Understanding the Discrete Fourier Transform

1 author:

R.N. Mutagi
Indus University

Understanding the Discrete Fourier Transform

DFT equations, without insight into what the summations signify, often look formidable to many engineers. DFT can be interpreted in terms of *spot correlation* to understand the physical meaning of the transform.

By R.N. Mutagi

Discrete Fourier Transform (DFT) is used extensively in signal processing applications such as communications, broadcasting, entertainment and many other areas. The DFT equations often discourage engineers in other fields from understanding the digital signal processing (DSP) literature. With the development in analog to digital conversion (ADC) and DSP technology, more and more of the analog circuits are brought into the arena of DSP. For example, in today's software-defined radio, the RF signal leaves the analog domain in the early stage of the receiver and is converted to digital at intermediate frequency (IF) stage. A typical DSP system is shown in Figure 1.

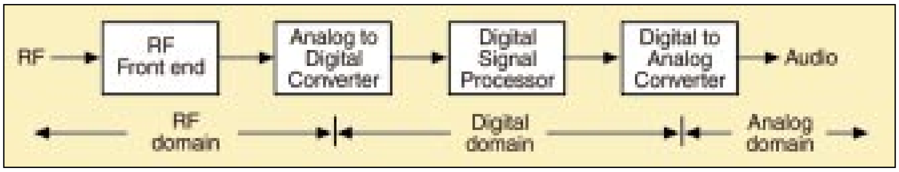

Figure 1. Typical DSP system.

The analog signal is sampled and quantized in an ADC and fed to a DSP as a sequence of numbers. A programmable DSP, with a suitable architecture dealing with data and instructions simultaneously, does the number crunching with hardware, executing signal processing tasks. Understanding DFT is imperative for RF and analog engineers.

The Fourier transform converts a signal representation from time-domain to frequency domain for frequency analysis. However, the Fourier transform is a continuous function of frequency and is not suitable for computation with a DSP. DFT representation of the continuous-time signal permits the computer analysis. Here, we develop DFT, beginning with a simple concept of correlation.

Spot correlation

Suppose we try to find which among the waveforms 2 and 3 in Figure 2 match closely to waveform 1. We may approach the problem intuitively, but we may not always be correct. However, a method exists for measuring the similarity between two functions or waveforms. In signal processing we call it correlation.

We use correlation to estimate a received signal with its known characteristics. For example, we recover the carrier from a noisy received signal by correlating with a locally generated 'clean' carrier and the clock from the noisy data by correlating with a 'clean' clock. In both cases, the error, the difference between the received signal and its estimate, corrects the oscillator phase and frequency. In spread-spectrum modulation, we correlate a locally generated pseudo noise (PN) pattern with the received PN pattern and correct its phase before demodulating the data.

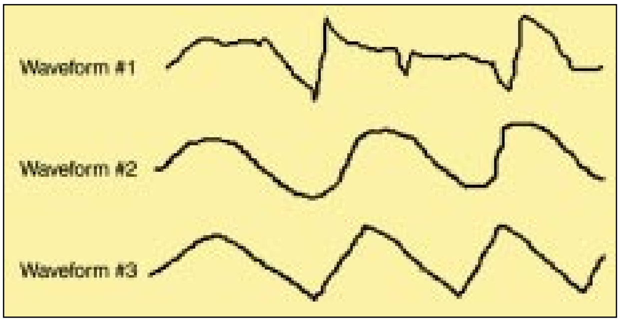

Figure 2. Finding similarity among waveforms.

In the DSP literature the correlation is mathematically described by the equation

$$y(\tau) = \int_{-\infty}^{\infty} x_1(t).x_2(t+\tau)dt \qquad (1)$$

Equation 1 provides the correlation $y(\tau)$ between two functions $x_1(t)$ and $x_2(t)$ as a function of the time shift τ. For any value of time shift τ, the two functions are multiplied and the result is integrated over infinity. For periodic functions this equation is modified as

$$y(\tau) = \frac{1}{T_p}\int_0^T x_1(t).x_2(t+\tau)dt \qquad (2)$$

As you notice, the integration and averaging are done over one period T_p because the signals are now power signals. We now introduce a simplified version of Equation 2 wherein we take away the time shift t and call it as spot correlation.

$$y = \frac{1}{T_p}\int_0^T x_1(t).x_2(t)dt \qquad (3)$$

The spot correlation Equation 3, which is simply the average of the product of two periodic signals over the period, is the measure of similarity between $x_1(t)$ and $x_2(t)$, with no time shift τ between any of them, i.e. the correlation on the spot. Thus, spot correlation provides the pattern matching for the two functions. This interpretation comes in handy when we try to understand the DFT equations. While correlation can be viewed as one function searching to find itself in another, spot correlation can be interpreted as the measure of how much one function is embedded in another. It is the average of the product-sum of two functions, or signals, over the period of interest.

Discrete spot correlation

Sampling a continuous signal $x(t)$, shown in Figure 3(a), with a sampling signal at a regular interval T as in Figure 3(b) gives discrete-time signal with non-zero values at instants nT as shown in Figure 3(c).

We can write the spot correlation equation for two discrete-time signals of N samples as

$$y = \frac{1}{N}\sum_{n=0}^{N-1} x_1(nT)x_2(nT) \qquad (4)$$

Comparing Equation 4 with Equation 2 we notice that the integration over the interval T for the continuous-time signals is replaced by summation of N terms for discrete-time signals. We can drop the sampling interval T from Equation 4 for simplicity and rewrite it as

$$y = \frac{1}{N}\sum_{n=0}^{N-1} x_1(n).x_2(n) \qquad (5)$$

The terms $x_1(n)$ and $x_2(n)$ now represent sequences of N numbers, and y is their product-sum over the interval T. Equation 5 represents the *discrete spot correlation*.

Fourier series

Transformation is an effort to represent an arbitrary signal with a set of signals (called basis functions) with known characteristics like amplitude, frequency and phase. The sum of all components in the set provides an approximation of the arbitrary signal. The difference between the original and the replica is an error, measured in terms of the mean squared error (MSE).

According to Fourier, any arbitrary periodic signal, $\tilde{x}(t)$, is formed by adding up an infinite number of sinusoids with frequencies harmonically related to a fundamental frequency ω_0, with proper amplitude and phase. In other words, any periodic signal can be decomposed into an infinite number of sinusoids whose frequencies are integer multiples of the fundamental frequency ω_0 of the periodic signal. The Fourier series expansion of $\tilde{x}(t)$is given by

$$\tilde{x}(t) = a_0 + \sum_{n=0}^{\infty}[a_n \cos(n\omega_0 t) + b_n \sin(n\omega_0 t)] \qquad (6)$$

Each sine or cosine term has a part, or is embedded, in $x(t)$. The constant term a_0 corresponds to the DC value (zero frequency) of $x(t)$. The coefficient a_n and b_n tell us the amplitudes of sine and cosines, i.e. how much each is contributing to $x(t)$. How do we find the values of a_n and b_n ? Since a_n is the amplitude of the n^{th} harmonic of the cosine signal that is embedded in the signal $x(t)$,we recognize that it is nothing but the spot correlation of the cosine and $x(t)$. Hence, we obtain a_n and b_n by using Equation 3.

$$a_n = \frac{1}{T}\int_{-T/2}^{T/2} \tilde{x}(t).\cos(n\omega_0 t)dt \qquad (7)$$

and

$$b_n = \frac{1}{T}\int_{-T/2}^{T/2} \tilde{x}(t).\sin(n\omega_0 t)dt \qquad (8)$$

Equation 6 can be written in exponential form as

$$\tilde{x}(t) = \sum_{n=0}^{\infty} c_n e^{jn\omega_0 t} \qquad (9)$$

and

$$c_n = \frac{1}{T}\int_{-T/2}^{T/2} \tilde{x}(t)e^{-jn\omega_0 t}dt \qquad (10)$$

A continuous, periodic signal can be decomposed into an infinite set, called the Fourier series, of harmonically related frequencies, the fundamental frequency being equal to the inverse of the period.

Fourier series to transform

Fourier series provides frequency domain representation for only periodic signals. Unfortunately all signals are not periodic. Speech, music or video signals are examples of non-periodic signals. We must have a method for obtaining frequency domain representation of non-periodic signals. This is precisely what Fourier transform provides. We can get Fourier transform equations from Fourier series equations following the simple steps in Figure 4.

We begin with a rectangular pulse train with period T_p and width t as shown in Figure 4(a). We can obtain the frequency domain representation for this signal using the Fourier series expression in Equation 10. The Fourier series, shown in Figure 4 (b), has line frequencies

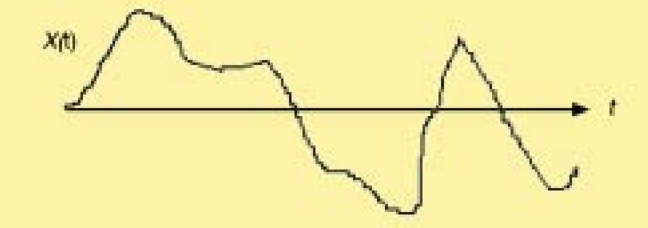

Figure 3(a). Continuous-time signal

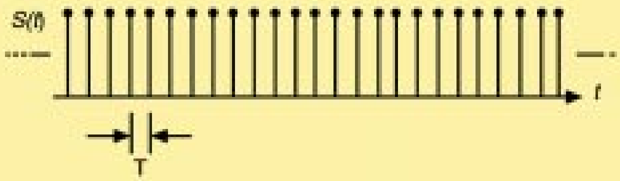

Figure 3(b). Sampling signal

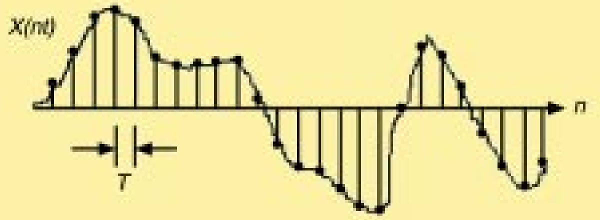

Figure 3(c). Discrete-time signal

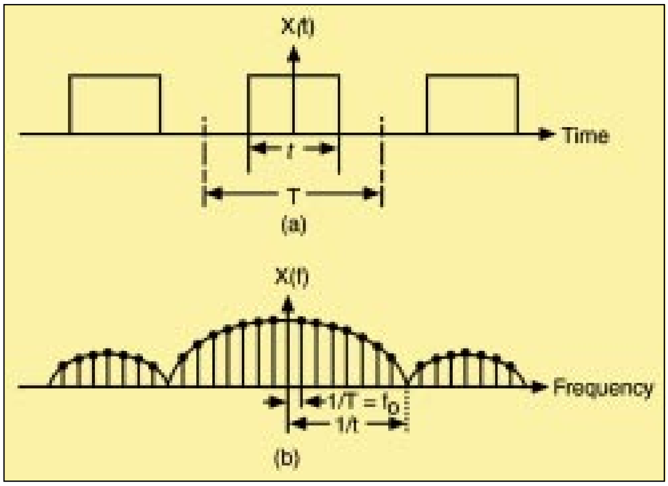

Figure 4. (a) A periodic pulse train and (b) its Fourier series.

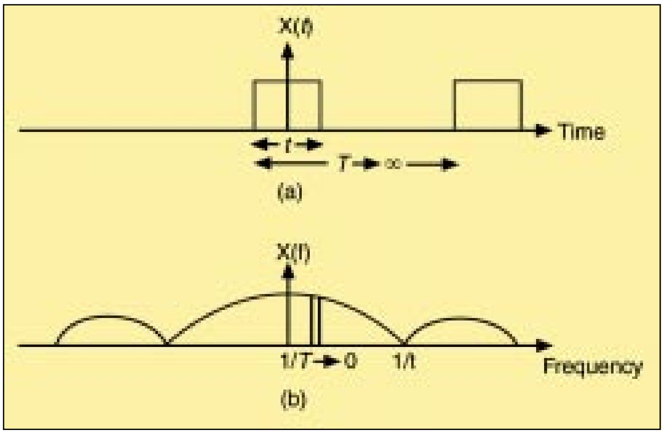

Figure 5. (a) As the periodic tends to infinity the periodic signal becomes aperiodic and (b) the frequency spacing in the Fourier series reduces to zero resulting in Fourier transform.

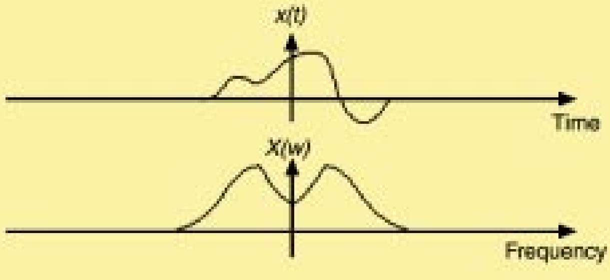

Figure 6(a). Non-periodic continuous-time signal with its non-periodic continous-frequency spectrum

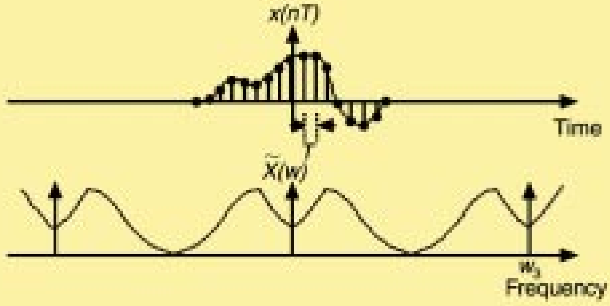

Figure 6(b). Non-periodic discrete-time signal with its periodic continous-frequency spectrum

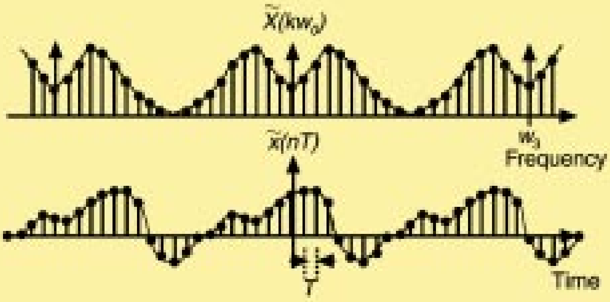

Figure 6(c). Periodic discrete-time signal and its periodic discrete-frequency spectrum

spaced at $1/T_P$ Hz and its envelope has the shape of *sin(x)/(x)*, with the null at *1/t* Hz. This example helps us in developing intuitively the Fourier transform from the series when the signal is not periodic. If the time period T_P is increased, it is reflected in the Fourier series with reduced frequency spacing because the spacing between the harmonics is equal to the fundamental frequency $f_0 = 1/T$. In the limit, if T is increased to infinity the periodic signal becomes aperiodic as shown in Figure 5(a) and the frequency spacing in the spectrum is reduced to zero as shown in Figure 5(b), making it continuous. This is essentially the Fourier Transform.

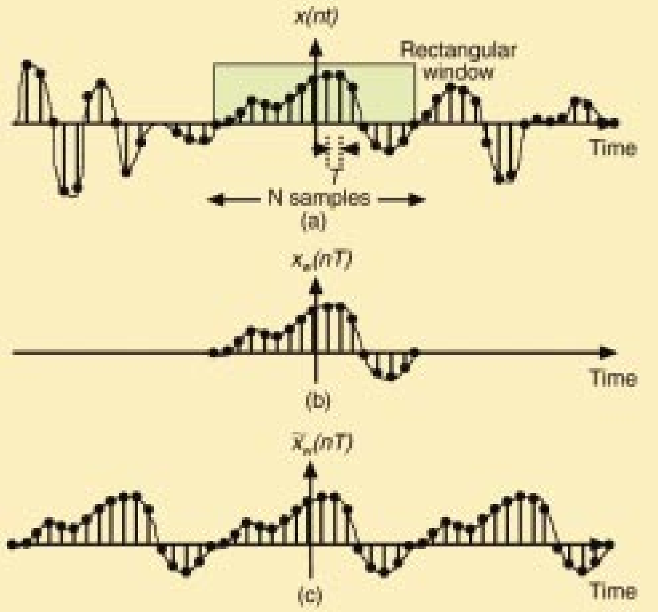

Figure 7. An aperiodic discrete-time signal (a) is windowed (b) and the windowed segment is replicated (c) to make it periodic for which DFT can be computed.

The Fourier series in Equation 6 then needs to be modified by changing the summation to integration and the discrete frequency $n\omega_0$ to continuous frequency ω. Equation 6 reduces to Fourier transform as given in Equation 11.

$$x(t) = \int_{w=-\infty}^{\infty} [a_n \cos(\omega t) + b_n \sin(\omega t)] d\omega \quad (11)$$

which may be written in exponential form as

$$x(t) = \int_{\omega=-\infty}^{\infty} X(w) e^{j\omega t} d\omega \quad (12)$$

The coefficient, or the magnitude, *X(ω)* of the exponent $(e^{j\omega})$ minimizes the error between the actual signal and its approximation through the exponents and is again obtained by using the spot correlation equation as below,

$$X(\omega) = \int_{t=-\infty}^{\infty} x(t) e^{-j\omega t} dt \quad (13)$$

Thus, *X(ω)* tells us how much of $e^{-j\omega t}$ is present in *x(t)*. Equations 12 and 13 form the Fourier transform pairs.

Discrete Fourier Transform

Let us take a close look at Equations 12 and 13. We note that to find *x(t)* or *X(ω)* we need to integrate the product of two continuous functions over infinity. This is fine if we are only doing the math, but if we are interested in finding Fourier transform in real-time applications it is just not possible. We need to use some tricks to make it a usable solution. Firstly, we must have a finite duration signal and a finite spectral band. Secondly, the signal must have finite time samples and the spectrum must have finite frequency components. This is precisely what we are going to have, following the steps given here.

Let us begin with a continuous-time non-periodic signal *x(t)* as shown in Figure 6(a). Its Fourier transform would be *X(ω)*, which is continuous and non-periodic in frequency domain. Now, we sample the signal at intervals of *T* making it discrete in time. One of the theorems in signal processing is that multiplication in time domain is equivalent to convolution in frequency domain and vice versa. Since sampling is multiplication of *x(t)* with sampling signal *s(t)* in time domain, it results in the convolution of the signal spectrum with the spectrum of *s(t)*, which is a series of harmonics of f_s as shown in Figure 6(b). If you remember the amplitude modulation spectrum you will recognize that the spectrum in Figure 6(b) is just the modulation of each harmonic of f_s by *s(t)*. Convolution is just that process.

The next step is to make this spectrum discrete by sampling it at ω_0 Hz. Again this process in frequency domain affects our time-domain signal by making it periodic as shown in Figure 6(c). This is the result of convolution of the signal with an impulse train with a period $1/f_0$. We now have a discrete-time signal and its discrete Fourier transform. Both these extend to infinity but are periodic. Hence, we take one period in each domain by multiplying by a rectangular function. This is called windowing. Windowing in one domain will, of course, affect the signal in another domain. Thus, we consider one period of the final discrete spectrum as the discrete Fourier transform of one period of the discrete periodic signal. Equations 14 and 15 provide DFT synthesis and analysis equations.

$$x(n) = \frac{1}{N}\sum_{k=0}^{N-1} X(k)e^{j(2\pi/N)kn} \tag{14}$$

$$X(k) = \sum_{n=0}^{N-1} x(n)e^{-j(2\pi/N)kn} \tag{15}$$

The synthesis equation suggests that a set of *N* time samples can be decomposed to *N* exponentials (sine and cosine terms) whose magnitude is given by the analysis equation. Thus, each *X(k)* representing the magnitude of sine and cosine components is obtained by the spot correlation of the *N* samples of the signal segment and *N* samples of the sine and cosine components. Computing each value of *X(k)* requires *N* multiplications and as there are *N* values for k, we need N^2 multiplications to compute the DFT.

When we try to analyze the signals using DFT in practice, we first obtain a discrete signal by sampling as shown in Figure 7(a), and take a segment of *N* samples by and windowing as shown in Figure 7(b). Next, we imagine that the segment is periodic, i.e. the segment is extended on either side as shown in Figure 7(c). For this signal we compute the coefficients of one period of the discrete Fourier transform using Equation 15. Then, we move onto the next segment in the sequence of samples and repeat our computation of DFT coefficients. The representation through samples of the Fourier transform is in effect a representation of the finite-duration sequence by a periodic sequence, one period of which is the finite-duration sequence we wish to represent.

The physical meaning of the transform, if understood in terms of spot correlation, can help us interpret DFT, and can take away the intimidation factor of DFT equations. RFD

ABOUT THE AUTHOR

R.N.Mutagi holds a B.E. in Telecommunications from Karnatak University and a diploma in Electronics Design technology from the Indian Institute of Science, both from India. He worked at the Indian Space Research Organization, Ahmedabad as the Head of Baseband Processing Division developing satellite communication systems.

In 1999 he moved to Canada and worked at EMS Technologies at Montreal as Design Manager and Systems Engineer. His interests include DSP, Digital Communications and Signal Compression. He has authored many articles and papers.

Mutagi may be contacted at:
mutagi@ieee.org