Intuitive Guide to Fourier Analysis

Charan Langton
Victor Levin

Much of this book relies on math developed by important persons in the field over the last 200 years. When known or possible, the authors have given the credit due. We relied on many books and articles and consulted many articles on the internet and often many of these provided no name for credits. In this case, we are grateful to all who make the knowledge available free for all on the internet.

The publisher offers discounts on this book when ordered in quantity for bulk purchase or special sales. We can also make available on special or electronic version applicable to your business goals, such as training, marketing and branding issues. For more information, please contact us.

mntcastle@comcast.net

Website for this book: complextoreal.com/fftbook

ISBN- 13: 978-0-913063-26-2

3 | Discrete-time Signals and Fourier series representation

Peter Gustav Lejeune Dirichlet
13 February 1805 – 5 May 1859

Johann Peter Gustav Lejeune Dirichlet was a German mathematician who made deep contributions to number theory, and to the theory of Fourier series and other topics in mathematical analysis; he is credited with being one of the first mathematicians to give the modern formal definition of a function. Dirichlet published in 1829 a famous memoir giving the conditions, showing for which functions the convergence of the Fourier series holds. Before Dirichlet's solution, not only Fourier, but also Poisson and Cauchy had tried unsuccessfully to find a rigorous proof of convergence. The memoir introduced Dirichlet's test for the convergence of series. It also introduced the Dirichlet function as an example that not any function is integrable (the definite integral was still a developing topic at the time) and, in the proof of the theorem for the Fourier series, introduced the Dirichlet kernel and the Dirichlet integral. – From Wikipedia

In the previous two chapters, we discussed Fourier series as applied to continuous-time signals. We saw that the Fourier series can be used to create a representation of any periodic signal. This representation is done using the sine and cosine functions or with complex exponentials. Both forms are equivalent. In the previous two chapters, our discussion was limited to continuous-time (CT) signals. In this chapter we will discuss Fourier series analysis as applied to discrete-time (DT) signals.

Discrete signals are different from analog signals

Although some data is naturally discrete such as stock prices, number of students in a class etc. many electronic signals we work with are sampled from analog signals. Examples of sampled signals are voice, music, and medical/biological signals. The discrete signals are generated from analog signals by a process called **sampling**. This is also known as **Analog-to-Digital** conversion. The generation of a discrete signal from an analog signal is done by an instantaneous measurement of the analog signal amplitude at uniform intervals.

Discrete vs. digital

In general terms, a **discrete signal** is *continuous* in amplitude but is *discrete* in time. This means that it can have any value whatsoever for its amplitude but is defined or measured only at *uniform* time intervals. Hence the term *discrete* applies to the time dimension and not to the *amplitude*. For purposes of Fourier analysis, we assume that the sampling is done at constant time intervals between the samples.

A discrete signal is often confused with the term **digital signal**. Although in common language they are thought of as the same thing, a digital signal is a special type of discrete signal. Like any discrete signal, it is defined only at specific time intervals, but its amplitude is constrained to specific values. We have binary digital signals where the amplitude is limited to only two values, $\{+1, -1\}$ or $\{0, 1\}$. A M-level signal can take on just one of 2^M preset amplitudes only. Hence a *digital* signal is a specific type of discrete signal with *constrained amplitudes*. In this chapter we will be talking about general discrete signals which include digital signals. Both of these types of signals are called discrete-time (DT) signals. We call a general sampling time, the *sampling instant*. How fast or slow a signal is sampled is specified in terms of its *sampling frequency*, which is given in terms of the number of samples per second.

Generating discrete signals

In mathematics there is often a need to distinguish between a continuous-time (CT) and a discrete-time (DT) signal. The convention is that a discrete-time signal is written with square brackets around the time index, n, whereas the continuous-time signal is written the usual way with round brackets around the time index t. We will be using n as the index of

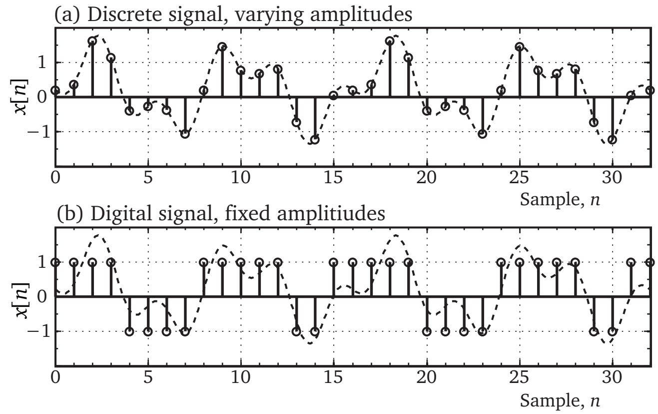

Figure 3.1: Discrete sampling collects the actual amplitudes of the signal at the sampling instant, whereas digital sampling rounds the values to the nearest allowed value. In (b), the sampling values are limited to just 2 values, $+1$ or -1. Hence, each value from (a) has been rounded to either a $+1$ or -1 to create a binary digital signal.

discrete time for a DT signal and t as the index of time for a continuous-time signal.

$$\begin{array}{ll} x(t) & \text{Continuous} \\ x[n] & \text{Discrete} \end{array}$$

We can create a discrete signal by multiplying a continuous signal with a comb-like sampling signal, as shown in Fig. 3.2(b). The sampling signal shown in this figure is an impulse train but we will give it a generic name of $p(t)$. We write the sampled function, x_s as a function of time as the product of the continuous signal and the sampling signal.

$$x_s(t) = x(t)p(t) \tag{3.1}$$

The time between the samples, or the sampling time is referred to as T_s, and the sampling frequency or rate is defined as the inverse of this sample time.

If we are given a CT signal of frequency f_0, and this is being sampled at M samples per second, we would compute the discrete signal from the continuous signal with this Matlab code. Here time t has been replaced with n/F_s.

```
xc = sin(2*pi*f0*t)
Fs = 24
n = -48: 47
xd = sin(2*pi n/Fs)}
```

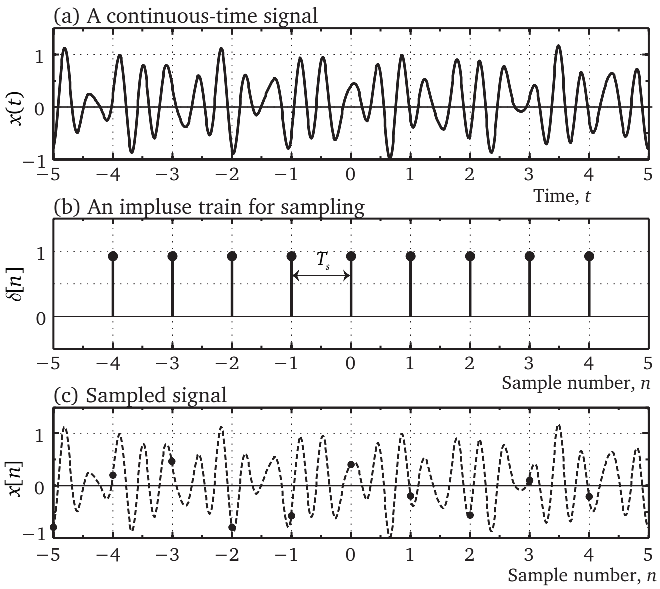

Figure 3.2: A continuous-time signal sampled at uniform intervals T_s with an ideal sampling function. The discrete signal in (c) $x[n]$ consists only of the discrete samples and nothing else. The continuous signal is shown in dashed line for reference only. The receiver has no idea what it is. All it sees are the samples.

Sampling and interpolation

Ideal sampling

Let's assume we have an impulse train, $p(t)$ with period T_s as the sampling function. Multiplying the impulse train with the signal, as in Eq. (3.1) we get a continuous signal with non-zero samples at discrete-time sample points, referred to by nT_s, or n/F_s. Hence, the absolute time is the ordinal sample number times the time in between each sample. For the discrete signals, the sample time is left out as a parameter and we only talk about n, the sample number. The sample time T_s, becomes an independent parameter. Hence if we have two discrete signals with exactly the same sample values, are the signals identical? No, because the sampling interval may be different.

For a signal sampled with sampling function, $p(t)$, an impulse train, we write the sampled signal per Eq. (3.1) as

$$\begin{aligned} p(t) &= \sum_{n=-\infty}^{\infty} \delta(t - nTs) \\ x_s(t) &= x(t) \sum_{n=-\infty}^{\infty} \delta(t - nTs) \end{aligned} \tag{3.2}$$

We write the expression for a discrete signal of a sampled version of the CT signal.

$$x[n] = x_s(t)|_{t=nT_s)} = x(nT_s) \tag{3.3}$$

The term $x(nT_s)$ with round brackets is continuous since it is just the value of the continuous-time signal at time nT_s. The term $x[n]$ however is discrete, because the index n is an integer by definition. The discrete signal $x[n]$ has values only at points $t = nT_s$ where n is the integer sample number. It is undefined at all non-integers *unlike* the continuous-time signal. The sampling time T_s relative to the signal frequency determines how coarse or fine the sampling is. The discrete signal can of course be real or complex. The individual value $x[n]$ is called the n-th sample of the sequence.

Reconstruction of the analog signal from discrete samples

Why sample a signal? We sample a signal for one big reason, to reduce its bandwidth. The other benefit we get from sampling is that signal processing on digital signals is “easier”. However, once sampled, processed and transmitted, this signal must often then be converted back to its analog form. The process of reconstructing a signal from its discrete samples is called **interpolation**. This is the same thing we do when we plot a function. We compute a few values at some selected points and then connect those points to plot the representation of the function. The reconstruction by machines however is not as straight forward and requires giving them an algorithm that they are able to do. This is where things get complicated.

First of all, we note that there are two conditions for ideal reconstruction. One is that the signal must have been ideally sampled to start with, i.e. by an impulse train such that the sampled values represent the true amplitudes of the signal. Ideal sampling is hard to achieve but for our purposes, we will assume it can be done.

The second is that the signal must not contain any frequencies above one-half of the sampling frequency. This second condition can be met by first filtering the signal by an anti-aliasing filter, a filter with a cutoff frequency that is one-half the sampling frequency prior to sampling. Or we can assume that the sampling frequency chosen is large enough to encompass all the important frequencies in the signal. Let's assume this is done also.

For the purposes of reconstruction, we chose an arbitrary pulse shape, $h(t)$. The idea is that we will replace each discrete sample with this pulse shape, and we are going to do this by convolving the pulse shape with the sampled signal. We write the sampled signal (in large parenthesis) convolved with an arbitrary shape, $h(t)$ as

$$x_r(t) = \left\{ \sum_{n=-\infty}^{\infty} x(nT_s)\delta(t - nT_s) \right\} * h(t) \tag{3.4}$$

The subscript r indicates that this is a reconstructed signal. At each sample n, we convolve the sample (a single value) by $h(t)$ (a little wave of some sort, lasting some time). This convolution in Eq. (3.4) centers the "little wave" at the sample location. All these arrayed waves are then added in time. (Note that they are in continuous-time.) Depending on the $h(t)$ or the little wave selected, we will get a reconstructed signal which may or may not be a good representation of the original signal.

Simplifying this equation by completing the convolution of $h(t)$ with the impulse train, we write this somewhat simpler equation for the reconstructed signal.

$$x_r(t) = \sum_{n=-\infty}^{\infty} x(nT_s)h(t - nT_s) \tag{3.5}$$

To examine the possibilities for shapes, $h(t)$, we pick the following three; a rectangular pulse, a triangular pulse and a sinc function. It turns out that these three pretty much cover most of what is used in practice. Each of these "shapes" has a distinctive frequency response as shown in Fig. 3.3. We use the frequency response to determine the effect these shapes will have on the reconstructed signal. Of course, we have not yet said what a *frequency response* is. A frequency response is actually the spectrum, but it has a slightly different interpretation. It is meant to imply that a system can be identified in this manner, or what it does by what its frequency output looks like.

Method 1: Zero-order hold

Fig. 3.3(a) shows a square pulse. We will replace each sample with a square pulse of amplitude equal to the sample value. This basically means that the sample amplitude is held to the next sampling instant in a flat line. The hold time period is T. This form of reconstruction is called **sample-and-hold** or **zero-order-hold (ZOH)** method of signal reconstruction. Zero in ZOH is the slope of the interpolation function, a straight line of zero

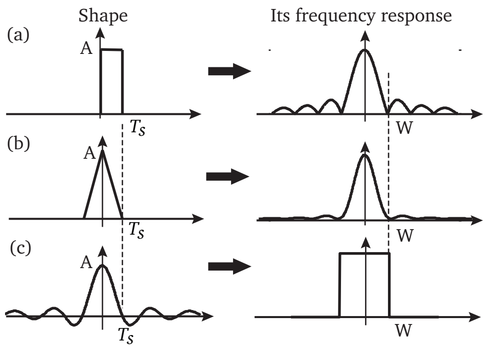

Figure 3.3: We will reconstruct the analog signal by replacing each sample by one of these shapes on the left: rectangular, triangular or the complicated looking sinc function. Each has a distinct frequency response as shown on the right.

slope connecting one sample to the next. We may think this as a simplistic method but if done with small enough resolution, that is a very narrow rectangle in time, ZOH can do a decent job of reconstructing the signal. The shape function $h(t)$ in this case is a rectangle.

$$h(t) = \text{rect}(t - nT_s) \tag{3.6}$$

The reconstructed signal is now given using the general expression of Eq. (3.4), where we substitute the [rect] shape into Eq. (3.5).

$$x_r[n] = \sum_{n=-\infty}^{\infty} x(nT_s)\text{rect}(t - nT_s) \tag{3.7}$$

We show the index as going from $-\infty < n < +\infty$ as the general form. In Fig. 3.4(a), we see a signal reconstructed using a ZOH circuit. The rectangular pulse is scaled and repeated at each sample.

Method 2: First order hold (linear interpolation)

In this case, we replace each sample with a pulse shape that looks like a triangle of width $2T_s$ as given by the expression

$$h(t) = \begin{cases} 1 - t/T & 0 < t < T_s \\ 1 + t/T & T_s < t < 2T_s \\ 0 & \text{else} \end{cases} \tag{3.8}$$

This function is shaped like a triangle and the reconstructed signal equation from Eq. (3.5) now becomes

$$x_s[n] = \sum_{n=-\infty}^{\infty} x(nT_s)\ \text{tri}\left(\frac{t - nT_s}{T_s}\right) \tag{3.9}$$

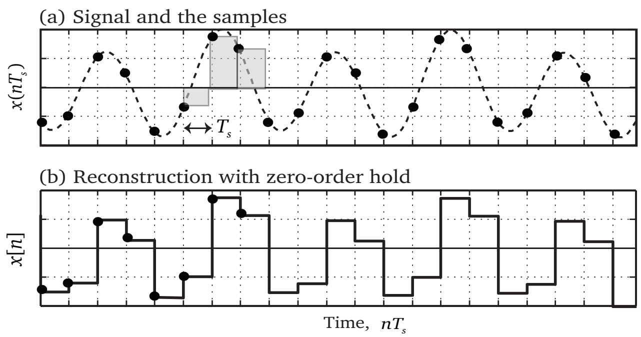

Figure 3.4: A zero-order-hold method of reconstructing the original analog signal. In (a), we see the samples, in (b) we hold each sample value to the next sample time.

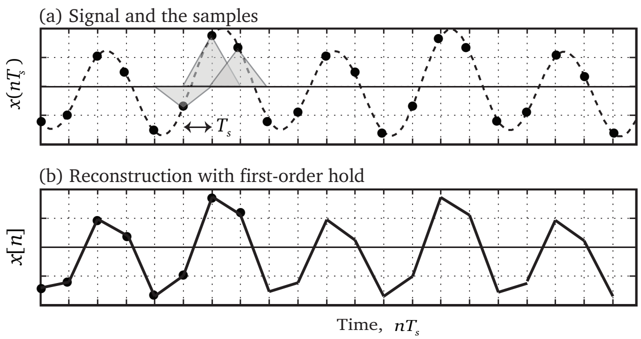

Figure 3.5: A First-order-hold (FOH) method of reconstruction by replacing each sample with a triangle of the same amplitude but twice the sample time width. The summation of overlapping triangles results in a signal that appears to linearly connect the samples.

We see in Fig 3.5(b) that instead of non-overlapping rectangles, as in ZOH we have overlapping triangles. That's because we set the width of the triangle as twice the sample time. This double-width does two things; it keeps the amplitude the same as the case of the rectangle and it fills the in-between points in a linear fashion. This method is also called a **linear interpolation** as we are just connecting the points. This is also called the First-order-hold (FOH) because we are connecting the adjacent samples with a line of linear slope. Why you may ask use triangles when we can just connect the samples? Machines cannot "see" the samples nor "connect" the samples. Addition is about all they can do well. Hence this method replaces linear interpolation as you and I might do visually with a simple addition of displaced triangles. It also gives us a hint as to how we can use any shape we want and in fact of any length, not just two times the sample time! sinc pulse is such a shape.

Method 3: Sinc interpolation

We used triangles in FOH and that seems to produce a better looking reconstructed signal than ZOH. We can in-fact use just about any shape we want to represent a sample, from the rectangle to one composed of a complex shape, such as a sinc function. A sinc function seems like an unlikely choice since it is non-causal (as it extends into the future) but it is in-fact an extension of the idea of the first two methods. Both zero-order and first-order holds are forms of polynomial curve fit. The first-order is a linear polynomial, and we continue in this fashion with second-order on up to infinite orders to represent just about any type of wiggly shape we can think of. A sinc function, an infinite order polynomial, is the basis of perfect reconstruction. The reconstructed signal becomes a sum of scaled, shifted sinc functions same as we did with triangular shapes. Even though the sinc function is an infinitely long function, it is zero-valued at regular intervals. This interval is equal to the sampling period. Since each sinc pulse lobe crosses zero at only the sampling instants, the summed signal where each sinc is centered at a different time, adds no interference (quantity of its own amplitude) to other sinc pulses centered at other times. Hence this shape is considered to be free of **inter-symbol interference (ISI)**.

The equation we get for the reconstructed signal in this case is similar to the first two cases, with the reconstructed signal summed with each sinc located at nT_s.

$$\begin{aligned} h(t) &= \mathrm{sinc}(t - nT_s) \\ x_r[n] &= \sum_{n=-\infty}^{\infty} x(nT_s)\,\mathrm{sinc}(t - nT_s) \end{aligned} \tag{3.10}$$

In Fig. 3.6 we see the sinc reconstruction process for a signal with each sample being replaced by a sinc function and the resulting reconstructed signal compared to the original signal in (c).

Clearly we can see that the sinc construction in Fig. 3.6(c) does a very good job. How to tell which of these three methods is better? Clearly ZOH is kin dof rough. But to properly assess these methods requires the full understanding of the Fourier transform, a topic yet to be covered in Chapter 4. So we will drop this subject now with recognition that a signal can be perfectly reconstructed using the linear superposition principal using many different shapes, with sinc function being one example, albeit a really good one, the one we call the "perfect reconstruction".

Sinc function detour

We will be coming across the sinc function a lot. It is the most versatile and also most used piece of mathematical concept in signal processing. Hence we examine the sinc function

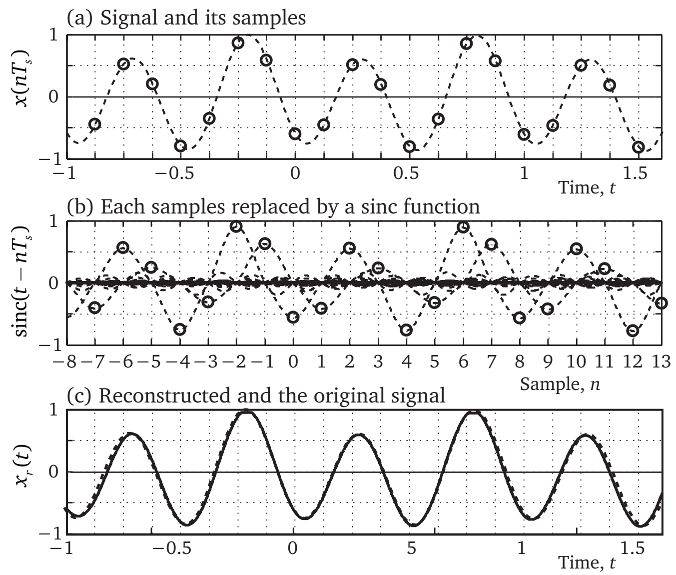

Figure 3.6: Reconstruction with sinc pulses means replacing the sample with a scaled sinc function. (a) the signal samples are effectively replaced by a scaled sinc as in (b) to create a perfect signal in (c).

in a bit more detail now. In Fig. 3.7, we see the function plotted in time-domain. This form is called the normalized sinc function. It is a continuous function of time, t and it is not periodic.

$$h(t) = \begin{cases} 1 & t = 0 \\ \dfrac{\sin(\pi t)}{\pi t} & t \neq 0 \end{cases} \tag{3.11}$$

At $t = 0$, its value is 1.0. As we see from this equation, the function is zero for all integer values of t, because sine of an integer multiple of π is zero. In Matlab this function is given as: sinc(t). No π is needed as it is already programmed in. The Matlab plot would yield first zero crossing at ± 1 and as such the width of the main lobe is 2 units. We can create any main lobe width by inserting a variable T_s into the equation Eq. (3.12). The generic sinc function of lobe width T_s (main lobe width $= 2T_s$) is given by

$$h(t) = \frac{\sin(\pi t / T_s)}{\pi t / T_s} \tag{3.12}$$

In Matlab, we would create a sinc function as follows.

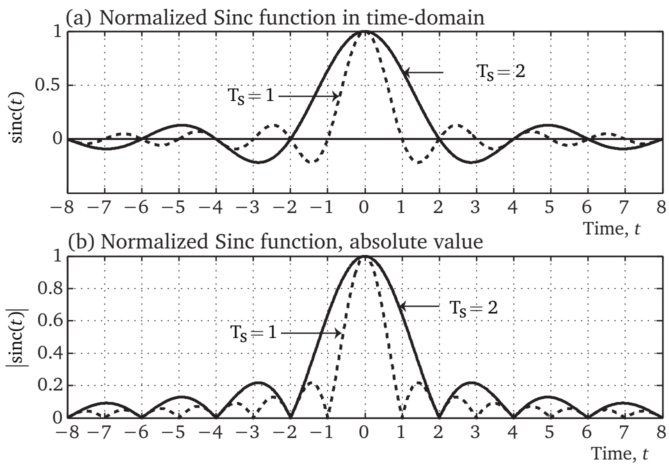

Figure 3.7: The sinc function in time-domain. The signal is non-periodic. Its peak value is 1 for the normalized form. Note that the main lobe of the sinc function spans two times the parameter T_s. In (b) we plot the absolute value of this function so the lobes on the negative side have flipped to the top.

```
% A sinc function in two forms
    t = -6: .01: 6;
    Ts = 2;
    h = sinc(t/Ts);
    habs = abs(h1)
    plot(t, h, t, habs)
```

Fig. 3.7 shows this function for two different values of, $T_s = 2$ and $T_s = 1$ (same as the normalized case.) The sinc function is often plotted in the second style as in Fig. 3.7(b) with amplitudes shown as absolute values. This style makes it easy to see the lobes and the zero crossings. Note that zero-crossings occur every T_s seconds. It is the preferred style in frequency domain, but not in time domain. Note that the function has a main lobe that is 2 times T_s seconds wide. All the other lobes are T_s seconds wide.

The sinc function has some interesting and useful properties. First one is that the area under it is equal to 1.0.

$$\int_{-\infty}^{\infty} \text{sinc}(2\pi t/T_s)dt = \text{rect}(0) = 1.$$

The second interesting and very useful property, from Eq. (3.12) is that as T_s decreases, the sinc function approaches an impulse. This is of course obvious from Fig. 3.7. A smaller value of T_s means a narrower lobes. Narrow main lobe makes the central part impulse-like

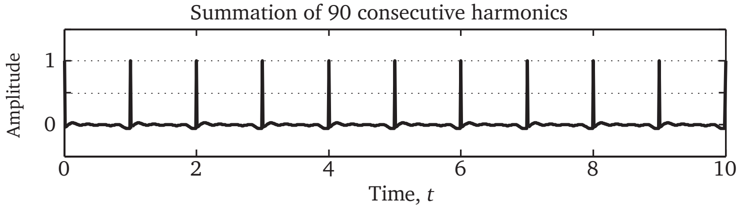

Figure 3.8: In this figure we add 90 consecutive harmonics ($f = 1,\ 2, \ldots 90$)*, and what we get looks very nearly like an impulse train.*

and hence we note that as $T_s \to 0$ goes to zero, the function approaches an impulse. Another interesting property is that the sinc function is equivalent to the summation of all complex exponentials. This is a magical property in that it tells us how Fourier transform works by scaling these exponentials. We showed this effect in Chapter 1 by adding many harmonics together and noting that result approaches an impulse train.

$$\text{sinc}(t) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{j\omega t} d\omega \tag{3.13}$$

This property is best seen in Fig. 3.8, which shows what we get when we add a large number of harmonic complex exponentials together. The signal looks very much like an impulse train.

The sinc function is also the frequency response to a square pulse. We can say that it is a representation of a square pulse in the frequency domain. If we take a square pulse (also called a rectangle, probably a better name anyway) in time domain, then its Fourier series representation will be a sinc, alternately, if we take a time domain sinc function as we are doing here, then its frequency representation is a rectangle, which says that it is bounded in bandwidth. We learn from this that a square pulse has very large (or in fact infinite) bandwidth.

Sampling rate

How do we determine an appropriate sampling rate for an analog signal? In Fig. 3.9 we show an analog signal sampled at two different rates, in (a) the signal is sampled slowly and in (b) sampled rapidly. At this point, our idea of slow and rapid is arbitrary.

It is obvious by looking at the samples in Fig. 3.9(a) that the rate is not quick enough to capture all the ups and downs of the signal. Some high and low points have been missed. But the rate in (b) looks like it might be too fast as it is capturing more samples than we may need. Can we get by with a smaller rate? Is there is an optimum sampling rate that captures just enough information such that the sampled analog signal can still be reconstructed faithfully from the discrete samples?

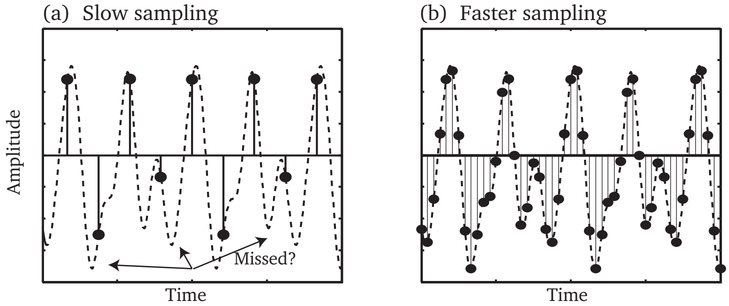

Figure 3.9: The sampling rate is an important parameter, (a) analog signal sampled probably too slowly, (b) probably too fast.

Shannon's Theorem

There is an *optimum sampling rate*. This optimum sampling rate was established by Harry Nyquist and Claude Shannon and others before them. But the theorem has come to be attributed to Shannon and is called the **Shannon's Theorem**. Although Shannon is often given credit for this theorem, it has a long history. Even before Shannon, Harry Nyquist (a Swede who immigrated to USA in 1907 and did all his famous work in the USA) had already established the **Nyquist Rate**. Shannon took it further and applied the idea to reconstruction of discrete signals. And even before Nyquist, the sampling theorem was used and specified in its present form by a Russian scientist by the name of **V. A. Kotelnikov** in 1933. And even he may have not been the first. So simple and yet so profound, the theorem is a very important concept for all types of signal processing.

The theorem says:

For any analog signal containing among its frequency content a maximum frequency of $f_{\max}$, then the analog signal can be represented faithfully by N equally spaced samples, provided the sampling rate is at least two times $f_{\max}$ samples per second.

We define the **Sampling frequency**, F_s as the number of samples collected per second. For a faithful representation of an analog signal, the sampling rate F_s must be equal or greater than two times the maximum frequency contained in the analog signal. We write this rule as

$$\boxed{F_s \geq 2f_{\max}} \tag{3.14}$$

The **Nyquist Rate** is defined as the case of sampling frequency F_s exactly equal to 2 times $f_{\max}$. This is also called the ***Nyquist threshold*** or ***Nyquist frequency***. T_s is defined as the time period between the samples, and is the inverse of the sampling frequency, F_s.

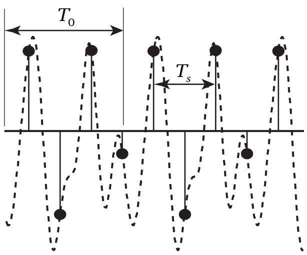

Figure 3.10: There is no relationship between the sampling period and the fundamental period of the signal. They are independent quantities.

A real life signal will have many frequencies. In setting up the Fourier series representation, we define the lowest of all its frequencies, f_0 as its *fundamental frequency*. The fundamental period of the signal, T_0 is the inverse of the fundamental frequency we defined in chapter 1.

The ***maximum frequency***, $f_{\max}$ contained within the signal is used to determine an appropriate sampling frequency for the signal, F_s. An important thing to note is that the fundamental frequency, f_0 is not related to the maximum frequency of the signal. Hence there is no relationship *whatsoever* between the fundamental frequency, f_0 of the analog signal , the maximum frequency, $f_m ax$, and the sampling frequency, F_s picked to create a discrete signal from the analog signal. The same is true for the fundamental period, T_0 of the analog signal and the sampling period T_s. They are not related either. This point can be confusing. T_0 is a property of the signal, whereas T_s is something chosen externally for sampling purposes. The maximum frequency similarly indicates the bandwidth of the signal, from $f_m ax - f_0$.

The Shannon theorem applies, strictly speaking, only to baseband signals or what we call the low-pass signals. There is a complex-envelope version where even though the center frequency of a signal is high due to having been modulated and up-converted to a higher carrier frequency, the signal can still be sampled at twice its bandwidth and be perfectly reconstructed. This is called the band-pass sampling theorem. We won't go into it in this book.

Aliasing of discrete signals

In Fig. 3.11(a) we see discrete samples of a signal and in (b) we see that these points fit several of the waves shown. So which wave or signal did they come from?

The samples in Fig. 3.11(a) could have in fact come from an infinite number of others which are not shown. This is a troubling property of discrete signals. This effect, that many

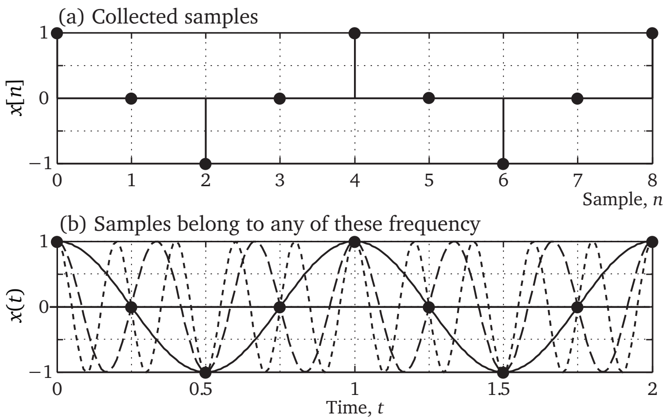

Figure 3.11: Three signals of frequency 1, 3 and 5 Hz all pass through the same discrete samples shown in (a). How can we tell which frequency was transmitted?

different frequencies can be mapped to the same samples, is called **aliasing**. This effect, caused by improper sampling of the analog signal leads to erroneous conclusions about the signal. Later we will discuss how the spectrum of a discrete signal repeats, and it repeats precisely for this reason that we do not know the real frequency of the signal.

Bad sampling

If a sinusoidal signal of frequency f_0 (since a sine wave only has one frequency, both its highest and its lowest frequencies are the same) is sampled at less than two times the maximum frequency, $F_s < 2f_0$, then the signal that is re-constructed although passing through all the samples is an alias, which means it is not the correct one. The expression for all aliases possible for a set of samples is given by

$$y(t) = \sin(2\pi(f_0 - mF_s)t) \tag{3.15}$$

Here m is a positive integer satisfying this equation

$$\left|f_0 - mF_s\right| \leq \frac{F_s}{2} \tag{3.16}$$

These equations are very important but they are not intuitive. So let's take a look at an example.

Example 3.1. Take the signal with $f_0 = 5$ Hz and $F_s = 8$ Hz or samples per second or samps. We use Eq. (3.15) to find the possible alias frequencies. Here are first three for

$(m = 1, 2, 3, \ldots)$ aliases.

$$m = 1 : y(t) = \sin(2\pi(5 - 1 \times 8)t) = \sin(2\pi\underline{3}t)$$
$$m = 2 : y(t) = \sin(2\pi(5 - 2 \times 8)t) = \sin(2\pi\underline{11}t)$$
$$m = 3 : y(t) = \sin(2\pi(5 - 3 \times 8)t) = \sin(2\pi\underline{19}t)$$

The first three alias frequencies are 3, 11, and 19 Hz, all varying by 8 Hz, the sampling frequency. The samples fit all of these frequencies. The significance of m, the order of the aliases is as follows. When the signal is reconstructed, we need to filter it by an anti-aliasing filter to remove all higher frequency aliases. Setting $m = 1$ implies the filter is set at frequency of $mF_s/2$ or in this case 4 Hz. So we only see the frequencies that fit those samples below this number. Higher order aliases although present are filtered out.

Fig. 3.12 which is a spectrum of the reconstructed signal shows the Eq. (3.15) in action. Each m in this expression represents a shift. For $m = 1$, the cutoff point is 4 Hz, which only lets one see the 3 Hz frequency but not 11 Hz or higher. Note the first set of components are at ±3 Hz from the center.

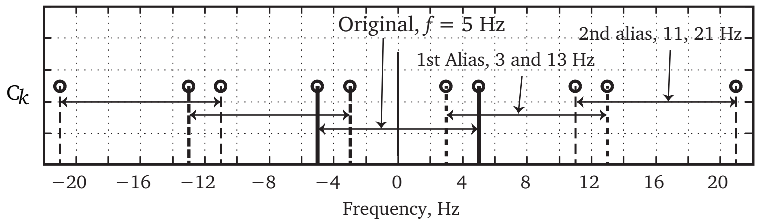

Figure 3.12: The spectrum of the signal repeats with sampling frequency of 8 Hz. Only the 3 Hz signal is below the 5 Hz cutoff.

The fundamental pair of components (the real signal before reconstruction) are at +5 and −5 Hz. Now from Eq. (3.15), this spectrum (the bold pair of impulses at ±5 Hz) repeats with a sampling frequency of 8 Hz. Hence the first move of the pair centered at 0 Hz is now centered at 8 Hz (dashed lines). The lower component falls at $8 - 5 = 3$ Hz and the upper one at $8 + 5 = 13$ Hz. The second shift centers the components at 16 Hz, with lower component at $16 - 5 = 11$ Hz and the higher at $16 + 5 = 21$ Hz. The same thing happens on the negative side. All of these are called alias pairs. They are all there unless the signal is filtered to remove these.

Good sampling

The sampling theorem states that you must sample a signal at twice or higher times its maximum frequency in order to properly reconstruct the signal from the samples. The consequence of not doing that is that we get aliases (from Eq. (3.15)) at wrong frequencies. But what if we do sample at twice or greater rate. Does that have an effect and what is it?

Example 3.2.

$$x(t) = 0.2\sin(2\pi t) + \sin(4\pi t) + 0.7\cos(6\pi t) + 0.4\cos(8\pi t)$$

The signal has four frequencies, which are 1, 3, 3 and 4 Hz. Let's take the signal as shown in Fig. 3.13(a). The highest frequency is 4 Hz. We sample this signal at 20 Hz and then again at 10 Hz. Both of these sampling frequencies are above the Nyquist rate, so that is good. The spectrum as computed by the Fourier series coefficients (FSC) of the 4 frequencies in this signal is shown in Fig. 3.13(b). (We have not yet discussed how to compute this discrete spectrum, but the idea is exactly the same as for the continuous-time case.)

A very important fact for discrete signals is that the Fourier series coefficients repeat with integer multiple of the sampling frequency F_s. The entire spectrum is copied and shifted to a new center frequency to create an *alias* spectrum. This continues forever on both sides of the principal alias, shown in a dashed box in the center in Fig. 3.14. The spectrum around the zero frequency is called the *Principal alias*. Usually this is the one we are looking for.

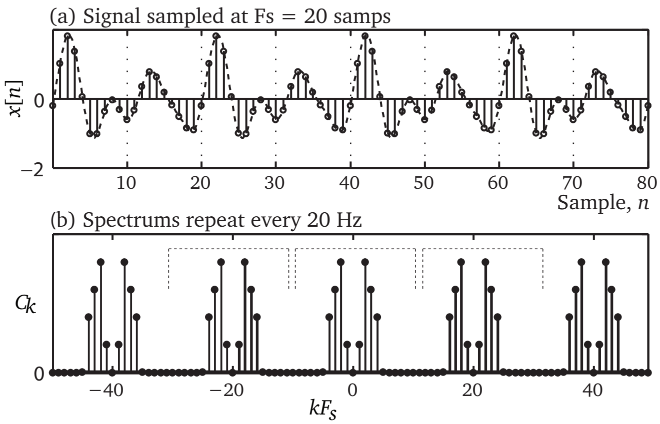

Figure 3.13: A composite signal of several sinusoid is sampled at twice the highest frequency. In (b), we see the discrete coefficients (what we call the spectrum) repeating with the sampling frequency, $F_s = 20$ Hz.The higher the sampling frequency, the further apart are the copies of the spectrum.

In Fig. 3.13, we see the signal sampled at 20 Hz, and we see that there is plenty of distance between the copies. This is because the bandwidth of the signal is only 8 Hz, hence we have 12 Hz between the copies. Fig. 3.14(b) shows the spectrum for the signal when sampled at 10 Hz. The spectrum itself is only 8 Hz wide so there is no overlap but now the spectrum are close together with only 2 Hz between the copies.

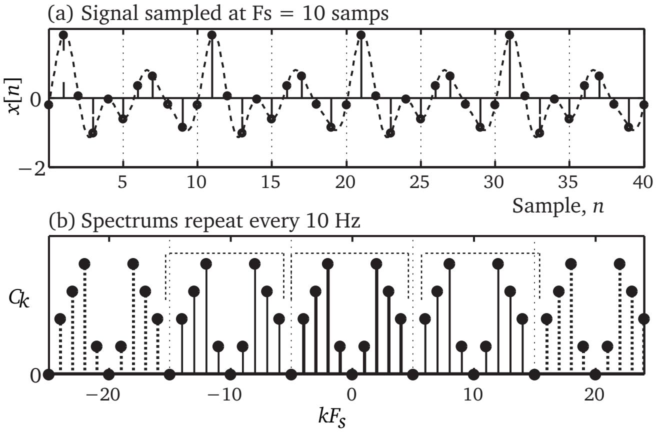

Figure 3.14: The same signal sampled at $F_s = 10$ in (a), results in much closer replications in (b).

Decreasing the sampling rate decreases the spacing between the alias spectrum. The copies would start to overlap if they are not spaced at least 2 times the highest frequency of the signal. In such a case, we would not be able to separate one spectrum from another, making the original signal impossible to reconstruct. When non-linearities are present, the sampling rate must be higher than Nyquist threshold to allow the spectrum to spread but not overlap. The same is true for the effect of the roll-off from the anti-aliasing filter. Since practical filters do not have sharp cutoffs, some guard band has to be allowed. This guard band needs to be taken into account when choosing a sampling frequency.

If the spectrum do *alias* or overlap, the effect cannot be gotten rid of by filtering. Since we do not a-priori have knowledge of the signal spectrum, we are not likely to be aware of any aliasing if it happens. We always hope that we have correctly guessed the highest frequency in the signal and hence have picked a reasonably large sampling frequency to avoid this problem.

However, usually we do have a pretty good idea about the target signal frequencies. We allow for uncertainties by sampling at a rate that is higher than twice the maximum frequency, and usually much higher than twice this rate. For example, take audio signals which range in frequency from 20-20,000 Hz. When recording these signals, they are typically sampled at 44.1 kHz (CD), or 48 kHz (professional audio), or 88.2 kHz, or 96 kHz rates depending on quality desired. Signals subject to non-linear effects spread in bandwidth after transmission and require sampling rates of 4 to 16 times the highest frequency to cover the spreading of the signal.

Discrete signal parameters

There are important differences between discrete and analog signals. An analog signal is defined by parameters of frequency and time. To retain this analogy of time and frequency for discrete signals, we use n, the sample number as the unit of discrete-time. The frequency however gives us a problem. If in discrete-time, time has units of *sample*, then the frequency of a discrete signal must have units of *radians per sample*.

The frequency of a discrete signal is indeed a different type of frequency than the traditional frequency of continuous signals. We call it **Digital Frequency** and use the symbol Ω to designate it. We can show the similarity of this frequency to the analog frequency by noting how we write these two forms of signals.

$$\begin{aligned} \text{Analog signal}: \ x(t) &= \sin(2\pi f_0 t) \\ \text{Discrete signal}: \ x[n] &= \sin(2\pi f_0 n T_s) \end{aligned} \tag{3.17}$$

The first expression is a continuous signal and the second a discrete signal. For the discrete signal, we replace continuous-time, t with nT_s. Alternately we can write the discrete signal as in Eq. (3.18) by noting that the sampling time is inverse of the sampling frequency. (We *always* have the issue of sampling frequency even if the signal is naturally discrete and was never sampled from a continuous signal. In such a case, the sampling frequency is just the inverse of time between the samples.)

$$x[n] = \sin\left(\frac{2\pi f_0}{F_s} n\right). \tag{3.18}$$

Digital frequency, only for discrete signals

Now define the *Digital frequency*, Ω by this expression.

$$\Omega = \frac{2\pi f_0}{F_s} \tag{3.19}$$

Substitute this definition of digital frequency into Eq. (3.18) and we get a sampled sinusoid.

$$x[n] = \sin(\Omega n)$$

Now we have two analogous expressions for a sinusoid, a discrete and a continuous form.

$$\begin{aligned} \text{Analog signal}: x(t) &= \sin(\omega t) \\ \text{Discrete signal}: x[n] &= \sin(\Omega n) \end{aligned} \tag{3.20}$$

The digital frequency Ω is equivalent in concept to the analog frequency, but these two "frequencies" have different units. The analog frequency has *units of radians* per second whereas the digital frequency has units of *radians per sample*.

The fundamental period of a discrete signal is defined as a certain number of samples, N_0. This is equivalent in concept to the fundamental period of an analog signal, T_0. To be considered periodic, a discrete signal must repeat after N_0 samples. In the continuous domain, a period represents 2π radians. To retain equivalence in both domains, N_0 samples hence also cover 2π radians, from which we have this relationship.

$$\Omega_0 N_0 = 2\pi \tag{3.21}$$

The units of the fundamental digital frequency Ω_0 are radians/sample and units of N_0 are just samples. The digital frequency is a measure of the number of radians the signal moves per sample. And when we multiply it by the fundamental period N_0, we get an integer multiple of 2π. Hence a periodic discrete signal repeats with a frequency of 2π which is the same condition as for an analog signal.

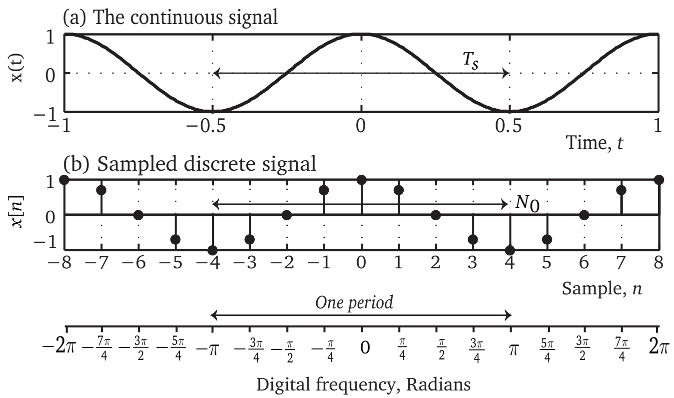

Figure 3.15: A discrete signal in time domain can be referred by its sample numbers, n (1 to N) or by the digital frequency phase advance. Each sample advances the phase by $2\pi/N$ radians. In this example, N is 8. In (a), the x-axis is in terms of real time. In (b) it is in terms of sample identification number or n. In (c) we note the radians that pass between each sample such that total excursion over one period is π.

There are three ways to specify a sampled signal. In Fig. 3.15(a), we show two periods of a signal. This is a continuous signal, hence the x-axis is continuous-time, t. Now we sample this signal. Each cycle is sampled with eight samples, so we show a total of 17 samples, numbered from -8 to $+8$ in Fig. 3.15(b). This is the discrete representation of signal $x(t)$ in terms of samples which are identified by the sample number, n. This is one way of showing a discrete signal. Each sample has a number to identify it.

We can replace the sample number with a phase value for an alternate way of showing the discrete signal. In Fig. 3.15(c), there are 8 samples over each 2π radians or equivalently a discrete angular frequency of $2\pi/8$ radians per sample. This is the digital frequency, Ω_0 that is pushing the signal forward by this many radians. Each sample moves the signal further in phase by $\pi/4$ radians from the previous sample, with two cycles or 16 samples covering 4π radians. Hence we can label the samples in radians. Both forms, using n or the phase are equivalent but the last form (using the phase) is more common for discrete signals, particularly in text books, however it tends to be non-intuitive and confusing.

Are discrete signals periodic?

Fourier series representation requires that the signal be *periodic*. So can we assume that a discrete signal if it is sampled from a periodic signal is also periodic? The answer is strangely enough, no. Here we look at the conditions of periodicity for a continuous and a discrete signal.

$$\begin{aligned} \text{Continuous signal}: \ & x(t) = x(t+T) \\ \text{Discrete signal}: \ & x[n] = x[n+N] \end{aligned} \tag{3.22}$$

This expression says that if the values of a signal repeat after a certain number of samples, N for the discrete case and a certain period of time T for the continuous case, then the signal is periodic. The smallest value of N that satisfies this condition is called the *fundamental period of the discrete signal*. Since we use sinusoids as basis functions for Fourier analysis, let's apply this general condition to a sinusoid. To be periodic, a discrete sinusoid which is defined in terms of the digital frequency and time sample n, must repeat after N samples, hence it must meet this condition.

$$\cos(\Omega_0 n) = \cos\big(\Omega_0(n+N)\big) \tag{3.23}$$

We expand Eq. 3.23 using this trigonometric identity.:

$$cos(A+B) = cos(A)cos(B + sin(A)sin(B)$$

To examine under which condition this expression is true. we set

$$\cos\big(\Omega_0(n+N)\big) = \cos(\Omega_0 n)\cos(\Omega_0 N) - \sin(\Omega_0 n)\sin(\Omega_0 N). \tag{3.24}$$

For this expression to be true, we first need the underlined parts on the RHS to be equal to 1 and 0 respectively.

$$\cos(\Omega_0 n) = \cos(\Omega_0 n)\underbrace{\cos(\Omega_0 N)}_{=1} - \sin(\Omega_0 n)\underbrace{\sin(\Omega_0 N)}_{=0} \tag{3.25}$$

For these two conditions to be true, we must have

$$\begin{aligned}\Omega_0 N &= 2\pi k \text{ or}\\ \frac{\Omega_0}{2\pi} &= \frac{k}{N}\end{aligned} \tag{3.26}$$

We conclude that a discrete sinusoid is periodic if and only if its digital frequency is a rational multiple of 2π based on the smallest period N. This implies that discrete signals are not periodic for all values of Ω_0, nor for all values of N. For example if $\Omega_0 = 1$, then no integer value of N or k can be found to make the signal periodic per Eq. (3.26).

We write the expression for the fundamental period of a periodic discrete signal as

$$\boxed{N = \frac{2\pi k}{\Omega_0}} \tag{3.27}$$

The smallest integer k, resulting in an integer N, gives the fundamental period of the periodic sinusoid, if it exists. Hence for $k = 1$, we get $N = N_0$.

Example 3.3. What is the digital frequency of this signal? What is its fundamental period?

$$x[n] = \cos\left(\frac{2\pi}{5}n + \frac{\pi}{3}\right)$$

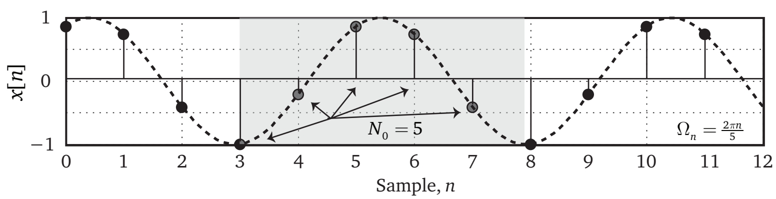

Figure 3.16: Signal of Example 3.3.

The digital frequency of this signal is $2\pi/5$ because that is the coefficient of time index n. The fundamental period N_0 is equal to 5 samples which we find using Eq. (3.27) setting $k = 1$.

$$N_0 = \frac{2\pi}{\Omega_0} = \frac{2\pi}{2\pi/5} = 5.$$

Example 3.4. What is the period of this discrete signal? Is it periodic?

$$x[n] = \sin\left(\frac{3\pi}{4}n + \frac{\pi}{4}\right)$$

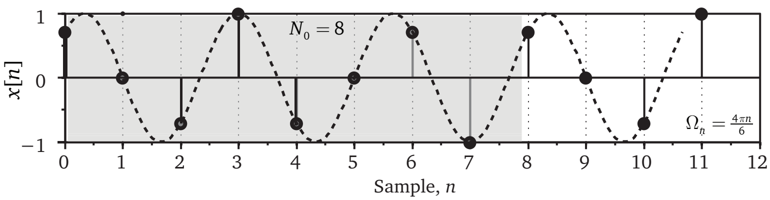

Figure 3.17: Signal of Example 3.4. The period of this signal is 8 samples.

The digital frequency of this signal is $3\pi/4$. The fundamental period is equal to

$$N_0 = \frac{2\pi k}{\Omega_0} = \frac{2\pi(k=3)}{3\pi/4} = 8 \text{ samples}$$

The period is 8 samples but it takes 6π radians to get the same sample values again. As we see, the signal covers 3 cycles in 8 samples. As long as we get an integer number of samples in any integer multiple of 2π, the signal is considered periodic.

Example 3.5. Is this discrete signal periodic?

$$x[n] = \sin\left(\frac{1}{2}n + \pi\right)$$

The digital frequency of this signal is $1/2$. Its period from Eq. (3.26) is equal to

$$N = \frac{2\pi k}{\Omega_0} = 4\pi k$$

Since k must be an integer, this number will always be irrational hence it will never result in repeating samples. The continuous signal is of course periodic but as we can see in Fig. 3.18, there is no periodicity in the discrete samples. They are all over the place, with no regularity.

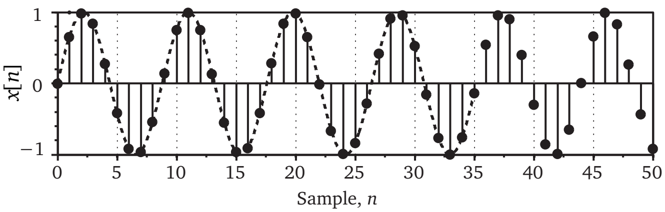

Figure 3.18: Signal of Example 3.5 that never achieves an integer number of samples in any integer multiple of 2π.

Discrete complex exponentials as basis of DT Fourier series

The **continuous-time Fourier series** (CTFS) is written in terms of trigonometric functions or complex exponentials. Because these functions are harmonic and hence orthogonal to each other, both trigonometric and complex exponentials form a basis set for complex Fourier analysis. The coefficients can be thought of as scaling of the basis functions. We are now going to look at the Fourier series representation for discrete-time signals using discrete-time complex exponentials as the basis functions.

A discrete complex exponential is written by replacing t in continuous-time domain with n, and ω with the digital frequency Ω. Now we have these two forms of the CE just as we wrote the two forms of the sinusoid in Eq. (3.20).

$$\text{Continuous form of a CE} : e^{j\omega_0 t}$$
$$\text{Discrete form of a CE} : e^{j\Omega_0 n}$$

We expand the discrete form of the fundamental as follows. The harmonic factor k has not yet been included in this equation.

$$e^{j\Omega_0 n} = e^{j\left(\frac{2\pi}{N}\right)n} \tag{3.28}$$

Harmonics of a discrete fundamental CE

For an analog signal we define its harmonics by multiplying its frequency directly by a multiplier k. Can we do the same for discrete signals? Do we just multiply the fundamental frequency by the index k? Well, no. If a signal has fundamental digital frequency of $\pi/5$, then is frequency $2\pi/5$ the next harmonic? No, this is not how we specify the harmonics of a discrete signal. The range of the digital frequency is just 2π. To obtain its harmonic, we increment its frequency by adding an integer multiple of 2π to it. Hence the frequency of a kth harmonic of a discrete signal is $\Omega_0 + 2\pi k$ (or $(12\pi/5)$). This is a very important point. The analog and discrete harmonics have equivalent definitions for purposes of Fourier analysis. We will see, however that they do not display the same behavior. We can not use these traditionally defined discrete harmonics for Fourier analysis.

$$\text{Discrete fundamental} : e^{j(2\pi/N)n}$$
$$\text{Discrete harmonic} : e^{j(2\pi/N+2\pi k)n}$$

Repeating Harmonics of a discrete signal

Where each and every harmonic of an analog signal looks different, i.e. has higher frequency, shows more ups and downs etc., the discrete harmonics defined by $e^{j(2\pi/N+2\pi k)n}$ are not distinct from each other. They are said to repeat for each harmonic index, k. This is easy to see from the following proof.

$$\begin{aligned}\phi_k[n] &= e^{j(2\pi/N+2k\pi)n}\\ &= e^{jk(2\pi/N)n} + \underbrace{e^{j(2k\pi n)}}_{=0}\\ &= \phi_0[n]\end{aligned} \tag{3.29}$$

Each increment of the harmonic by $2\pi k$ causes the harmonic factor to cancel and result is we get right back to the fundamental!

Example 3.6. Show the first two harmonics of a discrete exponential of frequency $2\pi/6$ if it is being sampled with a sampling period of 0.25 seconds.

The discrete frequency of this signal is $\pi/6$. For an exponential given by $e^{-j\omega_0 t}$, we replace ω_0 with $\pi/6$ and t with $n/4$. ($T_s = .25$) We write this discrete signal as

$$x[n] = e^{-j\frac{\pi}{24}n}$$

Let's plot this signal along with its next two harmonics, which are:

$$\begin{aligned}\text{Fundamental} &: e^{-j\left(\frac{\pi}{24}\right)n}\\ \text{Harmonic 1} &: e^{-j\left(\frac{\pi}{24}+2(k=1)\pi\right)n} = e^{-j\left(\frac{\pi}{24}+2\pi\right)n}\\ \text{Harmonic 2} &: e^{-j\left(\frac{\pi}{24}+2(k=2)\pi\right)n} = e^{-j\left(\frac{\pi}{24}+4\pi\right)n}\end{aligned} \tag{3.30}$$

We plot these two harmonics along with the fundamental in Fig. 3.19. Why is there only one plot in this figure? Simply because the three signals from Eq. (3.30) are identical and indistinguishable.

This example says that for a discrete signal the concept of harmonic frequencies does not lead to meaningful harmonics. All harmonics are the same. But then how can we do Fourier series analysis on a discrete signal if all basis signals are identical? So far we only looked at discrete signals that differ by a phase of 2π. Although the harmonics obtained this way are harmonic in a mathematical sense, they are pretty much useless in the practical

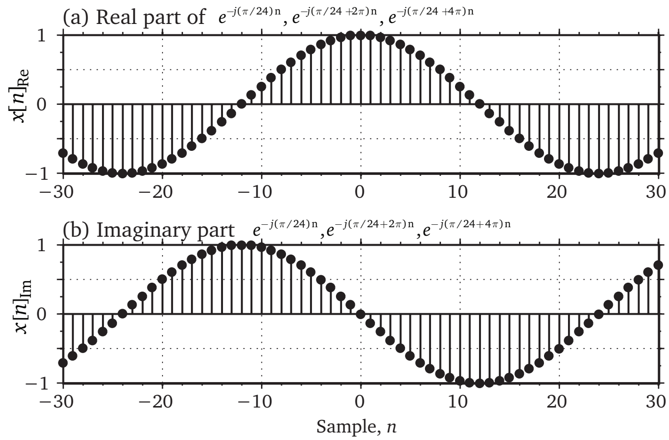

Figure 3.19: Signal of Example 3.6 the imaginary part is a sine wave, (b) the real part, which is of course a cosine. The picture is same for all integer values of k.

sense, being non-distinct. So where are the distinct harmonics that we can use for Fourier analysis?

Here is the secret hiding place of discrete harmonics. They hiding *inside* the 2π range. Here we find N unique harmonics, perfectly suitable for Fourier analysis! These N sub-frequencies are indeed distinct. But there are only N of them, with N being the fundamental period.

Given a discrete signal of period N, the signal will have only N unique harmonics. Each such harmonic frequency is given by

$$\Omega_0[n] = \frac{2\pi}{N}n$$
$$\Omega_k[n] = k\frac{2\pi}{N}n \quad \text{for } k = 0, 1, \ldots, N-1.$$

Increasing k beyond $N-1$ will give the same harmonic as for $k = 0$ again.

Example 3.7. Let's see what happens as digital frequency of a signal is varied just within the 0 to 2π range instead of as integer increment of 2π. Take this signal

$$x[n] = e^{j\frac{2\pi}{6}n}$$

Its digital frequency is $2\pi/6$ and its period N_0 is equal to 6 samples. We now know that the signals of digital frequencies $2\pi/6$ and $14\pi/6$ $(2\pi/6 + 2\pi)$ are exactly the same. So we will increase the digital frequency not by 2π but instead in 6 steps, each time increasing it by $2\pi/6$ so that after 6 steps, the total increase will be 2π as we go from $2\pi/6$ to $14\pi/6$. We do not jump from $2\pi/6$ to $14\pi/6$ but instead are moving in between. We can start with zero frequency or from $2\pi/6$ or 2π as it makes no difference where you start. Starting with 0th harmonic, if we move in six steps, we get these 6 unique signals.

$$\begin{aligned}
\phi_0 &= 2\pi(k=0)/6 = 0 \\
\phi_1 &= 2\pi(k=1)/6 = 2\pi/6 \\
\phi_2 &= 2\pi(k=2)/6 = 4\pi/6 \\
&\vdots \\
\phi_5 &= 2\pi(k=5)/6 = 10\pi/6
\end{aligned}$$

The variable k, the index of the harmonics, steps from 0 to $K-1$. Index n remains the index of the sample or time. Note that since the signal is periodic with N_0 samples, K is equal to N_0.

We can visualize this process as shown in Fig. 3.20 for $N = 6$. This is our not-so-secrete set of N harmonics (within any 2π range) that are unique and used the basis set for discrete Fourier analysis.

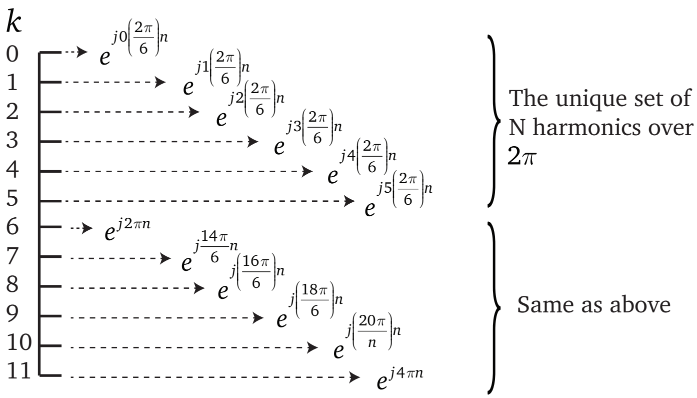

Figure 3.20: Discrete harmonic frequencies in the range of $2\pi m$ to $2\pi(m+1)$.

In Fig. 3.21, we plot the discrete complex exponential so we can examine these. There are two columns in this figure, with left containing the real and right the imaginary part, together representing the complex-exponential harmonic. The analog harmonics are shown in dashed lines for elucidation. The discrete frequency appears to increase (more oscillations of the samples) at first but then after 3 steps (half of the period, N) start to back down again. Reaching the next harmonic at 2π the discrete signal is back to where it started. Further increases repeat the same cycle.

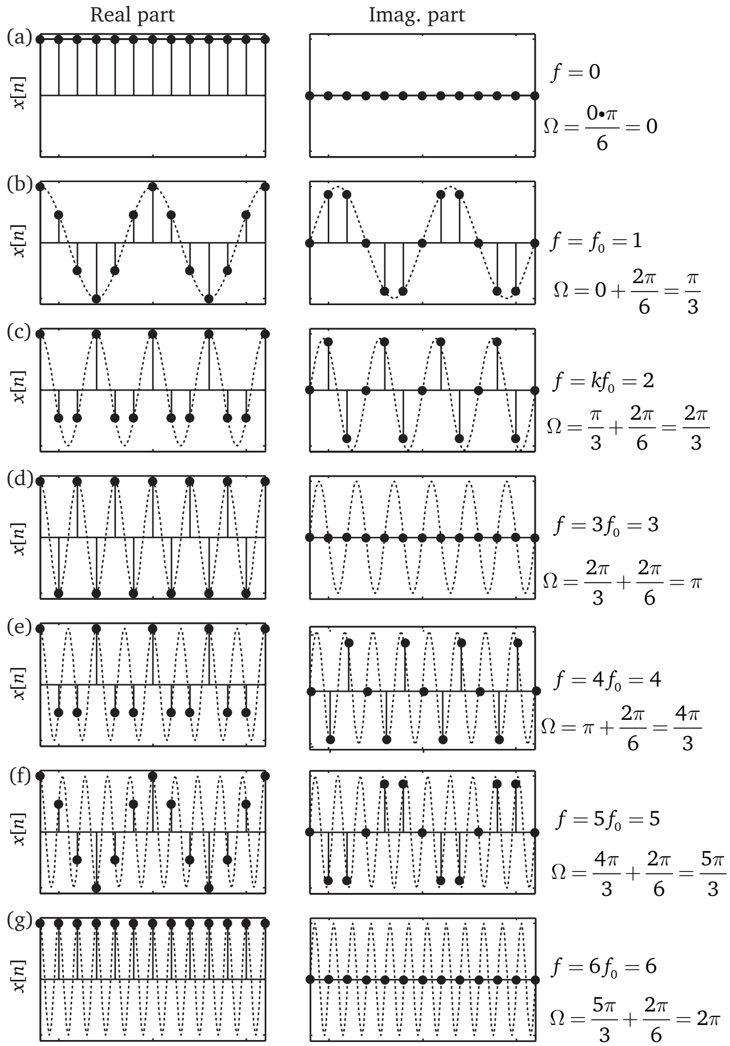

Figure 3.21: The real and the imaginary component of the discrete signal harmonics. They are all different. The discrete frequency appears to increase (more oscillations of the samples) at first but then after 3 steps (half of the period, N) start to back down again. Reaching the next harmonic at 2π, the discrete signal is back to where it started. Further increases repeat the same cycle.

Let's take a closer look. The first row in Fig. 3.21, shows a zero frequency harmonic. All real samples are 1.0, since this is a cosine. In (b), the continuous signal is of frequency 1 Hz, the discrete samples come from $\cos(2\pi/6)$. In (c), we see a continuous signal of 2 Hz and discrete samples from $\cos(2\pi/6)$. We see that by changing the phase from (a) to (g) we have gone through a complete 2π cycle. In (g) the samples are identical to case (a) yet continuous frequency is much higher. The samples for case (g) look exactly the same as for case (a) and case (h) looks exactly the same as case (b) and so on.

These harmonics are an orthogonal basis set and can be used to create a Fourier series representation of a discrete signal. The weighted sum of these N special signals form the discrete Fourier series representation of the signal. Unlike the continuous-time signal, here the meaningful range of the harmonic signal is limited to a finite number of harmonics which is equal to the period of the discrete signal in samples. Hence the number of unique coefficients is finite and equal to the period N.

Discrete-time Fourier Series (DTFS) representation

The **Discrete-time Fourier series (DTFS)** is the discrete representation of a discrete-time periodic signal by a linear weighted combination of these N_0 unique complex exponentials. These unique orthogonal exponentials exist within just one cycle of the signal with cycle defined as a 2π phase shift. Since the number of harmonics available is discrete, the spectrum is also discrete, just as it is for a continuous signal. We write the Fourier representation of the discrete signal $x[n]$ as the weighted sum of these harmonics.

$$\boxed{x[n] = \frac{1}{N_0}\sum_{k=0}^{K_0-1} C_k e^{jk\frac{2\pi}{N_0}n}} \tag{3.31}$$

Eq. (3.31) is the Fourier series representation of a discrete periodic signal. C_k are the complex Fourier series coefficients, which we discussed in Chapter 2. There are k harmonics, hence you see the coefficients with index k. The DTFS coefficients of the $k-th$ harmonic are exactly the same as the coefficient for a harmonic that is an integer multiple of N_0 samples so that:

$$C_k = C_{k+mN_0}$$

The kth coefficient is equal to

$$\boxed{C_k = \sum_{n=0}^{N_0-1} x[n] e^{-j\Omega_0 nk}} \tag{3.32}$$

The $(k+mK_0)$th coefficient is given by

$$\begin{aligned} C_{k+mK_0} &= \sum_{n=0}^{N_0-1} x[n]e^{-j(k+mN_0)\Omega_0 n} \\ &= \sum_{n=0}^{N_0-1} x[n]e^{-jk\Omega_0 n}e^{-jmN_0\Omega_0 n} \end{aligned}$$

The second part of the signal is equal to 1.

$$e^{-jmN_0\Omega_0 n} = e^{-jm2\pi n} = 1$$

(Because the value of the complex exponential at integer multiples of 2π is equal to 1.0). So we have:

$$C_{k+mN_0} = \sum_{n=0}^{N_0-1} x[n]e^{-j(k+mK_0)\Omega_0 n} = C_k \tag{3.33}$$

This proves to us that the harmonics repeat, and hence the coefficients also repeat. The spectrum of the discrete signal (comprised of the coefficients) keeps repeating after every integer multiple of the K_0 samples, or by the sampling frequency. In practical sense, this means we can limit the computation to just the first K_0 harmonics. Again, K_0 is equal to N_0.

This is the reason the discrete Fourier analysis of a sampled signal results in replicating spectrum. It happens because after the first K_0 harmonics, the harmonics repeat and so do their coefficients. We get nothing new. This is a very different situation from the continuous signals, which do not have such behavior. The continuous-time coefficients are unique for all values of k. Discrete signal coefficients repeat because as we allow k to vary over all values of digital frequencies from $-k\pi$ to $+k\pi$, we can no longer tell the harmonics apart. We end up looking at the same K_0 numbers over and over again.

DTFS Examples

Example 3.8. Find the discrete-time Fourier series coefficients of this signal.

$$x[k] = 1 + \sin\left(\frac{2\pi}{10}k\right)$$

The fundamental period of this signal is equal to 10 samples from observation. Hence it can only have at most 10 unique coefficients.

Now we write the signal in the complex exponential form.

$$x[n] = 1 + \frac{1}{2j}e^{j\frac{2\pi}{10}n} - \frac{1}{2j}e^{-j\frac{2\pi}{10}n}$$

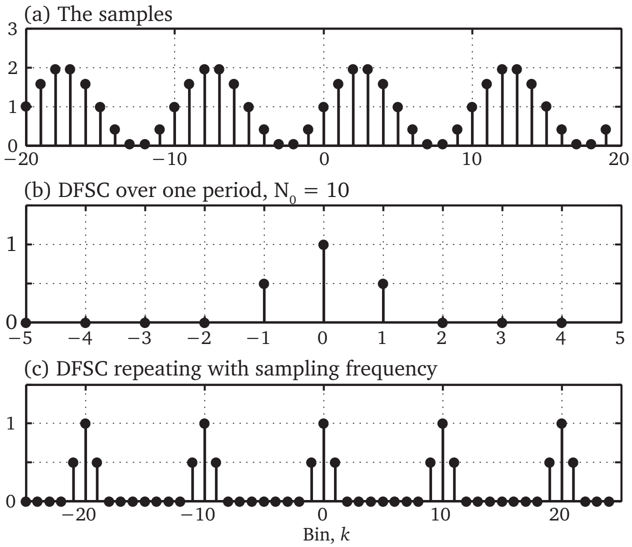

Figure 3.22: Signal of example 3.8 and its Fourier coefficients (a) The discrete signal with period = 10, (b) the fundamental spectrum, (c) the true repeating spectrum.

Note that because the signal has only three components, corresponding to index $k = -1, 0, 1$, for zero frequency and $k = \pm 1$, which corresponds to the fundamental frequency, the coefficients for remaining harmonics are zero. We can write the coefficients as

$$
\begin{aligned}
C_k &= \frac{1}{N_0} \sum_{n=0}^{N_0-1} x[n] e^{-jk\Omega_0 n} \\
&= \frac{1}{10} \sum_{n=0}^{9} x[n] e^{-jk\frac{2\pi}{10}n} \\
C_0 &= \frac{1}{10} \sum_{n=0}^{9} x[n] e^{-j(k=0)\frac{2\pi}{10}n} \\
&= \frac{1}{10} \sum_{n=0}^{9} x[n] \\
&= 1
\end{aligned}
$$

In computing the next coefficient, we compute the value of the complex exponential for $k = 1$ and then for each value of k, we use the corresponding $x[n]$ and the value of the complex exponential. The summation will give us these values.

$$C_1 = \frac{1}{10}\sum_{n=0}^{9} x[n]e^{-j(k=1)\frac{2\pi}{10}n} = \frac{1}{2j}$$
$$C_{-1} = \frac{1}{10}\sum_{n=0}^{9} x[n]e^{-j(k=-1)\frac{2\pi}{10}n} = -\frac{1}{2j}$$

Of course, we can see the coefficients directly in the complex exponential form of the signal. The rest of the coefficients from C_2 to C_9 are zero. However, the coefficients repeat after C_9 so that $C_{1+9k} = C_1$ for all k. We see this in the spectrum of the signal in Fig. 3.23(c). In Fig. 3.23(b) we show only the fundamental spectrum, but in fact the true spectrum is the one in (c). The spectrum repeats every 10 samples forever.

Example 3.9. Compute the DTFSC of this discrete signal.

$$x[n] = \frac{5}{2} + 3\cos(\frac{2\pi}{5}n) - \frac{3}{2}\sin(\frac{2\pi}{4}n)$$

The period of the second term, cosine is 5 samples and the period of sine is 4 samples. Period of the whole signal is 20 samples because it is the least common multiple of 4 and 5. This signal repeats after every 20 samples. The fundamental frequency of this signal is

$$\Omega_0 = \frac{2\pi}{20} = \frac{\pi}{10}$$

We calculate the coefficients as

$$\begin{aligned} C_n &= \frac{1}{N_0}\sum_{n=0}^{N_0-1} x[n]e^{-jk\Omega_0 n} \\ &= \frac{1}{20}\sum_{n=0}^{19} x[n]e^{jk\frac{\pi}{10}n} \\ \Rightarrow C_0 &= \frac{1}{20}x[0]e^{-j\frac{2\pi}{5}} \end{aligned}$$

The Fourier series coefficients of this signal repeat with a period of 20. Each harmonic exponential varies in digital frequency by $\pi/20$. Based on this knowledge, we can see that the $2\pi/5$ exponential falls at $k = 4$, $2(\pi/10) = 2\pi/5$ and exponential $2\pi/4$ falls at $k = 5$, $5\pi/10 = \pi/2$.

We can also write this signal as

$$x[n] = \frac{5}{2} + \left(\frac{3}{2}e^{\frac{2\pi}{5}n} - \frac{3}{2}e^{-\frac{2\pi}{5}n}\right) + j\left(\frac{3}{4}e^{\frac{2\pi}{4}n} - \frac{3}{4}e^{-\frac{2\pi}{4}n}\right)$$

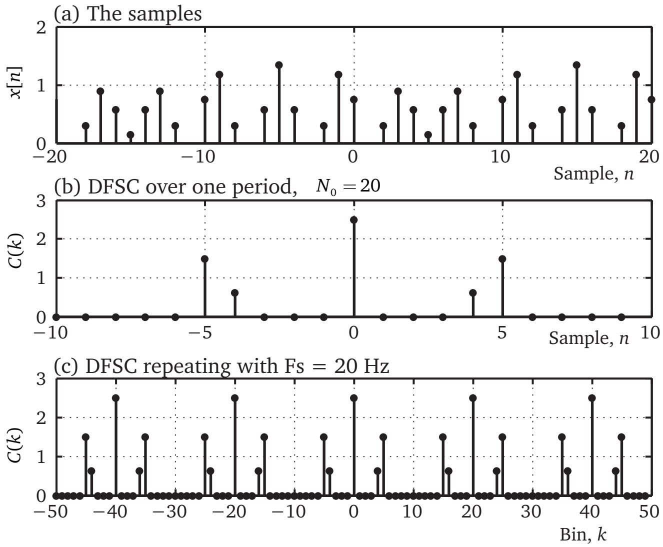

Figure 3.23: Signal of example 3.9 (a) The discrete signal with period = 20, (b) the fundamental spectrum, (c) the true repeating spectrum.

From here, we see that the zero-frequency harmonic has a coefficient of 5/2. The frequency $\pm 2\pi/5$ has coefficients of 3/2 and frequency $\pm 2\pi/4$ has a coefficient of 3/4 as shown in Fig. 3.23(b).

Example 3.10. Compute the DTFSC of a periodic discrete signal that repeats with period = 4 and has two impulses of amplitude 2 and 1, as shown in Fig. 3.24(a).

The period of this signal is 4 samples as we can see and its fundamental frequency is

$$\Omega_0 = \frac{2\pi}{4} = \frac{\pi}{2}.$$

We write expression for the DTFSC from Eq.(3.32).

$$C_k = \sum_{n=0}^{3} x[n] e^{jk\frac{\pi}{2}n}$$

Solving this summation in closed form is hard. In nearly all such problems we need to know series summations or the equation has to be solved numerically. In this case we do know the relationship. We first express the complex exponential in its Euler form. We know values of the complex exponential for argument $\pi/4$ are 0 and 1 respectively for the cosine and sine. We write it in a concise way as

$$\cos\left(\frac{\pi}{2}\right) = 0 \text{ and } \sin\left(\frac{\pi}{2}\right) = 1.$$

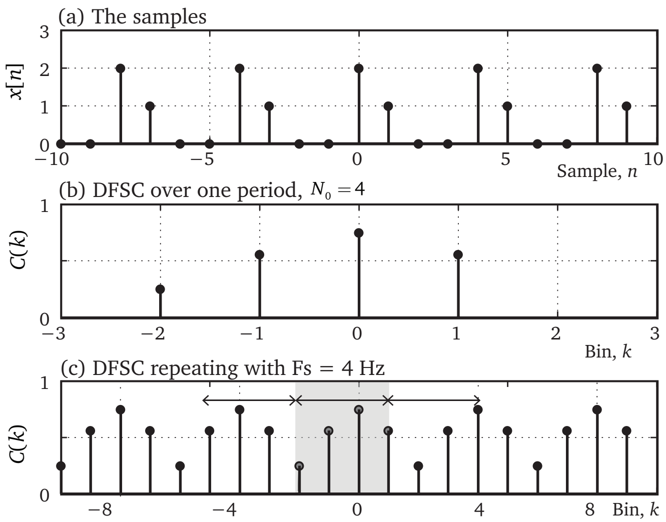

Figure 3.24: Signal of example 3.10. The discrete signal with period = 4 (b) the fundamental spectrum, (c) the true repeating spectrum.

We get for the above exponential

$$e^{-jn\frac{\pi}{2}k} = \left(\cos\frac{\pi}{2} - j\sin\frac{\pi}{2}\right) = (-j)^{nk}.$$

Now substitute this into the DTFSC equation and calculate the coefficients, knowing there are only 4 harmonics in the signal because the number of harmonics are equal to the fundamental period of the signal.

$$C_0 = \left|\frac{1}{4}(2-1)\right| = 0.25$$

$$C_1 = \left|\frac{1}{4}(2-j1)\right| = 0.56$$

$$C_2 = \left|\frac{1}{4}(2+1)\right| = 0.75$$

$$C_3 = \left|\frac{1}{4}(2+j1)\right| = 0.56$$

We see these four values repeated in Fig. 3.24(c).

Example 3.11. Find the DTFSC of the following sequence.

$$x[n] = \{0, 1, 2, 3, 0, 1, 2, 3, \ldots\}$$

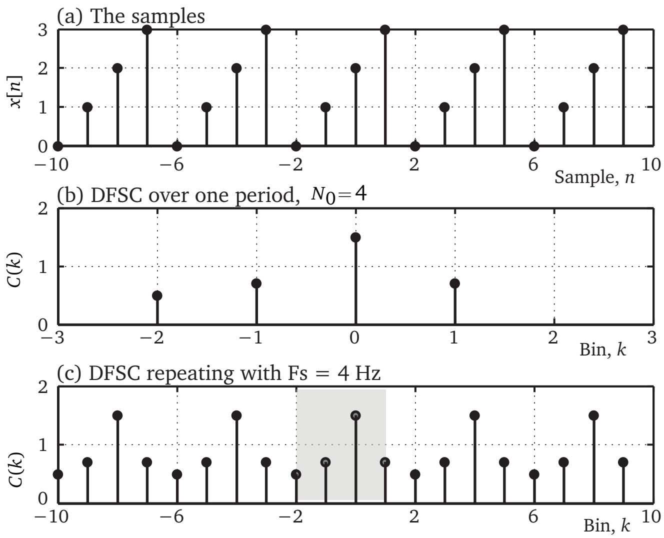

Figure 3.25: (a) The discrete signal with period = 4, (b) the fundamental spectrum, (c) the true repeating spectrum.

The fundamental period of this series is equal to 4 samples by observation. We will now use a compact form of the exponentials to write out the solution.

$$W_4 = e^{-j\frac{2\pi}{4}} = \underbrace{\cos\left(\frac{2\pi}{4}\right)}_{0} - j\underbrace{\sin\left(\frac{2\pi}{4}\right)}_{1} = -j.$$

Note that this W is not a variable but a constant. Its value for these parameters is equal to $-j$. Now we write the coefficients as

$$C_k = \sum_{n=0}^{3} x[n] W_4^{nk}, \quad k = 0, \pm 1, \pm 2, \ldots.$$

From here we get

$$
\begin{aligned}
C_0 &= \sum_{n=0}^{3} x[n]W_4^{0\cdot n} = 0+1+2+3 = 6 \\
C_1 &= \sum_{n=0}^{3} x[n]W_4^{1\cdot n} = \sum_{k=0}^{3} x[n](-j)^n = 0-j-2+j3 = -2+j2 \\
C_2 &= \sum_{n=0}^{3} x[n]W_4^{2\cdot n} = \sum_{k=0}^{3} x[n](-j)^{2n} = -2 \\
C_3 &= \sum_{n=0}^{3} x[n]W_4^{3\cdot n} = \sum_{k=0}^{3} x[n](-j)^{3n} = -2-j2.
\end{aligned}
$$

DTFSC of a repeating square pulse signal

Example 3.12. Find the discrete-time Fourier series coefficients of a square pulse signal of width L samples and duty cycle L/N. Here N is the period of the pulse and L is the width of the pulse itself. The duty cycle is defined as the length of the occupied signal to signal period. It is a ratio from 0 to 1. The pulse is not centered at the origin.

To compute the DTFSC, we will use this important property of geometric series.

$$
\boxed{\sum_{n=0}^{Ł} a^n = \frac{1-a^L}{1-a}, \quad |a| < 1.} \tag{3.34}
$$

This is a very important property and will use to compute the coefficients.

We write the coefficients of this signal as

$$
\begin{aligned}
C_k &= \sum_{n=0}^{L-1} x[n]e^{-j\frac{2\pi}{N}nk} + \underbrace{\sum_{n=L}^{N-1} x[n]e^{-j\frac{2\pi}{N_0}nk}}_{=0} \\
&= \sum_{n=0}^{L-1} 1\cdot e^{-j\frac{2\pi}{N}nk} \\
&= \sum_{n=0}^{L-1} \left(e^{-j\frac{2\pi}{N}n}\right)^k
\end{aligned}
$$

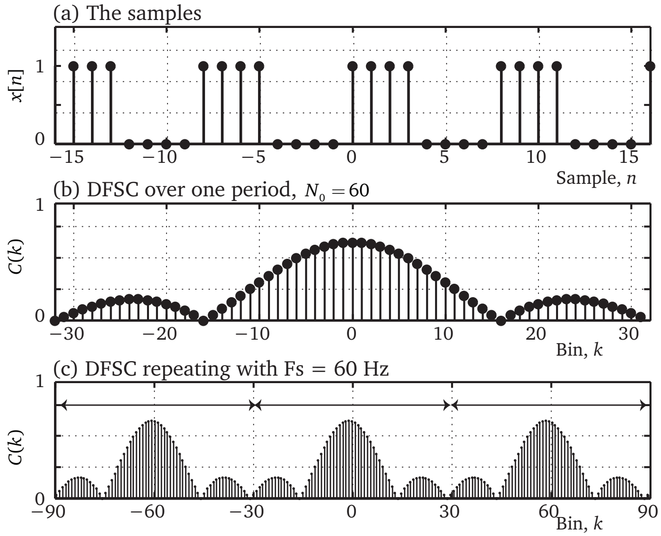

Figure 3.26: (a) The discrete signal, (b) discrete-time Fourier series (DTFSC) coefficients, (c) the true repeating spectrum

Now we use the geometric series by setting a to $(e^{-jk\frac{2\pi}{N}})$. We now have

$$a^n = (e^{-jk\frac{2\pi}{N}})^n$$

Using this term in of Eq.(3.34), we get the following expression for the coefficients.

$$C_k = \frac{1}{N}\frac{1-e^{-jk\frac{2\pi}{N}L}}{1-e^{-jk\frac{2\pi}{N}}}$$

Now we pull out a common term from the numerator to write it as

$$(e^{-jkL\frac{2\pi}{N}\frac{1}{2}})\underline{(e^{jkL\frac{2\pi}{N}\frac{1}{2}} - e^{-jkL\frac{2\pi}{N}\frac{1}{2}})}$$

The underlined part is equal to $2j\ \sin(Lk\frac{\pi}{N})$. Similarly by pulling out this common term, $e^{-jk\frac{2\pi}{N}frac12}$, from the denominator, we get

$$(e^{-jk\frac{2\pi}{N}\frac{1}{2}})\underline{(e^{jk\frac{2\pi}{N}\frac{1}{2}} - e^{-jk\frac{2\pi}{N}\frac{1}{2}})}$$

The underlined term here is similarly equal to $2j\ \sin(k\frac{\pi}{N})$. Note the missing parameter L. From these we now write the coefficients as

$$\frac{1}{N}\frac{1-e^{-j\frac{2\pi}{N}Lk}}{1-e^{-j\frac{2\pi}{N}k}} = \frac{1}{N}\frac{e^{-j\frac{\pi}{N}Lk}\left(e^{j\frac{\pi}{N}Lk}-e^{-j\frac{\pi}{N}Lk}\right)}{e^{-j\frac{\pi}{N}k}\left(e^{j\frac{\pi}{N}k}-e^{-j\frac{\pi}{N}k}\right)}$$

We manipulate this expression a bit more to get the (3.35).

$$C_k = \frac{1}{N}\frac{\sin(kL\pi/N)}{\sin(k\pi/N)}\underline{e^{-jk\frac{\pi}{N}(L-1)}} \tag{3.35}$$

It does not look much simpler! But if we look only at its magnitude (the front part), we get about as simple as we can get in DSP, which is to say not a lot. The DTFSC magnitude for a general square pulse signal of width L and period N samples is given by front part and the phase b The underlined part. Now if you try to plot this function for $k = 0$, you will have a singularity, so for this point, we can compute the value of the function using the L'Hopital's rule, which gives the value of this function as L/N, or the average valve of the function over the period of N samples and certainly that makes sense from what we know of the a_0 value of a Fourier series. It is the DC value.

In Chapter 2 we computed the spectrum of a square pulse signal. The spectrum of that signal was a sinc function, whereas this one repeats because this is a discrete signal. This is an important new development and worth understanding. The function here which is different from sinc is called the Drichlet function. It is a periodic form of the sinc function. In Matlab we would plot the response using the diric function as follows.

```
% DTFSC a square pulse train
N = 10; % Period
n = -15:14
n2 = -3*pi: .01: 3*pi'
L = 5;  % Width of the pulse
mag = abs(diric(2*pi*n/N, L)); % Discrete
mag2 = abs(diric(2*pi*n2/N, L)); % Continuous function
phase = exp(- 1i * n * (L -1) * pi/N)
stem(n, mag); grid on;
hold on;
plot(n2, mag2, '-.b')
```

We plot in Fig **??** the coefficients of various such square pulse signals. These should be studied so you develop an intuitive feel for what happens as the sampling rate increases as well the effect of the duty cycle, i.e. the width of the pulse vs. the period. Note that all of the spectrum repeat with the sampling frequency.

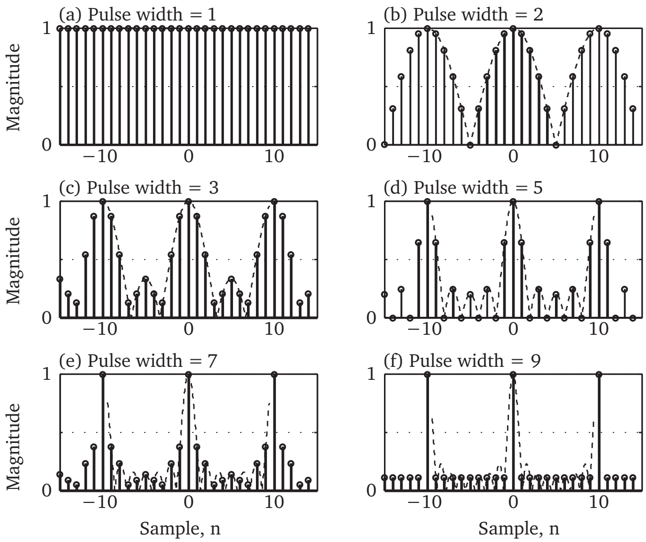

Figure 3.27: Spectrum of a periodic discrete square pulse signal. We see the DTFSC change as the duty cycle (the width of the pulse) of the square pulse increases relative to the period. In (a) all we have are single impulses and hence the response is a flat line. In (f) as the width gets larger, the DTFSC take on an impulse like shape. Note that the underlying sinc does not change.

Fig. 3.27 shows the DTFSC for various widths of the square pulse. As the pulses get wider, the response gets narrower, and when the pulse width is equal to the period, hence it is all impulses, we get an impulse train for a response. This is a very important effect to know.

Power spectrum

In all the examples in this chapter, we have been plotting either the magnitude or the amplitude spectrum. The Power spectrum is a different thing. by Parseval's theorem it is defined as

$$\sum_{n=-\infty}^{\infty} |x[n]|^2 = \sum_{k=-\infty}^{\infty} \left|X[\tfrac{F_s}{n}k]\right|^2 \tag{3.36}$$

For a discrete signal, we can square the time-domain sample amplitudes. If the unit of the amplitude is voltage, the units become voltage-squared. If we add all of these, we get the total power in the signal (assuming a unity resistance.) Or we can sum the square of the amplitudes of each harmonic after computing the coefficients. This also gives us units of voltage-squared and can be used to convert a DTFSC into a Power spectrum by plotting the quantity $(C_k)^2$), instead of the coefficients.

Matrix method for computing Fourier series coefficients

DTFSC are computed in closed form equations for homework problems only. For practical applications, we use Matlab and other digital devices. We will now look at a matrix method of computing the DTFSC. Let's first define this common term. It looks strange and confusing, but it is a very simple idea, we are using it to separate out things that are constant.

$$W_{N_0} \triangleq e^{-j\frac{2\pi}{N_0}}. \tag{3.37}$$

For a given N_0, the signal period in samples, this term W, also called the **Twiddle factor**) is a *constant* and is given a shorthand notation to make the equation-writing easier. We can write the discrete Fourier series (DFS) and inverse discrete Fourier series (IDFS) using this factor as

$$\begin{aligned} \text{DFS: } X[k] &= \sum_{n=0}^{N_0-1} x[n] W_{N_0}^{nk} \\ \text{IDFS: } x[n] &= \frac{1}{N_0} \sum_{k=0}^{N_0-1} X[k] W_{N_0}^{-nk} \end{aligned} \tag{3.38}$$

In this form, the terms $W_{N_0}^{nk}$ and $W_{N_0}^{-nk}$ are same as

$$\begin{aligned} W_{N_0}^{nk} &\triangleq e^{jk\frac{2\pi}{N_0}n} \\ W_{N_0}^{-nk} &\triangleq e^{-jk\frac{2\pi}{N_0}n} \end{aligned}$$

These terms can be pre-computed and stored to make computation quicker. They are the basic idea behind the FFT algorithm which speeds up computation. We can setup the DTFSC equation in matrix form by using the Twiddle factor and writing it in terms of the two variables, the index n and k.

Now we write

$$C_k = \frac{1}{K_0} x[n] \begin{vmatrix} W^{-0\times0} & W^{-1\times0} & W^{-2\times0} & W^{-3\times0} \\ W^{-0\times1} & W^{-1\times1} & W^{-2\times1} & W^{-3\times1} \\ W^{-0\times2} & W^{-1\times2} & W^{-2\times2} & W^{-3\times2} \\ W^{-0\times3} & W^{-1\times3} & W^{-2\times3} & W^{-3\times3} \end{vmatrix} \tag{3.39}$$

Here we have assumed that $N_0 = 4$. Each column represents the harmonic index k and each row the time index, n. It takes 16 exponentiations, 16 multiplications and four summations to solve this equation. We will come back to this matrix methodology again when we talk about DFT and FFT in Chapter 6.

Summary of Chapter 3

In this chapter, we examine discrete signals, the requirements for sampling set by Shanon and Nyquist and methods of reconstruction. Discrete signals resultin frequency ambiguity, as many analog frequencies can fit through the same samples. If not sampled at the proper rate, we get aliasing. The Fourier series representation of discrete signals uses discrete basis functions. If a discrete signal has a period of N samples, then it only has N discrete harmonics within a 2π range. In this chapter, we develop the Fourier series coefficients for discrete signals and look at the inverse process. The coefficients computed repeat with the sampling frequency.

The terms we introduced in this chapter:

- **Discrete signals** - Defined only at specific uniform time intervals.
- **Digital signal** - Discrete signals the amplitude of which is constrained to certain values. A binary digital signal can only take on two values, a 0 or 1.
- **Nyquist rate** - Two times the maximum frequency in the signal
- **Aliasing** - Given a set of discrete samples, a frequency ambiguity exists since infinite number of frequencies can pass through these samples. This effect is called aliasing.

1. An ideal discrete signal is generated by sampling a continuous signal with an impulse train of desired sampling frequency. The time between the impulses is called the sample time.
2. The sampling frequency of an analog signal should be greater than two times the highest frequency in the signal in order to accurately reconstruct the signal.
3. The fundamental period of a discrete periodic signal, given by N_0 must be an integer number of samples for the signal to be periodic in a discrete sense.
4. The fundamental discrete frequency of the signal, given by Ω_0 is equal to $2\pi/N_0$.
5. The period of the digital frequency is defined as any integer multiple of 2π. Harmonic discrete frequencies that vary by integer multiple of 2π, such as $2\pi k$ and $2\pi k + 2\pi n$ are identical.
6. Because discrete harmonic frequencies are identical, we cannot use them to create a Fourier series representation. Instead we divide the range from 0 to 2π by N_0 and use these sub-frequencies as the basis set. They are harmonic and unique.
7. There are only K_0 harmonics available to represent a discrete signal. The number of available harmonics for the Fourier representation, K_0 is exactly equal to the fundamental period of the signal, N_0.
8. Beyond the 2π range of harmonic frequencies, the discrete-time Fourier series coefficients, (DTFSC) repeat because the harmonics themselves are identical.

9. In contrast, the continuous-time signal coefficients are aperiodic and do not repeat. This is because all harmonics of a continuous signal are unique.
10. The coefficients of a discrete periodic signal are discrete just as they are for the continuous-time signals. The coefficients around the zero frequency are called the principal fundamental alias or principal spectrum.
11. The principal fundamental spectrum of a discrete periodic signal, repeats with sampling frequency, F_s.
12. Sometimes we can solve the coefficients using closed from solutions but in a majority of the cases, matrix methods are used to find the coefficients of a signal.
13. Matrix method is easy to setup but is computationally intensive. Fast matrix methods are used to speed up the calculations. One such method is called FFT or Fast Fourier Transform.

Questions

1. Give three examples each of a naturally occurring continuous, discrete and a binary signal.
2. If a signal must be ideally sampled for perfect reconstruction, then how is it we can use binary or M-level sampling and still reconstruct the signal?
3. If sampling consists of multiplying a CT signal with an impulse train, then for reconstruction we are convolving a shape with the sampled function, See Eq. (3.4)? What is the convolution doing?
4. If the largest frequency in a signal is f_1 and the lowest is f_2, then what is the minimum frequency at which this signal should be sampled to be consistent with Shannon's Theorem?
5. What is the difference between Shannon's Theorem and the Nyquist rate.
6. For following three signals, what sampling frequency should be used so that it meets the Nyquist rate.
   (a) $x = \cos(50t)$
   (b) $y = \sin(30\pi t)$
   (c) $z = \sin(31.4(t+2)) - \cos(40\pi t)$
7. What is the digital frequency of a signal given by these samples: $x = [1\ 1\ -1\ 1]$.
8. What is the fundamental period of a sinusoid $\cos(\Omega n + \phi)$, the digital frequency of which is given by: 0.4π, 0.5π, 0.6π, and 0.75π.
9. What is the value of a sinc function at $t = 0; 2; 10$, when $T_s = 2.0$.
10. A discrete signal repeats after 37 samples. Is it periodic? A discrete signal repeats with digital frequency of $2\pi/5$, is it periodic?

11. Stock data is noted every 5 seconds. What is its sampling frequency?
12. Temperature is measured every 10 minutes during the day and every 15 minutes during the night. 108 samples are collected over one day. Can we compute Fourier series coefficients of this data?
13. A signal has a period of 8 samples. What is its fundamental digital frequency?
14. Why would we want to recreate a signal from its samples?
15. Are we able to transmit discrete data over an rf(analog) link?
16. We have a sequence of alternating 0's and 1's. What might its DFS coefficients look like?
17. If you have a CT signal $x(t) = \cos(5t)$ and are told to sample it at 4 times the fundamental frequency, at what rate would you sample this signal?
18. A CT signal is given by $x(t) = \sin(5\pi t)$, if we sample it at a sampling frequency of 20 samples per second, how would you write the discrete version of this signal?
19. A signal is sampled at the rate of 15 samples per second. What frequency is represented by the harmonic index $k = 3$ if the harmonics range from -7 to +7.
20. Digital frequency is limited to a range of 0 to 2π. Why?
21. What is the minimum number of harmonics needed to represent this CT signal; $x(t) = \sin(4\pi t + \pi/5)$, with $F_s = 25$.
22. A discrete signal repeats after every 12 samples. What is its digital frequency? What is its fundamental frequency? What is its period?
23. A signal consists of three sinusoids of periods N = 7, 9 and 11 samples. What is the fundamental period of the composite signal?
24. Why is a signal recreated using sinc reconstruction considered ideal?
25. If $x[n] = 2\cos[2\pi n/5]$, what are its DFS coefficients?
26. If $x[n] = 1 - \sin[2\pi n/8]$, what are its DFS coefficients?
27. What happens to the spectrum of a train of square pulses as the pulses get narrower?
28. Why does the spectrum of a periodic discrete signal repeat? The repetition occurs over what frequency?