Exstrom Laboratories LLC
P.O. Box 7651
Longmont, Colorado 80501
http://www.exstrom.com

Calculating the Frequency Spectrum of a Signal

by
Stefan Hollos, Richard Hollos (Exstrom Laboratories LLC, stefan@exstrom.com, richard@exstrom.com)

Introduction

The problem of accurately determining the frequency spectrum of a signal is a common one in engineering and science. Speech recognition, sonar, radar, and nuclear magnetic resonance (NMR) are just some of the areas where being able to accurately determine frequency content is essential. As an example, a proton precession magnetometer can measure the value of the Earth's magnetic field only as accurately as the frequency of the signal from the precessing protons can be determined.

This article is about calculating the Fourier transform (FT) of a measured signal, which is the standard method for determining its frequency content. The FT is a function of frequency. It is, generaly, a complex function, having both a magnitude and a phase. Its value at a particular frequency is a direct indicator of the magnitude and phase of that frequency in the signal.

The most widely used method for calculating the FT of a measured or sampled signal is called the discrete fourier transform (DFT). The process of measurement and calculation will invariably lead to distortion in the DFT.

The DFT can really only give an estimate of the actual FT of the signal. If the DFT is properly implemented the estimate can be quite good.

We will discuss the step by step process of going from a continuous signal to a calculation of its DFT. The way in which each step in the process affects the FT is discussed. We hope that this will make clear how distortion can appear in the final DFT and how best to implement the process in order to minimize it.

A note on terminology: throughout the article the terms, "spectrum" and "Fourier transform" are used interchangably. In many places we use the angular frequency ω which is equal to $2\pi f$ where f is the frequency.

Calculating the Spectrum

Suppose we have a continuous time signal $s(t)$, whose spectrum or Fourier transform (FT) we want to determine. Mathematically the FT is given by equation 1.

$$S(\omega) = \int_{-\infty}^{+\infty} s(t) e^{-j\omega t} dt \tag{1}$$

To use equation 1 directly, we would need to be able to express $s(t)$ in a mathematical form for which the integral could be solved. Even if possible this approach would be slow, tedious, and not easily automated. Figure 1 shows the outline of a computational approach, which if implemented carefully can give a good approximation of the spectrum of $s(t)$ at a set of evenly spaced frequencies.

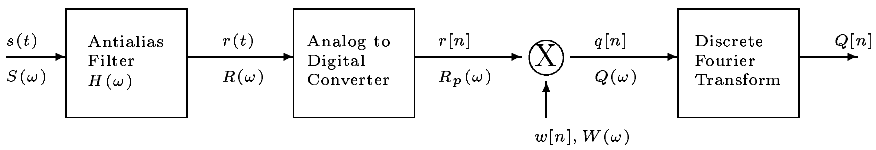

Figure 1: Computational approach for approximating the spectrum of $s(t)$

The first step in this process shows the signal $s(t)$ being fed through a filter with a frequency response $H(\omega)$ to produce the filtered signal $r(t)$ whose FT is given by $R(\omega) = H(\omega)S(\omega)$. The filter is called an antialias filter and it has a low pass frequency response. To understand the purpose of the filter and the reason for its name we need to jump ahead one step and look at the effect that converting a continuous signal into a sequence of samples of the signal has on the spectrum.

In the second step, the continuous signal $r(t)$ is sampled at a fixed time interval T, called the sampling period. This produces a sequence of numbers $r[n] = r(nT)$, where n is an integer. This sampling is usually done with an analog to digital converter (ADC). The ADC converts the signal amplitude (usually in the form of a voltage) at time intervals T into a fixed length binary number.

The conversion to a fixed length binary number leads to what is known as quantization error. As an example, consider an 8 bit ADC with an input voltage range of -10V to +10V. The 8 bits can only represent 256 values. The ADC can therefore only distinguish between sample values that differ by more than 78mV. The effect of this is to introduce some additional noise into the final calculated spectrum. We will ignore the effect of quantization noise in this article and assume that the ADC produces an exact sample of the signal.

There is a minimum sampling period T at which the ADC can effectively operate, and this sets the maximum number of samples per second, $sps = 1/T$. It also places a limit on the largest frequency that can be present in the signal. The reason for this is that the spectrum of the sampled signal, $r[n]$, is equal to a periodic extension of the spectrum of the continuous signal $r(t)$. See figure 2.

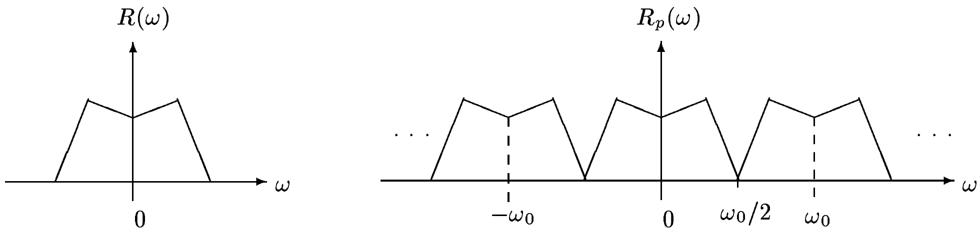

Figure 2: $R(\omega)$ and $R_p(\omega)$

$R(\omega)$ is the spectrum of the continuous signal, and $R_p(\omega)$ is the spectrum of the sampled signal. $R_p(\omega)$ consists of copies of $R(\omega)$ placed at intervals of $\omega_0 = 2\pi/T$ up and down the frequency axis. Mathematically $R(\omega)$ and $R_p(\omega)$ are related by equation 2 (for a full mathematical derivation of all the results of this article, see reference [1]).

$$R_p(\omega) = \frac{1}{T} \sum_{n=-\infty}^{+\infty} R(\omega - n\omega_0) \qquad \omega_0 = 2\pi/T \tag{2}$$

It is clear now why the sampling rate places a limit on the largest frequency that can be present in the signal. If you look at figure 2 you can see that if $R(\omega)$ has a frequency greater than $\omega_0/2$, then the copies of $R(\omega)$ that are present in $R_p(\omega)$ will overlap, and the spectrum of $r(t)$ can then no longer be recovered from the spectrum of the samples.

The purpose of the low pass filtering of the signal $s(t)$ in step 1 should now also be clear. Frequencies in $s(t)$ greater than $\omega_0/2$ have to be eliminated or at least greatly reduced to prevent overlap in the frequency spectrum of the sampled signal. This overlap is called "aliasing" which is the reason for the name "antialiasing filter".

$R_p(\omega)$ is a periodic function of ω with period equal to ω_0, $R_p(\omega + \omega_0) = R_p(\omega)$. A periodic function (the ones that we are interested in) can be written in the form of a Fourier series. A Fourier series represents a function as a weighted sum of complex exponentials or equivalently as a sum of sines and cosines. The Fourier series representation of $R_p(\omega)$ is given in equation 3.

$$R_p(\Omega) = \sum_{n=-\infty}^{+\infty} r[n] e^{-jn\Omega} \qquad \Omega = \omega T \tag{3}$$

This equation is written in terms of the dimensionless frequency $\Omega = \omega T$, and it gives the spectrum of the sampled signal directly in terms of the samples. You can also see from the equation that there is no longer any explicit dependence upon the sampling interval T. The spectrum is now in terms of the dimensionless variable Ω, and $R_p(\Omega)$ now has a period of 2π. From this point on, there is no longer an explicit time dependence. This is not surprising, since there is no way to determine the sampling rate from the samples themselves. After the spectrum has been calculated in terms of Ω, the actual frequency can be determined from $\omega = \Omega/T$.

We are now almost ready to calculate the spectrum, but if you look at equation 3 you will notice a problem. The limits in the summation go from $-\infty$ to $+\infty$. Since we don't have the time to wait that long for that many samples, the computation will have to be performed with a smaller set of samples. Mathematically, we can take into account the effect of a limited set of samples by multiplying the sequence of samples $r[n]$ by another sequence $w[n]$, which has nonzero values for a limited range of n. The sequence of numbers $w[n]$ are samples of what is called a window function. Let N denote the total number of samples used in the calculation of the spectrum, then $w[n]$ will be chosen so as to have nonzero values from $n = 0$ to $N - 1$. The obvious choice is to just let $w[n] = 1$ for $n = 0$ to $N - 1$, which would then select the corresponding samples of $r[n]$ unaltered. This is called a rectangular window for obvious reasons. It is often times better however to taper the window gradually to zero at the beginning and end of the interval. To understand the reason for this, we need to look at the effect that the multiplication of the window function has on the spectrum of $r[n]$.

When the two sequences $r[n]$ and $w[n]$ are multiplied together to form a new sequence $q[n]$, the spectra are related by an operation called convolution. Mathematically we write this as $Q(\Omega) = R(\Omega) * W(\Omega)$ which should not be confused with multiplication. The effect of convolution can best be illustrated by an example. In figure 3, the spectrum $W(\Omega)$ is concentrated in a narrow lobe, and the spectrum $R(\Omega)$ is a constant over an interval from $-\sigma$ to $+\sigma$ and 0 outside the interval.

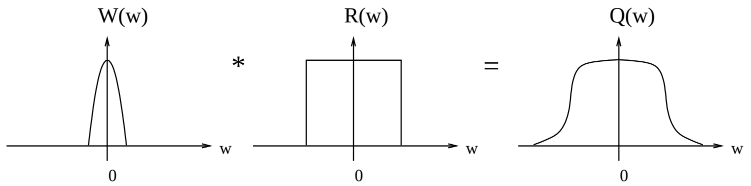

Figure 3: Convolution of two transforms

The effect of the convolution is to smear out the abrupt edges in $R(\Omega)$ causing the spectrum to spread out. The more narrow $W(\Omega)$ is, the less it smears out the spectrum of $R(\Omega)$. If $W(\Omega)$ could be made infinitesimally narrow then it would not cause any spreading in the spectrum of $R(\Omega)$, and $Q(\Omega)$ would be an exact copy of $R(\Omega)$. A window function with such a narrow spectrum would be ideal since it does not alter the spectrum of the sequence $r[n]$ that we are trying to determine. Figure 4 shows two window functions, the simple rectangular window, and a window called a Hamming window, along with their spectra. The spectrum of the rectangular window shows a narrow main lobe and relatively large side lobes. As we will see in a later example, these large side lobes can result in a misleading spectrum. The Hamming window on the other hand, has a broader main lobe but much smaller side lobes which will usually give a truer representation of the spectrum. The Hamming window sequence is given by equation 4.

$$w[n] = \begin{cases} 0.54 - 0.46\cos(2\pi n/(N-1)), & 0 \le n \le N-1 \\ 0, & otherwise \end{cases} \tag{4}$$

After windowing, we have a finite number of nonzero samples $q[n] = r[n]w[n]$ and we can now calculate the spectrum. The spectrum of $q[n]$ is given by equation 5.

$$Q(\Omega) = \sum_{n=0}^{N-1} q[n]e^{-jn\Omega} \qquad \Omega = \omega T \tag{5}$$

This equation can be used to calculate the spectrum for any value of Ω but it is still, computationally, somewhat tedious. A tremendous simplification is possible if we restrict the calculation to N evenly spaced points in the interval $\Omega = 0$ to 2π, we can then use equation 6 to calculate the kth point in the spectrum.

$$Q[k] = Q(\frac{2\pi k}{N}) = \sum_{n=0}^{N-1} q[n]e^{-j2\pi kn/N} \qquad k = 0, 1 \ldots N-1 \tag{6}$$

This equation is known as a discrete Fourier transform (DFT). It allows N samples of the spectrum $Q[k]$ to be determined from N samples of the signal $q[n]$ and vice versa. The simplification comes when the number of samples N is a power of 2, the DFT can then be calculated using a very fast and efficient algorithm called a fast Fourier transform (FFT). The FFT performs the multiplications in equation 6 in a very efficient manner and it does them in parallel so that we get all the samples of the spectrum, $Q[k]$, at once.

Now that we have arrived at a method for calculating a spectrum, it would be good to summarize the entire process. We start with a signal $s(t)$ whose spectrum we want to determine. $s(t)$ must first be low pass filtered to remove frequencies greater than 1/2 the subsequent sampling frequency. The filtered signal $r(t)$ is then sampled by an ADC. The spectrum of the sampled signal is composed of periodic copies of the continuous signal, with a periodicity equal to the sampling frequency. This periodicity is the reason for the initial low pass filtering. The sampled signal is then multiplied by a window function $w[n]$ to yield a set of N nonzero samples. This has the effect of convolving, or smearing out, the spectrum of $r[n]$ with the spectrum of the window function. The form of the window function must then be chosen so that as much as possible of the information of interest in the original spectrum is preserved. Once the sampled signal has been windowed, the spectrum can be calculated at N evenly spaced points using the DFT. If N is a power of 2, then a fast and efficient algorithm for calculating the DFT, called the FFT, can be used. Now it is time to look at a few examples.

Let's start by looking at the simplest possible case of a single pure sine wave. Mathematically, the signal is given by $s(t) = sin(\omega_s t)$. The FT of this can of course be calculated directly using equation 1. The magnitude of the FT consists of two impulses or lines at $-\omega_s$ and $+\omega_s$. Assuming that the sampling frequency is greater than $2w_s$, the spectrum of the sampled signal will just be composed of periodic copies of these two lines. The convolution of any spectrum with an impulse or line just places a copy of the spectrum at the position of the line. Therefore the spectrum of the windowed samples of the sine wave will consist of copies of the spectrum of the window centered at $-w_s$ and $+w_s$ and at the other periodic copies of these lines. When we calculate the DFT, what we will get then is samples of the spectrum of the window centered around the frequency of the sine wave. Looking at the spectrum of the rectangular and Hamming window in figure 4, you can see that samples of these spectra will not necessarily reproduce the single line spectrum of the sine wave.

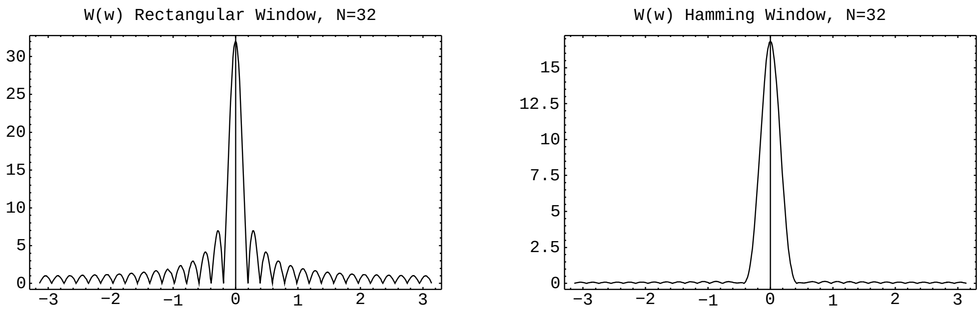

Figure 4: $W(\omega)$ for Rectangular and Hamming windows, $N = 32$

Now let's use some actual numbers and see what we get. Let the frequency of the sine wave be 10 Hz and the sampling rate be 64 sps. If we use a rectangular window 64 samples long, we will then have 64 samples of the sine wave. This corresponds to 1 second worth of data which is 10 complete cycles of the sine wave. The sampled sine wave and its FFT are shown in figure 5.

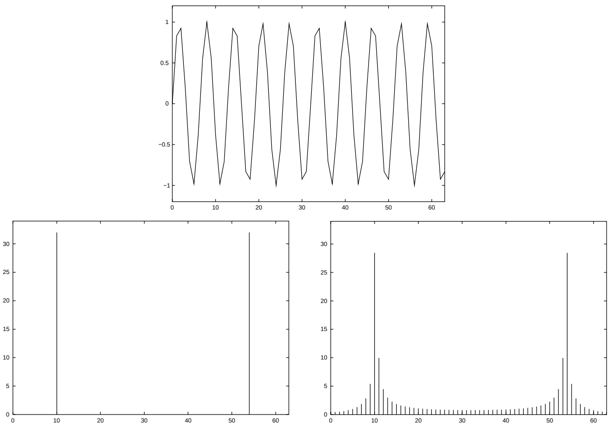

Figure 5: Sampled sine wave (10 Hz) using rectangular window, FFT with 64 sps (left), and FFT with 62.44 sps (right)

In this case the FFT has reproduced the spectrum of the sine wave exactly. There are only 2 lines at points corresponding to +10 Hz and -10 Hz. What has happened in this case is that the FFT has given us samples of the rectangular window spectrum at the peak of the main lobe and at the zeros of the spectrum located between the lobes. For a rectangular window, this will always happen when an integer number of periods of the sine wave fit inside the window. There is another more subtle way of looking at this. The samples of the spectrum of the windowed sine wave, that are calculated by the FFT, are actually the spectrum of copies of the windowed sine wave placed end to end. Since there are an integer number of periods of the sine wave in the window, these copies just reproduce the original sine wave.

Now let's look at what happens when the sampling rate is reduced slightly to 64/1.025 sps, and the number of samples is kept at 64. Now there will be 1.025 seconds of data, which is 10.25 cycles of the 10 Hz sine wave. The FFT for this case is also shown in figure 5. Now we no longer have the single line spectrum of a pure sine wave. This will always be the case when a noninteger number of periods of the sine wave fit inside the window. If copies of the windowed data are placed end to end, they do not reproduce the sine wave. There will be a discontinuity in the resulting waveform which results in the extra frequencies present in the spectrum. Or you can look at it from the point of view of sampling the spectrum of the rectangular window. In this case the samples do not fall on the zeros of the spectrum as they did in the previous example.

What these two examples illustrate is that it is almost never a good idea to use a rectangular window for a general signal. A signal $s(t)$ will generally be composed of many sine waves of varying frequencies and phases, and almost none of them will be properly windowed to give an exact spectral line. To avoid the discontinuities induced by the rectangular window, it is better to use a window such as the Hamming window.

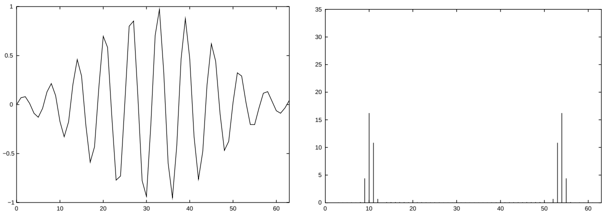

Figure 6: Sampled sine wave (10 Hz) using Hamming window, and FFT with 62.44 sps

Figure 6 shows the previous example using a Hamming window instead of a rectangular window. You can see that the window tapers the sine wave samples to zero at the ends. This removes the sharp discontinuity present in the rectangular window when a noninteger number of periods of the sine wave are present. You can see now that the spectrum is more narrow and gives a truer representation of the single line spectrum of the sine wave. The Hamming window is only one alternative to the rectangular window but it is the simplest and most widely used alternative. Many other windows have been developed for use in different circumstances. For more information on windows see one of the references.

We have tried to keep the level of mathematical detail in this article at an absolute minimum. The goal was to provide a practical guide and not a mathematical treatise on Fourier analysis. If you are interested in a more mathematically in depth treatment of the topics in this article see reference [1]. For a detailed discussion of many aspects of discrete time signal processing see reference [2]. If you are interested in the mathematics of Fourier analysis at a more advanced level see references [3] and [4].

References

[1] http://www.exstrom.com/journal/sigproc/index.html

[2] Oppenheim, A. V., and Schafer, R. W., *Discrete-Time Signal Processing*, Prentice-Hall Englewood Cliffs, NJ, 1989.

[3] Papoulis, A., *The Fourier Integral and Its Applications*, McGraw-Hill Book Company, New York 1962.

[4] Papoulis, A., *Signal Analysis*, McGraw-Hill Book Company, New York 1977.