WSS/LTI Spectral Theory Notes ©

Keith M. Chugg
University of Arizona
ECE Dept.
ECE 422C – (520) 626-2095
chugg@ece.arizona.edu

November 19, 1995

Contents

Scope and Status

These notes represent a work in progress – more accurately, they are a first-cut at providing the student with as good a text for the second-half of the semester as is available for the first-half. This excellent reference for the first half is primarily the *"Supplemental Notes"* by R.A. Scholtz. I will often refer to these notes, and the other existing handouts written by me (*"Cholesky Factorization," "Transform Theory,"* and *"MMSE Estimation"*). Portions of this text which are to be written or need substantial expansion will be marked by the symbol ♣. The notes are basically my "pad notes" for the second half of the course, or a least as much as I could type in four days. My lecture notes, in turn, are also based largely on the lecture notes and "Old Supplemental Notes" of Prof. Scholtz; so a lot of what you are about to read I learned from him.

At this stage, I view these notes as a sort of hard-copy version of share-ware – use them, and if you find them useful take the time to send me suggestions, comments and corrections.

1 Motivation

You have just completed a fairly intensive study of second moment descriptions of random vectors. The central aspect of this study was the effects of linear systems. You all know by heart that when a random vector $\mathbf{x}(u)$ is operated on by an arbitrary linear system, represented by a matrix $\mathbf{H}$, the output process has second moment description given by

$$\mathbf{m}_{\mathbf{y}} = \mathbf{H}\mathbf{m}_{\mathbf{y}} \quad \mathbf{K}_{\mathbf{y}} = \mathbf{H}\mathbf{K}_{\mathbf{x}}\mathbf{H}^{\dagger} \quad \widetilde{\mathbf{K}}_{\mathbf{y}} = \mathbf{H}\widetilde{\mathbf{K}}_{\mathbf{x}}\mathbf{H}^{\mathrm{t}}. \tag{1}$$

This result is probably more powerful than you realize; notice that this covers the effects of an arbitrary linear system defined on a finite index set.

Consider expanding the results you have learned about random vectors to other index sets, specifically discrete or continuous time. It is most likely that you are not even comfortable characterizing the effects of arbitrary linear systems on *deterministic signals,* let alone random signals. You are probably comfortable with a special class of linear systems, namely Linear Time-Invariant (LTI) systems. These notes concentrate on LTI systems and a specific class of second moment random processes, namely Wide-Sense Stationary (WSS) random processes.

The reason that we concentrate on these special cases is simple and the significance is difficult to overstate. This reason is summarized by the following example.[1] Suppose that, on a finite index set for simplicity, the linear system $\mathbf{H}$ and the input covariance operator, represented by the matrix $\mathbf{K}_{\mathbf{x}}$, have the same orthonormal set of eigenvectors

$$\mathbf{H}\mathbf{e}_f = H(f)\mathbf{e}_f \qquad \mathbf{K}_{\mathbf{x}}\mathbf{e}_f = \lambda_x(f)\mathbf{e}_f \qquad f \in \mathcal{F} = \mathcal{T} = \{0, 1, 2, \ldots N-1\}. \tag{2}$$

In matrix form this is equivalent to

$$\mathbf{H} = \mathbf{E}\mathbf{\Lambda}_{\mathbf{H}}\mathbf{E}^{\dagger} \qquad \mathbf{K}_{\mathbf{x}} = \mathbf{E}\mathbf{\Lambda}_{\mathbf{x}}\mathbf{E}^{\dagger}, \tag{3}$$

[1] This is just the solution to Scholtz problem 11.

where $\mathbf{\Lambda_H} = \mathbf{diag}(H(0), H(1), \ldots H(N-1))$ and $\mathbf{\Lambda_x} = \mathbf{diag}(\lambda_x(0), \lambda_x(1), \ldots \lambda_x(N-1))$. In this case the output covariance matrix is

$$\mathbf{K_y} = \mathbf{H}\mathbf{K_x}\mathbf{H}^\dagger = \mathbf{E}\mathbf{\Lambda_H}\mathbf{\Lambda_x}\mathbf{\Lambda_H^\dagger}\mathbf{E}^\dagger. \tag{4}$$

In other words, $\mathbf{K_y}$ has the same set of orthonormal eigenvectors and eigenvalues given by

$$\lambda_y(f) = |H(f)|^2\lambda_x(f) \quad f \in \mathcal{F}. \tag{5}$$

Also consider the simplifications obtained when the mean vector $\mathbf{m_x}$ is a scalar multiple of one of these eigenfunctions

$$\mathbf{m_x} = m_x\mathbf{e}_f \quad \text{for some } f \in \mathcal{F}. \tag{6}$$

For concreteness, consider the case $\mathbf{m_x} = m_x\mathbf{e}_0$, with m_x a known scalar. It is simple to verify that the output mean is then $\mathbf{m_y} = H(0)m_x\mathbf{e}_0$. This implies that one may keep track of the system effect on the mean by the relation $\mathbf{m_y} = m_y\mathbf{e}_0$ with $m_y = H(0)m_x$. Another simplification which occurs when $\mathbf{m_x} = m_x\mathbf{e}_0$ is that the correlation matrix has the same eigenvectors as the covariance matrix

$$\mathbf{R_x}\mathbf{e}_f = S_x(f)\mathbf{e}_f \quad f \in \mathcal{F}, \tag{7}$$

with corresponding eigenvalues $S_x(f)$ related to those of the covariance operator by

$$S_x(f) = \begin{cases} \lambda_x(0) + |m_x|^2 & f = 0 \\ \lambda_x(f) & f \neq 0. \end{cases} \tag{8}$$

This example illustrates the following principle:

If we restrict our attention to a class of linear systems and covariance (correlation) operators with the same known eigenfunctions and with mean-functions which are scalar multiples of one of these eigenfunctions, then the effects of the system on the second moment description can be fully characterized in the eigenvalue (spectral) domain.

The covariance operator of a WSS process has the same eigenfunctions as LTI systems, namely complex exponentials. We will use the above principle to characterize the effects of LTI systems on the second moment description of WSS processes in the eigenvalue domain, via Fourier transforms. It follows that for the infinite index sets we cover only a special case of the theory developed for a finite index set.

An additional aspect which should be considered for infinite index sets is the topic of *stochastic convergence.* The fundamental question involved in this topic is “What is meant by an infinite linear combination of random variables, or more generally, a limit of a sequence of random variables?” While we motivate the “detour” into stochastic convergence theory by the need to handle these linear systems, it is, in any event, a topic deserving of attention in a class such as EE562a.

This motivates the three topics for the “second-half” – LTI systems, WSS random processes, and stochastic convergence.

$\mathcal{T}$	Description	Addition Operator
$\mathcal{Z}_N = \{0, 1, \ldots N-1\}$	Time-Limited (periodic) Discrete Time	$\oplus = \oplus_N$ Addition Modulo N
$\mathcal{Z} = \{0, \pm 1, \pm 2 \ldots\}$	Discrete Time	$\oplus = +$ Standard Integer Addition
$\mathcal{R} = (-\infty, \infty)$	Continuous Time	$\oplus = +$ Standard Real Addition
$\mathcal{R}_T = [0, T)$	Time-Limited (periodic) Continuous Time	$\oplus = \oplus_T$ Addition Modulo T

Table 1: Index sets considered and the corresponding addition operator.

2 ♣ LTI System Theory

This material is covered in the *Supplemental Notes* (Sections 5.1–5.2) and the *Transform Theory Notes* (Section 2).[2] There are basically two approaches: the abstract space approach and the concrete signal approach. The desired result is the same; to consider several index sets, and in each case show that the eigenfunctions of LTI systems are complex exponentials.

In the abstract space approach, taken in the *Supplemental Notes,* this is shown in one elegant proof. In the abstract notation, a deterministic signal defined on the index set $\mathcal{T}$ is represented as a point in an abstract space

$$\tilde{\boldsymbol{x}} \in \mathcal{S}_{\mathcal{T}} \qquad \Longleftrightarrow \qquad \tilde{\boldsymbol{x}} = \{x(t) : t \in \mathcal{T}\}. \tag{9}$$

A system is represented as a transformation on this space $\mathbb{H}$, so that the system output $y(t)$ is represented by $\tilde{\boldsymbol{y}} \in \mathcal{S}_{\mathcal{T}}$ where $\tilde{\boldsymbol{y}} = \mathbb{H}\tilde{\boldsymbol{x}}$. The system is linear if it has the superposition property

$$\mathbb{H}(\alpha\tilde{\boldsymbol{x}}_1 + \beta\tilde{\boldsymbol{x}}_2) = \alpha\mathbb{H}\tilde{\boldsymbol{x}}_1 + \beta\mathbb{H}\tilde{\boldsymbol{x}}_2. \tag{10}$$

Four index sets will be discussed, these are summarized in Table 1. For the general case we will denote addition and subtraction by $\oplus$ and $\ominus$, respectively. The specific form of the addition operator for each choice of $\mathcal{T}$ is defined in order to ensure that the index set is closed under addition. Since for the bounded index sets $\mathcal{Z}_N$ and $\mathcal{R}_T$ the addition is modulo, it is often helpful to think of signals defined on these index sets as periodic.

The system is *time invariant* if it commutes with the shift operator

$$\mathbb{H} \text{ is time invariant} \qquad \Longleftrightarrow \qquad \mathbb{H}\mathbb{T}_\tau\tilde{\boldsymbol{x}} = \mathbb{T}_\tau\mathbb{H}\tilde{\boldsymbol{x}} \quad \forall\, \tau \in \mathcal{T},\ \tilde{\boldsymbol{x}} \in \mathcal{S}_{\mathcal{T}}, \tag{11}$$

where the shift operator is defined as

$$\tilde{\boldsymbol{y}} = \mathbb{T}_\tau\tilde{\boldsymbol{x}} \quad \longleftrightarrow \quad y(t) = x(t \oplus \tau). \tag{12}$$

The *Supplemental Notes* use this abstract notation to prove the following two points

[2] Be aware that everything in this section, and the related portions of the *Supplemental Notes* and *Transform Theory Notes*, concern *deterministic signal processing.* It is a general statement of the material from a standard “signals and systems” undergraduate level class.

$\mathcal{T}$	Frequency Set $\mathcal{F}$	Eigenfunction Expansion
$\mathcal{Z}_N$	$\mathcal{F} = \{k/N : k = 0, 1, \ldots N-1\}$ (or any N consecutive points)	Discrete Fourier Transform (periodic)
$\mathcal{Z}$	$\mathcal{F} = [-1/2, 1/2)$ (or any other interval of length 1)	Discrete Time Fourier Transform (periodic)
$\mathcal{R}$	$\mathcal{F} = \mathcal{R}$	Fourier Transform
$\mathcal{R}_T$	$\mathcal{F} = \{k/T : k \in \mathcal{Z}\}$	Fourier Series

Table 2: Index sets considered and the corresponding frequency sets.

- The eigenfunctions of the shift operator are the complex exponentials $\tilde{\boldsymbol{e}}_f \in \mathcal{S}_{\mathcal{T}}$, for $f \in \mathcal{F}$, defined as $\tilde{\boldsymbol{e}}_f = \{\exp(j2\pi f t) : t \in \tau\}$.
- LTI systems and the shift operator have the same eigenfunctions, namely $\tilde{\boldsymbol{e}}_f$.

The second result is not surprising – recall from linear algebra that two matrices which commute have the same eigenvectors. These two results imply that the frequency set $\mathcal{F}$ associated with a given index set $\mathcal{T}$ can be found by determining the values which provide true eigenfunctions for the shift operator.

The approach taken in the *Transform Theory Notes* is more concrete, but the results are the same. For each of the four index sets considered, the eigenfunctions are shown to be complex exponentials and $\mathcal{F}$ is described. The results obtained in the *Supplemental Notes* and the *Transform Theory Notes* are summarized in Table 2.

3 Wide-Sense Stationary Random Processes

A WSS random process is one for which the second moment description does not change with time. Specifically, a second moment random process $x(u,t)$ is Wide-Sense Stationary on the index set $\mathcal{T}$ if $x(u,t) \overset{\text{ws}}{=} x(u, t \oplus \tau)$ for all $\tau \in \mathcal{T}$.

Recall that two random processes $x(u,t)$ and $y(u,t)$ are *wide-sense equivalent* (denoted $x(u,t) \overset{\text{ws}}{=} y(u,t)$) when they have the same mean and covariance function. Therefore, the conditions for $x(u,t)$ to be WSS are

$$m_x(t) = m_x(t \oplus \tau) \quad \forall\, t, \tau \in \mathcal{T} \tag{13a}$$

$$K_x(t_1, t_2) = K_x(t_1 \oplus \tau, t_2 \oplus \tau) \quad \forall\, t_1, t_2, \tau \in \mathcal{T}. \tag{13b}$$

The condition in (13a) implies that $m_x(t) = m_x \;\forall\; t \in \mathcal{T}$ – i.e., the mean function is a constant for all values of t. The condition in (13b) means that the covariance function only depends on the difference between t_1 and t_2. Consider $\mathcal{T} = \mathcal{R}$ for concreteness; (13b) requires that

$$K_x(3,1) = K_x(5,3) = K_x(12.2, 10.2) = K_x(2,0). \tag{14}$$

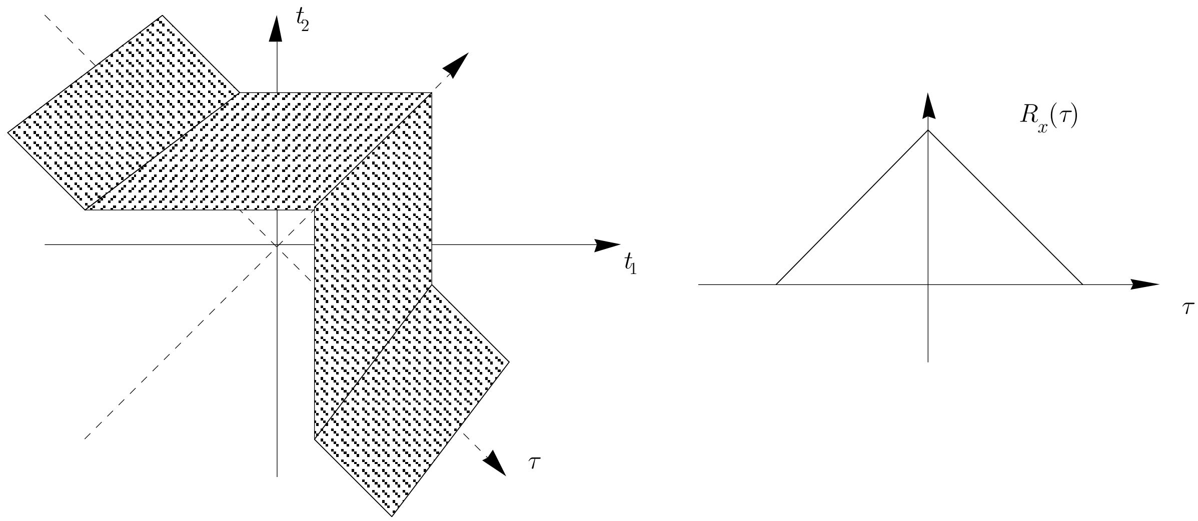

Figure 1: An example WSS correlation function for $\mathcal{T} = \mathcal{R}$.

For a WSS process $x(u,t)$ we have $K_x(t_1,t_2) = K_x(t_1 \ominus t_2, 0)$. A natural convention is then to drop the second argument "0," and use the notation $K_x(t_1,t_2) = K_x(t_1 \ominus t_2) = K_x(\tau)$, where $\tau = t_1 \ominus t_2$. Notice that this is similar to the convention used for LTI systems. A general linear system is characterized by its time-varying impulse response $h(t_1,t_2)$. For an LTI system this impulse response is only a function of the delta-application and output-observation times: $h(t_1,t_2) = h(t_1 \ominus t_2, 0) = h(t)$, where $t = t_1 \ominus t_2$.[3]

A necessary and sufficient condition for $x(u,t)$ to be WSS is that $m_x(t) = m_x$ and $K_x(t_1,t_2) = K_x(t_1 \ominus t_2)$.

A contour plot for a WSS covariance function is shown in Figure 1.

A few facts regarding WSS processes (which that you should verify) are listed below

- $R_x(t_1,t_2) = R_x(t_1 \ominus t_2)$, in fact $R_x(\tau) = K_x(\tau) + |m_x|^2$.
- $R_x(0) = \mathbb{E}\left\{|x(u,t)|^2\right\} \geq 0$ and $K_x(0) = \text{var}\left[x(u,t)\right] \geq 0$ (i.e., real and non-negative).
- Hermitian Symmetry: $R_x(\tau) = R_x^*(-\tau)$ (same for $K_x(\tau)$).
- Non-Negative Definite: see properties of Power Spectral Density.
- Cauchy-Schwartz: $|R_x(\tau)| \leq R_x(0) \quad \forall\, \tau \in \mathcal{T}$.

[3] I know of no good reason why the convention in LTI theory is to use t for the time difference, while for WSS theory τ is usually used.

3.1 Concepts Related to WSS

Three concepts related to Wise-Sense Stationarity are explained in this section: (i) Fully-WSS, (ii) jointly-WSS, and (iii) Strict Stationary.

A complex random process $z(u,t)$ is said to be Fully-WSS if $z(u,t) \stackrel{\text{fws}}{=} z(u, t \oplus \tau) \;\; \forall\, \tau \in \mathcal{T}$. A necessary and sufficient condition for $z(u,t)$ to be Fully-WSS is that $z(u,t)$ is WSS and $\widetilde{K}_z(t_1, t_2) = \widetilde{K}_z(t_1 \ominus t_2)$. In the common case of circular complex processes (i.e., $\widetilde{K}_z(t_1, t_2) = 0$), WSS and Fully-WSS is equivalent.

Two random processes $x(u,t)$ and $y(u,t)$ are Jointly-WSS if they are each WSS and $K_{xy}(t_1, t_2) = K_{xy}(t_1 \ominus t_2)$. If a complex process is denoted by $z(u,t) = x(u,t) + jy(u,t)$, where $x(u,t)$ and $y(u,t)$ are real random processes, then

$$z(u,t) \text{ is Fully-WSS} \qquad \Longleftrightarrow \qquad x(u,t) \text{ and } y(u,t) \text{ are Jointly-WSS.} \tag{15}$$

One may also encounter the term Jointly-Fully-WSS processes, which has the obvious meaning.

Wide-Sense Stationarity is a weakened version of Strict-Sense Stationarity (also referred to as simply "Stationarity"). A random process $x(u,t)$ is said to be (strictly) stationary if its complete statistical description does not change with time. Mathematically, for all integer $N > 0$ and choices of $t_1, t_2, \ldots t_N \; \in \mathcal{T}$ we must have

$$F_{x(u,t_1),x(u,t_2),\ldots x(u,t_N)}(z_1, z_2, \ldots z_N) = F_{x(u,t_1 \oplus \tau),x(u,t_2 \oplus \tau),\ldots x(u,t_N \oplus \tau)}(z_1, z_2, \ldots z_N) \quad \forall\, \tau \in \mathcal{T}. \tag{16}$$

The following facts are simple results from this these definitions

- Stationarity in the strict sense $\Rightarrow$ stationarity in the wide-sense and Fully-WSS (converse is untrue in general!).
- For a real Gaussian process Strict-Stationarity and WS-Stationarity are equivalent.
- For a complex Gaussian process Strict-Stationarity and FWS-Stationarity are equivalent.

4 WSS/LTI Connection – Abstract Version

It has already been claimed that WSS processes and LTI systems are related in a very special manner. It is shown in this section that when a WSS process is input into an LTI system, the output is also WSS. This development will be repeated in concrete notation for each of the index sets in the following sections.

A random process defined on the index set $\mathcal{T}$ can be viewed as a random element of $\mathcal{S}_{\mathcal{T}}$, and denoted by

$$\tilde{\boldsymbol{x}}(u) \in \mathcal{S}_{\mathcal{T}} \qquad \Longleftrightarrow \qquad \tilde{\boldsymbol{x}}(u) = \{x(u,t) : t \in \mathcal{T}\}. \tag{17}$$

Let $y(u,t)$ be the output of an LTI system driven by $x(u,t)$, then abstractly, $\tilde{\boldsymbol{y}}(u) = \mathbb{H}\tilde{\boldsymbol{x}}(u)$. What happens when this output is operated on by the shift operator?

$$\begin{aligned}
\mathbb{T}_\tau \tilde{\boldsymbol{y}}(u) &= \mathbb{T}_\tau(\mathbb{H}\tilde{\boldsymbol{x}}(u)) && && (18)\\
&= \mathbb{H}(\mathbb{T}_\tau \tilde{\boldsymbol{x}}(u)) && (\mathbb{H} \text{ is TI}) && (19)\\
&= \mathbb{H}(\stackrel{\text{ws}}{=} \tilde{\boldsymbol{x}}(u)) && (\tilde{\boldsymbol{x}}(u) \text{ is WSS}) && (20)\\
&\stackrel{\text{ws}}{=} \mathbb{H}\tilde{\boldsymbol{x}}(u) = \tilde{\boldsymbol{y}}(u). && (\mathbb{H} \text{ is LTI}) && (21)
\end{aligned}$$

This means that the output is WS equivalent to a shifted version of itself, or that $y(u,t)$ is WSS. There are two sticky points in declaring this a valid proof

1. As mentioned earlier, if $\mathbb{H}$ involves an infinite linear combination, the definition of the output is not clear at this point – i.e., there are issues of stochastic convergence.

2. In the last step, it has been assumed that WS equivalence is preserved by an LTI system. This implies that the second moment description of the output of a linear system depends only on the second moment description of the input process. We omit a proof of this property, but your intuition from the first-half of the course should convince you that it is reasonable.

4.1 The Covariance Operator

It is useful to maintain a dual concept of the covariance function. Primarily, it defines the covariance of the random variables drawn from the process – it is a measure of how much $x(u,t_1)$ "looks like" $x(u,t_2)$, or how fast the process changes with time. The secondary interpretation is to view the deterministic covariance function as a linear operator. This second interpretation may not seem logical, but recall how much was learned about a random vector and the effects of a linear system by studying the properties of the covariance matrix. For example, finding the eigenvectors of the covariance matrix (operator) allows one to determine directional preference.

The eigenfunctions and the corresponding eigenvalues (spectrum) of the covariance operator play an important role in the following material. The deterministic, abstract linear operator defined by the covariance function $K(t_1,t_2)$ will be denoted by $\mathbb{K}$ – on some occasions, we will work with the correlation operator $\mathbb{R}$ instead.

5 WSS/LTI Processing on $\mathcal{T} = \mathcal{Z}_N$

There are three properties of the index set $\mathcal{T} = \mathcal{Z}_N$ which distinguish it from the other three

- $\mathcal{Z}_N$ is finite, therefore convergence issues do not arise.

- We already know how to handle all linear systems and second moment random vectors.

- All random variables in the process can be listed and used as a notation.

For these reasons, this will serve as our "bridge" from the first-half of the course to the second-half. Three notations can be used, the abstract, and "signal" and "vector" concrete notations. The first steps are to apply what is known about LTI systems and WSS processes to this special case.

5.1 LTI System Theory

The input/output relation for a linear system defined on $\mathcal{Z}_N$ can be written as[4]

$$\tilde{\boldsymbol{y}} = \mathbb{H}\tilde{\boldsymbol{x}} \tag{22a}$$

$$y(n_1) = \sum_{n_2=0}^{N-1} h(n_1, n_2)x(n_2) \qquad n_1 \in \mathcal{Z}_N \tag{22b}$$

$$\underbrace{\begin{bmatrix} y(0) \\ y(1) \\ \vdots \\ y(N-1) \end{bmatrix}}_{\mathbf{y}} = \underbrace{\begin{bmatrix} h(0,0) & h(0,1) & \cdots & h(0,N-1) \\ h(1,0) & h(1,1) & \cdots & h(1,N-1) \\ \vdots & \vdots & \ddots & \vdots \\ h(N-1,0) & h(N-1,1) & \cdots & h(N-1,N-1) \end{bmatrix}}_{\mathbf{H}} \underbrace{\begin{bmatrix} x(0) \\ x(1) \\ \vdots \\ x(N-1) \end{bmatrix}}_{\mathbf{x}} \tag{22c}$$

where each of the expressions (22) are equivalent. The abstract, signal and vector notations are defined in (22a), (22b), and (22c) respectively.

What is the restriction on the matrix $\mathbf{H}$ if $\mathbb{H}$ is an LTI system? To answer this, consider what the output is when a Kronecker delta function is applied at time $n = 0$

$$\mathbf{H} \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} h(0,0) \\ h(1,0) \\ h(2,0) \\ \vdots \\ h(N-1,0) \end{bmatrix}. \tag{23}$$

In other words, the system impulse response is the first column of $\mathbf{H}$. Since $\mathbb{H}$ is LTI, if a delta function is applied at time $n = 1$ the output must be a (circularly) shifted version of the impulse response

$$\mathbf{H} \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} h(0,1) \\ h(1,1) \\ h(2,1) \\ \vdots \\ h(N-1,1) \end{bmatrix} = \begin{bmatrix} h(N-1,0) \\ h(0,0) \\ h(1,0) \\ \vdots \\ h(N-2,0) \end{bmatrix}. \tag{24}$$

[4] For discrete time n will be used for the "t" variable to emphasize that it must take only integer values.

It follows that, for an LTI system, $h(n_1, n_2) = h(n_1 \ominus_N n_2, 0)$. Analogous to the convention for WSS theory, we will drop the "0" argument for simplicity. Carrying out this argument on the other columns results in the following *circulant* matrix structure

$$\mathbf{H} = \begin{bmatrix} h(0) & h(N-1) & h(N-2) & \cdots & h(1) \\ h(1) & h(0) & h(N-1) & \cdots & h(2) \\ h(2) & h(1) & h(0) & \cdots & h(3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ h(N-1) & h(N-2) & h(N-3) & \cdots & h(0) \end{bmatrix}. \tag{25}$$

The notation adopted in (25) takes into account the mod N addition; for example

$$h(1 \ominus_N N) = h(N \ominus_N 1) = h(N-1). \tag{26}$$

Circular matrices are a special class of Toeplitz matrices, which have an (i, j) element which depends only on $i - j$ (i.e., regular integer subtraction).

Thus, multiplication by a circulant matrix characterizes the effects of a linear system in the vector notation. In the signal notation, the circulant matrix multiplication corresponds to *circular convolution*

$$y(n) = \sum_{i=0}^{N-1} h(n \ominus_N i) x(i) = h(n) \circledast_N x(n) \qquad n \in \mathcal{Z}_N \qquad \text{(Circular Convolution)} \tag{27}$$

This can be considered a book-keeping technique, since the class of circulant matrices are closed under multiplication (*WHY?*) and a circulant matrix are completely defined by its first column.

The eigensignals for LTI systems are known to be complex exponentials — in the three notations:

$$\mathbb{H}\tilde{\boldsymbol{e}}_k = H(k)\tilde{\boldsymbol{e}}_k \qquad \text{(Abstract Notation)} \tag{28}$$

$$h(n) \circledast_N e_k(n) = H(k) e_k(n) \qquad n \in \mathcal{Z}_N \qquad \text{(Signal Notation)} \tag{29}$$

$$\mathbf{H}\mathbf{e}_k = H(k)\mathbf{e}_k, \qquad \text{(Vector Notation)} \tag{30}$$

which holds for $k \in \mathcal{Z}_n$, or all integers if the periodic interpretation is taken. The specific form of the orthonormal eigen-signals/vectors are

$$e_k(n) = \frac{1}{\sqrt{N}} e^{j2\pi \frac{k}{N} n} \qquad n \in \mathcal{Z}_N \qquad \text{(Signal Notation)} \tag{31}$$

$$\mathbf{e}_k = \frac{1}{\sqrt{N}} \begin{bmatrix} 1 \\ e^{j2\pi \frac{k}{N}} \\ e^{j2\pi \frac{2k}{N}} \\ \vdots \\ e^{j2\pi \frac{(N-1)k}{N}} \end{bmatrix} \qquad \text{(Vector Notation).} \tag{32}$$

As shown in the *Transform Theory Notes*, the eigenvalues of the system $\{H(k)\}_{k=0}^{N-1}$ are the *Discrete Fourier Transform* of the system impulse response

$$H(k) = \mathbb{DFT}\{h(n)\} = \sum_{n=0}^{N-1} h(n) e^{-j2\pi\frac{k}{N}n} \quad k \in \mathcal{Z}_N. \tag{33}$$

It may seem surprising that we have written down an orthonormal set of eigenvectors for any circulant matrix. However, this is equivalent to stating that complex exponentials are the eigenfunctions of LTI systems. The importance of DFT theory in deterministic signal processing is that when the input and output signals are expanded in terms of these eigensignals (i.e., take the DFT), the effects of the LTI system are characterized by multiplication by $H(k)$ – see the *Transform Theory Notes* for the details.

5.2 WSS Processes

In this section the structure of the covariance matrix for WSS processes on $\mathcal{T} = \mathcal{Z}_N$ is described. If $x(u,n)$ is a WSS process on $\mathcal{Z}_N$, then by definition[5]

$$K_x(n_1, n_2) = K_x(m) \qquad m = n_1 \ominus_N n_2 \in \mathcal{Z}_N \tag{34}$$

$$m_x(n) = m_x \qquad n \in \mathcal{Z}_N. \tag{35}$$

In the vector notation, this implies

$$\mathbf{m_x} = m_x \mathbf{1} = m_x(\sqrt{N}\mathbf{e}_0) \tag{36}$$

$$\mathbf{K_x} = \begin{bmatrix} K_x(0) & K_x(N-1) & K_x(N-2) & \cdots & K_x(1) \\ K_x(1) & K_x(0) & K_x(N-1) & \cdots & K_x(2) \\ K_x(2) & K_x(1) & K_x(0) & \cdots & K_x(3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ K_x(N-1) & K_x(N-2) & K_x(N-3) & \cdots & K_x(0) \end{bmatrix}, \tag{37}$$

where $\mathbf{e}_0$ is the $k = 0$ eigenvector of circulant matrices (i.e., all components are $1/\sqrt{N}$). It follows that a WSS process on $\mathcal{Z}_N$ has a circulant covariance matrix and a mean vector which is a multiple of an eigenvector of $\mathbf{K_x}$. Also, since any covariance matrix is Hermitian Symmetric (HS), Non-Negative Definite (NND), the covariance matrix of a WSS process belongs to a special class of circulant matrices. For example, from (37) it is clear that $K_x(N-1) = K_x^*(1)$.

[5] For discrete time we will use $m = n_1 \ominus n_2$ and reserve $\tau = t_1 \ominus t_2$ for general discussions and continuous time.

Example for $N = 4$: A WSS process on $\mathcal{Z}_4$ has the following covariance matrix form

$$\mathbf{K_x} = \begin{bmatrix} K_x(0) & K_x(3) & K_x(2) & K_x(1) \\ K_x(1) & K_x(0) & K_x(3) & K_x(2) \\ K_x(2) & K_x(1) & K_x(0) & K_x(3) \\ K_x(3) & K_x(2) & K_x(1) & K_x(0) \end{bmatrix} \tag{38}$$

$$= \begin{bmatrix} K_x(0) & K_x^*(1) & K_x(2) & K_x(1) \\ K_x(1) & K_x(0) & K_x^*(1) & K_x(2) \\ K_x(2) & K_x(1) & K_x(0) & K_x^*(1) \\ K_x^*(1) & K_x(2) & K_x(1) & K_x(0) \end{bmatrix}. \tag{39}$$

Notice, in this case $K_x(3) = K_x^*(1)$ and $K_x(2)$ is real. It follows that the covariance function of a WSS process on $\mathcal{Z}_N$ is defined by less than N numbers (*Exercise: How many?*).

5.2.1 The Covariance Operator

When the covariance function is thought of as a linear operator, as suggested in Section 4.1, its action is defined as follows

$$\tilde{\boldsymbol{w}} = \mathbb{K}_x \tilde{\boldsymbol{v}} \qquad \text{(Abstract Notation)} \tag{40}$$

$$w(n_2) = \sum_{n_2=0}^{N-1} K(n_2, n_1) v(n_1) \quad n_2 \in \mathcal{Z}_N \qquad \text{(Signal Notation)} \tag{41}$$

$$\mathbf{w} = \mathbf{K_x v}. \qquad \text{(Vector Notation)} \tag{42}$$

It follows that if the process $x(u, n)$ is WSS, then $\mathbb{K}_x$ represents a LTI operator. Therefore, its eigenvectors are known, and the eigenvalues may be found via the DFT. Specifically, the eigenvalues of $\mathbf{K_x}$ are

$$\lambda_x(k) = \mathbb{DFT}\{K_x(m)\} = \sum_{m=0}^{N-1} K_x(m) e^{-j2\pi \frac{k}{N} m} \quad k \in \mathcal{Z}_N. \tag{43}$$

Since the mean of a WSS process is a multiple of the $k = 0$ eigenvector, we also have that the eigenvalues of the correlation matrix are

$$S_x(k) = \mathbb{DFT}\{R_x(m)\} = \sum_{m=0}^{N-1} R_x(m) e^{-j2\pi \frac{k}{N} m} \quad k \in \mathcal{Z}_N \tag{44}$$

$$= \begin{cases} \lambda_x(0) + N|m_x|^2 & k = 0 \\ \lambda_x(k) & k = 1, 2, \dots N-1. \end{cases} \tag{45}$$

Example $N = 2$: Consider a real WSS process on $\mathcal{Z}_2$ with covariance matrix given by

$$\mathbf{K_x} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}. \tag{46}$$

This matrix came up often in the first-half of the course (e.g., see the Chugg self-test), the eigenvectors and eigenvalues are known to be

$$\mathbf{e}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad \lambda_1 = 3 \tag{47a}$$

$$\mathbf{e}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix} \qquad \lambda_2 = 1. \tag{47b}$$

With the additional knowledge about circulant matrices, this can be seen as a special case. In fact, any WSS process on $\mathcal{Z}_N$ has a covariance matrix of the form

$$\mathbf{K_x} = \begin{bmatrix} K_x(0) & K_x(1) \\ K_x(1) & K_x(0) \end{bmatrix}, \tag{48}$$

with the same eigenvectors as in (47), and eigenvalues given by the 2-point DFT

$$\lambda_x(0) = K_x(0) + K_x(1) \tag{49}$$

$$\lambda_x(1) = K_x(0) - K_x(1). \tag{50}$$

An interesting fact can then be stated: since the eigenvalues of a HS-NND matrix are non-negative, the DFT of $K_x(m)$ must be real and non-negative! The converse is true as well; if $\{\lambda(k)\}_{k=0}^{N-1}$ is a real non-negative sequence, then $K(m) = \mathbb{DFT}^{-1}\{\lambda(k)\}$ is a valid WSS covariance function on $\mathcal{Z}_N$.

5.3 WSS/LTI Spectral Relationship

We have now demonstrated a concrete example of the type discussed in Section 1, namely a class of processes (WSS on $\mathcal{Z}_N$) and linear systems (LTI on $\mathcal{Z}_N$) for which the mean is a system eigenvector and the covariance has the same eigenvectors as the system.

It follows from the principle outlined in Section 1, that if the input to an LTI system $\mathbb{H}$ is a WSS process $x(u,n)$, the output $y(u,n)$ is also WSS and the following holds

$$m_y = H(0)m_x y \tag{51}$$

$$\lambda_y(k) = |H(k)|^2 \lambda_x(k) \quad k \in \mathcal{Z}_N \tag{52}$$

$$S_y(k) = |H(k)|^2 S_x(k) \quad k \in \mathcal{Z}_N. \tag{53}$$

The covariance function of the output process $y(u,n)$ can then be obtained by taking the inverse DFT

$$K_y(m) = \mathbb{DFT}^{-1}\{\lambda_y(k)\} = \frac{1}{N}\sum_{k=0}^{N-1} \lambda_y(k) e^{j2\pi\frac{k}{N}m} \quad m \in \mathcal{Z}_N. \tag{54}$$

Example $N = 2$**:** Consider what occurs when the $N = 2$ real WSS random process $x(u, n)$ with covariance matrix given in (46) and with $m_x = 4$ is passed through the LTI system represented by

$$\mathbf{H} = \begin{bmatrix} -3 & 6 \\ 6 & -3 \end{bmatrix}. \tag{55}$$

The 2-point DFT of $h(n)$ is

$$H(0) = -3 + 6 = 3 \tag{56}$$
$$H(1) = -3 - 6 = -9. \tag{57}$$

Note that since $\mathbf{H}$ is not NND, it has a negative eigenvalue. According to the theory, the spectrum of the output covariance function is

$$\lambda_y(0) = |H(0)|^2 \lambda_x(0) = 27 \tag{58}$$
$$\lambda_y(1) = |H(1)|^2 \lambda_x(1) = 81. \tag{59}$$

Taking the inverse 2-point DFT yields the output covariance

$$K_y(0) = (27 + 81)/2 = 54 \tag{60}$$
$$K_y(1) = (27 - 81)/2 = -27. \tag{61}$$

The mean of the output process is $m_y = H(0)m_x = 3m_x = 12$. It follows that in vector notation we have

$$\mathbf{K_y} = \begin{bmatrix} 54 & -27 \\ -27 & 54 \end{bmatrix} \qquad \mathbf{m_y} = \begin{bmatrix} 12 \\ 12 \end{bmatrix}. \tag{62}$$

It is straightforward to check these results using the techniques of the first-half of the class – i.e.,

$$\mathbf{K_y} = \mathbf{H}\mathbf{K_x}\mathbf{H}^{\mathrm{t}} = \begin{bmatrix} -3 & 6 \\ 6 & -3 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} -3 & 6 \\ 6 & -3 \end{bmatrix} \tag{63}$$
$$\mathbf{m_y} = \mathbf{H}\mathbf{m_x} = \begin{bmatrix} -3 & 6 \\ 6 & -3 \end{bmatrix} \begin{bmatrix} 4 \\ 4 \end{bmatrix}. \tag{64}$$

The last example raises an interesting question: "Why should we use this fancy DFT-based trick for computing $\mathbf{H}\mathbf{K_x}\mathbf{H}^{\mathrm{t}}$, if we already know how the method for direct computation?" The answer is simple; consider a case where $N = 1024$. For large values of N using the spectral technique, along with the computational *Fast Fourier Transform (FFT)* algorithm, will be orders of magnitude less computationally intensive than the matrix multiplication technique. Also, for the other index sets the matrix technique is not a practical option, so this builds an intuitive bridge.

5.3.1 An Example Application: Affine MMSE Estimation

Consider the effects of this WSS/LTI theory on the Affine estimation problem for $\mathcal{T} = \mathcal{Z}_N$. Recall that the estimate of the desirable $\mathbf{z}(u)$ from the observation $\mathbf{x}(u)$ is

$$\hat{\mathbf{z}}(u) = \mathbf{G}_{\text{opt}}(\mathbf{x}(u) - \mathbf{m_x}) + \mathbf{m_z} = \mathbf{K_{zx}}\mathbf{K_x}^{-1}(\mathbf{x}(u) - \mathbf{m_x}) + \mathbf{m_z}. \tag{65}$$

Now consider the result when $x(u,n)$ and $z(u,n)$ are Jointly-WSS; that is $\mathbf{K_x}$, $\mathbf{K_z}$, and $\mathbf{K_{zx}}$ are all $(N \times N)$ circulant matrices and the means are $\mathbf{m_x} = m_x\mathbf{1}$ and $\mathbf{m_z} = m_z\mathbf{1}$. this will be the case, for example, when $x(u,n)$ and $z(u,n)$ are both WSS and are related by $\mathbf{x}(u) = \mathbf{H}\mathbf{z}(u) + \mathbf{n}(u)$, with $\mathbf{H}$ circulant and $\mathbf{n}(u)$ WSS. It follows that $\mathbf{G}_{\text{opt}}$ is circulant since

$$\begin{aligned}\mathbf{G}_{\text{opt}} &= \mathbf{K_{zx}}\mathbf{K_x}^{-1} && (66)\\ &= \mathbf{K_{zx}}\mathbf{E}\mathbf{\Lambda_x}^{-1}\mathbf{E}^{\dagger} && (67)\\ &= \mathbf{E}\mathbf{\Lambda_{zx}}\mathbf{\Lambda_x}^{-1}\mathbf{E}^{\dagger}, && (68)\end{aligned}$$

where $\mathbf{E}$ is the matrix of the circulant eigenvectors defined in (32), and $\mathbf{\Lambda}_{zx}$ is the diagonal matrix of eigenvalues of $K_{zx}(m)$

$$\lambda_{zx}(k) = \mathbb{DFT}\{K_{zx}(m)\}. \tag{69}$$

The spectrum of the optimal $\mathbb{G}$ is then defined by $\mathbf{\Lambda_G} = \mathbf{\Lambda_{zx}}\mathbf{\Lambda_x}^{-1}$, which has the diagonal elements

$$G_{\text{opt}}(k) = \frac{\lambda_{zx}(k)}{\lambda_x(k)} \qquad k \in \mathcal{Z}_N. \tag{70}$$

The above development, coupled with the structure of the mean vectors implies that the estimator may be obtained by

$$\hat{z}(u,n) = g_{\text{opt}}(n) \circledast_N (x(u,n) - m_x) + m_z, \tag{71}$$

where the LTI filter impulse response is $g_{\text{opt}}(n) = \mathbb{DFT}^{-1}\{G_{\text{opt}}(k)\}$. Thus, the affine estimation problem in the WSS/LTI environment reduces to an LTI filter design problem which can be most easily solved in the spectral domain.

Before leaving the finite index set we should note that its most important practical application may be as an approximation to the index sets $\mathcal{T} = \mathcal{Z}$ and $\mathcal{T} = \mathcal{R}$. Specifically, a process defined on $\mathcal{T} = \mathcal{Z}$, may be approximated as WSS $\mathcal{T} = \mathcal{Z}_N$ in order to take advantage of the computational advantages of the FFT. Thus, a good understanding of this processing yields a more complete understanding of the underlying approximation.

5.4 Some Related Problems/Questions

- Why are circulant matrices closed under multiplication – why is the product of two circulant matrices circulant? Do circulant matrices commute?

- How many unique values define a WSS covariance function $K_x(m)$ on $\mathcal{T} = \mathcal{Z}_N$?

- What type of symmetry properties does $\lambda_x(k)$ have if the random process $x(u,n)$ is real?
- How would the whitening and/or simulation be done in the LTI/WSS setting.
- How are the KL–expansion and the DFT related for WSS processes on $\mathcal{T} = \mathcal{Z}_N$.
- Can a WSS hypothesis testing problem be formulated in the spectrum? If so, what is the form of the filter used for the correlation detector?
- What does a singular WSS correlation function look like in the spectral domain.
- Given the spectrum of $K_x(m)$, what is the directional preference of the WSS process?
- If the covariance operator is singular, what does the pseudo inverse look like in the spectral domain? How would this affect the solution in the Affine MMSE estimation example?
- Express $\mathbb{E}\{\|\mathbf{x}(u)\|^2\}$ in terms of $S_x(k)$ for a WSS process. How is this related to $\mathbb{E}\{|x(u,n)|^2\} = R_x(0)$?
- If $S_x(0) = 0$, what can be said about m_x? What if $S_x(0) > 0$, then what can be said about the mean of $x(u,n)$?
- Will this list ever end? ;-)

6 Stochastic Convergence Theory

The objective of stochastic convergence theory is to define what is meant by

$$x(u) = \lim_{n\to\infty} x_n(u). \tag{72}$$

Specifically, this is a limit of functions of u, so do we require that the limit holds for all $u \in \mathcal{U}$?, or is an average measure more appropriate?

To give an appreciation of the problem, consider the Fourier Series which represents a limit of functions of t

$$x(t) = \lim_{n\to\infty} \sum_{k=-n}^{n} X_k e^{j2\pi \frac{k}{T} t}. \tag{73}$$

Do you know in what sense this limit holds? It certainly doesn't hold for all values of t – witness the "ringing" effects for a square wave expansion (i.e., the Gibbs phenomenon). It actually represents convergence in the sense that, in the limit, the difference has no power. So you have dealt with a problem involving convergence of functions and have a least accepted the Fourier Series result as a reasonable criterion for convergence. The following represents an extension along these lines.

Before describing stochastic convergence, it is helpful to quickly review what it means for a sequence of real *deterministic* numbers to converge.

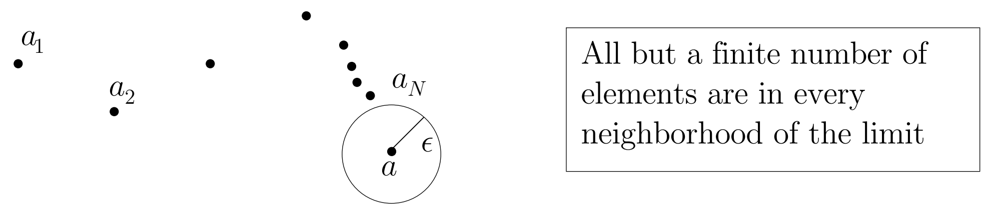

Figure 2: The geometric interpretation of a convergent sequence.

Definition: A sequence of real numbers $a_1, a_2, a_3, \ldots$ is said to converge to a limit $a \in \mathcal{R} = (-\infty, \infty)$ if for every choice of $\epsilon > 0$, there exists an integer N_ϵ such that $n > N_\epsilon$ ensures that $|a - a_n| < \epsilon$.

The shorthand for this statement (to be used henceforth) is: $\{a_n\}$ converges if $\exists\ a \in \mathcal{R} \ni \forall\ \epsilon > 0,\ \exists\ N_\epsilon \ni n > N_\epsilon \Rightarrow |a - a_n| < \epsilon$. Another way of stating this is that if any ball around a contains all but a finite number of the elements of $\{a_n\}$ – this concept is illustrated in Figure 2.

Finally, note that the limit must be a real number (∞ and $-\infty$ don't count!).

6.1 Modes of Stochastic Convergence

There are many "types" or "modes" of stochastic convergence which can, and have, been defined. Below is a partial list, each using the definition of a deterministic limit as defined above. In these definitions, for simplicity, the sequence of random variables is assumed to be real – extensions to complex sequences are straightforward.

Sure Convergence This is point-wise convergence; $x_n(u)$ converges to $x(u)$ *surely* if and only if (iff)

$$\lim_{n\to\infty} x_n(u_0) = x(u_0) \quad \forall\ u_0 \in \mathcal{U}. \tag{74}$$

In other words, for fixed $u_0 \in \mathcal{U}$ the question is one of whether a sequence of real numbers converge, so *sure convergence* requires that the limit holds for every choice of u_0. It should be noted that sure convergence seems to be an idea cooked-up to help engineers to understand the next concept, *almost sure convergence* – nobody uses sure convergence.

Almost Sure Convergence This is the next best thing to sure convergence. The definition of $x_n(u)$ converging to $x(u)$ *almost surely* is

$$x_n(u) \stackrel{\text{as}}{\to} x(u) \qquad \Longleftrightarrow \qquad \lim_{n\to\infty} x_n(u_0) = x(u_0) \quad \forall\ u_0 \in A \subset \mathcal{U} \text{ with } P(A) = 1. \tag{75}$$

In English, we cannot assure that the limit is $x(u)$ for all choices of $u \in \mathcal{U}$, but if all the values of u for which it is false are collected, they have probability zero. This is also written as $\Pr\{x(u) = \lim_{n\to\infty} x_n(u)\} = 1$.

An example of *almost sure* convergence is provided by the *Strong Law of Large Numbers*.

Convergence in Probability The definition of $x_n(u)$ converging to $x(u)$ *in probability* is

$$x_n(u) \xrightarrow{\mathrm{p}} x(u) \qquad \Longleftrightarrow \qquad \lim_{n\to\infty} \mathrm{P\scriptstyle R}\left\{|x(u) - x_n(u)| > \epsilon\right\} = 0 \quad \forall\, \epsilon > 0. \tag{76}$$

This means that if one is willing to wait, it can be assured that the probability that $x_n(u)$ is close to $x(u)$ (as close as one likes) is made arbitrarily close to 1.

An example of convergence in probability is provided by the *Weak Law of Large Numbers*. Convergence in probability is sometimes call"*p*–convergence."

Convergence in Distribution The definition of $x_n(u)$ converging to $x(u)$ *in distribution* is

$$x_n(u) \xrightarrow{\mathrm{d}} x(u) \qquad \Longleftrightarrow \qquad \lim_{n\to\infty} F_{x_n(u)}(z) = F_{x(u)}(z) \quad \forall\, z \in \mathcal{R} \quad \ni \; F_{x(u)}(z) \text{ continuous at } z. \tag{77}$$

Convergence in distribution deals only with the complete statistical description. For example, if $x_n(u)$ converges in distribution to a mean-zero, unit-variance Gaussian random variable, then it converges to *every* mean-zero, unit-variance Gaussian random variable. This means that the limit random variable has nothing to do with the random variable $x_n(u)$, it could be independent.

An example of convergence in distribution is the *Central Limit Theorem*. Convergence in distribution is sometimes called "weak convergence."

Convergence in the Mean-Square Sense This is the most important mode of convergence for second moment theory. It is analogous to the Fourier Series convergence. The definition of $x_n(u)$ converging to $x(u)$ *in the mean-square sense (mss)* is

$$x_n(u) \xrightarrow{\mathrm{mss}} x(u) \qquad \Longleftrightarrow \qquad \lim_{n\to\infty} \mathbb{E}\left\{|x(u) - x_n(u)|^2\right\} = 0. \tag{78}$$

Intuitively, mss convergence corresponds to vanishing average power in the difference between $x_n(u)$ and the limit random variable.

We'll see many examples of mss convergence in the second-half of the class. For the record, mss convergence is often referred to as the "limit in the mean" and denoted by

$$x(u) = \mathrm{l.i.m.}_{n\to\infty} x_n(u) \quad \longleftrightarrow \quad x_n(u) \xrightarrow{\mathrm{mss}} x(u). \tag{79}$$

This notation will not be used in this text. Mean-square convergence is also referred to "$\mathcal{L}_2$ convergence."

The obvious question is: Which of these modes is the best, or at least the strongest? The answer is provided by the Venn diagram of Figure 3.[6] The conclusion is that there is no

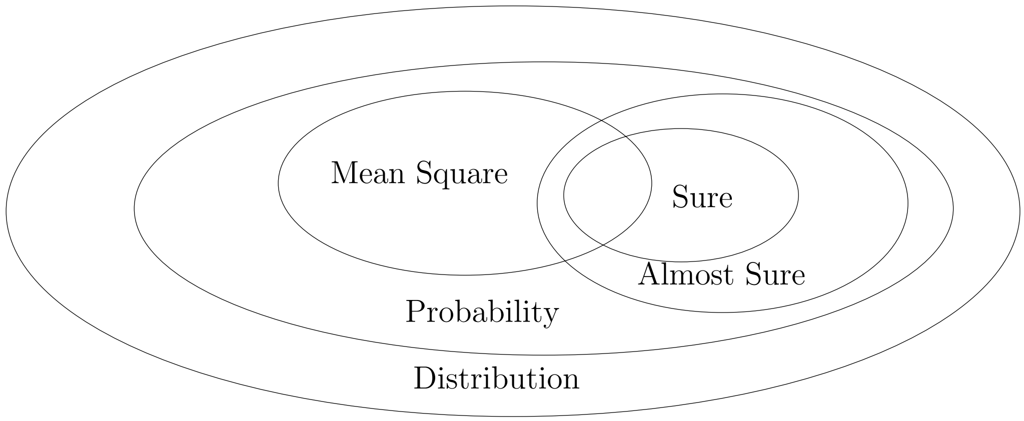

Figure 3: The relation between various modes of convergence.

strongest mode of convergence, although either mss convergence or almost sure convergence imply convergence in probability, which in turn implies convergence in distribution.

It seems strange that we can have almost sure or even sure convergence without mss convergence. Examples of this sort are usually fairly artifical, with the trait that for large n, some of the probability mass drifts off to infinity.

6.2 Convergence in an Arbitrary Metric Space

Recall that a metric space $\mathcal{X}$ is an abstract space where distances can be measured. The distance between $\tilde{\boldsymbol{x}}$ and $\tilde{\boldsymbol{y}} \in \mathcal{X}$, denoted by $d(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}})$, must satisfy four properties (see the *MMSE handout*, problem 1, or the *Supplemental Notes*, Figure 1.5). Convergence in a metric space is defined in a manner analogous to convergence on the real line.

Definition A sequence $\{\tilde{\boldsymbol{x}}_n\}$ in a metric space $\mathcal{X}$ converges iff $\exists\ \tilde{\boldsymbol{x}} \in \mathcal{X} \ni \forall\ \epsilon > 0$, $\exists\ N_\epsilon$ and $\ni n > N_\epsilon \Rightarrow d(\tilde{\boldsymbol{x}}_n, \tilde{\boldsymbol{x}}) < \epsilon$.

Put simply,

$$\tilde{\boldsymbol{x}}_n \to \tilde{\boldsymbol{x}} \qquad \Longleftrightarrow \qquad \lim_{n\to\infty} d(\tilde{\boldsymbol{x}}_n, \tilde{\boldsymbol{x}}) = 0. \tag{80}$$

The interpretation of Figure 2 still holds, with the size of the ball now defined by the distance function. Note that most EE students get confused because they forget that the limit $\tilde{\boldsymbol{x}}$ must be a point in $\mathcal{X}$.

Some examples should help solidify the concept

[6] Believe it or not, the picture in Papoulis' classic book (at least thru the 2nd Ed.) is wrong! – nobody's perfect.

- $\mathcal{X} = \mathcal{R}$, with $d(a,b) = |a-b|$. In this case

$$a_n \to a \qquad \Longleftrightarrow \qquad \lim_{n\to\infty} |a - a_n| = 0. \tag{81}$$

  Convergence in this space is the standard convergence of real sequences.

- $\mathcal{X} = \mathcal{C}^m$, with $d(\mathbf{x},\mathbf{y}) = \|\mathbf{x}-\mathbf{y}\| = \sqrt{(\mathbf{x}-\mathbf{y})^\dagger(\mathbf{x}-\mathbf{y})}$. In this example

$$\mathbf{x}_n \to \mathbf{x} \qquad \Longleftrightarrow \qquad \lim_{n\to\infty} \|\mathbf{x}_n - \mathbf{x}\| = 0, \tag{82}$$

  a reasonable notion of convergence for complex vectors.

- $\mathcal{X} = \mathcal{L}_2[0,T]$, the space of square-integrable functions on $[0,T]$, with

$$d(\tilde{\boldsymbol{f}},\tilde{\boldsymbol{g}}) = \sqrt{\int_0^T |f(t)-g(t)|^2 dt}. \tag{83}$$

  Convergence of the Fourier Series is of this form.

- $\mathcal{X} = \mathcal{W}$, with $d_{\mathcal{W}}(\tilde{\boldsymbol{x}},\tilde{\boldsymbol{y}}) = \sqrt{\mathbb{E}\left\{|x(u)-y(u)|^2\right\}}$. Recall, this is the space of second moment random variables defined on the sample space $\mathcal{U}$. Convergence in this metric space is defined by

$$\begin{aligned} \tilde{\boldsymbol{x}}_n \to \tilde{\boldsymbol{x}} \;&\Longleftrightarrow\; \lim_{n\to\infty} \sqrt{\mathbb{E}\left\{|x_n(u)-x(u)|^2\right\}} = 0 && (84)\\ &\Longleftrightarrow\; \lim_{n\to\infty} \mathbb{E}\left\{|x_n(u)-x(u)|^2\right\} = 0. && (85) \end{aligned}$$

  It follows that convergence in this abstract space corresponds to mean-square sense convergence of the concrete random sequence.

The last example provides a link between the abstract space of random variables $\mathcal{W}$, with which we are familiar, and a particular mode of stochastic convergence, namely mss convergence.

6.3 A MSS Law of Large Numbers

Time for an example; consider a sequence of uncorrelated complex random variables $y_k(u)$, each with mean m, and variance σ^2, both finite. Define the sequence of *sample mean random variables* as

$$x_n(u) = \frac{1}{n}\sum_{k=1}^{n} y_k(u) \qquad n = 1,2,3,\ldots \tag{86}$$

It is a simple exercise to show that

$$\mathbb{E}\left\{x_n(u)\right\} = m \qquad \mathrm{var}\left[x_n(u)\right] = \frac{\sigma^2}{n}. \tag{87}$$

Since as n gets large, the variance of $x_n(u)$ goes to zero, while the mean remains m, we might guess that the limit of $x_n(u)$ is $x(u) = m \ \forall\, u \in \mathcal{U}$. In fact this is true for convergence in probability (i.e., the Weak Law of Large Numbers). If the random variables $\{y_n(u)\}$ are independent (not just uncorrelated), then m is the limit in the almost sure sense as well (i.e., the Strong Law of Large Numbers[7]).

Is m the limit in the mss too? Let's check the definition

$$\mathbb{E}\left\{|x_n(u) - x(u)|^2\right\} = \mathbb{E}\left\{|x_n(u) - m|^2\right\} = \mathrm{var}\left[x_n(u)\right] = \frac{\sigma^2}{n}. \tag{88}$$

Since this limit of this quantity is zero, we have the "MSS Law of Large Numbers:"

$$\frac{1}{n}\sum_{k=1}^{n} y_k(u) \stackrel{\mathrm{mss}}{\to} m. \tag{89}$$

We may call this an "Equally Strong Law of Large Numbers" for the following reason: the Strong Law implies the Weak Law since almost sure convergence implies convergence in probability. However, neither the Strong nor Weak Laws imply this new law since mss convergence is not implied by either of the other two modes. In fact, the mss result assumes only uncorrelated random variables, while the Strong Law of Large Numbers requires an independent sequence – in this sense our new law is even stronger than the strong law!

This example should help to convince you that mss convergence is reasonable, but it raises a serious issue. How does one guess the limit random variable if it is not a trivial constant, as was the case for the Law of Large Numbers? The answer is the *Cauchy Criterion.*

6.4 Cauchy Sequences and Complete Spaces

The term "complete space" has been mentioned before, but rather than explaining it in depth, it was described as a "space with nice convergence properties." This is clarified in this section.

Definition A *complete space* is one in which all Cauchy sequences converge.

Definition A sequence $\{\tilde{\boldsymbol{x}}_k\}$ in a metric space $\mathcal{X}$ is called a Cauchy[8] sequence if

$$\lim_{n,m\to\infty} d(\tilde{\boldsymbol{x}}_n, \tilde{\boldsymbol{x}}_m) = 0. \tag{90}$$

The formal definition of a limit on two arguments (i.e., n and m) is

$$\lim_{n,m\to\infty} d(\tilde{\boldsymbol{x}}_n, \tilde{\boldsymbol{x}}_m) = 0 \qquad \Longleftrightarrow \qquad \forall\ \epsilon\ > 0,\ \exists\ N_\epsilon\ \ni\ n,\ m > N_\epsilon \Rightarrow d(\tilde{\boldsymbol{x}}_n, \tilde{\boldsymbol{x}}_m) < \epsilon. \tag{91}$$

[7] Since this is a sequence of complex random variables, the Strong Law should be modified to handle the complex case. This can be done by requiring that the real and imaginary parts are independent

[8] By the way the answer to the question you're afraid to ask is: "A Cauchy sequence and the Cauchy-Schwartz are completely unrelated, except that Cauchy (apparently a very smart guy) is credited for both concepts.

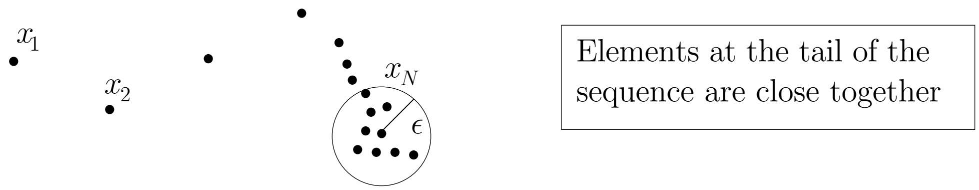

Figure 4: The geometric interpretation of a Cauchy sequence.

A Cauchy sequence is illustrated in Figure 4. For large values of n, all the points of a Cauchy sequence are close together. If you think practically (i.e., an engineer as opposed to a mathematician) and understand the Cauchy sequence definition, you should have the following question in mind: "How can a Cauchy sequence *not* converge? – its going somewhere, so that must be the limit." The answer is the Cauchy sequence leaves the space. Remember the limit must be a point in $\mathcal{X}$, so if the sequence is leaving the space it has no limit. A complete space is one in which a Cauchy sequence cannot leave the space.[9]

It is a simple exercise to show that any sequence which converges is a Cauchy sequence, thus *in a complete space a sequence converges if and only if it is a Cauchy sequence.* In a complete space, convergence (or lack thereof) can be checked by applying the "Cauchy Test" – i.e., if the sequence is Cauchy it converges, otherwise it does not converge.

Here are some examples to illustrate the concept of completeness:

- $\mathcal{X} = (0, 1]$, with $d(a, b) = |a - b|$. This space is incomplete. For example the sequence $x_n = 1/n$ is a Cauchy sequence, but it has no limit in $\mathcal{X}$. It is tempting to claim that the limit is 0, but this is not in the space – the Cauchy sequence left the space.

- $\mathcal{X} = \mathcal{Q}$, the space of rational numbers with $d(p, q) = |p - q|$. It is possible to show that the sequence defined by
$$q_n = \sum_{k=0}^{n} \frac{1}{k!}, \tag{92}$$
is a Cauchy sequence. This sequence has no limit in the rational numbers. Once again, the sequence leaves the space – this sequence has a limit in the real numbers, namely e, the natural log base.

- $\mathcal{X} = \mathcal{R}$, with $d(a, b) = |a - b|$ is a complete space. It follows that
$$a_n \to a \qquad \Longleftrightarrow \qquad \lim_{n,m\to\infty} |a_n - a_m| = 0. \tag{93}$$
This is a fact that is learned (and shortly forgotten) in freshman Calculus.

[9] In fact, an incomplete space can be "completed" by grouping all sequences which seem to leave the space at the same point together into an equivalence class.

- $\mathcal{X} = \mathcal{W}$, with $d_{\mathcal{W}}(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}) = \sqrt{\mathbb{E}\{|x(u) - y(u)|^2\}}$. This is a complete space,[10] so that

$$\tilde{\boldsymbol{x}}_n \rightarrow \tilde{\boldsymbol{x}} \in \mathcal{W} \iff \lim_{n,m\to\infty} \sqrt{\mathbb{E}\{|x_n(u) - x_m(u)|^2\}} = 0 \tag{94}$$

$$\iff \lim_{n,m\to\infty} \mathbb{E}\left\{|x_n(u) - x_m(u)|^2\right\} = 0. \tag{95}$$

It follows that the Cauchy test can be used to check for mss convergence.

In a complete space, a sequence converges if and only if it is a Cauchy sequence. The space of second moment random variables $\mathcal{W}$ is a complete space, and convergence in this space corresponds to mss convergence. Therefore,

$$x_n(u) \stackrel{\text{mss}}{\rightarrow} x(u) \qquad \iff \qquad \lim_{n,m\to\infty} \mathbb{E}\left\{|x_n(u) - x_m(u)|^2\right\} = 0. \tag{96}$$

6.5 What Good is MSS Convergence?

The Cauchy Criterion provides a method for determining whether a sequence converges in the mss. If a sequence of random variables is shown to converge by the Cauchy test, the obvious question is: "What good is it to know that the sequence converges if the limit random variable is unknown?" In other words, if $\{x_n(u)\}$ is Cauchy then it has a limit, say $x(u)$, but $x(u)$ is still unknown. The answer to this question is that *the second moment description of a mss limit can be found by taking the limit of the second moment description.* This is stated formally in the following theorem.

Theorem If $x_n(u) \stackrel{\text{mss}}{\rightarrow} x(u)$ and $y_n(u) \stackrel{\text{mss}}{\rightarrow} y(u)$ then

$$\mathbb{E}\{x(u)\} = \lim_{n\to\infty} \mathbb{E}\{x_n(u)\} \tag{97a}$$

$$\text{var}[x(u)] = \lim_{n\to\infty} \text{var}[x_n(u)] \tag{97b}$$

$$\mathbb{E}\left\{|x(u)|^2\right\} = \lim_{n\to\infty} \mathbb{E}\left\{|x_n(u)|^2\right\} \tag{97c}$$

$$\mathbb{E}\{x(u)y^*(u)\} = \lim_{n\to\infty} \mathbb{E}\{x_n(u)y_n^*(u)\} \tag{97d}$$

Proof First, we present proofs for (97a) and (97c), which together imply (97b). We use the shorthand notation

$$m_n = \mathbb{E}\{x_n(u)\} \qquad m = \mathbb{E}\{x(u)\} \tag{98}$$

$$\beta_n^2 = \mathbb{E}\left\{|x_n(u)|^2\right\} \qquad \beta^2 = \mathbb{E}\left\{|x(u)|^2\right\} \tag{99}$$

$$r_n = \mathbb{E}\{x_n(u)x(u)\} \qquad e_n(u) = x(u) - x_n(u). \tag{100}$$

[10]This is an important fact, but it is stated without proof. A proof is beyond the scope of this text.

To prove (97a), we note that

$$\text{var}\left[e_n(u)\right] = \mathbb{E}\left\{|e_n(u)|^2\right\} - \left|\mathbb{E}\left\{e_n(u)\right\}\right|^2 \geq 0 \tag{101}$$

$$\mathbb{E}\left\{e_n(u)\right\} = \mathbb{E}\left\{x(u) - x_n(u)\right\} = m - m_n, \tag{102}$$

$$\tag{103}$$

so that

$$0 \leq \left|\mathbb{E}\left\{e_n(u)\right\}\right|^2 = |m - m_n|^2 \leq \mathbb{E}\left\{|e_n(u)|^2\right\} = \mathbb{E}\left\{|x(u) - x_n(u)|^2\right\} \to 0. \tag{104}$$

The last equality in (104) follows from the definition of $x_n(u) \stackrel{\text{mss}}{\to} x(u)$. As a result of (104) we know that $|m - m_n|^2$ converges to 0, thus $m_n \to m$. This proves (97a).

The proof of (97c) follows in a similar manner. First note that

$$0 \leq |\beta - \beta_n|^2 = \beta^2 - 2\beta\beta_n + \beta_n^2. \tag{105}$$

By the Cauchy-Schwartz inequality, $\Re\left\{r_n\right\} \leq |r_n| \leq \beta\beta_n$, so that

$$0 \leq |\beta - \beta_n|^2 \leq \beta^2 - 2\Re\left\{r_n\right\} + \beta_n^2 = \mathbb{E}\left\{|x(u) - x_n(u)|^2\right\} \to 0. \tag{106}$$

It follows that $|\beta - \beta_n|^2 \to 0$, or that β is the limit of β_n.

The proof of (97d) is as follows:

$$\begin{aligned}
\left|\mathbb{E}\left\{x(u)y^*(u)\right\} - \mathbb{E}\left\{x_n(u)y_n^*(u)\right\}\right| &= \left|\mathbb{E}\left\{x(u)y^*(u) - x_n(u)y_n^*(u)\right\}\right| && \text{(107a)}\\
&= \left|\mathbb{E}\left\{(x(u) - x_n(u))y^*(u) + x_n(u)(y(u) - y_n(u))^*\right\}\right| && \text{(107b)}\\
&\leq \left|\mathbb{E}\left\{(x(u) - x_n(u))y^*(u)\right\}\right| + \left|\mathbb{E}\left\{x_n(u)(y(u) - y_n(u))^*\right\}\right| && \text{(107c)}\\
&\leq \sqrt{\mathbb{E}\left\{|x(u) - x_n(u)|^2\right\}\mathbb{E}\left\{|y^*(u)|^2\right\}}\\
&\quad + \sqrt{\mathbb{E}\left\{|x_n(u)|^2\right\}\mathbb{E}\left\{|y(u) - y_n(u)|^2\right\}}, && \text{(107d)}
\end{aligned}$$

which goes to zero as $n \to \infty$. Note that (107c) follows from teh triangle inequality and (107d) follows from the Cauchy-Schwartz inequality.

Mean-Square convergence allows us to exchange limits and expectations up to second moments.

6.5.1 A Hitchhiker's Guide to the Cauchy Test

Many EE students can read about Cauchy Tests and mss convergence, but have difficulty applying the results. In this section, a simple "recipe" for checking for mss convergence is described.

An important fact to realize is that the limit theorem of Section 6.5 provides a quick and simple method for demonstrating that a sequence does not converge in the mss. Specifically, if $\{x_n(u)\}$ converges in the mss, then the mean and variance of the limit random variable must be finite (i.e., it must be a point in $\mathcal{W}$). Therefore, if the limits of the mean and variance of $x_n(u)$ do not exist, then the $x_n(u)$ does not converge in the mss. *The converse is not true; if the limit of the mean and variance of $x_n(u)$ exist and are finite, this does not imply that $x_n(u)$ converges in the mss.*[11] An example where the limit of the second moments exist and are finite, but mss convergence does not occur is provided by the Central Limit Theorem (i.e., the CLT does not hold in the mss – see Chugg problem 29).

It is usually much easier to check for the limit of the second moments, then to check the Cauchy Test. Thus, the following is an easy procedure for checking for mss:

---

1. *First compute $m_n = \mathbb{E}\{x_n(u)\}$ and $\sigma_n^2 = \mathrm{var}[x_n(u)]$. If one or each of these does not converge as $n \to \infty$, then conclude that $x_n(u)$ does not converge in the mss. If m_n and σ_n^2 both have a finite limit as $n \to \infty$, conclude nothing and go to step 2.*

2. *Now you must check to see if $x_n(u)$ is Cauchy. Compute $E_{n,m} = \mathbb{E}\{|x_n(u) - x_m(u)|^2\}$ and check to see if $E_{n,m}$ converges to zero as n and m go to infinity. You must be careful in performing this step, specifically:*

   (a) *If you are trying to show that $E_{n,m} \to 0$ (i.e., mss convergence of $x_n(u)$), you cannot assume a relationship between n and m other than $n \geq m$, or $n \leq m$.*

   (b) *If you are trying to show that $E_{n,m}$ does not converge to zero (i.e, no mss convergence of $x_n(u)$), you may assume a relationship between n and m. For example it is sufficient to show that $E_{n,n}$ does not converge to 0 as $n \to \infty$.*

To illustrate step 2(a), consider the case where $E_{n,m} = \frac{1}{m} + \frac{1}{n}$. Assuming only that $n \geq m$ implies

$$E_{n,m} = \frac{1}{m} + \frac{1}{n} \leq \frac{2}{m}. \qquad n \geq m \tag{108}$$

Since this goes to zero as $m \to \infty$ (with $n \geq m$), you may conclude that $E_{n,m} \to 0$ as $n, m \to \infty$.

To illustrate step 2(b), consider the case where $E_{n,m} = \left|1 - \frac{m}{n}\right|$. It is sufficient to demonstrate only one way in which $n, m \to \infty$ so that $E_{n,m}$ doesn't converge to 0. Simply take $n = 2m$, so that

$$E_{n,m} = 1 - \frac{m}{n} = \frac{1}{2}. \qquad n = 2m \tag{109}$$

It follows that for $\epsilon = 1/4$, there is no integer $N_{1/4}$ such that $n, m > N_{1/4}$ implies $E_{n,m} < 1/4$ – i.e., $E_{n,m}$ does not converge to 0 and therefore $x_n(u)$ does not converge in the mss.

[11] Interestingly, if it is known that $x_n(u)$ converges to $x(u)$ *almost surely*, then this condition does imply mss convergence.

6.6 Relation to Linear Systems on $\mathcal{T} = \mathcal{Z}$.

In this section, we will end our temporary detour into stochastic convergence theory by applying mss convergence results to discrete time linear systems. We will present the results for arbitrary linear systems and second moment random processes, then specialize the results to the LTI/WSS special case.

Consider an arbitrary discrete time linear system defined by

$$\tilde{\boldsymbol{y}} = \mathbb{H}\tilde{\boldsymbol{x}} \quad \longleftrightarrow \quad y(t) = \sum_{i=-\infty}^{\infty} h(t,i)x(i) \quad t \in \mathcal{Z}. \tag{110}$$

If the input to this system is a random sequence $x(u,t)$, then the output at time t is the limit of

$$y_n(u,t) = \sum_{i=-n}^{n} h(t,i)x(u,i) \quad t \in \mathcal{Z}, \tag{111}$$

as $n \to \infty$. This is a stochastic limit, and must be investigated for each fixed value of t (i.e., fix t and think of this as a random sequence in n).

The objective is to eventually obtain the second moment description of the output from the input's second moments and the system impulse response. Since in the presence of mss convergence we can obtain this by computing the limit of the second moment description of $y_n(u,t)$, the mean-square sense mode of convergence is a logical choice to investigate.

For a fixed $t \in \mathcal{T} = \mathcal{Z}$, $y_n(u,t)$ converges in the mss iff it is Cauchy. Explicitly, the output exists in the mss iff the following quantity goes to zero as $n, m \to \infty$

$$E_{n,m}(t) = \mathbb{E}\left\{|y_m(u,t) - y_n(u,t)|^2\right\}. \tag{112}$$

Assuming only that $m \geq n$, this may be expressed as

$$E_{n,m}(t) = \mathbb{E}\left\{\left|\sum_{i\in\mathcal{A}(n,m)} h(t,i)x(u,i)\right|^2\right\}, \tag{113}$$

where the region of summation is defined as

$$\mathcal{A}(n,m) = \{i : n < |i| \leq m\}. \tag{114}$$

This region is illustrated in Figure 5.

An upper bound to $E_{n,m}(t)$ will yield a sufficient condition for the output to exist in the mss at time t. A series of upper bounds follows from the definition

$$0 \leq E_{n,m}(t) = \sum_{i\in\mathcal{A}(n,m)} \sum_{j\in\mathcal{A}(n,m)} h(t,i)h^*(t,j)R_x(i,j) \tag{115a}$$

$$\leq \sum_{i\in\mathcal{A}(n,m)} \sum_{j\in\mathcal{A}(n,m)} |h(t,i)||h(t,j)||R_x(i,j)| \tag{115b}$$

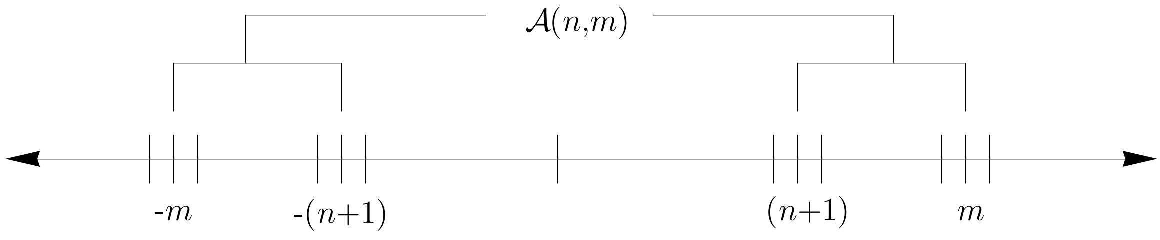

Figure 5: The region of sumation $\mathcal{A}(n,m)$.

$$\leq \sum_{i\in\mathcal{A}(n,m)} \sum_{j\in\mathcal{A}(n,m)} |h(t,i)||h(t,j)|\sqrt{R_x(i,i)}\sqrt{R_x(j,j)} \tag{115c}$$

$$= \left[\sum_{i\in\mathcal{A}(n,m)} |h(t,i)|\sqrt{R_x(i,i)}\right]^2 \tag{115d}$$

$$\leq \left[\sum_{|i|>n} |h(t,i)|\sqrt{R_x(i,i)}\right]^2 \tag{115e}$$

The reasoning behind each step is:

- (115a) follows from expanding (113) and taking expectation.
- (115b) follows from the fact that absolute value of a sum is less than sum of absolute values of the individual terms (i.e., $|a+b| \leq |a| + |b|$).
- (115c) follows from the Cauchy-Schwartz inequality; specifically $|R_x(i,j)| \leq \sqrt{R_x(i,i)}\sqrt{R_x(j,j)}$. The double sum may then be expressed as a square of a single sum as given in (115d).
- (115e) holds since $\mathcal{A}(n,m) \subset \{i : |i| > n\}$ and the terms of the sum are non-negative.

The series of upper bounds given in (115) yields a series of *sufficient* conditions for mss convergence of $y_n(u,t)$. The last is the easiest to check and leads to the following theorem

Theorem *(The Stability Theorem) If* the following two conditions are met

$$\sum_{i=-\infty}^{\infty} |h(t,i)| = M_t < \infty \qquad \text{(stable system at time } t\text{)} \tag{116a}$$

$$R_x(i,i) < R_{\max} < \infty \quad \forall\ i \in \mathcal{Z}, \qquad \text{(uniformly bounded input power)} \tag{116b}$$

then the output exists in the mss at time t.

Proof By (115e),

$$E_{n,m}(t) \leq R_{\max} \left[M_t - \sum_{i=-n}^{n} |h(t,i)| \right]^2, \tag{117}$$

which converges to zero as $n \to \infty$ by the stable system condition.

Notice what this theorem does *not* say – i.e., the converse is not true. If the two conditions of the above theorem are not met, this does not imply that the sequence does not converge in the mss. If the two conditions are not met, one should go back to (115) and test the other conditions. If each successive condition does not hold, the Cauchy Criterion should be tested directly before any conclusions are drawn.

The *Stability Theorem* has a particularly simple form when the process is WSS and the system is LTI.

The output of an LTI system defined on $\mathcal{Z}$ exists in the mss for all values of time when

$$\sum_{n=-\infty}^{\infty} |h(n)| = M < \infty \qquad \textit{(stable system)} \tag{118a}$$

$$R_x(0) < \infty. \qquad \textit{(finite power)} \tag{118b}$$

In other words, the output of a stable LTI system exists in the mss when the input is a finite power WSS process.

6.7 ♣ Mean-Square Calculus

See *Stark and Woods* for a good description of this subject. It will be covered near the end of the semester.

6.8 ♣ Advanced Topics

6.8.1 The Loève Criterion for MSS Convergence

The Loève test for mss convergence is equivalent to the Cauchy Criterion. In some cases, the Loève test is simpler to check. This test deals with the correlation function of the sequence

$$R_x(n_1, n_2) = \mathbb{E}\left\{ x_{n_1}(u) x_{n_2}^*(u) \right\}. \tag{119}$$

Theorem *(Loève Criterion)* A second moment sequence of random variables $x_n(u)$ converges in the mss iff $\lim_{n,m\to\infty} R_x(n,m) = r$, where r is a fixed (finite, real, non-negative) constant.

Proof First note that

$$E_{n,m} = \mathbb{E}\left\{|x_n(u) - x_m(u)|^2\right\} = R_x(n,n) + R_x(m,m) - 2\Re\left\{R_x(n,m)\right\}. \tag{120}$$

It follows that if $R_x(n,m)$ goes to r, then $x_n(u)$ converges in the mss. To prove the other direction, if $x_n(u)$ converges in the mss, then by the theorem of Section 6.5, $R_x(n,n)$ converges to $r = \mathbb{E}\{|x(u)|^2\}$. It follows that, for any $\epsilon > 0$, there exists an N_ϵ, so that $n, m > N_\epsilon$ ensures

$$\begin{aligned} |E_{n,m}| &< \epsilon & (121)\\ |r - R_x(n,n)| &< \epsilon & (122)\\ |r - R_x(m,m)| &< \epsilon. & (123) \end{aligned}$$

For such values of n and m, we have

$$|r - \Re\left\{R_x(n,m)\right\}| < \frac{3\epsilon}{2}, \tag{124}$$

so that $r - \Re\{R_x(n,m)\}$ is a real Cauchy sequence, and thus $\Re\{R_x(n,m)\}$ converges to r. Using this with the Cauchy-Schwartz inequality implies that $R_x(n,m)$ converges to r. Specifically,

$$\Re\left\{R_x(n,m)\right\} \le |R_x(n,m)| \le \sqrt{R_x(n,n)}\sqrt{R_x(m,m)}, \tag{125}$$

and since both the LHS and RHS converge to r, so does $|R_x(n,m)|$. Since $|R_x(n,m)|$ and $\Re\{R_x(n,m)\}$ both converge to the same value, so does $R_x(n,m)$.

6.8.2 ♣ Non-562a Topics from an Engineering Perspective

It would be nice to say more about the relationship between different types of convergence. Specifically:

- Conditions under which a.s. convergence implies mss convergence (counter examples).
- $\mathcal{L}_1$ convergence: implied by mss convergence (via Jensen's inequality), and the relation to a.s. convergence via Uniform Integrability
- Borel-Cantelli Lemmas and SLLN proof

7 LTI/WSS Processing on $\mathcal{T} = \mathcal{Z}$

With the mss convergence theory in place (i.e., the LTI/WSS stability theorem), we are prepared to obtain results for $\mathcal{T} = \mathcal{Z}$ which are analogous to those obtained for $\mathcal{Z}_N$ in Section 5.

7.1 LTI Systems

Because $\mathcal{Z}$ is infinite, the matrix notation is not practical, so we will rely on the signal notation exclusively. Any linear system on $\mathcal{S}_{\mathcal{Z}}$, $\mathbb{H}$ may be represented by the superposition sum

$$\tilde{\boldsymbol{y}} = \mathbb{H}\tilde{\boldsymbol{x}} \quad \longleftrightarrow \quad y(n_1) = \sum_{n_2=-\infty}^{\infty} h(n_1, n_2)x(n_2). \tag{126}$$

If the system is LTI, it can be shown that $h(n_1, n_2) = h(n_1 - n_2, 0)$. Again, the "0" argument is dropped for compactness and the superposition sum becomes a convolution sum

$$\tilde{\boldsymbol{y}} = \mathbb{H}\tilde{\boldsymbol{x}} \quad \longleftrightarrow \quad y(n_1) = \sum_{n_2=-\infty}^{\infty} h(n_1 - n_2)x(n_2). \tag{127}$$

The frequency set for these LTI systems is $\mathcal{F} = [1/2, 1/2)$, $\mathcal{F} = [0, 1)$, or any other interval of length one. This is why the periodic frequency domain interpretation is useful. It follows that

$$\mathbb{H}\tilde{\boldsymbol{e}}_\nu = H(\nu)\tilde{\boldsymbol{e}}_\nu \quad \longleftrightarrow \quad \sum_{n_2=-\infty}^{\infty} h(n_1 - n_2)e_\nu(n_2) = H(\nu)e_\nu(n_1), \tag{128}$$

where[12]

$$e_\nu(n) = e^{j2\pi\nu n}, \tag{129}$$

and the system eigenvalues are the DTFT of the system impulse response

$$H(\nu) = \mathbb{DTFT}\left\{h(n)\right\} = \sum_{n=-\infty}^{\infty} h(n)e^{-j2\pi\nu n}. \tag{130}$$

7.2 WSS Processes

A WSS process on $\mathcal{Z}$ has mean $m_x(n) = m_x$ and correlation function $R_x(n_1, n_2) = R_x(m)$, with $m = n_1 - n_2$. The correlation function for a WSS sequence defines an LTI operator by

$$\tilde{\boldsymbol{w}} = \mathbb{R}_x\tilde{\boldsymbol{v}} \quad \longleftrightarrow \quad w(n_1) = \sum_{n_2=-\infty}^{\infty} R_x(n_1, n_2)v(n_2) = \sum_{n_2=-\infty}^{\infty} R_x(n_1 - n_2)v(n_2). \tag{131}$$

So the correlation operator defines an LTI system with eigenvalues $S_x(\nu)$ defined by

$$R_x(m) * e_\nu(m) = S_x(\nu)e_\nu(m) \tag{132}$$

$$S_x(\nu) = \mathbb{DTFT}\left\{R_x(m)\right\} = \sum_{m=-\infty}^{\infty} R_x(m)e^{-j2\pi m\nu}. \tag{133}$$

Because the mean function is a multiple of an eigenfunction (i.e., $m_x(n) = m_x e_0(n) = m_x$) the correlation and covariance operators have the same eigenfunctions. The eigenvalues of $K_x(m)$ are

$$\lambda_x(\nu) = S_{x_0}(\nu) = \mathbb{DTFT}\left\{K_x(m)\right\}. \tag{134}$$

Due to this relation, the convention is to drop the $\lambda_x(\nu)$ notation and use $S_{x_0}(\nu)$.

[12] The notation $\nu \in \mathcal{F}$ will be used for the discrete time case to emphasize the normalized frequency of the DTFT.

7.3 LTI/WSS Spectral Relationship

Consider a WSS process $x(u,n)$ passed through a stable LTI system with impulse response $h(n)$. The output process $y(u,n)$ exists in the mss by the WSS Stability Theorem. This mss limit is

$$y(u,n) = h(n)*x(u,n) = \sum_{k=-\infty}^{\infty} h(k)x(u,n-k). \tag{135}$$

Since the convergence is mss, the second moment description of $y(u,n)$ can be found by exchanging limits and expectations.

The mean of $y(u,n)$ is

$$\begin{aligned}
\mathbb{E}\{y(u,n)\} &= \mathbb{E}\left\{\sum_{k=-\infty}^{\infty} h(k)x(u,n-k)\right\} && (136)\\
&= \sum_{k=-\infty}^{\infty} h(k)m_x(n-k) && (137)\\
&= m_x\left[\sum_{k=-\infty}^{\infty} h(k)\right]. && (138)
\end{aligned}$$

Notice that the output mean is not a function of time.

The correlation function of $y(u,n)$ is

$$\begin{aligned}
R_y(n_2+m,n_2) &= \mathbb{E}\{y(u,n_2+m)y^*(u,n_2)\} && (139)\\
&= \sum_{k=-\infty}^{\infty}\sum_{l=-\infty}^{\infty} h(k)h^*(l)R_x(n_2+m-k,n_2-l) && (140)\\
&= \sum_{k=-\infty}^{\infty}\sum_{l=-\infty}^{\infty} h(k)h^*(l)R_x(m+l-k) && (141)\\
&= \sum_{l=-\infty}^{\infty} h^*(l)\left[\sum_{k=-\infty}^{\infty} h(k)R_x(m+l-k)\right] && (142)\\
&= \sum_{l=-\infty}^{\infty} h^*(l)f(m+l), && (143)
\end{aligned}$$

where $f(m) = R_x(m)*h(m)$. Note that $R_y(n_2+m,n_2)$ depends only on m. Continuing, with $j=-l$

$$\begin{aligned}
R_y(m) &= \sum_{j=-\infty}^{\infty} h^*(-j)f(m-j) && (144)\\
&= f(m)*h^*(-m) && (145)\\
&= h(m)*R_x(m)*h^*(-m). && (146)
\end{aligned}$$

As expected from the general discussion in Section 4, the output process is also WSS. The standard spectral relation also holds, namely

$$m_y = H(0)m_x \tag{147}$$

$$S_y(\nu) = \mathbb{DTFT}\left\{R_y(m)\right\} = |H(\nu)|^2 S_x(\nu), \tag{148}$$

which can be found directly by transforming (146) (see the *Transform Theory Notes*, especially the "spectral conjugation property") and noting that the coefficient of m_x in (138) is $H(0)$.

These results are reasonably intuitive. The mean is a constant, deterministic signal, so it gets scaled by the DC-response of the system. The relation in (146) is directly analogous to (1), and the spectral relation is the special case of (5).[13]

7.4 Power Spectral Density: Properties and Examples

The set of eigenvalues of $R_x(m)$, namely its DTFT $S_x(\nu)$, is referred to as the *Power Spectral Density (PSD)* of $x(u, n)$. The PSD has the same properties as the eigenvalues of the correlation operator on finite index sets, namely it is real and non-negative. These properties follow directly form the Hermitian Symmetry and NND properties directly.

The PSD is Real: For a WSS process the Hermitian Symmetry property is $R_x(m) = R_x^*(-m)$. Therefore,

$$[S_x(\nu)]^* = \left[\sum_{m=-\infty}^{\infty} R_x(m)e^{-j2\pi\nu m}\right]^* \tag{149}$$

$$= \sum_{l=-\infty}^{\infty} R_x^*(-l)e^{-j2\pi\nu l} \qquad (l = -m) \tag{150}$$

$$= \sum_{l=-\infty}^{\infty} R_x(l)e^{-j2\pi\nu l} \qquad (R_x(m) = R_x^*(-m)) \tag{151}$$

$$= S_x(\nu), \tag{152}$$

so that the PSD is real. This is just a proof of the "HS $\sim$ Real" property of the DTFT (see *Transform Theory Notes*).

The PSD is Non-Negative: The NND property can be used to show that the PSD is non-negative:

$$0 \leq Q_N = \frac{1}{2N+1} \sum_{n_1=-N}^{N} \sum_{n_2=-N}^{N} e^{-j2\pi\nu n_1} R_x(n_1, n_2) e^{j2\pi\nu n_2} \tag{153}$$

[13] In fact, it may be thought of as this exact relation for the first column of an infinite dimensional circulant matrix.

$$= \frac{1}{2N+1} \sum_{n_1=-N}^{N} \sum_{n_2=-N}^{N} e^{-j2\pi\nu(n_1-n_2)} R_x(n_1 - n_2). \tag{154}$$

Changing variables to $m = n_1 - n_2$ and $l = n_1 + n_2$ yields

$$Q_N = \frac{1}{2N+1} \sum_{m=-2N}^{2N} e^{-j2\pi\nu m} R_x(m) \sum_{l \in \mathcal{A}(m)} 1 \tag{155}$$

$$= \frac{1}{2N+1} \sum_{m=-2N}^{2N} e^{-j2\pi\nu m} R_x(m)[2N + 1 - |m|], \tag{156}$$

where the region $\mathcal{A}(m)$ is illustrated in Figure 6. Adopting the notation from the *Transform Theory Notes* for a triangular function, we have

$$Q_N = \frac{1}{2N+1} \sum_{m=-\infty}^{\infty} \mathrm{Dtrian}_N(m) R_x(m) e^{-j2\pi\nu m} \tag{157}$$

$$= \frac{1}{2N+1} \mathbb{DTFT}\left\{\mathrm{Dtrian}_N(m) R_x(m)\right\} \tag{158}$$

$$= (2N+1)[\mathrm{dinc}_N(\nu)]^2 \circledast_1 S_x(\nu), \tag{159}$$

where the "Modulation" property of the DTFT has been used. The non-negative property of the PSD follows since

$$\lim_{N\to\infty} (2N+1)[\mathrm{dinc}_N(\nu)]^2 = \delta_D(\nu) \qquad \nu \in [-1/2, 1/2). \tag{160}$$

which is easily seen in the time domain since $\frac{1}{2N+1}\mathrm{Dtrian}_N(m)$ becomes 1 for all M as $N \to \infty$. The result is that

$$\lim_{N\to\infty} Q_N = S_x(\nu), \tag{161}$$

which is non-negative because Q_N is non-negative. A proof of this property which adds more engineering insight is outlined in Scholtz problem 52.

These are the two fundamental properties of the PSD. Other properties follow from the properties of the DTFT (e.g., $S_x(\nu + 1) = S_x(\nu)$). One notable property is that if $R_x(m)$ is real (which can hold even for complex $x(u, n)$), then the PSD is even (i.e., $S_x(\nu) = S_x(-\nu)$).

The Wiener-Khintchine Theorem for Random Sequences: If one were to try to compute the average power in $x(u, n)$ at frequency ν, the following expression would be a logical choice

$$\frac{1}{2N+1} \mathbb{E}\left\{\left|\sum_{n=-N}^{N} x(u,n) e^{-j2\pi\nu n}\right|^2\right\}. \tag{162}$$

Note that this is precisely the quantity Q_N defined in (153). The result that $\lim_{N\to\infty} Q_N = S(\nu)$ is known as the Wiener-Khintchine Theorem.

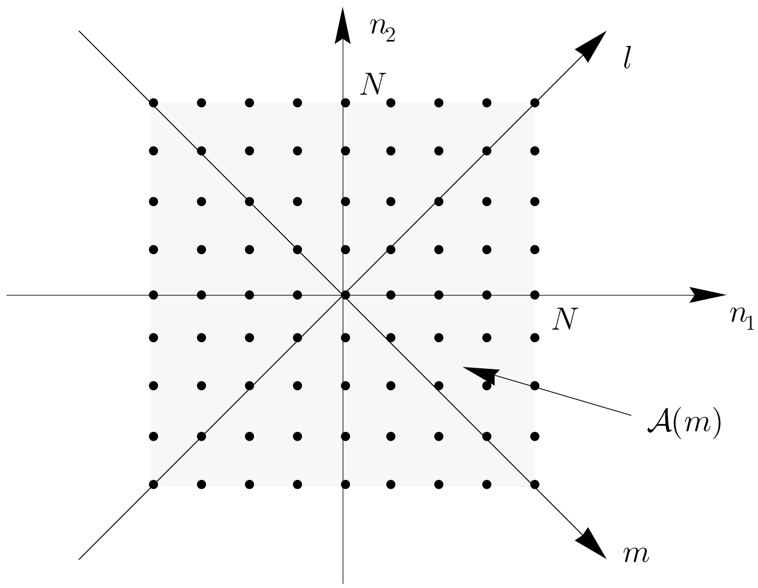

Figure 6: The change of variables from (n_1, n_2) to (m, l).

The Wiener-Khintchine Theorem says that the PSD can be obtained in two equivalent ways:

$$S_x(\nu) = \lim_{N\to\infty} \frac{1}{2N+1} \mathbb{E}\left\{ \left| \sum_{n=-N}^{N} x(u,n) e^{-j2\pi\nu n} \right|^2 \right\} = \mathbb{DTFT}\left\{R_x(m)\right\}. \tag{163}$$

Often times, the PSD is defined by the limit expression and the DTFT relation is obtained by this theorem.

The Wiener-Khintchine Theorem also provides the physical interpretation of the PSD.

The PSD $S_x(\nu)$ is the (ensemble) average power in the the process $x(u,n)$ contained at frequency ν.

This fact is reminiscent of the "directional preference" of a random vector. In this case if $S_x(\nu)$ takes on a maximum value at $\nu = \nu_0$, then the preferred direction is $e_{\nu_0}(n)$, and the mean-square projection coefficient in this direction is $S_x(\nu_0)$. Below are several examples to develop your intuition. The PSD for each case is sketched in Figures 7

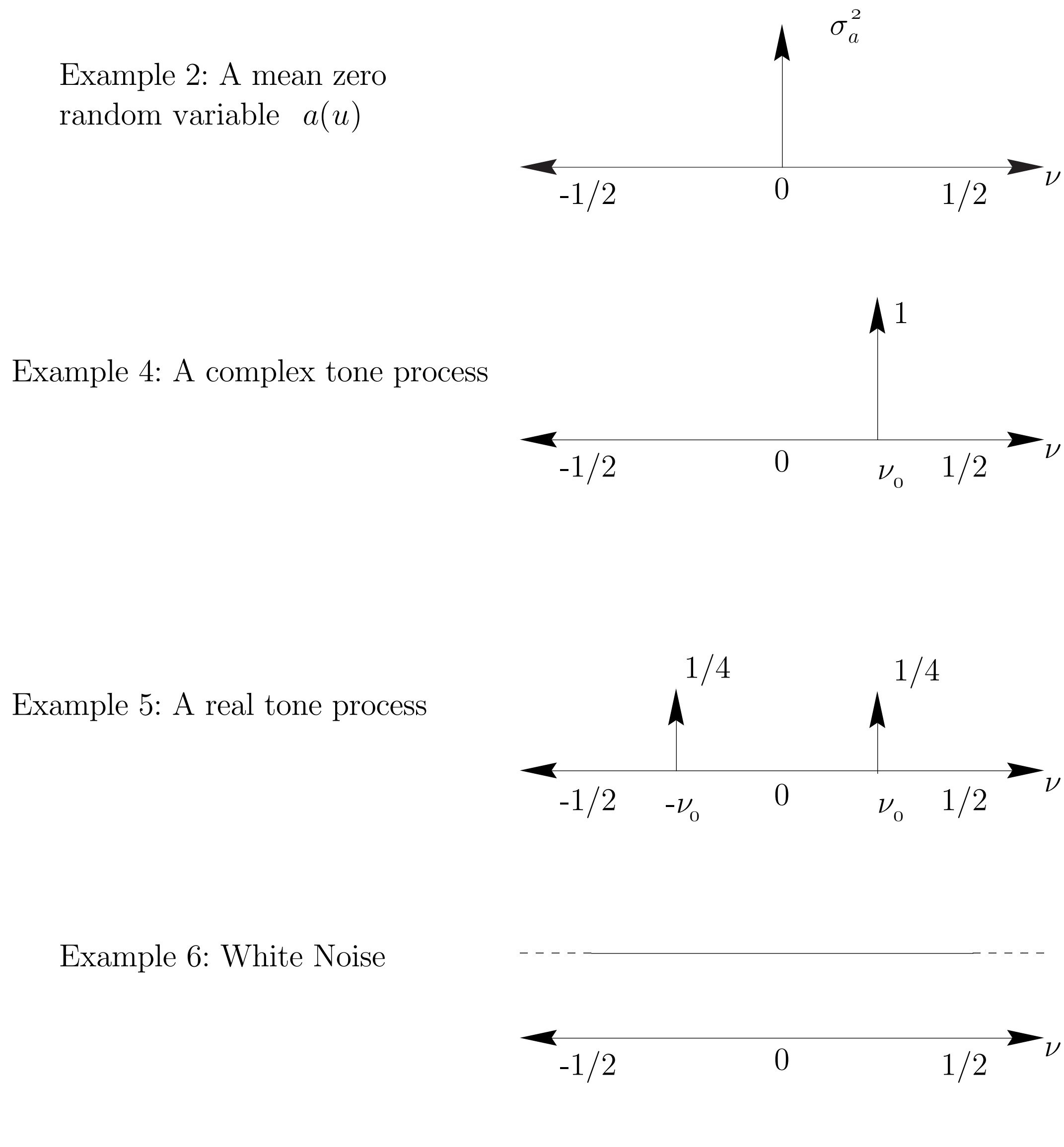

Figure 7: The PSD for the example processes.

1. The total average power of a WSS process is[14]

$$\text{Total Average Power in } x(u,n) = \mathbb{E}\left\{|x(u,n)|^2\right\} = R_x(0) = \int_1 S_x(\nu)d\nu. \tag{164}$$

   This follows by evaluating the inverse DTFT expression at $m = 0$. This makes sense: the total power is the integral of the PSD over all frequencies.

2. Let $x(u,n) = a(u)$, a random variable, with $\mathbb{E}\{a(u)\} = 0$ and $\mathbb{E}\{|a(u)|^2\} = \sigma_a^2$. Since $x(u,n)$ does not change with time, its power should be concentrated at $\nu = 0$. To see this, note that $R_x(m) = \sigma_a^2$ and $m_x = 0$, so

$$S_x(\nu) = \sigma_a^2 \delta_D(\nu) \qquad \nu \in [-1/2, 1/2). \tag{165}$$

   So, as expected, all the power is concentrated at DC. One comment on notation is that this may also be written as

$$S_x(\nu) = \sigma_a^2 \sum_{k=-\infty}^{\infty} \delta_D(\nu - k), \tag{166}$$

   with the periodic DTFT convention. In some cases it is easier to specify the PSD on one period, such as done in (165).

3. Since $K_x(m)$ is a correlation function (i.e., $K_x(m) = R_{x_0}(m)$), its DTFT has the same properties. The relation is

$$\begin{aligned} S_x(\nu) &= \mathbb{DTFT}\{R_X(m)\} && (167)\\ &= \mathbb{DTFT}\{K_x(m) + |m_x|^2\} && (168)\\ &= S_{x_0}(\nu) + |m_x|^2 \delta_D(\nu) \qquad \nu \in [-1/2, 1/2). && (169) \end{aligned}$$

   It follows that if $S_x(\nu)$ does not have a Dirac delta at $\nu = 0$, then $m_x = 0$. The previous example illustrates that the converse is not true: if $S_x(\nu)$ does have a delta at $\nu = 0$, the mean may or may not be zero.

4. Consider $z(u,n) = e^{j(2\pi\nu_0 n + \theta(u))}$, with $\theta(u)$ uniform over $[0, 2\pi)$ and $\nu_0 \in [0, 1/2)$. It is simple to show that

$$\begin{aligned} m_z &= 0 && (170)\\ R_z(m) &= \mathbb{E}\left\{e^{j(2\pi\nu_0(m+n)+\theta(u))} e^{-j(2\pi\nu_0 n+\theta(u))}\right\} = e^{j2\pi\nu_0 m} && (171)\\ \widetilde{R}_z(n_1, n_2) &= 0. && (172) \end{aligned}$$

   It follows that the PSD is

$$S_z(\nu) = \delta_D(\nu - \nu_0) \qquad \nu \in [-1/2, 1/2). \tag{173}$$

   As is expected, all the energy is concentrated at frequency ν_0.

[14] The subscript 1 on the integral means integration over an interval of length 1. This emphasizes the periodic interpretation of the DTFT.

5. Let $x(u,n) = \cos(2\pi\nu_0 n + \theta(u)) = \Re\{z(u,n)\}$, where $z(u,n)$ is the tone process from the last example. Therefore,

$$m_x = \Re\{m_z\} = 0 \tag{174}$$

$$R_x(n_1,n_2) = \frac{1}{2}\Re\left\{R_z(n_1,n_2) + \widetilde{R}_z(n_1,n_2)\right\} \tag{175}$$

$$= \frac{1}{2}\cos(2\pi\nu_0(n_1 - n_2)). \tag{176}$$

The PSD of $x(u,n)$ is then

$$S_x(\nu) = \frac{1}{4}\left[\delta_D(\nu - \nu_0) + \delta_D(\nu + \nu_0)\right] \qquad \nu \in [-1/2, 1/2). \tag{177}$$

This process has all its power concentrated at $\nu = \pm\nu_0$.

6. *Discrete Time White Noise:* Consider $w(u,n)$ which is a sequence of uncorrelated, mean zero random variables, each with variance σ^2. Mathematically, $m_w = 0$ and

$$R_w(m) = \sigma^2\delta_K(m). \tag{178}$$

This process has PSD $S_w(\nu) = \sigma^2$. In other words, white noise has equal power at all frequencies (i.e., no directional preference).

7.4.1 An Example of LTI/WSS Processing

Consider a simple real first-order feedback system driven by white noise, as shown in Figure 8. This system is governed by the following difference equation

$$\tilde{\boldsymbol{s}} = \mathbb{H}\tilde{\boldsymbol{v}} \quad \longleftrightarrow \quad s(n) - as(n-1) = v(n). \tag{179}$$

Let $v(n)$ be an eigenfunction, and it is clear that the system eigenvalue is

$$H(\nu) = \frac{1}{1 - ae^{-j2\pi\nu}}. \tag{180}$$

This system is stable iff $|a| < 1$. The impulse response is

$$h(n) = a^n \mathrm{u}(n) = \begin{cases} a^n & n \geq 0 \\ 0 & n < 0. \end{cases} \tag{181}$$

It follows that the output process has PSD

$$S_y(\nu) = |H(\nu)|^2 S_w(\nu) = |H(\nu)|^2 \tag{182}$$

$$= \left(\frac{1}{1 - ae^{-j2\pi\nu}}\right)\left(\frac{1}{1 - ae^{+j2\pi\nu}}\right) \tag{183}$$

$$= \frac{1}{1 - 2a\cos(2\pi\nu) + a^2}. \tag{184}$$

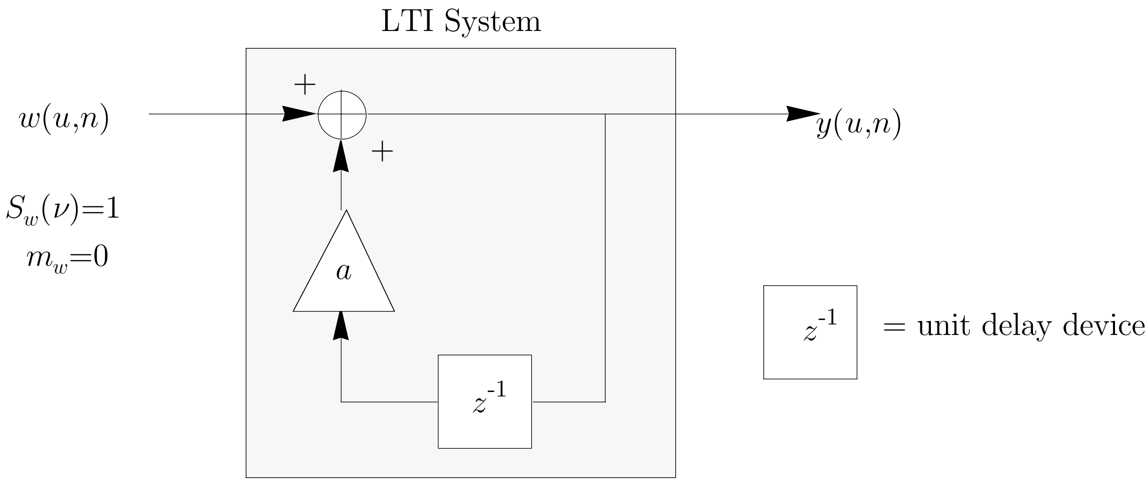

Figure 8: A first-order feed-back (autoregressive (AR)) discrete time system.

Note that since a is real the PSD is even. The corresponding correlation function is

$$R_y(m) = \frac{1}{1-a^2} a^{|m|}. \tag{185}$$

This could be obtained directly as $R_y(m) = h(m)*h(-m)$ with some effort. The correlation and PSD are plotted in Figures 10 and 9, respectively. These plots illustrate, that for this process, the faster $R_y(m)$ decays, the more power $x(u,n)$ contains at high frequency.

7.5 Simulation and Whitening

The simulation and whitening problems are analogous to those considered in the random vector case. In the discrete time WSS simulation problem we begin with a white sequence $w(u,n)$, (i.e., $S_w(\nu) = 1$) and design an LTI system with impulse response $h(n)$ so that $y(u,n) = h(n)*w(u,n) + c$. The objective is to choose $h(n)$ and c such that $y(u,n) \overset{\text{ws}}{=} x(u,n)$ (i.e., $m_y = m_x$ and $K_y(m) = K_x(m)$), where $x(u,n)$ is the WSS process to be simulated. This problem is illustrated in Figure 11.

The solution to this problem is provided by determining the mean and covariance of $y(u,n)$

$$m_y = \mathbb{E}\{h(n)*x(u,n) + c\} = H(0)m_w + c \tag{186}$$
$$= c \tag{187}$$
$$K_y(m) = h(m)*h^*(-m)*K_w(m) \tag{188}$$
$$= h(m)*h^*(-m). \tag{189}$$

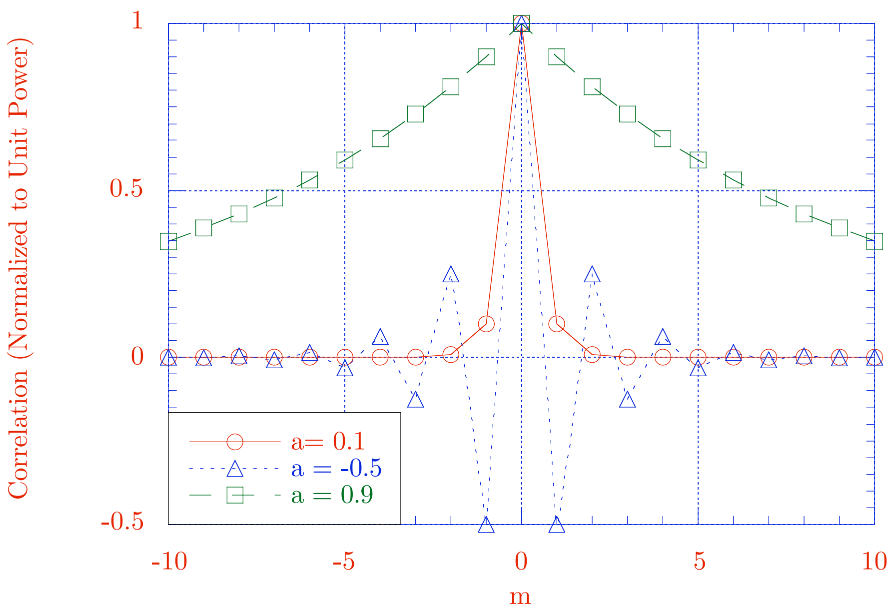

Figure 9: Correlation function of output of first order AR system with white noise input.

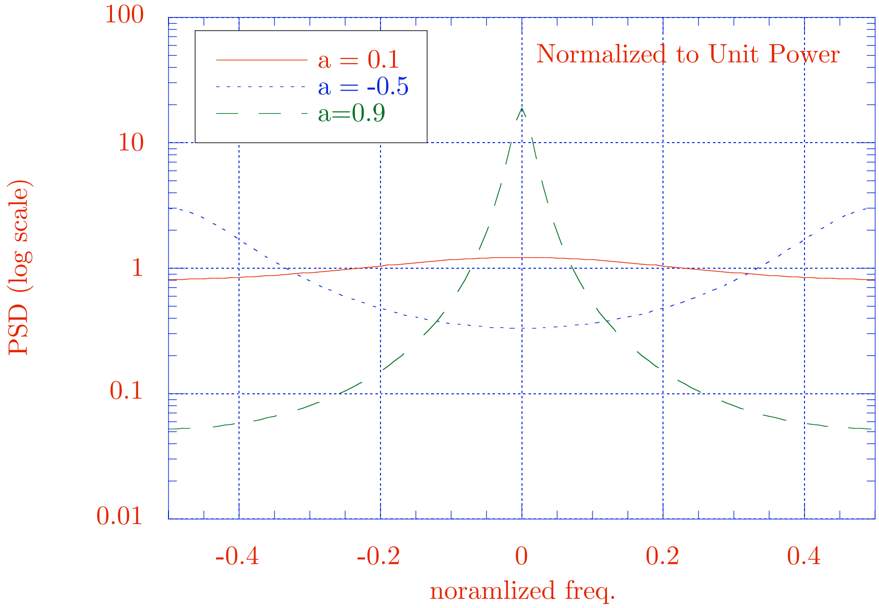

Figure 10: PSD of output of first order AR system with white noise input.

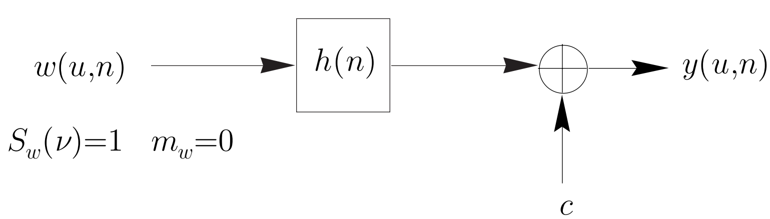

Figure 11: The LTI/WSS simulation problem.

The appropriate choice for is $c = m_x$. The problem of finding $h(m)$ so that $K_x(m) = h(m)*h^*(-m)$ is analogous to the $\mathbf{K_x} = \mathbf{H}\mathbf{H}^\dagger$ problem for random vectors. Solving for $h(m)$ is more easily accomplished in the frequency domain

$$S_y(\nu) = |H(\nu)|^2 S_w(\nu) = |H(\nu)|^2 = S_x(\nu). \tag{190}$$

Thus, the simulation problem is solved by designing an LTI filter with frequency response $H(\nu)$ with squared magnitude equally to $S_x(\nu)$.

The whitening problem is the inverse of the simulation problem. Specifically, given a WSS sequence $x(u,n)$, design an LTI system $g(n)$ so that $w(u,n) = g(n)*(x(u,n) - m_x)$ is a white sequence. From the above development, it is clear that the spectrum of $g(n)$ should satisfy

$$1 = |G(\nu)|^2 S_{x_0}(\nu) \qquad \Longleftrightarrow \qquad |G(\nu)|^2 = \frac{1}{S_{x_0}(\nu)}. \tag{191}$$

Notice that the whitening problem is ill-defined if the covariance operator is singular (i.e., one of the eigenvalues is zero – $S_{x_0}(\nu) = 0$).

7.6 Spectral Factorization

Both the simulation and whitening problems require factorization of a PSD; in this section we consider techniques for performing this spectral factorization.

7.6.1 Desired Properties of a Spectral Factor

Consider a specific example for the simulation problem:

$$S_x(\nu) = \frac{1}{1 - 2a\cos(2\pi\nu) + a^2}, \tag{192}$$

where a is a real number with magnitude less than unity. Because there is no delta function at $\nu = 0$, this process has mean zero, so that $c = 0$. An obvious choice for $H(\nu)$ is provided by the example in Section 7.4.1 – in fact, the solution is obvious because we have seen that this is the PSD of the output of the filter in Figure 8 when the input is white noise. However, this solution is clearly not unique. Consider the following valid choices for the *spectral factor* of $S_x(\nu)$

$$H(\nu) = \frac{1}{1 - ae^{-j2\pi\nu}} \qquad h(n) = a^n \mathrm{u}(n) \tag{193a}$$

$$H(\nu) = \frac{1}{1 - ae^{j2\pi\nu}} \qquad h(n) = a^{|n|} \mathrm{u}(-n) \tag{193b}$$

$$H(\nu) = \frac{1}{\sqrt{1 - 2a\cos(2\pi\nu) + a^2}} \qquad h(n) = ? \tag{193c}$$

$$H(\nu) = \frac{e^{j2\pi\nu(3)}}{1 - ae^{-j2\pi\nu}} \qquad h(n) = a^{n+3} \mathrm{u}(n+3) \tag{193d}$$

$$H(\nu) = \frac{e^{-j2\pi\nu(3)}}{1 - ae^{-j2\pi\nu}} \qquad\qquad h(n) = a^{n-3}\mathrm{u}(n-3). \tag{193e}$$

In fact there are infinitely many choices for $H(\nu)$ so that $S_x(\nu) = |H(\nu)|^2$.

The obvious question arises: "Is there a best choice for a spectral factor?" There is in fact a best choice and in the present example, your intuition should suggest that the choice in (193a) is the best. What are the qualities that we would like $H(\nu)$ to have? Below is a list of desirable features.

1. It should be easy to find $h(n) = \mathbb{DTFT}^{-1}\{H(\nu)\}$ and/or to build a filter with frequency response $H(\nu)$.

2. $H(\nu)$ should be the spectrum of a causal system.

3. In addition to causality, the filter should have no unnecessary delays – i.e., the impulse should start at $n = 0$.

4. Whenever possible, $H(\nu)$ should be selected so that $G(\nu) = 1/H(\nu)$ also has the above three properties.

The first feature eliminates the spectral factor in (193c) (I don't know how to invert this – do you?). The causality constraint eliminates the $H(\nu)$ in (193b) and (193d), while the third criterion eliminates the spectral factor in (193e). So we have have quantified our intuition – the choice of $H(\nu)$ in (193a) satisfies these reasonable criteria.

Now consider the problem of whitening the process $x(u, n)$. Again the solution is not unique, in fact, taking $G(\nu) = 1/H(\nu)$ for each choice of $H(\nu)$ in (193) provides a solution. However, the only choice for which $G(\nu)$ is a causal, stable filter with no unnecessary internal delay is (193a). Thus, selecting the spectral factor in (193a) has the additional advantage that the inverse filter has these same desirable properties.

The above example has one additional aspect which makes the spectral factorization practical and simple: it is a rational function of $z = e^{j2\pi\nu}$ (i.e., the ratio of two polynomials in z). We concentrate on factorization of rational PSD's for the following two practical reasons:

- When $S_x(\nu)$ is rational, a rational spectral factor $H(\nu)$ can be found and easily implemented using delays and adders.

- An arbitrary spectrum may be approximated by a rational function.

In fact, it is possible that you may never encounter a case where a non-rational spectral factorization is necessary. A technique for the general (non-rational) spectral factorization problem is described in Section 7.6.3.

7.6.2 A Recipe for Minimum-Phase Causal Spectral Factorization of Rational PSD

A spectral factor of a rational $S_x(\nu)$ which has the desired properties described in Section 7.6.1 is referred to as a minimum-phase causal (spectral) factor of $S_x(\nu)$. In this section, a specific procedure for performing minimum-phase factorization of rational PSD functions is described.

To begin the development, answer the following question:

$$\text{Is } F(\nu) = \frac{1 - qe^{-j2\pi\nu}}{1 - pe^{-j2\pi\nu}} \text{ a valid PSD?} \tag{194}$$

the answer is *NO*, since this function is not real. In fact, since $S_x(\nu)$ is real, if it contains a term like $F(\nu)$, then it must also contain the term $F^*(\nu)$. It follows, that any rational PSD has the following form

$$S_x(\nu) = C^2 \frac{\prod_{i=1}^{Q}(1 - q_i e^{-j2\pi\nu})(1 - q_i^* e^{j2\pi\nu})}{\prod_{i=1}^{P}(1 - p_i e^{-j2\pi\nu})(1 - p_i^* e^{j2\pi\nu})}, \tag{195}$$

where C is a real constant and $\{p_i\}$ and $\{q_i\}$ are known (possibly complex, and non-distinct) parameters. In words, if $S_x(z)$ has a pole at $z = p$, then it also has a pole at $1/p^*$, with a similar symmetry required for zero locations.[15] This structure, which is illustrated in Figure 12, makes it simple to identify the minimum phase causal factor in the frequency domain.[16]

The general procedure is most easily developed by starting with an example. Consider the specific case of

$$S_x(\nu) = \frac{10 + 6\cos(2\pi\nu)}{[\frac{17}{16} - \frac{1}{2}\cos(2\pi\nu)][\frac{5}{4} + \cos(2\pi\nu)]}. \tag{196}$$

The first step is to express this in the pole-zero form of (195). This is accomplished via the identity

$$(1 - pe^{-j2\pi\nu})(1 - p^* e^{j2\pi\nu}) = 1 + a^2 - 2a\cos(2\pi\nu - \theta), \tag{197}$$

with $p = ae^{j\theta}$ and $a > 0$ a real number. Applying this identity to the example at hand (not both poles are real here) yields

$$S_x(\nu) = \frac{(1 + 3e^{j2\pi\nu})(1 + 3e^{-j2\pi\nu})}{[(1 - \frac{1}{4}e^{j2\pi\nu})(1 - \frac{1}{4}e^{-j2\pi\nu})][(1 + \frac{1}{2}e^{j2\pi\nu})(1 + \frac{1}{2}e^{-j2\pi\nu})]} \tag{198}$$

Since the poles pairs in (195) are of the form $z = ae^{j\theta}$ and $z = \frac{1}{a}e^{j\theta}$, one may be associated with $H(\nu)$ and the other with $H^*(\nu)$. The task is then to split up these pole (and zero) pairs in such a way that $H(\nu)$ corresponds to a stable, causal, minimum-phase system. This step is most easily performed by converting to a Z-transform.

---

[15] The notation $S_x(z)$ is used to denote $S_x(\nu)$ expressed as a function of $z = \exp(j2\pi\nu)$.

[16] In the method presented, poles and/or zeros at $z = 0$, or $z = \infty$ are not considered until the last step of the factorization.

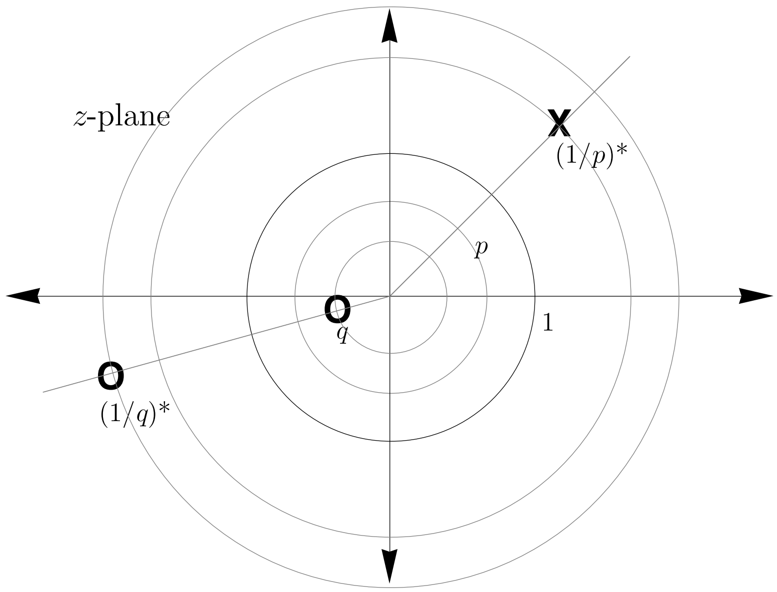

Figure 12: The pole/zero symmetry for $S_x(z)$.

The Z-transform may be seen as an extension of the DTFT; alternately, the DTFT may be viewed as the Z-transform evaluated on the unit circle (i.e., $z = e^{j2\pi\nu}$). It follows that the Z-transform of $R_x(m)$, denoted $S_x(z)$, has a region of convergence (ROC) which includes the unit circle in the z-plane. It is also useful to characterize an LTI system by its Z-transform, and the corresponding ROC. Specifically, the ROC of the Z-transform corresponding to a minimum-phase, stable, causal system has unique properties.

The following facts may be used to recognize a stable, causal, minimum-phase system in the z domain:

1. A stable system has a DTFT which exists, so the ROC of the corresponding Z-transform includes the unit circle.

2. A "right-handed sequence" is one in which $h(n) = 0$ for all $n < n_0$ for some n_0. The Z-transform of a right-handed sequence has an ROC which is the outside of a circle.

3. For a causal system,
$$\lim_{z\to\infty} H(z) = h(0). \tag{199}$$
This property is sometimes summarized as "the ROC of of a causal system includes the point $z = \infty$." This is easily seen from the definition; for a causal system ($h(n) = 0$ for $n < 0$)
$$H(z) = \sum_{n=-\infty}^{\infty} h(n)z^{-n} = h(0) + \frac{h(1)}{z} + \frac{h(2)}{z^2} + \ldots \tag{200}$$

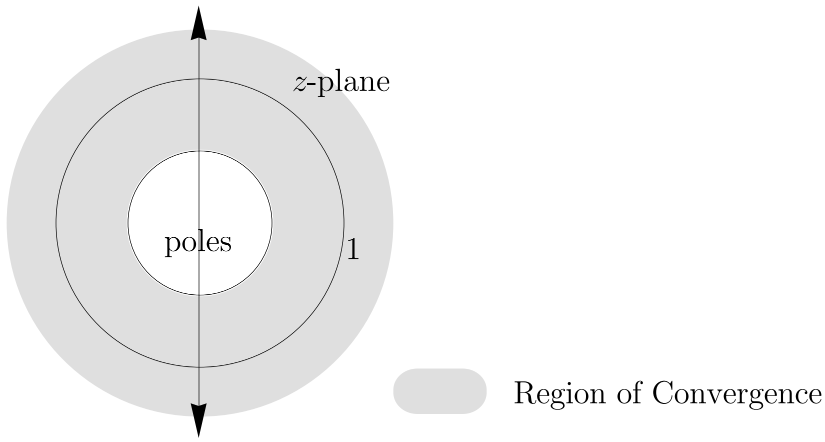

Figure 13: The ROC of $H(z)$, when $\mathbb{H}$ is a stable, causal, minimum-phase system.

If the system is non-causal, $H(z)$ will include positive powers of z and thus, as $z \to \infty$, $H(z)$ diverges.

These facts determine the ROC of the causal, minimum-phase, stable factor.

If $H(z)$ is the Z-transform of a stable, causal, minimum-phase system, then the ROC has the form $\{a < z \leq \infty\}$, where $0 \leq a < 1$, and

$$\lim_{z\to\infty} H(z) = h(0) \neq 0. \tag{201}$$

In other words, all of the poles are inside the unit circle, and $\lim_{z\to\infty} H(z)$ is finite and non-zero

The ROC of a minimum phase system is illustrated in Figure 13.

With this fact, the next step in the example is to convert to $S_x(z)$, so that

$$S_x(z) = \frac{(1+3z)(1+3z^{-1})}{[(1-\frac{1}{4}z)(1-\frac{1}{4}z^{-1})][(1+\frac{1}{2}z)(1+\frac{1}{2}z^{-1})]}. \tag{202}$$

The pole/zero plot for this function is illustrated in Figure 14. Next, split the Z-transform into

$$S_x(z) = H_r(z)\ H_r^*(\cdot)|_{z^{-1}} \tag{203}$$

so that all of the poles and zeroes of $H_r(z)$ are inside the unit circle. The notation $H_r^*(\cdot)|_{z^{-1}}$ is used to represent the function $H^*(\nu)$ evaluated at $z = e^{j2\pi\nu}$ – i.e., conjugate all parameters in $H_r(\cdot)$ and then evaluate at z^{-1}. If all of the poles and zeros of $H_r(z)$ are real, then this is simply $H_r(z^{-1})$.

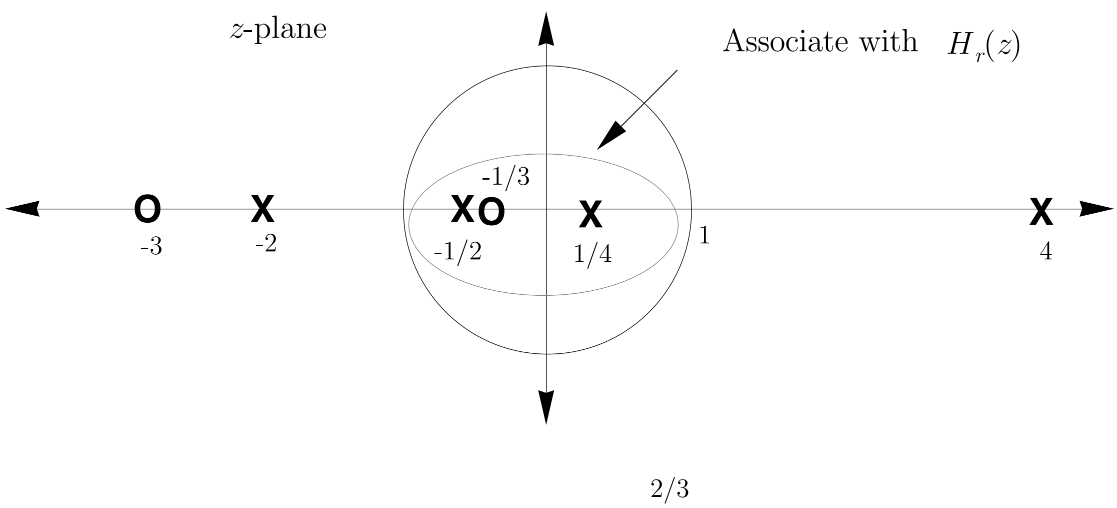

Figure 14: The pole/zero plot for the example $S_x(z)$.

In our example, $S_x(z)$ has poles at $z = \frac{1}{4}, 4, \frac{-1}{2}$, and -2, and zeros at $z = -3$ and $\frac{-1}{3}$. We therefore choose $H_r(z)$ to be

$$H_r(z) = \frac{(1+3z)}{(1-\frac{1}{4}z^{-1})(1+\frac{1}{2}z^{-1})}. \tag{204}$$

We have selected $H_r(z)$ so that it satisfies all the properties of the desired spectral factor except possibly the causality/minimum-phase criterion. The final step is to choose the spectral factor as

$$H(z) = z^m H_r(z), \tag{205}$$

with the integer m selected so that $\lim_{z\to\infty} H(z)$ is finite and non-zero. This can be accomplished by selecting m so that the numerator and denominator polynomials have the same order. Thus, in this example, the proper choice is $m = -1$, so that

$$H(z) = \frac{z^{-1}(1+3z)}{(1-\frac{1}{4}z^{-1})(1+\frac{1}{2}z^{-1})} = \frac{z(1+3z)}{(z-\frac{1}{4})(z+\frac{1}{2})}. \tag{206}$$

The result of our example is that $H(\nu)$ represents a causal, stable, minimum-phase filter, which has squared magnitude equal to $S_x(\nu)$. In fact, it is simple to show that

$$h(n) = \frac{4}{3}\left[\frac{7}{4}\left(\frac{1}{4}\right)^n + \frac{1}{2}\left(\frac{-1}{2}\right)^n\right] \mathrm{u}(n). \tag{207}$$

Notice that if we would have selected $m > -1$, the system would have been non-causal (i.e., $h(-1) \neq 0$), and if we selected $m < -1$, the system would not be minimum-phase (i.e., $h(0) = 0$).

Also notice that for this spectral factor, $G(\nu) = 1/H(\nu)$ is also a causal stable system. The procedure given provides this nice result whenever there are no zeros on the unit circle.

Let's summarize the procedure:

Below are the steps to factor a rational PSD $S_x(\nu)$ into $|H(\nu)|^2$, where $H(\nu)$ is the frequency response of a stable, causal, minimum-phase system:

1. *Convert $S_x(\nu)$ into pole/zero form (i.e., see (195)).*
2. *Convert to $S_x(z)$ using $z = e^{j2\pi\nu}$.*
3. *Split $S_x(z) = H_r(z)\ H_r^*(\cdot)|_{z^{-1}}$ so that all of the poles and zeroes of $H_r(z)$ are inside the unit circle.*
4. *Set $H(z) = z^m H_r(z)$ and select m so that $\lim_{z\to\infty} H(z)$ is finite and non-zero – i.e., select m so that the numerator and denominator of $H(z)$ have the same degree.*

With this procedure the choice of $H(z)$ is unique up to a unit magnitude scale factor. If you like, you may add the last step of multiplying by a constant α, with $|\alpha| = 1$, so that $h(0)$ is real and positive.

7.6.3 ♣ General Spectral Factorization Considerations

There are a few more general cases which we should mention before leaving spectral factorization.

Non-Rational PSD The first generalization is factorization of a general PSD – i.e., including non-rational functions. There is a general procedure for this, which begins by noting that

$$S_x(\nu) = |H(\nu)|^2 \quad \Rightarrow \quad \log(S_x(\nu)) = \log(H(\nu)) + \log(H^*(\nu)). \tag{208}$$

The next step involves forming the Fourier Series of $\log(S_x(\nu))$ in ν – i.e., computing the inverse DTFT of $\log(S_x(\nu))$. From this expansion the contribution of $\log(H(\nu))$ and $\log(H^*(\nu))$ can be identified and separated. The result is that one can identify the causal, stable factor of $S_x(\nu)$ even for non-ration PSD. However, the resulting filter is not simple to build, as is the case for a rational spectra.

♣ Eventually, I want to include the details for this factorization.

The main result is that if and only if the Fourier Series of $\log(S_x(\nu))$ exists. A necessary and sufficient condition for this is given below.

The Discrete Time Paley-Wiener Criterion *The PSD $S_x(\nu)$ may be factored into $S_x(\nu) = |H(\nu)|^2$, where $H(\nu)$ is the DTFT of a causal stable system if and only if*

$$\int_1 |\log(S_x(\nu))| d\nu < \infty. \tag{209}$$

For example, the following spectrum has no causal, stable spectral factor:

$$S_x(\nu) = \begin{cases} 1 & |\nu| < \frac{1}{4} \\ 0 & \frac{1}{4} \le |\nu| < \frac{1}{2}. \end{cases} \tag{210}$$

In general, any PSD which is zero over some interval does not have a causal, stable factor.

Non-WSS Simulation The second point is that we have solved only a special case of the factorization of $R_x(n_1, n_2)$. Specifically, we developed a technique for WSS process only. Consider the direct generalization of the Cholesky factorization for random vectors. In this general setting, one would attempt to design a time-varying, causal linear system with impulse response $h(n_1, n_2)$, so that the covariance of the output process is $K_x(n_1, n_2)$ when the input is white noise. This is a general operator factorization problem, and is not considered here.

♣ How can the Cholesky factorization technique be adopted to obtain a finite-memory system which closely approximates the desired solution?

7.7 Cross-PSD and the Two-Filter Formula

The cross-PSD of two jointly-WSS processes is defined by

$$S_{xy}(\nu) = \mathbb{DTFT}\{R_{xy}(m)\}. \tag{211}$$

The symmetry of the cross-correlation function (i.e., $R_{xy}(m) = R^*_{yx}(-m)$, implies

$$S_{xy}(\nu) = (S_{yx}(\nu))^*. \tag{212}$$

A useful result may be stated regarding the system in Figure 15, where two jointly-WSS processes $v(u, n)$ and $w(u, n)$ drive LTI systems so that $x(u, n) = h(n)*v(u, n)$ and $y(u, n) = g(n)*w(u, n)$. Using the techniques applied in Section 7.3, the following results are obtained

$$R_{xy}(m) = h(m)*R_{vw}(m)*g^*(-m) \tag{213}$$
$$S_{xy}(\nu) = H(\nu)G^*(\nu)S_{xy}(\nu). \tag{214}$$

The single-input/single-output result of Section 7.3 is a special case of this "two-filter formula." This result is extremely useful and should be committed to memory.

A special case of the two filter formula yields the cross-PSD of the input and output of a linear system. Specifically, let $w(u, n) = v(u, n)$ and $G(\nu) = 1$; in this case $S_{xw}(\nu) = H(\nu)S_w(\nu)$.

Similar results may be obtained for the pseudo-correlation and its DTFT when the signals are fully-WSS complex processes.

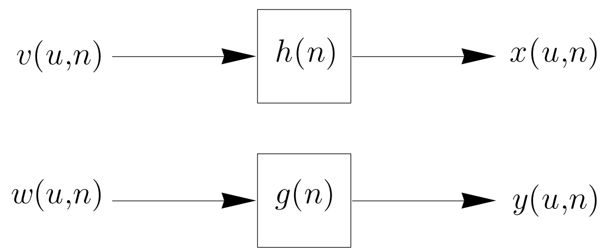

Figure 15: The set-up for the two-filter formula.

7.8 ♣ MMSE Estimation on $\mathcal{T} = \mathcal{Z}$

This will be covered in class; only the results are listed here. In the results listed, it is assumed the the observed sequence $x(u,n)$ and the desired sequence $z(u,n)$ are jointly-WSS processes.

7.8.1 The Affine Constraint

Based on observing $\{x(u,i)\}_{i=-\infty}^{\infty}$, the best affine estimator of $z(u,n)$ is

$$\hat{z}(u,n) = g_{\text{opt}}(n) * (x(u,n) - m_x) + m_z, \tag{215}$$

where the Wiener filter is defined by

$$G_{\text{opt}}(\nu) = \frac{S_{z_0 x_0}(\nu)}{S_{x_0}(\nu)}. \tag{216}$$

The corresponding minimum MSE is

$$MMSE = \mathbb{E}\left\{|z(u,n)|^2\right\} - \mathbb{E}\left\{|\hat{z}(u,n)|^2\right\} \tag{217}$$

$$= \int_1 S_{z_0}(\nu) - \frac{|S_{z_0 x_0}(\nu)|^2}{S_{x_0}(\nu)} d\nu. \tag{218}$$

7.8.2 The Causal Affine Constraint

Based on observing $\{x(u,i)\}_{i=-\infty}^{n}$, the best affine estimator of $z(u,n)$ is

$$\hat{z}(u,n) = g_{\text{opt}}(n) * (x(u,n) - m_x) + m_z, \tag{219}$$

where the Causal Wiener filter is defined by

$$G_{\text{opt}}(\nu) = \mathbb{C}\left\{\frac{S_{z_0 x_0}(\nu)}{H_{x_0}^*(\nu)}\right\} \frac{1}{H_{x_0}(\nu)}, \tag{220}$$

where $H_x(\nu)$ is the causal, minimum phase factor of $S_{x_0}(\nu)$. The causal part operator is defined by

$$\mathbb{C}\left\{P(\nu)\right\} = \mathbb{DTFT}\left\{\mathrm{u}(m)\mathbb{DTFT}^{-1}\left\{P(\nu)\right\}\right\}, \tag{221}$$

The corresponding minimum MSE is

$$MMSE = \mathbb{E}\left\{|z(u,n)|^2\right\} - \mathbb{E}\left\{|\hat{z}(u,n)|^2\right\} \tag{222}$$

$$= \int_1 S_{z_0}(\nu) - \left|\mathbb{C}\left\{\frac{S_{z_0 x_0}(\nu)}{H^*_{x_0}(\nu)}\right\}\right|^2 d\nu. \tag{223}$$

7.9 ♣ Inverting Transforms the Easy Way

8 LTI/WSS Processing on $\mathcal{T} = \mathcal{R}$

In this section results for the index set $\mathcal{T} = \mathcal{R}$ which are analogous to those obtained for $\mathcal{T} = \mathcal{Z}$ in Section 7 are developed. Due to the strong similarity of these results, the derivations and discussions are more terse.

8.1 LTI Systems

Any linear system on $\mathcal{S}_{\mathcal{R}}$, $\mathbb{H}$ may be represented by the superposition integral

$$\tilde{\boldsymbol{y}} = \mathbb{H}\tilde{\boldsymbol{x}} \quad \longleftrightarrow \quad y(t_1) = \int_{-\infty}^{\infty} h(t_1, t_2)x(t_2)dt_2. \tag{224}$$

If the system is LTI, it can be shown that $h(t_1, t_2) = h(t_1 - t_2, 0) = h(t_1 - t_2)$, so that

$$\tilde{\boldsymbol{y}} = \mathbb{H}\tilde{\boldsymbol{x}} \quad \longleftrightarrow \quad y(t_1) = \int_{-\infty}^{\infty} h(t_1 - t_2)x(t_2)dt_2. \tag{225}$$

The frequency set for these LTI systems is $\mathcal{R}$, so that

$$\mathbb{H}\tilde{\boldsymbol{e}}_f = H(f)\tilde{\boldsymbol{e}}_f \quad \longleftrightarrow \quad \int_{-\infty}^{\infty} h(t_1 - t_2)e_f(t_2) = H(f)e_f(t_1), \tag{226}$$

where

$$e_f(t) = e^{j2\pi f t}, \tag{227}$$

and the system eigenvalues are the Fourier Transform (FT) of the system impulse response

$$H(f) = \mathbb{FT}\left\{h(t)\right\} = \int_{-\infty}^{\infty} h(t)e^{-j2\pi f t}dt. \tag{228}$$

8.2 WSS Processes

A WSS process on $\mathcal{R}$ has mean $m_x(t) = m_x$ and correlation function $R_x(t_1, t_2) = R_x(\tau)$, with $\tau = t_1 - t_2$. The correlation function for a WSS sequence defines an LTI operator by

$$\tilde{\boldsymbol{w}} = \mathbb{R}_x \tilde{\boldsymbol{v}} \quad \longleftrightarrow \quad w(t_1) = \int_{-\infty}^{\infty} R_x(t_1, t_2) v(t_2) dt_2 = \int_{-\infty}^{\infty} R_x(t_1 - t_2) v(t_2) dt_2. \tag{229}$$

So the correlation operator defines an LTI system with eigenvalues $S_x(f)$ defined by

$$R_x(\tau) * e_f(\tau) = S_x(f) e_f(\tau) \tag{230}$$
$$S_x(f) = \mathbb{FT}\{R_x(\tau)\} = \int_{-\infty}^{\infty} R_x(\tau) e^{-j2\pi f t} dt. \tag{231}$$

Again, $S_x(f)$ is referred to as the PSD of $x(u,t)$ and it has the same interpretation as the PSD of a random sequence – i.e., the power in $x(u,t)$ at frequency f.

8.3 LTI/WSS Spectral Relationship

Consider a WSS process $x(u,t)$ passed through a stable LTI system with impulse response $h(t)$. The output process is

$$y(u,t) = h(t) * x(u,t) = \int_{-\infty}^{\infty} h(\alpha) x(u, t-\alpha) d\alpha. \tag{232}$$

The mean of $y(u,t)$ is

$$\mathbb{E}\{y(u,t)\} = \mathbb{E}\left\{\int_{-\infty}^{\infty} h(\alpha) x(u, t-\alpha) d\alpha\right\} \tag{233}$$
$$= \int_{-\infty}^{\infty} h(\alpha) m_x(t-\alpha) d\alpha \tag{234}$$
$$= m_x \left[\int_{-\infty}^{\infty} h(\alpha) d\alpha\right]. \tag{235}$$

The correlation function of $y(u,t)$ is

$$R_y(t_2+\tau, t_2) = \mathbb{E}\{y(u, t_2+\tau) y^*(u, t_2)\} \tag{236}$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(\alpha) h^*(\beta) R_x(\tau - \alpha + \beta) d\alpha d\beta \tag{237}$$
$$= h(\tau) * R_x(\tau) * h^*(-\tau). \tag{238}$$

After transforming, we obtain the familiar result

$$m_y = H(0) m_x \tag{239}$$
$$S_y(f) = \mathbb{FT}\{R_y(\tau)\} = |H(f)|^2 S_x(f). \tag{240}$$

This result should be sinking by now; when the system is LTI, and the input is WSS, the output will also be WSS with PSD as stated above.

8.4 Power Spectral Density: Properties and Examples

Once again, we have that the PSD is real and non-negative. Some properties are summarized below

- $R_x(\tau) = R_x^*(-\tau) \Rightarrow S_x(f)$ is real. Shown using the "HS $\sim$ Real" property of FT.
- $S_x(f) \geq 0$ for all $f \in \mathcal{R}$. Again, this is a direct consequence of the NND property, but the proof is a little tedious. The outline is:
  - Start with the fact that the NND property implies

$$Q_T = \frac{1}{2T}\int_{-T}^{T}\int_{-T}^{T} e^{-j2\pi f(t_1-t_2)} R_x(t_1-t_2)dt_1dt_2 \geq 0. \tag{241}$$

  - Change variables to $\tau = t_1 - t_2$ and $\beta = t_1 + t_2$ to show that

$$\begin{aligned} Q_T &= \int_{-2T}^{2T} e^{-j2\pi f\tau} R_x(\tau)\left[1-\frac{|\tau|}{2T}\right]d\tau & (242)\\ &= \mathbb{FT}\left\{R_x(\tau)\mathrm{trian}(\tau/(2T))\right\} & (243)\\ &= S_x(f)*[(2T)\mathrm{sinc}^2(2Tf)]. & (244)\end{aligned}$$

  - The final step is to note that

$$\lim_{T\to\infty}(2T)\mathrm{sinc}^2(2Tf) = \delta_D(f), \tag{245}$$

    so that $Q_T \to S_x(f)$ as $T \to \infty$. Since $Q_T \geq 0$ for all values of T, the limit must also be non-negative.

- The above development also provides the Wiener-Khintchine Theorem for continuous time processes:

$$S_x(f) = \lim_{T\to\infty}\frac{1}{2T}\mathbb{E}\left\{\left|\int_{-T}^{T} Nx(u,t)e^{-j2\pi ft}\right|^2\right\} = \mathbb{FT}\left\{R_x(\tau)\right\}. \tag{246}$$

- If $R_x(\tau)$ is real, then $S_x(f) = S_x(-f)$.
- If $S_x(f)$ does not have a Dirac delta at $f = 0$, then $m_x = 0$. The converse is false – i.e., a delta at zero implies nothing about the mean.
- Total Average Power:

$$\mathbb{E}\left\{|x(u,t)|^2\right\} = R_x(0) = \int_{-\infty}^{\infty} S_x(f)df. \tag{247}$$

8.5 Continuous Time White Noise – An Engineering View

Describing a white noise process on $\mathcal{T} = \mathcal{R}$ is more difficult than one might imagine. We have characterized white noise in the other index sets by either of the following two properties

- White noise decorrelates as fast as possible. In discrete time this means the white noise is a sequence of uncorrelated random variables.
- White noise has no directional preference; it has equal power in each eigen-direction.

The first characterization is difficult to extend to continuous time. Specifically, if we try

$$R_w(\tau) = \begin{cases} 1 & \tau = 0 \\ 0 & \text{otherwise}, \end{cases} \tag{248}$$

we run into the ambiguity that $S_w(f) = 0$ for all f, or equivalently,

$$\int_{-\infty}^{\infty} |R_w(\tau)|^2 d\tau = 0. \tag{249}$$

If we attempt to extend the second characterization to the continuous time case, we have

$$\text{Continuous Time White Noise:} \quad S_w(f) = \frac{N_0}{2} \quad \longleftrightarrow \quad R_w(\tau) = \frac{N_0}{2}\delta_D(\tau). \tag{250}$$

While this is the working definition that we will use, it is not without its own conceptual flaws. Most significantly, *continuous time white noise is not a second moment process*:

$$\text{Ave. Power in } w(u,t)\text{:} \;\; \mathbb{E}\left\{|w(u,t)|^2\right\} = R_w(0) = \int_{-\infty}^{\infty} \frac{N_0}{2} df \to \infty. \tag{251}$$

For this reason $N_0/2$ is not the "variance," but is often referred to as the *intensity* of the white noise. We will use the descriptive term *(two-sided) spectral level* when referring to $N_0/2$.

This is a a disturbing concept, but not unlike others you have accepted in the past. For example, a Dirac delta function does not exist, but it is useful to consider the response of an LTI system to such a non-existent signal (i.e, you cannot produce one in the lab.). In fact, if the system is stable, the output signal (i.e., the impulse response) is well-defined. To measure the system impulse response of an LTI system, one would use a very narrow pulse with unit energy in place of the delta. In practice, "very narrow" means narrow relative to the time constant of the system. In the frequency domain, the pulse spectrum should be flat across the bandwidth of the system being probed. White noise plays a similar modeling role in random signal processing.

Consider the output PSD of a stable LTI system with frequency response $H(f)$, when the input is our (non-existent) white noise (let $N_0/2 = 1$):

$$S_y(f) = |H(f)|^2 S_w(f) = |H(f)|^2. \tag{252}$$

Since this system is stable the output process has finite power. Also, if the actual input process has PSD which is approximately flat across the bandwidth of $H(f)$, the output process would have the same PSD. This is the value of the white noise model.

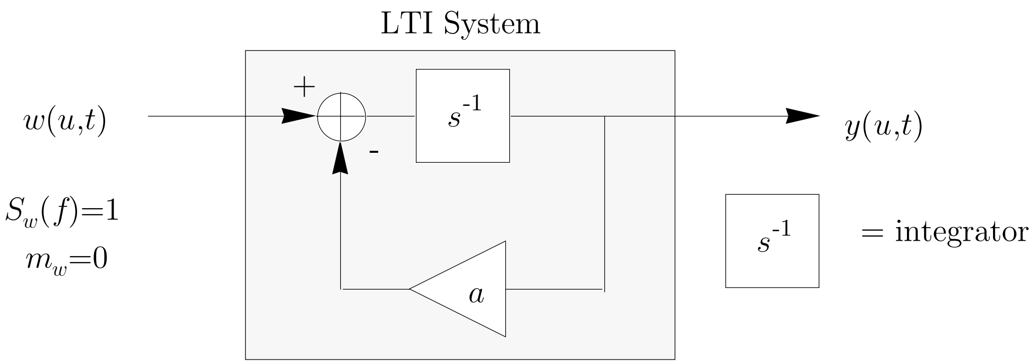

Figure 16: A first-order feed-back (autoregressive) continuous time system.

8.5.1 An LTI/WSS Example and White Noise

As a concrete example, consider a simple real first-order feedback system as shown in Figure 16. This system is governed by the following differential equation

$$\tilde{\boldsymbol{s}} = \mathbb{H}\tilde{\boldsymbol{v}} \quad \longleftrightarrow \quad as(t) + \dot{s}(t) = v(t), \tag{253}$$

where the derivative has been denoted by a dot, and a is a real constant. The frequency response of this system is

$$H(f) = \frac{1}{a + j2\pi f}. \tag{254}$$

This system is stable iff $a > 0$. The impulse response is

$$h(t) = e^{-at}\mathrm{u}(t) = \begin{cases} e^{-at} & t \geq 0 \\ 0 & t < 0. \end{cases} \tag{255}$$

If the input is white noise with spectral level 1, the output process has PSD

$$S_y(f) = |H(f)|^2 \tag{256}$$

$$= \left(\frac{1}{a + j2\pi f}\right)\left(\frac{1}{a - j2\pi f}\right) \tag{257}$$

$$= \frac{1}{a^2 + (2\pi f)^2}. \tag{258}$$

Note that since a is real the PSD is even. The corresponding correlation function is

$$R_y(\tau) = \mathbb{FT}^{-1}\left\{S_y(f)\right\} = \frac{1}{2a}e^{-a|\tau|}, \tag{259}$$

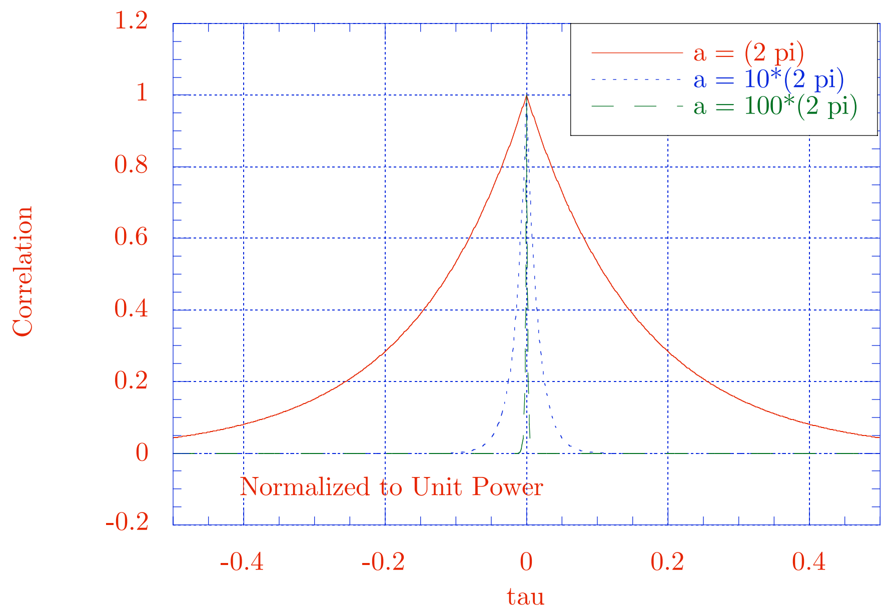

Figure 17: Correlation function of output of first-order AR system with white noise input.

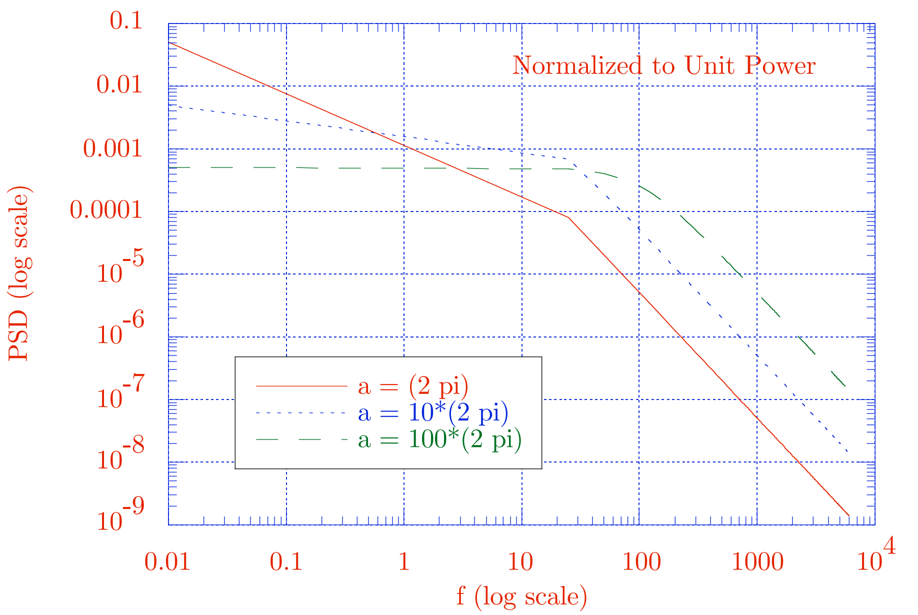

Figure 18: PSD of output of first-order AR system with white noise input.

which could have also been found by $R_y(\tau) = h(\tau)*h(-\tau)$. The correlation and PSD are plotted in Figures 17 and 18, respectively. These plots illustrate, that for this process, the faster $R_y(\tau)$ decays, the more power $x(u,t)$ contains at high frequency.

Consider what would happen if the input was a relatively broad-band process $x(u,t)$, with PSD

$$S_x(f) = \frac{1}{1 + (f/Nf_0)^2}, \tag{260}$$

where

$$f_0 = \frac{a}{2\pi}. \tag{261}$$

The PSD of $x(u,t)$ is fairly flat for $|f| < Nf_0$, then it falls off toward zero as $f \to \infty$. The "half-power" bandwidth of $H(f)$ is f_0, so as the factor N becomes large, we expect the the output process PSD to look similar to that in (258). This can be seen by the relation

$$S_y(f) = |H(f)|^2 S_x(f) = \frac{1}{a^2 + (2\pi f)^2}\left[\frac{1}{1 + (f/Nf_0)^2}\right], \tag{262}$$

Which holds when $x(u,t)$ is the input. The corresponding correlation function is

$$R_y(\tau) = \frac{N^2}{2a(N+1)(N-1)}\left[e^{-a|\tau|} + \frac{1}{N}e^{-aN|\tau|}\right]. \tag{263}$$

The correlation function and PSD for various values of N are plotted in Figures 19 and 20. Notice that as the bandwidth of the input PSD becomes flat across the bandwidth of the system, the white noise approximation becomes very accurate. Also note how much more simple it is to obtain the result when the white noise assumption is made.

White Noise Assumption: *If the input PSD is much broader than the system bandwidth, and is flat over this bandwidth, then the white noise assumption is valid – the input process may be modeled as white noise.*

An equivalent, but more rigorous approach to white noise is defined by the so-called Wiener process. This process, and its relation to the white noise process is described in Section 8.11.

8.6 Simulation and Whitening

The simulation and whitening problems are identical to those considered in Section 7. The notation is also carried over for this discussion. The solution to the simulation problem is obtained by producing $y(u,t) = h(t)*w(u,t) + m_y$, so that $y(u,t) \overset{\text{ws}}{=} x(u,t)$. The simulation filter must solve

$$K_x(\tau) = h(\tau)*h^*(-\tau) \qquad S_{x_0}(f) = |H(f)|^2. \tag{264}$$

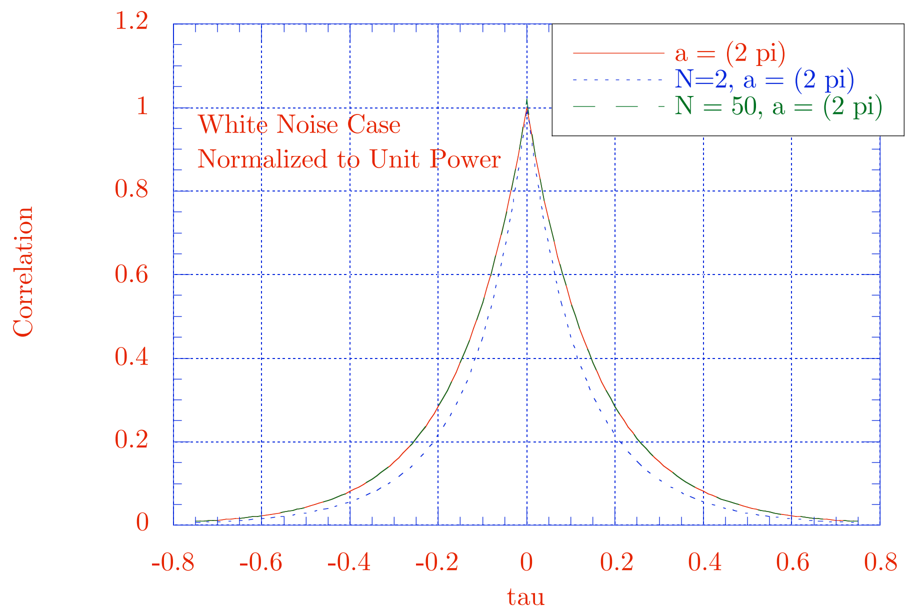

Figure 19: The output process correlation for the first-order AR system when the input is broad-band.

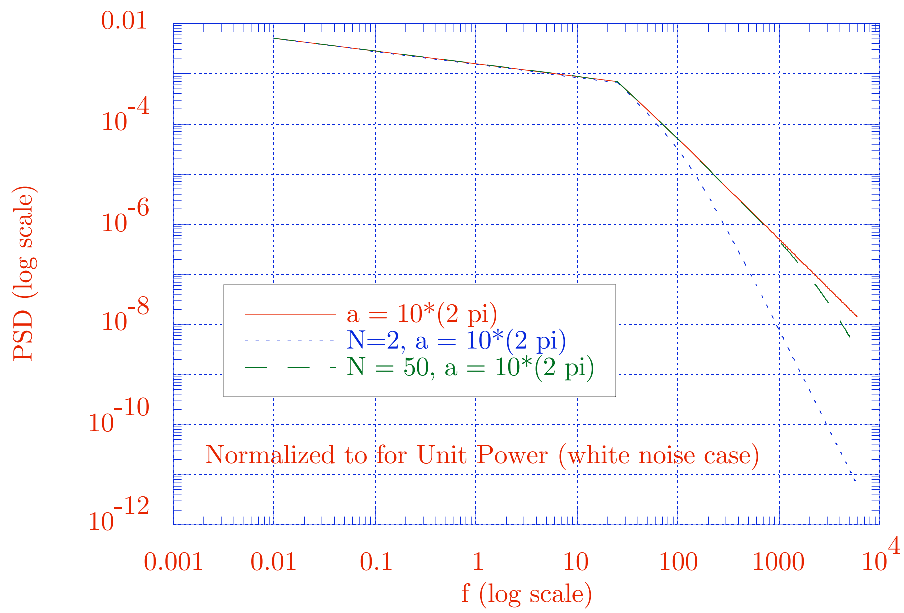

Figure 20: The output process PSD for the first-order AR system when the input is broadband.

The whitening problem requires the design of a filter $g(t)$ with frequency response

$$|G(f)|^2 = \frac{1}{S_{x_0}(f)}. \tag{265}$$

In most applications the whitening filter is cascaded (combined) with other system filters, so that all signals are finite power.

8.7 Spectral Factorization

8.7.1 Desired Properties of a Spectral Factor

Consider a specific example for the simulation problem:

$$S_x(f) = \frac{1}{a^2 + (2\pi f)^2}, \tag{266}$$

where $a < 0$ is a real number. Again, the obvious choice for $H(f)$ is provided by the example in Section 8.5.1, but the solution is not unique.

Consider the following valid choices for the *spectral factor* of $S_x(f)$

$$H(f) = \frac{1}{a + j2\pi f} \qquad h(t) = e^{-at}\mathrm{u}(t) \tag{267a}$$

$$H(f) = \frac{1}{a - j2\pi f} \qquad h(t) = e^{at}\mathrm{u}(-t) \tag{267b}$$

$$H(f) = \frac{1}{\sqrt{a^2 + (2\pi f)^2}} \qquad h(t) = ? \tag{267c}$$

$$H(f) = \frac{e^{+j2\pi 3f}}{a + j2\pi f} \qquad h(t) = e^{-a(t+3)}\mathrm{u}(t+3) \tag{267d}$$

$$H(f) = \frac{e^{-j2\pi 3f}}{a + j2\pi f} \qquad h(t) = e^{-a(t-3)}\mathrm{u}(t-3) \tag{267e}$$

The desirable properties are the same as those for the discrete time case. Namely, we should select a causal, stable, minimum-phase filter. In this case we have only on choice, the filter in (267a). This choice also has an a causal inverse (Is it stable?).

The above example is a rational function of $s = j2\pi f$. We concentrate on factorization of these rational PSD's because they are simple to factor, the filters are easy to construct using adders and integrators, and a non-rational PSD can be approximated using a rational PSD. A technique for the general (non-rational) spectral factorization problem is described in Section 8.7.3.

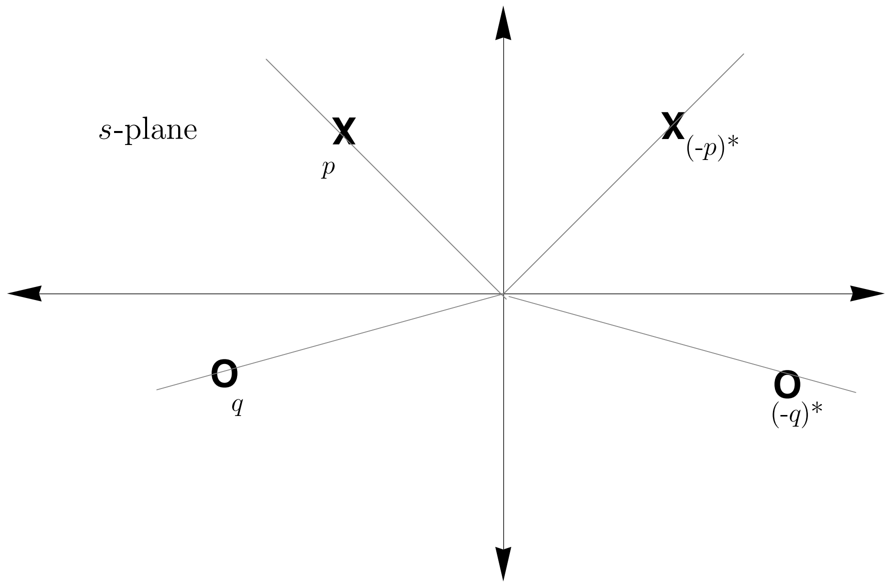

Figure 21: The pole zero symmetry of the PSD.

8.7.2 A Recipe for Minimum-Phase Causal Spectral Factorization of Rational PSD

The procedure for finding the causal, minimum-phase factor for the continuous time case is actually simpler than the one given for discrete time. This is because a delay does not enter the frequency response as a polynomial in s (e.g., see (267d) and (267e)).

To begin the development, answer the following question:

$$\text{Is } M(f) = \frac{q - j2\pi f}{p - j2\pi f} \text{ a valid PSD?} \tag{268}$$

Again, since the PSD must be real, the answer is *NO* – if $S_x(f)$ contains a term like $M(f)$, then it must also contain the term $M^*(f)$. It follows, that any rational PSD has the following form

$$S_x(f) = C^2 \frac{\prod_{i=1}^{Q}(q_i - j2\pi f)(q_i^* + j2\pi f)}{\prod_{i=1}^{P}(p_i - j2\pi f)(p_i^* + j2\pi f)}, \tag{269}$$

where C is a real constant and $\{p_i\}$ and $\{q_i\}$ are known (possibly complex, and non-distinct) parameters. In words, if $S_x(s)$ has a pole at $s = p$, then it also has a pole at $-p^*$, with a similar symmetry required for zero locations.[17] This structure is illustrated in Figure 21.

Again, its easiest to start with an example:

$$S_x(f) = \frac{9 + (2\pi f)^2}{[25 + (2\pi f)^2][4 + (2\pi f)^2]}. \tag{270}$$

[17] The notation $S_x(s)$ is used to denote $S_x(f)$ expressed as a function of $s = j2\pi f$.

The first step is to express this in the pole-zero form of (269). This is accomplished via the identity

$$(p - j2\pi f)(p^* + j2\pi f) = a^2 + (2\pi f - b)^2, \tag{271}$$

with $p = a + jb$ and a, b real numbers. Applying this identity to the example at hand (note both poles are real here) yields

$$S_x(f) = \frac{(3 + (2\pi f))(3 - (2\pi f))}{[(5 + (2\pi f))(5 - (2\pi f))(2 + (2\pi f))(2 - (2\pi f))}. \tag{272}$$

The Laplace Transform may be seen as an extension of the FT; alternately, the FT may be viewed as the Laplace Transform evaluated on the imaginary axis (i.e., $s = j2\pi f$). It follows that the Laplace Transform of $R_x(\tau)$, denoted $S_x(s)$, has a region of convergence (ROC) which includes the imaginary axis in the s-plane. It is also useful to characterize an LTI system by its Laplace Transform, and the corresponding ROC. Specifically, the ROC of the Laplace Transform corresponding to a minimum-phase, stable, causal system has unique properties.

The following facts may be used to recognize a stable, causal, minimum-phase system in the s-domain:

1. A stable system has a FT which exists, so the ROC of the corresponding Laplace Transform includes the imaginary axis.

2. A "right-handed signal" is one in which $h(t) = 0$ for all $t < t_0$ for some t_0. The Laplace Transform of a right-handed sequence has an ROC which is a right-half plane.

These facts determine the ROC of the causal, minimum-phase, stable factor.

If $H(s)$ is the Laplace Transform of a stable, causal, minimum-phase system, then the ROC has the form $\{s : \Re\{s\} > a\}$, where $a < 0$. In other words, all of the poles and zeros are in the left half plane (LHP).

The ROC of typical stablem, causal, minimum-phase system is illustated in Figure 22.

With this fact, the next step in the example is to convert to $S_x(s)$, so that

$$S_x(s) = \frac{(3 + s)(3 - s)}{(5 + s)(5 - s)(2 + s)(2 - s)}. \tag{273}$$

The pole/zero plot for this function is illustrated in Figure 23.

Next, split the Laplace Transform into

$$S_x(s) = H(s)\ H^*(\cdot)|_{-s} \tag{274}$$

so that all of the poles and zeroes of $H(s)$ are in the LHP. The notation $H^*(\cdot)|_{-s}$ is used to represent the function $H^*(f)$ evaluated at $s = j2\pi f$ – if all of the poles and zeros of $H(s)$ are real, then this is simply $H(-s)$.

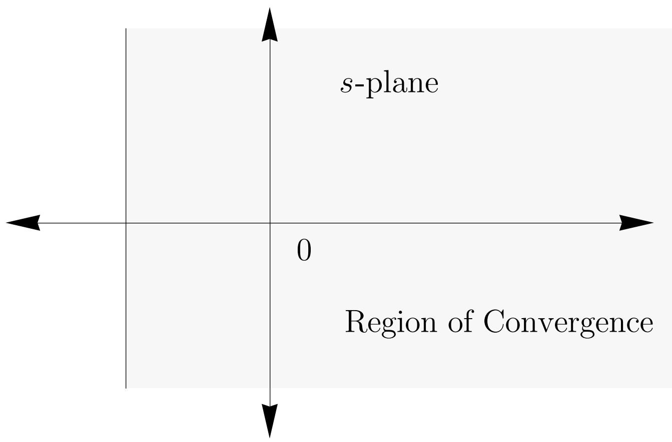

Figure 22: The ROC for the Z-Transform corresponding to a minimum phase system.

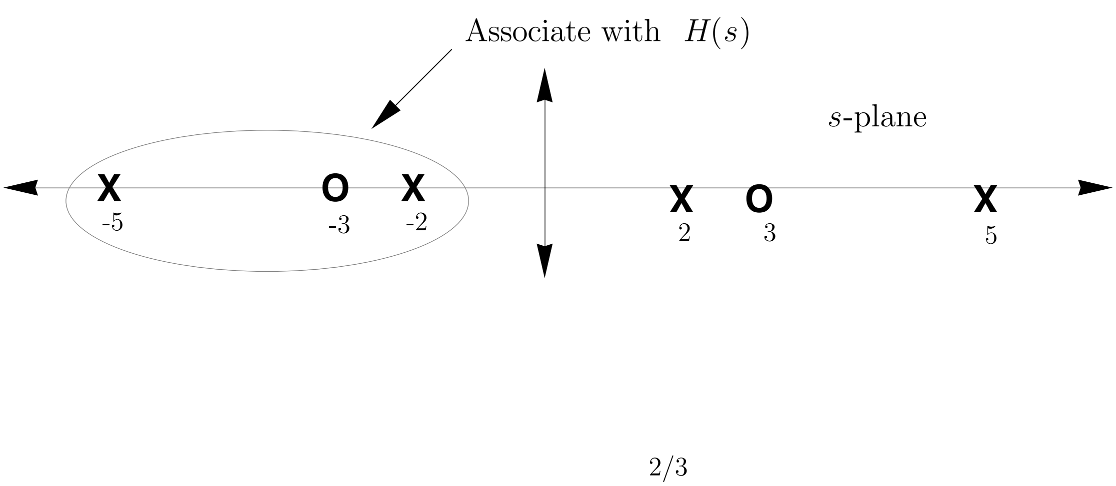

Figure 23: The pole zero locations for the example PSD.

In our example, $S_x(s)$ has poles at $s = \pm 2, \pm 5$ and zeros at $s = \pm 3$. Thus, the desired spectral factor is

$$H(s) = \frac{(3+s)}{(5+s)(2+s)}. \tag{275}$$

Below are the steps to factor a rational PSD $S_x(f)$ into $|H(f)|^2$, where $H(f)$ is the frequency response of a stable, causal, minimum-phase system:

1. *Convert $S_x(f)$ into pole/zero form (i.e., see (269)).*
2. *Convert to $S_x(s)$ using $s = j2\pi f$.*
3. *Split $S_x(s) = H(s)\ H^*(\cdot)|_{-s}$ so that all of the poles and zeroes of $H(s)$ are in the LHP.*

Again, this produces a rational spectral factor unique up to a unit magnitude scalar.

8.7.3 ♣ General Spectral Factorization Considerations

The Paley-Wiener Criterion *The PSD $S_x(f)$ may be factored into $S_x(f) = |H(f)|^2$, where $H(f)$ is the frequency response of a causal stable system if and only if*

$$\int_{-\infty}^{\infty} \frac{|\log(S_x(f))|}{1 + (2\pi f)^2} df < \infty. \tag{276}$$

♣ Eventually, include an engineering level proof and general factorization algorithm.

8.8 Cross-PSD and the Two-Filter Formula

The cross-PSD of two jointly-WSS processes is defined by

$$S_{xy}(f) = \mathbb{FT}\left\{R_{xy}(\tau)\right\}. \tag{277}$$

The symmetry of the cross-correlation function (i.e., $R_{xy}(\tau) = R^*_{yx}(-\tau)$), implies

$$S_{xy}(f) = \left(S_{yx}(f)\right)^*. \tag{278}$$

A useful result may be stated regarding the system in Figure 24, where two jointly-WSS processes $v(u,t)$ and $w(u,t)$ drive LTI systems so that $x(u,t) = h(t) {*} v(u,t)$ and $y(u,t) = g(t) {*} w(u,t)$. Using the techniques applied in Section 8.3, the following results are obtained

$$R_{xy}(\tau) = h(\tau) {*} R_{vw}(\tau) {*} g^*(-\tau) \tag{279}$$

$$S_{xy}(f) = H(f) G^*(f) S_{xy}(f). \tag{280}$$

Again, an example special case is $S_{xw}(f) = H(f) S_w(f)$.

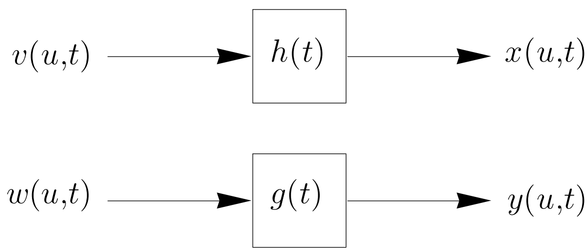

Figure 24: The set-up for the two-filter formula.

8.9 ♣ MMSE Estimation on $\mathcal{T} = \mathcal{R}$

This will be covered in class; only the results are listed here. In the results listed, it is assumed the the observed sequence $x(u,t)$ and the desired sequence $z(u,t)$ are jointly-WSS processes.

8.9.1 The Affine Constraint

Based on observing $\{x(u,\tau) : -\infty < \tau < \infty\}$, the best affine estimator of $z(u,t)$ is

$$\hat{z}(u,t) = g_{\text{opt}}(t) * (x(u,t) - m_x) + m_z, \tag{281}$$

where the Wiener filter is defined by

$$G_{\text{opt}}(f) = \frac{S_{z_0 x_0}(f)}{S_{x_0}(f)}. \tag{282}$$

The corresponding minimum MSE is

$$\text{MMSE} = \mathbb{E}\left\{|z(u,t)|^2\right\} - \mathbb{E}\left\{|\hat{z}(u,t)|^2\right\} \tag{283}$$

$$= \int_{-\infty}^{\infty} S_{z_0}(f) - \frac{|S_{z_0 x_0}(f)|^2}{S_{x_0}(f)} df. \tag{284}$$

8.9.2 The Causal Affine Constraint

Based on observing $\{x(u,\tau) : 0 < \tau < \infty\}$, the best affine estimator of $z(u,t)$ is

$$\hat{z}(u,t) = g_{\text{opt}}(t) * (x(u,t) - m_x) + m_z, \tag{285}$$

where the Causal Wiener filter is defined by

$$G_{\text{opt}}(f) = \mathbb{C}\left\{\frac{S_{z_0 x_0}(f)}{H^*_{x_0}(f)}\right\} \frac{1}{H_{x_0}(f)}, \tag{286}$$

where $H_{x_0}(f)$ is the causal, minimum phase factor of $S_{x_0}(f)$. The causal part operator is defined by

$$\mathbb{C}\left\{P(f)\right\} = \mathbb{FT}\left\{\mathrm{u}(\tau)\mathbb{FT}^{-1}\left\{P(f)\right\}\right\}, \tag{287}$$

The corresponding minimum MSE is

$$\mathrm{MMSE} = \mathbb{E}\left\{|z(u,t)|^2\right\} - \mathbb{E}\left\{|\hat{z}(u,t)|^2\right\} \tag{288}$$

$$= \int_{-\infty}^{\infty} S_{z_0}(f) - \left|\mathbb{C}\left\{\frac{S_{z_0 x_0}(f)}{H^*_{x_0}(f)}\right\}\right|^2 df. \tag{289}$$

8.10 ♣ Inverting Transforms the Easy Way

8.11 ♣ The Wiener Process

9 ♣ LTI/WSS Processing on $\mathcal{T} = \mathcal{R}_T$

Usually not covered in EE562a, but by now you could figure it out yourself!