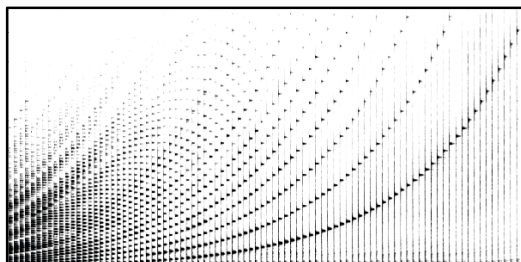


Chapter 2

Fourier Analysis of Signals



As we have seen in the last chapter, music signals are generally complex sound mixtures that consist of a multitude of different sound components. Because of this complexity, the extraction of musically relevant information from a waveform constitutes a difficult problem. A first step in better understanding a given signal is to decompose it into building blocks that are more accessible for the subsequent processing steps. In the case that these building blocks consist of sinusoidal functions, such a process is also called Fourier analysis. Sinusoidal functions are special in the sense that they possess an explicit physical meaning in terms of frequency. As a consequence, the resulting decomposition unfolds the frequency spectrum of the signal—similar to a prism that can be used to break light up into its constituent spectral colors. The Fourier transform converts a signal that depends on time into a representation that depends on frequency. Being one of the most important tools in signal processing, we will encounter the Fourier transform in a variety of music processing tasks.

In Section 2.1, we introduce the main ideas of the Fourier transform and summarize the most important facts that are needed for understanding the subsequent chapters of the book. Furthermore, we introduce the required mathematical notions. A good understanding of Section 2.1 is essential for the various music processing tasks to be discussed. In Section 2.2 to Section 2.5, we cover the Fourier transform in greater mathematical depth. The reader who is mainly interested in the music processing applications may skip these more technical sections on a first reading.

In Section 2.2, we take a closer look at signals and discuss their properties from a more abstract perspective. In particular, we consider two classes of signals: analog signals that give us the right physical interpretation and digital signals that are needed for actual digital processing by computers. The different signal classes lead to different versions of the Fourier transform, which we introduce with math-

ematical rigor along with intuitive explanations and numerous illustrating examples (Section 2.3). In particular, we explain how the different versions are interrelated and how they can be approximated by means of the discrete Fourier transform (DFT). The DFT can be computed efficiently by means of the fast Fourier transform (FFT), which will be discussed in Section 2.4. Finally, we introduce the short-time Fourier transform (STFT), which is a local variant of the Fourier transform yielding a time–frequency representation of a signal (Section 2.5). By presenting this material from a different perspective as typically encountered in an engineering course, we hope to refine and sharpen the understanding of these important and beautiful concepts.

2.1 The Fourier Transform in a Nutshell

Let us start with an audio signal that represents the sound of some music. For example, let us analyze the sound of a single note played on a piano (see Figure 2.1a). How can we find out which note has actually been played? Recall from Section 1.3.2 that the pitch of a musical tone is closely related to its fundamental frequency, the frequency of the lowest partial of the sound. Therefore, we need to determine the frequency content, the main periodic oscillations of the signal. Let us zoom into the signal considering only a 10-ms section (see Figure 2.1b). The figure shows that the signal behaves in a nearly periodic way within this section. In particular, one can observe three main crests of a sinusoidal-like oscillation (see also Figure 2.1c). Having approximately three oscillation cycles within a 10-ms section means that the signal contains a frequency component of roughly 300 Hz.

The main idea of **Fourier analysis** is to compare the signal with sinusoids of various¹ frequencies $\omega \in \mathbb{R}$ (measured in Hz). Each such sinusoid or pure tone may be thought of as a prototype oscillation. As a result, we obtain for each considered frequency parameter $\omega \in \mathbb{R}$ a magnitude coefficient $d_\omega \in \mathbb{R}_{\geq 0}$ (along with a phase coefficient $\varphi_\omega \in \mathbb{R}$, the role of which is explained later). In the case that the coefficient d_ω is large, there is a high similarity between the signal and the sinusoid of frequency ω , and the signal contains a periodic oscillation at that frequency (see Figure 2.1c). In the case that d_ω is small, the signal does not contain a periodic component at that frequency (see Figure 2.1d).

Let us plot the coefficients d_ω over the various frequency parameters $\omega \in \mathbb{R}$. This yields a graph as shown in Figure 2.1f. In this graph, the highest value is assumed for the frequency parameter $\omega = 262$ Hz. By (1.1), this is roughly the center frequency of the pitch $p = 60$ or the note C4. Indeed, this is exactly the note played in our piano example. Furthermore, as illustrated by Figure 2.1e, one can also observe a

¹ In the following, we also consider *negative frequencies* for mathematical reasons without explaining this concept in more detail. In our musical context, negative frequencies are redundant (having the same interpretation as positive frequencies), but simplify the mathematical formulation of the Fourier transform.

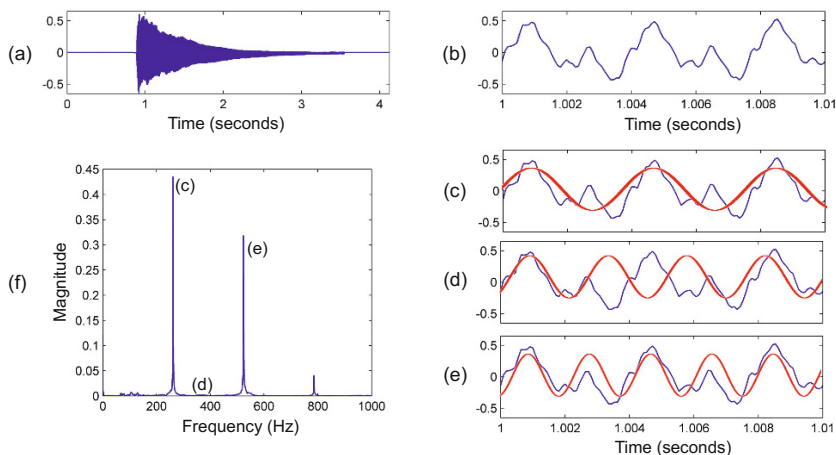


Fig. 2.1 (a) Waveform of a note C4 (261.6 Hz) played on a piano. (b) Zoom into a 10-ms section starting at time position $t = 1$ sec. (c–e) Comparison of the waveform with sinusoids of various frequencies ω . (f) Magnitude coefficients d_ω in dependence on the frequency ω .

high similarity between the signal and the sinusoid of frequency $\omega = 523$ Hz. This is roughly the frequency for the second partial of the tone C4.

With this example, we have already seen the main idea behind the **Fourier transform**. The Fourier transform breaks up a signal into its frequency components. For each frequency $\omega \in \mathbb{R}$, the Fourier transform yields a coefficient d_ω (and a phase φ_ω) that tells us to which extent the given signal matches a sinusoidal prototype oscillation of that frequency.

One important property of the Fourier transform is that the original signal can be reconstructed from the coefficients d_ω (along with the coefficients φ_ω). To this end, one basically superimposes the sinusoids of all possible frequencies, each weighted by the respective coefficient d_ω (and shifted by φ_ω). This weighted superposition is also called the **Fourier representation** of the original signal. The original signal and the Fourier transform contain the same amount of information. This information, however, is represented in different ways. While the signal displays the information across time, the Fourier transform displays the information across frequency. As put by Hubbard [9], the signal tells us when certain notes are played in time, but hides the information about frequencies. In contrast, the Fourier transform of music displays which notes (frequencies) are played, but hides the information about when the notes are played.

In the following sections, we take a more detailed look at the Fourier transform and some of its main properties.

2.1.1 Fourier Transform for Analog Signals

In Section 1.3.1, we saw that a signal or sound wave yields a function that assigns to each point in time the deviation of the air pressure from the average air pressure at a specific location. Let us consider the case of an **analog** signal, where both the time as well as the amplitude (or deviation) are continuous, real-valued parameters. In this case, a signal can be modeled as a function $f: \mathbb{R} \rightarrow \mathbb{R}$, which assigns to each time point $t \in \mathbb{R}$ an amplitude value $f(t) \in \mathbb{R}$. Plotting the amplitude over time, one obtains a graph of this function that corresponds to the waveform of the signal (see Figure 1.17).

The term **function** may need some explanation. In mathematics, a function yields a relation between a set of input elements and a set of output elements, where each input element is related to exactly one output element. For example, a function can be a polynomial $f: \mathbb{R} \rightarrow \mathbb{R}$ that assigns for each input element $t \in \mathbb{R}$ an output element $f(t) = t^2 \in \mathbb{R}$. At this point, we want to emphasize that one needs to differentiate between a function f and its output element $f(t)$ (also referred to as the **value**) at a particular input element t (also referred to as the **argument**). In other words, mathematicians think of a function f in an abstract way, where the symbol or physical meaning of the argument does not matter. As opposed to this, engineers often like to emphasize the meaning of the input argument and loosely speak of a function $f(t)$, even though this is strictly speaking an output value. In this book, we assume the viewpoint of a mathematician.

2.1.1.1 The Role of the Phase

After this side note, let us turn towards the spectral analysis of a given analog signal $f: \mathbb{R} \rightarrow \mathbb{R}$. As explained in our introductory example, we compare the signal f with prototype oscillations that are given in the form of sinusoids. In Section 1.3.2 and Figure 1.19, we have already encountered such sinusoidal signals. Mathematically, a **sinusoid** is a function $g: \mathbb{R} \rightarrow \mathbb{R}$ defined by

$$g(t) := A \sin(2\pi(\omega t - \varphi)) \quad (2.1)$$

for $t \in \mathbb{R}$. The parameter A corresponds to the **amplitude**, the parameter ω to the **frequency** (measured in Hz), and the parameter φ to the **phase** (measured in normalized radians with 1 corresponding to an angle of 360°). In Fourier analysis, we consider prototype oscillations that are normalized with regard to their power (average energy) by setting $A = \sqrt{2}$. Thus for each frequency parameter ω and phase parameter φ we obtain a sinusoid $\cos_{\omega, \varphi}: \mathbb{R} \rightarrow \mathbb{R}$ given by

$$\cos_{\omega, \varphi}(t) := \sqrt{2} \cos(2\pi(\omega t - \varphi)) \quad (2.2)$$

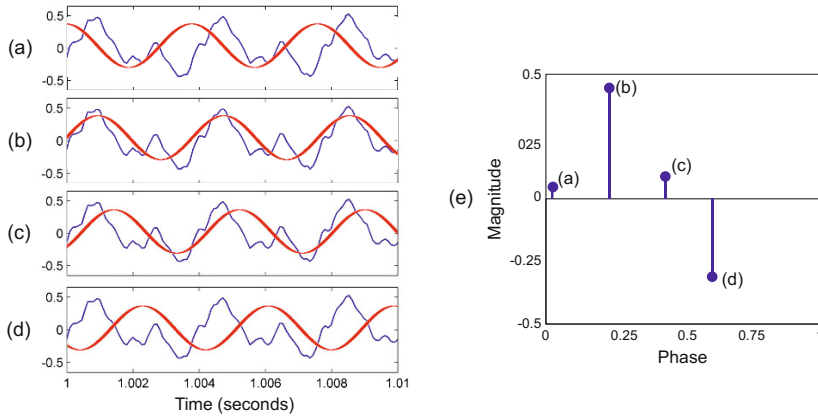


Fig. 2.2 (a–d) Waveform and different sinusoids of a fixed frequency $\omega = 262$ Hz but different phases $\varphi \in \{0.05, 0.24, 0.45, 0.6\}$. (e) Values that express the degree of similarity between the waveform and the four different sinusoids.

for $t \in \mathbb{R}$. Since the cosine function is periodic, the parameters φ and $\varphi + k$ for integers $k \in \mathbb{Z}$ yield the same function. Therefore, the phase parameter only needs to be considered for $\varphi \in [0, 1)$.

When measuring how well the given signal coincides with a sinusoid of frequency ω , we have the freedom of shifting the sinusoid in time. This degree of freedom is expressed by the phase parameter φ . As illustrated by Figure 2.2, the degree of similarity between the signal and the sinusoid of fixed frequency crucially depends on the phase. What have we done with the phase when computing the coefficients d_ω as illustrated by Figure 2.1? The procedure outlined in the introduction was only half the story. When comparing the signal f with a sinusoid $\cos_{\omega, \varphi}$ of frequency ω , we have implicitly used the phase φ_ω that yields the maximal possible similarity. To understand this better, we first need to explain how we actually compare the signal and a sinusoid or, more generally, how we compare two given functions.

2.1.1.2 Computing Similarity with Integrals

Let us assume that we are given two functions of time $f : \mathbb{R} \rightarrow \mathbb{R}$ and $g : \mathbb{R} \rightarrow \mathbb{R}$. What does it mean for f and g to be similar? Intuitively, one may agree that f and g are similar if they show a similar behavior over time: if f assumes positive values, then so should g , and if f becomes negative, the same should happen to g . The joint behavior of these functions can be captured by forming the integral of the product of the two functions:

$$\int_{t \in \mathbb{R}} f(t) \cdot g(t) dt. \quad (2.3)$$

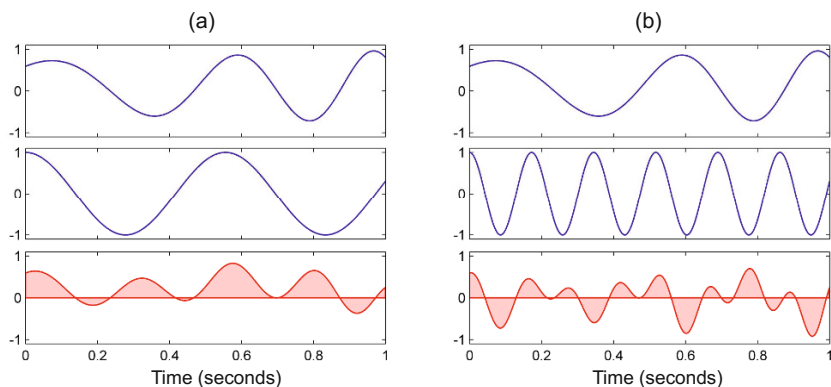


Fig. 2.3 Measuring the similarity of two functions f (top) and g (middle) by computing the integral of the product (bottom). **(a)** Two functions having high similarity. **(b)** Two functions having low similarity.

The integral measures the area delimited by the graph of the product $f \cdot g$, where the negative area (below the horizontal axis) is subtracted from the positive area (above the horizontal axis) (see Figure 2.3). In the case that f and g are either both positive or both negative at most time instances, the product is positive for most of the time and the integral becomes large (see Figure 2.3a). However, if the two functions are dissimilar, then the overall positive and the overall negative areas cancel out, yielding a small overall integral (see Figure 2.3b). Further examples are discussed in Exercise 2.1.

There are many more ways for comparing two given signals. For example, the integral of the absolute difference between the functions also yields a notion of how similar the signals are. In the formulation of the Fourier transform, however, one encounters the measure as considered in (2.3), which generalizes the **inner product** known from linear algebra (see 2.37). We continue this discussion in Section 2.2.3.

2.1.1.3 First Definition of the Fourier Transform

Based on the similarity measure (2.3), we compare the original signal f with sinusoids $g = \cos_{\omega, \varphi}$ as defined in (2.2). For a fixed frequency $\omega \in \mathbb{R}$, we define

$$d_{\omega} := \max_{\varphi \in [0, 1)} \left(\int_{t \in \mathbb{R}} f(t) \cos_{\omega, \varphi}(t) dt \right), \quad (2.4)$$

$$\varphi_{\omega} := \operatorname{argmax}_{\varphi \in [0, 1)} \left(\int_{t \in \mathbb{R}} f(t) \cos_{\omega, \varphi}(t) dt \right). \quad (2.5)$$

As previously discussed, the magnitude coefficient d_{ω} expresses the intensity of frequency ω within the signal f . Additionally, the phase coefficient $\varphi_{\omega} \in [0, 1)$ tells

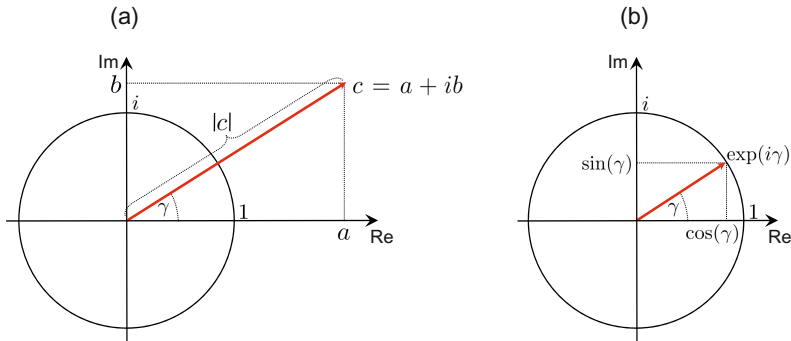


Fig. 2.4 (a) Polar coordinate representation of a complex number $c = a + ib$. (b) Definition of the exponential function.

us how the sinusoid of frequency ω needs to be displaced in time to best fit the signal f . The **Fourier transform** of a function $f : \mathbb{R} \rightarrow \mathbb{R}$ is defined to be the “collection” of all coefficients d_ω and ϕ_ω for $\omega \in \mathbb{R}$. Shortly, we will state this definition in a more formal way.

The computation of d_ω and ϕ_ω feels a bit awkward, since it involves an optimization step. The good news is that there is a simple solution to this optimization problem, which results from the existence of certain trigonometric identities that relate phases and amplitudes of certain sinusoidal functions. Using the concept of complex numbers, these trigonometric identities become simple and lead to an elegant formulation of the Fourier transform. We discuss such issues in more detail in Section 2.3. In the following, we introduce the standard complex-valued formulation of the Fourier transform without giving any proofs.

2.1.1.4 Complex Numbers

Let us first review the concept of complex numbers. The complex numbers extend the real numbers by introducing the imaginary number $i := \sqrt{-1}$ with the property $i^2 = -1$. Each complex number can be written as $c = a + ib$, where $a \in \mathbb{R}$ is the real part and $b \in \mathbb{R}$ the imaginary part of c . The set of all complex numbers is written as \mathbb{C} , which can be thought of as a two-dimensional plane: the horizontal dimension corresponds to the real part, and the vertical dimension to the imaginary part. In this plane, the number $c = a + ib$ is specified by the Cartesian coordinates (a, b) . As illustrated by Figure 2.4a, there is another way of representing a complex number, which is known as the **polar coordinate** representation. In this case, a complex number c is described by its absolute value $|c|$ (distance from the origin) and the angle γ between the positive horizontal axis and the line from the origin and c . The polar coordinates $|c| \in \mathbb{R}_{\geq 0}$ and $\gamma \in [0, 2\pi)$ (given in radians) can be derived from the coordinates (a, b) via the following formulas:

$$|c| := \sqrt{a^2 + b^2}, \quad (2.6)$$

$$\gamma := \text{atan2}(b, a). \quad (2.7)$$

Further details on polar coordinates and the function atan2 , which is a variant of the inverse of the tangent function, are explained in Section 2.3.2.2. To regain the complex number c from its polar coordinates, one uses the **exponential function**, which maps an angle $\gamma \in \mathbb{R}$ (given in radians) to a complex number defined by

$$\exp(i\gamma) := \cos(\gamma) + i\sin(\gamma) \quad (2.8)$$

(see also Figure 2.4b). The values of this function turn around the unit circle of the complex plane with a period of 2π (see Section 2.3.2.1). From this, we obtain the following **polar coordinate representation** for a complex number c :

$$c = |c| \cdot \exp(i\gamma). \quad (2.9)$$

2.1.1.5 Complex Definition of the Fourier Transform

What have we gained by bringing complex numbers into play? Recall that we have obtained a positive coefficient $d_\omega \in \mathbb{R}_{\geq 0}$ from (2.4) and a phase coefficient $\varphi_\omega \in [0, 1)$ from (2.5). The basic idea is to use these coefficients as polar coordinates and to encode both coefficients by a single complex number. Because of some technical reasons (a normalization issue that becomes clearer when discussing the mathematical details), one introduces some additional factors and a sign in the phase to yield the complex coefficient

$$c_\omega := \frac{d_\omega}{\sqrt{2}} \cdot \exp(2\pi i(-\varphi_\omega)). \quad (2.10)$$

This complex formulation directly leads us to the Fourier transform of a real-valued function $f: \mathbb{R} \rightarrow \mathbb{R}$. For each frequency $\omega \in \mathbb{R}$, we obtain a complex-valued coefficient $c_\omega \in \mathbb{C}$ as defined by (2.4), (2.5), and (2.10). This collection of coefficients can be encoded by a complex-valued function $\hat{f}: \mathbb{R} \rightarrow \mathbb{C}$ (called “ f hat”), which assigns to each frequency parameter the coefficient c_ω :

$$\hat{f}(\omega) := c_\omega. \quad (2.11)$$

The function \hat{f} is referred to as the **Fourier transform** of f , and its values $\hat{f}(\omega) = c_\omega$ are called the **Fourier coefficients**. One main result in Fourier analysis is that the Fourier transform can be computed via the following compact formula:

$$\hat{f}(\omega) = \int_{t \in \mathbb{R}} f(t) \exp(-2\pi i \omega t) dt \quad (2.12)$$

$$= \int_{t \in \mathbb{R}} f(t) \cos(-2\pi \omega t) dt + i \int_{t \in \mathbb{R}} f(t) \sin(-2\pi \omega t) dt. \quad (2.13)$$

In other words, the real part of the complex coefficient $\hat{f}(\omega)$ is obtained by comparing the original signal f with a cosine function of frequency ω , and the imaginary part is obtained by comparing with a sine function of frequency ω . The absolute value $|\hat{f}(\omega)|$ is also called the **magnitude** of the Fourier coefficient. Similarly, the real-valued function $|\hat{f}| : \mathbb{R} \rightarrow \mathbb{R}$, which assigns to each frequency parameter ω the magnitude $|\hat{f}(\omega)|$, is called the **magnitude Fourier transform** of f .

In the standard literature on signal processing, the formula (2.12) is often used to define the Fourier transform \hat{f} and, then, the physical interpretation of the Fourier coefficients is discussed. In particular, the real-valued coefficients d_ω in (2.4) and φ_ω in (2.5) can be derived from $\hat{f}(\omega)$. Using (2.10), one obtains

$$d_\omega = \sqrt{2}|\hat{f}(\omega)|, \quad (2.14)$$

$$\varphi_\omega = -\frac{\gamma_\omega}{2\pi}, \quad (2.15)$$

where $|\hat{f}(\omega)|$ and γ_ω are the polar coordinates of $\hat{f}(\omega)$.

2.1.1.6 Fourier Representation

As mentioned above, the original signal f can be reconstructed from its Fourier transform. In principle, the reconstruction is straightforward: one superimposes the sinusoids of all possible frequency parameters $\omega \in \mathbb{R}$, each weighted by the respective coefficient d_ω and shifted by φ_ω . Both kinds of information are encoded in the complex Fourier coefficient c_ω . In the analog case considered so far, we are dealing with a continuum of frequency parameters, where the superposition becomes an integration over the parameter space. The reconstruction is given by the formulas

$$f(t) = \int_{\omega \in \mathbb{R}_{\geq 0}} d_\omega \sqrt{2} \cos(2\pi(\omega t - \varphi_\omega)) d\omega \quad (2.16)$$

$$= \int_{\omega \in \mathbb{R}} c_\omega \exp(2\pi i \omega t) d\omega, \quad (2.17)$$

first given in the real-valued formulation, and then given in the complex-valued formulation with $c_\omega = \hat{f}(\omega)$. As said before, the representation of a signal in terms of a weighted superposition of sinusoidal prototype oscillations is also called the **Fourier representation** of the signal. Notice that the formula (2.12) for the Fourier transform and the formula (2.17) for the Fourier representation are nearly identical. The main difference is that the roles of the time parameter t and frequency parameter ω are interchanged. The beautiful relationship between these two formulas will be further discussed in later sections of this chapter.

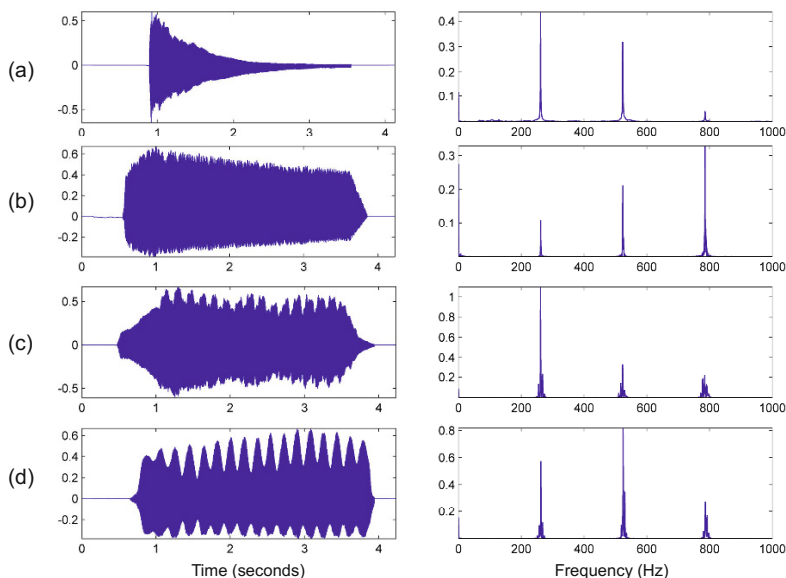


Fig. 2.5 Waveform and magnitude Fourier transform of a tone C4 (261.6 Hz) played by different instruments (see also Figure 1.23). (a) Piano. (b) Trumpet. (c) Violin. (d) Flute.

2.1.2 Examples

Let us consider some examples including the one introduced in Figure 2.1. Figure 2.5 shows the waveform and the magnitude Fourier transform for some audio signals, where a single note C4 is played on different instruments: a piano, a trumpet, a violin, and a flute. We have already encountered this example in Figure 1.23 of Section 1.3.4, where we discussed the aspect of timbre. Recall that the existence of certain partials and their relative strengths have a crucial influence on the timbre of a musical tone. In the case of the piano tone (Figure 2.5a), the Fourier transform has a sharp peak at 262 Hz, which reveals that most of the signal's energy is contained in the first partial or the fundamental frequency of the note C4. Further peaks (also beyond the shown frequency range from 0 to 1000 Hz) can be found at integer multiples of the fundamental frequency corresponding to the higher partials.

Figure 2.5b shows that the same note played on a trumpet results in a similar frequency spectrum, where the peaks appear again at integer multiples of the fundamental frequency. However, most of the energy is now contained in the third partial, and the relative heights of the peaks are different compared with the piano. This is one reason why a trumpet sounds different from a piano. For a violin, as shown by Figure 2.5c, most energy is again contained in the first partial. Observe that the peaks are blurred in frequency, which is the result of the vibrato (see also Figure 1.23b). The time-dependent frequency modulations of the vibrato are aver-

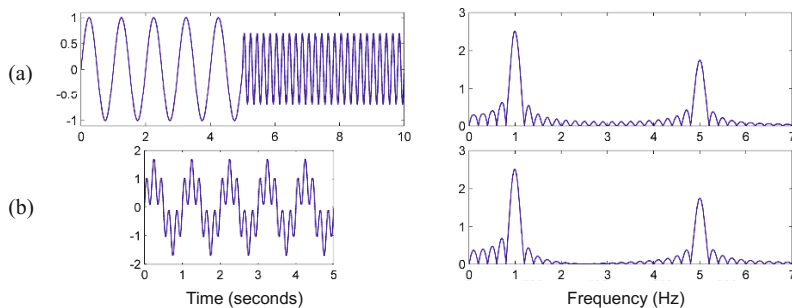


Fig. 2.6 Missing time information of the Fourier transform illustrated by two different signals and their magnitude Fourier transforms. **(a)** Two subsequent sinusoids of frequency 1 Hz and 5 Hz. **(b)** Superposition of the same sinusoids.

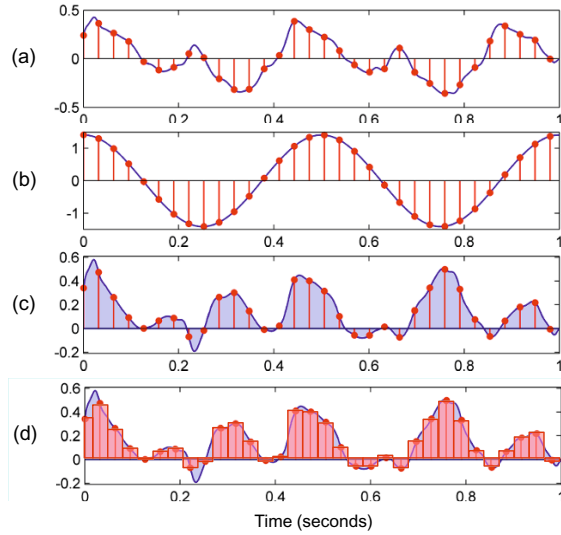
aged by the Fourier transform. This yields a single coefficient for each frequency independent of spectro-temporal fluctuations. A similar explanation holds for the flute tone shown in Figure 2.5d.

We have seen that the magnitude of the Fourier transform tells us about the signal’s overall frequency content, but it does not tell us at which time the frequency content occurs. Figure 2.6 illustrates this fact, showing the waveform and the magnitude Fourier transform for two signals. The first signal consists of two parts with a sinusoid of $\omega = 1$ Hz and amplitude $A = 1$ in the first part and a sinusoid of $\omega = 5$ Hz and amplitude $A = 0.7$ in the second part. Furthermore, the signal is zero outside the interval $[0, 10]$. In contrast, the second signal is a superposition of these two sinusoids, being zero outside the interval $[0, 5]$. Even though the two signals are different in nature, the resulting magnitude Fourier transforms are more or less the same. This demonstrates the drawbacks of the Fourier transform when analyzing signals with changing characteristics over time. In Section 2.1.4 and Section 2.5 we discuss a short-time version of the Fourier transform, where time information is recovered at least to some degree. Besides the two peaks, one can observe in Figure 2.6 a large number of small “ripples.” Such phenomena as well as further properties of the Fourier transform are discussed in Section 2.3.3.

2.1.3 Discrete Fourier Transform

When using digital technology, only a finite number of parameters can be stored and processed. To this end, analog signals need to be converted into finite representations—a process commonly referred to as **digitization**. One step that is often applied in an analog-to-digital conversion is known as **equidistant sampling**. Given an analog signal $f : \mathbb{R} \rightarrow \mathbb{R}$ and a positive real number $T > 0$, one defines a function $x : \mathbb{Z} \rightarrow \mathbb{R}$ by setting

Fig. 2.7 Illustration of the sampling process using a sampling rate of $F_s = 32$. The waveforms of the analog signals are shown as curves and the sampled versions as stem plots. **(a)** Signal f . **(b)** Sinusoid $\cos_{\omega, \varphi}$ with $\omega = 2$ and $\varphi = 0$. **(c)** Product $f \cdot \cos_{\omega, \varphi}$ and its area. **(d)** Approximation of the integral by a Riemann sum obtained from the sampled version.



$$x(n) := f(n \cdot T). \quad (2.18)$$

Since x is only defined on a discrete set of time points, it is also referred to as a **discrete-time** (DT) signal (see Section 2.2.2.1). The value $x(n)$ is called a **sample** taken at time $t = n \cdot T$ of the original analog signal f . This procedure is also known as **T -sampling**, where the number T is referred to as the **sampling period**. The inverse

$$F_s := 1/T \quad (2.19)$$

of the sampling period is also called the **sampling rate** of the process. It specifies the number of samples per second and is measured in Hertz (Hz). Figure 2.7a shows an example of sampling an analog signal using $F_s = 32$ Hz.

In general, one loses information in the sampling process. The famous **sampling theorem** says that the original analog signal f can be reconstructed perfectly from its sampled version x , if f does not contain any frequencies higher than

$$\Omega := F_s/2 = 1/(2T) \text{ Hz}. \quad (2.20)$$

In this case, we also say that f is an **Ω -bandlimited** signal, where the frequency Ω is known as the **Nyquist frequency**. In the case that f contains higher frequencies, sampling may cause artifacts referred to as **aliasing** (see Section 2.2.2 for details). The sampling theorem will be further discussed in Exercise 2.28.

In the following, we assume that the analog signal f satisfies suitable requirements so that the sampled signal x does not contain major artifacts. Now, having a discrete number of samples to represent our signal, how do we calculate the Fourier transform? Recall that the idea of the Fourier transform is to compare the signal with a sinusoidal prototype oscillation by computing the integral over the point-

wise product (see (2.12)). Therefore, in the digital domain, it seems reasonable to sample the sinusoidal prototype oscillation in the same fashion as the signal (see Figure 2.7b). By multiplying the two sampled functions in a pointwise fashion, we obtain a sampled product (see Figure 2.7c). Finally, integration in the analog case becomes summation in the discrete case, where the summands need to be weighted by the sampling period T . As a result, one obtains the following approximation:

$$\sum_{n \in \mathbb{Z}} T f(nT) \exp(-2\pi i \omega nT) \approx \hat{f}(\omega). \quad (2.21)$$

In mathematical terms, the sum can be interpreted as the overall area of rectangular shapes that approximates the area corresponding to the integral (see Figure 2.7d). Such an approximation is also known as a **Riemann sum**. As we will show in Section 2.3.4, the quality of the approximation is good for “well-behaved” signals f and “small” frequency parameters ω .

One defines a discrete version of the Fourier transform for a given DT-signal $x : \mathbb{Z} \rightarrow \mathbb{R}$ by setting

$$\hat{x}(\omega) := \sum_{n \in \mathbb{Z}} x(n) \exp(-2\pi i \omega n). \quad (2.22)$$

In this definition, where a simple 1-sampling (i.e., T -sampling with $T = 1$) of the exponential function is used, one does not assume that one knows the relation between x and the original signal f . If one is interested in recovering the relation to the Fourier transform \hat{f} , one needs to know the sampling period T . Based on (2.21), an easy calculation shows that

$$\hat{x}(\omega) \approx \frac{1}{T} \hat{f}\left(\frac{\omega}{T}\right). \quad (2.23)$$

In this approximation, the frequency parameter ω used for \hat{x} corresponds to the frequency ω/T for \hat{f} . In particular, $\omega = 1/2$ for \hat{x} corresponds to the Nyquist frequency $\Omega = 1/(2T)$ of the sampling process. Therefore, assuming that f is bandlimited by $\Omega = 1/(2T)$, one needs to consider only the frequencies with $0 \leq \omega \leq 1/2$ for \hat{x} . In the digital case, all other frequency parameters are redundant and yield meaningless approximations.

For doing computations on digital machines, we still have some problems. One problem is that the sum in (2.22) involves an infinite number of summands. Another problem is that the frequency parameter ω is a continuous parameter. For both problems, there are some pragmatic solutions. Regarding the first problem, we assume that most of the relevant information of f is limited to a certain duration in time.² For example, a music recording of a song hardly lasts for more than ten minutes. Having a finite duration means that the analog signal f is assumed to be zero outside a compact interval. By possibly shifting the signal, we may assume that this interval starts at time $t = 0$. This means that we only need to consider a finite number of

² Strictly speaking, this assumption is problematic since it conflicts with the requirement of f being bandlimited. A mathematical fact states that there are no functions that are both limited in frequency (bandlimited) and limited in time (having finite duration).

samples $x(0), x(1), \dots, x(N-1)$ for some suitable number $N \in \mathbb{N}$. As a result, the sum in (2.22) becomes finite.

Regarding the second problem, one computes the Fourier transform only for a finite number of frequencies. Similar to the sampling of the time axis, one typically samples the frequency axis by considering the frequencies $\omega = k/M$ for some suitable $M \in \mathbb{N}$ and $k \in [0 : M-1]$. In practice, one often couples the number N of samples and the number M that determines the frequency resolution by setting $N = M$. Note that the two numbers N and M refer to different aspects. However, the coupling is convenient. It not only makes the resulting transform invertible, but also leads to a computationally efficient algorithm, as we will see in Section 2.4.3. Setting $X(k) := \hat{x}(k/N)$ and assuming that $x(0), x(1), \dots, x(N-1)$ are the relevant samples (all others being zero), we obtain from (2.22) the formula

$$X(k) = \hat{x}(k/N) = \sum_{n=0}^{N-1} x(n) \exp(-2\pi i k n / N) \quad (2.24)$$

for integers $k \in [0 : M-1] = [0 : N-1]$. This transform is also known as the **discrete Fourier transform** (DFT), which is covered in Section 2.4.

Next, let us have a look at the frequency information supplied by the Fourier coefficient $X(k)$. By (2.23) the frequency ω of \hat{x} corresponds to ω/T of \hat{f} . Therefore, the index k of $X(k)$ corresponds to the physical frequency

$$F_{\text{coef}}(k) := \frac{k}{N \cdot T} = \frac{k \cdot F_s}{N} \quad (2.25)$$

given in Hertz. As we will discuss in Section 2.4.4, the coefficients $X(k)$ need to be taken with care. First, the approximation quality in (2.23) may be rather poor, in particular for frequencies close to the Nyquist frequency. Second, for a real-valued signal x , the Fourier transform fulfills certain symmetry properties (see Exercise 2.24). As a result, the upper half of the Fourier coefficients are redundant, and one only needs to consider the coefficients $X(k)$ for $k \in [0 : \lfloor N/2 \rfloor]$. Note that, in the case of an even number N , the index $k = N/2$ corresponds to $F_{\text{coef}}(k) = F_s/2$, which is the Nyquist frequency of the sampling process.

Finally, we consider some efficiency issues when computing the DFT. To compute a single Fourier coefficient $X(k)$, one requires a number of multiplications and additions linear in N . Therefore, to compute all coefficients $X(k)$ for $k \in [0 : N/2]$ one after another, one requires a number of operations on the order of N^2 . Despite being a finite number of operations, such a computational approach is too slow for many practical applications, in particular when N is large.

The number of operations can be reduced drastically by using an efficient algorithm known as the **fast Fourier transform** (FFT). The FFT algorithm, which was discovered by Gauss and Fourier two hundred years ago, has changed whole industries and is now being used in billions of telecommunication and other devices. The FFT exploits redundancies across sinusoids of different frequencies to jointly compute all Fourier coefficients by a recursion. This recursion works particularly well in the case that N is a power of two. As a result, the FFT reduces the overall number of

operations from the order of N^2 to the order of $N \log_2 N$. The savings are enormous. For example, using $N = 2^{10} = 1024$, the FFT requires roughly $N \log_2 N = 10240$ instead of $N^2 = 1048576$ operations in the naive approach—a savings factor of about 100. In the case of $N = 2^{20}$, the savings amount to a factor of about 50000 (see Exercise 2.6). In Section 2.4.3, we discuss the algorithmic details of the FFT.

2.1.4 Short-Time Fourier Transform

The Fourier transform yields frequency information that is averaged over the entire time domain. However, the information on *when* these frequencies occur is hidden in the transform. We have already seen this phenomenon in Figure 2.6a, where the change in frequency is not revealed when looking at the magnitude of the Fourier transform. To recover the hidden time information, Dennis Gabor introduced in the year 1946 the **short-time Fourier transform** (STFT). Instead of considering the entire signal, the main idea of the STFT is to consider only a small section of the signal. To this end, one fixes a so-called **window function**, which is a function that is nonzero for only a short period of time (defining the considered section). The original signal is then multiplied with the window function to yield a **windowed signal**. To obtain frequency information at different time instances, one shifts the window function across time and computes a Fourier transform for each of the resulting windowed signals.

This idea is illustrated by Figure 2.8, which continues our example from Figure 2.6a. To obtain local sections of the original signal, one multiplies the signal with suitably shifted rectangular window functions. In Figure 2.8b, the resulting local section only contains frequency content at 1 Hz, which leads to a single main peak in the Fourier transform at $\omega = 1$. Further shifting the time window to the right, the resulting section contains 1 Hz as well as 5 Hz components (see Figure 2.8c). These components are reflected by the two peaks at $\omega = 1$ and $\omega = 5$. Finally, the section shown in Figure 2.8d only contains frequency content at 5 Hz.

Already at this point, we want to emphasize that the STFT reflects not only the properties of the original signal but also those of the window function. First of all, the STFT depends on the length of the window, which determines the size of the section. Then, the STFT is influenced by the shape of the window function. For example, the sharp edges of the rectangular window typically introduce “ripple” artifacts. In Section 2.5.1, we discuss such issues in more detail. In particular, we introduce more suitable, bell-shaped window functions, which typically reduce such artifacts.

In Section 2.5, one finds a detailed treatment of the analog and discrete versions of the STFT and their relationship. In the following, we only consider the discrete case and specify the most important mathematical formulas as needed in practical applications. Let $x : \mathbb{Z} \rightarrow \mathbb{R}$ be a real-valued DT-signal obtained by equidistant sampling with respect to a fixed sampling rate F_s given in Hertz. Furthermore, let $w : [0 : N - 1] \rightarrow \mathbb{R}$ be a sampled window function of length $N \in \mathbb{N}$. For example,

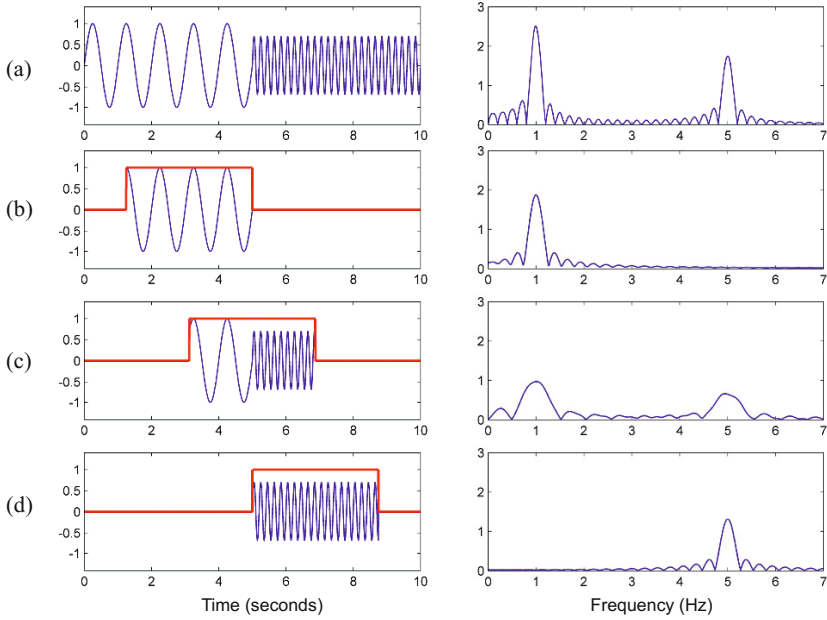


Fig. 2.8 Signal and Fourier transform consisting of two subsequent sinusoids of frequency 1 Hz and 5 Hz (see Figure 2.6a). **(a)** Original signal. **(b)** Windowed signal centered at $t = 3$. **(c)** Windowed signal centered at $t = 5$. **(d)** Windowed signal centered at $t = 7$.

in the case of a rectangular window one has $w(n) = 1$ for $n \in [0 : N - 1]$. Implicitly, one assumes that $w(n) = 0$ for all other time parameters $n \in \mathbb{Z} \setminus [0 : N - 1]$ outside this window. The length parameter N determines the duration of the considered sections, which amounts to N/F_s seconds. One also introduces an additional parameter $H \in \mathbb{N}$, which is referred to as the **hop size**. The hop size parameter is specified in samples and determines the step size in which the window is to be shifted across the signal.

With regard to these parameters, the **discrete STFT** \mathcal{X} of the signal x is given by

$$\mathcal{X}(m, k) := \sum_{n=0}^{N-1} x(n + mH)w(n) \exp(-2\pi i k n / N) \quad (2.26)$$

with $m \in \mathbb{Z}$ and $k \in [0 : K]$. The number $K = N/2$ (assuming that N is even) is the frequency index corresponding to the Nyquist frequency. The complex number $\mathcal{X}(m, k)$ denotes the k^{th} Fourier coefficient for the m^{th} time frame. Note that for each fixed time frame m , one obtains a **spectral vector** of size $K + 1$ given by the coefficients $\mathcal{X}(m, k)$ for $k \in [0 : K]$. The computation of each such spectral vector amounts to a DFT of size N as in (2.24), which can be done efficiently using the FFT.

What have we actually computed in (2.26) in relation to the original analog signal f ? As for the temporal dimension, each Fourier coefficient $\mathcal{X}(m, k)$ is associated with the physical time position

$$T_{\text{coef}}(m) := \frac{m \cdot H}{F_s} \quad (2.27)$$

given in seconds. For example, for the smallest possible hop size $H = 1$, one obtains $T_{\text{coef}}(m) = m/F_s = m \cdot T$ sec. In this case, one obtains a spectral vector for each sample of the DT-signal x , which results in a huge increase in data volume. Furthermore, considering sections that are only shifted by one sample generally yields very similar spectral vectors. To reduce this type of redundancy, one typically relates the hop size to the length N of the window. For example, one often chooses $H = N/2$, which constitutes a good trade-off between a reasonable temporal resolution and the data volume comprising all generated spectral coefficients. As for the frequency dimension, we have seen in (2.25) that the index k of $\mathcal{X}(m, k)$ corresponds to the physical frequency

$$F_{\text{coef}}(k) := \frac{k \cdot F_s}{N} \quad (2.28)$$

given in Hertz.

Before we look at some concrete examples, we first introduce the concept of a **spectrogram**, which we denote by \mathcal{Y} . The spectrogram is a two-dimensional representation of the squared magnitude of the STFT:

$$\mathcal{Y}(m, k) := |\mathcal{X}(m, k)|^2. \quad (2.29)$$

It can be visualized by means of a two-dimensional image, where the horizontal axis represents time and the vertical axis represents frequency. In this image, the spectrogram value $\mathcal{Y}(m, k)$ is represented by the intensity or color in the image at the coordinate (m, k) . Note that in the discrete case, the time axis is indexed by the frame indices m and the frequency axis is indexed by the frequency indices k .

Continuing our running example from Figure 2.8, we now consider a sampled version of the analog signal using a sampling rate of $F_s = 32$ Hz. Having a physical duration of 10 sec, this results in 320 samples (see Figure 2.9a). Using a window length of $N = 64$ samples and a hop size of $H = 8$ samples, we obtain the spectrogram as shown in Figure 2.9b. In the image, the shade of gray encodes the magnitude of a spectral coefficient, where darker colors correspond to larger values. By (2.27), the m^{th} frame corresponds to the physical time $T_{\text{coef}}(m) = m/4$ sec. In other words, the STFT has a time resolution of four frames per second. Furthermore, by (2.28), the k^{th} Fourier coefficient corresponds to the physical frequency $F_{\text{coef}}(k) := k/2$ Hz. In other words, one obtains a frequency resolution of two coefficients per Hertz. The plots of the waveform and the spectrogram with the physically correct time and frequency axes are shown in Figure 2.9c and Figure 2.9d, respectively.

Let us consider some typical settings as encountered when processing music signals. For example, in the case of CD recordings one has a sampling rate of $F_s = 44100$ Hz. Using a window length of $N = 4096$ and a hop size of $H = N/2$,

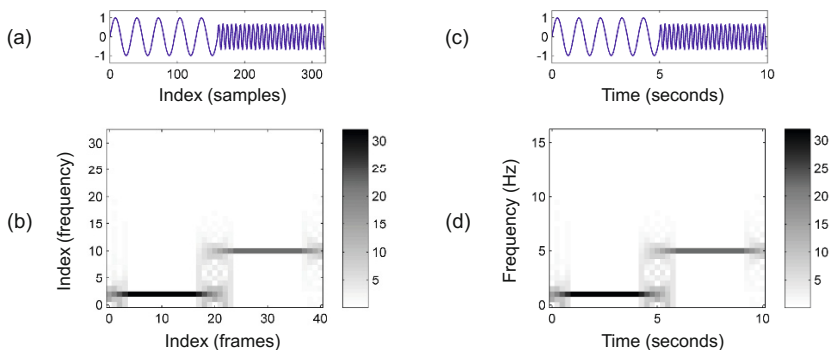


Fig. 2.9 DT-signal sampled with $F_s = 32$ Hz and STFT using a window length of $N = 64$ and a hop size of $H = 8$. **(a)** DT-signal with time axis given in samples. **(b)** STFT with time axis given in frames and frequency axis given in indices. **(c)** DT-signal with time axis given in seconds. **(d)** STFT with time axis given in seconds and frequency axis given in Hertz.

this results in a time resolution of $H/F_s \approx 46.4$ ms by (2.27) and a frequency resolution of $F_s/N \approx 10.8$ Hz by (2.28). To obtain a better frequency resolution, one may increase the window length N . This, however, leads to a poorer localization in time so that the resulting STFT loses its capability of capturing local phenomena in the signal. This kind of trade-off is further discussed in Section 2.5.2 and in the exercises.

We close this section with a further example shown in Figure 2.10, which is a recording of a C-major scale played on a piano. The first note of this scale is C4, which we have already considered in Figure 2.1. In Figure 2.10c, the spectrogram representation of the recording is shown, where the time and frequency axes are labeled in a physically meaningful way. The spectrogram reveals the frequency information of the played notes over time. For each note, one can observe horizontal lines that are stacked on top of each other. As discussed in Section 1.3.4, these equally spaced lines correspond to the partials, the integer multiples of the fundamental frequency of a note. Obviously, the higher partials contain less and less of the signal's energy. Furthermore, the decay of each note over time is reflected by the fading out of the horizontal lines. To enhance small sound components that may still be perceptually relevant, one often uses a logarithmic dB scale (see Section 1.3.3). Figure 2.10d illustrates the effect when applying the dB scale to the values of the spectrogram. Besides an enhancement of the higher partials, one can now observe vertical structures at the notes' onset positions. These structures correspond to the noise-like transients that occur in the attack phase of the piano sound (see Section 1.3.4).

This concludes our “nutshell section” covering the most important definitions and properties of the Fourier transform as needed for the subsequent chapters of this book. In particular, the formula (2.26) of the discrete STFT as well as the physical interpretation of the time parameter (2.27) and the frequency parameter (2.28) are

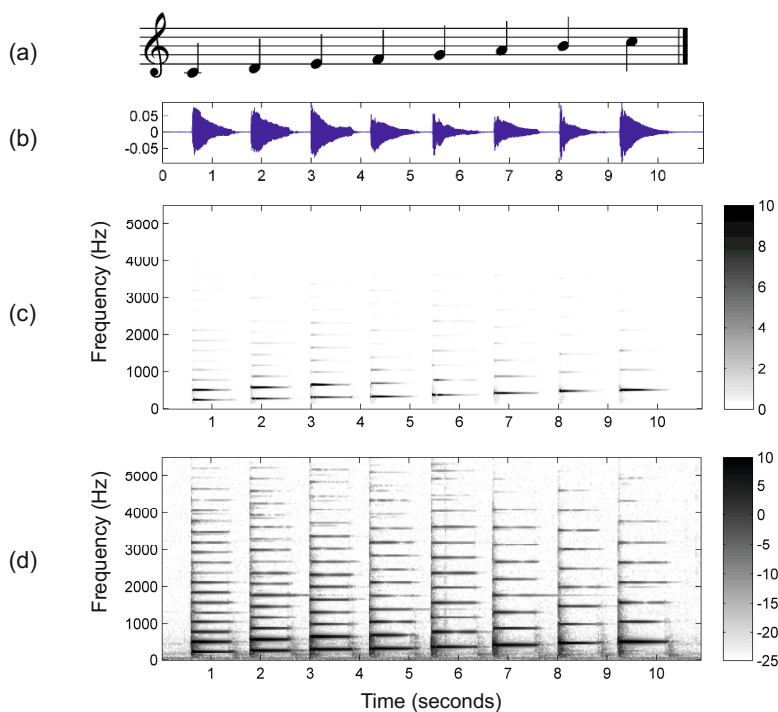


Fig. 2.10 Waveform and spectrogram of a music recording of a C-major scale played on a piano. (a) The recording's underlying musical score. (b) Waveform. (c) Spectrogram. (d) Spectrogram with the magnitudes given in dB.

of central importance for most music processing applications to be discussed. As said in the introduction, we provide in the subsequent sections of this chapter some deeper insights into the mathematics underlying the Fourier transform. In particular, we explain in more detail the connection between the various kinds of signals and associated Fourier transforms.

2.2 Signals and Signal Spaces

In technical fields such as engineering or computer science, a **signal** is a function that conveys information about the state or behavior of a physical system. For example, a signal may describe the time-varying sound pressure at some place, the motion of a particle through some space, the distribution of light on a screen representing an image, or the sequence of images as in the case of a video signal. In the following, we consider the case of audio signals as discussed in Section 1.3. We have seen that such a signal can be graphically represented by its waveform, which

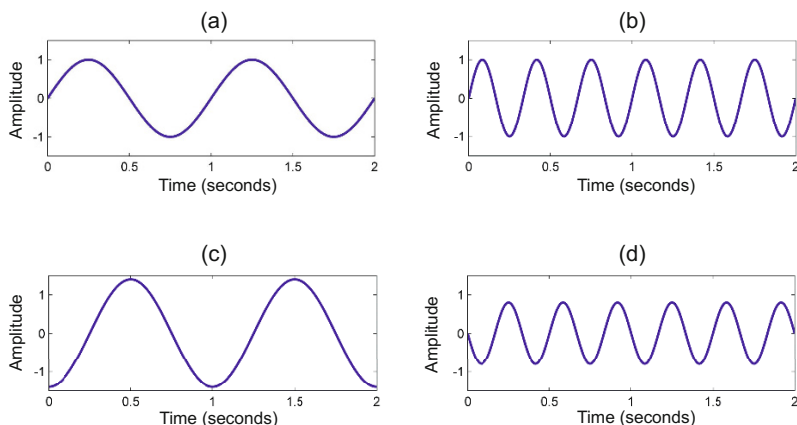


Fig. 2.11 The sinusoid $f(t) = A \sin(2\pi(\omega t - \varphi))$ displayed for $t \in [0, 2]$ and for various values of A , ω , and φ . (a) $A = 1$, $\omega = 1$, $\varphi = 0$. (b) $A = 1$, $\omega = 3$, $\varphi = 0$. (c) $A = 1.4$, $\omega = 1$, $\varphi = 0.25$. (d) $A = 0.8$, $\omega = 3$, $\varphi = 0.5$.

depicts the amplitude of the air pressure over time. In the following, we introduce the mathematical notation that is necessary to formally model such a signal. Doing so, we distinguish between two different types of signals: **analog signals** as occur around us in the real world and **digital signals** as are processed by computers. We show how signals can be modified and combined to yield new signals by applying mathematical operations. Some operations can be applied only if the involved signals satisfy certain properties. This leads us to the concept of **signal spaces**, a kind of universe that comprises signals that share a certain property.

2.2.1 Analog Signals

As already defined in Section 2.1.1, an **analog** signal is a function $f: \mathbb{R} \rightarrow \mathbb{R}$, which assigns an amplitude value $f(t) \in \mathbb{R}$ to each time point $t \in \mathbb{R}$. In the analog case, both the time domain as well as the range of the amplitude values are represented by the set \mathbb{R} of real numbers, which is a continuous range of values. This makes it possible to model infinitesimally small changes in both time and amplitude. In the case of having a continuous time axis (given by \mathbb{R}), one also speaks of **continuous-time** (CT) signals. A signal f is called **periodic** with **period** $\lambda \in \mathbb{R}_{>0}$ if $f(t) = f(t + \lambda)$ holds for all $t \in \mathbb{R}$. If there exists a least positive constant with this property, it is called the **prime period** of the signal (see Exercise 2.7 and Exercise 2.8).

In Section 1.3.2 and Section 2.1.1.1, we have already encountered an entire class of analog signals: the **sinusoids**. Recall from (2.1) that a sinusoid is a periodic function f defined by $f(t) := A \sin(2\pi(\omega t - \varphi))$, $t \in \mathbb{R}$. The parameter A describes the **amplitude**, the parameter ω the **frequency**, and the parameter φ the **phase**. The

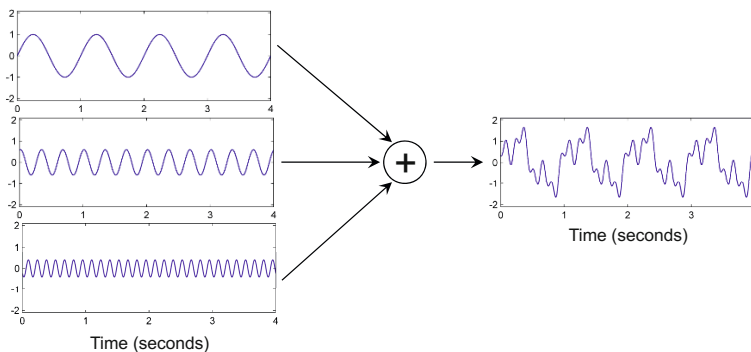


Fig. 2.12 Superposition of three analog signals.

frequency parameter ω determines the period of the sinusoid, which is $\lambda = 1/\omega$. In other words, a sinusoid of frequency ω repeats every $\lambda = 1/\omega$ unit times. In the following, we use seconds as the units of time if not specified otherwise. Figure 2.11 shows various sinusoids resulting from different parameter settings.

Besides having a compact description, sinusoids also have an explicit physical meaning with a perceptual correspondence: the amplitude A corresponds to the loudness and the frequency ω to the pitch of a sinusoidal sound. Only the phase ϕ , which indicates the relative position of an oscillation within its cycle, does not have a direct perceptual correspondence. Note that, because of the periodicity of a sinusoid, a phase shift by $\phi + 1$ has the same effect as a phase shift by ϕ . In other words, integer shifts leave a sinusoid unaltered and the parameter ϕ needs to be considered only in the interval $[0, 1)$.

Regarding a signal as a mathematical function is convenient, since this allows us to express modifications of signals in terms of mathematical operations. For example, the **superposition** of two signals f and g can be expressed by the sum $f + g$ defined as pointwise addition

$$(f + g)(t) := f(t) + g(t) \quad (2.30)$$

for $t \in \mathbb{R}$. Similarly, the **scaling** of a signal f by a real factor a is the scalar multiple af , which is also defined pointwise by

$$(af)(t) := a \cdot f(t). \quad (2.31)$$

Figure 2.12 shows an example of a superposition of three signals. We have seen in Section 2.1 that the Fourier transform can be regarded as a kind of inverse operation, where a given signal is decomposed into a weighted superposition of elementary signals.

2.2.2 Digital Signals

Analog signals have a continuous range of values in both time and amplitude, which generally leads to an infinite number of values. Since a computer can only store and process a finite number of values, one has to convert the waveform into some **discrete** representation—a process commonly referred to as **digitization**. Some analog signals such as sinusoids are already characterized by a small number of parameters, which can be used to represent the signal, but for general analog signals one needs other ways for deriving a model that can be described by a finite number of parameters. Furthermore, it should be possible to perform signal manipulations directly in the parameter domain such that computations become feasible and efficient. The most common approach for digitizing audio signals consists of two steps called **sampling** and **quantization** (see Figure 2.13 for an illustration). We now explain these two steps in more detail.

2.2.2.1 Sampling

In signal processing, the term **sampling** refers to the process of reducing a continuous-time (CT) signal to a **discrete-time** (DT) signal, which is defined only on a discrete subset of the time axis. By means of a suitable encoding, one often assumes that this discrete set is a subset I of the set \mathbb{Z} of integers. Then a DT-signal is defined to be a function $x: I \rightarrow \mathbb{R}$, where the domain I corresponds to points in time. Since one can extend any DT-signal from the domain I to the domain \mathbb{Z} simply by setting all values to zeros for points in $\mathbb{Z} \setminus I$, we may assume $I = \mathbb{Z}$. The most common sampling procedure to transform a CT-signal $f: \mathbb{R} \rightarrow \mathbb{R}$ into a DT-signal $x: \mathbb{Z} \rightarrow \mathbb{R}$ is known as **equidistant sampling**. For convenience, we repeat the definitions from Section 2.1.3. Fixing a positive real number $T > 0$, the DT-signal x is obtained by setting

$$x(n) := f(n \cdot T) \quad (2.32)$$

for $n \in \mathbb{Z}$. The value $x(n)$ is called the **sample** taken at time $t = n \cdot T$ of the original analog signal f . In short, this procedure is also called **T -sampling**. The number T is referred to as the **sampling period** and the inverse $F_s := 1/T$ as the **sampling rate**. The sampling rate specifies the number of samples per second and is measured in Hertz (Hz).

Figure 2.13 shows an illustrative example, where the DT-signal x is represented by the red stem plot. In this example, one has 13 samples in the first two seconds. Thus, the sampling rate is roughly 6.5 Hz and the sampling period 0.154 seconds. In practical applications, typical sampling rates are 8 kHz (8,000 Hz) for telephony, 32 kHz for digital radio, 44.1 kHz for CD recordings, and 48 kHz up to 96 kHz for professional studio technology.

In general, sampling is a **lossy** operation in the sense that information is lost in this process and that the original analog signal cannot be recovered from its sampled version. Only if the analog signal has additional properties in terms of its frequency

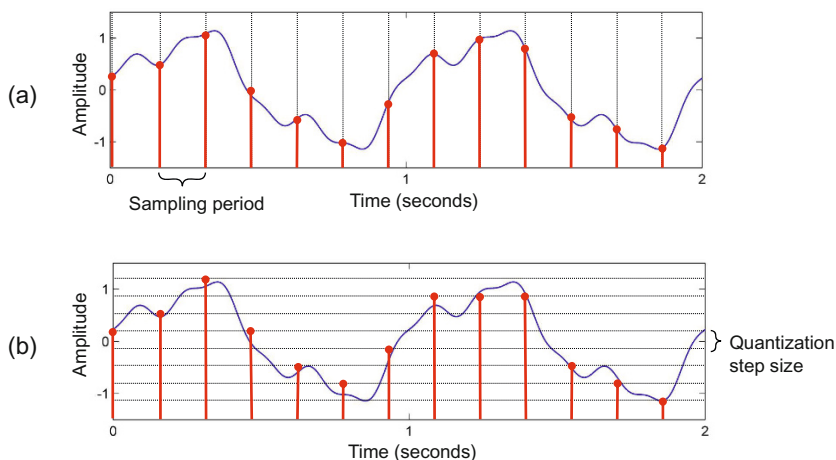


Fig. 2.13 Two steps of a digitization process to transform an analog signal (solid curve) into a digital signal (stem plot). **(a)** Sampling. **(b)** Quantization.

spectrum is a perfect reconstruction possible. This is the assertion of the famous **sampling theorem**, which we discuss in Exercise 2.28 in more detail. Without such additional properties, sampling may cause an effect known as **aliasing**, where certain frequency components of the signal become indistinguishable. This effect is illustrated by Figure 2.14, which shows an analog signal that is the superposition of two sinusoids. Using a high sampling rate as in Figure 2.14a, the analog signal can be reconstructed with high accuracy. However, when decreasing the sampling rate, the higher-frequency component is not captured well and only a coarse approximation of the original signal remains (see Figure 2.14c).

2.2.2.2 Quantization

We have seen how sampling transforms a continuous time axis (encoded by \mathbb{R}) into a discrete time axis (encoded by \mathbb{Z}). This is only the first step in an analog-to-digital conversion of a signal. In the second step, one needs to replace the continuous range of possible amplitudes (again encoded by \mathbb{R}) by a discrete range of possible values (encoded by a discrete set $\Gamma \subset \mathbb{R}$). This process is commonly known as **quantization**. Such a quantization can be modeled by a function $Q: \mathbb{R} \rightarrow \Gamma$, referred to as the **quantizer**, which assigns to each amplitude value $a \in \mathbb{R}$ a value $Q(a) \in \Gamma$. Many of the quantizers used simply round off or truncate the analog value to some units of precision. For example, a typical uniform quantizer with a **quantization step size** equal to some value Δ can be defined by

$$Q(a) := \text{sgn}(a) \cdot \Delta \cdot \left\lfloor \frac{|a|}{\Delta} + \frac{1}{2} \right\rfloor \quad (2.33)$$

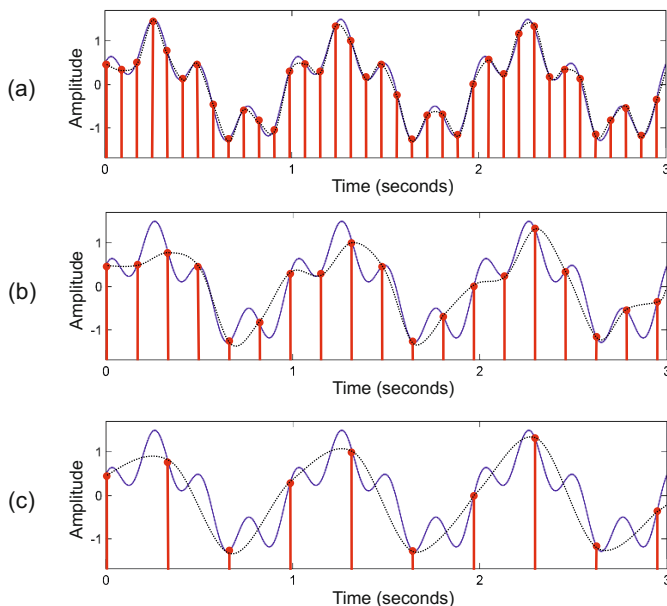


Fig. 2.14 Illustration of the aliasing effect when reducing the sampling rate. The figures show the original analog signal (solid curve), the sampled version (stem plot), and the reconstructed analog signal (dotted curve) for sampling rates of (a) 12 Hz, (b) 6 Hz, and (c) 3 Hz.

for $a \in \mathbb{R}$, where $\text{sgn}(\cdot)$ is the signum function that yields the sign of a real number and the brackets $\lfloor \cdot \rfloor$ truncate a real number to yield the largest integer below this number. Note that, in the case of $\Delta = 1$, the quantizer Q is simple rounding to the nearest integer. Like sampling, quantization is generally a lossy operation, because different analog values may be mapped to the same digital value. The difference between the actual analog value and the quantized value is called the **quantization error** (see Exercise 2.9). Reducing the quantization step size Δ typically leads to smaller quantization errors. However, at the same time, the number of quantized values (and therefore also the number of bits needed to encode these values) increases. Figure 2.13b shows the result after sampling and quantizing an analog signal. In this example, the quantization step size $\Delta = 1/3$ is used, resulting in 8 different quantization values for the given signal. Hence, a 3-bit coding scheme may be used to represent the quantized values. For CD recordings, a 16-bit coding scheme is used, which allows representation of 65536 possible values.

In summary, after using an analog-to-digital conversion based on sampling and quantization, it is generally not possible to reconstruct the original waveform from the digital representation. Aliasing and quantization may introduce audible sound artifacts such as harsh buzzing sounds or noise. For digital representations as used for CDs, however, the sampling rate as well as the quantization resolution are chosen

in such ways that the degradation of the waveform is not noticeable by the human ear.

2.2.3 Signal Spaces

In the previous sections, we considered analog and digital signals, which were modeled as CT-signals $f: \mathbb{R} \rightarrow \mathbb{R}$ and as DT-signals $x: \mathbb{Z} \rightarrow \mathbb{R}$, respectively. In the following discussion, we use the symbols f and g to denote CT-signals and the symbols x and y to denote DT-signals. For the time parameter, we typically use the parameter t in the CT case and the parameter n in the DT case.

2.2.3.1 Complex Numbers

In view of the complex-valued formulation of the Fourier transform one needs to extend the range \mathbb{R} of real numbers to the range \mathbb{C} of **complex numbers**. Recall from Section 2.1.1.4 that each complex number $c \in \mathbb{C}$ can be regarded as a pair $(a, b) \in \mathbb{R}^2$ of real numbers, where $a = \text{Re}(c)$ denotes the real part and $b = \text{Im}(c)$ the imaginary part of c . One also often writes $c = a + ib$, where i is the imaginary unit. The complex number field \mathbb{C} possesses a multiplication that extends the multiplication of the real number field \mathbb{R} . Given two complex numbers $c_1 = a_1 + ib_1, c_2 = a_2 + ib_2 \in \mathbb{C}$, the product is defined by

$$c_1 \cdot c_2 = a_1 a_2 - b_1 b_2 + i(a_1 b_2 + a_2 b_1). \quad (2.34)$$

Furthermore, the **complex conjugate** \bar{c} of a complex number $c = a + ib \in \mathbb{C}$ is defined as

$$\bar{c} = a - ib. \quad (2.35)$$

Various computation rules for complex numbers are discussed in Exercise 2.12. Extending the notion of real-valued signals, a complex-valued CT-signal is a function $f: \mathbb{R} \rightarrow \mathbb{C}$ and a complex-valued DT-signal a function $x: \mathbb{Z} \rightarrow \mathbb{C}$. As is the case with complex numbers, each complex-valued signal can be considered as a pair of two real-valued signals. Furthermore, each real-valued signal can be regarded as a complex-valued signal simply by defining the imaginary part to be zero. In the following, we therefore only consider the more general complex-valued case, which includes the real-valued case.

2.2.3.2 Vector Spaces

A general principle in mathematics is to form suitable spaces that comprise all objects under consideration. These spaces can then be equipped with additional structures that can be used to manipulate and organize the objects. For example, for a

given natural number $N \in \mathbb{N}$, one may consider the space \mathbb{R}^N consisting of all real-valued N -tuples. This space can be equipped with an addition and a scalar multiplication such that \mathbb{R}^N becomes a **vector space** over \mathbb{R} . Similarly, one can define the space \mathbb{C}^N , which consists of all complex-valued N -tuples. In our case, the objects under consideration are complex-valued CT- and DT-signals. The resulting signal spaces are defined as

$$\mathbb{C}^{\mathbb{R}} := \{f|f: \mathbb{R} \rightarrow \mathbb{C}\} \quad \text{and} \quad \mathbb{C}^{\mathbb{Z}} := \{x|x: \mathbb{Z} \rightarrow \mathbb{C}\}, \quad (2.36)$$

for the CT and DT case, respectively. We have already seen in (2.30) and (2.31) how one can define an addition of two signals and a scalar multiplication of a real factor and a signal. These definitions directly carry over to the case of complex-valued signals using complex summation and multiplication, which makes $\mathbb{C}^{\mathbb{R}}$ a vector space over \mathbb{C} . Similarly, one can define addition and scalar multiplication in the DT case, making $\mathbb{C}^{\mathbb{Z}}$ a vector space over \mathbb{C} .

One may need to get used to the fact that elements (the “points”) of a space such as $\mathbb{C}^{\mathbb{R}}$ or $\mathbb{C}^{\mathbb{Z}}$ can be entire signals. As opposed to the case \mathbb{C}^N , which defines a vector space of (complex) dimension N , the vector spaces $\mathbb{C}^{\mathbb{R}}$ and $\mathbb{C}^{\mathbb{Z}}$ have infinite dimension. Still, many of the geometric structures known for the finite-dimensional space \mathbb{C}^N can be transferred to suitably defined infinite-dimensional subspaces of $\mathbb{C}^{\mathbb{R}}$ and $\mathbb{C}^{\mathbb{Z}}$. This is what we show next.

2.2.3.3 Inner Products

We start by reviewing some concepts from linear algebra. Usually, an element $x \in \mathbb{C}^N$ is thought of as a column vector of size N . The **transposed** vector, which we denote by x^\top , is then the corresponding row vector. The vector space \mathbb{C}^N can be equipped with an additional structure called an **inner product**. This additional structure associates to each pair of vectors a scalar quantity which is called the inner product of the two vectors. Mathematically, the inner product of \mathbb{C}^N is a mapping $\langle \cdot | \cdot \rangle: \mathbb{C}^N \times \mathbb{C}^N \rightarrow \mathbb{C}$ defined by

$$\langle x | y \rangle := \sum_{n=0}^{N-1} x(n) \overline{y(n)} \quad (2.37)$$

for $x = (x(0), x(1), \dots, x(N-1))^\top \in \mathbb{C}^N$ and $y = (y(0), y(1), \dots, y(N-1))^\top \in \mathbb{C}^N$. The inner product satisfies three mathematical properties, which are also used for an axiomatic definition of general inner products. First, it is **positive definite**; i.e., $\langle x | x \rangle \geq 0$ and $\langle x | x \rangle = 0$ if and only if x is the all-zero vector. Second, it is **conjugate symmetric**; i.e., $\langle x | y \rangle = \overline{\langle y | x \rangle}$. And third, it is **\mathbb{C} -linear** in the first argument; i.e., $\langle x_1 + x_2 | y \rangle = \langle x_1 | y \rangle + \langle x_2 | y \rangle$ and $\langle cx | y \rangle = c \langle x | y \rangle$ for any $x_1, x_2, x, y \in \mathbb{C}^N$ and $c \in \mathbb{C}$.

The importance of inner products is that they allow the introduction of intuitive geometrical notions such as the length of a vector, the angle between two vectors, and orthogonality between vectors (see Figure 2.15 for an illustration). More pre-

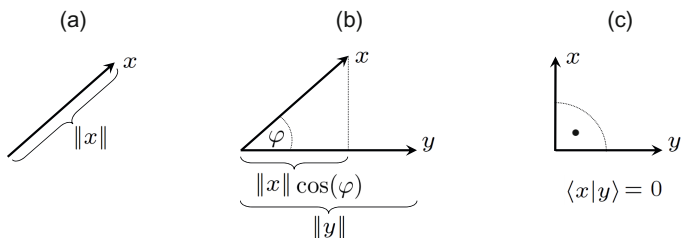


Fig. 2.15 Geometrical notions defined in terms of the inner product. **(a)** Length of a vector. **(b)** Angle between two vectors. **(c)** Orthogonality of two vectors.

cisely, the inner product induces a norm on \mathbb{C}^N via

$$\|x\| := \sqrt{\langle x|x \rangle}. \quad (2.38)$$

In general, a **norm** satisfies $\|x\| = 0$ if and only if $x = 0$, $\|ax\| = |a|\|x\|$ for any $a \in \mathbb{C}$ (positive scalability), and $\|x + y\| \leq \|x\| + \|y\|$ for any vectors x and y (**triangle inequality**). The positive number $\|x - y\|$ is also called the **distance** between the vectors x and y . The relation between the inner product and the angle φ between two vectors x and y is given by

$$\cos(\varphi) = \frac{|\langle x|y \rangle|}{\|x\| \cdot \|y\|}. \quad (2.39)$$

In other words, the angle φ is determined by the inner product: it is given by taking the inverse of the cosine of the absolute value of the inner product of the normalized vectors. The basis for this relation is the **Cauchy–Schwarz inequality**

$$|\langle x|y \rangle| \leq \|x\| \|y\|, \quad (2.40)$$

which is an indispensable mathematical tool for many estimations. Finally, two vectors $x, y \in \mathbb{C}^N$ are said to be **orthogonal** if $\langle x|y \rangle = 0$ (see Figure 2.15c). This concept can then be used to define orthogonal subspaces, orthogonal complements, projection operators, and so on.

2.2.3.4 The Space $\ell^2(\mathbb{Z})$

Given an arbitrary vector space, one can introduce the same geometric concepts once one has an inner product. It turns out that the signal spaces $\mathbb{C}^{\mathbb{R}}$ or $\mathbb{C}^{\mathbb{Z}}$ are too general. One strategy is to only consider signals with certain properties by passing over to suitable signal subspaces. We make this point clearer by first considering the space $\mathbb{C}^{\mathbb{Z}}$ of DT-signals. One idea for defining an inner product on this space is to simply extend the definition of (2.37) for \mathbb{C}^N . However, in contrast to \mathbb{C}^N , there may be an infinite number of nonzero summands in the case of $\mathbb{C}^{\mathbb{Z}}$, with the consequence

that the sum may be infinite. This leads to the following definitions: First, we define the **energy** $E(x)$ of a signal $x \in \mathbb{C}^{\mathbb{Z}}$ to be

$$E(x) := \sum_{n \in \mathbb{Z}} |x(n)|^2. \quad (2.41)$$

Then the space $\ell^2(\mathbb{Z}) \subset \mathbb{C}^{\mathbb{Z}}$ is defined to be the set of all signals having finite energy:

$$\ell^2(\mathbb{Z}) := \{x: \mathbb{Z} \rightarrow \mathbb{C} \mid E(x) < \infty\}. \quad (2.42)$$

In mathematical terms, $\ell^2(\mathbb{Z})$ is also referred to as the space of **square-summable** sequences. Obviously, there are many DT-signals that do not have finite energy. For example, the sampled sinusoid x given by $x(n) = \sin(\pi n/16)$ is not square-summable since it assumes the value 1 for infinitely many n . On the other hand, any DT-signal with a finite number of nonzero entries obviously has finite energy. The space \mathbb{C}^N for arbitrary $N \in \mathbb{N}$ can be regarded as a subspace of $\ell^2(\mathbb{Z})$ by extending a vector $x = (x(0), x(1), \dots, x(N-1))^T \in \mathbb{C}^N$ to a sequence by setting $x(n) = 0$ for all $n < 0$ and $n \geq N$. Furthermore, it is not hard to show that $\ell^2(\mathbb{Z})$ is a vector space (see Exercise 2.13). For the restricted space $\ell^2(\mathbb{Z}) \subset \mathbb{C}^{\mathbb{Z}}$, it is now possible to introduce an inner product that extends the one for \mathbb{C}^N . Indeed, one can show that

$$\langle x|y \rangle := \sum_{n \in \mathbb{Z}} x(n) \overline{y(n)} \quad (2.43)$$

is finite and hence well defined for any two signals $x, y \in \ell^2(\mathbb{Z})$ (see again Exercise 2.13). From this point on, everything works as in the finite-dimensional case \mathbb{C}^N . The inner product satisfies the Cauchy–Schwarz inequality (2.40), one can define an angle as in (2.39), one can talk about signals being orthogonal, and so on.

2.2.3.5 The Space $L^2(\mathbb{R})$

For the space $\mathbb{C}^{\mathbb{R}}$ of CT-signals, an inner product is defined in a similar fashion. However, technically, the definitions become more sophisticated in the continuous case, where summation becomes integration. In order to define an integral for a signal $f \in \mathbb{C}^{\mathbb{R}}$, it needs to fulfill certain integrability conditions, which in turn depend on the notion of integration to be used. For example, the notion of the well-known **Riemann integral** turns out to be too weak for many mathematical constructions. The technical deficiencies in Riemann integration can be remedied with the **Lebesgue integral**, which can be defined for a class of signals called **measurable**. At this point, since we may assume that basically all signals that we encounter are measurable, we do not want to go further into this issue. Similarly to the case of DT-signals, the **energy** $E(f)$ of a measurable signal $f \in \mathbb{C}^{\mathbb{R}}$ is defined by

$$E(f) := \int_{t \in \mathbb{R}} |f(t)|^2 dt. \quad (2.44)$$

Furthermore, the space $L^2(\mathbb{R}) \subset \mathbb{C}^{\mathbb{R}}$ is defined to be the set of all signals of finite energy:

$$L^2(\mathbb{R}) := \{f: \mathbb{R} \rightarrow \mathbb{C} \mid f \text{ measurable and } E(f) < \infty\}. \quad (2.45)$$

In mathematical terms, $L^2(\mathbb{R})$ is also referred to as the Lebesgue space³ of square-integrable functions. Again, there are many CT-signals that do not have finite energy. For example, any nonzero sinusoid has infinite energy. As with the DT case, it is not hard to show that $L^2(\mathbb{R})$ is a vector space. In the CT case, the inner product is defined by

$$\langle f|g \rangle := \int_{t \in \mathbb{R}} f(t) \overline{g(t)} dt \quad (2.46)$$

for any $f, g \in L^2(\mathbb{R})$. Again this makes it possible to introduce the geometric concepts known from linear algebra.

2.2.3.6 The Space $L^2([0, 1])$

Finally, we want to consider another class of CT-signals of fundamental importance: the class of **periodic signals**. As already mentioned above, nonzero periodic functions⁴ are not contained in $L^2(\mathbb{R})$. However, also for periodic functions one can define a suitable signal subspace of $\mathbb{C}^{\mathbb{R}}$ that possesses an inner product. Recall from Section 2.2.1 that a signal $f: \mathbb{R} \rightarrow \mathbb{C}$ is periodic with period $\lambda \in \mathbb{R}_{>0}$ if $f(t) = f(t + \lambda)$ holds for all $t \in \mathbb{R}$. A λ -periodic signal f can be transformed into a 1-periodic signal $t \mapsto f(\lambda \cdot t)$ by applying the linear transform $t \mapsto \lambda \cdot t$. Hence, in the following discussion, we only consider the case $\lambda = 1$. Obviously, any 1-periodic function f is already known when restricted to the interval $[0, 1)$. In contrast, any function $g: [0, 1) \rightarrow \mathbb{C}$ can be extended in an obvious fashion to a 1-periodic function $f: \mathbb{R} \rightarrow \mathbb{C}$. In other words, there is a one-to-one correspondence between the 1-periodic functions in $\mathbb{C}^{\mathbb{R}}$ and the signal space $\mathbb{C}^{[0,1)} := \{f: [0, 1) \rightarrow \mathbb{C}\}$. Similar to the nonperiodic case, one can define the **energy** $E_{[0,1)}(f)$ by

$$E_{[0,1)}(f) := \int_{t \in [0,1)} |f(t)|^2 dt \quad (2.47)$$

and the space $L^2([0, 1)) \subset \mathbb{C}^{[0,1)}$ by

$$L^2([0, 1)) := \{f: [0, 1) \rightarrow \mathbb{C} \mid f \text{ measurable and } E_{[0,1)}(f) < \infty\}. \quad (2.48)$$

Furthermore, one can show that the inner product

$$\langle f|g \rangle := \int_{t \in [0,1)} f(t) \overline{g(t)} dt \quad (2.49)$$

³ From a strict technical point of view, $L^2(\mathbb{R})$ is defined as a quotient space, where all functions that are zero almost everywhere are identified.

⁴ Strictly speaking, we mean here periodic functions that are not zero almost everywhere.

is well defined for any $f, g \in L^2([0, 1))$. Generalizing these definitions, one can introduce a space $L^2([a, b))$ with an inner product for any $a, b \in \mathbb{R}$, $a < b$, which consists of λ -periodic signals with $\lambda = b - a$.

2.2.3.7 Hilbert Spaces

In summary, we have introduced the signal spaces $\ell^2(\mathbb{Z})$, $L^2(\mathbb{R})$, and $L^2([0, 1))$, which all possess an inner product similar to the one of the finite-dimensional vector space \mathbb{C}^N . All of these spaces are special cases of what is known as **Hilbert space**. By definition, a Hilbert space is a vector space \mathcal{H} equipped with an inner product $\langle \cdot | \cdot \rangle: \mathcal{H} \times \mathcal{H} \rightarrow \mathbb{C}$ satisfying the three axiomatic conditions mentioned in Section 2.2.3. Furthermore, one requires that a Hilbert space is **complete** in the sense that every Cauchy sequence⁵ in \mathcal{H} converges in \mathcal{H} . Intuitively, a space is complete if no points are missing from it. For example, the set of rational numbers is not complete, because there are numbers such as $\sqrt{2}$ missing from it, even though one can construct Cauchy sequences of rational numbers that converge to such irrational numbers. As one can show, this nontrivial completeness condition is satisfied for the signal spaces $\ell^2(\mathbb{R})$, $L^2(\mathbb{R})$, and $L^2([0, 1))$. As we will see in the next sections, the geometric concepts provided by the inner product help to develop our intuition and to simplify the formulation of the Fourier transform.

A particularly important concept that generalizes from the finite-dimensional space \mathbb{C}^N to arbitrary Hilbert spaces is the existence of orthonormal bases. Let I be a countable set, then a subset $(x_i)_{i \in I}$ of \mathcal{H} is called an **orthonormal basis** (ON-basis) if the following three conditions hold:

$$\langle x_i | x_j \rangle = 0 \quad \text{for} \quad i, j \in I, i \neq j, \quad (2.50)$$

$$\|x_i\|^2 = 1 \quad \text{for} \quad i \in I, \quad (2.51)$$

$$x = \sum_{i \in I} \langle x | x_i \rangle x_i \quad \text{for} \quad x \in X. \quad (2.52)$$

The first condition means that any two distinct elements x_i and x_j are orthogonal, and the second one that each of the elements x_i has unit energy. The third condition, also referred to as the **completeness condition**, requires that any element of $x \in \mathcal{H}$ can be represented as a weighted superposition of the **basis vectors** x_i , $i \in I$. Intuitively, completeness means that everything in \mathcal{H} can be captured by the basis vectors. Furthermore, the weights are given by the inner products $\langle x | x_i \rangle$. One can show that for a Hilbert space there always exists an ON-basis and, in general, even a very large number of different ON-bases. As we will see, the Fourier transforms for DT-signals and periodic CT-signals are based on very specific choices of such ON-bases.

⁵ A Cauchy sequence is a sequence whose elements become arbitrarily close to each other as the sequence progresses. More precisely, given any small positive distance, all but a finite number of elements of the sequence are less than that given distance from each other.

2.3 Fourier Transform

The Fourier transform is the most important mathematical tool in audio signal processing. As discussed in Section 2.1, the Fourier transform converts a time-dependent signal into a frequency-dependent function. The inverse process is realized by the Fourier representation, which represents a signal as a weighted superposition of independent elementary functions. Each of the weights expresses the extent to which the corresponding elementary function contributes to the original signal, thus revealing a certain aspect of the signal. Because of their explicit physical interpretation in terms of frequency, sinusoids are particularly suited to serve as elementary functions. Each of the weights is then associated to a frequency value and expresses the degree to which the signal contains a periodic oscillation of that frequency. The Fourier transform can be regarded as a way to compute the frequency-dependent weights.

In the following, depending on the underlying signal space, we introduce several variants of the Fourier transform and its inverse, the Fourier representation. We start with the signal space $L^2([0, 1])$ consisting of 1-periodic finite-energy CT-signals (Section 2.3.1). We continue by showing how the formulation of the Fourier transform in terms of complex-valued exponential functions (instead of real-valued sinusoids) makes the mathematical handling much more convenient (Section 2.3.2). We then discuss the Fourier transform for the signal space $L^2(\mathbb{R})$ (Section 2.3.3) as well as for the signal space $\ell^2(\mathbb{Z})$ (Section 2.3.4). It is important to note that each of these signal spaces possesses its own Fourier transform and the mathematical concepts needed to prove the existence and properties of the respective Fourier transform are different for the variants. While giving mathematically rigorous definitions of the various Fourier transforms, we do not provide the proofs. In particular for the analog case, the proofs require results from measure and integration theory, which are outside the scope of this book. Instead, we will try to give some intuitive explanations while highlighting the meaning and the interrelations of the various variants.

2.3.1 Fourier Transform for Periodic CT-Signals

We start our discussion by considering the case of all **real-valued** signals in $L^2([0, 1])$. Let us denote this subspace by $L^2_{\mathbb{R}}([0, 1]) \subset L^2([0, 1])$. Note that any constant as well as any $(1/k)$ -periodic function for an integer $k \in \mathbb{N}$ is 1-periodic too. The sinusoid $t \mapsto \sqrt{2}\cos(2\pi kt)$ may be regarded as the archetype of a $(1/k)$ -periodic function, which represents a pure tone of k Hz. The factor $\sqrt{2}$ is introduced to normalize the sinusoid to have unit energy or, equivalently, to have norm one (see Exercise 2.14). Of course, also the sinusoid $t \mapsto \sqrt{2}\sin(2\pi kt)$ or all phase-shifted versions $t \mapsto \sqrt{2}\cos(2\pi(kt - \varphi))$ have the same interpretation. One important theorem in Fourier analysis is that any real-valued signal $f \in L^2_{\mathbb{R}}([0, 1])$ can be written as a superposition

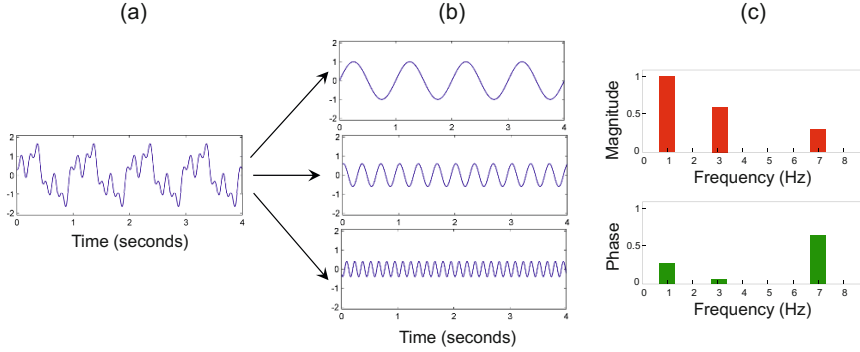


Fig. 2.16 (a) Analog 1-periodic signal. (b) Decomposition of the signal into three sinusoids. (c) Magnitude and phase coefficients of the Fourier transform.

$$f(t) = d_0 + \sum_{k \in \mathbb{N}} d_k \sqrt{2} \cos(2\pi(kt - \varphi_k)) \quad (2.53)$$

of 1-periodic sinusoids with suitable amplitudes $d_k \in \mathbb{R}_{\geq 0}$ and phases $\varphi_k \in [0, 1)$. The superposition exhibits the frequency content of f as follows: the coefficient d_k , also referred to as the **magnitude**, reflects the contribution of the sinusoid of k Hz, whereas the coefficient φ_k , also referred to as the **phase**, shows how the sinusoid has to be shifted to best “explain” or “match” the original signal. Note that the phase coefficients are determined only up to an integer and can therefore be assumed to lie in the interval $[0, 1)$. Figure 2.16 shows an example of a 1-periodic signal and the resulting magnitude and phase coefficients. The superposition in (2.53) is the **Fourier representation** of the signal f , whereas the magnitude and phase coefficients are called the **Fourier coefficients**.

In our first reformulation, we exploit the fact that any sinusoid with arbitrary phase can be represented as a weighted sum of two specific sinusoids of the same frequency having fixed phases. Indeed, using the trigonometric identity $\cos(\alpha - \beta) = \cos(\alpha)\cos(\beta) + \sin(\alpha)\sin(\beta)$ for arbitrary angles α and β , one obtains

$$\cos(2\pi(kt - \varphi)) = \cos(2\pi kt) \cos(2\pi \varphi) + \sin(2\pi kt) \sin(2\pi \varphi) \quad (2.54)$$

when setting $\alpha = 2\pi kt$ and $\beta = 2\pi \varphi$. Let $\mathbf{cos}_k, \mathbf{sin}_k \in L^2_{\mathbb{R}}([0, 1))$ be the two specific sinusoids defined by

$$\mathbf{cos}_k(t) := \sqrt{2} \cos(2\pi kt), \quad (2.55)$$

$$\mathbf{sin}_k(t) := \sqrt{2} \sin(2\pi kt), \quad (2.56)$$

for $k \in \mathbb{N}$. Then plugging (2.54) into (2.53), one obtains the following Fourier representation, which is also known as the **Fourier series**:

$$f(t) = a_0 + \sum_{k \in \mathbb{N}} a_k \mathbf{cos}_k(t) + \sum_{k \in \mathbb{N}} b_k \mathbf{sin}_k(t). \quad (2.57)$$

It readily follows that the **Fourier coefficients** a_0 , a_k , and b_k are given by

$$a_0 = d_0, \quad (2.58)$$

$$a_k = \cos(2\pi\varphi_k)d_k, \quad (2.59)$$

$$b_k = \sin(2\pi\varphi_k)d_k \quad (2.60)$$

for $k \in \mathbb{N}$. Vice versa, the magnitudes and phases can be computed from the a_k and b_k via

$$d_k = \sqrt{a_k^2 + b_k^2}, \quad (2.61)$$

$$\varphi_k = \frac{1}{2\pi} \text{atan2}(b_k, a_k). \quad (2.62)$$

The atan2 function, which is a variant of the inverse of the tangent function, will be explained in Section 2.3.2.2. A nice property of the Fourier representation in (2.57) is that its Fourier coefficients can be easily computed using Hilbert space theory. To this end, one needs to show that the set

$$\{\mathbf{1}, \mathbf{cos}_k, \mathbf{sin}_k | k \in \mathbb{N}\}, \quad (2.63)$$

is an ON-basis of the Hilbert space $L^2_{\mathbb{R}}([0, 1))$, where $\mathbf{1}$ denotes the all-one signal (i.e., $\mathbf{1}(t) = 1$ for $t \in [0, 1)$). The two conditions specified in (2.50) and (2.51) follow from trigonometric identities (see Exercise 2.14). Only the completeness condition specified in (2.52) is harder to show and requires some more involved mathematical tools that are outside the scope of this book. From (2.52), one not only recovers the Fourier series in (2.57), but also a formula for how to compute the Fourier coefficients as inner products of the signal f with the basis functions of the ON-basis:

$$a_0 = \langle f | \mathbf{1} \rangle = \int_{t \in [0, 1)} f(t) dt, \quad (2.64)$$

$$a_k = \langle f | \mathbf{cos}_k \rangle = \sqrt{2} \int_{t \in [0, 1)} f(t) \cos(2\pi kt) dt, \quad (2.65)$$

$$b_k = \langle f | \mathbf{sin}_k \rangle = \sqrt{2} \int_{t \in [0, 1)} f(t) \sin(2\pi kt) dt. \quad (2.66)$$

2.3.2 Complex Formulation of the Fourier Transform

As often in mathematics, the transfer of a problem from the real into the complex world can lead to significant simplifications. A famous example is the problem of finding solutions of polynomial equations. The equation $z^2 - 1 = 0$ has the two solutions $z = +1$ and $z = -1$, however the equation $z^2 + 1 = 0$ does not have any solution when only considering real numbers. Extending \mathbb{R} to \mathbb{C} , however, one also finds for the second equation two solutions given by $z = +i$ and $z = -i$, where

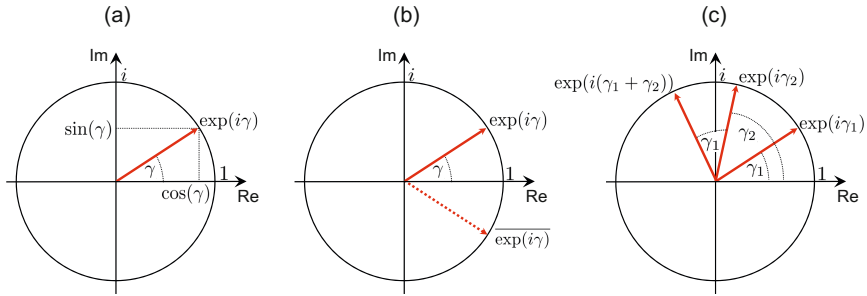


Fig. 2.17 Illustration of the complex exponential function.

i denotes the complex unit. Considering polynomial equations over \mathbb{C} makes the problem much easier to understand. In general, an extension of the real numbers to the complex numbers not only gives a broader view but also provides additional tools and structures. For example, the complex multiplication as defined by (2.34), which extends the usual multiplication of real numbers, yields such a powerful tool. Also, the trigonometric identities are considerably simplified when using a complex formulation.

2.3.2.1 Exponential Function

Converting the Fourier transform from the real into the complex domain has several advantages. First, the concept of Fourier series can be naturally generalized from real-valued to complex-valued signals. Second, one obtains compact and elegant formulas, where the magnitude and phase are naturally expressed by a single complex Fourier coefficient. Recall from Section 2.1.1.4 that the **exponential function** combines the two real-valued sinusoids given by the cosine and sine into a single complex-valued function:

$$\exp(i\gamma) = \cos(\gamma) + i\sin(\gamma). \quad (2.67)$$

This equation, which can be used as a defining relation, is also known as **Euler's formula**. However, there are many other ways in which the exponential function may be characterized, e.g., in terms of a power series expansion or by means of a differential equation. The exponential function has some important properties, which are also illustrated by Figure 2.17:

$$\exp(i\gamma) = \exp(i(\gamma + 2\pi)), \quad (2.68)$$

$$|\exp(i\gamma)| = 1, \quad (2.69)$$

$$\overline{\exp(i\gamma)} = \exp(-i\gamma), \quad (2.70)$$

$$\exp(i(\gamma_1 + \gamma_2)) = \exp(i\gamma_1)\exp(i\gamma_2) \quad (2.71)$$

for $\gamma, \gamma_1, \gamma_2 \in \mathbb{R}$. For a proof of these properties, we refer to Exercise 2.15. The property (2.68) means that the exponential function is 2π -periodic. The property (2.69) implies that all values of this function live on the unit circle of \mathbb{C} . By successively increasing the angle γ starting with $\gamma = 0$ and ending with $\gamma = 2\pi$, one travels exactly once along the unit circle in a counterclockwise fashion. The property (2.70) shows that complex conjugation results in changing the direction of this travel. Finally, the property (2.71) is the complex formulation of the real-valued trigonometric identities that hold for the cosine and sine functions (see also Exercise 2.15).

2.3.2.2 Polar Coordinates

A complex number $c = a + ib \in \mathbb{C}$ is specified by its Cartesian coordinates $(a, b) \in \mathbb{R}^2$ in the two-dimensional plane. The complex exponential function makes it possible to represent a complex number in the form of **polar coordinates**, which we discussed in Section 2.1.1.4. In the polar coordinate system, the point $c = a + ib$ is determined by the distance $|c|$ from the origin and the angle γ (in radians) between the positive horizontal axis and the point given by the coordinates (a, b) (see Figure 2.4). Repeating the formulas from (2.6) and (2.7), we obtain the following relations between Cartesian and polar coordinates:

$$|c| = \sqrt{a^2 + b^2}, \quad (2.72)$$

$$\gamma = \text{atan2}(b, a), \quad (2.73)$$

$$a = |c| \operatorname{Re}(\exp(i\gamma)) = |c| \cos(\gamma), \quad (2.74)$$

$$b = |c| \operatorname{Im}(\exp(i\gamma)) = |c| \sin(\gamma). \quad (2.75)$$

The atan2 function is a generalization of the arctangent function (denoted as \arctan), which is the inverse of the principal branch of the tangent function (see Figure 2.18b). The \arctan function requires a real-valued argument $v \in \mathbb{R}$ and computes an angle $\arctan(v) \in (-\pi/2, \pi/2)$ (given in radians), which is called the principal value. As opposed to the \arctan function, the atan2 function has two real-valued arguments. This makes it possible to capture the quadrant of the computed angle, which is not possible for the single-argument \arctan function. In terms of the standard \arctan function, the atan2 function is given by

$$\text{atan2}(b, a) := \begin{cases} \arctan(b/a), & a > 0, \\ \arctan(b/a) + \pi, & b \geq 0, a < 0, \\ \arctan(b/a) - \pi, & b < 0, a < 0, \\ +\pi/2, & b > 0, a = 0, \\ -\pi/2, & b < 0, a = 0, \\ \text{undefined} & b = 0, a = 0 \end{cases} \quad (2.76)$$

for $(a, b) \in \mathbb{R}^2$ (see Figure 2.18c). The angle computed by the atan2 function is positive for complex numbers $c = a + ib$ with positive imaginary part $b > 0$ (upper half-plane) and negative for those with negative imaginary part $b < 0$ (lower half-plane).

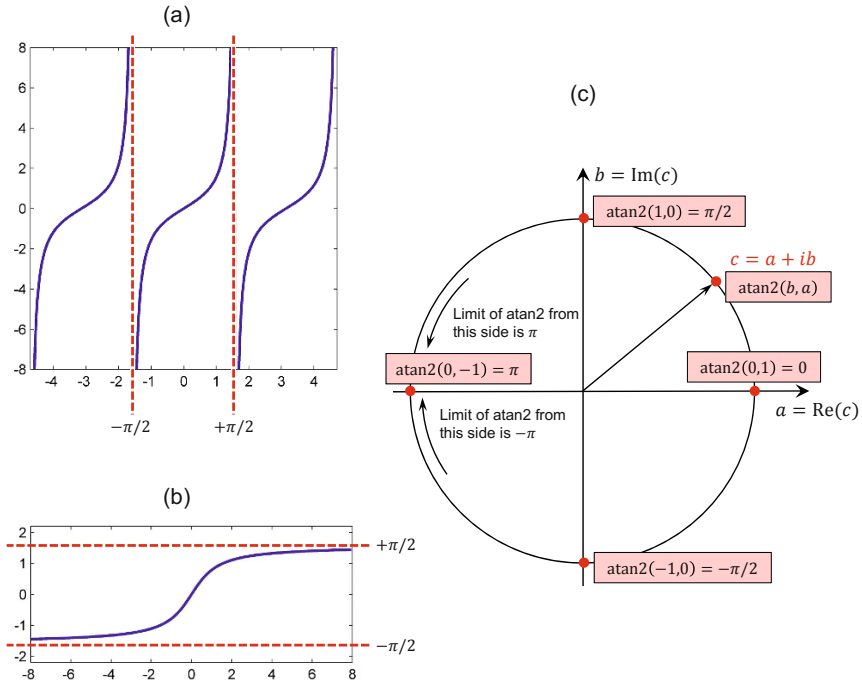


Fig. 2.18 (a) Tangent function with different branches. (b) Arctangent function inverting the principal branch of the tangent function. (c) Illustration of the values assumed by the atan2 function.

The range $(-\pi, \pi]$ of angles can be mapped to $[0, 2\pi)$ by adding 2π to negative values. Further properties of the atan2 function are discussed in Exercise 2.17.

2.3.2.3 Complex Fourier Series

We are now ready for the complex formulation of the Fourier series. To this end, we replace in (2.57) the real-valued sinusoids \cos_k and \sin_k defined for $k \in \mathbb{N}$ by the complex-valued exponential functions $\exp_k : [0, 1) \rightarrow \mathbb{C}$ defined by

$$\exp_k(t) := \exp(2\pi i k t). \quad (2.77)$$

Obviously, \exp_k is a $(1/k)$ -periodic signal for $k \neq 0$ and \exp_k is the all-one signal $\mathbf{1}$ for $k = 0$. Furthermore, as in (2.63), it can be shown that the set

$$\{\exp_k \mid k \in \mathbb{Z}\} \quad (2.78)$$

is an ON-basis of the (complex) Hilbert space $L^2[0, 1)$. The properties $\|\exp_k\| = 1$ for $k \in \mathbb{Z}$ and $\langle \exp_k | \exp_\ell \rangle = 0$ for $k \neq \ell$, $k, \ell \in \mathbb{Z}$, are shown in Exercise 2.16. Again,

as in the real-valued case, the completeness property is more difficult to prove and is not discussed in this book. The resulting expansion of a signal $f \in L^2([0, 1])$ with respect to this ON-basis leads to the equality⁶

$$f(t) = \sum_{k \in \mathbb{Z}} c_k \mathbf{exp}_k(t) = \sum_{k \in \mathbb{Z}} c_k \exp(2\pi i k t), \quad (2.79)$$

which is also referred to as the (complex) **Fourier series**. The corresponding (complex) **Fourier coefficients** $c_k \in \mathbb{C}$ are given by

$$c_k = \langle f | \mathbf{exp}_k \rangle = \int_{t \in [0, 1]} f(t) \overline{\mathbf{exp}_k(t)} dt = \int_{t \in [0, 1]} f(t) \exp(-2\pi i k t) dt, \quad (2.80)$$

where we used (2.70) in the last equation. As in (2.11), the function

$$\hat{f} : \mathbb{Z} \rightarrow \mathbb{C}, \quad \hat{f}(k) := c_k \quad (2.81)$$

is called the **Fourier transform** of $f \in L^2([0, 1])$. Note that, in this case, a 1-periodic *continuous-time* signal f is mapped to a *discrete-time* signal \hat{f} . Furthermore, one can show that the Fourier transform is **energy preserving** in the sense that the energy of \hat{f} is the same as the energy of f :

$$\|f\|_{L^2([0, 1])} = \|\hat{f}\|_{\ell^2(\mathbb{Z})}. \quad (2.82)$$

At this point, using the signal spaces as subscripts of the norms, we want to emphasize that the energy of \hat{f} is measured in the space $\ell^2(\mathbb{Z})$ and the energy of f is measured in $L^2([0, 1])$. Mathematically, such an energy-preserving map between Hilbert spaces is also called an **isometry**. As a consequence, the inverse mapping $\hat{f} \mapsto f$ given by the Fourier series (2.79) is again an isometry. We will see that the Fourier transforms for the other finite-energy signal spaces have similar properties.

2.3.2.4 Relation Between Complex and Real Fourier Series

Note that the complex Fourier series can be used to represent **complex-valued** signals, thus extending the Fourier series of (2.57) for **real-valued** signals. Being a special case of a complex-valued function, a real-valued signal $f \in L^2_{\mathbb{R}}([0, 1]) \subset L^2([0, 1])$ can also be represented using a complex Fourier series. In this case, each signal value $f(t)$ coincides with its complex conjugate $\overline{f(t)}$. Using the computation rules for complex numbers (see Exercise 2.12) and (2.70), one obtains

$$\sum_{k \in \mathbb{Z}} c_k \mathbf{exp}_k(t) = f(t) = \overline{f(t)} = \overline{\sum_{k \in \mathbb{Z}} c_k \mathbf{exp}_k(t)} = \sum_{k \in \mathbb{Z}} \overline{c_k} \mathbf{exp}_{-k}(t). \quad (2.83)$$

⁶ Strictly speaking, this equality only holds for almost all $t \in [0, 1]$. In the following, even though a bit sloppy in a strict mathematical sense, we do not further mention such issues.

This implies $c_{-k} = \overline{c_k}$ for $k \in \mathbb{Z}$. In other words, for real-valued signals, the coefficients with negative indices are redundant. Furthermore, the complex coefficients c_k of a real-valued signal relate to the real coefficients a_k and b_k of the Fourier series in (2.57) in the following way:

$$a_0 = c_0, \quad (2.84)$$

$$a_k = \sqrt{2} \operatorname{Re}(c_k), \quad (2.85)$$

$$b_k = -\sqrt{2} \operatorname{Im}(c_k) \quad (2.86)$$

for $k \in \mathbb{N}$. To see this, one needs to use $c_{-k} = \overline{c_k}$ and the definitions (2.77) of \mathbf{exp}_k , (2.55) of \mathbf{cos}_k , and (2.56) of \mathbf{sin}_k . Since the proof is an instructive example of how to compute with complex numbers, we conduct the calculation in detail:

$$\begin{aligned} f(t) &= \sum_{k \in \mathbb{Z}} c_k \mathbf{exp}_k(t) \\ &= c_0 + \sum_{k=1}^{\infty} c_k \mathbf{exp}_k(t) + \sum_{k=1}^{\infty} c_{-k} \mathbf{exp}_{-k}(t) \\ &= c_0 + \sum_{k=1}^{\infty} (c_k \mathbf{exp}_k(t) + \overline{c_k \mathbf{exp}_k(t)}) \quad (2.87) \\ &= c_0 + \sum_{k=1}^{\infty} 2 \operatorname{Re}(c_k \mathbf{exp}_k(t)) \\ &= c_0 + \sum_{k=1}^{\infty} (2 \operatorname{Re}(c_k) \cos(2\pi kt) - 2 \operatorname{Im}(c_k) \sin(2\pi kt)) \\ &= c_0 + \sum_{k=1}^{\infty} \sqrt{2} \operatorname{Re}(c_k) \mathbf{cos}_k(t) + \sum_{k=1}^{\infty} (-\sqrt{2} \operatorname{Im}(c_k)) \mathbf{sin}_k(t). \end{aligned}$$

Comparing coefficients with (2.57) yields the assertion.

Finally, let us come back to our first version of the Fourier series in (2.53), where we introduced the magnitude coefficients d_k and phase coefficients φ_k . How are these coefficients related to the complex Fourier coefficients c_k in the case of real-valued signals? This question can be easily answered when using (2.61) and (2.62) in combination with the polar coordinate representation $c_k = |c_k| \exp(i\gamma_k)$ and the above identities:

$$d_k = \sqrt{a_k^2 + b_k^2} = \sqrt{2 \operatorname{Re}(c_k)^2 + 2 \operatorname{Im}(c_k)^2} = \sqrt{2} |c_k|, \quad (2.88)$$

$$\begin{aligned} \varphi_k &= \frac{1}{2\pi} \operatorname{atan2}(b_k, a_k) = \frac{1}{2\pi} \operatorname{atan2}(-\sqrt{2} \operatorname{Im}(c_k), \sqrt{2} \operatorname{Re}(c_k)) \\ &= \frac{1}{2\pi} \operatorname{atan2}(-\operatorname{Im}(c_k), \operatorname{Re}(c_k)) = -\frac{\gamma_k}{2\pi}. \end{aligned} \quad (2.89)$$

In the last equations, we used the fact that $\operatorname{atan2}$ is invariant under scaling with a nonzero constant and assumes the negative angle for the conjugate of a complex number (see Exercise 2.17). These identities correspond to (2.14) and (2.15).

2.3.3 Fourier Transform for CT-Signals

The general idea of the Fourier transform carries over from the case of periodic to the case of nonperiodic signals in $L^2(\mathbb{R})$. In the nonperiodic case, however, the exponential functions \mathbf{exp}_k of integer frequency $k \in \mathbb{Z}$ do not suffice to “describe” a signal. Instead, one needs exponential functions

$$\mathbf{exp}_\omega : \mathbb{R} \rightarrow \mathbb{C}, \quad \mathbf{exp}_\omega(t) := \exp(2\pi i \omega t) \quad (2.90)$$

for all frequencies $\omega \in \mathbb{R}$. Then, replacing summation by integration one obtains the following nonperiodic analog of the Fourier representation:

$$f(t) = \int_{\omega \in \mathbb{R}} c_\omega \mathbf{exp}_\omega(t) d\omega = \int_{\omega \in \mathbb{R}} c_\omega \exp(2\pi i \omega t) d\omega \quad (2.91)$$

for $t \in \mathbb{R}$. The coefficients c_ω have the same interpretation as the Fourier coefficients c_k . The frequency-dependent function $\hat{f} : \mathbb{R} \rightarrow \mathbb{C}$ defined by

$$\hat{f}(\omega) := c_\omega = \int_{t \in \mathbb{R}} f(t) \overline{\mathbf{exp}_\omega(t)} dt = \int_{t \in \mathbb{R}} f(t) \exp(-2\pi i \omega t) dt \quad (2.92)$$

is called the **Fourier transform** of f . Again, it can be shown that the Fourier transform is energy preserving. In other words, if $f \in L^2(\mathbb{R})$, then $\hat{f} \in L^2(\mathbb{R})$ and $\|f\|_{L^2(\mathbb{R})} = \|\hat{f}\|_{L^2(\mathbb{R})}$.

Strictly speaking, there are some mathematical issues that need to be considered for the nonperiodic case. Recall that, in the periodic case, the elementary functions \mathbf{exp}_k have finite energy over the interval $[0, 1)$ and are therefore elements of $L^2([0, 1))$. This is the reason why the Fourier transform and the Fourier representation can be expressed by means of inner products. Unfortunately, this is no longer the case for the nonperiodic case, since the elementary functions \mathbf{exp}_ω do not have finite energy over the real time axis \mathbb{R} and are therefore *not* elements in the space $L^2(\mathbb{R})$. As a consequence, the inner product is not defined between a signal $f \in L^2(\mathbb{R})$ and \mathbf{exp}_ω . Furthermore, the integrals in (2.91) and (2.92) need to be defined as limits over increasing finite integration domains. For example,

$$\hat{f}(\omega) := \lim_{N \rightarrow \infty} \int_{t \in [-N, N]} f(t) \exp(-2\pi i \omega t) dt. \quad (2.93)$$

Similarly, one has to define the Fourier representation. However, these technical issues will not play any further role in this book. Furthermore, most of the signals we consider in this book have **compact support**; i.e., they are zero outside an interval of finite length. For such signals, no problems occur in the integrals even from a strict mathematical point of view.

The Fourier representation in (2.91) yields a quite surprising result. It states that every **nonperiodic** function of finite energy can be represented as a weighted (infinitesimal) superposition of **periodic** elementary frequency functions \mathbf{exp}_ω that continue out to infinity without decaying. For example, even noise-like short-

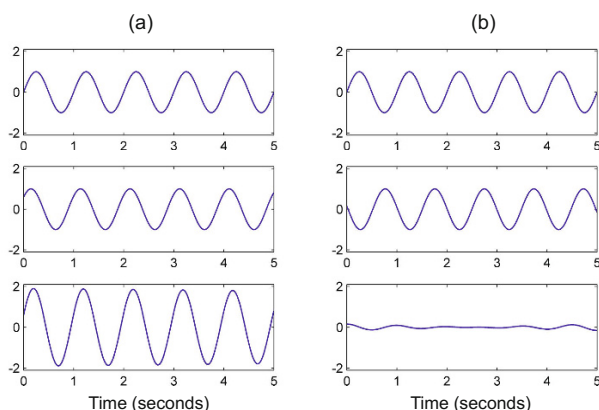


Fig. 2.19 Interference of two sinusoids of similar frequency. **(a)** Constructive interference. **(b)** Destructive interference.

duration sounds such as transients, which often occur in the attack phase of a tone, can be represented by ceaselessly oscillating sinusoids.

2.3.3.1 Interference

In Section 2.1.2, we have already discussed some real as well as synthetic signals to illustrate important properties of the Fourier transform. In the following, we take a closer look at some of the encountered phenomena. Let us start with the example from Figure 2.6b. Besides the two peaks, we could observe in the magnitude Fourier transform $|\hat{f}|$ a number of “ripples” of decreasing amplitude. Where do these ripples come from? In the figure, the analog signal f is shown only for the time interval $[0, 5]$ and is (implicitly) assumed to be zero outside this compact interval. The ripples in the spectrum come from a phenomenon known as destructive interference, where many different frequency components are involved for generating the compact support of f .

In general, **interference** occurs when a wave is superimposed with another wave of similar frequency. When a crest of one wave meets a crest of the other wave at some point, then the individual magnitudes add up for a certain period of time, which is known as **constructive interference** (see Figure 2.19a). Vice versa, when a crest of one wave meets a trough of the other wave, then the magnitudes cancel out for a certain period of time, which is known as **destructive interference** (see Figure 2.19b).

Coming back to Figure 2.6b, one needs the sinusoids of frequency $\omega = 1$ Hz and $\omega = 5$ Hz to generate the main components of the signal f within the interval $[0, 5]$. Note that these two sinusoids also oscillate outside the visualized interval $[0, 5]$, where the signal is assumed to be zero. Therefore, to cancel out these oscillations

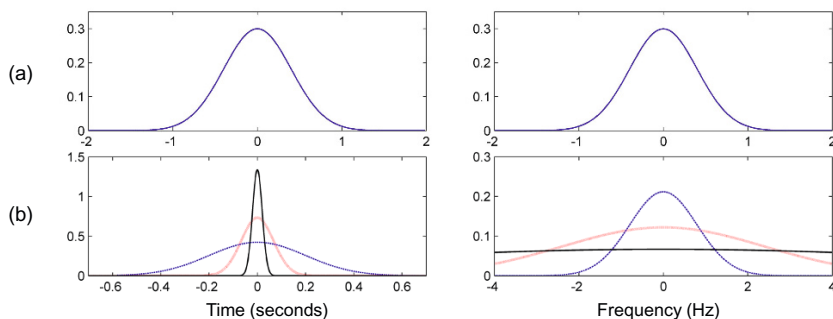


Fig. 2.20 (a) Gaussian function (left) and its Fourier transform (right). (b) Dirac sequence (left) with corresponding Fourier transforms (right).

outside $[0, 5]$ by destructive interference, one needs to add many more sinusoids of different frequencies and weights. These additional sinusoidal components are reflected by the ripples. Interference effects are further discussed in Exercise 2.19 and in the subsequent examples.

2.3.3.2 Fourier Transform for Impulses

The synthetic signals shown in Figure 2.20 illustrate further properties of the Fourier transform. First of all, the **Gaussian function** defined by the formula

$$f(t) = (2\pi)^{-\frac{1}{2}} \pi^{-\frac{1}{4}} \exp(-\pi t^2) \quad (2.94)$$

has the remarkable property that it coincides with its Fourier transform (see Figure 2.20a). In particular, its Fourier transform is real-valued and positive. Therefore, it agrees with its magnitude Fourier transform. The Fourier representation (2.91) tells us that the Gaussian function is obtained as an (infinitesimal) weighted superposition of periodic sine waves, where the weights are again given by the Gaussian function. The next question we consider is how the Fourier transform behaves, if we start to make the Gaussian function somewhat narrower (see Figure 2.20b). This leads to the notion of a **Dirac sequence**, which is a sequence of functions $(f_n)_{n \in \mathbb{N}}$ of norm $\|f_n\| = 1$ such that for increasing n the functions f_n “concentrate” more and more around the point $t = 0$. The limit of this sequence is the **Dirac delta function** or **impulse function** (often denoted by the symbol δ), which can be thought of as a function that is zero everywhere except for $t = 0$. At $t = 0$, it has an infinitely narrow spike of infinite height, which integrates to a value of one. Strictly speaking, this impulse is not a function, but a so-called **distribution**. As illustrated by Figure 2.20b, the magnitude Fourier transform of a Dirac sequence becomes broader and broader. This scaling property of the Fourier transform is shown in Exercise 2.20. In the limit case, the Fourier transform approaches a con-

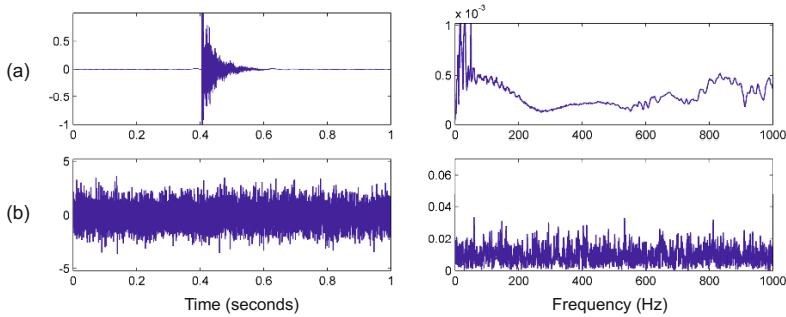


Fig. 2.21 Waveform and its magnitude Fourier transform for (a) a clapping sound and (b) white noise.

stant function, where the magnitudes of all frequency components have the same, yet infinitesimally small value.

The interpretation of this property is important in view of practical applications. It says that impulse-like sounds such as a drum hit or a transient as occurring in the attack phase of a musical tone (see Section 1.3.4) lead to a flat magnitude Fourier transform with many small, yet nonzero Fourier coefficients. In other words, for a sudden sharp sound, the signal's energy is spread across the entire spectrum of frequencies. This is also illustrated by Figure 2.21a, which shows the waveform and its magnitude Fourier transform for a real clapping sound. Another type of sound that results in an energy spread across the entire frequency spectrum are noise-like signals. Generally speaking, random signals such as white noise also remain random when transformed into the Fourier domain. For example, Figure 2.21b shows white Gaussian noise and its magnitude Fourier transform, which also looks like noise that is equally spread over the entire frequency range.

2.3.3.3 Translation and Modulation

As a final example, which is shown in Figure 2.22, we consider the **rectangular function**

$$f(t) := \begin{cases} 1, & \text{if } -0.5 \leq t \leq 0.5, \\ 0, & \text{otherwise.} \end{cases} \quad (2.95)$$

Its Fourier transform is the **sinc function**, which is defined by

$$\text{sinc}(t) := \begin{cases} \frac{\sin \pi t}{\pi t}, & \text{if } t \neq 0, \\ 1, & \text{if } t = 0. \end{cases} \quad (2.96)$$

For the proof of this fact, we refer to Exercise 2.21. The rectangular and the sinc function play an important role in the sampling theorem (see Exercise 2.28). In the case that the rectangular function is centered around $t = 0$, its Fourier transform is

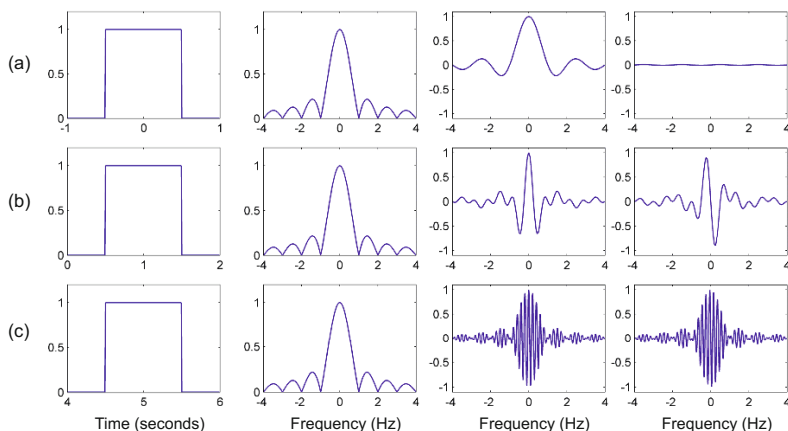


Fig. 2.22 Behavior of Fourier transform under translations. From left to right, the signal as well as the magnitude, real part, and imaginary part of the Fourier transform are shown. **(a)** Rectangular function. **(b)** Translation by one second. **(c)** Translation by five seconds.

a real-valued function (see Figure 2.22a). However, this is no longer the case if we start to shift the rectangle in time. For example, translating the rectangular function one second to the right, as illustrated by Figure 2.22b, leaves the magnitude of the Fourier transform unchanged. However, the translation has a significant impact on the phase as well as on the real and imaginary parts of the Fourier transform. This again demonstrates that time information is not revealed by the magnitude, but that it is encoded in the phase of the Fourier transform. Let us have a more general look at this phenomenon. Let $f \in L^2(\mathbb{R})$ be a signal, then the function f_{t_0} defined by

$$f_{t_0}(t) := f(t - t_0) \quad (2.97)$$

is called the **translation** of f by $t_0 \in \mathbb{R}$, and the function f^{ω_0} defined by

$$f^{\omega_0}(t) := \exp(2\pi i \omega_0 t) f(t) \quad (2.98)$$

is called the **modulation** of f by $\omega_0 \in \mathbb{R}$. It is not hard to show (Exercise 2.22) that for the Fourier transform one obtains

$$\widehat{f_{t_0}}(\omega) = \exp(-2\pi i \omega t_0) \hat{f}(\omega) \quad (2.99)$$

and

$$\widehat{f^{\omega_0}}(\omega) = \hat{f}(\omega + \omega_0). \quad (2.100)$$

In other words, a translation of the signal in the time domain leads to a modulation in the Fourier domain, and vice versa.

2.3.4 Fourier Transform for DT-Signals

We finally introduce the Fourier transform for the signal space $\ell^2(\mathbb{Z})$, which consists of the finite-energy DT-signals. Recall from (2.32) that the most common discretization procedure to transform a CT-signal $f: \mathbb{R} \rightarrow \mathbb{R}$ into a DT-signal $x: \mathbb{Z} \rightarrow \mathbb{R}$ is equidistant sampling, where the samples are defined by $x(n) = f(n \cdot T)$, $n \in \mathbb{Z}$, for a given sampling rate $F_s = 1/T$ and sampling period $T > 0$.

Let $x \in \ell^2(\mathbb{Z})$ be an arbitrary DT-signal of finite energy, then the **Fourier representation** of x is

$$x(n) = \int_{\omega \in [0,1)} c_\omega \exp_\omega(n) d\omega = \int_{\omega \in [0,1)} c_\omega \exp(2\pi i \omega n) d\omega \quad (2.101)$$

for $n \in \mathbb{Z}$. Furthermore, the coefficients c_ω are given by the frequency-dependent function $\hat{x}: [0, 1) \rightarrow \mathbb{C}$ defined by

$$c_\omega = \hat{x}(\omega) := \sum_{n \in \mathbb{Z}} x(n) \overline{\exp_\omega(n)} = \sum_{n \in \mathbb{Z}} x(n) \exp(-2\pi i \omega n), \quad (2.102)$$

which is called the **Fourier transform** of x . Both the Fourier representation as well as the Fourier transform are nontrivial facts that require mathematical proofs. Although similar in nature, the Fourier transform for DT-signals cannot be directly derived from the Fourier transform for CT-signals. However, as we will see, the case of DT-signals can be regarded to be dual to the case of periodic CT-signals. Also, the Fourier transform of a sampled analog signal can be regarded as a kind of approximation of the Fourier transform of the analog signal.

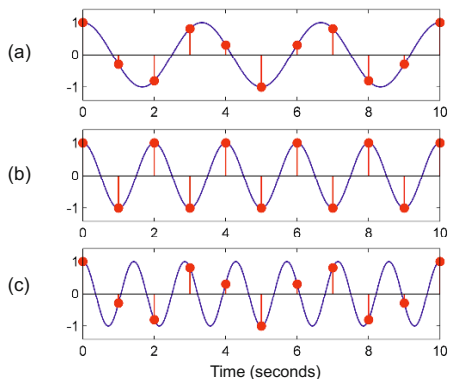
2.3.4.1 Periodicity and Aliasing

The Fourier representation (2.101) says that the signal x can be represented as an infinitesimal superposition of the elementary frequency functions \exp_ω sampled with $T = 1$ (see (2.32)). In this case, only the frequencies $\omega \in [0, 1)$ are needed. Intuitively, the restriction of the frequency parameters to the set $[0, 1)$ can be explained as follows: For an integer frequency parameter $k \in \mathbb{Z}$ and sampling points $n \in \mathbb{Z}$ one has $\exp(2\pi i k n) = 1$. Therefore,

$$\exp_{\omega+k}(n) = \exp(2\pi i(\omega + k)n) = \exp(2\pi i \omega n) \exp(2\pi i k n) = \exp_\omega(n). \quad (2.103)$$

In other words, two exponential functions with an integer difference in their frequency parameter coincide on the set of sampling points $n \in \mathbb{Z}$. Consequently, they cannot be distinguished when considered as 1-sampled DT-signals. We have encountered this **aliasing** phenomenon already in Figure 2.14 of Section 2.2.2. Using a sampling rate of 1 Hz, the **Nyquist frequency** is $\omega = 0.5$ Hz. All oscillations with a frequency above this rate are not captured by 1-sampling and lead to the

Fig. 2.23 Sinusoids of different frequencies ω sampled at a rate of $F_s = 1$ Hz. (a) $\omega = 0.3$ Hz. (b) $\omega = 0.5$ Hz. (c) $\omega = 0.7$ Hz.



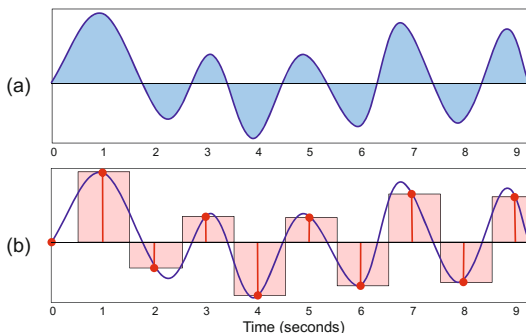
same samples as oscillations of lower frequencies. This fact is also illustrated by Figure 2.23.

Next, let us have a closer look at the Fourier transform (2.102). Note that (2.103) implies that the function $\omega \mapsto \exp(-2\pi i \omega n)$ is 1-periodic for all $n \in \mathbb{Z}$. Being a superposition of 1-periodic functions, also the Fourier transform \hat{x} is 1-periodic. Furthermore, one can show that the Fourier transform is energy preserving, i.e., $\|x\|_{\ell^2(\mathbb{Z})} = \|\hat{x}\|_{L^2([0,1])}$. Note that this is exactly the reverse of the situation we have seen for 1-periodic signals $f \in L^2([0,1])$, where the Fourier transform was a DT-signal $\hat{f} \in \ell^2(\mathbb{Z})$. Replacing the frequency parameter ω by the time parameter t , the formula (2.102) for the Fourier transform of $\ell^2(\mathbb{Z})$ becomes (up to a sign in the exponential function) the formula (2.79) for the Fourier representation of $L^2([0,1])$. A similar relation holds between the Fourier representation (2.101) for $\ell^2(\mathbb{Z})$ and the Fourier transform (2.80) for $L^2([0,1])$. From this it also follows that the Fourier transform for $\ell^2(\mathbb{Z})$ applied to the Fourier transform \hat{f} of a signal $f \in L^2([0,1])$ gives back the 1-periodic signal f up to a sign, i.e., $\hat{\hat{f}}(t) = f(-t)$. In mathematics, the close relation between the spaces $\ell^2(\mathbb{Z})$ and $L^2([0,1])$ and their Fourier transforms is also referred to as **duality**.

2.3.4.2 Riemann Approximation

Let us now investigate the relation between the Fourier transform of $L^2([0,1])$ and the one of $\ell^2(\mathbb{Z})$. Starting with a CT-signal $f \in L^2(\mathbb{R})$, let x be its T -sampled version. Then one obtains

Fig. 2.24 Approximation of the integral of an analog signal by a Riemann sum obtained from a 1-sampling. **(a)** Integral. **(b)** Riemann sum.



$$\begin{aligned}
 \hat{x}(\omega) &= \sum_{n \in \mathbb{Z}} x(n) \exp(-2\pi i \omega n) \\
 &= \sum_{n \in \mathbb{Z}} f(nT) \exp(-2\pi i \omega n) \\
 &\approx \int_{t \in \mathbb{R}} f(tT) \exp(-2\pi i \omega t) dt \quad (2.104) \\
 &= \frac{1}{T} \int_{t \in \mathbb{R}} f(t) \exp\left(\frac{-2\pi i \omega t}{T}\right) dt \\
 &= \frac{1}{T} \hat{f}\left(\frac{\omega}{T}\right),
 \end{aligned}$$

where we have used the substitution rule for indefinite integrals to replace tT by t . The approximation sign expresses that the value $\hat{x}(\omega)$ obtained by a sum has roughly the same size as the value $\hat{f}(\omega/T)/T$ obtained by an integral. This is a special case of the Riemann sum approximation, which we explain next.

Recall that the integral of a function is the (weighted) area determined by the function's graph and the time axis. In case of a complex-valued function, the complex-valued integral is defined by the integral of the real part and of the imaginary part of the function. For many functions, the integral can be approximated by partitioning the time axis into small intervals, picking the function value at the mid-point of each interval, and then summing up the interval lengths weighted by the respective value (see Figure 2.24). The resulting sum is also called the **Riemann sum** for the integral. The accuracy of the approximation very much depends on the resolution of the partition (the finer, the better the approximation) and the properties of the integrand (the slower it oscillates, the better the approximation).

In our case, the intervals of the partitioning have length one. Furthermore, the integrand is the function $h : \mathbb{R} \rightarrow \mathbb{C}$ defined by $h(t) := f(tT) \exp(-2\pi i \omega t)$, which basically is the product of the signal and an exponential function. Because of aliasing effects, in particular arising from the factor $\exp(-2\pi i \omega t)$, the Riemann sum does not yield a meaningful approximation for $\omega \in \mathbb{R} \setminus [-\frac{1}{2}, \frac{1}{2}]$. In particular, while \hat{x} is 1-periodic, the function $\omega \mapsto \hat{f}(\omega/T)/T$ is nonperiodic and approaches zero for $\omega \rightarrow \pm\infty$. Within the interval $[-\frac{1}{2}, \frac{1}{2}]$, however, in particular when approaching the

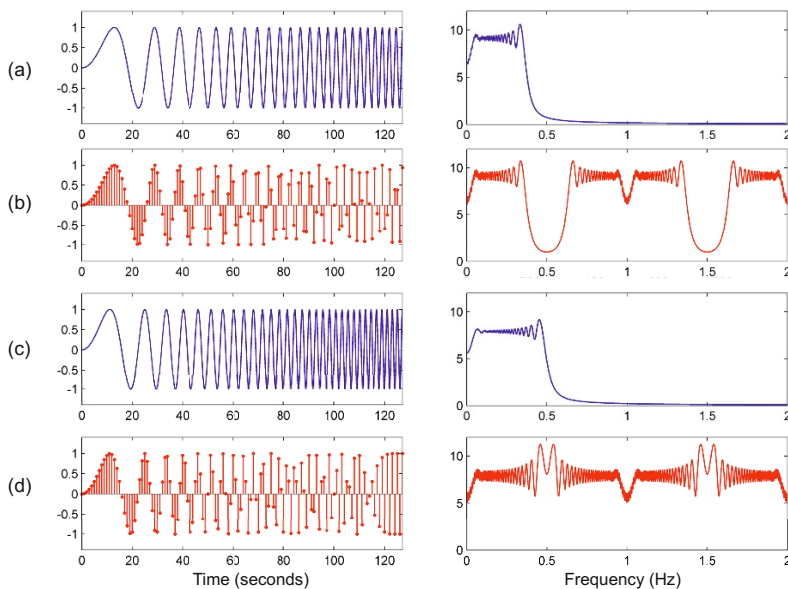


Fig. 2.25 Relation between the Fourier transform of a CT-signal and that of the DT-signal obtained by 1-sampling. Each row shows a signal (left) and its magnitude Fourier transform (right). **(a)** Analog chirp signal with $\lambda = 0.003$ and **(b)** its 1-sampled version. **(c)** Analog chirp signal with $\lambda = 0.004$ and **(d)** its 1-sampled version showing strong aliasing artifacts around the Nyquist frequency.

frequency $\omega = 0$, the Riemann sum $\hat{x}(\omega)$ approximates the value $\hat{f}(\omega/T)/T$ with increasing accuracy.

2.3.4.3 Chirp Signal Example

To further illustrate the relation between CT- and DT-signals and their Fourier transforms, we consider a signal in which the frequency increases with time. Such a signal is also called a **chirp signal** or **sweep signal**. In particular, for a given positive constant $\lambda > 0$, the function

$$f(t) := \begin{cases} \sin(\lambda \cdot \pi t^2), & \text{for } t \geq 0, \\ 0, & \text{for } t < 0, \end{cases} \quad (2.105)$$

defines a **linear chirp**, which is a sinusoidal wave that increases in frequency linearly over time. It can be shown that the **instantaneous frequency** at time $t = t_0$ is $\omega_0 = \lambda t_0$, which is the derivative of the phase divided by 2π . Figure 2.25 shows two chirp signals for different values of λ . In the first case (Figure 2.25a), the main frequencies are below $\omega \leq 0.4$, which is also shown by the magnitude Fourier transform. As a result, there is little aliasing when 1-sampling the signal (Figure 2.25b).

The Fourier transform \hat{x} of the resulting DT-signal x yields a good approximation of the Fourier transform \hat{f} in the range $[0, 0.5)$. Note that \hat{x} is 1-periodic whereas \hat{f} is not. Now, increasing the constant λ results in a chirp signal with frequency components above the Nyquist frequency of 0.5 (Figure 2.25c). Therefore, when 1-sampling the signal, there are aliasing artifacts where frequencies $0.5 + \omega$ are identified with frequencies $0.5 - \omega$ (see Figure 2.25d). In this case, the Riemann sum (2.104) yields a poor approximation of the actual integral.

2.4 Discrete Fourier Transform (DFT)

Computing the Fourier transform of signals involves the evaluation of integrals or infinite sums, which is, in general, computationally infeasible. In practice, as we have already discussed in Section 2.1.3, one typically approximates the Fourier transform by finite sums. Furthermore, the Fourier transform is evaluated only for a finite number of frequencies. In this section, we show how the finite sums and the Fourier coefficients must be chosen to obtain a linear transform known as the **discrete Fourier transform** (DFT). The important point is that the DFT can be computed efficiently by means of an algorithm, the famous **fast Fourier transform** (FFT). The FFT is considered one of the most important algorithms, being widely used for many applications in engineering and mathematics. In the following, we introduce the case of finite-length signals and their Fourier transform, which can then be formulated in terms of the DFT. We then describe in detail the FFT algorithm and discuss its computational complexity.

2.4.1 Signals of Finite Length

To derive the DFT, we start to reinvestigate the Fourier transform for a DT-signal $x \in \ell^2(\mathbb{Z})$. We assume that the energy of x is concentrated in the interval $[0 : N - 1]$, i.e., $x(n) \approx 0$ for $n \in \mathbb{Z} \setminus [0 : N - 1]$. Then we obtain from (2.102)

$$\hat{x}(\omega) = \sum_{n \in \mathbb{Z}} x(n) \overline{\exp_{\omega}(n)} \approx \sum_{n=0}^{N-1} x(n) \overline{\exp_{\omega}(n)} \quad (2.106)$$

for a frequency parameter ω . Recall that since \hat{x} is 1-periodic only the frequencies $\omega \in [0, 1)$ need to be considered. In practice, one often computes the Fourier transform only for a finite subset of frequencies. In particular, fixing a number $K \in \mathbb{N}$, one considers the frequencies $\omega = k/K$ for $k \in [0 : K - 1]$, which corresponds to a $1/K$ -sampling of the frequency space $[0, 1)$. Even though the number N of points in time and the number K of frequencies are not related at all, it is convenient to assume $N = K$. This assumption, as we will see, leads to a compact matrix-theoretic

formulation of the Fourier transform along with an efficient algorithm for computing the transform.

In the following, we assume $N = K$. Furthermore, let $x \in \ell^2(\mathbb{Z})$ be a signal that is zero outside the interval $[0 : N - 1]$ so that one obtains equality in (2.106). Such DT-signals are also referred to as **finite-length** signals, where N is the **length** of the signal. Each such signal x can be identified with a vector $\mathbf{x} := (x(0), x(1), \dots, x(N-1))^T \in \mathbb{C}^N$. This way, we can regard \mathbb{C}^N as a subspace of $\ell^2(\mathbb{Z})$, where the inner product (2.43) of $\ell^2(\mathbb{Z})$ reduces to the inner product (2.37) of \mathbb{C}^N . Not all frequencies $\omega \in [0, 1)$ are needed to characterize a signal of length N . Indeed, only the frequencies k/N for $k \in [0 : N - 1]$ suffice to represent such signals. To see this, we define a vector $\mathbf{u}_k \in \mathbb{C}^N$ for each $k \in [0 : N - 1]$ by setting

$$\mathbf{u}_k(n) := \exp_{k/N}(n) = \exp(2\pi i k n / N), \quad (2.107)$$

$n \in [0 : N - 1]$. In other words, the vector \mathbf{u}_k consists of the first N samples of the exponential function $\exp_{k/N}$. Then (2.106) can be expressed as

$$\hat{x}(k/N) = \sum_{n=0}^{N-1} x(n) \overline{\exp_{k/N}(n)} = \mathbf{x}^T \overline{\mathbf{u}_k} = \langle \mathbf{x} | \mathbf{u}_k \rangle. \quad (2.108)$$

Thus, the Fourier transform of a signal of length N can be obtained by inner products with the sampled and truncated exponential functions \mathbf{u}_k . We now show that these exponential functions (after rescaling) form an ON-basis of the Hilbert space \mathbb{C}^N . First, we define the number $\rho := \exp(2\pi i / N)$. Obviously, $\rho^N = 1$ and $\rho^k \neq 1$ for $k \in [1 : N - 1]$. Such a number is also called a **primitive** N^{th} root of unity (see also Exercise 2.23). Using the properties (2.70) and (2.71) of the exponential function, one obtains

$$\langle \mathbf{u}_k | \mathbf{u}_\ell \rangle = \sum_{n=0}^{N-1} \exp(2\pi i k n / N) \overline{\exp(2\pi i \ell n / N)} \quad (2.109)$$

$$= \sum_{n=0}^{N-1} \exp(2\pi i (k - \ell) n / N) = \sum_{n=0}^{N-1} \rho^{(k-\ell)n}. \quad (2.110)$$

for $k, \ell \in [0 : N - 1]$. In the case $k = \ell$, this implies $\|\mathbf{u}_k\|^2 = \langle \mathbf{u}_k | \mathbf{u}_k \rangle = N$. In the case $k \neq \ell$, one has $\rho^{(k-\ell)} \neq 1$. Therefore, one can apply the sum formula

$$\sum_{n=0}^{N-1} a^n = (1 - a^N) / (1 - a) \quad (2.111)$$

for geometric series, which holds for any complex number $a \neq 1$ (see Exercise 2.18). Setting $a = \rho^{(k-\ell)}$, one obtains

$$\langle \mathbf{u}_k | \mathbf{u}_\ell \rangle = \frac{1 - \rho^{N(k-\ell)}}{1 - \rho^{(k-\ell)}} = 0. \quad (2.112)$$

This shows that

$$\{\mathbf{u}_k/\sqrt{N} | k \in [0 : N-1]\} \quad (2.113)$$

is an ON-basis of the complex Hilbert space \mathbb{C}^N . In particular, from (2.52), one obtains the Fourier representation

$$\mathbf{x} = \frac{1}{N} \sum_{k=0}^{N-1} \langle \mathbf{x} | \mathbf{u}_k \rangle \mathbf{u}_k. \quad (2.114)$$

In other words, a finite-length signal can be represented as a weighted superposition of sampled and truncated exponential functions \mathbf{u}_k , where the weights are the Fourier coefficients given by (2.108). Next, we show how the Fourier transform and Fourier representation for finite-length signals relate to the discrete Fourier transform (DFT).

2.4.2 Definition of the DFT

Recall from (2.108) that the Fourier coefficients of a signal x of finite length N are given by

$$X(k) := \langle \mathbf{x} | \mathbf{u}_k \rangle = \sum_{n=0}^{N-1} x(n) \exp(-2\pi i k n / N) \quad (2.115)$$

for $k \in [0 : N-1]$. Let $\mathbf{X} := (X(0), X(1), \dots, X(N-1))^T \in \mathbb{C}^N$ denote the vector of Fourier coefficients. By definition, the **discrete Fourier transform** (DFT) is the mapping $\mathbb{C}^N \rightarrow \mathbb{C}^N$ that maps the input vector \mathbf{x} to the output vector \mathbf{X} . From (2.115) it is clear that this is a linear mapping, which can be described by the $(N \times N)$ matrix DFT_N given by

$$\text{DFT}_N(n, k) = \exp(-2\pi i k n / N). \quad (2.116)$$

One crucial observation is that there are many relations between the numbers $\exp(2\pi i k n / N)$ for $k, n \in [0, N-1]$. Using the primitive N^{th} root of unity $\rho = \exp(2\pi i / N)$ as well as the relations $\rho^{kn} = \exp(2\pi i k n / N)$ and $\omega := \bar{\rho} = \exp(-2\pi i / N)$, one obtains $\text{DFT}_N(n, k) = \omega^{kn}$. This yields the famous matrix

$$\text{DFT}_N = \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & \omega & \omega^2 & \cdots & \omega^{N-1} \\ 1 & \omega^2 & \omega^4 & \cdots & \omega^{2(N-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \omega^{N-1} & \omega^{2(N-1)} & \cdots & \omega^{(N-1)(N-1)} \end{pmatrix}. \quad (2.117)$$

Obviously, DFT_N is a symmetric matrix. Its columns are given by $\bar{\mathbf{u}}_k$ and its rows by $\bar{\mathbf{u}}_k^T$. In summary, we have seen that the Fourier transform \hat{x} of a DT-signal x of finite length N can be computed for frequencies $\omega = k/N$, $k \in [0 : N-1]$ by a single matrix–vector product $\mathbf{X} = \text{DFT}_N \cdot \mathbf{x}$.

The Fourier representation given by (2.114) is the inverse of the Fourier transform. For a spectral vector \mathbf{X} , it outputs the original signal \mathbf{x} . Again, being a linear mapping $\mathbb{C}^N \rightarrow \mathbb{C}^N$, the Fourier representation is given by a matrix, the inverse of the matrix DFT_N . From (2.113) it directly follows that

$$\text{DFT}_N^{-1} = \frac{1}{N} \overline{\text{DFT}_N}^\top = \frac{1}{N} \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & \rho & \rho^2 & \cdots & \rho^{N-1} \\ 1 & \rho^2 & \rho^4 & \cdots & \rho^{2(N-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \rho^{N-1} & \rho^{2(N-1)} & \cdots & \rho^{(N-1)(N-1)} \end{pmatrix}. \quad (2.118)$$

In other words, the inverse essentially coincides with the DFT matrix up to some normalizing factor and complex conjugation.

2.4.3 Fast Fourier Transform (FFT)

Note that the usual computation of the matrix–vector product $\mathbf{X} = \text{DFT}_N \cdot \mathbf{x}$ requires $O(N^2)$ multiplications and additions, which is too many for most applications. For example, having a signal with one thousand samples ($N = 10^3$) would require already a number of operations on the order of a million ($N^2 = 10^6$). In many cases one has to deal with much larger $N \gg 10^5$, which makes a naive computation of a DFT infeasible. The good news is that the DFT matrix is highly structured, which can be exploited when computing a matrix–vector product. The main idea lies in a factorization of the DFT matrix into a product of $O(\log N)$ sparse matrices, each of which can be evaluated with $O(N)$ operations. This leads to an efficient algorithm, the so-called **fast Fourier transform** (FFT), which only requires $O(N \log N)$ multiplications and additions. The FFT algorithm was originally found by Gauss in about 1805 and then rediscovered by Cooley and Tukey in 1965.

The FFT algorithm is based on the observation that applying a DFT of even size $N = 2M$ can be expressed in terms of applying two DFTs of half the size M . Let $\omega_N = \exp(-2\pi i/N)$ be the primitive root of unity used in DFT_N so that $\text{DFT}_N(n, k) = \omega_N^{kn}$ for $n, k \in [0 : N-1]$. Similarly, we define $\omega_M = \exp(-2\pi i/M)$ so that $\text{DFT}_M(n, k) = \omega_M^{kn}$ for $n, k \in [0 : M-1]$. Obviously, $\rho_M = \rho_N^2$. Let $\mathbf{x} \in \mathbb{C}^N$ be an input vector and $\mathbf{X} = \text{DFT}_N \cdot \mathbf{x}$ as before. Then for the first M entries $X(k)$, $k \in [0 : M-1]$ one has

$$X(k) = \sum_{n=0}^{N-1} x(n) \omega_N^{kn} \quad (2.119)$$

$$= \sum_{n=0}^{M-1} x(2n) \omega_N^{k2n} + \sum_{n=0}^{M-1} x(2n+1) \omega_N^{k(2n+1)} \quad (2.120)$$

$$= \sum_{n=0}^{M-1} x(2n) \omega_M^{kn} + \omega_N^k \sum_{n=0}^{M-1} x(2n+1) \omega_M^{kn}. \quad (2.121)$$

In other words, the first M entries of \mathbf{X} are obtained by first applying a DFT_M on the even-indexed entries of \mathbf{x} as well as a DFT_M on the odd-indexed entries of \mathbf{x} . The final result is then obtained by adding up the two output vectors, where the second one is adjusted by the factors ω_N^k , which are also known as **twiddle factors**. Similarly, for the last M entries $X(M+k)$, $k \in [M-1]$ one has

$$X(M+k) = \sum_{n=0}^{N-1} x(n) \omega_N^{(M+k)n} \quad (2.122)$$

$$= \sum_{n=0}^{M-1} x(2n) \omega_N^{(M+k)2n} + \sum_{n=0}^{M-1} x(2n+1) \omega_N^{(M+k)(2n+1)} \quad (2.123)$$

$$= \sum_{n=0}^{M-1} x(2n) \omega_M^{kn} - \omega_N^k \sum_{n=0}^{M-1} x(2n+1) \omega_M^{kn}, \quad (2.124)$$

where we have used $\omega_N^{M(2n+1)} = -1$. This shows that the last M entries of \mathbf{X} are obtained by the same computation scheme as the first M ones, except for using the twiddle factors $-\omega_N^k$ instead of ω_N^k . The following matrix factorization summarizes this result:

$$\text{DFT}_N \cdot \begin{pmatrix} x(0) \\ x(1) \\ \vdots \\ x(N-1) \end{pmatrix} = \begin{pmatrix} \text{id}_M & \Delta_M \\ \text{id}_M & -\Delta_M \end{pmatrix} \begin{pmatrix} \text{DFT}_M & 0 \\ 0 & \text{DFT}_M \end{pmatrix} \begin{pmatrix} x(0) \\ x(2) \\ \vdots \\ x(N-2) \\ x(1) \\ x(3) \\ \vdots \\ x(N-1) \end{pmatrix}. \quad (2.125)$$

The matrix $\text{id}_M = \text{diag}(1, 1, \dots, 1)$ denotes the $(M \times M)$ identity matrix and $\Delta_M = \text{diag}(1, \omega_N, \dots, \omega_N^{M-1})$ the $(M \times M)$ diagonal matrix containing the twiddle factors. The rearrangement of the input vector into components with an even and components with an odd index can be expressed by an additional permutation matrix. Altogether, this leads to a factorization of the DFT_N matrix into a product of sparse matrices (having only few nonzero coefficients) and DFT_M matrices of half the size.

Algorithm: FFT**Input:** The length $N = 2^L$ with N being a power of twoThe vector $(x(0), \dots, x(N-1))^T \in \mathbb{C}^N$ **Output:** The vector $(X(0), \dots, X(N-1))^T = \text{DFT}_N \cdot (x(0), \dots, x(N-1))^T$ **Procedure:** Let $(X(0), \dots, X(N-1)) = \text{FFT}(N, x(0), \dots, x(N-1))$ denote the general form of the FFT algorithm.If $N = 1$ then

$$X(0) = x(0).$$

Otherwise compute recursively:

$$(A(0), \dots, A(N/2 - 1)) = \text{FFT}(N/2, x(0), x(2), x(4), \dots, x(N-2)),$$

$$(B(0), \dots, B(N/2 - 1)) = \text{FFT}(N/2, x(1), x(3), x(5), \dots, x(N-1)),$$

$$C(k) = \omega_N^k \cdot B(k) \text{ for } k \in [0 : N/2 - 1],$$

$$X(k) = A(k) + C(k) \text{ for } k \in [0 : N/2 - 1],$$

$$X(N/2 + k) = A(k) - C(k) \text{ for } k \in [0 : N/2 - 1].$$

Table 2.1 Recursive version of the FFT algorithm.

The FFT algorithm is again summarized by the compact recursive version shown in Table 2.1.

What have we gained when evaluating the DFT_N by means of this procedure? Let $\mu(N)$ be the number of multiplications and additions⁷ needed to compute the matrix–vector product $\text{DFT}_N \cdot \mathbf{x}$. By (2.125), one needs to evaluate two DFT_M , which takes $2\mu(M)$ operations. Furthermore, at first sight, one seems to require $2M = N$ multiplications for the twiddle factors and $2M = N$ additions to sum up the output vectors from the DFT_M step. A closer look shows that one can do even better. First note that the first twiddle factor ($k = 1$) is $\omega_N^k = 1$, thus causing no multiplication cost. Furthermore, multiplication with the other twiddle factors ($k \in [1 : M - 1]$) needs to be done only once, but can be used twice (see $C(k)$ in Table 2.1, where it is used once in $X(k) = A(k) + C(k)$ and once in $X(N/2 + k) = A(k) - C(k)$). As a result, one requires only $M - 1$ multiplications for the twiddle factors (instead of $2M = N$). Altogether, one obtains the estimate

$$\mu(N) \leq 2\mu(N/2) + 1.5N. \quad (2.126)$$

Now, this procedure unfolds its full effect when applied recursively. To this end, one assumes that $N = 2^L$ is a power of two. Obviously $\mu(1) = 0$, since in the case $N = 1$ nothing has to be done. This leads to the following overall estimate:

⁷ In the following, subtractions are counted as additions.

$$\mu(N) \leq 2 \cdot \mu(N/2) + 1.5N \quad (2.127)$$

$$\leq 4 \cdot \mu(N/4) + 1.5N + 1.5N \quad (2.128)$$

$$\leq \dots \quad (2.129)$$

$$\leq 2^L \cdot \mu(1) + \underbrace{1.5N + 1.5N + \dots + 1.5N}_{L=\log_2(N) \text{ times}} \quad (2.130)$$

$$= 1.5N \log_2(N). \quad (2.131)$$

This equation can also be formally shown by a simple induction (see Exercise 2.26). The savings obtained from the FFT algorithm are huge, in particular for large N . For example, in the case $N = 10^3$, the FFT algorithm requires $2 \cdot 10^4$ operations instead of 10^6 as needed for the naive approach, which is a reduction of operations by a factor of 50. For $N = 10^5$, this factor is already 3,000, and for $N = 10^6$, it reaches 25,000. In this case, if the FFT requires a second of computing time, the naive approach would require 7 hours.

2.4.4 Interpretation of the DFT

Let us summarize the results obtained so far. We started with a CT-signal $f \in L^2(\mathbb{R})$ and derived a DT-signal x by T -sampling. Fixing a number $N \in \mathbb{N}$ of samples, we computed $\mathbf{X} = \text{DFT}_N \cdot \mathbf{x}$ for $\mathbf{x} = (x(0), \dots, x(N-1))^T$. What is the meaning of the Fourier coefficients $\mathbf{X} = (X(0), \dots, X(N-1))^T$ in relation to the original analog signal f ? To answer this question, we need to combine the results induced by the DFT approximation (2.106) and the Riemann approximation (2.104):

$$X(k) \approx \hat{x}\left(\frac{k}{N}\right) \approx \frac{1}{T} \hat{f}\left(\frac{k}{N} \cdot \frac{1}{T}\right). \quad (2.132)$$

In other words, to obtain the “correct” physical interpretation of the coefficient $X(k)$ one needs to know the window size N and the sampling rate $1/T$. First, $X(k)$ needs to be scaled by the factor T . Second, the index k corresponds to the frequency $\omega = k/(NT)$. In other words, the DFT computes the frequencies only on a linear grid of frequencies with a resolution of $1/(NT)$ Hz.

However, the approximations in (2.132) need to be taken with care. The first approximation is only good if the samples of $x(n)$ are close to zero outside the interval $[0 : N-1]$. Obviously, this is the case if the analog signal f is close to zero outside the interval $[0, (N-1)/T]$. Furthermore, recall that the second approximation is only good if f does not contain frequency components above the Nyquist frequency $1/(2T)$ Hz. Also, the approximation becomes poor for large k corresponding to high frequencies of the exponential functions. Assuming that f is real-valued, one can easily check that $\hat{f}(\omega) = \overline{\hat{f}(-\omega)}$, $\hat{x}(\omega) = \overline{\hat{x}(-\omega)}$, and $X(k) = \overline{X(N-k)}$ (see (2.83) and Exercise 2.24). Therefore, the coefficients $X(k)$ are redundant for

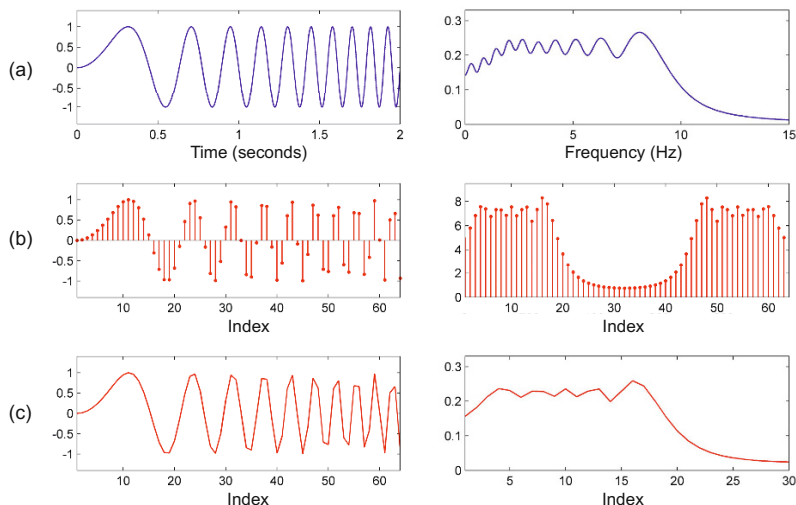


Fig. 2.26 DFT approximation of the Fourier transform. **(a)** Analog chirp signal and its Fourier transform. **(b)** Sampled signal using $T = 1/32$ and DFT coefficients using $N = 64$. **(c)** Interpolation of sampled signal and of DFT coefficients.

$k = \lfloor \frac{N}{2} \rfloor + 1, \dots, N-1$, and one only needs to consider the coefficients $X(k)$ for $k = 0, 1, \dots, \lfloor \frac{N}{2} \rfloor$.

As an example, let us consider the analog chirp signal shown in Figure 2.26a, where we assume that the signal is zero outside the shown interval $[0, 2]$. The Fourier transform is shown for frequencies $\omega \in [0, 15]$. Next, we sample the chirp signal using a sampling rate of $F_s = 32$ Hz and obtain a finite-length signal \mathbf{x} of length $N = 64$. Applying a DFT_N results in a complex-valued vector $\mathbf{X} = \text{DFT}_N \cdot \mathbf{x}$, the magnitude values of which are shown in Figure 2.26b. By (2.132), we obtain $X(k)/32 \approx \hat{f}(k/2)$. For example, the index $k = 30$ corresponds to the frequency $\omega = 15$ (see Figure 2.26c). The resulting frequency resolution is 0.5 Hz.

2.5 Short-Time Fourier Transform (STFT)

The Fourier transform \hat{f} of a signal $f \in L^2(\mathbb{R})$ describes the frequency content of the signal. Comparing the signal with a periodic exponential function $t \mapsto \exp(2\pi i \omega t)$ results in a coefficient $\hat{f}(\omega)$ that exhibits the overall intensity of oscillations at ω Hz occurring in the signal. However, because of the nonlocal nature of the analysis function, the frequency information is always averaged over the entire time domain. Sudden changes and local variations of the signal such as the beginning and the end of events cannot be detected well by the Fourier transform. Local phenomena of the signal become global phenomena in the Fourier transform. In contrast, small

changes in the phase of the Fourier transform can have considerable effects in the time domain.

To remedy the drawbacks of the Fourier transform, as we have already discussed in Section 2.1.4, Dennis Gabor introduced in the year 1946 the modified Fourier transform, now known as the **short-time Fourier transform** (STFT). This transform is a compromise between a time- and a frequency-based representation, determining the sinusoidal frequency and phase content of local sections of a signal as it changes over time. In this way, the STFT does not only tell which frequencies are “contained” in the signal but also at which points of times or, to be more precise, in which time intervals these frequencies appear. In the following, we start by introducing the STFT for the case of analog signals. From the STFT one can derive a spectrogram, which visually represents the time–frequency content of a signal. Finally, we introduce a discrete version of the STFT as it is typically used in practice. This is the version of the STFT we have already encountered in Section 2.1.4.

2.5.1 Definition of the STFT

For a given signal, we want to find a transform that exhibits the frequency content of f in a neighborhood of each point in time t . The basic idea is to consider only a small section of the signal around a point t , where the influence of a point within the section decreases with increasing distance from t . Mathematically, this weighting is modeled by multiplying the signal with a **window function**, which can be thought of as a weighting (often bell-shaped) function that localizes around t . Instead of using a different window function for each point t , one uses a single window function that localizes around the point $t = 0$. This function is then shifted across time. If $f \in L^2(\mathbb{R})$ is a signal and $g : \mathbb{R} \rightarrow \mathbb{R}$ is such a window function, then the function $f_{g,t}$ localized at point t is defined by

$$f_{g,t}(u) := f(u)g(u-t). \quad (2.133)$$

Figure 2.27 shows a chirp signal f as well as the resulting localized signals $f_{g,t}$ when using a bell-like window function g centered at zero for the shift parameters $t = 0.5$, $t = 1$, and $t = 1.5$, respectively.

In view of a general mathematical formulation, one often admits complex-valued window functions $g : \mathbb{R} \rightarrow \mathbb{C}$ and requires $g \in L^2(\mathbb{R})$ as well as $\|g\|_2 \neq 0$. Extending (2.133), the function $f_{g,t}$ is defined by

$$f_{g,t}(u) := f(u)\overline{g}(u-t). \quad (2.134)$$

Note that the complex conjugate does not play any role in case of a real-valued window g , which will always be the case in this book. Also, note that from a technical point of view, g does not need to have a particular shape.

Given a signal $f \in L^2(\mathbb{R})$ as well as a window function $g \in L^2(\mathbb{R})$, the (continuous-time) **short-time Fourier transform** (STFT) is a function $\tilde{f}_g : \mathbb{R} \times \mathbb{R} \rightarrow$

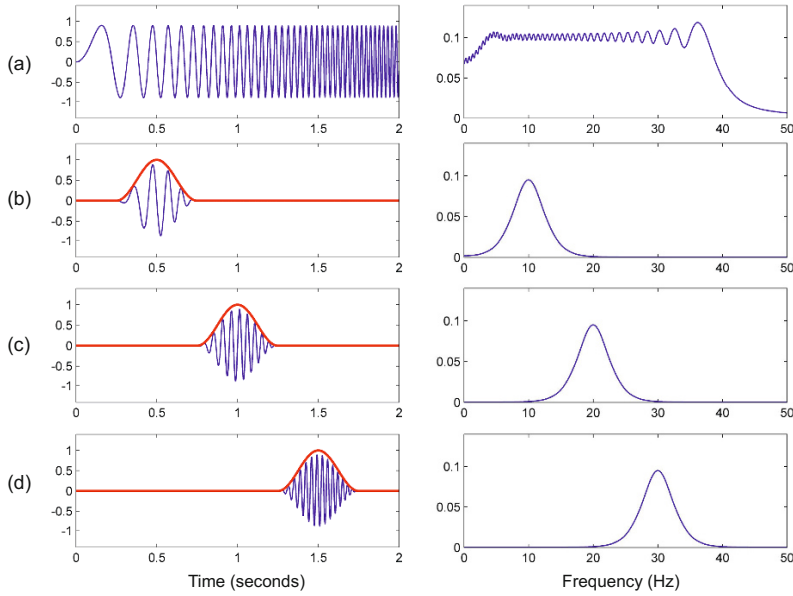


Fig. 2.27 Chirp signal and windowed versions along with their magnitude Fourier transforms. (a) Original signal. (b) Window centered at $t = 0.5$. (c) Window centered at $t = 1.0$. (d) Window centered at $t = 1.5$.

\mathbb{C} defined by

$$\widetilde{f}_g(t, \omega) := \widehat{f_{g,t}}(\omega) = \int_{u \in \mathbb{R}} f(u) \overline{g(u-t)} \exp(-2\pi i \omega u) du. \quad (2.135)$$

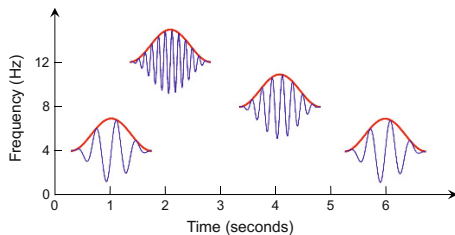
In other words, $\widetilde{f}_g(t, \cdot)$ coincides with the Fourier transform of the localized signal $f_{g,t}$ for a fixed time instance $t \in \mathbb{R}$.

As an illustration, let us continue with the example of Figure 2.27, which shows the chirp signal $f(t) = \sin(20 \cdot \pi t^2)$ for $t \in [0, 2]$. As we mentioned after (2.105), the instantaneous frequency at time t is $\omega = 20t$. Therefore, when considering the localized signal $f_{g,t}$ one may expect frequencies around $\omega = 20t$ Hz. Indeed, the Fourier transform $\widehat{f_{g,t}}$ reveals a peak at 10 Hz for $t = 0.5$ (Figure 2.27b), a peak at 20 Hz for $t = 1$ (Figure 2.27c), and a peak at 30 Hz for $t = 1.5$ (Figure 2.27d).

2.5.1.1 Alternative Definition of the STFT

When considering the short-time Fourier transform, one can assume a different viewpoint, which leads to a slightly different definition. In the above definition, we first windowed the original signal f with the time-shifted window g_t to obtain the localized signal $f_{g,t}$, which was then compared against the exponential func-

Fig. 2.28 Illustration of four different “musical notes” $g_{t,\omega}$ located in the time–frequency plane: $(t, \omega) = (1, 4)$, $(t, \omega) = (2, 12)$, $(t, \omega) = (4, 8)$, and $(t, \omega) = (6, 4)$.



tions \exp_{ω} . A different viewpoint is to construct localized elementary functions $g_{t,\omega} : \mathbb{R} \rightarrow \mathbb{C}$ by defining

$$g_{t,\omega}(u) := \exp(2\pi i \omega(u-t))g(u-t), \quad (2.136)$$

$u \in \mathbb{R}$. In other words, $g_{t,\omega}$ is obtained by first modulating the window g by ω Hz, which is a frequency shift in the Fourier domain (see (2.100)). The resulting modulated window is then shifted in time by t sec (see (2.97)). Intuitively, $g_{t,\omega}$ may be thought of as a “musical note” of frequency ω that is active in a neighborhood of t . The parameters t and ω allow for shifting the musical note in the time–frequency plane (see Figure 2.28).

It is not hard to see that $\|g_{t,\omega}\| = \|g\|$ for a window function $g \in L^2(\mathbb{R})$ (see Exercise 2.22). Therefore, as opposed to the exponential functions \exp_{ω} , which do not have finite energy, one has $g_{t,\omega} \in L^2(\mathbb{R})$. Therefore, we can define a function $\tilde{f}^g : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{C}$ by setting

$$\tilde{f}^g(t, \omega) = \langle f | g_{t,\omega} \rangle = \int_{u \in \mathbb{R}} f(u) \overline{g(u-t)} \exp(-2\pi i \omega(u-t)) du. \quad (2.137)$$

The inner product $\langle f | g_{t,\omega} \rangle$ measures the similarity between the signal f and the musical note $g_{t,\omega}$. If f and $g_{t,\omega}$ oscillate with the same frequency within the window, the inner product $\langle f | g_{t,\omega} \rangle$ has a large absolute value. Vice versa, if f has no frequency components around ω , the inner product is close to zero and f and $g_{t,\omega}$ are more or less orthogonal. The signal

$$u \mapsto \langle f | g_{t,\omega} \rangle g_{t,\omega}(u) \quad (2.138)$$

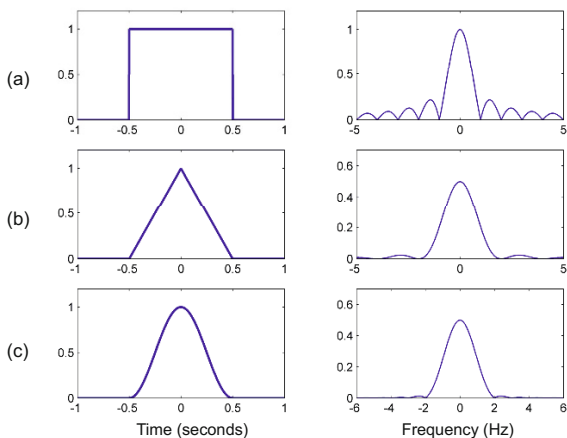
can be considered as the “projection” of the signal f in the direction of the musical note $g_{t,\omega}$ (see Figure 2.15).

The original STFT \tilde{f}_g defined by (2.135) and the version \tilde{f}^g defined by (2.137) coincide up to some time-dependent modulation factor:

$$\tilde{f}_g(t, \omega) = \tilde{f}^g(t, \omega) \exp(2\pi i \omega t). \quad (2.139)$$

In the first version only the window is shifted, whereas in the second version also the exponential function is shifted along with the window. Often \tilde{f}_g is used for the ana-

Fig. 2.29 Window functions and their Fourier transforms. (a) Rectangular window. (b) Triangular window. (c) Hann window.



log case, whereas \tilde{f}^g corresponds to what is used for the discrete Fourier transform (see for example (2.26)). We will come back to this issues in Section 2.5.3.

2.5.1.2 Role of the Window Function

We now discuss the role of the window function g , which plays an important role from a signal processing point of view. Typically, a window function is chosen to be zero-valued outside of some chosen section, so that when a signal is multiplied by the window function, the product is also zero-valued outside the section. The finite-length signal that is left can be regarded as a “view through the window.” The definition (2.135) shows that the STFT depends on both the signal as well as the window function, although one is typically interested only in the signal’s properties. The design of suitable window functions and their influence is a science by itself, which is outside the scope of this book. In the following, we discuss some examples that illustrate how the window may affect the spectral estimate computed by the STFT.

The seemingly simplest way to obtain a local view on the signal f is to leave it unaltered within the desired section and to set all values to zero outside the section. Such a localization is realized by a **rectangular window** as defined in (2.95) and again shown in Figure 2.29a. However, using the rectangular window has major drawbacks, since it generally leads to discontinuities at the section’s boundaries in the localized signal $f_{g,t}$. As we have discussed before, such abrupt changes lead to artifacts due to interferences which are spread over the entire frequency spectrum. Rather than being part of the original signal f , these frequency components come from the properties of the rectangular window (see Figure 2.29a). Recall that the Fourier transform of the rectangular window is the sinc function defined in (2.96), which shows slowly decaying ripples across the entire spectrum. These ripples also become visible in the STFT of a chirp signal as demonstrated by Figure 2.30a.

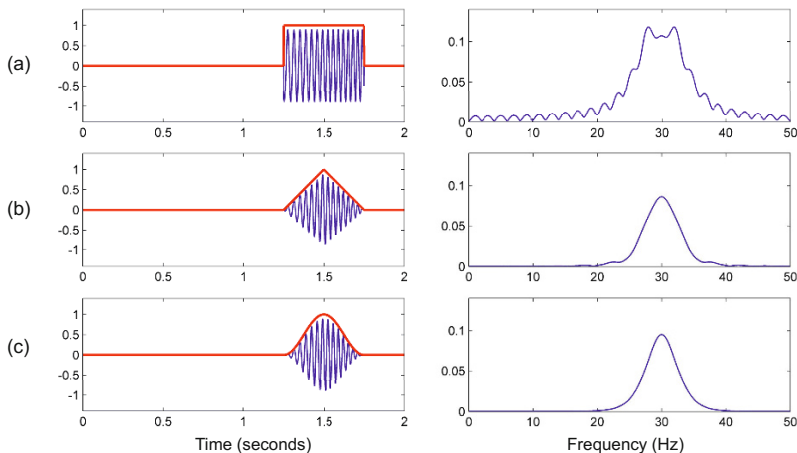


Fig. 2.30 Windowed chirp signal and its magnitude Fourier transform using different window functions. (a) Rectangular window. (b) Triangular window. (c) Hann window.

To attenuate the boundary effects, one often uses windows that are nonnegative within the desired section and continuously fall to zero towards the section's boundaries. One such example is the **triangular window** (Figure 2.29b), which leads to much smaller ripple artifacts (Figure 2.30b). A window often used in signal processing is the **Hann window** (also known as the **Hanning window**) named after Julius von Hann. The Hann window g is a raised cosine window defined by

$$g(u) := \begin{cases} (1 + \cos(\pi u))/2 & \text{if } -0.5 \leq u \leq 0.5 \\ 0 & \text{otherwise} \end{cases} \quad (2.140)$$

(see Figure 2.29c). Dropping smoothly to zero at the section boundaries, the above-mentioned artifacts in the Fourier transform of the windowed signal are softened. This is also illustrated by Figure 2.30c. However, on the downside, the Hann window introduces some smearing of frequencies. As a result, the Fourier transform of a signal's windowed section may look smoother than the signal's properties suggest. In other words, the reduction of ripple artifacts introduced by the window is achieved at the expense of a poorer spectral localization. Similarly, as we will see in the next section, the size of the window crucially affects the STFT.

2.5.2 Spectrogram Representation

The STFT of a signal f yields for each point in time t and frequency ω a complex number $\tilde{f}_g(t, \omega)$. This information is often visualized by means of a **spectrogram**, which is a two-dimensional representation of the squared magnitude:

$$\text{Spec}(t, \omega) = |\tilde{f}_g(t, \omega)|^2 = |\tilde{f}^g(t, \omega)|^2. \quad (2.141)$$

For the definition of the spectrogram, the version of the STFT in (2.139) does not matter, since the modulation factor has a magnitude of one. When generating an image of a spectrogram, the horizontal axis represents time, the vertical axis is frequency, and the dimension indicating the spectrogram value of a particular frequency at a particular time is represented by the intensity or color in the image. There are many variations in visualizing a spectrogram. Sometimes the vertical and horizontal axes are switched, so time runs up and down. Sometimes the amplitude is represented as the height of a 3D surface instead of color or intensity. To emphasize musical or tonal relationships, the frequency axis is often plotted in a logarithmic fashion, which yields a **log-frequency representation** as we will encounter in the subsequent chapters. A logarithmic frequency axis also accounts for the fact that human perception of pitch is logarithmic in nature (see Section 1.3.2). Finally, in the case of audio signals, the amplitude values are also often visualized using a logarithmic scale, for example, by using a decibel scale. In this way, small intensity values of perceptual relevance become visible in the image. In the following, if not specified otherwise, we use in our visualizations a linear frequency axis and a logarithmic scale to represent amplitudes. The specific scale is not of importance, but only serves the purpose of enhancing the qualitative properties of the visualization.

In our first example, we again consider a chirp signal f defined by $f(t) = \sin(400\pi t^2)$ for $t \in [0, 1]$, which is smoothly faded out towards $t = 1$ (see Figure 2.31a). For this chirp, the instantaneous frequency linearly raises from $\omega = 0$ Hz at $t = 0$ to $\omega = 400$ Hz at $t = 1$. For computing the STFT, we use a Hann window having a size of 62.5 ms. The resulting spectrogram is shown in Figure 2.31b. The logarithmic amplitude values are encoded by different gray levels, which are lighter for small values and darker for large values. Note that each column of the spectrogram corresponds to a plot of a Fourier transform as, for example, shown in Figure 2.30c.

The image of the spectrogram shows a strong diagonal stripe starting at the time–frequency point $(t, \omega) = (0, 0)$ and ending at $(t, \omega) = (1, 400)$, which reveals the linear frequency increase of the chirp signal. This diagonal stripe has a substantial width (roughly 40 Hz), which can be explained as follows: First, recall that at a given point t the STFT exhibits the frequency content of an entire neighborhood (a windowed section of the signal) around t , and the STFT averages the frequency information across this section. Second, as discussed in Section 2.5.1, the window introduces some additional smearing of frequencies in the Fourier domain. The artifacts introduced by the window function also explain the weaker diagonal stripes that run below and above the strong diagonal stripe. These weaker stripes correspond to the ripples occurring in the Fourier transform of the window function. As opposed to Figure 2.30c, where no such ripples can be seen for the Hann window, the ripples become visible in the visualization of the spectrogram only because we have used a logarithmic magnitude scale. We have already seen in Figure 2.30a that the ripple artifacts become much stronger when using a rectangular window instead of a Hann window. This phenomenon is illustrated by Figure 2.31c, which shows

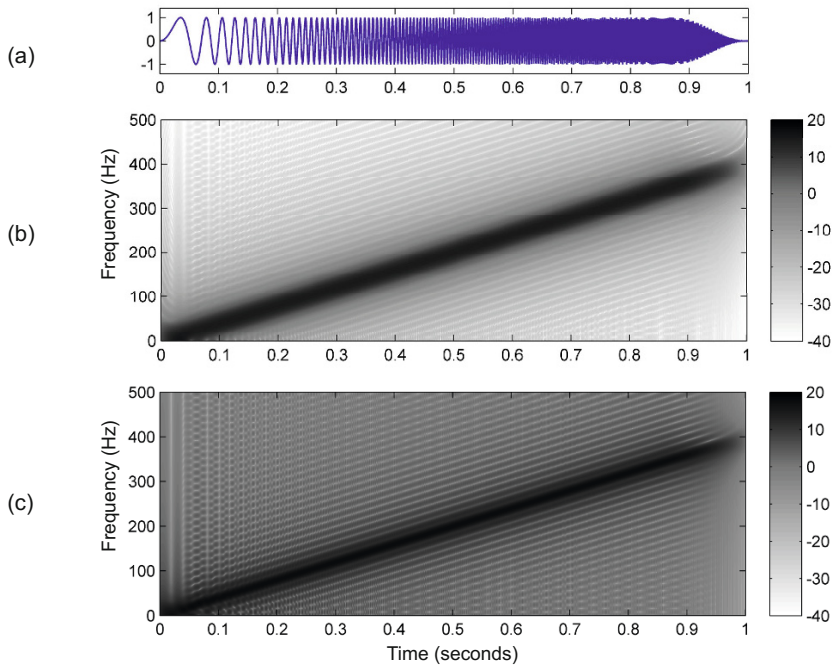


Fig. 2.31 Spectrogram of a chirp signal using two different window types. (a) Signal. (b) Spectrogram with Hann window of size 62.5 ms. (c) Spectrogram with rectangular window of size 62.5 ms.

a corresponding spectrogram. This visualization demonstrates the importance of choosing a suitable window function. In general, it is not easy to distinguish the characteristics of the signal and the effects introduced by the window function.

With the next example, we discuss the role of the size of the window function g . To this end, we consider the signal f shown in Figure 2.32a, which is defined by

$$f(t) = \sin(800\pi t) + \sin(900\pi t) + \delta(t - 0.45) + \delta(t - 0.5) \quad (2.142)$$

for $t \in [0, 1]$. In this interval, f is a superposition of two sinusoids of frequency 400 and 450 Hz, respectively. Furthermore, two impulses are added at the points $t = 0.45$ and $t = 0.5$ sec. Again we assume that f is zero outside the shown interval $[0, 1]$. This signal is interesting since it contains two components that are close in time (the two impulses that are 50 ms apart) and two components that are close in frequency (the two sinusoids that are 50 Hz apart). Figure 2.32b shows the spectrogram when using a Hann window of size 32 ms. The image contains a horizontal stripe in the region between 375 and 475 Hz, which corresponds to the sinusoids, as well as two vertical stripes at $t = 0.45$ and $t = 0.5$ sec, which correspond to the impulses. As illustrated by Figure 2.20b, each of the impulses results in many nonzero Fourier

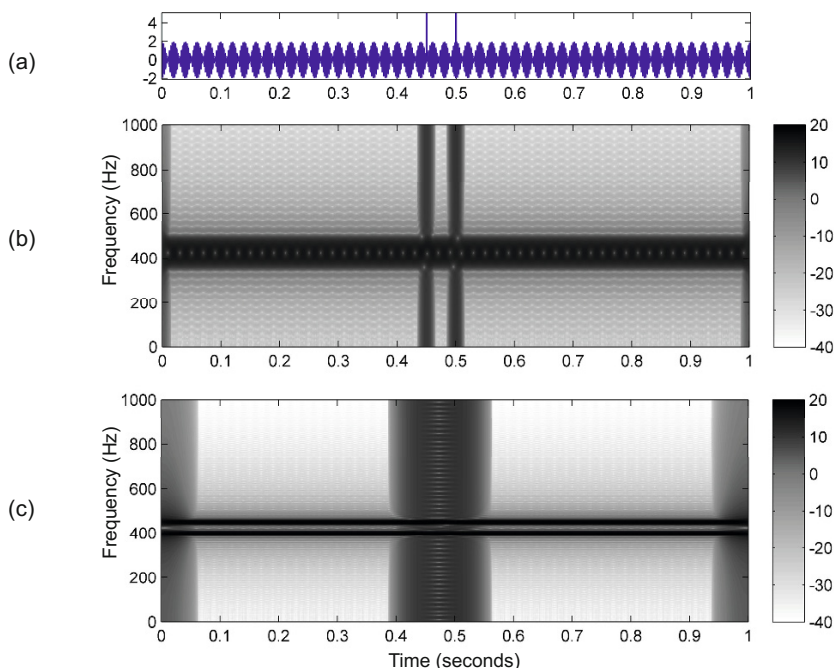


Fig. 2.32 Spectrogram using different window sizes. (a) Signal. (b) Spectrogram with short Hann window (32 ms). (c) Spectrogram with long Hann window (128 ms).

coefficients spread across the entire spectrum, which explains the vertical stripes. Since the window size of 32 ms implies that in each window there is at most one of the impulses, the two impulses can be clearly separated by the STFT. However, the STFT is not able to separate the two frequency components at $\omega = 400$ Hz and $\omega = 450$ Hz. The reason is that the chosen window introduces frequency smearing. The scaling property of the Fourier transform (Exercise 2.20) says that reducing the size by temporally compressing the window leads to a broadening of its Fourier transform. This, in turn, implies that the frequency smearing becomes more severe. Therefore, to separate the two frequency components, one strategy is to increase the window size, thus reducing the frequency smearing. Indeed, using a Hann window of size 128 ms results in a clear separation as shown by the two horizontal stripes (see Figure 2.32c). However, increasing the window size goes along with an increased smearing in the time domain. As a result, the two impulses are not separated any longer. As a side remark, we want to point to the two vertical stripes showing up at $t = 0$ and $t = 1$. An explanation is to be given in Exercise 2.27.

In summary, using a large window size results in a good localization in frequency, but a poor localization in time, whereas using a small window size has the opposite effect. Increasing the window size leads to an STFT which averages the frequencies of the signal over a greater time interval, resulting in a loss of time information. In

the limit case of an “infinite window size” one ends up with the usual Fourier transform, which averages the frequencies over the entire time domain \mathbb{R} . Vice versa, successively decreasing the window size results in a Dirac sequence, where, in the limit case of g being an impulse, the STFT gives back the original signal: perfect time localization, no frequency localization.

The time localization property of the STFT depends on the temporal spread of the window function g , whereas the frequency localization property of the STFT depends on the spectral spread of the Fourier transform \hat{g} . We want to mention that one cannot have both properties at the same time. A variant of the **Heisenberg uncertainty principle** says that there is no window function that simultaneously localizes in time and frequency with arbitrary precision.

2.5.3 Discrete Version of the STFT

So far, we have discussed the STFT and spectrogram in the case of analog signals. In practice, one uses sampled signals and computes the STFT only on a finite time–frequency grid. Because of efficiency issues, one typically employs DFTs which can be computed by means of the FFT algorithm. As before, let x be a DT-signal obtained from a CT-signal f by T -sampling. Furthermore, let w be a sampled version of an analog window function g . In the discrete case, the window can be shifted only in a sample-wise fashion. Because of efficiency issues, one often shifts the window in even larger steps, which are specified by some **hop size** parameter $H \in \mathbb{N}$ (given in samples). Following the alternative definition (2.137) in the analog case, we define the (discrete-time) STFT \tilde{x}^w of the DT-signal x with respect to the window function w by

$$\tilde{x}^w(m, \omega) := \sum_{n \in \mathbb{Z}} x(n) \bar{w}(n - mH) \exp(-2\pi i \omega(n - mH)) \quad (2.143)$$

$$= \sum_{n \in \mathbb{Z}} x(n + mH) \bar{w}(n) \exp(-2\pi i \omega n) \quad (2.144)$$

for $m \in \mathbb{Z}$ and $\omega \in [0, 1)$. Now, if the sampled window function w is a finite signal, the sum in (2.144) becomes finite, and we can apply the DFT to compute the discrete STFT for certain frequencies.

In the analog case, we assumed that the window function g was centered at time zero. To simplify the formulas in the discrete case, we assume that the support of the window function is contained only in the positive part of the time axis centered at half the window length (i.e., the window is shifted by half a window length to the right compared with the zero-centered case). The zero-centered case can be easily restored by also shifting the original signal by half a window length.

Having said this, we assume that the nonzero samples of the discrete window w are $w(n)$ for $n \in [0 : N - 1]$. For each frame index $m \in \mathbb{Z}$, we define the vector $\mathbf{x}_m = (x_m(0), \dots, x_m(N - 1))^T \in \mathbb{C}^N$ with

$$x_m(n) = x(n + mH)\bar{w}(n) \quad (2.145)$$

for $n \in [0 : N - 1]$ and compute the vector $\mathbf{X}_m = (X_m(0), \dots, X_m(N - 1))^T \in \mathbb{C}^N$ via a DFT of size N :

$$\mathbf{X}_m = \text{DFT}_N \cdot \mathbf{x}_m. \quad (2.146)$$

Then one obtains

$$\begin{aligned} \tilde{x}^w(m, k/N) &= \sum_{n=0}^{N-1} x(n + mH)\bar{w}(n) \exp(-2\pi i k n / N) \\ &= \sum_{n=0}^{N-1} x_m(n) \exp(-2\pi i k n / N) \\ &= X_m(k) \end{aligned} \quad (2.147)$$

for $k \in [0 : N - 1]$. Thus, we have shown that, for each time frame $m \in \mathbb{Z}$, one can compute the discrete STFT at frequencies $\omega = k/N$ for $k \in [0 : N - 1]$ by means of a DFT_N . In the case that N is a power of two, this can be done efficiently using the FFT.

2.5.3.1 Summary

Altogether, we have reached exactly the version of the discrete STFT already introduced in Section 2.1.4. Let us again summarize the main results. Let x be a DT-signal obtained by T -sampling. Furthermore, let w be a discrete window of finite length N with coefficients $w(n)$ for $n \in [0 : N - 1]$. Then

$$\mathcal{X}(m, k) = \tilde{x}^w(m, k/N) = \sum_{n=0}^{N-1} x(n + mH)\bar{w}(n) \exp(-2\pi i k n / N) \quad (2.148)$$

is the **discrete STFT** or simply the STFT of x (see also (2.26)). Each spectral vector for some time frame $m \in \mathbb{Z}$ can be computed by using a DFT_N , which can be evaluated efficiently by using an FFT if N is a power of two. The coefficients $\mathcal{X}(m, k)$ have a similar interpretation as discussed in Section 2.4.4. First recall that the upper half of the frequency coefficients are redundant if x and w are real-valued. In this case, one only considers the coefficients $k \in [0 : N/2]$. By (2.132), the index k corresponds to the frequency

$$F_{\text{coef}}(k) := \frac{k \cdot F_s}{N} \quad (2.149)$$

(see also (2.28)). In particular, the index $k = N/2$ corresponds to the Nyquist frequency $\omega = 1/(2T)$.

Next, we discuss how the index m is to be interpreted. The interpretation is not straightforward since m refers to an entire windowed section of the signal rather than a specific point in time. In signal processing, such a windowed section is also called

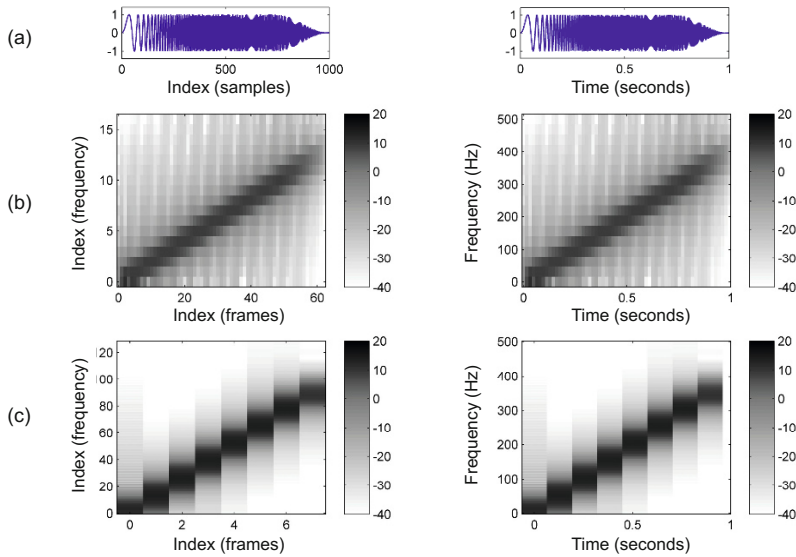


Fig. 2.33 Spectrogram representation of Discrete STFT. Shown are the original indices (left) and their physical interpretation (right). **(a)** Signal using $1/T = 1000$ Hz. **(b)** Spectrogram using $N = 32$ and $H = 16$. **(c)** Spectrogram using $N = 256$ and $H = 128$.

a **frame** and m is called the **frame index**. The physical duration of a frame is NT seconds. There are no strict conventions for associating a physical time position to a frame. When assuming that the window is centered at zero, as we did in the case of the continuous-time version of the STFT, one should take the center of the frame as a physical reference point. When assuming that the window starts at time position zero centered at half the window length, one may take the start of the frame as a physical reference point. As said before, the second convention can be transferred into the first one by shifting the original signal by half a window length. In the following, we want to adapt the second convention such that the frame index m is associated to the physical time position

$$T_{\text{coef}}(m) := \frac{m \cdot H}{F_s} \quad (2.150)$$

(see also (2.27)). Using this convention, the index $m = 0$ is associated with $t = 0$.

2.5.3.2 Examples

In Figure 2.9 we have already seen an example of how to interpret the frame and frequency indices in terms of physical units such as seconds and Hertz. Let us consider a second example to illustrate the effect of different parameter settings.

Figure 2.33a shows a DT-signal based on a sampling rate of $F_s = 1/T = 1000$ Hz. To compute the spectrogram of Figure 2.33b, a window length of $N = 32$ and a hop size of $H = 16$ were used. This yields a frame size of $NT = 32$ ms and the frame index m corresponds to time $T_{\text{coef}}(m) = mTH = m \cdot 16$ ms, which is also the time resolution of the STFT. In particular, frame index $m = 62$ corresponds to $T_{\text{coef}}(m) = 0.992 \approx 1$ sec. Furthermore, the frequency index k corresponds to frequency $F_{\text{coef}}(k) = k/(NT) = k \cdot 31.25$ Hz. In particular $k = 16$ yields the Nyquist frequency $F_{\text{coef}}(16) = 500$ Hz. A second parameter setting using $N = 256$ and $H = 128$ is shown in Figure 2.33c.

2.6 Further Notes

In this chapter, we have studied fundamental techniques for analyzing signals by means of elementary sinusoidal functions, which possess an explicit physical meaning in terms of frequency. We have considered various types of signals including analog or CT-signals as well as DT-signals or more general digital signals, which were obtained by sampling and quantization. Generally speaking, the CT-domain gives the “right” interpretation of physical phenomena, whereas the DT-domain is used to do the actual computations. Being the most important tool for processing audio signals, we have introduced different variants of the Fourier transform for the CT- as well as for the DT-domain. The **Fourier transform** converts a time-dependent signal into frequency-dependent coefficients, each of which indicates the strength of the respective elementary exponential function. The process of decomposing a signal into frequency components is also called **Fourier analysis**. In contrast, we have seen that the **Fourier representation** rebuilds a signal from the elementary functions, a process also called **Fourier synthesis**. The Fourier transform and the Fourier representation are closely related, leading to very similar formulas (see Table 2.2 for an overview). Many of these formulas can be expressed by inner products, which makes it also possible to use the same geometric language one knows from finite-dimensional Euclidean spaces.

We now give some references and pointers to literature for further reading. This chapter is a vastly expanded version of a summary on the Fourier transform given in [12, Section 2.2]. The basic definitions and main properties of the Fourier transform are covered in most introductory books on signal processing. As example references, we want to mention the classical textbook on *Signals and Systems* by Oppenheim et al. [13] or the book on *Digital Signal Processing* by Proakis and Manolakis [14]. Most signal processing software contains implementations of the Fourier transform. For example, all figures shown in this chapter have been generated using MATLAB [11]. An entertaining and nontechnical introduction to the main ideas of time–frequency analysis can be found in the book *The World According to Wavelets* by Hubbard [9]. Also Wikipedia contains many interesting articles, which have served as a source of inspiration for this chapter.

Signal space	$L^2(\mathbb{R})$	$L^2([0, 1))$	$\ell^2(\mathbb{Z})$
Inner product	$\langle f g \rangle = \int_{t \in \mathbb{R}} f(t) \overline{g(t)} dt$	$\langle f g \rangle = \int_{t \in [0, 1)} f(t) \overline{g(t)} dt$	$\langle x y \rangle = \sum_{n \in \mathbb{Z}} x(n) \overline{y(n)}$
Norm	$\ f\ _2 = \sqrt{\langle f f \rangle}$	$\ f\ _2 = \sqrt{\langle f f \rangle}$	$\ x\ _2 = \sqrt{\langle x x \rangle}$
Definition	$L^2(\mathbb{R}) := \{f : \mathbb{R} \rightarrow \mathbb{C} \mid \ f\ _2 < \infty\}$	$L^2([0, 1)) := \{f : [0, 1) \rightarrow \mathbb{C} \mid \ f\ _2 < \infty\}$	$\ell^2(\mathbb{Z}) := \{f : \mathbb{Z} \rightarrow \mathbb{C} \mid \ x\ _2 < \infty\}$
Elementary frequency function	$\mathbb{R} \rightarrow \mathbb{C}$ $t \mapsto \exp(2\pi i \omega t)$	$[0, 1) \rightarrow \mathbb{C}$ $t \mapsto \exp(2\pi i k t)$	$\mathbb{Z} \rightarrow \mathbb{C}$ $n \mapsto \exp(2\pi i \omega n)$
Frequency parameter	$\omega \in \mathbb{R}$	$k \in \mathbb{Z}$	$\omega \in [0, 1)$
Fourier representation	$f(t) = \int_{\omega \in \mathbb{R}} c_\omega \exp(2\pi i \omega t) d\omega$	$f(t) = \sum_{k \in \mathbb{Z}} c_k \exp(2\pi i k t)$	$x(n) = \int_{\omega \in [0, 1)} c_\omega \exp(2\pi i \omega n) d\omega$
Fourier transform	$\hat{f} : \mathbb{R} \rightarrow \mathbb{C}$ $\hat{f}(\omega) = c_\omega = \int_{t \in \mathbb{R}} f(t) \exp(-2\pi i \omega t) dt$	$\hat{f} : \mathbb{Z} \rightarrow \mathbb{C}$ $\hat{f}(k) = c_k = \int_{t \in [0, 1)} f(t) \exp(-2\pi i k t) dt$	$\hat{x} : [0, 1) \rightarrow \mathbb{C}$ $\hat{x}(\omega) = c_\omega = \sum_{n \in \mathbb{Z}} x(n) \exp(-2\pi i \omega n)$

Table 2.2 Overview of the signal spaces $L^2(\mathbb{R})$, $L^2([0, 1))$, and $\ell^2(\mathbb{Z})$ and their respective Fourier representation and Fourier transform.

In this chapter, we have used clear mathematical modeling which is necessary when one wants to understand the relation between the CT- and DT-domain. Additional mathematical structures such as the inner products or the complex formulation of the Fourier transform lead to compact and intuitive formulas. A geometric approach to signal processing can be found in the two recent books by Vetterli et al. [16, 17], which build on each other. Although we have used the notion of Lebesgue spaces, we have not introduced them with rigor. In particular in the case of CT-signals, the definition of Lebesgue spaces becomes a bit tricky, since one needs the notion of measurability of the functions in order for the integrals to be defined. For a mathematically rigorous treatment of measure and Lebesgue theory, we refer to the book *Real Analysis* by Folland [5]. As we have already indicated before, the spaces $L^2(\mathbb{R})$ and $L^2([0, 1))$ are actually quotient spaces where two functions f and g are considered to coincide if $\|f - g\|_2 = 0$, i.e., if they differ only up to a null set. The equality in the Fourier representation and in the Fourier transform is just an equality in the L^2 -sense, which is a weaker notion than pointwise equality. Under additional assumptions on f one also obtains pointwise equality. For example, if f is a continuously differentiable periodic CT-signal, the Fourier series converges uniformly to f on the interval $[0, 1)$ and one obtains pointwise equality. We have also mentioned before that the integral in the definition (2.92) of the Fourier transform of a signal $f \in L^2(\mathbb{R})$ does not exist in general. Instead, one needs to define the integral by some limit process (2.93). The existence of the limit is based on the so-called

Hahn–Banach theorem [5]. One main problem in the CT case is that the exponential functions $\exp_{\omega} : \mathbb{R} \rightarrow \mathbb{C}$ are not contained in $L^2(\mathbb{R})$. Therefore, the integral in (2.92) cannot be written as an inner product as is possible for the Fourier coefficients (2.80). Finally, note that we have not given any proofs for the existence and correctness of the considered Fourier transforms and Fourier representations. These proofs are outside the scope of this book and can be found in Folland [5]. In particular, in the case of periodic signals and DT-signals, the completeness property (2.52) is more difficult to prove and requires some quite technical machinery.

In this chapter, we have only scratched the topics of sampling and aliasing, which are of crucial importance for digital signal processing. In general, there are many ways to approximate a CT-signal and to describe it by a finite number of discrete parameters. For example, the discrete set of parameters could be the Fourier coefficients (for periodic signals), the coefficients of polynomials (when representing a function by its Taylor series), or the values of a CT-signal at a finite number of points in time. In all cases there are certain requirements on the original CT-signal, e.g., periodicity or differentiability, to guarantee certain bounds on the approximation error. In the case of sampling, these requirements concern the frequency content of the original signal. The famous **sampling theorem** says that an Ω -**bandlimited** signal $f \in L^2(\mathbb{R})$ (i.e., where the Fourier transform \hat{f} vanishes for $|\omega| > \Omega$ for a real number $\Omega > 0$) can be reconstructed perfectly from the T -sampling of f with $T := 1/(2\Omega)$ (see [13, 14]). In Exercise 2.28, we cover this important result in more detail. The sampling theorem is often associated with the names Harry Nyquist and Claude Shannon. It is interesting to note that the theorem was also discovered independently by Edmund Taylor Whittaker, Vladimir Kotelnikov, and others (see [1, 8] for an overview and historical notes).

There also exists a vast literature on the discrete Fourier transform (DFT) and its companion algorithm, the fast Fourier transform (FFT). In the original article by Cooley and Tukey [3], the authors describe an algorithm that works in case that the length N of the DFT is a power of two. By applying several tricky modifications of the FFT, this result can be extended to an algorithm for evaluating a DFT of arbitrary length $N \in \mathbb{N}$ with time complexity of $O(N \log N)$. A detailed description of this result can be found in the book *Fast Fourier Transforms* by Clausen and Baum [2], which treats this topic from an algebraic point of view. In particular, Section 2.4.3 closely follows [2, Section 1.3].

The short-time Fourier transform (STFT), which is also often referred to as the **windowed Fourier transform**, was pioneered in the year 1946 by Dennis Gabor for use in communication theory [6]. We have seen that the STFT is a compromise between a time- and a frequency-based representation of the signal. For a detailed discussion of the role of the window function used in the STFT calculation, we refer to [7]. One main drawback of the STFT is that the window function g implies a kind of rigid time–frequency resolution. As a result, properties of a signal that are much shorter than the window size are “synthesized” in the frequency domain, whereas properties of the signal that are much longer than the window size are “synthesized” in the time domain. In both cases many of the “notes” $g_{\omega,t}$ are needed to represent the phenomena of the signal. To remedy this problem, numerous alternatives

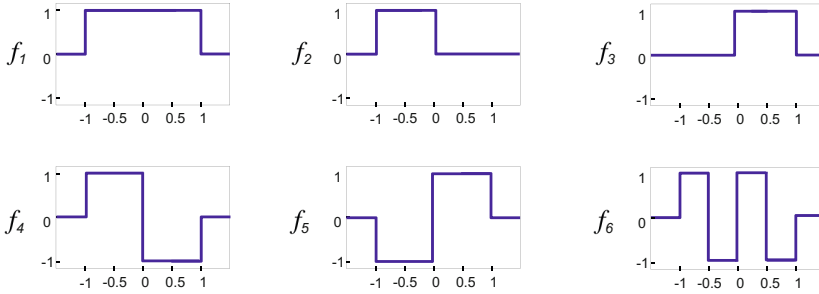
have been suggested, including time–frequency representations based on **wavelets**. For further reading and links on this topic, we refer to [4, 10, 15, 17]. Parts of Section 2.5 including the notation and the association of $g_{\omega,t}$ to “musical notes” were inspired by [10, Chapter 2]. Finally, we want to mention that so far we have mainly looked at the magnitude of the Fourier coefficients. In the later chapters, we will also have a closer look at the phase information, which can be used to refine the frequency estimation. Furthermore, the phase becomes important when reconstructing a signal from a modified STFT.

References

1. P. L. BUTZER, W. SPLETTSTÖSSER, AND R. L. STENS, *The sampling theorem and linear prediction in signal analysis*, Jahresbericht der Deutschen Mathematiker-Vereinigung, 90 (1988), pp. 1–70.
2. M. CLAUSEN AND U. BAUM, *Fast Fourier Transforms*, BI Wissenschaftsverlag, 1993.
3. J. W. COOLEY AND J. W. TUKEY, *An algorithm for the machine calculation of complex Fourier series*, Mathematics of Computation, 19 (1965), pp. 297–301.
4. I. DAUBECHIES, *Ten lectures on wavelets*, Society for Industrial and Applied Mathematics (SIAM), 1992.
5. G. B. FOLLAND, *Real Analysis*, John Wiley & Sons, 1984.
6. D. GABOR, *Theory of communication*, Journal of the Institution of Electrical Engineers (IEE), 93 (1946), pp. 429–457.
7. F. J. HARRIS, *On the use of windows for harmonic analysis with the discrete Fourier transform*, Proceedings of the IEEE, 66 (1978), pp. 51–83.
8. J. R. HIGGINS, *Five short stories about the cardinal series*, Bulletin of the American Mathematical Society, 12 (1985), pp. 45–89.
9. B. B. HUBBARD, *The world according to wavelets*, AK Peters, Wellesley, Massachusetts, 1996.
10. G. KAISER, *A Friendly Guide to Wavelets*, Modern Birkhäuser Classics, 2011.
11. MATLAB, *High-performance numeric computation and visualization software*. The MathWorks Inc., <http://www.mathworks.com>, 2013.
12. M. MÜLLER, *Information Retrieval for Music and Motion*, Springer Verlag, 2007.
13. A. V. OPPENHEIM, A. S. WILLSKY, AND H. NAWAB, *Signals and Systems*, Prentice Hall, 1996.
14. J. G. PROAKIS AND D. G. MANOLAKIS, *Digital Signal Processing*, Prentice Hall, 1996.
15. G. STRANG AND T. NGUYEN, *Wavelets and Filter Banks*, Wellesley-Cambridge Press, 2nd ed., 1996.
16. M. VETTERLI, J. KOVACEVIC, AND V. K. GOYAL, *Foundations of Signal Processing*, Cambridge University Press, <http://fourierandwavelets.org/>, 1st ed., 2013.
17. ———, *Fourier and Wavelet Signal Processing*, Cambridge University Press, <http://fourierandwavelets.org/>, 1st ed., 2014.

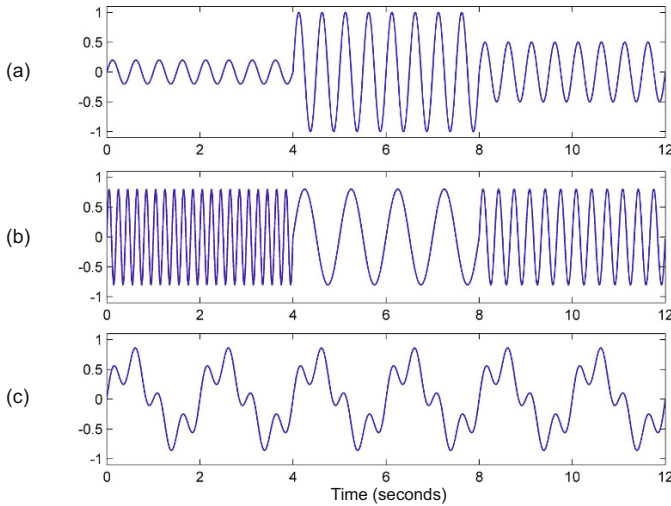
Exercises

Exercise 2.1. Let $\langle f|g \rangle := \int_{\mathbb{R}} f(t) \cdot g(t) dt$ be the similarity measure for two functions $f : \mathbb{R} \rightarrow \mathbb{R}$ and $g : \mathbb{R} \rightarrow \mathbb{R}$ as defined in (2.3). Consider the following six functions $f_n : \mathbb{R} \rightarrow \mathbb{R}$ for $n \in [1 : 6]$, which are defined to be zero outside the shown interval:



Determine the similarity values $\langle f_n | f_m \rangle$ for all pairs $(n, m) \in [1 : 6] \times [1 : 6]$.

Exercise 2.2. Sketch the magnitude Fourier transform of the following signals assuming that the signals are zero outside the shown intervals (see Figure 2.6 for similar examples):



Exercise 2.3. Based on (2.27) and (2.28), compute the time resolution (in ms) and frequency resolution (in Hz) of a discrete STFT based on the following parameter settings:

(a) $F_s = 22050$, $N = 1024$, $H = 512$

(b) $F_s = 48000$, $N = 1024$, $H = 256$

(c) $F_s = 4000$, $N = 4096$, $H = 1024$

What are the respective Nyquist frequencies?

Exercise 2.4. Let $F_s = 44100$, $N = 2048$, and $H = 1024$ be the parameter settings of a discrete STFT \mathcal{X} as defined in (2.26). What is the physical meaning of the Fourier coefficients $\mathcal{X}(1000, 1000)$, $\mathcal{X}(17, 0)$, and $\mathcal{X}(56, 1024)$, respectively? Why is the coefficient $\mathcal{X}(56, 1024)$ problematic?

Exercise 2.5. Sketch the magnitude Fourier transform (as in Figure 2.9) for each of the three signals shown in Exercise 2.2. Assume a window length that corresponds to a physical duration of about one second.

Exercise 2.6. The naive approach for computing a DFT requires about N^2 operations, while the FFT requires about $N \log_2 N$ operations. Compute the factor for the savings when using the FFT for various N . In particular, consider $N = 2^n$ for $n = 5, 10, 15, 20, 25, 30$.

Exercise 2.7. Let f_1 and f_2 be two periodic analog signals with integer periods $\lambda_1 \in \mathbb{N}$ and $\lambda_2 \in \mathbb{N}$, respectively. Show that $g = f_1 + f_2$ is periodic with periods that are integer multiples of λ_1 as well as λ_2 . In general, g may have additional periods not necessarily being integer multiples of λ_1 and λ_2 . As an example, specify two signals f_1 and f_2 with prime period $\lambda_1 = \lambda_2 = 2$ such that $g = f_1 + f_2$ is periodic with prime period $\lambda = 1$.

Exercise 2.8. In this exercise, we show that there are periodic functions that do not have a prime period (i.e., that do not have a least positive constant being a period). The easiest example of such a function is a constant function. Show that the function $f: \mathbb{R} \rightarrow \mathbb{R}$ defined by

$$f(t) := \begin{cases} 1, & \text{for } t \in \mathbb{Q}, \\ 0, & \text{for } t \in \mathbb{R} \setminus \mathbb{Q} \end{cases}$$

is also periodic without having a prime period.

[**Hint:** In this exercise, we assume that the reader is familiar with the properties of rational numbers (\mathbb{Q}) and irrational numbers ($\mathbb{R} \setminus \mathbb{Q}$).]

Exercise 2.9. Sketch the graph of the quantization function $Q: \mathbb{R} \rightarrow \mathbb{R}$ defined by

$$Q(a) := \operatorname{sgn}(a) \cdot \Delta \cdot \left\lfloor \frac{|a|}{\Delta} + \frac{1}{2} \right\rfloor$$

for $a \in \mathbb{R}$ and some fixed quantization step size $\Delta > 0$ (see (2.33)). Furthermore, sketch the graph of the absolute quantization error.

Exercise 2.10. In mathematics, the term “operator” is used to denote a mapping from one vector space to another. Let V and W be two vector spaces over \mathbb{R} . An operator $M: V \rightarrow W$ is called **linear** if $M[a_1 v_1 + a_2 v_2] = a_1 M[v_1] + a_2 M[v_2]$ for any $v_1, v_2 \in V$ and $a_1, a_2 \in \mathbb{R}$. Show that $V := \{f: \mathbb{R} \rightarrow \mathbb{R}\}$ and $W := \{x: \mathbb{Z} \rightarrow \mathbb{R}\}$ are vector spaces. Fixing a sampling period $T > 0$, consider the operator M that maps a CT-signal $f \in V$ to the DT-signal $M[f] := x \in W$ obtained by T -sampling as defined in (2.32). Show that this defines a linear operator.

Exercise 2.11. Show that the quantization operator $Q: \mathbb{R} \rightarrow \mathbb{R}$ as defined in Exercise 2.9 and (2.33) is *not* a linear operator.

Exercise 2.12. In this exercise we discuss various computation rules for complex numbers and their conjugates. The complex multiplication is defined by $c_1 \cdot c_2 = a_1 a_2 - b_1 b_2 + i(a_1 b_2 + a_2 b_1)$ for two complex numbers $c_1 = a_1 + ib_1, c_2 = a_2 + ib_2 \in \mathbb{C}$ (see (2.34)). Furthermore, complex conjugation is defined by $\bar{c} = a - ib$ for a complex number $c = a + ib \in \mathbb{C}$ (see (2.35)). Finally, the absolute value of a complex number c is defined by $|c| = \sqrt{a^2 + b^2}$. Prove the following identities:

- (a) $\operatorname{Re}(c) = (c + \bar{c})/2$
- (b) $\operatorname{Im}(c) = (c - \bar{c})/(2i)$
- (c) $\overline{c_1 + c_2} = \bar{c}_1 + \bar{c}_2$
- (d) $\overline{c_1 \cdot c_2} = \bar{c}_1 \cdot \bar{c}_2$
- (e) $c\bar{c} = a^2 + b^2 = |c|^2$
- (f) $1/c = \bar{c}/(c\bar{c}) = \bar{c}/(a^2 + b^2) = \bar{c}/(|c|^2)$

Exercise 2.13. We have seen in Section 2.2.3.2 that the set $\mathbb{C}^{\mathbb{Z}} = \{x | x: \mathbb{Z} \rightarrow \mathbb{C}\}$ of complex-valued DT-signals defines a vector space. Show that the subset $\ell^2(\mathbb{Z}) \subset \mathbb{C}^{\mathbb{Z}}$ of DT-signals of finite energy is a linear subspace. To this end, you need to show that $x + y \in \ell^2(\mathbb{Z})$ and $ax \in \ell^2(\mathbb{Z})$ for any $x, y \in \ell^2(\mathbb{Z})$ and $a \in \mathbb{C}$.

Exercise 2.14. In Section 2.3.1, we defined the set $\{\mathbf{1}, \sin_k, \cos_k \mid k \in \mathbb{N}\} \subset L^2_{\mathbb{R}}([0, 1])$. Prove that this set is an orthonormal set in $L^2_{\mathbb{R}}([0, 1])$, i.e., that it satisfies (2.50) and (2.51).

[Hint: Use the following trigonometric identities:

- (a) $\cos(\alpha)^2 + \sin(\alpha)^2 = 1$
- (b) $\cos(\alpha)\cos(\beta) = (\cos(\alpha + \beta) + \cos(\alpha - \beta))/2$
- (c) $\sin(\alpha)\sin(\beta) = (\cos(\alpha - \beta) - \cos(\alpha + \beta))/2$
- (d) $\sin(\alpha)\cos(\beta) = (\sin(\alpha + \beta) + \sin(\alpha - \beta))/2$

To show (2.51), use (a) and the fact that \cos_k^2 and \sin_k^2 have the same area over a full period. The proof of (2.50) is a bit cumbersome, but not difficult when using (b), (c), and (d).]

Exercise 2.15. Let $\exp(i\gamma) := \cos(\gamma) + i\sin(\gamma)$, $\gamma \in \mathbb{R}$, be the complex exponential function as defined in (2.67). Prove the following properties (see (2.68) to (2.71)):

- (a) $\exp(i\gamma) = \exp(i(\gamma + 2\pi))$
- (b) $|\exp(i\gamma)| = 1$
- (c) $\overline{\exp(i\gamma)} = \exp(-i\gamma)$
- (d) $\exp(i(\gamma_1 + \gamma_2)) = \exp(i\gamma_1)\exp(i\gamma_2)$
- (e) $\frac{d\exp(i\gamma)}{d\gamma} = i\exp(i\gamma)$

[Hint: To prove (d), you need the trigonometric identities $\cos(\alpha + \beta) = \cos(\alpha)\cos(\beta) - \sin(\alpha)\sin(\beta)$ and $\sin(\alpha + \beta) = \cos(\alpha)\sin(\beta) + \sin(\alpha)\cos(\beta)$. In (e), note that the real (imaginary) part of a derivative of a complex-valued function is obtained by computing the derivative of the real (imaginary) part of the function.]

Exercise 2.16. In (2.77), we defined for each $k \in \mathbb{Z}$ the complex-valued exponential function $\mathbf{exp}_k: [0, 1] \rightarrow \mathbb{C}$ by $\mathbf{exp}_k(t) := \cos(2\pi kt) + i\sin(2\pi kt)$, $t \in \mathbb{R}$. As in Exercise 2.14, show that the set $\{\mathbf{exp}_k \mid k \in \mathbb{Z}\} \subset L^2([0, 1])$ is an orthonormal set, i.e., $\|\mathbf{exp}_k\|^2 = 1$ for $k \in \mathbb{Z}$ (see (2.51)) and $\langle \mathbf{exp}_k, \mathbf{exp}_\ell \rangle = 0$ for $k \neq \ell$, $k, \ell \in \mathbb{Z}$ (see (2.50)).

[Hint: Use the properties of the exponential function introduced in Exercise 2.15. Furthermore, note that the real (imaginary) part of an integral of a complex-valued function is obtained by integrating the real (imaginary) part of the function.]

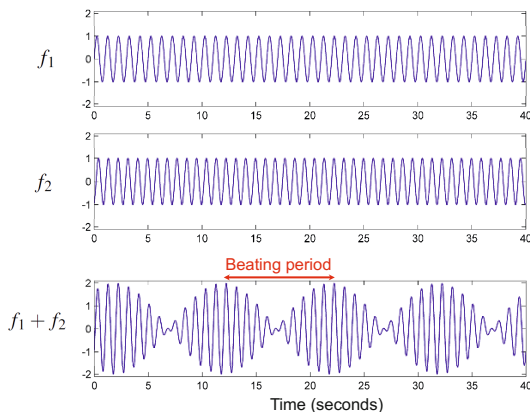
Exercise 2.17. Let atan2 be the function as defined in (2.76). For a complex number $c = a + ib \in \mathbb{C}$, we set $\text{atan2}(c) := \text{atan2}(b, a)$. Show that $\text{atan2}(\lambda \cdot c) = \text{atan2}(c)$ for any positive constant $\lambda \in \mathbb{R}_{>0}$. Furthermore, show that $\text{atan2}(\bar{c}) = -\text{atan2}(c)$.

[Hint: Use the fact that the arctan function is an odd function, i.e., $\arctan(-v) = -\arctan(v)$ for $v \in \mathbb{R}$.]

Exercise 2.18. In this exercise, we consider the geometric series for complex numbers, which is needed in (2.112). Prove that $\sum_{n=0}^{N-1} a^n = (1 - a^N)/(1 - a)$ for any complex number $a \neq 1$.

[Hint: For the proof, use mathematical induction on N .]

Exercise 2.19. We have seen that two sinusoids of similar frequency may add up (constructive interference) or cancel out (destructive interference); see Figure 2.19. Let $f_1(t) = \sin(2\pi\omega_1 t)$ and $f_2(t) = \sin(2\pi\omega_2 t)$ be two such sinusoids with distinct but nearby frequencies $\omega_1 \approx \omega_2$. In the following figure, for example, $\omega_1 = 1$ and $\omega_2 = 1.1$ is used.



The figure also shows that the superposition $f_1 + f_2$ of these two sinusoids results in a function that looks like a single sine wave with a slowly varying amplitude, a phenomenon also known as *beating*. Determine the rate (reciprocal of the period) of the beating in dependency on ω_1 and ω_2 . Compare this result with the plot of $f_1 + f_2$ in the figure.

[Hint: Use the trigonometric identity $\sin(\alpha) + \sin(\beta) = 2 \cos\left(\frac{\alpha-\beta}{2}\right) \sin\left(\frac{\alpha+\beta}{2}\right)$ for $\alpha, \beta \in \mathbb{R}$.]

Exercise 2.20. Let $f \in L^2(\mathbb{R})$ be a signal of unit energy $\|f\|^2 = 1$. Show that the scaled signal g defined by $g(t) := s^{1/2}f(s \cdot t)$ also has unit energy for a positive real scaling factor $s > 0$. Furthermore show that $\hat{g}(\omega) = s^{-1/2}\hat{f}(\omega/s)$ for $\omega \in \mathbb{R}$. Discuss this result. Describe how one can obtain a Dirac sequence by changing the parameter s (see Section 2.3.3.2).

Exercise 2.21. Show that the Fourier transform of the rectangular function in (2.95) is the sinc function in (2.96). Also prove that the sinc function is continuous at $t = 0$.

[Hint: Use the fact that the derivative of $t \mapsto \exp(-2\pi i \omega t)$ is given by $t \mapsto -2\pi i \omega \exp(-2\pi i \omega t)$; see Exercise 2.15. From this, one can derive the indefinite integral of the exponential function. To prove the continuity at $t = 0$, look at the first terms of the Taylor series of the sine function.]

Exercise 2.22. For a signal $f \in L^2(\mathbb{R})$, consider the translation f_{t_0} defined by $f_{t_0}(t) := f(t - t_0)$ for $t \in \mathbb{R}$ (see (2.97)) and the modulation f^{ω_0} defined by $f^{\omega_0}(t) := \exp(2\pi i \omega_0 t) f(t)$ for $t \in \mathbb{R}$ (see (2.98)). Show that $\|f\| = \|f_{t_0}\| = \|f^{\omega_0}\|$. Furthermore, prove the properties (2.99) and (2.100):

$$\widehat{f_{t_0}}(\omega) = \exp(-2\pi i \omega t_0) \hat{f}(\omega) \quad \text{and} \quad \widehat{f^{\omega_0}}(\omega) = \hat{f}(\omega + \omega_0)$$

for $\omega \in \mathbb{R}$.

Exercise 2.23. Any complex number $c \in \mathbb{C}$ with $c^N = 1$ for a given $N \in \mathbb{N}$ is called an N^{th} **root of unity**. If in addition $c^k \neq 1$ for $1 < k < N$, the root c is called **primitive**. Show that $\rho_N := \exp(-2\pi i/N)$ defines a primitive N^{th} root of unity. Furthermore, describe *all* N^{th} roots of unity. Which of these roots are primitive? Determine for $N \in \{4, 7, 12\}$ all primitive N^{th} roots of unity.

[Hint: In this exercise, one needs to know that a (nonzero) polynomial of degree N has at most N different roots, where a **root** of a function is an input value that produces an output of zero.]

Exercise 2.24. Let $\mathbf{x} = (x(0), \dots, x(N-1))^{\top}$ be a real-valued vector consisting of samples $x(n) \in \mathbb{R}$ for $n \in [0 : N-1]$. Show that

$$\mathbf{X} = \text{DFT}_N \cdot \mathbf{x}$$

with $\mathbf{X} = (X(0), \dots, X(N-1))^{\top}$ fulfills the symmetry property $X(k) = \overline{X(N-k)}$ for all $k \in [1 : N-1]$ and $X(0) \in \mathbb{R}$. This shows that the upper half of the frequency coefficients are redundant

if \mathbf{x} is real-valued. Furthermore, show the converse. Given a spectral vector \mathbf{X} with $X(0) \in \mathbb{R}$ and $X(k) = \overline{X(N-k)}$ for all $k \in [1 : N-1]$, then

$$\mathbf{x} = \text{DFT}_N^{-1} \cdot \mathbf{X}$$

is a real-valued vector (see (2.118)).

[Hint: Use the computation rules for complex numbers from Exercise 2.12.]

Exercise 2.25. Specify the DFT_N matrix explicitly for $N \in \{1, 2, 4\}$. Count the number of multiplications and additions when performing the usual matrix–vector product $\text{DFT}_4 \cdot \mathbf{x}$ for a vector $\mathbf{x} = (x_1, x_2, x_3, x_4)^\top$. Then conduct all steps of the FFT algorithm (two recursions are needed) and again count the overall number of multiplications and additions needed to compute $\text{DFT}_4 \cdot \mathbf{x}$.

Exercise 2.26. Let $N = 2^n$ be a power of two. In (2.127), we derived the estimate $\mu(N) \leq 2\mu(N/2) + 1.5N$ for the number of multiplications and additions needed to compute the matrix–vector product $\text{DFT}_N \cdot \mathbf{x}$. Using $\mu(1) = 0$ (the case $n = 0$), show by a mathematical induction on n that this implies $\mu(N) \leq 1.5N \log_2(N)$.

Exercise 2.27. In the spectrograms shown in Figure 2.32 one can notice vertical stripes at $t = 0$ and $t = 1$. Why?

Exercise 2.28. In this exercise, we prove the **sampling theorem**. A CT-signal $f \in L^2(\mathbb{R})$ is called **Ω -bandlimited** if the Fourier transform \hat{f} vanishes for $|\omega| > \Omega$, i.e., $\hat{f}(\omega) = 0$ for $|\omega| > \Omega$. Let $f \in L^2(\mathbb{R})$ be an Ω -bandlimited function and let x be the T -sampled version of f with $T := 1/(2\Omega)$, i.e., $x(n) = f(nT)$, $n \in \mathbb{Z}$. Then f can be reconstructed from x by

$$f(t) = \sum_{n \in \mathbb{Z}} x(n) \text{sinc}\left(\frac{t-nT}{T}\right) = \sum_{n \in \mathbb{Z}} f\left(\frac{n}{2\Omega}\right) \text{sinc}(2\Omega t - n),$$

where the sinc function is defined in (2.96). In other words, the CT-signal f can be perfectly reconstructed from the DT-signal obtained by equidistant sampling if the bandlimit is no greater than half the sampling rate.

[Hint: Note that one may assume $\Omega = 1/2$ (and $T = 1$) by considering the scaled function $t \mapsto f(t/\Omega)$. In this case, f is $1/2$ -bandlimited and can be extended to a 1-periodic function g . Represent g by its Fourier series (2.79) and compute the Fourier coefficients $c_n = \langle g | \exp_n \rangle$, $n \in \mathbb{Z}$. Compare these coefficients with the Fourier representation (2.91) of f evaluated at $t = n$ for $n \in \mathbb{Z}$ (again using the fact that f is $1/2$ -bandlimited). As a result, one obtains $c_n = f(-n)$. Finally, reconstruct f from the Fourier series of g . To this end, you need the result of Exercise 2.21.]

Fundamentals of Music Processing

Audio, Analysis, Algorithms, Applications

Müller, M.

2015, XXIX, 487 p. 249 illus., 30 illus. in color.,

Hardcover

ISBN: 978-3-319-21944-8