Spring, 2001 Data Processing and Analysis (GEOP 505)

Rick Aster

January 16, 2001

Introduction to Linear Systems, Part 1: The Time Domain

Our primary goal in this course is to understand methods of analyzing temporal and spatial series, especially as applied to *linear systems*, both in continuous and sampled (discrete) time, and to demonstrate applications to important problems in geophysics and other physical sciences. Much of the demonstration and homework in this course will be done using *Matlab*. You are thus encouraged to demo and/or refamiliarize yourself with this package at the earliest opportunity. *Matlab* is also available in a student version that you may wish to procure for your personal PC or other computer. There is also a Matlab primer on the class web page .

We will be primarily concerned with an important class of physical situations which can be adequately characterized by *linear systems*. A linear system is any functional transformation, ϕ, which converts some *input* signal, $x(t)$ to an *output* signal, $y(t)$

$$y(t) = \phi[x(t)] \tag{1}$$

and which follows the principles of superposition

$$\phi[x(t) + y(t)] = \phi[x(t)] + \phi[y(t)] \tag{2}$$

and amplitude scaling

$$\phi[\alpha x(t)] = \alpha\phi[x(t)] \tag{3}$$

where α is a scalar. Note that for positive integer values of α (3) is equivalent to (2). (3) also implies that the output of the system is zero when there is no input

$$\phi[0] = 0 \ . \tag{4}$$

Many of the phenomena which we wish to study in geophysics are linear, often because we study very weak signals (e.g., small gravity variations, seismic disturbances far away from the source; effects due to small fluctuations in the magnetic field) and the linear approximation is valid because the system is not perturbed very far from equilibrium. Examples where linearity does not hold up are generally instances of large amplitude (e.g., high strain elastic waves near an underground nuclear explosion or earthquake; ocean waves breaking at a shoreline). In these cases the physics of the problem depends strongly on the amplitude of the perturbation, so that superposition (2) and scaling (3) do not hold, and probably aren't even acceptable approximations.

Many physical systems are also *time-invariant*, i.e., ϕ is not a function of t (in some cases we look for time variation, for example, earthquake prediction researchers hope that this is not the case for some aspect of the earth response over short time intervals). In general, we will be primarily concerned with time-invariant systems.

A linear system is said to be *causal* if the output at some time t_0 depends only on values of the input for $t \leq t_0$. Note that all physical processes are causal (acausal systems propagate information backwards in time!) It is easy, however, to mathematically construct non-causal mathematical systems, and they may be useful in processing stored information. Also keep in mind that physical spatial phenomena (e.g. spatial filters) need not obey "causality" constraints.

A linear system is said to be *stable* if every bounded (in amplitude) input produces a bounded output. While obvious for physical systems (which will eventually become non-linear rather than produce infinite responses) this can an important consideration in mathematical models of active systems (i.e., systems that have feedback).

The simple rules defining linear systems provide profound and very useful constraints on the mathematical characterization of the system. Most remarkably, they lead to an elegant, tractable, and very useful set of analysis tools, embodied in *Fourier Theory* for describing linear systems in the complementary domains of time and frequency.

It may at first appear remarkable that the input to output transformation of *any* linear, time-invariant system can be characterized by a basic integral relation (a *convolution*). To derive this result straightforwardly, we must first define the *Dirac delta* or *impulse function*. This function is discontinuous; it is nonzero only at one value of its argument, where it is infinite. The trick to making the delta function rigorously mathematically useful is to define it as a limiting set of functions so that the area of the function remains finite. One definition (e.g., Bracewell) is:

$$\delta(t) = \lim_{\tau \to 0} \; \tau^{-1} \Pi(t/\tau) \tag{5}$$

where $\tau^{-1}\Pi(t/\tau)$ is the unit-area rectangle or *boxcar* function of height τ^{-1} and width τ. The limit of 5 as τ approaches zero is an infinite narrow function centered on $t = 0$, with unit area. It can be shown that one need not start with the rectangle function to obtain the same functional limit, we could just as easily have considered a limit of any set of unit-area functions (e.g., an appropriately scaled set of Gaussian envelopes). Although the delta function may seem outrageously artificial, it actually has a plethora of uses in characterizing physical and theoretical systems.

The usefulness of $\delta(t)$ in our time domain context here largely arises from its *sifting property*, whereby it can retrieve a functional value at a particular argument from witin an integration:

$$\int_a^b f(t)\delta(t-t_0)dt = f(t_0) \tag{6}$$

$$= f(t_0) \quad a \ \leq \ t_0 \ \leq \ b \tag{7}$$

$$= 0 \quad \text{elsewhere} \tag{8}$$

for any $f(t)$ continuous at finite $t = t_0$.

The delta function is one of several related discontinuous functions which will be of use to us. Another is the *step function*

$$\mathrm{H}(t-t_0) = \int_{-\infty}^{t} \delta(\tau - t_0)d\tau \tag{9}$$

which is 0 for $t < t_0$, 1 for $t > t_0$, and takes a discontinuous step at $t = t_0$. The step function is a useful mathematical construction fo "turning on" a system at $t = t_0$.

We can also define the *boxcar function*, $\Pi(t)$, and *sign function*, $\mathrm{sgn}(t)$, in terms of $\mathrm{H}(t)$, as

$$\Pi(t) = \mathrm{H}(t+1/2) - \mathrm{H}(t-1/2) \ . \tag{10}$$

$$\mathrm{sgn}(t) = \frac{|t|}{t} = 2\mathrm{H}(t) - 1 \ . \tag{11}$$

$\mathrm{sgn}(t)$ is also sometimes referred to as the *signum* function.

The *impulse response* of a system is the output produced by an impulse input

$$h(t) = \phi[\delta(t)] \ . \tag{12}$$

We will now show the important result that the response of a linear, time-invariant system to an arbitrary input is characterizable in a simple manner (via a convolution integral) in terms of its impulse response. First, note that any input signal, $f(t)$, can be written as a summation of impulse functions (because of the sifting property of the delta function)

$$f(t) = \int_{-\infty}^{\infty} f(\tau)\delta(t-\tau) \ d\tau \tag{13}$$

Thus, for a general linear system characterized by an operator, ϕ, the response, $g(t)$, to an arbitrary input, $f(t)$, is just that operator acting on (13)

$$g(t) = \phi\left[\int_{-\infty}^{\infty} f(\tau)\delta(t-\tau)d\tau\right] \tag{14}$$

or, from the definition of the integral,

$$g(t) = \phi[\lim_{\Delta\tau \to 0} \sum_{n=-\infty}^{\infty} f(\tau_n)\delta(t-\tau_n)\Delta\tau] \ . \tag{15}$$

Because ϕ characterizes a linear process, we can move it inside of the summation using the scaling relation (3), where the $f(\tau_n)$ are now weights

$$g(t) = \lim_{\Delta\tau \to 0} \sum_{n=-\infty}^{\infty} f(\tau_n)\phi[\delta(t-\tau_n)]\Delta\tau \ . \tag{16}$$

Now note that (16) defines the integral

$$g(t) = \int_{-\infty}^{\infty} f(\tau)\phi(t-\tau)d\tau \tag{17}$$

which is the *convolution* of $f(t)$ and $\phi(t)$, often written in shorthand as

$$g(t) = f(t) * \phi(t) \ . \tag{18}$$

Thus, convolution of a general input signal with an appropriate impulse response exactly describes the corresponding output signal for *any* linear physical process. An important observation along these lines is that a convolution describes the smearing action of a linear measurement tool of limited resolving power. A moment's reflection reveals that a measurement apparatus which records signals from the outside world exactly would need to have a delta function impulse response (so that its output exactly matched the signal in the external world). This should be clear once one realizes that that (13) is itself a convolution, as convolution with a delta function simply returns the input signal, shifted in time (delayed or advanced) by the delta function's origin time

$$f(t) * \delta(t-t_0) \ = \int_{-\infty}^{\infty} f(\tau)\delta(t-t_0-\tau) \ d\tau = f(t-t_0) \ . \tag{19}$$

As all functions can be thought of as continuous integral superpositions of delta functions, it is clear that a necessary and sufficient condition for system stability is that the impulse response be bounded for all t.

Convolution with a step function

$$\int_{-\infty}^{\infty} f(\tau)\mathrm{H}(t-\tau) \ d\tau = \int_{-\infty}^{\infty} f(\tau) \int_{-\infty}^{t} \delta(\xi-\tau)d\xi \ d\tau \tag{20}$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{t} f(\tau)\delta(\xi-\tau)d\xi \ d\tau \tag{21}$$

$$= \int_{-\infty}^{t} \int_{-\infty}^{\infty} f(\tau)\delta(\xi-\tau) \ d\tau d\xi = \int_{-\infty}^{t} f(\tau) \ d\tau \tag{22}$$

is the definite integral of f from $t = -\infty$ up to time t. Thus, while convolution with a delta function returns the system impulse response, convolution with a step function performs operation of definite integration.

$\delta(t)$ can usefully be regarded as the time derivative of $\mathrm{H}(t)$. The significance of convolution with the time derivative of $\delta(t)$ is left as an exercise.

Another useful function that we will use is the *sampling function* or Bracewell's *shah* function:

$$r\text{III}(rt) = \sum_{n=-\infty}^{\infty} r\delta(rt-n) \ . \tag{23}$$

Multiplication by $\text{III}(rt)$ produces a time-continuous representation of a *sampled* time series, with nonzero impulse values at $t = (\ldots,\ -2/r,\ -1/r,\ 0,\ 1/r,\ 2/r,\ \ldots)$, which are scaled by the values of the function at those points. r is referred to as the *sampling rate* (the multiplicative factor of r is required to maintain unit-area delta functions). Such sampled time series (not necessarily in one dimension, but frequently in 2 or more dimensions, and usually uniformly sampled in time or space) make up the vast majority of geophysical and many other types of scientific data.

Time domain interpretation of convolution. A way to develop an intuitive feel for convolution is to graphically examine the operation of the convolution integral

$$c(t) = f_1(t) * f_2(t) = \int_{-\infty}^{\infty} f_1(\tau) f_2(t-\tau)\ d\tau \ . \tag{24}$$

The procedure is as follows:

1. Plot both $f_1(\tau)$ and $f_2(t-\tau)$ on the τ-axis. Note that this operation flips the function $f_2(\tau)$ about the τ-axis and shifts it by an amount t (which is the independent variable of the output function $c(t)$).

2. Visualize that as t advances, $f_2(t-\tau)$ slides along the τ-axis.

3. For each t, the convolution integral (24) is just the area under the product of the two functions, $f_1(\tau)$ and $f_2(t-\tau)$.

As a simple example, consider the convolution of $\Pi(t)$ and the truncated exponential $e^{-t}\text{H}(t)$.

$$c(t) = \int_{-\infty}^{\infty} \Pi(\tau)\text{H}(t-\tau)e^{-(t-\tau)}\ d\tau \ . \tag{25}$$

Because of the discontinuities in $\Pi(t)$, the solution is most easily found by examining three cases:

- Case (a) $t \ \leq \ -1/2$

  The functions do not overlap, so $c(t) = 0$ here.

- Case (b) $-1/2 \ \leq \ t \ \leq 1/2$

  Here, the sliding exponential partially overlaps the boxcar function. The appropriate integral is

$$c(t) = \int_{-1/2}^{t} 1 \cdot e^{-(t-\tau)}\ d\tau = 1 - e^{-(t+1/2)} \ . \tag{26}$$

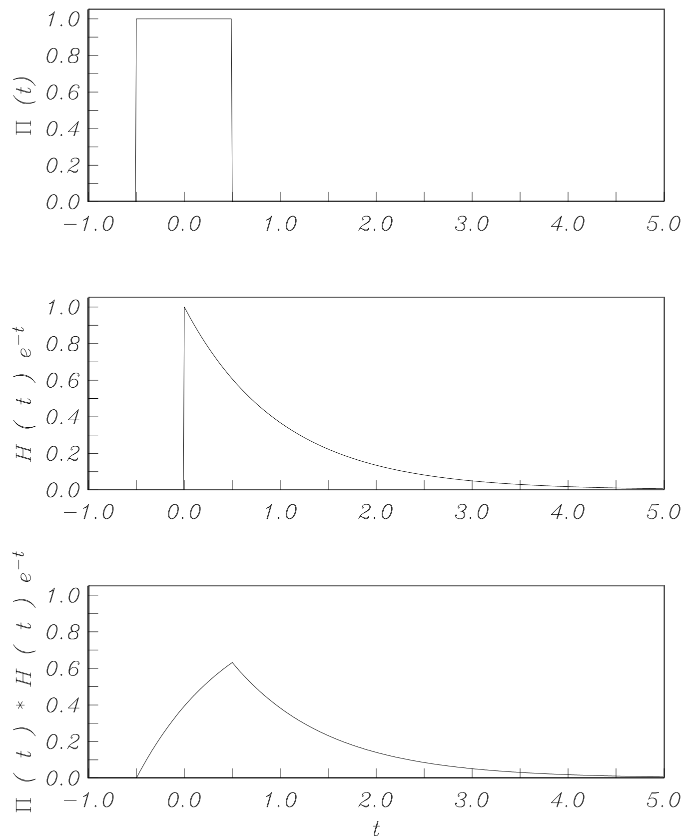

Figure 1: Convolution Example

- Case (c) $t \geq 1/2$

  For this range of t, the sliding exponential completely overlaps the boxcar function and the integral is

$$c(t) = \int_{-1/2}^{1/2} 1 \cdot e^{-(t-\tau)} \, d\tau = e^{-(t-1/2)} - e^{-(t+1/2)} \ . \tag{27}$$

The result of this convolution is plotted in Figure 1.
Note that we could have equivalently written the convolution as

$$c(t) = \int_{-\infty}^{\infty} \Pi(t-\tau)\mathrm{H}(\tau)e^{-\tau} \, d\tau \tag{28}$$

and obtained the same answer with somewhat different integrals. A more efficient and elegant way of evaluating convolutions will become apparent after we learn how to examine functions in the *frequency domain*, rather than the *time domain*.

Autocorrelation and crosscorrelation. Several additional integral operations, closely related to convolution, are commonly used in time series analysis. *Autocorrelation* is similar to *autoconvolution*

$$f(t) * f(t) = \int_{-\infty}^{\infty} f(\tau) f(t - \tau) \; d\tau \tag{29}$$

except that one of the functional components in the τ-domain is not reversed. The autocorrelation of a real function, $f(t)$, is thus

$$A(t) = \int_{-\infty}^{\infty} f(\xi) f(\xi - t) \; d\xi = \int_{\infty}^{-\infty} f(\xi - t) f(\xi) \; (-d\xi) \tag{30}$$

which is, if we let $\xi - t = -\tau$,

$$= \int_{-\infty}^{\infty} f(-\tau) f(t - \tau) \; d\tau = f(-t) * f(t) = f(t) * f(-t) \; . \tag{31}$$

If $f(t)$ is symmetric in time (an *even function*; $f(t) = f(-t)$), then the autoconvolution and autocorrelation are equal. Also, because the autocorrelation integral (31) is unchanged when we interchange $\pm t$, we see that any autocorrelation must be an even function.

It is often convenient to divide (31) by the signal energy to obtain a normalized form for the autocorrelation

$$a(t) = \frac{A(t)}{\int_{-\infty}^{\infty} f^2(\tau) \; d\tau} \tag{32}$$

which has a value that is bounded on the interval $[-1, \; 1]$. Note that for (32) and (31) to converge, the signal energy

$$E = A(0) = \int_{-\infty}^{\infty} f^2(\tau) \; d\tau \tag{33}$$

must be finite. It is thus necessary for $f^2(t)$ to have finite area (zero mean is not sufficient).

The *crosscorrelation* of two functions, $f_1(t)$ and $f_2(t)$ (often referred to simply as the *correlation*) is

$$C(t) = \int_{-\infty}^{\infty} f_1(\tau) f_2(\tau - t) d\tau = \int_{-\infty}^{\infty} f_1(\tau + t) f_2(\tau) d\tau \equiv f_1(t) \star f_2(t) \tag{34}$$

If (34) is divided by the *cross-signal energy* we have a normalized version of the crosscorrelation

$$c(t) = \frac{C(t)}{\int_{-\infty}^{\infty} f_1(\tau) f_2(\tau) d\tau} \; . \tag{35}$$

The autocorrelation and correlation operations have important applications in power spectra, coherency, signal detection and timing, and array processing.

Correlations and Crosscorrelations in Matlab. Matlab has built in convolution *conv*, and crosscorrelation (*xcorr*) functions. The numerical part of Matlab, of course, only operates on finite time series (or *sampled*) representations of functions stored as vectors or arrays of numbers which hopefully adequately represent a continuous function in nature (we will examine the issues associated with sampled functions in much detail later in the course.). The *conv* function thus calculates a sample-by-sample moving dot-product rather than an integral. You are encouraged to experiment with these and other Matlab functions at the earliest opportunity (e.g., try to reproduce the example of Figure 1). Note that if you have two Matlab time series, a_1 and a_2, which are of length n_1 and n_2 samples, respectively, then the convolution output from *conv*, $a_1 * a_2$ will be of length $(n_1 + n_2 + 1)$.