In this lecture, I state and prove an important result called the Convolution Theorem which relates the convolution operator and the Fourier transform. To prove this theorem, I’ll need to define convolution slightly differently. The basic issue is that some of the functions we care about, such as the local difference $D(x)$ and local average $B(x)$, are defined on values of $x$ outside of $0, 1, \ldots, N-1$, whereas the Fourier transform of a function requires a summation over the range $0, 1, \ldots, N-1$.

**Re-defining functions on the integers mod $N$**

In order to re-define these functions on $0, 1, \ldots, N-1$ so that we can apply the definition of the Fourier transform, we treat the functions as if they were defined on the integers “mod $N$”:

$$x = -1 \mod N = N - 1,$$

and so

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

Its Fourier transform is:

$$\hat{D}(k) = \sum_{x=0}^{N-1} D(x)e^{-i \frac{2\pi}{N} k x}$$

$$= \frac{1}{2}(-e^{-i \frac{2\pi}{N} k} + e^{-i \frac{2\pi}{N} (N-1)})$$

$$= \frac{1}{2}(-e^{-i \frac{2\pi}{N} k} + e^{i \frac{2\pi}{N} k})$$

$$= \frac{1}{2}(- \cos(\frac{2\pi}{N} k) - i \sin(\frac{2\pi}{N} k)) + (\cos(\frac{2\pi}{N} k) + i \sin(-\frac{2\pi}{N} k))$$

$$= i \sin(\frac{2\pi}{N} k)$$

$$= e^{i \frac{\pi}{2}} \sin(\frac{2\pi}{N} k)$$

Notice that $\hat{D}(k)$ is purely imaginary.

Similarly,

$$B(x) = \begin{cases} 
\frac{1}{2}, & x = 0 \\
\frac{1}{4}, & x = N - 1 \\
\frac{1}{4}, & x = 1 \\
0, & \text{otherwise}
\end{cases}$$
Taking its Fourier transform,
\[
F B(x) = \frac{1}{2} + \frac{1}{4} (e^{-i \frac{2\pi}{N} k} + e^{-i \frac{2\pi}{N} k (N-1)})
\]
\[
= \frac{1}{2} + \frac{1}{4} (e^{-i \frac{2\pi}{N} k} + e^{i \frac{2\pi}{N} k}),
\]
\[
= \frac{1}{2} (1 + \cos(\frac{2\pi}{N} k))
\]

Notice that \( \hat{B}(k) \) is real, i.e. it has no imaginary component, and indeed it is non-negative.

Circular convolution

Let \( I(x) \) and \( h(x) \) be two functions defined on the integers modulo \( N \). The circular convolution of \( I \) and \( h \) is defined:
\[
h(x) \ast I(x) \equiv \sum_{x' = 0}^{N-1} h(x') I((x - x') \mod N).
\]

Note that we don’t need to bother writing \( \mod N \) for the \( x' \) argument of \( h(\cdot) \) since it is automatically in \( 0, 1, \ldots, N - 1 \).

You can check for yourself that circular convolution obeys the commutative, associative and distributive laws, just as our earlier definition of convolution did.

Circular convolution handles the boundaries of a function differently from the previous definition of convolution. For example, suppose we take a local derivative. How is this operation defined at the boundary of the image? There is no way to look at the value of left neighbor of \( x = 0 \) or the right neighbor of \( x = N - 1 \) since these values just do not exist. Previously, we padded the image \( I(x) \) with zeros outside the range \( 0 \) to \( N - 1 \). Circular convolution does something different, namely it treats the underlying image as if the left neighbor of \( x = 0 \) were \( x = N - 1 \), and it treats the right neighbor of \( x = N - 1 \) as if it were \( x = 0 \).

Of course, the values of \( I \ast h(x) \) are not meaningful at the boundary, since the left neighbor of \( x = 0 \) is undefined, etc. One needs to keep this in mind when interpreting the computed values of the convolution. e.g. You don’t want to talk about the derivative at \( x = 0 \) when it has been computed using circular convolution. But its fine to talk about the derivative at \( x = 1 \) since it depends on the values of \( x = 0 \) and \( x = 2 \). This was true when we padded with zeros as well, namely the values of \( I(x) \ast h(x) \) were also junk at the image borders.

Convolution Theorem

We are now ready to prove the Convolution Theorem, namely, for any two functions \( I(x) \) and \( h(x) \) defined on the integers mod \( N \),
\[
F(I(x) \ast h(x))) = FI(x) \cdot Fh(x).
\]

The proof of this theorem only works for circular convolution.
Proof of convolution theorem:

\[ F \, I \ast h(x) = \sum_{x=0}^{N-1} e^{-i \frac{2\pi}{N} k x} \sum_{x'=0}^{N-1} I(x - x') h(x'), \text{ by definition} \]

\[ = \sum_{x'=0}^{N-1} h(x') \sum_{x=0}^{N-1} e^{-i \frac{2\pi}{N} k x} I(x - x'), \text{ by switching order of sums} \]

\[ = \sum_{x'=0}^{N-1} h(x') \sum_{u=0}^{N-1} e^{-i \frac{2\pi}{N} k(u + x')} I(u), \text{ where } u = x - x' \]

\[ = \sum_{x'=0}^{N-1} h(x') e^{-i \frac{2\pi}{N} k x'} \sum_{u=0}^{N-1} e^{-i \frac{2\pi}{N} k u} I(u) \]

\[ = \hat{h}(k) \hat{I}(k) \]

Example (impulse function)
Recall the impulse function \( \delta(x) \) and its property that \( I \ast \delta(x) = I(x) \), and \( \hat{\delta}(k) = 1 \) for all \( k \). These relationships make sense in terms of the convolution theorem, i.e.

\[ F \, I \ast \delta(x) = \hat{I}(k) \hat{\delta}(k) = \hat{I}(k). \]

Example (blur)
Notice that \( \hat{B}(k) = 0 \) when \( k = \frac{N}{2} \). How can we interpret this? Consider a cosine function with frequency of \( k = \frac{N}{2} \). Its Fourier transform is \( N \delta(k - \frac{N}{2}) \). (Exercise.) According to the convolution theorem, convolving the image with \( B(x) \) would remove this frequency component. What is the intuition here? Such a frequency component is of the form \( (\ldots, -1, 1, -1, 1, -1, \ldots) \), i.e. \( k = \frac{N}{2} \) corresponds to a wavelength (or “period”) of two pixels. By inspection, convolving such a (sampled) cosine with \( B(x) \) produces 0 everywhere.

Also notice that \( B(k) = 1 \) when \( k = 0 \). The frequency component for the case \( k = 0 \) is a constant function i.e. \( \cos(0) = 1 \), which is purely real. For any image \( I(x) \) which we blur to get \( I(x) \ast B(x) \), the \( k = 0 \) component of this image doesn’t change, that is, \( \hat{I}(0) \hat{B}(0) = \hat{I}(0) \cdot 1 = \hat{I}(0) \). This makes sense, since \( \hat{I}(0) \) is the sum of the values of \( I(x) \) over all \( x \) and this sum won’t change if we take a local average of the intensities.

What if we convolve an image \( m \) times with \( B(x) \)? Check for yourself, for example, that for \( m = 2 \),

\[ B(x) \ast B(x) = \begin{cases} \frac{1}{16}, & x \pm 2 \\ \frac{1}{4}, & x = \pm 1 \\ \frac{3}{8}, & x = 0 \\ 0, & \text{otherwise} \end{cases} \]

We can easily write this function on the integers \( \text{mod } N \) instead.

From the convolution theorem, for the case of general \( m \), we have

\[ F I(x) \ast B(x) \ast \cdots \ast B(x) = \hat{I}(k) \hat{B}(k)^m \]

\[ = \left( \frac{1}{2} \left( 1 + \cos \left( \frac{2\pi k}{N} \right) \right) \right)^m \]
The case of $m = 1, 2$ are sketched in the slides. As $m$ is increased, more and more frequencies get attenuated.

See the Matlab code:

$$\text{http://www.cim.mcgill.ca/~langer/646/MATLAB/plotB_1D.m}$$

which generates the figure

$$\text{http://www.cim.mcgill.ca/~langer/646/MATLAB/plotB_1D.jpg}$$

which shows $B(x) \ast \cdots \ast B(x)$ and its Fourier transform, for various $m$.

**Filtering and bandwidth**

Suppose we convolve an image $I(x)$ with a function $h(x)$. We have refered to $h(x)$ as an impulse response function, and we did so to emphasize what is the contribution of the value $I(x)$ at each pixel $x$. Another way to refer to $h(x)$ is as a linear filter. This term is used to describe how $h(x)$ attenuates, amplifies, or shifts each frequency component, as dictated by the convolution theorem. The Fourier transform of the filter $h(x)$ can be written

$$\hat{h}(k) = |\hat{h}(k)| e^{i\phi(k)}$$

where $|\hat{h}(k)|$ is usually called the amplitude spectrum and $\phi(k)$ is called the phase spectrum. By the convolution theorem,

$$\mathbf{F}I(x) = \mathbf{F}(I(x) \ast h(x)) = \hat{I}(k) |\hat{h}(k)| e^{i\phi(k)}$$

and so the amplitude $|\hat{h}(k)|$ attenuates/amplifies frequency component $\hat{I}(k)$, and the phase $\phi(k)$ of $h(\cdot)$ shifts each frequency component of $I(x)$.

We can characterize filters by which frequency components they preserve/amplify/attenuate. Here we are concerned only with the amplitude spectrum. Let’s first address the case of “ideal” filters.

- $h(x)$ is a low pass filter if there exists a frequency $k_0$ such that

$$\hat{h}(k) = \begin{cases} 
1, & 0 \leq k \leq k_0 \\
0, & k_0 < k \leq \frac{N}{2}
\end{cases}$$

- $h(x)$ is a high pass filter if there exists $k_0$ such that

$$\hat{h}(k) = \begin{cases} 
0, & 0 \leq k < k_0 \\
1, & k_0 \leq k \leq \frac{N}{2}
\end{cases}$$

- $h(x)$ is a bandpass filter if there exists two frequencies $k_0$ and $k_1$ such that

$$\hat{h}(k) = \begin{cases} 
0, & 0 \leq k < k_0 \\
1, & k_0 \leq k \leq k_1 \\
0, & k_1 < k \leq \frac{N}{2}
\end{cases}$$
Note that these definitions above only concern $k \in \{0, \ldots, \frac{N}{2}\}$. Frequencies above $k = \frac{N}{2}$ are ignored in the definition because the values of $\hat{h}(k)$ of these frequencies are determined by the conjugacy property, which determined high frequencies. (Recall that the conjugacy properties applies only if the filter is real valued function.)

We typically work with filters that are not ideal i.e. filters that only approximately satisfy the above definitions. For example, the blurring function $B(x)$ is approximately low pass. The local differencing filter $D(x)$ was approximately high pass.

If we have an approximately bandpass filter, then we would like to describe the width of this filter i.e. the range of frequencies that it lets through. One often does this by considering the frequencies at which $|\hat{h}(k)|$ reaches half its maximum value. The bandwidth at half-height is defined to be $k_1 - k_0$, where $k_0 < k_1$ and

$$|\hat{h}(k_0)| = |\hat{h}(k_1)| = \frac{1}{2} \max_{k\in[0,\frac{N}{2}]} |\hat{h}(k)|$$

ASIDE (not mentioned in class): Alternatively, we can describe the bandwidth as (the log of) the ratio of $k_1$ to $k_0$. The octave bandwidth at half height is:

$$\log_2\left(\frac{k_1}{k_0}\right) = \log_2(k_1) - \log_2(k_0)$$

For example, a filter with a bandwidth of one octave means that the $k_1$ frequency is twice the $k_0$ frequency.

**Periodicity property of the Fourier transform**

The definition of low, high, bandpass only concern $k$ in $0, \ldots, \frac{N}{2}$. If we are only concerned with the amplitude spectrum, then these frequencies are enough. But if we are also concerned with the phase, then we might want to keep all of the frequencies $k \in 0, \ldots, N - 1$. Alternatively, we sometimes make our plots for $k \in -\frac{N}{2}, \ldots, 0, \ldots, \frac{N}{2} - 1$.

The reason this is possible is that we can regard $\hat{I}(k)$ as being well-defined for any integer $k$. In particular, it is periodic in $k$ with period $N$,

$$\hat{I}(k) = \hat{I}(k + mN)$$

since, for any integer $m$,

$$e^{i\frac{2\pi}{N}mN} = 1$$
and so
\[ e^{\frac{2\pi}{N} kx} = e^{\frac{2\pi}{N} (k + mN)x}. \]
Therefore, if we use \( k + mN \) instead of \( k \) in the definition of the Fourier transform, we get the same result.

Note that this "periodicity property" is quite different from the conjugacy properties. First, the conjugacy property relates the Fourier coefficients of two frequencies \( k \) and \( N - k \), which lie symmetrically on opposite sides of \( k = N/2 \), whereas the periodicity property relates two frequencies separated by a multiple of \( N \). Second, the conjugacy property requires that the function \( I(x) \) whose Fourier transform we are considering is real valued. The periodicity property holds regardless of the function is real or not.

**Fourier transform of a Gaussian**

A very good approximation for the Fourier transform of a Gaussian \( G(x, \sigma) \) is:

\[
\hat{G}(k, \sigma) = F \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}} \approx e^{-\frac{1}{2} \left( \frac{2\pi k}{N} \right)^2 \sigma^2}
\]

Interestingly, this approximation becomes exact in the limit as \( N, \sigma \to \infty \), with \( \frac{\sigma}{N} \) held constant. (The proof of that statement is beyond the scope of this course.)

If you wish to see this approximation for yourself, run the Matlab script

http://www.cim.mcgill.ca/~langer/646/MATLAB/plotFourierTransformGaussian.m

which generates the figure

http://www.cim.mcgill.ca/~langer/646/MATLAB/plotFourierTransformGaussian.jpg

Two key properties to notice are:

- if the standard deviation of the Gaussian in the space \((x)\) domain is \( \sigma \) then the standard deviation of the Gaussian in the frequency \((k)\) domain is proportional to \( \frac{1}{\sigma} \)

- while \( \hat{G}(k, \sigma) \) has a Gaussian shape, it does not integrate to 1, namely there is no scaling factor present. The max value occurs at \( k = 0 \) and the max value is always 1.

In class, I briefly discussed the implications of these properties for the DOG function which is a (non-ideal) bandpass filter.