The Intuitive Guide to

Fourier Analysis and Spectral Estimation

Charan Langton

Victor Levin

Mountcastle Academic

The Intuitive Guide to Fourier Analysis and Spectral Estimation

The authors gratefully acknowledge so many websites, that freely share their knowledge on the Internet, making this a rich world to both showcase our work and to learn from others.

www.complextoreal.com/fftguide
MATLAB code and other supplemental material will be available at this website.

The Library of Congress cataloging data has been applied for.

ISBN-13: eBook: 9780913063262
ISBN-10: eBook: 0913063266
1. Fourier transform 2. Signal processing 3. Stochastic signal analysis 4. Spectral estimation

First eBook Edition, August 2016
1 2 3 4 5 6
Printed in the United States of America

Book cover design by Josep Blanco
Book interior design and Latex typesetting by Patricio Prada
Editing by Grima Sharma

Contents

Preface **vii**

1 Trigonometric Representation of CT Periodic Signals **1**
What is Fourier Analysis . . . 2
Frequency and Time Domain Views of a Signal . . . 4
Spectrum of a signal . . . 4
Fundamental waves and their harmonics . . . 6
Harmonics as Basis Functions . . . 10
Evenness and oddness of sinusoids . . . 11
Making waves . . . 12
Square waves . . . 12
Gibbs phenomenon . . . 15
Creating a sawtooth wave . . . 15
Generalizing the Fourier Series Equation . . . 16
Multiple ways of writing the Fourier series equation . . . 17
Fourier series in complex exponential form . . . 18
The Fourier Analysis . . . 19
Computing a_0, the DC coefficient . . . 19
Computing b_k, the coefficients of sine harmonics . . . 21
Computing a_k, the coefficients of cosine harmonics . . . 25
Coefficients become the spectrum . . . 29
The power spectrum . . . 32

2 Complex Representation of Continuous-Time Periodic Signals **37**
Euler's Equation . . . 38
The complex exponential . . . 38
Projections of a complex exponential . . . 40

The Sinusoids . . . 42
Fourier Series Representation using Complex Exponentials . . . 45
Double-sided spectrum . . . 50
Power spectrum . . . 58
Appendix: A little bit about complex numbers . . . 64

3 Discrete-Time Signals and Fourier Series Representation 69
Fourier Series and Discrete-time Signals . . . 70
Discrete signals are different . . . 70
Discrete vs. digital . . . 70
Generating discrete signals . . . 71
Sampling and Interpolation . . . 72
Ideal sampling . . . 72
Reconstruction of an analog signal from discrete samples . . . 73
Method 1: Zero-Order-Hold . . . 75
Method 2: First-Order-Hold (linear interpolation) . . . 76
Method 3: Sinc interpolation . . . 77
Sinc function detour . . . 78
Sampling rate . . . 81
Shannon's Theorem . . . 81
Aliasing of discrete signals . . . 83
Bad sampling . . . 83
Good sampling . . . 85
Discrete Signal Parameters . . . 87
Digital frequency, only for discrete signals . . . 88
Are discrete signals periodic? . . . 90
Discrete Complex Exponentials as Basis of DT Fourier Series . . . 92
Harmonics of a discrete fundamental CE . . . 93
Repeating harmonics of a discrete signal . . . 93
Discrete-Time Fourier Series Representation . . . 98
DTFS Examples . . . 99
DTFSC of a repeating square pulse signal . . . 105
Power spectrum . . . 109
Matrix method for computing FSC . . . 109

4 Fourier Transform of aperiodic and periodic signals 115
Applying Fourier Series to Aperiodic Signals . . . 116

Extending the period to infinity . . . 116
Continuous-Time Fourier Transform . . . 118
Comparing FSC and the Fourier Transform . . . 121
CTFT of Important Aperiodic Functions . . . 123
CTFT of an impulse function . . . 123
CTFT of a constant . . . 126
CTFT of a sinusoid . . . 127
CTFT of a complex exponential . . . 128
Time-shifting a function . . . 130
Duality with frequency shift . . . 132
Convolution property . . . 133
CTFT of a Gaussian function . . . 135
CTFT of a square pulse . . . 136
Fourier Transform of Periodic Signals . . . 138
CTFT of a periodic square pulse train . . . 140

5 DT Fourier Transform of Aperiodic and Periodic Signals 147
Discrete Time Fourier Transform . . . 149
DTFT of aperiodic signals . . . 150
Comparing DTFT with CTFT . . . 151
Obtaining a DTFT from CTFT . . . 152
DTFT of a delayed impulse . . . 153
DTFT of superposition of impulses . . . 154
DTFT of a square pulse . . . 156
Dirichlet detour . . . 158
Applying time-shift property to the DTFT of a square pulse . . . 160
Time expansion property . . . 160
DTFT of a triangular pulse . . . 162
DTFT of a raised-cosine pulse . . . 163
DTFT of a Gaussian pulse . . . 164
DTFT of periodic signals . . . 165
DTFT of important periodic signals . . . 168

6 Discrete Fourier Transform 175
Four Cases of Fourier Transform . . . 176
The Discrete Fourier Transform . . . 178
DFT from DTFT . . . 178

DFT frequency resolution . . . 179
Understanding the X-axis of a DFT . . . 181
Periodic signal assumption . . . 182
Computing DFT . . . 183
Computing DFT, the hard way with integration . . . 183
Computing DFT, the easier way with matrix math . . . 186
Computing DFT, the Easy Way With Matlab . . . 191
Zero-padding . . . 192
Better interpolation density vs. better resolution . . . 195
Fundamental Functions and Their DFT . . . 196
DFT of an impulse function . . . 196
DFT of a delayed impulse . . . 197
DFT of a sinusoid . . . 197
DFT of a complex exponential . . . 197
DFT of a square pulse . . . 200
DFT of a finite sinc function . . . 201
DFT repeats with sampling frequency . . . 202
DFT Periodicity and Leakage . . . 202
Truncation effect . . . 204
Truncation causes leakage . . . 205
Reducing leakage by doing a longer DFT . . . 206
Reducing leakage by windowing . . . 207
DFT and Multi-rate Processing . . . 209
Up-sampling and seeing images . . . 209
Down-sampling . . . 212
Interpolation . . . 213
Convolution by DFT . . . 213
Short-Time Fourier Transform . . . 216
DFT in Real-Life . . . 217
Fast Fourier Transform . . . 219

7 Leakage Mitigation with Windows 225
Smearing and Leakage Due to Truncation . . . 226
The Rectangular Truncation Window . . . 228
The rectangular window is always present . . . 229
Window Quality Metrics . . . 233
Filtering with Windows . . . 234

Spectral Estimation 235
Comparing windows 236
The Kaiser window 238

8 Fourier Analysis of Random Signals 245
Fourier Analysis of Deterministic vs. Random Signals 246
Conditions when Fourier analysis is valid 247
Energy and Power Signals 250
Energy spectral density 251
Power dpectral density 254
Characteristics of Random Signals 256
An ensemble and a realization 257
Measure of randomness 257
Stationarity 259
Strict and not so strict stationarity 259
Ergodicity 260
Understanding Properties of Random Signals 262
Time average vs. ensemble Average 262
Probability distribution 263
Moments 265
The Auto-correlation Function 267
Ways of looking at ACF 268
Varieties of auto-correlation function 268
Properties of the ACF 275
Comparing auto-correlation with convolution 277
Some Interesting Signals and Their ACF 279
Barker codes 279
Chirp signal 280
Fourier Transform of the ACF 280

9 Power Spectrum of Random Signals 287
Power Spectrum of a Deterministic signal 288
Spectrum of a Random Signal 289
Nonparametric Spectral Estimation via Autopower 290
ACF bias caused by finite signal length 291
The Measure of an Estimate, MSE, Bias and Variance 293
The Autopower Method and its parameters 294

What is a reasonable length for the signal? . . . 295
Biased vs. unbiased ACF . . . 296
Windowing the ACF . . . 298
How to choose a window . . . 298
Shortening the ACF . . . 299
Blackman-Tukey Algorithm to improve the Autopower . . . 301
The Periodogram . . . 302
Periodogram vs. length of signal . . . 304
Improving the Periodogram . . . 307
Reducing Periodogram variance using the Bartlett Method . . . 309
Reducing Periodogram variance with the Welch Method . . . 311
Bartlett method . . . 311
Welch method . . . 312
Comparing Autopower and Periodogram . . . 312
Figure of Merit of a Spectrum . . . 313

Preface

The Fourier transform shows up in so many different fields, that there isn't an another concept that spans across so many disciplines. Even the famous Heisenberg uncertainty principle is just a restating of the Fourier transform. There is no other mathematical concept that gives one as much bang for the buck as does the Fourier analysis. If you have a strong grasp of this concept, it can help you solve problems in interference, communications, probability theory, cryptography, acoustics, optics, control systems, and the list goes on. Like e and the golden ratio, it is one of those concepts, which defies boundaries and show up everywhere.

In college, not enough time is devoted to this subject. All electrical engineering students either take a Transform theory class or are exposed to it in bits and pieces in a DSP course. In DSP classes, the application of Fourier transform is thrown in with other material such as linear systems analysis. However, DSP only makes sense when you understand Fourier analysis well. All fundamental concepts of DSP are based on Fourier analysis. In mathematics and physics, this subject is covered only cursorily and left as an optional class in other sciences.

This book is intended to give both students and practicing engineers a deeper understanding of Fourier analysis, as a stand-alone topic from its emergence as Fourier series, its application to both analog and discrete signals and finally to spectral estimation using the Fourier transform. While this subject is at the pinnacle of human achievement, engineering education, due to the limited time available, often fails to impart its beauty and scope to students. It often takes years to appreciate this and happens mostly on the job when you have to apply these concepts to something "real."

Our main goal in this book is to help you fully understand what is happening when you compute an FFT of a discrete signal. It is intended to deepen your understanding of the transform theory on which much of numerical analysis is built. There is a whole long story behind this apparently easy-to-compute but hard-to-understand concept. We start the story with the Fourier series in its original trigonometric form as imagined by Baron Fourier, and then progress through all its developments with contributions from other notables along the way to the end point, the spectral estimation of random signals using the discrete Fourier transform. In the last two chapters of this book, we cover application of the Fourier analysis to the non-parametric spectral analysis of random signals.

The Fourier transform, a special case of the Laplace transform, is a fundamental tool for the analysis of stationary signals. In this book, we only cover Fourier analysis and although it leads to all sorts of other important transforms, we feel it is best not to confuse the issue by introducing other transforms. They all deserve a book of their own.

One of the hardest parts of writing this book was deciding where to start. To explain any concept in signal processing requires a good understanding of what comes before it. And of course, to build that understanding means you have to be comfortable with the fundamentals of that concept. To understand the ins and outs of modern-day digital spectral estimation, we start with the basics of sine waves. It is like starting a history lesson about WWII at the Middle ages! But this background is needed to really get into the topic.

Another difficult part is deciding how to describe these concepts. Do we go with the mantra that an equation speaks a thousand words or should a thousand words accompany the equation? Should we repeat ourselves? Take for example, the following sentence which is completely true about the discrete time Fourier transform (DTFT):

“The DTFT is a transformation that maps a discrete-time signal onto a continuous function of ω that spans from 0 to 2π. Alias spectrum appear outside this range.”

It is completely true and yet, transfers no quick shiny nugget of knowledge. Translated into English, it may read as:

“The DTFT is a mathematical machine that takes an input signal as a function of time and produces an output signal as a function of frequency. The input signal is discrete, meaning it only has amplitude values (called samples) at equally spaced time intervals. The output signal is a function of frequency ω and is continuous, so there is a value, although it might be zero, at all frequencies. The output signal is called the spectrum. The output signal ranges from 0 to ∞ radians, but because the spectrum repeats outside of 0 to 2π, we only need this section. The repetition occurs because if there is a sine wave of frequency ω_x that fits the samples, then $(\omega_x + 2\pi k)$ will also fit the samples.”

That is a lot work to write out, and even this is not sufficient to convey full understanding. However, that is what is needed to develop an intuitive understanding. Unfortunately short dense sentences and pages of equations is what students get in the textbooks. Most of the focus in school is also on solving hard, tricky problems or remembering transform pairs. This part of learning is important, but it is to solve homework problems without developing an intuitive understanding is work half-done.

This is a wonderfully simple but deep subject. We want this book to be a supplement for your education so that you will come to appreciate its beauty and depth. We explain what is going on behind the equations. We try to make these *complex* ideas come *alive* with the use of plots. A picture may be worth a thousand words, but we have decided to *also* give you the thousand words. Hence, you may say we examine the signals in a brand new domain, called the *Word Domain* in which we explore our signals in plain language. True understanding and an intuitive feel comes only when you can describe a mathematical concept in *words*.

Our goal is to help you master Fourier analysis from its beginning with the Fourier series all the way to the discrete Fourier transform (DFT) and spectral estimation, in a painless manner.

The first five chapters set the stage for the DFT. We start with the easy to understand trigonometric form of the Fourier series in Chapter 1, and then its more *complex* form in Chapter 2. From there, we go to discrete time signals in Chapter 3 which introduce new complexity to the topic. The development of the Fourier transform from the Fourier series, specifically the continuous time Fourier transform (CTFT) is discussed next. We combine the last two chapters to get to the discrete-time Fourier transform (DTFT) in Chapter 5. From here, it is a manageable leap to the DFT, our main quarry in Chapter 6. From there we spend the last three chapters on how the Fourier transform is used in “real life”. Chapter 7 explains how windows can improve the spectrum by mitigating leakage. Chapters 8 and 9 explain spectral estimation of stationary signals, specifically the non-parametric spectral estimation of random signals.

Altogether this book should help fill in the details and big concepts in Fourier analysis and, importantly, how to use them with comfort and ease. At the end of each chapter, we include some questions to test your conceptual understanding. These questions do not require much calculation, and should be answered verbally as much as possible. You may discuss the answers to these questions at the website for the book, www.complextoreal.com/fftguide.

We want to thank some important people who offered comments, and encouragement. At Loral, we would like to thank Tom Watson and John Walker for helpful technical discussions and advice in many fields over the years. I would like to thank Rick Lyons, the author of my favorite DSP book, who read some of the material and helped me think through a few topics. His book, "Understanding Digital Signal Processing" is my model of how all engineering books should be written. I credit him for being my inspiration. Rena Tishman, my good friend and a lapsed engineer, read, edited, and corrected the chapters over several iterations of the book. I can't thank her enough for her life-long friendship and support. Ryan Bahneman, Stephen Paine and Paul Johnson read the chapters and provided valuable comments. I am also grateful to Dr. Jerry Gibson at UCSB and Dr. Yannis Tsividis at Columbia University for providing valuable comments. An unexpected source of help came from Patricio Parada. Patricio not only designed the book in Latex, being an engineer himself, also corrected errors. We are very grateful for his help. For editing, we would like to thank Grima Sharma, who added linguistic polish to the book. Nina Levin, the publisher and our taskmaster, made sure we got the book done in this century.

We also want to thank The Mathworks, Inc. for their support with MATLAB. The MATLAB code for many of the plots in this book as well as other related materials will be available at the book site.

This is the first edition of this book. Errors are bound to be there! May we ask your consideration in dropping us an email as you find any errors in this version. Thank you.

Charan Langton
Victor Levin
www.complextoreal.com
charldsp@gmail.com

Chapter 1

Trigonometric Representation of Continuous-Time Periodic Signals

Jean-Baptiste Joseph Fourier
1768 – 1830

Jean-Baptiste Joseph Fourier was a French mathematician and physicist. He was appointed to the École Normale Supérieure, and subsequently succeeded Joseph-Louis Lagrange at the École Polytechnique. He is best known for developing the Fourier series and its applications to problems of heat transfer and vibrations. The Fourier transform and Fourier's Law are named in his honor. Fourier also did very important work in the field of astronautics, as well as discovering the greenhouse effect for which he is not so well known. – From Wikipedia

What is Fourier Analysis

When sunlight hits rain-soaked air, an interesting phenomenon happens. Water drops take the ostensibly pure white light and split it into multiple colors. The mathematical description of this process, the subject of this book, was first tackled by Isac Newton approximately 400 years ago. However, even though Newton was able to show that white light is, in fact, composed of other colors, he was unable to make the jump to the idea that light can be described as waves. He called the component colors of white light "specter" or ghosts, from which we get the word spectrum. It took the development of trigonometric series and the recognition of the fact that light can be thought of as a composite wave before Fourier could apply these ideas to the problem of heat transfer. Although the concept of harmonic trigonometric series already existed by the time he worked on the heat transfer problem, Fourier's contribution is considered so important that the whole field of trigonometric waveform analysis and synthesis now bears his name, *Fourier analysis*.

Fourier developed the following partial differential equation called the *Diffusion* equation to describe heat transfer through solids and other media. Here, v is a function representing the measure of heat and K is the heat diffusion constant of the material.

$$\frac{\partial v}{\partial t} = K \frac{\partial^2 v}{\partial x^2} \tag{1.1}$$

Fourier observed that the most general solution to this equation was given as a linear summation of sinusoids, i.e., sine and cosine waves of the form:

$$v(k,x) = \sum_{k=0}^{\infty} \left(a_k \sin kx + b_k \cos kx \right) \tag{1.2}$$

This led Fourier to conclude that an arbitrary wave can be represented as a sum of an infinite number of weighted sinusoids, i.e., sine and cosine waves. This sinusoid summation concept is now known as the **Fourier series**. This book is all about this simple but important idea.

Fourier analysis is applicable to a wide variety of disciplines and not just signal processing, where it is now an essential tool. In addition, Fourier analysis is used in image processing, geothermal and seismic studies, stochastic biological processes, quantum mechanics, acoustics, and even finance.

The Fourier analysis of waves or signals is similar to the concept of compound analysis in chemistry. Instead of atoms coming together to form a myriad of compounds, in signal processing sinusoids can be thought of as doing the same thing. A particular set of these sinusoids is called the **basis set**. Just as a compound may consist of two units of one element and four units of another, an arbitrary wave can consist of two units of one base wave and four units of another. Hence, we can create a particular wave by putting together some basis waves from the set. This process is called **Synthesis**. Conversely, the process of decomposing an arbitrary wave into a set of basis waves is termed **Analysis**. These two complementary and linear processes fall under the name of **Fourier analysis** and its analog, the **Fourier transform**. In Fourier analysis, the basis set of waves is periodic sinusoids.

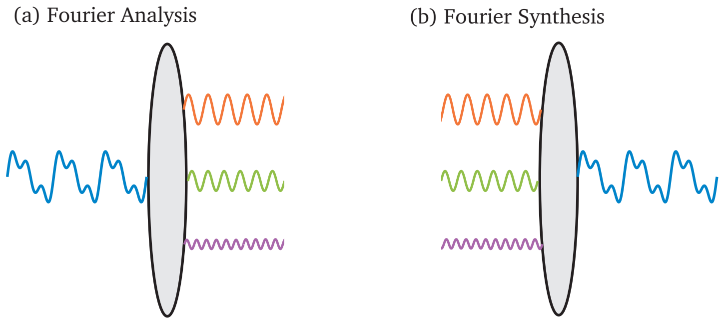

Figure 1.1: Fourier analysis is used to understand composite waves. (a) Analysis: breaking a given signal into sine and cosine components and (b) Synthesis: adding certain sine and cosines to create a desired signal.

Possibly, it was the solution of Eq. (1.1) that led Fourier to notice that summation of harmonic sine and cosine waves leads to some interesting looking periodic waves. From this, he posited that conversely it is also possible to create any periodic signal by the summation of a particular set of harmonic sinusoids. This may not seem like a big thing now but it was a revolutionary insight at the time.

Fourier's discovery was met with incredulity at first. Rightfully so, many of his contemporaries did not accept that his idea was truly **general** and applied to *all* signals. After some years of work by Fourier as well as other famous mathematicians of the age, his theorem was upheld, albeit not under all conditions and not for all types of signals. Subsequent development led to the Fourier transform, the extension of Fourier's original idea to *nonperiodic* signals. However, this computationally demanding concept languished for over 100 years, until the development of the Fast Fourier Transform (FFT), by J.W. Cooley and John Tukey in 1965. The FFT, an algorithmic technique, made the computation of Fourier series simpler and quicker and finally allowed Fourier analysis to be recognized and used widely. It is now the premier tool of analysis in many fields.

Frequency and Time Domain Views of a Signal

Consider the wave in Fig. 1.2. One would be hard pressed to guess its equation. Yet, it is just a sum of three waves as shown in Fig. 1.3(a), 1.3(b) and 1.3(c) of differing frequencies and amplitudes.

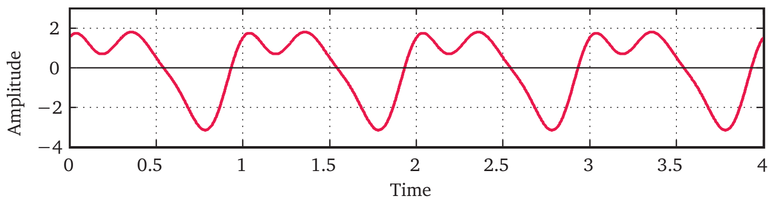

Figure 1.2: An arbitrary periodic wave for which we would like to know its equation.

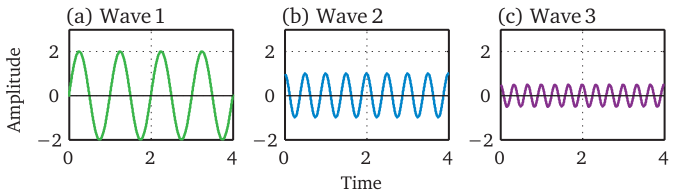

Figure 1.3: The components of the arbitrary wave of figure 1.2.

Therefore, although we know that the wave of Fig. 1.2 is created from the sum of three regular-looking sinusoids of Fig. 1.3, how could we figure this out, if we did not already know the answer?

Spectrum of a signal

Let us look at Fig. 1.4 which shows a general signal in three dimensions. When we see a signal in time, we are looking at it in what is called, the **Time Domain**. What we are actually observing is a *composite* signal. It is a sum of components which we are unable to see. In this three-dimensional view, when we look at a signal from the *side view*, each component appears as a single vertical line at its own discrete frequency. This *side view* of the signal is called the **Frequency Domain**. Another name for this view is the **Signal Spectrum**.

The spectrum is a way to quantify the component frequencies. By *quantify* we mean identifying the amplitude of each of the components. The spectrum of the signal shown in Fig. 1.2 is composed of just three frequencies and can be drawn as illustrated in Fig. 1.5(a). This is called a **one-sided amplitude spectrum**. The x-axis in Fig. 1.5(a) represents the component frequencies of the signal, whereas the y-axis is the amplitude of those frequencies. Fig. 1.5(a)

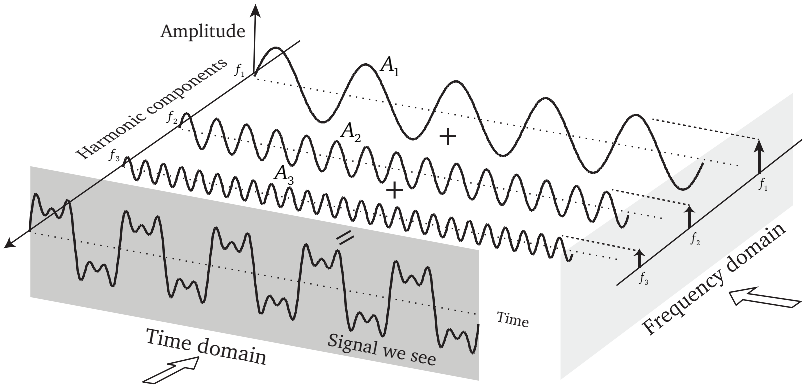

Figure 1.4: A time domain wave seen from the frequency domain provides useful information. The view from the frequency domain is called the spectrum of the wave.

shows the amplitude spectrum with the amplitude of each frequency, whereas Fig. 1.5(b) shows an alternate form of the spectrum called the power spectrum. The **power spectrum**, a more typical representation of the spectrum shows instead of amplitude, the quantity amplitude-squared (which is equal to the instantaneous power), in units of Decibels (dB).

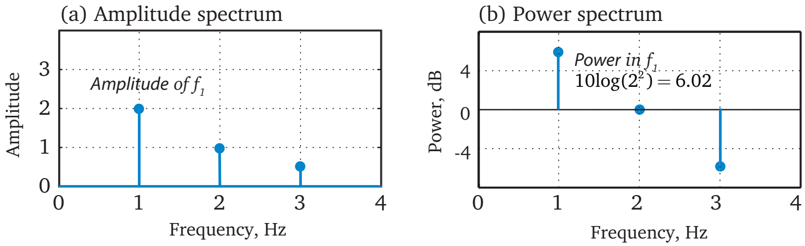

Figure 1.5: The frequency view of the arbitrary wave of Fig. 1.2. (a) The amplitude; and (b) the power in dBs.

We note that Fig. 1.5(a) is in fact the side view from Fig. 1.4, showing the amplitude of each of the components, f_1, f_2 and f_3. The graphical representation, such as the one shown in Fig. 1.5 is a unique *signature* of the signal at a *particular time*. This spectrum is a distinctive and variable quality of a signal. It is *not* a static thing but changes as the signal and its components change. Signals have distinctive spectrum and we can tell a lot about a signal by looking only at its spectrum. The signals for which Fourier analysis is considered valid, must have a non-changing spectrum. This property is generally called **stationarity**.

Fundamental waves and their harmonics

The basic building blocks of Fourier analysis are a set of **harmonic sinusoids**, called the **basis set**. The basis set is our tinker-toy from which we can construct a variety of waves. The set contains an infinite number of sinusoids of differing frequencies related in a special way known as *harmonic*.

The top row of Fig. 1.6 shows a sinusoid of an arbitrary frequency, f_0. Let us call this arbitrary frequency the *fundamental frequency*. We specify some more waves based on this wave called the **harmonics**. Each harmonic frequency is an integer multiple of the frequency of the fundamental. Fig. 1.6 shows that the second wave has half the wavelength and twice the frequency of the first one and so on. Each kth wave has a wavelength of T_0/k and a frequency of kf_0 with k being consecutive integers greater than or equal to 1. All such waves for $k > 1$ are called harmonics of the fundamental. The frequency of the fundamental is of course *arbitrary*, it can be any number whatsoever, but its harmonics are strictly integer multiples of the fundamental frequency, such as: f_0, $2f_0$, $3f_0$, ... kf_0.

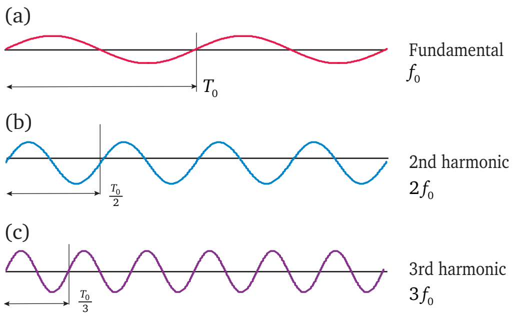

Figure 1.6: An arbitrary fundamental wave and its harmonics.

Let us start with the expression of a set of complex sinusoids; a pair of cosine and a sine of a certain frequency:

$$\begin{aligned} c(t) &= \cos(2\pi f_0 t) \\ s(t) &= \sin(2\pi f_0 t) \end{aligned} \tag{1.3}$$

Here, f_0 is an arbitrary frequency (measured in cycles/second or Hz). We will call this frequency the **fundamental**. The *basic period* T, of this sinusoid is the inverse of the frequency. Note that if a sinusoid is periodic with period T, then it is also periodic with period $2T$, $3T$, etc. for integer multiples of the basic period. From this we obtain the definition of harmonic waves.

A set of waves is harmonic if its frequency is an integer multiple of the fundamental wave's frequency. We can write this as a set.

$$\begin{aligned} c(t) &= \cos(2\pi f_k t) \\ s(t) &= \sin(2\pi f_k t) \end{aligned} \tag{1.4}$$

We have introduced an index, k in Eq. (1.4) such that each harmonic frequency is equal to k times the fundamental frequency, $f_k = kf_0$, with k any arbitrary integer, $0, 1, 2, \ldots, \infty$.

In Fourier series formulation, the index k spans all positive integers to infinity, including zero. Note that the fundamental wave or the first harmonic is often defined as the one for $k = 1$. Hence, the frequency of the first harmonic and the fundamental frequency are the same. The wave obtained for $k = 0$ is of course nothing but a flat line. The wave for $k = 2$ is called the second harmonic and so on for higher values of the index k.

Let us rewrite the definition of the harmonics allowing the **phase** and the **amplitude** to vary. We give each harmonic a unique amplitude and phase and rewrite the harmonic signals as:

$$\begin{aligned} c_k(t) &= a_k \cos(2\pi k f_0 t + \phi_k) \\ s_k(t) &= b_k \sin(2\pi k f_0 t + \phi_k) \end{aligned} \tag{1.5}$$

The wave $c_k(t)$ is a cosine wave of kth harmonic frequency or kf_0, its amplitude being a_k and the phase in radians being ϕ_k. The signal $s_k(t)$ is a sine wave with similarly unique amplitude and phase for the same harmonic frequency, kf_0. Now although the frequencies are still related by multiples of integer k, we are allowing the amplitude and the phase of each harmonic to be different. Such waves are still considered harmonic. The amplitude coefficient a_k for a cosine and b_k for a sine are now arbitrary.

What is **amplitude**? Amplitude is often thought of as the value of a wave's height above its mean value at any particular time. The amplitude can be either positive or negative, indicating where it is being measured, above or below the mean value. If a sinusoid is given by the expression: $a\cos(\omega t)$, then the instantaneous value of the sinusoid (± 1) is scaled by the coefficient, a. The peak value is never more than a, hence this coefficient is the peak amplitude. This coefficient, is called the amplitude of the wave. It is one half of the wave's full excursion above and below its mean. In Eq. (1.5), the terms a_k and b_k are the amplitudes of the waves.

What is **phase**? The argument of a sinusoid $\sin(\theta)$, is in fact an angle, or the term θ. However, in signal processing, we often need to represent a sinusoid as a function of time. We do this by writing the sinusoid as $\sin(\omega t)$, where ω is the radial frequency defined in terms of angles per time. Both forms of the sinusoid argument, the θ and its equivalent form

ωt, are called the *instantaneous phase* of the sinusoid. We can shift a sinusoid, or in fact any wave in time, by adding a term, ϕ, to the instantaneous phase, θ, writing the sinusoid as $\sin(\theta + \phi)$. This second term, ϕ, is assumed to be fixed as a function of time (for linear systems) and is commonly called *the phase*. This can be confusing, because we have two quantities here, both called phase. The total phase at any time is made up of these two parts, one a fixed quantity and the other changing with time. However, generally when we say phase, we are referring not to the instantaneous quantity ωt, but to the fixed quantity, ϕ. This term is more properly called the *phase shift* but the qualifier word *shift* is often left off. In Eq. (1.5), the term ϕ_k is the phase (shift) for the kth wave.

Figure 1.7(a) shows that the amplitude of the sinusoid is 1.0, even though the full excursion is 2.0. Figure 1.7(b) shows that a phase shift of $\phi = \pi/2$ has turned the sine wave into a cosine wave. This phase shift has no effect on the amplitude or the frequency of the wave, hence phase, amplitude and frequency are independent terms.

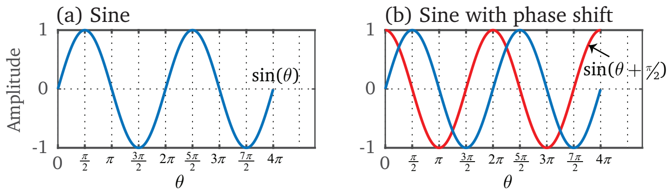

Figure 1.7: The phase of a sinusoid consists of its instantaneous and changing phase, ($\theta = \omega t$), plus a constant phase term, ϕ, called the phase shift as shown in (b).

The sine and cosine waves differ in phase by $\pi/2$ radians, or 90°. When two signals differ in phase by −90° or +90°, they are said to be in *quadrature*, hence sine and cosine waves are in quadrature. For linear Fourier analysis, it is assumed that phase (the ϕ term) and frequency are not a function of time, i.e., both the phase and the frequency of a signal do not change over time. This is one of the fundamental assumptions of Fourier analysis. All signals subjected to Fourier analysis are assumed to have fixed and unchanging frequencies as well as the phase term, ϕ. This assumption implies that the signal is *stationary*, a concept discussed in Chapter 8.

In Fig. 1.8, two harmonics of a sinusoid of frequency 1 Hz are seen, with sines on the left side and cosines on the right. Both the sines and cosines have a peak amplitude of 1. Depending on your field of interest, the units of amplitude can be pretty much anything. In this book, the signal amplitude units will be in volts and phase in radians.

The cosine always reaches its peak amplitude at time $t = 0$, and has a phase of $\pi/2$ at that time. The sine, no matter what the frequency, always has an amplitude of zero at time $t = 0$, which is equivalent to a phase of 0 radians. Hence, no matter how many sine

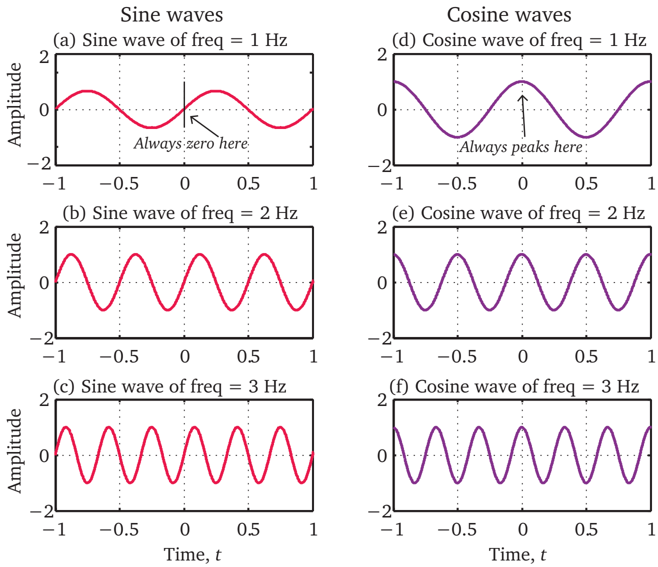

Figure 1.8: The fundamental of $f_0 = 1$ Hz and its two cosine and sine harmonics. All sines start at time $t = 0$, at amplitude $= 0$ and all cosines at the peak amplitude.

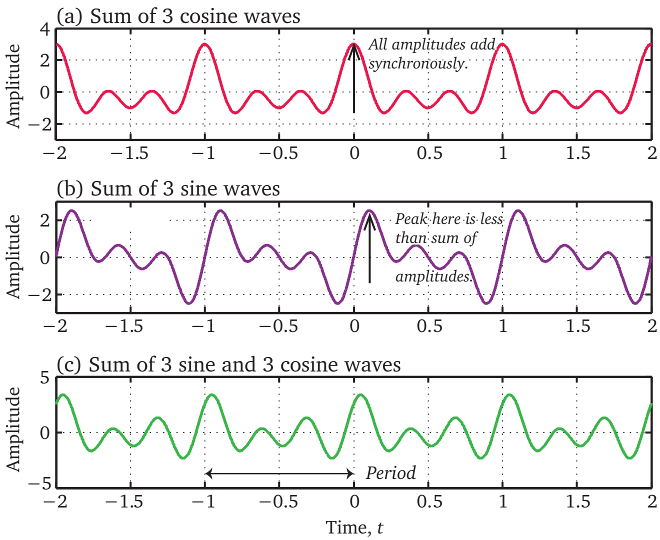

Figure 1.9: Sums of constant amplitude harmonics. (a) Sum of three cosines starts at an amplitude of 3; (b) sum of sines starts at 0; and (c) summation of both sines and cosines also starts at 3 since sines add nothing at $t = 0$.

waves of different frequencies and amplitudes are added together, they will never achieve any amplitude other than zero at time = 0, and can represent only those waves that are zero-valued at time $t = 0$. Similarly, the addition of various cosines will not achieve an amplitude of 0 at time $t = 0$.

Harmonics as Basis Functions

In Fig. 1.9 three cosine and three sine waves with amplitude of 1 are added together, respectively. After the addition of these three waves, it is seen that the cosines in Fig. 1.9(a) add such that the peak is equal to 3. This happens because the cosine is an *even* wave. The cosines add *constructively* at time $t = 0$ and at other times that are integer-periods away. The sum of the three sine waves added together in Fig. 1.9(b) however, looks strange. This asymmetry is a consequence of the sine wave being an *odd* wave. In Fig. 1.9(c), the three sine and three cosine waves are all added together. The behavior of this summation defies an easy explanation.

To examine this effect further, we add together an even larger number, 20 cosine and 20 sine harmonics; each of the same amplitude and 0 phase. Fig. 1.10(a) shows clearly that the cosine sum is symmetric about the time $t = 0$ point. Similarly, Fig. 1.10(b), shows that the sine sum is non-symmetric. Now, we add both of these waves per the following equation:

$$s(t) = \sum_{k=0}^{\infty} \Big| \cos(k\omega_0 t) + \sin(k\omega_0 t) \Big|$$

The summation of harmonic sine and cosine waves of equal amplitudes gives a function that is approaching an impulse train, i.e., a signal consisting of very narrow pulses. This result in Fig. 1.10(c) gives a clue to an important property of sinusoids that will be used in subsequent chapters. This property says that a summation of an infinite number of harmonic sinusoids (sine and cosine) results in an impulse train.

What if we allow the amplitudes and phases of these sinusoid to vary? An example of a wave created using this idea is shown in Fig. 1.11, where each sine and cosine wave has a different amplitude and a different phase. Note that the frequency of the composite wave is equal to the frequency of the fundamental, which is 1 Hz.

$$\begin{aligned} s(t) = 0.1\cos(2\pi t - 0.5) + 0.3\cos(4\pi t) - 0.4\cos(6\pi t - 0.1) \\ - 0.5\sin(2\pi t + 0.1) - 0.8\sin(4\pi t - 0.3) + 0.67\sin(6\pi t + .19) \end{aligned} \tag{1.6}$$

The most important thing to note is that by adding any number of harmonics, and allowing the amplitude and phase of each to vary, we can create or mimic many other waves. Figure

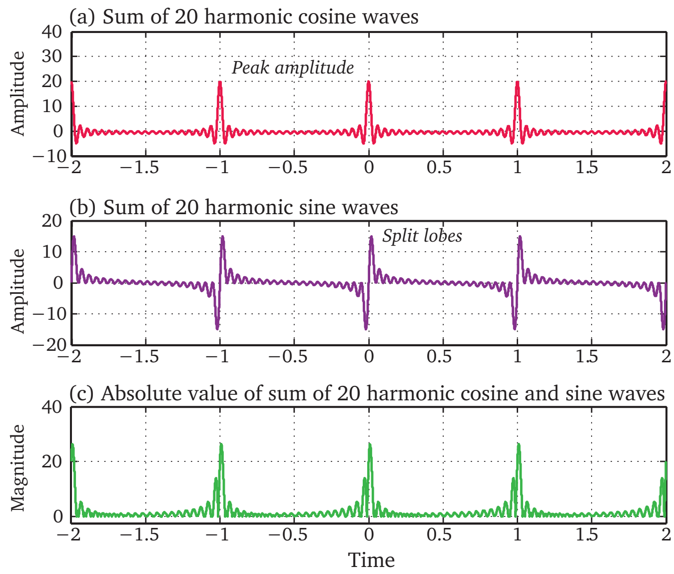

Figure 1.10: Sums of large numbers of both sine and cosine harmonics approach an impulse train. Note that the sum of cosines creates an even blip (a), whereas for sines it is an odd blip (b). In (c), we see the absolute value of the sum (a) and (b).

1.11 shows an example of just one such "interesting" looking wave created by using only three different sinusoids of distinctly different amplitudes and phases. This is main the idea behind Fourier synthesis and analysis.

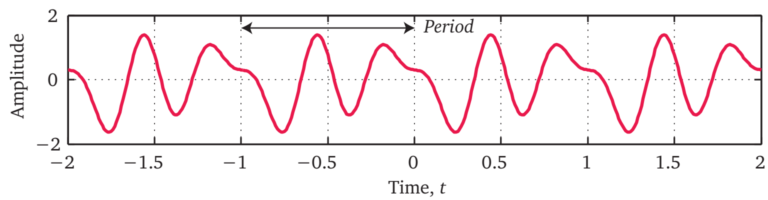

Figure 1.11: A wave comprising arbitrary amplitude harmonics of Eq. (1.6) *begins to look like a real signal.*

Evenness and oddness of sinusoids

Sinusoids have many interesting properties, a few of which are very useful in Fourier analysis. One important property of harmonic sine and cosine waves is asymmetry or *oddness* of the wave. All *sine waves* are considered odd functions because they obey the following definition of an odd function.

$$f(x) = -f(-x) \tag{1.7}$$

If you look at a sine wave, you see that it starts with an amplitude of 0 at time $t = 0$. If we were to flip it about the y-axis, the images would not overlap. However, if one of the sides was first flipped about the x-axis, then they do overlap. This describes the *oddness* of signals. It requires two flips for values to coincide, as we can see from the two negative signs in Eq. (1.7).

The cosine waves on the other hand are called *even* functions by a similar definition. The two sides of a cosine wave, if flipped about the y-axis, would overlap, hence there is only one negative in the equation below for even symmetry.

$$f(x) = f(-x) \tag{1.8}$$

By the superposition principle, if multiple odd waves are summed together, the resulting wave will remain *odd*. In contrast, if multiple *even* waves are summed, then the resulting wave will remain even, and a mixture will have no symmetry. This becomes important when synthesizing, which is the process of putting some waves together to make a desired wave. If a wave to be synthesized is purely an odd, or an even function, then it will only contain sines, or cosines, respectively, depending on its symmetry.

Making waves

Using the idea of harmonic summation, we can create a variety of waveforms. All we need to do is to change the amplitudes and the phases of the harmonics as we see fit. Certain combinations of these parameters lead to great-looking and useful waves that are periodic with the frequency of the fundamental.

Square waves

Now we examine the construction of a square wave. A square wave is not actually square in any particular sense. It is a wave where each period looks somewhat like two rectangles of opposite signs, as seen in Fig. 1.12(d). Yes, it has wiggles in it, and, it is not actually square. The square waves are very useful in signal processing and are used for data transmission. It is amazing that we can create them by just adding a bunch of sinusoids. The more sinusoids we add to the summation, the better the wave looks, with the wiggles getting smaller.

In Fig. 1.12, we have created a square wave by adding together only a few harmonic sinusoids. As the square wave shown in Fig. 1.12(a) is an odd function, we know that only sine waves are needed in its construction, a process also called synthesis. In Fig. 1.12(b), we have added to the fundamental of frequency ω_0, only one sine wave of frequency $3\omega_0$ and amplitude of $1/3$. In Fig. 1.12(c), we add one more sine wave of frequency $5\omega_0$ and of amplitude $1/5$, and in the last case another sine wave of frequency $7\omega_0$ and of amplitude $1/7$ is added. This last result looks good.

Figure 1.13(a) shows the same type of wave but shifted so that it is an even wave, per Eq. (1.8). This even wave can be created using just the cosines. For both cases, we start with a sinusoid of the fundamental frequency. This seems like a good start. But why? Because a periodic wave created from the addition of harmonics will always be periodic with the period of the fundamental, the lowest frequency of the signal. We then add more harmonic sine waves for the odd version of the square wave in Fig. 1.12(a) and cosine waves for the even version in Fig. 1.13(a) and watch the evolution of the square wave. With only three terms in the addition, the results are pretty decent looking square waves.

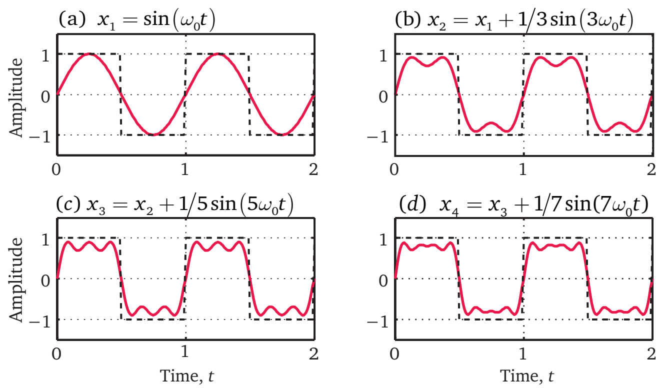

Figure 1.12: Synthesizing an odd square wave by adding odd harmonics of differing amplitudes. (a) Start with fundamental of 1 Hz, (b) add harmonic of 3 Hz, (c) add harmonic of 5 Hz, (d) add harmonic of 7 Hz.

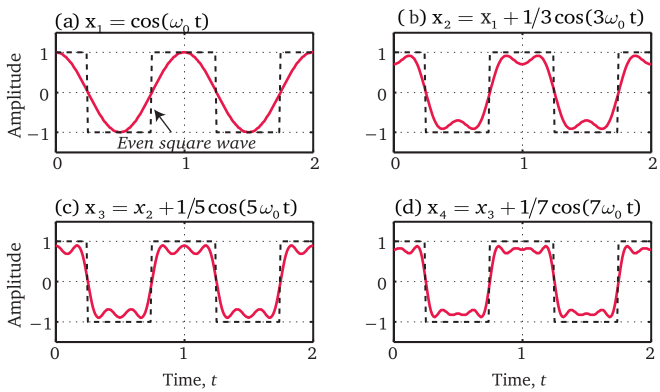

Figure 1.13: Synthesizing an even square wave by adding even harmonics of differing amplitudes.

Each addition of a sine wave (or a cosine) with a specific frequency and amplitude makes the synthesized wave appear closer to a square wave. We are, in fact, cooking up interesting recipes for making all kinds of waves using specific "quantities" of sinusoids. The quantities we vary are the amplitude and the phase of each harmonic. Collectively, the *amplitude* and the *phase* of a particular harmonic is called its **coefficient**. So to create a particular wave, we are controlling or changing the coefficients of the harmonics.

Here are the *recipes* for the two types of square waves, the odd and the even. The ingredient list is limited, as we used only three terms beyond the fundamental, which is underlined.

$$\begin{aligned} x_1 &= \underline{\sin(\omega_0 t)} + \frac{1}{3}\sin(3\omega_0 t) + \frac{1}{5}\sin(5\omega_0 t) + \frac{1}{7}\sin(7\omega_0 t) \\ x_2 &= \underline{\cos(\omega_0 t)} - \frac{1}{3}\cos(3\omega_0 t) + \frac{1}{5}\cos(5\omega_0 t) - \frac{1}{7}\cos(7\omega_0 t) \end{aligned} \tag{1.9}$$

Of note, in both of these "recipes", only odd harmonics (i.e. the harmonic index k is odd.) are used. This is true for both the odd and even versions of the square wave. The reason why only odd harmonics are used is that sinusoids of even harmonics cancel the odd harmonics as we see in Fig. 1.14 and hence mixing of odd and even harmonics does not allow us to create useful waves. We see in Fig. 1.14 that the two even harmonics of frequencies 2 and 4 Hz are poles apart from the odd harmonics of frequencies 1 and 3 Hz (at t = .5 sec.). We find that summation of consecutive harmonics begins to approach an impulse-like signal, as we see in Fig. 1.10 due to this destructive behavior.

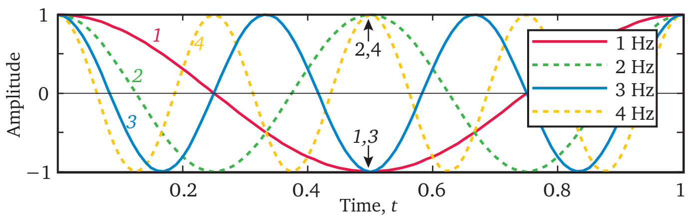

Figure 1.14: Even order harmonics are destructive and hence are not used in combination with odds to create most communications signals such as square waves. Note here that frequencies 1 and 3 Hz are antipodal to frequencies of 2 and 4 Hz.

The square wave is probably the single most important wave in signal processing and we will keep coming back to it in each chapter. The summation with k as the odd integer can go from $-\infty$ to $+\infty$ to form a really great looking square wave. The more terms we add, the closer we get to what we are trying to achieve. But in fact for square waves, we are never able to create a perfect square wave, as Gibbs phenomena takes hold near the corners. The corners never do become the true right angles as we would wish. This tells us that this form of harmonic representation, despite our best efforts, may not result in a perfect reconstruction for every signal, for example square waves.

Gibbs phenomenon

The square waves created in Fig. 1.12 and Fig. 1.13 have a fair amount of ripple (formal name for the wiggle) in the flat parts. We can increase the number of terms to see if it goes away. Figure 1.15 shows that even as K, the number of terms in the square wave summation is increased, the overshoot at the corners never goes away. This behavior, called the **Gibbs phenomenon** is a clear demonstration that Fourier representation can not accurately reconstruct all waves. Hence, waveforms that have hard discontinuities in amplitude, are avoided in signal processing. Instead of sharp-edged square waves, we use shaped waves with gentle corners and curves, ones that are possible to represent with summation of sinusoids.

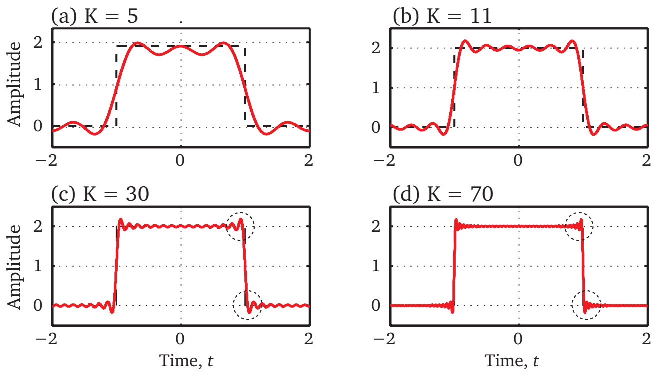

Figure 1.15: Gibbs phenomenon does not allow for a perfect Fourier representation of a square wave.

Creating a sawtooth wave

Let us look at one more special signal, a sawtooth wave. The sawtooth wave is an odd function, hence, composed only of sine waves. Its equation is given as

$$\text{sawtooth}(t) = \frac{2}{\pi}\sum_{k=1}^{\infty}(-1)^{k+1}\frac{\sin(2\pi k f_0 t)}{k} \tag{1.10}$$

Note that in Eq. (1.10) the coefficient, $\frac{(-1)^{k+1}}{k}$ is inversely proportional to the harmonic index, k and hence the contribution of higher frequencies is decreasing. This is also true for the

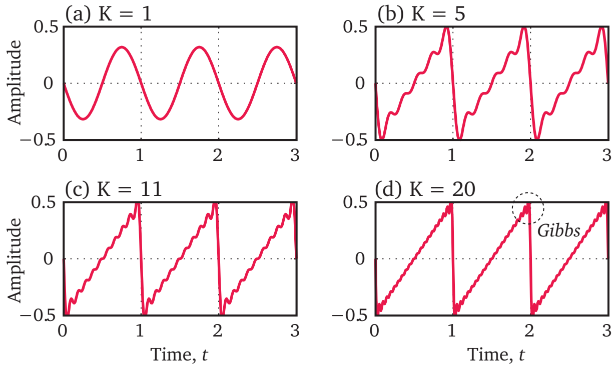

Figure 1.16: Evolution of a sawtooth wave, (a) fundamental sinusoid of desired frequency, (b) addition of five harmonics, (c) addition of eleven harmonics, and (d) with addition of 20 harmonics, a pretty decent looking saw-tooth wave presents itself.

square waves. We write this equation allowing the index k to go to ∞. In real life, we can often get by with just a few terms. Notice that Gibbs phenomena is present at the corners for this representation also. Hence, smooth waves lend to best Fourier synthesis.

Generalizing the Fourier Series Equation

A Fourier series is a general equation consisting of the summation of weighted harmonics, whereby manipulating the weighting (hence coefficients) we can represent any periodic wave. We called it a recipe but its mathematical name is **Fourier representation**.

$$f(t) = \sum_{k=1}^{K} a_k \cos(2\pi k f_0 t) + \sum_{k=1}^{K} b_k \sin(2\pi k f_0 t) \tag{1.11}$$

The coefficient a_k represents the coefficient of the kth cosine wave, and b_k of the kth sine wave. What is K? This is the largest harmonic index we use in any particular summation. If we use only 10 terms (or harmonics), then $K = 10$. We can control this parameter depending on the accuracy desired. For a general representation, we set K to ∞.

The sum of sines and cosines is always symmetrical about the x-axis so there is no possibility of representing a wave with a DC offset using the form in Eq. (1.11). (*The term DC comes from direct current but it is used in signal processing to mean a constant.*) To create a wave of nonzero mean, a new term must be added to Eq. 1.11. The constant, a_0 is included in

Eq. (1.12) so we can create waves that can move up (or down) from the x-axis.

$$\boxed{f(t) = a_0 + \sum_{k=1}^{\infty} a_k \cos(2\pi f_k t) + \sum_{k=1}^{\infty} b_k \sin(2\pi f_k t)} \tag{1.12}$$

Eq. (1.12) is called the **Fourier series equation**. The coefficients a_0, a_k, b_k are called the **Fourier Series Coefficients (FSC)**. The process of Fourier analysis consists of computing these three types of coefficients, given an arbitrary periodic wave, $f(t)$.

Multiple ways of writing the Fourier series equation

There are several different forms of the Fourier series equation in literature, and that can make understanding this equation harder.

The following representation uses the radial frequency, $\omega_k = 2\pi f_k$ to make the equation simpler to type. We can write this form of Fourier series as:

$$\boxed{f(t) = a_0 + \sum_{k=1}^{\infty} a_k \cos(\omega_k t) + \sum_{k=1}^{\infty} b_k \sin(\omega_k t)} \tag{1.13}$$

Now, we define T_0 as the period of the fundamental frequency.

$$T_0 = \frac{1}{f_0}$$

Then the period of the kth harmonic becomes T_0/k and its frequency, $f_k = k/T_0$. We can alternately write the Fourier series equation by adopting this form of the frequency in Eq. (1.14).

$$\boxed{f(t) = a_0 + \sum_{k=1}^{\infty} a_k \cos\left(\frac{2\pi k}{T_0} t\right) + \sum_{k=1}^{\infty} b_k \sin\left(\frac{2\pi k}{T_0} t\right)} \tag{1.14}$$

We can shorten the Fourier series equation by starting at zero frequency, hence index k starts at 0 instead of 1. Now the DC term disappears as it is included as the zero frequency coefficient obtained by setting the index $k = 0$. Here is this form with the DC term gone:

$$\boxed{f(t) = \sum_{k=0}^{\infty} a_k \cos(2\pi f_k t) + b_k \sin(2\pi f_k t)} \tag{1.15}$$

This can be simplified further by getting rid of the sine term altogether. Sines and cosines are really the same thing, one is just the shifted version of the other. The representation in Eq. (1.13) can be written solely with cosines with a shift. This way sine becomes a cosine with a $\pi/2$ shift and we get this form of the equation.

$$\boxed{f(t) = a_0 + \sum_{k=1}^{\infty} c_k \cos(\omega_k t + \phi_k)} \tag{1.16}$$

In this form, each harmonic, whether a sine or a cosine, can be thought of as a cosine of some phase. Hence each harmonic is represented by two cosines, one with a zero phase shift and the other with a shift of $\pi/2$. Now the whole expression uses only cosine wave and the index, k spans not from 0 to 1, but from $-\infty$ to $+\infty$.

$$f(t) = c_1 \cos(\omega_1 t \pm \phi_1) + c_2 \cos(\omega_2 t \pm \phi_2) + c_3 \cos(\omega_3 t \pm \phi_3) + \ldots$$

Fourier series in complex exponential form

In its *most* important representation, the *complex representation*, the Fourier series is written as:

$$\boxed{f(t) = \sum_{k=-\infty}^{\infty} C_k \underline{e^{j2\pi \frac{k}{T_0} t}}} \tag{1.17}$$

Here, we introduce a new term, the complex exponential, as underlined in Eq. (1.17). The complex exponential (CE) represents both a sine and a cosine in one concise form. In the next chapter, we will discuss this function in detail. The expanded form of the Fourier series in terms of the complex exponential looks like this:

$$f(t) = C_0 + C_1 e^{j2\pi \frac{1}{T_0} t} + C_2 e^{j2\pi \frac{2}{T_0} t} + C_3 e^{j2\pi \frac{3}{T_0} t} + \ldots$$

This form shows that we can create a periodic function by summing together complex exponentials. Although the complex form of the Fourier series is scary looking, it is the most commonly used form. In the next chapter, we will look at how it is derived and why we use it in Fourier analysis. All these different representations of the Fourier series are *identical* and mean exactly the same thing. They are all different ways in which you see the Fourier series equation written in books.

The Fourier Analysis

The process of adding together a bunch of sinusoids to create useful waves is called the **Synthesis** process. Synthesis of waveforms, is of course, interesting but what is really useful is the reverse process, that is, to take an arbitrary periodic signal and figure out its components. It's like trying to figure out the ingredients of a particular dish. By ingredients, we mean frequencies in the signal that contain significant power or amplitudes. This is the main use of Fourier analysis. It is called, not surprisingly, the **Analysis** part. What this involves is to make a guess of the fundamental frequency, f_0, and then computing the amplitudes (coefficients) of a certain number of harmonics. There is no guarantee that the fundamental chosen will result in finding all the main signal components exactly. Nevertheless, in most cases, we have a pretty good idea of signal components a-priori. So the process works well enough.

The usefulness of the process can be seen in the equation of the sawtooth wave in Eq. (1.10). The Fourier series allows us to create an estimate of the wave using a few or many terms. Hence, Fourier series represents an *estimate* of the true representation, the accuracy of which depends on the number of terms used.

The Fourier analysis process consists of finding the **series coefficients**. When we talk about Fourier series coefficients (FSC), we are talking about the amplitudes of the sine and cosine harmonics, and nothing else. Once we decide on a fundamental frequency -a starting point for the analysis- we already know all the harmonic frequencies since they are integer multiples of the fundamental frequency. All we have to do now is to compute the coefficients. The following are the three types of coefficients we need to compute.

1. The DC offset or the coefficient of the 0th frequency, $k = 0$.
2. Coefficients of the cosine $a_k \cos(2\pi k f_0 t)$ with $k = 1, 2, 3, \ldots, \infty$.
3. Coefficients of the sine $b_k \sin(2\pi k f_0 t)$ with $k = 1, 2, 3, \ldots, \infty$.

We will discuss each of these three types of coefficients separately and see how to compute them.

Computing a_0, the DC coefficient

We are given an arbitrary periodic signal, $f(t)$, of period T. The Fourier series says that the signal $f(t)$ is equivalent to a summation of K sinusoidal harmonics. Our task is to find the coefficients of each of these harmonics starting with $k = 0$ to $k = K$. This task is called Fourier analysis.

$$f(t) = \boxed{a_0} + \sum_{k=1}^{\infty} a_k \cos(2\pi f_k t) + b_k \sin(2\pi f_k t) \tag{1.18}$$

The constant a_0 in the Fourier series equation represents the DC offset. If our target wave has a nonzero DC component (if its average amplitude value is not zero), then we know that $a_0 \neq 0$. But before we compute it, let's take a look at a useful property of sine and cosine waves. Both sine and cosine waves are symmetrical about the x-axis. When you integrate

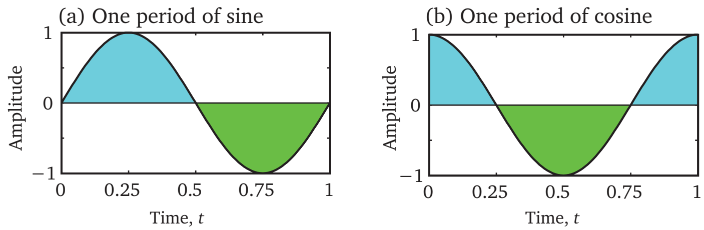

Figure 1.17: The area under both sine and cosine over one period is zero, no matter what their frequency.

a sine or a cosine wave over one period, you always get zero. The area above the x-axis cancels out the area below it. This is always true over one period as we can see in Fig. 1.17. The same is also true for the sum of sine and cosine waves. Any wave made by summing sine and cosine waves also has zero area over one period. If we were to integrate the given signal $f(t)$ over one period, as in Eq. (1.18), the area obtained will have to come from the coefficient a_0 only. None of the sinusoids makes any contribution to the integral and they will all fall out. Hence, the calculation of the DC term becomes easy simply because the integral of the harmonics is zero.

$$\int_0^{T_0} f(t)\,dt = \int_0^{T_0} a_0\,dt + \cancelto{0}{\int_0^{T_0} \left(\sum_{k=1}^{\infty} a_k \cos(2\pi f_k t) + b_k \sin(2\pi f_k t) \right) dt}$$

We compute a_0 by computing the integral of the wave over one period. The area under one period of this wave is equal to

$$\int_0^{T_0} f(t)\,dt = \int_0^{T_0} a_0\,dt$$

Integrating this simple equation, we get,

$$\int_0^{T_0} f(t)\,dt = a_0 T_0$$

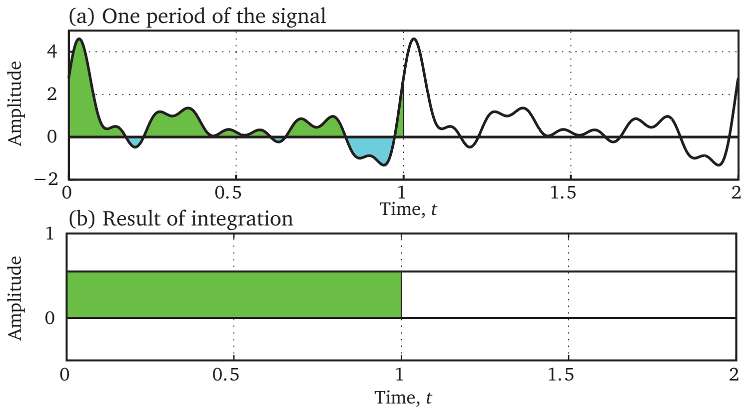

Figure 1.18: (a) The periodic signal before integration. (b) After integration of exactly one period, only the DC component is left.

We can now write a very easy equation for the first coefficient, a_0

$$\boxed{a_0 = \frac{1}{T_0}\int_0^{T_0} f(t)\, dt} \tag{1.19}$$

Summary: To compute the DC coefficient, integrate the target signal $f(t)$ over one period. The result of the integration, normalized by the period, is equal to the 0th coefficient. Hence the area under the given signal comes only from the 0th coefficient.

Computing b_k, the coefficients of sine harmonics

We are assuming that our target signal is composed of sines and cosine harmonics. Now we multiply the target signal by just one harmonic of kth frequency. We will get various different types of combinations. We will get sine harmonics multiplied by sine and cosine harmonics of both the same and different frequencies, or in other words a lot of terms to solve.

Here is a target signal we wish to represent with just two harmonics (1 and 3).

$$f(t) = \mathrm{a}_1\cos(\omega_0 t) + \mathrm{b}_1\sin(\omega_0 t) + \mathrm{a}_3\cos(3\omega_0 t) + \mathrm{b}_3\sin(3\omega_0)$$

We want to compute the coefficient of the first sine harmonic or the term b_1. To do that we multiply the signal $f(t)$ by this same harmonic. We get in the product, various combinations of sines and cosines:

$$\int_t^{t+T_0} \cos(\omega_0 t) \times \sin(\omega_0 t)\, dt$$
$$\int_t^{t+T_0} \sin(\omega_0 t) \times \sin(\omega_0 t)\, dt$$
$$\int_t^{t+T_0} \cos(3\omega_0 t) \times \sin(\omega_0 t)\, dt$$
$$\int_t^{t+T_0} \sin(3\omega_0 t) \times \sin(\omega_0 t)\, dt$$

Now compute the coefficient, a_1, we multiply the representation by $\cos(\omega_0 t)$. We get these various types of products:

$$\int_t^{t+T_0} \cos(\omega_0 t) \times \cos(\omega_0 t)\, dt$$
$$\int_t^{t+T_0} \sin(\omega_0 t) \times \cos(\omega_0 t)\, dt$$
$$\int_t^{t+T_0} \cos(3\omega_0 t) \times \cos(\omega_0 t)\, dt$$
$$\int_t^{t+T_0} \sin(3\omega_0 t) \times \cos(\omega_0 t)\, dt$$

From these two examples, we see that in total there exist, just six different types of products, no matter how many harmonics we want to use to represent the target signal. These types, in terms of sine and cosine products are:

1. Cosine times a sine of the same frequency, $\cos(k\omega_0 t) \times \sin(k\omega_0 t)$
2. Sine times a sine of same frequency, $\sin(k\omega_0 t) \times \sin(k\omega_0 t)$
3. Cosine times a sine of a different frequency, $\cos(k\omega_0 t) \times \sin(m\omega_0 t)$
4. Sine times a sine of a different frequency, $\sin(k\omega_0 t) \times \sin(m\omega_0 t)$
5. Cosine times a cosine of same frequency, $\cos(k\omega_0 t) \times \cos(k\omega_0 t)$
6. Cosine times a cosine of a different frequency, $\cos(k\omega_0 t) \times \cos(m\omega_0 t)$

We need to compute the integral of each of these types of products over one period for the Fourier equation. Conveniently, it turns out, that the integral of most of these sinusoidal products, over one period is zero, except for Type 2 and Type 5, when the frequencies of the sinusoids coincide. This makes the problem so much more tractable!

It is easy to show that these terms are zero. However, let us examine first what happens when we integrate a Type 2 product, a sine times a sine of the same frequency, over one period. We notice that the product of the two sine waves of the same frequency, as shown in Fig. 1.19, lies entirely above the x-axis and has a net positive area which is proportional to the coefficient b_k. From integral tables we can compute this area as equal to

$$\int_0^{T_0} b_k \sin(k\omega_0 t) \times \sin(k\omega_0 t)\, dt = b_k \frac{T_0}{2}$$

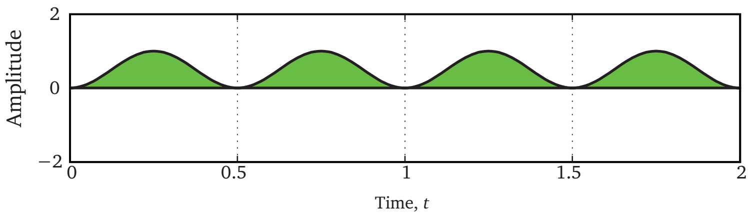

Figure 1.19: A sine wave multiplied by a sine wave of the same frequency has nonzero area under one period.

Now let us take a look at the integral of a Type 4 product, i.e. a sine wave multiplied by a sine wave of a different frequency.

$$\sin(k\omega_0 t) \times \sin(\underline{m}\omega_0 t)$$

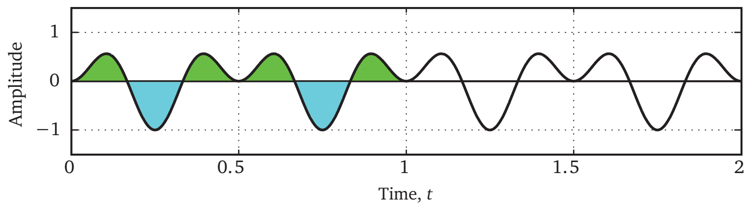

Figure 1.20: A sine wave multiplied by a sine wave of a different frequency has zero area under one period.

Figure 1.20 shows in the time domain the product of two sine waves of different frequencies. It may not be obvious from the figure, but the total area in one period of this product is zero. We make an important observation that the area in one period of a sine wave multiplied by *any* of its harmonics is zero. We conclude that when we multiply a signal by any of its harmonics, and integrate the product over one period, then the result can be used to find the contribution of just that harmonic and none other. The integral tells us how much of that frequency is present in the target signal. All other sinusoids in the signal contribute nothing. The individual contribution to the signal $f(t)$ by the kth harmonic can be written as:

$$\int_0^{T_0} b_k \sin(k\omega_0 t) \times \sin(m\omega_0 t)\, dt = 0 \quad \text{for } k \neq m.$$
$$\int_0^{T_0} b_k \sin(k\omega_0 t) \times \sin(m\omega_0 t)\, dt = \frac{T_0}{2} b_k \quad \text{for } k = m.$$

The same is true for cosine waves or Types 5 and 6.

$$\int_0^{T_0} a_k \cos(k\omega_0 t) \times \cos(m\omega_0 t)\, dt = 0 \quad \text{for } k \neq m.$$
$$\int_0^{T_0} a_k \cos(k\omega_0 t) \times \cos(m\omega_0 t)\, dt = \frac{T_0}{2} a_k \quad \text{for } k = m.$$

We see that the result of the integration of the product of two harmonics when their frequencies are unequal is **zero**. It is nonzero only when the waves have the *same* frequency. Hence, if we multiply our signal by its K harmonics, and integrate K times, the result for each case is the coefficient of the harmonic being used for multiplication.

Let us now look at the product of a sine wave and a cosine wave of the same frequency, or Type 1. For this case, as shown in Fig. 1.21, the net area under the product is also zero.

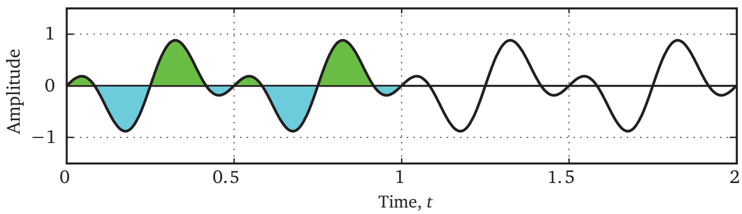

Figure 1.21: A sine wave multiplied by a cosine has total area of zero under one period.

The integral of Type 3 products is also zero. Hence, we note this important result; *the area under the product of a sine and a cosine over one period is zero whether the frequencies are the same or not.* Sines and cosines just don't agree. This is also the concept of orthogonality. We say that these waves are orthogonal to each other as they contribute nothing to the integral. Summarizing the results:

$$\int_0^{T} b_k \cos(k\omega_0 t) \times \sin(m\omega_0 t)\, dt = 0 \tag{1.20}$$

A practical interpretation of these properties is that sine and cosine waves act as filtering signals. When a signal, is multiplied by a sinusoid of a particular frequency, the integral is proportional to the content of that multiplying sinusoid, (by which we mean its amplitude). Hence, sinusoids behave as narrow-band filters. This is the fundamental concept of a *filter*.

Here are the key results we will be using in calculating the coefficients of the Fourier equation:

1. If you multiply a sine or a cosine wave by any of its harmonics, the area under the product is zero.
2. If you multiply a sine or a cosine of a particular frequency by itself, the area under the product is proportional to the Fourier coefficient of that frequency.
3. The area under a sine wave multiplied by a harmonic cosine is always zero. (Because sine and cosine are orthogonal!)

We use these observations to compute the b_k coefficients. We successively multiply the target signal, $f(t)$ by a sine wave of a specific harmonic frequency and then integrate over one period as in the equation below. All terms are zero except one. This term is given by:

$$\int_0^{T_0} b_k \sin(k\omega_0 t)\sin(k\omega_0 t)\, dt = \frac{b_k T_0}{2}$$

From this we obtain the coefficient of the sine, b_k as follows

$$\boxed{b_k = \frac{2}{T_0}\int_0^{T_0} f(t)\sin(k\omega_0 t)\, dt} \tag{1.21}$$

The coefficient b_k is computed by taking the target signal over one period, successively multiplying it with a sine wave of kth harmonic frequency and then integrating. The result of the integration is then multiplied (or normalized) by $2/T_0$ to obtain the coefficient for that particular harmonic. If we do this K times, we get K independent b_k coefficients.

Computing a_k, the coefficients of cosine harmonics

Now we need to do nearly the same thing for cosine coefficients. The process of computing the coefficients of the cosine harmonics is exactly the same as the one used for the sine waves. Instead of multiplying the target signal, $f(t)$ by a sine wave, we multiply the target signal sequentially by a cosine wave of frequency, $k\omega_0$. We get exactly the same result as when we compute the sine coefficients. Only one term will remain in the big long multiplication for each value of k. That term is

$$\int_0^{T_0} a_k \cos(k\omega_0 t)\cos(k\omega_0 t)\, dt = \frac{a_k T_0}{2}$$

The coefficient a_k hence can be calculated by multiplying $f(t)$ by the kth harmonic and integrating the expression. The result is proportional to the coefficient. This expression is nearly identical to Eq. (1.21) for the b_k coefficients.

$$\boxed{a_k = \frac{2}{T_0}\int_0^{T_0} f(t)\cos(k\omega_0 t)\, dt} \tag{1.22}$$

In all, we do this $2K$ times, with K computations each for the sine and, K for the cosine.

Conceptually, the process of computing the coefficients consists of filtering the target signal, one frequency at a time. Hence, a spectrum can be seen as tiny little filter fingers that pull out the quantity (amplitude) of that harmonic in the signal.

In some simple cases, when given a signal that is composed of sinusoids, we do not need to do any calculations, as the coefficients are there in plain sight for us to see as in the following example.

Example 1.1. What are the FSC of this signal:

$$f(t) = 0.8\sin(6\pi t) - 0.3\cos(6\pi t) + .75\cos(12\pi t)$$

The signal is already in the form of a Fourier series. Just by looking we can tell that the coefficient a_0 is 0. We set the fundamental frequency at 1 Hz, because then both frequencies in the signal, 3 Hz and 6 Hz fall on a harmonic. The coefficients by observation are: $a_0 = 0$, $a_3 = -0.3$, $b_3 = 0.8$ and $a_6 = 0.75$. Of course we could have set the fundamental as 3 Hz, in which case the coefficients would be given as: $a_0 = 0$, $a_1 = -0.3$, $b_1 = 0.8$ and $a_2 = 0.75$. Nothing has changed except the index.

We just did our first Fourier analysis without lifting a pencil! Now something a little more complicated.

Example 1.2. What are the FSC of this signal:

$$f(t) = \cos^3(2\pi t)$$

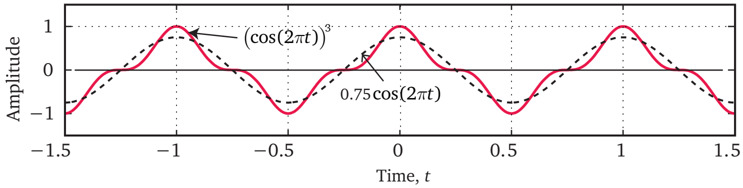

Figure 1.22: The signal $\cos^3(2\pi t)$*, in solid curve, and its first order representation in dashed line.*

From Fig. 1.22, the plot of this signal in the time-domain, we can spot two pieces of information. The first is that the period of the wave is 1 second (its frequency is 1 Hz.), and second, the symmetry of the signal is even. Because it is an even signal, it can be represented solely with cosine components. (But, of course, that makes sense as the function is cosine-cubed.) But because this a cubed function, at first we might not know what to do next. Fortunately, we can expand this function and put it in the Fourier series form, in which case the coefficients would become obvious. But instead, we will find the coefficients the hard way, and that means integration. First we calculate the a_0 coefficient using Eq. (1.19) to determine the DC offset.

$$a_0 = \int_0^1 \cos^3(2\pi t)\, dt = 0$$

This integration gives us a zero and the result agrees with the graph in Fig. 1.22. The function has a symmetry about the x-axis meaning there is no DC offset. Now using Eq. (1.22) we can determine the a_k coefficients (of cosine) starting with the fundamental harmonic. We set fundamental frequency f_0 equal to 1 Hz.

$$a_1 = 2\int_0^1 \cos^3(2\pi t)\cos(1 \times 2\pi t)\, dt = 0.75$$

The Fourier representation of this signal is now given by $0.75\cos(2\pi t)$. Figure 1.22 plots in the dashed line the representation of the signal using only the a_1 coefficient. It is clear that one coefficient is not enough to properly represent the signal. We should calculate a few more coefficients. However, it is amazing, how close we are with just one term!

$$a_2 = 2\int_0^1 \cos^3(2\pi t)\cos(2 \times 2\pi t)\, dt = 0$$

$$a_3 = 2\int_0^1 \cos^3(2\pi t)\cos(3 \times 2\pi t)\, dt = 0.25$$

With $a_0 = 0$, $a_1 = 0.75$, $a_2 = 0$, and $a_3 = 0.25$, the representation is given as:

$$f(t) = 0.75\cos(2\pi t) + 0.25\cos(6\pi t)$$

This is the Fourier series representation of the signal given to us as $\cos^3(2\pi t)$. If we continue to calculate more coefficients (for $k > 3$), we will see that they are all zero after a_3. From this, we can say that function cosine cubed is, in fact, made up of only two cosines of frequency 1 and 3 Hz. In this case, Fourier series representation is an exact representation of the signal, but this is not always the case.

Example 1.3. What are the FSC of this signal?

$$f(t) = \sin^2(2\pi t) + \sin(2\pi t)$$

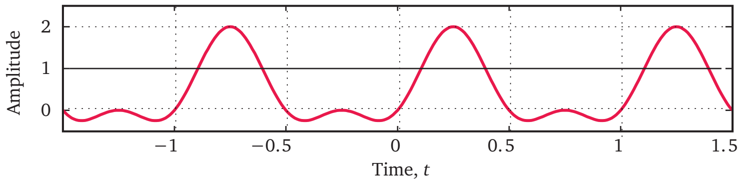

Figure 1.23: The signal, $f(t) = \sin^2(2\pi t) + \sin(2\pi t)$.

The analysis of this signal uses the ideas from the first two examples. The signal has a linear part, $\sin(2\pi t)$, and a non-linear part, $\sin^2(2\pi t)$. The best way to compute its Fourier series representation is to convert the non-linear term to a linear term. Using trigonometric identities, we do this conversion and rewrite the signal with all linear terms as:

$$f(t) = \frac{1}{2} + \sin(2\pi t) - \frac{1}{2}\cos(4\pi t)$$

From here, now that the signal is very nearly in Fourier form, we can expect to have nonzero a_0, b_1 and a_2 coefficients. The values of these will match the coefficients of the above equation. Solving for these the hard way, using integration, we show that this is indeed the case. The Fourier representation of this signal is exact. This only happens if the original signal is composed only of harmonic sinusoids. In real life, this situation is unlikely. Practical signals are almost never composed solely of harmonic sinusoids. That is our challenge. When we compute the Fourier representation, we are indeed *estimating* the form of the signal. And for this estimation, we use Fourier analysis as our tool.

$$a_0 = \int_0^1 \left(\sin^2(2\pi t) + \sin(2\pi t)\right) dt = \frac{1}{2}$$
$$a_1 = 2\int_0^1 \left(\sin^2(2\pi t) + \sin(2\pi t))\cos(1\cdot 2\pi t)\right) dt = 0$$
$$a_2 = 2\int_0^1 \left(\sin^2(2\pi t) + \sin(2\pi t))\cos(2\cdot 2\pi t)\right) dt = -\frac{1}{2}$$
$$b_1 = 2\int_0^1 \left(\sin^2(2\pi t) + \sin(2\pi t))\sin(1\cdot 2\pi t)\right) dt = 1$$
$$b_2 = 2\int_0^1 \left(\sin^2(2\pi t) + \sin(2\pi t))\sin(2\cdot 2\pi t)\right) dt = 0$$

Note that there is a $\frac{1}{2}$ in the equation of the signal, and this is the same thing as a DC offset, hence this is the a_0 coefficient. There is one cosine at a frequency of 2 Hz with amplitude of $-\frac{1}{2}$, which means that the coefficient a_2 would be $-\frac{1}{2}$. There is a sine at frequency of 1 Hz of amplitude 1, hence b_1 should be 1 and, it is. We could have solved this problem by observation only, without doing any calculations.

Coefficients become the spectrum

How do we go from coefficients to a spectrum? Assume that we have a signal which we have analyzed and have found that it has nine harmonics with $f_0 = 2.5$ Hz. The coefficients of the nine harmonics starting with $k = 1$ are given by:

$$a_k = [0.25, 0.2, 0.7, 0.5, -0.2, 0.2, 0.1, -0.05, 0.02]$$
$$b_k = [0.4, -0.3, 0.7, 0.7, 0.3, 0.275, 0.25, 0.2, 0.1]$$

Here, a_k are the coefficients of the cosine harmonics and b_k are the coefficients of the sine harmonics. We plot these in a bar graph in Fig. 1.24 as a function of the index, k. Both sine and cosine of the same frequency are plotted next to each other. This is a spectrum that in essence displays the recipe of the signal. It tells us how much of each harmonic, i.e., its amplitude, we need to recreate the signal. Given this spectrum, you should be able to write the Fourier series equation for this signal.

Commonly a spectrum is given in terms of "power" but the coefficients we compute via Fourier analysis are not power. They are *amplitudes*. The term spectrum is often used synonymously with power spectrum but is not the same as the power spectrum.

In signal processing, the coefficients computed for cosine are called *Real* and those for sine, are called *Imaginary*. Of course, there is nothing imaginary about the coefficients of sine, they are just as real as the cosine coefficients. It is just one of the many confusing terms we come across in signal processing.

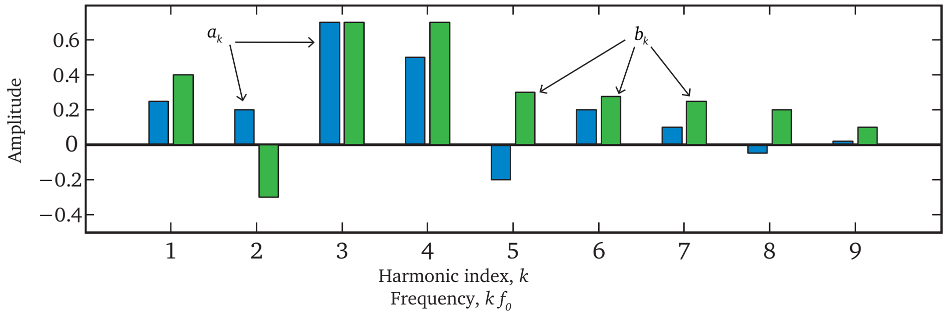

Figure 1.24: The coefficients of the harmonics.
For each harmonic index k representing a frequency of kf_0, there are two coefficients, one for sine and the other for cosine. These coefficients can be positive or negative and represent a formula for creating a Fourier representation of the test wave.

While the spectrum in Fig. 1.24 is in terms of sine and cosine coefficients, you may say that this is not the way we usually see a spectrum in books. A spectrum ought to have just one number for each frequency. So, we *combine* the sine and cosine coefficients into two real-world terms, called **magnitude and phase**.

As harmonic sine and cosine are orthogonal to each other, and have the same frequency, we can compute their *sum* by thinking of them as vectors. We can compute this by doing the root-sum-square (RSS) of the two coefficients. This RSS term is called the ***magnitude***. Now we plot the modified spectrum using magnitude on the y-axis instead of amplitudes of the sine and cosine. The magnitude is *one* number for each frequency and is computed by this expression.

$$\text{Magnitude: } c_k = \sqrt{a_k^2 + b_k^2} \tag{1.23}$$

While the amplitude can indeed be negative, the magnitude is always positive. The effect of the sign of the amplitude is now seen in a term called phase, which we calculate by

$$\text{Phase: } \phi_k = \tan^{-1} \frac{b_k}{a_k} \tag{1.24}$$

This phase term is different from the previous two types of phases we mentioned, as this is a combination of the phase of the sine and the cosine harmonic of a particular frequency. The range of arctan is from $-\pi/2$ to $+\pi/2$, so it must be unwrapped to compute the full phase. What it represents is the relative phase shift between the cosine and the sine harmonic. The phase spectrum is usually quite noisy due to computational effects.

We show these two quantities on a pair of plots, one for the magnitude and the other for this new phase term, both plotted as a function of the harmonic frequencies or the index, as shown in Fig. 1.25. The first one is called the **magnitude spectrum** and the second is called the **phase spectrum**.

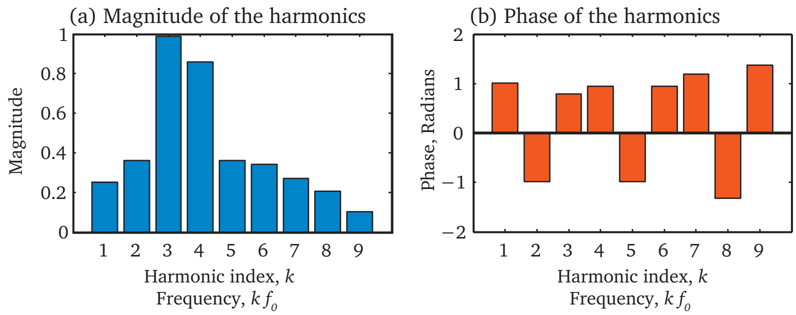

Figure 1.25: The magnitude and phase spectrum.
In (a), we see the RSS of both the sine and cosine coefficients, representing the magnitude of that harmonic frequency. In (b), we plot the combined phase for each frequency. The magnitude on the right is usually quite instructive but phase is hard to comprehend as it changes quickly from one frequency to the next.

Hence, a spectrum can be plotted in two ways: as a pair of plots of the coefficients of the sine and the cosine harmonics, or, as the magnitude and phase. (In Fig. 1.24, we have plotted both the real and imaginary coefficients on just on one figure.) The spectrum based on the real(cosine) and imaginary(sine) coefficients, changes depending on the starting point of the analysis. Whereas in the second form, the plot of magnitude and the phase are not a function of the starting point. Hence practicing engineers prefer the second form.

The magnitude spectrum, computed from Eq. (1.23) is the preferred form in industry. Its companion, the phase spectrum usually does not offer information we can use quickly, and often goes ignored. Confusingly, most people make no distinction when talking about amplitude or magnitude, and often these two terms are used interchangeably.

The power spectrum

The process of doing Fourier analysis consists of computing the amplitude of each harmonic and then converting it to the magnitude and phase spectrum. Fig. 1.26 shows a power spectrum with the y-axis given in dBs. To convert a magnitude spectrum to a power spectrum, we use Parseval's theorem. This tells us that the total power in a signal is the sum of the powers in each harmonic. The power of each harmonic is defined as the square of its magnitude. Hence to represent power, you square each magnitude value and then compute its dB value by $10 \cdot \log 10(c_k)^2$. Alternatively, you can compute the base 10 log of the coefficient and then multiply it by 20.

$$\text{Power in the } k\text{th harmonic} = 20 \log(c_k) \tag{1.25}$$

The power spectrum is often normalized to maximum power, such as **bin 3** in Fig. 1.26. (The bin is a form of identifier of the harmonic and is equivalent to the harmonic index k.) The level of each component is the dB equivalent of its ratio to the maximum power. All component levels relative to the maximum power, are said to be a certain number of dBs below the maximum. The maximum level is zero dB, with all other values shown as negative. These values are also called **power spectral density** (PSD) because they are a form of density of the power across a bandwidth.

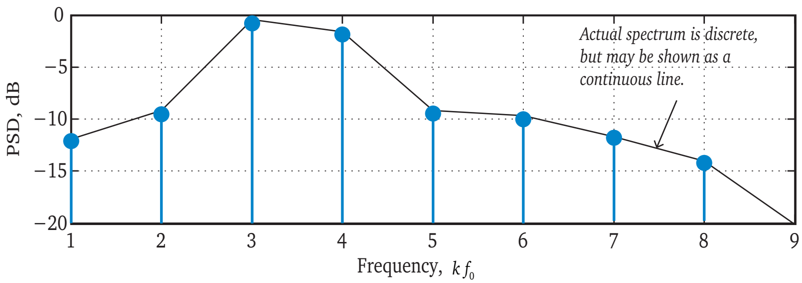

Figure 1.26: A traditional Power spectrum created from the Fourier coefficients.

Looking at the FSC, note that there are just K (the largest k) number of such coefficients, one for each of the harmonics. The spectrum from the FSC is thus considered *discrete*. In the time domain the signal is continuous, however, in the frequency domain, the one-sided spectrum developed from the FSC is *discrete*. It has only K terms. Since the index k is positive, the plot starts at 0 and hence, is called, a **one-sided spectrum**.

This is an important property of the Fourier series. A one-sided spectrum consists only of a finite number of components, each an independent value. The x-axis is given either by the index k, but more commonly, by the actual frequency ranging from 0, f_0, $2f_0$, $3f_0$,..., etc.

Fourier analysis applies only to periodic waves. And if that is all we could do with Fourier analysis, then it would not have a lot of use. For real signals we can never tell what the period is nor where it starts. Usually no periodicity can be seen. In fact, a real signal may not be periodic at all. In this case, further developments of the theory allow us to extend the "period" to infinity so we just pick any section of a signal or even the whole signal and call it "The Period" representing the whole signal. This idea is the basis of the Fourier transform that we will discuss in Chapter 4.

In this chapter we discussed the Fourier series representation of an arbitrary periodic wave in terms of weighted harmonic sine and cosine waves. In the next chapter, we will look at how we can do the same using the confusing and scary-looking complex exponentials.

Summary of Chapter 1

In this chapter, we introduced the concept of using sinusoids to represent an arbitrary periodic wave. We also introduced the concept of the fundamental frequency and its harmonics. The Fourier series representation consists of finding unique weightings of these harmonics to represent a particular periodic wave. These weights are called the Fourier series coefficients (FSC). When plotted as a function of frequency, these coefficients represent the spectrum of the signal. The spectrum calculated using FSC is discrete although the signal is continuous in time.

Terms used in this chapter:

- **Fundamental frequency** - The smallest frequency of the signal to be represented by Fourier series.
- **Harmonic frequencies** - All integer multiples of the fundamental frequency.
- **Sinusoids** - Sine or cosine wave.
- **Harmonic coefficients** - The amplitude of a harmonic.
- **Real and imaginary** - Cosine is said to exist in the real plane and sine in the imaginary plane.
- **Magnitude** - The RSS value of the amplitudes of the sine and cosine of a particular harmonic. It is always positive.
- **Phase** - The fixed phase shift value of a wave at $t = 0$, often specified in radians.
- **One-sided spectrum** - A spectrum for positive harmonic index, $k = 0, 1, 2$, etc.
- **Power spectrum** - Showing distribution of power rather than amplitude.

1. The most common trigonometric form of the Fourier series is given by

$$f(t) = a_0 + \sum_{k=1}^{\infty} a_k \cos(2\pi f_k t) + \sum_{k=1}^{\infty} b_k \sin(2\pi f_k t)$$

2. The coefficients of the Fourier series are easily computed by

$$a_0 = \frac{1}{T_0} \int_0^{T_0} f(t)\, dt$$
$$b_k = \frac{2}{T_0} \int_0^{T_0} f(t) \sin(k\omega_0 t)\, dt$$
$$a_k = \frac{2}{T_0} \int_0^{T_0} f(t) \cos(k\omega_0 t)\, dt$$

3. Many periodic signals can be represented by the weighted sum of harmonics sinusoids. The representation is an estimation and may not be an exact replication.
4. Harmonic sinusoids are orthogonal to each other; hence, the integral of their products (or cross product) is zero.
5. The linearity property of the Fourier series implies that a change of the coefficients of one harmonic does not affect the coefficients of the other harmonics.
6. A time or phase shift of the signal does not affect the magnitude of the coefficients.
7. To synthesize a signal based on the Fourier series, we pick a fundamental frequency first. All harmonics are integer multiples of the fundamental.
8. We designate harmonics by the letter k. Hence, for all integer k, all kf_0 frequencies are harmonics of the fundamental frequency f_0.
9. Fourier analysis means to find the Fourier coefficients of the Fourier series.
10. The Fourier coefficients are discrete because the harmonics are discrete. Hence the spectrum of a periodic continuous-time signal is discrete.
11. A spectrum can be a spectrum of amplitude, magnitude or power. These are all different and used for different purposes.

Questions

1. Can you state in words the principle behind the Fourier series.
2. What is the first and the third harmonic of a sinusoid of frequency 3 Hz?
3. Can any signal have harmonics?
4. When is one harmonic orthogonal to another? What trigonometric property tells us that this is true?
5. Three non-harmonic sinusoids are added together. What is the period of the summed wave?
6. If we add three non-harmonic sinusoids together, is the resulting signal periodic?
7. Is the sum of N harmonics also harmonic to the fundamental?
8. Are harmonics within a harmonic set *also* harmonic to each other?
9. If we make the amplitude of a signal, a function of time, what effect will that have on its frequency?
10. A change in phase of a cosine wave means it is still orthogonal to a sine wave of the same frequency; true or false?
11. If the phase of a harmonic is changed, does it remain a harmonic of an another wave? item If the amplitude of a harmonic is changed, does it remain a harmonic of an another wave?
12. What is the maximum amplitude of N harmonic cosine waves added together. What is it for sine waves?
13. We need to represent a wave that starts at time $t = 0$. What type of harmonics will be in its representation?
14. The summation of odd and even waves can be used to create any waveform we want, true or not?
15. Is Fourier series representation an accurate representation of a wave? Why not?
16. Can we create a Fourier series representation of any wave?
17. Why do we consider the set of harmonics a *basis set*? What constitutes a basis set?
18. Sine and cosine waves are a basis set for Fourier analysis. Can you give an example of another set of basis functions.
19. What quality of sine and cosine makes them suitable as a basis set?
20. Fourier series analysis is considered a linear process. Why?
21. What do the coefficients of a Fourier series represent? What does the a_0 coefficient represent?

22. What are the Fourier series coefficients of this signal? $s = A + B\sin(2\pi f t)$.
23. We want to compute the FSC of this signal. $s = \sin(6.5t) - \cos(4.75t)$. What should we pick as the fundamental frequency so as to accurately represent this signal.
24. If the target wave is shifted by a certain phase, what happens to its coefficients?
25. How many coefficients would you need to describe this wave? $x(t) = 2 + B\sin(4\pi t + \pi/2) - \cos(12\pi t)$. Find the coefficients of this signal.
26. What is the fundamental period of this Fourier series representation? $x(t) = (1/2)(\sin(4\pi t) + (1/3)\sin(12\pi t) + (1/5)\sin(20\pi t) + (1/7)\sin(28\pi t)$. What are the coefficients of the first four cosine and sine harmonics?
27. Given these equations, what are the Fourier series coefficients, a_0, a_1, a_2 for each case.
    (a) $y = \frac{1}{2} + \frac{3}{4}\sin(\pi x) - \frac{3}{5}\cos(2\pi x)$.
    (b) $y = \frac{3}{4}\cos(2\pi x) - \frac{3}{5}\cos(3\pi x)$.
28. What is the difference between amplitude and magnitude?
29. The amplitude of a harmonic varies from -1 v to +1 v. What is its peak amplitude and its peak-to-peak amplitude? What is its magnitude? What is the power of the harmonic? What is the value of the power in dB?
30. Examine Fig. 1.13 and give the first four coefficients of the even square wave.

Chapter 2

Complex Representation of Continuous-Time Periodic Signals

Leonhard Euler
1707 - 1783

Leonhard Euler was a pioneering Swiss mathematician and physicist. He made important discoveries in fields as diverse as infinitesimal calculus and graph theory. He also introduced much of the modern mathematical terminology and notation, particularly for mathematical analysis, such as the notion of a mathematical function. Euler is considered to be one of the greatest mathematicians to have ever lived. A student of Johann Bernoulli, Euler was the foremost scientist of his day. Born in Switzerland, he spent his later years at the University of St. Petersburg in Russia. He perfected plane and solid geometry, created the first comprehensive approach to complex numbers. Euler was the first to introduce the concept of $\log x$ *and* e^x *as functions and it was his efforts that made the use of* $e, \imath$ *and* π *the common language of mathematics. Among his other contributions were the consistent use of the trigonometric sine, and cosine functions and the use of a symbol for summation. A father of 13 children, he was a prolific man in all* *aspects, languages, medicine, botany, geography and physical sciences and has left his mark on our scientific thinking.– From Wikipedia*

Euler's Equation

$$\boxed{e^{j\omega t} = \cos\omega t + j\sin\omega t}$$

This equation is called the Euler's equation. Bertrand Russell and Richard Feynman both gave this equation plentiful praise with words such as "the most beautiful, profound and subtle expression in mathematics" and "the most amazing equation in all of mathematics." This perplexing equation was developed by Euler (pronounced Oiler) in the early 1800s.

The $e^{j\omega t}$ in Euler's equation is a decidedly confusing concept. What exactly is the role of j in $e^{j\omega t}$? We know from algebra that it stands for $\sqrt{-1}$ but what is it doing here with the sine and cosine? Can we even visualize this function?

The complex exponential

The function $e^{j\omega t}$ goes by the name of **complex exponential** (CE). This function is of the greatest importance in signal processing and Fourier analysis. We are going to discuss its conceptual nature and its application to Fourier analysis.

$$e^{j\omega t} = \underline{\cos\omega t + j\sin\omega t} \tag{2.1}$$

Looking at Eq. (2.1), we see the complex exponential on the left side. It is called the **positive complex exponential**, for the simple reason that the exponent of e is positive. On the right of the equal sign, underlined, is the expression containing a sine and a cosine. For now, ignore the complex exponential $e^{j\omega t}$ on the left-hand side and examine only the right-hand side of this equation, containing the sine and cosine waves, with the complex operator j thrown in.

This is a complex function which often means that it is a 3D function. It can be plotted by assigning a particular value to the radial frequency ω, and then for a range of time t, calculating both $\cos\omega t$ and $\sin\omega t$ values. Because the value of ω is held constant, we have three values, the independent time variable, t, and associated sin and cos values, $\cos(\omega t)$ and $\sin(\omega t)$. With these three values, t, $\cos(\omega t)$ and $\sin(\omega t)$, we can create the 3D plot shown in Fig. 2.1. Time is plotted on the x-axis, and y and z axis contain the values of the function $\cos(\omega t)$ and $\sin(\omega t)$. Each plotted helix is for a fixed value of frequency, ω. Changing the frequency will this figure.

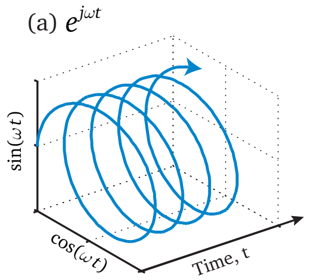

Figure 2.1: $e^{j\omega t}$ when plotted looks like is a helix. It is a 3-D function of three values, time t, $sin(\omega t)$ and $cos(\omega t)$ for a fixed frequency ω. The exponent of the CE indicates direction of advance or movement, (a) positive exponent and (b) negative exponent.

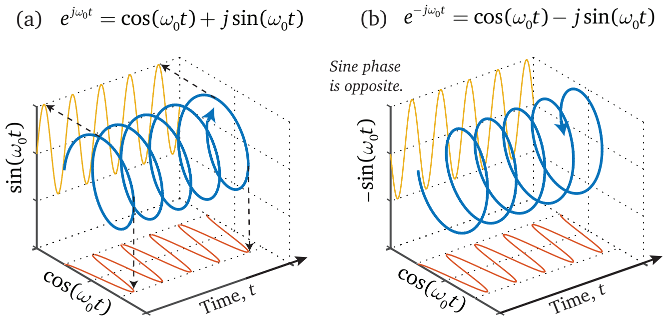

Figure 2.2: The projections of the complex exponential are sinusoids. (a) $e^{j\omega t}$ and its two projections, (b) $e^{-j\omega t}$ and its two projections. Note that the sine wave in (b) has different phase than one for the positive CE in (a). That is the only difference between the two CEs.

The expression for the **negative complex exponential** (negative exponent) is written similar to Eq. (2.1) as:

$$e^{-j\omega t} = \cos \omega t - j \sin \omega t \tag{2.2}$$

We can plot these figures in Matlab using this code.

```
t = 0:0.01:5;
y=exp((j*(2*pi))*t); % positive exponent
subplot(1,2,1);
plot3(t,real(y),imag(y), 'linewidth', 1);
subplot(1,2,2);
y2=exp((-j*(2*pi))*t); % negative exponent
plot3(t,real(y2),imag(y2), 'linewidth', 1);
```

Projections of a complex exponential

As the CE is a complex function, we examine its projections on the real and the imaginary axes. In Fig. 2.2(a) we plot the projections of the helix on the *Real* and the *Imaginary* planes. In Cartesian terms, these would be called the (X, Y) and (Z, Y) planes. The projections of the complex exponential on these two planes are sine and cosine waves. The Real projection of the complex exponential is a cosine wave and Imaginary projection is a sine wave.

These projections can be plotted using the following Matlab code.

```
t = 0:0.01:5;
y=exp((j*(2*pi))*t); % Positive exponential
subplot(1,2,1);
plot3(t,real(y),imag(y), 'linewidth', 1);
th2 = linspace(-2,-2, length(real(y))); hold on;
plot3(t, real(y), th2,'linewidth', .5); hold on;
th = linspace(2,2, length(real(y)))
plot3(t, th, imag(y), 'linewidth', .5)
subplot(1,2,2);
y2=exp((-j*(2*pi))*t); % Negative exponential
plot3(t,real(y2),imag(y2), 'linewidth', 1);
th2 = linspace(-2,-2, length(real(y2))); hold on;
plot3(t, real(y2), th2,'linewidth', .5); hold on;
th = linspace(2,2, length(real(y2)))
plot3(t,th, imag(y2), 'linewidth', .5)
```

For the negative exponent, or the so-called negative complex exponential, the projection, sine wave, is flipped 180° degrees from the positive exponential, as shown in Fig. 2.2(a). Often this exponential is referred to as having a negative frequency; however, it is not really the frequency that is negative. From the definition of the negative exponent exponential in Eq. (2.2), we see that negative sign of the exponential results in the imaginary projection, the sine wave doing a 180° phase change, or equivalently being multiplied by -1.

The real part of the positive CE as well as the negative CE is a positive cosine wave.

$$\mathrm{Re}\left(e^{j\omega t}\right) = \cos\omega t \qquad \mathrm{Re}\left(e^{-j\omega t}\right) = \cos\omega t$$

The imaginary part of the positive CE is a positive sine but is negative for the negative CE.

$$\mathrm{Im}\left(e^{j\omega t}\right) = \sin\omega t \qquad \mathrm{Im}\left(e^{-j\omega t}\right) = -\sin\omega t$$

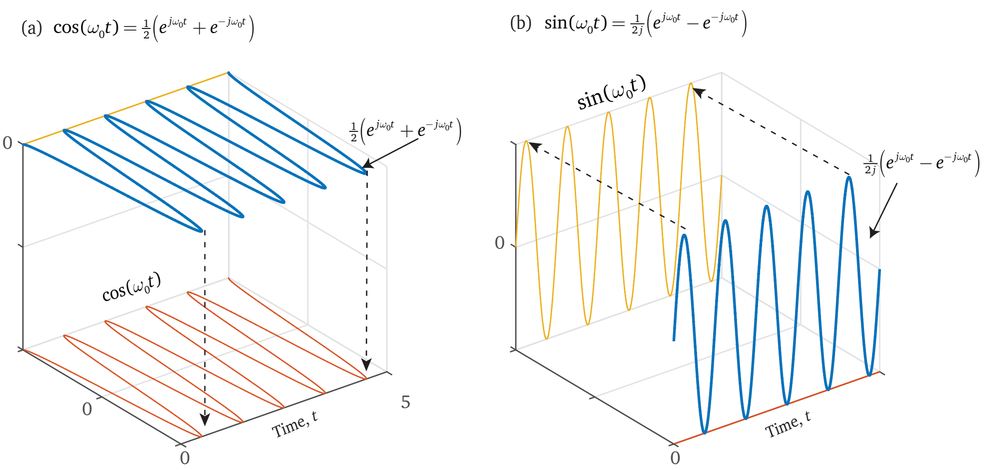

Figure 2.3: (a) Plotting $(e^{+j\omega t} + e^{-j\omega t})/2$ gives a cosine wave with zero projection on the imaginary plane (b) Plotting $(e^{+j\omega t} - e^{-j\omega t})/2j$ gives us a sine wave with zero projection on the real axis.

While the negative exponential has as its imaginary part a negative sine wave, the positive exponential has a positive sine as its imaginary part. The real part, which is a cosine, is same for both. We do not see any negative frequencies here, an idea generally associated with the negative complex exponential.

Adding and subtracting the complex exponentials, $e^{j\omega t}$ and $e^{-j\omega t}$, and then after a little rearrangement, we get these new ways of expressing a sine and a cosine.

$$\begin{aligned}\frac{1}{2}\left(e^{+j\omega t} + e^{-j\omega t}\right) &= \frac{1}{2j}\left(\cos\omega t + \cancel{j\sin\omega t} + \cos\omega t - \cancel{j\sin\omega t}\right)\\ &= \cos(\omega t)\end{aligned} \tag{2.3}$$

$$\begin{aligned}\frac{1}{2j}\left(e^{+j\omega t} - e^{-j\omega t}\right) &= \frac{1}{2j}\left(\cancel{\cos\omega t} + j\sin\omega t - \cancel{\cos\omega t} + j\sin\omega t\right)\\ &= \sin(\omega t)\end{aligned} \tag{2.4}$$

Let's see graphically what Eq. (2.3) and Eq. (2.4) look like. When we plot the two composite exponentials, we get the two plots as shown in Fig. 2.3. Fig. 2.3(a) shows that this composite exponential has a real projection of a cosine and Fig. 2.3(b), only the sine. The helix is gone, it has collapsed into a cosine or a sine. Hence, the sine and cosine can be said to be composed of complex exponentials. So a CE is composed of a cosine and a sine and conversely a sinusoid is also composed of two CE.

We are so used to thinking of sine and cosine as sort of atomic functions, divinely given that it seems hard to believe that they can be created from other functions. But Eq. (2.3) and

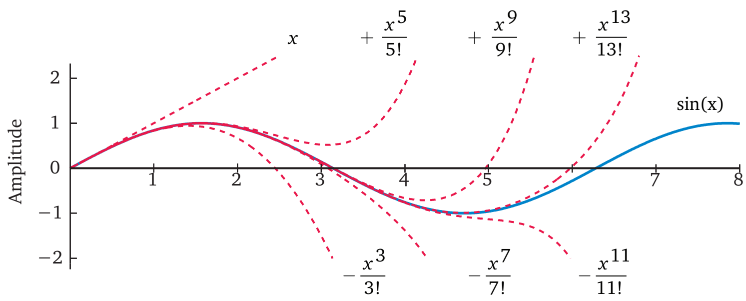

Figure 2.4: Sine wave as a sum of many exponentials of different weights.

Eq. (2.4) tell us that both sine and cosine can be created by adding complex exponentials. This is a case of two 3D functions coming together to create a 2D sinusoid.This sounds strange but it's actually not an unfamiliar concept. We can add two 2D functions and get a 1D function. An example is when we add a sine and a 180° shifted sine, we get a straight line, a 1D function. So a 2D function created by two 3D functions should not be a big stumbling block.

The Sinusoids

So how did Euler's equation come about and why is it so important to signal processing? We will try to answer that by first looking at Taylor series representations of the exponential e^x, sines, and the cosines. The Taylor series expansion for the two sinusoids is given in 2D by the infinite series as:

$$\begin{aligned} \cos x &= 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \dots \\ \sin x &= x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \dots \end{aligned} \tag{2.5}$$

Note that each one of these series is composed of many individual exponential functions. Thus, sine waves really are composed of exponentials! However, these are ***real*** exponentials that are nonperiodic and not the same thing as the complex exponentials. Complex exponentials are a special class of real exponentials and are used as an alternate to the sinusoids in Fourier analysis, as they are periodic and offer ease of expression and calculation, which is not obvious when we encounter them at first.

Real exponentials are used in Laplace analysis as the basis set, instead of the complex exponentials we use in Fourier analysis. Real exponentials are more general than complex exponentials and allow analysis of nonperiodic and transient signals. Laplace analysis is a general case of which Fourier is a special case applicable only to periodic or *mostly* periodic signals.

The Taylor series expansion for a *real* exponential e^x gives this expression

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \ldots \tag{2.6}$$

Both of these equations, Eq. (2.5) and Eq. (2.6) are straightforward concepts after you have digested them a bit. And indeed, if these functions were plotted, we would get just what we are expecting, the exponential of e and the sinusoids. How close our plots come to the continuous function depends on the number of terms that are included in the summation.

The similarity between the exponential and the sinusoids series in Eq. (2.5) and Eq. (2.6) shows clearly that there is a relationship here. Now let us change the exponent in Eq. (2.6) from x to $j\theta$. Note that the term θ is used here instead of ωt simply to keep the equation concise. Now we have by simple substitution, the $e^{j\theta}$ series:

$$e^{j\theta} = 1 + j\theta + \frac{(j\theta)^2}{2!} + \frac{(j\theta)^3}{3!} + \frac{(j\theta)^4}{4!} + \frac{(j\theta)^5}{5!} + \ldots \tag{2.7}$$

We know that $j^2 = -1$ and $j^4 = 1$, $j^6 = -1$, etc., substituting these values, we rewrite this series as

$$e^{j\theta} = 1 + j\theta - \frac{\theta^2}{2!} - \frac{j\theta^3}{3!} + \frac{\theta^4}{4!} + \frac{j\theta^5}{5!} - \frac{\theta^6}{6!} - \frac{j\theta^7}{7!} + \ldots \tag{2.8}$$

We can separate out every other term with j as a coefficient to create a two-part series, one without the j and the other with

$$\begin{aligned} e^{j\theta} &= 1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \frac{\theta^6}{6!} + \ldots \qquad \text{This is cosine} \\ &+ j\theta - \frac{j\theta^3}{3!} + \frac{j\theta^5}{5!} - \frac{j\theta^7}{7!} + \ldots \qquad \text{This is } j \text{ times sine} \end{aligned} \tag{2.9}$$

It can be seen that the first part of the series is a cosine per Eq. (2.5) and the second part with j as its coefficient is the series for a sine wave. Hence, we see that Euler's equation is really quite understandable! It came from the Taylor series. The concept of a Taylor series was formulated by the Scottish mathematician James Gregory but introduced into the world by the English mathematician Brook Taylor in 1715, so he gets his name on it instead.

$$e^{j\theta} = \cos(\theta) + j\sin(\theta)$$

We can now derive some interesting results from the Euler's equation, such as the following. By setting $\theta = \pi/2$, we can show that

$$\begin{aligned} e^{j\pi/2} &= \cancel{\cos(\pi/2)} + j\sin(\pi/2) \\ &= 0 + j \cdot 1 \\ &= j \end{aligned}$$

By setting $\theta = 3\pi/2$, we can show that

$$\begin{aligned} e^{j3\pi/2} &= \cancel{\cos(3\pi/2)} + j\sin(3\pi/2) \\ &= 0 + j \cdot -1 \\ &= -j \end{aligned}$$

And an another interesting result:

$$e^{j\pi} = \cos(\pi) + j\sin(\pi) = -1$$

From this, we can write this amazing looking equation!

$$e^{j\pi} + 1 = 0$$

So this complex looking function is not so complicated after all. These are some of the useful properties of a CE that we will be using. The function maintains a wondrous and mysterious quality, with added tinge of fear for some. However, we need to get over our fear of this equation and learn to love it. But, the question now is why bring up the Euler's equation in context of Fourier analysis at all? Why all this rigmarole about the complex exponential, why are sines and cosines not good enough? They are certainly easier to visualize.

In Fourier analysis, we computed the coefficients of sines and cosines (the harmonics) separately. In Chapter 1, we discussed the three different formulations of the Fourier series using sines and cosines, only with cosines, and with complex exponentials. Fourier analysis using the trigonometric form is not easy in practice. Trig functions are easy to understand but hard to manipulate. In fact, adding and multiplying them is a pain. On the other hand, doing math with exponentials is considerably easier. (See examples in Appendix A.)

Using a single exponential representing both sinusoids can simplify calculations in Fourier series. This is the main advantage of switching to complex exponentials in using the complex form of the Fourier series. The math looks hard but is actually easier. However, complex exponentials bring with them some conceptual difficulties, which is that they are hard to visualize and are confusing at first.

Typically, when we decompose something, we do it into a simpler form but here seemingly a more complex form is being employed. A simpler quantity, a cosine wave is now decomposed into two complex functions. However, the net result is that it will make Fourier analysis simpler. We will go from simplicity to complexity and then to simplicity again.

Let us take this sinusoid that has a phase term to complicate things and present it in complex form.

$$
\begin{aligned}
x(t) &= A\cos(\omega t + \theta) \\
&= \frac{A}{2}e^{j(\omega t+\theta)} + \frac{A}{2}e^{-j(\omega t+\theta)} \\
&= \frac{A}{2}e^{j\omega t}e^{j\theta} + \frac{A}{2}e^{-j\omega t}e^{-j\theta}
\end{aligned}
$$

In the last row, we separate the exponential into its powers. If we expand this expression into trigonometric domain using Euler's equation, we see that indeed we do get back the trigonometric cosine wave we started with.

$$
\begin{aligned}
&= \frac{A}{2}\Big(\cos(\omega t + \theta) + \cancel{j\sin(\omega t + \theta)} + \cos(\omega t + \theta) - \cancel{j\sin(\omega t + \theta)}\Big) \\
&= A\cos(\omega t + \theta)
\end{aligned}
$$

Fourier Series Representation using Complex Exponentials

In Chapter 1, we used trigonometric harmonics (the sine and cosine) as a basis set to develop the Fourier series representation. The target signal was "mapped" onto a k set of sinusoidal harmonics, such as these based on fundamental frequency of ω_0.

$$S = \left[\,\sin\omega_0 t,\ \cos\omega_0 t,\ \sin 2\omega_0 t,\ \cos 2\omega_0 t, \ldots, \cos k\omega_0 t, \ldots,\right]$$

A set of complex exponentials given by set S, can then be used alternately as a basis set for creating a complex (but preferred) form of the Fourier series.

$$S = \left[\,\ldots, e^{-3j\omega_0 t},\ e^{-2j\omega_0 t},\ e^{-1j\omega_0 t},\ 1\,, e^{1j\omega_0 t}, \ldots, e^{kj\omega_0 t}, \ldots\right]$$

The big difference between these two forms is that whereas for the trigonometric form, the index k is strictly positive, for the complex exponential form, it covers both positive and negative integers. This is an important difference as we shall see in this chapter. These complex exponentials are an orthogonal set, making them easy to separate from each other. This is the main reason why we pick orthogonal signals to represent something. Just as our 3D world is defined along three orthogonal axes, X, Y, and Z, our signals can be similarly projected on a K dimensional orthogonal set. We can not imagine these K orthogonal axes, but we know mathematically that they exist.

As mentioned earlier, the Fourier series is a sum of weighted sinusoids. By weighted we mean that each sinusoid has its own amplitude and starting phase. The time is continuous but frequency resolution is not, frequency takes on discrete harmonic values. If the fundamental frequency is ω_0, then each ω_k is an integer multiple of ω_0 and is discrete no matter how large k gets. The "distance" between each harmonic remains the same, ω_0.

Let us examine the trigonometric Fourier series equation again.

$$f(t) = a_0 + \sum_{k=1}^{K} a_k \cos(\omega_k t) + \sum_{k=1}^{K} b_k \sin(\omega_k t) \tag{2.10}$$

The coefficients a_0, a_k, b_k (which we call the trigonometric coefficients) are given by (from Chapter 1):

$$\begin{aligned} a_0 &= \frac{1}{T_0}\int_0^{T_0} f(t)\, dt \\ a_k &= \frac{2}{T_0}\int_0^{T_0} f(t)\cos(k\omega_0 t)\, dt \\ b_k &= \frac{2}{T_0}\int_0^{T_0} f(t)\sin(k\omega_0 t) dt \end{aligned} \tag{2.11}$$

The presence of an integral tells us that time is continuous. Now substitute Eq. (2.3) and Eq. (2.4), as the definition of sine and cosine into Eq. (2.10), to get

$$f(t) = a_0 + \sum_{k=1}^{K} \frac{a_k}{2}\left(e^{jk\omega t} + e^{-jk\omega t}\right) + \sum_{k=1}^{K} \frac{b_k}{2j}\left(e^{jk\omega t} - e^{-jk\omega t}\right) \tag{2.12}$$

Rearranging Eq. (2.12) so that each complex exponential is separated, we get

$$f(t) = a_0 + \sum_{k=1}^{\infty} \frac{1}{2}(a_k - jb_k)e^{jk\omega t} + \sum_{k=1}^{\infty} \frac{1}{2}(a_k + jb_k)e^{-jk\omega t} \tag{2.13}$$

Let us redefine two new terms, called the complex coefficients, A_k and B_k as:

$$A_k = \frac{1}{2}(a_k - jb_k) \tag{2.14}$$

$$B_k = \frac{1}{2}(a_k + jb_k) \tag{2.15}$$

Substituting these new coefficients into Eq. (2.13), we get this representation.

$$f(t) = a_0 + \sum_{k=1}^{\infty} A_k e^{jk\omega t} + \sum_{k=1}^{\infty} B_k e^{-jk\omega t} \tag{2.16}$$

By substituting the complex exponential for the sinusoids, we can rewrite the trigonometric coefficients from Eq. (2.11) as:

$$a_k = \frac{2}{T_0} \int_0^{T_0} f(t) \frac{1}{2}\left(e^{jk\omega t} + e^{-jk\omega t}\right) dt \tag{2.17}$$

$$b_k = \frac{2}{T_0} \int_0^{T_0} f(t) \frac{1}{2j}\left(e^{jk\omega t} - e^{-jk\omega t}\right) dt \tag{2.18}$$

These can be expanded as follows.

$$a_k = \frac{1}{T_0} \int_0^{T_0} f(t) e^{jk\omega t} dt + \frac{1}{T_0} \int_0^{T_0} f(t) e^{-jk\omega t}\, dt \tag{2.19}$$

$$b_k = \frac{1}{T_0} \int_0^{T_0} f(t) e^{jk\omega t} dt - \frac{1}{T_0} \int_0^{T_0} f(t) e^{-jk\omega t}\, dt \tag{2.20}$$

You can see that the trigonometric coefficient is split into two parts now, one for each of the exponentials. With some manipulation and using some simple complex math, we can show that the new complex coefficients are given by:

$$A_k = \frac{1}{T_0} \int_0^{T_0} f(t) e^{jk\omega t}\, dt \tag{2.21}$$

$$B_k = \frac{1}{T_0} \int_0^{T_0} f(t) e^{-jk\omega t}\, dt \tag{2.22}$$

Hence, A_k can be thought of as the coefficient of the positive exponential and B_k the coefficient of the negative exponential, analogous to the trigonometric coefficients belonging to the cosine and the sine harmonics, a_k and b_k. The equivalence between the complex coefficients and the trigonometric coefficients takes the exact same form as the Euler equation. We will see in the example problems the effect of the trigonometric coefficients splitting into two parts. The total magnitude at a particular frequency is the same for both complex and the trigonometric forms, but appears split in the spectrum because of the sign of the index.

We still have to contend with the term a_0, the DC component in Eq. (2.16). We can get rid of this DC term by including it inside the summation. We do that by expanding the range of the index k from 0 to ∞ instead of starting at 1. Rewrite Eq. (2.16) as

$$\boxed{f(t) = \sum_{k=0}^{\infty} A_k e^{jk\omega t} + \sum_{k=0}^{\infty} B_k e^{-jk\omega t}} \tag{2.23}$$

Going forward, Eq. (2.23) can be simplified even further by extending the range of coefficients from $-\infty$ to ∞. Since the index now expands from positive domain to negative, we no longer need the negative CE in Eq. (2.23). Now both terms can be combined into one positive CE with a two-sided index so that we get a much more compact and elegant equation for the Fourier series. There is no negative exponential in this equation because the index takes care of that. And here is a much shorter equation for the Fourier series in the complex domain.

$$\boxed{f(t) = \sum_{k=-\infty}^{\infty} C_k e^{jk\omega}} \tag{2.24}$$

The coefficient C_k in Eq. (2.24) is given by

$$\boxed{C_k = \frac{1}{T_0} \int_{-T_0/2}^{T_0/2} f(t) e^{-jk\omega_0 t}\, dt} \tag{2.25}$$

The equations Eq. (2.25) and Eq. (2.24) are called the complex form of the Fourier series. They are rigorously related to the trigonometric form. The magnitude calculated using the trigonometric form is exactly the same as the magnitude from this form. These two equations are the most used form of the Fourier series. The complex coefficients, C_k are of course equivalent to the trigonometric coefficients by this relationship.

$$\begin{aligned} C_k &= \frac{1}{2}(a_k - jb_k) \text{ if } k \geq 0 \\ &= \frac{1}{2}(a_k + jb_k) \text{ if } k < 0 \end{aligned} \tag{2.26}$$

As explained in Chapter 1, for the trigonometric form, the index k is always positive and therefore the spectrum for the Fourier series using the trigonometric form is one-sided. The x-axis for the one-sided spectrum is plotted against frequency starting at zero frequency or the "positive" fundamental frequency, which means that all k integer multiples of the fundamental frequency are positive. Because the index is positive, all frequencies are said to be positive. Now the index is allowed to be negative and this gives rise to the idea that the frequency has become negative, an effect of doing math in the complex domain.

If the two forms are equivalent, does the frequency actually become negative when we use complex exponentials? With complex formulation, the index k spans from $-\infty$ to $+\infty$. We start with the negative index, go through calculations of all negative exponent exponentials and then the positive ones. Note that at no point is the fundamental frequency ever negative. Hence it is not the frequency of the exponential that is negative but *just* the index. The exponential with the negative index k is different from the positive exponential in that the sign of the imaginary part is negative. There is nothing here that implies that the frequency itself is negative.

$$\begin{aligned} e^{+jk\omega t} &= \cos(k\omega t) + j\sin(k\omega t) \\ e^{-jk\omega t} &= \cos(k\omega t) - j\sin(k\omega t) \end{aligned} \tag{2.27}$$

We now equate this form with the trigonometric form that seemingly had only positive frequencies. To accomplish this, we look at what it takes to represent a sine and a cosine using complex exponentials.

$$\begin{aligned} \cos(k\omega t) &= \frac{1}{2}\left(e^{jk\omega t} + e^{-jk\omega t}\right) \\ \sin(k\omega t) &= \frac{1}{2j}\left(e^{jk\omega t} - e^{-jk\omega t}\right) \end{aligned} \tag{2.28}$$

We require both a negative-index exponential and a positive-index exponential to describe both the sines and cosines. Where index k is always positive on the left-hand side of this equation, it is both negative and positive on the right side. This traps us into thinking that frequency has changed sign. While in trigonometric form a positive index is enough to fully and completely represent the signal, in complex form it takes a double-sided index. The spectrum is plotting the product of the index and the frequency, $(k\omega)$, and not just the frequency, ω, on the x-axis. But we very quickly lose sight of this fact. We start talking about positive and negative frequencies because we confuse the *range* of the index with the sense of the frequency.

Double-sided spectrum

This gives rise to the **double-sided spectrum** which spans from values of k from -∞ to $+\infty$. In this type of spectrum, the x-axis represents frequency, but that isn't necessarily so. What we actually plot is the product of the harmonic index and the frequency. Calling it frequency gives us some intuitive comfort but then we have to worry about what a negative frequency means. When we say *negative frequency*, we have, in fact, unknowingly converted a complex idea into simple everyday language. Because of the plotting convention, the sign of the index is often forgotten and the axis is referred to as the frequency axis, spanning both positive and negative domains. In this book, we maintain that there is no such thing as a negative frequency. The idea comes from confusion caused by what the x-axis represents in the complex domain.

The complex coefficient values are one-half of what they are calculated in the trigonometric domain. Sometimes students think that this is because the frequency is being split into two parts, a negative part and positive part with each getting half the coefficient. This is how some authors try to explain the conundrum of positive and negative frequency in relation to Fourier analysis. However, they are just trying to explain a plotting convention. The real story is that we are not plotting frequency on the x-axis of a spectrum but the term $\pm k\omega$. First we call this "frequency" and then we try to explain this error. A CE is a helical function with a sense of rotational direction in addition to its frequency. The index k indicates this direction.

The reason the amplitude values are split in half can be explained intuitively. We have let the index k go from $-\infty$ to $+\infty$, thus, each frequency is now multiplied by both a positive k and a negative k. However, in reality, each frequency has only a certain amount of finite energy, so to make it all work out, the energy of this frequency is split between these two indices. Each index gets half.

There are certain things that are defined only as positive quantities, such as volume, mass, age, etc., and frequency as a physical property is one of those things. It is never negative. In complex domain, we look at the signal in three dimensions, and hence a signal has a frequency and a rotational direction. So the term frequency, although a scalar, may be thought to include this other rotational parameter.

Example 2.1. Compute the complex coefficients of a cosine wave.

$$\begin{aligned} f(t) &= A\cos(\omega_0 t) \\ &= \frac{A}{2}e^{j(k=1)\omega_0 t} + \frac{A}{2}e^{j(k=-1)\omega_0 t} \end{aligned} \tag{2.29}$$

This example is so simple that we can easily deduce the trigonometric coefficients just by looking at the expression. In fact, the equation is itself a perfect Fourier representation. The first thing we do is to write the equation in terms of its Euler form using Eq. (2.28). This form has the two CEs with the appropriate coefficients. The complex coefficients of the sinusoids are of magnitude $A/2$ located at $k=1$ and $k=-1$. We plot the coefficients, C_k, in Fig. 2.5 as the single-sided spectrum as well as the exponential coefficients, C_k, as the double-sided form. The x-axis variable is $k\omega$. Since, ω is a constant, we are really plotting, k, the index. Note, it is not the frequency that is negative but the harmonic index k. However, in a typical plot, the x-axis is labeled as frequency. In these plots, we have labeled it specifically as what it is, the term $k\omega$.

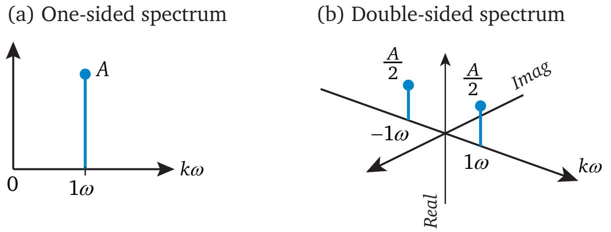

Figure 2.5: The spectrum of $A\cos\omega t$. The trigonometric form gives us a one-sided spectrum with one component located at ω. The complex form shows two components of half the total amplitude in the real plane. The vertical axis is amplitude for both forms. The total energy in both forms is the same, despite being split in exponential form in two parts.

Example 2.2. Compute complex coefficients of a sine wave

$$\begin{aligned} f(t) &= A\sin\omega_0 t \\ &= \frac{A}{2j}e^{j(k=1)\omega_0 t} - \frac{A}{2j}e^{j(k=-1)\omega_0 t} \end{aligned} \tag{2.30}$$

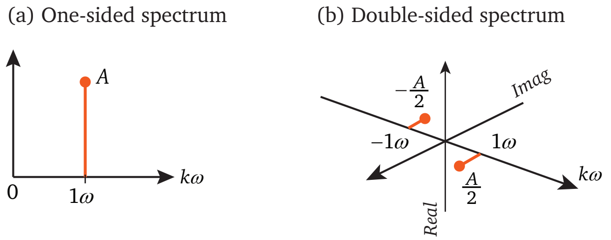

Figure 2.6: The spectrum of $A\sin\omega t$. The trigonometric form gives us a one-sided spectrum with one component located the ω. The complex form shows two components of half the total amplitude of opposite signs in the imaginary plane.

This example is just the same as the cosine example. The single-sided spectrum is easy. It is simply a harmonic of magnitude A and located at $k = 1$ just as it is for the cosine wave in Ex. 2.1. In Eq. (2.30), we write the complex form based on Eq. (2.28), with the amplitudes of the two complex exponentials as $A/2$ and $-A/2$ located at $k = \pm 1$. While the amplitudes were positive for cosine in Ex. 2.1, here they are of opposite signs.

However, there is also the dreaded j in the denominator of Eq. (2.30). What to do with this? The presence of j tells us that the coefficients are on the imaginary axis, so they are to be plotted on the imaginary plane, right-angle to the plane on which a cosine lies. Drawn in 2D form it has no computational effect, only that the vertical axis is called the imaginary axis. However, when there are both cosines and sine waves present in a signal, the coefficients of these two have to be combined not linearly but as a vector sum as seen in Fig. 2.7. This is because the harmonics are orthogonal to each other. When plotting the magnitude, it no longer falls in just the *Real* or the *Imaginary* planes but somewhere in between. In the next example, we combine both of these sinusoids in a single function.

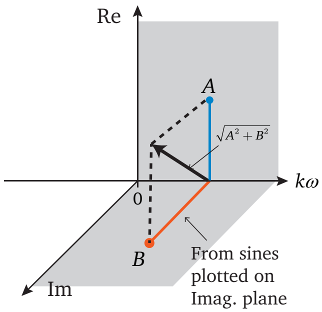

Figure 2.7: Magnitude of the resultant vector if a signal contains both a sine and a cosine.

Example 2.3. Compute the coefficients of $f(t) = A(\cos\omega t + \sin\omega t)$.

We can write this wave as

$$f(t) = \frac{A}{2}e^{j\omega t} + \frac{A}{2}e^{-j\omega t} + \frac{A}{2j}e^{j\omega t} - \frac{A}{2j}e^{-j\omega t} \tag{2.31}$$

Once again, we convert this function to its complex form using Eq. (2.28). From this form, we can pick out the trigonometric coefficients easily. It is simply A for the cosine and A for sine with magnitude equal to square root of $\sqrt{2}A$ located at $\omega = 1$.

The coefficients of the complex exponentials, too can be obtained by looking at the coefficients of the two exponentials in Eq. (2.31). The $e^{j\omega t}$ exponential has two coefficients, at 90° to each other, each of magnitude $A/2$. The vector sum of these is $A/\sqrt{2}$. Same for the negative exponential, except the amplitude contribution from the sine is negative. However, the vector sum or the magnitude for both is the same and always positive. This is shown in Fig. 2.8(c) drawn in a more conventional style showing only the vector sum. Note that the total energy (sum of magnitudes) of the one-sided spectrum is exactly the same as that of complex.

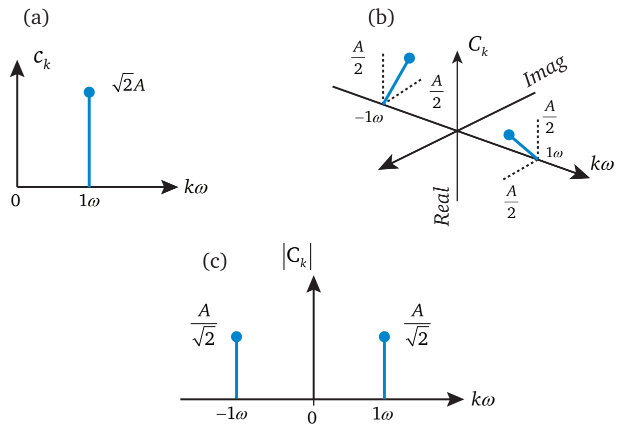

Figure 2.8: Amplitude spectrum of $A\sin\omega t + A\cos\omega t$.

Example 2.4. Compute coefficients of the complex signal $f(t) = A(\cos\omega t + j\sin\omega t)$.

This function is different from the one in Ex. 2.3 in that the sin is located in the imaginary plane. This is a complex signal. We can write this in CE form as:

$$f(t) = \frac{A}{2}e^{j\omega t} + \frac{A}{2}e^{-j\omega t} + \frac{A}{2}e^{j\omega t} - \frac{A}{2}e^{-j\omega t} = Ae^{j\omega t}$$

The coefficients from the sine and the cosine for the negative exponential cancel each other. On the positive side, the contribution from sine and cosine are coincident and add. Therefore, we see a single value at the positive index of $k = 1$ only. For this signal, both the single-sided and double-sided spectrum are identical. This is a surprising and perhaps a counter-intuitive result.

Important observation: *Only real signals have symmetrical spectrum about the origin. Complex signals do not.*

Example 2.5. Compute the coefficients of this signal, $f(t) = A$.

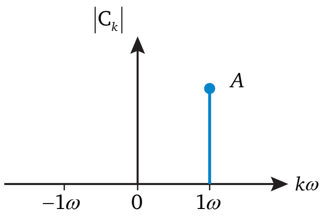

Figure 2.9: Double-sided spectrum of $A\cos\omega t + jA\sin\omega t$.

We can write this function as an exponential of zero frequency.

$$
\begin{aligned}
f(t) &= A\cos(\omega = 0)t \\
&= \frac{A}{2}e^{j(\omega=0)t} + \frac{A}{2}e^{-j(\omega=0)t} \\
&= A.
\end{aligned}
$$

The trigonometric coefficient is A at $\omega = 0$. For the complex representation, we get two complex coefficients, both of amplitude $A/2$ but both at $k = 0$ so their sum is A, that is exactly the same as in the trigonometric representation. The function $f(t)$, a constant, is a non-changing function of time and we classify it as a DC signal. The DC component, if any, always shows up at the origin. The single-sided and double-sided spectrum here are same as well. This signal has only one component at $\omega = 0$; hence, it has only one, coefficient a_0, all the others are zero. Therefore, that is what we are seeing in Fig. 2.10. Just the a_0 coefficient plotted.

Important observation: *A component at zero frequency means that the signal is not zero-mean.*

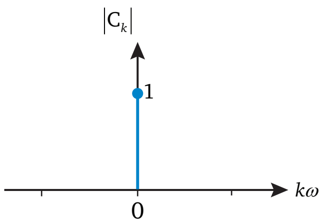

Figure 2.10: Double-sided spectrum of a constant signal of amplitude A. A constant signal is same as a DC value, hence, its spectrum always appears as an impulse at the origin.

Example 2.6. Compute the coefficients of $x(t) = 2\cos^2(\omega t)$.

We express this function in complex form as:

$$\begin{aligned} x(t) &= 2\left(\frac{e^{j\omega t} + e^{-j\omega t}}{2}\right)^2 \\ &= 1 + \frac{1}{2}e^{j2\omega t} + \frac{1}{2}e^{-j2\omega t} \end{aligned}$$

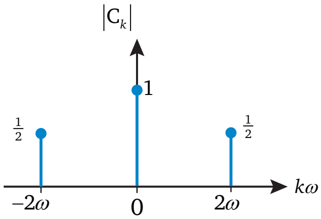

Figure 2.11: Double-sided amplitude spectrum of $2\cos^2(\omega t)$.

The coefficients become obvious when we put the trigonometric form into the complex exponential form. There is just one frequency, 2ω and it has amplitude of $\pm\frac{1}{2}$ at k of ± 2. The coefficient of zero frequency is 1.

Observation: *A squared signal by definition is always positive so the spectrum has a zero-frequency component at the origin.*

Example 2.7. Compute the coefficients of $x(t) = 2\cos(\omega t)\cos(2\omega t)$.

We can express this signal in complex form by making use of this trigonometric identity: $\cos(a+b) = \cos(B)\cos(A) + \sin(A)\sin(B)$.

$$\begin{aligned} x(t) &= \cos(\omega t) + \cos(3\omega t) \\ &= \frac{1}{2}e^{j\omega t} + \frac{1}{2}e^{-j\omega t} + \frac{1}{2}e^{j3\omega t} + \frac{1}{2}e^{-j3\omega t} \end{aligned}$$

Of course, doing this in trigonometric form would have been just as easy; however that is not always true. We draw the spectrum as in Fig. 2.12.

This signal has has four coefficients, $a_{\pm 1}$ and $a_{\pm 3}$; all the others are zero. So, that is what we are seeing in Fig. 2.12.

Example 2.8. Compute the complex coefficients of this real signal.

$$f(t) = \sin(4\pi t) + .8\cos(8\pi t) + .3\sin(14\pi t)$$

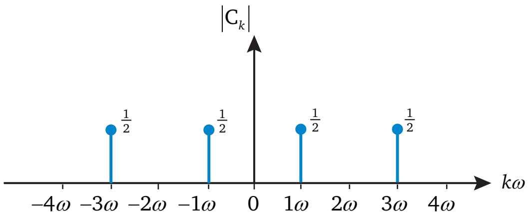

Figure 2.12: Double-sided amplitude spectrum of $2\cos(\omega t)\cos(2\omega t)$.

We want all these frequencies to fall on a harmonic, so we select as the fundamental frequency, $f = 1$. The signal has six components, at $k = \pm 2, \pm 4, \pm 7$. We can write this equation in complex form as:

$$\begin{aligned} f(t) = &\frac{1}{2}e^{j2\pi(k=2)t} + \frac{1}{2}e^{j2\pi(k=-2)t} \\ &+ \frac{0.8}{2}e^{j2\pi(k=4)t} + \frac{0.8}{2j}e^{j2\pi(k=-4)t} \\ &+ \frac{0.3}{2j}e^{j2\pi(k=7)t} + \frac{0.3}{2}e^{j2\pi(k=-7)t} \end{aligned}$$

The contributions at $k = 2$ comes only from a sine, at $k = 4$ from a cosine and at $k = 7$ only from a sine. Note, we plot these on the same line at full magnitude as if j does not exist in the equation. (We will drop mentioning the index k and call it frequency to fall in line with common usage. However, note that it is this sloppiness in terms that causes us to question our sanity and start asking: what is a negative frequency?)

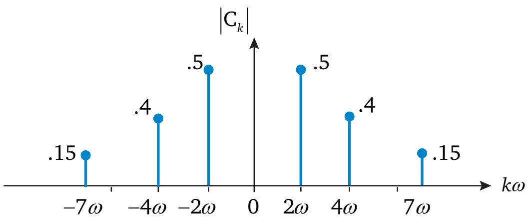

Figure 2.13: Two-sided symmetrical magnitude spectrum.

This signal has six coefficients, $a_{\pm 2}$, $a_{\pm 4}$ and $a_{\pm 7}$. All the others are zero. So that is what we are seeing here in Fig. 2.13. Just those three coefficients plotted on each side, with half the amplitudes from the time domain equation.

Example 2.9. Compute the complex coefficients of this real signal with phase terms. Then compute its power spectrum.

$$x(t) = 3 + 6\cos(4\pi t + 2) + j\sin(4\pi t + 3) - j6\sin(10\pi t + 1.5)$$

We convert this to the complex form as

$$\begin{aligned} x(t) &= 3 + \left(3e^{j4\pi t}e^{2} + 3e^{-j4\pi t}e^{-j2}\right) + \left(2e^{j4\pi t}e^{3} - 2e^{-j4\pi t}e^{-j3}\right) + \left(3e^{j10\pi t}e^{j1.5} - 3e^{-j10\pi t}e^{-j1.5}\right) \\ &= 3 + e^{j4\pi t}\left(3e^{j2} + 2e^{j3}\right) + e^{-j4\pi t}\left(3e^{-j2} - 2e^{-j3}\right) + 3e^{j10\pi t}\left(e^{j1.5}\right) + 3e^{-j10\pi t}\left(e^{-j1.5}\right) \end{aligned}$$

The magnitudes of the exponentials come from the terms in parenthesis. To add, we need to convert them first to rectangular form as follows. (see Appendix A). The CE $e^{j4\pi t}$ has the following coefficients:

$$\begin{aligned} e^{j4\pi t} &\rightarrow \left(3e^{j2} + 2e^{j3}\right) \\ \Rightarrow \left|3e^{j2} + 2e^{j3}\right| &= \sqrt{\left(3\cos(2) + 2\cos(3)\right)^2 + \left(3\sin(2) + 2\sin(3)\right)^2} \\ &= 4.414 \end{aligned}$$

Similarly, the coefficient of the negative exponential is

$$\begin{aligned} e^{-j4\pi t} &\rightarrow \left(3e^{-j2} - 2e^{-j3}\right) \\ \Rightarrow \left|3e^{-j2} - 2e^{-j3}\right| &= \sqrt{\left(3\cos(2) - 2\cos(3)\right)^2 + \left(3\sin(2) + 2\sin(3)\right)^2} \\ &= 3.098 \end{aligned}$$

We draw the spectrum in Fig. 2.14 and note that the spectrum is not symmetric because the signal is complex.

Important Observation: *Most signals we work with are complex and their spectrum are rarely symmetrical.*

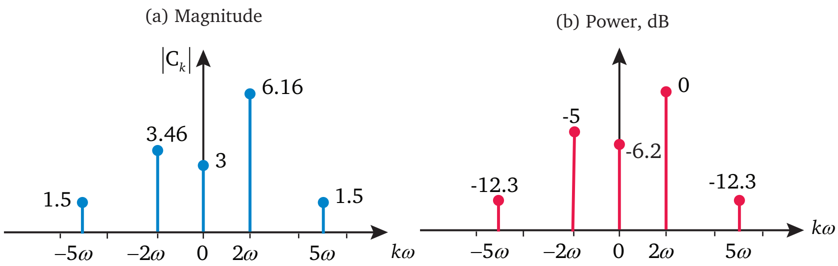

Figure 2.14: (a) Two-sided Magnitude spectrum of a complex signal, (b) its power spectrum computed by squaring each component and converting to dBs.

Power spectrum

You are familiar with this expression of power from circuits. The power is defined as:

$$P = V^2/R$$

Here, V is the voltage or the amplitude of the signal and R the resistance. Assume that R is equal to 1.0; in this case, the normalized power is equal to the voltage squared. If we square the peak voltage, we get peak power, and if we square mean voltage, then mean power is obtained. This idea is exactly the same as Parseval's theorem, which states that the power in a particular harmonic is equal to the square of its amplitude. So, for this particular example, to obtain the Power spectrum, we just square each magnitude, convert it into dBs, and then normalize for maximum power. The result is shown in Fig. 2.14(b).

Now a difficult but an important example, a periodic signal of square pulses.

Example 2.10. Compute the Fourier coefficients of a periodic signal consisting of square pulses. The square pulse is of amplitude $1v$ that lasts τ seconds and repeats every T seconds. (Note that τ and T are different and independent quantities.)

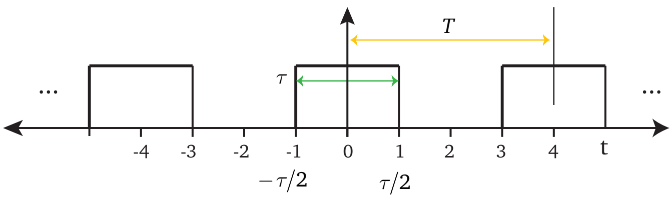

Figure 2.15: A square wave of period T and duty cycle $2/T$.

First, we compute the coefficients for a general case for pulse time equal to τ seconds and repeat time, or the period of the wave, equal to T seconds. The term τ/T is called the duty cycle of the wave.

$$C_k = \frac{1}{T}\int_{-\tau/2}^{\tau/2} 1 \cdot e^{-j2k\pi f_0 t}\, dt$$

Note that outside of $-\tau/2 < t < \tau/2$, the function is zero. This integral is given by

$$\begin{aligned} C_k &= \frac{1}{T}\frac{e^{-(j2\pi k/T)t}}{-j2\pi k/T}\Bigg|_{-\tau/2}^{\tau/2} \\ &= \frac{1}{T}\left(\frac{e^{(j2\pi k/T)\frac{\tau}{2}} - e^{(-j2\pi k/T)\frac{\tau}{2}}}{j2}\right)\frac{1}{\pi k/T} \\ &= \frac{\tau}{T}\frac{\sin(k\pi\tau/T)}{k\pi(\tau/T)} \end{aligned}$$

Replacing the duty cycle term τ with term r, the equation becomes easier to understand.

$$C_k = r\frac{\sin(k\pi r)}{k\pi r} = r\,\mathrm{sinc}(k\pi r).$$

Let us set the duty cycle of this signal to 0.5.

$$r = \frac{\tau}{T} = \frac{1}{2}$$

Substituting r in the above equation, we can write the expression for the coefficients of this signal as:

$$C_k = \frac{1}{2}\mathrm{sinc}(k\pi/2) \tag{2.32}$$

This is the *sinc* function. It comes up so often in signal processing that it is probably the second most important equation in DSP after the Euler's equation.

Now, we can plot the coefficients of the repeating continuous-time square pulse coefficients for various duty cycles. Note how the peakedness of the main lobe changes inversely with the duty cycle. A narrow pulse relative to the period in Fig. 2.16(a), has a shallower frequency response than one that takes up more of the period. The zero crossings occur at inverse of the duty cycle. For $r = 0.5$, the zero crossing occurs at $k = 2$, for $r = 0.25$, the crossing is at $k = 4$, and for $r = 0.75$, the crossing occurs at $n = 1.33$. At $r = 1.0$, the pulse would be a flat line and it will have an impulse as its frequency response. For very small r, the pulse is delta function-like and the response will go to a flat line. Note the usage of words, **frequency response**. This is just another name for the *spectrum*.

Although the equation for this function is fairly easy, it takes a little while to develop intuitive feeling. We cannot overemphasize the importance of this signal and you ought to spend some time playing around with the parameters to understand the effect. We will of course keep coming back to it in the next few chapters.

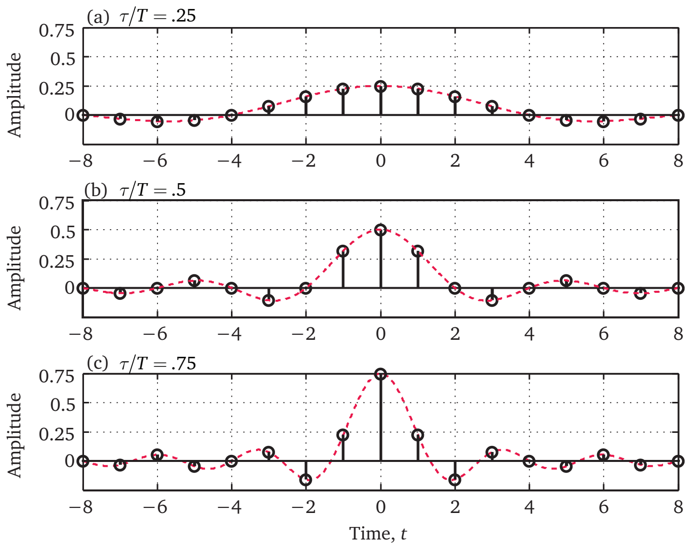

Figure 2.16: Fourier coefficients of a square pulse.
Note that as the pulse gets narrower, its frequency response gets shallower. However, as the duty cycle increases, such as in (c), the response is becoming narrower.

In this chapter, we covered the complex form of the Fourier series as a prelude to the next topic, the Fourier transform. We see that even though the time domain function is continuous and periodic, the Fourier series coefficients (FSC) and hence the spectrum developed is discrete. In Chapter 3, we will see how that affects the Fourier analysis.

Summary of Chapter 2

In this chapter, we looked at the complex exponential as a concise way of representing the Fourier series equation. The complex exponentials make the Fourier series math easier. The spectrum however is now shown as double-sided, which means that the frequency index spans from $-\infty$ to $+\infty$. This has the effect of splitting the trigonometric coefficients into half.
Terms used in this chapter:

- **Euler's equation**
- **Continuous-time complex exponential**, $e^{j\omega t}$ of frequency ω.
- **Complex coefficients of the Fourier series**, C_k
- **Double-sided spectrum**

1. The Euler equation defines a complex exponential as a 3D function consisting of a cosine and sine in quadrature.

$$e^{j\omega t} = \cos\omega t + j\sin\omega t$$
$$e^{-j\omega t} = \cos\omega t - j\sin\omega t$$

2. The Euler expression of a CE can have a positive or a negative exponent. The sign change indicates a change of direction of the function.
3. The sine and cosine alternately can be represented by two CE of different signs. We use the following expressions to represent them in the Fourier series to obtain a complex form of the Fourier series equation.

$$\cos(k\omega t) = \frac{1}{2}\left(e^{jk\omega t} + e^{-jk\omega t}\right)$$
$$\sin(k\omega t) = \frac{1}{2j}\left(e^{jk\omega t} - e^{-jk\omega t}\right)$$

4. To represent a periodic signal using the complex exponentials requires a double-sided harmonic index k, unlike the trigonometric case where the harmonic index is positive.
5. The harmonic index extends from $-K \leq k \leq +K$. K can span from -∞ to $+\infty$.
6. The x-axis now represents values from $-K\omega_0 \leq k\omega_0 \leq +K\omega_0$ and this is often read as representing negative frequency when, in fact, it is the index that is negative.
7. The FSC instead of being of three types can now be represented by a single equation:

$$C_k = \frac{1}{T_0}\int_{-T_0/2}^{T_0/2} f(t)e^{-jk\omega_0 t}\,dt$$

8. The fundamental properties remain the same, the time in this representation is continuous and frequency is discrete with index k, which is an integer.
9. The amplitude spectrum obtained from the complex representation looks different from the one-sided spectrum. We call this spectrum a two-sided spectrum. The amplitude value for a particular harmonic is now split in half for the positive index (so-called positive frequency) and half for the negative index (so-called negative frequency). The 0th component, however, remains the same. This conserves energy and makes both forms equivalent.

Questions

1. What is a complex exponential?
2. What is the expression for a sine in the complex form?
3. If we look along this axis, the CE appears as a circle. What axis is that? Which parameter is not visible in this view?
4. What is the value of $e^{-j\pi}$, $e^{j2\pi}$, and $e^{-j\frac{3}{4}\pi}$.
5. Why is this equation true? $e^{j\pi}+1=0$
6. What is the difference between two complex exponential of the same exponent but different signs, such as $e^{j\omega_k t}$ and $e^{-j\omega_k t}$. If we add these two signals, what do we get in the complex domain?
7. What dimensional space is required to plot a complex exponential?
8. The term phasor is often used in relation to complex exponentials, what is it?
9. If you plot a sinusoid plus its shifted version, $\sin(2\omega t)+\sin(2\omega t+\phi_0)$, what is the phase of the new signal?
10. Represent this sinusoid as a complex exponential: $\cos\left(\omega t+\frac{\pi}{4}\right)$.
11. What is the relationship of the Taylor series to a sinusoid?
12. What is the advantage of using complex exponential as a basis set instead of sines and cosines?
13. Given these CEs, give their expression in the Euler form: e^{-j3}, $(e^{-j4}+e^{-j2})$, and $e^{-j\frac{\pi}{2}t}$.
14. How would you express phase in the complex exponential form?
15. Write the CE form of these signals:
    (a) $\sin(7\pi t+\frac{\pi}{4})$
    (b) $\cos(7\pi t+\frac{\pi}{4})-\sin(7\pi t-\frac{\pi}{4})$
    (c) $\cos(3\pi t+\frac{\pi}{2})$
16. What is the magnitude of this complex signal, e^{-5jt}?
17. The real part of a CE has a peak amplitude of 1 and the imaginary part has the peak amplitude of 2. What is the peak and the average power of this signal?
18. How do we transmit a complex signal? What does it look like?
19. What does division by j mean?
20. What does multiplication by j mean?
21. What is the magnitude and phase of these signals:
    (a) $\left(\sin(\omega t)-\cos(\omega t)\right)$
    (b) $\left(\frac{1}{2}\cos(\omega t)-\sin(\omega t)\right)$
    (c) $\left(2\cos(\omega t)-\frac{1}{2}\sin(\omega t)\right)$
22. What is a single-sided spectrum? What does it represent?
23. Given the amplitude spectrum, how would you compute the power spectrum?

24. What is two-sided spectrum of these signals?
    (a) $f(t) = -\cos(2\pi t) + \cos(9\pi t) + \sin(12\pi t)$
    (b) $f(t) = 2\cos(9\pi t) + 2\cos(18\pi t)$
25. When plotting a two-sided spectrum, what does the x axis represent?
26. If you are given the real and imaginary components of a signal, how do you compute the phase? Is phase changing with time or frequency?
27. For a complex signal, both real and imaginary signals can have nonzero phase, so what is the phase of a complex signal? How is it different from the phases of the components?
28. What is the relationship of the trigonometric coefficients to the complex coefficients?
29. How do explain the idea of negative frequency?
30. Why the spectrum of a complex signal is always one-sided?
31. What is the mean and the peak power of a sinusoid of amplitude 1.0?
32. If we have a periodic signal of square pulses with a duty cycle of 0.1. How much wider is its spectrum as compared to a pulse that has a duty cycle of 0.5?
33. What happens to the magnitude spectrum if phase of the signal changes?

Appendix A: A little bit about complex numbers

We can use complex numbers to denote quantities that have more than one parameter associated with them. A point in a plane is one example. It has a y coordinate and a x coordinate. Another example is a sine wave: it has a frequency and a phase. The two parts of a complex number are denoted by the terms Real and Imaginary, but the Imaginary part is just as real as the Real part. Both are equally important because they are needed to nail down a physical signal.

The signals traveling through air are real signals and it is only the processing that is done in the complex domain. There is a very real analogy that will make this clear. When you hear a sound, the processing is done by our brains with two orthogonally placed receivers, the ears. The ears hear the sound with slightly different phase and time delay. The received signal by the two ears is different and from this our brains can derive fair amount of information about the direction, amplitude and frequency of the sound. So although yes, most signals are real, the processing is often done in complex plane if we are to drive maximum information.

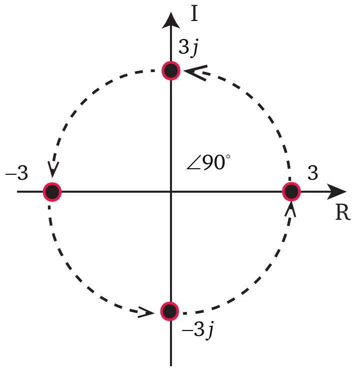

Figure A.1: Multiplication by j shifts the location $0f$ a point on a plane by $90°$.

The concept of complex numbers starts with real numbers as a point on a line. Multiplication of a number by -1 rotates that point $180°$ about the origin on the number line. If a point is 3, then multiplication by -1 makes it -3 and it is now located $180°$ from $+3$ on the number line. Multiplication by -1 can be seen as $180°$ shift. Multiplying this rotated number again by -1, gives the original number back, which is to say by adding another $180°$ shift. Therefore, multiplication by $(-1)^2$ results in a $360°$ shift. What do we have to do to shift a number off the line, say by $90°$? This is where j comes in. Multiply 3 by j, so it becomes $3j$. Where do we plot it now? Herein lies our answer to what multiplication with j does. Multiplication by j moves the point off the line.

Question: What does division by j mean?

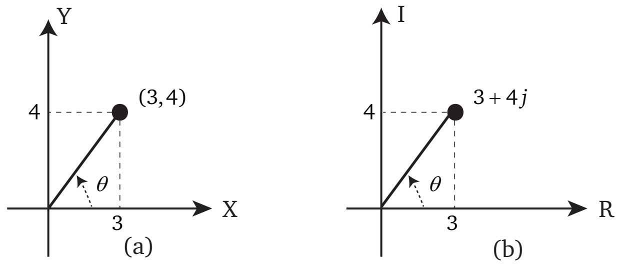

Figure A.2: a. A point is space on a Cartesian diagram. b. Plotting a complex function on a complex plane.

Answer: It is same as multiplying by $-j$.

$$\frac{x}{j} \times \frac{j}{j} = \frac{jx}{-1} = -jx.$$

This is essentially the concept of complex numbers. Complex numbers often perceived as "complicated numbers" follow all the common rules of mathematics. Perhaps, a better name for complex numbers would have been 2D numbers. To further complicate matters, the axes, which were called X and Y in Cartesian mathematics are now called *Real* and *Imaginary*, respectively. Why so? Is the quantity $j3$ any less *real* than 3? This semantic confusion is the unfortunate result of the naming convention of complex numbers and helps to make them confusing, complicated and, of course, complex.

Now let us compare how a number is represented in the complex plane. Plot a complex number, $3 + j4$. In a Cartesian plot, we have the usual $X - Y$ axes and we write this number as $(3,4)$ indicating 3 units on the X-axis and 4 units in the Y-axis. We can represent this number in a complex plane in two ways. One form is called the rectangular form and is given as:

$$z = x + jy$$

The part with the j is called the imaginary part (although of course it is a real number) and the one without is called the real part. Here 3 is the Real part of z and 4 is the Imaginary part. Both are real numbers of course. Note that when we refer to the imaginary part, we do not include j. The symbol j is there to remind you that this part (the imaginary part) lies on a different axis.

$$Re(z) = x$$
$$Im(z) = y$$

Alternate form of a complex number is the **polar form**.

$$z = M\angle\theta$$

where M is the magnitude and θ its angle with the real axis. The polar form that looks like a vector and in essence it is, is called a ***Phasor*** in signal processing. This idea comes from circuit analysis and is very useful in that realm. We also use it in signal processing but it seems to cause some conceptual difficulty. Mainly because, unlike in circuit analysis, in signal processing time is important. We are interested in signals in time domain and the phasor, which is a time-independent concept, is confusing. The phase as the term is used in signal processing is kind of the initial value of phase, where it is an angle in vector terminology. **Question**: If $z = Ae^{j\omega t}$ then what is its rectangular form? **Answer**: $z = A\cos\omega t + jA\sin\omega t$. We just substituted the Euler's equation for the complex exponential $e^{j\omega t}$. Think of $e^{j\omega t}$ as a shorthand functional notation for the expression $\cos\omega t + j\sin\omega t$. The real and imaginary parts of z are given by

$$\begin{aligned} Re(z) &= A\cos\omega t \\ Im(z) &= A\sin\omega t. \end{aligned}$$

Converting forms

1. Given a rectangular form $z = x + jy$ then its polar form is equal to

$$M\angle\theta = \begin{cases} \sqrt{x^2+y^2}\angle\tan^{-1}\frac{y}{x} & \text{if } x \geq 0 \\ \sqrt{x^2+y^2}\angle\left(tan^{-1}\frac{y}{x}+\pi\right) & \text{if } x < 0 \end{cases}$$

2. Given a polar form $M\angle\theta$ then its rectangular form is given by

$$x + jy = M\cos\theta + jM\sin\theta$$

Example 2.11. Convert $z = 5\angle 0.927$ to rectangular form

$$\begin{aligned} Re(z) &= 5\cos(0.927) = 3 \\ Im(z) &= 5\sin(0.927) = 4 \\ &\Rightarrow z = 3 + j4 \end{aligned}$$

Example 2.12. Convert $z = -1 - j$ to polar form

$$\begin{aligned} M &= \sqrt{(-1)^2 + (-1)^2} = \sqrt{2} \\ \theta &= \arctan\frac{y}{x} + \pi, \quad \text{since } x < 0 \\ &= \arctan\frac{-1}{-1} + \pi = \frac{3}{4}\pi \\ \Rightarrow z &= \sqrt{2}\angle\frac{3}{4}\pi. \end{aligned}$$

Example 2.13. Convert $z = 1 + j$ to polar form.

$$\begin{aligned} M &= \sqrt{(1)^2 + (1)^2} = \sqrt{2} \\ \theta &= \arctan\frac{y}{x}, \quad \text{since } x > 0 \\ &= \arctan\frac{1}{1} = \frac{1}{4}\pi \\ \Rightarrow z &= \sqrt{2}\angle\pi/4. \end{aligned}$$

Adding and Multiplying

Add in rectangular form, multiply in polar. Its easier this way.

1. Given $z_1 = a + jb$ and $z_2 = c + jd$ then $z_1 + z_2 = (a + c) + j(b + d)$.
2. Given $z_1 = M_1\angle\theta_1$ and $z_2 = M_2\angle\theta_2$, then $z_1 \cdot z_2 = M_1 M_2\angle(\theta_1 + \theta_2)$.

Example 2.14. Add $z_2 = \sqrt{2}\angle 0.785$ and $z_2 = 5\angle 0.927$.

Convert both to rectangular form.

$$\begin{aligned} z_1 &= 1 + j \text{ and } z_2 = 3 + j4 \\ \Rightarrow z_3 &= (1 + 3) + j(1 + 4) = 4 + j5. \end{aligned}$$

Example 2.15. Multiply $z_1 = 1 + j$ and $z_2 = 3 + j4$.

First convert to polar form and then multiply. Although multiplying these two complex numbers in rectangular format looks easy, in general that is not the case. Polar form is better for multiplication and division.

$$z_1 \cdot z_2 = \sqrt{2}\angle 0.785 \times 5\angle 0.927 = 5\sqrt{2}\angle 1.71$$

Example 2.16. Divide $z_1 = 1 + j$ and $z_2 = 3 + j4$.

$$\frac{z_1}{z_2} = \frac{\sqrt{2}\angle 0.785}{5\angle .927} = \frac{5}{\sqrt{2}}\angle 0.142$$

Conjugation

The conjugate for a complex number z, is given by $z^* = x - jy$. For a complex exponential $e^{j\omega t}$ is the complex conjugate of $e^{-j\omega t}$. In polar format, the complex conjugate is same phasor but rotating in the opposite direction.

Rule: If $z = M\angle\theta$, then $z^* = M\angle - \theta$.

Useful properties of complex conjugates

$$|z|^2 = zz^*$$

This relationship is used to compute the power of the signal. The magnitude of the signal can be computed by half the sum of the signal and its complex conjugate. Note, the imaginary part cancels out in this sum.

$$|z| = \frac{1}{2}\big(z + z^*\big).$$

Chapter 3

Discrete-Time Signals and Fourier Series Representation

Peter Gustav Lejeune Dirichlet
1805 – 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. In 1829, Dirichlet published 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, it introduced the Dirichlet kernel and the Dirichlet integral. – From Wikipedia

Fourier Series and Discrete-time Signals

In the previous two chapters, we discussed the Fourier series as applied to CT signals. We saw that the Fourier series can be used to create a representation of any periodic signal. This representation is made using the sine and cosine functions or with complex exponentials. Both forms are equivalent. In the previous two chapters, the 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

Although some data are naturally discrete such, as stock prices, number of students in a class, etc., many electronic signals we work with are sampled from analog signals, for example, 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 the Fourier analysis, we assume that the sampling is done at uniform time intervals among 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. There are 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 discussing general discrete signals that include *digital signals*. Both of these types of signals are called discrete time (DT) signals. We call the time of a sampling event, 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 captured.

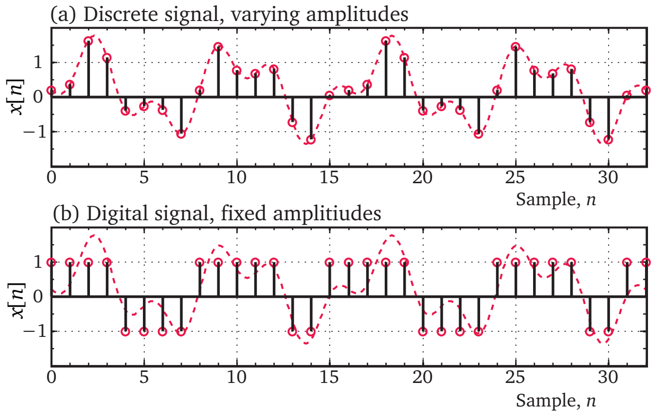

Figure 3.1: Discrete sampling collects the actual amplitudes of the signal at the sampling instant, whereas digital sampling rounds the values to the nearest allowed value. In (b), the sampling values are limited to just 2 values, $+1$ or -1. Hence, each value from (a) has been rounded to either a $+1$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

Generating discrete signals

We often need to distinguish between a CT and a DT signal. First way to distinguish the two is to note that we use letter n as the index of discrete time for a DT signal, whereas we use letter t for the index of time for a CT signal. The second way is that a DT signal is written with square brackets around the time index, n, whereas the CT signal is written with round brackets around the time index, t. The two types, CT and DT, are written as follows:

$x(t)$ A continuous time signal

$x[n]$ A discrete time signal

We can create a discrete signal by multiplying a continuous signal with a sampling signal, as shown in Fig. 3.2(b). This type of signal is called an impulse train and has a mathematical equation in terms of an infinite number of delta functions located at uniform intervals. This is a very special type of sampling function that is not only easy to visualize, but is also considered the ideal sampler. We give it a generic designation of $p(t)$ for the following discussion. The sampled function, x_s, is a simply the product of the CT signal and the sampling function, $p(t)$. We write the sampled function as:

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

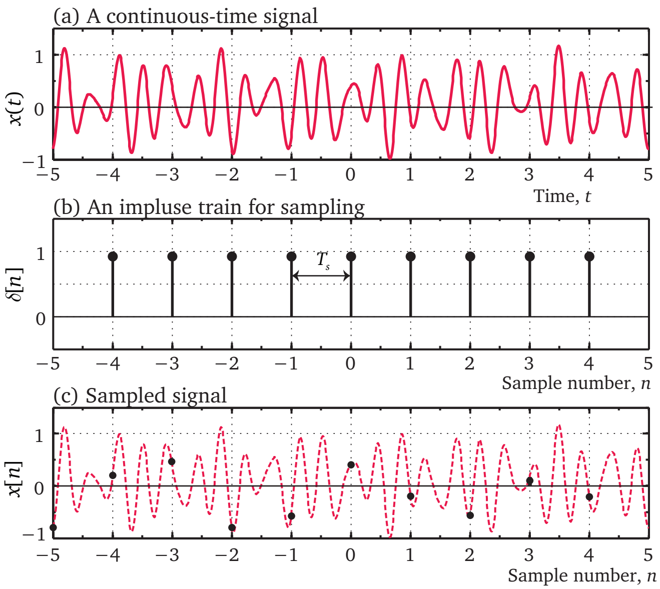

Figure 3.2: A CT signal sampled at uniform intervals T_s with an ideal sampling function. The discrete signal in (c) $x[n]$ consists only of the discrete samples and nothing else. The continuous signal is shown in dashed line for reference only. The receiver has no idea what it is. All it sees are the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

Sampling and Interpolation

Ideal sampling

Let us assume we have an impulse train, $p(t)$ with period T_s as the sampling function. Multiplying this signal with the CT signal, as shown in Eq. (3.1), we get a continuous signal with nonzero samples at the sample instants, referred to as nT_s or n/F_s. Hence, the absolute time is the sample number times the time in between each sample.

The sample time T_s is an independent parameter. The inverse of the sampling time, T_s, is called the **sampling frequency**, F_s, given in samples per second. For a signal sampled with the ideal sampling function, an impulse train, the sampled signal is written per Eq. (3.1) as:

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

The expression for a discrete signal of a sampled version of the CT signal is written as:

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

The term $x(nT_s)$ with round brackets is continuous, as it is just the value of the CT 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 CT 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 nth sample of the sequence.

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)}
```

Reconstruction of an analog signal from discrete samples

Why sample signals? The signals are sampled for one big reason, to reduce their bandwidth. The other benefit 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 discrete samples is called **interpolation**. This is the same idea as plotting a function. We compute a few values at some selected points and then connect those points to plot the continuous representation of the function. The reconstruction by machines, however, is not as straightforward and requires giving them an algorithm that they are able to do. This is where the subject gets complicated.

First, 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 true amplitudes of a signal. Ideal sampling is hard to achieve but for textbook purposes, we assume it can be done. In reality, lack of ideal sampling introduces distortions.

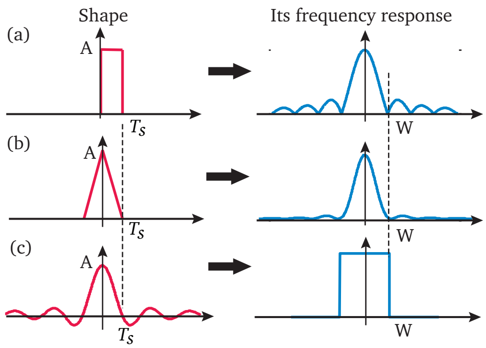

Figure 3.3: We can reconstruct the analog signal by replacing each sample by one of these shapes on the left: rectangular, triangular, or the complicated-looking sinc function. Each has a distinct frequency response as shown on the right.

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 antialiasing filter, a filter with a cutoff frequency that is one-half the sampling frequency prior to sampling. Alternatively we can assume that the sampling frequency chosen is large enough to encompass all the important frequencies in the signal. Let us assume this is also done.

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. Accordingly, the sampled signal $x(nT_s)$ (which is same as $x[n]$) convolved with an arbitrary shape, $h(t)$, is written 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 in Eq. (3.4) 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 packets of waves are then arrayed and added in time. (Note that they are continuous in time.) Depending on the $h(t)$ or the little wave selected, we 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 an impulse train, we write this somewhat simpler equation for the reconstructed signal as:

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

To examine the possibilities for shapes, $h(t)$, following three options are picked; 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. The frequency response is used to determine the effect these shapes will have on the reconstructed signal. Of course, we have not yet fully explained what a *frequency response* is. A frequency response is meant to identify or characterize physical systems. It is done by injecting an impulse into a system and then noting the output. This output is called the **frequency response** of the system. It is often characterized by the magnitude of the response, and the phase, so it is very similar to the idea of a spectrum. In industry these two terms are used interchangeably.

The sampled signals often require that we reconvert them back to their analog form. We discuss three main ways this is done. Intuitively speaking, the process consists of replacing each sample with a little wave.

Method 1: Zero-Order-Hold

Figure 3.3(a) shows a single square pulse. The idea is to 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_s. 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 this interpolation function, a straight line of zero slope connecting one sample to the next. It is 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) to get:

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

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, in which the rectangular pulse is scaled and repeated at each sample.

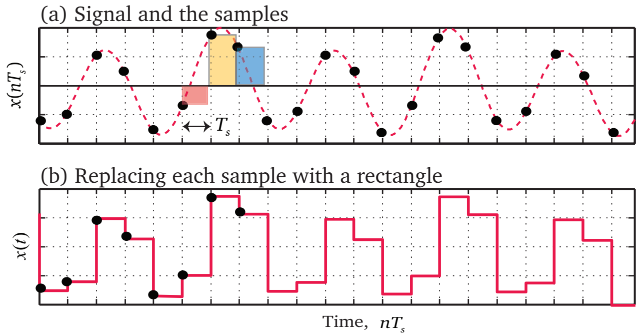

Figure 3.4: A zero-order-hold: hold each sample value to the next sample time.

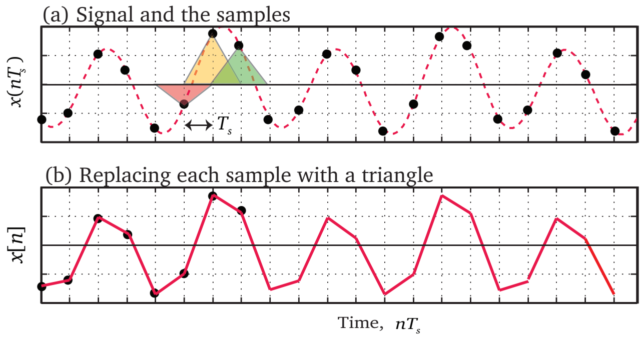

Figure 3.5: A First-order-hold (FOH): replace each sample with a triangle of twice the time-width.

Method 2: First-Order-Hold (linear interpolation)

Zero-order hold gives us a stair-step like signal. Now, instead of a square pulse, we replace each sample with 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

$$x_r(t) = \sum_{n=-\infty}^{\infty} x(nT_s)\,\mathrm{tri}\left(\frac{t - nT_s}{T_s}\right) \tag{3.9}$$

Figure 3.5(b) shows that instead of nonoverlapping rectangles as in ZOH, we use overlapping triangles. This is because the width of the triangle is set to 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? The answer to this query is that machines cannot "see" the samples nor "connect" the samples. Addition is about all they can do well. Hence, this method replaces the 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 as it is noncausal (as it extends into the future) but it is infact an extension of the idea of the first two methods. Both ZOH and FOH are forms of polynomial curve fit. The FOH 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 was done 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. As 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 obtained 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.

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

Figure 3.6 shows 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 Fig. 3.6(b).

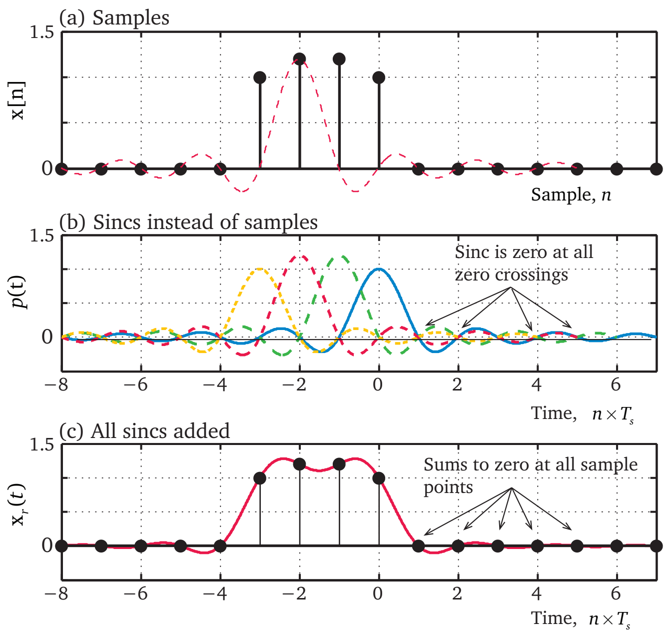

Figure 3.6: Sinc reconstruction: replace each sample with a sinc pulse

Clearly the sinc construction in Fig. 3.6(c) does a very good job. How to tell which of these three methods is better? The ZOH is kind of rough. But to properly assess these methods requires full understanding of the Fourier transform, a topic that will be covered in Chapter 4. Therefore, we will drop this subject now with recognition that a signal can be 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 not only the most versatile, but also most used piece of mathematical concept in signal processing. Hence, we examine the sinc function in a bit more detail here. In Fig. 3.7, the function is plotted in time-domain. This form is called the normalized sinc function. The sinc function is a continuous function of time, t, and is not periodic. It is that nice-looking single-peak signal that oscillates and eventually damps out.

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

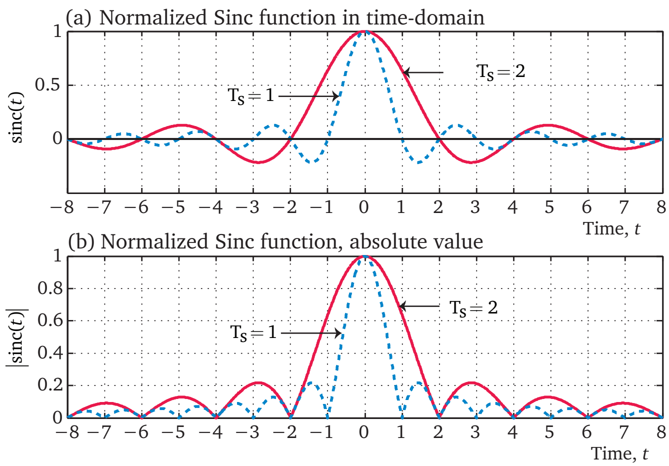

Figure 3.7: The sinc function in time-domain.
The signal is nonperiodic. Its peak value is 1 for the normalized form. The main lobe of the function spans two times the parameter T_s. In (b) the absolute values are shown so the lobes on the negative side are flipped up.

At $t = 0$, its value is 1.0. As seen in Eq. (3.11), 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. By inserting a variable T_s into equation Eq. (3.12), any main lobe width can be created. 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:

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

Figure 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, 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. Furthermore 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.

$$\int_{-\infty}^{\infty} \mathrm{sinc}(2\pi t/T_s)\, dt = \mathrm{rect}(0) = 1. \tag{3.13}$$

The second interesting and useful property, from Eq. (3.12) is that as T_s decreases, the sinc function approaches an impulse. This is seen in Fig. 3.7. A smaller value of T_s means a narrower lobes. Narrow main lobe makes the central part impulse-like and hence it is noted 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 have shown this effect in Chapter 1 by adding many harmonics together and noting that the result approaches an impulse train.

$$\mathrm{sinc}(t) = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{j\omega t}\, d\omega \tag{3.14}$$

This property is best seen in Fig. 3.8, which shows the result of an addition of a large number of harmonic complex exponentials together. The signal looks very much like an impulse train.

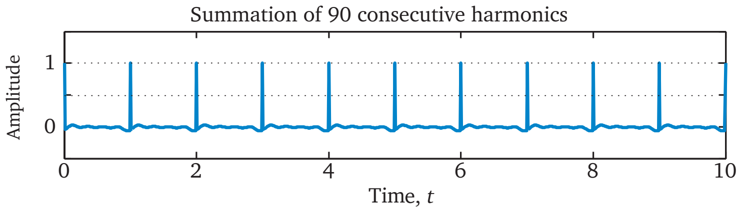

Figure 3.8: The addition of 90 consecutive harmonics ($f = 1, 2, \ldots 90$), *result in very nearly an impulse train.*

The sinc function is also the frequency response to a square pulse. It can be said that it is a representation of a square pulse in the frequency domain. If a square pulse (also called a rectangle, probably a better name anyway) is taken in time domain, then its Fourier series representation will be a sinc, alternately, the sinc function has a frequency representation of a rectangle, which says that it is absolutely bounded in bandwidth. We learn from this that a square pulse in time domain has very large (or in fact infinite) bandwidth and is not a desirable pulse type to transmit.

Sampling rate

How do we determine an appropriate sampling rate for an analog signal? Figure 3.9 shows an analog signal sampled at two different rates; the signal is sampled slowly and sampled rapidly. At this point, our idea of slow and rapid is arbitrary.

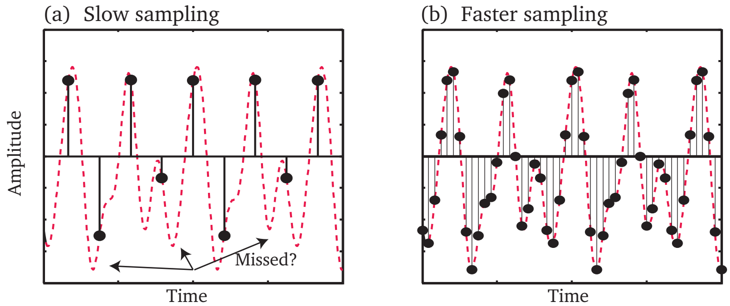

Figure 3.9: The sampling rate is an important parameter, (a) analog signal sampled probably too slowly, (b) probably too fast.

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. However the rate in Fig. 3.9(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 an optimum sampling rate that captures just enough information such that the sampled analog signal can still be reconstructed faithfully from the discrete samples?

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. However the theorem has come to be attributed to Shannon and is thus 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 Swedish scientist 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 **V. A. Kotelnikov** in 1933. In fact 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

$$\boxed{F_s \geq 2f_{\max}} \tag{3.15}$$

The **Nyquist rate** is defined as the case of sampling frequency F_s exactly equal to two 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.

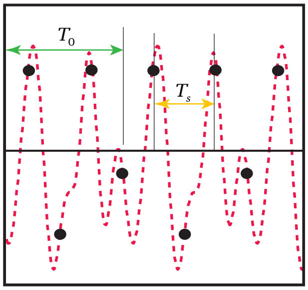

Figure 3.10: There is no relationship between the sampling period and the fundamental period of the signal. They are independent quantities.

A real life-signal will have many frequencies. In setting up the Fourier series representation, we defined 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 as 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_{max}, 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, that is $f_{max} - f_0$.

The Shannon theorem applies, strictly speaking, only to baseband signals or 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. This will not be taken up in this book.

Aliasing of discrete signals

Figure 3.11(a) shows discrete samples of a signal whereas Fig. 3.11(b) shows that these points fit several waves shown. So which wave did they really they come from?

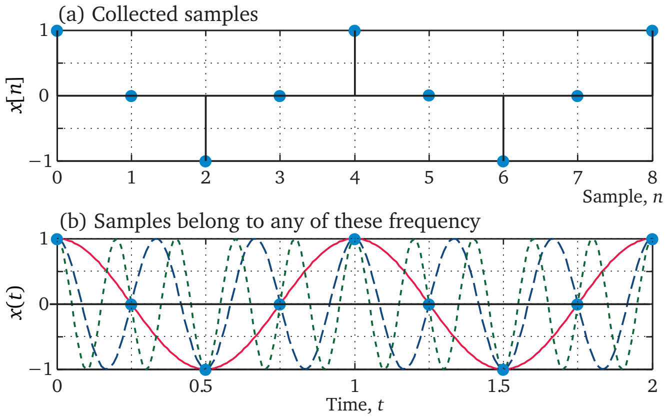

Figure 3.11: Three signals of frequency 1, 3 and 5 Hz all pass through the same discrete samples shown in (a). How can we tell which frequency was transmitted?

The samples in Fig. 3.11(a) could have, in fact, come from an infinite number of other waves which are not shown. This is a troubling property of discrete signals. This effect, that many 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. Its a way of the math telling us that it does not know the real answer among many seemingly correct solutions.

Bad sampling

If a sinusoidal signal of frequency f_0 (because 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 reconstructed, although passing through all the samples, is erroneously mapped to a wrong frequency. This wrong frequency, an *alias*, is not the one that we started with. Given a sampling frequency, we can identify all possible alias frequencies by this expression.

$$y(t) = \sin(2\pi(f_0 - mF_s)t) \tag{3.16}$$

Here, m is a positive integer satisfying this equation

$$\left|f_0 - mF_s\right| \leq \frac{F_s}{2} \tag{3.17}$$

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

Example 3.1. Take the signal with $f_0 = 5$ Hz and $F_s = 8$ Hz or 8 samples per second or samps. We use Eq. (3.16) to find the possible alias frequencies. Here are first three for $(m = 1, 2, 3, \ldots)$ aliases.

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

The first three alias frequencies computed 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 antialiasing filter to remove all higher frequency aliases. Setting $m = 1$ implies the filter is set at frequency of $F_s/2$ or in this case 4 Hz, as per the limitation set on index m by Eq. (3.19). Therefore, we only see alias frequencies that are below this number. Higher order aliases although present are not seen. In computing the possible set of alias frequencies, the value of m is limited by Eq. (3.17).

Figure 3.12 shows the Eq. (3.16) 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 alias frequency but not 11 Hz or higher.

The fundamental pair of components (the real signal before reconstruction) are at +5 and −5 Hz. Now from Eq. (3.16), this spectrum (the bold pair of impulses at ±5 Hz) repeats with a sampling frequency of 8 Hz. Hence, the pair centered at 0 Hz is also 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.

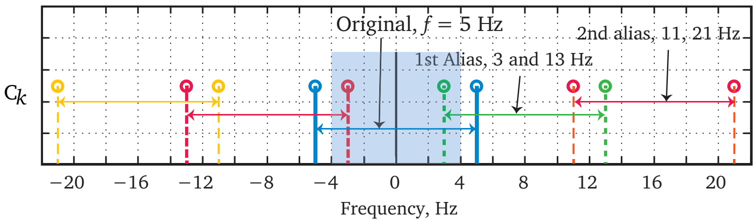

Figure 3.12: The spectrum of the signal repeats with sampling frequency of 8-Hz. Only the 3-Hz signal is below the 4-Hz cutoff.

However, the limit on what we can see is placed by the sampling frequency and Eq. (3.17). A system of sampling frequency of 8 Hz will allow us to detect only the one incorrect alias of 3 Hz as shown in highlighted part of Fig. 3.19.

Good sampling

The sampling theorem states that you must sample a signal at twice or higher times its maximum frequency to properly reconstruct it from its samples. The consequence of not doing this is we get aliases (from Eq. (3.16)) 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.

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

Let us take this signal as shown in Fig. 3.13(a). The signal has four frequencies, which are 1, 2, 3 and 4 Hz. The highest frequency is 4 Hz. We sample this signal at 20 Hz and also at 10 Hz. Both of these frequency choices are above the Nyquist rate, so that is good. The spectrum as computed by the FSC of the four frequencies in this signal is shown in Fig. 3.13(b). (We have not yet discussed how to compute this discrete spectrum and will do so soon, but the idea is the same as for the CT case.)

A very important fact for discrete signals is that the FSC 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 theoretically continues forever on both sides of the principal alias, shown in a dashed box in the center in Fig. 3.14. The spectrum centered at the zero frequency is called the *Principal alias*.

The DT version of the CT signal is given by setting CT time t to n/F_s in the following expression. Here n is the sample number and F_s, the sampling frequency.

$$x(t) = 0.2\sin(2\pi\tfrac{n}{F_s}) + \sin(4\pi\tfrac{n}{F_s}) + 0.7\cos(6\pi\tfrac{n}{F_s}) + 0.4\cos(8\pi\tfrac{n}{F_s})$$

Figure 3.13 shows the signal sampled at 20 Hz, and we see that there is plenty of distance

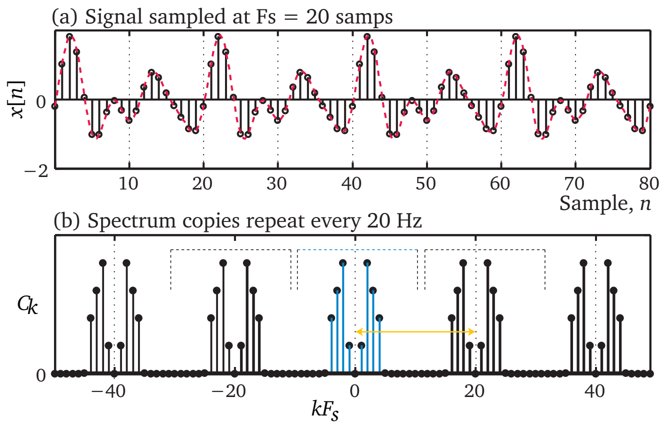

Figure 3.13: (a) A composite signal of several sinusoid is sampled at twice the highest frequency. (b), the discrete coefficients repeating with the sampling frequency, $F_s = 20$ Hz.

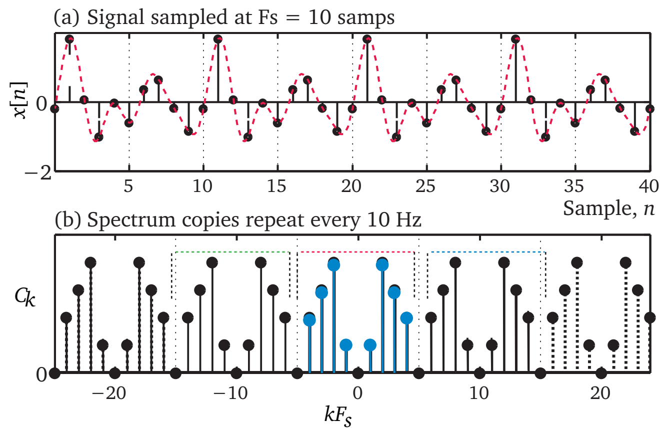

Figure 3.14: The signal sampled at $F_s = 10$ in (a), results in much closer replications in (b).

between the copies. This is because the bandwidth of the signal is only 8 Hz, hence, there is 12 Hz between the copies. Figure 3.14(b) shows the spectrum for the signal when sampled at 10 Hz. The spectrum is 8 Hz wide but now the spectrum are close together with only 2 Hz between the copies.

Decreasing the sampling rate decreases the spacing between the alias spectrum. The copies would start to overlap if they are not spaced at least two times the highest frequency of the signal. In such a case, separation of one spectrum from another becomes impossible. When nonlinearities 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 antialiasing filter. As 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 overlap, the effect cannot be gotten rid of by filtering. As we do not 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, often 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 that range in frequency from 20 to 20,000 Hz. When recording these signals, they are typically sampled at 44.1 kHz (CD), 48 kHz (professional audio), 88.2 kHz, or 96 kHz rates depending on quality desired. Signals subject to nonlinear 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 DT, 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. It is given a special name of its own, **digital frequency** and we use the symbol Ω to designate it. We can show the similarity of this frequency to an analog frequency by noting first how these two forms of signals are written.

$$\begin{aligned}\text{Analog signal}: \ & x(t) = \sin(2\pi f_0 t) \\ \text{Discrete signal}: \ & x[n] = \sin(2\pi f_0 n T_s)\end{aligned} \tag{3.18}$$

The first expression is a continuous signal and the second a discrete signal. For the discrete signal, replace CT, t, with nT_s. Alternately, write the discrete signal, as in Eq. (3.19) 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.)

$$x[n] = \sin\left(\frac{2\pi f_0}{F_s} n\right). \tag{3.19}$$

Digital frequency, only for discrete signals

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

$$\Omega = \frac{2\pi f_0}{F_s} \tag{3.20}$$

Substitute this definition of digital frequency into Eq. (3.19) to obtain a expression for a sampled sinusoid in discrete time.

$$x[n] = \sin(\Omega n) \tag{3.21}$$

Here are two analogous expressions for a sinusoid, a discrete and a continuous form.

$$\begin{aligned} \text{Analog signal} : x(t) &= \sin(\omega t) \\ \text{Discrete signal} : x[n] &= \sin(\Omega n) \end{aligned} \tag{3.22}$$

The digital frequency Ω is equivalent in concept to an 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, given in real time. 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 must also cover 2π radians, from which we have this relationship.

$$\Omega_0 N_0 = 2\pi \tag{3.23}$$

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. Furthermore, when it is multiplied by the fundamental period N_0, an integer multiple of 2π is obtained. Hence, a periodic discrete signal repeats with a frequency of 2π, which is the same condition as for an analog signal. The only difference being that analog frequency is defined in terms of time it takes to cover 2π radians and digital frequency in terms of samples needed to cover the same.

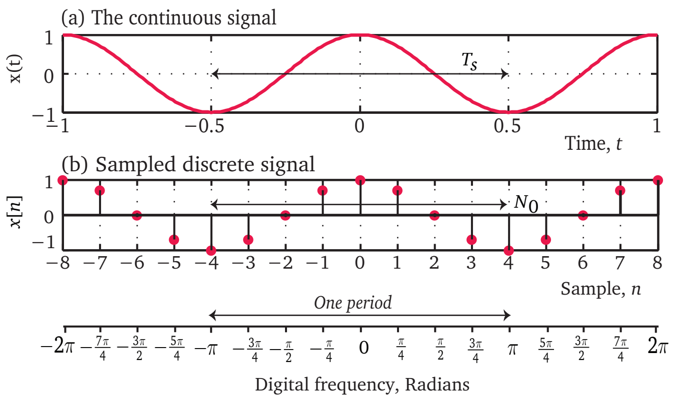

Figure 3.15: A discrete signal in time domain can be referred by its sample numbers, n (1 to N) or by the digital frequency phase advance. Each sample advances the phase by $2\pi/N$ radians. In this example, N is 8. In (a), the x-axis is in terms of real time. In (b) the x-axis is in terms of sample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

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

The sample number can be replaced with a instantaneous phase value for an alternate way of showing the discrete signal. Figure 3.15(c) shows that there are 8 samples over each 2π radians or equivalently a discrete frequency of $2\pi/8$ radians per sample. This is the digital frequency, Ω_0 which is pushing the signal forward by these 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 a signal to be *periodic*. Therefore, 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.

$$\begin{aligned} \text{Continuous signal: } & x(t) = x(t+T) \\ \text{Discrete signal: } & x[n] = x[n+N] \end{aligned} \tag{3.24}$$

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*. As we use sinusoids as basis functions for the Fourier analysis, let us apply this general condition to a sinusoid. To be periodic, a discrete sinusoid that is defined in terms of the digital frequency and time sample, n, must repeat after N samples, hence, it must meet this condition.

$$\cos(\Omega_0 n) = \cos\big(\Omega_0(n+N)\big) \tag{3.25}$$

We expand the right-hand-side of Eq. 3.25 using this trigonometric identity:

$$\cos(A+B) = \cos(A)\cos(B) - \sin(A)\sin(B) \tag{3.26}$$

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

$$\cos\big(\Omega_0(n+N)\big) = \cos(\Omega_0 n)\cos(\Omega_0 N) - \sin(\Omega_0 n)\sin(\Omega_0 N). \tag{3.27}$$

For Eq. (3.25) to be true, we need the underlined terms on the RHS to be equal to 1 and 0 respectively.

$$\cos(\Omega_0 n) = \cos(\Omega_0 n)\underbrace{\cos(\Omega_0 N)}_{=1} - \sin(\Omega_0 n)\underbrace{\sin(\Omega_0 N)}_{=0} \tag{3.28}$$

For these two conditions to be true, we must have

$$\begin{aligned} \Omega_0 N &= 2\pi k \text{ or} \\ \frac{\Omega_0}{2\pi} &= \frac{k}{N} \end{aligned} \tag{3.29}$$

It is concluded 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 neither 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. (3.29).

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

$$\boxed{N = \frac{2\pi k}{\Omega_0}} \tag{3.30}$$

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?

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

The digital frequency of this signal is $2\pi/5$ because that is the coefficient of time index n.

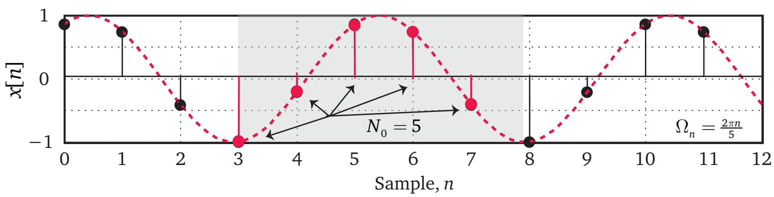

Figure 3.16: Signal of Ex. 3.3.

The fundamental period N_0 is equal to 5 samples that we find using Eq. (3.30) 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?

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

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

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

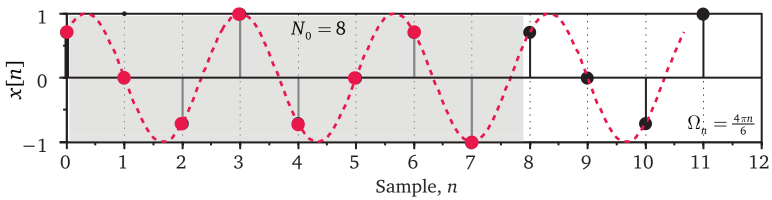

Figure 3.17: Signal of Ex. 3.4. The period of this signal is 8 samples.

The period is 8 samples but it takes 6π radians to get the same sample values again. As we see, the signal covers three 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. (3.30) is equal to

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

As 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.

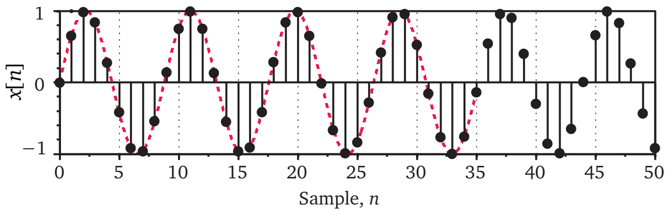

Figure 3.18: Signal of Ex. 3.5 that never achieves an integer number of samples in any integer multiple of 2π.

Discrete Complex Exponentials as Basis of DT Fourier Series

The **CT 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 DT signals using DT complex exponentials as the basis functions.

A discrete complex exponential is written by replacing t in CT 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. (3.22).

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

The discrete form of the fundamental is expanded as follows. We show a *single* CE of digital frequency $\frac{2\pi}{N}$ for variable n, the DT index of time. The harmonic factor k has not yet been included in this equation.

$$e^{j\Omega_0 n} = e^{j\left(\frac{2\pi}{N}\right)n} \tag{3.31}$$

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, such as $2k\pi/N$ for all k? Well, no. If a signal has fundamental digital frequency of $\pi/5$, then is frequency $2\pi/5$ the next harmonic? Yes, and no, because this method leads us into problems.

The range of the digital frequency is 2π. To obtain its next 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 $(\pi/5 + 2\pi) = 12\pi/5$ for $k = 1$. This is a very important point. The analog and discrete harmonics have equivalent definitions for purposes of the Fourier analysis. We will see, however, that they do not display the same behavior. We cannot use these traditionally defined discrete harmonics for the 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 $e^{j(2\pi/N+2\pi k)n}$ for each k

are not distinct from each other. They are said to be non-distinct for each harmonic index, k. This is easy to see from this proof.

$$\begin{aligned} \phi_k[n] &= e^{j(2\pi/N+2\pi k)n} \\ &= e^{jk(2\pi/N)n} \underbrace{e^{j(2\pi kn)}}_{=1} \end{aligned} \tag{3.32}$$

Note that the second part can be written as in Eq. (3.33) and is equal to 1.0 because the cosine wave at any integer $2k\pi$ radians is always 1 and the sine for the same is always 0.

$$e^{j(2k\pi n)} = \cos(2kn\pi)_{=1} + j\sin(2kn\pi)_{=0} = 1 \tag{3.33}$$

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! Hence this method of getting at distinct harmonics is for naught!

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

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 $n/4$. ($T_s = 0.25$ hence, $F_s = 4$) We write this discrete signal as:

$$x[n] = e^{-j\frac{2\pi}{24}n}$$

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

$$\begin{aligned} \text{Fundamental}: & \; e^{-j\left(\frac{\pi}{24}\right)n} \\ \text{Harmonic 1}: & \; e^{-j\left(\frac{\pi}{24}+2(k=1)\pi\right)n} = e^{-j\left(\frac{\pi}{24}+2\pi\right)n} \\ \text{Harmonic 2}: & \; e^{-j\left(\frac{\pi}{24}+2(k=2)\pi\right)n} = e^{-j\left(\frac{\pi}{24}+4\pi\right)n} \end{aligned} \tag{3.34}$$

We plot all three of these in Fig. 3.19. Why is there only one plot in this figure? Simply because the three signals from Eq. (3.34) are identical and indistinguishable.

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

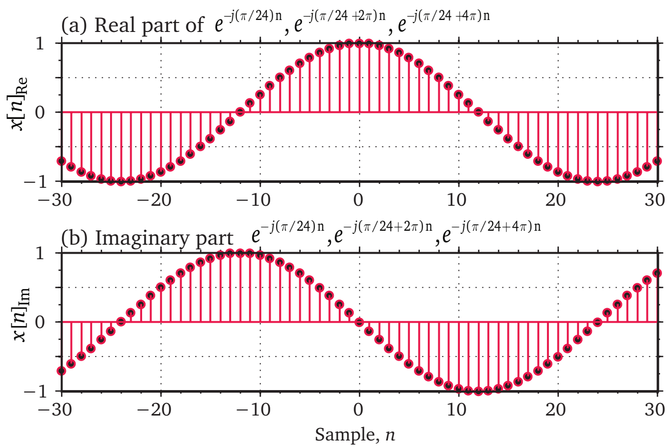

Figure 3.19: Signal of Example 3.6 the imaginary part is a sine wave, (b) the real part, which is of course a cosine. The picture is same for all integer values of k.

The secret hiding place of the discrete harmonics is *inside* the 2π range, 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 in number of samples.

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

$$\Omega_k[n] = k\frac{2\pi}{N}n \quad \text{for } k = 0, 1, \ldots, N-1. \tag{3.35}$$

Increasing k beyond $N-1$ will give the same harmonic as for $k=0$ again. This is of course equivalent to a full 2π radian phase traversal. Hence discrete harmonics are obtained by not by an increment of 2π but by the digital frequency itself, which is a small portion of 2π.

Example 3.7. Let us check this idea 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 is equal to 6 samples. We now know that the signals of digital frequencies $2\pi/6$ and $14\pi/6$ (which is $2\pi/6+2\pi$) are exactly the same. So, the digital frequency is increased, not by 2π but instead in six steps, each time increasing by $2\pi/6$ so that after six 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 move 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

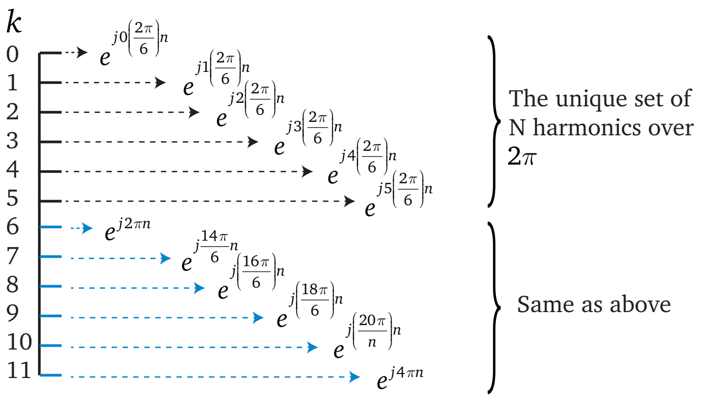

Figure 3.20: Discrete harmonic frequencies in the range of $2\pi m$ to $2\pi(m+1)$.

harmonic, if we move in six steps, we get these six 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, steps from 0 to $K-1$, where K is used as the total number of harmonics. 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.

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

Figure 3.21 plots these discrete complex exponentials so we can examine them. 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) at first but then after three increments (half of the period, N) starts 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 us take a closer look. The first row in Fig. 3.31 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

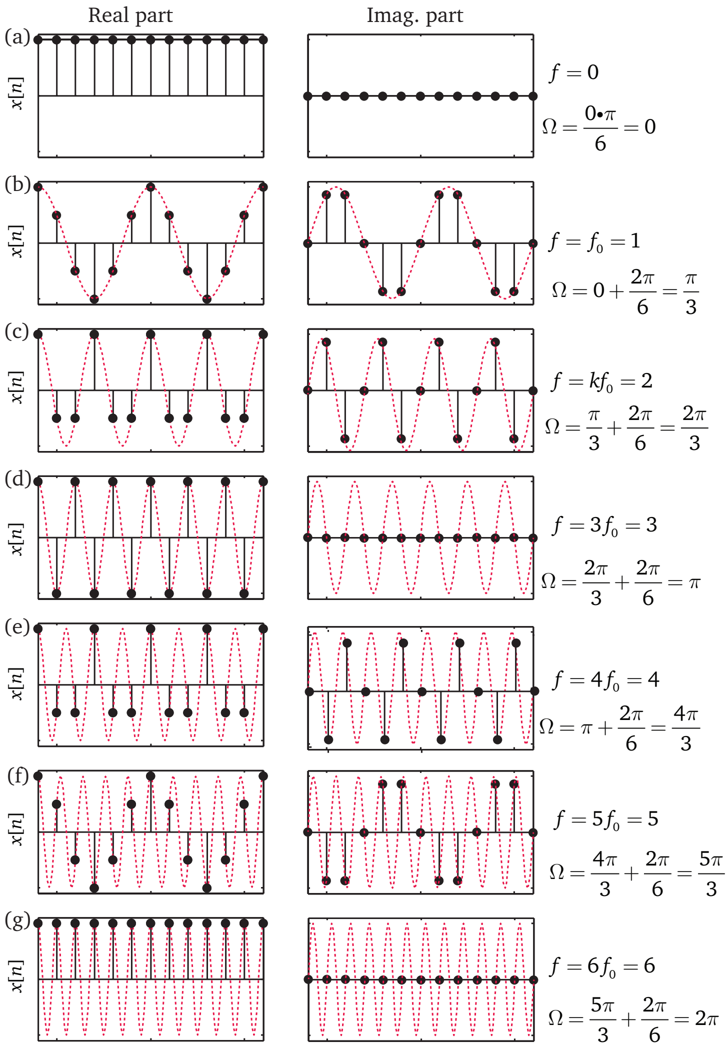

Figure 3.21: The real and the imaginary component of the discrete signal harmonics. They are all different.

Hz, the discrete samples come from $\cos(2\pi/6)$. In (c), we see a continuous signal of 2 Hz and discrete samples from $\cos(2\pi/3)$. It is seen 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 intermediate 6 harmonic in a 2π range are distinct where those obtained by increasing the frequency by 2π are not!

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 forms the discrete Fourier series (DFS) representation of the signal. Unlike the CT 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 Representation

The **Discrete-time Fourier series (DTFS)** is the discrete representation of a DT periodic signal by a linear weighted combination of these K_0 distinct complex exponentials. These distinct orthogonal exponentials exist within just one cycle of the signal with cycle defined as a 2π phase shift. As 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.

$$\boxed{x[n] = \frac{1}{N_0}\sum_{k=0}^{K_0-1} C_k e^{jk\frac{2\pi}{N_0}n}} \tag{3.36}$$

Equation (3.36) is the Fourier series representation of a discrete periodic signal. C_k are the complex Fourier series coefficients (FSC), which we discussed in Chapter 2. There are k harmonics, hence you see the coefficients with index k. In the exponent to the CE, the term $\frac{2\pi}{N_0}$ is incremental frequency change, or the fundamental digital frequency of the signal. Term k increments this fundamental frequency, with n is the index of time for the DT signal.

The DTFS coefficients of the kth 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 Inverse DTFSC equation is given by the kth coefficient and hence, the kth coefficient is given by

$$\boxed{C_k = \sum_{n=0}^{N_0-1} x[n]e^{-j\Omega_0 nk}} \tag{3.37}$$

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). Therefore, we have:

$$C_{k+mN_0} = \sum_{n=0}^{N_0-1} x[n]e^{-j(k+mK_0)\Omega_0 n} = C_k \tag{3.38}$$

This says that the harmonics repeat, and hence the coefficients also repeat. The spectrum of the discrete signal (comprising 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, where, K_0 is equal to N_0.

This is a very different situation from the continuous signals, which do not have such behavior. The CT coefficients are unique for all values of k.

DTFS Examples

Example 3.8. Find the DT 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.

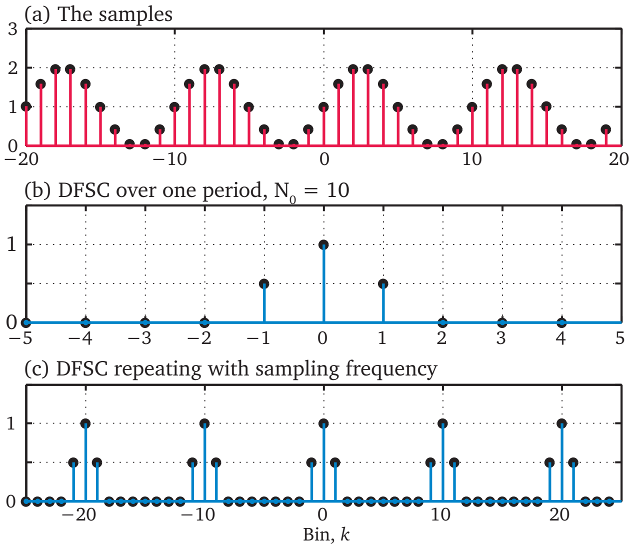

Figure 3.22: Signal of Ex. 3.8 and its Fourier coefficients (a) The discrete signal with period = 10, (b) the fundamental spectrum, (c) the true repeating spectrum.

Now we write the signal in complex exponential form.

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

Note that because the signal has only three components, corresponding to index $k = -1, 0,$ and 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, 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, the coefficients can be seen 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. This in shown in the spectrum of the signal in Fig. 3.23(c). In Figure 3.23(b), only the fundamental spectrum is shown, but infact the spectrum repeats every 10 samples, forever.

Example 3.9. Compute the DTFSC of this discrete signal.

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

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 FSC of this signal repeat with a period of 20. Each harmonic exponential varies in digital frequency by $2\pi/20$. Based on this knowledge, it can be shown that the $2\pi/5$ exponential falls at $k = 4$, $2(\pi/10) = 2\pi/5$ and exponential $2\pi/4$ falls at $k = 5$, $5\pi/10 = \pi/2$.

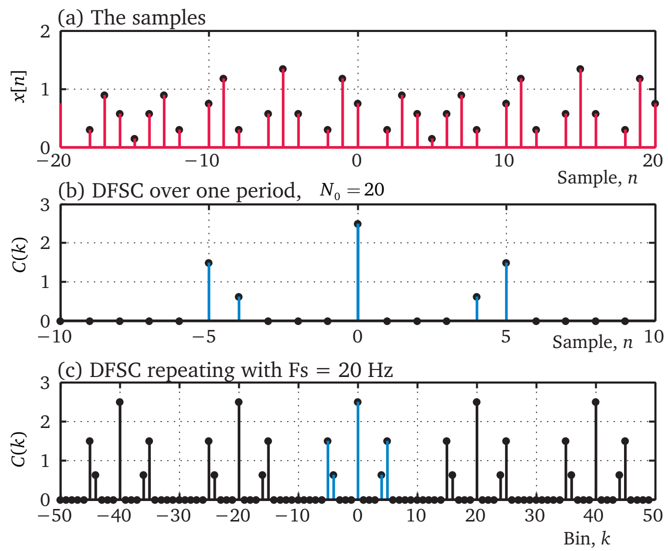

Figure 3.23: Signal of example 3.9 (a) The discrete signal with period = 20, (b) the fundamental spectrum, (c) the true repeating spectrum.

We can also write this signal as

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

From here, we see that the zero-frequency harmonic has a coefficient of 5/2. The frequencies $\pm 2\pi/5$ and $\pm 2\pi/4$ have coefficients of 3/2 and 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.(3.37).

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

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, the relationship is unknown. We first express the complex exponential in its Euler form. As known already

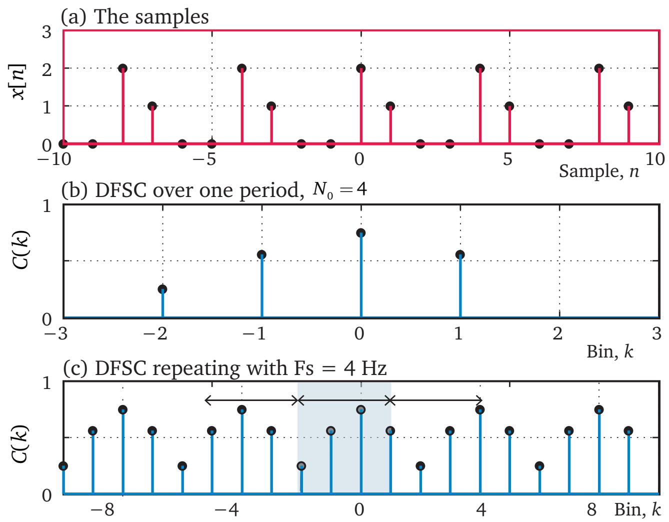

Figure 3.24: Signal of Ex. 3.10. The discrete signal with period = 4 (b) the fundamental spectrum, (c) the true repeating spectrum.

(from Chapter 2), the values of the complex exponential for argument $\pi/4$ are 0 and 1 respectively for the cosine and sine. It can be written in a concise form as:

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

We get for the above exponential

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

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

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

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

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

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

These four values can be seen repeated in Fig. 3.24(c).

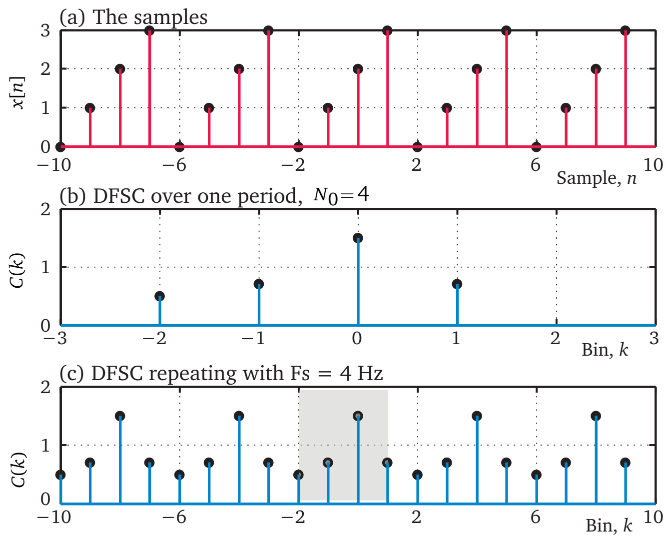

Figure 3.25: (a) The discrete signal with period = 4; (b) the fundamental spectrum; (c) the true repeating spectrum.

Example 3.11. Find the DTFSC of the following sequence.

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

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

$$C_k = \sum_{n=0}^{3} x[n] W_4^{nk}, \quad k = 0, \pm1, \pm2, \ldots.$$

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 DTFSC of a square pulse signal of width L samples and period N.

Once again, we examine the FSC of a square pulse signal. The signal is discrete now, which means that it has a certain number of non-zero samples along with the zeros for the rest, making up one period of DT data. We set the period of the pulse to L samples, which is the width of the pulse. The length of the period is N samples. The duty cycle is defined as the ratio of the pulse width, L and the period, N. The pulse is not centered at the origin in this case, as shown in Fig. 3.26. This has the effect of introducing a phase term, as we shall see in the result.

To compute the DTFSC, the following important property of geometric series is used.

$$
\boxed{\sum_{n=0}^{M} a^n = \frac{1-a^M}{1-a}, \quad |a| < 1.} \tag{3.39}
$$

The coefficients of this signal are given as:

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

Now, we use the geometric series in Eq. (3.39) 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 Eq. (3.39), 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, 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, $e^{-jk\frac{2\pi}{N}\frac{1}{2}}$, 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 Eq. (3.40).

$$C_k = \frac{1}{N}\frac{\sin(kL\pi/N)}{\sin(k\pi/N)}\underline{e^{-jk\frac{\pi}{N}(L-1)}} \tag{3.40}$$

It does not look much simpler! However, if we look only at its magnitude (the front part) it is 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 the front part and the phase by the underlined part. If you try to plot this function for $k = 0$, you will get 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.

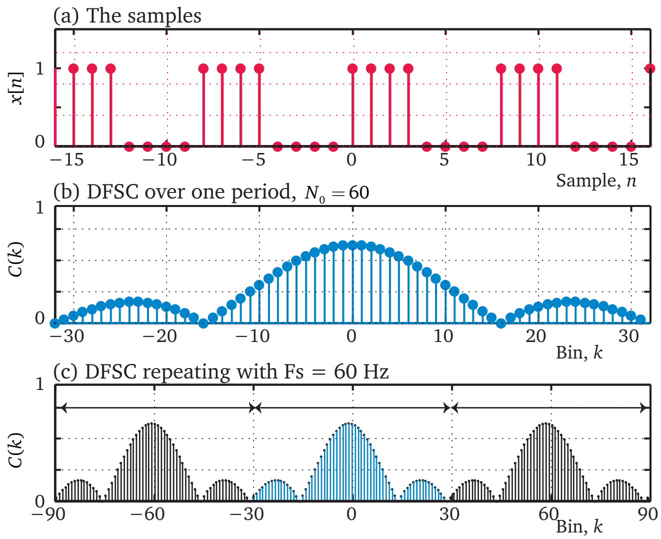

Figure 3.26: (a) The discrete signal; (b) DT Fourier series (DTFSC) coefficients, (c) the true repeating spectrum.

In Chapter 2, we computed the spectrum of a square pulse signal. The spectrum was *nonperiodic* sinc function, whereas this spectrum is *periodic*, i.e. repeats, a consequence of discrete time domain. This is an important new development and worth understanding. This function is called the **Dirichlet function** and is a periodic form of the sinc function. In Matlab, the Dirichlet function is plotted as follows:

```
% DTFSC a square pulse train
N = 10; % Period
```

```
n = -15:14
n2 = -3*pi: .01: 3*pi'
L = 5;  % Wi\dt h 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')
```

In Figure 3.27 the coefficients of various square pulse signals are plotted. These should be studied to 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 the spectrum repeat with the sampling frequency. This was not the case for CT signal. The Fig. 3.27 shows that as the pulses get wider, the response gets narrower. When the pulse width is equal to the period, hence it is all impulses, an impulse train for a response. This is a very important effect to know.

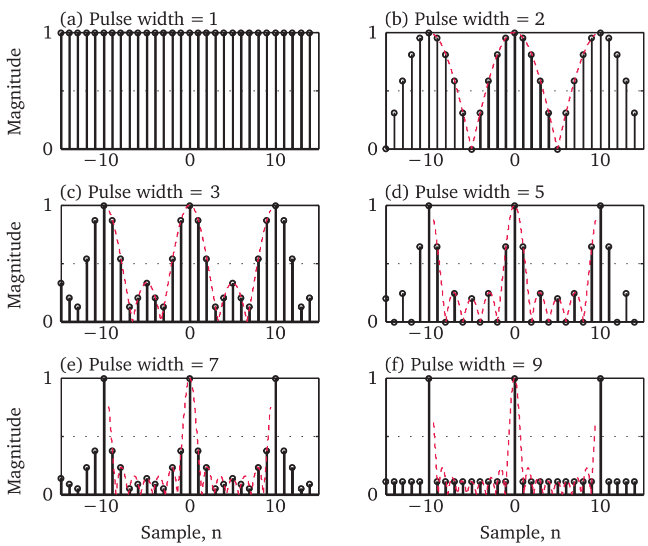

Figure 3.27: Spectrum of a periodic discrete square pulse signal. We see the DTFSC change as the duty cycle (the width of the pulse) of the square pulse increases relative to the period. In (a), all we have are single impulses and, hence, the response is a flat line. In (f), as the width gets larger,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

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. As per the Parseval's theorem, it is defined as:

$$\sum_{n=-\infty}^{\infty} |x[n]|^2 = \sum_{k=-\infty}^{\infty} \left|X[\tfrac{F_s}{n}k]\right|^2 \tag{3.41}$$

Similar to the idea from circuits, to obtain power, we square the time-domain sample amplitude to obtain the instantaneous power at that time instant. If the unit of the amplitude is voltage, the units become voltage-squared. If all such individual squared-amplitudes are summed, we get the total power in the signal (assuming a unity resistance.) Alternately, we can sum the square of the coefficients of each harmonic. They have units of amplitude too. This also gives us units of voltage-squared and can be used to convert a DTFSC into a Power spectrum by plotting the quantity $(C_k)^2$ as a function of the index, k.

Matrix method for computing FSC

DTFSC are computed in closed-form for homework problems only. For practical applications, Matlab and other software and hardware devices are used. We will now look at a matrix method of computing the DTFSC.

Let us first define this common term. It looks strange and confusing, but it is a very simple idea. We use it to separate out the constant terms.

$$W_{N_0} \triangleq e^{-j\frac{2\pi}{N_0}}. \tag{3.42}$$

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. Using this factor, the DFS and the inverse DFS (IDFS) can be written as:

$$\begin{aligned} \text{DFS: } X[k] &= \sum_{n=0}^{N_0-1} x[n] W_{N_0}^{nk} \\ \text{IDFS: } x[n] &= \frac{1}{N_0} \sum_{k=0}^{N_0-1} X[k] W_{N_0}^{-nk} \end{aligned} \tag{3.43}$$

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

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

These terms can be precomputed and stored to make computation quicker. They are the basic idea behind the FFT algorithm which speeds up computation. The DTFSC equation can be setup in matrix form using the Twiddle factor and writing it in terms of two variables, the index n and k. Now we write:

$$C_k = \frac{1}{K_0} x[n] \begin{vmatrix} W^{-0\times0} & W^{-1\times0} & W^{-2\times0} & W^{-3\times0} \\ W^{-0\times1} & W^{-1\times1} & W^{-2\times1} & W^{-3\times1} \\ W^{-0\times2} & W^{-1\times2} & W^{-2\times2} & W^{-3\times2} \\ W^{-0\times3} & W^{-1\times3} & W^{-2\times3} & W^{-3\times3} \end{vmatrix} \tag{3.44}$$

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 4 summations to solve this equation. We will come back to this matrix methodology again when we discuss DFT and FFT in Chapter 6.

Summary of Chapter 3

In this chapter, we examined discrete signals, the requirements for sampling set by Shanon and Nyquist, and methods of reconstruction. Discrete signals can experience frequency ambiguity, as many analog frequencies can fit through the same samples. If not sampled at a rate higher then Nyquist 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 developed the FSC for discrete signals and looked at the inverse process.

The terms we introduced in this chapter:

- **Discrete signals** - Defined only at specific uniform time intervals.
- **Digital signals** - A discrete signals the amplitude of which is constrained to certain values. A binary digital signal can only take on two values, 0 or 1.
- **Nyquist rate** - Two times the maximum frequency in the signal.
- **Digital frequency** - Measured in samples per radian.

- **Aliasing** - Given a set of discrete samples, a frequency ambiguity exists, as infinite number of frequencies can pass through these samples. This effect is called aliasing.
- **Discrete time Fourier series coefficients** - DTFSC which repeat are the spectrum of a discrete signal.

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, and its inverse is called the sampling frequency.
2. The sampling frequency of an analog signal should be greater than two times the highest frequency in the signal to accurately represent 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 frequency (or digital frequency) of a discrete 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. We increment digital frequency by itself or $2\pi/N_0$, and use the N_0 sub-frequencies resulting as the basis set. These frequencies are harmonic and distinct.
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 DT Fourier series coefficients, (DTFSC) repeat because the harmonics themselves are identical.
9. In contrast, the CT signal coefficients are aperiodic and do not repeat. This is because all harmonics of a continuous signal are unique.
10. The discrete Fourier series is written as:

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

11. The coefficients of the DTFS are written as:

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

12. The coefficients of a discrete periodic signal are discrete just as they are for the CT signals.

13. The coefficients around the zero frequency are called the principal fundamental alias or principal spectrum.
14. The spectrum of a discrete periodic signal, repeats with sampling frequency, F_s.
15. Sometimes, DTFSC coefficients can be solved using closed form solutions but in a majority of the cases, matrix methods are used to find the coefficients of a signal.
16. Matrix method is easy to set up 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. Given this CT signal, being sampled at $F_s = 25$ Hz, write its discrete form. Is the sampling frequency above the Nyquist rate? $x(t) = \sin(2\pi t) + 0.5\cos(7\pi t)$
2. 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 the Shannon's theorem?
3. Why would a signal be sampled at a rate higher than two times its maximum frequency?
4. A Gaussian signal does not have a clear maximum frequency. What frequency do you choose for sampling such a signal?
5. The bandwidth of a square pulse signal is infinite. How do you choose a sampling frequency for such a signal? What can you do to reduce the bandwidth of this signal?
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. A discrete signal repeats after 37 samples. Is it periodic? A discrete signal repeats with digital frequency of $2\pi/5$, is it periodic?
10. Stock data is noted every 5 s. What is its sampling frequency?
11. Temperature is measured every 10 min during the day and every 15 min during the night. 108 samples are collected over one day. Can we compute Fourier series coefficients of this data?
12. A signal has a period of 8 samples. What is its fundamental digital frequency?
13. Why would we want to recreate a signal from its samples?
14. Are we able to transmit discrete data over an rf(analog) link?
15. We have a sequence of alternating 0's and 1's. What might its DFS coefficients look like?
16. If you have a CT signal $x(t) = \cos(5t)$ and are told to sample it at four times the fundamental frequency, at what rate would you sample this signal?
17. 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?
18. 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, $k =$ -7 to +7.

19. Digital frequency is limited to a range of 0 to 2π. Why?
20. What is the minimum number of harmonics needed to represent this CT signal; $x(t) = \sin(4\pi t + \pi/5)$, with $F_s = 25$.
21. A discrete signal repeats after every 12 samples. What is its digital frequency? What is its fundamental frequency? What is its period?
22. A signal consists of three sinusoids of periods, $N = 7$, 9, and 11 samples. What is the fundamental period of the signal?
23. Why is a signal recreated using sinc reconstruction considered ideal?
24. If $x[n] = 2\cos[(2\pi/5)n]$, what are its DFS coefficients?
25. If $x[n] = 1 - \sin[(2\pi/8)n]$, what are its DFS coefficients?
26. What happens to the spectrum of a train of square pulses as the pulses get narrower?
27. Why does the spectrum of a periodic discrete signal repeat? The repetition occurs over what frequency?

Chapter 4

Continuous-Time Fourier Transform of Aperiodic and Periodic Signals

Harry Nyquist
February 7, 1889 - April 4, 1976

Harry Nyquist, was a Swedish-born American electronic engineer who made important contributions to communication theory. He entered the University of North Dakota in 1912 and received B.S. and M.S. degrees in electrical engineering in 1914 and 1915, respectively. He received a Ph.D. in physics at Yale University in 1917. His early theoretical work on determining the bandwidth requirements for transmitting information laid the foundations for later advances by Claude Shannon, which led to the development of information theory. In particular, Nyquist determined that the number of independent pulses that could be put through a telegraph channel per unit time is limited to twice the bandwidth of the channel. This rule is essentially a dual of what is now known as the Nyquist-Shannon sampling theorem. – From Wikipedia

Applying Fourier Series to Aperiodic Signals

In previous chapters, we discussed the Fourier series as it applies to the representation of continuous and discrete signals. We discussed the concept of harmonic sinusoids as basis functions, first the trigonometric version of sinusoids and then the complex exponentials as a more compact form for representing the basis signals. The analysis signal is "projected" on to these basis signals, and the "quantity" of each basis function is interpreted as spectral content, commonly known as the spectrum.

Fourier series discussions assume that the signal of interest is periodic. However, a majority of signals we encounter in signal processing are not periodic. Many that we think are periodic are not really so. Furthermore, we have many signals that are bunch of random bits with no pretense of *periodicity*. This is the real world of signals and Fourier series comes up short for these types of signals. This was, of course, noticed right away by the contemporaries of Fourier when he first published his ideas is 1822. The Fourier series is great for periodic signals but how about stand-alone nonperiodic, also called *aperiodic* signals like this one?

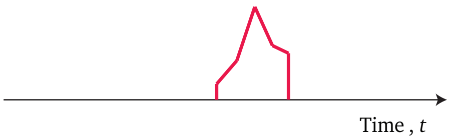

Figure 4.1: Can we compute the Fourier series coefficients of this aperiodic signal?

Taking some liberty with history, Fourier, we are sure, must have been quite disappointed receiving a very unenthusiastic response to his work upon first publishing it. He was denied membership into the French Academy, as the work was not considered rigorous enough. His friends and foes, who are now as famous as he is (Laplace, Lagrange etc.) objected to his overreaching original conclusion about the Fourier series that it can represent *any* signal. They correctly guessed that series representation would not work universally, such as for exponential signals as well as for signals that are not periodic. Baron Fourier, disappointed but not discouraged, came back 20 years later with something even better, the *Fourier transform*. (If you are having a little bit of difficulty understanding all this on first reading, this is forgivable. Even Fourier took 20 years to develop it.)

Extending the period to infinity

In this chapter, we will look at the mathematical trick Fourier used to modify the Fourier series such that it could be applied to signals that are transient or are not strictly periodic. Take the signal in Fig. 4.1. Let us say that this little signal, as shown, has been collected and the data show no periodicity. Being engineers, we want to compute its spectrum using Fourier analysis, even though we have been told that the signal must be periodic. What to do?

Well, we can pretend that the signal in Fig. 4.1 is actually a periodic signal, but we are only seeing one period, the length of which is longer than the length of the data at hand. We surmise that if the length of our signal is 4 s, then maybe the signal looks like the top row of this figure with signal repeating with a period of 5 s. Of course, this is arbitrary. We have no idea what the period of this signal is, or if it even has one.

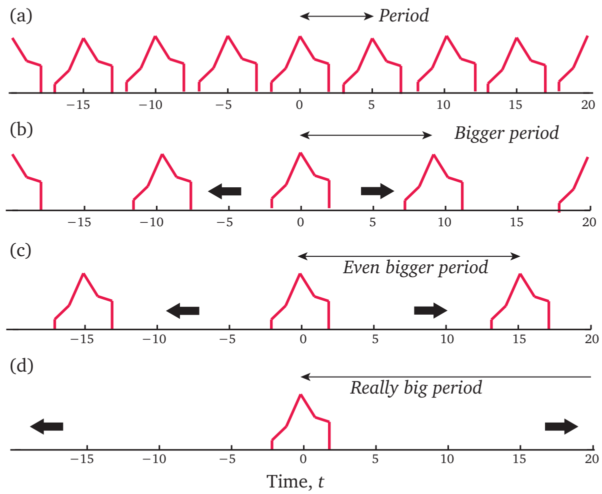

Figure 4.2: Going from periodic to aperiodic signal by extending the period.

Because 5 s is an assumed number anyway, let us just increase it some more by pushing these assumed copies out, increasing the time in-between. We can indeed keep doing this such that the time goes on forever on each side and effectively the period becomes infinitely long. The signal is now just by itself with zeros extending to infinity on each side. We declare that this is now a periodic signal but with a period extending to ∞. We have turned an aperiodic signal into a periodic signal with this assumption.

We can apply the Fourier analysis to this extended signal because it is ostensibly *periodic*. Mathematically, we have let the period T go to infinity so that the assumed periods of the little signal move so far apart that we see neither hide nor hair of them. The single piece of the signal is then *one period* of a periodic signal, the other periods of which we can not see. With this assumption, the signal becomes periodic in a mathematical sense, and its Fourier series coefficients (FSC) can be computed by setting its period to, $T = \infty$. This conceptual trickery is needed because a signal must be periodic for Fourier series representation to be valid.

Continuous-Time Fourier Transform

It was probably this same observation that led Fourier to the Fourier transform. We can indeed apply Fourier series analysis to an aperiodic signal by assuming that the period of an arbitrary aperiodic signal is *very* long and hence we are seeing only one period of the signal. The aperiodic data represents *one* period of a presumed periodic signal, $\tilde{x}(t)$. But if the period is infinitely long, then the *fundamental frequency* defined as the inverse of the period becomes infinitely small. The harmonics are still integer multiples of this infinitely small fundamental frequency but they are so very close to each other that they approach a continuous function of frequency. So a key result of this assumption is that the spectrum of an aperiodic signal becomes a continuous function of the frequency and is no longer discrete as are the FSC.

Figure 4.3(a) shows a pulse train with period T_0. The FSC of the pulse train are plotted next to it (See Ex. 2.10). Note that as the pulses move further apart in Fig. 4.3(b and c), the spectral lines or the harmonics are moving closer together.

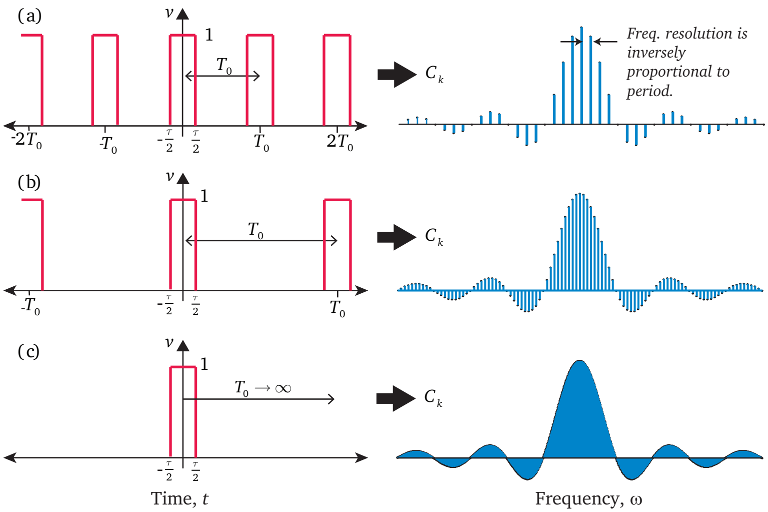

Figure 4.3: Take the pulse train in (a), as we increase its period, i.e., allow more time between the pulses, the fundamental frequency gets smaller, which makes the spectral lines move closer together as in (c). In the limiting case, where the period goes to ∞, the spectrum would become continuous.

We will now go through the math to show how the **Fourier transform** (FT) is directly derived from the FSC. Like much of the math in this book, it is not complicated, only confusing. However, once you have clearly understood the concepts of fundamental frequency, period, and the harmonic frequencies, the rest gets easier.

After we discuss the continuous-time Fourier Transform (CTFT), we will then look at the discrete-time Fourier transform (DTFT) in Chapter 5. Of course, we are far more interested in a yet to be discussed transform, called the discrete Fourier transform (DFT). However, it is much easier to understand DFT if we start with the continuous-time case first. Although you will come across CTFT only in books and school, it is essential for the full understanding of this topic.

In Equation (4.1) the expression for the FSC of a continuous-time signal is repeated from Chapter 2.

$$\boxed{C_k = \frac{1}{T_0}\int_{-T_0/2}^{T_0/2} x(t)e^{-jk\omega_0 t}\, dt} \tag{4.1}$$

To apply this to an aperiodic case, we let T_0 go to ∞. In Eq. (4.1) as the period gets longer, we are faced with division by infinity. Putting the period in form of frequency avoids this problem. Thereafter, we only have to worry about multiplication by zero. We write the period as a function of the frequency.

$$\frac{1}{T_0} = \frac{\omega_0}{2\pi} \tag{4.2}$$

If T_0 is allowed to go to infinity, then ω_0 is becoming tiny. In this case, we write frequency ω_0 as $\Delta\omega$ instead, to show that it is becoming infinitesimally smaller. Now we write the period in the limit as

$$\lim_{T_0\to\infty} \frac{1}{T_0} \approx \frac{\Delta\omega}{2\pi} \tag{4.3}$$

We rewrite Eq. (4.1) by substituting Eq. (4.3).

$$C_k = \frac{\Delta\omega}{2\pi}\int_{-T_0/2}^{T_0/2} x(t)e^{-jk\omega_0 t}\, dt \tag{4.4}$$

But now as T_0 goes to infinity, $\Delta\omega$ approaches zero, and the whole expression goes to zero. To get around this problem, we start with the time-domain Fourier series representation of $x(t)$, as given by

$$\boxed{x(t) = \sum_{k=-\infty}^{\infty} C_k e^{j\omega_0 t}} \tag{4.5}$$

Now substitute Eq. (4.4) into Eq. (4.5) for the value of C_k to write

$$x(t) = \sum_{k=-\infty}^{\infty} \underbrace{\left\{ \frac{\Delta\omega}{2\pi} \int_{-T_0/2}^{T_0/2} x(t) e^{-j\omega t}\, dt \right\}}_{C_k} e^{jk\omega_0 t} \tag{4.6}$$

We now change the limits of the period, T_0, from a finite number to ∞. We also change $\Delta\omega$ to $d\omega$, and $k\omega_0$ to just ω, the continuous frequency and the summation in Eq. (4.6) now becomes an integral. Furthermore, the factor $1/2\pi$ is moved outside. Now, we rewrite Eq. (4.6) incorporating these ideas as:

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \underline{\left\{ \int_{-\infty}^{\infty} x(t) e^{-j\omega t}\, dt \right\}} e^{j\omega t}\, d\omega \tag{4.7}$$

We give the underlined part a special name, calling it the **Fourier transform** and refer to it by the expression, $X(\omega)$. Substituting this nomenclature in Eq. (4.7) for the underlined part, we write it in a new form. This expression is called the **inverse Fourier transform** and is equivalent to the Fourier series representation or the synthesis equation.

$$\boxed{x(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \underline{X(\omega)} e^{j\omega t}\, d\omega} \tag{4.8}$$

The CTFT is defined as the underlined part in Eq. (4.8) and is equal to

$$\boxed{X(\omega) = \int_{-\infty}^{\infty} x(t) e^{-j\omega t}\, dt} \tag{4.9}$$

In referring to the Fourier transform, the following terminology is often used. If $x(t)$ is a time function, then its Fourier transform is written with a capital letter. Such as for time-domain signal, $y(t)$ the CTFT would be written as $Y(\omega)$. These two terms are called a **transform pair** and often written with a bidirectional arrow in between them such as here.

$$\begin{array}{ccccccc} y(t) & \longleftrightarrow & Y(\omega) & \quad & y(t) & \overset{\mathfrak{F}}{\longleftrightarrow} & Y(\omega) \\ c(t) & \longleftrightarrow & C(\omega) & \quad & c(t) & \overset{\mathfrak{F}}{\longleftrightarrow} & C(\omega) \end{array}$$

The symbol $\mathfrak{F}\{\cdot\}$ is also used to denote the Fourier transform. The symbol $\mathfrak{F}^{-1}\{\cdot\}$ is used to denote the inverse transform such that

$$Y(\omega) = \mathfrak{F}\{y(t)\}$$
$$g(t) = \mathfrak{F}^{-1}\{G(\omega)\}$$

The CTFT is generally a complex function. We can plot the real and the imaginary parts of the transform, or we can compute and plot the magnitude, referred to as $|X(\omega)|$ and the phase, referred to as $\angle X(\omega)$. The magnitude is computed by taking the square root of the product $X(\omega)X^*(\omega)$ and phase by the arctan of the ratio of the imaginary and the real parts. We can also write the transform this way, separating out the magnitude and the phase spectrum.

$$X(\omega) = |X(\omega)|e^{j\angle X(\omega)}$$

Here

- Magnitude Spectrum: $|X(\omega)|$
- Phase Spectrum: $\angle X(\omega)$.

If the two components of a CE each have an amplitude 1.0 each, then its magnitude is equal to the square-root of 2. The phase of a CE is constant and equal to 0.

Comparing FSC and the Fourier Transform

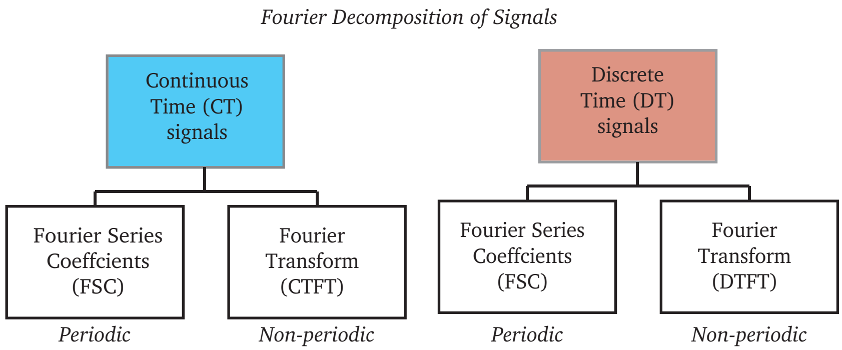

Figure 4.4: Fourier series and the Fourier transform

The Fourier series analysis can be used with both discrete and continuous-time signals as long as they are periodic. When a signal is aperiodic, the premium tool of analysis is the **Fourier transform**. Just as the Fourier series can be applied to continuous and discrete signals, the Fourier transform can also be applied to continuous and discrete signals. The discrete version of the Fourier transform is called the discrete-time Fourier transform, (DTFT) and will be discussed in the next chapter.

Let us compare the CTFT and the FSC equations. Recall that we are trying to determine the amplitudes of each of the harmonics used to represent the signal. The FSC and the CTFT are given as:

$$\begin{aligned} \text{FSC: } C_k &= \frac{1}{T_0}\int_{-T_0/2}^{T_0/2} x(t)e^{-jk\omega_0 t}\,dt \\ \text{CTFT: } X(\omega) &= \int_{-\infty}^{\infty} x(t)e^{-j\omega t}\,dt \end{aligned} \tag{4.10}$$

In the CTFT expression, we note that the time no longer extends over a period, but extends to infinity. That is because the period itself now extends to infinity. We see that the period T_0 in the front of FSC is missing from the latter. Where did it go and does it have any significance? We started development of CTFT by stretching the period and allowing it to go to infinity. We also equated $1/T_0$ to $d\omega/2\pi$ which was then associated with the time-domain formula or the inverse transform (it is not included in the center part of Eq. (4.8), which became the Fourier transform). Therefore, it moved to the inverse transform as the factor 2π.

Notice now the difference between the time-domain signal representation as given by the Fourier series and the Fourier transform.

$$\begin{aligned} \text{FSC: } x(t) &= \sum_{k=-\infty}^{\infty} C_k e^{jk\omega_0 t} \\ \text{CTFT: } x(t) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} X(\omega)e^{j\omega t}\,d\omega \end{aligned} \tag{4.11}$$

We no longer see the harmonic index k in the CTFT equation as compared to the FSC expression. This is because the frequency resolution ω_0 is infinitesimally small for CTFT and essentially the term $k\omega_0$ is continuous. For example, if ω_0 is equal to 0.001, then integer multiple of this number are very close together and hence nearly continuous. The summation of the FSC over the harmonic index k hence, becomes an integral over ω for the CTFT. Hence what is a summation for FSC, is an integral for the CTFT.

In Fourier series representation, to determine the quantity of a particular harmonic, the signal is multiplied by a harmonic, the product integrated over one period and result normalized by the fundamental period T_0. This gives the amplitude of that harmonic. Infact, this is done for all K harmonics, each divided by T_0. In Fourier transform, however, we do not divide by the period because we *don't know what it is*. It is assumed that it is ∞, but we would not want to divide by that either. Therefore, we just ignore it, and hence, we are not determining the signal's *true amplitude*. We are computing a measure of its content but it is not the actual content. Moreover, we are missing the same term from all coefficients, hence, the Fourier transform determines *relative amplitudes*. Very often, we are only interested in the relative levels of harmonic signal powers. The Fourier spectrum gives us the *relative distribution* of power among the various harmonic frequencies in the signal. In practice, we often normalize the maximum power to 0 dB such that the relative levels are consistent among all frequency components.

CTFT of Important Aperiodic Functions

Now, we will take a look at some important aperiodic signals and their transforms, also called transform pairs. In the process, we will use the properties listed in Table 4.1. which can be used to compute the Fourier transform of many functions. These properties do not need proof as they are well known, and we will refer to them as needed for the following important examples. (In most cases, they are easy to prove.) The examples in this section cover some fundamental functions that come up both in workplace DSP as well in textbooks, so they are worth understanding and memorizing. We will use the properties listed in Table 4.1 to compute the CTFTs in the subsequent examples in this and the following chapters. All following examples assume that the signal is aperiodic and is specified in continuous-time. The Fourier transform in these examples is referred to as CTFT.

CTFT of an impulse function

Example 4.1.

$$x(t) = \delta(t) \tag{4.12}$$

This is the most important function in signal processing. The delta function can be considered a continuous (Dirac delta function) or a discrete function (Kronecker delta function), but here we treat it as a continuous function. We use the CTFT equation, Eq. (4.9) and substitute delta

Table 4.1: Important CTFT properties

Zero value	$X(0) = \int_{-\infty}^{\infty} x(t)\, dt$
Duality	If $x(t) \longleftrightarrow X(\omega)$, then $X(t) \longleftrightarrow x(\omega)$
Linearity	$ax(t) + by(t) \longleftrightarrow aX(\omega) + bY(\omega)$
Time Shift	$x(t - t_0) \longleftrightarrow e^{-j\omega t_0} X(\omega)$
Frequency Shift	$e^{j\omega_0 t} x(t) \longleftrightarrow X(\omega - \omega_0)$
Time Reversal	$x(-t) \longleftrightarrow X(-\omega)$
Time expansion or contraction	$x(at) \longleftrightarrow \frac{1}{\lvert a \rvert} X\left(\frac{\omega}{a}\right)$, $a \neq 0$.
Derivative	$\frac{d}{dt} x(t) \longleftrightarrow j\omega X(\omega)$
Convolution in time	$x(t) * h(t) \longleftrightarrow X(\omega) H(\omega)$
Multiplication in Time	$x(t) y(t) \longleftrightarrow X(\omega) * Y(\omega)$
Power Theorem	$\int_{-\infty}^{\infty} \lvert x(t) \rvert^2\, dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} \lvert X(\omega) \rvert^2\, d\omega$

function for function $x(t)$. We compute its CTFT as follows.

$$
\begin{aligned}
X(\omega) &= \int_{-\infty}^{\infty} x(t) e^{-j\omega t}\, dt \\
&= \int_{-\infty}^{\infty} \delta(t) e^{-j\omega t}\, dt \\
&= e^{-j\omega(t=0)} \\
&= 1
\end{aligned}
$$

In the third step, the sifting property of the delta function is used. The sifting property states that the integral of the product of a CT signal with a delta function isolates the value of the signal at the location of the delta function per Eq. (4.13).

$$
\int_{-\infty}^{\infty} \delta(t - a) x(t)\, dt = x(a). \tag{4.13}
$$

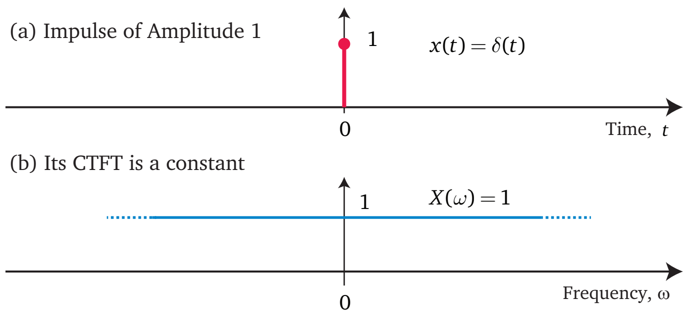

Figure 4.5: CTFT of a delta function located at time 0 is a constant.

If $a = 0$, then the isolated value of the complex exponential is 1.0, at the origin. The integrand becomes a constant, so it is no longer a function of frequency. Hence, CTFT is constant for all frequencies. We get a flat line for the spectrum of the delta function.

The delta function was defined by Dirac as a summation of an infinite number of exponentials.

$$\delta(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{j\omega t}\, d\omega \tag{4.14}$$

The same equation in frequency domain is given by:

$$\delta(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{j\omega t}\, dt \tag{4.15}$$

The general version of the Dirac delta function with a shift for time and frequency is given as:

$$\delta(t - t_0) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{j\omega(t-t_0)}\, d\omega \tag{4.16}$$

$$\delta(\omega - \omega_0) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{j(\omega-\omega_0)t}\, dt \tag{4.17}$$

In the transform of a delta function, we see a spectrum that encompasses whole of the frequency space to infinity, hence, a flat line from $-\infty$ to $+\infty$. Infact when in Chapter 1, Fig. 1.9, we added a whole bunch of sinusoids, this is just what we were trying to get at. This is a very important property to know and understand. It encompasses much depth and if you understand it, the whole of signal processing becomes easier.

Now, if the CTFT or $X(\omega) = 1$, then what is the inverse of this CTFT? We want to find the time-domain function that produced this function in frequency domain. It ought to be a delta function but let us see if we get that. Using the inverse CTFT Eq. (4.8), we write

$$\begin{aligned} x(t) &= \mathfrak{F}^{-1}\{1\} \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} 1 \cdot e^{j\omega t}\, d\omega \\ &= \delta(t) \end{aligned}$$

Substituting in the second step, the definition of the delta function from Eq. (4.14), we get the function back, a perfect round trip. The CTFT of a delta function is 1 in frequency-domain, and the inverse CTFT of 1 in frequency-domain is the delta function in time-domain.

$$\delta(t) \xrightarrow{\text{CTFT}} 1 \xrightarrow{\text{Inverse CTFT}} \delta(t)$$

CTFT of a constant

Example 4.2. What is the Fourier transform of the time-domain signal, $x(t) = 1$.

This case is different from Ex. 4.1. Here, the time-domain signal is a constant and not a delta function. It continues forever in time and is not limited to one single time instant as is the first case of a single delta function at the origin. Using Eq. (4.9), we write the CTFT as:

$$\begin{aligned} X(\omega) &= \int_{-\infty}^{\infty} x(t)e^{-j\omega t}\, dt \\ &= \int_{-\infty}^{\infty} 1 \cdot e^{-j\omega t}\, dt \end{aligned}$$

Using Eq. (4.15) for the expression of the delta function, we get the CTFT of the constant 1 as:

$$X(\omega) = 2\pi\delta(\omega)$$

It can be a little confusing as to why there is a 2π factor, but it is coming from the definition of the delta function, per Eq. (4.15).

If the time-domain signal is a constant, then its Fourier transform is the delta function and if we were to do the inverse transform of $2\pi\delta(\omega)$, we would get back $x(t) = 1$. We can write this pair as:

$$1 \longleftrightarrow 2\pi\delta(\omega)$$

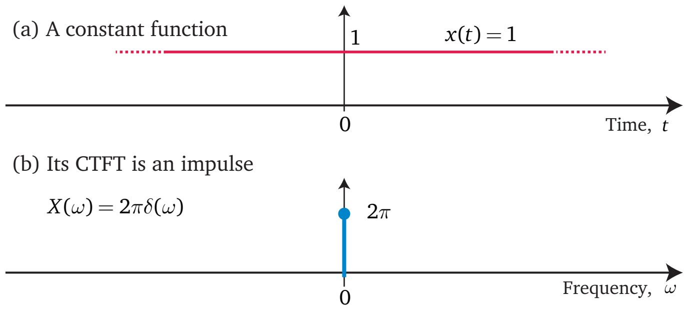

Figure 4.6: CTFT of a constant function that shows reciprocal relationship with Ex. 4.1.

Note in Ex. 4.1, we had this pair $\delta(t) \longleftrightarrow 1$, which is confusingly similar but is not the same thing. Note that in this case the FT is 2π larger than the result from Ex. 1. In this case the signal continues forever so we ought to expect this FT to be larger.

CTFT of a sinusoid

Example 4.3. Since a sinusoid is a periodic function, we will select only one period of it to make it *aperiodic*. Here, we have just a piece of a sinusoid. We make no assumption about what happens outside the selected time frame. The cosine wave shown in Fig. 4.7 has a frequency of 3 Hz, hence, you see one period of the signal lasting 0.33 s.

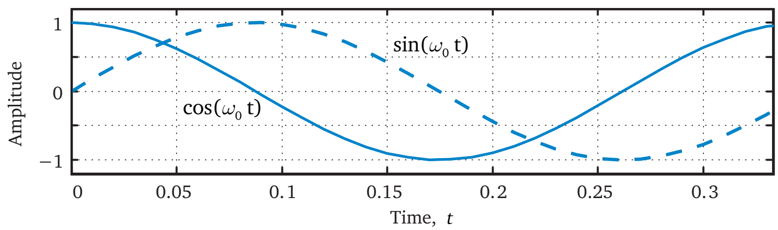

Figure 4.7: A piece of a sinusoid is an aperiodic signal.

We can compute the CTFT of this little piece of cosine as:

$$\begin{aligned} X(\omega) &= \mathfrak{F}\{\cos(\omega_0 t)\} \\ &= \int_{-\infty}^{\infty} \frac{1}{2}\left(e^{j\omega_0 t} + e^{-j\omega_0 t}\right)e^{j\omega t}\, dt \\ &= \int_{-\infty}^{\infty} \frac{1}{2} e^{j(\omega+\omega_0)t}\, dt + \int_{-\infty}^{\infty} \frac{1}{2} e^{j(\omega-\omega_0)t}\, dt \end{aligned}$$

Note that each of these integrals can be represented by a shifted delta function in frequency domain. We use Equation (4.17) to write this result as:

$$X(\omega) = \pi\delta(\omega + \omega_0) + \pi\delta(\omega - \omega_0) \tag{4.18}$$

The only difference between the CTFT of a cosine wave and FSC in Ex. 4.1 is the scaling. In the case of FSC, we get two delta functions of amplitude 1/2 for each or a total of 1. The amplitude of each component in this case is, however, π , or 2π times the amplitude of the FSC.

By similarity, the Fourier transform of a sine is given by

$$X(\omega) = \mathfrak{F}\{\sin(\omega_0 t)\} = j(\pi\delta(\omega + \omega_0) - \pi\delta(\omega - \omega_0)) \tag{4.19}$$

The presence of j in front just means that this transform is in the imaginary plane. It has no effect on the amplitude. We do however, see that the component at $-\omega_0$ frequency has a negative sign as compared to the cosine CTFT in Eq. (4.18) where both components are positive. This will come into play in the next example when we do the CTFT of a complex exponential.

CTFT of a complex exponential

Example 4.4. Now we calculate the CTFT of a very important function, the complex exponential.

$$x(t) = e^{j\omega_0 t} = \cos(\omega_0 t) + j\sin(\omega_0 t)$$

A CE is really two functions, one a cosine of frequency ω_0 and the other a sine of the same frequency, both orthogonal to each other.

We have already calculated the CTFT of a sine and a cosine given by:

$$\mathfrak{F}\{\cos(\omega_0 t)\} = \pi\delta(\omega + \omega_0\} + \pi\delta(\omega - \omega_0) \tag{4.20}$$

$$\mathfrak{F}\{\sin(\omega_0 t)\} = j\pi\delta(\omega + \omega_0) - j\pi\delta(\omega + \omega_0)) \tag{4.21}$$

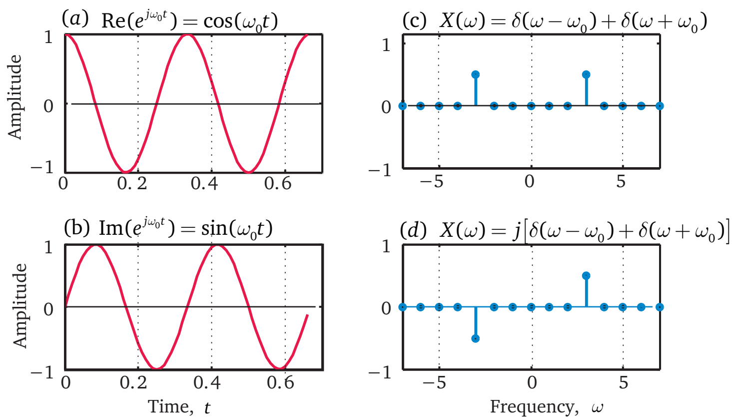

Figure 4.8: The CTFT of a complex exponential.
The Real part is a cosine, hence the spectrum looks like Fig. 4.8(a) and the Imaginary part is a sine, and hence this plot is exactly the same as Fig. 4.8(b).

By the linearity principle, we write the Fourier transform of the CE keeping the sine and cosine separate.

$$
\begin{aligned}
\mathfrak{F}\{e^{j\omega_0 t}\} &= \mathfrak{F}\{\cos(\omega_0 t)\} + j\mathfrak{F}\{\sin(\omega_0 t)\} \\
&= \underbrace{\pi\delta(\omega+\omega_0)+\pi\delta(\omega-\omega_0)}_{\text{cos}} + \underbrace{j^2\big(\pi\delta(\omega+\omega_0)-j^2\pi\delta(\omega-\omega_0)\big)}_{\text{sin}} \\
&= \underbrace{\pi\delta(\omega-\omega_0)-j^2\pi\delta(\omega-\omega_0)}_{\text{add}} + \underbrace{\pi\delta(\omega+\omega_0)+j^2\pi\delta(\omega+\omega_0)}_{\text{cancel}} \\
&= 2\pi\delta(\omega-\omega_0)
\end{aligned}
\tag{4.22}
$$

The result is a single delta function located at ω_0. This happens because of an addition and a cancellation of the components. We see in Fig. 4.9 that the sine component at frequency $+\omega$, because it is multiplied by a j, rotates up or counter clockwise and adds to the cosine component. On the other side, at frequency $-\omega$ the sine component coming out of the paper also rotates counter clockwise, because of its multiplication by j, which puts it directly in opposition to the cosine component and they both cancel. All we are left is a double component at the positive frequency. Hence, we get an asymmetrical result.

Example 4.5. Compute the iCTFT of a single impulse located at frequency ω_1.

$$X(\omega) = \delta(\omega - \omega_1)$$

We want to know what time-domain function produced this spectrum. We take the iCTFT per Eq. (4.8).

$$
\begin{aligned}
x(t) &= \mathfrak{F}\{\delta(\omega-\omega_1)\} \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} \delta(\omega-\omega_1)e^{j\omega t}\, d\omega = \frac{1}{2\pi}e^{j\omega_1 t}
\end{aligned}
$$

The result is a complex exponential of frequency ω_1 in time-domain. Because this is a complex signal, it has a non-symmetrical frequency response that consists of just one impulse located at the CE's frequency. In Fig. 4.9, we see why it is one sided.

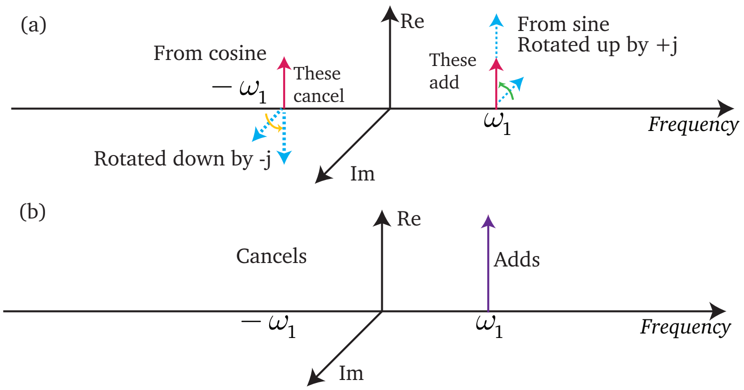

Figure 4.9: The asymmetrical spectrum of a complex exponential.

Here, we write the two important CTFT pairs. The CTFT of a CE is one-sided, an impulse at its frequency. (The CTFT of all complex functions are asymmetrical.)

$$e^{j\omega_0 t} \longleftrightarrow 2\pi\delta(\omega-\omega_0) \tag{4.23}$$

$$e^{-j\omega_0 t} \longleftrightarrow 2\pi\delta(\omega+\omega_0) \tag{4.24}$$

Time-shifting a function

Discrete signals can be constructed as summation of time-shifted delta functions. Hence, we ask, what is the CTFT of a delta function shifted by a time shift, t_0? This case is very important to further understanding of discrete signals.

We can determine the response of a delayed signal by noting the time-shift property in Table 4.1. The property says that if a function is delayed by a time period of t_0, then in frequency domain, the original response of the undelayed signal is multiplied by a CE of frequency

$e^{j\omega t_0}$. In this signal, time is constant and, hence, this is a frequency-domain signal, with frequency being the variable.

We write the shifted signal as $x(t) = \delta(t - t_0)$. Now we calculate the Fourier transform of this function from Eq. (4.9) as:

$$\begin{aligned}
X(\omega) &= \int_{-\infty}^{\infty} x(t)e^{-j\omega t}\, dt \\
&= \int_{-\infty}^{\infty} \delta(t - t_0)e^{-j\omega t}\, dt \\
&= e^{-j\omega t_0}
\end{aligned}$$

In Ex. 4.1, for a undelayed delta function, the CTFT was computed as 1.0. Here the result is the exponential due to the delay, hence proving the time-shift property. The CTFT of a delayed delta function is a CE. This CE has the form $e^{-j\omega t_0}$ and might be confusing. That is because we are not used to seeing exponentials in frequency domain.

This is a really simple case so we may ask, what is the effect of delay on an arbitrary CTFT? Delaying a signal does not change its amplitude (the main parameter by which we characterize signals.) Its frequency also does not change, but what does change is its phase. If a sine wave is running and we arrive to look at it at time t_0 after it has started, we are going to see an instantaneous phase at that time that will be different depending on when we arrive on the scene. That is all a time shift does. It changes the observed phase (or the starting point of the wave) of the signal.

Figure 4.10 shows the effect of time-delay. In Fig. 4.10(a), a signal with an arbitrary spectrum centered at frequency of 2 Hz is shown. The time-domain signal is not shown, only its Fourier transform. You only need to note its shape and center location on the frequency axis. Now we delay this signal by 2 s (we do not know what the signal is, but that does not matter.) and want to see what happens to the spectrum.

In Fig. 4.10(b) we draw the CE $e^{j\omega t_0}$ with $t_0 = 2$ (both sine and cosine are shown). In Fig. 4.10(c), we see the effect of multiplying this CE by the spectrum in Fig. 4.10(a). The magnitude is unchanged. But when we look at Fig. 4.10(d) we see the phase. Since we do not know what the previous phase was, no statement can be made about it yet. Now examine the second column. In this case, the signal is delayed by 4 seconds. Once again in Fig. 4.10(g) we see no change in magnitude but we see that phase in Fig. 4.10(h) has indeed changed from previous case in Fig. 4.10(d). The phase change is directly related to the delay.

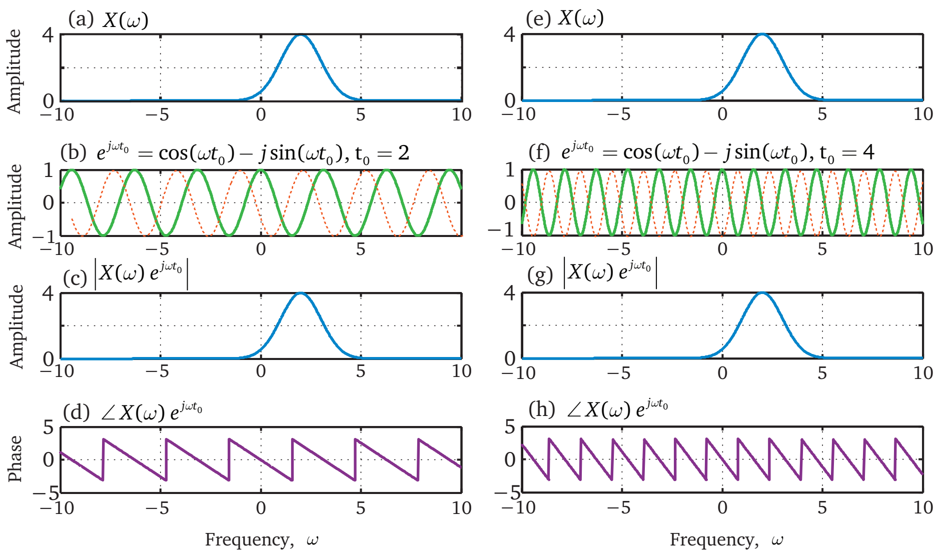

Figure 4.10: Signal delay causes only the phase response to change. In (a) and (e) we see spectrum of an arbitrary signal undelayed, (b) the signal delayed by 2 s, (f) same signal delayed by 4 s. Both cases have the same magnitude as in (c) and (g) but the phase is different as in (d) and (h).

Duality with frequency shift

If a signal is shifted in time, the response changes for phase but not for frequency. Now what if we shift the spectrum by a certain frequency, such as shifting $X(\omega)$ to $X(\omega-\omega_0)$, i.e, the response is to be shifted by a constant frequency shift of ω_0. We can do this by using the frequency shift property. In order to effect this change or frequency-shift, we need to change the time-domain signal as:

$$X(\omega-\omega_0) \longleftrightarrow e^{j\omega_0 t}x(t) \tag{4.25}$$

Hence, if a time-domain signal is multiplied by a CE of a certain frequency, the result is a shifted frequency response by this frequency. We show this here.

$$\begin{aligned}\mathfrak{F}\{e^{j\omega_0 t}x(t)\} &= \int_{-\infty}^{\infty} e^{j\omega_0 t}x(t)e^{-j\omega t}\,dt \\ &= \int_{-\infty}^{\infty} x(t)e^{-j(\omega-\omega_0)t}\,dt \\ &= X(\omega-\omega_0).\end{aligned} \tag{4.26}$$

The frequency shift property shown here is also called the **modulation** property. Modulation, also called upconversion, can be thought of as multiplying, in time-domain, a signal by another single-frequency signal (called a *carrier*) and infact if you look at Eq. (4.25), that is exactly what we are doing. A time-domain signal multiplied by a CE, $e^{j\omega_0 t}$ results in the signal transferring to the frequency of the CE, ω_0 without change in its amplitude.

Convolution property

The most important result from the Fourier transform is the convolution property. Infact the Fourier transform is often used to perform convolution in hardware instead of doing convolution in time-domain. The convolution property is given by

$$x(t) * h(t) = \int_{-\infty}^{\infty} x(\tau)h(t-\tau)\, d\tau \tag{4.27}$$

In time-domain, convolution is a resource-heavy computation. However, convolution can be done using less computational effort using the Fourier transform. The convolution property states:

$$x(t) * h(t) \longleftrightarrow X(\omega)H(\omega) \tag{4.28}$$

This states that the convolution of two signals can be computed by multiplying their individual Fourier transforms and then taking the inverse transform of the product. Let us see why this is possible. We write the time-domain expression for the convolution and then take its Fourier transform. Yes, it does look messy and requires fancy calculus.

$$\begin{aligned}\mathfrak{F}\{x(t) * h(t)\} &= \mathfrak{F}\left\{\int_{-\infty}^{\infty} x(\tau)h(t-\tau)\, d\tau\right\} \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x(\tau)h(t-\tau) d\tau e^{-j\omega t}\, dt\end{aligned}$$

Now, we interchange the order of integration to get this from

$$\mathfrak{F}\{x(t) * h(t)\} = \int_{-\infty}^{\infty} x(\tau)\left(\int_{-\infty}^{\infty} h(t-\tau)e^{-j\omega t}\, dt\right) d\tau$$

We make a variable change by setting $u = t - \tau$, hence we get

$$\mathfrak{F}\{x(t) * h(t)\} = \int_{-\infty}^{\infty} x(\tau)\left(\int_{-\infty}^{\infty} h(u)e^{-j\omega(u+\tau)} du\right) d\tau$$

This can be written as:

$$\mathfrak{F}\{x(t)*h(t)\} = \int_{-\infty}^{\infty} x(\tau)\left(\int_{-\infty}^{\infty} h(u)e^{-j\omega u}e^{-j\omega\tau}du\right)d\tau$$

Now we move the $e^{-j\omega\tau}$ term out of the inner integral because, it is not function of u, to get the desired result and complete the convolution property proof.

$$\begin{aligned}\mathfrak{F}\{x(t)*h(t)\} &= \left(\int_{-\infty}^{\infty} x(\tau)e^{-j\omega\tau}\,d\tau\right)\left(\int_{-\infty}^{\infty} h(u)e^{-j\omega u}du\right)\\ &= X(\omega)H(\omega)\end{aligned}$$

The duality property of the Fourier transform then implies that if we multiply two signals in time-domain, then the Fourier transform of their product would be equal to convolution of the two transforms.

$$x(t)h(t) \longleftrightarrow \frac{1}{2\pi}X(\omega)*H(\omega) \tag{4.29}$$

This is an efficient way to compute convolution rather than the standard way we learn by reversing and multiplying the signals. Convolution is hard to visualize. The one way to think of it is as smearing or a smoothing process. The convolution process produces the smoothed version of both of the signals as we can see in Fig. 4.11. Both of the pulses in Fig. 4.11(a and d) have been smoothed out by their convolution by the center pulse. They have also spread in time as in Fig. 4.11(c and f).

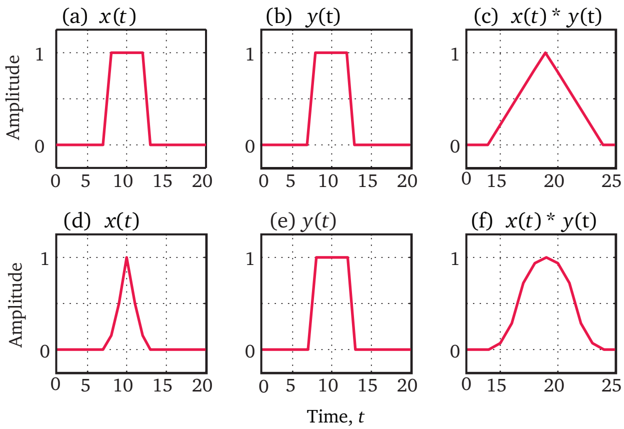

Figure 4.11: (a) The convolution of signals $x(t)$ and $y(t)$ in (c) is done using Fourier transform. In each case, the result is smoother than either of the original signals. Hence, convolution can be thought of as a filter.

CTFT of a Gaussian function

Example 4.6. Now we examine the CTFT of a really unique and useful function, the Gaussian. The zero-mean Gaussian function is given by

$$x(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-t^2/(2\sigma^2)} \tag{4.30}$$

where σ^2 is the signal variance and σ the standard deviation of the signal. This form of the Gaussian is particular to signal processing. A general Gaussian signal is often written in this form.

$$f(t) = ae^{-(t-b)^2/2c^2} \tag{4.31}$$

Here a is the peak height of the familiar bell curve, b is the center point of the curve and c is the standard deviation.

The CTFT of the Gaussian function is very similar to the function itself.

$$\begin{aligned} X(\omega) &= \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-t^2/(2\sigma^2)} e^{-j\omega t}\, dt \\ &= \frac{1}{2\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-t^2/2\sigma^2} e^{-j\omega t}\, dt \end{aligned}$$

This is a difficult integral to solve but fortunately smart people have already done it for us. The result is

$$X(\omega) = \frac{1}{2\sigma\sqrt{2\pi}} e^{-\frac{\sigma^2\omega^2}{2}} \tag{4.32}$$

As σ is a constant, the shape of this curve is a function of the square of the frequency, same as it is in time-domain where it is square of time. Hence, it is often said that the CTFT of the Gaussian function is same as itself, but what they really mean is that the shape is the same. This property of the Gaussian function is very important in nearly all fields. However, as we see in Eq. (4.32), there is no hard cutoff in the response and hence the bandwidth of this signal is not finite. How fast it decays depends on the parameter, σ, the standard deviation. If the x-axis is normalized by the standard deviation, the response becomes the normal distribution. In signal processing, we find that often noise can be modeled as a Gaussian process. We also find that when Gaussian signals are added or convolved, their joint distribution retains its Gaussian distribution.

CTFT of a square pulse

Example 4.7. Now, we examine the CTFT of a square pulse of amplitude 1, with a period of τ, centered at time zero. This case is different from the ones in Chapters 2 and 3 in that here we have just a single solitary pulse. This is not a case of repeating square pulses as in this section we are considering only aperiodic signals.

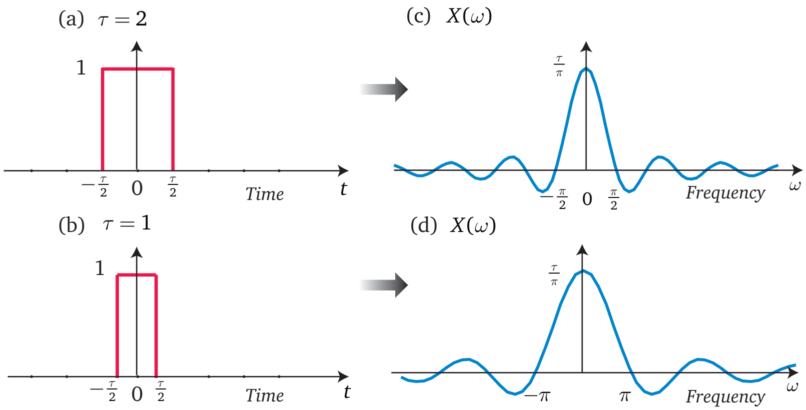

Figure 4.12: Spectrum along the frequency line. A square pulse has a sinc-shaped spectrum. (a) its time-domain shape, and (b) its CTFT.

We write the CTFT as given by Eq. (4.9). The function has an amplitude of 1.0 for the duration of the pulse.

$$
\begin{aligned}
X(\omega) &= \int_{-\infty}^{\infty} x(t)e^{-j\omega t}\, dt \\
&= \int_{-\tau/2}^{\tau/2} 1 \cdot e^{-j\omega t}\, dt \\
&= -\frac{e^{-j\omega t}}{j\omega}\Bigg|_{-\tau/2}^{\tau/2} \\
&= -\frac{1}{j\omega}\left(e^{-j\omega\tau/2} - e^{j\omega\tau/2}\right) \\
&= \frac{2}{\omega}\sin\left(\frac{\omega\tau}{2}\right)
\end{aligned}
$$

This can be simplified to

$$
X(\omega) = \tau \operatorname{sinc}\left(\frac{\omega\tau}{2\pi}\right) \tag{4.33}
$$

The spectrum is shown on the right side in Fig. 4.12 for $t = 1s$ and $2s$. Note that as the pulse gets wider, the spectrum gets narrower. As the sinc function is zero for all values that are integer multiple of 2π, the zero crossings occur whenever $\omega\tau = k\pi$, where k is an even integer larger than 2. For $\tau = 2$, the zeros would occur at radial frequency equal to $\pi, 2\pi, \ldots$. If the pulse were to become infinitely wide, the CTFT would become an impulse function. If they were infinitely narrow as in Ex. 4.1, the frequency spectrum would be flat.

Now assume that instead of the time-domain square pulse shown in Fig. 4.12, we are given a frequency response that looks like a square pulse. The spectrum is flat from $-W$ to $+W$ Hz. This can be imagined as the frequency response of an ideal filter. Notice, that in the time-pulse case, we defined the half-width of the pulse as $\tau/2$, but here we define the half bandwidth by W and not by $W/2$. The reason is that in time-domain, when a pulse is moved, its period is still τ. However, bandwidth is designated as a positive quantity only. There is no such thing as a negative bandwidth. In this case, the bandwidth of the signal (because it is centered at 0 is said to be W Hz and not $2W$ Hz. However, if this signal was moved to a higher frequency such that the whole signal was in the positive frequency range, it would be said to have a bandwidth of $2W$ Hz. This crazy definition gives rise to the concepts of low-pass and band-pass bandwidths. Low-pass is defined as being centered at the origin so it has half the bandwidth of band-pass.

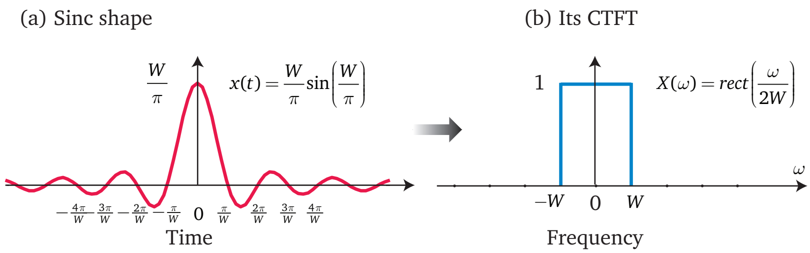

Figure 4.13: Time-domain signal corresponding to the rectangular frequency response.

Example 4.8. What time-domain signal produces a rectangular frequency response shown in Fig. 4.13(b)? The frequency response is limited to a certain bandwidth, W Hz.

$$X(\omega) = \begin{cases} 1 & |\omega| \leq W \\ 0 & |\omega| > W \end{cases}$$

We compute the time-domain signal by the inverse CTFT equation.

$$\begin{aligned} x(t) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} X(\omega)e^{j\omega t}\, d\omega \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} 1\cdot e^{j\omega t}\, d\omega \\ &= \frac{1}{2\pi}\frac{e^{j\omega t}}{jt}\bigg|_{-W}^{W} \end{aligned}$$

Which can be simplified to

$$x(t) = \frac{W}{\pi}\mathrm{sinc}\left(\frac{Wt}{\pi}\right) \tag{4.34}$$

Again we get a sinc function, but now in time-domain. This is the duality principle at work. This is a very interesting case and of fundamental importance in communications.

The frequency spectrum shown in Fig. 4.13(b) is, of course, a very desirable frequency response. We want the frequency response to be tightly constrained. The way to get this type of spectrum is to use a time-domain signal that has a sinc pulses. But a sinc function looks strange as a time-domain signal because it is of infinite length. However, because it is "well-behaved," which means it crosses zeros at predicable points, we can use it as a signal shape, at least in theory. In practice, it is impossible to build a signal shape of infinite time duration. It has to be truncated, however, truncation causes distortion and we do not get the perfect brick-wall frequency response. An alternate shape with similar properties is the **raised cosine**, most commonly used signal shape in communications. The raised cosine shape is also truncated to shorten its length but its distortion is manageable because it decays faster than a sinc shape.

In Fig. 4.14 some important Fourier transforms of aperiodic signals are given. A good engineer should know all of these by heart.

Fourier Transform of Periodic Signals

The Fourier transform came about so that the Fourier series could be made rigorously applicable to aperiodic signals. The signals we examined so far in this chapter are all *aperiodic*, even the cosine wave, which we limited to one period. Can we use the CTFT for a *periodic* signal? Our intuition says that this should be the same as the Fourier series. Let us see if that is the case.

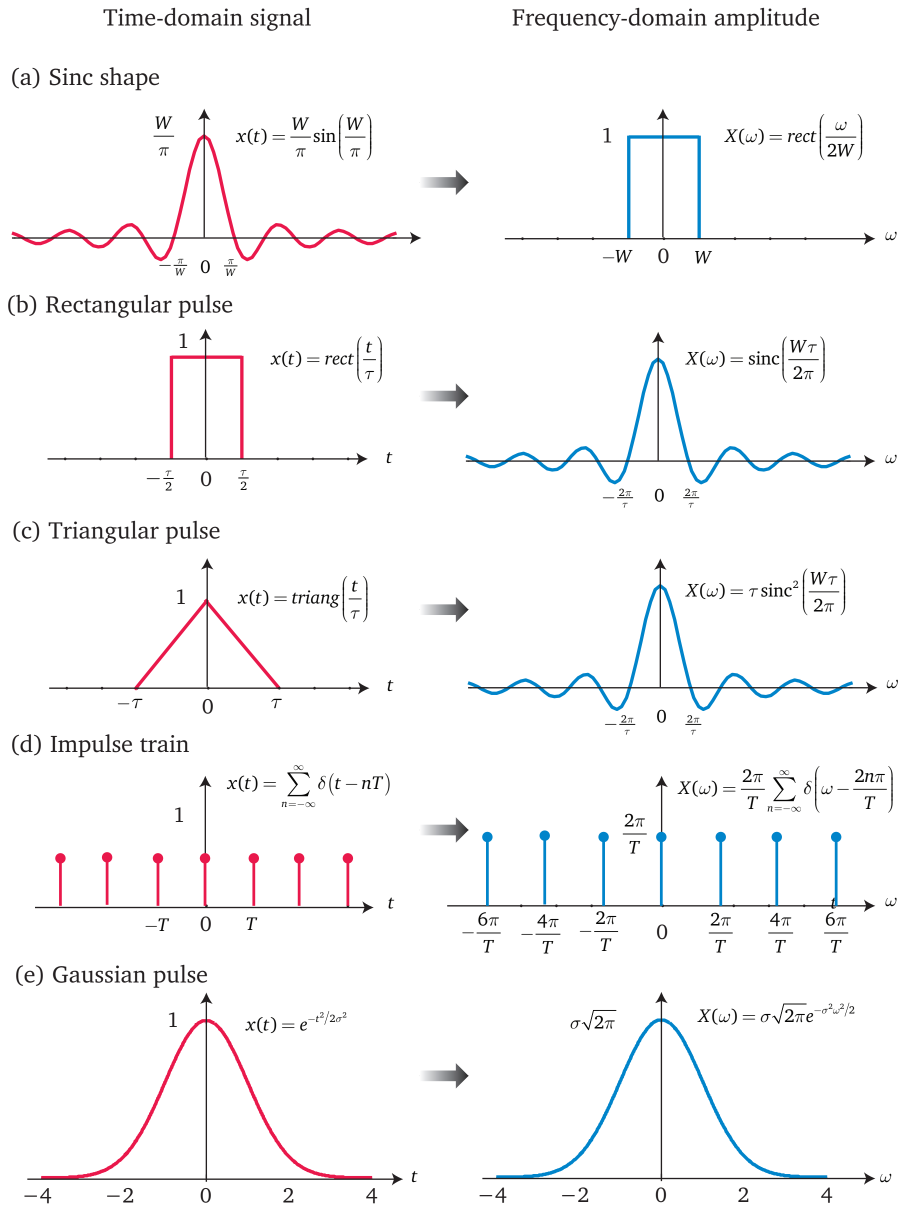

Figure 4.14: Response to a Sinc shaped time domain signal.

Take a periodic signal $x(t)$ with fundamental frequency of $\omega_0 = 2\pi/T_0$ and write its FS representation.

$$x(t) = \sum_{k=-\infty}^{\infty} C_k e^{j\omega_0 t}$$

Taking the CTFT of both sides of this equation, we get

$$X(\omega) = \mathfrak{F}\{x(t)\} = \mathfrak{F}\left\{ \sum_{k=-\infty}^{\infty} C_k e^{j\omega_0 t} \right\}$$

We can move the coefficients out of the CTFT because they are not function of frequency. They are just numbers.

$$X(\omega) = \sum_{k=-\infty}^{\infty} C_k \ \mathfrak{F}\{\underline{e^{j\omega_0 t}}\}$$

The Fourier transform of the complex exponential $e^{j\omega_0 t}$ is a delta function located at the frequency ω_0 as in Ex. 4.4. Making the substitution, we get

$$\boxed{X(\omega) = 2\pi \sum_{k=-\infty}^{\infty} C_k \delta(\omega - k\omega_0)} \tag{4.35}$$

This equation says that the CTFT of a periodic signal is a *sampled* version of the FSC. The FSC are being sampled at frequency of the signal, ω_0, with k the index of repetition. However, the FSC are already discrete! Thus, the only thing the Fourier transform does is change the scale. The magnitude of the CTFT of a periodic signal is 2π times bigger than that computed with FSC as seen by the factor 2π front of Eq. (4.35).

Important observation: *The CTFT of an aperiodic signal is aperiodic and continuous whereas the CTFT of a periodic signal is aperiodic but discrete.*

CTFT of a periodic square pulse train

Example 4.9. Now we examine the CTFT of the periodic square pulse. For the Fourier transform of this periodic signal, we will use Eq. (4.35)

$$X(\omega) = 2\pi \sum_{k=-\infty}^{\infty} C_k \delta(\omega - k\omega_0)$$

The FSC of a periodic pulse train with duty cycle = 1/2 are computed in Chapter 2 and given as

$$C_k = \frac{1}{2}\text{sinc}(k\pi/2)$$

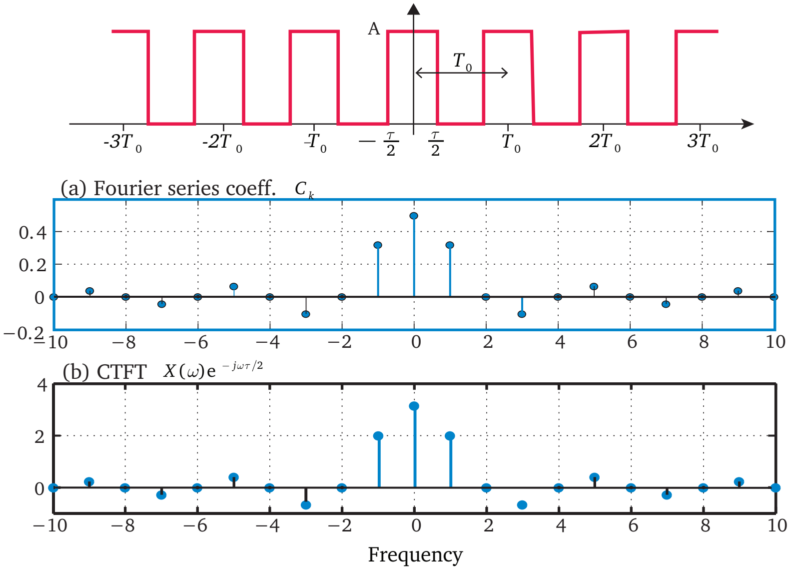

Figure 4.15: The periodic square wave with duty cycle of 0.5. (b) Its FSC and (c) its CTFT. Only the scale is different.

We plot these FSC in Fig. 4.15(b). To compute CTFT, we set $\omega_0 = 1$ and now we write the CTFT expression as:

$$X(\omega) = 2\pi \sum_{k=-\infty}^{\infty} C_k \delta\big(\omega - k(\omega_0 = 1)\big)$$

The result is the sampled version of the FSC scaled by 2π (See Ex. 2.10) which are of course themselves discrete.

What if the square pulse was not centered at 0 but shifted some amount. We can compute the CTFT of this periodic function by applying the time-shift property to the CTFT of the unshifted square wave.

This periodic function is same as Fig. 4.15 but is time-shifted. We can write it as:

$$y(t) = x(t - \tau/2)$$

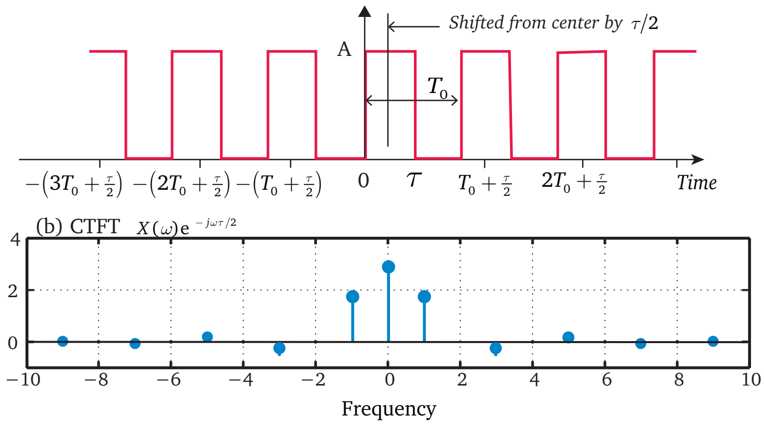

Figure 4.16: A time-shifted square pulse train.

By the time-shift property, we can write the CTFT of this signal by multiplying the CTFT of the unshifted case by $e^{j\omega\tau/2}$. Hence

$$Y(\omega) = X(\omega)e^{j\omega\tau/2}$$

which is

$$Y(\omega) = \left(2\pi \sum_{k=-\infty}^{\infty} C_k \delta(\omega - k)\right) e^{j\omega\tau/2}$$

This time shift has no effect on the shape of the response at all, just as we would expect. Only the phase gets effected by the time shift.

The main reason we do a Fourier transform rather than the Fourier series representation and its coefficients, is that the Fourier transform can be used for *aperiodic* and *periodic* signals. There is, however, a key difference between the FSC and the CTFT. We calculate actual amplitudes with FSC. However, in developing the Fourier transform, we dropped the concept of a period, hence, the results are useful in a relative sense only. The Fourier transform is not a tool for measuring the real signal amplitudes but is mostly used as a qualitative tool for assessing relative amplitudes, power and issues of bandwidth occupation.

Summary of Chapter 4

In this Chapter we looked at aperiodic signals and their frequency representations. The FS concept is extended so that Fourier analysis can be applied to aperiodic signals. In a manner similar to computing the coefficients, we call the process of computing the coefficients of aperiodic signal the Fourier transform. The spectrum of continuous signals using the Fourier transform is continuous, where the Fourier transform of a periodic signal is discrete.

Terms used in this chapter:

- **Fourier Transform, FT**
- **Continuous-time Fourier Transform, CTFT**
- **Discrete-time Fourier Transform, DTFT**
- **Transform pair** - The signal in one domain and its Fourier transform in the other domain are called a Fourier transform pair.

1. Aperiodic signals do not have mathematically valid FSC.
2. Fourier transform is developed by assuming that a aperiodic signal is actually periodic but with an infinitely long period.
3. Whereas the spectrum of a periodic signal is represented by the Fourier series coefficients, the spectrum of an aperiodic signal is called the Fourier transform.
4. The term Fourier transform applies not just to aperiodic signals but to periodic signals as well.
5. The FSC are discrete whereas the CTFT of an aperiodic signal is continuous in frequency.
6. The CTFT of aperiodic signals is aperiodic.
7. The CTFT and the iCTFT are computed by:

$$\text{CTFT } X(\omega) = \int_{-\infty}^{\infty} x(t)e^{-j\omega t}\, dt$$

$$\text{iCTFT } x(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} X(\omega)e^{j\omega t}\, d\omega$$

8. A function and its Fourier transform are called a transform pair.

$$x(t) \longleftrightarrow X(\omega)$$

9. The CTFT of a periodic signal is given by the expression

$$X(\omega) = 2\pi \sum_{k=-\infty}^{\infty} C_k \delta(\omega - k\omega_0)$$

10. The CTFT of a periodic signal can be considered a sampled version of the FSC.
11. Unlike the CTFT of an aperiodic signal, the CTFT of a periodic signal is discrete, just as are the FSC for a periodic signals.

Questions

1. What is the conceptual difference between the Fourier series and the Fourier transform?
2. Why is the CTFT continuous? Why are the CTFSC coefficients discrete?
3. What is the CTFT magnitude of these impulse functions:
   $\delta(t-1)$, $\delta(t-2)$, $\delta(t-T)$.
4. Give the expression for the CTFT of a cosine and a sine. What is the difference between the two?
5. Given a sinusoid of frequency 5 Hz. What does its CTFT look like?
6. What is the difference between the Fourier transform magnitude of a sine and a cosine of equal amplitudes?
7. What is the CTFT (amplitude) of these sinusoids: $\sin(-800\pi t)$, $-\cos(250\pi t)$, $0.25\sin(25\pi t)$. What is the magnitude spectrum of these sinusoids?
8. What is the value of $\sin(500\pi t)\delta(t)$, $\cos(100\pi t)\delta(t-\pi)$?
9. What is the CTFT of: $x(t) = 6\sin(10\pi t) - 4\cos(4\pi t)$?
10. If the FT of a signal is being multiplied by this CE: $e^{-j6\pi f}$, what is the resultant effect in time domain?
11. We multiply a signal in time domain by this CE: $e^{-j12\pi t}$, what is the effect in frequency domain on the FT of the signal?
12. The summation of complex exponentials represents what function?
13. What is the value of $\cos(6\pi t) * \delta(t-4)$?
14. What is the CTFT of the constant π?
15. A sinc function crosses first zero at π/B. What is its time domain equation? What does the spectrum look like and what is its bandwidth?
16. A sinc function crosses first zero at $t = 1$, give its time domain equation? What does the spectrum look like and what is its bandwidth?
17. What is the CTFT of $\sin(5\pi t) * \delta(t-5)$?

18. A signal of frequency 4 Hz is delayed by 10 s. By what CE do you multiply the unshifted CTFT to get the CTFT of the shifted signal?
19. Given $x(t) = \text{sinc}(t\pi)$, at what times does this function cross zeros?
20. The first zero-crossing of a sinc function occurs at time = B s; 0.5 s; 2 s. What is the bandwidth of each of these three cases?
21. What is the main lobe width of the CTFT of square pulses of widths: T_s, $T_s/2$, $\pi/2$, and 3 s.
22. If the main lobe width of a sinc function (one sided) is equal to $\pi/2$, then how wide is the square pulse in time?
23. What is the CTFT of an impulse train with period equal to 0.5 s. Is this a periodic signal?
24. Convolution in time domain of two sequences represents what in frequency domain?
25. What is the Fourier transform of an impulse of amplitude 2 V in time domain?
26. If a signal is shifted by 2 s, what happens to its CTFT?
27. The CTFT of a periodic signal is continuous while the CTFSC is discrete. True or false?
28. If the CTFSC of a signal at a particular harmonic is equal to 1/2, then what is the value obtained via CTFT at the same harmonic?
29. Given the FSC of a signal, how do you calculate the FT of this signal?
30. What is modulation and how is it accomplished?
31. Why do the base-band and pass-band bandwidths of a signal differ?

Chapter 5

Discrete-Time Fourier Transform of Aperiodic and Periodic Signals

Paul Adrien Maurice Dirac
1902 – 1984

Paul Adrien Maurice Dirac was an English theoretical physicist who made fundamental contributions to the early development of both quantum mechanics and quantum electrodynamics. He was the Professor of Mathematics at the University of Cambridge, and spent the last decade of his life at Florida State University. Among his discoveries, he formulated the Dirac equation, which describes the behavior of fermions and predicted the existence of antimatter. Dirac shared the Nobel Prize in Physics with Erwin Schrödinger. He also did work that forms the basis of modern attempts to reconcile general relativity with quantum mechanics. Paul Dirac in his influential 1930 book The Principles of Quantum Mechanics, introduced the "delta function" which he used as a continuous analogue of the discrete Kronecker delta. – From Wikipedia