Lecture 15

Fourier Transforms (cont'd)

Here we list some of the more important properties of Fourier transforms. You have probably seen many of these, so not all proofs will not be presented. (That being said, most proofs are quite straightforward and you are encouraged to try them.) You may find derivations of all of these properties in the book by Boggess and Narcowich, Section 2.2, starting on p. 98.

Let us first recall the definitions of the Fourier transform (FT) and inverse FT (IFT) that will be used in this course. Given a function $f(t)$, its Fourier transform $F(\omega)$ is defined as

$$F(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(t) e^{-i\omega t} dt. \tag{1}$$

And given a Fourier transform $F(\omega)$, its inverse FT is given by

$$f(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} F(\omega) e^{i\omega t} d\omega. \tag{2}$$

In what follows, we assume that the functions $f(t)$ and $g(t)$ are differentiable as often as necessary, and that all necessary integrals exist. (This implies that $f(t) \to 0$ as $|t| \to \infty$.) Regarding notation: $f^{(n)}(t)$ denotes the nth derivative of f w.r.t. t; $F^{(n)}(\omega)$ denotes the nth derivative of F w.r.t. ω.

1. **Linearity of the FT operator and the inverse FT operator:**

$$\begin{aligned} \mathcal{F}(f+g) &= \mathcal{F}(f) + \mathcal{F}(g) \\ \mathcal{F}(cf) &= c\,\mathcal{F}(f), \quad c \in \mathbf{C} \text{ (or } \mathbf{R}), \end{aligned} \tag{3}$$

$$\begin{aligned} \mathcal{F}^{-1}(f+g) &= \mathcal{F}^{-1}(f) + \mathcal{F}^{-1}(g) \\ \mathcal{F}^{-1}(cf) &= c\,\mathcal{F}^{-1}(f), \quad c \in \mathbf{C} \text{ (or } \mathbf{R}), \end{aligned} \tag{4}$$

These follow naturally from the integral definition of the FT.

2. **Fourier transform of a product of f with t^n:**

$$\mathcal{F}(t^n f(t)) = i^n F^{(n)}(\omega). \tag{5}$$

We consider the case $n = 1$ by taking the derivative of $F(\omega)$ in Eq. (1):

$$F'(\omega) = \frac{d}{d\omega}\left[\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(t) e^{-i\omega t} dt\right]$$

$$
\begin{aligned}
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(t) \frac{d}{d\omega}\left[e^{-i\omega t}\right] dt \qquad \text{(Leibniz Rule)} \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(t)(-it) e^{-i\omega t} dt \\
&= -i \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t f(t) e^{-i\omega t} dt \\
&= -i\mathcal{F}(tf(t)). \qquad (6)
\end{aligned}
$$

Repeated applications of the differentiation operator produces the result,

$$F^{(n)}(\omega) = (-i)^n \mathcal{F}(t^n f(t)), \qquad (7)$$

from which the desired property follows.

3. **Inverse Fourier transform of a product of F with ω^n:**

$$\mathcal{F}^{-1}(\omega^n F(\omega)) = (-i)^n f^{(n)}(t). \qquad (8)$$

Here, we start with the definition of the inverse FT in Eq. (2) and differentiate both sides repeatedly with respect to t. The reader will note a kind of reciprocity between this result and the previous one.

4. **Fourier transform of an nth derivative:**

$$\mathcal{F}(f^{(n)}(t)) = (i\omega)^n F(\omega). \qquad (9)$$

This is a consequence of 3. above.

5. **Inverse Fourier transform of an nth derivative:**

$$\mathcal{F}^{-1}(F^{(n)}(\omega)) = (-it)^n f(t). \qquad (10)$$

This is a consequence of 2. above.

6. **Fourier transform of a translation ("Spatial- or Temporal-shift Theorem"):**

$$\mathcal{F}(f(t-a)) = e^{-i\omega a} F(\omega). \qquad (11)$$

Because this result is very important, we provide a proof, even though it is very simple:

$$\mathcal{F}(f(t-a)) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(t-a) e^{-i\omega t} dt$$

$$
\begin{aligned}
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(s) e^{-i\omega(s+a)} ds \qquad (s = t - a, \quad dt = ds, \ etc.) \\
&= e^{-i\omega a} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(s) e^{-i\omega s} ds \\
&= e^{-i\omega a} F(\omega).
\end{aligned} \tag{12}
$$

7. **Fourier transform of a "modulation" ("Frequency-shift Theorem":**

$$
\mathcal{F}(e^{i\omega_0 t} f(t)) = F(\omega - \omega_0). \tag{13}
$$

The proof is very simple:

$$
\begin{aligned}
\mathcal{F}(e^{i\omega_0 t} f(t)) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{i\omega_0 t} f(t) e^{-i\omega t} dt \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(t) e^{-i(\omega - \omega_0)t} dt \\
&= F(\omega - \omega_0).
\end{aligned} \tag{14}
$$

We'll return to this important result to discuss it in more detail.

8. **Fourier transform of a scaled function ("Scaling Theorem"):** For a $b > 0$,

$$
\mathcal{F}(f(bt)) = \frac{1}{b} F\left(\frac{\omega}{b}\right). \tag{15}
$$

Once again, because of its importance, we provide a proof, which is also quite simple:

$$
\begin{aligned}
\mathcal{F}(f(bt)) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(bt) e^{-i\omega t} dt \\
&= \frac{1}{b} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(s) e^{-i\omega s/b} ds \qquad (s = bt, \quad dt = \frac{1}{b} ds, \ etc.) \\
&= \frac{1}{b} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(s) e^{-i\left(\frac{\omega}{b}\right)s} ds \\
&= \frac{1}{b} F\left(\frac{\omega}{b}\right).
\end{aligned} \tag{16}
$$

We'll also return to this result to discuss it in more detail.

9. **Convolution Theorem:** We denote the *convolution* of two functions f and g as "$f * g$," defined as follows:

$$
\begin{aligned}
(f * g)(t) &= \int_{-\infty}^{\infty} f(t - s) g(s) \, ds \\
&= \int_{-\infty}^{\infty} f(s) g(t - s) \, ds.
\end{aligned} \tag{17}
$$

This is a continuous version of the convolution operation introduced for the discrete Fourier transform. In special cases, it may be viewed as a local averaging procedure, as we'll show below.

Theorem: If $h(t) = (f * g)(t)$, then the FT of H is given by

$$H(\omega) = \sqrt{2\pi} F(\omega) G(\omega). \tag{18}$$

Another way of writing this is

$$\mathcal{F}(f * g) = \sqrt{2\pi} FG. \tag{19}$$

By taking inverse FTs of both sides and moving the factor $\sqrt{2\pi}$,

$$\mathcal{F}^{-1}(FG) = \frac{1}{\sqrt{2\pi}} f * g. \tag{20}$$

Proof: By definition, the Fourier transform of h is given by

$$\begin{aligned} H(\omega) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(t-s) g(s)\, ds\, e^{-i\omega t}\, dt \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(t-s) g(s)\, ds\, e^{-i\omega(t-s)} e^{-i\omega s}\, dt. \end{aligned} \tag{21}$$

Noting the appearance of $t-s$ and s in the exponentials and in the functions f and g, respectively, we make the following change of variables,

$$u(s,t) = s, \qquad v(s,t) = t - s. \tag{22}$$

The Jacobian of this transformation is

$$J = \begin{vmatrix} \partial u/\partial s & \partial u/\partial t \\ \partial v/\partial s & \partial v/\partial t \end{vmatrix} = \begin{vmatrix} 1 & 0 \\ -1 & 1 \end{vmatrix} = 1. \tag{23}$$

Therefore, we have $ds\, dt = du\, dv$. Furthermore, the above double integral is separable and may be written as

$$\begin{aligned} H(\omega) &= \frac{1}{\sqrt{2\pi}} \left(\int_{-\infty}^{\infty} f(v) e^{-i\omega v}\, dv \right) \left(\int_{-\infty}^{\infty} f(u) e^{-i\omega u}\, du \right) \\ &= \sqrt{2\pi}\, \frac{1}{\sqrt{2\pi}} \left(\int_{-\infty}^{\infty} f(v) e^{-i\omega v}\, dv \right) \frac{1}{\sqrt{2\pi}} \left(\int_{-\infty}^{\infty} f(u) e^{-i\omega u}\, du \right) \\ &= \sqrt{2\pi}\, F(\omega) G(\omega), \end{aligned} \tag{24}$$

which completes the derivation.

A more detailed look at some properties of Fourier transforms

The "Frequency-shift Theorem"

There are two interesting consequences of the Frequency-shift Theorem,

$$\mathcal{F}(e^{i\omega_0 t} f(t)) = F(\omega - \omega_0). \tag{25}$$

We may be interested in computing the FT of the product of either $\cos\omega_0 t$ or $\sin\omega_0 t$ with a function $f(t)$, which are known as *(amplitude) modulations* of $f(t)$. In this case, we express these trigonometric functions in terms of appropriate complex exponentials and then employ the above Shift Theorem.

First of all, we start with the relations

$$\begin{aligned} \cos\omega_0 t &= \frac{1}{2}\left[e^{\omega_0 t} + e^{-\omega_0 t}\right] \\ \sin\omega_0 t &= \frac{1}{2i}\left[e^{\omega_0 t} - e^{-\omega_0 t}\right]. \end{aligned} \tag{26}$$

From these results and the Frequency Shift Theorem, we have

$$\begin{aligned} \mathcal{F}(\cos\omega_0 t f(t)) &= \frac{1}{2}\left[F(\omega - \omega_0) + F(\omega + \omega_0)\right] \\ \mathcal{F}(\sin\omega_0 t f(t)) &= \frac{1}{2i}\left[F(\omega - \omega_0) - F(\omega + \omega_0)\right], \end{aligned} \tag{27}$$

where $F(\omega)$ denotes the FT of $f(t)$.

To show how these results may be helpful in the computation of FTs, let us revisit Example 2 on Page 6, namely the computation of the FT of the function $f(t)$ defined as the function $\cos 3t$, but restricted to the interval $[-\pi, \pi]$. We may view $f(t)$ as a product of two functions, i.e.,

$$f(t) = \cos 3t \cdot b(t), \tag{28}$$

where $b(t)$ denotes the boxcar function of Example 1. From the Frequency Shift Theorem,

$$F(\omega) = \mathcal{F}(f(t)) = \frac{1}{2}\left[B(\omega - 3) + B(\omega + 3)\right], \tag{29}$$

where

$$B(\omega) = \sqrt{2\pi}\,\mathrm{sinc}(\pi\omega) = \sqrt{2\pi}\,\frac{\sin(\pi\omega)}{\pi\omega} = \sqrt{\frac{2}{\pi}}\,\frac{\sin(\pi\omega)}{\omega} \tag{30}$$

is the FT of the boxcar function $b(t)$. Substitution into the preceeding equation yields

$$\begin{aligned} F(\omega) &= \frac{1}{\sqrt{2\pi}}\left[\frac{\sin(\pi(\omega-3))}{\omega-3} - \frac{\sin(\pi(\omega+3))}{\omega+3}\right] \\ &= \frac{1}{\sqrt{2\pi}}(-2\omega)\,\frac{\sin(\pi\omega)}{\omega^2-9} \\ &= \sqrt{\frac{2}{\pi}}\,\frac{\omega\sin(\pi\omega)}{9-\omega^2}, \end{aligned} \tag{31}$$

in agreement with the result obtained for Example 2 earlier.

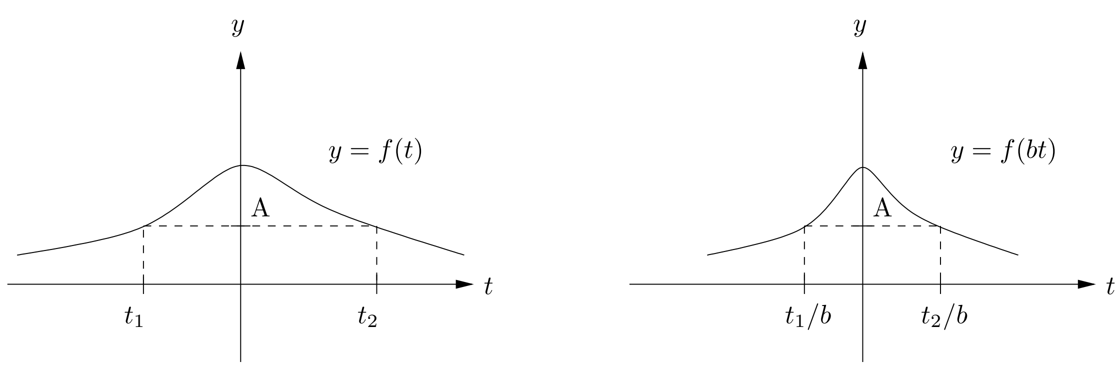
That being said, it is probably easier now to understand the plot of the Fourier transform that was presented with Example 2. Instead of trying to decipher the structure of the plot of the function $\sin(\pi\omega)$ divided by the polynomial $9-\omega^2$, one can more easily visualize a sum of two shifted sinc functions.

The "Scaling Theorem"

Let's try to get a picture of the Scaling Theorem:

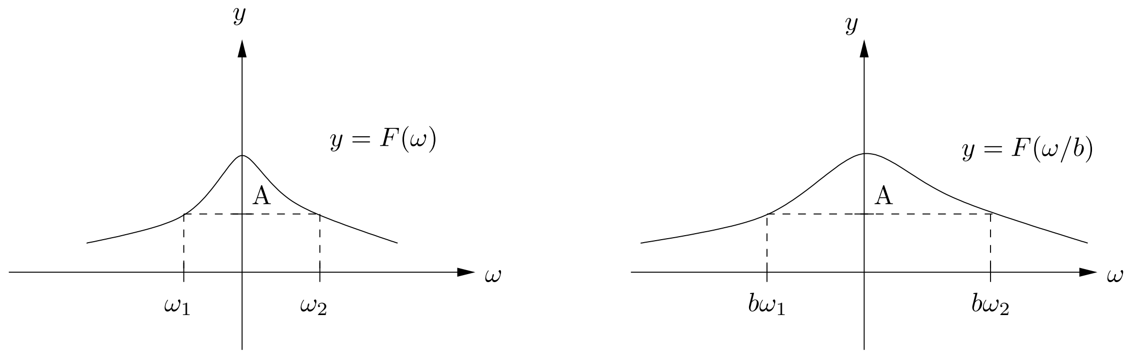
$$\mathcal{F}(f(bt)) = \frac{1}{b}F\left(\frac{\omega}{b}\right). \tag{32}$$

Suppose that $b>1$. Then the graph of the function $g(t)=f(bt)$ is obtained from the graph of $f(t)$ by *contracting* the latter horizontally toward the y-axis by a factor of $1/b$, as sketched in the plots below.



Left: Graph of $f(t)$, with two points t_1 and t_2, where $f(t)=A$. Right: Graph of $f(bt)$ for $b>1$. The two points at which $f(t)=A$ have now been contracted toward the origin and are at t_1/b and t_2/b, respectively.

On the other hand, the graph $G(\omega)=F\left(\frac{\omega}{b}\right)$ is obtained from the graph of $F(\omega)$ by *stretching* it outward away by a factor of b from the y-axis, as sketched in the next set of plots below.



Left: Graph of Fourier transform $F(\omega)$, with two points ω_1 and ω_2, where $F(\omega) = A$. Right: Graph of $F(\omega/b)$ for $b > 0$. The two points at which $F(\omega) = A$ have now been expanded away from the origin and are at $b\omega_1$ and $b\omega_2$, respectively.

The contraction of the graph of $f(t)$ along with an expansion of the graph of $F(\omega)$ is an example of the *complementarity* of time (or space) and frequency domains. We shall return to this point shortly.

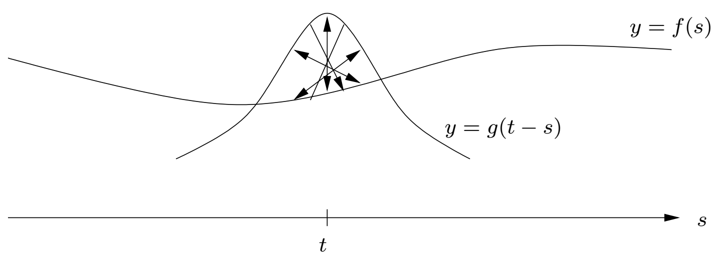
As in the case of discrete Fourier series, the magnitude $|F(\omega)|$ of the Fourier transform of a function must go to zero as $|\omega| \to \infty$. Assume once again that $b > 1$. Suppose most of the energy of a Fourier transform $F(\omega)$ of a function $f(t)$ is situated in the interval $[-\omega_0, \omega_0]$: For ω values outside this interval, $F(\omega)$ is negligible. But the Fourier transform of the function $f(bt)$ is now $F(\omega/b)$, which means that it lies in the interval $[-b\omega_0, b\omega_0]$, which represents an *expansion* of the interval $[-\omega_0, \omega_0]$. This implies that the FT of $f(bt)$ contains higher frequencies than the FT of $f(t)$. Does this make sense?

The answer is, "Yes," because the compression of the graph of $f(t)$ to produce $f(bt)$ will produce gradients of higher magnitude – the function will have greater rates of decrease or increase. As a result, it must have higher frequencies in its FT representation. (We saw this in the case of Fourier series.)

Of course, in the case that $0 < b < 1$, the situation is reversed. The graph of $f(bt)$ will be a *horizontally-stretched* version of the graph of $f(t)$, and the corresponding FT, $F(\omega/b)$ will be a *horizontally-contracted* version of the graph of $F(\omega)$.

The Convolution Theorem

The Convolution Theorem, Eq. (17), is extremely important in signal/image processing, as we shall see shortly. Let us return to the idea that a convolution may represent a local averaging process. If the function $g(t)$ is a Gaussian-type function, with peak at the origin, then the second line of the definition in (17) may be viewed as in the sketch below. For a given value of t, the function $(f * g)(t)$ will be produced by a kind of "reverse inner product" process where values $f(s)$ are multiplied by values $g(t-s)$ and added together. The function $g(t-s)$ peaks at $s=t$, so most of the value of $(f * g)(t)$ comes from $f(t)$, with lesser weights on $f(s)$ as s moves away from t.



A schematic illustration of the contributions to the convolution integral in Eq. (17) to produce a local averaging of f.

There is also a "reverse" Convolution Theorem: If you multiply two functions, i.e., $h(t) = f(t)g(t)$, then the FT of h is a convolution of the FTs of f and G. We'll return to this result shortly.

Lecture 16

Fourier transforms (cont'd)

We continue with a number of important results involving FTs.

Plancherel Formula

Assume that $f, g \in L^2(\mathbf{R})$. Then

$$\langle f, g\rangle = \langle F, G\rangle, \tag{33}$$

where $\langle \cdot, \cdot \rangle$ denotes the complex inner product on $\mathbf{R}$, i.e.,

$$\int_{-\infty}^{\infty} f(t)\overline{g(t)}\, dt = \int_{-\infty}^{\infty} F(\omega)\overline{G(\omega)}\, d\omega. \tag{34}$$

An important consequence of this result is the following: If $f = g$, then $\langle f, f\rangle = \langle F, F\rangle$, implying that

$$\|f\|_2^2 = \|F\|_2^2, \quad \text{or} \quad \|f\|_2 = \|F\|_2. \tag{35}$$

This result may also be written as follows,

$$\|f\|_2 = \|\mathcal{F}(f)\|_2. \tag{36}$$

In other words, the symmetric form of the Fourier transform operator is *norm-preserving*: The L^2 norm of f is equal to the norm of $F = \mathcal{F}(f)$. (For the unsymmetric forms of the Fourier transform, we would have to include a multiplicative factor involving 2π or its square root.)

Eq. (35) may be viewed as the continuous version of *Parseval's equation* studied earlier in the context of complete orthonormal bases. Recall the following: Let H be a Hilbert space. If the orthonormal set $\{e_k\}_0^\infty \subset H$ is complete, then for any $f \in H$,

$$f = \sum_{k=0}^{\infty} c_k e_k, \quad \text{where} \;\; c_k = \langle f, e_k\rangle, \tag{37}$$

and

$$\|f\|_{L^2} = \|c\|_{l^2}, \tag{38}$$

i.e.,

$$\|f\|^2 = \sum_{k=0}^{\infty} |c_k|^2 \qquad \text{(Parseval's equation)}. \tag{39}$$

On the other hand, Parseval's equation may be viewed as a discrete version of Plancherel's formula.

Proof of Plancherel's Formula

We first express the function $f(t)$ in terms of the inverse Fourier transform, i.e.,

$$f(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} F(\omega) e^{i\omega t} d\omega. \tag{40}$$

Now substitute for $F(\omega)$ using the definition of the Fourier transform:

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(s) e^{-i\omega s} ds \; e^{i\omega t} d\omega. \tag{41}$$

Now take the inner product of $f(t)$ with $g(t)$:

$$\langle f, g \rangle = \int_{-\infty}^{\infty} \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x) e^{-i\omega s} ds \; e^{i\omega t} d\omega \; \overline{g(t)} \; dt. \tag{42}$$

We now assume that f and g are sufficiently "nice" so that Fubini's Theorem will allow us to rearrange the order of integration. The result is

$$\begin{aligned} \langle f, g \rangle &= \int_{-\infty}^{\infty} \left(\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(s) e^{-i\omega s} ds \right) \left(\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \overline{g(t)} e^{i\omega t} dt \right) d\omega \\ &= \int_{-\infty}^{\infty} F(\omega) \overline{G(\omega)} \; d\omega, \\ &= \langle F, G \rangle. \end{aligned} \tag{43}$$

The Fourier transform of a Gaussian (in t-space) is a Gaussian (in ω-space)

This is a fundamental result in Fourier analysis. To show it, consider the following function,

$$f(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{t^2}{2\sigma^2}} \; . \tag{44}$$

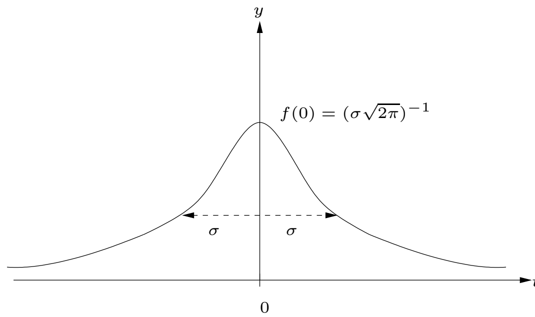
You might recognize this function: It is the *normalized Gaussian distribution function*, or simply the *normal distribution* with zero-mean and standard deviation σ or variance σ^2. The factor in front of the integral normalizes it, since

$$\int_{-\infty}^{\infty} f(t) \; dt = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{t^2}{2\sigma^2}} \; dt = 1. \tag{45}$$

(The above result implies that the L^1 norm of f is 1, i.e., $\|f\|_1 = 1$.)

Just in case you are not fully comfortable with this result, we mention that all of these results come from the following fundamental integral:

$$\int_{-\infty}^{\infty} e^{-x^2} \; dx = \sqrt{\pi}. \tag{46}$$



Sketch of normalized Gaussian function $f(t) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{t^2}{2\sigma^2}}$.

From this result, we find that

$$\int_{-\infty}^{\infty} e^{-A^2x^2}\,dx = \frac{1}{A}\sqrt{\pi}, \tag{47}$$

by means of the change of variable $y = Ax$.

This leads to another important point. The above Gaussian function is said to be *normalized*, but in the L^1-sense and not in the L^2 sense. You may wish to confirm that, for general σ,

$$\langle f, f\rangle = \|f\|_2^2 \neq 1. \tag{48}$$

For simplicity, let us compute the FT of the function

$$g(t) = e^{-\frac{t^2}{2\sigma^2}}\ dt. \tag{49}$$

Our desired FT for $f(t)$ will then be given by

$$F(\omega) = \frac{1}{\sigma\sqrt{\pi}}G(\omega). \tag{50}$$

From the definition of the FT,

$$\begin{aligned} G(\omega) &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{t^2}{2\sigma^2}}e^{-i\omega t}\,dt \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{t^2}{2\sigma^2}}[\cos(\omega t) - i\sin(\omega t)]\,dt \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{t^2}{2\sigma^2}}\cos(\omega t)\,dt. \end{aligned} \tag{51}$$

The $\sin(\omega t)$ term will not contribute to the integration because it is an odd function: Its product with the even Gaussian function is an odd function which, when integrated over $(-\infty, \infty)$, yields a zero result.

The resulting integral in (51) is not straightforward to evaluate because the antiderivative of the Gaussian does not exist in closed form. But we may perform a trick here. Let us differentiate both sides of the equation with respect to the parameter ω. The differentiation of the integrand is permitted by "Leibniz' Rule", so that

$$G'(\omega) = -\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t e^{-\frac{t^2}{2\sigma^2}} \sin(\omega t)\, dt. \tag{52}$$

The appearance of the "t" in the integrand now permits an antidifferentiation using integration by parts: Let $u = \sin(\omega t)$ and $dv = -t\exp(-t^2/(2\sigma^2))$, so that $v = \sigma^2 \exp(-t^2/(2\sigma^2))$ and $du = \omega\cos(\omega t)$. Then

$$G'(\omega) = \frac{1}{\sqrt{2\pi}} \left[\sin(\omega t) e^{-\frac{t^2}{2\sigma^2}} \right]_{-\infty}^{\infty} - \sigma^2 \omega \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{t^2}{2\sigma^2}} \cos(\omega t)\, dt. \tag{53}$$

The first term is zero, because the Gaussian function $e^{-\frac{t^2}{2\sigma^2}} \to 0$ as $t \to \pm\infty$. And the integral on the right, which once again involves the $\cos(\omega t)$ function, along with the $1/\sqrt{2\pi}$ factor, is our original $G(\omega)$ function. Therefore, we have derived the result,

$$G'(\omega) = -\sigma^2 \omega G(\omega). \tag{54}$$

This is a first order DE in the function $G(\omega)$. It is also a separable DE and may be easily solved to yield the general solution (details left for reader):

$$G(\omega) = D e^{-\frac{1}{2}\sigma^2\omega^2}, \tag{55}$$

where D is the arbitrary constant. From the fact that

$$G(0) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{t^2}{2\sigma^2}}\, dt = \frac{1}{\sqrt{2\pi}} \sqrt{2\pi}\sigma = \sigma, \tag{56}$$

we have that $D = \sigma$, implying that

$$G(\omega) = \sigma e^{-\frac{1}{2}\sigma^2\omega^2}. \tag{57}$$

From (50), we arrive at our goal, $F(\omega)$, the Fourier transform of the Gaussian function

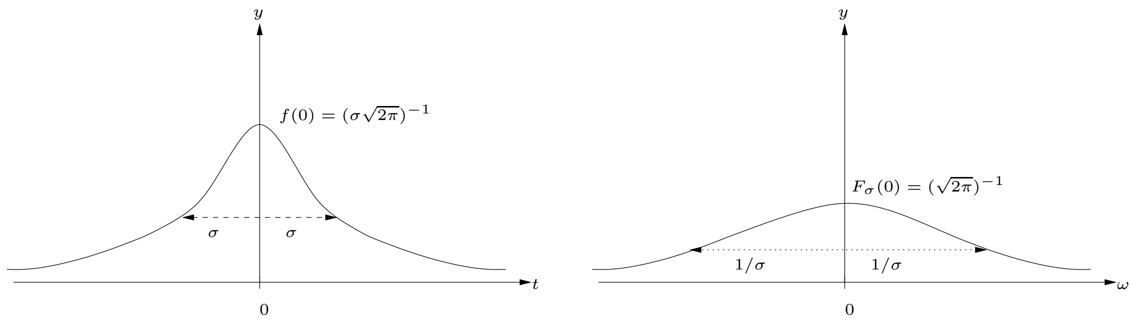
$$f(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{t^2}{2\sigma^2}} \tag{58}$$

is

$$F(\omega) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\sigma^2\omega^2}. \tag{59}$$

As the title of this section indicated, the Fourier transform $F(\omega)$ is a Gaussian in the variable ω.

There is one fundamental difference, however, between the two Gaussians, $F(\omega)$ and $f(t)$. The standard deviation of $f(t)$ is σ. But the standard deviation of $F(\omega)$, i.e., the value of ω at which $F(\omega)$ assumes the value $e^{-1/2}F(0)$, is σ^{-1}. In other words, if σ is small, so that the Gaussian $f(t)$ is a thin peak, then $F(\omega)$ is broad. This relationship, which is a consequence of the *complementarity* of the time (or space) and frequency domains, is sketched below.



Generic sketch of normalized Gaussian function $f_\sigma(t) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{t^2}{2\sigma^2}}$, with standard deviation σ (left), and its Fourier transform $F_\sigma(\omega) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\sigma^2\omega^2}$ with standard deviation $1/\sigma$ (right).

Of course, the reverse situation also holds: If σ is large, so that $f_\sigma(t)$ is broad, then $1/\sigma$ is small, so that $F_\sigma(\omega)$ is thin and peaked. These are but special examples of the "Uncertainty Principle" that we shall examine in more detail a little later.

The limit $\sigma \to 0$ and the Dirac delta function

Let's return to the Gaussian function originally defined in Eq. (44),

$$f(t) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{t^2}{2\sigma^2}} \ . \tag{60}$$

With a little work, it can be shown that in the limit $\sigma \to 0$,

$$f_\sigma(t) \to 0 \quad t \neq 0, \qquad f_\sigma(0) \to \infty. \tag{61}$$

This has the aroma of the "Dirac delta function," denoted as $\delta(t)$. As you may know, the Dirac delta function is not a function, but rather a *generalized function* or *distribution* which must be understood in the context of integration with respect to an appropriate class of functions. Recall that for any continuous function $g(t)$ on $\mathbf{R}$, and an $a \in \mathbf{R}$

$$\int_{-\infty}^{\infty} g(t)\delta(t)\, dt = g(0). \tag{62}$$

It can indeed be shown that

$$\lim_{\sigma \to 0} \int_{-\infty}^{\infty} g(t) f_\sigma(t)\, dt = g(0) \tag{63}$$

for any continuous function $g(t)$. As such we may state, rather loosely here, that in the limit $\sigma \to 0$, the Gaussian function $f_\sigma(t)$ becomes the Dirac delta function. (There is a more rigorous treatment of the above equation in terms of "generalized functions" or "distributions".)

Note also that in the limit $\sigma \to 0$, the Fourier transform of $f_\sigma(t)$, i.e., the Gaussian function $G_\sigma(\omega)$, becomes the constant function $1/\sqrt{2\pi}$. This agrees with the formal calculation of the Fourier transform of the Dirac delta function $\delta(t)$:

$$\mathcal{F}[\delta(t)] = \Delta(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \delta(t) e^{-i\omega t}\, dt = \frac{1}{\sqrt{2\pi}}. \tag{64}$$

From a signal processing perspective, this is the worst kind of Fourier transform you could have: all frequencies are present and there is no decay as $\omega \to \pm\infty$. Note that this is *not* in disagreement with previous discussions concerning the decay of Fourier transforms since neither the Dirac delta function, $\delta(t)$, nor its Fourier transform, $\Delta(\omega)$, are $L^2(\mathbf{R})$ functions.

Some final notes regarding Dirac delta functions and their occurrence in Fourier analysis

As mentioned earlier, the "basis" functions $e^{i\omega t}$ employed in the Fourier transform are not $L^2(\mathbf{R})$ functions. Nevertheless, they may be considered to obey an orthogonality property of the form,

$$\langle e^{i\omega_1 t}, e^{i\omega_2 t} \rangle = \int_{-\infty}^{\infty} e^{i(\omega_1 - \omega_2)t}\, dt = \begin{cases} 0, & \omega_1 \neq \omega_2, \\ \infty, & \omega_1 = \omega_2, \end{cases} \tag{65}$$

This has the aroma of a Dirac delta function. But the problem is that the infinity on the RHS is unchanged if we multiply it by a constant. Is the RHS simply the Dirac delta function $\delta(\omega_1 - \omega_2)$? Or is it a constant times this delta function?

To answer this question, we can go back to the Fourier transform result in Eq. (64). The inverse Fourier transform associated with this result is

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \left(\frac{1}{\sqrt{2\pi}} \right) e^{i\omega t}\, d\omega = \delta(t), \tag{66}$$

which implies that

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\omega t}\, d\omega = \delta(t). \tag{67}$$

Note that there is a symmetry here: We could have integrated $e^{i\omega t}$ with respect to t to produce a $\delta(\omega)$ on the right, i.e.,

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{i\omega t}\, dt = \delta(\omega). \tag{68}$$

But we can replace ω above by $\omega_1 - \omega_2$, i.e.,

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{i(\omega_1-\omega_2)t}\, dt = \delta(\omega_1-\omega_2). \tag{69}$$

The left-hand side of the above equation may be viewed as an inner product so that we have the following orthonormality result,

$$\left\langle \frac{1}{\sqrt{2\pi}}\, e^{i\omega_1 t}, \frac{1}{\sqrt{2\pi}}\, e^{i\omega_2 t} \right\rangle = \delta(\omega_1-\omega_2), \quad \omega_1, \omega_2 \in \mathbf{R}. \tag{70}$$

In other words, the functions

$$\frac{1}{\sqrt{2\pi}}\, e^{i\omega t}, \tag{71}$$

with **continuous index** $\omega \in \mathbf{R}$, may be viewed as forming an **orthonormal set** over the real line $\mathbf{R}$ according to the orthonormality relation in Eq. (70).

Note: In Physics, the functions in Eq. (71) are known as **plane waves** with (angular) frequency ω. These functions are **not** square integrable functions on the real line $\mathbf{R}$. In other words, they are **not** elements of $L^2(\mathbf{R})$.

And just to clean up loose ends, Eq. (70) indicates that the orthogonality relation in Eq. (65) should be written as

$$\langle e^{i\omega_1 t}, e^{i\omega_2 t}\rangle = \int_{-\infty}^{\infty} e^{i(\omega_1-\omega_2)t}\, dt = 2\pi\delta(\omega_1-\omega_2), \quad \omega_1, \omega_2 \in \mathbf{R}, \tag{72}$$

Finally, it may seem quite strange – and perhaps a violation of Plancherel's Formula – that in the limit $\sigma \to 0$, the area under the curve of the Fourier transform $F(\omega)$ becomes infinite, whereas the area under the curve $f_\sigma(t)$ appears to remain finite. Recall, however, that Plancherel's Formula relates the L^2 norms of the functions f and g. It is left as an exercise to see what is going on here.

The Convolution Theorem revisited – "Convolution Theorem Version 2"

Recall the Convolution Theorem for Fourier transforms: Given two functions $f(t)$ and $g(t)$ with Fourier transforms $F(\omega)$ and $G(\omega)$, respectively. Then the Fourier transform of their convolution,

$$\begin{aligned} h(t) = (f * g)(t) &= \int_{-\infty} f(t-s)g(s)\,ds \\ &= \int_{-\infty} f(s)g(t-s)\,ds, \end{aligned} \tag{73}$$

is given by

$$H(\omega) = \sqrt{2\pi}F(\omega)G(\omega). \tag{74}$$

Let us now express the above result as follows,

$$\mathcal{F}(f * g) = \sqrt{2\pi}FG. \tag{75}$$

There is a symmetry between the Fourier transform and its inverse that we shall now exploit. With an eye toward Eqs. (17) and (21), let us now consider a convolution – in frequency space – between the two Fourier transforms $F(\omega)$ $G(\omega)$:

$$(F * G)(\omega) = \int_{-\infty}^{\infty} F(\omega - \lambda)G(\lambda)\,d\lambda. \tag{76}$$

We now take the *inverse Fourier transform* of this convolution, i.e.,

$$\begin{aligned} [\mathcal{F}^{-1}(F * G)](t) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (F * G)(\omega)\, e^{i\omega t}\,d\omega \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(\omega-\lambda)g(\lambda)\,d\lambda\, e^{i\omega}\,d\omega \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(\omega-\lambda)g(\lambda)\,d\lambda\, e^{i(\omega-\lambda)}e^{i\lambda}\,d\omega \end{aligned} \tag{77}$$

In the same way as we did before, we consider the change of variable

$$u(\omega,\lambda) = \lambda, \quad v(\omega,\lambda) = \omega - \lambda, \tag{78}$$

with Jacobian $J = 1$, implying that $\omega\ \lambda = uv$. The above integral separates, and the result is, as you might have guessed,

$$[\mathcal{F}^{-1}(F * G)](t) = \sqrt{2\pi}f(t)g(t). \tag{79}$$

We may also express this result as

$$\mathcal{F}[f(t)g(t)] = \frac{1}{\sqrt{2\pi}}(F * G)(\omega). \tag{80}$$

In other words, the Fourier transform of a product of functions is, up to a constant, the same as the convolution of their Fourier transforms. We shall refer to this result as *Convolution Theorem – Version 2*.

You might well wonder, "When would we be interested in the Fourier transform of a product of functions?" We'll return to this question a little later. For the moment, it is interesting to ponder these two versions of the Convolution Theorem as manifestations of a single theme:

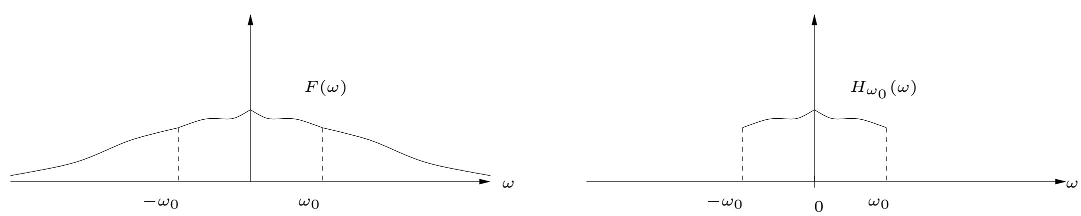
A product of functions in one representation is equivalent to a convolution in the other.

Using the Convolution Theorem to reduce high-frequency components

As we saw in our discussion of the discrete Fourier transform, many denoising methods work on the principle of reducing high-frequency components of the noisy signal. Here we examine two simple methods applied to the Fourier transform which are essentially continuous versions of what was discussed for DFTs.

Frequency thresholding

We return to the idea of "frequency thresholding." For continuous FTs, the idea is basically the same: Given a threshold frequency $\omega_0 > 0$, you "wipe out" the Fourier transform $F(\omega)$ for $|\omega| > \omega_0$. This process is sketched below.



Frequency thresholding: Eliminating all Fourier transform components $F(\omega)$ for $|\omega| > \omega_0$.

Mathematically, this operation may be expressed as follows: Given an FT $F(\omega)$, we construct a new FT $H_{\omega_0}(\omega)$ as follows,

$$H_{\omega_0}(\omega) = F(\omega)G_{\omega_0}(\omega), \tag{81}$$

where $G_{\omega_0}(\omega)$ is a boxcar-type function,

$$G_{\omega_0}(\omega) = \begin{cases} 1, & -\omega_0 \leq \omega \leq \omega_0, \\ 0, & \text{otherwise.} \end{cases} \tag{82}$$

From Eq. (81), this has the effect of producing a new signal $h_{\omega_0}(t)$ from $f(t)$ in the following way

$$h_{\omega_0}(t) = \sqrt{2\pi}\,(f * g_{\omega_0})(t), \tag{83}$$

where $g_{\omega_0}(t) = \mathcal{F}^{-1}[G_{\omega_0}]$ is a scaled sinc function. (The exact form of G_{ω_0} is left as an exercise.)

Now, if you were faced with the problem of denoising a noisy signal $f(t)$, you have two possibilities of constructing the denoised signal $h_{\omega_0}(t)$:

1. (i) From $f(t)$, construct its FT $F(\omega)$. Then (ii) perform the "chopping operation" of Eq. (81) to produce $H_{\omega_0}(\omega)$. Finally, (iii) perform an inverse FT on H_{ω_0} to produce $h_{\omega_0}(t)$.

2. For each t, compute the convolution in Eq. (83).

Clearly, the first option represents less computational work. In the second case, you would have to perform an integration – in practice over N data points – for *each* value of t. This gives you an indication of why signal/image processors prefer to work in the frequency domain.

Gaussian weighting

The above method of frequency thresholding, i.e., "chopping the tails of $F(\omega)$," may be viewed as a rather brutal procedure, especially if many of the coefficients of the tail are not negligible. One may wish to employ a "smoother" procedure of reducing the contribution of high-frequency components. One method is to multiply the FT $F(\omega)$ with a Gaussian function, which we shall write as follows,

$$G_\kappa(\omega) = e^{-\frac{\omega^2}{2\kappa^2}}. \tag{84}$$

Note that we have *not* normalized this Gaussian in order to ensure that that $G_\kappa(0) = 1$: For ω near 0, $G_\kappa(\omega) \approx 1$, so that low-frequency components of $F(\omega)$ are not affected very much. Of course, by varying the standard deviation κ, we perform different amounts of high-frequency reduction: κ measures the spread of $G_\kappa(\omega)$ in ω-space.

We denote the Gaussian-weighted FT by

$$H_\kappa(\omega) = F(\omega)G_\kappa(\omega). \tag{85}$$

From the Convolution Theorem, it follows that the resulting signal will be given by

$$h_\kappa(t) = \sqrt{2\pi}\,(f * g_{\kappa 0})(t), \tag{86}$$

where

$$g_\kappa(t) = \mathcal{F}^{-1}[G_\kappa]. \tag{87}$$

Note that in this case, the convolution kernel $g_\kappa(t)$ is a *Gaussian function*. Moreover, the standard deviation of $g_\kappa(t)$ is $1/\kappa_0$. (Remember our derivation that the FT of a Gaussian is a Gaussian, and vice versa.)

Convolution with a Gaussian is a kind of *averaging* procedure, where most of the weight is put at the t-value around which the convolution is being performed. Note that for very large κ, most of the averaging is performed over a tiny interval centered at each t. Indeed, in the limit $\kappa \to \infty$, the averaging will involve only one point, i.e., $f(t)$. This is consistent with the fact that in the limit $\kappa \to \infty$, $G_\kappa(\omega) = 1$, i.e., the Fourier transform $F(\omega)$ is left unchanged. As a result, the signal is unchanged.

On the other hand, for κ very small, i.e., significant damping of higher-frequency components of $F(\omega)$, the averaging of $f(t)$ is performed over a larger domain centered at t.

Once again, it would seem more efficient to perform such a procedure in the frequency domain. That being said, many image processing packages allow you to perform such Gaussian filtering in the time or spatial/pixel domain. In the case of images, the result is a *blurring*, with perhaps some denoising, of the image.

The Gaussian function $G_\kappa(\omega)$ presented above is but one example of a *low-pass filter*. ("Low-pass" means that lower frequencies are allowed to pass through relatively untouched.) Any function $G(\omega)$ that satisfies the following properties,

1. $\lim_{\omega \to 0} G(\omega) = 1$,

2. $G(\omega) \to 0$ as $|\omega| \to \infty$,

qualifies as a low-pass filter. (In the next section, the constant 1 will be modified because of the $\sqrt{2\pi}$ factors that occur in the definition of the Fourier transform.)

A return to Convolution Theorem - Version 2

We now return to the question, "When would the second version of the Convolution Theorem, i.e.,

$$\mathcal{F}[f(t)g(t)] = \frac{1}{\sqrt{2\pi}}(F * G)(\omega). \tag{88}$$

which involves the product of two functions $f(t)$ and $g(t)$, be useful/interesting?"

One rather simple situation that may be modelled by the product of two functions is the *clipping* of a signal, i.e., taking a signal $f(t)$ defined over $\mathbf{R}$, and setting it to zero outside an interval of interest, say, $[0, T]$. This is quite analogous to the frequency thresholding method discussed earlier.

Setting $f(t)$ to zero outside the interval $[0, T]$ is equivalent to multiplying $f(t)$ by the boxcar-type function

$$g(t) = \begin{cases} 1, & 0 \le t \le T, \\ 0, & \text{otherwise.} \end{cases} \tag{89}$$

In the frequency domain, the operation $h(t) = f(t)g(t)$ becomes, from the second version of the Convolution Theorem,

$$H(\omega) = \frac{1}{\sqrt{2\pi}}(F * G)(\omega). \tag{90}$$

Here $G(\omega)$ is a sinc function.

Once again, such a clipping of the signal may be viewed as a rather brutal procedure and the convolution of its FT, $F(\omega)$, with a sinc function attests to this. In fact, an example of this procedure was already seen in Example No. 2 of Lecture 14, where we "clipped" the function $\cos 3t$ and considered it to be zero outside the interval $[-\pi, \pi]$. The result was a Fourier transform $F(\omega)$ - which was explicitly computed - that was nonzero over the entire real line. The function $F(\omega)$ may also be computed by means of a convolution of a suitable sinc function with the Fourier transform of the function $\cos 3t$. (The latter is actually a sum of two Dirac delta functions.)

Now that you have produced a function $h(t)$ that is defined on a finite interval $[a, b]$, it is better to analyze it in terms of Fourier series, i.e., a discrete version of the Fourier transform. This, of course, will implicitly define a T-periodic extension of $h(t)$. As we have seen earlier in this course, in order to bypass problems with discontinuities at the endpoints, it is better to consider the even extension of $h(t)$ to the interval $[-T, T]$ and then the 2π-periodic extension of this function. This is done by simply using a Fourier cosine series expansion for $h(t)$, using the data on $[0, T]$.

The above operation of "clipping" may be considered as a primitive attempt to study the local properties of $f(t)$ over the interval $[0, T]$. From our discussions on frequency damping, one may wish to consider less brutal localizations, for example, the multiplication of $f(t)$ by a Gaussian function $g_{\sigma_0,t_0}(t)$ with variance σ_0^2 and centered at the point t_0, i.e.,

$$h(t) = f(t) g_{\sigma_0,t_0}(t). \tag{91}$$

The Gaussian function g is a kind of "envelope function." By varying t_0, we may examine the properties of $h(t)$ at various places in the signal. This is the basis of *windowed* transforms, for example, the so-called "Gabor transform," which is essentially a Fourier transform of such a Gaussian modulated function. We shall examine this method later in the course.

Lecture 17

Unfortunately, the Friday, February 15 lecture had to be cancelled. The plan was to present the following material, which is intended to be supplementary only. You are are **not** responsible for this material for the midterm or final examinations. That being said, it is hoped that you will go through this section because of its relevance to the course. It should give you a bigger picture of the effectiveness of several areas of mathematics, including Fourier transforms, PDEs and numerical methods in PDEs, in signal and image processing.

Most of the following material was taken from the instructor's lecture notes in AMATH 353, Partial Differential Equations I.

Supplementary reading: The heat equation and the "smoothing" of signals and images

In AMATH 231, and possibly other courses, you encountered the **heat** or **diffusion equation**, a very important example of a **partial differential equation**. In the case of a single spatial variable x and the time variable t, the heat equation for the function $u(x,t)$ has the form

$$\frac{\partial u}{\partial t} = k\frac{\partial^2 u}{\partial x^2}\,. \tag{92}$$

With appropriate boundary conditions and initial conditions, the function $u(x,t)$ could represent the temperature at a position x of a thin rod at time t.

In the case of three spatial dimensions, the heat equation for the function $u(\mathbf{x},t)$ has the form

$$\frac{\partial u}{\partial t} = k\nabla^2 u, \tag{93}$$

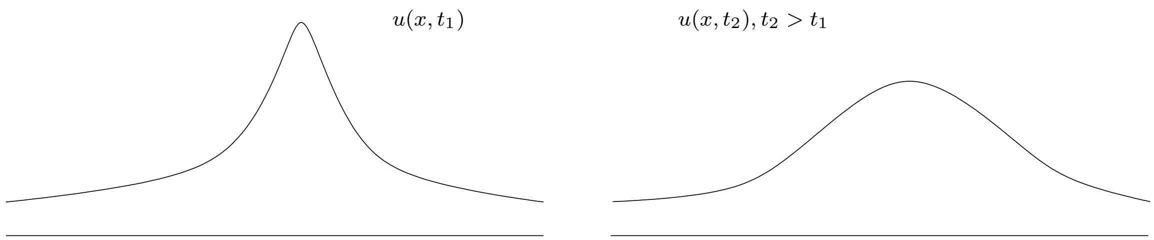
where ∇^2 denotes the Laplacian operator,

$$\nabla^2 u = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2}. \tag{94}$$

Here, the function $u(\mathbf{x},t)$ could represent the temperature at a point $\mathbf{x} \in \mathbf{R}^3$ at time t. Since this PDE also models diffusion, the function $u(\mathbf{x},t)$ could represent the concentration of a chemical in a solution.

It is not the purpose here to discuss the heat/diffusion equation in any detail. We simply wish to show that it is actually relevant to signal and image processing. Historically, it was the Fourier transform solution of the heat equation that inspired a great deal of research into the use of PDEs in signal and image processing. This research has led to many quite sophisticated mathematical methods of signal and image processing, including denoising and deblurring.

We first recall that solutions to the heat equation tend to "spread out" in time. (From this point onward, we shall simply write "heat equation" instead of "heat/diffusion equation.") In other words, any variations in temperature or concentration functions will become diminished in time. In the figure below, we show the graph of a temperature distribution that exhibits a concentration of heat energy around the position $x = a$. At a time $t_2 > t_1$, this heat energy distribution has been somehwat smoothened – some of the heat in the region of concentration has been transported to points farther away from $x = a$: This behaviour is to be expected if recall the derivation of the heat equation, which



was based on the physically-observated principle that heat moves from regions of higher temperature to regions of lower temperature. To be a little more precise, "Fourier's law of heat conduction" states that the **heat flux** vector at a point $\mathbf{x}$ is given by

$$\mathbf{h}(\mathbf{x}) = -\kappa \vec{\nabla} T(\mathbf{x}), \tag{95}$$

where $T(\mathbf{x})$ is the temperature at $\mathbf{x}$, and κ is the heat conductivity (possibly dependent on $\mathbf{x}$). In other words, the instantaneous direction of heat flow at $\mathbf{x}$ is the negative gradient of the temperature at $\mathbf{x}$.

In a first course on PDEs such as AMATH 353, one learns the actual mathematics behind the "smoothening" of temperature functions as they evolve according to the heat equation. With the help of the Fourier transform, one can show that solutions $u(x, t)$ to the heat equation are obtained by convolving the initial distribution $u(x, 0)$ with a Gaussian function that spreads in time. We shall first state the result for the one-dimensional heat equation and then derive it. The derivation provides an interesting opportunity to employ some of the properties of FTs discussed previously in a not-too-complicated manner.

Solution to 1D heat equation by convolution with a Gaussian heat kernel: Given the heat/diffusion equation on $\mathbf{R}$,

$$\frac{\partial u}{\partial t} = k\frac{\partial^2 u}{\partial x^2}, \qquad -\infty < x < \infty, \tag{96}$$

with initial condition

$$u(x,0) = f(x). \tag{97}$$

The solution to this initial value problem is

$$u(x,t) = \int_{-\infty}^{\infty} f(s)h_t(x-s)\,ds, \qquad t > 0, \tag{98}$$

where the "heat kernel" function $h_t(x)$ is given by

$$h_t(x) = \frac{1}{\sqrt{4\pi kt}}e^{-x^2/4kt}, \qquad t > 0. \tag{99}$$

This is a fundamental result – it states that the solution $u(x,t)$ is the *spatial convolution* of the functions $f(x)$ and $h_t(x)$.

Derivation of the above result: The derivation of Eq. (98) will be accomplished by taking Fourier transforms of the heat in Eq. 96). But we first have to clearly define what we mean by the Fourier transform of a solution $u(x,t)$ to the heat equation. At a given time $t \geq 0$, the FT of the solution function $u(x,t)$ is given by

$$\mathcal{F}(u(x,t)) = U(\omega,t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} u(x,t)e^{-i\omega x}dx. \tag{100}$$

Note that in this case the Fourier transform is given by an integration **over the independent variable** "x" and **not** "t" as was done in previous sections. This means that we may consider the FT of the solution function $u(x,t)$ as a function of both the frequency variable ω and the time t.

Let us now take Fourier transforms of both sides of the heat equation in (96), i.e.,

$$\mathcal{F}\left(\frac{\partial u}{\partial t}\right) = k\mathcal{F}\left(\frac{\partial^2 u}{\partial x^2}\right), \tag{101}$$

where we have acknowledged the linearity of the Fourier transform in moving the constant k out of the transform. It now remains to make sense of these Fourier transforms.

From our definition of the FT in Eq. (100), we now consider $\dfrac{\partial u}{\partial t}$ as a function of x and t and thereby write

$$\mathcal{F}\left(\frac{\partial u}{\partial t}\right) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} \frac{\partial u}{\partial t}(x,t)e^{-i\omega x}dx. \tag{102}$$

The only t-dependence in the integral on the right is in the integrand $u(x,t)$. As a result, we may write

$$\begin{aligned} \mathcal{F}\left(\frac{\partial u}{\partial t}\right) &= \frac{\partial}{\partial t}\left[\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} u(x,t)e^{-i\omega x}dx\right] \\ &= \frac{\partial}{\partial t}U(\omega,t). \end{aligned} \tag{103}$$

In other words, the partial time-derivative of $u(\omega,t)$ is simply the partial time-derivative of $U(\omega,t)$. Once again, we can bring the partial derivative operator $\partial/\partial t$ outside the integral because it has nothing to do with the integration over x.

We must now make sense of the right hand side of Eq. (101). Recalling that x is the variable of integration and time t is simply another variable, we may employ the property for Fourier transforms of derivatives of functions: Given a function $f(x)$ with Fourier transform $F(\omega)$, then

$$\mathcal{F}(f'(x)) = (-i\omega)F(\omega). \tag{104}$$

(Note once again that x is our variable of integration now, not t.) Applying this result to our function $u(x,t)$, we have

$$\mathcal{F}\left(\frac{\partial u}{\partial x}\right) = -i\omega\mathcal{F}(u) = -i\omega U(\omega,t). \tag{105}$$

We apply this result one more time to obtain the FT of $\dfrac{\partial^2 u}{\partial x^2}$:

$$\begin{aligned} \mathcal{F}\left(\frac{\partial^2 u}{\partial x^2}\right) &= -\omega^2\mathcal{F}(u) \\ &= = \omega^2 U(\omega,t). \end{aligned} \tag{106}$$

Now substitute the results of these calculations into Eq. (101) to give

$$\frac{\partial U(\omega,t)}{\partial t} = -k\omega^2 U(\omega,t). \tag{107}$$

Taking Fourier transforms of both sides of the heat equation converts a PDE involving both partial derivatives in x and t into a PDE that has only partial derivatives in t. This means that we can solve Eq. (107) as we would an **ordinary differential equation** in the independent variable t – in essence, we may ignore any dependence on the ω variable.

The above equation may be solved in the same way as we solve a first-order linear ODE in t. (Actually it is also separable.) We write it as

$$\frac{\partial U(\omega,t)}{\partial t} + k\omega^2 U(\omega,t) = 0, \tag{108}$$

and note that the integrating factor is $I(t) = e^{k\omega^2 t}$ to give

$$\frac{\partial}{\partial t}\left[e^{k\omega^2 t}U(\omega, t)\right] = 0. \tag{109}$$

Integrating (partially) with respect to t yields

$$e^{k\omega^2 t}U(\omega, t) = C \tag{110}$$

which implies that

$$U(\omega, t) = Ce^{-k\omega^2 t}. \tag{111}$$

Here, C is a "constant" with respect to partial differentiation by t. This means that C can be a function of ω. As such, we'll write

$$U(\omega, t) = C(\omega)e^{-k\omega^2 t}. \tag{112}$$

You can differentiate this expression partially with respect to t to check that it satisfies Eq. (107).

Now notice that at time $t = 0$,

$$U(\omega, 0) = C(\omega). \tag{113}$$

In other words, $C(\omega)$ is the FT of the function $u(x, 0)$. But this is the initial temperature distribution $f(x)$, i.e.,

$$C(\omega) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} f(x)e^{-i\omega x}dx = F(\omega). \tag{114}$$

Therefore, the solution Eq. (112) may be rewritten as

$$U(\omega, t) = F(\omega)e^{-k\omega^2 t}. \tag{115}$$

The solution of the heat equation in (96) may now be obtained by taking inverse Fourier transforms of the above result, i.e.,

$$u(x, t) = \mathcal{F}^{-1}(U(\omega, t)) = \mathcal{F}^{-1}(F(\omega)e^{-k\omega^2 t}). \tag{116}$$

Note that we must take the inverse FT of a product of FTs. This should invoke the memory of the Convolution Theorem (Version 1) from Lecture 15:

$$\mathcal{F}(f * g)(x) = \sqrt{2\pi}\,F(\omega)G(\omega), \tag{117}$$

which we may rewrite as

$$\mathcal{F}^{-1}(F(\omega)G(\omega)) = \frac{1}{\sqrt{2\pi}}(f * g)(x). \tag{118}$$

(Note that we have now used x to denote the independent variable instead of t.) From Eq. (116), the solution to the heat equation is given by

$$u(x,t) = \frac{1}{\sqrt{2\pi}}(f * g)(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} f(s)g(x-s)\,ds, \tag{119}$$

where $g(x,t)$ is the inverse Fourier transform of the function,

$$G(\omega,t) = e^{-k\omega^2 t} = e^{-(kt)\omega^2}. \tag{120}$$

We have grouped the k and t variables together to form a constant that multiplies the ω^2 in the Gaussian.

The next step is to find $g(x,t)$, the inverse FT of $G(\omega,t)$ - recall that t is an auxiliary variable. We can now invoke another earlier result – that the FT of a Gaussian, this time in x-space, is a Gaussian in ω-space. In Lecture 16, we showed that the FT of the Gaussian function,

$$f_\sigma(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{x^2}{2\sigma^2}}, \tag{121}$$

(note again that we are using x as the independent variable) is

$$F(\omega) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\sigma^2\omega^2}. \tag{122}$$

We can use this result to find our desired inverse FT of $G(\omega,t)$. First note that both $f_\sigma(x)$ and $F(\omega)$ have a common factor $1/\sqrt{2\pi}$. We now ignore this factor which allows us to write that

$$\mathcal{F}^{-1}(e^{-\frac{1}{2}\sigma^2\omega^2}) = \frac{1}{\sigma}e^{-\frac{x^2}{2\sigma^2}}. \tag{123}$$

We now compare Eqs. (122) and (120) and let

$$\frac{1}{2}\sigma^2 = kt, \tag{124}$$

which implies that

$$\sigma = \sqrt{2kt}. \tag{125}$$

We substitute this result into Eq. (123) to arrive at our desired result, namely, that

$$g(x,t) = \mathcal{F}^{-1}(e^{-kt\omega^2}) = \frac{1}{\sqrt{2kt}}e^{-\frac{x^2}{4kt}}. \tag{126}$$

From Eq. (119), and the definition of the convolution, we arrive at our final result for the solution to the heat equation,

$$
\begin{aligned}
u(x,t) &= \frac{1}{\sqrt{2\pi}}(f*g)(x) \\
&= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} f(s)g(x-s)\,ds \\
&= \int_{-\infty}^{\infty} f(s)h_t(x-s)\,ds, \qquad (127)
\end{aligned}
$$

where

$$h_t(x) = \frac{1}{\sqrt{2\pi}}g(x,t) = \frac{1}{\sqrt{4\pi kt}}e^{-x^2/4kt} \qquad (128)$$

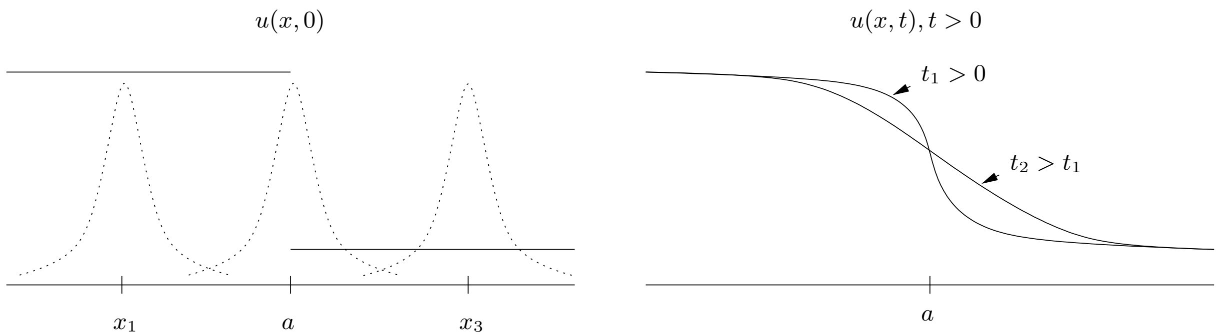
denotes the so-called **heat kernel**. This concludes the derivation.

The heat kernel $h_t(x)$ is a Gaussian function that "spreads out" in time: its standard deviation is given by $\sigma(t) = \sqrt{2kt}$. The convolution in Eq. (127) represents a *weighted averaging* of the function $f(s)$ about the point x – as time increases, this weighted averaging includes more points away from x, which accounts for the increased smoothing of the temperature distribution $u(x,t)$.

In the figure below, we illustrate how a discontinuity in the temperature function gets smoothened into a continuous distribution. At time $t = 0$, there is a discontinuity in the temperature function $u(x,0) = f(x)$ at $x = a$. Recall that we can obtain all future distributions $u(x,t)$ by convolving $f(x)$ with the Gaussian heat kernel. For some representative time $t > 0$, but not too large, three such Gaussian kernels are shown at points $x_1 < a$, $x_2 = a$ and $x_3 > a$. Since t is not too large, the Gaussian kernels are not too wide. A convolution of $u(x,0)$ with these kernels at x_1 and x_3 will not change the distribution too significantly, since $u(x,0)$ is constant over most of the region where these kernels have significant values. However, the convolution of $u(x,0)$ with the Gaussian kernel centered at the point of discontinuity $x = x_2$ will involve values of $u(x,0)$ to the left of the discontinuity as well as values of $u(x,0)$ to the right. This will lead to a significant averaging of the temperature values, as shown in the figure. At a time $t_2 > t_1$, the averaging of $u(x,0)$ with even wider Gaussian kernels will produce even greater smoothing.

The smoothing of temperature functions (and image functions below) can also be seen by looking at what happens in "frequency space," i.e., the evolution of the Fourier transform $U(\omega,t)$ of the solution $u(x,t)$, as given in Eq. (115), which we repeat below:

$$U(\omega,t) = F(\omega)e^{-k\omega^2 t}, \quad t \geq 0. \qquad (129)$$



At any fixed frequency $\omega > 0$, as t increases, the multiplicative Gaussian term $e^{-k\omega^2 t}$ decreases, implying that the frequency component, $U(\omega, t)$, decreases. For higher ω values, this decrease is faster. Therefore, higher frequencies are dampened more quickly than lower frequencies. We saw this phenomenon at the end of the the previous lecture.

We now move away from the idea of the heat or diffusion equation modelling the behaviour of temperatures or concentrations. Instead, we consider the function $u(x, t)$, $x \in \mathbf{R}$, as representing a time-varying *signal*. And in the case of two spatial dimensions, we may consider the function $u(x, y, t)$ as representing the time evolution of an *image*. Researchers in the signal and image processing noted the smoothing behaviour of the heat/diffusion equation quite some time ago and asked whether it could be exploited to accomplish desired tasks with signals and images. We outline some of these ideas very briefly below for the case of images, since the results are quite dramatic visually.

First of all, images generally contain many points of discontinuity – these are the *edges* of the image. In fact, edges are considered to define an image to a very large degree, since the boundaries of any objects in the image produce edges. In what follows, we shall let the function $u(x, y, t)$ denote the evolution of a (non-negative) image function under the heat/diffusion equation,

$$\frac{\partial u}{\partial t} = k\nabla^2 u = k\left[\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}\right], \qquad u(x, y, 0) = f(x, y). \tag{130}$$

Here we shall not worry about the domain of definition of the image. In practice, of course, image functions are defined on a finite set, e.g., the rectangular region $[a, b] \times [c, d] \subset \mathbf{R}^2$.

Some additional notes on the representation of images

Digitized images are defined over a finite set of points in a rectangular domain. Therefore, they are essentially $m \times n$ arrays of greyscale or colour values – the values that are used to assign brightness values to pixels on a computer screen. In what follows, we may consider such matrices to define images that are piecewise constant over a region of $\mathbf{R}^2$. We shall also be looking only at black-and-white (BW) images. Mathematically, we may consider the *range* of *greyscale values* of a BW image function to be an interval $[A, B]$, with A representing black and B representing white. In mathematical analysis, one usually assumes that $[A, B] = [0, 1]$. As you probably know, the practical storage of digitized images also involves a quantization of the greyscale values into discrete values. For so-called "8 bit-per-pixel" BW images, where each pixel may assume one of $2^8 = 256$ discrete values (8 bits of computer memory are used to the greyscale value at each pixel), the greyscale values assume the values (black) $0, 1, 2, \cdots, 255$ (white). This *greyscale range*, $[A, B] = [0, 255]$ is used for the analysis and display of the images below.

From our earlier discussions, we expect that edges of the input image $f(x, y)$ will become more and more smoothened as time increases. The result will be an increasingly *blurred* image, as we show in Figure 1 below. The top image in the figure is the input image $f(x, y)$. The bottom row shows the image $u(x, y, t)$ for two future times. The solutions $u(x, y, t)$ we computed by means of a 2D finite-difference scheme using forward time difference and centered difference for the Laplacian. It assumes the following form,

$$u_{ij}^{(n+1)} = u_{ij}^{(n)} + s[u_{i-1,j}^{(n)} + u_{i+1,j}^{(n)} + u_{i,j-1}^{(n)} + u_{i,j+1}^{(n)} - 4u_{ij}^{(n)}], \qquad s = \frac{k\Delta t}{(\Delta x)^2}. \tag{131}$$

(For details, you may consult the text by Haberman, Chapter 6, p. 253.) This scheme is numerically stable for $s < 0.25$: The value $s = 0.1$ was used to compute the images in the figure.

Heat/diffusion equation and "deblurring"

You may well question the utility of blurring an image: Why would one wish to degrade an image in this way? We'll actually provide an answer very shortly but, for the moment, let's make use of the blurring result in the following way: If we know that an image is blurred by the heat/diffusion equation as time *increases*, i.e., as time proceeds *forward*, then perhaps a blurry image can be deblurred by letting it evolve as time proceeds *backward*. The problem of "image deblurring" is an important one, e.g., acquiring the license plate numbers of cars that are travelling at a very high speed.

If everything proceeded properly, could a blurred edge could possibly be restored by such a procedure? In theory, the answer is "yes", provided that the blurred image is known for all (continuous)

Figure 1. Image blurring produced by the heat/diffusion equation. **Top:** 2288×1712 pixel (8 bits=256 grayscale levels/pixel) image as initial data function $f(x, y) = u(x, y, 0)$ for heat/diffusion equation on $\mathbf{R}^2$. **Bottom:** Evolution of image function $u(x, y, t)$ under discrete 2D finite-difference scheme (see main text). **Left:** After $n = 100$ iterations. **Right:** After $n = 500$ iterations.

values of x and y. In practice, however, the answer is "generally no", since we know only discrete, sampled values of the image. In addition, running the heat/diffusion equation backwards is an unstable process. Instead of the exponential damping that we saw very early in the course, i.e., an eigensolution in one dimension, $u_n(x,t) = \phi_n h_n(t)$ evolving as follows,

$$u_n(x,t) = \phi_n(x) e^{-k(n\pi/L)^2 t}, \tag{132}$$

we encounter exponential *increase*: replacing t with $-t$ yields,

$$u_n(x,t) = \phi_n(x) e^{k(n\pi/L)^2 t}. \tag{133}$$

As such, any inaccuracies in the function will be *amplified*. As a result, numerical procedures associated with running the heat/diffusion equation backwards are generally unstable.

To investigate this effect, the blurred image obtained after 100 iterations of the first experiment was used as the initial data for a heat/diffusion equation that was run backwards in time. This may be done by changing Δt to $-\Delta t$ in the finite difference scheme, implying that s is replaced by $-s$. The result is the following "backward scheme,"

$$u_{ij}^{(n-1)} = u_{ij}^{(n)} - s[u_{i-1,j}^{(n)} + u_{i+1,j}^{(n)} + u_{i,j-1}^{(n)} + u_{i,j+1}^{(n)} - 4u_{ij}^{(n)}]. \tag{134}$$

The first blurred image of the previous experiment (lower left image of Figure 1) was used as input into the above backward-time scheme. It is shown at the top of Figure 2. After five iterations, the image at the lower left of Figure 2 is produced. Some deblurring of the edges has been accomplished. (This may not be visible if the image is printed on paper, since the printing process itself introduces a degree of blurring.) After another five iterations, additional deblurring is achieved at some edges but at the expense of some severe degradation at other regions of the image. Note that much of the degradation occurs at smoother regions of the image, i.e., where spatially neighbouring values of the image function are closer to each other. This degradation is an illustration of the numerical instability of the backward-time procedure.

Image denoising under the heat/diffusion equation

We now return to the smoothing effect of the heat/diffusion equation and ask whether or not it could be useful. The answer is "yes" – it may be useful in the *denoising* of signals and/or images. In many applications, signals and images are degraded by noise in a variety of possible situations, e.g., (i) atmospheric disturbances, particularly in the case of images of the earth obtained from satellites or astronomical images obtained from telescopes, (ii) the channel over which such signals are transmitted are noisy, These are part of the overall problem of signal/image *degradation* which may include both

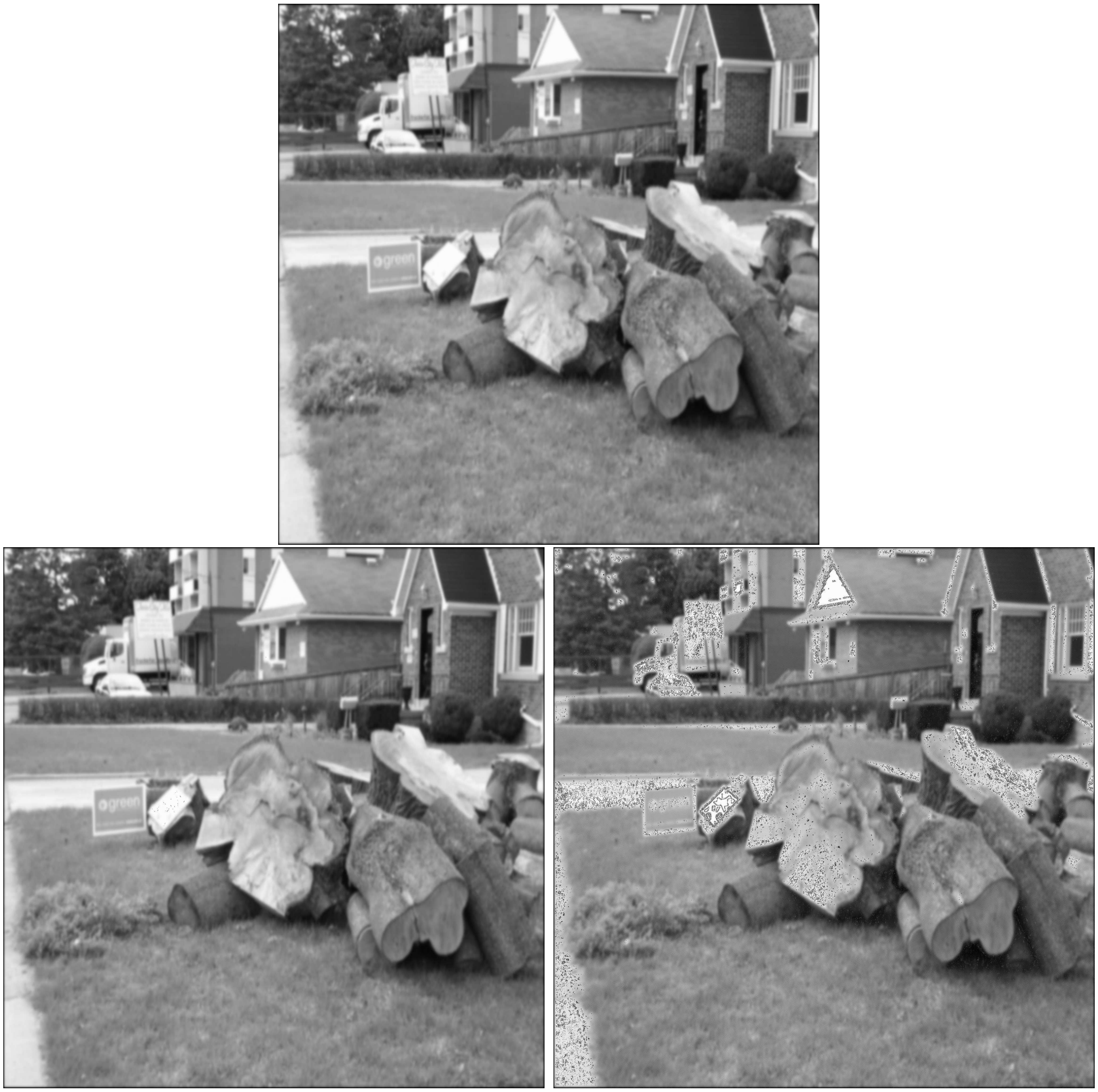
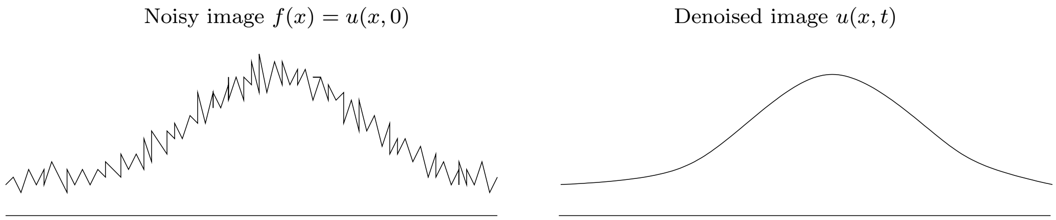


Figure 2. Attempts to deblur images by running the heat/diffusion equation backwards. **Top:** Blurred image $u_{ij}^{(100)}$ from previous experiment, as input into "backward" heat/diffusion equation scheme, with $s = 0.1$. **Bottom left:** Result after $n = 5$ iterations. Some deblurring has been achieved. **Bottom right:** Result after $n = 10$ iterations. Some additional deblurring but at the expense of degradation in some regions due to numerical instabilities.

blurring as well as noise. The removal of such degradations, which is almost always only partial, is known as *signal/image enhancement*.

A noisy signal may look something like the sketch at the left of Figure 3 below. Recalling that the heat/diffusion equation causes blurring, one might imagine that the blurring of a noisy signal may produce some deblurring, as sketched at the right of Figure 3. This is, of course, a very simplistic idea, but it does provide the starting point for a number of signal/image denoising methods.



A noisy signal (left) and its denoised counterpart (right).

In Figure 4 below, we illustrate this idea as applied to image denoising. The top left image is our original, "noiseless" image u. Some noise was added to this image to produce the noisy image $\tilde{u}$ at the top right. Very simply,

$$\tilde{u}(i,j) = u(i,j) + n(i,j), \tag{135}$$

where $n(i,j) \in \mathbf{R}$ was chosen randomly from the real line according to the normal Gaussian distribution $\mathcal{N}(0,\sigma)$, i.e., zero-mean and standard deviation σ. In this case, $\sigma = 20$ was used. The above equation is usually written more generally as follows,

$$\tilde{u} = u + \mathcal{N}(0,\sigma). \tag{136}$$

For reference purposes, the (discrete) L^2 error between u and $\tilde{u}$ was computed as follows,

$$\|\tilde{u} - u\|_2 = \sqrt{\frac{1}{512^2} \sum_{i,j=1}^{512} [\tilde{u}(i,j) - u(i,j)]^2} = 20.03. \tag{137}$$

This is the "root mean squared error" (RMSE) between the discrete functions $\tilde{u}$ and u. (You first compute the average value of the squares of the differences of the greyscale values at all pixels. Then take the square root of this average value.) In retrospect, it should be close to the standard deviation σ of the added noise. In other words, the average magnitude of the error between $u(i,j)$ and $\tilde{u}(i,j)$ should be the σ-value of the noise added.

The noisy image $\tilde{u}$ was then used as the input image for the diffusion equation, more specifically, the 2D finite difference scheme used earlier, i.e.,

$$u_{ij}^{(n+1)} = u_{ij}^{(n)} + s[u_{i-1,j}^{(n)} + u_{i+1,j}^{(n)} + u_{i,j-1}^{(n)} + u_{i,j+1}^{(n)} - 4u_{ij}^{(n)}], \tag{138}$$

with $s = 0.1$.

After five iterations (lower left), we see that some denoising has been produced, but at the expense of blurring, particularly at the edges. The L^2 distance between this denoised/blurred image and the original noiseless image u is computed to be

$$\|u_5 - u\|_2 = 16.30. \tag{139}$$

We see that from the viewpoint of L^2 distance, i.e., the denoised image u_5 is "closer" to the noiseless image u than the noisy image $\tilde{u}$. This is a good sign – we would hope that the denoising procedure would produce an image that is closer to u. But more on this later.

After another five iterations, as expected, there is further denoising but accompanied by additional blurring. The L^2 distance between this image and u is computed to be

$$\|u_{10} - u\|_2 = 18.23. \tag{140}$$

Note that the L^2 distance of this image is *larger* than that of u_5 – in other words, we have done *worse*. One explanation is that the increased blurring of the diffusion equation has degraded the image farther away from u than the denoising has improved it.

We now step back and ask: Which of the above results is "better" or "best"? In the L^2 sense, the lower left result is better since its L^2 error (i.e., distance to u) is smaller. But is it "better" visually? Quite often, a result that is better in terms of L^2 error is is poorer visually. And are the denoised images visually "better" than the noisy image itself. You will recall that some people in class had the opinion that the noisy image $\tilde{u}$ actually looked better than any of the denoised/blurred results. This illustrates an important point about image processing – the L^2 distance, although easy to work with, is not necessarily the best indicator of visual quality. Psychologically, our minds are sometimes more tolerant of noise than degradation in the edges – particularly in the form of blurring – that define an image.

Image denoising using "anisotropic diffusion"

We're not totally done with the idea of using the heat/diffusion equation to remove noise by means of blurring. Once upon a time, someone got the idea of employing a "smarter" form of diffusion – one which would perform blurring of images but which would leave their edges relatively intact. We could do this by making the diffusion parameter k to be sensitive to edges – when working in the vicinity of an edge, we restrict the diffusion so that the edges are not degraded. As we mentioned earlier, edges represent discontinuities – places where the magnitudes of the gradients become quite

Figure 4. Image denoising using the heat/diffusion equation. **Top left:** 512×512 pixel (8 bits=256 grayscale levels/pixel) *San Francisco* test image u. **Top right:** Noisy image, $\tilde{u} = u + \mathcal{N}(0, \sigma)$ (test image plus zero-mean Gaussian noise, $\sigma = 20$), which will serve as the initial data function for heat/diffusion equation on $\mathbf{R}^2$. L^2 error of noisy image: $\|\tilde{u} - u\|_2 = 20.03$.

Bottom: Evolution under discrete 2D finite-difference scheme (forward time difference scheme), **Left:** After $n = 5$ iterations, some denoising along with some blurring, $\|u_5 - u\|_2 = 16.30$. **Right:** After $n = 10$ iterations, some additional denoising with additional blurring, $\|u_{10} - u\|_2 = 18.23$.

large. (Technically, in the continuous domain, the gradients would be undefined. But we are working with finite differences, so the gradients will be defined, but large in magnitude.)

This implies that the diffusion parameter k would depend upon the position (x, y). But this is only part of the process – since k would be sensitive to the gradient $\vec{\nabla} u(x, y)$ of the image, it would, in fact, be dependent upon the image function $u(x, y)$ itself!

One way of accomplishing this selective diffusion, i.e., slower at edges, is to let $k(x, y)$ be *inversely proportional* to some power of the gradient, e.g.,

$$k = k(\|\vec{\nabla} u\|) = C\|\vec{\nabla} u\|^{-\alpha}, \qquad \alpha > 0. \tag{141}$$

The resulting diffusion equation,

$$\frac{\partial u}{\partial t} = k(\|\vec{\nabla} u\|)\nabla^2 u, \tag{142}$$

would be a *nonlinear* diffusion equation, since k is now dependent upon u, and it multiplies the Laplacian of u. And since the diffusion process is no longer constant throughout the region, it is no longer homogeneous but nonhomogeneous or *anisotropic*. As such, Eq. (142) is often called the *anisotropic diffusion equation*.

To illustrate this process, we have considered a very simple example, where

$$k(\|\vec{\nabla} u\|) = \|\vec{\nabla} u\|^{-1/2}. \tag{143}$$

Some results are presented in Figure 5. This simply anisotropic scheme works well to preserve edges, therefore producing better denoising of the noisy image used in the previous experiment. The denoised image u_{20} is better not only in terms of L^2 distance but also from the perspective of visual quality since its edges are better preserved.

Needless to say, a great deal of research has been done on nonlinear, anisotropic diffusion and its applications to signal and image processing.

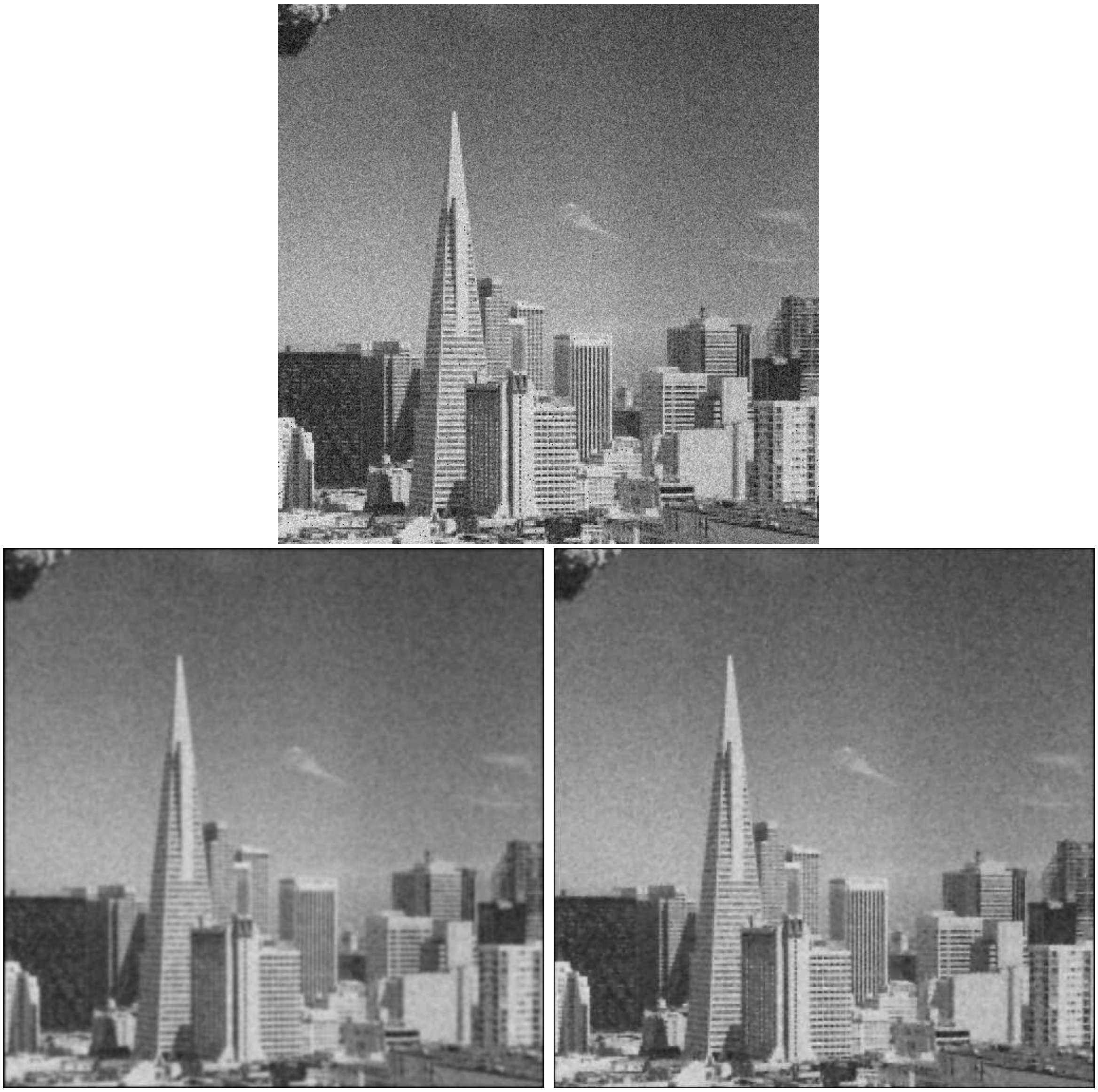


Figure 5. Image denoising and edge preservation via "anisotropic diffusion." **Top:** Noisy image, $\tilde{u} = u + \mathcal{N}(0, \sigma)$ (test image plus zero-mean Gaussian noise, $\sigma = 20$). L^2 error of noisy image: $\|\tilde{u} - u\|_2 = 20.03$.
Bottom left: Denoising with (isotropic) heat/diffusion equation, $u_t = k\nabla^2 u$, reported earlier. Finite-difference scheme, $s = 0.1$, $n = 5$ iterations. L^2 of denoised image: $\|u_5 - u\|_2 = 16.30$. **Bottom right:** Denoising with anisotropic heat equation, Finite-difference scheme, $s = 0.1$, $k(\|\vec{\nabla} u\|) = \|\vec{\nabla} u\|^{-1/2}$, $n = 20$ iterations. There is denoising but much less blurring around edges. L^2 error: $\|u_{20} - u\|_2 = 15.08$. Not only is the result from anisotropic diffusion better in the L^2 sense but it is also better visually, since edges have been better preserved.