CHAPTER 1

Review of Prerequisite Mathematics

This chapter reviews, mostly without proof, a number of mathematical topics that you should have encountered before. The material may be read rapidly, and referred to later as you have need. For a comprehensive treatment of any of these items, consult your previous texts.

1.1. Set notation

Defining sets

$\mathbb{R}$	The real numbers
$\mathbb{C}$	The complex numbers
$\mathbb{Z}$	The integers
$\mathbb{N}$	The natural numbers (positive integers)
$\mathbb{Q}$	The rational numbers
$\emptyset$	The empty set, $\{\}$
$x \in A$	x is an element of the set A, *e.g.*, $0 \in \{-1, 0, 1\}$
$x \notin A$	x is not an element of the set A, *e.g.*, $2 \notin \{-1, 0, 1\}$
(a, b)	Open interval on the real line: $x \in (a, b)$ means $a < x < b$. (a, b) also denotes an ordered pair of numbers, *e.g.*, $(1, 2)$ is the point in the xy plane with coordinates $x = 1$ and $y = 2$.
$[a, b]$	Closed interval on the real line: $x \in [a, b]$ means $a \leq x \leq b$.
$(a, b], [a, b)$	Half-open intervals on the real line
$\{x \mid \textit{condition}\}$	Set-builder notation, *e.g.*, $\{x \in \mathbb{R} \mid x > a\}$ denotes the set of all real x such that x is greater than a (*i.e.*, the open interval (a, ∞)). $\left\{f \mid \int \lvert f(x) \rvert dx < \infty\right\}$ denotes the set of functions f which are absolutely integrable.

Relationships between sets

$A \subseteq B$	The set A is a subset of the set B, *e.g.*, $\mathbb{N} \subseteq \mathbb{Z}$.
$A = B$	The sets A and B are equal: $A \subseteq B$ and $B \subseteq A$

Operations on sets

$A \cap B$	The intersection of sets A and B, $A \cap B = \{x \mid x \in A \text{ and } x \in B\}$, *e.g.*, $\{0, 1\} \cap \{1, 2\} = \{1\}$. Sets A and B are *disjoint* if $A \cap B = \emptyset$.
$A \cup B$	The union of sets A and B, $A \cup B = \{x \mid x \in A \text{ or } x \in B\}$, *e.g.*, $\{0, 1\} \cup \{1, 2\} = \{0, 1, 2\}$.
$A \setminus B$	The difference of sets A and B, $A \setminus B = \{x \mid x \in A \text{ and } x \notin B\}$, *e.g.*, $\{1, 2, 3\} \setminus \{1, 3\} = \{2\}$, but $\{1, 3\} \setminus \{1, 2, 3\} = \emptyset$.
$A \times B$	The product of sets A and B, $\{\text{ordered pairs } (a, b) \mid a \in A,\ b \in B\}$, *e.g.*, $\{0, 1\} \times \{2, 3\} = \{(0, 2), (0, 3), (1, 2), (1, 3)\}$; $\mathbb{R} \times \mathbb{R}$, also known as $\mathbb{R}^2$, is the real plane; $[a, b] \times [c, d]$ is a rectangular region in the plane, a subset of $\mathbb{R}^2$.
$f : A \to B$	A mapping f which assigns each element of A to one or more elements of B, *i.e.*, for each $x \in A$, there is $f(x) \in B$.

1.2. Vectors in space

In everyday life, we can measure the distance between two points in space. This distance is nonnegative, and zero only if the two points are identical. For two points a and b, it makes

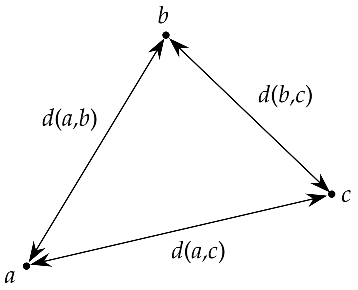

FIGURE 1.1: The triangle inequality: $d(a,c) \leq d(a,b) + d(b,c)$.

sense that the distance d from a to b is the same as the distance from b to a: $d(a,b) = d(b,a)$. Three points in space, a, b, c, constitute the vertices of a triangle. The distances between them, $d(a,b)$, $d(b,c)$, and $d(a,c)$, are the lengths of the sides of the triangle. It is always true that one side of a triangle is no longer than the sum of the lengths of the other two sides, *e.g.*,

$$d(a,c) \leq d(a,b) + d(b,c) \tag{1.1}$$

(with equality if the three points are collinear). This relationship is called the *triangle inequality* (Figure 1.1).

Distance enables us to partially orient ourselves in space relative to other objects. We can say if an object is near or far, and specify sets of objects which are within a particular distance (radius) of us. But distance doesn't tell direction; we can only say "how far", not "which way".

Vectors specify "which way" as well as "how far". Each point in space is uniquely positioned at the tip of a vector whose tail is fixed at a common reference point, or origin. The vector represents the displacement of the point from the origin. The length of a vector $\mathbf{v}$ is a nonnegative real number called the *norm*, denoted $\|\mathbf{v}\|$. The norm of a vector is equal to zero only if $\mathbf{v}$ is the zero vector. A *unit vector* is a vector whose length is one. A vector is *normalized*, made into a unit vector, by dividing it by its norm: $\mathbf{v}/\|\mathbf{v}\|$. Multiplying a vector by a ordinary real number, or *scalar*, c changes ("scales") its length: $\|c\mathbf{v}\| = |c|\ \|\mathbf{v}\|$. If $c > 0$, the direction of the vector is unchanged, but if $c < 0$, the direction is reversed: $\mathbf{v}$ and $-\mathbf{v} = -1\mathbf{v}$ have the same length, but point in opposite directions. A vector $\mathbf{v}$ can be represented as the product of a nonnegative scalar equal to $\mathbf{v}$'s norm, and a unit vector pointing in $\mathbf{v}$'s direction.

If I am located by vector $\mathbf{p}$ and you are located by vector $\mathbf{q}$, our respective distances from the origin are $\|\mathbf{p}\|$ and $\|\mathbf{q}\|$. We may define a third vector $\mathbf{r}$ that runs from the tip of $\mathbf{p}$ to the tip of $\mathbf{q}$ and directs me to you. Adding $\mathbf{r}$ to my position puts me at your position: $\mathbf{q} = \mathbf{p}+\mathbf{r}$, and so $\mathbf{r} = \mathbf{q} - \mathbf{p}$. The path from you to me is the opposite vector $\mathbf{p} - \mathbf{q} = -\mathbf{r}$. The norm $\|\mathbf{r}\| = \|-\mathbf{r}\|$ is the distance between us. The vectors $\mathbf{p}$, $\mathbf{q}$, and $\mathbf{q}-\mathbf{p}$ form a triangle, and their lengths, which are distances between points, also obey the triangle inequality: $\|\mathbf{q} - \mathbf{p}\| \leq \|\mathbf{p}\| + \|\mathbf{q}\|$. (Figure 1.2)

The *dot product* of two vectors is a scalar quantity defined

$$\mathbf{v} \cdot \mathbf{w} = \|\mathbf{v}\|\ \|\mathbf{w}\| \cos\theta\ . \tag{1.2}$$

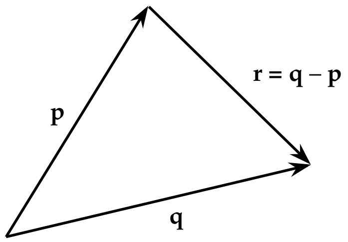

FIGURE 1.2: Addition and subtraction of vectors. The vector **r** is the difference of **p** and **q**. The vector **q** is the sum of **p** and **r**. The lengths of **p**, **q**, and **r** obey the triangle inequality.

where θ is the angle between **v** and **w**. The dot product of two nonzero vectors is zero if the angle θ is $\frac{\pi}{2}$; the vectors are then said to be *orthogonal.* Orthogonal unit vectors are said to be *orthonormal.* The dot product of a vector with itself is the square of its norm,

$$\mathbf{v}\cdot\mathbf{v} = \|\mathbf{v}\|\ \|\mathbf{v}\|\cos(0) = \|\mathbf{v}\|^2\ . \tag{1.3}$$

Algebraically, the dot product behaves like multiplication. For vectors **u**, **v**, and **w**, and scalar c,

$$\mathbf{v}\cdot\mathbf{w} = \mathbf{w}\cdot\mathbf{v} \tag{1.4a}$$

$$(c\mathbf{v})\cdot\mathbf{w} = c(\mathbf{v}\cdot\mathbf{w}) \tag{1.4b}$$

$$\mathbf{u}\cdot(\mathbf{v}+\mathbf{w}) = \mathbf{u}\cdot\mathbf{v}+\mathbf{u}\cdot\mathbf{w} \tag{1.4c}$$

A vector **v** in a plane may be expanded (or decomposed) into the sum (also called the *resultant*) of two orthogonal vectors, which are called the *components* of **v**. In a Cartesian coordinate system, the orthogonal directions are frequently denoted x and y. We denote unit vectors pointing in the x and y directions by $\mathbf{e}_x$ and $\mathbf{e}_y$. Then, the vector decomposition is

$$\mathbf{v} = \mathbf{v}_x + \mathbf{v}_y = v_x\mathbf{e}_x + v_y\mathbf{e}_y$$

where v_x and v_y are (scalar) coefficients. To calculate v_x and v_y, write the vector in the form $\mathbf{v} = v_x\mathbf{e}_x + v_y\mathbf{e}_y$ and take the dot product of both sides with $\mathbf{e}_x$ and $\mathbf{e}_y$, respectively.

$$\mathbf{v}\cdot\mathbf{e}_x = v_x(\mathbf{e}_x\cdot\mathbf{e}_x) + v_y(\mathbf{e}_y\cdot\mathbf{e}_x)$$
$$\mathbf{v}\cdot\mathbf{e}_y = v_x(\mathbf{e}_x\cdot\mathbf{e}_y) + v_y(\mathbf{e}_y\cdot\mathbf{e}_y)$$

Because the unit vectors are orthonormal, $\mathbf{e}_x\cdot\mathbf{e}_x = 1$, $\mathbf{e}_y\cdot\mathbf{e}_y = 1$, and $\mathbf{e}_x\cdot\mathbf{e}_y = \mathbf{e}_y\cdot\mathbf{e}_x = 0$. Hence,

$$\mathbf{v}\cdot\mathbf{e}_x = v_x$$
$$\mathbf{v}\cdot\mathbf{e}_y = v_y$$

The components $\mathbf{v}_x = v_x\mathbf{e}_x$ and $\mathbf{v}_y = v_y\mathbf{e}_y$ are also called the *orthogonal projections* of **v** along $\mathbf{e}_x$ and $\mathbf{e}_y$, respectively.

The expression $v_x\mathbf{e}_x + v_y\mathbf{e}_y$ is called a *linear combination* of $\mathbf{e}_x$ and $\mathbf{e}_y$. The orthonormal vectors $\mathbf{e}_x$ and $\mathbf{e}_y$ are called a *basis* for the plane, and are said to *span* the plane, because *any* vector in the plane can be written as a linear combination of $\mathbf{e}_x$ and $\mathbf{e}_y$.

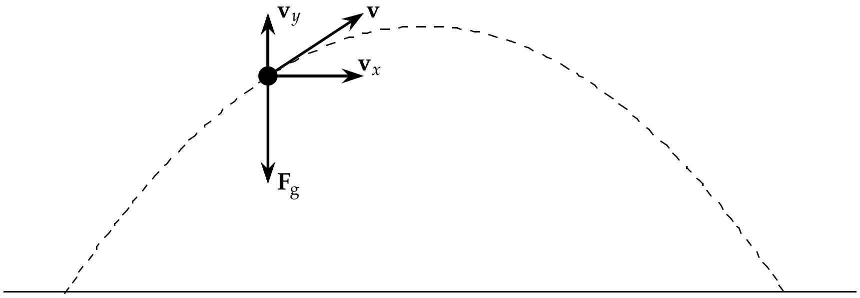

FIGURE 1.3: Horizontal and vertical components of velocity are orthogonal and independent.

Physically, orthogonal vectors represent noninteracting actions or motions. For example, in the parabolic motion of a projectile, gravity acts to decelerate/accelerate the vertical component of motion, so $\frac{dv_y}{dt} = -g$. On the other hand, the x component of velocity is unaffected by gravity: $v_x = \text{const}$ (Figure 1.3).

Expressing vectors in terms of their components relative to a common basis greatly simplifies vector calculations. Let $\mathbf{v}$ and $\mathbf{w}$ two vectors in the plane, expressed in terms of their orthogonal components, $\mathbf{v} = v_x\mathbf{e}_x + v_y\mathbf{e}_y$ and $\mathbf{w} = w_x\mathbf{e}_x + w_y\mathbf{e}_y$, and let c be a scalar. Then,

$$\begin{aligned}
c\mathbf{v} &= c(v_x\mathbf{e}_x + v_y\mathbf{e}_y) = (cv_x)\mathbf{e}_x + (cv_y)\mathbf{e}_y && (1.5a)\\
\mathbf{v} + \mathbf{w} &= v_x\mathbf{e}_x + v_y\mathbf{e}_y + w_x\mathbf{e}_x + w_y\mathbf{e}_y \\
&= (v_x + w_x)\mathbf{e}_x + (v_y + w_y)\mathbf{e}_y && (1.5b)\\
\mathbf{v} \cdot \mathbf{w} &= (v_x\mathbf{e}_x + v_y\mathbf{e}_y) \cdot (w_x\mathbf{e}_x + w_y\mathbf{e}_y) \\
&= v_x w_x(\mathbf{e}_x \cdot \mathbf{e}_x) + v_x w_y(\mathbf{e}_x \cdot \mathbf{e}_y) + v_y w_x(\mathbf{e}_y \cdot \mathbf{e}_x) + v_y w_y(\mathbf{e}_y \cdot \mathbf{e}_y) \\
&= v_x w_x + v_y w_y && (1.5c)\\
\mathbf{v} \cdot \mathbf{v} &= v_x^2 + v_y^2 = \|v\|^2 \ . && (1.5d)
\end{aligned}$$

Basis vectors are not unique. Instead of $\mathbf{e_x}$ and $\mathbf{e_y}$, we could use orthonormal vectors $\mathbf{e}_{x'}$ and $\mathbf{e}_{y'}$, which are rotated by 45° from $\mathbf{e}_x$ and $\mathbf{e}_y$ (Figure 1.4). The coefficients $v_{x'}$ and $v_{y'}$ will be different from v_x and v_y, because the $\mathbf{e}_{x'}$ and $\mathbf{e}_{y'}$ point in different directions than $\mathbf{e}_x$ and $\mathbf{e}_y$ (Figure 1.4). However, the resultant vector $\mathbf{v}$ is unchanged by the choice of basis, and vector calculations (sum, difference, dot product, norm) carried out with components relative to either basis will yield the same results, *e.g.*, $v_x^2 + v_y^2 = v_{x'}^2 + v_{y'}^2 = \|\mathbf{v}\|^2$. For modelling a physical quantity, one basis may be preferred over others. A good example is the case of a body moving in a circular path under the influence of a central force, *e.g.*, a satellite orbiting a planet. (Figure 1.5) The "polar" form, based on $\mathbf{e}_r$ and $\mathbf{e}_\theta$, is more natural than the Cartesian form, and the equations of motion are simpler than in Cartesian coordinates.

Vectors in the plane are easily generalized to vectors in three-dimensional space by adding a third basis vector orthogonal to the plane and a third component along that basis vector.

$$\mathbf{v} = v_x\mathbf{e}_x + v_y\mathbf{e}_y + v_z\mathbf{e}_z \ .$$

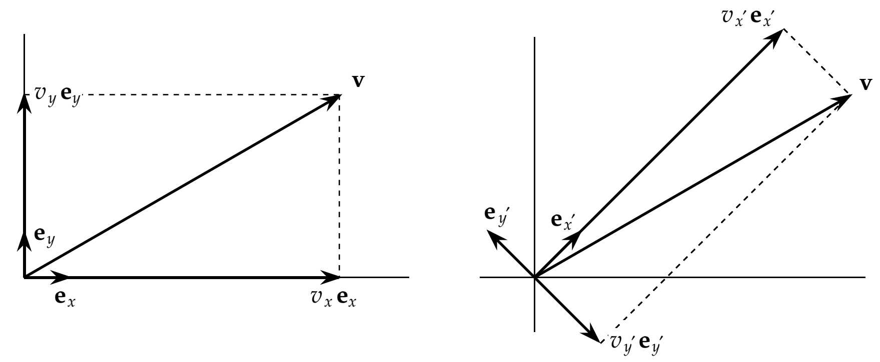

FIGURE 1.4: The decomposition of a vector depends on the choice of basis vectors.

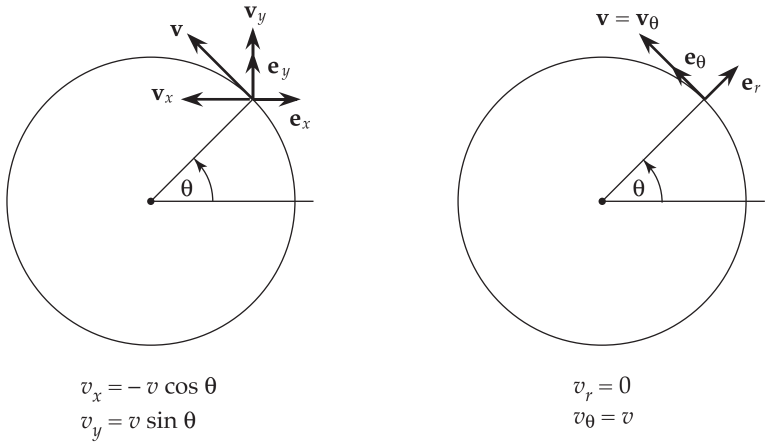

FIGURE 1.5: Circular motion is best described in polar coordinates.

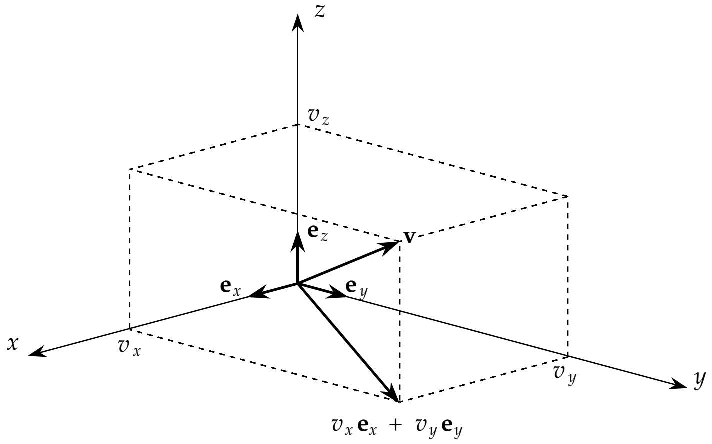

FIGURE 1.6: A vector in three-dimensional space.

(Figure 1.6). The length of the vector is

$$\|\mathbf{v}\| = \sqrt{v_x^2 + v_y^2 + v_z^2}\ ,$$

generalizing the Pythagorean formula. Vectors in three dimensions possess yet another multiplication operation called the cross product, which is useful in physics but has no relevance to Fourier theory and will not be considered further.

In Fourier analysis, an arbitrary function (waveform) is expressed as a sum of simple (sine and cosine) functions, which behave like orthogonal basis vectors in a generalized sense. The next chapter will lay the foundation for this important generalization of vector ideas.

1.3. Complex numbers

Complex numbers arose initially in the study of roots of certain algebraic equations. For example, you know that the square roots of 1, which are the solutions of the quadratic equation $x^2 - 1 = 0$, are $+1$ and -1. With a change of sign, the equation becomes $x^2 + 1 = 0$, which has no real-valued solution. We define an “imaginary number”, i, to be the square root of -1. This number has the property that $i^2 = -1$, and also that $-i \cdot -i = (-1)(-1)i^2 = -1$. With this invention, the equation $x^2 + 1 = 0$ has two roots, $x = +i$ and $-i$.

In mathematics and physics, the symbol i is used for the square root of -1. In engineering, j is frequently used instead of i. This is because electrical engineers in particular are accustomed to using i to denote electric current. Physicists, on the other hand, use j to stand for current density, and i for $\sqrt{-1}$.

The product of i with a real number y is an imaginary number iy. The combination of a real number x and an imaginary number iy is called a complex number, and is written as a sum, $z = x + iy$. The quantities x and y are called the real and imaginary parts of z, respectively, denoted $x = \mathcal{R}e\ z$ and $y = \mathcal{I}m\ z$. (Be careful here. It is a common error to include the i in the imaginary part, writing $\mathcal{I}m\ z = iy$. The imaginary part of a complex number is a *real* number.)

A complex number $z = x + iy$ defines a point (x, y) in a plane. It is convenient to think of this point as the tip of a vector extending from the origin (Figure 1.7). The length of this vector, $\sqrt{x^2 + y^2}$, is called the *modulus*, or *magnitude*, of the complex number, and is denoted $|z|$. If z is purely real, then $|z|$ is simply an absolute value. The angle from the real (x) axis to the vector is called the *argument* of the complex number, written $\arg z$. A complex number may be specified either by its real and imaginary parts, $z = x + iy$, or by its modulus and argument, $z = r\angle\theta$, where $r = |z|$ and $\theta = \arg z$. These are known, respectively, as the rectangular (or Cartesian) and polar forms.

By elementary trigonometry, we see $\tan\arg z = y/x$, or $\arg z = \arctan(y/x)$. We must be careful, however, to locate the angle in the proper quadrant of the complex plane. For the complex number $1 + i$, the ratio y/x is 1, and the angle is, by inspection, $\pi/4$. However, the number $-1 - i$, which lives in the third quadrant and has angle $-3\pi/4$, also yields $y/x = (-1)/(-1) = 1$. (Figure 1.8). When asked to calculate an arctangent, your pocket calculator will give a result between $-\pi/2$ and $\pi/2$, the so-called *principal value* of the arctangent. To

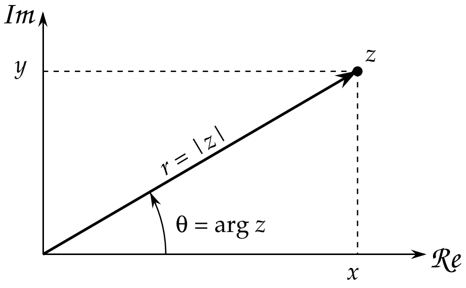

FIGURE 1.7: A complex number may be visualized as a vector in a plane.

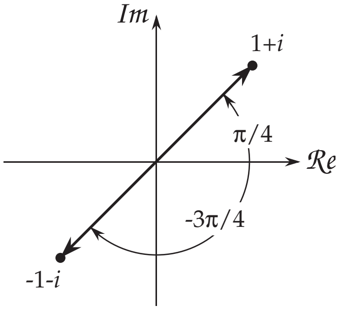

FIGURE 1.8: Although $\arg z = \arctan(\mathcal{I}m\ z/\mathcal{R}e\ z)$, one must be careful to locate the arctangent in the proper quadrant. The numbers $1{+}i$ and $-1{-}i$ both yield $\mathcal{I}m\ z/\mathcal{R}e\ z = 1$, yet their arguments are different.

properly calculate the argument, however, requires an arctangent function that respects the signs of the real and imaginary parts. Many calculators and computer languages have such a function; in MATLAB, for example, it is called `atan2`. MATLAB also has a function called `angle`, which takes a complex number directly, so you don't have to separate the imaginary and real parts. The MATLAB functions are illustrated below.

```
z = [1+i, -1-i];
>atan(imag(z)./real(z))
ans =
    0.7854  0.7854

>atan2(imag(z), real(z))
ans =
    0.7854  -2.3562

>angle(z)
ans =
    0.7854  -2.3562
```

A complex number in polar form, $z = r\angle\theta$, is easily converted to Cartesian form by the equations (see Figure 1.7):

$$\begin{aligned} x &= r\cos\theta \\ y &= r\sin\theta \end{aligned} \tag{1.6}$$

The rectangular form is more convenient for adding and subtracting complex numbers,

$$z_1 + z_2 = (x_1 + x_2) + i(y_1 + y_2) \tag{1.7}$$

while the polar form is better for multiplication and division,

$$z_1\, z_2 = r_1 r_2 \angle(\theta_1 + \theta_2). \tag{1.8}$$

(The proof of this is easy using complex exponentials—see Section 1.7). Both forms have physical interpretations in engineering and science.

The argument is multivalued (see Section 1.5). If $z = r\angle\theta$, then it is also true that $z = r\angle(\theta + 2\pi k)$, where k is any integer. We define the *principal value* of the argument, Arg z, to be the one between $-\pi$ and π. All other values are obtained by adding integer multiples of 2π to the principal value.

$$\begin{aligned} &\text{Arg } z \in (-\pi, \pi] \\ &\arg z = \text{Arg } z + 2\pi k, \quad k = 0,\ \pm 1,\ \pm 2, \ldots \end{aligned} \tag{1.9}$$

The complex number $z^* = x - iy$, obtained by changing i to $-i$, is called the *complex conjugate* of z. The product $zz^* = (x + iy)(x - iy) = x^2 + y^2$ is $|z|^2$, the squared modulus. Because of the change of sign in z^*, $\arg z^* = \arctan[(-y)/x] = -\arg z$. The sum $z + z^* = x + iy + x - iy = 2x$, and the difference $z - z^* = 2iy$. This leads to the relationships:

$$\begin{aligned} x &= \frac{z + z^*}{2} \\ y &= \frac{z - z^*}{2i} \end{aligned} \tag{1.10}$$

These are particularly useful in calculating the real and imaginary parts of complex-valued functions.

It is almost always advisable to simplify a complex fraction by "rationalizing the denominator" with the complex conjugate.

$$\frac{z_1}{z_2} = \frac{z_1 z_2^*}{z_2 z_2^*} = \frac{z_1 z_2^*}{|z_2|^2} \tag{1.11}$$

It follows that

$$\arg(1/z) = \arg z^* = -\arg z\ . \tag{1.12}$$

1.4. Matrix algebra

A matrix is an array of numbers, which may be real or complex.

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1c} \\ x_{21} & x_{22} & \cdots & x_{2c} \\ \vdots & \vdots & \ddots & \vdots \\ x_{r1} & x_{r2} & \cdots & x_{rc} \end{bmatrix}$$

The dimensions of the array are expressed "$r \times c$" (read "r by c"), where r is the the number of rows and c is the number of columns. Particularly important special cases are

- $1 \times n$, called a row vector

$$\begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}$$

- $n \times 1$, called a column vector

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

- $n \times n$, a square matrix

$$\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2c} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{r2} & \cdots & x_{nn} \end{bmatrix}$$

- A diagonal matrix

$$\begin{bmatrix} x_{11} & & & 0 \\ & x_{22} & & \\ & & \ddots & \\ 0 & & & x_{nn} \end{bmatrix}$$

- **I**, the identity matrix

$$\begin{bmatrix} 1 & & & 0 \\ & 1 & & \\ & & \ddots & \\ 0 & & & 1 \end{bmatrix}$$

The familiar three-dimensional vectors from physics (Section 1.2), expressed in terms of orthogonal components relative to a basis, are compactly written as arrays. The following are equivalent representations:

$$\mathbf{v} = v_1\mathbf{e}_1 + v_2\mathbf{e}_2 + v_3\mathbf{e}_3$$

$$\mathbf{v} = \begin{bmatrix} v_1 & v_2 & v_3 \end{bmatrix}$$

$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$$

When a vector is written as an array, the presence of an underlying basis is always understood.

The *transpose* of an array, denoted $\mathbf{X}^T$, is obtained by exchanging the rows and columns:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^T = \begin{bmatrix} a & c \\ b & d \end{bmatrix} .$$

The transpose of a row vector is a column vector, and vice versa. The complex conjugate of an array is made by taking the complex conjugate of each element in the array. The *adjoint* of an array, denoted $\mathbf{X}^+$, is the complex conjugate of the transpose,

$$\mathbf{X}^+ = \left(\mathbf{X}^T\right)^* = \left(\mathbf{X}^*\right)^T$$

$$\begin{bmatrix} 1 & 2i \\ 1-3i & 4 \end{bmatrix}^+ = \begin{bmatrix} 1 & 1+3i \\ -2i & 4 \end{bmatrix}$$

(In MATLAB, the command `x'` computes the adjoint of x. If you just want a transpose without complex conjugation, use instead `x.'`. If x is real, then `x'` = `x.'`.)

Matrix algebra generalizes scalar algebra. Anything which is true for matrices and vectors remains true when all the matrices and vectors are replaced by scalars. On the other hand, the generalization from scalar algebra to matrix algebra must be approached carefully.

Arrays may be added and subtracted if they are of the same dimension:

$$\begin{bmatrix} 1 \\ 2 \end{bmatrix} + \begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 3 \\ 5 \end{bmatrix}$$

is a valid operation, but

$$\begin{bmatrix} 1 \\ 2 \end{bmatrix} + \begin{bmatrix} 2 & 3 \end{bmatrix}$$

is undefined.

An array may be multiplied by a scalar,

$$c\mathbf{x} = c\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} cx_1 \\ cx_2 \\ \vdots \\ cx_n \end{bmatrix} , \qquad c\mathbf{A} = c\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \begin{bmatrix} ca_{11} & ca_{12} \\ ca_{21} & ca_{22} \end{bmatrix}$$

Two arrays may be multiplied if they have compatible dimensions. Let **X**'s dimensions be $r_x \times c_x$ and **Y**'s be $r_y \times c_y$. The product **XY** is of the form

$$\left[\, r_x \times c_x \,\right] \; \left[\, r_y \times c_y \,\right] .$$

The dimensions c_x and r_y are called the inner dimensions. If they are the same, $c_x = r_y$, then the product **XY** may be calculated. The resulting array has the outer dimensions, $r_x \times c_y$. Here's a good way to remember this:

$$\bigl[\, r_x \times \underbrace{c_x \,\bigr] \; \bigl[\, r_y}_{\text{inner}} \times c_y \,\bigr] = \bigl[\, \underbrace{r_x \times c_y}_{\text{outer}} \,\bigr] .$$

In the special case of a row vector $(1 \times n)$ times a column vector $(n \times 1)$, the result is a 1×1 array that we take to be a scalar, by analogy with the idea of the dot product. For two column vectors **x** and **y**, the *inner product* is the scalar defined by

$$\mathbf{x}^{+}\mathbf{y} = x_1^* y_1 + x_2^* y_2 + \cdots + x_n^* y_n \tag{1.13}$$

The norm of a vector is given by its inner product with itself,

$$\|\mathbf{x}\| = \sqrt{\mathbf{x}^{+}\mathbf{x}} = \sqrt{|x_1|^2 + |x_2|^2 + \cdots + |x_n|^2}$$

Notice how these definitions generalize the dot product and norm for real-valued vectors in two- and three-dimensional space.

The product of a matrix and a vector is calculated by repeated row-column products. The r^{th} element of the result is the product of the r^{th} row or column of the matrix and the vector.

$$\begin{bmatrix} \cdot & \cdot & \cdot \\ * & * & * \\ \cdot & \cdot & \cdot \end{bmatrix} \begin{bmatrix} * \\ * \\ * \end{bmatrix} = \begin{bmatrix} \cdot \\ * \\ \cdot \end{bmatrix}$$

or

$$\begin{bmatrix} * & * & * \end{bmatrix} \begin{bmatrix} \cdot & * & \cdot \\ \cdot & * & \cdot \\ \cdot & * & \cdot \end{bmatrix} = \begin{bmatrix} \cdot & * & \cdot \end{bmatrix}$$

The product of two matrices is just more of the same. The product of the r^{th} row and c^{th} column is the (r, c) element of the result.

$$\begin{bmatrix} \cdot & \cdot & \cdot & \cdot \\ * & * & * & * \\ \cdot & \cdot & \cdot & \cdot \end{bmatrix} \begin{bmatrix} \cdot & \cdot & * \\ \cdot & \cdot & * \\ \cdot & \cdot & * \\ \cdot & \cdot & * \end{bmatrix} = \begin{bmatrix} \cdot & \cdot & \cdot \\ \cdot & \cdot & * \\ \cdot & \cdot & \cdot \end{bmatrix}$$

In general, matrix multiplication does not commute: $\mathbf{AB} \neq \mathbf{BA}$.

The identity matrix generalizes the scalar multiplicative identity element, 1. For any array **A** and identity matrices of appropriate dimension, $\mathbf{IA} = \mathbf{A}$ and $\mathbf{AI} = \mathbf{A}$.

An $m \times n$ matrix **A** transforms an n-dimensional vector **x** into an m-dimensional vector **y** through the product $\mathbf{y} = \mathbf{Ax}$. When **A** is square, **x** and **y** have the same dimensions. If, in addition, $\mathbf{y} = \mathbf{Ax} = \lambda\mathbf{x}$, where λ is a complex scalar, we say that **x** is an *eigenvector* of **A**, and λ is the *eigenvalue* associated with **x**. Eigenvalues and eigenvectors have numerous applications in engineering and physics, and there are good numerical methods for computing them (for example, the MATLAB function `eig`).

The combination $\mathbf{x}^+\mathbf{A}\mathbf{x}$, where $\mathbf{A}$ is a square matrix, is called a *quadratic form.* If $\mathbf{A}$ is a diagonal matrix, then

$$\mathbf{x}^+\mathbf{A}\mathbf{x} = a_{11}|x_1|^2 + a_{22}|x_2|^2 + \cdots + a_{nn}|x_n|^2$$

which appears to be an n-dimensional generalization of the simple quadratic $a|x|^2$. The most general quadratic form can also include cross-terms such as $a_{12}x_1^*x_2$. If a is positive in the scalar case, $a|x|^2$ is also positive, for all nonzero x. In the matrix case, if $\mathbf{x}^+\mathbf{A}\mathbf{x} > 0$ for all nonzero $\mathbf{x}$, then $\mathbf{A}$ is said to be *positive definite.*

Matrix division is defined in a very restricted sense. If $\mathbf{A}$ is square, and $\mathbf{Ax} = \mathbf{y}$, then we may be able to solve for $\mathbf{x}$. In the scalar algebraic equation $ax = y$, we can calculate $x = y/a$ if a is nonzero. Otherwise, the division is undefined. In the matrix situation, the *determinant*, denoted $\det \mathbf{A}$ or $|\mathbf{A}|$, must be nonzero. The determinant will be nonzero if the rows of $\mathbf{A}$ are linearly independent (no row can be expressed as a nontrivial linear combination of the other rows). If this is the case, then $\mathbf{Ax} = \mathbf{y}$ represents n simultaneous equations in n unknowns, which have a unique solution. So, if $\det \mathbf{A} \neq 0$, the matrix inverse $\mathbf{A}^{-1}$ may be calculated, and the unique solution to the equation $\mathbf{Ax} = \mathbf{y}$ is $\mathbf{x} = \mathbf{A}^{-1}\mathbf{y}$. Practical computational algorithms for solving $\mathbf{Ax} = \mathbf{y}$ are readily available. A matrix for which $\det \mathbf{A} = 0$ (or, a practical matter, numerically very close to zero) cannot be inverted and is called *singular.*

If $\mathbf{A}$ is nonsingular, then $\mathbf{A}^{-1}$ exists and the products $\mathbf{A}^{-1}\mathbf{A}$ and $\mathbf{A}\mathbf{A}^{-1}$ exist and are both equal to $\mathbf{I}$, the identity matrix (analogous to the scalar case $a \times \dfrac{1}{a} = 1$). If, in addition to being linearly independent, the rows of a square matrix are also orthogonal, then it is called an *orthogonal* matrix. The product of an orthogonal matrix and its adjoint, $\mathbf{A}\mathbf{A}^+$, is a diagonal matrix. If, further, the rows of the matrix are orthonormal, then $\mathbf{A}\mathbf{A}^+ = \mathbf{I}$, and $\mathbf{A}$ is called a *unitary* matrix. For a unitary matrix, $\mathbf{A}^+ = \mathbf{A}^{-1}$.

1.5. Mappings and functions

A *mapping* is a rule which assigns to every point x in a set X a point y in a set Y. The set X is called the *domain* of the mapping. Frequently, we write $f : X \to Y$ to say "f is a mapping from X to Y". We can write $f : \mathbb{R} \to \mathbb{R}$ as shorthand for "f is a real-valued mapping of a real variable", and $f : \mathbb{R} \to \mathbb{C}$ to say "f is a complex-valued mapping of a real variable". For each point $x \in X$, the corresponding point $y \in Y$ is called the *image* of x, and is denoted $y = f(x)$. If each point x in the domain has only one image point, then the mapping f is called a *function.* A function is a *single-valued* mapping. In this text, when a function's domain is a subset of the integers, $X \subseteq \mathbb{Z}$, we will denote the image of n by $f[n]$ rather than $f(n)$. A function whose domain is a set of successive integers, *e.g.*, $\{1, 2, \ldots, N\}$, is also called a *sequence.*

The set of all image points, $f(X) = \{y \in Y \mid y = f(x), x \in X\}$, is called the *range* of f. The range of f is a subset of Y; there may be points in Y which are not "f of something in

X". If a point $y \in Y$ is the image of a point $x \in X$, we call x the *preimage* of y. An *inverse* f^{-1} can be defined using preimages. For a point $y \in Y$, $f^{-1}(y) = \{\text{preimages of } y \text{ in } X\} = \{x \in X \mid f(x) = y\}$. The inverse may or may not be a function, in that y may have more than one preimage.

If every y in the range of f has a unique preimage, we say f is *one-to-one*. If the range of f is identically Y, that is, if every $y \in Y$ has a preimage in X, we say that f is *onto*. If a function is both one-to-one and onto, so that there is a one-to-one correspondence between the points in X and the points in Y, then the inverse $f^{-1} : Y \to X$ is a function, with the property $f^{-1}(f(x)) = x$ and $f\left(f^{-1}(y)\right) = y$ (Figure 1.9).

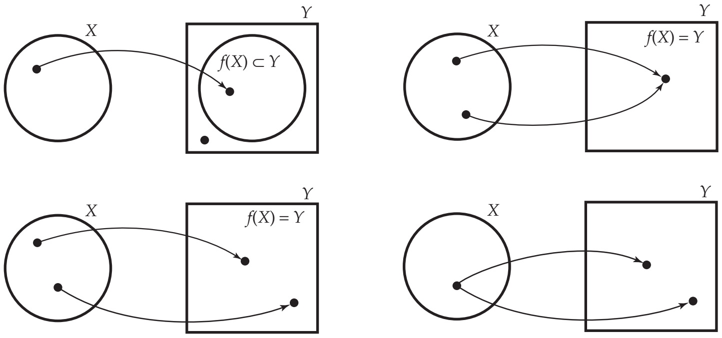

FIGURE 1.9: Mappings f from points in a set X to points in a set Y. Clockwise, from top left: One-to-one, but not onto; Onto, but not one-to-one; Multivalued; One-to-one and onto.

Here are some examples.

- The function $f : \mathbb{R} \to \mathbb{R}$ defined by $y = 2x + 3$ is both one-to-one and onto. The inverse is $x = \dfrac{y-3}{2}$.
- The function $f : \mathbb{R} \to \mathbb{R}$ defined by $y = x^2$ is one-to-one, but not onto, because there is no $x \in \mathbb{R}$ for which $f(x)$ is negative. Each positive real y has two preimages, $+\sqrt{y}$ and $-\sqrt{y}$, the positive and negative square roots of y.
- The logarithm function, $y = \log x$, is undefined for $x = 0$, and real-valued only for $x > 0$. We write $f : \mathbb{R}^+ \to \mathbb{R}$ to indicate that the function maps positive reals into the reals. With this restriction of the domain, the range is all of $\mathbb{R}$, and f is one-to-one and onto. The inverse is $x = \exp\ y$.
- The cosine function, $f : \mathbb{R} \to \mathbb{R}$, defined by $y = \cos x$, is not onto because its range is just the interval $[-1, 1]$ rather than all of $\mathbb{R}$. It is not one-to-one because there are many x values that map to the same y, *e.g.*, $\cos 2\pi k = 1,\ k \in \mathbb{Z}$. If the domain is restricted to the interval $[0, \pi]$, then cosine is one-to-one, but still not onto. If we further restrict the sets so $f : [0, \pi] \to [-1, 1]$, cosine is both one-to-one and onto, and the inverse, arccos, is a well-defined function.

The inverse of $y = x^2$ is not a function because it assigns two values, $\pm\sqrt{y}$. By restricting the range of the square root to either the nonnegative or nonpositive real numbers, we can make it single-valued; these restrictions are called the positive and negative *branches* of the square root. Engineers and scientists, including the author, usually call the square root a *multivalued function* rather than a mapping, although the former term is, strictly speaking, an oxymoron. We shall have much more to say about multivalued functions in a later chapter.

An *even* function is one whose graph is symmetric across the origin, $f(-x) = f(x)$. An *odd* function's graph is antisymmetric across the origin, $f(-x) = -f(x)$. An arbitrary function can be expressed as the sum of an even part and an odd part, $f(x) = f_e(x) + f_o(x)$, where

$$\begin{aligned} f_e(x) &= \frac{f(x) + f(-x)}{2} \\ f_o(x) &= \frac{f(x) - f(-x)}{2} \end{aligned} \tag{1.14}$$

If f is an odd function,

$$\int_{-a}^{a} f(x)dx = 0 \ ,$$

and if f is an even function,

$$\int_{-a}^{a} f(x)dx = 2\int_{0}^{a} f(x)dx.$$

A complex-valued function, $f : \mathbb{R} \to \mathbb{C}$, is *Hermitian* if $f(-x) = f^*(x)$. The real part of a Hermitian function is even, and its imaginary part is odd:

$$\begin{aligned} &f(-x) = f_r(-x) + if_i(-x) = f^*(x) = f_r(x) - if_i(x) \\ &f_r(-x) = f_r(x), \quad f_i(-x) = -f_i(x) \end{aligned}$$

A function f is *bounded above* if there is a finite $M \in \mathbb{R}$ such that $f(x) \le M$ for all x in f's domain. A function is *bounded below* if there is a finite $m \in \mathbb{R}$ such that $f(x) \ge m$ for all x in f's domain. If a function is bounded above, then there is a particular value of M that is the smallest possible upper bound. This value is called the *supremum* of f, denoted $\sup f$. Likewise, if a function is bounded below, then there is greatest lower bound, which is called the *infimum* of f, denoted $\inf f$.

Sometimes the supremum and infimum are just the largest and smallest values that f takes on in its domain. For example, $\sup\ \cos = 1$ and $\inf\ \cos = -1$ on the interval $[-\pi, \pi]$. Other functions never attain their bounds. For example, the "saturating exponential" $f(x) = 1 - \exp(-x)$ approaches one as $x \to \infty$, so $\sup f = 1$. A sequence of numbers can also have a supremum and an infimum. For example, for the sequence $x = \{\frac{1}{2}, \frac{1}{4}, \ldots, \frac{1}{2^n}, \ldots\}$, $\sup x = 1/2$ and $\inf x = 0$. The sequence attains its supremum, but not its infimum.

For a function $f : \mathbb{R} \to \mathbb{C}$ (because $\mathbb{R} \subset \mathbb{C}$, this includes both real- and complex-valued functions), if

$$\lim_{x \to a^+} f(x) = \lim_{x \to a^-} f(x) = f(a),$$

where $x \to a^+$ and $x \to a^-$ mean that x approaches a from above ($x > a$) and below ($x < a$), respectively, we say that f is *continuous* at a. For example, while the parabolic function $f(x) =$

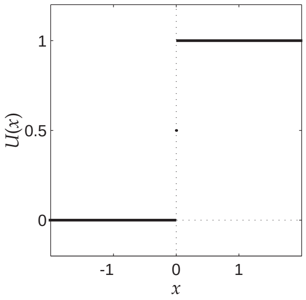

FIGURE 1.10: The unit step function is piecewise continuous. It has a finite jump at the origin.

x^2, is continuous for all finite x, the function

$$g(x) = \begin{cases} 1, & x = 0 \\ x^2, & x \neq 0 \end{cases}$$

is not continuous at $x = 0$. Even though $\lim_{x\to 0} g(x) = 0$ from both sides, the limit is not equal to $g(0)$.

A function $f : \mathbb{R} \to \mathbb{C}$ is said to be *piecewise continuous* if it is continuous everywhere except perhaps on a finite set of points $\{x_k\}_{k=1}^N$, and at these points $f(x_k^-)$ and $f(x_k^+)$ are finite. That is, f is piecewise continuous if it has at most a finite number of finite jump discontinuities. The unit step function $U(x)$, defined

$$U(x) = \begin{cases} 1, & x > 0 \\ \frac{1}{2}, & x = 0 \\ 0, & x < 0 \end{cases} .$$

is piecewise continuous (Figure 1.10).

If f is finite and continuous at x, and the limit

$$\lim_{\Delta x \to 0} \frac{f(x + \Delta x) - f(x)}{\Delta x} \tag{1.15}$$

exists, we say that f is *differentiable* at x. The limit is the derivative, $f'(x)$. For the parabolic function $f(x) = x^2$, which is continuous everywhere,

$$\begin{aligned} \frac{f(x + \Delta x) - f(x)}{\Delta x} &= \frac{x^2 + 2x\Delta x + (\Delta x)^2 - x^2}{\Delta x} \\ &= 2x + \Delta x. \end{aligned}$$

As $\Delta x \to 0$ from above or below, the limit is $2x$, which we recognize as the derivative of x^2.

On the other hand, the step function $U(x)$ is discontinuous at the origin. The limits as the origin is approached from below and above are 0 and 1, respectively. They are not equal to each other, nor are they equal to $U(0)$, which we have defined to be $\frac{1}{2}$. If we attempt to evaluate the limit (1.15) we obtain, on the one hand,

$$\lim_{\Delta x \to 0^-} \frac{0 - \frac{1}{2}}{\Delta x} = \lim_{\Delta x \to 0^-} \frac{-1}{2\Delta x}$$

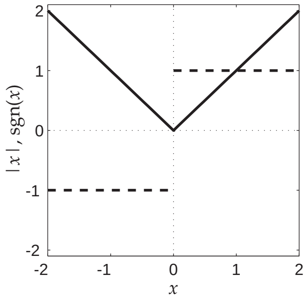

FIGURE 1.11: The absolute value function, $f(x) = |x|$, is piecewise linear and piecewise smooth. The derivative does not exist in the ordinary sense at $x = 0$. If $f'(0)$ is defined to be zero, then $f'(x) = \text{sgn } x$.

and, on the other hand,

$$\lim_{\Delta x \to 0^+} \frac{1 - \frac{1}{2}}{\Delta x} = \lim_{\Delta x \to 0^+} \frac{1}{2\Delta x} ,$$

neither of which exists.

A function $f : \mathbb{R} \to \mathbb{C}$ is *piecewise smooth* if it is piecewise continuous and its derivative f' is also piecewise continuous. The absolute value function, $f(x) = |x|$, is one example. It is everywhere continuous, even at the origin, where it has a "corner" (Figure 1.11). For $x < 0$, the derivative is -1, and for $x > 0$, the derivative is $+1$. The derivative at $x = 0$ does not exist according to the definition given in (1.15). But this is a single point, and because $f'(0-) = -1$ and $f'(0+) = 1$ are finite, the derivative is piecewise continuous.

Even though the derivative of $f(x) = |x|$ is undefined at the origin according to (1.15) we could assign it a value anyway, say $f'(0) = 0$. If you imagine that f is infinitesimally rounded at $x = 0$, this definition even makes intuitive sense. In a later chapter, we will see how to make it precise, so that $f'(x)$ is defined everywhere by the "signum function,"

$$\text{sgn } x = \begin{cases} 1, & x > 0 \\ 0, & x = 0 \\ -1, & x < 0 \end{cases} .$$

1.6. Sinusoidal functions

The basic sinusoid, or "sine wave" function of time is

$$f(t) = \sin \omega t = \sin 2\pi\nu t,$$

where ω is the angular frequency of the wave, expressed in units of radians per second, or rads/sec. Equivalently, we may use the frequency ν, expressed in cycles per second, or Hertz (Hz). The two forms of frequency are related by $\omega = 2\pi\nu$.

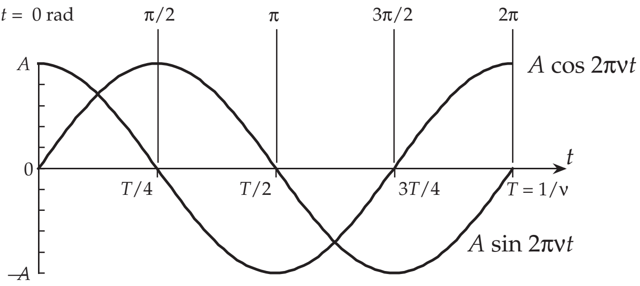

FIGURE 1.12: The sine and cosine functions.

The period, T, of the wave is the amount of time it takes for the wave to go through one complete cycle, *i.e.*, $\omega T = 2\pi\nu T = 2\pi$. The period is related to the frequency by $T = 1/\nu = 2\pi/\omega$.

Among the many trigonometric identities, two of the most useful are the sum formulae,

$$\begin{aligned} \sin(A+B) &= \sin A\,\cos B + \cos A\,\sin B \\ \cos(A+B) &= \cos A\,\cos B - \sin A\,\sin B \end{aligned} \tag{1.16}$$

The cosine function, $f(t) = \cos\omega t = \cos 2\pi\nu t$, is often preferred for modelling sinusoids. Sine and cosine are related by a phase shift of 90°, or $\pi/2$ radians (Figure 1.12):

$$\sin(2\pi\nu t + \pi/2) = \sin 2\pi\nu t\ \ \cos\pi/2 + \cos 2\pi\nu t\ \ \sin\pi/2 = \cos 2\pi\nu t$$

and similarly,

$$\cos(2\pi\nu t + \pi/2) = -\sin 2\pi\nu t\ .$$

A sinusoid of arbitrary amplitude and phase, $C\cos(2\pi\nu t + \phi)$, can always be written as the sum of a pure sine and cosine of the same frequency, but different amplitudes,

$$C\cos(2\pi\nu t + \phi) = A\cos 2\pi\nu t + B\sin 2\pi\nu t. \tag{1.17}$$

Using the formula for the cosine of the sum of two angles,

$$\begin{aligned} C\cos(2\pi\nu t + \phi) &= C(\cos\phi\cos 2\pi\nu t - \sin\phi\sin 2\pi\nu t) \\ &= (C\cos\phi)\cos 2\pi\nu t + (-C\sin\phi)\sin 2\pi\nu t \end{aligned}$$

from which it follows that

$$\begin{aligned} &A = C\cos\phi, \quad B = -C\sin\phi \\ &C = \sqrt{A^2+B^2}, \quad \tan\phi = -B/A \end{aligned} \tag{1.18}$$

This is illustrated in Figure 1.13.

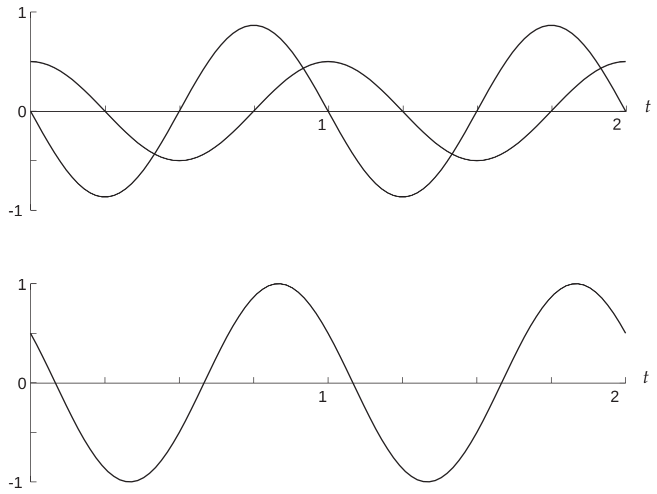

FIGURE 1.13: A cosine of arbitrary amplitude and phase may be formed from the sum of a pure sine and cosine. *Top:* The functions $\frac{1}{2}\cos 2\pi t$ and $-\frac{\sqrt{3}}{2}\sin 2\pi t$. *Bottom:* Their sum, $\cos(2\pi t + \pi/3)$

1.7. Complex exponentials

The sine and cosine are unified in the complex exponential function, $e^{i\theta}$.

$$e^{i\theta} = \cos\theta + i\sin\theta \tag{1.19}$$

The complex conjugate is $e^{-i\theta} = \cos\theta - i\sin\theta$. Combining these two expressions leads to the Euler equations,

$$\begin{aligned} \cos\theta &= \frac{e^{i\theta} + e^{-i\theta}}{2} \\ \sin\theta &= \frac{e^{i\theta} - e^{-i\theta}}{2i} \end{aligned} \tag{1.20}$$

It is often advantageous to reduce the sum and difference of two complex exponentials, $e^{i\theta_1} \pm e^{i\theta_2}$, by creating a symmetric form through factorization:

$$\begin{aligned} e^{i\theta_1} + e^{i\theta_2} &= 2e^{i\frac{\theta_1+\theta_2}{2}} \cos\left(\frac{\theta_1 - \theta_2}{2}\right) \\ e^{i\theta_1} - e^{i\theta_2} &= 2ie^{i\frac{\theta_1+\theta_2}{2}} \sin\left(\frac{\theta_1 - \theta_2}{2}\right) \end{aligned} \tag{1.21}$$

In particular,

$$1 - e^{-i\theta} = 2ie^{-i\theta/2}\sin(\theta/2), \tag{1.22}$$

a result that will come in handy later.

Because the real part of a complex number $z = r\angle\theta$ is $r\cos\theta$ and the imaginary part is $r\sin\theta$, the complex exponential provides another way to express the polar form of a complex number:

$$z = x + iy = r\cos\theta + ir\sin\theta = re^{i\theta}.$$

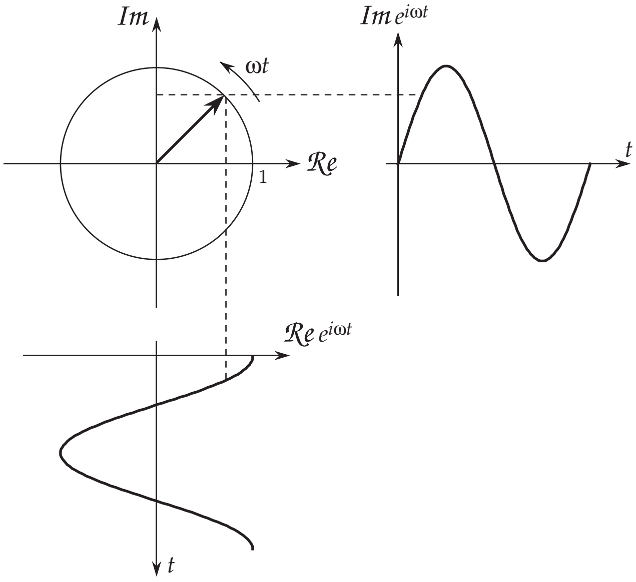

FIGURE 1.14: The complex exponential $e^{i\omega t}$ is a unit vector rotating counterclockwise with angular velocity ω. The projections of this vector along the real and imaginary axes trace out $\cos\omega t$ and $\sin\omega t$, respectively.

It also leads to a useful interpretation of sine and cosine waves. Consider

$$e^{i\omega t} = \cos\omega t + i\sin\omega t.$$

Recalling that the real and imaginary parts of a complex number define a vector in the complex plane, we may interpret $e^{i\omega t}$ as a rotating unit vector. The projection of this vector onto the real axis gives a cosine wave, and the projection onto the imaginary axis gives a sine wave (Figure 1.14).

The complex exponential is useful for solving ordinary differential equations (ODEs). For example, consider the following linear ODE with constant coefficients

$$\frac{dy}{dt} + cy = \cos\omega t\ .$$

The steady-state solution (output of the system) in response to a sinusoidal driving function is a sinusoid having the same frequency, but with a different amplitude and phase, $y(t) = A\cos(\omega t + \phi)$. Following the usual approach, we substitute this into the equation (using the alternative form, $a\cos\omega t + b\sin\omega t$), obtaining:

$$\begin{aligned}\frac{dy}{dt} + cy &= \frac{d}{dt}(a\cos\omega t + b\sin\omega t) + c(a\cos\omega t + b\sin\omega t)\\ &= -\omega a\sin\omega t + \omega b\cos\omega t + ca\cos\omega t + cb\sin\omega t\\ &= (-\omega a + cb)\sin\omega t + (\omega b + ca)\cos\omega t\\ &= \cos\omega t.\end{aligned}$$

From here we solve the simultaneous equations

$$\begin{aligned} -\omega a + cb &= 0 \\ \omega b + ca &= 1 \end{aligned}$$

which yield

$$\begin{aligned} a &= \frac{c}{\omega^2 + c^2} \\ b &= \frac{\omega}{\omega^2 + c^2} \ , \end{aligned}$$

from which we obtain the more useful form (1.18),

$$\begin{aligned} A &= \frac{1}{\sqrt{\omega^2 + c^2}} \\ \tan\phi &= -\omega/c \ . \end{aligned}$$

We will now present a simpler approach using the complex exponential. Were we to drive the equation instead with $\sin\omega t$ instead of $\cos\omega t$, the solution would be $A\sin(\omega t + \phi)$. And, by the linearity property of the differential equation, were we to drive the equation with the complex sum, $\cos\omega t + i\sin\omega t = e^{i\omega t}$, the solution would be $A\cos(\omega t + \phi) + iA\sin(\omega t + \phi) = A\exp\left(i(\omega t + \phi)\right)$. Now, $e^{i\omega t}$ has the convenient property $\dfrac{d}{dt}e^{i\omega t} = i\omega e^{i\omega t}$; in the language of linear algebra, we say that $e^{i\omega t}$ is an *eigenfunction* of the differential operator $\dfrac{d}{dt}$, with eigenvalue $i\omega$.

We drive the differential equation with $e^{i\omega t}$ and assume $y = Ae^{i(\omega t + \phi)}$. The differential equation then becomes a simple algebraic equation:

$$\begin{aligned} \frac{dy}{dt} + cy &= \frac{d}{dt}A\exp\left(i(\omega t + \phi)\right) + cA\exp\left(i(\omega t + \phi)\right) \\ &= i\omega Ae^{i\phi}e^{i\omega t} + cAe^{i\phi}e^{i\omega t} \\ &= e^{i\omega t} \end{aligned}$$

Collecting terms,

$$(i\omega Ae^{i\phi} + cAe^{i\phi} - 1)e^{i\omega t} = 0,$$

and because $|e^{i\omega t}| = 1$, the only way this can be true for all values of time is if

$$i\omega Ae^{i\phi} + cAe^{i\phi} - 1 = 0.$$

The solution of this algebraic equation is

$$Ae^{i\phi} = \frac{1}{c + i\omega} \ .$$

(If you have had a course in dynamic systems, you may recognize this as the transfer function of first-order system like a resistor-capacitor or a mass-damper.)

The amplitude A and phase ϕ of the actual solution, $y(t) = A\cos(\omega t + \phi)$, are given by calculating the modulus and argument of the complex function $\dfrac{1}{c + i\omega}$.

$$\begin{aligned} A &= \left|\frac{1}{c+i\omega}\right| = \left(\frac{1}{(c+i\omega)(c-i\omega)}\right)^{1/2} = \frac{1}{\sqrt{\omega^2+c^2}} \\ \tan\phi &= -\omega/c \end{aligned}$$

The solution may also be written

$$y(t) = \mathcal{R}e\ \left\{Ae^{i\phi}e^{i\omega t}\right\} = \mathcal{R}e\ \left\{\frac{1}{c+i\omega}e^{i\omega t}\right\}.$$

1.8. Geometric series

A series of the form $\sum_{n=0}^{N-1} x^n$ is called a *geometric* series. The sum of this series is an important result which should be committed to memory:

$$\sum_{n=0}^{N-1} x^n = \frac{1-x^N}{1-x} \tag{1.23}$$

It is easily proved (Problem 6). If $|x| < 1$, the series will converge as $N \to \infty$:

$$\sum_{n=0}^{\infty} x^n = \lim_{N\to\infty} \frac{1-x^N}{1-x} = \frac{1}{1-x}, \qquad |x| < 1 \tag{1.24}$$

This resembles the result of the integral,

$$\int_0^T e^{at}dt = \frac{e^{aT}-1}{a},$$

where, if $a < 0$, the limit as $T \to \infty$ exists and is equal to

$$\int_0^\infty e^{at}dt = \lim_{T\to\infty} \frac{e^{aT}-1}{a} = -\frac{1}{a}, \qquad a < 0.$$

1.9. Results from calculus

Here are some more items from elementary calculus which we shall need frequently.

Taylor series

In the vicinity of a point $x = a$, a function $f(x)$ may be approximated by a polynomial, by keeping the first few terms of the function's Taylor series:

$$f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \frac{f'''(a)}{3!}(x-a)^3 + O\{(x-a)^4\},$$

provided, of course, that the necessary derivatives exist at $x = a$. For example, the Taylor series for $\sin x$ about $x = 0$ is $x - x^3/3! + x^5/5! - \cdots$. Keeping the first term, the well-known "small angle approximation" approximates the function by a line, and keeping the next term approximates the sine by a cubic polynomial. Both approximations are tangent to the graph of the sine at the origin. (Figure 1.15).

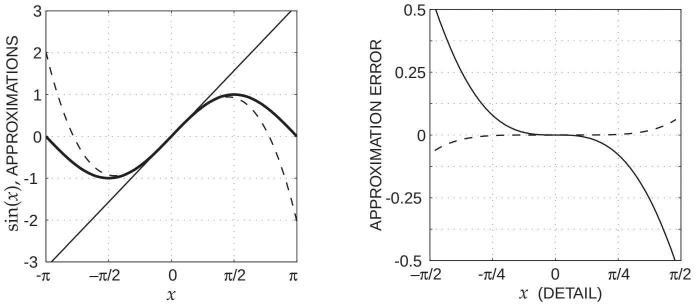

FIGURE 1.15: Low-order polynomial approximations to $\sin x$ created by truncating the Taylor series. *Left:* x (solid line) and $x - x^3/6$ (dashed line). *Right:* Approximation error over the range $\pi/2 > x > -\pi/2$. $\sin(x) - x$ (solid line) and $\sin(x) - (x - x^3/6)$ (dashed line).

L'Hospital's (or L'Hôpital's) rule

Sometimes, when evaluating $\lim\limits_{x\to a} \dfrac{f(x)}{g(x)}$, you find that $\lim\limits_{x\to a} f(x)$ and $\lim\limits_{x\to a} g(x)$ are both zero or both infinite. It is quite possible that the limit of the quotient will still exist, however, if both numerator and denominator are decreasing or increasing at the same rate. If $|\lim\limits_{x\to a} f'(x)| < \infty$ and $0 < |\lim\limits_{x\to a} g'(x)| < \infty$, then

$$\lim_{x\to a} \frac{f(x)}{g(x)} = \frac{\lim\limits_{x\to a} f'(x)}{\lim\limits_{x\to a} g'(x)}.$$

For example,

$$\lim_{x\to 0} \frac{\sin(\pi x)}{\pi x} = \frac{\lim\limits_{x\to 0} \pi\cos(\pi x)}{\lim\limits_{x\to 0} \pi} = 1 \tag{1.25}$$

If you find that both limits are still zero or infinite, keep differentiating until they are different. For example,

$$\begin{aligned}
\lim_{x\to 0} \frac{\sin^2 x}{1-\cos x} = \frac{\lim\limits_{x\to 0} \sin^2 x}{\lim\limits_{x\to 0} 1-\cos x} &= \frac{0}{0} \\
\frac{\lim\limits_{x\to 0} 2\sin x\ \cos x}{\lim\limits_{x\to 0} \sin x} &= \frac{0}{0} \\
\frac{\lim\limits_{x\to 0} 2\cos^2 x - 2\sin^2 x}{\lim\limits_{x\to 0} \cos x} &= \frac{2}{1} = 2
\end{aligned}$$

The Taylor series gives more insight into what is going on with these limits. In the first example, dividing the expansion for $\sin(\pi x)$ by πx gives

$$\frac{\sin(\pi x)}{\pi x} = \frac{\pi x - \dfrac{(\pi x)^3}{6} + \dfrac{(\pi x)^5}{120} - \cdots}{\pi x} = 1 - \frac{(\pi x)^2}{6} + \frac{(\pi x)^4}{120} \cdots$$

This series tends to 1 as $x \to 0$, in agreement with L'Hospital's rule. In the second example, expanding the numerator and denominator of $\dfrac{\sin^2 x}{1-\cos x}$ in their Taylor series gives

$$\frac{\sin^2 x}{1-\cos x} = \frac{x^2 - \dfrac{x^4}{3} + \cdots}{\dfrac{x^2}{2} - \dfrac{x^4}{24} + \cdots} = 2 - \frac{x^2}{3} + \cdots$$

and this goes to 2 as $x \to 0$.

Chain rule

$$\frac{d}{dx} f(u(x)) = f'(u(x)) \frac{du}{dx} = \frac{df}{du}\frac{du}{dx}$$

Integration by parts

Integration by parts is the inverse of the product rule for differentiation, $d(uv) = udv + vdu$.

$$\int udv = uv - \int vdu$$

Differentiating under the integral sign

If $f(x,y)$ is continuous on the region $[a,b] \times [c,d]$, and $\dfrac{\partial f}{\partial y}$ exists and is continuous on this region, then the integral

$$\int_a^b f(x,y)dx$$

is differentiable for all $y \in (c,d)$, and

$$\frac{d}{dy}\int_a^b f(x,y)dx = \int_a^b \frac{\partial f}{\partial y} dx \ .$$

Double integrals

If the function $f(x,y)$ is integrable on the region $[a,b] \times [c,d]$, the function $f(x_0,y)$ is integrable on $[c,d]$ for all $x_0 \in (a,b)$, and the function $f(x,y_0)$ is integrable on $[a,b]$ for all $y_0 \in (c,d)$, then the double integral may be written in terms of single integrals (Fubini's theorem):

$$\int_c^d \int_a^b f(x,y)dxdy = \int_c^d \left[\int_a^b f(x,y)dx\right] dy = \int_a^b \left[\int_c^d f(x,y)dy\right] dx$$

A particularly important form of this, which we shall use later, is:

$$\int_c^d \int_a^b f(x)g(y)h(x,y)dxdy = \int_c^d g(y) \left[\int_a^b f(x)h(x,y)dx\right] dy \tag{1.26a}$$

$$= \int_a^b f(x) \left[\int_c^d g(y)h(x,y)dy\right] dx \tag{1.26b}$$