Introduction to Communication Systems

Chapter 1

Introduction

A communication system is used to transmit information from an information source to some distant agent requiring the information. We wish to reliably transmit this information with high fidelity. In other words, we would like the receiver to obtain the actual information or something that is highly correlated to it. In the case of analog communications, we try to minimize the mean square error between the transmitted signal and the received one. In the case of digital communications, we try to minimize the symbol error rate. To achieve this goal, we must design signals to represent the information we are trying to transmit as well as receivers that can easily identify these signals, even when they are corrupted by interference or noise.

Thermal noise is caused by random electron movement in the circuits of the receiver. Interference is caused by the operation of other electrical systems in the vicinity of our communication system. Noise and interference limit our ability to communicate error-free.

The block diagram of a typical communication system is shown below in Figure 1.1. The modulator converts the information signal, $m(t)$, into a signal, $s(t)$, that is suitable for transmission over the communications channel. The channel is the physical link between the transmitter and receiver, such as coaxial cable, or the air in a wireless communications system. Typical channel models include additive noise, to model the effect of the receiver's thermal noise contribution to the received signal. The demodulator attempts to decipher the received signal so that it can produce an estimate of what was actually transmitted, $m_{est}(t)$. When designing a communications system, the communications engineer must consider how much power is to be allocated to the transmitted signal, the system's bandwidth, the effect of noise and interference on the receiver's ability to detect and demodulate the signal, and the overall cost and complexity of the system. Usually some tradeoffs are required, such as sacrificing performance for cost.

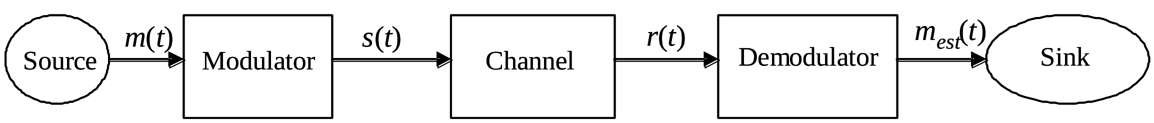


Figure 1.1 : Block diagram of a generic communication system.

1.1 Analog Communications

The information source produces an information signal $m(t)$, where $m(t)$ is the message signal. The modulator converts this signal into an analog signal $s(t)$. An analog signal is a signal that is continuous in both time and amplitude. At the receiver, the demodulator attempts to retrieve $m(t)$ from $r(t)$, which is the channel's response to input $s(t)$. Usually $r(t)$ is a corrupted form of $s(t)$, such as an attenuated version of $s(t)$ plus noise, which is expressed by $r(t) = \alpha s(t) + n(t)$, where α is the attenuation introduced by the channel (this attenuation may be time-varying, as is the case in wireless communications) and $n(t)$ is the noise from the receiver as well as any interference that may come from the channel. This received signal is also analog. Usually, the bandwidth of the noise signal is greater than that of the transmitted signal, $s(t)$, thus we can filter $r(t)$ to eliminate out-of-band noise, then we can use this filtered signal to estimate $m(t)$. Examples of analog communication systems that are commonly used are amplitude modulation (AM) and frequency modulation (FM) radio.

1.2 Digital Communications

A source produces an information signal that is made up of discrete symbols. For example, a computer in a local area network will produce a signal for transmission that is made up of bits. An analog source may be converted to a digital one by employing an analog to digital (A/D or A to D) converter. This consists of a sampler, quantizer and encoder. The modulator must convert this discrete time signal to one that is suitable for transmission over the channel. Discrete time signals have infinite bandwidth, therefore these signal must be converted to something that is continuous in time. However, the data symbols are usually represented by discrete voltages, amplitude levels, carrier phase values, frequencies or some combination of these. On a given signaling interval, the modulator output $s(t)$ is a waveform that comes from a discrete set of possible waveforms. At each new signaling interval, the waveform is changed to reflect the new input data symbol. The transmitted signal is then a sum of time shifted waveforms that come from a finite set of waveforms.

The receiver must observe the received signal $r(t)$ over each signaling interval. The received waveform is compared to all possible waveforms in order to determine which waveform was most likely transmitted. Since the transmitted signal on any given signaling interval can be only one of a finite set of possible waveforms, when the receiver correctly identifies which signal was transmitted, the noise and interference are completely eliminated. It is for this reason that digital communication systems relay information with a greater fidelity than analog ones.

In Figure 1.2, we demonstrate both analog and digital signals in the presence of noise. In the analog case, we can filter the out of-band noise by using a lowpass filter. The value of the transmitted signal comes from an infinite set of values between some minimum and some maximum. Thus, the filtered signal plus noise is not exactly the same as the original signal and so, there is a difference between the original signal and the filtered received one. In the digital case, the transmitted signal may only take on the values of ± 1 and it can only change values at the end of each signaling interval. Thus, when the signal is received with noise, as can be seen in Figure 1.2, it is easy to determine which value was most likely transmitted. We can then reconstruct the transmitted signal without error. Occasionally, the structure of the noise makes it difficult to determine which symbol was transmitted, and in these cases we sometimes incorrectly identify the transmitted symbol. It is the communication engineer’s task to minimize the mean square error between the transmitted signal and the received one in the analog case, and to minimize the symbol error probability in the digital case.

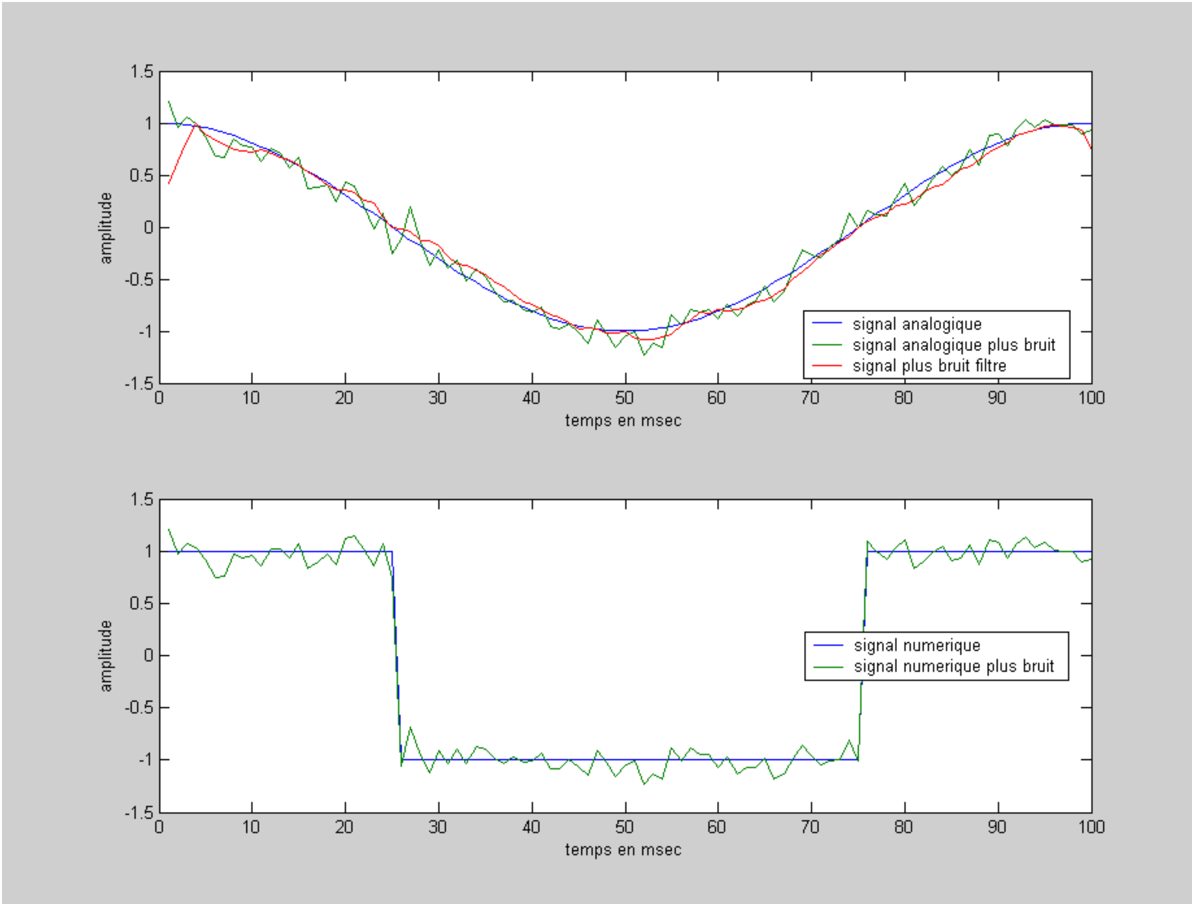


Figure 1.2 : Analog and Digital Communication Signals in the Presence of Noise.

1.3 Organization of the course

This course is organized as follows:

In chapter 2, we will review the concepts of signals and systems, focusing on linear time invariant (LTI) systems. In Chapter 3, we will discuss the representation of signals in the frequency domain. This includes the Fourier Series, the Fourier Transform, Frequency Response of LTI systems, Power and Energy Spectrum and Sampling Theory. In Chapter 4 we will examine some analog modulation techniques including amplitude modulation (AM), frequency modulation (FM) and their variants. In Chapter 5, we will introduce baseband digital modulation methods such as pulse amplitude modulation (PAM) and pulse coded modulation (PCM). We will discuss analog to digital (A/D) conversion and quantization as well as the Nyquist criteria. We will also discuss carrier based digital modulation schemes. In Chapter 6, we will introduce information and coding theory.

Chapter 2

Review of Signals and Systems

Communication theory relies heavily on the principles of signals and systems. Before we can introduce the concepts of communications, we must review the basic concepts from the signals and systems course as well as introduce some notations that will be used throughout the course.

2.1 Useful Signals

There are many signals that we use often in this course to demonstrate certain concepts. These signals are defined in this section.

1) **Impulse $\delta(t)$ (Dirac Delta Function)**

The impulse $\delta(t) = 0$ for $t \neq 0$, yet $\int_{-\infty}^{\infty} \delta(t)dt = 1$. The signal only has non-zero value at $t = 0$. Thus at $t = 0$ the function has infinite value but the duration of the impulse is infinitesimal. The impulse function is shown in figure 2.1.

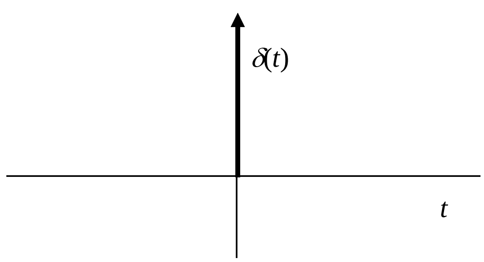


Figure 2.1 : Impulse Function.

Some properties of the impulse function

The impulse function is 0 for $t \neq 0$, but its integral is 1 if the limits of integration contain $t = 0$. Thus we can see that:

$$x(t)\delta(t) = x(0)\delta(t) \tag{2.1}$$

$$x(t)\delta(t-\tau) = x(\tau)\delta(t-\tau) \tag{2.2}$$

$$\int_{-\infty}^{\infty} x(t)\delta(t-\tau)dt = x(\tau) \tag{2.3}$$

$$\int_{-\infty}^{t} \delta(\lambda)d\lambda = \begin{cases} 0, & t < 0 \\ 1, & t > 0 \end{cases} \tag{2.4}$$

2) **Step Function $u(t)$**

The step function is used to describe switched signals. For $t < 0$, the switch is open and no current flows through the circuit. When $t = 0$, the switch is closed and current flows through the circuit. To describe this, for $t < 0$, $u(t) = 0$ and for $t > 0$, $u(t) = 1$. Therefore:

$$u(t) = \begin{cases} 0, & t < 0 \\ 1, & t > 0 \end{cases} = \int_{-\infty}^{t} \delta(\lambda)d\lambda \tag{2.5}$$

At $t = 0$, there is a discontinuity in the step function. Generally the value of a function at a discontinuity is the midpoint between the value of the function just before the discontinuity and the value just after the discontinuity. Thus $u(0) = ½\, u(0^-) + ½\, u(0^+) = ½$.

The step function is shown in figure 2.2.

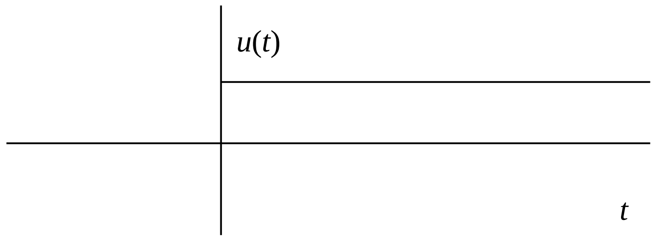


Figure 2.2 Step Function.

3) Rectangular Impulse $\Pi(t)$

The rectangular impulse function is generally used to mathematically describe some digital communication signals or the frequency response of ideal filters. It is given by the expression below:

$$\Pi(t) = \begin{cases} 1 & -\frac{1}{2}<t<\frac{1}{2} \\ 0 & \text{Otherwise} \end{cases} \tag{2.6}$$

Graphically, the rectangular impulse is shown in Figure 2.3.

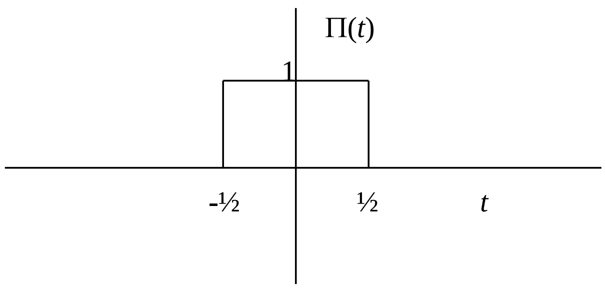


Figure 2.3: Rectangular Impulse Function.

4) Triangular Impulse $\Lambda(t)$

The triangular impulse function appears often in the autocorrelation function of digital communication signals. The autocorrelation function is used to determine a random signal's spectral density. It is given by the following expression:

$$\Lambda(t) = \begin{cases} 1-|t| & -1 < t < 1 \\ 0 & \text{Otherwise} \end{cases} \tag{2.7}$$

The triangular impulse is shown in Figure 2.4.

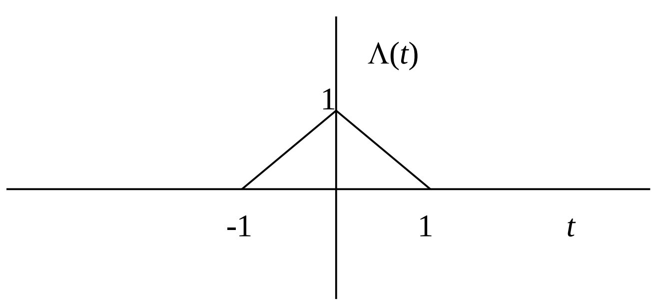


Figure 2.4 : Triangular Impulse Function.

5) Cardinal sine impulse (sinc)

The cardinal sine impulse sinc(t) is given by :

$$\mathrm{sinc}(t) = \frac{\sin(\pi t)}{\pi t} \tag{2.8}$$

The sinc(t) pulse shape is shown figure 2.5.

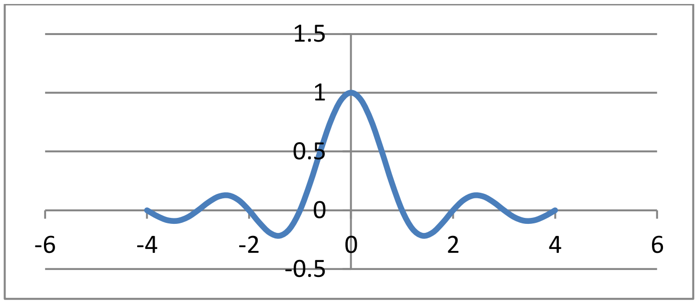


Figure 2.5 : sinc(t).

6) Squared carinal sine impulse

The pulse shape $\mathrm{sinc}^2(t)$ is given by :

$$sinc^2(t) = \frac{\sin^2(\pi t)}{(\pi t)^2} \tag{2.8}$$

It is shown in Fig. 2.6.

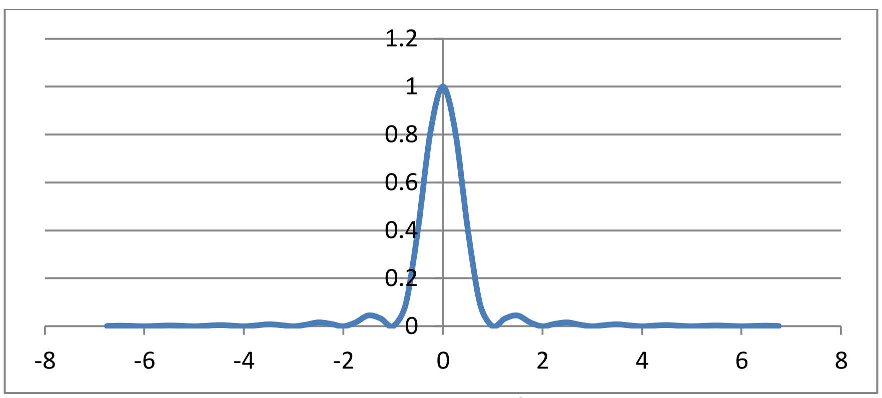


Figure 2.6 : $\mathrm{sinc}^2(t)$.

2.2 Convolution

We often use convolution in the analysis of communication signals. Given two functions, $x(t)$ and $y(t)$, the convolution of these two functions, $x(t)*y(t)$ is given by:

$$x(t) * y(t) = \int_{-\infty}^{\infty} x(\lambda) y(t-\lambda) d\lambda = \int_{-\infty}^{\infty} y(\lambda) x(t-\lambda) d\lambda = y(t) * x(t) \tag{2.9}$$

Exercise 2.1

Show that x(t)*y(t) = y(t)*x(t). (2.10)

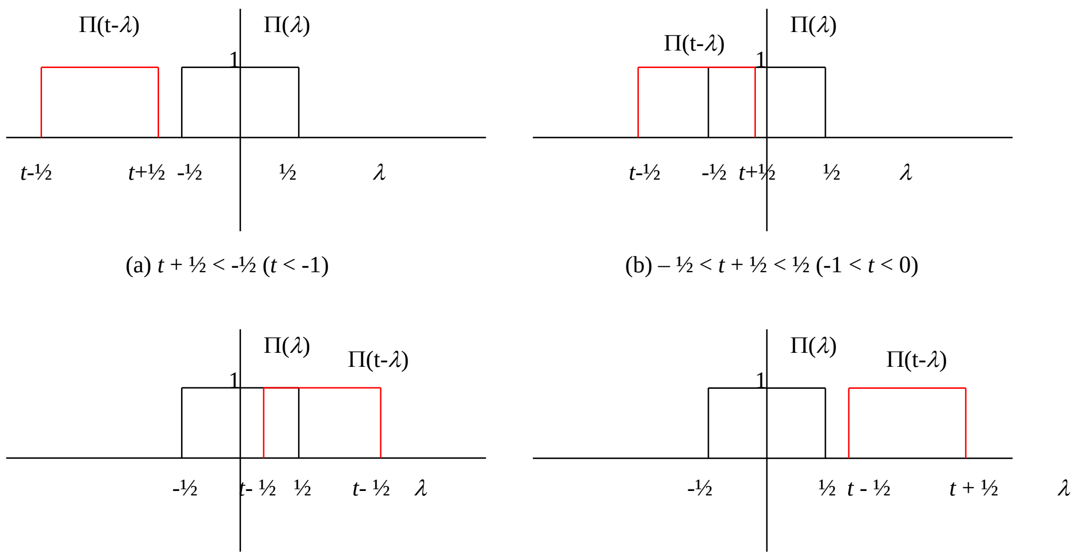
Exercise 2.2

Show that x(t)*(y(t)+z(t)) = x(t)*y(t) + x(t)*z(t) (2.11)

Example 2.1

Do the convolution $\Pi(t)*\Pi(t)$.

In Fig. 2,7, we show $\Pi(\lambda)$ and $\Pi(t\text{-}\lambda)$ as functions of λ for different values of t. We see that $\Pi(\lambda)\Pi(t\text{-}\lambda)$ depends on both t and λ.



(a) t + ½ < -½ (t < -1)

(b) – ½ < t + ½ < ½ (-1 < t < 0)

(c) – ½ < t - ½ < ½ (0 < t < 1)

(d) t - ½ > ½ (t > 1)

Figure 2.7 : $\Pi(\lambda)\Pi(t\text{-}\lambda)$ for (a) $t \le -1$, (b)$-1 \le t \le 0$, (c) $0 \le t \le 1$, (d) $t \ge 1$.

For $t < -1$ and $t > 1$, $\Pi(\lambda)\Pi(t\text{-}\lambda) = 0$ for all λ. Therefore for $t < -1$ and $t > 1$:

$$\int_{-\infty}^{\infty} \Pi(\lambda)\Pi(t-\lambda)d\lambda = 0$$

For the case where -1 < t < 0, $\Pi(\lambda)\Pi(t\text{-}\lambda) = 1$ for $-1/2 < \lambda < t+1/2$ and $\Pi(\lambda)\Pi(t\text{-}\lambda) = 0$ for all other values of λ. Therefore

$$\int_{-\infty}^{\infty} \Pi(\lambda)\Pi(t-\lambda)d\lambda = \int_{-1/2}^{t+1/2} d\lambda = \lambda|_{-1/2}^{t+1/2} = 1+t.$$

Similarly, for $0 < t < 1$, $\Pi(\lambda)\Pi(t\text{-}\lambda) = 1$ for $t-1/2 < \lambda < 1/2$ and $\Pi(\lambda)\Pi(t\text{-}\lambda) = 0$ otherwise. Therefore

$$\int_{-\infty}^{\infty} \Pi(\lambda)\Pi(t-\lambda)d\lambda = \int_{t-1/2}^{1/2} d\lambda = \lambda|_{t-1/2}^{1/2} = 1-t.$$

Therefore $\Pi(t)*\Pi(t)$ is given by:

$$\Pi(t)*\Pi(t) = \begin{cases} 0 & t < -1 \\ 1+t & -1 < t < 0 \\ 1-t & 0 < t < 1 \\ 0 & t > 1 \end{cases} = \begin{cases} 1-|t| & -1 < t < 1 \\ 0 & \text{Otherwise} \end{cases} = \Lambda(t) \quad (2.12)$$

2.3 Linearity

An electrical system has input, $x(t)$ and output $y(t)$. The output $y(t)$ is some function of the input $x(t)$. We express the output $y(t)$ by the following equation:

$$y(t) = H(x(t)) \tag{2.13}$$

where H( ) is some function that is performed on $x(t)$.

We say that the system is linear if the principle of superposition applies. In other words, the system is linear if for any input $x_3(t) = \alpha x_1(t)+\beta x_2(t)$, the output $y_3(t)$ is given by :

$$\begin{aligned} y_3(t) &= H\big(x_3(t)\big) \\ &= H(\alpha x_1(t) + \beta x_2(t)) \\ &= \alpha H\big(x_1(t)\big) + \beta H(x_2(t)) \\ &= \alpha y_1(t) + \beta y_2(t) \end{aligned} \tag{2.14}$$

where $y_1(t) = \mathrm{H}(x_1(t))$ et $y_2(t) = \mathrm{H}(x_2(t))$. If (2.14) is not true then the system is not linear.

Example 2.2

Consider a system with the input-output relationship given by: $y(t) = x^2(t)$. For the input $x_1(t)$, the output is $y_1(t) = x_1^2(t)$ and for input $x_2(t)$, the output is $y_2(t) = x_2^2(t)$. Let $x_3(t) = \alpha x_1(t) + \beta x_2(t)$, then the output is $y_3(t) = x_3^2(t) = (\alpha x_1(t) + \beta x_2(t))^2 = \alpha^2 x_1^2(t) + 2\alpha\beta x_1(t)x_2(t) + \beta^2 x_2^2(t)$. For the system to be linear, $y_3(t)$ must equal $\alpha y_1(t) + \beta y_2(t) = \alpha x_1^2(t) + \beta x_2^2(t) \neq y_3(t)$; therefore the system is not linear.

Example 2.3

The input output relationship of a system is $y(t) = tx(t)$. Let $x_3(t) = \alpha x_1(t) + \beta x_2(t)$, therefore the output is, $y_3(t) = t(\alpha x_1(t) + \beta x_2(t)) = \alpha(tx_1(t)) + \beta(tx_2(t)) = \alpha y_1(t) + \beta y_2(t)$. Therefore the system is linear.