

Chapter 1

Introduction

We live in an analog world, but we would like our digital computers to interact with it. Indeed, digital signal processing (DSP) has become pervasive. It is the basis for most modern consumer electronics, medical imaging devices, cell phones, internet protocol telephony, multimedia standards, speech processing, and a myriad of other products. Digital algorithms, implemented with microprocessors, are less pricey, easier to control, more robust, and more flexible than their analog counterparts, so that nowadays analog circuits are often replaced by digital chips. Digital data is also far easier to store, transmit, and manipulate than analog data. Therefore, in modern applications, an increasing number of functions are being pushed forward to sophisticated software algorithms, leaving only delicate finely tuned tasks for the circuit level. Nowadays, it feels natural that a media player shows our favorite movie, or that our surround system synthesizes pure acoustics, as if sitting in the orchestra, and not in the living room. The digital world plays a fundamental role in our everyday routine, to such a point that we almost forget that we cannot “hear” or “watch” these streams of bits, running behind the scenes. The world around us is analog, yet most modern man-made means for exchanging information are digital. “I am an analog girl in a digital world,” sings Judy Gorman [One Sky, 1998], capturing the essence of the digital revolution.

Whether recording sounds, capturing images, or processing an electromagnetic wave, many sources of information are of analog or continuous-time nature. Therefore, DSP inherently relies on a sampling mechanism which converts continuous signals to discrete sequences of numbers, while preserving the information present in those signals. This conversion is performed using a device known as an *analog-to-digital converter (ADC)*. ADC devices translate physical information into a stream of numbers, enabling digital processing by sophisticated software algorithms. After processing, the samples are converted back to the analog domain via a *digital-to-analog converter (DAC)*. Consequently, sampling theories lie at the heart of DSP and play a major role in enabling the digital revolution.

The ADC task is inherently intricate: its hardware must hold a snapshot of a fast-varying input signal steady, while acquiring measurements. Since these measurements are spaced in time, the values between consecutive snapshots are lost. In general, therefore, there is no way to recover the analog input unless some prior information on its structure is incorporated.

1.1 Standard sampling

The simplest way to record an analog signal $x(t)$ is to sample its values $x(nT)$ at intervals of length T , as depicted in Fig. 1.1(a). This type of sampling is referred to as *pointwise* sampling. We use the block diagram in Fig. 1.1(b) to illustrate this operation.

Given the samples, an approximation of $x(t)$ can be obtained by using an appropriate interpolating function, which we denote by $w(t)$. Figure 1.2 demonstrates several possible interpolations of the samples in Fig. 1.1 using different functions $w(t)$: zero-order-hold, linear interpolation, cubic spline interpolation (third-order polynomial), and sinc interpolation. In each case recovery is obtained by modulating the chosen $w(t)$ by the sample values:

$$\hat{x}(t) = \sum_{n \in \mathbb{Z}} d[n]w(t - nT), \tag{1.1}$$

where in our setting $d[n] = x(nT)$ are the given samples and T is the sampling period. More generally, we can choose $d[n]$ to be a function of the samples $x(nT)$,

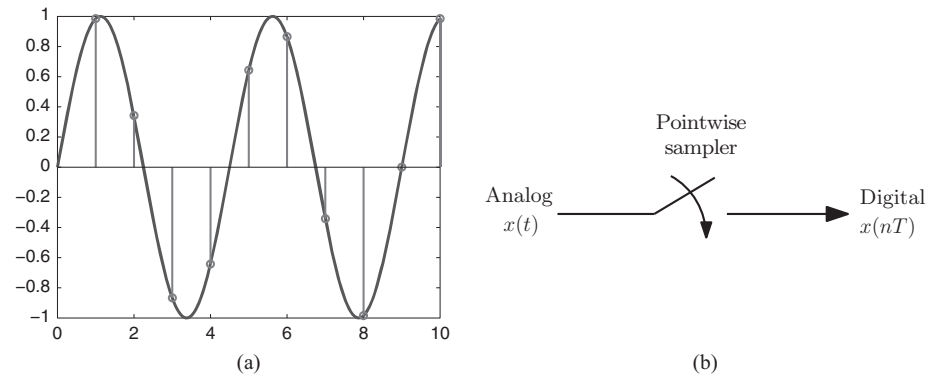


Figure 1.1 Pointwise sampling. (a) A continuous-time signal $x(t) = \sin(4\pi t/9)$ and its pointwise samples with period $T = 1$. (b) Block diagram of a pointwise sampler.

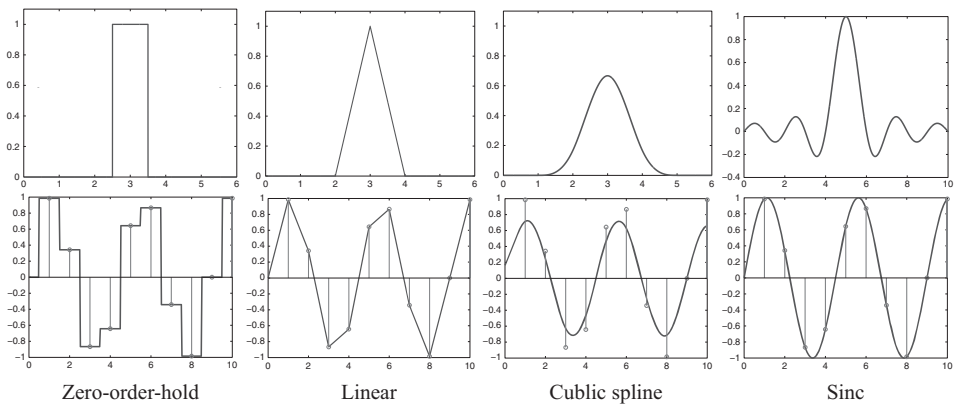


Figure 1.2 Signal reconstruction by various interpolation functions. Top: Interpolating function $w(t)$. Bottom: Recovery $\hat{x}(t)$.

designed to optimize the recovery process. In this case we refer to $d[n]$ as the *corrected samples*.

Clearly, the approximation quality depends on the interpolator chosen and on how well it matches the properties of the original input, and on the samples. Thus, a key component in any sampling theory is the information we have about our signal. Without incorporating prior knowledge, the problem of recovery from samples is ill-posed; there are always many curves that can pass through a set of points, as illustrated in Fig. 1.2. A challenge in practice is to find the “best” curve in some sense consistent with our prior information. Consequently, a large part of sampling theory deals with methods for optimizing $d[n]$ and for selecting $w(t)$ based on the input properties. An additional important design consideration is how small T has to be in order to ensure perfect recovery for certain classes of signals.

While here we have treated pointwise sampling, other more elaborate methods of sampling exist which we will discuss throughout the text. A straightforward generalization of pointwise sampling is depicted in Fig. 1.3, in which $x(t)$ is first filtered with a sampling filter $s(-t)$ that allows us to incorporate imperfections in the ideal sampler. The output is then pointwise sampled on a uniform grid leading to *generalized samples*, denoted by $c[n]$. We consider such sampling mechanisms in detail throughout the book. In particular, we will discuss methods for optimizing the sampling filter $s(t)$ based on the input properties.

Undoubtedly, the most-studied sampling theorem that has had a major influence on signal processing is the well-known Shannon–Nyquist theorem. This theorem was introduced formally into the information theory community by Shannon in [1], but Nyquist had already brought it to the attention of communication engineers in [2]. Kotelnikov is credited with introducing the theorem in the Russian literature [3]. In mathematics, the theorem was developed as part of the study of cardinal series in the works of E. T. Whittaker and his son J. M. Whittaker [4, 5]. The basic idea behind the bandlimited sampling theorem has also been attributed to Cauchy [6], who stated the essential result although without proof. A very illuminating review of the history and mathematics surrounding this theorem can be found in [7].

The Shannon–Nyquist theorem has become a landmark in both the mathematical and engineering literature and has had one of the most profound impacts on industrial development of DSP systems. It provides a method to compute $x(t)$ exactly from its pointwise samples, as long as the signal is sufficiently smooth. More precisely, to allow perfect recovery in (1.1), the sampling frequency, $1/T$, must be at least twice the highest frequency in the signal $x(t)$. This minimal rate is referred to as the *Nyquist rate*. The signal

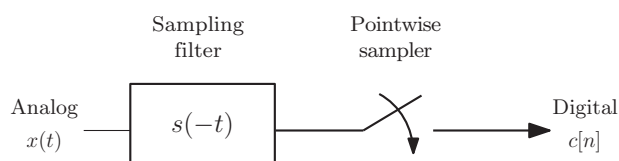


Figure 1.3 Generalized sampling. The signal $x(t)$ is first filtered with a sampling filter $s(-t)$ and then sampled by a pointwise sampler.

is then interpolated from its samples using shifts of the sinc function $w(t) = \text{sinc}(t/T)$, with $\text{sinc}(t) = \sin(\pi t)/(\pi t)$. This theorem assumes that the class of input functions is bandlimited to an appropriate frequency. The interpolating function $w(t)$ in (1.1) is then also a bandlimited function. Capitalizing on this result, much of signal processing has moved from the analog to the digital domain as it allows a continuous-time bandlimited function to be replaced by a discrete set of its samples without any loss of information.

To accommodate high operating rates while retaining low computational cost, efficient ADCs must be developed. While the Shannon–Nyquist theorem is extremely elegant and has had a major impact on DSP, it has several drawbacks. Unfortunately, real-world signals are rarely truly bandlimited. Even signals which are approximately bandlimited may have to be sampled at a fairly high Nyquist rate, requiring expensive sampling hardware and high-throughput digital machinery. Many natural signals, even if they are bandlimited, are often better represented (using fewer coefficients) in bases other than the Fourier basis. Bandlimiting also tends to introduce Gibbs oscillations which can be visually disturbing, for example in images. Finally, many classes of signals possess further structure that can be exploited in order to reduce sampling rates. Classical sampling theory, however, necessitates a high sampling rate whenever a signal has large bandwidth, even if the actual information content in the signal is small. For example, a piecewise linear signal is nondifferentiable; it is therefore not bandlimited, and moreover, its Fourier transform decays at a fairly slow rate. Nonetheless, such a signal is completely described by the location of its knots (transitions between linear segments) and the signal values at those positions, which can be far fewer parameters than the number of samples required by the Shannon–Nyquist theorem. It would therefore be more efficient to have a variety of sampling techniques, tailored to different signal models, such as bandlimited or piecewise linear signals. Such an approach echoes the fundamental quest of the recent field of compressive sampling, which is to capture only the essential information embedded in a signal.

Two other difficulties with the Shannon–Nyquist theorem are the assumptions of ideal pointwise sampling, and of sinc interpolation. Practical ADCs are usually not ideal, that is, they do not produce the exact signal values at the sampling locations. A common situation is that the ADC integrates the signal, usually over small neighborhoods surrounding the sampling points. Moreover, nonlinear distortions are often introduced during the sampling process. These various distortions need to be accounted for in the reconstruction. In addition, implementing the infinite sinc interpolating kernel as part of the DAC required by the Shannon–Nyquist theorem is difficult, since it has slow decay. In practice, much simpler kernels are used, such as linear interpolation. Another major obstacle in the context of modern imaging and communication systems is that signals today can be modulated up to several gigahertz (GHz), while standard ADCs have difficulty in accommodating such large analog bandwidths.

Therefore, to design sampling and interpolation methods that are adapted to practical scenarios, there are several issues that need to be properly addressed:

1. The sampling mechanism should be adequately modeled;
2. Relevant prior knowledge about the class of input signals should be taken into account;

3. Limitations should be imposed on the reconstruction algorithm in order to ensure robust and efficient recovery.

Throughout the book we treat each of these three essential components of the sampling scheme. For each of the elements we focus on several models, which commonly arise in signal processing and communication systems.

1.2 Beyond bandlimited signals

Following the introduction of the bandlimited sampling theorem by Shannon [1], sampling theory became an active area of research, reaching quite a mature state by the 1980s. Several thorough and beautiful tutorials were written on the topic around that time [8, 9]. At that point, research in the area of sampling became quite mathematical with less immediate impact on signal processing and communication applications, or on the actual design of ADCs. In the 1990s, sampling theory benefited from a surge of research due to the intense interest in wavelet theory and the connections made between the two fields. A lot of the theory developed for wavelet analysis is immediately applicable to sampling theorems. This led to many interesting interpretations of existing results along with new methods for sampling and processing signals that move away from the bandlimited paradigm, and instead consider more general signal models and sampling devices. An excellent summary of this perspective can be found in [10] (see also [11]).

In the past few years sampling has again been revived, this time by the vast interest in the area of compressed sensing (CS) [12, 13, 14] which suggests methods for reducing the number of measurements needed to represent sparse signals, or signals with certain types of structure. This framework has focused primarily on sampling of discrete-time signals and reconstruction techniques from a finite number of samples. Works in this area have shown that a high-dimensional vector with only a few nonzero elements is recoverable from a properly chosen underdetermined set of equations. Recovery can be obtained using a variety of different polynomial-time algorithms under appropriate conditions. These results suggest that sparse signals may be sampled at sub-Nyquist rates, which is crucial in modern communications settings.

The CS framework is mostly focused on discrete and finite settings, while sampling inherently deals with continuous-time signals. There are of course examples of analog signals that naturally possess finite representations, such as trigonometric polynomials. However, extending the ideas of CS to acquisition of more general continuous-time signals using practical hardware devices remains a difficult challenge despite the widespread literature in this area. Nonetheless, we will see that by combining analog sampling results with ideas from CS, a variety of efficient sub-Nyquist systems can be developed, leading to low-rate sampling of a broad set of analog signals. In addition, we show that often the acquired samples can be directly processed without having to interpolate them back to the high Nyquist grid, resulting in low-rate processing as well. Beyond developing the fundamental theory, we also discuss practical aspects of reduced-rate sampling and demonstrate prototype sub-Nyquist hardware realizations for a variety

of applications. These devices allow practical sampling and processing of many classes of signals at sub-Nyquist rates.

The key to developing low-rate analog sensing methods is relying on structure in the input. Signal processing algorithms have a long history of leveraging structure for various tasks. As an example, MUSIC [15] and ESPRIT [16] are popular techniques for spectrum estimation that exploit signal structure. Model-order selection methods in estimation [17], parametric estimation and parametric feature detection [18] are further examples where structure is heavily used. In our context, we are interested in utilizing signal models in order to reduce sampling rate. Classic approaches to sub-Nyquist sampling include carrier demodulation [19] and bandpass undersampling [20], which assume a linear model corresponding to a bandlimited input with predefined frequency support and fixed carrier frequencies. In the spirit of CS, where unknown nonzero locations result in a nonlinear model, we extend these classical results to analog inputs with unknown frequency support, as well as more broadly to scenarios that involve nonlinear input structures. The approach we take in this book follows the recently proposed Xampling framework [21, 22], which treats a nonlinear model of union of subspaces. In this structure, the input signal belongs to a single subspace out of multiple, possibly even infinitely many, candidate subspaces. The exact subspace to which the signal belongs is unknown a priori. This model encompasses a large variety of structured analog signals and paves the way to the development of practical sub-Nyquist sampling systems.

The importance of sampling theory is likely to continue to grow with the ongoing demand for more sophisticated and efficient DSP systems. The connection with CS offers yet another new perspective on sampling, and ways to better exploit the signal degrees of freedom. This relationship has also brought to the surface the need not only to develop sound mathematical frameworks but also to tie them to concrete hardware implementations so that the benefits predicted by the theory, such as reduced sampling rates, can be met in practice and have an impact on the ADC market.

1.3 Outline and outlook

In this book we consider many extensions of the Shannon–Nyquist theorem, which treat a wide class of input signals as well as nonideal sampling and nonlinear distortions. Our exposition is based on a Hilbert-space interpretation of sampling techniques, where the aim is to develop more traditional sampling theories, along with modern techniques emerging from the field of CS, in a unified framework. The roots of this framework can be found in the fundamentals of linear algebra, and rely on familiar engineering building blocks such as filters and Fourier analysis. This unification has provided new understandings of classical interpolation methods, and has set the stage for new and exciting frontiers.

The framework we consider is based on viewing sampling in a broader sense of projection onto appropriate subspaces, and then choosing the subspaces to yield interesting new possibilities. For example, the results we present can be used to uniformly sample nonbandlimited signals, and to compensate perfectly for nonlinear effects.

Chapter 2 is therefore dedicated to a detailed exposition of the linear algebra basics needed to derive our extended sampling framework, followed by a brief summary of relevant Fourier analysis tools in Chapter 3. The more recent concepts of sub-Nyquist sampling, or extensions of CS to the analog setting, require more background on CS which we will provide in Chapter 11. Here again we retain the subspace approach by viewing these problems within the broader framework of a union of subspaces.

In order to develop ADCs for a particular problem, we must have accurate models for the signals of interest. We devote Chapters 4 and 5 to a detailed exposition of the signal models we will be focusing on throughout the book, along with some of the fundamental mathematical properties associated with such signal sets. Much of classical signal processing is based on the notion that signals can be modeled as vectors living in an appropriate subspace. Chapter 6 is focused on sampling theorems for signals confined to an arbitrary subspace in the presence of possibly nonideal sampling. The methods we develop can also be used to reconstruct a signal using a given interpolation kernel that is easy to implement, with often only a minor loss in signal quality with respect to the optimal kernel matched to the input subspace properties. In Chapter 8 we extend this basic framework to include nonlinear distortions in the sampling process. Surprisingly, many types of nonlinearities that are encountered in practice do not pose any technical difficulty and can be completely compensated for despite their effect of bandwidth increase, without requiring higher sampling rates.

A more general and less restrictive formulation of the sampling problem is considered in Chapter 7 in which our prior knowledge on the signal is that it is smooth in some sense. Unlike subspace priors, a one-to-one correspondence between smooth signals and their sampled version does not exist since smoothness is a far less restrictive constraint than confining the signal to a subspace. Perfect recovery is therefore generally impossible. Instead, we focus on approximating the input as well as possible under several different design objectives. These concepts can also be used to develop effective rate conversion techniques between digital formats, as we discuss in Chapter 9.

Although linear models are very popular in sampling theory, and more generally in DSP, such simple models often fail to capture much of the structure present in many common classes of signals. For example, while it may be reasonable to model signals as vectors, in many cases not all possible vectors in the space represent valid signals. In response to these challenges, there has been a surge of interest in recent years, across many fields, in a variety of *low-dimensional signal models* that quantify the notion that the number of degrees of freedom in high-dimensional signals is often quite small compared with their ambient dimension. One path to developing a framework for sampling and processing of such signals is by using the *union of subspaces model* which is introduced more formally in Chapter 10. Probably the most well-studied example of a union of subspaces is that of a vector \mathbf{x} that is sparse in an appropriate basis. This model underlies the rapidly growing field of CS, which has attracted considerable attention in signal processing, statistics, and computer science, as well as the broader scientific community. A review of the essential CS concepts is provided in Chapter 11. In Chapters 12–15 we study how the fundamentals of CS can be expanded and extended to include richer structures in both analog and discrete-time signals, ultimately leading to sub-Nyquist

sampling techniques for a broad class of continuous-time signals. A more detailed outline of the book chapters can be found in the Preface.

The need and importance of sub-Nyquist techniques stems from the phenomenal success of DSP, thanks in large part to the Shannon–Nyquist theorem. This has spurred the digital revolution that is driving the development and deployment of new kinds of sensing systems with ever-increasing fidelity and resolution. As a result of this success, the amount of data generated by sensing systems has grown substantially. Unfortunately, in many important and emerging applications, the resulting sampling rate is so high that we end up with far too many samples that need to be transmitted, stored, and processed. In addition, in applications involving very wideband inputs it is often very costly, and sometimes even physically impossible, to build devices capable of acquiring samples at the necessary rate. Thus, despite extraordinary advances in sampling theory as well as computational power, the acquisition and processing of signals in application areas such as radar, wideband communications, imaging, video, medical imaging, remote surveillance, spectroscopy, and genomic data analysis continue to pose a tremendous challenge. Today, we are witnessing the outset of an interesting trend. Advances in related fields, such as wideband communication and radio-frequency technology, open a considerable gap with ADC devices. Conversion speeds which are twice the signal’s maximal frequency component have become more and more difficult to obtain. Consequently, alternatives to high-rate sampling are drawing considerable attention in both academia and industry.

Over the years, theory and practice in the field of sampling have developed in parallel routes. Contributions by many research groups have suggested a multitude of methods, other than uniform sampling, to acquire analog signals. The math has deepened, leading to abstract signal spaces and innovative sampling techniques, with the ability to treat a large class of input signals far beyond the standard bandlimited model associated with the Shannon–Nyquist theorem. At the same time, the market has adhered to the Nyquist paradigm; the footprints of Shannon–Nyquist are evident whenever conversion to digital takes place in commercial applications.

Throughout the book, wherever possible, we try to put an emphasis on practical aspects of sampling beyond the fundamental theory. Our goal is to bridge theory and practice, and to try to highlight where advances in sampling theory can have and already have had an impact on ADC design and on applications. This is particularly relevant in the second half of the book, which is targeted at solving a practical problem: reducing sampling and processing rates which are too costly in many modern applications. The exposition is aimed at trying to pinpoint the potential of sub-Nyquist strategies to emerge from the math to the hardware. In this spirit, we integrate contemporary theoretical viewpoints, which study signal modeling in a union of subspaces, together with a taste of practical aspects, including basic circuit design features. Our hope is that this combination of theory and practice will serve to further promote both academic and industrial advances in sampling theory.

As a final note before beginning our theoretical journey into linear algebra: We have very much enjoyed gathering and presenting the ideas in a unified way. We hope that the reader will share some of our enthusiasm for the material!

Chapter 2

Introduction to linear algebra

The process of sampling and reconstruction can be viewed as an expansion of a signal onto a set of vectors that span the space. Suppose we have a signal x that is defined on some domain, and has a series representation there of the form

$$x = \sum_n a[n]x_n, \quad (2.1)$$

where $\{a[n]\}$ is a countable set of coefficients, which depend on the input signal x , and $\{x_n\}$ is a fixed set of signals (or vectors). The expression in (2.1) implies that x is completely specified in terms of the coefficients $\{a[n]\}$, which we may think of as samples of x . We can therefore interpret (2.1) as the statement that x can be reconstructed from the samples $\{a[n]\}$ using the known vectors $\{x_n\}$. Series of the form (2.1), their generalizations and extensions, are the subject of this book.

When we consider sampling, or signal expansions, we need to clearly identify the class of possible inputs x , the expansion vectors $\{x_n\}$, and the relationship between the samples $\{a[n]\}$ and the original signal x . In this chapter we describe the mathematical machinery needed to explain series of the form (2.1). In particular we introduce vector spaces and Hilbert spaces which provide the setting for describing the class of input signals x and the domain in which the expansions take place. Some important concepts we will consider in detail are the linear transformation, its adjoint and the subspaces associated with it, projection operators, and the pseudoinverse, which are all essential for computing the representation coefficients. We also define bases in Hilbert spaces and focus on stable expansions which leads to the notion of a Riesz basis. At the end of the chapter we briefly discuss overcomplete representations and frames.

2.1 Signal expansions: some examples

Before delving into the mathematical notions underlying sampling theory, we begin by studying several simple examples that highlight some of the issues that arise when considering signal expansions.

One of the main concepts central to sampling theory is that prior knowledge regarding the signal structure is essential in order to enable recovery from a given set of samples. As we explained in more detail in the Introduction, the process of sampling reduces the continuous-time signal to a countable set of coefficients, so that without any prior

knowledge it is impossible to recover the full degrees of freedom describing the signal. To compensate for this dimensionality reduction, we must exploit knowledge regarding the signal structure. The following examples show how such information is incorporated into the recovery process. In all of the examples below we assume that the prior knowledge takes on the form of a subspace prior. The mathematics associated with such settings will be studied in much greater detail in Chapter 6.

Example 2.1 Suppose we are given two values $x(0)$ and $x(1)$ of a linear function $x(t) = at + b$, where a and b are unknown. Our goal is to evaluate $x(t)$ for any time t . Since the structure of $x(t)$ is known, to accomplish this goal all that is needed is to determine a and b . Therefore, our problem becomes that of computing a and b from the given samples.

It is easy to see that $b = x(0)$, and $a = x(1) - x(0)$. Therefore, for all t ,

$$x(t) = x(0)(1 - t) + x(1)t = \sum_{n=0}^1 a[n]x_n(t) \quad (2.2)$$

where $a[n] = x(n)$, and the expansion vectors are $x_0(t) = 1 - t$ and $x_1(t) = t$. Thus $x(t)$ can be represented by its samples $x(0)$ and $x(1)$. This example can be easily extended to allow recovery of a piecewise linear signal over the real line.

Although the previous example is very simple and almost trivial, it highlights some important features shared by many sampling theorems. First, we note that any function $x(t)$ from the given class of signals, in our case polynomials of degree 1, can be represented in the form (2.2) regardless of the values of a and b . For more general expansions, this corresponds to the statement that the expansion vectors $\{x_n\}$ are independent of the input x . Second, the ability to reconstruct the signal depends on our prior knowledge. In the example, we were able to recover $x(t)$ from its samples $x(0)$ and $x(1)$ since we knew that $x(t)$ is a polynomial of degree 1. On the other hand, note that the right-hand side of (2.2) always specifies a polynomial of degree 1; therefore, if $x(t)$ does not have this form, then it cannot be reconstructed using (2.2).

In Example 2.1 there were a finite number of expansion coefficients over each interval. Similar expansions are possible when there are infinitely many coefficients, as we show in the next example. The function we consider in the example below belongs to the well-known class of bandlimited signals, which leads to the Shannon–Nyquist theorem. We will revisit this example with more mathematical rigor in Chapter 4.

Example 2.2 Consider a signal $x(t)$ bandlimited to the frequency π/T . The famous Shannon–Nyquist theorem states that such a signal can be reconstructed from its samples $x(nT)$ using the expansion

$$x(t) = \sum_{n \in \mathbb{Z}} x(nT) \frac{\sin(\pi(t - nT)/T)}{\pi(t - nT)/T}. \quad (2.3)$$