathbf{r}_0 + \mathcal{M}_\alpha \left(\mathbf{r} - \mathbf{r}_0\right). \tag{10.4}$$

Applied to our example, this construction finally yields $A'\left(-\frac{1}{9}; \frac{32}{9}; -\frac{19}{9}\right)$.

There is another, rather physical, setting that involves reflection with respect to a vector, rather than a plane. Consider the Snell's law of geometric optics: the angle of reflection (with respect to the normal to the reflecting surface direction) is equal to the angle of incidence, i.e., $\mathbf{t}' = \mathcal{M}_{\mathbf{n}}\mathbf{t} = -\mathcal{M}_\alpha \mathbf{t}$, where $\mathbf{t}$ is a vector pointing the direction of the incident light ray. We shall consider the plane α from our previous example and let $\mathbf{t} = (1,\, 2,\, 1)^t$. Using the above calculations we easily find $\mathbf{t}' = -\frac{1}{9}\,(1,\, 24,\, 1)^t$, in which we may omit the pre-factor since only the direction is relevant. In order to obtain the trajectory of a particular reflected ray (or elastically scattered particle, e.g. billiard ball), we need some initial conditions, i.e., a source point - let this be again $A(3;2;1)$. Then, the incident ray has the parametric equations

$$\ell: \quad x = 3+\lambda, \quad y = 2(1+\lambda), \quad z = 1+\lambda, \qquad \lambda \in \mathbb{R}$$

and the point of incidence may easily be found as $B = \ell \cap \alpha$ via substitution

$$2(3+\lambda) - 2(1+\lambda) + 2(1+\lambda) + 1 = 2\lambda + 7 = 0 \quad \Rightarrow \quad \lambda = -\frac{7}{2}$$

which yields $B\left(-\frac{1}{2};\, 9;\, -\frac{5}{2}\right)$, so the equations of the reflected ray are

$$\ell': \quad x = -\frac{1}{2} + \mu, \quad y = 9 + 24\mu, \quad z = -\frac{5}{2} + \mu, \qquad \mu \in \mathbb{R}.$$

Note that the problem of geometric optics is not restricted to planar mirrors. On the contrary, one may consider a generic reflecting surface Σ usually given as an implicit function of the form $\Sigma:\ F(x,y,z) = 0$. The normal direction at each point is different (unless the surface is a plane) and therefore we talk

about a normal *vector field* $\mathbf{n}(x, y, z)$ rather than just a normal vector. It is not difficult to show that this normal field is actually proportional to the *gradient field* associated with the surface, defined as

$$\nabla F(x, y, z) = (F_x,\, F_y,\, F_z)^t \tag{10.5}$$

where the subscript denotes partial differentiation, e.g. $F_x = \dfrac{\partial F}{\partial x}$ etc. To see this, one only needs to note that the total differential of F vanishes

$$\mathrm{d}F = \frac{\partial F}{\partial x}\,\mathrm{d}x + \frac{\partial F}{\partial y}\,\mathrm{d}y + \frac{\partial F}{\partial z}\,\mathrm{d}z = 0 \tag{10.6}$$

since the function itself is constant on the surface. On the other hand, $(\mathrm{d}x,\, \mathrm{d}y,\, \mathrm{d}z)^t$ may be interpreted as an infinitesimal vector tangent to Σ and thus, the above vanishing scalar product asserts that ∇F is normal to it at each point. In physical considerations one may introduce a time parameter t and divide (10.6) by $\mathrm{d}t$, which yields $\dot{F} = \nabla F \cdot \mathbf{v} = 0$ where the dot denotes differentiation with respect to t and $\mathbf{v} = (\dot{x},\, \dot{y},\, \dot{z})^t$ stands for the instantaneous velocity of a moving particle bound to the surface Σ, which is therefore restricted to the tangent plane at each point. To illustrate the gradient field technique we consider a simple example of a quadratic surface

$$\Sigma: \quad F(x, y, z) = 3x^2 - y^2 + z^2 + 2xz - xy + yz + 1 = 0$$

which yields via straightforward differentiation

$$\nabla F(x, y, z) = (6x - y + 2z,\, -x - 2y + z,\, 2x + y + 2z)^t.$$

Now, let us take a light ray ℓ passing through the point $A(3; 2; 1)$ in the direction determined by the vector $\mathbf{t} = (2,\, 1,\, 2)^t$ and, as before, obtain the equation of the reflected ray ℓ'. To begin with, we have for the incident ray

$$\ell: \qquad x = 3 + 2\lambda, \quad y = 2 + \lambda, \quad z = 1 + 2\lambda, \qquad \lambda \in \mathbb{R}$$

and to find the point of incidence we need to substitute into the expression of Σ, which yields a quadratic equation for λ in the form

$$\ell \cap \Sigma: \qquad 23\lambda^2 + 50\lambda + 2z = 0 \quad \Rightarrow \quad \lambda_{1,2} = -1,\, \frac{25}{23}.$$

Let us take the intersection point $A_1(1; 1; -1)$, which corresponds to the root $\lambda_1 = -1$ and construct the normal $\mathbf{n}$ at this point as

$$\mathbf{n} = \nabla F(1, 1, -1) = (3,\, -4,\, 1)^t.$$

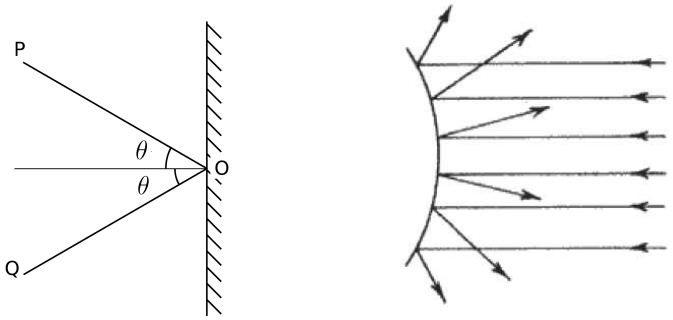

Figure 10.1: Reflection from planar (left) and curved (right) mirror surfaces.

Then, it is straightforward to obtain the mirror reflection as

$$\mathcal{M}_{\mathbf{n}} = 2\mathcal{P}_{\mathbf{n}}^{\parallel} - \mathcal{I} = -\frac{1}{13}\begin{pmatrix} 4 & 12 & -3 \\ 12 & -3 & 4 \\ -3 & 4 & 12 \end{pmatrix}$$

that yields the reflected direction vector $\mathbf{t}' = \mathcal{M}_{\mathbf{n}}\mathbf{t} = -\frac{1}{13}(14,\, 29,\, 22)^t$. Thus, omitting the irrelevant pre-factor, we finally obtain

$$\ell' : \qquad x = 1 + 14\mu, \quad y = 1 + 29\mu, \quad z = -1 + 22\mu, \qquad \mu \in \mathbb{R}.$$

There are more examples of this type at the end of the present chapter.

Rotations and the Galilean Group

As we already mentioned, all spacial motions are generated by reflections. In particular, the compositions of two reflections about intersecting planes (lines) is a rotation by twice the angle between them, while in the case of parallel planes/lines, it yields a translation by twice the distance between them (illustrate both cases with graphical examples). Since translations are reduced to vector summation, i.e., each spacial vector $\mathbf{u}$ is uniformly shifted by the translation vector $\mathbf{t}$ as $T : \ \mathbf{u} \ \rightarrow \ \mathbf{u}' = \mathbf{u} + \mathbf{t}$, we shall concentrate mostly on rotations. They have a particularly simple form in the planar case, as we already saw in the context of complex numbers, namely

$$\mathcal{R}_{\varphi} = \begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix} \tag{10.7}$$

which is a matrix representation of a complex number of unit modulus and thus, acts on vector in $\mathbb{R}^2$ as a rotation by an angle φ that may be seen directly as well. Both approaches clearly indicate that $\mathcal{R}_\varphi \mathcal{R}_\psi = \mathcal{R}_{\varphi+\psi}$ and in particular $\mathcal{R}_{-\varphi} = \mathcal{R}_\varphi^{-1}$ with $\mathcal{R}_0 = \mathcal{I}$. These properties turn the set of planar rotations into a (commutative) matrix group, i.e., the latter is closed under composition, with a neutral (unit) element and unique inverse for each element. Adding translations to that, which is allowing each radius-vector to be shifted arbitrarily, one obtains the group of Euclidean motions in the plane. Note, however, that the translational and rotational components of such motions are intertwined and hence, do not commute. For example, rotating a vector $\mathbf{u} \in \mathbb{R}^2$ by an angle φ and then, shifting it with a translation vector $\mathbf{t}$ clearly yields a compound motion $\mathbf{u} \to \mathbf{u}' = \mathcal{R}_\varphi \mathbf{u} + \mathbf{t}$, while reversing the order of the two transformations one ends up with $\mathbf{u}'' = \mathcal{R}_\varphi (\mathbf{u} + \mathbf{t}) \neq \mathbf{u}'$. We usually adopt the former convention (rotations first) and use block-matrices to represent the compound motion $\mathsf{G}(\varphi, \mathbf{t}) : \mathbf{u} \to \mathbf{u}'$ as

$$\begin{pmatrix} \mathbf{u}' \\ 1 \end{pmatrix} = \begin{pmatrix} \mathcal{R}_\varphi & \mathbf{t} \\ \mathbf{0}^t & 1 \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ 1 \end{pmatrix} = \begin{pmatrix} \cos\varphi & -\sin\varphi & t_1 \\ \sin\varphi & \cos\varphi & t_2 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ 1 \end{pmatrix}. \tag{10.8}$$

How is this construction generalized to the three-dimensional case? The block-matrix technique works similarly in every dimension and the extension of the translational counterpart is trivial, so the main difference is the group of spacial rotations compared to the planar case. To begin with, it is non-commutative, which makes it far more difficult to study. Each spacial rotation is again rotation in a plane (or about the axis, perpendicular to that plane), which cannot be said for higher dimensions. This plane, however, is not fixed but varies within the group. In general, the matrix of a rotation by an angle φ about the unit vector $\mathbf{n}$ is given by the Rodrigues' formula

$$\mathcal{R}(\varphi, \mathbf{n}) = \cos\varphi\, \mathcal{I} + (1 - \cos\varphi)\, \mathbf{n}\mathbf{n}^t + \sin\varphi\, \mathbf{n}^\times \tag{10.9}$$

where $\mathbf{n}^\times$ denotes the skew-symmetric linear transformation that maps each vector $\mathbf{v} \in \mathbb{R}^3$ to its cross product with $\mathbf{n}$, i.e., $\mathbf{n}^\times \mathbf{v} = \mathbf{n} \times \mathbf{v}$. Note that the above formula actually asserts that the parallel to the rotation axis component is preserved by the rotation and the normal one is rotated as in $\mathbb{R}^2$, which is obtained via decomposition in the normal to $\mathbf{n}$ plane with the aid of $\mathcal{P}_\mathbf{n}^\perp$ and $\mathbf{n}^\times$. In particular, rotations about the coordinate axes adopt a simple form, e.g. for a rotation about the z-axis one has from (10.9)

$$\mathcal{R}(\varphi, \hat{\mathbf{e}}_z) = \begin{pmatrix} \cos\varphi & -\sin\varphi & 0 \\ \sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Since the generic form of a spacial rotation (10.9) is far more complicated, in mechanics and navigation one often uses the so-called *Euler angles* (see Figure 10.2) representing an arbitrary $\mathcal{R}(\varphi, \mathbf{n})$ in the factorized form

$$\mathcal{R}(\phi, \theta, \psi) = \mathcal{R}(\psi, \hat{\mathbf{e}}_z)\mathcal{R}(\theta, \hat{\mathbf{e}}_x)\mathcal{R}(\phi, \hat{\mathbf{e}}_z).$$

The composition law for two consecutive rotations about non-parallel axes looks somewhat cumbersome in the three-dimensional case unless we take advantage of the Euler trigonometric substitution $\tau = \tan\dfrac{\varphi}{2}$ in (10.9) and denote for example $\mathbf{c} = \tau\mathbf{n}$, which is usually referred to as the *Rodrigues' vector*, or simply the *vector-parameter* of the corresponding rotation. Then, although this type of parametrization yields a singularity at $\varphi = \pi$ on the one hand, it also allows for expressing the matrix entries of $\mathcal{R}$ as rational functions and derive an efficient composition law in the form[3]

$$\langle \mathbf{c}_2, \mathbf{c}_1 \rangle = \frac{\mathbf{c}_1 + \mathbf{c}_2 + \mathbf{c}_2 \times \mathbf{c}_1}{1 - \mathbf{c}_1 \cdot \mathbf{c}_2}. \tag{10.10}$$

which yields the vector-parameter of the compound transformation

$$\mathcal{R} = \mathcal{R}(\mathbf{c}_2)\mathcal{R}(\mathbf{c}_1).$$

Baring in mind the above considerations about the rotation group in $\mathbb{R}^3$, one may still express spacial motions in pretty much the same block-matrix form as in $\mathbb{R}^2$, i.e., we have the composition of two motions in the form

$$\begin{pmatrix} \mathcal{R}_2 & \mathbf{t}_2 \\ \mathbf{0}^t & 1 \end{pmatrix} \begin{pmatrix} \mathcal{R}_1 & \mathbf{t}_1 \\ \mathbf{0}^t & 1 \end{pmatrix} = \begin{pmatrix} \mathcal{R}_2\mathcal{R}_1 & \mathcal{R}_2\mathbf{t}_1 + \mathbf{t}_2 \\ \mathbf{0}^t & 1 \end{pmatrix}$$

which defines the structure of the so-called *Galilean group* in $\mathbb{R}^3$. Similarly, one may allow for the upper-left block of the above matrices to contain inversions with respect to planes and spheres as well as scaling transformations, thus extending to the so-called *conformal* or *Möbius* group, which has its applications in various more complex geometric and mechanical problems.

Note that even if the rotation group is commutative, which is only in $\mathbb{R}^2$, introducing translations to it certainly destroys this property as the above composition formula clearly indicates. This is because translations and rotations are entangled in the Galilean motion group in a structure referred

[3] clearly, the cross product is responsible for the non-commutativity and when the two axes are collinear the expression is reduced to an addition formula for the tangent function.

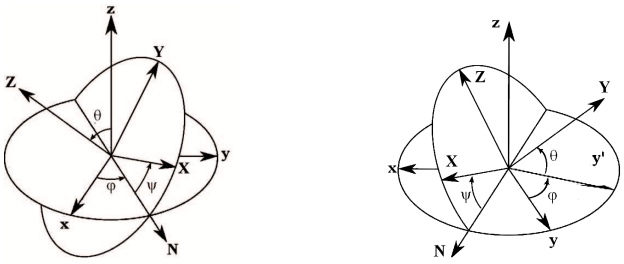

Figure 10.2: Parametrization of rotations with *Euler* and *Bryan* angles.

to as *semi-direct product* (in a direct product the two components would interact independently, and here only one of them affects the other). This peculiar behavior may be encoded in a conveniently chosen algebraic structure, which naturally adopts similar properties. One such structure is given by the so-called *dual quaternions* consisting of elements in the form

$$\tilde{\mathbf{q}} = \mathbf{q}_1 + \epsilon \mathbf{q}_0, \qquad \mathbf{q}_{0,1} \in \mathbb{H}, \;\; \epsilon^2 = 0$$

where the nilpotent element corresponds to the translational component. In mechanics this is also known as screw calculus. Similar algebraic constructions often make the description of cumbersome geometric and physical theories way easier and more productive, e.g. complexification of the above object is directly applicable to special relativity. Complex numbers, quaternions and groups themselves provide a good illustration of this idea.

Problem Session

1. Determine the rank and the nullity of a linear operator that projects n-dimensional vectors onto a k-dimensional subspace of $\mathbb{R}^n$.

2. If $\mathcal{R}$ is a spacial rotation, what is the rank of $\mathcal{I} \pm \mathcal{R}$ and does it depend on the angle/axis of $\mathcal{R}$?

3. Given an arbitrary basis $\{\mathbf{a}_k\}$ of $\mathbb{R}^n$, one has the so-called *Gram-Schmidt* procedure, which transforms it into an orthogonal basis. The idea is to choose a vector, say $\mathbf{a}_1$ and then modify $\mathbf{a}_2$ by subtracting its parallel to $\mathbf{a}_1$ component, i.e., $\mathbf{a}'_2 = \mathbf{a}_2 - \mathcal{P}_1^{\parallel}\mathbf{a}_2 = \mathcal{P}_1^{\perp}\mathbf{a}_2$. Then, for $\mathbf{a}'_3$ one subtracts two parallel projectors to make sure that $\mathbf{a}'_3$ is normal to both $\mathbf{a}'_2$ and $\mathbf{a}'_1 = \mathbf{a}_1$. Proceeding by induction, one may define

$$\mathbf{a}'_k = \mathbf{a}_k - \sum_{i=1}^{k-1} \mathcal{P}_i^{\parallel}\mathbf{a}_i$$

where $\mathcal{P}_i^{\parallel}$ denotes the parallel projector in the direction determined by $\mathbf{a}_i'$. Usually this is followed by a normalization that finally yields an orthonormal basis $\{\hat{\mathbf{a}}_k'\}$. Apply this technique to the set of vectors

$$\mathbf{a}_1 = (-1,\, 1,\, 1)^t, \qquad \mathbf{a}_2 = (1,\, -1,\, 1)^t, \qquad \mathbf{a}_3 = (1,\, 1,\, -1)^t.$$

4. Given the four points $A(1;2;3)$, $B(2;3;1)$, $C(3;1;2)$ and $D(0;0;0)$, find the coordinates of the reflections of the geometric center of the pyramid $ABCD$ with respect to its four sides.

5. A billiard ball with velocity vector $\mathbf{v} = (1, -1)^t$ hits the boundary of the billiard table, which has the shape on an astroid $x^{2/3} + x^{2/3} = 2$, at the point $A(1;1)$. Find the trajectory of the ball after the reflection.

Chapter 11

The Eigenvalue Problem

In many practical problems it is important to know the spacial directions, preserved by the action of a linear operator. This is especially useful in mechanical considerations including optimization, determining states of equilibria for complex systems etc. Denoting $\hat{\mathbf{v}}$ the unit vector along the fixed direction, we are faced with an equation of the form

$$A\hat{\mathbf{v}} = \lambda\hat{\mathbf{v}}, \qquad \lambda \in \mathbb{C} \tag{11.1}$$

known as the *characteristic equation* of the linear operator A. Actually, homogeneity with respect to $\hat{\mathbf{v}}$ implies that any vector $\mathbf{v} \parallel \hat{\mathbf{v}}$ is a solution and in particular the zero vector 0. However, we are interested only in nontrivial solutions, which we refer to as *eigenvectors*. To each eigenvector $\mathbf{v}$, or more precisely, invariant direction

$$\operatorname{Span}\mathbf{v} = \{\mu\mathbf{v},\, \mu \in \mathbb{R}\}$$

we may assign the unique complex number λ in (11.1) called an *eigenvalue*, which measures the deformation of vectors along this direction under the action of A. In particular, the construction of of eigenvectors and eigenvalues turns out to be of crucial importance for the study of (skew-)symmetric, orthogonal and unitary operators with their geometric and dynamical properties. For the description of the nontrivial solutions to equation (11.1) it is convenient to begin by writing it in the form

$$(A - \lambda\mathcal{I})\,\hat{\mathbf{v}} = \mathbf{0}$$

and point out that the necessary and sufficient condition for a homogenous linear system like this to have non-trivial solutions is to be singular, i.e.,

$$\det(A - \lambda\mathcal{I}) = 0. \tag{11.2}$$

Otherwise, inverting the matrix one inevitably ends up with $(A-\lambda\mathcal{I})^{-1}\,\mathbf{0}=\mathbf{0}$. With the aid of condition (11.2) we determine the set of eigenvalues $\{\lambda_k\}$ also referred to as the *spectrum* of A consisting of the zeroes of the above polynomial. Then, for each λ_k we solve the linear homogenous equation

$$(A-\lambda_k\mathcal{I})\,\mathbf{v}_k=0$$

and thus, after normalization, obtain the corresponding unit eigenvector $\hat{\mathbf{v}}_k$. For a simple root the corresponding solution contains one undetermined parameter, which corresponds to the freedom of scaling $\mathbf{v}$. Multiple roots, on the other hand, correspond to subspaces in $\mathbb{C}^n$ of dimension, equal to the multiplicity, in which all directions are preserved and thus, we may also "rotate" the eigenvectors. We shall clarify this part by the end of the chapter.

11.1 Spectral Properties of Operators

Let us now focus on some commonly used classes of linear operators and study the properties of their spectra. We begin with a simple but powerful

Theorem 12 *The eigenvalues of symmetric and hermitian operators are all real and the eigenvectors corresponding to different eigenvalues - orthogonal.*

Proof. Consider the hermitian bilinear form

$$\langle\mathbf{v}_i,A\mathbf{v}_j\rangle=\bar{\mathbf{v}}_i\cdot A\mathbf{v}_j=\lambda_j\langle\mathbf{v}_i,\mathbf{v}_j\rangle.$$

If A is hermitian, we also have from the conjugated version of formula (11.1)

$$\langle\mathbf{v}_i,A\mathbf{v}_j\rangle=\langle A\mathbf{v}_i,\mathbf{v}_j\rangle=\bar{\lambda}_i\langle\mathbf{v}_i,\mathbf{v}_j\rangle$$

and for $i=j$ this yields $\lambda_i=\bar{\lambda}_i\Rightarrow\lambda_i\in\mathbb{R}$. With this in mind, in the case $i\neq j$ and $\lambda_i\neq\lambda_j$ equality holds only for $\langle\mathbf{v}_i,\mathbf{v}_j\rangle=0$, which yields $\hat{\mathbf{v}}_i\perp\hat{\mathbf{v}}_j$. The symmetric case follows as a subset. □

Corollary 3 *The eigenvalues of skew-hermitian and skew-symmetric operators are purely imaginary and eigenvectors corresponding to different eigenvalues are orthogonal.*

In the above cases the eigenvectors constitute an orthonormal basis[1], in which the matrix of the corresponding linear operator has a diagonal form.

[1] in the case of multiple roots, one performs an additional normalization procedure.

Theorem 13 *The eigenvalues of unitary and orthogonal transformations lie on the unit circle* $|\lambda| = 1$ *and eigenvectors corresponding to different eigenvalues are orthogonal.*

Let U be unitary, and consider the hermitian form

$$v_{ij} = \langle U\mathbf{v}_i, U\mathbf{v}_j\rangle = \bar{\lambda}_i\lambda_j\langle \mathbf{v}_i, \mathbf{v}_j\rangle.$$

On the other hand, by definition $UU^\dagger = U^\dagger U = \mathcal{I}$, so

$$v_{ij} = \langle U^\dagger U\mathbf{v}_i, \mathbf{v}_j\rangle = \langle \mathbf{v}_i, \mathbf{v}_j\rangle$$

and for $i = j$ we have $v_{ii} \neq 0$ (assuming that $\mathbf{v}_i \neq 0$), thus $|\lambda_i| = 1$ follows. Moreover, $i \neq j$ yields $\bar{\lambda}_i\lambda_j = 1$ and since $\lambda_i \neq \lambda_j$ we come to a contradiction unless $\langle \mathbf{v}_i, \mathbf{v}_j\rangle = 0$. The orthogonal case obviously follows as a subset. □

The two above theorems are actually equivalent by a general result asserting that each unitary (orthogonal) matrix is an exponent of a skew-hermitian (skew-symmetric) one. Then, the equivalence follows as a consequence of a well-known property of operator spectra, given by the following

Theorem 14 *The matrices* A *and* $f(A)$ *have the same eigenvectors and if* $\{\lambda_k\}$ *are the eigenvalues of* A*, those of* $f(A)$ *are given by* $\{f(\lambda_k)\}$*.*

Proof. Let f be a smooth function. Then, it is expandable in a Taylor series

$$f(A) = f(0)\mathcal{I} + f'(0)A + \frac{f''(0)}{2!}A^2 + \ldots = \sum_{k=0}^{\infty}\frac{f^{(k)}(0)}{k!}A^k$$

which makes sense also for a matrix argument with $A^k = AA\ldots A$ (k times) and in particular, $A^0 = \mathcal{I}$. Therefore, the matrices A and $f(A)$ commute, so they determine identical sets of eigenvectors (prove this point alone!). As the latter constitute a basis, in which both A and are diagonal and it is not hard to show that $f:\ \mathrm{diag}\{\lambda_1, \lambda_2, \ldots \lambda_n\} \to \mathrm{diag}\{f(\lambda_1), f(\lambda_2), \ldots f(\lambda_n)\}$. Furthermore, for non-smooth functions f, one may always find an approximations with smooth ones and all properties we need endure the limit. □

Why is the zero commutator necessary and sufficient condition for a common set of eigenvectors? We shall leave the details of the proof to the reader providing only an motivational example. In the case of symmetric operators the eigenvectors constitute a basis, referred to as *canonical*, in which the

matrix of the operator has a diagonal form. Now, if two matrices are diagonal in the same basis, they obviously commute and vice versa - if one of them is diagonal in a given basis and the commutator vanishes, the other needs to be diagonal as well (if this statement is not obvious to you, work out the details). Our main example is the exponential function, which maps skew-symmetric (skew-hermitian) matrices into orthogonal (unitary) ones. This is also one more way to see that the eigenvalues of the latter lie on the unit circle (since for the former they are purely imaginary). Moreover, we may prove that each rotation, i.e., orientation-preserving orthogonal transformation (with positive determinant), in odd-dimensional space has an invariant axis defined as the linear span of eigenvectors corresponding to eigenvalue $\lambda = 1$. One way to see this is to take into account that if A is skew-symmetric, then

$$\det A = \det A^t = \det(-A) = (-1)^n \det A$$

and for $n = 2k + 1$ this yields $\det A = 0$, so in a certain basis the matrix of A has a vanishing row, hence, a zero eigenvalue. But then, the orthogonal transformation corresponding to it via the exponential map, has $\lambda = \mathrm{e}^0 = 1$ in the spectrum, which determines the invariant axis (the axis of rotation in the case $n = 3$). On the contrary, for $n = 2k$ the above equality of determinants is automatically satisfied so it does not imply a vanishing eigenvalue. Consider for example the matrix representation of the imaginary unit

$$\mathcal{J} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

whose characteristic equation $\lambda^2 + 1 = 0$ yields the pair of complex roots $\lambda_{1,2} = \pm i$. Now, one may jump directly to $n = 4$ with the block-matrix

$$A = \begin{pmatrix} \alpha\mathcal{J} & 0 \\ 0 & \beta\mathcal{J} \end{pmatrix}, \qquad \alpha, \beta \in \mathbb{R}$$

which obviously has eigenvalues $\{\pm i\alpha, \pm i\beta\}$ and its exponent yields a rotation simultaneously in the two mutually perpendicular x, y and z, w - planes by angles α and β, respectively. No axis in $\mathbb{R}^4$ is preserved by its action, only the origin (the rotation center) remains fixed. Note that rotations exist in all spacial dimensions but we can speak of rotation about an axis only for $n = 3$. In every other case we may encounter rotation in a plane at best, but this is something that does not happen often as it demands a set of highly non-trivial conditions to be satisfied by the corresponding matrix entries.

11.2 The Degenerate Case

So far we focused mostly on the case of simple eigenvalues, but the main results refer to the degenerate case as well. In particular, if a linear symmetric operator A has a multiple eigenvalue λ in the spectrum, it determines an invariant (under the action of A) subspace V_λ of dimension equal to the multiplicity of λ. The trivial example is the spherical operator $A = \lambda\mathcal{I}$, for which the multiplicity is maximal and thus, V_λ coincides with the ambient space. In the non-degenerate case all subspaces V_λ are one-dimensional and therefore we often speak of "invariant directions" rather than eigenspaces. Similarly, we have invariant planes and higher-dimensional subspaces in $\mathbb{R}^n$. Let us consider the case of symmetric operators starting with an example

$$A_{\{\hat{\mathbf{e}}_k\}} = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$$

in which it is not difficult to factorize the characteristic equation

$$\det(A - \lambda\mathcal{I}) = (2 - \lambda)(1 + \lambda)^2 = 0$$

which leads to one double and one simple root, namely $\lambda_{1,2} = -1$ and $\lambda_3 = 2$. The latter is almost trivial - the system of homogeneous equations yields

$$(A - 2\mathcal{I})\mathbf{v}_3 = 0 \quad \Rightarrow \quad \mathbf{v}_3 \sim (1,\, 1,\, 1)^t$$

and we may fix the arbitrary scalar factor to $1/\sqrt{3}$ in order to ensure unit length. On the other hand, for the double root one obtains two independent solutions in the form

$$\mathbf{v}_1 \sim (1,\, 1,\, -2)^t, \qquad \mathbf{v}_2 \sim (-2,\, 1,\, 1)^t$$

which are obviously not mutually perpendicular as one might expect. However, since the whole plane spanned by $\mathbf{v}_1$ and $\mathbf{v}_2$ (which, by the way, is the orthogonal complement of $\mathbf{v}_3$) is preserved by the action of A, we may choose an arbitrary pair of vectors in this plane and in particular, an orthonormal one. In order to do so, we only need to normalize $\mathbf{v}_1$ with a normalization factor $1/\sqrt{6}$ and use projectors in the Gram-Schmidt procedure[2] to eliminate the parallel to $\mathbf{v}_1$ component of $\mathbf{v}_2$, thus finally obtaining

$$\hat{\mathbf{v}}_1 = \frac{1}{\sqrt{6}}(1,\, 1,\, -2)^t, \qquad \hat{\mathbf{v}}_2 = \frac{1}{\sqrt{2}}(1,\, -1,\, 0)^t, \qquad \hat{\mathbf{v}}_3 = \frac{1}{\sqrt{3}}(1,\, 1,\, 1)^t.$$

[2] see the problem session of the previous chapter.

These unit vectors may be used to construct an orthonormal basis in which the matrix of the above operator takes the form $A_{\{\hat{\mathbf{v}}_k\}} = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 2 \end{pmatrix}$. As explained in Chapter 8, the components of the vectors $\hat{\mathbf{v}}_k$ yield the matrix of the orthogonal map that relates the two bases $T : \{\hat{\mathbf{e}}_k\} \to \{\hat{\mathbf{v}}_k\}$, i.e., the transition matrix. It may also be viewed as a similarity transformation linking the above matrix representations of the operator A, namely

$$A_{\{\hat{\mathbf{v}}_k\}} = T A_{\{\hat{\mathbf{e}}_k\}} T^{-1}.$$

This is easy to see keeping in mind that the transformation of vectors is $\mathbf{x}_{\{\hat{\mathbf{v}}_k\}} = T\mathbf{x}_{\{\hat{\mathbf{e}}_k\}}$ and the same for $\mathbf{x}' = A\mathbf{x}$, which yields

$$\mathbf{x}'_{\{\hat{\mathbf{v}}_k\}} = A_{\{\hat{\mathbf{v}}_k\}}\mathbf{x}_{\{\hat{\mathbf{v}}_k\}} = T A_{\{\hat{\mathbf{e}}_k\}} T^{-1} T\mathbf{x}_{\{\hat{\mathbf{e}}_k\}} = T\mathbf{x}'_{\{\hat{\mathbf{e}}_k\}}.$$

Similarly, if A is a matrix of a bilinear form, i.e., $\alpha(\mathbf{x}, \mathbf{y}) = \mathbf{x} \cdot A\mathbf{y}$, it may be shown (work out the details!) that it is transformed as

$$A_{\{\hat{\mathbf{v}}_k\}} = T A_{\{\hat{\mathbf{e}}_k\}} T^t$$

when the basis is changed. If we switch from one orthonormal basis to another, as we do in our example, the matrix T is orthogonal and the two transformations would be identical, but this is not so in the generic case. Now, let us consider a bit more exotic example, namely

$$B_{\{\hat{\mathbf{e}}_k\}} = \begin{pmatrix} 0 & 5 & -4 \\ -5 & 0 & 3 \\ -4 & 3 & 0 \end{pmatrix}$$

that appears in various physical contexts. The characteristic equation immediately yields $\lambda_1 = \lambda_2 = \lambda_3 = 0$, which is somewhat disturbing - does it mean that in some basis the matrix has only zero entries? Of course, not. In this case, the matrix we are dealing with is neither symmetric, nor skew-symmetric and its eigenvectors do not necessarily constitute a basis in $\mathbb{R}^3$. Actually, it is not difficult to see that its rank is equal to two: there is no pair of linearly dependent rows or columns, so we have only one relation for the three vectors. But then, it turns out that only one direction is preserved, namely the one given by the so-called Egyptian vector $\mathbf{v} = (3\,, 4\,, 5\,)^t$. The linear operator B acts on vectors in $\mathbb{R}^3$ as a cross product with $\mathbf{v}$ followed by a sign inversion of the first two components. The latter obviously does not affect zero vectors, so the kernel of B coincides with the kernel of $\mathbf{v} \times \cdot$ and we see that $\exp B$ is somewhat similar to a rotation. Such linear maps are usually referred to as *pseudo-rotations* or *Lorentz transformations* and have vital importance in hyperbolic geometry, relativity and quantum mechanics.

Problem Session

1. Obtain the eigenvectors of $\mathcal{J}$ and the block matrix A constructed with it above. Attempt to perform the Gram-Schmidt procedure relative to the Hermitian scalar product in the degenerate case $\alpha = \beta$.

2. What is the algebraic condition two matrices, say A and B need to satisfy, so that the eigenvalues of $A+B$, respectively AB, are the sum (product) of the eigenvalues of A and those of B.

3. Find the eigenvalues and eigenvectors of the matrix $A = \begin{pmatrix} 3 & -1 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 1 & 0 & 3 & 1 \\ 0 & 1 & 0 & 3 \end{pmatrix}$.

4. Given the matrix $A = \begin{pmatrix} 2 & -4 \\ -4 & 2 \end{pmatrix}$ find the inverse of $B = A^2 - 10\mathcal{I}$, as well as the eigenvalues and eigenvectors of both A and B^{-1}.

5. Find the eigenvalues and eigenvectors of a compound planar motion obtained as a composition of a (counterclockwise) rotation by an angle $\phi = \frac{\pi}{2}$, a reflection about the y-axis and a translation with a translation vector $\mathbf{t} = (1, -1)^t$. Change each of the above parameters separately to see how it affects the preserved directions. What happens if you include also scaling, e.g. stretch all distances twice?

6. Recall the construction of the cross product in $\mathbb{R}^3$ and note that fixing one of the factors it is reduced to a linear map on the other, i.e., $\mathbf{u} \times \mathbf{v} = \mathcal{A}\,\mathbf{v}$ and thus, it may be presented explicitly in a matrix form, namely as the skew-symmetric matrix (see if this is correct)

$$\mathcal{A} = \mathbf{u}^{\times} = \begin{pmatrix} 0 & -u_3 & u_2 \\ u_3 & 0 & -u_1 \\ -u_2 & u_1 & 0 \end{pmatrix}.$$

Obtain the eigenvectors and eigenvalues for arbitrary $\mathbf{u}$. Note also that if we choose $\mathbf{u} = \phi\,\mathbf{n}$ where $\mathbf{n}$ is a unit vector denoting the rotation axis and ϕ - the corresponding angle of rotation (this construction is usually referred to as the *rotation vector*), the $\mathbf{u}^{\times}$ has the meaning of an infinitesimal rotation and the compound transformation is obtained via exponentiation as $\mathcal{R}_t(\phi, \mathbf{n}) = \exp(t\mathcal{A})$, where t is the time parameter. Then, for a fixed rotation vector the eigenvectors do not depend on t, but the eigenvalues do. Show this dependence explicitly.

Chapter 12

The Geometry of Quadrics

Quadrics are a natural step beyond considering linear equations and as such, provide an intersection of linear algebra, projective and algebraic geometry with lots of practical applications. Examples emerge from various fields starting with elementary Euclidean geometry and arriving at rather advanced mathematical physics (e.g the Klein quadric and the twistor model).

12.1 Normal Forms and Canonical Axes

Let us begin with the almost trivial one-dimensional case, in which a generic quadric, i.e., a quadratic function of one variable, may be written as

$$\mathcal{Q}(x) = ax^2 + bx + c, \qquad a \neq 0. \tag{12.1}$$

What is the geometric meaning of this expression? If we have an extra dimension y to embed the graph of the function $y = \mathcal{Q}(x)$, it is well-known to be a parabola, which may be given in a simpler form by a linear change of coordinates $x \to \xi$, namely

$$\mathcal{Q}(\xi) = \alpha \pm \xi^2, \qquad \xi = \sqrt{|a|}\, x + \frac{b}{2\sqrt{|a|}}, \quad \alpha = c - \frac{b^2}{4|a|}$$

where the sign in front of ξ^2 equals the sign of the highest order coefficient a in the equation of the quadric (12.1). The above representation of $\mathcal{Q}(\xi)$ is called a *normal* of *canonical* form and the variable ξ itself - a *canonical variable*. Now, let us see how this idea may be generalized to higher dimensions. The reasonable thing to do here would be to consider the case $n = 2$, in which the generic expression of a quadric is given in the form

$$\mathcal{Q}(\mathbf{x}) = \mathbf{x} \cdot A\mathbf{x} + \mathbf{b} \cdot \mathbf{x} + c, \qquad \mathbf{x} = (x_1,\, x_2)^t \tag{12.2}$$

where $A \leftrightarrow \{a_{ij}\}$ is a nonzero 2×2 symmetric matrix[1], so the above may be written in components as

$$\mathcal{Q}(\mathbf{x}) = a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2 + b_1x_1 + b_2x_2 + c. \tag{12.3}$$

Then, one may consider the graph of the function $y = \mathcal{Q}(\mathbf{x})$ or, alternatively, setting $\mathcal{Q}(\mathbf{x}) = 0$, obtain the equation of the quadric as an implicit function[2] of x_1 and x_2. Here we may also find the canonical coordinates $\boldsymbol{\xi} = (\xi_1, \xi_2)^t$, in which the equation (12.3) takes its simplest (normal) form. In order to do so, one needs to introduce an Euclidean motion in the x_1, x_2-plane. Let us consider first the case in which only a rotation is sufficient, that is when the linear term is zero ($b_1 = b_2 = 0$) and one only needs to diagonalize the matrix A thus finding the *canonical axes* of the quadric $\mathcal{Q}$. This is a straightforward illustration of the eigenvalue problem considered above and we have

$$\mathcal{Q}(\boldsymbol{\xi}) = \lambda_1\,\xi_1^2 + \lambda_2\,\xi_2^2 + c, \qquad \boldsymbol{\xi} = T\mathbf{x} \tag{12.4}$$

where $\lambda_{1,2}$ are the eigenvalues of A and $T : \{\hat{\mathbf{e}}_k\} \to \{\hat{\mathbf{a}}_k\}$ is the rotation matrix constructed from its normalized eigenvectors $\hat{\mathbf{a}}_k$ that yield the so-called *canonical basis*. Let us see this with an example: consider the quadric

$$\mathcal{Q}: \qquad 2x_1x_2 = 1$$

and derive the matrix of its bilinear form $A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, for which the characteristic equation $\lambda^2 - 1 = 0$ yields the two simple roots $\lambda_{1,2} = \pm 1$ and thus, the canonical form of $\mathcal{Q}$ may be written as

$$\mathcal{Q}: \qquad \xi_1^2 - \xi_2^2 = 1.$$

It only remains to find how the initial (x_1, x_2)-coordinates are related to the canonical ones, ξ_1 and ξ_2. The linear transformation is given by the transition matrix between the two bases obtained by the normalized eigenvectors of A: $\hat{\mathbf{a}}_1 = \frac{1}{\sqrt{2}}(1, -1)^t$ and $\hat{\mathbf{a}}_2 = \frac{1}{\sqrt{2}}(1, 1)^t$, which finally yields the relation

$$\xi_1 = \frac{1}{\sqrt{2}}\,(x_1 + x_2), \qquad \xi_2 = \frac{1}{\sqrt{2}}\,(x_2 - x_1)$$

so the canonical axes in this case are the diagonals of the four quadrants.

[1] even if its skew-symmetric part is nonzero, it would not survive the contraction $\mathbf{x} \cdot A\mathbf{x}$.

[2] note that this approach would lead in the one-dimensional case to a set of two points.

Now, let us consider a more general type of quadric, e.g.

$$\mathcal{Q}: \quad x_1^2 + x_2^2 - 2x_1x_2 + x_1 + x_2 - 2 = 0$$

for which the canonical form would involve both rotation and translation. We shall approach this problem in the standard way - first by obtaining the eigenvalues of the matrix of the quadratic part $A = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$, which are obviously $\lambda_1 = 0$ and $\lambda_2 = 2$ and the corresponding (unit) eigenvectors, respectively $\hat{\mathbf{a}}_1 = \frac{1}{\sqrt{2}}(1, 1)^t$ and $\hat{\mathbf{a}}_2 = \frac{1}{\sqrt{2}}(-1, 1)^t$, so the transition matrix T is a *quarter-turn* (rotation by a right angle) and we have, similarly to the above case:

$$\xi = \frac{1}{\sqrt{2}}(x_1 - x_2), \qquad \eta = \frac{1}{\sqrt{2}}(x_1 + x_2).$$

Substituting $\mathbf{x} = T^{-1}\boldsymbol{\xi}$ in the expression of the quadric $\mathcal{Q}$, we obtain

$$\mathcal{Q}: \qquad 2\xi_2^2 - \sqrt{2}\,\xi_1 - 2 = 0$$

and then, easily get rid of the constant term by shifting the ξ_1 variable as

$$\xi_1 \to \tilde{\xi}_1 = \xi_1 + \sqrt{2}$$

which finally yields the equation of a parabola

$$\mathcal{Q}: \qquad \xi_2^2 + \frac{1}{\sqrt{2}}\,\tilde{\xi}_1 = 0.$$

Note that if we change only one sign in the above quadric, namely

$$\mathcal{Q}': \quad x_1^2 + x_2^2 - 2x_1x_2 + x_1 - x_2 - 2 = 0$$

the latter may be written for the variable $z = x_1 - x_2$ as

$$z^2 + z - 2 = 0$$

that has the two simple roots $z_1 = 1$, $z_2 = -2$ and thus, yields the two lines

$$g_1: \quad x_2 = x_1 - 1, \qquad g_2: \quad x_2 = x_1 + 2.$$

This splitting of the quadric into a pair of lines is referred to as *degeneracy*. Note that usually quadrics are considered in projective geometric perspective, but here we study their basic properties from Euclidean point of view.

12.2 Geometric and Optical Properties of Conics

Planar quadrics can be obtained as intersections of a (double infinite) cone with a plane and thus, often called *conic sections*. For example, if the plane is perpendicular to the cone's axis of symmetry, the intersection is obviously a circle, while if it is parallel to the cone's side then we end up with a parabola. In the generic setting the sections are ellipses or hyperbolae with various eccentricity, but there are also degenerate cases, in which it may be a line (the plane is tangent to the cone) or even a point (the cusp).

We shall assume a somewhat simplistic approach, which later will appear to be justified. Namely, let us declare the three types of non-degenerate conics with their canonical equations and set a clear classification criterion. Then, we shall use our knowledge on Euclidean motions in the plane to solve the problem of obtaining this canonical form of an arbitrary quadratic polynomial in x and y. There are three types of conics (see Figure 12.1):

- *Ellipses* with their canonical equation

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 \tag{12.5}$$

  where a and b are the so-called *semi-axes*[3].

- *Hyperbolae* described by either of the two equations

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = \pm 1. \tag{12.6}$$

- *Parabolas* given respectively as either of the equations

$$y^2 = 2px, \qquad x^2 = 2py \tag{12.7}$$

  where $p > 0$ is referred to as the *focal parameter*.

Note that one may choose only one of each pair of equations for the hyperbola and the parabola as a canonical representative since the other is obtained by a quarter-turn in the plane. However, it is sometimes convenient to have such freedom of choice. The classification theorem for quadratic planar curves states that they are all conic sections, i.e., one of the above types or possibly a degenerate case like the one considered above. An analogous

[3] respectively half of the minimal and maximal diameter (called just *axes*).

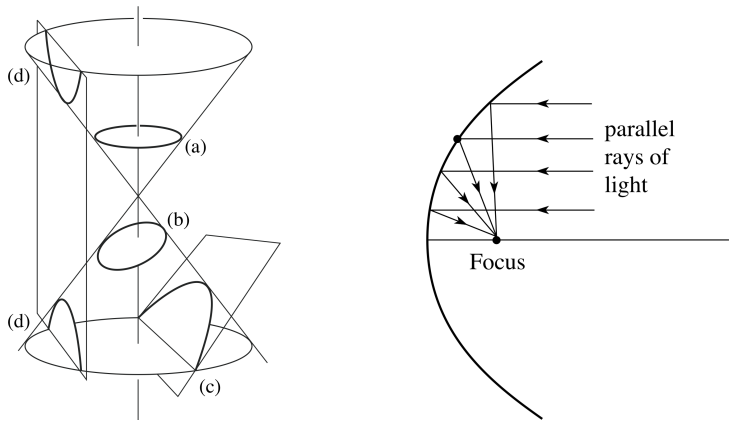

Figure 12.1: The basic non-degenerate conic sections: (a) the circle, (b) the ellipse, (c) the parabola and (d) the hyperbola (on the left); and the optical property of the parabola used in satellites (on the right).

result holds for quadratic surfaces in $\mathbb{R}^3$, which we discuss below as well. Moreover, all planar geometries have been classified in a similar manner as elliptic (e.g. geometry on the sphere and its projective versions), hyperbolic (such as the famous Lovachevsky model) and parabolic, also known as Euclidean. Then, if conics are that important, let us study their structure into more detail. It turns out that it is not necessary to intersect cones with planes each time we would like to generate such a quadric as we may define it also as *locus* of points. For instance, all points in the circle have the property that they are at the same distance R form the center. So, having that distance (radius) and the coordinates (x_0, y_0) of the center, we can generate the whole circle from the above property as the locus of points satisfying

$$(x - x_0)^2 + (y - y_0)^2 = R^2.$$

Similarly, for the ellipse (12.5), which obviously also has a center as it is symmetric with respect to reflection about both coordinate axes, we have two more important points referred to as *foci* (the plural of *focus*). The generalization of the above construction for ellipses sounds like this: the sum of the distances from the two foci $F_{1,2}$ is the same for each point of the ellipse. A similar construction works for the hyperbola, but there one takes the difference, rather than the sum of distances from the two foci. Of course, we need to determine the positions of the two foci in either case, and relate them to the canonical equation. Let us concentrate on the former for

now. It appears helpful to first define the *eccentricity* of the ellipse (12.5)

$$e = \sqrt{1 - \frac{b^2}{a^2}}, \qquad a \geq b. \tag{12.8}$$

This parameter may be used to determine the two foci as it yields the ratio of the distance $|F_1F_2|$ between them and the *major axis* $2a$ (here we assume that the ellipse is "horizontally" oriented, i.e., $a > b$). Hence, the two foci are placed on the major axis (in our case this is the x-axis) symmetrically with respect to the minor axis (and thus, the origin), and have coordinates

$$F_{1,2} : \left(\pm\sqrt{a^2 - b^2},\, 0\right).$$

Now, attempt to derive the canonical equation of the ellipse (12.5) from the restriction that the sum of the distances from each of its points to the two foci is constant. With a little more effort one may prove (do it!) the main optical property of ellipses, namely that the reflection (with respect to the ellipse, considered as a mirror surface) of a light ray passing through one of the foci, passes through the other. This very property holds also for the hyperbola if we consider light rays as infinite lines. In order to prove it, however, one needs to define the foci in this case and we do it again with the aid of the eccentricity e, which this time has the form

$$e = \sqrt{1 + \frac{b^2}{a^2}} \tag{12.9}$$

for a sideways opening hyperbola, i.e., equation (12.6) with the positive sign on the righthand side. The coordinates of the two foci are then determined similarly as (see Figure 12.2)

$$F_{1,2} : \left(\pm\sqrt{a^2 + b^2},\, 0\right)$$

and one may derive formula (12.6) from the condition that the difference between the distances from each point in the hyperbola from F_1 and F_2 is a constant. The value of this constant is equal to the major axis $2a$ both for the ellipse and the hyperbola. Note also that the hyperbola has one more convenient representation in the form $xy = c$, which may be written as an explicit function (that has no analogue in the elliptic or parabolic case).

For a thorough study of conic sections one needs to also introduce the notion of *directrices*. Those are two characteristic lines with special properties whose equations are given as (see Figure 12.2)

$$d_{1,2} : \quad x = \pm\frac{a}{e}. \tag{12.10}$$

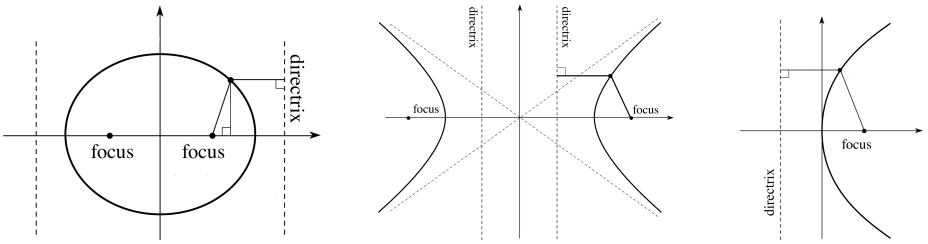

Figure 12.2: Directrices, foci and optical properties of conic sections.

The distance between the focus F_i and the directrix d_i is referred to as the *focal parameter* and in both cases it is equal to $p = b^2/ae$. Note that the eccentricity e and the focal parameter p are sufficient to determine the conic section. In the case of the ellipse we have for the two semi-axes

$$a = \frac{ep}{1-e^2}, \qquad b = \frac{ep}{\sqrt{1-e^2}} \tag{12.11}$$

and similarly for the hyperbola one has

$$a = \frac{ep}{e^2-1}, \qquad b = \frac{ep}{\sqrt{e^2-1}}. \tag{12.12}$$

The notion of a directrix is particularly useful when one considers the parabola, e.g., the one opening sideways: $y^2 = 2px$, for which the focus F and the directrix d are given respectively as $F(p/2,\, 0)$ and $d:\, x = -p/2$. Then, the parabola itself may be constructed as the locus of points at equal distance from F and d. Strictly speaking, in this case we have a second focus and directrix as well, but they are placed "at infinity". This concept appears a bit weird for the untrained student, but in fact quadrics and conics in particular appear most naturally in the context of projective geometry, were one has lines and points at infinity (the horizon is one good example).

Now, what is the identifying optical property of parabolas? As it turns out, each light ray through the focus is reflected by the parabola into a ray perpendicular to the directrix (parallel to the symmetry axis) and vice versa. This property is used in parabolic satellite dishes, which capture an almost parallel pencil of rays (this is a good approximation due to the long distance to the emitter) and direct them to the receiver placed at the focus (Figure 12.1). Satellites use rotational parabolic surfaces rather than plane parabolas, but it is not hard to see that the above property is invariant under rotation about the symmetry axis. Can you think of a technology that uses

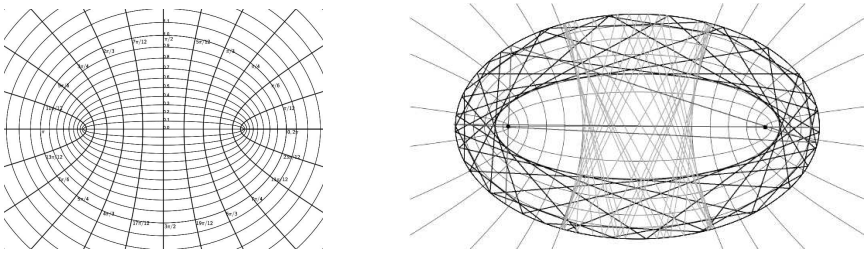

Figure 12.3: Foliation of the plane into mutually perpendicular families of confocal ellipses and hyperbolae given by the elliptic coordinates (on the left) and different types of closed trajectories of the elliptic billiard (on the right).

the same property in the reverse direction? Here is one more peculiar thing about conics: according to the famous Kepler's law of Newtonian gravity, the trajectories of heavy objects, whose motion is caused by the presence of a gravitational field, are either ellipses ($e < 1$), hyperbolae ($e > 1$) or parabolas ($e = 1$). The pole (e.g. a star causing the gravitational force) in each case is placed in one of the foci and the trajectory type is determined by the initial energy of the body, which is obviously related to the eccentricity. For example, ellipses are typical planetary or stellar orbits characterized by relatively low total energy (e.g. free falling). The hyperbolic case is known as *gravitational slingshot* - the initial energy of the body approaching the gravitational pole is large enough to overcome the force of attraction and escape the trap. Finally, at the boundary between the two cases ($e = 1$) the trajectory is a parabola. At first sight, this does not look much different from the hyperbolic escape, but there are major geometric distinctions, e.g. from a projective point of view, the parabola is a closed (at infinity) curve. Besides, all parabolas are similar, i.e., equal modulo a rotation, translation and/or a scale transformation, which is certainly not true for either hyperbolae or ellipses. Let us point out one more link, this time to dynamical systems, where it is often useful to consider toy models that mimic the behavior of more complex ones. In particular, billiard tables with different shapes may exhibit various properties, such as Lyapunov convergence (close trajectories tend to get exponentially closer with time) as it is in the elliptic case, orbital convergence (proximity between trajectories is conserved), e.g. rectangular or parabolic tables, as well as chaotic behavior (trajectories diverge at an exponential rate) as in the case of hyperbolic concave billiards.

More geometric properties: Since the circle is just a particular type of conic, it is relevant to ask whether polar coordinates are useful for the

description of other conics as well. It may be shown that the polar equation of a generic conic with eccentricity e and polar parameter p is given as

$$\rho = \frac{ep}{1 + e\cos\varphi} \tag{12.13}$$

where $\rho \in (0, \infty)$ is the radial variable (Euclidean distance to the origin) and $\varphi \in [0, 2\pi)$ stands for the polar angle. Apart from that, in the particular case of ellipses it is straightforward to generalize the polar change as

$$x = ar\cos\varphi, \qquad y = br\sin\varphi, \qquad \varphi \in [0, 2\pi) \tag{12.14}$$

and as the radial parameter $r > 0$ varies from 0 to 1 the above points fill the planar region enclosed by the ellipse given by equation[4] (12.5). However, the full orthogonal net in the plane for fixed foci $F_{1,2}(\pm\rho, 0)$ where $\rho = ae$ is the *linear eccentricity*, is generated by the so-called *elliptic coordinates*:

$$x = \rho\cos\vartheta\cosh\psi, \qquad x = \rho\sin\vartheta\sinh\psi \tag{12.15}$$

where the lines $\psi = \operatorname{arccosh}\frac{1}{e} = const.$ with $\vartheta \in [0, 2\pi)$ are confocal ellipses and respectively $\vartheta = const.$ with $\psi \in [0, \infty)$ - confocal hyperbolae, and both families of mutually perpendicular curves foliate the plane (Figure 12.3). In some considerations the latter turn out to be much more convenient compared to the usual Cartesian coordinates. Ellipses have plenty of other remarkable features, such as the isogonal property, which states that if O is the meeting point of the tangents at two arbitrary points $K_{1,2}$ in an ellipse with foci $F_{1,2}$, then the equality of angles $\measuredangle F_1OK_1 = \measuredangle F_2OK_2$ holds. It is interesting also to consider the envelope of all normals (the *evolute*) or that of closed billiard trajectories (Figure 12.3), which also play an important role in geometric optics and catastrophe theory. From a projective point of view we have several major theorems about conics (and quadrics in general) due to Pascal, Desargues Chasles, Apollonius and others considering the mutual positions of several points on a conic. As it turns out, a conic is completely determined by five points, i.e., it has five parameters[5], just as a line is determined by two or a plane - by three. There seems to be a contradiction at first sight as three points are sufficient to determine for instance a circle or a parabola, but this is only after we already know it is a circle or, respectively, a parabola, which alone fixes two of the parameters.

[4] show that a similar construction exploiting hyperbolic functions is valid for (12.6).

[5] three in the quadratic part and three in the linear one, modulo an overall scale factor.

12.3 Spacial Quadrics

Just as in the case of conic sections, we have a nice classification result for the bilinear forms of three real variables viewed as quadratic surfaces in $\mathbb{R}^3$. The non-degenerate cases fall into one of the following groups (Figure 12.4):

1. *Ellipsoids*, described by the canonical equation
$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1 \tag{12.16}$$
where a, b and c the *semi-axes* just as in the case of ellipses.

2. *One-sheeted hyperboloids*, given by the canonical equation
$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 1. \tag{12.17}$$

3. *Two-sheeted hyperboloids*, given respectively as
$$\frac{x^2}{a^2} - \frac{y^2}{b^2} - \frac{z^2}{c^2} = 1. \tag{12.18}$$

4. *Cones*, whose canonical equations may be written in the form
$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 0. \tag{12.19}$$
Apart from that, we have several cases that do not correspond to surfaces in $\mathbb{R}^3$, such as the the imaginary ellipsoid (sphere if $a = b = c$)
$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = -1$$
or the imaginary cone with equation
$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 0$$
whose only real point is the origin.

Note that we may have a coincidence of two or three semi-axes, the former case resulting in a surface of revolution, while the second one - possibly even in a spherical symmetry (in the case of the ellipsoid). Surfaces of revolution, such as the rotational ellipsoid for example, allow for dimensional reduction and are thus easier to study. The sphere, on the other hand, is just a sphere.

Next, we have the degenerate quadrics[6]:

[6] i.e., with a vanishing eigenvalue in the matrix of the bilinear form.

1. *Paraboloids* - respectively *elliptic* (+) and *hyperbolic* (−) ones:

$$z = \frac{x^2}{a^2} \pm \frac{y^2}{b^2} \tag{12.20}$$

2. *Cylinders* of different types:

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = \pm 1, \qquad \frac{x^2}{a^2} - \frac{y^2}{b^2} = 1, \qquad x^2 = 4pz \tag{12.21}$$

   the first one being referred to as either *elliptic* (+) or *imaginary* (−), the second - *hyperbolic* and the third one - *parabolic*.

3. Pairs of real (−) or imaginary (+) planes:

$$\frac{x^2}{a^2} \pm \frac{y^2}{b^2} = 0 \tag{12.22}$$

4. The most degenerate case is given by

$$x^2 = \alpha \tag{12.23}$$

   interpreted as either a pair of real parallel planes for $\alpha \geq 0$, coincident in the particular case $\alpha = 0$, or complex conjugated ones (for $\alpha < 0$).

Note that some of those cases are just the rotated analogues of the planar ones, e.g., each circular cone may clearly be interpreted as a rotated (about the bisector) configuration of two intersecting lines etc. Note also that in the cases of elliptic or hyperbolic paraboloids one may express the surface as an explicit function $z = z(x, y)$ and consider its *critical points*, i.e., those at which $z_x = z_y = 0$. To determine whether we have an *extremum* or a *saddle* at these points we consider the determinant of the second partial derivatives

$$\Delta = \begin{vmatrix} z_{xx} & z_{xy} \\ z_{yx} & z_{yy} \end{vmatrix} = z_{xx} z_{yy} - z_{xy}^2, \qquad z_{xy} = z_{yx}$$

and apply Sylvester's criterion to it, thus $\Delta < 0$ clearly yields a saddle point, i.e., we have $\lambda_1 \lambda_2 < 0$ for the eigenvalues, while $\Delta > 0$ - a local minimum or maximum[7]. However, $z(x, y)$ is quadratic, so Δ has constant coefficients, which means it determines the global behavior (constant curvature) of the surface. In the implicit function case we may work similarly with gradients.

[7] similarly, $\Delta = 0$ yields a *catastrophe*, such as the cylinder's *fold* or the cone's *cusp*.

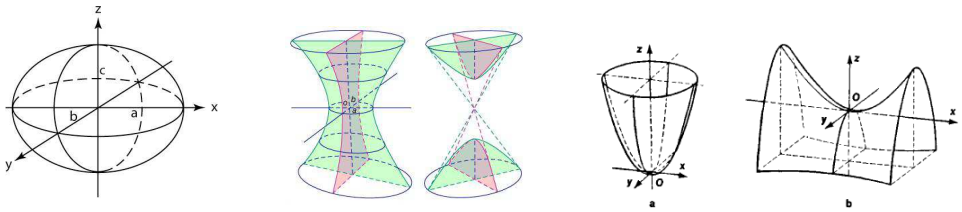

Figure 12.4: Spacial quadrics: ellipsoid, hyperboloids (both one- and two-sheeted) and paraboloids (circular and hyperbolic).

Problem Session

1. Obtain the canonical equations of the following conic sections

   a) $x^2 - 2xy + 2x + 2y + 1 = 0$

   b) $6x^2 + 9y^2 - 4xy - 4x - 32y - 6 = 0$

   c) $7x^2 + 4xy + 4y^2 - 40x - 32y + 5 = 0.$

2. Obtain the canonical equations of the following quadratic surfaces

   a) $x^2 + 2y^2 + 2z^2 + 2xy - 2x - 4y - 4z = 0$

   b) $y^2 + 3xy - 2yz + zx + 3x + y = 0$

   c) $x^2 + 2y^2 - z^2 + 2x - 4y + 2z + 1 = 0.$

3. Find the locus of the centers of the one-parameter set of curves

$$x^2 + 2xy - y^2 + 2\lambda x + 4\lambda y + 1 = 0, \qquad \lambda \in \mathbb{R}.$$

4. Determine the directrices and foci of the conics with equations

   a) $2x^2 + y^2 = 2y$

   b) $x2 - y^2 + 2xy - x + y = 0$

   c) $x^2 + 2x - y = 0.$

5. Determine the value of the parameter $\lambda \in \mathbb{R}$ so that the equation

$$x^2 + 3y^2 + 2zx + 2\lambda yz + 2zx - 2x - 8y - 2z - 3 = 0$$

describes a cone and find the coordinates of its cusp.

6. Derive the canonical equation and find the critical points of the surface

$$z(x, y) = x^2 + 2y^2 - 6xy + x - 3y + 1.$$

Obtain its normal vector. Hint: express the surface as an implicit function $F(x, y, z) = 0$ and use partial differentiation for $\mathbf{n} = (F_x,\, F_y,\, F_z)^t$.

7. Given the line $g:\ 2x - 5y + 1 = 0 \ \cap \ x + 3y - 2z + 3 = 0$, derive the equation of the tangent plane at the intersection point of each of the following quadrics with g (if they intersect):

   a) $4x^2 + 2y^2 + 12z^2 - 4xy + 8yz + 12zx + 14x - 10y + 7 = 0$

   b) $5x^2 + 9y^2 + 9z^2 - 12xy - 6zx + 12x - 36z = 0$

   c) $x^2 - 2y^2 + z^2 + 6yz - 4zx - 8x + 10y = 0$

   and determine the type of the corresponding quadric.

8. Explain how the curved surface of the one-sheeted hyperboloid can be constructed using only straight lines. Can you think of a practical application for this construction?

9. Explain why the acoustics in a concert hall with a hyperbolic (concave) shape would be bad. Where would you put the orchestra and the most expensive seats in the parabolic and elliptic cases? Are there any areas in the concert call that need to be avoided?

Bibliography

[1] Brinkmann H. and Klotz E., *Linear Algebra and Analytic Geometry*, Addison-Wesley Pub. Co., 1971.

[2] Rider P., *First-Year Mathematics for Colleges*, Second Edition, The Macmillan Company, New York 1962.

[3] Lichnerowicz A., *Elements of Tensor Calculus*, Methuen & Co. Ltd. 1962.

[4] Postnikov M., *Lectures in Geometry, Semester II* : *Linear Algebra and Differential Geometry*, Mir Publishers, Moscow 1982.

[5] Polyanin A., Manzhirov A., *Handbook of Mathematics for Engineers and Scientists*, Chapman & Hall/CRC Press, Boca Raton 2006.

[6] Tang K., *Mathematical Methods for Engineers and Scientists 2*, Springer-Verlag, Berlin 2007.

[7] P. Lounesto, *Clifford Algebras and Spinors* (Second Edition), Cambridge University Press, Cambridge 2001.