CSE 4214: Digital Communication

Lecture Notes

Andrew W. Eckford

Department of Computer Science and Engineering

York University, Toronto, Ontario, Canada

Contents

Review: Probability, Random Processes, and Linear Systems

1.1. Probability

In this section, we briefly review some necessary concepts of probability that will be used throughout this text.

1.1.1. Foundations. Basics of probability; joint probability; independence; conditional probability; Bayes' rule.

1.1.2. Discrete-valued random variables. Probability mass function; expected value; mean and variance; examples.

1.1.3. Continuous-valued random variables. Probability density function; expected value; mean and variance; examples.

1.1.4. The Gaussian distribution. Definition; properties (e.g., even function).

A Gaussian random variable x with with mean μ and variance σ^2 has a probability density function given by

$$(1.1) \qquad f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right).$$

Integrals over this pdf may be expressed in terms of the *error function complementary*, erfc($\cdot$), which is defined as

$$(1.2) \qquad \mathrm{erfc}(z) = \frac{2}{\sqrt{\pi}} \int_{t=z}^{\infty} \exp(-t^2)dt.$$

The function erfc($\cdot$) has the following mathematical interpretation: if t is a Gaussian random variable with mean $\mu = 0$ and variance $\sigma^2 = 1/2$, then $\mathrm{erfc}(z) = \Pr(|t| > z)$. Furthermore, due to the symmetry of the Gaussian pdf about the mean, we illustrate in Figure X that

$$(1.3) \qquad \Pr(t > z) = \Pr(t < z) = \frac{1}{2}\mathrm{erfc}(z).$$

Using a change of variables, erfc($\cdot$) may be used to calculate an arbitrary Gaussian integral. For instance, for the random variable x with pdf $f(x)$ in (1.1), suppose we want to calculate the probability $\Pr(x > z)$. This probability can be expressed as

$$(1.4) \qquad \Pr(x > z) \;=\; \int_{x=z}^{\infty} f(x)dx$$

$$(1.5) \qquad \qquad \;=\; \int_{x=z}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) dx.$$

Now we make the substitution

$$(1.6) \qquad t = \frac{x - \mu}{\sqrt{2\sigma^2}}.$$

To perform a change of variables in an integral, we need to replace both x and dx with the equivalent functions of t. Solving for x, we have that

$$(1.7) \qquad x = \sqrt{2\sigma^2} t + \mu,$$

so taking the first derivative of x with respect to t, dx is given by

$$(1.8) \qquad dx = \sqrt{2\sigma^2} dt.$$

Substituting (1.7)-(1.8) into (1.5), we get

$$(1.9) \qquad \Pr(x > z) \;=\; \int_{x=z}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x - \mu)^2 \right) dx$$

$$(1.10) \qquad =\; \int_{\sqrt{2\sigma^2}t + \mu = z}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-t^2 \right) \sqrt{2\sigma^2} dt$$

$$(1.11) \qquad =\; \int_{t=(z-\mu)/\sqrt{2\sigma^2}}^{\infty} \frac{1}{\sqrt{\pi}} \exp\left(-t^2 \right) dt$$

$$(1.12) \qquad =\; \frac{1}{2}\mathrm{erfc}\left(\frac{z - \mu}{\sqrt{2\sigma^2}} \right).$$

1.2. Discrete-Time Random Processes

There are many ways to define a random process, but for our purposes, the following is sufficient:

- A random process is a function of time $X(t)$, so that for each fixed time t^*, $X(t^*)$ is a random variable.

As a result, we can write the probability density function (pdf) of the random process at any given time. For example, $f_{X(t^*)}(x)$ represents the pdf of the random process at time t^*. Joint probability density functions measure the joint probability of the process at k different times; these are called k*th order statistics* of the random process. For example, for $k = 2$ and times t_1 and t_2, we can write the second order statistics as $f_{X(t_1),X(t_2)}(x_1, x_2)$.

1.2.1. Definition, Mean, and Variance. It's easy to imagine a random process in discrete time, as merely a sequence of random variables, one for each time interval. For instance, consider the following two random processes defined at integer times $t \in \{\ldots, -2, -1, 0, 1, 2, \ldots\}$:

EXAMPLE 1.1. *At each time* $t \in \{\ldots, -2, -1, 0, 1, 2, \ldots\}$, *a fair coin is flipped. If the coin shows heads after the flip at time t, then $X(t) = 1$; otherwise, $X(t) = -1$. Thus, for any integer t^*, we can write*

$$f_{X(t^*)}(x) = \begin{cases} 0.5, & x = +1; \\ 0.5, & x = -1; \\ 0 & \text{otherwise.} \end{cases}$$

Since, at each fixed time t, the random process is a random variable, we can calculate the mean and variance of the process at each fixed time as usual for random variables. Thus, for the process as a whole, the mean and variance for a

random process are calculated as functions of time. For instance, for the process in Example 1.1, the mean of this process is given by

$$\begin{aligned} \mu(t) &= \sum_{x \in \{+1,-1\}} x f_{X(t)}(x) \\ &= (+1)(0.5) + (-1)(0.5) \\ &= 0 \end{aligned}$$

for all t. The variance of the process is given by

$$\begin{aligned} \sigma^2(t) &= \sum_{x \in \{+1,-1\}} (x - \mu(t))^2 f_{X(t)}(x) \\ &= (+1 - 0)^2(0.5) + (-1 - 0)^2(0.5) \\ &= 1 \end{aligned}$$

for all t.

As an alternative, the following more compicated example has mean and variance that are non-trivial functions of time:

EXAMPLE 1.2. *Let $X(0) = 0$. For each $t \in \{1, 2, \ldots\}$, a fair coin is flipped. If the coin shows heads after the flip at time t, then $X(t) = X(t-1) + 1$; otherwise, $X(t) = X(t-1)$.*

For any t, it is clear that $X(t)$ is the number of heads in the previous t trials. Such random variables are represented by the binomial distribution [1]. *Thus,*

$$f_{X(t)}(x) = \binom{t}{x} \frac{1}{2^t}.$$

The mean of this random process is given by

$$\mu(t) = \frac{t}{2},$$

and the variance is given by

$$\sigma^2(t) = \frac{t}{4}.$$

The reader is asked to prove these values in the exercises.

Instances of the random processes from Examples 1.1 and 1.2 are given in Figure 1.1.

1.2.2. Autocorrelation. Suppose you wanted a measure of correlation between two random variables, X_1 and X_2, with the same mean $\mu = 0$ and the same variance $\sigma^2 > 0$. As a candidate for this measure, consider

(1.13) $$R = E[X_1 X_2].$$

If the random variables are independent (i.e., uncorrelated), then since $E[X_1 X_2] = E[X_1]E[X_2]$ for independent random variables, we would have

$$R = E[X_1]E[X_2] = \mu^2 = 0,$$

bearing in mind that each of the random variables are zero mean. On the other hand, if the two random variables are completely correlated (i.e., $X_1 = X_2$), we would have

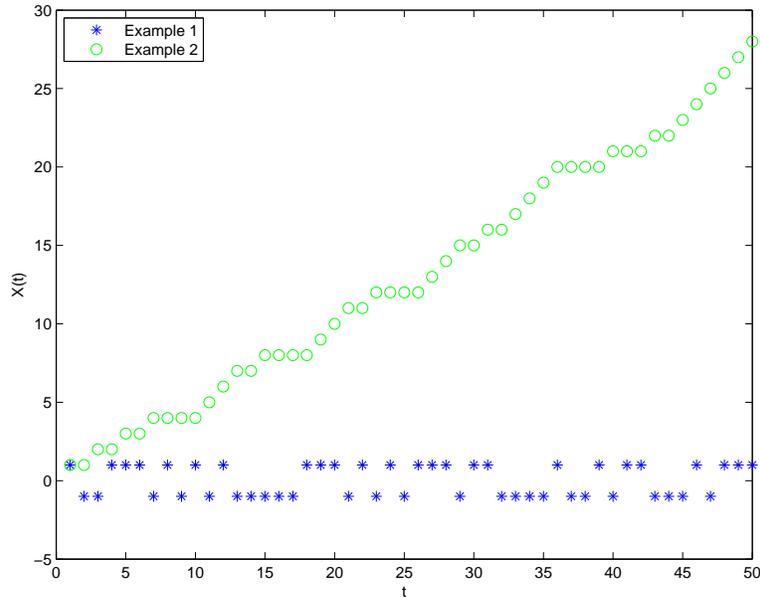$$R = E[X_1 X_2] = E[X_1^2] = \sigma^2.$$

FIGURE 1.1.   Illustration of the discrete-time random processes from Examples 1.1 and 1.2.

Further, if they were completely anticorrelated (i.e., $X_1 = -X_2$), it is easy to see that $R = -\sigma^2$.

This measure of correlation also has the following nice property:

THEOREM 1.1. *Given the above definitions, $|R| \leq \sigma^2$.*

Proof: Start with $E[(X_1 + X_2)^2]$. We can write:

$$
\begin{aligned}
E[(X_1 + X_2)^2] &= E[X_1^2 + 2X_1 X_2 + X_2^2] \\
&= E[X_1^2] + 2E[X_1 X_2] + E[X_2^2] \\
&= \sigma^2 + 2R + \sigma^2 \\
&= 2\sigma^2 + 2R.
\end{aligned}
$$

Since $(X_1 + X_2)^2 \geq 0$ for all X_1 and X_2, it is true that $E[(X_1 + X_2)^2] \geq 0$. Thus, $2\sigma^2 + 2R \geq 0$, so $R \geq -\sigma^2$. Repeating the same procedure but starting with $E[(X_1 - X_2)^2]$, we have that $R \leq \sigma^2$, and the theorem follows.   ∎
Since $R = 0$ when X_1 and X_2 are independent, $R = \sigma^2$ (the maximum possible value) when they are completely correlated, and $R = -\sigma^2$ (the minimum possible value) when they are completely anticorrelated, R is a good candidate for a correlation measure. The magnitude of R indicates the degree of correlation between X_1 and X_2, while the sign indicates whether the variables are correlated or anticorrelated. Properties of this correlation measure when the variances are unequal, or when the means are nonzero, are considered in the exercises.

We apply this correlation measure to different time instants of the same random process, which we refer to as the *autocorrelation*. In particular, let $X(t)$ be a discrete-time random process defined on $t \in \{\ldots, -2, -1, 0, 1, 2, \ldots\}$. Then the

autocorrelation between $X(t_1)$ and $X(t_2)$ is defined as

(1.14) $$R(t_1, t_2) = E[X(t_1)X(t_2)].$$

Note the similarity with (1.13), since $X(t)$ is merely a random variable for each time t. For the same reason, $R(t_1, t_2)$ has all the same properties as R.

1.2.3. Stationary random processes. A *stationary* discrete-time random process is a process for which the statistics do not change with time. Formally, a process is stationary if and only if

(1.15)
$$f_{X(t_1), X(t_2), \dots, X(t_k)}(x_1, x_2, \dots, x_k) = f_{X(t_1+\tau), X(t_2+\tau), \dots, X(t_k+\tau)}(x_1, x_2, \dots, x_k)$$

for all $k \in \{1, 2, \dots\}$ and all $\tau \in \{\dots, -2, -1, 0, 1, 2, \dots\}$. This does *not* imply that the process $X(t)$ is constant with respect to time, only that the statistical variation of the process is the same, regardless of when you examine the process. The process in Example 1.1 is stationary; intuitively, this is because we keep flipping the same unchanging coin, and recording the outcome in the same way at all t.

We now examine the effects of stationarity on the mean, variance, and autocorrelation of a discrete-time random process $X(t)$. The mean $\mu(t)$ is calculated as follows:

$$\begin{aligned} \mu(t) &= \int_x x f_{X(t)}(x) \mathrm{d}x \\ &= \int_x x f_{X(t+\tau)}(x) \mathrm{d}x \\ &= \mu(t+\tau), \end{aligned}$$

where the second line follows from the fact that $f_{X(t)} = f_{X(t+\tau)}$ for all $\tau \in \{\dots, -2, -1, 0, 1, 2, \dots\}$. Thus, $\mu(t) = \mu(t+\tau)$ for all τ, so $\mu(t)$ must be a constant with respect to t. Using a similar line of reasoning, we can show that $\sigma^2(t)$ is a constant with respect to t. Thus, for stationary random processes, we will write $\mu(t) = \mu$ and $\sigma^2(t) = \sigma^2$ for all t.

For the autocorrelation, we can write

$$\begin{aligned} R(t_1, t_2) &= E[X(t_1)X(t_2)] \\ &= \int_{x_1} \int_{x_2} x_1 x_2 f_{X(t_1), X(t_2)}(x_1, x_2) \mathrm{d}x_2 \mathrm{d}x_1 \end{aligned}$$

(1.16)

(1.17) $$= \int_{x_1} \int_{x_2} x_1 x_2 f_{X(t_1+\tau), X(t_2+\tau)}(x_1, x_2) \mathrm{d}x_2 \mathrm{d}x_1.$$

Let $\tau = \tau' - t_1$. Substituting back into (1.17), we have

$$\begin{aligned} R(t_1, t_2) &= \int_{x_1} \int_{x_2} x_1 x_2 f_{X(t_1+\tau'-t_1), X(t_2+\tau'-t_1)}(x_1, x_2) \mathrm{d}x_2 \mathrm{d}x_1 \end{aligned}$$

(1.18) $$= \int_{x_1} \int_{x_2} x_1 x_2 f_{X(\tau'), X(t_2-t_1+\tau')}(x_1, x_2) \mathrm{d}x_2 \mathrm{d}x_1.$$

However, in (1.18), since $X(t)$ is stationary, $f_{X(\tau'), X(t_2-t_1+\tau')}(x_1, x_2)$ does not change for any value of τ'. Thus, setting $\tau' = 0$, we can write

$$R(t_1, t_2) = \int_{x_1} \int_{x_2} x_1 x_2 f_{X(0), X(t_2-t_1)}(x_1, x_2) \mathrm{d}x_2 \mathrm{d}x_1,$$

which is not dependent on the exact values of t_1 or t_2, but only on the difference $t_2 - t_1$. As a result, we can redefine the autocorrelation function for stationary random processes as $R(t_2 - t_1)$; further, reusing τ to represent this difference, we will usually write $R(\tau)$, where

$$R(\tau) = E[X(t)X(t + \tau)]$$

for all t.

The properties that $\mu(t) = \mu$, $\sigma^2(t) = \sigma^2$, and $R(t_1, t_2) = R(t_2 - t_1)$ apply only to the first and second order statistics of the process $X(t)$. In order to verify whether a process is stationary, it is necessary to prove the condition (1.15) for every order of statistics. In general this is a difficult task. However, in some circumstances, only first and second order statistics are required. In this case, we define a *wide-sense stationary* (WSS) process as any process which satisfies the first and second order conditions of $\mu(t) = \mu$, $\sigma^2(t) = \sigma^2$, and $R(t_1, t_2) = R(t_2 - t_1)$. We have shown that all stationary processes are WSS, but it should seem clear that a WSS process is not necessarily stationary.

1.2.4. Power spectral density. For a wide-sense stationary random process, the *power spectral density* (PSD) of that process is the Fourier transform of the autocorrelation function:

$$(1.19) \qquad S_x(j\omega) = \mathcal{F}[R_x(\tau)] = \int_{\tau=-\infty}^{\infty} R_x(\tau)e^{-j\omega\tau}\,d\tau.$$

Properties of PSD:

$$(1.20) \qquad \mathrm{Var}(x[k]) = R_x(0) = \frac{1}{2\pi}\int_{-\infty}^{\infty} S_x(j\omega)\,d\omega.$$

1.3. Linear systems

1.3.1. Review of linear systems.

1.3.2. Linear systems and random processes. Apply a linear filter with frequency-domain transfer function $H(j\omega)$ to a wide-sense stationary random process with PSD $S_x(j\omega)$. The output is a random process with PSD $S_w(j\omega)$, where

$$(1.21) \qquad S_w(j\omega) = S_x(j\omega)|H(j\omega)|^2.$$

1.4. Problems

(1) For the random process in Example 1.2, show that $\mu(t) = t/2$, and $\sigma^2(t) = t/4$. Is this process stationary? Explain.

(2) Suppose X_1 and X_2 are zero-mean random variables with variances σ_1^2 and σ_2^2, respectively. For the correlation measure R defined in (1.13), show that
$$|R| \le \frac{\sigma_1^2 + \sigma_2^2}{2}.$$

(3) Suppose X_1 and X_2 have the same *nonzero* mean μ, and the same variance σ^2. For the correlation measure R defined in (1.13), show that $|R| \le \sigma^2 + \mu^2$.

(4) Give an example of a discrete-time random process for which $\mu(t) = \mu$ and $\sigma^2(t) = \sigma^2$ for all t, but there exist t_1 and t_2 such that $R(t_1, t_2) \ne R(t_2 - t_1)$.

(5) Calculate $\mu(t)$ and $R(t_1, t_2)$ for the continuous time random process given in Example **??**. Is this process stationary? Explain.

(6) Let $X(t) = X \sin(2\pi t)$, where X is a random variable corresponding to the result of a single fair coin flip: $X = 1$ if the coin is heads, and $X = -1$ is the coin is tails. Does $X(t)$ satisfy the definition of a continuous-time random process? If so, calculate $f_{X(t)}(x)$; if not, explain why not.

1.5. Laboratory Exercise: Probability and Random Processes

In this laboratory exercise, you will investigate the properties of discrete-valued random variables and random processes.

1.5.1. Generating arbitrary random variables. Let x be a discrete-valued random variable, taking values on 1, 2, ..., 6, with probability mass function $p(x)$.

- MATLAB provides a routine, `rand`, which generates uniformly distributed random variables on the range from 0 to 1. Given $p(x)$, propose a way to generate instances of x, with probabilities $p(x)$, from `rand`.
- Write a MATLAB function, called `xrand`, implementing the method you describe. The routine takes a 1×6 vector, where the first element of the vector is $p(1)$, the second is $p(2)$, and so on. The routine returns a value on 1, 2, ..., 6 at random according to the probabilities $p(x)$.

Discussion of empirical distributions.

Given a distribution, write a function to calculate the mean and variance, both empirically and theoretically.

Consider the following Gaussian random process: ... Plot the autocorrelation, both empirically and

Baseband Data Transmission

2.1. Hardware Model of Communication

2.1.1. Traditional communication system model.

2.1.2. Transmitter hardware model.

2.1.3. Receiver hardware model.

2.2. Noise

Although there are many sources of random distortion in communication systems, we will assume that the dominant source is *thermal noise*, arising from the random motion of electrons in electrical components. This random motion results in small current fluctuations, which can be significant in the presence of a very weak signal. Because there are many independently-moving electrons, all of which contribute randomly to the current fluctuations, the *central limit theorem* [**2**] can be used to model noise as a Gaussian random variable.

Gaussian noise has zero mean. The autocorrelation of a continuous-time Gaussian noise process $n(t)$ is given by

$$(2.1) \qquad R_n(\tau) = \frac{N_0}{2}\delta(\tau),$$

and is thus wide-sense stationary. Taking the Fourier transform of $R_n(\tau)$, its power spectral density is given by

$$(2.2) \qquad S_n(j\omega) = \frac{N_0}{2},$$

where N_0 is a constant proportional to the temperature of the device. From (2.2), the power spectrum is the same at all frequencies ω, so the noise is "white", analogously to white light; furthermore, the current fluctuations are added to whatever deterministic signal is present. Thus, we refer to this type of noise as *additive white Gaussian noise* (AWGN).

In discrete time, the received signal $y[k]$ is given by

$$(2.3) \qquad y[k] = As[k] + n[k],$$

where $s[k]$ represents the signal, A is a scaling factor representing signal attenuation/amplification, and $n[k]$ is the sampled version of the continuous-time AWGN process $n(t)$.

We will now determine the properties of $n[k]$. In the hardware model, we precede the A-to-D converter with an ideal lowpass filter having frequency-domain

transfer function $H_{LP}(j\omega)$, where

$$(2.4) \qquad H_{LP}(j\omega) = \begin{cases} 1, & |\omega| < \pi/T_s, \\ 0 & \text{otherwise.} \end{cases}$$

That is, the cutoff frequency of the filter is the Nyquist frequency $1/2T_s$. Let w(t) represent the noise random process at the output of the ideal lowpass filter, and let $S_w(j\omega)$ represent its PSD. From (1.21), $S_w(j\omega)$ is given by

$$(2.5) \qquad S_w(j\omega) \;=\; S_n(j\omega)|H_{LP}(j\omega)|^2$$

$$(2.6) \qquad \qquad =\; \begin{cases} N_0/2, & |\omega| < \pi/T_s, \\ 0 & \text{otherwise.} \end{cases}$$

Thus, from (1.20), the variance of $w(t)$ (and hence each sample $w[k]$) is given by

$$(2.7) \qquad \mathrm{Var}(w(t)) \;=\; \mathrm{Var}(w[k])$$

$$(2.8) \qquad \qquad =\; \frac{1}{2\pi}\int_{\omega=-\infty}^{\infty} S_w(j\omega)d\omega$$

$$(2.9) \qquad \qquad =\; \frac{1}{2\pi}\cdot\frac{2\pi}{T_s}\cdot\frac{N_0}{2}$$

$$(2.10) \qquad \qquad =\; \frac{N_0}{2T_s}.$$

We also need to know whether the samples $w[k]$ are independent. Taking the inverse Fourier transform of $S_w(j\omega)$, since $S_w(j\omega)$ is rectangular, we have that (see Appendix A)

$$(2.11) \qquad R_x(\tau) \;=\; \mathcal{F}^{-1}\left[S_w(j\omega)\right]$$

$$(2.12) \qquad \qquad =\; \frac{1}{T_s}\mathrm{sinc}\left(\frac{\tau}{T_s}\right).$$

If τ is a multiple of the sampling frequency, we have

$$(2.13) \qquad R_x(kT_s) \;=\; \frac{1}{T_s}\mathrm{sinc}\left(k\right)$$

$$(2.14) \qquad \qquad =\; \begin{cases} \frac{1}{T_s}, & k = 0, \\ 0, & k \neq 0. \end{cases}$$

Thus, recalling our discussion of correlation in Chapter 1, the noise process $w[k]$ is *uncorrelated* from sample to sample. Since $w[k]$ is a Gaussian random process, this is sufficient to show that it is *independent* from sample to sample.

2.3. Modulation and Detection

2.3.1. Modulation. Data can be represented as an arbitrarily long vector of binary $\{0,1\}$ symbols, as in Figure X. The goal of *modulation* is to transform such a vector into a function of time, which is necessary before the bit can be transmitted over the medium.

Since this book deals with discrete-time signal processing, the modulator will replace each bit with a non-overlapping sequence of samples; the function of time will then be generated in digital-to-analog conversion. From Section 2.1, each bit consists of $n_b = T_b/T_s$ samples, so we should define two sequences of n_b samples each: one to represent 0, and one to represent 1. Let $s_0[k]$ and $s_1[k]$ represent these sequences for bits 0 and 1, respectively.

The following examples illustrate modulation in detail.

EXAMPLE 2.1 (Polar nonreturn-to-zero). *In polar nonreturn-to-zero (NRZ), for any n_b, let*

$$(2.15) \qquad s_0[k] = \begin{cases} 1, & 1 \le k \le n_b, \\ 0, & \text{otherwise,} \end{cases}$$

and

$$(2.16) \qquad s_1[k] = -s_0[k].$$

EXAMPLE 2.2 (Binary phase shift keying). *In binary phase shift keying (BPSK), suppose for 0, the bit time T_b contains exactly one cycle of a sinusoid; thus, there would be exactly one cycle over n_b samples. We can write*

$$(2.17) \qquad s_0[k] = \begin{cases} \sin(2\pi k/n_b), & 1 \le k \le n_b, \\ 0, & \text{otherwise.} \end{cases}$$

Furthermore, let

$$(2.18) \qquad s_1[k] = \begin{cases} \sin(2\pi k/n_b + \pi), & 1 \le k \le n_b, \\ 0, & \text{otherwise,} \end{cases}$$

as depicted in Figure X. From (2.18), the phase is shifted by π radians in order to transmit a 1, hence the terminology. However, also note that $s_1[k] = -s_0[k]$.

Now let $\mathbf{b} = [0, 1, 1, 0, 1]$, and let $n_b = 5$. Replacing 0 and 1 with $s_0[k]$ and $s_1[k]$, respectively, from each example, we obtain discrete-time signals depicted in Figure X.

2.3.2. Detection. From (2.3), the received signal is corrupted by an AWGN random process $n[k]$. The detector's job is to extract the value of the bit, 0 or 1, from the noisy signal $y[k]$, as accurately as possible.

The detector consists of two components: a *filter*, which performs signal processing on $y[k]$, and a *decision device*, which takes the output of the filter and determines whether a 0 or 1 was sent. Typically, the decision device examines the output value of the filter after each bit has been sent (i.e., at integer multiples of the bit time T_b, or integer multiples of n_b in discrete time); we will call these values the *filter outputs*. We will assume that the filter is linear and time invariant, and has impulse response $h[k]$. The filter outputs $\phi(j)$ are given by

$$(2.19) \qquad \phi(j) = [y[k] \star h[k]]_{jn_b} \,,$$

where the notation $\star$ represents discrete-time convlution, and the notation $[\cdot]_{jn_b}$ indicates that the expression is evaluated at time jn_b.

Given $\phi(j)$, the decision device then assigns a bit, 0 or 1, to each possible output of the filter. This is usually done through a threshold (i.e., the bit is a 0 if the filter output exceeds the threshold, or a 1 if the filter output is less than the threshold). Let $d_z(x)$ represent the decision function on x with threshold z, where

$$(2.20) \qquad d_z(x) = \begin{cases} 0, & x \ge z, \\ 1, & x < z. \end{cases}$$

Thus, combining (2.19)-(2.20), the estimated bits $\hat{b}_j$ are given by

$$(2.21) \qquad \hat{b}_j = d_z(\phi(j)).$$

Selection of optimal $h[k]$ and z are non-trivial design problems, which will be discussed extensively in the remainder of the book. However, the following example illustrates a correctly designed detector, and its outputs in a noise-free channel.

EXAMPLE 2.3. *Let $s_0[t]$ and $s_1[t]$ be polar NRZ pulses, as defined in (2.15)-(2.16). Let*

$$(2.22) \qquad\qquad h[k] = \begin{cases} 1, & 0 \leq k < n_b, \\ 0, & \text{otherwise.} \end{cases}$$

Note that $h[k] = s_0[k+1]$.

In the absence of noise, the filter output in response to $s_0[k]$ at time n_b is

$$(2.23) \qquad\qquad [s_0[k] \star h[k]]_{n_b} = \sum_{i=-\infty}^{\infty} s_0[i]h[n_b - i]$$

$$(2.24) \qquad\qquad\qquad\qquad = \sum_{i=1}^{n_b} 1$$

$$(2.25) \qquad\qquad\qquad\qquad = n_b.$$

Since $s_1[k] = -s_0[k]$, the filter output in response to $s_1[k]$ is given by

$$(2.26) \qquad\qquad [s_1[k] \star h[k]]_{n_b} = -\sum_{i=-\infty}^{\infty} s_0[i]h[n_b - i]$$

$$(2.27) \qquad\qquad\qquad\qquad = -\sum_{i=1}^{n_b} 1$$

$$(2.28) \qquad\qquad\qquad\qquad = -n_b.$$

From now on, we will let s_0 represent the noise-free filter output when 0 is sent (and, respectively, s_1 when 1 is sent). Thus,

$$(2.29) \qquad\qquad s_0 = [s_0[k] \star h[k]]_{n_b} ,$$

and

$$(2.30) \qquad\qquad s_1 = [s_1[k] \star h[k]]_{n_b} .$$

Because the detection filter is linear, the effect of noise will be added to the noise-free output.

Now consider the effect of noise. From (2.3), since convolution distributes over addition, applying the filter $h[k]$ to $y[k]$ will result in

$$(2.31) \qquad\qquad y[k] \star h[k] = (s[k] + n[k]) \star h[k]$$

$$(2.32) \qquad\qquad\qquad\qquad = s[k] \star h[k] + n[k] \star h[k],$$

which consists of a signal term $s[k]\star h[k]$ and a noise term $n[k]\star h[k]$. The signal term in (2.32) can be obtained by calculating $h[k] \star s_0[k]$ and $h[k] \star s_1[k]$, as in Example 2.3 for polar NRZ. Considering the noise term, evaluating the filter output at time n_b, we can write

$$(2.33) \qquad\qquad [n[k] \star h[k]]_{n_b} = \sum_i h[i]n[n_b - i],$$

where the sum is over all possible values of i (which can be restricted to those values of i for which $h[i] \neq 0$, i.e., the "support" of $h[i]$). Since $n[k]$ is an AWGN random process with mean zero and variance $N_0/2T_s$ (from (2.10)), the sum in (2.33) is

a weighted sum of independent Gaussian random variables, with mean zero and variance

$$(2.34) \qquad \sigma^2 = \frac{N_0}{2T_s} \sum_i h[i]^2.$$

Recalling Chapter 1, the sum of Gaussian random variables is itself a Gaussian random variable. Thus, if the transmitted symbol is known to be 0, the filter output has mean $[s_0[k] \star h[k]]_{n_b}$ and variance σ^2. Similarly, if the transmitted symbol is known to be 1, the filter output has mean $[s_1[k] \star h[k]]_{n_b}$ and variance σ^2.

2.4. Error analysis

2.4.1. General form of the probability of error. An error occurs if $\hat{b}_j \neq b_j$ (we will refer to this event as error). Using the decision function $d_z(\phi[j])$ from (2.20), we have that $\hat{b}_j = 0$ if $\phi[j] \geq z$; thus, an error occurs if $b_j = 1$ and $\phi[j] \geq z$. Similarly, an error occurs if $b_j = 0$ and $\phi[j] < z$.

Remember that if b_j is known, then $\phi[j]$ is a Gaussian random variable. Thus, using properties of Gaussian random variables, we can calculate the conditional error probabilities $\Pr(\text{error}|b_j = 0)$ and $\Pr(\text{error}|b_j = 1)$. The average error probability is then given by

$$(2.35) \qquad \Pr(\text{error}) = \Pr(\text{error}|b_j = 0)\Pr(b_j = 0) + \Pr(\text{error}|b_j = 1)\Pr(b_j = 1).$$

Let's start with $\Pr(\text{error}|b_j = 1)$. Clearly

$$(2.36) \qquad \Pr(\text{error}|b_j = 1) = \Pr(\phi[j] \geq z)|b_j = 1).$$

This event is illustrated in Figure X. Given $b_j = 1$, $\phi[j]$ is a Gaussian random variable with mean s_1 (from (2.30)) and variance σ^2 (from (2.34)). The PDF of this random variable is given by

$$(2.37) \qquad f(\phi[k]|b_j = 1) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(\phi[k] - s_1)^2\right).$$

Thus, since $\phi[k]$ is a continuous-valued random variable,

$$(2.38) \quad \Pr(\text{error}|b_j = 1) = \Pr(\phi[j] \geq z)|b_j = 1)$$

$$(2.39) \qquad\qquad = \int_{\phi[k]=z}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(\phi[k] - s_1)^2\right) d\phi[k].$$

As we mentioned in Chapter 1, integrals over the Gaussian PDF, such as the one in (2.39), cannot be directly computed. However, we can use a special function known as the error function complementary, $\text{erfc}(\cdot)$, defined in (1.2), to represent this integral. With a change of variables, the integral in (2.39) can be expressed in terms of $\text{erfc}(\cdot)$ as

$$(2.40) \qquad \Pr(\text{error}|b_j = 1) = \frac{1}{2}\text{erfc}\left(\frac{z - s_1}{\sqrt{2\sigma^2}}\right).$$

By a similar derivation, $\Pr(\text{error}|b_j = 0)$ is given by

$$(2.41) \qquad \Pr(\text{error}|b_j = 0) = \frac{1}{2}\text{erfc}\left(\frac{s_0 - z}{\sqrt{2\sigma^2}}\right).$$

Showing the correctness of (2.40) and (2.41) are left as exercises for the reader. Finally, substituting into (2.35), we have

$$\text{Pr}(\text{error})$$

$$(2.42) \qquad = \quad \text{Pr}(\text{error}|b_j = 0)\text{Pr}(b_j = 0) + \text{Pr}(\text{error}|b_j = 1)\text{Pr}(b_j = 1)$$

$$(2.43) \qquad = \quad \frac{1}{2}\text{erfc}\left(\frac{s_0 - z}{\sqrt{2\sigma^2}}\right)\text{Pr}(b_j = 0) + \frac{1}{2}\text{erfc}\left(\frac{z - s_1}{\sqrt{2\sigma^2}}\right)\text{Pr}(b_j = 1).$$

2.4.2. Probability of error for Polar NRZ. Let's return to Example 2.3, which illustrated detection with Polar NRZ signals, and use a threshold $z = 0$. In the example, we showed that $s_0 = -s_1 = n_b$. Thus, substituting into (2.43), we have

$$\text{Pr}(\text{error})$$

$$(2.44) \qquad = \quad \frac{1}{2}\text{erfc}\left(\frac{n_b - 0}{\sqrt{2\sigma^2}}\right)\text{Pr}(b_j = 0) + \frac{1}{2}\text{erfc}\left(\frac{0 - (-n_b)}{\sqrt{2\sigma^2}}\right)\text{Pr}(b_j = 1)$$

$$(2.45) \qquad = \quad \frac{1}{2}\text{erfc}\left(\frac{n_b}{\sqrt{2\sigma^2}}\right)\text{Pr}(b_j = 0) + \frac{1}{2}\text{erfc}\left(\frac{n_b}{\sqrt{2\sigma^2}}\right)\text{Pr}(b_j = 1)$$

$$(2.46) \qquad = \quad \frac{1}{2}\text{erfc}\left(\frac{n_b}{\sqrt{2\sigma^2}}\right)(\text{Pr}(b_j = 0) + \text{Pr}(b_j = 1))$$

$$(2.47) \qquad = \quad \frac{1}{2}\text{erfc}\left(\frac{n_b}{\sqrt{2\sigma^2}}\right).$$

From (2.47), since $\text{erfc}(\cdot)$ is a decreasing function, it seems like increasing n_b leads to a decrease in $\text{Pr}(\text{error})$. However, by examining (2.34), we can show that the probability of error is independent of the sampling rate. In Example 2.3, we used a filter impulse response $h[k] = s_0[k]$, so (2.34) becomes

$$(2.48) \qquad \sigma^2 \quad = \quad \frac{N_0}{2T_s}\sum_{i=1}^{n_b} s_0[k]^2$$

$$(2.49) \qquad\qquad = \quad \frac{N_0}{2T_s}\sum_{i=1}^{n_b} 1$$

$$(2.50) \qquad\qquad = \quad \frac{N_0 n_b}{2T_s}.$$

However, the sample time T_s can be expressed as

$$(2.51) \qquad\qquad T_s = \frac{T_b}{n_b}.$$

Substituting into (2.50), we have

$$(2.52) \qquad\qquad \sigma^2 = \frac{N_0 n_b^2}{2T_b},$$

and finally, substituting into (2.47), we have

$$(2.53) \qquad \text{Pr}(\text{error}) \quad = \quad \frac{1}{2}\text{erfc}\left(\frac{n_b}{\sqrt{2(N_0 n_b^2)/(2T_b)}}\right)$$

$$(2.54) \qquad\qquad\qquad = \quad \frac{1}{2}\text{erfc}\left(\sqrt{\frac{T_b}{N_0}}\right).$$

Thus, in this example, the only important factors in determining the probability of error are the bit duration, T_b, and the AWGN power spectral density coefficient, N_0.

2.4.3. Summary and Examples. Our calculation of probability of error for Polar NRZ followed a procedure that can be generalized to other signalling schemes. Given modulation signals $s_0[k]$ and $s_1[k]$, detection filter $h[k]$, threshold z, and all relevant system parameters (e.g., i.e., probabilities $\Pr(b_j = 0)$ and $\Pr(b_j = 1)$, noise coefficient N_0, sample time T_s, samples per bit n_b), the following procedure may be used to calculate the probability of error:

(1) Calculate the noise-free filter outputs s_0 and s_1, using equations (2.29)-(2.30).
(2) Calculate the variance of the noise σ^2 at the output of the filter, using equation (2.34). Given that the input bit is 0 (or 1), the output of the filter is then a Gaussian random variable with mean $\mu = s_0$ (resp., s_1) and variance σ^2.
(3) Obtain probability of error by substituting all these quantities into equation (2.43).

Using this procedure, we now present two additional examples.

EXAMPLE 2.4 (Binary Phase Shift Keying). *Returning to example 2.2, we now calculate the probability of error for BPSK. Let*

$$(2.55) \qquad h[k] = \begin{cases} \sin\left(-\frac{2\pi k}{n_b}\right), & 1 \le k \le n_b \\ 0 & \text{otherwise} \end{cases},$$

and let the threshold $z = 0$. We now follow the procedure given above.
Noise-free filter outputs: Using (2.29), s_0 is given by

$$(2.56) \qquad s_0 \;=\; [s_0[k] \star h[k]]_{n_b}$$

$$(2.57) \qquad =\; \sum_{i=1}^{n_b} s_0[i] h[n_b - i]$$

$$(2.58) \qquad =\; \sum_{i=1}^{n_b} \sin\left(\frac{2\pi i}{n_b}\right) \sin\left(\frac{2\pi(i - n_b)}{n_b}\right)$$

$$(2.59) \qquad =\; \sum_{i=1}^{n_b} \sin\left(\frac{2\pi i}{n_b}\right) \sin\left(\frac{2\pi i}{n_b} + 2\pi\right)$$

$$(2.60) \qquad =\; \sum_{i=1}^{n_b} \sin\left(\frac{2\pi i}{n_b}\right)^2.$$

Furthermore, it is easy to show that

$$(2.61) \qquad s_1 = -s_0.$$

Variance of the noise: *Using (2.34), σ^2 is given by*

$$(2.62) \qquad \sigma^2 \;=\; \frac{N_0}{2T_s} \sum_{i=1}^{n_b} h[i]^2$$

$$(2.63) \qquad =\; \frac{N_0}{2T_s} \sum_{i=1}^{n_b} \sin\left(-\frac{2\pi k}{n_b}\right)^2$$

$$(2.64) \qquad =\; \frac{N_0}{2T_s} \sum_{i=1}^{n_b} \sin\left(\frac{2\pi k}{n_b}\right)^2$$

$$(2.65) \qquad =\; \frac{N_0}{2T_s} s_0,$$

where (2.64) follows from (2.63) since $\sin(-x) = -\sin(x)$.

Probability of error. *Substituting into (2.43), we have*

$$\text{Pr(error)}$$

$$(2.66) \qquad =\; \frac{1}{2}\mathrm{erfc}\left(\frac{s_0 - z}{\sqrt{2\sigma^2}}\right)\mathrm{Pr}(b_j = 0) + \frac{1}{2}\mathrm{erfc}\left(\frac{z - s_1}{\sqrt{2\sigma^2}}\right)\mathrm{Pr}(b_j = 1)$$

$$(2.67) \qquad =\; \frac{1}{2}\mathrm{erfc}\left(\sqrt{\frac{T_s s_0}{N_0}}\right)\mathrm{Pr}(b_j = 0) + \frac{1}{2}\mathrm{erfc}\left(\sqrt{\frac{T_s s_0}{N_0}}\right)\mathrm{Pr}(b_j = 1)$$

$$(2.68) \qquad =\; \frac{1}{2}\mathrm{erfc}\left(\sqrt{\frac{T_s s_0}{N_0}}\right).$$

To obtain specific numbers for this system, let $T_b = 10^{-4}$, $n_b = 8$, and $N_0 = 1.25 \cdot 10^{-5}$. Then from (2.60), $s_0 = 4$, so

$$(2.69) \qquad \text{Pr(error)} \;=\; \frac{1}{2}\mathrm{erfc}\left(\sqrt{\frac{(10^{-4}/8)\cdot 4}{1.25 \cdot 10^{-5}}}\right)$$

$$(2.70) \qquad =\; 0.00234.$$

EXAMPLE 2.5 (On-off keying). *Let $s_0[k]$ be the same as (2.15), and let $h[k] = s_0[k+1]$, as in Polar NRZ. However, let $s_1[k] = 0$ for all k. This is referred to as on-off keying, since the transmitter is "on" (all +1) to transmit 0, and "off" (all zero) to transmit 1. Suppose $\mathrm{Pr}(b_j = 0) = \mathrm{Pr}(b_j = 1) = 1/2$. We consider two cases: first, $z = 0$, and second, $z = n_b/2$.*

Noise-free filter outputs: *Since $s_0[k]$ and $h[k]$ are the same as in Polar NRZ, s_0 is also the same, so*

$$(2.71) \qquad s_0 = [s_0[k] \star h[k]]_{n_b} = n_b.$$

Since $s_1[k] = 0$, then

$$(2.72) \qquad s_1 = [0 \star h[k]]_{n_b} = 0.$$

Variance of the noise: *Since $h[k]$ is the same as in Polar NRZ, then*

$$(2.73) \qquad \sigma^2 = \frac{N_0}{2T_s}\sum_i h[k]^2 = \frac{N_0}{2T_s}n_b.$$

Probability of error. *Substituting into (2.43), and using $z = 0$, we have*

$\Pr(\mathsf{error})$

$$(2.74) \quad = \quad \frac{1}{2}\mathrm{erfc}\left(\frac{s_0 - z}{\sqrt{2\sigma^2}}\right)\Pr(b_j = 0) + \frac{1}{2}\mathrm{erfc}\left(\frac{z - s_1}{\sqrt{2\sigma^2}}\right)\Pr(b_j = 1)$$

$$(2.75) \quad = \quad \frac{1}{4}\mathrm{erfc}\left(\sqrt{\frac{T_s n_b}{N_0}}\right) + \frac{1}{4}\mathrm{erfc}\,(0)$$

$$(2.76) \quad = \quad \frac{1}{4}\left(\mathrm{erfc}\left(\sqrt{\frac{T_s s_0}{N_0}}\right) + 1\right).$$

On the other hand, using $z = n_b/2$, we have

$$(2.77) \qquad\qquad \Pr(\mathsf{error}) = \frac{1}{2}\mathrm{erfc}\left(\sqrt{\frac{n_b T_s}{4N_0}}\right).$$

It is straightforward to show that (2.77) is generally smaller (and therefore better) than (2.76).

2.5. Optimizing the decision threshold

In Example 2.5, we saw that the choice of threshold had an impact on the probability of error. Given a pair of signals $s_0[k]$ and $s_1[k]$, and a filter $h[k]$, the threshold z should obviously be selected so as to minimize $\Pr(\mathsf{error})$. How can we do this?

Let's return to equation (2.43), and differentiate with respect to z. Doing so, we get an expression closely related to the Gaussian integral: let

$$(2.78) \qquad\qquad f_T(t) = \frac{1}{\sqrt{\pi}}\exp(-t^2),$$

and let $F_T(t)$ represent the indefinite integral of $f_T(t)$ (which is not available in closed form). Then

$$(2.79) \qquad\qquad \mathrm{erfc}(z) \quad = \quad \int_z^\infty f_T(t)dt$$

$$(2.80) \qquad\qquad\qquad\quad = \quad F_T(\infty) - F_T(z).$$

However, by the fundamental theorem of calculus, it is true that

$$(2.81) \qquad\qquad f_T(t) = \frac{d}{dt}F_T(t),$$

so taking the first derivative of $\mathrm{erfc}(z)$ with respect to z, we get

$$(2.82) \qquad\qquad \frac{d}{dz}\mathrm{erfc}(z) = -f_T(z).$$

Applying this to (2.43), and using the chain rule for derivatives, we get

$\dfrac{d}{dz}\Pr(\mathsf{error})$

$$(2.83) \quad = \quad \frac{1}{2}\Pr(b_j = 0)\frac{d}{dz}\mathrm{erfc}\left(\frac{s_0 - z}{\sqrt{2\sigma^2}}\right) + \frac{1}{2}\Pr(b_j = 1)\frac{d}{dz}\mathrm{erfc}\left(\frac{z - s_1}{\sqrt{2\sigma^2}}\right)$$

$$(2.84) \quad = \quad \frac{1}{2\sqrt{\sigma^2}}\Pr(b_j = 0)f_T\left(\frac{s_0 - z}{\sqrt{2\sigma^2}}\right) - \frac{1}{2\sqrt{\sigma^2}}\Pr(b_j = 1)f_T\left(\frac{z - s_1}{\sqrt{2\sigma^2}}\right).$$

To find the minimum, we set the expression in (2.84) to zero. As a result, the minimizing value of z is the value satisfying

$$(2.85) \qquad \Pr(b_j = 0) f_T \left(\frac{s_0 - z}{\sqrt{2\sigma^2}} \right) = \Pr(b_j = 1) f_T \left(\frac{z - s_1}{\sqrt{2\sigma^2}} \right).$$

It is left as an exercise for the reader to show that this value is a minimum. Substituting into (2.85) with the expansion of $f_T(t)$, and collecting exponential terms, we have

$$(2.86) \qquad \exp\left(-\frac{1}{2\sigma^2}(s_0 - z)^2 + \frac{1}{2\sigma^2}(z - s_1)^2 \right) = \frac{\Pr(b_j = 1)}{\Pr(b_j = 0)}.$$

Taking the natural logarithm, log, of both sides results in

$$(2.87) \qquad \frac{1}{2\sigma^2} \left((z - s_1)^2 - (s_0 - z)^2 \right) = \log \frac{\Pr(b_j = 1)}{\Pr(b_j = 0)},$$

and collecting terms on the left, we have

$$(2.88) \qquad \frac{1}{2\sigma^2} \left(2(s_0 - s_1)z - (s_0^2 - s_1^2) \right) = \log \frac{\Pr(b_j = 1)}{\Pr(b_j = 0)}.$$

Finally, solving for z gives

$$(2.89) \qquad z = \frac{\sigma^2}{s_0 - s_1} \log \frac{\Pr(b_j = 1)}{\Pr(b_j = 0)} + \frac{1}{2}(s_0 + s_1).$$

Thus, in the polar NRZ example we presented above, $z = 0$ is indeed the optimal threshold, since $s_1 = -s_0$ and $\Pr(b_j = 0) = \Pr(b_j = 1) = 1/2$.

We make two remarks on (2.89). First, if the two binary values 0 and 1 are equiprobable (i.e., $\Pr(b_j = 0) = \Pr(b_j = 1) = 1/2$), then (2.89) reduces to

$$(2.90) \qquad z = \frac{1}{2}(s_0 + s_1),$$

which is exactly halfway between the two mean values s_0 and s_1. Thus, for any received value $y[k]$, the decision $\hat{b}[k]$ is made based on the *closest point* to $y[k]$, either s_0 or s_1. Second, if 0 and 1 are not equiprobable, then the threshold is biased towards the less likely bit, thereby expanding the region of $y[k]$ that maps to the more likely bit. This reduces error since, in case of uncertainty, it is safer to select the more likely bit. This is illustrated in Figure X.

In general, the optimal way to distinguish between two signals in noise is to employ the maximum *a posteriori* probability (MAP) criterion

2.6. Probability of error and energy per bit

2.7. Problems

(1) Making the changes of variables described in Chapter 1, show that (2.40) and (2.41) are correct.
(2) Show that the value of z satisfying (2.85) is a minimum of $\Pr(\mathsf{error})$.

2.8. Laboratory exercise

CHAPTER 3

Optimal signal detection: The matched filter

CHAPTER 4

Passband Data Transmission

4.1. Linear system aspects of passband communication

4.2. Hardware Model of Communication

4.3. Modulation

CHAPTER 5

Multiple Access Communication Systems

5.1. Interference-free spectrum sharing

5.1.1. Frequency division multiple access.

5.1.2. Time division multiple access.

5.2. Spread-spectrum techniques

CHAPTER 6

An Introduction to Information Theory

6.1. Error-control coding

6.1.1. Capacity.

6.1.2. Linear block codes.

6.1.3. Convolutional codes.

6.2. Data compression

6.2.1. Entropy.

6.2.2. Huffman codes.

Bibliography

[1] A. Leon-Garcia, *Probability and Random Processes for Electrical Engineering, 2nd ed.*, Reading, MA: Addison-Wesley, 1994.

[2] A. Papoulis, *Probability, Random Variables, and Stochastic Processes, 3rd ed.*, New York, NY: McGraw-Hill, 1991.

Fourier Transforms

A.1. Properties

A.2. Table of fourier transform pairs

A.2.1. Definitions.

(A.1)
$$\text{rect}(t) = \begin{cases} 1, & |t| < 1/2 \\ 1/2, & |t| = 1/2 \\ 0, |t| > 1/2 \end{cases}$$

(A.2)
$$\text{sinc}(t) = \frac{\sin(\pi t)}{\pi t}$$

A.2.2. Table.

Fourier transform $\leftrightarrow$ Time domain

(A.3)
$$\text{rect}(j\omega) \qquad \frac{1}{2\pi}\text{sinc}\left(\frac{t}{2\pi}\right)$$