J. A. Mea

1 Introduction

What is the Convolution?

The Convolution is a linear transformation operation, directly analogous to taking the dot product of two vectors; As for a vector v in a vector space V, convolution operates typically on a basis of time, or frequency, in an analytical vector space T governed by dimension with discrete or continuous rank. We refer to these analytical vector space transformations(think of a bilinear relational interaction) as systems- That is, linear, time-invariant systems. The resultant of the input convolved uniformly through the linear, time-invariant system is the system's output. Typically, convolution is that which we use to determine a predictable behavioral output for a given linear system, through which we run through a convolution input and right-multiply the system based on its impulse response. To obtain the impulse response for a linear, time-invariant system, we first convolve a Dirac Delta impulse function through the linear system(the Dirac impulse for the unit impulse response is sort of like a unit vector in the identity matrix); the output of this resultant convolution is thus the impulse response of the system, by which we can use to now predict any output given the system's input. Before we continue, we will review linearity of a vector space and linear transformations in the case of linear, time-invariant systems and then I will give a demonstration of a real-life usage of convolution on a simple physically-bound linear system. Now, onto the laws of linearity for an algebra over a field:

For a linear transformation $\mathbf{T}(\tilde{\mathbf{v}})$, the linear superposition principle holds:

$\mathbf{T}(\tilde{\mathbf{v}}) + \mathbf{T}(\tilde{\mathbf{u}}) = \mathbf{T}(\tilde{\mathbf{v}} + \tilde{\mathbf{u}})$

the scalar coefficient product for linear transformations on vectors also asserts that for a scalar constant $\mathbf{c}$:

$\mathbf{cT}(\tilde{\mathbf{v}}) = \mathbf{T}(\mathbf{c}\tilde{\mathbf{v}})$

$\mathbf{T}$ as a linear transformation follows the form of a linear mapping using the dot product as shown, where $\mathbf{n}$ designates the rank and $\tilde{\mathbf{t}}$ is a basis under the linear map defined in $\mathbf{T}$:

$\mathbf{T}(\tilde{\mathbf{v}}) = \mathbf{v_1}\tilde{\mathbf{t}}_\mathbf{1} + \mathbf{v_2}\tilde{\mathbf{t}}_\mathbf{2} + ... + \mathbf{v_n}\tilde{\mathbf{t}}_\mathbf{n}$

This definition will be extended to a matrix later on when Fourier transforms and tabular convolution are introduced. For the time being, it's important to note that convolution, being a linear transformation is itself applicable only to linear systems in particular, and we will later also come to recognize this central property in the usage of transforms from one domain to its covariant dual comprising the convolution theorem, from which its invertability emerges as a one-to-one and onto correspondence of linear independence for the bases of $\mathbf{T}$

Let's begin now with the discrete definition of the convolution:

$$(\mathbf{x} * \mathbf{h})[\mathbf{t}] = \sum_{\tau=-\infty}^{\infty} \mathbf{h}[\tau] \cdot \mathbf{x}[\mathbf{t} - \tau] = \sum_{\tau=-\infty}^{\infty} \mathbf{x}[\tau] \cdot \mathbf{h}[\mathbf{t} - \tau]$$

Note that the operation above is commutative; when the input to the system x[t] is convolved with the impulse response denoted h[t], the resultant is a sum of products corresponding to flipping the input and evaluating it at every delay, or shifted input, for the sum of the products of the overlapping values which coincide(superimpose) the system, for a given domain value at $\mathbf{t}$ in the resultant output for time = $\mathbf{t}$. The reason for which the kernel input gets flipped over the y-axis at each delay becoming critical to the definition is because when an impulse is sent through the system, the system responds in a specific manner as predicted by taking the dot product as the sum of products, at corresponding time delays one after another, in the sequence they occurred as asserted by the subsequent impulses happening at a later time except delayed, to be multiplied by an impulse with which an earlier input impulse had already undergone before the subsequent value. This makes more sense when you consider that the later values of an input signal, having sequentially greater time $\mathbf{t}$ values, happen after values which come beforehand designated with a lower value for $\mathbf{t}$. If you are in line waiting for your movie ticket at the kiosk, the person behind you mustn't cut ahead, and the person in front of you must receive their ticket ahead of you, earlier than you or the person behind you would. But instead of just one ticket, multiple different tickets are being distributed simultaneously per position in the line at different spots, and each position the line moves forward at $\mathbf{t}$ seconds asserts the total number of tickets received by everyone in their appropriate corresponding position in line at that time as it proceeds with time $\mathbf{t}$ being characteristic of the output at that time $\mathbf{t}$. What manifests this system as linear and time-invariant would be the fact that once the line has passed, every person has received exactly one of every unique ticket, and that at any given time $\mathbf{t}$, no two persons occupy the same spot in the line.

As noted prior, the impulse response is the output of a linear system convolved with an input of the Dirac Delta impulse, $\delta[\mathbf{t}]$. $\delta[\mathbf{t}]$ at $\mathbf{t} = \mathbf{0}$, or $\delta[\mathbf{0}]$, gives an output equal to 1; for all other values of $\mathbf{t}$, the output is equal to 0; we write this as:

$$\delta[\mathbf{t}] = \begin{cases} 1 & t = 0 \\ 0 & t \neq 0 \end{cases}$$

given this definition of the impulse function, to obtain the identity of the impulse response of a linear system $\mathbf{x}[\mathbf{t}]$, the input of the system is to be convolved with the system, giving the output to be equal to the linear identity of this system itself, the impulse response $\mathbf{h}[\mathbf{t}]$, for every linear, time-invariant system:

$$\mathbf{h}[\mathbf{t}] = (\mathbf{x} * \delta)[\mathbf{t}] = \mathbf{x}[\mathbf{t}] = \mathbf{y}[\mathbf{t}]$$

using this impulse response such that $\mathbf{y}[\mathbf{t}]$ is now the designated output for the impulse response, we can now perfectly evaluate the output at every time

$\mathbf{t}$ for the output of any linear system given its impulse response $\mathbf{h}[\mathbf{t}]$ and any input of scaled and summed impulses in $\mathbf{x}[\mathbf{t}]$. Values are not shifted because the offset of the input impulse kernel are not offset from 0.

Taking this newfound definition for the convolution as a scaled and shifted sum of products coinciding at flipped and delayed intervals, it now makes sense to consider a real-world analogy used in subject matters concerning transient and continuous response of a physical system; take for example, a capacitor in a circuit through which current is coursing in and out of. First, a more palatable model if we may, shall be demonstrated to realize this transfer of kinetic energy: a tank with a drain at the bottom, our physical system to model the capacitor with, and an adequately placed drain on the bottom, out from which a fluid such as water or crude oil may flow out of freely at a constant rate. This system is linear and time-invariant because water flows out as the output at a constant rate, no matter which time or delay is applied to this input of a fluid, and secondly, the volume of water total which has emptied from the drain is exactly proportional to the volume of water poured, or titrated, into the source of the tank's fluid output which it contains at any given moment in time. Our impulse function, in this specific case, can be understood as $\delta[\mathbf{t}] = \mathbf{dt}$, or 1 milliliter per second (mL/sec); in reality, for the sake of differential linearity, consider this as an arbitrarily transient titration of water added in an instant. If this droplet can be applied into the input of the system's source, we can measure the total output in the instant of time it is applied to the same input source location, as a function of time, yielding a time delay and an an instantaneous measure of the tank's water volume, with the output volume in that moment corresponding to the rate of change of the tank's fluid contents at that time. Suppose this impulse inputted is equivalent to a measure of 1 unit of water at time $\mathbf{t} = \mathbf{0}$; If we can accurately describe the exact amount of water we are adding water to the tank for each time $\mathbf{t}$, convolution allows us to calculate the total volume of water stored in the tank at each time $\mathbf{t}$. The motivating factor of this example lies beneath the behavior of a capacitor- that is, a capacitor with constant capacitance is considered to be a linear system, with the volume of water in a tank correlating to the charge and the output at a given time correlating to the current released.

2 The Transfer Function, the Fourier Transform

What is the Frequency Response in a System?

The Fourier Transform comes into play in an astoundingly magnificent manner to the study of impulse responses and filters; it can be derived when one considers merging, or rather convolving, an input kernel into and through a system- a kernel that just so happens to be none other than a complex exponential! The Frequency Response, which is simply the Fourier Transform of a system's impulse response, is a frequency-domain spectrum which emphasizes which oscillating input phasors, or rotating complex exponentials, with different angular frequencies coincide with the system, and by how much; this fact capitalizes on the notion of what is known as a harmonic resonant frequency of a given system, which can be described as a signal or even by its impulse response in the case of the frequency response. In other words, a complex exponential with a constant frequency, or in other words an oscillating phasor, produces a certain output of the system it is sent through, governed by its unique characteristic output that is its frequency response with(or for) that specific frequency. Each of these unique constant-valued frequency phasors yields its own characteristic constant magnitude output when sent through the system.

To work with complex exponentials, it is important to first understand both what they are and what they do:

$$\mathbf{Ae^{j(2\pi ft+\phi)} = Acos(2\pi ft + \phi) + jAsin(2\pi ft + \phi)}$$

The aforementioned is what is known as Euler's identity, discovered by the Swiss mathematician Leonhard Euler during the 18th century; It incorporates Euler's special constant $\mathbf{e}$, and associates it with a phase angle θ multiplied by the imaginary unit $\mathbf{j} = \sqrt{\mathbf{-1}}$ by raising $\mathbf{e}$ to the power of the product $\mathbf{j}\theta$. To prove this seemingly arbitrary relationship result, infinite power series are intermediately introduced as a method of approximating irrational constants for trigonometric functions and the definition of raising $\mathbf{e}$ itself to any power. This complex exponential is graphed as a sinusoidal helix projected parametrically as a phasor with amplitude(magnitude) $\mathbf{A}$ on the complex plane with constant frequency $\mathbf{f}$ over variable $\mathbf{t}$ at phase delayed ϕ radians. Notice how the period of this function is standardly $\mathbf{1/(2\pi f)}$. For the sake of simplification of notation, let's denote frequency from revolutions of the phasor, or oscillations, per second, to radians/sec or $\mathbf{rad * Hz = 2\pi f} = \omega$, following suit of phase where phase ϕ is given in terms with regards to radial frequency:

$$\mathbf{Ae^{j(2\pi f(t-\tau))} = Ae^{j(2\pi ft - 2\pi f\tau)}}; \phi = \mathbf{-2\pi f}\tau$$

$-\tau$ in this case is the constant of phase delay, in terms of phase shifted right along the time $\mathbf{t}$ axis as a period of the time it takes for the phasor, or complex sinusoid, to make a full revolution around the time $\mathbf{t}$ axis. Let's recall the radial phase angle frequency and rewrite the complex exponential in those terms:

$$\mathbf{Ae^{j(\omega t+\phi)} = Acos(\omega t + \phi) + jAsin(\omega t + \phi)}$$

For now, for sake of ease, instead of using normalized frequencies in the context of discrete-time sampling periods, let's stick to defining time in terms of $\mathbf{t}$ for impulse responses. Given the definition of our complex exponential kernel which will be used to yield the frequency response output of a linear system, let's define it as a discrete function of $\mathbf{t}$, with 0 delay phase shift and magnitude 1:

$$\mathbf{x[t] = e^{j(\omega t)}}$$

Next, we try running it through a convolution in the system that is our impulse response and observing what happens:

$$(\mathbf{x} * \mathbf{h})[\mathbf{t}] = \sum_{\tau=-\infty}^{\infty} \mathbf{h}[\tau] \cdot \mathbf{x}[\mathbf{t} - \tau] = \sum_{\tau=-\infty}^{\infty} \mathbf{h}[\tau] \cdot \mathbf{e}^{\mathbf{j}\omega(\mathbf{t}-\tau)}$$

If the above is expanded further, it ends up as:

$$\sum_{\tau=-\infty}^{\infty} \mathbf{h}[\tau] \cdot \mathbf{e}^{\mathbf{j}(\omega(\mathbf{t}-\tau))} = \sum_{\tau=-\infty}^{\infty} \mathbf{h}[\tau] \cdot \mathbf{e}^{\mathbf{j}\omega\mathbf{t}} \cdot \mathbf{e}^{-\mathbf{j}\omega\tau} = \mathbf{e}^{\mathbf{j}\omega\mathbf{t}} \cdot \sum_{\tau=-\infty}^{\infty} \mathbf{h}[\tau] \cdot \mathbf{e}^{-\mathbf{j}\omega\tau}$$

We moved $\mathbf{e}^{\mathbf{j}\omega\mathbf{t}}$ outside the sum as a scalar coefficient because the convolution is linear and operates on linear systems, and as stated prior, linearity holds that $\mathbf{cT}(\tilde{\mathbf{v}}) = \mathbf{T}(\mathbf{c}\tilde{\mathbf{v}})$, therefore we ended up with the original function $\mathbf{x}[\mathbf{t}] = \mathbf{e}^{\mathbf{j}(\omega\mathbf{t})}$ on the outside, with the convolution being a constant at each output, which asserts itself as a scalar coefficient of the linear system. Alas,

$$(\mathbf{x} * \mathbf{h})[\mathbf{t}] = \mathbf{e}^{\mathbf{j}\omega\mathbf{t}} \cdot \sum_{\tau=-\infty}^{\infty} \mathbf{h}[\tau] \cdot \mathbf{e}^{-\mathbf{j}\omega\tau} = \mathbf{e}^{\mathbf{j}\omega\mathbf{t}} \cdot \mathbf{H}(\omega)$$

Where $\mathbf{H}(\omega)$ is the frequency response of $\mathbf{h}[\mathbf{t}]$. What it is doing in this eigenfunction kernel convolution is running infinite-interval unbounded samples of complex rotating phasors according to their unique frequency in reverse, forward through the impulse response of the system in the same fashion in which convolution is executed, and outputting whether they coincide perfectly or imperfectly; perfect superimpositions when convolution is performed through the system diverge to a constant phasor with a magnitude known as the gain of the system, and imperfect occurrences converge to an arbitrarily small value and evaluate to 0 at infinity, overtaking asymptotal limiting behavior; these factors in the outputted dot product when observed as multiplier coefficient factors of the inputted complex exponential phasor signal in the outputted product either amplify the system's outputted input phasor magnitude by a coefficient and shift/delay its phase, or they simply don't! Plotting this transformation of domain on the axis of a dual-sided frequency domain spectrum is the linear transformation of analytically decomposed phasors on transformed analytical vector space basis vector coordinates. We can demonstrate this with a special type of matrix derived from the definition of convolution as a linear transformation and the complex exponential kernel as an eigenvector, but for now let's stick to extrapolating the definition of the convolution operation to a more workable and practical continuous-time domain, for analyzing analog signals in an analytical vector space. To accomplish this, first extend the convolution to an integrable working definition, with which the argument t is enclosed in parentheses:

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

This continuous analog for the summation form of the convolution definition is not only more workable when considering kernels that are continuous arguments themselves, but is also more applicable to realtime systems such as the water canister analogy for capacitors, and gives us non-aliased samples for filters of continuous-time impulse response and thus a clear-cut finite frequency-domain spectrum highlighting the phasors present in an analog signal which is also infinite! To demonstrate the frequency response in the same manner, just replace the summation blocks with integration- since integrate is after all just an infinitely precise sum over a finite interval which could be unbounded while preserving its precision:

$$\int_{-\infty}^{\infty} \mathbf{h}(\tau) \cdot \mathbf{e}^{\mathbf{j}(\omega(\mathbf{t}-\tau))}\,\mathbf{d}\tau = \int_{-\infty}^{\infty} \mathbf{h}(\tau) \cdot \mathbf{e}^{\mathbf{j}\omega\mathbf{t}} \cdot \mathbf{e}^{-\mathbf{j}\omega\tau}\,\mathbf{d}\tau = \mathbf{e}^{\mathbf{j}\omega\mathbf{t}} \cdot \int_{-\infty}^{\infty} \mathbf{h}(\tau) \cdot \mathbf{e}^{-\mathbf{j}\omega\tau}\,\mathbf{d}\tau$$

$$(\mathbf{x} * \mathbf{h})(\mathbf{t}) = \mathbf{e}^{\mathbf{j}\omega\mathbf{t}} \cdot \int_{-\infty}^{\infty} \mathbf{h}(\tau) \cdot \mathbf{e}^{-\mathbf{j}\omega\tau}\,\mathbf{d}\tau = \mathbf{e}^{\mathbf{j}\omega\mathbf{t}} \cdot \mathbf{H}(\omega)$$

The frequency response is simply:

$$\mathbf{H}(\omega) = \sum_{\tau=-\infty}^{\infty} \mathbf{h}[\tau] \cdot \mathbf{e}^{-\mathbf{j}\omega\tau}$$

for Finite-Impulse Response(FIR) Filters, and:

$$\mathbf{H}(\omega) = \int_{-\infty}^{\infty} \mathbf{h}(\tau) \cdot \mathbf{e}^{-\mathbf{j}\omega\tau}\,\mathbf{d}\tau$$

for a continuous analog impulse filter.

If you notice the naturally emergent behavior of the process by which we projectively scale each frequency phasor as an eigenfunction convolved and output as a scalar with the system, we can extrapolate the behavior of this coefficient scaling factor as the input kernels to materialize the Fourier Transform of any given system, provided that the transform gives a linear transformation between domains of linearly independent analytical vector spaces. The following is the Discrete-Time Fourier Transform (DTFT):

$$\mathbf{F}(\hat{\omega}) = \mathscr{F}\{\mathbf{f}[\mathbf{n}]\} = \sum_{\mathbf{n}=-\infty}^{\infty} \mathbf{f}[\mathbf{n}] \cdot \mathbf{e}^{-\mathbf{j}\hat{\omega}\mathbf{n}}$$

Where $\hat{\omega}$ is the normalized radial frequency; The following is the Continuous-Time Fourier Transform:

$$\mathbf{F}(\omega) = \mathscr{F}\{\mathbf{f}(\mathbf{t})\} = \sum_{\mathbf{n}=-\infty}^{\infty} \mathbf{f}[\mathbf{n}] \cdot \mathbf{e}^{-\mathbf{j}\hat{\omega}\mathbf{n}}$$