Introduction to Linear Systems

Today we begin the second part of the course, which addresses image analysis. We will start off with the basic elements of linear system theory. The main application that we will start with is to detect edges in the presence of image noise.

Example: local difference

One common image analysis operation involves taking a linear combination of the intensities of nearby pixels. An example is to take the local difference of intensities. Let's just look at the problem in 1D, and consider the operation D which takes an image $I(x)$ and computes an approximation of a derivative

$$DI(x) = \frac{1}{2}I(x+1) - \frac{1}{2}I(x-1).$$

What is the effect of such a transformation? One key idea is that such a derivative would be useful for marking positions where the intensity changes. Such a change is called an *edge*. It is important to detect edges in images because they often mark locations at which object properties change. These can include changes in illumination along a surface due to a shadow boundary, or a material change. They also include changes in depth, as when one object ends and another begins. (In this case, there is typically a material and illumination change.) The computational problem of finding intensity edges in images is called *edge detection*.

We could then look for positions at which either $DI(x)$ has a large negative or positive value. Large positive values indicate an edge that goes from low to high intensity, and large negative values indicate an edge that goes from high to low intensity. For example, suppose the image consists of a single edge:

$$I(x) = \begin{cases} 100, & x > x_0 \\ 70, & x = x_0 \\ 40, & x < x_0 \end{cases}$$

Then,

$$DI(x) = \begin{cases} 0, & x > x_0 + 1 \\ 15, & x = x_0 + 1 \\ 30, & x = x_0 \\ 15, & x = x_0 - 1 \\ 0, & x < x_0 - 1 \end{cases}$$

Thus, DI takes large values at $x = x_0$.

If we think of DI as the first derivative of the image, then we see that there is a maximum in the absolute value of DI at the edge, that is, *an edge produces a peak in the first derivative*. Intuitively, this is very easy to understand, since an edge gives a high slope from one intensity to the next.

(Discrete) Convolution

The discrete derivative is a linear transform of the image intensities. But it is a special kind of linear transform in which the weights of linear combination depend only on the neighborhood relations between the points. Such a linear transformation is called a *convolution*. Let's first define convolution formally for 1-D images, then afterwards we'll define it for 2-D images.

If we have a function $I(x)$ where x is integer valued, define the discrete convolution of $I(x)$ with another function $f(x)$ to be

$$f(x) * I(x) \equiv \sum_{x'} f(x') \, I(x - x').$$

For the local difference operator above, we have

$$f(x) = \begin{cases} \frac{1}{2}, & x = -1 \\ -\frac{1}{2}, & x = 1 \\ 0, & \text{otherwise} \end{cases}$$

Local averaging is another example of a convolution. Consider a another function

$$f(x) = \begin{cases} 1/4, & x = -1 \\ 1/2, & x = 0 \\ 1/4, & x = 1 \\ 0, & \text{otherwise} \end{cases}$$

Local averaging is sometimes called *smoothing*.

Here is yet another example of a convolution.

$$f(x) * I(x) = -3 \, I(x+2) + 4 \, I(x+1) + 2 \, I(x-2)$$

Now the function $f(x)$ is

$$f(x) = \begin{cases} -3, & x = -2 \\ 4, & x = -1 \\ 2, & x = 2 \\ 0, & \text{otherwise} \end{cases}$$

These functions $f(x)$ which we convolve with images are often called *filters*. We say that we are "filtering" the image with a filter $f(x)$.

Boundary conditions

In the above definition of convolution, we did not specify the indices x' over which we are summing. There are a few ways to define these indices.

One way is to treat the intensities $I(x)$ as 0 "beyond" the image boundary i.e. beyond the indices $0, \ldots, N-1$. This is called *pad with zeros*, meaning that we treat $f(x)$ and $I(x)$ as if they each have infinite domains $x \in \{\ldots, -1, 0, 1, \ldots, N-1, N, \ldots\}$ but the values of $I(x)$ are zero outside of $x = 0$ to $N - 1$.

Another way to define the indices is to treat x as periodic so that the index of x is $x \bmod N$. In this case, we define a *circular convolution*

$$(I * f)(x) \equiv \sum_{x'=0}^{N-1} f(\, x' \bmod N) \, I((x - x') \bmod N) \, .$$

Notice that the mod operator in $f(x' \bmod N)$ is not really necessary here. However, if you want to perform circular convolution, then you need to define $f(x)$ on values in $0, \ldots, N-1$, for example, $-1 mod N = N - 1$. Compare with the $f(x)$ functions above, which were defined on negative values of x.

Algebraic properties of convolution

One very important property the convolution operation is that it *commutative*: one can switch the order of the two functions I and f in the convolution without affecting the result. To prove that convolution is commutative, we pad $I(x)$ and $f(x)$ with zeros. This allows us to take the summation from $-\infty$ to ∞. Using the substitution $w = x - x'$, we have

$$f(x) * I(x) = \sum_{x'=-\infty}^{\infty} f(x')\, I(x - x') = \sum_{w=-\infty}^{\infty} f(x - w)I(w) = I(x) * f(x)$$

Be careful. The definition of convolution is NOT:

$$\sum_{x'} f(x')\, I(x' + x).$$

Such a linear operation is called the *cross-correlation* of $f(x)$ and $I(x)$. Convolution and cross-correlation are identical if $f(x)$ is symmetric ($f(x) = f(-x)$). But the $f(x)$ we will be considering are often not symmetric. The commutativity property doesn't hold in general for cross correlation.

A second important property of convolution is that it is *associative*:

$$I * (f_1 * f_2) = (I * f_1) * f_2$$

The proof is simple, and you should work it out for yourself.

Why are these properties useful? Often, in signal processing, we perform a sequence of operations on a set of images. For example, you might average the pixels in a local neighborhood, then take their derivative (or second derivative) in some direction. The algebraic properties just described give us some flexibility in the order of operations.

A third useful property of convolution is that it is *distributive*:

$$(I_1 + I_2) * f = I_1 * f + I_2 * f$$

(This is also simple to prove and I leave it to you as an exercise.) This property is also useful. For example, if $I_1 = I(x)$ is an image and $I_2 = n(x)$ is a noise function that is added to the image, then if we blur and take the derivative of the "image+noise," we get the same result as if we blur and take the derivative of the image and noise functions separately, and then add the results together. We will see many examples of this.

Impulse function

Another way to think about convolution is in terms of the function

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

which is called the (discrete) *impulse function*. It is straightforward to show that, for any $f(x)$,

$$\delta(x) * f(x) = f(x).$$

For this reason, we sometimes call the filter $f(x)$ an *impulse response function*, since it is the output that you get when you convolve the impulse function with $f(x)$.

Next notice that

$$f(x) * \delta(x - x_0) = f(x - x_0) \ .$$

This is interesting, because we can write any function $I(x)$ as a sum of shifted impulse functions, namely

$$I(x) = \sum_u I(u)\delta(x - u)$$

and so

$$
\begin{aligned}
(f * I)(x) &= f(x) * \sum_u I(u)\delta(x - u) \ . \\
&= \sum_u I(u)f(x) * \delta(x - u) \ . \\
&= \sum_u I(u)f(x - u)
\end{aligned}
$$

which is just the definition of convolution. This allows us to visualize what convolution does. It adds up the "responses" of the filter $f(x)$ to a sum of shifted impulse functions.

(Continuous) Convolution

It is often convenient to work with a continuous domain x rather than a discrete domain, since this allows one to use the tools of Calculus. Here I will sketch out the basic ideas, and next lecture we will see how to use them.

If we have functions $I(x)$ and $f(x)$ where x is real valued on $[\infty, \infty]$, define the continuouous convolution of $I(x)$ with $f(x)$ to be

$$(I * f)(x) \equiv \int_{-\infty}^{\infty} I(x') \ f(x - x')dx'.$$

You should look at this formula and imagine that you adding up infinitely many shifted functions $f(x - x')$ such that the one shifted by x' is weighted by $I(x')dx'$.

The continuous version of the impulse function is defined as follows.

$$
\delta_\epsilon(x) = \left\{ \begin{array}{ll} \frac{1}{\epsilon}, & |x| \leq \frac{\epsilon}{2} \\ 0, & \text{otherwise} \end{array} \right.
$$

and

$$\delta(x) \equiv lim_{\ \epsilon \to 0} \ \delta_\epsilon(x).$$

Notice that the following is well defined

$$\int_{-\infty}^{\infty} \delta(x) = 1$$

since the integral is 1 for any $\epsilon > 0$ and hence the integral is 1 in the limit too.

A function $f(x)$ is again called an impulse response function, since

$$f(x) * \delta(x) = \int_{-\infty}^{\infty} f(u)\delta(x-u) \ du \ = \ lim \ _{\epsilon \to 0} \int_{-\infty}^{\infty} f(u)\delta_\epsilon(x)du = f(x).$$

Make sure you understand what is happening in that last equation. The delta function $\delta_\epsilon(x)$ only "sees" an epsilon neighborhood of $f(x)$, and the integral computes the local average of values of $f(x)$ in that ϵ neighorhood. If the impulse response function $f(x)$ is continuous near x, then as ϵ shrinks to zero, the result is the value exactly at $f(x)$.

Let's now return to the issues of "edges". Consider an *unit step* function,

$$u(x) = \begin{cases} 1, & x > 0 \\ 0, & x \leq 0. \end{cases}$$

This function is not continuous, but we can write it as the limit of a continuous function, as $\epsilon \to 0$:

$$u_\epsilon(x) = \begin{cases} 1, & x > \frac{\epsilon}{2} \\ \frac{1}{2} + \frac{x}{\epsilon}, & |x| \leq \frac{\epsilon}{2} \\ 0, & \text{otherwise} \end{cases}$$

If we take the derivative of $u_\epsilon(x)$ (and ignore the two values of x where the derivative is not defined) and take the limit, we see that

$$\delta(x) = \frac{d}{dx}u(x)$$

that is, a delta function is equivalent to the derivative of a step edge.

One final point: What is the relationship between the derivative operation $\frac{d}{dx}$ and the discrete derivative D that we saw at the beginning of the lecture? Recall from Calculus,

$$\frac{d}{dx}g(x) = lim_{\epsilon \to 0}\frac{g(x+\epsilon) - g(x-\epsilon)}{2\epsilon}$$

So if we define a function

$$f_\epsilon(x) = \frac{1}{2\epsilon}\delta(x+\epsilon) - \frac{1}{2\epsilon}\delta(x-\epsilon)$$

then, for any $g(x)$, we have

$$\frac{dg(x)}{dx} = lim_{\epsilon \to 0}f_\epsilon(x) * g(x).$$

Setting $\epsilon = 1$ we get the discrete derivative D (see page 1).