Chapter 3

Second-order processes : L_2-theory

In this chapter we will study stochastic processes characterized by their first two moments or second-order processes. Such processes are termed $L_2(\mathbb{P})$ processes, i.e., processes for which $\mathbf{E}[|X_t|^2] < \infty$. It turns out that from the point of view of signal processing these assumptions are sufficient to yield useful results much in the same way as Fourier series and transforms do in the deterministic context. In fact second order or L_2 theory gives analogous results on representations of stochastic processes which are important not only in sampling and reconstruction of stochastic processes but also the related issue of filtering by which we reduce the effects of noise on signals. We will begin with some general results regarding covariance functions and then discuss the cases of discrete-time and continuous-time stochastic processes.

3.1 Properties of the Covariance

Let us first recall the definition of a second-order process. Sine there are many applications in which we need to work with complex-valued processes we assume that the processes take values in $\mathcal{C}$, i.e., they are complex-valued.

Definition 3.1.1 *A stochastic process $\{X_t\}_{t \in T}$ which takes values in $\mathcal{C}$ is said to be a second-order process if $\mathbf{E}[|X_t|^2] < \infty$ for all $t \in T$.*

As mentioned in Chapter 2, a second-order process $\{X_t\}$ is characterized in terms of its mean $m_t = \mathbf{E}[X_t]$ and its covariance $R(t,s) = \mathbf{E}[(X_t - m_t)(X_s - m_s)^*]$. A second-order process is said to be wide-sense stationary (w.s.s.) if $m_t = m$ a constant not depending on t and $R(t,s) = R(t-s)$ with $R(t) = R^*(-t)$ i.e. it is a symmetric function in the difference $t - s$.

Let us first study some elementary properties of the covariance function. Unless explicitly mentioned we will always take the mean of the process to be 0 without loss of generality. Also we take T to be continuous. The properties of the covariance function for discrete-time 2nd. order processes or sequences follows *mutatis mutandis*.

Proposition 3.1.1 *Let $\{X_t\}_{t \in T}$ be a complex valued second-order process with mean 0 and covariance $R(t,s)$. Then $R(t,s)$ satisfies the following properties:*

  i. $R(t,s) = R^*(s,t)$ and $R(t,t) \geq 0$.

ii. $R(t, s)$ is a non-negative definite function on $T \times T$ i.e. let $\{\lambda_i\}_{i=1}^N$ be any complex scalars and $\{t_i\}_{i=1}^N$ be any set of points in T. then:

$$\sum_{i=1}^{N} \sum_{j=1}^{N} R(t_i, t_j) \lambda_j^* \lambda_i \geq 0$$

for all N and complex scalars λ_i and points t_i.

iii. $|R(t, s)| \leq \sqrt{R(t, t)} \sqrt{R(s, s)}$

iv. Additivity i.e. if $R_1(t, s)$ and $R_2(t, s)$ are covariance functions then $R(t, s) = R_1(t, s) + R_2(t, s)$ is a covariance function.

v. Multiplicative property: i.e. if $R_1(t, s)$ and $R_2(t, s)$ are covariance functions then $R(t, s) = R_1(t, s) R_2(t, s)$ defines a covariance function.

vi. Given any function $\sigma(t)$ on T with values in C then $\sigma(t)\sigma^*(s)$ defines a covariance function. In particular if $\{C_i\}_{i=1}^n$ are positive constants and $\{\sigma_i(t)\}_{i=1}^n$ are continuous functions on T then $\sum_{i=1}^n C_i \sigma_i(t) \sigma_i^*(s)$ defines a covariance. If the limit as $n \to \infty$ of the above series is well defined in a $L_2(T)$ sense then the limit will be a covariance.

Proof: The proof of i) is a consequence of the definition.

For the proof of ii) define the r.v.

$$Y = \sum_{i=1}^{N} \lambda_i X_{t_i}$$

Then

$$\mathbf{E}[YY^*] = \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i R(t_i, t_j) \lambda_j^* \geq 0$$

The proof of iii) is just the Cauchy-Schwarz inequality.

For the proof of iv) let X_t be a Gaussian process with covariance $R_1(t, s)$ and Y_t be another Gaussian process independent of X_t with covariance $R_2(t, s)$. Then $Z_t = X_t + Y_t$ is a Gaussian process with covariance $R(t, s) = R_1(t, s) + R_2(t, s)$.

The proof of v) follows as above with $Z_t = X_t Y_t$.

For the proof of vi) define $X_t = \sigma(t)Z$ where Z is $N(0, 1)$ then X_t has the required covariance. The extension to finite sums and the limit is direct.

Corollary 3.1.1 If $\{X_t\}$ is w.s.s. then $R(\tau) = \mathbf{E}[X_t X_{t+\tau}^*]$ has the following properties: $R(0) \geq 0$, $R(\tau) = R^*(-\tau)$ and $|R(\tau)| \leq R(0)$.

We now study the continuity of a second-order process and show how the continuity properties can be obtained from the knowledge of the covariance function.

Let us first recall the definition of quadratic mean (q.m) or mean squared continuity.

Definition 3.1.2 A second-order stochastic process $\{X_t\}_{t \in T}$ is said to be q.m. continuous at t if :

$$\mathbf{E}[|X_{t+h} - X_t|^2] \overset{h \to 0}{\to} 0$$

If it is continuous for all $t \in T$ then we simply say that $\{X_t\}$ is q.m. continuous.

Proposition 3.1.2 *Let $\{X_t\}_{t \in T}$ be a second-order process with covariance $R(t, s)$ (wlog we take $m_t = \mathbf{E}[X_t] = 0$). Then*

a) $\{X_t\}$ is continuous at t in the q.m. if and only if $R(.,.)$ is continuous at the diagonal $R(t, t)$.

b) If $\{X_t\}$ is q.m. continuous (at every $t \in T$) then $R(.,.)$ is continuous at every point $(t, s) \in T \times T$.

c) If a non-negative definite function $R(t, s)$ on $T \times T$ is continuous at every diagonal point it is continuous everywhere on $T \times T$.

Proof: a) It $R(t, t)$ is continuous then

$$
\begin{aligned}
\mathbf{E}[|X_{t+h} - X_t|^2] &= R(t+h, t+h) - R(t, t+h) - R(t+h, t) + R(t, t) \\
&= (R(t+h, t+h) - R(t, t) - (R(t+h, t) - R(t, t)) - (R(t, t+h) - R(t, t)) \\
&\to 0 \ as \ h \to 0
\end{aligned}
$$

Conversely, if $\{X_t\}$ is q.m. continuous at t then:

$$
\begin{aligned}
|R(t+h, t+h') - R(t, t)| &= |\mathbf{E}[X_{t+h} X^*_{t+h'}] - \mathbf{E}[X_t X^*_t]| \\
&= |\mathbf{E}[(X_{t+h} - X_t) X^*_{t+h}] - \mathbf{E}[X_t (X_{t+h'} - X_t)^*]| \\
&\leq \sqrt{\mathbf{E}[|X_{t+h} - X_t|^2]\mathbf{E}[|X_{t+h'}|^2]} + \sqrt{\mathbf{E}[|X_{t+h'} - X_t|^2]\mathbf{E}[|X_t|^2]} \\
&\to 0 \ \ as \ h, h' \to 0
\end{aligned}
$$

b)

$$
\begin{aligned}
|R(t+h, s+h') - R(t, s)| &= \mathbf{E}[X_{t+h} X^*_{s+h'}] - \mathbf{E}[X_t X^*_s]| \\
&\leq \sqrt{\mathbf{E}[|X_{t+h} - X_t|^2]\mathbf{E}[|X_{s+h'}|^2]} + \sqrt{\mathbf{E}[|X_{s+h'} - X_s|^2]\mathbf{E}[|X_t|^2]} \\
&\to 0 \ \ as \ h, h' \to 0
\end{aligned}
$$

c) The proof of this part follows from parts a) and b) by noting that any non-negative definite function can be associated with the covariance of a second-order process.

3.2 Hilbert spaces associated with second-order processes

Second-order processes have a natural space on which they are defined. The fact that they are characterized by their second-order properties defined through their covariances gives rise to a Hilbert space $L_2(\mathbb{P})$, i.e. a linear vector space associated with random variables with finite second moments. We now recall some elementary facts about Hilbert spaces. A linear vector space is said to be complete if every Cauchy sequence has its limit in that space. Of course to speak of convergence we need the notion of a metric or 'distance' between two elements of that space. The metric is provided by the definition of a norm which should satisfy the properties below:

Definition 3.2.1 *The metric or distance measure, denoted $||.||$, defines a norm on a vector space if:*

i. For every $x \in X$, $||x|| \geq 0$ and the equality holds if and only if $x = 0$.

ii. For any scalar α, $||\alpha x|| = |\alpha|||x||$.

iii. If $x, y \in X$ then $||x + y|| \le ||x|| + ||y||$

Example: In $\Re^n$ the Euclidean distance metric denoted by $||x - y|| = \sqrt{\sum_1^n |x_i - y_i|^2}$ is a norm.

Definition 3.2.2 *A complete normed space is called a Banach space.*

Before giving the definition of a Hilbert space we need the notion of an inner-product.

Definition 3.2.3 *Given a vector space X, an inner-product in X, denoted by $[.,.]$, is a scalar (in general complex if X is)associated between any two elements $x, y \in X$ with the following properties:*

i. $[x, x] = ||x||^2$ i.e. the inner-product of an element with itself defines a norm.

ii. For any $x, y, z \in X$ and any scalars (could be complex) a and b:

$$[ax + by, z] = a[x, z] + b[y, z]$$

i.e. it is linear w.r.t. the first argument.

iii. $[x, y] = [y, x]^*$

We now define a Hilbert space.

Definition 3.2.4 *A complete inner-product space is called a Hilbert space.*

Example: In $\Re^n$ an inner-product is given by the usual scalar product defined by $[x, y] = \sum_1^n x_i y_i$.

If we consider the space of complex-valued square-integrable functions on $[0, T]$ it can be readily verified that an inner-product is given by $[f, g] = \int_0^T f(t) g^*(t) dt$. It is important to note that the inner-product (and hence the norm) need not be uniquely defined i.e. there could be several norms defined on a space. However if the space is finite-dimensional it can be shown that all norms are equivalent.

Having briefly discussed the notion of a Hilbert space we now show how we can associate a Hilbert space with second-order processes.

Let us first begin by seeing how these spaces arise. Let X and Y be two r.v's (mean 0 without loss of generality) with finite second-moments. Then the covariance operator:

$$cov(X, Y) = \mathbf{E}[XY^*]$$

defines an inner-product on the space of r.v's with finite second moment where the space H is taken to be the space of all linear combinations of second-order r.v's where

$$Y = \sum_{k=1}^n a_k X_k$$

and their limits in the mean square when they exist. Noting by the fact that if a r.v. has variance 0 it is almost surely a constant (0 in the case of mean 0 r.v's) then two elements are equivalent (in a probabilistic sense) if $\mathbf{E}[|X - Y|^2] = 0$. We have seen that if a r.v. is Cauchy in the mean square then

the limit exists and is a second-order r.v. and thus the space generated by all linear combinations of second-order r.v's becomes a Hilbert space (denoted $L_2(\mathbb{P})$ equipped with the inner-product

$$[X, Y]_{L_2(\mathbb{P})} = \mathbf{E}[XY^*]$$

The subscript on the inner-product is used to differentiate between the inner-product in $L_2(\mathbb{P})$ from the inner-product in $\Re^n$.

Associated with the space $L_2(\mathbb{P})$ we can define Hilbert subspaces. For example, given a sequence of second-order r.v's $\{X_n\}$ we can define a Hilbert sub-space generated by all linear combinations of the past of the process at time n denoted by : H_n^- defined by

$$\{Y : Y = \ linear\ span\ of\ X_{n-1}, X_{n-2}.....\}$$

and of the future denoted by H_n^+

$$\{Y : Y = \ linear\ span\ of\ X_{n+1}, X_{n+2}, ...\}$$

where the equalities are understood in the mean-squared sense.

Let us now study some consequences of the Hilbert spaces and their relation to representation of w.s.s. processes.

The geometry of these spaces plays a very important role in the problems of linear filtering, prediction and smoothing (linear estimation theory) which will be discussed in detail elsewhere.

Having defined the Hilbert space $L_2(\mathbb{P})$ with inner-product $[X, Y]$ (we drop the sub-script $L_2(\mathbb{P})$ when the meaning is unambiguous) we say that two elements X and Y are orthogonal if they are uncorrelated i.e. $[X - m_x, Y - m_y] = 0$ where $\mathbf{E}[x] = m_x$ and $\mathbf{E}[Y] = m_y$. As we saw in chapter 1 the conditional expectation can be viewed as a projection of one r.v. onto the other. If the two r.v's are jointly Gaussian then the conditional expectation is a linear function of the r.v. on which we condition.

In the context of L_2 r.v's we denote the linear projection of one r.v. onto a given subspace spanned by others as $\hat{\mathbf{E}}(X|H_1)$ knowing that if X and the r.v's which define H_1 are jointly Gaussian then the projection is indeed a conditional expectation.

Let $\{X_n\}$ be a w.s.s. sequence. Define $H_n = \overline{span}(\cdots, X_{n-1}, X_n)$, i.e., the linear space generated by the r.v.'s $\{X_k\}$ up to time n.

Define $\hat{X}_{n-1} = \hat{\mathbf{E}}(X_n|H_{n-1})$, i.e. the projection of X_n onto its past. Also define $\nu_n^x = X_n - \hat{X}_{n-1}$, the "error". Then from the projection theorem we know that ν_n^x is uncorrelated with all $\{X_k\}, k = n - 1, n - 2, n - 3, \ldots$, i.e $\nu_n^x \perp H_{n-1}$.

By definition:

$$X_n = \hat{X}_{n-1} + \nu_n^x$$

Define: $H_{-\infty} = \cap_{n \in Z} H_n$. This is the subspace which consists of the remote past of $\{X_n\}$. Similarly define $H_\infty = \overline{\cup_{n \in Z} H_n}$. This is the subspace associated with the entire sequence $\{X_n\}$.

Definition 3.2.5 *The process $\{\nu_n^x\}$ is called the innovations process.*

We drop the superscript x when the context is clear.

Definition 3.2.6 *A stochastic process $\{X_n\}$ is said to be deterministic if and only if for all n, $X_n \in H_{n-1}$. If this is not the case then we call the process non-deterministic. If $H_{-\infty} = \{0\}$, then $\{X_n\}$ is said to be purely non-deterministic or regular.*

Define $N_n = \overline{span}\{\dots, \nu_{n-1}, \nu_n\}$, i.e., N_n is the space spanned by the innovations process associated with $\{X_n\}$. Similarly define $N_{-\infty} = \bigcap_n N_n$ and $N_\infty = \overline{\bigcup_n N_n}$.

Lemma 3.2.1 *For any $m < n$ we have:*
a) $H_n = H_m + span(\nu_{m+1}, \dots, \nu_n)$

b) $H_m \perp span(\nu_{m+1}, \dots, \nu_n)$

c) $H_n = H_{-\infty} + N_n$, $N_{-\infty} = \{0\}$ *and* $H_{-\infty} \perp N_\infty$.

Proof: The proofs of parts a) and b) are direct from the definitions of ν_n and the fact that they are errors associated with projections. we will only prove the last part.

Let $h \in H_{-\infty}$, then by definition $h \in H_{n-1} \forall n \in Z$. So $h \perp \nu_n \ \forall n \in Z$. Now by the continuity of the inner-product, $h \perp \overline{span}\{\dots, \nu_{-1}, \nu_0, \dots\}$. so it follows that $H_{-\infty} \perp N_\infty$.

Now for any $p \in N_{-\infty}, \Rightarrow p \in N_n \subset H_n \ \forall n$ so $p \in H_{-\infty} \bigcap N_{-\infty}$. This implies that $p = 0$ since $N_{-\infty} \subset H_{-\infty}$. Also by the definition of N_n we have $H_n \supset H_{-\infty} + N_n$. Now for ant $h \in H_n$ we have:

$$
\begin{aligned}
\hat{\mathbf{E}}[h|H_{-\infty} + N_n] &= \hat{\mathbf{E}}[h|H_{-\infty}] + \hat{\mathbf{E}}[h|N_n](since\ H_{-\infty} \perp N_n) \\
&= \lim_{m \to -\infty} \hat{\mathbf{E}}[h|H_m] + \lim_{m \to -\infty} \hat{\mathbf{E}}[h|\nu_{m+1}, \cdots, \nu_n] \\
&= \lim_{m \to -\infty} \hat{\mathbf{E}}[h|H_n] \\
&= h
\end{aligned}
$$

This means that $H_n \subset H_{-\infty} + N_n$ and hence the result follows.

We now prove an important result associated with the representation of w.s.s. sequences due to Wold.

Theorem 3.2.1 *(Wold decomposition)*
Let $\{X_n\}$ be a $\Re$ valued w.s.s. process and $\{\nu_n\}$ be the corresponding innovations process. Then:

a) $X_n = U_n + V_n$ *where* $U_n = \hat{\mathbf{E}}[X_n|N_n]$, $V_n = \hat{\mathbf{E}}[X_n|H_{-\infty}]$ *and* $U_n \perp V_n, \ \forall n$.

b) *The process $\{U_n\}$ is a w.s.s. process and has the representation:*

$$U_n = \sum_{k=0}^{\infty} a_k \nu_{n-k}$$

where $\mathbf{E}[\nu_n \nu_m] = \sigma^2 \delta_{n,m}$. Furthermore,

$$0 \le \sum_{k=0}^{\infty} a_k^2 < \infty$$

and $a_0 = 1$

c) *$\{V_n\}$ is a deterministic w.s.s. process and*

$$\overline{span}(\cdots, V_{n-1}, V_n) = H_{-\infty} \ \forall n$$

Proof:

a) The proof of this part is immediate from the fact that $H_n = N_n + H_{-\infty}$.

b) By definition $\hat{\mathbf{E}}[\nu_n|H_m] = 0 \quad \forall\, n > m$. Now since $U_n \in N_n$ there exist a_n such that $U_n = \sum_{k=0}^{\infty} a_k \nu_{n-k}$ where:

$$
\begin{aligned}
a_k \sigma^2 &= \mathbf{E}[U_k \nu_{n-k}] \\
&= \hat{\mathbf{E}}[\hat{\mathbf{E}}[X_k|N_k]\nu_{n-k}] \\
&= \hat{\mathbf{E}}[\hat{\mathbf{E}}[X_k \nu_{n-k}|N_k]] \\
&= \hat{\mathbf{E}}[X_k \nu_{n-k}] = \mathbf{E}[X_0 \nu_{-k}]
\end{aligned}
$$

Hence:

$$
\begin{aligned}
\sigma^2 &= \mathbf{E}[X_0 \nu_0] \\
&= \mathbf{E}[(\nu_0 + \hat{\mathbf{E}}[X_0|H_{-1}])\nu_0] \\
&= \mathbf{E}[\nu_0^2]
\end{aligned}
$$

c) Now $N_n \subset H_n$, $V_m \in H_{-\infty} \subset H_n \; \forall m, n \in Z$.

Furthermore for all $m < n$, $X_m = U_m + V_m \in N_n + \overline{span}(..., V_{n-1}, V_n)$ and $span(..., V_{n-1}, V_n) \perp N_n$. Also since $N_n + \overline{span}(..., V_{n-1}, V_n) = H_n = N_n + H_{-\infty}$ and the orthogonal complement of N_n is unique it implies that $\overline{span}(..., V_{n-1}, V_n) = H_{-\infty}$. This completes the proof.

Remark 3.2.1 *The Wold decomposition holds for $\Re^n$- valued processes also. In that case a_k is a matrix and $\sigma^2 = \Sigma$ is a matrix. In that case if Σ is singular , then $A_0\Sigma = \Sigma = \Sigma A_0^*$. If Σ is non-singular then $A_0 = I$.*

The significance of the Wold decomposition becomes clear if we consider the following prediction problem.

Let $\{X_n\}$ be a 0 mean w.s.s. process. Denote $\hat{X}_{n|0} = \hat{\mathbf{E}}[X_n|H_0]$ where $H_0 = \overline{span}(..., X_{-1}, X_0)$ i.e. we want to predict the value of X_n knowing all the past values up to 0 i.e. prediction of X n steps into the future if we regard 0 as the present.

From the Wold decomposition:

$$
X_n = V_n + \sum_{k=0}^{\infty} a_k \nu_{n-k}
$$

with $\sum_{k=0}^{\infty} |a_k|^2 < \infty$.

Hence:

$$
\begin{aligned}
\hat{X}_{n|0} &= \hat{\mathbf{E}}[V_n + \sum_{k=0}^{\infty} a_k \nu_{n-k}|H_0] \\
&= \hat{\mathbf{E}}[V_n|H_0] + \sum_{k=0}^{\infty} a_k \hat{\mathbf{E}}[\nu_{n-k}|H_0] \\
&= V_n + \sum_{k=0}^{\infty} a_k \hat{\mathbf{E}}[\nu_{n-k}|N_0] \\
&= V_n + \sum_{k=n}^{\infty} a_k \nu_{n-k}
\end{aligned}
$$

The error in prediction is:

$$\mathbf{E}[X_n - \hat{X}_{n|0}]^2 = \sigma_n^2 = \sigma^2 \sum_{k=0}^{n-1} a_k^2$$

where $\sigma^2 = E[\nu_n^2]$. If $\{X_n\}$ is purely non-deterministic or regular then $V_n = 0$ and also:

$$\lim_{n\to\infty} \sigma_n^2 = \sigma^2 \sum_{k=0}^{\infty} a_k^2 = \mathbf{E}[X_n^2]$$

implying that if we try to predict further and further out we basically obtain 0 as the prediction since the error is the same as the variance of X_n.

Of course, the Wold decomposition depends on knowing the coefficients $\{a_k\}$. In practice we can measure the covariance $\mathbf{E}[X_n X_{n+k}^*] = R(k)$ (assuming 0 means). We will see later how this information allows us to determine the $\{a_k\}$. We conclude our discussion on the Hilbert space structure for w.s.s. processes here.

In the above we have considered Hilbert spaces associated with discrete-time second-order processes. In a similar way we now consider spaces generated by q.m continuous processes. To do so we have to consider the continuous-time equivalent of a linear mapping on the process with is defined through an integral. For this we need to define stochastic integrals and give conditions for them to exist.

3.2.1   Stochastic integrals of second-order processes

As mentioned above the analog of a linear transformation associated with a q.m continuous process is that of an integral defined w.r.t. to a sample-path of a stochastic process. We used them in Chapter 2 (in the context of ergodic processes assuming that they were well defined). These integrals arise when we want to study the output of a linear system whose input is a q.m. continuous process. Specifically, we would like to define processes of the type

$$Y_t = \int_0^t h(t,s) X_s ds$$

where $h(t,s)$ is the impulse response of a linear system. Of course the limits of integration need not be $[0,t)$ and could be semi-infinite intervals such as $(-\infty, t]$, $[0, \infty)$ etc..

Let $[a,b]$ be a finite-interval in T. Let $\{t_i\}_{i=1}^n$ be points of T such that :

$$J_n = \{t_i : a = t_1 < t_2 < \ldots < t_{n-1} < t_n = b\}$$

Then we define (if it exists):

$$\int_a^b f(s) X_s ds = \lim \ in \ q.m. \sum_{k=1}^{n-1} f(\tau_k) X_{\tau_k}(t_{k+1} - t_k); \quad t_k \leq \tau_k < t_{k+1} \qquad (3.2.\ 1)$$

where in the limit as $n \to \infty$ is such that $\sup |t_n - t_{n-1}| \to 0$ and the limit is independent of the choice of J_n.

Note as defined the partial sums are Riemann sums and thus the limit is a q.m. limit of Riemann sums or the integral is defined as a Riemann integral.

If X_t is q.m. continuous then we know that the covariance function $R(t,s)$ is continuous on $T \times T$ and hence the limit will exist in the q.m. sense if and only if the following integral exists (in a Riemann sense) :

$$\int_a^b \int_a^b f(t)f(s)^* R(t,s) dt ds < \infty \qquad (3.2.\ 2)$$

and since $[a,b]$ is finite it is sufficient if $f(t)$ is piece-wise continuous on $[a,b]$. The extension to the one or both end-points being infinite is to consider the finite intervals and then require the limit of the integral (3.2.2) to exist for the corresponding limits.

Remark: Stochastic integrals can also be defined pathwise (i.e. for almost every sample path) but to do so we need to introduce the notion of Lebesgue integrals which is beyond the scope of the course. Furthermore note that the integrand above is taken to be non-random since all we are defining is a linear operation on $\{X_t\}$. The extension of to random integrands is the purview of Ito stochastic integrals and once again beyond the scope of the course.

Hence, since by definition the integrals are q.m limits of sums of second-order processes the limit is a second-order r.v.. Hence we can associate a Hilbert space with q.m. continuous processes and Hilbert subspaces for defining the 'past' and 'future' as in the case of sequences. In particular we have:

$$\mathbf{E}[\int_a^b f(s)X_s ds (\int_a^b g(u)X_s du)^*] = \int_a^b \int_a^b f(t)g(s)^* R(t,s) dt ds = [AX, BX]_{L_2(\mathbb{P})}$$

where the 'operator' $AX = \int_a^b f(s)X_s ds$ and similarly for B.

3.2.2 More on the Hilbert space structure associated with second order processes

The Wold decomposition provides a way of representing any w.s.s. sequence in terms of uncorrelated sequences $\{\nu_n\}$. This provides an orthogonal decomposition for for the sequence $\{X_n\}$ where the inner-product between two r.v's X and Y is defined as: $[X, Y] = cov(X, Y)$.

We can develop a similar framework for any sequence of second order r.v's. This is the basis of linear estimation theory that we will see in the next chapter in detail.

Let $\{X_k\}_{k=0}^{\infty}$ be a 0 mean[1] second order sequence of r.v's with $var(X_k) = \sigma^k < \infty$.
Define:

$$L_k^X = span\{X_0, X_1, \cdots, X_n\} = \{Y_n : Y_n = \sum_{k=0}^{n} a_k X_k, a_k \in \Re \ with \ |a_k|^2 < \infty\}$$

Then for each $n < \infty$ the space L_n^X forms a linear space of 2-nd order r.v's that are formed from a linear combination of $\{X_k\}_{k=0}^{n}$.

Define $\hat{X}_0 = 0$ and $\nu_0 = X_0$. Then we have:

$$X_0 = \nu_0 \in L_0^X$$

.

Define $\hat{X}_{n|n-1} = Projection \ of \ X_n \ onto \ L_{n-1}^X$. Then in particular:

$$\hat{X}_{1|0} = \frac{[X_1, X_0]}{||X_0||^2} X_0 = \frac{\mathbf{E}[X_1 X_0]}{\sigma_0^2} X_0$$

---

[1]It is easy to extend the result to non-zero mean sequences

Denote $a_0 = \frac{\mathbf{E}[X_1 X_0]}{\sigma_0^2}$ Define $\nu_1 = X_1 - \hat{X}_{1|0}$. Then by definition of the projection $\nu_1 \perp X_0$, i.e., $[X_0, \nu_1] = \mathbf{E}[X_0 \nu_1] = \mathbf{E}[X_1 X_0] - \mathbf{E}[X_1 X_0] = 0$ or X_0 and ν_1 are uncorrelated. Now by the orthogonal decomposition:

$$X_1 = X_0 + \nu_1 = a_0 \nu_0 + \nu_1$$

and ν_0 and ν_1 are uncorrelated. Moreover, it is easy to check:

$$
\begin{aligned}
\mathbf{E}[X_1^2] &= \mathbf{E}[a_0^2 X_0^2] + \mathbf{E}[X_1^2] - \frac{(\mathbf{E}[X_1 X_0])^2}{\sigma_0^2} + \\
&= \sigma_1^2 + a_0^2 \sigma_0^2 - a_0^2 \sigma_0^2 = \sigma_1^2
\end{aligned}
$$

Hence, by definition: $L_1^X = L_1^\nu$ and L_1^ν is spanned by uncorrelated r.v's.

Now repeating this argument we define:

$$\nu_n = X_n - \hat{X}_{n|n-1}$$

By iterating it can readily be seen that:

$$X_n = \sum_{k=0}^{n} a_{k-1} \nu_k$$

and moreover $L_n^X = L_n^\nu \ \forall \ n$.

In other words we have represented X_n as the sum of $n + 1$ uncorrelated r.v's ν_n. or we have obtained an orthogonal decomposition for the Hilbert space L_n^X much as in $\Re^n$ for n-dimensional vectors through the Gram-Schmidt procedure.

Remark 3.2.2 *The sequence of uncorrelated r.v's $\{\nu_n\}$ is called the* innovations sequence *(or process) associated with $\{X_n\}$. It amounts to the* extra *information needed over and above knowing $(X_0, X_1, X_2 \cdots, X_{n-1})$ to reconstruct X_n in an L_2 sense.*

If the sequence of r.v's is jointly distributed gaussian then the equality can be interpreted in an a.s. sense as the linear projection is equivalent to the conditional expectation.

3.2.3 Differentiation of q.m. continuous processes

In order to develop a complete second-order calculus associated with q.m. continuous processes we need to define derivatives. As in the case of integrals and other operations since we are working with second-order processes we define them in the q.m. sense and seek conditions on the covariance function for them to exist.

Definition 3.2.7 *A q.m. continuous process $\{X_t\}$ is said to be differentiable at t if the following limit (denoted $\dot{X}_t$) exists in the q.m. sense:*

$$\dot{X}_t = \lim_{h \to 0} \ in \ q.m. \frac{X_{t+h} - X_t}{h} \tag{3.2. 3}$$

Remark: Since the derivative is the q,m. limit of second-order processes $X_t^h = \frac{X_{t+h} - X_t}{h}$ the limit is a second-order process and thus belongs to the Hilbert space associated with X.

Proposition 3.2.1 *If the covariance function $R(t, s)$ of a q.m. continuous stochastic process $\{X_t\}$ has mixed partial derviatives $\frac{\partial^2 R(t,s)}{\partial t \partial s}$ at the diagonal point (t, t) then the derivative $\dot{X}_t$ exists in the q.m. sense at $t \in T$ and conversely. If the condition holds for all $t \in T$ then we say the process is differentiable in the q.m. and then:*

$$cov(\dot{X}_t \dot{X}_s) = \frac{\partial^2 R(t, s)}{\partial t \partial s}$$

Proof: To show that the derivative exists in the q.m. case it is sufficient to show that:

$$\lim_{h,u \to 0} \mathbf{E}[\frac{(X_{t+h} - X_t)}{h} \frac{(X_{t+u} - X_t)^*}{u}] = \frac{\partial^2 R(t, t)}{\partial t^2}$$

by Problem 15 Chapter 2 on an equivalent necessary and sufficient condition for q.m. convergence.
Now:

$$
\begin{aligned}
\lim_{h,u \to 0} \mathbf{E}[\frac{(X_{t+h} - X_t)}{h} \frac{(X_{t+u} - X_t)^*}{u}] &= \lim_{h,u \to 0} \frac{R(t+h, t+u) - R(t+h, t) - R(t, t+u) + R(t, t)}{hu} \\
&= \lim_{h \to 0} \frac{1}{h} \left(\frac{\partial R(t+h, t)}{\partial t} - \frac{\partial R(t, t)}{\partial t} \right) \\
&= \frac{\partial^2 R(t, t)}{\partial t^2}
\end{aligned}
$$

Note in the above we first took the limit as $u \to 0$ and then $h \to 0$ but it can clearly be seen that the order does not matter.

From the definition it is clear that $\frac{\partial^2 R(t,s)}{\partial t \partial s}$ defines a covariance since it is the limit of covariance functions and the proof is complete.

Remark: If the process is w.s.s. then $R(t, s) = R(t - s)$ for $t > s$ and so the covariance of $\dot{X}_t$ denoted by $R^{(1)}(t) = -\frac{d^2}{dt^2} R(t)$ where the negative sign occurs because the partial serivative w.r.t. to s has a negative sign because of the negative sign associated with s.

As we have seen that under the condition that the covariance matrix is twice differentiable we can define the derivative of a 2nd. order stochastic process which itself is a second-order process. Thus if we formally think of $dX_s = \dot{X}_s ds$ then we can define the stochastic integral w.r.t to dX_s i.e. $\int_a^b f(s) dX_s$ is well defined as in Section 3.2.1 provided $\int_a^b \int_a^b f(t) f(s) \frac{\partial^2 R(t,s)}{\partial t \partial s} dt ds < \infty$.

A natural question to ask is how many orders of the derivatives for 2nd order processes are defined in a q.m. sense? Clearly this is related to how many derivatives of the covariance function exist. We will study this question a little later in this chapter although from above we can infer that the process would be twice differentiable (in the mean squared sense) if the fourth partial derivative of the covariance exists.

3.3 Spectral Theory of Wide Sense Stationary Processes

In this section we study the spectral theory for wide sense stationary processes. Spectral theory deals with the frequency content of signals and this in turn is related to Fourier representations of such processes. Recall a second-order process $\{X_t\}$ is said to be wide sense stationary (w.s.s) if :

11

i. $\mathbf{E}[X_t] = m$

ii. $R(t, s) = R(|t - s|)$

Let us first study the discrete-time case.

3.3.1 Herglotz Theorem for discrete-time w.s.s. processes

In the beginning of this Chapter we saw that covariance functions possess the property of non-negative definiteness. It is usually difficult to check for this condition from the definition. However, the non-negative definiteness of the covariance functions for w.s.s. processes can easily be established due to the following result due to Herglotz (called the Bochner theorem in the continuous-time case). Essentially it amounts to the property that the Fourier transform of non-negative definite functions is non-negative (implying that the transform is real). We state and prove this result below.

Proposition 3.3.1 *Let $R(n)$ be the covariance function of a w.s.s sequence. Then there exists a finite measure $F(\lambda)$ on $(-\frac{1}{2}, \frac{1}{2}]$ called the spectral measure such that*

$$R(n) = \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{i2\pi\lambda n} dF(\lambda) \qquad (3.3.\ 4)$$

where $F(\lambda)$ is non-decreasing on $(-\frac{1}{2}, \frac{1}{2}]$ with $F(-\frac{1}{2}) = 0$ and $F(\frac{1}{2}) = R(0)$. Moreover $F(\frac{1}{2})$ is unique.

If in addition $\sum_{n=0}^{\infty} |R(n)| < \infty$ then $F(\lambda)$ is differentiable in λ and

$$S(\lambda) = \frac{dF(\lambda)}{d\lambda}$$

is called the spectral density and $S(\lambda) \geq 0$ on $(-\frac{1}{2}, \frac{1}{2}]$.

Proof: We will prove the result under the assumption that a spectral density exists i.e. $\sum_{n=0}^{\infty} |R(n)| < \infty$ since the general case requires knowledge of advanced results from analysis.

First note by the non-negative definiteness of $R(n)$ defining:

$$\begin{aligned}
S_N(\lambda) &= \frac{1}{N} \sum_{j=0}^{N-1} \sum_{k=0}^{N-1} R(j-k) e^{-i2\pi\lambda(j-k)} \\
&= \sum_{k=-N+1}^{N-1} R(k) e^{-i2\pi\lambda k} \left(1 - \frac{|k|}{N}\right) \\
&\geq 0
\end{aligned}$$

Then it is easy to see that:

$$\begin{aligned}
\int_{-\frac{1}{2}}^{\frac{1}{2}} e^{in2\pi\lambda} S_N(\lambda) d\lambda &= R(n) \left(1 - \frac{|n|}{N}\right) \quad ; |n| \leq N \\
&= 0 \quad |n| > N
\end{aligned}$$

Now:

$$|S_N(\lambda)| \leq \sum_{n=-N}^{N} |R(n)| \leq \sum_{n=-\infty}^{\infty} |R(n)| < \infty$$

and therefore the limit $\lim_{N\to\infty} S_N(\lambda)$ exists and therefore for every n

$$R(n) = \lim_{N\to\infty} R(n)(1 - \frac{|n|}{N}) = \lim_{N\to\infty} \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{in2\pi\lambda} S_N(\lambda)d\lambda = \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{in2\pi\lambda} S(\lambda)d\lambda$$

where the lim on the r.h.s. is by bounded convergence.

Since $S(\lambda)$ is the limit of $S_N(\lambda)$ which is non-negative for all N it follows that the limit is non-negative and the representation of $R(n)$ as the Fourier coecfficient of $S(\lambda)$ is proved.

Now define:

$$F(\lambda) = \int_{-\frac{1}{2}}^{\lambda} S(\nu)d\nu$$

then $F(\lambda)$ is increasing with $F(-\frac{1}{2}) = 0$ and $F(\frac{1}{2}) = R(0)$ where $F(\lambda)$ is now called the spectral distribution or measure as in the first part of the statement. As mentioned earlier the existence of the derivative $S(\lambda)$ is not necessary for the Herglotz theorem to hold with $F(\lambda)$ having the given propeties.

Uniqueness states that if F and F' are two spectral measures corresponding to a given covariance function $R(k)$ then $F(\lambda) = F'(\lambda)$ on $(-\frac{1}{2}, \frac{1}{2}]$. This follows from the fact that if $f(\lambda) = \mathbf{1}_{(a,b]}(\lambda)$ with $(a,b] \subset (-\frac{1}{2}, \frac{1}{2}]$ then we can expand $f(\frac{1}{2})$ in a Fourier series i.e,:

$$f(\lambda) = \sum_{-\infty}^{\infty} f_n e^{in2\pi\lambda}$$

And hence:

$$\begin{aligned}
F(a,b] &= \int_{-\frac{1}{2}}^{\frac{1}{2}} f(\omega)dF(\omega) = \lim_{N\to\infty} \sum_{-N}^{N} f_n \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{in2\pi\lambda}dF(\lambda) \\
&= \lim_{N\to\infty} \sum_{-N}^{N} f_n \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{in2\pi\lambda}dF'(\lambda) \\
&= \int_{-\frac{1}{2}}^{\frac{1}{2}} f(\lambda)dF'(\lambda) = F'(a,b]
\end{aligned}$$

The limiting operation is valid by dominated convergence since F and F' are finite spectral measures. Since the above is true for all $(a,b]$ in $(-\frac{1}{2}, \frac{1}{2}]$ the proof is done i.e. $F(\lambda) = F'(\lambda)$ for all $\omega \in (-\frac{1}{2}, \frac{1}{2}]$.

Remark: In practical terms the condition that $\sum_{n=0}^{\infty} |R(n)| < \infty$ not being satisfied implies that the spectral distribution $F(\lambda)$ will contain discontinuities or in otherwords the spectral density will contain delta functions(called spectral lines). More generally the lack of summability of the covariance functions is related to a phenomenon of *long-range dependence* which we will see at the end of this chapter.

As an example of the lack of summability, consider the w.s.s. sequence defined by:

$$X_n = X \cos n2\pi\lambda_0 + Y \sin n2\pi\lambda_0$$

where X, Y are independent Gaussian r.v's with mean o and variance 1.

Then:

$$R(m) = \cos m 2\pi \lambda_0$$

and (formally)

$$dF(\lambda) = (\delta(\lambda - \lambda_0) + \delta(\lambda + \lambda_0))\, d\lambda$$

Suppose the condition $\sum_0^\infty |R(n)| < \infty$ is satisfied i.e. a spectral density exists then the spectral density is just given by the Fourier series on $(-\frac{1}{2}, \frac{1}{2}]$ with Fourier coefficients $R(n)$ i.e.

$$S(\lambda) = \sum_{-\infty}^{\infty} R(n)e^{-in2\pi\lambda}$$

and in the case of real processes noting that $R(n) = R(-n)$ we obtain:

$$S(\omega) = R(0) + 2\sum_{n=1}^{\infty} R(n)\cos(n2\pi\lambda)$$

We are now in a position to study the questions related to the output of a linear, time-invariant, causal and stable system whose input is a w.s.s. process. We will assume that the input process has a spectral density.

The requirement of stability just implies that if h_k denotes the discrete impulse response function of a causal system then $\sum_n |h_n| < \infty$ which in turn implies that the transfer function $H(z) = \sum_{n=0}^{\infty} h_n z^n$ is analytic on the entire unit circle. This is equivalent to saying that the transfer function has when viewed as a rational function of z (i.e. a numerator polynomial in z and denominator polynomial in z) has no "poles" (roots of the denominator polynomial) inside or on the unit circle. In practical terms, lack of stability means that the output will grow without bound and hence it is unrealistic to assume any steady state or w.s.s. property will hold for the output process. We give a formal definition of stable signals and the associated notion of finite-energy signals below.

Definition 3.3.1 *A discrete-time signal or sequence $\{h_k\}$ is said to be stable if it satisfies*

$$\sum_{k=-\infty}^{\infty} |h_k| < \infty \tag{3.3. 5}$$

A discrete-time signal $\{h_k\}$ is said to be a finite energy signal if

$$\sum_{k=-\infty}^{\infty} |h_k|^2 < \infty \tag{3.3. 6}$$

Remark: Causality imposes the condition that $h_k = 0$ for $k < 0$. In the parlance of analysis stable signals belong to ℓ_1 while finite energy signals belong to ℓ_2. Furthermore the two sided z-transform of stable signals defined with $H(z)$ analytic on $|z| = 1$

Let $\{X_n\}$ be a w.s.s. sequence (0 mean) with spectral density $S_X(\lambda)$. Let $\{Y_n\}$ denote the output process when $\{X_n\}$ is the input to a linear, time-invariant, causal (LTIC) and stable system with impulse response function $\{h_k\}$. Then,

$$Y_n = \sum_{-\infty}^{n} h_{n-k}X_k; \quad -\infty < n < \infty \tag{3.3. 7}$$

Note causality implies that $h_k = 0$ for $k < 0$ and stability implies that $H(z)$ is analytic for $|z| = 1$.

Proposition 3.3.2 *If a w.s.s. sequence (0 mean) with spectral density $S_X(\lambda)$ is the input to a linear, time-invariant, causal and stable system, then the output process $\{Y_n\}$ is a w.s.s. process with spectral density denoted $S_Y(\lambda)$ and*

$$S_Y(\lambda) = |H(e^{i2\pi\lambda})|^2 S_X(\lambda) \tag{3.3. 8}$$

where $H(z)$ denotes the z-transform of the impulse response function defined by:

$$H(z) = \sum_{n=0}^{\infty} h_n z^n$$

.

Proof:

Since $Y_n = \sum_{-\infty}^{n} h_{n-k} X_k = \sum_{k=0}^{\infty} h_k X_{n-k}$ then $\mathbf{E}[Y_n] = 0$ Define $R_Y(n+k, n) = \mathbf{E}[Y_{n+k} Y_n^*]$. Hence using the w.s.s. of $\{X_n\}$ we obtain,

$$R_Y(n+k, n) = \sum_{j=0}^{\infty} \sum_{m=0}^{\infty} h_j h_m R_X(k+m-j)$$

Multiply both sides by $e^{-i2\pi\lambda k}$ and sum. Then we have:

$$
\begin{aligned}
\sum_{k=-\infty}^{\infty} R_Y(n+k, n) e^{i2\pi\lambda k} &= \sum_{k=-\infty}^{\infty} \sum_{j=0}^{\infty} \sum_{m=0}^{\infty} h_j h_m R(k+m-j) e^{-i2\pi\lambda k} \\
&= \sum_{l=-\infty}^{\infty} \sum_{j=0}^{\infty} \sum_{m=0}^{\infty} h_j e^{-i2\pi\lambda j} h_m e^{i2\pi\lambda m} R(l) e^{-i2\pi\lambda l} \\
&= H(e^{i2\pi\lambda}) H(e^{-i2\pi\lambda}) S_X(\lambda)
\end{aligned}
$$

Hence, since the series on the lhs does not depend on n it implies that $R_Y(n+k, n) = R_Y(k)$ or the process is w.s.s. and its spectral density is as announced.

Let us consider the more general case. Let $\{h_n\}$ and $\{g_n\}$ be two stable signals. Let $\{X_n\}$ and $\{Y_n\}$ be two (zero mean) jointly w.s.s. processes with cross-covariance $R_{XY}(n) = \mathbf{E}[X_{k+n} Y_n^*]$ and corresponding spectral density $S_{XY}(\lambda)$. Then we have the following result which is called the the relation of isometry. We will use the notation $\sum_{k \in Z}$ to denote $\sum_{k=-\infty}^{\infty}$.

Proposition 3.3.3

$$\mathbf{E}[(\sum_{k \in Z} h_k X_k)(\sum_{k \in Z} g_k Y_k)^*] = \int_{-\frac{1}{2}}^{\frac{1}{2}} H(e^{i2\pi\lambda}) G(e^{-i2\pi\lambda}) S_{XY}(\lambda) d\lambda \tag{3.3. 9}$$

where $H(z)$ is the two-sided z–transform of $\{h_k\}$ and $G(z)$ is the two-sided z-transform of $\{g_k\}$ (which exist because of the stability hypotheses).

Proof:

The proof just follows by taking the expectations and then using the Herglotz theorem for the representation of the covariance. Indeed:

$$\mathbf{E}[(\sum_{k \in Z} h_k X_k)(\sum_{k \in Z} g_k Y_k)^*] = \sum_{k \in Z} \sum_{j \in Z} h_k g_j^* R_{XY}(k-j)$$

15

$$= \int_{-\frac{1}{2}}^{\frac{1}{2}} \sum_{k \in Z} \sum_{j \in Z} h_k g_j^* e^{i(k-j)2\pi\lambda} S_{XY}(\lambda) d\lambda$$

$$= \int_{-\frac{1}{2}}^{\frac{1}{2}} H(e^{i2\pi\lambda}) G(e^{-i2\pi\lambda}) S_{XY}(\lambda) d\lambda$$

where $H(e^{i2\pi\lambda}) = \sum_{k \in Z} h_k e^{i2\pi\lambda k}$ and similarly for $G(e^{-i2\pi\lambda})$. Note the interchange between the integral and the summations is justified by the use of Fubini's theorem.

Remark: We can derive the result concerning the spectral density of the output of a LTIC and stable system using the result above. For the sake of completeness let us do it.

First note that since the system is a time-invariant,causal and stable system $h_k = 0; k < 0$. Noting that :

$$Y_n = \sum_{-\infty}^{n} h_{n-k} X_k$$

we have $\sum_{-\infty}^{\infty} h_{n-k} e^{i2\pi\lambda k} = e^{i2\pi\lambda n} \sum_{k=0}^{\infty} h_k e^{-i2\pi\lambda k} = e^{i2\pi\lambda n} H(e^{-i2\pi\lambda})$. Similarly the corresponding expression for $G(e^{-i2\pi\lambda})$ with $g_k = h_{n-k}$ is $e^{-i2\pi\lambda n} H(e^{i2\pi\lambda})$. Hence

$$\mathbf{E}[Y_{n+k} Y_n^*] = \mathbf{E}[(\sum_{-\infty}^{n+k} h_{n+k-j} X_j)(\sum_{-\infty}^{n} h_{n-m} X_m)^*]$$

$$= \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{i(n+k)2\pi\lambda} H(e^{-i2\pi\lambda}) e^{-in2\pi\lambda} H(e^{i2\pi\lambda}) S_X(lambda) d\lambda$$

$$= \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{ik2\pi\lambda} |H(e^{i2\pi\lambda})|^2 S_X(\lambda) d\lambda$$

and the last line is just the Herglotz representation of the covariance of $\{Y_n\}$ with spectral density given by $|H(e^{i2\pi\lambda})|^2 S_X(\lambda)$.

This concludes our discussion of spectral theory for w.s.s. sequences. In the next chapter we will see some further consequences of the results we have developed.

3.3.2 Spectral theory for q.m. continuous w.s.s. processes

In this subsection we develop the analogous results for continuous-time w.s.s. stochastic processes. The analogous result to the Herglotz theorem is called the Bochner theorem. Bochner's theorem was actually developed to obtain the representation for characteristic functions (recall in Chapter 1 we showed that characteristic functions have the non-negative definiteness property) and we will comment on this later.

Let $\{X_t\}$ be a q.m. continuous w.s.s. process with mean 0 and covariance $R(\tau) = \mathbf{E}[X_{t+\tau} X_t^*]$. Recall, $R(t)$ is a non-negative definite function with $|R(t)| \leq R(0)$. Then we can state the following theorem.

Proposition 3.3.4 *(Bochner's Theorem)*
Let $R(t)$ be a continuous non-negative definite function on $(-\infty, \infty)$. Then there exists a unique, finite, non-decreasing function called the spectral measure such that:

$$R(t) = \int_{-\infty}^{\infty} e^{i2\pi\lambda t} dF(\lambda)) \tag{3.3. 10}$$

with $F(-\infty) = 0$ and $F(\infty) = R(0)$. *If in addition $\int_{-\infty}^{\infty} |R(t)| dt < \infty$ then $F(\lambda)$ is differentiable with*

$$S(\lambda) = \frac{dF(\lambda)}{d\lambda} \geq 0$$

and $S(\lambda)$ is called the spectral density of the process.

Proof: The proof closely resembles the proof of the Herglotz theorem. We give a proof when a spectral density exists i.e. under the condition that $\int_{-\infty}^{\infty} |R(t)| dt < \infty$.

Define the following integral:

$$S_T(\lambda) = \frac{1}{T} \int_0^T \int_0^T R(t-s) e^{-i2\pi\lambda(t-s)} dt ds$$

Then it is easy to see that $S_T(\lambda) = [Re^{i2\pi\lambda}, e^{i2\pi\lambda}]$ where $[.,.]$ denotes the inner-product in $L_2[0,T]$. By the non-negative definiteness of $R(t)$ it therefore follows that $S_T(\lambda) \geq 0$ for all λ.

Alternatively note that :

$$
\begin{aligned}
\frac{1}{T} \int_0^T \int_0^T R(t-s) e^{-i2\pi\lambda(t-s)} dt ds &= \frac{1}{T} \mathbf{E}[(\int_0^T X_t e^{-i2\pi\lambda t} dt)(\int_0^T X_s e^{-i\omega s} ds)^*] \\
&= \frac{1}{T} \mathbf{E}[|\int_0^T X_t e^{-i2\pi\lambda t} dt|^2] \geq 0
\end{aligned}
$$

Changing the double integral to a single integral we obtain:

$$S_T(\lambda) = \int_{-T}^{T} R(\tau)(1 - \frac{|\tau|}{T}) e^{-i2\pi\lambda\tau} d\tau \ for \ 0 < |\tau| \leq T$$

Now $|R_T(\tau) = R(\tau)(1 - \frac{|\tau|}{T})| \leq R(0)$ hence $R_T(\tau) e^{-i\omega\tau}$ converges to $R(\tau) e^{-i2\pi\lambda\tau}$ as $T \to \infty$ and since $\int_{-\infty}^{\infty} |R(t)| dt < \infty$ by dominated convergence

$$\int_{-T}^{T} R_T(\tau) e^{-i2\pi\lambda\tau} d\tau \to \int_{-\infty}^{\infty} R(\tau) e^{-i2\pi\lambda\tau} d\tau$$

Therefore $S_T(\lambda)$ converges to a finite limit for each $\lambda \in (-\infty, \infty)$ as $T \to \infty$ which we denote by $S(\lambda)$ and :

$$S(\lambda) = \int_{-\infty}^{\infty} R(t) e^{-i2\pi\lambda t} dt$$

or $S(\lambda)$ is the Fourier transform of $R(t)$. Furthermore since $S_T(\lambda) \geq 0$ for all T it implies that the limit $S(\lambda) \geq 0$. Also noting that $S_T(\lambda)$ is the Fourier transform of the truncated function $R_T(t)$ we have for every T

$$\int_{-\infty}^{\infty} S_T(\lambda) d\lambda = R(0)$$

and hence $\int_{-\infty}^{\infty} S(\lambda) d\lambda = R(0)$. Taking inverse Fourier transforms we obtain that:

$$R(t) = \int_{-\infty}^{\infty} e^{i2\pi\lambda t} S(\lambda) d\lambda$$

Uniqueness can be established as in the Herglotz theorem.

Remark: Let us see the connection between this result and that of characteristic functions. Define the normalized covariance $\frac{R(t)}{R(0)} = C(t)$. Now, $C(t)$ is a non-negative definite function and thus

by Bochner's theorem the spectral measure associated with $C(t)$ denoted by $F(x)$ will have the property that it is non-decreasing with $F(-\infty) = 0$ and $F(\infty) = C(0) = 1$ or $F(x)$ is a probability distribution function on $(-\infty, \infty)$ and thus $C(t)$ is the characteristic function of the r.v. X with $F(x)$ as its distribution function. By the argument of uniqueness as in the Herglotz theorem we obtain that the characteristic function completely specifies the probability distribution since if two characteristic functions are equal then they must correspond to the same distribution.

Similar to the discrete-time case we can obtain the isometry relationship in the continuous-time case. As before we define stable signals and finite energy signals.

Definition 3.3.2 *A continuous-time signal $f(t)$ $t \in (-\infty, \infty)$ is said to be stable if*

$$\int_{-\infty}^{\infty} |f(t)|dt < \infty$$

and of finite energy if

$$\int_{-\infty}^{\infty} |f(t)|^2 dt < \infty$$

Remark: In the parlance of analysis stable signals belong to $L_1(-\infty, \infty)$ while finite energy signals belong to $L_2(-\infty, \infty)$. Note if a signal is stable then it is of finite energy but not necessarily vice versa. A very simple consequence of the stability property is that the Fourier transform of stable signals is well defined and if it is a rational polynomial then it will have all its poles in the left half plane.

We now state the following isometry result. When $f = g$ and $X = Y$ then this is just the analog of the Parseval theorem.

Proposition 3.3.5 *Let $\{X_t\}$ and $\{Y_t\}$ be two zero mean jointly w.s.s q.m continuous stochastic processes with cross-spectral density $S_{XY}(\lambda)$. Let $\{f_t\}$ and $\{g_t\}$ be two real, stable signals. Then:*

$$\mathbf{E}[(\int_{\Re} f_t X_t dt)(\int_{\Re} g_t Y_t dt)^*] = \int_{\Re} F(-\lambda)G(\lambda)S_{XY}(\lambda)d\lambda \tag{3.3. 11}$$

where $F(\lambda)$ and $G(\lambda)$ correspond to the Fourier transforms of $\{f_t\}$ and $\{g_t\}$ respectively.

Proof: For the sake of completeness we include the proof. Note by the w.s.s. stationarity and the use of Bochner's theorem for representing the covariance:

$$
\begin{aligned}
\mathbf{E}[(\int_{\Re} f_t X_t dt)(\int_{\Re} g_t Y_t dt)^*] &= \int_{\Re} \int_{\Re} f_t g_s R_{XY}(t-s)dtds \\
&= \int_{\Re} \int_{\Re} \int_{\Re} f_t g_s e^{i2\pi\lambda(t-s)} S_{XY}(\lambda)d\lambda \\
&= \int_{\Re} F(-\lambda)G(\lambda)S_{XY}(\lambda)d\lambda
\end{aligned}
$$

where $F(\lambda) = \int_{\Re} f_t e^{-i2\pi\lambda t}dt$ and similarly for $G(\lambda)$.

Using this result we state the following result conncerning the output process of a stable LTIC system whose input is a w.s.s. process $\{X_t\}$ (zero mean) with spectral density $S_X(\lambda)$. We omit the proof since it follows by a direct application of the isometry theorem as in the discrete-time case.

18

Proposition 3.3.6 *Let $\{X_t\}$ be a zero mean w.s.s. process with spectral density $S_X(\lambda)$ which is the input to a linear,time-invariant, causal and stable system. Let $\{Y_t\}$ denote the output process. Then $\{Y_t\}$ is a zero mean w.s.s. process with spectral density, denoted $S_Y(\lambda)$:*

$$S_Y(\lambda) = |H(\lambda)|^2 S_X(\lambda) \tag{3.3. 12}$$

where $H(\lambda)$ is the Fourier transform of the impulse response function of the system.

We end this section by showing how Bochner's theorem can be used to study the existence of higher-order derivatives (in the mean squared sense) of second order processes.

Let us first recall proposition 3.2.1. If $\{X_t; -\infty < t < \infty\}$ is w.s.s. then the existence of the derviative in the mean squared sense requires that the second partial derivative of the covariance function $R(t)$ exists and the second-partial derivative then corresponds to the covariance of the derivative of the process. A natural question we posed is that can we say directly by measuring come characteristics of the process as to how many derivatives of the process exist? The answer to this is given by Bochner's theorem by noting that the spectral density is nothing but the Fourier transform of the covariance.

Let $\{X_t; -\infty < t < \infty\}$ be a w.s.s. process with spectral density $S(\lambda)$. We can now state the sufficient conditions for the existence of derivatives of order n.

Proposition 3.3.7 *Let $\{X_t; -\infty < t < \infty\}$ be a w.s.s. process with spectral density $S(\lambda); -\infty < \lambda < \lambda$.*

If

$$\int_{-\infty}^{\infty} |\lambda|^{2n} S(\lambda) d\lambda < \infty \tag{3.3. 13}$$

then the $n-th$ order derivative of $\{X_t\}$ exists in a mean squared sense and is a w.s.s. process with covariance given by

$$R^{(n)}(t) = \int_{-\infty}^{\infty} (4\pi^2 \lambda^2)^n S(\lambda) e^{i2\pi\lambda t} d\lambda \tag{3.3. 14}$$

Proof: First note that by Bochner's theorem:

$$R(t) = \int_{-\infty}^{\infty} e^{i2\pi\lambda t} S(\lambda) d\lambda$$

Hence:

$$-\frac{d^2}{dt^2} R(t) = \int_{-\infty}^{\infty} (2\pi\lambda)^2 e^{i2\pi\lambda t} S(\lambda) d\lambda$$

Since $S(\lambda) \geq 0$ for all λ it implies that $\lambda^2 S(\lambda) \geq 0$ or $4\pi^2\lambda^2 S(\lambda)$ is a spectral density and thus $-\frac{d^2}{dt^2} R(t)$ is a covariance which is the covariance function of $\dot{X}_t$ from Proposition 3.2.1. Let us denote it by $R^{(1)}(t)$. Hence $\int_{-\infty}^{\infty} \lambda^2 S(\lambda) d\lambda < \infty$ implies that $R^{(1)}(0) < \infty$. Noting that $|R(t)| \leq R(0)$ this means that the process is well defined as a second order process.

Similarly the negative of the second derivative of $R^{(1)}(t)$ will also define a covariance which by the argument in Proposition 3.2.1 can be identified as the covariance of the second derivative of the process which is well defined if $\int_{-\infty}^{\infty} (2\pi\lambda)^4 S(\lambda) d\lambda < \infty$ and so on. Repeating the arguments we see that the condition given assures that the $2n-th$ derivative of $R(t)$ is defined which can then be associated with the $n-th$ derivative of X_t.

19

Examples:

Suppose that the spectral density of a w.s.s. process is a rational function and the difference between the order of the denominator polynomial and the numerator polynomial (in powers of λ) is 2m. Then the process is $(m-1)$ times differentiable since multiplying the spectral density by $\lambda^{2(m-1)}$ will leave the order of the denominator polynomial larger by λ^2 which will make the integral finite.

Here is an example of a process for which a derivative cannot be defined.

Let $\{X_t\}$ be a w.s.s. Gaussian stochastic process with spectral density:

$$S(\lambda) = \frac{2k}{k^2 + 4\pi^2\lambda^2}; \quad -\infty < \lambda < \infty$$

Then clearly:

$$\int_{-\infty}^{\infty} \lambda^2 S(\lambda) d\lambda = \infty$$

or the derivative cannot be defined.

This is a well known process which we have seen earlier. Taking inverse Fourier transforms gives

$$R(t) = e^{-k|t|}$$

which implies that it is a Gauss Markov process.

Finally an example of a process which is infinitely differentiable is a Gaussian process with spectral density

$$S(\lambda) = e^{-|\lambda|}$$

In fact if the spectral density is a rational function then only a finite number of derivatives of the process can be defined as seen above.

Suppose X_t is a w.s.s. process whose spectral density $S(\lambda)$ is band-limited to $[-B, B]$, i.e. its spectrum has compact (finite) support. In that case it is easy to see that such a process possesses derivatives of all orders in an L_2 sense. Such a process is said to be analytic in L_2.

This concludes our introduction to spectral theory of processes. In the sequel we will apply these results in the context of estimation as well as signal representation.

3.3.3 Ergodic theorems revisited

In light of the results above we can re-state the conditions for ergodic theorems holding in the mean squared sense.

First note that the sufficient condition for the existence of spectral densities in both the Herglotz theorem and the Bochner theorem imply the necessary and sufficient condition for convergence of sample means to the true means (Proposition 2.6.3) in the mean square. Hence, the existence of a spectral density implies that the process satisfies the law of large numbers in the mean squared sense.

The condition that $R(k) \in \ell_1$ (or $R(t) \in L_1$) is actually stronger than the conditions as stated in Proposition 2.6.3. It turns out that the condition in (2.6.3) corresponds to continuity of the spectral distribution at the origin i.e. at 0. We state the result below.

Proposition 3.3.8 *A necessary and sufficient condition for the sample-mean of a w.s.s. process (discrete or continuous-time) to converge to the mean in the mean square is that the spectral distribution be continuous at $\{0\}$ i.e. $F(\{0\}) = 0$.*

Proof: We will only give the proof in the discrete case. The continuous-time analog is similar. By proposition 2.6.3. the necessary and sufficient condition is

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} R(k) = 0$$

Let us use the Herglotz theorem to rewrite this condition.

$$
\begin{aligned}
\frac{1}{n} \sum_{k=0}^{n-1} R(k) &= \frac{1}{n} \sum_{k=0}^{n-1} \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{ik2\pi\lambda} dF(\lambda) \\
&= \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{1}{n} \left(\sum_{k=0}^{n-1} e^{i2\pi\lambda k} \right) dF(\lambda)
\end{aligned}
$$

Define $\phi_n(\lambda) = \frac{1}{n} \sum_{k=0}^{n-1} e^{i2\pi\lambda k}$.

Then

$$
\begin{aligned}
\phi_n(\lambda) &= 1 \quad \lambda = 0 \\
&= \frac{1}{n} \frac{e^{in2\pi\lambda} - 1}{e^{i2\pi\lambda} - 1} \quad \lambda \neq 0
\end{aligned}
$$

But

$$
\begin{aligned}
|\phi_n(\lambda)| &\leq \left| \frac{\sin n\pi\lambda}{n \sin \pi\lambda} \right| \\
&\leq \frac{\pi}{2} \left| \frac{\sin n\pi\lambda}{n\pi\lambda} \right| \\
&\leq \frac{\pi}{2}
\end{aligned}
$$

where we used the following fact:

$$|\sin \lambda| \geq \frac{2}{\pi}|\lambda| \quad for \; |\lambda| \leq \frac{\pi}{2}$$

.

Therefore $\lim_{n\to\infty} \phi_n(\lambda) \to \mathbf{1}_{\{0\}}(\lambda)$. Therefore the condition that $\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} R(k) = 0$ is equivalent to

$$\lim_{n\to\infty} \int_{-\frac{1}{2}}^{\frac{1}{2}} \phi_n(\lambda) dF(\lambda) = \int_{-\frac{1}{2}}^{\frac{1}{2}} \mathbf{1}_{\{0\}}(\lambda) dF(\lambda) = 0$$

or the spectral measure is continuous at $\{0\}$. The interchange of the limit in the integral follows by dominated convergence since the spectral measure is finite.

Finally we conclude with an important result due to Maruyama which establishes strict ergodicity in the case when the process $\{X_t\}$ is Gaussian. Note since it is Gaussian, w.s.s implies strict sense stationarity. We state the result without proof.

Proposition 3.3.9 *Let $\{X_t\}$ be a stationary Gaussian process. If the spectral measure or distribution is continuous everywhere (in $(-\frac{1}{2}, \frac{1}{2}]$ in the discrete-case and in $(-\infty, \infty)$ in the continuous time case) then the process is ergodic and hence Proposition 2.6.5(b) holds.*

Remark: As a consequence of the above result, if a Gaussian process possesses a spectral density it is ergodic.

3.4 Nyquist sampling principle for w.s.s. processes

It is convenient to discuss this first in the context of deterministic signals. Suppose $\{X_t\}$ is a deterministic process whose Fourier transform has compact support in $[-B, B]$. i.e. the Fourier transform vanishes (equals 0) outside the given interval. Then $2B$ is referred to as the band-width of the signal. Now the well known theorem due to Nyquist states that sampling the process at instants faster than or equal to $\frac{1}{2B}$ (called the Nyquist rate) allows us to completely reconstruct the signal in an L_2 sense i.e. there is no loss of energy (by Parseval's theorem). This result has a stochastic counterpart for w.s.s. signals called the Shannon-Nyquist Sampling Theorem.

Proposition 3.4.1 *Let $\{X_t; -\infty < t < \infty\}$ be a zero mean w.s.s. process whose spectral density $S(\omega)$ vanishes outside $[-B, B]$. Then:*

$$X_t = l.i.m_{N\to\infty} \sum_{-N}^{N} X(\frac{n}{2B}) \frac{\sin \pi(2Bt - n)}{\pi(2Bt - n)} \qquad (3.4.\ 15)$$

where $X(\frac{n}{2B})$ denotes the samples of $\{X_t\}$ taken every $\frac{1}{2B}$ units of time and l.i.m denotes the limit in q.m.

Proof: The proof esentially uses the sampling theorem in the deterministic context and Bochner's theorem.

Let $P(\lambda)$ denote the spectral density. Then if $R(t)$ denotes the covariance of $\{X_t\}$ by Bochner's theorem:

$$R(t) = \int_{-B}^{B} e^{i2\pi\lambda t} P(\lambda)d\lambda$$

due to the compact support of $P(\lambda)$ in $[-B, B]$.

Now for $\lambda \in [-B, B]$ we expand $e^{i2\pi\lambda t}$ in a Fourier series in λ i.e.

$$e^{i2\pi\lambda t} = \sum_{n\in Z} a_n(t) e^{i\frac{2\pi n\lambda}{2B}}$$

where $\{a_n(t)\}$ denote the Fourier coefficients and are given by:

$$
\begin{aligned}
a_n(t) &= \frac{1}{2B} \int_{-B}^{B} e^{i2\pi\lambda(t - \frac{n}{2B})}d\lambda \\
&= \frac{\sin \pi(2Bt - n)}{\pi(2Bt - n)}
\end{aligned}
$$

Now substituting for $e^{i2\pi\lambda t}$ in the resperentation for $R(t)$ we obtain:

$$
\begin{aligned}
R(t) &= \int_{-B}^{B} \sum_{n\in Z} a_n(t) e^{i\frac{2\pi\lambda n}{2B}} P(\lambda)d\lambda \\
&= \sum_{n\in Z} a_n(t) \int_{-B}^{B} e^{i\frac{2\pi\lambda n}{2B}} P(\lambda)d\lambda \\
&= \sum_{n\in Z} a_n(t) R(\frac{n}{2B})
\end{aligned}
$$

The above is just a statement of the deterministic or Nyquist sampling theorem for the deterministic function $R(t)$ whose Fourier transform (the spectral density) has support in $[-B, B]$.

First let us note some properties of the functions $a_n(t)$. From the definition of $a_n(t)$ it follows that $a_n(t - \frac{m}{2B}) = a_{n+m}(t)$ and furthermore:

$$R(t - \frac{m}{2B}) = \sum_{n \in Z} a_n(t - \frac{m}{2B}) R(\frac{n}{2B}) = \sum_{n \in Z} a_n(t) R(\frac{n-m}{2B})$$

. Using the above properties it readily follows that

$$\sum_n a_n(t) a_{n+m}(t+s) = a_m(s)$$

and

$$\sum_n a_n(t) R(t - \frac{n}{2B}) = R(0)$$

noting that the implicit assumption is the Fourier series representation for $e^{i2\pi\lambda t}$ for $\lambda \in [-B, B]$.

Hence let us show that $E|X_t - \sum_{n=-N}^{N} a_n(t) X(\frac{n}{2B})|^2 \to 0$ as $N \to \infty$. Let $Z_N(t) = X_t - \sum_{n=-N}^{N} X(\frac{n}{2B}) a_n(t)$ then

$$
\begin{aligned}
E[Z_N^2(t)] &= E|X_t|^2 - 2 \sum_{n=-N}^{N} a_n(t) E[X_t X(\frac{n}{2B})] + \sum_{n=-N}^{N} \sum_{m=-N}^{N} a_n(t) a_m(t) E[X(\frac{n}{2B}) X(\frac{m}{2B})] \\
&= R(0) - 2 \sum_{n=-N}^{N} a_n(t) R(t - \frac{n}{2B}) + \sum_{n=-N}^{N} \sum_{m=-N}^{N} a_n(t) a_m(t) R(\frac{n-m}{2B})
\end{aligned}
$$

Noting that $\sum_{n \in Z} a_n(t) R(t - \frac{n}{2B}) = R(0)$ and

$$
\begin{aligned}
\sum_{n \in Z} \sum_{m \in Z} a_n(t) a_m(t) R(\frac{n-m}{2B}) &= \sum_{m \in Z} a_m(t) \left(\sum_{n \in Z} a_n(t) R(\frac{n-m}{2B}) \right) \\
&= \sum_{m \in Z} a_m(t) R(t - \frac{m}{2B}) \\
&= R(0)
\end{aligned}
$$

Note in the above the order of taking the summations does not matter.

Taking limits as $N \to \infty$ above we see that

$$\lim_{N \to \infty} E|X_t - \sum_{n=-N}^{N} a_n(t) X(\frac{n}{2B})|^2 = 0$$

and the proof is complete.

Remark 3.4.1 *From the definition of $a_n(t)$ and Plancherel-Parseval formula it is easy to see that:*

$$
\begin{aligned}
\int_{-\infty}^{\infty} a_n(t) a_m(t) dt &= 0 \quad n \neq m \\
&= Const(n) \quad n = m
\end{aligned}
$$

Which means that $\{a_n(t)\}$ form orthogonal functions on $L_2(-\infty, \infty)$, and indeed can be shown to be a basis for all $L_2(-\infty, \infty)$ functions whose Fourier transform vanishes outside $[-B, B]$.

Hence we have:

$$\int_{-\infty}^{\infty} X_t a_m(t) dt = X(\frac{m}{2B})$$

or the sampling theorem is a Fourier expansion of X_t in terms of the basis $\{a_n(t)\}$ with Fourier coefficients $X(\frac{n}{2B})$.

Let us now study the issue of sampling a w.s.s. process at intervals of $\frac{1}{2B}$ units of time when in fact the spectral density has a support of greater than $2B$. In fact we consider a process whose spectral density is not necessarily band limited. The interesting questions that arise are what is the resulting spectral density of the sampled process and the mean squared error?

Let $\hat{X}_t$ denote the approximation obtained based on sampling a w.s.s. process X_t every $\frac{1}{2B}$ units of time using the approximating formula:

$$\hat{X}_t = \sum_{n \in Z} X(\frac{n}{2B}) a_n(t)$$

where $a_n(t)$ are the functions defined above.

Let us first show that $\hat{X}_t$ is indeed a w.s.s. process.

Now,

$$
\begin{aligned}
\mathbf{E}[\hat{X}_t \hat{X}_{t+s}] &= \mathbf{E}[\sum_{n \in Z} \sum_{m \in Z} a_n(t) a_m(t+s) X(\frac{n}{2B}) X(\frac{m}{2B})] \\
&= \sum_{n \in Z} \sum_{m \in Z} a_n(t) a_m(t+s) R(\frac{n-m}{2B})
\end{aligned}
$$

We now use Bochner's theorem to represent $R(\frac{n-m}{2B}$ and then

$$\hat{R}(t, t+s) = \int_{-\infty}^{\infty} \sum_{n \in Z} \sum_{m \in Z} \left(a_n(t) a_m(t+s) e^{i\frac{2\pi\lambda}{2B}n} e^{-i\frac{2\pi\lambda}{2B}m} \right) P(\lambda) d\lambda$$

Now the term in the integrand in brackets is periodic in $2B$ since $e^{i\frac{2\pi\lambda}{2B}} = e^{i\frac{2\pi(\lambda+2kB)}{2B}}$ and therefore the above integral can be written is:

$$\hat{R}(t, t+s) = \int_{-B}^{B} \sum_{n \in Z} \sum_{m \in Z} e^{i\frac{2\pi\lambda}{2B}n} e^{-i\frac{2\pi\lambda}{2B}m} \sum_{k \in Z} P(\lambda + 2kB) d\lambda$$

Now using the fact that for $\lambda \in [-B, B]$ $\sum_{n \in Z} a_n(t) e^{i\frac{2\pi\lambda}{2B}n} = e^{i2\pi\lambda t}$ we have

$$\hat{R}(t, t+s) = \int_{-B}^{B} e^{12\pi\lambda s} \sum_{k \in Z} P(\lambda + 2kB) d\lambda$$

Noting that the r.h.s only depends on s and not t via Bochner's theorem we conclude that $\hat{R}(t, t+s) = \hat{R}(s)$ or the process is w.s.s. with spectral density $\sum_{k \in Z} P(\lambda + 2kB)$.

From the above form of the spectral density for the approximation it can readily be seen that if the support of the spectral density of X_t is contained in $[-B, B]$ then the process $\hat{X}_t$ has the same spectral density or is completely recovered in an L_2 sense while if the support is larger then there

is contribution of the spectral density outside $[-B, B]$ shifted by $2kB$. This phenomenon is termed *aliasing* and hence we do not recover the original process.

Let us now compute the mean squared error corresponding to the variance of the difference process $z_t = X_t - \hat{X}_t$.

Now $\mathbf{E}[|X_t - \hat{X}_t|^2 = \mathbf{E}[X_t^2] + \mathbf{E}[\hat{X}_t^2] - 2\mathbf{E}[X_t\hat{X}_t]$. Noting that

$$\mathbf{E}[X_t\hat{X}_t] = \sum_{n \in Z} a_n(t)R(t - \frac{n}{2B})$$

and

$$\mathbf{E}[\hat{X}_t^2] = \sum_{n \in Z}\sum_{m \in Z} a_n(t)a_m(t)R(\frac{n-m}{2B}) = R(0)$$

we obtain:

$$
\begin{aligned}
\mathbf{E}[z_t^2] &= 2R(0) - 2\sum_{n \in Z} a_n(t)R(t - \frac{n}{2B}) \\
&= \int_{-\infty}^{\infty}(2 - 2\sum_{n \in Z} a_n(t)e^{i2\pi\lambda(t - \frac{n}{2B})})P(\lambda)d\lambda \\
&= \int_{-\infty}^{\infty}(1 - \sum_{n} a_n(t)e^{i2\pi\lambda(t - \frac{n}{2B})})^2 P(\lambda)d\lambda \\
&= \int_{-B}^{B}\sum_{k \in Z}(1 - \sum_{n \in Z} a_n(t)e^{i2\pi(\lambda+2kB)(t - \frac{n}{2B})})^2 P(\lambda + 2kB)d\lambda \\
&= \int_{-B}^{B}\sum_{k \in Z}(1 - e^{i4\pi kBt})^2 P(\lambda + 2kB)d\lambda \\
&= 4\int_{-B}^{B}\sum_{k \in Z}\sin^2(2\pi kBt)P(\lambda + 2kB)d\lambda
\end{aligned}
$$

where we have used Bochner's theorem and the fact that for $\lambda \in [-B, B]$ the term $\sum_{n \in Z} a_n(t)e^{-i\frac{2\pi\lambda}{2B}n} = e^{-2\pi\lambda t}$ and $e^{i4\pi kn} = 1$.

It is interesting to note from above that although $\{X_t\}$ and $\{\hat{X}_t\}$ are individually w.s.s. they are not jointly so.

Let us now consider the problem of interpolation i.e. starting out with a discrete-time process we use the approximating formula to obtain a continuous time process. We state this result as a proposition.

Proposition 3.4.2 *Let $\{X_n\}_{n=-\infty}^{\infty}$ be a w.s.s. sequence with spectral density $P(\lambda), \lambda \in [-\frac{1}{2}, \frac{1}{2}]$. Define the continuous time process $\{X_t; -\infty < t < \infty\}$ by*

$$X_t = l.i.m \sum_{n=-N}^{N} X_n \frac{\sin \pi(2Bt - n)}{\pi(2Bt - n)} \tag{3.4. 16}$$

Then $\{X_t\}$ is w.s.s. process with spectral density:

$$
\begin{aligned}
S(\lambda) &= \frac{1}{2B}P(\frac{\lambda}{2B}) \quad |\lambda| \leq B \\
&= 0 \quad \text{otherwise}
\end{aligned}
$$

25

Proof: By definition of $\{X_t\}$ we have:

$$\mathbf{E}[X_t X_{t+s}] = \sum_{n=-N}^{N} \sum_{m=-N}^{N} a_n(t) a_m(t+s) R(n-m)$$

Using the Herglotz theorem to represent $R(n-m)$ we obtain:

$$
\begin{aligned}
\mathbf{E}[X_t X_s] &= \lim_{N\to\infty} \int_{-\frac{1}{2}}^{\frac{1}{2}} \sum_{n=-N}^{N} \sum_{m=-N}^{N} a_n(t) a_m(t+s) e^{i2\pi\lambda(n-m)} P(\lambda) d\lambda \\
&= \lim_{N\to\infty} \frac{1}{2B} \int_{-B}^{B} \sum_{n=-N}^{N} a_n(t) e^{i2\pi \frac{f}{2B} n} \sum_{m=-N}^{N} a_m(t+s) a^{-i2\pi \frac{f}{2B} m} P(\frac{f}{2B}) df \\
&= \frac{1}{2B} \int_{-B}^{B} \sum_{n\in Z} a_n(t) e^{i2\pi \frac{f}{2B} n} \sum_{m\in Z} a_m(t+s) e^{-i2\pi \frac{f}{2B} m} P(\frac{f}{2B}) df \\
&= \frac{1}{2B} \int_{-B}^{B} e^{i2\pi f s} P(\frac{f}{2B}) df
\end{aligned}
$$

and hence by Bochner's theorem X_t is w.s.s. with spectral density as announced.

Remark 3.4.2 *Note that by construction $\{X_t\}$ is a band-limited process with bandwidth $2B$.*

Shannon-Nyquist sampling provides the theoretical basis for most statistical signal processing algorithms because physical signals have finite energy and thus unless their spectral density vanishes at a sufficiently fast rate they will violate the finite energy constraint. Indeed most practical signals can be assumed to to have finite bandwidth.

Given the importance of Markov models in applications it will then come as a surprise to note that if X_t is a w.s.s. process and is Markov, then it cannot be reconstructed via Nyquist sampling in an L_2 sense or equivalently such processes cannot be bandlimited.

The simplest case to observe this is the Gauss-Markov model. We know from Proposition 2.4.2 that if X_t is a stationary Gauss-markov process then its covariance is given by $R(t) = Ce^{-\beta|t|}$ where $C = var(X_t)$ and $\beta = -\ln \frac{R(1)}{C}$. The spectral density is thus given by:

$$S(\lambda) = \frac{2C\beta}{\beta^2 + 4\pi^2\lambda^2}, \quad -\infty < \lambda < \infty$$

and thus the spectral density has infinite support and thus cannot be recovered without error using any finite sampling rate.

Let us show this result more generally for any w.s.s. Markov process.

Proposition 3.4.3 *Let $\{X_t, -\infty < t < \infty\}$ be a w.s.s. Markov process and $\{X_{nT}\}$ be the samples of $\{X_t\}$ sampled at instants $nT, n = 0, \pm 1, \pm 2\cdot$. The X_t cannot be recovered exactly in an L_2 sense by Shannon-Nyquist sampling.*

Proof:
The proof essentially follows from the Markov property and the minimality of the conditional expectation with respect to the mean squared error criterion.

Indeed let $f(X_{nT}, |n| \leq N)$ denote a functional consisting of $2N + 1$ samples that include t.

Then we know that $\mathbf{E}[(X_t - f(X_{nT}, |n| \leq N))^2]$ is minimized when :

$$f(X_{nT}, |n| \leq N) = \mathbf{E}[X_t | X_{nT}, |n| \leq N]$$

But from the Markov property the conditional distribution only depends on X_{mT} and $X_{(m+1)T}$ where $t \in [mT, (m+1)T]$:

$$\mathbf{E}[X_t | X_{nT}, |n| \leq N] = g(X_{mt}, X_{(m+1)T})$$

for some measurable function $g(.)$.

Therefore from the minimizing property of the conditional expectation we have:

$$0 < \mathbf{E}[(X_t - g(X_{mT}, X_{(m+1)T})^2] \leq \mathbf{E}[(X_t - \sum_{n=-N}^{N} a_n(t) X_{nT})^2]$$

since by definition the conditional distribution of X_t given $X_{mT}, X_{(m+1)T}$ is assumed to be strictly positive from the Markov property.

Now taking limits as $N \to \infty$ and noting that the l.h.s. does not depend on N we have that

$$\liminf_{N \to \infty} \mathbf{E}[(X_t - \sum_{n=-N}^{N} a_n(t) X_{nT})^2] > 0$$

which shows that the process cannot be recovered without error for any finite T in the $L_2(\mathbb{P})$ sense as we have shown for w.s.s. processes with spectral density of finite support.

We conclude this section by discussing the sampling principle for processes whose spectral density is defined only for certain pass-bands i.e. processes whose spectral density is non-zero only in the region $[-B - B_0, -B_0 + B]$ and $[B_0 - B, B_0 + B]$. Such processes arise in communications applications where the energy of a signal is concentrated around a given frequency called the base frequency. This is typically the case when a so-called base-band signal is modulated by a high frequency signal of frequency B_0. Of course we can use the above sampling theorem and conclude that we can reconstruct the process if we sample at rate $\frac{1}{2(B+B_0)}$. If B_0 is large then this will entail sampling at a very high rate and it does not take into account that the spectral density is 0 between $[-B_0 + B, B_0 - B]$. In fact it can be shown that the signal can be recovered by sampling at the rate $\frac{1}{2B}$ which can represent a considerable saving if B_0 is large in comparision to B. The basic idea is that by modulating the signal at a frequency of B_0 we can shift the spectrum to lie between $[-B, B]$ and then apply the above result.

3.5   Karhunen-Loeve expansions for 2nd. order processes

One can exploit the property that 2nd. order processes are defined on Hilbert spaces to obtain representations of 2nd order processes in terms of collections of uncorrelated random variables much akin to the Fourier series representations for deterministic signals defined on a finite interval.

Let us begin by studying a simple case. Let $\{X_t ; -\infty < t < \infty$ be a zero mean w.s.s. process whose covariance is periodic with period T i.e. $R(t + kT) = R(t)$. We can exploit this periodicity of the covariance to express the process $\{X_t\}$ in terms of a Fourier series whose coefficients are uncorrelated r.v's.

First note since $R(t)$ is periodic in T we can represent it by a Fourier series on $[0, T]$. Let $[f.g] = \frac{1}{T} \int_0^T f(t) \bar{g}(s) ds$ denote the inner-product on $L_2[0, T]$ where $\bar{g}(t)$ denotes the complex conjugate of

$g(t)$. Let $\{\phi_n(t)\}$ denote a C.O.N.S. (complete orthonormal system) on $L_2[0,T]$. In fact $\phi_n(t)$ can be taken as $e^{in\omega_0 t}; \omega_0 = \frac{2\pi}{T}$ which is complete in $L_2[0,T]$. Then

$$R(t) = \sum_{n \in Z} [R, \phi_n]\phi_n(t)$$

where $[R, \phi_n] = R_n$ denotes the nth. Fourier coefficient.

We can then show the following result:

Proposition 3.5.1 *Let $\{X_t ; -\infty < t < \infty\}$ be a zero mean w.s.s. process whose covariance $R(t)$ is periodic with period T.*

Then

$$X_t = l.i.m. \sum_{n=-N}^{N} X_n e^{in\omega_0 t} ; \quad \omega_0 = \frac{2\pi}{T} \tag{3.5. 17}$$

where $\{X_n\}$ is a collection of uncorrelated r.v's with $\mathbf{E}[X_n^2] = R_n$ where R_n is the Fourier coefficient corresponding to the Fourier series for $R(t)$.

Proof: Define $X_n = [X, \phi_n]$ where $\phi_n(t) = e^{in\omega_0 t}$. Then:

$$\begin{aligned}
\mathbf{E}[X_n X_m^*] &= \frac{1}{T^2} \int_0^T \int_0^T \mathbf{E}[X_t X_s] e^{i\omega_0(ms-nt)} dt ds \\
&= \frac{1}{T^2} \int_0^T \int_0^T R(t-s) e^{-in\omega_0 t} e^{im\omega_0 s} dt ds \\
&= \frac{1}{T} \int_0^T R(t) e^{-in\omega_0 t} dt [\phi_n, \phi_m] \\
&= R_n \delta_{n,m}
\end{aligned}$$

Thus showing that the $\{X_n\}'s$ form a collection of uncorrelated r.v's with $\mathbf{E}[X_n^2] = R_n$. Finally as in the previous section noting that $R(0) = \sum_{n \in Z} R_n$ we obtain that

$$\lim_{N \to \infty} \mathbf{E}[|X_t - \sum_{n=-N}^{N} X_n \phi_n(t)|^2] = 0$$

proving the result.

Actually such a representation holds for any second order process without the requirement that it be w.s.s. for every finite interval of time. This important result is known as the Karhunen-Loeve expansion and is the basis for decomposing signals in terms of uncorrelated r.v's which has important implications in vector quantization of signals to achieve minimum mean squared distortion.

Before we develop the Karhunen-Loeve expansion we will see some properties associated with covariance operators or kernels.

Let $\{X_t; t \in [0,T]\}$ be a zero mean second-order process with covariance $R(t,s) = \mathbf{E}[X_t X_s^*]$. Then $R(t,s)$ is defined for $(t,s) \in [0,T] \times [0,T]$. Then the fundamental property of a covariance operator is that it be a non-negative definite operator i.e. for every $f \in L_2[0,T]$ where $L_2[0,T] = \{f(.): \int_0^T |f(t)|^2 dt < \infty\}$ we have: $[Rf,f] = \int_0^T \int_0^T R(t,s)f(s)f^*(t)dsdt \geq 0$. Also the property that $\int_0^T \int_0^T R(t,s)dsdt < \infty$ implies that we can find a sequence of functions $\{\phi_k(t), 0 \leq t \leq T\}$ such that $||\phi_k||^2 = \int_0^T |\phi_k(t)|^2 dt = 1$ and $[\phi_k, \phi_m] = 0$ *for* $k \neq m$ and

$$R\phi_k = \int_0^T R(t,s)\phi_k(s)ds = \lambda_k \phi_k(t)$$

28

where $\lambda_k \geq 0$ i.e. λ_k is an eigenvalue corresponding to the eigenfunction $\phi_k(t)$ for the operator R. This is exactly analogous to the case when R is a non-negative definite matrix. Moreover it can be shown that the functions $\{\phi_k(t)\}$ are "complete" in $L_2[0,T]$ in the sense that for any function $f \in L_2[0,T]$ $[f,\phi_k] = 0$implies$||f||^2 = 0$. Another important property of such covariance kernels is that

$$Trace(R) = \int_0^T R(t,t)dt = \sum_{k=0}^{\infty} \lambda_k < \infty$$

which corresponds to the analog for matrices that the trace is the sum of the eigenvalues. This implies that $\lim_{k\to\infty} \lambda_k = 0$. Now let us examine what a zero eigenvalue for R means for the process.

$$\begin{aligned}
\mathbf{E}[[\phi_k, X]^2] &= \mathbf{E}[\int_0^T \int_0^T \phi_k(t)X_t^* \phi_k(s)^* X_s dt ds] \\
&= \int_0^T \int_0^T \phi_k(t)R(s,t)\phi_k^*(s)dsdt \\
&= \int_0^T \lambda_k \phi_k(s)\phi_k^*(s)ds \\
&= \lambda_k
\end{aligned}$$

and hence $\lambda_k = 0$ implies that the r.v. has variance 0 and hence is almost surely 0 and hence we can neglect the 0 eigenvalue terms.

Now let $\{X_t; t \in [0,T]\}$ be s second-order process with covariance $R(t,s)$. Suppose that we can find a representation of the form:

$$X_t = l.i.m. \sum_{n=0}^{N} Z_n \varphi_n(t)$$

where $\{\varphi_n(t)\}$ are orthonormal in $L_2[0,T]$ and $\{Z_n\}$ are uncorrelated r.v's.. Let us see what constraints it puts on the r.v's and the functions $\varphi_n(t)$. Let $\lambda_n = E[Z_n Z_n^*]$ the variance (assuming that X_t is zero mean) of Z_n. Then:

$$\mathbf{E}[X_t X_s^*] = \sum_{n=0}^{\infty} \lambda_n \varphi_n(t)\varphi_n^*(s)$$

and therefore

$$\begin{aligned}
\int_0^T R(t,s)\varphi_n(s)ds &= \sum_{m=0}^{\infty} \lambda_m \varphi_m(t)[\varphi_m, \varphi_n] \\
&= \lambda_n \varphi_n(t)
\end{aligned}$$

by the orthogonality of the functions $\varphi_n(t)$. Hence from above it implies that $\varphi_n(t) = \phi_n(t)$ where $\phi_n(t)$ are the eigenfunctions of $R(t,s)$ with λ_n the corresponding eigen values.

The above arguments motivate the Karhunen-Loeve expansion. Before we state the theorem we give another important result which can be seen as a result of the above. This result is called Mercer's theorem which we state separately since it is of independent interest.

Theorem 3.5.1 *Let $R(t,s); (t,s) \in [0,T] \times [0,T]$ be a continuous covariance kernel. Then*

$$R(t,s) = \sum_{n=0}^{\infty} \lambda_n \phi_n(t)\phi_n^*(s) \tag{3.5. 18}$$

uniformly in $0 \leq s, t \leq T$ where $\phi_n(t)$ are the eigenfunctions of $R(t,s)$ with corresponding eigenvalue λ_n.

The idea behind the proof is the following: by the property that $\int_0^T \int_0^T |R(t,s)| dt ds < \infty$ it can be shown that

$$\lambda_0 = \max_{||\phi||=1} [R\phi, \phi]$$

exists. Let ϕ_0 denote the function for which the max is achieved. Then $\int_0^T R(t,s)\phi_0(s)ds = \lambda_0\phi_0(t)$. Now define:

$$R_1(t,s) = R(t,s) - \lambda_0\phi_0(t)\phi_0^*(s)$$

Then $R_1(t,s)$ defines a covariance function by noting the following: define the process:

$$Y_t = X_t - \int_0^T X_s\phi_0^*(s)ds\phi_0(t)$$

Then the covariance of Y_t is just $R_1(t,s)$. Now since $R_1(t,s)$ is a covariance we can define the largest eigenvalue λ_1 as above and then:

$$R_2(t,s) = R_1(t,s) - \lambda_1\phi_1(t)\phi_1^*(s)$$

will define a covariance with the additional property that λ_1 will be an eigenvalue of $R(t,s)$ corresponding to the eigenfunction $\phi_1(t)$ while $\phi_0(t)$ will belong to the kernel (corresponding to the eigenvalue 0) of $R_1(t,s)$. Let us show this:

$$R_1\phi_0 = R\phi_0 - \lambda_0\phi_0[\phi_0, \phi_0] = \lambda_0\phi_0 - \lambda_0\phi_0 = 0$$

and hence $[R_1\phi_1, \phi_0] = \lambda_1[\phi_1, \phi_0] = [\phi_1, R_1\phi_0] = 0$ which implies that $[\phi_1, \phi_0] = 0$ from which it follows that $R\phi_1 = R_1\phi_1 = \lambda_1\phi_1$ or λ_1 is also an eigenvalue of R with corresponding eigenfunction ϕ_1. Hence repeating these arguments will show that :

$$R(t,s) = \sum_{n=0}^{N} \lambda_n\phi_n(t)\phi_n^*(s) + R_{N+1}(t,s)$$

and then by the nuclearity condition (sum of all the eigenvalues is finite) it follows that $\lim_{N\to\infty} R_{N+1}(t,s)$ goes to zero uniformly.

Using the above observations we now state the Karhunen-Loeve Theorem.

Proposition 3.5.2 *(Karhunen-Loeve Theorem)*
 Let $\{X_t; t \in [0,T]\}$ be a zero mean q.m.continuous process with covariance $R(t,s)$. Then

a) Let $\{\phi_n(t)\}$ be the eigenfunctions of $R(t,s)$ corresponding to eigenvalues $\{\lambda_n\}$ i.e.

$$R\phi_n = \int_0^T R(t,s)\phi_n(s)ds = \lambda_n\phi_n(t); \quad 0 \le t \le T$$

Define $Z_n(\omega) = \int_0^T X_s(\omega)\phi_n^*(s)ds$. Then:

$$X_t(\omega) = l.i.m \sum_{n=0}^{N} Z_n(\omega)\phi_n(t) \tag{3.5. 19}$$

where $\{Z_n(\omega)\}$ form a sequence of uncorrelated r.v's with $\mathbf{E}[Z_nZ_n^*] = \lambda_n$ and $\{\phi_n(t)\}$ are orthonormal in $L_2[0,T]$.

b) If $X_t(\omega)$ has the expansion above such that $\{\phi_n(t)\}$ are orthonormal and $\{Z_n\}$ are uncorrelated with variance λ_n then the pair $(\lambda_n, \phi_n(t))$ must be an eigenvalue, eigenfunction pair for the covariance $R(t,s)$ on $[0,T] \times [0,T]$

Proof: We will only prove part a) since the proof of part b) follows from the discussion before Mercer's theorem.

Let us first show that $\mathbf{E}[Z_n Z_m^*] = \lambda_n \delta_{n,m}$. Indeed,

$$
\begin{aligned}
\mathbf{E}[Z_n Z_m^*] &= \mathbf{E}\Big[\int_0^T \int_0^T X_t \phi_n^*(t) X_s^* \phi_m(s) dt ds\Big] \\
&= \int_0^T \int_0^T R(t,s)\phi_m(s)\phi_n^*(t) dt ds = \lambda_m[\phi_m, \phi_n] \\
&= \lambda_n \delta_{n,m}
\end{aligned}
$$

showing that the $Z_n's$ are uncorrelated with variance λ_n.

Now:

$$
\mathbf{E}|X_t - \sum_{n=0}^N Z_n \phi_n(t)|^2 = R(t,t) - \sum_{n=0}^N \lambda_n |\phi_n(t)|^2
$$

which goes to zero uniformly in t as $N \to \infty$ by Mercer's theorem.

Let us see some applications of the Karhunen-Loeve expansion.

Example 1: Let $\{X_t;\ t \in [0,T]\}$ be a zero mean q.m. continuous process with covariance $R(t,s)$. What is the best approximation of order 2 to the process $\{X_t\}$ in $L_2[0,T]$?

From the Karhunen-Loeve expansion the best mean-squared approximation is just

$$
\hat{X}_t = Z_0 \phi_0(t) + Z_1 \phi_1(t)
$$

where $E[Z_0 Z_0^*] = \lambda_0$ where $\lambda_0 = \max\{\lambda_k\}$ where λ_k's are the eigenvalues of $R(t,s)$ with corresponding eigenfunction $\phi_k(t)$. Similarly λ_1 should be taken as the next highest eigenvalue with $\phi_1(t)$ the corresponding eigenfunction. This follows from Mercer's theorem since the approximation error is just $Trace(R) - \sum_{n=0}^1 \lambda_n = \sum_{n>1} \lambda_n$.

Let us consider a more important example related to a series representation of the covariance of a Wiener process or Brownian motion.

Example 2: Let $\{X_t;\ t \in [0,T]\}$ be a standard Brownian motion. Recall that $\{X_t\}$ is a Gaussian process with zero mean and covariance $E[X_t X_s] = R(t,s) = min(s,t)$ denoted by s^t.

Let us begin by calculating the eigenvalues and eigenfunctions of the operator R given by:

$$
R\phi = \lambda\phi
$$

which is equivalent to:

$$
\int_0^T R(t,s)\phi(s)ds = \lambda\phi(t); \quad t \in [0,T]
$$

Substituting for $R(t,s)$ we obtain:

$$
\int_0^t s\phi(s)ds + \int_t^T t\phi(s)ds = \lambda\phi(t) \tag{3.5. 20}
$$

31

The above is a so-called Fredholm integral equation which we can readily solve by differentiating both sides and obtaining a differential equation as follows:

Differentiating the equation above once w.r.t t we obtain:

$$\int_t^T \phi(s)ds = \lambda\phi'(t)$$

which still leaves us with an integral and so we differentiate it once more to obtain:

$$-\phi(t) = \lambda\phi''(t)$$

which is a second-order differential equation. To solve this we need two initial conditions which can be obtained by inspection of the first two equations: i.e. $\phi(0) = 0$ and $\phi'(T) = 0$.

The solution to 3.5. 20 for initial condition 0 is given by:

$$\phi(t) = A\sin\frac{1}{\sqrt{\lambda}}t \tag{3.5. 21}$$

Applying the condition $\phi'(T) = 0$ gives

$$\cos\frac{1}{\sqrt{\lambda}}T = 0$$

from which we obtain that the eigenvalues are given by:

$$\lambda_n = \frac{T^2}{(n+\frac{1}{2})^2\pi^2} \quad ; \ n = 0, 1, 2, ... \tag{3.5. 22}$$

and the normalized eigenfunctions (norm 1) are given by:

$$\phi_n(t) = \sqrt{\frac{2}{T}}\sin\left[(n+\frac{1}{2})\pi\frac{t}{T}\right] \tag{3.5. 23}$$

Hence by the Karhunen-Loeve theorem we obtain:

$$X_t = \begin{array}{c} l.i.m \\ N \to \infty \end{array} \sum_{n=0}^{N} \sqrt{\lambda_n}\phi_n(t)Y_n$$

where Y_n are $i.i.d.$ N(0,1) gaussian random variables.

It is interesting to note that although the sample-paths of Brownian motion are not differentiable anywhere with respect to t from the Karhunen-Loeve theorem the r.h.s. is q.m. differentiable.

Finally we note by Mercer's theorem we obtain the following representation for the covariance:

$$\min(s,t) = \sum_{n=0}^{\infty} \frac{2T}{(n+\frac{1}{2})^2\pi^2}\sin\left[(n+\frac{1}{2})\pi\frac{t}{T}\right]\sin\left[(n+\frac{1}{2})\pi\frac{s}{T}\right] \tag{3.5. 24}$$

which is a purely analytical result which is hard to derive directly.

3.6 Spectral representation of w.s.s. processes

The Bochner theorem (for continuous-time processes) and the Herglotz theorem (for discrete-time processes) relate the covariances of w.s.s. processes to the spectral density in frequency domain i.e. the spectral density is the Fourier transform of the covariance. The question that is natural to ask is whether the Fourier transform of the original process is well defined when we view the process as a function of the time index? This section is devoted to the study of this issue.

Let us begin by considering the continuous-time case. First note that for any deterministic function of time such that $\int_{-\infty}^{\infty} |f_t|^2 dt < \infty$ the Fourier transform $\hat{F}(\lambda) = \int_{-\infty}^{\infty} f_t e^{-i2\pi\lambda t} dt$ is well defined. The original function can be recovered by taking the inverse Fourier transform given by $\int_{-\infty}^{\infty} e^{i2\pi\lambda t} \hat{F}(\lambda) d\lambda$.

In the stochastic case the basic problem is that $\int_{-\infty}^{\infty} e^{-i2\pi\lambda t} X_t dt$ need not exist (because of the infinite interval) and thus we cannot define the Fourier transform directly. To do so we need to define a so-called generalized Fourier transform which is well defined called the spectral process. Then the original process can be viewed as the inverse Fourier transform of the spectral process where the equality holds in a quadratic mean sense.

We will first see the result in the discrete-time case.

Let us first recall the notion of an orthogonal increment process. Let $Z(\lambda)$ be a zero-mean, complex-valued stochastic process. Then $Z(\lambda)$ is said to be a process with orthogonal increments if: given $(\lambda_1, \lambda_2] \bigcap (\lambda_3, \lambda_4] = \emptyset$ then: $E[(Z(\lambda_2) - Z(\lambda_1))(Z(\lambda_4) - Z(\lambda_3)^*] = 0$ i.e. the process on the non-overlapping intervals is uncorrelated. Let us furthermore impose the following condition: $E[(Z(\lambda_2) - Z(\lambda_1))(Z(\lambda_2) - Z(\lambda_1))^*] = E|Z(\lambda_2) - Z(\lambda_1)|^2 = |\lambda_2 - \lambda_1|$ i.e. the process has stationary, orthogonal increments.

Let $f(\lambda)$ be a continuous function on $-\frac{1}{2} \leq \lambda \leq \frac{1}{2}$. Define the stochastic integral:

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} f(\lambda) dZ(\lambda)$$

where $Z(\lambda)$ is the orthogonal, increment process defined above. Such an integral is well defined if $\int_{-\frac{1}{2}}^{\frac{1}{2}} |f(\lambda)|^2 d\lambda < \infty$ as we have seen.

Choose as the function $f(\lambda)$ the function $e^{2\pi n\lambda}$ then let us define the following integral:

$$X_n = \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{i2\pi n\lambda} dZ(\lambda) \tag{3.6. 25}$$

By definition $E[X_n] = 0$ and

$$
\begin{aligned}
E[X_n X_m^*] &= E[\int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{i2\pi(n\lambda - m\nu)} dZ(\lambda) dZ(\nu)] \\
&= \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{i2\pi(n-m)\lambda} d\lambda \\
&= 0 \; if \; n \neq m \\
&= 1 \; if \; n = m
\end{aligned}
$$

implying that $\{X_n\}$ is a sequence of uncorrelated, random variables indexed by n with variance 1 or a *white noise sequence*. This is the heart of the spectral representation theorem which we state below:

Theorem 3.6.1 *Let $\{X_n\}$ be a w.s.s., zero mean sequence of r.v's with spectral density $P(\lambda)$. Then we can represent X_n in the mean squared sense as:*

$$X_n = \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{i2\pi n\lambda} \psi(\lambda) dZ(\lambda) \qquad (3.6.\ 26)$$

where $Z(\lambda)$ is a stationary, orthogonal increment process and

$$\psi(\lambda)\psi^*(\lambda) = P(\lambda) \qquad (3.6.\ 27)$$

It is enough to see that the l.h.s. and r.h.s are the same in quadratic mean via the definition of $Z(\lambda)$ and the Herglotz theorem.

The function $\psi(\lambda)$ is called the spectral factor of $P(\lambda)$ and the procedure is known as spectral factorization. In particular $\psi(\lambda)$ can be chosen to be stable when $P(\lambda)$ is rational.

To see this we note that $P(\lambda)$ being a spectral density must be non-negative for all $\lambda \in [-\frac{1}{2}, \frac{1}{2}]$ and hence both the denominator and numerator polynomials (in λ) must be even. Furthermore if $e^{2\pi\lambda_0}$ is a root of the denominator polynomial then $e^{-2\pi\lambda_0}$ must also be a root which implies that if there exists a root within the unit circle there exists a corresponding root outside the unit circle. Hence we can choose as $\psi(\lambda)$ the rational function with all its roots in the denominator polynomial and numerator polynomial within the unit circle implying that it is a stable, minimum phase (all the roots of the numerator polynomial are within the unit circle).

Let us now see an example on how we can identify the "transfer function" $\psi(\lambda)$".

Example: Let $\{X_n\}$ be a stationary process defined by:

$$\sum_{k=0}^{m} a_k X_{n-k} = W_n \qquad (3.6.\ 28)$$

where $\{W_n\}$ is a white noise sequence. Clearly since the process $\{X_n\}$ is shationary and $\{W_n\}$ is stationary it is necessary that the "system" defined by the sequence $\{a_k\}$ must be stable i.e. the transfer function $A(z) = \sum_{k=0}^{m} a_k z^{-k}$ must have all its roots within the unit circle . (This is just the Wold decomposition for a purely non-deterministic process if we re write X_n in terms of the $\{W_n\}$). Such a process is called a stationary AR (Auto Regressive) process.

Let us use the spectral representation theorem substituting for X_n to obtain:

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} \sum_{k=0}^{m} a_k e^{i2\pi(n-k)\lambda} \psi(\lambda) dz(\lambda = \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{i2\pi n\lambda} dZ(\lambda)$$

The equality of the lhs. and the r.h.s implies that $\psi(\lambda) = \left[\sum_{k=0}^{m} a_k e^{-i2\pi k\lambda}\right]^{-1}$. Defining $z = e^{i2\pi\lambda}$ we see that all the roots of the denominator polynomial corresponding to $\psi(z) = \frac{1}{A(z)}$ are within the unit circle implying that $\psi(z)$ is stable system. From the isometry theorem we have:

$$A(e^{-i2\pi\lambda})A(e^{i2\pi\lambda})P(\lambda) = 1$$

In an analogous way we can define the spectral representation of a w.s.s. continuous-time process. We state the result below.

Proposition 3.6.1 *Let $\{X_t, -\infty < t < \infty\}$ be a w.s.s. continuous-time stochastic process whose spectral density $S(\lambda)$ satisies $\int_{-\infty}^{\infty} S(\lambda)d\lambda < \infty$. Then $\{X_t\}$ can be written as:*

$$X_t = \int_{-\infty}^{\infty} e^{i2\pi\lambda t}\psi(\lambda)d\hat{X}(\lambda) \tag{3.6. 29}$$

where:

$$\psi(\lambda)\psi(-\lambda) = S(\lambda) \tag{3.6. 30}$$

and $\{\hat{X}(\lambda); -\infty < \lambda < \infty\}$ is an orthogonal increment process satisfying:

$$E|\hat{X}(\lambda) - \hat{X}(\nu)|^2 = |\lambda - \nu| \tag{3.6. 31}$$

The use of the spectral representation is in the context of Wiener-Hopf filtering and other signal processing issues which we will discuss elsewhere.

3.7   Long-range dependence and self-similarity

Long-range dependence is a phenomenon by which correlations in a process persist over long durations of time. In mathematical terms it means that the correlations between X_n and X_{n+k} decay slowly in that even for k large the correlations are not negligible, i.e., when $\sum_{k=1}^{\infty} R(k) = \infty$. When correlations die away slowly they lead to special characteristics which can indeed have very important consequences in practice. For example we have seen in Chapter 2 that if $\sum_{k=0}^{\infty} R(k) < \infty$ where $R(k)$ is the covariance of a w.s.s. process then the SLLN holds for the empirical sum of the random sequence. We also saw that we needed a second-order condition for the CLT to hold. On the other hand when the correlations persist one would need inordinately large sample sizes to obtain good estimates of means and variances from observed samples. We will now study these characteristics in some detail and provide some precise definitions.

Definition 3.7.1 *Let $\{X_k\}$ denote a w.s.s. sequence. For each $m = 1, 2, \ldots,$, define the $m - level$ aggregated process $\{X_k^{(m)}\}_{k=1}$ formed by taking the average of m samples at a time given by:*

$$X_k^{(m)} = \frac{X_{m(k-1)+1} + \cdots + X_{km}}{m}, \quad k = 1, 2, \ldots \tag{3.7. 32}$$

The sequence $\{X_k^{(m)}\}$ is thus formed from the original sequence averaging over non-overlapping blocks of length m. Thus we can think of $\{X_k^{(m)}\}$ as viewed at a larger time scale of mk, i.e. each unit of time or sample of the sequence $X_k^{(m)}$ is actually composed of m samples of the original sequence. This process can be repeated by averaging over m samples f the sequence $X_k^{(m)}$ to obtain another sequence viewed at the time scale $m^2 k$ and so on. This process of averaging is important because it will allow us to study the phenomenon of long-range dependence.

First let us note that the sequence $X_k^{(m)}$ is also a second-order sequence with mean, $\mathbf{E}[X_k^{(m)}] = \mathbf{E}[X_k]$ and variance:

$$
\begin{aligned}
R^{(m)}(0) &= \frac{R(0)}{m} + \frac{2}{m^2} \sum_{k=1}^{m} R(k)(m-k) \\
&= \frac{R(0)}{m} + \frac{2}{m} \sum_{l=1}^{m-1} \sum_{k=1}^{l} R(l)
\end{aligned}
$$

35

Now let us define the difference operator $\delta^2 g_n = g_{n+1} - 2g_n + g_{n-1}$. Then we have:

$$\delta^2(m^2 R^m(0)) = (m+1)^2 R^{(m+1)}(0) - 2m^2 R^{(m)}(0) + (m-1)^2 R^{(m-1)}(0)$$
$$= 2R(m)$$

From this we see that specifying $R^{(m)}(0)$ is equivalent to specifying $R(m)$ or equivalently $S(\lambda)$ the spectral density.

Using the above we can also compute the covariance of the sequence $X_k^{(m)}$ as:

$$R^{(m)}(k) = \frac{1}{2}\delta^2\left(k^2 R^{(km)}(0)\right)$$

Let us first note some simple consequences of the fact $\sum_k R(k) < \infty$. From the Herglotz theorem we have $S(0) < \infty$. Furthermore $\lim_{m\to\infty} mR^{(m)}(0) \to R(0) + 2\sum_{k=1}^{\infty} R(k) < \infty$. One consequence of this is that if $\{X_k\}$ is a sequence of i.i.d. mean 0 variance 1 r.v's then as $m \to \infty$ the m-aggregated process goes to 0 in the mean square. We now define the notion of long-range dependence and study some consequences.

Definition 3.7.2 *A w.s.s. stochastic sequence or process $\{X_n\}$ is said to be long-range dependent if any of the following equivalent conditions hold:*

i. $\sum_{k=0}^{\infty} R(k) = \infty$

ii. $S(\lambda)$ *diverges at the origin*

iii. $mR^{(m)}(0) \to \infty$ *as* $m \to \infty$.

Remark 3.7.1 *A very important class of processes is one whose covariance decays as a power-law for large k i.e. $R(k) \sim ck^{-\beta}$ for $\beta \in (0,1)$. For such processes it can be seen that $R^{(m)}(0) \sim m^{-\beta}c$, where c is a constant, as $m \to \infty$. Such processes are usually termed asymptotically long-range dependent processes.*

Closely related to the notion of long-range dependence is the notion of self-similarity. In a crude sense what self-similarity implies is that the process *looks the same* over all time scales i.e. no matter whether the time-scale is long or short the process behaves statistically the same there is no *averaging out* effect. Once again we need a few definitions:

Let $H > 0$ and let us define the following m-level aggregated process as in equation 3.7. 32 except that we divide the m samples by m^H rather than m. i.e. let:

$$X_k^{(m^H)} = \frac{X_k^{(m)}}{m^{H-1}} \tag{3.7. 33}$$

Definition 3.7.3 *A discrete-time stochastic process $\{X_n\}$ is said to be (exactly) self-similar with self-similarity parameter $H > 0$, if for all $m = 1, 2, \cdots,$, the process $\{X_k^{(m^H)}\}$ has the same finite dimensional (joint) distributions as $\{X_k\}$.*

Requiring exact self-similarity is usually too much. In most applications we need a weaker form of 2nd. order self-similarity which we define below.

Definition 3.7.4 *A w.s.s. process $\{X_k\}$ is said to be 2nd. order self-similar with self-similarity parameter $H > 0$ if $\{X_k^{(m^H)}\}$ has the same covariance as $\{X_k\}$.*

36

Suppose that $\{X_k\}$ is a second-order self-similar process. Since $X_k^{(m)} = m^{H-1}X_k^{(m^H)}$ it follows that:

$$R^{(m)}(0) = m^{2H-2}R^{m^H}(0) = m^{2H-2}R(0)$$

Let $\beta \in (0,2)$ and let $H = 1 - \frac{\beta}{2}$. Then we see that $R^{(m)}(0) = m^{-\beta}R(0)$. From the remark 3.7.1 we see that if $\beta \in (0,1)$ the process is also long-range dependent.

An equivalent definition of (exact) second-order self-similarity is that $\{X_k\}$ is 2nd order self-similar with self-similarity parameter $H = 1 - \frac{\beta}{2}$ for $\beta \in (0,2)$ if $R^{(m)}(0) = m^{-\beta}R(0)$.

Using the fact that $R^{(m)}(k) = \frac{\delta^2}{2}(k^2 R^{km}(0))$ we can show that if a process is second-order self-similar with parameter $H = 1 - \frac{\beta}{2}$ we have: $R^{(m)}(k) = \frac{1}{2}\delta^2(k^{2H})$. Hence:

$$R^{(m)}(k) = \frac{1}{2}\left(|k+1|^{2H} - 2|k|^{2H} + |k-1|^{2H}\right), k = 0, \pm 1, \pm 2, \cdots,$$

Note when $H = 1$ the process is un-correlated.

The definition of long-range dependence and self-similarity for continuous-time processes is similar except that we now do not need m to be an integer. We define the notion of self-similarity below:

Definition 3.7.5 *A continuous time process $\{X_t\}_{t\geq 0}$ is self-similar with parameter $H > 0$, if for all $a > 0$ the finite dimensional distributions of $\{X_{at}\}$ and $\{a^H X_t\}$ are identical. If the sorresponding result holds only for moments of order up to 2 then the process is said to be second-order self-similar.*

A canonical example of a self-similar and long-range dependent process is a Gaussian process called *Fractional Brownian Motion*. It is self-similar with parameter H and long-range dependent if $H \in (\frac{1}{2}, 1)$.

Definition 3.7.6 *A stochastic process $\{W_t^H\}_{t\geq 0}$ is said to be a Fractional Brownian Motion (abbreviated as fbm) if:*

i. For each t, W_t^H is Gaussian with zero mean.

ii. $W_0^H = 0$ a.s..

$\{W_t^H\}$ has stationary increments and the increment $W_{t+s}^H - W_s^H$ is distributed $N(0, \sigma^2 t^{2H})$, $H > 0$

From the definition, it can be easily seen that if $H = \frac{1}{2}$ then W^H is a Brownian motion. The parameter H related to the covariance plays an important role in the theory of long-range dependent and self-similar processes. It is called the Hurst parameter, after a British scientist who studied the level of the Nile river in the Sudan in the late 1800's.

A discrete-time counter part of fbm constructed by defining $X_n = W_n^H - W_{n-1}^H, n = 1, 2, ...$ is called fractional Gaussian noise.

A natural question that one should ask is whether self-similarity and long-range dependent processes are common. It turns out, self-similar processes are natural limits of normalized partial sums of strictly stationary random variables. This important result is due to Lamperti which we state below.

Proposition 3.7.1 *Let $\{Y_t\}$ be a stochastic process such that $Y_1 \neq 0$ with positive probability. Assume that $\{T_t\}$ is the limit in distribution of the following normalized partial sums :*

$$Y_t = a_n^{-1}\sum_{k=1}^{\lfloor nt\rfloor} X_i, \quad n = 1, 2, \cdots, \tag{3.7. 34}$$

where $\lfloor nt \rfloor$ denotes the integer part of nt and $\{X_k\}$ is a stationary sequence of r.v's and $a_1, a_2, \cdots$, is a sequence of positive normalizing constants such that $\log a_n \to \infty$ as $n \to \infty$. Then there exists an $H > 0$ such that for any $u > 0$

$$\lim_{n \to \infty} \frac{a_{nu}}{a_n} = u^H$$

and $\{Y_t\}$ is a self-similar process with stationary increments and self-similarity parameter H. Additionally, all self-similar processes with stationary increments and $H > 0$ can be obtained as limits in distribution of such partial sums.

Thus long-range dependence can obtained as a limit of sums of random variables which have heavy-tails by which we mean $P(X > t) \sim ct^{-\beta}$ for some $\beta \in (0, 2)$.

This study of long-range dependence and self-similarity is a very rich and active subject but we will leave our discussion here and defer to more advanced books.

This concludes our study of the basic results in the study of second-order processes. In the sequel we will apply these results to problems arising in the filtering,prediction and smoothing of random signals in the presence of additive noise.

Problems

1. Test whether each of these functions can be the covariance function of some w.s.s. process.

   (a) $R(t,s) = 1 - |t - s|$, $0 \leq |t - s| \leq 1$ and $R(t,s) = 0$ otherwise.

   (b) $R(t,s) = e^{-a|t-s|}$, $-\infty < t, s < \infty$

   (c) $R(t,s) = 1$, $0 \leq |t - s| \leq 1$ and $R(t,s) = 0$ otherwise.

2. Let $R(t)$ be the covariance of a w.s.s. process and $\int_{-\infty}^{\infty} |R(t)| dt < \infty$.

   Define:
   $$R_T(t) = R(t) \mathbf{1}_{[|t| \leq T]}$$

   i.e. we truncate the covariance to a finite interval. Show that $R_T(t)$ need not define a covariance function.

   Hint: Consider $R_T(t) = 1, |t| \leq T$ and $R_T(t) = 1 - \frac{|t|}{T}$

3. Show that the following function cannot be the covariance function of any discrete-time process:

$$
\begin{aligned}
R(n) &= \quad \pi \ n = 0 \\
&= \quad 2 \ n = \pm 5 \\
&= \quad 3 \ n = \pm 15 \\
&= \quad 0 \ for \ all \ other \ n
\end{aligned}
$$

4. Let $\{S_n\}$ be a stationary stochastic sequence with spectral density $p(\lambda)$. Define, for fixed B, the continuous-time stochastic process $\{X_t\}$ by:

$$X_t = \sum_{k \in Z} S_n \frac{\sin \pi(2Bt - n)}{\pi(2Bt - n)} \quad -\infty < t < \infty$$

   where the convergence of the infinite series is in the mean squareed sense. Find the covariance of $X_{k\Delta}$ for fixed Δ and $2B\Delta \leq 1$. Specialize the result when

$$p(\lambda) = 1 \ \lambda \in [-\frac{1}{2}, \frac{1}{2}]$$

5. Show that if $R(t,s)$, $-\infty < t, s < \infty$ is a covariance function so is $aR(at, as)$ for any $a > 0$.

6. Let $\{W_t\}$ be a standard Brownian motion. Let:

$$X(t) = \sin(2\pi ft + W_t), \ t \geq 0$$

   Calculate the mean and covariance of $X(t)$.

   Show that:

$$
\begin{aligned}
\lim_{t \to \infty} E[X(t)] &= \quad 0 \\
\lim_{T \to \infty} R(T, T+t) &= \quad = \frac{\cos(2\pi ft)}{2} e^{-\frac{|t|}{2}}
\end{aligned}
$$

   or the process is asymptotically w.s.s.

7. A stochastic process is called lognormal if it is of the form:

$$Y(t) = e^{X(t)}, \quad -\infty < t < \infty$$

where $X(t)$ is a Gaussian process with mean m and covariance $R(\tau) = cov[X(t)X(t+\tau)]$.

Find the mean and covariance of $y(t)$. Does the spectral density exist? If so, compute it.

8. Let ξ_k be an iid sequence with mean 0 and variance σ^2. Let $x(t)$ be a stationary Gaussian process with mean 0 and spectral density $S(\lambda)$.

Define:

$$X_n = X(\sum_{k=1}^{n} \xi_k)$$

Find the mean and covariance of $\{X_n\}$.

The process $\{X_n\}$ is a randomly sampled version of $X(t)$.

9. Let $X(t)$ be a 0 mean w.s.s. process with covariance $R(\tau)$.

Show that if:

$$R(0) = R(1)$$

then:

$$R(t) = \sum_{0}^{\infty} a_n^2 \cos 2\pi nt; \quad \sum_{0}^{\infty} a_n^2 < \infty$$

10. Let $R(t), \ -\infty < t < \infty$ bethe covariance of a w.s.s. process with $\int_{-\infty}^{\infty} |R(t)| dt < \infty$. Assume $R''(t)$ exists.

Show that:

$$\frac{R'(t)}{t} \geq R''(0)$$

and

$$R''(t) \geq R''(0), \ t \geq 0$$

11. Let $\{X_n\}$ be a 0 mean w.s.s process with covariance function $R(\tau)$. and spectral density $S(\lambda)$.

Define:

$$Y(t) = X(t+C) - X(t-C); \quad C > 0$$

Find $R_Y(\tau)$ and $S_Y(\lambda)$ the covariance and spectral density of $Y(.)$.

12. Let $n(t)$ be a white noise process with mean 0 and covariance $\delta(t-s)$ (Dirac delta). Suppose $n(.)$ is the input to two LTIC systems with impulse responses $h(t)$ and $g(t)$ respectively with corresponding outputs $X(.)$ and $Y(.)$ respectively.

Show that $X(.)$ and $Y(.)$ are jointly w.s.s. and find $E[X(t)Y(t+\tau)]$.

13. Let $\{X_n\}$ be a real-valued stationary process with mean 0 and covariance $E[X_n X_{n+k}] = R(k)$ with $R(0) < \infty$.

Let

$$Y_n = \sum_{k=0}^{\infty} a_k X_{n-k}, \ n = 0, \pm 1, \ldots,$$

be a stochastic process defined from the stationary sequence $\{X_k\}$ with $\sum_{k=0}^{\infty} |a_k| < \infty$. Find the covariance $R_y(k)$ of $\{Y_n\}$ and show that the sequence $\{Y_n\}$ satisfies the SLLN.

Hint: $\sum_{k=0}^{\infty} |a_k| < \infty$ implies that $\sum_{k=0}^{\infty} |a_k|^2 < \infty$

14. Let $X(t), -\infty < t < \infty$ be a zero mean w.s.s. process with spectral density $S(\lambda)$ that is not necessarily bandlimited.

Define the following process:

$$\hat{X}(t) = \sum_{n=-\infty}^{\infty} X(\frac{n}{2B})a_n(t), \quad -\infty < t < \infty$$

where $a_n(t) = \frac{\sin(\pi(2Bt-n))}{2\pi(2Bt-n)}$.

Find the variance of $\hat{X}(t)$. What do you observe?

15. Show that the function

$$R(t) = e^{-t^2}$$

can be the covariance function of a w.s.s. process.

16. Let $\{X_n\}$ be a 0 mean stationary Gauss-Markov sequence with covariance function

$$R_x(k) = \rho^{|k|} \; ; \; k = 0, \pm 1, \pm 2, \dots,$$

with $\rho < 1$. The sequence $\{X_n\}$ is the input to a linear, time-invariant and causal (LTIC) system whose impulse response is

$$h_k = \frac{1}{a^k}, \; k \geq 0$$

with $a > 1$. The corresponding output is denoted by $\{Y_n\}$.

Find the variance $R_Y(0)$. Show that the output $\{Y_n\}$ satisfies the Strong Law of Large Numbers (SLLN) (the a.s. version) and, moreover, is ergodic.

17. Let $\{X_t\}$ be a zero mean w.s.s. process with spectral density $S_X(\omega)$. Suppose that $S_X(\omega) = 0$ if $|\omega| \geq 2\pi B$. Let $T = \frac{1}{2B}$. Suppose we want to find the best linear mean squared estimate of X_t given $\{X_{nT}\}_{n=-N}^{N}$ in the form:

$$\hat{X}_t = \sum_{-N}^{N} h(t - nT)X_{nT}$$

Find $h(t - nT)$.

This problem shows that the sampling theorem can be interpreted as a least squares representation on the space $L_2[(-\infty, \infty), \mathbb{P})$.

18. Let $R(t, s)$ given by the function in Problem 1 (a). Find the eigenvalues and eigenfunctions of $R(t, s)$.

Hint: Show that $R(t)$ must satisfy:

$$\frac{d^2 R(t)}{dt^2} = -2\delta(t)$$

REFERENCES

The basic material in this chapter can be found in any standard text on stochastic processes. However the following books are of particular note because of their comprehensive treatment of the topics and any serious reader who wishes to learn more about the rich and powerful theory as well as its applications are encouraged to consult them.

1. A. V. Balakrishnan, *Introduction to random processes in engineering*, J. Wiley, N.Y., 1995

   This book is a *tour de force* especially in its treatment of the second-order theory as well as the engineering applications.

2. H. Cramer and H. Leadbetter, *Stationary and related stochastic processes. Sample function properties and their applications.* John Wiley & Sons, Inc., 1967.

   This book is a classic but unfortunately out of print.

3. E. Wong and B. Hajek, *Stochastic processes in engineering systems*, Springer-Verlag, N.Y., 1985.

   This book is excellent for its treatment of second order processes.

4. J. Beran, *Statistics for long-memory processes*, Chapman and Hall, 1994

5. G. Samorodnitsky and M. S. Taqqu,*Stable non-Gaussian random processes. Stochastic models with infinite variance. Stochastic Modeling.* Chapman and Hall, New York, 1994.

   The last two books should be consulted for more details on long-range dependence, self-similarity, and other issues.