

Signals And Systems by Alan V. Oppenheim: Notes

Malcolm

Started 17 April 2025

Contents

1	Introduction	2
1.1	Signal Energy and Power	2
1.2	Even and Odd signals	5
1.3	Differences between continuous and discrete periodic complex exponentials	6
1.4	Intuition for discrete-time periodicity	10
1.5	Difference in harmonic relations in discrete and continuous periodic exponentials	12
1.6	More on complex exponential and sinusoidal signals	13
1.7	Unit impulse and Unit step functions	15
1.8	Basic system properties 1	21
1.9	Basic system properties 2	26
2	Linear Time-Invariant Systems	32
2.1	Discrete-time convolution sum	32
2.2	Discrete-time convolution sum—examples	37
2.3	Continuous-time—the convolution integral	45
2.4	Convolution integral—examples	49
2.5	Properties of LTI systems 1	54
2.6	Properties of LTI systems 2	58
2.7	The Unit Step Response of an LTI system	65
2.8	Causal LTI systems described by differential and difference equa- tions	66
2.9	Block diagram representations of first-order systems described by differential and difference equations	70
2.10	The Unit impulse as an idealised short pulse	73
2.11	Defining the unit impulse through convolution	76
A	Other proofs	78
A.1	Sum of geometric series	78

Chapter 1

Introduction

1.1 Signal Energy and Power

Motivation and definition

In many but not all, applications, the signals considered directly related to physical quantities capturing power and energy in a physical system. (for instance v^2/R for the power across a resistor)

As such it is a common and worthwhile convention to use similar terminology for power and energy for *any* continuous-time signal, denoted $x(t)$, or any discrete-time signal $x[n]$. In this case, the total energy over the time interval $t_1 \leq t \leq t_2$ in a continuous signal $x(t)$ is defined as

$$\int_{t_1}^{t_2} |x(t)|^2 dt$$

where $|x|$ denotes the magnitude of the (possibly complex) number x ; see that the time-averaged signal can be obtained by dividing by $(t_2 - t_1)$. Similarly for a discrete signal $x[n]$ over the interval $n_1 \leq n \leq n_2$ the total energy is

$$\sum_{n=n_1}^{n_2} |x[n]|^2$$

with the average power calculated by dividing by $(n_2 - n_1 + 1)$.

It is important to remember that the terms ‘power’ and ‘energy’ are used here *independently* of their relation to physical energy (they clearly don’t correlate since their units or scalings would differ). Nevertheless we will find it convenient to use these terms in a general fashion.

Power and energy over infinite intervals

Considering signals over an infinite time interval, meaning for $-\infty < t < +\infty$ or $-\infty < n < +\infty$. Here we define the total energy as the limits of the aforementioned equations increase without bound; in continuous time,

$$E_{\infty} \triangleq \lim_{T \rightarrow \infty} \int_{-T}^T |x(t)|^2 dt = \int_{-\infty}^{+\infty} |x(t)|^2 dt$$

and in discrete time,

$$E_{\infty} \triangleq \lim_{N \rightarrow \infty} \sum_{n=-N}^{+N} |x[n]|^2 = \sum_{n=-\infty}^{+\infty} |x[n]|^2$$

Note that these expressions may not converge; for instance say $x(t)$ or $x[n]$ equal some nonzero constant for all time: such signals have infinite energy, while signals with $E_{\infty} < \infty$ have finite energy.

(next page)

Cont.

Analagously, we can define the time-averaged power over an infinite interval as

$$P_{\infty} \triangleq \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T |x(t)|^2 dt$$

and

$$P_{\infty} \triangleq \lim_{N \rightarrow \infty} \frac{1}{2N+1} \sum_{n=-N}^{+N} |x[n]|^2$$

In continuous and discrete time respectively.

Intuition

See that with these definitions, we can identify three classes of signals: first those with finite total energy, meaning $E_{\infty} < \infty$. See that such a signal would have zero average power:

$$P_{\infty} = \lim_{T \rightarrow \infty} \frac{E_{\infty}}{2T} = 0$$

Second would be signals with finite average power P_{∞} ; see from the above expression that for $P_{\infty} > 0$, this requires that $E_{\infty} = \infty$.

Last would be signals for which neither P_{∞} nor E_{∞} are finite. An example of this might be $x(t) = t$.

Note on discrete signals

It is important to note that the discrete-time signal $x[n]$ is defined *only* for *integer* values of the independent variable.

1.2 Even and Odd signals

Definition

A continuous-time signal is *even* if

$$x(-t) = x(t)$$

while a discrete-time signal is *even* if

$$x[-n] = x[n]$$

These signals are referred to as *odd* if

$$x(-t) = -x(t)$$

$$x[-n] = -x[n]$$

Note that an odd signal must be 0 at $t = 0$ or $n = 0$ since the equations require that $x(0) = -x(0)$ and $x[0] = -x[0]$.

Decomposition

An important fact is that any signal can be broken into a sum of two signals, where one is even and the other odd. To see this, consider

$$\text{Ev}\{x(t)\} = \frac{1}{2}[x(t) + x(-t)]$$

which is referred to as the *even part* of $x(t)$. Similarly, the *odd part* of $x(t)$ is given by

$$\text{Od}\{x(t)\} = \frac{1}{2}[x(t) - x(-t)]$$

See that $x(t)$ is the sum of the two. Exactly analogous definitions hold in the discrete time case.

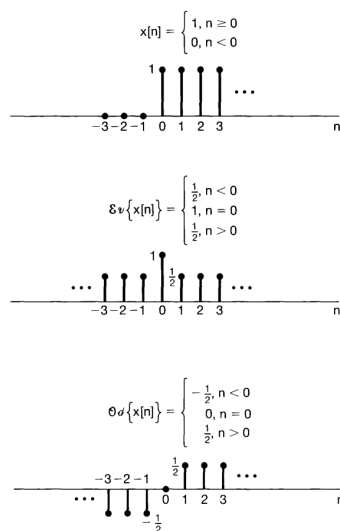


Figure 1.18 Example of the even-odd decomposition of a discrete-time signal.

1.3 Differences between continuous and discrete periodic complex exponentials

The continuous-time *complex exponential signal* is of the form

$$x(t) = Ce^{at}$$

where C and a are, in general, complex numbers. An important class of complex exponentials is obtained by constraining a to be purely imaginary:

$$x(t) = e^{i\omega t}$$

Periodicity and harmonic relations (purely imaginary power)

An important property of this signal is that it is periodic; recall that $x(t)$ will be periodic with period T if

$$e^{i\omega t} = e^{i\omega(t+T)}$$

this means

$$e^{i\omega(t+T)} = e^{i\omega t} e^{i\omega T} \implies e^{i\omega T} = 1$$

If $\omega = 0$ then this is satisfied for any T . If $\omega \neq 0$, see that the *fundamental period* T_0 of $x(t)$ —that is, the smallest positive value of T for which this holds—is

$$T_0 = \frac{2\pi}{|\omega|}$$

(the signals $e^{i\omega_0 t}$ and $e^{-i\omega_0 t}$ have the same fundamental period) Naturally, there is a set of exponentials periodic to a common period T_0 . These are said to be *harmonically related* complex exponentials; the necessary condition they satisfy is

$$e^{i\omega T_0} = 1$$

which implies that

$$\omega T_0 = 2\pi k, \quad k = 0, \pm 1, \pm 2, \dots$$

(next page)

Cont.

We had

$$\omega T_0 = 2\pi k, \quad k = 0, \pm 1, \pm 2, \dots$$

if we define

$$\omega_0 = \frac{2\pi}{T_0}$$

this means that the harmonic frequencies ω must be integer multiples of ω_0 :

$$\phi_k(t) = e^{ik\omega_0 t}, \quad k = 0, \pm 1, \pm 2, \dots$$

For $k = 0$, $\phi_k(t)$ is a constant, while for any other value of k , $\phi_k(t)$ is periodic with fundamental frequency $|k|\omega_0$ and fundamental period

$$\frac{2\pi}{|k|\omega_0} = \frac{T_0}{|k|}$$

Each $\phi_k(t)$ itself defines a fundamental frequency and a corresponding fundamental period. (see that $|k|\omega_0 \cdot T_0/|k| = 2\pi$, so this scaled down period is the corresponding period for this scaled up frequency. Each frequency is unique, point here is that they are also periodic with T_0 , but with fundamental periods getting proportionally smaller.)

Note that the k th harmonic $\phi_k(t)$ is still periodic with T_0 ; it goes through exactly $|k|$ of its fundamental periods during any time interval of length T_0 . (the term ‘harmonic’ is consistent with its use in music, where it refers to tones resulting from variations in acoustic pressure at frequencies that are integer multiples of a fundamental frequency)

(next page)

Discrete case

As in continuous time, an important signal in discrete time is the *complex exponential signal*, defined as

$$x[n] = C\alpha^n$$

where C and α are, in general, complex numbers. See that this could also be expressed as

$$x[n] = Ce^{\beta n}$$

where $\alpha = e^\beta$. See that we can constrain β to be purely imaginary:

$$x[n] = e^{i\omega_0 n}$$

Periodicity properties of Discrete-time complex exponentials

While there are many similarities between continuous and discrete-time signals, there are a number of important differences. For the continuous time signal $e^{i\omega_0 t}$, we know that

- The larger the magnitude of ω_0 , the higher the rate of oscillation of the signal
- $e^{i\omega_0 t}$ is periodic for any value of ω_0

These properties are different in the discrete-time case.

Given the first property, consider the discrete-time complex exponential with frequency $\omega_0 + 2\pi$:

$$e^{i(\omega_0+2\pi)n} = e^{i2\pi n}e^{i\omega_0 n} = e^{i\omega_0 n}$$

(see that this is a direct result of the fact that we iterate through discrete time as integers) The exponential at frequency $\omega_0 + 2\pi$ is the *same* as that at frequency ω_0 . This is unlike the continuous-time case where each distinct ω_0 represents a distinct signal.

In discrete time, the signal with frequency ω_0 is identical to the signals with frequencies $\omega_0 \pm 2\pi, \omega_0 \pm 4\pi$, and so on. Therefore when considering discrete time complex exponentials, see that we need only consider a frequency interval of length 2π in which to choose ω_0 , such as $0 \leq \omega_0 < 2\pi$ or $-\pi \leq \omega_0 < \pi$.

Also see that because of this the discrete exponential $e^{i\omega_0 n}$ does *not* have a continually increasing rate of oscillation as ω_0 increases in magnitude; the signals will oscillate faster until we reach $\omega_0 = \pi$, after which the rate of oscillation decreases until we reach $\omega_0 = 2\pi$, at which the same constant sequence as $\omega_0 = 0$ is produced.

(next page)

Cont.

The second property we wish to consider concerns the periodicity of the discrete time complex exponential. In order for the signal $e^{i\omega_0 n}$ to be periodic with period $N > 0$ we must have

$$e^{i\omega_0(n+N)} = e^{i\omega_0 n}$$

or equivalently

$$e^{i\omega_0 N} = 1$$

For this to hold, $\omega_0 N$ must be a multiple of 2π . That is, there must be an integer m such that

$$\omega_0 N = 2\pi m$$

or equivalently

$$\frac{\omega_0}{2\pi} = \frac{m}{N}$$

The signal $e^{i\omega_0 n}$ is periodic if $\omega_0/2\pi$ is a rational number and is not periodic otherwise.

Fundamental period

Recall the idea of a *fundamental period*; in this case it would mean the smallest N such that $\omega_0 N = 2\pi m$ holds (this is unlike the continuous case where there is always some T where $\omega_0 T = 2\pi$); see that this occurs when m and N do not have any factors in common.

See that from this we can derive a *fundamental frequency* as

$$\frac{2\pi}{N} = \frac{\omega_0}{m}$$

(see that this frequency is always equal or lower—intuitively, to have a different wave that completes one oscillation in N time, its frequency will either be equal or lower)

To summarize

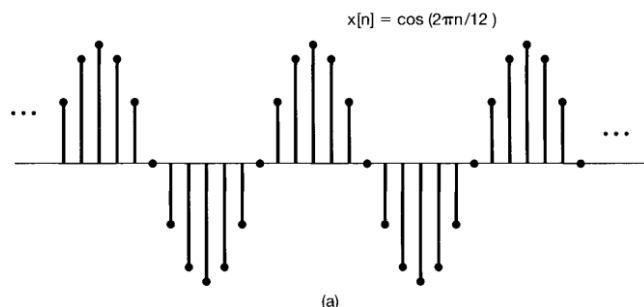
TABLE 1.1 Comparison of the signals $e^{i\omega_0 t}$ and $e^{i\omega_0 n}$.

$e^{i\omega_0 t}$	$e^{i\omega_0 n}$
Distinct signals for distinct values of ω_0	Identical signals for values of ω_0 separated by multiples of 2π
Periodic for any choice of ω_0	Periodic only if $\omega_0 = 2\pi m/N$ for some integers $N > 0$ and m .
Fundamental frequency ω_0	Fundamental frequency* ω_0/m
Fundamental period $\omega_0 = 0$: undefined $\omega_0 \neq 0$: $\frac{2\pi}{\omega_0}$	Fundamental period* $\omega_0 = 0$: undefined $\omega_0 \neq 0$: $m \left(\frac{2\pi}{\omega_0} \right)$

*Assumes that m and N do not have any factors in common.

1.4 Intuition for discrete-time periodicity

Consider the sequence $x[n] = \cos(2\pi n/12)$:



we can think of this as a set of samples of the continuous-time sinusoid $x(t) = \cos(2\pi t/12)$ at integer time points. In this case, see that both $x(t)$ and $x[n]$ are periodic with fundamental period 12. That is, the values of $x[n]$ repeat every 12 points, exactly in step with the fundamental period of $x(t)$.

Now consider the signal $x[n] = \cos(8\pi n/31)$:



This can also be viewed as a set of samples of $x(t) = \cos(8\pi t/31)$ at integer points in time. But now see that in this case $x(t)$ is periodic with fundamental period $31/4$, while $x[n]$ is periodic with fundamental period 31.

This difference stems from the fact that the discrete-time signal is defined only for integer values of the independent variable—there is no sample at time $t = 31/4$, when $x(t)$ completes one period, or at $t = 2 \cdot 31/4$ or $t = 3 \cdot 31/4$, when $x(t)$ has completed two or three periods. Only at sample $t = 4 \cdot 31/4 = 31$, when $x(t)$ has completed *four* periods is the discrete sequence defined.

This manifests as the pattern of $x[n]$ not repeating with each cycle of positive and negative values, but rather only after four of such cycles, specifically 31 points.

(next page)

Cont.

Finally consider the signal $x[n] = \cos(n/6)$:



In this case, the values of $x(t)$ at integer sample points *never repeat*, as these sample points never span an interval that is an exact multiple of the period, 12π , of $x(t)$.

Thus, $x[n]$ is *not periodic*, although the eye visually interpolates between the sample points, suggesting *the envelope* $x(t)$ which is periodic.

1.5 Difference in harmonic relations in discrete and continuous periodic exponentials

As in continuous time, it is also of considerable value in discrete-time to consider sets of harmonically related periodic exponentials—that is, *periodic exponentials with a common period N* .

We know that these are precisely the signals which are at frequencies which are multiples of $2\pi/N$; that is

$$\phi_k[n] = e^{ik(2\pi/N)n}, \quad k = 0, \pm 1, \dots$$

In the continuous-time case, all the harmonically related complex exponentials $e^{ik(2\pi/T_0)t}$, $k = 0, \pm 1, \pm 2, \dots$ are distinct. However, recall that for discrete signals we have

$$e^{i(\omega_0 + 2\pi)n} = e^{i2\pi n} e^{i\omega_0 n} = e^{i\omega_0 n}$$

(this is a direct result of the fact that we iterate through discrete time as integers) As such the harmonically related complex exponentials *are not all unique in discrete time*; specifically,

$$\begin{aligned} \phi_{k+N}[n] &= e^{i(k+N)(2\pi/N)n} \\ &= e^{ik(2\pi/N)n} e^{i2\pi n} = \phi_k[n] \end{aligned}$$

See that this implies that there are only N distinct periodic exponentials in the set of $\phi_k[n]$; meaning

$$\phi_0[n] = 1, \phi_1[n] = e^{i(2\pi/N)n}, \phi_2[n] = e^{i2(2\pi/N)n}, \dots, \phi_{N-1}[n] = e^{i(N-1)(2\pi/N)n}$$

are all distinct, but any other $\phi_k[n]$ would just be identical to one of them. (for instance $\phi_N[n] = \phi_0[n]$ or $\phi_{-1}[n] = \phi_{N-1}[n]$.)

1.6 More on complex exponential and sinusoidal signals

Continuous case

A continuous-time *complex exponential signal* is of the form

$$x(t) = Ce^{at}$$

where C and a are, in general, complex numbers.

Euler identity and ‘combined’ sinusoidal form

Recall euler’s identity:

$$e^{i\omega_0 t} = \cos(\omega_0 t) + i \sin(\omega_0 t)$$

See that the scaled and phase-delayed sinusoid can be written in terms of these periodic complex exponentials with the same fundamental period:

$$A \cos(\omega_0 t + \phi) = \frac{A}{2} e^{i\phi} e^{i\omega_0 t} + \frac{A}{2} e^{-i\phi} e^{-i\omega_0 t}$$

We can also express

$$A \cos(\omega_0 t + \phi) = A \operatorname{Re}\{e^{i(\omega_0 t + \phi)}\}$$

and

$$A \sin(\omega_0 t + \phi) = A \operatorname{Im}\{e^{i(\omega_0 t + \phi)}\}$$

Energy and power

Periodic signals—and in particular, the complex periodic exponential signal—are examples of signals with infinite total energy but finite average power. Calculating the total energy and of the periodic exponential signal over one period:

$$\begin{aligned} E_{\text{period}} &= \int_0^{T_0} |e^{i\omega_0 t}|^2 dt \\ &= \int_0^{T_0} 1 dt = T_0 \end{aligned}$$

(The absolute value of a complex number is its magnitude. Think of the absolute value as the (possibly multidimensional) distance from zero.) Calculating the average power:

$$P_{\text{period}} = \frac{1}{T_0} E_{\text{period}} = 1$$

Since there are an infinite number of periods as t ranges from $-\infty$ to $+\infty$, the total energy integrated over all time is infinite. However, since the average power over each period is 1, averaging over multiple periods always yields an average power of 1. That is, the complex periodic exponential signal has finite average power equal to

$$P_{\infty} = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T |e^{i\omega_0 t}|^2 dt = 1$$

(next page)

General continuous complex exponential signals

In the most general case Ce^{at} where both C and a are complex, see that since C and a can be just

$$C = |C|e^{i\theta}, \quad a = r + i\omega_0$$

we can express the general complex signal as

$$Ce^{at} = |C|e^{i\theta}e^{(r+i\omega_0)t} = |C|e^{rt}e^{i(\omega_0 t + \theta)}$$

we can expand this further as

$$Ce^{at} = |C|e^{rt} \cos(\omega_0 t + \theta) + i|C|e^{rt} \sin(\omega_0 t + \theta)$$

Discrete case

As in continuous time, the discrete time *complex exponential signal* is defined by

$$x[n] = C\alpha^n$$

where C and α are, in general, complex numbers. This could alternatively be expressed in the form

$$x[n] = Ce^{\beta n}$$

where $\alpha = e^\beta$.

Euler identity and ‘combined’ sinusoidal form

As with the continuous case, constraining β to be purely imaginary, we have Euler’s identity

$$e^{i\omega_0 n} = \cos \omega_0 n + i \sin \omega_0 n$$

and

$$A \cos(\omega_0 n + \phi) = \frac{A}{2} e^{i\phi} e^{i\omega_0 n} + \frac{A}{2} e^{-i\phi} e^{-i\omega_0 n}$$

General discrete complex exponential signals

As with the continuous case, for complex C and α , we have

$$C = |C|e^{i\theta}, \quad \alpha = |\alpha|e^{i\omega_0}$$

so the general complex exponential signal can be expressed as

$$C\alpha^n = |C||\alpha|^n \cos(\omega_0 n + \theta) + i|C||\alpha|^n \sin(\omega_0 n + \theta)$$

1.7 Unit impulse and Unit step functions

Discrete-Time

One of the simplest discrete-time signals is the *unit impulse/unit sample*, defined as

$$\delta[n] = \begin{cases} 0, & n \neq 0 \\ 1, & n = 0 \end{cases}$$



Figure 1.28 Discrete-time unit impulse (sample).

Another basic discrete-time signal is the discrete-time *unit step*, denoted by $u[n]$ and defined by

$$u[n] = \begin{cases} 0, & n < 0 \\ 1, & n \geq 0 \end{cases}$$



Figure 1.29 Discrete-time unit step sequence.

See that the discrete-time unit impulse is the *first difference* of the discrete-time step:

$$\delta[n] = u[n] - u[n-1]$$

Conversely, the discrete-time unit step is the *running sum* of the unit sample:

$$u[n] = \sum_{m=-\infty}^n \delta[m]$$



Figure 1.30 Running sum of eq. (1.66): (a) $n < 0$; (b) $n > 0$.

Since the only nonzero value of the unit sample is at 0, the running sum is 0 for $n < 0$ and 1 for $n \geq 0$.

(next page)

Alternative form

we had the discrete-time unit step as the running sum of the unit sample:

$$u[n] = \sum_{m=-\infty}^n \delta[m]$$

See that by changing the variable of summation from m to $k = n - m$, we can rewrite this as

$$u[n] = \sum_{k=-\infty}^0 \delta[n - k]$$

and equivalently

$$u[n] = \sum_{k=0}^{\infty} \delta[n - k]$$

An interpretation of this is a superposition of delayed impulses; we can view the unit step as the sum of unit impulses $\delta[n]$ (nonzero at $n = 0$), $\delta[n - 1]$ (nonzero at $n = 1$), and all other $\delta[n - k]$ for integer k extending to infinity.

Sampling property

See that the unit impulse can also be used to sample the value of a signal at $n = 0$; since $\delta[n]$ is nonzero (and equal to 1) only for $n = 0$, it follows that

$$x[n]\delta[n] = x[0]\delta[n]$$

More generally, if we consider a unit impulse $\delta[n - n_0]$ at $n = n_0$, then

$$x[n]\delta[n - n_0] = x[n_0]\delta[n - n_0]$$

(next page)

Continuous-Time

The continuous-time *unit step function* $u(t)$ is defined in a similar manner to its discrete-time counterpart:

$$u(t) = \begin{cases} 0, & t < 0 \\ 1, & t > 0 \end{cases}$$



Figure 1.32 Continuous-time unit step function.

Note that the unit step is *discontinuous* at $t = 0$. The continuous-time *unit impulse* $\delta(t)$ is related to the unit step in a manner analagous to that of their discrete counterparts; in particular, the continuous-time unit step is the *running integral* of the unit impulse:

$$u(t) = \int_{-\infty}^t \delta(\tau) d\tau$$

It also follows that the continuous-time unit impulse can be thought of as the *first derivative* of the continuous-time unit step:

$$\delta(t) = \frac{du(t)}{dt}$$

In contrast to discrete-time, there is some formal difficulty with this equation as a representation of the unit impulse—since $u(t)$ is discontinuous at $t = 0$ and consequently is not formally differentiable.

(next page)

Cont.

We get around this by considering an approximation to the unit step $u_{\Delta}(t)$, rising from the value 0 to 1 in a short time interval of length Δ :



Figure 1.33 Continuous approximation to the unit step, $u_{\Delta}(t)$.

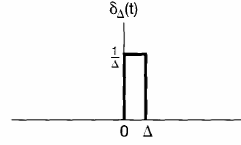


Figure 1.34 Derivative of $u_{\Delta}(t)$.

Also considering the derivative:

$$\delta_{\Delta}(t) = \frac{du_{\Delta}(t)}{dt}$$

The unit step changes from value 0 to 1 instantaneously and can be thought of an idealisation of $u_{\Delta}(t)$ for Δ so short that its duration doesn't matter for any practical purpose. Formally, $u(t)$ is the limit of $u_{\Delta}(t)$ as $\Delta \rightarrow 0$.

Consider the derivative again; $\delta_{\Delta}(t)$ is a short pulse, of duration Δ and unit area for any value of Δ . As $\Delta \rightarrow 0$, $\delta_{\Delta}(t)$ becomes narrower and higher while maintaining its unit area. Its limiting form

$$\delta(t) = \lim_{\Delta \rightarrow 0} \delta_{\Delta}(t)$$

can then be thought of as an idealisation of the short pulse $\delta_{\Delta}(t)$ as the duration Δ becomes insignificant:



Figure 1.35 Continuous-time unit impulse.



Figure 1.36 Scaled impulse.

$\delta(t)$ has, in effect, no duration but unit area. The arrow at $t = 0$ indicates the area of the pulse is concentrated at $t = 0$ and the height of the arrow and the '1' next to the arrow is used to represent the *area* of the impulse. More generally, a scaled impulse $k\delta(t)$ will have an area k , and thus

$$ku(t) = \int_{-\infty}^t k\delta(\tau)d\tau$$

A scaled impulse, where the height of the arrow is chosen to be proportional to the area of the impulse.

(next page)

Alternative form

As with discrete time, the relationship between the continuous time unit step and impulse can be rewritten in a different form, by changing the variable of integration from τ to $\sigma = t - \tau$:

$$u(t) = \int_{-\infty}^t \delta(\tau) d\tau = \int_{\infty}^0 \delta(t - \sigma)(-d\sigma)$$

or equivalently

$$u(t) = \int_0^{\infty} \delta(t - \sigma) d\sigma$$

(This negation after inversion of the limits of the integral can be derived from the first fundamental theorem. For a more intuitive understanding, consult the definition of the Riemann sum:

$$\sum_{i=0}^{n-1} f(t_i)(x_{i+1} - x_i)$$

Considering the $d\tau$, or respectively $d\sigma$, as the width of the step between two arguments of the sum, if we change the direction in which we integrate, the step also changes its sign. In the discrete sum, the summands are not multiplied by this step.)

Sampling property

As with discrete-time, the continuous-time impulse has a very important sampling property; consider

$$x_1(t) = x(t)\delta_{\Delta}(t)$$

By construction, $x_1(t)$ is zero outside the interval $0 \leq t \leq \Delta$. See that for Δ sufficiently small so that $x(t)$ is approximately constant over this interval:

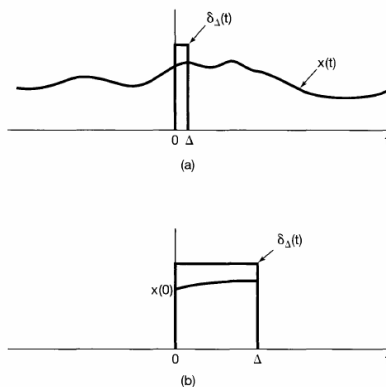


Figure 1.39 The product $x(t)\delta_{\Delta}(t)$: (a) graphs of both functions; (b) enlarged view of the nonzero portion of their product.

so we have

$$x(t)\delta_{\Delta}(t) \approx x(0)\delta_{\Delta}(t)$$

(next page)

Cont.

We had, for small Δ ,

$$x(t)\delta_{\Delta}(t) \approx x(0)\delta_{\Delta}(t)$$

since $\delta(t)$ is the limit as $\Delta \rightarrow 0$ of $\delta_{\Delta}(t)$, it follows that

$$x(t)\delta(t) = x(0)\delta(t)$$

By the same argument, we have an analogous expression for an impulse concentrated at an arbitrary point, say t_0 . That is,

$$x(t)\delta(t - t_0) = x(t_0)\delta(t - t_0)$$

1.8 Basic system properties 1

General form

A general formula for a continuous first-order linear differential equation is

$$\frac{dy(t)}{dt} + ay(t) = bx(t)$$

where $x(t)$ is the input, $y(t)$ the output, and a, b constants. An example of this might be

$$\frac{dv(t)}{dt} + \frac{\rho}{m}v(t) = \frac{1}{m}f(t)$$

Discrete cases have general first-order linear difference equations of the form

$$y[n] + ay[n-1] = bx[n]$$

Considering the earlier example, if we let $v[n] = v(n\Delta)$ and $f[n] = f(n\Delta)$ and approximate $dv(t)/dt$ at $t = n\Delta$ by the first backward difference:

$$\frac{v(n\Delta) - v((n-1)\Delta)}{\Delta}$$

we can obtain

$$\begin{aligned}\frac{v[n] - v[n-1]}{\Delta} + \frac{\rho}{m}v[n] &= \frac{1}{m}f[n] \\ v[n] - v[n-1] + \frac{\rho\Delta}{m}v[n] &= \frac{\Delta}{m}f[n] \\ v[n] \left(1 + \frac{\rho\Delta}{m}\right) - v[n-1] &= \frac{\Delta}{m}f[n] \\ v[n] \frac{m + \rho\Delta}{m} - v[n-1] &= \frac{\Delta}{m}f[n] \\ v[n] - \frac{m}{m + \rho\Delta}v[n-1] &= \frac{\Delta}{m + \rho\Delta}f[n]\end{aligned}$$

which is in the general form as described above.

(next page)

Basic system properties

Memory

A system is said to be *memoryless* if its output for each value of the independent variable is dependent only on the *input at that same time*.

For instance, the system

$$y[n] = (2x[n] - x^2[n])^2$$

is memoryless, since the value of $y[n]$ at any particular time n_0 depends only on the value of $x[n]$ at that time. Other examples of memoryless systems include the input-output relationship of a resistor, with input $x(t)$ taken to be current and output $y(t)$ voltage:

$$y(t) = Rx(t)$$

where R is the resistance. The *identity system*, whose output is identical to its input, is also memoryless

$$y(t) = x(t)$$

Written in discrete time:

$$y[n] = x[n]$$

An example of a discrete-time system *with memory* is an *accumulator* or *summer*

$$y[n] = \sum_{k=-\infty}^n x[k]$$

see that the accumulator must store or ‘remember’ information about the past inputs up to current time. Another example of a system with memory is a *delay*

$$y[n] = x[n - 1]$$

While the concept of memory in a system would typically suggest storing *past* input and output values, our formal definition also leads to our referring to a system as having memory if the current output is dependent on *future* values of the input and output (noncausal systems).

(next page)

Causality

A system is *causal* if the output at any time depends only on values of the input at the present time and in the past. Such a system is often referred to as being *nonanticipative*, as the system output does not ‘anticipate’ future values of the input.

Consequently, if two inputs to a causal system are identical up to some point in time t_0 or n_0 , the corresponding outputs must also be equal up to this same time. For instance the capacitor:

$$y(t) = \frac{1}{C} \int_{-\infty}^t x(\tau) d\tau$$

is a causal system, with input taken to be current, output voltage, and C capacitance. The capacitor voltage responds only to the present and past values of the input.

Examples of *noncausal* systems might be

$$y[n] = x[n] - x[n+1]$$

or

$$y(t) = x(t+1)$$

See that all memoryless systems are causal, since the output responds only to the current value of the input in such systems.

An example of a noncausal system might be when averaging data over an interval to smooth out fluctuations and keep only the trend. Such an averaging system might look like

$$y[n] = \frac{1}{2M+1} \sum_{k=-M}^{+M} x[n-k]$$

More examples

Consider

$$y[n] = x[-n]$$

see that for $n < 0$ the output depends on a future value of the input, and hence the system is not causal. Another example: consider

$$y(t) = x(t) \cos(t+1)$$

This can be rewritten as

$$y(t) = x(t)g(t)$$

Thus only the current value of the input $x(t)$ influences the current value of the output $y(t)$. This system is causal (and, in fact, memoryless).
(next page)

Invertibility and inverse systems

A system is said to be *invertible* if distinct inputs lead to distinct outputs. If a system is invertible, then an *inverse* system exists that, when cascaded with the original system, yields an output $w[n]$ equal to the input $x[n]$ in the first system:



Figure 1.45 Concept of an inverse system for: (a) a general invertible system; (b) the invertible system described by eq. (1.97); (c) the invertible system defined in eq. (1.92).

See that the series interconnection between the invertible system and its inverse system has an overall input-output relationship same as that for the identity system. As illustrated in the second graph, an example of an invertible continuous-time system is

$$y(t) = 2x(t)$$

for which the inverse system is

$$w(t) = \frac{1}{2}y(t)$$

Another example is illustrated in the third block diagram, see that the accumulator is an invertible system:

$$y[n] = \sum_{k=-\infty}^n x[k]$$

with inverse system

$$w[n] = y[n] - y[n-1]$$

Examples of noninvertible systems include

$$y[n] = 0$$

or

$$y(t) = x^2(t)$$

where a single output can correspond to different inputs.
(next page)

Stability

Another important system property is *stability*. Informally, a stable system is one in which small inputs lead to responses that do not diverge (grow uncontrollably).

More formally, if the input to a stable system is bounded (if its magnitude of the input does not grow without bound), then the output must also be bounded and therefore cannot diverge. (this is the definition to take note of)

For instance consider the averaging system brought up earlier

$$y[n] = \frac{1}{2M+1} \sum_{k=-M}^{+M} x[n-k]$$

Suppose that the input $x[n]$ is bounded in magnitude by some number, say B for all values of n . Then the largest possible magnitude for $y[n]$ is also B since it is the average. Therefore, $y[n]$ is bounded and the system is stable.

On the other hand, the accumulator sums all of the past values of the input, so the output will grow continually even if $x[n]$ is bounded—it is an unstable system.

A useful strategy to verify that a system is unstable is to look for a specific bounded input that leads to an unbounded output. For instance consider

$$y(t) = tx(t)$$

See that a constant input $x(t) = 1$ yields $y(t) = t$, which is unbounded. Since for any finite constant bound, $|y(t)|$ will exceed that bound at some t , this system is unstable.

Now consider a different system

$$y(t) = e^{x(t)}$$

See that for an arbitrary positive number B , if $x(t)$ is bound by B , that is

$$|x(t)| < B$$

or

$$-B < x(t) < B$$

for all t , then $y(t)$ must satisfy

$$e^{-B} < |y(t)| < e^B$$

Any input to this system bounded by an arbitrary positive number B has a output guaranteed to be bounded by e^B —this system is stable.

1.9 Basic system properties 2

Time invariance

A system is *time invariant* if a time shift in the input signal results in an identical time shift in the output signal.

That is, if $y[n]$ is the output of a discrete-time, time-invariant system with input $x[n]$, then $y[n - n_0]$ is the output when $x[n - n_0]$ is applied. In continuous time with $y(t)$ the output corresponding to the input $x(t)$, a time-invariant system will have $y(t - t_0)$ as the output when $x(t - t_0)$ is the input.

Examples

Consider the continuous-time system

$$y(t) = \sin[x(t)]$$

To check that this system is time invariant, we must determine whether the time-invariance property holds for *any* input and *any* time shift t_0 . Thus, letting $x_1(t)$ be an arbitrary input to the system, with corresponding output

$$y_1(t) = \sin[x_1(t)]$$

Now consider a second input x_2 obtained by shifting $x_1(t)$ in time

$$x_2(t) = x_1(t - t_0)$$

The output corresponding to this input is

$$y_2(t) = \sin[x_2(t)] = \sin[x_1(t - t_0)]$$

Now consider translating the output y_1 , see that

$$y_1(t - t_0) = \sin[x_1(t - t_0)]$$

Since y_2 , the output of the shifted input, is the same as $y_1(t - t_0)$, which is if we had shifted the output instead, this system is time invariant.
(next page)

More examples

As a second example, consider

$$y[n] = nx[n]$$

See that time shifting the input doesn't correspond with an equivalent shift in the output—this is a time-varying system. This system represents one with a time-varying gain, even if we know the input value, we cannot determine the output value without knowing the current time.

For a counterexample, consider having input $x_1[n] = \delta[n]$, which yields an output $y_1[n] = 0$ (since $n\delta[n] = 0$). However, the input $x_2[n] = \delta[n - 1]$ yields the output $y_2[n] = n\delta[n - 1] = \delta[n - 1]$ —while $x_2[n]$ is a shifted version of $x_1[n]$, $y_2[n]$ is *not* a shifted version of $y_1[n]$.

For a final example, consider the system

$$y(t) = x(2t)$$

This system represents a time scaling. That is, $y(t)$ is a time-compressed (by a factor of 2) version of $x(t)$. Intuitively then, any time shift in the input will also be compressed by a factor of 2, and it is for this reason that the system is not time invariant. Consider a counterexample:



Figure 1.47 (a) The input $x_1(t)$ to the system in Example 1.16; (b) the output $y_1(t)$ corresponding to $x_1(t)$; (c) the shifted input $x_2(t) = x_1(t - 2)$; (d) the output $y_2(t)$ corresponding to $x_2(t)$; (e) the shifted signal $y_1(t - 2)$. Note that $y_2(t) \neq y_1(t - 2)$, showing that the system is not time invariant.

see that if we shift the input signal by 2, the resulting output is not the same as if we had shifted the output signal by 2.

(next page)

Linearity

A *linear system*, in continuous or discrete time, is a system that possesses the important property of superposition: If an input consists of the weighted sum of several signals, then the output is the superposition—that is, the weighted sum—of the responses of the system to each of those signals.

More precisely, letting $y_1(t)$ be the response of a continuous-time system to an input $x_1(t)$, and $y_2(t)$ the response to $x_2(t)$, the system is linear if:

1. The response to $x_1(t) + x_2(t)$ is $y_1(t) + y_2(t)$.
2. The response to $ax_1(t)$ is $ay_1(t)$, where a is any complex constant.

The first property is known as the *additivity* property, while the second the *scaling* or *homogeneity* property; the same definition holds for continuous time. Note that a system can be linear without being time invariant, and it can be time invariant without being linear.

The two properties defining a linear system can be combined into a single statement:

$$\text{continuous time: } ax_1(t) + bx_2(t) \rightarrow ay_1(t) + by_2(t)$$

$$\text{discrete time: } ax_1[n] + bx_2[n] \rightarrow ay_1[n] + by_2[n]$$

where a and b are any complex constants.

From this see that for a set of inputs $x_k[n]$, $k = 1, 2, 3, \dots$ to a discrete-time linear system with corresponding outputs $y_k[n]$, $k = 1, 2, 3, \dots$, then the response to a linear combination of these inputs given by

$$x[n] = \sum_k a_k x_k[n] = a_1 x_1[n] + a_2 x_2[n] + a_3 x_3[n] + \dots$$

is

$$y[n] = \sum_k a_k y_k[n] = a_1 y_1[n] + a_2 y_2[n] + a_3 y_3[n] + \dots$$

This is known as the *superposition property*, which holds for linear systems in both continuous and discrete time.

See that a direct consequence of the superposition property is that, for linear systems, an input which is zero for all time results in an output which is zero for all time; if $x[n] \rightarrow y[n]$, then by the homogeneity property

$$0 = 0 \cdot x[n] \rightarrow 0 \cdot y[n] = 0$$

(next page)

Examples

Consider the system

$$y(t) = tx(t)$$

To determine whether or not it is linear, we consider two arbitrary inputs $x_1(t)$ and $x_2(t)$.

$$x_1(t) \rightarrow y_1(t) = tx_1(t)$$

$$x_2(t) \rightarrow y_2(t) = tx_2(t)$$

Now consider $x_3(t)$, a linear combination of $x_1(t)$ and $x_2(t)$:

$$x_3(t) = ax_1(t) + bx_2(t)$$

where a and b are arbitrary scalars. See that for input $x_3(t)$, we have the output

$$\begin{aligned} y_3(t) &= tx_3(t) \\ &= t(ax_1(t) + bx_2(t)) \\ &= atx_1(t) + btx_2(t) \\ &= ay_1(t) + by_2(t) \end{aligned}$$

So we conclude that the system is linear (also see that it is not time invariant).

For another example, consider the system

$$y(t) = x^2(t)$$

as defining $x_1(t)$, $x_2(t)$, and $x_3(t)$ as in the previous example, we have

$$x_1(t) \rightarrow y_1(t) = x_1^2(t)$$

$$x_2(t) \rightarrow y_2(t) = x_2^2(t)$$

and

$$\begin{aligned} x_3(t) \rightarrow y_3(t) &= x_3^2(t) \\ &= (ax_1(t) + bx_2(t))^2 \\ &= a^2x_1^2(t) + b^2x_2^2(t) + 2abx_1(t)x_2(t) \\ &= a^2y_1(t) + b^2y_2(t) + 2abx_1(t)x_2(t) \end{aligned}$$

which isn't a superposition of the inputs, thus the system is not linear.
(next page)

More examples

It is important to remember that the scaling constants of the superposition criteria are allowed to be complex. Consider the system

$$y[n] = \text{Re}\{x[n]\}$$

This system is additive, but does not satisfy the homogeneity property, consider input

$$x_1[n] = r[n] + is[n]$$

the corresponding output is

$$y_1[n] = r[n]$$

Now consider scaling $x_1[n]$ by a complex number, for instance $a = i$; defined as x_2 :

$$\begin{aligned} x_2[n] &= ix_1[n] = i(r[n] + is[n]) \\ &= -s[n] + ir[n] \end{aligned}$$

The corresponding output then is

$$y_2[n] = \text{Re}\{x_2[n]\} = -s[n]$$

which is not equal to the scaled version of $y_1[n]$:

$$ay_1[n] = ir[n]$$

thus the system violates the homogeneity property and is not linear.
(next page)

Example—incrementally linear system

Consider the system

$$y[n] = 2x[n] + 3$$

This system is not linear, see that it violates the additivity property:

$$x_1[n] \rightarrow y_1[n] = 2x_1[n] + 3$$

$$x_2[n] \rightarrow y_2[n] = 2x_2[n] + 3$$

where the response to $x_3[n] = x_1[n] + x_2[n]$ is

$$y_3[n] = 2(x_1[n] + x_2[n]) + 3$$

while

$$y_1[n] + y_2[n] = 2(x_1[n] + x_2[n]) + 6$$

See that this system violates the property of linear systems where zero input yields zero output.

It may be surprising that the system here is nonlinear since it describes a linear equation. Intuitively, see that the output of this system can be represented as the sum of the output of a linear system

$$x[n] \rightarrow 2x[n]$$

and another signal equal to the *zero-input response* of the system (when the input is zero),

$$y_0[n] = 3$$



Figure 1.48 Structure of an incrementally linear system. Here, $y_0[n]$ is the zero-input response of the system.

There are large classes of systems that can be represented like this, for which the overall system output consists of the superposition of the response of a linear system with a zero-input response; these systems correspond to the class of *incrementally linear systems*.

The *difference* between the responses to any two inputs to an incrementally system is a linear (additive and homogeneous) function of the *difference* between the two inputs. For example, for inputs $x_1[n]$, $x_2[n]$ and corresponding outputs $y_1[n]$, $y_2[n]$ we have

$$y_1[n] - y_2[n] = 2x_1[n] + 3 - (2x_2[n] + 3) = 2[x_1[n] - x_2[n]]$$

Chapter 2

Linear Time-Invariant Systems

2.1 Discrete-time convolution sum

Representing a discrete signal in terms of impulses

The key idea in visualising how the discrete-time unit impulse can be used to construct any discrete-time signal is to think of a discrete-time signal as a sequence of individual impulses:

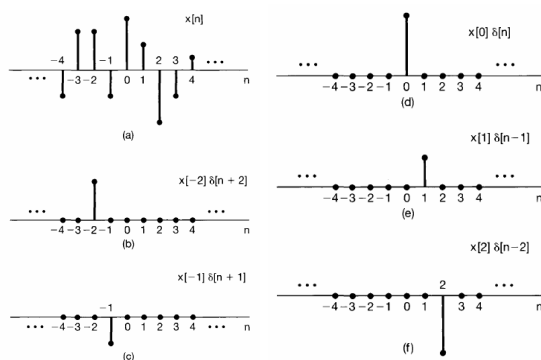


Figure 2.1 Decomposition of a discrete-time signal into a weighted sum of shifted impulses.

See that we can express each value of $x[n]$ as an individual scaled, shifted impulse; for instance

$$\begin{aligned}x[-1]\delta[n+1] &= \begin{cases} x[-1], & n = -1 \\ 0, & n \neq -1 \end{cases} \\x[0]\delta[n] &= \begin{cases} x[0], & n = 0 \\ 0, & n \neq 0 \end{cases} \\x[1]\delta[n-1] &= \begin{cases} x[1], & n = 1 \\ 0, & n \neq 1 \end{cases}\end{aligned}$$

More generally, see that we can write

$$\begin{aligned}x[n] &= \dots + x[-3]\delta[n+3] + x[-2]\delta[n+2] + x[-1]\delta[n+1] + x[0]\delta[n] \\ &\quad + x[1]\delta[n-1] + x[2]\delta[n-2] + x[3]\delta[n-3] + \dots\end{aligned}$$

For any value of n , only one of the terms on the right-hand side of the equation is nonzero, and the scaling associated with that term is precisely $x[n]$. Writing this summation in a more compact form, we have

$$x[n] = \sum_{k=-\infty}^{+\infty} x[k]\delta[n-k]$$

This equation is called the *sifting property* of the discrete-time unit impulse; the summation ‘sifts’ through $x[k]$ and preserves only the value corresponding to $k = n$.

(next page)

Convolution-sum representation of LTI systems

The fact that $x[n]$ can be represented as a superposition of scaled versions of (time shifted) impulses, means that the response of a linear system to $x[n]$ will be a *superposition of the scaled responses of the system* to each of these shifted impulses.

Moreover, the property of time invariance tells us that that the *responses of a time-invariant system to the time-shifted unit impulses are simply time-shifted versions of one another*. The convolution-sum representation for discrete-time systems that are both linear and time invariant results from putting these two basic facts together.

Linear, not necessarily time invariant response

Consider the response of a linear (but possibly time-varying) system to an arbitrary input $x[n]$. We can represent $x[n]$ as a linear combination of shifted impulses.

Letting $h_k[n]$ denote the response of the linear system to the shifted unit impulse $\delta[n - k]$, see that the superposition property of linear systems means that the response $y[n]$ of a linear system to the input $x[n]$ is simply the weighted linear combination of these basic responses:

$$y[n] = \sum_{k=-\infty}^{+\infty} x[k]h_k[n]$$

The response value at a specific time n of a linear system is the superposition of the ‘contributions’ to that output point from each input (from all points in time).

(next page)

Illustrated

For instance, given the input signal $x[n]$ to a linear (non-time invariant) system with the responses $h_{-1}[n]$, $h_0[n]$, and $h_1[n]$ to the signals $\delta[n+1]$, $\delta[n]$, and $\delta[n-1]$:

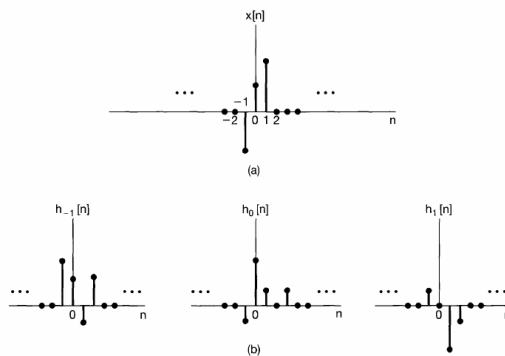


Figure 2.2 Graphical interpretation of the response of a discrete-time linear system as expressed in eq. (2.3).

Since $x[n]$ can be written as a linear combination of the shifted impulses. The system *responds* to these time shifted impulses, h , scaled accordingly by their respective input value, sum to produce $y[n]$.

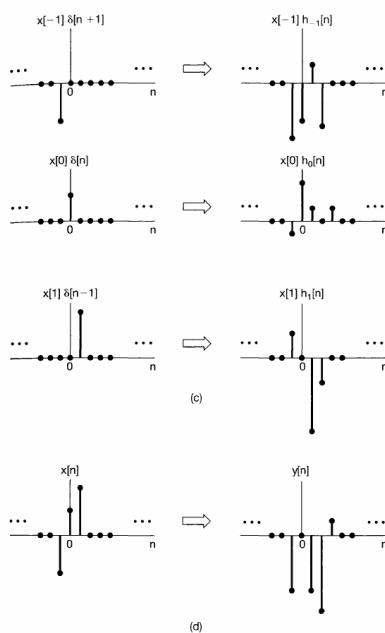


Figure 2.2 Continued

(This illustration only considers three points, but see that the summation on the previous page accounts for the possibility that every input could contribute to a particular output point in time, thus the limits $-\infty$ to ∞)
(next page)

LTI response—convolution sum

In the previous case, the responses $h_k[n]$ need not be related to each other for different values of k . However, if the linear system is also *time invariant*, then these responses to time-shifted unit impulses are all *time-shifted versions of each other*, meaning

$$h_k[n] = h_0[n - k]$$

For notational convenience, we drop the subscript on $h_0[n]$ and define the *unit impulse (sample) response* as

$$h[n] = h_0[n]$$

That is, when $h[n]$ is the output of the LTI system when $\delta[n]$ is the input, we have

$$y[n] = \sum_{k=-\infty}^{+\infty} x[k]h[n - k]$$

This result is referred to as the *convolution sum* or *superposition sum*, and the operation on the right-hand side is known as the *convolution* of the sequences $x[n]$ and $h[n]$, represented symbolically as

$$y[n] = x[n] * h[n]$$

See that an LTI system is completely characterised by its response to a single signal, the unit impulse. As before, the actual output is a superposition of the responses, but due to time invariance this time every response is just a time shift of each other.

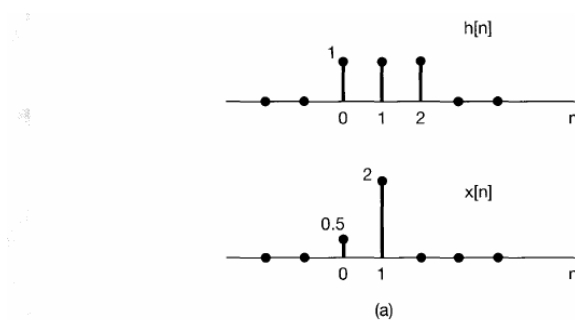
2.2 Discrete-time convolution sum—examples

These examples are crucial for a deeper understanding of the discrete convolution. For reference, the definition of the convolution sum: When $h[n]$ is the output of the LTI system when $\delta[n]$ is the input, we have output $y[n]$

$$y[n] = \sum_{k=-\infty}^{+\infty} x[k]h[n-k]$$

Example 1

Consider an LTI system with impulse response $h[n]$ and input $x[n]$, as illustrated



In this case, since only $x[0]$ and $x[1]$ are nonzero, the output (via convolution), simplifies to the expression

$$y[n] = x[0]h[n-0] + x[1]h[n-1] = 0.5h[n] + 2h[n-1]$$

The sequences $0.5h[n]$ and $2h[n-1]$ are two ‘echoes’ of the impulse response needed for the superposition involved in generating $y[n]$, obtained by summing up each echo for each value of n .

(next page)

Example 1 illustrated

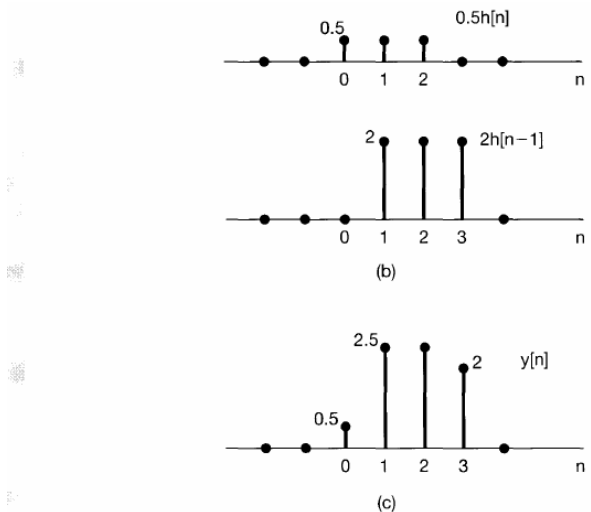


Figure 2.3 (a) The impulse response $h[n]$ of an LTI system and an input $x[n]$ to the system; (b) the responses or "echoes," $0.5h[n]$ and $2h[n-1]$, to the nonzero values of the input, namely, $x[0] = 0.5$ and $x[1] = 2$; (c) the overall response $y[n]$, which is the sum of the echos in (b).

The output for point 0 depends on only on the first point of the impulse response at point 0, $x[0]h[0]$ (in this case there aren't any inputs before this point to produce other impulse responses to affect this). In contrast, the output for point 1 depends on both the *second* point of the impulse response at point 0, $x[0]h[1-0]$ and the *first* point of the impulse response at point 1, $x[1]h[1-1]$. (see how this relates to the structure of the convolution sum)

Alternative view of the convolution

Consider the evaluation of the output value at a specific time n , a particularly convenient way of displaying this calculation graphically begins with the two signals $x[k]$ and $h[n-k]$ viewed as functions of k .

Multiplying these two functions, we obtain a sequence $g[k] = x[k]h[n-k]$, which, at each time k , can be seen as representing the contribution of $x[k]$ to the output at time n . Summing all the samples in the sequence of $g[k]$ yields the output value at the selected time n .

Thus, to calculate $y[n]$ for all values of n requires repeating this procedure for each value of n . Fortunately, changing the value of n has a very simple graphical interpretation for the two signals $x[k]$ and $h[n-k]$, viewed as functions of k . The following examples illustrate this.

(next page)

Example 2—view as a function of k

Let us consider again the convolution problem encountered in the previous example. Now we have the sequences $x[k]$ and $h[n-k]$, for n fixed and viewed as a function of k ; see that the graph of $h[n-k]$ is a time reversal and shift—to the *right* if n is positive and *left* if n is negative:



Having sketched $x[k]$ and $h[n-k]$ for a given n , we multiply the two signal and sum over all values of k (like a dot product) to get $y[n]$. (We multiply and sum up the points where the graphs ‘overlap’)

(Intuit that the view in the first example is the same as ‘reversing’ the impulse response and moving it along the input like a sliding window, which is essentially what is happening here.)

For $n < 0$, neither of the graphs overlap at points where both are nonzero, so $y[n] = 0$ for $n < 0$. For $n = 0$, since the product of the sequence $x[k]$ and the sequence $h[0-k]$ has only one nonzero sample (at $k = 0$), we have

$$y[0] = \sum_{k=-\infty}^{\infty} x[k]h[0-k] = x[0]h[0] = 0.5$$

(next page)

cont. from example 2

The other examples have more ‘overlapping’ points:

$$y[1] = \sum_{k=-\infty}^{\infty} x[k]h[1-k] = 0.5 + 2.0 = 2.5$$

$$y[2] = \sum_{k=-\infty}^{\infty} x[k]h[2-k] = 0.5 + 2.0 = 2.5$$

$$y[3] = \sum_{k=-\infty}^{\infty} x[k]h[3-k] = 2.0$$

(next page)

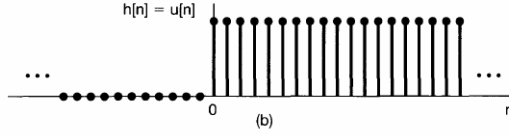
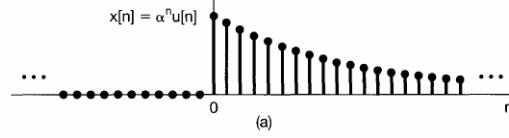
Example 3

Consider input $x[n]$ and unit impulse $h[n]$ given by

$$x[n] = \alpha^n u[n]$$

$$h[n] = u[n]$$

where $0 < \alpha < 1$:



As before, consider $x[k]$ and $h[n-k]$ for different n :



For $n < 0$ there is no overlap between nonzero points in $x[k]$ and $h[n-k]$, so $y[n] = 0$ for $n < 0$. See that for $n \geq 0$,

$$x[k]h[n-k] = \begin{cases} \alpha^k, & 0 \leq k \leq n \\ 0, & \text{otherwise} \end{cases}$$

Thus for $n \geq 0$, the usual infinite limits are superfluous, and we can write

$$y[n] = \sum_{k=0}^n \alpha^k$$

this is a geometric series, we can write

$$y[n] = \sum_{k=0}^n \alpha^k = \frac{1 - \alpha^{n+1}}{1 - \alpha} \quad \text{for } n \geq 0$$

(see appendix for sum of geometric series formula)
(next page)

cont. example 3

We had

$$y[n] = \sum_{k=0}^n \alpha^k = \frac{1 - \alpha^{n+1}}{1 - \alpha} \quad \text{for } n \geq 0$$

which can be rewritten as

$$y[n] = \left(\frac{1 - \alpha^{n+1}}{1 - \alpha} \right) u[n]$$

Illustrated:

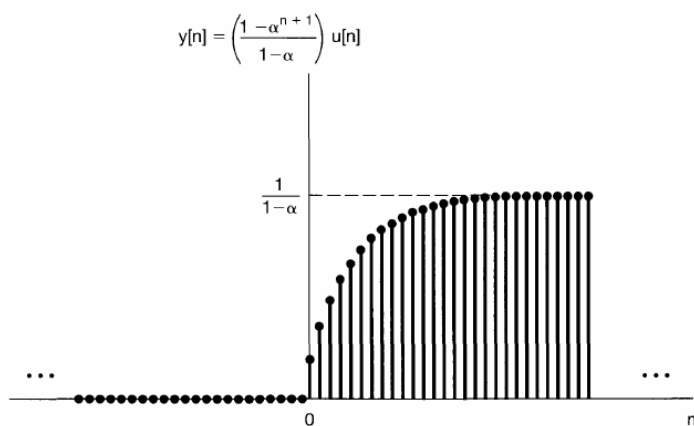


Figure 2.7 Output for Example 2.3.

See that the convolution operation can be described as a ‘sliding’ of the sequence $h[n - k]$ past $x[k]$ (as we increase n the $h[n - k]$ graph just gets shifted to the right).

For example, suppose we have evaluated $y[n]$ for some particular value of n , say $n = n_0$, meaning we have sketched $h[n_0 - k]$, multiplied it by $x[k]$, and summed the results over all k . To evaluate $y[n]$ at the next value of n , so $n = n_0 + 1$, we sketch $h[(n_0 + 1) - k]$; however, we can do this by simply translating $h[n_0 - k]$ to the right by one point.

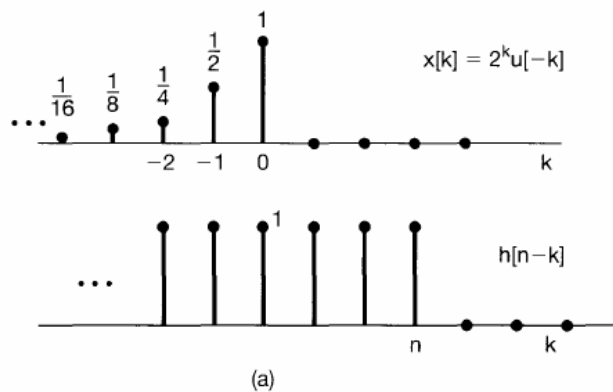
Repeating this for each successive n demonstrates this ‘sliding’ idea.
(next page)

Example 4

Consider an LTI system with input $x[n]$ and unit impulse response $h[n]$ specified as

$$x[n] = 2^n u[-n]$$

$$h[n] = u[n]$$



See that $x[k]$ is zero for $k > 0$ and $h[n-k]$ is zero for $k > n$. Also see that regardless of n , the sequence $x[k]h[n-k]$ always has nonzero samples along the k -axis, and when $n \geq 0$, $x[k]h[n-k]$ has nonzero samples in the interval $k \leq 0$. As such we can write, for $n \geq 0$

$$y[n] = \sum_{k=-\infty}^0 x[k]h[n-k] = \sum_{k=-\infty}^0 2^k$$

and for $n < 0$, $x[k]h[n-k]$ has nonzero samples for $k \leq n$. It follows that, for $n < 0$

$$y[n] = \sum_{k=-\infty}^n x[k]h[n-k] = \sum_{k=-\infty}^n 2^k$$

(next page)

cont. example 4

In both cases we can simplify using the the formula for summing geometric series (see appendix)

$$\sum_{k=0}^{\infty} \alpha^k = \frac{1}{1 - \alpha}$$

Where for the $n \geq 0$ case, by performing a change of variable $r = -k$,

$$y[n] = \sum_{k=-\infty}^0 2^k = \sum_{r=0}^{\infty} \left(\frac{1}{2}\right)^r = \frac{1}{1 - (1/2)} = 2$$

For the $n < 0$ case, by performing a change of variable $l = -k$, and then $m = l + n$,

$$\begin{aligned} y[n] &= \sum_{k=-\infty}^n 2^k = \sum_{l=-n}^{\infty} \left(\frac{1}{2}\right)^l = \sum_{m=0}^{\infty} \left(\frac{1}{2}\right)^{m-n} \\ &= \left(\frac{1}{2}\right)^{-n} \sum_{m=0}^{\infty} \left(\frac{1}{2}\right)^m = 2^n \cdot 2 = 2^{n+1} \end{aligned}$$



2.3 Continuous-time—the convolution integral

Representating a continuous signal in terms of impulses

We develop a continuous-time counterpart of the discrete-time ‘sifting’ property. Consider a riemann-sum like approximation, $\hat{x}(t)$, to a continuous-time signal $x(t)$:



This approximation can be expressed as a linear combination of delayed rectangular pulses. If we define

$$\delta_{\Delta}(t) = \begin{cases} \frac{1}{\Delta}, & 0 \leq t < \Delta \\ 0, & \text{otherwise} \end{cases}$$

then since $\Delta\delta_{\Delta}(t)$ has unit amplitude, we have the expression

$$\hat{x}(t) = \sum_{k=-\infty}^{+\infty} x(k\Delta)\delta_{\Delta}(t - k\Delta)\Delta$$

(δ_{Δ} ‘turns on’ a specific rectangle, x specifies its height, and Δ the base length) See that for any t , only one term in the summation on the right is nonzero.

As we let Δ approach 0, the approximation $\hat{x}(t)$ becomes better, and in the limit equals $x(t)$, therefore we write

$$x(t) = \lim_{\Delta \rightarrow 0} \sum_{k=-\infty}^{+\infty} x(k\Delta)\delta_{\Delta}(t - k\Delta)\Delta$$

Also, as $\Delta \rightarrow 0$, the summation approaches an integral; see that since the limit of $\delta \rightarrow 0$ of $\delta_{\Delta}(t)$ is the unit impulse function $\delta(t)$, we have

$$x(t) = \int_{-\infty}^{+\infty} x(\tau)\delta(t - \tau)d\tau$$

This should be viewed as an idealisation in the sense that, for any Δ ‘small enough’, the approximation by this summation is essentially exact for any practical purpose. The integral then represents an idealisation of the summation by taking Δ to be vanishingly small.

(next page)

Viewed as a function of τ

See that when the signal $\delta(t - \tau)$ is viewed as a function of τ with t fixed, it is still a unit impulse located at $\tau = t$:

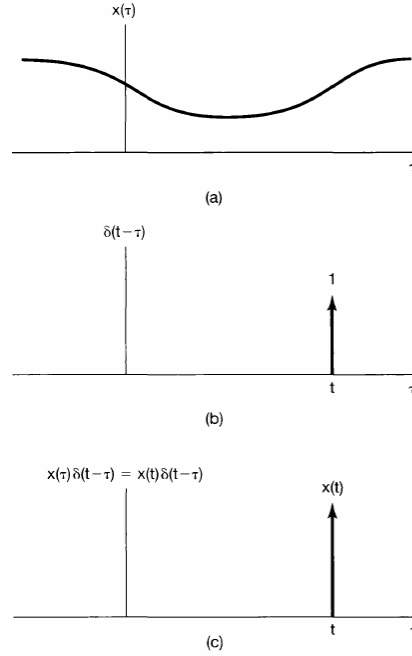


Figure 2.14 (a) Arbitrary signal $x(\tau)$; (b) impulse $\delta(t - \tau)$ as a function of τ with t fixed; (c) product of these two signals.

As such, we have

$$x(\tau)\delta(t - \tau) = x(t)\delta(t - \tau)$$

which is just a scaled impulse at $\tau = t$ with area equal to $x(t)$. See that this can be used to show the statement on the previous page

$$\int_{-\infty}^{+\infty} x(\tau)\delta(t - \tau)d\tau = \int_{-\infty}^{+\infty} x(t)\delta(t - \tau)d\tau = x(t) \int_{-\infty}^{+\infty} \delta(t - \tau)d\tau = x(t)$$

Emphasising the interpretation of the $x(t)$ being represented as a sum (or more precisely, an integral) of weighted, shifted impulses.

(next page)

Continuous-time impulse response

We could represent $\hat{x}(t)$ as a sum of scaled and shifted versions of the basic pulse signal $\delta_\Delta(t)$.

$$\hat{x}(t) = \sum_{k=-\infty}^{+\infty} x(k\Delta)\delta_\Delta(t - k\Delta)\Delta$$

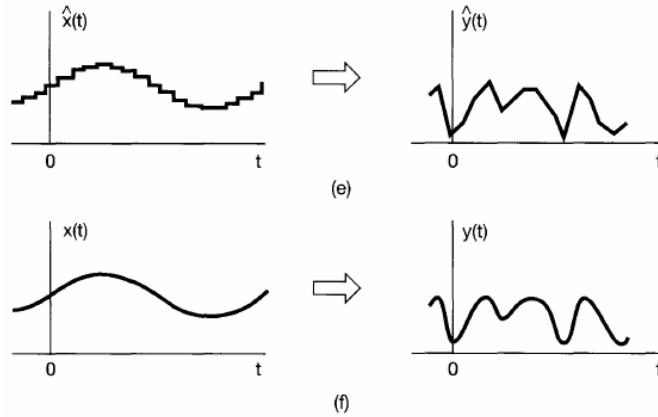
Consequently, the response $\hat{y}(t)$ of a linear system to this signal will be the superposition of the responses to the scaled and shifted versions of $\delta_\Delta(t)$.

Specifically, let us define $\hat{h}_{k\Delta}(t)$ as the response of an LTI system to the input $\delta_\Delta(t - k\Delta)$, then by superposition we have

$$\hat{y}(t) = \sum_{k=-\infty}^{+\infty} x(k\Delta)\hat{h}_{k\Delta}(t)\Delta$$

Now consider what happens as Δ becomes vanishingly small, meaning $\Delta \rightarrow 0$. $\hat{x}(t)$ becomes an increasingly good approximation to $x(t)$, and in fact, the two coincide as $\Delta \rightarrow 0$.

Consequently, the response to $\hat{x}(t)$, namely, $\hat{y}(t)$, must converge to $y(t)$, the response to the actual input $x(t)$. Either way, we say that for Δ ‘small enough’, the duration of the pulse $\delta_\Delta(t - k\Delta)$ is of no significance, in that, as far as the system is concerned, the response to this pulse is essentially the same as the response to a unit impulse as $\Delta \rightarrow 0$:



Since the pulse $\delta_\Delta(t - k\Delta)$ corresponds to a shifted unit impulse as $\Delta \rightarrow 0$, the response $\hat{h}_{k\Delta}(t)$ to this input also becomes the impulse response in the limit. As such we have

$$y(t) = \lim_{\Delta \rightarrow 0} \sum_{k=-\infty}^{+\infty} x(k\Delta)\hat{h}_{k\Delta}(t)\Delta$$

(next page)

cont.

We had

$$y(t) = \lim_{\Delta \rightarrow 0} \sum_{k=-\infty}^{+\infty} x(k\Delta) \hat{h}_{k\Delta}(t) \Delta$$

As $\Delta \rightarrow 0$, the summation on the right-hand side becomes an integral:

$$y(t) = \int_{-\infty}^{+\infty} x(\tau) h_{\tau}(t) d\tau$$

This interpretation is analagous to the one earlier, where any input $x(t)$ can be represented as

$$x(t) = \int_{-\infty}^{+\infty} x(\tau) \delta(t - \tau) d\tau$$

$x(t)$ is essentially represented as a ‘sum’, which, by linearity, leads to $y(t)$ being represented as a superposition of responses to each impulse in the ‘sum’.

Convolution integral

$y(t)$ as decribed above represents the general form of the response of a linear system in continuous time. If, in addition to being linear, the system is also time invariant, then the response to the impulse $\delta(t - \tau)$ is just the response to $\delta(t)$ shifted by τ from the origin, meaning $h_{\tau}(t) = h_0(t - \tau)$.

For notational convenience, we can then drop the subscript and define the unit impulse response $h(t)$ (the response to $\delta(t)$) as

$$h(t) = h_0(t)$$

In this case, the expression for the $y(t)$ above becomes

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

This is referred to as the *convolution integral* or the *superposition integral*, the continuous-time counterpart of the convolution sum, representing a continuous-time LTI system in terms of its response to a unit impulse, written symbolically as

$$y(t) = x(t) * h(t)$$

See that a continuous-time LTI system is completely characterised by its impulse response—a single elementary signal.

Evaluation by visualisation in terms of τ

The procedure for evaluating the convolution integral is similar to the discrete-time counterpart: we first obtain the signal $h(t - \tau)$ regarded as a function of τ with t fixed; this is a reflection about the origin, shifted to the right by t if $t > 0$ or left if $t < 0$. Next multiply $x(\tau)$ and $h(t - \tau)$ and integrate the product over infinity. (see examples)

2.4 Convolution integral—examples

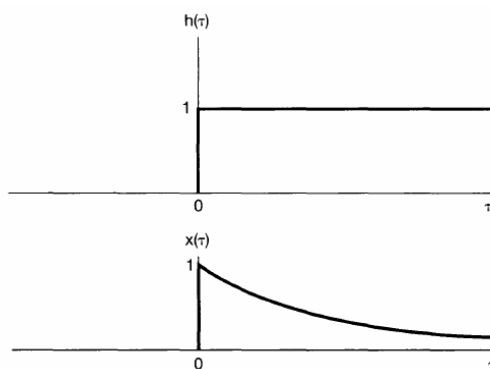
Example 1

Consider letting $x(t)$ be the input to an LTI system with unit impulse response $h(t)$, where

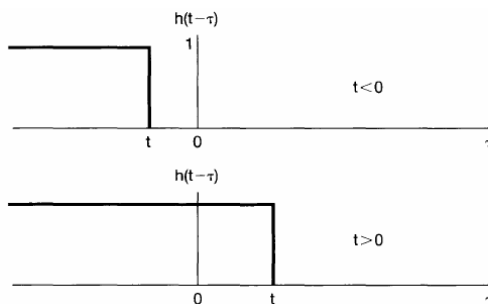
$$x(t) = e^{-at}u(t), \quad a > 0$$

and

$$h(t) = u(t)$$



See that plotting $h(t - \tau)$ with τ as the horizontal axis:



See that for $t < 0$, the product of $x(\tau)$ and $h(t - \tau)$ is zero, and consequently, $y(t)$ is zero. While for $t > 0$,

$$x(\tau)h(t - \tau) = \begin{cases} e^{-a\tau}, & 0 < \tau < t \\ 0, & \text{otherwise} \end{cases}$$

As such, the convolution integral to compute $y(t)$ for $t > 0$ looks like

$$\begin{aligned} y(t) &= \int_0^t e^{-a\tau} d\tau = -\frac{1}{a} e^{-a\tau} \Big|_0^t \\ &= \frac{1}{a} (1 - e^{-at}) \end{aligned}$$

(next page)

cont. example 1

We had $y(t) = 0$ for $t < 0$, and for $t > 0$ we had

$$y(t) = \frac{1}{a}(1 - e^{-at})$$

for all t , we can write

$$y(t) = \frac{1}{a}(1 - e^{-at})u(t)$$

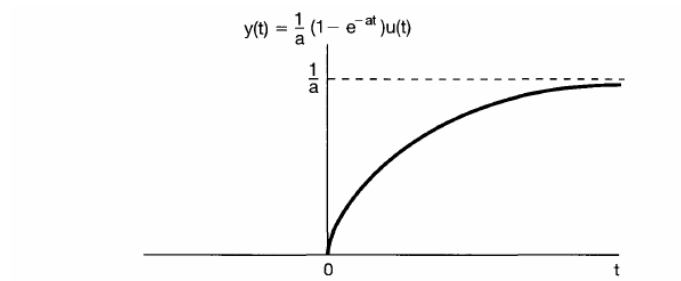


Figure 2.18 Response of the system in Example 2.6 with impulse response $h(t) = u(t)$ to the input $x(t) = e^{-at}u(t)$.

(next page)

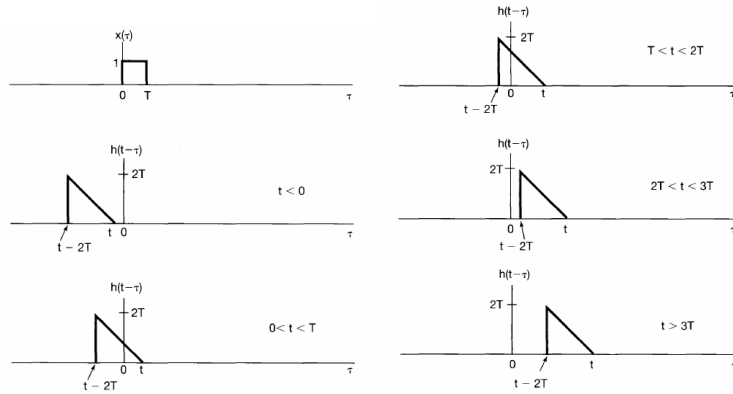
Example 2

Consider the convolution of the following two signals

$$x(t) = \begin{cases} 1, & 0 < t < T \\ 0, & \text{otherwise} \end{cases}$$

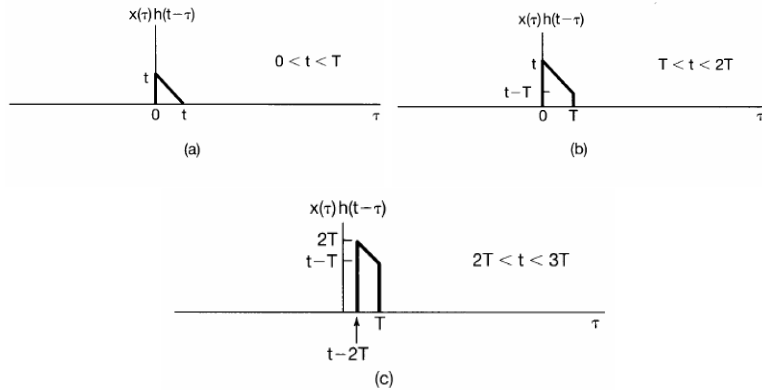
$$h(t) = \begin{cases} t, & 0 < t < 2T \\ 0, & \text{otherwise} \end{cases}$$

Now consider the graphs of $x(\tau)$ and $h(t - \tau)$ plotted for different values of t :



See that the convolution integral is different depending on the value of t :

$$y(t) = \begin{cases} 0, & t < 0 \\ \int_0^t -\tau + t d\tau, & 0 < t < T \\ \int_0^T -\tau + t d\tau, & T < t < 2T \\ \int_{t-2T}^T -\tau + t d\tau, & 2T < t < 3T \\ 0, & 3T < t \end{cases}$$

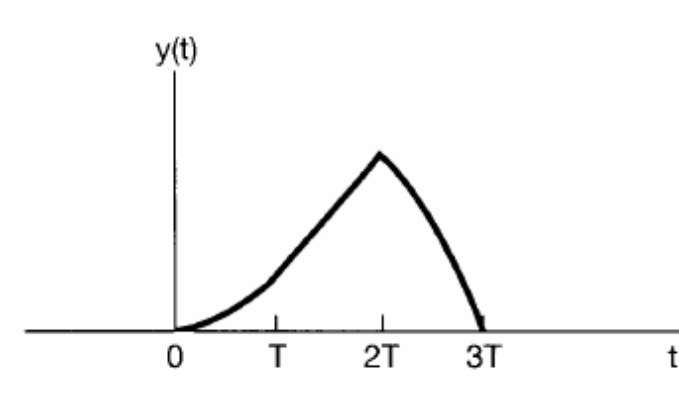


(next page)

cont. example 2

Each convolution integral evaluates to

$$y(t) = \begin{cases} 0, & t < 0 \\ \frac{1}{2}t^2, & 0 < t < T \\ Tt - \frac{1}{2}T^2, & T < t < 2T \\ -\frac{1}{2}t^2 + Tt + \frac{3}{2}T^2, & 2T < t < 3T \\ 0, & 3T < t \end{cases}$$



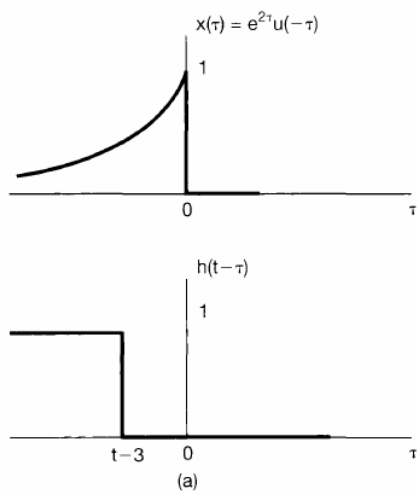
(next page)

Example 3

Consider the convolution between the two signals

$$x(t) = e^{2t}u(-t)$$

$$h(t) = u(t - 3)$$

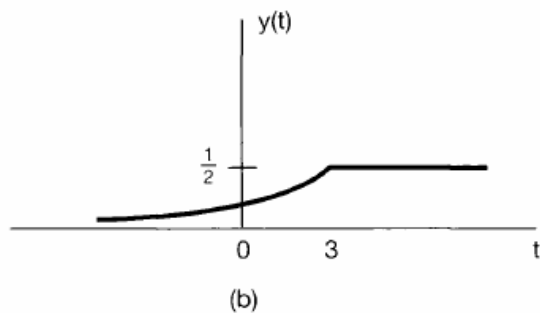


Observe that $x(\tau)$ and $h(t - \tau)$ have regions of nonzero overlap regardless of the value of t . When $t - 3 \leq 0$, the product of $x(\tau)$ and $h(t - \tau)$ is nonzero for $-\infty < \tau < t - 3$; the convolution integral becomes

$$y(t) = \int_{-\infty}^{t-3} e^{2\tau} d\tau = \frac{1}{2}e^{2(t-3)}$$

For $t - 3 \geq 0$, the product $x(\tau)h(t - \tau)$ is only nonzero for $-\infty < \tau < 0$, so the convolution integral is

$$y(t) = \int_{-\infty}^0 e^{2\tau} d\tau = \frac{1}{2}$$



2.5 Properties of LTI systems 1

For reference, we can represent continuous and discrete-time LTI systems in terms of their unit impulse responses; in discrete-time this takes the form of the convolution sum, while its continuous-time counterpart is the convolution integral:

$$y[n] = \sum_{k=-\infty}^{+\infty} x[k]h[n-k] = x[n] * h[n]$$

$$y(t) = \int_{-\infty}^{+\infty} x(\tau)h(t-\tau)d\tau = x(t) * h(t)$$

It is important to emphasise that this property holds in general *only* for LTI systems (since the entire derivation is based on this).

Commutativity

Convolution in both continuous and discrete time is *commutative*. That is, in discrete time

$$x[n] * h[n] = h[n] * x[n] = \sum_{k=-\infty}^{+\infty} h[k]x[n-k]$$

and in continuous time

$$x(t) * h(t) = h(t) * x(t) = \int_{-\infty}^{+\infty} h(\tau)x(t-\tau)d\tau$$

These expressions can be verified by means of a substitution of variables. In the discrete-time case, substituting $r = n - k$:

$$\begin{aligned} x[n] * h[n] &= \sum_{k=-\infty}^{+\infty} x[k]h[n-k] = \sum_{r=-\infty}^{-\infty} x[n-r]h[r] \\ &= \sum_{r=-\infty}^{\infty} x[n-r]h[r] = h[n] * x[n] \end{aligned}$$

and in the continuous case, substituting $r = t - \tau$:

$$\begin{aligned} x(t) * h(t) &= \int_{-\infty}^{+\infty} x(\tau)h(t-\tau)d\tau = \int_{+\infty}^{-\infty} -x(t-r)h(r)dr \\ &= \int_{-\infty}^{+\infty} x(t-r)h(r)dr = h(t) * x(t) \end{aligned}$$

In the context of computing the convolution, this means that the step where a signal is reflected and shifted according to n or t can be done with either x or h in both discrete and continuous time. One form may be easier to visualise, but both forms always result in the same answer.

(next page)

Distributivity

Convolution is also *distributive*. Specifically, convolution distributes over addition, meaning in discrete time

$$x[n] * (h_1[n] + h_2[n]) = x[n] * h_1[n] + x[n] * h_2[n]$$

and in continuous time

$$x(t) * [h_1(t) + h_2(t)] = x(t) * h_1(t) + x(t) * h_2(t)$$

This can be verified by direct substitution into the convolution sum/integral.

See this means that the system

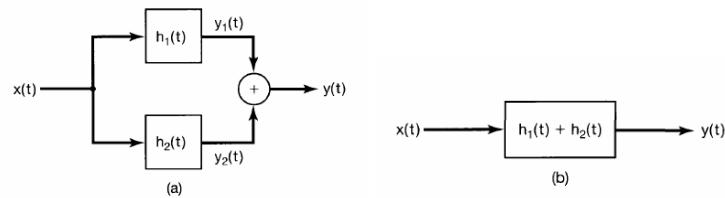
$$y(t) = x(t) * h_1(t) + x(t) * h_2(t)$$

and the system

$$y(t) = x(t) * [h_1(t) + h_2(t)]$$

are identical. There is an analogous interpretation in discrete time.

Illustrated using block diagrams, the two are identical:



See also that as a consequence of both the commutative and distributive properties, we have

$$[x_1[n] + x_2[n]] * h[n] = x_1[n] * h[n] + x_2[n] * h[n]$$

and

$$[x_1(t) + x_2(t)] * h(t) = x_1(t) * h(t) + x_2(t) * h(t)$$

this means that the response of an LTI system to the sum of two inputs must equal the sum of the responses to each input individually.

(next page)

Associativity

Convolution is *associative*; that is, in discrete time

$$x[n] * (h_1[n] * h_2[n]) = (x[n] * h_1[n]) * h_2[n]$$

and in continuous time

$$x(t) * [h_1(t) * h_2(t)] = [x(t) * h_1(t)] * h_2(t)$$

This property can be proven by straightforward manipulation of the summations and integrals. In discrete time (the infinite limit case obscures understanding, so assuming the convolution product is nonzero only for $0 < k < n$ (or t)),

$$\begin{aligned} ((f * g) * h)[n] &= \sum_{k=0}^n (f * g)[k] h[n - k] \\ &= \sum_{k=0}^n \left(\sum_{l=0}^k f[l] g[k - l] \right) h[n - k] \\ &= \sum_{0 \leq l \leq k \leq n} f[l] g[k - l] h[n - k] \\ &= \sum_{l=0}^n \sum_{k=l}^n f[l] g[k - l] h[n - k] \\ &= \sum_{l=0}^n f[l] \left(\sum_{r=0}^{n-l} g[r] h[n - r - l] \right) \\ &= \sum_{l=0}^n f[l] (g * h)[n - l] \\ &= (f * (g * h))[n] \end{aligned}$$

In the third last step we substitute $r = k - l$. The continuous case is analogously:

$$\begin{aligned} ((f * g) * h)(t) &= \int_0^t (f * g)(\tau) h(t - \tau) d\tau \\ &= \int_{\tau=0}^t \left(\int_{s=0}^{\tau} f(s) g(\tau - s) ds \right) h(t - \tau) d\tau \\ &= \iint_{0 \leq s \leq \tau \leq t} f(s) g(\tau - s) h(t - \tau) ds d\tau \\ &= \int_{s=0}^t f(s) \left(\int_{\tau=s}^t g(\tau - s) h(t - \tau) d\tau \right) ds \\ &= \int_{s=0}^t f(s) \left(\int_{r=0}^{t-s} g(r) h(t - r - s) dr \right) ds \\ &= \int_{s=0}^t f(s) (g * h)(t - s) ds = (f * (g * h))(t) \end{aligned}$$

(next page)

Associativity cont.

See that as a consequence of associativity, the expressions

$$y[n] = x[n] * h_1[n] * h_2[n]$$

and

$$y(t) = x(t) * h_1(t) * h_2(t)$$

are unambiguous, since it doesn't matter in what order we convolve the signals. It also follows that the signals

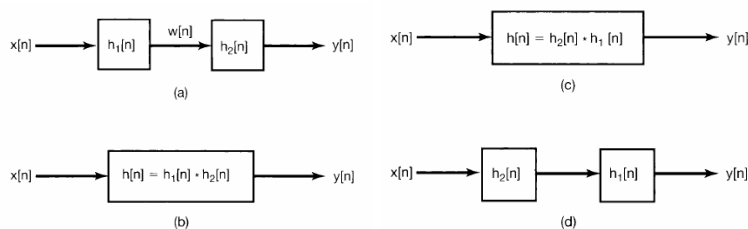
$$\begin{aligned} y[n] &= w[n] * h_2[n] \\ &= (x[n] * h_1[n]) * h_2[n] \end{aligned}$$

and

$$\begin{aligned} y[n] &= x[n] * h[n] \\ &= x[n] * (h_1[n] * h_2[n]) \end{aligned}$$

are equivalent. This can be generalised to an arbitrary number of LTI systems in cascade, and the analogous interpretation and conclusion also hold in continuous time.

Further, consider these block diagrams



Using the associative property, we can show (a)=(b); using the commutative property we know (b)=(c), and finally again the associative property shows (c)=(d).

Consequently, the unit impulse response of a cascade of an arbitrary number of LTI systems does not depend on the order in which they are cascaded. The same conclusions hold in continuous time.

2.6 Properties of LTI systems 2

Memory

Recall that a system is *memoryless* if its output at any time depends only on the value of the input at that same time.

See that the only way this can be true for a discrete-time LTI system is if $h[n] = 0$ for $n \neq 0$. The impulse response would be of the form

$$h[n] = K\delta[n]$$

Where $K = h[0]$ is a constant. (the response to each input only affects the input itself—the response at a single point is not influenced by any other points.)

See that the convolution sum reduces to the relation

$$y[n] = Kx[n]$$

If a discrete-time LTI system has an impulse response $h[n]$ that is not identically zero for $n \neq 0$, then the system has memory.

Similar properties hold for continuous-time LTI systems, which are memoryless if $h(t) = 0$ for $t \neq 0$. Such a memoryless LTI system has the form

$$y(t) = Kx(t)$$

for some constant K ; it has the impulse response

$$h(t) = K\delta(t)$$

Note that if $K = 1$, then these systems become *identity systems*, with output equal to the input and with unit response equal to the unit impulse. In this case, the convolution sum and integral imply that

$$\begin{aligned}x[n] &= x[n] * \delta[n] \\x(t) &= x(t) * \delta(t)\end{aligned}$$

(next page)

Invertibility

A system is *invertible* only if an *inverse system* exists that, when cascaded in series with the original system, produces an output equal to the input to the first system. Consider the following block diagrams

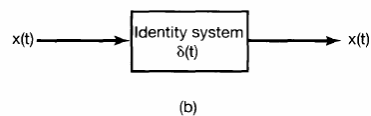
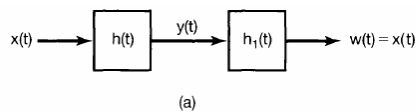


Figure 2.26 Concept of an inverse system for continuous-time LTI systems. The system with impulse response $h_1(t)$ is the inverse of the system with impulse response $h(t)$ if $h(t) * h_1(t) = \delta(t)$.

Given a system with impulse response $h(t)$, the inverse system, with impulse response $h_1(t)$, results in $w(t) = x(t)$ —such that the series interconnection is essentially identical to the identity system.

See (by considering associativity) that the overall impulse response convolved with $x(t)$ is $(h(t) * h_1(t))$. As such we have a condition that $h_1(t)$ must satisfy for it to be the impulse response of the inverse system, namely,

$$h(t) * h_1(t) = \delta(t)$$

Similarly in discrete time, the impulse response $h_1[n]$ of the inverse system for an LTI system with impulse response $h[n]$ must satisfy

$$h[n] * h_1[n] = \delta[n]$$

(next page)

Invertibility example

Consider the LTI system consisting of a pure time shift

$$y(t) = x(t - t_0)$$

Such a system is a *delay* if $t_0 > 0$ and an *advance* if $t_0 < 0$. If $t_0 = 0$, the system is the identity system and thus is memoryless; for any other value of t_0 , this system has memory, (since it responds to an input value at a time other than the current time).

By taking input equal to the $\delta(t)$ we get the impulse response

$$h(t) = \delta(t - t_0)$$

See that as per the definition of the convolution (or by visualising the convolution)

$$y(t) = x(t - t_0) = x(t) * \delta(t - t_0)$$

See that to invert the system, all that is required is to shift the output back. If we consider another system

$$y(t) = x(t + t_0)$$

(see that this is just another time shift in the opposite direction) That would have the impulse response

$$h_1(t) = \delta(t + t_0)$$

then

$$h(t) * h_1(t) = \delta(t - t_0) * \delta(t + t_0) = \delta(t)$$

Similarly in discrete time, a pure time shift with impulse response $\delta[n - n_0]$ would have an inverse consisting of a time shift in the opposite direction by the same amount—another LTI system with the impulse response $\delta[n - n_0]$.

(next page)

Causality

Recall the *causality* property: the output of a causal system depends only on the present and past values of the input to the system.

In the context of convolution, see that for a discrete-time LTI system to be causal, $y[n]$ must not depend on $x[k]$ for $k > n$, and that in order for this to be true we must have

$$h[n] = 0 \quad \text{for } n < 0$$

(try to visualise the convolution to see this) This means that the impulse response of a causal LTI system must be zero before the impulse occurs, which is consistent with the intuitive concept of causality.

More generally, causality for a linear system is equivalent to the condition of *initial rest*: if the input to a causal system is 0 up to some point in time, then the output must also be 0 up to that time. It is important to emphasise that the equivalence of causality and the condition of initial rest applies only to linear systems.

(For example, the system $y[n] = 2x[n] + 3$ is not linear but is causal. However if $x[n] = 0$, $y[n] = 3 \neq 0$, so it does not satisfy the condition of initial rest.)

Once again visualising the convolution sum, see that for a causal discrete-time LTI system, the $h[n] = 0$ for $n < 0$ implies that it can be simplified to

$$y[n] = \sum_{k=-\infty}^n x[k]h[n-k]$$

and the alternative equivalent form (by commutativity) becomes (visualise as reflecting the input signal in time instead)

$$y[n] = \sum_{k=0}^{\infty} h[k]x[n-k]$$

Similarly, a continuous-time LTI system is causal if

$$h(t) = 0 \quad \text{for } t < 0$$

where the convolution integral is given by

$$y(t) = \int_{-\infty}^t x(\tau)h(t-\tau)d\tau = \int_0^{\infty} h(\tau)x(t-\tau)d\tau$$

While causality is a property of systems, it is common terminology to refer to a signal as being causal if it is zero for $n < 0$ or $t < 0$. The motivation for this comes from the conditions for causality as defined above: Causality for an LTI system is equivalent to its impulse response being a ‘causal’ signal.
(next page)

Stability

Recall that a system is *stable* if every bounded input produces a bounded output. We want to derive a condition to determine whether a candidate LTI system is stable; consider an input $x[n]$ that is bounded in magnitude:

$$|x[n]| < B \quad \text{for all } n$$

Suppose we apply this input to an LTI system with unit impulse response $h[n]$. Then, using the convolution sum, we obtain an expression for the *magnitude* of the output as

$$|y[n]| = \left| \sum_{k=-\infty}^{+\infty} h[k]x[n-k] \right|$$

Since the magnitude of the sum of a set of numbers is no larger than the sum of the magnitudes of the numbers, it follows that

$$|y[n]| \leq \sum_{k=-\infty}^{+\infty} |h[k]| |x[n-k]|$$

We know that $|x[n-k]| < B$ for all values of k and n . This implies that

$$|y[n]| \leq B \sum_{k=-\infty}^{+\infty} |h[k]| \quad \text{for all } n$$

We conclude that the impulse response is *absolutely summable* if

$$\sum_{k=-\infty}^{+\infty} |h[k]| < \infty$$

See that if the impulse response is absolutely summable, $y[n]$ will be bounded in magnitude, and hence the system is stable. This means that the above is a sufficient condition to guarantee the stability of a discrete-time LTI system. (see that this condition is also necessary, as if it were not satisfied then there would be unbounded outputs)
(next page)

Stability cont.

In continuous time we have an analogous characterisation of stability in terms of the impulse response of an LTI system; if $|x(t)| < B$ for all t , then it follows that

$$\begin{aligned} |y(t)| &= \left| \int_{-\infty}^{+\infty} h(\tau)x(t-\tau)d\tau \right| \\ &\leq \int_{-\infty}^{+\infty} |h(\tau)||x(t-\tau)|d\tau \\ &\leq B \int_{-\infty}^{+\infty} |h(\tau)|d\tau \end{aligned}$$

Therefore, the system is stable if it is *absolutely integrable*:

$$\int_{-\infty}^{+\infty} |h(\tau)|d\tau < \infty$$

As with the discrete case this is actually a necessary condition for stability, as if it were unsatisfied, it would lead to unbounded outputs.

Stability example 1—stable

Consider a system that is a pure time shift in either continuous or discrete time. Then in discrete time

$$\sum_{n=-\infty}^{+\infty} |h[n]| = \sum_{n=-\infty}^{+\infty} |\delta[n - n_0]| = 1$$

while in continuous time

$$\int_{-\infty}^{+\infty} |h(\tau)|d\tau = \int_{-\infty}^{+\infty} |\delta(\tau - t_0)|d\tau = 1$$

and we can conclude both systems are stable.

(next page)

Stability example 2—unstable

Consider now the accumulator, with an impulse response equal to $u[n]$; see that the impulse response in this case is not absolutely summable:

$$\sum_{n=-\infty}^{\infty} |u[n]| = \sum_{n=0}^{\infty} u[n] = \infty$$

Similarly consider the integrator (the continuous-time counterpart of the accumulator):

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

This is also an unstable system; see that the impulse response for the integrator is:

$$h(t) = \int_{-\infty}^t \delta(\tau) d\tau = u(t)$$

and

$$\int_{-\infty}^{+\infty} |u(t)| d\tau = \int_0^{+\infty} |u(t)| d\tau = \infty$$

Since the impulse response is not absolutely integrable, the system is not stable.

2.7 The Unit Step Response of an LTI system

There is another signal that is also used quite often in describing the behaviour of LTI systems: the *unit step response* $s[n]$ or $s(t)$; as implied, it corresponds to the output when $x[n] = u[n]$ or $x(t) = u(t)$.

From the convolution-sum representation, see that the step response can be found by convolution of the unit step with the impulse response:

$$s[n] = u[n] * h[n]$$

Also see that by commutativity, $s[n] = h[n] * u[n]$, and therefore $s[n]$ can be viewed as the response to the input $h[n]$ of a discrete-time LTI system with unit response $u[n]$. Since $u[n]$ is the unit impulse response of the *accumulator*, this means that an input $h[n]$ to an accumulator yields the unit step response:

$$s[n] = \sum_{k=-\infty}^n h[k]$$

The fact that we know the inverse system for the accumulator means we know then that $h[n]$ can be recovered from $s[n]$ using the relation

$$h[n] = s[n] - s[n-1]$$

That is, the step response of a discrete-time LTI system is the *running sum of its impulse response* (try to visualise this!). Conversely, the impulse response of a discrete-time LTI system is the first difference of its step response.

Similarly in continuous time, the step response of an LTI system with impulse response $h(t)$ is given by $s(t) = u(t) * h(t) = h(t) * u(t)$. The *integrator* has $u(t)$ as its impulse response, so this is equivalent to inputting $h(t)$ into an integrator:

$$s(t) = \int_{-\infty}^t h(\tau) d\tau$$

as such see that the unit impulse response can be recovered as the first derivative

$$h(t) = \frac{ds(t)}{dt} = s'(t)$$

The unit step response of a continuous-time LTI system is the running integral of its impulse response, and the unit impulse response is the first derivative of the unit step response.

In that sense, the unit step response can also be used to characterise an LTI system, since we can calculate the unit impulse response from it.

2.8 Causal LTI systems described by differential and difference equations

Linear constant-coefficient differential equations

An important point about differential equations is that they provide an *implicit* specification of a system. That is, they describe a relationship between the input and the output, rather than an *explicit* expression for the system output as a function of the input. (meaning they don't immediately tell you the output from a given input value, unlike say a function)

Auxiliary conditions—Initial rest

Recall that the formation of an explicit solution to a differential equation requires the specification of certain *auxiliary/initial conditions* in order to fully evaluate the *homogeneous solution* or the *natural response* of the system. Also recall that different choices of auxiliary conditions will lead to different relationships between the input and output.

In some cases this auxiliary condition might take the form of the condition of *initial rest*, where if $x(t) = 0$ for $t < t_0$, then $y(t)$ must also equal 0 for $t < t_0$. It is important to emphasise that a condition of initial rest does not specify a zero initial condition at a fixed point in time, but rather adjusts this point in time so that the *response is zero until the input becomes nonzero*. (thus, if $x(t) = 0$ for $t \leq t_0$, then we could use $y(t_0) = 0$ to solve for the output for $t > t_0$)
(next page)

General form

A general N th-order linear constant-coefficient differential equation is given by

$$\sum_{k=0}^N a_k \frac{d^k y(t)}{dt^k} = \sum_{k=0}^M b_k \frac{d^k x(t)}{dt^k}$$

The *order* refers to the highest derivative of the output $y(t)$ appearing in the equation. In the case where $N = 0$, see that this reduces to

$$y(t) = \frac{1}{a_0} \sum_{k=0}^M b_k \frac{d^k x(t)}{dt^k}$$

In this case, $y(t)$ is an explicit function of the input $x(t)$ and its derivatives. For $N \geq 1$, the equation would specify an implicit relationship in terms of the input. In this case, the analysis of the equation would proceed in a way analogous to that of a first-order DE, with a solution consisting of two parts—a particular solution plus a solution to the homogeneous differential equation

$$\sum_{k=0}^N a_k \frac{d^k y(t)}{dt^k} = 0$$

As in the first-order case, the general form does not completely specify the output in terms of the input, and we need to identify auxiliary conditions to determine completely the input-output relationship for the system. Should we consider initial rest, we assume then that if $x(t) = 0$ for $t \leq t_0$, then $y(t) = 0$ for $t \leq t_0$, and therefore, the response for $t > t_0$ can be calculated with the initial conditions

$$y(t_0) = \frac{dy(t_0)}{dt} = \dots = \frac{d^{N-1}y(t_0)}{dt^{N-1}} = 0$$

(For an N th order DE we'll need N initial conditions. This comes from the fact that the characteristic polynomial will have N roots, so N exponentials where any linear combination of them would satisfy the homogeneous equation, thus warranting the need for the initial conditions to determine the coefficients of this linear combination.)

Under the condition of initial rest, the system described in the general form above is causal and LTI.

(next page)

Linear constant-coefficient difference equations—General form

The discrete-time counterpart of the previous general form is the N th-order linear constant-coefficient difference equation

$$\sum_{k=0}^N a_k y[n-k] = \sum_{k=0}^M b_k x[n-k]$$

All equations of this type can be solved in a manner analagous to that for differential equations. As before, the solution $y[n]$ can be written as the sum of a particular solution to the above plus a solution to the homogeneous equation

$$\sum_{k=0}^N a_k y[n-k] = 0$$

As in the continuous-time case, the general form is not sufficient to completely specify an output, and requires specification of auxiliary conditions (in this case this might be $y[n-1], \dots, y[n-N]$). The condition of initial rest can also be applied here, where if $x[n] = 0$ for $n < n_0$, then $y[n] = 0$ for $n < n_0$ as well. With initial rest, the system would be LTI and causal.

Recursive equations

Although all of these properties can be developed following an approach that directly parallels those for differential equations, the discrete-time case offers an alternative path. See that the general form can be rearranged as

$$y[n] = \frac{1}{a_0} \left\{ \sum_{k=0}^M b_k x[n-k] - \sum_{k=1}^N a_k y[n-k] \right\}$$

See that the output at time n can be expressed in terms of previous values of the input and output. An equation of this form is called a *recursive equation*, since it specifies a recursive procedure for determining the output in terms of the input and previous outputs. In the special case where $N = 0$, this reduces to

$$y[n] = \sum_{k=0}^M \frac{b_k}{a_0} x[n-k]$$

Here $y[n]$ is an explicit function of the present and previous values of the input. For that reason this is often called a *nonrecursive equation* (since we do not recursively use previously computed values to compute the present output).

Although we do not require auxiliary conditions for the case of $N = 0$, such conditions are needed for the recursive case when $N \geq 1$.

(next page)

Difference equation example

Consider the difference equation

$$y[n] - \frac{1}{2}y[n-1] = x[n]$$

rewritten as

$$y[n] = x[n] + \frac{1}{2}y[n-1]$$

this highlights the idea that we need the previous value of the output, $y[n-1]$, to calculate the current value. Thus to begin the recursion we need an initial condition.

Suppose we impose the condition of initial rest and consider the input

$$x[n] = K\delta[n]$$

In this case, since $x[n] = 0$ for $n \leq -1$ (because of the delta function), the condition of initial rest implies that $y[n] = 0$ for $n \leq -1$, so we have as an initial condition $y[-1] = 0$. Starting from the initial condition, see that we can recursively solve for successive values of $y[n]$ for $n \geq 0$ as follows:

$$\begin{aligned}y[0] &= x[0] + \frac{1}{2}y[-1] = K \\y[1] &= x[1] + \frac{1}{2}y[0] = \frac{1}{2}K \\y[2] &= x[2] + \frac{1}{2}y[1] = \left(\frac{1}{2}\right)^2 K \\&\vdots \\y[n] &= x[n] + \frac{1}{2}y[n-1] = \left(\frac{1}{2}\right)^n K\end{aligned}$$

See additionally that by setting $K = 1$ we obtain the impulse response for the system as

$$h[n] = \left(\frac{1}{2}\right)^n u[n]$$

(see that some initial conditions may lead to a non-LTI system, in which obtaining the impulse response be much less useful)

2.9 Block diagram representations of first-order systems described by differential and difference equations

Discrete-time

Consider the causal system described by the first-order difference equation

$$y[n] + ay[n - 1] = bx[n]$$

Note that the evaluation of $y[n]$ requires three basic operations: addition, multiplication by a coefficient, and a delay (to capture the relationship between $y[n]$ and $y[n - 1]$). We define these three basic network elements as follows:

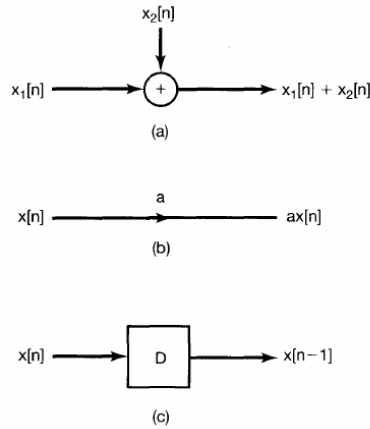


Figure 2.27 Basic elements for the block diagram representation of the causal system described by eq. (2.126): (a) an adder; (b) multiplication by a coefficient; (c) a unit delay.

Rewriting the equation in a form that directly suggests a recursive algorithm for computing successive values of $y[n]$:

$$y[n] = -ay[n - 1] + bx[n]$$

see that we can represent this algorithm using the basic elements as

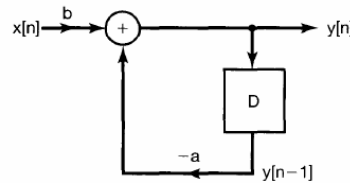


Figure 2.28 Block diagram representation for the causal discrete-time system described by eq. (2.126).

This is an example of a *feedback* system (since the output is fed back after a delay and multiplication as an addition to the next input). See that the presence of feedback is a direct consequence of the recursive nature of the system. The block diagram makes clear the required memory in this system and the consequent need for initial conditions (the initial memory element).

(next page)

Continuous-time

Consider next the causal continuous-time system described by a first-order differential equation:

$$\frac{dy(t)}{dt} + ay(t) = bx(t)$$

which we can rewrite as

$$y(t) = -\frac{1}{a} \frac{dy(t)}{dt} + \frac{b}{a} x(t)$$

The three operations involved here are addition, multiplication by a coefficient, and differentiation. Therefore, if we define the three network elements as follows

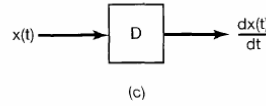
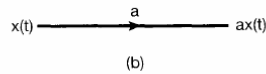
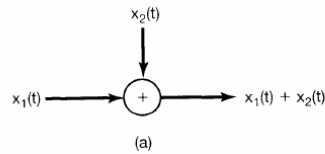


Figure 2.29 One possible set of basic elements for the block diagram representation of the continuous-time system described by eq. (2.128): (a) an adder; (b) multiplication by a coefficient; (c) a differentiator.

we can consider representing the system as follows

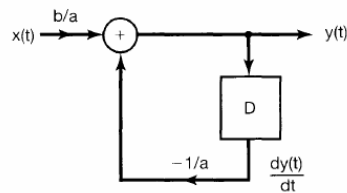


Figure 2.30 Block diagram representation for the system in eqs. (2.128) and (2.129), using adders, multiplications by coefficients, and differentiators.

While such a configuration is a valid representation of the causal system as described, it is not the representation that is most frequently or practically used, since differentiators are both difficult to implement and extremely sensitive to errors and noise.

(next page)

Continuous-time cont.

The previously proposed representation of the system

$$y(t) = -\frac{1}{a} \frac{dy(t)}{dt} + \frac{b}{a} x(t)$$

is valid but impractical, since differentiators are difficult to implement. An alternative implementation that is much more widely used can be obtained by first rewriting the system as

$$\frac{dy(t)}{dt} = bx(t) - ay(t)$$

and then *integrating* from $-\infty$ to t . Specifically, if we assume that the value of $y(-\infty)$ is zero, then the integral of $dy(t)/dt$ from $-\infty$ to t is precisely $y(t)$. Consequently, we obtain the equation

$$y(t) = \int_{-\infty}^t [bx(\tau) - ay(\tau)] d\tau$$

In this form, see that this system can be implemented using the adder and coefficient multiplier as indicated earlier, together with an *integrator*:

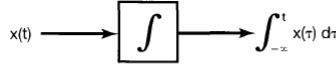


Figure 2.31 Pictorial representation of an integrator.

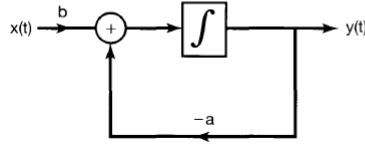


Figure 2.32 Block diagram representation for the system in eqs. (2.128) and (2.131), using adders, multiplications by coefficients, and integrators.

Integrators are much more readily implemented. In the continuous-time case it is the integrator that represents the memory storage element of the system. This is perhaps more readily seen if we consider integrating the rearranged equation from a finite point in time t_0 :

$$y(t) = y(t_0) + \int_{t_0}^t [bx(\tau) - ay(\tau)] d\tau$$

See that an initial condition $y(t_0)$ must be specified; this is the value that the integrator stores at time t_0 .

2.10 The Unit impulse as an idealised short pulse

We know that the unit impulse $\delta(t)$ is the impulse response of the identity system. That is,

$$x(t) = x(t) * \delta(t)$$

for any signal $x(t)$. Therefore if we take $x(t) = \delta(t)$, we have

$$\delta(t) = \delta(t) * \delta(t)$$

This is a basic property of the unit impulse, and it also has a significant implication for our interpretation of the unit impulse as an idealised pulse. For instance, recall that we initially considered $\delta(t)$ to be the limiting form of a rectangular pulse $\delta_\Delta(t)$; now consider

$$r_\Delta(t) = \delta_\Delta(t) * \delta_\Delta(t)$$

Then $r_\Delta(t)$ looks like

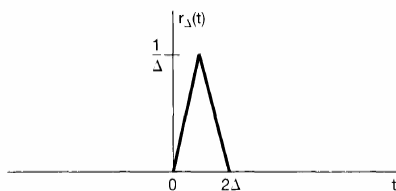


Figure 2.33 The signal $r_\Delta(t)$ defined in eq. (2.135).

See that if we wish to interpret $\delta(t)$ as the limit as $\Delta \rightarrow 0$ of $\delta_\Delta(t)$, then by the second statement, the limit as $\Delta \rightarrow 0$ of $r_\Delta(t)$ *must also be the unit impulse*.

Since we can argue that the limits as $\Delta \rightarrow 0$ for $r_\Delta(t)$ must also be a unit impulse, a similar argument can be made that as $\Delta \rightarrow 0$, $r_\Delta(t) * r_\Delta(t)$ or $r_\Delta(t) * \delta_\Delta(t)$ must also be unit impulses and so on.

Thus see that if we define the unit impulse as being the limiting form of some signal, there are an unlimited number of very dissimilar looking signals that all behave like an impulse in the limit.

(next page)

Cont.

We found that if we define the unit impulse as being the limiting form of some signal, there are an unlimited number of very dissimilar looking signals that all behave like an impulse in the limit.

The idea here is that they all ‘behave like an impulse’, meaning that the response an LTI system to all of these signals is essentially identical as long as the pulse is ‘short enough’, meaning Δ is ‘small enough’. To illustrate this, consider an LTI system described by the first order differential equation

$$\frac{dy(t)}{dt} + 2y(t) = x(t)$$

assuming initial rest. The responses of this system to $\delta_\Delta(t)$, $r_\Delta(t) * \delta_\Delta(t)$, and $r_\Delta(t) * r_\Delta(t)$ for several values are illustrated here for several values of Δ :

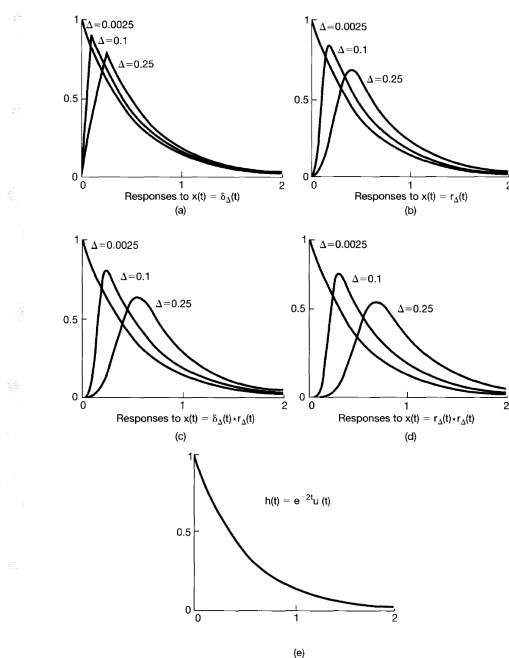


Figure 2.34 Interpretation of a unit impulse as the idealization of a pulse whose duration is “short enough” so that, as far as the response of an LTI system to this pulse is concerned, the pulse can be thought of as having been applied instantaneously: (a) responses of the causal LTI system described by eq. (2.136) to the input $\delta_\Delta(t)$ for $\Delta = 0.25, 0.1$, and 0.0025 ; (b) responses of the same system to $r_\Delta(t)$ for the same values of Δ ; (c) responses to $\delta_\Delta(t) * r_\Delta(t)$; (d) responses to $r_\Delta(t) * r_\Delta(t)$; (e) the impulse response $h(t) = e^{-2t}u(t)$ for the system. Note that, for $\Delta = 0.25$, there are noticeable differences among the responses to these different signals; however, as Δ becomes smaller, the differences diminish, and all of the responses converge to the impulse response shown in (e).

See that for Δ large enough, the responses differ noticeably. However, for Δ sufficiently small, they all have a limiting form of $e^{-2t}u(t)$. The limit for each of these signals as $\Delta \rightarrow 0$ is the unit impulse, so we conclude that $e^{-2t}u(t)$ is the impulse response for this system.

(next page)

Cont.

One important point to be emphasised here is what we mean by ‘ Δ small enough’ depends on the particular LTI system to which the preceding pulses are applied. For instance consider instead the causal LTI system

$$\frac{dy(t)}{dt} + 20y(t) = x(t)$$

plotted as per the previous instance for different values of Δ :

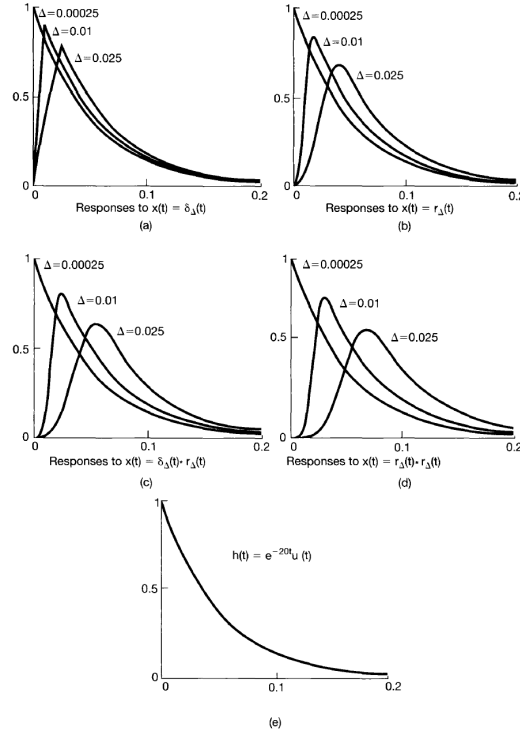


Figure 2.35 Finding a value of Δ that is “small enough” depends upon the system to which we are applying inputs: (a) responses of the causal LTI system described by eq. (2.137) to the input $\delta_\Delta(t)$ for $\Delta = 0.025, 0.01$, and 0.00025 ; (b) responses to $r_\Delta(t)$; (c) responses to $\delta_\Delta(t) * r_\Delta(t)$; (d) responses to $r_\Delta(t) * r_\Delta(t)$; (e) the impulse response $h(t) = e^{-20t}u(t)$ for the system. Comparing these responses to those in Figure 2.34, we see that we need to use a smaller value of Δ in this case before the duration and shape of the pulse are of no consequence.

In this case see that we need a smaller value of Δ for the responses to be indistinguishable from each other and from the impulse response $h(t) = e^{-20t}u(t)$ for the system.

As such see that what we mean by ‘ Δ small enough’ differs depending on the system. In that sense the unit impulse is the idealisation for a short pulse whose duration is short enough for *all systems*.

2.11 Defining the unit impulse through convolution

As outlined previously, for Δ small enough, the signals $\delta_\Delta(t)$, $r_\Delta(t)$, $r_\Delta(t) * \delta_\Delta(t)$ and $r_\Delta(t) * r_\Delta(t)$, all act like impulses when applied to an LTI system (naturally there are an infinite number of signals for which this is true).

This suggests that we should think of the unit impulse as how an LTI system responds to it. While usually a function or signal is defined by what it is at each time point, the primary importance of the unit impulse is not that, but rather what it *does under convolution*. As such we may alternatively define the unit impulse as the signal which, when applied to an LTI system, yields the impulse response. That is, we define $\delta(t)$ as the signal for which

$$x(t) = x(t) * \delta(t)$$

for any $x(t)$. The signals that satisfy this property in the are all said to behave like a unit impulse.

All the previously mentioned properties of the unit impulse we need can be obtained from the *operational definition* given above. For instance, should we let $x(t) = 1$ for all t , then

$$\begin{aligned} 1 = x(t) &= x(t) * \delta(t) = \delta(t) * x(t) = \int_{-\infty}^{+\infty} \delta(\tau) x(t - \tau) d\tau \\ &= \int_{-\infty}^{+\infty} \delta(\tau) d\tau \end{aligned}$$

showing that the unit impulse has unit area.

(next page)

Cont.
It is

Appendix A

Other proofs

A.1 Sum of geometric series

Here we show

$$S_n = \sum_{i=0}^{n-1} \alpha^i = \frac{\alpha^n - 1}{\alpha - 1}$$

This can be seen from

$$\begin{aligned} (\alpha - 1)S_n &= \alpha \sum_{i=0}^{n-1} \alpha^i - \sum_{i=0}^{n-1} \alpha^i \\ &= \sum_{i=1}^n \alpha^i - \sum_{i=0}^{n-1} \alpha^i \\ &= \alpha^n + \sum_{i=1}^{n-1} \alpha^i - (1 + \sum_{i=1}^{n-1} \alpha^i) \\ &= \alpha^n - 1 \end{aligned}$$

and so

$$S_n = \frac{\alpha^n - 1}{\alpha - 1}$$