Università degli Studi di Torino

Master's Degree in Materials Science

Pre-course for Quantum effects in materials: from Theory to Modelling

Andrea Cavaglià, Paolo Torrielli

Contents

Chapter 1

Basics in complex numbers

1.1 Definitions and properties

A complex number z is an ordered pair of real numbers x, y

$$z = x + iy \equiv (x, y)\,, \qquad x, y \in \mathbb{R}, \quad z \in \mathbb{C}\,, \tag{1.1}$$

where i is the imaginary unit satisfying

$$i^2 = -1\,. \tag{1.2}$$

The numbers x and y are called the *real* and *imaginary* part of z, respectively:

$$x = \mathrm{Re}(z)\,, \qquad x = \mathrm{Im}(z)\,. \tag{1.3}$$

Note that the imaginary part y is a real number.
Two basic operations are defined on the set $\mathbb{C}$ of complex numbers: *addition* and *multiplication*, with the following definition and properties, identical to those of real numbers.

1) **Addition (+)**

$$z_1 = x_1 + iy_1\,, \quad z_2 = x_2 + iy_2 \quad \Longrightarrow \quad z_1 + z_2 = (x_1 + x_2) + i(y_1 + y_2)\,. \tag{1.4}$$

Associativity: $z_1 + (z_2 + z_3) = (z_1 + z_2) + z_3$.
Commutativity: $z_1 + z_2 = z_2 + z_1$.
The neutral element under addition is 0 (namely $0 + i0$): $z + 0 = 0 + z = z$.
The inverse of z under addition is $-z = -x - iy$, such that $z + (-z) = (-z) + z = 0$.

2) **Multiplication ($\cdot$)**

$$\begin{aligned} z_1 \cdot z_2 &= (x_1 + iy_1) \cdot (x_2 + iy_2) \\ &= (x_1x_2 - y_1y_2) + i(x_1y_2 + x_2y_1)\,. \end{aligned} \tag{1.5}$$

Associativity: $z_1 \cdot (z_2 \cdot z_3) = (z_1 \cdot z_2) \cdot z_3$.
Commutativity: $z_1 \cdot z_2 = z_2 \cdot z_1$.
The neutral element under multiplication is 1 (namely $1 + i0$): $z \cdot 1 = 1 \cdot z = z$.
The inverse of z under multiplication is z^{-1}, such that $z \cdot (z^{-1}) = (z^{-1}) \cdot z = 1$. For $z = x + iy$, its inverse is found as follows

$$z^{-1} = \frac{1}{x + iy} = \frac{x - iy}{(x + iy)(x - iy)} = \frac{x - iy}{x^2 + y^2} = \frac{x}{x^2 + y^2} - i\frac{y}{x^2 + y^2}\,. \tag{1.6}$$

A distributive property holds as well:

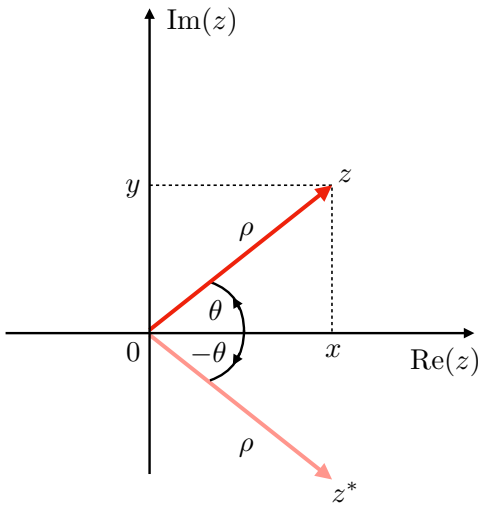
$$z_1 \cdot (z_2 + z_3) \;=\; z_1 \cdot z_2 + z_1 \cdot z_3 \,. \tag{1.7}$$

Exercise 1

Given $z_1 = 3 + 5i$, $z_2 = 2 - 4i$, $z_3 = 3i$, compute the quantities z_1^2 , z_3^3 , $z_2^2 - z_3$, and find the real number a such that $2z_1 - 3z_2 + az_3 = 0$.

1.2 Geometric representations of complex numbers

A complex number $z = x + iy$ can be graphically represented as a point in the complex plane of z, where the the real and imaginary parts of z are associated to the x and y axes, respectively, see plot. This is called *cartesian representation* of z, with x and y representing the projections of z on the two axes.



An equivalent representation is the *polar representation*

$$z \;=\; \rho\, e^{i\theta} \;=\; \rho\,(\cos\theta + i\sin\theta) \;=\; \rho\cos\theta + i\rho\sin\theta\,, \tag{1.8}$$

where we have used Euler formula $e^{i\theta} = \cos\theta + i\sin\theta$. The real number ρ represents the length of the red arrow, and is called the *modulus* (or *absolute value*, or *magnitude*) of z:

$$\rho \;\equiv\; |z| \;=\; \sqrt{x^2 + y^2}\,. \tag{1.9}$$

The real number θ is the angle of the arrow with respect to the real axis, and is called the *argument* (or *phase*) of z:

$$\theta \;\equiv\; \arg(z) \;=\; \arctan(y/x)\,. \tag{1.10}$$

The relations between cartesian and polar coordinates are

$$x \;=\; \rho\cos\theta\,, \qquad y \;=\; \rho\sin\theta\,. \tag{1.11}$$

As cartesian coordinates are suited to ease the computation of *sums* of complex numbers, polar coordinates make it straightforward to compute *products* of complex numbers. Indeed, given $z_1 = \rho_1\, e^{i\theta_1}$ and $z_2 = \rho_2\, e^{i\theta_2}$, their product is $z_1 \cdot z_2 = \rho_1\, e^{i\theta_1} \rho_2\, e^{i\theta_2} = \rho_1\rho_2\, e^{i(\theta_1+\theta_2)}$, namely a complex number whose modulus is the product of the two moduli, and whose phase is the sum of the two phases.
The *complex conjugate* of number z, denoted with z^*, is a complex number with the same real part as z, $\mathrm{Re}(z^*) = \mathrm{Re}(z)$ and opposite imaginary part, $\mathrm{Im}(z^*) = -\mathrm{Im}(z)$; equivalently, it has the same modulus, $|z^*| = |z|$, and opposite phase, $\arg(z^*) = -\arg(z)$:

$$z \,=\, x + iy \,=\, \rho\, e^{i\theta} \qquad \Longrightarrow \qquad z^* \,=\, x - iy \,=\, \rho\, e^{-i\theta}\ . \tag{1.12}$$

Note that $z \cdot z^* = (x+iy)(x-iy) = x^2 + y^2 = |z|^2$ which is a real number, at variance with $z^2 = (x+iy)(x+iy) = x^2 - y^2 + 2ixy$, which is complex.
Note also that, as is clear from the polar representation, increasing the phase of z by 2π just amounts to rotating counterclockwise by a full round angle the arrow representing z, so that the arrow after rotation is the same as before. The same happens if one increases the phase by 4π, 6π, .. or in general by $2n\pi$, with integer n. In formulae

$$z \cdot e^{i2n\pi} \,=\, \rho\, e^{i\theta} e^{i2n\pi} \,=\, \rho\, e^{i(\theta+2n\pi)} = z\,. \tag{1.13}$$

In other words, a single cartesian representation corresponds to (infinitely) many polar representations, all differing by $e^{i2n\pi}$. For example $e^{i0} = e^{i2\pi} = e^{-i16\pi} = 1$.

Exercise 2

Given $z_1 = \sqrt{3} + i$, $z_2 = 5 - 5i$, write down in cartesian and polar representation the following quantities: z_1, z_2, $z_1 \cdot z_2^*$, z_1/z_2.

Exercise 3

Write down $\sin\theta$ and $\cos\theta$ in terms of complex exponentials $e^{\pm i\theta}$, using Euler formula.

Exercise 4

Use the polar representation of complex numbers to compute $\cos(a+b)$ and $\sin(a-b)$ in terms of sines and cosines of a and b.

Exercise 5

Given two complex waves $\psi_1 = a_1\, e^{ib_1}$ and $\psi_2 = a_2\, e^{ib_2}$, with $a_{1,2}, b_{1,2}$ real numbers, compute the intensity I of the superposition wave, defined as $I = |\psi_1 + \psi_2|^2$.

Exercise 6

Use Euler formula, and the properties of complex numbers, to show that

$$\int_A^{A+4\pi} d\theta\, \cos\theta \,=\, 0\,, \qquad \int_A^{A+2\pi} d\theta\, \cos^2\theta \,=\, \pi\,, \qquad \text{for generic } A\,. \tag{1.14}$$

Chapter 2

Basics of dimensional analysis

In physics, one is dealing with measurable quantities carrying their own units of measurement. Specifying the latter corresponds to specifying the *physical dimensions* of the quantities under analysis. For instance, velocity has the dimension of length/time (m/s in the International System of Units (IS)), surface has the dimension of square length (m^2 in the IS), electric current is charge divided by time (C/s = A in the IS), the number 34 is dimensionless, and so on.
Since physical dimensions are an intrinsic feature of physical quantities, equations relating such quantities must be dimensionally consistent, verifying which is the purpose of *dimensional analysis*. Dimensional consistency may seem just a further set of constraints that complicates physics with respect to pure mathematics, however the information carried by such constraints on the nature of the involved quantities can be exploited as a relevant hint towards the solution of physical problems.

The basic rules of dimensional analysis state that:

(a) quantities with different physical dimensions cannot be equal;

(b) quantities with different physical dimensions cannot be added together.

Item (a) is a particular case of the general rule that objects with different features cannot be equal (e.g. a vector cannot be equal to a matrix, as well as an apple cannot be equal to an orange), namely an equality must match all the features of the involved elements. This works for inequalities as well (when saying that an apple is equal to or bigger than an orange, one is implicitly talking about their volume, which is a common feature that can be meaningfully compared).
Item (b) is a particular case of the logical rule that objects with different features cannot be summed together (e.g. a vector plus a number has no mathematical sense, as well as an apple plus an orange does not have logical sense). For instance, the sum of an object having the dimension of a length with one having the dimension of a surface does not have any physical meaning, since the resulting object could not be assigned a specific physical dimension. Note that 0 and ∞ can assume any physical dimension.

The impossibility to add elements with different dimensions has a deep consequence when dealing with functions of dimensionful quantities. Consider quantity x having whatever physical dimension, for instance length, and consider a function $f(x)$, regular at $x = 0$. Being regular, $f(x)$ can be expanded as a Taylor series in powers of x as

$$f(x) \,=\, f_0 + f_1\, x + f_2\, x^2 + \dots\,, \tag{2.1}$$

where f_0, f_1, f_2, are dimensionless coefficients (i.e. pure numbers). As such, the above Taylor expansion does not have any physical meaning, since it is adding together objects with different dimensions (namely, the number f_0 plus the length $f_1\,x$ plus the square length $f_2\,x^2$, and so on). This implies that

(c) the argument of functions must in general be dimensionless (only exception is when the function is a monomial, e.g. $f(x) = x^3$: x can be dimensionful since no addition is performed).

Condition (c) is particularly useful for establishing the dimension of physical quantities: every time a function appears in a physical problem, the involved parameters must organise themselves to give rise to a dimensionless argument. In case this is not enough, one can apply conditions (a) and (b), as well as resort to well known physics equations to assign a physical dimension to the quantities involved in the problem at hand.

Exercise 7

Find the physical dimensions of the vacuum permittivity ϵ_0, of Boltzmann's constant k_B, and of Planck's constant h. Does the latter have the same dimension as angular momentum?

Exercise 8

The potential energy U of a spring is given by $U = \frac{1}{2}kx^2$. Determine the physical dimension of k, being x a length. Since the kinetic energy of the spring is $K = \frac{1}{2}m(dx/dt)^2$, determine by sole dimensional analysis the typical oscillation frequency of the spring.

Exercise 9

Given $a = 1$ m, $b = 100$ m, and

$$P \,=\, \int_0^\infty dx\, 2\, e^{-x/b} + \int_0^a dx\, \frac{1}{4 + 2\sin(x/a) + \cos^{3/2}(x/a)}\,, \tag{2.2}$$

estimate the value of P with $\sim 1\%$ accuracy just by means dimensional analysis and trivial computations.

Chapter 3

Linear algebra

As we will see in the course, the mathematical language of quantum mechanics is complex linear algebra. Linear algebra is a very important part of Mathematics which is a crucial tool in many sciences, from Physics and Chemistry to Statistics, Engineering and Computer Science. It is a mathematical formalism that concerns two basic entities: *vectors* and *matrices* (also called *linear operators*).

In this chapter, we will first introduce the basic notions using examples from the properties of familiar geometrical vectors. Then, we will pass to an abstract generalisation of the concepts of *vector*, *vector space*, *scalar product* and *linear operator*.[1]

We will introduce in particular *complex vector spaces* with a complex scalar product, also known as *Hilbert spaces*: these are the arena of Quantum Mechanics. We will also learn the crucial concept of *Hermitian matrix*, and how to *diagonalize* a matrix.

3.1 Example of 2D geometrical vectors

3.1.1 Vectors and scalar products

We start by considering the example of vectors in 2D Cartesian geometry. They are arrows connecting the origin to a given point in the plane. We can represent any such vector as a **linear combination** of the unit vectors $\vec{e}_1$ and $\vec{e}_2$, parallel to the two axes, i.e. any vector can be written as

$$\vec{v} = \alpha\vec{e}_1 + \beta\vec{e}_2, \quad \alpha, \beta \in \mathbb{R}. \tag{3.1}$$

for two real numbers (α, β).

We call the vectors $\{\vec{e}_1\, \vec{e}_2\}$ a **basis** for the vectors in the plane: this means that we can write **any** vector $\vec{v}$ in the form above (with some appropriate coefficients α and β, which are called the *components* of the vector $\vec{v}$ in the given basis). In the following, we will discuss more about the important concept of basis. At the moment, it is important to notice that the choice of a basis is not unique: we could for example rotate the axis by a given angle and choose two new reference vectors $\{\vec{e}'_1\, \vec{e}'_2\}$, corresponding to the rotation of $\vec{e}_1$ or $\vec{e}_2$.

[1] In fact, we will see that in Quantum Mechanics an abstract *vector* will represent the state of the system.

Column vector notation. Since the coefficients α, β completely specify the vector $\vec{v}$, we can also express it just as a *column vector* giving the list of these two numbers, i.e.,

$$\vec{v} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} . \tag{3.2}$$

Notice that, similarly, we can write the basis vectors as

$$\vec{e}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} , \qquad \vec{e}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} . \tag{3.3}$$

Notice that the coefficients (α, β) depend on a choice of the basis vectors. We will discuss changes of bases later (see in particular the exercise 14 below). For the moment, we will keep the basis fixed and represent vectors in components, in order to understand their operations.

Linear combinations of vectors. Vectors in the plane have two important properties.

First, we can multiply them by any real number λ. This gives a vector with redefined length, but parallel to the original vector. It is obtained by multiplying all the components by the same real number:

$$\vec{v} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \to \lambda\vec{v} = \begin{pmatrix} \lambda\alpha \\ \lambda\beta \end{pmatrix} . \tag{3.4}$$

Secondly, we can take the sum of vectors. This is obtained with the familiar *parallelogram rule*. In components, we simply take the sum of each component. Suppose we want to take the sum of two vectors,

$$\vec{v} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} , \text{ and } \vec{w} = \begin{pmatrix} \gamma \\ \delta \end{pmatrix} , \tag{3.5}$$

then the result will be

$$\vec{v} + \vec{w} = \begin{pmatrix} \alpha + \gamma \\ \beta + \delta \end{pmatrix} . \tag{3.6}$$

We can combine the two operations and take a generic linear combination of two vectors with generic coefficients.

Notice that there is a special element of the space which is the vector with all components zero. This vector is called the **null vector** or **zero vector**; it is the equivalent of the number 0 for ordinary algebra.

Exercise 10

Consider two vectors $\vec{v}_1 = \vec{e}_1 + 4.2\,\vec{e}_2$ and $\vec{v}_2 = 3\vec{e}_1 + 7\vec{e}_2$. Compute the vector $\vec{v}_3 = 2\vec{v}_1 - \vec{v}_2$. Express the result in terms of the original basis vectors, and write its equivalent form as a column vector.

The concept of linear dependence and independence. We have seen that the couple of vectors $\{\vec{e}_1, \vec{e}_2\}$ form a basis of the space, i.e. any vector can be written in a unique way as their combination. It is no accident that the basis has 2 elements, which is the same as the dimension of the space. We want to understand why the number of elements in a basis is fixed, and how this is linked to the important concept of **linear independence**.

To see this, let us try to hypothetically build a basis made of 3 elements, by adding a third vector to the basis:

$$\left\{\vec{e}_1, \vec{e}_2, \vec{f}\right\}. \tag{3.7}$$

Would this also be a basis of the 2D vector space? No, because the decomposition of a 2D vector $\vec{v}$ into a linear combination *would not be unique*. To explain why this is the case, consider an example, e.g. $\vec{f} = \vec{e}_1 + \vec{e}_2$, where we used the fact that we already know that any third vector can be written as a linear combination of $\vec{e}_1$ and $\vec{e}_2$. Then, for any new vector $\vec{v} = v_1\vec{e}_1 + v_2\vec{e}_2$, we can write a non-unique combination

$$\vec{v} = \gamma\vec{f} + (v_1 - \gamma)\vec{e}_1 + (v_2 - \gamma)\vec{e}_2, \tag{3.8}$$

where the coefficients depend on an *arbitrary constant* γ. Thus, the set $\left\{\vec{e}_1, \vec{e}_2, \vec{f}\right\}$ is not a basis since it has too many elements, which implies that the decomposition is not unique. The precise statement is that the vectors of a basis should be **linearly independent**: a set of vectors is called **linearly independent** if no vector in the set is a linear combination of the other vectors. A basis must be made of linearly independent vectors.

On the other hand, consider trying to remove elements from a basis, e.g. consider only the set

$$\mathcal{B}' = \{\vec{e}_1\}. \tag{3.9}$$

This basis generates only a **subset** of the full 2D vector space, because it cannot generate any vector $\vec{v} = v_1\vec{e}_1 + v_2\vec{e}_2$ with $v_2 \neq 0$. A basis that can generate a full vector space is called a **complete** basis. The number of elements of a complete basis is the maximal number of linearly independent vectors which can be found in the space. This is also called the **dimension** of the space, and in our case, this number is 2 and coincides with our intuition of the geometrical dimension. The reduced basis $\mathcal{B}'$ can be seen as a complete basis of a subspace, which is the 1D set of vectors of the form $\alpha\vec{e}_1$, $\alpha \in \mathbb{R}$. As another intuitive examples of a subspace, think of a plane in 3D space: the 3D space is generated by a basis of 3 vectors, while the plane is a subspace generated by a smaller basis of just 2 elements.

Summary so far. We give a brief summary of the intuitions we have collected so far, suggesting their generalisations which will be useful later.

- Vectors are mathematical objects that can be combined by taking their linear combination (this is the *mathematical definition* of what vectors are). The space containing all possible linear combinations of a given group of vectors is called a *vector space*.
- A **basis** is a set of vectors such that any vector of the space can be decomposed **uniquely** as linear combination of the basis elements.
- The elements of a basis must be **linearly independent**. This means that no element can be obtained as linear combination of the other elements.

- The number of elements of a complete basis is called the **dimension** of the vector space.
- Vectors can always be represented as N-dimensional lists of numbers, with N equal to the dimension of the space. This list of numbers represents the components of the vector decomposed in a given basis. Given the basis, the components identify uniquely the vector.
- Vectors can also have a certain concrete interpretation, e.g. as geometrical vectors in the example above. The interpretation can be much more abstract, but the mathematics is however always the same.

Scalar product. Another important operation on vectors in the plane is the scalar product. It takes two vectors and gives back a number. Consider two vectors, with components in the usual Cartesian basis $\{\vec{e}_1\, \vec{e}_2\}$ given by,

$$\vec{v} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}, \text{ and } \vec{w} = \begin{pmatrix} w_1 \\ w_2 \end{pmatrix}. \tag{3.10}$$

Then their scalar product is defined by

$$\vec{v} \cdot \vec{w} = \sum_{i=1}^{2} v_i w_i = v_1 w_1 + v_2 w_2. \tag{3.11}$$

The scalar product has an important mathematical property: it is *linear* with respect to the two arguments. This means that,

$$(\lambda \vec{v}) \cdot \vec{w} = \lambda\, \vec{v} \cdot \vec{w}, \quad (\vec{v}_1 + \vec{v}_2) \cdot \vec{w} = \vec{v}_1 \cdot \vec{w} + \vec{v}_2 \cdot \vec{w}, \tag{3.12}$$

where $\vec{v}_1$, $\vec{v}_2$, $\vec{v}$, $\vec{w}$ are arbitrary vectors and $\lambda \in \mathbb{R}$ is an arbitrary real number.

The square root of the scalar product of a vector with itself is called the **norm** of the vector. In the case of geometrical vectors it measures their square length.

$$\vec{v}.\vec{v} \equiv ||\vec{v}||^2. \tag{3.13}$$

The scalar product between two different vectors, instead measures the angle between them:

$$\vec{v} \cdot \vec{w} = ||\vec{v}||\, ||\vec{w}|| \cos(\theta), \tag{3.14}$$

where θ is the angle between $\vec{v}$ and $\vec{w}$. Vectors forming a 90° angle are called orthogonal, and this means that their scalar product is zero. Generalising this notion, in more abstract vector spaces **two vectors are called orthogonal if their scalar product vanishes**. The formula (3.14) implies an important inequality, that also holds also in more abstract situations where we lose the immediate geometrical intuition:

$$|\vec{v} \cdot \vec{w}|^2 \leq ||\vec{v}||^2 ||\vec{w}||^2. \tag{3.15}$$

This is called the Schwarz inequality. Finally, notice that the scalar products between elements of the Cartesian basis are

$$\vec{e}_i \cdot \vec{e}_j = \delta_{ij}, \quad i = 1, 2, \tag{3.16}$$

where δ_{ij} denotes the Kronecker delta, i.e. it returns 1 for $i = j$ and zero otherwise. Eq. (3.16) means that the basis vectors $\vec{e}_1$, $\vec{e}_2$ are orthogonal and have unit norm. A set of vectors having the properties (3.16) is called a **orthonormal basis**, if their number is equal to the dimension of space, in this case 2. Orthonormal bases are special: they are the easiest to work with, because they encode the scalar product in the very simple expression (3.11). The expression would be the same in any other orthonormal basis, but would be different if we had chosen a non-orthogonal basis.

Exercise 11

Suppose $\vec{v}$, $\vec{w}$, $\vec{z}$ are three vectors, such that $\vec{v} \cdot \vec{w} = 3$ and $\vec{v} \cdot \vec{z} = 5$.

- Compute the scalar product between $\vec{v}$ and $\vec{y} \equiv \lambda_1 \vec{w} + \lambda_2 \vec{z}$ as a function of the two coefficients λ_1 and $\lambda_2 \in \mathbb{R}$.
- What are the possible values of λ_1 and λ_2 such that $\vec{v}$ and $\vec{y}$ are orthogonal?

Exercise 12

Consider a vector $\vec{v} = \alpha \vec{e}_1 + \beta \vec{e}_2$ with two generic coefficients α, $\beta \in \mathbb{R}$. Construct a second vector $\vec{w}$ that is orthogonal to $\vec{v}$ for any α, β. Use this result to construct a new orthonormal basis where one of the elements is proportional to $\vec{v}$.

3.1.2 Matrices: linear transformations of vectors

The example of rotations. Consider the transformation R_φ that rotates vectors by an angle φ. This rotation will send the basis vectors to

$$R_\varphi : \vec{e}_1 \quad \rightarrow \quad \vec{e}\,'_1 = \cos\varphi \vec{e}_1 + \sin\varphi \vec{e}_2, \tag{3.17}$$

$$R_\varphi : \vec{e}_2 \quad \rightarrow \quad \vec{e}\,'_2 = -\sin\varphi \vec{e}_1 + \cos\varphi \vec{e}_2. \tag{3.18}$$

Similarly, it will send a general vector $\vec{v} = v_1 \vec{e}_1 + v_2 \vec{e}_2$ to

$$\vec{v} \rightarrow (R_\varphi \circ \vec{v}) = v_1 \vec{e}\,'_1 + v_2 \vec{e}\,'_2. \tag{3.19}$$

This property follows from the fact that a rotation acts **linearly** on vectors.

Notice that this transformation can be represented in terms of a 2×2 **matrix**:

$$R_\varphi \equiv \begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix}, \tag{3.20}$$

so that the components of the transformed vector v are given explicitly by

$$R_\varphi \circ \vec{v} = \vec{v}' = \begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix} \cdot \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} \cos\varphi v_1 - \sin\varphi v_2 \\ \sin\varphi v_1 + \cos\varphi v_2 \end{pmatrix}. \tag{3.21}$$

Exercise 13

Verify that this rule gives the correct result for the basis vectors $\vec{e}_1$ and $\vec{e}_2$.

Exercise 14

Consider the vector $\vec{v} = \vec{e}_1 + 2\vec{e}_2$. Find its coefficients as a linear combination in the rotated basis. i.e. find the constants α, β such that

$$\vec{v} = \alpha\vec{e}'_1 + \beta\vec{e}'_2, \tag{3.22}$$

where $\vec{e}'_1$, $\vec{e}'_2$ are the rotated vectors defined in (3.17). This is a very important exercise as it teaches us to change between one basis and another to represent the same vector, and this is extremely useful. The problem can be solved by noticing that (3.22), together with the definition of $\vec{e}'_i$, gives a linear system of equations for the coefficients α, β.

In the solution of this exercise in section 5, it is also explained how to solve this linear system easily using the technology of matrix inversion, explained below in section 3.3.2.

Exercise 15

An example of linear transformation A in $N = 3$ dimensions is given by a rotation around the z axis by an angle θ:

$$A = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{3.23}$$

Verify that this indeed rotates 3-dimensional vectors in the expected way. What would be the form of a matrix representing rotations around the x-axis? And about the y-axis?

Definition of general matrix multiplication. Above, in (3.21) we have assumed familiarity with the notion of multiplication between a matrix and a vector. Let us give the precise definition of the multiplication between two matrices. The multiplication of a matrix and a vector is a special case, since a m-dimensional column vector can be seen as a $m \times 1$ matrix.

Consider two matrices A and B, of dimensions respectively $m \times n$ and $n \times k$. We will now define their product $A \cdot B$. Notice that, in order to define their multiplication, the number of columns in the first matrix should be the same as the number of rows in the second matrix - other wise they cannot be multiplied! The product is defined as

$$A = \begin{pmatrix} A_{11} & A_{12} & \dots & A_{1n} \\ A_{21} & A_{22} & \dots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m1} & A_{m2} & \dots & A_{mn} \end{pmatrix}, \qquad B = \begin{pmatrix} B_{11} & B_{12} & \dots & B_{1k} \\ B_{21} & B_{22} & \dots & B_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ B_{n1} & B_{n2} & \dots & B_{nk} \end{pmatrix} \tag{3.24}$$

Then, the product $A \cdot B$, defines a new matrix of dimension $m \times k$ defined as

$$A \cdot B \equiv C = \begin{pmatrix} C_{11} & C_{12} & \dots & C_{1k} \\ C_{21} & C_{22} & \dots & C_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ C_{m1} & C_{m2} & \dots & C_{mk} \end{pmatrix}, \tag{3.25}$$

with elements given by

$$C_{ij} = \sum_{l=1}^{n} A_{il} B_{lj}, \quad \text{for } 1 \leq i \leq m, \;\; 1 \leq j \leq k. \tag{3.26}$$

This is also called the rule of multiplication row-by-column.

Exercise 16

Compute the product $A \cdot B$ between the following two matrices:

$$A = \begin{pmatrix} 3 & 0 & -2 \\ 1 & -1 & -4 \end{pmatrix}, \qquad B = \begin{pmatrix} 1 & 9 & 10 \\ 0 & -7 & 0 \\ 6 & 0 & 0 \end{pmatrix}. \tag{3.27}$$

Note. The scalar product between two n-dimensional vectors can be written, using the rules of matrix multiplication, as

$$\vec{v}.\vec{w} = (\vec{v})^T \cdot (\vec{w}) = \begin{pmatrix} v_1 & v_2 & \dots & v_N \end{pmatrix} \cdot \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_N \end{pmatrix} \tag{3.28}$$

as a matrix product between the row-vector $(\vec{v})^T$, obtained by **transposing** $\vec{v}$, and $\vec{w}$.
Definition of transposed matrix. In general, we define the **transposed** of a $m \times n$ matrix A, denoted by A^T, as the matrix with elements

$$(A^T)_{ij} \equiv A_{ji}.$$

Notice that if A is an $m \times n$ matrix, then A^T is $n \times m$.

The transpose has a very important property: under the multiplication of two matrices it behaves as

$$(A \cdot B)^T = B^T \cdot A^T. \tag{3.29}$$

Notice that it changes the order of the matrices in the product!
Note. Notice that matrices have a very important difference as compared to ordinary numbers: suppose matrices A and B have dimensions such that they can be multiplied in any order. Then, it can happen that

$$A \cdot B \neq B \cdot A! \tag{3.30}$$

An useful object to quantify this mismatch is the **commutator**

$$[A, B] \equiv A \cdot B - B \cdot A, \tag{3.31}$$

which is generally different from zero. In some special cases, it can happen that $[A, B] = 0$. In this case, we say that A and B **commute**.

Exercise 17

Consider the two matrices

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{3.32}$$

Compute the **commutator**

$$[\sigma_1, \sigma_3] \equiv \sigma_1 \cdot \sigma_3 - \sigma_3 \cdot \sigma_1. \tag{3.33}$$

and show that σ_1 and σ_3 **do not commute**.

Matrices as linear maps. Let us come back to the example of vectors in 2D space. We have introduced a specific matrix (3.21) and shown that it represents a rotation in this space. We could also consider a more general matrix, parametrised by more parametrs, e.g. a 2D matrix

$$M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \tag{3.34}$$

with some coefficients a, b, c, d. By considering the matrix multiplication with vectors, we see that this matrix defines a certain more general **linear map** which sends vectors to vectors:

$$\vec{v} \to \vec{v} = M \cdot \vec{v}. \tag{3.35}$$

The adjective "linear" means, as usual, that the map respects linear combinations

$$M \cdot (\lambda_1 \vec{v}_1 + \lambda_2 \vec{v}_2) = \lambda_1 M \cdot \vec{v}_1 + \lambda_2 M \cdot \vec{v}_2. \tag{3.36}$$

Now consider two matrices, M_1 and M_2. Each represents a linear map. The matrix $M_1 \cdot M_2$ represents the composition of the maps, namely it gives the result of performing first the map M_2 and then the map M_1 on the resulting vector.

Exercise 18

Verify that the map sending vectors to themselves is given uniquely by a matrix

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{3.37}$$

This is called the identity matrix.

Exercise 19

Verify that by multiplying two rotation matrices of angles φ_1 and φ_2 in 2D, one obtains a rotation matrix corresponding to the angle $\varphi_1 + \varphi_2$.

The special case of orthogonal matrices. What do rotation matrices have special in 2D? They are the only matrices that leave the scalar product unchanged.

Such a matrix is called an **orthogonal matrix**, and should satisfy

$$M^T M = I, \tag{3.38}$$

where T denotes the **transposed matrix**, where we swap rows and columns: $(M^T)_{ij} = (M)_{ji}$.

In fact, let us consider two vectors $\vec{v}$, $\vec{w}$. The transformed vectors under the map are $\vec{v}' = M \circ \vec{v}$, $\vec{w}' = M \circ \vec{w}$. The requirement that the scal;ar product of the original vectors is the same as the scalar product of the transformed vectors reads

$$(M \circ \vec{v})^T \cdot (M \circ \vec{w}) = (\vec{v})^T \cdot (\vec{w}). \tag{3.39}$$

This condition is true for every choice of $\vec{v}$ and $\vec{w}$ if and only if the condition (3.38) is satisfied.

In the case of complex vector spaces that we consider in Quantum Mechanics, the concept of orthogonal matrices will be replaced by the one of **unitary matrix**, which we will discuss later.

Exercise 20

Verify that rotation matrices satisfy the identity (3.38), i.e. they are orthogonal.

Summary on linear operators.

- Maps from a vector space to itself which act linearly on vectors are called **linear operators**.
- In a given basis, linear operators can be represented by $N \times N$ matrices, where N is the dimension of the vector space.
- Matrices can be multiplied using the rules of row-by-column multiplication. The composition of two linear operators is represented by the product of the corresponding matrices.
- Rectangular matrices, of size $N_1 \times N_2$ with $N_1 \neq N_2$ represent linear maps between two vector spaces of different dimensions.

3.2 Complex vector spaces

In the previous section, we have seen some properties of vector spaces in a concrete example of 2D vectors. It is already becoming clear that we can generalize our notion of vector space in various ways:

- We can consider vector spaces of **different dimensions**. For example, think of vectors in 3D rather than 2D. However, we can consider easily an arbitrary dimension N (and even infinite dimension!).
- We can consider *abstract* vectors which no longer need to have a connection to vectors in the ordinary space around us.
- We can allow ourselves to take linear combinations with **complex numbers instead of real numbers**. In this case, we use a certain complex version of scalar product.

In quantum mechanics, we need all these generalisations. We use complex vector spaces with a complex scalar product, where the vectors have an interpretation as "states of the system" (which we will discuss in the course). Before discussing the generalization to complex vector spaces, let us make a digression, to familiarize ourselves to the fact that vectors can represent abstract objects.

3.2.1 Digression: vectors with abstract interpretations

Example 1: vectors as a list of different features. A botanist wants to select flowers based on how similar they are to a reference flower. She is interested in 3 core characteristics: length of the flower, width of the petals, shade of yellow. The botanist knows how to assign scores to every flower based on these 3 features, such that (in the chosen units) the reference flower has a score $(1, 1, 1)$.

The botanist has to decide a way to measure how close a given flower (say, with the score (f_1, f_2, f_3)) is to the reference flower. Then, the botanist defines a function that "measures this distance between two flowers", for example given by the norm[2]

$$\text{dist}(\vec{f}, (1, 1, 1)) \equiv \sqrt{(f_1 - 1)^2 + (f_2 - 1)^2 + (f_3 - 1)^2}. \tag{3.40}$$

Given three flowers with the characteristics:

$$\texttt{Flower}_1 = (0.5, 2.1, 1.1), \quad \texttt{Flower}_2 = (0.7, 1.1, 1.9), \quad \texttt{Flower}_3 = (0.9, 1.2, 1.1),$$

which one is the closest to the reference one? You can verify that, according to the metric (3.40), the botanist would select the 3rd flower.

Example 2: use of matrices and vectors in an economical model [*This example is taken from the Wikipedia page on matrix multiplication*]

A fictitious factory uses 4 kinds of basic ingredients, b_1, b_2, b_3, b_4, to produce 3 kinds of intermediate goods, m_1, m_2, m_3, which in turn are used to produce 3 kinds of final products, f_1, f_2, f_3. The matrices

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

provide the amount of basic commodities needed for a given amount of intermediate goods, and the amount of intermediate goods needed for a given amount of final products, respectively.

[2] The botanist could decide to give different importance to the 3 features, for example defining $\text{dist}(\vec{f}, (1, 1, 1)) \equiv \sqrt{w_1(f_1 - 1)^2 + w_2(f_2 - 1)^2 + w_3(f_3 - 1)^2}$, where $w_i \geq 0$ are some weights which measure how important a given feature is. This would correspond to a different definition of scalar product, and it can be brought to the framework of the standard definition just by rescaling the components.

For example, the first column of $\mathbf{A}$ tells us that, to produce one unit of intermediate good m_1, we would need one unit of basic commodity b_1, two units of b_2, no units of b_3, and one unit of b_4.

The matrix

$$\mathbf{A}\cdot\mathbf{B} = \begin{pmatrix} 5 & 4 & 3 \\ 8 & 9 & 5 \\ 6 & 5 & 3 \\ 11 & 9 & 6 \end{pmatrix};$$

directly provides the amounts of basic commodities needed for given amounts of final goods. For example, the bottom left entry of $\mathbf{A}\cdot\mathbf{B}$ is computed as $1\cdot 1 + 1\cdot 2 + 2\cdot 4 = 11$, reflecting that 11 units of b_4 are needed to produce one unit of f_1. Indeed, one b_4 unit is needed for m_1, one for each of 2 m_2, and 2 for each of the 4 m_3 units that go into the f_1 unit.

In order to produce e.g. 100 units of the final product f_1, 80 units of f_2, and 60 units of f_3, the necessary amounts of basic goods can be computed as

$$(\mathbf{A}\cdot\mathbf{B})\cdot\begin{pmatrix} 100 \\ 80 \\ 60 \end{pmatrix} = \begin{pmatrix} 1000 \\ 1820 \\ 1180 \\ 2180 \end{pmatrix},$$

that is, 1000 b_1, 1820 b_2, 1180 units of b_3, 2180 units of b_4 are needed. Similarly, the product matrix $\mathbf{A}\cdot\mathbf{B}$ can be used to compute the needed amounts of basic goods for other final-good amount data.

Example 3: vector spaces of functions. The two previous examples illustrated some abstract situations where it might be useful to use vectors and matrices. However, in the above examples one crucial property of vector spaces - the possibility to take linear combination of vectors - did not play a crucial role. Let us give now an example which fits perfectly the mathematical definition of vector space: the concept of vector space of functions.

Consider the following space

$$\mathcal{C}([0,1]) = \{\texttt{set of all continuous real functions on the interval } [0,1]\}\,.$$

- For any real function $f(x)\in\mathcal{C}([0,1])$, $\lambda f(x)$ is still a real continuous function for any constant $\lambda\in\mathbb{R}$.
- For any two functions $f_1(x)$ $f_2(x)\in\mathcal{C}([0,1])$, their sum $f_1(x)+f_2(x)$ is still a function in $\mathcal{C}([0,1])$.

The two properties above tell us that $\mathcal{C}([0,1])$ can be understood as a vector space! We can generalise also the concept of linear operators and of scalar product. For example, a possible definition of scalar product between two functions is given by

$$\langle f_1, f_2\rangle = \int_0^1 f_1(x) f_2(x) dx. \tag{3.41}$$

Notice that the space of functions is a continuous space - in particular, we cannot expect to find a basis with a finite number of elements. This is in fact a vector space of **infinite dimension**! Also in infinite-dimensional vector spaces, when some conditions are satisfied, it still makes sense to talk of a basis. The basis will be infinite, and vectors can be represented as infinite lists of numbers - the components in a given basis.

In general, in quantum mechanics we are going to need vector spaces of infinite dimension to represent the space of a system. In fact, in many setups we will use spaces of functions similar

to the one of the example above (although they will be a bit different[3]) to the one described in this example. In quantum mechanics these functions $f(x)$, called wave functions, will have a very particular meaning.

Now we introduce a very important generalization of all features seen until now (which is also crucial in quantum mechanics) - we start considering complex vectors.

3.2.2 Complex vector spaces.

Vectors in a N-dimensional complex vector space can be represented as **lists of N complex numbers**. We can take linear combinations of vectors with coefficients in $\mathbb{C}$, and this gives a vector in the same space. Any vector can be decomposed in terms of a canonical basis of vectors $\vec{e}_i$, for $i = 1, ..., N$, where for instance we can take

$$\vec{e}_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \qquad \vec{e}_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \qquad \cdots, \qquad \vec{e}_N = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}, \tag{3.42}$$

so that in general

$$\vec{v} = \sum_{i=1}^{N} v_i \, \vec{e}_i = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_N \end{pmatrix}, \tag{3.43}$$

where $v_i \in \mathbb{C}$ is the component of $\vec{v}$ along the i-th direction described by $\vec{e}_i$.

The concepts of **basis** and **linear independence** are the same as in the real case - with the obvious differences due to the use of complex vectors. We repeat here the discussion more generally. Let us call the space of complex vectors described above

$$\mathcal{H} = \{\text{set of complex linear combinations of } \vec{e}_i, \;\; i = 1, \ldots, N\}.$$

By definition, all vectors of $\mathcal{H}$ can be written uniquely as linear combinations of elements of the basis $\vec{e}_i$: we say that the basis $\{e_i\}$ forms a **complete basis** of the space, or equivalently that it **spans** or **generates** the whole $\mathcal{H}$. The number of vectors in a complete basis is the **dimension** of the vector space: in this case, N is the dimension of $\mathcal{H}$, since we cannot generate the same space of vectors using a basis of less than N elements.

[3] A warning for the readers interested in mathematical precision: the space of states used in quantum mechanics will be mathematically different from the space $\mathcal{C}([0,1])$ described here - not only because in quantum mechanics we will use complex numbers rather than only real numbers when taking linear combination. A key difference is that in quantum mechanics we will need an additional mathematical condition (relevant in the infinite-dimensional space) making the space into a so-called "Hilbert space" - and $\mathcal{C}([0,1])$ does not satisfy this property. A Hilbert space is a complex vector space with a scalar product which is "complete" in a certain sense under the convergence of infinite series. In order to make $\mathcal{C}([0,1])$ complete in this sense, we would need to enlarge it, to also include functions which have isolated points of discontinuity. This is due to the fact that there exist series of continuous functions which can converge to discontinuous functions - therefore also the latter should be included. The resulting space, obtained by "completing" $\mathcal{C}([0,1])$ with the addition of pointwise discontinuous functions, is called by mathematicians $L^2([0,1])$; you can find its precise definition in formal books on quantum mechanics, or books in mathematical analysis. Notice that there are no such subtleties in the finite-dimensional case: an N-dimensional complex vector space with a complex inner product is always by definition a full-fledged Hilbert space.

Another important concept, which we repeat, is that of **subspace**. For instance, the set of $\vec{e}_i$, with $i = 1, \cdots, N-1$, is not a complete basis for $\mathcal{H}$, since vectors with a component along the N-th direction cannot be written as linear combinations of elements of that set alone. Instead, this reduced basis is a complete basis for a *subspace* of $\mathcal{H}$ containing all vectors with null N-th component.

Exercise 21

Consider the two complex vectors $\vec{v}_1 = \vec{e}_1 + (2+3i)\vec{e}_2$ and $\vec{v}_2 = (7-i)\vec{e}_1 + 2i\vec{e}_2$.

- Compute $\vec{w} = \vec{v}_1 + \vec{v}_2$.
- Express the vector $\vec{w}$ in a new basis $\left\{\vec{f}_1, \vec{f}_2\right\}$, where $\vec{f}_1 = \vec{e}_1 + \vec{e}_2$ and $\vec{f}_2 = \vec{e}_1 - 2\vec{e}_2$.

Complex scalar product. We now would like to define a complex version of the scalar product on $\mathcal{H}$, which we will also call a complex **inner product**.[4] It is useful to first define

$$\vec{v}^\dagger = (v_1^*, \cdots, v_N^*),$$

the so-called **hermitian conjugate** of vector $\vec{v}$, which is obtained by transposing the vector and taking the conjugate of its elements. In particular, for the basis elements,

$$\begin{aligned} \vec{e}_1^\dagger &= (1, 0, \cdots, 0)\,, \\ \vec{e}_2^\dagger &= (0, 1, \cdots, 0)\,, \\ &\cdots \\ \vec{e}_N^\dagger &= (0, 0, \cdots, 1)\,, \end{aligned} \tag{3.44}$$

and the row vector can be decomposed as

$$\vec{v}^\dagger = \sum_{i=1}^{N} v_i^*\, \vec{e}_i^\dagger\,. \tag{3.45}$$

The scalar product between $\vec{v}$ and $\vec{w}$ is defined as

$$\vec{v}^\dagger \cdot \vec{w} = \sum_{i=1}^{N}\sum_{j=1}^{N} v_i^* w_j\, \vec{e}_i^\dagger \cdot \vec{e}_j = (\vec{w}^\dagger \cdot \vec{v})^* \quad \in \mathbb{C}\,, \tag{3.46}$$

where we have decomposed both vectors $\vec{v}$, $\vec{w}$ on the canonical basis.
The basis $\vec{e}_i$ is **orthonormal** , namely the scalar product of basis elements is

$$\vec{e}_i^\dagger \cdot \vec{e}_j = \begin{cases} 1 & \text{if} \quad i = j\,, \\ 0 & \text{if} \quad i \neq j\,, \end{cases} \qquad \equiv \delta_{ij} = \delta_{ji} \quad (\text{Kronecker } \delta)\,, \tag{3.47}$$

[4] While "scalar product" is the term used for geometrical vectors, for an abstract vector space it is more common to use the term "inner product". They are anyway interchangeable terms

just as we saw before. Notice that, as compared to the scalar product in real vector spaces, the complex scalar product does not treat the two vectors on the same footing (since one is conjugated, the other is not). This means that the complex scalar product is linear in the second argument,

$$\vec{v}^{\,\dagger} \cdot (\lambda_1 \vec{w}_1 + \lambda_2 \vec{w}_2) = \lambda_1 \; \vec{v}^{\,\dagger} \cdot \vec{w}_1 + \lambda_2 \; \vec{v}^{\,\dagger} \cdot \vec{w}_2, \tag{3.48}$$

but is *anti-linear* in the first argument, i.e.

$$(\lambda_1 \vec{v}_1 + \lambda_2 \vec{v}_2)^{\,\dagger} \cdot \vec{w} = \lambda_1^* \; \vec{v}_1^{\,\dagger} \cdot \vec{w} + \lambda_2^* \; \vec{v}_2^{\,\dagger} \cdot \vec{w}, \tag{3.49}$$

for arbitrary vectors and complex constants λ_1, $\lambda_2 \in \mathbb{C}$. The scalar product allows to express the squared norm of a vector as

$$||\vec{v}||^2 \equiv \vec{v}^{\,\dagger} \cdot \vec{v} \; = \; \sum_{i=1}^{N} v_i^* v_i \; = \; \sum_{i=1}^{N} |v_i|^2 \geq 0 \,, \tag{3.50}$$

whose non-negativity is implied by the presence of a complex-conjugate row vector in the definition of scalar product (i.e. defining the scalar product as $\vec{v}^{\,\mathrm{T}} \cdot \vec{v}$ would result in negative norms for certain $\vec{v}$! This is why the definition with complex conjugation is so important).

It can be shown that the scalar product satisfies the same Schwarz inequality that we already encountered:

$$|\vec{v}^{\,\dagger} \cdot \vec{w}|^2 \; \leq \; ||\vec{v}||^2 ||\vec{w}||^2 \,. \tag{3.51}$$

We can now introduce a very important nomenclature: a N-dimensional complex vector space with a complex inner product is called a **Hilbert space**.

Exercise 22

Constructing a orthonormal basis. Construct a orthonormal basis of 2-dimensional complex vector space, such that one of the basis vectors is proportional to $\vec{f} = (\vec{e}_1 + (3+i)\vec{e}_2)$. You can do this in two steps

- First, find a second vector that is orthogonal to $\vec{f}$. To do this, take a generic vector $\vec{f'} = \alpha \vec{e}_1 + \beta \vec{e}_2$, and constrain the coefficients so that it is orthogonal to $\vec{f}$
- Form the new basis as $\left\{ \lambda \vec{f}, \; \kappa \vec{f'} \right\}$, where the constants λ and κ should be chosen to ensure that the basis elements have unit norm.

3.2.3 The completeness relation

There is a very convenient way to decompose a vector in an orthonormal basis. The trick is the observation that **the components of a generic vector $\vec{v}$ in an orthonormal basis can be computed as scalar products**:

$$\vec{e}_j^{\,\dagger} \cdot \vec{v} \; = \; \sum_{i=1}^{N} v_i \, \vec{e}_j^{\,\dagger} \cdot \vec{e}_i \; = \; \sum_{i=1}^{N} v_i \, \delta_{ji} \; = \; v_j \,, \tag{3.52}$$

where we used the fact that $\vec{e}_i^{\,\dagger} \cdot \vec{e}_j = \delta_{ij}$. This is very useful in practice when we want to decompose a known vector in a new orthonormal basis. You can see it directly in the following exercise.

Exercise 23

Consider the vectors $\vec{v} = -\vec{e}_1 + (2+3i)\vec{e}_2$, $\vec{f_1} = \frac{1}{\sqrt{5}}(\vec{e}_1 + 2e^{i\frac{\pi}{3}}\vec{e}_2)$ and $\vec{f_2} = \frac{1}{\sqrt{5}}(-2e^{-i\frac{\pi}{3}}\vec{e}_1 + \vec{e}_2)$. where $\vec{e}_1$ and $\vec{e}_2$ form a orthonormal basis.

- Verify that the two vectors $\vec{f_1}$ and $\vec{f_2}$ form a new orthonormal basis.
- Express the vector $\vec{v}$ in this new basis. Since it is an orthonormal basis, you can do this using the rule explained above. Verify that this gives the correct decomposition.

There is a very useful and elegant way to write the property above, which is called **completeness relation** or **resolution of the identity**.

$$I = \sum_{i=1}^{N} (\vec{e}_i)(\vec{e}_i)^\dagger. \tag{3.53}$$

In fact, notice that this defines an identical transformation when acting on any vector:

$$\left(\sum_{i=1}^{N} (\vec{e}_i)(\vec{e}_i)^\dagger \right) \cdot \vec{v} = \sum_{i=1}^{N} (\vec{e}_i)(\vec{e}_i^\dagger \vec{v}) = \vec{v}, \tag{3.54}$$

where the last equality follows from the fact that, as we saw before, $\vec{e}_i^\dagger \vec{v} = v_i$ gives the component of the decomposition of the vector $\vec{v}$ in the basis.
The identity (3.53) might seem a bit magical at first. Notice that it is true for *any* choice of orthonormal basis. In the following exercise, you will verify that it actually works!

Exercise 24

- Verify the completeness relation using the canonical basis elements as column vectors in a complex space of dimension 2.
- Verify the completeness relation for $\vec{f_1} = \frac{1}{\sqrt{5}}(\vec{e}_1 + 2e^{i\frac{\pi}{3}}\vec{e}_2)$ and $\vec{f_2} = \frac{1}{\sqrt{5}}(-2e^{-i\frac{\pi}{3}}\vec{e}_1 + \vec{e}_2)$.

3.2.4 The Dirac notation

In quantum mechanics it is customary to use a notation due to Dirac to denote vectors and their hermitian conjugates. Since it will be important throughout the course, here we introduce it.

In this notation, one indicates a column vector $\vec{v}$ with the symbol $|v\rangle$ (notice that we omit the arrow!), called a **ket**, and its hermitian conjugate $\vec{v}^\dagger$ with $\langle v|$, called **bra** (the origin of these strange names is explained below). Let us translate our results above in this new notation. First, we will denote a canonical orthonormal basis as

$$\vec{e}_i \leftrightarrow |i\rangle, \quad i = 1, \ldots, N, \tag{3.55}$$

correspondingly their hermitian conjugates are denoted as

$$\vec{e}_i^{\,\dagger} \leftrightarrow \langle i|, \quad , i = 1, \ldots, N. \tag{3.56}$$

Denoting with $|v\rangle$, $|w\rangle$ two generic vectors, the properties we have discussed in the sections above now can be rewritten as:

$$\begin{aligned}
&\text{decomposition} \qquad |v\rangle \;=\; \sum_{i=1}^{N} v_i |i\rangle \,, \\
&\text{orthonormal basis} \qquad \langle i|j\rangle \;=\; \delta_{ij} \,, \\
&\text{scalar product} \qquad \langle v|w\rangle \;=\; \langle w|v\rangle^* = \sum_{i=1}^{N}\sum_{j=1}^{N} v_i^* w_j \, \langle i|j\rangle \;=\; \sum_{i=1}^{N} v_i^* w_i \,, \\
&\text{squared norm} \qquad \langle v|v\rangle \;=\; \sum_{i=1}^{N} |v_i|^2 \,, \\
&\text{Schwarz inequality} \qquad |\langle v|w\rangle|^2 \;\leq\; \langle v|v\rangle \, \langle w|w\rangle \,, \\
&\text{vector components} \qquad v_i \;=\; \langle i|v\rangle \,, \qquad v_i^* \;=\; \langle v|i\rangle \,.
\end{aligned} \tag{3.57}$$

Notice that the relation between bras and kets is $\langle v| = (|v\rangle)^\dagger$, stating that, where $|v\rangle$ represents a column vector with components v_i, then $\langle v|$ is the row vector with components v_i^*.

Notice that **the scalar product always involves a bra contracted with a ket** – together they form a *bracket* $\langle v|w\rangle$ - this wordplay explains the strange names introduced by Dirac. We will also see in the next sections that matrix elements of generic linear operators in an orthonormal basis are given by $\langle i|A|j\rangle$.

Exercise 25

- Given a orthonormal basis $|i\rangle$, $i = 1, 2$, with $\langle i|j\rangle = \delta_{ij}$, form a new orthonormal basis $|i'\rangle$ where one of the basis vectors is proportional to $|1'\rangle \propto |1\rangle + i|2\rangle$.
- Express the vector $|v\rangle = (\frac{2}{7} + 2i)|1\rangle + (-7 + 3i)|2\rangle$ in the new basis.
- Compute the scalar products $\langle w|v\rangle$, $\langle v|w\rangle$, $\langle z|w\rangle$, $\langle w|z\rangle$, where $|w\rangle = i|1\rangle + |2\rangle$, and $|z\rangle = (3 + 2i)|1'\rangle$.

3.2.5 Linear maps on complex vector spaces.

We have already encountered the concept of a linear map, and how it is represented by a matrix. This is precisely the same on a complex vector space: a linear operator A sends vectors to vectors, respecting linear combinations. In the Dirac notation, this condition of linearity is written as

$$A \text{ linear} \qquad \Longleftrightarrow \qquad A\Big(a|v\rangle + b|w\rangle\Big) \;=\; a\, A|v\rangle + b\, A|w\rangle \,, a, b \in \mathbb{C}. \tag{3.58}$$

Linear transformations acting on the vectors of $\mathcal{H}$, of dimension N, are represented by square $N \times N$ matrices, and in Quantum Mechanics they are customarily called *operators*.

Notice that, in the Dirac notation, we do not write explicitly "$\cdot$" or "$\circ$" for the action of the operator on vectors. There is no danger to confuse linear operators with vectors, since vectors are marked by their special bra and ket notation. Sometimes, when there is some danger of confusion, we will mark linear operators with a hat, i.e. $\hat{A}$, just to distinguish them from numerical constants.

Action on vectors. Matrix elements (which are now generally complex numbers) are, just like we saw before, singled out by contracting the matrix with a pair of elements of the canonical basis

$$A_{ij} \;=\; \langle i|A|j\rangle\,, \tag{3.59}$$

(note: $\langle i|$ selects the i-th row in A, as it happens in the latter of eq. (3.57) on $|v\rangle$, while $|j\rangle$ selects the j-th column) and the action of operators on generic kets $|v\rangle$ and bras $\langle v|$ can be expressed in components as

$$\begin{aligned}
\langle i|A|v\rangle \;&=\; \langle i|A\sum_{j=1}^{N} v_j|j\rangle \;=\; \sum_{j=1}^{N} v_j\langle i|A|j\rangle \;=\; \sum_{j=1}^{N} A_{ij}v_j\,,\\
\langle v|A|i\rangle \;&=\; \sum_{j=1}^{N}\langle j|v_j^*A|i\rangle \;=\; \sum_{j=1}^{N} v_j^*\langle j|A|i\rangle \;=\; \sum_{j=1}^{N} v_j^*A_{ji}\,.
\end{aligned} \tag{3.60}$$

These formulae specify that A acts as a matrix multiplying the vector $\vec{v}$ from the left:

$$A|v\rangle \leftrightarrow A\cdot\vec{v} \;=\; \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1N}\\ \vdots & \vdots & \ddots & \vdots\\ A_{N1} & A_{N2} & \cdots & A_{NN}\end{pmatrix}\begin{pmatrix} v_1\\ \vdots\\ v_N\end{pmatrix}\,, \tag{3.61}$$

while the bra $\langle v|A$ is obtained through matrix multiplication on the right,

$$\langle v|A \leftrightarrow \vec{v}^{\,\dagger}\cdot A \;=\; (v_1^*,\cdots,v_N^*)\begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1N}\\ \vdots & \vdots & \ddots & \vdots\\ A_{N1} & A_{N2} & \cdots & A_{NN}\end{pmatrix}\,. \tag{3.62}$$

The identical transformation, transforming each vector $|v\rangle$ into itself, is represented by the identity matrix I:

$$I|v\rangle \;=\; |v\rangle\,, \qquad\qquad \langle i|I|j\rangle \;=\; \langle i|j\rangle \;=\; \delta_{ij} \;=\; \begin{pmatrix} 1 & 0 & \cdots & 0\\ 0 & 1 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & 1\end{pmatrix}\,. \tag{3.63}$$

We have already seen that, **on a complete orthonormal basis $|i\rangle$, the identity matrix has the following representation**:

$$I|v\rangle \;=\; |v\rangle \;=\; \sum_{i=1}^{N} v_i|i\rangle \;=\; \sum_{i=1}^{N}|i\rangle\langle i|v\rangle \quad \text{for all } |v\rangle\,, \tag{3.64}$$

$$\Longrightarrow \qquad I \;=\; \sum_{i=1}^{N}|i\rangle\langle i|\,, \tag{3.65}$$

which takes the name of **completeness relation** .

As we have already remarked, subsequent linear transformations are products of operators. In general, the composite action of transformations depends on their ordering, and the *commutator* is defined as

$$[A_1, A_2] \equiv A_1 A_2 - A_2 A_1 \,. \tag{3.66}$$

Analogously, the *anticommutator*, useful in some applications is defined as

$$\{A_1, A_2\} \equiv A_1 A_2 + A_2 A_1 \,, \tag{3.67}$$

so that in general

$$A_1 A_2 = \frac{1}{2}\Big(A_1 A_2 - A_2 A_1\Big) + \frac{1}{2}\Big(A_1 A_2 + A_2 A_1\Big) = \frac{1}{2}[A_1, A_2] + \frac{1}{2}\{A_1, A_2\}. \tag{3.68}$$

Decomposing a general linear operator in any orthonormal basis. A very nice feature of the Dirac notation is that we it does not assume any particular choice of basis. When we denote a linear operator as A, for example, this means the linear map on the vector space.

We can however always use the completeness relation to decompose A in a given orthonormal basis. In fact, since the identity commutes with any operator, we can write $A = IAI$, and thus, using the resolution of the identity twice,

$$A = I \cdot A \cdot I = \left(\sum_{i=1}^{N} |i\rangle\langle i|\right) A \left(\sum_{j=1}^{N} |j\rangle\langle j|\right), \tag{3.69}$$

from which we find

$$A = \sum_{i,j=1}^{N} |i\rangle\langle i|A|j\rangle\langle j| = \sum_{i,j=1}^{N} A_{ij}\, |i\rangle\langle j|. \tag{3.70}$$

Exercise 26

Consider a linear operator[5] σ_2 of 2-dimensional complex space, which acts on a orthonormal basis $|1\rangle$, $|2\rangle$ in the following way: $\sigma_2|1\rangle = i|2\rangle$, $\sigma_2|2\rangle = -i|1\rangle$, where i is the imaginary unit.

- Write down the action of σ_2 on a generic vector $|v\rangle = \alpha|1\rangle + \beta|2\rangle$, with α, $\beta \in \mathbb{C}$.
- Represent σ_2 as a 2×2 matrix, in the basis $|1\rangle$, $|2\rangle$.
- Represent σ_2 as a 2×2 matrix, in the different orthonormal basis $|1'\rangle = \frac{1}{\sqrt{2}}(|1\rangle + i|2\rangle)$, $|2'\rangle = \frac{1}{\sqrt{2}}(|1\rangle - i|2\rangle)$.

3.2.6 Hermitian and unitary matrices

Hermitian conjugation, and hermitian operators. Since operators, as well as vectors, are complex objects, they are associated with a notion of complex conjugation. In particular, given an

operator A, its **hermitian conjugate** $A^\dagger$ is defined through the following relation, valid for generic kets $|v\rangle$, $|w\rangle$:

$$\langle v|A^\dagger|w\rangle \;=\; \Big(\langle w|A|v\rangle\Big)^* . \tag{3.71}$$

In components, the two sides of the above definition are

$$\begin{aligned}
&\langle v|A^\dagger|w\rangle \;=\; \langle v|IA^\dagger I|w\rangle \;=\; \sum_{i,j=1}^{N} \langle v|j\rangle\langle j|A^\dagger|i\rangle\langle i|w\rangle \;=\; \sum_{i,j=1}^{N} v_j^*(A^\dagger)_{ji}\, w_i\,, \\
&\Big(\langle w|A|v\rangle\Big)^* \;=\; \Big(\sum_{i,j=1}^{N} w_i^* A_{ij}\, v_j\Big)^* \;=\; \sum_{i,j=1}^{N} v_j^* A_{ij}^*\, w_i\,.
\end{aligned} \tag{3.72}$$

The equality of the two sides for all $|v\rangle$ and $|w\rangle$ implies

$$(A^\dagger)_{ji} \;=\; A_{ij}^*\,, \tag{3.73}$$

namely the operation consists in taking the complex conjuate together with the transpose, $A^\dagger = (A^T)^*$.

Note that the definition of the hermitan conjugate operator is such that the following important relation holds

$$\Big(A|v\rangle\Big)^\dagger \;=\; \langle v|A^\dagger\,. \tag{3.74}$$

An operator equal to its hermitian conjugate, $A = A^\dagger$, is said to be **h**ermitian.

Exercise 27

Which of these matrices is hermitian?

- $M_1 = \begin{pmatrix} 2 & 3 \\ 1 & 2 \end{pmatrix}$.
- $M_2 = \begin{pmatrix} 2+i & 3+i \\ 3-i & 5 \end{pmatrix}$.
- $M_3 = \begin{pmatrix} 2 & 2-i5 \\ 2+5i & 1/3 \end{pmatrix}$.
- $M_4 = \begin{pmatrix} 1 & 1/2-i & 2 \\ 1/2+i & 0 & i \\ 2 & -i & 74 \end{pmatrix}$
- $M_5 = \begin{pmatrix} 0 & e^{i\pi/5}+1 \\ 1-e^{i\pi/5} & 0 \end{pmatrix}$.

Unitary operators. An operator satisfying $AA^\dagger = A^\dagger A = I$ is called **unitary**.
Unitary operators generalize the notion of orthogonal operators to the complex context. In fact, they are characterised by the fact that they **preserve the value of the scalar product between vectors**.

In fact, consider the two kets $|v\rangle$, and $|w\rangle$, with scalar product $\langle w|v\rangle = (\langle v|w\rangle)^*$. Applying the operator A, they transform into $|v'\rangle = A|v\rangle$, and $|w'\rangle = A|w\rangle$. The scalar product of the transformed vectors is

$$\langle w'|v'\rangle = \underbrace{\langle w|A^\dagger}_{=(A|w\rangle)^\dagger} A|v\rangle = \langle w|A^\dagger A|v\rangle = \underbrace{\langle w|I|v\rangle}_{\text{by } AA^\dagger = I} = \langle w|v\rangle, \tag{3.75}$$

so it is invariant provided that $AA^\dagger = I$.

Notice that there is a certain close relation between hermitian and unitary matrices. To find it, let us consider the simplest example of $\times 1$ matrices, i.e. just normal complex numbers.

> Consider 1×1 matrices. What are the conditions for such matrices to be hermitian or unitary?

Solution: You should have discovered that hermitian 1×1 numbers are simply *real* numbers. Unitary 1×1 matrices z are simply numbers satisfying $|z| = 1$, i.e. they can be written in the form $z = e^{i\alpha}$, with α a real number. Therefore we can say that, at least in the 1×1 case, we have the relation

$$A \text{ is hermitian } \leftrightarrow e^{iA} \text{ is unitary .} \tag{3.76}$$

In fact **this relation is valid in any dimension**, for $N \times N$ matrices! To verify we should explain how we can consider functions of a matrix, such as taking its exponential. It turns out that we can give a precise meaning to these operations for hermitian matrices. This will be explained in detail in the next section, where we treat the diagonalization of a matrix.

Summary.

- We have learnt the definition of complex vector spaces with a complex scalar product linear in the second argument, anti-linear in the first argument. These spaces are are also called **Hilbert spaces**.
- We have seen the Dirac notation, where vectors are represented as kets $|v\rangle$ and their hermitian conjugate are bras $\langle v|$.
- Decomposing the identity operator in an orthonormal basis gives the completeness relation $I = \sum_i |i\rangle\langle i|$.
- The matrix elements of a linear operator A in a orthonormal basis are obtained as $\langle i|A|j\rangle$.
- Hermitian operators satisfy $A = A^\dagger$, i.e. $\langle i|A|j\rangle = (\langle j|A|i\rangle)^*$.

3.3 Operations on matrices

We close this section by introducing some very important operations on $N \times N$ matrices: the **inverse**, and diagonalisation.

3.3.1 Determinant and trace of a square matrix

.

The determinant. The determinant is a number associated to a $N \times N$ matrix. Before giving the general definition, let us give its definition for 2×2 and 3×3 matrices:

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - cb, \tag{3.77}$$

$$\det \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & j \end{pmatrix} = a \cdot e \cdot j - a \cdot f \cdot h - b \cdot d \cdot j + b \cdot g \cdot f + c \cdot d \cdot h - c \cdot g \cdot e. \tag{3.78}$$

To give its general definition we have to consider different orderings of the columns of the matrix. Different orderings of N distinct elements are called by mathematicians *permutations*, i.e. they are maps of the form

$$\sigma : \begin{pmatrix} 1 & 2 & \dots & N \\ \sigma(1) & \sigma(2) & \dots & \sigma(N) \end{pmatrix}. \tag{3.79}$$

A natural and important quantity is the *sign* of the permutation, denoted as $(-1)^\sigma$. This is a number ± 1 associated to any permutation which can be obtained by noticing that every permutation can be decomposed as a sequence of exchanges of the position of two elements. The sign is $+1$ if this decomposition contains an even number of exchanges, and -1 for an odd number.

For example, for the permutation sending 1234 to 1432:

$$\sigma \equiv \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 4 & 3 & 2 \end{pmatrix}, \tag{3.80}$$

can be obtained by a single exchange $4 \leftrightarrow 2$. Then the sign is $(-1)^\sigma = -1$.

As another example, consider now the permutation

$$\sigma \equiv \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 1 & 3 & 2 \end{pmatrix}. \tag{3.81}$$

This permutation can be obtained, for example, by first exchanging $4 \leftrightarrow 2$, and then exchanging $1 \rightarrow 4$: two exchanges. Alternatively, we could obtain it by first exchanging $1 \leftrightarrow 2$, and then exchanging $2 \leftrightarrow 4$. Notice that the decomposition itself is not unique. What is well defined is the sign, since in all decompositions into single exchanges, we must have an even number of exchanges. Thus, in this case the sign is $(-1)^\sigma = +1$.

With this definition, **the determinant of a $N \times N$ matrix is defined as**

$$\det A = \sum_\sigma (-1)^\sigma \, \langle 1|A|\sigma(1)\rangle \, \langle 2|A|\sigma(2)\rangle \, \dots \langle N|A|\sigma(N)\rangle. \tag{3.82}$$

Notice that the definition of determinant depends on a choice of basis. However, it can be proved that the determinant of a matrix computed in two different orthonormal basis is the same.

Exercise 28

Prove that this definition of determinant agrees with the one given above for 2×2 and 3×3 matrices.

The determinant is a crucial quantity because it is needed to compute the eigenvalues of a matrix, as we see below.

The trace. The trace of a matrix is another useful quantity: it is just the sum of the elements on the diagonal of the matrix

$$\mathrm{Tr}A = \sum_{i=1}^{N} A_{ii} = \sum_{i=1}^{N} \langle i|A|i\rangle. \tag{3.83}$$

3.3.2 Inverse of a square matrix

An operator A is invertible if its determinant is $\neq 0$. In such a case there exists another operator, called the **inverse** operator A^{-1}, that undoes the action of A, namely such that

$$AA^{-1} \,=\, A^{-1}A \,=\, I\,. \tag{3.84}$$

The inverse of a product of invertible operators works a bit like the transposed: it is the product of inverses *in reversed order*:

$$\big(A_1 \underbrace{A_2\big)\big(A_2^{-1}}_{=I} A_1^{-1}\big) \,=\, \underbrace{A_1\, A_1^{-1}}_{=I} \,=\, I \qquad \Longrightarrow \qquad A_2^{-1}A_1^{-1} \,=\, \big(A_1\, A_2\big)^{-1}\,. \tag{3.85}$$

There is a general formula for the inverse of a matrix. You are not asked to memorise it but just to know that it exists. The element (i, j) of the inverse matrix can be computed as:

$$A^{-1} = \frac{1}{\det(A)}\ \mathrm{Adj}(A), \tag{3.86}$$

where $\mathrm{Adj}(A)$ is the so-called **adjugate matrix**: the element (i, j) of the adjugate matrix is obtained by the following procedure:

- Delete the row j and the column i of the matrix A, and the the determinant of the resulting $N-1 \times N-1$ matrix.
- Multiply the result by $(-1)^{i+j}$.

In particular, for a 2×2 matrix we get

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad-bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}. \tag{3.87}$$

Exercise 29

Verify the formula above for the inverse in the case $N = 2$ is correct: namely, verify that (3.84) is satisfied. Write an explicit formula for the inverse of a 3×3 matrix using the general formula above.

3.3.3 Diagonalisation: eigenvalues and eigenvectors.

A crucial property of operators, both at the mathematical level, and especially for their applications in Quantum Mechanics, is their diagonalisability. An $N \times N$ matrix is *diagonalisable* if one can find a basis of N vectors $|v_k\rangle$, with $k = 1, \cdots, N$, that satisfy the *eigenvalue equation*

$$A|v_k\rangle \;=\; \lambda_k |v_k\rangle \,, \tag{3.88}$$

for certain numbers $\lambda_k \in \mathbb{C}$. If this is the case, the vectors $|v_k\rangle$ are called *eigenvectors* of A, while λ_k are the *eigenvalues* of A. In the basis of eigenvectors $|v_k\rangle$ the matrix A assumes a diagonal form:

$$A_{\text{diag}} \;=\; \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ 0 & 0 & \cdots & \lambda_N \end{pmatrix} . \tag{3.89}$$

Diagonalising a matrix means finding a complete basis of its eigenvectors (with the respective eigenvalues). This operation is crucial in quantum mechanics. Let us give practical recipe on how this can be done.

How to diagonalize a matrix

- First, solve the **indicial equation**. This is the equation given by

$$\det\left(A - \lambda I\right) = 0, \tag{3.90}$$

  where I is the $N \times N$ identity matrix, and $\lambda \in \mathbb{C}$ is the unknown. This is a polynomial equation of the N-th order for λ. Therefore it will in general have N complex roots, which however may be repeated. These N roots λ_i are candidates to be the eigenvalues.

- For every solution of the indicial equation, look for an eigenvector which satisfies the eigenvalue equation (3.88) with the eigenvalue λ_i. In practice this is a linear system of equations for the coefficients of $|v\rangle$ in the chosen basis.

- Repeat the process for the other eigenvalues until (if possible) a complete basis of eigenvectors is built.

This process can only encounter some issues in case there are repeated roots of the equation, i.e. $\lambda_i = \lambda_j$, for $i \neq j$. In this case it *can happen* that one of the eigenvectors cannot be found, i.e. the matrix is not diagonalisable.

Example of diagonalisation. Given the matrix

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

its eigenvalues λ_k, for $k = 1, 2$, are found by solving for λ the *characteristics (or secular) equation*

$$\det(A - \lambda I) = 0\,, \tag{3.91}$$

where in this case I is the 2×2 identity operator:

$$A - \lambda I = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} -\lambda & 1 \\ 1 & -\lambda \end{pmatrix},$$
$$\det \begin{pmatrix} -\lambda & 1 \\ 1 & -\lambda \end{pmatrix} = \lambda^2 - 1 = 0 \qquad \Longrightarrow \qquad \lambda_1 = 1\,, \qquad \lambda_2 = -1\,.$$

The eigenvector corresponding to $\lambda_1 = 1$ is found imposing its eigenvalue equation. Setting $|v_1\rangle = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$, one has

$$\underbrace{\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}}_{A} \underbrace{\begin{pmatrix} \alpha \\ \beta \end{pmatrix}}_{|v_1\rangle} = \underbrace{\lambda_1}_{=1} \underbrace{\begin{pmatrix} \alpha \\ \beta \end{pmatrix}}_{|v_1\rangle} \qquad \Longrightarrow \qquad \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} \beta \\ \alpha \end{pmatrix} \qquad \Longrightarrow \qquad \beta = \alpha\,.$$

Each vector $|v_1\rangle$ with two equal components is eigenvector of A with eigenvalue 1. Normalising the eigenvector to unit norm we get

$$1 = \langle v_1 | v_1 \rangle = (\alpha^*, \alpha^*) \begin{pmatrix} \alpha \\ \alpha \end{pmatrix} = 2|\alpha|^2 \qquad \Longrightarrow \qquad |\alpha| = \frac{1}{\sqrt{2}} \qquad \Longrightarrow \qquad \alpha = \frac{e^{i\phi}}{\sqrt{2}}\,.$$

By choosing $\phi = 0$, namely a real eigenvector, we finally obtain

$$|v_1\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

With analogous computations, relative to eigenvalue $\lambda_2 = -1$, one arrives at

$$|v_2\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

Note that there is no possibility to distinguish, from the diagonalisation procedure, two eigenvectors by their phase, namely $|v_1\rangle$ and $|v_1'\rangle = e^{i\phi}|v_1\rangle$ do satisfy the same properties (they are both eigenvectors with eigenvalue $+1$, and normalised to unity), so unit eigenvectors are always defined up to a phase.

More formally, a matrix A is diagonalisable if there exists an invertible matrix P such that $P^{-1}AP = A_{\text{diag}}$, where A_{diag} is diagonal has the eigenvalues of A as entries in the diagonal, i.e. $(A_{\text{diag}})_{ij} = \lambda_j \delta_{ij}$, see eq. (3.89). P is the matrix whose k-th column is the k-th eigen-ket $|v_k\rangle$ of A, namely $P_{ik} = \langle i | v_k \rangle = (v_k)_i$. To show that this is the case, we can recall that, by definition of inverse, we have $P^{-1}P = I$, namely $\sum_n (P^{-1})_{in} P_{nk} = \sum_n (P^{-1})_{in} (v_k)_n = \delta_{ik}$, and that, by definition of eigenvectors, $\sum_m A_{nm} (v_k)_m = \lambda_k (v_k)_n$. This yields: $(P^{-1}AP)_{ik} = \sum_{n,m} (P^{-1})_{in} A_{nm} P_{mk} = \sum_{n,m} (P^{-1})_{in} A_{nm} (v_k)_m = \sum_n (P^{-1})_{in} \lambda_k (v_k)_n = \lambda_k \delta_{ik} = (A_{\text{diag}})_{ik}$, which concludes the proof.

Exercise 30

Diagonalise the matrices

$$M_1 = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \qquad M_2 = \begin{pmatrix} 1 & 2 & 1 \\ 0 & 2 & 0 \\ 1 & -2 & 1 \end{pmatrix}. \tag{3.92}$$

Find a decomposition of these matrices of the form $PA_{\text{diag}}P^{-1}$, computing P and A_{diag} explicitly.

Exercise 31

- Diagonalise the matrices

$$\sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \qquad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{3.93}$$

- Find a decomposition of these matrices of the form $PA_{\text{diag}}P^{-1}$, computing P and A_{diag} explicitly.

With respect to diagonalisation, hermitian operators enjoy several remarkable properties, of which the most important are the following:

1) **every hermitian operator A is diagonalisable**, and it is always possible to find an **orthonormal basis of eigenvectors of A**;

2) hermitian operators have *real* **eigenvalues**;

3) **eigenvectors of a hermitian operator corresponding to different eigenvalues are orthogonal**.

We show explicitly the second and third properties. Given A hermitian, and $|v_i\rangle$ one of its eigenvectors, corresponding to eigenvalue λ_i, one has

$$\lambda_i \langle v_i|v_i\rangle = \langle v_i|A|v_i\rangle = \langle v_i| \underbrace{A^\dagger}_{A=A^\dagger} |v_i\rangle = \Big(\langle v_i|A|v_i\rangle\Big)^* = \Big(\lambda_i \langle v_i|v_i\rangle\Big)^* = \lambda_i^* \langle v_i|v_i\rangle\,, \tag{3.94}$$

which implies $\lambda_i^* = \lambda_i$, namely real eigenvalues (property 2)).

Moreover,

$$\lambda_j \langle v_i|v_j\rangle = \langle v_i|A|v_j\rangle = \Big(\langle v_j|A^\dagger|v_i\rangle\Big)^* = \Big(\langle v_j|A|v_i\rangle\Big)^* = \Big(\lambda_i \langle v_j|v_i\rangle\Big)^* = \underbrace{\lambda_i}_{\lambda_i^*=\lambda_i} \langle v_i|v_j\rangle\,, \tag{3.95}$$

which implies $(\lambda_i - \lambda_j)\langle v_i|v_j\rangle = 0$. If $\lambda_i \neq \lambda_j$ it is thus $\langle v_i|v_j\rangle = 0$, namely eigenvectors corresponding to different eigenvalues are orthogonal (property 3)).

We also note that a hermitian operator can always be diagonalised by means of a unitary transformation, i.e. if A is hermitian, the P matrix achieving $P^{-1}AP = A_{\rm diag}$ is unitary. Being unitary, $P^{-1} = P^\dagger$, hence one has $P^\dagger AP = A_{\rm diag}$. Matrix $P^\dagger$ is such that its k-th row is the k-th eigen-bra of A, namely $(P^\dagger)_{ki} = \langle v_k|i\rangle$. Thus $(P^\dagger AP)_{ij} = \sum_{m,n}(P^\dagger)_{im}A_{mn}P_{nj} = \sum_{m,n}\langle v_i|m\rangle\langle m|A|n\rangle\langle n|v_j\rangle = \langle v_i|A|v_j\rangle = (A_{\rm diag})_{ij}$, having used $\sum_n |n\rangle\langle n| = I$. This specifies the meaning of the statement 'in the basis of eigenvectors $|v_k\rangle$ the matrix A assumes a diagonal form', before eq. (3.89).

Another property of hermitian operators, particular important for the developments of the formalism of quantum mechanics, is the following:

4) given two **commuting hermitian operators** A and B acting on vectors of $\mathcal{H}$, one can always find a **complete orthonormal basis of common eigenvectors**: that is, a set of N orthonormal eigenvectors of *both* operators that span the whole space $\mathcal{H}$.

We will not show this property in general, but just give a flavour of how it comes about. Suppose as a first case that A features only non-degenerate eigenvalues:

$$A|v_k\rangle \,=\, \lambda_k|v_k\rangle \qquad k \,=\, 1,\cdots,N, \qquad \lambda_1 \neq \lambda_2 \neq \cdots \neq \lambda_N\,.$$

Assuming commutativity between A and B, one has

$$A\big(B|v_k\rangle\big) \,=\, BA|v_k\rangle \,=\, B\lambda_k|v_k\rangle \,=\, \lambda_k\big(B|v_k\rangle\big)\,.$$

Since $|v_k\rangle$ is, by hypothesis, the sole eigenvector corresponding to λ_k, and $B|v_k\rangle$ satisfies the same eigenvalue equation as $|v_k\rangle$, they must be proportional:

$$B|v_k\rangle \,=\, \beta_k|v_k\rangle\,.$$

Hence the orthonormal set of $|v_k\rangle$ is a common basis of eigenvectors of A and B, that span the complete space $\mathcal{H}$ (see first property of hermitian operators under diagonalisation).
Suppose now that A has one doubly-degenerate eigenvalue λ, common to eigenvectors $|v_1\rangle$ and $|v_2\rangle$, so that

$$A|v_1\rangle \,=\, \lambda|v_1\rangle\,, \qquad\qquad A|v_2\rangle \,=\, \lambda|v_2\rangle\,.$$

In this case, one has

$$A\big(B|v_i\rangle\big) \,=\, BA|v_i\rangle \,=\, B\lambda|v_i\rangle \,=\, \lambda\big(B|v_i\rangle\big)\,, \qquad i=1,2\,,$$

so one must have

$$B|v_1\rangle \,=\, \alpha|v_1\rangle + \beta|v_2\rangle\,, \qquad\qquad B|v_2\rangle \,=\, \gamma|v_1\rangle + \delta|v_2\rangle\,,$$

since $|v_1\rangle$ and $|v_2\rangle$ are the only two eigenvectors with eigenvector λ, by hypothesis. Note that, by hermiticity of B, the coefficients satisfy $\alpha,\delta \in \mathbb{R}$, and $\gamma = \beta^*$. In particular, defining

$$r_\pm \,=\, \frac{1}{2\beta}\left(\alpha - \delta \pm \sqrt{(\alpha-\delta)^2 + 4|\beta|^2}\right)\,,$$

the two kets

$$|w_\pm\rangle \,=\, \frac{r_\pm|v_1\rangle + |v_2\rangle}{\sqrt{1+|r_\pm|^2}}\,, \qquad\qquad \langle w_-|w_+\rangle \,=\, 0\,,$$

are orthonormal eigenvectors of A with eigenvalue λ, and of B with eigenvalues $r_\pm\beta + \delta$, hence the basis $|w_+\rangle$, $|w_-\rangle$, $|v_3\rangle$, $\cdots$, $|v_N\rangle$, is a common basis of eigenvectors of A and B, that span the complete space $\mathcal{H}$.
Analogously to what done in these simple cases, one can generalise to arbitrary degeneracy patterns: in the end one is always able to find a set of orthonormal eigenvectors common to A and B spanning the entire $\mathcal{H}$.

Functions of diagonalizable matrices. It is possible compute functions of diagonalizable matrices. The idea to define functions of a matrix is to use the Taylor expansion, for example

$$\sin(A) = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} A^{2n+1}, \tag{3.96}$$

where A^k denotes the matrix multiplication of A with itself k times.

Let us consider first the simplest case: when A is already a diagonal matrix. Then, computing these powers is very simple, because they simply amount to taking the powers of the eigenvalues!

$$A_{\text{diag}} = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \dots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_N \end{pmatrix} \to A^k_{\text{diag}} = \begin{pmatrix} \lambda_1^k & 0 & \dots & 0 \\ 0 & \lambda_2^k & \dots & 0 \\ \vdots & \dots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_N^k \end{pmatrix}. \tag{3.97}$$

Then, when we take a series, we find that this becomes equivalent to taking the corresponding series for each elements on the diagonal: for example,

$$\sin\left(A_{\text{diag}}\right) = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} \begin{pmatrix} \lambda_1^{2n+1} & 0 & \dots & 0 \\ 0 & \lambda_2^{2n+1} & \dots & 0 \\ \vdots & \dots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_N^{2n+1} \end{pmatrix}, \tag{3.98}$$

which means we can just sum the series for each element separately. This simply means that a function of a diagonal matrix, is simply given by a diagonal matrix, where each element is the function of the corresponding eigenvalue:

$$\sin\left(A_{\text{diag}}\right) = \begin{pmatrix} \sin(\lambda_1) & 0 & \dots & 0 \\ 0 & \sin(\lambda_2) & \dots & 0 \\ \vdots & \dots & \ddots & \vdots \\ 0 & 0 & \dots & \sin(\lambda_N) \end{pmatrix}, \tag{3.99}$$

and more generally, for any function $f(A)$ with a Taylor expansion,

$$F\left(A_{\text{diag}}\right) = \begin{pmatrix} F(\lambda_1) & 0 & \dots & 0 \\ 0 & F(\lambda_2) & \dots & 0 \\ \vdots & \dots & \ddots & \vdots \\ 0 & 0 & \dots & F(\lambda_N) \end{pmatrix}. \tag{3.100}$$

What about matrices A that are not diagonal, but can be diagonalised? Then, we know that we can find a decomposition $A = PA_{\text{diag}}P^{-1}$. Then, we can still easily compute matrix powers like A^k. In fact notice that $A^2 = (PA_{\text{diag}}P^{-1})(PA_{\text{diag}}P^{-1}) = P(A_{\text{diag}})^2P^{-1}$ (since $PP^{-1} = I$) and similarly $A^3 = (PA_{\text{diag}}P^{-1})(PA_{\text{diag}}P^{-1})(PA_{\text{diag}}P^{-1}) = P(A_{\text{diag}})^3P^{-1}$. It is simple to convince yourself of the general formula

$$A^k = P(A_{\text{diag}})^k P^{-1}. \tag{3.101}$$

This then extends to the whole series! We can rewrite for example

$$\sin(A) = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} A^{2n+1} = P\left(\sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} (A_{\text{diag}})^{2n+1}\right) P^{-1} = P\sin\left(A_{\text{diag}}\right) P^{-1}, \tag{3.102}$$

and more generally for any Taylor-expandable function,

$$F(A) = PF\left(A_{\text{diag}}\right)P^{-1}. \tag{3.103}$$

This is great because we have already learnt how to compute a function of a diagonal matrix! Therefore, we have learnt how to compute any function (with a Taylor expansion) of a diagonalizable matrix, in particular Hermitian matrices.

Exercise 32

Compute $\sin(\theta\sigma_i)$ and $e^{i\theta\sigma_i}$, where σ_i $(i = 2, 3)$ are the two matrices you have already diagonalised in exercise (31), and θ is a parameter. Verify that for $\theta \in \mathbb{R}$, then $U_i \equiv e^{i\theta\sigma_i}$ is a unitary matrix, i.e. satisfies $U_i U_i^\dagger = I$.

Important concepts of this section.

- Determinant, trace and inverse of a square matrix. Matrices are invertible if and only if they have nonzero determinant.
- Definition of diagonalization, eigenvalues and eigenvectors.
- Method to diagonalize a matrix.
- Special diagonalization properties of Hermitian matrices.
- How to compute functions of a diagonalizable matrix.

3.4 Continuous limit

So far, we have recollected notions of linear algebra in case the dimension of $\mathcal{H}$ is $N < \infty$. One can generalise everything in case $N \to \infty$, and also to the *continuous case*, namely the case in which the index labelling the basis elements is not a discrete integer (as was for i above), but a real number $x \in (-\infty, \infty)$. This can be visualised as if at each 'position' x of the range there were a different 'dimension' represented by a unit vector $|x\rangle$ along which a generic vector $|v\rangle$ may have a non-zero component. Thus vector $|v\rangle$ is decomposed as

$$|v\rangle \;=\; \int_{-\infty}^{\infty} dx\, v(x)|x\rangle\,, \tag{3.104}$$

which is the continuous generalisation of the first of (3.57). The component $v(x)$ is extracted by contracting $|v\rangle$ with a basis bra:

$$\langle x'|v\rangle \;=\; \int_{-\infty}^{\infty} dx\, v(x)\langle x'|x\rangle\,, \tag{3.105}$$

where, analogously to $\langle j|i\rangle = \delta_{ji} = \delta_{ij}$, for an orthonormal basis $|x\rangle$ one has

$$\langle x'|x\rangle \;=\; \delta(x'-x) \;=\; \delta(x-x') \qquad (\text{Dirac } \delta)\,, \tag{3.106}$$

and the Dirac delta is

$$\delta(x'-x) \;=\; \begin{cases} +\infty & \text{if} \quad x \,=\, x' \,, \\ 0 & \text{if} \quad x \,\neq\, x' \,, \end{cases} \tag{3.107}$$

so that

$$\int_{-\infty}^{\infty} dx\, \delta(x-x') \;=\; 1 \,, \qquad\qquad \int_{-\infty}^{\infty} dx\, \delta(x-x')\, f(x) \;=\; f(x') \,, \tag{3.108}$$

which are the analogues of the discrete relations $\sum_i \delta_{ij} = 1$, and $\sum_i \delta_{ij} f_i = f_j$. The Dirac $\delta(x-x')$ can be conveniently thought of as a degenerate normalised gaussian distribution centred in $x = x'$, in the limit in which its width $\sigma \to 0$, and its height $h \to \infty$, so to maintain unit normalisation.
Given the above relations, the component of $|v\rangle$ along direction $|x'\rangle$ is thus

$$\langle x'|v\rangle \;=\; \int_{-\infty}^{\infty} dx\, v(x) \langle x'|x\rangle \;=\; \int_{-\infty}^{\infty} dx\, v(x) \delta(x-x') \;=\; v(x') \,, \tag{3.109}$$

analogous to the last of relations (3.57).
The completeness relation of basis $|x\rangle$ is expressed by

$$I \;=\; \int_{-\infty}^{\infty} dx\, |x\rangle\langle x| \,, \tag{3.110}$$

analogously to (3.64), and the scalar product is defined as

$$\langle v|w\rangle \;=\; \int_{-\infty}^{\infty} dx\, \langle v|x\rangle\langle x|w\rangle \;=\; \int_{-\infty}^{\infty} dx\, v(x)^* w(x) \,, \tag{3.111}$$

again analogous to (3.57).
The squared norm of vector $|v\rangle$ is

$$\langle v|v\rangle \;=\; \int_{-\infty}^{\infty} dx\, |v(x)|^2 \,, \tag{3.112}$$

which is $< \infty$ *only* if $v(\pm\infty) = 0$.
As a final generalisation, one can consider a number of continuous dimensions larger than 1, namely

$$|v\rangle \;=\; \int_{-\infty}^{\infty} dx_1 \cdots dx_n\, v(x_1, \cdots, x_n)\, |x_1, \cdots, x_n\rangle \;\equiv\; \int_{-\infty}^{\infty} d^n x\; v(\vec{x}\,)\, |\vec{x}\,\rangle \,, \tag{3.113}$$

with obvious extensions for the expression of the components and of the completeness relation:

$$v(\vec{x}\,) \;=\; \langle \vec{x}\,|v\rangle \,, \qquad\qquad I \;=\; \int_{-\infty}^{\infty} d^n x\; |\vec{x}\,\rangle\langle \vec{x}\,| \,. \tag{3.114}$$

To conclude the linear-algebra recollection we consider two specific examples of operators A and B acting on the bras at fixed 'position' $\langle x|$, defined as

$$\begin{aligned} \langle x|A \;=\; \langle x|x \;=\; x\langle x| \,, \qquad (x \in \mathbb{R}) \,, \qquad &\Longrightarrow \qquad \langle x|A|v\rangle \;=\; x\langle x|v\rangle = x\, v(x) \,, \\ \langle x|B \;=\; -i\hbar \frac{d}{dx} \langle x| \,, \qquad &\Longrightarrow \qquad \langle x|B|v\rangle \;=\; -i\hbar \frac{d}{dx} \langle x|v\rangle = -i\hbar \frac{dv(x)}{dx} \,. \end{aligned} \tag{3.115}$$

These operators have a crucial relevance in Quantum Mechanics, as they are linked to the position and the momentum of particles, respectively. We want to investigate the commutator $[A, B]$, and the hermiticity properties of A and B. The commutator reads

$$\begin{aligned} \langle x|AB|v\rangle \;&=\; x\, \langle x|B|v\rangle \;=\; -i\hbar x \frac{dv(x)}{dx} \,, \\ \langle x|BA|v\rangle \;=\; -i\hbar \frac{d}{dx} \Big(\langle x|A|v\rangle \Big) \;&=\; -i\hbar \frac{d}{dx} \Big(x\, v(x) \Big) \;=\; -i\hbar x \frac{dv(x)}{dx} - i\hbar v(x) \,, \end{aligned} \tag{3.116}$$

thus, for all $|v\rangle$,

$$\langle x|[A,B]|v\rangle \;=\; -i\hbar x\frac{dv(x)}{dx} + i\hbar x\frac{dv(x)}{dx} + i\hbar v(x) \;=\; i\hbar v(x) \;=\; i\hbar\langle x|I|v\rangle \qquad \Longrightarrow \qquad [A,B] \;=\; i\hbar I\,. \tag{3.117}$$

The hermiticity of A and B can be readily checked. Given generic vectors $|v\rangle$, $|w\rangle$, one has

$$\begin{aligned}\langle v|A|w\rangle \;&=\; \int_{-\infty}^{\infty} dx\,\langle v|x\rangle\langle x|A|w\rangle \;=\; \int_{-\infty}^{\infty} dx\,v(x)^* x\,\langle x|w\rangle \;=\; \int_{-\infty}^{\infty} dx\,\Big(x\,v(x)\Big)^*\langle x|w\rangle \\ &=\; \int_{-\infty}^{\infty} dx\,\Big(\langle x|A|v\rangle\Big)^*\langle x|w\rangle \;=\; \int_{-\infty}^{\infty} dx\,\langle v|A^\dagger|x\rangle\langle x|w\rangle \;=\; \langle v|A^\dagger|w\rangle\,,\end{aligned} \tag{3.118}$$

which implies $A = A^\dagger$. Analogously

$$\begin{aligned}\langle v|B|w\rangle \;&=\; \int_{-\infty}^{\infty} dx\,\langle v|x\rangle\langle x|B|w\rangle \;=\; \int_{-\infty}^{\infty} dx\,v(x)^*(-i\hbar)\frac{d}{dx}\langle x|w\rangle \\ &=\; \underbrace{-i\hbar v(x)^* w(x)\Big|_{-\infty}^{\infty}}_{=0 \text{ since } v,w(\pm\infty)=0} + \int_{-\infty}^{\infty} dx\,\Big(i\hbar\frac{dv(x)^*}{dx}\Big)\langle x|w\rangle \;=\; \int_{-\infty}^{\infty} dx\,\Big(-i\hbar\frac{dv(x)}{dx}\Big)^*\langle x|w\rangle \\ &=\; \int_{-\infty}^{\infty} dx\,\Big(\langle x|B|v\rangle\Big)^*\langle x|w\rangle \;=\; \int_{-\infty}^{\infty} dx\,\langle v|B^\dagger|x\rangle\langle x|w\rangle \;=\; \langle v|B^\dagger|w\rangle\,,\end{aligned} \tag{3.119}$$

implying $B = B^\dagger$. Hence A and B are hermitian operators, and their commutator is proportional to the identity operator I.

Chapter 4

Introduction to probabilities

Quantum Mechanics deals with probabilities for certain events to happen, i.e. for measurements of dynamical variables of the physical systems. In the following we thus recollect basic notions of probability theory, in order to set the nomenclature used in the main part of the course.

4.1 Probabilities in the discrete case

Let a quantity be s, for instance the outcome of a coin toss, and suppose it can assume only *discrete* values $s_1, \cdots, s_n$ with probabilities $\rho(s_1), \cdots, \rho(s_n)$. The function $\rho(s)$ is the *probability distribution function* (or PDF) of s, and is normalised so that the sum of all probabilities is the certainty:

$$\sum_{i=1}^{n} \rho(s_i) \, = \, 1 \, . \tag{4.1}$$

Examples.

- In an unbiased-coin toss one has $n = 2$ possibilities, namely $s_1 = 0 =$ heads and $s_2 = 1 =$ tails, and $\rho(s_1) = \rho(s_2) = 1/2$, corresponding to the fact that both heads or tails are equally probable.
- Suppose we have performed 10000 coin tosses, and we got heads in 9500, while tails only in 500 cases. We can deduce, a posteriori, that the coin is biased, since $\rho(s_1) = 0.95 \gg \rho(s_2) = 0.05$.
- In an unbiased-dice toss there are $n = 6$ possibilities, and $\rho(s_i) = 1/6$ for all $i = 1, \cdots, 6$.
- If one wants to measure the momentum of a classical particle of mass m and energy E elastically (i.e. without losing kinetic energy) bouncing against the walls of a one-dimensional box, the two possible values $s_1 = \sqrt{2mE}$, $s_2 = -s_1$ are measured with equal probability $\rho(s_i) = 1/2$.

Given a PDF $\rho(s)$, the *expectation value* (or average, or mean value) of s is defined as

$$\langle s \rangle \, \equiv \, \sum_{i=1}^{n} s_i \, \rho(s_i) \, , \tag{4.2}$$

namely it is the sum of possible values s_i that the variable s can assume, each weighed by the probability $\rho(s_i)$ for it to occur.
The expectation value gives an idea of the *position* (in the range in which s can vary) of the region

where the main contributions to the probability distribution are concentrated, as it weighs more values with larger $\rho(s)$.
Examples of expectation-value calculations.

- The unbiased coin has $i = 1, 2$, $s_1 = 0$ (heads), $s_2 = 1$ (tails), $\rho(s_i) = 1/2$:

$$\langle s \rangle = \frac{1}{2} \times 0 + \frac{1}{2} \times 1 = \frac{1}{2} \,. \tag{4.3}$$

  The expectation value is at the same distance from the two possible values.

- As for the biased coin, we have $i = 1, 2$, $s_1 = 0$ (heads), $s_2 = 1$ (tails), $\rho(s_1) = 0.95$, $\rho(s_2) = 0.05$

$$\langle s \rangle = 0.95 \times 0 + 0.05 \times 1 = 0.05 \,. \tag{4.4}$$

  In this case, the expectation value is much closer to 0 (heads) than to 1 (tails), since the probability to get heads is much larger, i.e. $\langle s \rangle$ signals where the peak of probabilities is.

- The unbiased dice has $i = 1, 6$, $s_1 = 1$, $s_2 = 2$, ..., i.e. $s_i = i$, $\rho(s_i) = 1/6$:

$$\langle s \rangle = \sum_{i=1}^{6} \frac{1}{6} i = \frac{7}{2} \,, \tag{4.5}$$

  which again, as in the case of the unbiased coin, is in the middle of the allowed range for s.

- For the bouncing particle $i = 1, 2$, $s_1 = \sqrt{2mE}$, $s_2 = -\sqrt{2mE}$, while $\rho(s_1) = \rho(s_2) = 1/2$:

$$\langle s \rangle = \frac{1}{2}\sqrt{2mE} - \frac{1}{2}\sqrt{2mE} = 0 \,. \tag{4.6}$$

From the examples above it is clear that $\langle s \rangle$ is not necessarily equal to one of the allowed values s_i, however it is always inside the allowed range for s. Moreover, if one of the s_i is such that $\rho(s_i) = 1$, then $\langle s \rangle = s_i$, since $\rho(s_j) = 0$ for all $j \neq i$. For instance, if a dice is totally biased, such that $\rho(s_3) = 1$ and $\rho(s_{i\neq 3}) = 0$, corresponding to a dice whose outcome is always $s_3 = 3$, the expectation value equals the only possible outcome, 3:

$$\langle s \rangle = 0 \times 1 + 0 \times 2 + 1 \times 3 + 0 \times 4 + 0 \times 5 + 0 \times 6 = 3 \,. \tag{4.7}$$

It is useful to define expectation values for functions $f(s)$ of the variable s, through

$$\langle f(s) \rangle \equiv \sum_{i=1}^{n} f(s_i)\, \rho(s_i) \,, \tag{4.8}$$

and particularly important are the expectation values of the powers s^k

$$\langle s^k \rangle \equiv \sum_{i=1}^{n} s_i^k\, \rho(s_i) \,, \tag{4.9}$$

where $\langle s^k \rangle$ is called k-th *moment* of the distribution $\rho(s)$.
The *variance* $(\Delta s)^2 \geq 0$ is defined in terms of the second moment:

$$\begin{aligned}(\Delta s)^2 &\equiv \sum_{i=1}^{n} \Big(s_i - \langle s \rangle\Big)^2 \rho(s_i) = \sum_{i=1}^{n} \Big(s_i^2 + \langle s \rangle^2 - 2 s_i \langle s \rangle\Big) \rho(s_i) \\ &= \underbrace{\sum_{i=1}^{n} s_i^2 \, \rho(s_i)}_{\equiv \langle s^2 \rangle} + \langle s \rangle^2 \underbrace{\sum_{i=1}^{n} \rho(s_i)}_{\equiv 1} - 2 \langle s \rangle \underbrace{\sum_{i=1}^{n} s_i \, \rho(s_i)}_{\equiv \langle s \rangle} \\ &= \langle s^2 \rangle + \langle s \rangle^2 - 2 \langle s \rangle^2 = \langle s^2 \rangle - \langle s \rangle^2 \,, \end{aligned} \tag{4.10}$$

where we have used (4.2) and the fact that $\langle s \rangle$ is independent of i. Note from its very definition that the variance is a non-negative quantity, which implies that $\langle s^2 \rangle \geq \langle s \rangle^2$ is always satisfied. As a consequence, the square root of the variance is a positive quantity, called *standard deviation* Δs.
Examples of variance calculation.

- For the unbiased coin, we have

$$\langle s^2 \rangle = \frac{1}{2} \times 0^2 + \frac{1}{2} \times 1^2 = \frac{1}{2} \,, \qquad \langle s \rangle^2 = \frac{1}{4} \quad \Longrightarrow \quad (\Delta s)^2 = \frac{1}{4} \,. \tag{4.11}$$

- The biased coin gives

$$\langle s^2 \rangle = 0.95 \times 0^2 + 0.05 \times 1^2 = 0.05 \,, \qquad \langle s \rangle^2 = 0.0025 \quad \Longrightarrow \quad (\Delta s)^2 = 0.0475 \,. \tag{4.12}$$

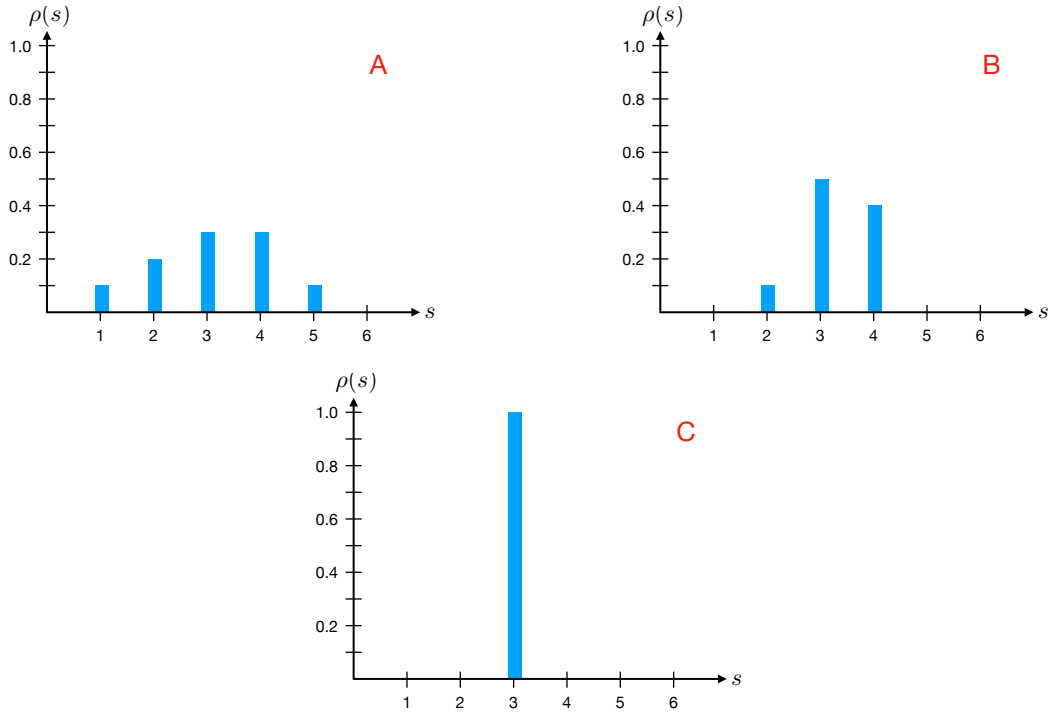
- As for the unbiased dice

$$\langle s^2 \rangle = \sum_{i=1}^{6} \frac{1}{6} i^2 = \frac{91}{6} \,, \qquad \langle s \rangle^2 = \frac{49}{4} \quad \Longrightarrow \quad (\Delta s)^2 = \frac{35}{12} \,. \tag{4.13}$$

- For the bouncing particle we have

$$\langle s^2 \rangle = \frac{1}{2} (\sqrt{2mE})^2 + \frac{1}{2} (\sqrt{2mE})^2 = 2mE \,, \qquad \langle s \rangle^2 = 0 \quad \Longrightarrow \quad (\Delta s)^2 = 2mE \,. \tag{4.14}$$

From its definition, the variance is the expectation value of the squared distance between s_i and the average $\langle s \rangle$, hence it represents a measure of the *broadness* of the probability distribution $\rho(s)$ around its average value: if s_i are all close (or at least if the ones associated with large $\rho(s_i)$ are), then the average distance from $\langle s \rangle$, and consequently the variance, is small, and $\rho(s)$ is 'peaked' around $\langle s \rangle$; the opposite happens if the values s_i and their probabilities are more evenly distributed. In order to visualise how the variance and standard deviation quantify the broadness of the $\rho(s)$ distribution we consider three cases of biased dice, whose PDFs are represented in figure as blue bins. We compute mean and standard deviations for the three cases:



$$
\begin{aligned}
\text{A}: \quad & \langle s \rangle \;=\; 0.1 \times 1 + 0.2 \times 2 + 0.3 \times 3 + 0.3 \times 4 + 0.1 \times 5 \;=\; 3.1\,, \\
& \langle s^2 \rangle \;=\; 0.1 \times 1 + 0.2 \times 4 + 0.3 \times 9 + 0.3 \times 16 + 0.1 \times 25 \;=\; 10.9 \\
& \Longrightarrow \qquad \Delta s \;=\; \sqrt{\langle s^2 \rangle - \langle s \rangle^2} \;=\; 1.136\,, \\
\\
\text{B}: \quad & \langle s \rangle \;=\; 0.1 \times 2 + 0.5 \times 3 + 0.4 \times 4 \;=\; 3.3\,, \\
& \langle s^2 \rangle \;=\; 0.1 \times 4 + 0.5 \times 9 + 0.4 \times 16 \;=\; 11.3 \\
& \Longrightarrow \qquad \Delta s \;=\; \sqrt{\langle s^2 \rangle - \langle s \rangle^2} \;=\; 0.640\,, \\
\\
\text{C}: \quad & \langle s \rangle \;=\; 1 \times 3 \;=\; 3\,, \\
& \langle s^2 \rangle \;=\; 1 \times 9 \;=\; 9 \\
& \Longrightarrow \qquad \Delta s \;=\; \sqrt{\langle s^2 \rangle - \langle s \rangle^2} \;=\; 0\,. \qquad (4.15)
\end{aligned}
$$

We see that in case C the standard deviation is 0: there is only one possible outcome, no variability in the result, hence we are certain that we will get 3 after a dice toss. The probability distribution is perfectly peaked at the only possible value. In case B the variance is non-zero, corresponding to the fact that the result of a dice toss can deviate from the mean, i.e. the probability distribution is broader than in case C. Case A has an even broader distribution, and correspondingly a larger variance. In the case of an unbiased dice the PDF features six bins of height $1/6$, hence it is broader than case A: indeed its standard deviation is easily computed to be $\Delta s = \sqrt{35/12} = 1.707$.

4.2 Measuring probabilities

The above examples are sufficient not only to introduce basic statistical concepts, but also to give an idea of how to *measure* probability distributions, which underlies many topics in Quantum Mechanics.
The logic is the one used in the above examples to establish $\rho(s)$ for the biased coin: one simply repeats N times the same 'experiment' under the same conditions, and reads $\rho(s_i)$ as the number N_i of experiments which gave s_i as an outcome, divided by the total number of experiments, i.e. $\rho(s_i) \sim N_i/N$. In the case of a coin, an 'experiment' is just a single coin toss, and collecting the frequency with which the different outcomes occur one builds information on the underlying probability distributions.
There are two assumptions underlying this strategy.

- The 'measured' system and the experimental conditions in which the 'measurement' is performed are identical at each new experiment: in the case of a coin, identical experimental conditions are achieved if one is repeatedly tossing *the same* coin.
  In the case of a physical experiment, one cannot repeatedly use *the same* system (e.g. the same electron), however one always sets oneself in the condition of having N *identical* physical systems (e.g. N identical silver atom coming out of the same oven with the same kinetic energy and angular momentum, ...), measured with the same experimental apparatus. If this is the case, measuring probabilities on physical systems proceeds in the same way as measuring probabilities on coins or dice, namely counting frequencies.

- In order to deduce $\rho(s)$ from N identical experiments, the number of experiments should be sufficiently *large*, so that formally one would write $\rho(s_i) = \lim_{N\to\infty} N_i/N$. For small N, the effect of random fluctuations is important, and the deduced PDF is barely reliable. For instance, if we toss an unbiased coin only twice, and get twice the same value (either heads or tails – this happens in 50% of cases), we would be tempted to say that the coin is biased, namely it would be quite easy to deduce a wrong information on $\rho(s)$. If we keep tossing, the frequencies N_i/N with which the various outcomes occur tend to stabilise around $\rho(s_i)$, and the relative difference $[N_i/N - \rho(s_i)]/\rho(s_i)]$ scales as $1/\sqrt{N}$, i.e. the larger N, the smaller the effect of random fluctuations, and the better the estimate of $\rho(s_i)$ by means of N_i/N.

4.3 Continuous case

It is possible to generalise the concepts exposed above to the *continuous* case, namely to the case in which s varies in a continuous range, $s \in [a, b]$. For instance, for the above mentioned particle bouncing elastically back and forth in a one-dimensional box of width L, we could let s be the particle's position, which can range in $[0, L]$.
We can start by introducing the PDF $\rho(s)$, subject to the normalisation condition

$$\int_a^b ds\, \rho(s) \;=\; 1\,, \tag{4.16}$$

which is just the continuous version of (4.1), in which the sum over all possible discrete values s_i has become an integral over the whole interval in which the continuous variable s can range.
The normalisation condition immediately tells that $\rho(s)$ has dimensions of $1/s$ (e.g. if s is position, as in the case of the particle in a box, $\rho(s)$ has dimension of inverse length), namely it represents the probability *per unit of* s to find a give value of s.

The normalisation condition states the *certainty* that s be within the $[a, b]$ range. The probability that s be in a reduced range $[a', b']$ is instead

$$P(a' \leq s \leq b') \,=\, \int_{a'}^{b'} ds\, \rho(s)\,, \tag{4.17}$$

and, if we let the interval be thinner and thinner, $\rho(s)\, ds$ represents the probability to find the variable in range $[s, s + ds]$. This highlights that, upon knowing $\rho(s)$, one is able to measure probabilities for generic intervals in the allowed range $[a, b]$.
In the continuous case, the definitions for the moments of ρ proceed as in the discrete one:

$$\begin{aligned}
\langle s \rangle &= \int_a^b ds\, s\, \rho(s)\,, \\
\langle f(s) \rangle &= \int_a^b ds\, f(s)\, \rho(s)\,, \\
\langle s^k \rangle &= \int_a^b ds\, s^k\, \rho(s)\,, \\
\big(\Delta s\big)^2 &= \int_a^b ds\, \big(s - \langle s \rangle\big)^2 \rho(s) \,=\, \langle s^2 \rangle - \langle s \rangle^2\,.
\end{aligned} \tag{4.18}$$

As an example, we can imagine a the above particle trapped in a box, with position $s \in [0, L]$, bouncing elastically between the walls. The particle spends on average the same amount of time in each unit length inside the box, whence its PDF is $\rho(s) = k$, with k constant. The normalisation condition allows to determine k:

$$1 \,=\, \int_0^L ds\, k \,=\, k\, L\,, \qquad \Longrightarrow \qquad k \,=\, 1/L \,=\, \rho(s)\,. \tag{4.19}$$

The average position is

$$\langle s \rangle \,=\, \int_0^L ds\, s \frac{1}{L} \,=\, \frac{L}{2}\,,$$

while the variance is

$$\big(\Delta s\big)^2 \,=\, \int_0^L ds\, s^2 \frac{1}{L} - \left(\int_0^L ds\, s \frac{1}{L}\right)^2 \,=\, \frac{L^2}{3} - \frac{L^2}{4} = \frac{L^2}{12} \qquad \Longrightarrow \qquad \Delta s \,=\, \frac{L}{2\sqrt{3}}\,.$$

In order to measure $\rho(s)$ in the continuous case, one would just proceed in the same way as explained in the discrete case, by repeating experiments many times in the same conditions. For instance, in the case of classical particle in a box, one would ideally look inside the box at N subsequent random times, and record each time the position of the particle. The probability for particle to be in range $[s, s + ds]$ would just be the number of times the particle was found in that range divided by the total number of counts N, and the smaller the considered range ds, the larger N should be, in order to get a precise determination of the corresponding $\rho(s)$.

Chapter 5

Solutions to the exercises

Exercise 1

$z_1^2 = (3+5i)(3+5i) = -16+30i$;

$z_3^3 = (3i)(3i)(3i) = -9(3i) = -27i$;

$z_2^2 - z_3 = (2-4i)(2-4i) - 3i = (-12-16i) - 3i = -12-19i$;

$2z_1 - 3z_2 + az_3 = 6+10i-6+12i+3ai = 22i+3ai = 0$, so $a = -22/3$.

Exercise 2

$z_1 = \sqrt{3}+i$, $|z_1| = \sqrt{3+1} = 2$; $\arg(z) = \arctan(1/\sqrt{3}) = \pi/6$, so $z_1 = 2\,e^{i\pi/6}$.

$z_2 = 5-5i$, $|z_2| = \sqrt{25+25} = 5\sqrt{2}$; $\arg(z) = \arctan(-1) = -\pi/4$, so $z_2 = 5\sqrt{2}\,e^{-i\pi/4}$.

$z_1 \cdot z_2^* = (\sqrt{3}+i)5(1+i) = 5[(\sqrt{3}-1)+i(\sqrt{3}+1)]$; in polar coordinates $z_1 \cdot z_2^* = 2\,e^{i\pi/6}\,5\sqrt{2}\,e^{i\pi/4} = 10\sqrt{2}\,e^{i5\pi/12}$.

$z_1/z_2 = \frac{\sqrt{3}+i}{5(1-i)} = \frac{(\sqrt{3}+i)(1+i)}{5(1-i)(1+i)} = \frac{\sqrt{3}-1+i(\sqrt{3}+1)}{10}$; in polar coordinates $z_1/z_2 = \frac{2\,e^{i\pi/6}}{5\sqrt{2}\,e^{-i\pi/4}} = \frac{\sqrt{2}}{5}e^{i5\pi/12}$.

Exercise 3

Given $e^{\pm i\theta} = \cos\theta \pm i\sin\theta$, one has $e^{i\theta}+e^{-i\theta} = 2\cos\theta$, while $e^{i\theta}-e^{-i\theta} = 2i\sin\theta$, whence

$$\cos\theta = \frac{e^{i\theta}+e^{-i\theta}}{2}, \qquad \sin\theta = \frac{e^{i\theta}-e^{-i\theta}}{2i}. \tag{5.1}$$

Exercise 4

Using that $\cos\theta = \mathrm{Re}(e^{i\theta})$, and $\sin\theta = \mathrm{Im}(e^{i\theta})$, one has

$$\begin{aligned}
\cos(a+b) &= \mathrm{Re}(e^{i(a+b)}) = \mathrm{Re}(e^{ia}e^{ib}) = \mathrm{Re}[(\cos a + i\sin a)(\cos b + i\sin b)] \\
&= \mathrm{Re}[\cos a\cos b - \sin a\sin b + i(\cos a\sin b + \sin a\cos b)] \\
&= \cos a\cos b - \sin a\sin b\,, \qquad (5.2)\\
\sin(a-b) &= \mathrm{Im}(e^{i(a-b)}) = \mathrm{Im}(e^{ia}e^{-ib}) = \mathrm{Im}[(\cos a + i\sin a)(\cos b - i\sin b)] \\
&= \mathrm{Im}[\cos a\cos b + \sin a\sin b + i(-\cos a\sin b + \sin a\cos b)] \\
&= -\cos a\sin b + \sin a\cos b\,. \qquad (5.3)
\end{aligned}$$

Exercise 5

The squared modulus of $\psi_1 + \psi_2$ reads

$$\begin{aligned} |\psi_1+\psi_2|^2 &= (\psi_1+\psi_2)(\psi_1+\psi_2)^* = (\psi_1+\psi_2)(\psi_1^*+\psi_2^*) \\ &= |\psi_1|^2 + |\psi_2|^2 + \psi_1\psi_2^* + \psi_1^*\psi_2 = a_1^2 + a_2^2 + a_1a_2\, e^{i(b_1-b_2)} + a_1a_2\, e^{i(b_2-b_1)} \\ &= a_1^2 + a_2^2 + 2a_1a_2\frac{e^{i(b_1-b_2)} + e^{-i(b_1-b_2)}}{2} = a_1^2 + a_2^2 + 2a_1a_2\cos(b_1-b_2)\,. \end{aligned} \tag{5.4}$$

The total intensity is not just the sum of the two intensities (a_1^2 and a_2^2), but there is an extra modulation term that depends on the phase difference between the two waves. If such a difference is 0 (π) the total intensity reaches its maximum (minimum), corresponding to the phenomenon of constructive (destructive) *interference*.

Exercise 6

$$\begin{aligned} \int_A^{A+4\pi} d\theta\, \cos\theta &= \frac{1}{2}\int_A^{A+4\pi} d\theta\, (e^{i\theta} + e^{-i\theta}) = \frac{1}{2i}\left[e^{i\theta} - e^{-i\theta}\right]_A^{A+4\pi} \\ &= \frac{1}{2i}\left[e^{iA}e^{i4\pi} - e^{iA} - e^{-iA}e^{-i4\pi} + e^{-iA}\right]. \end{aligned} \tag{5.5}$$

Since $e^{\pm i4\pi} = 1$, the above result is

$$\int_A^{A+4\pi} d\theta\, \cos\theta = \frac{1}{2i}\left[e^{iA} - e^{iA} - e^{-iA} + e^{-iA}\right] = 0\,. \tag{5.6}$$

$$\begin{aligned} \int_A^{A+2\pi} d\theta\, \cos^2\theta &= \frac{1}{4}\int_A^{A+2\pi} d\theta\, (e^{i\theta} + e^{-i\theta})^2 = \frac{1}{4}\int_A^{A+2\pi} d\theta\, (e^{2i\theta} + e^{-2i\theta} + 2) \\ &= \frac{1}{4}\Big[\frac{e^{2iA+4i\pi} - e^{2iA}}{2i} + \frac{e^{-2iA-4i\pi} - e^{-2iA}}{-2i} + 2(A+2\pi-A)\Big] \\ &= \frac{1}{4}\Big[\frac{e^{2iA} - e^{2iA}}{2i} + \frac{e^{-2iA} - e^{-2iA}}{-2i} + 4\pi\Big] = \pi\,. \end{aligned} \tag{5.7}$$

Exercise 7

The physical dimension of the vacuum permittivity ϵ_0 can be deduced by recalling the expression of the electric potential generated at a distance r by a pointlike charge q, namely $V \sim q/(\epsilon_0 r)$, where we have neglected dimensionless constants, irrelevant for this problem. The dimension of ϵ_0 is thus that of a charge divided by potential divided by length, or equivalently square charge divided by energy divided by length.

Boltzmann's constant k_B appears in the state equations of perfect gases, $PV = Nk_BT$, hence it is easily deduced to have the physical dimension of pressure times volume divided by temperature, namely energy divided by temperature. Analogously one can recall it appearing in thermodynamics in Boltzmann's suppression factors e^{-E/k_BT}, whence one reaches the same conclusion about its dimension.

Planck's constant h appears in Planck's law $E = h\nu$, hence it has the dimension of energy divided by frequency, or equivalently energy times time. Angular momentum is a length times a momentum, namely length times mass times velocity, or equivalently energy times time, hence h has the same physical dimensions as angular momentum.

Exercise 8

Since kx^2 has the dimension of an energy, and x of a length, then the physical dimension of k is energy over square length, or equivalently mass over square time, $\mathrm{kg/s^2}$ in the IS.
The parameters characterising the spring are k and m, whence the oscillation frequency is a function of those. Since k has the dimension of mass over square time, and m that of a mass, the only way to combine the two to get a frequency (namely an inverse time) is dividing k by m and taking the square root, hence the typical oscillation frequency is $\nu \sim \sqrt{k/m}$.

Exercise 9

The integration variable x has the dimension of a length, as can be deduced by the fact that x/b and x/a must be dimensionless (they appear as arguments of functions). Since dx (as x) is a length, P itself has the dimension of a length.
Concerning the first integral, we know it has the dimension of a length, and it just depends on the b parameter, hence its result will be b times a dimensionless number. Dimensional analysis is not able to get this number, which however can be first estimated to be of order 1 (since the pure numbers involved in the integrand are of this order), and then explicitly obtained with a trivial calculation of the integral itself. The result of the first integral is $2\,b$.
The second integral would involve a complicated calculation. However we know that (i) it has the dimension of a length (otherwise it could not be summed to the first integral, nor equated to P), and (ii) it just depends on a. The result of the second integral will then be, purely on dimensional grounds, $k\,a$, with k a dimensionless number, of order 1 (again, the involved numbers in the integrand are such). The second integral contributes very little to P with respect to the first, namely $k\,a/(2b) = k/200$, and for a $O(1\%)$ estimate of P it can be safely neglected.
Dimensional analysis (plus the trivial computation of the first integral) has thus immediately guided us to $P \sim 2b = 200$ m. The computation of the second integral would have shown that $k = 0.19175$, hence our estimate, based on physical dimensions, is accurate at the permille level!

Exercise 10

In the original basis, $\vec{v}_3 = -\vec{e}_1 + 1.4\vec{e}_2$, represented as column vector:

$$\begin{pmatrix} -1 \\ 1.4 \end{pmatrix}. \tag{5.8}$$

Exercise 11

By linearity, $\vec{v} \cdot \vec{y} = 3\lambda_1 + 5\lambda_2$. They are orthogonal for $\lambda_1 = -\frac{5}{3}\lambda_2$, (with $\lambda_1^2 + \lambda_2^2 \neq 0$ in order to have a nontrivial vector).

Exercise 12

- Imposing orthogonality we find $\vec{w} = \lambda\,(-\alpha_2\vec{e}_1 + \alpha_1\vec{e}_2)$ is orthogonal to $\vec{v}$, with arbitrary $\lambda \in \mathbb{R}$, $\lambda \neq 0$.
- We can construct a orthonormal basis with $\vec{e}'_1 \propto \vec{v}$, $\vec{e}'_2 \propto \vec{w}$ by normalising these two vectors:

$$\vec{e}'_1 = \pm\frac{\alpha_1\vec{e}_1 + \alpha_2\vec{e}_2}{\sqrt{\alpha_1^2+\alpha_2^2}}, \quad \vec{e}'_2 = \pm\frac{-\alpha_2\vec{e}_1 + \alpha_1\vec{e}_2}{\sqrt{\alpha_1^2+\alpha_2^2}}. \tag{5.9}$$

where the $\pm$ signs are arbitrary (any choice of signs still gives an orthonormal basis). It is easy to verify $\vec{e}'_i \cdot \vec{e}'_2 = \delta_{ij}$.

Exercise 13

With the rules of matrix multiplication it is easy to verify

$$\begin{pmatrix}\cos\varphi & -\sin\varphi\\ \sin\varphi & \cos\varphi\end{pmatrix}\cdot\begin{pmatrix}1\\0\end{pmatrix} = \begin{pmatrix}\cos\varphi\\ \sin\varphi\end{pmatrix}, \quad \begin{pmatrix}\cos\varphi & -\sin\varphi\\ \sin\varphi & \cos\varphi\end{pmatrix}\cdot\begin{pmatrix}0\\1\end{pmatrix} = \begin{pmatrix}-\sin\varphi\\ \cos\varphi\end{pmatrix}. \tag{5.10}$$

Exercise 14

We can just impose

$$\vec{v} = \alpha\vec{e}'_1 + \beta\vec{e}'_2 \to \begin{pmatrix}1\\2\end{pmatrix} = \begin{pmatrix}\alpha\cos\phi - \beta\sin\phi\\ \alpha\sin\phi + \beta\cos\phi\end{pmatrix} \tag{5.11}$$

. From the first equation we obtain $\alpha = \beta\tan\phi - 1/\cos\phi$, and substitution in the second equation gives $\alpha = \cos\phi + 2\sin\phi$ and $\beta = 2\cos\phi - \sin\phi$.

Alternative (more powerful method) A quicker way to obtain this result is the following. We have a linear system written as:

$$\begin{pmatrix}1\\2\end{pmatrix} = M\cdot\begin{pmatrix}\alpha\\\beta\end{pmatrix}, \tag{5.12}$$

with the matrix of coefficients

$$M = \begin{pmatrix}\cos\phi & -\sin\phi\\ \sin\phi & \cos\phi\end{pmatrix}. \tag{5.13}$$

We can multiply by the inverse matrix[1]

$$M^{-1} = \begin{pmatrix}\cos\phi & \sin\phi\\ -\sin\phi & \cos\phi\end{pmatrix}, \tag{5.14}$$

then the linear system becomes

$$M^{-1}\cdot\begin{pmatrix}1\\2\end{pmatrix} = \begin{pmatrix}\alpha\\\beta\end{pmatrix}, \tag{5.15}$$

[1]Check that it can be obtained with the rules explained in the main text for computing the inverse of a matrix. You can easily check that it satisfies $M\cdot M^{-1} = M^{-1}\cdot M = I$.

and from this, using M^{-1} given above, we immediately just *read* the solution for α and β. Notice that here we do not need to do any computation. All the computation is encoded in our calculation of the inverse matrix. This shows why matrices are so useful - using them we can solve linear systems of equations without needing to solve all equations one by one. We just need to follow the rules to invert the matrix of coefficients.

This method is more powerful, especially, because it lets us solve, without needing to do any new computation, the following more general question:

- What are the coefficients (α, β) such that $\vec{v} = \alpha\vec{e}'_1 + \beta\vec{e}'_2$, where $\vec{v}$ is a *generic* vector $\vec{v} = \sum_i v_i \vec{e}_i$?

In fact, the answer is given by solving the linear system

$$\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = M \cdot \begin{pmatrix} \alpha \\ \beta \end{pmatrix}, \tag{5.16}$$

and we can just solve it as before using the inverse matrix:

$$\begin{pmatrix} \alpha \\ \beta \end{pmatrix} = M^{-1} \cdot \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} \cos\phi v_1 + \sin\phi v_2 \\ -\sin\phi v_1 + \cos\phi v_2 \end{pmatrix}. \tag{5.17}$$

Notice that the matrix of coefficients M (and thus M^{-1}) is the same as before! Therefore, we solve in one go the problem of rewriting *any* vector $\vec{v}$ in the new basis.

Exercise 15

The given matrix rotates a 3D vector as

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} \to \begin{pmatrix} \cos\theta x - \sin\theta y \\ \cos\theta y + \sin\theta x \\ z \end{pmatrix}, \tag{5.18}$$

therefore it performs a rotation of angle θ in the xy-plane, clockwise if we are looking in the direction of the z-axis.

If we want to perform a rotation around the x-axis, we should use the matrix

$$R_x(\theta) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{pmatrix}. \tag{5.19}$$

while a rotation around the y-axis is given by

$$R_y(\theta) = \begin{pmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{pmatrix}. \tag{5.20}$$

Exercise 16

$$A \cdot B = \begin{pmatrix} -9 & 27 & 30 \\ -23 & 16 & 10 \end{pmatrix}. \tag{5.21}$$

Exercise 17

$$[\sigma_1, \sigma_3] = \begin{pmatrix} 0 & -2 \\ 2 & 0 \end{pmatrix}. \tag{5.22}$$

Exercise 18

Considering a matrix with generic coefficients $M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, we should impose $M \cdot \vec{v}$ for any vector. It is enough to consider two cases: $\vec{v} = \vec{e}_1$ and $\vec{v} = \vec{e}_2$. These conditions give equations which completely fix the coefficients of M: one finds $a = d = 1$, $c = d = 0$, namely $M = Id$.

Exercise 19

$$\begin{aligned} R_{\phi_1} \cdot R_{\phi_2} &= \begin{pmatrix} \cos\phi_1 \cos\phi_2 - \sin\phi_1 \sin\phi_2 & -\cos\phi_1 \sin\phi_2 - \sin\phi_1 \cos\phi_2 \\ \sin\phi_1 \cos\phi_2 + \cos\phi_1 \sin\phi_2 & -\sin\phi_1 \sin\phi_2 + \cos\phi_1 \cos\phi_2 \end{pmatrix} && (5.23) \\ &= \begin{pmatrix} \cos(\phi_1 + \phi_2) & -\sin(\phi_1 + \phi_2) \\ \sin(\phi_1 + \phi_2) & \cos(\phi_1 + \phi_2) \end{pmatrix}. && (5.24) \end{aligned}$$

Exercise 20

$$R_\phi \cdot R_\phi^T = \begin{pmatrix} \cos^2\phi + (-1)^2 \sin^2\phi & -\cos\phi \sin\phi + \sin\phi \cos\phi \\ \sin\phi \cos\phi - \cos\phi \sin\phi & \sin^2\phi + \cos^2\phi \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{5.25}$$

Exercise 21

- $\vec{w} = (8 - i)\vec{e}_1 + (2 + 5i)\vec{e}_2$.
- This is again a problem of writing a known vector in a new basis. It can be solved with the same method as above (see Exercise 13), just using complex numbers. The result is $\vec{w} = \alpha \vec{f}_1 + \beta \vec{f}_2$ with $\alpha = 6 + i$, $\beta = 2 - 2i$.

Exercise 22

Following the instructions we construct an orthonormal basis of the form

$$\vec{f}_1 = e^{i\alpha_1} \frac{\vec{e}_1 + (3 + i)\vec{e}_2}{\sqrt{11}}, \quad \vec{f}_2 = e^{i\alpha_2} \frac{(3 - i)\vec{e}_1 - \vec{e}_2}{\sqrt{11}}, \tag{5.26}$$

where α_1, α_2 are arbitrary real phases. The orthonormality $\vec{f}_i^\dagger \cdot \vec{f}_j = \delta_{ij}$ is easy to check.

Exercise 23

- It is immediate the verify that $\vec{f}_i^{\dagger} \cdot \vec{f}_j = \delta_{ij}$.
- $\vec{v} = \alpha \vec{f}_1 + \beta \vec{f}_2$ with $\alpha = \vec{f}_1^{\dagger} \cdot \vec{v} = \frac{1+(4+6i)e^{-\frac{i\pi}{3}}}{\sqrt{5}}$, $\beta = \vec{f}_2^{\dagger} \cdot \vec{v} = \frac{(2+3i)-2e^{\frac{i\pi}{3}}}{\sqrt{5}}$.

Exercise 24

The matrix form of the completeness relation, for the two bases described in the exercise, is:

- $$\begin{pmatrix}1\\0\end{pmatrix} \cdot \begin{pmatrix} 1 & 0 \end{pmatrix} + \begin{pmatrix}0\\1\end{pmatrix} \cdot \begin{pmatrix} 0 & 1 \end{pmatrix} = \begin{pmatrix}1 & 0\\0 & 1\end{pmatrix}.$$
- $$\frac{1}{5}\begin{pmatrix}1\\2e^{i\pi/3}\end{pmatrix} \cdot \begin{pmatrix} 1 & 2e^{-i\pi/3} \end{pmatrix} + \frac{1}{5}\begin{pmatrix}-2e^{-i\pi/3}\\1\end{pmatrix} \cdot \begin{pmatrix} -2e^{i\pi/3} & 1 \end{pmatrix} = \begin{pmatrix}1 & 0\\0 & 1\end{pmatrix}.$$

Exercise 25

- The basis is given by $|1'\rangle' = e^{i\alpha_1}\frac{|1\rangle+i|2\rangle}{\sqrt{2}}$, $|2'\rangle = e^{i\alpha_2}\frac{|1\rangle-i|2\rangle}{\sqrt{2}}$, with $\alpha_i \in \mathbb{R}$ arbitrary phases. We can make any choice of these phases: we will set $\alpha_i = 0$ in the following.
- Using the completeness relation $I = |1'\rangle\langle 1'| + |2'\rangle\langle 2'|$, we find $|v\rangle = |1'\rangle(\langle 1'|v\rangle) + |2'\rangle(\langle 2'|v\rangle)$, so the coefficients of the decomposition are
$$\langle 1'|v\rangle = \frac{\frac{23}{7}+9i}{\sqrt{2}}, \quad \langle 2'|v\rangle = -\frac{\frac{19}{7}+5i}{\sqrt{2}}. \tag{5.27}$$
- $$\langle w|v\rangle = (-i\langle 1| + \langle 2|)\left((\frac{2}{7}+2i)|1\rangle + (-7+3i)|2\rangle\right) = -i\frac{2}{7} + 2 + (-7+3i) = -5 + \frac{19}{7}i,$$
$$\langle v|w\rangle = (\langle w|v\rangle)^* = -5 - \frac{19}{7}i,$$
$$\langle z|v\rangle = (2-3i)\langle 1'|v\rangle = e^{-i\alpha_1}(2-3i)\frac{\frac{23}{7}+9i}{\sqrt{2}} = \frac{\frac{235}{7}+\frac{57i}{7}}{\sqrt{2}},$$
$$\langle v|z\rangle = (\langle z|v\rangle)^* = \frac{\frac{235}{7}-\frac{57i}{7}}{\sqrt{2}}.$$

Exercise 26

- $\sigma_2|v\rangle = iv_1|2\rangle - iv_2|1\rangle$.
- The matrix elements are given by $(\sigma_2)_{ij} = \langle i|\sigma_2|j\rangle$, hence the matrix form is
$$\langle i|\sigma_2|j\rangle = \begin{pmatrix}0 & -i\\ i & 0\end{pmatrix}. \tag{5.28}$$
- In the other basis $|1'\rangle$, $|2'\rangle$, the matrix elements are given by $\langle i'|\sigma_2|j'\rangle$, which leads to the matrix
$$\langle i'|\sigma_2|j'\rangle = \begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}. \tag{5.29}$$

Exercise 27

The hermitian matrices in this list are M_3 and M_4.

Exercise 28

For the case $N = 2$ we can consider the permutations of two elements. There are two possible permutations:

$$\sigma_1 : \begin{pmatrix} 1 & 2 \\ 1 & 2 \end{pmatrix} \longrightarrow (-1)^{\sigma_1} = 1, \tag{5.30}$$

or the exchange:

$$\sigma_2 : \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} \longrightarrow (-1)^{\sigma_2} = -1, \tag{5.31}$$

Applying the formula we then get:

$$\det A = (+1)A_{1\sigma_1(1)}A_{2\sigma_1(2)} - A_{1\sigma_2(1)}A_{2\sigma_1(2)} = A_{11}A_{22} - A_{12}A_{21}, \tag{5.32}$$

which agrees with the formula given in the main text.

For $N = 3$, we have six independent orderings: three associated with the plus sign, i.e. the permutations sending 123 into 123, 231 and 312, and three permutations with negative sign, which send 123 to 213, 321 or 132.

The determinant is then the sum of six terms:

$$\det A = A_{11}A_{22}A_{33} + A_{12}A_{23}A_{31} + A_{13}A_{21}A_{32} - A_{12}A_{21}A_{33} - A_{11}A_{23}A_{32} - A_{13}A_{22}A_{31}, \tag{5.33}$$

also matching the formula given in the main text.

Exercise 29

For a 2×2 matrix

$$M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \tag{5.34}$$

the adjugate matrix is (following the definition given in the main text)

$$\mathrm{adj}M = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}, \tag{5.35}$$

we then see that the inverse matches formula (3.87).

For a 3×3 matrix

$$M = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}, \tag{5.36}$$

the adjugate matrix is

$$\mathrm{adj}M = \begin{pmatrix} ei - fh & ch - bi & bf - ce \\ fg - di & ai - cg & cd - af \\ dh - eg & bg - ah & ae - bd \end{pmatrix}, \tag{5.37}$$

and dividing by the determinant we find the inverse

$$M^{-1} = \frac{1}{-ceg + bfg + cdh - afh - bdi + aei} \begin{pmatrix} ei - fh & ch - bi & bf - ce \\ fg - di & ai - cg & cd - af \\ dh - eg & bg - ah & ae - bd \end{pmatrix}. \tag{5.38}$$

Exercise 30

The matrix M_1 has eigenvalues $\lambda_1 = 2$, $\lambda_2 = 0$, corresponding respectively to the eigenvectors

$$|e_1\rangle = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad |e_2\rangle = \begin{pmatrix} 1 \\ -1 \end{pmatrix}. \tag{5.39}$$

Now we want to find the decomposition $M_1 = PA_{\text{diag}}P^{-1}$. From the eigenvalues we see that[2]

$$A_{\text{diag}} = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}. \tag{5.40}$$

The matrix P is obtained as the matrix which has the eigenvectors as columns. Thus,

$$P = \begin{pmatrix} \langle 1|e_1\rangle & \langle 1|e_2\rangle \\ \langle 2|e_1\rangle & \langle 2|e_2\rangle \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \tag{5.41}$$

It is immediate to verify that indeed $M_1 = PA_{\text{diag}}P^{-1}$.

Repeating the analysis for M_2, you should find the eigenvalues $\lambda_1 = 2$, $\lambda_2 = 2$, $\lambda_3 = 0$ (the order is conventional), corresponding to the eigenvectors (the normalisation is conventional),

$$|e_1\rangle = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \quad |e_2\rangle = \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix}, \quad |e_3\rangle = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} \tag{5.42}$$

We then have a decomposition $M_2 = PA_{\text{diag}}P^{-1}$ with

$$A_{\text{diag}} = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{5.43}$$

and

$$P = \begin{pmatrix} 1 & 2 & -1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}. \tag{5.44}$$

Exercise 31

The matrix σ_2 has two eigenvalues $\lambda_\pm = \pm 1$, and corresponding eigenvectors

$$|+\rangle = \begin{pmatrix} -i \\ 1 \end{pmatrix}, \quad |-\rangle = \begin{pmatrix} i \\ 1 \end{pmatrix}. \tag{5.45}$$

[2]Notice that we chose a conventional order of the two eigenvalues. We could have chosen the reverse order, and this would simply have meant a reordering of the columns of P and of the elements on the diagonal of A_{diag}.

Thus, it has a decomposition $\sigma_2 = PA_{\text{diag}}P^{-1}$ with

$$A_{\text{diag}} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \qquad P = \begin{pmatrix} -i & i \\ 1 & 1 \end{pmatrix}. \tag{5.46}$$

The other matrix σ_3 is already diagonal, with eigenvalues ± 1 and eigenvectors

$$|+\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad |-\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \tag{5.47}$$

so clearly in this case P is the identity matrix and $A_{\text{diag}} = \sigma_3$.

Exercise 32

Let us consider first σ_3, which is diagonal. Then, we find simply

$$\sin(\theta\sigma_3) = \begin{pmatrix} \sin(\theta) & 0 \\ 0 & \sin(-\theta) \end{pmatrix} = \sin(\theta) \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \tag{5.48}$$

and similarly

$$e^{i\theta\sigma_3} = \begin{pmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{pmatrix}. \tag{5.49}$$

Notice that for θ a real number this matrix is unitary. This is indeed as it should be, since it is the exponential of a Hermitian matrix.

Let us now consider the same functions but evaluated on σ_2. We have already found the decomposition $\sigma_2 = PA_{\text{diag}}P^{-1}$ (see the previous exercise), i.e.

$$\sigma_2 = \underbrace{\begin{pmatrix} -i & i \\ 1 & 1 \end{pmatrix}}_{=P} \cdot \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \cdot \underbrace{\frac{1}{2}\begin{pmatrix} i & 1 \\ -i & 1 \end{pmatrix}}_{=P^{-1}} \tag{5.50}$$

Then we find

$$\sin(\theta\sigma_2) = \begin{pmatrix} -i & i \\ 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} \sin(\theta) & 0 \\ 0 & -\sin(\theta) \end{pmatrix} \cdot \frac{1}{2}\begin{pmatrix} i & 1 \\ -i & 1 \end{pmatrix} = \sin(\theta) \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \tag{5.51}$$

and

$$e^{i\theta\sigma_2} = \begin{pmatrix} -i & i \\ 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{pmatrix} \cdot \frac{1}{2}\begin{pmatrix} i & 1 \\ -i & 1 \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}. \tag{5.52}$$

Again, we see that for θ real, then $e^{i\theta\sigma_2}$ is unitary, consistent with the fact that it is the exponential of a Hermitian matrix.