1 Matrices and vector spaces

In so far as vector algebra is concerned (see the summary in Section A.9 of Appendix A), a *vector* can be considered as a geometrical object which has both a magnitude and a direction, and may be thought of as an arrow fixed in our familiar three-dimensional space. This space, in turn, may be defined by reference to, say, the fixed stars. This geometrical definition of a vector is both useful and important since it is *independent* of any coordinate system with which we choose to label points in space.

In most specific applications, however, it is necessary at some stage to choose a coordinate system and to break down a vector into its *component vectors* in the directions of increasing coordinate values. Thus for a particular Cartesian coordinate system (for example) the component vectors of a vector $\mathbf{a}$ will be $a_x\mathbf{i}$, $a_y\mathbf{j}$ and $a_z\mathbf{k}$ and the complete vector will be

$$\mathbf{a} = a_x\mathbf{i} + a_y\mathbf{j} + a_z\mathbf{k}. \tag{1.1}$$

Although for many purposes we need consider only real three-dimensional space, the notion of a vector may be extended to more abstract spaces, which in general can have an arbitrary number of dimensions N. We may still think of such a vector as an "arrow" in this abstract space, so that it is again *independent* of any (N-dimensional) coordinate system with which we choose to label the space. As an example of such a space, which, though abstract, has very practical applications, we may consider the description of a mechanical or electrical system. If the state of a system is uniquely specified by assigning values to a set of N variables, which could include angles or currents, for example, then that state can be represented by a vector in an N-dimensional space, the vector having those values as its components.[1]

In this chapter we first discuss general *vector spaces* and their properties. We then go on to consider the transformation of one vector into another by a linear operator. This leads naturally to the concept of a *matrix*, a two-dimensional array of numbers. The properties of matrices are then developed and we conclude with a discussion of how to use these properties to solve systems of linear equations and study some oscillatory systems.

1 This is an approach often used in control engineering.

1.1 Vector spaces

A set of objects (vectors) $\mathbf{a}, \mathbf{b}, \mathbf{c}, \ldots$ is said to form a *linear vector space* V if:

(i) the set is closed under commutative and associative addition, so that

$$\mathbf{a} + \mathbf{b} = \mathbf{b} + \mathbf{a}, \tag{1.2}$$

$$(\mathbf{a} + \mathbf{b}) + \mathbf{c} = \mathbf{a} + (\mathbf{b} + \mathbf{c}); \tag{1.3}$$

(ii) the set is closed under multiplication by a scalar (any complex number) to form a new vector $\lambda\mathbf{a}$, the operation being both distributive and associative so that

$$\lambda(\mathbf{a} + \mathbf{b}) = \lambda\mathbf{a} + \lambda\mathbf{b}, \tag{1.4}$$

$$(\lambda + \mu)\mathbf{a} = \lambda\mathbf{a} + \mu\mathbf{a}, \tag{1.5}$$

$$\lambda(\mu\mathbf{a}) = (\lambda\mu)\mathbf{a}, \tag{1.6}$$

where λ and μ are arbitrary scalars;

(iii) there exists a *null vector* $\mathbf{0}$ such that $\mathbf{a} + \mathbf{0} = \mathbf{a}$ for all $\mathbf{a}$;

(iv) multiplication by unity leaves any vector unchanged, i.e. $1 \times \mathbf{a} = \mathbf{a}$;

(v) all vectors have a corresponding *negative vector* $-\mathbf{a}$ such that $\mathbf{a} + (-\mathbf{a}) = \mathbf{0}$. It follows from (1.5) with $\lambda = 1$ and $\mu = -1$ that $-\mathbf{a}$ is the same vector as $(-1) \times \mathbf{a}$.

We note that if we restrict all scalars to be real then we obtain a *real vector space* (an example of which is our familiar three-dimensional space); otherwise, in general, we obtain a *complex vector space*. We note that it is common to use the terms "vector space" and "space", instead of the more formal "linear vector space".

The *span* of a set of vectors $\mathbf{a}, \mathbf{b}, \ldots, \mathbf{s}$ is defined as the set of all vectors that may be written as a linear sum of the original set, i.e. all vectors

$$\mathbf{x} = \alpha\mathbf{a} + \beta\mathbf{b} + \cdots + \sigma\mathbf{s} \tag{1.7}$$

that result from the infinite number of possible values of the (in general complex) scalars $\alpha, \beta, \ldots, \sigma$. If $\mathbf{x}$ in (1.7) is equal to $\mathbf{0}$ for some choice of $\alpha, \beta, \ldots, \sigma$ (not *all* zero), i.e. if

$$\alpha\mathbf{a} + \beta\mathbf{b} + \cdots + \sigma\mathbf{s} = \mathbf{0}, \tag{1.8}$$

then the set of vectors $\mathbf{a}, \mathbf{b}, \ldots, \mathbf{s}$, is said to be *linearly dependent*. In such a set at least one vector is redundant, since it can be expressed as a linear sum of the others. If, however, (1.8) is not satisfied by *any* set of coefficients (other than the trivial case in which all the coefficients are zero) then the vectors are *linearly independent*, and no vector in the set can be expressed as a linear sum of the others.

If, in a given vector space, there exist sets of N linearly independent vectors, but no set of $N + 1$ linearly independent vectors, then the vector space is said to be N-dimensional. In this chapter we will limit our discussion to vector spaces of finite dimensionality.

1.1.1 Basis vectors

If V is an N-dimensional vector space then *any* set of N linearly independent vectors $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_N$ forms a *basis* for V. If $\mathbf{x}$ is an arbitrary vector lying in V then it can be written as a linear sum of these basis vectors:

$$\mathbf{x} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + \cdots + x_N\mathbf{e}_N = \sum_{i=1}^{N} x_i\mathbf{e}_i, \tag{1.9}$$

for some set of coefficients x_i. Since any $\mathbf{x}$ lying in the span of V can be expressed in terms of the *basis* or *base vectors* $\mathbf{e}_i$, the latter are said to form a *complete* set.

The coefficients x_i are called the *components* of $\mathbf{x}$ with respect to the $\mathbf{e}_i$-basis. They are *unique*, since if both

$$\mathbf{x} = \sum_{i=1}^{N} x_i\mathbf{e}_i \qquad \text{and} \qquad \mathbf{x} = \sum_{i=1}^{N} y_i\mathbf{e}_i, \qquad \text{then} \qquad \sum_{i=1}^{N}(x_i - y_i)\mathbf{e}_i = \mathbf{0}. \tag{1.10}$$

Since the $\mathbf{e}_i$ are linearly independent, each coefficient in the final equation in (1.10) must be individually zero and so $x_i = y_i$ for all $i = 1, 2, \ldots, N$.

It follows from this that *any* set of N linearly independent vectors can form a basis for an N-dimensional space.[2] If we choose a different set $\mathbf{e}'_i$, $i = 1, \ldots, N$ then we can write $\mathbf{x}$ as

$$\mathbf{x} = x'_1\mathbf{e}'_1 + x'_2\mathbf{e}'_2 + \cdots + x'_N\mathbf{e}'_N = \sum_{i=1}^{N} x'_i\mathbf{e}'_i, \tag{1.11}$$

but this does not change the vector $\mathbf{x}$. The vector $\mathbf{x}$ (a geometrical entity) is independent of the basis – it is only the components of $\mathbf{x}$ that depend upon the basis.

1.1.2 The inner product

This subsection contains a working summary of the definition and properties of inner products; for a fuller mathematical treatment the reader is referred to Appendix B.

To describe how two vectors in a vector space "multiply" (as opposed to add or subtract) we define their *inner product*, denoted in general by $\langle\mathbf{a}|\mathbf{b}\rangle$. This is a scalar function of vectors $\mathbf{a}$ and $\mathbf{b}$, though it is not necessarily real. Alternative notations for $\langle\mathbf{a}|\mathbf{b}\rangle$ are $(\mathbf{a}, \mathbf{b})$, or simply $\mathbf{a}\cdot\mathbf{b}$.

The scalar or dot product, $\mathbf{a}\cdot\mathbf{b} \equiv |\mathbf{a}||\mathbf{b}|\cos\theta$, of two vectors in real three-dimensional space (where θ is the angle between the vectors) is an example of an inner product. In effect the notion of an inner product $\langle\mathbf{a}|\mathbf{b}\rangle$ is a generalization of the dot product to more abstract vector spaces. The inner product has the following properties (in which, as usual,

2 All bases contain *exactly* N base vectors. A (putative) alternative base with M ($< N$) vectors would imply that there is no set of more than M linearly independent vectors – but the original base is just such a set, giving a contradiction. Equally, $M > N$ would imply the existence of a linearly independent set with more than N members – contradicting the specification for the original base set. Hence $M = N$.

a * superscript denotes complex conjugation):[3]

$$\langle \mathbf{a}|\mathbf{b}\rangle = \langle \mathbf{b}|\mathbf{a}\rangle^* , \tag{1.12}$$

$$\langle \mathbf{a}|\lambda\mathbf{b} + \mu\mathbf{c}\rangle = \lambda \langle \mathbf{a}|\mathbf{b}\rangle + \mu \langle \mathbf{a}|\mathbf{c}\rangle , \tag{1.13}$$

$$\langle \lambda\mathbf{a} + \mu\mathbf{b}|\mathbf{c}\rangle = \lambda^* \langle \mathbf{a}|\mathbf{c}\rangle + \mu^* \langle \mathbf{b}|\mathbf{c}\rangle , \tag{1.14}$$

$$\langle \lambda\mathbf{a}|\mu\mathbf{b}\rangle = \lambda^*\mu \langle \mathbf{a}|\mathbf{b}\rangle . \tag{1.15}$$

Following the analogy with the dot product in three-dimensional real space, two vectors in a general vector space are defined to be *orthogonal* if $\langle \mathbf{a}|\mathbf{b}\rangle = 0$.

In the same way, the *norm* of a vector $\mathbf{a}$, defined by $||\mathbf{a}|| = \langle \mathbf{a}|\mathbf{a}\rangle^{1/2}$, is clearly a generalization of the length or modulus $|\mathbf{a}|$ of a vector $\mathbf{a}$ in three-dimensional space. In a general vector space $\langle \mathbf{a}|\mathbf{a}\rangle$ can be positive or negative; however, we will be concerned only with spaces in which $\langle \mathbf{a}|\mathbf{a}\rangle \geq 0$ and which are therefore said to have a *positive semi-definite norm*. In such a space $\langle \mathbf{a}|\mathbf{a}\rangle = 0$ implies $\mathbf{a} = \mathbf{0}$.

It is usual when working with an N-dimensional vector space to use a basis $\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2, \ldots, \hat{\mathbf{e}}_N$ that has the desirable property of being *orthonormal* (the basis vectors are mutually orthogonal and each has unit norm), i.e. a basis that has the property

$$\left\langle \hat{\mathbf{e}}_i \,|\, \hat{\mathbf{e}}_j \right\rangle = \delta_{ij}. \tag{1.16}$$

Here δ_{ij} is the *Kronecker delta* symbol, defined by the properties

$$\delta_{ij} = \begin{cases} 1 & \text{for } i = j, \\ 0 & \text{for } i \neq j. \end{cases}$$

Using the above basis, any two vectors $\mathbf{a}$ and $\mathbf{b}$ can be written as

$$\mathbf{a} = \sum_{i=1}^{N} a_i \, \hat{\mathbf{e}}_i \qquad \text{and} \qquad \mathbf{b} = \sum_{i=1}^{N} b_i \, \hat{\mathbf{e}}_i .$$

Furthermore, *in such an orthonormal basis* we have, for any $\mathbf{a}$,

$$\left\langle \hat{\mathbf{e}}_j | \mathbf{a} \right\rangle = \sum_{i=1}^{N} \left\langle \hat{\mathbf{e}}_j | a_i \, \hat{\mathbf{e}}_i \right\rangle = \sum_{i=1}^{N} a_i \left\langle \hat{\mathbf{e}}_j \,|\, \hat{\mathbf{e}}_i \right\rangle = a_j . \tag{1.17}$$

Thus the components of $\mathbf{a}$ are given by $a_i = \langle \hat{\mathbf{e}}_i | \mathbf{a}\rangle$. Note that this is *not* true unless the basis is orthonormal.

We can write the inner product of $\mathbf{a}$ and $\mathbf{b}$ in terms of their components in an orthonormal basis as

$$\begin{aligned} \langle \mathbf{a}|\mathbf{b}\rangle &= \langle a_1 \, \hat{\mathbf{e}}_1 + a_2 \, \hat{\mathbf{e}}_2 + \cdots + a_N \, \hat{\mathbf{e}}_N | b_1 \, \hat{\mathbf{e}}_1 + b_2 \, \hat{\mathbf{e}}_2 + \cdots + b_N \, \hat{\mathbf{e}}_N \rangle \\ &= \sum_{i=1}^{N} a_i^* b_i \, \langle \hat{\mathbf{e}}_i \,|\, \hat{\mathbf{e}}_i \rangle + \sum_{i=1}^{N} \sum_{j \neq i}^{N} a_i^* b_j \left\langle \hat{\mathbf{e}}_i \,|\, \hat{\mathbf{e}}_j \right\rangle \\ &= \sum_{i=1}^{N} a_i^* b_i , \end{aligned}$$

3 It is a useful exercise in close analysis to deduce properties (1.14) and (1.15), on a justified step-by-step basis, using only those given in (1.12) and (1.13) and the general properties of complex conjugation.

where the second equality follows from (1.15) and the third from (1.16). This is clearly a generalization of the expression for the dot product of vectors in three-dimensional space.

The extension of the above results to the case where the base vectors $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_N$ are *not* orthonormal is more mathematically complicated and given in Appendix B.

1.1.3 Some useful inequalities

For a set of objects (vectors) forming a linear vector space in which $\langle \mathbf{a}|\mathbf{a}\rangle \geq 0$ for all $\mathbf{a}$, there are a number of inequalities that often prove useful. Here we only list them; for the corresponding proofs the reader is referred to Appendix C.

(i) *Schwarz's inequality* states that

$$|\langle \mathbf{a}|\mathbf{b}\rangle| \leq ||\mathbf{a}||\,||\mathbf{b}||, \tag{1.18}$$

where the equality holds when $\mathbf{a}$ is a scalar multiple of $\mathbf{b}$, i.e. when $\mathbf{a} = \lambda\mathbf{b}$. It is important here to distinguish between the *absolute value* of a scalar, $|\lambda|$, and the *norm* of a vector, $||\mathbf{a}||$.

(ii) The *triangle inequality* states that

$$||\mathbf{a}+\mathbf{b}|| \leq ||\mathbf{a}|| + ||\mathbf{b}|| \tag{1.19}$$

and is the intuitive analogue of the observation that the length of any one side of a triangle cannot be greater than the sum of the lengths of the other two sides.

(iii) *Bessel's inequality* states that if $\hat{\mathbf{e}}_i$, $i = 1, 2, \ldots, N$ form an orthonormal basis in an N-dimensional vector space, then

$$||\mathbf{a}||^2 \geq \sum_i^M |\langle \hat{\mathbf{e}}_i|\mathbf{a}\rangle|^2, \tag{1.20}$$

where the equality holds if $M = N$. If $M < N$ then inequality results, unless the basis vectors omitted all have $a_i = 0$. This is the analogue of $|\mathbf{x}|^2$ for a three-dimensional vector $\mathbf{v}$ being equal to the sum of the squares of all its components, and if any are omitted the sum may fall short of $|\mathbf{x}|^2$.

To these inequalities can be added one equality that sometimes proves useful. The *parallelogram equality* reads

$$||\mathbf{a}+\mathbf{b}||^2 + ||\mathbf{a}-\mathbf{b}||^2 = 2\left(||\mathbf{a}||^2 + ||\mathbf{b}||^2\right), \tag{1.21}$$

and may be proved straightforwardly from the properties of the inner product.

1.2 Linear operators

We now discuss the action of *linear operators* on vectors in a vector space. A linear operator $\mathcal{A}$ associates with every vector $\mathbf{x}$ another vector

$$\mathbf{y} = \mathcal{A}\mathbf{x},$$

in such a way that, for two vectors $\mathbf{a}$ and $\mathbf{b}$,

$$\mathcal{A}(\lambda\mathbf{a} + \mu\mathbf{b}) = \lambda\mathcal{A}\mathbf{a} + \mu\mathcal{A}\mathbf{b},$$

where λ, μ are scalars. We say that $\mathcal{A}$ "operates" on $\mathbf{x}$ to give the vector $\mathbf{y}$. We note that the action of $\mathcal{A}$ is *independent* of any basis or coordinate system and may be thought of as "transforming" one geometrical entity (i.e. a vector) into another.

If we now introduce a basis $\mathbf{e}_i$, $i = 1, 2, \ldots, N$, into our vector space then the action of $\mathcal{A}$ on each of the basis vectors is to produce a linear combination of the latter; this may be written as

$$\mathcal{A}\mathbf{e}_j = \sum_{i=1}^{N} A_{ij}\mathbf{e}_i, \tag{1.22}$$

where A_{ij} is the ith component of the vector $\mathcal{A}\mathbf{e}_j$ in this basis; collectively the numbers A_{ij} are called the components of the linear operator in the $\mathbf{e}_i$-basis. *In this basis* we can express the relation $\mathbf{y} = \mathcal{A}\mathbf{x}$ in component form as

$$\mathbf{y} = \sum_{i=1}^{N} y_i\mathbf{e}_i = \mathcal{A}\left(\sum_{j=1}^{N} x_j\mathbf{e}_j\right) = \sum_{j=1}^{N} x_j \sum_{i=1}^{N} A_{ij}\mathbf{e}_i,$$

and hence, in purely component form, in this basis we have

$$y_i = \sum_{j=1}^{N} A_{ij}x_j. \tag{1.23}$$

If we had chosen a different basis $\mathbf{e}'_i$, in which the components of $\mathbf{x}$, $\mathbf{y}$ and $\mathcal{A}$ are x'_i, y'_i and A'_{ij} respectively then the geometrical relationship $\mathbf{y} = \mathcal{A}\mathbf{x}$ would be represented in this new basis by

$$y'_i = \sum_{j=1}^{N} A'_{ij}x'_j.$$

We have so far assumed that the vector $\mathbf{y}$ is in the same vector space as $\mathbf{x}$. If, however, $\mathbf{y}$ belongs to a different vector space, which may in general be M-dimensional ($M \neq N$) then the above analysis needs a slight modification. By introducing a basis set $\mathbf{f}_i$, $i = 1, 2, \ldots, M$, into the vector space to which $\mathbf{y}$ belongs we may generalize (1.22) as

$$\mathcal{A}\mathbf{e}_j = \sum_{i=1}^{M} A_{ij}\mathbf{f}_i,$$

where the components A_{ij} of the linear operator $\mathcal{A}$ relate to both of the bases $\mathbf{e}_j$ and $\mathbf{f}_i$.

The basic properties of linear operators, arising from their definition, are summarized as follows. If $\mathbf{x}$ is a vector and $\mathcal{A}$ and $\mathcal{B}$ are two linear operators then

$$(\mathcal{A} + \mathcal{B})\mathbf{x} = \mathcal{A}\mathbf{x} + \mathcal{B}\mathbf{x},$$
$$(\lambda\mathcal{A})\mathbf{x} = \lambda(\mathcal{A}\mathbf{x}),$$
$$(\mathcal{A}\mathcal{B})\mathbf{x} = \mathcal{A}(\mathcal{B}\mathbf{x}),$$

where in the last equality we see that the action of two linear operators in succession is associative. However, the product of two general linear operators is not commutative, i.e. $\mathcal{AB}\mathbf{x} \neq \mathcal{BA}\mathbf{x}$ in general.[4]

In an obvious way we define the null (or zero) and identity operators by

$$\mathcal{O}\mathbf{x} = \mathbf{0} \qquad \text{and} \qquad \mathcal{I}\mathbf{x} = \mathbf{x},$$

for any vector $\mathbf{x}$ in our vector space. Two operators $\mathcal{A}$ and $\mathcal{B}$ are equal if $\mathcal{A}\mathbf{x} = \mathcal{B}\mathbf{x}$ for all vectors $\mathbf{x}$. Finally, if there exists an operator $\mathcal{A}^{-1}$ such that

$$\mathcal{AA}^{-1} = \mathcal{A}^{-1}\mathcal{A} = \mathcal{I}$$

then $\mathcal{A}^{-1}$ is the *inverse* of $\mathcal{A}$. Some linear operators do not possess an inverse and are called *singular*, whilst those operators that do have an inverse are termed *non-singular*.

1.3 Matrices

We have seen that in a particular basis $\mathbf{e}_i$ both vectors and linear operators can be described in terms of their components with respect to the basis. These components may be displayed as an array of numbers called a *matrix*. In general, if a linear operator $\mathcal{A}$ transforms vectors from an N-dimensional vector space, for which we choose a basis $\mathbf{e}_j$, $j = 1, 2, \ldots, N$, into vectors belonging to an M-dimensional vector space, with basis $\mathbf{f}_i$, $i = 1, 2, \ldots, M$, then we may represent the operator $\mathcal{A}$ by the matrix

$$\mathsf{A} = \begin{pmatrix} A_{11} & A_{12} & \ldots & A_{1N} \\ A_{21} & A_{22} & \ldots & A_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ A_{M1} & A_{M2} & \ldots & A_{MN} \end{pmatrix}. \tag{1.24}$$

The *matrix elements* A_{ij} are the components of the linear operator with respect to the bases $\mathbf{e}_j$ and $\mathbf{f}_i$; the component A_{ij} of the linear operator appears in the ith row and jth column of the matrix. The array has M rows and N columns and is thus called an $M \times N$ matrix. If the dimensions of the two vector spaces are the same, i.e. $M = N$ (for example, if they are the same vector space) then we may represent $\mathcal{A}$ by an $N \times N$ or *square* matrix of *order* N. The component A_{ij}, which in general may be complex, is also commonly denoted by $(\mathsf{A})_{ij}$.

In a similar way we may denote a vector $\mathbf{x}$ in terms of its components x_i in a basis $\mathbf{e}_i$, $i = 1, 2, \ldots, N$, by the array

$$\mathsf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{pmatrix},$$

4 Consider a two-dimensional linear vector space in which a typical vector is $\mathbf{x} = (x_1, x_2)$, with linear operators $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ defined by $\mathcal{A}\mathbf{x} = (2x_1 + x_2, x_2)$, $\mathcal{B}\mathbf{x} = (x_1, x_1 + 2x_2)$ and $\mathcal{C}\mathbf{x} = (x_1 - x_2, 2x_2)$. Show that, although $\mathcal{A}$ and $\mathcal{C}$ commute, $\mathcal{A}$ and $\mathcal{B}$ do not.

which is a special case of (1.24) and is called a *column matrix* (or conventionally, and slightly confusingly, a *column vector* or even just a *vector* – strictly speaking the term "vector" refers to the geometrical entity $\mathbf{x}$). The column matrix x can also be written as

$$\mathsf{x} = (x_1 \quad x_2 \quad \cdots \quad x_N)^{\mathrm{T}},$$

which is the *transpose* of a *row matrix* (see Section 1.6).

We note that in a different basis $\mathbf{e}'_i$ the vector $\mathbf{x}$ would be represented by a *different* column matrix containing the components x'_i in the new basis, i.e.

$$\mathsf{x}' = \begin{pmatrix} x'_1 \\ x'_2 \\ \vdots \\ x'_N \end{pmatrix}.$$

Thus, we use x and x' to denote different column matrices which, in different bases $\mathbf{e}_i$ and $\mathbf{e}'_i$, represent the *same* vector $\mathbf{x}$. In many texts, however, this distinction is not made and $\mathbf{x}$ (rather than x) is equated to the corresponding column matrix; if we regard $\mathbf{x}$ as the geometrical entity, however, this can be misleading and so we explicitly make the distinction. A similar argument follows for linear operators; the same linear operator $\mathcal{A}$ is described in different bases by different matrices A and A', containing different matrix elements.

1.4 Basic matrix algebra

The basic algebra of matrices may be deduced from the properties of the linear operators that they represent. In a given basis the action of two linear operators $\mathcal{A}$ and $\mathcal{B}$ on an arbitrary vector $\mathbf{x}$ (see towards the end of Section 1.2), when written in terms of components using (1.23), is given by

$$\begin{aligned}
\sum_j (\mathsf{A}+\mathsf{B})_{ij} x_j &= \sum_j A_{ij} x_j + \sum_j B_{ij} x_j, \\
\sum_j (\lambda \mathsf{A})_{ij} x_j &= \lambda \sum_j A_{ij} x_j, \\
\sum_j (\mathsf{AB})_{ij} x_j &= \sum_k A_{ik} (\mathsf{Bx})_k = \sum_j \sum_k A_{ik} B_{kj} x_j.
\end{aligned}$$

Now, since $\mathbf{x}$ is arbitrary, we can immediately deduce the way in which matrices are added or multiplied, i.e.[5]

$$(\mathsf{A}+\mathsf{B})_{ij} = A_{ij} + B_{ij}, \tag{1.25}$$

5 Express the operators appearing in footnote 4 in matrix form and then use (1.27) to demonstrate their commutation or otherwise. Do operators $\mathcal{B}$ and $\mathcal{C}$ commute?

$$(\lambda \mathsf{A})_{ij} = \lambda A_{ij}, \tag{1.26}$$

$$(\mathsf{AB})_{ij} = \sum_k A_{ik} B_{kj}. \tag{1.27}$$

We note that a matrix element may, in general, be complex. We now discuss matrix addition and multiplication in more detail.

1.4.1 Matrix addition and multiplication by a scalar

From (1.25) we see that the sum of two matrices, $\mathsf{S} = \mathsf{A} + \mathsf{B}$, is the matrix whose elements are given by

$$S_{ij} = A_{ij} + B_{ij}$$

for every pair of subscripts i, j, with $i = 1, 2, \ldots, M$ and $j = 1, 2, \ldots, N$. For example, if A and B are 2×3 matrices then $\mathsf{S} = \mathsf{A} + \mathsf{B}$ is given by

$$\begin{pmatrix} S_{11} & S_{12} & S_{13} \\ S_{21} & S_{22} & S_{23} \end{pmatrix} = \begin{pmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \end{pmatrix} + \begin{pmatrix} B_{11} & B_{12} & B_{13} \\ B_{21} & B_{22} & B_{23} \end{pmatrix}$$

$$= \begin{pmatrix} A_{11} + B_{11} & A_{12} + B_{12} & A_{13} + B_{13} \\ A_{21} + B_{21} & A_{22} + B_{22} & A_{23} + B_{23} \end{pmatrix}. \tag{1.28}$$

Clearly, for the sum of two matrices to have any meaning, the matrices must have the same dimensions, i.e. both be $M \times N$ matrices.

From definition (1.28) it follows that $\mathsf{A} + \mathsf{B} = \mathsf{B} + \mathsf{A}$ and that the sum of a number of matrices can be written unambiguously without bracketing, i.e. matrix addition is *commutative* and *associative*.

The difference of two matrices is defined by direct analogy with addition. The matrix $\mathsf{D} = \mathsf{A} - \mathsf{B}$ has elements

$$D_{ij} = A_{ij} - B_{ij}, \qquad \text{for } i = 1, 2, \ldots, M,\ j = 1, 2, \ldots, N. \tag{1.29}$$

From (1.26) the product of a matrix A with a scalar λ is the matrix with elements λA_{ij}, for example

$$\lambda \begin{pmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \end{pmatrix} = \begin{pmatrix} \lambda A_{11} & \lambda A_{12} & \lambda A_{13} \\ \lambda A_{21} & \lambda A_{22} & \lambda A_{23} \end{pmatrix}. \tag{1.30}$$

Multiplication by a scalar is distributive and associative.

The following example illustrates these three elementary properties or definitions.

Example The matrices A, B and C are given by

$$\mathsf{A} = \begin{pmatrix} 2 & -1 \\ 3 & 1 \end{pmatrix}, \qquad \mathsf{B} = \begin{pmatrix} 1 & 0 \\ 0 & -2 \end{pmatrix}, \qquad \mathsf{C} = \begin{pmatrix} -2 & 1 \\ -1 & 1 \end{pmatrix}.$$

Find the matrix $\mathsf{D} = \mathsf{A} + 2\mathsf{B} - \mathsf{C}$.

Dealing separately with the elements in each particular position in the various matrices, we have

$$\begin{aligned} \mathsf{D} &= \begin{pmatrix} 2 & -1 \\ 3 & 1 \end{pmatrix} + 2\begin{pmatrix} 1 & 0 \\ 0 & -2 \end{pmatrix} - \begin{pmatrix} -2 & 1 \\ -1 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 2 + 2\times 1 - (-2) & -1 + 2\times 0 - 1 \\ 3 + 2\times 0 - (-1) & 1 + 2\times(-2) - 1 \end{pmatrix} = \begin{pmatrix} 6 & -2 \\ 4 & -4 \end{pmatrix}. \end{aligned}$$

As a reminder, we note that for the question to have had any meaning, A, B and C all had to have the same dimensions, 2×2 in practice; the answer, D, is also 2×2. ◀

From the above considerations we see that the set of all, in general complex, $M \times N$ matrices (with fixed M and N) provide an example of a linear vector space – one whose elements have no obvious "arrow-like" qualities.

The space is of dimension MN. One basis for it is the set of $M \times N$ matrices $\mathsf{E}^{(p,q)}$ with the property that $E_{ij}^{(p,q)} = 1$ if $i = p$ and $j = q$ whilst $E_{ij}^{(p,q)} = 0$ for all other values of i and j, i.e. each matrix has only one non-zero entry, and that equals unity. Here the pair (p, q) is simply a label that picks out a particular one of the matrices $E^{(p,q)}$, the total number of which is MN.

1.4.2 Multiplication of matrices

Let us consider again the "transformation" of one vector into another, $\mathbf{y} = \mathcal{A}\mathbf{x}$, which, from (1.23), may be described in terms of components with respect to a particular basis as

$$y_i = \sum_{j=1}^{N} A_{ij} x_j \qquad \text{for } i = 1, 2, \ldots, M. \tag{1.31}$$

Writing this in matrix form as $\mathsf{y} = \mathsf{A}\mathsf{x}$ we have

$$\begin{pmatrix} y_1 \\ \boxed{y_2} \\ \vdots \\ y_M \end{pmatrix} = \begin{pmatrix} A_{11} & A_{12} & \ldots & A_{1N} \\ \boxed{A_{21} \quad A_{22} \quad \ldots \quad A_{2N}} \\ \vdots & \vdots & \ddots & \vdots \\ A_{M1} & A_{M2} & \ldots & A_{MN} \end{pmatrix} \begin{pmatrix} \boxed{\begin{matrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{matrix}} \end{pmatrix} \tag{1.32}$$

where we have highlighted with boxes the components used to calculate the element y_2: using (1.31) for $i = 2$,

$$y_2 = A_{21}x_1 + A_{22}x_2 + \cdots + A_{2N}x_N.$$

All the other components y_i are calculated similarly.