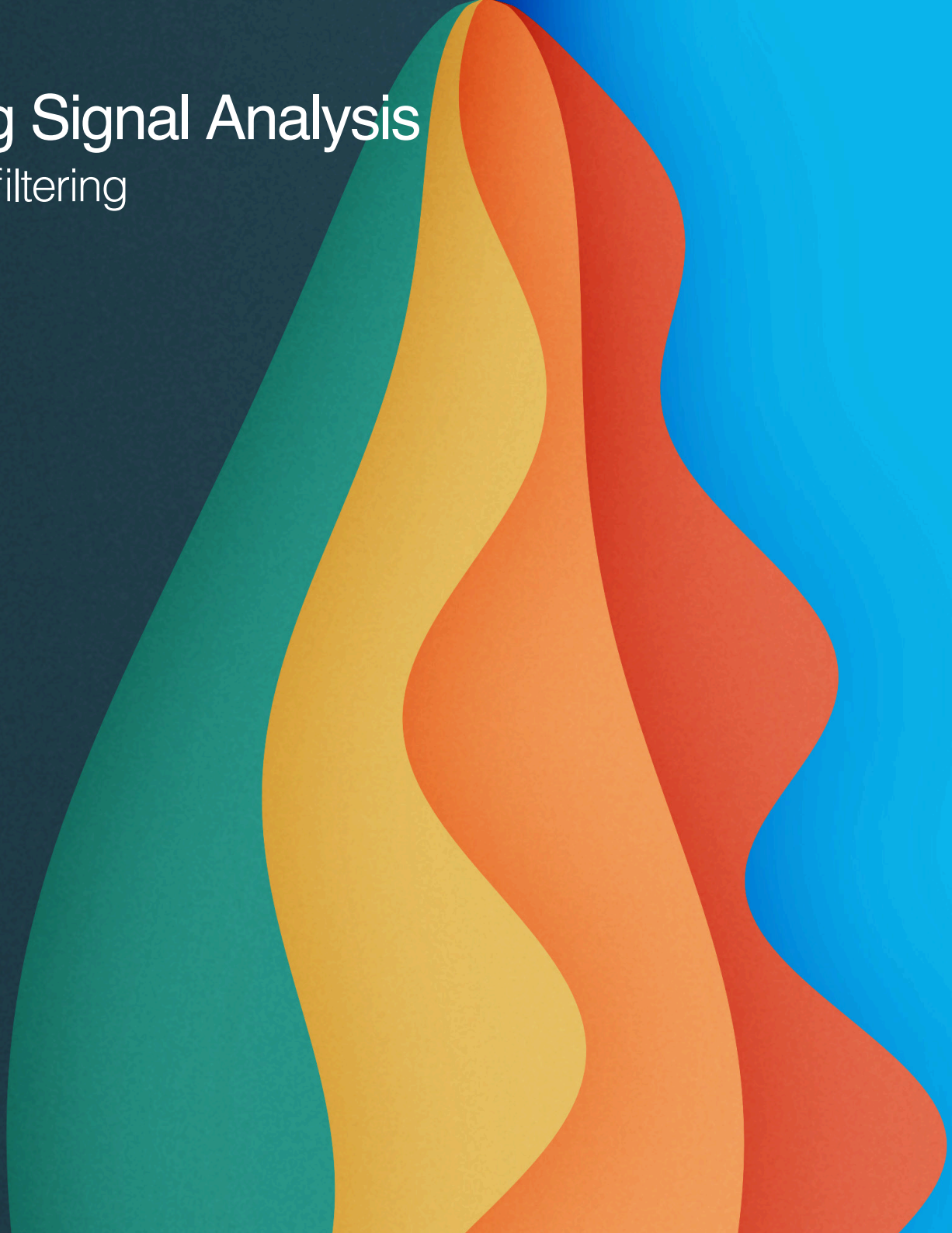
Engineering Signal Analysis

from Fourier to filtering

Christian Tiberius
Max Mulder



Theory

TU Delft OPEN Publishing

Engineering Signal Analysis

– from Fourier to filtering –

Theory

Christian Tiberius & Max Mulder

Cover design by **CYANETICA** (cyanetica.com), Sebastiaan de Stigter

Engineering Signal Analysis – Theory
Christian Tiberius, Max Mulder
April 2026, first edition, version 1.1.

Publisher:
TU Delft OPEN Publishing TU Delft Open Textbooks
Delft University of Technology – The Netherlands

keywords:
Fourier series, Fourier transform, sampling, Discrete Fourier Transform, spectral estimation, linear systems, filtering

ISBN (softback/paperback): 978-94-6384-901-2
ISBN (e-book): 978-94-6518-235-3
DOI: https://doi.org/10.59490/mt.247

The text has been typeset using the MikTex 2.9 implementation of LaTeX. Graphs have been created in Python, and drawings and diagrams with Inkscape.

Contents

Preface

This textbook provides an introduction to the subject of signal analysis in the frequency domain. Although many excellent textbooks already exist on this topic, there were a number of reasons that motivated us to write our own book. We aim to provide an introductory volume: concise, with a clear engineering-oriented storyline, and highlighting the practical use of the theory. With 18 chapters, well under 200 pages, and written in an informal manner, the book is explicitly meant as an introduction. The content is aimed to be taught at the bachelor (undergraduate) level of an academic engineering curriculum. After studying this book, the student should have an understanding of the basic concepts in signal analysis and be able to apply them to actual signals.

While providing a theoretical framework, the book's main focus is on explaining how to perform signal analysis in practice. It focuses on analyzing continuous-time signals, as in many engineering applications signals are indeed, in physical reality, continuous-time. Since much of the measuring, processing and analyzing of signals takes place using a computer, in discrete time, a significant part of the book explains the effects of sampling and working with discrete-time signal representations of the continuous-time phenomena.

The book has five parts. In Part I, the subject of signal analysis and the reasons to study signals in the frequency domain are discussed, including a chapter with (mathematical) preliminaries and definitions relevant to signal analysis. In Part II the transformations of continuous-time signals to the frequency domain are introduced: the Fourier series and Fourier transform. The principle of signal convolution is explained, as well as effects of finite signal duration – inevitable when measuring signals in practice, causing spectral leakage. In Part III signal sampling and reconstruction are explained, the Discrete-Time Fourier Transform is derived and the most common signal transformation applied in practice, the Discrete Fourier Transform, is introduced. In Parts IV and V, respectively, spectral estimation and linear systems are introduced. A number of appendices contain supplementary material, to review prior knowledge, offer further insight, and provide guidelines for practical application.

Although this book, **Theory**, already includes many worked-out examples, it is best used together with its companion book, **Exercises**. This second volume contains more exercises, more (practical) examples and computer-based problems with Python.

The authors welcome corrections and suggestions for improvement, as well as feedback in general. Please e-mail ESAbook-TUD@tudelft.nl for this purpose.

Acknowledgments

First and foremost, our students deserve our gratitude. Teaching this material has been a tremendous inspiration. Working with our students, the interaction with young minds, the feedback received, and their many clever questions have made us better teachers, and hopefully has led to a book that will inspire future generations.

Christian would like to acknowledge professor Roland Klees for sharing valuable insights during several years of co-teaching classes on the subject of signal analysis. Many colleagues of the faculty of Civil Engineering and Geosciences contributed indirectly, in particular from the departments of Engineering Structures, Hydraulic Engineering, Water Management, Geoscience and Engineering, and Geoscience and Remote Sensing.

Max would like to thank his teachers, mentors and colleagues from the faculties of Aerospace and Mechanical Engineering. Emeritus professors Bob Mulder and Henk Stassen for their in-

spiring lectures, academic guidance and mentorship, which motivated him to spend his life working on topics related to signals, systems and control. Ton van Lunteren, Hans van der Vaart, Jan-Willem van Staveren, Qi Ping Chu and René van Paassen for expanding his knowledge and insight and fostering his curiosity.

The authors would like to thank Eric Verschuur and Rowenna Wijlens for reviewing major parts of this book and providing valuable feedback. All remaining errors are our own.

Thank you, Sebastiaan de Stigter (cyanetica.com), for designing such an inspiring book cover.

Thank you, Saskia Roselaar (Roselaar Tekstadvies), for your meticulous proofreading.

Finally, we have appreciated all support from the TU Delft OPEN Publishing team: thank you to Jacqueline Michielen-van de Riet, Michiel Munnik and Kees Moerman.

Christian Tiberius, Max Mulder
Delft, January 2026

Note to version 1.1, April 2026

Several typos were corrected, minor inconsistencies have been resolved.

I

Introduction

1

Introduction

Signals are everywhere around us. Signals, and the associated systems which operate on these signals, are of paramount importance in science, engineering and society, see Figure 1.1. In engineering and technology, examples are countless, ranging from smartphones, communication, radar and sonar, noise and acoustics, monitoring and maintenance, remote sensing, media and entertainment, robotics, vehicles on Earth, in the air and in space, to the transition to clean energy. Also in the biological, neuroscience, medical and sociological domains, the capability to dissect the matter at hand (cells, brain, body, groups of people) into smaller components, sub-systems, and to identify and measure the signals with which these components interact in various command, control and communication hierarchies in order to achieve a particular goal, is crucial. The 'signals and systems perspective' is perhaps the most versatile model of the world that exists in science. This book provides a brief introduction to this broad topic, focusing in particular on the analysis of signals in engineering applications.

Figure 1.1: Applications of signals and systems are everywhere. All images from [1], public domain or open license; titles, attributions and licenses are summarized on page 232.

1.1 Signals and systems

A generic signals and systems view on the world is illustrated in Figure 1.2 in its most basic form. In engineering, a system is often considered as some device or algorithm which performs an operation on a signal, the input signal $x(t)$, to create an output signal, $y(t)$. Systems can be connected to create larger systems, with signals passing to and from the system components.

Systems can exist in physical reality (mechanical, electrical or computerized, and combinations of these) or as just a script running on a computer. In most cases, signals are functions of time, either in continuous time, $x(t)$ or, when computers or other discrete-time devices are involved, in discrete time, x_n. While systems often have multiple inputs and outputs, this book limits the discussion to single-input single-output systems, i.e., SISO systems.

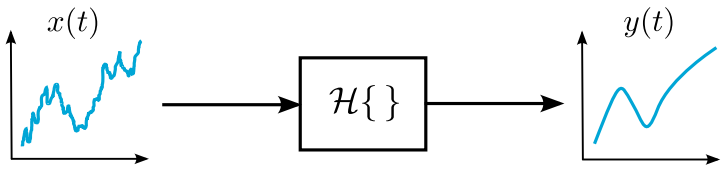


Figure 1.2: A system turns input signal $x(t)$ into output signal $y(t)$.

The operation $\mathcal{H}\{\}$ that the system performs on the input signal can represent many possible actions or phenomena. Systems can be designed to perform a specific operation, such as in a filter, where the system operates on the frequency content of the input signal, e.g., to create an output signal in which the high-frequency components of the input signal have been removed, as illustrated in Figure 1.2. But we could also apply the same perspective to investigate the dynamic characteristics of a device or structure that exists in physical reality, by studying the response of that device to a cleverly designed input signal. For instance, in a wind tunnel we can study how the blade of a wind turbine (the system studied) bends (the output of the system) as the result of applying variable wind loads on that blade (the input to the system), revealing the dynamic response of this wind turbine blade.

Whereas we can measure, characterize and analyze system input and output signals, and often also the operation performed by the system, in the time domain, it turns out that this analysis is generally much easier and provides more and more useful information when performing it in the frequency domain. Both signals (functions of time) *and* systems (their dynamic response in time) can be transformed to the frequency domain.

1.2 Signals in time and frequency

The analysis of signals, i.e., functions of time, in the time domain is often the first thing we do. For instance, after performing an experiment we plot the measured signal as a function of time, with time on the horizontal axis, and signal value on the vertical axis, to 'see what the signal looks like'. In many applications, however, such as in analyzing speech or other vibration-related signals, the signal may be difficult to interpret in time. Effects of measurement noise and the complexity of the signal studied may be prohibitive in detecting the underlying phenomenon. For example, periodic or oscillatory behavior (such as heart rhythms or machinery vibrations) may be hidden within seemingly random signals.

This is where spectral [1] analysis, that is, studying the signal in the frequency domain, has proven to be extremely useful. Spectral analysis involves the *decomposition of a signal into its constituent frequency components*, as if we 'build' the signal using basic building blocks. Compare this operation to placing a triangular glass prism (the system) into a beam of white

[1] From the Latin noun *spectrum* (plural: *spectra*), meaning 'an appearance, an image'; the noun comes from the Latin verb *specere*, meaning 'to look at, to view'.

light (the input to the system): the refraction of the prism (its output) will be a triangular beam of light with all visible colors separated, like a rainbow. This is because white light 'contains' all visible colors (wavelengths and hence frequencies), and the prism, through frequency-dependent refraction, splits the white light into its constituent frequency components, the individual colors. The prism splits the input signal and thus performs a spectral decomposition, allowing for spectral analysis. The prism allows us to 'see' what white light is composed of.

1.2.1 Spectral analysis principles

Sinusoids act as our gateway between the time domain and frequency domain. Sinusoid functions are very convenient: they are simple, they represent a pure oscillation with a specific frequency, they are continuous functions of time (without discontinuities) and they differentiate and integrate again into sinusoid functions.

In spectral analysis, requiring the transformation of signals from the time to the frequency domain, the waveform that acts as the main building block is the cosine function $b(t) = A\cos(2\pi f_0 t + \theta)$. This function is characterized by three parameters: its frequency f_0 in [Hz][2], its amplitude A and its phase θ in [rad].

Plotting this building block as a function of time t will show an oscillation, cosine-shaped, ranging between values $+A$ and $-A$, i.e., a periodic function with period T_0 in [s] equal to $\frac{1}{f_0}$. Plotting this building block as a function of frequency f requires two graphs: one that shows the amplitude of the signal as a function of frequency $A(f)$, and one that shows the phase of the signal as a function of frequency $\theta(f)$. For the cosine function, the amplitude and phase functions are zero for all frequencies, except *at* the frequency f_0 where the amplitude function equals A, $A(f_0) = A$, and where the phase function equals θ, $\theta(f_0) = \theta$.

An arbitrary periodic signal $x(t)$, with period T_0 and corresponding fundamental frequency f_0, can be built by *adding* cosine function building blocks with different frequencies, amplitudes and phases, resulting in a 'model' of this signal, $\tilde{x}(t)$. The frequency of each building block must be an integer multiple of the fundamental frequency f_0 for that building block to be periodic with T_0. They are all *harmonic* frequencies, with f_0 known as the 'first harmonic' and kf_0 the kth harmonic, with $k \in \mathbb{N}^+$. As will be discussed in Chapter 3, this also means that the building blocks are orthogonal. The model $\tilde{x}(t)$ of signal $x(t)$ can be written as:

$$\begin{aligned}\tilde{x}(t) &= A_0 + A_1\cos(2\pi 1 f_0 t + \theta_1) + A_2\cos(2\pi 2 f_0 t + \theta_2) + A_3\cos(2\pi 3 f_0 t + \theta_3) + \ldots \\ &= A_0 + \sum_{k=1}^{K-1} A_k\cos(2\pi k f_0 t + \theta_k) \end{aligned} \tag{1.1}$$

The value of constant A_0 (representing the average of $x(t)$) and the amplitudes and phases of the other $K-1$ building blocks $b_k(t) = A_k\cos(2\pi k f_0 t + \theta_k)$, for $k = 1, \ldots, K-1$, need to be chosen in such a way that the model $\tilde{x}(t)$ represents the signal $x(t)$ in the best possible way (e.g., through minimizing the mean square of error signal $e(t) = \tilde{x}(t) - x(t)$).

Each kth building block is a component of the signal, with amplitude A_k and phase θ_k, at frequency kf_0. Constructing a model of signal $x(t)$ is equivalent to decomposing that signal into its constituent frequency components. It allows us to 'see' what components are important to build the signal, i.e., at what frequencies the signal 'is happening with how much amplitude'; in other words, we 'see' the signal spectrum.

This process of signal (de)composition is illustrated in Figure 1.3. The top figure shows signal $x(t)$ as a black dashed line, a square wave which alternates between 1 and -1 from $t = -\infty$ to ∞ with a period of 2 seconds. Obviously, this signal also repeats itself every 4 s

[2] After the German physicist H.R. Hertz (1857-1894)

and 6 s et cetera, but the *smallest* repetition interval is 2 s, referred to as the *fundamental* period T_0 of this waveform. Its fundamental frequency f_0 then equals $f_0 = \frac{1}{T_0} = \frac{1}{2}$ Hz.

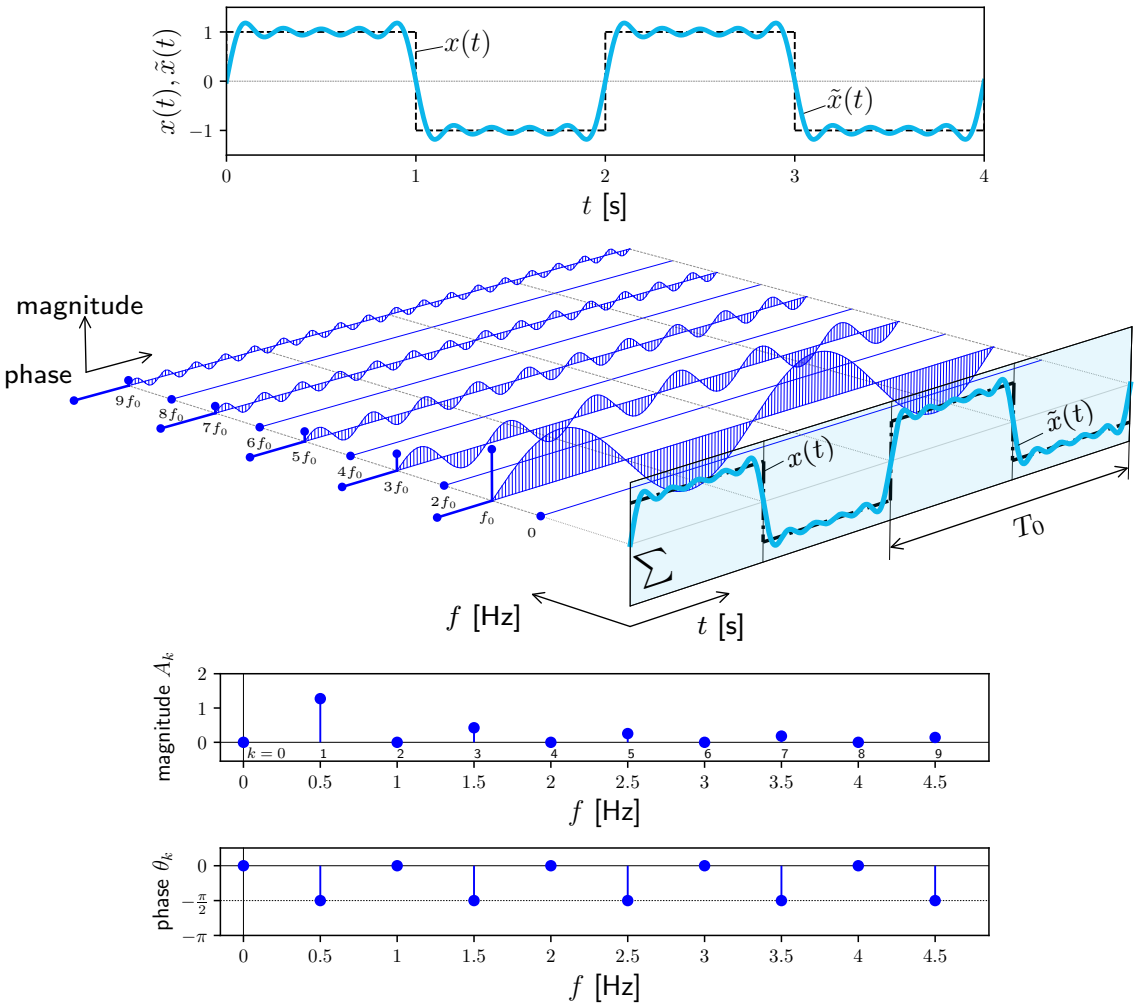


Figure 1.3: Transforming a periodic square wave $x(t)$ (period $T_0 = 2$ s) from the time domain (top) to the frequency domain (below). The middle figure (inspired by [2]) shows, in front, the square wave (black dashed curve) assembled as the sum $\tilde{x}(t)$ (light blue curve) of ten cosine 'building blocks' $A_k \cos(2\pi k f_0 t + \theta_k)$ with $k = 0, 1, 2, \ldots, 9$. The cosine waves are harmonically related as their frequencies are integer multiples of $f_0 = \frac{1}{T_0} = \frac{1}{2}$ Hz. At these frequencies, the cosine waves can have different amplitudes A_k (magnitude) and phases θ_k, as indicated by the stems on the left. The figure below shows the spectrum of $x(t)$.

The middle figure in Figure 1.3 shows in a three-dimensional plot how the same signal can be considered as a function of time t in [s], shown on the right axis, and as a function of frequency f in [Hz], on the left axis. The signal in time is decomposed into cosine building blocks, together forming the signal model $\tilde{x}(t)$. The frequency of each building block is an integer multiple of the fundamental frequency f_0 of the periodic waveform; they are all harmonics. The amplitude of each building block depends on how much of this building block is needed to construct signal $x(t)$. The phase of each building block depends on how much the building block must be shifted forward or backward in time to help construct signal $x(t)$.

The signal model $\tilde{x}(t)$ thus constructed uses $K = 10$ building blocks, $k = 0, 1, 2, \ldots, 9$, with frequencies equal to integer multiples of f_0, as indicated by the frequency axis f in [Hz]. Five of these building blocks are zero. Adding the five non-zero building blocks results in the light

blue signal shown in the top and middle graphs. The signal model alternates in a similar way as the square wave but is clearly not exactly the same. Using more building blocks, with higher frequencies (and typically smaller amplitudes) will increase the accuracy of our model.

The bottom figure in Figure 1.3 shows the amplitude (or magnitude) and phase (in [rad]) of the signal model $\tilde{x}(t)$ as a function of frequency. Given that we are modeling a periodic signal, with fundamental frequency f_0, both functions of frequency can only have non-zero values at integer multiples of this frequency. The signal model is relatively simple and although it can be further improved we 'see' at a glance how the square wave is constructed.

Apparently, for this particular square wave, the building blocks for all even-indexed integer multiples of f_0 ($0f_0$, $2f_0$, $4f_0$, $6f_0$, $8f_0$) are zero. All non-zero amplitude building blocks have a phase equal to $-\frac{\pi}{2}$. This is because for this signal, switching to $+1$ at $t = 0$, we exclusively need sine functions to construct it, and a sine function can be described as a cosine function having phase $-\frac{\pi}{2}$: $\sin\alpha = \cos\left(\alpha - \frac{\pi}{2}\right)$. These latter two findings are the result of the symmetry properties of this square wave and will be further explained in Chapter 3.

The spectral decomposition just performed (at a basic level) allows for spectral analysis and provides insight into what the dominant frequencies of this waveform are, i.e., what 'main components' signal $x(t)$ consists of. For aperiodic signals, which do not repeat themselves in time, we follow the same reasoning but let the period T_0 go to infinity, as if the signal *does* repeat itself after an infinite number of seconds. Then the building block fundamental frequency becomes infinitesimally small, the building block components at the different frequencies will lie infinitesimally close to each other in the frequency domain, ultimately forming a *continuous function of frequency*, known as the Fourier[3] transform $X(f)$ of signal $x(t)$.

In Part II the theory of signal analysis will be introduced for periodic signals, as we have done in this subsection, but also for aperiodic signals, using the Fourier transform. Signal analysis theory will be explained first for the continuous-time case, and then in Part III for discrete-time signals, the most common application, using the Discrete Fourier Transform.

1.2.2 Spectral analysis applications

A plethora of operations and applications opens up once a signal has been transformed to the frequency domain, but only three will be discussed here.

The spectral representation of a signal is often (much) more compact and meaningful than the time-domain representation, especially for a signal that has repeating patterns. By just focusing on the dominant frequency components of a signal, spectral analysis allows for easier interpretation, operation, compression, storage and transmission. For instance, in communication, when sending a measured signal from one place to another, the communicated signal could either be composed using the measured signal values in time *or* we could first perform a frequency transform and then send the signal spectrum, that is, the values of the signal in frequency. In many cases, the latter will require far less communication bandwidth. When compressing a signal, we could first transform it to the frequency domain and then cut off the higher-frequency components, thereby not losing much of the signal, given that for most signals the higher-frequency components will indeed become smaller and smaller in amplitude, as illustrated in Figure 1.3. Signal compression, such as composing a JPEG image, is indeed a common application of signal analysis, illustrated in Figure 1.4. Despite a significant compression, by a factor of 32, the object in the photo can still be recognized.

Measurements often contain some form of randomness caused by measurement noise, which often manifests itself at high frequencies. After transforming the signal to the frequency domain, it becomes easier to identify, isolate and remove the unwanted noise or interference from the desired signal. Filters can be designed to target the elimination of certain

[3] After the French mathematician and physicist J.-B.J. Fourier (1768-1830).

Figure 1.4: Photo of TU Delft's laboratory aircraft PH-LAB, operated together with the Netherlands Aerospace Center NLR, compressed to $\frac{1}{4}$ (center) and $\frac{1}{32}$ (right) of the original (copyrighted) at left.

frequency components of a signal. For instance, a so-called low-pass filter will pass the lower frequency band of its input signal and block the higher frequency band of this same signal. In entertainment applications, such as music or cinema, an equalizer can be used to boost the bass in a song or movie soundtrack through increasing the amplitudes of the lower-frequency components while keeping the existing amplitudes of the higher-frequency components.

Spectral analysis enables us to better understand and interpret the underlying structure of signals by examining their frequency components, which is – apart from mostly trivial cases – impossible in the time domain. A signal's dominant frequencies, possible harmonics and distribution in frequency of the signal component amplitudes can be studied. This allows potentially harmful resonant frequencies in mechanical systems to be identified and anomalies in electrical circuits to be detected. Spectral analysis of vibrations or acoustics can reveal information about machinery health and enables fault diagnosis, often in a non-destructive manner. Specific patterns in the frequency domain and changes in these patterns can indicate wear, imbalance, defects or other issues which may be impossible to 'see' in the time domain.

1.3 Systems in time and frequency

While the focus of this book is on signal analysis in the frequency domain, in Part V we introduce systems. Modeling and understanding a system's input-output relation, in time and frequency, is a key competence of an engineer and is briefly explained by a simple example.

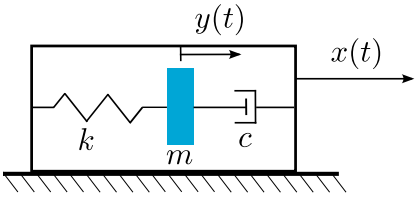


Figure 1.5: Mass-spring-damper system: a proof mass (in blue) is suspended by a spring and connected to a dashpot damper; it is positioned on a flat, perfectly horizontal table, such that effects of gravity can be ignored. This system turns external motion, input acceleration $\ddot{x}$, into output displacement y of the proof mass.

Consider the single-degree-of-freedom mass-spring-damper system illustrated in Figure 1.5, an elementary physical dynamic system. Its behavior is governed by a second-order differential equation:

$$\ddot{y}(t) + 2\zeta\omega_0\dot{y}(t) + \omega_0^2 y(t) = -\ddot{x}(t) \tag{1.2}$$

with both the input x and output y axis positive pointing right. In this equation ζ is the damping ratio (which is dimensionless), and equals $\zeta = \frac{c}{2\sqrt{mk}}$, ω_0 is the undamped natural angular frequency in [rad/s], $\omega_0 = \sqrt{\frac{k}{m}}$, with m the mass in [kg], c the damper constant in

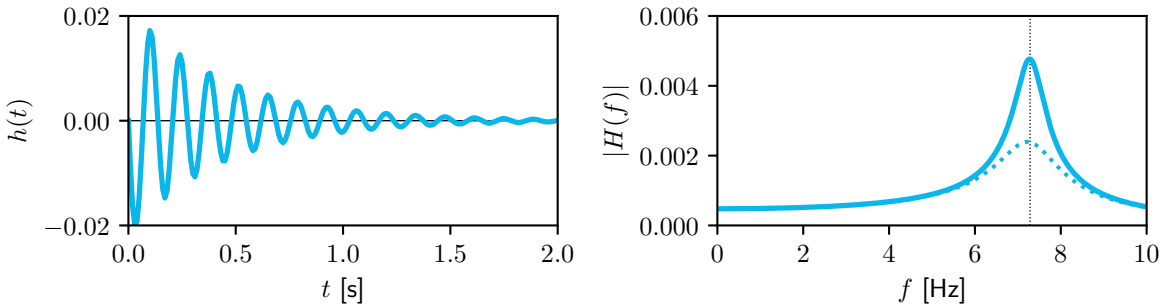


Figure 1.6: At left: impulse response $h(t)$ of mass-spring-damper system as a function of time t, with mass $m = 1.0$ kg, spring constant $k = 2100$ N/m and damping ratio $\zeta = 0.05$. At right: corresponding FRF $H(f)$ as a function of frequency f, with $\omega = 2\pi f$. The dotted vertical line marks the damped natural frequency $f_d = 7.28$ Hz. The dotted blue curve shows the FRF for a system with damping ratio that is twice as large: $\zeta = 0.10$.

[kg/s], and k the spring constant in [N/m].

Through solving (1.2) we find that the response of this system to a so-called impulse input at $t = 0$ (such as hitting it with a hammer on one of its sides) equals a damped harmonic described as a function of time t:

$$h(t) = -\frac{1}{\omega_d} e^{-\zeta\omega_0 t} \sin(\omega_d t) \tag{1.3}$$

for $t \geq 0$, with $\omega_d = \omega_0\sqrt{1-\zeta^2}$ the damped natural angular frequency in [rad/s]. This response to an impulse input, $h(t)$, is referred to as the impulse response function; it is shown in Figure 1.6 at left. The impulse response function fully characterizes the system response in the time domain. When $h(t)$ is known, we can compute the response $y(t)$ of this system to *any* input signal $\ddot{x}(t)$ through the procedure of convolution: $y(t) = h(t) * \ddot{x}(t)$, as will be explained in Chapter 7.

Having mastered the theory on signal analysis discussed in Part II, we will learn how to *also* characterize the system response as a function of frequency, known as the Frequency Response Function (FRF). Fourier-transforming the impulse response function $h(t)$ leads to the FRF $H(f)$. A crucial advantage of working with systems in the frequency domain relative to the time domain is that the tedious procedure of convolution, required to compute the system's response to an input, becomes a simple *multiplication* in the frequency domain: $Y(f) = H(f)\ddot{X}(f)$, with $Y(f)$ and $\ddot{X}(f)$ the Fourier transforms of the output and input signals, respectively. In most cases, we calculate $y(t)$ by first moving the entire problem to the frequency domain, using the Fourier transform, solve it there, and then inverse Fourier-transform $Y(f)$ back to $y(t)$.

The fact that signals and systems can both be transformed to the frequency domain, where the interplay of signals and systems can be described using algebraic equations, is extremely valuable. It allows for a quick and easy analysis of this interplay which is (apart from trivial cases) not possible in the time domain. Again, for a particular class of systems, the impulse response function $h(t)$ and its frequency domain equivalent, the FRF $H(f)$, fully characterize the dynamic behavior of that system and allow engineers to compute its response to arbitrary input signals; for instance, to predict the behavior of a structure (e.g., a wind turbine blade, a car suspension system, a bridge deck, et cetera) under a dynamic load.

In particular the FRF provides clear access to a structure's natural frequencies, which is of crucial importance as this allows to prevent excessive and possibly devastating resonance of the structure. The modulus of the FRF, relating input acceleration to output displacement of the mass-spring-damper system, is shown in Figure 1.6 at right as a function of frequency f

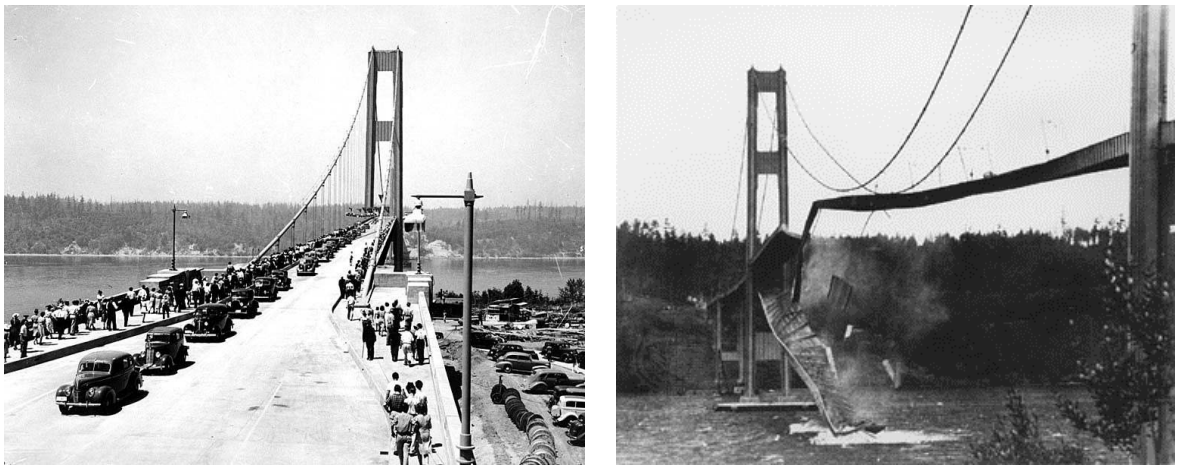


Figure 1.7: The Tacoma Narrows bridge, a suspension bridge in the US state of Washington that spanned Puget Sound, opened to traffic on July 1st, 1940, at left, and dramatically collapsed into Puget Sound on November 7th of the same year, at right, because of variable wind loads acting on the bridge deck. The Tacoma Narrows bridge is a classical and popular example of resonance in engineering, though the actual cause of the collapse was more intricate. Images taken from Wikimedia Commons and Wikimedia Commons [1], public domain.

in [Hz], and can be computed as:

$$H(\omega) = \frac{1}{\omega^2 - 2j\zeta\omega_0\omega - \omega_0^2} \tag{1.4}$$

with angular frequency $\omega = 2\pi f$, and j the imaginary unit with $j^2 = -1$. The FRF shows the response of the structure to a sinusoidal input over the entire range of frequencies; it gives a direct and clear view on how that structure amplifies or dampens vibration at different frequencies. For the mass-spring-damper system this graph shows a clear peak for the natural frequency f_0; the dotted vertical line is positioned at the damped natural frequency f_d. The width of the peak with regard to its height is driven by damping ratio ζ. For comparison, the graph also shows the FRF for a damping ratio twice as large: $\zeta = 0.10$. For the undamped case with $\zeta = 0.0$ the peak will go to infinity and the system will start to resonate when excited at that frequency. For small frequencies, $f \to 0$: $|H(f)| \to \frac{m}{k}$; for large frequencies, $|H(f)| \propto \frac{1}{f^2}$; for $f \to \infty$, $|H(f)|$ approaches zero. The FRF is a key element in structural dynamics for system identification, model validation and monitoring, with applications in aerospace engineering, civil engineering (see Figure 1.7) and many other fields.

1.4 This book

1.4.1 Definitions and notations

In this book, a signal is by default a function of time t in seconds. The Fourier transform of a signal is by default a function of frequency f in Hertz. Note that a signal can instead be a function of spatial position, or another running variable. In this book we stick to time t in [s] and frequency f in [Hz].

Signals in the time domain are written using lowercase letters, like $x(t)$, while that same signal in the frequency domain (its Fourier transform) is written using the uppercase equivalent, $X(f)$.

We focus on one-dimensional signals and their transforms. For transforms in two or more dimensions, like in imaging applications, the reader is referred to other books, e.g., [3].

The unit of a signal (when provided) is given in square brackets [ ], for instance [V] if the signal is a voltage measured by an electrical sensor.

1.4.2 Structure

The book consists of five parts, see Figure 1.8.

In Part I the book is introduced (this chapter) and the main signal analysis definitions and concepts are explained in Chapter 2. Appendices A–D present common mathematical formulas, define elementary functions, and explain complex algebra and the procedure of signal quantization, respectively.

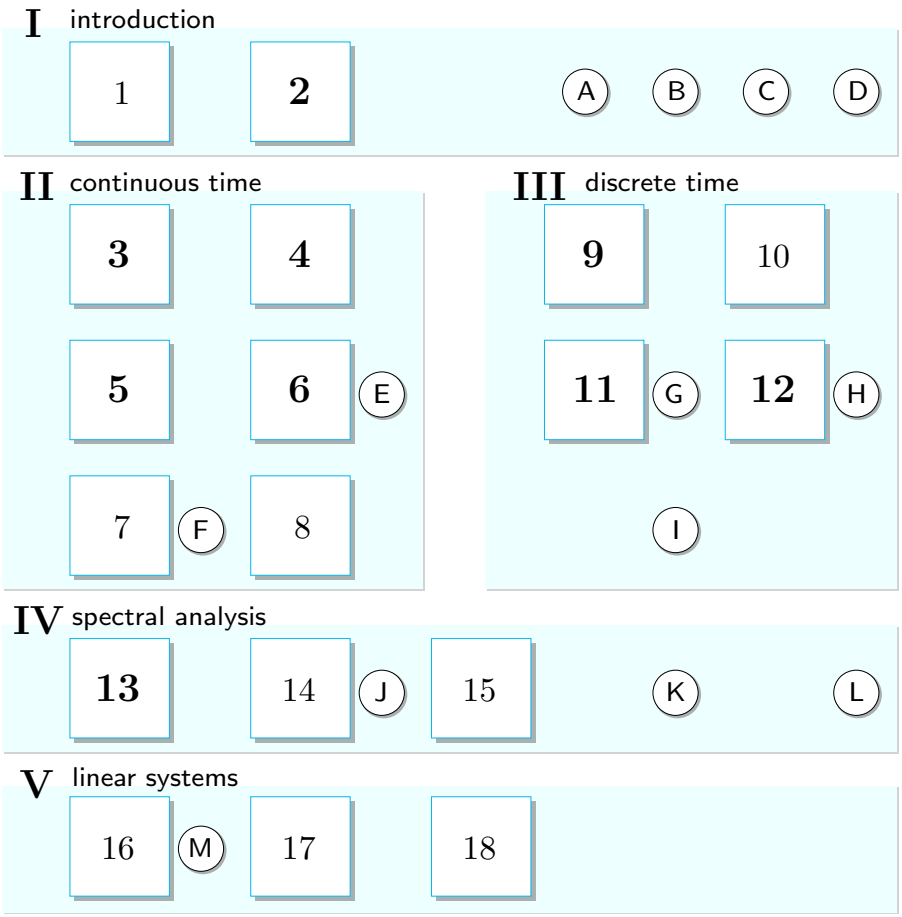


Figure 1.8: The book consists of five parts (light blue rectangles), eighteen chapters (squares) and thirteen appendices (circles). Boldface chapter numbers indicate essential material. The theory on continuous-time signals forms the main thread of this book. When working with discrete-time signals, the analogy with the continuous-time counterpart is maintained, for the reason of interpretation of the results and implications in physical reality.

Parts II and III contain essential material for engineers to understand and apply signal analysis for, respectively, continuous-time and discrete-time signals.

In many engineering applications the physical phenomena studied and the resulting (measurement) signals are continuous-time: $x(t)$, with $t \in \mathbb{R}$. Part II forms the theoretical heart of the book and focuses on the Fourier transform of continuous-time signals. Chapters 3 and 4 first discuss the real and complex exponential Fourier series models for periodic signals. Chapters 5 and 6 then introduce the Fourier transform and its main theorems, for aperiodic signals and, in-the-limit, for periodic signals. Chapter 7 explains the concept and properties of signal convolution, in the time and frequency domains. In Chapter 8 the consequences of time-windowing a signal on the Fourier transform of that signal are explained. Though typically not extensively covered in other sources, this chapter is a major step from infinite-duration signal analysis *theory* to finite-duration signal analysis *practice*. Some Fourier transform pairs are

listed in Appendix E; the convolution of discrete-time sequences is discussed in Appendix F.

Since nowadays, most, if not all, of measuring, processing and analyzing signals takes place on the computer, in discrete time, Part III discusses how to work with the discrete-time representation of continuous-time signal $x(t)$: the sequence of samples x_n, with $n \in \mathbb{Z}$. First the effects of sampling a continuous-time signal – the analog-to-digital conversion – on the Fourier transform of that signal are discussed in Chapter 9, using the impulse train sampling model. Chapter 10 discusses the reverse process – the digital-to-analog conversion – using ideal reconstruction and reconstruction in practice. Chapter 11 then derives the Fourier transform of the infinite-duration discrete-time sequence x_n, the Discrete-Time Fourier Transform (DTFT). Evaluating the DTFT of a time-windowed, and therefore finite-duration discrete-time sequence, at a limited number of frequencies, leads to the Discrete Fourier Transform (DFT), the most common signal transformation applied in practice, discussed in Chapter 12. Common DTFT and DFT pairs are summarized in, respectively, Appendices G and H. The full procedure, from the Fourier transform of a continuous-time infinite-duration signal to the DFT of a discrete-time finite-duration sequence, is summarized in Appendix I.

Ultimately, one of the main goals of signal analysis is to compute and analyze the spectrum of a signal. Spectral analysis is the topic of Part IV. In this part, first the energy spectral density (ESD) and power spectral density (PSD) functions of signals are defined in Chapter 13, for continuous-time signals $x(t)$ and discrete-time sequences x_n. The estimation of signal spectra for random but stationary, finite-duration discrete-time sequences is the topic of Chapter 14. The estimation of signal spectra for potentially non-stationary, finite-duration discrete-time sequences, the spectrogram, is briefly discussed in Chapter 15. Appendix J defines the confidence interval of the spectral estimate and Appendices L and K discuss, very briefly, the concepts of (cross-)correlation and white noise.

While this book focuses on the analysis of signals, in Part V the analysis of continuous-time systems is introduced. Chapter 16 defines system properties and an important class of systems, namely the linear time-invariant (LTI) system. The dynamic response of an LTI system is fully characterized in the time domain by its impulse response function, and in the frequency domain by the Fourier transform of this function, the frequency response function (FRF). The signals and systems perspective on practical measurements is discussed in Chapter 17, after which the book concludes with a chapter on filtering, the most common purpose of systems from the perspective of signal analysis, covered in Chapter 18. Appendix M explains, for a simple example, how to obtain the impulse response function of an LTI system from the system's differential equation.

2

Signal preliminaries

In this chapter we introduce models and definitions to characterize signals as functions of time. Some mathematical signal symmetry properties are discussed. We show how the sinusoid signal building block can be easily studied in the *frequency domain*, using phasors, to obtain the single-sided or double-sided spectrum. Then the so-called singularity functions are introduced as basis functions. The rules of signal operation in time are discussed, and it will be shown how new signals can be built using basis functions. We define the concepts of energy and power of signals in time, and explain what it means to move from mathematical definitions to real-life measurements: all signals become energy signals.

This chapter is accompanied by Appendices B–D, which provide background on some of the mathematical foundations and practical implications.

2.1 Signal models

We discuss deterministic and random (or stochastic) signals, and the continuous-ness and discrete-ness of signals both in time and in amplitude. We define a crucial characteristic, that is, whether a signal is periodic or aperiodic, and introduce the most important 'building block' in signal analysis: the sinusoidal signal.

2.1.1 Deterministic and random signals

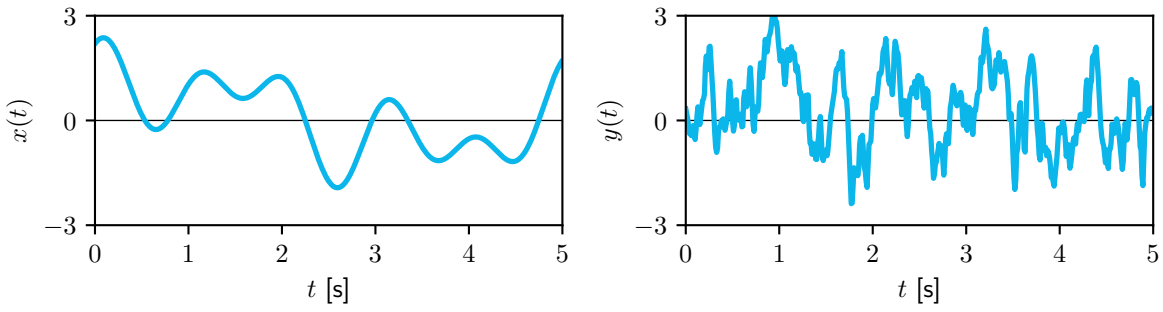


Figure 2.1: Deterministic (left) and random (right) signal; one realization of the random signal is shown.

Figure 2.1 illustrates an example of a deterministic and a random signal. A *deterministic* signal $x(t)$ is a specified function of time t. That is, when we know t, we know exactly what value x the signal has: $x(t)$. Examples are mathematical expressions, e.g.,

$$x(t) = 3t^2 + 2t - 4,$$

sinusoids, and the *unit pulse* function $\Pi(t)$ (width of 1 s) as defined in Appendix B:

$$\Pi(t) = \begin{cases} 1 & |t| < \frac{1}{2} \\ 0.5 & |t| = \frac{1}{2} \\ 0 & |t| > \frac{1}{2} \end{cases} \tag{B.1}$$

A *random* (or stochastic) signal has random values at any moment in time, and can only be modeled in a probabilistic way. Random signal $y(t)$ in Figure 2.1 (at right) will look different every time we measure it. In real life, randomness in signals (some random appearance) occurs constantly, because of the inherent variable nature of the phenomenon being observed and because of measurement noise. A large portion of this book focuses on deterministic signals. In reality, most signals are random in nature, and we return to this in Chapter 14.

2.1.2 Continuous and discrete signals

A second important characterization of a signal is whether it is continuous or discrete, both in terms of what *values* a signal can take, as well as in the independent variable time t, that is, at what *times* does the signal 'exist'. This is illustrated in Figure 2.2.

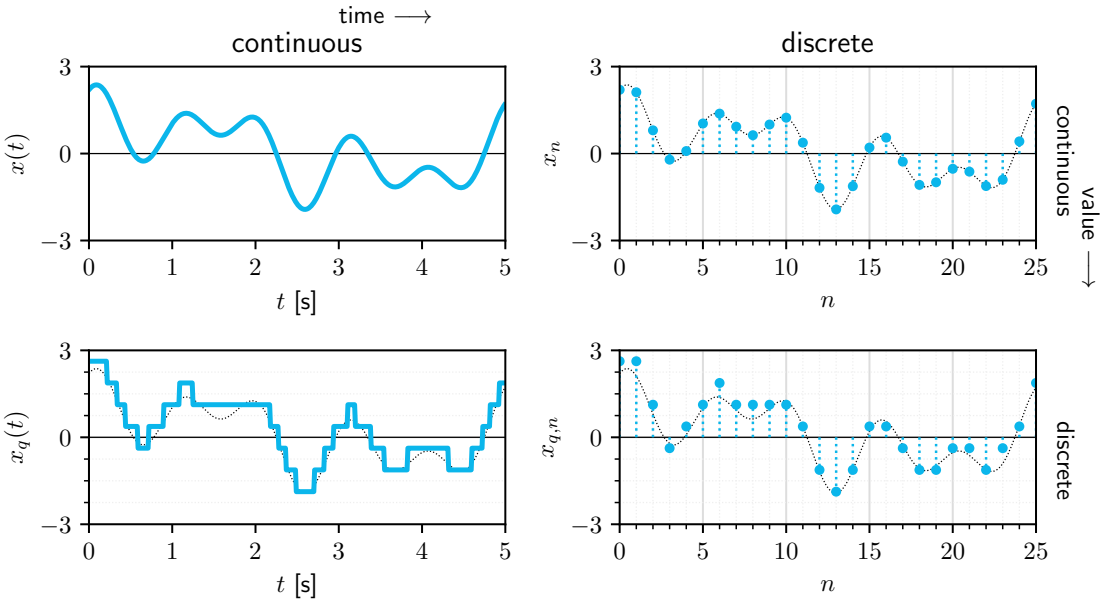


Figure 2.2: Signal continuity in time (continuous at left, discrete at right) and in value (continuous on top, discrete at bottom).

Most of our theory aims at analyzing signals that are continuous in time (they exist at all times $t \in \mathbb{R}$) and value (they can have all values in $\mathbb{R}$), as shown in the top-left of Figure 2.2. In real life, the signals (measurements) we process in our computers are *sampled* (we only have their values at discrete moments in time) and *quantized* (the samples cannot take on any value because we have a limited number of bits to represent them (here, 3, so $2^3 = 8$ possible values)). The signals we deal with in practice resemble the bottom right of Figure 2.2.

The principle of quantization is explained in Appendix D. Because the number of bits can be large (typically 14-16 bits or even more) the effects of quantization are ignored in this book. This means that our theory considers a discrete-time signal as illustrated at the top right of Figure 2.2. In other words, all signals in this book will, from now on, as an approximation, be considered continuous in value: $x_q(t) \approx x(t)$ and $x_{q,n} \approx x_n$.

Regarding notation, a continuous-time signal is indicated as $x(t)$; its discrete-time equivalent (the sampled version of $x(t)$) is written as x_n. While t can have any value $t \in \mathbb{R}$, n is the *index* of the sample, $n \in \mathbb{N}$, or $\mathbb{Z}$. In this sense, 'time is lost', and only when multiplying the index by the sample interval Δt (0.2 s in Figure 2.2) we retrieve time on the horizontal axis.

2.1.3 Periodic and aperiodic signals

A signal $x(t)$ is *periodic* if and only if:

$$\boxed{x(t+T_0) = x(t)}, \qquad -\infty < t < \infty \tag{2.1}$$

with T_0 the *fundamental period*. T_0 marks the *smallest* value in time after which the signal repeats itself. The *fundamental frequency* in [Hz] is then $f_0 = \frac{1}{T_0}$.

Any signal not satisfying (2.1) for any value of T_0 is called *aperiodic*. An example of an aperiodic signal is the unit pulse (B.1).

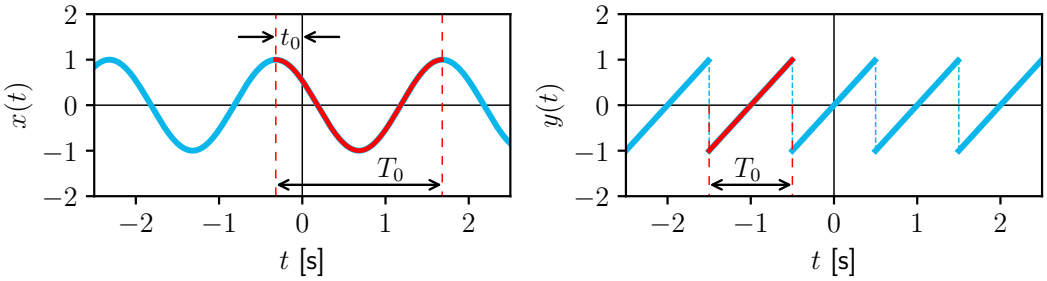


Figure 2.3: Periodic signals: sinusoid (left) and saw-tooth (right). T_0 indicates the period.

A special case is when the signal is a constant, i.e., $x(t) = c$, with $c \in \mathbb{R}$. This signal is periodic for *any* T_0; hence, the fundamental period cannot be determined.

Sinusoidal signals

The *sinusoidal* signal, the quintessential periodic signal that acts as the main building block of signal analysis, is defined as follows:

$$\boxed{x(t) = A\cos\left(2\pi f_0 t + \theta\right)}, \qquad -\infty < t < \infty \tag{2.2}$$

Here, A, f_0 and θ are constants referred to as the *amplitude*, *frequency* (in [Hz]) and *(relative) phase* (in [rad]), respectively. Amplitude $A > 0$ carries the *unit* of the signal, [unit]. It is important to note that the zero-phase, positive-amplitude *cosine* is taken as the reference for defining the phase, not the sine. A signal that is constant ($x(t) = c$, $c \in \mathbb{R}$) can be interpreted as a cosine with amplitude c, zero phase, and frequency equal to zero, $f_0 = 0$ Hz.

An example ($A = 1$, $f_0 = 0.5$ Hz, $\theta = 1$ rad) is illustrated in Figure 2.3. The phase shift θ (in [rad]) can also be considered as a *time* shift $t_0 = \frac{\theta}{2\pi f_0} = \frac{\theta}{2\pi} T_0$ (in [s]):

$$x(t) = A\cos\left(2\pi f_0 t + \theta\right) = A\cos\left(2\pi f_0\left(t + \frac{\theta}{2\pi f_0}\right)\right) = A\cos\left(2\pi f_0 (t + t_0)\right) \tag{2.3}$$

This book focuses on signal analysis; sinusoids are mostly characterized in [Hz], the linear (or ordinary) frequency f_0. This number represents the number of periods (or, as we will see, full rotations in the complex plane) per second. In books focusing on (dynamic) *system* analysis, sinusoids are often characterized using the angular (or *radial*) frequency ω_0, in [rad/s], where $\omega_0 = 2\pi f_0$.

Commensurability

A continuous-time sinusoidal signal $x(t)$ with frequency f_0 is periodic. One often constructs periodic signals using sinusoidal components, e.g.:

$$x(t) = A_1 \sin(2\pi f_1 t) + A_2 \sin(2\pi f_2 t) \tag{2.4}$$

with (A_1, A_2) and (f_1, f_2) the amplitudes and frequencies of the two components, respectively. While the individual components are both periodic, what can we say about the periodic-ness of their sum? The concept of *commensurability* provides the answer. Two frequencies f_1 and f_2 are commensurable if they have a common measure. That is, there is a number f_0 contained in each of them an *integer* number of times: $f_1 = n_1 f_0$ and $f_2 = n_2 f_0$, with $n_1, n_2 \in \mathbb{N}$. The *highest* number f_0 for which this is valid is the *fundamental frequency*. Integers n_1 and n_2 are then the *smallest* integers defining the ratio, and $T_0 = \frac{1}{f_0}$ is the *fundamental period*.

In other words, for commensurable frequencies the ratio of the frequencies of the components $\frac{f_1}{f_2} = \frac{T_2}{T_1} = \frac{n_1}{n_2}$ is a rational number, $\frac{n_1}{n_2} \in \mathbb{Q}$. *Only* when the frequencies of the components (two or more) are commensurable, their summation will yield a periodic signal.

EX 2.1 Suppose in (2.4) $f_1 = 1.5$ Hz and $f_2 = 2.0$ Hz. What is the fundamental frequency?

Solution Their ratio equals $\frac{f_1}{f_2} = \frac{1.5}{2}$ which, when written using (the smallest possible) integers, equals $\frac{3}{4}$, a rational number. Then: $\frac{f_1}{f_2} = \frac{3}{4} = \frac{n_1}{n_2} = \frac{n_1 f_0}{n_2 f_0}$. Hence, $f_1 = n_1 f_0 = 3f_0 = 1.5$ Hz, so $f_0 = 0.5$ Hz. In conclusion, the sum of components is periodic and we obtain fundamental frequency $f_0 = 0.5$ Hz and period $T_0 = 2$ s.

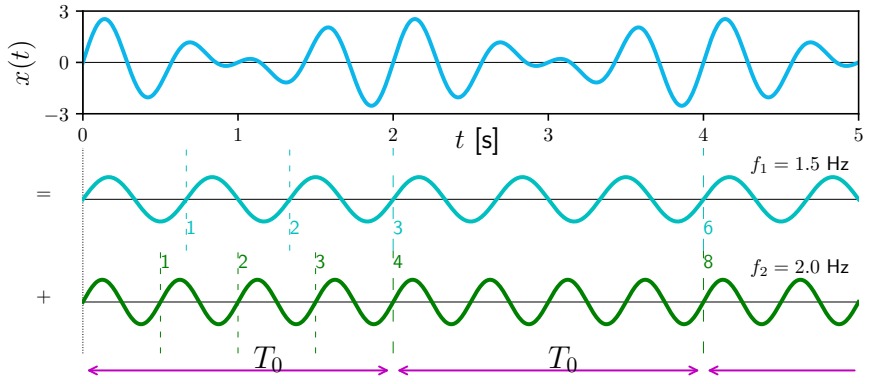


Figure 2.4: Finding the fundamental frequency and period when adding two harmonic components.

Figure 2.4 shows the summation $x(t)$ in the top row, and the two components in the middle and bottom row. At $t = 0$ the two components are synchronized (same phase), while the first time they *both* complete an integer number of periods (that is, the first time they are 'starting again' *together*) is after 2 seconds. At that moment the first component has completed 3 full periods ($= n_1$), the second component 4 periods ($= n_2$). After 4 s this pattern repeats again; the first component has completed 6 periods ($= 2n_1$), the second component 8 periods ($= 2n_2$). Etcetera. The *first* time the two components repeat themselves together corresponds to the period. The fundamental frequency is the *highest* common frequency f_0, as this belongs to the *smallest* time T_0, the first time the resulting signal repeats itself.

Suppose $f_1 = 3.5$ Hz and $f_2 = 6.1$ Hz. What is the fundamental frequency? EX 2.2

Solution The highest common frequency equals $\frac{f_1}{f_2} = \frac{3.5}{6.1} = \frac{35}{61} = \frac{n_1 f_0}{n_2 f_0}$, so $f_1 = 3.5 = 35 f_0$ and $f_2 = 6.1 = 61 f_0$, hence $f_0 = 0.1$ Hz, $T_0 = 10$ s. At that time, the first component repeated itself 35 times, the second component 61 times. It is the *first* time this happens; after 20 s (then $f_0 = 0.05$ Hz), 40 s (then $f_0 = 0.025$ Hz) etc. it happens again. We need the *greatest* common divisor to find the fundamental frequency.

Assume $f_1 = 2$ Hz and $f_2 = \pi$ Hz. What is the fundamental frequency? EX 2.3

Solution Their ratio equals $\frac{f_1}{f_2} = \frac{2}{\pi}$, an irrational number $\in \mathbb{R}$, not $\in \mathbb{Q}$. While the components are periodic, their summation will not be periodic.

Do the magnitudes of the components matter here, or their relative phases? Consider Figure 2.4, and you can see that the answer is no. Only the frequencies matter.

Singularity functions

The sinusoidal signal, (2.2), is periodic, oscillates about zero and runs from $t = -\infty$ to $+\infty$. But how can we describe phenomena that happen once and last for instance just a second, or switch from one value to another in some time interval? This is where the *aperiodic* singularity functions, defined in Appendix B, come into play: the unit impulse $\delta(t)$ (Dirac[1], or delta function), unit step $u(t)$ and unit ramp $r(t)$ functions, see Figure 2.5. Especially the *sifting property* of the Dirac function, see (B.12) and Figure B.4, is important.

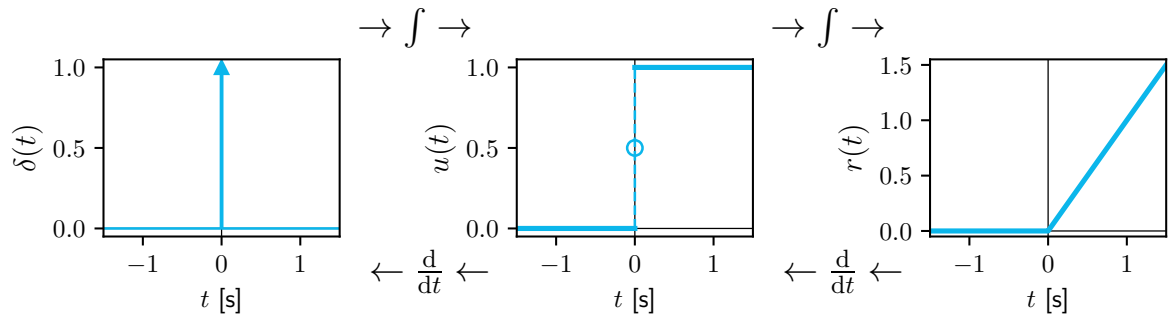


Figure 2.5: Unit impulse $\delta(t)$, unit step $u(t)$, unit ramp $r(t)$ functions, and how they relate through integration (from left to right) and differentiation (from right to left). Note that the value of a function with a sudden jump is, in this book, defined to lie 'half-way the jump', Appendix B. The dashed line and small circle at $t = 0$ s in $u(t)$ are drawn here to accentuate this point, but will be replaced with a solid line in later chapters.

More singularity functions exist, such as the unit parabolic function, the unit pulse function $\Pi(t)$ (B.1) and unit triangular function $\Lambda(t)$ (B.2). Note that, whereas they are introduced here as functions of time, these same functions can *also* have frequency f as their 'running variable', or any running variable for that matter. They are 'just' mathematical functions often used in signal analysis, in time *and* frequency. To prevent confusion, it is important to stick to the notation (u for step function, etc.) and use lowercase for functions of time ($u(t)$) and, in later chapters, uppercase for functions of frequency ($U(f)$).

[1] After the British theoretical physicist P.A.M. Dirac (1902-1984).

2.2 Signal symmetry

2.2.1 Definitions

Signals can exhibit characteristics that show a particular behavior in time, the most important of which in our context is whether a signal is even, odd, or half-wave odd. These symmetry properties are illustrated in Figure 2.6 and are defined in mathematical terms in Table 2.1.

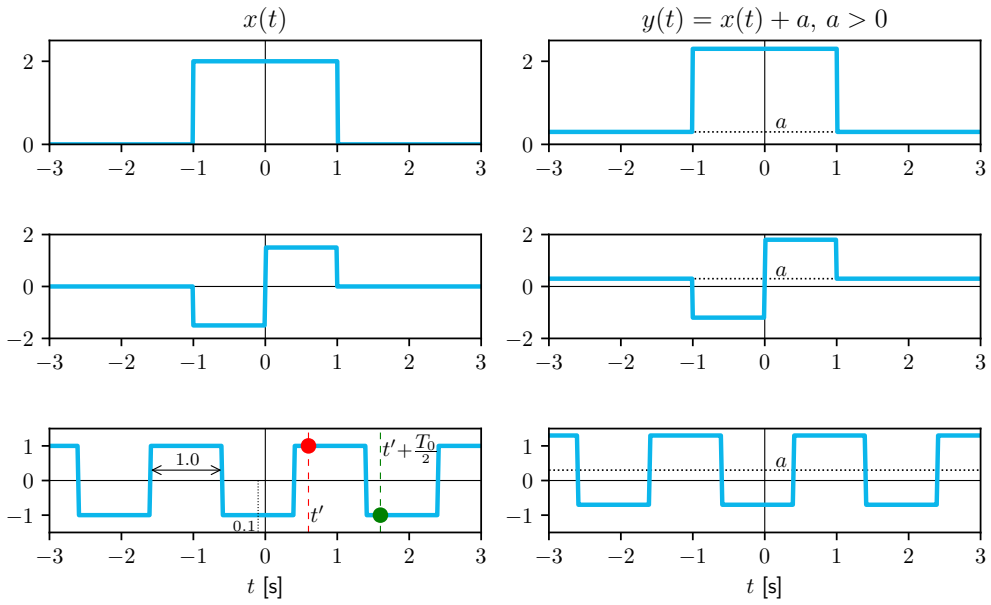


Figure 2.6: Even, odd and half-wave odd symmetry: pulse (top), doublet (middle), square wave (bottom).

The left-hand side of Figure 2.6 shows the even, odd and half-wave odd signals $x(t)$ which correspond to the mathematical definitions of Table 2.1. Evenness and oddness express symmetry with respect to $t = 0$. The half-wave odd property (bottom left) is strictly for periodic signals, here $T_0 = 2$ s. Imagine taking a sample from this signal at any arbitrary moment in time t' (the red dot in Figure 2.6), and then taking another sample half of a period later (or earlier) $t' \pm \frac{T_0}{2}$ yields the same signal value but with opposite sign (green dot).

The right-hand side of Figure 2.6 shows signal $y(t)$, constructed by adding a non-zero constant a to signal $x(t)$: $y(t) = x(t) + a$. Adding the constant does not 'destroy' the evenness of the function, $y(t)$ is even. The same cannot be said about the oddness and half-wave oddness property, which do not hold anymore for $y(t)$. Odd signals and half-wave odd signals must have a zero average. When subtracting the mean (a) from the signal $y(t)$ we obtain $x(t)$. Hence, non-zero mean signals can have characteristics of oddness or half-wave oddness 'about their average'. This will be discussed further in Chapters 3 and 4 on Fourier series.

Table 2.1: Mathematical definitions of signal symmetry properties.

even	odd	half-wave odd
$x(t) = x(-t)$	$x(t) = -x(-t)$	$x(t \pm \frac{T_0}{2}) = -x(t)$
	zero average	zero average
		periodic (T_0)

When considering the half-wave odd signal $x(t)$ in the bottom left of Figure 2.6, we see that this signal is neither odd nor even. Shifting it a little bit (0.1 s) to the right would make

it an even signal. Shifting it 0.4 s to the left would make it an odd signal. In both cases the signal remains half-wave odd. In fact, no matter how much we shift this signal to the left or right, it will maintain this property. Signals can therefore be odd and half-wave odd, or even and half-wave odd; these properties are independent.

2.2.2 Even and odd signal parts

Note that *any* signal $x(t)$ can be written as the sum:

$$\boxed{x(t) = x_e(t) + x_o(t)} \tag{2.5}$$

of an even signal component $x_e(t)$ and an odd signal component $x_o(t)$, which in turn can be computed as follows:

$$x_e(t) = \frac{1}{2}\left(x(t) + x(-t)\right) \tag{2.6}$$

$$x_o(t) = \frac{1}{2}\left(x(t) - x(-t)\right) \tag{2.7}$$

This finding has important consequences, as discussed in the following chapters. When $x(t)$ is even, the odd component $x_o(t)$ is zero; when $x(t)$ is odd, the even component $x_e(t)$ is zero. Except these special cases, signals typically have both even and odd components.

As a first example, consider the half-wave odd signal $x(t)$ at the bottom left of Figure 2.6. Figure 2.7 shows how the even and odd components of this signal are constructed. Note how the components $x_e(t)$ and $x_o(t)$ share the same amplitude and period as $x(t)$.

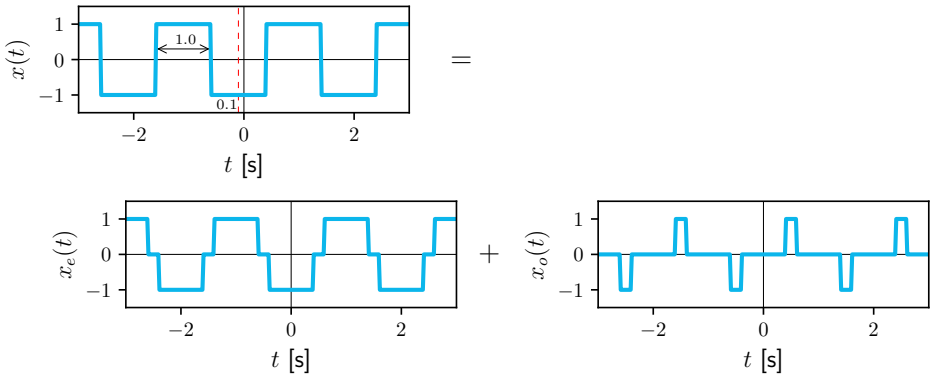


Figure 2.7: Even and odd parts of half-wave odd signal of Figure 2.6: $x(t) = x_e(t) + x_o(t)$.

As a second example, take the sinusoidal signal $x(t) = A\cos(2\pi f_0 t + \theta)$ introduced in Section 2.2. Using the trigonometric relation (A.7) we obtain:

$$A\cos\left(2\pi f_0 t + \theta\right) = \underbrace{(A\cos\theta)\cos(2\pi f_0 t)}_{\text{even part}} + \underbrace{(-A\sin\theta)\sin(2\pi f_0 t)}_{\text{odd part}} \tag{2.8}$$

with a zero phase cosine and zero phase sine, respectively, as a function of time. Hence, a plain cosine and sine together, with the same frequency, can create a cosine with any phase θ. Figure 2.8 illustrates the even and odd components of this signal for four values of θ.

2.3 Sinusoid building block

The sinusoid (2.2) is the most important building block in signal analysis. In Chapter 3 it will be shown that periodic signals can be 'constructed' using a Fourier series, a summation

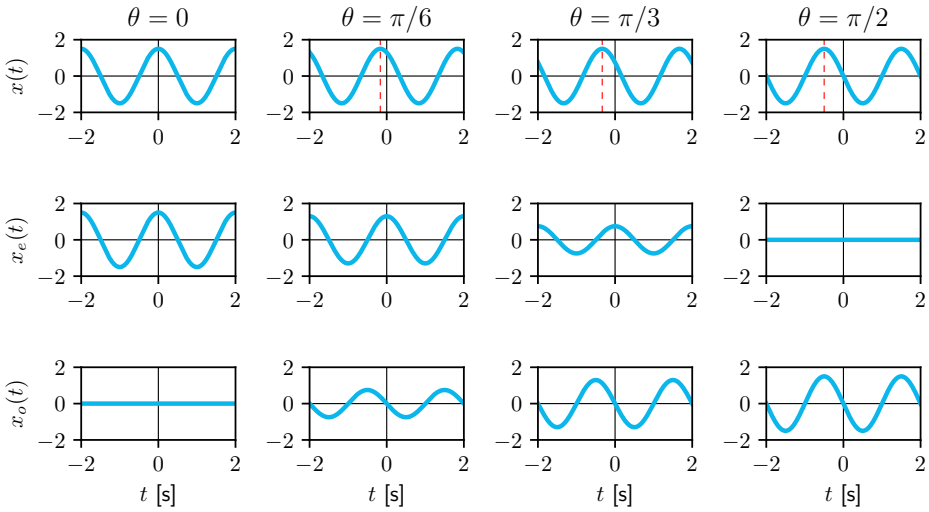


Figure 2.8: Even and odd parts of sinusoidal signal, (2.8), for four values of phase shift θ ($A = \frac{3}{2}$, $f_0 = 0.5$ Hz).

of sines and cosines that share a common frequency. Sinusoidal signals have many features that make them easy to use, such as orthogonality, discussed in Appendix A. But another useful property of sinusoidal signals is that they can be easily described in (in other words: *'transformed to...'*) the frequency domain. While in this book we only deal with real-valued signals, it is customary to represent these real-valued signals in terms of complex quantities.

2.3.1 Phasor description

The common way to describe a sinusoidal signal in the frequency domain is to use *phasors*, as discussed in Appendix C. In short, we use Euler's[2] formula (C.1) to relate *real-valued* sinusoidal signals to the *complex-valued* exponential basis function, as illustrated in Figure 2.9.

The complex exponential basis function $e^{j2\pi f_0 t}$, a complex number representing a unit vector in the complex plane, rotates counterclockwise in the complex plane (top left) as t advances. Its magnitude is one; its phase at $t = 0$ is zero. The frequency f_0 defines how fast it rotates; when $f_0 = 1$ Hz, it makes a full rotation of 2π radians in 1 second. The *real* part of this function is the cosine $\cos(2\pi f_0 t)$ (bottom left), the *imaginary* part of this function is the sine $\sin(2\pi f_0 t)$ (top right). When $t' = T_0 = \frac{1}{f_0}$ a full rotation is completed and the function starts all over again, it is periodic. The complex *conjugate* of this function, $e^{-j2\pi f_0 t}$, rotates clockwise in the complex plane; the real part is the same cosine, the imaginary part is the negative sine, Figure C.4. This representation is at the very heart of Fourier analysis.

The phasor description appears when representing a sinusoidal signal using the *real* part of the complex exponential function:

$$x(t) = A\cos\left(2\pi f_0 t + \theta\right) = \mathrm{Re}\left(Ae^{j(2\pi f_0 t+\theta)}\right) = \mathrm{Re}\Big(\underbrace{Ae^{j\theta}}_{\text{phasor } \vec{X}} e^{j2\pi f_0 t}\Big) \tag{2.9}$$

The phasor $\vec{X}$, (C.11), is (usually) a complex number, scaling the complex exponential basis function (a unit vector) magnitude from 1 to A, and setting its phase angle at $t = 0$ seconds, from 0 to θ. In other words, the phasor $\vec{X}$ is the initial position ($t = 0$) of the complex vector $Ae^{j(2\pi f_0 t+\theta)}$ in the complex plane. Together with the frequency f_0 of the basis function, $x(t)$ is fully described by the phasor, (2.9).

[2] After the Swiss mathematician L. Euler (1707-1783).

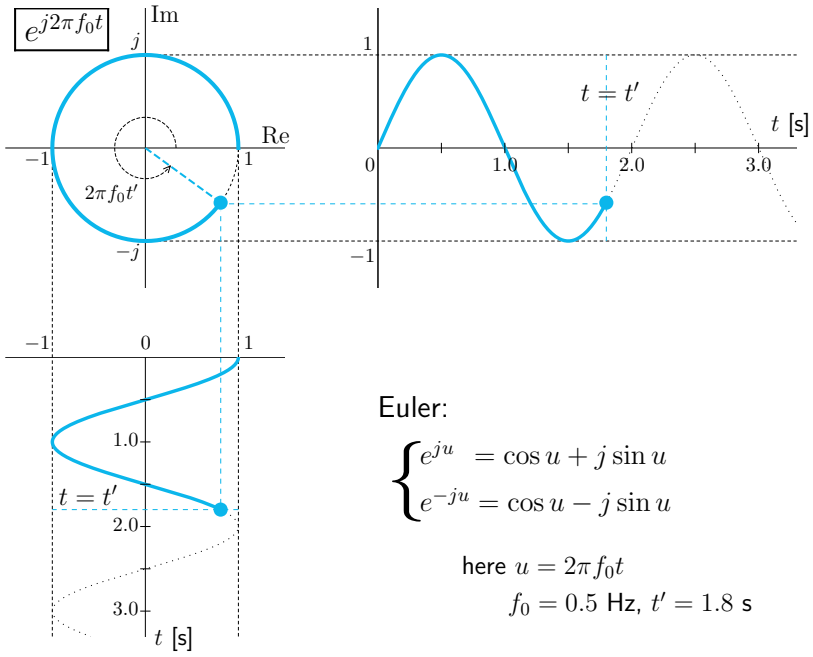


Figure 2.9: The complex exponential basis function $e^{j2\pi f_0 t}$, uniting the sine (top right) and cosine (bottom left) time functions in a unit vector rotating in counterclockwise direction in the complex plane.

2.3.2 Single-sided spectrum

A real-valued signal can be described as the projection of the complex rotating exponential, multiplied by the phasor, on the real axis. To obtain the single-sided spectrum, we simply show at what frequency (or frequencies) a signal $x(t)$ is active, i.e., has a non-zero amplitude. At that frequency (or frequencies) the signal has an amplitude and phase, resulting in an amplitude 'as a function of frequency', the *amplitude spectrum*, and a phase 'as a function of frequency', the *phase spectrum*. This is illustrated in the left column of Figure 2.10, for a sinusoidal signal with $A = 1.5$, $\theta = \frac{\pi}{2}$ and $f_0 = 2$ Hz.

2.3.3 Double-sided spectrum

A real-valued signal can also be described as the summation of the complex rotating exponential, multiplied by the phasor, and its complex conjugate, divided by two:

$$\begin{aligned} x(t) &= A\cos\left(2\pi f_0 t + \theta\right) &&= \frac{1}{2}A\left(e^{j(2\pi f_0 t+\theta)} + e^{-j(2\pi f_0 t+\theta)}\right) \\ &= \frac{1}{2}Ae^{j\theta}e^{j2\pi f_0 t} + \frac{1}{2}Ae^{-j\theta}e^{-j2\pi f_0 t} &&= \frac{1}{2}\vec{X}\,e^{j2\pi f_0 t} + \frac{1}{2}\vec{X}^*\,e^{-j2\pi f_0 t} \end{aligned} \tag{2.10}$$

This representation, in terms of two *conjugate*, oppositely rotating complex exponential functions is the most common one in signal analysis (Figure C.4). It can be considered as the sum of two rotating complex vectors, one with a *positive* frequency and one with a *negative* frequency. Of course, negative frequencies do not exist in physical reality. It is a mathematical abstraction that results from describing real-valued signals (which occur on the real axis of the complex plane) using complex conjugate exponential basis functions.

The double-sided spectrum of the same signal as above ($A = 1.5$, $\theta = \frac{\pi}{2}$ and $f_0 = 2$ Hz) is illustrated in the right column of Figure 2.10. Note that the amplitude spectrum has an *even* symmetry, and the phase spectrum has an *odd* symmetry. This is simply the consequence of using two conjugate rotating phasors to obtain a *real-valued* signal. Their magnitudes *must*

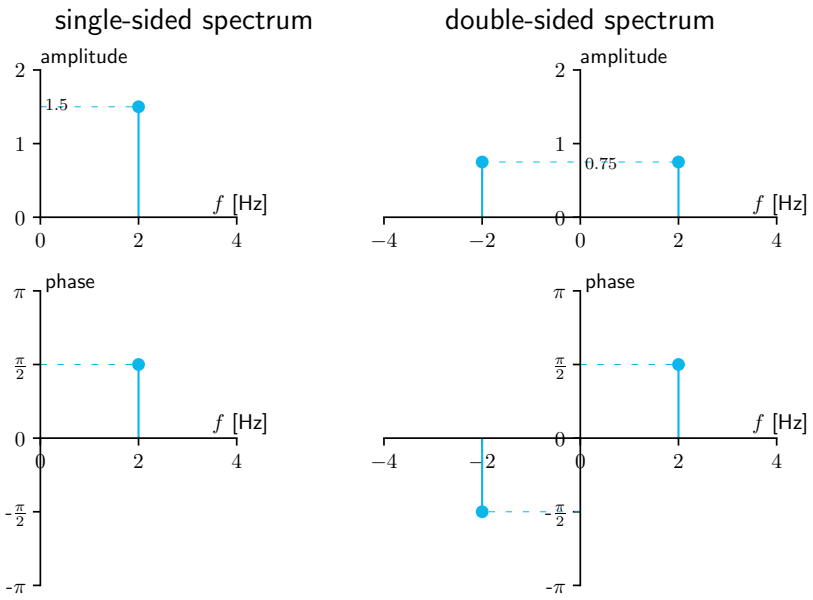


Figure 2.10: Single-sided (left) and double-sided (right) amplitude (top) and phase (bottom) spectra of $x(t) = 1.5\cos(4\pi t + \frac{\pi}{2})$. The convention is to show the phase spectrum below the amplitude spectrum.

be the same and their phases *must* be equal, but opposite, in sign. This also follows directly from (2.10). See also Appendix C, (C.2) and (C.13).

Clearly, these relations and findings are elementary. However, as we will see later in this book, they have massive consequences for signal analysis. The convention here is the use of the *double-sided* spectrum, although with the above-mentioned symmetries of amplitude and phase, we may only show one half, namely, the part for *positive frequencies*. A common mistake is then to name this half of the double-sided spectrum the *single*-sided spectrum, as only one side is shown. But as we can clearly see in Figure 2.10 the amplitudes of the double-sided spectrum are *half* of the amplitudes of the single-sided spectrum (the phase is identical). As a final note, for most periodic signals the amplitude and phase spectra will only have non-zero values at some frequencies, and these spectra are referred to as *line spectra*.

EX 2.4

Consider the following signal: $x(t) = 4\sin(30\pi t) + 3\cos(70\pi t + \frac{\pi}{4})$. Construct the single-sided and double-sided spectra of this signal.

Solution This signal has two components with frequencies $f_1 = 15$ and $f_2 = 35$ Hz. When drawing the single-sided spectrum, we first need to write the sine component as a cosine, because the sine component of the rotating complex exponential is *imaginary*. Using $\sin u = \cos(u - \frac{\pi}{2})$, the first component becomes $4\cos(30\pi t - \frac{\pi}{2})$.

The single-sided spectrum then quickly follows, from:

$$x(t) = \mathrm{Re}\left(4e^{j(30\pi t - \frac{\pi}{2})} + 3e^{j(70\pi t + \frac{\pi}{4})}\right)$$

Similarly, for the double-sided spectrum we obtain:

$$x(t) = 2e^{j(30\pi t - \frac{\pi}{2})} + 2e^{-j(30\pi t - \frac{\pi}{2})} + \frac{3}{2}e^{j(70\pi t + \frac{\pi}{4})} + \frac{3}{2}e^{-j(70\pi t + \frac{\pi}{4})}$$

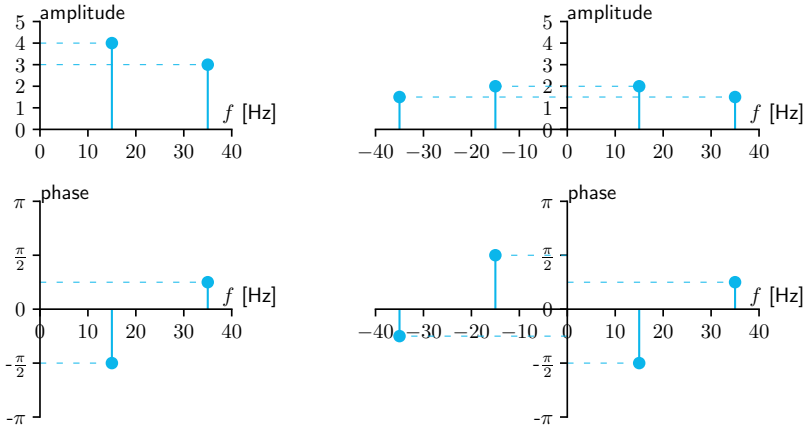


Figure 2.11: Single-sided (left) and double-sided (right) spectra of Example 2.4.

2.3.4 Spectrum conventions

By convention, the *cosine* function is taken as the *reference*: sine functions are 'converted to' cosines, a zero-phase sine is a cosine with a phase of $-\frac{\pi}{2}$.

In the amplitude spectrum, amplitudes are taken to be positive; a minus sign in front of the amplitude is accounted for through adding π to (or subtracting π from) the phase.

In the phase spectrum, the phase lies within $[-\pi, \pi]$. While a signal's (unwrapped) phase θ may be smaller than $-\pi$ or larger than π, it is wrapped into the region $[-\pi, \pi]$ by adding (integer multiples $k \in \mathbb{Z} \setminus \{0\}$ of) 2π. Visualizing complex number $X = Ae^{j\theta}$ in the complex plane (Figure C.1) implies an *ambiguity*: $Ae^{j\theta}$ and $Ae^{j(\theta+k2\pi)}$ are mapped on the same vector.

As an example, before drawing its spectrum, signal $x(t) = -5\cos(2\pi t + \frac{\pi}{4})$ is converted to $x(t) = 5\cos(2\pi t - \frac{3\pi}{4})$, rather than $x(t) = 5\cos(2\pi t + \frac{5\pi}{4})$.

The commensurability of the signal components is *not* required to draw its spectrum.

2.4 Signal operations

While we focus on operating on the argument of functions of time, the same rules hold for operated functions of frequency. Clever operations on building blocks allows us to construct an infinite variety of new functions of time (or frequency). One must first define the 'rules of precedence' and apply these rules consistently. This can be compared to agreeing on, and following the 'order of operations' in mathematics. E.g., $4 + 2 \times 3$ equals 10 rather than 18. This is because we defined the operation of multiplication to take precedence over addition.

2.4.1 Rules

Imagine that we have an arbitrary function of time $x(t)$. Suppose we would perform some operation on the time-argument, like $\beta t + \alpha$; parameter β is dimensionless, parameter α has unit [s]. Then what does signal $y(t) = x(\beta t + \alpha)$, with argument $\beta t + \alpha$, rather than just t, look like? To answer this question, first rewrite $y(t)$:[3]

$$y(t) = x(\beta t + \alpha) = x\left(\beta\left(t + \frac{\alpha}{\beta}\right)\right) \tag{2.11}$$

and then apply the following three rules.

[3]Note that we aim to obtain $1t$ again, i.e., $x(\beta(1t + \frac{\alpha}{\beta}))$, similar as in (2.3).

(1)	$\beta < 0$	:	function is flipped around $t = 0$
(2)	$\lvert\beta\rvert \neq 1$	:	function is scaled in time:
		$\lvert\beta\rvert > 1$	signal is *compressed* (signal goes faster)
		$\lvert\beta\rvert < 1$	signal is *expanded* (signal goes slower)
(3)	$\alpha \neq 0$	:	function is shifted in time, with $t_0 = \frac{\alpha}{\beta}$:
		$t_0 > 0$	signal is *shifted to the left* (signal comes earlier, time advance)
		$t_0 < 0$	signal is *shifted to the right* (signal comes later, time delay)

The rules define the order of precedence, first (1), then (2), and then (3). Consistency in applying the rules (in time and frequency) is extremely important.

Finally we note that a signal can also be transformed in terms its amplitude: $y(t) = \gamma x(\beta t + \alpha)$, where the amplitude is scaled by a factor of γ.

2.4.2 Examples

EX 2.5 What does signal $x(t) = r(-3t - 6)$ look like?

Solution This is an operation of the unit ramp function $r(t)$. First re-write $x(t)$:

$$x(t) = r(-3t - 6) = r(\underbrace{-3}_{=\beta}(t \underbrace{+2}_{=\frac{\alpha}{\beta}})) \tag{2.12}$$

First step: we see that $\beta < 0$, so the function $r(t)$ is mirrored about $t = 0$. Second step: $\lvert\beta\rvert = 3$, larger than 1, the signal compresses and runs three times as fast, i.e., the $r(t)$ function rises more rapidly over time. Third step: $\frac{\alpha}{\beta} = +2$, so $t_0 > 0$, meaning that the signal shifts *left* by 2 seconds. The resulting signal is shown in Figure 2.12 (left), with the blue circle added to check that $x(t)$ rises by $+3$ units per second.

EX 2.6 What does signal $y(t) = \Pi(-2t + 3)$ look like?

Solution This is an operation of the unit pulse function $\Pi(t)$. First re-write $y(t)$:

$$y(t) = \Pi(-2t + 3) = \Pi(\underbrace{-2}_{=\beta}(t \underbrace{- \frac{3}{2}}_{=\frac{\alpha}{\beta}})) \tag{2.13}$$

First step: we see that $\beta < 0$, so the function $\Pi(t)$ is mirrored about $t = 0$, which has no effect in this case, as $\Pi(t)$ is even. Second step: $\lvert\beta\rvert = 2$, larger than 1, so the signal compresses, it runs twice as fast and the $\Pi(t)$ function width shrinks to half a second. Third step: $\frac{\alpha}{\beta} = -\frac{3}{2}$, so $t_0 < 0$, meaning that the signal shifts *right* by $\frac{3}{2}$ seconds. The resulting signal is shown in Figure 2.12 (right).

To check one's answer regarding where on the time axis the operated signal appears, we set everything within parentheses to zero and check where the feature of the original signal at $t = 0$ shows up within the operated signal. For instance, in Example 2.5: $(-3t - 6) = 0 \Rightarrow t = -2$ seconds, where the 'bend' of the operated $r(t)$ appears.

In Example 2.6 the signal was compressed by a factor of 2, and the pulse width changed

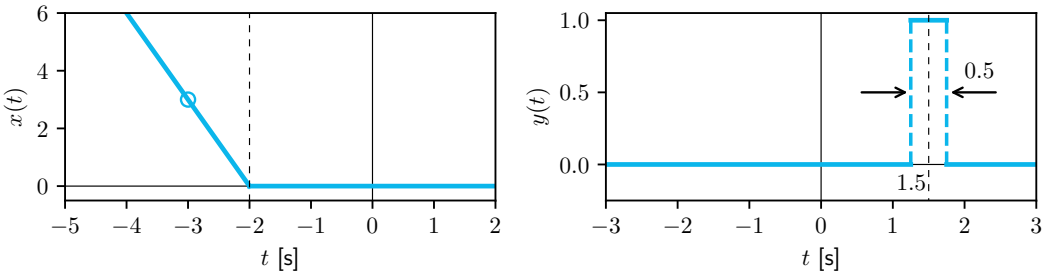


Figure 2.12: Illustrating $x(t) = r(-3t - 6)$ and $y(t) = \Pi(-2t + 3)$ from Examples 2.5 and 2.6.

from 1 to half a second. An alternative way to define the pulse function is to write it as $\Pi\left(\frac{t}{\tau}\right)$, as then *dimensionless* scale factor $\beta = \frac{1}{\tau}$ indicates that the pulse has a width of τ seconds. Then $\Pi(t) = \Pi\left(\frac{t}{1}\right)$ is a unit pulse with a width of one second. In Example 2.6, $y(t)$ can be re-written as $y(t) = \Pi\left(\frac{t-\frac{3}{2}}{\frac{1}{2}}\right)$. One can see at a glance that $y(t)$ is a pulse with a width of half a second, centered at $t = \frac{3}{2}$ s. Similarly, the triangular function can be written as $\Lambda\left(\frac{t}{\tau}\right)$, again with $\beta = \frac{1}{\tau}$, where 2τ is the width of the triangle, as it runs from $t = -\tau$ to $+\tau$ s, see the unit triangular function in Figure B.1. Pulse and triangle widths are defined to be positive, $\tau > 0$.

Note that when these widths are defined using parameters that have a unit, the scaling of the functions is performed using the parameter *value* only. This is because the widths correspond to the scale factor, parameter β in (2.11), which is dimensionless.

2.4.3 Constructing signals using building blocks

With the rules established, we can build new signals using building blocks and even define building blocks using others. E.g., the unit pulse with a width of τ seconds can be expressed:

$$\Pi\left(\frac{t}{\tau}\right) = u(t + \frac{\tau}{2}) - u(t - \frac{\tau}{2}) \tag{2.14}$$

using the unit-step function $u(t)$. An important use of writing a signal using building blocks will appear later, e.g., when we compute the Fourier transform of a signal $x(t)$. Imagine that we could construct $x(t)$ as the summation of two other signals $y(t)$ and $z(t)$, i.e., $x(t) = y(t) + z(t)$. It could well be that, whereas the Fourier transform of $x(t)$ is tedious, the Fourier transforms of $y(t)$ and $z(t)$ are relatively easy. In that case, it is easier to first write the signal as the sum of components, Fourier transform the components, and (using the linearity property of the Fourier transform discussed in Chapter 6), add the Fourier-transformed components.

EX 2.7

What does signal $x(t) = 2u(-t)\cos(2\pi t)$ look like?

Solution This is an operation of the unit step function $u(t)$ multiplied by a 1 Hz cosine function, multiplied by 2. First re-write $u(-t)$:

$$u(-t) = u(-1(t + 0)) \tag{2.15}$$

First step: $\beta < 0$, so the function $u(t)$ is mirrored about $t = 0$ s. Second step: $|\beta| = 1$, which has no effect. Third step: $\frac{\alpha}{\beta} = t_0 = 0$, which again has no effect. We multiply the 'flipped' unit step by the cosine, with amplitude equal to 2. The resulting signal is shown in Figure 2.13 (left).

EX 2.8

What does signal $y(t) = -2r(t) + r(t-4) + r(t+4)$ look like?

Solution This is a summation of three (operations of) unit ramp functions $r(t)$. Apply the check stated above to determine at what times the three components have their 'bend'. First component: at $t = 0$ s; second component: at $t = 4$ s; third component: at $t = -4$ s. Rewrite $y(t)$ to put them in their order of appearance:

$$y(t) = r(t+4) - 2r(t) + r(t-4) \tag{2.16}$$

The first component is a ramp function that is zero before $t = -4$ s and then rises by 1 unit every second. The second ramp function is zero before $t = 0$ s and then descends by 2 units every second. The third ramp function is zero before $t = 4$ s and then rises by 1 unit every second. The signal is shown in Figure 2.13 (right).

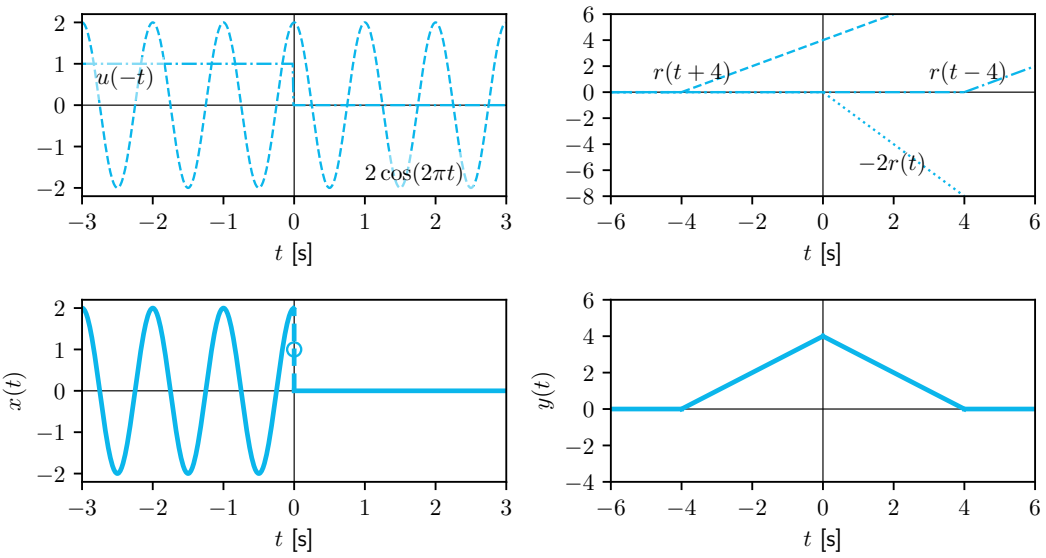


Figure 2.13: Constructing $x(t) = 2u(-t)\cos(2\pi t)$ (left) and $y(t) = -2r(t) + r(t-4) + r(t+4)$ (right) using building blocks (top row), from Examples 2.7 and 2.8.

2.5 Energy and power

2.5.1 Definitions

An important classification of signals used throughout this book is that of energy signals and power signals. For an arbitrary real-valued signal $x(t)$, its (total) *energy* is defined as:

$$E = \lim_{T\to\infty} \int_{-\frac{T}{2}}^{+\frac{T}{2}} x^2(t)\,\mathrm{d}t \tag{2.17}$$

Suppose the unit of the signal is [unit], then the energy of that signal has unit [unit^{2}s].The *average power* of that signal is defined as:

$$P = \lim_{T\to\infty} \frac{1}{T} \int_{-\frac{T}{2}}^{+\frac{T}{2}} x^2(t)\,\mathrm{d}t \tag{2.18}$$

The unit of the power of a signal is [unit^2].[4]

Three signal classes can be identified, as listed in Table 2.2. A signal can be an *energy* signal, a *power* signal, or *neither an energy nor power signal*, with an example of each shown in Figure 2.14. In this figure (and unless noted otherwise, in *all* figures in this book), note that only part of the time axis is shown. These signals range from $t = -\infty$ to $+\infty$.

Table 2.2: Three signal classes.

energy signal	$0 < E < \infty$	$P = 0$
power signal	$E = \infty$	$0 < P < \infty$
signal is neither energy nor power		

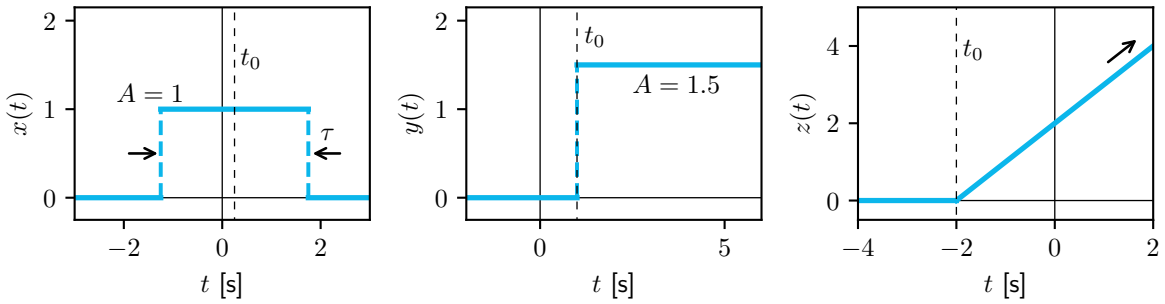


Figure 2.14: Illustrating the three signal classes: energy (left), power (middle), neither energy nor power (right). Here $x(t) = \Pi\left(\frac{t-t_0}{\tau}\right)$, $y(t) = \frac{3}{2}u(t-t_0)$ and $z(t) = r(t+t_0)$, with $t_0 > 0$ an arbitrary time shift.

Clearly, signal $x(t)$ in Figure 2.14 is an energy signal; its energy equals $A^2\tau = \tau$, (2.17), its average power is zero. Signal $y(t)$ is a power signal; it has infinite energy and its average power equals $\frac{A^2}{2} = \frac{9}{8}$, (2.18). Signal $z(t)$ is neither energy nor power, as both its energy, (2.17), and its power, (2.18), are infinite.

Four insights follow from these examples. First, energy and power scale with A^2, so *only* the amplitude matters. Second, shifting signals left or right has no effect on their energy or power: the time shift t_0 in Figure 2.14 is irrelevant. Third, for signals to be either energy or power signals they must be bounded in amplitude. Fourth, signals with a non-zero average, over $t = -\infty$ to ∞, cannot be energy signals.

Periodic signals are an important class of signals, Section 2.1. From the above definition, it is clear that periodic signals are *power* signals: they repeat forever and their energy is infinite. Then, since a periodic signal repeats every period of T_0 seconds, the energy contained in each period is the same, and the average power can be computed *using one period*:

$$P = \frac{1}{T_0}\int_{T_0} x^2(t)\,\mathrm{d}t \tag{2.19}$$

[4] In electrical engineering, where the concepts of energy and power originate, signals are often in [Volt] (or [Ampere]). For power signals, when *normalized to unit resistance* (i.e., 1 Ω, in [Ohm]), and with Ohm's law, the power (2.18) is then expressed in [Watt]. For energy signals, when normalized to unit resistance, the energy (2.17) is expressed in [Joule]=[Watt s]. Section 13.1 provides a more extensive explanation. These units were named after, respectively, the Italian chemist and physicist A.G.A.A. Volta (1745-1827), the French physicist and mathematician A.-M. Ampère (1775-1836), the German physicist G.S. Ohm (1789-1854), the Scottish engineer J. Watt (1736-1819) and British physicist J.P. Joule (1818-1889).

Decibel notation

Energy and power values are often expressed in decibel, dB[5]. It is not a unit but a representation of a numerical value as a (base-10) logarithmic value. To represent a signal's power P in dB, use $P_{dB} = 10 \log_{10} P$. The other way around: $P = 10^{P_{dB}/10}$. The same holds for signal energy. Expressing values in decibel is extremely useful when these values are either very small, or very large. E.g., when a voltage signal has power $P = 0.0000000001$ W ($= 10^{-10}$ W), we write: $P_{dB} = -100$ dBW. The unit remains Watt, the number is expressed in decibel.

2.5.2 Examples

EX 2.9 Consider signal $x(t) = Ae^{-\alpha t}u(t)$, with $\alpha > 0$, where A and α are constants. Determine to what class of signals $x(t)$ belongs, according to Table 2.2.

Solution Note that it is often useful to first sketch the signal, if possible, see Figure 2.15 (left). First compute the energy of the signal, using (2.17):

$$E_x = \lim_{T\to\infty} \int_{-\frac{T}{2}}^{\frac{T}{2}} (Ae^{-\alpha t}u(t))^2 \, \mathrm{d}t = A^2 \lim_{T\to\infty} \int_{0}^{\frac{T}{2}} e^{-2\alpha t} \, \mathrm{d}t = \frac{A^2}{-2\alpha} \lim_{T\to\infty} (e^{-\alpha T} - 1) = \frac{A^2}{2\alpha}$$

Because energy is limited, $x(t)$ is an energy signal. When $\alpha \to 0$ we obtain signal $y(t) = Au(t)$, with infinite energy, see middle graph of Figure 2.15. Its average power can be computed using (2.18):

$$P_y = \lim_{T\to\infty} \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} (Au(t))^2 \, \mathrm{d}t = A^2 \lim_{T\to\infty} \frac{1}{T} \int_{0}^{\frac{T}{2}} 1^2 \, \mathrm{d}t = A^2 \lim_{T\to\infty} \frac{1}{T} (\frac{T}{2} - 0) = \frac{A^2}{2}$$

EX 2.10 Consider signal $z(t) = A\cos(2\pi f_0 t + \theta)$, where A, f_0 and θ are constants. Determine to what class of signals $z(t)$ belongs, according to Table 2.2.

Solution Signal $z(t)$ is periodic (see Figure 2.15 at right); hence it cannot be an energy signal as its energy is infinite. The average power can be computed using either (2.18) or (2.19). Using the latter, with $T_0 = \frac{1}{f_0}$, and using $\cos^2 u = \frac{1}{2}(1 + \cos(2u))$, (A.4):

$$\begin{aligned} P_z &= \frac{1}{T_0} \int_0^{T_0} (A\cos(2\pi f_0 t + \theta))^2 \, \mathrm{d}t = \frac{1}{T_0} A^2 \int_0^{T_0} \frac{1}{2}(1 + \cos(2(2\pi f_0 t + \theta))) \, \mathrm{d}t \\ &= \frac{A^2}{T_0} \left(\frac{T_0}{2} + \frac{1}{8\pi f_0} \sin(4\pi f_0 t + 2\theta) \Big|_0^{T_0} \right) = \frac{A^2}{2} + \frac{A^2}{8\pi} (\sin(4\pi + 2\theta) - \sin(2\theta)) \\ &= \frac{A^2}{2} \qquad \text{since } \sin(4\pi + 2\theta) = \sin(2\theta) \end{aligned}$$

The average power of a sinusoidal signal only depends on its amplitude. Its phase θ *and* frequency f_0 do not matter in energy or power calculations.

[5] After A.G. Bell (1847-1922), a Canadian-American engineer who patented the first practical telephone.

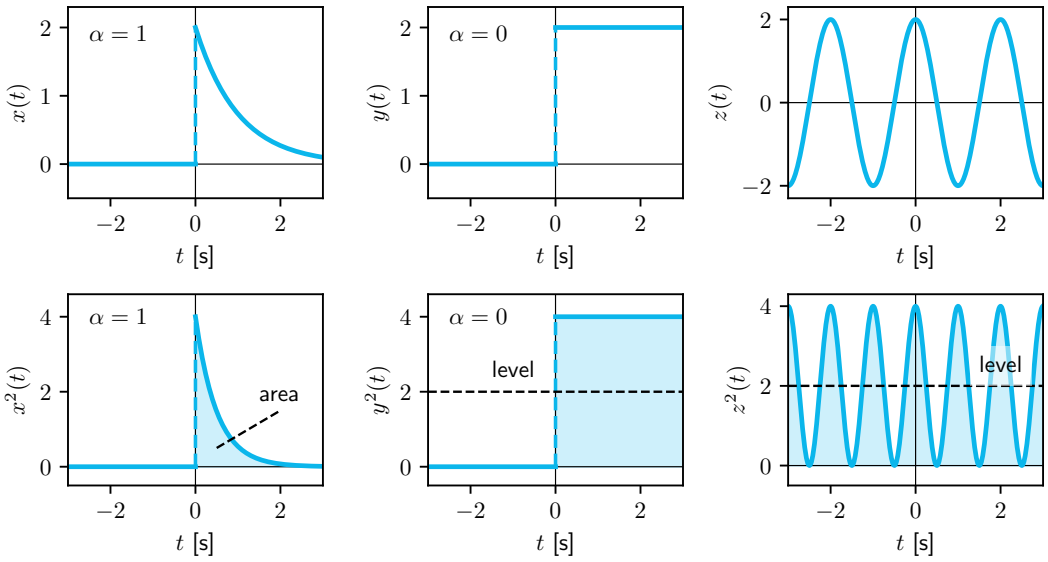


Figure 2.15: Sketching the signals (top row) and their energy or power (bottom) of Example 2.9 and Example 2.10. Here $x(t) = Ae^{-\alpha t}u(t)$ (with $A = 2$, $\alpha = 1$), $y(t) = Ae^{-\alpha t}u(t)$ (with $A = 2$, $\alpha = 0$) and $z(t) = A\cos(2\pi f_0 t + \theta)$ (with $A = 2$,$f_0 = \frac{1}{2}$ and $\theta = 0$).

Figure 2.15 illustrates signals $x(t)$, $y(t)$ and $z(t)$ from Examples 2.9 and 2.10 (for $A = 2$, $\alpha = 1$ and $\alpha = 0$, $f_0 = 0.5$ Hz and $\theta = 0$). Again, in energy and power calculations it is wise to sketch the signal first before doing any calculations. From the sketch it often quickly becomes obvious to what class of signals the signal belongs, and what calculations (if any!) are necessary.

For instance, consider energy signal $x(t)$, shown in Figure 2.14 at left. Suppose we define a new signal, $w(t)$, which equals $x(t)$ plus a non-zero (and possibly negative) constant a. In that case, $w(t)$ is a *power* signal with average power a^2. Whereas $x(t)$ has a zero average, $w(t)$ has a non-zero average, a. Because of this, its energy becomes infinite; its average power is a^2. The pulse becomes irrelevant, as can be seen in the next example.

Compute the average power of signal $w(t) = a + A\Pi(\frac{t}{\tau})$, with a and A non-zero constants, and $\tau > 0$. EX 2.11

Solution

$$P_w = \lim_{T\to\infty}\frac{1}{T}\int_{-\frac{T}{2}}^{+\frac{T}{2}} w^2(t)\,\mathrm{d}t = \lim_{T\to\infty}\frac{1}{T}\int_{-\frac{T}{2}}^{+\frac{T}{2}} (a^2 + A^2\Pi(\frac{t}{\tau}) + 2aA\Pi(\frac{t}{\tau}))\,\mathrm{d}t$$

$$= \lim_{T\to\infty}\frac{1}{T}\left(a^2T + \int_{-\frac{\tau}{2}}^{+\frac{\tau}{2}} (A^2 + 2aA)\,\mathrm{d}t\right) = a^2 + \lim_{T\to\infty}\frac{1}{T}(A^2 + 2aA)\tau = a^2 + 0 = a^2$$

2.6 Practical implications

The concepts defined in this chapter are fundamental to the analysis of signals in the time and frequency domains. Before moving on to the foundations of signal analysis in the frequency domain, the main topic of this book, it is important to briefly look ahead on how signal analysis

is performed in 'real life' and the consequences for the concepts considered.

The most important difference between the mathematical view on signals and what signal analysis entails in real life, is that *measurements do not last forever*. The observation of a signal starts at some time t_1 and ends some later time t_2. Typically we set t_1 to be equal to 0, and then $t_2 - t_1$ equals the measurement duration or time T_{meas}. What does this mean for some of the signal characteristics we discussed up to now? For example, do *even* and *odd* signals exist in real life? Since these characterize the potential symmetry properties of a signal around $t = 0$, see Section 2.2, and we just agreed that measurements in practice *start at* $t = 0$, it is clear that we do not have evenness or oddness of measured signals. These are mathematical abstractions, which offer insight and are helpful in signal analysis, but in most cases not relevant when processing measurements.

Regarding *periodicity*, periodic signals are defined from $t = -\infty$ to $+\infty$, (2.1), but measurements do not last that long. The periodic-ness of signals disappears completely. The measurement $x_w(t)$ that we have of a (potentially) periodic signal $x(t)$, or *any* signal for that matter, is a *window* on the signal, lasting T_{meas} seconds. This is illustrated in Figure 2.16 for measuring a deterministic (top) and random (bottom) signal. The rectangular window is shown in red in Figure 2.16 (top); it equals value 1 during the window, and 0 outside.

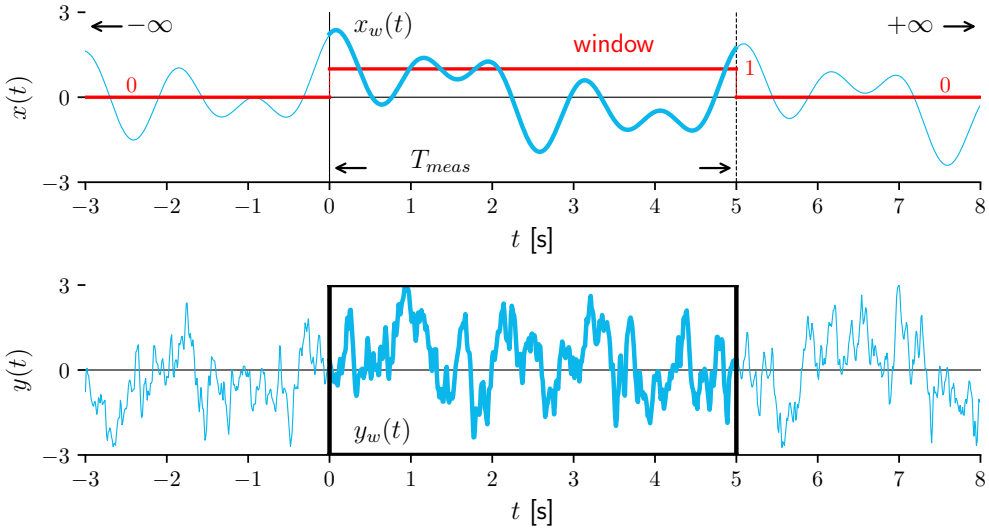


Figure 2.16: The signals of Figure 2.1 considered as measurements of infinite-duration signals (for T_{meas} = 5 s).

Figure 2.16 illustrates that while signals $x(t)$ and $y(t)$ shown in Figure 2.1 (in theory) run from $t = -\infty$ to $+\infty$, after measuring these signals for $T_{meas} = 5$ seconds, *we work with* $x_w(t)$ and $y_w(t)$, the resulting signals after multiplication by the window. This has important consequences: measurements are by definition *energy* signals, since their energy is limited, their average power equals zero. Moving our mathematical concepts and techniques to practical applications means that we have to account for the effects caused by the (inevitable) time window. We return to the subject of finite signal duration in Chapter 8, and to energy and power in this respect in Chapter 13.

This chapter concludes Part I, which introduced the topic of signal analysis and some signal models and definitions, all in the time domain. In Part II, we discuss how periodic and aperiodic signals can be mathematically described in the frequency domain using the Fourier series and Fourier transform, respectively.

II

Continuous time

3

Real Fourier series

In this chapter we perform a signal *de-composition*, expressing a *periodic* real-valued signal as a linear combination of orthogonal basis functions. The Fourier series expansion relies on *sinusoids* as its basis function, and expresses a periodic signal as a sum of harmonic, zero-phase cosines and sines. We first consider the real, or trigonometric, Fourier series; in Chapter 4 the same decomposition but then in terms of complex exponentials will be discussed, the complex Fourier series. The real Fourier series allows us to easily obtain the single-sided amplitude and phase spectra of periodic signals. Chapter 2 showed that obtaining the spectrum of a sinusoid is straightforward. De-composing a periodic signal into a sum of sinusoid components means that the spectrum of that signal is simply the sum of the spectra of these components.

3.1 Rationale and definition

The Fourier series and Fourier transform have been named in honour of French mathematician and physicist Jean-Baptiste Joseph Fourier (1768-1830). Several influential mathematicians played a role in the development, and notably also Carl Friedrich Gauss (1777-1855) [4].

Consider signal $x(t)$ which is periodic with period T_0 and fundamental (linear) frequency $f_0 = \frac{1}{T_0}$ in [Hz]. The fundamental angular frequency is then $\omega_0 = 2\pi f_0 = \frac{2\pi}{T_0}$, in [rad/s]. The rationale of the real Fourier series is to decompose this signal $x(t)$ into a series of K pure sine wave components, each with its own amplitude and with a frequency that is an integer multiple of the fundamental frequency ω_0. Adding these K components together 'builds' the signal $x(t)$, as if we obtain a *model* $\tilde{x}(t)$ of signal $x(t)$ using sinusoidal building blocks:

$$\tilde{x}(t) = \sum_{k=0}^{K-1} (a_k \cos(k\omega_0 t) + b_k \sin(k\omega_0 t))$$

Because $\cos 0 = 1$ and $\sin 0 = 0$, the b_0 component is discarded and a_0 is taken out of the summation, yielding:

$$\tilde{x}(t) = a_0 + \sum_{k=1}^{K-1} (a_k \cos(k\omega_0 t) + b_k \sin(k\omega_0 t))$$

Then, when $x(t)$ adheres to certain conditions (Section 3.3), the model $\tilde{x}(t)$ will approximate signal $x(t)$ for $K \to \infty$. In what follows, the real Fourier series will be defined for this ultimate

case, yet in most examples and all practical applications of the real Fourier series we work with a finite number of components K, corresponding to $2K-1$ coefficients.

The real, or trigonometric Fourier series of signal $x(t)$ (for $K \to \infty$) is written as:

$$\boxed{x(t) = a_0 + \sum_{k=1}^{k=\infty} a_k \cos(k\omega_0 t) + \sum_{k=1}^{k=\infty} b_k \sin(k\omega_0 t)} \tag{3.1}$$

Expanding the summations, the Fourier series becomes:

$$x(t) = a_0 + a_1 \cos(\omega_0 t) + a_2 \cos(2\omega_0 t) + \ldots + b_1 \sin(\omega_0 t) + b_2 \sin(2\omega_0 t) + \ldots \tag{3.2}$$

Periodic signal $x(t)$, for $-\infty < t < \infty$, is built, or constructed from a sum of harmonically related sines and cosines, with the coefficients given by a_0 and $\{a_k, b_k\}$ for $k \in \mathbb{N}^+$. Harmonically related means that the frequencies of the sinusoidal components are all *integer multiples* of the fundamental (angular) frequency ω_0; the frequencies are commensurable and the resulting signal is periodic with T_0. The kth cosine and sine components, that is, the kth harmonic of the fundamental frequency, fit an integer k times in the period T_0. Regarding the *unit* of the coefficients, from (3.1) it follows that they have the same unit [unit] as the signal.

3.2 Derivation Fourier series coefficients

In this section we derive expressions, in terms of periodic signal $x(t)$, for the coefficients a_0, a_k and b_k, for $k = 1, 2, 3, \ldots$ in the real Fourier series.

First, coefficient a_0 is found by integrating all terms of (3.2) over one period:

$$\begin{aligned} \int_{T_0} x(t)\,\mathrm{d}t \;=\; & \int_{T_0} a_0\,\mathrm{d}t \;+ \int_{T_0} a_1 \cos(\omega_0 t)\,\mathrm{d}t + \int_{T_0} a_2 \cos(2\omega_0 t)\,\mathrm{d}t + \ldots \\ & + \int_{T_0} b_1 \sin(\omega_0 t)\,\mathrm{d}t + \int_{T_0} b_2 \sin(2\omega_0 t)\,\mathrm{d}t + \ldots \end{aligned}$$

Now, all terms on the right-hand side, except the first one, are zero, as for each integral term we integrate a sine or cosine over one period T_0. For any $\cos(k\omega_0 t)$ and $\sin(k\omega_0 t)$, with integer k, as much 'area' appears above the time axis as below it, in one period T_0, and the integrals are zero. We arrive at $\int_{T_0} x(t)\,\mathrm{d}t = a_0 T_0$, and therefore:

$$\boxed{a_0 = \frac{1}{T_0} \int_{T_0} x(t)\,\mathrm{d}t} \tag{3.3}$$

Hence, a_0 is simply the *average value* of the signal. Another interpretation of a_0 is that it represents a cosine with zero frequency, i.e., an ultimately slowly varying cosine signal. Note that the rest of the terms in (3.1) are oscillations about this average.

A general expression for the coefficients a_k, for $k \in \mathbb{N}^+$, is found by multiplying both sides of (3.1) by $\cos(m\omega_0 t)$, for $m \in \mathbb{N}^+$, and then integrating over one period T_0:

$$\begin{aligned} \int_{T_0} x(t)\cos(m\omega_0 t)\,\mathrm{d}t \;=\; & a_0 \int_{T_0} \cos(m\omega_0 t)\,\mathrm{d}t \\ & + \int_{T_0} \left(\sum_{k=1}^{k=\infty} a_k \cos(k\omega_0 t) \right) \cos(m\omega_0 t)\,\mathrm{d}t \\ & + \int_{T_0} \left(\sum_{k=1}^{k=\infty} b_k \sin(k\omega_0 t) \right) \cos(m\omega_0 t)\,\mathrm{d}t \end{aligned}$$

The first term on the right-hand side is zero. With the two other terms, we move the term $\cos(m\omega_0 t)$ inside the summation and integrate term-by-term. Essentially, we interchange the

order of summation and integration as follows:

$$\int_{T_0} x(t)\cos(m\omega_0 t)\,\mathrm{d}t = \sum_{k=1}^{k=\infty} a_k \int_{T_0} \cos(k\omega_0 t)\cos(m\omega_0 t)\,\mathrm{d}t + \sum_{k=1}^{k=\infty} b_k \int_{T_0} \sin(k\omega_0 t)\cos(m\omega_0 t)\,\mathrm{d}t$$

Applying the *orthogonality* properties of integrals involving products of sines and cosines, (A.11)-(A.13), results in each term in the second series on the right-hand side to be zero, for all values of k and m (using integral I_3, (A.13)). Each term in the first series on the right-hand side is also zero, except for the one term where k equals m, where the integral equals $T_0/2$ (integral I_2, (A.12)). We obtain, with $k = m$:

$$\int_{T_0} x(t)\cos(m\omega_0 t)\,\mathrm{d}t = \int_{T_0} x(t)\cos(k\omega_0 t)\,\mathrm{d}t = a_k\left(\frac{T_0}{2}\right)$$

which leads to our main result:

$$a_k = \frac{2}{T_0}\int_{T_0} x(t)\cos(k\omega_0 t)\,\mathrm{d}t \quad \text{with } k \in \mathbb{N}^+ \tag{3.4}$$

Similarly, the values for coefficients b_k are found, by multiplying the series (3.1) by $\sin(m\omega_0 t)$, integrating it over one period T_0 and using the integral properties I_1 and I_3:

$$b_k = \frac{2}{T_0}\int_{T_0} x(t)\sin(k\omega_0 t)\,\mathrm{d}t \quad \text{with } k \in \mathbb{N}^+ \tag{3.5}$$

One way to interpret integrals (3.4) and (3.5), is that for each component (representing the kth harmonic of the fundamental frequency ω_0) we consider the *commonality* of signal $x(t)$ *with that component* (harmonic). For instance, if signal $x(t)$ is dominated by the kth cosine component in the Fourier series, integral (3.4) and therefore the corresponding a_k coefficient will be larger than other coefficients. Apparently, this particular building block is essential to construct $x(t)$. Conversely, when signal $x(t)$ has *nothing* in common with the kth sine component in the Fourier series, the corresponding b_k coefficient will be *zero*: this building block is not needed to construct signal $x(t)$.

3.3 Conditions

For expanding a signal $x(t)$ in a Fourier series, this signal must be *absolutely integrable* over the finite interval $(-\frac{T_0}{2}, \frac{T_0}{2})$, i.e., over one period:

$$\int_{-\frac{T_0}{2}}^{\frac{T_0}{2}} |x(t)|\,\mathrm{d}t < \infty$$

In this case the coefficients a_k and b_k have finite values, see [5].

If signal $x(t)$ has only a finite number of maxima and minima and a finite number of (finite jump) discontinuities in the interval $(-\frac{T_0}{2}, \frac{T_0}{2})$ (that is, signal $x(t)$ is of *bounded variation*), then the Fourier series, for $K \to \infty$, *converges* to $x(t)$ for all values of t, and to the 'average' of the two values forming a discontinuity, see [5]. In this case, i.e., for indices $k \to \infty$, the coefficients a_k and b_k will go to zero.

These conditions are sufficient, implying that they are more restrictive than necessary. Most practical periodic signals do not contain sudden jumps and the Fourier series gives an adequate approximation of the signal when using a sufficient finite number of components K.

3.4 Example: Fourier series of a square wave

EX 3.1

Compute the real Fourier series of the periodic square wave signal $x(t)$, given as:

$$x(t) = \begin{cases} 1 & \text{for } 0 \le t < \frac{T_0}{2} \\ -1 & \text{for } \frac{T_0}{2} \le t < T_0 \end{cases} \tag{3.6}$$

and periodically extended outside this interval, as shown in red in Figure 3.1 at left.

Solution Expressing $x(t)$ as a Fourier series requires the calculation of coefficients a_0, a_k and b_k. The first coefficient a_0 is found with (3.3):

$$a_0 = \frac{1}{T_0}\int_{T_0} x(t)\,\mathrm{d}t = \frac{1}{T_0}\left(\int_0^{\frac{T_0}{2}} 1\,\mathrm{d}t + \int_{\frac{T_0}{2}}^{T_0} (-1)\,\mathrm{d}t\right)$$

The integral is taken over period T_0; we have full freedom to set the lower and upper integral bounds, as long as exactly one period is taken. Here we choose to integrate over the interval $[0, T_0]$:

$$a_0 = \frac{1}{T_0}\left([t]_0^{\frac{T_0}{2}} + [-t]_{\frac{T_0}{2}}^{T_0}\right) = \frac{1}{T_0}\left(\frac{T_0}{2} - 0 - T_0 + \frac{T_0}{2}\right) = 0 \tag{3.7}$$

This is no surprise as the square wave $x(t)$ has a zero mean, see Figure 3.1.

The coefficients a_k are found with (3.4):

$$\begin{aligned} a_k &= \frac{2}{T_0}\int_{T_0} x(t)\cos(k\omega_0 t)\,\mathrm{d}t = \frac{2}{T_0}\left(\int_0^{\frac{T_0}{2}} \cos(k\omega_0 t)\,\mathrm{d}t - \int_{\frac{T_0}{2}}^{T_0} \cos(k\omega_0 t)\,\mathrm{d}t\right) \\ &= \frac{2}{T_0}\left(\left[\frac{1}{k\omega_0}\sin(k\omega_0 t)\right]_0^{\frac{T_0}{2}} - \left[\frac{1}{k\omega_0}\sin(k\omega_0 t)\right]_{\frac{T_0}{2}}^{T_0}\right) \end{aligned}$$

and with $\omega_0 = \frac{2\pi}{T_0}$ we obtain:

$$\begin{aligned} a_k &= \frac{2}{T_0}\frac{T_0}{k2\pi}\left(\sin(k\frac{2\pi}{T_0}\frac{T_0}{2}) - \sin(k\frac{2\pi}{T_0}0) - \sin(k\frac{2\pi}{T_0}T_0) + \sin(k\frac{2\pi}{T_0}\frac{T_0}{2})\right) \\ &= \frac{1}{k\pi}\left(\sin(k\pi) - \sin 0 - \sin(k2\pi) + \sin(k\pi)\right) \;=\; 0 \end{aligned} \tag{3.8}$$

All cosine coefficients $a_k = 0$, for $k \in \mathbb{N}^+$. This is also not a surprise, since the square wave is odd-symmetric about $t = 0$, see Table 2.1, and hence an even function such as a cosine is of no use at all to construct this square wave.

The coefficients b_k are found using (3.5):

$$\begin{aligned} b_k &= \frac{2}{T_0}\int_{T_0} x(t)\sin(k\omega_0 t)\,\mathrm{d}t = \frac{2}{T_0}\left(\int_0^{\frac{T_0}{2}} \sin(k\omega_0 t)\,\mathrm{d}t - \int_{\frac{T_0}{2}}^{T_0} \sin(k\omega_0 t)\,\mathrm{d}t\right) \\ &= \frac{2}{T_0}\left(\left[-\frac{1}{k\omega_0}\cos(k\omega_0 t)\right]_0^{\frac{T_0}{2}} + \left[\frac{1}{k\omega_0}\cos(k\omega_0 t)\right]_{\frac{T_0}{2}}^{T_0}\right) \end{aligned}$$

and again with $\omega_0 = \frac{2\pi}{T_0}$ we obtain:

$$\begin{aligned} b_k &= \frac{2}{T_0}\frac{T_0}{k2\pi}\left(-\cos(k\frac{2\pi}{T_0}\frac{T_0}{2}) + \cos(k\frac{2\pi}{T_0}0) + \cos(k\frac{2\pi}{T_0}T_0) - \cos(k\frac{2\pi}{T_0}\frac{T_0}{2})\right) \\ &= \frac{1}{k\pi}\left(-\cos(k\pi) + \cos 0 + \cos(k2\pi) - \cos(k\pi)\right) = \frac{2}{k\pi}\left(1 - \cos(k\pi)\right) \end{aligned}$$

Note that $\cos(k\pi) = 1$ for integer k even, and $\cos(k\pi) = -1$ for integer k odd. Hence:

$$b_k = \begin{cases} \frac{4}{k\pi} & \text{for } k \text{ odd} \\ 0 & \text{for } k \text{ even} \end{cases} \tag{3.9}$$

with $k \in \mathbb{N}^+$. All even-indexed coefficients are zero, a consequence of this signal being half-wave odd symmetric. This will be discussed in more detail in Section 3.5.

With the above coefficients we can write the Fourier series for the square wave as:

$$x(t) = \frac{4}{\pi}\left(\frac{1}{1}\sin(\omega_0 t) + \frac{1}{3}\sin(3\omega_0 t) + \frac{1}{5}\sin(5\omega_0 t) + \ldots\right) \tag{3.10}$$

The left-hand side of Figure 3.1 shows the square wave (in red), period $T_0 = 1$ s, and the real Fourier series approximation (in blue) for $K = 5$ (top) and $K = 11$ (bottom).

Figure 3.1 at right shows the Fourier series coefficients $a_0, \{a_k, b_k\}$, as a function of k on the horizontal axis, representing angular frequency $k\omega_0$. In other words, the horizontal axis represents *frequency* and the coefficients $a_0, \{a_k, b_k\}$ are shown *as a function of frequency*. When adding more components, the signal approximation becomes better (compare the graph at bottom ($K = 11$) with the one on top ($K = 5$)). Adding components (i.e., higher-frequency building blocks) has no effect on the coefficients of components already computed. Typically, the coefficients become smaller for larger k (for this signal, see (3.9)), and reduce to 0 when $k \to \infty$.

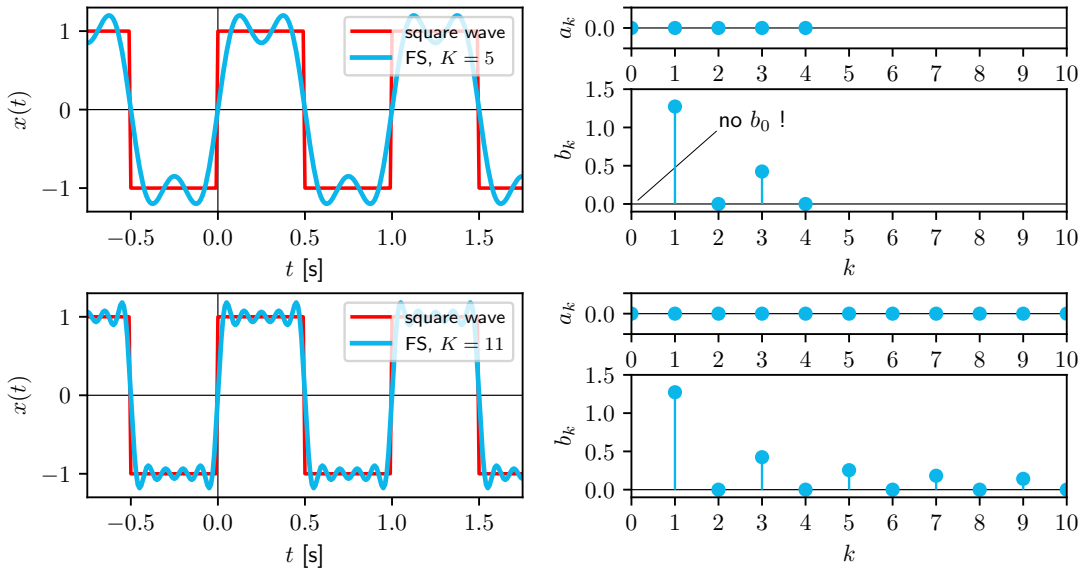


Figure 3.1: Left: square wave signal $x(t)$ and its Fourier series approximation. Right: the $2K - 1$ Fourier series coefficients $a_0, \{a_k, b_k\}$ as a function of k, for $K = 5$ (top row) and $K = 11$ (bottom row).

3.5 Effects of symmetry

When considering the Fourier series coefficients of the square wave signal, Example 3.1, some patterns emerged. All cosine coefficients were shown to be zero, an effect caused by the oddness of the square wave. All even-indexed coefficients were zero as well. These patterns can be explained using the three symmetry properties introduced in Section 2.2.

As shown in (2.5), all signals can be described as the summation of an even and odd signal part: $x(t) = x_e(t) + x_o(t)$. The same holds for the real Fourier series, (3.1):

$$x(t) = \underbrace{a_0 + \sum_{k=1}^{k=\infty} a_k \cos(k\omega_0 t)}_{\text{even part } x_e(t)} + \underbrace{\sum_{k=1}^{k=\infty} b_k \sin(k\omega_0 t)}_{\text{odd part } x_o(t)}$$

When considering the Fourier series of arbitrary periodic signals from this perspective, it is clear that for even signals all sine components are zero, whereas for odd signals all cosine components are zero. When signals are neither even nor odd, *both* sine and cosine components are required to construct the signal. Note that the signal average, a_0, belongs to the even part as the odd part must be zero for $t = 0$, see Table 2.1.

Finally, it can be shown (see Exercise *e*.3-7) that for signals which have half-wave odd symmetry (Section 2.2) all the *even-indexed* coefficients ($k = 2, 4, 6, \ldots$) are zero.

EX 3.2

What would happen to the Fourier series coefficients of Example 3.1 if we were to shift the square wave $x(t)$ a quarter of a period ($\frac{T_0}{4}$) to the left?

Solution From visual inspection of Figure 3.1, we learn that the shifted signal is even, and hence all sine coefficients will be zero. In this way the signal can be constructed with cosines. The signal remains half-wave odd, as time shifts have no effect on this symmetry property, and all even-indexed coefficients remain zero.

EX 3.3

What would happen to the Fourier series coefficients of Example 3.1 if we were to add a non-zero constant c to the square wave $x(t)$? That is, what are the Fourier series coefficients $a_0', \{a_k', b_k'\}$ of periodic signal $y(t) = x(t) + c$?

Solution Strictly mathematically, $y(t)$ is neither odd nor half-wave odd. When comparing the real Fourier series coefficients of $x(t)$ and $y(t)$, the *only* difference will occur in the average: $a_0' = a_0 + c = 0 + c = c$. All other coefficients will *remain the same*: $a_k' = a_k$ and $b_k' = b_k$, for $k \in \mathbb{N}^+$. The Fourier series coefficients will show the same patterns as those obtained for the original signal $x(t)$: all cosine coefficients are zero, all even-indexed coefficients are zero. As discussed in Section 2.2, non-zero-mean signals can have characteristics of oddness and half-wave oddness 'about their average'.

3.6 Single-sided spectrum

The real Fourier series gives an array of real numbers (a_0, and $\{a_k, b_k\}$ for $k \in \mathbb{N}^+$), which can be converted to *single-sided* amplitude and phase spectra.

3.6.1 Derivation

Consider the kth harmonic of the Fourier series ($k \neq 0$), and suppose we want to express it as a single cosine function with amplitude A_k and phase θ_k:

$$a_k \cos(k\omega_0 t) + b_k \sin(k\omega_0 t) = A_k \cos(k\omega_0 t + \theta_k)$$

Applying $\cos(u+v) = \cos u \cos v - \sin u \sin v$, (A.7), to the right-hand side yields:

$$a_k \cos(k\omega_0 t) + b_k \sin(k\omega_0 t) = A_k \cos(k\omega_0 t) \cos\theta_k - A_k \sin(k\omega_0 t) \sin\theta_k$$

Thus:

$$a_k = A_k \cos\theta_k \quad \text{and} \quad b_k = -A_k \sin\theta_k \tag{3.11}$$

which allows us to obtain the amplitude A_k of the kth harmonic:

$$\boxed{A_k = \sqrt{a_k^2 + b_k^2} \quad \text{with } k \in \mathbb{N}^+} \tag{3.12}$$

and phase θ_k of the kth harmonic:[1]

$$\boxed{\theta_k = \arctan\left(-\frac{b_k}{a_k}\right) \quad \text{with } k \in \mathbb{N}^+} \tag{3.13}$$

For the signal average, the 'zero-frequency' Fourier series component ($k = 0$): $A_0 = |a_0|$. When $a_0 > 0$, the average lies on the positive real axis of the complex plane, $\theta_0 = 0$. When $a_0 < 0$, the average lies on the negative real axis, $\theta_0 = \pm\pi$. When $a_0 = 0$, $\theta_0 = 0$.

EX 3.4

Draw the single-sided spectrum of the square wave signal of Example 3.1.

Solution Plotting the K amplitudes A_k and phases θ_k as a function of k, for all values of $k = 0, 1, 2, \ldots, K-1$, we obtain the single-sided amplitude and phase spectra, respectively. Figure 3.2 at left illustrates a_0 and coefficients a_k and b_k (for $K = 17$); at right are shown the single-sided amplitude (A_k, top) and phase (θ_k, bottom) spectra of the square wave of Example 3.1. As only discrete frequencies occur in the spectra, all integer multiples of ω_0, these spectra are known as *line spectra*.

All even-indexed coefficients are zero, and so are their amplitudes and phases. The odd-indexed b_k coefficients are non-zero (for finite K), their phases are all equal to $-\frac{\pi}{2}$ as they represent 'pure' sines. Remember from Chapter 2: the *cosine* is our reference, a zero-phase sine is a 'cosine with phase equal to $-\frac{\pi}{2}$', $\sin u = \cos(u - \frac{\pi}{2})$.

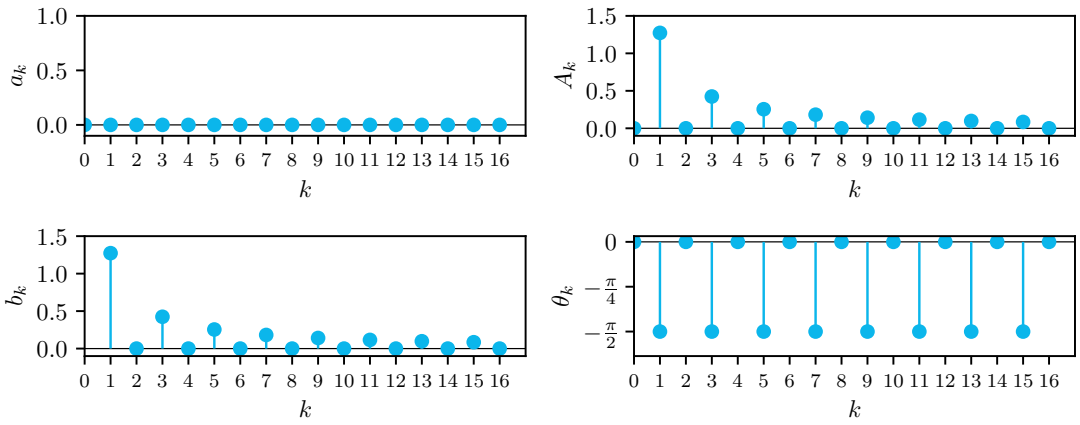


Figure 3.2: Signal $x(t)$; first $K = 17$ Fourier series coefficients (left column); single-sided amplitude and phase spectra (right column).

[1] Numerically, e.g., in Python, you need to use the four-quadrant arctan function.

3.6.2 Alternative definition of real Fourier series

With A_k and θ_k defined, an alternative expression for the real Fourier series, (3.1) reads:

$$x(t) = a_0 + \sum_{k=1}^{k=\infty} A_k \cos(k\omega_0 t + \theta_k) \tag{3.14}$$

This expression shows the Fourier series directly as single-sided amplitude and phase spectra. It allows for easy analysis of what happens to a signal spectrum when shifting that signal in time. Define new signal $y(t) = x(t + t_0)$ with $t_0 > 0$ some time shift. Substituting $t + t_0$ for t in (3.14) yields the amplitudes A'_k and phases θ'_k of the single-sided spectrum of $y(t)$:

$$A'_k = A_k \quad \text{(unchanged)} \quad \text{and} \quad \theta'_k = \theta_k + k\omega_0 t_0 \tag{3.15}$$

Clearly, the amplitudes remain unaffected by the time shift. Recall from Section 2.1 that, for a sinusoid with frequency ω_0, a time shift t_0 corresponds to a phase shift $\theta = \omega_0 t_0$, (2.3). Hence, the phase of the first harmonic, the fundamental frequency, increases with $\omega_0 t_0$, and the phase of the kth harmonic increases with k times this phase shift, $k\omega_0 t_0$. Using (3.11), the real Fourier series coefficients $\{a'_k, b'_k\}$ of new signal $y(t)$ can be computed as:

$$a'_k = a_k \cos(k\omega_0 t_0) + b_k \sin(k\omega_0 t_0) \quad \text{and} \quad b'_k = -a_k \sin(k\omega_0 t_0) + b_k \cos(k\omega_0 t_0)$$

The average remains unaffected by the time shift; $a'_0 = a_0$, $A'_0 = A_0$ and $\theta'_0 = \theta_0$.

EX 3.5

Given (3.14), reconsider Example 3.2, in other words, what happens to the Fourier series coefficients and single-sided spectrum when shifting signal $x(t)$ by $\frac{T_0}{4}$ to the left?

Solution Since $y(t) = x(t + \frac{T_0}{4})$, $t_0 = \frac{T_0}{4}$; with $\omega_0 = \frac{2\pi}{T_0}$, $k\omega_0 t_0 = k\frac{2\pi}{T_0}\frac{T_0}{4} = k\frac{\pi}{2}$. Hence:

$$A'_k = A_k \quad \text{and} \quad \theta'_k = \theta_k + k\frac{\pi}{2}$$

The amplitude spectrum remains the same; the average remains zero. The even-indexed amplitudes and phases remain zero. The phase of the first harmonic ($k = 1$) increases with $\frac{\pi}{2}$, a quarter of a wavelength, corresponding to $\frac{\frac{\pi}{2}}{2\pi}T_0 = \frac{T_0}{4}$, a quarter of the period, see (2.3). The phases of the other odd-indexed kth harmonics increase with k times this phase shift: $k\frac{\pi}{2}$. The coefficients of $y(t)$ become (recall, in Example 3.1 we found for signal $x(t)$ that $a_k = 0\ \forall\ k$):

$$a'_k = b_k \sin(k\frac{\pi}{2}) \quad \text{and} \quad b'_k = b_k \cos(k\frac{\pi}{2})$$

Since $b_k = 0$ for even indices k, and $\cos(k\frac{\pi}{2}) = 0$ for odd indices k, we conclude that $b'_k = 0\ \forall\ k$: signal $y(t)$ is even, no sines are needed. Hence, using (3.9):

$$a'_k = \begin{cases} +\frac{4}{k\pi} & \text{for } k = 1, 5, 9, \ldots \\ -\frac{4}{k\pi} & \text{for } k = 3, 7, 11, \ldots \\ 0 & \text{for } k \text{ even} \end{cases}$$

The half-wave odd symmetry is maintained. Figure 3.3 shows the first $K = 17$ Fourier series coefficients (left) and single-sided amplitude and phase spectra (right) of $y(t)$. The signal $y(t)$ is even, and all sines in $x(t)$ (3.10) become cosines in $y(t)$, with positive coefficients a'_k for $k = 1, 5, 9, \ldots$ and negative coefficients for $k = 3, 7, 11, \ldots$.

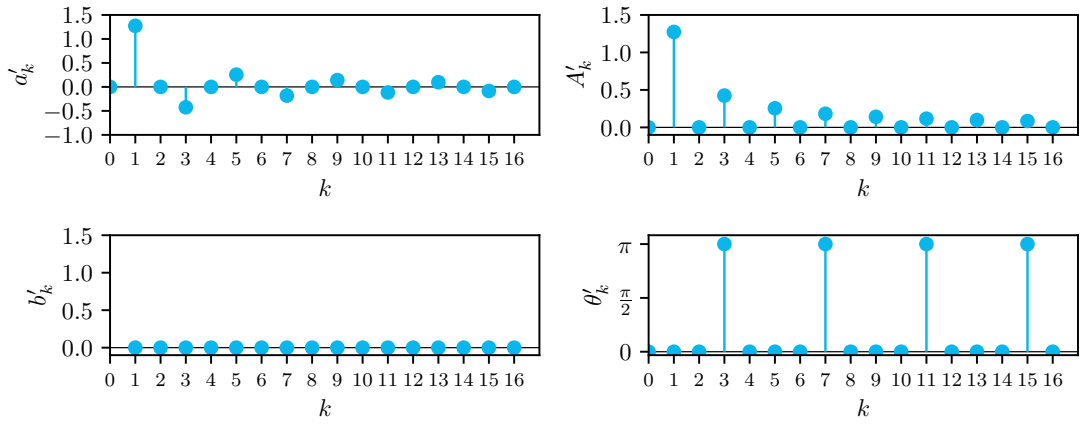


Figure 3.3: Signal $y(t) = x(t + \frac{T_0}{4})$; first $K = 17$ Fourier series coefficients (left column); single-sided amplitude and phase spectra (right column).

This is because for $k = 1, 5, 9, \ldots$ the sines become positive cosines ($k = 1$: $\sin(1\omega_0 t + \frac{\pi}{2}) = \cos(1\omega_0 t + \frac{\pi}{2} - \frac{\pi}{2}) = \cos(\omega_0 t)$; $k = 5$: $\sin(5\omega_0 t + 5\frac{\pi}{2}) = \cos(5\omega_0 t + 5\frac{\pi}{2} - \frac{\pi}{2}) = \cos(5\omega_0 t + 2\pi) = \cos(5\omega_0 t)$). For $k = 3, 7, 11, \ldots$ the sines become negative cosines ($k = 3$: $\sin(3\omega_0 t + 3\frac{\pi}{2}) = \cos(3\omega_0 t + 3\frac{\pi}{2} - \frac{\pi}{2}) = \cos(3\omega_0 t + \pi)$, etc.).
Hence, in the phase spectrum, when limiting phases to $[-\pi, \pi]$, the phases of the odd-indexed coefficients become either zero (for $k = 1, 5, 9, \ldots$) or π (for $k = 3, 7, 11, \ldots$).

EX 3.6

Given $x(t)$ from Example 3.1, elaborate on the coefficients $\{a''_k, b''_k\}$ and spectrum $\{A''_k, \theta''_k\}$ of signal $z(t) = x(t + \frac{T_0}{3})$, which is neither odd nor even, Figure 3.4.

Similar to Figure 3.3, shifting a signal in time causes the coefficients $\{a_k, b_k\}$ to change. Considering the spectrum, the amplitude spectrum remains the same ($A''_k = A_k$). Effects of a time shift become visible *only* in the phase spectrum. For this signal, which is neither odd nor even, the a''_k, b''_k and θ''_k graphs become more irregular, which is the default in practical applications. The signal remains half-wave odd.

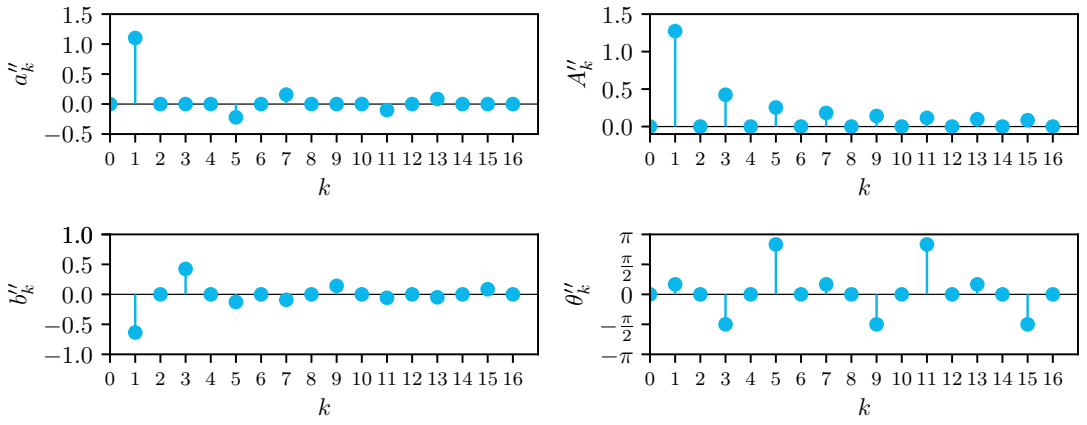


Figure 3.4: Signal $z(t) = x(t + \frac{T_0}{3})$; first $K = 17$ Fourier series coefficients (left column); single-sided amplitude and phase spectra (right column).

EX 3.7

Figure 3.5 shows the Fourier series coefficients $a_0, \{a_k, b_k\}$ of periodic signal $v(t)$. Describe $v(t)$ in mathematical terms.

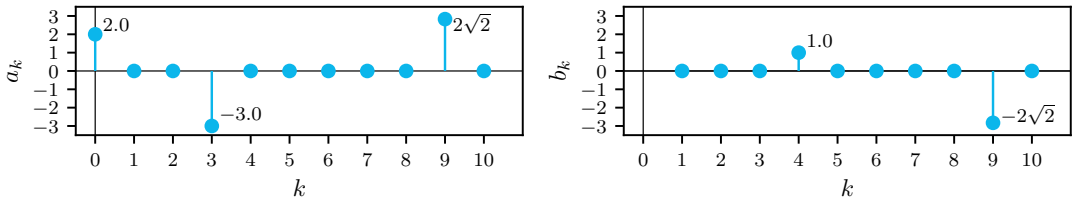


Figure 3.5: Fourier series coefficients of signal $v(t)$.

Solution The non-zero coefficients can be read from the figure: $a_0 = 2$, $a_3 = -3$, $a_9 = 2\sqrt{2}$, $b_4 = 1$ and $b_9 = -2\sqrt{2}$. The average of the signal is 2, we have a pure negative cosine with amplitude 3 at $3\omega_0$ ($A_3 = 3, \theta_3 = \pi$), a pure sine with amplitude 1 at $4\omega_0$ ($A_4 = 1, \theta_4 = -\frac{\pi}{2}$) and a shifted cosine at $9\omega_0$. With (3.12) and (3.13), respectively, the amplitude and phase at the latter frequency yield $A_9 = \sqrt{8+8} = 4, \theta_9 = \arctan(1) = \frac{\pi}{4}$. Signal $v(t)$ can be written as:

$$v(t) = 2 + 3\cos(3\omega_0 t + \pi) + 1\sin(4\omega_0 t) + 4\cos(9\omega_0 t + \frac{\pi}{4})$$

The sine component can also be written as $1\cos(4\omega_0 - \frac{\pi}{2})$. A spectrum like Figure 3.5 does *not* provide information on ω_0, which remains unknown.

EX 3.8

Figure 3.6 shows the amplitude and phase spectrum $\{A_k, \theta_k\}$ of periodic signal $w(t)$ that has period $\frac{1}{6}$ s. Describe $w(t)$ in mathematical terms.

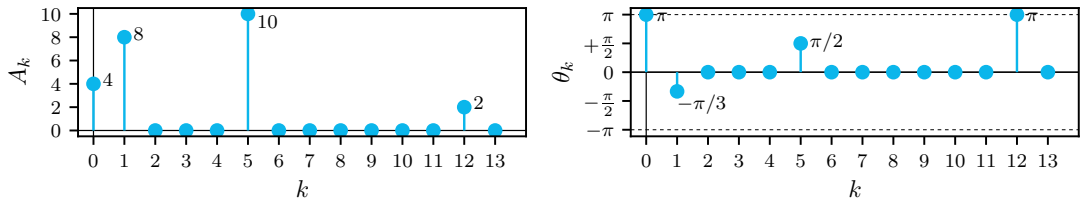


Figure 3.6: Single-sided amplitude and phase spectra of signal $w(t)$.

Solution The fundamental frequency $\omega_0 = 12\pi$ rad/s. Signal $w(t)$ can be written as:

$$\begin{aligned} w(t) &= 4\cos(0t + \pi) + 8\cos(12\pi t - \frac{\pi}{3}) + 10\cos(60\pi t + \frac{\pi}{2}) + 2\cos(144\pi t + \pi) \\ &= -4 + 8\cos(12\pi t - \frac{\pi}{3}) - 10\sin(60\pi t) - 2\cos(144\pi t) \end{aligned}$$

3.6.3 Diagram of coefficients

With reference to (3.12) and (3.13), the relation between the $2K-1$ real Fourier series coefficients and the single-sided spectrum of the K components is shown in Figure 3.7. Throughout this book, similar diagrams, e.g., for the complex exponential Fourier series, will be included.

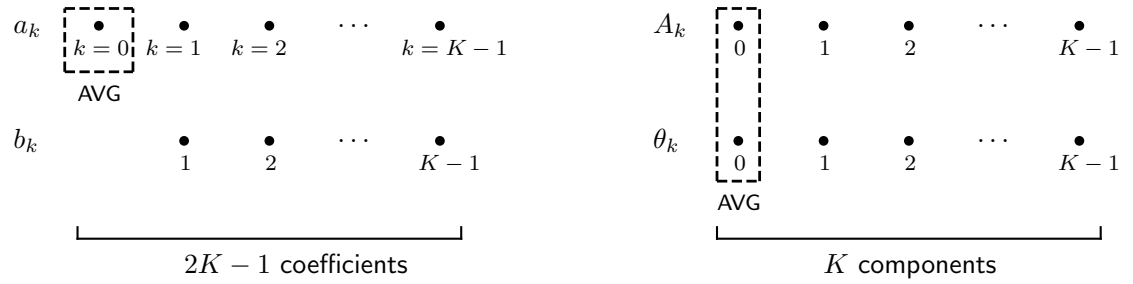


Figure 3.7: Real Fourier series coefficients (left) and the single-sided amplitude and phase spectra (right), they are related through (3.12) and (3.13) In this diagram, AVG means average.

3.7 Parseval: real Fourier series

Here we deal with periodic signals, which are by definition *power* signals, since their energy is infinite. Power can be calculated in time, but *also* in frequency, using *Parseval's theorem*:[2]

$$P = \frac{1}{T_0}\int_{t_0}^{t_0+T_0} x^2(t)\,\mathrm{d}t = a_0^2 + \frac{1}{2}\sum_{k=1}^{\infty}(a_k^2+b_k^2) = A_0^2 + \frac{1}{2}\sum_{k=1}^{\infty}A_k^2 \tag{3.16}$$

using (3.12) and computing the average power using one period T_0, as in (2.19). The power of the signal average equals $a_0^2 = A_0^2$ and the power of the kth harmonic in the Fourier series equals $\frac{1}{2}(a_k^2+b_k^2) = \frac{1}{2}A_k^2$. These follow from the discussion in Section 2.5, see Example 2.10.

As shown in (3.16), *only* the amplitudes matter, and the phases and frequencies are irrelevant when computing signal power. The amplitude spectrum contains *all* the information required to compute the signal power. Finally, note that to compute a signal's power, we need to include *all* components of the Fourier series, i.e., $K \to \infty$, see the following example.

EX 3.9

Compute the average power of the square wave $x(t)$ of Example 3.1 using (3.16). Assume that $x(t)$ is a voltage signal, its unit is [V].

Solution The square wave average power can be easily obtained in time:

$$P_x = \frac{1}{T_0}\left((1)^2\frac{T_0}{2} + (-1)^2\frac{T_0}{2}\right) = 1\ \text{W}$$

Half of the period the signal value equals $+1$, and half of the period it equals -1. Calculating the average power in frequency using Parseval's theorem is more complicated. Time shifts do not affect the amplitude spectrum, the only thing that matters in the power calculation, so we can use the $a_0, \{a_k, b_k\}$ coefficients calculated in Example 3.1, with (3.12), to directly obtain the amplitude spectrum A_k:

$$A_k = \begin{cases} \frac{4}{k\pi} & \text{for } k \text{ odd} \\ 0 & \text{for } k \text{ even} \end{cases}$$

with $k \in \mathbb{N}^+$ and $A_0 = a_0 = 0$. Then, using (3.16) we get:

$$P_x = A_0^2 + \frac{1}{2}\sum_{k=1}^{\infty}A_k^2 = 0^2 + \frac{1}{2}\frac{16}{\pi^2}\sum_{\ell=1}^{\infty}\frac{1}{(2\ell-1)^2} = \frac{8}{\pi^2}\frac{\pi^2}{8} = 1\ \text{W}$$

[2] After the French mathematician M.-A. Parseval (1755-1836).

Here, we first expressed the series in its odd-indexed coefficients only and then used (A.22), a result from power series [6]. Clearly, this signal's average power is easier to calculate in time. Given that time shifts do not matter in power calculations, the power of signals $y(t)$ and $z(t)$ in, respectively, Examples 3.5 and 3.6, is the same as P_x.

With the Fourier series, however, we can compute which components contribute most to the signal power. Figure 3.8 shows the 'cumulative build-up of power' P as we add more and more components in our calculation above. Signal component $k = 1$ contributes 81% of the power, and after adding component $k = 7$ we already reach 95% of the total signal power P_x. Using a Fourier series model, we can '*see*' how the signal power is built up as a function of frequency.

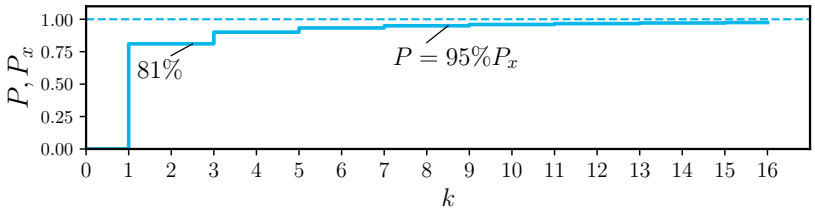


Figure 3.8: Power build-up: the more harmonic components are used in the Fourier series, the closer the power P of the series gets to the power P_x of signal $x(t)$ (horizontal dashed line). Here $K \leq 16$.

EX 3.10 Compute the average power of signals $v(t)$ and $w(t)$ from Examples 3.7 and 3.8. You can assume that both are voltage signals, unit [V].

Solution For signal $v(t)$, use (3.16):

$$P_v = a_0^2 + \frac{1}{2}\sum_{k=1}^{\infty}(a_k^2 + b_k^2) = (2)^2 + \frac{1}{2}\left((3)^2 + (1)^2 + (2\sqrt{2})^2 + (-2\sqrt{2})^2\right) = 4 + 13 = 17 \text{ W}$$

Similarly, for $w(t)$:

$$P_w = A_0^2 + \frac{1}{2}\sum_{k=1}^{\infty} A_k^2 = (4)^2 + \frac{1}{2}\left((8)^2 + (10)^2 + (2)^2\right) = 16 + 84 = 100 \text{ W}$$

Both signals are a summation of sinusoids. Example 2.10 showed that the power of a sinusoid with amplitude A equals $\frac{A^2}{2}$. The power of $v(t)$ and $w(t)$ can be obtained by simply adding up the powers of all components, and the power of the average.

In the next chapter, the complex exponential Fourier series will be discussed, which is much more common in signal analysis than the real Fourier series. The complex exponential Fourier series forms a convenient stepping stone towards the Fourier transform.

4

Complex exponential Fourier series

In Chapter 3 we decomposed a real-valued, periodic signal $x(t)$ (period T_0) into a sum of harmonic cosines and sines, known as the real or trigonometric Fourier series. In this chapter, we perform the *same* signal decomposition but express it differently, using complex algebra, yielding the complex exponential Fourier series. The complex Fourier series allows us to easily obtain the double-sided amplitude and phase spectra of periodic signals. We will discuss the relationships between the real and complex Fourier series, the effects of signal symmetry on the complex Fourier series coefficients, and the time shift and differentiation theorems.

4.1 Derivation of the complex exponential Fourier series

Using Euler's formula (C.1) from Appendix C, we have:

$$\cos(k\omega_0 t) = \frac{1}{2}(e^{-jk\omega_0 t} + e^{jk\omega_0 t}) \quad \text{and} \quad \sin(k\omega_0 t) = j\frac{1}{2}(e^{-jk\omega_0 t} - e^{jk\omega_0 t})$$

Substituting these in (3.1) we obtain (for an infinite number of components K):

$$x(t) = a_0 + \sum_{k=1}^{k=\infty} a_k \frac{1}{2}(e^{-jk\omega_0 t} + e^{jk\omega_0 t}) + \sum_{k=1}^{k=\infty} b_k \frac{j}{2}(e^{-jk\omega_0 t} - e^{jk\omega_0 t})$$

Collecting the terms associated with the positive and negative exponents, we get:

$$x(t) = a_0 + \sum_{k=1}^{k=\infty} \underbrace{\frac{1}{2}(a_k - jb_k)}_{X_k} e^{jk\omega_0 t} + \sum_{k=1}^{k=\infty} \underbrace{\frac{1}{2}(a_k + jb_k)}_{X_{-k}} e^{-jk\omega_0 t} \tag{4.1}$$

Here we define new coefficients, X_k and X_{-k}, for $k \in \mathbb{N}^+$. This is the complex exponential form of the Fourier series:

$$x(t) = a_0 + \sum_{k=1}^{k=\infty} X_k e^{jk\omega_0 t} + \sum_{k=1}^{k=\infty} X_{-k} e^{-jk\omega_0 t}$$

Defining $X_0 = a_0$ (recall $e^{j0\omega_0 t} = 1$), and re-defining the index in the second summation,

$$x(t) = X_0 + \sum_{k=1}^{k=\infty} X_k e^{jk\omega_0 t} + \sum_{k=-1}^{k=-\infty} X_k e^{jk\omega_0 t}$$

yields a more common form of the complex Fourier series, the *synthesis*[1] equation:

$$\boxed{x(t) = \sum_{k=-\infty}^{k=\infty} X_k e^{jk\omega_0 t}} \tag{4.2}$$

The coefficients X_k can be computed using the *analysis* equation:

$$\boxed{X_k = \frac{1}{T_0}\int_{T_0} x(t)e^{-jk\omega_0 t}\,\mathrm{d}t} \tag{4.3}$$

with $k \in \mathbb{Z}$, or, using the real Fourier series coefficients a_k and b_k, (4.1):

$$X_0 = a_0 \quad \text{and} \quad X_k = \tfrac{1}{2}(a_k - jb_k) \quad \text{and} \quad X_{-k} = \tfrac{1}{2}(a_k + jb_k) \quad \text{for} \quad k \in \mathbb{N}^+ \tag{4.4}$$

Hence, $\mathrm{Re}(X_k) = \frac{a_k}{2}$ and $\mathrm{Im}(X_k) = -\frac{b_k}{2}$ for $k \in \mathbb{N}^+$. The complex exponential Fourier series coefficients X_k and X_{-k} are conjugate complex numbers,

$$\boxed{X_{-k} = X_k^*} \tag{4.5}$$

in a similar way as that the complex exponential function $e^{jk\omega_0 t}$ is the complex conjugate of complex exponential function $e^{-jk\omega_0 t}$. Hence, whereas the right-hand side of (4.2) consists of the summation of complex numbers multiplied by the complex exponential basis function, the result of this summation is the *real-valued* signal $x(t)$. The *unit* of the coefficients X_k equals the unit [unit] of the signal.

4.2 Interpretation, polar form

The building block of the complex exponential Fourier series is, as its name implies, the complex exponential basis function. Its components have *positive* frequencies, $k\omega_0$, and *negative* frequencies, $-k\omega_0$, a convenient mathematical construct which stems from the sum of a complex signal and its complex conjugate. The summation for k in (4.2) runs from $k = -\infty$ to ∞, referring to (angular) frequencies $k\omega_0$ from $-\infty$ to ∞ (similarly, for linear frequencies kf_0, with $\omega_0 = 2\pi f_0$). The complex Fourier series is said to be *double-sided*, in line with the construction of a real signal as the sum of a complex signal and the complex conjugate of that signal, see (2.10) in Section 2.3.3 and Appendix C.

The complex Fourier series coefficients X_k are, in general, complex-valued, and can therefore also be expressed in *polar* form, see Appendix C:

$$\boxed{X_k = |X_k|e^{j\theta_k} \quad \text{with } k \in \mathbb{Z}} \tag{4.6}$$

The magnitude of the complex coefficient equals:

$$\boxed{|X_k| = \sqrt{\mathrm{Re}^2(X_k) + \mathrm{Im}^2(X_k)} \qquad \left(= \tfrac{1}{2}\sqrt{a_k^2 + b_k^2} = \tfrac{1}{2}A_k \quad \text{with } k \in \mathbb{N}^+\right)} \tag{4.7}$$

and the phase of the complex coefficient equals:[2]

$$\boxed{\theta_k = \arctan\left(\frac{\mathrm{Im}(X_k)}{\mathrm{Re}(X_k)}\right) \qquad \left(= \arctan\left(-\frac{b_k}{a_k}\right) \quad \text{with } k \in \mathbb{N}^+\right)} \tag{4.8}$$

[1] From ancient Greek: *súnthesis*, meaning 'the combination of parts so as to form a whole'; *analusis*, meaning 'separation of a whole into its component parts'.

[2] Numerically, e.g., in Python, you need to use the four-quadrant arctan function.

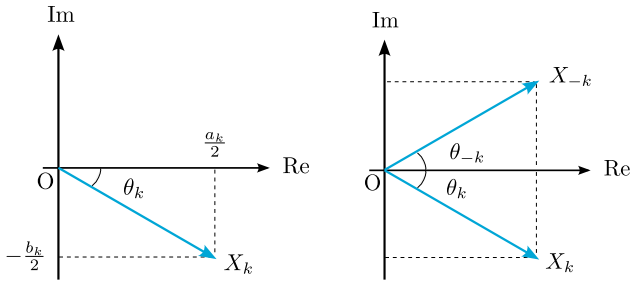


Figure 4.1: Left: complex Fourier series coefficient X_k represented by a vector in the complex plane (for $k > 0$, $a_k > 0$ and $b_k > 0$), with magnitude $|X_k|$. Phase θ_k is shown here as a negative angle. Right: complex Fourier series coefficients X_k and X_{-k} represented by vectors in the complex plane.

Note that $X_0 = a_0$ represents the signal's average, with $|X_0| = |a_0|$ and $\theta_0 = 0$ ($a_0 \geq 0$) or $\theta_0 = \pm\pi$ ($a_0 < 0$). Because $X_{-k} = X_k^*$ it follows that $|X_{-k}| = |X_k|$ and $\theta_{-k} = -\theta_k$.

The coefficient X_k is the *phasor* description, as explained in Section 2.3.1. It scales the complex exponential basis function $e^{jk\omega_0 t}$ magnitude from 1 to $|X_k|$, and sets its phase angle at $t = 0$ seconds from 0 to θ_k, see (4.6). Its conjugate, coefficient X_{-k}, scales the complex exponential basis function $e^{-jk\omega_0 t}$ magnitude from 1 to $|X_{-k}|$, and sets its phase angle at $t = 0$ seconds from 0 to θ_{-k}. The sum of the two complex conjugate phasor signals, $X_k e^{jk\omega_0 t}$ and $X_{-k} e^{-jk\omega_0 t}$, as they always occur in pairs in (4.2), yields the real-valued kth harmonic component. The coefficients X_k and X_{-k}, for $k > 0$, are illustrated in Figure 4.1.

4.3 Effects of symmetry

Similar to the real Fourier series coefficients, the complex Fourier series coefficients show particular patterns when the signal being constructed has a symmetry property. From (4.4) we can directly derive that when a signal is even (only cosines are needed, all b_k's are zero) the complex Fourier series coefficients are real-valued; when a signal is odd (only sines are needed, all a_k's are zero) the coefficients are imaginary, and when a signal is half-wave odd (all even-indexed $\{a_k, b_k\}$'s are zero) the even-indexed coefficients are zero.

This also follows from re-writing the analysis equation, (4.3), as follows:

$$X_k = \underbrace{\frac{1}{T_0}\int_{T_0} x(t)\cos(k\omega_0 t)\,\mathrm{d}t}_{\text{real part } \mathrm{Re}(X_k)} + j\underbrace{\left(\frac{(-1)}{T_0}\int_{T_0} x(t)\sin(k\omega_0 t)\,\mathrm{d}t\right)}_{\text{imaginary part } \mathrm{Im}(X_k)} \tag{4.9}$$

When $x(t)$ is even, $x(-t) = x(t)$, the imaginary part in (4.9) is zero, as it integrates the multiplication of an even function with an odd function, which results in an odd function, the integral of which over one period equals zero. Similarly, when $x(t)$ is odd, $x(-t) = -x(t)$, the real part in (4.9) is zero. In general, signals are neither even nor odd, and (some or all of) their complex Fourier series coefficients are complex-valued.

4.4 Double-sided spectrum

The complex exponential Fourier series gives an array of complex numbers $\{X_k\}$, for $k \in \mathbb{Z}$, which can be directly converted to the *double-sided* amplitude and phase spectra. This is because the complex Fourier series is already written as a sum of complex conjugate phasors, (2.10). For instance, consider the kth harmonic, with (4.6):

$$X_k e^{jk\omega_0 t} + X_{-k} e^{-jk\omega_0 t} = |X_k| e^{j\theta_k} e^{jk\omega_0 t} + |X_{-k}| e^{j\theta_{-k}} e^{-jk\omega_0 t}$$

For the positive frequency $k\omega_0$ we have amplitude $|X_k|$ and phase θ_k; for the negative frequency $-k\omega_0$ we have amplitude $|X_{-k}| = |X_k|$ and phase $\theta_{-k} = -\theta_k$. Hence, the amplitude spectrum is an *even* function of frequency, while the phase spectrum is an *odd* function of frequency. Note that for the amplitude spectrum we use the modulus or magnitude of X_k, and hence, 'magnitude spectrum' would be a more appropriate term.

Comparing the magnitude and phases of the double-sided spectrum, (4.7) and (4.8), with the expressions obtained for the single-sided spectrum, (3.12) and (3.13), we see that, for positive frequencies, the phases are the same, and the amplitudes are *halved* in the double-sided spectrum. At the zero frequency, $k = 0$, i.e., the average, the value remains the same.

4.4.1 Diagram of coefficients

In practical applications, we have K components and the complex exponential Fourier series gives an array of $2K-1$ complex numbers $\{X_k\}$, for $k = -(K-1), \ldots, -1, 0, 1, \ldots, K-1$. The relation between the $2K-1$ complex Fourier series coefficients and the double-sided spectrum of the K components is illustrated in Figure 4.2. Compare this diagram with the diagram of the real Fourier series, Figure 3.7, and study the differences.

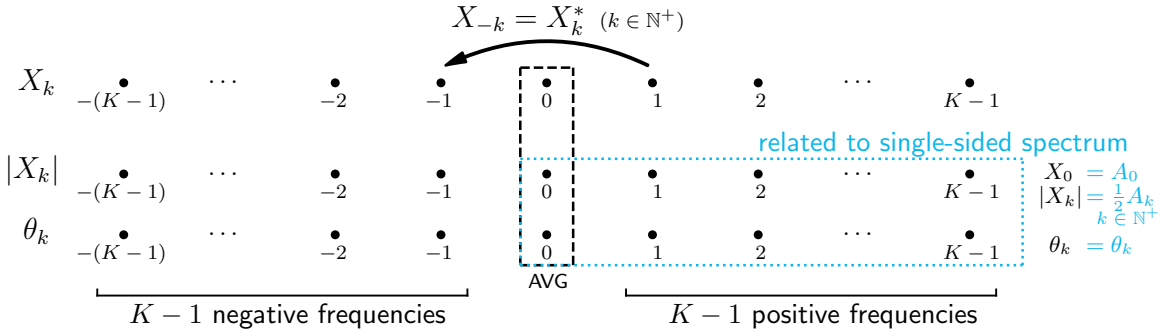


Figure 4.2: Complex Fourier series coefficients (top row) and the double-sided amplitude (middle) and phase (bottom) spectra, see (4.5), (4.7) and (4.8). In this diagram, AVG means average.

4.4.2 Example

EX 4.1

Compute the complex Fourier series coefficients of the square wave of Example 3.1, using the real Fourier series coefficients. Show the complex coefficients as a function of frequency, together with the double-sided amplitude and phase spectra.

Solution In Example 3.1, we found that $a_0 = 0$, $a_k = 0\ \forall\ k$, $b_k = \frac{4}{k\pi}$ for odd k and $b_k = 0$ for even k. Using (4.4) we obtain $X_0 = a_0 = 0$ and, for $k \in \mathbb{N}^+$:

$$X_k = \begin{cases} -\frac{2j}{k\pi} & \text{for } k \text{ odd} \\ 0 & \text{for } k \text{ even} \end{cases} \quad \text{and} \quad X_{-k} = \begin{cases} +\frac{2j}{k\pi} & \text{for } k \text{ odd} \\ 0 & \text{for } k \text{ even} \end{cases} \tag{4.10}$$

One is advised to start with X_k, $k \in \mathbb{N}^+$, then take the complex conjugate of X_k to obtain X_{-k}. Figure 4.3 shows the real and imaginary parts of X_k, the double-sided amplitude $|X_k|$ and phase θ_k spectra, for $k \in \mathbb{Z}$, using $K = 17$ components. Compare this spectrum with its single-sided equivalent, Figure 3.2. To facilitate this comparison, Figure 4.3 shows the A_k values in light gray (row 3). Amplitude and phase spectra are, respectively, even and odd functions of frequency. They are referred to as *line spectra* because only discrete frequencies occur in them, namely integer multiples of ω_0.

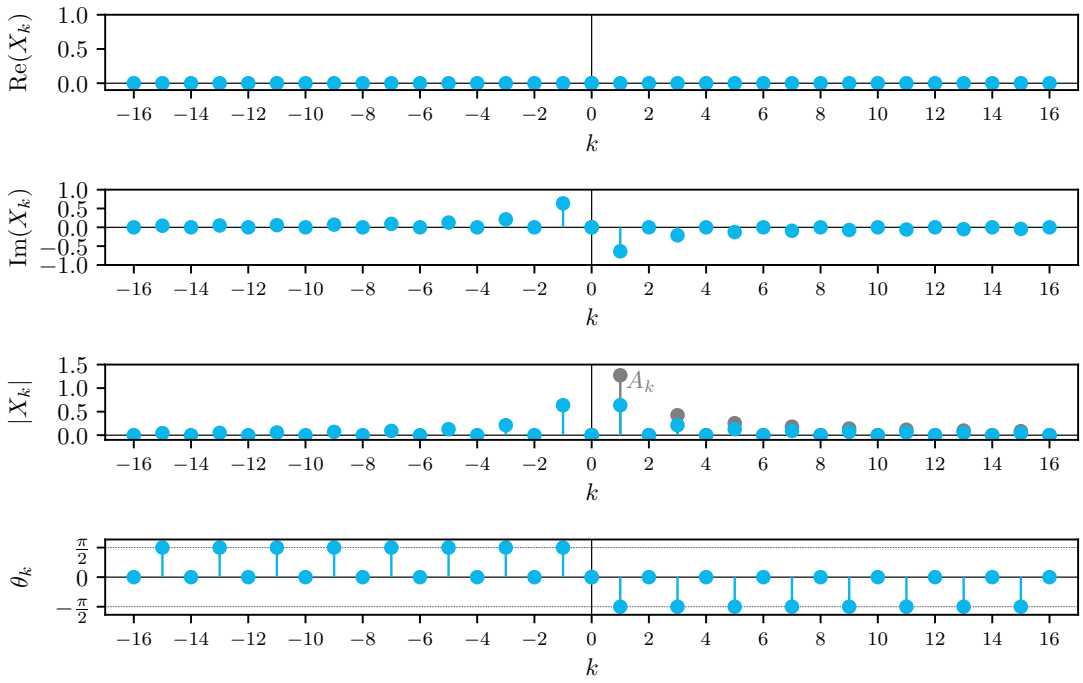


Figure 4.3: Signal $x(t)$: real and imaginary part of the complex Fourier series coefficients (rows 1 and 2); double-sided amplitude (row 3) and phase (row 4) spectra, for the first $K = 17$ components.

For the complex Fourier series, showing the amplitude and phase spectra is more common than showing the real and imaginary parts of the coefficients as a function of frequency. When considering the Fourier series coefficients corresponding to the kth harmonic, X_k and X_{-k}, the amplitude spectrum gives *no* indication about whether the coefficient is real-, imaginary-, or complex-valued. This information is *all* contained in the phase spectrum. Phases equal to 0 and $\pm\pi$ indicate pure cosines (real), phases equal to $\pm\frac{\pi}{2}$ indicate pure sines (imaginary), and any other phases indicate shifted cosines (complex). In Figure 4.3, the phases reveal that all non-zero complex Fourier series coefficients are imaginary-valued: they all represent pure sines.

4.5 Examples

EX 4.2

Consider a periodic signal $y(t)$ which is composed of three cosines and two sines:

$$y(t) = 1\cos(\pi t) + 2\cos(2\pi t) + 1\cos(4\pi t) + 1\sin(\pi t) - 2\sin(3\pi t)$$

Determine the fundamental frequency f_0 and period T_0. Compute the real Fourier series coefficients a_0, $\{a_k, b_k\}$ and, from these, the complex Fourier series coefficients X_k.

Solution The five components are commensurable and their highest common frequency equals $f_0 = 0.5$ Hz; this signal is periodic with period $T_0 = 2$ s. The signal is already written as a Fourier series, so we can quickly obtain its real Fourier series coefficients through writing $y(t)$ like (3.2), with $\omega_0 = 2\pi f_0 = \pi$ rad/s:

$$y(t) = 1\cos(1\omega_0 t) + 2\cos(2\omega_0 t) + 1\cos(4\omega_0 t) + 1\sin(1\omega_0 t) - 2\sin(3\omega_0 t)$$

Then: $a_0 = 0$, $a_1 = 1$, $a_2 = 2$, $a_4 = 1$, $b_1 = 1$ and $b_3 = -2$. All other coefficients are zero. The complex Fourier series coefficients can be computed using (4.4):

$$\begin{aligned} Y_0 = 0, \quad Y_1 &= \tfrac{1}{2}(1-j), \quad Y_2 = 1, \quad Y_3 = j, \quad Y_4 = \tfrac{1}{2} \\ Y_{-1} &= \tfrac{1}{2}(1+j), \quad Y_{-2} = 1, \quad Y_{-3} = -j, \quad Y_{-4} = \tfrac{1}{2} \end{aligned}$$

Figure 4.4 shows the signal, the real Fourier series coefficients and the double-sided amplitude and phase spectra. Verify their magnitudes and phases using (4.7) and (4.8).

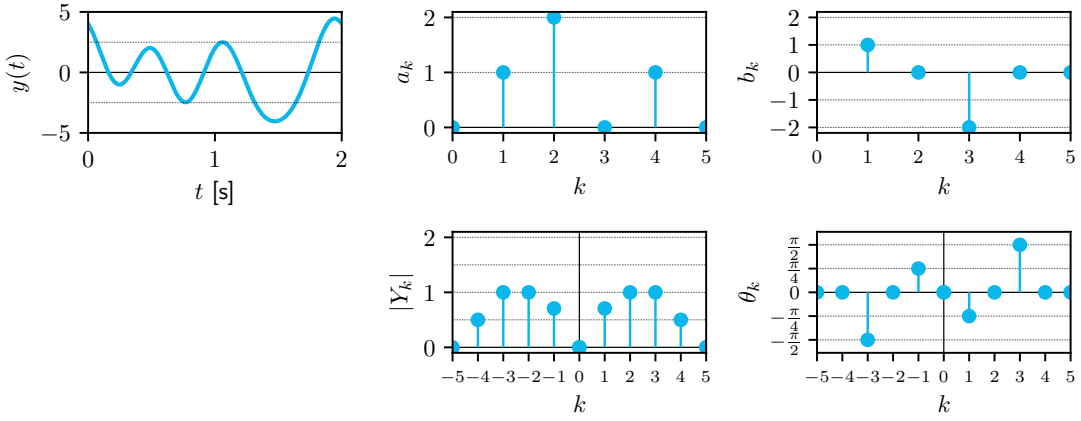


Figure 4.4: One period of signal $y(t)$ (left), its real Fourier series coefficients a_k and b_k (right, top) and its double-sided amplitude and phase spectra, $|Y_k|$ and θ_k (right, bottom).

EX 4.3

Compute the complex exponential Fourier series coefficients X_k for the pulse train signal $x(t)$ illustrated in Figure 4.5.

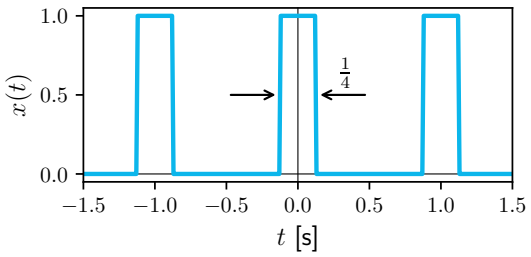


Figure 4.5: Pulse train signal $x(t)$, with amplitude $A = 1$, period $T_0 = 1$ s and pulse width $\tau = \frac{1}{4}$ s.

Solution Though not really needed, we first provide an expression for $x(t)$. The building block of this signal is $\Pi(t)$, the unit pulse function, (B.1). We scale this pulse to a width of τ seconds, by considering $\Pi(\frac{t}{\tau})$, still centered at $t = 0$ s, and we set its amplitude to A. This pulse then repeats itself every T_0 seconds:

$$x(t) = \sum_{n=-\infty}^{n=\infty} A\Pi\left(\frac{t - nT_0}{\tau}\right) \tag{4.11}$$

In Figure 4.5, $A = 1$, $T_0 = 1$ s and $\tau = \frac{1}{4}$ s. We compute the coefficients using the analysis equation, (4.3). We choose the integration interval to be symmetric about

$t = 0$ s, that is, from $t = -\frac{T_0}{2}$ to $\frac{T_0}{2}$:

$$X_k = \frac{1}{T_0}\int_{-\frac{T_0}{2}}^{\frac{T_0}{2}} x(t)e^{-jk\omega_0 t}\,\mathrm{d}t = \frac{1}{T_0}\int_{-\frac{\tau}{2}}^{\frac{\tau}{2}} Ae^{-jk\omega_0 t}\,\mathrm{d}t$$

for all $k \in \mathbb{Z}$. The integral is evaluated as:

$$X_k = \frac{A}{T_0}\left[-\frac{1}{jk\omega_0}e^{-jk\omega_0 t}\right]_{-\frac{\tau}{2}}^{\frac{\tau}{2}} = \frac{A}{jk\omega_0 T_0}\underbrace{\left(-e^{-jk\omega_0\frac{\tau}{2}} + e^{jk\omega_0\frac{\tau}{2}}\right)}_{=2j\sin(k\omega_0\frac{\tau}{2})}$$

Using (C.3), and substituting $\omega_0 = \frac{2\pi}{T_0}$, we obtain:

$$X_k = \frac{2A}{k\omega_0 T_0}\sin(k\omega_0\frac{\tau}{2}) = \frac{A}{k\pi}\sin(\frac{k\pi\tau}{T_0})$$

Multiplying the right-hand side by $1 = \frac{\tau}{T_0}\frac{T_0}{\tau}$ and then re-arranging terms yields:

$$X_k = \frac{A}{k\pi}\sin(\frac{k\pi\tau}{T_0})\frac{\frac{\tau}{T_0}}{\frac{\tau}{T_0}} = \frac{A\tau}{T_0}\frac{\sin(\frac{\pi k\tau}{T_0})}{\frac{\pi k\tau}{T_0}} = \frac{A\tau}{T_0}\operatorname{sinc}\left(\frac{k\tau}{T_0}\right) \tag{4.12}$$

with the sinc function defined in Appendix B, (B.3).

In this case, with a real, even function of time $x(t)$, all complex Fourier series coefficients are real-valued; their imaginary parts are zero. The complex Fourier series coefficients represent a real, even, discrete, function of frequency $k\omega_0$, with $k \in \mathbb{Z}$. This follows from the symmetry properties discussed in Section 4.3.

Figure 4.6 shows the double-sided magnitude (amplitude) spectrum $|X_k|$, at left, and the double-sided phase spectrum $\theta_k = \arg(X_k)$ at right, for the first $K = 11$ components, in blue. In the amplitude spectrum, it *also* shows the values of X_k, in gray, which is possible in this example because X_k only has a real part.

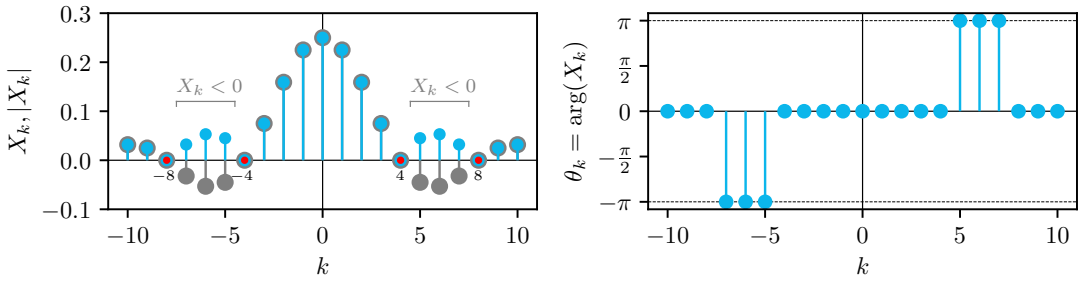


Figure 4.6: At left, magnitude (amplitude) spectrum $|X_k|$ (blue). It also shows the (here, real-valued) complex Fourier series coefficients X_k (gray). At right, phase spectrum showing the angle (or argument) of the coefficients $\theta_k = \arg(X_k)$, in radians, for the pulse train in Figure 4.5.

For the signal average, we substitute $k = 0$ in (4.12) and obtain, since $\operatorname{sinc}(0) = 1$, $X_0 = \frac{A\tau}{T_0} = \frac{1}{4}$. This makes sense, as the signal illustrated in Figure 4.5 equals 1 during $\tau = \frac{1}{4}$ s and 0 during $T_0 - \tau = \frac{3}{4}$ s, for every $T_0 = 1$ s period.

The sinc function $\text{sinc}(\frac{k\tau}{T_0})$ has its zero-crossings when $\frac{k\tau}{T_0} = \pm 1, \pm 2, ...$, see Appendix B. When τ fits an integer m number of times in T_0, $m\tau = T_0$, then the complex Fourier series coefficients will indeed be zero for $k = \pm m, \pm 2m, \pm 3m, ...$ In this example, $m = 4$, so $X_k = 0$ for $k = \pm 4, \pm 8, ...$, indicated by the red dots in Figure 4.6.

In Figure 4.6, some X_k coefficients have negative values (e.g., X_5). These coefficients lie on the negative real axis in the complex plane; their phase is $\pm\pi$. Here, we have *chosen* the phase of these coefficients for $k > 0$ to be $+\pi$. Then, because of the oddness of the phase spectrum, the phase of these coefficients for $k < 0$ equals $-\pi$.

The phases of all positive-valued X_k coefficients are zero, as all these coefficients lie on the positive real axis in the complex plane. The phase of the average, X_0, is zero as the average is a positive real number, $\frac{1}{4}$.

4.6 Two Fourier series theorems

The complex Fourier series comes with a number of theorems which can reduce the number of calculations needed to find the series expansion. Two theorems are discussed here.

4.6.1 Time shift theorem

Imagine that we have already obtained the complex Fourier series coefficients X_k of a periodic signal $x(t)$. The *time shift theorem* of the complex Fourier series allows us to quickly calculate the complex Fourier series coefficients Y_k of the time-shifted signal $y(t) = x(t + t_0)$:

$$\boxed{Y_k = X_k e^{jk\omega_0 t_0} \quad \text{with } k \in \mathbb{Z}} \tag{4.13}$$

Because of the time shift, the phase of the complex Fourier coefficients changes: $\arg(Y_k) = \arg(X_k) + k\omega_0 t_0$; recall (3.15). Their magnitudes remain the same: $|Y_k| = |X_k|$. The complex Fourier series coefficients, the phasors, *rotate* in the complex plane.

EX 4.4

Using the time shift theorem, compute the complex Fourier series coefficients Y_k of signal $y(t) = x(t + \frac{T_0}{4})$, with $x(t)$ the square wave of Example 4.1.

Solution Coefficients X_k have been computed, (4.10). To obtain $y(t)$, signal $x(t)$ is shifted by $t_0 = \frac{T_0}{4}$ s ($t_0 > 0$, to the left). Applying the time shift theorem, (4.13), yields:

$$Y_k = X_k e^{jk\omega_0 \frac{T_0}{4}} = X_k e^{jk\frac{2\pi}{T_0}\frac{T_0}{4}} = X_k e^{jk\frac{\pi}{2}}$$

While this equation holds for $k \in \mathbb{Z}$, one is advised to work with $k \in \mathbb{N}$ and first obtain the coefficients at the positive frequencies, then take the complex conjugate to obtain the coefficients at the negative frequencies. Since $e^{jk\frac{\pi}{2}} = (j)^k$, once we know X_k we can readily compute Y_k. We can show this in a table, for $k \geq 0$:

k	0	1	2	3	4	5	6	7	...
X_k	0	$\frac{-2j}{\pi}$	0	$\frac{-2j}{3\pi}$	0	$\frac{-2j}{5\pi}$	0	$\frac{-2j}{7\pi}$	
$(j)^k$	1	j	-1	$-j$	1	j	-1	$-j$	
Y_k	0	$\frac{2}{\pi}$	0	$\frac{-2}{3\pi}$	0	$\frac{2}{5\pi}$	0	$\frac{-2}{7\pi}$	

As expected, all complex Fourier series coefficients Y_k are real-valued, and signal $y(t)$ is an even function. Taking the complex conjugate to calculate Y_{-k} leads to the same

Fourier series coefficient in this case: $Y_{-k} = Y_k$. In other words:

$$Y_k = \begin{cases} +\frac{2}{|k|\pi} & \text{for } k = \pm 1, \pm 5, \pm 9, \ldots \\ -\frac{2}{|k|\pi} & \text{for } k = \pm 3, \pm 7, \pm 11, \ldots \\ 0 & \text{for } k \text{ even} \end{cases} \tag{4.14}$$

As expected, the magnitudes remain the same: $|Y_k| = |X_k|$. The complex coefficients are *rotated* by $k\frac{\pi}{2}$ in the complex plane. When considering the coefficients Y_k as a function of frequency $k\omega_0$, we obtain a real and even function of frequency. This is because $y(t)$ is real and even. One can check the complex Fourier series coefficients in (4.14) using the real Fourier series coefficients computed in Example 3.5.

4.6.2 Differentiation theorem

Imagine that we have obtained the complex Fourier series coefficients X_k of a periodic signal $x(t)$. The *differentiation theorem* allows us to quickly calculate the complex Fourier series coefficients Y_k of signal $y(t) = \frac{\mathrm{d}^n x(t)}{\mathrm{d}t^n}$ (with $n \in \mathbb{N}^+$), the nth derivative of $x(t)$:

$$\boxed{Y_k = (jk\omega_0)^n X_k \quad \text{with } k \in \mathbb{Z}} \tag{4.15}$$

EX 4.5

Compute the complex Fourier series coefficients Z_k of signal $z(t)$, defined as the first derivative of $y(t)$ in Example 4.2: $z(t) = \frac{\mathrm{d}}{\mathrm{d}t} y(t)$.

Solution With the Fourier series coefficients Y_k of $y(t)$, and $\omega_0 = \pi$ rad/s, we obtain, using (4.15) with $n = 1$, that $Z_k = (jk\omega_0)^1 Y_k = (jk\pi)Y_k$. For $k = 0$ we obtain $Z_0 = (0)Y_0 = 0$: differentiating a constant with respect to time yields zero. That is, $z(t)$ will always have a zero average. Consider positive frequencies, $k > 0$: then, each coefficient Y_k is multiplied by $jk\pi$, an imaginary number with magnitude $k\pi$ and phase $\frac{\pi}{2}$. Hence, the magnitude $|Z_k| = k\pi|Y_k|$, and the phase $\arg(Z_k) = \arg(Y_k) + \frac{\pi}{2}$. All coefficients Y_k are *rotated* counterclockwise in the complex plane by $+\frac{\pi}{2}$, and multiplied by $k\pi$ to produce Z_k. For the negative frequencies, $k < 0$, the Y_k-s are rotated by $-\frac{\pi}{2}$, i.e., clockwise, and multiplied by $|k|\pi$. This can be shown in a table, for $k \geq 0$:

k	0	1	2	3	4
Y_k	0	$\frac{1}{2}(1-j)$	1	j	$\frac{1}{2}$
$jk\pi$	0	$j\pi$	$j2\pi$	$j3\pi$	$j4\pi$
Z_k	0	$\frac{\pi}{2}(1+j)$	$j2\pi$	-3π	$j2\pi$

We see that, e.g., $Y_2 = 1$ lies on the positive real axis and is rotated to the positive imaginary axis (and multiplied by 2π) to become $Z_2 = j2\pi$. The cosine component with magnitude 2 and frequency $2\omega_0 = 2\pi$ rad/s becomes a sine component with magnitude 4π and the same frequency.[a] Now, of course, the relationship $Z_{-k} = Z_k^*$ still holds, as $z(t)$ remains a real-valued signal. We obtain:

$$\begin{aligned} Z_0 = 0, \quad Z_1 &= \tfrac{\pi}{2}(1+j), \quad Z_2 = j2\pi, \quad Z_3 = -3\pi, \quad Z_4 = j2\pi \\ Z_{-1} &= \tfrac{\pi}{2}(1-j), \quad Z_{-2} = -j2\pi, \quad Z_{-3} = -3\pi, \quad Z_{-4} = -j2\pi \end{aligned}$$

Figure 4.7 shows $z(t)$ and its double-sided amplitude and phase spectra. Verify their magnitudes and phases using (4.7) and (4.8).

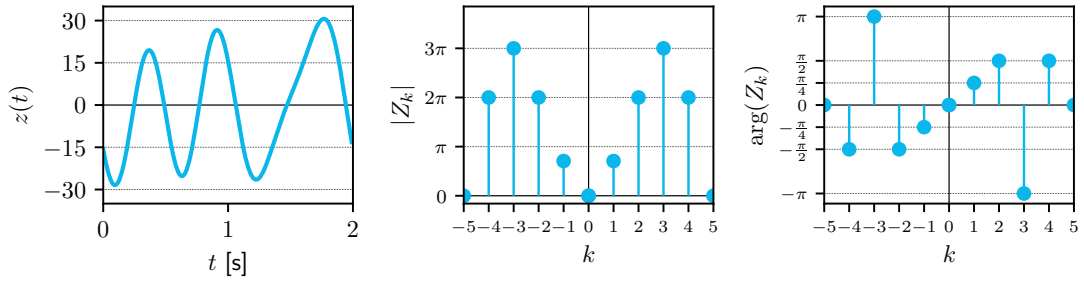


Figure 4.7: Signal $z(t)$ (left) and its double-sided amplitude (center) and phase spectra (right). All other coefficients, magnitudes and phases, are zero.

[a] In plain mathematics: $\frac{\mathrm{d}}{\mathrm{d}t}(2\cos(2\pi t)) = -4\pi\sin(2\pi t) = -4\pi\cos(2\pi t - \frac{\pi}{2}) = 4\pi\cos(2\pi t + \frac{\pi}{2})$.

4.7 Parseval: complex exponential Fourier series

Similar to the real Fourier series, the average power of a periodic signal $x(t)$ can be computed with Parseval's theorem (Section 13.2.1) using the complex Fourier series coefficients:

$$P = \frac{1}{T_0}\int_{t_0}^{t_0+T_0} x^2(t)\,\mathrm{d}t = \sum_{k=-\infty}^{\infty}|X_k|^2 = X_0^2 + 2\sum_{k=1}^{\infty}|X_k|^2 \tag{4.16}$$

Again, *only* the amplitudes matter; the phases and frequencies are irrelevant when computing signal power. The amplitude spectrum contains *all* the information needed to compute the signal power. We do need to include *all* components of the Fourier series, i.e., $K \to \infty$.

EX 4.6

Compute the average power of the square wave $x(t)$ of Example 4.1 using (4.16). Assume that $x(t)$ is a voltage signal, and its unit is [V].

Solution The square wave average power was computed to be 1 W in Example 3.9. Using Example 4.1 the complex Fourier series coefficients magnitudes can be calculated: $|X_0| = 0$, $|X_k| = 0$ for even k, $|X_k| = |\frac{2}{k\pi}|$ for odd k. Using (4.16) we get:

$$P_x = X_0^2 + 2\sum_{k=1}^{\infty}|X_k|^2 = 0^2 + 2\frac{4}{\pi^2}\sum_{\ell=1}^{\infty}\frac{1}{(2\ell-1)^2} = \frac{8}{\pi^2}\frac{\pi^2}{8} = 1\ \text{W},$$

where we applied the same reasoning as in Example 3.9.

With the complex Fourier series covered, it is a small step to the Fourier transform, the quintessential tool in signal analysis, which can deal with aperiodic signals. The basic idea is that we increase the period T_0 of the periodic signal $x(t)$; in fact, we make it infinitely large, and then look at what happens to the complex Fourier series coefficients X_k. Chapter 5 will show that we get a *continuous function of frequency* $X(f)$.

5

Fourier transform

Previously we covered the real Fourier series and the complex exponential Fourier series, which are both signal decompositions of periodic signals. The Fourier transform applies to *aperiodic* signals, and can be considered as to evolve from the complex exponential Fourier series for a periodic signal, when the period T_0 of that signal becomes infinitely large. In this chapter, the Fourier transform will be derived, effects of signal symmetry on the transform investigated, and the double-sided spectrum for aperiodic signals defined. The relationship with the complex Fourier series will be established, and it will be shown that periodic signals can *also* be Fourier-transformed, in-the-limit, yielding Dirac delta functions in frequency.

5.1 Rationale: development towards aperiodic signals

Recall Example 4.3, where we studied the complex exponential Fourier series for the pulse train signal $x(t)$, shown in Figure 4.5 (amplitude $A = 1$, period $T_0 = 1$ second and pulse width $\tau = \frac{1}{4}$ s). The complex exponential Fourier series coefficients X_k of this periodic signal are:

$$X_k = \frac{A\tau}{T_0}\,\mathrm{sinc}\left(\frac{k\tau}{T_0}\right), \quad \forall\ k\ \in\ \mathbb{Z} \tag{4.12}$$

Because signal $x(t)$ is real and even, the coefficients X_k represent a real, even, discrete function of frequency $k\omega_0$ (or kf_0), with $k \in \mathbb{Z}$. Now, suppose that the amplitude A and pulse width τ remain constant: what would happen to this function when the time between the pulses, the signal period T_0, increases?

Figure 5.1 illustrates $x(t)$ for two periods, $T_0 = 1$ s (top, left) and 5 s (top, right). At bottom, for both periods, it shows T_0X_k as a function of frequency $f = kf_0$ in [Hz], as stems, and the continuous function $X(f) = A\tau\,\mathrm{sinc}(\tau f)$, dashed. Clearly, the spacing of the Fourier series coefficients along the frequency axis decreases, and the bottom-right graph becomes much more dense. This is because when T_0 increases, the fundamental frequency f_0 *decreases*, and since we show X_k as a function of kf_0, the coefficients move closer and closer in frequency. For $T_0 = 1$ s, the coefficients are 1 Hz apart; for $T_0 = 5$ s, they are $\frac{1}{5}$ Hz apart.

Eventually, for $T_0 \to \infty$, the complex Fourier series coefficients merge to become a *continuous* function of frequency, and together form the dashed envelope function $X(f)$ in Figure 5.1: $T_0X_k \to X(f = kf_0)$. As we will discover in Example 5.1, $X(f)$ is the Fourier transform of one single pulse, here shown with amplitude $A = 1$, width $\tau = \frac{1}{4}$ s, and its center at $t = 0$ s.

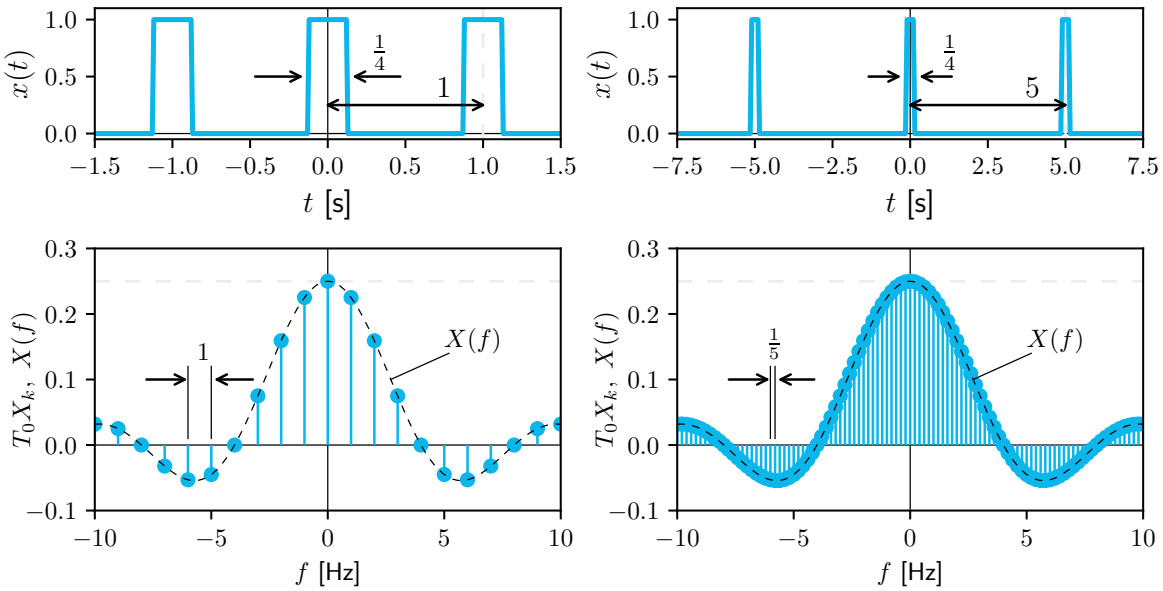


Figure 5.1: Two pulse trains (top row) differing in their period $T_0 = 1$ s (left column) and $T_0 = 5$ s (right column); pulse duration $\tau = \frac{1}{4}$ s and amplitude $A = 1$. The bottom row shows complex Fourier series coefficients scaled by T_0, i.e., T_0X_k, as a function of frequency f in [Hz], with $f = kf_0$; and the envelope function, $X(f)$, the Fourier transform of one single pulse with duration $\tau = \frac{1}{4}$ s and amplitude $A = 1$ (dashed).

5.2 Derivation of the Fourier transform

To derive the Fourier transform from the complex exponential Fourier series in a heuristic fashion, we return to the synthesis and analysis equations, (4.2) and (4.3):

$$x(t) = \sum_{k=-\infty}^{k=\infty} X_k e^{j2\pi k f_0 t} \tag{4.2}$$

where we substituted $\omega_0 = 2\pi f_0$, and:

$$X_k = \frac{1}{T_0} \int_{-T_0/2}^{T_0/2} x(t) e^{-j2\pi k f_0 t}\,\mathrm{d}t \quad \text{with } k \in \mathbb{Z} \tag{4.3}$$

where the integral over one period T_0 is conveniently chosen to be symmetrical about zero.

With $f_0 = \frac{1}{T_0}$ we obtain, by substituting the second equation above into the first one:

$$x(t) = \sum_{k=-\infty}^{\infty} \left(f_0 \int_{-T_0/2}^{T_0/2} x(t) e^{-j2\pi k f_0 t}\,\mathrm{d}t \right) e^{j2\pi k f_0 t}$$

Now, when $T_0 \to \infty$, then f_0 becomes infinitesimally small: $f_0 \to \mathrm{d}f$, and the product kf_0 approaches the continuous frequency variable f. The *summation* from $k = -\infty$ to $k = \infty$ becomes an *integration* in f from $f = -\infty$ to $f = \infty$:

$$x(t) = \int_{f=-\infty}^{\infty} \left(\mathrm{d}f \int_{t=-\infty}^{\infty} x(t) e^{-j2\pi f t}\,\mathrm{d}t \right) e^{j2\pi f t}$$

Re-arranging the terms yields:

$$x(t) = \int_{f=-\infty}^{\infty} \underbrace{\left(\int_{t=-\infty}^{\infty} x(t) e^{-j2\pi f t} \, \mathrm{d}t \right)}_{= X(f)} e^{j2\pi f t} \, \mathrm{d}f$$

The integration within the parentheses, known as the *Fourier integral*, is defined as the (continuous-time) Fourier transform of aperiodic signal $x(t)$:

$$\boxed{X(f) = \mathcal{F}\{x(t)\} = \int_{t=-\infty}^{\infty} x(t) e^{-j2\pi f t} \, \mathrm{d}t} \tag{5.1}$$

The inverse Fourier transform follows readily as:

$$\boxed{x(t) = \mathcal{F}^{-1}\{X(f)\} = \int_{f=-\infty}^{\infty} X(f) e^{j2\pi f t} \, \mathrm{d}f} \tag{5.2}$$

Similar to the complex exponential Fourier series, (5.1) is known as the *analysis* equation, and (5.2) as the *synthesis* equation.

The *unit* of the Fourier transform of a signal, $X(f)$, equals [unit seconds] or, equivalently, [unit/Hz], with [unit] the unit of signal $x(t)$. Fourier transform $X(f)$ is a *density* function[1] in the frequency domain. According to (5.2) we need to *integrate* $X(f)$, multiplied by the complex exponential $e^{j2\pi f t}$, over frequency, from $f = -\infty$ to ∞ in order to obtain the original signal $x(t)$ in the time domain.

The complex-valued coefficients X_k with the complex Fourier series for periodic signals, pertaining to discrete frequencies equal to k times the fundamental frequency of that signal (i.e., $k\omega_0$ or kf_0), have become a complex-valued, continuous *function* of frequency $X(f)$ with the Fourier transform. Apparently, for aperiodic signals, we *may* need components with *all* possible frequencies f to construct these signals using the complex exponential basis function.

The continuous-time signal $x(t)$ and its continuous-frequency Fourier transform $X(f)$ are referred to as a Fourier transform *pair*, and denoted by $x(t) \overset{\mathcal{F}}{\leftrightarrow} X(f)$, or, as $\{x(t), X(f)\}$.

An overview of frequently used Fourier transform *pairs* can be found in Appendix E. The next chapter will elaborate on the most important Fourier transform *theorems*. One of these theorems, *linearity*, states that the Fourier transform is a linear operation:

$$\mathcal{F}\{a_1 x_1(t) + a_2 x_2(t)\} = a_1 X_1(f) + a_2 X_2(f) \tag{6.1}$$

for arbitrary constants a_1 and a_2. This property will already be used throughout this chapter.

5.3 Conditions

Not all (aperiodic) signals can be Fourier-transformed. The conditions that guarantee that the Fourier transform of a signal $x(t)$ exists, are called the Dirichlet[2] conditions [3]:

1. Signal $x(t)$ has a finite number of finite discontinuities,

[1] Similar to the probability density function in statistics, the 'signal content' at a single discrete frequency f_1, equals zero, that is, $\int_{f=f_1}^{f=f_1} X(f) e^{j2\pi f t} \mathrm{d}f = 0$.

[2] After the German mathematician J.P.G. Lejeune Dirichlet (1805-1859).

2. Signal $x(t)$ has a finite number of minima and maxima, and
3. Signal $x(t)$ is *absolutely integrable* over the infinite interval $(-\infty, \infty)$, that is:

$$\int_{-\infty}^{\infty} |x(t)|\,\mathrm{d}t < \infty \tag{5.3}$$

Typically, but not invariably [5], these conditions mean that $x(t)$ will decay to zero as $t \to \pm\infty$.

Conditions 1-3 above are *sufficient*, meaning that there are signals that violate at least one of the conditions, but still have Fourier transforms, [7]. Several examples of these signals, e.g., the Dirac function $\delta(t)$ and sinusoidal signals, will be shown in Section 5.8, where Fourier transforms in-the-limit are discussed.

While signals that are absolutely integrable, (5.3), have finite energy, the reverse is *not* always true [3]. Hence, a weaker condition, relevant in this book, is that real signal $x(t)$ has finite energy:

$$E_x = \int_{-\infty}^{\infty} x^2(t)\,\mathrm{d}t < \infty \tag{5.4}$$

Most practical signals, with finite energy, can be Fourier-transformed.

5.4 Interpretation, polar form

The Fourier transform has much in common with the complex exponential Fourier series. Both have the complex exponential basis function as their building block, are evaluated at positive and negative frequencies (i.e., are double-sided) and, for real signals, always consist of pairs of complex conjugates.

Similar to the complex Fourier series, the Fourier transform $X(f)$ of a signal is, in general, complex-valued, and can be expressed in polar form:

$$\boxed{X(f) = |X(f)|\,e^{j\theta(f)}} \tag{5.5}$$

where for the phase $\theta(f)$ the notation with the angle symbol, $\angle X(f)$, is also often used, and sometimes the argument $\arg(X(f))$.

The magnitude of the complex-valued Fourier transform, at each frequency f, equals:

$$\boxed{|X(f)| = \sqrt{\mathrm{Re}(X(f))^2 + \mathrm{Im}(X(f))^2}} \tag{5.6}$$

and the phase of the complex-valued Fourier transform, at each frequency f, equals:[3]

$$\boxed{\angle X(f) = \arctan\left(\frac{\mathrm{Im}(X(f))}{\mathrm{Re}(X(f))}\right)} \tag{5.7}$$

The magnitude $|X(f)|$ is a *density* function. Suppose signal $x(t)$ has unit Volt [V], then $|X(f)|$ has unit [V/Hz]. $|X(f)|$ represents the 'level of signal magnitude *per frequency*'. To compute the contribution, in terms of magnitude, of the harmonics in a band of frequencies $[f_1, f_2]$ (in [Hz]) to signal $x(t)$, we must *integrate* $|X(f)|$ from f_1 to f_2. This will be further discussed in Section 5.9.

[3] Numerically, e.g., in Python, you need to use the four-quadrant arctan function.

Note that $X(0)$ does *not* represent the signal's average (substitute $f = 0$ in (5.1)):

$$X(0) = \int_{t=-\infty}^{\infty} x(t)\,\mathrm{d}t \tag{5.8}$$

This is an important difference with the complex exponential Fourier series, where X_0 equals the signal mean over one period. Note that energy signals *must* have zero mean, see Example 2.11. For real-valued signals $x(t)$, $X(0)$ lies on the real axis, its magnitude is $|X(0)|$, and its phase equals either 0 (when $X(0) \geq 0$) or $\pm\pi$ (when $X(0) < 0$).

5.5 Effects of symmetry

The Fourier transform, like the Fourier series, shows particular patterns when the signal being transformed has a symmetry property. Aperiodic signals have two possible symmetries, evenness and oddness, see Table 2.1. All real-valued signals $x(t)$ can be written as the sum of an even signal $x_e(t)$ and an odd signal $x_o(t)$, (2.5). Then, $\mathcal{F}\{x(t)\} = \mathcal{F}\{x_e(t) + x_o(t)\} = \mathcal{F}\{x_e(t)\} + \mathcal{F}\{x_o(t)\}$. Re-writing the analysis equation, (5.1), we obtain:

$$X(f) = \underbrace{\int_{-\infty}^{\infty} x(t)\cos(2\pi f t)\,\mathrm{d}t}_{\text{real part } \mathrm{Re}(X(f)) = \mathcal{F}\{x_e(t)\}} + j \underbrace{\left((-1)\int_{-\infty}^{\infty} x(t)\sin(2\pi f t)\,\mathrm{d}t\right)}_{\text{imaginary part } \mathrm{Im}(X(f)) = \mathcal{F}\{x_o(t)\}} \tag{5.9}$$

The real part of $X(f)$ is the Fourier transform of the *even* part of a signal, $x_e(t)$. The imaginary part of $X(f)$ is the Fourier transform of the *odd* part of that signal, $x_o(t)$. This is because the product of an even function and an odd function results in an odd function, and the integration of an odd function symmetrically about zero yields zero.

5.6 Double-sided spectrum

The Fourier transform of signal $x(t)$ generally gives a complex-valued, continuous function of frequency, $X(f)$, which can be converted to the double-sided amplitude and phase spectra. For real-valued signals $x(t)$, the subject of this book, the value of its Fourier transform at any frequency f, and corresponding frequency $-f$, are complex conjugate numbers:

$$\boxed{X(-f) = X^*(f)} \tag{5.10}$$

This result is similar to the complex Fourier series, (4.5).

Proof Eq. (5.10) can be proven easily, for real-valued signals $x(t)$:

$$X^*(f) = \int_{-\infty}^{\infty} \left(x(t)e^{-j2\pi f t}\right)^* \mathrm{d}t = \int_{-\infty}^{\infty} x(t)e^{+j2\pi f t}\,\mathrm{d}t = X(-f)$$

For real-valued signals it then follows that $|X(-f)| = |X(f)|$ and $\angle X(-f) = -\angle X(f)$. The amplitude spectrum is an *even* function of frequency, and the phase spectrum an *odd* function of frequency. Note that for the amplitude spectrum we use the modulus or magnitude of $X(f)$, so that 'magnitude spectrum' would be a more appropriate term.

From (5.9) it follows that the Fourier transform of an even signal is a real-valued, even function of frequency. Similarly, the Fourier transform of an odd signal is an imaginary-valued,

odd function of frequency. A signal that has no symmetry properties has a complex-valued Fourier transform. This, and more, will be explained in the following examples.

5.7 Examples

EX 5.1 Compute the Fourier transform of the pulse function $x(t) = \Pi\left(\frac{t}{\tau}\right)$, (2.14).

Solution To compute $X(f)$ we simply substitute $x(t)$ in (5.1):

$$\begin{aligned} X(f) &= \mathcal{F}\left\{\Pi\left(\frac{t}{\tau}\right)\right\} = \int_{-\infty}^{\infty} \Pi\left(\frac{t}{\tau}\right) e^{-j2\pi ft}\,\mathrm{d}t = \int_{-\frac{\tau}{2}}^{\frac{\tau}{2}} 1e^{-j2\pi ft}\,\mathrm{d}t \\ &= \left[\frac{1}{-j2\pi f} e^{-j2\pi ft}\right]_{-\frac{\tau}{2}}^{\frac{\tau}{2}} = -\frac{1}{j2\pi f} \underbrace{\left(e^{-j2\pi f\frac{\tau}{2}} - e^{j2\pi f\frac{\tau}{2}}\right)}_{=-2j\sin(2\pi f\frac{\tau}{2})} \\ &= \frac{\sin(\pi f\tau)}{\pi f}\frac{\tau}{\tau} = \tau\,\mathrm{sinc}(\tau f) \end{aligned}$$

with the sinc function from Appendix B. We obtained our first Fourier transform pair:

$$\boxed{\Pi\left(\frac{t}{\tau}\right) \overset{\mathcal{F}}{\leftrightarrow} \tau\,\mathrm{sinc}(\tau f)} \tag{5.11}$$

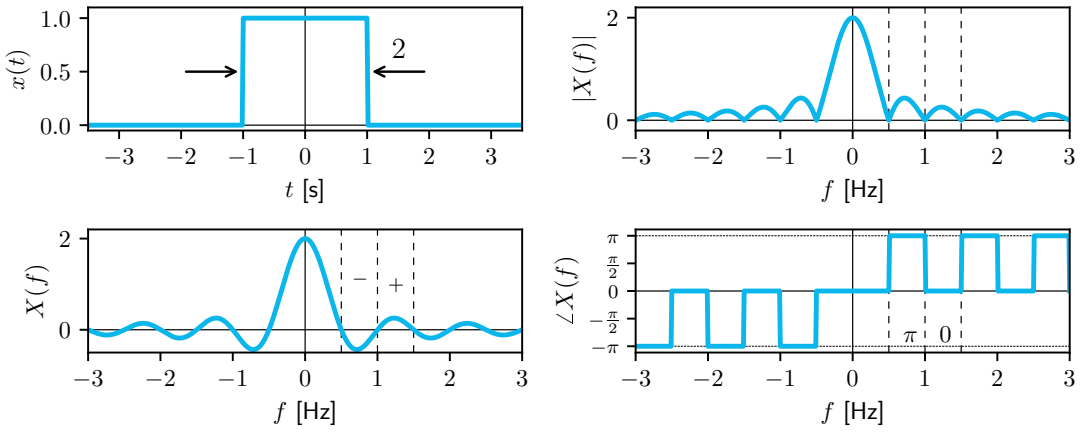


Figure 5.2: Left column: unit pulse function $x(t) = \Pi\left(\frac{t}{2}\right)$ ($\tau = 2$) (top) and its Fourier transform, $X(f) = 2\,\mathrm{sinc}(2f)$ (bottom). Right column: double-sided amplitude (top) and phase spectra (bottom).

Figure 5.2 shows $x(t)$ (top, left), for $\tau = 2$. Signal $x(t)$ is real and even, and so is its Fourier transform, see Figure 5.2 (bottom, left), for $\tau = 2$.

Clearly, $X(0) = 2$, while this signal has a zero average over the interval $(-\infty, \infty)$. The zero-crossings of $\tau\,\mathrm{sinc}(\tau f)$ occur at $f = \pm\frac{1}{\tau}, \pm\frac{2}{\tau}, \ldots$, i.e., with $\tau = 2$ at $f = m\frac{1}{2}$ Hz $\forall\ m \in \mathbb{Z} \setminus \{0\}$.

Figure 5.2 (at right) shows the amplitude (top) and phase (bottom) spectra, which are even and odd functions of frequency, respectively. When $X(f) \geq 0$ (+), this value lies on the positive real axis, and its phase is 0. When $X(f) < 0$ (−), its phase equals $\pm\pi$. Here, we *choose* $\angle X(f)$ to be $+\pi$ when $X(f) < 0, f > 0$. Then, $\angle X(f)$ *must* be $-\pi$ when $X(f) < 0, f < 0$.

EX 5.2

Compute the amplitude and phase specta of the unit doublet $y(t) = \Pi(t-\frac{1}{2}) - \Pi(t+\frac{1}{2})$.

Solution To obtain the spectra, we first compute the Fourier transform of $y(t)$:

$$\begin{aligned} Y(f) &= \int_{-1}^{0} (-1)e^{-j2\pi ft}\,\mathrm{d}t + \int_{0}^{1} 1e^{-j2\pi ft}\,\mathrm{d}t = \frac{1}{j2\pi f}\left([e^{-j2\pi ft}]_{-1}^{0} - [e^{-j2\pi ft}]_{0}^{1}\right) \\ &= \frac{1}{j2\pi f}(2-2\cos(2\pi f)) = \frac{-j}{2\pi f}4\sin^2(\pi f) = -j2\pi f\,\mathrm{sinc}^2(f) \end{aligned} \tag{5.12}$$

Because signal $y(t)$ is real and odd, its Fourier transform is imaginary-valued and odd.

The magnitude equals $2\pi|f|\,\mathrm{sinc}^2(f)$, an even function of frequency. The phase is $\arctan\left(\frac{-2\pi f\,\mathrm{sinc}^2(f)}{0}\right)$ which equals $\arctan(-\infty) = -\frac{\pi}{2}$ for $f > 0$ and $\arctan(+\infty) = +\frac{\pi}{2}$ for $f < 0$. For $f = 0$, $Y(0) = 0$, and the phase is defined to be 0.

Figure 5.3 shows $y(t)$ and $Y(f)$ in the left column (with an imaginary-valued vertical axis), and the spectra in the right column. Note that function $Y(f)$ equals zero at all integer multiples of 1 Hz; at these frequencies its phase should be 0 (not shown).

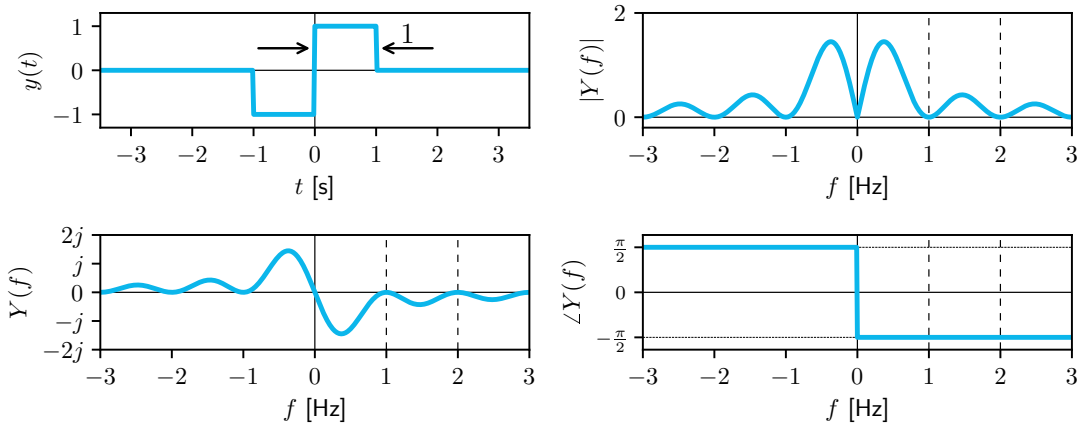


Figure 5.3: Left column: unit doublet $y(t) = \Pi(t-\frac{1}{2}) - \Pi(t+\frac{1}{2})$ (top) and its Fourier transform, $Y(f)$ (bottom). Right column: double-sided amplitude (top) and phase (bottom) spectra.

EX 5.3

Compute the Fourier transform of the exponential decay function $z(t) = e^{-\alpha t}u(t)$ for $\alpha > 0$ and with $u(t)$ the unit step function from Appendix B. Compute its real and imaginary parts and show the amplitude and phase spectra.

Solution The Fourier transform $Z(f)$ can be computed through direct substitution:

$$\begin{aligned} Z(f) &= \int_{-\infty}^{\infty} e^{-\alpha t}u(t)e^{-j2\pi ft}\,\mathrm{d}t = \int_{0}^{\infty} e^{-(\alpha+j2\pi f)t}\,\mathrm{d}t \\ &= \left[-\frac{1}{\alpha+j2\pi f}e^{-(\alpha+j2\pi f)t}\right]_{0}^{\infty} = \frac{1}{\alpha+j2\pi f} \end{aligned} \tag{5.13}$$

where $t \to \infty$ yields $e^{-\alpha t} \downarrow 0$ (for $\alpha > 0$), and $e^{-j2\pi ft}$ is bounded (its magnitude is $1\ \forall\ t$). Because $z(t)$ has no symmetry properties, its Fourier transform $Z(f)$ is a

complex-valued function of f. We compute the real and imaginary parts as follows:

$$Z(f) = \frac{1}{\alpha + j2\pi f}\frac{\alpha - j2\pi f}{\alpha - j2\pi f} = \frac{\alpha - j2\pi f}{\alpha^2 + (2\pi f)^2}$$

Then $\mathrm{Re}(Z(f)) = \frac{\alpha}{\alpha^2+(2\pi f)^2}$ and $\mathrm{Im}(Z(f)) = \frac{-2\pi f}{\alpha^2+(2\pi f)^2}$. The amplitude and phase follow from, respectively, (5.6) and (5.7):

$$|Z(f)| = \frac{1}{\sqrt{\alpha^2 + (2\pi f)^2}} \quad \text{and} \quad \angle Z(f) = \arctan(\frac{-2\pi f}{\alpha})$$

Then $|Z(0)| = \frac{1}{\alpha}$, $|Z(\pm\infty)| = 0$ and $\angle Z(0) = 0$, $\angle Z(\infty) = -\frac{\pi}{2}$, $\angle Z(-\infty) = +\frac{\pi}{2}$.

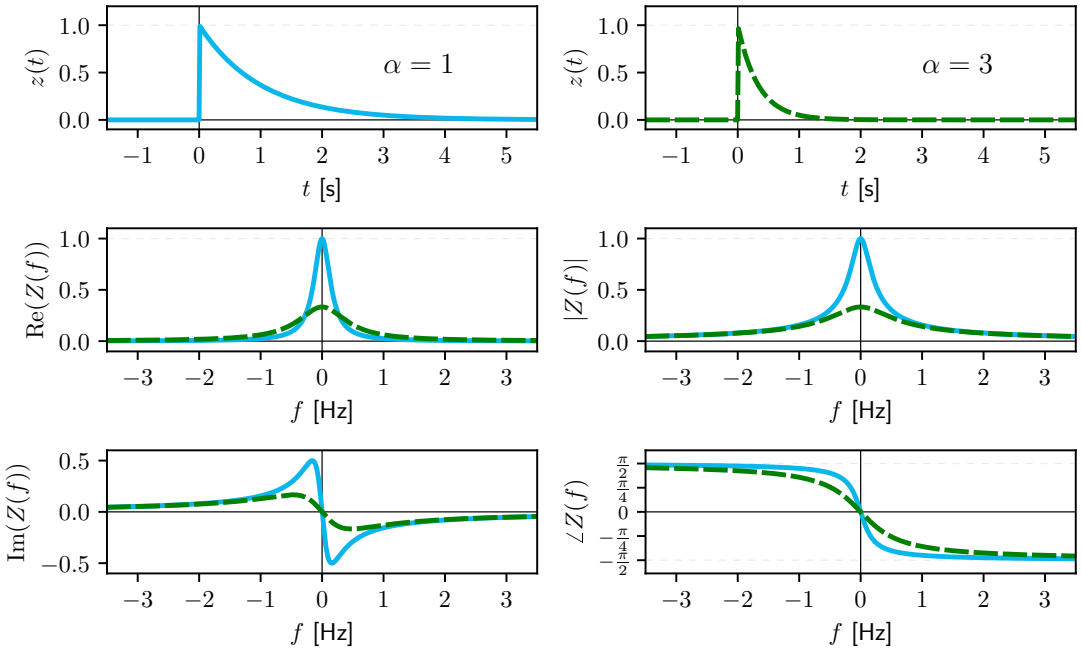


Figure 5.4: Top row: exponential decay function $z(t) = e^{-\alpha t}u(t)$ for $\alpha = 1$ (left) and $\alpha = 3$ (right). The middle and bottom rows show the real and imaginary parts of $Z(f)$ (at left), and the double-sided amplitude and phase spectra (at right).

Figure 5.4 shows $z(t)$ and $Z(f)$, for two values of $\alpha = 1$ (blue) and $\alpha = 3$ (dashed green). The top row shows the time signals; the middle and bottom rows show the real and imaginary parts of $Z(f)$ (at left), and the amplitude and phase spectra (at right). The real part of $Z(f)$ is an even function of frequency; the imaginary part is an odd function of frequency. When α increases, the magnitude of $Z(f)$ is smaller for lower frequencies, but for higher frequencies the value of α does not make a difference.

5.8 Fourier transform in-the-limit

Some signals $x(t)$ fail to meet the Dirichlet conditions, yet have Fourier transforms. In this section, we introduce a few basic and very useful signals belonging to this category.

5.8.1 Dirac delta function and constant

$$\boxed{\delta(t) \overset{\mathcal{F}}{\leftrightarrow} 1} \tag{5.14}$$

Both functions are even, and real-valued. The extremely 'narrow' Dirac delta function in time, when Fourier-transformed, becomes extremely 'broad' in frequency, as $X(f) = 1 \; \forall \; f$.

Proof Although the Dirac function has non-finite discontinuities, its Fourier transform is easily found through direct substitution, then using the sifting property, (B.12):

$$X(f) = \mathcal{F}\{\delta(t)\} = \int_{-\infty}^{\infty} \delta(t) e^{-j2\pi ft} \, \mathrm{d}t = e^0 = 1$$

$$\boxed{1 \overset{\mathcal{F}}{\leftrightarrow} \delta(f)} \tag{5.15}$$

Both functions are even, and real-valued. The extremely 'broad' constant function in time, $x(t) = 1 \; \forall \; t$, when Fourier-transformed, becomes extremely 'narrow' in frequency.

Proof Signal $x(t)$ is clearly not absolutely integrable, yet has a Fourier transform. Again we use the sifting property and now start with the inverse Fourier transform:

$$x(t) = \mathcal{F}^{-1}\{\delta(f)\} = \int_{-\infty}^{\infty} \delta(f) e^{j2\pi ft} \, \mathrm{d}f = e^0 = 1$$

We have obtained two important Fourier transform pairs: $\{\delta(t), 1\}$ and $\{1, \delta(f)\}$. These pairs follow when using the *duality* Fourier theorem, Section 6.5. As the Fourier transform is linear, both sides of these pairs can, for instance, be multiplied by constant A.

5.8.2 Cosine and sine

The cosine function $x(t) = A\cos(2\pi f_0 t)$ has as its Fourier transform:

$$\boxed{X(f) \; = \; \mathcal{F}\{A\cos(2\pi f_0 t)\} \; = \; \frac{A}{2}\Big(\delta(f + f_0) + \delta(f - f_0)\Big)} \tag{5.16}$$

Both functions are even, and real-valued.

Proof We prove (5.16) in two steps. First, a frequency-translated impulse $X_1(f) = \delta(f + f_0)$ is inverse Fourier-transformed:

$$x_1(t) = \int_{-\infty}^{\infty} X_1(f) e^{j2\pi ft} \, \mathrm{d}f = \int_{-\infty}^{\infty} \delta(f + f_0) e^{j2\pi ft} \, \mathrm{d}f = e^{-j2\pi f_0 t}$$

and similarly $X_2(f) = \delta(f - f_0)$:

$$x_2(t) = \int_{-\infty}^{\infty} X_2(f) e^{j2\pi ft} \, \mathrm{d}f = \int_{-\infty}^{\infty} \delta(f - f_0) e^{j2\pi ft} \, \mathrm{d}f = e^{j2\pi f_0 t}$$

We then add these two time signals together, multiply by $\frac{A}{2}$, and use (C.1) to obtain:

$$x(t) = \frac{A}{2}(x_1(t) + x_2(t)) = \frac{A}{2}\left(e^{-j2\pi f_0 t} + e^{j2\pi f_0 t}\right) = A\cos(2\pi f_0 t)$$

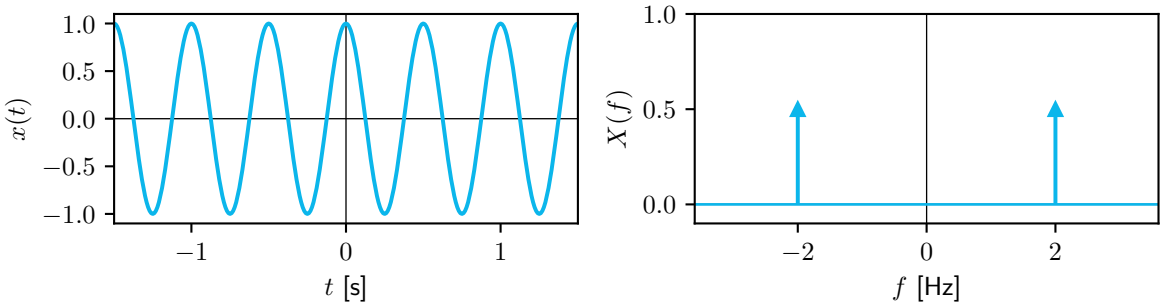


Figure 5.5: Fourier transform in-the-limit. At left signal $x(t) = \cos(2\pi f_0 t)$, with $f_0 = 2$ Hz, as a function of time, and at right its Fourier transform $X(f) = \frac{1}{2}(\delta(f + f_0) + \delta(f - f_0))$, as a function of frequency.

Similarly, adding the two Fourier transforms together and multiplying by $\frac{A}{2}$ yields:

$$X(f) = \frac{A}{2}(X_1(f) + X_2(f)) = \frac{A}{2}(\delta(f + f_0) + \delta(f - f_0))$$

Properties of the Fourier transform, such as linearity and frequency-translation used above, are discussed in detail in Chapter 6. Signal $x(t)$ and its Fourier transform $X(f)$ are shown in Figure 5.5. $X(f)$ is an even, real-valued, continuous function of frequency. It equals zero for all frequencies, except at $\pm f_0$, where it equals a Dirac pulse multiplied by $\frac{A}{2}$ (here $A = 1$).

The Fourier transform of the sine function $x(t) = A\sin(2\pi f_0 t)$ can be obtained similarly:

$$X(f) = \mathcal{F}\{A\sin(2\pi f_0 t)\} = j\frac{A}{2}\Big(\delta(f + f_0) - \delta(f - f_0)\Big) \tag{5.17}$$

Both functions are odd, the Fourier transform is imaginary-valued.

5.8.3 Relation with the complex exponential Fourier series

In the discussion of Fourier transforms in-the-limit, three of the four example signals were in fact *periodic* signals, for which the complex Fourier series coefficients can be (easily) obtained.

First, consider the constant in time, function $x(t) = 1$. Recall that this is indeed a *periodic* function, Section 2.1. Its complex Fourier series has just one non-zero coefficient, $X_0 = 1$, the signal average. This coefficient belongs to $k = 0$, i.e., *only* at (discrete) frequency $f = 0$ we have a non-zero number. In contrast, the Fourier transform of the constant, $X(f)$, is a *continuous* function of frequency. It equals zero at all frequencies, except at $f = 0$, where it has a Dirac pulse multiplied by 1.

Second, consider the cosine function $x(t) = A\cos(2\pi f_0 t)$. In the real Fourier series, it has one non-zero coefficient: $a_1 = A$, at (discrete) frequency $1f_0$. The complex Fourier series has two non-zero coefficients: $X_1 = \frac{A}{2}$ (at $1f_0$) and $X_{-1} = \frac{A}{2}$ (at $-1f_0$). In contrast, the Fourier transform is a *continuous* function of frequency. It is zero at all frequencies, except at $f = \pm f_0$, where it has a Dirac pulse multiplied by $\frac{A}{2}$.

Third, consider the sine function $x(t) = A\sin(2\pi f_0 t)$. In the real Fourier series, it has one non-zero coefficient: $b_1 = A$, at (discrete) frequency $1f_0$. The complex Fourier series has two non-zero coefficients: $X_1 = -j\frac{A}{2}$ (at $1f_0$) and $X_{-1} = j\frac{A}{2}$ (at $-1f_0$). In contrast, the Fourier transform is a *continuous* function of frequency. It is zero at all frequencies, except at $f = \pm f_0$, where it has a Dirac pulse, multiplied by $-j\frac{A}{2}$ (at $1f_0$) or $j\frac{A}{2}$ (at $-1f_0$).

Fourier transforms in-the-limit of *periodic* time signals lead to the inclusion of Dirac delta functions in the frequency domain. When Dirac delta functions appear in the Fourier transform of a signal, that signal is a *power* signal. Section 6.12 will formalize the relation of the Fourier transform of periodic functions with their complex exponential Fourier series.

5.8.4 Examples

The principle of performing a Fourier transform in-the-limit is illustrated using two examples. In the first example, the Dirac function $\delta(t)$ and the constant signal $x(t) = 1$ will both be expressed as limiting cases of a unit pulse function (see Appendix B), and then Fourier-transformed. In the second example, a signal $x(t)$ which is *not* absolutely integrable will be approximated by means of an auxiliary signal $y(t) = x(t)e^{-\alpha|t|}$, for $\alpha > 0$, which *is* absolutely integrable. Whereas $X(f)$ cannot be computed directly, $Y(f)$ can. We then take the limit $\alpha \downarrow 0$, where $\lim_{\alpha \downarrow 0} e^{-\alpha|t|} = 1$. Then $y(t)$ approximates $x(t)$ and $Y(f)$ approximates $X(f)$.

EX 5.4

Express the Dirac function $\delta(t)$ as the limiting case of a scaled pulse function and show that, in-the-limit, its Fourier transform equals $1 \; \forall \; f$. Repeat the analysis for a constant function $y(t) = 1$ and show that, in-the-limit, its transform equals $\delta(f)$.

Solution Using (B.10) we can write $x(t) = \lim_{\epsilon \downarrow 0} \frac{1}{\epsilon} \Pi\left(\frac{t}{\epsilon}\right) = \delta(t)$. Fourier-transforming this equation yields:

$$\mathcal{F}\{\delta(t)\} = \mathcal{F}\{\lim_{\epsilon \downarrow 0} \frac{1}{\epsilon} \Pi\left(\frac{t}{\epsilon}\right)\} = \lim_{\epsilon \downarrow 0} \frac{1}{\epsilon} \epsilon \, \mathrm{sinc}(\epsilon f) = \lim_{\epsilon \downarrow 0} \frac{\sin(\pi \epsilon f)}{\pi \epsilon f} = \frac{0}{0}$$

Using L'Hôpital's[4] rule, we quickly find: $\mathcal{F}\{\delta(t)\} = 1 \; \forall \; f$. This is a heuristic, alternative way to derive (5.14).

In a similar way, the Fourier transform of a constant can be found, using:

$$y(t) = \lim_{\epsilon \to \infty} \Pi\left(\frac{t}{\epsilon}\right) = 1$$

Fourier-transforming this equation yields, see (5.15):

$$\mathcal{F}\{1\} = \mathcal{F}\{\lim_{\epsilon \to \infty} \Pi\left(\frac{t}{\epsilon}\right)\} = \lim_{\epsilon \to \infty} \epsilon \, \mathrm{sinc}(\epsilon f) \approx \delta(f)$$

Figure 5.6 shows the transforms in-the-limit for both functions, for different values of ϵ. The top row, for $x(t) = \frac{1}{\epsilon} \Pi\left(\frac{t}{\epsilon}\right)$, shows that when ϵ is smaller, the function in time approximates the Dirac function better; it becomes narrower and higher, its Fourier transform $X(f) = \mathrm{sinc}(\epsilon f)$ becomes broader and, in the limit, $\epsilon \downarrow 0$, approaches $1 \; \forall \; f$.

The bottom row, for $y(t) = \Pi\left(\frac{t}{\epsilon}\right)$ shows that when ϵ is larger, the function in time becomes broader and its Fourier transform $Y(f) = \epsilon \, \mathrm{sinc}(\epsilon f)$ narrower. At $f = 0$, $Y(0) = \epsilon \, \mathrm{sinc}(\epsilon 0) = \epsilon$, which becomes ∞ when $\epsilon \to \infty$. The zero-crossings of the sinc function occur at $f = k\frac{1}{\epsilon}$, with $k \in \mathbb{Z} \setminus \{0\}$, and get closer and closer to $f = 0$ when $\epsilon \to \infty$.

We obtain a function that has similar properties as the Dirac delta function: zero $\forall \; f$ except for ∞ at $f = 0$, and its integral equal to 1. The latter point can be understood by considering the, by approximation, triangular shape of the sinc function around $f = 0$ (bottom right plot, in red): its area equals $\frac{1}{2} \frac{2}{\epsilon} \epsilon = 1$.

[4] After the French mathematician G.F.A. marquis de L'Hôpital (1661-1704).

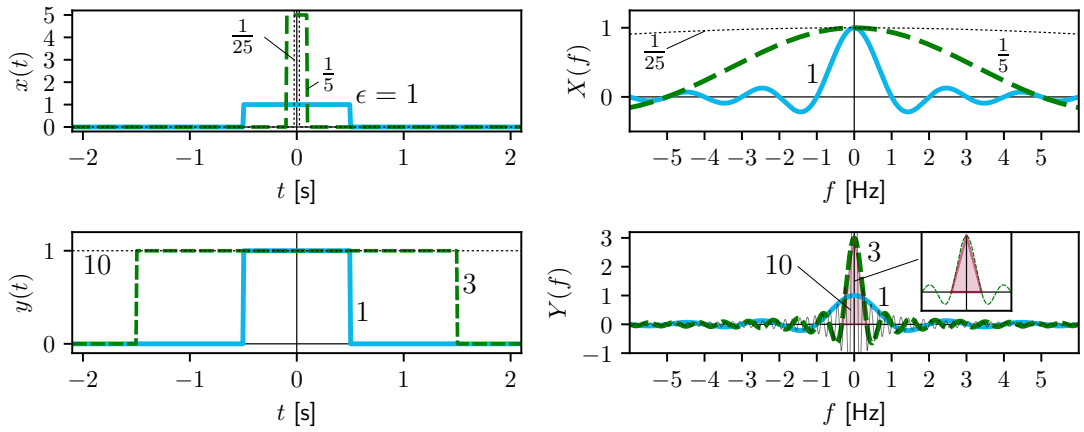


Figure 5.6: Left column: time signals $x(t) = \frac{1}{\epsilon}\Pi\left(\frac{t}{\epsilon}\right)$ (top), and $y(t) = \Pi\left(\frac{t}{\epsilon}\right)$ (bottom). The right column shows their Fourier transforms. Signal $x(t)$ and $X(f)$ are shown for $\epsilon = 1$ (blue), $\frac{1}{5}$ (green, dashed) and $\frac{1}{25}$ (black, dotted). Signal $y(t)$ and $Y(f)$ are shown for $\epsilon = 1$ (blue), 3 (green, dashed) and 10 (black, dotted). The bottom right plot shows the triangle (red) referred to in the text.

EX 5.5

Compute the Fourier transform of signal $v(t) = -\frac{1}{2}e^{\alpha t}u(-t) + \frac{1}{2}e^{-\alpha t}u(t)$.[a] Show that, in-the-limit, we can compute the Fourier transform of the unit step function $u(t)$ using the expression $u(t) = v(t) + \frac{1}{2}$.

[a] For $\alpha = 0$, signal $\text{sgn}(t) = 2v(t)$ is known as the *sign* or *signum* signal.

Solution Recall Example 5.3 where a similar function $z(t)$ was Fourier-transformed, for $\alpha > 0$. The Fourier transform theorems, discussed in Chapter 6, would allow for a quick calculation of $V(f)$ from $Z(f)$; instead, we apply direct substitution here:

$$\begin{aligned} V(f) &= -\frac{1}{2}\int_{-\infty}^{0} e^{\alpha t}e^{-j2\pi ft}\,\mathrm{d}t + \frac{1}{2}\int_{0}^{\infty} e^{-\alpha t}e^{-j2\pi ft}\,\mathrm{d}t \\ &= -\frac{1}{2}\frac{1}{\alpha - j2\pi f} + \frac{1}{2}\frac{1}{\alpha + j2\pi f} = \frac{-j2\pi f}{\alpha^2 + (2\pi f)^2} \end{aligned}$$

In the limit case, when $\alpha \downarrow 0$, we get $V(f) = -\frac{j}{2\pi f} = \frac{1}{j2\pi f}$.

Now, it is given that $u(t) = v(t) + \frac{1}{2}$, so $U(f) = V(f) + \frac{1}{2}\delta(f)$, and we obtain the Fourier transform of the unit step function $u(t)$:

$$U(f) = \mathcal{F}\{u(t)\} = \frac{1}{2}\delta(f) + \frac{1}{j2\pi f} \tag{5.18}$$

Compare this result with the Fourier transform of $z(t) = e^{-\alpha t}u(t)$ for $\alpha > 0$, $Z(f) = \frac{1}{\alpha + j2\pi f}$. Apparently, we *cannot* obtain the Fourier transform of $u(t)$ by simply substituting $\alpha = 0$ in the expression for $Z(f)$! This is because $Z(f)$ can be calculated in the manner performed in Example 5.3 *only* for $\alpha > 0$. When $\alpha > 0$, $z(t)$ and $v(t)$ are energy signals. When $\alpha = 0$, $z(t)$ and $v(t)$ become *power* signals, with averages $\frac{1}{2}$ and 0, respectively. Clearly, we must be careful with Fourier transforms in-the-limit.

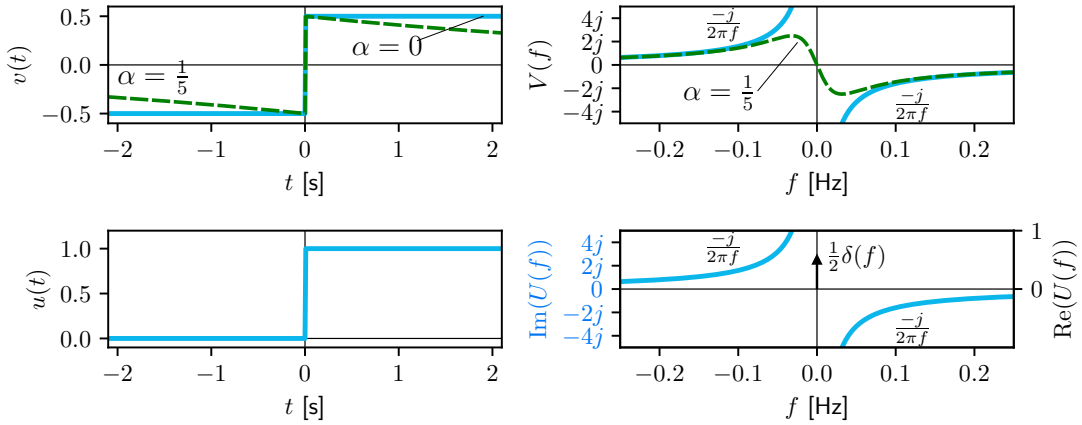


Figure 5.7: Fourier transform in-the-limit of signal $v(t) = -\frac{1}{2}e^{\alpha t}u(-t) + \frac{1}{2}e^{-\alpha t}u(t)$ (top row) and the unit step $u(t)$ (bottom row). Signal $v(t)$ is shown for $\alpha = 0$ (blue), and $\alpha = \frac{1}{5}$ (green). While $V(f)$ only has an imaginary part, $U(f)$ is complex-valued. Its real part is shown in black, its imaginary part in blue.

Figure 5.7 shows $v(t)$ and $u(t)$ at left and their Fourier transforms $V(f)$ and $U(f)$ at right. Signal $v(t)$ is shown for two values of α, namely $\alpha = 0$ and $\alpha = \frac{1}{5}$.

5.9 Parseval: Fourier transform

The Fourier transform exists for signals that fulfill the Dirichlet conditions, Section 5.3. Generally, these signals are *energy* signals: energy is limited, the average power is zero. Energy can be calculated in time, but also in frequency, using Parseval's theorem (Section 13.2.2):

$$E = \int_{-\infty}^{\infty} x^2(t)\,\mathrm{d}t = \int_{-\infty}^{\infty} |X(f)|^2\,\mathrm{d}f \tag{5.19}$$

For a signal with unit [unit], the energy has unit [unit^2 s]. The fact that we *integrate* $|X(f)|^2$ over all frequencies f indicates that, like $|X(f)|$, $|X(f)|^2$ is also a *density* function.

Compute the energy of signal $z(t)$, from Example 5.3; its unit is [V]. EX 5.6

Solution Although computing this signal's energy is easier using (2.17), see Example 2.9, here we compute it in the frequency domain, using (5.19):

$$E_z = \int_{-\infty}^{\infty} |Z(f)|^2\,\mathrm{d}f = \int_{-\infty}^{\infty} \frac{1}{\alpha^2 + (2\pi f)^2}\,\mathrm{d}f = \frac{2}{\alpha^2}\int_{0}^{\infty} \frac{1}{1 + (\frac{2\pi f}{\alpha})^2}\,\mathrm{d}f$$

$$= \frac{2}{\alpha^2}\frac{\alpha}{2\pi}\int_{0}^{\infty} \frac{1}{1+\nu^2}\,\mathrm{d}\nu = \frac{1}{\pi\alpha}\left[\arctan(\nu)\right]_0^{\infty} = \frac{1}{\pi\alpha}\frac{\pi}{2} = \frac{1}{2\alpha}\ \mathrm{J}$$

In the first line we used the fact that the integrand is an even function of frequency (and so are *all* amplitude spectra of real-valued signals) and in the second line we used $\nu = \frac{2\pi f}{\alpha}$, so $\mathrm{d}f = \frac{\alpha}{2\pi}\mathrm{d}\nu$. When $\alpha \to 0$, energy becomes infinite and $z(t)$ is a power signal. These results correspond to the answers obtained in Example 2.9 for $A = 1$.

EX 5.7 Compute the energy of $w(t) = \Pi\left(\frac{t}{2}\right) + 1$; its unit is [V].

Solution The Fourier transform can be swiftly obtained from (5.11) and (5.15):

$$W(f) = 2\,\mathrm{sinc}(2f) + \delta(f)$$

where we used the property that the Fourier transform is a linear operation. A Dirac pulse appears in $W(f)$: this signal has infinite energy, it is a *power* signal and its average power P_w is $1^2 = 1$ W. The average of this signal equals 1; the pulse $\Pi\left(\frac{t}{2}\right)$ at $t = 0$ is irrelevant. Recall Example 2.11, with here $A = 1$, $\tau = 2$ and $a = 1$.

EX 5.8 Compute the energy of unit pulse $x(t) = \Pi\left(\frac{t}{\tau}\right)$ (unit [V], with $\tau = 1$) in time and frequency. In frequency, compute the contribution to the energy of the unit pulse from: (a) all frequencies between ± 1 Hz, and (b) *at* frequency 0.5 Hz.

Solution The energy of a unit pulse $x(t) = \Pi\left(\frac{t}{\tau}\right)$ can be easily calculated in time: $E_x = (1)^2\tau = \tau$ J. In frequency, we use (5.19):

$$\begin{aligned} E_x &= \int_{-\infty}^{\infty} |X(f)|^2\,\mathrm{d}f = \int_{-\infty}^{\infty} |\tau\,\mathrm{sinc}(\tau f)|^2\,\mathrm{d}f = \tau^2 \int_{-\infty}^{\infty} \left(\frac{\sin(\pi\tau f)}{\pi\tau f}\right)^2 \mathrm{d}f \\ &= 2\tau^2 \frac{1}{\pi\tau} \int_{0}^{\infty} \left(\frac{\sin(\pi\tau f)}{\pi\tau f}\right)^2 \mathrm{d}(\pi\tau f) = 2\tau^2 \frac{1}{\pi\tau}\frac{\pi}{2} = \tau \text{ J} \end{aligned}$$

Here, we used the fact that the sinc function is an even function of frequency, and (A.15). For $\tau = 1$, the energy is 1 J. Energy in time equals energy in frequency.

To calculate the contribution to the energy of the unit pulse from all frequencies between ± 1 Hz (a), consider Figure 5.2 (bottom, left). When starting at $f = 0$ and moving left or right on the frequency axis, $X(f)$ has its first zero-crossings at $f = \pm\frac{1}{\tau} = \pm 1$ Hz. Hence, we are requested to compute the contribution of the central lobe of $X(f)$, centered at 0 Hz, to the total energy of $x(t)$. We obtain:

$$E_x|_{f\in[-1,1]} = \int_{-1}^{1} |X(f)|^2\,\mathrm{d}f = 2\int_{0}^{1} \left(\frac{\sin(\pi f)}{\pi f}\right)^2 \mathrm{d}f \approx 0.9028 \text{ J}$$

through numerical integration. The central lobe contains $\approx 90.3\%$ of the pulse energy.

To calculate the contribution to the energy of the pulse *at* 0.5 Hz (b), (5.19) shows that we integrate $|(X(f)|^2$ from $f = 0.5$ to 0.5, which results in 0 J.

This, and the previous calculation, illustrates that $|X(f)|^2$ is a *density* function. *Only* when a signal's Fourier transform $X(f)$ contains a Dirac pulse at some frequency f_0, which makes that signal a power signal, we obtain a non-zero contribution *at* that frequency f_0 to the signal's average power.

With the Fourier transform and its inverse defined, the following chapter will discuss the main Fourier transform theorems. These will prove to be essential tools in signal analysis.

6

Fourier transform theorems

In the previous chapter, we introduced the Fourier transform, allowing a description of a continuous-time signal $x(t)$ as a continuous-frequency function $X(f)$:

$$X(f) = \mathcal{F}\{x(t)\} = \int_{-\infty}^{\infty} x(t)e^{-j2\pi ft}\,\mathrm{d}t \tag{5.1}$$

Suppose we have the Fourier transform of signal $x(t)$ available, and we require the Fourier transform of a signal $y(t)$, which is obtained by applying some operation on $x(t)$. Examples are applying a time shift, i.e., $y(t) = x(t - t_0)$, or a time scaling, $y(t) = x(at)$. Then, is it possible to obtain the Fourier transform of $y(t)$, in an easy way, rather than having to evaluate the Fourier integral again? Fortunately, this is indeed the case, and in this chapter we demonstrate, for a collection of basic operations on a signal in the time domain, how these operations carry over to the Fourier transform of that signal, in the frequency domain. As will become clear in the remainder of this book, the Fourier transform theorems will prove to be instrumental in signal analysis in continuous time as well as in discrete time.

6.1 Linearity

$$a_1x_1(t) + a_2x_2(t) \overset{\mathcal{F}}{\leftrightarrow} a_1X_1(f) + a_2X_2(f) \tag{6.1}$$

The Fourier transform is a linear operation. The Fourier transform of a linear combination (also called *superposition*) of two (or more) signals $x_1(t)$ and $x_2(t)$, with, respectively, Fourier transforms $X_1(f)$ and $X_2(f)$, is just this linear combination of the Fourier transforms.

Proof Fourier-transforming $y(t) = a_1x_1(t) + a_2x_2(t)$ (with a_1 and a_2 arbitrary constants), we obtain through direct substitution:

$$\begin{aligned} Y(f) &= \int_{-\infty}^{\infty} y(t)e^{-j2\pi ft}\,\mathrm{d}t = \int_{-\infty}^{\infty} \left(a_1x_1(t) + a_2x_2(t)\right)e^{-j2\pi ft}\,\mathrm{d}t \\ &= a_1\int_{-\infty}^{\infty} x_1(t)e^{-j2\pi ft}\,\mathrm{d}t + a_2\int_{-\infty}^{\infty} x_2(t)e^{-j2\pi ft}\,\mathrm{d}t = a_1X_1(f) + a_2X_2(f) \end{aligned}$$

We used this property in Chapter 5 several times, for instance while explaining the symmetry properties, in Section 5.5, where $\mathcal{F}\{x(t)\} = \mathcal{F}\{x_e(t) + x_o(t)\} = \mathcal{F}\{x_e(t)\} + \mathcal{F}\{x_o(t)\}$.

6.2 Time shift

$$x(t+t_0) \overset{\mathcal{F}}{\leftrightarrow} X(f)e^{j2\pi f t_0} \tag{6.2}$$

A time shift (a time delay $t_0 < 0$, or a time advance $t_0 > 0$) of signal $x(t)$ in time causes its Fourier transform $X(f)$ to be multiplied by the complex exponential $e^{j2\pi f t_0}$. Since the magnitude of the latter is $1\ \forall\ f$, the magnitude of $X(f)$ is not affected, only its phase changes.

Proof With $y(t) = x(t+t_0)$, the Fourier transform reads:

$$Y(f) = \int_{-\infty}^{\infty} y(t)e^{-j2\pi f t}\,\mathrm{d}t = \int_{-\infty}^{\infty} x(t+t_0)e^{-j2\pi f t}\,\mathrm{d}t$$

Substitute $t+t_0 = \tau$ then $t = \tau - t_0$, $\mathrm{d}t = \mathrm{d}\tau$. When $t \to -\infty$ then $\tau \to -\infty$, and when $t \to \infty$ then $\tau \to \infty$. We obtain:

$$\begin{aligned} Y(f) &= \int_{-\infty}^{\infty} x(\tau)e^{-j2\pi f(\tau-t_0)}\,\mathrm{d}\tau = \int_{-\infty}^{\infty} x(\tau)e^{-j2\pi f\tau}e^{j2\pi f t_0}\,\mathrm{d}\tau \\ &= \int_{-\infty}^{\infty} x(\tau)e^{-j2\pi f\tau}\,\mathrm{d}\tau\ e^{j2\pi f t_0} = X(f)e^{j2\pi f t_0} \end{aligned}$$

where in the second line we used the fact that the complex exponential $e^{j2\pi f t_0}$ does not depend on τ and can therefore be taken out of the integral in τ.

The magnitude is invariant under a time shift: $|Y(f)| = |X(f)|$, but its phase has changed: $\arg(Y(f)) = \arg(X(f)) + 2\pi f t_0$. The effect is referred to as a phase rotation. Compare with (3.15) and (4.13).

6.3 Time scale change

$$x(at) \overset{\mathcal{F}}{\leftrightarrow} \frac{1}{|a|}X\left(\frac{f}{a}\right) \text{ for } a \neq 0 \tag{6.3}$$

Changing the scale of the running variable (time t) inversely scales the running variable (frequency f) of the Fourier transform. When $|a| > 1$ the signal gets compressed in time, and expanded in frequency; conversely, when $|a| < 1$ the signal gets expanded in time, and compressed in frequency. Recall Section 2.4 and the rules of signal operations.

Proof With $y(t) = x(at)$, the Fourier transform reads:

$$Y(f) = \int_{-\infty}^{\infty} y(t)e^{-j2\pi f t}\,\mathrm{d}t = \int_{-\infty}^{\infty} x(at)e^{-j2\pi f t}\,\mathrm{d}t$$

We distinguish two cases: $a > 0$ and $a < 0$. For positive a, we apply a change of variable $\tau = at$; then $t = \frac{\tau}{a}$ and $\mathrm{d}t = \frac{1}{a}\mathrm{d}\tau$:

$$Y(f) = \int_{\tau=-\infty}^{\infty} x(\tau)e^{-j2\pi f\frac{\tau}{a}}\frac{1}{a}\,\mathrm{d}\tau = \frac{1}{a}\int_{-\infty}^{\infty} x(\tau)e^{-j2\pi\frac{f}{a}\tau}\,\mathrm{d}\tau$$

which, with $\frac{f}{a}$ the frequency variable in the Fourier transform of $x(t)$, yields:

$$Y(f) = \frac{1}{a}X\left(\frac{f}{a}\right)$$

For negative a, we recognize that $at = -|a|t$:

$$Y(f) = \int_{-\infty}^{\infty} x(-|a|t)e^{-j2\pi ft}\,\mathrm{d}t$$

Substitute $-|a|t = \tau$, then $t = \frac{\tau}{-|a|}$ and $\mathrm{d}t = -\frac{1}{|a|}\mathrm{d}\tau$; when $t \to -\infty$ then $\tau \to +\infty$; when $t \to \infty$ then $\tau \to -\infty$:

$$Y(f) = \int_{\tau=\infty}^{-\infty} x(\tau)e^{j2\pi f\frac{\tau}{|a|}}(-\tfrac{1}{|a|})\,\mathrm{d}\tau = \tfrac{1}{|a|}\int_{\tau=-\infty}^{\tau=\infty} x(\tau)e^{-j2\pi\frac{-f}{|a|}\tau}\,\mathrm{d}\tau$$

which, with $-\frac{f}{|a|}$ the frequency variable in the Fourier transform of $x(t)$, yields:

$$Y(f) = \tfrac{1}{|a|}X(-\tfrac{f}{|a|}) = \tfrac{1}{|a|}X(\tfrac{f}{a})$$

as here $a < 0$. Combining the results for $a > 0$ and $a < 0$ leads to (6.3).

6.4 Time reversal

$$\boxed{x(-t) \overset{\mathcal{F}}{\leftrightarrow} X(-f)} \tag{6.4}$$

This theorem is a special case of the scale change theorem, (6.3), with $a = -1$.

6.5 Duality

$$\boxed{X(t) \overset{\mathcal{F}}{\leftrightarrow} x(-f)} \tag{6.5}$$

In other words, if $\{x(t), X(f)\}$ is a Fourier transform pair, then $\{X(t), x(-f)\}$ is also a Fourier transform pair. For every transform that we find, duality gives us *two* transform pairs.

Proof The expression for the Fourier transform, (5.1), and the one for the inverse Fourier transform, (5.2), are very similar. The main differences are the variable of integration and the sign of the complex exponential function. This suggests some sort of symmetry in the Fourier transform pair, referred to as duality. Once we have:

$$X(f) = \int_{t=-\infty}^{\infty} x(t)e^{-j2\pi ft}\,\mathrm{d}t$$

we can change variables as $f = u$ and $t = v$, and the following relation holds:

$$X(u) = \int_{v=-\infty}^{\infty} x(v)e^{-j2\pi uv}\,\mathrm{d}v$$

Now we change variables again, according to $u = t$ and $v = -f$, and we get:

$$X(t) = \int_{-f=-\infty}^{\infty} x(-f)e^{j2\pi tf}\,\mathrm{d}(-f) = -\int_{f=\infty}^{-\infty} x(-f)e^{j2\pi ft}\,\mathrm{d}f = \int_{-\infty}^{\infty} x(-f)e^{j2\pi ft}\,\mathrm{d}f$$

where the last expression on the right is the inverse Fourier transform of $x(-f)$, the result of which is $X(t)$ as a function of time.

6.6 Frequency translation

$$x(t)e^{j2\pi f_0 t} \overset{\mathcal{F}}{\leftrightarrow} X(f-f_0) \tag{6.6}$$

This theorem is also known as the frequency shift theorem. It forms a *dual* of the time shift theorem, (6.2). Note that multiplying a real-valued signal $x(t)$, as assumed in this book, by $e^{j2\pi f_0 t}$, results in a complex-valued signal.

Proof When we take the Fourier transform of $y(t) = x(t)e^{j2\pi f_0 t}$ we find:

$$\begin{aligned} Y(f) &= \int_{-\infty}^{\infty} y(t)e^{-j2\pi ft}\,\mathrm{d}t = \int_{-\infty}^{\infty} x(t)e^{j2\pi f_0 t}e^{-j2\pi ft}\,\mathrm{d}t \\ &= \int_{-\infty}^{\infty} x(t)e^{-j2\pi(f-f_0)t}\,\mathrm{d}t \;=\; X(f-f_0) \end{aligned}$$

such that $X(f)$ has been frequency-shifted by f_0.

6.7 Modulation

$$x(t)\cos(2\pi f_0 t) \overset{\mathcal{F}}{\leftrightarrow} \tfrac{1}{2}X(f+f_0) + \tfrac{1}{2}X(f-f_0) \tag{6.7}$$

Signal $x(t)$ is said to be (amplitude-)*modulated* onto the carrier signal $\cos(2\pi f_0 t)$.

Proof With Euler's formula, (C.1), we obtain:

$$\cos(2\pi f_0 t) = \tfrac{1}{2}\left(e^{j2\pi f_0 t} + e^{-j2\pi f_0 t}\right)$$

The Fourier transform of $y(t) = x(t)\cos(2\pi f_0 t)$ can be found by applying the frequency translation theorem, (6.6), twice. This shifts one copy of $X(f)$ by f_0 and shifts one copy of $X(f)$ by $-f_0$. Both copies are multiplied by $\frac{1}{2}$:

$$Y(f) = \tfrac{1}{2}\left(X(f-f_0) + X(f+f_0)\right) = \tfrac{1}{2}X(f+f_0) + \tfrac{1}{2}X(f-f_0)$$

6.8 Convolution

$$x_1(t) * x_2(t) \overset{\mathcal{F}}{\leftrightarrow} X_1(f)X_2(f) \tag{6.8}$$

The convolution of two signals $y(t) = x_1(t) * x_2(t)$ is an operation defined by the following integral (as will be explained in more detail in Chapter 7):

$$y(t) = \int_{-\infty}^{\infty} x_1(\tau)x_2(t-\tau)\,\mathrm{d}\tau$$

The convolution integral will play an important role in the analysis of linear time-invariant systems, Chapter 16.

Proof Starting with the convolution integral:

$$y(t) = \int_{-\infty}^{\infty} x_1(\tau)x_2(t-\tau)\,\mathrm{d}\tau$$

we substitute for $x_2(t-\tau)$ the inverse Fourier transform of $X_2(f)$ (using $t-\tau$ as the time variable, rather than t):

$$x_2(t-\tau) = \int_{-\infty}^{\infty} X_2(f) e^{j2\pi f(t-\tau)}\,\mathrm{d}f$$

We obtain:

$$y(t) = \int_{-\infty}^{\infty} x_1(\tau) \int_{-\infty}^{\infty} X_2(f) e^{j2\pi f(t-\tau)}\,\mathrm{d}f\,\mathrm{d}\tau = \int_{\tau=-\infty}^{\infty} \int_{f=-\infty}^{\infty} x_1(\tau) X_2(f) e^{j2\pi ft} e^{-j2\pi f\tau}\,\mathrm{d}f\,\mathrm{d}\tau$$

where we moved the term $x_1(\tau)$ inside the integral in f. Change the order of integration:

$$y(t) = \int_{f=-\infty}^{\infty} \int_{\tau=-\infty}^{\infty} x_1(\tau) X_2(f) e^{j2\pi ft} e^{-j2\pi f\tau}\,\mathrm{d}\tau\,\mathrm{d}f = \int_{-\infty}^{\infty} X_2(f) e^{j2\pi ft} \int_{-\infty}^{\infty} x_1(\tau) e^{-j2\pi f\tau}\,\mathrm{d}\tau\,\mathrm{d}f$$

where we moved the term $X_2(f)e^{j2\pi ft}$ outside the integral in τ. The inner integral in τ equals the Fourier transform of $x_1(\tau)$:

$$y(t) = \int_{f=-\infty}^{\infty} X_2(f) e^{j2\pi ft} X_1(f)\,\mathrm{d}f = \int_{-\infty}^{\infty} X_1(f) X_2(f) e^{j2\pi ft}\,\mathrm{d}f$$

which presents the inverse Fourier transform of $Y(f) = X_1(f)X_2(f)$.

We often explain this theorem as 'a convolution in the time domain is equivalent to a multiplication in the frequency domain'. Its dual is the multiplication theorem, discussed next.

6.9 Multiplication

$$x_1(t)x_2(t) \overset{\mathcal{F}}{\leftrightarrow} X_1(f) * X_2(f) \tag{6.9}$$

We explain this theorem as 'a multiplication in the time domain is equivalent to a convolution in the frequency domain'.

Proof The convolution of two signals in the frequency domain reads:

$$Y(f) = X_1(f) * X_2(f) = \int_{-\infty}^{\infty} X_1(\nu) X_2(f-\nu)\,\mathrm{d}\nu$$

For $X_2(f-\nu)$ we substitute the Fourier transform of $x_2(t)$ (using $f-\nu$ as the frequency variable, rather than f):

$$X_2(f-\nu) = \int_{-\infty}^{\infty} x_2(t) e^{-j2\pi(f-\nu)t}\,\mathrm{d}t$$

We obtain:

$$Y(f) = \int_{-\infty}^{\infty} X_1(\nu) \int_{-\infty}^{\infty} x_2(t) e^{-j2\pi(f-\nu)t}\,\mathrm{d}t\,\mathrm{d}\nu = \int_{\nu=-\infty}^{\infty} \int_{t=-\infty}^{\infty} X_1(\nu) x_2(t) e^{-j2\pi ft} e^{j2\pi\nu t}\,\mathrm{d}t\,\mathrm{d}\nu$$

where we moved the term $X_1(\nu)$ inside the integral in t. Change the order of integration:

$$Y(f) = \int_{t=-\infty}^{\infty} \int_{\nu=-\infty}^{\infty} X_1(\nu) x_2(t) e^{-j2\pi ft} e^{j2\pi\nu t}\,\mathrm{d}\nu\,\mathrm{d}t = \int_{-\infty}^{\infty} x_2(t) e^{-j2\pi ft} \int_{-\infty}^{\infty} X_1(\nu) e^{j2\pi\nu t}\,\mathrm{d}\nu\,\mathrm{d}t$$

where we moved the term $x_2(t)e^{-j2\pi ft}$ outside the integral in ν. The inner integral in ν equals the inverse Fourier transform of $x_1(t)$:

$$Y(f) = \int_{t=-\infty}^{\infty} x_2(t)e^{-j2\pi ft}x_1(t)\,\mathrm{d}t = \int_{-\infty}^{\infty} x_1(t)x_2(t)e^{-j2\pi ft}\,\mathrm{d}t$$

which presents the Fourier transform of $y(t) = x_1(t)x_2(t)$.

6.10 Differentiation

$$\boxed{\frac{\mathrm{d}^n x(t)}{\mathrm{d}t^n} \overset{\mathcal{F}}{\leftrightarrow} (j2\pi f)^n X(f)} \tag{6.10}$$

This Fourier theorem holds, assuming that the derivative(s) exist(s).

Proof With $y(t) = \frac{\mathrm{d}x(t)}{\mathrm{d}t}$, its Fourier transform is found through:

$$Y(f) = \int_{-\infty}^{\infty} y(t)e^{-j2\pi ft}\,\mathrm{d}t = \int_{-\infty}^{\infty} \frac{\mathrm{d}x(t)}{\mathrm{d}t}e^{-j2\pi ft}\,\mathrm{d}t$$

We then use integration by parts:

$$Y(f) = \left[x(t)e^{-j2\pi ft}\right]_{-\infty}^{\infty} + j2\pi f \underbrace{\int_{-\infty}^{\infty} x(t)e^{-j2\pi ft}\,\mathrm{d}t}_{=X(f)} = (j2\pi f)X(f)$$

The first term equals zero, because $x(t)$ has to be absolutely integrable (Dirichlet condition 3, (5.3)), hence $x(t)$ has to go to zero both for $t \to -\infty$ and $t \to \infty$, and $e^{-j2\pi ft}$ is bounded. Integration by parts can be applied repeatedly to prove (6.10) for n differentiations.

6.11 Integration

$$\boxed{\int_{-\infty}^{t} x(\tau)\,\mathrm{d}\tau \overset{\mathcal{F}}{\leftrightarrow} \frac{1}{j2\pi f}X(f) + \frac{1}{2}X(0)\delta(f)} \tag{6.11}$$

Note that when $X(0) = \int_{-\infty}^{\infty} x(t)\mathrm{d}t = 0$, only the first part remains, implying the direct counterpart of (6.10).

Proof Integration of $x(t)$ from $-\infty$ to t can be represented as the convolution of $x(t)$ with the unit step function $u(t)$:

$$y(t) = x(t) * u(t) = \int_{-\infty}^{\infty} x(\tau)u(t-\tau)\,\mathrm{d}\tau = \int_{-\infty}^{t} x(\tau)\,\mathrm{d}\tau$$

Fourier-transforming this equation and using the convolution theorem, (6.8), yields:

$$\mathcal{F}\{\int_{-\infty}^{t} x(\tau)\,\mathrm{d}\tau\} = \mathcal{F}\{x(t) * u(t)\} = X(f)U(f)$$

In Example 5.5, using a Fourier transform in-the-limit, we derived $U(f)$:

$$U(f) = \mathcal{F}\{u(t)\} = \tfrac{1}{2}\delta(f) + \frac{1}{j2\pi f} \tag{5.18}$$

Multiplication by $X(f)$:

$$\mathcal{F}\{\int_{-\infty}^{t} x(\tau)\,\mathrm{d}\tau\} = X(f)U(f) = X(f)\left(\tfrac{1}{2}\delta(f) + \tfrac{1}{j2\pi f}\right) = \frac{X(0)}{2}\delta(f) + \frac{X(f)}{j2\pi f}$$

6.12 Relation with the complex exponential Fourier series

For *periodic* signals $x(t)$, with fundamental frequency f_0 and (finite) period T_0, the following relationship holds between the Fourier transform and the complex Fourier series:

$$x(t) \overset{\mathcal{F}}{\leftrightarrow} \sum_{k=-\infty}^{\infty} X_k \delta(f - kf_0) \tag{6.12}$$

Here, the X_k's refer to the complex Fourier series coefficients of periodic signal $x(t)$, (4.3). This relation formalizes the discussion in Section 5.8.3.

Proof The Fourier series representation of *periodic* signal $x(t)$ equals:

$$x(t) = \sum_{k=-\infty}^{\infty} X_k e^{j2\pi k f_0 t} \tag{4.2}$$

Each complex exponential basis function with frequency kf_0 can be written as an integral of a Dirac delta function, positioned at $f = kf_0$, using the sifting property, (B.12):

$$e^{j2\pi k f_0 t} = \int_{-\infty}^{\infty} \delta(f - kf_0) e^{j2\pi f t}\,\mathrm{d}f$$

Substituting in the Fourier series representation, we obtain:

$$x(t) = \sum_{k=-\infty}^{\infty} X_k \int_{-\infty}^{\infty} \delta(f - kf_0) e^{j2\pi f t}\,\mathrm{d}f = \sum_{k=-\infty}^{\infty} \int_{-\infty}^{\infty} X_k \delta(f - kf_0) e^{j2\pi f t}\,\mathrm{d}f$$

Changing the order of summation and integration yields:

$$x(t) = \int_{-\infty}^{\infty} \sum_{k=-\infty}^{\infty} X_k \delta(f - kf_0) e^{j2\pi f t}\,\mathrm{d}f$$

which is the inverse Fourier transform of $\sum_{k=-\infty}^{\infty} X_k \delta(f - kf_0)$.

6.13 Examples

In this section two examples are presented on the use of the Fourier transform theorems.

EX 6.1

In Chapter 5, the Fourier transform of pulse function $x(t) = \Pi\left(\frac{t}{\tau}\right)$ was computed:

$$x(t) = \Pi\left(\frac{t}{\tau}\right) \overset{\mathcal{F}}{\leftrightarrow} X(f) = \tau\,\mathrm{sinc}(\tau f) \tag{5.11}$$

Using duality, obtain the Fourier transform of $x(t) = \tau\,\mathrm{sinc}(\tau t)$.

Solution We have the Fourier transform pair $\{x(t), X(f)\}$, then duality states that $\{X(t), x(-f)\}$ is also a Fourier transform pair. Hence:

$$X(t) = \tau\,\mathrm{sinc}(\tau t) \overset{\mathcal{F}}{\leftrightarrow} x(-f) = \Pi\left(\frac{-f}{\tau}\right)$$

Because $\Pi\left(\frac{-f}{\tau}\right)$ is even, $x(-f) = x(f)$. The second pair is then:

$$x(t) = \tau\,\mathrm{sinc}(\tau t) \overset{\mathcal{F}}{\leftrightarrow} X(f) = \Pi\left(\frac{f}{\tau}\right) \tag{6.13}$$

In words: 'a pulse in time becomes a sinc in frequency; a pulse in frequency becomes a sinc in time'. Figure 6.1 illustrates this fundamental result, for $\tau = 2$.

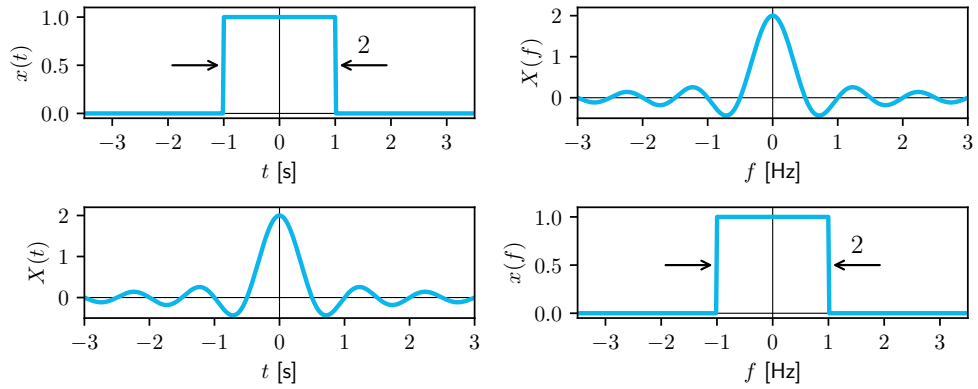


Figure 6.1: Duality of the Fourier transform pairs $\{\Pi\left(\frac{t}{\tau}\right), \tau\,\mathrm{sinc}(\tau f)\}$ and $\{\tau\,\mathrm{sinc}(\tau t), \Pi\left(\frac{f}{\tau}\right)\}$.

EX 6.2

Recall Examples 5.3 and 5.5, where the Fourier transforms were obtained for, respectively, decay functions $z(t) = e^{-\alpha t}u(t)$ and $v(t) = -\frac{1}{2}e^{\alpha t}u(-t) + \frac{1}{2}e^{-\alpha t}u(t)$. When $Z(f) = \frac{1}{\alpha + j2\pi f}$, obtain $V(f)$ in a quick way. Assume $\alpha > 0$.

Solution Using the linearity and time reversal theorems, we immediately obtain:

$$V(f) = -\frac{1}{2}Z(-f) + \frac{1}{2}Z(f) = -\frac{1}{2}\frac{1}{\alpha - j2\pi f} + \frac{1}{2}\frac{1}{\alpha + j2\pi f} = \frac{-j2\pi f}{\alpha^2 + (2\pi f)^2}$$

The Exercises book contains many more examples of how to use the Fourier transform theorems in a clever way; these will prove to be instrumental in the remainder of this book.

Before we discuss signal analysis in discrete time, the following two chapters will provide background on the concept of convolution (in the time and frequency domains) and the consequences of finite signal duration on the Fourier transform.

7

Convolution

Convolution, denoted by the '$*$' symbol, is a mathematical operation on two functions (x and h) that produces a new function ($x * h$). In the previous chapter, we have seen that Fourier-transforming the product of two functions in time, $\mathcal{F}\{x(t)h(t)\}$, yields the *convolution* of the Fourier transforms of the individual signals, (6.9). Its dual states that Fourier-transforming the convolution of two functions in time, $\mathcal{F}\{x(t)*h(t)\}$, yields the *product* of the Fourier transforms of the individual signals, (6.8). But what exactly *is* convolution, what are its properties, and why is it useful in signal analysis? These questions will be answered in this chapter.

7.1 Definition and rationale

This book discusses the analysis of signals in the time and frequency domain, and in both domains the convolution operation can be applied. In this section, we first discuss the convolution operation on signals in the time domain. Consider two signals, functions of time $x(t)$ and $h(t)$; their convolution $x(t) * h(t)$ yields another signal $y(t)$:

$$y(t) \;=\; x(t) * h(t) \;=\; \int_{\tau=-\infty}^{\infty} x(\tau)h(t-\tau)\,\mathrm{d}\tau \tag{7.1}$$

Here, we use τ as the (arbitrary) running variable in the integral. With $x(t)$ and $h(t)$ running from $t = -\infty$ to ∞, their convolution yields a new signal $y(t)$ which also runs from $t = -\infty$ to ∞. An important example where convolution plays a major role is when computing the response $y(t)$ of a so-called Linear Time-Invariant (LTI) system to an input signal $x(t)$, where the response of that system is completely characterized by its impulse response function $h(t)$. These will be explained in Chapter 16.

The operation of convolution can be split into five steps:

1. Consider both signals as a function of time τ, here $x(\tau)$ and $h(\tau)$,
2. Swap, or time-reverse signal $h(\tau)$ to yield $h(-\tau)$,
3. Shift signal $h(-\tau)$ by t to yield $h(t-\tau) = h(-(\tau - t))$,
4. Compute the integral of the product of functions $x(\tau)$ and $h(t-\tau)$, and
5. Repeat steps 3 and 4 for every value of t, i.e., from $t = -\infty$ to ∞.

This eventually results in a new signal, $y(t)$, for all times $t \in (-\infty, \infty)$.

With two simple signals, we evaluate integral (7.1) and demonstrate step-by-step the convolution operation. Assume signal $x(t)$ to be the unit step function:

$$x(t) = u(t)$$

and signal $h(t)$ to be a pulse function:

$$h(t) = 2\Pi\left(\frac{t-3}{2}\right)$$

with a duration of 2 seconds, centered at $t = 3$ s, and with amplitude equal to 2.

In the first step, both signals are considered as functions of running variable τ rather than time t. Figure 7.1 shows $x(\tau) = u(\tau)$ in blue at left, and $h(\tau) = 2\Pi\left(\frac{\tau-3}{2}\right)$ in green at right. To illustrate convolution, we mark three points ((a), (b) and (c)) in the function $h(\tau)$, which will be swapped and shifted, with (c) at the center. Point (d) marks function $x(\tau)$ at $\tau = 0$.

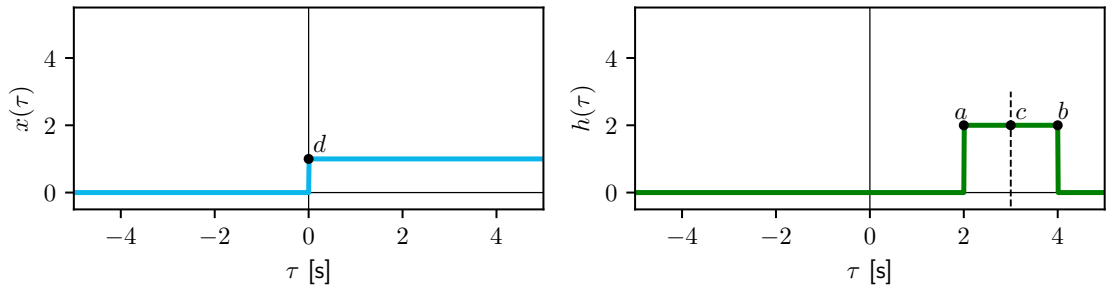


Figure 7.1: Unit step function $x(\tau) = u(\tau)$ at left, and pulse function $h(\tau) = 2\Pi\left(\frac{\tau-3}{2}\right)$ at right.

In the second step, we swap signal $h(\tau)$ about $\tau = 0$, resulting in $h(-\tau)$ (time reversal, Section 2.4), see Figure 7.2 at left. Note that this signal equals $h(t-\tau)$ for time shift $t = 0$ s.

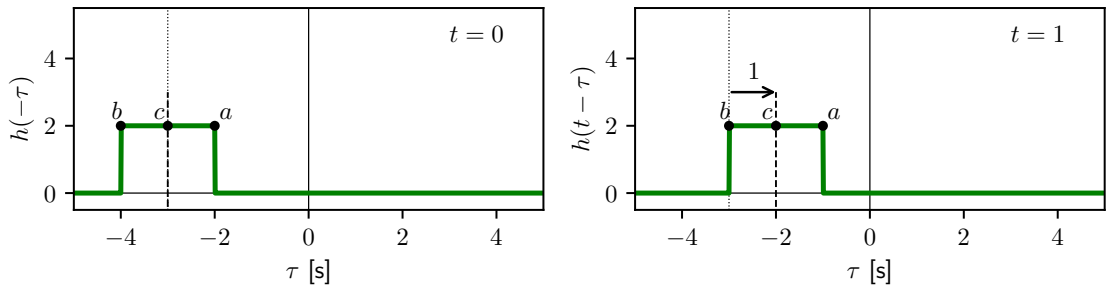


Figure 7.2: Swapped pulse function $h(-\tau) = 2\Pi\left(\frac{-\tau-3}{2}\right)$ at left, and swapped and shifted pulse function $h(t-\tau) = 2\Pi\left(\frac{t-\tau-3}{2}\right)$, for shift $t = 1$ s, at right.

In the third step, we shift signal $h(-\tau)$ by t, resulting in signal $h(t-\tau) = h(-(\tau-t))$ which is shown, for $t = 1$ s, in Figure 7.2 at right. The center of the pulse (c), shifted by t, now occurs at $\tau = -3 + t = t - 3$. The edges of the pulse, shifted by t, now occur at $\tau = -4 + t = t - 4$ (b) and $\tau = -2 + t = t - 2$ (a), see Figure 7.2 at right for $t = 1$.

Now, the fourth step is broken into three parts. First, we note that for any $t - 2 < 0$, hence $t < 2$, the product of $x(\tau)$ and $h(t-\tau)$ is everywhere equal to zero, and hence the integral of this product is zero. This is because, for $t < 2$, point (a) lies to the left of $\tau = 0$, where $u(\tau)$

starts to get a non-zero value (point (d)). Hence, the two functions $x(\tau)$ and $h(t-\tau)$ do not have any *overlap*; their product is zero, and so is the integral (7.1): $y(t) = 0$ for $t < 2$.

In the second part, for $t \geq 2$, pulse function $h(t-\tau)$ 'enters' into overlap with the unit step function. This is shown in Figure 7.3 at left for $t = 3$. This 'entering' covers the range from $t - 2 \geq 0$ (point (a) passes point (d) at $\tau = 0$) to $t - 4 \leq 0$ (point (b) passes point (d) at $\tau = 0$); hence $2 \leq t \leq 4$ for shift t. When $t = 3$, we integrate the product of 1 (from $u(\tau)$) and 2 (from $h(t-\tau)$) from $\tau = 0$ (when $u(\tau)$ is non-zero) to $\tau = t - 2 = 3 - 2 = 1$ (point (a)). For the 'entering' range we obtain:

$$y(t) = \int_0^{t-2} 2 \, \mathrm{d}\tau = [2\tau]_0^{t-2} = 2(t-2)$$

for $2 \leq t \leq 4$. Again, for shift $t = 3$, as we can see in Figure 7.3 at left, the area of the product function equals 2, and $y(t = 3) = 2$, see Figure 7.3 at right.

In the third and last part, shift t satisfies $t - 4 > 0$ (point (b) has passed point (d) at $\tau = 0$) and the pulse function $h(t-\tau)$ is fully overlapping the unit step function. The boundaries of the integral of the product function are the left edge of the pulse at $\tau = t - 4$ (point (b)) and the right edge at $\tau = t - 2$ (point (a)). The result becomes:

$$y(t) = \int_{t-4}^{t-2} 2 \, \mathrm{d}\tau = [2\tau]_{t-4}^{t-2} = 2(t-2) - 2(t-4) = 4$$

for $t > 4$. In this way, we have also covered the fifth step of evaluating the integral of the product function for every value of shift t.

The resulting function $y(t)$ is shown in Figure 7.3 at right, and is given by:

$$y(t) = \begin{cases} 0 & \text{for } t < 2 \\ 2(t-2) & \text{for } 2 \leq t \leq 4 \\ 4 & \text{for } t > 4 \end{cases} \tag{7.2}$$

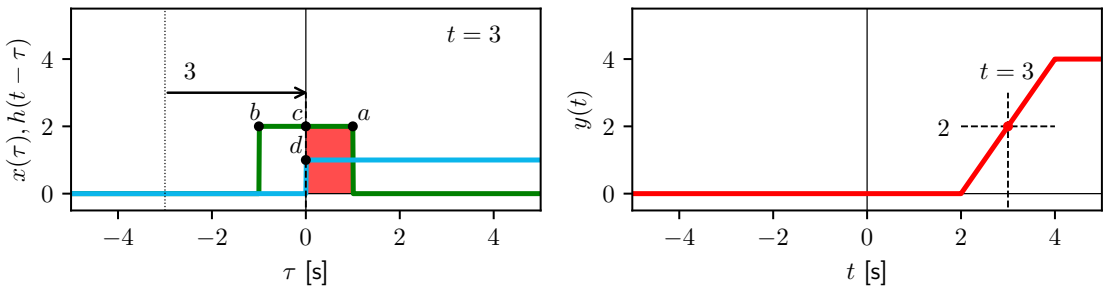


Figure 7.3: Swapped and shifted pulse function $h(t-\tau) = 2\Pi\left(\frac{t-\tau-3}{2}\right)$, shown here for $t = 3$ in green, and unit step function $x(\tau) = u(\tau)$ in blue, at left; convolution output $y(t)$ in red at right.

Figure 7.4 illustrates the process for these two signals once more. Basically, one function in the convolution integral 'remains where it is' (here $x(\tau) = u(\tau)$). The other function (here $h(\tau)$) is 'flipped around $\tau = 0$', shifted to $-\infty$ and then moved right, through changing t, ultimately to $+\infty$, and, for each time shift t, the product $u(\tau)h(t-\tau)$ is integrated *over all values of* τ. The outcome of this integral for *that* value of t then becomes the value of the new function y *at* that time t, $y(t)$.

7.2 Properties of convolution

The convolution operation has a number of properties, here stated in the time domain but equally valid for convolutions in the frequency domain.

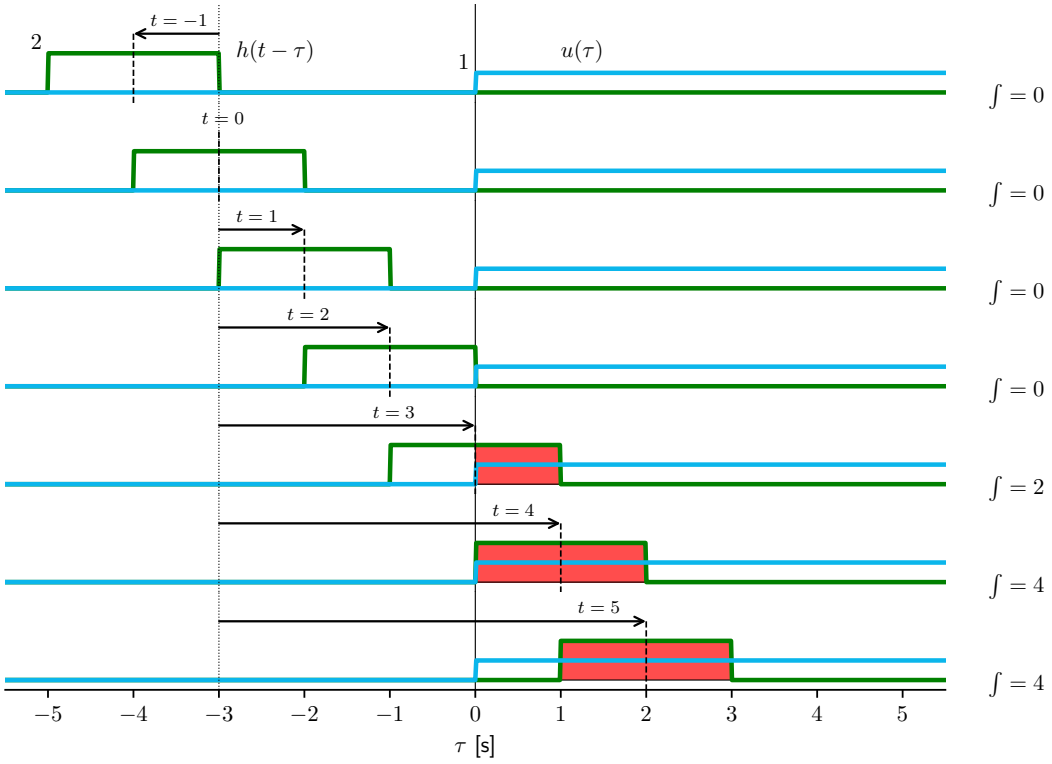


Figure 7.4: Illustration of the convolution of $x(t) = u(t)$ in blue and $h(t) = 2\Pi\left(\frac{t-3}{2}\right)$ in green, for several shift times t. The values of the convolution integral computed at these shift times t are indicated at right.

First of all, convolution is:

1. commutative: $x(t) * h(t) = h(t) * x(t)$,
2. distributive: $(x_1(t) + x_2(t)) * h(t) = x_1(t) * h(t) + x_2(t) * h(t)$, and
3. associative: $x(t) * (h_1(t) * h_2(t)) = (x(t) * h_1(t)) * h_2(t)$.

The commutative property means that:

$$y(t) = x(t) * h(t) = \int_{-\infty}^{\infty} x(\tau)h(t-\tau)\,\mathrm{d}\tau = h(t) * x(t) = \int_{-\infty}^{\infty} h(\tau)x(t-\tau)\,\mathrm{d}\tau \tag{7.3}$$

which can be easily proven by applying a change of variable. In words: when convolving two functions, it does not matter which of the two functions 'remains where it is', while the other is reversed and shifted. For the two signals used in Section 7.1, this is illustrated in Figure 7.5.

Furthermore, it can be proven that:

4. for any constant or factor a: $ay(t) = ax(t) * h(t) = x(t) * ah(t)$, and
5. for any time shift t_0: $y(t + t_0) = x(t + t_0) * h(t) = x(t) * h(t + t_0)$

From the properties of distributivity and associativity with a scalar, it follows that convolution is a *linear* operation.

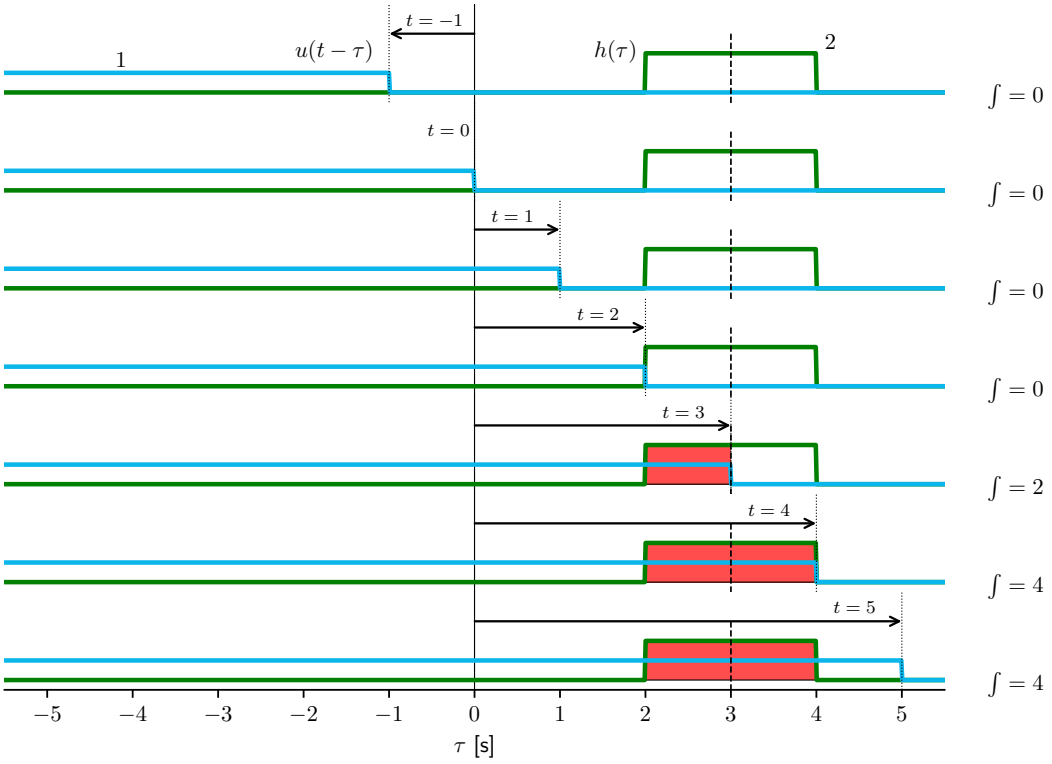


Figure 7.5: Illustration of the commutative property of convolution, using $x(t)$ and $h(t)$ from Section 7.1.

7.3 Convolution with Dirac pulse

The convolution operation is, arguably, not an easy one, and can rapidly become difficult for more complicated functions. There is, however, one notable exception, and that is when one of the functions in the convolution integral is a Dirac delta function. Suppose we would convolve an arbitrary signal $x(t)$ with a Dirac pulse positioned at $t = t_0$, with $t_0 > 0$:

$$y(t) = x(t) * \delta(t - t_0) = x(t - t_0) \tag{7.4}$$

Proof This can be easily proven by direct substitution in the convolution integral, (7.1):

$$y(t) \;=\; x(t) * \delta(t - t_0) \;=\; \int_{-\infty}^{\infty} x(\tau)\delta(t - \tau - t_0)\,\mathrm{d}\tau \;=\; x(t - t_0)$$

where we used the sifting property of the Dirac delta function, (B.12).

Convolution with a Dirac pulse at $t = t_0$ yields a signal which equals $x(t)$, but which occurs t_0 seconds *later* (because here we defined t_0 as larger than 0). Convolving a function $x(t)$ with the Dirac pulse $\delta(t-t_0)$ creates a *copy* of that function, where the function value at $t = 0$ gets *positioned at* the Dirac pulse position, here $t = t_0$.

7.4 Convolution in frequency

Convolution is an operation on two functions which can be functions of time, of frequency, or any other variable of interest. Hence, convolution in the frequency domain is exactly the same operation, follows exactly the same principles, and has exactly the same properties as a convolution in the time domain. For the sake of completeness, a convolution of two functions

$X(f)$ and $H(f)$ in frequency can be stated as:

$$Y(f) = X(f) * H(f) = \int_{\nu=-\infty}^{\infty} X(\nu)H(f-\nu)\,\mathrm{d}\nu \tag{7.5}$$

We use ν as the running variable in the integral. With $X(f)$ and $H(f)$ running from $f = -\infty$ to ∞, their convolution yields a new signal transform $Y(f)$ which also runs from $f = -\infty$ to ∞.

Similar to the time domain, a special case is convolution with a Dirac pulse, that is when $H(f)$ equals $\delta(f - f_0)$:

$$Y(f) = X(f) * \delta(f - f_0) = X(f - f_0) \tag{7.6}$$

This relationship will show to be extremely useful, and can be used to quickly prove several of the Fourier transform theorems introduced in Chapter 6.

7.5 Examples

EX 7.1 Compute the convolution of $x(t) = 4\Pi\left(\frac{t}{4}\right)$ with $y(t) = \delta(t+6)$: $z(t) = x(t) * y(t)$.

Solution

$$z(t) = x(t) * y(t) = 4\Pi\left(\frac{t}{4}\right) * \delta(t+6) = 4\Pi\left(\frac{t+6}{4}\right)$$

Pulse $x(t)$ is centered at $t = 0$ s, pulse $z(t)$ is centered at $t = -6$ s, the position of the Dirac pulse $y(t)$. The shape of the pulse (amplitude and width) does not change.

EX 7.2 Prove the modulation theorem, (6.7), using the multiplication theorem, (6.9).

Solution Suppose $y(t) = x(t)\cos(2\pi f_0 t)$, with f_0 the carrier frequency. Then, with $h(t) = \cos(2\pi f_0 t)$, and $X(f)$ and $H(f)$ the respective Fourier transforms, we obtain:

$$Y(f) = X(f) * H(f) = X(f) * \left(\frac{1}{2}\left(\delta(f+f_0) + \delta(f-f_0)\right)\right) = \frac{1}{2}\left(X(f+f_0) + X(f-f_0)\right)$$

Multiplication in time of a signal $x(t)$ with another signal, in this case a cosine function with frequency f_0, is equivalent to a convolution in frequency of their Fourier transforms. Convolving $X(f)$ with a Dirac pulse at $f = -f_0$ creates a copy of $X(f)$ centered at $f = -f_0$: $X(f + f_0)$. Convolving $X(f)$ with a Dirac pulse at $f = f_0$ yields $X(f - f_0)$.

EX 7.3 Consider signal $x(t) = \tau\,\mathrm{sinc}(\tau t)$ with Fourier transform $X(f) = \Pi\left(\frac{f}{\tau}\right)$, see Example 6.1. Obtain the Fourier transform $Y(f)$ of signal $y(t) = x(t)x(t) = \tau^2\,\mathrm{sinc}^2(\tau t)$.

Solution Fourier-transforming a multiplication of two functions in time results in a convolution of the Fourier transforms of the individual signals in frequency, see (7.5):

$$Y(f) = X(f) * X(f) = \int_{\nu=-\infty}^{\infty} X(\nu)X(f-\nu)\,\mathrm{d}\nu = \int_{-\infty}^{\infty} \Pi\left(\frac{\nu}{\tau}\right)\Pi\left(\frac{f-\nu}{\tau}\right)\mathrm{d}\nu$$

The first pulse function is centered at $\nu = 0$, the second pulse function is centered at $\nu = f$. We consider four domains for f:

1. if $f < -\tau$, then $Y(f) = 0$ as the two pulse functions do not overlap,
2. if $-\tau < f < 0$, then $Y(f) = \int_{-\frac{\tau}{2}}^{f+\frac{\tau}{2}} (1)\,\mathrm{d}\nu = f + \tau$,
3. if $0 < f < \tau$, then $Y(f) = \int_{f-\frac{\tau}{2}}^{\frac{\tau}{2}} (1)\,\mathrm{d}\nu = \tau - f$, and
4. if $f > \tau$, then $Y(f) = 0$ as the two pulse functions do not overlap.

The first and fourth item can be summarized as $Y(f) = 0$ for $|f| > \tau$, and with the triangular function defined in Appendix B we have obtained:

$$Y(f) = \tau\Lambda(\frac{f}{\tau})$$

Dividing $y(t)$ and $Y(f)$ by τ results in a common Fourier transform pair:

$$\boxed{y(t) = \tau\,\mathrm{sinc}^2(\tau t) \overset{\mathcal{F}}{\leftrightarrow} Y(f) = \Lambda\left(\frac{f}{\tau}\right)} \tag{7.7}$$

The convolution of $X(f)$ with $X(f)$ is illustrated in Figure 7.6, for $\tau = 2$.

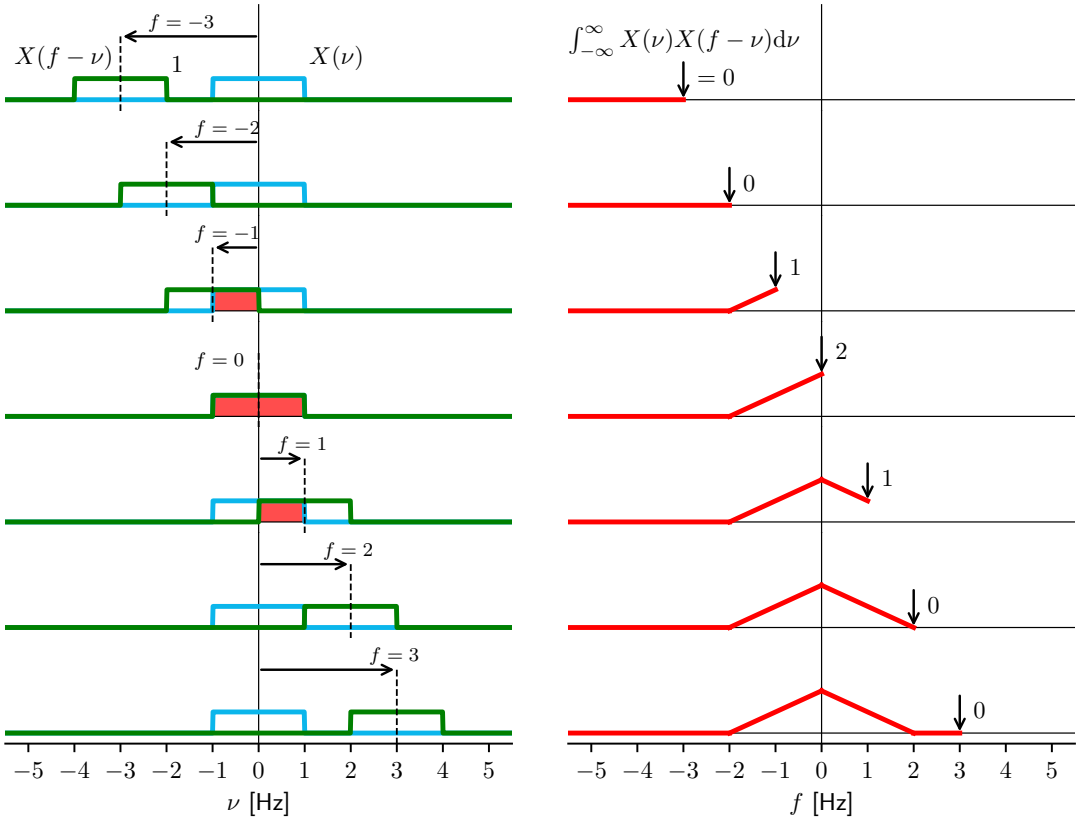


Figure 7.6: Illustration of the convolution of $X(f) = \Pi\left(\frac{f}{\tau}\right)$ with itself (here $\tau = 2$).

EX 7.4

Evaluate the convolution of the unit step function $x(t) = u(t)$ with the exponential decay function $h(t) = \alpha e^{-\alpha t} u(t)$, with $\alpha > 0$. Both functions are shown in Figure 7.7.

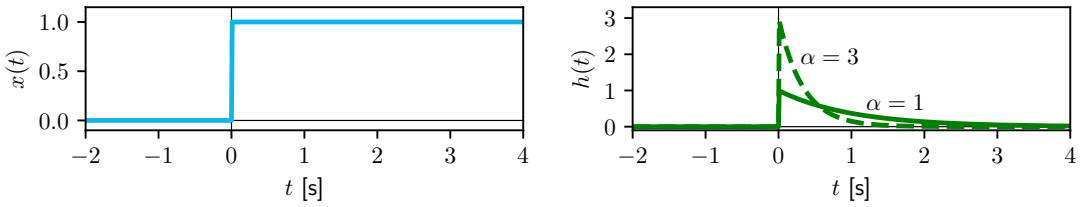


Figure 7.7: Unit step function $x(t) = u(t)$ (left) and exponential decay function $h(t) = \alpha e^{-\alpha t} u(t)$ (right), for two values of α (1 and 3).

Solution Because convolution is commutative we can write:

$$y(t) = x(t) * h(t) = h(t) * x(t) = \int_{-\infty}^{\infty} h(\tau) x(t-\tau)\, \mathrm{d}\tau$$

We choose $h(\tau)$ to 'remain where it is' and time-reverse and time-shift $x(\tau)$. Then, the result after the third step is that we have the exponential decay as a function of τ, $h(\tau) = \alpha e^{-\alpha \tau} u(\tau)$, and the unit step function has been time-reversed and shifted by t; hence $x(t-\tau) = u(t-\tau)$ as a function of τ. The exponential decay function $h(\tau)$ equals zero for $\tau < 0$ due to the presence of $u(\tau)$. The time-reversed and time-shifted unit step function equals zero, i.e., $x(t-\tau) = u(t-\tau) = 0$, for $t - \tau < 0$, i.e., for $\tau > t$. The integration bounds become $\tau = 0$ and $\tau = t$, with $t > 0$, as otherwise there is no overlap between $h(\tau)$ and $x(t-\tau)$, so $y(t) = 0$.

The fourth and fifth steps yield:

$$y(t) = \int_{\tau=0}^{t} \alpha e^{-\alpha \tau}\, \mathrm{d}\tau = \left[-e^{-\alpha \tau}\right]_0^t = -e^{-\alpha t} + e^0 = 1 - e^{-\alpha t}$$

for $t > 0$. The solution can also be written as: $y(t) = (1 - e^{-\alpha t}) u(t)$, for $t \in \mathbb{R}$.

The resulting signal is shown in Figure 7.8, for the two values of α. We return to this example in Chapter 16.

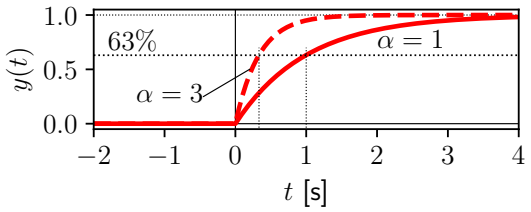


Figure 7.8: Output $y(t)$ as the result of the convolution $x(t) = u(t)$ and $h(t) = \alpha e^{-\alpha t} u(t)$, for two values of α. At $t = \frac{1}{\alpha}$, $y(t)$ equals 63% $(= 1 - e^{-1})$ of its final value, 1; see Chapter 16.

This chapter discussed the convolution of continuous-time signals, in the time and frequency domains. The convolution of two discrete-time sequences is covered in Appendix F.

Before starting with Part III, on discrete-time signals, first the effects of (inevitably) limited observation times will be discussed in the next chapter.

8

Finite signal duration, leakage and windowing

With the Fourier transform in Chapter 5 we work, in general, with signals $x(t)$ in the time domain, for $-\infty < t < \infty$, and hence with signals which extend to infinity, see (5.1). In practice, however, a signal $x(t)$ will be available and measured only for a *finite* time duration T_{meas}, Section 2.6. In this chapter we elaborate on the consequences of limiting the time duration of a signal on spectral analysis: the Fourier transform of a *limited-duration* signal may differ from the Fourier transform of the corresponding infinite-time signal, a phenomenon called *spectral leakage*. We discuss ways to mitigate this inevitable practical problem.

8.1 Finite signal duration

Measurements do not last forever, and typically a recording of a signal $x(t)$ for only T_{meas} seconds is available. Limiting a signal's duration can mathematically be described by multiplication of the infinite-duration signal $x(t)$ by a (dimensionless) *time-window* function $w(t)$. Here we use a rectangular pulse $w(t) = \Pi\left(\frac{t}{T}\right)$ with width $T = T_{meas}$. During T seconds, from $t = -\frac{T}{2}$ to $\frac{T}{2}$ s, this window function $w(t)$ equals 1, and it is 0 for all other times.

The windowed signal $x_w(t)$ equals the original signal $x(t)$ inside the window of duration T, and it equals 0 outside. In other words, we select only a *part* of the original signal:

$$x_w(t) = w(t)x(t) \tag{8.1}$$

The multiplication Fourier transform theorem, (6.9), states that the product of two time signals corresponds to a *convolution* of their Fourier transforms:

$$X_w(f) = W(f) * X(f) \tag{8.2}$$

The Fourier transform of $w(t) = \Pi\left(\frac{t}{T}\right)$ equals $W(f) = T\,\mathrm{sinc}(Tf)$, (5.11). We deliberately set the window to be symmetric about $t = 0$ for convenience, such that $w(t)$ is an even function, and therefore $W(f)$ is a real (and even) function. Both functions are shown in Figure 8.1, for three different values of window duration T. Note that $W(0) = T$ and that the zero-crossings of $W(f)$ occur at $f = k\frac{1}{T}$ for $k \in \mathbb{Z} \setminus \{0\}$. Hence, the longer the window duration T, the higher and narrower the main lobe of the sinc function becomes.

To demonstrate the effects of limiting a signal to a finite duration on its Fourier transform, we continue with Example 7.3. Signal $x(t) = \tau\,\mathrm{sinc}^2(\tau t)$ and its Fourier transform $X(f)$ are shown in Figure 8.2. The Fourier transform is a triangular function, $X(f) = \Lambda\left(\frac{f}{\tau}\right)$, (7.7).

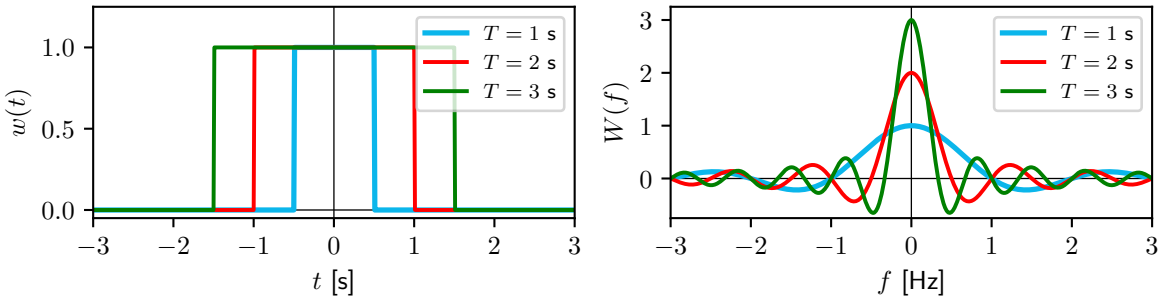


Figure 8.1: Rectangular window $w(t)$ with three different values for duration T, at left, and the corresponding Fourier transforms $W(f)$, at right.

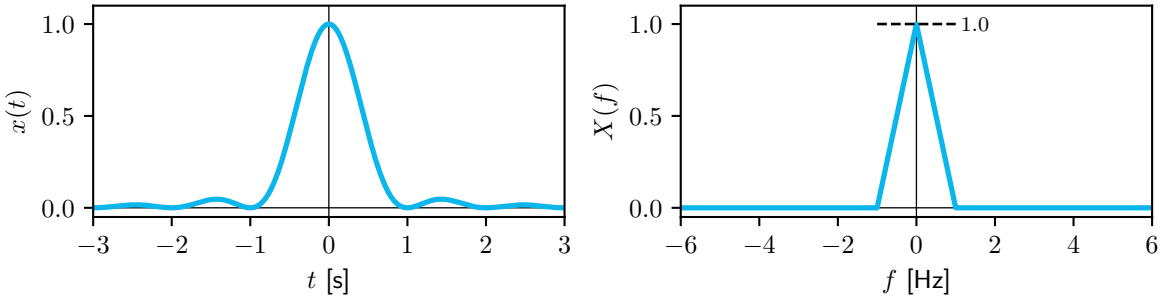


Figure 8.2: Signal $x(t) = \tau\,\mathrm{sinc}^2(\tau t)$ at left, and its Fourier transform $X(f) = \Lambda\left(\frac{f}{\tau}\right)$ at right, for $\tau = 1$.

We apply a rectangular window $w(t) = \Pi\left(\frac{t}{T}\right)$, with a duration of $T = 3$ seconds, and obtain $x_w(t)$, Figure 8.3 (left). The Fourier transform $X_w(f)$, the result of the convolution of $X(f)$ with $W(f)$, shown in Figure 8.3 at right, looks a bit 'wobbly'. Overall the shape of function $X_w(f)$ is still quite similar to the triangular function of $X(f)$, shown as the dotted black shape, though the details are different. Through the width of the main peak of the sinc function, in Figure 8.1 at right, we get a kind of 'spreading' effect. The sharp peak of the triangular function has become smoothed, rounded and smaller. The side lobes of the sinc function cause scalloping, a 'wobbly' effect visible in particular in the tails of $X_w(f)$.

Of course, differences in the Fourier transforms can be expected, as we truncated the tails of the original signal $x(t)$ on both sides, $x_w(t) \neq x(t)$. These effects are referred to as *spectral leakage*: signal content in the frequency domain gets spread, or moved to, or *leaks into*, neighboring frequencies.

In practice, with a sufficiently long measurement duration T, we hope to minimize leakage and achieve $X_w(f) \approx X(f)$. One can imagine that using a longer measurement duration T will be better, having less impact on the Fourier transform, as the sinc function to convolve with, will be narrower, see Figure 8.1. Ideally, $T \to \infty$, and $W(f)$ would become by approximation a Dirac delta function, the convolution with which would yield $X(f)$: $X_w(f) = W(f) * X(f) \approx \delta(f) * X(f) = X(f)$, the original Fourier transform. Obviously, a window of infinite length has no impact at all on the Fourier transform of a signal, that is, its spectral representation.

Again note that, for convenience, both signal $x(t)$ and window $w(t)$ have been chosen, on purpose, to be real and even functions, such that their Fourier transforms $X(f)$ and $W(f)$ are real and even. Defining $w(t) = 1$ for $t = [0, T]$ would mean that $W(f)$ is complex-valued. Hence, in general, $X(f)$ and $W(f)$ are complex-valued and the modulus needs to be taken to show these transforms in a graph.

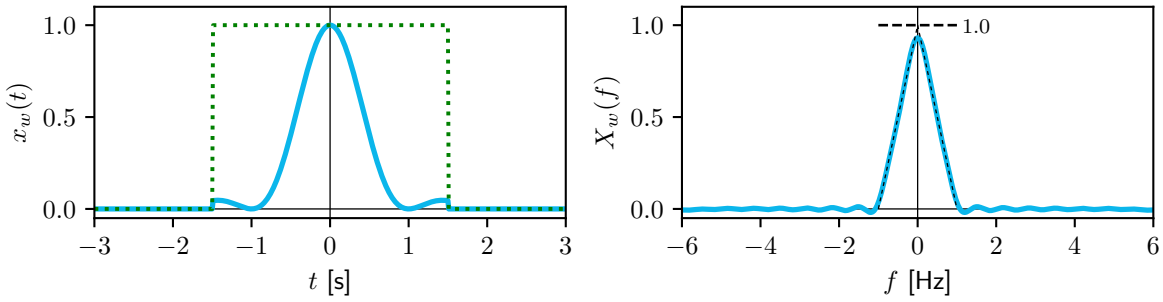


Figure 8.3: Signal $x_w(t) = w(t)x(t)$ at left, with time window function $w(t) = \Pi\left(\frac{t}{T}\right)$ shown in green dots (for $T = 3$ s), and corresponding Fourier transform $X_w(f) = W(f) * X(f)$ at right (obtained numerically, see Appendix F).

8.2 Leakage of harmonics

When applying a time window to a cosine signal, the effects of leakage can be very clearly demonstrated. The Fourier transform of a cosine signal only exists in-the-limit, (5.16), and its spectral representation only consists of two Dirac pulses, Figure 5.5.

The cosine function $x(t) = A\cos(2\pi f_0 t)$ has as its Fourier transform $X(f) = \frac{A}{2}(\delta(f + f_0) + \delta(f - f_0))$. We realize the time window through multiplication of $x(t)$ by window function $w(t) = \Pi\left(\frac{t}{T}\right)$. The Fourier transform of $w(t)$ is $W(f) = T\,\mathrm{sinc}(Tf)$, and the Fourier transform of the time-windowed cosine is found through:

$$\begin{aligned} X_w(f) &= W(f) * X(f) = T\,\mathrm{sinc}(Tf) * \left(\frac{A}{2}\left(\delta(f + f_0) + \delta(f - f_0)\right)\right) \\ &= \frac{AT}{2}\Big(\mathrm{sinc}(T(f + f_0)) + \mathrm{sinc}(T(f - f_0))\Big) \end{aligned} \tag{8.3}$$

We obtain two sinc functions, both with amplitude $\frac{AT}{2}$, one centered at $-f_0$, the other at f_0, see Figure 8.4. The spreading and scalloping effects are clearly visible, as are the widening of peaks and the wobbles. Signal $x_w(t)$, in red, is no longer periodic, and this is reflected in its Fourier transform; components over a *continuous range of frequencies* are needed to construct it.

Finally, note that windowing a harmonic has yet another effect. The harmonic goes on for ever (instead of fading out when $t \to \pm\infty$). Taking a longer window means that we get 'more signal'. The window duration T appears in the amplitude of the resulting sinc functions, see (8.3). The amplitude in the frequency domain is amplified by the window duration.

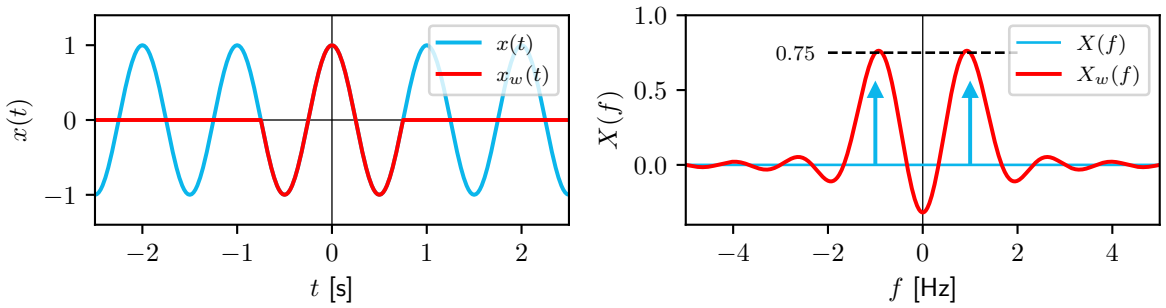


Figure 8.4: Signal $x(t) = A\cos(2\pi f_0 t)$ at left in blue, with $f_0 = 1$ Hz and $A = 1$, and windowed $x_w(t)$ in red, with $T = 1.5$ s. The corresponding Fourier transform $X(f)$ is shown at right in blue, and $X_w(f)$ in red.

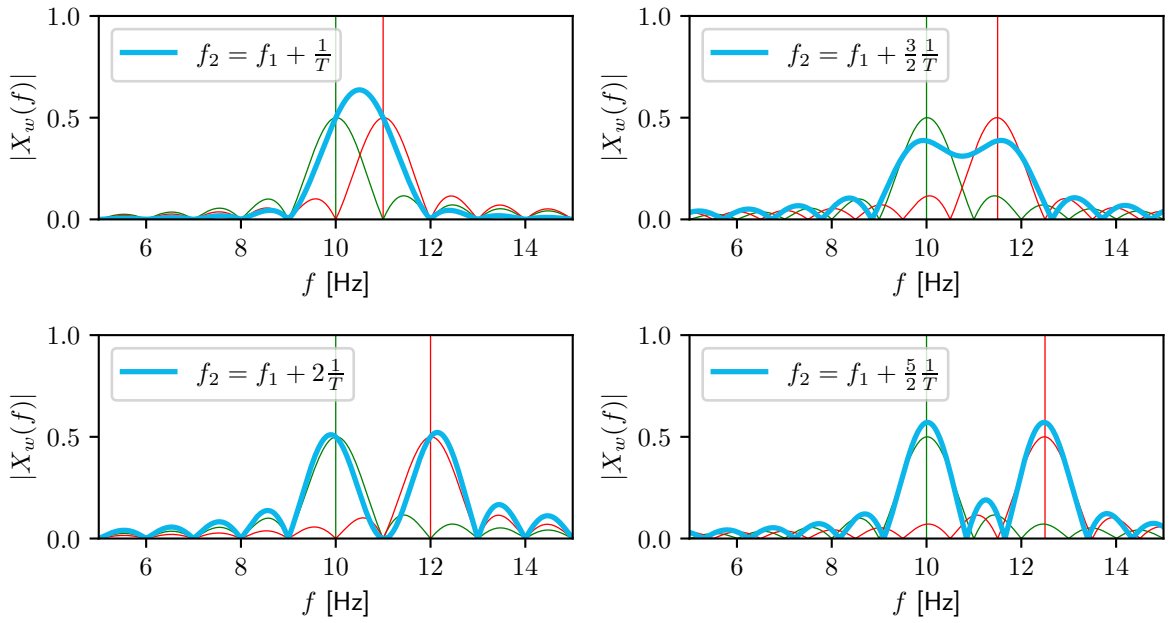


Figure 8.5: Magnitude spectrum of time-windowed signal $x_w(t) = w(t)x(t)$, with $x(t)$ the sum of two zero-phase cosines $x_1(t) = A_1\cos(2\pi f_1 t)$ and $x_2(t) = A_2\cos(2\pi f_2 t)$, with $A_1 = A_2 = 1$, $f_1 = 10$ Hz, and f_2 slightly offset from f_1, as indicated; window length $T = 1$ s (rectangular window $w(t)$). The magnitude spectra of the two (time-windowed) cosines are shown separately in green and red curves, for, respectively, $|W(f) * X_1(f)|$ and $|W(f) * X_2(f)|$; the resulting spectrum $|X_w(f)|$ is shown in blue.

8.3 Identifying neighboring frequencies

In this book, we generally present simple examples with signals consisting of a single harmonic at a specific frequency. Theoretically, with infinite observation times, if we have a signal consisting of multiple harmonics, even at slightly different frequencies, it would be possible to identify each harmonic separately in the frequency domain. The Fourier transform (in-the-limit) of a cosine (frequency f_0) consists of two Dirac pulses: one at $-f_0$, and one at f_0. For a signal with K separate harmonics, we would see $2K$ Dirac pulses in the frequency domain, for example in the magnitude spectrum.

In practice, however, we deal with finite observation length T and consequently *always* face the phenomenon of leakage. As shown in Figure 8.4 at right, signal amplitude 'leaks into' neighboring frequencies, and this may pose problems when the signal contains multiple harmonics at closely separated frequencies: they will start to overlap in the frequency domain. In this section we briefly elaborate on this problem, using two examples, and provide initial guidelines with regard to the ability to identify (harmonics with) neighboring frequencies.

Two cosines are considered with equal amplitude, for a window length of $T = 1$ s. The frequency of the first cosine is fixed, $f_1 = 10$ Hz, and the frequency of the second one, f_2, is varied, such that the separation $f_2 - f_1$ equals $\frac{1}{T}$, $\frac{3}{2}\frac{1}{T}$, $2\frac{1}{T}$, or $\frac{5}{2}\frac{1}{T}$.[1] Figure 8.5 shows, in blue, the magnitude spectrum $|X_w(f)|$ of the time-windowed signal (double-sided spectrum, although only relevant part of positive frequencies is shown). In green and red, the spectra of the two individual (time-windowed) cosines are shown.

When the two frequencies are close together (only $\frac{1}{T}$ apart; top left graph), they cannot be separately identified; their spectra merge into a single peak, shown in blue, halfway the two individual frequencies. When they are separated by $\frac{3}{2}\frac{1}{T}$ (top right graph), two separate peaks

[1] In Chapter 12, $\Delta f = \frac{1}{T}$ will be introduced as the *frequency step* of the Discrete Fourier Transform (DFT).

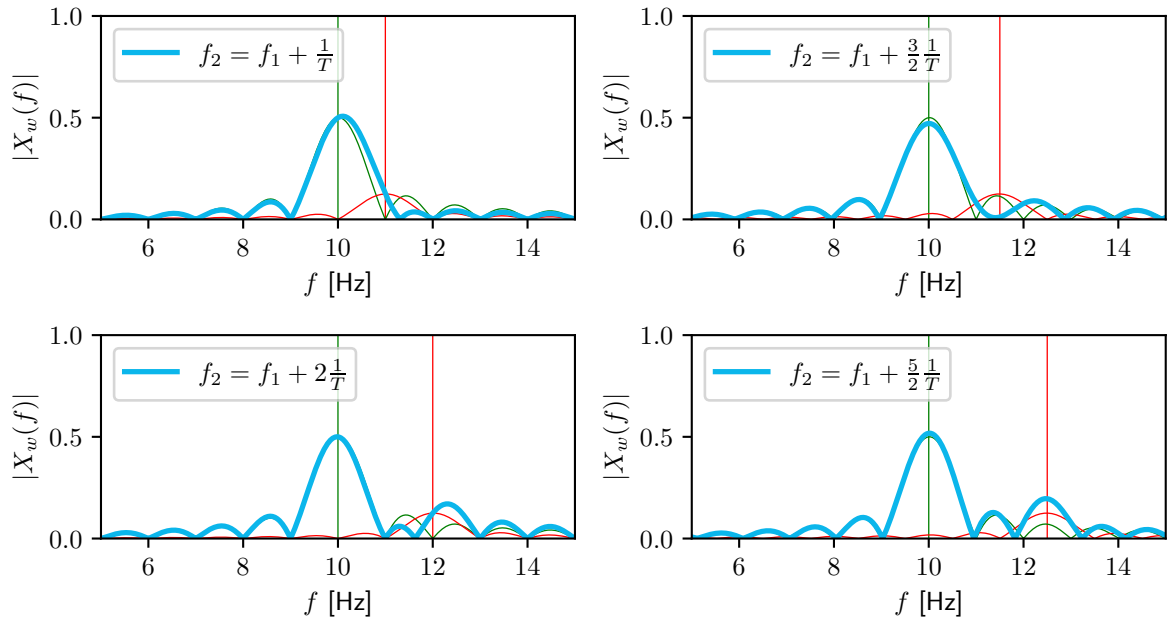


Figure 8.6: Magnitude spectrum of time-windowed signal $x_w(t) = w(t)x(t)$, with $x(t)$ the sum of two zero-phase cosines $x_1(t) = A_1\cos(2\pi f_1 t)$ and $x_2(t) = A_2\cos(2\pi f_2 t)$, with $A_1 = 1$, $A_2 = 0.25$, $f_1 = 10$ Hz, and f_2 slightly offset from f_1, as indicated; window length $T = 1$ s (rectangular window $w(t)$). The magnitude spectra of the two (time-windowed) cosines are shown separately in green and red curves, for, respectively, $|W(f) * X_1(f)|$ and $|W(f) * X_2(f)|$; the resulting spectrum $|X_w(f)|$ is shown in blue.

start to become visible, and when they are $2\frac{1}{T}$ apart (bottom left graph), they both can be clearly identified, though the attentive reader will notice that the two peaks of the blue curve are slightly offset (biased) from the actual frequencies, as indicated by the vertical lines.

When we consider a weak harmonic near a strong one, the problem of identifying them separately becomes even more challenging. We continue the example and keep the amplitude of the first cosine at $A_1 = 1$, and set the amplitude of the second one to $A_2 = 0.25$. Figure 8.6 shows, in blue, the magnitude spectrum $|X_w(f)|$ of the time-windowed signal. In green and red, the spectra of the two individual (time-windowed) cosines are shown.

When the frequency separation is $\frac{1}{T}$ (top left graph), the weak harmonic is not visible at all in the resulting spectrum; the strong harmonic masks the weak one. The latter introduces only a slight bias in the peak of the blue curve toward the right. When the strong and weak components are separated by $2\frac{1}{T}$ (bottom left graph), the weak harmonic becomes visible, although with considerable bias (the peak is off from the correct frequency of 12 Hz). Whereas in this example the two harmonics can be clearly identified at a separation of $\frac{5}{2}\frac{1}{T}$ (bottom right graph), this finding depends on the amplitude ratio $\frac{A_2}{A_1}$ of the two harmonics.

8.4 Consequences for energy and power

Note that windowing a cosine, or any function, periodic or aperiodic, has another effect as well. The resulting windowed signal is always an *energy* signal.

Example 2.10 showed that a cosine $x(t)$ is a power signal, with average power $P_x = \frac{A^2}{2}$. Windowing the cosine with window length T turns it into an *energy* signal $x_w(t)$, $E_{x_w} = \frac{A^2}{2}T$. Recall that taking a longer window means that we get 'more signal' and hence *also* more energy. The average power of $x_w(t)$ is zero, however. To obtain its average power, *calculated*

for the finite window, we divide its energy by T, i.e., the duration of the time window, yielding $P_{x_w} = \frac{A^2}{2}$ for the cosine signal. The implicit assumption in this average power calculation is that the original signal $x(t)$, running from $t = -\infty$ to ∞, 'behaves the same' as the measured signal $x_w(t)$, before and beyond the time window of T seconds.

We conclude that *all* windowed signals, and therefore all *measured* signals, are energy signals. For an arbitrary signal, to obtain the average power during the measurement window, we divide the energy accumulated during the window by the window time duration T:

$$\boxed{P_{x_w} = \frac{1}{T} E_{x_w}} \tag{8.4}$$

8.5 Mitigating leakage: windowing

Measurements have finite duration T, so in all practical cases we need to be aware of potential effects of leakage in the Fourier transform of the measured signal $x_w(t)$. Fortunately, many techniques exist to mitigate leakage effects, other than increasing measurement time. A prominent category of techniques uses *window functions*, other than the rectangular window $w(t)$ that we have used to model and demonstrate the effects of finite duration. Most, if not all, of these window functions represent smooth curves which start and end at value zero.

A large family of window functions is based on a *raised cosine function*. For instance, the so-called Hann window reads:[2]

$$w(t) = \Pi\left(\frac{t}{T}\right)\left(\frac{1}{2} + \frac{1}{2}\cos\left(2\pi\frac{t}{T}\right)\right) \tag{8.5}$$

This function is shown in red in Figure 8.7 at left, for $T = 2$ s. It is based on a cosine with a period of T seconds, $\cos(2\pi\frac{t}{T})$, raised by adding $\frac{1}{2}$, and fit to the interval $[0, 1]$ in terms of function value; it is shown in blue in Figure 8.7. Only one period of the cosine is selected by multiplication by the rectangular pulse in green (dotted). The Hann window, shown in red, just like the rectangular window of Figure 8.1, limits the time duration of signal $x(t)$ to T, through multiplication of $x(t)$ by $w(t)$, (8.1).

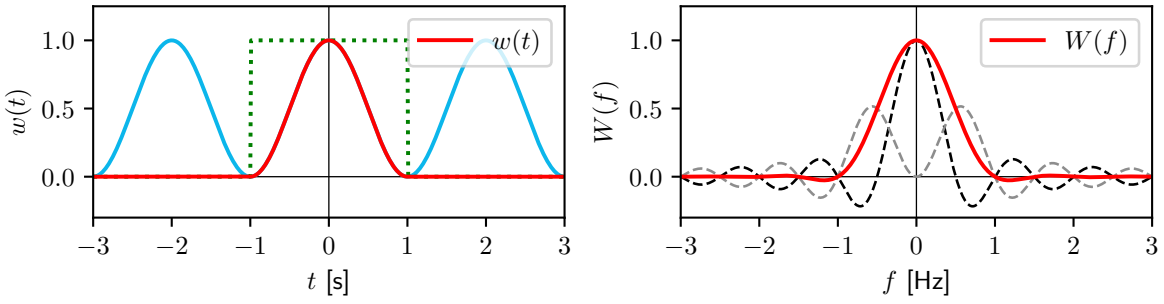


Figure 8.7: Hann window $w(t) = \Pi\left(\frac{t}{T}\right)\left(\frac{1}{2} + \frac{1}{2}\cos(2\pi\frac{t}{T})\right)$ with duration $T = 2$ s in red at left, and corresponding Fourier transform $W(f)$ in red at right. At right, the black and gray dashed curves show, respectively, the first and second part of (8.6).

The essential difference with the rectangular window, is that the Hann window gradually starts from zero and gradually returns to zero at the end, rather than an abrupt switch-on at the beginning and switch-off at the end. In other words, the Hann window gradually 'pushes'

[2] After the Austrian meteorologist J.F. von Hann (1839-1921).

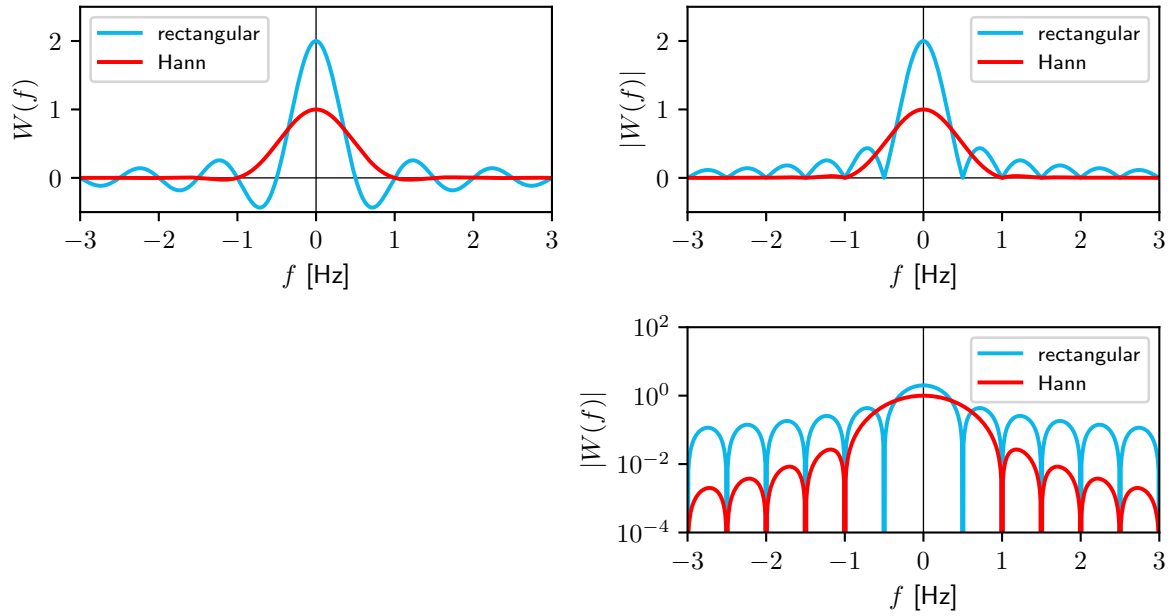


Figure 8.8: Fourier transform $W(f)$ of rectangular window, (5.11), in blue and of Hann window, (8.6), in red (left); window duration $T = 2$ s. Magnitude $|W(f)|$ (right) on a linear (top) and logarithmic (bottom) magnitude scale.

the beginning and end values of the windowed signal $x_w(t)$ to zero at $t = \pm\frac{T}{2}$. Only *then* we compute the Fourier transform of $x_w(t)$.

Figure 8.7 at right shows, in red, the corresponding Fourier transform $W(f)$:

$$\begin{aligned} W(f) &= T\,\mathrm{sinc}(Tf) * \left(\frac{\delta(f)}{2} + \frac{1}{4}(\delta(f+\frac{1}{T}) + \delta(f-\frac{1}{T}))\right) \\ &= \frac{T}{2}\,\mathrm{sinc}(Tf) + \frac{T}{4}\left(\mathrm{sinc}(T(f+\frac{1}{T})) + \mathrm{sinc}(T(f-\frac{1}{T}))\right) \end{aligned} \tag{8.6}$$

found through applying the multiplication theorem, (6.9). The first sinc function, $\frac{T}{2}\,\mathrm{sinc}(Tf)$, and the sum of the latter two sinc functions are shown in, respectively, black and gray dashed curves in Figure 8.7 at right. Since $w(t)$ is a real, even function of time, $W(f)$ is a real, even function of frequency.

Figure 8.8 compares the rectangular window and Hann window in the frequency domain, showing $W(f)$ (left) and $|W(f)|$ (right) on linear (top) and logarithmic (bottom) magnitude scales. The Hann window has a larger width of the main or central peak, twice as wide as that of the rectangular window, but the level of the side peaks is (much) smaller. Practically, the Hann window will introduce more spreading and smoothing than the rectangular window, but it will introduce smaller 'wobbles' in the spectral representation of the windowed signal. This is illustrated in Figure 8.9, where we extend the discussion of Section 8.2 in two ways.

First, we show the effects of using the Hann window (in green) versus the rectangular window (in red). It is clear that the Hann window yields a broader transform $X_w(f)$ (i.e., more smoothing) without much scalloping (i.e., fewer fluctuations). Second, we show the effects of extending the window duration T from 1.5 s (top row) to 3.0 s (bottom row). Clearly, the peaks of the Fourier transforms become higher, the sinc functions narrower, and, in case of a rectangular window, more scalloping occurs. Longer windows will result in even narrower and higher sinc functions, approximating Dirac delta functions when $T \to \infty$, when ultimately there will be no leakage.

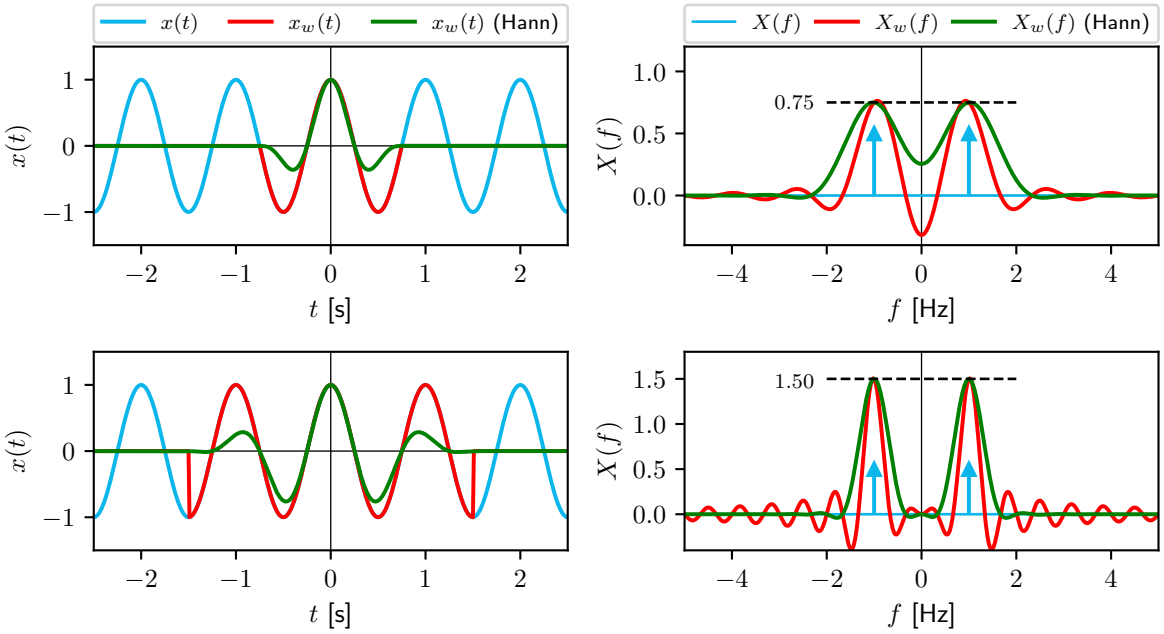


Figure 8.9: At left: signal $x(t) = A\cos(2\pi f_0 t)$ with $f_0 = 1$ Hz and $A = 1$, in blue, and windowed signal $x_w(t)$ (rectangular window: red, Hann window: green) with $T = 1.5$ s (top) and $T = 3$ s (bottom). At right, the corresponding Fourier transforms $X(f)$ (blue) and $X_w(f)$ (rectangular window: red, Hann window: green).

The Hann window is just one member of a family of raised cosine window functions. Many alternative window functions to mitigate effects of leakage have been developed. There is no window function that best serves all applications; it generally comes down to a trade-off between smoothing and scalloping.

As a final note: when taking a measurement of some phenomenon in practice, the effects of the rectangular time window, predominantly leakage, *always* occur. The rectangular window represents our 'model' of performing a finite-duration measurement; it is *always* there. To mitigate the effects of leakage, some opportunities exist, like the Hann window discussed here, but their use is *optional*. In short, whereas the rectangular window is *inevitable* for any practical measurement, applying additional time-windowing techniques is a *choice*.

This chapter concludes Part II, which covered the analysis of continuous-time signals and their representations in the frequency domain. The techniques discussed in this part form the theoretical core of signal analysis, but need to be reconsidered before performing signal analysis in real life. In any modern application of signal analysis, we work with algorithms on a computer, most notably the Discrete Fourier Transform (DFT). Before applying these algorithms, the continuous-time signals must be 'put into the computer' first, a process referred to as sampling and quantization, see Section 2.1.2. Part III of this book will focus on the analysis of discrete-time signals and their representations in the frequency domain.

III

Discrete time

9

Sampling

In Chapter 5 we considered aperiodic, continuous-time signals $x(t)$, of infinite duration, and transformed them to the frequency domain by means of the Fourier transform. In practice, we do not work with continuous-time signals but instead with discrete-time signals: signal $x(t)$ has been *sampled*, with sample frequency f_s Hz. In addition, measurements have a *limited duration* $T = T_{meas}$, yielding a finite sequence of N samples in a computer, x_n. A key question then arises: in what way does the spectrum of these samples x_n, calculated by the computer, differ from the spectrum $X(f)$ of the original, infinite-duration, continuous-time signal $x(t)$?

This question will be answered in three steps. First, in this chapter and Chapter 10, we assume that we have an infinite-duration measurement and will, respectively, discuss the effects of sampling and its reverse process, signal reconstruction, using a *continuous-time* model. Second, in Chapter 11 we introduce the Fourier transform for infinite-duration, discrete-time *sampled* signals, the Discrete-Time Fourier Transform (DTFT). In the third step, Chapter 12, we discuss the effects of limited observation time, and introduce the corresponding Fourier transform, the Discrete Fourier Transform (DFT).

9.1 Sampling

Consider Figure 9.1, which shows an infinite-duration continuous-time signal $x(t)$, with Fourier transform $X(f)$ (the latter not shown). All signal analysis theory and tools discussed in this book so far apply to this category of signals. While in the past analog electrical circuits were used to analyze a continuous-time signal in the frequency domain, nowadays computers are used to process a discrete-time, *sampled* version of that signal.

On a computer, all that we have is a sequence of numbers x_n, with n the index of that number. Computers are *agnostic* when it comes to the time a number 'belongs to': time is lost. As a consequence, they are also agnostic when it comes to the frequency of an oscillation: frequency is lost. Time and frequency disappear altogether, and it is our job to restore time and frequency at some point, e.g., when plotting a signal or its spectrum, or when performing calculations on (or transformations of) signals on a computer.

We define sampling to be performed at a fixed rate, the sampling frequency f_s in [Hz]. A sample is then obtained every sampling interval $\Delta t = \frac{1}{f_s}$ s. Since we assume infinite-time duration signals $x(t)$, we get an infinite number of samples x_n, where $n \in \mathbb{Z}$. While this is not realistic, this assumption will be useful to obtain crucial insights about what happens when you sample a signal. In Chapter 12 we abandon this assumption and consider the practical case: a limited-duration signal, working with a limited number of samples.

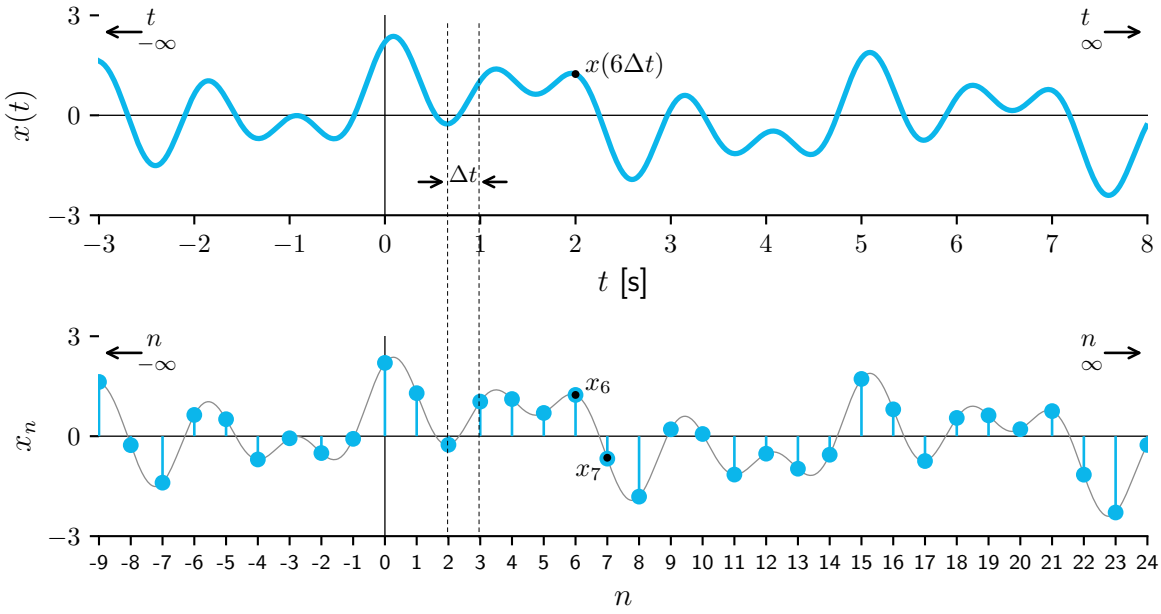


Figure 9.1: Top: continuous-time signal $x(t)$, of infinite duration, $t \in \mathbb{R}$. Bottom: the samples obtained from $x(t)$, with sampling interval Δt, producing an infinite sequence of numbers x_n with $n \in \mathbb{Z}$.

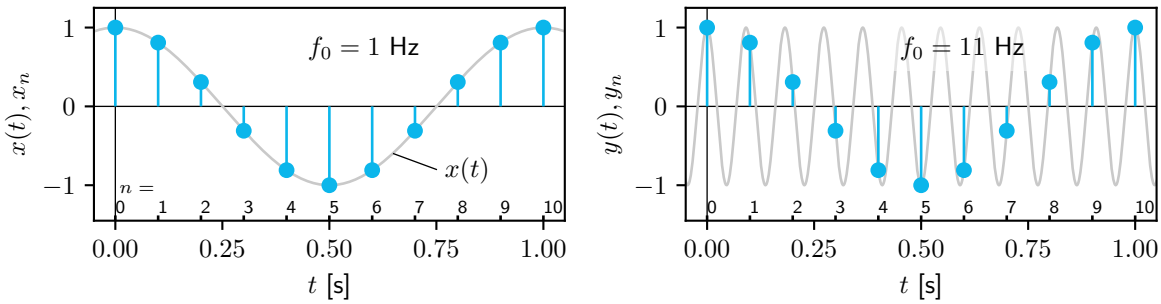


Figure 9.2: Left: continuous-time signal $x(t)$, a cosine with frequency $f_0 = 1$ Hz, sampled with $f_s = 10$ Hz, yielding x_n. Right: signal $y(t)$, a cosine with frequency $f_0 = 11$ Hz, also sampled with $f_s = 10$ Hz, yielding y_n. The horizontal axis shows time t in [s], and index n.

At each moment $n\Delta t$ a sample is taken, i.e., at each sample time $t = n\Delta t$, the value of the sample x_n equals the value of the continuous-time signal at that time:

$$\boxed{x_n = x(t = n\Delta t)}, \qquad \forall\ n\ \in\ \mathbb{Z} \tag{9.1}$$

The effects of signal quantization (for which, see Section 2.1 and Appendix D) are ignored. As illustrated in Figure 9.1, sample x_6 equals $x(6\Delta t)$, et cetera. From x_n we move to x_{n+1}, and in the same way from x_6 to x_7. In between the samples, the sampled discrete-time signal is *undefined*. When having two samples on the computer, we do not know what the continuous-time signal looks like in between the samples. Information is lost; an inevitable consequence of sampling. However, when signal $x(t)$ 'behaves well' in between the samples – for instance it does not oscillate rapidly or show a sudden peak – one could argue that its samples x_n represent signal $x(t)$ 'well enough' to analyze it using the samples.

Figure 9.2 illustrates a 'well-sampled' signal (left), and the *ambiguity* that results from

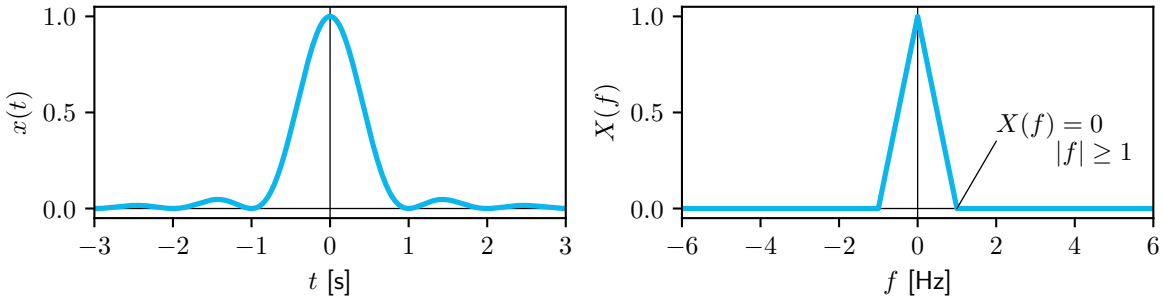


Figure 9.3: Aperiodic continuous-time signal $x(t) = \text{sinc}^2(t)$ at left, and its Fourier transform $X(f) = \Lambda(f)$ at right. As signal $x(t)$ is even, the Fourier transform $X(f)$ is even and real-valued.

sampling a signal too slowly (right). From the samples y_n we might conclude that $y(t) = x(t)$, since we cannot distinguish between $x(t)$ and $y(t)$. Signal $x(t)$ is therefore called an *alias* of $y(t)$. In fact, an infinite number of signals exist (e.g., a cosine with frequency $f_0 = 9$ or 21 Hz) which, when improperly sampled, yield exactly the same sample values. Clearly, this must be prevented, and it will be shown that we must sample 'fast enough' to obtain a sequence of samples that truly represent the signal we intend to analyze, and not an alias. With infinite observation times, the requirement is that we can *perfectly reconstruct* a continuous-time signal from its samples. This leads to the *sampling theorem*, discussed in Section 9.4. When adhering to this theorem, no information is lost, a remarkable fact given that we do not know the signal in between the samples. Signal reconstruction is the topic of Chapter 10.

Throughout this and the following four chapters, we take signal $x(t) = \text{sinc}^2(t)$ as an example. Its Fourier transform $X(f)$ equals $\Lambda(f)$, (7.7) (here $\tau = 1$); both are shown in Figure 9.3. We deliberately use an *even* signal as an example, because then $X(f)$ is *real*-valued, and we only need to show the real part of the Fourier transform. Of course, in general $x(t)$ will not be even, and the Fourier transform $X(f)$ is complex-valued. Another useful property of this signal, as we will see in Section 9.4, is that it is *band-limited*: its Fourier transform equals 0 for frequencies greater than or equal to 1 Hz; see Figure 9.3 at right. band-limited signals are constrained in how rapidly they change in (an interval of) time, a signal property that will play a crucial role.

In the next section, we first develop a mathematical model for the sampling process, the *impulse train sampling model*. This model enables us to obtain and understand the spectrum of a sampled signal, while using the (continuous-time) Fourier transform.

9.2 Impulse train sampling model

A mathematical model for the sampling process is developed, as an ideal abstraction of reality: the *impulse train sampling* model. We continue to work in continuous time, and construct a conceptual signal $x_s(t)$, which preserves the information in $x(t)$ at just the sample instants. Considering signal $x(t)$ only at discrete, equitemporal time instants, $t = n\Delta t$ is materialized through multiplication of that signal by the impulse train sampling function $p(t)$:

$$x_s(t) = x(t)p(t) \tag{9.2}$$

where:

$$p(t) = \Delta t \sum_{n=-\infty}^{\infty} \delta(t - n\Delta t) \tag{9.3}$$

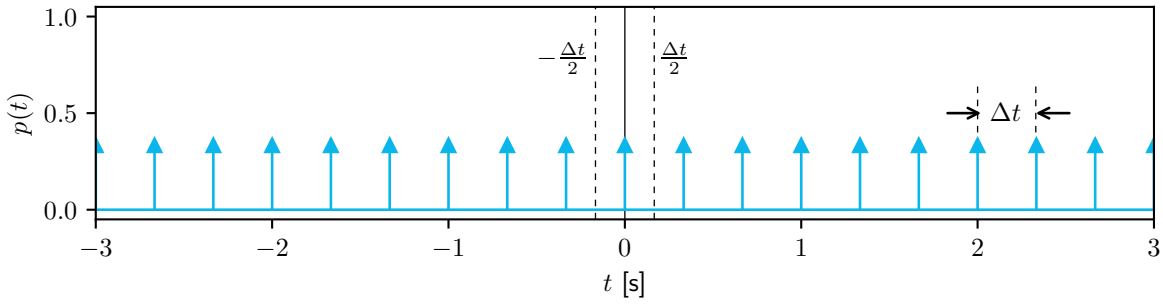


Figure 9.4: Signal $p(t) = \Delta t \sum_{n=-\infty}^{\infty} \delta(t - n\Delta t)$ is the impulse train, here $\Delta t = \frac{1}{3}$ s.

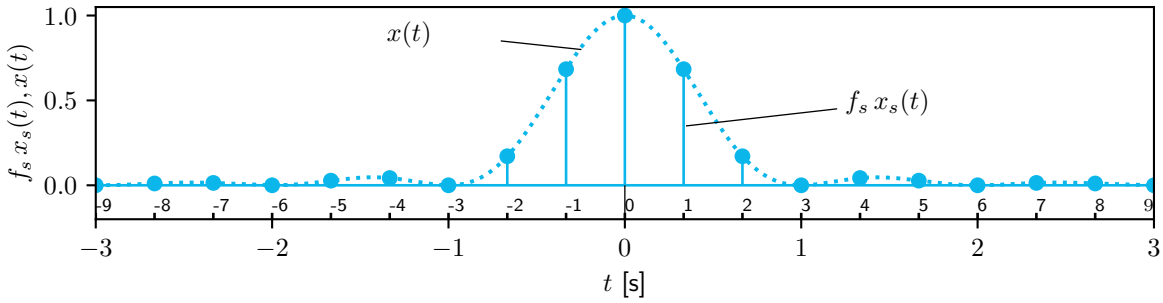


Figure 9.5: Signal $x(t) = \mathrm{sinc}^2(t)$ (dotted) and the model of the samples, $x_s(t)$ multiplied by f_s (stems, and zero otherwise); sampling frequency $f_s = \frac{1}{\Delta t} = 3$ Hz.

with $n \in \mathbb{Z}$. The impulse train, also known as a 'Dirac comb', is a continuous-time, periodic function with a Dirac delta impulse every Δt s; its period is Δt. It is zero at all times, except at infinitesimal moments of time which coincide with the sampling times $n\Delta t$, see Figure 9.4.

Signal $x_s(t)$ does not exist in real life but acts as a convenient 'model of the samples', and plays a crucial role in this and the following chapter. It is a *continuous-time* signal, and has a value at *all* times $t \in \mathbb{R}$. In between the time of the samples, it equals zero. *At* the time of a sample $t = n\Delta t$, the weight of the Dirac pulse at that time equals Δt times the value of $x(t)$ at that sample time, $x(n\Delta t)$. Because the height of the Dirac pulse itself at that time has infinite value, an interpretation of (9.2), e.g., in terms of signal value, is troublesome.

We used the Dirac function many times before in order to 'put a number at a frequency', for instance in the Fourier transform-in-the-limit of a sinusoidal signal, (5.16). In $x_s(t)$, we use a Dirac comb to 'put' the sample values $x(n\Delta t)$ *at* the right times, while for all other times the signal is zero. Note that continuous-time signal $x_s(t)$ is *not* the same as the discrete-time sequence of samples, x_n, used in Chapters 11 and 12.

The scaling with Δt in (9.3) may seem arbitrary. Other textbooks on signal analysis typically do *not* include it in their impulse train sampling model, e.g. [3, 7, 8]. For multiple reasons, we introduce the scaling here. As a start, recall that the Dirac delta impulse $\delta(t)$ has dimension of one-over-time, see Appendix B. This means that through multiplication by Δt in (9.2) we ensure that $x(t)$ and $x_s(t)$ share the same unit, [unit].

Scaling $x_s(t)$ with Δt has many other advantages, but one disadvantage: when plotting $x_s(t)$ and $x(t)$ together in one figure, $x_s(t)$ would, since Δt is typically (much) smaller than 1, lie well below $x(n\Delta t)$. To avoid this, we plot $f_s x_s(t)$ in all figures showing $x_s(t)$. For our example, the resulting uniformly-sampled signal is shown in Figure 9.5, with the stems representing $x(n\Delta t)$ as the weights of the Dirac pulses.

9.3 Derivation of Fourier transform of sampled signal

Because $x_s(t)$ is a continuous-time signal, we can compute its Fourier transform, $X_s(f)$. We first find the Fourier series of the periodic impulse train function $p(t)$, and then develop an expression for $X_s(f)$. The impulse train function $p(t)$ is *periodic* with period Δt and its fundamental frequency equals f_s (instead of symbol T_0 for the period and f_0 for the fundamental frequency, respectively, as used in Chapter 3). With (4.2) the complex exponential Fourier series of the impulse train function $p(t)$ reads:

$$p(t) = \sum_{k=-\infty}^{k=\infty} P_k e^{j2\pi k f_s t} \tag{9.4}$$

with Fourier coefficients (4.3), for $k \in \mathbb{Z}$:

$$P_k = \frac{1}{\Delta t}\int_{\Delta t} p(t)e^{-j2\pi k f_s t}\,\mathrm{d}t \;=\; \frac{1}{\Delta t}\int_{-\frac{\Delta t}{2}}^{\frac{\Delta t}{2}} p(t)e^{-j2\pi k f_s t}\,\mathrm{d}t \tag{9.5}$$

We chose the integration over one period Δt to be computed symmetrically about $t = 0$, see Figure 9.4. In this interval only the Dirac delta at $t = 0$ (scaled by Δt) occurs:

$$P_k \;=\; \frac{1}{\Delta t}\int_{-\frac{\Delta t}{2}}^{\frac{\Delta t}{2}} \Delta t\,\delta(t)e^{-j2\pi k f_s t}\,\mathrm{d}t \;=\; \frac{1}{\Delta t}\Delta t e^{-j2\pi k f_s 0} \;=\; 1\;\forall\; k \tag{9.6}$$

With $P_k = 1\;\forall\; k$, the impulse train function $p(t)$, (9.4), can be expressed as:

$$p(t) = \sum_{k=-\infty}^{k=\infty} e^{j2\pi k f_s t} \tag{9.7}$$

Substituting (9.7) into (9.2) yields:

$$x_s(t) = x(t)p(t) = x(t)\sum_{k=-\infty}^{k=\infty} e^{j2\pi k f_s t} \;=\; \sum_{k=-\infty}^{k=\infty} x(t)e^{j2\pi k f_s t} \tag{9.8}$$

Fourier-transforming this equation results in:

$$X_s(f) = \mathcal{F}\{x_s(t)\} = \mathcal{F}\left\{\sum_{k=-\infty}^{\infty} x(t)e^{j2\pi k f_s t}\right\} = \sum_{k=-\infty}^{\infty} \mathcal{F}\left\{x(t)e^{j2\pi k f_s t}\right\} \tag{9.9}$$

because the Fourier transform is a linear operator, (6.1). Using the frequency translation property of the Fourier transform, (6.6), we obtain our final result:

$$\boxed{X_s(f) = \sum_{k=-\infty}^{\infty} X(f - k f_s)} \tag{9.10}$$

The Fourier transform $X_s(f)$ is a *continuous* function of frequency f, with $f \in \mathbb{R}$. It is *periodic* with period f_s: $X_s(f + m f_s) = X_s(f)\;\forall\; m \in \mathbb{Z}$. Apparently, the result of impulse train sampling a signal in time with sampling frequency f_s is that the Fourier transform of that signal is copied to all integer multiples of that sampling frequency $k f_s$ with $k \in \mathbb{Z}$.

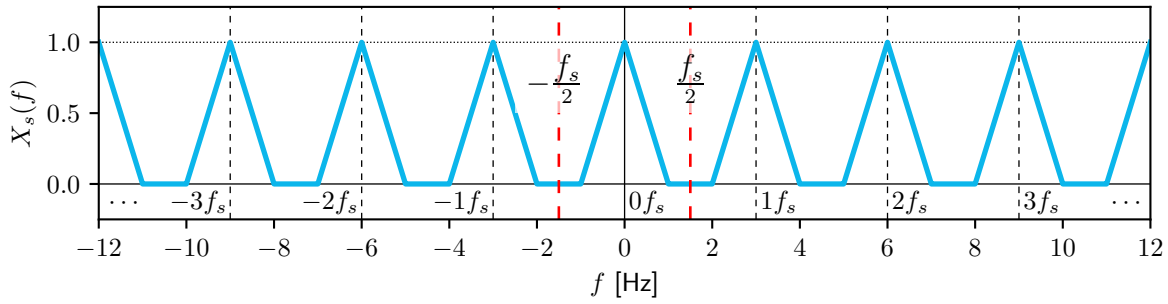


Figure 9.6: Fourier transform $X_s(f)$ of $x_s(t)$. The sampling frequency $f_s = \frac{1}{\Delta t} = 3$ Hz.

Returning to our example signal, the Fourier transform $X_s(f)$ of $x_s(t)$ is shown in Figure 9.6. Compare this figure with the Fourier transform of the original continuous-time signal $x(t)$, $X(f)$, Figure 9.3 at right. Clearly, for $k = 0$, (9.10) or, equivalently, between $f = -\frac{f_s}{2}$ and $f = \frac{f_s}{2}$, $X_s(f)$ equals the Fourier transform $X(f)$ of the original continuous-time signal $x(t)$. In addition, *copies* of $X(f)$ appear at *all* integer multiples of the sampling frequency, that is, at kf_s with $k = -\infty, \ldots, -1, 0, 1, 2, 3, \ldots, \infty$. This represents the *ambiguity* we discussed before, with Figure 9.2.

Apparently, a consequence of sampling is that the infinite frequency axis is cut into equal bands of frequency, all with a width of f_s, where copies of the original spectrum $X(f)$ are positioned. In Chapter 11 we analyze with more mathematical rigor *why* the Fourier transform of a sampled signal repeats itself. In later chapters, by default we consider only one 'period' of the spectrum $X_s(f)$ of a sampled signal, either from $[0, f_s)$ or from $[-\frac{f_s}{2}, \frac{f_s}{2})$.

Scaling the Dirac comb in (9.3) with Δt, as we do, avoids scaling of $X_s(f)$. As mentioned above, many other textbooks do not scale the Dirac comb with Δt, but then (because P_k would equal $\frac{1}{\Delta t}$ $\forall$ k in (9.6)) their expression of $X_s(f)$ is scaled by $\frac{1}{\Delta t}$. Again, a big advantage of scaling the Dirac comb by Δt is that the spectrum of the samples, $X_s(f)$, when properly sampled, will equal $X(f)$ for $-\frac{f_s}{2} \leq f \leq \frac{f_s}{2}$.

9.4 Sampling theorem

The result of sampling a continuous-time signal $x(t)$ is that its Fourier transform $X(f)$ gets 'copied' in the frequency domain. An infinite number of copies is created, centered at integer multiples of the sampling frequency f_s. Each copy is assigned a band of frequencies, with a width of f_s. But, what would happen when $X(f)$ is broader than this band, that is, when $X(f)$ is non-zero beyond $\pm\frac{f_s}{2}$?

Sampling faster, with higher f_s, will assign larger frequency bands to each copy; the copies are positioned further apart, Figure 9.7. For all sampling frequencies, the copy put at 0 Hz, will maintain its position, at $0f_s$. When changing the sampling frequency, all other copies will move. Sampling slower, with lower f_s, will assign smaller frequency bands to each copy; they are positioned closer together. When sampling too slowly, the copies will *overlap*, Figure 9.8. Within each frequency band the energy in the copies will merge, and the Fourier transform $X_s(f)$ of the sampled signal $x_s(t)$ will be *different* from the original Fourier transform $X(f)$. This is known as *aliasing*, and should be prevented. When it occurs, the original Fourier transform $X(f)$ cannot be reconstructed from the Fourier transform of the sampled signal $X_s(f)$. Instead of reconstructing the original signal from the sampled signal, an *alias* is obtained, leading to false conclusions about the measured signal.

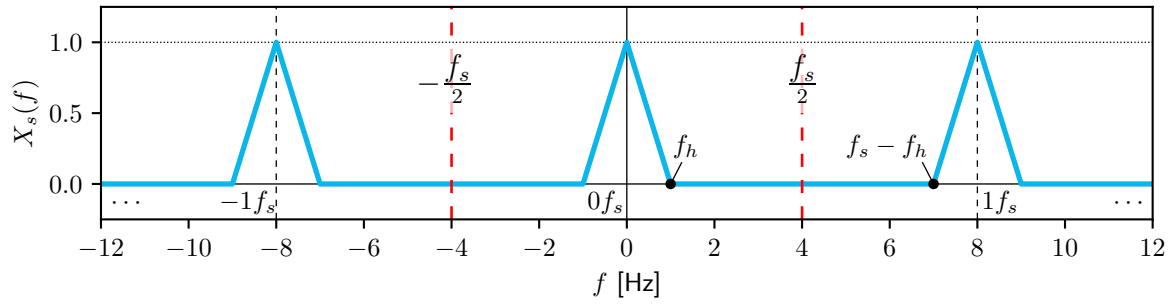


Figure 9.7: Fourier transform $X_s(f)$ of $x_s(t)$; sampling frequency $f_s = 8$ Hz.

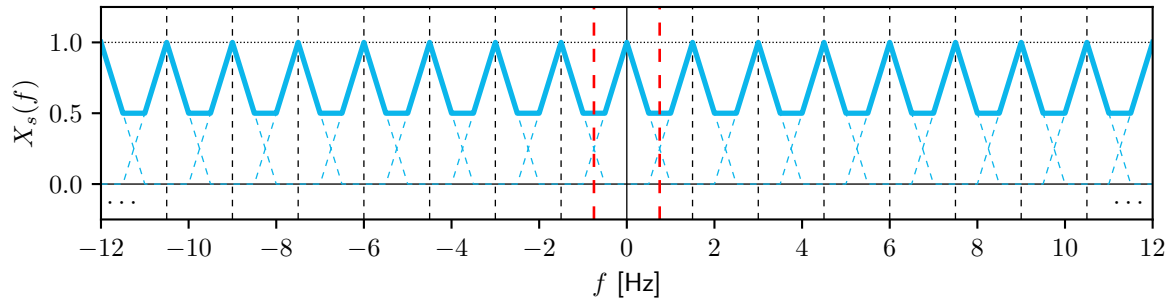


Figure 9.8: Fourier transform $X_s(f)$ of $x_s(t)$; sampling frequency $f_s = 1.5$ Hz. The individual copies, occurring at integer multiples of 1.5 Hz, are shown using dashed lines in blue. To prevent clutter, no further annotations are shown in the figure.

In our example, the highest frequency present in the signal, referred to as f_h, is $f_h = 1$ Hz, see Figure 9.7. From this figure, we can determine for what sampling frequency f_s the copy positioned at $0f_s$ will 'touch' the copy positioned at $1f_s$. This happens when:

$$f_h = f_s - f_h \Rightarrow f_s = 2f_h \tag{9.11}$$

No overlap occurs when $f_s > 2f_h$, which leads to the *sampling theorem*:

A *band-limited* signal $x(t)$, having no frequency components beyond f_h, i.e., $X(f > f_h) = 0$, is completely specified by its samples, when sampled at a uniform rate larger than $2f_h$. In other words, the sampling frequency f_s should meet:

$$\boxed{f_s > 2f_h} \tag{9.12}$$

with f_h the largest frequency occurring in the Fourier transform $X(f)$ of signal $x(t)$. The value $2f_h$ is a characteristic of the *signal* to be sampled and is referred to as the Nyquist *rate*.[1] Another, equivalent way to phrase requirement (9.12) is:

$$\boxed{\frac{f_s}{2} > f_h} \tag{9.13}$$

The value $\frac{f_s}{2}$ is a characteristic of the *sampling system* and is referred to as the Nyquist *frequency*. Whereas the Nyquist rate is a property of the *signal*, the Nyquist frequency is a

[1] This inequality, and also the sampling theorem itself, is named after the Swedish-American physicist and electronic engineer H.T. Nyquist (1889-1976), though be attributed also to British mathematician Sir E.T. Whittaker (1873-1956) and American mathematician C.E. Shannon (1916-2001).

property of the *sampler*. As will become clear in the next chapter on signal reconstruction, the band of frequencies between $-\frac{f_s}{2}$ and $\frac{f_s}{2}$, also referred to as 'from negative Nyquist to positive Nyquist', will play a prominent role. Although in principle one could use any copy in $X_s(f)$ to reconstruct a signal from the sampled signal (and account for the frequency translation of that copy relative to $f = 0$), typically the copy at $f = 0f_s$ is used. This is also why this region is indicated using the vertical red dashed lines in Figures 9.6–9.8.

Note that when $X(f \geq f_h) = 0$, i.e., $X(f)$ is zero *at* f_h, the sampling theorem becomes:

$$f_s \geq 2f_h \text{ or, equivalently, } \frac{f_s}{2} \geq f_h \tag{9.14}$$

In our example, with $f_h = 1$ Hz, we could sample the signal $x(t)$ with sampling frequency f_s *equal to* 2 Hz, and still be able to perfectly reconstruct it from the sampled signal.

Hence, two cases exist for setting the minimum required sampling frequency:

- When $X(f = f_h) \neq 0$: the minimum sampling frequency is *strictly larger than* $2f_h$;
- When $X(f = f_h) = 0$: the minimum sampling frequency is *equal to, or larger than* $2f_h$.

To decide on the choice for f_s, we have to calculate (beforehand) the Fourier transform of the signal to be sampled, which in practice we may not know. In practice, we typically sample at (much) higher frequencies than the Nyquist rate, i.e., *over-sampling*.

When the above requirement, stated as either (9.12) or (9.13), is not met, we are *under-sampling* the signal and *aliasing* will occur, as demonstrated in the next section.

9.5 Aliasing

In this section the effects of sampling, including aliasing, are shown using two examples in which we aim to sample a plain cosine signal:

$$x(t) = 2\cos(2\pi f_0 t) \tag{9.15}$$

with $f_0 = 11$ Hz, and Fourier transform-in-the-limit $X(f) = \delta(f+11) + \delta(f-11)$, consisting of two Dirac pulses. Signal $x(t)$ is chosen to be even, such that its Fourier transform is real and even, and we can just show the real part. The highest frequency of the Fourier transform of this signal, f_h, equals 11 Hz, and *at* this frequency we have a Dirac pulse, i.e., $X(f = f_h) \neq 0$. We must apply a sampling frequency that is *strictly higher* than $2f_h$: $f_s > 22$ Hz.

EX 9.1

Sample signal $x(t)$ of (9.15) with $f_s = 30$ Hz. Show the Fourier transforms $X(f)$, $X_s(f)$ and $X_r(f)$ of, respectively, the original signal $x(t)$, the sampled signal $x_s(t)$ and the reconstructed signal $x_r(t)$. Assume that we use the band from $f = -\frac{f_s}{2}$ to $\frac{f_s}{2}$ Hz of $X_s(f)$ for $X_r(f)$.

Solution Figure 9.9 illustrates all Fourier transforms. The top image shows $X(f)$, two Dirac pulses at ± 11 Hz, with a weight equal to 1. When sampling $x(t)$ with $f_s = 30$ Hz, the Fourier transform of the sampled signal, $X_s(f)$, will consist of copies of $X(f)$ at all integer multiples of f_s, i.e. at $k30$ Hz, with $k \in \mathbb{Z}$, see (9.10).

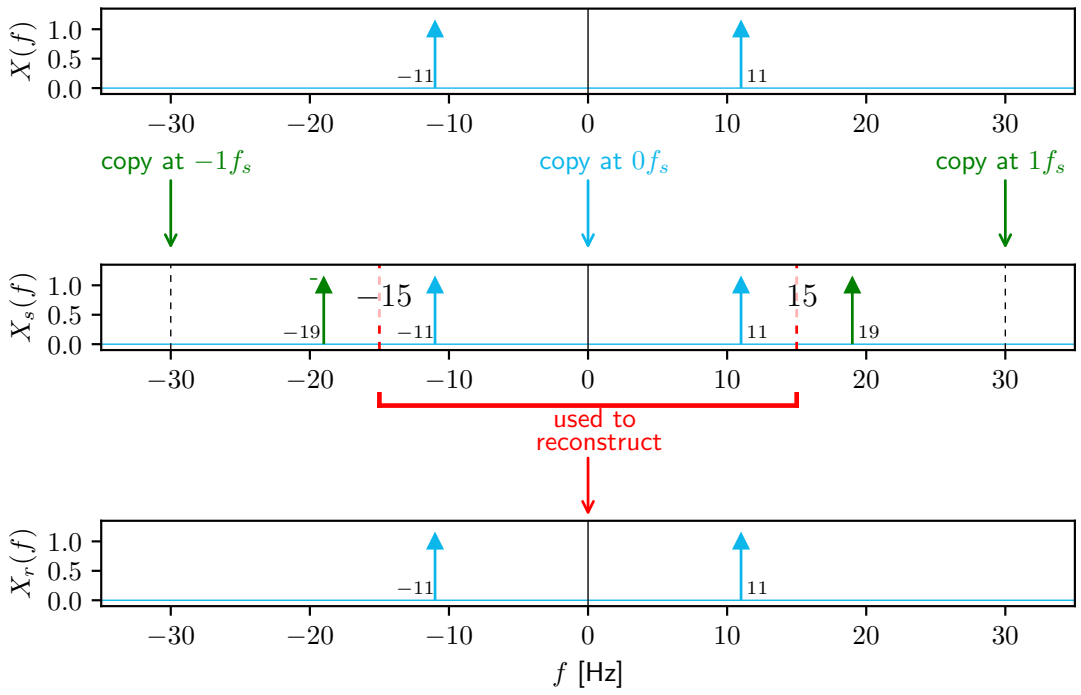


Figure 9.9: Fourier transforms of original signal, $X(f)$ (top), sampled signal $X_s(f)$ (middle) and reconstructed signal $X_r(f)$ (bottom). The original signal is a zero-phase cosine with amplitude 2 and frequency $f_0 = 11$ Hz. The signal is sampled with $f_s = 30$ Hz.

In the second row of Figure 9.9, the first three copies are shown. The first copy is positioned at $f = 0f_s$ (for $k = 0$), with two Dirac pulses at ± 11 Hz (blue). One copy is positioned at $f = 1f_s$ ($k = 1$), yielding Dirac pulses at $30 - 11 = 19$ Hz (green) and $30 + 11 = 41$ Hz (not shown). One copy is positioned at $f = -1f_s$ ($k = -1$), yielding Dirac pulses at $-30 - 11 = -41$ Hz (not shown) and $-30 + 11 = -19$ Hz (green).

An infinite number of other copies occur, which are not shown, as the frequency axis runs here from $f = -35$ to 35 Hz. For instance, the 'next' copy at positive frequencies would be placed at $2f_s = 60$ Hz (for $k = 2$), resulting in Dirac pulses at $60 - 11 = 49$ Hz and at $60 + 11 = 71$ Hz.

In the bottom image we used $X_s(f)$ between $\pm\frac{f_s}{2} = \pm 15$ Hz to reconstruct the signal. What results is $X_r(f)$, which is *identical* to $X(f)$ shown at the top. We have sampled the signal properly, and so it can be perfectly reconstructed from its samples. Signal reconstruction will be discussed in more detail in Chapter 10.

Repeat Example 9.1, but use sampling frequency $f_s = 10$ Hz. Explain the result. EX 9.2

Solution According to the sampling theorem, (9.12), the sampling frequency is too low, and aliasing will occur. Figure 9.10 shows all Fourier transforms. Again, the top image shows $X(f)$, with two Dirac pulses at ± 11 Hz, with weight 1. The second image, for the sake of explanation, shows an *intermediate* result to the Fourier transform, namely the Fourier transform of the sampled signal when copying $X(f)$ to $f = 0f_s$ ($k = 0$, blue), to $f = 1f_s = 10$ Hz ($k = 1$, green) and to $-1f_s = -10$ Hz ($k = -1$, green). You will see that Dirac pulses appear at, respectively, ± 11 Hz, at $10 - 11 = -1$ and $10 + 11 = 21$ Hz, and at $-10 - 11 = -21$ and $-10 + 11 = 1$ Hz.

The third image shows the complete Fourier transform $X_s(f)$, for the range of frequencies in the figure, with different colors to distinguish between the copies (purple: copies at $\pm 2f_s = \pm 20$ Hz ($k = \pm 2$), yellow: copies at $\pm 3f_s = \pm 30$ Hz ($k = \pm 3$), black: copies at $\pm 4f_s = \pm 40$ Hz ($k = \pm 4$)) and the numbers above the Dirac pulse show the 'integer' of the integer multiple k in kf_s.

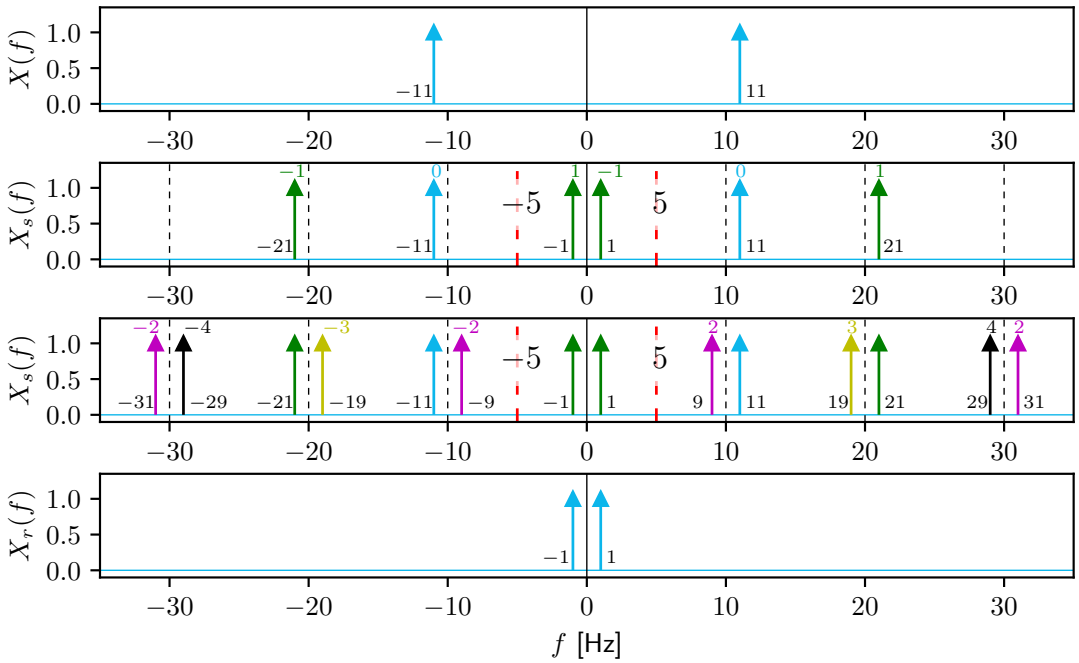


Figure 9.10: Fourier transforms of original signal, $X(f)$ (top), sampled signal $X_s(f)$ (second and third images) and reconstructed signal $X_r(f)$ (bottom). The original signal is a zero-phase cosine with amplitude 2 and frequency $f_0 = 11$ Hz. The signal is sampled with $f_s = 10$ Hz.

On the bottom image we used $X_s(f)$ between $\pm\frac{f_s}{2} = \pm 5$ Hz to reconstruct the signal. What results is $X_r(f)$, which is clearly different from $X(f)$ in the top row. We obtain a zero-phase cosine with amplitude 2 and frequency 1 Hz, i.e., the situation shown in Figure 9.2 (though with unit amplitude). We did not sample the signal properly, and so, when we reconstruct it from the samples, we obtain instead an *alias*, not the original signal.

In this example we know the original continuous-time signal and its Fourier transform, but in practice we typically do not, and therefore drawing conclusions based on $X_r(f)$ could lead to serious mistakes.

In these examples, and also in the rationale leading to the sampling theorem, we already alluded to the possibility that, after obtaining a sequence of samples x_n of continuous-time signal $x(t)$, at some point it may be important to perform the inverse operation. This is known as *signal reconstruction* and will be discussed in more detail in the next chapter.

10

Signal reconstruction

In Chapter 9 we discussed how a continuous-time signal $x(t)$, for infinite measurement duration, can be sampled, ultimately yielding a discrete-time sequence x_n in a computer. In this chapter, the reverse process, called signal reconstruction, is discussed. That is, is it possible to *reconstruct* the original continuous-time signal $x(t)$ from the sampled signal? This may seem an impossible task, as we do not know the value of $x(t)$ in between sample times. However, when we assume that the signal is *band-limited* and has been sampled with a sampling frequency f_s according to the sampling theorem (Section 9.4) it is shown that the signal can indeed be perfectly reconstructed from the sampled signal. This *ideal* signal reconstruction is a limiting case in theory, and cannot be applied in real-time applications. Hence, several other signal reconstruction procedures have been developed which *can* be used in real time, such as the Zero-Order Hold (ZOH) technique.

10.1 Ideal signal reconstruction

The transformation of a sampled signal back to a continuous-time signal is called signal reconstruction. In the following, we assume that a continuous-time signal $x(t)$ has been sampled with a sufficiently high sampling frequency f_s, such that *no aliasing* occurs. We will work with the continuous-time model of our samples, $x_s(t)$, to explain the principles, and continue with example signal $x(t) = \mathrm{sinc}^2(t)$ from Figure 8.2, as it has a convenient (real and even) Fourier transform. The reconstructed signal is referred to as $x_r(t)$ and its Fourier transform $X_r(f)$.

Figure 10.1 (top) illustrates the Fourier transform $X_s(f)$ of $x_s(t)$, showing the (infinite) repetition of the Fourier transform $X(f)$ at integer multiples of the sampling frequency f_s, where each copy lies in a band of frequencies with a width of f_s. A straightforward way to obtain the desired $X_r(f)$ from $X_s(f)$ is to multiply $X_s(f)$, (9.10), by a unit-amplitude pulse $H_r(f)$ with width f_s:

$$X_r(f) = H_r(f)X_s(f) \tag{10.1}$$

with:

$$H_r(f) = \Pi\left(\frac{f}{f_s}\right) \tag{10.2}$$

This operation is known as *filtering* and will be discussed in detail in Chapter 18. In fact, $H_r(f)$ is a so-called *ideal* filter: it perfectly passes a part of the spectrum (here, all frequencies between $-\frac{f_s}{2}$ and $+\frac{f_s}{2}$) and perfectly 'blocks' all other parts, see Figure 10.1 (middle). Expressed as in (10.2), $H_r(f)$ is known as the *ideal reconstruction filter*.

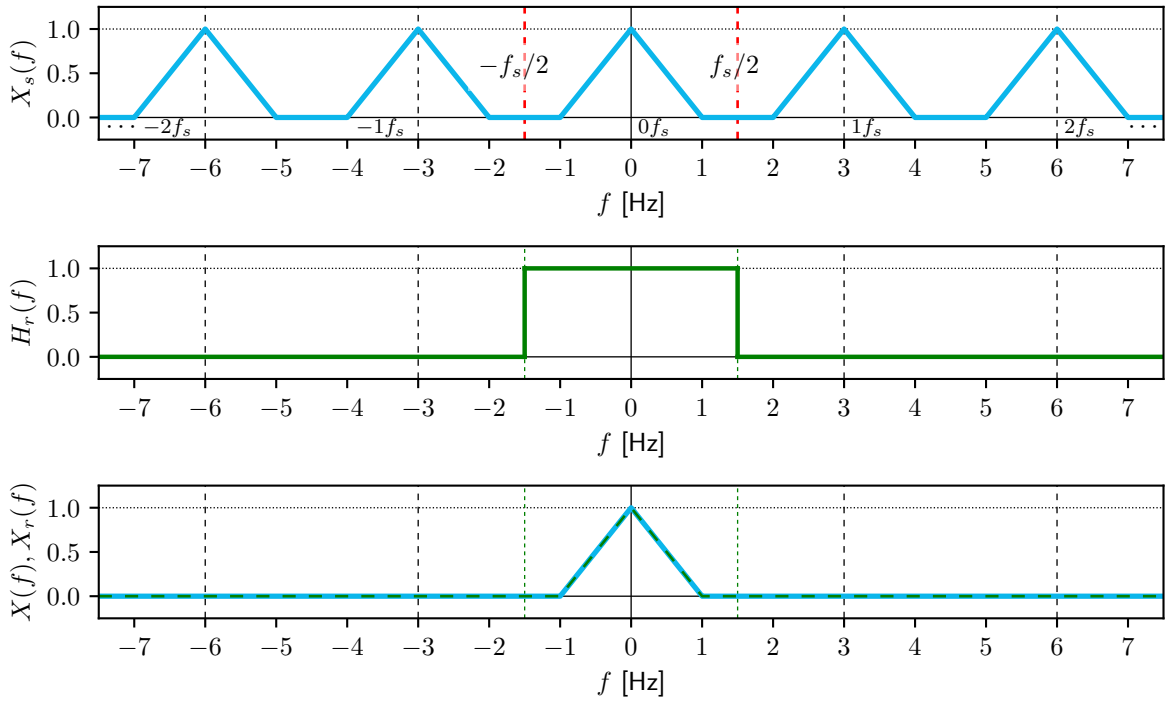


Figure 10.1: Top: Fourier transform of $x_s(t)$, with sampling frequency $f_s = 3$ Hz. Middle: Ideal reconstruction filter $H_r(f)$. Bottom: Fourier transform $X_r(f)$, (10.1), of the reconstructed signal $x_r(t)$ (green, dashed) and the original, $X(f)$ (blue), with $x(t) = \text{sinc}^2(t)$. Note that all signal transforms shown here are real and even.

Using the ideal reconstruction filter, for infinite signal duration, the original Fourier transform $X(f)$ is reconstructed perfectly, see Figure 10.1 (bottom):

$$X_r(f) = \Pi\left(\frac{f}{f_s}\right) \sum_{k=-\infty}^{\infty} X(f - kf_s) \;=\; X(f) \tag{10.3}$$

Because the Fourier transform $X_r(f)$ of the reconstructed signal $x_r(t)$ is identical to the Fourier transform $X(f)$ of $x(t)$, the reconstructed time signal $x_r(t)$ is identical to the original $x(t)$. Whereas in the frequency domain this is a trivial result, it is perhaps less trivial in order to understand what happens in the time domain. After all, we only have the sampled signal and do not know the values of $x(t)$ in between the samples.

Recall that a multiplication of the Fourier transforms of two signals in the frequency domain is equivalent to Fourier-transforming a *convolution* of these two signals in the time domain, (6.8). Generally we try to avoid working with convolution integrals; with one exception, and that is when one of the signals being convolved is a (sum of) Dirac delta function(s), which is the case here, as the model of our samples $x_s(t)$ is an infinite sum of Dirac functions, (9.2).

Therefore, since $X_r(f) = H_r(f)X_s(f)$, then $x_r(t) = h_r(t) * x_s(t)$. We can obtain $h_r(t)$ through inverse Fourier-transforming $H_r(f)$:

$$h_r(t) = \mathcal{F}^{-1}\{\Pi\left(\frac{f}{f_s}\right)\} \;=\; f_s\,\text{sinc}(f_s t) = \frac{1}{\Delta t}\,\text{sinc}\left(\frac{t}{\Delta t}\right) \tag{10.4}$$

This function is the so-called *impulse response function* of the ideal filter, and will be discussed in more detail in Chapter 18. Figure 10.2 (at left) illustrates that $h_r(0) = f_s$ and that $h_r(t)$ has its zero-crossings at all integer multiples of Δt, except at $t = 0$:

$$h_r(t) = 0 \quad \text{for} \quad t = m\Delta t, \quad m \in \mathbb{Z} \setminus \{0\} \tag{10.5}$$

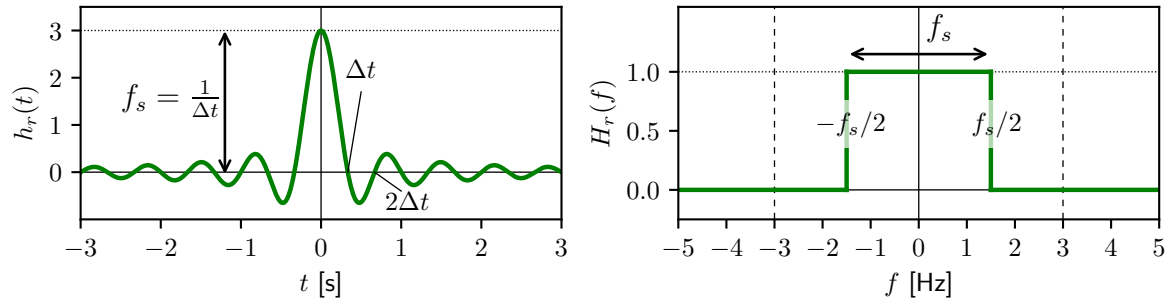


Figure 10.2: Left: impulse response of ideal filter $h_r(t)$; right: ideal filter $H_r(f)$ (sampling frequency $f_s = 3$ Hz).

Convolving $h_r(t)$ with $x_s(t)$ yields:

$$x_r(t) = h_r(t) * x_s(t) = \sum_{n=-\infty}^{\infty} x(n\Delta t)\,\mathrm{sinc}\left(\frac{t - n\Delta t}{\Delta t}\right) \tag{10.6}$$

which is a summation of all samples $x(n\Delta t)$ multiplied by the function $\Delta t\, h_r(t)$ located *at* the position $t = n\Delta t$ of each individual sample on the continuous-time axis.

Proof With $h_r(t) = \frac{1}{\Delta t}\,\mathrm{sinc}\left(\frac{t}{\Delta t}\right)$ and $x_s(t) = \Delta t x(t) \sum_{n=-\infty}^{\infty} \delta(t - n\Delta t)$ we obtain:

$$\begin{aligned}
x_r(t) &= \frac{1}{\Delta t}\,\mathrm{sinc}\left(\frac{t}{\Delta t}\right) * \left(\Delta t x(t) \sum_{n=-\infty}^{\infty} \delta(t - n\Delta t)\right) \\
&= \int_{\tau=-\infty}^{\infty} \left(\mathrm{sinc}\left(\frac{t-\tau}{\Delta t}\right) x(\tau) \sum_{n=-\infty}^{\infty} \delta(\tau - n\Delta t)\right) \mathrm{d}\tau \\
&= \sum_{n=-\infty}^{\infty} \int_{\tau=-\infty}^{\infty} \mathrm{sinc}\left(\frac{t-\tau}{\Delta t}\right) x(\tau)\delta(\tau - n\Delta t)\,\mathrm{d}\tau \\
&= \sum_{n=-\infty}^{\infty} x(n\Delta t)\,\mathrm{sinc}\left(\frac{t - n\Delta t}{\Delta t}\right)
\end{aligned}$$

Here we changed the order of integration and summation in the third step, and used the sifting property of the Dirac delta function in the final step.

Figures 10.3 and 10.4 illustrate the process of building $x_r(t)$ using an *infinite* summation of sinc functions, (10.6). Figure 10.3 shows the multiplication of sample value $x(n\Delta t) = x_n$, (9.1), by the sinc function for five individual samples, $n = 0, \ldots, 4$. Starting with the sample at $n = 0$ (top), the sinc function (green curve) is scaled by the value of the sample at $n = 0$, x_0. We obtain a continuous-time function which *at* the time of this sample, $t = 0\Delta t$, exactly equals the value of this sample, x_0. In addition, because its zero-crossings occur at all integer multiples of Δt, (10.5), this sinc function will *not* affect the value of the reconstructed signal at any other sample time. Repeating this procedure for the samples, $n = 1, \ldots, 4$, shown in the other plots, we obtain a scaled sinc function *at* each sample time, with its maximum exactly equal to the value *of* that sample, *not* affecting the value at any other sample time.

Figure 10.4 illustrates, as part of the summation (10.6), the summation of the sinc functions obtained for samples $n = -1$, $n = 0$ and $n = 1$ (top), starting from the left and assuming that

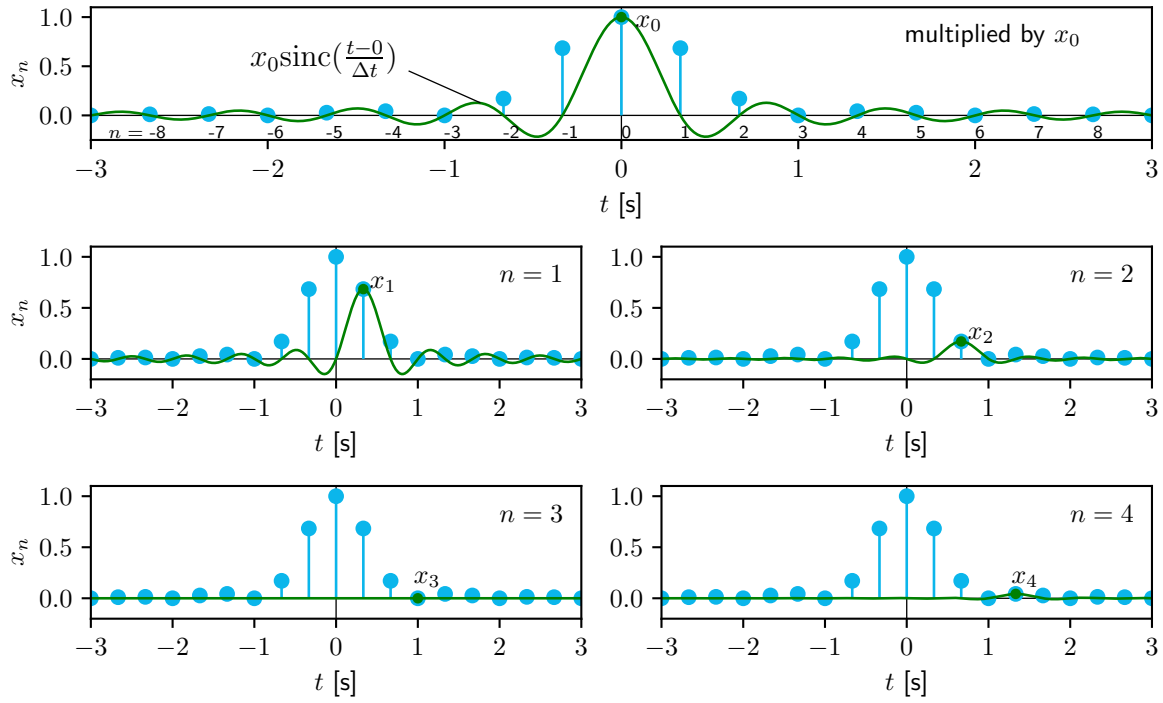


Figure 10.3: Multiplication of the shifted sinc function $h_r(t)$, (10.4), times Δt (green curve) by sample values x_n (blue stems), for $n = 0, \ldots, 4$, to construct $x_r(t)$, (10.6). Only the top plot shows indices n; these are absent in the other plots to prevent clutter.

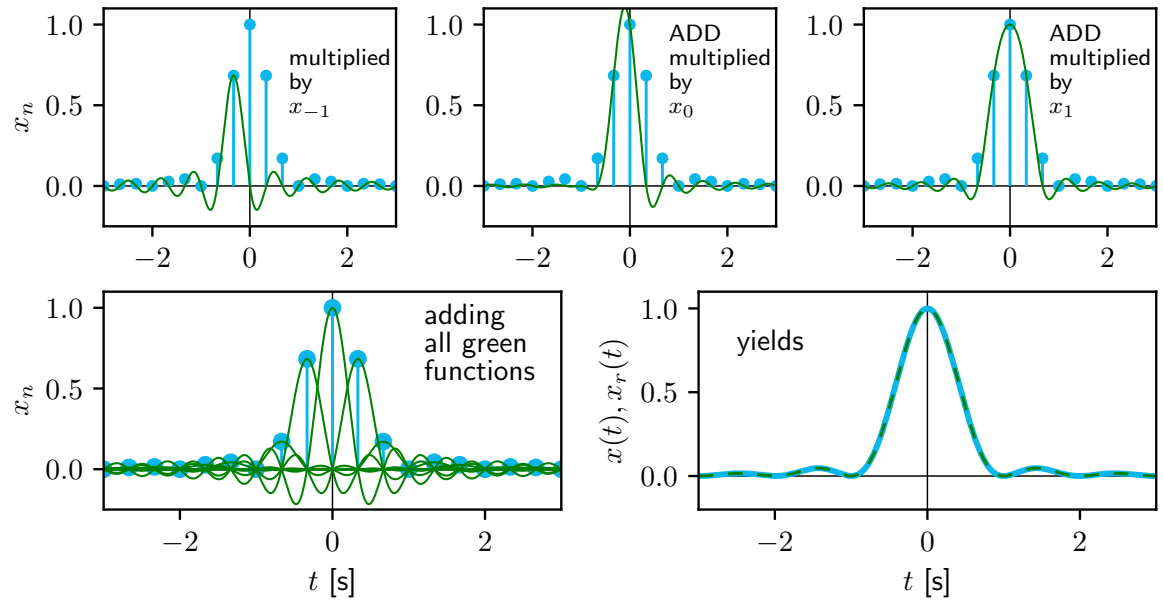


Figure 10.4: Top: multiplication of shifted sinc function $\Delta t\, h_r(t)$ by x_{-1} (left), then adding to that function the multiplication of sinc function $\Delta t\, h_r(t)$ by x_0 (center), then adding to that function the multiplication of shifted sinc function $\Delta t\, h_r(t)$ by x_1 (right). Bottom left: result of multiplication of shifted sinc functions $\Delta t\, h_r(t)$ by all individual samples. Bottom right: adding all green functions in the bottom left figure yields $x_r(t)$, (10.6), (dashed green), the perfectly reconstructed $x(t)$ (blue). Note that the indices n are not shown to prevent clutter.

the contribution of all samples before $n = -1$ is small (which is valid in this case). We see that adding the sinc functions belonging to just these three samples (green curves) already approximates this particular continuous-time function $x(t)$ quite well. The final result, obtained when adding the convolutions for *all* samples (bottom left), for infinite signal duration, is shown at the bottom right of Figure 10.4, where, ultimately, the original continuous-time signal $x(t)$ is obtained. Through (10.6), $x(t)$ has been perfectly reconstructed from the sampled signal: $x_r(t) = x(t)$.

EX 10.1

Recall Figure 9.8, which shows the Fourier transform $X_s(f)$ of the sampled version of signal $x(t) = \text{sinc}^2(t)$ when sampling too slowly, $f_s = \frac{3}{2}$ Hz < 2 Hz; hence aliasing occurs. Use the ideal reconstruction filter to reconstruct the signal, leading to $x_r(t)$. What does $x_r(t)$ look like?

Solution Multiplication of $X_s(f)$ in Figure 9.8 by $H_r(f)$ leads to $X_r(f)$ shown at right in Figure 10.5. This real and even function can be described as:

$$X_r(f) = \frac{1}{2}\Pi\left(\frac{f}{\frac{3}{2}}\right) + \frac{1}{2}\Lambda\left(\frac{f}{\frac{1}{2}}\right) \tag{10.7}$$

Inverse Fourier-transforming (10.7) results in $x_r(t)$:

$$x_r(t) = \frac{3}{4}\,\text{sinc}(\frac{3}{2}t) + \frac{1}{4}\,\text{sinc}^2(\frac{1}{2}t) \tag{10.8}$$

which is, clearly, not equal to the original signal $x(t) = \text{sinc}^2(t)$. We obtain an *alias* of $x(t)$, $x_r(t)$, the samples of which are *exactly equal* to the samples of $x(t)$. The time-domain signal $x_r(t)$ and its Fourier transform $X_r(f)$ are illustrated in Figure 10.5.

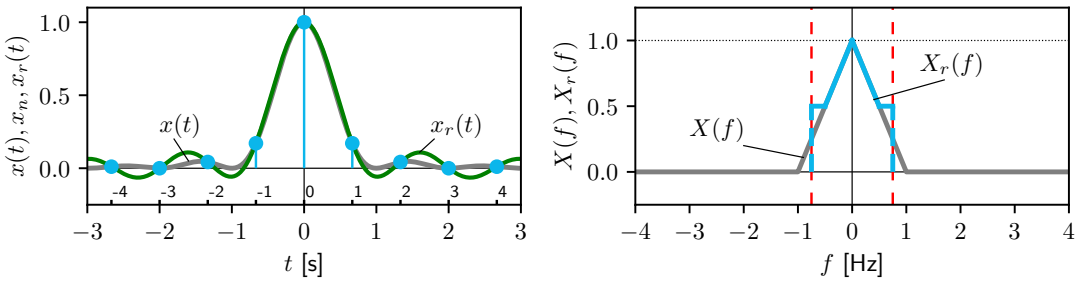


Figure 10.5: Left: $x(t)$ (gray), sampled with $f_s = 1.5$ Hz (blue), and reconstructed signal $x_r(t)$ (green). Right: Fourier transforms $X(f)$ of $x(t) = \text{sinc}^2(t)$ (gray), and $X_r(f)$ of reconstructed signal $x_r(t)$ (blue).

10.2 Zero-order hold reconstruction

Given that perfect signal reconstruction is possible, we may ask: is there a need for alternative methods? The answer is yes, and lies in the fact that an ideal filter is a *non-causal* filter, Chapter 16. For now, it is sufficient to know that non-causal filters need past, present and *future* values of a signal to work properly. In practical, real-time applications we only know the current and past values of a signal and need to apply a *causal* filter.

The most elementary (and obvious) way to reconstruct a continuous-time signal from past and present samples in practice is to *hold* the present sample value x_n up until the next sample,

x_{n+1}, becomes available, also known as sample-and-hold:

$$x_r(t) = x_n \quad \text{for} \quad n\Delta t \leq t < (n+1)\Delta t \tag{10.9}$$

This process is known as *Zero-Order Hold* (ZOH) reconstruction, and can be modeled as the convolution, in time, of the sampled signal $x_s(t)$ with a scaled pulse, with a width Δt, and shifted to the right by $\frac{\Delta t}{2}$:

$$x_r(t) = h_r(t) * x_s(t) \tag{10.10}$$

with:

$$h_r(t) = \frac{1}{\Delta t}\Pi\left(\frac{t - \frac{\Delta t}{2}}{\Delta t}\right) \tag{10.11}$$

Because the derivation is very similar to the ideal reconstruction case (Section 10.1) we present, without proof, the formulation for the reconstructed signal $x_r(t)$ using ZOH:

$$x_r(t) = \sum_{n=-\infty}^{\infty} x(n\Delta t)\Pi\left(\frac{t - n\Delta t - \frac{\Delta t}{2}}{\Delta t}\right) \tag{10.12}$$

which is a piecewise-continuous function of time.

The process is illustrated in Figures 10.6 and 10.7, following the same reasoning as for the ideal signal reconstruction (Figures 10.3 and 10.4). The ZOH reconstruction is more crude than the ideal reconstruction, as it results in a staircase continuous-time function. Sampling faster improves the reconstruction, as illustrated in Figure 10.8. This makes perfect sense: when the sampling interval becomes smaller, the 'step size' of the staircase becomes smaller as well, thereby more closely approaching $x(t)$.

ZOH reconstruction is easy to understand in the time domain. Clearly, a loss of accuracy results: the reconstructed signal will *not* be exactly equal to the original signal $x(t)$. Hence, its Fourier transform $X_r(f)$ will *not* equal $X(f)$.

To illustrate what happens in the frequency domain, recall that the Fourier transform of a convolution of two time signals equals the *multiplication* of the Fourier transforms of those signals, (6.8). Therefore, since $x_r(t) = h_r(t) * x_s(t)$, then $X_r(f) = H_r(f)X_s(f)$, as in (10.1). We can obtain $H_r(f)$ through Fourier-transforming $h_r(t)$ from (10.11):

$$H_r(f) = \text{sinc}(\Delta t f)e^{-j2\pi f\frac{\Delta t}{2}} \tag{10.13}$$

where we used (5.11) and (6.2). The ZOH reconstruction filter $H_r(f)$ is a complex-valued function of frequency. The impulse response function of this filter, $h_r(t)$, as well as the magnitude of its Fourier transform, $|H_r(f)|$, are illustrated in Figure 10.9. The ZOH reconstruction filter approximates the ideal filter well for small frequencies (near the maximum of the sinc function), but *extends beyond* the range $[-\frac{f_s}{2}, \frac{f_s}{2}]$ which means that, when using this filter to reconstruct the signal from its samples, part of the (energy contained in the) copies of $X_s(f)$ ends up in $X_r(f)$.

This is illustrated in Figure 10.10. The left and right columns show the ZOH reconstruction for sampling frequencies of 3 and 10 Hz, respectively. Sampling with 3 Hz creates copies at integer multiples of 3 Hz, which are then multiplied by $H_r(f)$ to obtain $X_r(f)$.

The amplitude spectrum of the reconstructed signal, $|X_r(f)|$, is similar to the amplitude spectrum of the original signal, $|X(f)|$, at lower frequencies, but contains scallops around

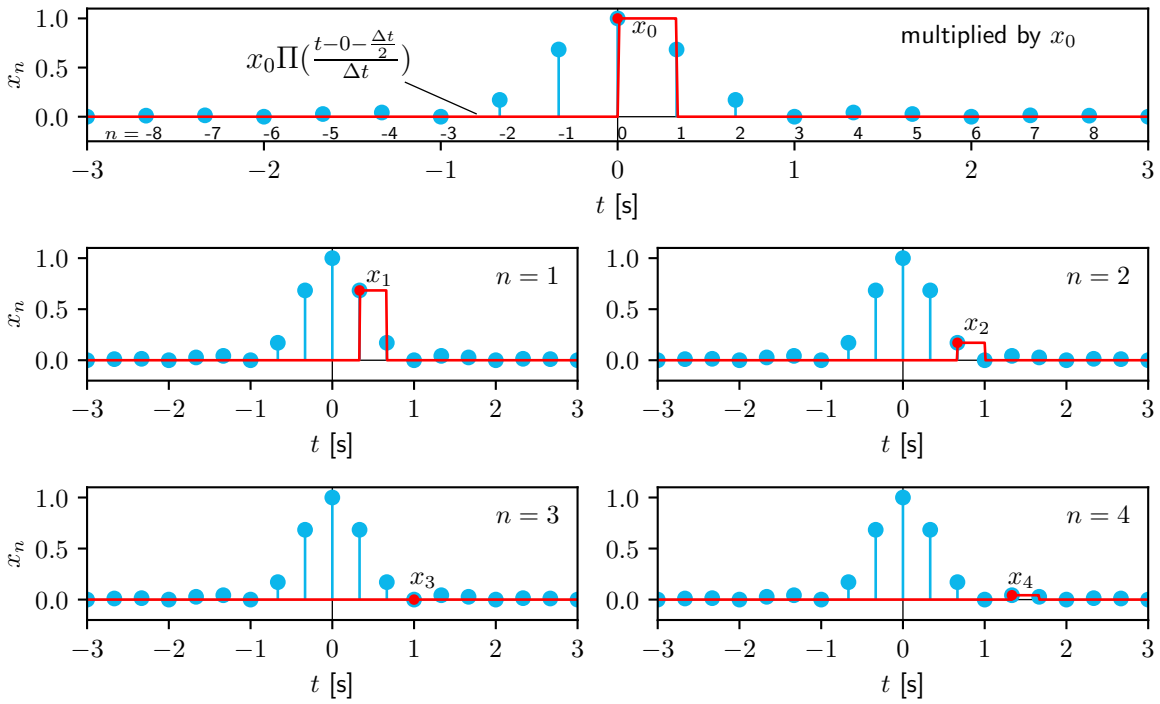


Figure 10.6: Multiplication of the shifted pulse $h_r(t)$, (10.11), times Δt (red line) by sample values x_n (blue stems), for $n = 0, \ldots, 4$. Only the top plot shows indices n; these are absent in the other plots to prevent clutter.

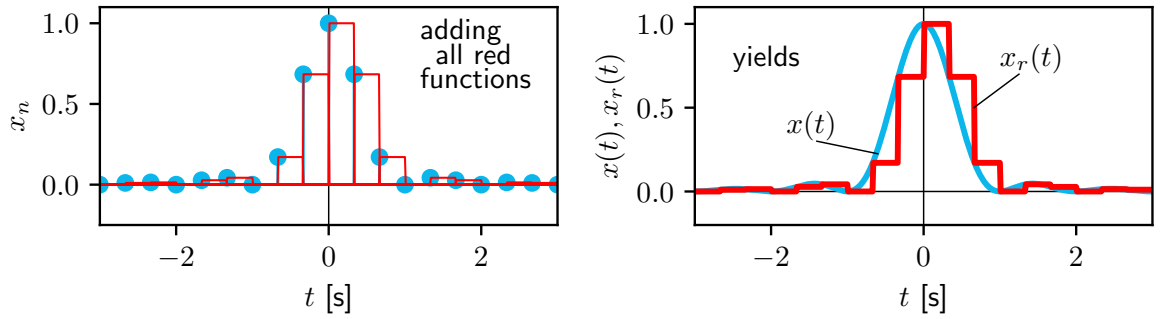


Figure 10.7: Adding all red functions in the left figure yields the reconstructed signal $x_r(t)$ (red) at right, with original $x(t)$ in blue. Note that in this figure, and the following, the indices n are not shown to prevent clutter.

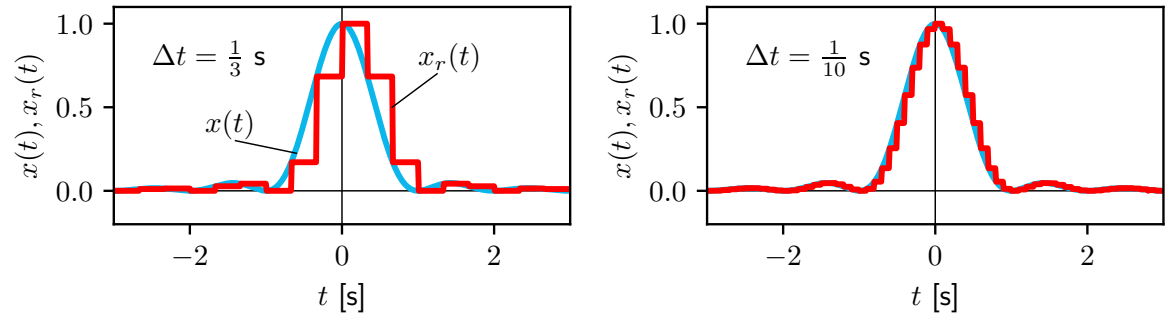


Figure 10.8: ZOH reconstruction with sampling frequency $f_s = 3$ Hz (left) and $f_s = 10$ Hz (right). The reconstructed signal $x_r(t)$, (10.12), is shown in red, the original $x(t) = \mathrm{sinc}^2(t)$ in blue.

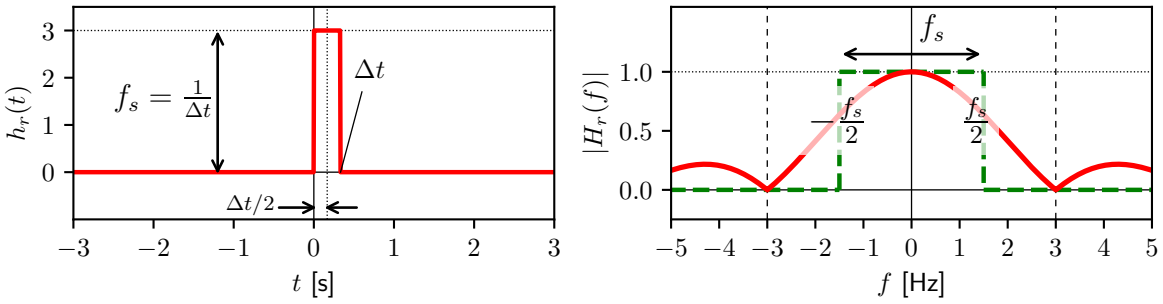


Figure 10.9: Left: impulse response of ZOH reconstruction filter $h_r(t)$ (sampling frequency $f_s = 3$ Hz). Right: magnitude of Fourier transform of ZOH filter, $|H_r(f)|$ (red); the ideal filter is also shown (green, dashed).

integer multiples of 3 Hz where the energy of the copies ends up in $|X_r(f)|$. Apparently, the staircase shape of $x_r(t)$ in the ZOH reconstruction in Figure 10.8 requires its amplitude spectrum $|X_r(f)|$ to have energy at the higher frequencies to build the 'steps' in time.

Sampling faster makes these steps smaller and the amplitude spectrum of $x_r(t)$ more closely resembles the amplitude spectrum of $x(t)$, Figure 10.10 at right. Sampling with 10 Hz creates copies at integer multiples of 10 Hz, to be multiplied by $H_r(f)$ to obtain $X_r(f)$. The spectrum of the reconstructed signal is very similar to the spectrum of the original signal for a larger range of frequencies around $f = 0$, compared to the graph at left. Smaller scallops occur around integer multiples of 10 Hz where the energy of the copies ends up in $|X_r(f)|$.

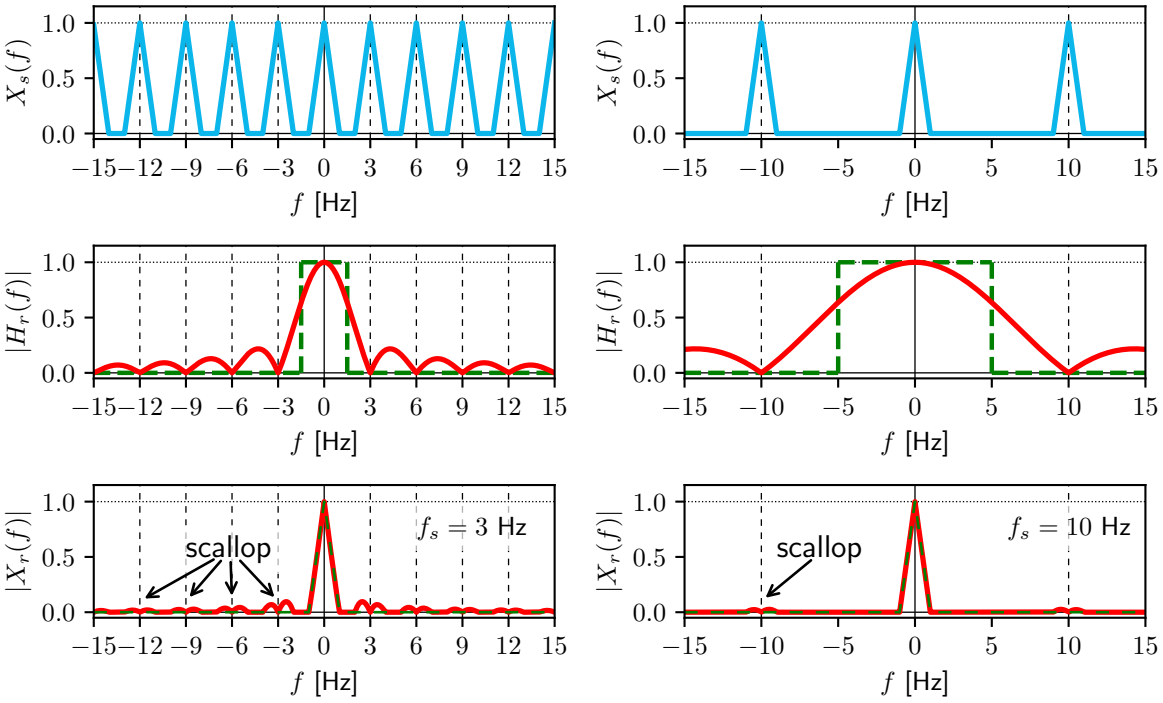


Figure 10.10: Left and right columns show results for sampling with, respectively, $f_s = 3$ Hz and 10 Hz. Top: Fourier transform $X_s(f)$ of the sampled signal. Middle: magnitude of Fourier transform of ZOH filter, $|H_r(f)|$ (red), and ideal filter (green, dashed). Bottom: magnitude of Fourier transforms of the reconstructed signal, $|X_r(f)|$, using ZOH (red) and ideal filter (green, dashed); the latter is identical to $|X(f)|$.

EX 10.2

A real and even cosine signal $x(t) = 2\cos(2\pi t)$ has been sampled with sampling frequency $f_s = 3$ Hz. Use a ZOH filter to reconstruct a continuous-time signal from the sampled signal, $x_r(t)$, with Fourier transform $X_r(f)$. Show the magnitude and phase spectra of this Fourier transform and explain the result. Repeat the example for a sampling frequency of 10 Hz and explain the differences.

Solution The cosine has a frequency $f_0 = 1$ Hz; its Nyquist rate is 2 Hz and no aliasing occurs for both sampling frequencies. Its Fourier transform equals $X(f) = \delta(f+1) + \delta(f-1)$. The Fourier transform of the impulse train model of the samples, $X_s(f)$, will have copies of $X(f)$ at all integer multiples of the sampling frequency. When reconstructing using a ZOH filter, the Dirac pulses in $X_s(f)$ (all with a weight of 1) are multiplied by the ZOH filter $H_r(f)$, (10.13), yielding $X_r(f)$, see (10.1). This latter function is zero at all frequencies, except at the frequencies of the Dirac pulses that occur in $X_s(f)$. At these frequencies, it equals $X_s(f)H_r(f)$, the Dirac pulses get a *complex*-valued weight, with magnitude $|X_r(f)| = |X_s(f)H_r(f)| = |H_r(f)|$ and phase $\angle X_r(f) = \angle(X_s(f)H_r(f)) = \angle H_r(f)$. Hence, at all frequencies in $X_s(f)$ that contain Dirac pulses, $X_r(f)$ equals $H_r(f)$.

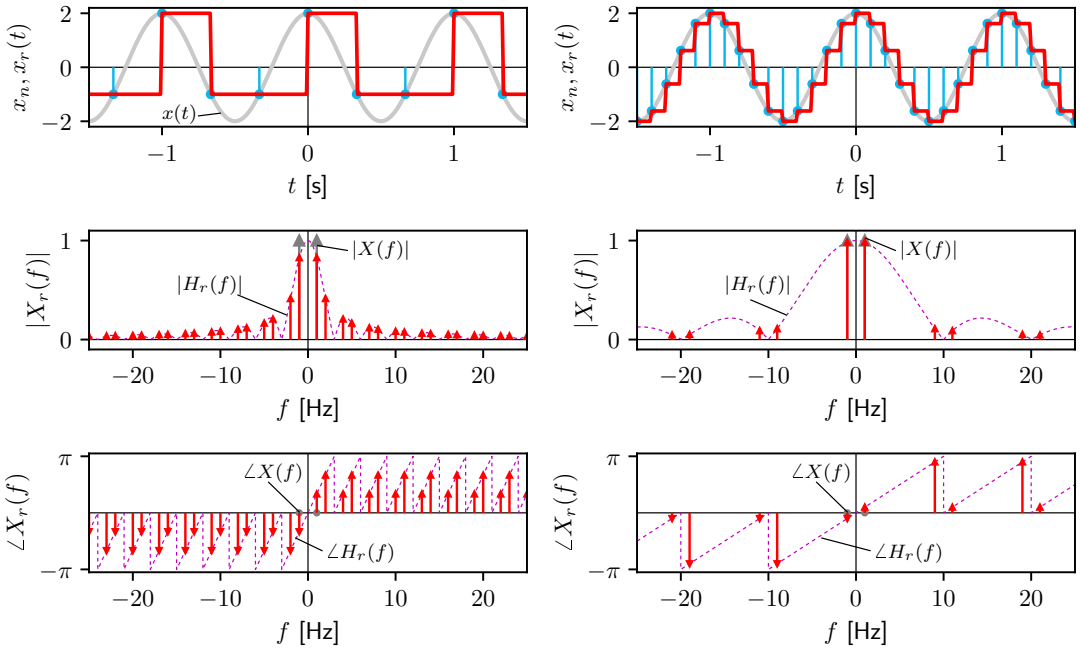


Figure 10.11: Left and right columns show results for sampling with, respectively, $f_s = 3$ Hz and 10 Hz. Top: time-domain signal $x(t)$ (gray), the samples x_n (blue stems) and reconstructed signal $x_r(t)$ (red). Middle: amplitude spectra of $X(f)$ (gray), $H_r(f)$ (purple, dashed) and $X_r(f)$ (red). Bottom: phase spectra of $X(f)$ (gray; just at $f = \pm 1$ Hz), $H_r(f)$ (purple, dashed) and $X_r(f)$ (red).

Figure 10.11 illustrates the results for sampling frequencies $f_s = 3$ (at left) and 10 Hz (at right). Starting with the left column, the top figure shows the original cosine $x(t)$ in gray, its samples x_n using blue stems and the reconstructed signal $x_r(t)$ in red. Clearly, although meeting the sampling theorem, the signal has been poorly sampled and reconstructed: $x_r(t)$ differs greatly from $x(t)$, it resembles a block wave. This is reflected in its Fourier transform $X_r(f)$, shown in the middle (amplitude spectrum) and bottom (phase spectrum) plots. We obtain a large array of Dirac pulses at all frequencies to which the original Fourier transform $X(f)$ is copied: $\pm 1, \pm 2, \pm 4, \pm 5, \ldots$ Hz.

These Dirac pulses, at each frequency are multiplied by the magnitude of $H_r(f)$ at that frequency. The phase at each frequency equals the phase of $H_r(f)$ at that frequency, as $\angle X(f) = \angle X_s(f) = 0 \; \forall \; f$ of the Dirac pulses. Apparently, we need a large number of shifted cosines, with these frequencies, to 'model' $x_r(t)$.

Sampling faster yields a better reconstruction, shown in Figure 10.11 at right. Although still far from perfect, $x_r(t)$ is clearly a cosine wave, and its Fourier transform represents $X(f)$ (shown in gray) more closely. However, also in this case, the ZOH-reconstructed signal is not even, and its Fourier transform is complex-valued.

Several other (more advanced) reconstruction techniques exist. Examples are the First-Order Hold (FOH), which applies a piecewise linear interpolation between consecutive samples using the triangular Λ-function for $h_r(t)$, and higher-order hold techniques. Sampled, discrete-time signals can also be converted to analog, continuous-time signals using (low-pass) electronic filters, discussed in Chapter 18.

The Exercises book contains extensions to the theory explained in Chapters 9 and 10. For instance, what about sampling signals which have a band-limited, *rectangular* spectrum, such as the sinc function. And, perhaps more importantly, sampling signals that are *non*-band-limited; these are mostly periodic signals such as the square wave function.

In this and the previous chapter we discussed the Fourier transform $X_s(f)$ of the impulse train model of the samples $x_s(t)$ for infinite-duration signals. The consequences of sampling and signal reconstruction could all be introduced while still working in continuous time. In the next chapter, we move our discussion to discrete time, and introduce the Fourier transform of an infinite-duration *sequence* of samples x_n, the Discrete-Time Fourier Transform (DTFT). In Chapter 12 we abandon the infinite signal duration assumption and discuss the practical case: working with a finite number of samples, using the Discrete Fourier Transform (DFT). A summary of all steps is presented in Appendix I.

11

Discrete-Time Fourier Transform

In Chapter 9 the process and consequences of sampling an infinite-duration, continuous-time signal $x(t)$ have been discussed. The sequence of samples x_n was modeled by introducing the impulse train-sampled signal $x_s(t)$, (9.2). This signal is defined solely by the samples and the sample interval Δt; it loses any information of the values of $x(t)$ in between the sampling times. It was shown that $x_s(t)$ can be Fourier-transformed, yielding:

$$X_s(f) = \sum_{k=-\infty}^{\infty} X(f - kf_s) \tag{9.10}$$

This Fourier transform is *periodic* in frequency with sampling frequency f_s: $X_s(f + mf_s) = X_s(f)\ \forall\ m \in \mathbb{Z}$. $X_s(f)$ consists of an infinite number of *copies* of the Fourier transform $X(f)$ of the original signal $x(t)$, at integer multiples of f_s. To avoid the copies overlapping, the *sampling theorem*, (9.12), states that, for band-limited signals, the sampling frequency f_s should be larger than twice the largest frequency f_h occurring in $X(f)$.

But clearly, something is missing. After all, $x_s(t)$ is still a continuous-time signal, while on the computer we have a discrete-time sequence of sample values x_n, with the index $n \in \mathbb{Z}$ as the 'running variable'. Time and frequency disappear altogether.

In this chapter we introduce the Fourier transform of an infinite-duration discrete-time sequence x_n: the Discrete-Time Fourier Transform (DTFT). It will be shown that our formulation of this transform is *identical* to $X_s(f)$. In addition, it will be explained *why* $X_s(f)$ is periodic with f_s, as a logical, inevitable consequence of moving from continuous time to discrete time.

11.1 Derivation of the DTFT

Let us return to the definition of the Fourier transform of a continuous-time signal $x(t)$:

$$X(f) = \mathcal{F}\{x(t)\} = \int_{t=-\infty}^{\infty} \underbrace{x(t)e^{-j2\pi ft}}_{z(t)}\,\mathrm{d}t \tag{5.1}$$

the integral of a complex function $z(t) = x(t)e^{-j2\pi ft}$. Suppose we approximate this Fourier integral through its Riemann sum, considering this function $z(t)$ only at discrete time instants $t = n\Delta t$. The integration differential $\mathrm{d}t$ becomes Δt, the sampling interval:

$$X(f) \approx \sum_{n=-\infty}^{\infty} \underbrace{x(n\Delta t)e^{-j2\pi fn\Delta t}}_{z(n\Delta t)=z_n}\Delta t = \Delta t \sum_{n=-\infty}^{\infty} x(n\Delta t)e^{-j2\pi fn\Delta t} \tag{11.1}$$

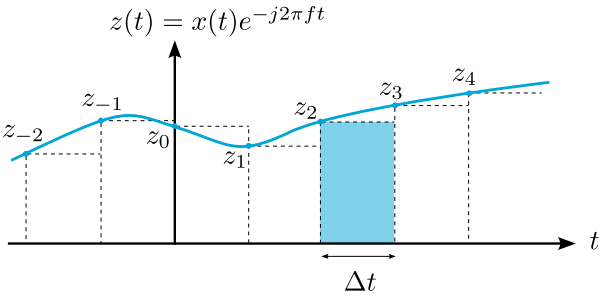


Figure 11.1: Discretization of the Fourier transform integral, (5.1), through a Riemann sum approximation, (11.1). The blue curve illustrates (the real part of) function $z(t)$, discretized at time instants $t = n\Delta t$.

In doing so, the Fourier transform integral has been 'discretized', resembling, but not being equal to, the process of sampling. In this approximation, we assume $z(t)$ to be *constant* between time instants $n\Delta t$ and $(n+1)\Delta t$, as if we would approximate it using a ZOH (Section 10.2). Figure 11.1 illustrates (the real part of) the product of signal $x(t)$ and the complex exponential $e^{-j2\pi ft}$ (for some value of frequency f); this product $z(t)$ is generally complex. In summation (11.1), the integrand $z(t)$ is considered only at discrete time instants $t = n\Delta t$, with sampling interval Δt.

The approximation of $X(f)$ would be perfect when Δt becomes infinitesimally small, or when the corresponding sampling frequency $f_s = \frac{1}{\Delta t}$ becomes infinitely large. But that is because we attempted to approximate $X(f)$, the Fourier transform of the continuous-time signal $x(t)$. What we need, however, is the Fourier transform of the sequence of samples x_n, where the value of x_n for any n not an integer is not constant, but *undefined*.

A formulation for the Discrete-Time Fourier Transform (DTFT) can be obtained by examining the summation at the right-hand side of (11.1) a little closer. For the time being, we refer to this summation as $X_{\Delta t}(f)$:

$$X_{\Delta t}(f) = \Delta t \sum_{n=-\infty}^{\infty} x(n\Delta t)e^{-j2\pi fn\Delta t} = \Delta t \sum_{n=-\infty}^{\infty} x_n e^{-j2\pi f\Delta tn} \tag{11.2}$$

This is a continuous function of frequency f, which *only* uses the sample sequence values x_n, in a similar way as the Fourier transform: the summation represents a linear combination of the multiplications of samples x_n (all real-valued) with 'building block' $e^{-j2\pi f\Delta tn}$. But what would this 'Fourier transform of the sample sequence x_n' look like? The proof below shows that $X_{\Delta t}(f)$ is *identical* to the Fourier transform $X_s(f)$ of the impulse train-sampled signal $x_s(t)$, (9.10):

$$X_{\Delta t}(f) = \sum_{k=-\infty}^{\infty} X(f - kf_s) \; =, X_s(f) \tag{11.3}$$

Proof Define continuous-time signal $y(t) = x(t)e^{-j2\pi f_1 t}$ for some frequency f_1 which we specify later. Then $Y(f) = X(f + f_1)$, (6.6). Next, define the following function, using $y(t)$ and sample time Δt:

$$\psi(t) = \sum_{n=-\infty}^{\infty} y(t + n\Delta t)$$

This function is periodic, with period Δt: $\psi(t+\Delta t)=\psi(t),\ \forall\ t\in\mathbb{R}$ and $n\in\mathbb{Z}$. Hence, $\psi(t)$ can be written as a complex exponential Fourier series, with $m\in\mathbb{Z}$:

$$\psi(t)=\sum_{m=-\infty}^{\infty}\Psi_m e^{j2\pi m f_s t} \tag{11.4}$$

with fundamental frequency $f_s=\frac{1}{\Delta t}$. The complex exponential Fourier series coefficients Ψ_m equal $f_s\,Y(mf_s)$, (G.1), with $Y(f)$ the Fourier transform of $y(t)$, see Section G.1. We obtain:

$$\psi(t)=\sum_{n=-\infty}^{\infty}y(t+n\Delta t)=f_s\sum_{m=-\infty}^{\infty}Y(mf_s)e^{j2\pi m f_s t}$$

Set $t=0$:

$$\psi(0)=\sum_{n=-\infty}^{\infty}y(n\Delta t)=f_s\sum_{m=-\infty}^{\infty}Y(mf_s)$$

Substitute the expressions for $y(t)$ and $Y(f)$ in terms of, respectively, $x(t)$ and $X(f)$:

$$\sum_{n=-\infty}^{\infty}x(n\Delta t)e^{-j2\pi f_1 n\Delta t}=f_s\sum_{m=-\infty}^{\infty}X(mf_s+f_1)$$

Shifting over *any* frequency $f_1=f$, we obtain:

$$\sum_{n=-\infty}^{\infty}x(n\Delta t)e^{-j2\pi f n\Delta t}=f_s\sum_{m=-\infty}^{\infty}X(mf_s+f)=f_s\sum_{m=-\infty}^{\infty}X(f+mf_s)$$

Substituting this equation in (11.2), and with $k=-m$, we obtain our final result:

$$X_{\Delta t}(f)=\Delta t\sum_{n=-\infty}^{\infty}x(n\Delta t)e^{-j2\pi f n\Delta t}=\sum_{k=-\infty}^{\infty}X(f-kf_s)$$

a relationship known as the *Poisson summation* [9].

The Fourier transform of the discrete-time sample sequence x_n is thus obtained, yielding the expression for the Discrete-Time Fourier Transform (DTFT):

$$\boxed{X_{\Delta t}(f)=\text{DTFT}\{x_n\}=\Delta t\sum_{n=-\infty}^{\infty}x_n e^{-j2\pi f\Delta t n}} \tag{11.5}$$

We obtain a *continuous* function of frequency f when DTFT-ing the infinite-duration sequence of samples x_n. The DTFT is *identical* to $X_s(f)$ and has the same properties. Its unit equals [unit seconds] or, equivalently, [unit/Hz], with [unit] the unit of sequence x_n.

The approximation of $X(f)$ in (11.1) is perfect *only* when the sampling frequency f_s becomes infinitely large (hence, ultimately staying in continuous time). All copies of $X(f)$ occurring at non-zero integer multiples of f_s will be positioned at infinitely-high frequencies, and $X_{\Delta t}(f)$ will indeed be identical to $X(f)$. With a limited sampling frequency f_s, the Fourier transforms of $x(t)$ and x_n are different, as the result of the latter transform consists of *copies* of the former transform centered at integer multiples of that sampling frequency f_s (11.3).

11.2 Alternative derivation and formulation

An alternative way to arrive at the definition of the DTFT (11.5) is to Fourier-transform the continuous-time impulse train model of our samples $x_s(t)$, (9.2):

$$X_s(f) = \int_{-\infty}^{\infty} x_s(t) e^{-j2\pi ft}\,\mathrm{d}t \quad \text{with} \quad x_s(t) = \Delta t\, x(t) \sum_{n=-\infty}^{\infty} \delta(t - n\Delta t) \tag{11.6}$$

We obtain:

$$\begin{aligned} X_s(f) &= \int_{-\infty}^{\infty} \left(\Delta t\, x(t) \sum_{n=-\infty}^{\infty} \delta(t - n\Delta t) \right) e^{-j2\pi ft}\,\mathrm{d}t \\ &= \Delta t \sum_{n=-\infty}^{\infty} \int_{t=-\infty}^{\infty} x(t)\delta(t - n\Delta t) e^{-j2\pi ft}\,\mathrm{d}t \;=\; \Delta t \sum_{n=-\infty}^{\infty} x(n\Delta t) e^{-j2\pi fn\Delta t} \\ &= \Delta t \sum_{n=-\infty}^{\infty} x_n e^{-j2\pi f\Delta tn} = X_{\Delta t}(f) = \mathrm{DTFT}\{x_n\} \end{aligned}$$

Note that in this alternative derivation we continue to work with the continuous-time *model* of the samples and corresponding Dirac pulses, while in the derivation of the DTFT in Section 11.1 we abandoned this model and worked directly, and *only*, with the *sequence of samples* itself.

The definition of the DTFT in (11.5) differs from the definition used in other textbooks on signal analysis, e.g., [8]. First, as explained in Section 9.2, in these books the impulse train sampling function (9.3) is *not* scaled by Δt, while we do this, for multiple reasons. Second, the term $2\pi f\Delta t$ is replaced by Ω (from $2\pi f = \omega$, hence $2\pi f\Delta t = \omega\Delta t = \Omega$). As a result, these books state the following definition of the DTFT:

$$\mathrm{DTFT}\{x_n\} = X(\Omega) = \sum_{n=-\infty}^{\infty} x_n e^{-j\Omega n} \tag{11.7}$$

Since $\Omega = 2\pi f\Delta t$, its unit is [rad] and we can verify that $X(\Omega)$ is periodic with 2π: $X(\Omega + m2\pi) = X(\Omega)$, in a similar way as $X_{\Delta t}(f)$ is periodic with f_s, (11.3). The only difference between (11.7) and our definition (11.5) is the scaling with Δt in the latter: $X_{\Delta t}(f) = \Delta t X(\Omega)$.

11.3 Inverse DTFT

For the sake of completeness, we state the expression for the inverse DTFT, returning the sequence of samples x_n from the DTFT $X_{\Delta t}(f)$:

$$x_n = \mathrm{iDTFT}\{X_{\Delta t}(f)\} = \int_{f_s} X_{\Delta t}(f) e^{j2\pi f\Delta tn}\,\mathrm{d}f \tag{11.8}$$

The proof, as well as some common DTFT pairs, is included in Appendix G. Equation (11.8) shows that the infinite-duration sequence of samples x_n is constructed by integrating the product of $X_{\Delta t}(f)$ and complex exponential $e^{j2\pi f\Delta tn}$ over all frequencies f *within a band* with a width of f_s; $X_{\Delta t}(f)$ is periodic. A consequence of sampling is that only a *band* of frequencies, with a width f_s, is available to construct the sequence x_n from its DTFT.

Similar to the complex exponential Fourier series and the Fourier transform, (11.5) is known as the analysis equation and (11.8) as the synthesis equation of the DTFT.

11.4 DTFT properties

The DTFT plays a small role in this book. Only the properties which are important in relation to the Discrete Fourier transform (DFT), see Chapter 12, are described. For more DTFT properties (e.g., linearity), theorems and pairs the reader is referred to other sources, e.g., [8].

The DTFT (of sample sequence x_n) and the Fourier transform (of continuous-time signal $x(t)$) have much in common. Consider the following DTFT properties:

- $X_{\Delta t}(f)$ is a complex-valued, continuous function of frequency f, and can be expressed in polar form, see (5.5).
- $X_{\Delta t}(f)$ has a magnitude $|X_{\Delta t}(f)|$ and phase $\angle X_{\Delta t}(f)$, and plotting these as a function of frequency yields the amplitude and phase spectra, see (5.6) and (5.7).
- $X_{\Delta t}(0)$ equals the sum of all samples multiplied by Δt, a crude integration of the discrete-time sequence x_n, see (5.8).
- $X_{\Delta t}(f)$ has a real part and an imaginary part which, respectively, correspond to the DTFT of the even and odd parts of sequence x_n, see (5.9).

Three other properties are stated here. First, as holds *identically* for $X(f)$, see (5.10):

$$\boxed{X_{\Delta t}(-f) = X^*_{\Delta t}(f)} \tag{11.9}$$

for real-valued signals, which directly follows from (11.5).

Second, clearly *in contrast* to $X(f)$, $X_{\Delta t}(f)$ is *periodic* with f_s:

$$\boxed{X_{\Delta t}(f + mf_s) = X_{\Delta t}(f)}, \qquad \forall\ m \in \mathbb{Z} \tag{11.10}$$

This latter property was already shown in Figure 9.6 for $X_s(f)$, follows from (11.3) and can be easily proven using (11.5).

Proof

$$\begin{aligned} X_{\Delta t}(f + mf_s) &= \Delta t \sum_{n=-\infty}^{\infty} x_n e^{-j2\pi(f+mf_s)\Delta t n} = \Delta t \sum_{n=-\infty}^{\infty} x_n e^{-j2\pi f \Delta t n} e^{-j2\pi m \overbrace{f_s \Delta t}^{=1} n} \\ &= \Delta t \sum_{n=-\infty}^{\infty} x_n e^{-j2\pi f \Delta t n} \underbrace{e^{-j2\pi m n}}_{=1\ \forall\ m,n \in \mathbb{Z}} = \Delta t \sum_{n=-\infty}^{\infty} x_n e^{-j2\pi f \Delta t n} = X_{\Delta t}(f) \end{aligned}$$

Third, applying (11.10), for $f = -f$ and with $m = 1$, to (11.9) yields:

$$\boxed{X_{\Delta t}(f_s - f) = X^*_{\Delta t}(f)} \tag{11.11}$$

Figure 11.2 illustrates the three properties for the DTFT $X_{\Delta t}(f)$ of example signal $x(t) = \mathrm{sinc}^2(t)$, sampled with $f_s = 3$ Hz (no aliasing), see Figure 9.6. Because $x(t)$ is real and even, the imaginary part of the DTFT of its samples is zero, and $X^*_{\Delta t}(f) = X_{\Delta t}(f)$. This signal is therefore less suited as an example to show the effects of a complex conjugate, but we added graphical elements to clarify. The periodicity of the DTFT, (11.10), is shown at the top.

Because $X_{\Delta t}(f)$ is periodic with f_s, (11.10), *any* frequency range with a width of f_s (i.e., one period) contains all information. Considering $X_{\Delta t}(f)$ within the frequency range $[-\frac{f_s}{2}, \frac{f_s}{2})$

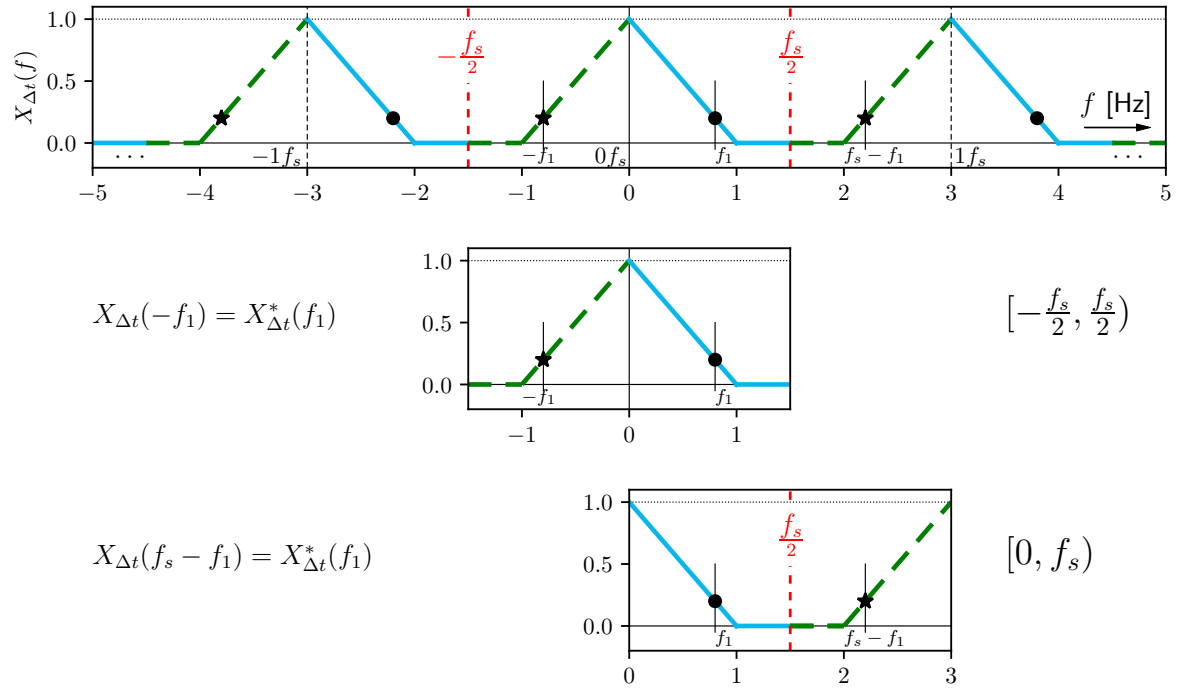


Figure 11.2: DTFT $X_{\Delta t}(f)$ of x_n, the discrete-time sequence obtained when sampling $x(t) = \text{sinc}^2(t)$; sampling frequency $f_s = 3$ Hz (no aliasing), $f_1 = 0.8$ Hz. The three parts of this figure illustrate the full frequency range (top), the range $[-\frac{f_s}{2}, \frac{f_s}{2})$ (middle) and the range $[0, f_s)$ (bottom).

illustrates property (11.9), identical to the continuous-time Fourier transform property (5.10). At frequency $f = f_1 > 0$ the DTFT equals $X_{\Delta t}(f_1)$, marked by the black dot. At frequency $f = -f_1$ the DTFT equals $X_{\Delta t}(-f_1) = (X_{\Delta t}(f_1))^* = X^*_{\Delta t}(f_1)$, marked by the black star. We highlight the DTFT values for the 'positive' frequencies ($f = f_1 > 0$) using blue continuous lines, and those for the 'negative' frequencies ($f = -f_1$) using green dashed lines.

Considering $X_{\Delta t}(f)$ within frequency range $[0, f_s)$ illustrates property (11.11). The DTFT coefficient at a frequency $f = f_s - f_1$ (black star), which lies beyond the Nyquist frequency $\frac{f_s}{2}$ equals the complex conjugate of the DTFT coefficient at frequency $f = f_1$ (black dot). Within this interval, at frequencies beyond the Nyquist frequency, the DTFT evaluates the 'negative' frequencies of the *copy* of $X(f)$ positioned at $1f_s$.

11.5 Discrete-time complex exponential

Properties (11.9)–(11.11) are studied again in this section, to show that these are *all* the consequence of the behavior of the DTFT building block $e^{-j2\pi f \Delta t n}$ appearing in (11.2).

In continuous time, the Fourier transform uses the continuous-time complex exponential building block $c(t, f) = e^{-j2\pi f t}$, a vector with magnitude 1 and phase $-2\pi f t$ [rad], (5.1). Fixing the frequency to $f = f_1$, $f_1 \in \mathbb{R}$, results in a harmonic, complex-valued function of *time*, $c(t, f_1)$: the unit vector rotates in the complex plane, describing a unit circle centered at zero, see Figure 2.9. This function $c(t, f)$ has three properties:

1. It is periodic in time for *any* frequency f_1,
2. For any two different frequencies f_1 and f_2, we get two *different* functions, and
3. When frequency f_1 increases, the vector rotates *faster* in the complex plane.

In discrete time, the DTFT uses the discrete-time complex exponential building block $c_n(f) = e^{-j2\pi f \Delta t n}$, (11.5), where continuous time $t \in \mathbb{R}$ is replaced with discrete time $n\Delta t, n \in \mathbb{Z}$. For any frequency $f = f_1$, the time discretization causes $c_n(f_1)$ to *step* along the unit circle as a function of index n, with step size $2\pi f_1 \Delta t = 2\pi \frac{f_1}{f_s}$. This 'stepping' along the unit circle has major implications: none of the three properties stated above for $c(t, f_1)$ are true for $c_n(f_1)$.

First, $c_n(f_1)$ is defined to be periodic in time index n with period N ($N > 0$), when:

$$c_{n+N}(f_1) = c_n(f_1) \tag{11.12}$$

The smallest value of $N \in \mathbb{N}^+$ for which (11.12) holds is the fundamental period in terms of index. Hence, for $c_n(f_1)$ to be periodic, $c_{n+N}(f_1) = c_n(f_1)$ must hold, which is only true when:

$$\frac{f_1}{f_s} = \frac{\ell}{N}, \quad \forall\, \ell \in \mathbb{Z} \tag{11.13}$$

Proof $e^{-j2\pi f_1 \Delta t(n+N)} = e^{-j2\pi f_1 \Delta t n} \Rightarrow e^{-j2\pi f_1 \Delta t n} e^{-j2\pi f_1 \Delta t N} = e^{-j2\pi f_1 \Delta t n}$
$\Rightarrow e^{-j2\pi f_1 \Delta t N} = 1 \quad \Rightarrow 2\pi f_1 \Delta t N = \ell 2\pi \quad \Rightarrow f_1 \Delta t = \frac{\ell}{N} \quad \Rightarrow \frac{f_1}{f_s} = \frac{\ell}{N}$

In discrete time, when frequency f_1 is not a rational number multiplied by the sampling frequency f_s, $f_1 \neq \frac{\ell}{N} f_s$, the building block $c_n(f_1)$ is *aperiodic*. In this case, when stepping along the unit circle in the complex plane, $c_n(f_1)$ *never* assumes the same position.

EX 11.1

Considering Figure 11.2, describe the building block $c_n(f)$ using (11.13), for three frequencies: $f_a = 0.8$, $f_b = \sqrt{2}$ and $f_c = 1.12$ Hz. The sampling frequency $f_s = 8$ Hz.

Solution At frequency f_a, $c_n(f_a)$ is periodic: $\frac{0.8}{8} = \frac{1}{10}$, $\ell = 1$ and $N = 10$. At this frequency, with this sampling frequency, $c_n(f_a)$ repeats itself 1 time every 10 samples.

At frequency f_b, $c_n(f_b)$ is *a*periodic, because $\frac{\sqrt{2}}{8}$ is an irrational number.

At frequency f_c, $c_n(f_c)$ is periodic: $\frac{1.12}{8} = \frac{112}{800} = \frac{7}{50}$, $\ell = 7$ and $N = 50$. At this frequency, with this sampling frequency, $c_n(f_c)$ repeats itself 7 times every 50 samples. The sketch below (the sine part of $c_n(f_c)$) shows that after 50 samples $c_n(f_c)$ will attain exactly the same position in the complex plane, after completing 7 full cycles.

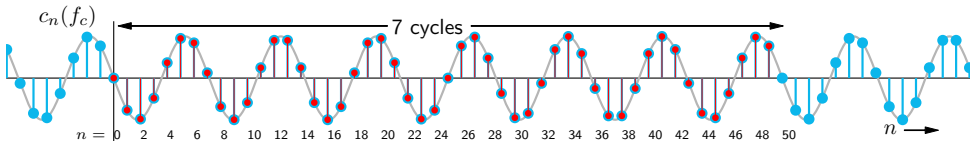


Figure 11.3: Sine part of DTFT building block $c_n(f_c)$, with f_s = 8 Hz and f_c = 1.12 Hz.

Second, the rotation in the complex plane described by $c_n(f_1)$, as a function of time index n, is *identical* for frequencies that are separated by an integer multiple of the sampling frequency:

$$c_n(f_1 + mf_s) = c_n(f_1), \quad \forall\, m \in \mathbb{Z} \tag{11.14}$$

Proof $e^{-j2\pi(f_1+mf_s)\Delta t n} = e^{-j2\pi f_1 \Delta t n} e^{-j2\pi m \overbrace{f_s \Delta t}^{=1} n} = e^{-j2\pi f_1 \Delta t n} 1 = e^{-j2\pi f_1 \Delta t n}$

We used this property several times, e.g., in proving (11.10).

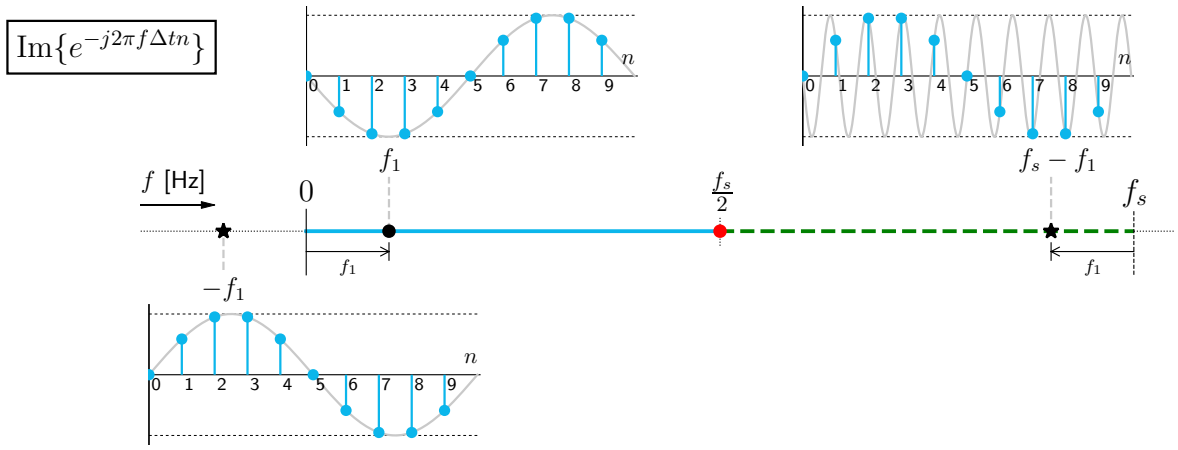


Figure 11.4: The imaginary part of the DTFT building block, the complex exponential function $c_n(f)$, is shown for three frequencies: $f = f_1$ ($\ell = 1$ in (11.13)), $-f_1$ ($\ell = -1$) and $f_s - f_1$ ($\ell = 9$), with $f_1 = 0.8$ Hz and $f_s = 8$ Hz, as a function of discrete time index $n = 0, \ldots, N-1$ ($c_n(f)$ is periodic, with $N = 10$, see Example 11.1). The gray continuous curves show the *continuous*-time building block $c(t, f)$ (sine part) for the same frequencies. The central horizontal line shows a part of the (infinite, continuous) frequency axis f, emphasizing $f \in [0, f_s)$, using the same line and color coding as in Figure 11.2; the red marker shows the Nyquist frequency.

Third, the *apparent* rate of rotation in the complex plane described by $c_n(f_1)$ will increase when f_1 progresses from 0 to $\frac{f_s}{2}$, and will then *decrease* when f_1 progresses from $\frac{f_s}{2}$ to f_s:

$$c_n(f_s - f_1) = c_n^*(f_1) \tag{11.15}$$

Proof $e^{-j2\pi(f_s - f_1)\Delta tn} = e^{-j2\pi \overbrace{f_s\Delta t}^{=1} n} e^{j2\pi f_1 \Delta tn} = 1e^{j2\pi f_1\Delta tn} = \left(e^{-j2\pi f_1\Delta tn}\right)^*$

The latter two properties are illustrated in Figure 11.4 for frequency $f_1 = 0.8$ Hz and sampling frequency $f_s = 8$ Hz. Function $c_n(-f_1)$ (bottom left) is the complex conjugate of $c_n(f_1)$ (top left). Function $c_n(f_s - f_1)$ (top right) equals $c_n(-f_1)$, (11.14), and is the complex conjugate of $c_n(f_1)$, (11.15). Although $c_n(f_s - f_1)$ rotates faster in the clockwise direction than $c_n(f_1)$ (the gray continuous curves), its *apparent* rotation (the blue stems) is a *slower* rotation in the *counter*clockwise direction, *identical* to $c_n(-f_1)$. Hence, $c_n(f_s - f_1) = c_n(-f_1) = c_n^*(f_1)$, (11.15). The DTFT building block at any frequency $f_s - f_1$ is the complex conjugate of the DTFT building block at f_1 Hz: the highest rate of apparent rotation occurs at $f = \frac{f_s}{2}$.

DTFT properties (11.10) and (11.11) are the consequence of the properties of the DTFT building block $c_n(f) = e^{-j2\pi f\Delta tn}$, (11.14) and (11.15). This function $c_n(f)$ cannot distinguish between frequencies f_s Hz apart: at these frequencies it will have *identical* values. When evaluating the sequence x_n in (11.5) at these frequencies, the DTFT performs *the same calculation*, so that $X_{\Delta t}(f)$ will have *identical* values, explaining the DTFT periodicity, (11.10).

In short, the consequences of sampling in time on the Fourier transform can be explained by considering the properties of the DTFT building block $c_n(f)$. In time, we have no information in between the samples. In frequency, the building block can only uniquely describe harmonic oscillations within a frequency band with a width of f_s Hz.

In this and the previous two chapters, signals were assumed to last forever. Clearly, this is not very practical. In the next chapter we discuss the effects of limited observation time and introduce the Discrete Fourier Transform (DFT). The DFT is our main practical signal analysis tool and is used in all subsequent chapters on spectral analysis and estimation.

12

Discrete Fourier Transform

In Chapter 5 we considered aperiodic continuous-time signals $x(t)$, and transformed them to the frequency domain $X(f)$ by means of the Fourier transform. In Chapter 9 we studied the Fourier transform $X_s(f)$ of a *sampled* signal $x_s(t)$ (though still employing a continuous-time model of the signal). In Chapter 11 we presented the Discrete-Time Fourier Transform (DTFT) $X_{\Delta t}(f)$ of a *discrete-time sequence* x_n, defined in (9.1). We are two steps away from applying the Fourier transform in practice, using the Discrete Fourier Transform (DFT).

12.1 Finite length - discrete frequency

First, continuous-time signals $x(t)$ and their discrete-time sample sequences x_n last forever. This is not practical. The phenomenon of interest may only exist or be available for a limited amount of time, and we can definitely observe or measure it only for a finite amount of time. Hence, in the first step, in the time domain, we apply a *finite-length* measurement *window*, with duration T or T_{meas}, see (8.1). In this chapter we work with samples in discrete time, and the resulting windowed or time-truncated discrete-time sequence is indicated by $x_{n,T}$. It follows from the infinite-length sequence x_n as:

$$x_{n,T} = \begin{cases} x_n & \text{for } n = 0, \dots, N-1 \\ 0 & \text{otherwise} \end{cases} \tag{12.1}$$

This operation can be interpreted as an element-wise multiplication of sequence x_n by a dimensionless window sequence w_n (with $w_n = 1$ for $n = 0, \dots, N-1$, and 0 otherwise): $x_{n,T} = w_n x_n$. The samples $x_{n,T}$ for $n \notin \{0, \dots, N-1\}$, being outside the window, equal zero, and the result $x_{n,T}$ effectively is a length N discrete-time sequence (of finite length), corresponding to a duration of T seconds. The samples are spaced by Δt seconds, such that $T = N\Delta t$ (in line with the sample-and-hold convention discussed in Chapter 10). In Appendix I it is discussed that it does not matter whether we sample first and then window, or the other way around.

Second, the Discrete-Time Fourier Transform (DTFT) $X_{\Delta t}(f)$ (11.5) of sequence x_n is a continuous function of frequency f, with $f \in \mathbb{R}$. This does not lend itself well to digital signal processing on the computer. Therefore, in the second step we *choose* to evaluate the DTFT $X_{\Delta t}(f)$ (in fact, $X_{\Delta t,T}(f)$, as we apply the time window) only at a *selection* of *discrete frequencies*, with a frequency step size of Δf Hz. Recall that through the action of sampling, the DTFT $X_{\Delta t}(f)$ becomes *periodic* with period f_s, see (11.3) and (11.10). Therefore, the selection of discrete frequencies has to cover only the interval $[0, f_s)$, as beyond this interval $X_{\Delta t}(f)$ simply repeats itself, see Figure 11.2. Although we may, equivalently, select different intervals,

such as $[-\frac{f_s}{2}, \frac{f_s}{2})$, the interval $[0, f_s)$ is the default in computer algorithms. When Fourier-transforming data sequence $x_{n,T}$, effectively a sequence of length N with $n = 0, \ldots, N-1$, we prefer to neither lose information nor introduce redundant information. Hence, the selection of frequencies will contain N discrete frequencies, exactly as many as we have samples. With an equidistant spacing on interval $[0, f_s)$ this leads to frequency step:

$$\boxed{\Delta f = \frac{f_s}{N} = \frac{1}{N\Delta t} = \frac{1}{T}} \tag{12.2}$$

in the frequency domain. Within interval $[0, f_s)$ the selection of N discrete frequencies then becomes $0, \Delta f, 2\Delta f, \ldots, (N-1)\Delta f$; these are referred to as the 'analysis frequencies' of the Discrete Fourier Transform (DFT). The continuous-frequency function $X_{\Delta t,T}(f)$ is evaluated at $f = k\Delta f$ with $k = 0, \ldots, N-1$, and consequently denoted as $X_k = X_{\Delta t,T}(k\Delta f)$.

Starting from the DTFT $X_{\Delta t}(f)$ of sampled sequence x_n, implementing these two steps yields the Discrete Fourier Transform. The DFT transforms a finite-length discrete-time sequence $x_{n,T}$ into a finite-length discrete-frequency sequence X_k, of same length by default.

EX 12.1

We continue the example of Figure 9.3 with even signal $x(t)$, sampled in Figure 9.5 with sampling frequency $f_s = 3$ Hz. Apply a time window from $t = -2$ to $t = 2$ s ($T = 4$ s) to obtain $x_{n,T}$ and evaluate its DTFT $X_{\Delta t,T}(f)$ at the $N = f_s T = 12$ analysis frequencies.

Solution Figure 12.1 at left shows $x_{n,T}$ with the blue stems. According to the sample-and-hold convention in Chapter 10 on signal reconstruction, there are $N = 12$ samples at an interval of $\Delta t = \frac{1}{3}$ s (each sample represents a duration of Δt s), such that $T = N\Delta t$, and the sample at $t = 2$ s is *not* included.

Figure 12.1 at right shows the (modulus of the) Fourier transform $X_{\Delta t,T}(f)$ evaluated at the $N = 12$ frequencies, at an interval of $\Delta f = \frac{1}{T} = \frac{1}{4}$ Hz, such that $f_s = N\Delta f$, and the frequency $f = f_s = 3$ Hz is *not* included.

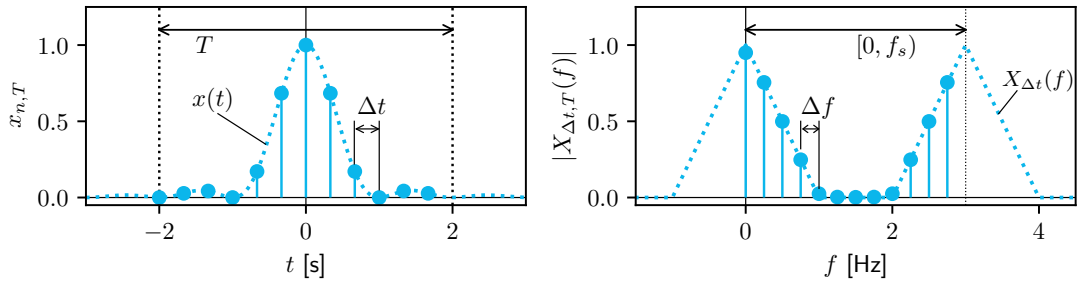


Figure 12.1: Left: sampled and time-windowed sequence $x_{n,T}$, with $\Delta t = \frac{1}{3}$ s, $N = 12$ and window length $T = N\Delta t = 4$ s, shown by the stems; $x(t)$ is shown by the dotted curve as a function of time t, as in Figure 9.3. Right: the modulus of Fourier transform $X_{\Delta t,T}(f)$, evaluated at discrete frequencies (the stems) with $\Delta f = \frac{1}{T} = \frac{1}{4}$ Hz; the DTFT $X_{\Delta t}(f) = X_s(f)$ is shown by the dotted curve, as in Figure 9.6, but here for a limited frequency range.

Apart from the zero frequency, the lowest frequency for which we evaluate $X_{\Delta t,T}(f)$ is $f = \Delta f$, the frequency of a harmonic which completes one full 'revolution' or cycle within the observation window of T seconds. We can imagine that trying to evaluate the spectrum at (much) lower frequencies is challenging at best, as our window contains only a (very) small part of the full cycle of the corresponding harmonic, and the rest of its cycle is unknown to us. The graph at right also shows the DTFT $X_{\Delta t}(f) = X_s(f)$,

dotted, and the stems correspond well; spectral leakage due to the finite time window length is very limited here; see Figure 8.3 for a shorter window of $T = 3$ s.

12.2 Definition of the DFT and inverse DFT

12.2.1 DFT derivation and definition

With the goal of obtaining the Fourier transform of a finite-length discrete-time sequence $x_{n,T}$, we begin with the Discrete-Time Fourier Transform (DTFT), (11.5), of the sampled sequence x_n. The time window of duration T turns the infinite sum in (11.5) into a finite sum from $n = 0$ to $n = N - 1$. And *in* this window, $x_{n,T}$ is identical to x_n, (12.1):

$$X_{\Delta t,T}(f) = \text{DTFT}\{x_{n,T}\} = \Delta t \sum_{n=-\infty}^{\infty} x_{n,T} e^{-j2\pi f \Delta t n} = \Delta t \sum_{n=0}^{N-1} x_n e^{-j2\pi f \Delta t n} \tag{12.3}$$

Next, the resulting Fourier transform $X_{\Delta t,T}(f)$ is evaluated at N discrete frequencies, which cover the range $[0, f_s)$: $f = k\Delta f$ with $\Delta f = \frac{f_s}{N} = \frac{1}{N\Delta t}$, and $k = 0, \ldots, N-1$:

$$X_{\Delta t,T}(k\Delta f) = \Delta t \sum_{n=0}^{N-1} x_n e^{-j2\pi k \Delta f \Delta t n} = \Delta t \sum_{n=0}^{N-1} x_n e^{-j2\pi k \frac{1}{N\Delta t} \Delta t n} \tag{12.4}$$

This results in the definition of the Discrete Fourier Transform:

$$X_k = \text{DFT}\{x_n\} = \Delta t \sum_{n=0}^{N-1} x_n e^{-jk\frac{2\pi}{N}n} \tag{12.5}$$

We obtain N complex-valued coefficients, X_k, with index $k = 0, \ldots, N-1$, which together represent the evaluation of $X_{\Delta t,T}(f)$ at N discrete frequencies. Their unit equals [unit seconds] or, equivalently, [unit/Hz], with [unit] the unit of sequence x_n.

In the remainder of this book, unless stated otherwise, we omit the T index with the (finite-length) discrete-time sequence: we use (again) x_n instead of $x_{n,T}$. In practice it is evident that the sequence can never be of infinite length: it is of finite length, N samples. Our main tool, the DFT, only considers the time interval for $n = 0, \ldots, N-1$, where $x_{n,T} = x_n$ (12.1).

The DFT is a computer algorithm that has as its input N (in this book real-valued) samples of $x(t)$, denoted by x_n with $n = 0, \ldots, N-1$, and has as its output N complex-valued numbers X_k with $k = 0, \ldots, N-1$; this operation is also referred to as the 'N-point DFT'.

EX 12.2

We re-formulate Example 12.1 in the context of the DFT, with DFT input sequence x_n ($n = 0, \ldots, N-1$) and DFT output sequence X_k ($k = 0, \ldots, N-1$).

Solution Figure 12.1 from Example 12.1 becomes Figure 12.2; the twelve numerical values in the two figures at left are *identical*, and so are the twelve values at right (the latter are generally complex-valued; therefore the modulus is shown).

In Figure 12.1 the time dimension has been restored in the graph at left, and the frequency dimension in the graph at right. A computer does not know time and frequency; it merely has the indices n and k, representing the nth sample and the kth DFT frequency, respectively, as illustrated in Figure 12.2.

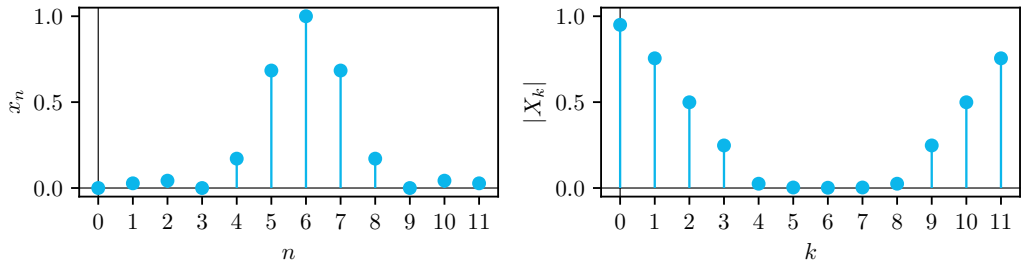


Figure 12.2: The sampled and time-windowed signal is a discrete-time sequence x_n with $n = 0, \ldots, N-1$, consisting of N data points (at left), and the DFT turns these into N numerical evaluations of the Fourier transform $X_{\Delta t,T}(f)$ at $f = k\Delta f$ with $k = 0, \ldots, N-1$, denoted by $|X_k|$ (at right), for the example in Figure 12.1, with $N = 12$.

Figure 12.2 shows the DFT input sequence x_n and output sequence X_k. Since the DFT by default considers x_n with $n = 0, \ldots, N-1$, according to (9.1), corresponding to time window $t \in [0, T]$, we apply a time shift of $t_0 = -\frac{T}{2}$ (see (6.2)) to the samples shown in Figure 12.1 at left. This does not affect the magnitude of $X_{\Delta t,T}(f)$, shown at right, but it *would* impact its phase, (6.2). Therefore, $|X_k|$ is shown in Figure 12.2.

12.2.2 Inverse DFT

The inverse DFT is given by:

$$x_n = \text{iDFT}\{X_k\} = \frac{1}{\Delta t}\frac{1}{N}\sum_{k=0}^{N-1} X_k e^{jk\frac{2\pi}{N}n} \tag{12.6}$$

for $n = 0, \ldots, N-1$. The proof, as well as some common DFT pairs, is included in Appendix H. Similar to the complex exponential Fourier series, the Fourier transform and the DTFT, (12.5) is known as the analysis equation and (12.6) as the synthesis equation of the DFT.

With (12.5) and (12.6) we can go from x_n to X_k and back, without losing any information. In practice, we typically apply the DFT to a finite-length discrete-time signal, turning the $N = 12$ blue dots of Figure 12.2 at left into the $N = 12$ blue dots in the graph at right.

Starting from an infinite-length continuous-time signal, Appendix I summarizes the full procedure of first applying a window in the time domain, then sampling the signal, and eventually evaluating the Fourier transform in the frequency domain at discrete frequencies.

12.2.3 Alternative definition

Note that in many textbooks and programming environments, such as Python, the DFT (12.5) and its inverse (12.6) are defined *without* the factor Δt and $\frac{1}{\Delta t}$, respectively:

$$X_k = \sum_{n=0}^{N-1} x_n e^{-jk\frac{2\pi}{N}n} \tag{12.7}$$

and:

$$x_n = \frac{1}{N}\sum_{k=0}^{N-1} X_k e^{jk\frac{2\pi}{N}n} \tag{12.8}$$

meaning that it is up to the user to restore the time and frequency dimension.

In this book, by default, we refer to the DFT (12.5) and its inverse (12.6), respectively, with Δt and $\frac{1}{\Delta t}$ *included*, as we did in the previous chapter on the DTFT, see (11.5). In this way, the *time* interval or step size Δt and *frequency* step size $\Delta f = \frac{1}{N\Delta t}$ become immediately apparent, and the unit ([unit/Hz]) of a Fourier transform is maintained, see Section 5.2.

12.2.4 Fast Fourier transform (FFT)

An extremely efficient algorithmic implementation of the DFT is the Fast Fourier Transform (FFT) [10, 11], available in Python through the `numpy.fft.fft` for (12.7) and `numpy.fft.ifft` for (12.8). The FFT was published by Cooley and Tukey in 1965 [10][1] and, especially in times of limited computer power, revolutionized spectral analysis, as it substantially reduces the number of operations needed to obtain the DFT.

When the number of samples N is a power of 2 ($N = 2^m$) the DFT needs 2^{2m} operations, while the FFT needs 'only' $m2^m$ operations. While computing the *same* N coefficients X_k from N samples, the FFT algorithm is $\frac{2^m}{m}$ times faster. For instance, for $N = 4{,}096$ samples $m = 12$, the DFT requires $16{,}777{,}216$ operations, while the FFT requires $49{,}152$ operations, which makes it 341 times faster than a plain DFT algorithm for this number of samples.

The FFT implementation is not discussed further in this book.

12.3 DFT frequencies, properties, and diagram

The properties of the DFT are a logical consequence of its close relation with the DTFT of a time-windowed discrete-time sequence, as outlined in the previous section.

12.3.1 DFT frequencies

The complex exponential building block $e^{-jk\frac{2\pi}{N}n}$ of the DFT is always *periodic* in time, as the analysis frequencies are all a rational number multiplied by the sampling frequency f_s:

$$f = k\Delta f = k\frac{f_s}{N} = \frac{k}{N}f_s \tag{12.9}$$

see (11.13). The building block assumes that the period of the kth analysis frequency harmonic fits exactly k times in the (duration of the) sequence of N samples. This means that the DFT by its coefficients X_k always yields a *periodic* 'model' of the finite data sequence x_n, as if $x_{n+N} = x_n$ (see Figure I.7) which is usually *not* the case with the original signal.

The DFT building block does not know about time and frequency. Recall the DTFT building block $c_n(f) = e^{-j2\pi f\Delta tn}$, introduced in Section 11.5. Since in the DFT frequency f is restricted to $f = k\frac{f_s}{N} = k\frac{1}{N\Delta t}$, the Δt disappears altogether, (12.4): the DFT building block becomes $e^{-jk\frac{2\pi}{N}n}$. When time index n progresses from 0 to $N-1$, this complex exponential steps along the unit circle, in steps that are an integer multiple k of 'fundamental frequency' $\frac{2\pi}{N}$; the latter can be considered to divide the unit circle into N equal pieces. For $k = 1$, the first harmonic, it takes the building block N steps to reach the same position again in the complex plane. This means that for the DFT building block, *only* the number of samples N matters: it does not matter whether the DFT (12.5) is evaluating a sequence of 100 numbers that results from sampling a signal for 1 s at 100 Hz, or from sampling a signal for 10 s at 10 Hz.

Figure 12.3 illustrates the imaginary part (the sine component) of the building block for a sequence of $N = 8$ samples. The same properties occur as were shown in Figure 11.4. All DFT frequencies are shown: all are an integer multiple k of the fundamental frequency $\frac{2\pi}{8}$.

[1] It was later discovered that Cooley and Tukey together had independently re-invented an algorithm already known to Carl Friedrich Gauss around 1805 [4].

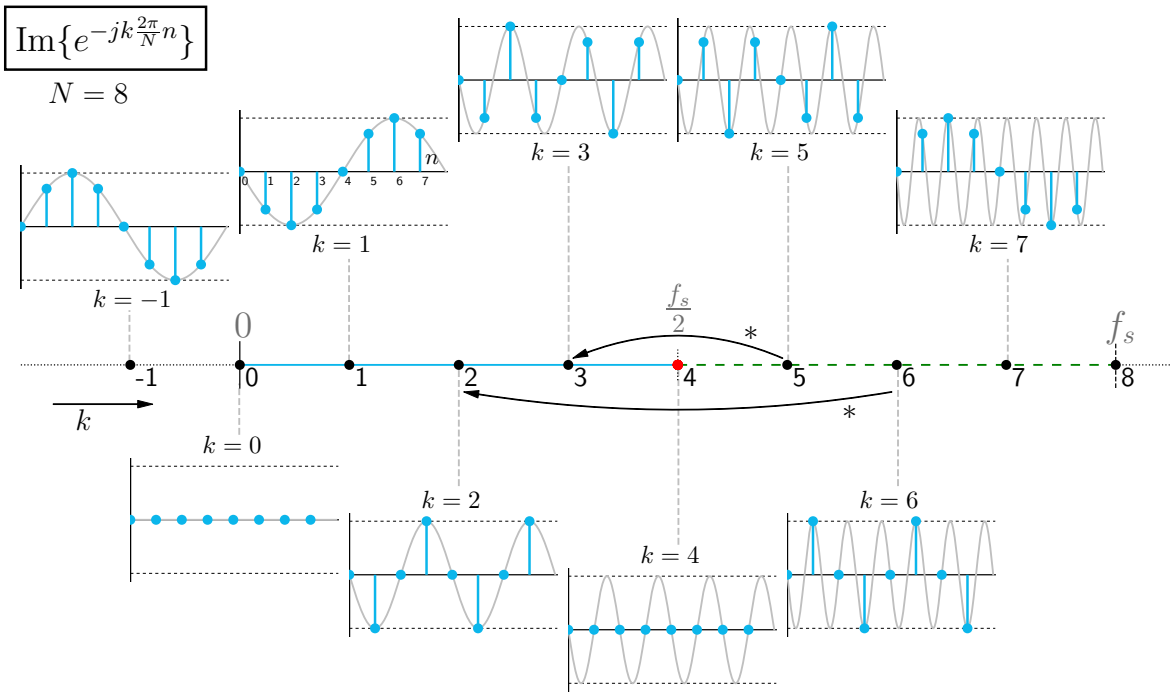


Figure 12.3: Imaginary part of DFT building block $e^{-jk\frac{2\pi}{N}n}$, for $N = 8$, with frequency indices $k = \{-1, 0, 1, 2, \ldots, 7\}$, when discrete-time index n progresses from $n = 0$ to $n = N - 1 = 7$. Similar to Figure 11.4, the gray continuous curves show the continuous-time sine for the same frequencies. Note the resemblance with Figure 11.4 and the resulting symmetry properties, for instance, $X_5 = X_3^*$, and $X_6 = X_2^*$.

Multiplying, for each frequency index k, the sequence x_n by the building block, then adding, times Δt, yields the DFT coefficient X_k. The real part of coefficient X_k equals the 'integrated' product of sequence x_n and the real part of the kth building block, the cosine component. The imaginary part of coefficient X_k equals the 'integrated' product of sequence x_n and the imaginary part of the kth building block, the sine component. In doing so, the DFT evaluates or measures the amount of similarity between input sequence x_n and the kth building block, in a very similar way as the complex exponential Fourier series, discussed in Chapter 4.

12.3.2 DFT properties

The DFT has much in common with the complex exponential Fourier series, see Chapter 4. The DFT fundamental frequency equals $\frac{f_s}{N} = \frac{1}{T}$ Hz, with the 'fundamental period' T seconds. The DFT results in N complex-valued coefficients X_k at integer multiples $k = 0, 1, \ldots N-1$ of this fundamental frequency. The coefficients X_k can be expressed in polar form, see (4.6), with a magnitude $|X_k|$, (4.7), and phase $\angle X_k$, (4.8). When shown as a function of discrete frequency $k\Delta f$, $|X_k|$ and $\angle X_k$ result in, respectively, the amplitude and phase spectra.

For real-valued signals, and because of the equivalence of the DTFT and DFT at the analysis frequencies, DTFT properties (11.9)–(11.11) become for the DFT, respectively:

$$\boxed{X_{-k} = X_k^*} \tag{12.10}$$

$$\boxed{X_{k+mN} = X_k}, \qquad \forall\, m \in \mathbb{Z} \tag{12.11}$$

$$\boxed{X_{N-k} = X_k^*} \tag{12.12}$$

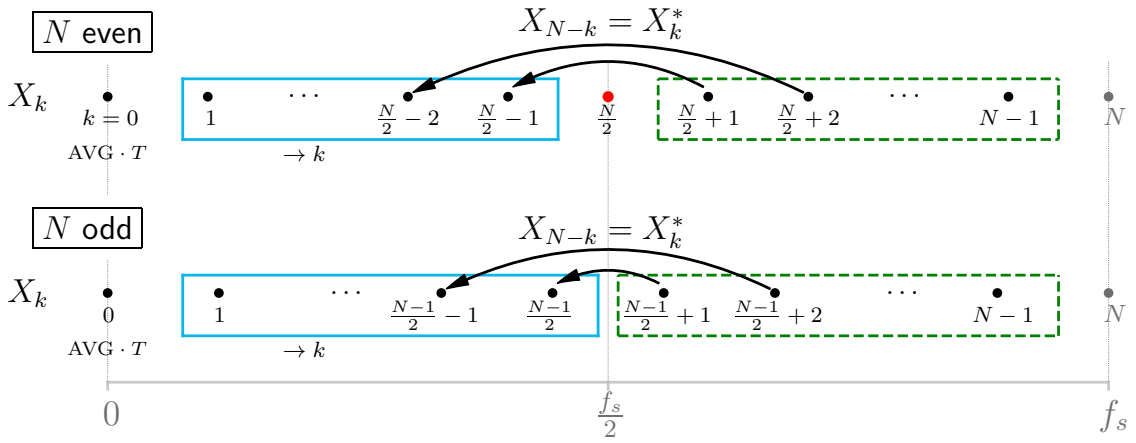


Figure 12.4: DFT coefficients X_k and their relationships for an even (top) and odd (bottom) number of samples N; k represents the discrete frequency index. The blue box shows the positive frequencies, the green dashed box the negative frequencies, similar to Figure 11.2. The Nyquist frequency is shown by the red dot, for even N. The gray horizontal line shows the continuous frequency interval $f \in [0, f_s)$ corresponding to the discrete frequencies.

12.3.3 DFT diagram of coefficients

Figure 12.4 shows the diagram for the DFT coefficients X_k. At $k = 0$, the DFT coefficient X_0 with (12.5) yields the sum of all samples multiplied by Δt; $\frac{X_0}{T}$ equals the average (AVG) of the sequence. When the number of samples N is even, the Nyquist frequency is positioned at $k = \frac{N}{2}$, and the DFT yields an evaluation of $X_{\Delta t,T}(f)$ at $\frac{N}{2} - 1$ positive frequencies. When N is odd, the analysis frequencies do *not* include the Nyquist frequency, and the DFT yields an evaluation of $X_{\Delta t,T}(f)$ at $\frac{N-1}{2}$ positive frequencies. The number of negative frequencies always equals the number of positive frequencies, and at these negative frequencies the DFT coefficients are the complex conjugate of the coefficients at positive frequencies, see (12.12).

When investigating a real-valued signal, the negative frequencies are often omitted, and when N is even also the Nyquist frequency: the frequencies from $k = 0$ to $k = \frac{N}{2} - 1$ (for even N) contain all information. The DFT turns N real-valued numbers x_n into N complex-valued numbers X_k, the latter all consisting of a real and an imaginary part, but there is no additional information created or contained, due to (12.12).

12.4 DFT example ($N = 8$)

In this section it is demonstrated how the DFT of a discrete-time sequence x_n, with length $N = 8$ samples, is constructed. The DFT fundamental frequency $\frac{2\pi}{N}$ becomes $\frac{2\pi}{8}$, and the DFT building block $e^{-jk\frac{2\pi}{8}n}$ can assume 8 positions on the unit circle, as a function of frequency index k and time index n, see Figure 12.5. We compute and discuss some of the DFT coefficients, (12.5). Note that computing them in Python relies on (12.7) (without factor Δt).

$$X_0 = \Delta t \sum_{n=0}^{7} x_n e^{-j0\frac{2\pi}{8}n} = \Delta t(x_0 + x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7) \quad \text{zero frequency}$$

$$X_1 = \Delta t \sum_{n=0}^{7} x_n e^{-j1\frac{2\pi}{8}n} = \ldots \quad \text{one cycle in } N = 8$$

$$X_2 = \Delta t \sum_{n=0}^{7} x_n e^{-j2\frac{2\pi}{8}n} = \Delta t(x_0 - jx_1 - x_2 + jx_3 + x_4 - jx_5 - x_6 + jx_7) \quad \text{two cycles in } N = 8$$

$$X_3 = \Delta t \sum_{n=0}^{7} x_n e^{-j3\frac{2\pi}{8}n} = \ldots \quad \text{three cycles}$$

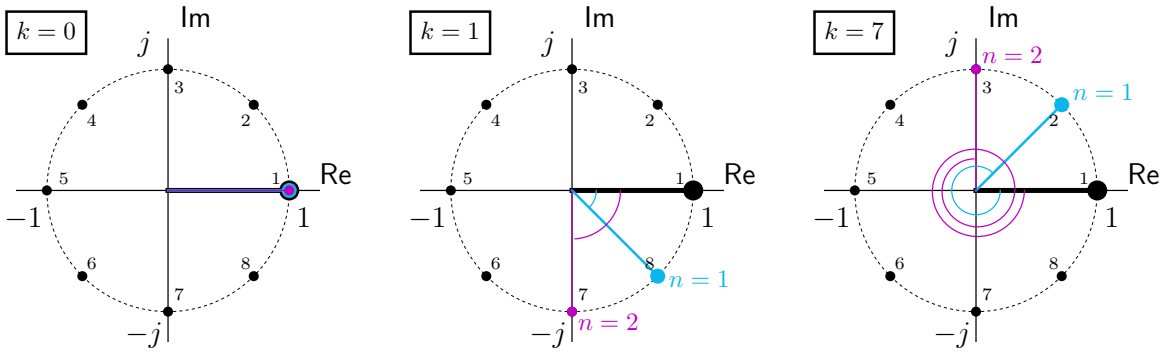


Figure 12.5: Positions of the DFT building block $e^{-jk\frac{2\pi}{N}n}$ in the complex plane, for $N = 8$, indicated by the small black dots, representing integer multiples of $\frac{2\pi}{8}$. Three values of frequency index k are shown, from left to right $k = 0$, $k = 1$ and $k = 7$. For each frequency, the black, blue and purple dots and lines show the positions of the DFT building block, for $n = 0$, $n = 1$ and $n = 2$, respectively.

$$
\begin{aligned}
X_4 = \Delta t \sum_{n=0}^{7} x_n e^{-j4\frac{2\pi}{8}n} &= \Delta t(x_0 - x_1 + x_2 - x_3 + x_4 - x_5 + x_6 - x_7) && \text{four cycles} \\
&= \Delta t((x_0 + x_2 + x_4 + x_6) - (x_1 + x_3 + x_5 + x_7)) && \text{Nyquist frequency} \\
X_5 = \Delta t \sum_{n=0}^{7} x_n e^{-j5\frac{2\pi}{8}n} &= \ldots && \text{five cycles} \\
X_6 = \Delta t \sum_{n=0}^{7} x_n e^{-j6\frac{2\pi}{8}n} &= \Delta t(x_0 + jx_1 - x_2 - jx_3 + x_4 + jx_5 - x_6 - jx_7) && \text{six cycles} \\
X_7 = \Delta t \sum_{n=0}^{7} x_n e^{-j7\frac{2\pi}{8}n} &= \ldots && \text{seven cycles}
\end{aligned}
\tag{12.13}
$$

The zero-frequency coefficient X_0 is a *real* number equal to the sum of all samples multiplied by Δt. Figure 12.5 at left illustrates that for $k = 0$ the complex exponential maintains its position in the complex plane, at 1, when time index n progresses. Hence, the sequence x_n is 'integrated', as in the continuous-time Fourier transform for the zero frequency, (5.8). Dividing X_0 by T yields the average of the sequence: $\frac{1}{T}X_0 = \frac{1}{N\Delta t}\Delta t \sum_{n=0}^{N-1} x_n = \frac{1}{N}\sum_{n=0}^{N-1} x_n$.

At $k = \frac{N}{2}$ (for even N) we obtain the coefficient at the Nyquist frequency. It is a *real* number equal to the sum of the even-indexed samples minus the sum of the odd-indexed samples, multiplied by Δt. This number is of no practical use and typically discarded. When sampled properly, i.e., fast enough, (9.13), there should be 'nothing of value' at this frequency.

The coefficients for $k = 1$, 2 and 3 are complex numbers representing the 'positive' frequencies; they correspond to the product of the sequence x_n with the DFT building block that makes, respectively, 1, 2 and 3 cycles through the complex plane, rotating in clockwise direction, when time index n progresses from 0 to 7 ($= N - 1$). Figure 12.5, middle, illustrates the first two steps of the complex exponential for $k = 1$, from $n = 0$ to $n = 2$.

The coefficients for $k = 5$, 6 and 7 are complex numbers representing the 'negative' frequencies; they correspond to the product of the sequence x_n with the DFT building block that makes, respectively, 5, 6 and 7 cycles through the complex plane, rotating in clockwise direction, when n progresses from 0 to 7. The coefficients obtained here are the complex conjugates of those obtained for, respectively, $k = 3$, 2 and 1, see (12.12). Figure 12.5 at right illustrates the first two steps of the complex exponential for $k = 7$, from $n = 0$ to $n = 2$.

In (12.13), verify that $X_6 = X_2^*$. This is because the complex exponential that rotates 6 times $\frac{2\pi}{8}$ in the complex plane every time step, in clockwise direction, is equal to the conjugate of the complex exponential that rotates 2 times $\frac{2\pi}{8}$ every time step, in *counter*clockwise direction. This is illustrated in Figure 12.5 for $k = 1$ and $k = 7$; as a consequence, $X_7 = X_1^*$.

12.5 Two more examples: DFT of cosine

EX 12.3

Compute the DFT of a unit-amplitude, zero-phase cosine signal $x(t)$ with frequency $f_0 = 3$ Hz, after sampling this signal with $f_s = 10$ Hz, for time duration $T = 2$ s.

Solution With $f_s = 10$ Hz and $T = 2$ s, $N = 20$ samples are obtained. The DFT frequency step Δf equals $\frac{1}{T} = \frac{1}{2}$ Hz, which means that the signal frequency f_0 equals 6 times the frequency step size Δf. The sequence x_n completes exactly 6 cycles in the time duration $T = 2$ s, and so it fits perfectly.

In Appendix H it is shown that the DFT of a unit-amplitude, zero-phase cosine signal with frequency f_0 equal to ℓ times frequency step Δf, i.e., $x_n = \cos[2\pi f_0 \Delta t n] = \cos[2\pi(\ell\frac{1}{N\Delta t})\Delta t n] = \cos[\ell\frac{2\pi}{N}n]$ equals:

$$X_k = \frac{\Delta t N}{2}\Big(\delta_{k-\ell} + \delta_{k-(N-\ell)}\Big) \tag{H.7}$$

All X_k coefficients are zero, except for $k = \ell$ and $k = N - \ell$. Here, with $N = 20$ and $\ell = 6$, we expect that the DFT of x_n results in two non-zero coefficients:

$$X_6 = \frac{\Delta t N}{2} = \frac{T}{2} = 1 \quad \text{and} \quad X_{14} = 1$$

Figure 12.6 shows the result of Python's `numpy.fft.fft` on sequence x_n, multiplied by Δt: DFT coefficients X_k, (12.5). The $N = 20$ real-valued time samples x_n are transformed to $N = 20$ complex-valued numbers X_k in discrete frequency. Since x_n is real and even, the imaginary parts of X_k are very close to 0, in the order of 10^{-16}, caused by the finite computer precision.

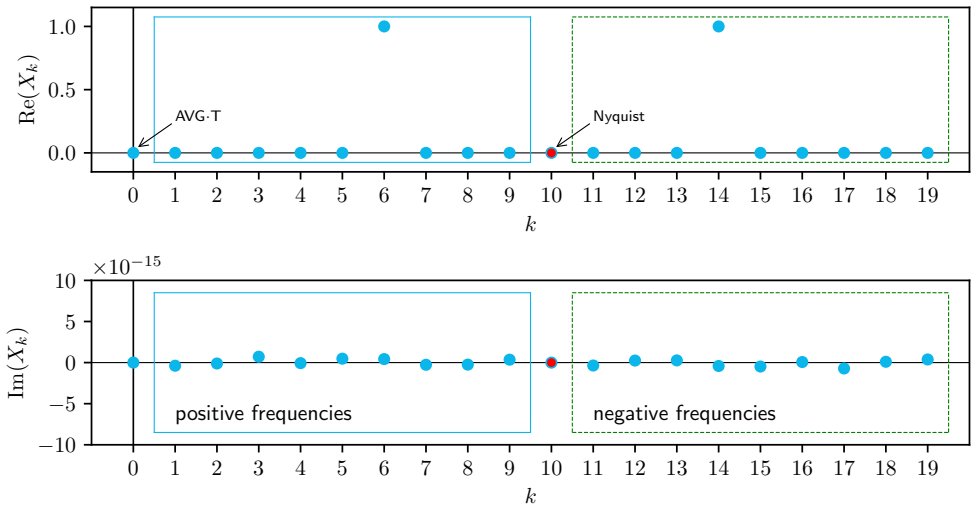


Figure 12.6: Result of DFT of unit-amplitude, zero-phase cosine with frequency $f_0 = 3$ Hz, sampled with $f_s = 10$ Hz for a duration of $T = 2$ s. All X_k coefficients are shown as blue dots, except for the coefficient at the Nyquist frequency, which is shown in red.

Considering the real part of X_k, at $k = 0$ we obtain 0, the average of the cosine sequence x_n, with $n = 0, \ldots, N-1$, a perfect fit of $\ell = 6$ cycles in duration T. Since N is even, we get a Nyquist frequency, at $k = \frac{N}{2} = 10$, at which $X_{k=10} = 0$ also in this case. All X_k's are zero, except at $k = 6$ $(= \ell)$ and $k = 14$ $(= N - \ell)$, where $X_k = 1$, as expected.

For the sake of explanation, and to facilitate a comparison with the DFT diagram of Figure 12.4, the coefficient at the Nyquist frequency is shown by means of a red dot. The coefficients at the positive frequencies, $k = 1, \ldots, 9$, are enclosed by the blue box, those at the negative frequencies, $k = 11, \ldots, 19$, by the green, dashed box.

Figure 12.7 shows the result of using Python's `numpy.fft.fftshift` function on sequence X_k, which changes the order of the coefficient sequence as to cover the interval $[-\frac{f_s}{2}, \frac{f_s}{2})$, symmetric about $f = 0$. The frequency axis is restored by multiplying the index by the frequency step size Δf: the peaks occur exactly at $f = \pm 3$ Hz. The Nyquist frequency is put at $f = -\frac{f_s}{2}$, but is typically discarded.

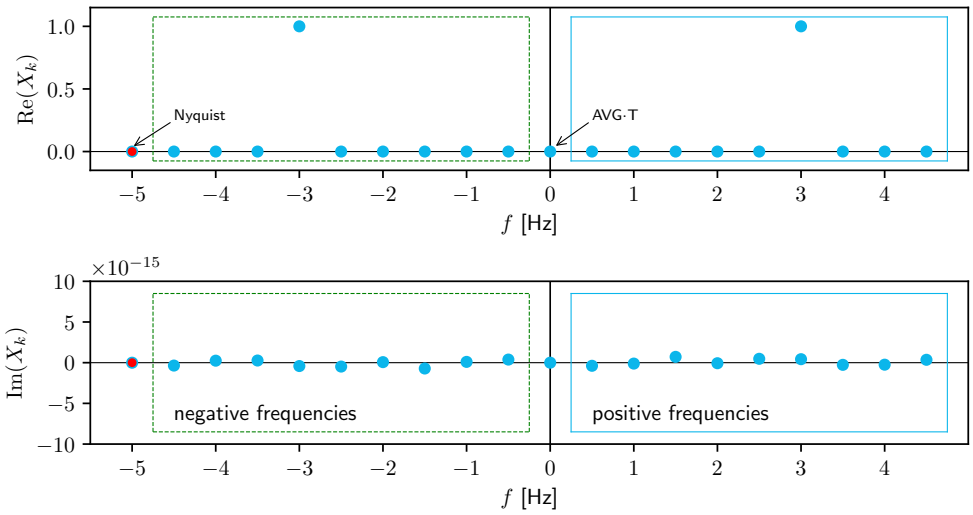


Figure 12.7: Creating a symmetric spectrum for DFT result of Figure 12.6: moving the negative frequencies to the left and restoring the frequency axis through multiplying k by frequency step size Δf.

EX 12.4 Repeat Example 12.3 for two cases. First, now use sampling frequency $f_s = 20$ Hz for the same duration $T = 2$ s, yielding the finite-duration sequence y_n. Second, use the original sampling frequency $f_s = 10$ Hz, but increase duration T to 4 s, resulting in sequence z_n. Compute Y_k and Z_k and comment on the differences with X_k from Example 12.3.

Solution Since the cosine is a real, even function, the imaginary part of X_k, Y_k and Z_k is zero (for all cases) and is discarded here. Figure 12.8 shows the three sequences x_n, y_n and z_n at left and their DFTs X_k, Y_k and Z_k at right, after performing the frequency shift operations explained in Example 12.3. The Nyquist frequency is discarded.

The middle row shows that doubling the sampling frequency f_s for constant signal duration T results in twice the number of samples N. The frequency step, however, does not change and Δf remains $\frac{1}{T} = \frac{1}{2}$ Hz. What *does* change is that we can study the

DFT Y_k of sequence y_n from $f = -\frac{f_s}{2} = -10$ Hz to $\frac{f_s}{2} = 10$ Hz, twice the range of the analysis frequencies compared to X_k (top). Increasing the sampling frequency does *not* yield smaller DFT frequency steps. Rather, it places the copies of the spectrum at higher frequencies, allowing a larger range of frequencies for evaluation.

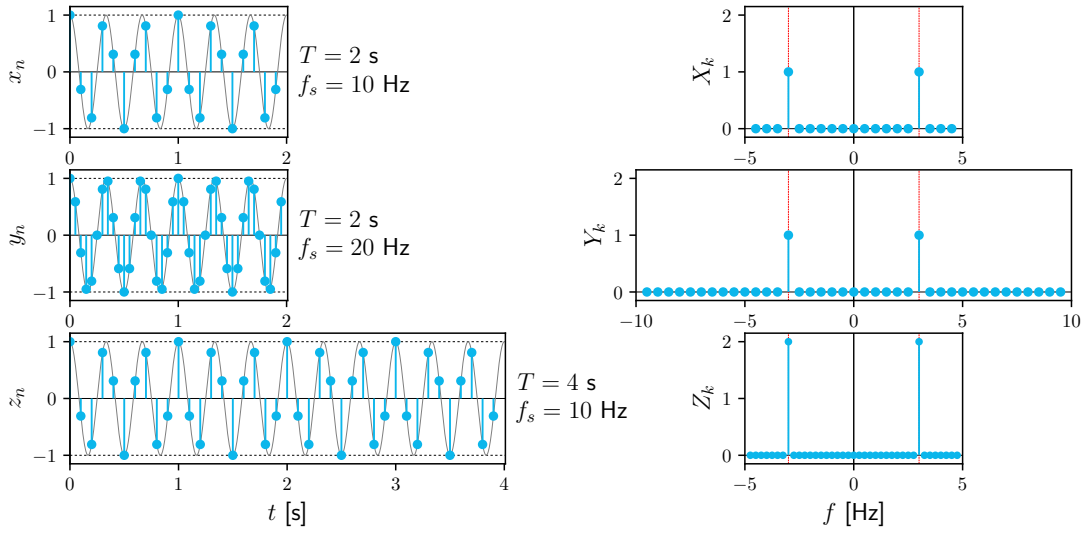


Figure 12.8: Effects on the DFT of a unit-amplitude, zero-phase cosine. Only the real parts are shown, the imaginary parts are close to zero. Top: sequence x_n and its DFT X_k from Example 12.3. Middle: effect of increasing sampling frequency f_s, sequence y_n. Bottom: effect of enlarging measurement time T, sequence z_n. The coefficient at the Nyquist frequency is discarded.

The bottom row shows that doubling the signal duration T for constant sampling frequency f_s also yields twice the number of samples. The frequency resolution *has* improved, step size Δf becomes $\frac{1}{4}$ Hz, half the step size of X_k (top row). What does *not* change, is the range of analysis frequencies, which remains $f \in (-5, 5)$ Hz. Increasing the signal duration leads to a better DFT frequency resolution, we get 'more DFT analysis frequencies per Hz', the DFT frequency step size becomes smaller.[2]

In these examples, the cosine signal completed exactly an integer number of cycles ℓ within duration T; in other words, its frequency was an integer multiple ℓ of the DFT frequency step size Δf, (12.2). This *periodic* signal in N could be perfectly modeled using the *intrinsically periodic* DFT building blocks, resulting in very clean DFTs with the 'spikes' occurring at exactly the right positions in frequency. In the next section the more general case will be discussed, involving the – in practice often inevitable – effects of spectral leakage.

12.6 Window and leakage – revisited

Close inspection of the graph in Figure 12.1 at right reveals that the dots of the stems (evaluation of $|X_{\Delta t,T}(f)|$) are close, but not all exactly centered on the dotted lines (representing $X_{\Delta t}(f)$). We can imagine that the Fourier transform of a *part* of the signal, the time-windowed signal $x_T(t)$ may differ from the Fourier transform of the full (infinite-length) signal $x(t)$.

As discussed in Chapter 8, in practice, a signal $x(t)$ is only available for a *finite* time duration, and consequently the phenomenon of *leakage* was introduced. In Chapter 8 we

[2] Therefore, when $\frac{1}{T}$ is the DFT frequency step, $T = \frac{N}{f_s}$ is the DFT frequency resolution.

discussed only continuous-time, continuous-frequency signals. In Figure 8.3 at right, we observed that the Fourier transform of the time-windowed signal looked a bit 'wobbly'.

In the present chapter, we work with time-windowed, discrete-time (sampled) signals, and evaluate the Fourier transform $X_{\Delta t,T}(f)$ at discrete frequencies. In Figure 12.1 we see that $|X_{\Delta t,T}(f)|$ of the *sampled* and *time-windowed* signal, evaluated at the analysis frequencies, indicated by the stems, is very close to $X_{\Delta t}(f) = X_s(f)$ (dotted), the Fourier transform of the sampled signal derived in Chapter 9. There are tiny differences due to *leakage*, as a consequence of the (rectangular) time window. In this example, the window is quite long compared to the main features of the signal, so the differences are small. As a general guideline we can assume that the longer the time window is, the smaller the differences, i.e., the smaller the impact of the window on the spectrum.

EX 12.5

We return to Examples 12.3 and 12.4 to demonstrate the consequences of time-windowing a signal in the context of the DFT.

Solution Recall that the Fourier transform of a unit-amplitude, zero-phase cosine function with frequency f_0, $x(t) = \cos(2\pi f_0 t)$, consists of two Dirac delta impulses, both with a weight of a half, one at frequency $f = -f_0$ and one at f_0 (5.16). The effects of time-windowing were demonstrated in Figure 8.4 and expressed by (8.3). Here we consider a unit-amplitude, zero-phase cosine at $f_0 = 4$ Hz and one at $f_0 = 4.5$ Hz, and we use a rectangular time window of $T = 1$ s. The magnitude spectrum $|X_w(f)|$ of the time-windowed signal is shown at left in Figure 12.9, for continuous-time continuous-frequency. Only a part of the (double-sided) spectrum is shown. As the window length $T = 1$ s, $W(f) = T\,\mathrm{sinc}(Tf)$ has a main peak width equal to 2 Hz and the side lobes have a width equal to 1 Hz, (5.11).

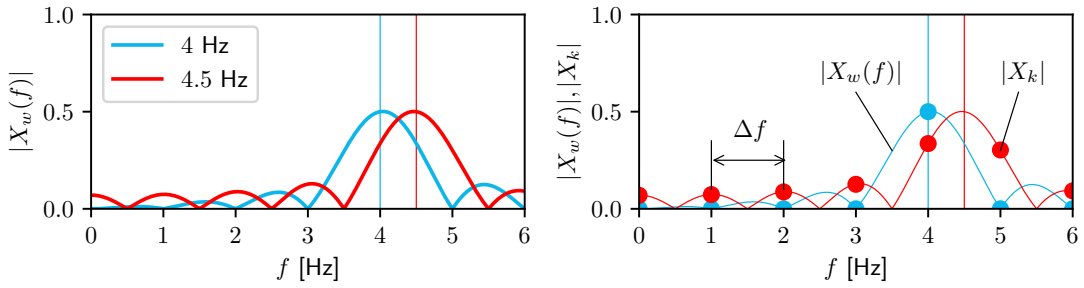


Figure 12.9: Magnitude spectrum $|X_w(f)|$ of time-windowed unit-amplitude, zero-phase cosine at left, for a window length of $T = 1$ s, in blue for $f_0 = 4$ Hz and in red for $f_0 = 4.5$ Hz ($f_s = 64$ Hz). The DFT results of the sampled and time-windowed cosine, in magnitude, i.e., $|X_k|$, are shown as dots at right.

The graph in Figure 12.9 at right, using dots in blue and red, shows the magnitude spectrum $|X_k|$ for the two cosines as obtained with the DFT (12.5) (and divided by $N\Delta t$, see (8.3), see also (H.7); this scaling maintains the analogy with the Fourier transform of continuous-time $x(t)$). The cosine signal was sampled at $f_s = 64$ Hz (intentionally a large value was chosen, such that copies of the spectrum, see (9.10), are far away, and do not impact our analysis of leakage). As the window duration is $T = 1$ s, the frequency step is $\Delta f = 1$ Hz.

The spectrum of the 4 Hz cosine, in blue, looks perfect: all values equal zero, except for a clear peak, with a value of 0.5, at $f = 4$ Hz, exactly as we would expect. Leakage will always occur, as soon as a finite window duration is used (as shown by the underlying blue curve), but in this case we were 'lucky' to 'sample' the spectrum

exactly at the peak and at the zeros of the spectrum of the windowed signal $X_w(f)$; working with the DFT we do not notice leakage at all. The cosine frequency $f_0 = 4$ Hz is an integer multiple ℓ of the frequency step $\Delta f = \frac{1}{T}$. This cosine wave fits exactly an integer number of $\ell = 4$ times in the window; for the period T_0 of this cosine, $4T_0 = T$ and hence $f_0 = 4\frac{1}{T}$.

For the cosine with a frequency of $f_0 = 4.5$ Hz (red), the situation is different. With a frequency step of $\Delta f = \frac{1}{T} = 1$ Hz, the analysis frequencies do *not* coincide with the peak at 4.5 Hz; instead, the spectrum is evaluated at $f = 4$ and 5 Hz. The other $|X_k|$-values are non-zero, as we happen to evaluate the spectrum at the peaks of the side lobes of the windowed signal $X_w(f)$. The spectrum, shown by the red dots, looks quite different from the ideal situation, substantial spectral *leakage* occurs.

Obviously, when performing spectral analysis in practice, we typically do *not* know in advance whether the harmonics of interest will fit an integer number of times in window T; we need to be prepared to deal with leakage.

12.7 Zero-padding

The DFT uses as input N time samples of the signal, x_n with $n = 0, \ldots, N-1$, and evaluates the DTFT $X_{\Delta t,T}(f)$ at N discrete frequencies in the frequency domain. The DTFT of the time-windowed and sampled signal, sequence x_n, is a *continuous* function of frequency f.

We could evaluate this function $X_{\Delta t,T}(f)$ at many more discrete frequencies than the default $k = 0, \ldots, N-1$ times Δf; this would require changes to the code for the numerical implementation of the DFT, for instance in Python. As a quick work-around we could instead add samples with zero value to the DFT input sequence, and use again the default DFT algorithm. This is referred to as zero-padding.

In zero-padding, we add to the time-windowed and sampled signal, sequence x_n, samples with a value equal to zero, *extending* the signal, yielding a larger time-window duration T_x, with $T_x > T$, and thereby *reducing* the frequency step $\Delta f = \frac{1}{T}$ to $\Delta f_x = \frac{1}{T_x} < \frac{1}{T}$. No new information is being added (only zeros), and we are interpolating the spectrum, i.e., evaluating the underlying function $X_{\Delta t,T}(f)$ with a smaller frequency step: as a result, resolution improves.

Another application of zero-padding is to make the size of the input sequence N equal to a power of two: $N = 2^m$. In that case, the much faster FFT algorithm can be used to compute the DFT coefficients (see Section 12.2.4).

We continue with Example 12.1 and zero-pad sequence x_n from $T = 4$ s to $T_x = 10$ s. EX 12.6

Solution The result of zero-padding x_n is shown as a function of (restored, continuous) time t on the horizontal axis in Figure 12.10.

The corresponding DFT, by its magnitude $|X_k|$, as a function of frequency f on the horizontal axis, is shown in Figure 12.11 as stems. The frequency range covers $[0, f_s)$ in both cases, but the frequency step size is clearly smaller for the zero-padded signal at right: the graph is more dense. The dotted curves show $|X_{\Delta t,T}(f)|$ as a continuous function of frequency for $f_s = 3$ Hz and $T = 4$ s.

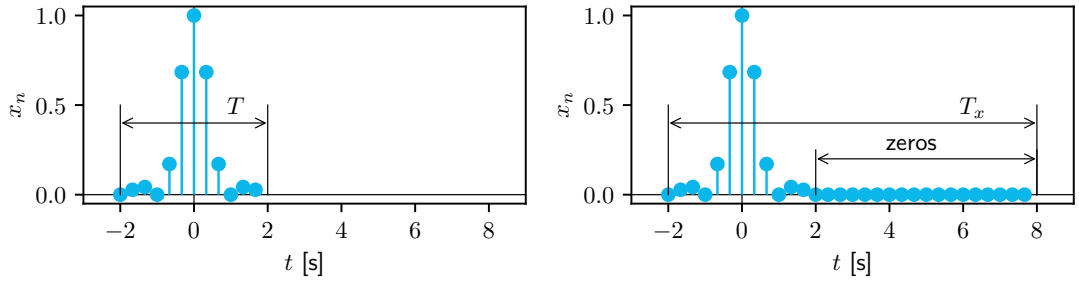


Figure 12.10: Time-windowed and sampled signal, sequence x_n for $n = 0, \ldots, N-1$, with $f_s = 3$ Hz, $N = 12$ and time window length $T = N\Delta t = 4$ s, at left, shown by the stems (identical to Figure 12.1 at left), and zero-padded signal with a total window duration of $T_x = 10$ s, $N_x = 30$, at right. Samples are shown as a function of the 'restored' continuous time on the horizontal axis, i.e., $t = n\Delta t$.

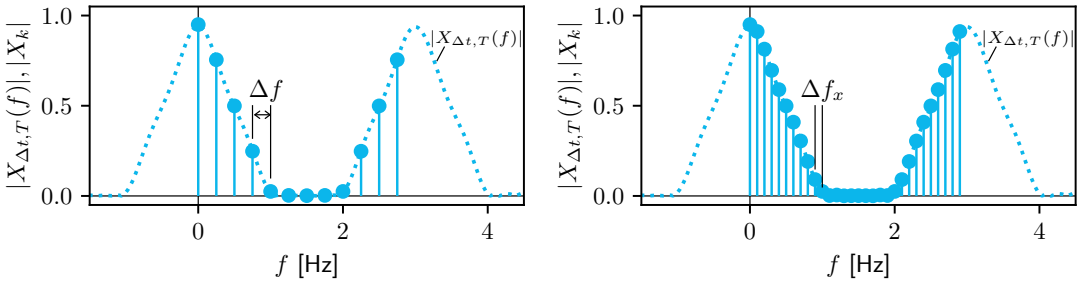


Figure 12.11: DTFT $|X_{\Delta t,T}(f)|$ (dotted), evaluated at discrete frequencies shown by the stems $|X_k|$, at left, based on DFT of x_n with $T = 4$ s and $\Delta f = \frac{1}{T} = \frac{1}{4}$ Hz (identical to Figure 12.1 at right), and, at right, based on zero-padded input sequence with $T_x = 10$ s and $\Delta f_x = \frac{1}{T_x} = \frac{1}{10}$ Hz; sampling frequency $f_s = 3$ Hz. DFT coefficients are shown as a function of 'restored' continuous frequency on the horizontal axis, i.e., $f = k\Delta f$.

With the DFT defined, the entire procedure of Fourier-transforming finite-length, discrete-time signals, typically sampled measurements, has been covered. In Part IV, we discuss how the DFT is used to determine power spectral density functions of deterministic signals in Chapter 13, of random signals in Chapter 14, and of non-stationary signals in Chapter 15.

IV

Spectral estimation

13

Energy and power spectral density

In this chapter we introduce the Energy Spectral Density (ESD) $G(f)$ and the Power Spectral Density (PSD) $S(f)$; both are a continuous function of frequency f. They show, respectively, how energy and power of a signal are distributed over frequency. We start in Chapter 13 with *deterministic* signals, both in continuous time and discrete time. Chapter 14 covers spectral *estimates* of finite-duration discrete-time sample sequences of *random* signals.

13.1 Introduction

The concepts of energy and power, in the context of signal processing and analysis, originate from electrical engineering. Energy refers to the total amount of *work* a signal can do. The energy of a continuous-time signal $x(t)$ equals the integral of the square of the signal over time, see (2.17). Power is the *rate* at which energy is transferred, delivered or dissipated. As (2.18) shows, power equals energy per unit time. Obviously, signal *amplitude* drives energy and power. For a cosine signal, doubling the amplitude increases the power by a factor of 4.

Figure 13.1 shows a basic electrical circuit with a power source (e.g., a battery) and a resistor (e.g., a light bulb). When a voltage $u(t)$ is applied over a resistor (resistance R), current $i(t)$ starts to flow and the resistor will dissipate heat. According to Ohm's law $u(t) = i(t)R$, the electric current $i(t)$ is directly proportional to the voltage $u(t)$, with resistance R being the proportionality constant. For convenience we refer, in spectral analysis, to a $R = 1\ \Omega$ resistor, such that $u(t) = i(t) = x(t)$. This is referred to as 'unit resistance'.

With signal $x(t)$ being a voltage in [V] (Volt) or a current in [A] (Ampere), energy E in (2.17) is expressed in Joule [J], and power P in (2.18) in Watt [W]; Watt is equal to Joule per second (all in relation to unit resistance). Signal power is the product of voltage and current; (2.18) expresses the time-averaged value of the instantaneous power.

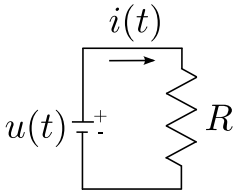


Figure 13.1: Electrical circuit with a power source at left and a resistor at right. According to Ohm's law, voltage $u(t)$, current $i(t)$ and resistance R are related through $u(t) = i(t)R$.

13.2 Continuous-time signals

We start with signals in continuous time and discuss periodic signals, which are typically power signals, and aperiodic signals, which are generally energy signals. These signals have, respectively, a power spectral density and energy spectral density, functions which describe how signal power and energy are distributed over frequencies. We then consider signals which have a limited duration in time, which generally have limited energy and *zero* power, and show how a power spectral density can still be defined for these signals.

13.2.1 Power spectrum and power spectral density

The average power of *periodic* signal $x(t)$ with period T_0 (see also (2.19)) is:

$$P = \frac{1}{T_0}\int_{T_0} |x(t)|^2 \,\mathrm{d}t = \sum_{k=-\infty}^{k=\infty} |X_k|^2 \tag{4.16}$$

Proof For a real, periodic signal $x(t)$ we have:

$$\begin{aligned} P &= \frac{1}{T_0}\int_{T_0} x^2(t)\,\mathrm{d}t = \frac{1}{T_0}\int_{T_0} x(t)x(t)\,\mathrm{d}t \\ &= \frac{1}{T_0}\int_{T_0} x(t)\left(\sum_{k=-\infty}^{k=\infty} X_k e^{jk\omega_0 t}\right)\mathrm{d}t = \sum_{k=-\infty}^{k=\infty} X_k\left(\frac{1}{T_0}\int_{T_0} x(t)e^{jk\omega_0 t}\,\mathrm{d}t\right) \end{aligned} \tag{13.1}$$

where we substituted the complex exponential Fourier series, (4.2), and then changed the order of summation and integration. Substitute the complex conjugate of X_k for the expression between parentheses on the right-hand side of (13.1) and we obtain:

$$P = \sum_{k=-\infty}^{k=\infty} X_k X_k^* = \sum_{k=-\infty}^{k=\infty} |X_k|^2$$

For periodic signals, plotting $|X_k|^2$ as a function of kth harmonic kf_0, $k \in \mathbb{Z}$, yields the (discrete) power spectrum, a line spectrum (Section 3.6). The average power is obtained by adding up the contributions at the discrete frequencies of the line spectrum, see (4.16).

The power spectral *density* function $S(f)$ of periodic signal $x(t)$ is a *continuous* function of frequency, defined as:

$$S(f) = \sum_{k=-\infty}^{\infty} |X_k|^2 \delta(f - kf_0) \tag{13.2}$$

The Dirac delta function is used to put the contribution to the (average) power of the kth harmonic $|X_k|^2$ *at* frequency kf_0, while for all other frequencies $S(f)$ is equal to zero. Compare this result with (6.12). The proof is given in [12].

The (average) power P follows from *integrating* the power spectral density function $S(f)$:

$$P = \int_{-\infty}^{\infty} S(f)\,\mathrm{d}f \tag{13.3}$$

When signal $x(t)$ has unit [V], its average power P (on a 1 Ohm basis) has unit Watt [W] and the unit of the power spectral density function $S(f)$ is [W/Hz].

EX 13.1

Show that signal $x(t) = A\cos(2\pi f_0 t)$, with period $T_0 = \frac{1}{f_0}$, has a line spectrum, and determine the PSD $S(f)$ of this signal.

Solution The complex Fourier series coefficients are $X_{-1} = X_1 = \frac{A}{2}$; all other coefficients are zero. Signal $x(t)$ has a double-sided amplitude line spectrum, with amplitude $\frac{A}{2}$ at frequency $-f_0$ and amplitude $\frac{A}{2}$ at frequency f_0.

Showing $|X_k|^2$ as a function of k, which refers to frequency kf_0, yields the (discrete) power spectrum – in this example, $|X_{-1}|^2 = |X_1|^2 = \frac{A^2}{4}$, see Figure 13.2 at left. In Section 4.4 it has been shown that, for a real signal $x(t)$, the amplitude spectrum is *even*, $|X_k| = |X_{-k}|$, and so is the power spectrum.

The (average) power P follows from (4.16): $P = \frac{A^2}{4} + \frac{A^2}{4} = \frac{A^2}{2}$. If signal $x(t)$ is a voltage [V], power P (normalized to 1 Ohm) has unit Watt [W]. Introducing a non-zero phase θ into the signal $x(t) = A\cos(2\pi f_0 t + \theta)$ does *not* change the amplitude and power spectra.

The power spectral *density* function of cosine signal $x(t)$ can be written as $S(f) = \frac{A^2}{4}(\delta(f + f_0) + \delta(f - f_0))$ – see [12] for a derivation – and it is shown in Figure 13.2 at right. The Dirac delta function appears in the frequency domain: $\delta(f)$. It has an infinitesimally short 'duration', and yet a unit area (see Appendix B). With signal $x(t)$ a voltage, power spectral density function $S(f)$ has unit [W/Hz].

The (average) power P follows from (13.3): $P = \frac{A^2}{4} + \frac{A^2}{4} = \frac{A^2}{2}$ W.

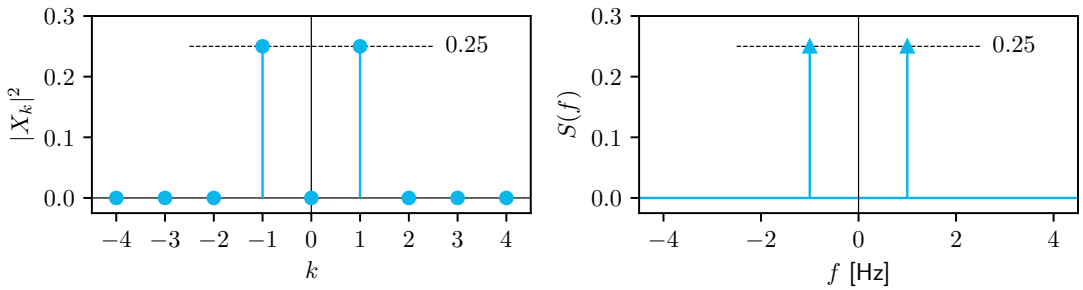


Figure 13.2: Power (line) spectrum $|X_k|^2$ of cosine signal $x(t) = A\cos(2\pi f_0 t)$ at left, and its power spectral density $S(f)$ at right, for $A = 1$ and $f_0 = 1$ Hz.

13.2.2 Energy spectral density

Aperiodic signals $x(t)$ have a Fourier transform $X(f)$. Their energy can be calculated as:

$$E = \int_{-\infty}^{\infty} |x(t)|^2 \,\mathrm{d}t = \int_{-\infty}^{\infty} |X(f)|^2 \,\mathrm{d}f \tag{5.19}$$

where the energy spectral density function $G(f)$ is defined as:

$$\boxed{G(f) = |X(f)|^2} \tag{13.4}$$

The total energy E is obtained by *integrating* $G(f)$ over all frequencies:

$$E = \int_{-\infty}^{\infty} G(f) \,\mathrm{d}f \tag{13.5}$$

When signal $x(t)$ has unit [V], $X(f)$ in (5.1) has unit [V/Hz]; its energy E (on a 1 Ohm basis) has unit Joule [J] and the unit of the energy spectral density function $G(f)$ is [J/Hz].

Proof For a real, aperiodic signal $x(t)$ the energy can be expressed as:

$$E = \int_{t=-\infty}^{\infty} x(t)^2 \,\mathrm{d}t = \int_{t=-\infty}^{\infty} x(t)x(t) \,\mathrm{d}t = \int_{t=-\infty}^{\infty} x(t) \left(\int_{f=-\infty}^{\infty} X(f) e^{j2\pi ft} \,\mathrm{d}f \right) \mathrm{d}t$$

where we substituted the inverse Fourier transform, (5.2). We can move $x(t)$ within the integral (in f) between parentheses and then change the order of integration:

$$E = \int_{t=-\infty}^{\infty} \int_{f=-\infty}^{\infty} x(t)X(f)e^{j2\pi ft} \,\mathrm{d}f \,\mathrm{d}t = \int_{f=-\infty}^{\infty} \left(\int_{t=-\infty}^{\infty} x(t)X(f)e^{j2\pi ft} \,\mathrm{d}t \right) \mathrm{d}f$$

We can now move $X(f)$ out of the integral (in t) between parentheses and obtain:

$$E = \int_{f=-\infty}^{\infty} X(f) \int_{t=-\infty}^{\infty} x(t)e^{j2\pi ft} \,\mathrm{d}t \,\mathrm{d}f = \int_{f=-\infty}^{\infty} X(f)X^*(f) \,\mathrm{d}f = \int_{f=-\infty}^{\infty} |X(f)|^2 \,\mathrm{d}f$$

using the Fourier transform's complex conjugate, (5.1), with $x(t) = x^*(t)$ for a real signal.

EX 13.2 Determine the ESD $G(f)$ and the total energy E of continuous-time, deterministic, aperiodic voltage signal $x(t) = \mathrm{sinc}^2(t)$.

Solution The Fourier transform is $X(f) = \Lambda(f)$ (7.7), see Figure 8.2. The energy spectral density $G(f) = |X(f)|^2$ is quickly found as:

$$G(f) = \begin{cases} (1+f)^2 & \text{for} \quad -1 \le f < 0 \\ (1-f)^2 & \text{for} \quad 0 \le f \le 1 \\ 0 & \text{other} \quad f \end{cases}$$

Signal $x(t)$ and its energy spectral density $G(f)$ are shown in Figure 13.3.

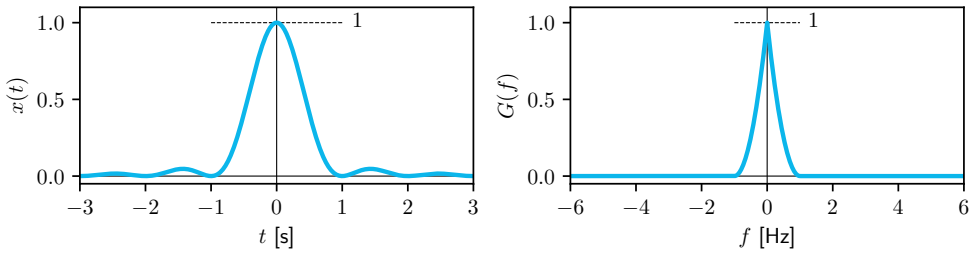


Figure 13.3: Aperiodic signal $x(t) = \mathrm{sinc}^2(t)$ at left, and its energy spectral density $G(f)$ at right.

The total energy E of $x(t)$ is found with (13.5):

$$E = \int_{-\infty}^{\infty} G(f) \,\mathrm{d}f = \int_{-1}^{0} (1+f)^2 \,\mathrm{d}f + \int_{0}^{1} (1-f)^2 \,\mathrm{d}f = \frac{1}{3} + \frac{1}{3} = \frac{2}{3} \text{ J}$$

This aperiodic signal $x(t) = \mathrm{sinc}^2(t)$ is clearly a finite-energy signal, and has zero (average) power: $P = 0$.

In Section 5.6 it has been shown that for a real signal $x(t)$ the amplitude spectrum is an *even* function of frequency, $|X(f)| = |X(-f)|$, and so is the energy spectral density $G(f)$. The default in this book is that we work with *two-sided*, or *double-sided* spectra, although we may show only one side, typically for positive frequencies $f \geq 0$, to save space. We must not forget that in calculating the power (or energy) of a signal to include the positive *and* negative frequencies: we must integrate from $f = 0$ to ∞, then multiply by 2.

13.2.3 Finite-duration signals

For an infinite signal duration or record, $T \to \infty$, which is only possible in theory, as the signal's total energy is finite the average power is zero, (2.18), and so would be its PSD $S(f)$. For a *finite* signal duration, however, with power defined as energy per unit time, we often still consider the PSD $S(f)$, obtained by dividing the ESD $G(f)$ by the signal duration T:

$$S(f) = \lim_{T\to\infty} \frac{|X_T(f)|^2}{T} \tag{13.6}$$

(see for a derivation [5] and [12]). Here, $X_T(f)$ indicates the Fourier transform of signal $x(t)$ *time-truncated* to duration T, i.e., $x_T(t) = \Pi(\frac{t}{T})x(t)$, and therefore it equals zero outside the interval $[-\frac{T}{2}, \frac{T}{2}]$, as in Chapter 8, and $x_T(t) \overset{\mathcal{F}}{\leftrightarrow} X_T(f)$.

With (13.6) the average power P of signal $x_T(t)$ over interval T can be computed using (13.3). Note that only power signals truly possess a power spectral density, but since they have infinite energy this challenges the existence of the Fourier transform integral (5.3). Definition (13.6) uses time truncation and can be applied to physically-realizable signals.

EX 13.3

We continue with Example 13.2, consider the (finite) time duration of the signal $x(t) = \text{sinc}^2(t)$ and compute the energy and power spectral densities.

Solution Figure 13.3 at left shows that signal $x(t) = \text{sinc}^2(t)$, though in principle running from $t = -\infty$ to $t = \infty$, decays to zero on both sides; therefore it can be Fourier-transformed and integrated. In practice we work with a *finite* signal duration T; the top image of Figure 13.4 shows the ESD $G(f)$ for two time-windowed versions of $x(t)$: $T = 2$ s (left) and $T = 10$ s (right); these graphs were created in discrete time. Considering the signal over a longer time window yields more energy.

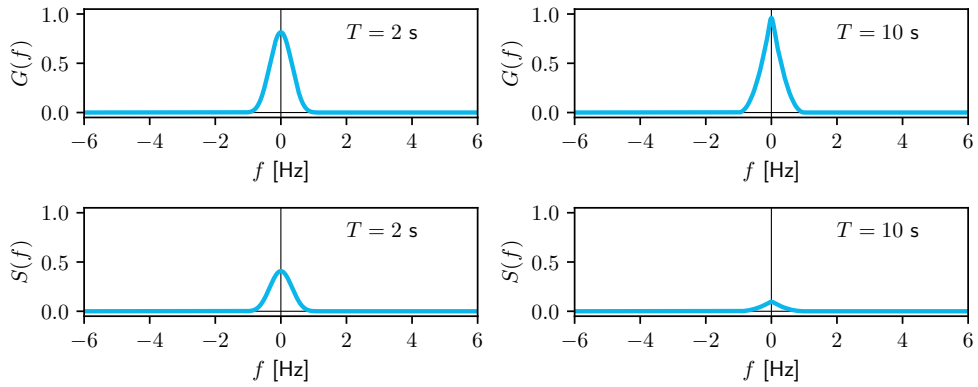


Figure 13.4: ESD $G(f)$ (top) and PSD (bottom) of signal $x(t) = \text{sinc}^2(t)$, time-windowed to $T = 2$ s at left and to $T = 10$ s at right.

Dividing $G(f)$ by T yields the PSD $S(f)$, the bottom image of Figure 13.4.

13.3 Discrete-time signals

We limit the discussion to finite-duration discrete-time signals, which have the most practical relevance. For infinite-duration discrete-time signals and applications of the DTFT in computing their energy and power spectral densities, we refer to other books, e.g., [3].

In practice, we deal with finite signal duration $T = N\Delta t$: N samples x_n, with $n = 0, \ldots, N-1$, of signal $x(t)$, with fixed sampling interval Δt. Because these signals have finite duration and are, in most practical situations, limited in amplitude, these signals are by definition *energy* signals (Section 2.6). Therefore, we start with defining energy and the energy spectral density, and then use the insights gained in Section 13.2.3 to move on to defining power and the power spectral density for limited-duration, discrete-time sequences x_n.

13.3.1 Energy spectral density

In discrete time, energy can be defined through discretizing integral (2.17):

$$E = \lim_{N\to\infty} \sum_{n=-\frac{N}{2}}^{\frac{N}{2}} |x_n|^2 \Delta t = \Delta t \lim_{N\to\infty} \sum_{n=-\frac{N}{2}}^{\frac{N}{2}} x_n^2$$

with $t = n\Delta t$ and $T = N\Delta t$ (and assuming even N for convenience). For a finite-duration, real discrete-time sequence, x_n, running from $n = 0$ to $n = N-1$, energy can be computed as:

$$\boxed{E = \Delta t \sum_{n=0}^{N-1} x_n^2} \tag{13.7}$$

Parseval's theorem, stating that energy can also be computed in the frequency domain (see Section 5.9), also holds in discrete time. We can discretize integral (5.19):

$$E = \sum_{k=0}^{N-1} |X_k|^2 \Delta f \tag{13.8}$$

with X_k the DFT of sequence x_n, (12.5), and $\Delta f = \frac{1}{T}$. We obtain:

$$\boxed{E = \frac{1}{T} \sum_{k=0}^{N-1} |X_k|^2} \tag{13.9}$$

This leads to the following definition of the energy spectral density in discrete frequency:

$$\boxed{G_k = |X_k|^2} \tag{13.10}$$

The energy E can be obtained in the frequency domain by adding up G_k for $k = 0, \ldots, N-1$, then multiplying by frequency spacing $\Delta f = \frac{1}{T}$, (13.9).

Proof Consider (13.7); for a real sequence we have $x_n = x_n^*$. Substitute for x_n the inverse DFT of X_k, (12.6), and for x_n^* also its inverse DFT but using index ℓ rather than k:

$$x_n^2 = x_n x_n = x_n x_n^* = \frac{1}{\Delta t}\frac{1}{N}\left(\sum_{k=0}^{N-1} X_k e^{jk\frac{2\pi}{N}n}\right)\frac{1}{\Delta t}\frac{1}{N}\left(\sum_{\ell=0}^{N-1} X_\ell^* e^{-j\ell\frac{2\pi}{N}n}\right)$$

Multiplying the two summations, and expanding them (and for the moment leaving out factor $\frac{1}{(N\Delta t)^2}$ in front), leads to two sums. The first is a sum of N terms with equal indices ($k = \ell$): $X_k X_k^*$ for $k = 0, \ldots, N-1$, where the associated complex exponentials cancel each other out ($e^{jk\frac{2\pi}{N}n}e^{-jk\frac{2\pi}{N}n} = e^0 = 1$). The second sum is a sum of $N^2 - N = N(N-1)$ cross terms with unequal indices ($k \neq \ell$): $X_k X_\ell^* e^{j(k-\ell)\frac{2\pi}{N}n}$ for $k, \ell = 0, \ldots, N-1$ with $k \neq \ell$.

Including summation $\sum_{n=0}^{N-1}$ of (13.7), we get for the first sum:

$$\sum_{n=0}^{N-1} x_n^2 = \frac{1}{(N\Delta t)^2}\sum_{n=0}^{N-1}\sum_{k=0}^{N-1} X_k X_k^* = \frac{1}{(N\Delta t)^2} N \sum_{k=0}^{N-1} |X_k|^2 = \frac{1}{\Delta t}\frac{1}{T}\sum_{k=0}^{N-1} |X_k|^2$$

proving expression (13.9) for energy E, provided that the summation over n of all the cross terms, i.e., the second sum, equals zero.

Including summation $\sum_{n=0}^{N-1}$ of (13.7), all cross-terms together yield (leaving out constant factors):

$$\sum_{n=0}^{N-1}\sum_{k=0}^{N-1}\sum_{\substack{l=0 \\ l\neq k}}^{N-1} X_k X_l^* e^{j(k-\ell)\frac{2\pi}{N}n}$$

Focusing on the first summation and the complex exponential, we find that $\sum_{n=0}^{N-1} e^{j(k-\ell)\frac{2\pi}{N}n} = 0$ for $k \neq \ell$, with $k, \ell \in \{0, \ldots, N-1\}$, similarly as in Appendix H, by using the geometric series identity (A.22) with $a = e^{j\frac{2\pi}{N}(k-\ell)}$ and using n as the summation index (rather than k).

13.3.2 Power spectral density

Finite-duration signals are, by definition, energy signals: their average power is zero. The average power of a discrete-time sequence x_n during the signal *duration* can be computed through dividing the signal energy E, (13.7), by signal duration $T = N\Delta t$:

$$P = \frac{E}{T} = \frac{1}{N\Delta t}\left(\Delta t \sum_{n=0}^{N-1} x_n^2\right) = \frac{1}{N}\sum_{n=0}^{N-1} x_n^2 \tag{13.11}$$

The power of a discrete-time sequence (N samples) can also be computed in the frequency domain, the equivalent of (13.9):

$$\boxed{P = \frac{1}{T}\sum_{k=0}^{N-1}\frac{|X_k|^2}{T}} \tag{13.12}$$

This leads to the following definition of the power spectral density at discrete frequencies, for a discrete-time sequence x_n:

$$\boxed{S_k = \frac{|X_k|^2}{T}} \tag{13.13}$$

with X_k from the DFT, (12.5). This expression is also referred to as the *periodogram*.

The average power P can be obtained in the frequency domain by adding up S_k for $k = 0, \ldots, N-1$, then multiplying by the frequency spacing $\Delta f = \frac{1}{T}$, see (13.12).

EX 13.4

We continue with Example 13.2 and consider $N = 12$ discrete-time samples of signal $x(t) = \text{sinc}^2(t)$, with $\Delta t = \frac{1}{3}$ s and window length $T = 4$ s, as shown in Figure 13.5 at left, and determine the periodogram.

Solution The periodogram for $N = 12$ discrete-time samples of signal $x(t) = \text{sinc}^2(t)$ is shown in Figure 13.5 at right. The dotted curves show the infinite-duration, continuous-time signal (at left) and the corresponding theoretical, infinite-duration signal energy spectral density divided by signal duration, $\frac{G(f)}{T}$ (at right).

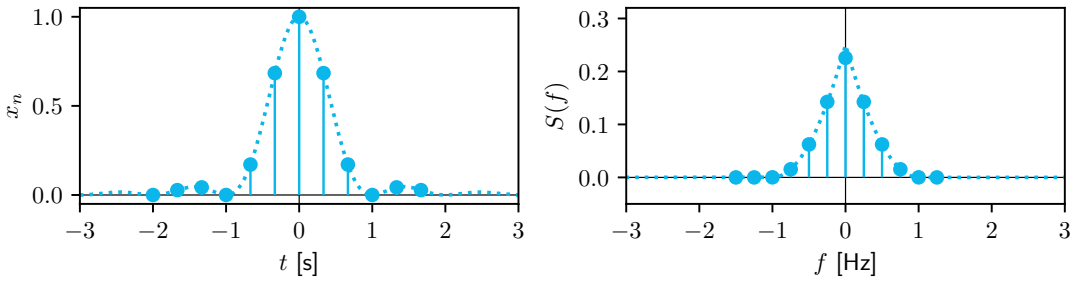


Figure 13.5: Discrete-time signal x_n (left), as a result of sampling signal $x(t) = \text{sinc}^2(t)$ – the same signal as in Figure 12.1 ($N = 12$, $f_s = 3$ Hz, $T = 4$ s) – and its periodogram (right), evaluating the PSD $S(f)$ at N discrete frequencies, with $\Delta f = \frac{1}{T}$, from $[-\frac{f_s}{2}, \frac{f_s}{2})$, rather than $[0, f_s)$ in Figure 12.1 at right.

This chapter has defined the energy spectral density and power spectral density functions for deterministic signals, in continuous time and discrete time. The following chapter discusses how to obtain *estimates* of these spectra, especially the PSD, from a finite-duration, discrete-time sequence x_n of a *random* signal.

14

Spectral estimation

So far, we worked with deterministic signals. However, signals measured in practice are subject to (uncontrolled) variability, whether small or large, both in the phenomenon itself and in the measurement of it. We refer to this variability as *noise*. Signals which include noise are called *random* signals. Appendix K presents a few statistical concepts for random signals, but by no means embodies a full account of the subject; the reader is referred to, e.g., [5] or [13].

In this chapter we work with random signals and *estimate* the Power Spectral Density (PSD). The resulting PSD estimate is denoted with a 'hat' symbol, $\hat{S}(f)$ or $\hat{S}_k$, to indicate that it results from a measured signal which exhibits variability due to noise. Repeating the experiment and determining the PSD again will result in a *different* estimate due to the uncontrolled variability. The most practically relevant statistical properties of the spectral estimator are covered in this chapter.

14.1 Assumptions

The random signal $x(t)$ is assumed to be *ergodic*. This means that a time average of a realization of the signal equals the corresponding ensemble average (the statistical expectation). Ergodicity implies *stationarity*, see Appendix K. The random signal is assumed to be stochastically continuous and to have a mean of zero [5]. The latter means in practice that a trend in the measured time series of the signal (e.g., an offset and/or slope) should be removed from the signal prior to spectral analysis.

14.2 Power spectral density

As [5] points out, the very nature of 'stationarity' suggests that a realization of a random signal will almost certainly not 'decay' at infinity, and therefore violate (5.3). Therefore, as in the previous chapter, a time-truncated version of signal $x(t)$ is considered: $x_T(t) = \Pi\left(\frac{t}{T}\right) x(t)$, which equals to zero outside the interval $[-\frac{T}{2}, \frac{T}{2}]$; the signal has been time-truncated to a duration of T seconds. $X_T(f)$ denotes the Fourier transform of random signal $x_T(t)$ and is itself a random function as well; $X_T(f)$ considered at a single, discrete frequency is a random variable, and its variability is described by a statistical distribution.

Every time we observe a new realization of the random signal, and subsequently compute the Fourier transform of the time-truncated realization, it looks different. This aspect needs to be dealt with in the definition of the PSD.

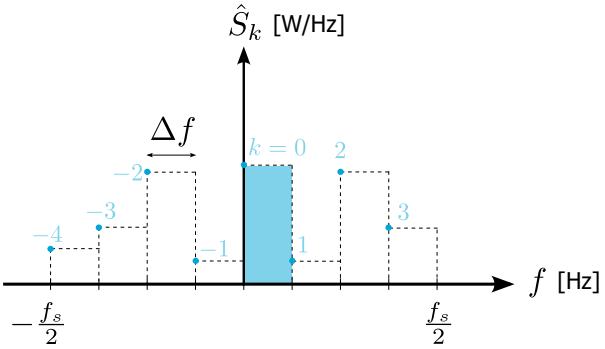


Figure 14.1: Example of a periodogram of a discrete-time signal, consisting of $N = 8$ samples. The estimated periodogram concerns a *density*, and is evaluated at $N = 8$ discrete frequencies $k\Delta f$, with $k = -4, -3, -2, -1, 0, 1, 2, 3$, covering the frequency range $[-\frac{f_s}{2}, \frac{f_s}{2})$.

For a continuous-time random, time-truncated signal $x_T(t)$, the PSD[1] is defined as:

$$S(f) = \lim_{T\to\infty} E\left(\frac{|X_T(f)|^2}{T}\right) \tag{14.1}$$

see [5], [12]. The above expression is similar to (13.6) for a continuous-time deterministic signal in the previous chapter, though in the above expression the statistical expectation operation E is introduced, as we deal with a *random* signal here. This statistical definition would practically imply taking the mean over all realizations of the random signal (rather than depending on just a single realization).

In practice, we have to be satisfied with an *estimate* for the PSD, indicated by the hat symbol, computed on the basis of a single realization, of finite duration, of the random signal:

$$\hat{S}(f) = \frac{|X_T(f)|^2}{T}$$

with $X_T(f)$ the Fourier transform of the measured finite-duration signal.

14.3 Periodogram

We focus on discrete-time, random signals (sequences), as this is applicable in practice. Based on N samples $x_0, \ldots, x_{N-1}$ of a *single realization* (of finite duration) of the random signal, we compute the DFT $X_0, \ldots, X_{N-1}$ (12.5), and the *estimate* for the PSD reads:

$$\hat{S}(f_k) = \hat{S}_k = \frac{1}{T}|X_k|^2 \tag{14.2}$$

with $k = 0, \ldots, N-1$ for frequencies $k\Delta f$ – the analysis frequencies – with $\Delta f = \frac{1}{T}$ (see Chapter 12). This expression is identical to (13.13) in the previous chapter. Estimate (14.2) is referred to as the *periodogram* (in Python: `scipy.signal.periodogram`).

Figure 14.1 shows an example of a periodogram, for $N = 8$. By default, the DFT through `fft` yields X_k for $k = 0, \ldots, N-1$, as shown, for a different example, in Figure 12.2 at right.

[1] In practice it is common to denote, by means of indices, to which signal the PSD refers, hence $S_{xx}(f)$, in this case with signal $x(t)$ or actually the time-truncated signal $x_T(t)$; a double index is used, because the *cross*-spectral density of two different signals, e.g., $x(t)$ and $y(t)$, might also be considered (this is beyond the scope of this book). (14.1) is sometimes referred to as the *auto*-PSD. In this book, as we work solely with signal $x(t)$ or its time truncation, we omit these indices.

In Figure 14.1, the estimates $\hat{S}_k$ for $k = 4, 5, 6, 7$ have been moved to the left (e.g., using `fftshift`), now representing respectively $k = -4, -3, -2, -1$, to yield a symmetric graph, centered at $f = 0$, as explained in Figure 11.2. As (13.3) states, the (average) power P follows as the integral over the PSD $S(f)$. Equation (13.12) states that, with a discrete frequency estimate, as with periodogram (14.2), the power is obtained through discretization of the integral as the sum of the estimates $\hat{S}_k$ multiplied by frequency step $\Delta f = \frac{1}{T}$.

The term periodogram was introduced in [14] with the purpose of searching for periodicities in a measured signal, hence concerning discrete spectra based on Fourier series.

EX 14.1

The application of the periodogram is demonstrated for a periodic discrete-time voltage signal. We use a plain cosine signal $x(t) = \cos(2\pi f_0 t)$, with $f_0 = 1$ Hz, without and with noise, shown in Figure 14.2. The sampling frequency is $f_s = 7$ Hz and the measurement duration $T = 10$ s. Note that the discrete-time samples (black dots) are connected here by line segments for clearer visualization (it is *not* a continuous-time signal).

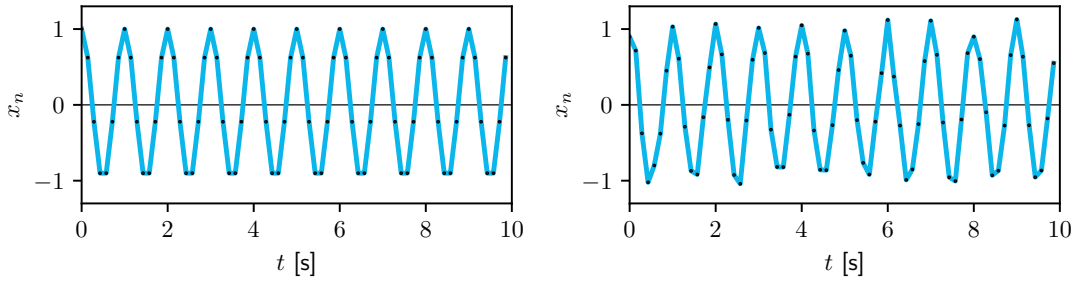


Figure 14.2: Cosine signal in discrete time: at left, deterministic, without noise, and at right with noise (additive white Gaussian noise with standard deviation $\sigma = 0.1$); $f_0 = 1$ Hz, unit amplitude.

Solution Figure 14.3 shows the estimated periodogram, on a linear scale (top row) and logarithmic scale (bottom row, expressed in dB (Section 2.5)). At left for the deterministic cosine signal, and at right for the random cosine signal (i.e., the cosine with noise added). The periodogram estimates (the black dots at the discrete frequencies $k\Delta f$) are connected by blue line segments for clearer visualization.

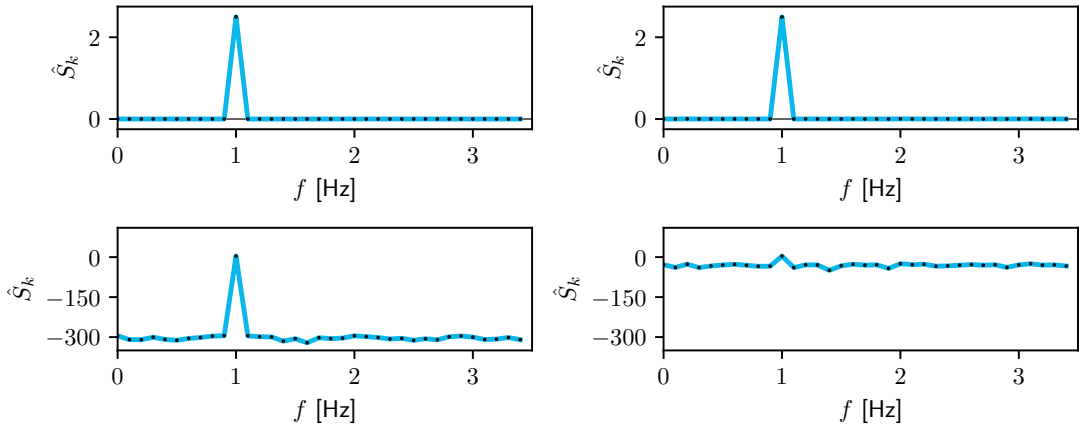


Figure 14.3: Periodograms of a cosine signal: at left for a deterministic signal, at right with noise (this corresponds to Figure 14.2). The PSD magnitude is shown on a linear scale (top row, in [W/Hz]) and logarithmic scale (bottom row, in [dBW/Hz]).

The theoretical PSD of a continuous-time 1 Hz cosine signal was shown in Fig-

ure 13.2 at right, with a single 'peak' at a discrete frequency, here 1 Hz (and equally at the negative frequency). In all graphs in Figure 14.3 we see a peak at 1 Hz, with a magnitude of 2.5 W/Hz (exactly at left, approximately at right). The effect of the noise can only clearly be seen when showing the periodogram on a logarithmic axis.

Recall that the periodogram shows a power spectral *density*. Here, the frequency step equals $\Delta f = \frac{1}{T} = 0.1$ Hz. The corresponding power *at* a frequency of 1 Hz equals 2.5 multiplied by $\Delta f = 0.1$, and hence is 0.25 W. By default we work with double-sided spectra, therefore the total power of this signal equals 0.5 W.

Figure 14.3 at left shows that the PSD at other frequencies is around -300 dBW/Hz (which is very, very small; mind the logarithmic scale). The numerical precision of digital computing is typically up to 15 significant digits: the X_k's for the other frequencies are close to zero and numerically at the 10^{-15} level, while the peak is at the 10^{0} level. As a result, $|X_k|^2$ is approximately 10^{-30}, which on a logarithmic dB-scale becomes -300 (and you would still need to take into account the effect of the measurement duration T for a density). In the graph at right, the spectral density is at the -30 dBW/Hz level for other frequencies, due to the addition of white noise (with standard deviation $\sigma = 0.1$).

EX 14.2

This example shows the periodogram applied to a zero-mean Gaussian white noise discrete-time voltage signal, see Figure 14.4 at left (later we will use a longer time window T). White noise is discussed in more detail in Appendix L.

Solution The theoretical autocorrelation function, see Appendix K, in continuous time, for white noise, is $R(\tau) = \sigma^2\delta(\tau)$, with a noise standard deviation of σ. The corresponding theoretical PSD is $S(f) = \sigma^2$, leading to a flat spectrum, such as the dashed line in Figure 14.4 at right. All frequencies contribute equally to the signal, in theory from $f = -\infty$ to ∞ (see Section L.1).

The dashed line actually refers to a discrete-time white noise sequence, hence band-limited white noise, see Section L.2. The power equals the noise variance $\sigma^2 = N_0 B$, and the theoretical PSD is given by (L.4), with one-sided bandwidth $B = \frac{f_s}{2}$ (the Nyquist frequency; to either side), and the total two-sided bandwidth $2B = f_s$.

The PSD estimate using the periodogram (14.2), shown in blue, is very noisy. On average it is close to the dashed line, but the estimated values at the individual discrete frequencies $[0.0, 0.1, \ldots, 1.5)$ Hz with $\Delta f = \frac{1}{T} = 0.1$ Hz show a large degree of variability.

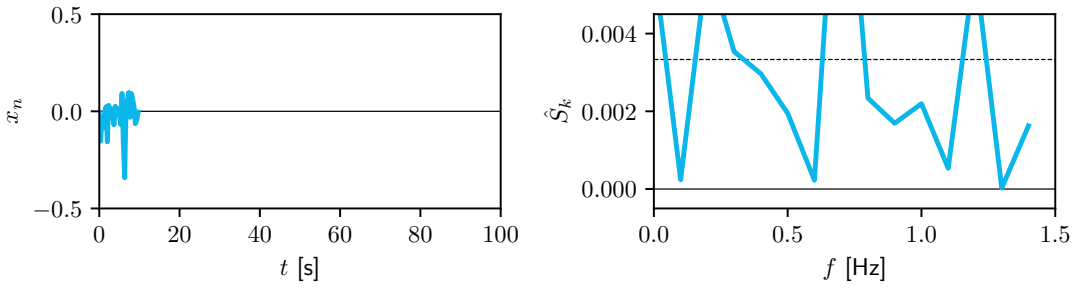


Figure 14.4: White noise signal in discrete time, $f_s = 3$ Hz, $T = 10$ s, simulated based on a Gaussian distribution with $\sigma = 0.1$; at left as a function of time (sequence), at right the estimated periodogram of this sequence, with the discrete-time theoretical PSD, (L.4), as a dashed horizontal line (in [W/Hz]).

EX 14.3

We continue Example 14.2 and apply a low-pass filter to the white noise discrete-time signal, and again consider the periodogram.

Solution A first-order Butterworth low-pass filter with frequency response function $H(f)$ given by (18.1), is applied to the white noise sequence. The result is shown in Figure 14.5: at left the time series, where only a slow behavior remains (the high-frequency variability of Figure 14.4 at left was removed), and at right the estimated periodogram with the theoretical PSD (dashed curve, obtained using (16.15)).

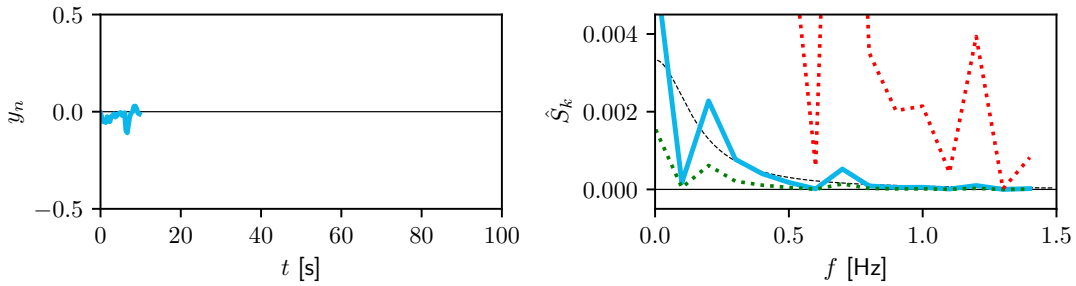


Figure 14.5: Filtered white noise signal in discrete time, $f_s = 3$ Hz, $T = 10$ s, and standard deviation $\sigma = 0.1$, with first-order Butterworth low-pass filter: at left as a function of time (sequence), at right the estimated periodogram (in blue) and theoretical (dashed black) PSD, both in [W/Hz]. The bounds of the 95% confidence interval about the estimate are shown in green (lower limit) and red (upper limit).

The (filtered) signal is subject to variability, and so is the periodogram estimate based on samples of this signal. The periodogram estimate in blue generally follows the dashed black curve of the theoretical PSD, though with a large degree of variability. A confidence interval is used to show the uncertainty in the PSD estimate (J.8). In Figure 14.5 at right the lower limit of the confidence interval is shown by the green dotted line, and the upper limit by the red dotted line, for $\alpha = 0.05$, i.e., a 95% confidence interval.

14.4 Statistical properties of the periodogram

In Appendix J, especially (J.8), we assume that the periodogram is an unbiased estimator of the spectral density (practically meaning that the estimates on average yield the true PSD). Section 6.2 in [12] states, however, that the periodogram is a biased estimator due to the truncation to T, but that it becomes unbiased as $T \to \infty$, i.e., the periodogram is *asymptotically unbiased* (see also Section 6.2 in [5]).

In Appendix J, (J.9) states that the standard deviation of the periodogram estimator, as a measure for the uncertainty in the estimate, equals the target value of the estimate itself, meaning that the uncertainty is as large as the quantity of interest itself. Therefore the periodogram is practically not of much use.

In Figure 14.5 at right we see that for frequencies $f > 1$ Hz, the variability of the blue line is very small, but in this case the dashed curve is already close to zero. For frequencies up to 0.5 Hz, the variability (and deviation from the dashed curve) is much larger. The variability scales directly with the magnitude of the target quantity.

An expression for the confidence interval for the periodogram is derived in Appendix J, using the Chi-squared distribution. The expression for the confidence interval and the standard deviation of the periodogram estimator in Appendix J, derived from Gaussian-distributed samples of the signal, are independent of the measurement duration T. This means that a

longer measurement time T does *not* improve the periodogram estimator, i.e., the extent of variability remains unchanged.

EX 14.4

We continue Example 14.3 with filtered white noise and demonstrate that a longer measurement time T does not help in reducing the periodogram variability.

Solution The first-order Butterworth-filtered white noise used in Example 14.3 is particularly suited to demonstrate variability in the periodogram estimate, as we *know* that the theoretical PSD, the dashed curve in Figure 14.5, is a continuous, smooth curve (as a function of frequency). Hence, the variability in the periodogram estimate (in blue) with respect to the true curve (in black) is clearly shown.

Initially, we may be inclined to think that using more data (i.e., a longer record) will improve the matter, i.e., reduce the variability. This is, however, not the case, as stated above. A longer measurement time will merely lead to an increased frequency resolution: 'more points' of the PSD to be estimated. The variance (per 'point') is not reduced by this; it remains the same.

The example of Figure 14.5 is repeated, but now for $T = 100$ s (instead of $T = 10$ s as before; we actually add 90 s of signal samples to the previous signal record). The extended signal record is shown in Figure 14.6 at left, and the corresponding periodogram at right. It is clear that a longer measurement record does *not* improve the variance of the periodogram estimator (the same level of variability is present as in Figure 14.5 at right in Example 14.3). However, we do see that the frequency resolution has improved (smaller frequency step size), by a factor of 10, to $\Delta f = \frac{1}{T} = 0.01$ Hz, and so the graph has become more dense.

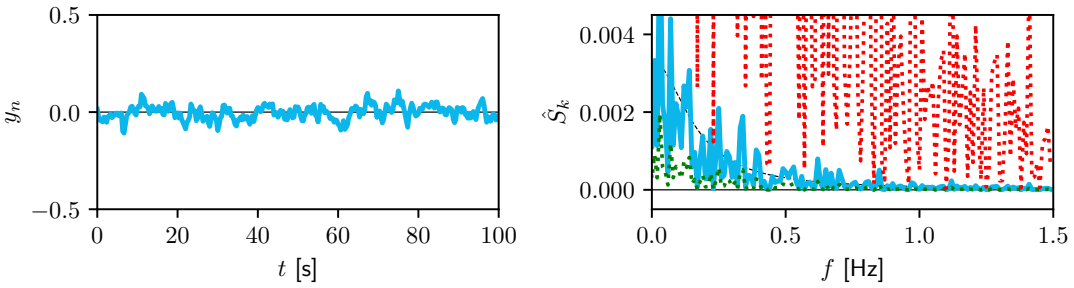


Figure 14.6: Filtered white noise signal in discrete time, $f_s = 3$ Hz, $T = 100$ s, with standard deviation $\sigma = 0.1$ and first-order Butterworth low-pass filter: at left as a function of time (sequence), at right the estimated periodogram (in blue) and theoretical (dashed, black) PSD, both in [W/Hz]. The limits of the 95% confidence interval about the estimate are shown in green (lower limit) and red (upper limit).

14.5 Welch periodogram

Several approaches exist to improve the periodogram estimator. In this book we only present the Welch periodogram [15], as it is often used in practice.[2]

The cornerstone of improving the periodogram is the use of *segments*. The total data record of length N is divided into M segments, each of length $\frac{N}{M}$ (assuming here for convenience that N can be divided by M using integer arithmetic), and a corresponding duration of $\frac{T}{M} = \frac{N}{M}\Delta t$. Segment averaging is suggested by [16].

[2] After the American scientist and researcher P.D. Welch (1929-2023).

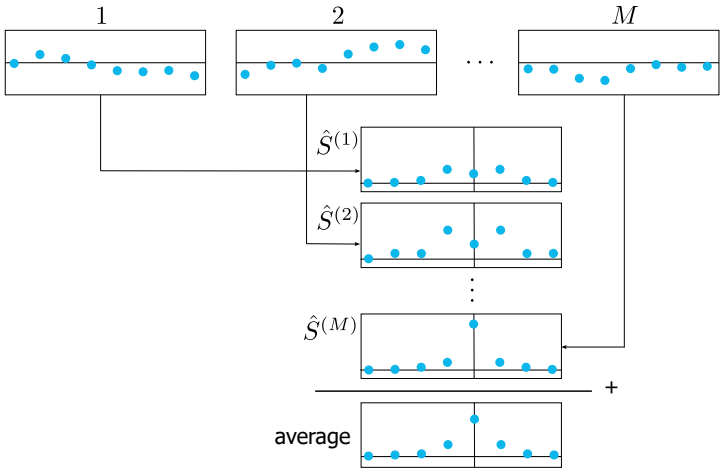


Figure 14.7: Diagram of the computation of the periodogram through averaging over M segments, as featured in the Welch periodogram. The time (top row) and frequency (other rows) indices are absent in these plots to prevent clutter.

For each segment, a periodogram is computed, with the frequency step *increased* to $\Delta f = \frac{1}{\frac{N}{M}\Delta t} = \frac{M}{T}$. Then, the M periodograms are *averaged* (per frequency). The variance reduces by a factor of M, but the price to pay is a poorer frequency resolution.

Figure 14.7 illustrates the process of first dividing the measured discrete-time signal of length N into M segments, shown on top for $1, 2, \ldots, M$, and then determining a periodogram for *each* of them, with estimates $\hat{S}^{(1)}$, $\hat{S}^{(2)}$, through $\hat{S}^{(M)}$. The final periodogram $\hat{S}_k$ is then the average, per discrete frequency $k\Delta f$, with $k = 0, \ldots, \frac{N}{M} - 1$, of the M segment-wise periodograms, with $\Delta f = \frac{M}{T}$:

$$\hat{S}_k = \frac{1}{M} \sum_{m=1}^{M} \hat{S}_k^{(m)} \tag{14.3}$$

In addition to segment averaging, the Welch periodogram [15] features the use of a window in each segment, e.g., a Hann window (see Chapter 8) and overlapping segments. The Welch periodogram is available in Python through `scipy.signal.welch`. The confidence interval for the Welch periodogram, with non-overlapping segments, is given by (J.10).

EX 14.5

We continue Example 14.4 and demonstrate the Welch periodogram, for which the measurement record is divided into segments.

Solution The measurement record of $T = 100$ s at $f_s = 3$ Hz (Figure 14.6 at left), is divided into $M = 10$ segments of $\frac{N}{M} = 30$ samples each. Therefore, the measurement duration of each segment is 10 s, and hence the frequency step size in the graph at right, Figure 14.8, is $\Delta f = 0.1$ Hz.

The graph of the Welch periodogram shows that, compared to the default periodogram at left (for $T = 100$ s), the frequency resolution has become worse (larger step size), but the variance of the estimator has become smaller by a factor of M, thanks to the averaging.

The graph at right shows a much reduced variability in the periodogram estimates and also the confidence interval: the limits, shown in green and red, are much tighter.

In other words, the Welch periodogram is closer to the true PSD, represented by the dashed black curve.

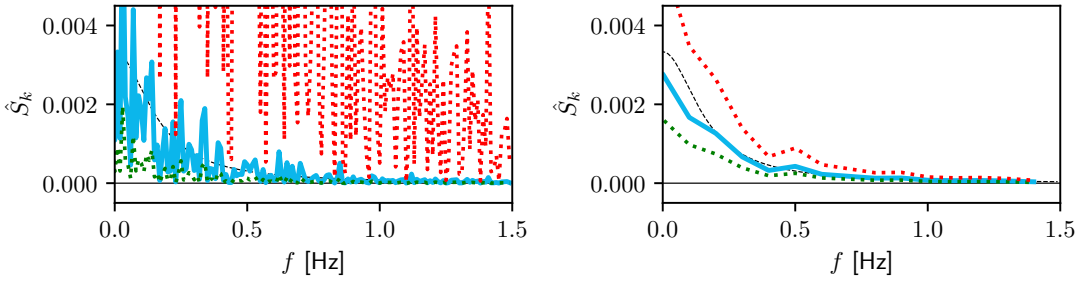


Figure 14.8: Filtered white noise signal in discrete time, $f_s = 3$ Hz, $T = 100$ s, with standard deviation $\sigma = 0.1$ and first-order Butterworth low-pass filter. At left, the standard periodogram with $\Delta f = 0.01$ Hz; at right, the Welch periodogram (measurement record divided into $M = 10$ segments, each of length $\frac{N}{M} = 30$ samples, with no overlap) with $\Delta f = 0.1$ Hz (in blue). The theoretical PSD is shown with the black, dashed curve. All PSDs are in [W/Hz]. The 95% confidence interval limits about the estimate are shown in green (lower limit) and red (upper limit).

We covered spectral density computations for random and deterministic signals in this and the previous chapter, assuming that the signal properties do not change over time. In Chapter 15 we abandon this assumption and discuss the spectrogram, allowing for analysis of non-stationary signals.

15

Spectrogram

The basic periodogram estimate was given as (14.2) in the previous chapter. A *non-stationary* signal is not suited as input to compute a periodogram straightaway, as the properties of the signal *change* over time. In this chapter, we present the spectrogram as a simple means to deal with non-stationary signals in practice.

Figure 15.1 shows how a spectrogram is computed. The diagram looks very similar to Figure 14.7, but the key difference is that here each segment is treated *separately*, whereas in Figure 14.7 the M segments were *averaged* to obtain the eventual spectral estimate. When using a spectrogram we assume that a non-stationary signal may still, by approximation, be assumed to be 'stationary over a short time span'. Per segment a periodogram is estimated, the results are stacked together and a time axis is added to the figure. Along the time axis we can see how the signal periodogram evolves as a function of time.

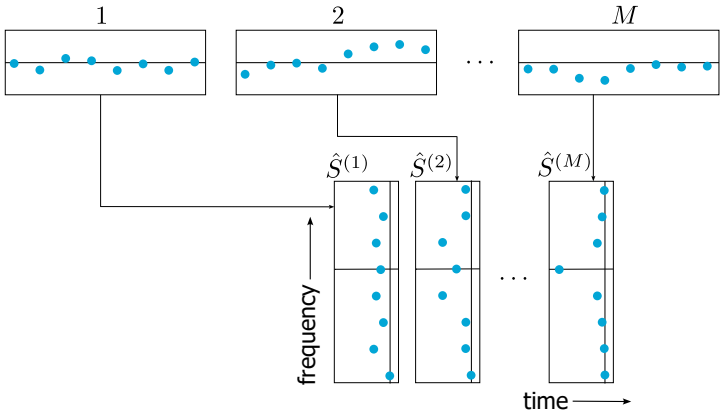


Figure 15.1: Diagram of computation of a spectrogram created by dividing the total signal measurement record into M segments, and computing a periodogram for each segment separately. The time and frequency indices are absent in these plots to prevent clutter.

The DFT frequency step size Δf depends on the measurement duration T: $\Delta f = \frac{1}{T}$ using the entire measurement duration. With M segments, for each segment the step size Δf equals $\frac{M}{T}$. This immediately demonstrates that the price to pay for improved frequency resolution (smaller M) is reduced *temporal* resolution (longer segments), and vice versa. For improved frequency resolution (a smaller step size Δf), each segment, and therefore the periodogram, represents a longer time span, resulting in less frequent 'updating' along the time axis.

In Figure 15.2 the spectrogram at left, with $M = 4$, shows a better frequency resolution (smaller frequency steps) than the one at right. The spectrogram at right, with a larger M,

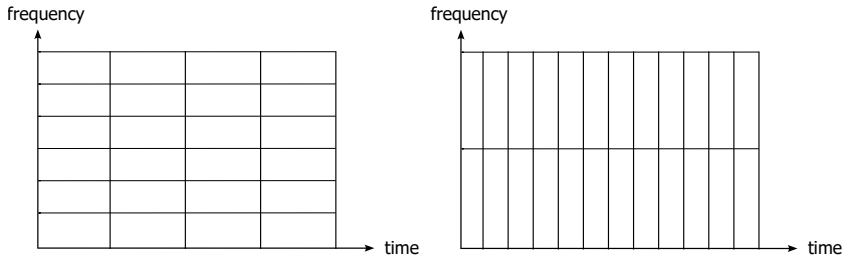


Figure 15.2: Trade-off between temporal and spectral resolution with the spectrogram. The total signal measurement record of duration T is divided into M segments, and a periodogram is computed for each segment separately. The segment duration (defining the temporal resolution) is $\frac{T}{M}$ and the frequency step size (defining the spectral resolution) is $\Delta f = \frac{M}{T}$. At left $M = 4$; at right $M = 12$.

here $M = 12$, has a better temporal resolution (shorter segments). Once the total signal measurement duration T is fixed, we cannot have a high resolution in both the time and frequency domains, and the segment length needs to satisfy the assumption or approximation of the signal being (quasi-) stationary.

EX 15.1

We consider the spectrogram for a signal measured by a vibration sensor. The sensor was fixed (tightly bolted) onto a four-cylinder combustion engine. A signal was measured for a duration of $T = 7$ seconds with the engine running, and the throttle was used to 'speed up' or 'rev' the engine.

Solution The measurement record is divided into $M = 7$ one-second segments; the frequency step $\Delta f = \frac{7}{7} = 1$ Hz. Figure 15.3 shows the resulting periodograms, computed for each segment, as 'traces'. No overlap was used, and a rectangular window.

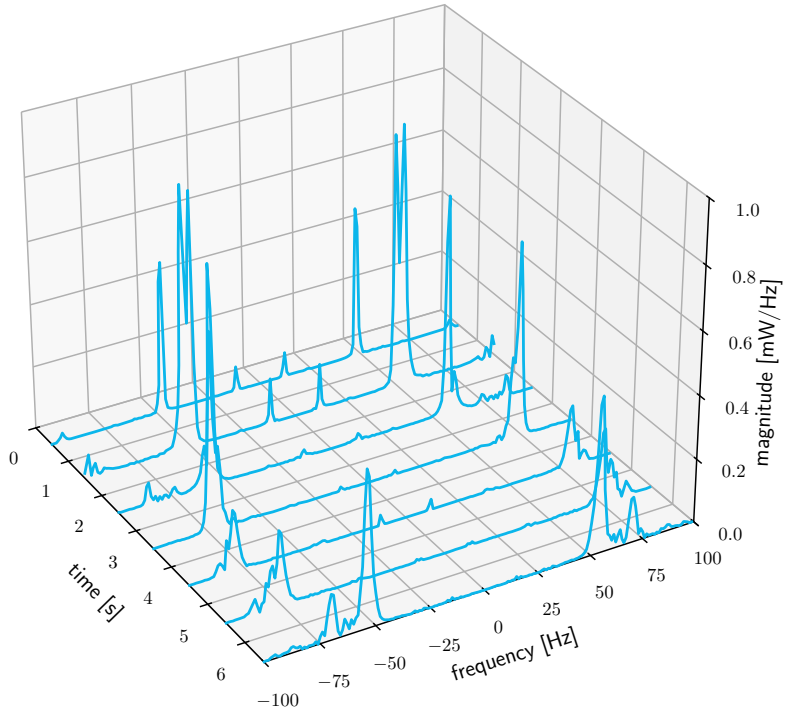


Figure 15.3: Spectrogram of a signal measured with a vibration sensor on a running combustion engine. The $M = 7$ periodograms are shown by means of traces, next to one other; the sampling frequency f_s was 200 Hz, the measurement time $T = 7$ s, and the segment duration $\frac{T}{M} = 1$ s.

A bird's eye view of Figure 15.3 is shown in Figure 15.4; the (double-sided) periodograms are shown in vertical direction, stacked next to each other along the time axis. The value of the PSD is shown by the shade of blue, see the color bar at right. This type of graph is commonly referred to as a 'heat map'. The PSD values are expressed in [dBW/Hz]; values smaller than -60 dBW/Hz have been removed to allow for a clear visualization of the peaks.

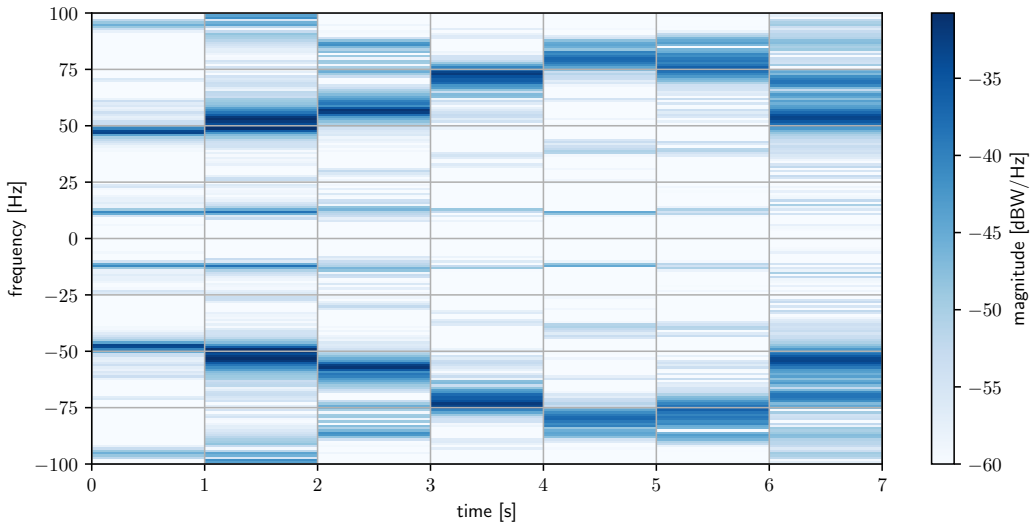


Figure 15.4: Spectrogram of a signal measured with a vibration sensor on a running combustion engine. A bird's eye view of the time-frequency area in Figure 15.3 is shown using a color bar for the values of the PSD (in [dBW/Hz]).

During the time span from 1 to 4 seconds, the engine was revved, so that the frequency of the peak in the spectrum increased, from around 50 Hz to over 75 Hz, and also the signal amplitude became larger (the larger power density is shown as darker blue). After 5 seconds the throttle was released again.

The complete double-sided spectrum is shown in Figures 15.3 and 15.4. In practice, we show only the part for positive frequencies, as the spectrum is symmetric.

Python offers the short-time FFT `scipy.signal.ShortTimeFFT`, which computes sequential FFTs (DFTs) by sliding a time window over the input signal. In this way, it computes the coefficients X_k for each segment, and the spectrogram (`spectrogram`), part of the above `ShortTimeFFT` class, is then readily obtained through (14.2). Segments may partially overlap and a window may be used per segment, other than just a rectangular one.

Now that the spectrogram has been discussed, the practical details of performing spectral estimation have been covered. In Part V, we briefly discuss *systems*, focusing in particular on the process of *filtering* signals.

V

Linear systems

16

Linear time-invariant systems

This chapter presents a brief introduction into the subject of systems in the context of signal processing and analysis. The basic concept of a system is illustrated in Figure 16.1.

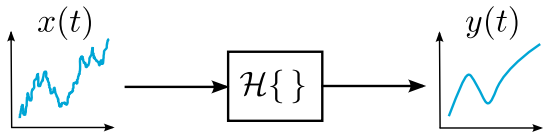


Figure 16.1: A system turns input signal $x(t)$ into output signal $y(t)$ through operation $\mathcal{H}$: $y(t) = \mathcal{H}\{x(t)\}$.

Some device, structure, or algorithm turns input signal $x(t)$ into output signal $y(t)$. A system can be considered as performing some 'operation' on $x(t)$, to produce $y(t)$:

$$y(t) = \mathcal{H}\{x(t)\} \tag{16.1}$$

First, some general system properties are introduced. The scope is then restricted to the class of linear time-invariant (LTI) systems, which have one input and one output, in continuous time. For these single-input single-output (SISO) LTI systems, the relation between input and output can be fully captured by the system impulse response function $h(t)$ which, when Fourier-transformed, is known as the system frequency response function $H(f)$. Plotting the magnitude and phase of $H(f)$ as a function of frequency yields the system's Bode plot, a common, useful graphical illustration of the LTI system's dynamic response.

16.1 System properties

System models mathematically describe the interaction of signals (the arrows in Figure 16.1) and systems (the box in Figure 16.1) and the relationships between *causes* (the system *inputs*) and *effects* (the system *outputs*) for a system. Systems can have many inputs and outputs, but we focus on SISO systems.

Modeling the operation or behavior of a system often starts by developing a differential equation (DE), e.g., based on physical laws that govern the system behavior. The resulting relationship can then be used to characterize the system, in a variety of ways:

1. Instantaneous or dynamic,
2. Time-invariant or time-varying,
3. Linear or nonlinear,

4. Causal or non-causal, and
5. Continuous-time or discrete-time.

We briefly discuss these properties, following [7]. Depending on the system characteristics, different mathematical techniques are available to analyze the system behavior.

16.1.1 Instantaneous and dynamic systems

A system for which the output is a function of the input at the present time t only is said to be *instantaneous* or *static*. When this is not the case, the system is called *dynamic*.

Instantaneous systems are modeled by relations such as:

$$y(t) = Ax(t) + B \quad \text{or} \quad y(t) = Ax(t) + Bx^3(t)$$

for constants A and $B \in \mathbb{R}$. We refer to instantaneous systems as 'memory-less' or 'zero-memory': their output does not depend on past (or future) input.

For *dynamic* systems, the output depends on past and/or future values of the input, generally in addition to present time. These systems 'have memory' and can in some cases be modeled by relationships such as:

$$a_p \frac{\mathrm{d}^p y(t)}{\mathrm{d}t^p} + a_{p-1} \frac{\mathrm{d}^{p-1} y(t)}{\mathrm{d}t^{p-1}} + \cdots + a_0 y(t) = b_q \frac{\mathrm{d}^q x(t)}{\mathrm{d}t^q} + b_{q-1} \frac{\mathrm{d}^{q-1} x(t)}{\mathrm{d}t^{q-1}} + \cdots + b_0 x(t) \quad (16.2)$$

where $a_p, \ldots, a_0, b_q, \ldots, b_0$ are constants $\in \mathbb{R}$ and non-zero initial conditions may apply. Whereas (16.2) describes a system governed by a linear, constant-coefficient, ordinary differential equation (ODE) of order p, this is more exception than rule. For real-life systems, the mathematical relationship between input and output is often more complicated.

The DE may also contain integral terms on input and output. The *order* of a system is defined as the highest derivative of output $y(t)$ when *all* integrals in the DE are removed through differentiation of that DE, see Example 16.2.

EX 16.1 Consider the system described by $y(t) = 2x(t^2)$. Is it instantaneous or dynamic? What is the order of this system?

Solution Testing the input-output relationship for various values of time t quickly shows that this system is dynamic. E.g., for $t = 2$ s we find $y(2) = 2x(4)$: the output at $t = 2$ s equals 2 times the input at $t = 4$ s. The order of the system is zero.

EX 16.2 Consider the system described by the DE:

$$\frac{\mathrm{d}y(t)}{\mathrm{d}t} + 2y(t) = \int_{-\infty}^{t} x(\tau)\,\mathrm{d}\tau$$

Is this system instantaneous or dynamic? What is the order of this system?

Solution This system is clearly dynamic, as the integral on the right-hand side shows that the output depends on all past values of the input. To find the order, we first need to get rid of all integrals, which is done by computing the time derivative of the DE:

$$\frac{\mathrm{d}^2 y(t)}{\mathrm{d}t^2} + 2\frac{\mathrm{d}y(t)}{\mathrm{d}t} = x(t) \quad \Rightarrow \quad \text{The order of the system is } 2.$$

16.1.2 Time-invariant and time-varying systems

A system is time-invariant, or *fixed*, if its input-output relationship does not change with time. Otherwise it is time-varying. In terms of (16.1), a system is time-invariant if and only if:

$$\boxed{\mathcal{H}\{x(t-\tau)\} = y(t-\tau)} \tag{16.3}$$

for any $x(t)$ and any $\tau \in \mathbb{R}$, and with the system considered to be 'at rest' (all derivatives of $y(t)$ are zero) before the input is applied. In other words, a system is time-invariant when, no matter at what time t the input is given, its output stays the same (though equally time-shifted as the input).

EX 16.3

Consider system I described by $\frac{\mathrm{d}y(t)}{\mathrm{d}t} + t^2 y(t) = 5x(t)$ and system II described by $\frac{\mathrm{d}y(t)}{\mathrm{d}t} + y(t)y(t) = 5x(t)$. Which of these two systems is time-invariant?

Solution System I is time-varying. When applying an input $x(t)$ at $t = 0$ s, the system dynamics at that time equal $\frac{\mathrm{d}y(t)}{\mathrm{d}t} = 5x(t)$. When applying the same input at $t = 4$ s, the system dynamics at that time equal $\frac{\mathrm{d}y(t)}{\mathrm{d}t} + 16y(t) = 5x(t)$. Clearly, the system dynamics *change* over time and the system will respond *differently* to the same input signal. System II is time-invariant, since its DE coefficients are all constants.

16.1.3 Linear and nonlinear systems

For a linear system, the *superposition principle* holds:

$$\mathcal{H}\{\alpha_1 x_1(t) + \alpha_2 x_2(t)\} = \alpha_1 \mathcal{H}\{x_1(t)\} + \alpha_2 \mathcal{H}\{x_2(t)\} \tag{16.4}$$

for any inputs $x_1(t)$ and $x_2(t)$ and any constants α_1 and $\alpha_2 \in \mathbb{R}$.

When the response of the system to $x_1(t)$ equals $y_1(t)$ and the response of the system to $x_2(t)$ equals $y_2(t)$, the superposition principle is stated as:

$$\boxed{\mathcal{H}\{\alpha_1 x_1(t) + \alpha_2 x_2(t)\} = \alpha_1 y_1(t) + \alpha_2 y_2(t)} \tag{16.5}$$

When (16.5) does not hold, the system is said to be *nonlinear*.

EX 16.4

Prove that the (time-varying) system described by the DE $\frac{\mathrm{d}y(t)}{\mathrm{d}t} + ty(t) = x(t)$ is linear.

Solution The response $y_1(t)$ to $x_1(t)$ satisfies $\frac{\mathrm{d}y_1(t)}{\mathrm{d}t} + ty_1(t) = x_1(t)$ and the response $y_2(t)$ to $x_2(t)$ satisfies $\frac{\mathrm{d}y_2(t)}{\mathrm{d}t} + ty_2(t) = x_2(t)$. Multiplying by arbitrary constants α_1 and $\alpha_2 \in \mathbb{R}$ and then adding them together (omit (t) for brevity) yields:

$$\alpha_1 \frac{\mathrm{d}y_1}{\mathrm{d}t} + \alpha_2 \frac{\mathrm{d}y_2}{\mathrm{d}t} + \alpha_1 t y_1 + \alpha_2 t y_2 = \alpha_1 x_1 + \alpha_2 x_2$$

$$\Rightarrow \frac{\mathrm{d}}{\mathrm{d}t}(\alpha_1 y_1 + \alpha_2 y_2) + t(\alpha_1 y_1 + \alpha_2 y_2) = (\alpha_1 x_1 + \alpha_2 x_2)$$

Thus, the response to the input $\alpha_1 x_1(t) + \alpha_2 x_2(t)$ equals $\alpha_1 y_1(t) + \alpha_2 y_2(t)$. The superposition principle (16.5) holds, and therefore the system is linear.

16.1.4 Causal and non-causal systems

A system is *causal* if its response to an input does not depend on *future* values of that input.

An easy way to determine whether a system is causal is to note that $y(t)$ cannot do anything in anticipation of $x(t)$. It is *non-anticipatory*: the system output *reacts* to the input. All systems that represent real-world phenomena, such as (combinations of) mechanical, biological, and electrical systems, are causal. Chapter 18 on filters discusses examples of causal and non-causal filters. A system must be causal to be applicable in *real time*.

As an example, the input-output relation $y(t) = x(t+3)$ describes a non-causal system, since the output depends on future input.

16.1.5 Continuous-time and discrete-time systems

A *continuous-time* system operates with continuous-time signals, and a *discrete-time* system with discrete-time signals. Continuous-time systems and discrete-time systems are often governed by, respectively, differential equations (such as (16.2)) and difference equations. Chapter 18 discusses examples of both types of systems.

16.1.6 Linear time-invariant systems

Combinations of system properties yield particular system classes, with the *linear time-invariant* (LTI) system as a common example. As the definition says, an LTI system is *linear* (the superposition principle holds) and *time-invariant* (its input-output relation does not change over time). LTI systems can be continuous-time or discrete-time, and may be instantaneous or dynamic. In the remainder of this chapter, the discussion is restricted to LTI systems in continuous time, often characterized by (16.2). LTI systems are used extensively in signal and system analysis for two reasons: mathematical techniques to study these systems are well established, e.g., the Fourier and Laplace transforms, and in real life many systems are *by approximation* linear and time-invariant.

16.2 System response

Once the differential equation is given, e.g., in the form of (16.2), how do we find the response $y(t)$ of the LTI system, given an input $x(t)$? Several ways exist to answer this question.

The first approach is to solve the DE through calculus. For a simple example, this approach is demonstrated in Appendix M for the system discussed in Example 16.5.

The second approach is through superposition. Without proof we state that, for LTI systems, output $y(t)$ is found as the *convolution* of input $x(t)$ with the so-called *impulse response function* $h(t)$ of the system:

$$y(t) = x(t) * h(t) = \int_{-\infty}^{\infty} x(\tau)h(t-\tau)\,\mathrm{d}\tau \tag{16.6}$$

This convolution integral is also known as a *superposition integral* (and as Duhamel's integral[1] in the context of structural dynamics). For the DE in Appendix M, the convolution expression above yields (M.8), with the impulse response function $h(t) = \alpha e^{-\alpha t}u(t)$.

The third approach is using the Fourier transform. Recall that the Fourier transform of a convolution of two signals in the time domain equals the multiplication of the Fourier transforms of these two signals, (6.8). Fourier-transforming (16.6) yields:

$$Y(f) = X(f)H(f) \tag{16.7}$$

[1] After the French mathematician J.-M.C. Duhamel (1797-1872).

with $H(f)$ the Fourier transform of the impulse response function $h(t)$:

$$\boxed{H(f) = \mathcal{F}\{h(t)\}} \tag{16.8}$$

$H(f)$ is referred to as the *frequency response function* (FRF) of the LTI system. The response of an LTI system is completely characterized by either $h(t)$ or $H(f)$; these functions play an essential role in the analysis of systems.

Regarding the units of $H(f)$ and $h(t)$, the following can be stated. When input signal $x(t)$ has unit [in] and output signal $y(t)$ unit [out], then their Fourier transforms $X(f)$ and $Y(f)$ have units [in/Hz] and [out/Hz], respectively. From (16.7) it follows that $H(f) = \frac{Y(f)}{X(f)}$, which means that $H(f)$ has [out/in] as its unit. $H(f)$ is the Fourier transform of $h(t)$, so the unit of the latter is [out/in/second]. In Chapter 17 some practical examples are given.

In forward modeling, with a given input $x(t)$ and system impulse response function $h(t)$, after applying the Fourier transform to both, output $Y(f)$ can be computed through (16.7), then transformed back to the time domain to obtain $y(t)$: $y(t) = \mathcal{F}^{-1}\{Y(f)\}$.

In case we first need to find impulse response function $h(t)$, e.g., from the differential equation, this DE is first Fourier-transformed to the frequency domain. Through the relation $H(f) = \frac{Y(f)}{X(f)}$, the frequency response function $H(f)$ is found, and can be inverse Fourier-transformed to the time domain to obtain $h(t)$. This is demonstrated in the next example.

EX 16.5

For a system, described by the first-order ODE (from Appendix M):

$$\frac{1}{\alpha}\frac{\mathrm{d}y(t)}{\mathrm{d}t} + y(t) = x(t) \tag{16.9}$$

with $\alpha > 0$, find the impulse response function $h(t)$ by transforming the equation into the frequency domain.

Solution With the Fourier transforms of, respectively, input $x(t)$ and output $y(t)$ denoted as $X(f)$ and $Y(f)$, Fourier-transforming the above differential equation is easily achieved by using the differentiation Fourier transform theorem, (6.10):

$$\frac{1}{\alpha}(j2\pi f)\,Y(f) + Y(f) = X(f)$$

which can be rewritten as:

$$\left(\frac{\alpha + j2\pi f}{\alpha}\right) Y(f) = X(f)$$

The frequency response function equals:

$$H(f) = \frac{Y(f)}{X(f)} = \frac{\alpha}{\alpha + j2\pi f} \tag{16.10}$$

With Example 5.3 the impulse response function $h(t)$ is found as:

$$h(t) = \alpha e^{-\alpha t}u(t) \quad \text{for} \quad \alpha > 0 \tag{16.11}$$

This function is identical to the expression for $h(t)$ derived in Appendix M. It was shown in Figure 7.7 at right, for $\alpha = 1$ and $\alpha = 3$.

In an experimental approach, when no differential equation of the system is available, we can attempt to recover the system impulse response function $h(t)$ by measuring both output and input signals (or by measuring the output to a known input), Fourier-transforming both, performing a so-called *de-convolution*, $H(f) = \frac{Y(f)}{X(f)}$, and inverse Fourier-transforming $H(f)$. This procedure is easily defined in theory, but not straightforward to apply in practice.

16.3 Impulse response function

Function $h(t)$ is the impulse response function of the system, and for an LTI system it fully describes the input-output relation of that system.

To explain this name of $h(t)$, consider the system response $y(t)$ to the Dirac delta function, an impulse, at $t = 0$, i.e., $x(t) = \delta(t)$. The convolution (16.6) becomes:

$$y(t) = \int_{-\infty}^{\infty} \delta(\tau) h(t - \tau)\, \mathrm{d}\tau = h(t)$$

where we used the sifting property of the Dirac delta function, (B.12). The result implies that function $h(t)$ is the response of the system to a Dirac delta impulse as input, applied at $t = 0$; hence the name *impulse response* function (of the system).

Using the impulse response function $h(t)$, we can easily determine a system's causality: $h(t)$ should be *zero* for $t < 0$, because $h(t)$ equals the system response to an impulse input *at* $t = 0$. For instance, the ideal reconstruction filter introduced in Chapter 10 is a non-causal system; see Figure 10.2 (left). The ZOH reconstruction filter, Figure 10.9 (left), is causal.

EX 16.6

With the impulse response function of a system given as $h(t) = e^{-t}u(t)$, compute the output $y(t)$ to the unit step function as input, i.e., to $x(t) = u(t)$.

Solution This problem has been solved in Example 7.4 for $\alpha = 1$. Figure 7.8 shows that, from $t = 0$ onwards the system gradually and asymptotically adjusts to the new state, i.e., the unit step function $x(t) = u(t)$ applied at $t = 0$. The system is causal.

16.4 Frequency response function

Function $H(f)$ is the frequency response function of the system, the Fourier transform of the impulse response function $h(t)$. These functions form a Fourier transform pair: $h(t) \overset{\mathcal{F}}{\leftrightarrow} H(f)$. The frequency response function plays a pivotal role in the description of the LTI system response to a harmonic input.

We start with a cosine with frequency f_0, initial phase φ, and amplitude A as input:

$$x(t) = A\cos(2\pi f_0 t + \varphi) \tag{16.12}$$

and rewrite this using Euler's formula (C.1):

$$x(t) = \frac{A}{2}\left(e^{j(2\pi f_0 t + \varphi)} + e^{-j(2\pi f_0 t + \varphi)}\right)$$

Compute the output through convolution, (16.6), using the commutative property, (7.3):

$$y(t) = \int_{-\infty}^{\infty} \frac{A}{2}\left(e^{j(2\pi f_0 (t-\tau) + \varphi)} + e^{-j(2\pi f_0 (t-\tau) + \varphi)}\right) h(\tau)\, \mathrm{d}\tau$$

which yields:

$$y(t) = \frac{A}{2} e^{j2\pi f_0 t} e^{j\varphi} \underbrace{\int_{-\infty}^{\infty} e^{-j2\pi f_0 \tau} h(\tau)\, \mathrm{d}\tau}_{= H(f_0)} + \frac{A}{2} e^{-j2\pi f_0 t} e^{-j\varphi} \underbrace{\int_{-\infty}^{\infty} e^{j2\pi f_0 \tau} h(\tau)\, \mathrm{d}\tau}_{= H(-f_0)} \tag{16.13}$$

obtaining the Fourier transform (5.1) of $h(t)$, once evaluated at frequency f_0 and once at $-f_0$.

Next, we denote the frequency response function $H(f)$ in polar form, (5.5), $H(f) = |H(f)|e^{j\theta(f)}$ with modulus $|H(f)|$ and phase $\theta(f) = \arg H(f)$ (also written as $\angle H(f)$). If impulse response $h(t)$ is real-valued (as in the case of any physically realizable system), the magnitude is even $|H(f)| = |H(-f)|$ and the phase is odd, $\theta(f) = -\theta(-f)$. Rewriting (16.13):

$$y(t) = \frac{A}{2} e^{j2\pi f_0 t} e^{j\varphi} \, |H(f_0)| e^{j\theta(f_0)} \; + \; \frac{A}{2} e^{-j2\pi f_0 t} e^{-j\varphi} \, |H(f_0)| e^{-j\theta(f_0)}$$

and rearranging terms yields:

$$y(t) = \frac{A}{2} |H(f_0)| \left(e^{j(2\pi f_0 t + \varphi + \theta(f_0))} \; + \; e^{-j(2\pi f_0 t + \varphi + \theta(f_0))} \right)$$

which, with Euler's formula, (C.1), becomes:

$$y(t) = A\, |H(f_0)| \, \cos\left(2\pi f_0 t + \varphi + \theta(f_0)\right) \tag{16.14}$$

We arrive at an important conclusion, namely that the (steady-state) response of an LTI system to a harmonic on the input is a harmonic with the *same* frequency f_0. With input (16.12), the amplitude changes from A to $A\,|H(f_0)|$ and the phase from φ to $\varphi + \theta(f_0)$ in output (16.14).

16.4.1 System response: magnitude

In the frequency domain, the LTI system response follows as in (16.7). In Appendix C we have the product of two complex vectors as $XZ = RSe^{j(\theta+\varphi)}$, and hence $|XZ| = |X||Z| = RS$, so that in terms of magnitude or amplitude spectrum we have:

$$|Y(f)| = |H(f)X(f)| = |H(f)|\,|X(f)| \tag{16.15}$$

Using this relationship, it is straightforward to relate the energy spectral density functions of the input and output signals:

$$\underbrace{|Y(f)|^2}_{= G_y(f)} = |H(f)|^2 \underbrace{|X(f)|^2}_{= G_x(f)} \tag{16.16}$$

A similar relationship holds for the power spectral density functions of the input and output signals: $S_y(f) = |H(f)|^2 S_x(f)$, see [17].

16.4.2 System response: phase

For convenience and clarity we set initial phase φ to zero in (16.12). The output (16.14) reads:

$$y(t) \;=\; A\,|H(f_0)|\,\cos\left(2\pi f_0 t + \theta(f_0)\right) \;=\; A\,|H(f_0)|\,\cos\left(2\pi f_0 \left(t + \frac{\theta(f_0)}{2\pi f_0}\right)\right)$$

clearly telling us that passing a harmonic through an LTI system will cause a phase shift $\theta(f_0)$. A phase shift can also be interpreted as a *time* shift, (2.3): $t_0 = \frac{\theta(f_0)}{2\pi f_0}$. When $\theta(f_0) > 0$, $t_0 > 0$, the output will be *ahead* of the input, a time *advance*. Vice versa, when $\theta(f_0) < 0$, $t_0 < 0$, the output will be *lagging behind* the input, a time *delay*.

In system analysis we distinguish three types of phase responses:

1. Zero phase: $\theta(f_0) = 0$, for any f_0. This is true when $H(f)$ is real and positive. Hence, $h(t)$ is required to be an even function, which is only true for non-causal systems.
2. Linear phase: $\theta(f_0) = af_0$ with constant $a \in \mathbb{R}$: the phase shift depends linearly on frequency f_0; as a result, time shift $\frac{\theta(f_0)}{2\pi f_0} = \frac{a}{2\pi}$ is constant and the same for all frequencies.
3. Nonlinear phase: any other behavior.

In a system with a nonlinear phase response, signal components at different frequencies may get time-delayed by different amounts, and as a result the output signal may be distorted.

16.4.3 Bode plot

When the frequency response function of the LTI system is known, a common and practical way to illustrate the system dynamic response is using a Bode plot[2]. In a Bode plot, both the magnitude response $|H(f)|$ and phase response $\angle H(f)$ are shown as a function of frequency f, with the magnitude at the top and the phase at the bottom. Only positive frequencies are shown, $f > 0$. The magnitude response is an even function of frequency. For the phase response we should remember that this is an *odd* function of frequency.

A logarithmic frequency axis (the horizontal axis) is often used. Magnitude is also typically shown on a logarithmic axis. Phase is shown on a linear axis in [rad] or [deg]. Figure 16.2 shows the Bode plot for the LTI system of Example 16.5, for two values of α.

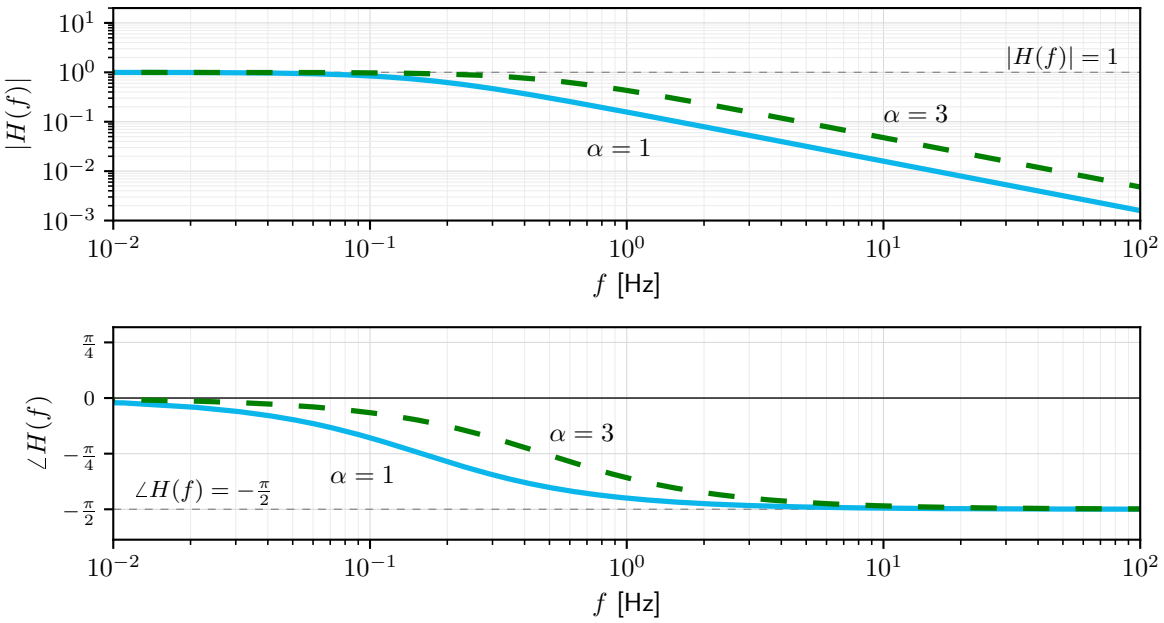


Figure 16.2: Bode plot of the FRF of the LTI system characterized by DE (16.9) from Example 16.5 for two values of α, namely $\alpha = 1$ (blue curves) and 3 (green dashed curves); phase is shown in [rad].

The main application of LTI systems in this book is filtering, which is the subject of Chapter 18. But before discussing filters, Chapter 17 will first explain what typical measurement set-ups look like, including the (dynamic) characteristics of sensors.

[2]Named after H.W. Bode (1905-1981), an American engineer of Dutch ancestry.

17

Measuring signals

Setting up an experiment and successfully collecting the intended measurement signal is an art. It can be a serious challenge, requiring knowledge, insight and (practical) skills. As this subject deserves a book on its own (and a lot of practice), we provide a concise introduction, in continuous time, just touching upon the most important aspects. Throughout this chapter we use an accelerometer as an example.

17.1 Sensor

A sensor is a device which delivers an output signal, most commonly an electrical voltage, proportional to a physical quantity of interest, i.e., the phenomenon to be measured. A sensor is a *transducer*, which, more generally, is a device that converts a signal of one type of energy into another. For instance, the (classic) accelerometer in Figure 17.1 at left turns physical acceleration (motion) into an electric voltage.

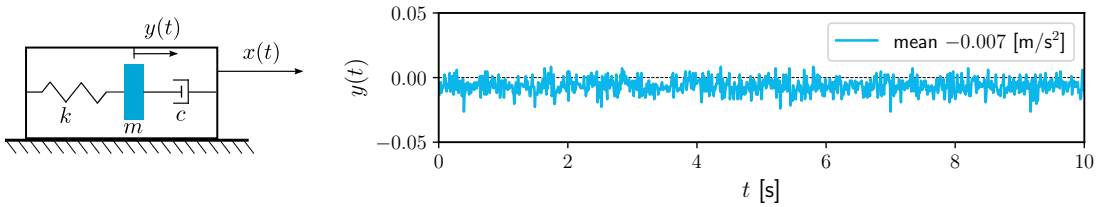


Figure 17.1: Left: basic outline of an accelerometer; $x(t)$ represents the motion of the housing and $y(t)$ is the displacement of the proof mass relative to the housing; the absolute motion of the proof mass is $x(t) + y(t)$. Right: accelerometer measurements (in [m/s^2]) obtained with a smartphone, lying still on a flat and level concrete floor. The measurements, on one of the horizontal axes, were obtained with a STMicroelectronics K6DS3TR sensor at $f_s = 100$ Hz, for a duration of 10 seconds.

Within its housing (rectangle), a proof mass (in blue, mass m) is suspended by a spring (spring constant k) and connected to a dashpot damper (damper constant c). The accelerometer is put on a horizontal surface, aligned with its input-output axis of interest. It senses external motion, (input) acceleration $\ddot{x}(t)$ of the housing in [m/s^2], through measuring the *displacement* $y(t)$ of the proof mass with respect to the housing (along one axis), a so-called single degree-of-freedom system, and outputs a proportional electric voltage [V]. The proportionality is the *sensitivity* of the sensor: the ratio of change in sensor output (voltage) and the change in sensor input (acceleration). In case of the accelerometer, the sensitivity is expressed in [V/m/s^2]. Ideally, the sensor input-output shows a linear relation, materialized by just a single scale factor.

The sensitivity is one of the key parameters of a sensor determining the measurement precision. With a digital signal, *resolution* also plays a role, as the sensor outputs a discrete signal value (with a certain 'step size' Δx); see Appendix D on quantization. Another key parameter of a sensor is the *dynamic range* D, indicating to which values of the input the sensor sensitivity is restricted, i.e., which values of the physical quantity can be 'handled properly' by the sensor. Beyond this measurement range, the sensor is saturated, values may get 'clipped' and the measured signal may become seriously distorted.

Figure 17.1 at right shows 10 seconds of accelerator measurements at a sampling rate of $f_s = 100$ Hz, collected by a three-axis accelerometer sensor in a smartphone, lying still and flat on a level and solid concrete floor, for one of the horizontal axes. The mean is -0.007 m/s^2; close to zero, as expected, but not exactly zero.

There is a small systematic offset, a *bias*. This is common in practice, and the experimenter needs to be aware of it. In this case it could be caused by a slight misalignment of the sensor's axes (in its housing or the smartphone casing, such that it is not perfectly horizontal, and actually senses a tiny component of gravity). In addition, it is not uncommon to see, over long time spans, some slow trend or *drift* in the observed signal, for instance due to temperature effects in the sensor. A calibration may remedy a bias as well as drift effects.

Finally, random fluctuations in the output signal occur, a phenomenon which we call *noise*, for instance caused by random motion of electrons in electrical circuitry (thermal noise). The measured acceleration signal in Figure 17.1 is noisy; the empirical standard deviation is 0.005 m/s^2 (actual minor background vibrations in the set-up cannot be entirely ruled out, but likely the measurement noise is dominant in this experiment). Note that in practice, due to noise, the precision of the sensor is (considerably) worse than its resolution – in the accelerometer example the difference is about a factor of 5.

17.2 Frequency response

A sensor may be effectively sensitive only to input signal values in a certain range (dynamic range D, as mentioned above). The sensitivity may also depend on the *frequency* of the input. Consequently, the frequency response function of the sensor is an important element of the sensor specification. In this section we study the FRF of an (ideal) accelerometer.

We start from the theoretical model of motion for an ideal accelerometer, as shown in Figure 17.1 at left, on a perfectly horizontal surface. In this section and the next one, as is common in dynamics, we use angular frequency $\omega = 2\pi f$, rather than linear frequency f. The motion of the mass-spring-damper system is described by a second-order ODE:

$$\ddot{y}(t) + 2\zeta\omega_0\,\dot{y}(t) + \omega_0^2\,y(t) \;=\; -\ddot{x}(t) \tag{17.1}$$

with both the positive x and y axis pointing right. The sign on the right-hand side is negative, because an acceleration of the housing to the right causes the proof mass to move to the *left* relative to its housing. In (17.1), $\zeta = \frac{c}{2\sqrt{mk}}$ is the (dimensionless) damping ratio and $\omega_0 = \sqrt{\frac{k}{m}}$ the undamped natural angular frequency in [rad/s]. Parameter m is the mass in [kg], c the damper constant in [kg/s] (= [Ns/m]), and k the spring constant in [N/m].

As explained in Chapter 16, this linear, constant coefficient differential equation can be solved by transformation to the frequency domain. We are interested in the relation of *displacement* output $y(t)$ and *acceleration* input $\ddot{x}(t)$. Using the differentiation Fourier transform theorem, (6.10), differential equation (17.1) can be written as:

$$(j\omega)^2 Y(\omega) + 2\zeta\omega_0\,(j\omega)Y(\omega) + \omega_0^2\,Y(\omega) = -\ddot{X}(\omega) \tag{17.2}$$

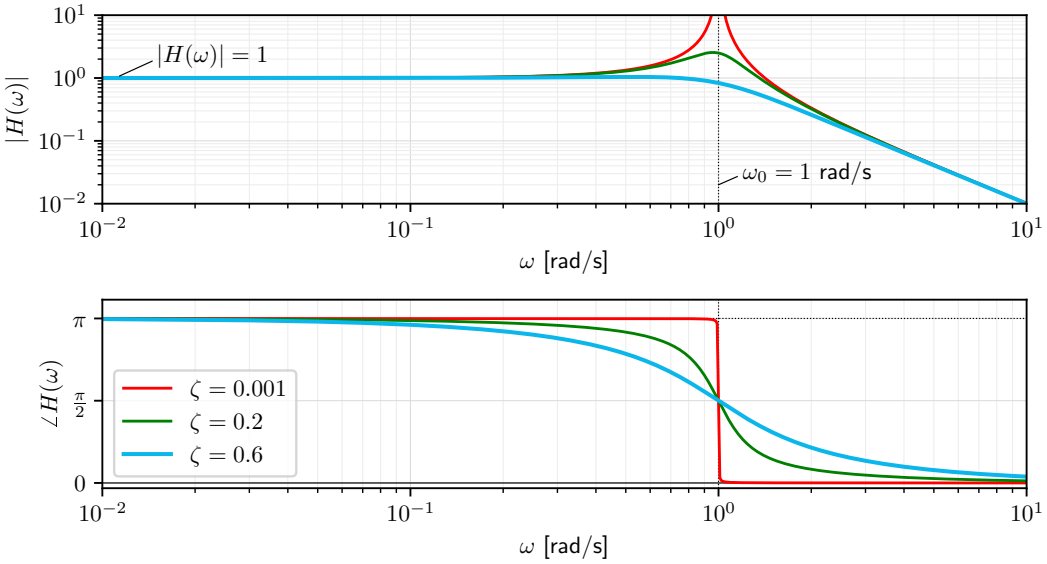


Figure 17.2: Bode plot of FRF $H(\omega)$ of the accelerometer of Figure 17.1 (left), with magnitude spectrum $|H(\omega)|$ (top) and phase spectrum $\angle H(\omega)$ (bottom, in [rad]), for three values of damping ratio ζ; the undamped natural frequency is set to $\omega_0 = 10^0 = 1$ rad/s.

with $Y(\omega)$ the Fourier transform of $y(t)$, and $\ddot{X}(\omega)$ of $\ddot{x}(t)$. This is an algebraic equation in ω, from which the frequency response function can be obtained:

$$H(\omega) = \frac{Y(\omega)}{\ddot{X}(\omega)} = \frac{1}{\omega^2 - \omega_0^2 - 2\zeta\omega_0(j\omega)} \tag{17.3}$$

The (ideal) accelerometer is a linear time-invariant system, with order 2. The modulus and argument of $H(\omega)$ are shown in Figure 17.2, a Bode plot, for three values of damping ratio ζ and $\omega_0 = 1$ rad/s. We see that, with extremely low damping ($\zeta = 0.001$, shown in red), and subject to input acceleration with frequency $\omega = \omega_0$ rad/s (see (16.14)) the accelerometer proof mass gets into resonance at that frequency. In practice, the resonance frequency of an accelerometer is in the order of tens of kHz (when expressed in linear frequency f).

For frequencies ω less than $\omega_0 = 1$ rad/s (in Figure 17.2 this is shown to the left of the vertical dotted line), $|H(\omega)|$ has an ideal and flat behavior: $|H(\omega)| \approx 1$. When $\omega \to 0$, then $|H(\omega)| \to \frac{1}{\omega_0^2}$ ($= 1$ in this example). The amplitude attenuation or amplification is the same for frequencies from zero up to $\approx 0.3\omega_0$, which means that for that range of frequencies the accelerometer sensitivity is independent of frequency. This is the 'useful' frequency application range of the accelerometer, i.e., its *bandwidth*, as within this range we get a proportional input-output relation between acceleration and displacement, with a small phase lag $\angle H(\omega) \approx \pi$. This phase angle π is because of the presence of a minus sign in (17.1); otherwise the phase angle would be zero.

17.3 Impulse response

Frequency response function (17.3) can be split into two parts:

$$H(\omega) = \frac{\frac{1}{2j\omega_d}}{\zeta\omega_0 + j\omega_d + j\omega} - \frac{\frac{1}{2j\omega_d}}{\zeta\omega_0 - j\omega_d + j\omega}$$

with the damped natural frequency $\omega_d = \omega_0\sqrt{1-\zeta^2}$, and with $0 < \zeta < 1$. Fourier transform (5.13), with $\alpha > 0$, can be shown to hold also for complex-valued constant α; working with

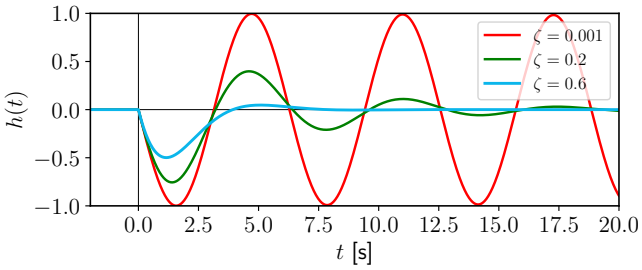


Figure 17.3: Accelerometer impulse response function $h(t)$, with $\omega_0 = 1$ rad/s, for three values of ζ.

angular frequency ω, we have $e^{-\alpha t}u(t) \overset{\mathcal{F}}{\leftrightarrow} \frac{1}{\alpha+j\omega}$. Applying this twice (and using the linearity property of the Fourier transform (6.1)) leads to:

$$h(t) = -\left\{\frac{1}{2j\omega_d}e^{-\zeta\omega_0 t}e^{j\omega_d t}\,u(t) - \frac{1}{2j\omega_d}e^{-\zeta\omega_0 t}e^{-j\omega_d t}\,u(t)\right\}$$

Using Euler's formula, (C.1), we obtain for the impulse response function:

$$h(t) = -\frac{1}{\omega_d}e^{-\zeta\omega_0 t}\sin(\omega_d t)\,u(t) \tag{17.4}$$

a damped sinusoid, see Figure 17.3. The accelerometer is a causal system: $h(t) = 0$ for $t \leq 0$. The impulse response shows the (in this case horizontal) motion of the proof mass after applying an impulse at $t = 0$. When the accelerometer is accelerated to the right, i.e., $\ddot{x}(t) > 0$, its proof mass will initially move *left* with respect to the housing: $y(t) < 0$.

17.4 Example

In this example we compare two accelerometers, with the same damping ratio $\zeta = 0.6$, but different undamped natural frequencies, respectively, $\omega_0 = 1$ rad/s (in blue) and $\omega_0 = 10$ rad/s (in green); see the Bode plot of the FRFs in Figure 17.4. To facilitate the comparison, the FRFs have been multiplied by ω_0^2, that is:

$$H(\omega) = \frac{Y(\omega)}{\ddot{X}(\omega)} = \frac{\omega_0^2}{\omega^2 - \omega_0^2 - 2\zeta\omega_0(j\omega)}$$

This multiplication can be considered as applying a scale factor to convert proof mass displacement $y(t)$ into an output voltage, scaling the sensor proportionality. Both accelerometers then have unit modulus, $|H(\omega)| = 1$, at $\omega = 0$, allowing for a straight comparison.

We apply to both sensors either a sinusoidal acceleration input signal $\ddot{x}_1(t) = \cos(\omega_1 t)$ with $\omega_1 = 0.2$ rad/s, or a sinusoidal input signal $\ddot{x}_2(t) = \cos(\omega_2 t)$, with $\omega_2 = 2$ rad/s. Figure 17.5 shows the *steady-state* responses $y_{ss}(t)$ of the two accelerometers to the input signals $\ddot{x}_1(t)$ (red, dashed) at the top, and $\ddot{x}_2(t)$ (orange, dashed) at the bottom. For convenient visualization, the negative of the responses is shown because of the minus sign in (17.1).

Starting with the low-frequency input signal $\ddot{x}_1(t)$, Figure 17.5 (at top) shows that both accelerometers, in steady state, provide an output with the same amplitude as the input. While for accelerometer 'green' there is no discernible time delay, the output of accelerometer 'blue' is lagging behind the input by approximately 1.4 s. These findings can be easily verified with the Bode plot in Figure 17.4. At frequency $\omega_1 = 0.2$ rad/s (vertical dashed line in red), the FRF magnitude of both accelerometers is $|H(\omega)| \approx 1$. This means the output in terms of amplitude matches the input exactly. At this frequency, the FRF phase $\angle H(\omega)$ of accelerometer 'green'

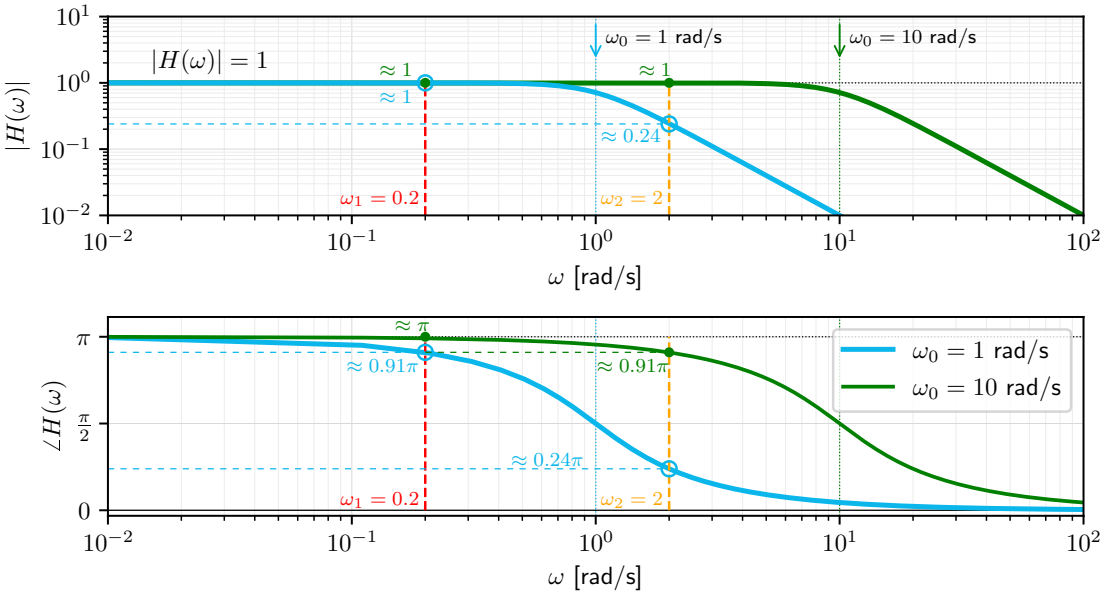


Figure 17.4: Bode plot of two accelerometers, with $\omega_0 = 1$ rad/s in blue and $\omega_0 = 10$ rad/s in green. The vertical dashed lines show the two input signal frequencies, $\omega_1 = 0.2$ rad/s (red) and $\omega_2 = 2$ rad/s (orange).

is $\approx \pi$, the FRF phase of accelerometer 'blue' is $\approx 0.91\pi$. While the output of accelerometer 'green' has a negligible delay, the output of accelerometer 'blue' will *lag behind* the input with a phase lag of $\pi - 0.91\pi = 0.09\pi$, equivalent to a time delay of $\frac{0.09\pi}{0.2} = 1.4$ s.

For the high-frequency input signal, $\ddot{x}_2(t)$, Figure 17.5 (at bottom) shows that, in steady state, only accelerometer 'green' provides an output with the same amplitude as the input signal. The output of accelerometer 'blue' is much smaller in amplitude. Whereas for accelerometer 'green' there is no discernible time delay, the output of accelerometer 'blue' is lagging behind the input considerably. Again, these responses can be directly obtained from the Bode plot, Figure 17.4, at the dashed line in orange. At frequency $\omega_2 = 2$ rad/s, the FRF magnitude of accelerometer 'green' $|H(\omega)| \approx 1$, the FRF magnitude of accelerometer 'blue' is ≈ 0.24, explaining the amplitude of its output. At this frequency, the FRF phase $\angle H(\omega)$ of accelerometer 'green' is $\approx 0.91\pi$, meaning a phase lag of 0.09π which corresponds to a time delay of $\frac{0.09\pi}{2} = 0.14$ s. The FRF phase of accelerometer 'blue' equals $\approx 0.24\pi$, resulting in a phase lag of 0.76π, equivalent to a time delay of $\frac{0.76\pi}{2} = 1.19$ s.

The Bode plot allows the amplitude and phase of the *steady-state* responses to sinusoidal input signals to be easily obtained. It immediately shows that accelerometer 'green' is superior to accelerometer 'blue': its useful range of frequencies (bandwidth) is much larger.

17.5 Sensor and object

This section serves to create awareness of the fact that in practice a sensor cannot function in ideal, isolated conditions. On the contrary: the sensor is attached or fixed to an object or structure of interest. The sensor actually becomes *part* of the object or structure.

This object or structure can be considered a system as well, that is, upon an input signal (a certain load or vibration), it produces a response or output signal. The sensor is also a (small) system, with a certain characteristic relationship between input and output (including the effects of its mounting on the structure[1]). The interplay of structure and sensor is illustrated

[1]Ideally, mounting the sensor on the structure does not affect the structure's dynamic response in any way.

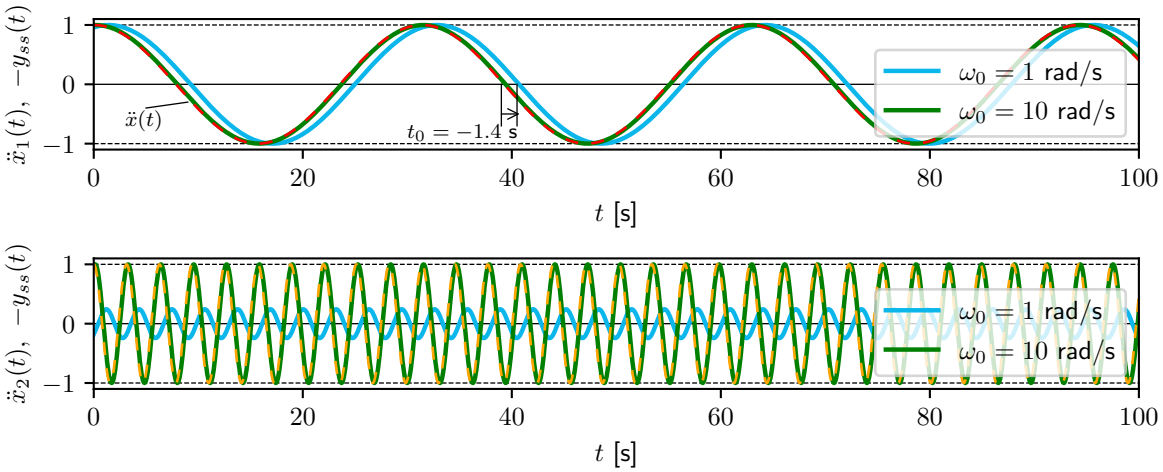


Figure 17.5: Steady-state time response $y_{ss}(t)$ of two accelerometers with $\omega_0 = 1$ rad/s in blue and $\omega_0 = 10$ rad/s in green (shown as negative, i.e., $-y_{ss}$), to input cosines $\ddot{x}_1(t)$ (red, dashed, $\omega_1 = 0.2$ rad/s) (top) and $\ddot{x}_2(t)$ (orange, dashed, $\omega_2 = 2$ rad/s) (bottom).

in Figure 17.6. The goal of this set-up typically is to determine or to validate, by means of experiment, the characteristics of the structure.

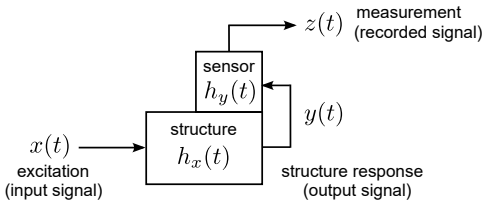


Figure 17.6: A structure turns excitation, input signal $x(t)$, into a response, output signal $y(t)$. Next, a sensor measures this output signal $y(t)$, and, being a system as well, it responds through producing output signal $z(t)$.

Assuming that both structure and sensor are LTI systems, the structure is fully characterized by its impulse response function $h_x(t)$ and the sensor by its impulse response $h_y(t)$. The eventual sensor output is:

$$z(t) = h_y(t) * y(t) = h_y(t) * h_x(t) * x(t) \tag{17.5}$$

with $y(t) = h_x(t) * x(t)$, or, in the frequency domain:

$$Z(f) = H_y(f)Y(f) = H_y(f)H_x(f)X(f)$$

When the structure is subject to an ideal impulse 'hammer' input, $x(t) = \delta(t)$, then its output is, by definition, $y(t) = h_x(t)$.

Next, if we work with an *ideal* sensor, i.e., $h_y(t) = \delta(t)$, then the sensor output directly delivers the impulse response of the system of interest $z(t) = y(t)$, and in this 'ideal hammer input' case $z(t) = h_x(t)$. Referring to the Bode plot, the ideal sensor $h_y(t) = \delta(t)$ shows as $H(\omega) = 1$, that is, with unit amplitude $|H(\omega)| = 1$ and zero phase $\angle H(\omega) = 0$ for all frequencies ω.

Now that the LTI system properties have been explained and a typical measurement set-up discussed, the next chapter presents a major application of systems in signal analysis: filters.

Typically, to achieve this, the sensor must be much smaller than the structure.

18

Filters

This chapter presents an introduction to signal filtering in the frequency domain, as an application of linear time-invariant systems. The filter, as a system, was conceptually already shown in Figure 16.1. An input signal $x(t)$ is fed to a filter, which performs some *frequency-selective* operation, resulting in output signal $y(t)$. A typical application is the use of a so-called low-pass filter to remove unwanted high-frequency noise components from an observed signal.

The operation of a filter is most conveniently explained with continuous-time signals. The last section of this chapter is dedicated to the application of digital filters to discrete-time sequences, i.e., filtering in the discrete time domain.

We limit the discussion to the *principle* of a filter. In practice many different filters exist, which we do not cover. Furthermore, the design of filters is beyond the scope of this book.

18.1 Ideal filters

The operation of a filter is most directly characterized by its response in the frequency domain. Figure 18.1 shows the frequency response function $H(f)$ of three so-called *ideal* filters: the ideal low-pass filter (LPF), the ideal high-pass filter (HPF) and the ideal bandpass filter (BPF). The latter is a combination of the first two, and is characterized by $f_2 > f_1$.

The frequency response functions of these three ideal filters are specified as follows:

- Low-pass filter: $H(f) = 1$ for $|f| \leq f_2$, and zero otherwise,
- High-pass filter: $H(f) = 1$ for $|f| \geq f_1$, and zero otherwise, and
- Bandpass filter: $H(f) = 1$ for $f_1 \leq |f| \leq f_2$, and zero otherwise.

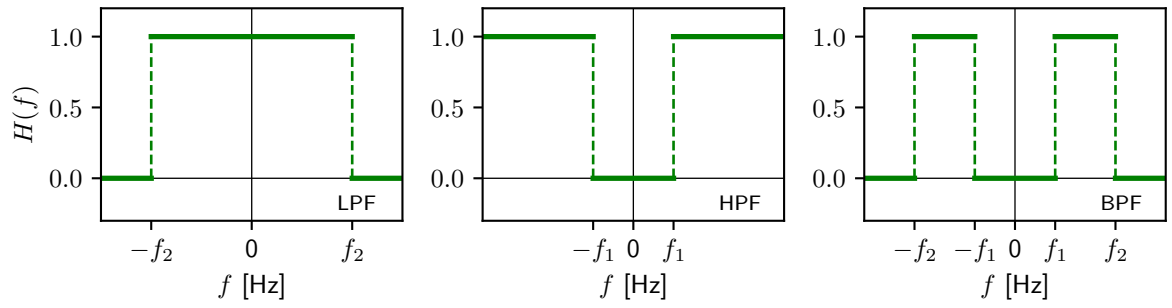


Figure 18.1: FRF of three ideal filters: left, low-pass filter; middle, high-pass filter; right, bandpass filter.

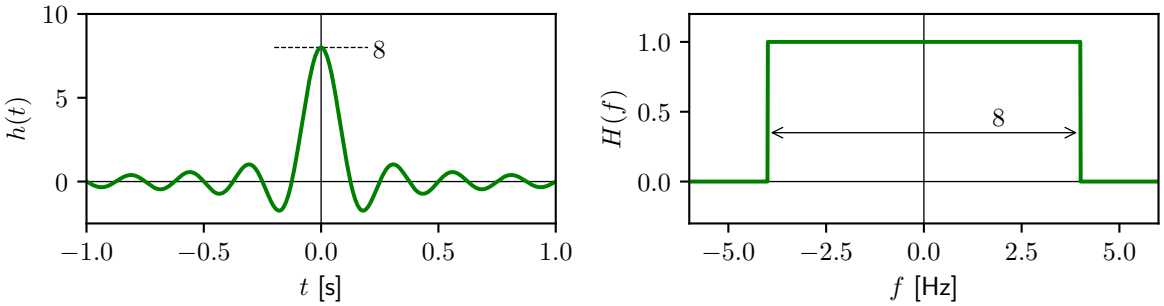


Figure 18.2: Ideal low-pass filter with $f_2 = 4$ Hz. At right, frequency response $H(f) = \Pi\left(\frac{f}{2f_2}\right)$; at left, corresponding impulse response $h(t) = 2f_2\,\mathrm{sinc}(2f_2 t)$.

These filters are *ideal*, as in the selected frequency range, the *pass band*, they let input signal components pass, i.e., $|H(f)| = 1$ and $\angle H(f) = 0$, see (16.14), whereas outside this range, they completely block all input signal components: the *stop band*, where $|H(f)| = 0$. With ideal filters, the transition between pass band(s) and stop band(s) is abrupt.

Frequencies f_1 and f_2 are known as the *bandwidth* frequencies of, respectively, the ideal high-pass and low-pass filters. For the ideal bandpass filter, the one-sided bandwidth equals $f_2 - f_1$, centered about center frequency $\frac{f_1+f_2}{2}$, see Figure 18.1.

The magnitude of the ideal filters can be changed to an arbitrary value through multiplication by a constant $G \in \mathbb{R}\setminus\{0\}$, the so-called filter *gain* (rather than the default value of 1). Then, in the pass band, $|H(f)| = |G|$ and $\angle H(f) = 0$ for $G > 0$ and $\angle H(f) = \pm\pi$ for $G < 0$.

18.1.1 Causality

Ideal filters are *non-causal*, i.e., $h(t) \neq 0$ for $t < 0$. This is because their FRFs $H(f)$ are all real, even functions of frequency, see Figure 18.1. Their impulse response functions $h(t)$ are then all real, even functions in time.

For example, consider the ideal low-pass filter, with FRF $H(f) = \Pi\left(\frac{f}{2f_2}\right)$. Its impulse response function $h(t) = 2f_2\,\mathrm{sinc}(2f_2 t)$ extends from $t = -\infty$ to $t = \infty$, see Figure 18.2 at left. This filter cannot be realized in practice; we can only work with an approximation of this, by means of a finite length impulse response function. Second, because the filter is non-causal, we would need *future* signal values or samples for the filter to work in real time, which is impossible. Only in post-processing, an ideal filter can be applied, with the above approximation.

18.1.2 Impact of ideal filter

The impact of applying an ideal (brickwall) low-pass filter is demonstrated, using as input signal a unit pulse function $x(t) = \Pi(t)$, shown at left in Figure 18.3, and its Fourier transform $X(f) = \mathrm{sinc}(f)$ is shown at right. The low-pass filter has FRF $H(f) = \Pi\left(\frac{f}{2f_2}\right)$, shown at right in Figure 18.2, with $f_2 = 4$ Hz. The corresponding impulse response, $h(t) = 2f_2\,\mathrm{sinc}(2f_2 t)$, see (6.13), is shown at left.

The filter output is found in the frequency domain through multiplication $Y(f) = X(f)H(f)$, (16.7), shown at right in Figure 18.4, and in the time domain through convolution $y(t) = x(t) * h(t)$, (16.6), shown at left. In the graph at right we can see that $Y(f)$ is indeed zero beyond $f_2 = 4$ Hz; compare this to $X(f)$ in Figure 18.3 at right. At left, we observe that 'sharp' features of the input signal, representing high-frequency components, are lost in the output, due to low-pass filtering. Instead, we get 'wobbles' or ripples in the signal, similar to the

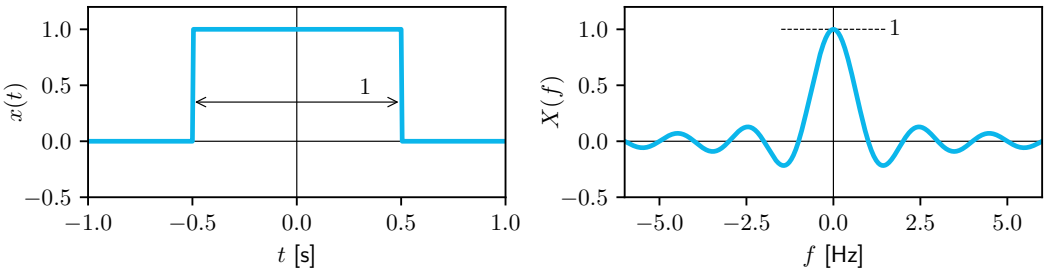


Figure 18.3: Left: input signal $x(t) = \Pi(t)$, the unit pulse function; right: corresponding Fourier transform $X(f) = \text{sinc}(f)$.

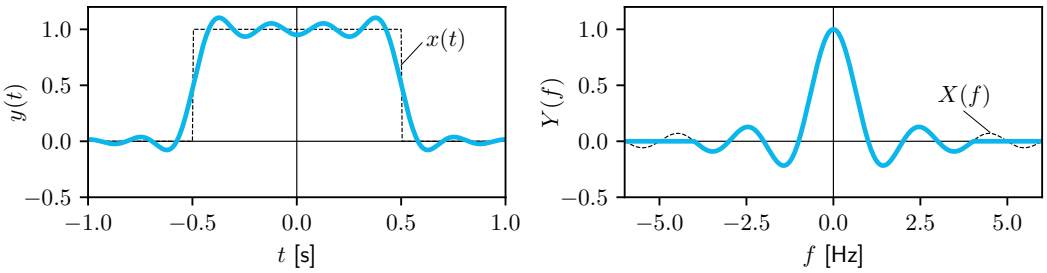


Figure 18.4: Left: output signal $y(t)$, the response of the ideal low-pass filter of Figure 18.2 ($f_2 = 4$ Hz) to a unit pulse input; right: corresponding Fourier transform $Y(f)$. To highlight the working of the filter, the thin black dashed curves show the input signal $x(t)$ (left) and its Fourier transform $X(f)$ (right) from Figure 18.3.

approximation of a square wave by just the first few terms of the Fourier series, Figure 3.1.

The purpose of a low-pass filter is often to filter out high-frequency noise in an observed signal. Similarly, we may want to filter out all (non-relevant) high-frequency components of a signal before sampling that signal, so-called *pre-sample filtering*.

As an example, a signal with (simulated) random noise is shown in Figure 18.5. Figure 18.6 shows the output of the filter, at left for $f_2 = 4$ Hz and at right for $f_2 = 20$ Hz. With $f_2 = 4$ Hz, the ideal LPF clearly is effective in removing the high-frequency noise, but unavoidably, also in removing the high-frequency components of the desired signal $\Pi(t)$; the 'sharp' edges are lost, just like in Figure 18.4 at left. With $f_2 = 20$ Hz, part of the high-frequency noise (in this example, band-limited at 50 Hz, see Appendix L) remains in the output, but at the same time the 'sharp' characteristics of the unit pulse input are also maintained. Applying filters in practice often means finding a good compromise.

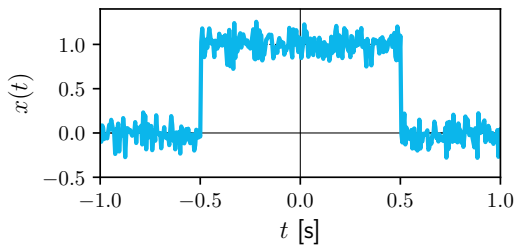


Figure 18.5: Input signal $x(t) = \Pi(t)$, with additive noise, simulated, zero mean, and normally distributed with standard deviation equal to 0.1.

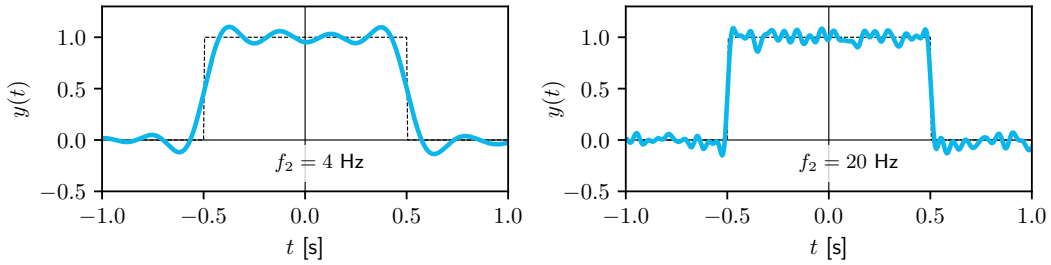


Figure 18.6: Left: output signal $y(t)$ as a result of ideal low-pass filtering the unit pulse input contaminated with noise (Figure 18.5) with filter bandwidth $f_2 = 4$ Hz; right: with filter bandwidth $f_2 = 20$ Hz. The thin black dashed lines show the unperturbed input signal $x(t) = \Pi(t)$ from Figure 18.3 (at left).

EX 18.1 Consider as input voltage signal $x(t) = 10\,\mathrm{sinc}(10t)$, fed through an ideal high-pass filter with bandwidth frequency $f_1 = 3$ Hz. Compute the energy of output signal $y(t)$.

Solution The Fourier transform of $x(t)$ equals $X(f) = \Pi\left(\frac{f}{10}\right)$, (6.13), and is shown in Figure 18.7 at the top. The frequency response of the ideal high-pass filter is shown in the middle of this figure. Through $Y(f) = H(f)X(f)$, (16.7), we find that $Y(f) = \Pi\left(\frac{f+4}{2}\right) + \Pi\left(\frac{f-4}{2}\right)$, shown at the bottom of Figure 18.7.

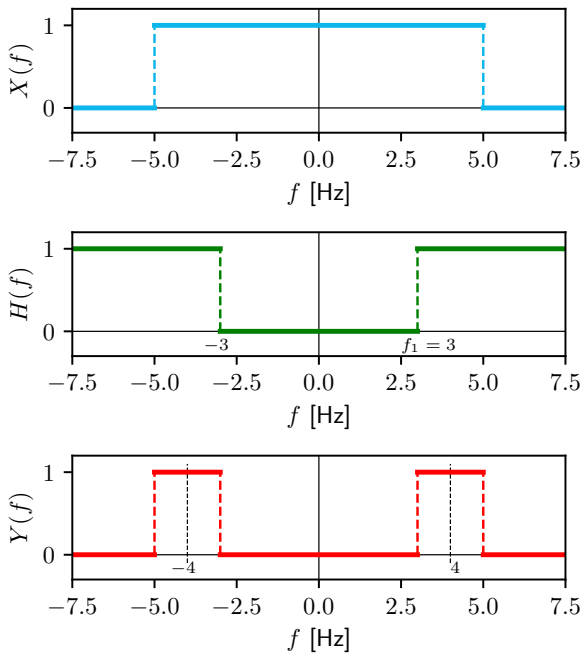


Figure 18.7: Fourier transform of input signal $x(t)$, $X(f)$ (top), FRF of the ideal high-pass filter $H(f)$ (middle) and Fourier transform of the output signal $y(t)$, $Y(f)$ (bottom).

The energy of the output signal can be computed with (5.19):

$$E_y = \int_{-\infty}^{\infty} |Y(f)|^2 \,\mathrm{d}f = (1)^2 2 + (1)^2 2 = 4 \text{ J}$$

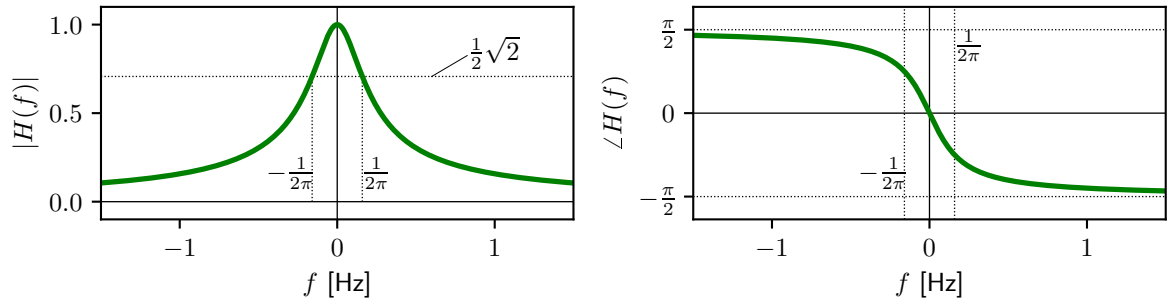


Figure 18.8: FRF of first-order Butterworth low-pass filter (18.1): magnitude response $|H(f)|$ at left and phase response $\angle H(f)$ at right. The vertical dotted lines indicate $f = \pm\frac{1}{2\pi}$ Hz.

18.2 Practical filters

The FRF of an ideal filter abruptly changes, causing a discontinuity in $H(f)$ at some frequency. For instance, for the ideal low-pass filter, $H(f)$ changes from $H(f) = 1$ to $H(f) = 0$ at $f = f_2$. In order to arrive at causal filters, we must allow for a *gradual change*, in the frequency domain, from pass band to stop band (and vice versa).

A basic, causal low-pass filter is the first-order Butterworth filter.[1] The frequency response function of this filter, a first-order LTI system, reads:

$$H(f) = \frac{1}{1 + j2\pi f} \tag{18.1}$$

or $H(\omega) = \frac{1}{1+j\omega}$ in terms of angular frequency ω. The magnitude and phase response of the filter are shown in Figure 18.8. The magnitude response reads:

$$|H(f)| = \frac{1}{\sqrt{1 + (2\pi f)^2}} \tag{18.2}$$

The phase response equals:

$$\angle H(f) = -\arctan(2\pi f) \tag{18.3}$$

For frequencies f close to zero, $|H(f)| \approx 1$ and $\angle H(f) \approx 0$; for large frequencies $|H(f)| \approx 0$ and $\angle H(f) \approx -\frac{\pi}{2}$. The pass band and stop band of the filter become clearer when considering the Bode plot of the filter (see Figure 18.9) with logarithmic frequency and magnitude scales. In contrast to the (in practice non-realizable) ideal low-pass filter, there is indeed a *gradual* transition from pass band to stop band.

18.2.1 Cut-off frequency

The width of the pass band of the first-order Butterworth filter can be tuned by 'scaling' the frequency variable. We rewrite $H(f)$, (18.1), as:

$$H(f) = \frac{1}{1 + j\frac{f}{f_3}} \tag{18.4}$$

with $f_3 = \frac{1}{2\pi}$. At frequency $f = f_3$, we have $|H(f_3)| = \frac{\sqrt{2}}{2}$ or $|H(f_3)|^2 = \frac{1}{2}$. Therefore, as is shown in (16.16), an input signal component at this frequency is attenuated by the filter, such

[1] After the British physicist S. Butterworth (1885-1958), who proposed an entire family of filters.

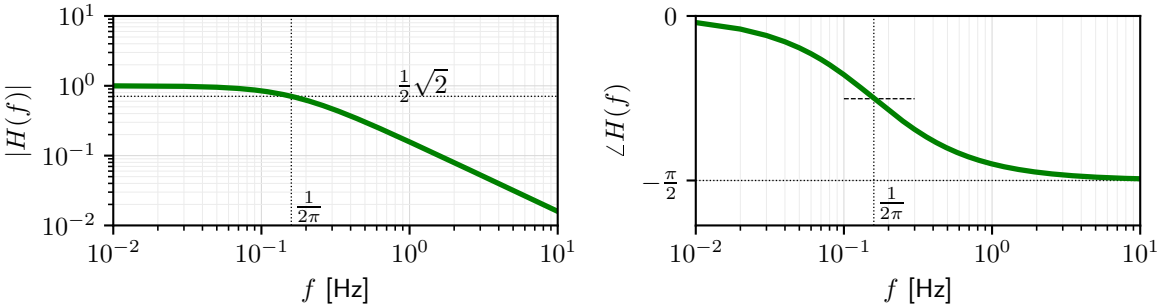


Figure 18.9: Bode plot of the same first-order Butterworth low-pass filter as in Figure 18.8. Note that usually the phase response is shown below the magnitude response.

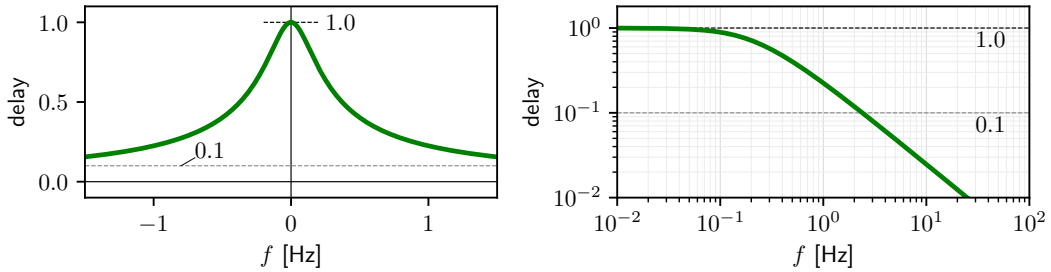


Figure 18.10: Magnitude of time delay (in [s]) of the first-order Butterworth low-pass filter (18.1) as a function of frequency, on a linear (left) and logarithmic (right) axis.

that half of its power or energy remains at the output. By convention, the gradual transition from pass band to stop band is characterized by this 'half power' frequency, referred to as the *cut-off frequency*, or *3dB-frequency*. The subscript '3' in f_3 is chosen on purpose to mark that a reduction by half of the power corresponds to a loss of 3 dB; as $10\log_{10}(\frac{1}{2}) \approx -3$. The cut-off frequency f_3 is indicated by the dotted vertical lines in Figures 18.8 and 18.9.

18.2.2 Frequency response function – phase revisited

The phase response of (18.1) is $\angle H(f) = \theta(f) = -\arctan(2\pi f)$, (18.3) and shown in Figures 18.8 and 18.9 at right. The output of a linear time-invariant system to a cosine input was given by (16.14) and the phase response was discussed in Section 16.4.2.

For the first-order Butterworth filter, the phase $\theta(f)$ clearly depends on frequency f in a nonlinear way. For an input signal with frequency $f = f_0$, the output lags behind the input (negative time shift, $t_0 < 0$) with time delay $t_0 = \frac{\theta(f_0)}{2\pi f_0} = -\frac{\arctan(2\pi f_0)}{2\pi f_0}$; this is shown in Figure 18.10 on linear (left) and logarithmic (right) axes. For small frequencies, the filter output lags behind the input by exactly one second.

18.2.3 Low-pass filter analogy: mechanical system

By means of an example, we show a mechanical analogy of the first-order Butterworth filter, (18.1). That is, we present a mechanical system which, upon a load or input force, responds exactly like this low-pass filter.

Figure 18.11 shows a simple mass-damper system. The motion of the proof mass is the result of the external force $F(t) = m\ddot{x}(t)$, the input to the system, and the force of the damper, which delivers a linear drag, a force proportional to the velocity $\dot{y}(t)$ of the proof mass. The mass is assumed to slide without friction over the supporting surface.

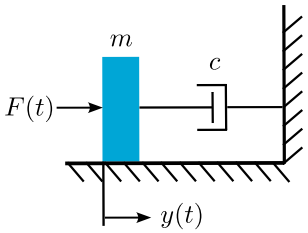


Figure 18.11: Basic outline of a mass-damper system: a proof mass (in blue) is connected to a dashpot damper. An external force (input) $F(t) = m\ddot{x}(t)$ is applied to the proof mass, resulting in motion (displacement) of the proof mass $y(t)$. The damper delivers a linear drag with a force proportional to the velocity of the proof mass.

The corresponding differential equation follows from $m\ddot{y}(t) = m\ddot{x}(t) - c\dot{y}(t)$ as:

$$\frac{m}{c}\ddot{y}(t) + \dot{y}(t) = \frac{m}{c}\ddot{x}(t)$$

with mass m in [kg] and damper coefficient c in [Ns/m]. With constant $\alpha = \frac{c}{m}$ we obtain:

$$\frac{1}{\alpha}\ddot{y}(t) + \dot{y}(t) = \frac{1}{\alpha}\ddot{x}(t) \tag{18.5}$$

This is an LTI system of order 2. To obtain the relation between the output *velocity* $\dot{y}(t)$ and input acceleration $\ddot{x}(t)$, we observe that the DE above is similar to the one in Example 16.5, (16.9); it differs by just a factor of $\frac{1}{\alpha}$ on the right-hand side. Solving the resulting *first*-order differential equation in the frequency domain yields:

$$H(f) = \frac{\dot{Y}(f)}{\ddot{X}(f)} = \frac{1}{\alpha + j2\pi f} \tag{18.6}$$

For $\alpha = 1$ the FRF is indeed identical to the first-order Butterworth low-pass filter FRF, (18.1).

In Example 5.3 the impulse response function $h(t)$ was found to be:

$$h(t) = e^{-\alpha t}u(t) \text{ for } \alpha > 0 \tag{18.7}$$

shown in Figure 5.4 for $\alpha = 1$ and $\alpha = 3$. As is shown, due to the presence of the unit step function $u(t)$ in (18.7), this system is *causal*: $h(t) = 0$ for $t < 0$.

Example 7.4 demonstrates for $\alpha = 1$ the response of this system to a unit step input. Figure 7.8 shows that the output is indeed zero for $t < 0$ and then features a gradual transition to the new state. At time $t = \frac{1}{\alpha}$, the output reaches 63% of its final value.

The 3dB frequency of the FRF (18.6) equals $\frac{\alpha}{2\pi}$ with $\alpha = \frac{c}{m}$. Through varying the dashpot damper coefficient c and/or the proof mass m we can change the bandwidth of the resulting low-pass filter.

18.2.4 FIR and IIR

In the context of filters, two expressions are often used to distinguish between two types of impulse responses of the filter system:

- Infinite Impulse Response (IIR), where $h(t)$ continues indefinitely, and
- Finite Impulse Response (FIR), where $h(t) = 0$ for $t > t_0$, for some finite t_0; $h(t)$ has a finite duration and settles to zero in finite time.

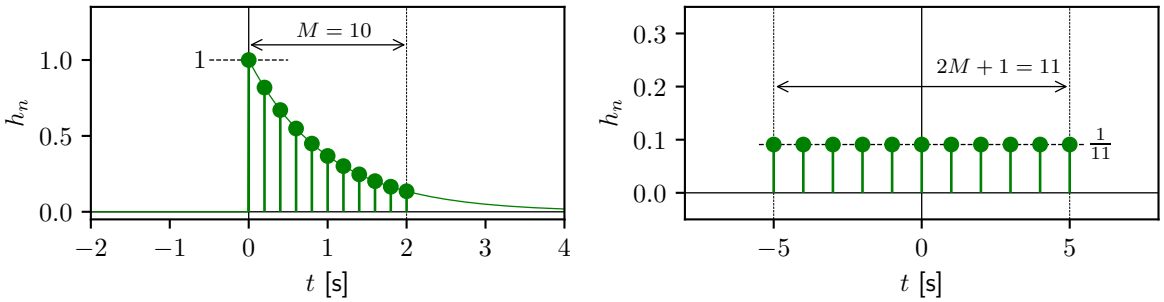


Figure 18.12: Left: impulse response sequence h_n in discrete time of causal first-order Butterworth low-pass filter ($\alpha = 1$) (with $M = 10$ and $\Delta t = 0.2$ s); right: non-causal moving-average filter (with $M = 5$ and $\Delta t = 1$ s).

The first-order Butterworth low-pass filter is an IIR, with impulse response $h(t) = e^{-t}u(t)$ ($\alpha = 1$), shown in Figure 7.7 at right. This function asymptotically approaches zero only for $t \to \infty$. We may approximate an IIR by a FIR: when $h(t)$ gets close to zero from a certain point onwards, we simply set these non-zero impulse response values to zero.

The moving-average filter, presented in the next section, is a FIR. Its impulse response has a *finite* duration. The FIR filter or system output $y(t)$ to input $x(t)$ is given by:

$$y(t) = \int_0^{t_0} h(\tau)x(t-\tau)\,\mathrm{d}\tau \tag{18.8}$$

where we used the commutative property of convolution, (7.3). Output $y(t)$ at time t relies on input signal $x(t)$ with $t \in [t - t_0, t]$ (a finite interval). We assumed the filter (i.e., the system) to be causal, $h(t) = 0$ for $t < 0$, such that the lower integral bound in (18.8) is $\tau = 0$.

18.3 Digital filters

In continuous time, filters are relatively easy to understand and apply, mostly using analog electrical circuits. In practice, however, we often apply *digital* filtering to discrete-time sequences. The filter output, sequence y_n, follows as the convolution of input sequence x_n with filter impulse response sequence h_n: $y_n = x_n * h_n$; see also Appendix F.

Now that we covered continuous-time filter impulse response functions for LTI systems, in this section we take a pragmatic approach to briefly introduce digital filters. The discrete-time version of the filter impulse response h_n follows from *sampling* the continuous-time impulse response $h(t)$. In case of an IIR we take an additional approximation and *truncate* the continuous-time impulse response to finite length.

18.3.1 Digital Butterworth filter

We apply this approach for the first-order Butterworth low-pass filter, for which the impulse response $h(t)$ was given by (18.7) with $\alpha = 1$ (Figure 7.7 at right). This is a causal filter. The corresponding discrete-time impulse response is obtained by:

$$h_n = \begin{cases} h(n\Delta t) & 0 \le n \le M \\ 0 & \text{otherwise} \end{cases} \tag{18.9}$$

with finite M, and hence we approximate the IIR first-order Butterworth by a FIR. The impulse response function $h_n = e^{-n\Delta t}$ goes to zero quite quickly (see Figure 18.12 at left, for $M = 10$ and $\Delta t = 0.2$ s).

18.3.2 Moving-average filter

A moving-average (MA) operation can also be considered a low-pass filter, as it tends to smooth (short-term) fluctuations in the input signal. At each time instant $t = n\Delta t$, the output signal is computed as the average over a set of adjacent (past, and possibly also *future*) input signal values; this means the window is sliding over the input data sequence. For instance:

$$y_n = \frac{1}{2M+1}(x_{n-M} + x_{n-(M-1)} + \ldots + x_n + x_{n+1} + \ldots + x_{n+M}) \tag{18.10}$$

for a centralized moving-average filter with window length $2M + 1$. This operation can be expressed as a discrete convolution $y_n = x_n * h_n$ with impulse response sequence h_n:

$$h_n = \begin{cases} \frac{1}{2M+1} & |n| \leq M \\ 0 & \text{otherwise} \end{cases} \tag{18.11}$$

This is a length $2M + 1$ sequence, shown for $M = 5$ and $\Delta t = 1$ s in Figure 18.12 at right.

The moving-average filter has a FIR. The centralized moving-average filter as above is non-causal, as $h_n \neq 0$ for $n < 0$. The filter can be made causal through shifting the impulse response by M positions such that it only considers present and past values:

$$y_n = \frac{1}{2M+1}(x_{n-2M} + x_{n-(2M-1)} + \ldots + x_{n-1} + x_n)$$

still with a length $2M + 1$ impulse response sequence; then $h_n = 0$ for $n < 0$. The impulse response shown in Figure 18.12 at right would be shifted by $M = 5$ positions to the right, and would be applicable in real time.

Appendices

A

Common mathematical formulas

A.1 Trigonometric identities

$$\sin^2 u + \cos^2 u = 1 \tag{A.1}$$

$$\cos(2u) = \cos^2 u - \sin^2 u \tag{A.2}$$

$$\sin(2u) = 2 \sin u \cos u \tag{A.3}$$

$$\cos^2 u = \frac{1}{2}(1 + \cos(2u)) \tag{A.4}$$

$$\sin^2 u = \frac{1}{2}(1 - \cos(2u)) \tag{A.5}$$

$$\sin(u \pm v) = \sin u \cos v \pm \cos u \sin v \tag{A.6}$$

$$\cos(u \pm v) = \cos u \cos v \mp \sin u \sin v \tag{A.7}$$

$$\sin u \sin v = \frac{1}{2}(\cos(u - v) - \cos(u + v)) \tag{A.8}$$

$$\cos u \cos v = \frac{1}{2}(\cos(u - v) + \cos(u + v)) \tag{A.9}$$

$$\sin u \cos v = \frac{1}{2}(\sin(u - v) + \sin(u + v)) \tag{A.10}$$

A.2 Orthogonality of sines and cosines

Assume ℓ and m to be non-zero integers, $\ell, m \in \mathbb{N}^+$and T_0 the period corresponding to the fundamental angular frequency $\omega_0 = \frac{2\pi}{T_0}$. The *orthogonality* properties of integrals involving products of sines and cosines can then be defined as follows:

$$I_1 = \int_{T_0} \sin(\ell\omega_0 t) \sin(m\omega_0 t)\, \mathrm{d}t = \begin{cases} 0 & \text{for } \ell \neq m \\ \frac{T_0}{2} & \text{for } \ell = m \neq 0 \end{cases} \tag{A.11}$$

$$I_2 = \int_{T_0} \cos(\ell\omega_0 t) \cos(m\omega_0 t)\, \mathrm{d}t = \begin{cases} 0 & \text{for } \ell \neq m \\ \frac{T_0}{2} & \text{for } \ell = m \neq 0 \end{cases} \tag{A.12}$$

$$I_3 = \int_{T_0} \sin(\ell\omega_0 t) \cos(m\omega_0 t)\, \mathrm{d}t = 0 \qquad \text{for all } \ell, m \tag{A.13}$$

These integrals can be proven using trigonometric identities such as (A.8).

A.3 Definite integrals

$$\int_0^\infty \frac{\sin(ax)}{x}\,\mathrm{d}x = \frac{\pi}{2}, \quad a > 0 \tag{A.14}$$

$$\int_0^\infty \frac{\sin^2 x}{x^2}\,\mathrm{d}x = \frac{\pi}{2} \tag{A.15}$$

A.4 Indefinite integrals

$$\int x\sin(ax)\,\mathrm{d}x = \frac{1}{a^2}(\sin(ax) - ax\cos(ax)) \tag{A.16}$$

$$\int x\cos(ax)\,\mathrm{d}x = \frac{1}{a^2}(\cos(ax) + ax\sin(ax)) \tag{A.17}$$

$$\int e^{ax}\sin(bx)\,\mathrm{d}x = e^{ax}\frac{1}{a^2+b^2}(a\sin(bx) - b\cos(bx)) \tag{A.18}$$

$$\int e^{ax}\cos(bx)\,\mathrm{d}x = e^{ax}\frac{1}{a^2+b^2}(a\cos(bx) + b\sin(bx)) \tag{A.19}$$

A.5 Series expansions

Sum of squared reciprocal integers [6]:

$$\sum_{k=1}^{\infty}\frac{1}{k^2} = \frac{1}{1^2} + \frac{1}{2^2} + \frac{1}{3^2} + \ldots = \frac{\pi^2}{6} \tag{A.20}$$

Sum of squared reciprocal odd integers [6]:

$$\sum_{k=1}^{\infty}\frac{1}{(2k-1)^2} = \frac{1}{1^2} + \frac{1}{3^2} + \frac{1}{5^2} + \ldots = \frac{\pi^2}{8} \tag{A.21}$$

Geometric series identity:

$$\sum_{k=0}^{N-1} a^k = \begin{cases} \dfrac{1-a^N}{1-a} & \text{for } a \neq 1 \\ N & \text{for } a = 1 \end{cases} \tag{A.22}$$

B

Elementary functions

In this appendix we introduce a couple of elementary continuous-time functions. They are frequently used as signals in examples to illustrate and elaborate on the theory.

B.1 Unit pulse function

The unit pulse function, with symbol Π resembling the function shape, is defined as:

$$\Pi(t) = \begin{cases} 1 & |t| < \frac{1}{2} \\ 0.5 & |t| = \frac{1}{2} \\ 0 & |t| > \frac{1}{2} \end{cases} \tag{B.1}$$

The unit pulse function is shown in Figure B.1 at left. The unit pulse is an even function with amplitude equal to 1 and extending to half a second on both sides of the origin. The unit pulse $\Pi(t)$ has a width of 1 second. At $t = \pm\frac{1}{2}$ s we *define* the function value to be at $\frac{1}{2}$ (the blue circles in Figure B.1). This function is also known as the boxcar or rectangular function.

B.2 Unit triangular function

The unit triangular function, with symbol Λ resembling the function shape, is defined as:

$$\Lambda(t) = \begin{cases} 1 - |t| & \text{for } |t| \leq 1 \\ 0 & \text{for } |t| > 1 \end{cases} \tag{B.2}$$

The unit triangular function is shown in Figure B.1 at right. The triangular function is an even function with amplitude equal to 1 at maximum for $t = 0$ s; it reaches zero on both sides of the origin at $t = \pm 1$ s; its width is 2 seconds.

B.3 Sine cardinal function (sinc)

The sine cardinal function, or sinc function for short, is defined as:

$$\operatorname{sinc}(t) = \frac{\sin(\pi t)}{\pi t} \tag{B.3}$$

The sinc function is illustrated in Figure B.2. It has a maximum of 1 at $t = 0$ s (using L'Hôspital's rule) and equals zero at $t = \pm 1, \pm 2, \pm 3, \ldots$, which is implied by the sine in the numerator. Some

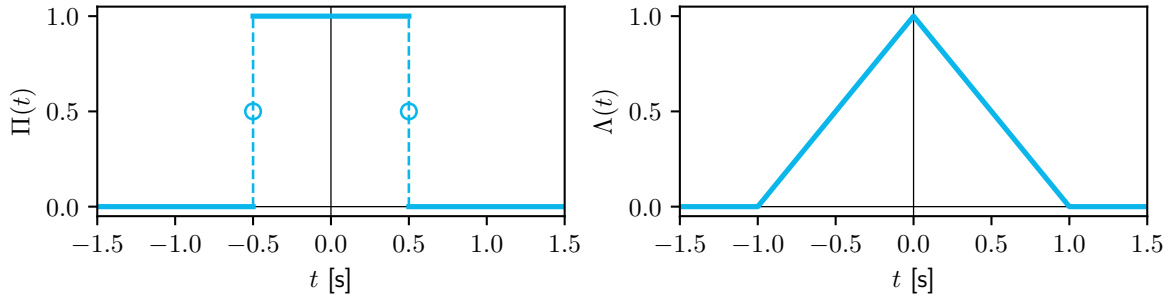


Figure B.1: Unit pulse function $\Pi(t)$ at left, and unit triangular function $\Lambda(t)$ at right.

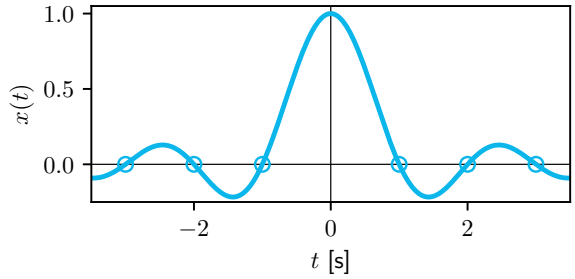


Figure B.2: Sine cardinal, or sinc function $x(t) = \mathrm{sinc}(t)$.

zero-crossings are shown in Figure B.2 using blue circles. From (B.3) it follows that:

$$\lim_{t \to \pm\infty} \mathrm{sinc}(t) = 0 \tag{B.4}$$

The definition in (B.3) is sometimes referred to as the *normalized* sinc function (to distinguish it from the definition without π in both the numerator and denominator).

B.4 Singularity functions

Three common singularity functions are introduced here: the unit step function, the unit ramp function and the Dirac delta function. The unit step function is the derivative of the unit ramp function, and the Dirac delta function is the derivative of the unit step function.

B.4.1 Unit step function

The unit step function is defined as:

$$u(t) = \begin{cases} 0 & t < 0 \\ 0.5 & t = 0 \\ 1 & t > 0 \end{cases} \tag{B.5}$$

Figure B.3 (left) illustrates the unit step function. At $t = 0$ s its value is *defined* to lie half-way, at $\frac{1}{2}$ (blue circle). The unit step function represents a signal that switches on at $t = 0$ s and stays on indefinitely. It is also known as the Heaviside[1] step function.

[1] After the British mathematician O. Heaviside (1850-1925).

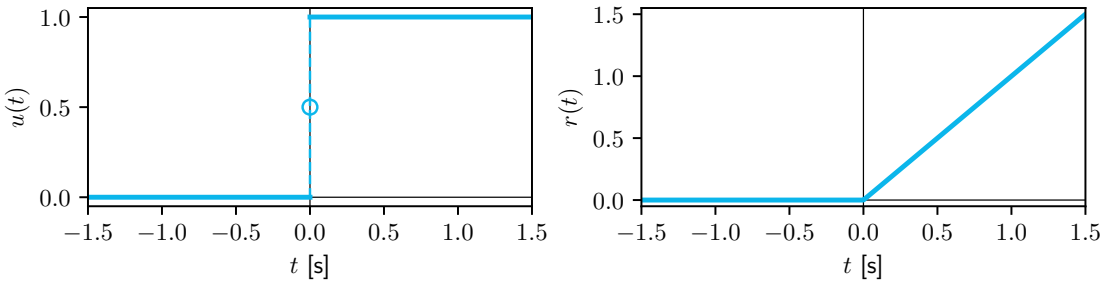


Figure B.3: Unit step function $u(t)$ at left, and unit ramp function $r(t)$ at right.

B.4.2 Unit ramp function

The unit ramp function is defined as:

$$r(t) = \begin{cases} 0 & t < 0 \\ t & t \geq 0 \end{cases} \tag{B.6}$$

and shown in Figure B.3 at right. This function grows indefinitely for $t > 0$.

The unit ramp function is related to the unit step function as:

$$r(t) = \int_{-\infty}^{t} u(\sigma)\,\mathrm{d}\sigma \tag{B.7}$$

B.4.3 Dirac delta function

The unit impulse or Dirac delta function is definitely the most frequently used singularity function in Fourier analysis. It is defined to be zero everywhere, except for $t = 0$:

$$\boxed{\delta(t) = 0 \quad t \neq 0} \tag{B.8}$$

with unit area:

$$\boxed{\int_{-\infty}^{\infty} \delta(t)\,\mathrm{d}t = 1} \tag{B.9}$$

which is obtained in an infinitesimal interval of time. The Dirac delta function resembles an extremely narrow, extremely high spike or peak at $t = 0$ s, see Figure B.4. Its value *at* $t = 0$ s is infinity; its width is zero. When multiplied by a number a, the area of the Dirac delta function, also defined as its *weight*, becomes a. The Dirac delta function is used to represent a signal or phenomenon that occurs (almost) instantaneously, e.g., the impact of a hammer on a clock bell, or the impact of a billiard ball on another one when hitting it.

The Dirac delta function is often referred to as a Dirac pulse. In many practical cases, the Dirac delta function can be considered as a limiting case of the pulse function, (B.1):

$$\delta(t) = \lim_{\epsilon \downarrow 0} \frac{1}{\epsilon} \Pi\left(\frac{t}{\epsilon}\right) \tag{B.10}$$

This function attains value $\frac{1}{\epsilon}$ over the interval from $t = -\frac{\epsilon}{2}$ to $t = \frac{\epsilon}{2}$ s, and has unit area for all values of ϵ. This approximation, known as a 'nascent delta function', demonstrates that the Dirac delta function is an *even* function.

The unit step function is related to the unit impulse function as:

$$u(t) = \int_{-\infty}^{t} \delta(\sigma)\,\mathrm{d}\sigma \tag{B.11}$$

Note that the Dirac delta function is not a function in the traditional sense. With the above definition of taking on a non-zero value only for $t = 0$ s, a Riemann integral over the Dirac delta function would yield zero (rather than 1, as defined). The above integration must instead be understood in the sense of so-called generalized functions, and a proper explanation can be done using the theory of distributions [13].

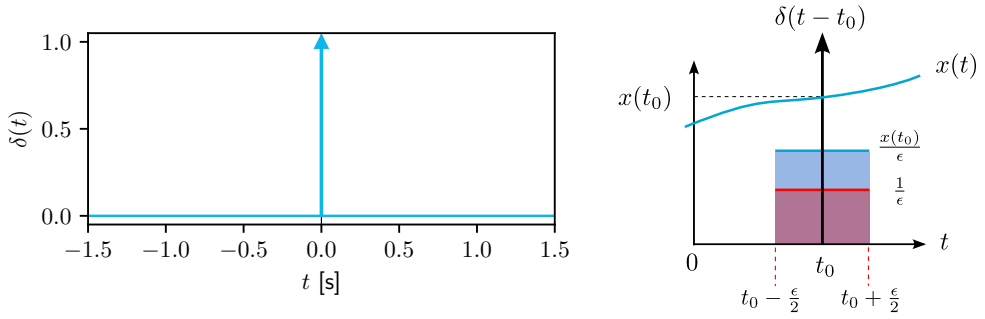


Figure B.4: At left: unit impulse or Dirac delta function $\delta(t)$. At right: sifting property of the Dirac delta function, at $t = t_0$, in an integral over smooth function $x(t)$. Here, the Delta pulse approximation is shown in red, centered at time t_0 with duration ϵ and amplitude $\frac{1}{\epsilon}$, and smooth function $x(t)$ in blue. The integral of the sifting property is represented by the area in blue, and equal to $x(t_0)$.

The Dirac delta function is a so-called generalized function. It is not defined in terms of values, but rather how it acts in an integral, when multiplied by a smooth function. This leads to the *sifting* property, with a Dirac delta function occurring at $t = t_0$:

$$\boxed{\int_{-\infty}^{\infty} x(t)\delta(t - t_0)\,\mathrm{d}t = x(t_0)} \tag{B.12}$$

The result of this integral is the value of function $x(t)$ *at* $t = t_0$; the Dirac delta function *sifts out* the value of $x(t)$ at $t = t_0$. This is illustrated in Figure B.4.

The Dirac delta function has the inverse dimension of its argument. Consider $\delta(t)$ with time t in seconds, then the unit of $\delta(t)$ is one-over-seconds $[\frac{1}{\mathrm{s}}]$. The area under this 'function' equals 1 (dimensionless), see (B.9). This result also follows from the approximation $\frac{1}{\epsilon}\Pi\left(\frac{t}{\epsilon}\right)$, where ϵ has the same unit as t. Similarly, the unit of $\delta(f)$, with f in [Hz], is one-over-Hertz, or seconds, [s].

When changing the 'running variable' in a function which includes a Dirac delta function, we need to assure that (B.9) holds. Generally, it can be shown that:

$$\boxed{\delta(t) = |a|\delta(at)} \tag{B.13}$$

for some non-zero constant $a \in \mathbb{R}$ and here with a Dirac delta function in time. For instance, in the frequency domain, when expressing $X(f)$ as $X(\omega)$, with angular frequency $\omega = 2\pi f$, we substitute $2\pi\delta(\omega)$ for each occurrence of $\delta(f)$ in $X(f)$. When expressing $X(2\pi f\Delta t)$, with discrete frequency $2\pi f\Delta t$, as $X(f)$, we substitute $\delta(2\pi f\Delta t) = \frac{1}{2\pi\Delta t}\delta(f)$ for each occurrence of $\delta(2\pi f\Delta t)$ in $X(2\pi f\Delta t)$.

C

Complex algebra

In signal analysis we heavily rely on the use of complex algebra. Complex numbers can be considered an extension of real numbers. As shown in Table C.1, complex numbers are of the form $a + jb$, with both a and b real numbers, and j the imaginary number or unit, for which holds that $j^2 = -1$. The complex number $X = a + jb$ has a real part, a, and an imaginary part, b. Hence: $a = \mathrm{Re}(X)$ and $b = \mathrm{Im}(X)$.

Table C.1: Overview of number systems.

symbol	number system	examples	comment
$\mathbb{N}$	natural numbers	$0, 1, 2, 3, \ldots$	
$\mathbb{Z}$	integers	$\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots$	
$\mathbb{Q}$	rational numbers	$-\frac{3}{2}$, $\frac{17}{6}$, ...	ratio of two integers
$\mathbb{R}$	real numbers	-10, -0.33, 0, $\sqrt{2}$, π, ...	
$\mathbb{C}$	complex numbers	$2 - j3$, $3 + j4$, ...	$a + jb$

C.1 Complex plane

A complex number X can be interpreted as a vector in the complex plane, see Figure C.1, also known as an Argand diagram.[1] The length, or magnitude, of the vector can be written as R and equals the modulus of the complex number: $R = |X|$. The complex absolute value is a Euclidean[2] norm: $R = \sqrt{a^2 + b^2}$. The angle of this vector with the real axis is $\theta = \arg(X) = \angle X$, and equals $\theta = \arctan(\frac{b}{a})$.[3] The angle goes counterclockwise, its unit is radians.

When the imaginary part of a complex number is zero, this number lies somewhere on the real axis of the complex plane. When the real part of that number is positive, it lies on the positive real axis, its phase is 0. When the real part is negative, it lies on the negative real axis, its phase is $\pm\pi$. When the real part is also zero, the number lies at the origin of the complex plane, its magnitude is zero, its phase is *defined* zero.

Complex number X can also be written as $X = R\cos\theta + jR\sin\theta$, which later will be turned into $X = Re^{j\theta}$, using the complex exponential $e^{j\theta}$. Note that formally, a complex exponential has the form e^Z with Z a complex number, i.e., $Z = c + jd$, hence $e^Z = e^c e^{jd}$, though in this book we restrict to complex numbers on the unit circle in the complex plane, hence $c = 0$.

[1] After the Swiss mathematician J.-R. Argand (1768-1822)
[2] After the Greek mathematician Euclid (≈300 BC).
[3] Numerically, e.g., in Python, you need to use the four-quadrant arctan function.

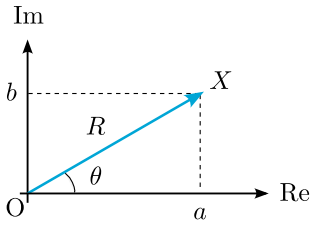


Figure C.1: Complex number $X = a + jb$ interpreted as a vector in the complex plane. The vector has magnitude R, and angle θ with the real axis; $a = R\cos\theta$ and $b = R\sin\theta$.

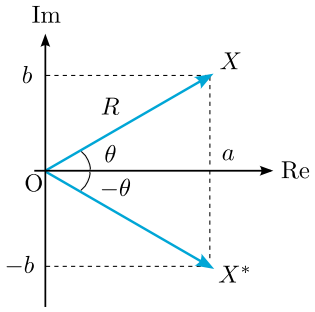


Figure C.2: Complex number $X = a + jb$ and its complex conjugate $X^* = a - jb$. Complex number X and its complex conjugate X^* have equal magnitude R, but opposite angle θ.

Complex number $X = a + jb$ is referred to as expressed in *Cartesian*[4] coordinates a and b. That same complex number X can be expressed in *polar* coordinates R and θ: $X = Re^{j\theta}$.

C.2 Euler's formula

Using e, the base of the natural logarithm, Euler's formula reads:

$$\boxed{e^{j\theta} = \cos\theta + j\sin\theta} \tag{C.1}$$

With this equation we can easily find that:

$$\frac{1}{2}(e^{-j\theta} + e^{j\theta}) = \cos\theta \tag{C.2}$$

and:

$$j\frac{1}{2}(e^{-j\theta} - e^{j\theta}) = \sin\theta \tag{C.3}$$

C.3 Complex conjugate

Complex number $X = a + jb = Re^{j\theta}$ has a conjugate defined as $X^* = a - jb = Re^{-j\theta}$. The complex conjugation operation is the reflection with respect to the real axis, see Figure C.2. It is easy to prove that $R = |X| = \sqrt{XX^*} = \sqrt{X^*X}$, $X + X^* = 2\mathrm{Re}(X)$, and $|X| = |X^*|$.

[4] After the French mathematician/philosopher R. Descartes (1596-1650).

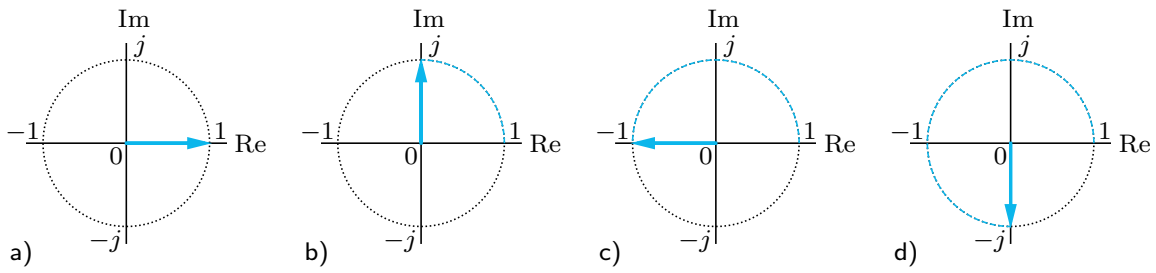


Figure C.3: Complex number a) $X = 1 + j0$, b) multiplied by j, hence $jX = j$, c) again multiplied by j, hence $jj = -1$, and d) multiplied another time by j, hence $-j$.

C.4 Operations

The addition and difference of two complex numbers are easiest obtained using the Cartesian form. With $X = a + jb$ and $Z = c + jd$ we have:

$$X + Z = (a + c) + j(b + d) \tag{C.4}$$

and:

$$X - Z = (a - c) + j(b - d) \tag{C.5}$$

Multiplication and division are easiest done in the polar form, with $X = Re^{j\theta}$ and $Z = Se^{j\phi}$:

$$XZ = RSe^{j(\theta+\phi)} \tag{C.6}$$

and:

$$\frac{X}{Z} = \frac{R}{S}e^{j(\theta-\phi)} \tag{C.7}$$

Using these relationships, it is relatively easy to understand what happens to a complex number X when you multiply it by j:

$$jX = e^{j\frac{\pi}{2}}Re^{j\theta} = Re^{j(\theta+\frac{\pi}{2})} \tag{C.8}$$

When being multiplied by j, the complex number X is rotated in the complex plane by $+\frac{\pi}{2}$ radians (counterclockwise), as illustrated in Figure C.3. Similarly, dividing a complex number X with j means multiplying that number by $-j$:

$$\frac{X}{j} = \frac{j}{j}\frac{X}{j} = -jX = e^{-j\frac{\pi}{2}}Re^{j\theta} = Re^{j(\theta-\frac{\pi}{2})} \tag{C.9}$$

The complex number is rotated in the complex plane by $-\frac{\pi}{2}$ (clockwise).

C.5 Phasor

To start relating complex numbers to the description of signals, we introduce the rotating phasor signal as a time-variant vector in the complex plane:

$$\tilde{x}(t) = Ae^{j(\omega_0 t+\theta)} = Ae^{j\theta}e^{j\omega_0 t} \tag{C.10}$$

This is a vector with magnitude A and (time-variant) angle $\omega_0 t + \theta$ through time t in the exponent. The radial frequency ω_0 (in radians/second) follows from the linear (or ordinal)

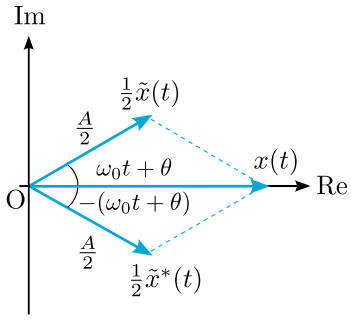


Figure C.4: Real signal $x(t) = A\cos(\omega_0 t + \theta)$ is constructed as the sum of complex vector $\frac{1}{2}\tilde{x}(t)$ and its complex conjugate $\frac{1}{2}\tilde{x}^*(t)$.

frequency f_0 (in [Hz], i.e., 'per second') using $\omega_0 = 2\pi f_0$. In a time duration of $T_0 = \frac{1}{f_0} = \frac{2\pi}{\omega_0}$ seconds this vector completes a full circle (2π radians) in the complex plane.

The initial position of this vector at $t = 0$ s is referred to as the *phasor*:

$$\boxed{\vec{X} = Ae^{j\theta}} \tag{C.11}$$

The phasor describes the amplitude A and initial phase θ, *not* the frequency.

Though a nice and very convenient mathematical description, complex-valued signals do not exist in reality. Physical systems always interact with real signals. To obtain a real signal from a complex signal, we can simply take only the real part of the complex signal:

$$x(t) = \mathrm{Re}(\tilde{x}(t)) = A\cos(\omega_0 t + \theta) \tag{C.12}$$

or exploit the symmetry with the complex conjugate, see Figure C.4 and (C.2):

$$x(t) = \frac{1}{2}\tilde{x}(t) + \frac{1}{2}\tilde{x}^*(t) = A\cos(\omega_0 t + \theta) \tag{C.13}$$

D

Quantization

Figure 2.2 in Chapter 2 shows all four possible appearances of a signal. The independent variable time t can be either continuous (left column) or discrete (right column), and the signal value x can be either continuous (top row) or discrete (bottom row). The signal on top-left is referred to as an *analog* signal, e.g., a voltage from an electrical device, the signal at bottom-right as a *digital* signal. The latter follows from the first by sampling and quantization.

The process of *sampling*, going from left to right in Figure 2.2, is covered extensively in Chapter 9. Continuous time $t \in \mathbb{R}$ is converted to discrete time $t = n\Delta t$ with $n \in \mathbb{Z}$, see (9.1).

The process of *quantization*, going from top to bottom in Figure 2.2, is discussed here. Continuous signal value $x \in \mathbb{R}$ is converted to discrete signal value $x_q = q\Delta x$, with $q \in \mathbb{Z}$ and $\Delta x \in \mathbb{R}$ the step size or increment in signal value, and $x_q(t)$ denoting the quantized signal (in continuous time). Step size Δx determines the *resolution* of the quantized signal; a large step size yields a poor resolution (much signal 'detail' may get lost in the quantization).

A simple and straightforward example of quantization is rounding to the nearest integer. The value of signal $x(t)$ is a real number $x \in \mathbb{R}$, and it gets converted, or mapped to $x_q \in \mathbb{Z}$. In this case $q \in \mathbb{Z}$ and $\Delta x = 1$. All real values in the interval $x \in [16.50, 17.49)$ get mapped to integer $x_q = 17$. Clearly, quantization is a nonlinear, irreversible process; a range of values for x gets mapped to a single value for x_q.

Similar to restricting the time duration of the signal, from infinite duration to T or T_{meas}, done in Chapter 8, we also restrict the range of signal values in practice, from an expected minimum signal value to an expected maximum signal value, the *dynamic range* D:

$$D = \max(x(t)) - \min(x(t)) \tag{D.1}$$

We also restrict or truncate the countable set from $q = -\infty$ to ∞ to a set with a finite number of values $q = 0, \ldots, Q-1$, typically with $Q = 2^p$, such that in total there are $Q = 2^p$ possible values to represent the signal. This last step is referred to as *encoding*, with p the number of bits (binary 0/1-values) to represent the signal value in computer memory. With $p = 3$ there are $Q = 2^3 = 8$ possible signal values, which are stored using 3 bits. Note that the term quantization is often used to refer to both quantization *and* encoding.

When the number of bits p is known, the *quantization step size* Δx is given by:

$$\Delta x = \frac{D}{2^p} \tag{D.2}$$

For the example signal $x(t)$ of Figure 2.2, we assume that its values lie in between -3 and 3: the dynamic range $D = 6$. Figure D.1 illustrates the quantization process when $p = 3$

bits would be available: we allow for $Q = 2^3 = 8$ possible (discrete) signal values, the step size $\Delta x = \frac{6}{8} = 0.75$. Discretization takes place along the vertical axis. Usually the quantization is uniform (equispaced), Δx is constant. Quantized signal $x_q(t)$ is still continuous in time, changes from one quantized signal level to another may occur at any time.

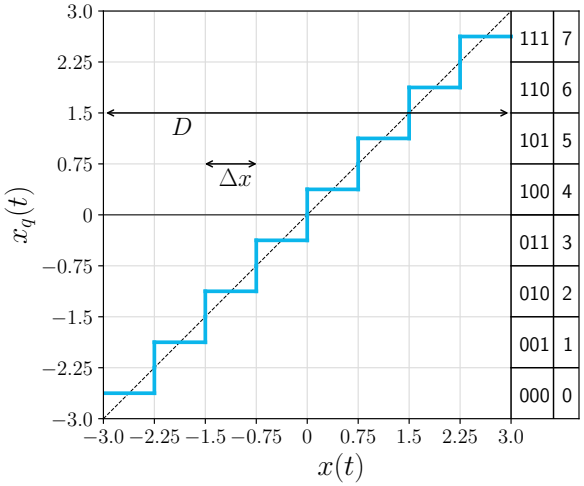


Figure D.1: The process of quantization: (input) signal value $x(t)$ is along the horizontal axis, covering the dynamic range D, the (output) quantized signal value $x_q(t)$ along the vertical axis.

In the example, we chose to deploy the dynamic range symmetrically about $x = 0$, from -3 to 3. Offering signal values outside this range would lead to saturation: the quantized signal is clipped to its minimum or maximum value, here, respectively, -2.625 or 2.625.

We introduce signal value offset x_q^0, set to $x_q^0 = -2.625$ in this example, to allow for $q = 0, \ldots, 7$, according to:

$$x_q = x_q^0 + q\Delta x \qquad \text{with } q = 0, \ldots, Q-1$$

omitting the time-argument, as this holds for both continuous and discrete time. These x_q-values represent the center or mid-values of the quantization intervals, see Figure D.1. The corresponding thresholds or boundaries of the interval are $[x_q^0 + q\Delta x - \frac{\Delta x}{2}, x_q^0 + q\Delta x + \frac{\Delta x}{2})$.

The quantization error is the difference between original signal value x and the quantized signal value x_q. The magnitude of this error is half the step size $\frac{\Delta x}{2}$ at most, much similar to a round-off error. Usually quantization is considered to add a (zero mean) contribution to the signal's noise. In this book we assume that a high resolution is used for quantization (large value for p, giving very small steps Δx), and hence that the effect is negligible.

EX D.1

Assume a voltage signal $x(t) = 2 + 3\cos(2\pi t + 0.5)$ which needs to be quantized with a step size of at most 0.05 V. What is the minimum number of bits needed?

Solution The dynamic range D of this signal is 6 V, as the signal oscillates between -1 and $+5$ V. Dividing D by the required step size $\Delta x = 0.05$ yields 120 signal levels. With $p = 7$ bits we get $Q = 2^p = 128$ possibilities, which is just sufficient, it leads to a step size Δx of $\frac{6}{128} = 0.0469$ V. Taking $p = 6$ bits would yield a step size of $\Delta x = \frac{6}{64} = 0.094$ V, which is too large. The minimum number of bits p is 7.

E

Fourier transform pairs

Table E.1: Common Fourier transform pairs.

$x(t)$	$X(f)$	remark
unit pulse and sinc		
$\Pi\left(\frac{t}{\tau}\right)$	$\tau\,\mathrm{sinc}(\tau f)$	see (5.11)
$\tau\,\mathrm{sinc}(\tau t)$	$\Pi\left(\frac{f}{\tau}\right)$	see (6.13)
unit triangular and sinc^2		
$\Lambda\left(\frac{t}{\tau}\right)$	$\tau\,\mathrm{sinc}^2(\tau f)$	duality (see (6.5)) with next pair
$\tau\,\mathrm{sinc}^2(\tau t)$	$\Lambda\left(\frac{f}{\tau}\right)$	see (7.7) with both $y(t)$ and $Y(f)$ divided by factor τ
exponential		
$e^{-\alpha t}u(t)$ with $\alpha > 0$	$\frac{1}{\alpha + j2\pi f}$	see (5.13)
$te^{-\alpha t}u(t)$ with $\alpha > 0$	$\frac{1}{(\alpha + j2\pi f)^2}$	see Exercise e.5-5(a)
$e^{-\alpha\lvert t\rvert}$ with $\alpha > 0$	$\frac{2\alpha}{\alpha^2 + (2\pi f)^2}$	see Exercise e.6-7(b)
$e^{-\pi(\frac{t}{\tau})^2}$	$\tau e^{-\pi(\tau f)^2}$	see Exercise e.5-5(d)
in-the-limit		
$\delta(t)$	1	see (5.14)
1	$\delta(f)$	see (5.15)
$\delta(t - t_0)$	$e^{-j2\pi f t_0}$	time shift and pair $\{\delta(t), 1\}$
$e^{j2\pi f_0 t}$	$\delta(f - f_0)$	duality with previous pair
$\cos(2\pi f_0 t)$	$\frac{1}{2}\Big(\delta(f + f_0) + \delta(f - f_0)\Big)$	see (5.16)
$\sin(2\pi f_0 t)$	$j\frac{1}{2}\Big(\delta(f + f_0) - \delta(f - f_0)\Big)$	see (5.17)

F

Discrete convolution

In Chapter 7, the concept of convolution of two continuous-time functions was introduced, (7.1), resulting in a new continuous-time function. In a similar way, the convolution of two discrete-time sequences x_n and h_n (with index n, of infinite length, and pertaining to the same time interval Δt) can be defined, resulting in a new discrete-time sequence y_n:

$$y_n = x_n * h_n \tag{F.1}$$

again with the asterisk '$*$'-symbol, a short-hand notation for the following summation:

$$y_n = \sum_{k=-\infty}^{\infty} x_k h_{n-k} \tag{F.2}$$

The convolution integral, (7.1), has been replaced by a discrete summation, with k as the running variable.

Similar to the convolution in continuous time, Section 7.1, a discrete-time convolution can be split into five steps:

1. Consider both sequences with index k, hence x_k and h_k,
2. Mirror or flip sequence h_k about the origin ($k = 0$) to get h_{-k},
3. Shift sequence h_k by n to yield h_{n-k},
4. Compute output sequence y_n for one specific index value n by summing, over index k, product $x_k h_{n-k}$ from $k = -\infty$ to ∞, and
5. Repeat steps 3 and 4 for every value of n,

to obtain the full output sequence y_n, as a function of time index n.

When the two discrete-time sequences have been obtained through sampling (with sampling interval Δt), and we would like to maintain the analogy with continuous-time convolution (7.1), the integral is to be discretized. The above convolution, (F.2), becomes:

$$y_n = \Delta t \sum_{k=-\infty}^{\infty} x_k h_{n-k} \tag{F.3}$$

In this case, when convolving with a Dirac delta function $h(t) = \delta(t)$ in discrete time, we should use $h_n = \frac{1}{\Delta t}$ for $n = 0$ and $h_n = 0$ otherwise, rather than the commonly-defined

Kronecker[1] delta function, $\delta_n = 1$ for $n = 0$ and $\delta_n = 0$ otherwise. This is done to maintain the analogy with the continuous-time Dirac delta function having unit area, (B.9).

EX F.1

Convolve discrete sequence $x_n = [1, 2, 3]$ with sequence $h_n = [4, 5, 6]$, according to $y_n = x_n * h_n$, (F.2).

Solution Because x_n and h_n are *finite* length sequences (as usual in practice), the summation bounds for k in (F.2), as well as the shift n in the third step above, have to be adjusted to the lengths of x_n and h_n, here $N = 3$.

In Table F.1 sequence h_n has been mirrored, and in Table F.2 it has been shifted, in this case by $n = 1$; the convolution output is $y_{n=1} = 5 + 8 = 13$ (there is 'overlap' only for $k = 0$ and $k = 1$). The full result is shown in Table F.3.

Convolution of discrete sequences can be done in Python through `numpy.convolve`; it can output either the 'full' result, as in Table F.3, or the so-called central or 'same' part $y_n = [13, 28, 27]$ for index values $n = 0, 1, 2$ just like the given input.

Table F.1: Convolution of discrete sequences x_n and h_n: sequence h_n has been mirrored.

k	-2	-1	0	1	2
x_k			1	2	3
h_{-k}	6	5	4		

Table F.2: Convolution of sequences x_n and h_n: mirrored sequence h_n has been shifted by $n = 1$. The bottom row lists the products $x_k h_{n-k}$, which – summed over k – yields the convolution output for index $n = 1$.

k	-2	-1	0	1	2
x_k			1	2	3
h_{1-k}		6	5	4	
$x_k h_{1-k}$			5	8	

Table F.3: Result of convolution of sequences x_n and h_n: $y_n = x_n * h_n$.

n	0	1	2	3	4
y_n	4	13	28	27	18

[1] After the German mathematician L. Kronecker (1823-1891).

G

DTFT addendum

In Chapter 11 the Discrete-Time Fourier Transform (DTFT), (11.5), and its inverse, (11.8), were introduced and discussed. In this appendix, a missing step in the DTFT proof is provided, the proof of the inverse DTFT is given, and some DTFT pairs are presented.

G.1 Addition to proof of DTFT

In Section 11.1 the Discrete-Time Fourier Transform was derived, where (11.4) introduced the complex exponential Fourier series coefficients Ψ_m of signal $\psi(t)$, stating without proof that:

$$\Psi_m = f_s\, Y(mf_s), \quad m \in \mathbb{Z} \tag{G.1}$$

with sampling frequency f_s, and $Y(f)$ the Fourier transform of $y(t)$, evaluated at frequency $f = mf_s$. The proof is as follows.

Proof Since $\psi(t)$ is periodic with Δt, its fundamental frequency is f_s. We can compute the complex exponential Fourier series coefficients of $\psi(t)$, (4.3):

$$\Psi_m = \frac{1}{\Delta t}\int_0^{\Delta t} \psi(t) e^{-j2\pi m f_s t}\,\mathrm{d}t = f_s \int_0^{\Delta t} \sum_{n=-\infty}^{\infty} y(t+n\Delta t) e^{-j2\pi m f_s t}\,\mathrm{d}t$$

$$= f_s \sum_{n=-\infty}^{\infty} \int_0^{\Delta t} y(t+n\Delta t) e^{-j2\pi m f_s t}\,\mathrm{d}t$$

Substitute $t + n\Delta t = \tau$, then $t = \tau - n\Delta t$ and $\mathrm{d}\tau = \mathrm{d}t$; when $t = 0$, $\tau = n\Delta t$; when $t = \Delta t$, $\tau = \Delta t + n\Delta t = (n+1)\Delta t$:

$$\Psi_m = f_s \sum_{n=-\infty}^{\infty} \int_{n\Delta t}^{(n+1)\Delta t} y(\tau) e^{-j2\pi m f_s(\tau - n\Delta t)}\,\mathrm{d}\tau$$

The expression $e^{-j2\pi m f_s(\tau - n\Delta t)}$ in the integrand can be written as:

$$e^{-j2\pi m f_s(\tau - n\Delta t)} = e^{-j2\pi m f_s \tau} e^{j2\pi m n \overbrace{f_s \Delta t}^{=1}} = e^{-j2\pi m f_s \tau} \underbrace{e^{j2\pi m n}}_{=1 \;\forall\; m,n \in \mathbb{Z}} = e^{-j2\pi m f_s \tau}$$

Hence:

$$\Psi_m = f_s \sum_{n=-\infty}^{\infty} \int_{n\Delta t}^{(n+1)\Delta t} y(\tau) e^{-j2\pi m f_s \tau}\,\mathrm{d}\tau = f_s \int_{-\infty}^{\infty} y(\tau) e^{-j2\pi m f_s \tau}\,\mathrm{d}\tau = f_s\, Y(mf_s)$$

G.2 Proof of inverse DTFT

In Section 11.3 the inverse Discrete-Time Fourier Transform was given:

$$x_n = \text{iDTFT}\{X_{\Delta t}(f)\} = \int_{f_s} X_{\Delta t}(f) e^{j2\pi f \Delta t n} \, \mathrm{d}f \tag{11.8}$$

The proof is as follows.

Proof We copy the expression for the DTFT (11.5) but, for the sake of the proof we use time index ℓ instead of n (both n and $\ell \in \mathbb{Z}$).

$$X_{\Delta t}(f) = \Delta t \sum_{\ell=-\infty}^{\infty} x_\ell e^{-j2\pi f \Delta t \ell} \tag{11.5}$$

Insert this expression for $X_{\Delta t}(f)$ in the definition of the inverse DTFT, (11.8):

$$x_n = \int_{f_s} \left(\Delta t \sum_{\ell=-\infty}^{\infty} x_\ell e^{-j2\pi f \Delta t \ell} \right) e^{j2\pi f \Delta t n} \, \mathrm{d}f = \Delta t \int_{f_s} \sum_{\ell=-\infty}^{\infty} x_\ell e^{-j2\pi f \Delta t (\ell - n)} \, \mathrm{d}f$$

$$= \Delta t \sum_{\ell=-\infty}^{\infty} x_\ell \int_{f_s} e^{-j2\pi f \Delta t (\ell - n)} \, \mathrm{d}f$$

When $\ell = n$ the integral on the right-hand side equals f_s, and when $\ell \neq n$ the integral is zero. When, for instance, $\ell - n = 1$, we integrate the complex exponential function $e^{-j2\pi f \Delta t} = e^{-j2\pi \frac{f}{f_s}}$ from $f = 0$ to f_s: the complex exponential function completes one cycle in the complex plane and the result of the integration is zero. The same is true for other values of $(\ell - n)$, then the complex exponential makes $\ell - n$ cycles, the integration of which is also zero. Hence, only one term of the summation remains, and we obtain:

$$x_n = \Delta t \, x_n \, f_s = x_n$$

G.3 DTFT pairs

Because the DTFT plays a marginal role in this book, only a few DTFT transform pairs are given. Table G.1 lists some common pairs, stating the expression for $X_{\Delta t}(f)$ for frequencies $-\frac{f_s}{2} \leq f < \frac{f_s}{2}$. Recall that $X_{\Delta t}(f)$ is periodic with f_s.

When consulting other resources, for more DTFT pairs, care must be taken regarding the scaling of the DTFT, and especially the scaling of the Dirac delta function in discrete time, when applicable. Consider for example the DTFT of a (properly sampled, i.e., $f_0 < \frac{f_s}{2}$) cosine function, the mathematical expression of which reads:

$$x_n = \cos[2\pi f_0 \Delta t n]$$

With $\Omega = 2\pi f \Delta t$ and $\Omega_0 = 2\pi f_0 \Delta t$, [8] states that its DTFT (within the range $-\pi \leq \Omega < \pi$, which is, because $f = \frac{\Omega}{2\pi \Delta t} = \frac{\Omega}{\pi} \frac{f_s}{2}$, equivalent to $-\frac{f_s}{2} \leq f < \frac{f_s}{2}$):

$$X(\Omega) = \pi \left(\delta(\Omega + \Omega_0) + \delta(\Omega - \Omega_0) \right)$$

Our definition of the DTFT, (11.5), *scales* this $X(\Omega)$ with Δt, and Dirac pulses need to be

Table G.1: Common DTFT pairs.

x_n	$X_{\Delta t}(f)$ for $-\frac{f_s}{2} \le f < \frac{f_s}{2}$	remarks
pulse, sinc and window		
$\begin{cases} 1 & -M \le n \le M \\ 0 & \text{else} \end{cases}$	$\Delta t \, \frac{\sin\left(2\pi f \Delta t \left(\frac{2M+1}{2}\right)\right)}{\sin\left(\frac{2\pi f \Delta t}{2}\right)}$	pulse, $M \ge 0$, length $2M+1$; see proof below; note that $X_{\Delta t}(0) = \Delta t(2M+1)$
$f_0 \, \mathrm{sinc}[f_0 \Delta t n]$	$\Pi\left(\frac{f}{f_0}\right)$	see proof below; $0 < f_0 < \frac{f_s}{2}$
$\begin{cases} 1 & 0 \le n \le M-1 \\ 0 & \text{else} \end{cases}$	$\Delta t \, \frac{\sin\left(2\pi f \Delta t \left(\frac{M}{2}\right)\right)}{\sin\left(\frac{2\pi f \Delta t}{2}\right)} e^{-j2\pi f \Delta t \left(\frac{M-1}{2}\right)}$	shifted pulse, length $M \ge 1$; see Exercise e.11-1; note that $X_{\Delta t}(0) = \Delta t M$
in-the-limit		
δ_n	Δt	direct substitution in (11.5)
$1 \; \forall \; n$	$\delta(f)$	inverse proof, direct substitution in (11.8)
δ_{n-n_0}	$\Delta t \, e^{-j2\pi f \Delta t n_0}$	direct substitution in (11.5)
$e^{j2\pi f_0 \Delta t n}$	$\delta(f - f_0)$	inverse proof, use (11.8); $-\frac{f_s}{2} < f_0 < \frac{f_s}{2}$
$\cos[2\pi f_0 \Delta t n]$	$\frac{1}{2}\left(\delta(f+f_0) + \delta(f-f_0)\right)$	inverse proof, use (11.8); $f_0 < \frac{f_s}{2}$
$\sin[2\pi f_0 \Delta t n]$	$j\frac{1}{2}\left(\delta(f+f_0) - \delta(f-f_0)\right)$	inverse proof, use (11.8); $f_0 < \frac{f_s}{2}$

re-scaled when changing the running variable, (B.13). Hence:

$$\begin{aligned} X_{\Delta t}(f) &= \Delta t \, \pi \left(\delta(\Omega + \Omega_0) + \delta(\Omega - \Omega_0)\right) = \Delta t \, \pi \frac{1}{2\pi \Delta t} \left(\delta(f + f_0) + \delta(f - f_0)\right) \\ &= \frac{1}{2} \left(\delta(f + f_0) + \delta(f - f_0)\right) \end{aligned}$$

which equals $X(f)$, (5.16), for this band of frequencies.

Note that in Table G.1 the DTFT in-the-limit of the Kronecker delta function δ_n ($\delta_n = 1$ for $n = 0$ and $\delta_n = 0$ for other n) yields Δt. Recall from Appendix F that in situations where we would like to maintain the analogy with the continuous-time Dirac delta function, having unit area, we should multiply δ_n by $\frac{1}{\Delta t}$. The DTFT of the *scaled* Kronecker delta, $\frac{1}{\Delta t}\delta_n$, equals 1, see (5.14). The same holds for the DTFT of the shifted Kronecker delta function δ_{n-n_0}.

Proof of DTFT of pulse function (width $2M$, an even function in n, with $2M+1$ (odd number) samples):

Proof Direct substitution of the pulse function in the DTFT, (11.5) results in:

$$X_{\Delta t}(f) = \Delta t \sum_{n=-\infty}^{\infty} x_n e^{-j2\pi f \Delta t n} = \Delta t \sum_{n=-M}^{M} e^{-j2\pi f \Delta t n}$$

Substitute $\ell = n + M$, then $n = \ell - M$, we get:

$$X_{\Delta t}(f) = \Delta t \sum_{\ell=0}^{2M} e^{-j2\pi f \Delta t(\ell - M)} = \Delta t e^{j2\pi f \Delta t M} \sum_{\ell=0}^{2M} e^{-j2\pi f \Delta t \ell}$$

Define $2M = N - 1$, then $N = 2M + 1$ and $M = \frac{N-1}{2}$:

$$X_{\Delta t}(f) = \Delta t\, e^{j2\pi f \Delta t\left(\frac{N-1}{2}\right)} \sum_{\ell=0}^{N-1} \left(e^{-j2\pi f \Delta t}\right)^{\ell}$$

on the right-hand side we obtain the geometric series identity, (A.22), with $a = e^{-j2\pi f \Delta t}$ and $k = \ell$. Hence:

$$X_{\Delta t}(f) = \Delta t\, e^{j2\pi f \Delta t\left(\frac{N-1}{2}\right)} \left(\frac{1 - e^{-j2\pi f \Delta t N}}{1 - e^{-j2\pi f \Delta t}}\right) \tag{G.2}$$

The denominator can be written as $e^{-j\pi f \Delta t}\left(e^{j\pi f \Delta t} - e^{-j\pi f \Delta t}\right) = e^{-j\pi f \Delta t} 2j \sin(\frac{2\pi f \Delta t}{2})$. Similarly, the numerator can be written as $e^{-j\pi f \Delta t N} 2j \sin(\frac{2\pi f \Delta t N}{2})$. Substitute in (G.2):

$$X_{\Delta t}(f) = \Delta t\, e^{j2\pi f \Delta t\left(\frac{N-1}{2}\right)} \frac{e^{-j\pi f \Delta t N} 2j \sin(\frac{2\pi f \Delta t N}{2})}{e^{-j\pi f \Delta t} 2j \sin(\frac{2\pi f \Delta t}{2})} = \Delta t \frac{\sin\left(2\pi f \Delta t \left(\frac{N}{2}\right)\right)}{\sin\left(\frac{2\pi f \Delta t}{2}\right)} \tag{G.3}$$

which, with $N = 2M + 1$, completes the proof. $X_{\Delta t}(f)$ is real, even in f.[1]

Proof of DTFT of discrete-time sinc function:

Proof Use the inverse DTFT, (11.8):

$$\begin{aligned} x_n &= \int_{f_s} X_{\Delta t}(f) e^{j2\pi f \Delta t n}\, \mathrm{d}f = \int_{f_s} \Pi\left(\frac{f}{f_0}\right) e^{j2\pi f \Delta t n}\, \mathrm{d}f = \int_{-\frac{f_0}{2}}^{\frac{f_0}{2}} e^{j2\pi f \Delta t n}\, \mathrm{d}f \\ &= \frac{1}{j2\pi \Delta t n}\left[e^{j2\pi f \Delta t n}\right]_{-\frac{f_0}{2}}^{\frac{f_0}{2}} = \frac{1}{j2\pi \Delta t n}\left(e^{j\pi f_0 \Delta t n} - e^{-j\pi f_0 \Delta t n}\right) = \frac{2j \sin[\pi f_0 \Delta t n]}{j2\pi \Delta t n} \\ &= \frac{\sin[\pi f_0 \Delta t n]}{\pi \Delta t n} \frac{f_0}{f_0} = f_0 \frac{\sin[\pi f_0 \Delta t n]}{\pi f_0 \Delta t n} = f_0\, \mathrm{sinc}[f_0 \Delta t n] \end{aligned}$$

A real, even function in n, because $X_{\Delta t}(f)$ is real, even in f.

[1] The right-hand term in (G.3), without the Δt, is known as the Dirichlet kernel.

H

DFT addendum

In Chapter 12 the Discrete Fourier Transform (DFT), (12.5), and its inverse, (12.6), were introduced and discussed. In this appendix a formal proof of the inverse DFT is given, a matrix interpretation of the DFT presented and some common DFT pairs discussed.

H.1 Proof of inverse DFT

Proof We copy the expression for the DFT (12.5), but, for the sake of the proof we use time index ℓ instead of n:

$$X_k = \Delta t \sum_{\ell=0}^{N-1} x_\ell e^{-jk\frac{2\pi}{N}\ell} \tag{H.1}$$

Insert this expression for X_k in the definition of the inverse DFT (12.6):

$$x_n = \frac{1}{\Delta t}\frac{1}{N}\sum_{k=0}^{N-1}\left(\Delta t \sum_{\ell=0}^{N-1} x_\ell e^{-jk\frac{2\pi}{N}\ell}\right)e^{jk\frac{2\pi}{N}n} = \frac{1}{N}\sum_{k=0}^{N-1}\left(\sum_{\ell=0}^{N-1} x_\ell e^{-jk\frac{2\pi}{N}(\ell-n)}\right) \tag{H.2}$$

where the exponential $e^{jk\frac{2\pi}{N}n}$ could be put into the second summation (over ℓ), because it is not a function of ℓ. Changing the order of summation yields:

$$x_n = \frac{1}{N}\sum_{\ell=0}^{N-1}\sum_{k=0}^{N-1} x_\ell e^{jk\frac{2\pi}{N}(n-\ell)} = \frac{1}{N}\sum_{\ell=0}^{N-1} x_\ell \sum_{k=0}^{N-1} e^{jk\frac{2\pi}{N}(n-\ell)} \tag{H.3}$$

where x_ℓ could be taken out of the second summation (over k) because x_ℓ is not a function of k. To compute the summation over k at the right-hand side of (H.3), we use the geometric series identity:

$$\sum_{k=0}^{N-1} a^k = \begin{cases} \dfrac{1-a^N}{1-a} & \text{for } a \neq 1 \\ N & \text{for } a = 1 \end{cases} \tag{A.22}$$

With $a = e^{j\frac{2\pi}{N}(n-\ell)}$ in (H.3), this leads to:

$$x_n = \frac{1}{N}\sum_{\ell=0}^{N-1} x_\ell \frac{1-e^{j2\pi(n-\ell)}}{1-e^{j\frac{2\pi}{N}(n-\ell)}}$$

with $n, \ell \in \{0, \ldots, N-1\}$. Assume that $\ell \neq n$, then $e^{j2\pi(n-\ell)} = 1$, and $e^{j\frac{2\pi}{N}(n-\ell)} \neq 1$, as $n - \ell \in \{-(N-1), \ldots, -1, 1, \ldots, N-1\}$: the numerator becomes zero, the denominator does not, the expression is zero hence the summation is zero for $\ell \neq n$.

This leaves only the term for $\ell = n$, then $a = 1$, the summation becomes N, we obtain:

$$x_n = \frac{1}{N}(x_n N) = x_n \tag{H.4}$$

The inverse DFT has been proven.

H.2 DFT matrix

Recall that the Discrete Fourier Transform turns N samples of real-valued signal $x(t)$, that is, x_n, with $n = 0, 1, \ldots, N-1$, into N numerical evaluations of DTFT $X_{\Delta t,T}(f)$, i.e., N complex numbers X_k, with $k = 0, 1, \ldots, N-1$. The computation of each single element X_k implies a summation over n in (12.5), $n = 0, 1, \ldots, N-1$, and this can be cast as the multiplication of row k of a matrix W with vector $(x_0, x_1, \ldots, x_{N-1})^T$:

$$X_k = \Delta t \begin{pmatrix} e^{-jk\frac{2\pi}{N}0} & e^{-jk\frac{2\pi}{N}1} & e^{-jk\frac{2\pi}{N}2} & \cdots & e^{-jk\frac{2\pi}{N}(N-1)} \end{pmatrix} \begin{pmatrix} x_0 \\ x_1 \\ x_2 \\ \vdots \\ x_{N-1} \end{pmatrix}$$

for $k = 0, \ldots, N-1$. The whole operation can then be described in a vector-matrix-vector form:

$$\begin{pmatrix} X_0 \\ X_1 \\ X_2 \\ \vdots \\ X_{N-1} \end{pmatrix} = \Delta t \underbrace{\begin{pmatrix} e^{-j0\frac{2\pi}{N}0} & e^{-j0\frac{2\pi}{N}1} & e^{-j0\frac{2\pi}{N}2} & \cdots & e^{-j0\frac{2\pi}{N}(N-1)} \\ e^{-j1\frac{2\pi}{N}0} & e^{-j1\frac{2\pi}{N}1} & e^{-j1\frac{2\pi}{N}2} & \cdots & e^{-j1\frac{2\pi}{N}(N-1)} \\ e^{-j2\frac{2\pi}{N}0} & e^{-j2\frac{2\pi}{N}1} & e^{-j2\frac{2\pi}{N}2} & \cdots & e^{-j2\frac{2\pi}{N}(N-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ e^{-j(N-1)\frac{2\pi}{N}0} & e^{-j(N-1)\frac{2\pi}{N}1} & e^{-j(N-1)\frac{2\pi}{N}2} & \cdots & e^{-j(N-1)\frac{2\pi}{N}(N-1)} \end{pmatrix}}_{= W} \begin{pmatrix} x_0 \\ x_1 \\ x_2 \\ \vdots \\ x_{N-1} \end{pmatrix} \tag{H.5}$$

Matrix W, referred to as 'the DFT-matrix', is square and has dimensions N by N. As both indices k (frequency) and n (time) run from 0 to $N-1$, the above matrix W is symmetric, hence $W = W^T$, the matrix equals its transpose.

In fact, with (12.5) and (12.6) we can verify that $W^{-1} = \frac{1}{N}W^H$, where the superscript 'H' denotes the *Hermitian* of a matrix, meaning that the elements of matrix W need to be transposed, denoted by 'T', *and* complex conjugated, denoted by '$*$'. The complex exponential $e^{-jk\frac{2\pi}{N}n}$ as element (k, n) of matrix W in (12.5) becomes $e^{jn\frac{2\pi}{N}k}$ as element (n, k) of matrix W^{-1} according to (12.6), with a factor of $\frac{1}{N}$ in front of the matrix (included in W^{-1}).

Applying the DFT (12.5) and then the inverse DFT (12.6) means that the original N time domain samples are returned, i.e., $W^{-1}W = I_N$. Multiplying (materialized by the sum over k) row n of W^{-1} by column ℓ of W, yields zero if $n \neq \ell$, and yields one if $n = \ell$.

Element (k, n) of matrix W reads $e^{-jk\frac{2\pi}{N}n}$, and element $((N-k), n)$ reads $e^{-j(N-k)\frac{2\pi}{N}n}$. The latter can be written as $e^{-jN\frac{2\pi}{N}n}e^{jk\frac{2\pi}{N}n} = e^{jk\frac{2\pi}{N}n}$, because $e^{-j2\pi n} = 1$ for all $n \in \mathbb{Z}$. This

means that $W_{k,n} = W^*_{(N-k),n}$, or, that the kth row of matrix W equals the complex conjugate of the $(N-k)$th row.

Consider $N = 6$ and evaluate the complex exponentials in matrix W, (H.5). EX H.1

Solution The resulting W-matrix reads:

$$W = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & \frac{1}{2} - j\frac{\sqrt{3}}{2} & -\frac{1}{2} - j\frac{\sqrt{3}}{2} & -1 & -\frac{1}{2} + j\frac{\sqrt{3}}{2} & \frac{1}{2} + j\frac{\sqrt{3}}{2} \\ 1 & -\frac{1}{2} - j\frac{\sqrt{3}}{2} & -\frac{1}{2} + j\frac{\sqrt{3}}{2} & 1 & -\frac{1}{2} - j\frac{\sqrt{3}}{2} & -\frac{1}{2} + j\frac{\sqrt{3}}{2} \\ 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & -\frac{1}{2} + j\frac{\sqrt{3}}{2} & -\frac{1}{2} - j\frac{\sqrt{3}}{2} & 1 & -\frac{1}{2} + j\frac{\sqrt{3}}{2} & -\frac{1}{2} - j\frac{\sqrt{3}}{2} \\ 1 & \frac{1}{2} + j\frac{\sqrt{3}}{2} & -\frac{1}{2} + j\frac{\sqrt{3}}{2} & -1 & -\frac{1}{2} - j\frac{\sqrt{3}}{2} & \frac{1}{2} - j\frac{\sqrt{3}}{2} \end{pmatrix}$$

The above-mentioned property of W, namely that its kth row equals the complex conjugate of the $N-k$th row clearly shows for rows $k = 1$ and $k = N - 1 = 5$, and for rows $k = 2$ and $k = N - 2 = 4$. Note that for $k = 3$ we have $N - k = 3$, so that this row equals the complex conjugate of itself. Recall that, in discrete-time, counting starts at zero.

H.3 DFT pairs

Some common DFT pairs are listed in Table H.1. In this table, the Kronecker delta function is used in the discrete time (δ_n) and discrete frequency (δ_k) domains. Further note that in the DFT of the complex exponential function, the cosine and sine functions, we consider only the frequency f_0 equal to ℓ times the frequency step size Δf. The function then completes an integer number ℓ cycles in the time duration $T = N\Delta t$. With $f_0 = \ell\Delta f = \ell\frac{1}{N\Delta t}$ the DFT shows no spectral leakage, see Example 12.5. For instance, the cosine function equals:

$$\cos[2\pi f_0 \Delta t n] \;=\; \cos[2\pi(\ell\frac{1}{N\Delta t})\Delta t n] \;=\; \cos[\ell\frac{2\pi}{N}n] \tag{H.6}$$

Table H.1 shows that the DFT of (H.6) equals:

$$X_k = \frac{\Delta t N}{2}\left(\delta_{k-\ell} + \delta_{k-(N-\ell)}\right) \tag{H.7}$$

meaning that all X_k are zero, except for $k = \ell$, specifying the 'positive frequency' $\ell\frac{2\pi}{N}$, and for $k = (N-\ell)$, specifying the 'negative frequency' $(N-\ell)\frac{2\pi}{N}$. This follows from:

$$\begin{aligned} \cos[\ell\frac{2\pi}{N}n] &= \frac{1}{2}\left(e^{j\ell\frac{2\pi}{N}n} + e^{-j\ell\frac{2\pi}{N}n}\right) \\ &= \frac{1}{2}\left(e^{j\ell\frac{2\pi}{N}n} + e^{-j\ell\frac{2\pi}{N}n}\underbrace{e^{N\frac{2\pi}{N}n}}_{=1}\right) \\ &= \frac{1}{2}\left(e^{j\ell\frac{2\pi}{N}n} + e^{j(N-\ell)\frac{2\pi}{N}n}\right) \end{aligned}$$

Table H.1: Common DFT pairs.

x_n for $0 \le n \le N-1$	X_k for $0 \le k \le N-1$	remarks
window, periodic sinc		
$\begin{cases} 1 & 0 \le n \le M-1 \\ 0 & \text{else} \end{cases}$	$\Delta t \, \dfrac{\sin\left[k\frac{2\pi}{N}\left(\frac{M}{2}\right)\right]}{\sin\left[\frac{k\frac{2\pi}{N}}{2}\right]} e^{-jk\frac{2\pi}{N}\left(\frac{M-1}{2}\right)}$	shifted pulse, length $1 \le M \le N$; see Exercise *e*.12-1; note that $X_0 = \Delta t M$
$\frac{1}{N\Delta t} \dfrac{\sin\left[n\frac{2\pi}{N}\left(\frac{M}{2}\right)\right]}{\sin\left[\frac{n\frac{2\pi}{N}}{2}\right]} e^{-jn\frac{2\pi}{N}\left(\frac{M-1}{2}\right)}$	$\begin{cases} 1 & 0 \le k \le M-1 \\ 0 & \text{else} \end{cases}$	periodic, phase-rotated sinc; dual of DFT of window; note that $x_0 = \frac{M}{N\Delta t}$
in-the-limit		
δ_n	$\Delta t \;\forall\; k$	direct substitution in (12.5)
$1 \;\forall\; n$	$\Delta t N \, \delta_k$	inverse proof, direct substitution in (12.6)
δ_{n-n_0}	$\Delta t \, e^{-jk\frac{2\pi}{N}n_0}$	direct substitution in (12.5); $0 \le n_0 \le N-1$
$e^{j\ell\frac{2\pi}{N}n}$	$\Delta t N \, \delta_{k-\ell}$	inverse proof, use (12.6); $1 \le \ell \le N-1$
$\cos[\ell\frac{2\pi}{N}n]$	$\frac{\Delta t N}{2}\left(\delta_{k-\ell} + \delta_{k-(N-\ell)}\right)$	inverse proof, use (12.6); $1 \le \ell \le \frac{N}{2}-1$ (even N); $1 \le \ell \le \frac{N-1}{2}$ (odd N); see Figure 12.4
$\sin[\ell\frac{2\pi}{N}n]$	$-j\frac{\Delta t N}{2}\left(\delta_{k-\ell} - \delta_{k-(N-\ell)}\right)$	inverse proof, use (12.6); $1 \le \ell \le \frac{N}{2}-1$ (even N); $1 \le \ell \le \frac{N-1}{2}$ (odd N); see Figure 12.4

I

Spectral analysis in practice: full procedure

In this appendix we summarize all steps, starting from a theoretical (deterministic) continuous-time signal, to spectral analysis in practice. That is, we discuss the full procedure from measuring continuous-time signal $x(t)$ for T seconds, to computing the Discrete Fourier Transform (DFT) coefficients X_k, which typically form the basis for further spectral analysis, such as estimating the signal's Power Spectral Density (PSD). We go through the following steps:

1. Apply time window on continuous-time signal $x(t)$, yielding $x_w(t)$,
2. Sample $x_w(t)$ to obtain $x_{sw}(t)$, and
3. Compute and evaluate Fourier transform $X_{sw}(f)$, at discrete frequencies, in the frequency domain.

I.1 Window - time domain

As an example we take (again) signal $x(t) = \text{sinc}^2(t)$, with as its Fourier transform $X(f) = \Lambda(f)$, (7.7) (here we set $\tau = 1$); both are shown in Figure I.1.

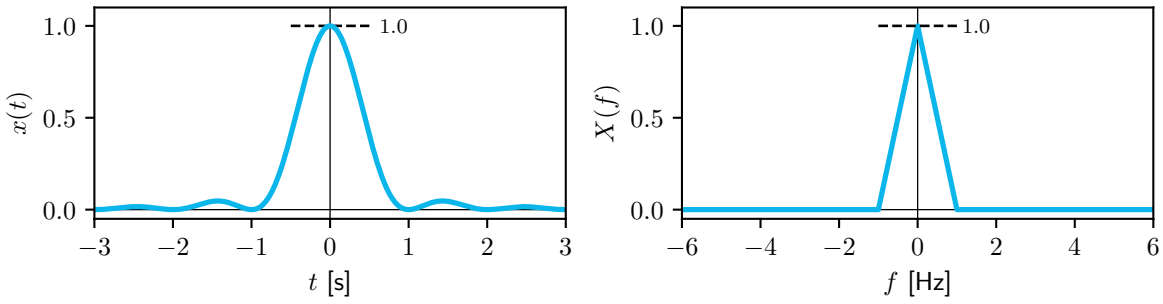


Figure I.1: Aperiodic continuous-time signal $x(t) = \text{sinc}^2(t)$ at left, and its Fourier transform $X(f) = \Lambda(f)$ at right. Because signal $x(t)$ is real and even, its Fourier transform $X(f)$ is also real and even.

In the first step we apply a rectangular window with duration $T = 4$ s. The window function $w(t)$ and its Fourier transform $W(f)$ are shown in Figure I.2.

The time-windowed signal $x_w(t)$ and its Fourier transform $X_w(f)$ are shown in Figure I.3. The consequence of applying the time window is leakage: the peak gets rounded, and small

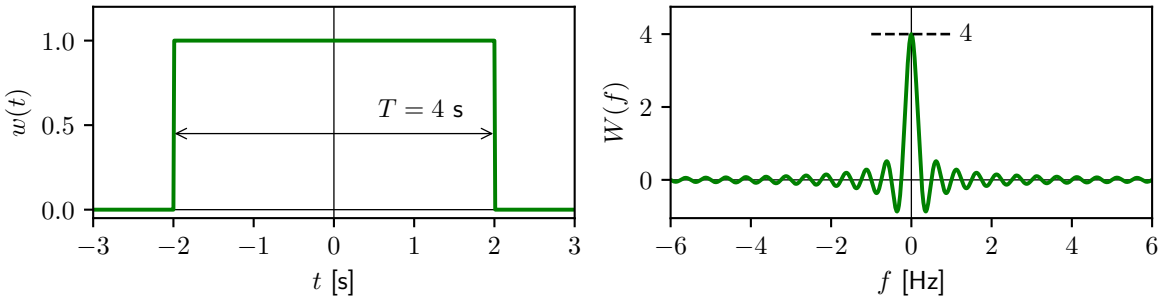


Figure I.2: Rectangular window $w(t)$ at left, with $T = 4$ s, and its Fourier transform $W(f)$ at right. Because signal $w(t)$ is real and even, its Fourier transform $W(f)$ is also real and even.

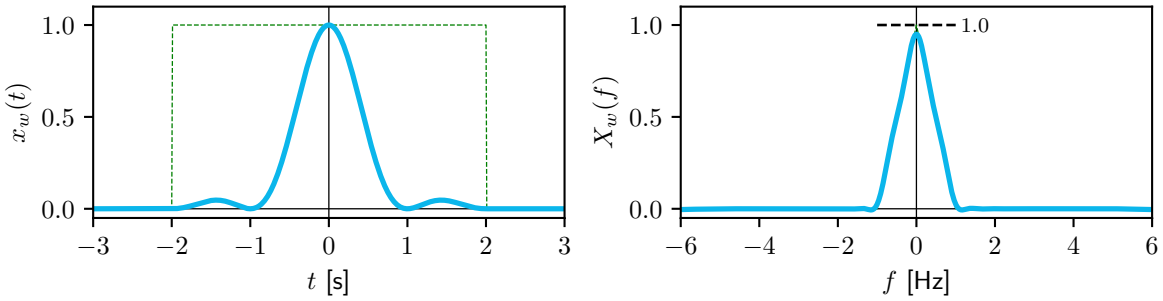


Figure I.3: Time-windowed signal $x_w(t) = w(t)x(t)$ at left, and its Fourier transform $X_w(f) = W(f) * X(f)$ at right. Both $x_w(t)$ and $X_w(f)$ are real and even.

'wobbles' appear in the tails, see Figure I.3 at right, and carefully compare with Figure I.1 at right. Leakage is covered in Chapter 8 and similarly shown in Figure 8.3 at right, though there a shorter window was used, $T = 3$ s, as to emphasize the effects of leakage.

I.2 Sampling - time domain

With (9.7) in Chapter 9 we found the Fourier series expression for the *periodic* impulse train function $p(t)$, (9.4), to be:

$$p(t) = \sum_{k=-\infty}^{k=\infty} e^{j2\pi k f_s t} \tag{I.1}$$

In Chapter 5 we found the Fourier transform (in-the-limit) of $x(t) = 1$ to be $X(f) = \delta(f)$, (5.15). Applying the frequency translation theorem of Chapter 6, (6.6), to $x(t) = 1$ yields:

$$x(t)e^{j2\pi f_0 t} = e^{j2\pi f_0 t} \quad \overset{\mathcal{F}}{\leftrightarrow} \quad X(f) = \delta(f - f_0) \tag{I.2}$$

We apply this result to every term of the above Fourier series expression for the periodic impulse train function, and, using linearity, obtain its Fourier transform:

$$P(f) = \sum_{k=-\infty}^{\infty} \delta(f - k f_s) \tag{I.3}$$

We find that the Fourier transform of a periodic impulse train function in the time domain, (9.3), is *also* a periodic impulse train function, in the frequency domain. Both are shown in

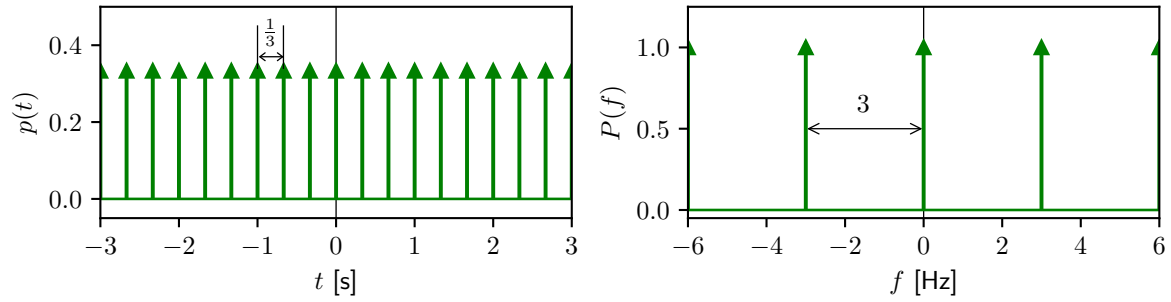


Figure I.4: Periodic impulse train $p(t)$, (9.3), as sampling function at left, with sampling frequency $f_s = 3$ Hz, sampling interval $\Delta t = \frac{1}{3}$ s, and corresponding Fourier transform $P(f)$ at right, with period f_s.

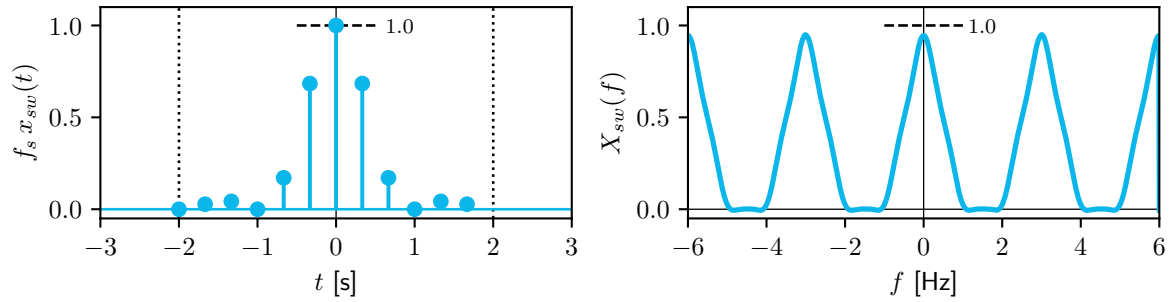


Figure I.5: Continuous-time model of the windowed and sampled signal $x_{sw}(t)$ at left ($f_s = 3$ Hz, $\Delta t = \frac{1}{3}$ s, $T = 4$ s and $N = 12$), shown here as $f_s x_{sw}(t)$ (stems representing $x(n\Delta t)$ as the weights of the Dirac pulses), and Fourier transform $X_{sw}(f)$ at right.

Figure I.4, for a sampling frequency $f_s = 3$ Hz, sampling interval $\Delta t = \frac{1}{3}$ s, and, consequently, the period of $P(f)$ at right is $f_s = 3$ Hz.

The continuous-time model of the windowed and sampled signal $x_{sw}(t)$ is obtained by multiplication of $x_w(t)$ by $p(t)$, (9.2). Its corresponding Fourier transform $X_{sw}(f)$ can then be obtained through convolution: $X_{sw}(f) = X_w(f) * P(f)$, (6.8), see Figure I.5. Similar as found in Chapter 9 we obtain copies of the spectrum, see (9.10), though there the result was derived without making use of convolution. A careful inspection of, and comparison with Figure 9.6 reveals small differences. Recall that with Figure 9.6 we considered an *infinite length* signal (no window was applied yet). In Figures I.3 and I.5 at right one can notice small effects of leakage due to time-windowing.

We first applied a time window $x_w(t) = w(t)x(t)$, and then sampled the signal to obtain $x_{sw}(t) = x_w(t)p(t) = w(t)x(t)p(t)$. The corresponding Fourier transforms are $X_w(f) = W(f) * X(f)$ and $X_{sw}(f) = X_w(f) * P(f) = W(f) * X(f) * P(f)$. The order of the operations can be changed, as both multiplication and convolution are commutative.

I.3 Evaluating Fourier transform at discrete frequencies

Fourier transform $X_{sw}(f)$ is periodic in frequency, with period f_s. Therefore, considering only one 'period' suffices, for instance the range of frequencies between 0 and f_s, $f \in [0, f_s)$. The last step of the entire procedure is then to evaluate the Fourier transform only at discrete frequencies. This can be considered as multiplication of $X_{sw}(f)$, a function continuous in frequency, by an periodic impulse train in the *frequency* domain. In Chapter 12 the frequency step size was set to $\Delta f = \frac{f_s}{N} = \frac{1}{T}$ Hz. The Discrete Fourier Transform (DFT) turns N time

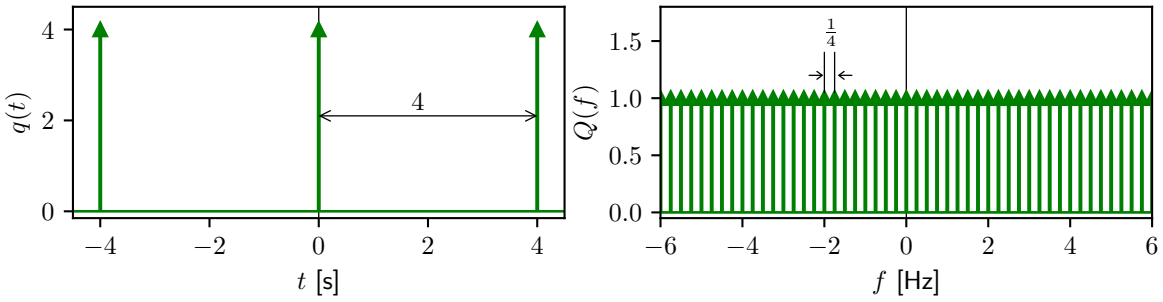


Figure I.6: Periodic impulse train $Q(f)$ in frequency domain at right, with period $\Delta f = \frac{1}{T} = \frac{1}{4}$ Hz, and corresponding time domain function $q(t)$ at left, with period $T = 4$ s.

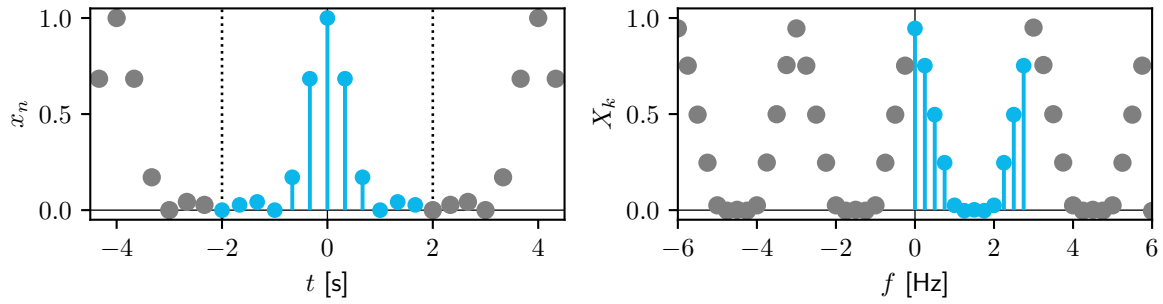


Figure I.7: The DFT evaluates Fourier transform $X_{sw}(f)$ only at discrete frequencies, with step size $\Delta f = \frac{1}{T} = \frac{1}{4}$ s, its coefficients X_k are shown at right. Correspondingly, the discrete time sequence x_n in the time domain at left, is considered to be *periodic* by the DFT, with period $T = 4$ s.

domain samples into the evaluation of the corresponding Fourier transform at N discrete frequencies in the interval $[0, f_s)$. The periodic impulse train $Q(f) = \sum_{k=-\infty}^{\infty} \delta(f - k\Delta f)$ is shown in Figure I.6 at right, for $\Delta f = \frac{1}{4}$ Hz. For completeness, also the corresponding time domain representation $q(t)$ is shown at left, with a period of $T = 4$ s. The inverse Fourier transform of the periodic impulse train function in frequency is a periodic impulse train function in time, in this case scaled by one over the frequency step size Δf.

Though typically of lesser interest to the practical user, the consequence of evaluating the Fourier transform only at discrete frequencies is that the DFT actually considers its input, the discrete-time sequence x_n to be *periodic*, with period T, see Figure I.7.

Multiplication of $X_{sw}(f)$ by $Q(f)$ in the frequency domain at right, is equivalent to convolution of $x_{sw}(t)$ and $q(t)$ in the time domain at left. This can also be observed from the expression for the inverse DFT (12.6). From this equation follows: $x_n = x_{n+N}$, as $e^{jk\frac{2\pi}{N}n} = e^{jk\frac{2\pi}{N}(n+N)}$ with both $n, k \in \mathbb{Z}$, see also Section 12.3.

J

Confidence interval of periodogram

In Chapter 14 the periodogram was presented as an *estimate* of the spectral density of a continuous-time signal $x(t)$, based on N discrete-time samples x_n of that signal. The periodogram is computed at a set of frequencies $f_k = k\Delta f = k\frac{f_s}{N}$ as (14.2):

$$\hat{S}(f_k) = \frac{1}{T}|X_k|^2 \tag{14.2}$$

Here, X_k are the DFT coefficients of the DFT-ed sequence x_n:

$$X_k = \Delta t \sum_{n=0}^{N-1} x_n e^{-jk\frac{2\pi}{N}n} \tag{12.5}$$

Using Euler's formula, (C.1), X_k can be written as:

$$X_k = \Delta t \left(\sum_{n=0}^{N-1} x_n \cos(k\frac{2\pi}{N}n) - j \sum_{n=0}^{N-1} x_n \sin(k\frac{2\pi}{N}n) \right) \tag{J.1}$$

Substituting this in (14.2) with the squared modulus, and $T = N\Delta t$:

$$\hat{S}(f_k) = \frac{\Delta t}{N} \left\{ \left(\sum_{n=0}^{N-1} x_n \cos(k\frac{2\pi}{N}n) \right)^2 + \left(\sum_{n=0}^{N-1} x_n \sin(k\frac{2\pi}{N}n) \right)^2 \right\} \tag{J.2}$$

In practice, we work with *random* signals: signals with noise. As a consequence the periodogram estimate will not be exact, but instead be subject to uncertainty and exhibit variability upon repetition of the experiment. To judge the significance of findings on the basis of the periodogram, a statistical confidence interval for the periodogram is developed.

J.1 Estimating amplitudes

Our input $x_0, \ldots, x_{N-1}$ consists of N random variables, for which we assume here that they are normally distributed, with a zero mean, and that they are all uncorrelated and have the same variance, such that the variance matrix is the identity matrix I_N, scaled by variance σ^2.

Casting the N random variables in a vector, we have:

$$\begin{pmatrix} x_0 \\ x_1 \\ \vdots \\ x_{N-1} \end{pmatrix} \sim \mathcal{N}(\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \sigma^2 \underbrace{\begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}}_{I_N}) \tag{J.3}$$

in terms of an N-dimensional, multivariate normal distribution $\mathcal{N}()$.

Next, we observe that expression (J.2) for the periodogram is closely related to the (simultaneous) least-squares estimation of the amplitude of a cosine and the amplitude of a sine, both with frequency f_k, based on the input data.

The functional model, in terms of the expectation E (mean), reads:

$$\underbrace{E\begin{pmatrix} x_0 \\ x_1 \\ \vdots \\ x_{N-1} \end{pmatrix}}_{x} = \underbrace{\begin{pmatrix} \cos(2\pi f_k \Delta t 0) & \sin(2\pi f_k \Delta t 0) \\ \cos(2\pi f_k \Delta t 1) & \sin(2\pi f_k \Delta t 1) \\ \vdots & \vdots \\ \cos(2\pi f_k \Delta t (N-1)) & \sin(2\pi f_k \Delta t (N-1)) \end{pmatrix}}_{C} \underbrace{\begin{pmatrix} a_k \\ b_k \end{pmatrix}}_{c}$$

with $N \times 2$ matrix C, and 2×1 vector c.

Evaluating the spectrum, as we do with the DFT, for frequencies $f_k = k\Delta f = k\frac{f_s}{N} = k\frac{1}{N\Delta t}$ with $k = 0, 1, \ldots, N-1$ (and in the following we exclude $k = 0$ (the zero frequency), and $k = \frac{N}{2}$ (the Nyquist frequency) for N even), matrix C becomes:

$$C = \begin{pmatrix} \cos(k\frac{2\pi}{N}0) & \sin(k\frac{2\pi}{N}0) \\ \cos(k\frac{2\pi}{N}1) & \sin(k\frac{2\pi}{N}1) \\ \vdots & \vdots \\ \cos(k\frac{2\pi}{N}(N-1)) & \sin(k\frac{2\pi}{N}(N-1)) \end{pmatrix}$$

The least-squares estimator for vector c follows as $\hat{c} = (C^T C)^{-1} C^T x$, where superscript '$T$' indicates matrix transpose. We first note that, surprisingly simple:

$$C^T C = \frac{N}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

as with $\cos^2\alpha = \frac{1}{2} + \frac{1}{2}\cos(2\alpha)$, (A.4), and similar goniometric identities:

$$\sum_{n=0}^{N-1} \cos^2(k\frac{2\pi}{N}n) = \frac{N}{2} + \frac{1}{2}\sum_{n=0}^{N-1} \cos(2k\frac{2\pi}{N}n) = \frac{N}{2}$$

$$\sum_{n=0}^{N-1} \sin^2(k\frac{2\pi}{N}n) = \frac{N}{2} - \frac{1}{2}\sum_{n=0}^{N-1} \cos(2k\frac{2\pi}{N}n) = \frac{N}{2}$$

$$\sum_{n=0}^{N-1} \sin(k\frac{2\pi}{N}n)\cos(k\frac{2\pi}{N}n) = \frac{1}{2}\sum_{n=0}^{N-1} \sin(2k\frac{2\pi}{N}n) = 0$$

where, for $N > 2$, the sum of N equidistant samples ($\frac{n}{N}$ runs from 0 to $\frac{N-1}{N}$), taken over $2k$ periods of a cosine or a sine, which both are zero mean functions, equals zero (with integer $k = 1, \ldots, N-1$, excluding $k = 0$ and $k = \frac{N}{2}$ if N even).

The least-squares estimator for c becomes:

$$\hat{c} = \begin{pmatrix} \hat{a}_k \\ \hat{b}_k \end{pmatrix} = \frac{2}{N} \begin{pmatrix} \sum_{n=0}^{N-1} x_n \cos(k\frac{2\pi}{N}n) \\ \sum_{n=0}^{N-1} x_n \sin(k\frac{2\pi}{N}n) \end{pmatrix} \tag{J.4}$$

With the assumption of (J.3), observing that both amplitude estimators $\hat{a}_k$ and $\hat{b}_k$ in $\hat{c}$ are *linear* combinations of the signal samples $x_0, x_1, \ldots, x_{N-1}$, and applying the mean propagation law, we find that vector $\hat{c}$ has zero mean, and through the variance propagation law we find variance matrix:

$$Q_{\hat{c}} = \sigma^2 (C^T C)^{-1} = \sigma^2 \frac{2}{N} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{J.5}$$

Normality as a distribution is maintained, see Section 2.13 in [5], hence the amplitude estimators are distributed as:

$$\underbrace{\begin{pmatrix} \hat{a}_k \\ \hat{b}_k \end{pmatrix}}_{\hat{c}} \sim \mathcal{N}(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \sigma^2 \frac{2}{N} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}) \tag{J.6}$$

J.2 Confidence interval for periodogram

We find, under the assumed (J.3) and consequently (J.6), that the quadratic form $\hat{c}^T Q_{\hat{c}}^{-1} \hat{c} \sim \chi^2(2,0)$ has a central Chi-squared distribution with 2 degrees-of-freedom.

We use (J.4) and (J.5):

$$\hat{c}^T Q_{\hat{c}}^{-1} \hat{c} = \frac{1}{\sigma^2} \frac{2}{N} \left\{ \left(\sum_{n=0}^{N-1} x_n \cos(k\frac{2\pi}{N}n) \right)^2 + \left(\sum_{n=0}^{N-1} x_n \sin(k\frac{2\pi}{N}n) \right)^2 \right\} \sim \chi^2(2,0) \tag{J.7}$$

so that we can evaluate the probability P, as $1-\alpha$, of this quadratic form lying in between the two bounds as:

$$P\left(\chi^2_{1-\frac{\alpha}{2}}(2,0) \leq \hat{c}^T Q_{\hat{c}}^{-1} \hat{c} \leq \chi^2_{\frac{\alpha}{2}}(2,0) \right) = 1-\alpha$$

where $\chi^2_{\frac{\alpha}{2}}$ is the threshold value of the Chi-squared distribution for a right-sided tail probability of $\frac{\alpha}{2}$, and $\chi^2_{1-\frac{\alpha}{2}}$ for a right-sided tail probability of $1-\frac{\alpha}{2}$ (and consequently for a left-sided tail probability of $\frac{\alpha}{2}$); they are, respectively, the upper and lower $100(\frac{\alpha}{2})\%$ points.

The above $\hat{c}^T Q_{\hat{c}}^{-1} \hat{c}$ (J.7) is close to, but not exactly identical to $\hat{S}(f_k)$ (J.2). We have:

$$\hat{c}^T Q_{\hat{c}}^{-1} \hat{c} = \frac{2}{\sigma^2 \Delta t} \hat{S}(f_k)$$

and the above probability interval can be turned into:

$$P\left(\frac{\sigma^2 \Delta t}{2} \chi^2_{1-\frac{\alpha}{2}}(2,0) \leq \hat{S}(f_k) \leq \frac{\sigma^2 \Delta t}{2} \chi^2_{\frac{\alpha}{2}}(2,0) \right) = 1-\alpha$$

or:

$$P\left(\frac{2}{\sigma^2 \Delta t} \frac{1}{\chi^2_{1-\frac{\alpha}{2}}(2,0)} \geq \frac{1}{\hat{S}(f_k)} \geq \frac{2}{\sigma^2 \Delta t} \frac{1}{\chi^2_{\frac{\alpha}{2}}(2,0)} \right) = 1-\alpha$$

and with:

$$P\left(\frac{2\hat{S}(f_k)}{\chi^2_{\frac{\alpha}{2}}(2,0)} \leq \sigma^2\Delta t \leq \frac{2\hat{S}(f_k)}{\chi^2_{1-\frac{\alpha}{2}}(2,0)}\right) = 1-\alpha$$

The expectation or mean of a central Chi-squared distributed random variable with q degrees-of-freedom equals q, and hence with (J.7) $E(\hat{c}^T Q_{\hat{c}}^{-1}\hat{c}) = 2$, so $E(\hat{S}(f_k)) = \sigma^2\Delta t = S(f_k)$, where the latter follows from assuming the periodogram to be an unbiased estimator of the spectral density. This yields a *confidence interval* for the *periodogram*:

$$P\left(\frac{2\hat{S}(f_k)}{\chi^2_{\frac{\alpha}{2}}(2,0)} \leq S(f_k) \leq \frac{2\hat{S}(f_k)}{\chi^2_{1-\frac{\alpha}{2}}(2,0)}\right) = 1-\alpha \tag{J.8}$$

see also Section 6.2 in [5] and Section 8.5.4 in [17].

This confidence interval expresses the uncertainty in the spectral density estimator $\hat{S}(f_k)$, shown as the variability of the value along the vertical axis in, e.g., Figure 14.4 at right, due to the stochastic model assumed in (J.3), basically considering the noise in the signal samples, quantified by variance σ^2. Confidence interval (J.8) is evaluated per frequency f_k.

J.3 Variance of periodogram

The variance of a central Chi-squared distributed random variable with q degrees-of-freedom equals $2q$, and hence $\sigma^2_{\hat{c}^T Q_{\hat{c}}^{-1}\hat{c}} = 4$, so $\sigma^2_{\hat{S}(f_k)} = (\frac{\sigma^2\Delta t}{2})^2 4 = \sigma^4\Delta t^2 = (E(\hat{S}(f_k)))^2$, with the above expectation. This is an important finding, stating:

$$\sigma_{\hat{S}(f_k)} = E(\hat{S}(f_k)) \tag{J.9}$$

the standard deviation of the periodogram estimator, as a measure for the uncertainty in the estimate, *equals* the target value of the estimate *itself* (with E the expectation); the uncertainty is as large as the quantity of interest.

J.4 Advanced spectral estimation

To improve the precision of spectral estimation the data record is split into M disjoint sub-records or segments, each with $\frac{N}{M}$ samples (for convenience assuming N an integer multiple of M) and record/window length $\frac{T}{M}$ (we assume the segments to be non-overlapping and adjacent). The periodogram for each segment still satisfies the above confidence interval (J.8) as the distribution does not depend on N nor T (see (J.7)), although the frequency step size of the estimator increases from $\frac{1}{T}$ for segment-length N, to $\frac{M}{T}$ with a length $\frac{N}{M}$ segment; resolution gets poorer.

We can then *average* the M per-segment-periodogram estimates to produce one final estimate, and averaging brings down the uncertainty: by averaging the variance reduces by a factor of M. The confidence interval for a spectral density estimator with averaging over M segments (see Section 14.5), see Section 8.5.4 in [17], reads:

$$P\left(\frac{2M\hat{S}(f_k)}{\chi^2_{\frac{\alpha}{2}}(2M,0)} \leq S(f_k) \leq \frac{2M\hat{S}(f_k)}{\chi^2_{1-\frac{\alpha}{2}}(2M,0)}\right) = 1-\alpha \tag{J.10}$$

This equation reduces to (J.8) for $M = 1$.

K

Correlation

In this appendix we define the *cross*-correlation function of two real-valued signals $x(t)$ and $y(t)$, which reduces to the *auto*-correlation for $y(t) = x(t)$. First we cover deterministic signals, then we cover random signals, both in continuous time and discrete time.

K.1 Deterministic signals

K.1.1 Continuous-time signals

The cross-correlation function of two deterministic, continuous-time, and real-valued signals $x(t)$ and $y(t)$, see [12], reads:

$$R_{xy}(\tau) = \lim_{T\to\infty} \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} x(t)y(t+\tau)\,\mathrm{d}t = \lim_{T\to\infty} \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} x(t-\tau)y(t)\,\mathrm{d}t \tag{K.1}$$

and it is a function of time shift τ, referred to as the time *lag* between the two signals. In the first integral, signal $y(t)$ is shifted by τ to the left, and, in the second integral signal $x(t)$ is shifted by τ to the right. Both lead to exactly the same situation of the two signals with respect to each other. Time lag $\tau \in \mathbb{R}$, and thus can also be negative.

The cross-correlation function is a measure, as a function of time lag τ, of the *similarity* between the two signals $x(t)$ and $y(t)$. For each value of τ, it equals the time average of the product of $x(t)$ and $y(t+\tau)$.

When both signals $x(t)$ and $y(t)$ share the same unit, e.g., [V], then the cross-correlation function has the square of that unit, e.g., [V^2].

The autocorrelation function of $x(t)$ is obtained by replacing $y(t+\tau)$ by $x(t+\tau)$ in the first expression of (K.1), or replacing $y(t)$ by $x(t)$ in the second expression. Then we measure the *similarity* of signal $x(t)$ with a time-shifted version of *itself*. Considering the autocorrelation function for zero time shift $\tau = 0$ yields:

$$P = R_{xx}(\tau = 0) = \lim_{T\to\infty} \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} x^2(t)\,\mathrm{d}t \tag{K.2}$$

the average (normalized) power of signal $x(t)$.

Definition (K.1) is common in statistics, though alternatively we may find:

$$R'_{xy}(\tau) = \lim_{T\to\infty} \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} x(t+\tau)y(t)\,\mathrm{d}t \tag{K.3}$$

see in particular [13] and it is often implemented accordingly with discrete-time signals in programming environments, such as Python. The two definitions are related through $R'_{xy}(\tau) = R_{xy}(-\tau) = R_{yx}(\tau)$, hence by swapping x and y.

Finally, we comment that we use the *time average* in the above expressions for the cross-correlation function. Later, with stochastic or random signals we use the *ensemble* average in the definition of $R(\tau)$, and these two types of averages are equivalent if the random signals are *ergodic*. In many textbooks you may find, however, the cross-correlation function defined without time averaging (i.e., without the factor $\frac{1}{T}$ in front of the integral in (K.1) and (K.3)).

EX K.1 Compute the autocorrelation of signal $x(t) = A\sin(2\pi f_0 t)$.

Solution Using (K.1) we get:

$$R_{xx}(\tau) = \lim_{T\to\infty} \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} A\sin(2\pi f_0 t)\; A\sin(2\pi f_0(t+\tau))\,\mathrm{d}t$$

Using trigonometric identity (A.8), this becomes:

$$R_{xx}(\tau) = \frac{A^2}{2} \lim_{T\to\infty} \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} (\cos(2\pi f_0\tau) - \cos(4\pi f_0 t + 2\pi f_0\tau))\,\mathrm{d}t$$

The integral is broken into two parts:

$$\begin{aligned} R_{xx}(\tau) &= \frac{A^2}{2} \lim_{T\to\infty} \left\{ \frac{1}{T}\cos(2\pi f_0\tau)\,[t]_{-\frac{T}{2}}^{\frac{T}{2}} - \frac{1}{T}\left[\frac{1}{4\pi f_0}\sin(4\pi f_0 t + 2\pi f_0\tau)\right]_{-\frac{T}{2}}^{\frac{T}{2}} \right\} \\ &= \frac{A^2}{2}\cos(2\pi f_0\tau) + \frac{A^2}{2} \underbrace{\lim_{T\to\infty} \left\{ -\frac{1}{T}\frac{1}{2\pi f_0}\sin(2\pi f_0 T)\cos(2\pi f_0\tau) \right\}}_{=0} \end{aligned}$$

where we used (A.10) in the last step. The end-result reads:

$$R_{xx}(\tau) = \frac{A^2}{2}\cos(2\pi f_0\tau) \tag{K.4}$$

Autocorrelation measures the similarity of a signal with a time-shifted version of itself. For $\tau = 0$ we get maximum similarity $R_{xx}(\tau = 0) = \frac{A^2}{2}$. The sine is periodic, and hence for a time shift or lag of $\tau = \frac{1}{f_0}$ (the period of the sine), the autocorrelation is again at maximum. The autocorrelation function $R_{xx}(\tau)$ is *periodic* in τ.

K.1.2 Discrete-time signals

Discretizing the integral in (K.1) with step size Δt, the cross-correlation for discrete time, infinite length sequences x_n and y_n becomes:

$$R_{xy}(m) = \lim_{N\to\infty} \frac{\Delta t}{N\Delta t} \sum_{n=-\frac{N}{2}}^{\frac{N}{2}} x_n y_{n+m} = \lim_{N\to\infty} \frac{1}{N} \sum_{n=-\frac{N}{2}}^{\frac{N}{2}} x_n y_{n+m} \tag{K.5}$$

with $T = N\Delta t$, $t = n\Delta t$ and shift $\tau = m\Delta t$, and assuming N to be even for convenience. Index $m \in \mathbb{Z}$ denotes the shift between the two sequences.

Working with finite length N sequences (both sequences of equal length), we arrive at:

$$R_{xy}(m) = \frac{1}{N} \sum_{n} x_n y_{n+m} \tag{K.6}$$

where the bounds of summation over n depend on the lengths of x_n and y_n, and on the shift m (we need to satisfy both $0 \le n \le N-1$ and $0 \le n+m \le N-1$). And thereby the shift itself is bound as well ($|m| \le N-1$). The result $R_{xy}(m)$ is a length $2N-1$ sequence.

EX K.2

Compute the cross-correlation of discrete sequence $x_n = [1, 2, 3]$ with sequence $y_n = [4, 5, 6]$ according to (K.6).

Solution Both x_n and y_n are, as usual in practice, *finite* length sequences, here with length $N = 3$.

For a shift of $m = -1$ we have the situation as shown in Table K.1. Sequence y_n is delayed by one step. The overlap (and hence summation index n) runs from $n = 1$ to $n = 2$. The result is $R_{xy}(m = -1) = 2 \times 4 + 3 \times 5 = 23$.

Table K.1: Cross-correlation of discrete-time sequences x_n and y_n, for shift $m = -1$.

n	0	1	2	3	4
x_n	1	2	3		
y_{n-1}		4	5	6	

Table K.2 shows the situation for a shift of $m = 2$. Sequence y_n is advanced by two steps. The overlap is limited to one position and the summation has only one term, namely for $n = 0$. The result is $R_{xy}(m = 2) = 1 \times 6 = 6$.

Table K.2: Cross-correlation of discrete-time sequences x_n and y_n, for shift $m = 2$.

n	-2	-1	0	1	2	3
x_n			1	2	3	
y_{n+2}	4	5	6			

The full cross-correlation result is shown in Table K.3.

Table K.3: Cross-correlation $R_{xy}(m)$ of discrete-time sequences x_n and y_n.

m	-2	-1	0	1	2
$R_{xy}(m)$	12	23	32	17	6

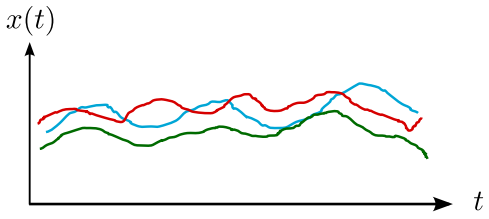


Figure K.1: Three sample functions, also referred to as realizations, of a continuous-time random signal.

Cross-correlation of discrete-time sequences can be done in Python through `numpy.correlate`. The 'full' output is obtained from `correlate(x,y,"full")` and results in $R'_{xy}(m) = [6, 17, 32, 23, 12]$, which equals the result in Table K.3, except that the output sequence is just reversed. We get $R'_{xy}(m) = R_{yx}(m) = R_{xy}(-m)$, as the implementation in Python is based on cross-correlation defined in (K.3). The indices m for output R'_{xy} in Python actually run from 0 to $2(N-1)$, in this example $[0, 1, 2, 3, 4]$.

As a sidenote, we observe that in the Example in Appendix F on discrete convolution, we used the same two data sequences $x_n = [1, 2, 3]$ and $h_n = [4, 5, 6]$. Convolution and correlation are very similar operations. Both involve shifting or sliding one sequence with respect to the other, and for each shift, taking the sum of the element-wise product of the two sequences. The key difference is that with convolution, according to (F.2), sequence h_n is flipped or mirrored prior to the sliding. The result of convolution $x_n * h_n$ was given in Table F.3. Exactly the same result is obtained in the present example, with Python, through first reversing the order of the second sequence h_n, and then obtain the cross-correlation $R'_{xh'}$, where h' refers to sequence h in reversed or flipped order, hence `correlate(x,h[::-1],"full")`.

Finally, the attentive reader may have noticed that we 'forgot' to divide by N in this example, see (K.6). In Table K.1, for $m = -1$, the overlap limited to two positions, and in Table K.2, for $m = 2$, to only one position. In general the overlap is limited to $N - |m|$ positions, and hence the factor in front of the summation in (K.6) should be adjusted to $\frac{1}{N-|m|}$, taking the actual overlap between the two sequences into account.

K.2 Random signals

Having covered deterministic signals, we now turn our attention to random signals, which result from observing stochastic processes. A stochastic process is a phenomenon which is subject to uncontrolled variability and associated uncertainty, and additional variability (noise) is typically involved in the observation of the process, see for example Figure 17.1.

A random signal is a random variable which is a function of time t. The variability in outcomes of a random variable is mathematically modeled through a probability density function. Similarly we do so for the variability in outcomes of the (observable) process, and the probability density function generally includes time as a parameter.

Would we draw samples (i.e., repeat the experiment under identical circumstances), for instance, from a normally-distributed random variable with mean equal to 7.2 and standard deviation equal to 0.2, we could obtain: 7.1, 7.4, 7.6, 6.9 and 7.0. Similarly, we can draw samples or *realizations* of a random signal, and several different functions of time would result, as shown in Figure K.1. The set of all possible realizations is referred to as the *ensemble*. The ensemble forms the random signal.

Cross-correlation, and power spectral density are defined for random signals, and these

definitions involve an expectation (mean), in which the probability density function is involved.

K.2.1 Continuous-time signals

The cross-correlation of two random signals for x at time t_1 and y at time t_2 is defined as:

$$R_{xy}(t_1, t_2) = E(x(t_1)y(t_2)) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f_{xy}(x, y; t_1, t_2)\,\mathrm{d}x\mathrm{d}y \tag{K.7}$$

with $f_{xy}(x, y)$ being the joint probability density function of x and y. Expectation $E(\cdot)$ refers to the ensemble average, in this case of the product $x(t_1)y(t_2)$.

When random signals are (jointly) wide-sense stationary, as we assume throughout this book, then the mean and cross-correlation do not depend on time (in brief, stationarity means that the statistical properties of the signals do not change over time, see [5]). The cross-correlation does not depend on time instants t_1 and t_2 in an absolute sense, but only in a relative sense on time lag $\tau = t_2 - t_1$, and we have:

$$R_{xy}(\tau) = E(x(t)y(t+\tau)) = E(x(t-\tau)y(t)) \tag{K.8}$$

K.2.2 Discrete-time signals

In discrete time, a random signal can be interpreted as a sequence of random variables (also referred to as a random sequence). Considering a random signal at a single discrete time instant yields a random variable, which variability is described by a statistical distribution, e.g., the normal or Gaussian distribution.

The cross-correlation of two (jointly wide-sense stationary) random sequences x_n and y_n is defined as:

$$R_{xy}(m) = E(x_n y_{n+m}) \tag{K.9}$$

with index m denoting the shift between the two sequences, similar as in (K.6).

K.2.3 Cross-correlation estimation in practice

In practice, all we have is a single sample or realization of the random signal, i.e., one time series of discrete-time measurements of the stochastic process, and we compute an *estimate* for the cross-correlation or the power spectral density. It is key to understand that this estimate is subject to *uncertainty*: the outcome is not exact, nor perfect. Would we repeat the measurement under identical circumstances, then the second realization will differ from the first one, see Figure K.1, and hence also the estimate computed from it.

For the above approach to work, we also need to assume that the random signal is *ergodic*, [13]. Ergodicity means that time averages of a sample function equal the corresponding ensemble averages of the random signal (in practice we do have access only to the first one, and we are after the second one). Ergodicity implies stationarity (only stationary random processes can be ergodic).

An estimate (indicated by the hat-symbol) for the cross-correlation, based on discrete time sequences x and y, as realizations of their respective random sequences, see [17], reads:

$$\hat{R}_{xy}(m) = \frac{1}{N-m}\sum_{n=0}^{N-1-m} x_n y_{n+m} \quad \text{for } m \geq 0 \tag{K.10}$$

and for $m < 0$ we have:

$$\hat{R}_{xy}(m) = \frac{1}{N+m}\sum_{n=-m}^{N-1} x_n y_{n+m} \quad \text{for } m < 0$$

though in practice it may be more convenient to use again (K.10), but swapping x_n and y_n, and exploit $R_{xy}(-m) = R_{yx}(m)$, as noted with (K.3).

The factor $\frac{1}{N-m}$ in front of the summation in (K.10) yields the *unbiased* cross-correlation estimator, [17]. Using instead a factor of $\frac{1}{N}$, as in (K.6), yields the *biased* estimator.

L

White noise

In this appendix we present a theoretical continuous-time white noise random signal, and then a band-limited white noise signal which can be realized in discrete time.

L.1 Theoretical white noise

A random signal is said to be white noise, when it has a flat power spectral density:

$$S(f) = \frac{N_0}{2} \tag{L.1}$$

the PSD is *constant* over the entire range of frequencies from $f = -\infty$ to ∞, with N_0 a positive constant, commonly used to denote the noise power spectral density, see [12].

The autocorrelation of the (continuous-time) signal is then:

$$R(\tau) = \frac{N_0}{2}\delta(\tau) \tag{L.2}$$

a Dirac delta function, with weight $\frac{N_0}{2}$. The spectral density $\frac{N_0}{2}$, e.g., in [W/Hz] for a voltage signal, appears in the autocorrelation function and the Dirac delta function has the inverse dimension of its argument, see Appendix B. Both functions are shown in Figure L.1.

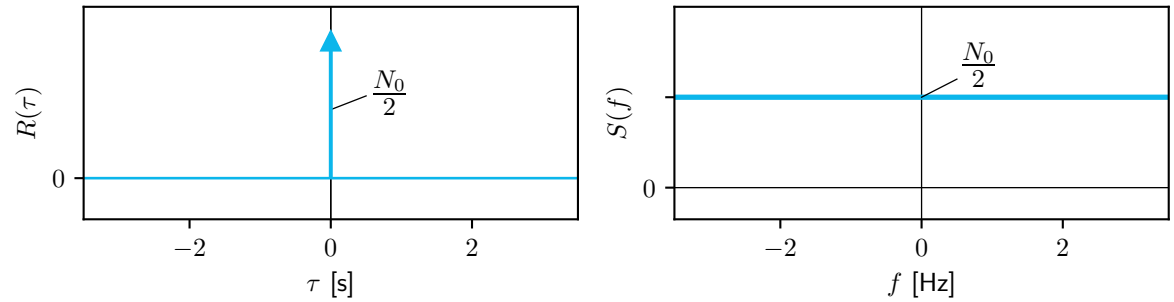


Figure L.1: Autocorrelation function $R(\tau)$ at left, and PSD $S(f)$ at right, of a continuous-time white noise signal.

The autocorrelation function $R(\tau)$ and the power spectral density $S(f)$ are a Fourier transform pair $R(\tau) \overset{\mathcal{F}}{\leftrightarrow} S(f)$, according to the Wiener-Khintchine[1] theorem for a wide-sense stationary random signal, see [12], a subject which is beyond the scope of this book.

[1] After the American mathematician and philosopher N. Wiener (1894-1964) and the Russian mathematician A.Y. Khinchin (1894-1959).

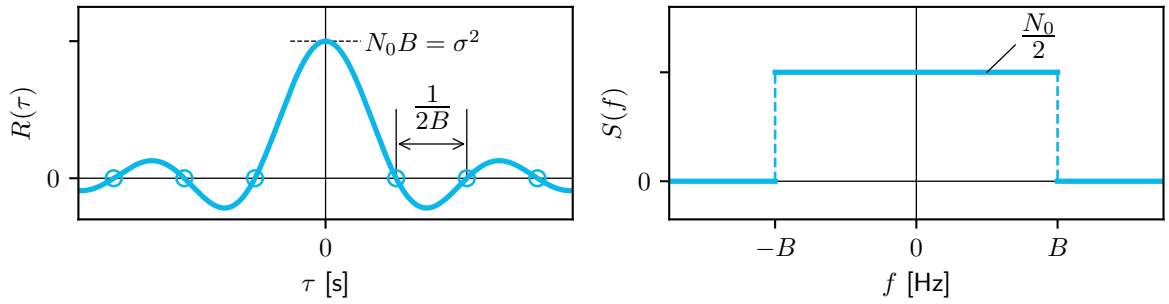


Figure L.2: Autocorrelation function $R(\tau)$ at left, and PSD $S(f)$ at right, of band-limited white noise signal.

Considering a white noise signal at two time instants, no matter how far apart or close, the above given autocorrelation function $R(\tau)$ implies that the two corresponding random variables are *uncorrelated*. If they are not only uncorrelated but also *independent*, then the signal is referred to as strictly white noise [13]. Ref. [5] states that the definition of white noise based on uncorrelatedness is believed to be the more widely used one.

If noise is not white, one refers to *colored* noise. The name white noise is an analogy with white light being composed of all colors of light (all wavelengths/frequencies).

A continuous-time white noise signal is a theoretical construct and cannot exist in reality — it would have infinite power P, see (13.3).

L.2 Band-limited white noise

A band-limited white noise signal has a constant power spectral density over a bandwidth of $2B$ (with $0 < B < \infty$) as:

$$S(f) = \begin{cases} \frac{N_0}{2} & \text{for } |f| \le B \\ 0 & \text{otherwise} \end{cases} \tag{L.3}$$

The (double-sided) power spectral density can then also be expressed using the pulse function:

$$S(f) = \frac{N_0}{2} \Pi\left(\frac{f}{2B}\right) \tag{L.4}$$

covering a band from $f = -B$ to B. It can be obtained from (L.1) by means of ideal low-pass filtering, see also Figure 18.2. The corresponding autocorrelation function reads:

$$R(\tau) = N_0 B \,\mathrm{sinc}(2B\tau) \tag{L.5}$$

which for $B \to \infty$ leads to (L.2), referred to as infinite bandwidth white noise.

Figure L.2 shows the autocorrelation function $R(\tau)$ and the power spectral density $S(f)$ of a band-limited white noise signal. The noise spectral density at right has a bandwidth of $2B$, and correspondingly the autocorrelation function $R(\tau)$ shows its first zero-crossing at $\tau = \frac{1}{2B}$.

The (average) power equals:

$$P = \int_{-\infty}^{\infty} S(f)\,\mathrm{d}f = \frac{N_0}{2} 2B = N_0 B$$

and this equals the noise variance σ^2, therefore:

$$\sigma^2 = R(\tau = 0) = N_0 B$$

Of practical interest is the discrete-time white noise signal with samples taken an interval of $\Delta t = \frac{1}{2B}$ apart, that is, with sampling frequency $f_s = 2B$. This results in a sequence of samples drawn from *uncorrelated* random variables; it is a white noise sequence (as we assumed with (J.3)). Ref. [5] refers to a sequence of uncorrelated random variables as a purely random process; the process has 'no memory'.

The autocorrelation for samples of two random variables $\tau = \frac{1}{2B}$ seconds apart (and integer multiples of it), with (L.5), is $R(\tau = \frac{1}{2B}) = 0$. The sample instants then occur exactly at the so-called nulls of the autocorrelation function (the sinc function) in Figure L.2 at left: these samples are uncorrelated. For other values of τ there *is* correlation between the samples.

L.3 Additive White Gaussian Noise (AWGN)

Related to the topic of white noise, one often encounters the acronym AWGN for Additive White Gaussian[2] Noise.

A measured or observed signal is often considered to consist of a deterministic signal of interest, and added to that, a random white noise signal.

White noise refers to (L.2) and (L.1), and does not imply anything else on the statistical distribution of the random variable or process. Hence, white noise can also be Gaussian, but it does not necessarily need to be.

Gaussian noise is often assumed and used though, and the stochastic process considered at a single moment in time is a random variable, which is then *normally distributed*.

[2]After the German mathematician J.C.F. Gauss (1777-1855).

M

Solving first-order differential equation – example

In this appendix we demonstrate, by means of an example, how to analytically solve, in the time domain, a first-order ordinary differential equation (ODE). The ODE is given as:

$$\frac{1}{\alpha}\frac{\mathrm{d}y(t)}{\mathrm{d}t} + y(t) = x(t) \tag{16.9}$$

with input $x(t)$ and output $y(t)$, and $\alpha > 0$. We aim to express output $y(t)$ explicitly in terms of (arbitrary) input $x(t)$, where we assume that $x(t)$ is applied at time $t = t_0$ and that $y(t_0) = y_0$. This linear, first-order, constant coefficient, ordinary differential equation fully describes a dynamic first-order system that is linear, time-invariant and causal. The example is taken from [7].

The above differential equation can describe how the temperature of an object, as a function of time, is related to the ambient temperature, or the velocity of a mass due to an applied force in a mass-damper system in mechanical engineering, or the output voltage as a response to the input voltage in an elementary resistor-capacitor (RC) circuit in electrical engineering.

M.1 Homogeneous solution

The solution to the homogeneous differential equation:

$$\frac{1}{\alpha}\frac{\mathrm{d}y(t)}{\mathrm{d}t} + y(t) = 0 \tag{M.1}$$

is found by assuming a solution of the form $y(t) = Ae^{pt}$. Substituting this in (M.1) leads to $p = -\alpha$. The homogeneous solution reads:

$$y(t) = Ae^{-\alpha t} \tag{M.2}$$

M.2 Total solution

In order to find the total solution we use the technique of 'variation of parameters', which consists of assuming a solution of the form of the above homogeneous solution, but with undetermined coefficient A replaced by a function of time $A(t)$, which is to be found. Thus:

$$y(t) = A(t)e^{-\alpha t} \tag{M.3}$$

Differentiating (and using the chain rule for differentiation), leads to:

$$\frac{\mathrm{d}y(t)}{\mathrm{d}t} = \left(\frac{\mathrm{d}A(t)}{\mathrm{d}t} - \alpha A(t)\right) e^{-\alpha t} \tag{M.4}$$

Next, substituting the assumed solution (M.3) and its derivative (M.4) in the original differential equation (16.9), we obtain:

$$\frac{1}{\alpha} e^{-\alpha t} \frac{\mathrm{d}A(t)}{\mathrm{d}t} = x(t) \quad \Rightarrow \quad \frac{\mathrm{d}A(t)}{\mathrm{d}t} = x(t)\, \alpha e^{\alpha t} \tag{M.5}$$

Solving for $\frac{\mathrm{d}A(t)}{\mathrm{d}t}$, i.e., integrating the right-hand side of (M.5) yields:

$$A(t) - A(t_0) = \alpha \int_{t_0}^{t} x(\tau) e^{\alpha \tau}\, \mathrm{d}\tau$$

Using (M.3) at time t_0: $y(t_0) = y_0 = A(t_0) e^{-\alpha t_0}$, or $A(t_0) = y_0\, e^{\alpha t_0}$, we find the time-varying parameter $A(t)$ as:

$$A(t) = \alpha \int_{t_0}^{t} x(\tau) e^{\alpha \tau}\, \mathrm{d}\tau + y_0\, e^{\alpha t_0} \tag{M.6}$$

This result can be substituted in the assumed solution (M.3), resulting in:

$$y(t) = y_0\, e^{-\alpha(t-t_0)} + \alpha \int_{t_0}^{t} x(\tau) e^{-\alpha(t-\tau)}\, \mathrm{d}\tau$$

Assuming that the input $x(t)$ is applied at $t = -\infty$, hence $t_0 = -\infty$, and that $y_0 = y(t_0) = y(t = -\infty) = 0$, we obtain:

$$y(t) = \int_{-\infty}^{t} x(\tau) \alpha e^{-\alpha(t-\tau)}\, \mathrm{d}\tau \tag{M.7}$$

M.3 Solution

Output $y(t)$ to input $x(t)$ can be found through solving the above integral (M.7). In order to align with (16.6) we introduce the unit step function $u(t)$ and slightly rewrite (M.7):

$$y(t) = \int_{-\infty}^{\infty} x(\tau) \underbrace{\alpha e^{-\alpha(t-\tau)} u(t-\tau)}_{= h(t-\tau)}\, \mathrm{d}\tau \tag{M.8}$$

where $u(t-\tau)$ 'turns off' at $\tau = t$ (and stays off). Output $y(t)$ follows as the convolution of input $x(t)$ and impulse response $h(t)$, (16.6). The interpretation of function $h(t) = \alpha e^{-\alpha t} u(t)$ is discussed in Chapter 16.

Bibliography

[1] *Wikimedia Commons,* (n.d.), media file repository, for public domain and freely licensed educational media content.

[2] T. Bořil, *Fourier series 3d - interactive demonstration,* (2016), website.

[3] J. G. Proakis and D. K. Manolakis, *Digital signal processing*, 4th ed. (Pearson Education Limited, 2014).

[4] M. T. Heideman, D. H. Johnson, and C. S. Burrus, *Gauss and the history of the fast Fourier transform,* Archive for History of Exact Sciences **34**, 265 (1985).

[5] M. B. Priestley, *Spectral analysis and time series* (Academic Press, 1981).

[6] M. R. Spiegel, S. Lipschutz, and J. Liu, *Mathematical handbook of formulas and tables*, 3rd ed. (McGraw-Hill, 2009).

[7] R. E. Ziemer, W. H. Tranter, and D. R. Fannin, *Signals & systems – continuous and discrete*, 4th ed. (Prentice Hall, 1998).

[8] A. V. Oppenheim, A. S. Willsky, and I. T. Young, *Signals and systems*, 1st ed. (Prentice Hall International Editions, 1983).

[9] A. Papoulis, *The Fourier integral and its applications*, 2nd ed. (McGraw-Hill, 1962).

[10] J. W. Cooley and J. W. Tukey, *An algorithm for the machine calculation of complex Fourier series,* Mathematics of Computation **19**, 297 (1965).

[11] G. D. Bergland, *A guided tour of the fast Fourier transform,* IEEE Spectrum **6**, 41 (1969).

[12] L. W. Couch, *Digital and analog communication systems*, 7th ed. (Pearson Prentice Hall, 2007).

[13] A. Papoulis, *Probability, random variables, and stochastic processes*, 2nd ed. (McGraw-Hill, 1984).

[14] A. Schuster, *On the investigation of hidden periodicities with application to a supposed 26 day period of meteorological phenomena,* Terrestrial Magnetism **3**, 13 (1898).

[15] P. Welch, *The use of fast fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms,* IEEE Transactions on Audio and Electroacoustics **15**, 70 (1967).

[16] M. Bartlett, *Smoothing periodograms from time-series with continuous spectra,* Nature **161**, 686 (1948).

[17] J. S. Bendat and A. G. Piersol, *Random data analysis and measurement procedures*, 4th ed. (Wiley, 2010).

Attribution of photos in Chapter 1

All photos, except for the Citation II (Figure 1.4), were obtained from Wikimedia Commons:
https://commons.wikimedia.org/

In the following table the photo rights are summarized, for details we refer to [1].

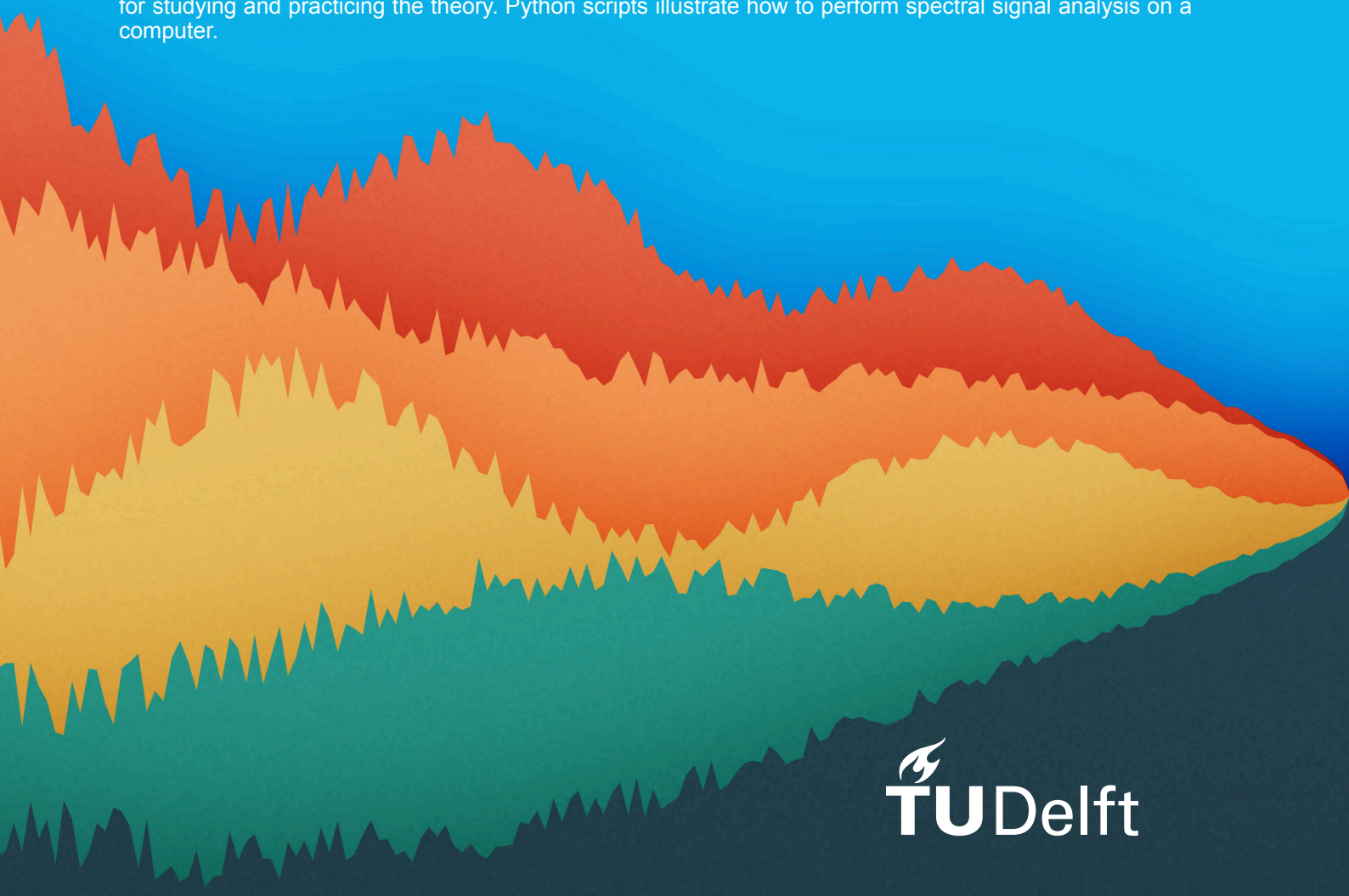
(provisional)Title	URL (click)	license
Figure 1.1 (from top left to bottom right)		
ISS astronaut	Wikimedia Commons	public domain
X-ray of hand	Wikimedia Commons	CC0
Music studio	Wikimedia Commons	CC BY SA
Valkyrie robot	Wikimedia Commons	public domain
Wind turbines	Wikimedia Commons	public domain
Smartphone use	Wikimedia Commons	CC0
PSX controller	Wikimedia Commons	public domain
MRI head	Wikimedia Commons	public domain
Quadcopter	Wikimedia Commons	CC0
Self-driving car	Wikimedia Commons	CC0
A380 cockpit	Wikimedia Commons	CC BY 2.0
Mount Pleasant radio telescope	Wikimedia Commons	public domain
GPS satellite	Wikimedia Commons	public domain
Mars exploration rover	Wikimedia Commons	public domain
Figure 1.4		
Citation II laboratory aircraft	not applicable	copyrighted TU Delft & NLR
Figure 1.7		
Opening day of the Tacoma Narrows Bridge, Tacoma, Washington, July 1, 1940	Wikimedia Commons	public domain
Tacoma-narrows-bridge-collapse	Wikimedia Commons	public domain

Engineering Signal Analysis
from Fourier to filtering

Theory

This book, Theory, is an introductory textbook on the analysis of signals in time and frequency. It takes an engineer's perspective and discusses how to characterize, analyze and operate on signals. The basic theoretical concepts, Fourier series and transform, are x in continuous time. It then introduces discrete-time signals, addressing how sampling and finite signal duration affect spectral analysis. It discusses the discrete Fourier transform and its use in spectral estimation. The book concludes with an introduction to linear systems and signal filtering.

The companion book, Exercises, contains hundreds of exercises, including answers and worked examples, for studying and practicing the theory. Python scripts illustrate how to perform spectral signal analysis on a computer.



TU Delft

Christian Tiberius
TU Delft
Faculty of Civil Engineering and Geosciences
Geoscience and Remote Sensing

Max Mulder
TU Delft
Faculty of Aerospace Engineering
Control and Operations

ISBN 978-94-6518-235-3
DOI https://doi.org/10.59490/mt.247
https://books.open.tudelft.nl

Cover design:
CYANETICA
Sebastiaan de Stigter