CHAPTER 4

Linear systems

4.1 Introduction

The word system is defined [Hanks] as 'a group or combination of inter-related, inter-dependent, or interacting elements forming a collective entity; ...'. In the context of a digital communications system the interacting elements, for example electronic amplifiers, mixers, detectors etc., are themselves subsystems made up of components such as resistors, capacitors and transistors. An understanding of how systems behave, and are described, is therefore important to the analysis of electronic communications equipment.

This chapter reviews the properties of the most analytically tractable, but also most important, class of system (i.e. linear systems) and applies concepts developed in Chapters 2 and 3 to them. In particular convolution is used to provide a time domain description of the effect of a system on a signal and the convolution theorem is used to link this to the equivalent description in the frequency domain. Towards the end of the chapter the effect of *memoryless non*-linear systems on the pdf of signals and noise is briefly discussed.

4.2 Linear systems

Linear systems constitute one, restricted, class of system. Electronic communications equipment is predominantly composed of interconnected linear subsystems.

4.2.1 Properties of linear systems

Before becoming involved in the mathematical description of linear systems there are two important questions which should be answered:

1. What is a linear system?
2. Why are linear systems so important?

In answering the first question it is almost as important to say what a linear system is not, as to say what it is. Figure 4.1 shows the input/output characteristic of a system which is specified mathematically by the straight line equation:

$$y(t) = mx(t) + C \tag{4.1}$$

This system, perhaps surprisingly, is non-linear (providing $C \neq 0$). A definition of a linear system can be given as follows:

A system is linear if its response to the sum of any two inputs is the sum of its responses to each of the inputs alone.

This property is usually called the principle of *superposition* since responses to component inputs are superposed at the output. In this context linearity and superposition are synonymous. If $x_i(t)$ are inputs to a system and $y_i(t)$ are the corresponding outputs then superposition can be expressed mathematically as:

$$y(t) = \sum_i y_i(t) \tag{4.2(a)}$$

when:

$$x(t) = \sum_i x_i(t) \tag{4.2(b)}$$

Proportionality (also called homogeneity) is a property which follows directly from linearity. It is defined by:

$$y(t) = my_1(t) \tag{4.3(a)}$$

when:

$$x(t) = mx_1(t) \tag{4.3(b)}$$

The system described by Figure 4.1 and equation (4.1) would have this property if $C = 0$ and in this special case is, therefore, linear. For $C \neq 0$, however, the system does not obey proportionality and therefore cannot be linear. (Equations (4.3) represent a necessary and sufficient condition for linearity providing the system is *memoryless* i.e. its instantaneous output depends only on its instantaneous input.) A further property which systems often

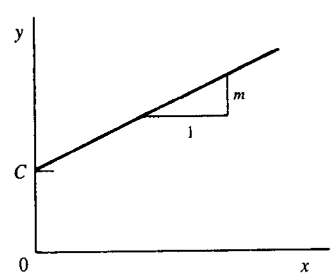


Figure 4.1 *Input/output characteristic of a non-linear system.*

have is *time invariance*. This means that the output of a system does not depend on when the input is applied (except in so far as its location in time). More precisely time invariance can be defined by:

$$y(t) = y_1(t-T) \tag{4.4(a)}$$

when:

$$x(t) = x_1(t-T) \tag{4.4(b)}$$

The majority of communications subsystems obey both equations (4.2) and (4.4) and are therefore called *time invariant linear systems* (TILS).

Time invariant linear systems can be defined using a single formula which also explicitly recognises proportionality, i.e.:

$$\text{If} \quad y_1(t) = S\{x_1(t)\} \quad \text{and} \quad y_2(t) = S\{x_2(t)\} \tag{4.5(a)}$$

$$\text{then} \quad S\{ax_1(t-T) + bx_2(t-T)\} = ay_1(t-T) + by_2(t-T) \tag{4.5(b)}$$

where $S\{\ \}$ represents the functional operation of the system.

4.2.2 Importance of linear systems

The importance of linear systems in engineering cannot be overstated. It is interesting to note, however, that (like periodic signals) linear systems constitute a conceptual ideal that cannot be strictly realised in practice. This is because any device behaves non-linearly if excited by signals of large enough amplitude. An obvious example of this in electronics is the transistor amplifier which saturates when the amplitude of the output approaches the power supply rail voltages (Figure 4.2). Such an amplifier is at least approximately linear, however, over its normal operating range. It is ironic, therefore, that whilst no systems are linear if driven by large enough signals many non-linear systems are at least approximately linear when driven by small enough signals. This is because the transfer characteristic of a non-linear (memoryless) system can normally be represented by a polynomial of the form:

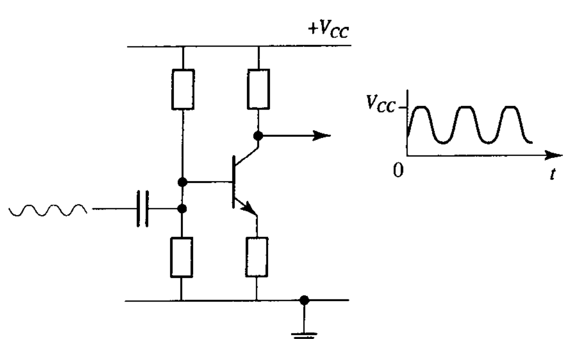


Figure 4.2 *Non-linear behaviour of a simple transistor amplifier.*

$$y(t) = ax(t) + bx^2(t) + cx^3(t) + \cdots \tag{4.5(c)}$$

For small enough input signals (and providing $a \neq 0$) only the first term in equation (4.5(c)) is significant and the system therefore behaves linearly.

An important property of linear systems is that they respond to sinusoidal inputs with sinusoidal outputs of the same frequency (i.e. they conserve the shape of sinusoidal signals).

Other compelling reasons for studying and using linear systems are that:

1. The electric and magnetic properties of free space are linear, i.e.:

$$\mathbf{D} = \varepsilon_o \mathbf{E} \quad (\mathrm{C\,m^{-2}}) \tag{4.6(a)}$$

$$\mathbf{B} = \mu_o \mathbf{H} \quad (\mathrm{Wb\,m^{-2}}) \tag{4.6(b)}$$

   (Since free space is memoryless, proportionality is sufficient to imply linearity.)
2. The electric and magnetic properties of many materials are linear over a large range of field strengths, i.e.:

$$\mathbf{D} = \varepsilon_r \varepsilon_o \mathbf{E} \quad (\mathrm{C\,m^{-2}}) \tag{4.7(a)}$$

$$\mathbf{B} = \mu_r \mu_o \mathbf{H} \quad (\mathrm{Wb\,m^{-2}}) \tag{4.7(b)}$$

$$\mathbf{J} = \sigma \mathbf{E} \quad (\mathrm{A/m^{-2}}) \tag{4.7(c)}$$

   where ε_r, μ_r and σ are constants. (There are notable exceptions to this, of course, e.g. ferromagnetic materials.)
3. Many general mathematical techniques are available for describing, analysing and synthesising linear systems. This is in contrast to non-linear systems for which few, if any, general techniques exist.

EXAMPLE 4.1

Demonstrate the linearity or otherwise of the systems represented by the diagrams in Figure 4.3.

(a) For input $x_1(t)$ output is $y_1(t) = x_1(t) + f(t)$
For input $x_2(t)$ output is $y_2(t) = x_2(t) + f(t)$
For input $x_1(t) + x_2(t)$ output is $x_1(t) + x_2(t) + f(t) \neq y_1(t) + y_2(t)$
i.e. superposition does not hold and system (a) is, therefore, not linear.

(b) For input $x(t) = x_1(t) + x_2(t)$ the output $y(t)$ is:

$$y(t) = f(t)[x_1(t) + x_2(t)]$$

$$= f(t)x_1(t) + f(t)x_2(t) = y_1(t) + y_2(t)$$

which is the superposition of the outputs due to $x_1(t)$ and $x_2(t)$ alone. System (b) is, therefore, linear.

(c) For input $x(t) = x_1(t) + x_2(t)$ output $y(t)$ is:

$$y(t) = \frac{d}{dt}\left[x_1(t) + x_2(t)\right]$$

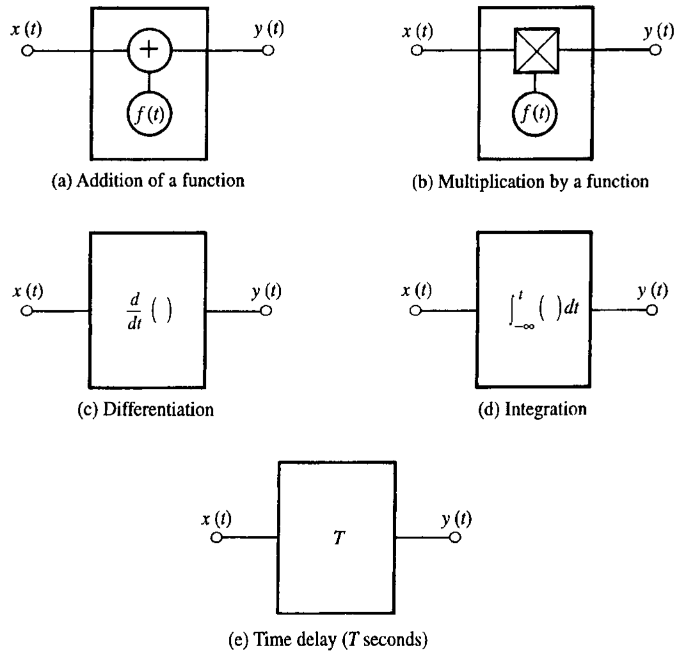


Figure 4.3 *Systems referred to in Example 4.1.*

$$= \frac{d}{dt}\, x_1(t) + \frac{d}{dt}\, x_2(t)$$

$$= y_1(t) + y_2(t)$$

i.e. system (c) is linear.

(d) For input $x(t) = x_1(t) + x_2(t)$ output $y(t)$ is:

$$y(t) = \int_{-\infty}^{t} [x_1(t') + x_2(t')]\, dt'$$

$$= \int_{-\infty}^{t} x_1(t')\, dt' + \int_{-\infty}^{t} x_2(t')\, dt' = y_1(t) + y_2(t)$$

i.e. system (d) is linear.

(e) For input $x(t) = x_1(t) + x_2(t)$ output is:

$$y(t) = x_1(t - T) + x_2(t - T)$$

$$= y_1(t) + y_2(t)$$

i.e. system (e) is linear.

There is an apparent paradox involved in the consideration of the linearity of additive and multiplicative systems. This is that the operation of addition is, by definition, linear, i.e. if two inputs are added the output is the sum of each input by itself. The reason system (a) is not linear is that $f(t)$ is considered to be part of the system, not an input. Conversely multiplication is a non-linear operation in the sense that if two inputs are multiplied then the output is not the sum of the outputs due to each input alone (which would both be zero since each input alone would be multiplied by zero). When $f(t)$ in Figure 4.3(b) is considered to be part of the system, however, superposition holds and the system is therefore linear.

Another point to note is that systems must be either non-linear or time varying (or both) in order to generate frequency components at the output which do not appear at the input. Multiplying by $f(t)$ in Figure 4.3(b) will result in new frequencies at the output providing $f(t)$ is not a constant. This is because in this case we have a time varying linear system. If $f(t)$ = constant then system (b) is a time invariant linear system (in fact a linear amplifier) and no new frequencies are generated.

4.3 Time domain description of linear systems

Just as signals can be described in either the time or frequency domain, so too can systems. In this section time domain descriptions are addressed and the close relationship between linear systems and linear equations is demonstrated.

4.3.1 Linear differential equations

Any system which can be described by a linear differential equation, of the form:

$$\begin{aligned} a_0 y + a_1 \frac{dy}{dt} + a_2 \frac{d^2y}{dt^2} + \cdots + a_{N-1} \frac{d^{N-1}y}{dt^{N-1}} \\ = b_0 x + b_1 \frac{dx}{dt} + b_2 \frac{d^2x}{dt^2} + \cdots + b_{M-1} \frac{d^{M-1}x}{dt^{M-1}} \end{aligned} \tag{4.8}$$

always obeys the principle of superposition and is therefore linear. If the coefficients a_i and b_i are constants then the system is also time invariant. The response of such a system to an input can be defined in terms of two components. One component, the *free response*, is the output, $y_{free}(t)$, when the input (or forcing function) $x(t) = 0$. (Since $x(t)$ is zero for all t then all the derivatives $d^n x(t)/dt^n$ are also zero.) The free response is therefore the solution of the homogeneous equation:

$$a_0 y + a_1 \frac{dy}{dt} + a_2 \frac{d^2y}{dt^2} + \cdots a_{N-1} \frac{d^{N-1}y}{dt^{N-1}} = 0 \tag{4.9}$$

subject to the value of the output, and its derivatives, at $t = 0$, i.e.:

$$y(0), \frac{dy}{dt}\Big|_{t=0}, \frac{d^2y}{dt^2}\Big|_{t=0}, \ldots, \frac{d^{N-1}y}{dt^{N-1}}\Big|_{t=0}$$

These values are called the *initial conditions*. The second component, the *forced response*, is the output, $y_{forced}(t)$, when the input, $x(t)$, is applied but the initial conditions are set to zero, i.e. it is the solution of equation (4.8) when:

$$y(0) = \frac{dy}{dt}\Big|_{t=0} = \frac{d^2y}{dt^2}\Big|_{t=0} = \cdots = \frac{d^{N-1}y}{dt^{N-1}}\Big|_{t=0} = 0 \tag{4.10}$$

The total response of the system (unsurprisingly, since superposition holds) is the sum of the free and forced responses, i.e.:

$$y(t) = y_{free}(t) + y_{forced}(t) \tag{4.11}$$

An alternative decomposition of the response of a linear system is in terms of its steady state and transient responses. The steady state response is that component of $y(t)$ which does not decay (i.e. tend to zero) as $t \to \infty$. The transient response is that component of $y(t)$ which does decay as $t \to \infty$, i.e.:

$$y(t) = y_{steady}(t) + y_{transient}(t) \tag{4.12}$$

4.3.2 Discrete signals and matrix algebra

Consider a linear system with discrete (or sampled) input $x_1, x_2, x_3, \cdots, x_N$ and discrete output $y_1, y_2, y_3, \cdots, y_M$ as shown in Figure 4.4. Each output is then given by a weighted sum of all the inputs [Spiegel]:

$$\begin{bmatrix} y_1 \\ y_2 \\ .. \\ .. \\ y_M \end{bmatrix} = \begin{bmatrix} G_{11} & G_{12} & \cdots & G_{1N} \\ G_{21} & G_{22} & \cdots & G_{2N} \\ .. & .. & \cdots & .. \\ .. & .. & \cdots & .. \\ G_{M1} & .. & \cdots & G_{MN} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ .. \\ .. \\ x_N \end{bmatrix} \tag{4.13}$$

i.e.:

$$y_i = \sum_{j=1}^{N} G_{ij} x_j \tag{4.14}$$

(If the system is a physical system operating in real time then $G_{ij} = 0$ for all values of x_j occurring after y_i.)

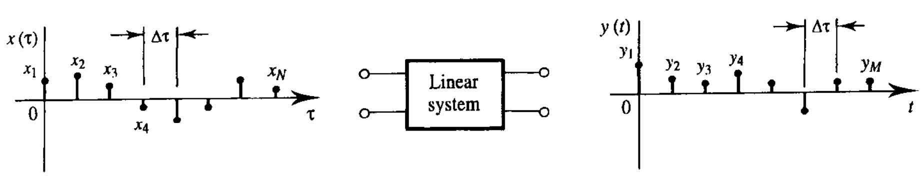


Figure 4.4 *Linear systems with discrete input and output.*

4.3.3 Continuous signals, convolution and impulse response

If the discrete input and output of equation (4.14) are replaced with continuous equivalents, i.e.:

$$y_i \rightarrow y(t)$$

$$x_j \rightarrow x(\tau)$$

(the reason for keeping the input and output variables, τ and t, separate will become clear later) then the discrete summation becomes continuous integration giving:

$$y(t) = \int_0^{N\Delta\tau} G(t,\tau)x(\tau)\, d\tau \tag{4.15}$$

The limits of integration in equation (4.15) assume that x_1 occurs at $\tau = 0$, and the N input samples are spaced by $\Delta\tau$ seconds. Once again, for physical systems operating in real time, it is obvious that future values of input do not contribute to current, or past, values of output. The upper limit in the integral of equation (4.15) can therefore be replaced by t without altering its value, i.e.:

$$y(t) = \int_0^{t} G(t,\tau)x(\tau)\, d\tau \tag{4.16}$$

Furthermore, if input signals are allowed which start at a time arbitrarily distant in the past then:

$$y(t) = \int_{-\infty}^{t} G(t,\tau)x(\tau)\, d\tau \tag{4.17}$$

Systems described by equations (4.16) and (4.17) are called *causal* since only past and current input values affect (or cause) outputs. Equations (4.15) to (4.17) are all examples of integral transforms (of $x(\tau)$) in which $G(t,\tau)$ is the transform kernel. Replacing the input to the system described by equation (4.17) with a (unit strength) impulse, $\delta(\tau)$, results in:

$$h(t) = \int_{-\infty}^{t} G(t,\tau)\delta(\tau)\, d\tau \tag{4.18}$$

(The symbol $h(t)$ is traditionally used to represent a system's impulse response.) If the impulse is applied at time $\tau = T$ then, assuming the system is time invariant, the output will be:

$$h(t-T) = \int_{-\infty}^{t} G(t,\tau)\delta(\tau - T)\, d\tau \tag{4.19}$$

The sampling property of $\delta(\tau - T)$ under integration means that $G(t,T)$ can be interpreted as the response to an impulse applied at time $\tau = T$, i.e.:

$$h(t-T) = G(t,T) \tag{4.20}$$

the surface $G(t,\tau)$ therefore represents the responses for impulses applied at all possible times (Figure 4.5). Replacing T with τ in equation (4.20) (which is a change of notation only), and substituting into equation (4.17) gives:

$$y(t) = \int_{-\infty}^{t} h(t-\tau)x(\tau)\,d\tau \tag{4.21}$$

If non-causal systems are allowed then equation (4.21) is rewritten as:

$$y(t) = \int_{-\infty}^{\infty} h(t-\tau)x(\tau)\,d\tau \tag{4.22}$$

Equations (4.21) and (4.22) can be recognised as convolution, or superposition, integrals. The output of a time invariant linear system is therefore given by the convolution of the system's input with its impulse response, i.e.:

$$y(t) = h(t) * x(t) \tag{4.23}$$

Note that the commutative property of convolution means that equations (4.22) and (4.23) can also be written as:

$$\begin{aligned} y(t) &= x(t) * h(t) \\ &= \int_{-\infty}^{\infty} h(\tau)x(t-\tau)\,d\tau \end{aligned} \tag{4.24}$$

Note also that equations (4.21) to (4.24) are consistent with the definition of an impulse response since in this case:

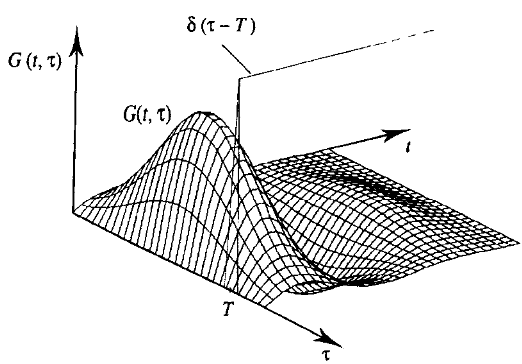


Figure 4.5 *$G(t,\tau)$ for a hypothetical system. Response for an impulse applied at time $\tau = T$ is the curve formed by the intersection of $G(t,\tau)$ with the plane containing $\delta(\tau - T)$.*

$$y(t) = h(t) * \delta(t) = h(t) \tag{4.25}$$

4.3.4 Physical interpretation of $y(t) = h(t) * x(t)$

The input signal $x(t)$ can be considered to consist of many closely spaced impulses, each impulse having a strength, or weight, equal to the value of $x(t)$ at the time the impulse occurs times the impulse spacing. The output is then simply the sum (i.e. superposition) of the responses to all the weighted impulses. This idea is illustrated schematically in Figure 4.6. It essentially represents a decomposition of $x(t)$ into a set of (orthogonal) impulse functions. Each impulse function is operated on by the system to give a (weighted, time shifted) impulse response and the entire set of impulse responses is then summed to give the (reconstituted) response of the system to the entire input signal. In this context equation (4.22) can be reinterpreted as:

$$y(t) = \int_{-\infty}^{\infty} h(t - \tau)[x(\tau)\, d\tau] \tag{4.26}$$

where $[x(\tau)d\tau]$ is the weight of the impulse occurring at the input at time τ and $h(t - \tau)$ is the 'fractional' value to which $[x(\tau)d\tau]$ has decayed at the system output by time t (i.e. $t - \tau$ seconds after the impulse occurred at the input). As always, for causal systems, the upper limit in equation (4.26) could be replaced by t corresponding to the condition (see Figure 4.7):

$$h(t - \tau) = 0, \qquad \text{for } t < \tau \tag{4.27}$$

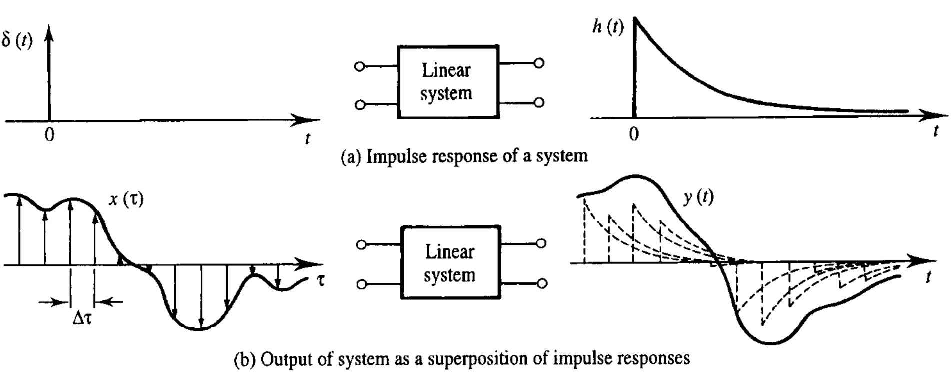


Figure 4.6 *Decomposition of input into (orthogonal) impulse functions and output formed as a sum of weighted impulse responses.*

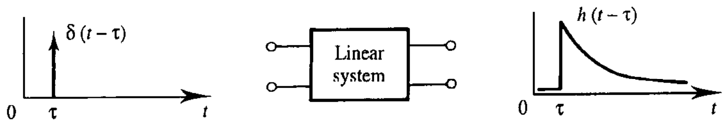


Figure 4.7 *Causal impulse response of a baseband system.*

EXAMPLE 4.2

Find the output of a system having a rectangular impulse response (amplitude A volts, width τ seconds) when driven by an identical rectangular input signal.

Figure 4.8 shows the evolution of the output as $h(t-\tau)$ moves through several different values of t. The result is a triangular function. For a discretely sampled input signal using the standard z-transform notation [Mulgrew and Grant], where $h(n) = x(n) = A + Az^{-1} + Az^{-2} + Az^{-3}$ the output signal is discretely sampled with values $A^2z^{-1} + 2A^2z^{-2} + 3A^2z^{-3} + 4A^2z^{-4} + 3A^2z^{-5}$, etc. Note in this example that, as the impulse response is symmetrical, time reversal of $h(\tau)$ to form $h(-\tau)$ produces a simple shift along the τ-axis.

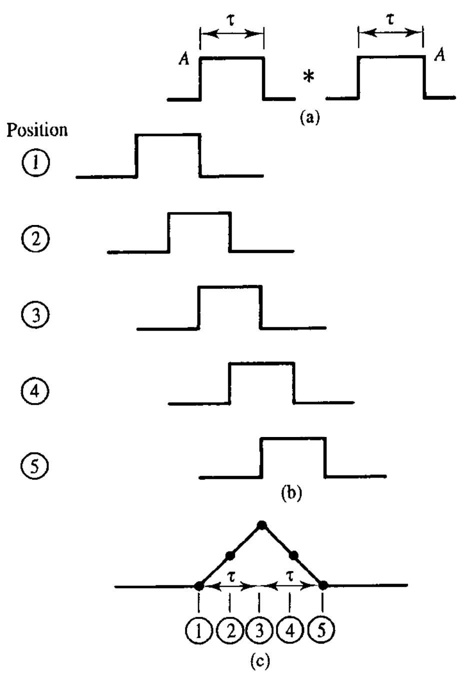
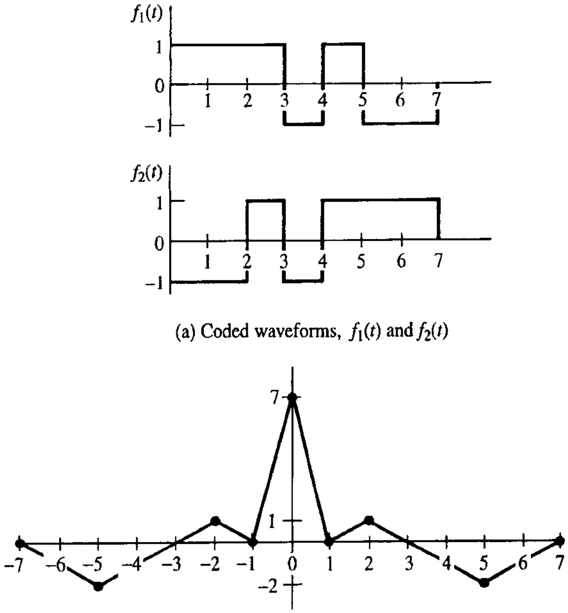


Figure 4.8 *Convolution of two rectangular pulses: (a) pulses; (b) movement of second pulse with respect to first; and (c) values of the convolved output.*



(a) Coded waveforms, $f_1(t)$ and $f_2(t)$

(b) Convolution result, $f_1(t) * f_2(t)$

Figure 4.9 *Convolution of two coded waveforms (Example 4.3).*

EXAMPLE 4.3
Sketch the convolution of the binary coded waveforms $f_1(t)$ and $f_2(t)$ shown in Figure 4.9(a).

This is obtained by time reversing one waveform, e.g. $f_2(t)$, and then sliding it past $f_1(t)$. As each time unit overlaps then, as we are using rectangular pulses, the convolution result is piece-wise linear. The waveforms have been deliberately chosen so that when $f_1(t)$ is time aligned with $f_2(-t)$ then they are identical. The convolution result is shown in Figure 4.9(b). It takes 7 time units to reach the maximum value and another 7 time units to decay again to the final zero value. If $f_2(t)$ is the impulse response of a filter, this represents an example of a matched filter receiver. This type of optimum receiver is discussed in detail later (section 8.3.1).

4.3.5 Step response

Consider the system impulse response shown in Figure 4.10. If the system is driven with a step signal, $u(t)$ (sometimes called the Heaviside step) defined by:

$$u(t) = \begin{cases} 1.0, & t > 0 \\ 0.5, & t = 0 \\ 0\ , & t < 0 \end{cases} \tag{4.28}$$

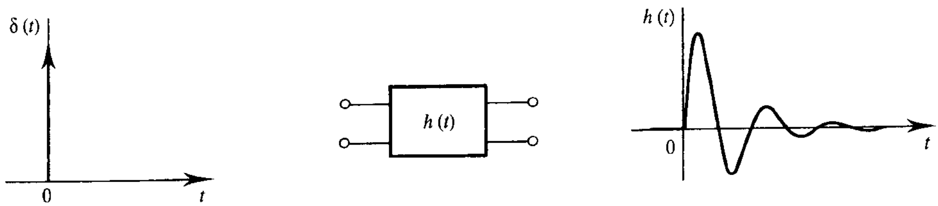


Figure 4.10 *Causal impulse response of a bandpass system.*

then the output of the system (i.e. its step response) is given by:

$$q(t) = \int_{-\infty}^{\infty} h(\tau)u(t-\tau)\, d\tau \qquad (4.29)$$

A graphical interpretation of the integrand in equation (4.29) is shown in Figure 4.11. Since $u(t-\tau) = 0$ for $\tau > t$ and $h(\tau) = 0$ for $\tau < 0$, equation (4.29) can be rewritten as:

$$q(t) = \int_{0}^{t} h(\tau)u(t-\tau)\, d\tau \qquad (4.30)$$

Furthermore, in the region $0 < \tau < t$, $u(t-\tau) = 1.0$, i.e.:

$$q(t) = \int_{0}^{t} h(\tau)\, d\tau \qquad (4.31)$$

The step response is therefore the integral of the impulse response, Figure 4.12. Conversely, of course, the impulse response is the derivative of the step response, i.e.:

$$h(t) = \frac{d}{dt} q(t) \qquad (4.32)$$

Equation (4.32) is particularly useful if the step response of a system is more easily measured than its impulse response.

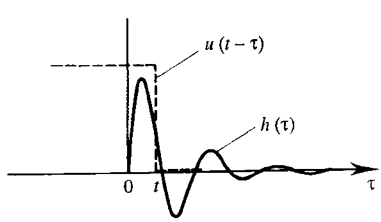


Figure 4.11 *Elements of integrand in equation (4.29).*

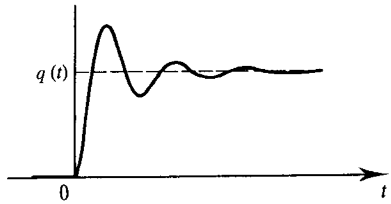


Figure 4.12 *Step response corresponding to impulse response in Figure 4.10.*

EXAMPLE 4.4
Find and sketch the impulse response of the system which has the step response $\Lambda(t-1)$.

The step response is:

$$u(t) = \Lambda(t-1) = \begin{cases} t, & 0 \le t \le 1 \\ 2-t, & 1 \le t \le 2 \\ 0, & \text{elsewhere} \end{cases}$$

Therefore the impulse response is:

$$h(t) = \frac{d}{dt}[\Lambda(t-1)] = \begin{cases} 1, & 0 < t < 1 \\ -1, & 1 < t < 2 \\ 0, & \text{elsewhere} \end{cases}$$

A sketch of $h(t)$ is shown in Figure 4.13.

4.4 Frequency domain description

In the time domain the output of a time invariant linear system is the convolution of its input and its impulse response i.e.:

$$y(t) = h(t) * x(t) \tag{4.33}$$

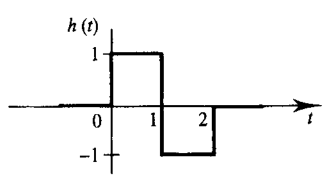


Figure 4.13 *Impulse response of a system with triangular step response (Example 4.4).*

The equivalent frequency domain expression is found by taking the Fourier transform of both sides of equation (4.33) and using the convolution theorem (see Table 2.5):

$$\begin{aligned} \mathrm{FT}\{y(t)\} &= \mathrm{FT}\{h(t) * x(t)\} \\ &= \mathrm{FT}\{h(t)\}\,\mathrm{FT}\{x(t)\} \end{aligned} \tag{4.34}$$

$$\text{i.e. } Y(f) = H(f)\,X(f) \tag{4.35}$$

In equation (4.35), $Y(f)$ is the output voltage spectrum, $X(f)$ is the input voltage spectrum and $H(f)$ is the frequency response of the system. All three quantities are generally complex and can be plotted as either amplitude and phase or real and imaginary components. At a particular frequency, f_o, the frequency response is a single complex number giving the voltage gain (or attenuation) and phase shift of a sinusoid of frequency f_o as it passes from system input to output, i.e.:

$$H(f_o) = A(f_o)e^{j\phi(f_o)} \tag{4.36}$$

For a sinusoidal input, $x(t) = \cos 2\pi f_o t$, the output is therefore given by:

$$y(t) = A(f_o)\cos[2\pi f_o t + \phi(f_o)] \tag{4.37}$$

It follows directly from the Fourier transform relationship between $H(f)$ and $h(t)$ that the frequency responses of systems with real impulse responses have *Hermitian* symmetry, i.e.:

$$\Re\{H(f)\} = \Re\{H(-f)\} \tag{4.38(a)}$$

$$\Im\{H(f)\} = -\Im\{H(-f)\} \tag{4.38(b)}$$

where $\Re/\Im$ indicate real/imaginary parts. Equivalently:

$$|H(f)| = |H(-f)| \tag{4.38(c)}$$

$$\phi(f) = -\phi(-f) \tag{4.38(d)}$$

EXAMPLE 4.5

A linear system with the impulse response shown in Figure 4.14(a) is driven by the input signal shown in Figure 4.14(b). Find (i) the voltage spectral density of the input signal, (ii) the frequency response of the system, (iii) the voltage spectral density of the output signal, (iv) the (time domain) output signal.

(i) The input signal is given by the difference between two rectangular functions:

$$v_{in}(t) = 3.0\,\Pi\left(\frac{t-1}{2}\right) - 3.0\,\Pi\left(\frac{t-3}{2}\right)$$

The voltage spectral density of the input is given by the Fourier transform of this:

$$V_{in}(f) = \mathrm{FT}\{v_{in}(t)\}$$

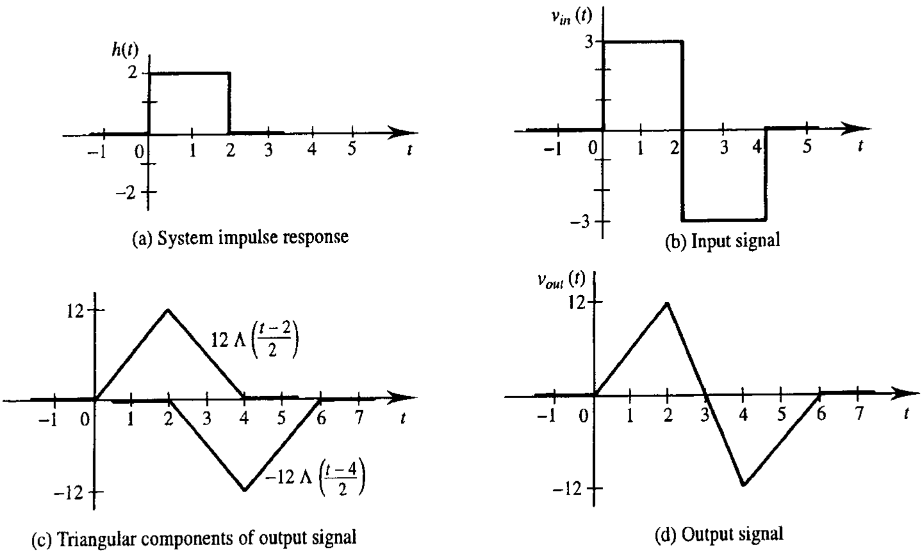


Figure 4.14 *Functions for Example 4.5.*

$$= 3.0\,\mathrm{FT}\left\{\Pi\left(\frac{t-1}{2}\right)\right\} - 3.0\,\mathrm{FT}\left\{\Pi\left(\frac{t-3}{2}\right)\right\}$$

$$= 3.0\left[\,2\,\mathrm{sinc}(2f)\,e^{-j\omega 1}\,\right] - 3.0\left[\,2\,\mathrm{sinc}(2f)\,e^{-j\omega 3}\,\right]$$

$$= 6\,\mathrm{sinc}(2f)\left[\,e^{j\omega} - e^{-j\omega}\,\right]e^{-j\omega 2}$$

$$= 6\,\mathrm{sinc}(2f)\,2j\,\sin\omega\,e^{-j2\omega} = 12j\,\mathrm{sinc}(2f)\,\sin(2\pi f)\,e^{-j4\pi f}$$

$$= 12\,\frac{\sin^2(2\pi f)}{2\pi f}\,e^{-j\left(4\pi f - \frac{\pi}{2}\right)}$$

(ii) The frequency response of the system is:

$$H(f) = \mathrm{FT}\,\{h(t)\}$$

$$= \mathrm{FT}\left\{\,2.0\,\Pi\left(\frac{t-1}{2}\right)\right\}$$

$$= 2.0\left[2\,\mathrm{sinc}(2f)\,e^{-j\omega 1}\right] = 4\,\mathrm{sinc}(2f)\,e^{-j2\pi f}$$

(iii) The voltage spectral density of the output signal is given by:

$$V_{out}(f) = V_{in}(f)\,H(f)$$

$$= j12\,\frac{\sin^2(2\pi f)}{2\pi f}\,e^{-j4\pi f}\,4\,\frac{\sin(2\pi f)}{2\pi f}\,e^{-j2\pi f}$$

$$= 12 \frac{\sin^3(2\pi f)}{(\pi f)^2} e^{-j\left(6\pi f - \frac{\pi}{2}\right)} = 12 \frac{\sin^3(2\pi f)}{(\pi f)^2} e^{-j\pi(6f - 0.5)}$$

(iv) The time domain output signal could be found as the inverse Fourier transform of $V_{out}(f)$. It is easier in this example, however, to find the output by convolving the input and impulse response, i.e.:

$$v_{out}(t) = v_{in}(t) * h(t) = \int_{-\infty}^{\infty} v_{in}(\tau)h(t - \tau)\, d\tau$$

Furthermore the problem can be simplified if $v_{in}(t)$ is split into its component parts:

$$v_{out}(t) = \left[3\Pi\left(\frac{t-1}{2}\right) - 3\Pi\left(\frac{t-3}{2}\right)\right] * 2\Pi\left(\frac{t-1}{2}\right)$$

$$= 6\left[\Pi\left(\frac{t-1}{2}\right) * \Pi\left(\frac{t-1}{2}\right)\right] - 6\left[\Pi\left(\frac{t-3}{2}\right) * \Pi\left(\frac{t-1}{2}\right)\right]$$

We know that the result of convolving two rectangular functions of equal width gives a triangular function. Furthermore the peak value of the triangular function is numerically equal to the area under the product of the aligned rectangular functions and occurs at the time shift of the reversed function which gives this alignment. The half width of the triangular function is the same as the width of the rectangular function. Thus:

$$v_{out}(t) = 6\left[2\Lambda\left(\frac{t-2}{2}\right)\right] - 6\left[2\Lambda\left(\frac{t-4}{2}\right)\right]$$

$$= 12\Lambda\left(\frac{t-2}{2}\right) - 12\Lambda\left(\frac{t-4}{2}\right)$$

The two triangular functions making up $v_{out}(t)$ are shown in Figure 4.14(c) and their sum, $v_{out}(t)$, is shown in Figure 4.14(d).

4.5 Causality and the Hilbert transform

All physically realisable systems must be causal, i.e.:

$$h(t) = 0, \qquad \text{for } t < 0 \tag{4.39(a)}$$

This is intuitively obvious since physical systems should not respond to inputs before the inputs have been applied. An equivalent way of expressing equation (4.39(a)) is:

$$h(t) = u(t)h(t) \tag{4.39(b)}$$

where $u(t)$ is the Heaviside step function. The frequency response of a causal system with real impulse response must therefore satisfy:

$$H(f) = \text{FT}\{u(t)\} * \text{FT}\{h(t)\}$$

$$= \left[\frac{1}{2}\,\delta(f) + \frac{1}{j2\pi f} \right] * H(f)$$

$$= \frac{1}{2}\,H(f) + \left[\frac{1}{j2\pi f} * H(f) \right] \tag{4.40(a)}$$

$$\text{i.e.}\quad H(f) = \frac{1}{j\pi f} * H(f) \tag{4.40(b)}$$

Equation (4.40(b)) is precisely equivalent to equation (4.39(b)).

A necessary and sufficient condition for an amplitude response, $A(f) = |H(f)|$, to be *potentially* causal is:

$$\int_{-\infty}^{\infty} \frac{|\ln A(f)|}{1 + f^2}\, df < \infty \tag{4.41}$$

The expression potentially causal, in this context, means that a system satisfying this criterion will be causal *given that it has a suitable phase response.* Equation (4.41) is called the Paley-Wiener criterion. It has the important implication that a causal system can only have isolated zeros in its amplitude response, i.e. $A(f)$ cannot be zero over a finite band of frequencies.

Returning to the causality condition of equation (4.40(b)), if $H(f)$ is expressed as real and imaginary parts:

$$H_{\Re}(f) + j\,H_{\Im}(f) = \frac{1}{j\pi f} * [H_{\Re}(f) + j\,H_{\Im}(f)]$$

$$= \left[\frac{1}{j\pi f} * H_{\Re}(f) \right] + \left[\frac{1}{\pi f} * H_{\Im}(f) \right] \tag{4.42}$$

and real and imaginary parts are equated, then:

$$H_{\Re}(f) = \frac{1}{\pi f} * H_{\Im}(f) \tag{4.43(a)}$$

$$H_{\Im}(f) = -\frac{1}{\pi f} * H_{\Re}(f) \tag{4.43(b)}$$

The relationship between real and imaginary parts of $H(f)$ in equation (4.43(a)) is called the *inverse* (frequency domain) *Hilbert* transform which can be written explicitly as:

$$H_{\Re}(f) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{H_{\Im}(f')}{f - f'}\, df' \tag{4.44}$$

Equation (4.43(b)) is the *forward* Hilbert transform often denoted by:

$$H_{\Im}(f) = \hat{H}_{\Re}(f) \tag{4.45}$$

In the time domain (since the real part of $H(f)$ transforms to the even part of $h(t)$ and the imaginary part of $H(f)$ transforms to the odd part of $h(t)$) the equivalent operations to equations (4.43) are:

$$h_{even}(t) = j\ \text{sgn}(t) h_{odd}(t) \tag{4.46}$$

$$h_{odd}(t) = -j\ \text{sgn}(t) h_{even}(t) \tag{4.47}$$

Notice that, unlike the Fourier transform, the Hilbert transform does *not* change the domain of the function being transformed. It can therefore be applied either in the frequency domain (as in equation (4.43(b)) or in the time domain. Table 4.1 summarises the frequency and time domain Hilbert transform relationships.

Table 4.1 *Summary of frequency and time domain Hilbert transform relationships.*

$-j\ \text{sgn}(t)x(t)$	FT $\rightleftarrows$	$\hat{X}(f) = \frac{-1}{\pi f} * X(f)$
$\downarrow\uparrow$ HT_f		HT_f $\uparrow\downarrow$
$x(t)$	FT $\rightleftarrows$	$X(f)$
$\uparrow\downarrow$ HT_t		HT_t $\downarrow\uparrow$
$x(t) = \frac{-1}{\pi t} * x(t)$	FT $\rightleftarrows$	$+j\ \text{sgn}(f)X(f)$

(HT_t *is the time domain Hilbert transform,* HT_f *is the frequency domain Hilbert transform.*)

The time domain Hilbert transform is sometimes called the quadrature filter since it represents an all-pass filter which shifts the phase of positive frequency components by +90° and negative frequency components by −90°. This operation is useful in the representation of bandpass signals and systems as equivalent baseband processes (see section 13.2). It also makes obvious the property that a function and its Hilbert transform are orthogonal.

EXAMPLE 4.6

Establish which of the following systems are causal and which are not: (i) $h(t) = \Lambda(t-3)$; (ii) $h(t) = e^{-(t-10)^2}$; (iii) $h(t) = u(t)e^{-t}$; (iv) $H(f) = e^{-f^2}$; (v) $H(f) = \Pi(f)$; (vi) $H(f) = \Lambda(f-3) + \Lambda(f+3)$; (vii) $H(f) = (1 - jf)/(1 + f^2)$.

(i) $\Lambda(t-3)$ represents a triangular function which is centred on $t = 3$ and which is zero for $t < 2$ and $t > 4$. It is therefore a causal impulse response.

(ii) $e^{-(t-10)^2}$ represents a Gaussian function centred on $t = 10$. Since it only tends to zero as $t \rightarrow \pm\infty$ it represents an acausal impulse response.
(iii) $u(t)e^{-t}$ is causal by definition since the Heaviside factor ensures it is zero for $t < 0$.
(iv) e^{-f^2} represents a Gaussian frequency response. Since Gaussian functions in one domain transform to Gaussian functions in the other domain the impulse response of this system is Gaussian. The system is therefore acausal as in (ii).
(v) $\Pi(f)$ is a strictly bandlimited frequency response. The impulse response cannot therefore be time limited and is thus acausal. (The impulse response is, of course, sinc(t).)
(vi) $\Lambda(f-3)+\Lambda(f+3)$ represents a bandpass triangular amplitude response. It is strictly bandlimited and therefore an acausal system as in (v).
(vii)To test whether $H(f)$ is causal we can find out if $H_{\Im}(f)$ is the Hilbert transform of $H_{\Re}(f)$.

$$H_{\Im}(f) = \frac{-f}{1+f^2}, \quad H_{\Re}(f) = \frac{1}{1+f^2}$$

In the absence of Hilbert transform tables:

$$\hat{H}_{\Re}(f) = -\frac{1}{\pi f} * \frac{1}{1+f^2} = -\int_{-\infty}^{\infty} \frac{1}{\pi\phi} \frac{1}{1+(f-\phi)^2}\, d\phi$$

$$= -\frac{1}{\pi}\int_{-\infty}^{\infty} \frac{1}{\phi\,(\phi^2 - 2f\phi + f^2 + 1)}\, d\phi$$

Using a table of standard integrals (e.g. Dwight, 4th edition, Equation 161.11):

$$\hat{H}_{\Re}(f) = -\frac{1}{\pi}\left[\frac{1}{2(f^2+1)} \ln\left(\frac{\phi^2}{\phi^2 - 2f\phi + f^2 + 1}\right) + \frac{2f}{2(f^2+1)} \int \frac{1}{\phi^2 - 2f\phi + f^2 + 1}\, d\phi\right]_{-\infty}^{\infty}$$

The logarithmic factor in the first term in square brackets above tends to zero as $\phi \rightarrow \pm\infty$. The integral in the second term is also standard (e.g. Dwight, 4th edition, equation (160.01)) giving:

$$\hat{H}_{\Re}(f) = -\frac{1}{\pi}\left[\frac{f}{(f^2+1)} \frac{2}{\sqrt{4(f^2+1) - 4f^2}} \tan^{-1}\left(\frac{(2\phi - 2f)}{\sqrt{4(f^2+1) - 4f^2}}\right)\right]_{-\infty}^{\infty}$$

$$= \frac{-1}{\pi} \frac{f}{f^2+1}\left[\tan^{-1}(\phi - f)\right]_{-\infty}^{\infty} = -\frac{f}{f^2+1}$$

$$= H_{\Im}(f)$$

Thus $H_{\Im}(f)$ is the Hilbert transform of $H_{\Re}(f)$ and $H(f)$ therefore represents a causal system.

4.6 Random signals and linear systems

The effect of a linear system on a deterministic signal is specified completely by:

$$y(t) = h(t) * x(t) \tag{4.48}$$

or, alternatively, by:

$$Y(f) = H(f)X(f) \tag{4.49}$$

Random signals cannot, by definition, be specified as deterministic functions either in the time domain or in the frequency domain. It follows that neither equation (4.48) nor equation (4.49) is particularly useful for information bearing signals or noise. In practice, however, the two properties of such signals which must most commonly be specified are their power spectra and their probability density functions. The effects of linear systems on these signal characteristics are now described.

4.6.1 PSDs and linear systems

The most direct way of deriving the relationship between the power spectral density at the input and output of a linear system is to take the square magnitude of equation (4.49), i.e.:

$$\begin{aligned}|Y(f)|^2 &= |H(f)X(f)|^2 \\ &= |H(f)|^2|X(f)|^2 \end{aligned} \tag{4.50}$$

Since $|Y(f)|^2$ and $|X(f)|^2$ are power spectral densities equation (4.50) can be rewritten as:

$$G_y(f) = |H(f)|^2 G_x(f) \quad (\text{V}^2/\text{Hz}) \tag{4.51}$$

If the system input is an energy signal then the power spectral densities, equations (2.52) and (3.49), are replaced by energy spectral densities, equation (2.53):

$$E_y(f) = |H(f)|^2 E_x(f) \quad (\text{V}^2\text{s/Hz}) \tag{4.52}$$

The equivalent time domain description is obtained by taking the inverse Fourier transform of equation (4.50):

$$\text{FT}^{-1}\{Y(f)\,Y^*(f)\} = \text{FT}^{-1}\{H(f)H^*(f)\} * \text{FT}^{-1}\{X(f)X^*(f)\} \tag{4.53}$$

Using the Wiener-Kintchine theorem (or equivalently the conjugation and time reversal Fourier transform theorems, Table 2.5):

$$R_{yy}(\tau) = R_{hh}(\tau) * R_{xx}(\tau) \tag{4.54}$$

where R is the correlation function and the double subscript emphasises auto- or self-correlation. It is almost always the frequency domain description which is the most convenient in practice. As an example of the application of equations (4.51) and (4.52) the noise power spectral density and total noise power at the output of an *RC* filter are now calculated for the case of white input noise.

EXAMPLE 4.7

Find the output power spectral density for a simple *RC* filter when it is driven by white noise. What is the total noise power at the filter's output?

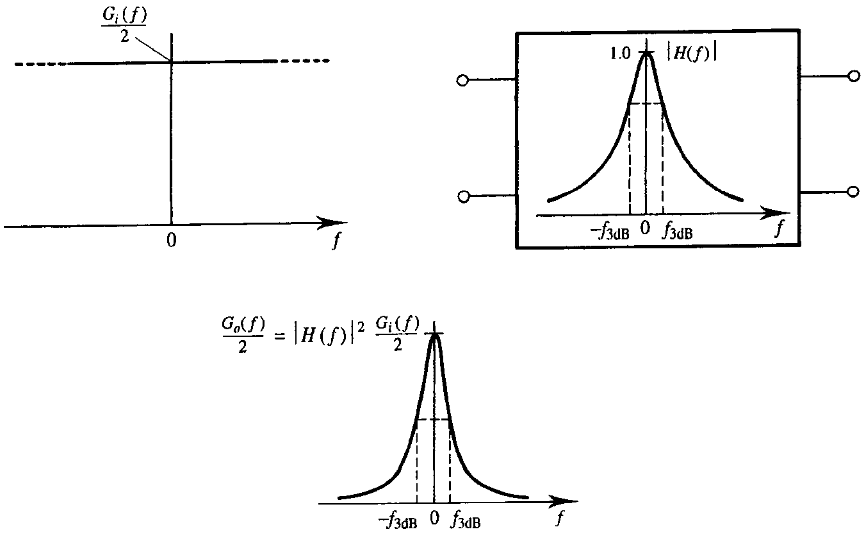


Figure 4.15 *NPSD at output of single pole RC filter driven by white noise. ($G(f)$ is the one sided power spectral density.)*

Figure 4.15 shows the problem schematically. The power spectral density at the filter output is:

$$G_o(f) = |H(f)|^2 \, G_i(f)$$

The frequency response of the filter is given by:

$$H(f) = \frac{1}{1 + j(f/f_{3dB})}$$

where f_{3dB} is the filter −3 dB, or cut-off, frequency. Substituting:

$$G_o(f) = \left| \frac{1}{1 + j(f/f_{3dB})} \right|^2 G_i(f)$$

$$= \frac{1}{1 + (f/f_{3dB})^2} \, G_i(f)$$

Interpreting $G_i(f)$ and $G_o(f)$ as one sided, the total noise power, N, at the filter output is:

$$N = \int_0^\infty G_o(f) \, df$$

$$= \int_0^\infty \frac{1}{1 + (f/f_{3dB})^2} \, G_i(f) \, df$$

Using the change of variable $u = f/f_{3dB}$ and remembering that the input noise is white (i.e. $G_i(f)$ is a constant, Figure 4.15) then:

$$N = G_i \int_0^\infty \frac{1}{1 + u^2} \, f_{3dB} \, du$$

$$= G_i f_{3dB} \int_0^\infty \frac{1}{1+u^2}\, du = G_i\, f_{3dB}\, [\tan^{-1} u]_0^\infty$$

$$= G_i\, f_{3dB}\, \pi/2 \quad (V^2)$$

4.6.2 Noise bandwidth

The noise bandwidth, B_N, of a filter is defined as that width which a rectangular frequency response would need to have to pass the same noise power as the filter, given identical white noise at the input to both. This definition is illustrated in Figure 4.16. It can be expressed mathematically as:

$$B_N = \int_0^\infty \frac{|H(f)|^2}{|H(f_p)|^2}\, df \tag{4.55}$$

where f_p is the frequency of peak amplitude response. Notice that noise bandwidth is not equal, in general, to the −3 dB bandwidth. (For the single pole lowpass filter noise bandwidth is larger than the −3 dB bandwidth by a factor $\pi/2$ (see Example 4.7). In this case, for white noise calculations, the use of the 3 dB bandwidth in place of noise bandwidth would therefore lead to a noise power error of 2 dB.)

4.6.3 Pdf of filtered noise

Not only the power spectral density of a random signal is changed when it is filtered but, in general, so is its probability density function. The effect of memoryless systems on the pdf of a signal is discussed in section 4.7. Most communications subsystems, however, have non-zero memory. Unfortunately in this case there is no general, analytical, method of deriving the pdf of the output from the pdf of the input. There is, however, an important exception to this, for which a general result can be derived. This is the pdf of filtered Gaussian noise. Consider Figure 4.17. The output noise, $n_o(t)$, is given by the convolution of the input noise, $n_i(t)$, with the impulse response of the filter or system, i.e.:

$$n_o(t) = \int_{-\infty}^{t} h(t-\tau)\, n_i(\tau)\, d\tau \tag{4.56}$$

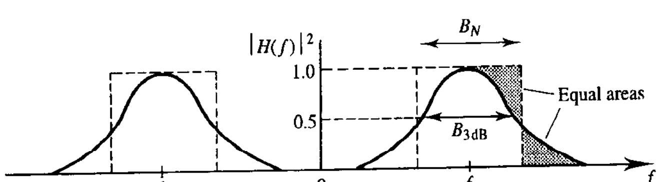


Figure 4.16 *Illustration of noise bandwidth, B_N.*

Notice that the output can be interpreted as a sum of weighted input impulses of strength $n_i(\tau)d\tau$, the weighting factor, $h(t-\tau)$, depending on the time at which the individual input impulses occurred. If $n_i(t)$ is white Gaussian noise then the adjacent impulses are independent Gaussian random variables (since the autocorrelation of $n_i(t)$ is impulsive). The output noise at any instant, e.g. $n_o(t_1)$, is therefore a linear sum of Gaussian random variables and is consequently, itself, a Gaussian random variable. Thus:

Filtered white Gaussian noise is Gaussian

The above result is easily generalised in the following way.

Consider the frequency response in Figure 4.17 to be split into two parts as shown in Figure 4.18. White Gaussian noise at the input of $H_1(f)$ has been shown to result in (non-white) Gaussian noise at the output of $H_2(f)$. Applying the same reasoning, however, the input to $H_2(f)$ is (non-white) Gaussian noise. It follows, by considering the

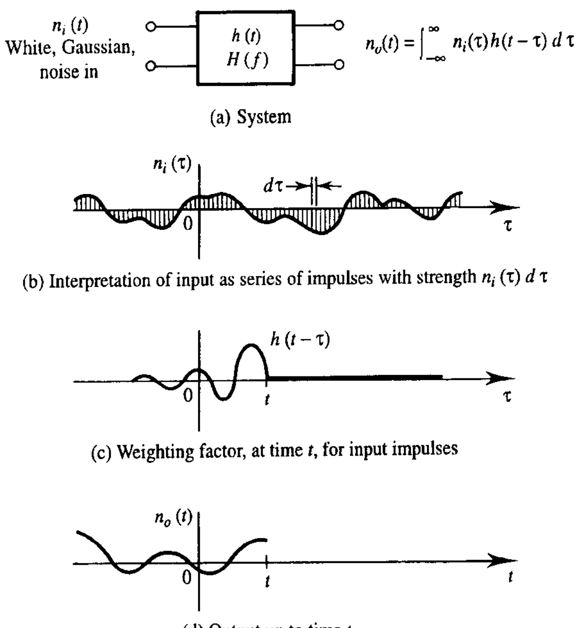


Figure 4.17 *Output as linear sum of many independent, Gaussian, random impulses. ($n_i(\tau)$ is white, Gaussian, noise but it can only be drawn as a bandlimited process.)*

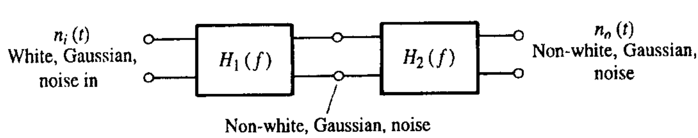


Figure 4.18 *Reinterpretation of $H(f)$ in Figure 4.17 as two cascaded sections.*

input and output of $H_2(f)$ that:

Filtered Gaussian noise is Gaussian

The above result is exact and, to some extent, obvious in that if it were not true then the Gaussian nature of thermal noise, for example, would be obscured by the many filtering processes it is normally subjected to before it is measured.

4.6.4 Spectrum analysers

Spectrum analysers are instruments which are used to characterise signals in the frequency domain. If the signal is periodic the characterisation is partial in the sense that phase information is not usually displayed. If the signal is random, as is the case for noise, the spectral characterisation is essentially complete. Figure 4.19 shows a simplified block diagram of a spectrum analyser and Figure 4.20 shows an alternative conceptual implementation of the same instrument. At a given frequency the display shows either the RMS voltage or mean square voltage which is passed by the filter when it is centred on that frequency. In practice the display y-axis is usually calibrated in dBμ given by $20 \log_{10} (V_{RMS}/10^{-6})$ or dBm given by $10 \log_{10}[(V_{RMS}^2/R_{in})/10^{-3}]$ where dBμ

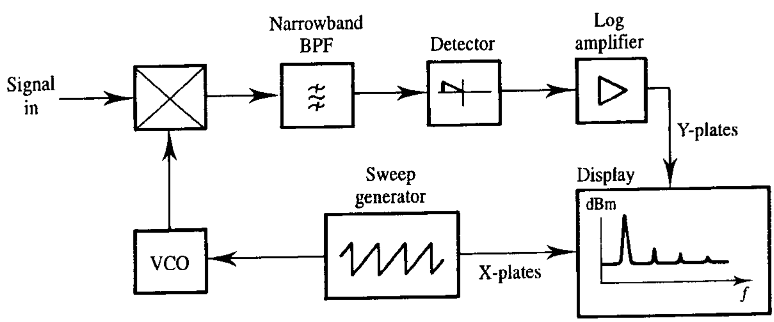


Figure 4.19 *Simplified block diagram of real time, analogue, spectrum analyser.*

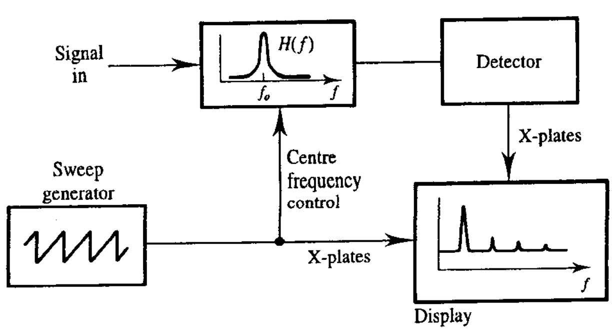


Figure 4.20 *Conceptual model of spectrum analyser.*

indicates dB with respect to a 1 μV and dBm indicates dB with respect to 1 mW. (R_{in}, often 50 Ω, is the input impedance which converts V^2 to W.)

If the signal being measured is periodic, and the filter has a bandwidth which adequately resolves the resulting spectral lines, then the (dBm) display is a faithful representation of the signal's discrete power spectrum. If the signal being measured is a stationary random process, however, then its power spectral density is continuous and the resulting display is the actual power spectral density *correlated* with the filter's squared amplitude response, $|H(f)|^2$ (see section 2.6 and equation (2.85)). If the bandwidth of the filter is narrow compared with the frequency scale over which the signal's power spectral density changes significantly, then the smearing of the spectrum in the correlation process is small and the shape of the spectrum is essentially unchanged. In this case the signal's power spectral density in W/Hz can be found by dividing the displayed spectrum (in watts) by the noise bandwidth, B_N, of the filter (in Hz). On a dB scale this corresponds to:

$$G(f)\ \text{(dB mW Hz}^{-1}\text{)} = \text{Display (dBm)} - 10\log_{10} B_N\ \text{(dB Hz)} \tag{4.57}$$

4.7 Non-linear systems and transformation of random variables

Non-linear systems are, in general, difficult to analyse. This is principally because superposition no longer applies. As a consequence complicated input signals cannot be decomposed into simple signals (on which the effect of the system is known) and the resulting modified components recombined at the output.

There is one signal characteristic, however, which can often be found at the output of *memoryless* non-linear systems without too much difficulty. This is the signal's probability density function. Mathematically this problem is called a transformation of random variables. An outline of this technique is given below.

Consider a pair of bivariate random variables X, Y and S, T which are related in some deterministic way. Every point in the x, y plane can be mapped into the s, t plane as shown in Figure 4.21. Now consider all the points (x_1, y_1), (x_2, y_2), $\cdots$, in the x, y plane which map into the rectangle centred on s_1, t_1. (There may be none, one, or more than one such point.) Each one of these points (x_n, y_n) has its own small area dA_n in the x, y plane which maps into the rectangle in (s, t). The probability that X, Y lies in any of the areas (x_n, y_n) is equal to the probability that S, T lies in the rectangle at (s_1, t_1), i.e.:

$$\sum_n p_{X,Y}(x_n, y_n)\, dA_n = p_{S,T}(s_1, t_1)\, ds\, dt \tag{4.58}$$

Equation (4.58) can be interpreted as a conservation of probability law.

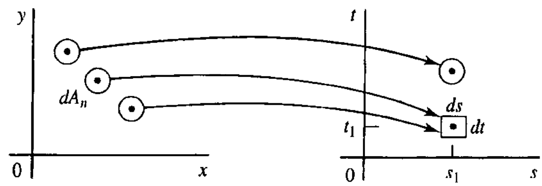


Figure 4.21 *Mapping of random variables using a transformation* $(x, y) \rightarrow (s, t)$.

4.7.1 Rayleigh probability density function

When Gaussian noise (Figure 4.22) is present at the input of an envelope (i.e. amplitude) detector the pdf of the noise at the output is Rayleigh distributed (Figure 4.23). The derivation of this distribution, given below, is a transformation of random variables and uses the conservation of probability law given in equation (4.58).

Let X, Y (quadrature noise components) be independent Gaussian random variables with equal standard deviations, σ, and zero means. Equation (3.33) then simplifies to:

$$p_{X,Y}(x, y) = \frac{1}{\sigma\sqrt{2\pi}} e^{\left(\frac{-x^2}{2\sigma^2}\right)} \frac{1}{\sigma\sqrt{2\pi}} e^{\left(\frac{-y^2}{2\sigma^2}\right)} \tag{4.59}$$

Let R, Θ (noise amplitude and phase) be a new pair of random variables related to X, Y by:

$$r = \sqrt{x^2 + y^2} \tag{4.60(a)}$$

$$\theta = \tan^{-1}(y/x) \tag{4.60(b)}$$

(r, θ can be interpreted as the polar coordinates of the point x, y as shown in Figure 4.24.) The area $d\theta dr$ in the R, Θ plane corresponds to an area $dA = r\ d\theta\ dr$ in the X, Y plane,

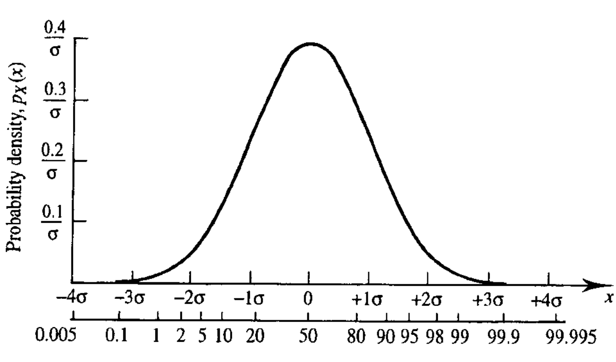


Figure 4.22 *Gaussian probability density function,* $p_X(x) = [1/(\sigma\sqrt{2\pi})]e^{-(x^2/2\sigma^2)}$ *and CD in per- cent.*

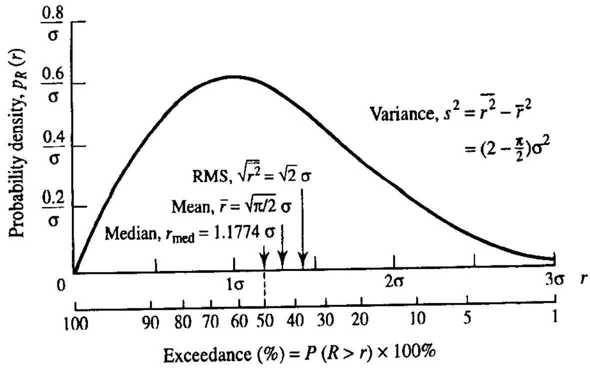


Figure 4.23 *Rayleigh probability density function, $p_R(r) = [r/\sigma^2]e^{-(r^2/2\sigma^2)}$, where σ is the standard deviation of either component in the parent bivariate Gaussian pdf.*

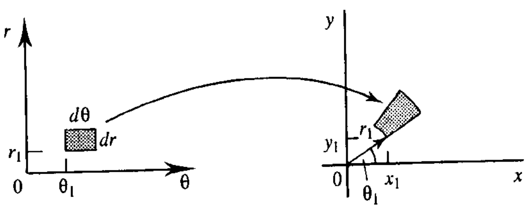


Figure 4.24 *Relationship between (r, θ) and (x, y).*

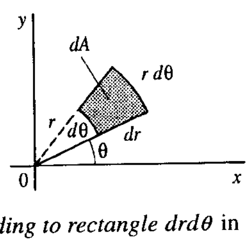


Figure 4.25 *Area in x, y corresponding to rectangle $drd\theta$* in r, θ.

Figure 4.25. Conservation of probability requires that:

$$p_{R,\Theta}(r, \theta)\; dr\; d\theta \;=\; p_{X,Y}(x, y)\; r\; d\theta dr \tag{4.61}$$

Therefore:

$$\begin{aligned} p_{R,\Theta}(r, \theta) &= p_{X,Y}(x, y)\; r \\ &= \frac{r}{\sigma^2 2\pi}\, e^{\left(-\frac{x^2+y^2}{2\sigma^2}\right)} = \frac{r}{2\pi\sigma^2}\, e^{\left(\frac{-r^2}{2\sigma^2}\right)} \end{aligned} \tag{4.62}$$

Equation (4.62) gives the joint pdf of R and Θ. The (marginal) pdf of R is now given by:

$$p_R(r) = \int_0^{2\pi} p_{R,\Theta}(r,\theta)\, d\theta$$

$$= \frac{r}{2\pi\sigma^2} e^{\left(\frac{-r^2}{2\sigma^2}\right)} \int_0^{2\pi} d\theta \tag{4.63}$$

$$\text{i.e. } p_R(r) = \frac{r}{\sigma^2} e^{\left(\frac{-r^2}{2\sigma^2}\right)} \tag{4.64}$$

Equation (4.64) is the Rayleigh probability density function shown in Figure 4.23. Since $p_{R,\Theta}(r,\theta)$ has no θ dependence the marginal probability density function of Θ is uniform, Figure 4.26, i.e.:

$$p_\Theta(\theta) = \frac{1}{2\pi} \tag{4.65}$$

(Strictly the RHS of equations (4.64) and (4.65) should be multiplied by the Heaviside step function, $u(r)$, and the rectangular function, $\Pi(\theta/2\pi)$, respectively since the probability densities are zero outside these ranges.)

4.7.2 Chi-square distributions

Another transformation of random variables common in electronic communication systems occurs when Gaussian noise is present at the input to a square law device. Let the random variable X representing noise at the input of a square law detector be Gaussianly distributed, i.e.:

$$p_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\left(\frac{-x^2}{2\sigma^2}\right)} \tag{4.66}$$

The detector (Figure 4.27) is characterised by:

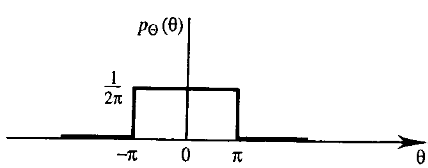


Figure 4.26 *Uniform distribution of Θ.*

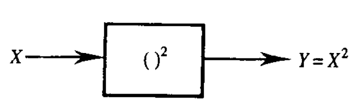


Figure 4.27 *Square-law detector.*

$$Y = X^2 \tag{4.67}$$

which means that two probability areas in X (i.e. $p_X(x)dx$ and $p_X(-x)dx$) both transform to the same probability area $p_Y(y)dy$. This is illustrated in Figure 4.28. Conservation of probability therefore requires that:

$$p_X(x)\,dx + p_X(-x)\,dx = p_Y(y)\,dy \tag{4.68}$$

And by symmetry this means that:

$$2\,p_X(x)\,dx = p_Y(y)\,dy \tag{4.69}$$

Thus:

$$p_Y(y) = 2\,p_X(x)\,\frac{dx}{dy} \tag{4.70}$$

and since $y = x^2$ then:

$$\frac{dy}{dx} = 2x \tag{4.71(a)}$$

and:

$$\frac{dx}{dy} = \frac{1}{2x} \tag{4.71(b)}$$

Therefore:

$$\begin{aligned} p_Y(y) &= 2\,p_X(x)\,\frac{1}{2x} \\ &= \frac{1}{\sigma\sqrt{2\pi}}\,\frac{1}{x}\,e^{\left(-\frac{x^2}{2\sigma^2}\right)} \end{aligned} \tag{4.72}$$

Using $x = \sqrt{y}$ gives:

$$p_Y(y) = \frac{1}{\sigma\sqrt{2\pi}}\,\frac{1}{\sqrt{y}}\,e^{\left(\frac{-y}{2\sigma^2}\right)}, \quad \text{for } y \geq 0 \tag{4.73}$$

(For $y < 0$, $P_Y(y) = 0$.) Equation (4.73) is, in fact, the special case for $N = 1$ of a more

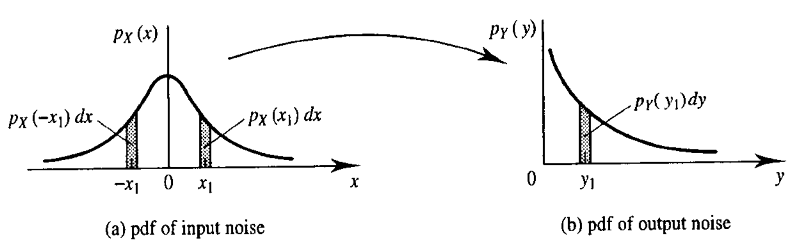


Figure 4.28 *Mapping of two input points to one output point for a square law detector: (a) pdf of input noise; (b) pdf of output noise.*

general distribution which results from the transformation $Y = \sum_{i=1}^{N} X_i^2$ where X_i are independent Gaussian random variables with equal variance, σ^2, and zero mean. N, here, is the number of degrees of freedom of the distribution. The mean and variance of this generalised chi-square, χ^2, distribution are, respectively:

$$\bar{Y} = N\sigma^2 \tag{4.74(a)}$$

and

$$\sigma_Y^2 = 2N\sigma^4 \tag{4.74(b)}$$

The pdf for a χ^2 distribution with various degrees of freedom and $\sigma^2 = 1$ is shown in Figure 4.29.

4.8 Summary

Linear systems obey the principle of superposition. Many of the subsystems used in the design of digital communications systems are linear over their normal operating ranges. Linear systems are useful and important because they can be described by linear differential equations.

It is a property of a linear system that its time domain output is given by its time domain input convolved with its impulse response. The impulse response of a linear system is the time derivative of its step response. The output (complex) voltage spectrum of a linear system is the voltage spectrum of its input multiplied by its (complex) frequency response. A system's impulse response and frequency response form a Fourier transform pair.

All physically realisable systems are causal, i.e. their outputs do not anticipate their inputs. The real and imaginary parts of the frequency response of a causal system form a (frequency domain) Hilbert transform pair. The PSD of a random signal at the output of a linear system is given by the PSD at its input multiplied by its squared amplitude response. The autocorrelation of a random signal at the output of a linear system is the

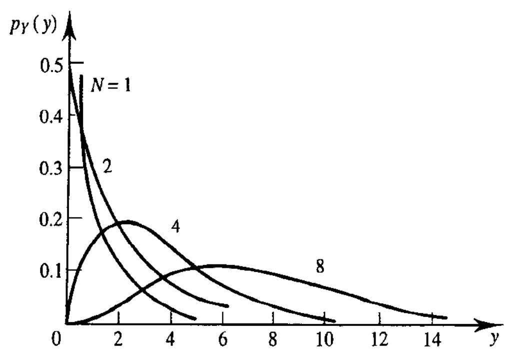


Figure 4.29 *Pdf of a chi-square distribution for several different degress of freedom* ($\sigma^2 = 1$).

convolution of the input signal's autocorrelation with the autocorrelation of the system's impulse response. The noise bandwidth of a system is equal to the width of an ideal rectangular amplitude response which passes the same noise power as the system, for identical white noise at the inputs.

The pdf at the output of memoryless systems, both linear and non-linear, can be found by the method of transformation of random variables. (Linear memoryless systems have a trivial effect since they represent simple scaling factors.) No general, analytical methods are currently known for predicting the pdf at the output of systems with memory. A useful result for the special case of Gaussian noise, however, is that filtered Gaussian noise is Gaussian.

Spectrum analysers are widely used to display the power spectrum of signals. For periodic signals a properly adjusted spectrum analyser will display the signal's discrete line spectrum, usually on a decibel scale of power. For random signals the spectrum analyser displays a good approximation to the signal's power spectral density. In this case care is needed in interpreting the absolute magnitude of the spectrum (in W/Hz) if this information is important. (Often only the shape of the spectrum is required, as the total power in the signal is already known.)

4.9 Problems

4.1. Classify the following systems (input $x(t)$, output $y(t)$, impulse response $h(t)$) as: linear or non-linear, time varying or time invariant, causal or non-causal, memoryless or non-zero memory. (N.B. $u(t)$ is the Heaviside step function.)
(a) $y(t) = 3.7x(t)$, (b) $y(t) = 3.7x(t - 6.2)$, (c) $y(t) = 3.7x(t + 10^{-20})$,
(d) $y(t) = 3.7[x(t - 6.2) + 0.01]$, (e) $y(t) = x(t)\cos(2\pi 50t)$, (f) $y(t) = x^{1.1}(t)e^{-t}$,
(g) $y(t) = \cos(2\pi 50t)[x(t) + x(t - 1)]$, (h) $h(t) = u(t)\cos[2\pi 100(t + 4)]e^{-t}$, (i) $y(t) = x(t)x(t - 2)$,
(j) $y(t) = d/dt\ [x(t + 1)]$, (k) $y(t) = x(t) * u(t)e^{-t}$, (l) $y(t) = x(t/3)$, (m) $y(t) = \int_0^t t'x(t')\, dt'$,
(n) $y(t) = 1/(1 + x(t))$, (o) $y(t) = x(t) + y(t - 1)$, (p) $h(t) = [1 - u(-t)]e^{-t^2}$, (q) $y(t) = \text{sgn}[x(t)]$

4.2. A circuit is described by the linear differential equation:

$$Ry(t) + 2L\frac{dy(t)}{dt} + RLC\frac{d^2y(t)}{dt^2} + L^2C\frac{d^3y(t)}{dt^3} = R\,x(t)$$

where R, L and C are constants, $x(t)$ is the input and $y(t)$ is the output. Find, by taking the Fourier transform of the differential equation, term by term, an expression for the frequency response of the system. What is the amplitude multiplication factor, and phase shift, of a sinusoidal input at the frequency $f = 1/(2\pi\sqrt{LC})$?

4.3. A linear system has the impulse response $h(t) = u(t) - u(t - 2)$. Sketch $h(t)$ and find the system output when its input is $x(t) = \frac{1}{2}\Pi((t - 1)/2) - \frac{1}{2}\Pi((t - 3)/2)$. What is the system's step response and what is its frequency response?

4.4. How might a system with the impulse response given in Problem 4.3 be implemented using integrators, delay lines, invertors (i.e. amplifiers with voltage gain $G_V = -1.0$) and adders?

4.5. The impulse response of a system is given by $h(t) = \Pi((t - 2)/2)$ and the system's input signal is given by $x(t) = (2/3)\ t\ \Pi((t - 1.5)/3)$. Find, and sketch, the system's output.

4.6. The impulse response of a time invariant linear system is $h(t) = u(t)/(1 + t^2)$ where $u(t)$ is the

Heaviside step function. Find, and sketch, the response of this system to a rectangular pulse of unit height and width.

4.7. The amplitude response of a rectangular low pass filter is given by $|H(f)| = \Pi(f/(2f_\chi))$. Find, and sketch, the impulse response of this filter if its phase response is: (a) $\phi(f) = 0$; and (b) $\phi(f) = \text{sgn}(f)\pi/2$.

(N.B. Problems 4.8 and 4.9 presuppose some knowledge of elementary circuit theory.)

4.8. Find the frequency response and impulse response of: (a) an *LR*; (b) an *RL*; and (c) a *CR* filter. (The input is across both components in series and the output is across the second, earthed, component alone. *L*, *R* and *C* denote inductors, resistors and capacitors respectively.)

4.9. An electrical system consists of an *RC* potential divider (input across series combination of *RC*, output across *C* alone) followed by an ideal differentiator described by $y(t) = d/dt\,[x(t)]$. For an impulse applied at the input to the potential divider find and sketch: (i) the response at the potential divider output; and (ii) the response at the differentiator output. Find the frequency response of the entire system and use convolution to calculate the system output when the system input is $x(t) = \Pi((t - T/2)/T)$. Sketch the output if the input pulse width equals the time constant of the potential divider, i.e. $T = RC$.

(N.B. Problem 4.10 presupposes some knowledge of elementary circuit theory and electronics.)

4.10. An ideal operational amplifier is driven at its non-inverting input by a signal via an R_1C potential divider. It is driven at its inverting input by the same signal via a series resistor R_2. Negative feedback is applied using a resistor R_3 connected across the operational amplifier's output and inverting input. Find the impulse response and frequency response of this electronic circuit which is commonly used in signal processing.

4.11. A system has an impulse response $h(t) = u(t)\,e^{-t/\tau}$ and an applied input signal $x(t) = \Pi((t - \tau/2)/\tau)t$. Find the system's output signal.

4.12. A raised cosine filter has the amplitude response:

$$|H(f)| = \begin{cases} \frac{1}{2}[1 + \cos(\pi f/2f_\chi)], & |f| \le 2f_\chi \\ 0, & |f| > 2f_\chi \end{cases}$$

Explain (in a few words) why (strictly) this filter is not physically realisable.

4.13. An electrical system consists of 10 cascaded *RC* filters. (Each filter is a potential divider with input across the series combination of *R* and *C* and output across *C* alone.) If operational amplifier impedance buffers are inserted between all *RC* filters deduce (without elaborate calculations) the approximate *shape* of the system's amplitude response. (The impedance buffers merely reduce the loading effect of each *RC* stage on the preceding stage to a negligible level.)

4.14. What is the −3 dB bandwidth of the system with a one sided exponential impulse response $h(t) = u(t)e^{-5t}$? If white Gaussian noise with one sided NPSD of 2.0×10^{-9} V^2/Hz is applied to the input of this system what is the PSD of the noise at the system output? What is the total noise power at the system output? What is the output noise power within the system's −3 dB bandwidth? What is the pdf of the noise at the system output?

4.15. If the noise at the output of the system described in Problem 4.14 is applied (after impedance buffering to avoid loading effects) to a second, identical, system, what will be the total noise power at the (second) system output? What proportion of this total noise power resides in the frequency band below 1.0 Hz? [6.4×10^{-12} V^2, 20%]

4.16. A mobile communications system, consisting of a transmitting mobile and receiving fixed base station, experiences noise at the receiving antenna. Assuming that the noise is spectrally white, and has variance σ^2, calculate the coherence (i.e. autocorrelation) function for the output of a lowpass *RC* filter attached to the antenna output when the spectral density of the transmitted signal is given by:

$$S_{xx}(\omega) = \begin{cases} \tfrac{1}{2}\,[1 + \cos(\pi\omega/10)], & |\omega| < 10 \\ 0, & \text{otherwise} \end{cases}$$

and the square magnitude of the channel frequency response is given by:

$$|H(\omega)|^2 = \frac{A^2}{B^2 + \omega^4}$$

What information does this give you about the system? How might it be measured in practice?

4.17. Find the equivalent noise bandwidth of the finite-time integrator whose impulse response is given by:

$$h(t) = \frac{1}{T}\,[\,u(t) - u(t - T)] \qquad \text{Hint: } \int_0^\infty \frac{\sin^2(ax)}{x^2}\,dx = |a|\,\frac{\pi}{2}$$

4.18. A linear system has the following impulse response: $h(t) = e^{-5t}$, when $t \geq 0$ and $h(t) = 0$ at other times. The input signal to the above system is a sample function from a random process which has the form:

$$X(t) = M, \qquad -\infty < t < \infty$$

in which M is a random variable that is uniformly distributed from -6 to $+18$. Find: (a) an expression for the output sample function; (b) the mean value of the output; and (c) the variance of the output. [$M/5$, 0.2, 1.92]

4.19. Find the cross-correlation function $R_{xy}(\tau)$ for a single-stage low-pass *RC* filter when the input $x(t)$ has the following autocorrelation function:

$$R_{xx}(\tau) = \frac{\beta N_0}{2}\,e^{-\beta|\tau|}, \qquad -\infty < \tau < \infty$$

[0.80 Hz, 1.0×10^{-10} V^2, 5.0×10^{-11} V^2]

4.20. A random variable x has a pdf: $p_X(x) = u(x)5e^{-5x}$ and a statistically independent random variable y has a pdf: $p_Y(y) = 2u(y)e^{-2y}$. For the random variable $Z = X + Y$ find: (a) $p_Z(0)$; (b) the modal value of z (i.e. that for which $p_Z(z)$ is a maximum); and (c) the probability that $z > 1.0$. [3.33 $\exp(-2z)\,(1 - \exp(-3z))$, 0, 0.305, 0.22]

4.21. Find the pdf of noise at the output of a full wave rectifier if Gaussian noise with a variance of 1 V^2 is present at its input. (Note the output $y(t)$ of a full wave rectifier is related to its input $x(t)$ by $y(t) = |x(t)|$.)

4.22. A signal with uniform pdf: $p_X(x) = 0.5\,\Pi((x - 1)/2)$ is processed by a, non-linear, memoryless system with input/output characteristic: $y(t) = 5x(t) + 2$. What is the pdf of the output signal?

4.23. A signal with pdf $p_X(x) = 1/(\pi(1 + x^2))$ is processed by a square law detector (characteristic $Y = X^2$). What is the pdf of processed signal?

Part Two

Digital communications principles

Part Two, by far the largest part of the book, uses the theoretical concepts of Part One to describe and analyse communications links which are robust in the presence of noise and other impairment mechanisms.

Chapter 5 starts with a discussion of sampling and aliasing and demonstrates the practical problems associated with representing an analogue signal in digital (pulse code modulated) form. This highlights the care that must be taken to achieve accurate reconstruction, without distortion, of the original analogue signal in a receiver. Chapter 5 also describes a variety of techniques by which the bandwidth of a PCM signal may be reduced in order to allow the effective use of bandlimited channels. Chapter 6 addresses the fundamentals of binary baseband transmission, covering important aspects of practical decision theory, and describes how the spectral properties of baseband digital signals can be altered by the use of different line coding schemes. Receiver equalisation, employed to overcome transmission channel distortion, is also discussed. Probability theory, is applied to receiver detection and decision processes in Chapter 7, along with a discussion of the Bayes and Neyman-Pearson decision criteria. Chapter 8 investigates optimum pulse shaping at the transmitter, and optimum filtering at the receiver, designed to minimise transmitted signal bandwidth whilst maximising the probability of correct symbol decisions.

Chapter 9 presents the fundamentals of information theory and source coding, introducing the important concepts of entropy and coding efficiency. It is shown how redundancy present in source data may be minimized, using variable length coding schemes to achieve efficient, low bit rate, digital speech, and other, signals. Chapter 10 describes the converse technique, in which transmitted data redundancy is increased to achieve error correction, or detection, in the presence of noise.

Chapter 11 analyses the bandpass binary modulation schemes which employ amplitude, frequency, or phase shift keying. Variants, and hybrid combinations, of these schemes are then examined which are especially spectrally efficient (e.g. QAM), are especially power efficient (e.g. MFSK), or have some other desirable property. Following a detailed discussion of the sources of noise in electronic circuits, Chapter 12 outlines the calculation of received signal power, noise power, and associated signal-to-noise ratio, for simple communications links. Finally Chapter 13 indicates how the performance of a complex communication system can be predicted by simulation before any hardware prototyping is attempted.