Fun with the Fourier Transform

Kolja Kähler (kkaehler@acm.org)

December 29, 2009

Introduction

Recently, I spent some time revisiting the Fourier transform, especially focusing on the connection between the continuous and discrete formulations. A wide range of reading material is available in print and on the Internet on this topic, so this text does not represent another introduction or even full treatment of the Fourier transform. Instead, it is a highly subjective selection of sometimes rather basic, at other times more advanced points that are often omitted, or too easily dismissed in textbook explanations — perhaps because most of the "roadblocks" quickly evaporate once you overcome the difficulties, and after that it's hard to imagine what the initial problem was.

If you already have a basic understanding of what the Fourier transform is, but want to know some of the fine print, I hope these notes can provide insight and pointers for further study.

The Continuous Fourier Transform

This first half of this text deals with the Fourier transform in the form

$$F(s) = \int_{-\infty}^{\infty} f(x) e^{-i2\pi sx} dx,$$

and its inverse

$$f(x) = \int_{-\infty}^{\infty} F(s) e^{i2\pi sx} ds.$$

Before talking about the transform itself, I would like to cover the one thing that might look a bit strange, if you've never seen this before: the $e^{i\cdots}$ term. What on earth does e^{ix} mean? How can you raise a real number to a complex exponent? The typical answer is "it is the same as $\cos(x) + i\sin(x)$". We could just live with that, but is there an explanation?

Sidetrack: The Complex Exponential Function

The story dates back to Leonard Euler. In his work *Introductio in analysin infinitorum*, published in 1748, he defined the function e^x by the limit:

$$e^x = \lim_{n\to\infty} \left(1 + \frac{x}{n}\right)^n,$$

and developed it into an infinite power series:

$$\lim_{n\to\infty}\left(1+\frac{x}{n}\right)^n = 1+\frac{x}{1!}+\frac{x^2}{2!}+\frac{x^3}{3!}+\ldots.$$

Euler recognized that, with x now not being an exponent anymore, it could be replaced by a complex number ix. What does the limit function then converge to? Using the definition $i=\sqrt{-1}$, we can rewrite the above as

$$\begin{aligned}
e^{ix} &= \lim_{n\to\infty}\left(1+\frac{ix}{n}\right)^n && (1)\\
&= 1+\frac{ix}{1!}+\frac{(ix)^2}{2!}+\frac{(ix)^3}{3!}+\ldots && (2)\\
&= 1+\frac{ix}{1!}-\frac{x^2}{2!}-\frac{ix^3}{3!}+\frac{x^4}{4!}+\ldots && (3)\\
& && (4)
\end{aligned}$$

By reordering the terms (ignoring the dangers of doing this on an infinite sum), real and imaginary terms are collected:

$$e^{ix} = \left(1-\frac{x^2}{2!}+\frac{x^4}{4!}-\frac{x^6}{6!}+\ldots\right)+i\left(x-\frac{x^3}{3!}+\frac{x^5}{5!}-\frac{x^7}{7!}\ldots\right).$$

This — as Euler already knew — is the sum of the power series of the cosine and sine functions, respectively. Thus, we finally arrive at the expression

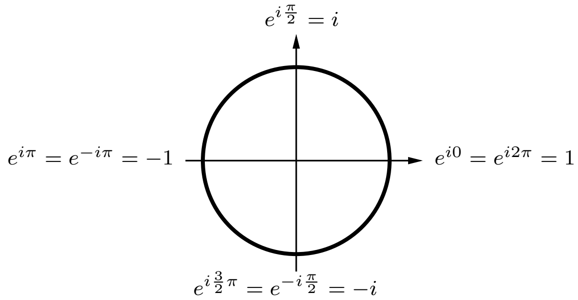
$$e^{ix} = \cos(x)+i\sin(x).$$

Since the cosine function is even, and sine is odd, e^{-ix} can be written

$$e^{-ix} = \cos(-x)+i\sin(-x) = \cos(x)-i\sin(x).$$

Interpretation and interesting Facts about e

The graph of e^{ix} describes a perfect unit circle in the complex plane, with a period of 2π in x:



The e notation is very convenient for describing a signal in terms of phase and amplitude, which is why it appears so frequently in electrical engineering texts. Any complex number $z = x+iy$ can be written in this form by multiplying in a radius $r = |z| = \sqrt{x^2+y^2}$.

With $\sin\phi = \frac{y}{r}$ and $\cos\phi = \frac{x}{r}$, we have

$$z = r(\cos\phi + i\sin\phi) = re^{i\phi},$$

The most beautiful result though, one of the most famous equations of all times, is this:

$$e^{i\pi} + 1 = 0.$$

I suspect you can buy T–shirts and coffee mugs imprinted with this equation somewhere, but, admittedly, I haven't checked.

Common Variations on the Fourier Transform Definition

The definition of the Fourier transform I gave in the beginning is frequently subtly altered by scaling coefficients added to the forward transform, the inverse transform, or both. Where do they come from?

These variations are merely different ways of scaling the frequency parameter. The two most common "systems" (forward and inverse transform pairs) are:

System I

$$F(s) = \int_{-\infty}^{\infty} f(x)e^{-i2\pi sx}dx \Longleftrightarrow f(x) = \int_{-\infty}^{\infty} F(s)e^{i2\pi sx}ds$$

Here, the units of the frequency parameter s can be understood as "cycles of the e function per unit of x". A signal repeating every 0.5 units corresponds to a frequency value $s = 2$.

System II

$$F(\omega) = \int_{-\infty}^{\infty} f(x)e^{-i\omega x}dx \Longleftrightarrow f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} F(\omega)e^{i\omega x}d\omega$$

This form is common when angular frequencies ω are used. ω can be read as "cycles per 2π units of x". A signal repeating every π units corresponds to a frequency value $\omega = 2$. Why the asymmetry between forward and inverse transforms? Compare s in the previous formulation to the frequency parameter $\omega = 2\pi s$: The graph of $F(\omega)$ is 2π times wider than the graph of $F(s)$. As a consequence, $\int F$ also becomes 2π times larger. To recover the original function f through the inverse transform, this scaling effect must be undone.

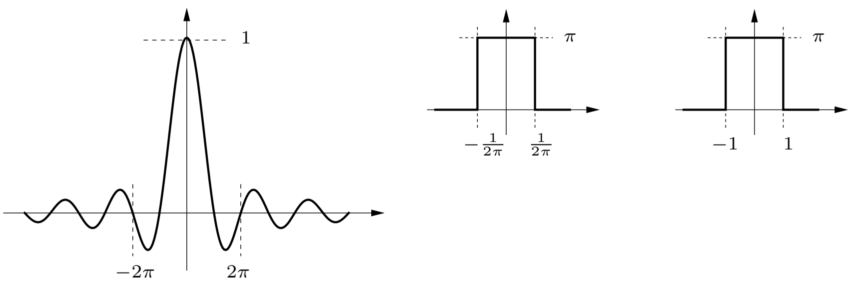
The Graph of the Transform

The plot of the Fourier transform of a function (the "spectrum" of the function) shows the frequencies the function is composed of. Questions that immediately come to mind, are:

- What do the function values mean?

- The spectrum shows frequencies — but what happens on the left side of the graph? What is a "negative frequency"?
- The plot seems to be symmetrical — is this always the case?

Let's look at a simple example of a Fourier transform plot. In the figure below, the sinc function $sinc(x) = \frac{sin(x)}{x}$ is shown on the left. Its transform is the rectangle function, shown for "system I" in the middle, and for "system II" on the right side. Note the different scaling by a factor of 2π, reflecting the explanations in the previous section. Both the original function and the transform are real (the values have no imaginary part).



The plots are read as follows: The $sinc$ function has a maximum frequency of $1/(2\pi)$ "cycles per unit of x" in system I, corresponding to the period of the sine function. The value $F(0) = \pi$ corresponds to the integral of the original function. In "system II", the frequency variable s is read as "1 cycle per 2π units of x'.

To answer the questions about the symmetry of the graph and these curious negative frequencies, we need to take a closer look at even and odd functions, and how these concepts apply to the Fourier transform.

Even and odd

Any function $f(x)$ can be expressed as the sum of an odd function and an even function: $f(x) = e(x) + o(x)$. Let's rewrite the Fourier transform using this notation:

$$\begin{aligned} F(s) &= \int_{-\infty}^{\infty} (e(x) + o(x))e^{-i2\pi sx}dx \\ &= \int_{-\infty}^{\infty} (e(x) + o(x))(\cos(2\pi sx) - i\sin(2\pi sx))dx \\ &= \int_{-\infty}^{\infty} (e(x)\cos(2\pi sx) - e(x)i\sin(2\pi sx) + \\ & \quad o(x)\cos(2\pi sx) - o(x)i\sin(2\pi sx))dx \end{aligned}$$

The transform is a sum of sine and cosine functions. Since the sine function is odd ($\sin(x) = -\sin(-x)$), the parts of $\int e(x)i\sin(2\pi sx)dx$ for negative and positive x cancel each other out, while the values for $x < 0$ in $\int o(x)\sin(2\pi sx)dx$ are the same as those for $x > 0$. The opposite is true for the cosine part — the cosine function is even: $\cos(x) = \cos(-x)$. Thus, the Fourier transform reduces to:

$$F(s) = 2\int_{0}^{\infty} e(x)\cos(2\pi sx)dx - 2i\int_{0}^{\infty} o(x)\sin(2\pi sx)dx.$$

This clearly shows how the cosine terms are associated with the real part of the transformed values, while the sine terms are imaginary. What does this tell us?

Symmetry

In most practical applications we deal with real–valued input functions. If $f(x)$ is a real function, $e(x)$ and $o(x)$ are real, too. Thus the even terms of the transform are all real, while the odd terms are imaginary. Consequently, the real part of the Fourier transform is even, and the imaginary part is odd (this property is called hermitian). If the real part of the spectrum is graphed, it is indeed mirrored at the y axis. If f is not only real, but also even, the transform is real and even, too.

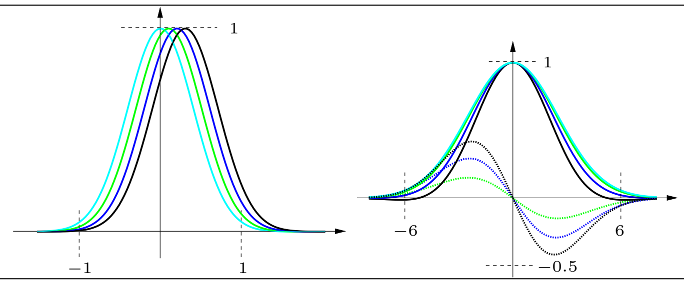
Negative Frequencies

The intuitive notion of a frequency is repetition in time or space, at a certain rate. A signal repeating every n units has a frequency of $1/n$. But what is a *negative* frequency supposed to mean?

Lets start by looking at the equation for the transform. We see that for $s < 0$, the sign of the exponent of the exponential changes. Otherwise, moving up the positive s axis is the same as moving down along negative s. The inverted sign causes $\cos(x) - i\sin(x)$ to turn into $\cos(x) + i\sin(x)$: the function $f(x)$ is analyzed for the same frequencies (as in "cycles per unit"), but with *inverted phase*.

The negative frequencies are no "additional" frequencies that exist independently of the positive ones; there's a tight connection between the positive and negative halves of the Fourier spectrum. The best way to get an intuitive grasp of the connection is again to think about even and odd functions, extending the thoughts from the previous section.

For a real function $f(x)$, we already found the real part of the Fourier transform to always be even. If f is real *and even*, the Fourier transform is purely real, with no imaginary part (the $i\sin$ terms cancel out). If, in contrast, f is real *and odd*, its transform is purely imaginary, and also odd. Now imagine a perfectly even function f, and its (purely real) Fourier spectrum. Step by step, shift f slightly along x, so it becomes "more odd". The Fourier transform will change from purely real to complex values, with increasing imaginary parts. The following figure shows shifted versions of the function $e^{-\pi x^2}$ and their transforms. On the right side, real and imaginary parts of the transform are depicted as solid and dashed lines, respectively.



Instead of separating the complex values into real and imaginary graphs, it is common to graph *magnitudes* and *phases* — the graph of absolute values of a shifted, transformed function remains the same. If you graph the real values of the transform pointing up, and the imaginary values out of the screen, the shift really causes a "twist" of the complex values around the s axis.

If the negative half of the Fourier transform is nothing more than a mirror copy of the positive half, do we actually need to compute it? For real–valued functions $f(x)$, the anwser is indeed: no, we don't. If we have $F(s)$ for $s > 0$, $F(-s)$ is simply the conjugate complex $\overline{F(s)} = Re(F(s)) - Im(F(s))$. This is not true for *complex* functions f, however! There always is a symmetry relation between the real and imaginary parts (with any combination of even- and oddness), but without further knowledge about the original function, you cannot reconstruct it from one half of the spectrum alone.

The Discrete Fourier Transform

For numerical applications of the Fourier transform, we must *discretize* the input function, and we also cannot use a continuous spectrum of frequencies for numerical purposes. Hence, the standard formulation of the discrete Fourier transform, or DFT, differs somewhat from the continuous formulation to take the discretiziation into account. The DFT works on an array of input data, i.e., samples from a finite range of values of an assumed underlying continuous function. These are transformed into an array representing the discrete spectrum of the data. There are many ways to introduce the DFT. The following explanations show how the discrete formulation can be derived from the continuous setting by applying the discretization steps (note the plural: there are actually *two* discretization processes!).

Discretize, and then again

Starting with the continuous formulation, we approximate the integration by an infinite sum, where dx turns into the sampling interval Δx:

$$F_s(s) = \sum_{n=-\infty}^{\infty} f(n\Delta x) e^{-i2\pi s n \Delta x} \Delta x,$$

We cut out a finite window of f, and also scale f so that the sampling interval becomes $\Delta x = 1$, thus eliminating Δx from the equation.

$$F_w(s) = \sum_{n=0}^{N-1} f(n) e^{-i2\pi s n}.$$

The function $F_w(s)$ is periodic with a period of 1: The exponential term reaches the same value for $s, s+1, s+2, \ldots$. *This is a consequence from sampling, and a notable difference from the continuous transform, where the frequency range is infinite.* But it is still a continuous function (it could be evaluated for arbitrary frequency values), so we need to sample the *spectrum*, too. It can be shown that there are only N independent frequency values s. (I'm not going into the derivation here, but it seems intuitively clear that for a signal with N degrees of freedom, there is only an equal amount of degrees of freedom in the frequency domain). We thus evaluate $F_w(s)$ at N equidistant points in the range $[0, \ldots, \frac{N-1}{N}]$:

$$F_w(\frac{k}{N}) = \sum_{n=0}^{N-1} f(n) e^{-i2\pi n \frac{k}{N}}, \qquad k \in [0, 1, 2, \ldots, N-1].$$

We now have N different frequency values, and N points where f is evaluated. We can simplify the notation some more by leaving out the factor $1/N$ from the argument of F_w:

$$F[k] = \sum_{n=0}^{N-1} f[n] e^{-i2\pi n \frac{k}{N}}.$$

This is the standard form of the DFT. I've used brackets here to emphasize the array–like discretization on integer values. The scaling of the frequencies by the factor N above must be corrected for in the inverse transform, to recover the original function values:

$$f[n] = \frac{1}{N} \sum_{k=0}^{N-1} F[k] e^{i2\pi n \frac{k}{N}}, \qquad n \in [0, 1, 2, \ldots, N-1].$$

This is essentially the same as the scaling for "system II" introduced earlier. The factor $1/N$ is sometimes already added to the forward transform. For interpretation of the values obtained by a numerical FFT package, one should thus check how the algorithm scales the values. In the form shown here, the frequency value 0 is equivalent to the *sum* of the signal coefficients (corresponding to the integral of the signal in the continuous case), in the other form it would be the *average* signal value.

Things to keep in Mind

The discretization processes (sampling the signal, and sampling the spectrum) are responsible for the different "look and feel" of the DFT compared to the continuous transform. Understanding the differences is important for correct usage of the DFT as a tool.

Spectrum and Reconstruction are periodic

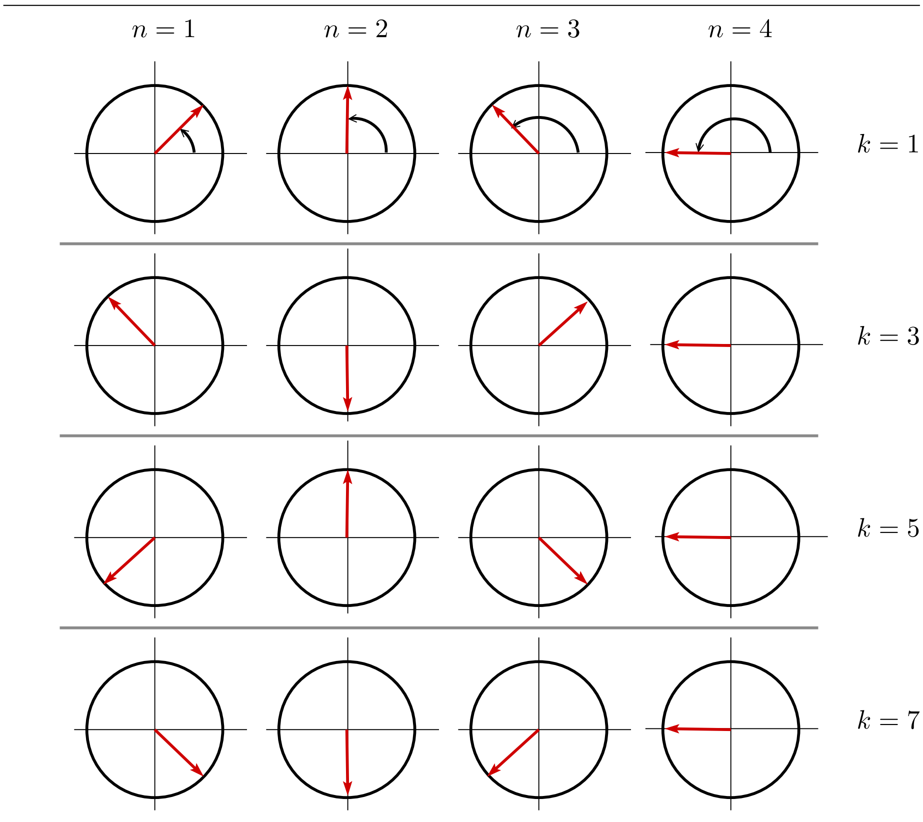
Both the forward and the inverse transforms are periodic: when the *signal* is sampled, we get a periodic, continuous spectrum. When the *spectrum* is sampled, we cause the reconstructed function to become periodic.

The DFT Spectrum wraps around

What in the continuous transform shows up as negative frequencies, is mapped to the second half of the positive frequency range in the DFT. The spectrum appears to be shifted and wrapped around at $k = N/2$. This effect can also be attributed to discrete sampling and the periodicity of the e function. An intuitive way to think about this, is to picture the e function in the DFT as a wheel with N spokes, one spoke per value of n. Spoke 0 corresponds to e^0, spoke $N/4$ is $e^{-i0.5\pi}$, $N/2$ is $e^{-i\pi}$, etc. At frequency $k = 1$, the samples of f are multiplied with spoke 0,1,.... At $k = 2$, only spokes $0, 2, \ldots$ are taken into the sum. Finally, At $k = N/2$, the e term in the sum alternates between the values e^0 and $e^{-i\pi}$ (spokes 0 and $N/2$). This is the highest positive frequency. When the value of k is increased above $N/2$, each iteration in the sum moves over more than half of the spokes. *This is equivalent to "moving around the wheel" in the other direction.* At $k = N - 1$, $N - 1$ spokes are skipped for each n in the sum. This is effectively the same as moving 1 spoke in the opposite direction: $k = N-1$ corresponds to the lowest negative frequency ($k=-1$). To illustrate this last case more formally:

$$e^{-i2\pi\frac{n(N-1)}{N}} = e^{-i2\pi\frac{nN}{N}+i2\pi\frac{n}{N}} = e^{-i2\pi n}e^{i2\pi\frac{n}{N}} = e^{i2\pi\frac{n}{N}}.$$

The following figure depicts some of the values $e^{-i2\pi\frac{kn}{N}}$ assumes for $N=8$: each row represents an integer frequency $k = [1,3,5,7]$, with four of the eight coefficients ($n = [1,2,3,4]$).



Note how frequencies $k > 4$ are equivalent to "moving around" in the other direction.

The Spectrum can be subsampled by padding the Signal

As shown above, the DFT has only $N/2$ positive frequencies for a signal with N samples. This is in line with the sampling theorem, which states that we must sample a function at twice its maximum frequency, if we wish to fully reconstruct it. With sampling $f(x)$ at N integers 0,1,2,. . . , we can recover a maximum frequency of 0.5 "cycles per x", which maps to $F[N/2]$. The lowest non-zero frequency at $k=1$ is $1/N$ "cycles per x", i.e., one cycle over the whole sampled range $[0, N-1]$. The frequency step size is $1/N$. The following table illustrates the correspondence of k and "real" frequency for $N=4$:

k	0	1	2	3
f	0	0.25	±0.5	-0.25

where f is the frequency in "cycles", and k is the index into the transformed signal array. At the 0 frequency lies the average value of the sampled signal.

In the DFT, the number of signal samples is the same as the number of samples from the spectrum. For a higher–resolution spectrum, we can get a finer sampling by extending the signal with 0 values. In the above example, if we set $N = 8$, the frequencies are:

k	0	1	2	3	4	5	6	7
f	0	0.125	0.25	0.375	± 0.5	-0.375	-0.25	-0.125

Take a look at the definition of the DFT: zero coefficients of the input signal to not add to the transformed function. The only effect is thus a denser sampling of the spectrum. This spectrum does *not* reveal any new information though: it is merely a smoother version of the original one. Vice versa, the number of non-zero coefficients in the input signal *does* make a difference in the spectrum: the spectrum of the sampled step function [0,0,0,0,1,1,1,1] is different from [0,0,0,0,0,0,1,1], even though one might think there is only a single step in both signals, after all. But in the discrete setting, there is not really a "step" in a function, there is only a number of "peaks", where any assumed values between the peaks do not enter the equation. Each sample in the signal corresponds to one complex sine wave in the spectrum. If the signal contains only two non-zero samples, the spectrum will be a sum of two exponentials. No matter how finely you sample this spectrum, no more detail will occur. This is very different from the continuous case, where a step function results in an infinite number of non-zero frequencies in the transform, and the function is evaluated at infinitely small ranges around the step.

References

There's a ton of literature on the Fourier transform, and I cannot possibly list even a substantial part of it. These references just contain a few books that have helped me tremendously in understanding what's going on.

[1] Maor, Eli. e: The story of a number. Princeton University Press, 1994.

[2] Bracewell, Ronald Newbold: the fourier transform and its applications, 3rd ed. McGraw–Hill, 2000.

[3] von Grüningen, Daniel Ch.: Digital Signalverarbeitung. Fachbuchverlag Leipzig, 2004.