Chapter 1

Superposition principle

The superposition principle is encountered in many branches of physics and engineering and can be employed to solve essentially all *linear problems*. This principle states that, for linear systems, *the output of a sum of inputs equals the sum of the respective outputs.* This property of linearity will be mathematically defined shortly; for now, we discuss the physical significance of this fundamental principle. Note that the term *input* is quite general: it can refer to a force applied to a mass on a spring, a voltage applied to a RLC circuit, or an optical field impinging on a piece of tissue. Similarly, the *output* can be anything from the displacement of the mass attached to the spring, the transport of charge through a wire, or the optical field scattered by the tissue. By *system* we understand the mechanism that transforms the input into output, e.g. a mass-spring ensemble, RLC circuit, or the tissue in the examples above. *The essential consequence of the superposition principle is that the solution (output) to a complicated problem (input) can be obtained by solving a number of simpler problems, the results of which can be summed up at the end.*

Figure 1.1 illustrates this idea with an example of two optical fields interacting simultaneously with a medium (system). In order to find the response to applying the two fields through the system, we have two choices: (i) add the two inputs $U_1 + U_2$ and solve for the output; (ii) find the individual outputs, U'_1, U'_2 and add them up, $U'_1 + U'_2$. Of course, it is the second option that relies on the principle of superposition. It is not clear *a priori* which of the two approaches provides a more direct access to the solution $(U'_1 + U'_2)$. However, the superposition principle allows us to decompose U_1 and U_2 into yet simpler signals, for which the solutions can be easily found. In the following we discuss two very common such decompositions that allow us to solve complicated (but linear!) problems very efficiently: *Green's* method and *Fourier's* method. These decompositions are employed all the time in solving optics problems.

doi:10.1088/2053-2563/aae121ch1

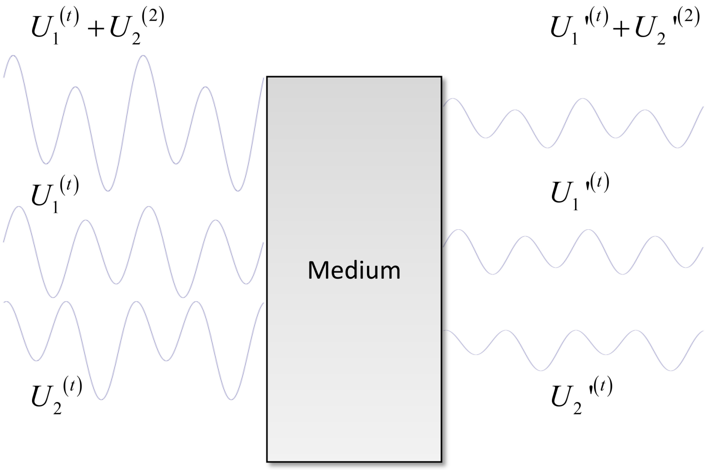


Figure 1.1. The superposition principle. The response of the system (e.g. a piece of glass) to the sum of two fields, $U_1 + U_2$, is the same as summing the individual outputs, $U_1' + U_2'$.

1.1 Green's function method

Green's method of solving linear problems refers to 'breaking down' the input signal into a succession of pulses that are infinitely thin, mathematically expressed by Dirac delta distributions. Figure 1.2 illustrates δ-functions in 1D, 2D, and 3D. Superficially, we can think of the Dirac delta distribution as a function, which is zero everywhere except at the origin, where it is infinite,

$$\delta(x) = \begin{cases} \infty, & x = 0 \\ 0, & x \neq 0, \end{cases}$$

where δ satisfies the normalization to unit area,

$$\int_{-\infty}^{\infty} \delta(x)dx = 1.$$

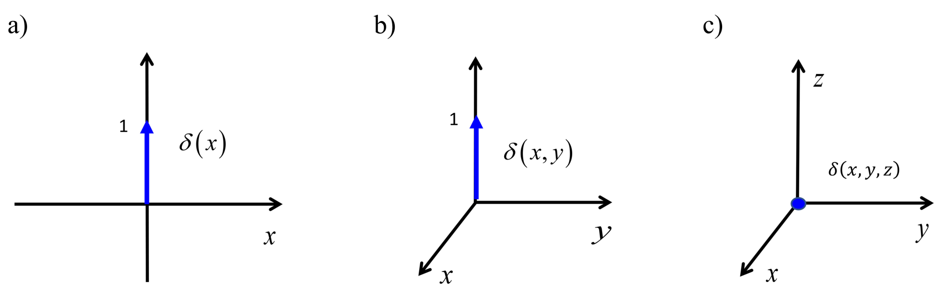


Figure 1.2. Delta function in 1D (a), 2D (b) and 3D (c).

Typically, these distributions are illustrated in 1D and 2D by arrows of unit length, to suggest their infinitely thin width and unit normalization (figure 1.2(a)). Note that representing $\delta(x,y,z)$ in 3D would require a 4D plot. Thus, we illustrate the 3D δ-function by a dot, keeping in mind that its amplitude is 1.

The Dirac delta is more rigorously described as a distribution whose properties are revealed when integrated against other functions. For example, for an arbitrary function f, the following property holds:

$$\int_{-\infty}^{\infty} \delta(x) f(x) dx = f(0).$$

Another related property of the Dirac delta is that, when multiplied by a function it 'samples' the function at the position of the delta,

$$f(x)\delta(x - a) = f(a)\delta(x - a).$$

This property will prove very useful in expressing complicated input signals as a sequence of delta impulses.

Throughout the book, we will deal with temporal signals, spatial signals, or a combination of the two. Like in most references, we will sometimes refer loosely to the delta distribution as 'function', keeping in mind the understanding above. Figure 1.3 illustrates how an arbitrary temporal (figure 1.3(a)) and spatial (figures 1.3(b) and (c)) input can be described as an ensemble of pulses. Using the basic property of δ-functions, the signal in figure 1.3(a) can be written as

$$\begin{aligned} U(t) &= \int_{-\infty}^{\infty} U(t')\delta(t - t')dt' \\ &= U(t) \circledcirc \delta(t) \end{aligned} \tag{1.1}$$

which defines $U(t)$ as a summation over infinitely short pulses, each characterized by a position in time, $t - t'$, and amplitude, $U(t')$. The type of integral in equation (1.1) is called *convolution*, which we are denoting by symbol $\circledcirc$. If the delta function is shifted to a certain coordinate, a, the result of equation (1.1) will also be shifted, i.e. $U(t) \circledcirc \delta(t - a) = U(t - a)$.

Similarly, this decomposition is applied to 2D and 3D signals (figures 1.3(b) and (c)). Exploiting the superposition principle, the response to this temporal signal can be obtained by finding the response to each impulse and summing up the results. This type of problem is useful in dealing with, for instance, the propagation of light pulses through various media.

Similarly, the response to the *spatial* 2D input, $U(x, y)$, shown in figure 1.3(b), can be obtained by solving the problem for each impulse and adding up the results, because U can be written as

$$U(x, y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} U(x', y')\delta(x - x', y - y')dx'dy' \ . \tag{1.2}$$

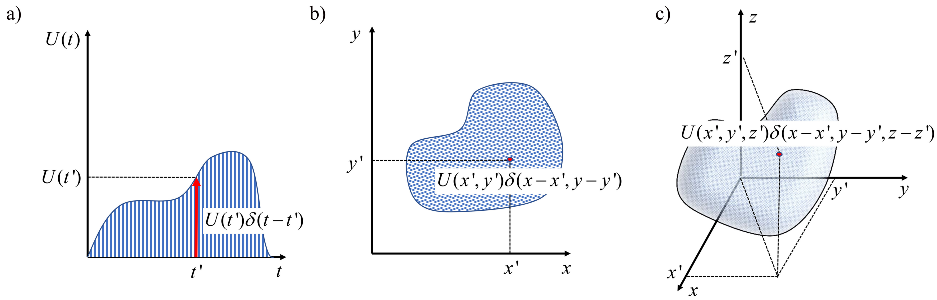


Figure 1.3. 1D (a), 2D (b), and 3D (c) signals can be described as an ensemble of impulses. The delta functions have their amplitudes equal to the signal evaluated at the position of the delta function, namely ($U(t')$), ($U(x',y')$), ($U(x', y',z')$).

This type of input is encountered often in problems related to imaging. Of course, the 3D case is similar (figure 1.3(c)),

$$U(x, y, z) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} U(x', y', z')\delta(x - x', y - y', z - z')dx'dy'dz' \quad . \quad (1.3)$$

Green's method is extremely powerful because solving linear problems with an 'impulse' input is typically an easy task. The response to such an impulse is called *Green's function* or the *impulse response* of the linear system. This fundamental property of linear systems will be described mathematically in more detail in chapters 2 and 3 and Green's method for solving linear problems will be used broadly throughout the book.

1.2 Fourier transform method

Another efficient way of decomposing an input into simpler bits is to break it down into sinusoidal signals of suitable frequencies, amplitudes, and phases. Essentially, any curve can be reconstructed by summing up such sine waves, as illustrated for both temporal and spatial input signals in figure 1.4. Again, the main advantage of such a decomposition is that solving a linear problem for a single sinusoid as input is a simple task. Thus, the output is simply the summation of all responses associated with these sinusoids, which typically can be calculated easily. The signals illustrated in figure 1.4 are real, i.e. the inputs are reconstructed from a summation of *cosine* signals. The *Fourier decomposition* of a signal is the generalization of this concept whereby a signal, which generally can be complex, is decomposed in a series of elements of the form $A(\omega)e^{-i\omega t}$ (for temporal signals), with ω the temporal angular frequency, and $A(\mathbf{k})e^{i\mathbf{k}\cdot\mathbf{r}}$ (for spatial signals), with $\mathbf{k} = (k_x, k_y, k_z)$ the angular spatial frequency. The superposition principle allows us to express an arbitrary input into a summation of sines and cosines, solve for the output of each sinusoidal and add up the results. We will see that, because the FT of a complex exponential is a δ-function,

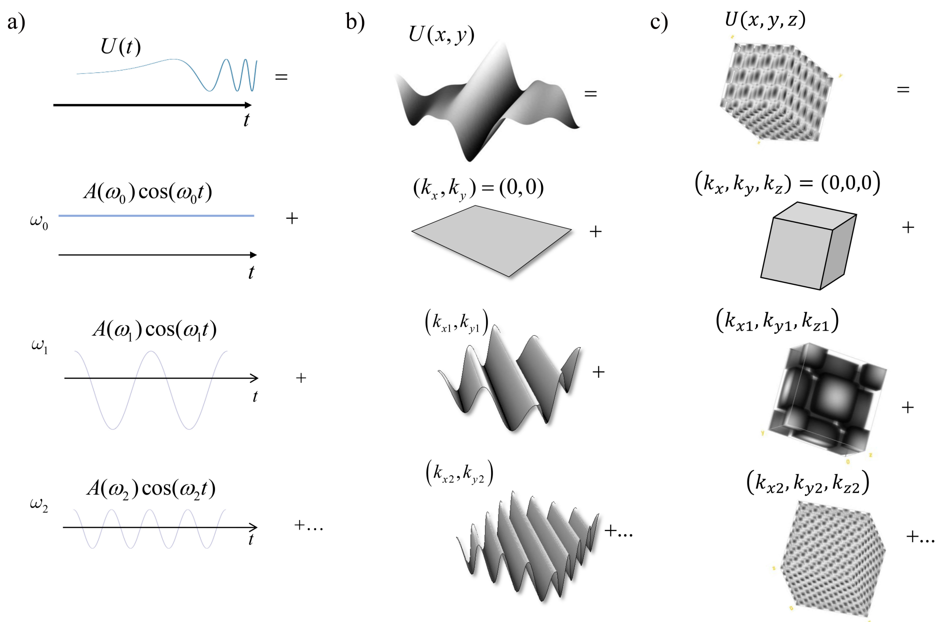


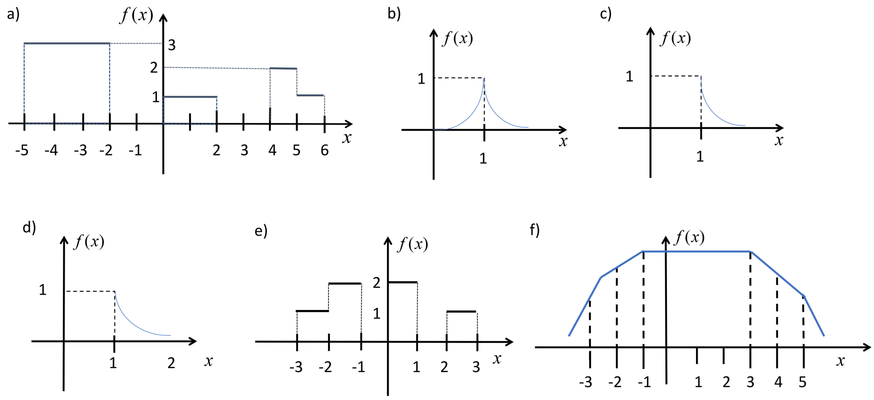
Figure 1.4. A signal can be decomposed into a sum of sinusoids: 1D (a), 2D (b), 3D (c).

the Green's and Fourier decomposition methods are in fact related. We will discuss much more on this subject in chapter 3, once we study linear systems in more depth.

1.3 Problems

1. Express the functions below in terms of convolutions with δ-functions. Write the results using the following symbols: Π (rectangular function), Γ (Heaviside step function), Λ (triangle function), sign, sinc, chapter 4. Plot all these functions.
   a) $\Pi(x - 5)$
   b) $\Pi(\frac{3x}{2} + 3) - \Pi(\frac{5x}{2} - 2)$
   c) $\Gamma(x - 2) + sinc(\frac{2x}{3} + 5)$
   d) $\sum_{n=1}^{5} \Lambda\left(\frac{x}{2} - 2n\right)$
   e) $sign(x - 7)$.

2. Use our typical symbols to describe the following signals.



3. Prove the *sampling property* of the delta function,

$$f(x)\delta(x - a) = f(a)\delta(x - a) \ .$$

4. Calculate the following expressions.
   a) $(x^3 + 2x + 4)\delta(x)$
   b) $(\ln(x^2) + e^x + 1)\delta(x - 1)$
   c) $\left(\dfrac{x^2 + 1}{x^2 - 1}\right)^{\ln(x)} \delta(x - 3)$
5. Calculate the following integrals.
   a) $\int_{-\infty}^{\infty} \delta(x)[x^2 + x + 1]dx$
   b) $\int_{-\infty}^{\infty} \delta(x)[e^{ixb}]dx$, *b real constant*
   c) $\int_{-\infty}^{\infty} \delta(x - 5)[x^2 + x + 1]dx$
   d) $\int_{-\infty}^{\infty} \delta(x - 5)e^{ixb}dx$, *b real constant*

Further reading

[1] Feynman R P, Leighton R B and Sands M L 1963 *The Feynman Lectures on Physics* (Reading, MA: Addison-Wesley)