Signal Processing of Discrete-time Signals

Andrew C. Singer and David C. Munson Jr.

January 26, 2009

Chapter 1

Overview of Discrete-time Signal Processing

- 1 DSP overview

- 2 Continuous-time signals

- 3 Discrete-time signals
  GOALS

  > 1. Overview of discrete-times signal processing
  >
  > 2. Introduction to discrete-time and continuous time signals

☐

1.1   DSP overview:

This is a textbook covering the topic of discrete-time signal processing. The tremendous advances in integrated circuit technologies over the last few decades have enabled a remarkable revolution in the way in which we measure and process information. The types if information, or the "signals" as we call them, that we have interest in processing can range widely from measurements of our environment, such as temperature, rainfall, geophysical or siesmic data, to signals governing man-made machines of interest, such as the velocity of a moving vehicle, or the controller modulating signals to the fuel injectors within our cars, to purely man-made sources of information, such as radio or television signals, or the MP3 encoding of recorded music.

   The abstract concept of a signal is that of a measurement of a quantity of interest that is indexed by an associated reference axis. Most often, we will refer to this reference axis as the "time" axis for the signal and refer to these as signals that vary as a function of time. For example, within an analog circuit, we may measure the voltage across a given resistor within the circuit fabric and label this "signal" $v_r(t)$ to denote that this is the voltage v of a particular resistor r taken at a particular point in time t. As time varies, the voltage may be plotted against the independent variable t to produce a voltage vs. time plot, where we typically assume that negative time is to the left, and positive time to the right. For most signals of interest, there is a common notion of a reference time, $t = 0$, such that signals may be plotted with this reference time in the center, so long as the reference time is known to those interpreting the graph.

   Another example of a signal may be the sequence of closing prices of the Dow Jones Industrial Averages (DJIA) as reported daily from the New York Stock Exchange. Unlike the resistor voltage example, the sequence of prices will be indexed by an integer reference axis indicating the specific day in a sequence of days at which the index closed at a particular value. We may write $p[n]$, to denote the closing price of the DJIA on the n^{th}day of trading, referenced to some common day $n = 0$. Note that while the resistor voltage $v_r(t)$ could be referenced to any time t taking on any real value, the price $p[n]$ can only be referenced to n taken as an integer. We refer to signals whose index t can take on any real value as continuous-time signals, since the "time" signal is assumed to be available over a continuum of values t. We refer to signals whose index n can only take on values from the integers, i.e. only discrete values, as discrete-time signals. While this text is largely about discrete-time signal processing, that is the processing and analysis of discrete-time signals, we will also be interested in understanding how similar properties of continuous-time signals relate to discrete-time signals, when the discrete-time signals themselves are related in a direct way to a continuous-time signal. A common means for this to occur is through the notion of sampling.

   When you play a song through your MP3 player, you are processing a discrete-time recording of music, which is stored as samples of the music signal in the memory on your player. Your player then converts this set of samples into a continuous-time electrical signal which is then transduced into an acoustic signal that you hear through your headphones. This is the process of converting a discrete-time signal into a continuous-time signal, which is done because our auditory processing system (our ears) want continuous-time signals. However the microprocessors and digital circuitry on your MP3 player work best with discrete-time samples of the music that can be stored sequentially in the memory on your device. Similarly, when the MP3 file was created, a continuous-time version of the music (either from a magnetic tape recording, or a live studio performance) was processed such that it could be sampled in time and stored digitally to be used by your MP3 player. This process of conversion from a continuous-time signal to a discrete-time signal is called sampling, and in its simplest form we could take the signal $v_r(t)$ and create a discrete-time signal from it by simply recording the values at a periodic rate $f_s = 1/T$, called the sampling rate to produce the discrete-time signal $v_r[n] = v_r(nT)$, which is precisely what is done on modern digital sampling oscilloscopes. If the sampling rate is sufficiently fast, then when plotted on the oscilloscope, the signal $v[n]$ with a properly normalized time axis will be virtually indistinguishable from the continuous-time signal $v_r(t)$. Of course the notion of "sufficiently fast" will depend heavily on the content of the signal $v_r(t)$, but we are getting ahead of ourselves, and will return to this notion again shortly, and then again more carefully when we revisit sampling in later chapters.

   This enables us to define or adopt some notation that will be convenient to use throughout this text. We will denote a continuous-time signal using the round braces notation we just employed for the resistor voltage, namely, $x_c(t)$, for a continuous-time signal x. Note that the term "analog" is often used synonymously to the term "continuous time," however there is an important difference between the two. A contunuous-time
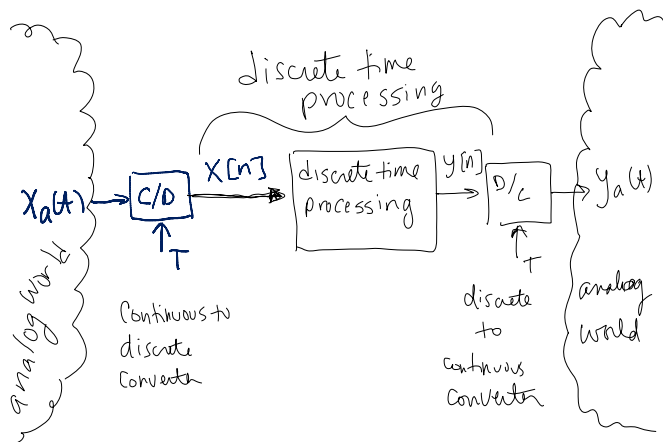
Figure 1.1: Discrete-time signal processing in an analog world. Note that in the "analog world" both the independent "time" variable is continuous as are the possible values of the signal $x_a(t)$. In the "discrete-time" world, after the C/D converter, the independent time variable n is now discrete, however the amplitudes $x[n]$ may still take on real-valued quantities. If the discrete-time processing takes place on a digital signal processor or standard microprocessor, then the discrete-time signal would need to also be a digital signal, i.e. its amplitude must also be quantized to a finite precision using a finite-precision analog-to-digital (A/D) converter.

signal is one for which the indepdendent variable (t in the case of $x_c(t)$) takes on any value in a continuum, i.e. any real number. An analog signal is one for which the *values* of the signal can take on any level within a continuum, i.e. $x_c(t) \in \Re$. Note that a continuoous-time signal could be an analog signal, if both t and $x_c(t)$ can take on values from the reals. However a discrete-time signal $x[n]$ could also be an analog signal, if its values can be arbitrary real numbers. The term "digital" referrs to signals that have been quantized such that their values must lie within a finite set, as could be represented in digital form, using a binary representation. Most of the signals we will be interested in for this text will be discrete-time analog signals, since we will not consider effects of numerical quantization, that is the effects of forcing the signals $x[n]$ to only take on values that can be represented using a binary representation with a finite number of bits. For simplicity, we will simply refer to these as discrete-time signals.

Discrete-time signals will be denoted using the square braces notation used for the DJIA prices, namely, discrete-time signal x will be denoted $x[n]$. In both the discrete-time and continuous-time cases we will refer to the independent variables t and n as the "time axis" though in some cases this may be something different such as the position of a vehicle along a track, or of the reading head along a magnetic tape, or the index labeling of the nth sheep to be sheared at a wool processing plant. While many examples of such discrete-time signals exist in which the independent axis is not simply a time variable, a vast majority of discrete-time signal processing is undertaken for the case when a continuous-time signal or measurement is sampled, and then processed digitally by a special purpose piece of hardware, to produce a new set of discrete-time output signals and these new output signals are then converted to continuous-time signals for use in our "continuous-time, analog world", as indicated in Figure 1.1.

In the figure, we note that many signals of interest in our "analog world" (where both time and amplitude can be taken from the real numbers) are indeed continuous in nature, our notion of time itself is indeed continuous and many such signals are indexed with respect to time. However, for a variety of reasons, it is more practical to build processors for these signals out of digital electronics, and therefore, we need discrete-time representations of signals. Much of this course will take you through the steps necessary to fully understand (and hopefully to build!) systems of the sort depicted in Figure 1.1. In the figure, note that the signals of interest begin within the analog world to the left, with the signal of interest $x(t)$. This signal is then processed using a "continuous-to-discrete" converter or "C/D" converter for short. In simple terms, the C/D converter simply samples the signal $x(t)$ at equally spaced samples, spaced T seconds apart in time to produce the discrete-time signal $x[n] = x(nT)$. We mentioned earlier that the necessary sampling rate $f_s = 1/T$ at which the C/D converter must operate was dependent upon the makeup of the signal $x(t)$.

$$x(t) \longrightarrow \boxed{\text{CT Filter}} \longrightarrow y(t)$$

Figure 1.2: A continuous time (CT) filter processing the continuous-time signal $x(t)$ to produce the continuous-time output $y(t)$.

We will elaborate on this a bit here, and in more detail later. You may recall from a course on continuous signals and systems the concept of the frequency content of continuous-time signals, as developed by Fourier through what we now call the Fourier series representation of periodic signals and the Fourier transform for aperiodic signals. For a large class of signals $x(t)$, we can define the following two expressions, which together make up what is referred to as the continuous-time Fourier transform (CTFT) representation for aperiodic signals,

$$
\begin{aligned}
x(t) &= \frac{1}{2\pi} \int\limits_{\omega=-\infty}^{\infty} X(\omega) e^{j\omega t} d\omega \\
X(\omega) &= \int\limits_{t=-\infty}^{\infty} x(t) e^{-j\omega t} dt.
\end{aligned}
$$

We refer to the signal $X(\omega)$ as the Fourier transform of the signal $x(t)$. The first relation above is called the Fourier transform synthesis equation, as it depicts the "recipe" or formula for synthesizing the signal $x(t)$ out of the class of periodic complex exponential sequences $e^{j\omega t} = \cos(\omega t) + j\sin(\omega t)$, where $j = \sqrt{-1}$. When the signal $X(\omega)$ satisfies $|X(\omega)| = 0$ for $|\omega| > B$, then we say that the signal $x(t)$ is "band limited" to $F = B/2\pi$ Hz, since there is no energy in the signal outside of the range from $-B < \omega < B$. We refer to twice the highest frequency in $x(t)$ as the Nyquist frequency, i.e. $2F$, which is also equal to the total width of the frequency band (or band width) of the signal, when you include the negative and the positive frequencies. When the C/D converter used in Figure 1.1 satisfies $(1/T) > 2F$, then it is possible to exactly recover the signal $x(t)$ from its samples. Namely, we can directly compute

$$x(t) = \sum_{-\infty}^{\infty} x[n] \text{sinc}(\frac{\pi}{T}(t - nT)), \tag{1.1}$$

where the "sinc" function is defined as

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

In practice, we sample slightly faster than $2F$, i.e. smaller values of T, to ensure that we can capture all of the information in the signal $x(t)$. The equation given in (1.1) is precisely the process that is undertaken in an ideal D/C, or discrete-to-continuous converter, as depicted in figure (1.1), in order to construct the signal $y(t)$ from the discrete-time signal $y[n]$, just prior to the signal $y(t)$ being sent back out into the analog world. Note that we could process the signals directly within the analog world. This is precisely what analog electronics is for and this is a huge industry. However, as the processing that is desired becomes more complex, the capacitors, inductors, and resistors, and transistors, and diodes, and other circuit components depicted in Figure (1.2) become increasingly difficult to design and manufacture and become more expensive.

The alternative to processing such signals while leaving them within the analog world, is to sample them, and process them in discrete-time, as depicted in Figure (1.3). The discrete-time (DT) filter shown in the figure depicts the process of mathematically manipulating the samples in $x[n]$, which essentially comprises multiplication, addition and movement through storage registers.

As we will see in later chapters, often such discrete-time operations can be described by what are known as difference equations operating on the signals $x[n]$ and $y[n]$, much like the governing equations for continuous-time filters can often be described in terms of differential equations. An example of such a difference equation might be

$$y[n] = b_0 x[n] + b_1 x[n-1] + \cdots + b_N x[n-N],$$

$$x(t) \longrightarrow \boxed{\text{C/D}}_{\uparrow T} \longrightarrow x[n] \longrightarrow \boxed{\text{DT Filter}} \longrightarrow y[n] \longrightarrow \boxed{\text{D/C}}_{\uparrow T} \longrightarrow y(t)$$

Figure 1.3: Discrete-time filtering alternative to the continuous-time filtering of a continuous-time signal.

where, since the output $y[n]$ can be computed directly from samples of the input, the difference equation is called "non recursive". This is in contrast to a recursive implementation that might take the form

$$y[n] = a_1 y[n-1] + a_2 y[n-2] + \cdots + a_N y[n-N], + b_0 x[n] + b_1 x[n-1] + \cdots + b_M x[n-M],$$

where the coefficients b_k and a_k are selected to produce the desired relationship between the input $x[n]$ and output $y[n]$.

The term DSP is often used to describe "digital signal processing" which describes not only discrete-time signal processing, but also the additional step of quantizing the output of the ideal C/D converter to create a "digital signal" which is not only a discrete-time signal, but one that is also capable of being represented by a fixed number of bits. The additional step of quantization that takes place within the C/D converter is something that we will explore in more detail in later chapters. However for now, we will choose to ignore these bits of detail. Some more sophisticated forms of DSP that have impacted the technology you use regularly, include telecommunications, in which the modulation, coding, echo cancellation and line equalization is done nearly completely in discrete-time in modern systems. Speech processing uses analysis and synthesis methods for low bit-rate speech transmission over cellular and other communications links, speech recognition systems are sophisticated digital algorithms that process sampled and digitized speech. Image processing that takes place for HD video and other digital image and multimedia applications, raging from coding for video conferencing, digitization and compression in digital cameras and camcorders (such as those on the mobile telephone in your pocket), fax transmission, HDTV, digital video recorders (DVRs), and image archiving; noise removal; deblurring; and object recognition. DSP has a massive footprint in consumer electronics from CD/DVD players, to Digital Satellite TVs, HD-DVRs, cellular phones, MP3/MP4 players. It's nearly impossible to find consumer electronics to do not contain DSP these days. Modern medical imaging is largely based on discrete-time signal processing technology, including the pervasive computed tomography (CT) and magnetic resonance imaging (MRI) - even X-ray technology uses an array of detectors to construct a digital representation of the received radiation, so X-ray films are as much a part of the past as film-based cameras, 8 track tapes, film-based movie cameras and more.

Digital representation and processing is often preferred over direct analog/continuous-time processing for three (or more) main reasons. First, discrete-time systems are more versatile – sophisticated processing, time varying and adaptive filtering, nonlinear processing, multidimensional signals (especially image processing) can all be captured with relative ease in such systems. Second, the guaranteed accuracy, as determined by the register lengths (number of bits) used to perform the computations, and not by the non-ideal integrated circuit components, including resistors capacitors, inductors, operational amplifiers, and other components whose performance is not only not ideal, but will vary over time and with temperature. Finally digital implementations are often smaller, cheaper, lower power, thanks in large part to the incredible scaling of integrated microelectronics through the process known as "Moore's Law", whereby such integration densities nearly double every 18 months. Due to speed limitation of analog-to-digital converters and computers, digital signal processing was initially employed (during 1960s) in application areas having low-bandwidth signals, such as telephone quality speech (approximately 3 KHz bandwidth). The practical frequency range for digital processing has increased vastly over the years and continues to climb. As an example, digital signal processing is now commonly employed in radar, where sampling rates may be tens or even hundreds of megahertz. Optical communications links are starting to use digital signal processing with sampling rates in excess of $50 GHz$! Note that here we use the units of Hz to represent samples-per-second, in addition to its more common meaning of cycles-per-second. This slight abuse of notation is common in the discrete-time signal processing industry and one with which we all must live.

1.2   Continuous-time (CT) signals

While the class we will refer to as continuous-time signals is defined by properties of the independent time axis, t, we can further break down this class of signals into subclasses that will be useful in various contexts. For example, we refer to finite-length continuous-time signals, as signals $x(t)$ for which the independent time axis is only defined over a finite interval, i.e., $x(t), t \in [0, 1]$, as might be convenient for measurements taken within a specific time interval, or for which the signals are not present outside of the interval of interest. For example, $x(t)$ might correspond to the temperature of the water column in the ocean from the surface, $t = 0$ to the ocean bottom at, say, $t = 25$ meters, for a shallow water environment. It does not make sense to consider the temperature as existing outside of this region, since there is no ocean outside this region at this location in the ocean.

   Having defined finite-length signals, we can similarly define infinite-length signals as those for which the time axis of interest, or over which the signals of interest are defined is indeed infinite, i.e. $x(t), t \in \Re$. Finally, we can also define continuous-time periodic signals as infinite-length signals for which the following relationship holds $x(t) = x(t + P)$ for some value of $P \in \Re$. We refer to the minimum value of $P > 0$ for which this relationship holds as the "fundamental period" or just the "period" of the periodic signal.

1.3   Discrete-time (DT) signals

We can repeat this exercise for discrete-time signals as well. While the class we will refer to as discrete-time signals is defined by properties of the independent time axis, n, we can further break down this class of signals into subclasses. Once again, we refer to finite-length discrete-time signals, as signals $x(t)$ for which the independent time axis is only defined over a finite interval, i.e. $,x[n], n \in [0, N]$, as might be convenient for a finite set of measurements, or for which the signals are not present outside of the interval of interest. Once again, $x[n]$ might correspond to the number of words on the nth page of a specific book that only has N pages. It does not make sense to consider pages that do not exist!

   Having defined finite-length signals, we can similarly define infinite-length discrete-time signals as those for which the time axis of interest, or over which the signals of interest are defined is indeed infinite, i.e. $x[n], n \in Z$, where Z refers to the set of integers. Finally, we can also define discrete-time periodic signals as infinite-length signals for which the following relationship holds $x[n] = x[n + P]$ for some value of $P \in Z$. We refer to the minimum value of $P > 0$ for which this relationship holds as the "fundamental period" or just the "period" of the periodic signal. We will return to the concept of discrete-time and continuous-time periodic signals in Chapter 2.