Convolutions
by Dr. Colton, Physics 441/471 (last updated: Fall 2019)

What is a convolution? Physical ideas to think about.

1. Spectrometer. A spectrometer allows you to separate light into its colors by means of a diffraction grating. It acts basically like a large prism. To measure the amount of light at the different colors, you place a detector in the diffracted beam (or in the refracted beam, in the case of a prism), and move it back and forth as you record the intensity. However, the detector is not infinitely narrow, even though a narrow slit is often used—so if you attempt to measure how much light is at wavelength λ_0, what you're *actually* measuring is the light intensity for a small range of wavelengths around λ_0. How is what you are actually measuring related to the true $I(\lambda)$?

2. Shadows. A light source creates a shadow of an object. If the light source is point-like, the shadow that is produced can be called the "true shadow". However, an actual light source is not a true point, but rather has a certain physical extent. How is the shadow created by an actual light source different than the true shadow?

How would the first example change if the detector were even larger? How would the second example change if the light source were even larger?

Definition

Convolutions frequently arise when you want to know a measurement for a point detector or point source, but have a source/detector of finite size. In that case, the thing you actually measure is said to be the true response *convolved* with a "kernel function" (or more simply, just a "kernel") that models the finite size of source/detector. The convolution is often represented by the symbol $\otimes$, and the mathematical definition of the convolution of two functions is:

$$g(t) = a(t) \otimes b(t) = \int_{-\infty}^{\infty} a(t')b(t - t')dt'$$

$g(t)$ is the convolution of the two functions $a(t)$ and $b(t)$.[1] Here $a(t)$ could be the true response and $b(t)$ the kernel, or vice versa—it doesn't matter because the convolution is commutative, as can readily be proved from the definition.

$$a(t) \otimes b(t) = b(t) \otimes a(t) \qquad \text{Commutative property}$$

The kernel, let's call it $b(t)$, is sometimes called a "weighting function", and can be thought of as a window through which we observe the function $a(t)$. In the spectrometer example, the kernel function is perhaps even a literal window, namely the entrance to the detector. When a convolution is done in time rather than space, a convolution could arise, for example, if your detector cannot response to your signal infinitely quickly; the kernel would in this case be a function describing how your detector responds to a quick change. Such a kernel is sometimes called the "instrument response function".

Performing a convolution generally smoothes the original function $a(t)$ such that sharp peaks are rounded and steep slopes are reduced. The amount of smoothing depends on the nature of the two functions $a(t)$

---

[1] These could obviously be functions of space rather than of time, but I'll use time as the independent variable here for simplicity.

and $b(t)$. Convolutions are used in the data processing world to smooth out noise (a type of filtering), to do digital image processing, and so forth. Wolfram MathWorld puts it like this:

> *A convolution is an integral that expresses the amount of overlap of one function as it is shifted over another function. It therefore "blends" one function with another.*[2]

Mathematica demonstration

I have created a demonstration in Mathematica to help you improve your intuitive understanding of what a convolution is. You can download it from the class website.

For a given kernel (tan function to the right) and a given function $f(x)$ (blue function to the right), it displays the convolution of the two (blue function on bottom graph). It does so by representing each point along the resulting convolution with a Manipulate setting, indicated by the x0 bar at the top of the graph, and by the red dashed lines. The value of the convolution function at a given point x0 is $f(x)$ multiplied by the kernel shifted to that point, then integrated from $-\infty$ to ∞ in order to calculate the area of overlap. The area is depicted in the middle graph.

Play around with the demonstration to improve your convolution intuition. I suggest that as a minimum you try these four things. Make sure you understand the results you get.

1. Change the kernel to be a narrower and a wider Gaussian function. In particular, what happens when the kernel approaches a delta function?
2. Change the peaks of $f(x)$ (size, width, number) while keeping the kernel the same.
3. Simulate some high frequency noise by adding something like $0.1\text{Sin}[100x]$ to $f(x)$; see how the convolution then acts a bit like a low pass filter.
4. See what happens when the kernel and/or $f(x)$ are square pulses, by changing one or both to $\text{UnitBox}[x]$ or $\text{UnitBox}[2x]$. Can you predict the results?
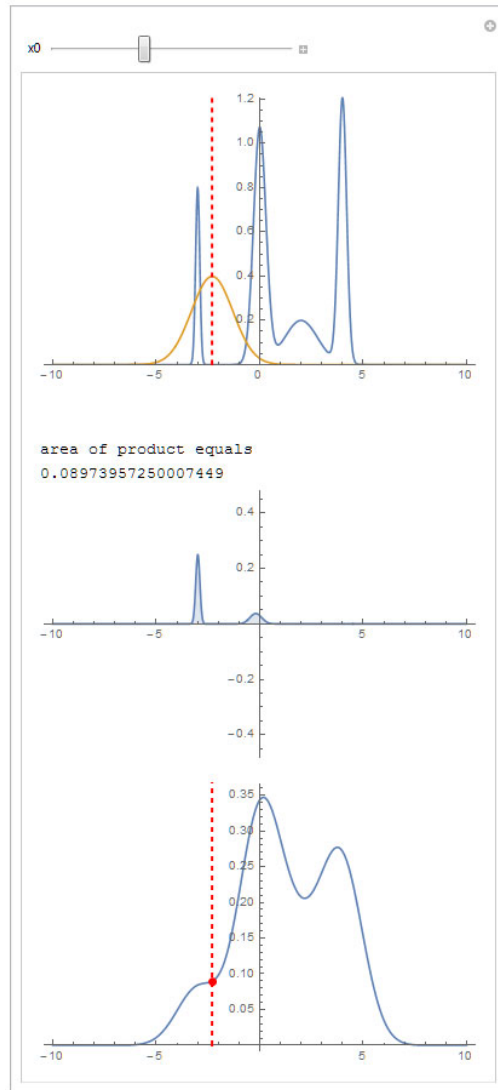
Convolution of a delta function

If you convolve a function $f(t)$ with a delta function at the origin, you obtain the following:

$$f(t) \otimes \delta(t) = \int_{-\infty}^{\infty} f(t')\delta(t-t')dt'$$
$$= f(t)$$

---

[2] http://mathworld.wolfram.com/Convolution.html

In other words, due to the sifting property of the delta function, you just get the original function back! Hopefully this matches what you found when playing with the Mathematica file. That is because, using the spectrometer example, the delta function is like using an infinitely narrow slit in front of your detector (while still allowing light through) instead of using a slit with finite width. An infinitely narrow slit will give you the true spectral response.

If the delta function is located at t_0 instead of at the origin, here's what happens:

$$f(t) \otimes \delta(t - t_0) = \int_{-\infty}^{\infty} f(t')\delta\big((t - t_0) - t'\big)dt'$$
$$= f(t - t_0)$$

The convolution gives you the original function back, shifted by t_0.

Convolution theorems

An interesting result occurs if you take the Fourier transform of a convolution. This is done in a homework problem in Peatross & Ware, *Physics of Light and Optics*, P0.26(a), and is called the Convolution Theorem:

$$FT\{a(t) \otimes b(t)\} = \sqrt{2\pi}\, FT\{a(t)\} \cdot FT\{b(t)\}$$

In other words, aside from the factor of $\sqrt{2\pi}$, **the Fourier transform of a convolution is the product of the Fourier transforms**.

Here's a related theorem from Peatross & Ware, P0.26(b), and is sometimes also called the Convolution Theorem:

$$FT\{a(t) \cdot b(t)\} = \frac{1}{\sqrt{2\pi}}\, FT\{a(t)\} \otimes FT\{b(t)\}$$

In other words, aside from the factor of $1/\sqrt{2\pi}$, **the Fourier transform of a product is the convolution of the Fourier transforms**.

Related forms of the above two theorems also apply to inverse Fourier transforms. Also, be careful about the factors of $\sqrt{2\pi}$. They depend on the specific definitions of Fourier transforms that are being used, and don't always show up in this same way.

The bolded statements summarizing the convolution theorems are so important that <u>they should be memorized</u>.

As an example: they can allow you to deduce the true response from a measured response, if you have a good idea (or model) of what your kernel function is. Suppose $a(t)$ is the true response you want to know, $b(t)$ is the kernel which perhaps you can measure or can look up in the instrument specs, and $a(t) \otimes b(t)$ is your measured data. Dropping the $\sqrt{2\pi}$ and the "(t)"s for simplicity, we can manipulate the first convolution theorem like this:

$$FT\{a \otimes b\} = FT\{a\} \cdot FT\{b\}$$

$$FT\{a\} = \frac{FT\{a \otimes b\}}{FT\{b\}}$$

$$a = FT^{-1}\left\{\frac{FT\{a \otimes b\}}{FT\{b\}}\right\}$$

$$true\ response = FT^{-1}\left\{\frac{FT\{measured\ data\}}{FT\{kernel\}}\right\}$$

In other words, to get the true response, you can take the Fourier transform of your measured data, divide by the Fourier transform of the kernel, and then take the inverse Fourier transform of the result. Think about what that implies: even if your detector cannot respond faster than, say, 1 nanosecond, you can still obtain valid data for times shorter than 1 ns, just as long as you know what the exact time response of your detector actually is. That's amazing! But it really works, and I've used that trick in my lab. We measured the kernel by recording the time response of our detector for a delta function-like impulse. Then we did precisely that calculation to determine the true response of the material we were measuring even though the true response was faster than our detector could measure. This is called doing a "deconvolution" of the data.

Disclaimer: If you have experimental noise in your data that process usually fails. In that case a better method is often to make a model for the true response, then convolute the model with the known kernel function, then adjust the model until it matches your data. This is called doing an "iterative reconvolution" fit.