Lectures on EE 3025:

Random Processes

John Kieffer
Department of Electrical and Computer Engineering
University of Minnesota Twin Cities
Minneapolis, MN 55455

Lecture 31

Random Processes Part 1

We are now entering the "random process" part of EE 3025, which consists of our last 12 lectures. During these lectures, we will examine material from Chapters 10 and 11 as well as a further development of some of the Chapter 9 material on mean square estimation.

31.1 Further Examples of Random Processes

In Lecture 30, we gave the Bernoulli (coin flipping) process as our first example of random process. Here are some further examples of RP's that we will refer back to now and again in order to examine certain properties of RP's that we will be learning.

Example 31.1. A discrete time process

$$X_n, \quad n = 1, 2, \cdots$$

is an *IID process* if the "component RV's" X_n are independent and they all have the same probability distribution. For example, the Bernoulli process is IID: every RV in the Bernoulli process has the same density

$$(1/2)\delta(x+1) + \delta(x-1).$$

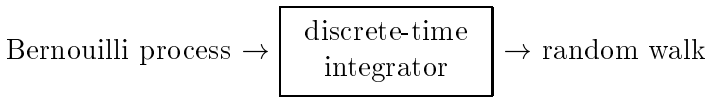
Another special case of IID process is when all the X_n's are Gaussian(0,1). This is called *Gaussian white noise*.

Example 31.2. Let $X_n,\ n = 1, 2, \cdots$ be the Bernouill process and then let $Y_n,\ n = 0, 1, 2, 3, \cdots$ be the random process such that

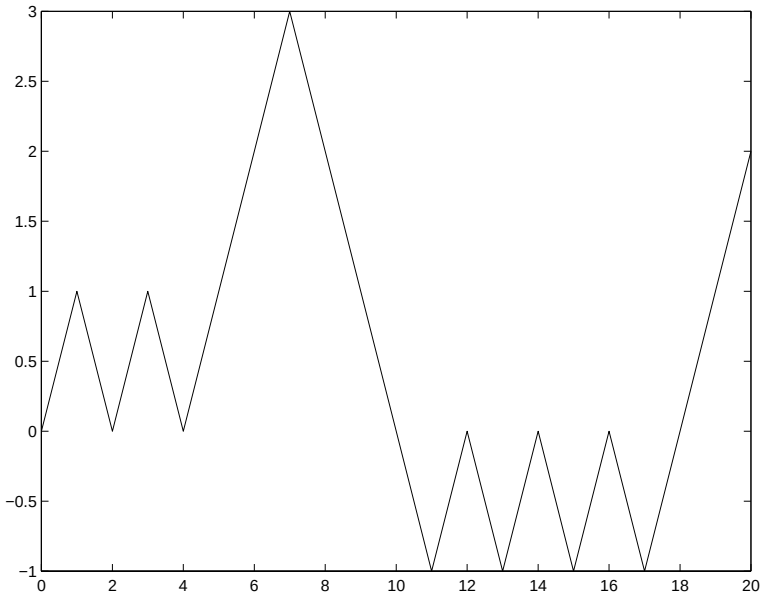
$$\begin{aligned} Y_0 &= 0 \\ Y_1 &= X_1 \\ Y_2 &= X_1 + X_2 \end{aligned}$$

$$\begin{aligned} Y_3 &= X_1 + X_2 + X_3 \\ \vdots &= \vdots \\ Y_n &= X_1 + X_2 + \cdots + X_n \\ \vdots &= \vdots \end{aligned}$$

The process Y_n is called *random walk*. One can conceptualize random walk as the process arising from passing the Bernoulli process through a discrete-time integrator which starts at time zero:



The ensemble of realizations of random walk consists of all discrete-time integer-valued signals $y_n,\ n = 0, 1, 2, \ldots$ whose plots start at the origin and evolve over time so that if the point (n, m) is part of the plot, then the next point in the plot is either $(n, m+1)$ or $(n, m-1)$. The realizations have some other properties which are rather subtle. For example, it can be shown in an advanced course that every realization of random walk (with probability one) crosses or touches the horizontal axis infinitely many times. Here is the beginning of one particular realization of the random walk process:



You can amusingly think of this plot as representing the staggering path of a man who has had to much to drink: he moves a step forward, a step backward, a step forward, a step backward, three steps forward, four steps backward, a step forward, etc. You can see from this why some people call random walk *drunkard's walk*! (Note: From the preceding plot, you can see that the convention has been adopted of connecting up the realization's plot ordinates at discrete times with straight lines, in order to see better what is going on. This is customarily done in plotting random walk realizations.)

Example 31.3. Let A be a random variable uniformly distributed in an interval $[-A^*, A^*]$, and let Θ be a random variable independent of A which is uniformly distributed in the interval $[0, 2\pi]$. The following defines a continuous-time random process called a *random sinusoid*:

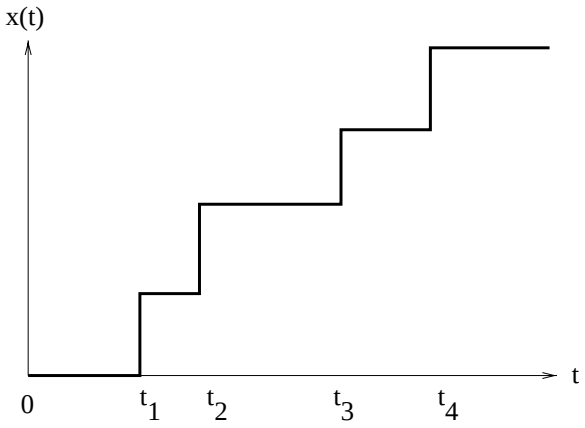
$$X(t) = A\cos(\omega_0 t + \Theta), \quad -\infty < t < \infty$$

The ensemble of realizations of the random sinusoid is easily seen to consist of all sinusoids with period $2\pi/\omega_0$ and amplitude at most A^*.

Example 31.4. The *Poisson process* is a very useful CT process $X(t)$, $t \geq 0$. It is sometimes called the Poisson *arrival* process, because it is intimately connected to the arrivals (of telephone calls, message packets, customers, or whatever the entity may be that is arriving) that occur in a queueing system. These arrivals occur at random times

$$0 < t_1 < t_2 < t_3 < t_4 < \cdots,$$

where we let t_i denote the i-th arrival time, $i = 1, 2, \cdots$. On each performance of the random experiment, one observes the infinite sequence of arrival times and then one determines a Poisson process realization $x(t)$, which is the following step function:



At any time $t \geq 0$, the value $x(t)$ is the number of arrivals that have occured in the time interval $[0, t]$. It follows that the realization $x(t)$ is a step function with unit steps: it is zero from time $t = 0$ up to the time t_1 of the first arrival; it takes the value 1 from time $t = t_1$ up to the time t_2 of the second arrival, increasing to value 2 at time t_2 over to the time t_3 of the third arrival, etc. There is an important parameter $\lambda > 0$ of a Poisson process which completely determines the probabilistic behavior of the process in a manner that we will investigate further in a future lecture; λ measures the average arrival rate (number of arrivals per unit time). For example, if $\lambda = 2$, we would expect about 20 arrivals in the first 10 seconds, averaged over different realizations of the Poisson process.

31.2 Cross-Sections of a Process

- Let $X(t), \quad -\infty < t < \infty$, be a random process and let t^* be a particular time. If we sample the process at time $t = t^*$, then we get a 1-D random variable $X(t^*)$. This is called a *one-dimensional cross-section* of the process. You can visualize the randomness in the cross-section $X(t^*)$ as arising in the following way. Every time you perform the experiment, you get a different realization, and therefore a different value of $X(t^*)$ by sampling that realization at time t^*. This interpretation is why the word "cross-section" is used. You can think of plots of all of the realizations, and then you take a "cross section" across these realizations at time $t = t^*$ (view this as taking a vertical line through $t = t^*$ on the t-axis, this vertical line cutting through all the realizations). Each one-dimensional cross-section $X(t^*)$ can be characterized in terms of its probability density function $f_{X(t^*)}(x)$.
- Now let t_1 and t_2 be any two fixed times. Sample the process $X(t), \quad -\infty < t < \infty$, at times $t = t_1$ and $t = t_2$. This gives us a 2-D random variable $(X(t_1), X(t_2))$ called a *two-dimensional cross-section* of the process. Each two-dimensional cross-section $(X(t_1), X(t_2))$ can be characterized in terms of its joint density $f_{X(t_1),X(t_2)}(x_1, x_2)$.
- There are also higher dimensional cross-sections of a process. For example, fix three times t_1, t_2, t_3. The samples $(X(t_1), X(t_2), X(t_3))$ form a three-dimensional cross-section of the process. Each three-dimensional cross-section would be characterized by a joint density function which is a function of three variables. We do not consider three and higher dimensional cross-sections in this course, because only one- and two-dimensional cross-sections are needed in order to design optimal linear filters. (We see more about this application towards the end of the course.)

Example 31.5. Consider the random process $X(t), \quad -\infty < t < \infty$ in which

$$X(t) = At + B,$$

where A, B are independent random variables which are each equidistributed over the set $\{-1, 1\}$. There are four realizations of this process:

$$x^{(1)}(t) \quad = \quad t + 1$$

$$\begin{aligned} x^{(2)}(t) &= t-1 \\ x^{(3)}(t) &= -t+1 \\ x^{(4)}(t) &= -t-1 \end{aligned}$$

Each of the realizations occurs with probability 1/4. Let us examine the 1-D cross-section $X(0)$ at time $t = 0$. Sampling the realizations at $t = 0$, we get

$$\begin{aligned} x^{(1)}(0) &= 1 \\ x^{(2)}(0) &= -1 \\ x^{(3)}(0) &= 1 \\ x^{(4)}(0) &= -1 \end{aligned}$$

Two of the realizations (with total probability $= 2/4 = 1/2$) yield the value 1 when sampled at $t = 0$, and the other two realizations yield the value -1 when sampled at $t = 0$. Therefore, the PMF of the 1-D cross-section $X(0)$ is

$$p_{X(0)}(x) = \begin{cases} 1/2, & x = 1 \\ 1/2, & x = -1 \end{cases}$$

Now let us look at the 1-D cross-section $X(1)$ at time $t = 1$:

$$\begin{aligned} x^{(1)}(1) &= 2 \\ x^{(2)}(1) &= 0 \\ x^{(3)}(1) &= 0 \\ x^{(4)}(1) &= -2 \end{aligned}$$

This gives us

$$p_{X(1)}(x) = \begin{cases} 1/4, & x = 2 \\ 1/2, & x = 0 \\ 1/4, & x = -2 \end{cases}$$

Notice that the cross-sections $X(0)$ and $X(1)$ do not have the same probability distribution. The process in this example is a simple example of a *time-varying* or *nonstationary* process, since its one-dimensional cross-sections have distributions that vary with time.

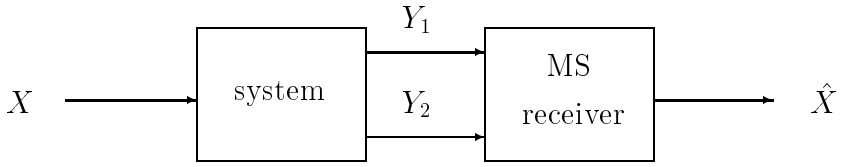
Exercise. Show that the following is the joint PMF of the 2-D cross section $(X(0), X(1))$ for the process $X(t)$ of Example 31.5:

$$\begin{array}{c} \\ X(0) = -1 \\ X(0) = 1 \end{array} \begin{array}{c} \begin{array}{ccc} X(1) = -2 & X(1) = 0 & X(1) = 2 \end{array} \\ \begin{pmatrix} 1/4 & 1/4 & 0 \\ 0 & 1/4 & 1/4 \end{pmatrix} \end{array}$$

The row and column sums should yield the PMF's of the 1-D cross-sections found in Example 31.5.

31.3 Another Look at Mean-Square Estimation

Consider the following block diagram:



We are attempting to form an estimate $\hat{X}$ of X based now on two inputs Y_1, Y_2 to the receiver, instead of just one input as we did earlier in the course. This means the estimate now takes the form

$$\hat{X} = AY_1 + BY_2,$$

where A, B are constants to be determined so that the MS estimation error

$$E[(X - \hat{X})^2] = E[(X - AY_1 - BY_2)^2]$$

is a minimum. To find A, B, you take the partial derivations of $E[(X - \hat{X})^2]$ with respect to A and B, and set each of them equal to zero. It is not hard to see that these lead to the system of equations:

$$E[(X - \hat{X})Y_1] = 0 \tag{31.1}$$
$$E[(X - \hat{X})Y_2] = 0 \tag{31.2}$$

For example,

$$\frac{\partial}{\partial A}E[(X - \hat{X})^2] = E\left[\frac{\partial}{\partial A}(X - \hat{X})^2\right] = E[2(X - \hat{X})\frac{\partial}{\partial A}(X - \hat{X})] = E[2(X - \hat{X})(-Y_1)].$$

Equations (31.1)-(31.2) are called *orthogonality relations*. They state that the estimation error $X - \hat{X}$ is orthogonal to each of the observations Y_i that are linearly combined to form the estimate $\hat{X}$. In our case here, we have just two Y_i's, and equations (31.1)-(31.2) reduce to the two equations

$$AE[Y_1^2] + BE[Y_1Y_2] = E[XY_1]$$
$$AE[Y_1Y_2] + BE[Y_2^2] = E[XY_2]$$

in the two unknowns A, B that one can solve simultaneously to see what the values of A, B should be.

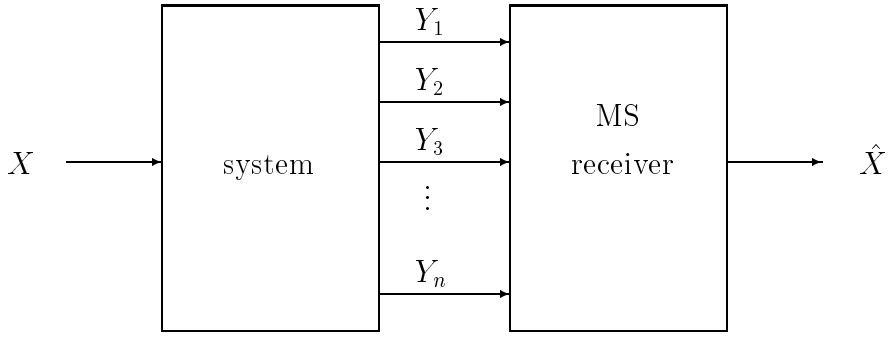
We shall say more about this during our next lecture.

Lecture 32

Random Processes Part 2

32.1 Orthogonality Principle

We can easily extend what we were doing in Section 31.3. Consider the following block diagram:



In this scenario, we are going to make use of receiver inputs $Y_1, Y_2, \cdots, Y_n$ to form a linear mean square estimate of X of form

$$\hat{X} = A_1Y_1 + A_2Y_2 + A_3Y_3 + \cdots + A_nY_n, \tag{32.1}$$

where $A_1, A_2, \cdots, A_n$ are constants that are chosen so that the mean square estimation error $E[(X - \hat{X})^2]$ is a minimum. (Previously, in Section 31.3, we took $n = 2$. Now, n can be any positive integer.) If you take the partial derivative of the mean square estimation error with respect to each A_i and set this derivative equal to zero, then you will obtain n linear equations in the n A_i's that can be solved simultaneously to obtain the unique solution for the A_i's. An easy way to remember these equations is to use the orthongonality principle, which says the following:

Orthogonality Principle: *The estimation error $X - \hat{X}$ is orthogonal to each of the Y_i's used to form the linear estimate (32.1).* That is,

$$E[(X - \hat{X})Y_i] = 0, \quad i = 1, 2, \cdots, n. \tag{32.2}$$

(Note: When you say that RV's U, V are *orthogonal*, you simply mean that $E[UV] = 0$, that is, their correlation is zero. The concept of orthogonal RV's is a convenient generalization of the concept of perpendicular geometric vectors.)

It is easy to work out in this case what the specific system of n equations in n unknowns is that arises from the orthogonality relations (32.2). The reader can verify that the relations (32.2) are reducible to:

$$R_Y \begin{bmatrix} A_1 \\ A_2 \\ A_3 \\ \vdots \\ A_n \end{bmatrix} = \begin{bmatrix} E[XY_1] \\ E[XY_2] \\ E[XY_3] \\ \vdots \\ E[XY_n] \end{bmatrix}, \tag{32.3}$$

where R_Y is the $n \times n$ correlation matrix of $Y_1, Y_2, \cdots, Y_n$, defined in an earlier lecture to be the matrix

$$R_Y = [E[Y_i Y_j]\,]_{i,j=1,2,3,\cdots,n}$$

(That is, the element in row i and column j of R_Y is the correlation $E[Y_i Y_j]$ between RV Y_i and RV Y_j.) The correlations $E[XY_i]$ appearing on the right side of (32.3) are called *cross-correlations*, because X is an input to the system and the Y_i's are outputs from this system (you are going *across* from input to outputs). The correlations $E[Y_i Y_j]$ appearing on the left side of (32.3) are called *autocorrelations*, because these are correlations computed internally among the Y_i's. In our coverage of random processes, you will ultimately see treatment of the concepts of autocorrelation and cross-correlation with regard to a random signal imposed as input to a system and the corresponding random signal received as output to the system. Such autocorrelations and cross-correlations are all you need to know when you design an optimal mean square filter at the receiving end of the system, where the purpose of the filter is to estimate the system input random signal.

Example 32.1. Again, let X be the random variable which is the input to our system. Suppose we have exactly three outputs from our system that are to be used to form our mean square estimate of X: Y_1, Y_2, Y_3. That is, our estimate takes the form

$$\hat{X} = A_1 Y_2 + A_2 Y_2 + A_3 Y_3.$$

Then the system of equations (32.2) reduce in this case to

$$\begin{bmatrix} E[Y_1^2] & E[Y_1 Y_2] & E[Y_1 Y_3] \\ E[Y_1 Y_2] & E[Y_2^2] & E[Y_2 Y_3] \\ E[Y_1 Y_3] & E[Y_2 Y_3] & E[Y_3^2] \end{bmatrix} \begin{bmatrix} A_1 \\ A_2 \\ A_3 \end{bmatrix} = \begin{bmatrix} E[XY_1] \\ E[XY_2] \\ E[XY_3] \end{bmatrix}.$$

Given the three different cross-correlations on the right side and the six different autocorrelations on the left side, you can solve to find the unique values of A_1, A_2, A_3.

Example 32.2. This example points out that the straight line receiver is a special case of the MS estimation theory that we have presented in this section. In this scenario, we have one system input X and one system output Y. The straight line receiver generates the estimate

$$\hat{X} = AY + B, \tag{32.4}$$

where the constants A, B are chosen so that the mean square estimation error $E[(X - \hat{X})^2]$ is minimized. We can regard this estimate as the special case of the estimate

$$\hat{Y} = AY_1 + BY_2 \tag{32.5}$$

when random variable Y_1 is taken to be Y and random variable Y is taken to be 1. The A, B needed to form the estimate (32.5) are found by solving

$$\begin{bmatrix} E[Y_1^2] & E[Y_1Y_2] \\ E[Y_1Y_2] & E[Y_2^2] \end{bmatrix} \begin{bmatrix} A \\ B \end{bmatrix} = \begin{bmatrix} E[XY_1] \\ E[XY_2] \end{bmatrix}.$$

Substituting $Y_1 = Y$ and $Y_2 = 1$, you obtain the system

$$\begin{bmatrix} E[Y^2] & E[Y] \\ E[Y] & 1 \end{bmatrix} \begin{bmatrix} A \\ B \end{bmatrix} = \begin{bmatrix} E[XY] \\ E[X] \end{bmatrix} \tag{32.6}$$

that is to be solved to find the A, B for the straight line estimate (32.4). It is not hard to solve the equation (32.6). You get:

$$A = \frac{\rho_{X,Y}\sigma_X}{\sigma_Y} \tag{32.7}$$

$$B = \mu_X - \frac{\rho_{X,Y}\sigma_X}{\sigma_Y} \tag{32.8}$$

32.1.1 Curious Fact

If RV's X, Y are bivariate Gaussian, then we know from Chapter 4 that

$$E[X|Y = y] = \mu_X + \frac{\rho_{X,Y}\sigma_X}{\sigma_Y}(y - \mu_Y). \tag{32.9}$$

If we re-write the right side of (32.9) in the "slope intercept form" $Ay + B$, then we see that the slope and intercept A, B are given by the equations (32.7)-(32.8). In other words, if we have any pair of RV's X, Y and we want to see what the straight line estimate of X will be when $Y = y$, then we just substitute the value y into the right side of (32.9).

Is what we have just pointed out simply a "curious fact"? Or is there some deep reason why this is so? Your answer should be: DEEP REASON. Here is why. Suppose we start with any random pair (X, Y). We want to find the straight line estimate $AY + B$ for X based on Y. We

know that A, B can be found by solving equation (32.6). We will prove by clever indirect reasoning that the solutions A, B to (32.6) are indeed given by equations (32.7)-(32.8). Pick another random pair (X^*, Y^*) which is *bivariate Gaussian* and which satisfies

$$\mu_{X^*} = \mu_X \tag{32.10}$$
$$\mu_{Y^*} = \mu_Y \tag{32.11}$$
$$\sigma_{X^*} = \sigma_X \tag{32.12}$$
$$\sigma_{Y^*} = \sigma_Y \tag{32.13}$$
$$\rho_{X^*,Y^*} = \rho_{X,Y} \tag{32.14}$$

We argue that $E[X^*|Y^*]$ is the straight line estimate of X^* based on Y^*. First, recall from an earlier lecture that $E[X^*|Y^*]$ is the estimate of X^* based on Y^* which gives the smallest possible mean square estimation error *among all possible estimates*. Secondly, since (X^*, Y^*) is bivariate Gaussian, we know from Chapter 4 that $E[X^*|Y^*]$ is of straight line form $A^*Y + B^*$ with

$$A^* = \frac{\rho_{X^*,Y^*}\sigma_{X^*}}{\sigma_{Y^*}} \tag{32.15}$$
$$B^* = \mu_{X^*} - \frac{\rho_{X^*,Y^*}\sigma_{X^*}}{\sigma_{Y^*}} \tag{32.16}$$

By definition, the straight line estimate of X^* based on Y^* is that estimate of straight line form which yields the smallest mean square estimation error among all estimates of straight line form. Since $E[X^*|Y^*]$ is of straight line form and yields the smallest possible mean square estimation error among *all possible estimates*, it must automatically yield the smallest possible mean square estimation error among *all straight line estimates*. Therefore, $E[X^*|Y^*] = A^*Y + B^*$ is the straight line estimate of X^* based on Y^*, where A^* and B^* satisfy the equations (32.15)-(32.16). The coefficients A^*, B^* must therefore also satisfy the equation

$$\begin{bmatrix} E[(Y^*)^2] & E[Y^*] \\ E[Y^*] & 1 \end{bmatrix} \begin{bmatrix} A^* \\ B^* \end{bmatrix} = \begin{bmatrix} E[X^*Y^*] \\ E[X^*] \end{bmatrix} \tag{32.17}$$

Because of equations (32.10)-(32.14), we have

$$E[Y^*] = E[Y]$$
$$E[X^*] = E[X]$$
$$E[(Y^*)^2] = E[Y^2]$$
$$E[X^*Y^*] = E[XY]$$

and therefore the equation (32.17) has the same solutions as the equation (32.6), that is, $A^* = A$ and $B^* = B$. Now the solutions A, B to equation (32.6) give the straight line estimate $AY + B$ of X based on Y. Since $A = A^*$ and $B = B^*$, we can conclude that A, B must be given by the right sides of the equations (32.15)-(32.16). But, because of (32.10)-(32.14), the right sides of (32.15)-(32.16) are the same as the right sides of (32.7)-(32.8). Therefore, we conclude that equations (32.7)-(32.8) do indeed give the constants A, B in the straight line estimate $AY + B$ of X based on Y.

32.2 More 1-D Cross-Section Examples

Example 32.3. Let random variable U be Uniform(0,1). Define the following continuous time random process:

$$X(t) = t - U, \quad t \geq 0.$$

Fix t, and let us determine the PDF of the 1-D cross-section random variable $X(t)$. Since U is Uniform over the interval $[0, 1]$, the random variable $-U$ must be Uniform over the interval $[-1, 0]$. (For *any* RV U, the PDF of $-U$ will always be the reflection of the PDF of U. If we reflect the interval $[0, 1]$, we obtain the interval $[-1, 0]$.) Writing

$$t - U = t + (-U),$$

we see that the PDF of the RV $t-U$ is obtained by translating the PDF of $-U$ t units to the right. If we translate the interval $[-1, 0]$ t units to the right, we obtain the interval $[t-1, t]$. Therefore, the RV $X(t) = t - U$ must be Uniform over the interval $[t - 1, t]$. Its density function is a rectangular pulse over this interval, of amplitude 1. Therefore, this density function is given by

$$f_{X(t)}(x) = u(x - [t - 1]) - u(x - t), \quad -\infty < x < \infty.$$

Example 32.4. We determine the probability distributions of the 1-D cross-sections of the Poisson process $X(t)$ whose arrival rate is the parameter λ. Let $T_1, T_2, T_3, \cdots$ be an infinite sequence of independent RV's which are each exponentially distributed with mean $1/\lambda$. Then we can model the random arrival times of the Poisson process $X(t)$ as:

$$\begin{aligned}
first\ arrival\ time &= T_1 \\
second\ arrival\ time &= T_1 + T_2 \\
third\ arrival\ time &= T_1 + T_2 + T_3 \\
\vdots &= \vdots \\
k - th\ arrival\ time &= T_1 + T_2 + \cdots + T_k \\
\vdots &= \vdots
\end{aligned}$$

(The RV's T_i are typically referred to as the *interarrival times* of the Poisson process, because T_i gives the length of time between the $(i-1)$-st arrival and the i-th arrival.) If we think of $X(t)$ as a random signal in terms of its variation in time (that is, think of $X(t)$ as a random realization of the process), then we can express $X(t)$ in terms of the T_i's as

$$X(t) = \sum_{i=1}^{\infty} u(t - [T_1 + T_2 + \cdots + T_i]), \quad 0 \leq t < \infty.$$

Fix a time $t > 0$, and take the 1-D cross-section RV $X(t)$. The random variable $X(t)$ is a discrete RV which takes as possible value any nonnegative integer

$$0, 1, 2, \cdots.$$

Let k be any one of these values and let us compute $P[X(t) = k]$. Note that the event $\{X(t) = k\}$ occurs if and only if the k-th arrival has occured at time $\leq t$ but the $(k+1)$-st arrival will occur at time $> t$. The time at which the k-th arrival occurs is

$$Y = T_1 + T_2 + \cdots + T_k,$$

and the time at which the $(k+1)$-st arrival occurs is

$$Y + T_{k+1} = (T_1 + T_2 + \cdots + T_k) + T_{k+1}.$$

We conclude that

$$\{X(t) = k\} = \{Y \leq t, \;\; Y + T_{k+1} > t\},$$

and therefore

$$P[X(t) = k] = P[Y \leq t, \;\; Y + T > t],$$

where T is the random variable T_{k+1}. Note that the RV's Y and T are independent. Apply the following linear transformation to the pair (Y, T) to obtain a new pair of RV's (U, V):

$$\begin{aligned} U &= Y \\ V &= Y + T \end{aligned}$$

By Section 28.3,

$$f_{U,V}(u, v) = f_Y(u) f_T(v - u) = f_Y(u) \lambda \exp(-\lambda(v - u)), \;\; 0 \leq u < v < \infty \; (zero \; elsewhere).$$

Therefore,

$$\begin{aligned} P[X(t) = k] &= P[U \leq t, \;\; V > t] \\ &= \int_0^t \int_t^\infty f_{U,V}(u, v) dv du \\ &= \int_0^t \int_t^\infty f_Y(u) \lambda \exp(-\lambda(v - u)) dv du \\ &= \int_0^t f_Y(u) \exp(-\lambda(t - u)) du \\ &= f_Y(t) * [\exp(-\lambda t) u(t)] \end{aligned}$$

To evaluate the convolution of the signal $f_Y(t)$ with the signal $\exp(-\lambda t)u(t)$, we can take the inverse Laplace transform of the product of the Laplace transforms of these two signals. The Laplace transform of $f_Y(t)$ is

$$\mathcal{L}[f_Y(t)] = \left(\frac{\lambda}{s+\lambda}\right)^k. \tag{32.18}$$

To see this, note that Y is the sum of k independent RV's all gaving the same density $\lambda \exp(-\lambda t)$. Therefore, $f_Y(t)$ is the convolution of k signals all of which have Laplace transform $\lambda/(s+\lambda)$, and so equation (32.18) follows. Therefore, the Laplace transform of $f_Y(t) * [\exp(-\lambda t)u(t)]$ is:

$$\left(\frac{\lambda}{s+\lambda}\right)^k \left(\frac{1}{s+\lambda}\right) = \frac{\lambda^k}{(s+\lambda)^{k+1}}.$$

We can now say that

$$\begin{aligned} P[X(t)=k] &= \mathcal{L}^{-1}\left[\frac{\lambda^k}{(s+\lambda)^{k+1}}\right] \\ &= \lambda^k \exp(-\lambda t)\mathcal{L}^{-1}\left[\frac{1}{s^{k+1}}\right] \\ &= \lambda^k \exp(-\lambda t)\frac{t^k}{k!} \\ &= = \frac{\exp(-\lambda t)(\lambda t)^k}{k!}. \end{aligned}$$

We conclude that the PMF of $X(t)$ is

$$P[X(t)=k] = \frac{\exp(-\lambda t)(\lambda t)^k}{k!}, \quad k = 0, 1, 2, 3, \cdots.$$

In other words, for each fixed $t > 0$, $X(t)$ is a Poisson RV with parameter λt. It is now understandable why the process $X(t)$ is called a Poisson process: this is because all the 1-D cross-sections have a Poisson distribution. (At this point, the reader who has forgotten the expression for the PMF of a Poisson RV should refer to Appendix A in order to refresh the memory, because we will need to talk about the Poisson process now and again.)

Exercise. Let Θ be a RV uniformly distributed over the interval $[0, 2\pi]$. Let process $X(t)$ be the following random sinusoid:

$$X(t) = \sin(t + \Theta), \quad -\infty < t < \infty.$$

Show using the "CDF Method" of Section 3.7 of the textbook that the PDF of the 1-D cross-section $X(0)$ is

$$f_{X(0)}(x) = \frac{1}{\pi\sqrt{1-x^2}}, \quad -1 < x < -1 \ (zero\ elsewhere).$$

32.3 Definition of Mean Function

Let $X(t)$ be any process. The mean function $\mu_X(t)$ of the process is defined by

$$\mu_X(t) \triangleq E[X(t)], \quad for\ all\ times\ t.$$

Thus, for each fixed time t, the value $\mu_X(t)$ of the mean function at time t is simply the expected value of the 1-D cross-section RV $X(t)$. If the density $f_{X(t)}(x)$ of the cross-section $X(t)$ is known, then the mean function can be computed as follows:

$$\mu_X(t) = \int_{-\infty}^{\infty} x f_{X(t)}(x) dx.$$

Fortunately, there are many examples of random processes for which we can compute the mean function $\mu_X(t)$ without needing to know what the density $f_{X(t)}(x)$ is.

Example 32.5. We determine the mean function of the "random straight line" process considered earlier in Example 31.5. This is the process

$$X(t) = At + B, \quad t \geq 0,$$

where A, B are independent RV's each taking the values ± 1 equiprobably. We have

$$\mu_X(t) = E[X(t)] = E[At + B].$$

In computing $E[At + B]$, t is a fixed time, so when you apply the expected value operator E, treat t exactly the same way you would a constant:

$$\mu_X(t) = E[At + B] = tE[A] + E[B] = t * 0 + 0 = 0.$$

We conclude that the mean function of our random straight line process is equal to 0 for all times t. There is a way to visualize this result physically: there are four realizations of this process and let us denote them by

$$x^{(1)}(t), x^{(2)}(t), x^{(3)}(t), x^{(4)}(t).$$

Since these realizations each occur with probability 1/4, we can express the mean function in terms of these realizations by the formula:

$$\mu_X(t) = (1/4)[x^{(1)}(t) + x^{(2)}(t) + x^{(3)}(t) + x^{(4)}(t)].$$

When you plot the four realizations, you will see that two of them lie above the t axis and the other two of them lie below the t axis. The two realizations below the t axis are the negatives of the two realizations above the t axis. Therefore, when you average up all four realizations, cancellation occurs and you wind up with a mean function whose plot stays along the t axis for all t, that is, the mean function is a signal that is always equal to zero.

Remark. Suppose one performs the experiment giving rise to a random process a large number of times, thereby obtaining a large number of realizations. If you average up these realization signals, you will obtain a good approximation of the mean function $\mu_X(t)$. This procedure is called "space averaging".

Example 32.6. Let $X(t)$ be the Poisson process with arrival rate λ. Then the mean function is given by

$$\mu_X(t) = \lambda t, \quad t \geq 0. \tag{32.19}$$

We showed in Example 32.4 that the 1-D cross-section $X(t)$ has a Poisson distribution with parameter λt. The mean of a Poisson distribution is its parameter. Thus, the expected value of $X(t)$ is λt, and so equation (32.19) is true. The fact that the mean of $X(t)$ is λt makes sense because we expect λ arrivals per second, so that in t seconds, we expect λt arrivals.

32.4 Definition of Autocorrelation Function

Let $X(t)$ be any process. Its autocorrelation function $R_X(t_1, t_2)$ is the function of two variables defined by

$$R_X(t_1, t_2) \triangleq E[X(t_1)X(t_2)], \quad \textit{for all times } t_1, t_2.$$

In other words, at any two fixed times t_1, t_2, you compute the correlation between the 1-D cross-section $X(t_1)$ and the 1-D cross-section $X(t_2)$. If the joint PDF $f_{X(t_1),X(t_2)}(x_1, x_2)$ of the 2-D cross-section $(X(t_1), X(t_2))$ is known, then one can compute the autocorrelation $R_X(t_1, t_2)$ as the double integral

$$R_X(t_1, t_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 f_{X(t_1),X(t_2)}(x_1.x_2) dx_1 dx_2.$$

Fortunately, there are many examples of random processes where one can compute $R_X(t_1, t_2)$ without having to find the joint density $f_{X(t_1),X(t_2)}(x_1, x_2)$. The following example illustrates this.

Example 32.7. Again, we look at the random straight line, that is, the process

$$X(t) = At + B, \quad t \geq 0,$$

where A, B are independent RV's each taking the values ± 1 equiprobably. We can compute the autocorrelation function $R_X(t_1, t_2)$ using the expected value operator as follows:

$$\begin{aligned} R_X(t_1, t_2) &= E[X(t_1)X(t_2)] \\ &= E[(At_1 + B)(At_2 + B)] \\ &= E[t_1 t_2 A^2 + BAt_2 + ABt_1 + B^2] \\ &= t_1 t_2 E[A^2] + (t_1 + t_2)E[AB] + E[B^2] \end{aligned}$$

At this point, observe that

$$E[A^2] = E[B^2] = 1,$$

because the values of A, B are ± 1 and $(\pm 1)^2 = 1$. Also, because A, B are independent, we may write

$$E[AB] = E[A]E[B] = 0 * 0 = 0.$$

Our formula for the autocorrelation function therefore reduces to

$$R_X(t_1, t_2) = t_1 t_2 + 1. \tag{32.20}$$

Exercise. For each realization $x(t)$ of the random straight line process, sample at times $t = t_1$ and $t = t_2$ to obtain the values $x(t_1)$ and $x(t_2)$, and then multiply, thereby obtaining the product $x(t_1)x(t_2)$. Average up these products over the four realizations. See if some cancellation occurs, resulting in the simple formula (32.20).

Remark. Suppose one performs the experiment giving rise to a random process a large number of times, thereby obtaining a large number of realizations. For each of these realizations $x(t)$, suppose you sample at times $t = t_1$ and $t = t_2$ and then form the product $x(t_1)x(t_2)$. Averaging up these products over all these realizations, you get a good approximation to the autocorrelation $R_X(t_1, t_2)$. This procedure is called "space averaging".

Lecture 33

Random Processes Part 3

33.1 Definition of WSS Process

A process $X(t)$ is *wide sense stationary* (WSS) if two things hold:

(a) The mean function $\mu_X(t)$ is constant for all t.

(b) Each autocorrelation $R_X(t_1, t_2)$ depends only on $t_1 - t_2$.

Example 33.1. We show that the random sinusoid is a WSS process. We have

$$X(t) = A\cos(\omega_0 t + \Theta), \quad -\infty < t < \infty$$

Let us suppose that the amplitude A is random with no assumption about the PDF of A. We also suppose that the phase Θ is uniformly distributed in the interval $[0, 2\pi]$. Finally, we ssume that A and Θ are independent. We have to verify *both* condition (a) and condition (b) in the WSS process definition.

Verification of (a): For each fixed t, we have

$$\begin{aligned}
\mu_X(t) &= E[X(t)] \\
&= E[A\cos(\omega_0 t + \Theta)] \\
&= E[A]E[\cos(\omega_0 t + \Theta)] \\
&= E[A]\int_0^{2\pi} \cos(\omega_0 t + \theta) f_\Theta(\theta) d\theta \\
&= E[A]\int_0^{2\pi} \cos(\omega_0 t + \theta)(1/2\pi) d\theta
\end{aligned}$$

The reader can easily do the integral from 0 to 2π and show that it is zero. (When you find the antiderivative, and plug in the upper and lower limits, you will find that the limits cancel.) We conclude that

$$\mu_X(t) = 0, \quad for\ all\ t.$$

The mean function takes the constant value 0. This completes the verification of (a).

Verification of (b): For each fixed t_1, t_2, we have

$$\begin{aligned} E[X(t_1)X(t_2)] &= E[A^2\cos(\omega_0 t_1 + \Theta)\cos(\omega_0 t_2 + \Theta)] \\ &= E[A^2]E[\cos(\omega_0 t_1 + \Theta)\cos(\omega_0 t_2 + \Theta)] \\ &= E[A^2]E[(1/2)\cos(\omega_0\{t_1 - t_2\}) + (1/2)\cos(\omega_0\{t_1 + t_2\} + 2\Theta)] \\ &= (E[A^2]/2)\left[\cos(\omega_0\{t_1 - t_2\}) + \int_0^{2\pi}\cos(\omega_0\{t_1 + t_2\} + 2\theta)(1/2\pi)d\theta\right] \\ &= (E[A^2]/2)\left[\cos(\omega_0\{t_1 - t_2\}) + 0\right] = (E[A^2]/2)\cos(\omega_0\{t_1 - t_2\}) \end{aligned}$$

In the preceding, we used the trig identity

$$\cos a \cos b = (1/2)\cos(a - b) + (1/2)\cos(a + b).$$

We've shown

$$R_X(t_1, t_2) = (E[A^2]/2)\cos(\omega_0\{t_1 - t_2\})$$

which depends only on $t_1 - t_2$. This completes the verification of (b). Since both (a) and (b) hold, the random sinusoid is WSS.

Example 33.2. The random straight line process $X(t)$ of Example 32.7 satisfies

$$\mu_X(t) = 0, \quad for\ all\ t.$$

That is, the mean function is constant. However, this is only condition (a) in the definition of WSS process. You would also have to check condition (b). We established in Example 32.7 that

$$R_X(t_1, t_2) = t_1 t_2 + 1.$$

This is not a function of $t_1 - t_2$. Since condition (b) fails, the process is not WSS.

Exercise. Let X_n be the random walk process of Example 31.2. First, show that the mean function is constant by showing that

$$E[X_n] = 0, \quad for\ all\ n.$$

Then show that the process fails to be WSS by showing that condition (b) fails.

Example 33.3. Let X_n be an IID process. We show that X_n is WSS. All the X_n's have the same mean μ and the same variance σ^2. Therefore,

$$\mu_X(n) = E[X_n] = \mu,$$

and the mean function takes the constant value μ. This verifies condition (a) for WSS process. Now we verify condition (b). If $n_1 \neq n_2$, then X_{n_1} and X_{n_2} are independent, and we have

$$R_X(n_1, n_2) = E[X_{n_1} X_{n_2}] = E[X_{n_1}]E[X_{n_2}] = \mu^2.$$

On the other hand, if $n_1 = n_2$, we have

$$R_X(n_1, n_2) = E[X_{n_1}^2] = \sigma^2 + \mu^2.$$

We can combine these two different cases into one by writing

$$R_X(n_1, n_2) = \mu^2 + \sigma^2 \delta[n_1 - n_2], \quad \textit{for all } n_1, n_2. \tag{33.1}$$

Here, $\delta[n]$ denotes the discrete time delta function, defined by

$$\delta[n] = \begin{cases} 1, & n = 0 \\ 0, & n \neq 0 \end{cases}$$

The one-line expression (33.1) tells us that $R_X(n_1, n_2)$ depends only on $n_1 - n_2$. Condition (b) for WSS process is therefore satisfied. We conclude that an IID process is WSS.

33.1.1 Notation μ_X and $R_X(\tau)$

If you have a WSS process $X(t)$, then μ_X is the notation for the constant value of the mean function of the process, and $R_X(\tau)$ is the notation for the autocorrelation function of the process. The τ in $R_X(\tau)$ is called the *lag variable*. If we sample process $X(t)$ at times t_1, t_2, then we take τ to be $|t_2 - t_1|$, and we have

$$R_X(t_1, t_2) = R_X(|t_1 - t_2|) = R_X(\tau).$$

Alternatively, you can define the single variable autocorrelation function $R_X(\tau)$ in terms of the two variable autocorrelation function $R_X(t_1, t_2)$ as

$$R_X(\tau) \triangleq R_X(0, \tau).$$

Example 33.4. For the random sinusoid process $X(t)$ of Example 33.1, we have

$$\begin{aligned} \mu_X &= 0 \\ R_X(\tau) &= (E[A^2]/2)\cos(\omega_0 \tau) \end{aligned}$$

Example 33.5. For the IID process X_n of Example 33.3, we have

$$\begin{aligned} \mu_X &= \mu \\ R_X(\tau) &= \mu^2 + \sigma^2 \delta[\tau] \end{aligned}$$

Example 33.6. The discrete-time white noise process X_n is the WSS process satisfying

$$\begin{aligned} \mu_X &= 0 \\ R_X(\tau) &= \sigma^2 \delta[\tau], \end{aligned}$$

where σ^2 is the common variance of all the X_n's. In other words, for DT white noise, all the component RV's of the process have 0 mean, the same variance, and they are uncorrelated.

Example 33.7. The continuous-time white noise process $X(t)$ is the WSS process satisfying

$$\begin{aligned} \mu_X &= 0 \\ R_X(\tau) &= A\delta(\tau), \end{aligned}$$

where $\delta(\tau)$ is the continuous-time delta function, and A is a positive constant. (Unlike with discrete-time white noise, the constant A in front of the delta function in the autocorrelation function is *not* the variance of each $X(t)$, because

$$E[X(t)^2] = variance(X(t)) = +\infty, \quad for\ all\ t \tag{33.2}$$

for CT white noise. Instead, A has some other interpretation, which we will encounter later. The fact that (33.2) holds for continuous-time white noise is a consequence of the fact that $\delta(t)$ blows up at $t = 0$.)

33.2 A Filtering Example

Let Z_n be the discrete time white noise process with unit variance. This process satisfies

$$\begin{aligned} E[Z_n] &= 0, \quad for\ all\ n \\ E[Z_n^2] &= 1, \quad for\ all\ n \\ E[Z_i Z_j] &= 0, \quad i \neq j \end{aligned}$$

Let us think of Z_n as a random signal. Let us pass Z_n through a FIR linear filter which does the following:

$$Z_n \rightarrow \boxed{\begin{array}{c}\text{FIR} \\ \text{linear} \\ \text{filter}\end{array}} \rightarrow X_n = aZ_n + bZ_{n-1} + cZ_{n-2}$$

This filter is determined by its three "tap weights" a, b, c. The output random signal X_n turns out to be WSS. Let us prove this and find μ_X and $R_X(\tau)$.

First, we must check condition (a) for WSS process:

$$\begin{aligned} \mu_X(n) &= E[X_n] = E[aZ_n + bZ_{n-1} + cZ_{n-2}] \\ &= aE[Z_n] + bE[Z_{n-1}] + cE[Z_{n-2}] \\ &= a*0 + b*0 + c*0 = 0 \end{aligned}$$

Therefore, the mean function of the X process takes the constant value 0. Now we must concern ourselves with condition (b). First, note that

$$\begin{aligned} R_X(n,n) &= E[X_n^2] = E[(aZ_n + bZ_{n-1} + cZ_{n-2})^2] \\ &= a^2E[Z_n^2] + b^2E[Z_{n-1}^2] + c^2E[Z_{n-2}^2] \\ &= a^2 + b^2 + c^2 \end{aligned}$$

Let's explain how to go from line 1 to line 2 above. Write the two factors in $(aZ_n + bZ_{n-1} + cZ_{n-2})^2$ above and below one another as

$$\begin{aligned} &aZ_n + bZ_{n-1} + cZ_{n-2} \\ &aZ_n + bZ_{n-1} + cZ_{n-2} \end{aligned}$$

When we multiply these two factors, we first obtain the following factors by multiplying "straight up and down":

$$a^2Z_n^2 + b^2Z_{n-1}^2 + c^2Z_{n-2}^2. \tag{33.3}$$

There are additional cross-product terms, which are each of the form a constant times $E[Z_iZ_j]$, where $i \neq j$. Since $E[Z_iZ_j] = 0$ for $i \neq j$, we can ignore these additional terms; it is only the expected value of the terms in (33.3) which we need to be concerned about, which will give $a^2 + b^2 + c^2$.

Now let us compute $R_X(n, n-1)$:

$$\begin{aligned} R_X(n,n-1) &= E[X_nX_{n-1}] \\ &= E[(aZ_n + bZ_{n-1} + cZ_{n-2})(aZ_{n-1} + bZ_{n-2} + cZ_{n-3})] \\ &= abE[Z_{n-1}^2] + bcE[Z_{n-2}^2] \\ &= ab + bc \end{aligned}$$

To see this, write the two factors as

$$\begin{aligned} &aZ_n + bZ_{n-1} + cZ_{n-2} \\ &\qquad aZ_{n-1} + bZ_{n-2} + cZ_{n-3} \end{aligned}$$

and then take the product of just the "straight up and down" terms:

$$abZ_{n-1}^2 + bcZ_{n-2}^2, \tag{33.4}$$

ignoring the cross-product terms. The expected value of these terms, which yields $ab+bc$, is all we need.

Now let us compute $R_X(n, n-2)$:

$$\begin{aligned} R_X(n,n-2) &= E[X_n X_{n-2}] \\ &= E[(aZ_n + bZ_{n-1} + cZ_{n-2})(aZ_{n-2} + bZ_{n-3} + cZ_{n-4})] \\ &= acE[Z_{n-2}^2] = ac \end{aligned}$$

It is apparent that this is all we get if we write

$$\begin{aligned} &aZ_n + bZ_{n-1} + cZ_{n-2} \\ &\qquad\qquad aZ_{n-2} + bZ_{n-3} + cZ_{n-4} \end{aligned}$$

and then just pick out the product of the "straight up and down terms" (i.e., the terms that have the same time index).

It should be apparent from the precding that

$$R_X(n, n-\tau) = 0, \quad \tau \geq 3.$$

We have shown that

$$R_X(n_1, n_2) = \begin{cases} a^2+b^2+c^2, & n_1 = n_2 \\ ab+bc, & |n_1 - n_2| = 1 \\ ac, & |n_1 - n_2| = 2 \\ 0, & |n_1 - n_2| \geq 3 \end{cases}$$

The autocorrelations $R_X(n_1, n_2)$ clearly only depend on $n_1 - n_2$. Thus, condition (b) is satisfied and the X process is therefore WSS.

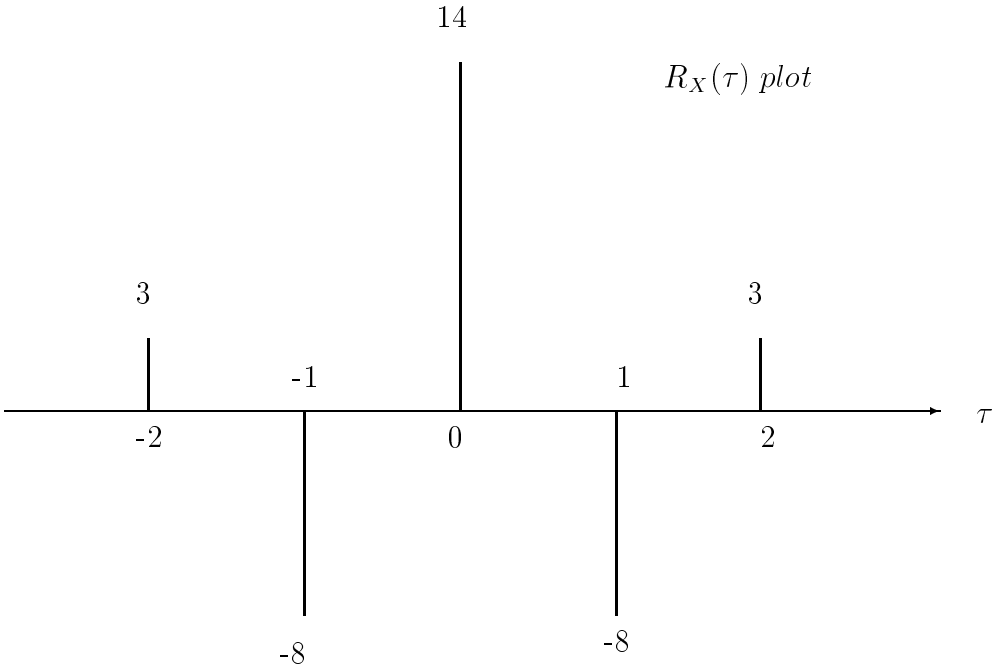
The parameter μ_X of the WSS process X_n is

$$\mu_X = 0.$$

The autocorrelation function $R_X(\tau)$ of the WSS process X_n is

$$R_X(\tau) = \begin{cases} a^2+b^2+c^2, & \tau = 0 \\ ab+bc, & \tau = \pm 1 \\ ac, & \tau = \pm 2 \\ 0, & \tau = \pm 3, \pm 4, \pm 5, \cdots \end{cases} \tag{33.5}$$

This expression is valid no matter what the three tap weights are. For example, if $a = 1$, $b = -2$, and $c = 3$, the plot of $R_X(\tau)$ would be:



Notice that $R_X(\tau)$ is of only finite extent (that is, it takes the value 0 outside a finite range of τ values). We will see later (as a consequence of Chapter 11) that this kind of autocorrelation function will always arise when you filter white noise using a time-invariant linear FIR filter—the output autocorrelation function will always be of finite extent.

Lecture 34

Random Processes Part 4

34.1 Poisson Process Properties

Let $X(t), \ \ t \geq 0$ be a Poisson process with arrival rate λ. In Example 32.4, we proved that for each fixed t, $X(t)$ is a Poisson RV with mean λt. This is the same thing as saying that the number of arrivals in the time interval $[0, t]$ is a Poisson RV with mean λt, because $X(t)$ is the number of such arrivals. In this section, we state two further properties of the Poisson process (without proof).

Property 1: Let t_1 and t_2 be any two times for which $0 < t_1 < t_2$. Then the number of Poisson process arrivals in the time interval from t_1 to t_2 is a Poisson RV with mean $\lambda(t_2 - t_1)$. This result will be true whether we include one or both or neither of the "endpoint" times t_1, t_2 in this time interval, because the probability of an arrival at precisely time t_1 or precisely time t_2 is zero. In terms of the Poisson process $X(t)$, the number of arrivals at times t satisfying $t_1 < t \leq t_2$ is $X(t_2) - X(t_1)$; note that in this time interval, we exclude endpoint t_1 and include endpoint t_2. We conclude that $X(t_2) - X(t_1)$ is a Poisson RV with mean $\lambda(t_2 - t_1)$.

Property 2: Choose times t_1, t_2, t_3, t_4 so that

$$0 \leq t_1 < t_2 \leq t_3 < t_4.$$

Then the RV's $X(t_2) - X(t_1)$ and $X(t_4) - X(t_3)$ are independent. In other words, we have statistical independence of the number of arrivals in the time interval from t_1 to t_2 and the number of arrivals in the time interval from t_3 to t_4. This property is called the *independent increments property* of the Poisson process. We have stated the independent increments property for two nonoverlapping time intervals. It extends more generally to any finite number of nonoverlapping time intervals.

Example 34.1. For the Poisson random process $X(t)$ with arrival rate λ, let's compute the probability that there are two occurences between $t = 0$ and $t = 1$ and three occurences between

$t = 1$ and $t = 2$. This is the probability that $X(1) - X(0) = 2$ and that $X(2) - X(1) = 3$. By Properties 1-2, the random variables $X(1) - X(0)$ and $X(2) - X(1)$ are independent Poisson RV's, each having mean λ. Therefore,

$$\begin{aligned} P[X(1) - X(0) = 2,\ X(2) - X(1) = 3] &= P[X(1) - X(0) = 2]P[X(2) - X(1) = 3] \\ &= \{\exp(-\lambda)\lambda^2/2\}\{\exp(-\lambda)\lambda^3/6\} \\ &= \exp(-2\lambda)\lambda^5/12 \end{aligned}$$

34.2 A Filter Design Example

In this example, we use what we learned in Section 33.2 to design a linear filter. We consider the class of all discrete-time 3-tap causal linear FIR filters. These are the filters whose impulse response function $h[n]$ takes the form

$$h[n] = a\delta[n] + b\delta[n-1] + c\delta[n-2],$$

where parameters a, b, c are the so-called "filter tap weights" and can be any real numbers whatsoever. In Section 33.2, we learned that if we use a general filter of this type to filter white noise with unit variance, then the filter output process X_n is WSS with autocorrelation function $R_X(\tau)$ given by equations (33.5).

In this example, we are going to require that the filter output autocorrelation function be

$$R_X(\tau) = \begin{cases} 10, & \tau = 0 \\ 3, & \tau = \pm 1 \\ -1, & \tau = \pm 2 \\ 0, & \tau = \pm 3, \pm 4, \pm 5, \cdots \end{cases} \tag{34.1}$$

Our job is to design the filter that will do this by figuring out what the filter tap weights a, b, c should be. Looking back at (33.5), we see that we must solve the equations

$$\begin{aligned} a^2 + b^2 + c^2 &= 10 \\ ab + bc &= 3 \\ ac &= -1 \end{aligned} \tag{34.2}$$

simultaneously for a, b, c. Unfortunately, these are simultaneous quadratic equations and not simultaneous linear equations. College algebra taught you how to solve simultaneous linear equations but not simultaneous quadratic equations. There is a general method for solving equations (34.2) called *spectral factorization*, but this method is beyond the scope of EE 3025. (We teach the spectral factorization method in our first year graduate level course in probability and random processes.) However, the equations (34.2) are sufficiently simple that we can use Matlab to find the solutions for us. Here is a simple Matlab script that you can run that will obtain a solution for a, b, c:

```
[a,b,c] = solve('a^2+b^2+c^2=10','a*b+b*c=3','a*c=-1');
a=eval(a(1)); b=eval(b(1)); c=eval(c(1));
a,b,c
```

Running this script, we obtained the solution:

$$\begin{aligned} a &= -0.34242364057618 \\ b &= 1.16372191220048 \\ c &= 2.92035911514989 \end{aligned}$$

There may be other solutions for a, b, c in which a, b, c are real numbers. Any alternate solution would also yield a filter delivering the required output autocorrelation function in response to white noise. (In addition, there may be solutions to (34.2) given by Matlab in which one or more of the a, b, c values is a complex number and not a real number. Such a solution would not be allowed in this particular application.)

34.3 Form of WSS Process Correlation Matrix

Suppose you have a discrete-time WSS process X_n. If you sample this process at N consecutive times, the resulting N RV's have correlation matrix which is the $N \times N$ matrix

$$[R_X(i-j)]_{i,j=1,2,\cdots,N}.$$

In other words, the entry in row i and column j of this correlation matrix is $R_X(i-j)$. To illustrate, suppose we take $N = 4$. The correlation matrix of 4 consecutive samples of process X_n then takes the form

$$\begin{bmatrix} R_X(0) & R_X(1) & R_X(2) & R_X(3) \\ R_X(1) & R_X(0) & R_X(1) & R_X(2) \\ R_X(2) & R_X(1) & R_X(0) & R_X(1) \\ R_X(3) & R_X(2) & R_X(1) & R_X(0) \end{bmatrix} \tag{34.3}$$

Examine the matrix (34.3) carefully. Notice that the entries on the main diagonal (going from upper left hand corner to lower right hand corner) are all equal to $R_X(0)$. Also, the entries on the two "subdiagonals" right below and right above the main diagonal are all equal to $R_X(1)$. There are two more subdiagonals right below and right above the subdiagonals containing the $R_X(1)$'s; all the entries in these two subdiagonals are equal to $R_X(2)$. There are just two remaining subdiagonals, consisting of just one entry each in the lower left corner and upper right corner—these two entries are taken to be $R_X(3)$. In general, an $N \times N$ correlation matrix for the WSS process samples at N consecutive times would have a similar "constant along subdiagonals" structure. You'd fill in the main diagonal first with $R_X(0)$ entries. Then you'd visit in order each pair of subdiagonals moving outward from the main diagonal; the constant entries along each of these pairs of subdiagonals

would be $R_X(1)$, $R_X(2)$, $R_X(3)$, etc., in that order. A matrix with this kind of "constant along subdiagonals" structure is called a *Toeplitz matrix*. There is a Matlab command `toeplitz` which can easily generate any Toeplitz matrix. We illustrate this Matlab command in the following example.

Example 34.2. Let the autocorrelation function of WSS process X_n be

$$R_X(\tau) = 2^{-|\tau|}, \quad for\ all\ integers\ \tau.$$

(You will eventually see how such an autocorrelation function might arise.) Suppose we sample this process at times $n = 1, 2, 3, 4$, thereby obtaining the RV's X_1, X_2, X_3, X_4. You can run the following Matlab script, which generated the 4×4 correlation matrix of these four RV's:

```
format rat
tau=0:3;
corr_matrix = toeplitz(2.^(-tau))
corr_matrix =

     1     1/2     1/4     1/8
    1/2     1      1/2     1/4
    1/4    1/2      1      1/2
    1/8    1/4     1/2      1
```

34.4 Properties of $R_X(\tau)$

Not any function of τ can be an autocorrelation function, because autocorrelation functions have very special properties. We state some facts about the structure of autocorrelation functions of WSS processes.

Fact 1: For any autocorrelation function $R_X(\tau)$, the value $R_X(0)$ is nonnegative and

$$|R_X(\tau)| \leq R_X(0), \quad for\ all\ \tau.$$

Fact 1 tells us that the *peak value* of an autocorrelation function $R_X(\tau)$ is always $R_X(0)$.

Fact 2: Any autocorrelation function $R_X(\tau)$ is an even function of τ, that is,

$$R_X(-\tau) = R_X(\tau), \quad for\ all\ \tau.$$

Fact 3: We point out a property of any autocorrelation function $R_X(\tau)$ called the *positive semidefiniteness property*. For a continuous-time WSS process $X(t)$, the positive semidefiniteness property says that for any positive real number T,

$$\int_0^T \int_0^T \phi(t_1)\phi(t_2)R_X(t_1 - t_2)dt_1dt_2 \geq 0$$

for *any* real-valued function ϕ for which the integral on the left side exists. For a discrete-time WSS process X_n, the positive semidefiniteness property says that for any positive integer N,

$$aR_X^N a^T \geq 0, \tag{34.4}$$

for any N-dimensional row vector a with real entries, where R_X^N is the $N \times N$ correlation matrix of process samples at N consecutive times. (This is the same thing as saying that all eigenvalues of the symmetric matrix R_X^N are ≥ 0; in matrix theory, such a symmetric matrix is said to be a positive semidefinite matrix, which explains why we call our property the positive semidefiniteness property.)

Fact 4: If the realizations of a WSS process X are all periodic with the same period T, then $R_X(\tau)$ is also periodic with period T.

Fact 5: This fact allows us to construct autocorrelation functions for either CT or DT WSS processes. If $\phi(\tau)$ is any continuous-time signal for which the convolution

$$\phi(\tau) * \phi(-\tau) \tag{34.5}$$

exists, then the function (34.5) is an autocorrelation function $R_X(\tau)$ for some continuous-time WSS process $X(t)$. Similarly, if $\phi[\tau]$ is any discrete-time signal for which the convolution

$$\phi[\tau] * \phi[-\tau] \tag{34.6}$$

exists, then the function (34.6) is an autocorrelation function $R_X(\tau)$ for some discrete-time WSS process X_n.

Remarks. Facts 1 and 5 will be proved later. Fact 2 is trivial. Fact 4 is easy to prove. I prove Fact 3 below.

Proof of Fact 3. I prove the discrete-time version of Fact 3. (The proof of the continuous-time version of Fact 3 is similar.) Let $X_1, X_2, \cdots, X_N$ be samples of WSS process X_n at times $n = 1, 2, \cdots, N$. Let a be an N-dimensional row vector

$$a = [a_1, a_2, \cdots, a_N].$$

Obviously,

$$E[(a_1X_1 + a_2X_2 + \cdots + a_NX_N)^2] \geq 0.$$

From Theorem 5.13 of Chapter 5 of your textbook,

$$E[(a_1X_1 + a_2X_2 + \cdots + a_NX_N)^2] = aR_X^N a^T.$$

Therefore, (34.4) is true.

Example 34.3. Consider again the random sinusoid

$$X(t) = A\cos(\omega_0 t + \Theta)$$

with autocorrelation function

$$R_X(\tau) = (E[A^2]/2)\cos(\omega_0 \tau). \tag{34.7}$$

Notice that all of the realizations of the random sinusoid are periodic with period $2\pi/\omega_0$. Fact 4 then tells us that the autocorrelation function should also be periodic with this same period $2\pi/\omega_0$. Checking equation (34.7), we see that this is true.

Example 34.4. Bill is working with a discrete-time WSS process X_n. He tells you that he has computed the autocorrelation function $R_X(\tau)$ to be

$$R_X(\tau) = \begin{cases} 1, & \tau = 0, \pm 1, \pm 2 \\ 0, & elsewhere \end{cases}$$

Is Bill possibly correct, or did he make a mistake in computing the autocorrelation function? Let's check the positive semidefiniteness property. The correlation matrix of four consecutive samples of the process would be

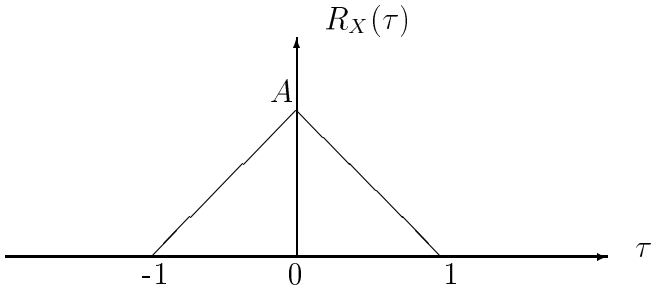
```
1  1  1  0
1  1  1  1
1  1  1  1
0  1  1  1
```

Executing the Matlab command

```
eig(toeplitz([1 1 1 0]))
```

we see that one of the four eigenvalues is -0.5616, which is negative. The positive semidefiniteness property is therefore violated. Consequently, the $R_X(\tau)$ function given above cannot possibly be an autocorrelation function. We conclude that Bill must have made a mistake in computing the autocorrelation function for his process.

Exercise. Using Fact 5, argue that the triangular pulse plotted below must be an autocorrelation function. (Hint: In EE 3015, recall what the convolution of a rectangular pulse with its reflection is.)



34.5 WSS Process Power and Variance

For a WSS process $X(t)$, power P_X and variance σ_X^2 are two important parameters of the process. These are defined by

$$\begin{aligned} P_X &\triangleq E[X(t)^2] = R_X(0) \\ \sigma_X^2 &\triangleq Var[X(t)] = P_X - \mu_X^2 \end{aligned}$$

Discussion. The parameter σ_X^2 is clear: it is the common variance of all the 1-D cross-sections of the process. The power figure P_X is a little more problematic for us at this point in time, if we try to correlate with the power concept you had in EE 3015. In EE 3015, you learned that the power generated by a power signal $x(t)$ could be computed as a time average

$$< x(t)^2 >= \lim_{T\to\infty} \frac{1}{T} \int_0^T x(t)^2 dt.$$

(We will customarily use brackets $< x(t)^2 >$ when we want to take a time average of whatever is inside the brackets; in this case, we are time-averaging $x(t)^2$, the square of the power signal $x(t)$.) It is true that the realizations of a WSS process are power signals, and it is tempting to try to define the power generated by the process as the power generated by a realization. Unfortunately, *different realizations can generate different power figures*, and if we want a power figure for the process as a whole, it must not be dependent on the realization. Taking the power figure P_X to be defined as $R_X(0)$ turns out to be the right thing to do, in the sense that if we take a large number of realizations and average up their power figures, then we will obtain $R_X(0)$, approximately. We will continue this discussion later on, after we have discussed time and space averaging and ergodic processes. At that point in time, we can make a full justification for using $R_X(0)$ as our power figure. For the time being, we can at least have you do power calculations using the $R_X(0)$ figure for the power, even if you do not completely understand as yet why this number should be related to power.

Example 34.5. For the random sinusoid

$$X(t) = A\cos(\omega_0 t + \Theta)$$

we worked out earlier that the process mean is $\mu_X = 0$. Therefore, σ_X^2 and P_X are the same for this process. The autocorrelation function is

$$R_X(\tau) = (E[A^2]/2)\cos(\omega_0 \tau),$$

and so

$$\sigma_X^2 = P_X = R_X(0) = E[A^2]/2.$$

Let's see if the power figure $E[A^2]/2$ makes sense for this process. A realization of this process would be the sinusoid

$$A\cos(\omega_0 t + \Theta)$$

in which A is fixed. You know from your basic sophomore circuits course that the power generated by a sinusoid is one half the square of the amplitude, which is $A^2/2$ in this case. Note that this figure is *random*, because the amplitude A changes from realization to realization. If we average this power figure over all the realizations, we obtain

$$E[A^2/2] = E[A^2]/2 = P_X.$$

So, at least for the random sinusoid, it does make sense to think of $P_X = R_X(0)$ as the average of the power figures across realizations of the process.

34.6 First Order Linear Predictor Design

Let $X(t)$ be a continuous-time or discrete-time WSS process with autocorrelation function $R_X(\tau)$. Let t be some fixed future time, and suppose that you have examined a realization of the process up to time $t-\tau$, which is τ seconds previous to time t. Based upon what the realization is doing at time $t-\tau$, your goal is to build a predictor to predict (estimate) what the realization will be doing at time t, as indicated in the following block diagram:

$$X(t-\tau) \rightarrow \boxed{\text{predictor}} \rightarrow \hat{X}(t)$$

The input to the predictor is the random variable $X(t-\tau)$, the process sampled at time $t-\tau$. The predictor output is your prediction $\hat{X}(t)$ of $X(t)$, the process sampled at time t. This is called a *first order* predictor, because we are using just the one sample $X(t-\tau)$ to form the prediction $\hat{X}(t)$. We will also require that the predictor be *linear*, meaning that the prediction $\hat{X}(t)$ must be some fixed constant multiple of $X(t-\tau)$:

$$\hat{X}(t) = AX(t-\tau). \tag{34.8}$$

To design the first order linear predictor, we must choose the constant A appropriately. We will choose A in order that the prediction $\hat{X}(t)$ of $X(t)$ provide the best possible mean square fit to $X(t)$. That is, we will choose A so that

$$E[(X(t) - \hat{X}(t))^2], \tag{34.9}$$

the so-called *mean square prediction error*, will be minimized. Note that predictor design falls within the framework of mean square estimation theory. That is, $\hat{X}(t)$ is a mean square estimate of $X(t)$ based on $X(t-\tau)$, and in this framework the prediction error (34.9) is just mean square estimation error. The orthogonality principle of linear mean square estimation theory (covered in Section 32.1) can now be used to find the optimum choice of A. In our problem here, the orthogonality principle says that

- *The prediction error $X(t) - \hat{X}(t)$ is orthogonal to $X(t-\tau)$, the only sample that will be used to form the linear prediction $\hat{X}(t)$ according to (34.8).*

We obtain one single orthogonality relation from the orthogonality principle, namely, the equation

$$E[(X(t) - \hat{X}(t))X(t-\tau)] = 0.$$

Substituting in for $\hat{X}(t)$, this equation becomes

$$E[(X(t) - AX(t-\tau))X(t-\tau)] = 0,$$

which simplifies to

$$E[X(t)X(t-\tau)] - AE[X(t-\tau)^2] = 0.$$

Note that

$$\begin{aligned} E[X(t)X(t-\tau)] &= R_X(\tau) \\ E[X(t-\tau)^2] &= R_X(0) \end{aligned}$$

Therefore, we have the following formula for our optimized prediction coefficient A, purely in terms of the autocorrelation function of the X process:

$$A = \frac{R_X(\tau)}{R_X(0)}.$$

In our next lecture, we will examine the mean square prediction error that results from the first order linear predictor, which will lead us to some interesting conclusions.

Lecture 35

Random Processes Part 5

35.1 MS Prediction Error for First Order Predictor

As in Section 34.6, we have a WSS process $X(t)$. In Section 34.6, we showed that

$$\hat{X}(t) = AX(t-\tau),$$

where

$$A = \frac{R_X(\tau)}{R_X(0)},$$

is the first order (linear) predictor for $X(t)$ based on $X(t-\tau)$. We are now going to work out the MS prediction error

$$E[(X(t) - \hat{X}(t))^2]$$

for the first order predictor. First, we write

$$E[(X(t) - \hat{X(t)})^2] = E[((X(t) - \hat{X}(t))X(t)] - E[((X(t) - \hat{X}(t))\hat{X}(t)].$$

The second term on the right hand side is zero, by the orthogonality principle. Therefore,

$$\begin{aligned} E[(X(t) - \hat{X(t)})^2] &= E[((X(t) - \hat{X}(t))X(t)] \\ &= E[(X(t) - AX(t-\tau))X(t)] \\ &= E[X(t)^2] - AE[X(t-\tau)X(t)] \\ &= R_X(0) - AR_X(\tau) \end{aligned}$$

Plugging in $A = R_X(\tau)/R_X(0)$ and simplifying, we obtain the following expression for the prediction error of the first order predictor:

$$first\ order\ MS\ prediction\ error\ = \frac{R_X(0)^2 - R_X(\tau)^2}{R_X(0)}. \tag{35.1}$$

We can draw two conclusions from the expression (35.1) for the first order MS prediction error:

Conclusion 1: Fact 1 about $R_X(\tau)$ given in Section 34.4 is now clearly seen to be true. (This is the fact that says that $R_X(0)$ is the peak value of $R_X(\tau)$.) Since MS prediction error is obviously a nonnegative quantity, we see from formula (35.1) that

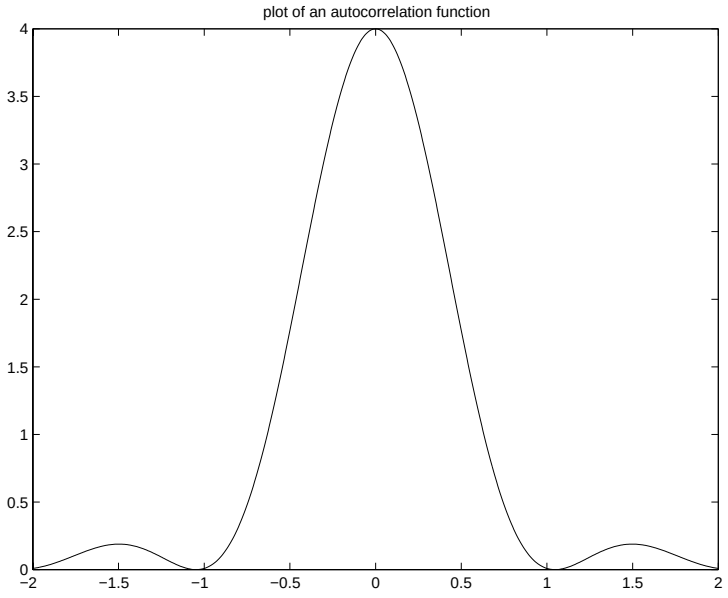
$$R_X(0)^2 \geq R_X(\tau)^2,$$

which reduces to

$$R_X(0) \geq R_X(\tau)$$

if you take the positive square root of both sides.

Conclusion 2: Suppose you want to predict $X(t)$ and you have available to you all observations $X(t-\tau)$ for $\tau \geq 1$. (In other words, you have all observations of the process 1 second or more in the past.) If you are designing a first order predictor, you can only use one of these past observations. Which one should you use? Looking at formula (35.1), you should clearly pick the observation $X(t-\tau^*)$, where τ^* is the value of $\tau \geq 1$ for which $|R_X(\tau)|$ is as close as possible to $R_X(0)$, for it is in this case that the MS prediction error will be as small as it can possibly be.



Example 35.1. Suppose the autocorrelation function $R_X(\tau)$ of WSS process $X(t)$ looks like the above. (We suppose that $R_X(\tau)$ vanishes outside the interval $-2 \leq \tau \leq 2$.) Suppose our first order

predictor must use one of the observations $X(t-\tau)$ to predict $X(t)$, where τ must be ≥ 1. From the plot, we see that $\tau = 1.5$ would be the best choice (this is the τ value where the secondary peak in $R_X(\tau)$ occurs). That is, if we take our predictor of the form

$$\hat{X}(t) = AX(t-1.5),$$

we will obtain the smallest possible MS prediction error if you are restricted to choosing an observation 1 second or more in the past on which to base your prediction. We also see from the plot that $\tau = 1$ would be a poor choice (don't use $X(t-1)$ to predict $X(t)$!). Choosing $\tau = 2$ would also be a poor choice. The choices $\tau = 1$ and $\tau = 2$ would give the worst possible MS prediction error.

35.2 k-th order Linear Predictor

In this section, we go beyond the first order predictor to "higher order" predictors. For simplicity, we now assume the process is discrete-time rather than continuous-time. We are given a DT WSS process X_n. For some fixed positive integer k, we want to build a predictor to form a prediction $\hat{X}_n$ of X_n based on

$$X_{n-1}, X_{n-2}, X_{n-3}, \cdots, X_{n-k},$$

the observations of the process at the k discrete times immediately previous to time n. This predictor is called the k-order linear predictor. We can conceptualize our k-th order predictor via the block diagram

$$X_{n-k}, \cdots, X_{n-3}, X_{n-2}, X_{n-1} \rightarrow \boxed{\begin{array}{c} k\text{-th order} \\ \text{linear} \\ \text{predictor} \end{array}} \rightarrow \hat{X}_n,$$

where the prediction $\hat{X}_n$ takes the form

$$\hat{X}_n = A_1 X_{n-1} + A_2 X_{n-2} + A_3 X_{n-3} + \cdots + A_k X_{n-k}. \tag{35.2}$$

In (35.2), we must choose the predictor coefficients $A_1, A_2, \cdots, A_k$ to minimize the MS prediction error $E[(X_n - \hat{X}_n)^2]$. The orthogonality principle says that

$$E[(X_n - \hat{X}_n)X_i] = 0, \quad i = n-1, n-2, n-3, \cdots, n-k.$$

This reduces to the following system of k linear equations in the k unknowns $A_1, A_2, \cdots, A_k$:

$$R_X \begin{bmatrix} A_1 \\ A_2 \\ A_3 \\ \vdots \\ A_k \end{bmatrix} = \begin{bmatrix} R_X(1) \\ R_X(2) \\ R_X(3) \\ \vdots \\ R_X(k) \end{bmatrix},$$

where R_X is the $k \times k$ correlation matrix of k consecutive X_n process samples. For example, for $k = 2$ (the second order predictor), this system is

$$\begin{bmatrix} R_X(0) & R_X(1) \\ R_X(1) & R_X(0) \end{bmatrix} \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} = \begin{bmatrix} R_X(1) \\ R_X(2) \end{bmatrix}$$

For $k = 2$ (the third order predictor), the system becomes

$$\begin{bmatrix} R_X(0) & R_X(1) & R_X(2) \\ R_X(1) & R_X(0) & R_X(1) \\ R_X(2) & R_X(1) & R_X(0) \end{bmatrix} \begin{bmatrix} A_1^* \\ A_2^* \\ A_3^* \end{bmatrix} = \begin{bmatrix} R_X(1) \\ R_X(2) \\ R_X(3) \end{bmatrix},$$

where I have labelled the three predictor coefficients as A_1^*, A_2^*, A_3^*, to emphasize the fact that the first two predictor coefficients A_1^* and A_2^* for the third order predictor will not coincide with the respective predictor coefficients A_1, A_2 of the second order predictor.

See Recitation 12, Experiment 5, where some examples of second and third order linear predictors are worked out.

35.3 An Example

The autocorrelation function of a WSS process can be used to directly compute the correlation between two samples of the process at different times. However, the autocorrelation function can be used to determine other things as well. I illustrate this in the following example.

Let X_n be a WSS process and let its mean μ_X and autocorrelation function $R_X(\tau)$ be given by

$$\begin{aligned} \mu_X &= 1 \\ R_X(\tau) &= 5(2^{-|\tau|}) + 1 \end{aligned}$$

Let's compute each of the following things:

(a): P_X

(b): σ_X^2

(c): $E[(X_5 - X_8)^2]$

(d): ρ_{X_5, X_8}

Solution to (a).

$$P_X = R_X(0) = 6.$$

Solution to (b).

$$\sigma_X^2 = P_X - \mu_X^2 = 6 - 1^2 = 5.$$

Solution to (c).

$$\begin{aligned} E[(X_5 - X_8)^2] &= E[X_5^2 - 2X_5X_8 + X_8^2] \\ &= E[X_5^2] - 2E[X_5X_8] + E[X_8^2] \\ &= R_X(0) - 2R_X(3) + R_X(0) = 8.75 \end{aligned}$$

Solution to (d).

$$Cov(X_5, X_8) = E[X_5X_8] - \mu_X^2 = R_X(3) - 1 = 5/8.$$

You divide this by

$$\sigma_{X_5}\sigma_{X_8} = \sqrt{\sigma_X^2}\sqrt{\sigma_X^2} = 5.$$

The correlation coefficient is therefore 1/8.

Exercise. In the preceding example, suppose that you are told that the correlation coefficient between two process samples is 1/128. How far apart are the two samples in time? (Assume that the samples X_n are taken one second apart.)

35.4 Introduction to Ergodic Processes

A discrete-time WSS process X_n is said to be *ergodic* if

$$\lim_{N\to\infty} \frac{\sum_{n=1}^N X_n}{N} = \mu_X \tag{35.3}$$

holds, and if

$$\lim_{N\to\infty} \frac{\sum_{n=1}^N X_n X_{n-\tau}}{N} = R_X(\tau) \tag{35.4}$$

holds for every integer τ. The limits in (35.3)-(35.4) are in the *stochastic convergence sense* which we defined back in the Statistics section of the course.

Intuitively, here is what the stochastic convergence statements (35.3)-(35.4) say: If the process X_n is ergodic, it is highly likely that the observed realization signal x_n will satisfy

$$\frac{\sum_{n=1}^N x_n}{N} \approx \mu_X \tag{35.5}$$

and

$$\frac{\sum_{n=1}^N x_n x_{n-\tau}}{N} \approx R_X(\tau) \tag{35.6}$$

if the number of samples N is large enough. In other words, for an ergodic process you can *time average* along just one realization to get μ_X and $R_X(\tau)$ estimates given by the left sides of (35.5) and (35.6), respectively.

It is easy to modify (35.5) and (35.6) to get the time averaging estimates for a continuous time ergodic WSS process $X(t)$. In this case, you are highly likely to get an observed realization signal $x(t)$ for which

$$\frac{\int_0^T x(t)dt}{T} \approx \mu_X$$

and

$$\frac{\int_0^T x(t)x(t-\tau)dt}{T} \approx R_X(\tau),$$

if T is large enough.

We need a criterion which will tell us when we have an ergodic process. Here is a criterion which will work for many Gaussian WSS processes (we cover Gaussian processes in Lecture 36). This result can be found in many textbooks.[1]

Useful Result: Suppose you have a WSS discrete-time or continuous-time Gaussian process X. Then the process is ergodic if

$$\lim_{\tau\to\infty} R_X(\tau) = \mu_X^2. \tag{35.7}$$

Remark. The reader may wonder why we required that the process be Gaussian. Unfortunately, if you have a nonGaussian WSS process, you cannot tell whether the process is ergodic by looking at $R_X(\tau)$ alone. You would need to look at more extensive statistical properties of the process to determine whether the process is ergodic—these properties would be too complicated to present in a first course on random processes such as EE 3025.

Example 35.2. Suppose you have a continuous-time Gaussian WSS process $X(t)$ with the following mean and autocorrelation properties:

$$\begin{aligned} \mu_X &= 0 \\ R_X(\tau) &= \exp(-|\tau|) \end{aligned}$$

We have

$$\lim_{\tau\to\infty} R_X(\tau) = \lim_{\tau\to\infty} \exp(-|\tau|) = 0 = \mu_X^2.$$

The condition (35.7) is satisfied. Therefore, this process is ergodic.

Exercise. Is an IID process ergodic? Why or why not?

I will talk further about Ergodic Processes during Lecture 36.

[1] See, for example, page 111 of *Introduction to Ergodic Theory*, by Ya. G. Sinai (Princeton University Press, 1976).

Lecture 36

Random Processes Part 6

36.1 More About Ergodic Processes

Example 36.1. An IID process X_n is ergodic. To see this, notice that

$$\lim_{n\to\infty} \frac{X_1 + X_2 + \cdots + X_n}{n} = \mu_X,$$

stochastically, by the law of large numbers, since the X_i's are all independent samples from the same probability distribution with mean μ_X. Then, notice that

$$\lim_{n\to\infty} \frac{X_1^2 + X_2^2 + \cdots + X_n^2}{n} = E[X_1^2] = R_X(0) = P_X,$$

stochastically, by the law of large numbers, since the X_i^2's are all independent and have the same probability distribution with mean $R_X(0)$. We conclude that we can do time-averaging along one fixed realization to estimate the process mean μ_X or the process power P_X. To complete the proof that the process X_n is ergodic, we also must show stochastic convergence of the average

$$\lim_{n\to\infty} \frac{\sum_{i=1}^{n} X_i X_{i+\tau}}{n}$$

to $R_X(\tau)$ for any positive integer τ. Let's prove this for $\tau = 1$. (The proof for $\tau > 1$ will be similar.) We will show that the average

$$\frac{1}{2n}\left[X_1X_2 + X_2X_3 + X_4X_5 + \cdots + X_{2n-1}X_{2n} + X_{2n}X_{2n+1}\right] \tag{36.1}$$

converges stochastically to $R_X(1)$. We cannot directly use the law of large numbers to conclude this, because the terms being summed up in (36.1) are *dependent*. But, we can express the average (36.1) as an average of two other averages as follows:

$$(1/2)\left\{\frac{1}{n}(X_1X_2 + X_3X_4 + \cdots + X_{2n-1}X_{2n}) + \frac{1}{n}(X_2X_3 + X_4X_5 + \cdots + X_{2n}X_{2n+1})\right\}.$$

Each of the two separate averages in the preceding expression involves independent terms, so that you can use the law of large numbers to conclude that each separate average converges stochastically to $R_X(1)$. The average of these two separate averages, which is (36.1), therefore converges stochastically to

$$(1/2)\{R_X(1) + R_X(1)\} = R_X(1),$$

too.

Example 36.2. A Gaussian white noise process (abbreviated *GWN process*) is ergodic, both in continuous and discrete time. Here is how you can see this. Suppose you have a discrete-time GWN process X_n. Then, it is an IID process and must therefore be ergodic by Example 36.1. Or, use the "Useful Result" given in Section 35.4. Both the discrete-time and continuous-time GWN processes satisfy

$$\lim_{\tau \to \infty} R_X(\tau) = 0,$$

because

$$R_X(\tau) = 0, \quad \tau \neq 0,$$

for such processes. (This is because $R_X(\tau)$ in this case is either a constant multiple of the discrete time delta function $\delta[\tau]$ or the continuous time delta function $\delta(\tau)$.) But, by definition, white noise has zero mean and so we can say that

$$\lim_{\tau \to \infty} R_X(\tau) = \mu_X^2.$$

This is condition (35.7) of the "Useful Result" in Section 35.4, and so that Useful Result tells us that GWN is ergodic. (The Useful Result applies to Gaussian processes and GWN is a special case of a Gaussian process. Gaussian processes are covered in the next section.)

Remark. Not every WSS process is ergodic. We give a couple of examples of WSS processes which are nonergodic (i.e., not ergodic).

Example 36.3. Consider the random sinusoid process

$$X(t) = A\cos(\omega_0 t + \Theta), \quad -\infty < t < \infty, \tag{36.2}$$

where A^2 is random and Θ is uniformly distributed in $[0, 2\pi]$ and independent of A. We know that this process is WSS. We argue that this process is nonergodic. For a particular realization of (36.2), you get a sinusoid in which the amplitude A and the phase angle Θ take fixed values. Sophomore level EE students learn that the power generated by a sinusoid is $1/2$ the square of the amplitude (it does not depend on the phase angle!). Therefore, the power generated by the random signal on the right side of equation (36.2) is $A^2/2$, a *random variable*. If the process were ergodic, all realizations (with probability one) would generate power equal to the same number. Therefore, the random sinusoid process $X(t)$ is nonergodic.

Example 36.4. This is perhaps the classic example of a nonergodic process. You have a box containing two coins, one unbiased, and the other biased with probability of heads equal to 2/3. Perform the following random experiment: First, select a coin from the box at random. Then, flip that coin an infinite number of times. On each flip ($n = 1, 2, 3, \cdots$), declare $X_n = 1$ if heads occurs and $X_n = 0$ if tails occurs. The resulting process X_n is WSS but not ergodic. (Hint: Show that the time-average behavior depends heavily on whether the realization is being generated according to the biased coin or the unbiased coin. If you get stuck, see Problem 3.4 of the Solved Problems on Random Processes.)

36.1.1 More Useful Facts about Ergodic Processes

In Section 35.4, we gave a "Useful Fact" that gave you some circumstances under which a WSS Gaussian process will be ergodic. Here are two more useful facts:

Fact(i): If a CT or DT wide-sense stationary process X is ergodic, and if the limit of $R_X(\tau)$ exists as $\tau \to \infty$, then it will automatically be true that

$$\lim_{\tau \to \infty} R_X(\tau) = \mu_X^2.$$

Fact(ii): Let $X(t), \quad -\infty < t < \infty$ be a continuous or discrete time WSS ergodic process and suppose we filter this random signal for all time $-\infty < t < \infty$ with a stable, linear-time, time-invariant filter. Then the filter output process $Y(t), \quad -\infty < t < \infty$ will also be WSS and ergodic. The following block diagram summarizes this result:

Example 36.5. As a consequence of Fact(i) above, we can say the following: Suppose we have a WSS process for which the limit of $R_X(\tau)$ exists as $\tau \to \infty$, but

$$\lim_{\tau \to \infty} R_X(\tau) \neq \mu_X^2.$$

Then we can say that the process must be nonergodic. To illustrate, re-examine the X_n process defined in Example 36.4. In Problem 3.4 of the Solved Problems on Random Processes, it is shown for this process that

$$\begin{aligned} \mu_X &= 7/12 \\ R_X(\tau) &= (17/72)\delta[\tau] + 25/72 \end{aligned}$$

Therefore,

$$\lim_{\tau\to\infty} R_X(\tau) = 25/72,$$

which is not the same as $\mu_X^2 = 49/144$. We conclude that this process must be nonergodic.

Example 36.6. Notice that in Fact(i), we assumed that we had an ergodic process for which the limit of $R_X(\tau)$ exists as $\tau \to \infty$. The purpose of this example is to point out that there do exist some ergodic processes for which this limit does not exist. We present the following example of this type: The discrete-time process X_n, $n = 1, 2, 3, \cdots$ has the three possible realizations indicated as follows, selected equiprobably:

$$\begin{array}{c} \\ \textit{realization } 1 \\ \textit{realization } 2 \\ \textit{realization } 3 \end{array} \begin{array}{c} \begin{array}{cccccccc} n=1 & n=2 & n=3 & n=4 & n=5 & n=6 & n=7 & \cdots \end{array} \\ \left(\begin{array}{cccccccc} 1 & 2 & 3 & 1 & 2 & 3 & 1 & \cdots \\ 2 & 3 & 1 & 2 & 3 & 1 & 2 & \cdots \\ 3 & 1 & 2 & 3 & 1 & 2 & 3 & \cdots \end{array}\right) \end{array}$$

Notice that each realization is periodic with period 3. This process is both WSS and ergodic. The process parameters μ_X and P_X can be obtained by time-averaging along one period of any realization x_n chosen from among the above three:

$$\begin{aligned} \mu_X &= \frac{x_1 + x_2 + x_3}{3} \\ P_X &= R_X(0) = \frac{x_1^2 + x_2^2 + x_3^2}{3} \end{aligned}$$

$R_X(1)$ can also be obtained by time-averaging along this arbitrary realization x_n:

$$R_X(1) = \frac{x_1x_2 + x_2x_3 + x_3x_4}{3}.$$

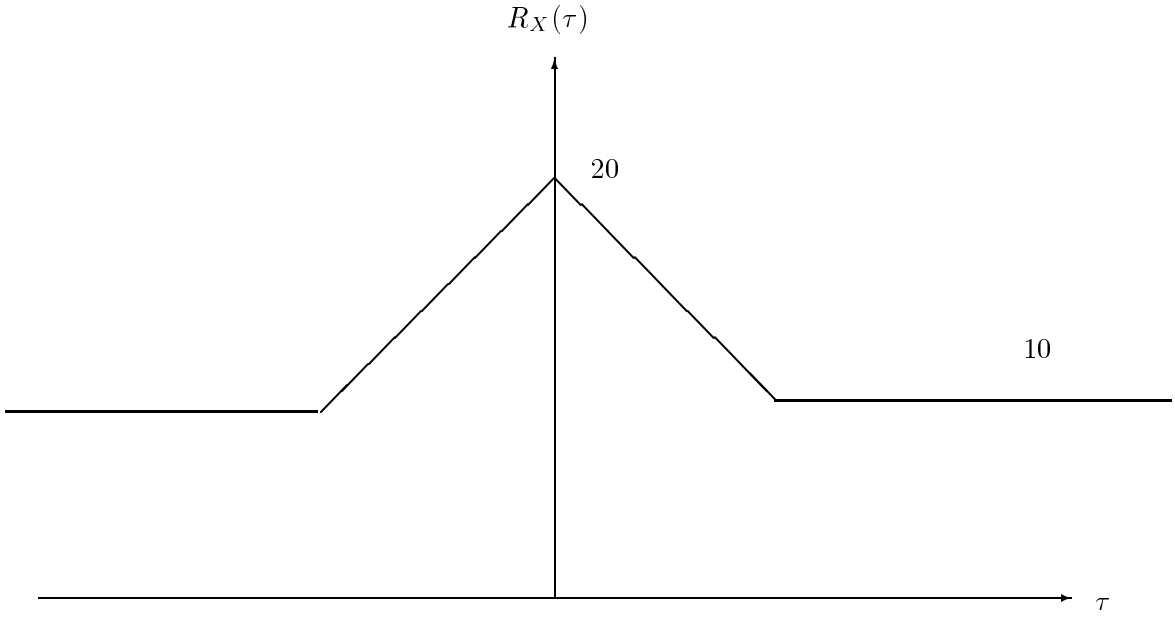
The reader can check that all the above values remain invariant no matter which of the three realizations is chosen. $R_X(\tau)$ can now be computed for any τ because $R_X(\tau)$ will be periodic with period 3 in this case: the values $R_X(-1) = R_X(1), R_X(0), R_X(1)$ constitute autocorrelation values over one period—translating to left and right, you then obtain $R_X(\tau)$ for any integer τ. For more about this example, see Problem 3.3 of the Solved Problems on Random Processes. Notice that $\lim_{\tau\to\infty} R_X(\tau)$ does not exist because for large τ, $R_X(\tau)$ keeps oscillating back and forth over the two values $R_X(1)$ or $R_X(0)$, never converging to one particular value. (The reader who is not yet convinced that this limit does not exist should sketch a plot of $R_X(\tau)$ versus τ.)

Example 36.7. In Section 33.2, we linearly filtered a Gaussian white noise process Z_n to obtain at the filter output a process X_n according to the following block diagram:

$$Z_n \to \boxed{\begin{array}{c}\text{FIR} \\ \text{linear} \\ \text{filter}\end{array}} \to X_n = aZ_n + bZ_{n-1} + cZ_{n-2}$$

The filtering operation is stable and time-invariant (no matter what the three filter tap weights a, b, c are), and the GWN process Z_n is WSS and ergodic. Therefore, by Fact(ii) of this section, the process X_n is WSS and ergodic. This fact allows you to time average along realizations of the process X_n in order to estimate μ_X, P_X, and $R_X(\tau)$.

Example 36.8. A WS ergodic process $X(t)$ has the following autocorrelation function:



Suppose we know that the process mean μ_X is less than 0. Then we can determine μ_X from Property(i). Observe from the plot that

$$\lim_{\tau \to \infty} R_X(\tau) = 10.$$

By Property(i), the value 10 must be μ_X^2. Since we are assuming that $\mu_X < 0$, we conclude that

$$\mu_X = -\sqrt{10}.$$

36.2 Space/Time Averaging Concepts

Recitation 12 covered thoroughly the two ways in which we can do averaging to estimate the parameters of a process: space averaging and time averaging. For completeness of these class notes, I include here the background material presented in the Recitation 12 instructions on space and time averaging.

36.2.1 Computing $\mu_X, P_X, R_X(\tau)$ via Space Averaging

Let $X(t)$ be a WSS continuous time or discrete time process. Suppose you have a large number of realizations of $X(t)$, and let these be

$$x_1(t), x_2(t), x_3(t), \ldots, x_N(t) \tag{36.3}$$

Fix any time t^*. Then, we can estimate μ_X, P_X and $R_X(\tau)$ as follows:

$$\begin{aligned} \mu_X &\approx N^{-1} \sum_{i=1}^{N} x_i(t^*) \\ P_X &\approx N^{-1} \sum_{i=1}^{N} x_i(t^*)^2 \\ R_X(\tau) &\approx N^{-1} \sum_{i=1}^{N} x_i(t^*) x_i(t^*+\tau) \end{aligned}$$

This type of averaging is called SPACE AVERAGING. You can visualize space averaging in this way. Plot the N realizations that you get by performing the experiment N times. At time $t = t^*$ (an arbitrary time held fixed by you), erect a vertical slice, which cuts through the N realizations and yields one signal sample for each realization. Average these N samples to obtain an approximation to μ_X. Average the squares of these samples to obtain an approximation to P_X.

36.2.2 Computing $\mu_X, P_X, R_X(\tau)$ via Time Averaging

The time average of a CT signal $\phi(t)$ is written $\langle \phi(t) \rangle$ and is defined by

$$\langle \phi(t) \rangle \triangleq \lim_{T \to \infty} T^{-1} \int_0^T \phi(t) dt$$

The time average $\langle \phi(t) \rangle$ is sometimes called the *DC value* of the signal $\phi(t)$. The average power generated by the signal $\phi(t)$ shall be written $\langle \phi(t)^2 \rangle$ and is defined by

$$\left\langle \phi(t)^2 \right\rangle \triangleq \lim_{T \to \infty} T^{-1} \int_0^T \phi(t)^2 dt$$

Example 36.9. If the signal $\phi(t)$ is periodic with period T, then

$$\begin{aligned} \langle \phi(t) \rangle &= T^{-1} \int_0^T \phi(t) dt \\ \left\langle \phi(t)^2 \right\rangle &= T^{-1} \int_0^T \phi(t)^2 dt \end{aligned}$$

For a DT signal $\phi[n]$, the DC value and the power would be computed by

$$\begin{aligned} \langle \phi[n] \rangle &\triangleq \lim_{N\to\infty} N^{-1} \sum_{n=1}^{N} x[n] \\ \left\langle \phi[n]^2 \right\rangle &\triangleq \lim_{N\to\infty} N^{-1} \sum_{n=1}^{N} x[n]^2 \end{aligned}$$

Let $X(t)$ be a CT or DT WSS process and suppose you have a large number of realizations (36.3). Then, $\mu_X, P_X, R_X(\tau)$ can be approximated as

$$\begin{aligned} \mu_X &\approx N^{-1} \sum_{i=1}^{N} \langle x_i(t) \rangle \\ P_X &\approx N^{-1} \sum_{i=1}^{N} \left\langle x_i(t)^2 \right\rangle \\ R_X(\tau) &\approx N^{-1} \sum_{i=1}^{N} \langle x_i(t) x_i(t+\tau) \rangle \end{aligned}$$

This type of averaging is called TIME AVERAGING. Here is what the preceding three relations say: given a large number of realizations, (1) you can compute the DC value for each realization as a time average, and then average these figures over the different realizations to get an approximation to μ_X; (2) you can compute power for each realization as a time average, and then average these power figures over the different realizations to get an approximation to P_X; and (3), to estimate $R_X(\tau)$ for a fixed τ, you can do a time average

$$\langle x(t)x(t+\tau) \rangle = \lim_{T\to\infty} \frac{1}{T} \int_0^T x(t)x(t+\tau)dt \tag{36.4}$$

for each realization $x(t)$ and then average the figures (36.4) over the realizations. (In the case of an ergodic process, just one of these realizations is needed, as discussed previously.)

36.3 Why Formula $P_X = R_X(0)$ Makes Sense

In this section, we will attempt to explain better why the formula

$$P_X = R_X(0)$$

for computing the power generated by a WSS process $X(t)$ makes sense. In terms of the physical nature of what power means, this is the reasonable way to define the power figure P_X:

$$P_X \triangleq E[< X(t)^2 >]. \tag{36.5}$$

Here is how we interpret the equation (32.11): In the right side of (32.11), let $X(t)$ represent a randomly selected realization signal of the process. We first perform a time average along this signal, obtaining the time average denoted by $< X(t)^2 >$. The quantity $< X(t)^2 >$ physically denotes the power generated by the randomly selected power signal $X(t)$. In general (in case we have a nonergodic process), the expression $< X(t)^2 >$ will be random, i.e., it will vary from realization to realization. Thus, we may think of $< X(t)^2 >$ as defining a random variable. When we have a random variable, we can talk about its expected value. The expected value $E[< X(t)^2 >]$ of the RV $< X(t)^2 >$ is then taken as the power P_X generated by the random process $X(t)$. $E[< X(t)^2 >]$ represents the average of the powers generated by different realizations, as we consider more and more realizations. If we select the different realizations in independent trials of our experiment, the average power across these realizations will converge (by the law of large numbers) to the number $E[< X(t)^2 >]$ as the number of realizations that we select goes to infinity. We hope this discussion has convinced the reader that the definition (36.5) is the proper and sensible way to define the power figrue P_X.

Modulo our discussion just given, why then is it reasonable to say that P_X should turn out to be the same thing as $R_X(0)$? Notice that in equation (36.5) there are two different averaging operations: the time average coming first, denoted by the bracket symbols $< >$, and then the averaging operation across realizations denoted by the expectation operator E. Let us suppose that we can interchange these two averaging operations; that is, let us declare that

$$E[< X(t)^2 >] =< E[X(t)^2] > . \tag{36.6}$$

In the right side of (36.6), the quantity $E[X(t)^2]$, measured at each time t, is easy to compute: by the fact that the process is WSS, $E[X(t)^2]$ does not depend on time t and takes the value $R_X(0)$ for all t. Therefore, the right side of (36.6) may be re-expressed as the time-average

$$< R_X(0) > . \tag{36.7}$$

We may interpret the quantity in (36.7) as the time average of a DC signal which takes the value $R_X(0)$ for all time t. Obviously, the time average of a DC signal is the DC value of the signal (the constant value that the signal takes on). Therefore,

$$< R_X(0) >= R_X(0)$$

and we are led to the conclusion that P_X must be the same thing as $R_X(0)$.

One can re-do the general argument just given, valid for any WSS process $X(t)$, in the case of any particular random process, in order to see that the argument makes sense for that particular process. This we do in the following example.

Example 36.10. Consider again the random sinusoid

$$X(t) = A\cos(\omega_0 t + \Theta).$$

Let us see what power figure P_X is given to us by the formula (36.5). We have

$$P_X = E[< X(t) >^2] = E[A^2/2] = E[A^2]/2.$$

Earlier, we derived the autocorrelation function to be

$$R_X(\tau) = (E[A^2]/2)\cos(\omega_0\tau).$$

Plugging in $\tau = 0$, we see that

$$R_X(0) = E[A^2]/2.$$

We see that P_X and $R_X(0)$ do indeed coincide in this case.

36.4 Gaussian Processes

Gaussian processes are possibly the most useful class of random processes. A continuous or discrete time process $(X(t))$ is said to be a Gaussian process if for each finite set of sampling times

$$t_1 < t_2 < \ldots < \ldots < t_n,$$

the vector of process samples $(X(t_1), X(t_2), \cdots, X(t_n))$ has an n-dimensional joint Gaussian distribution. This means that the joint density of $(X(t_1), X(t_2), \cdots, X(t_n))$ takes the form

$$f(x_1, x_2, \ldots, x_n) = C \exp\left[-(1/2)\sum_{i,j=1}^{n} A_{i,j}(x_i - \mu_i)(x_j - \mu_j)\right] \tag{36.8}$$

where C is a constant, μ_i is the mean of $X(t_i)$ $(i = 1, 2, \cdots, n)$ and $[A_{i,j}]$ is a symmetric $n \times n$ matrix whose inverse is the covariance matrix $[Cov(X(t_i), X(t_j)]$. In particular, every component RV $X(t)$ of a Gaussian process $(X(t))$ is Gaussian, and for any two times $t_1 < t_2$, $(X(t_1), X(t_2))$ has the bivariate Gaussian distribution described in Chapter 4 of your textbook.

The following are two important properties of Gaussian processes:

Property 1: Suppose that the mean function $\mu_X(t)$ of a Gaussian process $(X(t))$ has been specified and that every correlation $E[X(t_1)X(t_2)]$ has been specified. Then, all of the joint densities (36.8) are uniquely determined from these. In particular, for a WSS Gaussian process, the densities (36.8) are all uniquely derivable from the process mean μ_X and the autocorrelation function $R_X(\tau)$.

Property 2: If a Gaussian process is passed through a linear filter, then the output process is also a Gaussian process.

Example 36.11. Let $X(t)$ be a WSS Gaussian process for which

$$\begin{aligned} \mu_X &= -1 \\ R_X(\tau) &= 1 + 2^{-|\tau|} \end{aligned}$$

(a) Find the density $f(x)$ of $X(1)$.

(b) Find the joint density $f(x_1, x_2)$ of $(X(1), X(2))$

(c) Find the covariance matrix of $(X(1), X(2), X(3))$ that one inverts to compute the exponent of the multivariate density of $(X(1), X(2), X(3))$ as indicated in (36.8).

Solution to (a). The density of $X(1)$, being a Gaussian density, must take the form

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-(1/2)\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

where

$$\begin{aligned} \mu &= \mu_X = -1 \\ \sigma^2 &= \sigma_X^2 = R_X(0) - \mu_X^2 = 1 \end{aligned}$$

Solution to (b). The covariance between $X(1)$ and $X(2)$ is given by

$$Cov(X(1), X(2)) = R_X(2-1) - \mu_X^2 = 1.5 - 1 = 0.5$$

The covariance matrix of $(X(1), X(2))$ is therefore

$$\begin{bmatrix} Var(X(1)) & Cov(X(1), X(2)) \\ Cov(X(1), X(2)) & Var(X(2)) \end{bmatrix} = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$$

Inverting this matrix using Matlab, one obtains the matrix

$$\begin{bmatrix} 4/3 & -2/3 \\ -2/3 & 4/3 \end{bmatrix}$$

This inverse matrix determines the quadratic form

$$Q(x_1, x_2) = (4/3)(x_1+1)^2 - (4/3)(x_1+1)(x_2+1) + (4/3)(x_2+1)^2$$

(the variables are centered at the means and the two means are $-1, -1$). The joint density $f(x_1, x_2)$ takes the form

$$f(x_1, x_2) = C \exp(-0.5 Q(x_1, x_2))$$

Solution to (c). We have

$$Cov(X(i), X(j)) = E[X(i)X(j)] - \mu_X^2 = R_X(|i-j|) - 1 = 2^{-|i-j|}.$$

The desired covariance matrix is therefore

$$\begin{bmatrix} 1 & 1/2 & 1/4 \\ 1/2 & 1 & 1/2 \\ 1/4 & 1/2 & 1 \end{bmatrix}.$$

Example 36.12. Let $(Z_n : n = 1, 2, 3, \cdots)$ be the discrete-time process consisting of independent Gaussian(0,1) RV's (this process is called *discrete-time Gaussian white noise*). Let $(X_n : n = 1, 2, 3, \cdots)$ be the process obtained by recursive filtering of the (Z_n) process as follows:

$$\begin{aligned} X_1 &= Z_1 \\ X_n &= Z_n + aX_{n-1}, \quad n = 2, 3, 4, \cdots \end{aligned}$$

where a is a parameter. By Property 2, the process (X_n) is Gaussian. Let us find the joint density of $(X_1, X_2, \cdots, X_n)$. The joint density of $(Z_1, Z_2, \cdots, Z_n)$ is the product of the marginal densities of the Z_i's (by independence of the Z_i's), and is therefore given by

$$(1/\sqrt{2\pi})^n \exp[-0.5\{z_1^2 + z_2^2 + \cdots + z_n^2\}]$$

Making the substitutions

$$\begin{aligned} z_1 &= x_1 \\ z_2 &= x_2 - ax_1 \\ z_3 &= x_3 - ax_2 \\ &\cdots \\ z_n &= x_n - ax_{n-1} \end{aligned}$$

the following joint density for $(X_1, X_2, \cdots, X_n)$ is obtained:

$$f(x_1, x_2, \cdots, x_n) = (1/\sqrt{2\pi})^n \exp[-0.5\{x_1^2 + (x_2 - ax_1)^2 + (x_3 - ax_2)^2 + \cdots + (x_n - ax_{n-1})^2\}]$$

The process (X_n) is called a *Gauss-Markov process*. Gauss-Markov processes have many applications.

36.5 Power Spectral Density Definition (Cont Time)

Let $S_X(f)$ denote the power spectral density (PSD) of a continuous time WSS process $X(t)$. The PSD is the CT Fourier transform of the autocorrelation function, which gives us the definition

$$S_X(f) \triangleq \int_{-\infty}^{\infty} R_X(\tau) e^{-j(2\pi f)\tau} d\tau$$

Taking the inverse Fourier transform, we have

$$R_X(\tau) = \int_{-\infty}^{\infty} S_X(f) e^{j(2\pi f)\tau} df.$$

Plugging $\tau = 0$ into the preceding equation, we reach the conclusion that

$$P_X = R_X(0) = \int_{-\infty}^{\infty} S_X(f) df. \tag{36.9}$$

So, if we integrate the PSD over all frequencies, we can obtain the total power generated by the process. Later, we shall see that if we integrate the PSD in a narrow frequency band about some fixed frequency, we will obtain the approximate contribution to the total power due to sinusoidal components of the realizations with frequencies near the fixed frequency. Therefore, looking at the values of $S_X(\omega)$ for different frequencies ω tells us which sinusoidal components contribute most to the power (just see for what frequencies f the function $S_X(f)$ takes its biggest values).

We will give properties of PSD's and numerous examples of PSD's during Lecture 37. We conclude the present lecture with this first example on PSD's.

Example 36.13. For CT white noise $X(t)$, we have

$$R_X(\tau) = A\delta(\tau),$$

for some positive constant A. Taking the Fourier transform, we see that

$$S_X(f) = A, \quad -\infty < f < \infty.$$

In other words, white noise is characterized by a *flat power spectrum.* All sinusoidal components of white noise have an equal influence on the overall power. Doing the integral (36.9), we see that CT white noise has infinite power, and therefore cannot be a physically realizable process. However, one can easily obtain a process that is physically realizable and behaves approximately like white noise: just take a process $X(t)$ whose power spectrum is of the form

$$S_X(f) = \begin{cases} A, & -B \le f \le B \\ 0, & elsewhere \end{cases}$$

where B is very large. The white noise model is very useful in practice. Two such uses are: (1) white noise is typically used to model the ambient noise in electronic devices (the backgound noise); and (2) Gaussian white noise is typically used to model the additive noise occuring in the transmission of a signal through a communication system.

Lecture 37

Random Processes Part 7

37.1 Processes Through Systems

When we send a WSS random signal through a system, what can we say about the autocorrelation structure of the output signal? In this section, we consider three different types of systems for which we can answer this question.

37.1.1 Additive System

We are given two WSS processes $X(t)$ and $Y(t)$ which are independent, which means that any RV U computed from the X process realization is independent of any RV computed from the Y process realization. In particular, independence implies that for all times s, t:

$$E[X(s)Y(t)] = E[X(s)]E[Y(t)].$$

We consider the *additive system* whose output is

$$Z(t) = X(t) + Y(t).$$

The output process $Z(t)$ will be WSS and its process mean μ_Z and autocorrelation function $R_Z(\tau)$ are computable via the formulas:

$$\mu_Z = \mu_X + \mu_Y \tag{37.1}$$

$$R_Z(\tau) = R_X(\tau) + R_Y(\tau) + 2\mu_X\mu_Y \tag{37.2}$$

The proof of formula (37.1) is left to the reader. We give the proof of (37.2) below.

Proof of (37.2). We have

$$R_Z(\tau) = E[Z(t)Z(t+\tau)]$$

$$\begin{aligned} &= E[(X(t)+Y(t))(X(t+\tau)+Y(t+\tau))] \\ &= E[X(t)X(t+\tau)] + E[Y(t)Y(t+\tau)] + E[Y(t)X(t+\tau)] + E[X(t)Y(t+\tau)] \\ &= R_X(\tau) + R_Y(\tau) + E[Y(t)]E[X(t+\tau)] + E[X(t)]E[Y(t+\tau)] \\ &= R_X(\tau) + R_Y(\tau) + 2\mu_X\mu_Y \end{aligned}$$

Remark. The Gaussian additive noise channel, one of the most common communication system models, can be viewed as an additive system. The input $X(t)$ to this channel is the information-bearing signal you are trying to communicate, and the process $Y(t)$ is the channel noise which is modeled as a Gaussian process (usually a GWN process, but it may be correlated noise). The channel output is $Z(t) = X(t) + Y(t)$, the information-bearing signal corrupted by additive noise.

37.1.2 Multiplicative System

As in Section 37.1.1, we start with two independent WSS processes $X(t)$ and $Y(t)$. We now consider the *multiplicative system* whose output is

$$Z(t) = X(t)Y(t).$$

The output process $Z(t)$ will be WSS and its process mean μ_Z and autocorrelation function $R_Z(\tau)$ are computable via the formulas:

$$\mu_Z = \mu_X\mu_Y \tag{37.3}$$

$$R_Z(\tau) = R_X(\tau)R_Y(\tau) \tag{37.4}$$

The proof of formula (37.3) is left to the reader. We give the proof of (37.4) below.

Proof of (37.4). We have

$$\begin{aligned} R_Z(\tau) &= E[Z(t)Z(t+\tau)] \\ &= E[X(t)X(t+\tau)Y(t)Y(t+\tau)] \\ &= E[X(t)X(t+\tau)]E[Y(t)Y(t+\tau)] \\ &= R_X(\tau)R_Y(\tau) \end{aligned}$$

Remark. One common instance in which a multiplicative system is used is *amplitude modulation*. In this case, $X(t)$ is the information-bearing signal you are trying to communicate, and $Y(t)$ is a random sinusoid (the "carrier wave"). The output of the amplitude modulator is then $X(t)Y(t)$.

37.1.3 Linear Filtering

Suppose we have a stable linear time-variant filter with impulse response function $h(t)$. We apply as input to the filter a WSS random signal $X(t)$:

$$X(t) \rightarrow \boxed{h(t)} \rightarrow Y(t)$$

As indicated in the block diagram, the filter output is random signal $Y(t)$. The process $Y(t)$ will be WSS (to ensure this, run the filter for all time $-\infty < t < \infty$ to make sure that no transient component is present in $Y(t)$). We need to know how to compute μ_Y, $R_Y(\tau)$, and $S_Y(f)$ from μ_X, $R_X(\tau)$, and $S_X(f)$. Here are the formulas for doing this:

$$\mu_Y = \mu_X \left[\int_{-\infty}^{\infty} h(t)dt\right] \tag{37.5}$$

$$R_Y(\tau) = R_X(\tau) * h(\tau) * h(-\tau) \tag{37.6}$$

$$S_Y(f) = S_X(f)|H(f)|^2 \tag{37.7}$$

Equation (37.5) tells you how to compute the output mean from the input mean. Equation (37.6) tells you how to compute the output autocorrelation function from the input autocorrelation function. (Unfortunately, this is a triple convolution.) Equation (37.7) tells you how to compute the output power spectral density from the input power spectral density; in (37.7), $H(f)$ represents the frequency response function of the filter, which is the Fourier transform of $h(t)$:

$$H(f) = \int_{-\infty}^{\infty} h(t)\exp(-jt2\pi f)dt.$$

We point out that you can obtain (37.7) by taking the Fourier transform of both sides of (37.6). To see this, let $\mathcal{F}$ denote the Fourier transform operator and perform the following steps:

$$\begin{aligned} S_Y(f) &= \mathcal{F}[R_Y(\tau)] \\ &= \mathcal{F}[R_X(\tau) * h(\tau) * h(-\tau)] \\ &= \mathcal{F}[R_X(\tau)]\mathcal{F}[h(\tau)]\mathcal{F}[h(-\tau)] \\ &= S_X(f)H(f)H(-f) \\ &= S_X(f)H(f)H(f)^* \\ &= S_X(f)|H(f)|^2 \end{aligned}$$

(In the preceding $H(f)^*$ denotes the complex conjugate of $H(f)$.)

You can find formulas (37.5)-(37.7) in Chapter 11 of your textbook. They also hold for discrete-time filtering of discrete-time random signals, with the obvious modifications. (For example, in formula (37.5), you'd do a summation instead of an integral. In formula (37.6), you'd be doing discrete-time convolution.) In future lectures, I will attempt to give some reasoning behind why these formulas are true. Until that time, however, we can still feel free to use these formulas in our work.

37.2 Higher Moments of Gaussian RV's

Let X be any RV. Recall that its moment generating function $M_X(s)$ is defined by:

$$M_X(s) \triangleq E[e^{sX}].$$

I want to extend what we know about $M_X(s)$ a little further. The McClaurin Series expansion of e^t is:

$$e^t = 1 + t + (t^2/2) + (t^3/3!) + (t^4/4!) + (t^5/5!) + (t^6/6!) + \cdots. \tag{37.8}$$

Substituting $t = sX$, this becomes

$$e^{sX} = 1 + Xs + X^2(s^2/2) + X^3(s^3/3!) + X^4(s^4/4!) + X^5(s^5/5!) + X^6(s^6/6!) + \cdots.$$

Now take the expected value of both sides, doing the expected value of the right side term by term:

$$M_X(s) = 1+E[X]s+E[X^2](s^2/2)+E[X^3](s^3/3!)+E[X^4](s^4/4!)+E[X^5](s^5/5!)+E[X^6](s^6/6!)+\cdots. \tag{37.9}$$

Now let X be a Gaussian(0,1) RV. We know that the first two moments of X are:

$$E[X] = 0, \quad E[X^2] = 1.$$

We can use the expansion (37.9) to learn about the higher moments of X, that is, the moments of the form $E[X^k]$, for k an integer power bigger than two. Let's use the following formula from Chapter 6 of your textbook for the MGF of the standard Gaussian RV X:

$$M_X(s) = e^{s^2/2}.$$

Go to equation (37.8) and substitute in $t = s^2/2$. You get the expansion:

$$e^{s^2/2} = 1 + (s^2/2) + (s^4/8) + (s^6/48) + \cdots. \tag{37.10}$$

If we compare each power on the right side of (37.10) with the corresponding power on the right side of (37.9), we can easily obtain any moment of X that we wish.

Example 37.1. Let's compute $E[X^4]$ and $E[X^6]$ for a Gaussian(0,1) RV X. Picking off the coefficient of s^4 in (37.9) and then in (37.10), we obtain the equation

$$E[X^4]/4! = 1/8.$$

Solving for $E[X^4]$, we obtain

$$E[X^4] = 3.$$

Picking off the coefficient of s^6 in (37.9) and then in (37.10), we obtain the equation

$$E[x^6]/6! = 1/48,$$

from which one sees that

$$E[X^6] = 15.$$

The long way of doing this computation would be

$$E[X^6] = \int_{-\infty}^{\infty} x^6 f_X(x)dx = \int_{-\infty}^{\infty} x^6 \left(\frac{1}{\sqrt{2\pi}}\right) \exp(-x^2/2)dx,$$

an integral which some people might attempt to evaluate via repeated integration by parts. Impress your friends outside of EE 3025 by showing them how fast you can do the integral

$$\int_{-\infty}^{\infty} x^8 \exp(-x^2/2)dx.$$

37.3 Brownian Motion Process (Wiener Process)

The Brownian motion process $(W(t),\ \ t \geq 0)$ is a special type of Gaussian process that is important in many science and engineering contexts. It was discovered by Louis Bachelier in his 1900 thesis on financial theory. It was exploited by Norbert Wiener in his work on Brownian motion and other topics. (The Brownian motion process $W(t)$ is also called the Wiener process or the Bachelier-Wiener process. The W in $W(t)$ stands for Wiener.)

We describe a simple way of obtaining the Brownian motion process $W(t)$. Let $X(t)$ be a Gaussian white noise process with autocorrelation function

$$R_X(\tau) = A\delta(\tau).$$

We pass the process $X(t)$ through an integrator, as illustrated in the following block diagram:

$$X(t) \rightarrow \boxed{\int_0^t} \rightarrow W(t)$$

The integrator output process $W(t)$, defined for $t \geq 0$, is the Brownian motion process. Thus, we may take the following equation as the definition of the Brownian motion process:

$$W(t) \triangleq \int_0^t X(\alpha)d\alpha, \ \ t \geq 0.$$

In our earlier section on Gaussian processes, we learned that if we linearly filter a Gaussian process, then we obtain another Gaussian process. This is why the Brownian motion process $W(t)$ is a Gaussian process.

Component RV's of Brownian Motion Process. Fix a time $T > 0$. We find the probability density of the random variable $W(T)$. The component RV's of a Gaussian process are Gaussian

RV's. Therefore, the density of $W(T)$ is of Gaussian form, completely determined by the mean μ of $W(T)$ and the variance σ^2 of $W(T)$. Computation of μ is easy:

$$\begin{aligned}\mu &= E\left[\int_0^T X(\alpha)d\alpha\right] \\ &= \int_0^T E[X(\alpha)]d\alpha = \int_0^T 0\, d\alpha = 0.\end{aligned}$$

Since the mean of $W(T)$ is zero, the variance σ^2 coincides with the second moment $E[W(T)^2]$. Now $W(T)$ is an integral. We express $W(T)^2$, an integral times itself, as a double integral (this is a standard trick):

$$\begin{aligned}W(T)^2 &= \left[\int_0^T X(t)dt\right]\left[\int_0^T X(s)ds\right] \\ &= \int_0^T \int_0^T X(s)X(t)dsdt.\end{aligned}$$

Next, we take the expected value of both sides, at the same time moving the expected value inside the double integral (another standard trick):

$$\begin{aligned}\sigma^2 &= \int_0^T \int_0^T E[X(s)X(t)]dsdt \\ &= \int_0^T \int_0^T R_X(s-t)dsdt \\ &= \int_0^T \left[\int_0^T A\delta(s-t)ds\right]dt \\ &= \int_0^T A[u(T-t)-u(-t)]dt \\ &= AT.\end{aligned}$$

The last step followed from the fact that $u(T-t)-u(-t)$ is a rectangular pulse of amplitude one over the time interval $[0,T]$, the area under which is T. We can now express the PDF of $W(T)$ as follows:

$$f_{W(T)}(w) = \frac{1}{\sqrt{2\pi AT}}\exp\left[-\frac{w^2}{2AT}\right], \quad -\infty < w < \infty.$$

One can now compute any probability of the form $P[a \le W(T) \le b]$.

Facts About Brownian Motion Process

- Every Brownian motion realization starts at the origin (i.e., $W(0) = 0$) and is a continuous nowhere-differentiable function of time.[1]

[1] This is a deep result which is not easy to prove.

- If $t_1 < t_2$ are any two times, then the "increment random variable" $W(t_2) - W(t_1)$ is Gaussian with mean 0 and variance $A(t_2 - t_1)$. (As a special case, taking $t_1 = 0$ and $t_2 = t$, we see that $W(t)$ is Gaussian with mean 0 and variance At, as proved previously.)
- The Brownian motion process obeys the "independent increments property": If $t_1 < t_2 \leq t_3 < t_4$ are four times, then the increments $W(t_2) - W(t_1)$ and $W(t_4) - W(t_3)$ are independent RV's. (Recall that the Poisson process also satisfies the independent increments property.)
- The autocorrelation and autocovariance structure of the Brownian motion process is as follows:

$$R_W(t_1, t_2) = A \min(t_1, t_2). \tag{37.11}$$
$$Cov[W(t_1), W(t_2)] = A \min(t_1, t_2). \tag{37.12}$$

Exercise. Prove (37.11) using the "double integral trick" in similar fashion to our earlier derivation of $Var(W(T))$.

Example 37.2. For the Brownian motion process $W(t)$, let us compute

(a) $E[W(5)|W(3) = 2]$

(b) $E[W(6)W(9)]$

Computation (a).

$$\begin{aligned} E[W(5)|W(3) = 2] &= E[W(5) - W(3)|W(3) = 2] + E[W(3)|W(3) = 2] \\ &= E[W(5) - W(3)] + 2 \\ &= 0 + 2 = 2 \end{aligned}$$

We used the fact that $W(5) - W(3)$ and $W(3) = W(3) - W(0)$ are independent increments.

Computation (b).

$$\begin{aligned} E[W(6)W(9)] &= E[W(6)(W(9) - W(6))] + E[W(6)W(6)] \\ &= E[W(6)]E[W(9) - W(6)] + \mathrm{V}ar[W(6)] \\ &= 0 * 0 + 6A = 6A \end{aligned}$$

Remarks:

- The DT random walk process (X_n) has autocorrelation structure

$$R_X(n_1, n_2) = \min(n_1, n_2).$$

Comparing this to (37.11), we see that the Brownian motion process and random walk have similar autocorrelation structure. Brownian motion is often thought of as a continuous-time analogue of random walk.

- A Poisson process $X(t)$ with arrival rate λ satisfies

$$Cov(X(t_1), X(t_2)) = \lambda \min(t_1, t_2).$$

(Show this.) The Brownian motion process $W(t)$ has the similar autocovariance structure (37.12). But the realizations of the Poisson process are quite different from the realizations of the Brownian motion process (step functions versus continuous functions). Thus, the autocovariance structure of a random process does not determine the shape of its realizations.

- The 1-D Brownian motion process (the one we described here) can be used as a tool in expressing the solution to certain stochastic differential equations.

- Sometimes, people design a continuous-time filter to convert Gaussian white noise into a process with a desired autocorrelation structure. Unfortunately, Gaussian white noise is not physically realizable. However, the Brownian motion process is physically realizable. One could instead apply the designed filter to Brownian motion and then differentiate the output signal to arrive at the same result.

- The 2-D Brownian motion process (not covered here) is useful in EE because there is a method based on it for solving Laplace's partial differential equation

$$\partial^2 u/\partial^2 x + \partial^2 u/\partial^2 y = 0.$$

- The bouncing around of a pollen particle in a fluid, first observed by botanist Robert Brown in 1827 and later explained by Einstein in the early 1900's, is called Brownian motion, and can be thought of as a realization of a 3-D Brownian motion process.

Exercise. Fix any two times t_1, t_2 for which $0 < t_1 < t_2$. The 2-D cross section $(W(t_1), W(t_2))$ has a bivariate Gaussian density $f(w_1, w_2)$. Determine $f(w_1, w_2)$.

37.4 Properties of CT Power Spectral Density

Property 1: $S_X(f) \geq 0$ for all frequencies f.

Property 2: $S_X(-f) = S_X(f)$ (i.e., $S_X(f)$ is an even function).

Property 3: P_X is the area under the $S_X(f)$ curve.

Remarks. Property 3 has been remarked upon previously. (See equation (36.9).) Property 2 follows immediately from the EE 3015 fact that the Fourier transform of an even function is an even function. Property 1 is a nontrivial fact that I will attempt to prove later on.

Example 37.3. Let $X(t)$ be WSS with

$$R_X(\tau) = C\exp(-a|\tau|).$$

We shall see later that such a process arises from passing white noise through an RC filter. In a good Fourier transform table, one will find the Fourier transform pair

$$\exp(-a|\tau|) \leftrightarrow \frac{2a}{a^2 + (2\pi f)^2}.$$

Therefore

$$S_X(f) = \frac{2aC}{a^2 + (2\pi f)^2}. \tag{37.13}$$

As a test of our EE 3015 skills, let us see if we can derive this another way. (You are recommended to go through this discussion if your EE 3015 skills are rather abysmal.) Let $\phi(\tau)$ be the one-sided decaying exponential function

$$\phi(\tau) = \exp(-a\tau)u(\tau).$$

Then we have

$$R_X(\tau) = C[\phi(\tau) + \phi(-\tau)].$$

The Laplace transform of $\phi(\tau)$ is well known to be

$$\Phi(s) = \frac{1}{s+a}.$$

Applying the reflection property of Laplace transforms, we can say that the Laplace transform of $R_X(\tau)$ is

$$C[\Phi(s) + \Phi(-s)],$$

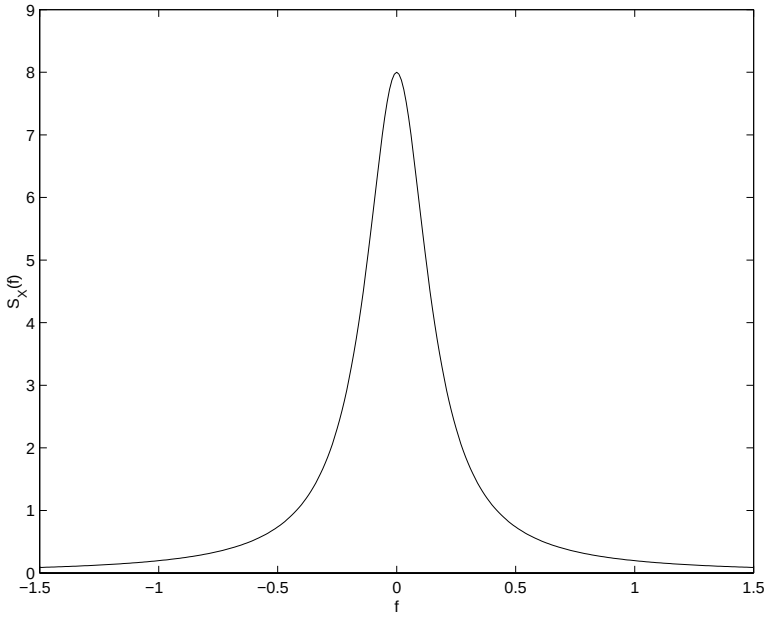
which we can simplify as

$$\frac{C}{s+a} + \frac{C}{-s+a} = \frac{2aC}{a^2 - s^2}.$$

We can go from this expression to the Fourier transform by substituting $s = j(2\pi f)$:

$$\left[\frac{2aC}{a^2 - s^2}\right]_{s=j(2\pi f)} = \frac{2aC}{a^2 + (2\pi f)^2}.$$

The formula (37.13) has therefore been established. Here is the plot of a typical $S_X(f)$ of this type:



Note the nonegativity (Property 1) and the evenness (Property 2). Also, note that $S_X(f)$ is most dominant at low frequencies f. Such a random signal is called a "lowpass random signal". Finally, let us verify that the power P_X generated by our random signal can be computed as

$$P_X = \int_{-\infty}^{\infty} S_X(f)df.$$

We make use of the calculus formula

$$\int \frac{a}{a^2+u^2}du = Tan^{-1}\left(\frac{u}{a}\right).$$

We obtain, making the change of variable $\omega = 2\pi f$,

$$\begin{aligned}
\int_{-\infty}^{\infty} S_X(f)df &= \int_{-\infty}^{\infty} \frac{2aC}{a^2+(2\pi f)^2}df \\
&= \frac{C}{\pi}\int_{-\infty}^{\infty} \frac{a}{a^2+\omega^2}d\omega \\
&= \frac{C}{\pi}\left[Tan^{-1}(\omega/a)\right]_{\omega\to-\infty}^{\omega\to\infty} \\
&= \frac{C}{\pi}[(\pi/2)-(-\pi/2)] = C.
\end{aligned}$$

We conclude that the power generated by this random signal must be

$$P_X = C.$$

This coincides with the result we get doing the power computation via the autocorrelation function:

$$P_X = R_X(0) = C\exp(-a * 0) = C.$$

37.5 Disc Time Power Spectral Density Definition/Properties

Let X_n be a discrete-time WSS random process. Then, the power spectral density is the discrete-time Fourier transform of $R_X(\tau)$:

$$S_X(f) = \sum_{\tau=-\infty}^{\infty} R_X(\tau)e^{-j(2\pi f)\tau}.$$

By taking the inverse discrete-time Fourier transform of the power spectral density, you get back to $R_X(\tau)$:

$$R_X(\tau) = \int_0^1 S_X(f)e^{j(2\pi f)\tau}df.$$

In particular, plugging in $\tau = 0$, we see that power is computed by

$$P_X = \int_0^1 S_X(f)df. \tag{37.14}$$

Here are some properties of the discrete-time power spectral density:

Property 1: $S_X(f) \geq 0$ for all frequencies f.

Property 2: $S_X(-f) = S_X(f)$ (i.e., $S_X(f)$ is an even function).

Property 3: $S_X(f)$ is periodic with period 1.

Property 4: P_X can be computed by integrating $S_X(f)$ over any one period. (In (37.14), we integrated from $f = 0$ to $f = 1$. We can equally well integrate from $f = -1/2$ to $f = 1/2$.)

Remarks. Properties 1 and 2 above should not be surprising, because they coincide with Properties 1 and 2 of the CT power spectrum. Property 3 is an additional property that the CT power spectrum did not have; Property 3 must be true because any discrete-time Fourier transform is periodic with period 1.

Example 37.4. Let X_n be discrete-time white noise. This means the autocorrelation function is of the form

$$R_X(\tau) = A\delta[\tau],$$

where A is the process variance (which coincides with P_X, since the process mean μ_X is zero for white noise). The DT Fourier transform of $\delta[\tau]$ is 1. Therefore,

$$S_X(f) = A, \quad -\infty < f < \infty.$$

Amusingly, this is the same expression for the power spectral density of continuous time white noise that we obtained in Example 36.13. However, keep in mind that the power characteristics of CT white noise and DT white noise are completely different. Our DT white noise process in this example has finite power, according to the calculation

$$P_X = \int_0^1 S_X(f)df = \int_0^1 Adf = A, \tag{37.15}$$

because we are integrating the constant power spectrum A over just a *finite range of frequencies*. For CT white noise, we'd be integrating this same power spectrum A over *all frequencies* $-\infty < f < \infty$, thereby giving us an infinite power calculation for CT white noise. Finally, note that the power figure A we got for our DT white noise process X_n via the "frequency domain" calculation (37.15) coincides with what we get by doing the power calculation in "time domain":

$$P_X = R_X(0) = A\delta[0] = A * 1 = A.$$

Lecture 38

Random Processes Part 8

38.1 Power Spectra of Periodic Processes

The random sinusoid is one example of a type of process called periodic process; periodic processes are common in many communication systems applications. For our purposes here, we will define a process to be a periodic process if all of its realizations are periodic signals with the same period. In this section, we want to examine the form of the power spectral densities of WSS periodic processes. First, some examples.

Example 38.1. Suppose

$$X(t) = A\cos(2\pi f_0 t + \Theta)$$

is the random sinusoid with A and Θ independent RV's and Θ uniformly distributed between 0 and 2π. All of the realizations are periodic with period $1/f_0$. Therefore, the random sinusoid is a periodic process. In addition, we already know that this process is WSS and that its autocorrelation function is

$$R_X(\tau) = (E[A^2]/2)\cos(2\pi f_0 \tau).$$

Taking the Fourier transform, we see that the power spectral density is

$$S_X(f) = (E[A^2]/4)\delta(f + f_0) + (E[A^2]/4)\delta(f - f_0).$$

This is a *discrete spectrum*. We shall ultimately see that every WSS periodic process has discrete spectrum.

Example 38.2. Let $\phi(t)$ be a deterministic periodic signal with period T_0. Let U be a random variable uniformly disributed in the interval $[0, T_0]$. Define the process

$$X(t) \triangleq \phi(t + U), \quad -\infty < t < \infty.$$

Notice that every realization of this process is period with period T_0. Therefore, $X(t)$ is a periodic process. We now show that this process is WSS. First, we compute the mean function:

$$E[X(t)] = E[\phi(t+U)] = \frac{1}{T_0}\int_0^{T_0} \phi(t+u)du.$$

Making the change of variable $v = t + u$, this becomes

$$E[X(t)] = \frac{1}{T_0}\int_t^{t+T_0} \phi(v)dv.$$

Since ϕ is periodic, its integral will be the same no matter what interval we integrate over that is of length equal to one period. In other words, integrating ϕ over the interval $[t, t+T_0]$ will yield the same result as integrating over the interval $[0, T_0]$. We conclude that

$$E[X(t)] = \frac{1}{T_0}\int_0^{T_0} \phi(u)du, \quad -\infty < t < \infty.$$

We have shown that the mean function is constant. Now let us compute autocorrelations. We have

$$E[X(t_1)X(t_2)] = E[\phi(t_1+U)\phi(t_2+U)] = \frac{1}{T_0}\int_0^{T_0} \phi(t_1+u)\phi(t_2+u)du.$$

Making the change of variable $v = t_1 + u$, this becomes

$$E[X(t_1)X(t_2)] = \frac{1}{T_0}\int_{t_1}^{t_1+T_0} \phi(v)\phi(v+t_2-t_1)dv.$$

For fixed t_1, t_2, the function $\phi(v)\phi(v+t_2-t_1)$, considered as a function of v, is periodic with period T_0, so we will obtain the same integral for this function whether we integrate over the interval $[t_1, t_1+T_0]$ of length one period, or whether we integrate over the interval $[0, T_0]$ of length one period. We conclude that

$$E[X(t_1)X(t_2)] = \frac{1}{T_0}\int_0^{T_0} \phi(u)\phi(u+t_2-t_1)du,$$

which depends only on the time difference $t_2 - t_1$. This completes the proof that the periodic process $X(t)$ is WSS. For future reference, the process mean μ_X and the autocorrelation function $R_X(\tau)$ of this process are given by

$$\begin{aligned} \mu_X &= \frac{1}{T_0}\int_0^{T_0} \phi(u)du \\ R_X(\tau) &= \frac{1}{T_0}\int_0^{T_0} \phi(u)\phi(u+\tau)du \end{aligned} \tag{38.1}$$

From equation (38.1), it follows that $R_X(\tau)$ is periodic with period T_0. The reader can easily prove this by showing that

$$\int_0^{T_0} \phi(u)\phi(u+\tau)du = \int_0^{T_0} \phi(u)\phi(u+\tau+T_0)du$$

is true as a consequence of the fact that ϕ is periodic with period T_0.

Remark. It is a useful fact that $R_X(\tau)$ is periodic if $X(t)$ is a WSS periodic process, and that the period of $R_X(\tau)$ is the common period possessed by all of the realizations of $X(t)$. We have so far already seen two special cases of this useful result to be true: (1) the random sinusoid, and (2) the class of WSS peridic processes discussed in Example 38.2.

We are now in a position to discuss the power spectrum of a WSS periodic process. Accordingly, let $X(t)$ be a WSS periodic process. Let T_0 be the common period of all the realizations of $X(t)$. As we have pointed out in the preceding remark, the autocorrelation function $R_X(\tau)$ is also periodic with period T_0. It has the following Fourier series expansion:

$$R_X(\tau) = \sum_{k=-\infty}^{\infty} a_k e^{jk2\pi f_0 \tau} \tag{38.2}$$

where f_0 is the fundamental frequency in cycles/second given by

$$f_0 = \frac{1}{T_0}$$

and each a_k is a Fourier coefficient given by

$$a_k = \frac{1}{T_0}\int_0^{T_0} R_X(\tau)e^{-jk2\pi f_0 \tau}d\tau.$$

Taking the Fourier transform of both sides of (38.2), we obtain the following formula for the power spectrum of our WSS periodic random signal $X(t)$:

$$S_X(f) = \sum_{k=-\infty}^{\infty} a_k \delta(f - kf_0), \quad -\infty < f < \infty$$

Conclusion. The fundamental conclusion of this section is that *every WSS periodic process has discrete power spectrum.*

Example 38.3. Let $X(t)$ be a WSS periodic process with

$$R_X(\tau) = 3\cos(2\pi f_0 \tau) + 5\cos(4\pi f_0 \tau). \tag{38.3}$$

The right side of (38.3) is already a linear combination of sinusoidal terms, so we do not need to find the Fourier series expansion of $R_X(\tau)$. (In fact, the right side of (38.3) is already the Fourier series expansion of $R_X(\tau)$.) Taking the Fourier transform of the right side of (38.3), we obtain the following power spectral density:

$$S_X(f) = (1.5)\delta(f - f_0) + (1.5)\delta(f + f_0) + (2.5)\delta(f - 2f_0) + (2.5)\delta(f + 2f_0).$$

Exercise. Let $X(t)$ be a WSS periodic process with

$$R_X(\tau) = 2\cos(2\pi f_0\tau) + \cos(2\pi f_0(\tau - 1)) + \cos(2\pi f_0(\tau + 1)).$$

Show that

$$S_X(f) = [1 + \cos(2\pi f_0)][\delta(f + f_0) + \delta(f - f_0)].$$

38.2 Some Random Signal Filtering Scenarios

In filtering a WSS random signal $X(t)$ via a LTI stable filter, we obtain a WSS random signal $Y(t)$ at the filter output. To finish the present lecture, we illustrate a couple of common filtering scenarios of this type, concentrating on continuous-time scenarios. In our next lecture, we will continue with further scenarios, both in continuous-time and discrete-time.

38.2.1 Ideal Low Pass Filtering

The filter input signal $X(t)$ and filter output signal $Y(t)$ are here taken to be CT WSS random signals. Suppose we take our filter to be an ideal low pass filter. We assume the filter frequency response function $H(f)$ takes the form:

$$H(f) = \begin{cases} 1, & -B \le f \le B \\ 0, & elsewhere \end{cases} \tag{38.4}$$

Using the formula

$$S_Y(f) = |H(f)|^2 S_X(f),$$

we obtain

$$S_Y(f) = \begin{cases} S_X(f), & -B \le f \le B \\ 0, & elsewhere \end{cases}$$

Therefore, we may express the output power as

$$P_Y = \int_{-\infty}^{\infty} S_Y(f)df = \int_{-B}^{B} S_X(f)df. \tag{38.5}$$

Example 38.4. WSS random signal $X(t)$ with $R_X(\tau) = C\exp(-\alpha|\tau|)$ is passed through an ideal low pass filter with frequency response (38.4). Let us find the filter bandwidth B so that the filter output signal $Y(t)$ will have 90% of the signal input power. The signal input power is $R_X(0) = C$. By formula (38.5), the signal output power is

$$\begin{aligned} P_Y &= \int_{-B}^{B} S_X(f)df \\ &= \int_{-B}^{B} \frac{2\alpha C}{(2\pi f)^2 + \alpha^2} df = \frac{2C}{\pi} Tan^{-1}\left(\frac{2\pi B}{\alpha}\right). \end{aligned}$$

Setting $P_Y = (0.9)P_X$, we must solve

$$\frac{2C}{\pi} Tan^{-1}(B/\alpha) = 0.9C$$

for B. We obtain

$$B = \alpha[\tan(0.45\pi)].$$

38.2.2 Delay Line Filtering

Again, the filter input signal $X(t)$ and filter output signal $Y(t)$ are here taken to be CT WSS random signals. "Delay line filtering" describes the scenario in which the filter impulse response function takes the form

$$h(t) = \sum_i c_i \delta(t - t_i),$$

for constants c_i and times t_i. The frequency response would be

$$H(f) = \sum_i c_i \exp(-jt_i 2\pi f).$$

The computation of the filter output power spectrum $S_Y(f)$ requires us to compute $|H(f)|^2$. Here is how we can do this:

$$\begin{aligned} |H(f)|^2 &= H(f)H(-f) \\ &= \sum_i c_i \exp(-jt_i 2\pi f) \sum_k c_k \exp(jt_k 2\pi f) \\ &= \sum_{i,k} c_i c_k \exp(-j[t_i - t_k]2\pi f) \end{aligned}$$

In the preceding double sum over i and k, you can sum over the pairs (i,k) for which $i = k$ first. This part yields

$$\sum_i c_i^2.$$

In the remaining part of the double sum, the best thing to do is to group the terms as follows:

$$\sum_{i<k} c_i c_k \{\exp(-j[t_i - t_k]2\pi f) + \exp(j[t_i - t_k]2\pi f)\} = 2\sum_{i<j} c_i c_k \cos([t_i - t_j]2\pi f).$$

This gives us output power spectrum expressible as

$$S_Y(f) = |H(f)|^2 S_X(f) = \left\{ \sum_i c_i^2 + 2\sum_{i<j} c_i c_k \cos([t_i - t_k]2\pi f) \right\} S_X(f). \tag{38.6}$$

Example 38.5. Let's take our delay line filter impulse response as

$$h(t) = \delta(t) - \delta(t-1).$$

There are two t_i's, namely, $t_1 = 0$ and $t_2 = 1$. The corresponding c_i's are $c_1 = 1$ and $c_2 = -1$. Plugging into (38.6), we see that

$$S_Y(f) = [2 - 2\cos(2\pi f)]S_X(f).$$

This gives us one way to compute the output power:

$$P_Y = \int_{-\infty}^{\infty} [2 - 2\cos(2\pi f)]S_X(f)df.$$

Another way to compute P_Y would be to use the time domain technique via which you compute P_Y as $R_Y(0)$. We have

$$R_Y(\tau) = R_X(\tau) * h(\tau) * h(-\tau).$$

Observe that

$$\begin{aligned} h(\tau) * h(-\tau) &= [\delta(\tau) - \delta(\tau - 1)] * [\delta(\tau) - \delta(\tau + 1)] \\ &= 2\delta(\tau) - \delta(\tau + 1) - \delta(\tau + 1). \end{aligned}$$

Convoluting this with $R_X(\tau)$, we obtain

$$R_Y(\tau) = 2R_X(\tau) - R_X(\tau + 1) - R_X(\tau - 1).$$

Finally,

$$P_Y = R_Y(0) = 2R_X(0) - 2R_X(1).$$

Lecture 39

Random Processes Part 9

39.1 Another Periodic Process Example

Let $X(t)$ be the periodic random signal

$$X(t) = A_0 + A_1 \cos(2\pi f_0 t + \Theta_1) + A_2 \cos(4\pi f_0 t + \Theta_2),$$

where A_0, A_1, Θ_1, A_2, Θ_2 are independent RV's. We assume further that

$$E[A_0^2] = 4, \quad E[A_1^2/2] = 1, \quad E[A_2^2/2] = 2,$$

and that Θ_1 and Θ_2 are each uniformly distributed in the interval $[0, 2\pi]$. Under the given assumptions, $X(t)$ is a WSS process. Let us compute $R_X(\tau)$ and $S_X(f)$. We can write

$$X(t) = X_0(t) + X_1(t) + X_2(t),$$

where $X_0(t)$ is the "random DC signal"

$$X_0(t) = A_0,$$

and $X_1(t)$ and $X_2(t)$ are the random sinusoids

$$\begin{aligned} X_1(t) &= A_1 \cos((2\pi f_0 t + \Theta_1) \\ X_2(t) &= A_2 \cos(4\pi f_0 t + \Theta_2) \end{aligned}$$

It is easy to show that

$$\begin{aligned} \mu_{X_0} &= E[A_0] \\ R_{X_0}(\tau) &= E[A_0^2] = 4. \end{aligned}$$

From our earlier work, we know that

$$\begin{aligned}
\mu_{X_1}(\tau) &= 0 \\
R_{X_1}(\tau) &= (E[A_1^2/2])\cos(2\pi f_0\tau) = \cos(2\pi f_0\tau) \\
\mu_{X_2} &= 0 \\
R_{X_2}(\tau) &= (E[A_2^2/2])\cos(4\pi f_0\tau) = 2\cos(4\pi f_0\tau)
\end{aligned}$$

It is easy to generalize the result of Section 37.1.1 to obtain

$$R_X(\tau) = R_{X_0}(\tau) + R_{X_1}(\tau) + R_{X_2}(\tau) + 2\mu_{X_0}\mu_{X_1} + 2\mu_{X_0}\mu_{X_2} + 2\mu_{X_1}\mu_{X_2}.$$

The last three terms drop out, and we obtain the following expression for $R_X(\tau)$ in simplest form:

$$R_X(\tau) = 4 + \cos(2\pi f_0\tau) + 2\cos(4\pi f_0\tau).$$

Fourier transforming, we see that

$$S_X(f) = 4 + 0.5\delta(f + f_0) + 0.5\delta(f - f_0) + \delta(f + 2f_0) + \delta(f - 2f_0).$$

Let us check whether we get the same value for P_X using both $R_X(\tau)$ and $S_X(f)$ to compute the power. The integral of a delta function is 1. Therefore

$$P_X = \int_{-\infty}^{\infty} S_X(f)df = 4 + 0.5 + 0.5 + 1 + 1 = 7.$$

Also,

$$P_X = R_X(0) = 4 + \cos(0) + 2\cos(0) = 7.$$

39.2 Filtering White Noise

Let $X(t)$ be CT white noise. This means $X(t)$ is WSS and

$$\begin{aligned}
\mu_X &= 0 \\
R_X(\tau) &= A\delta(\tau) \\
S_X(f) &= A
\end{aligned}$$

Suppose we filter $X(t)$ with a stable continuous-time LTI filter with impulse response $h(t)$:

$$X(t) \rightarrow \boxed{h(t)} \rightarrow Y(t)$$

In general, we have the following three formulas concerning the WSS filter output process $Y(t)$ valid for *any* WSS input $X(t)$:

$$\begin{aligned} R_Y(\tau) &= R_X(\tau) * h(\tau) * h(-\tau) \\ S_Y(f) &= S_X(f)|H(f)|^2 \\ P_Y &= \int_{-\infty}^{\infty} S_X(f)|H(f)|^2 df \end{aligned}$$

In the special case of our white noise input $X(t)$, these become

$$R_Y(\tau) = Ah(\tau) * h(-\tau) = A\int_{-\infty}^{\infty} h(t-\tau)h(t)dt \tag{39.1}$$

$$S_Y(f) = A|H(f)|^2 \tag{39.2}$$

$$P_Y = A\int_{-\infty}^{\infty} |H(f)|^2 df \tag{39.3}$$

Plugging $\tau = 0$ into the right side of (39.1) we also obtain the formula

$$P_Y = A\int_{-\infty}^{\infty} h(t)^2 dt, \tag{39.4}$$

which, combined with (39.3), gives us two ways to compute the output power in response to a white noise input.

We obtain similar results when we filter DT white noise. Suppose we have a DT white noise process X_n, which means X_n is WSS and

$$\begin{aligned} \mu_X &= 0 \\ R_X(\tau) &= A\delta[\tau] \\ S_X(f) &= A \end{aligned}$$

Suppose we filter X_n with a stable discrete-time LTI filter with impulse response $h[n]$:

$$X_n \rightarrow \boxed{h[n]} \rightarrow Y_n$$

Then one can establish the following four formulas concerning the WSS filter output process Y_n:

$$R_Y(\tau) = Ah(\tau) * h(-\tau) = A\sum_{n=-\infty}^{\infty} h[n-\tau]h[n] \tag{39.5}$$

$$S_Y(f) = A|H(f)|^2 \tag{39.6}$$

$$P_Y = A\int_0^1 |H(f)|^2 df \tag{39.7}$$

$$P_Y = A\sum_{n=-\infty}^{\infty} h[n]^2 \tag{39.8}$$

Warning. Only use formulas (39.1)-(39.8) when you have a white noise filter input!

Example 39.1. CT white noise $X(t)$ with autocorrelation function $R_X(\tau) = \delta(\tau)$ is passed through the filter with impulse response $h(t) = e^{-3t}u(t)$. Then the power generated by the filter output process $Y(t)$ is given by

$$P_Y = \int_{-\infty}^{\infty} h(t)^2 dt = \int_0^{\infty} e^{-6t} dt = 1/6.$$

Example 39.2. DT white noise X_n with autocorrelation function $R_X(\tau) = \delta[\tau]$ is passed through a so-called first order autoregressive filter. The resulting filter output process Y_n is related to X_n via the recursion

$$Y_n = X_n + (1/3)Y_{n-1}.$$

Let's compute the output power P_Y. By the recursion method from EE 3015, it is easy to see that the impulse response of the filter is

$$h[n] = (1/3)^n u[n].$$

Therefore,

$$P_Y = \sum_{n=-\infty}^{\infty} h[n]^2 = \sum_{n=0}^{\infty} (1/3)^{2n} = \sum_{n=0}^{\infty} (1/9)^n = \frac{1}{1-(1/9)} = 9/8.$$

Exercise. For the scenario in Example 39.2, show that the frequency response of the filter is

$$H(f) = \frac{3}{3 - \exp(-j2\pi f)}.$$

Use this result to show that

$$S_Y(f) = |H(f)|^2 = \frac{9}{10 - 6\cos(2\pi f)}.$$

Then use Matlab function `int` to verify that the result of the following integration is correct:

$$P_Y = \int_0^1 S_Y(f) = \int_0^1 \frac{9}{10 - 6\cos(2\pi f)} df = 9/8.$$

39.2.1 A Spectral Synthesis Application

Suppose we wish to design a filter so that if we filter white noise with the filter, then the filter output random signal will have a certain autocorrelation function (or, equivalently, a certain power spectral density function). This filter design problem is called the *spectral synthesis problem*. In this section, we work out a specific instance of the spectral synthesis problem.

Let $Z(t)$ be white noise with $R_Z(\tau) = \delta(\tau)$. Suppose we filter $Z(t)$, obtaining a WSS process $X(t)$:

$$Z(t) \rightarrow \boxed{h(t)} \rightarrow X(t)$$

We wish to design the filter impulse response $h(t)$ so that the process $X(t)$ will have autocorrelation function

$$R_X(\tau) = \exp(-a|\tau|),$$

where a is a positive real constant. Using a Fourier transform table, we see that the power spectral density of $X(t)$ must be

$$S_X(f) = \frac{2a}{a^2 + (2\pi f)^2}.$$

Since we are filtering a white noise process whose PSD is 1, we must have

$$S_X(f) = |H(f)|^2 = H(f)H(-f).$$

To find the frequency response $H(f)$ of the desired filter, we can find a facotrization of the form

$$\frac{2a}{a^2 + (2\pi f)^2} = H(f)H(-f). \tag{39.9}$$

It is easy in this case to see what the two factors on the right side of (39.9) should be:

$$\frac{2a}{a^2 + (2\pi f)^2} = \left(\frac{\sqrt{2a}}{a + j(2\pi f)}\right)\left(\frac{\sqrt{2a}}{a - j(2\pi f)}\right). \tag{39.10}$$

We can take either of the two factors on the right side of (39.10) to be $H(f)$. This will give us two solutions for $h(t)$. One of the solutions gives us a causal filter and the other one gives us an anticausal filter. Obviously, for implementation purposes, we prefer to choose the causal filter as our solution. The causal solution is

$$H(f) = \frac{\sqrt{2a}}{a + j(2\pi f)}.$$

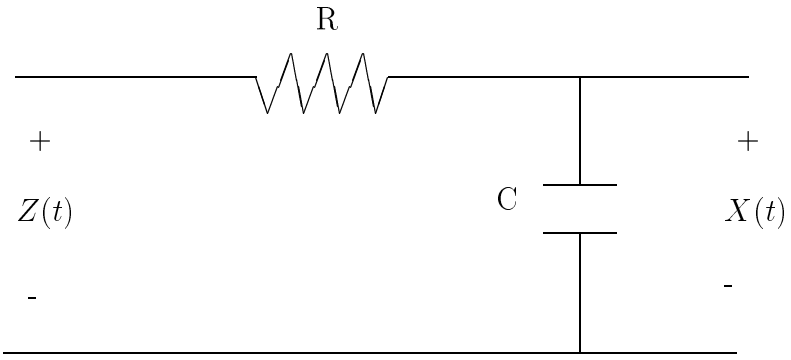
Taking the inverse Fourier transform, we see that

$$h(t) = \sqrt{2a}\exp(-at)u(t).$$

39.2.2 RC Filtering of White Noise

In this section, you learn that one possible solution to the spectral synthesis filter design problem of the preceding section is to use an RC filter.

We consider the following RC filter:



The input voltage $Z(t)$ is white noise with $R_Z(\tau) = \delta(\tau)$. Let us determine the autocorrelation function $R_X(\tau)$ of the random output voltage $X(t)$. From your earlier coursework, you know that the frequency response function of the RC filter is

$$H(f) = \frac{1}{RCj2\pi f + 1}$$

and therefore

$$S_X(f) = |H(f)|^2 = \frac{1}{(RC)^2(2\pi f)^2 + 1}.$$

Using the Fourier transform pair

$$\exp(-\alpha|\tau|) \leftrightarrow \frac{2\alpha}{\alpha^2 + (2\pi f)^2}$$

we see that

$$R_X(\tau) = (1/2RC)\exp(-|\tau|/RC).$$

Our conclusion is that if we use an RC filter to filter white noise, then we do obtain output autocorrelation function in the form of a two-sided decaying exponential function.

Let us go further with this example to see if we can compute P_X three different ways. The first way to do it is to use $R_X(\tau)$:

$$P_X = R_X(0) = \frac{1}{2RC}.$$

The second way to do it is to integrate the power spectrum:

$$P_X = \int_{-\infty}^{\infty} S_X(f)df = \int_{-\infty}^{\infty} \frac{1}{(RC)^2(2\pi f)^2 + 1}df = \frac{1}{2RC}.$$

Evaluate the integral in the preceding line using the calculus formula

$$\int \frac{a}{a^2 + u^2}du = Tan^{-1}(u/a).$$

The third way to do it is to use the impulse response function, which is

$$h(t) = \mathcal{F}^{-1}\left[\frac{1}{RCj2\pi f + 1}\right] = \left(\frac{1}{RC}\right)\exp(-t/RC)u(t).$$

We have

$$P_X = \int_{-\infty}^{\infty} h(t)^2 dt = \int_0^{\infty} \left(\frac{1}{(RC)^2}\right)\exp(-2t/RC)dt = \frac{1}{2RC}.$$

39.3 Filter Output Mean Function

Let $X(t)$ be any random process (not necessarily WSS). Suppose we pass $X(t)$ through an arbitrary linear filter as follows:

$$X(t) \rightarrow \boxed{\begin{array}{c}\text{linear}\\ \text{filter}\end{array}} \rightarrow Y(t)$$

The question we answer in this section is how to compute the mean function $\mu_Y(t)$ of the filter output process $Y(t)$. Suppose we perform our random experiment over and over again (independent trials) in order to obtain N realizations of $X(t)$, denoted as follows:

$$x^i(t), \quad i = 1, 2, \cdots, N.$$

Suppose we pass each realization $x^i(t)$ through the linear system:

$$x^i(t) \rightarrow \boxed{\begin{array}{c}\text{linear}\\ \text{filter}\end{array}} \rightarrow y^i(t)$$

The resulting output signal $y^i(t)$ can be regarded as a realization of the random output signal $Y(t)$. Suppose we apply as input to the linear system the signal

$$\frac{\sum_{i=1}^N x^i(t)}{N}. \tag{39.11}$$

Then the principle of superposition from EE 3025 tells us that the corresponding system response will be

$$\frac{\sum_{i=1}^N y^i(t)}{N}. \tag{39.12}$$

The law of large numbers tells us that for large N, it is highly likely that the signal (39.11) will be close to the signal $\mu_X(t)$ and that the signal (39.12) will be close to the signal $\mu_Y(t)$. If we pass to the limit as $N \rightarrow \infty$, we are forced to conclude that $\mu_Y(t)$ is the response of the linear system to input signal $\mu_X(t)$. This argument has given us the following useful result.

Useful Result: If random signal $Y(t)$ is the output from a linear system when the input is random signal $X(t)$, then the mean function $\mu_Y(t)$ will be the output from the linear system when the input is the signal $\mu_X(t)$. Schematically, we have

$$\mu_X(t) \rightarrow \boxed{\begin{array}{c}\text{linear}\\ \text{filter}\end{array}} \rightarrow \mu_Y(t)$$

Remark. The Useful Result is also true for discrete-time linear filtering of discrete-time random signals.

Example 39.3. Let A be a RV taking the values $0, 1$ with prob. $1/2$ each. Consider the process

$$X(t) = \begin{cases} At, & t \geq 0 \\ 0, & elsewhere \end{cases}$$

Let $Y(t)$ be the integrated process:

$$X(t) \rightarrow \boxed{\int_0^t} \rightarrow Y(t)$$

It is easily determined that $\mu_X(t) = tu(t)/2$. We get $\mu_Y(t)$ from a consideration of the block diagram:

$$\mu_X(t) = tu(t)/2 \rightarrow \boxed{\int_0^t} \rightarrow \mu_Y(t) = t^2u(t)/4$$

39.3.1 Special Case: Output Mean in LTI Filtering of WSS Process

Suppose now we take our linear system to be a stable LTI filter with impulse response function $h(t)$. As input to this filter, we apply a WSS random signal $X(t)$. The resulting output signal $Y(t)$ is also WSS. The mean function of process X is the constant μ_X, and the mean function of process Y is the constant μ_Y. Suppose we consider μ_X and μ_Y to be DC signals. Then from our Useful Result above, we see that the DC signal μ_Y must be the filter output when the filter input is the DC signal μ_X. Via convolution, we have

$$\mu_Y = \mu_X * h(t).$$

Doing the convolution on the right side, we obtain the following formula that allows us to compute μ_Y from μ_X:

$$\mu_Y = \mu_X \left\{ \int_{-\infty}^{\infty} h(t)dt \right\}.$$

We can also relate μ_Y to μ_X using the frequency response function $H(f)$. We know from EE 3015 that if the input to the filter is the complex sinusoid

$$C\exp(j2\pi f_o t),$$

then the corresponding filter output will be the complex sinusoid

$$CH(f_0)\exp(j2\pi f_0 t).$$

In particular, we can see what happens when $f_0 = 0$. Then the filter output when the filter input is the DC signal μ_X will be the DC signal $\mu_X H(0)$. This reasoning has given us the following second formula relating μ_Y and μ_X:

$$\mu_Y = \mu_X H(0).$$

We obtain a similar result relating μ_Y and μ_X for LTI discrete-time filtering of a WSS discrete-time process. We summarize our results as follows.

Useful Result 2:

(a): Suppose you filter a continuous-time WSS process $X(t)$ with continuous-time LTI stable filter having impulse response $h(t)$ and frequency response $H(f)$, thereby obtaining WSS process $Y(t)$ at the filter output. Then the input and output process means are related by the formulas

$$\begin{aligned}\mu_Y &= \mu_X\left\{\int_{-\infty}^{\infty} h(t)dt\right\}\\ \mu_Y &= \mu_X H(0)\end{aligned}$$

(b): Suppose you filter a discrete-time WSS process X_n with continuous-time LTI stable filter having impulse response $h[n]$ and frequency response $H(f)$, thereby obtaining WSS process Y_n at the filter output. Then the input and output process means are related by the formulas

$$\begin{aligned}\mu_Y &= \mu_X\left\{\sum_{n=-\infty}^{\infty} h[n]\right\}\\ \mu_Y &= \mu_X H(0)\end{aligned}$$

Example 39.4. Let X_n be a WSS process with $\mu_X = 1$. Let Y_n be the process arising from the filtering operation

$$Y_n = (1/2)Y_{n-1} + X_n.$$

The filter impulse response and frequency response are given by

$$\begin{aligned} h[n] &= (1/2)^n u[n] \\ H(f) &= \frac{\exp(j2\pi f)}{\exp(j2\pi f) - (1/2)}, \end{aligned}$$

as every good EE 3015 student knows. The mean μ_Y of process Y_n can then be computed in two different ways:

$$\begin{aligned} \mu_Y &= \mu_X \sum_n h[n] = \sum_{n=0}^{\infty} (1/2)^n = 2 \\ &= \mu_X H(0) = \frac{1}{1 - (1/2)} = 2 \end{aligned}$$

Lecture 40

Random Processes Part 10

40.1 Random DC Signal

The random DC signal is possibly the simplest of all WSS random signals, and I should have presented it to you earlier. (I forgot.)

Let A be a RV. Then

$$X(t) = A, \quad -\infty < t < \infty$$

defines a random DC signal. It is WSS and its mean and autocorrelation function (whose simple derivations we omit) are

$$\begin{aligned} \mu_X &= E[A] \\ R_X(\tau) &= E[A^2] \end{aligned}$$

Because of the Fourier transform pair

$$1 \leftrightarrow \delta(f),$$

the power spectral density of our random DC signal would then be

$$S_X(f) = E[A^2]\delta(f).$$

The power P_X generated by the random DC signal is

$$P_X = E[A^2].$$

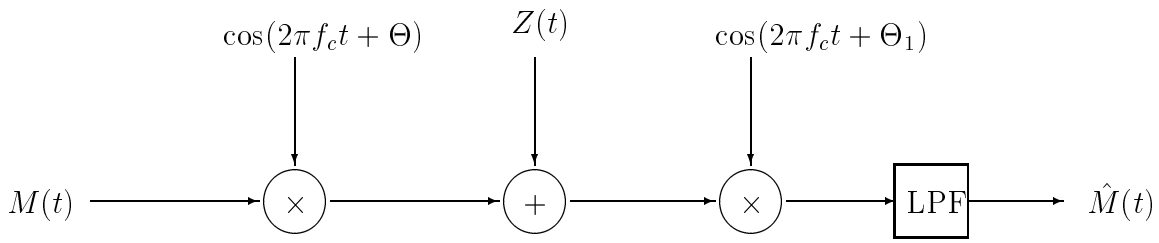
The process variance is

$$\sigma_A^2 = Var[A].$$

The random DC signal $X(t)$ is clearly nonergodic (unless A is a constant).

40.2 Amplitude Modulation

Let $M(t)$ be a random message signal, assumed to be WSS. Here is the block diagram of an amplitude modulation (AM) system for transmitting $M(t)$:



The signal $\cos(2\pi f_c t + \Theta)$ is called the carrier wave and f_c is the carrier frequency; in order that we may take the carrier wave to be a WSS random signal (which simplifies the analysis), we make the usual assumption that Θ is uniformly distributed between 0 and 2π. The random signal

$$X(t) = M(t)\cos(2\pi f_c t + \Theta)$$

is the AM modulated signal. The channel noise $Z(t)$ is assumed to be white and uncorrelated with $X(t)$. The filter marked LPF is a low pass filter.

Let us determine the power spectrum of the modulated signal $X(t)$. From our earlier work on random sinusoids, we know that the autocorrelation function of the carrier wave is $\cos(2\pi f_c \tau)/2$. Assuming the carrier wave and the message signal to be statistically independent, we know from Section 37.1.2 that the autocorrelation function of $X(t)$ is the product of the autocorrelation function of $M(t)$ and the autocorrelation function of the carrier wave:

$$R_X(\tau) = R_M(\tau)\cos(2\pi f_c \tau)/2$$

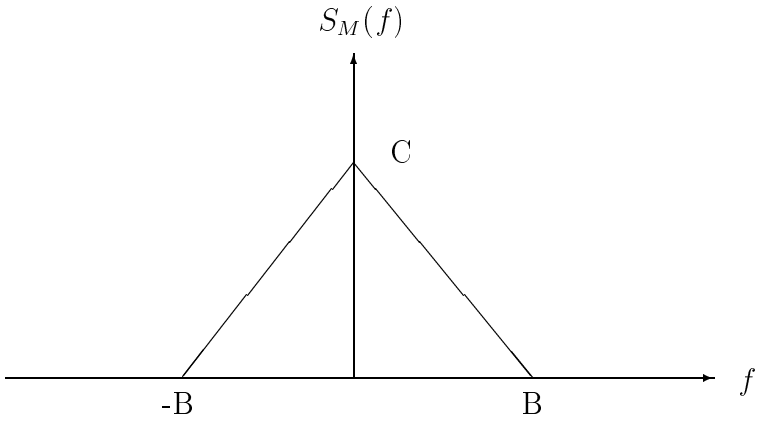
We can take the Fourier transform using the Fourier transform pair

$$\phi_1(t)\phi_2(t) \leftrightarrow \Phi_1(f) * \Phi_2(f)$$

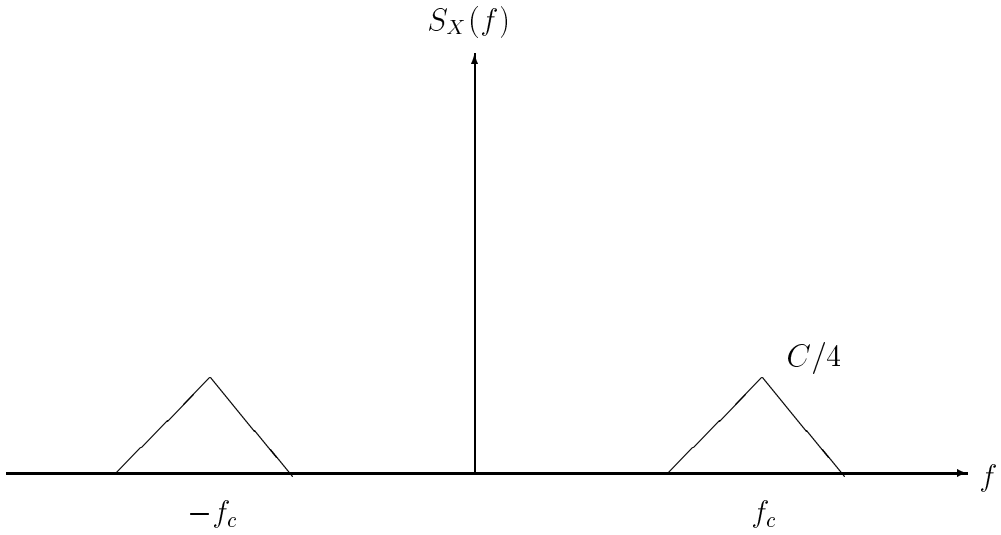
getting

$$\begin{aligned} S_X(\tau) &= S_M(f) * [\delta(f - f_c) + \delta(f + f_c)]/4 \\ &= S_M(f - f_c)/4 + S_M(f + f_c)/4 \end{aligned}$$

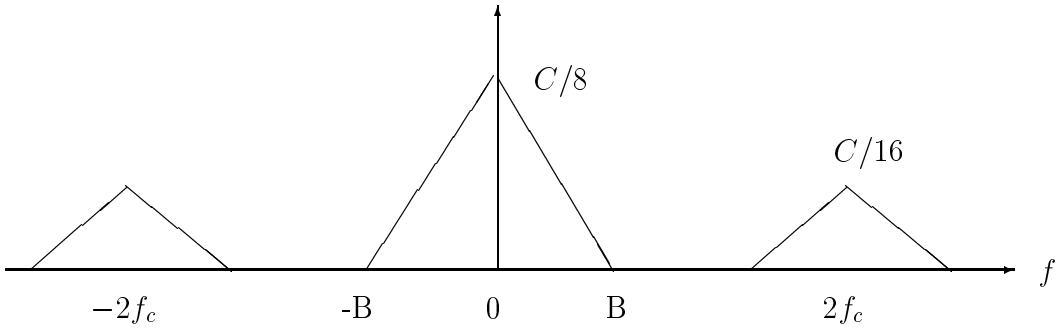
In other words, AM modulation "frequency shifts" the message signal power spectrum to the right and to the left an amount equal to f_c. For example, if the message power spectrum looks like



then the AM modulated power spectrum looks like

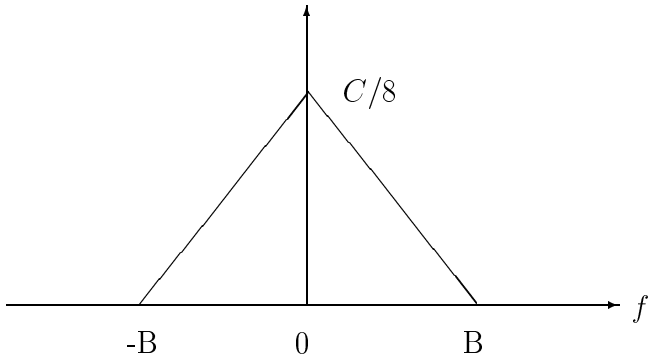


At the receiving end of the channel, note that we have a second multiplier in which we multiply by carrier wave $\cos(2\pi f_c + \Theta_1)$, where Θ_1 is uniformly distributed between 0 and 2π. (Since the carrier waves before and after transmission need not be synchronized with respect to phase, it is not necessarily true that $\Theta_1 = \Theta$; whether this is true or not makes no difference in our analysis.) Doing the "frequency shifting" a second time, we see that the power spectrum of the signal part of the second multiplier output will look like

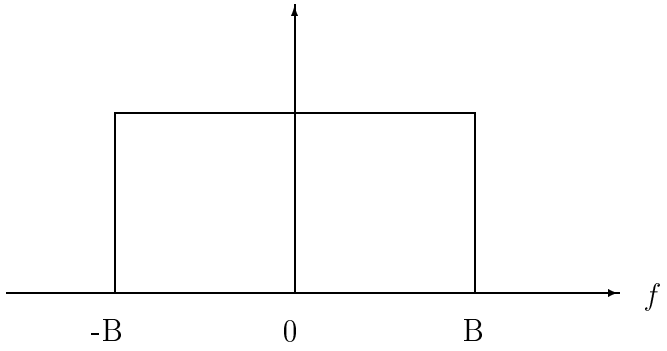


The noise component of the second multiplier output will still be white. Taking LPF to be an ideal low pass filter with bandwidth B (the message signal bandwidth), the power spectra of the signal and noise components of the low pass filter output then look like

LPF output power spectrum (signal part)



LPF output power spectrum (noise part)



In our analysis up to now, we have taken the amplitude of the carrier wave in the AM modulator to be 1. As we increase the amplitude of the carrier wave through various amplitude levels above

1, the noise power at the LPF output remains fixed, so that we may set the signal-to-noise ratio at the LPF output to any desired level by adjusting the carrier wave amplitude appropriately.

40.3 Receiver Design Methods For AWN Channels

The quintessential communication systems channel model is the additive white noise (AWN) channel, given by the following block diagram:

$$X(t) \rightarrow \boxed{\text{AWN channel}} \rightarrow Y(t) = X(t) + Z(t)$$

The input random signal $X(t)$ to the AWN channel is assumed to be WSS. The channel noise process $Z(t)$ is assumed to be white noise with autocorrelation function

$$R_Z(\tau) = A\delta(\tau).$$

We make the usual assumption that the X process and Z process are uncorrelated, so that

$$E[X(s)Z(t)] = E[X(s)]E[Z(t)] = 0,$$

for all times s, t. (The preceding "cross correlation" vanishes because $E[Z(t)] = 0$ is part of the white noise assumption.) To complete the design of the communication system, a receiver must be designed to process the channel output random signal $Y(t)$. We will only consider receivers which are LTI stable linear systems; let $h(t)$ denote the receiver's impulse response function. The following block diagram illustrates the scenario that occurs when the receiver processes $Y(t)$:

$$Y(t) = X(t) + Z(t) \rightarrow \boxed{\begin{array}{c}\text{receiver} \\ h(t)\end{array}} \rightarrow X^0(t) + Z^0(t)$$

Process $X^0(t)$ is the *signal component* of the receiver output, that is, random signal $X^0(t)$ is the receiver's response to the random signal $X(t)$. Process $Z^0(t)$ is the *noise component* of the receiver output, that is, random signal $Z^0(t)$ is the receiver's response to the noise $Z(t)$. In order to judge how well the receiver is doing, some people look at the SNR (signal to noise ratio) at the receiver output, which is defined by:

$$receiver\ output\ SNR \triangleq \frac{X^0(t)\ power}{Z^0(t)\ power}.$$

If you have a choice between two receivers, it might make sense (all other things being equal) to choose the receiver which gives the bigger SNR at the receiver output. We can express the receiver output SNR in terms of the frequency response $H(f)$ of the receiver as follows:

$$receiver\ output\ SNR = \frac{\int_{-\infty}^{\infty} S_X(f)|H(f)|^2 df}{\int_{-\infty}^{\infty} S_Z(f)|H(f)|^2 df} = \frac{\int_{-\infty}^{\infty} S_X(f)|H(f)|^2 df}{A\int_{-\infty}^{\infty} |H(f)|^2 df}$$

To conclude this section, we point out three common receiver design methodologies.

Method 1: Low Pass Filtering. Suppose the bandwidth of the signal $X(t)$ is B. That is,

$$S_X(f) = 0, \ \ |f| > B.$$

Then one can use as receiver a low pass filter will frequency response

$$H(f) = \begin{cases} 1, & -B \leq f \leq B \\ 0, & elsewhere \end{cases}$$

The advantage to this receiver is that it passes $X(t)$ through perfectly (that is, $X^0(t) = X(t)$) while suppressing the noise in the sense that the noise power spectrum at the receiver output will be limited strictly to the frequency band $-B \leq f \leq B$. The SNR at the receiver output for the low pass filter receiver is given by the formula

$$receiver\ output\ SNR\ = \frac{P_X}{2AB}.$$

The low pass filter receiver is sometimes regarded as a *prefiltering operation*. That is, one can design a second receiver to be applied to the output of the low pass filter. Since the low pass filter passes $X(t)$ through perfectly, there is no loss of generality if in our overall receiver design we first prefilter with a low pass filter.

Method 2: Max SNR Filtering. Suppose one has a class of possible $h(t)$'s which could potentially be used for the receiver's impulse response. In Max SNR Filtering, one would choose that $h(t)$ from the class for which the receiver output SNR is the biggest. We will consider Max SNR Filtering in the lecture notes for Lecture 41.

Method 3: Wiener Filtering. Suppose we wish to regard the receiver output as an estimate $\hat{X}(t)$ of the random signal $X(t)$:

$$Y(t) \rightarrow \boxed{\begin{array}{c}\text{receiver} \\ h(t)\end{array}} \rightarrow \hat{X}(t)$$

Suppose one has a class of possible $h(t)$'s that can be considered for the receiver impulse response. In the Wiener filtering approach, one chooses $h(t)$ from this class so that $\hat{X}(t)$ will be the *minimum mean square estimate* of $X(t)$, that is, we want the mean square estimation error

$$E[(X(t) - \hat{X}(t))^2] \tag{40.1}$$

to be minimized. Note that the quantity (40.1) does not depend on t because we assumed the signal $X(t)$ is WSS. As a result, there will be just one receiver that will provide the minimum

mean square estimation error (40.1) for all times t. (If we had to change the receiver's characteristics with time, we would obtain a time-varying filter by this approach and not a time-invariant filter.) This unique receiver is called a *Wiener filter*. In Section 40.4, I design a type of Wiener filter called *FIR Wiener filter*. In the notes for Lecture 41, I will derive another Wiener filter called the *noncausal Wiener filter*.

40.4 FIR Wiener Filter Design

I will do FIR Wiener Filter Design in the context of discrete-time signals and systems. We have a WSS random signal X_n. It is sent through the AWN channel and is then processed by a receiver with impulse response $h[n]$:

$$X_n \rightarrow \boxed{\text{AWN channel}} \rightarrow Y_n = X_n + Z_n \rightarrow \boxed{h[n]} \rightarrow \hat{X}_n$$

The channel noise process Z_n is white noise, and is uncorrelated with the random signal X_n.

Let k be any positive integer. I show you how to design the k-tap FIR Wiener filter. This means $\hat{X}_n$ has the form

$$\hat{X}_n = h[0]Y_n + h[1]Y_{n-1} + h[2]Y_{n-2} + \ldots + h[k-1]Y[n-k+1].$$

We must determine the k filter tap weights

$$h[0], h[1], \cdots, h[k-1]$$

so that the MS estimation error $E[(X_n - \hat{X}_n)^2]$ is minimized. (Since the filter we are designing is an FIR filter, the remaining $h[n]$ values are all equal to zero.) The *orthogonality principle* of linear MS estimation theory tells us that the estimation error $X_n - \hat{X}_n$ must be orthogonal to each of the k observations Y_{n-i} $(i = 0, 1, \ldots, k-1)$ that are being combined to form the estimate $\hat{X}_n$ of X_n. That is, we have the equations

$$E[(X_n - \hat{X}_n)Y_{n-i}] = 0, \quad i = 0, 1, \ldots, k-1.$$

These equations simplify to the equations

$$E[\hat{X}_n Y_{n-i}] = E[X_n Y_{n-i}], \quad i = 0, 1, \ldots, k-1. \tag{40.2}$$

Notice that

$$E[X_n Y_{n-i}] = E[X_n X_{n-i}] + E[X_n Z_{n-i}] = E[X_n X_{n-i}] + 0 = R_X(i).$$

This allows us to simplify equations (40.2) to

$$h[0]R_Y(i) + h[1]R_Y(i-1) + h[2]R_Y(i-2) + \ldots + h[k]R_Y(i-k) = R_X(i), \quad i = 0, 1, \ldots, k-1.$$

The autocorrelations $R_Y(0), R_Y(1), \ldots, R_Y(k-1)$ are computed according to the formula

$$R_Y[\tau] = R_X[\tau] + R_Z[\tau],$$

which we learned in Section 37.1.1. Let R_Y^k be the $k \times k$ matrix such that the element in row i and column j is $R_Y(i-j)$. (That is, R_Y^k is the correlation matrix of any k consecutive samples of the Y process.) We can write our system of equations in matrix form as:

$$R_Y^k \begin{bmatrix} h[0] \\ h[1] \\ h[2] \\ \vdots \\ h[k-1] \end{bmatrix} = \begin{bmatrix} R_X(0) \\ R_X(1) \\ R_X(2) \\ \vdots \\ R_X(k-1) \end{bmatrix}$$

You solve these equations to find the k-tap Wiener filter impulse response function $h[n]$.

Example 40.1. For a two-tap Wiener filter, the receiver output takes the form

$$\hat{X}_n = h[0]Y_n + h[1]Y_{n-1},$$

where we find the two tap weights $h[0], h[1]$ by solving the equations

$$\begin{bmatrix} R_Y[0] & R_Y[1] \\ R_Y[1] & R_Y[0] \end{bmatrix} \begin{bmatrix} h[0] \\ h[1] \end{bmatrix} = \begin{bmatrix} R_X[0] \\ R_X[1] \end{bmatrix}.$$

Let's suppose that

$$R_X(\tau) = (1/2)^{|\tau|} + (1/4)^{|\tau|}.$$

and that

$$R_Z(\tau) = \delta[\tau].$$

Then

$$\begin{aligned} R_X(0) &= 2 \\ R_X(1) &= 3/4 \\ R_Y(0) &= R_X(0) + R_Z(0) = 3 \\ R_Y(1) &= R_X(1) + R_Z(1) = 3/4. \end{aligned}$$

Plugging these values in the equations above, we have to solve

$$\begin{bmatrix} 3 & 3/4 \\ 3/4 & 3 \end{bmatrix} \begin{bmatrix} h[0] \\ h[1] \end{bmatrix} = \begin{bmatrix} 2 \\ 3/4 \end{bmatrix}.$$

The solutions are

$$h[0] = 29/45, \quad h[1] = 4/45.$$

Example 40.2. For a three-tap Wiener filter, the receiver output takes the form

$$\hat{X}_n = h[0]Y_n + h[1]Y_{n-1} + h[2]Y_{n-2},$$

where we find the three tap weights $h[0], h[1], h[2]$ by solving the equations

$$\begin{bmatrix} R_Y(0) & R_Y(1) & R_Y(2) \\ R_Y(1) & R_Y(0) & R_Y(1) \\ R_Y(2) & R_Y(1) & R_Y(0) \end{bmatrix} \begin{bmatrix} h[0] \\ h[1] \\ h[2] \end{bmatrix} = \begin{bmatrix} R_X(0) \\ R_X(1) \\ R_X(2) \end{bmatrix}.$$

Let's use the same X and Z process characteristics that we did in Example 40.1. Then the above system of equations becomes

$$\begin{bmatrix} 3 & 3/4 & 5/16 \\ 3/4 & 3 & 3/4 \\ 5/16 & 3/4 & 3 \end{bmatrix} \begin{bmatrix} h[0] \\ h[1] \\ h[2] \end{bmatrix} = \begin{bmatrix} 2 \\ 3/4 \\ 5/16 \end{bmatrix}.$$

The solution is

$$h[0] = 0.6437, \quad h[1] = 0.0851, \quad h[2] = 0.0158.$$

Lecture 41

Random Processes Part 11

41.1 FIR Max SNR Filter Design

In the block diagram

$$X[n] \rightarrow \boxed{\text{channel}} \rightarrow Y[n] = X[n] + Z[n] \rightarrow \boxed{h[n]} \rightarrow X_0[n] + Z_0[n]$$

we assume that the random signal $X[n]$ is WSS and that the random noise $Z[n]$ is white with $R_Z[\tau] = \delta[\tau]$. The filter with impulse response $h[n]$ is to be designed so that the filter output signal-to-noise ratio (SNR) is maximized:

$$\text{SNR} = \frac{X_0[n] \text{ power}}{Z_0[n] \text{ power}}$$

In the SNR ratio, $X_0[n]$ is the signal part of the filter output and $Z_0[n]$ is the noise part of the filter output. The resulting filter is called a max SNR filter. There is more than one max SNR filter, depending upon the form of the filter. For example, for each k, one could design a unique max SNR filter to be a FIR filter with k taps. The resulting k tap max SNR filter would provide an output satisfying

$$X_0[n] + Z_0[n] = h[0]Y[n] + h[1]Y[n-1] + \ldots + h[k-1]Y[n-k+1]$$

and the SNR would be expressible in the form

$$\text{SNR} = \frac{\sum_{i,j=0}^{k-1} h[i]h[j]R_X[i-j]}{\sum_{i=0}^{k-1} h[i]^2}$$

Notice that if we take a scalar multiple of $h[n]$, the SNR does not change. Therefore, in maximizing the SNR, we may assume that $h[0]^2 + h[1]^2 + \ldots + h[k-1]^2 = 1$. Let R be the $k \times k$ matrix such

that the element in row i and column j is $R_X[i-j]$, and let h be the k-dimensional column vector whose entries are $h[0], h[]1, \ldots, h[k-1]$. Then, to find the k-tap max SNR filter we must solve the optimization problem

$$\max_{||h||=1} h^T R h$$

where

$$||h|| = \sqrt{h[0]^2 + h[1]^2 + \ldots + h[k-1]^2}$$

is the length of the vector h. It is well known how to solve this problem: Let λ be the maximum eigenvalue of R. Then, h is the eigenvector of R of length one corresponding to the eigenvalue λ, obtained by solving the equation

$$Rh = \lambda h$$

(h is unique up to sign). Matlab can be used to quickly find the eigenvalues and eigenvectors of a matrix. Therefore, max SNR filter design is easily accomplished with Matlab, as the following example illustrates.

Example 41.1. Let $R_X[\tau] = 8(2^{-|\tau|})$. Let us find the 4-tap max SNR filter. We have

$$R = \begin{bmatrix} R_X[0] & R_X[1] & R_X[2] & R_X[3] \\ R_X[1] & R_X[0] & R_X[1] & R_X[2] \\ R_X[2] & R_X[1] & R_X[0] & R_X[1] \\ R_X[3] & R_X[2] & R_X[1] & R_X[0] \end{bmatrix} = \begin{bmatrix} 8 & 4 & 2 & 1 \\ 4 & 8 & 4 & 2 \\ 2 & 4 & 8 & 4 \\ 1 & 2 & 4 & 8 \end{bmatrix}$$

```
R=toeplitz([8 4 2 1]);
[a,b]=eig(R)
a =

    0.3162   -0.5573   -0.6325    0.4352
   -0.6325    0.4352   -0.3162    0.5573
    0.6325    0.4352    0.3162    0.5573
   -0.3162   -0.5573    0.6325    0.4352


b =

    3.0000         0         0         0
         0    4.3153         0         0
         0         0    8.0000         0
         0         0         0   16.6847
```

The diagonal elements of b are the eigenvalues of R. The largest one, 16.6847, is in the fourth position, so the fourth column of a is the eigenvector corresponding to this eigenvalue. We see that the desired impulse response is

$$h[0] = 0.4352, \quad h[1] = 0.5573, \quad h[2] = 0.5573, \quad h[3] = 0.4352$$

The maximum possible SNR for a 4-tap filter is the largest eigenvalue just found, which is 16.6847.

Max SNR filtering is also called *principal component filtering*.

41.2 More Time-Varying Linear Filtering of Gaussian White Noise

When we earlier defined the Brownian motion process in Section 37.3, we obtained it by performing a certain time-varying linear filtering operation on a Gaussian white noise random signal. In this section, we briefly point out that we can obtain other processes, which behave much like the Brownian motion process, by similar filtering operations.

Let $X(t)$ be Gaussian white noise. Let $\phi(t)$ be any deterministic signal. Then

$$Y(t) = \int_0^t \phi(u)X(u)du, \quad t \geq 0,$$

defines a random process $Y(t)$. You may visualize $Y(t)$ as being formed from $X(t)$ via the following two-step linear filtering operation on $X(t)$: (1) First multiply $X(t)$ by $\phi(t)$, and then (2) pass the resulting signal through the integrator $\int_0^t$.

By looking at different $\phi(t)$'s, one obtains many different processes $Y(t)$. One of these is the Brownian motion process (just take $\phi(t) = 1$ for all t). All of these $Y(t)$ processes share the following properties (which we noticed earlier were true of the Brownian motion process):

- Realizations of $Y(t)$ process start at origin, that is, $Y(0) = 0$.
- $Y(t)$ is not a WSS process.[1]
- $Y(t)$ is an independent increments process.
- $Y(t)$ is a Gaussian process with zero mean.

In Section 37.3, we employed a "double integral trick" to compute the variance of each component RV of the Brownian motion process. A similar trick can be exploited to compute autocorrelations for the $Y(t)$ process. Observe that

$$\begin{aligned} Y(t_1 Y(t_2) &= \left[\int_0^{t_1} \phi(t)X(t)dt\right] \left[\int_0^{t_2} \phi(s)X(s)ds\right] \\ &= \int_0^{t_1} \int_0^{t_2} \phi(t)\phi(s)X(s)X(t)dsdt. \end{aligned}$$

[1] There are trivial exceptions to this. For example, take $\phi(t) = 0$ for all t. Then $Y(t)$ is a trivial WSS process equal to zero at all times.

When you compute the expected value of $Y(t_1)Y(t_2)$, you can pull the expected value operator inside the double integral, yielding

$$R_Y(t_1,t_2) = E[Y(t_1)Y(t_2)] = \int_0^{t_1}\int_0^{t_2} E[\phi(s)\phi(t)X(s)X(t)]dsdt = \int_0^{t_1}\int_0^{t_2} \phi(s)\phi(t)E[X(s)X(t)]dsdt.$$

Let us suppose that

$$R_X(\tau) = A\delta(\tau).$$

Then we have

$$R_Y(t_1,t_2) = A\int_0^{t_1}\int_0^{t_2} \phi(t)\phi(s)\delta(s-t)dsdt = A\int_0^{\min(t_1,t_2)} \phi(t)^2 dt.$$

41.3 Periodogram Estimate of Power Spectrum

For an ergodic WSS continuous-time process $X(t)$ with realization $x(t)$, we want to be able to say the following:

$$S_X(f) \approx \frac{1}{T}\left|\int_0^T x(t)e^{-j2\pi jt}dt\right|^2 \quad (large\ T) \tag{41.1}$$

The right side of formula (41.1) is called the *periodogram estimate* of the power spectrum. Notice that the periodogram estimate is random because the process realization is random. Let us use $\hat{S}_X^T(f)$ to denote the periodogram estimate, that is, we define

$$\hat{S}_X^T(f) \triangleq \frac{1}{T}\left|\int_0^T X(t)e^{-j2\pi jt}dt\right|^2.$$

In the integrand on the right side, notice that we have written $X(t)$ instead of $x(t)$ to denote that the realization is random and therefore, for fixed frequency f, $\hat{S}_X^T(f)$ is a random variable.

With the tools available to us in EE 3025, we will not be able to prove that the approximation (41.1) is highly likely to be valid when we observe a realization $x(t)$ of $X(t)$. However, we will be able to prove the next best thing, namely, that $\hat{S}_X^T(f)$ is an *asymptotically unbiased estimator* of $S_X(f)$ as $T \to \infty$. That is, we shall prove that

$$E[\hat{S}_X^T(f)] \to S_X(f)\ as\ T \to \infty. \tag{41.2}$$

I will prove (41.2) via a "double integral trick" similar to what I did in Section 41.2. I start with $\hat{S}_X^{2T}(f)$ instead of $\hat{S}_X^T(f)$ because it is easier to work with. By wide-sense stationarity, we may write

$$E[\hat{S}_X^{2T}(f)] = \frac{1}{2T}E\left[\left|\int_0^{2T} X(t)e^{-j2\pi jt}dt\right|^2\right] = \frac{1}{2T}E\left[\left|\int_{-T}^{T} X(t)e^{-j2\pi jt}dt\right|^2\right].$$

We now do the double integral trick:

$$\left| \int_{-T}^{T} X(t) e^{-j2\pi f t} dt \right|^2 = \int_{-T}^{T} \int_{-T}^{T} X(t_1) X(t_2) e^{-j2\pi f (t_1 - t_2)} dt_1 dt_2.$$

We then take the expected value of both sides, moving the expected value into the double integral:

$$\begin{aligned} E\left[\int_{-T}^{T} \int_{-T}^{T} X(t_1) X(t_2) e^{-j2\pi f (t_1 - t_2)} dt_1 dt_2 \right] &= \int_{-T}^{T} \int_{-T}^{T} E[X(t_1) X(t_2)] e^{-j2\pi f (t_1 - t_2)} dt_1 dt_2 \\ &= \int_{-T}^{T} \int_{-T}^{T} R_X(t_1 - t_2) e^{-j2\pi f (t_1 - t_2)} dt_1 dt_2 \quad (41.3) \end{aligned}$$

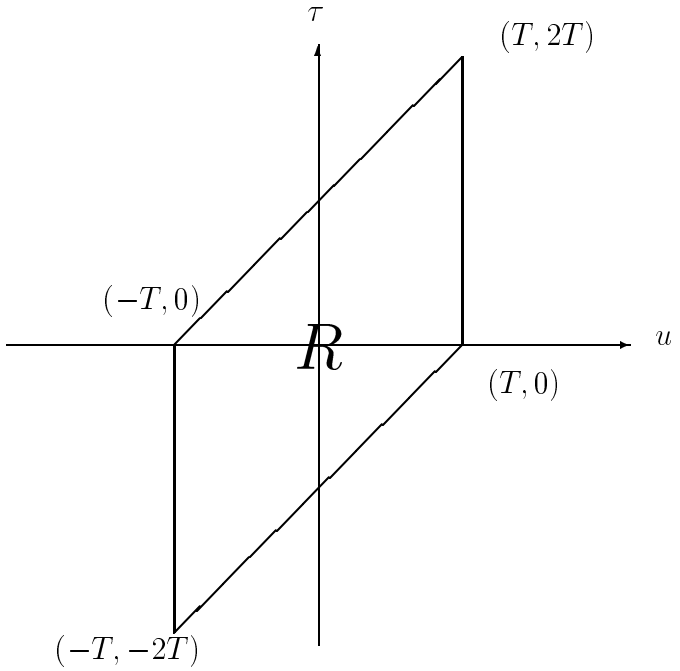
We then make the change of variable

$$\begin{aligned} u &= t_1 \\ \tau &= t_1 - t_2 \end{aligned}$$

In the double integral in (41.3), we are integrating in the (t_1, t_2) plane over the square

$$\{(t_1, t_2) : -T \le t_1 \le T, \ \ -T \le t_2 \le T\} \quad (41.4)$$

Under the above change of variable, we are integrating in the (u, τ) plane over the parallelogram R bounded by the four vertices $(T, 0), (T, 2T), (-T, 0), (-T, -2T)$ (transform the four vertices $(T, T), (T, -T), (-T, -T), (-T, T)$ of the square (41.4)). This region R is sketched as follows:



Under our change of variable, the integral in (41.3) becomes

$$\iint_R R_X(\tau)e^{-j2\pi f\tau}dud\tau$$

If we integrate in the u direction first, this integral becomes

$$\int_{-2T}^{2T} R_X(\tau)(2T-|\tau|)e^{-j2\pi f\tau}d\tau$$

(Because, if we fix a point τ between $-2T$ and $2T$ on the vertical axis of the preceding plot, the length of the horizontal slice through R at that point is $(2T-|\tau|)$.) We may split this integration into two parts:

$$2T\int_{-2T}^{2T} R_X(\tau)e^{-j2\pi f\tau}d\tau - \int_{-2T}^{2T}|\tau|R_X(\tau)e^{-j2\pi f\tau}d\tau. \tag{41.5}$$

Dividing the expression (41.5) by $2T$, we obtain $E[\hat{S}_X^{2T}(f)]$. We have shown that

$$E[\hat{S}_X^{2T}(f)] = \int_{-2T}^{2T} R_X(\tau)e^{-j2\pi f\tau}d\tau - \epsilon(T),$$

where

$$\epsilon(T) = \frac{1}{2T}\int_{-2T}^{2T}|\tau|R_X(\tau)e^{-j2\pi f\tau}d\tau.$$

It is not hard to see that $\epsilon(T)$ vanishes in the limit as $T\to\infty$, using the fact that

$$\left|\int_{-2T}^{2T}|\tau|R_X(\tau)e^{-j2\pi f\tau}d\tau\right| \le \int_{-\infty}^{\infty}|\tau|R_X(\tau)d\tau < \infty. \tag{41.6}$$

(We assume the integral on the right in (41.6) is finite; many autocorrelation functions decay sufficiently fast so that this will be true.) We have therefore shown that

$$\begin{aligned}\lim_{T\to\infty} E\left[\hat{S}_X^{2T}(f)\right] &= \lim_{T\to\infty}\int_{-2T}^{2T} R_X(\tau)e^{-j2\pi f\tau}d\tau \\ &= \int_{-\infty}^{\infty} R_X(\tau)e^{-j2\pi f\tau}d\tau = S_X(f),\end{aligned}$$

which is what we wanted to prove.

Discussion. The periodogram estimate of the power spectrum is not as good as some other power spectrum estimates. For example, Bartlett's method yields a better estimate. Bartlett's method partitions a long piece of the realization into nonoverlapping parts of equal length; a periodogram is computed for each part, and then the periodograms of the parts are averaged to obtain the Bartlett estimate of $S_X(f)$. Using the asymptotic unbiasedness of the periodogram estimate and the fact that process $X(t)$ is ergodic, one can show that the Bartlett power spectrum estimate will converge to $S_X(f)$ as we let the estimate be computed over a bigger and bigger piece of the realization of $X(t)$.

Discrete-Time Power Spectrum Estimation

For an ergodic WSS discrete-time process $X[n]$ with realization $x[n]$, the periodogram estimate of the power spectral density $S_X(f)$ is

$$\frac{1}{N}\left|\sum_{n=1}^{N} x[n]e^{-j2\pi n}\right|^2 .$$

One can show this estimate is asymptotically unbiased much as we did for the continuous-time case.

Recitation 14 gives you Matlab experiments in which realizations of discrete-time signals are simulated and their power spectra estimated from these realizations using periodogram estimates and Bartlett estimates.

Lecture 42

Random Processes Part 12

42.1 Noncausal Wiener Filter

In Section 40.4, we considered the FIR Wiener filter, which is the minimum MS receiver for the additive noise channel model which uses just finitely many of the received channel outputs to form the estimate of the random channel input signal. In this section, we develop the *noncausal Wiener filter*. This is the minimum MS receiver for the additive noise channel model which uses *all* of the channel outputs to estimate the channel input signal. Surprisingly, the noncausal Wiener filter is much easier to derive than the FIR Wiener filter. Although the noncausal Wiener filter is physically unrealizable, it is useful in that it gives the limiting performance of all minimum MS receiving filters with finite memory and delay. It is therefore the "granddaddy" of all possible MS receivers: it is the receiver whose performance one should try to get as close as possible to in receiver design.

As in Section 40.4, I will do Wiener Filter Design in the context of discrete-time signals and systems. We have a WSS random signal X_n. It is sent through an additive noise channel:

$$X_n \rightarrow \boxed{\text{additive noise channel}} \rightarrow Y_n = X_n + Z_n \rightarrow \boxed{h[n]} \rightarrow \hat{X}_n$$

The channel noise process Z_n need not be white: we will assume only that Z_n is WSS and is orthogonal to the channel input process in the sense that

$$E[X_n Z_i] = 0, \quad for\ all\ i, n.$$

This fact allows us to say the following two things about output autocorrelations and cross-correlations:

$$R_Y(\tau) = R_X(\tau) + R_Z(\tau) \tag{42.1}$$

$$E[X_n Y_i] = R_X(n-i), \quad for\ all\ i, n. \tag{42.2}$$

Let us now derive the noncausal Wiener filter impulse response function $h[n]$. Our starting point is the orthogonality principle. In this case, it says the the estimation error $X_n - \hat{X}_n$ is orthogonal to every Y_i, as i ranges through all integers. This gives us the equation

$$E[(X_n - \hat{X}_n)Y_i] = 0, \quad for\ all\ i, n.$$

We can re-write this as

$$E[\hat{X}_n Y_i] = E[X_n Y_i], \quad for\ all\ i, n.$$

Using (42.2) and the fact that the receiver output signal $\hat{X}_n$ is equal to the convolution $h[n] * Y_n$, this equation then becomes

$$E[(h[n] * Y_n)Y_i] = R_X(n-i), \quad for\ all\ i, n.$$

In the preceding equation, it suffices to consider the case $i = 0$ for us to be able to determine what $h[n]$ is; the $i = 0$ case gives us

$$E[(h[n] * Y_n)Y_0] = R_X(n), \quad for\ all\ n.$$

We can re-write the left hand side as

$$E[(h[n] * Y_n)Y_0] = E[(h[n] * \{Y_n Y_0\})] = h[n] * E[Y_n Y_0] = h[n] * R_Y(n).$$

We conclude that

$$h[n] * R_Y(n) = R_X(n), \quad for\ all\ n. \tag{42.3}$$

We remark that for the FIR Wiener filter, the two sides of equation (42.3) are equal for only finitely many n. Since the two sides of (42.3) are equal for all n in the case of the noncausal Wiener filter, we are allowed to take the Fourier transform of both sides of equation (42.3); it is this fact which makes it easier for us to find the noncausal Wiener than the FIR Wiener filter considered in Section 40.4. Fourier transforming both sides of (42.3), we obtain

$$H(f)S_Y(f) = S_X(f),$$

where $H(f)$ is the frequency response function of the noncausal Wiener filter. Using (42.3) and (42.1), we conclude that the frequency response function of the noncausal Wiener filter is given by the following simple formula:

$$H(f) = \frac{S_X(f)}{S_Y(f)} = \frac{S_X(f)}{S_X(f) + S_Z(f)}. \tag{42.4}$$

Remark. The preceding derivation goes through almost word for word for the derivation of the continuous-time noncausal Wiener filter. The formula (42.4) expresses both the frequency response

function of the discrete-time noncausal Wiener filter and the continuous-time noncausal Wiener filter.

Example 42.1. We derive a continuous-time noncausal Wiener filter. The channel input is a WSS process $X(t)$ with autocorrelation function

$$R_X(\tau) = \exp(-|\tau|).$$

The additive channel noise $Z(t)$ is assumed to have autocorrelation function

$$R_Z(\tau) = \delta(\tau) + 3\exp(-|\tau|).$$

The channel output process $Y(t) = X(t) + Z(t)$ is to be filtered by noncausal Wiener filter with frequency response function $h(t)$. Let's find $h(t)$. We have

$$\begin{aligned} S_X(f) &= \frac{2}{(2\pi f)^2 + 1} \\ S_Z(f) &= 1 + \frac{6}{(2\pi f)^2 + 1} \end{aligned}$$

Substituting these expressions into (42.4) we see that the frequency response function $H(f)$ of the noncausal Wiener filter is given by (after some algebraic simplification):

$$H(f) = \frac{2}{(2\pi f)^2 + 9},$$

from which it follows that

$$h(t) = (1/3)\exp(-3|t|).$$

Noncausal Wiener Filter MS Estimation Error

There is a nice formula for the MS estimation error of the noncausal Wiener filter. For the discrete-time noncausal Wiener filter, the formula is:

$$DT\ noncausal\ Wiener\ filter\ MS\ estimation\ error\ = \int_0^1 \left[\frac{S_X(f)S_Z(f)}{S_X(f) + S_Z(f)}\right] df. \tag{42.5}$$

For the continuous-time noncausal Wiener filter, the formula is:

$$CT\ noncausal\ Wiener\ filter\ MS\ estimation\ error\ = \int_{-\infty}^{\infty} \left[\frac{S_X(f)S_Z(f)}{S_X(f) + S_Z(f)}\right] df. \tag{42.6}$$

I finish this section with a derivation of the discrete-time formula (42.5). Obvious modifications in the proof can be made to obtain formula (42.6) for the continuous-time case.

For the discrete-time case, the MS estimation error is $E[(X_n - \hat{X}_n)^2]$, which can be rewritten as follows:

$$\begin{aligned} E[(X_n - \hat{X}_n)^2] &= E[(X_n - \hat{X}_n)(X_n - \hat{X}_n)] \\ &= E[(X_n - \hat{X}_n)X_n] - E[(X_n - \hat{X}_n)\hat{X}_n] \\ &= E[(X_n - \hat{X}_n)X_n] - 0 \end{aligned}$$

In the preceding, I used the fact that $E[(X_n - \hat{X}_n)\hat{X}_n] = 0$, in other words, the estimation error $X_n - \hat{X}_n$ is orthgonal to the estimate $\hat{X}_n$. This is true because the estimation error is orthogonal to every receiver input Y_i and therefore it must also be orthogonal to any linear combination of the Y_i's; in particular, the estimation error must be orthogonal to the estimate $\hat{X}_n$ itself. We can now simplify the MS estimation error further as follows:

$$\begin{aligned} E[(X_n - \hat{X}_n)X_n] &= E[X_n^2] - E[\hat{X}_n X_n] \\ &= R_X(0) - E[(h[n] * X_n)X_n] \\ &= R_X(0) - E[(\sum_{u=-\infty}^{\infty} h[n-u]X_u)X_n] \\ &= R_X(0) - E[\sum_{u=-\infty}^{\infty} h[n-u]X_u X_n] \\ &= R_X(0) - \sum_{u=-\infty}^{\infty} h[n-u]E[X_u X_n] \\ &= R_X(0) - \sum_{u=-\infty}^{\infty} h[n-u]R_X(n-u) \\ &= R_X(0) - \sum_{\tau=-\infty}^{\infty} h[\tau]R_X(\tau) \end{aligned}$$

We want to re-express this last expression in frequency domain. To do this, let $\phi[n]$ be the discrete-time signal

$$\phi[n] \triangleq h[n] * R_X(n).$$

It is not hard to see that

$$\phi[0] = \sum_{\tau=-\infty}^{\infty} h[\tau]R_X(\tau).$$

On the other hand, we know from EE 3015 Fourier transform theory that

$$\phi[0] = \int_0^1 \Phi(f)df = \int_0^1 H(f)S_X(f)df,$$

and so we can say that

$$\sum_{\tau=-\infty}^{\infty} h[\tau]R_X(\tau) = \int_0^1 H(f)S_X(f)df.$$

We conclude that

$$\begin{aligned} E[(X_n - \hat{X}_n)^2] &= R_X(0) - \int_0^1 H(f)S_X(f)df \\ &= \int_0^1 S_X(f)df - \int_0^1 H(f)S_X(f)df \\ &= \int_0^1 S_X(f)[1 - H(f)]df \end{aligned}$$

In this last expression, substitute the expression for the noncausal Wiener filter frequency response $H(f)$ on the right side of (42.4). After some simple algebra, you obtain the formula (42.5) for the discrete-time noncausal Wiener filter MS estimation error.

42.2 Single Server Queue Asymptotics

In Recitations 13 and 14, you worked a little bit with the single-server queueing system model. In this model, the arrivals of the message packets occur according to a Poisson process with arrival rate λ (packets/second). The single server serves packets (i.e., routes packets) at a rate of μ (packets/second). Suppose $\mu > \lambda$. (This is the case of a stable queue.) Let random variable W_n denote the waiting time of the n-th arriving packet, for $n = 1, 2, 3, \cdots$. (In other words, W_n is the length of time that packet n waits in the queue from its time of arrival until the time it starts to be served by the server.) In this section, I want to discuss a little bit the asymptotic behavior of the sequence of waiting times W_n. Since the queue is stable, the waiting times W_n will not blow up as $n \to \infty$. It is a well-known fact that can be found in any basic textbook on queueing systems that the probability density function of W_n for large n is approximately equal to the density

$$f(w) = \left(1 - \frac{\lambda}{\mu}\right)\delta(w) + \frac{\lambda}{\mu}(\mu - \lambda)e^{-(\mu-\lambda)w}u(w), \tag{42.7}$$

a mixed probability distribution. It is easy to see why this density function should be mixed: it may happen that when packet n arrives, the queue is empty, which means that the waiting time W_n takes the value 0, producing the delta function component at the beginning of $f(w)$; on the other hand, if the length of the queue is > 0 upon the packet's arrival, the packet will wait an amount of time which is continuously distributed (in fact, formula (42.7) tells us that this continuous conditional distribution is exponential with mean $1/(\mu - \lambda)$).

Here are some simple conclusions we can draw from the form of the asymptotic waiting time density (42.7):

- For any nonnegative real numbers $a \leq b$, we have

$$P[a \leq W_n \leq b] \approx \int_a^b f(w)dw, \quad for\ large\ n.$$

  In particular,

$$P[W_n = 0] \approx 1 - \frac{\lambda}{\mu}, \quad for\ large\ n.$$

- The expected waiting time satisfies

$$E[W_n] \approx \int_0^\infty wf(w)dw = \frac{\lambda}{\mu(\mu - \lambda)}, \quad for\ large\ n.$$

- For large n, the expected amount of time that packet n stays in the queueing system is approximately equal to $1/(\mu - \lambda)$. To see this, note that the amount of time spent by the packet in the system is the sum of its waiting time and service time, and therefore we have

$$E[waiting\ time] + E[service\ time] \approx \frac{\lambda}{\mu(\mu - \lambda)} + \frac{1}{\mu} = \frac{1}{\mu - \lambda}, \quad for\ large\ n.$$

42.3 Why $S_X(f) \geq 0$

Suppose we have a continuous-time WSS process $X(t)$. I have put off until now trying to explain why the power spectral density $S_X(f)$ takes nonnegative values at all frequencies f. In this section, I show you why this is true. Let f^* be a particular positive frequency at which we want to measure approximately what $S_X(f^*)$ is. Pick a narrowband idealized filter with center frequency f^* and bandwidth Δf. This means the frequency response of the filter is

$$H(f) = \begin{cases} 1, & f^* - \Delta f/2 \leq |f| \leq f^* + \Delta f/2 \\ 0, & elsewhere \end{cases}$$

Pass a realization $x(t)$ of $X(t)$ through this narrowband filter and then measure the filter output power with a power meter:

$$x(t) \rightarrow \boxed{\begin{array}{c}\text{narrowband}\\ \text{filter}\end{array}} \rightarrow \boxed{\begin{array}{c}\text{power}\\ \text{meter}\end{array}} \rightarrow \text{output power}$$

Assuming we have an ergodic process, approximately what will we obtain for our power meter measurement? By the ergodic process assumption, we know that the power measured by the power meter will be approximately equal to

$$2\int_{f^*-\Delta f/2}^{f^*+\Delta f/2} S_X(f)df. \tag{42.8}$$

Since Δf is close to zero, the usual calculus approximation tells us that (42.8) is approximately equal to

$$2(\Delta f)S_X(f^*).$$

Conclusion: The power measured at the output of the narrowband filter, divided by by $2\Delta f$, is approximately equal to $S_X(f^*)$. Since a power measurement will always give a nonegative result, we are forced to conclude that $S_X(f^*)$ must be nonnegative.

Remark. There is another way to prove that $S_X(f) \geq 0$. Back in Section 34.4, we introduced a property of autocorrelation function $R_X(\tau)$ called the *positive semidefiniteness property*. It turns out that $R_X(\tau)$ obeys the positive semidefiniteness property if and only if $S_X(f) \geq 0$ for all f. The positive semidefiniteness property says that the double integral

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(s)h(t)R_X(s-t)dsdt \tag{42.9}$$

is nonnegative for any function $h(t)$ for which the preceding double integral exists. The reader can easily check that the integral (42.9) is the same thing as the quantity

$$h(\tau) * h(-\tau) * R_X(\tau)$$

evaluated at $\tau = 0$, which is the same thing as the integral

$$\int_{-\infty}^{\infty} |H(f)|^2 S_X(f)df. \tag{42.10}$$

Saying that the quantity (42.10) is always nonnegative for every filter $h(t)$ is therefore the same thing as saying that the quantity (42.9) is always nonnegative. Saying that the quantity (42.10) is always nonnegative for every filter $h(t)$ is equivalent to saying that $S_X(f) \geq 0$ at all f. Therefore, saying that $S_X(f) \geq 0$ at all f is equivalent to saying that (42.9) is nonnegative for all $h(t)$, which is the positive semidefiniteness property.

42.4 Proof of Formula $R_Y(\tau) = h(\tau) * h(-\tau) * R_X(\tau)$

Suppose we pass WSS process $X(t)$ through LTI stable filter with impulse response $h(t)$, obtaining WSS filter output process $Y(t)$. Here, I finally prove for you that

$$R_Y(\tau) = h(\tau) * h(-\tau) * R_X(\tau)$$

holds.

To start the proof, we note that since

$$(X(t), Y(t))$$

is a filter input-output pair, so is

$$(Y(0)X(t), Y(0)Y(t)). \tag{42.11}$$

Notice that (42.11) is a random input-output pair. If we observe many realizations of input and the corresponding output, and then average, we obtain by the principle of superposition another input-output pair. If we then take a limit of such averages, we conclude by the law of large numbers that

$$(E[Y(0)X(t)], E[Y(0)Y(t)]) = (E[Y(0)X(t)], R_Y(t))$$

is an input-output pair, and then applying the convolution theorem we see that

$$R_Y(t) = h(t) * E[Y(0)X(t)]. \tag{42.12}$$

Now send random signal $X(-t)$ through the filter with impulse response $h(-t)$, thereby obtaining random output signal $Y(-t)$. That is,

$$(X(-t), Y(-t))$$

is an input-output pair for the filter $h(-t)$. Consequently,

$$(X(0)X(-t), X(0)Y(-t))$$

is also an input-output pair for the filter $h(-t)$. By using the principle of superposition and the law of large numbers as we did in our preceding argument, we conclude that

$$(E[X(0)X(-t)], E[X(0)Y(-t)]) = (R_X(t), E[X(0)Y(-t)])$$

is an input-output pair for the filter $h(-t)$, from which it follows that

$$E[X(0)Y(-t)] = h(-t) * R_X(t). \tag{42.13}$$

Note the following time-invariance property:

$$E[X(0)Y(-t)] = E[X(u)Y(u-t)], \quad for\ all\ t, u.$$

In particular, taking $u = t$, we can say that

$$E[X(0)Y(-t)] = E[X(t)Y(0)]. \tag{42.14}$$

If you stare at the three equations (42.12)-(42.14), you see that equation

$$R_Y(t) = h(t) * h(-t) * R_X(t)$$

pops out!