Chapter 2

Linear algebra

The modelling of linear systems is one of the most powerful and versatile tools available to the natural scientist, in both theoretical and applied areas. Broadly speaking, a system is referred to as *linear* if some measurable quantity of interest (*e.g.*, an electric or magnetic field) responds in a linear way – *i.e.*, loosely, 'proportionally' – to some other observable quantity (*e.g.*, the density of electric charge or current).

Notice that linear relationships may involve not only scalar quantities (represented by single real or complex numbers), but vector quantities as well, as is the case for the relationship between the electromagnetic fields and their sources, mentioned above and described by Maxwell equations (see Chapter 1). The relation between electromagnetic fields and their sources is 'linear' precisely because Maxwell equations are 'linear' (that is, they include only the first power of fields and sources).[1] Evidence of the linearity of electromagnetism is ubiquitous: a single microwave communication link is ordinarily exploited simultaneously for hundreds of phone calls, which superpose without disrupting each other. Travelling electromagnetic waves can hence be 'superposed': mathematically, this allows for their description in terms of a *linear space*. This description is especially powerful and fully accounts for interference phenomena, the typical signature of the wave-like behaviour of light.

A very similar description has been borrowed to describe the behaviour of microscopic, quantum mechanical systems. Here, as well, interference effects arise for the probabilities of particles to be found at different places or in different internal states. Here, as well, the 'states' of such particles are described by elements of a vector space (in this case called a *Hilbert space*).

Moreover, linear relationships often offer a good approximation of complex, generally non-linear problems, so that one is lead to consider approximated 'linearised' versions of such problems, which can be more or less accurate (if these approximations are accurate, one says the system is a 'linear regime'). The description of linear systems thus plays a central role in applications as well.

At the heart of the mathematical description of linear systems lies the concept of *linear transformation* between vectors. Such are the transformations that can be represented as matrices if the dimension of the linear space is finite (which will always be assumed to be the case in these lectures). The central aim of this chapter

[1] This does not mean that the *dynamics* derived from Maxwell equations is linear: in general, the movement of electric charges is affected by the electromagnetic fields and, in turn, generates them, thus inducing complex, non-linear dynamics. The relationships between density and current of charges and fields is still linear, though.

is introducing and developing a powerful characterisation of linear transformations, based on the definition of their *eigenvalues* and associated *eigenvectors*, which has bearings on most facets and areas of physics and the other natural sciences.

2.1 Real and complex vector spaces. Inner product.

A *complex* vector space $\mathcal{H}$ is a slight generalisation of a *real* vector space. It is defined by the property that

$$\forall\ \mathbf{v}_1, \mathbf{v}_2 \in \mathcal{H} \quad \text{and} \quad \forall\ a_1, a_2 \in \mathbb{C} \quad : \quad (a_1\mathbf{v} + a_2\mathbf{2}) \in \mathcal{H}\ ,$$

which can be worded as "for all vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ belonging to the vector space $\mathcal{H}$ and for all complex numbers a_1 and a_2, the linear combination $(a_1\mathbf{v}_1 + a_2\mathbf{v}_2)$ still belongs to $\mathcal{H}$". In other words, a vector space is defined by the property that by adding together vectors and by multiplying them by scalars one stays within the vector space.[2] The fact that a_1 and a_2 belong to $\mathbb{C}$ qualifies the vector space $\mathcal{H}$ as "complex". Had a_1 and a_2 belonged to $\mathbb{R}$, the vector space would have then been "real".

A set of n vectors $\{\mathbf{v}_j, j \in [1, \ldots, n]\}$ (where the label j is assumed to take integer values), is said to be "linearly independent" if for *any* set of scalar coefficients $\{a_j, j \in [1, \ldots, n]\}$, one has

$$\sum_{j=1}^{n} a_j\mathbf{v}_j = a_1\mathbf{v}_1 + \ldots + a_n\mathbf{v}_n \neq 0\ .$$

The *maximum* number of linearly independent vectors d is a characteristic property of the vector space, called its dimension.

For instance, physical space is said to have dimension three because given any four vectors in it one may always find real coefficients such that the linear combination of the given vectors equals zero.

A *basis* of a d-dimensional vector space $\mathcal{H}$ is then any set of d linearly independent vectors. Let us denote a basis by $\{\hat{\mathbf{e}}_j, j \in [1, \ldots, d]\}$. Any vector in $\mathcal{H}$ can be represented as a linear combination of basis vectors:

$$\forall \mathbf{v} \in \mathcal{H} \quad : \quad \mathbf{v} = \sum_{j=1}^{d} v_j\hat{\mathbf{e}}_j\ ,$$

where the coefficients $\{a_j\}$ are generally complex for a complex vector space, and real for a real space. *Given a fixed basis*, any vector $\mathbf{v}$ can then be identified with its set of d coefficients $\{v_j, j \in [1, \ldots, d]\}$, also called 'components'.

A scalar product between vectors may also be defined. In the case of real vector spaces, the most commonly adopted scalar product, and the only one relevant in these lectures, is the Euclidean scalar product, denoted by a dot "$\cdot$" (and also occasionally referred to as "dot product"). A convenient choice for a basis, and one which can always be made, is that of a *orthonormal* basis, such that

$$\hat{\mathbf{e}}_j \cdot \hat{\mathbf{e}}_k = \delta_{jk}\ ,$$

[2] Note that our definition is incomplete and not rigorous, in that we should have defined all the properties of the addition and scalar multiplication functions. This was intentional, in order to keep up the pace of the lectures and avoid cluttering them with technicalities. We assume the reader is familiar with the definition and properties of such basic operations.

where the "Kronecker delta" δ_{jk} is defined by

$$\delta_{jk} = \begin{cases} 1 & \text{if} \quad j = k \,, \\ 0 & \text{if} \quad j \neq k \,. \end{cases}$$

An expression for the scalar product $\mathbf{v} \cdot \mathbf{w}$ can then be easily derived as:

$$\mathbf{v} \cdot \mathbf{w} = \left(\sum_{j=1}^{d} v_j \hat{\mathbf{e}}_j \right) \cdot \left(\sum_{j=1}^{d} w_k \hat{\mathbf{e}}_k \right) = \sum_{j,k=1}^{d} v_j w_k \hat{\mathbf{e}}_j \cdot \hat{\mathbf{e}}_k = \sum_{j,k=1}^{d} v_j w_k \delta_{jk} = \sum_{j=1}^{d} v_j w_j \,,$$

where the double summation over j and k simplifies to a single one because $\delta_{jk} = 0$ if $j \neq k$. Also, the introduction of the scalar product allows for a systematic way to determine the components of a vector $\mathbf{v}$ with respect to an orthonormal basis $\{\hat{\mathbf{e}}_j, j \in [1, \ldots, d]\}$:

$$\hat{\mathbf{e}}_j \cdot \mathbf{v} = \hat{\mathbf{e}}_j \cdot \left(\sum_{j=1}^{d} v_k \hat{\mathbf{e}}_k \right) = \sum_{k=1}^{d} v_k \hat{\mathbf{e}}_j \cdot \hat{\mathbf{e}}_k = \sum_{k=1}^{d} v_k \delta_{jk} = v_j \,.$$

In the case of a complex vector space, the Euclidean product is generalised to another scalar product, also termed 'inner product' or 'sesquilinear scalar product',[3] which, for vectors $\mathbf{v}$ and $\mathbf{w}$, we will denote as $(\mathbf{v}, \mathbf{w})$. Orthonormal bases such that

$$(\hat{\mathbf{e}}_j, \hat{\mathbf{e}}_k) = \delta_{jk}$$

can be defined for the inner product as well. The inner product $(\mathbf{v}, \mathbf{w})$ can then be expressed in terms of the components of the vectors with respect to one of such bases as:

$$(\mathbf{v}, \mathbf{w}) = \sum_{j=1}^{d} v_j^* w_j \,.$$

Notice that the components of the first vector $\mathbf{v}$ *are complex conjugated.* Likewise, we can extend the formula to derive the components of $\mathbf{v}$ in an orthonormal basis:

$$(\hat{\mathbf{e}}_j, \mathbf{v}) = \hat{\mathbf{e}}_j \cdot \left(\sum_{j=1}^{d} v_k \hat{\mathbf{e}}_k \right) = \sum_{k=1}^{d} v_k \hat{\mathbf{e}}_j \cdot \hat{\mathbf{e}}_k = \sum_{k=1}^{d} v_k \delta_{jk} = v_j \,.$$

in the following, for convenience, we will also represent the inner product $(\mathbf{v}, \mathbf{w})$ in terms of the conjugate transposed vector $\mathbf{v}^\dagger$ (obtained by turning the column vector $\mathbf{v}$ into a row and by complex conjugating all its entries):

$$\mathbf{v}^\dagger \mathbf{w} = (\mathbf{v}, \mathbf{w}) \,. \tag{2.1}$$

[3] "Sesqui-" is a Latin root relating to "one and a half", like in Al_2O_3, which is also known as Aluminium "sesquioxide". "Sesquilinear" then means "one-and-a-half-linear", because this product is linear with respect to the second input vector but only half–linear with respect to the first one (in that its coefficients get complex conjugated).

2.2 Linear operators (reminder)

We will hence forth always refer to complex vector spaces. Unless otherwise noted, any of these statements hold for real spaces as well by replacing complex coefficients with real ones. A linear transformation M acting on a complex vector space $\mathcal{H}$ is defined by the following property

$$\forall\ \mathbf{v}_1, \mathbf{v}_2 \in \mathcal{H} \quad \text{and} \quad \forall\ a_1, a_2 \in \mathbb{C} \quad : \quad M(a_1\mathbf{v}_1 + a_2\mathbf{v}_2) = a_1 M(\mathbf{v}_1) + a_2 M(\mathbf{v}_2)\ .$$

That is, the transformation of a linear combination of vectors equals the corresponding linear combination of the transformed vectors. Because of this property, and because any vector can be expanded as a linear combination of basis vectors, a linear transformation is completely described by its action on a set of basis vectors $\{\hat{\mathbf{e}}_j, j \in [1, \ldots, d]\}$. Restricting for simplicity to linear transformations between a vector space and itself (technically known as "endomorphisms"), we can specify the action of M on each basis vector $\hat{\mathbf{e}}_k$ as follows:

$$M(\hat{\mathbf{e}}_k) = \sum_{j=1}^{d} M_{jk}\hat{\mathbf{e}}_j$$

(simply because the vector $M(\hat{\mathbf{e}}_k)$, like any other vector, can in turn be expanded in terms of the basis $\{\hat{\mathbf{e}}_j, j \in [1, \ldots, d]\}$). Given a basis, the transformation M is hence completely determined by the $d \times d$ complex numbers M_{jk}, which are usually organised in a square *matrix*:

$$M = \begin{pmatrix} M_{11} & M_{12} & \cdots & M_{1d} \\ M_{21} & M_{22} & & \vdots \\ \vdots & & \ddots & \vdots \\ M_{d1} & \cdots & \cdots & M_{dd} \end{pmatrix}$$

Therefore, $d \times d$ complex matrices 'represent' the whole set of linear transformations from a complex vector space to itself (so that, from now on, the same symbol will refer to both matrices and to the corresponding linear transformations).

Given a linear transformation M from the d-dimensional $\mathcal{H}$ into $\mathcal{H}$, one can define its 'inverse' M^{-1} as the linear transformation such that

$$M^{-1}M = MM^{-1} = \mathbb{1}\ ,$$

where the 'identity' operator $\mathbb{1}$ is simply the linear operator that leaves *any* vector unchanged:

$$\forall\ \mathbf{v} \in \mathcal{H} \quad : \quad \mathbb{1}(\mathbf{v}) = \mathbf{v}\ .$$

The matrix corresponding to $\mathbb{1}$ does not depend on the basis, its entries are always given by δ_{jk} (where $j, k \in [1, \ldots, d]$): they are all ones on the main diagonal and zeros elsewhere.

2.2.1 The determinant (reminder)

The *determinant* is a map between $n \times n$ (*square*) matrices and scalars – complex, if the vector space is complex, or real, if the vector space is real – numbers. The determinant of the matrix M will be denoted by $\det(M)$. Here, we will not dwell on how to evaluate determinants, but we will just remark that the determinant can be entirely defined by the following three properties:

- *The determinant is linear in all the rows and columns.*
- *The determinant is anti-symmetric under exchange of any two rows or columns.*
- *The determinant of the identity matrix* $\mathbb{1}$ *is* 1*, in any dimension:* $\det(\mathbb{1}) = 1$.

No other mapping between matrices and numbers fulfills all of these conditions.

We now recall that, given the matrix entries M_{jk} is a certain basis, a formula for the entries of the inverse matrix $(M^{-1})_{jk}$ can be constructed. We will not remind this formula here, but will just recall that the formula contains a factor $1/\det(M)$. From this fact, we can infer a first piece of information which will be valuable later on:

Proposition 1. *A linear transformation* M *on a finite vector space can be inverted if and only if*

$$\det(M) \neq 0$$

.

We intend now to relate the invertibility of M to another equivalent condition: we want to prove the following statement:

Proposition 2. *A linear transformation* M *on a finite vector space can be inverted if and only if no vector* $\mathbf{v}$ *such that* $|\mathbf{v}| \neq 0$ *exists such that* $M(\mathbf{v}) = 0$ *(where* 0 *stands here for the null vector):*

$$\exists M^{-1} : \forall \mathbf{v} \in \mathcal{H} : M^{-1}M\mathbf{v} = \mathbf{v} \quad \Leftrightarrow \quad M\mathbf{v} \neq 0 \quad \forall \mathbf{v} \in \mathcal{H} : |\mathbf{v}| \neq 0 \, .$$

Proof. The direction ($\Rightarrow$) is easily proven *ad absurdum*: if a non-trivial vector $\mathbf{v}$ such that $M\mathbf{v} = 0$ existed, then $M^{-1}M\mathbf{v} = M^{-1}0 = 0$ ($M0 = 0$ for any linear transformation M), so that the first condition would be violated too. The proof of the direction ($\Leftarrow$) is slightly more technical: consider the transformation of a generic vector $\mathbf{v} = \sum_{j=1}^{d} v_j \hat{\mathbf{e}}_j$, where the $\{\hat{\mathbf{e}}_j\}$ form a basis of $\mathcal{H}$. The second condition implies:

$$M\left(\sum_{j=1}^{d} v_j \hat{\mathbf{e}}_j\right) = \sum_{j=1}^{d} v_j M(\hat{\mathbf{e}}_j) \neq 0 \quad \forall \quad v_j \in \mathbb{C},$$

which is equivalent to stating that the d vectors $M(\hat{\mathbf{e}}_j)$ are also a basis of $\mathcal{H}$ (because a basis is precisely a set of d vectors satisfying the condition above for any choice of the coefficients $\{v_j, j \in [1, \ldots, d]\}$). Now, the linear transformation M^{-1} can always be *constructed* by imposing that

$$M^{-1}(M(\hat{\mathbf{e}}_j)) = \hat{\mathbf{e}}_j \quad \forall \, j \in [1, \ldots, d] \, ,$$

because the previous prescription corresponds simply to specifying how M^{-1} acts on a basis of $\mathcal{H}$, which determines uniquely the transformation M^{-1} and may always be done. Notice that the condition that $\{M(\hat{\mathbf{e}}_j), j \in [1, \ldots, d]\}$ constitutes a basis was crucial in this last step. *We could not have fixed the action of* M^{-1} *on a set of linearly* dependent *vectors*, precisely because this vectors, being mutually dependent, would not have allowed for an arbitrary choice.

Proposition 1 and 2 lead to the following conclusion:

Proposition 3. *Given a linear transformation* M*, a vector* $\mathbf{v}$ *with* $|\mathbf{v}| \neq 0$ *such that* $M\mathbf{v} = 0$ *exists if and only if* $\det(M) = 0$.

2.3 Eigenvalues and eigenvectors of a linear operator

The *eigenvalue* λ and associated *eigenvector* $\mathbf{v}$ of a matrix M are defined as the scalar and vector such that

$$M\mathbf{v} = \lambda\mathbf{v} \,. \tag{2.2}$$

In other words, eigenvectors represent the directions of the vector space along which the action of M yields vectors parallel (anti-parallel for negative eigenvalues) to the original ones but with lengths stretched (or contracted) by factors given by the eigenvalues of M.

Eigenvalues can be determined by applying Proposition 3 from the previous Section. In fact, Eq. (2.2) can be recast as follows:

$$M\mathbf{v} = \lambda\,\mathbb{1}\mathbf{v} \quad \Rightarrow \quad (M - \lambda\,\mathbb{1})\mathbf{v} = 0\,,$$

where we have just inserted the identity operator $\mathbb{1}$ in front of a vector, as can always be done. Now, Proposition 3 tells us that a non-zero vector $\mathbf{v}$ satisfying the previous equation exists if and only if $\det(M - \lambda\mathbb{1}) = 0$ (*note that the null vector is a trivial eigenvector of any linear operator: it will be henceforth always disregarded*). We have thus determined the condition for a scalar λ to be eigenvalue of M: λ must be a root of the so-called "characteristic equation", given by

$$\det(M - \lambda\,\mathbb{1}) = 0 \,. \tag{2.3}$$

Because of the definition of the determinant, the characteristic equation of a $d \times d$ matrix M is an algebraic equation of order d, and has hence d roots, each equal to an eigenvalue λ_j of M, in turn corresponding to an eigenvalue $\mathbf{v}_j$. As we will see in the following, more than one linearly independent eigenvector may correspond to the same eigenvalue λ_j. The polynomial $\det(M - \lambda\,\mathbb{1})$, of order d in λ, is referred to as the *characteristic polynomial* of M.

Notice also that multiplying an eigenvector by a given scalar yields another eigenvector associated to the same eigenvalue. This can be easily proven: assume that $\mathbf{v}_j$ is an eigenvector of M with eigenvalue λ_j (so that $M\mathbf{v}_j = \lambda_j\mathbf{v}_j$), and consider the action of M on the vector $c\mathbf{v_j}$, where $c \in \mathbb{C}$ is any scalar:

$$M(c\mathbf{v}_j) = cM\mathbf{v}_j = c\lambda_j\mathbf{v}_j = \lambda_j(c\mathbf{v}_j)\,,$$

which proves that $c\mathbf{v}_j$ is also an eigenvector with eigenvalue λ_j. This means that, as already mentioned above, eigenvectors identify *directions*, rather than individual vectors, along which the action of the linear operator corresponds to a dilation. Hence, eigenvectors can only be determined up to an arbitrary multiplicative factor. Usually it is convenient to determine *normalised* eigenvectors, *i.e.* eigenvectors of modulus 1.[4]

Let us now specify a systematic, pragmatic recipe to determine eigenvalues and eigenvectors of a linear operator:

1. Write down the characteristic equation (2.3) and solve it to find the eigenvalues λ_j.

[4]When dealing with complex vector spaces, normalised eigenvectors may be multiplied by arbitrary 'phase factors' $e^{i\varphi}$ for any $\varphi \in [0, 2\pi[$ and still remain normalised eigenvectors with the same associated eigenvalue. Such an ambiguity may not be avoided: any of those eigenvectors is a valid one.

2. For each eigenvalue λ_j, write the unknown eigenvector $\tilde{\mathbf{v}}_j$ as $\tilde{\mathbf{v}}_j = (a, b, c, \ldots)^\mathsf{T}$, where $\{a, b, c, \ldots\}$ are to be determined by solving the 'eigensystem':

$$M\tilde{\mathbf{v}}_j = \lambda_j \tilde{\mathbf{v}}_j \,. \tag{2.4}$$

This is a linear system of d equation for d unknown variables $\{a, b, c, \ldots\}$.

3. Since, as seen above, an eigensystem only determines vectors up to multiplicative factors, at least one of the d linear equation of the eigensystem (2.4) will be redundant (*i.e.*, it will be implied by the other $(d-1)$ equations).

4. Hence, one of the variables $\{a, b, c, \ldots\}$ has to be set arbitrarily: to do this, pick one variable, say a, and make sure that the eigensystem does not imply $a = 0$ (one can always find a variable like that, otherwise the eigenvector would be zero). Then, simply set $a = 1$ and solve the eigensystem.

5. Once the non-normalised eigenvector $\tilde{\mathbf{v}}_j$ has been determined, determine the normalised one $\mathbf{v}_j$ by the equation $\mathbf{v}_j = \tilde{\mathbf{v}}_j / \sqrt{(\tilde{\mathbf{v}}_j, \tilde{\mathbf{v}}_j)}$.

Let us now see a concrete example. Consider the 2×2 matrix A:

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

The characteristic equation $\det(A - \lambda \mathbb{1}) = 0$ reads (point **1**)

$$\det \begin{pmatrix} -\lambda & 1 \\ 1 & -\lambda \end{pmatrix} = \lambda^2 - 1 = (\lambda + 1)(\lambda - 1) = 0 \,.$$

The two solutions of the equation are hence $\lambda_1 = -1$ and $\lambda_2 = 1$. These are the two eigenvalues of A. To find the eigenvector $\tilde{\mathbf{v}}_1$ corresponding to λ_1, one has to set

$$\tilde{\mathbf{v}}_1 = \begin{pmatrix} a \\ b \end{pmatrix}$$

and solve the eigensystem $A\mathbf{v}_1 = \lambda_1 \tilde{\mathbf{v}}_1$ for the variables a and b (point **2**):

$$A\mathbf{v}_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} b \\ a \end{pmatrix} = \lambda_1 \tilde{\mathbf{v}}_1 = -\mathbf{v}_1 = \begin{pmatrix} -a \\ -b \end{pmatrix} .$$

Being a vector equation, the previous equality implies two scalar equations: $b = -a$ and $a = -b$. These two equations are identical, as we should have expected from point **3** above (one equation in the eigensystem is *always* redundant). The only relevant equation is then

$$b = -a \,.$$

This equation does not imply $a = 0$. Hence, according to point **4**, we can set $a = 1$ and have

$$\tilde{\mathbf{v}}_1 = \begin{pmatrix} 1 \\ -1 \end{pmatrix} .$$

Finally, the vector $\tilde{\mathbf{v}}_1$ is normalised as per point **5**:

$$\mathbf{v}_j = \tilde{\mathbf{v}}_j / \sqrt{(\tilde{\mathbf{v}}_j, \tilde{\mathbf{v}}_j)} = \begin{pmatrix} 1 \\ -1 \end{pmatrix} / \sqrt{1^2 + (-1)^2} = \begin{pmatrix} 1 \\ -1 \end{pmatrix} / \sqrt{2} \,.$$

The eigenvector $\mathbf{v}_2$ associated to $\lambda_2 = +1$ can be determined as

$$\mathbf{v}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} / \sqrt{2} \, .$$

The eigenvalues and eigenvectors of real matrices are not necessarily real: consider for example the matrix Ω:

$$\Omega = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} .$$

The characteristic equation $\det(\lambda \mathbb{1} - \Omega) = 0$ reads (point **1**)

$$\det \begin{pmatrix} \lambda & -1 \\ 1 & \lambda \end{pmatrix} = \lambda^2 + 1 = (\lambda - i)(\lambda + i) = 0 \, .$$

The two solutions of the equation are hence $\lambda_1 = -i$ and $\lambda_2 = i$. These are the two eigenvalues of Ω. To find the eigenvector $\tilde{\mathbf{v}}_1$ corresponding to λ_1, one has to set

$$\tilde{\mathbf{v}}_1 = \begin{pmatrix} a \\ b \end{pmatrix}$$

and solve the eigensystem $\Omega \mathbf{v}_1 = \lambda_1 \tilde{\mathbf{v}}_1$ for the variables a and b (point **2**):

$$\Omega \mathbf{v}_1 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} b \\ -a \end{pmatrix} = \lambda_1 \tilde{\mathbf{v}}_1 = -i \mathbf{v}_1 = \begin{pmatrix} -ia \\ -ib \end{pmatrix} .$$

As expected from from point **3**, the vector equation above corresponds to a single independent scalar equation (apparent if one multiplies one of the two rows by i). Setting $a = 1$, as per point **4**, determines $b = -i$, such that

$$\tilde{\mathbf{v}}_1 = \begin{pmatrix} 1 \\ -i \end{pmatrix} .$$

The vector $\tilde{\mathbf{v}}_1$ is normalised as per point **5**, recalling the rules for the inner product of complex vectors:

$$\mathbf{v}_1 = \tilde{\mathbf{v}}_1 / \sqrt{(\tilde{\mathbf{v}}_1, \tilde{\mathbf{v}}_1)} = \begin{pmatrix} 1 \\ -i \end{pmatrix} / \sqrt{1^2 + i(-i)} = \begin{pmatrix} 1 \\ -i \end{pmatrix} / \sqrt{2} \, .$$

It can be shown following the same steps that the normalised eigenvector $\mathbf{v}_2$ corresponding to $\lambda_2 = +i$ is given by

$$\mathbf{v}_2 = \begin{pmatrix} -i \\ 1 \end{pmatrix} / \sqrt{2} \, .$$

2.3.1 Degenerate eigenvalues

As already pointed out, the characteristic equation (2.3), which determines the eigenvalues of the linear operator M, is an algebraic equation of order d. If two or more solutions of the equation coincide, the characteristic polynomial is said to have *degenerate* roots. In that case, clearly, there will be less than d eigenvalues.

However, the eigenvalue corresponding to a degenerate root may have a number of corresponding eigenvectors up to its multiplicity (this is not necessarily the case!).

Such eigenvectors are determined by the eigensystem, like any eigenvector, but for them the eigensystem might feature more than one redundant equation, so that one may have to arbitrarily set more than one entry of the eigenvector.

This situation is better illustrated with a concrete example. Consider the matrix B:

$$B = \begin{pmatrix} 5 & 1 & 2 \\ 1 & 5 & -2 \\ 2 & -2 & 2 \end{pmatrix} .$$

The characteristic equation $\det(\lambda \mathbb{1} - B) = 0$ reads

$$\det \begin{pmatrix} \lambda - 5 & -1 & -2 \\ -1 & \lambda - 5 & 2 \\ -2 & 2 & \lambda - 2 \end{pmatrix} = \lambda^3 - 12\lambda^2 + 36\lambda = \lambda(\lambda - 6)^2 = 0 ,$$

with solutions 0 and 6. These are the only two eigenvalues of the matrix B. However, note that the eigenvalue 6 is a root of the characteristic equation with multiplicity 2 (in that the factor $(\lambda - 6)$ to the power 2 occurs in the characteristic polynomial). As anticipated above, the eigenvalue 6 may have up to 2 corresponding linearly independent eigenvectors, in which case it is said to be *degenerate* ('doubly' degenerate, in in this instance).

The eigenvector $\tilde{\mathbf{v}}_1$ corresponding to $\lambda_1 = 0$ is found by setting

$$\tilde{\mathbf{v}}_1 = \begin{pmatrix} a \\ b \\ c \end{pmatrix}$$

and solving the eigensystem $B\mathbf{v}_1 = \lambda_1 \tilde{\mathbf{v}}_1$ for the variables a, b and c:

$$B\mathbf{v}_1 = \begin{pmatrix} 5 & 1 & 2 \\ 1 & 5 & -2 \\ 2 & -2 & 2 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 5a + b + 2c \\ a + 5b - 2c \\ 2a - 2b + 2c \end{pmatrix} = \lambda_1 \tilde{\mathbf{v}}_1 = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} .$$

Adding up the first and second lines of the previous scalar equation yields

$$a + b = 0 ,$$

such that the third line becomes

$$4a + 2c = 0 .$$

We can then set $c = 1$ to get $a = -1/2$ and $b = 1/2$. Finally, the eigenvector $\tilde{\mathbf{v}}_1 = (-1/2, 1/2, 1)^{\mathsf{T}}$ can be normalised to obtain

$$\mathbf{v}_1 = \begin{pmatrix} -1 \\ 1 \\ 2 \end{pmatrix} / \sqrt{6} .$$

Let us now move on to the degenerate eigenvalue $\lambda_2 = 6$. Setting

$$\tilde{\mathbf{v}}_2 = \begin{pmatrix} a \\ b \\ c \end{pmatrix} ,$$

the eigensystem $B\mathbf{v}_2 = \lambda_2\tilde{\mathbf{v}}_2$ for the variables a, b and c reads:

$$B\mathbf{v}_1 = \begin{pmatrix} 5 & 1 & 2 \\ 1 & 5 & -2 \\ 2 & -2 & 2 \end{pmatrix}\begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 5a+b+2c \\ a+5b-2c \\ 2a-2b+2c \end{pmatrix} = \lambda_2\tilde{\mathbf{v}}_2 = \begin{pmatrix} 6a \\ 6b \\ 6c \end{pmatrix}.$$

The three equations from the vector equality above are all identical and equivalent to

$$a - b - 2c = 0\,. \tag{2.5}$$

All vectors satisfying such a relationship are eigenvectors associated to $\lambda_2 = 6$. Clearly, however, we are left with only one condition to determine the three variables a, b and c. Two of them will have hence to be *arbitrarily*. By setting $a = 1$ and $c = 1$, one gets

$$b = -1\,,$$

such that the first eigenvector corresponding to λ_2 is determined as

$$\tilde{\mathbf{v}}_2 = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix},$$

normalised as

$$\mathbf{v}_2 = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}/\sqrt{3}\,.$$

Setting $a = 1$ and $c = 0$ instead gives

$$\tilde{\mathbf{v}}_3 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix},$$

normalised as

$$\mathbf{v}_3 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}/\sqrt{2}\,.$$

The eigenvectors $\mathbf{v}_2$ and $\mathbf{v}_3$ are both associated to the same degenerate eigenvalue λ_2. Notice that the expressions of $\mathbf{v}_2$ and $\mathbf{v}_3$ depend on the choice we made for the coefficients when solving the einsystem for λ_2. Different choices could have been made, leading to different eigenvectors $\mathbf{v}_2$ and $\mathbf{v}_3$. The choice we made was such as to make $\mathbf{v}_2$ and $\mathbf{v}_3$ orthogonal to each other: this can always be made.

In fact, $\mathbf{v}_3$ could have been determined by imposing Eq. (2.5) together with the condition of being orthogonal to $\mathbf{v}_3$. The latter condition would read

$$0 = \mathbf{v}_2 \cdot \begin{pmatrix} a \\ b \\ c \end{pmatrix} = (a - b + c)/\sqrt{3}\,. \tag{2.6}$$

Eqs. (2.5) and (2.6) imply

$$\begin{aligned} a &= b\,, \\ c &= 0\,, \end{aligned}$$

which indeed correspond, up to normalisation, to the vector $\mathbf{v}_3$ above.

2.3.2 Diagonalisable matrices

A square matrix is said to be *diagonalisable* if and only if the eigenvectors of the matrix form a basis of the vector space on which it is defined. In practice, this condition can be directly tested: for a $d \times d$ matrix, if one can find d linearly independent eigenvalues then the matrix is diagonalisable.

As we are about to see, diagonalisability is equivalent to the existence of a basis of the vector space where the matrix can be expressed in a very simple, 'diagonal' form, with non-zero entries only on the main diagonal. This sort of simplification is extremely useful in a number of applications. In quantum mechanics, for instance, the dynamics of a system becomes extremely easy to handle (and is in fact considered to be 'solved') if one can diagonalise its Hamiltonian operator (which, roughly speaking, plays the role of the energy). In classical mechanics, when dealing with systems that can rotate along several axes, the diagonalisation of the inertia tensor provides one with knowledge about the 'principal axes' of rotation, and with the moments of inertia along those axes.

As illustrated by the following counterexample, *not all square matrices can be diagonalised.* Consider in fact the matrix J:

$$J = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} .$$

The characteristic equation for J reads $\lambda^2 = 0$, such that $\lambda_1 = 0$ is the only (potentially doubly degenerate) eigenvalue of J. In terms of the generic vector $\tilde{\mathbf{v}}_1 = (a, b)^\mathsf{T}$, the eigensystem for $\lambda_1 = 0$ is

$$J\tilde{\mathbf{v}}_1 = \begin{pmatrix} b \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} .$$

The only condition that can be extracted from the eigensystem is hence $b = 0$. The vector $\mathbf{v}_1 = (1, 0)^\mathsf{T}$ is in fact an eigenvector of J. However, no other linearly independent eigenvector exists. In particular, the vector $\mathbf{v}_2 = (0, 1)^\mathsf{T}$, orthogonal to $\mathbf{v}_1$ and which would form a basis with the latter, is not an eigenvector of J as can be directly verified:

$$J\mathbf{v}_2 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \neq \lambda_2 \mathbf{v}_2 .$$

A matrix like J is referred to as a *Jordan block*. Jordan blocks are never diagonalisable.

Let us now assume that a $d \times d$ matrix M is diagonalisable, with a basis of eigenvectors $\{\mathbf{v}_j, j \in [1, \ldots, d]\}$. Let λ_j be the eigenvalue associated to each $\mathbf{v}_j$: $M\mathbf{v}_j = \lambda_j \mathbf{v}_j$.[5] Let L be defined as the $d \times d$ matrix with columns equal to the d vectors $\mathbf{v}_j$'s. We shall adopt the following notation:

$$L = \left(\mathbf{v}_1, \ldots \mathbf{v}_j, \ldots \mathbf{v}_d\right) , \tag{2.7}$$

where each $\mathbf{v}_j$ represents the column of entries of the vector $\mathbf{v}_j$ in a generic basis. Since, per hypothesis, the $\mathbf{v}_j$'s are linearly independent, the matrix L must be invertible: a

[5] Note that we are not assuming anything about the degeneracy or multiplicity of each eigenvalue: it may well be that $\lambda_j = \lambda_k$ for some j and k. Also, we are *not* assuming here that the eigenvectors $\mathbf{v}_j$ are orthonormal.

matrix L^{-1} exists such that $L^{-1}L = \mathbb{1}$. Let us represent L^{-1} as a matrix of rows:

$$L^{-1} = \begin{pmatrix} \mathbf{w}_1^\dagger \\ \vdots \\ \mathbf{w}_k^\dagger \\ \vdots \\ \mathbf{w}_d^\dagger \end{pmatrix} . \tag{2.8}$$

Then $L^{-1}L = \mathbb{1}$ can be expressed as

$$\begin{pmatrix} \mathbf{w}_1^\dagger \\ \vdots \\ \mathbf{w}_k^\dagger \\ \vdots \\ \mathbf{w}_d^\dagger \end{pmatrix} \Big(\mathbf{v}_1, \dots \mathbf{v}_j, \dots \mathbf{v}_d \Big) = \begin{pmatrix} \mathbf{w}_1^\dagger \mathbf{v}_1 & \cdots & \mathbf{w}_1^\dagger \mathbf{v}_j & \cdots & \mathbf{w}_1^\dagger \mathbf{v}_d \\ \vdots & \ddots & & \iddots & \vdots \\ \mathbf{w}_k^\dagger \mathbf{v}_1 & \cdots & \mathbf{w}_k^\dagger \mathbf{v}_j & \cdots & \mathbf{w}_k^\dagger \mathbf{v}_d \\ \vdots & \iddots & & \ddots & \vdots \\ \mathbf{w}_d^\dagger \mathbf{v}_1 & \cdots & \mathbf{w}_d^\dagger \mathbf{v}_j & \cdots & \mathbf{w}_d^\dagger \mathbf{v}_d \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & \ddots & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} ,$$

where the notation for the inner product (2.1) has been used. The previous matrix equation can be represented as

$$\mathbf{w}_k^\dagger \mathbf{v}_j = \delta_{jk} , \tag{2.9}$$

where the 'Kronecker delta' δ_{jk} is defined by: $\delta_{jk} = 1$ if $j = k$ and $\delta_{jk} = 0$ if $j \neq 0$.

Now, let us apply the ordinary rules of matrix multiplication to evaluate the matrix $D = L^{-1}ML$, which represents the transformation M in the new basis of eigenvectors ('eigenbasis'):[6]

$$\begin{aligned} D &= L^{-1}M\Big(\mathbf{v}_1, \dots \mathbf{v}_j, \dots \mathbf{v}_d\Big) = L^{-1}\Big(M\mathbf{v}_1, \dots M\mathbf{v}_j, \dots M\mathbf{v}_d\Big) \\ &= L^{-1}\Big(\lambda_1\mathbf{v}_1, \dots \lambda_j\mathbf{v}_j, \dots \lambda_d\mathbf{v}_d\Big) = \begin{pmatrix} \mathbf{w}_1^\dagger \\ \vdots \\ \mathbf{w}_k^\dagger \\ \vdots \\ \mathbf{w}_d^\dagger \end{pmatrix} \Big(\lambda_1\mathbf{v}_1, \dots \lambda_j\mathbf{v}_j, \dots \lambda_d\mathbf{v}_d\Big) \\ &= \begin{pmatrix} \lambda_1\mathbf{w}_1^\dagger \mathbf{v}_1 & \cdots & \lambda_j\mathbf{w}_1^\dagger \mathbf{v}_j & \cdots & \lambda_d\mathbf{w}_1^\dagger \mathbf{v}_d \\ \vdots & \ddots & & \iddots & \vdots \\ \lambda_1\mathbf{w}_k^\dagger \mathbf{v}_1 & \cdots & \lambda_j\mathbf{w}_k^\dagger \mathbf{v}_j & \cdots & \lambda_d\mathbf{w}_k^\dagger \mathbf{v}_d \\ \vdots & \iddots & & \ddots & \vdots \\ \lambda_1\mathbf{w}_d^\dagger \mathbf{v}_1 & \cdots & \lambda_j\mathbf{w}_d^\dagger \mathbf{v}_j & \cdots & \lambda_d\mathbf{w}_d^\dagger \mathbf{v}_d \end{pmatrix} = \begin{pmatrix} \lambda_1 & 0 & 0 & 0 & 0 \\ 0 & \ddots & 0 & 0 & 0 \\ 0 & 0 & \lambda_j & 0 & 0 \\ 0 & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & \lambda_d \end{pmatrix} . \end{aligned}$$

As anticipated above, the expression of the diagonalisable linear transformation M in the basis given by its eigenvectors *is a diagonal matrix D, with the eigenvalues on the main diagonal and zero entries everywhere else*.

[6] Let us recall that, in general, given an invertible matrix L and a linear operator represented by the matrix M, the transformation $L^{-1}ML$ (known as 'similarity'), represents the linear operator in the new basis given by the columns of L. The fact that the columns of L form a basis, *i.e.* that they are linearly independent vectors, is equivalent to state that the square matrix L is invertible.

In practice, the matrix L which *diagonalises* the matrix M is just constructed as the matrix whose column are the eigenvalues of M, determined as detailed in section 2.3. For instance, for the matrix A of our previous example, the matrix L is given by

$$L = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} / \sqrt{2} \, .$$

The inverse L^{-1} is given by

$$L = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} / \sqrt{2} \, ,$$

and the diagonalisation can be checked directly

$$L^{-1}AL = \frac{1}{2}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} .$$

2.3.3 Invariants and eigenvalues

It is clear from the expression determined above

$$D = L^{-1}ML \tag{2.10}$$

(where D is the diagonal matrix containing the eigenvalues of M on the main diagonal), that the eigenvalues are invariant under a change of basis of the vector space. In fact, consider a change of basis described by the invertible matrix N, such that $M \to M' = N^{-1}MN$. One has $M = NM'N^{-1}$, which can be inserted into Eq. (2.10) to obtain

$$D = L^{-1}ML = L^{-1}NM'N^{-1}L = L'^{-1}M'L' \quad \text{with} \quad L' = N^{-1}L \, .$$

So, while the matrix of eigenvectors changes from L into $N^{-1}L$, the diagonal matrix of eigenvalues D stays the same under the change of basis.

The eigenvalues are related to other invariant quantities, which are generally much easier to evaluate. We will only mention two of them, arguably the most useful:

- The determinant $\det(M)$, whose invariance is easily proved.[7] The determinant equals the product of all the eigenvalues of a matrix: $\det(M) = \prod_1^d \lambda_j$.
- The 'trace' tr(M), defined as the sum of the elements on the main diagonal. The trace equals the sum of all the eigenvalues of a matrix: $\mathrm{tr}(M) = \sum_{j=1}^d \lambda_j$.

For instance, in the case of matrix A of our previous example, one has $\det(A) = -1$ and $\mathrm{tr}(A) = 0$, which are indeed, respectively, the product and the sum of the two eigenvalues -1 and $+1$.

2.3.4 Functions of operators (optional)

Let M be an operator that can be diagonalised, *i.e.* such that

$$M = LDL^{-1} \, ,$$

[7] From Binet's formula for the determinant of a product $\det(AB) = \det(A)\det(B)$, one has $\det(L^{-1}ML) = \det(L^{-1})\det(M)\det(L) = \frac{1}{\det(L)}\det(L)\det(M) = \det(M)$

where D is a diagonal matrix (with the eigenvalues $\{\lambda_j\}$ of M on the main diagonal and zeros elsewhere) and L is an invertible matrix (whose columns represent the eigenvectors of M).

Given the function of one variable $f(x)$, the matrix representation of the operator $f(M)$ is evaluated as:

$$f(M) = L^{-1} f(D) L \ ,$$

where $f(D)$ is a diagonal matrix with diagonal entries equal to $\{f(\lambda_j)\}$.

For $f(x) = x^{-1}$, this technique allows one to determine the inverse of the diagonalisable matrix M as follows:

$$D^{-1} = \begin{pmatrix} \frac{1}{\lambda_1} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{\lambda_d} \end{pmatrix}$$

and

$$M^{-1} = L D^{-1} L^{-1} \ .$$

Clearly, a matrix with any eigenvalue equal to zero has determinant zero (see Sec. 2.3.3) and is hence not invertible.

2.4 Special Matrices

In the previous section, we saw how to diagonalise (*i.e.*, how to find eigenvalues and eigenvectors and how to change the basis to obtain a diagonal matrix from the original matrix) a generic matrix. We will now describe special classes of matrices of particular mathematical and applicative relevance, and address their diagonalisability and the properties of their eigenvalues and eigenvectors.

Before proceeding, let us recall and introduce a couple of definitions:

- Given a matrix M, the *transpose* of M is denoted by M^{T} and is obtained by writing its rows as the column of M. In terms of matrix entries one has: $M^{\mathsf{T}}_{jk} = M_{kj}$ (swap rows and columns).
- Given a matrix M, the *hermitian conjugate* of M is denoted by $M^\dagger$ and is obtained by complex conjugation of the transpose M^{T}: $M^\dagger = M^{\mathsf{T}*}$. In terms of matrix entries one has: $M^\dagger_{jk} = M^*_{kj}$ (swap rows and columns and complex conjugate).

We can now list the classes of special matrices that we will deal with:

- A square matrix M is said to be *normal* if and only if $M^\dagger M = M M^\dagger$.
- A square matrix H is said to be *hermitian* if and only if $H^\dagger = H$.
- A square matrix S is said to be *real and symmetric* if and only if its entries are real and $S^{\mathsf{T}} = S$.
- A square matrix U is said to be *unitary* if and only if $U^\dagger U = U U^\dagger = \mathbb{1}$.
- A square matrix O is said to be *orthogonal* if and only if it has real entries and $O^{\mathsf{T}} O = O O^{\mathsf{T}} = \mathbb{1}$.

Let us first clarify the dependencies between the sets of matrices defined above:

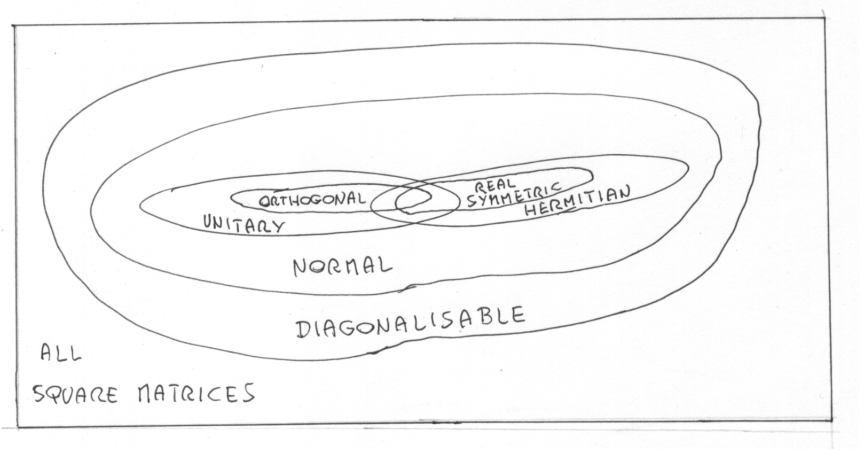

Figure 2.1: Set relationships between classes of special matrices.

- Since for real matrices hermitian conjugation is the same as transposition, *orthogonal matrices are unitary*: $O^\dagger O = O^\mathsf{T} O = \mathbb{1}$.
- Since for real matrices hermitian conjugation is the same as transposition, *real symmetric matrices are hermitian*: $S^\dagger = S^\mathsf{T} = S$.
- *Unitary matrices are normal*: $U^\dagger U = \mathbb{1} = UU^\dagger$.
- *Hermitian matrices are normal*: $H^\dagger H = HH = HH^\dagger$.

It can also be proven (a result known as *spectral theorem*) that *normal matrices are diagonalisable*. This is, in a sense, the broader sufficient condition known for diagonalisability. Notice that this condition is only sufficient and *not* necessary: there exist diagonalisable matrices which are not normal. In general, given a non-normal matrix, the only way to know whether it is diagonalisable or not is by trying to find its eigenvectors and verifying whether they form a basis of the vector space or not. Notice moreover that, since unitary, orthogonal, hermitian and real symmetric matrices are all normal, these are other (narrower) sufficient conditions for diagonalisability. The relations between these classes of matrices are depicted in Fig. 2.1.

2.4.1 Eigenvalues and eigenvectors of hermitian matrices

We just saw that any hermitian matrix $H = H^\dagger$ is diagonalisable. As we shall see, something more can be said about its eigenvalues and eigenvectors.

Because complex numbers are involved in the definition of hermiticity, we will need a bit of notation derived from the expression for the inner product (2.1). In general, given the vectors $\mathbf{v}$ and $\mathbf{w}$ and the square matrix M one has, because of the definition of hermitian conjugate $M^\dagger$:[8]

$$\mathbf{v}^\dagger M \mathbf{w} = \left(\mathbf{w}^\dagger M^\dagger \mathbf{v}\right)^* \ . \tag{2.11}$$

[8]Notice that the notation $\mathbf{v}^\dagger M \mathbf{w}$ represents a scalar complex number, given by the action of M on $\mathbf{w}$ and by the subsequent inner product with $\mathbf{v}$. The same number could be represented as $(\mathbf{v}, M\mathbf{w})$, according to the notation for the inner product introduced in section 2.1.

Let us then denote by λ_j and λ_k two generic eigenvalues of the hermitian matrix H, with corresponding eigenvectors $\mathbf{v}_j$ and $\mathbf{v}_k$:

$$H\mathbf{v}_j = \lambda_j \mathbf{v}_j \,, \tag{2.12}$$
$$H\mathbf{v}_k = \lambda_k \mathbf{v}_k \,. \tag{2.13}$$

Let multiply Eq. (2.12) on the left by $\mathbf{v}_k^\dagger$ to obtain

$$\mathbf{v}_k^\dagger H \mathbf{v}_j = \lambda_j \mathbf{v}_k^\dagger \mathbf{v}_j \,. \tag{2.14}$$

The hermitian conjugate of Eq. (2.13) reads

$$\mathbf{v}_k^\dagger H^\dagger = \mathbf{v}_k^\dagger H = \lambda_k^* \mathbf{v}_k^\dagger$$

(where we made use of the property $H = H^\dagger$), and can be multiplied on the right by $\mathbf{v}_j$ to get

$$\mathbf{v}_k^\dagger H \mathbf{v}_j = \lambda_k^* \mathbf{v}_k^\dagger \mathbf{v}_j \,. \tag{2.15}$$

The left hand sides of Eqs. (2.14) and (2.15) are the same, so that we can equate the right hand sides to obtain

$$(\lambda_j - \lambda_k^*)\mathbf{v}_k^\dagger \mathbf{v}_j = 0 \,. \tag{2.16}$$

Eq. (2.16) has two consequences:

- If $j = k$, one has $\mathbf{v}_k^\dagger \mathbf{v}_j = \mathbf{v}_j^\dagger \mathbf{v}_j = |\mathbf{v}_j|^2 \neq 0$ (because null vectors, with all zero entries, are trivial eigenvector of any matrix and are excluded by hypothesis when considering eigenvectors of linear operators), such that one is left with $\lambda_j = \lambda_j^*$: *the eigenvalues of hermitian matrices are always real.*

- If $j \neq k$ and $\lambda_j \neq \lambda_k$, then $\mathbf{v}_k^\dagger \mathbf{v}_j = 0$: *the eigenvector of hermitian operators associated to different eigenvalues are orthogonal.*

We saw in section 2.3.1 that, if an eigenvalue is degenerate, orthogonal eigenvectors may always be selected to form an *orthogonal* basis of the 'eigenspace' (the subspace of the vector space made up of eigenvectors associated to the same degenerate eigenvalue). Therefore, since as we saw above eigenvectors associated to district eigenvalues are orthogonal to each other, we find that *the eigenvectors of a hermitian operator may always be chosen to form an orthonormal basis.*

This important result may be rephrased in terms of the transformation that diagonalises H, which we will call U: $U^{-1}HU = D$. As we saw before, the matrix U is the matrix with columns equal to the eigenvectors of H: if such eigenvalues are orthonormal, the matrix U is given by

$$U = \left(\mathbf{v}_1, \dots \mathbf{v}_j, \dots \mathbf{v}_d\right) \,, \quad \text{with} \quad \mathbf{v}_j^\dagger \mathbf{v}_k = \delta_{jk} \,,$$

and the inverse L^{-1} can be immediately constructed:

$$U^{-1} = \begin{pmatrix} \mathbf{v}_1^\dagger \\ \vdots \\ \mathbf{v}_j^\dagger \\ \vdots \\ \mathbf{v}_d^\dagger \end{pmatrix} = U^\dagger \,.$$

The matrix U is hence unitary: *any hermitian matrix is diagonalisable and can be diagonalised by a unitary transformation.*

Also, as we determined above, the eigenvalues of hermitian matrices are real: during the early steps of quantum mechanics, this fact brought to identify hermitian operators with the simplest cases of 'observables', whose eigenvalues (collectively known as 'spectra') represent the possible outcomes of a given quantum measurement.

Real symmetric matrices being hermitian, their eigenvalues are bound to be real too. Moreover, along the same lines described here, it can be shown that real symmetric matrices can always be diagonalised by orthogonal (that is, essentially, real and unitary) transformations. Orthogonal transformations represent generalised rotations (reducing to the common spatial rotations in dimension 2 and 3): real symmetric matrices can hence always be diagonalised by rotating the basis of the vector space.

2.5 Real quadratic forms

Matrices do not only represent linear transformations, but also quadratic combinations of sets of variables. Let $\{x_1, \dots, x_d\}$ be a set of d variables, assumed for simplicity to be real, and let $\sum_{j,k=1}^{d} Q_{jk} x_j x_k$ be any generic quadratic combination of them, with real coefficients Q_{jk}. Notice that, since $x_j x_k = x_k x_j$ (scalar multiplication is commutative), the quantities Q_{jk} can always be chosen to be symmetric, such that $Q_{jk} = Q_{kj}$, without loss of generality for the possibility of representing any quadratic expression.

Let us then define the column vector $\mathbf{x} = (x_1, \dots, x_d)^\mathsf{T}$ and the $d \times d$ matrix Q with entries Q_{jk}, to write

$$\sum_{j,k=1}^{d} Q_{jk} x_j x_k = Q(\mathbf{x}) = \mathbf{x}^\mathsf{T} Q \mathbf{x} \,. \tag{2.17}$$

Real symmetric matrices hence also represent all the possible *quadratic forms* $Q(\mathbf{x}) \in \mathbb{R}$ acting on real vector spaces, which are fed a vector and output a real number. The term 'quadratic', refers to the additional property:

$$Q(a\mathbf{x}) = a^2 Q(\mathbf{x}) \quad \forall\, a \in \mathbb{R} \,,$$

which is immediately apparent in matrix form:

$$Q(a\mathbf{x}) = (a\mathbf{x}^\mathsf{T}) Q(a\mathbf{x}) = a^2 \mathbf{x}^\mathsf{T} Q \mathbf{x} = a^2 Q(\mathbf{x}) \,.$$

The applications of quadratic forms are beyond count: for instance, the potential energy of a set of coupled harmonic oscillators (springs) can be represented by a real quadratic form. Notice that because Q is symmetric, it may always be diagonalised, which is often handy. In the next section, we will solve the dynamics of a system of coupled oscillators: although we will not make explicit use of the potential energy in our solution (simply because we will tackle the equations of motion directly), we will solve the problem by a diagonalisation which is completely analogous to diagonalising the quadratic form for the potential energy of the coupled springs.

2.6 Normal modes of oscillation

To appreciate the usefulness of the algebraic techniques we learned, it is instructive to consider a relevant example where such techniques are applied to solve the dynamics of a classical (non-quantum) system.

Consider three particles of equal mass m joined by springs with elastic constant k and rest length l, and constrained to move on a line (Fig. 2.2). By Newton's and Hook's laws, the equations of motion governing the positions x_1, x_2 and x_3 of the three particles are:

$$m\ddot{x}_1 = k(x_2 - x_1 - l)\,, \tag{2.18}$$
$$m\ddot{x}_2 = -k(x_2 - x_1 - l) + k(x_3 - x_2 - l)\,, \tag{2.19}$$
$$m\ddot{x}_3 = -k(x_3 - x_2 - l)\,. \tag{2.20}$$

These second-order differential equations for the functions $x_1(t)$, $x_2(t)$ and $x_3(t)$ are *coupled*: each solution will hence depend on the other two and is in general not easy to find if the equations are tackled directly in the form given above. However, we will see that the diagonalisation of a matrix will allow us to write down a set of equations equivalent to the above, but much simpler to solve.

Notice that, by defining the vector of positions $\mathbf{x} = (x_1, x_2, x_3)^\mathsf{T}$ and the vector of derivatives $\ddot{\mathbf{x}} = (\ddot{x}_1, \ddot{x}_2, \ddot{x}_3)^\mathsf{T}$, the Eqs. (2.18-2.20) can be recast in vector notation as follows:

$$\ddot{\mathbf{x}} = \frac{k}{m}A\mathbf{x} + \frac{k}{m}\mathbf{x}_0\,, \tag{2.21}$$

where

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

and

$$\mathbf{x}_0 = \begin{pmatrix} -l \\ 0 \\ l \end{pmatrix}.$$

The matrix A is real and symmetric, and can hence be diagonalised. We leave the diagonalisation procedure to the reader, and just give the result in terms of the eigenvalues λ_1, λ_2 and λ_3 and associated normalised eigenvectors $\mathbf{v}_1$, $\mathbf{v}_2$ and $\mathbf{v}_3$:

$$\lambda_1 = 0\,, \quad \lambda_2 = -1\,, \quad \lambda_3 = -3\,,$$

$$\mathbf{v}_1 = \begin{pmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ 1/\sqrt{3} \end{pmatrix}, \quad \mathbf{v}_1 = \begin{pmatrix} 1/\sqrt{2} \\ 0 \\ -1/\sqrt{2} \end{pmatrix}, \quad \mathbf{v}_1 = \begin{pmatrix} 1/\sqrt{6} \\ -2/\sqrt{6} \\ 1/\sqrt{6} \end{pmatrix}.$$

As expected, since the matrix is real and symmetric, the eigenvalues are real and the eigenvectors are real and orthogonal. The transformation R which diagonalises A is therefore the orthogonal transformation given by

$$R = \begin{pmatrix} 1/\sqrt{3} & 1/\sqrt{2} & 1/\sqrt{6} \\ 1/\sqrt{3} & 0 & -2/\sqrt{6} \\ 1/\sqrt{3} & -1/\sqrt{2} & 1/\sqrt{6} \end{pmatrix},$$

with $R^{-1} = R^\mathsf{T}$ (orthogonality). One has then

$$R^\mathsf{T}AR = D\,, \quad \text{with} \quad D = \begin{pmatrix} 0 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -3 \end{pmatrix}. \tag{2.22}$$

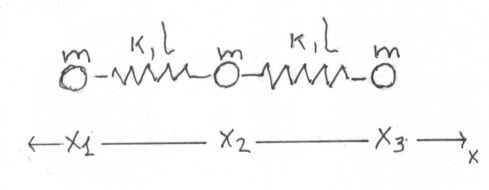

Figure 2.2: Three bodies of mass m coupled by two springs of elastic constant k and rest length l, constrained to move along one dimension (x).

The previous equation can be multiplied through on the left by R and on the right by R^{T} to obtain the equivalent relation

$$A = RDR^{\mathsf{T}} .$$

Let us now insert this expression for A into Eq. (2.21):

$$\ddot{\mathbf{x}} = \frac{k}{m} RDR^{\mathsf{T}}\mathbf{x} + \frac{k}{m}\mathbf{x}_0 ,$$

which can be multiplied through on the left by R^{T} and written as

$$R^{\mathsf{T}}\ddot{\mathbf{x}} = \frac{k}{m} DR^{\mathsf{T}}\mathbf{x} + \frac{k}{m} R^{\mathsf{T}}\mathbf{x}_0 . \tag{2.23}$$

Eq. (2.23) suggests the following definitions:

$$\mathbf{y} = R^{\mathsf{T}}\mathbf{x} , \quad \text{and} \quad \mathbf{y}_0 = R^{\mathsf{T}}\mathbf{x}_0 = \begin{pmatrix} 0 \\ \sqrt{2}l \\ 0 \end{pmatrix} \tag{2.24}$$

which, if inserted into it yields

$$\ddot{\mathbf{y}} = \frac{k}{m} D\mathbf{y} + \frac{k}{m}\mathbf{y}_0 . \tag{2.25}$$

Now, because the matrix D is diagonal, the three scalar differential equations contained in (2.25), are much simpler than the original system we set out to solve. In terms of the components y_1, y_2 and y_3 of $\mathbf{y}$, one has:

$$\ddot{y}_1 = 0 , \tag{2.26}$$

$$\ddot{y}_2 = -\frac{k}{m} y_2 + \sqrt{2}\frac{kl}{m} , \tag{2.27}$$

$$\ddot{y}_3 = -3\frac{k}{m} y_3 . \tag{2.28}$$

Each equation now depends on only one of the three unknown functions: the system of differential equation has been 'decoupled'. The identification of the new variables y_1, y_2 and y_3 which allowed for such a decoupling was made possible by the diagonalisation of the coupling matrix A. Notice that, although for simplicity we assumed all the masses, spring constants and rest lengths to be the same, the very same decoupling

would have been possible for any system of coupled harmonic oscillators, regardless of such details. The matrix A is in fact always symmetric, and hence diagonalisable, for these systems (a consequence, in a sense, of Newton's reaction principle).

The three variables y_1, y_2 and y_3, decoupling the dynamics of the coupled springs, are commonly known as *normal modes of oscillation*. The mode y_1 just represents the centre of mass of the three particles (here the sum of the three positions, as all the masses were assumed equal): in fact, its equation of motion is that of a free particle (the force acting on y_1 is zero), as expected since no external force is acting on the three particles. Modes y_2 and y_3 are instead at times referred to, respectively, as the 'breathing' and the 'Egyptian' mode. Normal modes play a central role in quantum mechanics as well, for instance in the description of quantized electromagnetic fields.

To complete our treatment, let us solve Eqs. (2.26-2.28):

$$y_1 = At + B \,, \tag{2.29}$$

$$y_2 = C\cos\left(\sqrt{\frac{k}{m}}t\right) + D\sin\left(\sqrt{\frac{k}{m}}t\right) + \frac{kl}{m}\frac{t^2}{\sqrt{2}} \,, \tag{2.30}$$

$$y_3 = E\cos\left(\sqrt{\frac{3k}{m}}t\right) + F\sin\left(\sqrt{\frac{3k}{m}}t\right) \,, \tag{2.31}$$

in terms of 6 integration constants A, B, C, D, E and F (which would have to be determined by the initial conditions, not given here). Finally, the general solution for the orignal variables x_1, x_2 and x_3 can be obtained by the relationship (due to the definition of $\mathbf{y}$):

$$\mathbf{x} = R\mathbf{y} \,.$$

Multiplying the vector of solutions $\mathbf{y}(t)$, determined by Eqs. (2.29-2.31), by R, one gets

$$\begin{aligned} x_1 &= \frac{1}{\sqrt{3}}(At+B) + \frac{1}{\sqrt{2}}\left(C\cos\left(\sqrt{\frac{k}{m}}t\right) + D\sin\left(\sqrt{\frac{k}{m}}t\right) + \frac{kl}{m}\frac{t^2}{\sqrt{2}}\right) \\ &\quad + \frac{1}{\sqrt{6}}\left(E\cos\left(\sqrt{\frac{3k}{m}}t\right) + F\sin\left(\sqrt{\frac{3k}{m}}t\right)\right) , \end{aligned}$$

$$x_2 = \frac{1}{\sqrt{3}}(At+B) - \sqrt{\frac{2}{3}}\left(E\cos\left(\sqrt{\frac{3k}{m}}t\right) + F\sin\left(\sqrt{\frac{3k}{m}}t\right)\right) ,$$

$$\begin{aligned} x_3 &= \frac{1}{\sqrt{3}}(At+B) - \frac{1}{\sqrt{2}}\left(C\cos\left(\sqrt{\frac{k}{m}}t\right) + D\sin\left(\sqrt{\frac{k}{m}}t\right) + \frac{kl}{m}\frac{t^2}{\sqrt{2}}\right) \\ &\quad + \frac{1}{\sqrt{6}}\left(E\cos\left(\sqrt{\frac{3k}{m}}t\right) + F\sin\left(\sqrt{\frac{3k}{m}}t\right)\right) . \end{aligned}$$

As apparent, such solutions would have been extremely difficult to find without applying matrix diagonalisation to decouple the dynamics of the system.