Chapter 1 Introduction

This chapter is devoted to providing students a set of general concepts of signals and systems. We will begin our development with the intuitive questions that may be raised by most of the students at the beginning of the first class of this course: what is a signal and what is a system? Mathematical descriptions and representations of signals and systems are the most important concepts throughout this course and also play a role of corn-stone for other more advanced subjects. We will build on the foundation for developing these concepts and discuss some properties of systems as well as the relationship between signals and systems in this chapter.

1.1 Overview of signals and systems

Signals and *systems* are two of the words which are heard most frequently in our daily life. These concepts arise in virtually all areas, and particularly, play a very important role in engineering. In fact, it can be argued that much of the recent development of high technology, which has brought our life to a new dimension, is a result of advancements in the theory and techniques of signals and systems.

1.1.1 What is a signal?

Examples of signals we encounter frequently include speech music signals, and, picture image signals, which, in the *signal processing* community, are usually referred to as *audio* and *video* signals, respectively. A signal is actually a variable used to carry *information.* For example, a speech signal from the speaker of a research seminar represents air pressure that varies with time and stimulates the audience's ears, and the information is reflected by the way how the air pressure (i.e., the signal) changes with time. A Signal can be represented in different forms Figure 1.1 shows the waveform of a recorded speech signal displayed on a computer screen. The physical meaning (i.e., information) carried by this signal is "di qiu" in Mandarin Chinese (it means "earth" in English) spoken by a Chinese male.

An image signal is the light intensity, also called *gray level*, that varies with two spatial coordinates (see Figure 1.2), while video signal in the televisions consists of a set of pictures that occur in sequence and hence are the light intensity that changes with the two spatial coordinates as well as the time.

A *signal* is formally defined as a *function* of one or more *independent* variables.

The speech signal in Figure 1.1 can then be denoted as $s(t)$ with the independent variable t representing the time, while the image signal in Figure 1.2, as $p(x, y)$ with x, y representing the horizontal and vertical coordinates, respectively.

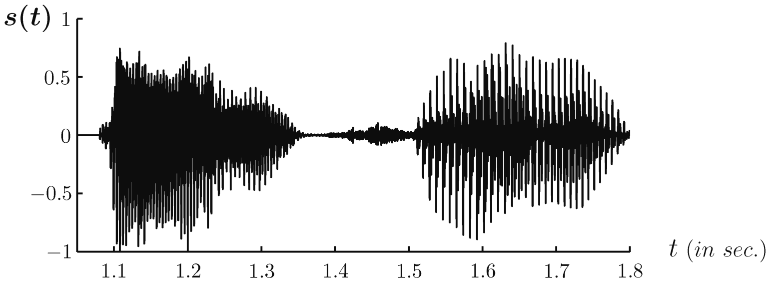


Figure 1.1 A recording of the speech for “di qiu” in Chinese (meaning “earth” in English) spoken by a Chinese male.

Figure 1.2 The scene of *female teacher* in San Qing Mountain.

It should be pointed out that most signals in our life are generated by natural means and hence correspond to physical phenomena. However, a signal can also be produced artificially, say by computer simulations. Such a signal does not have a specified physical meaning as it is not born from any natural phenomenon but we can always use it to represent a certain meaning (i.e., information). In fact, a signal, no matter how it is generated, is just a carrier that can be used to represent different information.

1.1.2 What is a system?

In the broadest sense, a *system* is an *entity* that is used to achieve a specified function. Figure 1.3 depicts a simplified rectifier circuit.

Such a system consists of two resistors r, R, one diode D, and a capacitor C as components. Denote $y(t)$ as the voltage across the capacitor, which is the response to the voltage source $x(t) = A\cos(2\pi F_0 t + \phi)$ applied. The function of this system is to make $y(t)$ constant, namely, invariant with time t. In this system, the source $x(t)$ and the capacitor voltage $y(t)$ are referred to as the *input signal* and *output signal*, respectively.

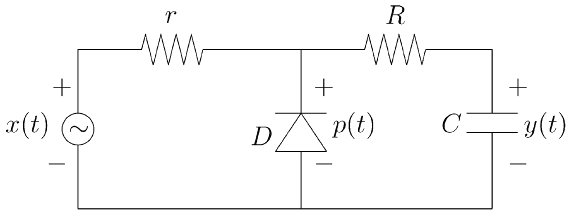


Figure 1.3 Block diagram of a rectifier circuit.

> A system is an interconnection of components or parts with terminals or access ports through which signals can be applied and extracted.

Let y be the output *responding to* the input x of a system. This fact is denoted as

$$x \rightarrow y$$

Such a notation focuses on the relationship between the input and output of the system rather than how all the components are connected. Frequently, a system can be viewed as a *black box*, in which the *input* signals are *processed* in some manner to yield the *output* signals. Figure 1.4(a) is the black box representation of the system $x(t) \rightarrow y(t)$, whose detailed structure is specified by Figure 1.3.

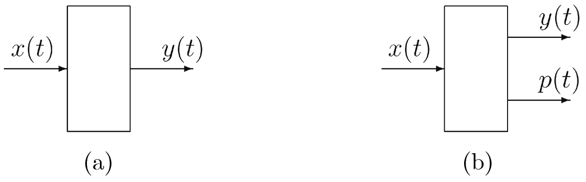


Figure 1.4 Black box representations of the circuit by Figure 1.3.

It should be pointed out that if we are also interested in the voltage $p(t)$ across the diode D in Figure 1.3, then the same circuit can be viewed as a system of one input and two outputs. See Figure 1.4(b).

A system may have M inputs and N outputs. Such a system is usually referred to as a multi-input multi-output (MIMO) system when M, N are all bigger than one. If $M = N = 1$, the system is a single-input single-output (SISO) system. In this book, most of the systems in our discussions belong to the catalog of SISO.

Examples of much more complicated systems can be easily found. Below are three important classes of systems that find a lot of applications in our daily life.

A. Communication systems

Figure 1.5 shows a simplified structure of communication systems. The function of such a system is to convey information from one point (the sender) to the other point (the destination).

Every communication system consists of three basic *sub-systems*: the emitter, the

Figure 1.5 A block-diagram of communication systems.

channel, and the receiver. The emitter, located at one point in space, is to generate a signal $x(t)$ that contains the message signal $s(t)$ (say a speech signal) produced by a source of information and to transmit it efficiently to the channel. The channel, as the physical medium, can be an optical fiber, a coaxial cable, or simply the air, and is to guide the transmitted signal $x(t)$ to the receiver that is located at some other point separate from the emitter. As the transmitted signal propagates over the channel, it is distorted before reaching the receiver due to many factors including the physical characteristics of the channel, noise and interfering signals from other sources. The objective of the receiver is to process the received signal $r(t)$, a distorted version of the transmitted signal $x(t)$, so as to yield a signal $y(t)$ which is in a recognizable form of the original message signal $s(t)$.

B. Control systems

Control engineering is another important area in which the theory and technology of *signals and systems* find successful applications. Such examples are ranged from simple appliances such as air-conditioners and refrigerators found in homes to very sophisticated engineering innovations such as aircraft autopilots, robots, paper mills, mass-transit vehicles, oil refineries, automobile engines, nuclear reactors, and power plants.

A block-diagram for a class of control systems is depicted in Figure 1.6:

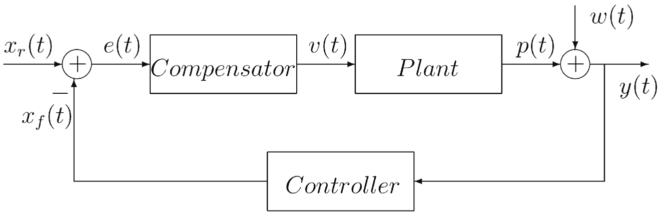


Figure 1.6 A block-diagram for a class of feedback control systems.

where the *plant* is the system to be controlled, $w(t)$ is the disturbance signal which plus the plant output forms the measurement signal $y(t)$, and the (feedback) *controller* is a system that consists of a sensor to collect the signal $y(t)$ and a micro-processor to generate the feedback signal $x_f(t)$. The latter is then to be compared with a *reference* signal $x_r(t)$ to produce an error signal $e(t) = x_r(t) - x_f(t)$. This error signal is then fed into the *compensator*, a system used to generate a signal $v(t)$ to control the plant such that a desired plant output $p(t)$ is achieved.

In an aircraft landing system, the plant refers to the aircraft's body and actuator.

The sensor system is used by the pilot to measure the lateral position of the aircraft. In this situation, $w(t)$ is the measurement error, while the reference input $x_r(t)$ corresponds to the desired landing path of the vehicle and the compensator is designed such that the output of the plant tracks $x_r(t)$ well.

C. Signal processing systems

As stated by Simon Haykin[1], signal processing is at its best when it successfully combines the unique ability of mathematics to generalize with both the insight and prior information gained from the underlying physics of the problem at hand.

In general, the functional form of a signal does not directly reveal the embedded information. One of the reasons for this is that due to factors such as measurement noise and channel distortion, the signal under processing is usually a corrupted version of the one which contains information. *Filtering*, one of the most important tools of signal processing, is a set of signal operations used to get rid of the disturbance such that information can be extracted easily.

In the system depicted in Figure 1.7, the input is a corrupted music signal $s(t) = s_0(t) + e(t)$, where $s_0(t)$ is the desired music signal and $e(t)$ is the noise attached somehow.

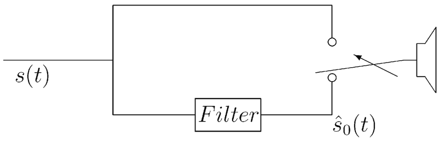


Figure 1.7 Block-diagram of a simplified audio play system.

None would enjoy listening to the sound from the louder speaker driven directly by $s(t)$. The filter is a system that is intended to block the noise $e(t)$ and let $s_0(t)$ pass through. It would be a completely different story if the speaker is excited by the output $\hat{s}_0(t) \approx s_0(t)$ of the filter.

A digital mixer is a more sophisticated audio instrument used for audio signal processing such as equalization, noise gating, and dynamic control. One of the most important parts of such a system is a set of filters, called filter bank.

1.2 Description and classification of signals

A signal, represented mathematically as a function of M independent variables, is usually referred to as an M-dimensional (M-D) signal. The speech signal $s(t)$ shown in Figure 1.1 is a 1-D signal, while the light density $p(x, y)$ of a picture (see Figure 1.2) is a 2-D signal.

---

[1] S.S. Haykin, "Signal processing: where physics and mathematics meet," *IEEE Signal Processing Magazine*, vol. 18, Issue 4, pp. 6-7, Jul., 2001.

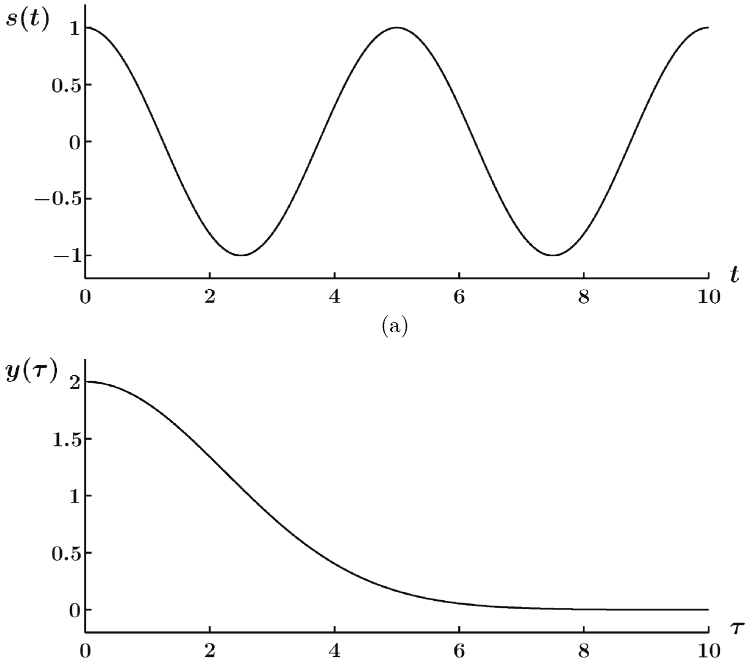
Define $\mathcal{R}$ as the set of all *real numbers*. In this book, we focus our attention mainly on 1-D signals that are defined as a *single-valued* function of an independent variable, say ξ, taking on a subset of $\mathcal{R}$, denoted as $\mathcal{R}_\xi$. *Single-valued* means that for any value of $\xi \in \mathcal{R}_\xi$, there is one and only one value of the function corresponding to it.

1.2.1 Continuous-time signals and discrete-time signals

This classification of signals is on the basis of how they are defined as a function of ξ. A signal x is said a *continuous-time* signal if the set $\mathcal{R}_\xi$ is continuous. Such a signal is denoted as $x(\xi)$.

Most of the signals generated naturally in our world are continuous. For example, a speech signal, represented by acoustic pressure as a function of the time t, is a continuous-time signal as the time variable t is always continuous. A continuous-time signal can also be generated artificially using computers with a given function. Figure 1.8 shows two continuous-time signals defined on $\mathcal{R}$

$$\left\{\begin{array}{lcl} s(t) & = & \cos(0.4\pi t),\ t \in \mathcal{R} \\ y(\tau) & = & 2\mathrm{e}^{-0.1\tau^2},\ \tau \in \mathcal{R} \end{array}\right. \tag{1.1}$$



(a)

(b)

Figure 1.8 Waveforms of the two signals defined in (1.1), both plotted for the interval $[0, 10]$.

It should be noted that a continuous-time signal does not necessarily imply that the underlying independent variable represents *time*. Let $x(h)$ be the variation of air pressure with altitude h of a spot on the earth. Such a signal is a continuous-time signal though the altitude h, which varies continuously, has nothing to do with the *time*. It is just for convenience that we generally refer to the independent variable of a signal as *time*.

Denote $\mathcal{Z}$ as the set of *all integers*, i.e.,

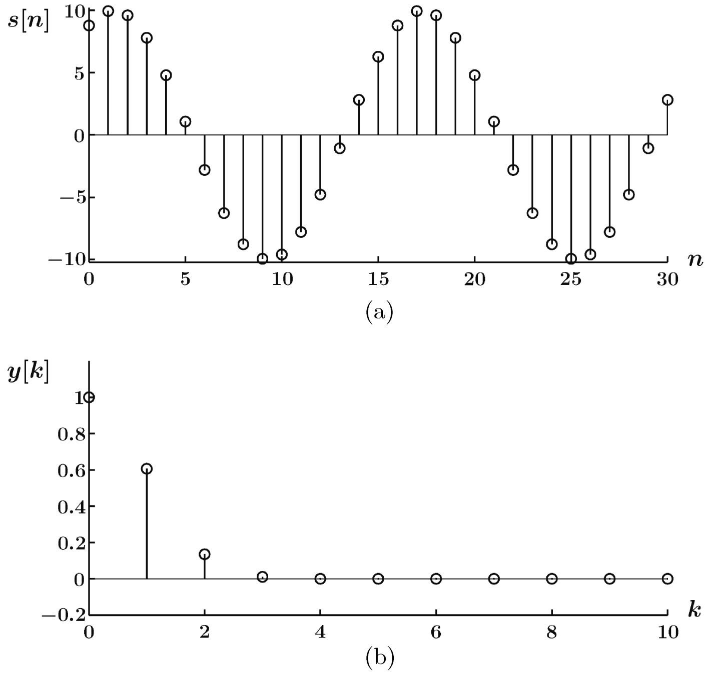
$$\mathcal{Z} \triangleq \{\cdots, -2, -1, 0, 1, 2, \cdots\}$$

A signal is said a *discrete-time* signal if the set $\mathcal{R}_\xi$ belongs to $\mathcal{Z}$.

Two examples of discrete-time signals defined on $\mathcal{Z}$ are given as

$$\begin{cases} s[n] &= 10\cos(0.125\pi n - 0.5), & n \in \mathcal{Z} \\ y[k] &= \mathrm{e}^{-0.5k^2}, & k \in \mathcal{Z} \end{cases} \tag{1.2}$$

and are shown in Figure 1.9, respectively.



(a)

(b)

Figure 1.9 Waveforms of the two discrete-time signals defined in Equation (1.2).

Roughly speaking, the set of discrete-time signals can be classified into two categories.

- In the first one, the discrete-time signals are inherently discrete. Figure 1.10 shows a daily averaged Shanghai Composite Index for a period of 40 days, where the independent variable n represents the nth day of that period. Another example of such discrete-time signals is an image signal obtained with a digital camera, where the gray level of the image is represented over a finite set of points, called *pixels*.

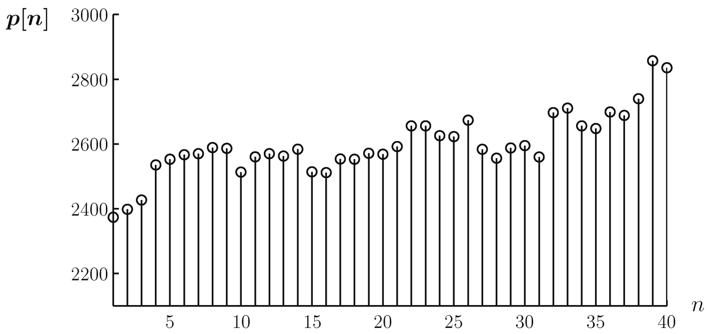


Figure 1.10 Daily average of Shanghai Composite Index from May 4 to July 1, 2010.

- In the second one, the discrete-time signals are obtained by sampling continuous-time signals $x(t)$ at some values of t. For example, if the continuous-time signal $x(t) = A\cos(10\pi t)$, defined on $[0, 1]$, is collected at the points $t_n = 1 - 2^{-n}$, then a discrete-time signal, denoted as $x[n]$, is obtained with $x[n] \triangleq x(t_n) = A\cos[10\pi(1 - 2^{-n})]$. Very often, sampling is done uniformly such as $t_n = nT_s$ for $n \in \mathcal{Z}$, where T_s is usually referred to as *sampling period.* Figure 1.11 shows $s(t) = 10\cos(20\pi t - 0.5)$ and its sampled version $s[n] = s(nT_s)$ with $T_s = 1/50$. More details for this topic will be given in Chapter 5.

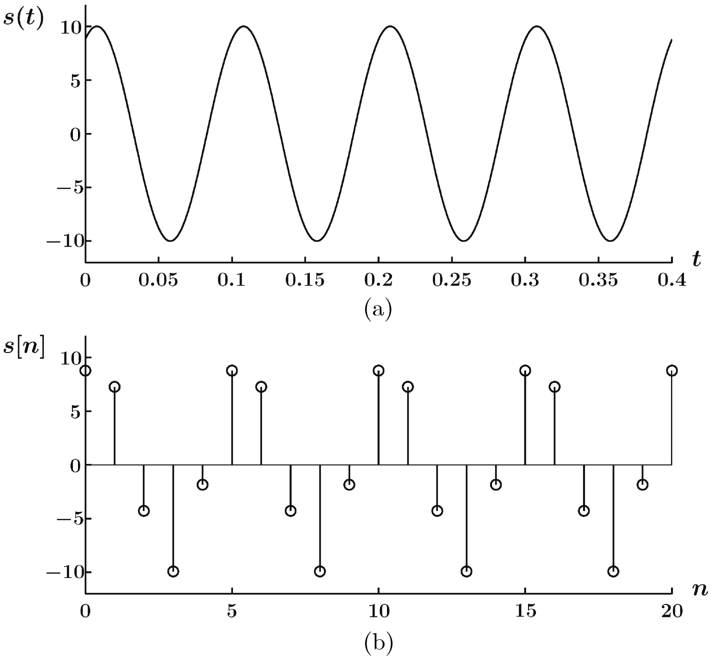


Figure 1.11 (a) $s(t) = 10\cos(20\pi t - 0.5),\ t \in [0,\ 0.4]$. (b) $s[n] = s(t_n)$ with $t_n = n/50$.

It should be noted that throughout this book, we use the letters like t, τ, f to denote continuous (independent) variables, while the letters like i, k, n, m, p, q for the independent variables of discrete-time signals. More importantly, as a convention we use the *parentheses* (.) and the *bracket* [.] to distinguish continuous-time (CT) signals from discrete-time (DT) signals.

1.2.2 Energy signals and power signals

In order to study signals, we have to find a proper *measure* or *norm* to quantify them. The absolute value, usually referred to as *magnitude*, is a proper quantity to measure how *big* a signal is if it is constant but such a measure can not give the overall view of the signal when it is time-varying.

Let $i(t)$ and $v(t)$ be the *current* and *voltage* which are through and across a resistor R of unit resistance (i.e., $R = 1$ ohm), respectively. As well known from physics, the *instantaneous power* is

$$p(t) \triangleq i(t)v(t) = v^2(t)$$

and the total *energy* expended/consumed by this resistor over the time interval (t_1, t_2) is given by

$$\int_{t_1}^{t_2} p(t)\mathrm{d}t = \int_{t_1}^{t_2} v^2(t)\mathrm{d}t$$

and the *average power* over (t_1, t_2) is

$$\frac{1}{t_2 - t_1}\int_{t_1}^{t_2} p(t)\mathrm{d}t = \frac{1}{t_2 - t_1}\int_{t_1}^{t_2} v^2(t)\mathrm{d}t$$

Borrowed from the above, we have the concepts of *energy* and *power* for a signal x, which are defined below

- *Energy* E_x: Denote

$$\left\{\begin{array}{lllr} E_x(T) & \triangleq & \int_{-T}^{T} |x(t)|^2\mathrm{d}t & - \quad \text{CT} \\ E_x[N] & \triangleq & \sum_{n=-N}^{N} |x[n]|^2 & - \quad \text{DT} \end{array}\right. \tag{1.3}$$

for $T > 0, N > 0$. The energy for a continuous-time signal $x(t)$ over $\mathcal{R}$ and a discrete-time signal $x[n]$ over $\mathcal{Z}$ is defined respectively as

$$\left\{\begin{array}{lllr} E_x & \triangleq & \lim_{T\to+\infty} E_x(T) = \int_{-\infty}^{+\infty} |x(t)|^2\mathrm{d}t & - \quad \text{CT} \\ E_x & \triangleq & \lim_{N\to+\infty} E_x[N] = \sum_{n=-\infty}^{+\infty} |x[n]|^2 & - \quad \text{DT} \end{array}\right. \tag{1.4}$$

- *Power* P_x: With $E_x(T), E_x[N]$ defined in (1.3), the power for a continuous-time signal $x(t)$ over $\mathcal{R}$ and a discrete-time signal $x[n]$ over $\mathcal{Z}$ is defined respectively as

$$\left\{\begin{array}{lllr} P_x & \triangleq & \lim_{T\to+\infty} \frac{E_x(T)}{2T} & - \quad \text{CT} \\ P_x & \triangleq & \lim_{N\to+\infty} \frac{E_x[N]}{2N+1} & - \quad \text{DT} \end{array}\right. \tag{1.5}$$

A signal x is said to be an *energy signal* if its energy satisfies $0 < E_x < +\infty$ and a *power signal*, if $0 < P_x < +\infty$.

The signals we have discussed so far are all real-valued. Though we get used to representing signals with real-valued functions, signal representations in complex-valued functions do have some advantages over the former, especially in signal analysis. This will be seen in the sequel.

As well known, a complex-valued number (or signal) $x \in \mathcal{C}$ consists of a real part and an imaginary part, which, for convenience, are denoted as $\mathcal{R}_e(x)$ and $\mathcal{I}_m(x)$, respectively. Let $\mathrm{j} \triangleq \sqrt{-1}$, then x can be represented into the *rectangular coordinates*, i.e., *Cartesian-form*:

$$x = \mathcal{R}_e(x) + \mathrm{j}\mathcal{I}_m(x) \quad \Rightarrow \quad |x|^2 = xx^* = \mathcal{R}_e^2(x) + \mathcal{I}_m^2(x) \tag{1.6}$$

where the subscript * denotes the complex conjugate operator[1]: $x^* = \mathcal{R}_e(x) - \mathrm{j}\mathcal{I}_m(x)$.

As well known, a *complex* number $c = \mathcal{R}_e(c) + \mathrm{j}\mathcal{I}_m(c)$ can be represented alternatively in the *Polar-form*

$$c = \rho \mathrm{e}^{\mathrm{j}\theta} \tag{1.7}$$

where both $\rho \geqslant 0$ and θ are real numbers, called *magnitude* and *phase*, respectively.

Using the famous *Euler's formula*:

$$\boxed{\mathrm{e}^{\mathrm{j}\theta} = \cos\theta + \mathrm{j}\sin\theta} \tag{1.8}$$

we have the following relationship between the *rectangular coordinates* and the *Polar-form*

$$\boxed{\left\{\begin{array}{lcl} \mathcal{R}_e(c) & = & \rho\cos\theta \\ \mathcal{I}_m(c) & = & \rho\sin\theta \end{array}\right. \Leftrightarrow \left\{\begin{array}{lcl} \rho & = & \sqrt{\mathcal{R}_e^2(c) + \mathcal{I}_m^2(c)} \\ \theta & = & \arctan\dfrac{\mathcal{I}_m(c)}{\mathcal{R}_e(c)} \end{array}\right.} \tag{1.9}$$

It should be noted that for a given c, the values of θ satisfying (1.7) are not uniques. This is because $\mathrm{e}^{\mathrm{j}(\theta+2\pi m)} = \mathrm{e}^{\mathrm{j}\theta}\mathrm{e}^{\mathrm{j}2\pi m} = \mathrm{e}^{\mathrm{j}\theta}$ for any integer m. Usually, *the phase* means the θ that satisfies

$$-\pi \leqslant \theta < \pi$$

It turns out from (1.8) that $\mathrm{e}^{-\mathrm{j}\theta} = \cos\theta - \mathrm{j}\sin\theta$ and hence

$$\boxed{\left\{\begin{array}{lcl} \cos\theta & = & \frac{1}{2}(\mathrm{e}^{\mathrm{j}\theta} + \mathrm{e}^{-\mathrm{j}\theta}) \\ \sin\theta & = & \frac{1}{\mathrm{j}^2}(\mathrm{e}^{\mathrm{j}\theta} - \mathrm{e}^{-\mathrm{j}\theta}) \end{array}\right.} \tag{1.10}$$

These expressions will be used frequently throughout this textbook.

---

[1] It is assumed throughout this textbook.