Chapter 4

Fourier Analysis

4.1 Motivation

At the beginning of this course, we saw that superposition of functions in terms of sines and cosines was extremely useful for solving problems involving linear systems. For instance, when we studied the forced harmonic oscillator, we first solved the problem by assuming the forcing function was a sinusoid (or complex exponential). This turned out to be easy. We then argued that since the equations were linear this was enough to let us build the solution for an arbitrary forcing function if only we could represent this forcing function as a sum of sinusoids. Later, when we derived the continuum limit of the coupled spring/mass system we saw that separation of variables led us to a solution, but only **if** we could somehow represent general initial conditions as a sum of sinusoids. The representation of arbitrary functions in terms of sines and cosines is called *Fourier analysis.*

Jean Baptiste Joseph Fourier. Born: 21 March 1768 in Auxerre. Died: 16 May 1830 in Paris. Fourier trained as a priest and nearly lost his head (literally) in the French revolution. He is best known for his work on heat conduction. Fourier established the equation governing diffusion and used infinite series of trigonometric functions to solve it. Fourier was also a scientific adviser to Napoleon's army in Egypt.

4.2 The Fourier Series

So, the motivation for further study of such a Fourier superposition is clear. But there are other important reasons as well. For instance, consider the data shown in Figure 4.1.

These are borehole tiltmeter measurements. A tiltmeter is a device that measures the local tilt relative to the earth's gravitational field. The range of tilts shown here is between -40 and 40 nanoradians! (There are 2 π radians in 360 degrees, so this range corresponds to about 8 millionths of a degree.) With this sensitivity, you would expect that the dominant signal would be due to earth tides. So buried in the time-series on the top you would expect to see two dominant frequencies, one that was diurnal (1 cycle per day) and one that was semi-diurnal (2 cycles per day). If we somehow had an automatic way of representing these data as a superposition of sinusoids of various frequencies, then might we not expect these characteristic frequencies to manifest themselves in the size of the coefficients of this superposition? The answer is yes, and this is one of the principle aims of Fourier analysis. In fact, the power present in the data at each frequency is called the power spectrum. Later we will see how to estimate the power spectrum using a Fourier transform.

You'll notice in the tiltmeter spectrum that the two peaks (diurnal and semi-diurnal seem to be split; i.e., there are actually two peaks centered on 1 cycle/day and two peaks centered on 2 cycles/day. Consider the superposition of two sinusoids of nearly the same frequency:

$$\sin((\omega - \epsilon)t) + \sin((\omega + \epsilon)t).$$

Show that this is equal to

$$2\cos(\epsilon t)\sin(\omega t).$$

Interpret this result physically, keeping in mind that the way we've set the problem up, ϵ is a small number compared to ω. It might help to make some plots. Once you've figured out the interpretation of this last equation, do you see evidence of the same effect in the tiltmeter data?

There is also a *drift* in the tiltmeter. Instead of the tides fluctuating about 0 tilt, they slowly drift upwards over the course of 50 days. This is likely a drift in the instrument and not associated with any tidal effect. Think of how you might *correct* the data for this drift.

As another example Figure 4.2 shows 50 milliseconds of sound (a low C) made by a soprano saxophone and recorded on a digital oscilloscope. Next to this is the estimated power spectrum of the same sound. Notice that the peaks in the power occur at integer multiples of the frequency of the first peak (the nominal frequency of a low C).

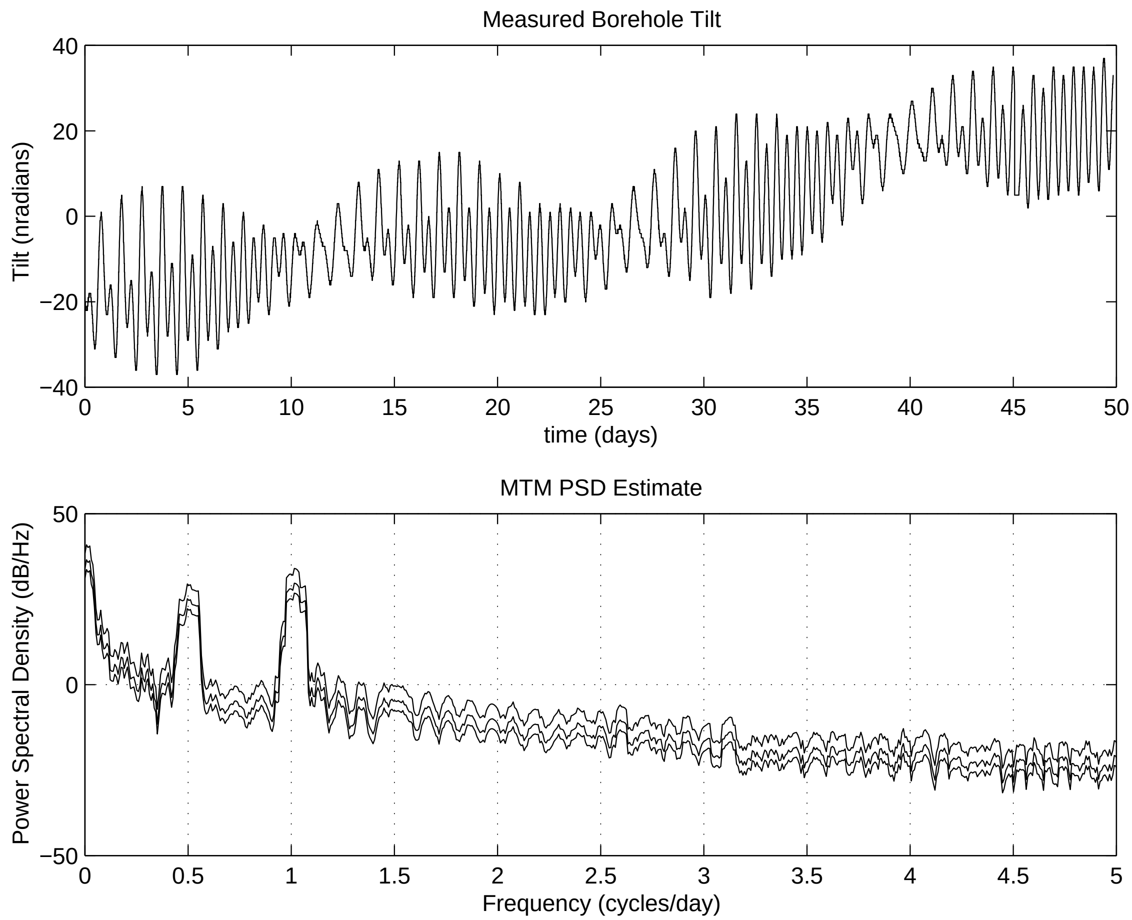


Figure 4.1: Borehole tiltmeter measurements. Data courtesy of Dr. Judah Levine (see [**?**] for more details). The plot on the top shows a 50 day time series of measurements. The figure on the bottom shows the estimated power in the data at each frequency over some range of frequencies. This is known as an estimate of the *power spectrum* of the data. Later we will learn how to compute estimates of the power spectrum of time series using the *Fourier transform*. Given what we know about the physics of tilt, we should expect that the diurnal tide (once per day) should peak at 1 cycle per day, while the semi-diurnal tide (twice per day) should peak at 2 cycles per day. This sort of analysis is one of the central goals of Fourier theory.

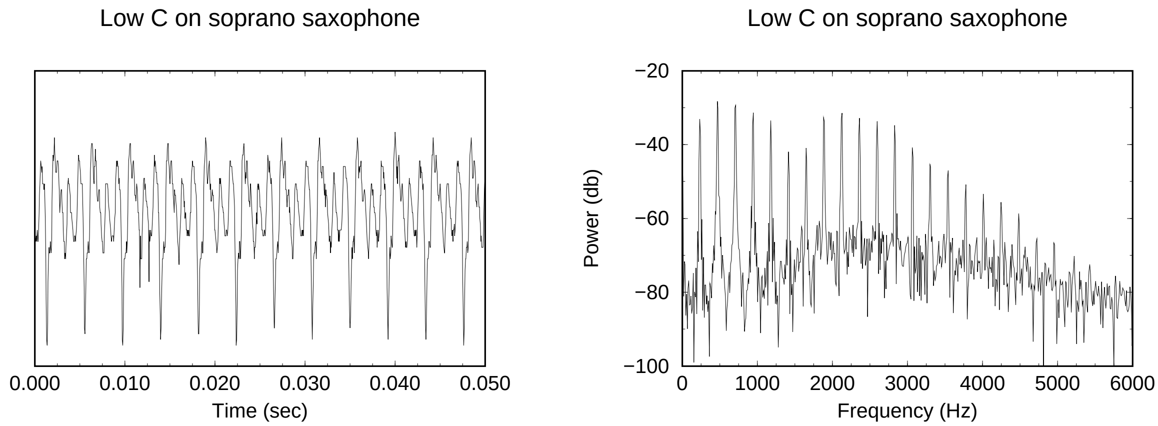


Figure 4.2: On the left is .05 seconds of someone playing low C on a soprano saxophone. On the right is the power spectrum of these data. We'll discuss later how this computation is made, but essentially what you're seeing the power as a function of frequency. The first peak on the right occurs at the nominal frequency of low C. Notice that all the higher peaks occur at integer multiples of the frequency of the first (fundamental) peak.

Definition of the Fourier Series

For a function periodic on the interval $[-l, l]$, the Fourier series is defined to be:

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos(n\pi x/l) + b_n \sin(n\pi x/l). \tag{4.2.1}$$

or equivalently,

$$f(x) = \sum_{n=-\infty}^{\infty} c_n e^{in\pi x/l}. \tag{4.2.2}$$

We will see shortly how to compute these coefficients. The connection between the real and complex coefficients is:

$$c_k = \frac{1}{2}(a_k - ib_k) \quad c_{-k} = \frac{1}{2}(a_k + ib_k). \tag{4.2.3}$$

In particular notice that the sine/cosine series has only positive frequencies, while the exponential series has both positive and negative. The reason is that in the former case each frequency has two functions associated with it. If we introduce a single complex function (the exponential) we avoid this by using negative frequencies. In other words, any physical vibration always involves two frequencies, one positive and one negative.

Later on you will be given two of the basic convergence theorems for Fourier series. Now let's look at some examples.

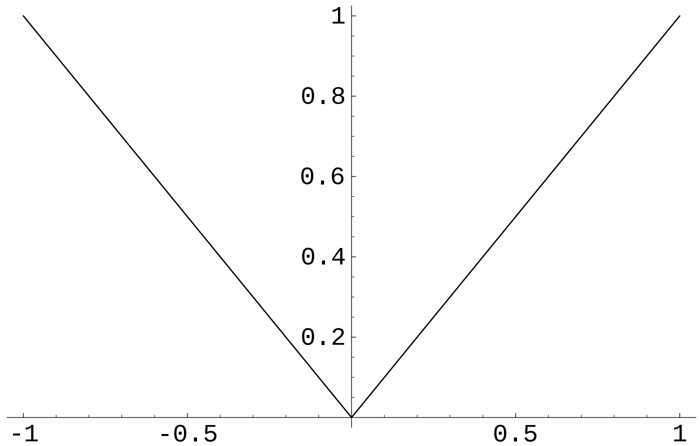


Figure 4.3: Absolute value function.

4.2.1 Examples

Let $f(x) = \mathrm{abs}(x)$, as shown in Figure 4.3. The first few terms of the Fourier series are:

$$\frac{1}{2} - \frac{4\,\cos(\pi\,x)}{\pi^2} - \frac{4\,\cos(3\,\pi\,x)}{9\,\pi^2} - \frac{4\,\cos(5\,\pi\,x)}{25\,\pi^2} \tag{4.2.4}$$

This approximation is plotted in Figure 4.3.

Observations

Note well that the convergence is slowest at the origin, where the absolute value function is not differentiable. (At the origin, the slope changes abruptly from -1 to +1. So the left derivative and the right derivative both exist, but they are not the same.) Also, as for any even function (i.e., $f(x) = f(-x)$) only the cosine terms of the Fourier series are nonzero.

Suppose now we consider an odd function (i.e., $f(x) = -f(-x)$), such as $f(x) = x$. The first four terms of the Fourier series are

$$\frac{2\,\sin(\pi\,x)}{\pi} - \frac{\sin(2\,\pi\,x)}{\pi} + \frac{2\,\sin(3\,\pi\,x)}{3\,\pi} - \frac{\sin(4\,\pi\,x)}{2\,\pi} \tag{4.2.5}$$

Here you can see that only the sine terms appear, and no constant (zero-frequency) term. A plot of this approximation is shown in Figure 4.4.

So why the odd behavior at the endpoints? It's because we've assume the function is periodic on the interval $[-1, 1]$. The *periodic extension* of $f(x) = x$ must therefore have

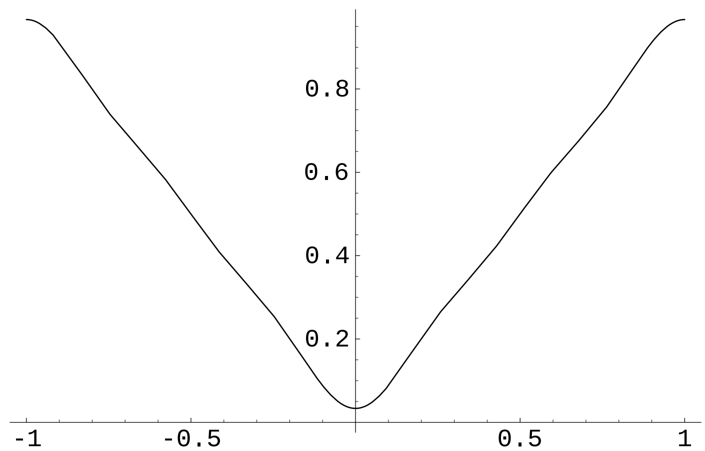


Figure 4.4: First four nonzero terms of the Fourier series of the function $f(x) = \text{abs}(x)$.

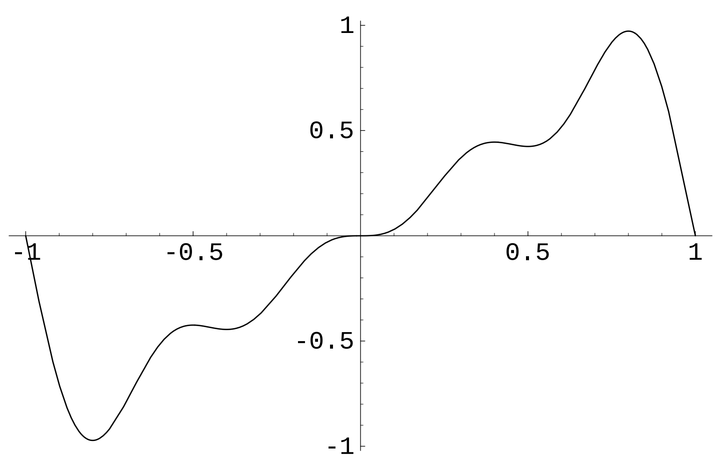


Figure 4.5: First four nonzero terms of the Fourier series of the function f(x) = x.

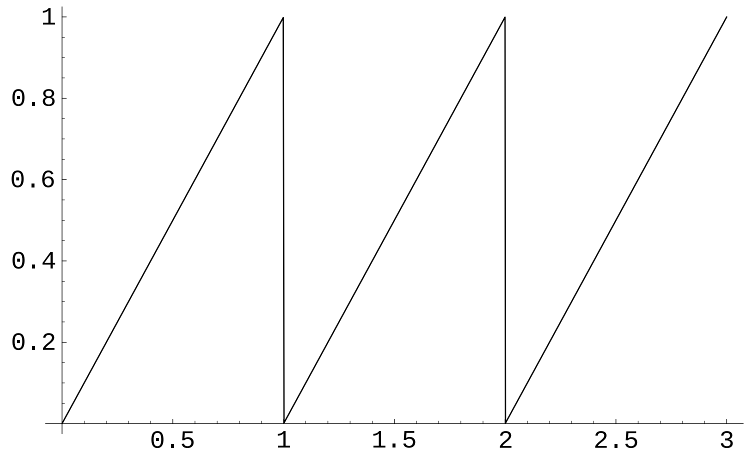


Figure 4.6: Periodic extension of the function $f(x) = x$ relative to the interval $[0, 1]$.

a sort of sawtooth appearance. In other words any non-periodic function defined on a finite interval can be used to generate a periodic function just by cloning the function over and over again. Figure 4.6 shows the periodic extension of the function $f(x) = x$ relative to the interval $[0, 1]$. It's a potentially confusing fact that the same function will give rise to different periodic extensions on different intervals. What would the periodic extension of $f(x) = x$ look like relative to the interval $[-.5, .5]$?

4.3 Superposition and orthogonal projection

Now, recall that for any set of N linearly independent vectors $\mathbf{x}_i$ in R^N, we can represent an arbitrary vector $\mathbf{z}$ in R^N as a superposition

$$\mathbf{z} = c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + \cdots + c_N\mathbf{x}_N, \tag{4.3.1}$$

which is equivalent to the linear system

$$\mathbf{z} = X \cdot \mathbf{c} \tag{4.3.2}$$

where X is the matrix whose columns are the $\mathbf{x}_i$ vectors and $\mathbf{c}$ is the vector of unknown expansion coefficients. As you well know, matrix equation has a unique solution $\mathbf{c}$ if and only if the $\mathbf{x}_i$ are linearly independent. But the solution is especially simple if the $\mathbf{x}_i$ are orthogonal. Suppose we are trying to find the coefficients of

$$\mathbf{z} = c_1\mathbf{q}_1 + c_2\mathbf{q}_2 + \cdots + \mathbf{q}_N, \tag{4.3.3}$$

when $\mathbf{q}_i \cdot \mathbf{q}_j = \delta_{ij}$. In this case we can find the coefficients easily by projecting onto the orthogonal directions:

$$c_i = \mathbf{q}_i \cdot \mathbf{z}, \tag{4.3.4}$$

or, in the more general case where the $\mathbf{q}$ vectors are orthogonal but not necessarily normalized

$$c_i = \frac{\mathbf{q}_i \cdot \mathbf{z}}{\mathbf{q}_i \cdot \mathbf{q}_i}. \tag{4.3.5}$$

We have emphasized throughout this course that functions are vectors too, they just happen to live in an infinite dimensional vector space (for instance, the space of square integrable functions). So it should come as no surprise that we would want to consider a formula just like 4.3.3, but with functions instead of finite dimensional vectors; e.g.,

$$f(x) = c_1 q_1(x) + c_2 q_2(x) + \cdots + c_n q_n(x) + \cdots. \tag{4.3.6}$$

In general, the sum will require an infinite number of coefficients c_i, since a function has an infinite amount of information. (Think of representing $f(x)$ by its value at each point x in some interval.) Equation 4.3.6 is nothing other than a Fourier series if the $q(x)$ happen to be sinusoids. Of course, you can easily think of functions for which all but a finite number of the coefficients will be zero; for instance, the sum of a finite number of sinusoids.

Now you know exactly what is coming. If the basis functions $q_i(x)$ are "orthogonal", then we should be able to compute the Fourier coefficients by simply projecting the function $f(x)$ onto each of the orthogonal "vectors" $q_i(x)$. So, let us define a dot (or inner) product for functions on an interval $[-l, l]$ (this could be an infinite interval)

$$(u, v) \equiv \int_{-l}^{l} u(x)v(x)dx. \tag{4.3.7}$$

Then we will say that two functions are orthogonal if their inner product is zero.

Now we simply need to show that the sines and cosines (or complex exponentials) are orthogonal. Here is the theorem. Let $\phi_k(x) = \sin(k\pi x/l)$ and $\psi_k(x) = \cos(k\pi x/l)$. Then

$$(\phi_i, \phi_j) = (\psi_i, \psi_j) = l\delta_{ij} \tag{4.3.8}$$

$$(\phi_i, \psi_j) = 0. \tag{4.3.9}$$

The proof, which is left as an exercise, makes use of the addition formulae for sines and cosines. (If you get stuck, the proof can be found in [2], Chapter 10.) A similar result holds for the complex exponential, where we define the basis functions as $\xi_k(x) = e^{ik\pi x/l}$.

Using Equations 4.3.8 and 4.3.9 we can compute the Fourier coefficients by simply projecting $f(x)$ onto each orthogonal basis vector:

$$a_n = \frac{1}{l}\int_{-l}^{l} f(x)\cos(n\pi x/l)dx = \frac{1}{l}(f, \psi_n), \tag{4.3.10}$$

and

$$b_n = \frac{1}{l}\int_{-l}^{l} f(x)\sin(n\pi x/l)dx = \frac{1}{l}(f, \phi_n). \tag{4.3.11}$$

Or, in terms of complex exponentials

$$c_n = \frac{1}{2l} \int_{-l}^{l} f(x) e^{-in\pi x/l} dx. \tag{4.3.12}$$

4.4 The Fourier Integral

For a function defined on any finite interval, we can use the Fourier series, as above. For functions that are periodic on some other interval than $[-l, l]$ all we have to do to use the above formulae is to make a linear change of variables so that in the new coordinate the function is defined on $[-l, l]$. And for functions that are not periodic at all, but still defined on a finite interval, we can fake the periodicity by replicating the function over and over again. This is called periodic extension.

OK, so we have a function that is periodic on an interval $[-l, l]$. Looking at its Fourier series (either Equation 4.2.1 or 4.2.2) we see straight away that the frequencies present in the Fourier synthesis are

$$f_1 = \frac{1}{2l}, \;\; f_2 = \frac{2}{2l}, \;\; f_3 = \frac{3}{2l}, \cdots, \;\; f_k = \frac{k}{2l} \cdots \tag{4.4.1}$$

Suppose we were to increase the range of the function to a larger interval $[-L, L]$ trivially by defining it to be zero on $[-L, -l]$ and $[l, L]$. To keep the argument simple, let us suppose that $L = 2l$. Then we notice two things straight away. First, the frequencies appearing in the Fourier synthesis are now

$$f_1 = \frac{1}{2L}, \;\; f_2 = \frac{2}{2L}, \;\; f_3 = \frac{3}{2L}, \cdots, \;\; f_k = \frac{k}{2L} \cdots \tag{4.4.2}$$

So the frequency interval is half what it was before. And secondly, we notice that half of the Fourier coefficients are the same as before, with the new coefficients appearing mid-way between the old ones. Imagine continuing this process indefinitely. The Fourier coefficients become more and more densely distributed, until, in the limit that $L \to \infty$, the coefficient sequence c_n becomes a continuous function. We call this function the Fourier transform of $f(x)$ and denote it by $F(k)$. In this case, our Fourier series

$$f(x) = \sum_{n=-\infty}^{\infty} c_n e^{in\pi x/l}$$

becomes

$$f(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} F(k) e^{ikx} dk \tag{4.4.3}$$

with the "coefficient" function $F(k)$ being determined, once again, by orthogonal projection:

$$F(k) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x) e^{-ikx} dx \tag{4.4.4}$$

Normalization

A function $f(t)$ is related to its Fourier transform $f(\omega)$ via:

$$f(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} F(\omega) e^{i\omega t} \, d\omega \tag{4.4.5}$$

and

$$F(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(t) e^{-i\omega t} \, dt \tag{4.4.6}$$

It doesn't matter how we split up the 2π normalization. For example, in the interest of symmetry we have defined both the forward and inverse transform with a $1/\sqrt{2\pi}$ out front. Another common normalization is

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega) e^{i\omega t} \, d\omega \tag{4.4.7}$$

and

$$F(\omega) = \int_{-\infty}^{\infty} f(t) e^{-i\omega t} \, dt. \tag{4.4.8}$$

It doesn't matter how we do this as long as we're consistent. We could get rid of the normalization altogether if we stop using circular frequencies ω in favor of f measured in hertz or cycles per second. Then we have

$$g(t) = \int_{-\infty}^{\infty} G(f) e^{2\pi i f t} \, df \tag{4.4.9}$$

and

$$G(f) = \int_{-\infty}^{\infty} g(t) e^{-2\pi i f t} \, dt \tag{4.4.10}$$

Here, using time and frequency as variables, we are thinking in terms of time series, but we could just as well use a distance coordinate such as x and a wavenumber k:

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(k) e^{ikx} \, dk \tag{4.4.11}$$

with the inverse transformation being

$$F(k) = \int_{-\infty}^{\infty} f(x) e^{-ikx} \, dx. \tag{4.4.12}$$

Invertibility: the Dirichlet Kernel

These transformations from time to frequency or space to wavenumber are invertible in the sense that if we apply one after the other we recover the original function. To see this plug Equation (4.4.12) into Equation (4.4.11):

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dk \int_{-\infty}^{\infty} f(x') e^{-ik(x'-x)} \, dx'. \tag{4.4.13}$$

If we define the kernel function $K(x - x', \mu)$ such that

$$K(x' - x, \mu) = \frac{1}{2\pi} \int_{-\mu}^{\mu} e^{-ik(x'-x)} \, dk = \frac{\sin \mu(x' - x)}{\pi(x' - x)} \tag{4.4.14}$$

then we have

$$f(x) = \int_{-\infty}^{\infty} f(x')K(x' - x)dx' \tag{4.4.15}$$

where $K(x' - x)$ is the limit (assuming that it exists) of $K(x' - x, \mu)$ as $\mu \to \infty$. In order for this to be true $K(x' - x)$ will have to turn out to be a *Dirac delta function.*

In one space dimension, the Dirac delta function is defined by the property that for any interval I, $f(x) = \int_I f(y)\delta(y - x)dy$ if x is in I and zero otherwise. (We can also write this as $f(0) = \int_I f(y)\delta(y)dy$.) No ordinary function can have this property since it implies that $\delta(y - x)$ must be zero except when $x = y$. If you try integrating any function which is finite at only one point (and zero everywhere else), then you always get zero. This means that $\int_I f(y)\delta(y - x)dy$ would always be zero if $\delta(0)$ is finite. So $\delta(0)$ must be infinite. And yet the function $\delta(x)$ itself must integrate to 1 since if we let $f(x) = 1$, then the basic property of the delta function says that: $1 = \int \delta(y)dy$. So we are forced to conclude that $\delta(x)$ has the strange properties that it is zero, except when $x = 0$, it is infinite when $x = 0$ and that it integrates to 1. This is no ordinary function.

The Dirac delta function is named after the Nobel prize-winning English physicist Paul A.M. Dirac (born August 1902, Bristol, England; died October 1984, Tallahassee, Florida). Dirac was legendary for making inspired physical predictions based on abstract arguments. His book *Principals of Quantum Mechanics* was one of the most influential scientific books of the 20th century. He got the Nobel Prize in Physics in 1933. Amazingly, Dirac had published 11 significant papers before his completed his PhD work. Along with Newton, Dirac is buried in Westminster Abbey.

We won't attempt to prove that the kernel function converges to a delta function and hence that the Fourier transform is invertible; you can look it up in most books on analysis. But Figure 4.7 provides graphical evidence. We show plots of this kernel function for $x = 0$ and four different values of μ, 10, 100, 1000, and 10000. It seems pretty clear that in the limit that $\mu \to \infty$, the function K becomes a Dirac delta function.

4.4.1 Examples

Let's start with an easy but interesting example. Suppose we want the compute the Fourier transform of a box-shaped function. Let $f(x)$ be equal to 1 for x in the interval

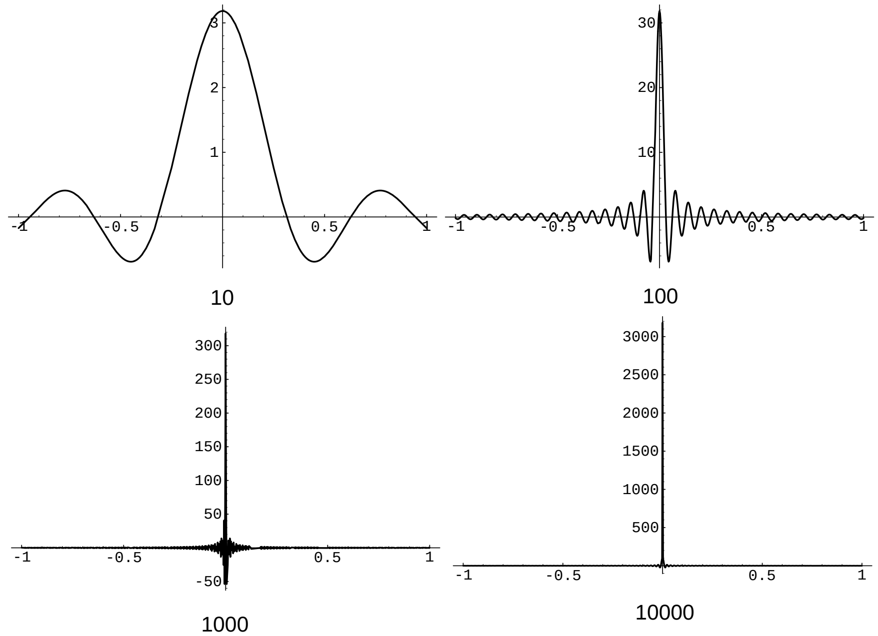


Figure 4.7: The kernel $\sin \mu x/\pi x$ for $\mu = 10$, 100, 1000, and 10000.

$[-1, 1]$ and 0 otherwise. So we need to compute

$$\int_{-1}^{1} e^{-ikx} dx = \frac{2 \sin\ k}{k}.$$

This function is shown in Figure 4.8 and is just the Dirichlet kernel for $\mu = 1$, centered about the origin.[1]

Here is a result which is a special case of a more general theorem telling us how the Fourier transform scales. Let $f(x) = e^{-x^2/a^2}$. Here a is a parameter which corresponds to the width of the bell-shaped curve. Make a plot of this curve. When a is small, the curve is relatively sharply peaked. When a is large, the curve is broadly peaked. Now compute the Fourier transform of f:

$$F(k) \propto \int_{-\infty}^{\infty} e^{-x^2/a^2} e^{-ikx} dx.$$

The trick to doing integrals of this form is to complete the square on the exponentials. You want to write the whole thing as an integral of the form

$$\int_{-\infty}^{\infty} e^{-z^2} dz.$$

As you'll see shortly, this integral can be done analytically. The details will be left as an exercise, here we will just focus on the essential feature, the exponential.

$$e^{-x^2/a^2} e^{-ikx} = e^{-1/a^2 \left[(x+ika^2/2)^2 + (ka^2/2)^2 \right]}.$$

[1] To evaluate the limit of this function at $k = 0$, use L'Hôpital's rule.

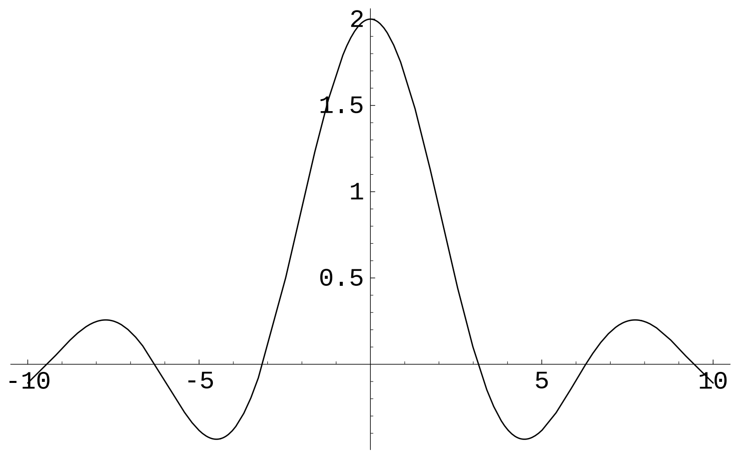


Figure 4.8: The Fourier transform of the box function.

So the integral reduces to

$$ae^{-k^2a^2/4}\int_{-\infty}^{\infty} e^{-z^2}dz = \sqrt{\pi}ae^{-k^2a^2/4}.$$

(The $\sqrt{\pi}$ will come next.) So we see that in the Fourier domain the factor of a^2 appears in the numerator of the exponential, whereas in the original domain, it appeared in the denominator. Thus, making the function more peaked in the space/time domain makes the Fourier transform more broad; while making the function more broad in the space/time domain, makes it more peaked in the Fourier domain. This is a very important idea.

Now the trick to doing the Gaussian integral. Since

$$H = \int_{-\infty}^{\infty} e^{-x^2}\ dx$$

$$H^2 = \left[\int_{-\infty}^{\infty} e^{-x^2}\ dx\right]\left[\int_{-\infty}^{\infty} e^{-y^2}\ dy\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(x^2+y^2)}\ dx\ dy.$$

Therefore

$$H^2 = \int_0^{\infty}\int_0^{2\pi} e^{-r^2}r\ dr\ d\theta = \frac{1}{2}\int_0^{\infty}\int_0^{2\pi} e^{-\rho}\ d\rho\ d\theta = \pi$$

So $H = \sqrt{\pi}$.

4.4.2 Some Basic Theorems for the Fourier Transform

It is very useful to be able think of the Fourier transform as an operator acting on functions. Let us define an operator Φ via

$$\Phi[f] = F \tag{4.4.16}$$

where

$$F(\omega) = \int_{-\infty}^{\infty} f(t)e^{-i\omega t}\ dt. \tag{4.4.17}$$

Then it is easy to see that Φ is a linear operator

$$\Phi[c_1 f_1 + c_2 f_2] = c_1\Phi[f_1] + c_2\Phi[f_2]. \tag{4.4.18}$$

Next, if $f^{(k)}$ denotes the k-th derivative of f, then

$$\Phi[f^{(k)}] = (i\omega)^k \Phi[f] \quad k = 1, 2, \ldots \tag{4.4.19}$$

This result is crucial in using Fourier analysis to study differential equations. Next, suppose c is a real constant, then

$$\Phi[f(t-c)] = e^{-icw}\Phi[f] \tag{4.4.20}$$

and

$$\Phi[e^{ict}f(t)] = F(t-c) \tag{4.4.21}$$

where $F = \Phi(f)$. And finally, we have the convolution theorem. For any two functions $f(t)$ and $g(t)$ with $F = \Phi(f)$ and $G = \Phi(g)$, we have

$$\Phi(f)\Phi(g) = \Phi[f * g] \tag{4.4.22}$$

where "*" denotes convolution:

$$[f * g](t) = \int_{-\infty}^{\infty} f(\tau)g(t-\tau)d\tau. \tag{4.4.23}$$

The convolution theorem is one of the most important in time series analysis. Convolutions are done often and by going to the frequency domain we can take advantage of the algorithmic improvements of the fast Fourier transform algorithm (FFT).

The proofs of all these but the last will be left as an exercise. The convolution theorem is worth proving. Start by multiplying the two Fourier transforms. We will throw caution to the wind and freely exchange the orders of integration. Also, let's ignore the normalization for the moment:

$$\begin{aligned}
F(\omega)G(\omega) &= \int_{-\infty}^{\infty} f(t)e^{-i\omega t}dt \int_{-\infty}^{\infty} g(t')e^{-i\omega t'}dt' && (4.4.24)\\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-i\omega(t+t')}f(t)g(t')dt\ dt' && (4.4.25)\\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-i\omega\tau}f(t)g(\tau-t)dt\ d\tau && (4.4.26)\\
&= \int_{-\infty}^{\infty} e^{-i\omega\tau}\left[\int_{-\infty}^{\infty} f(t)g(\tau-t)dt\right]d\tau. && (4.4.27)
\end{aligned}$$

This completes the proof, but now what about the normalization? If we put the symmetric $1/\sqrt{2\pi}$ normalization in front of both transforms, we end up with a left-over factor of

$1/\sqrt{2\pi}$ because we started out with two Fourier transforms and we ended up with only one and a convolution. On the other hand, if we had used an asymmetric normalization, then the result would be different depending on whether we put the $1/(2\pi)$ on the forward or inverse transform. This is a fundamental ambiguity since we can divide up the normalization anyway we want as long as the net effect is $1/(2\pi)$. This probably the best argument for using f instead of ω since then the 2πs are in the exponent and the problem goes away.

4.5 The Sampling Theorem

Now returning to the Fourier transform, suppose the spectrum of our time series $f(t)$ is zero outside of some symmetric interval $[-2\pi f_s, 2\pi f_s]$ about the origin.[2] In other words, the signal does not contain any frequencies higher than f_s hertz. Such a function is said to be *band limited*; it contains frequencies only in the band $[-2\pi f_s, 2\pi f_s]$. Clearly a band limited function has a finite inverse Fourier transform

$$f(t) = \frac{1}{2\pi} \int_{-2\pi f_s}^{2\pi f_s} F(\omega) e^{-i\omega t} \, d\omega. \tag{4.5.1}$$

sampling frequencies and periods

f_s is called the *sampling* frequency. Hence the sampling period is $T_s \equiv 1/f_s$. It is sometimes convenient to normalize frequencies by the sampling frequency. Then the maximum normalized frequency is 1:

$$\hat{f} = \frac{\hat{\omega}}{2\pi} = fT_s = f/f_s.$$

Since we are now dealing with a function on a finite interval we can represent it as a Fourier series:

$$F(\omega) = \sum_{n=-\infty}^{\infty} \phi_n e^{i\omega n/2f_s} \tag{4.5.2}$$

where the Fourier coefficients ϕ_n are to be determined by

$$\phi_n = \frac{1}{4\pi f_s} \int_{-2\pi f_s}^{2\pi f_s} F(\omega) e^{-i\omega n/2f_s} \, d\omega. \tag{4.5.3}$$

[2]In fact the assumption that the interval is symmetric about the origin is made without loss of generality, since we can always introduce a change of variables which maps an arbitrary interval into a symmetric one centered on 0.

Comparing this result with our previous work we can see that

$$\phi_n = \frac{f(n/2f_s)}{2f_s} \tag{4.5.4}$$

where $f(n/2f_s)$ are the samples of the original continuous time series $f(t)$. Putting all this together, one can show that the band limited function $f(t)$ is completely specified by its values at the countable set of points spaced $1/2f_s$ apart:

$$\begin{aligned} f(t) &= \frac{1}{4\pi f_s} \sum_{n=-\infty}^{\infty} f(n/2f_s) \int_{-2\pi f_s}^{2\pi f_s} e^{i(\omega n/2f_s - \omega t)}\, d\omega \\ &= \sum_{n=-\infty}^{\infty} f(n/2f_s) \frac{\sin(\pi(2f_s t - n))}{\pi(2f_s t - n)}. \end{aligned} \tag{4.5.5}$$

The last equation is known as the **Sampling Theorem**. Notice that the function $\sin x/x$ appears here too. Since this function appears frequently it is given a special name, it is called the sinc function:

$$\text{sinc}(x) = \frac{sinx}{x}.$$

And we know that the sinc function is also the Fourier transform of a box-shaped function. So the sampling theorem says take the value of the function, sampled every $1/2f_s$, multiply it by a sinc function centered on that point, and then sum these up for all the samples.

It is worth repeating for emphasis: any band limited function is completely determined by its samples chosen $1/2f_s$ apart, where f_s is the maximum frequency contained in the signal. This means that in particular, a time series of finite duration (i.e., any real time series) is completely specified by a finite number of samples. It also means that in a sense, the information content of a band limited signal is infinitely smaller than that of a general continuous function. So if our band-limited signal $f(t)$ has a maximum frequency of f_s hertz, and the length of the signal is T, then the total number of samples required to describe f is $2f_sT$.

A sampling exercise

Consider the continuous sinusoidal signal:

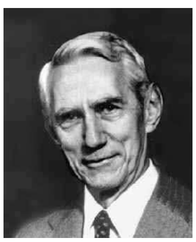
$$x(t) = A\cos(2\pi f t + \phi)$$

Suppose we sample this signal at a sampling period of T_s. Let us denote the discrete samples of the signal with square brackets:

$$x[n] \equiv x(nT_s) = A\cos(2\pi f n T_s + \phi).$$

Now consider a different sinusoid of the same amplitude and phase, but sampled at a frequency of $f + \ell f_s$, where ℓ is an integer and $f_s = 1/T_s$. Let the samples of this second sinusoid be denoted by $y[n]$. Show that $x[n] = y[n]$. This is an example of *aliasing*. These two sinusoids have exactly the same samples, so the frequency of one appears to be the same.

The sampling theorem is due to Harry Nyquist, a researcher at Bell Labs in New Jersey. In a 1928 paper Nyquist laid the foundations for the sampling of continuous signals and set forth the sampling theorem. Nyquist was born on February 7, 1889 in Nilsby, Sweden and emigrated to the US in 1907. He got his PhD in Physics from Yale in 1917. Much of Nyquist's work in the 1920's was inspired by the telegraph. In addition to his work in sampling, Nyquist also made an important theoretical analysis of thermal noise in electrical systems. In fact this sort of noise is sometimes called Nyquist noise. Nyquist died on April 4, 1976 in Harlingen, Texas.



A generation after Nyquist's pioneering work Claude Shannon, also at Bell Labs, laid the broad foundations of modern communication theory and signal processing. Shannon (Born: April 1916 in Gaylord, Michigan; Died: Feb 2001 in Medford, Massachusetts) was the founder of modern information theory. After beginning his studies in electrical engineering, Shannon took his PhD in mathematics from MIT in 1940. Shannon's *A Mathematical Theory of Communication* published in 1948 in the Bell System Technical Journal, is one of the profoundly influential scientific works of the 20th century. In it he introduced many ideas that became the basis for electronic communication, such as breaking down information into sequences of 0's and 1's (this is where the term *bit* first appeared), adding extra bits to automatically correct for errors and measuring the information or variability of signals. Shannon's paper and many other influential papers on communication are compiled in the book *Key papers in the development of information theory* [?].

4.5.1 Aliasing

As we have seen, if a time-dependent function contains frequencies up to f_s hertz, then discrete samples taken at an interval of $1/2f_s$ seconds completely determine the signal. Looked at from another point of view, for any sampling interval Δ, there is a special frequency (called the Nyquist frequency), given by $f_s = \frac{1}{2\Delta}$. The extrema (peaks and troughs) of a sinusoid of frequency f_s will lie exactly $1/2f_s$ apart. This is equivalent to saying that critical sampling of a sine wave is 2 samples per wavelength.

We can sample at a finer interval without introducing any error; the samples will be

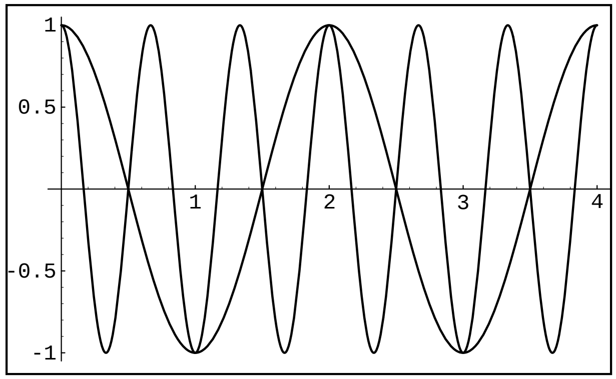


Figure 4.9: A sinusoid sampled at less than the Nyquist frequency gives rise to spurious periodicities.

redundant, of course. However, if we sample at a coarser interval a very serious kind of error is introduced called aliasing. Figure 4.9 shows a cosine function sampled at an interval longer than $1/2f_s$; this sampling produces an apparent frequency of 1/3 the true frequency. This means that any frequency component in the signal lying outside the interval $(-f_s, f_s)$ will be spuriously shifted into this interval. Aliasing is produced by under-sampling the data: once that happens there is little that can be done to correct the problem. The way to prevent aliasing is to know the true band-width of the signal (or band-limit the signal by analog filtering) and then sample appropriately so as to give at least 2 samples per cycle at the highest frequency present.

4.6 The Discrete Fourier Transform

Now we consider the third major use of the Fourier superposition. Suppose we have discrete data, not a continuous function. In particular, suppose we have data f_k recorded at locations x_k. To keep life simple, let us suppose that the data are recorded at N evenly spaced locations $x_k = 2\pi k/N$, $k = 0, 1, \ldots N-1$. Think of f_k as being samples of an unknown function, which we want to approximate. Now we write down a Fourier approximation for the unknown function (i.e., a Fourier series with coefficients to be determined):

$$p(x) = \sum_{n=0}^{N-1} c_n e^{inx}. \tag{4.6.1}$$

Now we will compute the coefficients in such a way that p interpolates (i.e., fits exactly) the data at each x_k:

$$f_k = p(x_k) = \sum_{n=0}^{N-1} c_n e^{in2\pi k/N}. \tag{4.6.2}$$

In theory we could do this for any linearly independent set of basis functions by solving a linear system of equations for the coefficients. But since sines/cosines are orthogonal, the c_n coefficients can be computed directly:

$$c_k = \frac{1}{N} \sum_{n=0}^{N-1} f_n e^{-in2\pi k/N}. \tag{4.6.3}$$

This is the discrete version of the Fourier transform (DFT). f_n are the data and c_k are the harmonic coefficients of a trigonometric function that interpolates the data. Now, of course, there are many ways to interpolate data, but it is a theorem that the only way to interpolate with powers of $e^{i2\pi x}$ is Equation 4.6.3.

Optional Exercise In the handout you will see some Mathematica code for computing and displaying discrete Fourier transforms. Implement the previous formula and compare the results with Mathematica's built in *Fourier* function. You should get the same result, but it will take dramatically longer than Mathematica would for 100 data points. The reason is that Mathematica uses a special algorithm called the FFT (Fast Fourier Transform). See Strang for an extremely clear derivation of the FFT algorithm.

4.7 The Linear Algebra of the DFT

Take a close look at Equation 4.6.3. Think of the DFT coefficients c_k and the data points f_n as being elements of vectors $\mathbf{c}$ and $\mathbf{f}$. There are N coefficients and N data so both $\mathbf{c}$ and $\mathbf{f}$ are elements of R^N. The summation in the Fourier interpolation is therefore a matrix-vector inner product. Let's identify the coefficients of the matrix. Define a matrix Q such that

$$Q_{nk} = e^{in2\pi k/N}. \tag{4.7.1}$$

N is fixed, that's just the number of data points. The matrix appearing in Equation 4.6.3 is the complex conjugate of Q; i.e., Q^*. We can write Equation 4.6.2 as

$$\mathbf{f} = Q \cdot \mathbf{c}. \tag{4.7.2}$$

The matrix Q is almost orthogonal. We have said that a matrix A is orthogonal if $AA^T = A^T A = I$, where I is the N-dimensional identity matrix. For complex matrices we need to generalize this definition slightly; for complex matrices we will say that A is orthogonal if $(A^T)^*A = A(A^T)^* = I$.[3] In our case, since Q is obviously symmetric, we have:

$$Q^*Q = QQ^* = I. \tag{4.7.3}$$

Once again, orthogonality saves us from having to solve a linear system of equations: since $Q^* = Q^{-1}$, we have

$$\mathbf{c} = Q^* \cdot \mathbf{f}. \tag{4.7.4}$$

Now you may well ask: what happens if we use fewer Fourier coefficients than we have data? That corresponds to having fewer unknowns (the coefficients) than data. So you wouldn't expect there to be an exact solution as we found with the DFT, but how about a least squares solution? Let's try getting an approximation function of the form

$$p(x) = \sum_{n=0}^{m} c_n e^{inx} \tag{4.7.5}$$

where now we sum only up to $m < N - 1$. Our N equations in m unknowns is now:

$$f_k = \sum_{n=0}^{m} c_n e^{in2\pi k/N}. \tag{4.7.6}$$

So to minimize the mean square error we set the derivative of

$$||\mathbf{f} - Q \cdot \mathbf{c}||^2 \tag{4.7.7}$$

with respect to an arbitrary coefficient, say c_j, equal to zero. But this is just an ordinary least squares problem.

4.8 The DFT from the Fourier Integral

In this section we will use the f (cycles per second) notation rather than the ω (radians per second), because there are slightly fewer factors of 2π floating around. You should be comfortable with both styles, but mind those 2πs! Also, up to now, we have avoided any special notation for the Fourier transform of a function, simply observing whether it was a function of space-time or wavenumber-frequency. Now that we are considering discrete transforms and real data, we need to make this distinction since we will generally have both the sampled data and its transform stored in arrays on the computer. So for this section we will follow the convention that if $h = h(t)$ then $H = H(f)$ is its Fourier transform.

[3] Technically such a matrix is called Hermitian or self-adjoint–the operation of taking the complex conjugate transpose being known at the adjoint–but we needn't bother with this distinction here.

We suppose that our data are samples of a function and that the samples are taken at equal intervals, so that we can write

$$h_k \equiv h(t_k), \qquad t_k \equiv k\Delta, \qquad k = 0, 1, 2, \ldots, N-1, \tag{4.8.1}$$

where N is an even number. In our case, the underlying function $h(t)$ is unknown; all we have are the digitally recorded time series. But in either case we can estimate the Fourier transform $H(f)$ at at most N discrete points chosen in the range $-f_s$ to f_s where f_s is the Nyquist frequency:[4]

$$f_n \equiv \frac{n}{\Delta N}, \qquad n = \frac{-N}{2}, \ldots, \frac{N}{2}. \tag{4.8.2}$$

The two extreme values of frequency $f_{-N/2}$ and $f_{-N/2}$ are not independent ($f_{-N/2} = -f_{N/2}$), so there are actually only N independent frequencies specified above.

A sensible numerical approximation for the Fourier transform integral is thus:

$$H(f_n) = \int_{-\infty}^{\infty} h(t) e^{-2\pi i f_n t}\, dt \approx \sum_{k=0}^{N-1} h_k e^{-2\pi i f_n t_k} \Delta. \tag{4.8.3}$$

Therefore

$$H(f_n) \approx \Delta \sum_{k=0}^{N-1} h_k e^{-2\pi i k n/N}. \tag{4.8.4}$$

Defining the *Discrete Fourier Transform* (DFT) by

$$H_n = \sum_{k=0}^{N-1} h_k e^{-2\pi i k n/N} \tag{4.8.5}$$

we then have

$$H(f_n) \approx \Delta H_n \tag{4.8.6}$$

where f_n are given by Equation (4.8.2).

Now, the numbering convention implied by Equation (4.8.2) has $\pm$ Nyquist at the extreme ends of the range and zero frequency in the middle. However it is clear that the DFT is periodic with period N:

$$H_{-n} = H_{N-n}. \tag{4.8.7}$$

As a result, it is standard practice to let the index n in H_n vary from 0 to $N-1$, with n and k varying over the same range. In this convention 0 frequency occurs at

[4]The highest frequency f_s in the Fourier representation of a time series sampled at a time interval of Δ is $\frac{1}{2\Delta}$. This maximum frequency is called the Nyquist frequency. You'll study this in detail in the digital course.

$n = 0$; positive frequencies from from $1 \le n \le N/2 - 1$; negative frequencies run from $N/2 + 1 \le n \le N - 1$. Nyquist sits in the middle at $n = N/2$. The inverse transform is:

$$h_k = \frac{1}{N} \sum_{n=0}^{N-1} H_n e^{2\pi i k n/N} \tag{4.8.8}$$

Mathematica, on the other hand, uses different conventions. It uses the symmetric normalization ($1/\sqrt{N}$ in front of both the forward and inverse transform), and defines arrays running from 1 to N in Fortran fashion. So in *Mathematica*, the forward and inverse transforms are, respectively:

$$H_n = \frac{1}{\sqrt{N}} \sum_{k=1}^{N} h_k e^{-2\pi i (k-1)(n-1)/N} \tag{4.8.9}$$

and

$$h_k = \frac{1}{\sqrt{N}} \sum_{n=1}^{N} H_n e^{2\pi i (k-1)(n-1)/N}. \tag{4.8.10}$$

If you are using canned software, make sure you know what conventions are being used.

4.8.1 Discrete Fourier Transform Examples

Here we show a few examples of the use of the DFT. What we will do is construct an unknown time series' DFT by hand and inverse transform to see what the resulting time series looks like. In all cases the time series h_k is 64 samples long. Figures 4.10 and 4.11 show the real (left) and imaginary (right) parts of six time series that resulted from inverse DFT'ing an array H_n which was zero except at a single point (i.e., it's a Kronecker delta: $H_i = \delta_{i,j} = 1$ if $i = j$ and zero otherwise; here a different j is chosen for each plot). Starting from the top and working down, we choose j to be the following samples: the first, the second, Nyquist-1, Nyquist, Nyquist+1, the last. We can see that the first sample in frequency domain is associated with the zero-frequency or DC component of a signal and that the frequency increases until we reach Nyquist, which is in the middle of the array. Next, in Figure 4.12, we show at the top an input time series consisting of a pure sinusoid (left) and the real part of its DFT. Next we add some random noise to this signal. On the left in the middle plot is the real part of the noisy signals DFT. Finally, at the bottom, we show a Gaussian which we convolve with the noisy signal in order to attenuate the frequency components in the signal. The real part of the inverse DFT of this convolved signal is shown in the lower right plot.

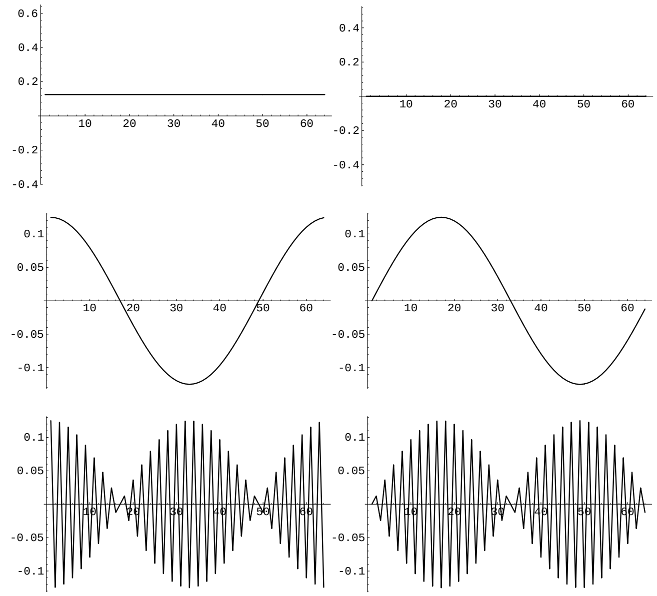


Figure 4.10: The real (left) and imaginary (right) parts of three length 64 time series, each associated with a Kronecker delta frequency spectrum. These time series are reconstructed from the spectra by inverse DFT. At the top the input spectrum is $\delta_{i,0}$, in the middle $\delta_{i,1}$, and at the bottom, $\delta_{i,64/2-1}$.

4.9 Convergence Theorems

One has to be a little careful about saying that a particular function is equal to its Fourier series since there exist piecewise continuous functions whose Fourier series diverge everywhere! However, here are two basic results about the convergence of such series.

Point-wise Convergence Theorem: If f is piecewise continuous and has left and right derivatives at a point c[5] then the Fourier series for f converges converges to

$$\frac{1}{2}\left(f(c-) + f(c+)\right) \tag{4.9.1}$$

where the + and - denote the limits when approached from greater than or less than c.

Another basic result is the **Uniform Convergence Theorem**: If f is continuous with period 2π and f' is piecewise continuous, then the Fourier series for f converges uniformly to f. For more details, consult a book on analysis such as *The Elements of Real Analysis* by Bartle [1] or *Real Analysis* by Haaser and Sullivan [**?**].

[5] A right derivative would be: $\lim_{t\to 0}(f(c+t) - f(c))/t$, $t > 0$. Similarly for a left derivative.

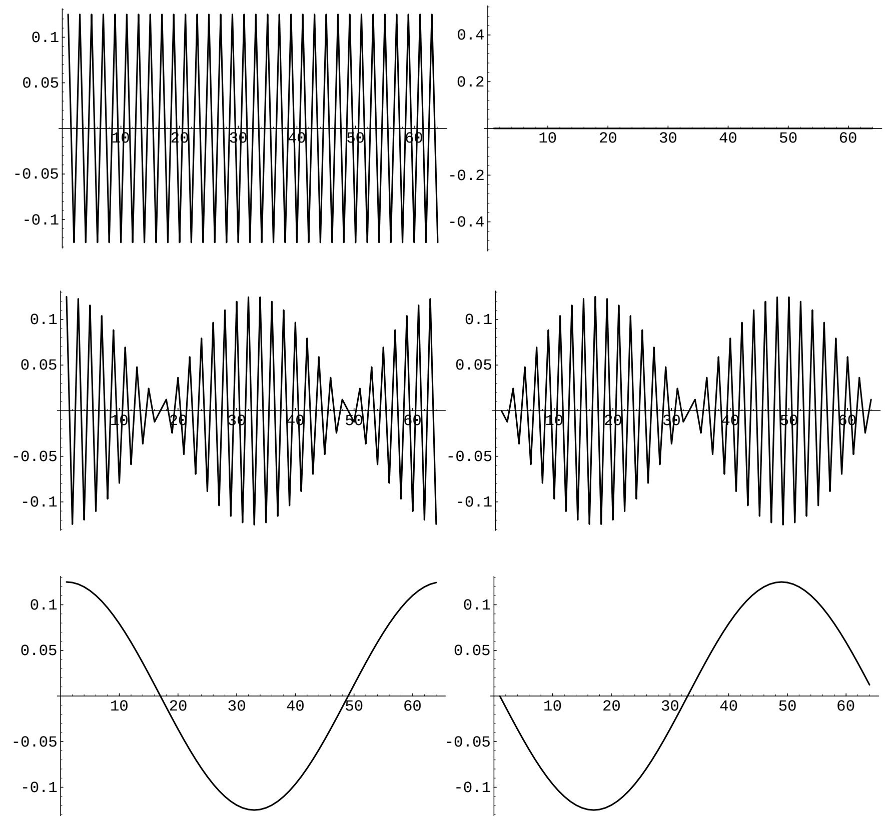


Figure 4.11: The real (left) and imaginary (right) parts of three time series of length 64, each associated with a Kronecker delta frequency spectrum. These time series are reconstructed from the spectra by inverse DFT. At the top the input spectrum is $\delta_{i,64/2}$, in the middle $\delta_{i,64/2+1}$, and at the bottom $\delta_{i,64}$.

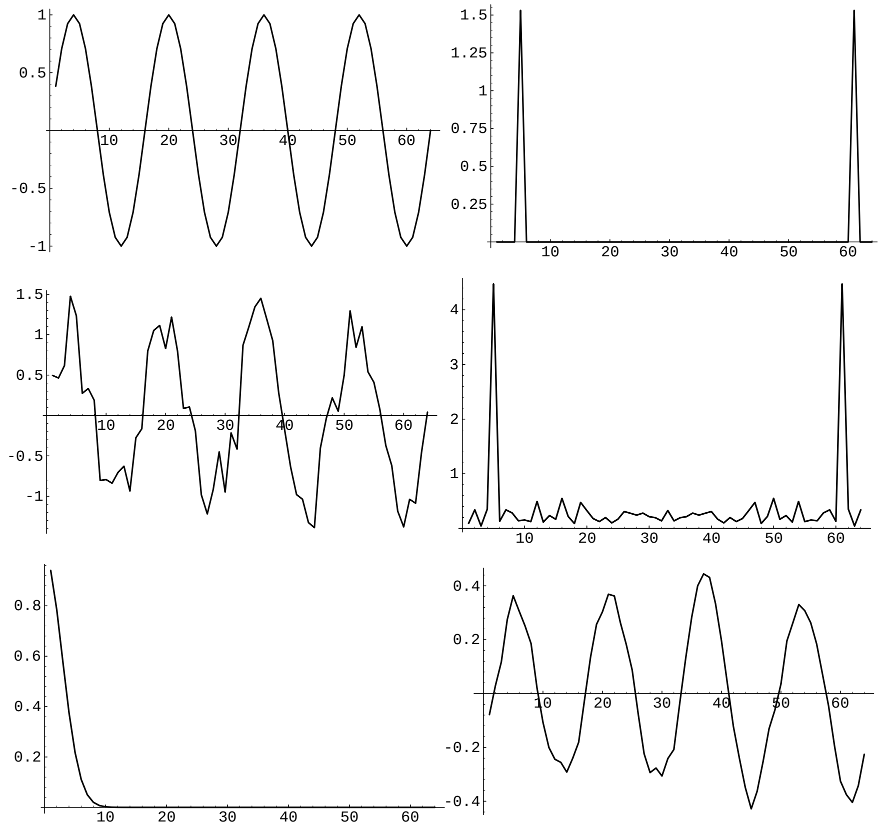


Figure 4.12: The top left plot shows an input time series consisting of a single sinusoid. In the top right we see the real part of its DFT. Note well the wrap-around at negative frequencies. In the middle we show the same input sinusoid contaminated with some uniformly distributed pseudo-random noise and its DFT. At the bottom left, we show a Gaussian time series that we will use to smooth the noisy time series by convolving it with the DFT of the noisy signal. When we inverse DFT to get back into the "time" domain we get the smoothed signal shown in the lower right.

4.10 Basic Properties of Delta Functions

Another representation of the delta function is in terms of Gaussian functions:

$$\delta(x) = \lim_{\mu \to \infty} \frac{\mu}{\sqrt{\pi}} e^{-\mu^2 x^2}. \tag{4.10.1}$$

You can verify for yourself that the area under any of the Gaussian curves associated with finite μ is one.

The spectrum of a delta function is completely flat since

$$\int_{-\infty}^{\infty} e^{-ikx} \delta(x)\ dx = 1. \tag{4.10.2}$$

For delta functions in higher dimensions we need to add an extra $1/2\pi$ normalization for each dimension. Thus

$$\delta(x, y, z) = \left(\frac{1}{2\pi}\right)^3 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{i(k_x x + k_y y + k_z z)}\ dk_x\ dk_y\ dk_z. \tag{4.10.3}$$

The other main properties of delta functions are the following:

$$\delta(x) = \delta(-x) \tag{4.10.4}$$

$$\delta(ax) = \frac{1}{|a|}\delta(x) \tag{4.10.5}$$

$$x\delta(x) = 0 \tag{4.10.6}$$

$$f(x)\delta(x-a) = f(a)\delta(x-a) \tag{4.10.7}$$

$$\int \delta(x-y)\delta(y-a)\ dy = \delta(x-a) \tag{4.10.8}$$

$$\int_{-\infty}^{\infty} \delta^{(m)} f(x)\ dx = (-1)^m f^{(m)}(0) \tag{4.10.9}$$

$$\int \delta'(x-y)\delta(y-a)\ dy = \delta'(x-a) \tag{4.10.10}$$

$$x\delta'(x) = -\delta(x) \tag{4.10.11}$$

$$\delta(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{ikx}\ dk \tag{4.10.12}$$

$$\delta'(x) = \frac{i}{2\pi}\int_{-\infty}^{\infty} k e^{ikx}\ dk \tag{4.10.13}$$

Exercises

4.1 Prove Equations 4.2.4 and 4.2.5.

4.2 Compute the Fourier transform of the following function. $f(x)$ is equal to 0 for $x < 0$, x for $0 \leq x \leq 1$ and 0 for $x > 1$.

4.3 Prove Equations 4.4.18, 4.4.19, 4.4.20, 4.4.22.

4.4 Compute the Fourier transform of $f(x) = e^{-x^2/a^2}$. If a is small, this bell-shaped curve is sharply peaked about the origin. If a is large, it is broad. What can you say about the Fourier transform of f in these two cases?

4.5 Let $f(x)$ be the function which is equal to -1 for $x < 0$ and +1 for $x > 0$. Assuming that

$$f(x) = \frac{a_0}{2} + \sum_{k=1}^{\infty} a_k \cos(k\pi x/l) + \sum_{k=1}^{\infty} b_k \sin(k\pi x/l),$$

compute a_0, a_1, a_2, b_1 and b_2 by hand, taking the interval of periodicity to be $[-1, 1]$.

4.6 For an odd function, only the sine or cosine terms appear in the Fourier series. Which is it?

4.7 Consider the complex exponential form of the Fourier series of a real function

$$f(x) = \sum_{n=-\infty}^{\infty} c_n e^{in\pi x/l}.$$

Take the complex conjugate of both sides. Then use the fact that since f is real, it equals its complex conjugate. What does this tell you about the coefficients c_n?

.