Appendix B

Review of random processes

B.0 INTRODUCTION

This appendix reviews some results from the theory of random processes, which will be essential when analyzing roundoff noise in filter banks (Chap. 9). These results will also be used in studying the effects of subband quantization in Appendix C. Detailed treatements of random processes can be found in a number of references, e.g., Papoulis [1965], Davenport [1970], Oppenheim and Schafer [1975], Peebles [1987], and Therrien [1992].

B.1 REAL RANDOM VARIABLES

We assume familiarity with the notions of probability and random variables, which are basic to all further discussions. Let X be a real *random variable* (abbreviated as rv or r.v.). † We will deal with *continuous random variables*. This means that X can take any value in a continuous range such as $a \leq X \leq b$, where a and b may not be finite. In this case, we have to describe the r.v. by a probability density function (rather than just probability), denoted as $f_X(x)$. This function is defined such that the integral

$$\int_{x_1}^{x_2} f_X(x)dx, \qquad (B.1.1)$$

represents the probability that X is in the range $x_1 \leq x \leq x_2$. It satisfies the following properties: (a) $f_X(x) \geq 0$ and (b) $\int_{-\infty}^{\infty} f_X(x)dx = 1$.

Figure B.1-1 shows an example of $f_X(x)$ called the uniform density function. It is easy to see that this satisfies the property $\int_{-\infty}^{\infty} f_X(x)dx = 1$. The density function $f_X(x)$ can exceed unity, since it is not a probability by itself. In the example shown, if $(b-a) < 1$, the value of $f_X(x)$ exceeds unity. Fig. B.1-2 shows another popular

† Formally, an r.v. is defined to be a mapping from a *sample space* to the real line (more generally the complex plane); see Papoulis [1965].

density called the Gaussian density. This is analytically given by

$$f_X(x) = \frac{1}{\sigma_X\sqrt{2\pi}} \exp\left(\frac{-(x-m_X)^2}{2\sigma_X^2}\right). \qquad (B.1.2)$$

The meanings of m_X and σ_X will be explained below. For the moment note that as m_X decreases, the center of the plot shifts to the left, whereas as σ_X decreases the plot gets narrower and taller. †

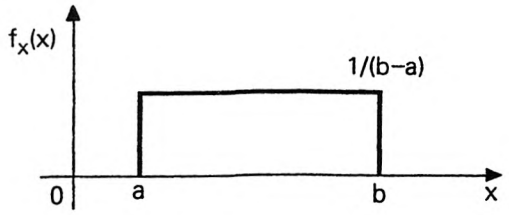

Figure B.1-1 The uniform probability density function.

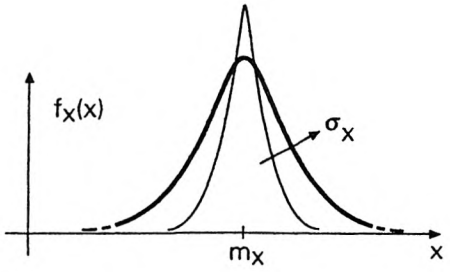

Figure B.1-2 The Gaussian density function.

Expected Values

Let $g(X)$ be a function of an r.v. X (e.g., $g(X) = X^2, \sin X$ etc.). In general $g(X)$ is itself a random variable. The expected value of $g(X)$ is defined to be

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) dx. \qquad (B.1.3)$$

Here are some standard examples: $m_X \triangleq E[X]$ is the expected value of the r.v. X. The quantity $E[X^2]$ is called the mean square value of X. The quantity $E[(X - m_X)^2]$ which is the mean square value of the r.v. $(X - m_X)$ is called the variance of X, denoted σ_X^2. Summarizing,

$$\begin{aligned} \text{Expected value } m_X &= E[X] \\ \text{Mean square value} &= E[X^2] \\ \text{Variance } \sigma_X^2 &= E[(X - m_X)^2]. \end{aligned} \qquad (B.1.4)$$

† In this section we have used upper case letters such as X for the random variables, and lower case letters such as x, x_1 for the values assume by the random variables. This is a useful notation, but sometimes becomes infeasible because of conflicting notational conventions, and other demands for upper case letters. In many cases we will leave it to the reader's judgement to make the distinction between an r.v. and the value assigned to it.

These are related as

$$E[X^2] = \sigma_X^2 + m_X^2. \tag{B.1.5}$$

m_X is also called the mean value or *mean* of X. The quantity σ_X (square root of variance) is called the *standard deviation* of X. From these definitions one can verify that for the Gaussian density, the mean and variance are indeed given by m_X and σ_X^2 appearing in (B.1.2). Refering to Fig. B.1-2 we see that the mean m_X agrees with the center of the plot. (This is not in general true for arbitrary density functions.) For the uniform density function one can verify that

$$m_X = \frac{b+a}{2}, \quad \sigma_X^2 = \frac{(b-a)^2}{12}. \tag{B.1.6}$$

In general, the function $f_X(x)$ can have impulses (Dirac delta functions) in it. These are required to take care of the discrete nature of the density function in some applications; for example, if the probability for X to take the value 7 is a half, then $f_X(x)$ must have a term $0.5\delta_a(x-7)$. In this text, we have no need to accomodate such impulses, so we assume $f_X(x)$ to be impulse-free.

Collection of Random Variables

Suppose X and Y are two random variables. These are jointly described by the *joint probability density* function $f_{XY}(x,y)$. This is defined such that

$$\int_{x=x_1}^{x_2} \int_{y=y_1}^{y_2} f_{XY}(x,y)dxdy, \tag{B.1.7}$$

is equal to the probability that X and Y are in the range $x_1 \leq X \leq x_2$ and $y_1 \leq Y \leq y_2$. The joint density is nonnegative and such that if we set $x_1 = y_1 = -\infty$ and $x_2 = y_2 = \infty$, the above integral reduces to unity. Moreover, the density function of one of the random variables, say that of X, can be recovered from $f_{XY}(x,y)$ by integrating over the other r.v. Y, that is,

$$f_X(x) = \int_{y=-\infty}^{\infty} f_{XY}(x,y)dy. \tag{B.1.8}$$

$f_Y(y)$ can be obtained similarly. $f_X(x)$ and $f_Y(y)$ are said to be *marginal density* functions.

Given a function $g(X,Y)$ of the two random variables, we define its expected value as

$$E[g(X,Y)] = \int_{x=-\infty}^{\infty} \int_{y=-\infty}^{\infty} g(x,y) f_{XY}(x,y)dxdy. \tag{B.1.9}$$

The quantity

$$R_{XY} \triangleq E[XY], \tag{B.1.10}$$

is called the *cross correlation* between the real random variables X and Y. The quantity

$$C_{XY} \triangleq E[(X-m_X)(Y-m_Y)], \tag{B.1.11}$$

is called the *cross covariance* between X and Y. These are related as

$$R_{XY} = C_{XY} + m_X m_Y, \tag{B.1.12}$$

a generalization of (B.1.5).

Depending on the behavior of $f_{XY}(x,y)$, R_{XY} and C_{XY}, two random variables are often classified into useful types. We say that X and Y are

1. *statistically independent* if $f_{XY}(x,y) = f_X(x)f_Y(y)$.
2. *uncorrelated* if $E[XY] = E[X]E[Y]$, that is, $R_{XY} = m_X m_Y$ i.e., $C_{XY} = 0$.
3. *orthogonal* if $E[XY] = 0$.

Note that uncorrelatedness and orthogonality are identical, if one of the random variables has zero mean. It can be shown that statistical independence implies uncorrelatedness but the converse is in general not true. For example, let $X = \cos\theta$ and $Y = \sin\theta$ where θ is a real r.v., with uniform density function in the range $0 \leq \theta < 2\pi$. Then X and Y are uncorrelated but not statistically independent. If $f_{XY}(x,y)$ is Gaussian (Sec. B.5), then it can be shown that the uncorrelated property implies independence.

In a manner identical to the above, if we are given several random variables $X_0, X_1, \ldots X_{M-1}$, we can define a joint density function and various expected values. See Sec. B.5.

B.2 REAL RANDOM PROCESSES

Let $\{X(n)\}$ be a real sequence such that each sample $X(n)$ is a random variable. We say that $\{X(n)\}$ is a random process (with braces ususally omitted). (Once again a formal definition based on a mapping from a "sample space" to the space of functions can be found in the references mentioned before). We use $X(n)$ to indicate the random process and $x(n)$ to denote a particular realization, i.e., $x(n)$ is the value taken by the random variable $X(n)$ in a particular measurement.

When we attempt to characterize the random process, several practical difficulties arise. Each sample $X(n)$ is characterized by a density function, and pairs of samples such as $X(n-1), X(n)$ are characterized by two dimensional joint density functions. In fact, given any set of samples such as $X(n_1), X(n_2), \ldots X(n_M)$, we have to characterize them by a M dimensional joint density. One can see that the complete characterization becomes very complicated. Note also that in general $x(n)$ does not have finite energy, and its z-transform defined in the usual way may not converge anywhere. So we require a different tool to understand and characterize these waveforms.

However, as we will see, a partial characterization based on expected values is sufficient for many applications, for example, noise analysis in digital filter banks. We can define various kinds of expected values for any random process. The quantity $E[X(n)]$ is the mean of the random process, and in general depends on n. The quantity

$$R(m,n) = E[X(m)X(n)], \tag{B.2.1}$$

is the cross correlation between the real random variables $X(m)$ and $X(n)$.

Wide sense stationary (WSS) processes

A wide sense stationary (WSS) random process is one for which $E[X(n)]$ is independent of n and $R(m,n)$ depends only on the difference $m-n$. We can therefore define the functions

$$\begin{aligned} m_X &= E[X(n)] = \text{mean value}, \\ R_{XX}(k) &= E[X(n)X(n-k)] = \text{autocorrelation sequence}. \end{aligned} \tag{B.2.2}$$

The quantity k in $R_{XX}(k)$ is called the lag variable (as it describes a difference in time index). We say that $R_{XX}(k)$ is the autocorrelation of the WSS process at lag k.

For applications such as noise analysis in linear systems, the noise source (such as the output noise of a roundoff quantizer), can often be modeled satisfactorily as a wide sense stationary random process [Oppenheim and Schafer, 1975]. In these applications, the quantity of main interest is *noise variance* at the system output. This can be calculated from the autocorrelation $R_{XX}(k)$ of the noise source. It is not necessary to know the higher dimensional density functions (or even the two dimensional density function) in order to perform this analysis. This simplifies our study of random processes, as well as noise analysis to a large extent. From this point on, we will concentrate entirely on WSS random processes.

From $R_{XX}(k)$ we can define several useful quantities: the energy of the random process is

$$E[X^2(n)] = R_{XX}(0), \tag{B.2.3}$$

and the variance of the process is

$$\sigma_X^2 = E[(X(n) - m_X)^2]. \tag{B.2.4}$$

These quantities are related as

$$\sigma_X^2 = R_{XX}(0) - m_X^2. \tag{B.2.5}$$

The covariance squence $C_{XX}(k)$ is defined as

$$C_{XX}(k) = E[(X(n) - m_X)(X(n-k) - m_X)], \tag{B.2.6}$$

and leads to

$$R_{XX}(k) = C_{XX}(k) + m_X^2. \tag{B.2.7}$$

$C_{XX}(k)$ is the autocorrelation of the zero mean process $X(n) - m_X$.

Power Spectral Density

The power spectrum (or power spectral density) of the WSS process is defined as the Fourier transform of $R_{XX}(k)$, that is,

$$S_{XX}(e^{j\omega}) = \sum_{k=-\infty}^{\infty} R_{XX}(k)e^{-j\omega k}. \tag{B.2.8}$$

For a real process, $R_{XX}(k)$ is symmetric (that is, $R_{XX}(-k) = R_{XX}(k)$), so that $S_{XX}(e^{j\omega})$ is real-valued. In fact it can be shown that $S_{XX}(e^{j\omega}) \geq 0$ (whether the process is real or complex). Note that we can write $S_{XX}(e^{j\omega})$ as a sum of two terms:

$$S_{XX}(e^{j\omega}) = \sum_{k=-\infty}^{\infty} C_{XX}(k)e^{-j\omega k} + m_X^2 2\pi \sum_{m=-\infty}^{\infty} \delta(\omega + 2\pi m). \tag{B.2.9}$$

From the relation (B.2 8), we can express $R_{XX}(0)$ as

$$R_{XX}(0) = \int_{-\pi}^{\pi} S_{XX}(e^{j\omega}) \frac{d\omega}{2\pi}. \tag{B.2.10}$$

In particular, for a zero mean WSS process, the variance is equal to the above integral.

In many cases, the correlation between the samples $x(n)$ and $x(n-k)$ becomes weaker as k grows. (This is not always true, for example if $R_{XX}(k)$ has periodic components). This means that $C_{XX}(k) \to 0$ as $k \to \infty$, i.e., $R_{XX}(k) \to m_X^2$ as $k \to \infty$. So $S_{XX}(e^{j\omega})$ has an impulse component $2\pi m_X^2 \delta_a(\omega)$, and a nonimpulsive component due to $C_{XX}(k)$ (Fig. B.2-1).

For our applications, $S_{XX}(e^{j\omega})$ will be assumed to be free from impulse functions except possibly at $\omega = 0$. This means that $R_{XX}(k)$ does not have any periodic components except the m_X^2 term which has period one. In this case, Fig. B.2-1(a) is a typical plot of $R_{xx}(k)$.

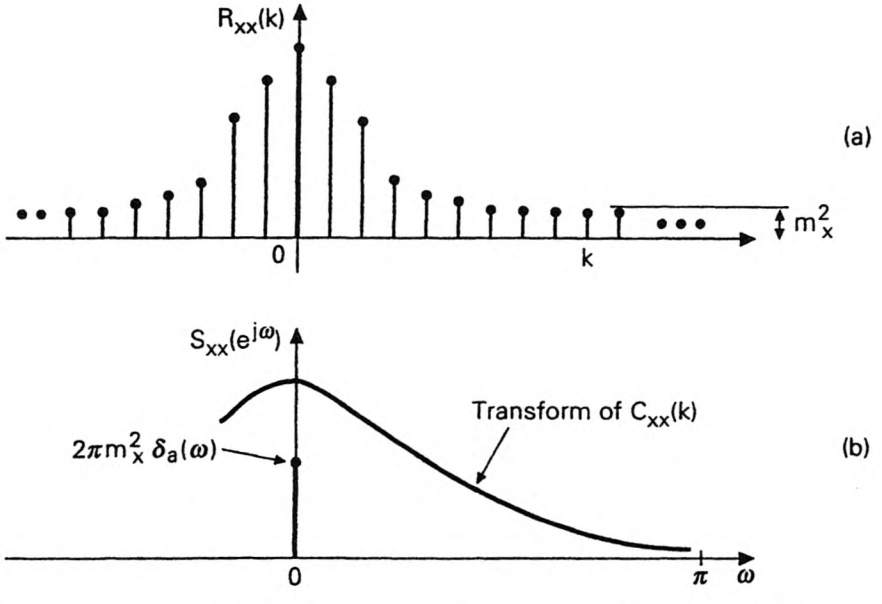

Figure B.2-1 (a) A typical autocorrelation sequence, asymptotically tending to m_x^2, and (b) the corresponding power spectrum, revealing an impulse at $\omega = 0$, representing nonzero mean.

White Random Process

A random process is said to be white if any pair of samples are uncorrelated, i.e., $E[X(n)X(m)] = E[X(n)]E[X(m)]$ for $m \neq n$. If a white process is WSS, then

$$R_{XX}(k) = \begin{cases} m_X^2 & k \neq 0 \\ \sigma_X^2 + m_X^2 & k = 0. \end{cases} \qquad (B.2.11)$$

This can be compactly expressed as

$$R_{XX}(k) = \sigma_X^2 \delta(k) + m_X^2. \qquad (B.2.12)$$

Evidently the covariance sequence is $C_{XX}(k) = \sigma_X^2 \delta(k)$. The power spectrum of a white WSS process is given by

$$S_{XX}(e^{j\omega}) = \sigma_X^2 + 2\pi m_X^2 \delta(\omega), \qquad (B.2.13)$$

in $-\pi \leq \omega < \pi$, and repeats with period 2π. In other words, it has a constant component with height σ_X^2 representing the variance of $X(n)$, and an impulse component representing m_X^2. A zero-mean WSS white random process is therefore characterized by a flat power spectrum (height σ_X^2) and an autocorrelation $R_{XX}(k) = \sigma_X^2 \delta(k)$. Fig. B.2-2 summarizes these. If a WSS process is not white, it is said to be *colored.*

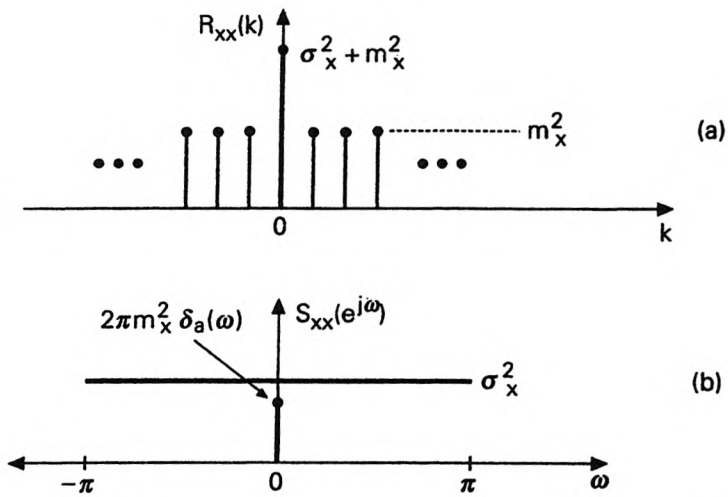

Figure B.2-2 (a) The autocorrelation sequence of a white WSS process, and (b) corresponding power spectral density.

Joint Description of Two or More Random Processes

In many applications, it is necessary to deal with the effect of several random signals at a time. In a digital filter structure, for example, there may exist a number of noise waveforms, generated by a number of quantizers, and it is necessary to find the effective noise power at the filter output.

Suppose $X(n)$ and $Y(n)$ are two real random processes. These are said to be *jointly WSS* if each process is WSS and if $E[X(n)Y(n-k)]$ is a function only of the lag k. In this case we define

$$R_{XY}(k) = E[X(n)Y(n-k)], \qquad (B.2.14)$$

which is called the cross-correlation between the processes. The two processes are said to be uncorrelated if $R_{XY}(k) = E[X(n)]E[Y(n-k)] = m_X m_Y$ for all k. If one of the processes has zero mean, then uncorrelatedness implies $R_{XY}(k) = 0$ for all k.

Ergodicity. A WSS process is said to be ergodic if the statistical averages (such as $E[X(n)]$, $E[X^2(n)]$ etc.), are equal to the corresponding time averages over any

single realization of the process. The ergodicity assumption enables us to estimate these expected values by using time averages.

B.3 PASSAGE THROUGH LTI SYSTEMS

Consider Fig. B.3-1 where $H(z) = \sum_n h(n)z^{-k}$ represents the transfer function of a (stable) LTI system with real impulse response $h(n)$. Suppose the input $X(n)$ is a real WSS random process with mean m_X, autocorrelation $R_{XX}(k)$, and power spectral density $S_{XX}(e^{j\omega})$. Then $Y(n)$ is a WSS random process. Letting m_Y, $R_{YY}(k)$ and $S_{YY}(e^{j\omega})$ be its mean, autocorrelation and power spectrum, we can make the following statements.

a) $m_Y = m_X \sum_n h(n)$.
b) $S_{YY}(e^{j\omega}) = S_{XX}(e^{j\omega})|H(e^{j\omega})|^2$.
c) $R_{YY}(k) = \sum_n R_{hh}(n)R_{XX}(k-n)$ where $R_{hh}(k) = \sum_n h(n)h(n-k)$.

$R_{hh}(k)$ is called the (deterministic) autocorrelation of $h(n)$. All summations above are from $-\infty$ to ∞. Thus $R_{YY}(k)$ is the convolution of $R_{hh}(k)$ with $R_{XX}(k)$.

As an application of this, suppose $x(n)$ is zero mean WSS white with variance σ_X^2. Then

$$R_{XX}(k) = \sigma_X^2 \delta(k), \quad \text{and} \quad S_{XX}(e^{j\omega}) = \sigma_X^2, \tag{B.3.1}$$

so that

$$R_{YY}(k) = \sigma_X^2 R_{hh}(k), \quad \text{and} \quad S_{YY}(e^{j\omega}) = \sigma_X^2 |H(e^{j\omega})|^2. \tag{B.3.2}$$

In other words, the output power spectrum is precisely equal to the magnitude-squared response of $H(z)$ (scaled by σ_X^2). The output process $Y(n)$ is therefore not white but becomes 'colored' by the LTI system $H(z)$ (unless $H(z)$ is allpass). The variance of the output process is

$$R_{YY}(0) = \sigma_X^2 R_{hh}(0) = \sigma_X^2 \sum_n h^2(n). \tag{B.3.3}$$

In other words, the variance gets amplified by the energy in the impulse response $h(n)$.

Figure B.3-1 Passing a random process through an LTI system.

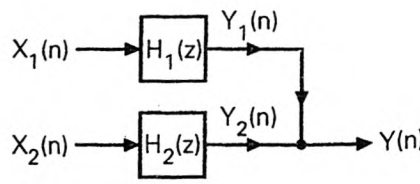

Figure B.3-2 Two uncorrelated WSS processes going through two systems.

Fig. B.3-2 shows another application of the concepts introduced above. Here $H_1(z) = \sum_n h_1(n)z^{-n}$ and $H_2(z) = \sum_n h_2(n)z^{-n}$ are two stable LTI systems with

real impulse responses. Suppose the inputs $X_1(n)$ and $X_2(n)$ are zero mean, white, real, jointly WSS random processes. Assume further that the two processes are uncorrelated as defined in the previous section. It can then be shown that $Y_1(n)$ and $Y_2(n)$ are zero mean, (possibly colored), real, jointly WSS random processes, *uncorrelated* to each other. The variances of $Y_1(n)$ and $Y_2(n)$ are, respectively,

$$\sigma_{Y_1}^2 = \sigma_{X_1}^2 \sum_n h_1^2(n), \quad \sigma_{Y_2}^2 = \sigma_{X_2}^2 \sum_n h_2^2(n). \tag{B.3.4}$$

Because of the fact that $Y_1(n)$ and $Y_2(n)$ are uncorrelated, it can be shown that the variance of $Y(n)$ in Fig. B.3-2 is the sum of variances of $Y_1(n)$ and $Y_2(n)$. So $Y(n)$ is a zero mean WSS random process with variance

$$\sigma_Y^2 = \sigma_{Y_1}^2 + \sigma_{Y_2}^2. \tag{B.3.5}$$

This kind of an analysis finds applications in studying roundoff noise in filter banks (Chap. 9).

B.4 THE COMPLEX CASE

A complex r.v. $X = X_r + jX_i$ is a complex quantity whose real and imaginary parts are random variables, possibly correlated. We can define the joint density function $f_{X_r,X_i}(x_r, x_i)$ to describe the pair of random variables (X_r, X_i). In defining the expected value of functions of X (such as $E[X]$, $E[|X|^2]$ and so on, we merely use this joint density in the appropriate integral. Quantities such as mean and variance generalize nicely. For example the variance is defined now as $E[|X - m_X|^2]$. To make the presentation more efficient, we shall discuss the joint properties of two complex random variables X and Y (such as cross correlation etc.), and set $X = Y$ to obtain the single complex r.v. case. The following is a summary of definitions and properties.

Definitions

1. $m_X \triangleq E[X]$ = mean value of X.
2. $R_{XY} \triangleq E[XY^*]$ = cross correlation between X and Y.
3. $C_{XY} \triangleq E[(X - m_X)(Y - m_Y)^*]$ = cross covariance between X and Y.
4. $\sigma_X^2 \triangleq C_{XX}$ = variance of X.
5. X, Y uncorrelated if $R_{XY} = m_X m_Y^*$.
6. X, Y orthogonal if $R_{XY} = 0$.
7. X, Y (statistically) independent if $f_{XY}(x, y) = f_X(x) f_Y(y)$.

Properties

1. $R_{XX} = \sigma_X^2 + |m_X|^2$.
2. $R_{XY} = R_{YX}^*$ and $C_{XY} = C_{YX}^*$.
3. $R_{XY} = C_{XY} + m_X m_Y^*$.
4. $R_{XX} = E[XX^*] = E[|X|^2]$ = mean square value.
5. X, Y uncorrelated if and only if $C_{XY} = 0$. Statistical independence implies uncorrelatedness, but converse is not true. Uncorrelatedness is same as orthogonality if m_X or m_Y is zero.

Complex Random Processes

A complex random process is defined in the same way as a real process except that $X(n)$ is now complex. We define the mean to be $E[X(n)]$ and autocorrelation to be $E[X(n)X^*(n-k)]$. The process is called WSS if the mean is constant and the autocorrelation is independent of n. Two processes $X(n)$ and $Y(n)$ are jointly WSS, if each of them is individually WSS, and the cross correlation defined as $E[X(n)Y^*(n-k)]$ is independent of n. For the jointly WSS case, we summarize various definitions and properties next.

Definitions

1. $m_X = E[X(n)]$ = mean value.
2. $R_{XY}(k) = E[X(n)Y^*(n-k)]$ = cross correlation between $X(n)$ and $Y(n)$.
3. $C_{XY}(k) = E[(X(n)-m_X)(Y(n-k)-m_Y)^*]$ = cross covariance between $X(n)$ and $Y(n)$.
4. $S_{XY}(e^{j\omega}) = \sum_{k=-\infty}^{\infty} R_{XY}(k)e^{-j\omega k}$ = cross power spectral density.

Properties

1. $R_{XY}(k) = C_{XY}(k) + m_X m_Y^*$.
2. $R_{YX}(k) = R_{XY}^*(-k)$.
3. $S_{YX}(e^{j\omega}) = S_{XY}^*(e^{j\omega})$.
4. In particular,
   $R_{XX}(k) = C_{XX}(k) + |m_X|^2$.
   $R_{XX}(k) = R_{XX}^*(-k)$ (Hermitian symmetric sequence).
   $S_{XX}(e^{j\omega})$ is real (in fact nonnegative).

Passage Through LTI Systems

Suppose $X(n)$ is a complex WSS process. Let this be input to a stable LTI system $H(z) = \sum_n h(n)z^{-n}$, to generate output $Y(n)$. Then $Y(n)$ is WSS. We will now state further properties of the process $Y(n)$. For this define $R_{hh}(k)$ to be the deterministic autocorrelation of $h(n)$, that is,

$$R_{hh}(k) = \sum_n h(n)h^*(n-k).$$

Or equivalently, $R_{hh}(k) = \sum_n h^*(n)h(n+k)$. Taking Fourier transforms on both sides, we get $\sum_k R_{hh}(k)e^{-j\omega k} = |H(e^{j\omega})|^2$. The following properties are true.

1. $m_Y = m_X \sum_n h(n)$.
2. $R_{YY}(k) = \sum_n R_{XX}(n)R_{hh}(k-n)$. (convolution)
3. $S_{YY}(e^{j\omega}) = S_{XX}(e^{j\omega})|H(e^{j\omega})|^2$.
4. $R_{YX}(k) = \sum_m h(m)R_{XX}(k-m)$. (convolution)
5. $S_{YX}(e^{j\omega}) = H(e^{j\omega})S_{XX}(e^{j\omega})$.

We can obtain the case of real processes (with real LTI systems) by dropping the conjugate signs.

B.5. THE VECTOR CASE

At the end of Sec. B.1 it was mentioned that a collection of random variables $X_0 \ldots X_{M-1}$ can be described by a joint density function. We denote this as $f_{\mathbf{X}}(\mathbf{x})$

where $\mathbf{X} = [X_0 \ldots X_{M-1}]^T$. The vector $\mathbf{X}$ is said to be a random vector, or just r.v. for simplicity. $f_{\mathbf{X}}(\mathbf{x})$ is a nonnegative scalar function of the vector $\mathbf{x}$ and has properties similar to $f_{XY}(x, y)$ in Sec. B.1.

The 'mean' or expectation $E[\mathbf{X}]$ is the vector $[E[X_0] \ldots E[X_{M-1}]]^T$. The vector $\mathbf{X}$ has zero mean if this expectation is the null vector. We define the *cross correlation matrix* between the vectors $\mathbf{X}$ and $\mathbf{Y}$ as $E[\mathbf{X}\mathbf{Y}^\dagger]$.

Note in particular that $\mathbf{R} = E[\mathbf{X}\mathbf{X}^\dagger]$ is the *autocorrelation matrix* of $\mathbf{X}$. The autocorrelation of $(\mathbf{X} - E[\mathbf{X}])$ is said to be the *covariance* matrix of $\mathbf{X}$, and is denoted $\mathbf{C}$. Both $\mathbf{R}$ and $\mathbf{C}$ are $M \times M$ Hermitian positive semidefinite matrices. We sometimes use subscripts (e.g., $\mathbf{R}_{\mathbf{xx}}$) when it is necessary to distinguish between more than one random vector. Notice that the k, m element of $\mathbf{C}$ is the cross covariance between X_k and X_m. If every pair of components of $\mathbf{X}$ are uncorrelated, then $\mathbf{C}$ becomes a diagonal matrix. If in addition all components have same variance σ^2 then $\mathbf{C} = \sigma^2\mathbf{I}$.

Gaussian random vector. An $M \times 1$ real random vector $\mathbf{X}$ with mean $\mathbf{m}$ and covariance $\mathbf{C}$ is said to be Gaussian if the joint density function is given by

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{M/2}[\det \mathbf{C}]^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x}-\mathbf{m})^T\mathbf{C}^{-1}(\mathbf{x}-\mathbf{m})\right). \qquad (B.5.1)$$

In this case we also say that the set of random variables X_k (components of $\mathbf{X}$) are jointly Gaussian. The implicit assumption in the above definition is that $\mathbf{C}$ is nonsingular. If the components of $\mathbf{X}$ are uncorrelated, then $\mathbf{C}$ is diagonal, and the above function can be written as the product $f_{X_0}(x_0) \ldots f_{X_{M-1}}(x_{M-1})$ where $f_{X_i}(x_i)$ are one-variable Gaussian density functions. This shows, in particular, that uncorrelated Gaussain random variables are also statistically independent.

Vector Random Processes

An $M \times 1$ vector random process $\{\mathbf{X}(n)\}$ (with curly braces often omitted) is a sequence where all the samples $\mathbf{X}(n)$ are random vectors. This is said to be a wide sense stationary (WSS) process if the vector $E[\mathbf{X}(n)]$ and the matrix $E[\mathbf{X}(n)\mathbf{X}^\dagger(n-k)]$ are independent of n. In this case, we define the autocorrelation to be

$$\mathbf{R}_{\mathbf{xx}}(k) = E[\mathbf{X}(n)\mathbf{X}^\dagger(n-k)]. \qquad (B.5.2)$$

This is an $M \times M$ matrix sequence. The power spectrum of the WSS process $\mathbf{X}(n)$ is defined as

$$\mathbf{S}_{\mathbf{xx}}(z) = \sum_k \mathbf{R}_{\mathbf{xx}}(k)z^{-k}. \qquad (B.5.3)$$

Thus, $\mathbf{S}_{\mathbf{xx}}(e^{j\omega})$ is the Fourier transform of $\mathbf{R}_{\mathbf{xx}}(k)$. Some of the key properties of these quantities are listed below.

1. $\mathbf{R}_{\mathbf{xx}}(k) = \mathbf{R}_{\mathbf{xx}}^\dagger(-k)$.
2. $\mathbf{S}_{\mathbf{xx}}(e^{j\omega}) = \mathbf{S}_{\mathbf{xx}}^\dagger(e^{j\omega})$.
3. The matrix $\mathbf{R}_{\mathbf{xx}}(0)$ is positive (semi)definite, and so is $\mathbf{S}_{\mathbf{xx}}(e^{j\omega})$ for *all* ω.

Two random processes $\mathbf{X}(n)$ and $\mathbf{Y}(n)$ are said to be jointly WSS if (i) each of them is WSS, and (ii) $E[\mathbf{X}(n)\mathbf{Y}^\dagger(n-k)]$ is independent of n. In this case, we

define the cross correlation $\mathbf{R_{xy}}(k)$ and cross power spectrum $\mathbf{S_{xy}}(z)$ as

$$\mathbf{R_{xy}}(k) = E[\mathbf{X}(n)\mathbf{Y}^\dagger(n-k)] \tag{B.5.4}$$

$$\mathbf{S_{xy}}(z) = \sum_k \mathbf{R_{xy}}(k)z^{-k}. \tag{B.5.5}$$

Passage through LTI systems. Let $\mathbf{X}(n)$ be a $r \times 1$ WSS random process. Let this be the input to a $p \times r$ LTI system $\mathbf{H}(z)$. Then the output $\mathbf{Y}(n)$ is a $p \times 1$ vector WSS process. The power spectra of the processes are related as

$$\mathbf{S_{yy}}(z) = \mathbf{H}(z)\mathbf{S_{xx}}(z)\widetilde{\mathbf{H}}(z). \tag{B.5.6}$$

An interesting example occurs when $\mathbf{X}(n)$ is a zero-mean process with

$$\mathbf{R_{xx}}(k) = \begin{cases} \mathbf{0} & k \neq 0 \\ \sigma^2\mathbf{I}, & k = 0 \end{cases} \tag{B.5.7}$$

This means that (a) any two samples $\mathbf{X}(k)$ and $\mathbf{X}(m)$ are uncorrelated, (b) any two components of $\mathbf{X}(k)$ (for any fixed k) are uncorrelated, and (c) each component of $\mathbf{X}(k)$ has the same variance σ^2. We can abbreviate (B.5.7) as $\mathbf{R_{xx}}(k) = \sigma^2\delta(k)\mathbf{I}$. In this case, $\mathbf{S_{xx}}(z) = \sigma^2\mathbf{I}$, so that

$$\mathbf{S_{yy}}(z) = \sigma^2\mathbf{H}(z)\widetilde{\mathbf{H}}(z). \tag{B.5.8}$$

Vector Processes from Scalar Processes

Suppose $X(n)$ is a scalar process and we form the vector process $\mathbf{X}(n)$ by partitioning $X(n)$ into successive blocks of M samples, that is,

$$\mathbf{X}(n) = [\,X(nM) \quad X(nM-1) \quad \ldots \quad X(nM-M+1)\,]^T \tag{B.5.9}$$

According to the notations of Chap. 10, this is the 'blocked version' of $X(n-M+1)$. If $X(n)$ is WSS, then so is the blocked version $\mathbf{X}(n)$. Furthermore it can be shown that if $X(n)$ is WSS, the power spectrum matrix $\mathbf{S_{xx}}(z)$ of $\mathbf{X}(n)$ has the pseudocirculant property (defined in Sec. 5.7.2).

Let $X(n)$ be real WSS, with autocorrelation $R(k)$. Let $\mathbf{R}_M(k)$ be the autocorrelation sequence of the blocked version $\mathbf{X}(n)$. (The subscript M is a reminder that $\mathbf{X}(n)$ is $M \times 1$.) Then $\mathbf{R}_M(0)$ is a (real) symmetric Toeplitz matrix (Appendix A). For example, with $M = 3$, we have

$$\mathbf{R}_3(0) = \begin{bmatrix} R(0) & R(1) & R(2) \\ R(1) & R(0) & R(1) \\ R(2) & R(1) & R(0) \end{bmatrix}. \tag{B.5.10}$$

$\mathbf{R}_M(0)$ is said to be the $M \times M$ autocorrelation matrix associated with the scalar process $X(n)$. This matrix plays a fundamental role in several problems, for example, optimal linear prediction [Jayant and Noll, 1984]. Clearly $\mathbf{R}_0(0) = R(0) > 0$ and is nonsingular. However, $\mathbf{R}_M(0)$ may become singular for sufficiently large M.

If $\mathbf{R}_M(0)$ is singular then so is $\mathbf{R}_n(0)$ for any $n > M$ (since $\mathbf{R}_M(0)$ is the upper left submatrix of $\mathbf{R}_n(0)$).

Harmonic processes. A scalar WSS process is Harmonic(N) if its power spectrum $S_{xx}(e^{j\omega})$ is zero everywhere except at N frequencies ω_m in the range $0 \leq \omega < 2\pi$. In other words,

$$S_{xx}(e^{j\omega}) = 2\pi \sum_{m=0}^{N-1} c_m \delta_a(\omega - \omega_m), \quad 0 \leq \omega < 2\pi, \tag{B.5.11}$$

with $c_m \neq 0$ for any m. Thus the autocorrelation is a sum of 'single frequency terms', that is,

$$R_{xx}(k) = \sum_{m=0}^{N-1} c_m e^{j\omega_m k}. \tag{B.5.12}$$

It can be shown that $X(n)$ is Harmonic(N) if $\mathbf{R}_{N+1}$ is singular but $\mathbf{R}_N$ nonsingular.