

EECS 451
Digital Signal Processing and Analysis
Lecture Notes
J. Fessler

DSP is everywhere, and hardly needs a motivating introduction these days: modems, cell phones, computer sound cards, digital video.

Course Overview

- 2 Discrete-time signals
- 2 Discrete-time systems (LTI, convolution)
- 3 z -transform
- 4 Discrete-time Fourier transform (DTFT)
- 8 Filter design
- 5 Discrete Fourier transform (DFT)
- 6 Fast Fourier transform (FFT)
- 9 A/D and D/A (sampling, quantization, reconstruction)
- Multirate signal processing (upsampling, downsampling)
- Image processing (as time permits)

Chapter 1

Introduction

Contents

Notation	1.1
Overview	1.2
Signals, Systems, and Signal Processing	1.2
Basic Elements of a DSP System	1.3
Advantages of Digital over Analog SP	1.3
Classification of Signals	1.4
Dimensionality	1.4
Time characteristics	1.5
Value characteristics	1.6
Other types of classifications	1.6
Deterministic vs Random signals	1.6
Periodicity	1.6
Energy and power	1.6
Frequency	1.7
Continuous-Time Sinusoidal Signals	1.7
Discrete-Time Sinusoidal Signals	1.8
Instantaneous frequency	1.10
Analog to Digital and Digital to Analog Conversion	1.11
Sampling Analog Signals	1.11
The Sampling Theorem	1.11
Quantization of continuous-amplitude signals	1.11
Quantization of sinusoidal signals	1.11
Coding of quantized samples	1.11
D/A conversion	1.12
Analysis of digital signals and systems vs discrete-time signals and systems	1.12
Summary	1.12

Notation

- set of real numbers: \mathbb{R}
- set of complex numbers: \mathbb{C}
- set of integers: \mathbb{Z}
- set of natural numbers: $\mathbb{N} = \{1, 2, \dots\}$
- square root of -1 : j

Overview

- Defines terms, *e.g.*, DSP
- Basic elements of a DSP system
- Basic types of signals
- Frequency
- Sampling A/D and D/A conversion

Much of this chapter is review of 206/306, so it will be covered only briefly in lecture.

Why digital over analog? Cheaper. More reliable. Programmable - flexible. Often higher precision.

These notes will follow organization, notation, and mathematical content of text fairly closely.

They will fill in the “whys,” and extend concepts to 2D for **image processing**.

1.1
Signals, Systems, and Signal Processing

Mathematically speaking, a **signal** is just a *function*. But usually we think of signals as representing some “physical” quantity that varies with time or space (or any other independent variable or variables).

Often when we discuss signals we refer to the mathematical representation of the physical quantity.

Example: An approaching ambulance siren produces a time-varying change in acoustic pressure that our ears perceive as sound. A simplified representation of the siren signal is

$$s(t) = (1 + t) \cos(2\pi[1000 + 10t + 300 \cos(2\pi 2t)]t)$$

The $(1 + t)$ amplitude term represents increasing loudness as the ambulance approaches, the sinusoidal term represents the siren oscillation, the frequency is time-varying; the $300 \cos(2\pi 2t)$ term represents the eeh-oooh-eeh-oooh periodic variation in pitch, and the $10t$ term represents increasing pitch due to the Doppler effect as the ambulance approaches. The point is that the primary physical attributes can often be captured in a mathematical representation that we call a signal.

Often, however, signals are very complicated and not easily described by a concise mathematical formula.

A **system** is a physical “device” that performs an operation on a signal.

Example. The human ear converts acoustic signals into electrical nerve synapses (another signal) that are processed by the brain. In this case the input and output signals for the system are different physical quantities.

Example. The tone controls (bass treble) in an audio amplifier perform filtering of an audio signal (represented as analog voltages in the amplifier). In this case the input and output signals are both voltage waveforms.

One of the main roles of electrical engineers is to design and analyze systems that take some input signal and produce some related (but almost always different) output signal. We refer such operations as **signal processing**.

Example. For an audio amplifier, ideally the output signal is “simply” an amplified version of the input signal. (On paper it is easy: $s_{\text{out}}(t) = a s_{\text{in}}(t)$. But implementing this in analog hardware with minimal distortion is nontrivial.)

Example. The input to a fingerprint recognition system is an image of a fingerprint. The “output” is (most likely) some bits within a digital device that correspond to the identity of the person with that fingerprint, or a list of likely matches.

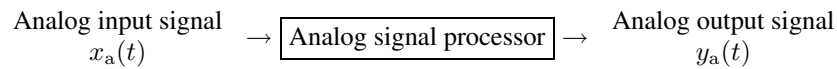
We call this type of application **signal analysis** or **signal classification**.

This course will emphasize signal processing methods for **digital signals** - to be defined soon.

1.1.1

Basic Elements of a DSP System

Typical analog signal processing system:



In EE (since after all the first E in EE is electrical) typically the input and output signals are voltages or currents.

Since most physical quantities are analog, a DSP system usually needs an **interface** between the analog physical world and the digital computer world.

Typical digital signal processing system



Often the final D/A step is not needed.

Example. In a speech recognition system the analog voice signal (from microphone) is digitized, and then the processing converts the digital signal into ASCII characters representing letters, numbers, punctuation, etc.

Although DSP systems have many advantages over analog SP, the A/D and D/A interfaces are the “Achilles heel” of DSP, since they limit the **speed** and **accuracy** of a DSP system. Of course the speed and accuracy of A/D converters is continually increasing with improvements in circuit technology, but faster and more accurate A/D chips are also usually more expensive, so a DSP system designer must consider the tradeoff between cost and performance.

Example. Music CD at 44kHz. Higher sampling rate might give higher quality, but less music on CD.

1.1.2

Advantages of Digital over Analog SP

- Programmable - flexible (“just” a software change rather than complete redesign)
- Often more accurate (higher precision) (*cf.* 10% tolerance resistors and other passive circuit elements)
- Can store, transmit, duplicate digital signals with no (additional beyond A/D) loss of fidelity.
- Cheaper (sometimes), *i.e.*, due to flexibility.
- More reliable (*cf.* temperature drifts of analog components).

1.2

Classification of Signals

A signal is a function of one or more independent variables.

1.2.1

Dimensionality

Domain dimension

We can classify signals by the **dimension of the domain** of the function, *i.e.*, how many arguments the function has.

- A **one-dimensional** signal is a function of a single variable, *e.g.*, time.
- An **M -dimensional** signal is a function of M independent variables.

Example. An audio signal $x_a(t)$ is a one-dimensional, or **1D**, signal, since it is a function of time.

Example. A color photograph is a two-dimensional (2D) signal, which we call an **image**, since it is a function of two spatial coordinates, *e.g.*, $f(x, y)$.

Example. What is an example of a 3D signal?

A “black and white” movie is a time-varying sequence of pictures $I(x, y, t)$, *i.e.*, it is a (scalar valued) function of two spatial coordinates x and y and of time t , so it is a 3D signal.

We will focus on **one-dimensional** signals in this course, generally considering the independent variable to be time t . We will also consider images, which are 2D signals.

Range dimension

We can also classify signals by the **dimension of the range** of the function, *i.e.*, the space of values the function can take.

- **scalar** or **single-channel** signals
 - real-valued scalar
 - complex-valued
- **multichannel** signals
 - real vector
 - complex vector

Example. A color TV picture can be described by a red, blue and green signal, so it is a three-channel signal, whereas a BW TV picture is scalar valued.

We will focus on **scalar** signals in this course, both real and complex.

Most of the design/analysis techniques generalize to multichannel and multidimensional signals.

1.2.2

Time characteristics (type of domain)

- **Continuous-time signals or analog signals**

Defined for all times $t \in (-\infty, \infty)$, or at least over some continuous interval (a, b) .

Example. $x_a(t) = e^{-t^2}$, $-\infty < t < \infty$. **Picture**

- **Discrete-time signals**

Defined only at certain specific values of time.

Typically use t_n , $n = 0, \pm 1, \pm 2, \dots$ to denote time instants where signal is defined.

In this course we focus on uniformly spaced time samples

$$t_n = nT,$$

where T denotes the time-spacing between samples.

In this case we can (and will) use the short hand $x[n]$ as follows:

$$x[n] = x_a(t_n) = x_a(nT).$$

Some authors use the notation x_n or $x(n)$.

Example.

$$x[n] = \begin{cases} 0, & n < 0, \\ 1 & n = 0, 1, 2, \\ 0, & n = 3, 4 \\ 0.5^n, & n = 5, 6, \dots \end{cases} \quad \textbf{Picture}$$

Discrete-time signals arise as follows.

- Sampling a continuous signal at discrete time instants.
- Accumulating a quantity over a period of time

Example. When counting number of heart attacks per month, n would index the month, and $x[n]$ would be the number.

For images, we refer to **continuous-space** functions $f(x, y)$ and **discrete-space** functions $x[n, m]$.

Example. When a “black and white” photograph with intensity $f(x, y)$ is **scanned** by a digital scanner, the output of the scanner is a digital image $x[n, m]$ consisting of uniformly-spaced samples of the original image:

$$x[n, m] = f(n \Delta, m \Delta),$$

where Δ denotes the sample spacing, *e.g.*, “72 dots per inch” means $\Delta = 1/72$ inches.

1.2.3

Value characteristics (type of range)

- A **continuous-valued signal** can take any value in some continuous interval, *e.g.*, voltage between 0 and 5 volts.
- A **discrete-valued signal** only takes values from a finite set of possible values.
Example. In heart attack example above, $x[n]$ could be 0, 1, 2, . . . , population_of_world.
- A **binary signal** only takes two values.

A **digital signal** is a **discrete-time** signal that is also **discrete-valued**.

If the input to a DSP system is originally an analog signal (*e.g.*, an acoustic voice signal that has been converted to a voltage signal by a microphone), then the A/D converter will convert the analog signal to digital form by **quantizing** its values to a finite discrete set of values.

Example. An 8-bit A/D converter can represent $2^8 = 256$ different values. Each value of the input signal must be rounded to the nearest of the 256 output values.

Other types of classifications

1.2.4

Deterministic vs Random signals

- **Deterministic signals** can be described by an explicit mathematical representation.
- **Random signals** evolve over time in an unpredictable manner.
Example: “Hiss” or “noise” in an audio system.

We will focus on deterministic signals, although reducing noise (eliminating a random component) is often a goal in a DSP system.

Periodicity

A discrete-time signal $x[n]$ is called **periodic** with period $N \in \mathbb{N}$, or N -periodic, if and only if

$$x[n] = x[n + N], \forall n \in \mathbb{Z}.$$

The smallest such N is called the **fundamental period** of the signal.

If no such N exists, the signal is called **aperiodic**.

A constant signal, *e.g.*, $x[n] = c$, is a degenerate type of periodic signal. In fact it is N -periodic for every $N \in \mathbb{N}$. From a time-domain perspective, it is natural to say that the “fundamental period” of this signal is $N = 1$. However, from a frequency-domain perspective, it is also reasonable to say that its fundamental period is ∞ . Which definition is used is not particularly important in practice.

Energy and power

For concepts that were introduced in detail in 1D in 206, these notes will sometimes present only the 2D versions. The 1D versions can be obtained easily from the 2D versions.

- The energy of an image $x[n, m]$ is defined as

$$E_x = \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} |x[n, m]|^2.$$

- If $E_x < \infty$, we say $x[n, m]$ is an **energy signal**.
- The average power of an image $x[n, m]$ is defined as

$$P_x = \lim_{N \rightarrow \infty} \frac{1}{(2N + 1)^2} \sum_{n=-N}^N \sum_{m=-N}^N |x[n, m]|^2.$$

- If $0 < P_x < \infty$, we say $x[n, m]$ is an **power signal**.

1.3

Frequency

There are similarities and differences between the meaning of frequency for continuous-time and discrete-time signals.

Point: to begin to see similarities and differences between discrete- and continuous-time.

1.3.1

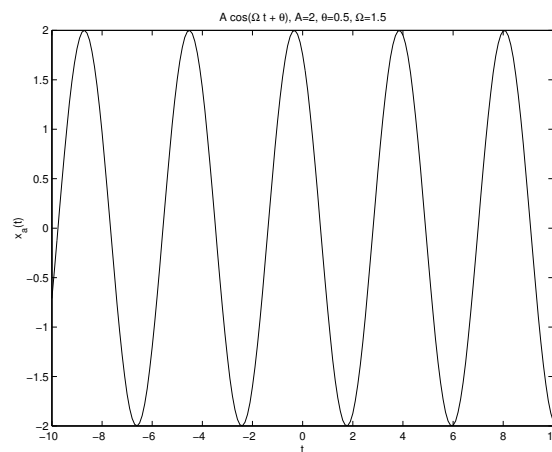
Continuous-Time Sinusoidal Signals

Consider an analog continuous-time sinusoidal oscillation

$$x_a(t) = A \cos(2\pi F_0 t + \phi), \quad \infty < t < \infty$$

- A is the **amplitude**
- ϕ is the **phase** in radians
- F_0 is the **frequency** in Hz (cycles per second)

Some books express the frequency in radians per second using $\Omega = 2\pi F_0$.



Properties of analog sinusoidal signals.

- A1 The signal $x_a(t)$ is **periodic** for any fixed value of F_0 , *i.e.*,

$$x_a(t + T_0) = x_a(t)$$

where $T_0 = 1/F_0$ is the **fundamental period** of the signal.

- A2 Continuous-time sinusoidal signals with distinct (different) frequencies are distinct. If $F_1 \neq F_2$ (and both have the same sign) then $\exists t_0$ s.t. $A \cos(2\pi F_1 t_0 + \phi) \neq A \cos(2\pi F_2 t_0 + \phi)$.
- A3 Increasing the frequency F_0 increases the **rate of oscillation**, *i.e.*, there will be more periods in a given time interval.

The above relationships also hold for **complex exponential** signals

$$x_a(t) = A e^{j(2\pi F_0 t + \phi)}$$

where by the **Euler identity**

$$e^{\pm j\theta} = \cos \theta \pm j \sin \theta.$$

For sinusoidal signals, the frequency F_0 is usually taken to be nonnegative.

But for complex exponential signals we must allow the frequency F_0 to be both positive and negative.

1.3.2

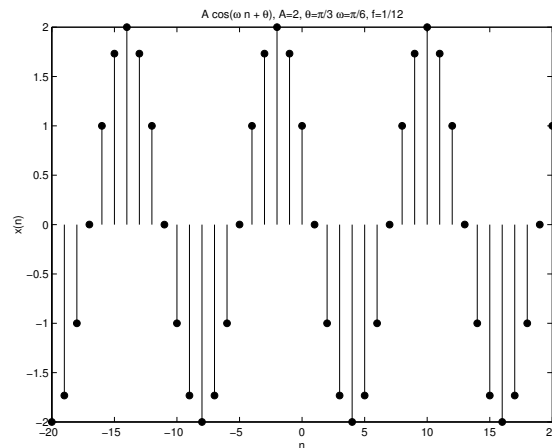
Discrete-Time Sinusoidal Signals

A discrete-time sinusoidal signal may be expressed

$$x[n] = A \cos(\omega n + \phi), \quad n = 0, \pm 1, \pm 2, \dots$$

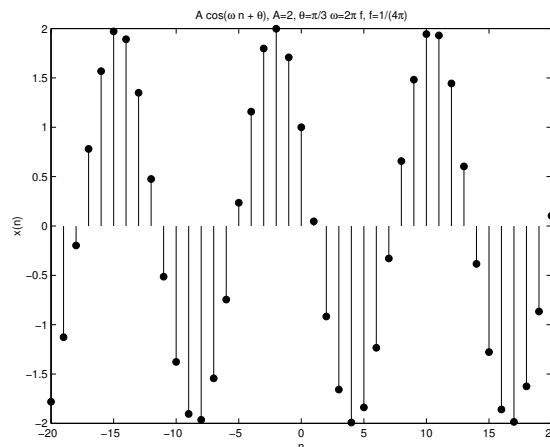
- n is an integer variable called the **sample number**
- A is the **amplitude**
- ϕ is the **phase** in radians
- ω is the **frequency** in radians per sample

Sometimes we express the frequency in cycles per sample using f , where $\omega = 2\pi f$.

1.12 Picture of discrete-time sinusoid

Properties of discrete-time sinusoidal signals.

- B1 A DT sinusoidal signal $x[n]$ is **periodic** if and only if its frequency ω is 2π times a **rational number**, i.e., $\omega = 2\pi M/N$ for $M, N \in \mathbb{Z}$.



Why periodic only for rational frequencies?

Proof. *skip*

Sinusoidal discrete-time signal with frequency ω_0 is periodic iff $\exists N$ s.t.

$$\cos(\omega_0 n + \phi) = \cos(\omega_0(n + N) + \phi) = \cos(\omega_0 n + \phi + \omega_0 N).$$

Since \cos is periodic with fundamental period 2π , the above relationship holds iff $\omega_0 N$ is an integer multiple of 2π , i.e., \exists an integer M s.t. $\omega_0 N = 2\pi M$, or equivalently $\omega_0 = 2\pi M/N$. Thus the frequency must be a ratio of two integers, and hence rational.

Skill: Finding the fundamental period of a DT sinusoidal signal.

Express $\omega_0 = 2\pi M/N$, where $M \in \mathbb{Z}$ and $N \in \mathbb{N}$ and M and N have no common divisors.

Then N will be the **fundamental period** (in samples).

If no such ratio, then $\frac{\omega_0}{2\pi}$ is irrational and the DT sinusoidal signal is **aperiodic**.

To convert a quantity, such as the period N , from units “samples” to time units (e.g., seconds), multiply by the sampling rate $T_s = 1/F_s$.

- B2 DT sinusoidal signals with frequencies separated by an integer multiple of 2π are identical (indistinguishable).

If $\omega_2 = \omega_1 + k2\pi$ then

$$\cos(\omega_2 n + \phi) = \cos((\omega_1 + 2\pi k)n + \phi) = \cos(\omega_1 n + \phi + 2\pi kn) = \cos(\omega_1 n + \phi)$$

since \cos is 2π periodic.

Discrete-time sinusoidal signals with frequencies in the range $0 \leq \omega \leq \pi$ are distinct.

Any DT sinusoidal signal with frequency outside the range $0 \leq \omega \leq \pi$ is identical to a DT sinusoidal signal with a frequency within that range (possibly with the negative phase). Hence we refer to such signals as **aliases**, since they are “the same signal with a different name.”

Example.

$$\begin{aligned} \cos\left(\frac{19\pi}{7}n + \frac{\pi}{3}\right) &= \cos\left(\frac{5\pi}{7}n + 2\pi n + \frac{\pi}{3}\right) = \cos\left(\frac{5\pi}{7}n + \frac{\pi}{3}\right) \\ \cos\left(\frac{9\pi}{7}n + \frac{\pi}{3}\right) &= \cos\left(\frac{-9\pi}{7}n - \frac{\pi}{3}\right) = \cos\left(\frac{5\pi}{7}n - 2\pi n - \frac{\pi}{3}\right) = \cos\left(\frac{5\pi}{7}n - \frac{\pi}{3}\right) \end{aligned}$$

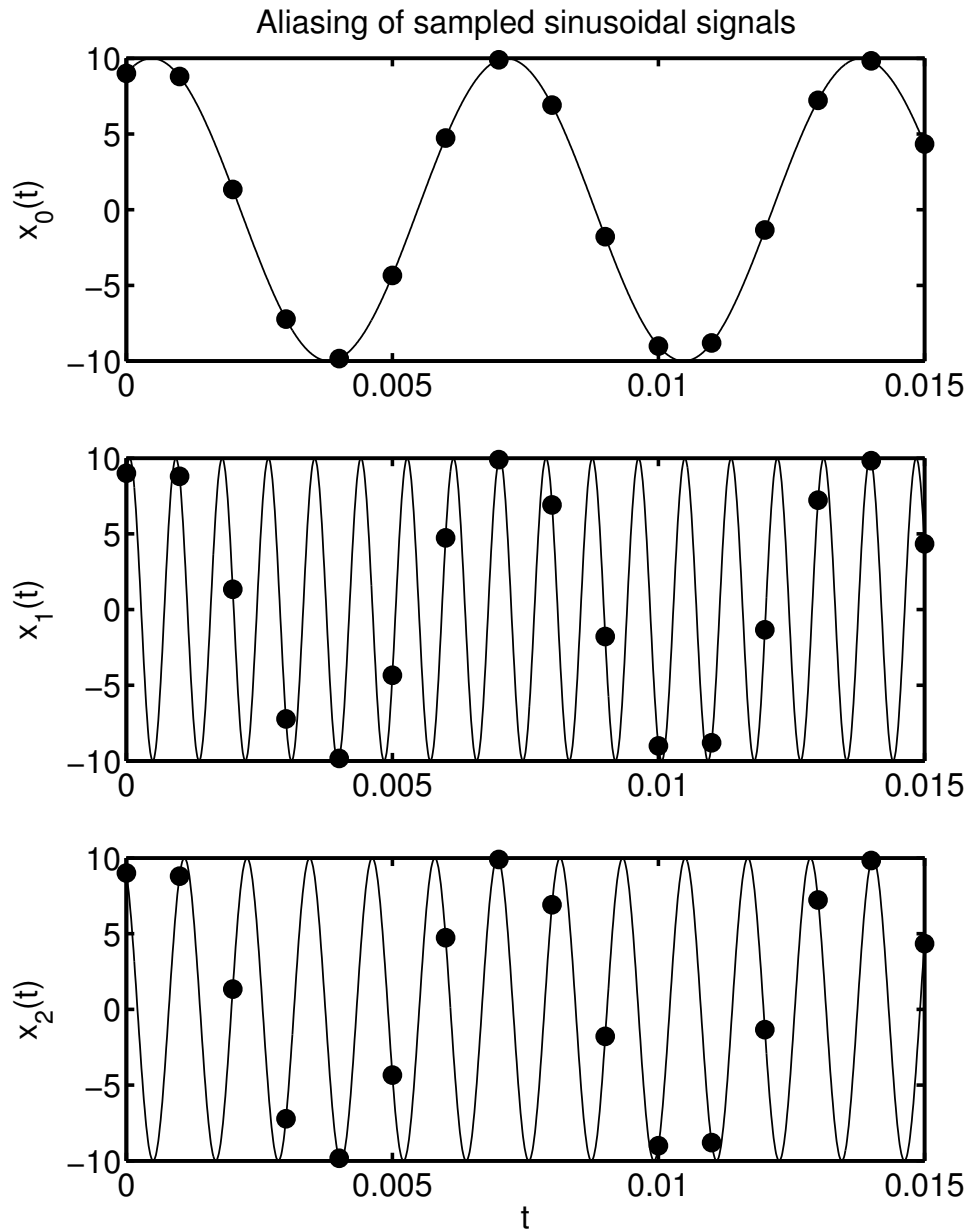
- B3 The highest **rate of oscillation** of a discrete-time sinusoid is attained when $\omega = \pi$. **Picture**

DT **complex exponential** signals:

$$x[n] = A e^{j(\omega n + \phi)}$$

For sinusoidal signals, we usually take the frequency to be nonnegative. For complex exponential signals, we allow the frequency ω to be both positive and negative.

Aliasing illustrated



Three distinct CT sinusoidal signals whose DT samples are identical.

Instantaneous frequency

For a “sinusoidal” function of the form

$$x_a(t) = \cos(2\pi g(t)),$$

the **instantaneous frequency** is given by

$$F_t = \frac{d}{dt}g(t).$$

Example. If $g(t) = F_0 t$, then $\frac{d}{dt}g(t) = F_0$, as usual.

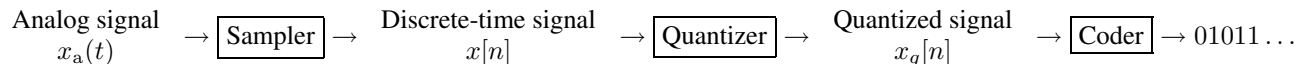
Example. If $g(t) = \alpha t^2$, then $\frac{d}{dt}g(t) = 2\alpha t$, meaning that the instantaneous frequency is increasing (linearly) with time. This is called a **chirp** signal.

1.4

Analog to Digital and Digital to Analog Conversion

Before an analog signal can be processed by a digital system, the signal must be converted into digital form. This process is called **analog-to-digital (A/D) conversion** and the corresponding devices are called **A/D converters** or **ADCs**.

An A/D converter has the following components, at least conceptually.



- **Sampling** the analog signal at discrete time instants, to form a discrete-time signal.
- **Quantizing** the sampled signal to form a discrete-time, discrete-valued signal. The values of $x_q[n]$ are from some finite set.
- **Coding** the discrete-valued signal using a unique binary sequence for each element in the set of possible values.

The difference between the unquantized signal $x[n]$ and the quantized signal $x_q[n]$ is called the **quantization error**.

1.4.1

Sampling Analog Signals

Ideal **periodic sampling** or **uniform sampling** is defined by

$$x[n] = x_a(nT_s)$$

where T_s is the **sampling period** or **sampling interval**. (See Ch. 9 for practical discussions of sampling.)

Its reciprocal, $F_s = 1/T_s$ is called the **sampling rate** or the **sampling frequency**.

Skim aliasing discussion for now. $F_s/2$ called **folding frequency**.

1.4.2

The Sampling Theorem

Is there loss of information when sampling a continuous-time signals? Yes in general, but not always.

The following is the **Nyquist sampling theorem**.

If an analog signal $x_a(t)$ is **band-limited**, *i.e.*, if the highest frequency in $x_a(t)$ is F_{\max} , *i.e.*, $X_a(F) = 0$ for $|F| > F_{\max}$, then

- it is sufficient to sample at any rate $F_s > 2F_{\max}$, and
- we can exactly recover (in principle) the analog signal $x_a(t)$ from such samples using the following formula:

$$x_a(t) = \sum_{n=-\infty}^{\infty} x[n] \frac{2F_{\max}}{F_s} \text{sinc}(2F_{\max}(t - n/F_s))$$

where $x[n] = x_a(n/F_s)$ and $\text{sinc}(t) \triangleq \frac{\sin(\pi t)}{\pi t}$.

The sampling rate $F_s = 2F_{\max}$ is called the **Nyquist rate**.

Because of this property, A/D converters are often preceded by an **anti-aliasing filter** that is a low-pass filter with cutoff frequency appropriately coupled to the sampling rate of the A/D converter.

Note: if $X_a(F)$ does not have any Dirac delta functions at $\pm F_{\max}$, then it is sufficient to sample at $F_s \geq 2F_{\max}$. But if $x_a(t) = \sin(2\pi F_{\max} t + \theta)$, then we need $F_s > 2F_{\max}$.

1.4.3

Quantization of continuous-amplitude signals

Quantization of audio signals illustrated in first computer assignment.

If signal quantized to 2^b levels, what is bit rate out of coder? bF_s (bits/sample) (samples/sec)

1.4.4

Quantization of sinusoidal signals

1.4.5

Coding of quantized samples

An n -bit code can represent 2^n discrete signal values.

1.4.6

D/A conversion

Example: CD players and PC sound cards...

Skim text for the above

1.4.7

Analysis of digital signals and systems vs discrete-time signals and systems

The quantization/coding step makes analysis complicated.

Therefore we focus initially on **discrete-time** but **continuous-valued** signals.

1.5

Summary

For simplicity, we will focus initially on analyzing discrete-time signals and systems, disregarding the connection to the real-world analog signal $x_a(t)$. Nevertheless, it is this connection that makes DSP useful in practice, so we will return to it later.

