convolution: *son et lumière*

Alex Barnett

October 5, 2009

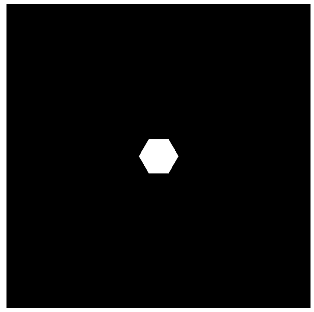
Exterior night, cityscape. Thousands of distant points of light lie still in the summer air. Below them a neon diner sign flickers silently across a nearby empty street. Suddenly the close up face of the main character moves into the frame, the focus pulls, bringing their silent profile into crisp silhouette, while blurring the urban backdrop into a splash of color. Each point of light blooms into a hexagon, the neon sign now an unreadable splotch of red. The main character pauses for a second, then moves on into the night. Fade to black.

Something special just happened in this imagined movie scene, when viewed through a scientific lens. The cityscape image, in its transition from crisp to out-of-focus, underwent a *convolution*, a process with a precise mathematical meaning that will become *our* main character in this short piece. I will try to explain as accurately as possible the concept of convolution, using familiar examples from our everyday lives. Convolution is in fact all around us—we just have to attune ourselves to it, to shift to a viewpoint in which lights and sounds and motions become patterns of numbers and signals. It could be argued that this type of shift, with the logical power to model and predict that it brings, is an idea that has opened the doors to centuries of tremendous human scientific and technological creation. I hope to leave you with a glimpse of how an engineer or scientist, or at least this applied mathematician, thinks about the world.[1]

The classic cinematic trope above, toying with depth of field, mirrors what our eyes and brain do daily, mostly unconsciously, as we shift attention in a complex three-dimensional visual world. The myopic among us may recreate the same blurring effect for free, simply by taking off our glasses and staring into the distance (here the hexagons are instead discs). Let's examine more carefully what took place. We may treat the cityscape image as a pattern of light intensity existing across a two-dimensional plane, as in a rectangular photographic still, or a grid of pixels on your computer screen. Just like Manhattan's grid of avenues and streets, each point in this plane has a location given by its horizontal distance from the left side (x coordinate) and vertical distance up from the bottom (y coordinate). At each such point the image has a brightness, which may treat as a number, zero for black, one for bright, two for very bright, ... and so forth. So we have a number at each point in the plane: mathematically this

[1] In footnotes I will add a flavor of the mathematical formulae, but these may be skipped without missing much of the story.

scene image f



aperture image g

convolution image of f and g

is called a function.[2]

Where did the hexagons come from? They are copies of the shape of the camera aperture or iris, the other ingredient in the mix. Return to the opening scene. If the cityscape had instead consisted of only a *single* bright point of light, you can picture what would happen upon pulling focus: the single point blurs into a single, uniformly-bright hexagon, surrounded by blackness. The key is to realise that this bright-hexagon-on-black-background is *also* itself an image (a brightness pattern existing over a two-dimensional plane)—we might call it the 'aperture image'. We now have two images: the cityscape 'scene image' and the aperture image (see figure). The act of blurring the scene image (a function, let's call it f) using the aperture image (another function, let's call it g), is precisely to *convolve* one by the other, making a new image written mathematically as,[3]

$$f * g$$

The rule is that each point of light in the scene is replaced by an identical copy of the hexagonal aperture image, and these individual copies are then layered on top of each other, or added up, to give the final blurry picture.[4] If you like, we are painting the original scene using a hexagon-shaped paintbrush. The act was performed by turning a focus ring on a camera, sending a flurry of photons into different places on the film. However, it may also be performed mathematically or computationally, and, once it has been defined precisely, abstracted to any images, functions or data, real or imagined: therein lies its power.

It is a beautiful and not obvious fact that if we swapped the roles of our two images, the picture produced would be exactly the same as before. In other words convolution has a *symmetry*,[5] and that is something that delights the mathematical glands, for both aesthetic and practical reasons. How may

[2] We could name this function f, in which case the brighness at a point (x, y) is $f(x, y)$.

[3] Note that the asterisk means convolution, not mere multiplication.

[4] The adding or summing up operation needs to be done for every single point in the scene image: there are a continuum of points, which mathematically turns the summation into an *integral*, and since there are two directions to sum over, a double integral. Hence the formula for a two-dimensional convolution, $(f * g)(x, y) = \int\int f(w, z)g(x - w, y - z)dw\,dz$.

[5] We may state $g * f = f * g$ as functions, for all f and g. We won't prove it rigorously here.

we convince ourselves of this fact by mental visualization? In the swapped situation we take each point of light within the aperture hexagon and replace it by a copy of the whole cityscape scene, suitably jiggled to be centered at the point on the hexagon from which it came. Again we add up all these copies. The brightness at any point is now a sum of intensities over all the nearby points in the scene image lying in a hexagon centered at that point. But, thinking back to the original situation, we realise this describes that too. Painting a scene using an aperture-shaped brush is equivalent to painting an aperture using a scene-shaped brush!

Let's switch senses from light to sound; we will find convolution lurking almost everywhere we listen. Speak aloud in a large echoey room and your usual speech is changed by the acoustics of the space into a boomy, muddied version: this is *convolution of two sounds*. What do our actors f and g signify here? A sound is air pressure or vibrational motion changing very rapidly in time, which we may describe as a function.[6] Speak aloud outdoors in still air and your voice sounds crisp and dry—this is, give or take, your voice's source function f. On the other hand, clap in the silent large room, and you will hear an echoey reverberation dying away—this is the room's 'response function' g. Now combine the two: emit your voice into the room and the resulting[7] audio signal is precisely $f * g$. To help visualize this the reader might enjoy messing around with a wonderful online interactive demonstration of convolving two hand-drawn functions [1].

It is worth drawing some connections between this and our original case of images. Muddying of your voice by a large echoey room is analogous to blurring of the scene by a wide out-of-focus camera aperture. In order to produce a final result that was simply g, we sent in for f a 'clap' (very short spike-like signal) in the audio case, or a 'bright point' (very narrow spike-like image) in the image case.[8] The power of applied mathematics is that these seemingly unrelated phenomena are described by (essentially) the same formulae.

The sonic example allows us to explain a final property of convolution—one which has many applications in signal processing—again using everyday experiences. I mentioned that sounds are pressure variations as a function of time; however, there is a complementary way to describe them by the strength (amplitude) with which each component *frequency* is present.[9] Many home stereos (and audio player applications) show with a set of illuminated bars how much signal strength is present in the ranges 50-250Hz, 250-1000Hz, etc, from the bottom to the top of our hearing range. Moreover, by adjusting the 'graphic equalizer' on your stereo (or pressing buttons such as 'bass boost') you are able to selectively amplify or reduce the strengths in each of these frequency

[6] Now the function depends on only one variable, the time t. It is often called a signal.

[7] The relevant formula for convolution in one variable is $(f * g)(t) = \int f(s)g(t-s)ds$.

[8] Mathematically these special short signals are known as *delta distributions*.

[9] Recall that frequency is how many times per second something vibrates, and, crudely speaking, corresponds to musical pitch. In isolation, a single frequency sounds like a pure tone such as that made by a tuning fork. Combinations of frequencies create all the wonderful variations in musical timbre. The frequency description is called the *Fourier transform*.

bands, 'shaping' or coloring the sound according to a curve. Mathematically this multiplies the amplitudes by a function which depends on frequency: for instance if you make this function larger at high frequencies than at low ones you will get a bright, tinny, or sibilant effect. If instead you choose a 'flat' or constant function you don't color the sound at all, just like the humble volume knob which merely multiplies the amplitude by the *same* number at all frequencies.

We now have a picture of coloring a sound by enhancing or subduing certain frequencies: it turns out that this is in fact *precisely* what an echoey room does to the sound of your voice in the previous example, when viewed through a 'frequency lens'. Probably the most convincing demonstration of this is by singing in the shower: a shower cabinet is a small echoey room so we know it has a response function (clap and you'll hear it); on the other hand you know that when you sing or hum you find that certain notes of the scale stand out as surprisingly loud and resonant compared to others, indicating heavy coloration of your voice. The remarkable property we have thus shown is that this process of coloring (or multiplying) in the frequency picture is identical to convolving the sound with some response function. All of this can be proven rigorously,[10] and used to manipulate audio with computers to great effect [2].

So, how does this mathematical toolbox help humanity? One real-world problem that repeatedly crops up is that of *deconvolution*: say you took an out-of-focus image, or your camera moved by accident during the exposure, turning your image into a riot of useless squiggly lines—how could you recover the original unsullied image? Applications vary from forensic investigation of blurred photographs of license plates, to astronomy [3], seismic imaging, and improving the MRI images used to diagnose cancer [4]. Maybe the discussion above has given you some hints: transform into the 'frequency picture', then undo the 'graphic equalizer' effect (by dividing rather than multiplying), finally transform back to the signal or image space. Barring certain computational details and tricks, this is in fact what is done: the results can be amazingly good. Problems and failures can occur, most notably when the aperture function is very smooth, or is unknown. Worse yet, there are situations where a convolution framework is simply not general enough, such as when blurriness or aperture size varies across a single image. Something beyond deconvolution would be required; mathematicians and computer scientists are hard at work on this and related signal recovery and imaging problems.

Convolution is in the sun's dappled shadows on the grass, in the vocal reverb you use to add pizazz to your demo CD, in the way heat spreads out on your dining table after you have placed your hot coffee mug on it, and in the multiplication of binary numbers happening every nanosecond in every computer on the planet.[11] Mathematics is woven through the world around us … if only we choose to bring it into focus.

[10] If the hat $\hat{}$ symbol signifies the *Fourier transform* (frequency picture) of a signal, then what I have just explained is known as the *convolution theorem*: $\widehat{f * g} = \hat{f}\hat{g}$.

[11] The first two of these examples are similar to the cases I described earlier; the latter two are more sophisticated, yet convolutions nonetheless.

The author thanks Liz Canner and Wendy Hui Kyong Chun for fruitful discussions.

References

[1] S. Crutchfield, *Joy of Convolution*, web applet (accessed 10/5/09), `http://www.jhu.edu/signals/convolve/`

[2] S. J. Orfanidis, *Introduction to Signal Processing* (Prentice Hall, Englewood Cliffs, NJ, 1996)

[3] J. L. Starck and F. Murtagh, *Astronomical Image and Data Analysis* (Springer, 2006)

[4] M. Bertero and P. Boccacci, *Introduction to Inverse Problems in Imaging* (Taylor & Francis, 1998)