Chapter 9
Random Processes through Linear Systems

In this chapter we study how random processes behave when they pass through linear time invariant systems. We will restrict ourselves to the class of W.S.S. random processes.

9.1 Review of Linear Systems

LTI System and Convolution

Recall that a linear time invariant system consists of two properties:

- Linearity: If $X_1(t) \rightarrow Y_1(t)$ and $X_2(t) \rightarrow Y_2(t)$, then
  
  $$aX_1(t) + bX_2(t) \rightarrow aY_1(t) + bY_2(t).$$

- Time invariant: If $X(t) \rightarrow Y(t)$, then
  
  $$X(t + \tau) \rightarrow Y(t + \tau).$$

All LTI systems satisfy the convolution property. That is, the output $Y(t)$ is a convolution of the input $X(t)$ with a filter (or kernel) $h(t)$:

$$Y(t) = h(t) * X(t),$$

where $*$ denotes the convolution operation, defined as

$$h(t) * X(t) = \int_{-\infty}^{\infty} h(\tau)X(t - \tau)d\tau.$$

For discrete time signals, convolution is defined as

$$h[n] * X[n] = \sum_{k=-\infty}^{\infty} h[k]X[n - k].$$
Continuous-time and Discrete-time Fourier Transforms

In this chapter we will mainly use continuous-time Fourier transform. For any function $h$ satisfying (See Oppenheim and Willsky)

- it is absolute-integrable function, i.e.,
  \[ \int_{-\infty}^{\infty} |h(t)| dt < \infty, \]
- $h$ has finite number of discontinuity
- $h$ has bounded variation

then the continuous Fourier transform exists and is defined as

\[ H(\omega) \overset{\text{def}}{=} \int_{-\infty}^{\infty} h(t) e^{j\omega t} dt. \]  

(9.2)

For notational simplicity, we write $H = \mathcal{F}(h)$. The inverse Fourier transform is $h = \mathcal{F}^{-1}(H)$.

**Example.** The “DC” value of the Fourier transform is

\[ H(0) = \int_{-\infty}^{\infty} h(t) e^{j0t} dt = \int_{-\infty}^{\infty} h(t) dt. \]

Occasionally we will use the discrete-time Fourier transform. Discrete-time Fourier transform of a function $h$ is defined as

\[ H(e^{j\omega}) \overset{\text{def}}{=} \sum_{n=-\infty}^{\infty} h[n] e^{j\omega n}. \]  

(9.3)

9.2 Mean and Autocorrelation through LTI Systems

**Mean Function through LTI System**

Consider a W.S.S. random process $X(t)$. Since we assume that $X(t)$ is W.S.S., the mean function of $X(t)$ stays a constant, i.e., $\mu_X(t) = \mu_X$.

**Proposition 1.** If $X(t)$ passes through an LTI system to yield $Y(t)$, then the mean function of $Y(t)$ is

\[ \mathbb{E}[Y(t)] = \mu_X \int_{-\infty}^{\infty} h(\tau) d\tau = \mu_X H(0). \]  

(9.4)
Proof.

\[ \mu_Y(t) = \mathbb{E}[Y(t)] = \mathbb{E} \left[ \int_{-\infty}^{\infty} h(\tau)X(t-\tau)d\tau \right] \]
\[ = \int_{-\infty}^{\infty} h(\tau)\mathbb{E}[X(t-\tau)]d\tau \]
\[ = \int_{-\infty}^{\infty} h(\tau)\mu_X d\tau \]
where the last equality holds because \( X(t) \) is W.S.S. so that \( \mathbb{E}[X(t-\tau)] = \mu_X \).

For discrete-time signals, the proposition becomes

\[ \mathbb{E}\left[ Y[n] \right] = \mu_X \sum_{k=-\infty}^{\infty} h[k] = \mu_X H(e^{j0}). \] (9.5)

### Autocorrelation Function through LTI System

**Proposition 2.** If \( X(t) \) passes through an LTI system to yield \( Y(t) \), then the autocorrelation function of \( Y(t) \) is

\[ R_Y(\tau) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(s)h(r)R_X(\tau + s - r)dsdr. \] (9.6)

**Proof.**

\[ R_Y(\tau) = \mathbb{E}[Y(t)Y(t+\tau)] \]
\[ = \mathbb{E} \left[ \int_{-\infty}^{\infty} h(s)X(t-s)ds \int_{-\infty}^{\infty} h(r)X(t+\tau-r)dr \right] \]
\[ \overset{(a)}{=} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(s)h(r)\mathbb{E} [X(t-s)X(t+\tau-r)dsdr] \]
\[ = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(s)h(r)R_X(\tau + s - r)dsdr, \]
where in (a) we assume that integration and expectation are interchangeable.

For discrete-time signals, the above proposition becomes

\[ R_Y[k] = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} h[m]h[n]R_X[k + m - n]. \] (9.7)
Proposition 3. If $X(t)$ passes through an LTI system to yield $Y(t)$, then the power spectral density of $Y(t)$ is
\[ S_Y(\omega) = |H(\omega)|^2 S_X(\omega). \] (9.8)

Proof. By definition, power spectral density $S_Y(\omega)$ is the Fourier transform of the autocorrelation function $R_Y(\omega)$. Therefore,
\[ S_Y(\omega) = \int_{-\infty}^{\infty} R_Y(\tau)e^{-j\omega\tau} \, d\tau \]
\[ = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(s)h(r)R_X(\tau+s-r)dsdr \, e^{-j\omega\tau} \, d\tau \]

Letting $u = \tau + s - r$, we have
\[ S_Y(\omega) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(s)h(r)R_X(u)e^{-j\omega(u+s-r)}dsdr \, du \]
\[ = \int_{-\infty}^{\infty} h(s)e^{j\omega s} \, ds \int_{-\infty}^{\infty} h(r)e^{-j\omega r} \, dr \int_{-\infty}^{\infty} R_X(u)e^{-j\omega u} \, du = \overline{H(\omega)}H(\omega)S_X(\omega), \]
where $\overline{H(\omega)}$ is the complex conjugate of $H(\omega)$.

For discrete-time signals, the above proposition becomes
\[ S_Y(e^{j\omega}) = |H(e^{j\omega})|^2 S_X(e^{j\omega}). \] (9.9)

Example 1: A W.S.S. process $X(t)$ has a correlation function
\[ R_X(\tau) = \text{sinc}(\pi \tau). \]
Suppose that $X(t)$ passes through an LTI system with input/output relationship
\[ 2 \frac{d^2}{dt^2} Y(t) + 2 \frac{d}{dt} Y(t) + 4Y(t) = 3 \frac{d^2}{dt^2} X(t) - 3 \frac{d}{dt} X(t) + 6X(t). \]
Find $R_Y(\tau)$.

Solution: The sinc function has a Fourier transform given by
\[ \text{sinc}(Wt) \leftrightarrow_{\mathcal{F}} \frac{\pi}{W} \text{rect}\left(\frac{\omega}{2W}\right). \]
Therefore, the auto-correlation function is
\[ R_X(\tau) = \text{sinc}(\pi \tau) \leftrightarrow_{\mathcal{F}} \frac{\pi}{\pi} \text{rect}\left(\frac{\omega}{2\pi}\right). \]
By taking the Fourier transform on both sides, we have

\[ S_X(\omega) = \begin{cases} 
1, & -\pi \leq \omega \leq \pi, \\
0, & \text{elsewhere}.
\end{cases} \]

The system response can be found from the differential equation as

\[ H(\omega) = \frac{3(j\omega)^2 - 3(j\omega) + 6}{2(j\omega)^2 + 2(j\omega) + 4} = \frac{3[(2 - \omega^2) - j\omega]}{2[(2 - \omega^2) + j\omega]}. \]

Taking the magnitude square yields

\[ |H(\omega)|^2 = \frac{3[(2 - \omega^2) - j\omega]}{2[(2 - \omega^2) + j\omega]} \cdot \frac{3[(2 - \omega^2) + j\omega]}{2[(2 - \omega^2) - j\omega]} = \frac{9(2 - \omega^2)^2 + \omega^2}{4(2 - \omega^2)^2 + \omega^2} = \frac{9}{4}. \]

Therefore, the output power spectral density is

\[ S_Y(\omega) = |H(\omega)|^2 S_X(\omega) = \frac{9}{4} S_X(\omega). \]

Taking the inverse Fourier transform, we have

\[ R_Y(\tau) = \frac{9}{4} \text{sinc}(\pi \tau). \]

**Example 2:** A random process \( X(t) \) has zero mean and \( R_X(t, s) = \min(t, s) \). Consider a new process \( Y(t) = e^t X(e^{-2t}) \).

(a) Is \( Y(t) \) W.S.S.?

(b) Suppose \( Y(t) \) passes through an LTI system \( \text{yo} \) yield an output \( Z(t) \) according to

\[ \frac{d}{dt} Z(t) + 2Z(t) = \frac{d}{dt} Y(t) + Y(t). \]

Find \( R_Z(\tau) \).

**Solution:**
(a) In order to verify whether \( Y(t) \) is W.S.S., we need to check the mean function and the auto-correlation function. The mean function is

\[
\mathbb{E}[Y(t)] = \mathbb{E} \left[ e^t X(e^{-2t}) \right] \\
= e^t \mathbb{E} \left[ X(e^{-2t}) \right].
\]

Since \( X(t) \) has zero mean, \( \mathbb{E}[X(t)] = 0 \) for all \( t \). This implies that if \( u = e^{-2t} \), then \( \mathbb{E}[X(u)] = 0 \) because \( u \) is just another time instant. Therefore, \( \mathbb{E}[X(e^{-2t})] = 0 \), and hence \( \mathbb{E}[Y(t)] = 0 \).

The auto-correlation can be found as

\[
\mathbb{E}[Y(t + \tau)Y(t)] = \mathbb{E} \left[ e^{t+\tau} X(e^{-2(t+\tau)})e^t X(e^{-2t}) \right] \\
= e^{2t+\tau}\mathbb{E} \left[ X(e^{-2(t+\tau)})X(e^{-2t}) \right] \\
= e^{2t+\tau} R_X(e^{-2(t+\tau)}, e^{-2t}) \\
= e^{2t+\tau} \min(e^{-2(t+\tau)}, e^{-2t}) \\
= e^{2t+\tau} \left\{ \begin{array}{ll}
  e^{-2(t+\tau)}, & \tau \geq 0 \\
  e^{-2t}, & \tau < 0
\end{array} \right.
\]

So, \( R_Y(\tau) = e^{-|\tau|} \). Since \( R_Y(\tau) \) is a function of \( \tau \), \( Y(t) \) is W.S.S.

(b) The system response is given by

\[
H(\omega) = \frac{1 + j\omega}{2 + j\omega}.
\]

The magnitude is therefore

\[
|H(\omega)|^2 = \frac{1 + \omega^2}{4 + \omega^2}.
\]

Hence, the output auto-correlation function is

\[
R_Y(\tau) = e^{-|\tau|} \Longleftrightarrow S_Y(\omega) = \frac{2}{1 + \omega^2},
\]

and

\[
S_Z(\omega) = |H(\omega)|^2 S_Y(\omega) = \frac{1 + \omega^2}{4 + \omega^2} \cdot \frac{2}{1 + \omega^2} = \frac{2}{4 + \omega^2}.
\]

Therefore,

\[
R_Z(\tau) = \frac{1}{2} e^{-2|\tau|}.
\]
9.3 Cross-Correlation through LTI Systems

Before we discuss the cross-correlation function passing through an LTI system, we first define jointly W.S.S. for two arbitrary random processes $X(t)$ and $Y(t)$.

**Jointly W.S.S. Processes**

**Definition 1.** Two random processes $X(t)$ and $Y(t)$ are jointly W.S.S. if

1. $X(t)$ is W.S.S. and $Y(t)$ is W.S.S.
2. $R_{X,Y}(t_1, t_2) = E[X(t_1)Y(t_2)]$ is a function of $t_1 - t_2$.

If $X(t)$ and $Y(t)$ are jointly W.S.S., then we write

$$R_{X,Y}(t_1, t_2) = R_{X,Y}(	au) \overset{\text{def}}{=} E[X(t + \tau)Y(\tau)].$$

The definition $R_{Y,X}(\tau)$ can be seen from the following Lemma.

**Lemma 1.** For any random processes $X(t)$ and $Y(t)$, the cross-correlation $R_{X,Y}(\tau)$ is related to $R_{Y,X}(\tau)$ as

$$R_{X,Y}(\tau) = R_{Y,X}(-\tau). \quad (9.10)$$

**Proof.** Recall the definition of $R_{Y,X}(-\tau) = E[Y(t - \tau)X(t)]$. It holds that

$$R_{Y,X}(-\tau) = E[Y(t - \tau)X(t)] = E[X(t)Y(t - \tau)]$$

$$= E[X(t' + \tau)Y(t')] = R_{X,Y}(\tau),$$

where we substituted $t' = t - \tau$. \[qed\]

**Example 3.** Let $X(t)$ and $N(t)$ be two independent W.S.S. random processes with expectations $E[X(t)] = \mu_x$ and $E[N(t)] = 0$, respectively. Let $Y(t) = X(t) + N(t)$. We want to show that $X(t)$ and $Y(t)$ are jointly W.S.S., and we want to find $R_{X,Y}(\tau)$.

**Solution.**

Before we show the joint W.S.S. property of $X(t)$ and $Y(t)$, we first show that $Y(t)$ is W.S.S.: 

$$E[Y(t)] = E[X(t) + N(t)] = \mu_x$$

$$R_Y(t_1, t_2) = E[(X(t_1) + N(t_1))(X(t_2) + N(t_2))]$$

$$= E[(X(t_1)X(t_2)] + E[(N(t_1)N(t_2)]$$

$$= R_X(t_1 - t_2) + R_N(t_1 - t_2).$$

Thus, $Y(t)$ is W.S.S.
To show that \( X(t) \) and \( Y(t) \) are jointly W.S.S., we need to check the cross-correlation function:

\[
R_{X,Y}(t_1, t_2) = \mathbb{E}[X(t_1)Y(t_2)] \\
= \mathbb{E}[X(t_1)(X(t_2) + N(t_2))] \\
= \mathbb{E}[X(t_1)(X(t_2)) + \mathbb{E}[X(t_1)N(t_2)]] \\
= R_X(t_1, t_2) + \mathbb{E}[X(t_1)]\mathbb{E}[N(t_2)] = R_X(t_1, t_2).
\]

Since \( R_{X,Y}(t_1, t_2) \) is a function of \( t_1 - t_2 \), and since \( X(t) \) and \( Y(t) \) are W.S.S., \( X(t) \) and \( Y(t) \) must be jointly W.S.S.

Finally, to find \( R_{X,Y}(\tau) \), we substitute \( \tau = t_1 - t_2 \) and obtain \( R_{X,Y}(\tau) = R_X(\tau) \).

We next define the cross power spectral density of two jointly W.S.S. processes as the Fourier transform of the cross-correlation function.

**Definition 2.** The cross power spectral density of two jointly W.S.S. processes \( X(t) \) and \( Y(t) \) is defined as

\[
S_{X,Y}(\omega) = \mathcal{F}[R_{X,Y}(\tau)] \\
S_{Y,X}(\omega) = \mathcal{F}[R_{Y,X}(\tau)]
\]

The relationship between \( S_{X,Y} \) and \( S_{Y,X} \) can be seen from the following Lemma.

**Proposition 4.** For two jointly W.S.S. random processes \( X(t) \) and \( Y(t) \), the cross power spectral density satisfies the property that

\[
S_{X,Y}(\omega) = \overline{S_{Y,X}(\omega)},
\]

where \( \overline{\cdot} \) denotes the complex conjugate.

**Proof.** Since \( S_{X,Y}(\omega) = \mathcal{F}[R_{X,Y}(\tau)] \) by definition, it follows that

\[
\mathcal{F}[R_{X,Y}(\tau)] = \int_{-\infty}^{\infty} R_{X,Y}(\tau)e^{-j\omega\tau} d\tau = \int_{-\infty}^{\infty} R_{Y,X}(-\tau)e^{-j\omega\tau} d\tau \\
= \int_{-\infty}^{\infty} R_{X,Y}(\tau')e^{j\omega\tau'} d\tau' = \overline{S_{Y,X}(\omega)}.
\]

© 2017 Stanley Chan. All Rights Reserved.
Cross-correlation Function through LTI System

We now study the special case where \(X(t)\) is the input to an LTI system, and \(Y(t)\) is the output of the LTI system.

**Proposition 5.** If \(X(t)\) passes through an LTI system to yield \(Y(t)\), then the cross-correlation is

\[
R_{Y,X}(\tau) = h(\tau) * R_X(\tau).
\]

**Proof.** Recall the definition of cross-correlation, we have

\[
R_{Y,X}(\tau) = \mathbb{E}[Y(t+\tau)X(t)] = \mathbb{E} \left[ X(t) \int_{-\infty}^{\infty} X(t+\tau-r)h(r)dr \right]
\]

\[
= \int_{-\infty}^{\infty} \mathbb{E}[X(t)X(t+\tau-r)]h(r)dr = \int_{-\infty}^{\infty} R_X(\tau-r)h(r)dr.
\]

\[
\square
\]

**Proposition 6.** If \(X(t)\) passes through an LTI system to yield \(Y(t)\), then the cross power spectral density is

\[
S_{Y,X}(\omega) = H(\omega)S_X(\omega)
\]

\[
S_{X,Y}(\omega) = H(\omega)S_X(\omega)
\]

**Proof.** By taking the Fourier transform on \(R_{Y,X}(\tau)\) we have that \(S_{Y,X}(\omega) = H(\omega)S_X(\omega)\). Since \(R_{X,Y}(\tau) = R_{Y,X}(-\tau)\), it holds that \(S_{X,Y}(\omega) = H(\omega)S_X(\omega)\).

\[
\square
\]

**Example 4.** Let \(X(t)\) be a W.S.S. random process with

\[
R_X(\tau) = e^{-\tau^2/2}, \quad H(\omega) = e^{-\omega^2/2}.
\]

Find \(S_{X,Y}(\omega)\), \(R_{X,Y}(\tau)\), \(S_Y(\omega)\) and \(R_Y(\tau)\).

**Solution.** First, by Fourier transform table we know that \(S_X(\omega) = \sqrt{2\pi}e^{-\omega^2/2}\). Since \(H(\omega) = e^{-\omega^2/2}\), we have

\[
S_{X,Y}(\omega) = H(\omega)S_X(\omega) = \sqrt{2\pi}e^{-\omega^2}.
\]

The cross-correlation function is

\[
R_{X,Y}(\omega) = \mathcal{F}^{-1}\left[\sqrt{2\pi}e^{-\omega^2}\right] = \frac{1}{\sqrt{2}}e^{-\frac{\tau^2}{4}}.
\]
The power spectral density of $Y(t)$ is
\[ S_Y(\omega) = |H(\omega)|^2 S_X(\omega) = \sqrt{2\pi} e^{-\frac{3\omega^2}{2}}. \]
Therefore, the autocorrelation function of $Y(t)$ is
\[ R_Y(\tau) = \mathcal{F}^{-1} \left[ \sqrt{2\pi} e^{-\frac{3\omega^2}{2}} \right] = \frac{1}{\sqrt{3}} e^{-\tau^2/6}. \]

9.4 Optimal Linear Systems

The (discrete-time) optimal linear systems concern about the following problem: Let $X[n]$ be the input signal and $Y[n]$ be the observed output signal. The exact relationship between $X[n]$ and $Y[n]$ is unknown. However, statistical properties of $X[n]$ and $Y[n]$ are known, or at least can be estimated. These properties include $R_X[k]$, $R_Y[k]$ and $R_{X,Y}[k]$. As suggested by its name, an optimal linear system imposes a linear model. Given $Y[n]$, we postulate that the best estimate of $X[n]$, denoted by $\hat{X}[n]$, can be found using a linear combination of the $Y[n]$’s. Depending on how far we want to go in time (past and future), there are three types of estimates:

1. Finite Impulse Response (FIR)
\[ \hat{X}[n] = \sum_{k=n-K+1}^{n} h[n-k]Y[k] \]

2. Wiener Filter
\[ \hat{X}[n] = \sum_{k=-\infty}^{\infty} h[n-k]Y[k] \]

3. Causal Filter
\[ \hat{X}[n] = \sum_{k=0}^{\infty} h[n-k]Y[k] \]

In these three equations, the filter $h[n]$ is the subject of interest. The goal is to find the optimal $h[n]$ such that the error between the estimate $\hat{X}[n]$ and the truth $X[n]$ is minimized, i.e.,
\[ \min_{h[n]} \mathbb{E} \left[ \left( X[n] - \hat{X}[n] \right)^2 \right]. \]
The following sections will describe solutions to each of these models.
A. Finite Impulse Response

The FIR model assumes that

$$
\hat{X}[n] = \sum_{k=n-K+1}^{n} h[n-k]Y[k] = \sum_{i=0}^{K-1} h[i]Y[n-i],
$$

(9.13)

where we substituted $i = n - k$. Our goal is to minimize the mean squared error

$$
\text{minimize}_{h[i]} \mathbb{E}\left[ (X[n] - \hat{X}[n])^2 \right].
$$

(9.14)

**Theorem 1.** The optimal filter $\{h[0], \ldots, h[K-1]\}$ of the FIR system is given by the solution of the matrix equation

$$
\begin{pmatrix}
R_{X,Y}[0] \\
R_{X,Y}[1] \\
\vdots \\
R_{X,Y}[K-1]
\end{pmatrix}
= \begin{pmatrix}
R_{Y}[0] & R_{Y}[1] & \ldots & R_{Y}[K-1] \\
R_{Y}[1] & R_{Y}[0] & R_{Y}[1] & \vdots \\
\vdots & \ddots & \ddots & \vdots \\
R_{Y}[K-1] & \ldots & R_{Y}[1] & R_{Y}[0]
\end{pmatrix}
\begin{pmatrix}
h[0] \\
h[1] \\
\vdots \\
h[K-1]
\end{pmatrix},
$$

(9.15)

which is known as the Yule-Walker Equation.

To prove this result, we first observe that

$$
\mathbb{E}\left[ (X[n] - \hat{X}[n])^2 \right] = \mathbb{E}\left[ \left( X[n] - \sum_{i=0}^{K-1} h[i]Y[n-i] \right)^2 \right].
$$

The first order derivative with respect to $h[i]$ is

$$
\frac{d}{dh[i]} \mathbb{E}\left[ (X[n] - \hat{X}[n])^2 \right] = \frac{d}{dh[i]} \mathbb{E}\left[ \left( X[n] - \sum_{i=0}^{K-1} h[i]Y[n-i] \right)^2 \right]
$$

$$
= \mathbb{E}\left[ 2 \left( X[n] - \sum_{i=0}^{K-1} h[i]Y[n-i] \right) Y[n-i] \right]
$$

Setting to zero, we arrive at the equation

$$
\mathbb{E}\left[ X[n]Y[n-i] \right] = \mathbb{E}\left[ \sum_{j=0}^{K-1} h[j]Y[n-j]Y[n-i] \right]
$$
Recall the definition of $R_{X,Y}$ and $R_Y$, the above equation can be written as

$$R_{X,Y}[i] = \sum_{j=0}^{K-1} h[j] R_Y[j+i], \quad i = 0, 1, \ldots, K - 1$$

More compactly, we write it in the matrix-vector form

$$
\begin{pmatrix}
R_{X,Y}[0] \\
R_{X,Y}[1] \\
\vdots \\
R_{X,Y}[K-1]
\end{pmatrix} = 
\begin{pmatrix}
\vdots & \ddots & \ddots & \vdots \\
R_Y[K-1] & \cdots & R_Y[1] & R_Y[0]
\end{pmatrix}
\begin{pmatrix}
h[0] \\
h[1] \\
\vdots \\
h[K-1]
\end{pmatrix}
$$

(9.16)

This completes the proof.

**Solving Yule-Walker Equation in Practice.** In practice, there are two issues we need to address when solving Yule-Walker equation:

- How to estimate $R_Y[k]$ and $R_{X,Y}[k]$?
- How to solve the linear system of equations?

Estimating $R_Y[k]$ and $R_{X,Y}[k]$ is problem-specific. To illustrate the idea, consider the following example:

**Example.** Assume that the true signal $X[n]$ and the estimated signal $\hat{X}[n]$ are

$$X[n] = Y[n]$$
$$\hat{X}[n] = \sum_{k=0}^{K-1} h[k] Y[n-k].$$

In this case, the auto-correlation and the cross-correlation can be found using

$$R_Y[k] \approx \frac{1}{N} \sum_{n=0}^{N-1} Y[n+k] Y[n]$$

$$R_{XY}[k] \approx \frac{1}{N} \sum_{n=0}^{N-1} X[n+k] Y[n]$$
$$= \frac{1}{N} \sum_{n=0}^{N-1} Y[n+k] Y[n] = R_Y[k].$$

These two results hold because of the W.S.S. assumption and the mean Ergodic theorem, which states that the statistical average is equivalent to the temporal average. With these estimates, the Yule-Walker Equation becomes
\[
\begin{pmatrix}
R_Y[1] & R_Y[0] & \cdots & \vdots \\
\vdots & \ddots & \ddots & \vdots \\
R_Y[K-1] & \cdots & R_Y[1] & R_Y[0]
\end{pmatrix}
\begin{pmatrix}
h[0] \\
h[1] \\
\vdots \\
h[K-1]
\end{pmatrix}
\]

(9.17)

The matrix in Yule-Walker equation is a Toeplitz matrix. There are many known algorithms in solving Toeplitz system. One of the most popular algorithms is called the Levisnon-Durbin algorithm which can be found in advanced random process textbooks.

**Remark:** The condition
\[
\mathbb{E}\left[ (X[n] - \hat{X}[n]) Y[n-i] \right] = 0
\]
is called the orthogonality condition. It is called “orthogonal” because the residue \(X[n] - \hat{X}[n]\) is orthogonal to the measurement \(Y[n-i]\).

## B. Non-Causal filter (Wiener)

In Wiener filter, the estimate is assumed to be
\[
\hat{X}[n] = \sum_{k=-\infty}^{\infty} h[n-k]Y[k] \\
= \sum_{i=-\infty}^{\infty} h[i]Y[n-i],
\]

where we substituted \(i = n-k\). The difference of the Wiener filter with the FIR filter is the limit of the summation, which now goes from \(-\infty\) to \(\infty\). Our goal is to find \(\{h[i]\}\) by minimizing the function
\[
\min_{h[i]} \mathbb{E}\left[ (X[n] - \hat{X}[n])^2 \right].
\]

To solve the minimization problem, we take differentiation of the objective function. This gives us
\[
\frac{d}{dh[i]} \mathbb{E}\left[ \left( X[n] - \sum_{j=-\infty}^{\infty} h[j]Y[n-j] \right)^2 \right] = 2\mathbb{E}\left[ (X[n] - \sum_{j=-\infty}^{\infty} h[j]Y[n-j])Y[n-i] \right].
\]
Setting the right hand side to zero yields
\[
\mathbb{E}[X[n]Y[n-i]] = \mathbb{E}\left[Y[n-i] \sum_{j=-\infty}^{\infty} h[j]Y[n-j]\right]
= \sum_{j=-\infty}^{\infty} h[j]\mathbb{E}[Y[n-i]Y[n-j]].
\]
Using the definition of auto-correlation function, the equation becomes the Wiener Equation
\[
R_{XY}[i] = \sum_{j=-\infty}^{\infty} h[j]R_Y[i-j]. \quad (9.18)
\]
Taking Fourier Transform on both sides yields
\[
S_{XY}(e^{-j\omega}) = H(e^{-j\omega})S_Y(e^{-j\omega}), \quad (9.19)
\]
which implies that the filter coefficients \(\{h[n]\}\) are
\[
h[n] = \mathcal{F}^{-1}\left(\frac{S_{XY}(e^{-j\omega})}{S_Y(e^{-j\omega})}\right). \quad (9.20)
\]
**Continuous-time**: Wiener filter can be extended to continuous-time signals. In this case, the linear relationship is given by the integration
\[
Y(t) = \int_{-\infty}^{\infty} h(\tau)X(t-\tau)d\tau. \quad (9.21)
\]
The Wiener filter is then
\[
h(t) = \mathcal{F}^{-1}\left(\frac{S_{XY}(\omega)}{S_Y(\omega)}\right). \quad (9.22)
\]
**C. Causal filter (Wiener Hopf)**
The Wiener Hopf model assumes that
\[
\hat{X}[n] = \sum_{i=0}^{\infty} h[i]Y[n-i].
\]
Following the same derivation as in the Wiener case, we arrive at the Wiener Hopf Equation by replacing the lower limit to 0.
\[
R_{XY}[i] = \sum_{j=0}^{\infty} h[j]R_Y[i-j]. \quad (9.23)
\]
The difficulty of solving the Wiener Hopf equation is that the limit now starts from zero. Fourier transform is no longer applicable.

The solution of Wiener Hopf is to first consider the special case where \( Y[n] \) is white noise. If \( Y[n] \) is white noise, then the auto-correlation becomes a Dirac delta function, i.e., \( R_Y[i-j] = \delta[i-j] \). Therefore,

\[
R_{XY}[i] = \sum_{j=0}^{\infty} h[j] \delta[i-j] \\
= h[i]. \tag{9.24}
\]

In this case, the optimal filter is \( h[i] = R_{XY}[i] \). To further ensure that \( h[i] \) is causal, we can set

\[
h[i] = \begin{cases} 
R_{XY}[i] = F^{-1}(S_{XY}(e^{-j\omega})), & i \geq 0, \\
0, & i < 0.
\end{cases} \tag{9.25}
\]

Now, suppose that \( Y[n] \) is not white noise. Then, in order to use the white property, we proceed to design a whitening filter \( G(e^{-j\omega}) \) which converts \( Y[n] \) to white noise, and a filter \( H_0(e^{-j\omega}) \) to achieve the desired filter response.

**Design of** \( G(e^{-j\omega}) \). Let \( U[n] \) be the impulse response of the output signal that \( Y[n] \) passes through the whitening filter \( G(e^{-j\omega}) \). Then, the power spectral density of \( U[n] \) satisfies

\[
S_U(e^{-j\omega}) = |G(e^{-j\omega})|^2 S_Y(e^{-j\omega}),
\]

where \( S_Y(e^{-j\omega}) \) is the power spectral density of the input \( Y[n] \).

Since \( G(e^{-j\omega}) \) is defined as a whitening filter, the output \( U[n] \) should have a power spectral density \( S_U(e^{-j\omega}) = 1 \). Consequently, we can write \( S_Y(e^{-j\omega}) \) as

\[
S_Y(e^{-j\omega}) = \frac{1}{G(e^{-j\omega}) G(e^{-j\omega})^*},
\]

where \( G(e^{-j\omega})^* \) is the complex conjugate. The result suggests that the whitening filter \( G(e^{-j\omega}) \) is found by factorizing the power spectral density \( S_Y(e^{-j\omega}) \). The factorization technique is called **spectral factorization**. An interesting property of spectral factorization is that \( G(e^{-j\omega}) \) is **causal**: \( g[i] = 0 \) for \( i < 0 \).

**Design of** \( H_0(e^{-j\omega}) \). The choice of the second filter \( H_0(e^{-j\omega}) \) follows from the fact that if \( U[n] \) is white noise, then the optimal \( H_0(e^{-j\omega}) \) should satisfy the condition

\[
h_0[n] = R_{XU}[n].
\]
Therefore, it remains to determine $R_{XU}[n]$. Since $U[n] = \sum_{j=0}^{\infty} g[j]Y[n - j]$, it follows that

\[
R_{XU}[i] = \mathbb{E} [X[n + i]U[n]]
= \mathbb{E} \left[ X[n + i] \sum_{j=0}^{\infty} g[j]Y[n - j] \right]
= \sum_{j=0}^{\infty} g[j]R_{XY}[i + j].
\]

The cross-correlation $R_{XY}[n]$ can be calculated from the observations. In case where only $S_{XY}(e^{-j\omega})$ is known, we can take (discrete-time) Fourier transform on both sides of the above equation to show that

\[
\sum_{k=-\infty}^{\infty} R_{XU}[k]e^{-j\omega k} = \sum_{n=0}^{\infty} \left( \sum_{k=-\infty}^{\infty} g[n]R_{XY}[k + n] \right) e^{-j\omega k}
= \sum_{n=0}^{\infty} g[n] \left( \sum_{k=-\infty}^{\infty} R_{XY}[k + n]e^{-j\omega k} \right)
= \sum_{n=0}^{\infty} g[n] \left( \sum_{\ell=-\infty}^{\infty} R_{XY}[\ell]e^{-j\omega(\ell - n)} \right)
= \sum_{n=0}^{\infty} g[n]e^{j\omega n} \sum_{\ell=-\infty}^{\infty} R_{XY}[\ell]e^{-j\omega \ell}
= G(e^{-j\omega})^* S_{XY}(e^{-j\omega}),
\]

where we substituted $\ell = k + n$. The last equality holds because $g[n]$ is causal. Therefore, the filter $h_0[n]$ is given by

\[
h_0[n] = \begin{cases} 
\mathcal{F}^{-1} \{ G(e^{-j\omega})^* S_{XY}(e^{-j\omega}) \}, & n \geq 0; \\
0, & n < 0.
\end{cases}
\] (9.26)

The overall Wiener-Hopf filter has the frequency response

\[
H(e^{-j\omega}) = H_0(e^{-j\omega})G(e^{-j\omega}),
\]

where $H_0(e^{-j\omega})$ is the frequency response of $h_0[n]$. 

**Continuous-time**: For continuous time signals, the linear relationship is

\[
Y(t) = \int_{0}^{\infty} h(\tau)X(t - \tau)\,d\tau.
\] (9.27)
The whitening filter $G(\omega)$ is the one that factorizes $S_Y(\omega)$:

\[ S_Y(\omega) = \frac{1}{G(\omega)} \frac{1}{G(\omega)^*}. \]

The second filter $H_0(\omega)$ has impulse response

\[ h_0(t) = \begin{cases} \mathcal{F}^{-1} \{ G(\omega)^* S_{XY}(\omega) \}, & t \geq 0, \\ 0, & t < 0. \end{cases} \]

The overall Wiener-Hopf filter is

\[ H(\omega) = H_0(\omega)G(\omega). \]