01 Signals and signal processing

The course involves the processing of signals and images through digital (numeric) methods, specifically, Digital Signal Processing (DSP).

A signal is a function of independent variables, such as time, distance, or position. For instance, voice and music signals represent air pressure as a function of time in a specific position in space. In the case of a black-and-white image, it represents light intensity as a function of two spatial coordinates. A TV video signal comprises a sequence of images, referred to as frames, and is a function of three variables: two spatial coordinates and time.

Most signals we encounter are generated by natural means (e.g., voice, music, images, seismic waves), but there are signals that are produced synthetically or generated by computers (e.g., musical synthesis, ...).

Signals carry information, and the purpose of digital signal processing is to extract the useful information carried by a signal.

Digital signal processing deals with the mathematical representation of signals, and the operations, algorithms, applied to signals in order to extract the information they carry. We will see that the signal representation can be done by basic functions in the original independent variable domain (the time domain), or can be performed by means of basic functions in a transformed domain. Similarly, digital signal processing can be carried out in the original independent variable domain (in the time domain) or in a transformed domain.

In our lectures, we will study both the representation and processing of signals using digital means.

01.01 Characterization and classification of signals

Signals can be classified based on the nature of the independent variables or the values of the function that defines the signal.

A signal can be *mono-dimensional* or *multi-dimensional* (1-D, 2-D, 3-D, ...). A mono-dimensional signal is a function of a single independent variable, while an M-dimensional signal is a function of M independent variables. For example, voice is a 1-D signal, a function of the independent variable 'time.' An image is a 2-D signal, a function of two spatial coordinates. A video is a 3-D signal, a function of three independent variables: two spatial coordinates and time.

A signal can also be classified as *mono-channel* or *multi-channel*. A mono-channel signal is a scalar function of the independent variable, whereas a multi-channel signal is a vector function of the independent variable. For example, a grayscale image is a scalar function of two spatial coordinates, represented as $I(x, y)$, while a color image is a vector function (with three components) of the spatial coordinates:

$$\begin{bmatrix} R(x,y) \\ G(x,y) \\ B(x,y) \end{bmatrix},$$

where R, G, and B represent the intensity of the three primary colors: red, green, and blue.

In the first part of the course, we will study only mono-dimensional and mono-channel signals.

Although the independent variable isn't necessarily time, it is a common practice to refer to it as *time* for these signals.

The signal value at a specific time instant is called *amplitude*.

The variation of amplitude as a function of the independent variable, i.e., as a function of time, is called *waveform*.

If the independent variable is continuous ($t \in \mathbb{R}$), the signal is called a *continuous-time signal*. If the independent variable is discrete ($t \in \mathbb{N}$ or $t \in \mathbb{Z}$), the signal is called a *discrete-time signal*.

A continuous-time signal is defined for each time instant. On the contrary, a discrete-time signal is defined only for certain time instants, while it does not exist between these time instants. A discrete time signal is a sequence of numbers.

Similarly, a signal can have a *continuous amplitude* or a *discrete amplitude*. A continuous amplitude signal can assume an infinite number of amplitude values (in $\mathbb{R}$ or $\mathbb{C}$). On the contrary, a signal with discrete amplitude can assume only a finite number of values (which can be coded with a finite number of binary digits).

A continuous-time signal with continuous amplitude is called an *analog signal*. Many natural signals are analog signals (e.g., voice, music, seismic waves).

A discrete-time signal with discrete amplitude is called a *digital signal*. Digital signals can be generated artificially, but there also exist natural digital signals (for example, the number of sunspots in one year or the number of cars passing on a road in the last month). Moreover, digital signals can be generated from analog signals through the following operations:

- sampling the signal,
- quantizing the samples,
- coding the quantized samples.

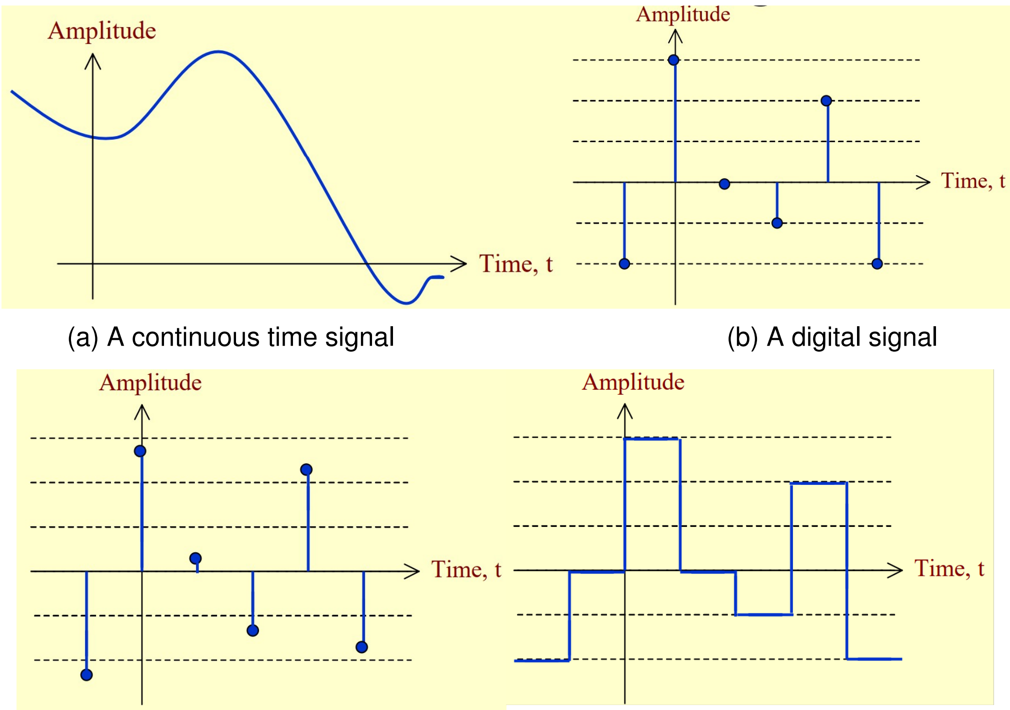
For example, the music stored in an audio CD is generated with this procedure.

A discrete-time signal with continuous amplitude is often referred to as a *sample-data signal*.

A continuous-time signal with discrete amplitude is commonly known as a *quantized boxcar signal*.

A signal can be classified as a *real signal* or a *complex signal* based on the amplitude of the signal, whether it belongs to the set of real numbers ($\mathbb{R}$) or the set of complex numbers ($\mathbb{C}$).

In the following, we will denote the independent variable of a continuous-time 1-D signal as t, and the independent variable of a discrete-time signal as n. We will treat continuous-time signals as continuous functions and discrete-time signals as sequences (of numbers). For example, $u_a(t)$ represents a 1-D continuous-time signal, and $\{v(n)\}$ represents a 1-D discrete-time signal. Each element $v(n)$ of the sequence $\{v(n)\}$ is called a *sample*. In many applications, a discrete-time signal is generated from a



(a) A continuous time signal (b) A digital signal

(c) A sample-data signal (d) A quantized boxcart signal

Figure 01.01: (From S. K. Mitra, "Digital signal processing: a computer based approach", McGraw Hill, 2011)

continuous-time signal by sampling the signal at uniformly spaced intervals T:

$$u_a(t) \to u_a(nT).$$

We will often represent the sampled signal $u_a(nT)$ simply as $u(n)$.

A final classification of signals arises from the way we know and treat them. We define a signal as *deterministic* when it is uniquely described by a mathematical expression, a look-up table, or a well-defined rule. For example,

$$x(n) = a^n.$$

We define a signal as *random* if it is generated in a random fashion and evolves in an unpredictable way. For most of the course, we will primarily focus on studying deterministic discrete-time signals.

Signals originating from the real world are typically analog signals, which means they have continuous-time and continuous amplitude. However, Digital Signal Processing focuses on digital signals, i.e., signals with discrete-time and discrete amplitude. For simplicity, we will consider these signals as discrete-time signals with continuous amplitude. Later in the course, we will study the effects of quantization on signals.

01.02 Digital signal processing: pros and cons

Processing a signal involves treating the signal with an appropriate system, such as a device, circuit, or algorithm implemented by a program on a computer or any other processor. Many of the signals found in nature are analog signals, and these signals can be directly processed by suitable analog systems. In this case, both the input and output signals of the analog system are analog:

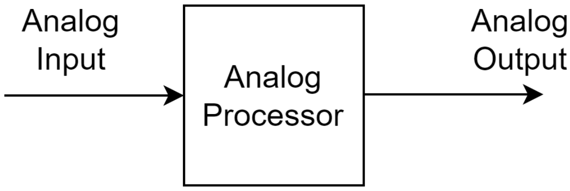


Figure 01.02: (From S. K. Mitra, “Digital signal processing: a computer based approach”, McGraw Hill, 2011)

Digital signal processing offers an alternative method for processing analog signals through three fundamental steps:

1. Analog to Digital (A/D) conversion.
2. Digital processing.
3. Digital to Analog (D/A) conversion.

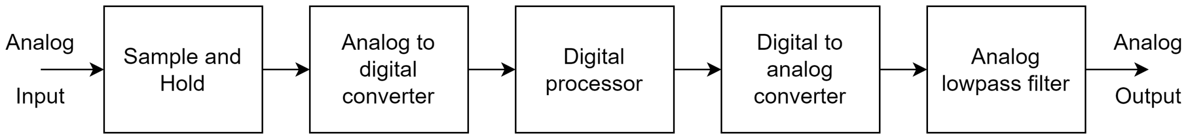


Figure 01.03: (From S. K. Mitra, “Digital signal processing: a computer based approach”, McGraw Hill, 2011)

The input signal undergoes processing through a Sample-and-Hold circuit, which samples the signal and maintains its value constant between samples, facilitating analog-to-digital conversion (this operation is often considered part of the A/D converter). A/D conversion involves quantizing and coding the samples. The coded samples can be numerically processed by a digital processor, such as a computer, a processor with architecture specialized for digital signal processing (a Digital Signal Processor), or a digital electronic circuit. Subsequently, the output signal of the digital processor is converted back to analog form with a D/A converter, typically generating an analog signal formed with continuous steps. An analog lowpass filter is employed to eliminate undesired high-frequency components (i.e., the steps) and generate the desired analog output signal. Refer to Fig. 01.04.
Conceptually, the entire process appears more complex than analog signal processing. Why do we prefer to process analog signals digitally?

Pros of digital signal processing

1. Digital circuits are less sensitive to component tolerances and are independent of temperature and aging effects on components. They can be produced in large volumes without the need for

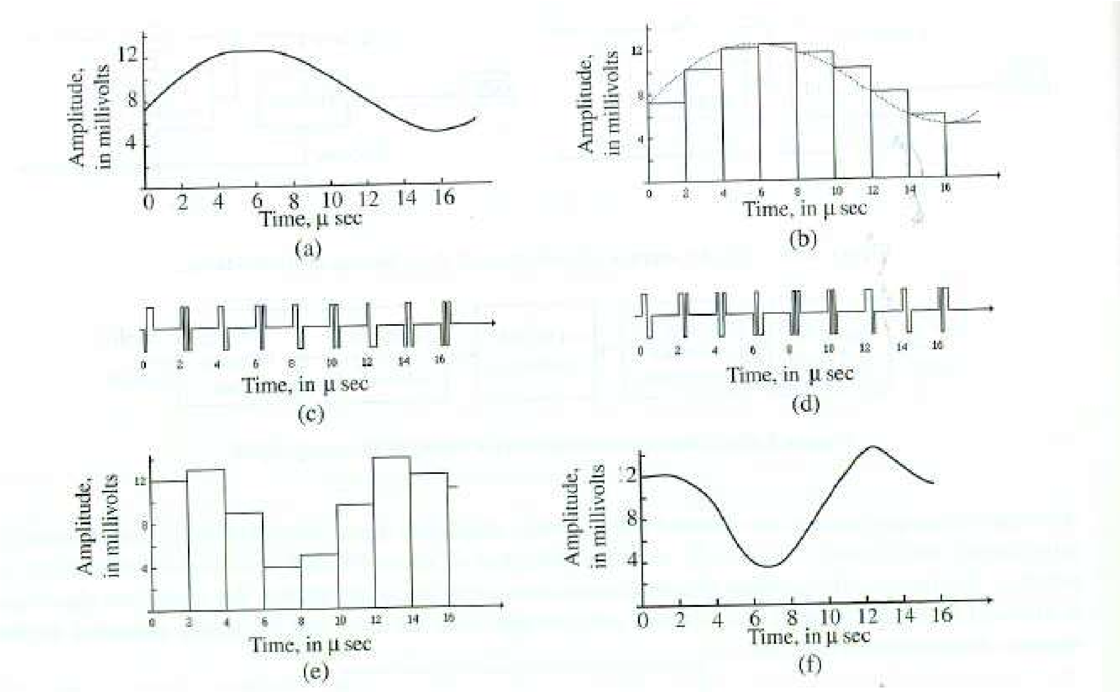


Figure 1.43: Typical waveforms of signals appearing at various stages in Figure 1.42. (a) Analog input signal, (b) output of the S/H circuit, (c) A/D converter output, (d) output of the digital processor, (e) D/A converter output, and (f) analog output signal. In (c) and (d), the digital HIGH and LOW levels are shown as positive and negative pulses for clarity.

Figure 01.04: (From S. K. Mitra, “Digital signal processing: a computer based approach”, McGraw Hill, 2011)

any adjustments during construction or later in use. They can be manufactured in the form of VLSI or ULSI circuits, enabling the implementation of very complex systems on a single chip.

2. Numeric filtering enables improved robustness and accuracy. By increasing the precision of operations, it becomes possible to arbitrarily reduce the noise introduced by the system.

3. Digital systems allow the implementation of much more complex signal processing algorithms than those of analog systems.

4. It becomes possible to have a software implementation of the system, which allows for easy reconfiguration of the digital processor functions (simply by changing the program). The same processor can process different signals through time-sharing of the processor.

5. It is easy to design systems that automatically adapt the internal parameters to the specific signal, known as adaptive filters.

6. Digital signals can be stored almost indefinitely without any loss of information.

7. Digital systems generally have a much lower economic cost than analog systems that implement the same operations.

Cons of digital signal processing

1. The system complexity is higher due to the pre- and post-processing operations (Sample-and-hold, A/D, D/A, lowpass filtering).
2. There are some limitations due to the speed of A/D and D/A converters and digital signal processors.
3. Digital signal processors necessarily use active devices that consume electric power.
4. There are some undesired effects originating from the use of algorithms implemented with finite-precision arithmetic.

Applications of digital signal processing

- Voice (voice coding, voice recognition, voice synthesis)
- Music (coding, filtering, equalization, audio effects, music transcription, synthesis)
- Images and Video (coding, filtering, pattern recognition, ...)
- Communications (modulation, demodulation, equalization, ...)
- Medicine (Electrocardiography, Electroencephalogram, Echography, TAC, ...)
- Seismology (seismic data analysis, petrol research, ...)
- Radar and sonar
- Vibration control systems in buildings or engines
- Automated guidance systems
- Economy and finance
- Big data
- ...

For more information study:

S. K. Mitra, ”Digital Signal Processing: a computer based approach,” 4th edition, McGraw-Hill, 2011
Chapter 1.1, pp. 1-3
Chapter 1.5, pp. 37-40