Prediction and Filtering of Stationary Processes:  
Yaglom’s Method and Minimax Filtering

by Philipp Mascher

A thesis submitted to the  
Department of Mathematics and Statistics  
in conformity with the requirements  
for the degree of Master of Science

Queen’s University  
Kingston, Ontario, Canada

August 5, 2015
Abstract

The aim of this work is to give a basic introduction to the theory of stationary stochastic processes, particularly to the somewhat specialized problem of prediction and filtering of such processes. The study of stationary processes arguably originated from Khinchin’s 1934 paper in the *Mathematische Annalen* [8]. Another common name for that theory, and in fact the original term used by Khinchin himself, is correlation theory.

As to the problem of predicting stationary processes, Kolmogorov [11, 12] was the first to make a contribution to its solution using involved mathematical theory. In the years following the publication of Wiener’s famous book [14], the theory gained considerable popularity from the applied sciences, particularly radio engineering. Wiener managed to obtain simple solutions for problems similar to those treated by Kolmogorov. Moreover, Wiener studied the problem of filtering, which is of great importance in engineering. Due to their impact in this field, the theory of prediction and filtering of stationary processes is often referred to as Wiener-Kolmogorov theory or Wiener theory.

As stated by Yaglom [15], Wiener’s approach tends to lack mathematical rigor by making “use of mathematically meaningless expressions.” In this work, we shall present Yaglom’s method to solving the problems considered in Wiener’s book. This alternative approach is entirely based on rather basic facts from Hilbert space theory and the theory of complex variables.

As it turns out, the theory of filtering of stationary processes heavily relies on spectral properties of the processes. In particular, Yaglom’s approach assumes complete knowledge of the spectral densities. In this work, however, we shall not be concerned with the problem of estimating such quantities based on a finite sample. For such matters, we refer the interested reader to the monograph [2] by Brillinger. Instead, in order to account for uncertainty as frequently encountered in practice, we shall discuss the problem of minimax filtering which has emerged from the practical need of allowing for incomplete knowledge about spectral properties.

This work is structured as follows: Part I briefly introduces the concept of stationary processes, and gives some preliminary technical results, such as spectral representations; Part II introduces the problem of prediction and filtering, and presents Yaglom’s solution of which with an application to rational spectral densities; Part III is dedicated to a review about minimax filtering.
Acknowledgement

I would like to thank Prof. Dr. Boris Levit for his support and helpful suggestions he provided throughout the course of writing this report. Further, I would like to thank the faculty and student body at the Department of Mathematics and Statistics at Queen’s University for the many fruitful discussions during my studies.
Contents

I  Introduction to Stationary Processes 6

1  Stochastic Processes and Stationarity 6

2  Spectral Representations 9
   2.1  Bochner-Khinchin Theorem  ................................................. 9
   2.2  Spectral Decomposition of Stationary Processes  ...................... 11

3  Linear Transformations 14
   3.1  Admissible Filters  .......................................................... 14
   3.2  Physical Realizability  ...................................................... 15
   3.3  Mean Square Differentiation  ............................................. 18

II  Prediction and Filtering 19

4  Prediction of Stationary Processes 19
   4.1  Formulation of the Problem  ............................................... 20
   4.2  Yaglom’s Solution  ........................................................... 23
      4.2.1  Sufficient Conditions on the Spectral Characteristic  ............... 24
      4.2.2  Examples  ..................................................................... 27
      4.2.3  Spectral Characteristic for Rational Spectral Densities .............. 29

5  Filtering of Stationary Processes 34
   5.1  Formulation of the Problem  ............................................... 35
   5.2  Solution for Non-Causal Filtering  ....................................... 38
   5.3  Yaglom’s Solution for Causal Filtering  .................................. 41
      5.3.1  Sufficient Conditions on the Spectral Characteristic ............... 41
      5.3.2  Example  ..................................................................... 42
## CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.3.3 Spectral Characteristic for Rational Spectral Densities</td>
<td>44</td>
</tr>
<tr>
<td>5.4 Filtering of Stationary Sequences</td>
<td>49</td>
</tr>
<tr>
<td>5.4.1 Yaglom’s Solution</td>
<td>49</td>
</tr>
<tr>
<td>5.4.2 Example</td>
<td>51</td>
</tr>
<tr>
<td>III Minimax Approach to Filtering</td>
<td>52</td>
</tr>
<tr>
<td>6 Minimax Filtering Problem and Solution</td>
<td>52</td>
</tr>
<tr>
<td>6.1 Practical Motivation</td>
<td>52</td>
</tr>
<tr>
<td>6.2 Problem Formulation</td>
<td>52</td>
</tr>
<tr>
<td>6.3 Least Favorable Spectral Densities</td>
<td>54</td>
</tr>
<tr>
<td>6.4 Examples</td>
<td>56</td>
</tr>
<tr>
<td>6.4.1 The p-Point Model</td>
<td>56</td>
</tr>
<tr>
<td>6.4.2 The Band Model and $\epsilon$-Contaminated Model</td>
<td>57</td>
</tr>
<tr>
<td>7 Analogy to Robust Hypotheses Testing</td>
<td>59</td>
</tr>
<tr>
<td>7.1 Robust Hypotheses Testing</td>
<td>59</td>
</tr>
<tr>
<td>7.2 Connection to Minimax Filtering</td>
<td>60</td>
</tr>
<tr>
<td>7.3 Example – Total-Variation Model</td>
<td>62</td>
</tr>
<tr>
<td>References</td>
<td>65</td>
</tr>
</tbody>
</table>
Part I

Introduction to Stationary Processes

1 Stochastic Processes and Stationarity

We start by reviewing some basic notions of the theory of stochastic processes. For a more comprehensive treatment, a reader can be recommended to the monograph [3] by Gikhman and Skorokhod.

Stochastic Processes

Definition 1.1. A stochastic process \( X = \{X(t)\}_{t \in T} \) is a collection of complex-valued random variables defined on a common probability space \( (\Omega, \mathcal{F}, P) \). The nature of \( T \) may be quite arbitrary.

We shall also use the term process when referring to a stochastic process. At times, we write \( X(t) \) when referring to the entire process \( X \); it should be clear from the context when we mean the value of \( X \) at \( t \) or the entire process \( X \). For the sake of convenience, we will refer to a stochastic process as a stochastic sequence in the discrete case, that is when \( T = \mathbb{Z} \). If we do not specify the nature of \( T \), we will usually use the term stochastic process. However, we will consider no other cases than \( T = \mathbb{Z} \) and \( T = \mathbb{R} \).

Remark 1.2. Oftentimes, it is convenient to understand a process \( X \) as some physical quantity evolving over time, such as voltage or current. In light of this, the random variable \( X(t) \) usually represents the value of that quantity at time \( t \). This being said, the mathematical notion of stochastic processes does not require physical justification. It is rather through such applications coming from diverse fields like physics, engineering, biology, finance and many others that the theory of stochastic processes is largely motivated and extensively being used.

Stationarity

The notion of stationarity is best illustrated by considering a physical quantity evolving over time. Assume an observer takes measurements of some quantity at different times, which can be understood as realizations, also called trajectories, of a process \( X \). Intuitively, we would expect those realizations to share common features over different instances of time. Mathematically, this corresponds to time-invariance of all finite distributions of \( X \), which is called strict-stationarity. In this work, however, we want to focus on a weaker concept called wide-sense stationarity.
Definition 1.3. A complex-valued stochastic process $X$ is said to be wide-sense stationary if

1. its second moment exists, that is $E|X(t)|^2 < \infty$,
2. its first moment is time-invariant, that is $E[X(t)] \equiv \text{const}$, and
3. the quantity $E[X(s)X(t)]$ depends on $s$ and $t$ only through the difference $s - t$.

Notation 1.4. Henceforth, $L^2(P)$ will denote the space of complex-valued random variables with finite second moment. Thus, (S1) says that, for a stationary process $X$, the random variable $X(t)$ belongs to $L^2(P)$.

Assumption 1.5. We shall assume $E[X(t)] \equiv 0$ without loss of generality, which is always possible by considering the process $X(t) - E[X(t)]$ instead of $X(t)$.

The concept of wide-sense stationarity is known under several different names in the mathematical literature, without clear preference for one or another. Common names alongside wide-sense stationarity are weak stationarity, second-order stationarity, and covariance stationarity. Since we are exclusively dealing with the concept of wide-sense stationarity throughout this work, we shall neglect the different terms altogether and instead write stationary with the implicit understanding that we mean wide-sense stationary.

Definition 1.6. The function $B(t) = E[X(t)\overline{X(0)}]$ is called the covariance function of the stationary process $X = \{X(t)\}$. If $X$ is indexed by real numbers, we let $B$ be defined on the real line; if $X$ is indexed by integers, that is when $X$ is a stationary sequence, we let $B$ be defined on the integers.

Remark 1.7. (i) We observe that the quantity in (S3) is the covariance of $X(s)$ and $X(t)$ since $E[X(t)] \equiv 0$ without loss of generality. As imposed in (S3), it depends on $s$ and $t$ only through $s - t$. Therefore, we obtain $B(t) = E[X(s + t)\overline{X(s)}]$ for all $s$.

(ii) There are several differing names for this function in the literature, such as auto-covariance function, correlation function, and auto-correlation function. Nonetheless, we shall stick to the term covariance function.

Proposition 1.8. The covariance function $B$ of a stationary process $X$ satisfies the following:

(i) $B(0)$ is a non-negative real number.

(ii) $B$ is Hermitian, that is $B(-t) = \overline{B(t)}$.

(iii) $B$ is bounded, namely $|B(t)| \leq B(0)$.
Proof. (i) follows from $B(0) = E[X(0)]^2 \geq 0$ and $E[X(0)]^2 < \infty$. (ii) follows from $B(-t) = E[X(-t)X(0)] = E[X(-t)X(0)] = B(0 - (-t)) = B(t)$. (iii) follows from an application of the Cauchy-Schwarz inequality: $|B(t)|^2 = |E[X(t)X(0)]|^2 \leq E[X(t)]^2 E[X(0)]^2 = B(0)^2$.

Assumption 1.9. In addition to these properties, we shall always assume that $B$ is continuous.

Remark 1.10. This is as a matter of fact both a desirable and in practice frequently encountered situation, for instance in physical applications where quantities such as voltage, current, or radio signals are thought of as continuous quantities.

As it turns out, the covariance function $B$ of a stationary process $X$ is continuous if and only if the process $X$ is mean square continuous.

Definition 1.11. A stochastic process $X$ is said to be mean square continuous at $t$ whenever

$$\lim_{h \to 0} E|X(t+h) - X(t)|^2 = 0.$$  \hfill (1)

If $X$ is mean square continuous at all $t$, then it is simply said to be mean square continuous.

At times, we shall simply write that a process is continuous with the implicit understanding that we actually mean that the process is mean square continuous.

Proposition 1.12. The covariance function $B$ of a stationary process $X$ is continuous at $t = 0$ if and only if $X$ is mean square continuous.

Proof. Let us first show the necessity part. For fixed $t$, we observe

$$E|X(t+h) - X(t)|^2$$
$$= E|X(t+h)|^2 + E|X(t)|^2 - E[X(t+h)X(t)] - E[X(t)X(t+h)]$$
$$= 2B(0) - B(h) - B(-h).$$

The latter goes to zero as $h \to 0$ by continuity of $B$ at $t = 0$. The sufficiency part goes as follows. For arbitrary $t$, we obtain by virtue of the Cauchy-Schwarz inequality

$$|B(t+h) - B(t)|^2 = \left|E[(X(t+h) - X(t))X(0)]\right|^2 \leq E|X(t+h) - X(t)|^2 E|X(0)|^2$$

where the first term in the product on the right-hand side converges to zero as $h \to 0$. In particular, $B$ is continuous at $t = 0$. \qed
2 Spectral Representations

The results in this section, most importantly the Bochner-Khinchin Theorem as well as the spectral decomposition of stationary processes, are drawn from the monograph [3] by Gikhman and Skorokhod. Recall that we always mean that a stationary processes is mean square continuous when merely using the term continuous.

2.1 Bochner-Khinchin Theorem

The following spectral representation of the covariance function of a stationary process is largely due to Bochner [1] and Khinchin [8], who proved this result independently around the same time.

**Theorem 2.1. (Bochner-Khinchin)** Let \( B \) be the covariance function of a continuous stationary process. Then there exists a bounded non-decreasing right-continuous function \( F : \mathbb{R} \to [0, \infty) \) such that

\[
B(t) = \int_{-\infty}^{\infty} e^{it\lambda} dF(\lambda).
\]

The choice of \( F \) is unique provided that \( F(u) \to 0 \) as \( u \to -\infty \).

**Definition 2.2.** The function \( F \) in the representation (2) of the covariance function of a continuous stationary process is called a **spectral function**. If \( F \) is absolutely continuous, a function \( f \) satisfying

\[
F(u) = \int_{-\infty}^{u} f(\lambda)d\lambda
\]

is called a **spectral density** associated with the covariance function, or with the underlying stationary process. Further, we let \( L^2(F) \) denote the space of complex-valued measurable functions \( g \) such that \( \int |g(\lambda)|^2 dF(\lambda) < \infty \).

**Assumption 2.3.** Henceforth, we shall always assume that the spectral function \( F \) associated with a stationary process satisfies \( F(u) \to 0 \) as \( u \to -\infty \).

**Proposition 2.4.** In the following, let \( B \) denote the covariance function of a continuous stationary process with associated spectral function \( F \). Then, the following holds:

(i) \( F(u) \to B(0) \) as \( u \to \infty \).

(ii) If \( f \) is a spectral density associated with \( B \), then

\[
B(t) = \int_{-\infty}^{\infty} e^{it\lambda} f(\lambda)d\lambda.
\]
(iii) If \( B \) is integrable, that is \( \int |B(t)|dt < \infty \), then \( B \) admits a spectral density \( f \) given by the inverse Fourier transform of \( B \), that is
\[
f(\lambda) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-it\lambda} B(t) dt. \tag{5}
\]
In particular, \( f \) is bounded and continuous in that case.

**The Discrete Case**

In the following, we shall give a version of the Bochner-Khinchin Theorem for covariance functions of stationary sequences. Recall that the covariance function is defined on the integers in this case.

**Theorem 2.5.** Let \( B \) be the covariance function of a stationary sequence. Then there exists a bounded non-decreasing right-continuous function \( F : (-\pi, \pi] \rightarrow [0, \infty) \) such that
\[
B(t) = \int_{-\pi}^{\pi} e^{it\lambda} dF(\lambda). \tag{6}
\]
The choice of \( F \) is unique provided that \( F(u) \rightarrow 0 \) as \( u \rightarrow -\pi \).

**Definition 2.6.** The function \( F \) in the representation (6) of the covariance function of a stationary sequence is called a spectral function. If \( F \) is absolutely continuous, a function \( f \) satisfying
\[
F(u) = \int_{-\pi}^{u} f(\lambda) d\lambda \tag{7}
\]
is called a spectral density associated with the covariance function, or with the underlying stationary sequence.

**Assumption 2.7.** Henceforth, we shall always assume that the spectral function \( F \) associated with a stationary sequence satisfies \( F(u) \rightarrow 0 \) as \( u \rightarrow -\pi \).

**Proposition 2.8.** In the following, let \( B \) denote the covariance function of a stationary sequence with associated spectral function \( F \). Then, the following holds:

(i) \( F(u) \rightarrow B(0) \) as \( u \rightarrow \pi \).

(ii) If \( f \) is a spectral density associated with \( B \), then
\[
B(t) = \int_{-\pi}^{\pi} e^{it\lambda} f(\lambda) d\lambda. \tag{8}
\]
2.2 Spectral Decomposition of Stationary Processes

(iii) If $B$ is summable, that is $\sum_{t \in \mathbb{Z}} |B(t)| < \infty$, then $B$ admits a spectral density $f$ given by the Fourier series

$$f(\lambda) = \frac{1}{2\pi} \sum_{t=\infty}^{\infty} e^{-it\lambda} B(t).$$

In particular, $f$ is bounded and continuous in that case.

2.2 Spectral Decomposition of Stationary Processes

At this point, it should be pointed out that there are two representations of the covariance function $B$ of a continuous stationary process $X$, namely

$$B(t) = E[X(t)X(0)]$$

and

$$B(t) = \int_{-\infty}^{\infty} e^{it\lambda} dF(\lambda).$$

Therefore, one might suspect some deeper connection between exponential functions $e^{it\lambda}$ on the real line and the random variables $X(t)$ which indeed exists, and is called the **spectral decomposition of stationary processes**. Its original formulation in terms of the geometry of the Hilbert space $L^2(P)$ is due to Kolmogorov [10, 11] and dates back to the 1940s. Before we can state the main result, Theorem 2.12 on the following page, we need to make a few definitions.

**Definition 2.9.** For a measure $\mu$ defined on some $\sigma$-algebra $\mathcal{L}$, let $\mathcal{L}_0$ denote the class of all sets $A \in \mathcal{L}$ such that $\mu(A) < \infty$. A mapping $Z$ from $\mathcal{L}_0$ to $L^2(P)$ is called **stochastic orthogonal measure** with reference measure $\mu$ if

(O1) $Z(\bigcup_{n \geq 1} A_n) = \sum_{n \geq 1} Z(A_n)$ with probability 1 for any countable collection of disjoint sets $\{A_n\}$ in $\mathcal{L}_0$;

(O2) $E[Z(A)\overline{Z(B)}] = \mu(A \cap B)$ for all $A, B \in \mathcal{L}_0$.

**Remark 2.10.** The term **orthogonal** comes from the fact that $E[Z(A)\overline{Z(B)}] = 0$ whenever $A$ and $B$ are disjoint, due to (O2).

For a detailed introduction to stochastic measures and stochastic integrals, we refer the interested reader to the monograph [3] by Gikhman and Skorokhod, Chapter 5 Section 3. At this point, we shall merely state an important property of stochastic orthogonal measures known as the **Ito isometry**; a proof of this result can be found in [3]. Let $Z$ be a stochastic orthogonal measure with reference measure $\mu$ defined on the real Borel sets. If $f$ and $g$ belong to $L^2(\mu)$ then

$$E \left[ \left( \int_{-\infty}^{\infty} g(\lambda) dZ(\lambda) \right) \left( \int_{-\infty}^{\infty} f(\lambda) d\overline{Z(\lambda)} \right) \right] = \int_{-\infty}^{\infty} g(\lambda) \overline{f(\lambda)} d\mu(\lambda).$$

(9)
Definition 2.11. For a stationary process $X$, let $H_X$ denote the closed linear subspace of $L^2(P)$ spanned by $\{X(t)\}$.

Theorem 2.12. ([3], Theorem 5.4.2) A continuous stationary process $X$ satisfying $E[X(t)] \equiv 0$ with spectral function $F$ has a representation of the form

$$X(t) = \int_{-\infty}^{\infty} e^{it\lambda} dZ(\lambda) \text{ for all } t$$

with probability 1, where $Z$ is some stochastic orthogonal measure with reference measure $\mu$ defined by $\mu((\infty, u]) = F(u)$. $Z$ is unique in $L^2(P)$ sense. Between $H_X$ and $L^2(F)$ there exists an isometric correspondence under which

(i) $X(t) \leftrightarrow e^{it\lambda}$ and $Z(A) \leftrightarrow I_A(\lambda)$, where $I_A$ denotes the indicator function of $A$;

(ii) if $V_k \leftrightarrow g_k$ for $k = 1, 2$ then

$$V_k = \int_{-\infty}^{\infty} g_k(\lambda)dZ(\lambda), \ E[V_k] = 0, \text{ and } E[V_1V_2] = \int_{-\infty}^{\infty} g_1(\lambda)\overline{g_2(\lambda)}dF(\lambda).$$

Definition 2.13. Formula (10) is called the spectral decomposition of the stationary process $X$, and the stochastic orthogonal measure $Z$ is called the stochastic spectral measure of the process. We shall refer to the last identity in (11) as the Ito isometry.

Corollary 2.14. It immediately follows from Theorem 2.12:

(i) The stochastic spectral measure $Z$ satisfies $E[Z(A)] \equiv 0$.

(ii) A function $g$ belongs to $L^2(F)$ if and only if the random variable $V = \int_{-\infty}^{\infty} g(\lambda)dZ(\lambda)$ belongs to $H_X$.

Conversely, Yaglom shows in his monograph [15], Chapter 2 Section 9 p39, that there is an inversion formula for recovering the stochastic spectral measure only using the stationary process to begin with. At this point, however, we shall not provide further details, but rather refer the interested reader to [15].

Spectral Decomposition when a Spectral Density Exists

In the following, let us assume the stationary process $X$ assumes a spectral density $f$. For this case, we shall give an alternative spectral decomposition of $X$ with similar distributional properties.
Definition 2.15. A stochastic orthogonal measure $W$ is called a *standard stochastic orthogonal measure* if

(SO1) the reference measure associated with $W$ is the Lebesgue measure on the real line;
(SO2) $E[W(A)] = 0$ for all real Borel sets $A$.

Having this definition in place, we at once arrive at the following Corollary to Theorem 2.12 on the preceding page.

Corollary 2.16. Let $X$ be a continuous stationary process satisfying $E[X(t)] \equiv 0$ with covariance function $B$ and spectral density $f$. For a standard orthogonal measure $W$, define a process $\tilde{X}$ by

$$\tilde{X}(t) = \int_{-\infty}^{\infty} e^{it\lambda} \sqrt{f(\lambda)} dW(\lambda).$$

Then, $\tilde{X}$ is a stationary process with $E[\tilde{X}(t)] \equiv 0$ and covariance function $B$.

Proof. Note that $E[|\tilde{X}(t)|^2] = \int f(\lambda) d\lambda < \infty$ for all $t$, and $E[\tilde{X}(t)] = 0$ for all $t$ since $E[\int g(\lambda) d\lambda] = 0$ for all integrable $g$. Further, for all $s$, an application of the Ito isometry and minding non-negativity of $f$ yields the equality

$$E[\tilde{X}(s + t)\tilde{X}(s)] = \int_{-\infty}^{\infty} e^{it\lambda} |\sqrt{f(\lambda)}|^2 d\lambda = \int_{-\infty}^{\infty} e^{it\lambda} f(\lambda) d\lambda = B(t).$$

Remark 2.17. All of the results in Section 2.2 hold analogously for stationary sequences, only that the bounds of integration are $-\pi$ and $\pi$ rather than $-\infty$ and $\infty$. 

3 Linear Transformations

As we shall see later in the theory of prediction and filtering, we will need to calculate integrals \( \int \phi(t-s)X(s)ds \) and derivatives \( X'(t) \) in some appropriate sense. In this section, we shall briefly summarize how these concepts are to be understood in the context of stationary processes. In particular, we will introduce the notion of an admissible filter and explain what it means for such a filter to be physically realizable. Throughout the entire section, we shall assume that all stationary processes considered are mean square continuous.

3.1 Admissible Filters

This section is largely based on the monograph \([3]\) by Gikhman and Skorokhod, Chapter 5 Section 5.

**Definition 3.1.** For a stationary process \( X = \{X(t)\} \), a stochastic process \( L = \{L(t)\} \) is called an admissible filter, or simply filter, with input \( X \) if there is a sequence \( \phi_n \in L^1(\mathbb{R}) \cap L^2(\mathbb{R}) \) such that, for all \( t \), \( L(t) \) is the \( L^2(P) \) limit of integrals of the form

\[
\int_{-\infty}^{\infty} \phi_n(t-s)X(s)ds \quad \text{as } n \to \infty.
\]

(12)

**Remark 3.2.** We understand the integral (12) as the \( L^2(P) \) limit of integrals of the form

\[
\int_{-a}^{b} \phi_n(t-s)X(s)ds
\]

as \( a, b \to \infty \). Further, note that expressions like (13) require that \( X \), which really is a function on \( \mathbb{R} \times \Omega \), be measurable with respect to the completion of the product sigma-algebra \( \mathcal{B}(\mathbb{R}) \times \mathcal{F} \); here, \( \mathcal{B}(\mathbb{R}) \) denotes the class of real Borel sets. At this point, we merely state that stationary processes satisfy this technicality; for more details, we refer the interested reader to Gikhman and Skorokhod \([3]\), Chapter 5 Section 3.

**Proposition 3.3.** ([3], Theorem 5.5.1 and 5.5.2) Let \( X \) be a stationary process with spectral function \( F \) and stochastic spectral measure \( Z \). A stochastic process \( L \) is a filter with input \( X \) if and only if there exists \( \Phi \in L^2(F) \) such that \( L \) is a stationary process with spectral decomposition and covariance function \( B_L \) given by

\[
L(t) = \int_{-\infty}^{\infty} e^{it\lambda} \Phi(\lambda) dZ(\lambda) \quad \text{and} \quad B_L(t) = \int_{-\infty}^{\infty} e^{it\lambda} |\Phi(\lambda)|^2 dF(\lambda).
\]

(14)
Definition 3.4. The function $\Phi$ associated with a filter $L$, such as is (14), is called the spectral characteristic of that filter.

The following result describes how the spectral characteristic of a filter can be calculated.

**Lemma 3.5.** Let $X$ be a stationary process with spectral function $F$ and stochastic spectral measure $Z$, let $L$ be a filter with input $X$, and let $\phi_n \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ be such that $\int_{-\infty}^{\infty} \phi_n(t-s)X(s)ds$ converges in $L^2(P)$ to $L(t)$ as $n \to \infty$ for all $t$. Then, the sequence of functions $\Phi_n$ defined by

$$
\Phi_n(\lambda) = \int_{-\infty}^{\infty} \phi_n(s)e^{-is\lambda}ds
$$

belongs to $L^2(F)$, and its $L^2(F)$ limit is the spectral characteristic of $L$.

**Proof.** First, $\Phi_n$ belongs to $L^2(F)$ because an application of the triangle inequality yields

$$
\int_{-\infty}^{\infty} \left| \int_{-\infty}^{\infty} \phi_n(s)e^{-is\lambda}ds \right|^2 dF(\lambda) \leq \left( \int_{-\infty}^{\infty} |\phi_n(s)|ds \right)^2 \left( \int_{-\infty}^{\infty} dF(\lambda) \right) < \infty,
$$

where the last strict inequality follows from $\phi_n \in L^1(\mathbb{R})$ and $\int dF(\lambda) = E|X(0)|^2 < \infty$.

Let $L_n$ be a stochastic process defined by $L_n(t) = \int \phi_n(t-s)X(s)ds$. On the basis of Lemma 5 of Chapter 5 Section 3 in Gikhman and Skorokhod [3] on pp197, we obtain

$$
L_n(t) = \int_{-\infty}^{\infty} \left( \int_{-\infty}^{\infty} \phi_n(t-s)e^{is\lambda}ds \right) dZ(\lambda)
$$

$$(s \to t-s) \quad = \int_{-\infty}^{\infty} \left( \int_{-\infty}^{\infty} \phi_n(s)e^{i(t-s)\lambda}ds \right) dZ(\lambda)
$$

$$
= \int_{-\infty}^{\infty} e^{it\lambda} \Phi_n(\lambda)dZ(\lambda).
$$

Let the $L^2(P)$ limit of $L_n(t)$ be $L(t) = \int e^{it\lambda} \Phi(\lambda)dZ(\lambda)$ for some $\Phi \in L^2(F)$, and note that $E|L(t) - L_n(t)|^2 \to 0$ if and only if $\int |\Phi - \Phi_n|^2dF \to 0$. Thus, $\Phi$ is the $L^2(F)$ limit of $\Phi_n$. This completes the proof. \(\square\)

### 3.2 Physical Realizability

**Definition 3.6.** Let $X$ be a stationary process. A stochastic process $L$ is called a physically realizable filter, or simply realizable filter, with input $X$ if $L$ satisfies Definition 3.1 on the previous page where the integral (12) is replaced by

$$
\int_{0}^{\infty} \phi_n(s)X(t-s)ds.
$$

15
In other words, the value $L(t)$ of a realizable filter $L$ at $t$ depends only on quantities $X(s)$ with $s \leq t$, and does not depend on any $X(s)$ with $s > t$.

**Remark 3.7.** Note that (16) may be rewritten as $\int_{-\infty}^{\infty} \phi_n(t - s)X(s)\,ds$ provided that we let $\phi_n(s) = 0$ for all $s < 0$.

**Definition 3.8.** Let $\mathcal{R}$ be the space comprised of all functions $\Phi$ for which there exists a function $\phi$ in $L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ such that $\Phi(\lambda) = \int_{-\infty}^{\infty} \phi(s)e^{-is\lambda}\,ds$ for all $\lambda$. The symbol $\mathcal{R}$ stands for “realizable.”

**Lemma 3.9.** $\mathcal{R}$ is a linear subspace of $L^2(F)$ for any spectral function $F$.

**Remark 3.10.** $\mathcal{R}$ is not necessarily closed in $L^2(F)$.

**Proof.** Let $\phi \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$. Then,

$$\int_{-\infty}^{\infty} \left( \int_{0}^{\infty} \phi(s)e^{-is\lambda}\,ds \right)^2 \,dF(\lambda) \leq \int_{-\infty}^{\infty} \left( \int_{0}^{\infty} |\phi(s)|\,ds \right)^2 \,dF(\lambda) < \infty$$

because $\phi \in L^1(\mathbb{R})$ and $\int dF(\lambda) < \infty$. Linearity is immediate.

**Definition 3.11.** Let $F$ be a spectral function. We define $L^2(F)^-$ to be the $L^2(F)$ closure of the linear span of functions of the form $\lambda \mapsto e^{is\lambda}$ for $s < 0$.

The following two results are due to the author of this work. Later on, we shall use those in proving one of the main results due to Yaglom presented in this work, namely Theorem 4.13 on page 24.

**Theorem 3.12.** Let $F$ be a spectral function. Then, $L^2(F)^-$ is the $L^2(F)$ closure of $\mathcal{R}$.

**Proof.** Let $\Phi \in L^2(F)^-$. We need to show that there exists a sequence $\phi_m \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ such that the one-sided Fourier transforms $\int_{0}^{\infty} \phi_m(u)e^{-iu\lambda}\,du$ converge in $L^2(F)$ to $\Phi$ as $m \to \infty$.

Let $N(s, \epsilon)$ denote the normal distribution with mean $s$ and variance $\epsilon$, and let $U_{s,\epsilon} \sim N(s, \epsilon)$ be random variables. For $\epsilon > 0$, let $K_\epsilon$ denote the probability density function of $N(0, \epsilon)$, that is

$$K_\epsilon(u) = \frac{1}{\sqrt{2\pi \epsilon}} \exp \left( -\frac{u^2}{2\epsilon} \right).$$

For the remainder, let $s > 0$ be arbitrary but fixed. It is known that $\int K_\epsilon(s - u)e^{-iu\lambda}\,du = \exp(-is\lambda - \frac{1}{2\epsilon} \lambda^2)$ since the integral is the characteristic function of $N(s, \epsilon)$ evaluated at $-\lambda$, that is $E[\exp(i(-\lambda)U_{s,\epsilon})]$. 


Note that \( \exp(-is\lambda - \frac{\lambda^2}{2}) \) converges uniformly in \( \lambda \) to \( e^{-is\lambda} \) as \( \epsilon \downarrow 0 \). Further, \( \int_{-\infty}^{0} K_\epsilon(s - u)e^{-iu\lambda}du \to 0 \) uniformly in \( \lambda \) as \( \epsilon \downarrow 0 \) because

\[
\left| \int_{-\infty}^{0} K_\epsilon(s - u)e^{-iu\lambda}du \right| \leq \int_{-\infty}^{0} K_\epsilon(s - u)du = P(U_{s,\epsilon} \leq 0)
\]

which converges to zero as \( \epsilon \downarrow 0 \) since \( s > 0 \). Let \( G_\epsilon(\lambda, s) = \int_{0}^{\infty} K_\epsilon(s - u)e^{-iu\lambda}du \). For all \( s > 0 \), we have observed

\[
\sup_{\lambda \in \mathbb{R}} |G_\epsilon(\lambda, s) - e^{-is\lambda}| \to 0 \quad \text{as} \quad \epsilon \downarrow 0.
\]

Now, \( \Phi \in L^2(F)^- \) if and only if there exists a sequence of functions \( \Phi_m \) which are linear combinations of \( \lambda \mapsto e^{is\lambda} \) for negative \( s \). That is, for each \( m \), there exists a positive integer \( n^{(m)} \) such that for all \( k = 1, \ldots, n^{(m)} \) there exist some \( s_k^{(m)} < 0 \) and complex numbers \( c_k^{(m)} \) so that

\[
\Phi_m(\lambda) = \sum_{k=1}^{n^{(m)}} c_k^{(m)} e^{is_k^{(m)}\lambda} \quad \text{for all} \quad \lambda.
\]

Next, let \( \phi_m \) be defined by \( \phi_m(u) = \sum_{k=1}^{n^{(m)}} c_k^{(m)} K_m(-s_k^{(m)} - u) \), and note that \( \int_{0}^{\infty} \phi_m(u)e^{-iu\lambda}du = \sum_{k=1}^{n^{(m)}} c_k^{(m)} G_\epsilon(\lambda, -s_k^{(m)}) \) by definition of \( G_\epsilon \).

Since \( G_\epsilon(\lambda, -s_k^{(m)}) \to e^{is_k^{(m)}\lambda} \) uniformly in \( \lambda \) as \( \epsilon \downarrow 0 \), we observe the uniform convergence

\[
\sup_{\lambda \in \mathbb{R}} \left| \sum_{k=1}^{n^{(m)}} c_k^{(m)} G_\epsilon(\lambda, -s_k^{(m)}) - \Phi_m(\lambda) \right| \leq \sum_{k=1}^{n^{(m)}} |c_k^{(m)}| \sup_{\lambda \in \mathbb{R}} |G_\epsilon(\lambda, -s_k^{(m)}) - e^{is_k^{(m)}\lambda}| \to 0 \quad \text{as} \quad \epsilon \downarrow 0.
\]

Now, let \( \delta > 0 \) be arbitrary but fixed. Choose \( m \) large enough such that \( \int |\Phi - \Phi_m|^2 d\mu < \delta^2 \).

For that \( m \), let \( \epsilon_m \) be small enough such that \( \sup_{\lambda} |\Phi_m(\lambda) - \int_{0}^{\infty} \phi_m(u)e^{-iu\lambda}du| < \delta \). Then, the triangle inequality yields

\[
\left( \int_{-\infty}^{\infty} \left| \Phi(\lambda) - \int_{0}^{\infty} \phi_m(u)e^{-iu\lambda}du \right|^2 d\mu(\lambda) \right)^{1/2} \leq \delta + \delta \left( \int_{-\infty}^{\infty} d\mu(\lambda) \right)^{1/2}.
\]

The right-hand side goes to zero as \( \delta \downarrow 0 \). Note that \( \phi_m \in L^1(\mathbb{R}) \cap L^2(\mathbb{R}) \) because this already holds for the Gaussian probability density function. This completes the proof.

**Corollary 3.13.** Let \( X \) be a stationary process with spectral function \( F \). A function \( \Phi \) is the spectral characteristic of a realizable filter \( L \) with input \( X \) if and only if \( \Phi \) belongs to \( L^2(F)^- \).
Proof. Note that a function $\Phi$ is the spectral characteristic of a realizable filter $L$ with input $X$ if and only if $\Phi$ is the $L^2(F)$ limit of functions $\Phi_n$ that all belong to $\mathcal{R}$, which follows from the definition of realizable filters and Lemma 3.5 on page 15. This means that $\mathcal{R}$ is dense (in $L^2(F)$ sense) in the space of spectral characteristics of all realizable filters with input $X$. On the other hand, $\mathcal{R}$ is also dense in $L^2(F)$ by Theorem 3.12 on page 16. Therefore, $L^2(F)$ coincides with the space of spectral characteristics of realizable filters with input $X$.  

3.3 Mean Square Differentiation

Let us now introduce the notion of mean square differentiability. The results in this section are drawn from Gikhman and Skorokhod [3], Chapter 5 Section 2, and from Yaglom [15], Chapter 2 Section 11.

Definition 3.14. A stochastic process $X$ is called mean square differentiable at $t$ if

$$X(t + h) - X(t)$$

converges in $L^2(P)$ as $h \to 0$. We shall denote its limit as $X'(t)$, that is

$$\lim_{h \to 0} E \left| \frac{X(t + h) - X(t)}{h} - X'(t) \right|^2 = 0. \tag{17}$$

If $X$ is mean square differentiable at all $t$, we shall say that $X$ is mean square differentiable, or simply differentiable. $X'(t)$ shall be called the derivative of $X(t)$.

Theorem 3.15. ([3], Corollary 5.2.2) A continuous stationary process $X$ with covariance function $B$ is mean square differentiable if and only if the second derivative of $B$ exists at $t = 0$. Then, the derivative $B''(t)$ exists for all $t$, the process $X'$ is stationary, and its covariance function $B^{[1]}$ is

$$B^{[1]}(t) = -B''(t).$$

Proposition 3.16. Let $X$ be a mean square continuous stationary process with covariance function $B$. Then, $B''(t)$ exists at $t = 0$ if and only if $\int \lambda^2 dF(\lambda) < \infty$, in which case the derivative $X'$, its covariance function $B^{[1]}$, and its spectral function $F^{[1]}$ are respectively given by

$$X'(t) = \int_{-\infty}^{\infty} e^{it\lambda} i\lambda dZ(\lambda), \quad B^{[1]}(t) = \int_{-\infty}^{\infty} e^{it\lambda} \lambda^2 dF(\lambda), \quad F^{[1]}(u) = \int_{-\infty}^{u} \lambda^2 dF(\lambda).$$

Further, the $n$th derivative $X^{(n)}$ exists if and only if $\int \lambda^{2n} dF(\lambda) < \infty$, in which case

$$X^{(n)}(t) = \int_{-\infty}^{\infty} e^{it\lambda} (i\lambda)^n dZ(\lambda), \quad B^{[n]}(t) = \int_{-\infty}^{\infty} e^{it\lambda} \lambda^{2n} dF(\lambda), \quad F^{[n]}(u) = \int_{-\infty}^{u} \lambda^{2n} dF(\lambda).$$
Part II

Prediction and Filtering

4 Prediction of Stationary Processes

Motivation In many cases of practical interest, the only knowledge about \( X \) concerns past observations \( \{ X(s) \}_{s \leq t} \). In this context, we shall understand the real number \( t \) as the present time. The problem of prediction is then to find a reasonable estimator \( L_\tau(t) \) for \( X(t + \tau) \) with positive lag \( \tau \) based on the quantities \( X(s) \) for \( s \leq t \). As a criterion of estimation quality, we shall employ the mean square error between \( X(t + \tau) \) and \( L_\tau(t) \).

Assumption 4.1. Henceforth, we will always assume that \( X = \{ X(t) \} \) is a mean square continuous stationary process. Without loss of generality, we assume \( E[X(t)] = 0 \). Moreover, we shall assume its covariance function \( B \) defined by \( B(t) = E[X(t)X(0)] \) is known.

We shall restrict ourselves to linear estimators: first, the theory of linear estimators for stationary processes is well developed; and second, in practice one often encounters Gaussian processes for which the best estimator is always linear. We proceed our discussion with a brief introductory example.

Example 4.2. Suppose we have observed \( X \) at \( n \) past times \( t_k \leq 0 \), and now we aim to estimate \( X(\tau) \), where \( \tau > 0 \). Any linear predictor is then of the form

\[
\alpha_1 X(t_1) + \ldots + \alpha_n X(t_n),
\]

with complex coefficients \( \alpha_k \). Since we are interested in the predictor minimizing the mean squared error, we need to study the quantity

\[
E[|X(\tau) - \alpha_1 X(t_1) - \ldots - \alpha_n X(t_n)|^2].
\]

Now, finding the minimum of this quantity with respect to the \( \alpha_k \) corresponds to — geometrically speaking — dropping a perpendicular from \( X(\tau) \) onto the linear subspace of \( L^2(P) \) spanned by \( \{ X(t_1), \ldots, X(t_n) \} \) whose elements are exactly of the form (18). Since the perpendicular is to be orthogonal to the whole subspace we require

\[
E \left[ (X(\tau) - \alpha_1 X(t_1) - \ldots - \alpha_n X(t_n))X(t_k) \right] = 0 \quad \text{for all } k.
\]

Hence, it suffices to find coefficients \( \alpha_k \) satisfying this relation.
Note that by using the covariance function $B$, we can rewrite
\[
E \left[ (X(\tau) - \alpha_1 X(t_1) - \ldots - \alpha_n X(t_n)) X(t_k) \right] = B(\tau - t_k) - \sum_{l=1}^{n} \alpha_k B(t_l - t_k).
\]

Now, setting the right-hand side of the above expression equal to zero results in a system of $n$ linear equations with $n$ unknowns, which can be easily solved.

### 4.1 Formulation of the Problem

#### Hilbert Space Formulation

Note that $L^2(P)$ is a Hilbert space with inner product defined by $\langle U, V \rangle = E[U \overline{V}]$ for $U$ and $V$ belonging to $L^2(P)$. As such, we may invoke the rich body of Hilbert space theory.

**Definition 4.3.** For a stationary process $X$, let $H_X(t)$ denote the closure of the linear subspace of $L^2(P)$ spanned by the random variables $\{X(s)\}_{s \leq t}$.

Note that $H_X(t)$ is a closed linear subspace of $L^2(P)$. We observe the following: $V$ belongs to $H_X(t)$ if and only if it is a linear combination of an arbitrary number of elements in $\{X(s)\}_{s \leq t}$, such as in (18), or if it is the $L^2(P)$ limit of such combinations.

**Definition 4.4.** Let $\sigma^2_\tau$ denote the smallest mean square error when approximating $X(t + \tau)$ from within $H_X(t)$, that is
\[
\sigma^2_\tau = \inf_{V \in H_X(t)} E |X(t + \tau) - V|^2. \tag{19}
\]

We shall also refer to $\sigma^2_\tau$ as the *prediction error*, or simply *error*. We shall see in Proposition 4.11 on page 23 that $\sigma^2_\tau$ does not depend on $t$.

Since we are dealing with Hilbert spaces, we observe that there exists a unique $L_\tau(t)$ in $H_X(t)$ minimizing the prediction error, that is
\[
\sigma^2_\tau = E |X(t + \tau) - L_\tau(t)|^2. \tag{20}
\]

Further, the perpendicular $X(t + \tau) - L_\tau(t)$ is orthogonal to $H_X(t)$ which is equivalent to
\[
E \left[ (X(t + \tau) - L_\tau(t)) \overline{X(t - s)} \right] = 0 \quad \text{for all } s \geq 0. \tag{21}
\]
**Definition 4.5.** The stochastic process \( L_\tau = \{L_\tau(t)\} \) comprised of the unique elements \( L_\tau(t) \) in \( H_X(t) \) satisfying (21) is said to be the best linear predictor with lag \( \tau \).

**Remark 4.6.** In general, we shall be interested in the problem of prediction when the entire past is known. This is a natural idealization of the setting when \( X(s) \) had been observed at times \( -T \leq s \leq t \) under the assumption of very large \( T \).

### Spectral Decomposition of the Best Linear Predictor

**Proposition 4.7.** The best linear predictor \( L_\tau \) is a stationary process with spectral decomposition and covariance function respectively given by

\[
L_\tau(t) = \int_{-\infty}^{\infty} e^{i\lambda t} \Phi_\tau(\lambda) \, dZ(\lambda) \quad \text{and} \quad B_\tau(t) = \int_{-\infty}^{\infty} e^{i\lambda t} |\Phi_\tau(\lambda)|^2 \, dF(\lambda)
\]  

(22)

where \( \Phi_\tau \) is some function belonging to \( L^2(F)^- \). (see Definition 3.11 on page 16)

**Proof.** Note that \( H_X(t) \) coincides with the closed linear subspace spanned by \( \{X(t)\}_{s<t} \) due to mean square continuity of \( X \). \( L_\tau(t) \) belongs to \( H_X(t) \), thus it is the \( L^2(P) \) limit of finite linear combinations of the form \( \sum_{k=1}^{n(m)} c_k^{(m)} X(t - s_k^{(m)}) \) as \( m \to \infty \) for some sequence of positive integers \( n(m) \) such that, for all \( m \) and \( k = 1, \ldots, n(m), c_k^{(m)} \) are complex numbers and \( s_k^{(m)} > 0 \).

The spectral decomposition associated to such a linear combination for fixed \( m \) is given by

\[
\sum_{k=1}^{n(m)} c_k^{(m)} \int_{-\infty}^{\infty} e^{i(t-s_k^{(m)})\lambda} \, dZ(\lambda) = \int_{-\infty}^{\infty} \sum_{k=1}^{n(m)} c_k^{(m)} e^{it\lambda} e^{-is_k^{(m)}\lambda} \, dZ(\lambda) = \int_{-\infty}^{\infty} e^{it\lambda} \Phi_{\tau,m}(\lambda) \, dZ(\lambda),
\]

where \( \Phi_{\tau,m}(\lambda) = \sum_{k=1}^{n(m)} c_k^{(m)} e^{-is_k^{(m)}\lambda} \). By virtue of the isometry between \( L^2(F) \) and \( H_X \), see Theorem 2.12 on page 12, the latter converges in \( L^2(P) \) to \( L_\tau(t) \) if and only if \( \Phi_{\tau,m} \) converges in \( L^2(F) \) to some \( \Phi_\tau \) as \( m \to \infty \). In particular, \( \Phi \in L^2(F)^- \) and \( L_\tau(t) = \int e^{it\lambda} \Phi_\tau(\lambda) \, dZ(\lambda) \).

The spectral representation of the covariance function follows from the Itô isometry since.

**Definition 4.8.** The function \( \Phi_\tau \) in (22) shall be called the spectral characteristic for prediction with lag \( \tau \). For the sake of brevity, we shall refer to \( \Phi_\tau \) simply as the spectral characteristic as well for the remainder of this section.

### Properties of the Spectral Characteristic of the Best Linear Predictor

We now reformulate condition (21) on the best linear predictor, which we deduced using geometrical arguments, in terms of conditions on its spectral characteristic.
Proposition 4.9. A function $\Phi_\tau$ is the spectral characteristic for prediction with lag $\tau > 0$ if and only if $\Phi_\tau$ belongs to $L^2(F)$ and if $\Phi_\tau$ satisfies the orthogonality requirement, that is

$$\int_{-\infty}^{\infty} e^{is\lambda} (e^{i\tau\lambda} - \Phi_\tau(\lambda)) dF(\lambda) = 0 \quad \text{for all } s \geq 0 \quad (23)$$

Proof. The necessity and sufficiency part can be essentially treated at once. The requirement that $\Phi_\tau \in L^2(F)$ has been shown in Proposition 4.7 on the preceding page. Considering the orthogonality requirement, recall that the best linear predictor satisfies $E[(X(t + \tau) - L_\tau(t))X(t - s)] = 0$ for all $s \geq 0$ which we may write as

$$B(\tau + s) - E[L_\tau(t)X(t - s)] = 0.$$ 

Using the spectral representation (2) of $B$, we obtain

$$B(\tau + s) = \int_{-\infty}^{\infty} e^{i\tau\lambda} e^{is\lambda} dF(\lambda).$$

The second term in the difference above can be rewritten as

$$E[L_\tau(t)X(t - s)] = E \left[ \left( \int_{-\infty}^{\infty} e^{it\lambda} \Phi_\tau(\lambda) dZ(\lambda) \right) \left( \int_{-\infty}^{\infty} e^{i(t-s)\lambda} dZ(\lambda) \right) \right] = \int_{-\infty}^{\infty} e^{is\lambda} \Phi_\tau(\lambda) dF(\lambda)$$

by virtue of the Ito isometry (11). Now, combining both terms back together we obtain

$$E[(X(t + \tau) - L_\tau(t))X(t - s)] = \int_{-\infty}^{\infty} e^{is\lambda} (e^{i\tau\lambda} - \Phi_\tau(\lambda)) dF(\lambda)$$

which completes the proof.

Remark 4.10. Observe that the orthogonality requirement (23) reduces to

$$\int_{-\infty}^{\infty} e^{is\lambda} (e^{i\tau\lambda} - \Phi_\tau(\lambda)) f(\lambda) d\lambda = 0 \quad \text{for all } s \geq 0 \quad (24)$$

provided that the spectral density $f$ exists.
Prediction Error

Let us find an integral expression for the prediction error (19). Using the isometry between $H_X$ and $L^2(F)$ we arrive at the following identity:

$$\sigma^2_\tau = E \left| \int_{-\infty}^{\infty} e^{i(t+\tau)\lambda} - e^{it\lambda}\Phi_\tau(\lambda) d\lambda \right|^2 = \int_{-\infty}^{\infty} |e^{i\tau\lambda} - \Phi_\tau(\lambda)|^2 dF(\lambda).$$  \hspace{1cm} (25)

**Proposition 4.11.** The prediction error $\sigma^2_\tau$ (19) does not depend on $t$, and can be rewritten as

$$\sigma^2_\tau = \int_{-\infty}^{\infty} (1 - \Phi_\tau(\lambda)e^{-i\tau\lambda}) dF(\lambda) = \int_{-\infty}^{\infty} (1 - |\Phi_\tau(\lambda)|^2) dF(\lambda)$$  \hspace{1cm} (26)

**Proof.** Recall that the perpendicular $X(t + \tau) - L_\tau(t)$ is orthogonal to $L_\tau(t)$ since the latter belongs to $H_X(t)$, that is $E[(X(t + \tau) - L_\tau(t))\overline{L_\tau(t)}] = 0$. This is equivalent to $\int(e^{i\tau\lambda} - \Phi_\tau(\lambda))\overline{\Phi_\tau(\lambda)} dF(\lambda) = 0$ by the Ito isometry (11). Using this fact in (25), we obtain $\sigma^2_\tau = \int(e^{i\tau\lambda} - \Phi_\tau(\lambda))e^{-i\tau\lambda} dF(\lambda)$ which is equivalent to the first equality.

For the second equality, we again use the fact that $\int e^{i\tau\lambda}\overline{\Phi_\tau(\lambda)} dF(\lambda) = \int |\Phi_\tau(\lambda)|^2 dF(\lambda)$. Further, note that $\sigma^2_\tau$ is real, so $\sigma^2_\tau = \overline{\sigma^2_\tau}$. Therefore,

$$\sigma^2_\tau = \int_{-\infty}^{\infty} (1 - \Phi_\tau(\lambda)e^{-i\tau\lambda}) dF(\lambda) = \int_{-\infty}^{\infty} (1 - \overline{\Phi_\tau(\lambda)}e^{i\tau\lambda}) dF(\lambda) = \int_{-\infty}^{\infty} (1 - |\Phi_\tau(\lambda)|^2) dF(\lambda).$$

\[ \square \]

**Remark 4.12.** Observe that the filtering error as derived in (26) reduces to

$$\sigma^2_\tau = \int_{-\infty}^{\infty} (1 - \Phi_\tau(\lambda)e^{-i\tau\lambda}) f(\lambda) d\lambda = \int_{-\infty}^{\infty} (1 - |\Phi_\tau(\lambda)|^2) f(\lambda) d\lambda$$  \hspace{1cm} (27)

provided that the spectral density $f$ exists.

To this end, we shall regard the problem of prediction as solved if we can find the spectral characteristic for prediction.

### 4.2 Yaglom’s Solution

In this section, we will deal with the problem of finding the spectral characteristic $\Phi_\tau$ under mild restrictions on its behavior in the complex plane when understood as a complex function. This will considerably simplify its derivation by allowing to use tools from the theory of complex functions. This approach is due to Yaglom and has been discussed in his monograph [15], Chapter 6 Section 30.
In the following, we assume that a spectral density \( f \) exists. The following Theorem states Yaglom’s main result which can be used for finding \( \Phi_\tau \) in many cases of practical interest.

### 4.2.1 Sufficient Conditions on the Spectral Characteristic

**Theorem 4.13.** Let the spectral density \( f \) of a continuous stationary process \( X \) be bounded. Let \( \Phi_\tau \) be a complex function, and let another complex function \( \Psi_\tau \) be given by

\[
\Psi_\tau(z) = (e^{iz\tau} - \Phi_\tau(z))f(z).
\]

If the function \( \Phi_\tau \) satisfies the three conditions (P1) to (P3), see below, then \( \Phi_\tau \) restricted to the real line is the spectral characteristic for prediction with lag \( \tau > 0 \).

- **(P1)** The function \( \Phi_\tau \) is analytic in the lower half-plane, and grows no faster than some power of \( |z| \) in the lower half-plane. That is, there exists \( q > 0 \) such that \( |z|^{-q}\Phi_\tau(z) \to 0 \) as \( |z| \to \infty \) on \( \text{Im}(z) \leq 0 \).

- **(P2)** The function \( \Psi_\tau \) is analytic in the upper half-plane, and falls off faster than \( |z|^{-1} \) in the upper half-plane. That is, there exists \( \epsilon > 0 \) such that \( |z|^{1+\epsilon}\Psi_\tau(z) \to 0 \) as \( |z| \to \infty \) on \( \text{Im}(z) \geq 0 \).

- **(P3)** The function \( \Phi_\tau \) belongs to \( L^2(F) \), that is

\[
\int_{-\infty}^{\infty} |\Phi_\tau(\lambda)|^2 f(\lambda) d\lambda < \infty.
\]

As we shall see, (P2) implies that \( \Phi_\tau \) satisfies the orthogonality requirement (23) in Proposition 4.9 on page 22; (P1) and (P3) imply that \( \Phi_\tau \) is the \( L^2(F) \) limit of a sequence of functions that belong to \( L^2(F)^- \), and thus \( \Phi_\tau \) belongs to \( L^2(F)^- \) as well, just as required in Proposition 4.9 on page 22. Before giving a complete proof, however, we need to establish two auxiliary results which are due to Yaglom [15], Chapter 6 Section 30.

**Lemma 4.14.** Let \( \Psi \) be a complex function that is analytic in the upper half-plane. If there exists some \( K, \epsilon > 0 \) such that \( |\Psi(z)| < K|z|^{-1-\epsilon} \) for all sufficiently large \( |z| \) on \( \text{Im}(z) \geq 0 \), then

\[
\int_{-\infty}^{\infty} e^{is\lambda}\Psi(\lambda) d\lambda = 0 \quad \text{for all} \ s \geq 0.
\]

**Proof.** For \( r > 0 \), define \( L_r = \{ z \in \mathbb{C} | \text{Re}(z) \geq 0, |z| = r \} \) and let \( C_r = L_r \cup [-r, r] \). We understand \( C_r \) as a path moving in positive orientation along the upper semi-circle with radius \( r \) and its base, the real line segment from \( -r \) to \( +r \). Since both \( \Psi \) and the exponential function
are analytic in the upper half-plane, we obtain \( \int_{C_r} e^{isz}\Psi(z) \, dz = 0 \) by the residue Theorem. Consequently,

\[
\int_{-r}^{r} e^{is\lambda} \Psi(\lambda) \, d\lambda = \int_{L_r} e^{isz} \Psi(z) \, dz.
\] (29)

By hypothesis, \(|\Psi(z)| < K|z|^{-1-\epsilon}\) for some \(K, \epsilon > 0\) on \(L_r\) whenever \(r\) is large enough. Further, note that \(|e^{isz}| \leq 1\) for \(s \geq 0\) on \(Im(z) \geq 0\) since, then, \(|e^{isz}| = |e^{-sIm(z)}| \leq 1\). Thus,

\[
\lim_{r \to \infty} \left| \int_{L_r} e^{is\lambda} \Psi(z) \, dz \right| = \lim_{r \to \infty} \left| \int_{0}^\pi \exp(re^{i\varphi})\Psi(re^{i\varphi})ire^{i\varphi} \, d\varphi \right| \leq \lim_{r \to \infty} \int_{0}^\pi \frac{K}{r^{1+\epsilon}} \, d\varphi = 0,
\]

where we parameterized \(z = re^{i\varphi}\) on \(L_r\) with \(0 \leq \varphi \leq \pi\). Due to (29), we finally deduce

\[
\int_{-\infty}^{\infty} e^{is\lambda} \Psi(\lambda) \, d\lambda = \lim_{r \to \infty} \int_{-r}^{r} e^{is\lambda} \Psi(\lambda) \, d\lambda = 0.
\]

Observe that the integral on the left-hand side is a Lebesgue integral. This is because \(\Psi\) is analytic, hence continuous, and decays like \(|\lambda|^{-1-\epsilon}\) as \(|\lambda| \to \infty\) on the real line. Therefore the last expression indeed means a Lebesgue integral by virtue of the Dominated Convergence Theorem.

**Lemma 4.15.** Let \(\Phi\) be a function in \(L^2(F)\), where we assume that the density \(f\) of \(F\) exists and is bounded. For some fixed positive integer \(r\) define a sequence of functions \(\tilde{\Phi}_n\) as

\[
\tilde{\Phi}_n(\lambda) = \frac{\Phi(\lambda)}{(1 + \lambda/n)^r}.
\] (30)

Then, \(\tilde{\Phi}_n\) converges to \(\Phi\) in \(L^2(F)\).

**Proof.** We need to show \(\int |\Phi(\lambda) - \tilde{\Phi}_n(\lambda)|^2 f(\lambda) \, d\lambda \to 0\) as \(n \to \infty\). For any \(L > 0\),

\[
\int_{-\infty}^{\infty} |\Phi(\lambda) - \tilde{\Phi}_n(\lambda)|^2 f(\lambda) \, d\lambda = \int_{|\lambda| \leq L} |\Phi(\lambda) - \tilde{\Phi}_n(\lambda)|^2 f(\lambda) \, d\lambda + \int_{|\lambda| > L} |\Phi(\lambda) - \tilde{\Phi}_n(\lambda)|^2 f(\lambda) \, d\lambda
\]

\[
= \int_{|\lambda| \leq L} \left| 1 - \left(1 + \frac{i\lambda}{n} \right)^{-r} \right|^2 |\Phi(\lambda)|^2 f(\lambda) \, d\lambda + \int_{|\lambda| > L} |\Phi(\lambda) - \tilde{\Phi}_n(\lambda)|^2 f(\lambda) \, d\lambda.
\]

The first integral on the right-hand side goes to zero as \(n \to \infty\) for any \(L\) by virtue of the Bounded Convergence Theorem; this is apparent by noting that \(1 - (1 + \frac{i\lambda}{n})^{-r} \to 0\) as \(n \to \infty\) and using
integrability of $|\Phi(\lambda)|^2 f(\lambda)$. The second integral on the right-hand side is upper bounded by

$$
\int_{|\lambda| > L} |\Phi(\lambda) - \Phi_n(\lambda)|^2 f(\lambda) \, d\lambda \leq \int_{|\lambda| > L} (|\Phi(\lambda)| + |\Phi_n(\lambda)|)^2 f(\lambda) \, d\lambda
$$

$$
\leq 4 \int_{|\lambda| > L} |\Phi(\lambda)|^2 f(\lambda) \, d\lambda
$$

due to the inequality $|\Phi_n(\lambda)| \leq |\Phi(\lambda)|$ which holds because

$$
1 \leq 1 + \frac{\lambda^2}{n^2} = \left| 1 + \frac{i\lambda}{n} \right|^2.
$$

Consequently, $\lim_n \int |\Phi(\lambda) - \Phi_n(\lambda)|^2 f(\lambda) \, d\lambda \leq 4 \int_{|\lambda| > L} |\Phi(\lambda)|^2 f(\lambda) \, d\lambda$ for arbitrary $L > 0$. The latter goes to zero as $L \to \infty$. Thus, $\lim_n \int |\Phi(\lambda) - \Phi_n(\lambda)|^2 f(\lambda) \, d\lambda = 0$. \qed

**Proof of Theorem 4.13**

Now we have the necessary tools for proving Theorem 4.13 on page 24. Even though Yaglom provides an idea for proving this result in his monograph [15], Chapter 6 Section 30, he does not give a rigorous proof. We shall fill that gap in what follows.

**Proof.** The orthogonality requirement (23) follows from Lemma 4.15 on the preceding page applied to $\Psi_r$. It remains to show that $\Phi_r$ belongs to $L^2(F)^\perp$. By virtue of Corollary 3.13 on page 17, this is equivalent to showing that $\Phi_r$ is the spectral characteristic of a realizable filter with input $X$. This in turn is true whenever we can find a sequence of functions in $R$ – see definition 3.8 on page 16 – which converges in $L^2(F)$ to $\Phi_r$.

Let $r$ be an integer satisfying $r > q + 1 + \epsilon$, and define functions $\bar{\Phi}_{r,n}$ analogously to (30). Observe that $|z|^{1+\epsilon} \bar{\Phi}_{r,n}(z) \to 0$ as $|z| \to \infty$ on $\text{Im}(z) \leq 0$ because of the following: first, for large $|\lambda|$, $|z|^{1+\epsilon} \left| \frac{\Phi_r(z)}{(1 + \frac{i\lambda}{n})^r} \right| = \frac{|z|^{1+\epsilon} |\Phi_r(z)|}{|z|^r \left| \frac{1}{z} + \frac{i\lambda}{n} \right|^r} \leq |z|^{-q} |\Phi_r(z)|$, where the last inequality stems from $1 + \epsilon - r \leq -q - 1$ and the inequality (31); and second, $|z|^{-q} \Phi_r(z) \to 0$ as $|z| \to \infty$ on $\text{Im}(z) \leq 0$ by (P1).

Further, such a function $\bar{\Phi}_{r,n}$ is analytic in the lower half-plane because its denominator can only vanish at $z = in$ and its numerator is already analytic on $\text{Im}(z) \leq 0$ by hypothesis. In particular, $\bar{\Phi}_{r,n}$ is continuous on $\mathbb{R}$; in conjunction with its rate of decay (faster than $|\lambda|^{-1-\epsilon}$) we deduce that $\bar{\Phi}_{r,n} \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$. 

26
Let us now invoke Lemma 29 on page 25 in a slightly modified version: instead of requiring analyticity in the upper half-plane, we do so in the lower half-plane from which we deduce

\[ \int_{-\infty}^{\infty} e^{is\lambda} \tilde{\Phi}_{\tau,n}(\lambda) \, d\lambda = 0 \quad \text{for all } s \leq 0. \]

This is indeed a Lebesgue integral since \( \tilde{\Phi}_{\tau,n} \in L^1(\mathbb{R}) \). Now, define the Fourier coefficients

\[ a_n(s) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{is\lambda} \tilde{\Phi}_{\tau,n}(\lambda) \, d\lambda, \]

and note that \( a_n(s) = 0 \) for \( s \leq 0 \). Recall two results from the theory of Fourier transforms: since \( \tilde{\Phi}_{\tau,n} \in L^1(\mathbb{R}) \cap L^2(\mathbb{R}) \), we deduce continuity of \( a_n \) and \( a_n \in L^2(\mathbb{R}) \), as well as

\[ \lim_{T \to \infty} \int_{-\infty}^{\infty} \left| \tilde{\Phi}_{\tau,n}(\lambda) - G_{n,T}(\lambda) \right|^2 d\lambda = 0 \quad \text{where} \quad G_{n,T}(\lambda) = \int_{0}^{T} a_n(s)e^{-is\lambda} ds \]

by virtue of Plancherel’s Theorem. Note that convergence in \( L^2(\mathbb{R}) \) already implies convergence in \( L^2(F) \) due to boundedness of the spectral density \( f \). An application of Lemma 4.15 on page 25 shows that \( \tilde{\Phi}_{\tau,n} \) converges in \( L^2(F) \) to \( \Phi_\tau \).

Now, fix an arbitrary \( \epsilon > 0 \). Choose \( n \) large enough such that \( \int \left| \Phi_\tau - \tilde{\Phi}_{\tau,n} \right|^2 dF < \epsilon^2 \), and for that \( n \) choose \( T_n \) large enough such that \( \int \left| \tilde{\Phi}_{\tau,n} - G_{n,T_n} \right|^2 dF < \epsilon^2 \). Then,

\[ \left( \int_{-\infty}^{\infty} \left| \Phi_\tau(\lambda) - G_{n,T_n}(\lambda) \right|^2 dF(\lambda) \right)^{1/2} < 2\epsilon, \]

which shows that \( G_{n,T_n} \) converges in \( L^2(F) \) to \( \Phi_\tau \) as \( n \to \infty \). Now, since \( a_n \) is continuous, it assumes a maximum on any compact interval, and hence \( a_n(s) \) is integrable on \([0, T]\) for each finite \( T \); further, \( a_n \in L^2(\mathbb{R}) \) as argued earlier. This shows that \( \Phi_\tau \) is indeed the \( L^2(F) \) limit of functions in \( R \). Thus, \( \Phi_\tau \) belongs to \( L^2(F)^{-} \). \( \square \)

### 4.2.2 Examples

In the following, we shall see how Theorem 4.13 on page 24 can be applied in particular cases.

**Example 4.16.** ([15], pp153) Let \( X \) have a covariance function of the form \( B(t) = Ce^{-\alpha |t|} \), where \( C, \alpha > 0 \). Since the spectral density \( f \) is the Fourier transform of \( B \), we obtain

\[ f(\lambda) = \frac{C\alpha}{\pi} \frac{1}{(\lambda - i\alpha)(\lambda + i\alpha)} \quad \text{and} \quad \Psi_{\tau}(z) = \frac{C\alpha}{\pi} \frac{e^{iz\tau} - \Phi_\tau(z)}{(z - i\alpha)(z + i\alpha)}. \]
4.2 Yaglom’s Solution

In order for \( \Psi_\tau(z) \) to be analytic in the upper half-plane – recalling condition (P2) – we need \( \Phi_\tau(\alpha) = e^{-\alpha \tau} \) to cancel out the zero at \( z = i\alpha \) in the denominator. Note also that all remaining conditions are met by letting \( \Phi_\tau(z) \equiv e^{-\alpha \tau} \) to be constant: (P1) \( \Phi_\tau \) is trivially analytic, and \( |\Phi_\tau(z)| < |z| \) for large \( |z| \); (P2) \( |\Psi_\tau(z)| \) falls off faster than \( |z|^{-2} \) as \( |z| \) grows on \( \text{Im}(z) \geq 0 \) because \( |e^{i \tau z}| = e^{-\tau \text{Im}(z)} \leq 1 \) for \( \tau > 0 \) and \( \text{Im}(z) \geq 0 \); and (P3) \( \int |\Phi_\tau(\lambda)|^2 f(\lambda) d\lambda < \infty \).

Recalling the spectral decompositions (22) and (10) yields the best linear predictor

\[
L_\tau(t) = \int_{-\infty}^{\infty} e^{it\lambda} e^{-\alpha \tau} dZ(\lambda) = e^{-\alpha \tau} X(t).
\]

Remark 4.17. In this example, the best linear predictor only depends on the last observed value at time \( t \), whereas all knowledge about preceding times are of no use. Due to this negligence of the past, stationary processes with covariance functions of that form are called wide-sense Markov processes.

Example 4.18. ([15], pp155) Consider the spectral density \( f \) given by

\[
f(\lambda) = \frac{C}{\lambda^4 + \alpha^4}
\]

where \( C, \alpha > 0 \) are constants. It has the four roots \( \{\pm \beta, \pm \bar{\beta}\} \) where \( \beta = \frac{1+i}{\sqrt{2}} \alpha \). This leads to

\[
\Psi_\tau(z) = C \frac{e^{i\tau z} - \Phi_\tau(z)}{(z - \beta)(z + \beta)(z - \bar{\beta})(z + \bar{\beta})},
\]

which we wish to be analytic in the upper half-plane. Therefore, the numerator in the definition of \( \Psi_\tau \) should vanish at \( z = \beta \) and \( z = -\bar{\beta} \) to compensate for the respective zeros of the denominator. This yields the two conditions

\[
\Phi_\tau(\beta) = \exp\left(-\frac{\alpha \tau}{\sqrt{2}} (1 - i)\right) \quad \text{and} \quad \Phi_\tau(-\bar{\beta}) = \exp\left(-\frac{\alpha \tau}{\sqrt{2}} (1 + i)\right).
\]

Moreover, \( \Phi_\tau \) must be entire. To see this, assume \( \Phi_\tau \) was not entire, that is \( \Phi_\tau \) had a singularity in the upper half-plane since we already require it to be analytic in the lower half-plane by condition (P1). Note that this singularity cannot occur at \( z = \beta \) nor \( z = -\bar{\beta} \) due to (32). Therefore, \( \Psi_\tau \) would have the same singularity. This would violate (P2), however.

We also observe that all conditions concerning the asymptotic behavior of \( \Phi_\tau(z) \) and \( \Psi_\tau(\lambda) \) as \( |z| \to \infty \) in the lower and upper half-plane, respectively, are satisfied if we set \( \Phi_\tau \) to be linear in \( z \), that is \( \Phi_\tau(z) = Az + B \) for some constants \( A, B \). Further, condition (P3), that is \( \int |\Phi_\tau(\lambda)|^2 f(\lambda) d\lambda < \infty \), would also hold. The two values of \( \Phi_\tau(z) \) at \( z = \beta \) and \( z = -\bar{\beta} \) (32)
then uniquely determine \( \Phi_\tau \), and it follows
\[
\Phi_\tau(z) = a_\tau iz + b_\tau,
\]
where
\[
a_\tau = \frac{\sqrt{2}}{\alpha} \exp\left(-\frac{\alpha \tau}{\sqrt{2}}\right) \sin\left(\frac{\alpha \tau}{\sqrt{2}}\right)
\]
and
\[
b_\tau = \exp\left(-\frac{\alpha \tau}{\sqrt{2}}\right) \left( \cos\left(\frac{\alpha \tau}{\sqrt{2}}\right) + \sin\left(\frac{\alpha |\tau|}{\sqrt{2}}\right) \right).
\]
Thus, recalling the spectral decomposition of the derivative \( X'(t) \), the best linear predictor assumes the form
\[
L_\tau(t) = \int_{-\infty}^{\infty} e^{i\lambda}(a_\tau i\lambda + b_\tau)dZ(\lambda) = a_\tau X'(t) + b_\tau X(t).
\]

4.2.3 Spectral Characteristic for Rational Spectral Densities

Previously, we studied an approach for finding \( \Phi_\tau \) under mild restrictions concerning its behavior in the complex plane, which we illustrated in two examples. In the following, we will specify this approach under the assumption that the spectral density of \( X \) is rational. We begin with a characterization of rational spectral densities.

**Proposition 4.19.** Let \( f \) be the spectral density of a real-valued stationary process \( X \) with covariance function \( B \). Further, let \( f \) be rational, that is \( f(\lambda) = \frac{P(\lambda)}{Q(\lambda)} \) for some real polynomials \( P \) and \( Q \), and let \( f \) vanish nowhere on \( \mathbb{R} \). Then there exists a rational function \( \phi \) such that
\[
\phi(\lambda) = \frac{B(\lambda)}{A(\lambda)} \quad \text{and} \quad f(\lambda) = s|\phi(\lambda)|^2,
\]
where \( B \) and \( A \) are real polynomials with leading coefficient 1, and \( s > 0 \). Further, it holds that:

(i) If \( P \) and \( Q \) have no common roots, then \( \deg Q \) is even and \( \phi \) has exactly \( \frac{1}{2} \deg Q \) poles.

(ii) The function \( \phi \) can be chosen such that all its zeros have non-negative imaginary part, and all its poles have positive imaginary part. This choice uniquely defines \( \phi \).

(iii) It follows that \( \phi \) is Hermitian, that is \( \overline{\phi(\lambda)} = \phi(-\lambda) \) for all real \( \lambda \).

**Proof.** Let \( m = \deg P \) and \( n = \deg Q \). Without loss of generality we can assume that \( Q \) has leading coefficient 1. Let \( x \) denote real-valued variables.

First, we shall show that \( P \) and \( Q \) are unique. Assume \( f = \frac{P'}{Q'} \) for some polynomials \( P' \) and \( Q' \) of degree \( m \) and \( n \), respectively. Further, let \( P' \) and \( Q' \) have no common roots, and let \( Q' \) have leading coefficient 1. Note that \( PQ' = Q'P' = QP \). Let \( x_0 \) be a zero of \( Q \) and note that \( P(x_0)Q'(x_0) \neq 0 \). But \( x_0 \) is no zero of \( P \) by hypothesis, so \( Q'(x_0) = 0 \). Conversely, if \( y_0 \) is a zero of \( Q' \), we obtain \( Q(y_0)P'(y_0) = 0 \). But \( y_0 \) is no zero of \( P' \) by hypothesis, so \( Q(y_0) = 0 \). Hence,
4.2 Yaglom’s Solution

$Q(x) = Q'(x)$ for all $x$ since $Q$ and $Q'$ both have leading coefficient 1. Therefore, $P(x) = P'(x)$ for all $x$. This shows uniqueness of $Q$ and $P$.

Next, we shall show that $f$ is even. Note that $B$ is real-valued since $X$ is real-valued, and so $B(t) = B(-t)$ for all $t$. By Proposition 1.8 on page 7, we also have $B(t) = B(-t)$ for all $t$. So, $B$ is even, that is $B(t) = B(-t)$ for all $t$. Recall that $f$ is the Fourier inverse of $B$ by virtue of (5). Consequently, $f$ is even because

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} B(t)e^{-itx}dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} B(-t)e^{-itx}dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} B(t)e^{-it(-x)}dt = f(-x).$$

Next, let $A$ be the set of distinct zeros of $Q$. Since $f$ is continuous it cannot have poles on $\mathbb{R}$, so $Q(x) \neq 0$ for all $x$, and hence $\text{Im}(a_j) \neq 0$. Assume that $n$ is odd; then $Q$ would have at least one real zero, which is a contradiction. Thus, $n = \deg Q$ is even.

Further, $Q$ is even because of $f(x) = \frac{P(x)}{Q(x)} = \frac{P(-x)}{Q(-x)} = f(-x)$ and uniqueness of $Q$. Similarly, $Q(x) = \overline{Q(x)}$ for all $x$ because $f$ is real-valued and $\frac{P(x)}{Q(x)} = \frac{P(x)}{Q(x)}$. Hence, if $a$ is a zero of $Q$, then so are $-a$ and $\pm \bar{a}$. The exact same observation holds for $P$ as well. In particular, $n$ is a multiple of 4. Let $A_- = \{a \in A | \text{Im}(a) < 0\}$ and $A_+ = \{a \in A | \text{Im}(a) > 0\}$.

Then, we may decompose $Q = Q_+ Q_-$ where

$$Q_+(x) = i^{n/2} \prod_{a \in A_+} (x - a) \quad \text{and} \quad Q_-(x) = i^{n/2} \prod_{a \in A_-} (x - a).$$

Note that $\overline{Q_+(x)} = Q_-(x)$ because $n$ is a multiple of 4 and $A_- = \{\bar{a} | a \in A_+\}$. Therefore, $Q(x) = Q_+(x)\overline{Q_+(x)}$ for all $x \in \mathbb{R}$. Now, we will show that $Q_+(-x) = Q_-(x)$ for all $x \in \mathbb{R}$. Since $(i)^k(-1)^k = i^k$ for any integer $k$, we observe

$$Q_+(-x) = (i)^{n/2}(-1)^{n/2} \prod_{a \in A_+} (-x - a) = Q_-(x) = (i)^{n/2} \prod_{a \in A_+} (x + a).$$

Now, note that

$$\prod_{a \in A_+} (x + a) = \prod_{a \in A_+ \text{ Re}(a) = 0} (x + a) \prod_{a \in A_+ \text{ Re}(a) < 0} (x + a) \prod_{a \in A_+ \text{ Re}(a) > 0} (x + a),$$

$$= \prod_{a \in A_- \text{ Re}(a) = 0} (x - a) \prod_{a \in A_- \text{ Re}(a) < 0} (x - a) \prod_{a \in A_- \text{ Re}(a) > 0} (x - a)$$

$$= \prod_{a \in A_-} (x - a).$$
So, we obtain \( Q_+(-x) = Q_-(x) \) for \( x \in \mathbb{R} \). Thus, \( Q_+ \) is Hermitian on \( \mathbb{R} \).

Next, we will derive an analogous decomposition for \( P \). First, we will show that each real root of \( P \) must have even multiplicity. Let \( B \) denote the set of zeros of \( P \), and let \( B_R = \{ b \in \mathbb{R} | \text{Im}(b) = 0 \} \) and \( B_+ = \{ b \in \mathbb{R} | \text{Im}(b) > 0 \} \) and \( B_- = \{ b \in \mathbb{R} | \text{Im}(b) < 0 \} \). Let \( P \) be given by \( P(x) = P_0 \prod_{b \in B} (x - b) \) for some real constant \( P_0 \). Recall that \( Q(x) \neq 0 \) on \( \mathbb{R} \), and \( n \) is even. So, since the leading coefficient of \( Q \) is 1, \( P(x) \) must be positive for all \( x \in \mathbb{R} \); so, \( P_0 > 0 \), and \( \deg P_R \) is even. Let \( P_R \) denote the real part of \( P \), that is

\[
P_R(x) = P_0 \prod_{b \in B_R} (x - b).
\]

Now, let \( B_- \cup B_+ = \{ b_1, \ldots, b_k \} \) denote the imaginary roots of \( P \). By our earlier observation, we know that \( m \) is even; so, \( k = m - \deg P_R \) is even as well. We may then decompose \( P = P_+ P_- \) where

\[
P_+(x) = \sqrt{P_R(x)} e^{ib/2} \prod_{b \in B_+} (x - b) \quad \text{and} \quad P_-(x) = \sqrt{P_R(x)} e^{-ib/2} \prod_{b \in B_-} (x - b).
\]

By the same arguments as applied to \( Q \) we can conclude that \( \overline{P_+(x)} = P_-(x) = P_+(x) \) and \( P(x) = P_+ P_+ P_- \) for real \( x \).

Finally, let \( A(x) = Q_+(x) \) and \( B(x) = (\frac{1}{P_R})^{1/2} P_+(x) \) and \( s = P_0 \), and define \( \phi(x) = B(x) A(x)^{1/2} \). For real \( x \), we obtain

\[
s|\phi(x)|^2 = \left| \frac{P_+(x)}{Q_+(x)} \right|^2 = \frac{P_+(x) \overline{P_+(x)}}{Q_+(x) \overline{Q_+(x)}} = f(x).
\]

Properties (i), (ii), and (iii) have been shown in the proof.

In the following, let us focus on rational spectral densities, that is when \( f \) is of the form

\[
f(\lambda) = s \left| \frac{B(\lambda)}{A(\lambda)} \right|^2
\]

for some \( s > 0 \) and polynomials \( A, B \) such as in Proposition 4.19 on page 29. Factoring \( A \) yields

\[
A(\lambda) = (\lambda - \alpha_1)^{n_1} \cdots (\lambda - \alpha_N)^{n_N},
\]

say, where the \( \alpha_k \) are the distinct roots of \( A \) with multiplicity \( n_k \), respectively. Note that \( n_1 + \cdots + n_N = \deg A \), and \( \text{Im}(\alpha_k) > 0 \) by Proposition 4.19 on page 29.

**Notation 4.20.** Henceforth, by writing \( \tilde{P}(\lambda) \) for a polynomial \( P(\lambda) = (\lambda - \alpha_1) \cdots (\lambda - \alpha_n) \) we mean \( \tilde{P}(x) = (\lambda - \bar{\alpha}_1) \cdots (\lambda - \bar{\alpha}_n) \), even when we consider complex-valued arguments \( \lambda \).
Theorem 4.21. Let the spectral density $f$ of a stationary process be of the form (33) and (34). Let $L = \deg A - 1$. Further, let the complex coefficients $c_l$ of a complex polynomial $h_\tau(z) = c_L z^L + \ldots + c_1 z + c_0$ be chosen such that the complex function $G_\tau$ defined by

$$G_\tau(z) = e^{i\tau z} B(z) - h_\tau(z)$$

(35)

has the same zeros as the polynomial $A$ with no smaller multiplicities. Then the spectral characteristic for prediction with lag $\tau > 0$ is the restriction to the real line of the complex function $\Phi_\tau$ defined by

$$\Phi_\tau(z) = \frac{h_\tau(z)}{B(z)}.$$

(36)

The idea behind this technically appearing Theorem is simply to construct a function $\Phi_\tau$ that satisfies Theorem 4.13 on page 24 so that it will be the spectral characteristic for prediction. Its reasoning is due to Yaglom’s monograph [15], Chapter 6 Section 32. There, the author gradually constructs a function satisfying Theorem 4.13 on page 24. However, we shall follow the reverse approach: we show that $\Phi_\tau$ defined by (36) satisfies Theorem 4.13 on page 24.

Proof. In the following, let $z$ denote complex variables. We need to check that $\Phi_\tau$ satisfies the requirements (P1) to (P3) from Theorem 4.13 on page 24. First, $|\Phi_\tau(z)|$ grows like $|z|^{L - \deg B}$ as $|z| \to \infty$ anywhere on the complex plane. The only possible singularities of $\Phi_\tau$ are at zeros of $B$ which are all non-negative by Proposition 4.19 on page 29. So, $\Phi_\tau$ is analytic in the lower half-plane, hence (P1) is satisfied.

Now, consider the function $\Psi_\tau$ as defined in (28), which can be rewritten as follows,

$$\Psi_\tau(z) = \left( e^{i\tau z} - \frac{h_\tau(z)}{B(z)} \right) s \frac{B(z)\bar{B}(z)}{A(z)\bar{A}(z)} = s \frac{G_\tau(z)\bar{B}(z)}{A(z)\bar{A}(z)}.$$

Since every zero of $A$ is also a zero of $G_\tau$, accounting for multiplicity, we deduce that $\Psi_\tau$ is analytic in the upper half-plane. Let us check the growth behavior of $|\Psi_\tau(\lambda)|$ as $|\lambda| \to \infty$ in the lower half-plane. The triangle inequality yields

$$|\Psi_\tau(z)| \leq s \left( |e^{i\tau z}| \left| \frac{B(z)}{A(z)} \right| + \left| \frac{h_\tau(z)}{A(z)} \right| \right) \left| \frac{B(z)}{A(z)} \right|.$$

Since $\deg B < \deg A$, we observe that there exists some $K_1 > 0$ such that $\left| \frac{B(z)}{A(z)} \right| \leq K_1 |z|^{-1}$ and $\left| \frac{B(z)}{A(z)} \right| \leq K_1 |z|^{-1}$ for large $|z|$. Further, because $\deg h_\tau = \deg A - 1$, there exists some $K_2 > 0$ such that $\left| \frac{h_\tau(z)}{A(z)} \right| \leq K_2 |z|^{-1}$ for large $|z|$. Finally, note that $|e^{i\tau z}| = e^{-\tau \text{Im}(z)} \leq 1$ whenever
\(\tau > 0\) and \(\text{Im}(z) \geq 0\). Hence, \(|\Psi_\tau(z)| \leq K|z|^{-2}\) for some \(K > 0\) for large \(|z|\) in the upper half-plane, which shows (P2).

It remains to show square integrability of \(\Phi_\tau\) with respect to \(f(\lambda)d\lambda\). In the following, let \(\lambda\) be real. Observe that

\[
|\Phi_\tau(\lambda)|^2 f(\lambda) = s \left| \frac{h_\tau(\lambda)}{A(\lambda)} \right|^2.
\]

The ratio on the right-hand side decays like \(|\lambda|^{-2}\) as \(|\lambda| \to \infty\). Further, it is bounded because \(A\) has no real zeros. This shows (P3), which completes the proof.

**Remark 4.22.** The condition on \(h_\tau\) in the previous Theorem is certainly fulfilled if for each \(k \leq N\) the first \(n_k\) derivatives of \(G_\tau\) vanish at \(z = \alpha_k\). Since this amounts to a total of \(n = \text{deg} \ A\) conditions we arrive at the problem of solving a linear system of \(n\) equations with \(n\) unknowns, the latter of which being the coefficients of \(h_\tau\).
5 Filtering of Stationary Processes

Motivation  Consider a signal $X$, for example a telephone message, which we assume to be a stationary process. Due to the unavoidable presence of noise during its transmission, it is natural to assume that the actual signal is corrupted by some noise.

We no longer observe the pure signal $X$, but rather the sum

$$Y = X + N$$  \hfill (37)

where $N = \{N(t)\}$ is some additive noise term. Clearly, $Y$ is a stationary process as well. In the following, we shall consider the problem of reconstructing the process $X$ at time $t + \tau$ only using observations of $Y$ up to finite time $t$. Intuitively speaking, $t$ denotes the present and $\tau$ denotes the lag of filtering. Since we restrict ourselves to the study of linear estimators, this problem is usually referred to as causal linear filtering. We shall, however, oftentimes drop the attribute causal. Further, if $\tau \geq 0$ we shall refer to this problem as filtering in the future or filtering with prediction; if $\tau < 0$ we shall refer to this problem as filtering in the past or smoothing.

When $Y(s)$ has been observed for all $-\infty < s < \infty$ and we are interested in attaining a reasonable estimator for $X(t)$ for some $t$, then we call this problem non-causal linear filtering.

Assumption 5.1. We always assume that the signal $X$ and noise $N$ are mean square continuous stationary processes. Without loss of generality, we shall assume that the processes $X$ and $N$ have zero mean, and that $X$ and $N$ are uncorrelated, that is

$$E[X(t)N(s)] = 0 \quad \text{for all } t, s.$$  \hfill (38)

Notation 5.2. It should be pointed out that we are now dealing with three stationary processes, each of which induces a covariance function $B$, a spectral function $F$, possibly a spectral density $f$, and a stochastic spectral measure $Z$. In order to underline which process we are considering, we shall add appropriate subindices to the respective object at hand; for instance, $B_X$ shall denote the covariance function of the process $X$.

Note that if the stationary processes $X$ and $N$ assume spectral densities $f_X$ and $f_N$, then

$$f_Y = f_X + f_N$$  \hfill (39)

is the spectral density of $Y$ due to the correlation assumption (38).
5.1 Formulation of the Problem

Hilbert Space Formulation

In accordance with the notation used in the section about prediction, we let $H_Y(t)$ denote the closure of the linear subspace of $L^2(P)$ spanned by $\{Y(s)\}_{s \leq t}$. For the remainder of this section, we shall redefine the error $\sigma^2_{\tau}$.

**Definition 5.3.** Let $\sigma^2_{\tau}$ denote the smallest mean square error when approximating $X(t + \tau)$ from within $H_Y(t)$, that is

$$\sigma^2_{\tau} = \inf_{V \in H_Y(t)} E |X(t + \tau) - V|^2. \quad (40)$$

We shall also refer to $\sigma^2_{\tau}$ as the *filtering error*, or simply error. Again, $\sigma^2_{\tau}$ does not depend on $t$, see Proposition 5.9 on page 37.

Since we are dealing with Hilbert spaces, we observe that there exists a unique $L_{\tau}(t)$ in $H_Y(t)$ minimizing the filtering error, that is

$$\sigma^2_{\tau} = E |X(t + \tau) - L_{\tau}(t)|^2. \quad (41)$$

Note that the perpendicular $X(t + \tau) - L_{\tau}(t)$ is orthogonal to $H_Y(t)$; this uniquely defines $L_{\tau}(t)$ and is equivalent to saying that

$$E \left[ (X(t + \tau) - L_{\tau}(t)) Y(t - s) \right] = 0 \quad \text{for all } s \geq 0. \quad (42)$$

**Definition 5.4.** The stochastic process $L_{\tau} = \{L_{\tau}(t)\}$ comprised of the $L_{\tau}(t)$ in $H_Y(t)$ satisfying (42) is said to be the *best linear filter* of $X$ given $Y$ with lag $\tau$.

Spectral Decomposition of the Best Linear Estimator For Filtering

Just as for the problem of prediction, we may rewrite the best linear filter in its spectral decomposition. We state the following results without proof due to the similarity to their counterparts in the section about prediction.

**Proposition 5.5.** The best linear filter $L_{\tau}$ is a stationary process with spectral decomposition and covariance function respectively given by

$$L_{\tau}(t) = \int_{-\infty}^{\infty} e^{it\lambda} \Phi_{\tau}(\lambda) dZ_Y(\lambda) \quad \text{and} \quad B_{\tau}(t) = \int_{-\infty}^{\infty} e^{it\lambda} |\Phi_{\tau}(\lambda)|^2 dF_Y(\lambda) \quad (43)$$
where $\Phi_\tau$ is some function belonging to $L^2(F_Y)$. (see Definition 3.11 on page 16)

**Definition 5.6.** The function $\Phi_\tau$ in the spectral decomposition of the best linear filter (43) is called the *spectral characteristic* for filtering with lag $\tau$. For the remainder of this section, we shall also refer to $\Phi_\tau$ simply as the *spectral characteristic*.

**Properties of the Spectral Characteristic of the Best Linear Filter**

We now reformulate condition (42) on the best linear filter, which we deduced using geometrical arguments, in terms of conditions on its spectral characteristic.

**Proposition 5.7.** A function $\Phi_\tau$ is the spectral characteristic for filtering with lag $\tau$ if and only if $\Phi_\tau \in L^2(F_Y)$ and if it satisfies the orthogonality condition

$$
\int_{-\infty}^{\infty} e^{i(t+s)\lambda} dF_X(\lambda) - \int_{-\infty}^{\infty} e^{i\lambda} \Phi_\tau(\lambda) dF_Y(\lambda) = 0 \quad \text{for all} \quad s \geq 0.
$$

(44)

**Proof.** The necessity and sufficiency part can be essentially treated at once. The part that $\Phi_\tau \in L^2(F_Y)$ is due to Proposition 5.5 on the preceding page. Considering the orthogonality requirement, recall that the best linear estimator satisfies $E[(X(t+\tau) - L_\tau(t))Y(t-s)] = 0$ for all $s \geq 0$. Let us invoke the spectral decompositions (10) and (43) and the fact $Z_Y = Z_X + Z_N$ to rewrite

$$
X(t+\tau)Y(t-s) = \left( \int_{-\infty}^{\infty} e^{i(t+s)\lambda} dZ_X(\lambda) \right) \left( \int_{-\infty}^{\infty} e^{i(t-s)\lambda} dZ_Y(\lambda) \right)
$$

$$
= \left( \int_{-\infty}^{\infty} e^{i(t+s)\lambda} dZ_X(\lambda) \right) \left( \int_{-\infty}^{\infty} e^{i(t-s)\lambda} dZ_X(\lambda) + \int_{-\infty}^{\infty} e^{i(t-s)\lambda} dZ_N(\lambda) \right).
$$

Note that $H_X$ and $H_W$ are orthogonal to each other since $X$ and $N$ are uncorrelated (38). Now, this fact and an application of the Ito isometry (11) yields

$$
E[X(t+\tau)Y(t-s)] = \int_{-\infty}^{\infty} e^{i(t+s)\lambda} dF_X(\lambda).
$$

Similarly, we obtain

$$
E[L_\tau(t)Y(t-s)] = \int_{-\infty}^{\infty} e^{i\lambda} \Phi_\tau(\lambda) dF_Y(\lambda).
$$
This completes the proof since

\[ E \left[ (X(t + \tau) - L_{\tau}(t)) Y(t - s) \right] = \int_{-\infty}^{\infty} e^{i(\tau + s)\lambda} dF_X(\lambda) - \int_{-\infty}^{\infty} e^{i\lambda \Phi_{\tau}(\lambda)} dF_Y(\lambda). \]

\[ \Box \]

Remark 5.8. Observe that the orthogonality requirement (44) reduces to

\[ \int_{-\infty}^{\infty} e^{is\lambda} (e^{i\tau\lambda} f_X(\lambda) - \Phi_{\tau}(\lambda) f_Y(\lambda)) d\lambda = 0 \quad \text{for all} \quad s \geq 0 \quad (45) \]

provided that the respective spectral densities exist.

**Prediction Error**

Let us now give some rather tractable expressions for the filtering error (41).

**Proposition 5.9.** The filtering error \( \sigma_{\tau}^2 \) is independent of \( t \) and may be rewritten as

\[ \sigma_{\tau}^2 = \int_{-\infty}^{\infty} dF_X(\lambda) - \int_{-\infty}^{\infty} |\Phi_{\tau}(\lambda)|^2 dF_Y(\lambda) = \int_{-\infty}^{\infty} |e^{i\tau\lambda} - \Phi_{\tau}(\lambda)|^2 dF_X(\lambda) + \int_{-\infty}^{\infty} |\Phi_{\tau}(\lambda)|^2 dF_N(\lambda) \quad (46) \]

**Proof.** First, we shall show that the first equality holds. Recall that \( X(t + \tau) - L_{\tau}(t) \) is orthogonal to \( L_{\tau}(t) \) since the latter is an element of \( H_Y(t) \). Thus, we obtain

\[ \sigma_{\tau}^2 = E \left[ (X(t + \tau) - L_{\tau}(t)) (X(t + \tau) - L_{\tau}(t))^\ast \right] \]

\[ = E \left[ (X(t + \tau) - L_{\tau}(t)) X(t + \tau) \right] \]

\[ = E |X(t + \tau)|^2 - E \left[ L_{\tau}(t) (X(t + \tau) - L_{\tau}(t) + L_{\tau}(t)) \right] \]

\[ = E |X(t + \tau)|^2 - E |L_{\tau}(t)|^2. \]

Using the spectral decomposition, see (10) and (43), and applying the Ito isometry (11) we obtain the first equality. Next, we will show that the first and third term coincide. Note that \( H_X \) and \( H_N \) are orthogonal to one another and \( Z_Y = Z_X + Z_N \) because \( X \) and \( N \) are uncorrelated.
Using the spectral decomposition and isometry property once again, we obtain

\[ \sigma_\tau^2 = E \left| \int_{-\infty}^{\infty} e^{i(t+\tau)\lambda} \Phi_\tau(\lambda) dZ_X(\lambda) - \int_{-\infty}^{\infty} e^{it\lambda} \Phi_\tau(\lambda) dZ_Y(\lambda) \right|^2 \]

\[ = E \left| \int_{-\infty}^{\infty} e^{it\lambda} (e^{i\tau\lambda} - \Phi_\tau(\lambda)) dZ_X(\lambda) - \int_{-\infty}^{\infty} e^{it\lambda} \Phi_\tau(\lambda) dZ_N(\lambda) \right|^2 \]

\[ = E \int_{-\infty}^{\infty} |e^{it\lambda} - \Phi_\tau(\lambda)|^2 f_X(\lambda) d\lambda + \int_{-\infty}^{\infty} |\Phi_\tau(\lambda)|^2 f_N(\lambda) d\lambda. \tag{47} \]

\[ \hat{\sigma}_\tau^2 = \int_{-\infty}^{\infty} f_X(\lambda) - |\Phi_\tau(\lambda)|^2 f_Y(\lambda) d\lambda = \int_{-\infty}^{\infty} |e^{it\lambda} - \Phi_\tau(\lambda)|^2 f_X(\lambda) + |\Phi_\tau(\lambda)|^2 f_N(\lambda) d\lambda \tag{48} \]

provided that the respective spectral densities exist.

5.2 Solution for Non-Causal Filtering

Let us consider the non-causal filtering problem. Considering the filtering error (47), we make the following definition.

**Definition 5.11.** For spectral densities \( f_X, f_N \) and \( \Phi \in L^2(F_Y) \), let the non-causal filtering error \( e(f_X, f_N; \Phi) \) be defined as

\[ e(f_X, f_N; \Phi) = \int_{-\infty}^{\infty} \left| 1 - \Phi(\lambda) \right|^2 f_X(\lambda) d\lambda + |\Phi(\lambda)|^2 f_N(\lambda) d\lambda. \tag{49} \]

Note that if \( \Phi \in L^1(\mathbb{R}) \cap L^2(\mathbb{R}) \), then we may define its Fourier inverse transform \( \phi(s) = \frac{1}{2\pi} \int \Phi(\lambda) e^{is\lambda} d\lambda \). The filtering error is then the mean square error, that is \( e(f, g; \Phi) = E|\hat{X}(0) - \hat{X}(0)|^2 \), where \( \hat{X} \) is a filter with input \( Y \) given by

\[ \hat{X}(t) = \int_{-\infty}^{\infty} e^{it\lambda} \Phi(\lambda) dZ_Y(\lambda) = \int_{-\infty}^{\infty} \phi(t-s)Y(s) ds. \]

The last equality follows from Lemma 3.5 on page 15.

**Definition 5.12.** The minimizer \( \Phi^* \) of the non-causal filtering error, that is

\[ \Phi^* = \arg \min_{\Phi \in L^2(F_Y)} e(f_X, f_N; \Phi), \]
5.2 Solution for Non-Causal Filtering

is called spectral characteristic of the Wiener filter, or short Wiener filter, associated with \( f_X \) and \( f_N \). The resulting error, that is

\[
e^*(f_X, f_N) = e(f_X, f_N; \Phi^*),
\]

is called the Wiener filtering error associated with \( f_X \) and \( f_N \).

**Proposition 5.13.** For given \( f_X \) and \( f_N \), the Wiener filter is given by

\[
\Phi^*(\lambda) = \frac{f_X(\lambda)}{f_X(\lambda) + f_N(\lambda)}.
\]

The Wiener filtering error is

\[
e^*(f_X, f_N) = \int_{-\infty}^{\infty} \frac{f_X(\lambda)f_N(\lambda)}{f_X(\lambda) + f_N(\lambda)} d\lambda.
\]

**Proof.** We are looking for the best linear filter \( L_0(t) \in H_Y \) or equivalently its corresponding spectral characteristic \( \Phi^* \in L^2(F_Y) \). Since \( Y(s) \) has been observed for all \(-\infty < s < \infty\), the orthogonality condition (45) becomes

\[
\int_{-\infty}^{\infty} e^{is\lambda}(f_X(\lambda) - \Phi^*(\lambda)f_Y(\lambda)) d\lambda = 0 \quad \text{for } -\infty < s < \infty.
\]

This means the Fourier transform of \( f_X(\lambda) - \Phi^*(\lambda)f_Y(\lambda) \) vanishes almost everywhere. So,

\[
\Phi^*(\lambda) = \frac{f_X(\lambda)}{f_Y(\lambda)} = \frac{f_X(\lambda)}{f_X(\lambda) + f_N(\lambda)}
\]

is the spectral characteristic for filtering. The filtering error (46) is hence

\[
e^*(f_X, f_N) = \int_{-\infty}^{\infty} f_X(\lambda) - \left( \frac{f_X(\lambda)}{f_X(\lambda) + f_N(\lambda)} \right)^2 (f_X(\lambda) + f_N(\lambda)) d\lambda
\]

\[
= \int_{-\infty}^{\infty} \frac{f_X(\lambda)f_N(\lambda)}{f_X(\lambda) + f_N(\lambda)} d\lambda.
\]

\[\square\]

Considering (50), we observe that perfect reconstruction of the signal \( X(t) \), that is when the Wiener filtering error vanishes, is possible if and only if \( f_X \) an \( f_N \) have disjoint support.
5.2 Solution for Non-Causal Filtering

Alternative Approach

Let us derive the solution to the non-causal filtering problem using an alternative approach. Let \( \hat{X} = \{ \hat{X}(t) \} \), where \( \hat{X}(t) \in H_Y \) for all \( t \), be some stationary process with spectral decomposition

\[
\hat{X}(t) = \int_{-\infty}^{\infty} e^{it\lambda} K(\lambda) dZ_Y(\lambda) = \int_{-\infty}^{\infty} e^{it\lambda} K(\lambda) (dZ_X(\lambda) + dZ_N(\lambda))
\]

for some \( K \in L^2(F_Y) \). Note that such \( K \) always exists since \( \hat{X}(t) \in H_Y \). Consider the mean square error

\[
E|\hat{X}(t) - X(t)|^2 = E \left| \int_{-\infty}^{\infty} e^{it\lambda} (K(\lambda) - 1) dZ_X(\lambda) + \int_{-\infty}^{\infty} e^{it\lambda} K(\lambda) dZ_N(\lambda) \right|^2.
\]

Now, we use the fact that \( X \) and \( N \) are uncorrelated and the Ito isometry to deduce

\[
E|\hat{X}(t) - X(t)|^2 = E \left( \int_{-\infty}^{\infty} e^{it\lambda} (K(\lambda) - 1) dZ_X(\lambda) \right)^2 + E \left( \int_{-\infty}^{\infty} e^{it\lambda} K(\lambda) dZ_N(\lambda) \right)^2 = \int_{-\infty}^{\infty} |K(\lambda) - 1|^2 f_X(\lambda) + |K(\lambda)|^2 f_N(\lambda) d\lambda.
\]

Next, we want to find \( K \in L^2(F_Y) \) that minimizes the mean square error. Let us write \( K(\lambda) = K_1(\lambda) + iK_2(\lambda) \) for some real-valued functions \( K_1 \) and \( K_2 \). Then,

\[
E|\hat{X}(t) - X(t)|^2 = \int_{-\infty}^{\infty} ((K_1(\lambda) - 1)^2 + K_2(\lambda)^2) f_X(\lambda) + (K_1(\lambda)^2 + K_2(\lambda)^2) f_N(\lambda) d\lambda. \tag{53}
\]

Now, we need to minimize (53) over all real-valued \( K_1, K_2 \in L^2(F_Y) \). In particular, it suffices to minimize the integrand if its point-wise minimum exists at all. Certainly, \( K_2(\lambda) \equiv 0 \) is necessary. For fixed \( \lambda \), the minimum of \( (K_1(\lambda) - 1)^2 f_X(\lambda) + K_1(\lambda)^2 f_N(\lambda) \), which is quadratic in \( K_2(\lambda) \), satisfies

\[
2(K_1(\lambda) - 1)f_X(\lambda) + 2K_1(\lambda)f_N(\lambda) = 0.
\]

Therefore, \( K(\lambda) = K_1(\lambda) = \frac{f_X(\lambda)}{f_X(\lambda) + f_N(\lambda)} \) minimizes the mean square error, and is hence the spectral characteristic for filtering. This is just the same result as obtained earlier.
5.3 Yaglom’s Solution for Causal Filtering

In this section, similarly to the section about prediction, we will deal with the problem of finding the spectral characteristic for filtering under mild restrictions on its behavior in the complex plane. This approach is due to Yaglom, and it has been presented in his monograph [15], Chapter 7 Section 33.

We shall assume that the spectral densities \( f_X \) and \( f_N \) exist. Therefore, \( f_Y \) exists as well.

5.3.1 Sufficient Conditions on the Spectral Characteristic

To find the spectral characteristic \( \Phi_\tau \), all considerations from the section about prediction are still valid in its essence, except that this time \( \Psi_\tau \) is defined as

\[
\Psi_\tau(z) = e^{i\tau z} f_X(z) - \Phi_\tau(z) f_Y(z)
\]

bearing in mind the orthogonality requirement (45). In particular, we require that the spectral density \( f_Y \) of \( Y \) is bounded. The following Theorem yields sufficient conditions on the spectral characteristic for filtering. Its proof is analogous to Theorem 4.13 on page 24 in the previous section, and will thus be omitted here.

**Theorem 5.14.** Let the spectral density \( f_Y \) of the stationary process \( Y = X + N \) be bounded. Let \( \Phi_\tau \) be a complex function, and let \( \Psi_\tau \) be a complex function defined by

\[
\Psi_\tau(z) = e^{i\tau z} f_X(z) - \Phi_\tau(z) f_Y(z).
\]

If the function \( \Phi_\tau \) satisfies the three conditions (F1) to (F3), see below, then \( \Phi_\tau \) restricted to the real line is the spectral characteristic for filtering with lag \( \tau \).

(F1) The function \( \Phi_\tau \) is analytic in the lower half-plane, and grows no faster than some power of \( |z| \) in the lower half-plane. That is, there exists \( q > 0 \) such that \( |z|^{-q} \Phi_\tau(z) \to 0 \) as \( |z| \to \infty \) on \( \text{Im}(z) \leq 0 \).

(F2) The function \( \Psi_\tau \) is analytic in the upper half-plane, and falls off faster than \( |z|^{-1} \) in the upper half-plane. That is, there exists \( \epsilon > 0 \) such that \( |z|^{1+\epsilon} \Psi_\tau(z) \to 0 \) as \( |z| \to \infty \) on \( \text{Im}(z) \geq 0 \).

(F3) The function \( \Phi_\tau \) belongs to \( L^2(F_Y) \), that is

\[
\int_{-\infty}^{\infty} |\Phi_\tau(\lambda)|^2 f_Y(\lambda) \, d\lambda < \infty.
\]
5.3.2 Example

In the following, we will illustrate the solution to a particular problem of causal filtering utilizing Theorem 5.14 on the previous page.

Example 5.15. ([15], pp170) We will deal with two cases: (I) \( \tau > 0 \), and (II) \( \tau \leq 0 \). In (III) we shall show that the two solutions for \( \Phi_\tau \) that we obtain in (I) and (II) coincide at \( \tau = 0 \). We shall consider stationary processes \( X \) and \( Y \) with spectral densities \( f_X \) and \( f_Y \) given by

\[
f_X(\lambda) = \frac{C}{\lambda^2 + \alpha_X^2} \quad \text{and} \quad f_Y(\lambda) = D\frac{\lambda^2 + \beta^2}{(\lambda^2 + \alpha_X^2)(\lambda^2 + \alpha_N^2)},
\]

(55)

where \( C, D, \alpha_X, \alpha_N, \beta > 0 \) are constants. According to (F2) from Theorem 5.14 on the preceding page, we require that the function

\[
\Psi_\tau(z) = e^{i\tau z} f_X(z) - \Phi_\tau(z) f_Y(z) = \frac{\chi_\tau(z)}{(z^2 + \alpha_X^2)(z^2 + \alpha_N^2)},
\]

(56)

where \( \chi_\tau(z) = C e^{i\tau \lambda} (z^2 + \alpha_X^2) - D \Phi_\tau(z) (z^2 + \beta^2) \), is analytic in the upper half-plane. Hence, the numerator \( \chi_\tau(z) \) must vanish at \( z = i\alpha_X^2 \) and \( z = i\alpha_N^2 \) to compensate for the zeros of the denominator with positive imaginary part. This implies

\[
\Phi_\tau(i\alpha_X) = \frac{C(\alpha_N^2 - \alpha_X^2)}{D(\beta^2 - \alpha_X^2)} e^{-\tau \alpha_X} \quad \text{and} \quad \Phi_\tau(i\alpha_N) = 0.
\]

(57)

(I) Filtering in the future. First, we shall consider the case \( \tau \geq 0 \); the case \( \tau < 0 \) will be dealt with hereafter. According to Theorem 5.14 condition (F1), we need that \( \Phi_\tau \) is analytic in the lower half-plane. However, \( \Psi_\tau \) cannot have common singularities with \( \Phi_\tau \); if they did, the analyticity condition in either (F1) or (F2) would be violated. This leaves only one potential singularity for \( \Phi_\tau \), namely at \( z = i\beta \). Altogether, we thus obtain

\[
\Phi_\tau(z) = A_\tau(z) \frac{z - i\alpha_N}{z - i\beta}
\]

for some entire function \( A_\tau \). Due to the requirement that \( \Phi_\tau \) shall merely grow polynomially, we may let \( A_\tau \) be a polynomial. Further, condition (F3) requires that \( |A_\tau(\lambda)|^2 f_Y(\lambda) \) fall off faster than \( |\lambda|^{-1} \) as \( |\lambda| \to \infty \) on the real line. Given that \( A_\tau \) is a polynomial, this can only hold true
when \( A_\tau(z) \equiv A_\tau \) is constant. As we already know the value of \( \Phi_\tau(z) \) at \( z = i\alpha_X \), we conclude
\[
A_\tau = \Phi_\tau(i\alpha_X) \frac{i\alpha_X - i\beta}{i\alpha_X - i\alpha_N} = -\frac{C(\alpha_N + \alpha_X)}{D(\beta + \alpha_X)} e^{-\tau \alpha_X}.
\]

Finally, we arrive at the following candidate for the spectral characteristic
\[
\Phi_\tau(z) = e^{-\tau \alpha_X} \frac{C(\alpha_N + \alpha_X)(\alpha_N + iz)}{D(\beta + \alpha_X)(\beta + iz)}.
\]

It remains to check the growth condition on \( \Psi_\tau \) in (F2). Let \( z = a + ib \) with \( b \geq 0 \), then
\[
|e^{i\tau z}| = e^{-\tau b}.
\]
This shows that if \( \tau < 0 \) then \( |\Psi_\tau(z)| \) grows exponentially; hence, (F2) cannot be satisfied with the preceding construction when \( \tau < 0 \). On the other hand, if \( \tau \geq 0 \) then \( |\Psi_\tau(z)| \) grows (or rather falls off) like \( |z|^{-2} \); hence, (F2) will be satisfied in that case. To conclude, the spectral characteristic for filtering in the future \( \Phi_\tau^{(F)} \) is given by
\[
\Phi_\tau^{(F)}(\lambda) = e^{-\tau \alpha_X} \frac{C(\alpha_N + \alpha_X)(\alpha_N + i\lambda)}{D(\beta + \alpha_X)(\beta + i\lambda)}. \tag{58}
\]

(II) Filtering in the past. Next, let us consider the case \( \tau < 0 \). The function \( \chi_\tau \) in (56) must be analytic in the upper half-plane as \( \Psi_\tau \) has to be so as well. Re-arranging \( \chi_\tau(z) \) we see that
\[
\Phi_\tau(z) = \frac{Ce^{i\tau z}(z^2 + \alpha_N^2) - \chi_\tau(z)}{D(z^2 + \beta^2)},
\]
and observe that \( \chi_\tau \) must be analytic in the lower half-plane as required by (F1). Recalling (56) and analyticity of \( \Psi_\tau \) in the upper half-plane, we deduce that \( \chi_\tau \) is an entire function. We noted earlier that \( \chi_\tau \) has zeros at \( i\alpha_X \) and \( i\alpha_W \), therefore
\[
\chi_\tau(z) = B_\tau(z)(z - i\alpha_X)(z - i\alpha_W)
\]
for some entire function \( B_\tau \). For what follows, it suffices that \( B_\tau \) is a constant polynomial. Given
\[
\chi_\tau(-i\beta) = Ce^{-\tau \beta}(\alpha_N^2 - \beta^2),
\]
we thus deduce
\[
\chi_\tau(z) = e^{\beta \tau}(\alpha_N - \beta)(z - i\alpha_X)(z - i\alpha_N).
\]

We observe that for large \( |\lambda| \) on the real line the integrand in condition (F3) of Theorem (5.14) behaves like
\[
|\Phi_\tau(\lambda)|^2 f_Y(\lambda) \approx \frac{K}{\lambda^2}
\]
for large $|\lambda|$ on the real line, where $K$ is some constant. Thus, (F3) is satisfied. We finally obtain that the spectral characteristic for filtering in the past $\Phi^{(P)}_\tau$ is given by

$$\Phi^{(P)}_\tau(\lambda) = \frac{C}{D} \frac{(\alpha_X + \beta)(\lambda - i\alpha_N)(\lambda - i\alpha_X)(\lambda - i\alpha_N)}{(\alpha_X + \beta)(\lambda + i\beta)(\lambda - i\beta)} e^{i\tau \lambda} + \frac{(\alpha_N - \beta)(\lambda - i\alpha_X)(\lambda - i\alpha_N) e^{i\beta \tau}}{(\alpha_X + \beta)(\lambda + \beta^2)}.$$

(III) Filtering without lag. As it turns out, the two formulae (58) and (59) coincide at $\tau = 0$:

$$\Phi^{(P)}_0(\lambda) = \frac{C}{D} \frac{(\alpha_X + \beta)(\lambda + i\alpha_N)(\lambda - i\alpha_N)(\lambda - i\alpha_X)(\lambda - i\alpha_N)}{(\alpha_X + \beta)(\lambda + i\beta)(\lambda - i\beta)} e^{i\lambda} + \frac{(\alpha_N - \beta)(\lambda + i\alpha_N)(\lambda - i\alpha_X)(\lambda - i\alpha_N) e^{i\beta \lambda}}{(\alpha_X + \beta)(\lambda + \beta^2)}$$

(IV) Spectral Characteristic. The spectral characteristic for filtering $\Phi_\tau$ is thus given by

$$\Phi_\tau(\lambda) = \begin{cases} \frac{C}{D} \frac{(\alpha_X + \beta)(\lambda - i\alpha_N)(\lambda - i\alpha_X)(\lambda - i\alpha_N)}{(\alpha_X + \beta)(\lambda + i\beta)(\lambda - i\beta)} e^{i\tau \lambda} + \frac{(\alpha_N - \beta)(\lambda - i\alpha_X)(\lambda - i\alpha_N) e^{i\beta \tau}}{(\alpha_X + \beta)(\lambda + \beta^2)} & \text{if } \tau \geq 0 \\ e^{-\tau \alpha_X} \frac{C}{D} \frac{(\alpha_N + \alpha_X)(\alpha_N + i\lambda)}{(\alpha_X + \beta)(\beta + i\lambda)} & \text{if } \tau \leq 0 \end{cases}$$

This construction is well-defined. Due to the rather complicated form of $\Phi_\tau$ there is little hope for finding a closed form expression for the resulting best linear filter $L_\tau(t)$.

5.3.3 Spectral Characteristic for Rational Spectral Densities

Let us now give a precise way for finding the spectral characteristic for filtering $\Phi_\tau$ in the case of rational spectral densities. We shall denote the densities $f_Y$ and $f_X$ by

$$f_Y(\lambda) = s_Y \left| \frac{B(\lambda)}{A(\lambda)} \right|^2 \quad \text{and} \quad f_X(\lambda) = s_X \left| \frac{D(\lambda)}{C(\lambda)} \right|^2$$

where $A$, $B$, $C$, and $D$ are polynomials with leading coefficients 1, and $s_Y$, $s_X > 0$.

By our general observation about rational spectral densities, see Proposition 4.19 on page 29, we know that $\deg B < \deg A$ and $\deg D < \deg C$. Furthermore, we recall that zeros of $B$ and
5.3 Yaglom’s Solution for Causal Filtering

$D$ have non-negative imaginary parts, and zeros of $A$ and $C$ have positive imaginary parts. In addition, we shall assume that the spectral density $f_Y$ vanishes nowhere on the real line, that is $B$ has no real zeros.

For the sake of convenience, we shall denote the distinct roots of $B$ and $C$ as $\beta_k$ and $\gamma_j$ occurring with multiplicities $m_k$ and $n_j$, respectively. Factoring $B$ and $C$ then yields

$$B(\lambda) = (\lambda - \beta_1)^{m_1} \cdots (\lambda - \beta_M)^{m_M} \quad \text{and} \quad C(\lambda) = (\lambda - \gamma_1)^{n_1} \cdots (\lambda - \gamma_N)^{n_N}. \quad (62)$$

It should be pointed out that $m_1 + \ldots + m_M = \deg B$ and $n_1 + \ldots + n_N = \deg C$.

As it turns out, the construction of the solution, that is the derivation of the spectral characteristic $\Phi_\tau$, significantly depends on whether the lag $\tau$ is positive or negative. We shall present the solutions in Theorem 5.16 and in Theorem 5.18 on page 47.

The idea behind these solutions is simply to construct a function $\Phi_\tau$ that satisfies Theorem 5.14 on page 41 so that it will be the spectral characteristic for filtering. Its reasoning is due to Yaglom’s monograph [15] on pp176 for the case $\tau \geq 0$ and on pp179 for the case $\tau < 0$.

Spectral Characteristic for Filtering in the Future

At this point, we would like to remind the reader of the notation $\tilde{P}(\lambda)$: for a polynomial $P(\lambda) = (\lambda - a_1) \cdots (\lambda - a_n)$, we let $\tilde{P}(\lambda) = (\lambda - \tilde{a}_1) \cdots (\lambda - \tilde{a}_n)$, even for complex-valued $\lambda$.

**Theorem 5.16.** Consider the problem of filtering with lag $\tau \geq 0$. Let $L = \deg C - 1$. Let the coefficients $c_l$ of a complex polynomial $h_\tau(z) = c_L z^L + \ldots + c_1 z + c_0$ be chosen such that the function

$$G_\tau(z) = e^{i\tau z} D(z) \tilde{D}(z) \tilde{A}(z) - h_\tau(z) \tilde{B}(z) \tilde{C}(z) \quad (63)$$

has the same zeros as $C$ with no smaller multiplicities. Then, the spectral characteristic for filtering with lag $\tau \geq 0$ is the restriction to the real line of the function $\Phi_\tau$ defined by

$$\Phi_\tau(z) = \frac{A(z) h_\tau(z)}{B(z)^{-1} C(z)}. \quad (64)$$

**Proof.** We need to prove the conditions (F1) to (F3) from Theorem 5.14 on page 41. In the following, let $z$ denote complex values. First, $\Phi_\tau$ clearly grows merely polynomially. Also, $\Phi_\tau$ has no poles other than zeros of $B$ and $C$ which lie in the upper half-plane, hence it is analytic in the lower half-plane. Thus, (F1) is satisfied.
Second, consider the function $\Psi_\tau$ as defined in (54). We can rewrite it as follows

$$\Psi_\tau(z) = \frac{s_X e^{i\tau z} D(z) \tilde{D}(z) A(z) \hat{A}(z) - s_Y \Phi_\tau(z) B(z) \tilde{B}(z) C(z) \hat{C}(z)}{C(z) \tilde{C}(z) A(z) A(z)}$$

$$= s_X \frac{e^{i\tau z} D(z) \tilde{D}(z) \hat{A}(z) - h_\tau(z) \tilde{B}(z) \hat{C}(z)}{C(z) \tilde{C}(z) A(z)}$$

$$= s_X \frac{G_\tau(z)}{C(z) \tilde{C}(z) A(z)}.$$

Hence, if every zero of $C$ is also a zero of $G_\tau$, counting multiplicity, then it follows that $\Psi_\tau$ has all its poles in the lower half-plane, specifically at zeros of $\tilde{C}$ and $\hat{A}$. From the above formula we can also bound the growth of $|\Psi_\tau(z)|$ for large $|z|$ with $\text{Im}(z) \geq 0$. Note that

$$|\Psi(z)| \leq s_X \left| \frac{D(z) \tilde{D}(z)}{C(z) \tilde{C}(z)} \right| + s_X \left| \frac{h_\tau(z) \tilde{B}(z)}{C(z) A(z)} \right|$$

because $|e^{i\tau z}| \leq 1$ for $\tau \geq 0$ and $\text{Im}(z) \geq 0$. The right-hand side of the above expression is a rational function with highest power of

$$\max \{2(\deg D - \deg C), \deg B - \deg A - 1 \} \leq -2.$$

Hence, we obtain $|\Psi_\tau(z)| \leq |z|^{-1-\varepsilon}$ for sufficiently large $|z|$, and $0 < \varepsilon < 1$. Thus, (F2) is satisfied.

Third, consider the growth of $|\Phi_\tau(\lambda)|^2 f_Y(\lambda)$ for real $\lambda$. We obtain

$$|\Phi_\tau(\lambda)|^2 f_Y(\lambda) = s_X \left| \frac{h_\tau(\lambda)}{C(\lambda)} \right|^2.$$

The right-hand side is uniformly bounded, say by $\delta$, because $C$ is polynomial with no real zeros, and $\deg C > \deg h_\tau$. Further, for large $|\lambda|$ on the real line, say for $|\lambda| > R$, there is a constant $M > 0$ such that $\frac{|h_\tau(\lambda)|}{|C(\lambda)|} \leq M |\lambda|^{-1}$ since $\deg C = \deg h_\tau + 1$. Now, splitting the integral of $|\Phi_\tau(\lambda)|^2 f_Y(\lambda)$ over the real line into the integral over $[-R, R]$ and $[-R, R]^C$ we obtain

$$\int_{-\infty}^{\infty} |\Phi_\tau(\lambda)|^2 f_Y(\lambda) \, d\lambda \leq 2R\delta + \int_{|\lambda| > R} M |\lambda|^{-1} \, d\lambda < \infty.$$

This proves (F3). Thus, under the imposed condition on $G_\tau$, we conclude that (64) is indeed the spectral characteristic for filtering in the future.

Remark 5.17. We point out that the condition on $G_\tau$ in the above Theorem is just to say that for each zero $\gamma_k$ of $C$, which appears with multiplicity $n_k$, the first $n_k$ derivatives of $G_\tau$ must
vanish at $z = \gamma_k$. This imposes a total of $n_1 + \ldots + n_N = L + 1$ conditions, and hence uniquely defines the $L + 1$ coefficients of $h_\tau$. This in turn uniquely defines $\Phi_\tau$. To conclude, we are left with solving a system of $L + 1$ linear equations in $L + 1$ unknowns in order to solve the problem of filtering in the future.

Spectral Characteristic for Filtering in the Past

**Theorem 5.18.** Consider the problem of filtering with lag $\tau < 0$. Let $L = \deg B + \deg C - 1$. Let the coefficients $c_i$ of a complex polynomial $h_\tau(z) = c_L z^L + \ldots + c_1 z + c_0$ be chosen such that the function

$$G_\tau(z) = s_X e^{i \tau z} \bar{D}(z) A(z) \bar{A}(z) - h_\tau(z) A(z) C(z)$$  \hspace{1cm} (65)

has the same zeros as $\bar{B}$ and $\bar{C}$ with no smaller multiplicities. Then, the spectral characteristic for filtering with lag $\tau < 0$ is the restriction to the real line of the function $\Phi_\tau$ given by

$$\Phi_\tau(z) = \frac{G_\tau(z)}{s_Y B(z) B(z) C(z) \bar{C}(z)}.$$  \hspace{1cm} (66)

**Proof.** We need to prove the conditions (F1) to (F3) from Theorem 5.14 on page 41. In the following, let $z$ denote complex values. First, on $Im(z) \leq 0$, we may bound $\Phi_\tau$ as follows

$$|\Phi_\tau(z)| \leq \frac{s_X |D(z) \bar{D}(z) A(z) \bar{A}(z)| + |h_\tau(z) A(z) C(z)|}{s_Y |B(z) B(z) C(z) \bar{C}(z)|}$$

because $|e^{i \tau z}| \leq 1$ for $\tau < 0$ and $Im(z) \leq 0$. The right-hand side is a rational function. This shows that $|\Phi_\tau(z)|$ grows no faster than some power of $|z|$ in the lower half-plane. Also, by hypothesis we assume that $G_\tau$ vanishes at zeros of $\bar{B}$ and $\bar{C}$ with sufficient multiplicity, hence $\Phi_\tau$ has no poles in the lower half-plane either. Thus, (F1) is satisfied.

Second, consider the function $\Psi_\tau$ as defined in (54), and rewrite it as follows

$$\Psi_\tau(\lambda) = \frac{s_X e^{i \tau z} D(z) \bar{D}(z) A(z) \bar{A}(z) - s_Y \Phi_\tau(z) B(z) \bar{B}(z) C(z) \bar{C}(z)}{A(z) A(z) C(z) \bar{C}(z)}$$

$$= \frac{s_X e^{i \tau z} D(z) \bar{D}(z) A(z) \bar{A}(z) - G_\tau(z)}{A(z) A(z) C(z) \bar{C}(z)}$$

$$= \frac{h_\tau(z)}{A(z) C(z)}.$$  

Note that $\Psi_\tau$ is a rational function. Counting the degrees of the involved polynomials yields $L - \deg C - \deg A \leq -2$. Hence, there exists some finite $K$ such that $|\Psi_\tau(z)| \leq K |z|^{-2}$ for sufficiently large $|z|$. Also, $\Psi_\tau$ is analytic in the upper half-plane. Thus, (F2) is satisfied.
Third, consider the growth of \( |\Phi_\tau(\lambda)|^2 f_Y(\lambda) \) for real \( \lambda \). We obtain

\[
|\Phi_\tau(\lambda)|^2 f_Y(\lambda) = \frac{s_X^2}{s_Y} \left| \frac{\bar{A}(\lambda)|D(\lambda)|^2 - h_\tau(\lambda) C(\lambda)}{B(\lambda)|C(\lambda)|^2} \right|^2 \leq \frac{s_X^2}{s_Y} \left( \left| \frac{A(\lambda)}{B(\lambda)} \right| \left| \frac{D(\lambda)}{C(\lambda)} \right|^2 + \left| \frac{h_\tau(\lambda)}{B(\lambda)C(\lambda)} \right| \right)^2
\]

The right-hand side grows like a rational function with highest power no greater than \(-2\) as \( |\lambda| \to \infty \) because

\[
(deg A - deg B) + 2(deg D - deg C) \leq -1 \quad \text{and} \quad deg h_\tau - (deg B + deg C) = -1.
\]

Also, the right-hand side is uniformly bounded in \( \lambda \) because \( B \) and \( C \) have no real zeros. Hence, the right-hand side is integrable. This shows (F3), and completes the proof.

**Remark 5.19.** Again, the condition on \( G_\tau \) in the above Theorem is to say that for each zero \( \bar{\beta}_j \) of \( \bar{B} \) and each zero \( \bar{\gamma}_k \) of \( \bar{C} \), which appear with multiplicity \( m_j \) and \( n_k \) respectively, the first \( m_j \) and \( n_k \) derivatives of \( G_\tau \) must vanish at \( z = \bar{\beta}_j \) and \( z = \bar{\gamma}_k \) respectively. This imposes a total of \( m_1 + \ldots + m_M + n_1 + \ldots + n_N = L + 1 \) conditions, and hence uniquely defines the \( L + 1 \) coefficients of \( h_\tau \). This uniquely defines \( \Phi_\tau \), and we are left with solving a system of \( L + 1 \) linear equations in \( L + 1 \) unknowns in order to solve the problem of filtering in the past.
5.4 Filtering of Stationary Sequences

Thus far, we have only considered the problems of prediction and filtering of stationary processes. However, one may as well consider the analogue for stationary sequences. At this point, note that all definitions and results from Section 5.1 essentially carry over; the only differences are that now \( \{ Y(t-1), Y(t-2), \ldots \} \) has been observed, and that the bounds of integration are now \(-\pi\) and \(\pi\) rather than \(-\infty\) and \(\infty\). Moreover, since prediction can be treated as a special case of filtering, we shall focus on the latter.

5.4.1 Yaglom’s Solution

The problem of finding the spectral characteristic for filtering with integer-valued lag \( \tau \) of a stationary sequence, which we again denote by \( \Phi_\tau \), is quite complicated in general; a complete solution is due to Kolmogorov [11]. Instead, we shall focus our attention on the case when \( \Phi_\tau \) is the uniform limit (in the ordinary sense) of a sequence of functions \( \Phi_{\tau,n} \) of the form

\[
\Phi_{\tau,n}(\lambda) = \sum_{k=1}^{n} c_k e^{-ik\lambda}.
\]

It is clear that \( \Phi_{\tau,n} \) then converges in \( L^2(F_Y) \) to \( \Phi_\tau \) given by \( \Phi_\tau(\lambda) = \sum_{k=1}^{\infty} c_k e^{-ik\lambda} \), which itself is equivalent to saying that the random variable

\[
\int_{-\pi}^{\pi} e^{it\lambda} \Phi_{\tau,n}(\lambda) dZ_Y(\lambda) = \sum_{k=1}^{n} c_k Y(t-k)
\]

converges in \( L^2(P) \) to

\[
\int_{-\pi}^{\pi} e^{it\lambda} \Phi_\tau(\lambda) dZ_Y(\lambda) = \sum_{k=1}^{\infty} c_k Y(t-k).
\]

Rational Spectral Densities

First, let us state a result that is analogous to Proposition 4.19 on page 29. A derivation of this result can be found in Yaglom [15], Chapter 4 Section 23 on pp121.

**Proposition 5.20.** Let \( f \) be the spectral density of a stationary sequence. If \( f \) is rational, then it can be represented in the form

\[
f(\lambda) = s \left| \frac{B(e^{i\lambda})}{A(e^{i\lambda})} \right|^2 = s \frac{(e^{i\lambda} - b_1) \cdots (e^{i\lambda} - b_M)}{(e^{i\lambda} - a_1) \cdots (e^{i\lambda} - a_N)}^2,
\]

(69)
where $B$ and $A$ are polynomials satisfying $M < N$ with leading coefficients 1, and $s \geq 0$. Further, all zeros of $A$ have absolute values strictly less than 1, while all zeros of $B$ have absolute values which do not exceed 1.

For the remainder of this section, we shall assume that the spectral densities $f_Y$ and $f_X$ associated with the stationary sequences $Y$ and $X$, respectively, satisfy Proposition 5.20 on the preceding page. In particular, let $f_X$ and $f_Y$ be given by

$$f_Y(\lambda) = s_Y \left| \frac{B(e^{i\lambda})}{A(e^{i\lambda})} \right|^2 \quad \text{and} \quad f_X(\lambda) = s_X \left| \frac{D(e^{i\lambda})}{C(e^{i\lambda})} \right|^2. \quad (70)$$

Also, we shall assume that $f_Y$ vanishes nowhere on the real line. Therefore, all zeros of $B$ have absolute values strictly less than 1.

Notation 5.21. Let $P(z) = (z - a_1) \cdots (z - a_n)$ be a complex polynomial. We define the complex function $\tilde{P}$ as $\tilde{P}(z) = (1 - \bar{a}_1 z) \cdots (1 - \bar{a}_n z)$. Note that $\tilde{P}(z) = z^{-n} (1 - \bar{a}_1 z) \cdots (1 - \bar{a}_n z)$.

Considering $f_Y$ and $f_X$, let the complex counterparts $f_Y^*$ and $f_X^*$ be defined by

$$f_Y^*(z) = s_Y \frac{B(z) \tilde{B}(z)}{A(z)A(z)} \quad \text{and} \quad f_X^*(z) = s_X \frac{D(z) \tilde{D}(z)}{C(z)C(z)}. \quad (71)$$

Yaglom’s Main Result

The following main result, which we state without proof, is due to Yaglom [15], Chapter 5 Section 26 on pp129.

Theorem 5.22. Let the spectral densities $f_Y$ and $f_X$ exist and be rational. Further, let the best linear filter $L_\tau$ assume a series expansion such as on the right-hand side in (68). Let $\Phi_\tau^*$ be defined by some power series $\Phi_\tau^*(z) = \sum_{k \geq 1} c_k z^{-k}$, and similarly let $\Psi_\tau^*$ be defined by $\Psi_\tau^*(z) = z^\tau f_X^*(z) - \Phi_\tau^*(z) f_Y^*(z)$.

If $\Phi_\tau^*$ satisfies the three conditions (F1*) to (F3*), see below, then the function $\Phi_\tau$ defined by $\Phi_\tau(\lambda) = \Phi_\tau^*(e^{i\lambda})$ is the spectral characteristic for filtering with integer-valued lag $\tau$. In particular,

$$L_\tau(t) = \sum_{k=1}^{\infty} c_k Y(t - k).$$

(F1*) $\Phi_\tau^*$ is analytic outside and on the boundary of the unit circle, that is on $|z| \geq 1$,

(F2*) $\Phi_\tau^*(z) \to 0$ as $z \to \infty$,

(F3*) $\Psi_\tau^*$ is analytic inside and on the boundary of the unit circle, that is on $|z| \leq 1$. 

50
5.4 Filtering of Stationary Sequences

5.4.2 Example

Let us illustrate Yaglom’s solution to the problem of filtering of stationary sequences when the lag $\tau$ is an integer no smaller than $-1$.

Example 5.23. \mbox{([15], pp133)} Suppose that the spectral densities $f_Y$ and $f_X$ are given by

$$f_Y(\lambda) = B \frac{|e^{i\lambda} - b|^2}{|e^{i\lambda} - a|^2} \quad \text{and} \quad f_X(\lambda) = \frac{A}{|e^{i\lambda} - a|^2},$$

respectively, where $a$ and $b$ satisfying $|a|, |b| < 1$ and $A, B > 0$ are real constants. The function $\Psi_\tau^*$ in Theorem 5.22 on the previous page is then given by

$$\Psi_\tau^*(z) = z^\tau \frac{Az}{(z-a)(1-az)} - \Phi_\tau^*(z) \frac{B(z-b)(1-bz)}{(z-a)(1-az)} = \frac{\chi_\tau(z)}{(z-a)(1-az)},$$

where $\chi_\tau(z) = Az^{\tau+1} - B(z-b)(1-bz)\Phi_\tau^*(z)$. Note that $\chi_\tau(z) = (z-a)\delta_\tau(z)$ where $\delta_\tau$ is analytic on $|z| \leq 1$ to compensate for the zero at $z = a$ of the denominator since we require that $\Psi_\tau^*$ is analytic on $|z| \leq 1$ due to (F3*). Then, $\Phi_\tau^*(z) = \frac{Az^{\tau+1} - (z-a)\delta_\tau(z)}{B(z-b)(1-bz)}$. Since $\Phi_\tau^*$ is required to be analytic $|z| \geq 1$, we need that $Az^{\tau+1} - (z-a)\delta_\tau(z) = B(1-bz)\gamma_\tau(z)$ for some $\gamma_\tau$ that is analytic on $|z| \geq 1$ in order to compensate for the zero at $z = b^{-1}$ in the denominator of $\Phi_\tau^*$. Consequently, since $z^{\tau+1}$ is an entire function for integers $\tau \geq -1$, we deduce that both $\delta_\tau$ and $\gamma_\tau$ are entire. Therefore,

$$\Phi_\tau^*(z) = \frac{\gamma_\tau(z)}{z-b}$$

for some entire function $\gamma_\tau$. Recall that $\Psi_\tau^*(z) = \frac{Az^{\tau+1} - B(1-bz)\gamma_\tau(z)}{(z-a)(1-az)}$ must be analytic on $|z| \leq 1$, which is possible only if its numerator vanishes at $z = a$ to compensate for the zero at $z = a$ of its denominator, that is $\gamma_\tau(a) = \frac{A}{B} \frac{a^\tau}{1-\bar{a}b}$. In order to satisfy (F2*), we need to let $\gamma_\tau$ be the constant function $\gamma_\tau(z) \equiv \gamma_\tau(a)$. Thus, we obtain

$$\Phi_\tau^*(z) = \frac{A}{B} \frac{a^\tau + 1}{1-ab} \frac{1}{z-b} = \frac{A}{B} \frac{a^\tau + 1}{1-ab} \sum_{k=1}^{\infty} \frac{b^{k-1}}{z^k},$$

which satisfies (F1*) to (F3*) by construction. To conclude, the spectral characteristic of filtering $\Phi_\tau$ and the best linear filter $L_\tau$ are respectively given by

$$\Phi_\tau(\lambda) = \frac{A}{B} \frac{a^\tau + 1}{1-ab} \sum_{k=1}^{\infty} b^{k-1}e^{-i\lambda k} \quad \text{and} \quad L_\tau(t) = \frac{A}{B} \frac{a^\tau + 1}{1-ab} \sum_{k=1}^{\infty} b^{k-1}Y(t-k).$$
Minimax Approach to Filtering

6 Minimax Filtering Problem and Solution

6.1 Practical Motivation

The classical theory of filtering of stationary processes due to Kolmogorov and Wiener assumes perfect knowledge about the spectral function. Since we only consider the case when the spectral function attains a spectral density, one may as well say that the classical theory depends on perfect knowledge about the respective spectral densities.

However, this is rarely the case in practice. Instead, usually one uses statistical methods to estimate the spectral density based on finite observations. As a result, one only has partial knowledge about the true spectral density; for instance, one might infer that the spectral density lies in some class of conceivable spectral densities. If one wishes to carry out some sort of filtering based on such partial knowledge, one is left with no supporting mathematical theory in the classical framework which could guarantee optimality.

One way to deal with this issue neglects such uncertainty altogether: instead, one simply constructs the optimal filter associated to whatever spectral densities one deems to be true. As was demonstrated by Kassam and Poor in [7], Section II A, this can as a matter of fact result in very poor performance – to a point where the filtered signal is noisier than the original one. It is thus reasonable to develop some sort of minimax filtering theory, such that the resulting filter performance is rather insensitive to small deviations from the assumed spectral densities. Another common name in the literature is robust filtering.

6.2 Problem Formulation

Let us consider the problem of filtering of a stationary process similar to Section 5.2. We shall consider the sum $Y$ of two stationary processes $X$ and $N$, that is

$$Y = X + N,$$

where the process $X$ denotes the signal that we are interested in, and the process $N$ denotes some noise that we wish to remove. We shall be concerned with the non-causal filtering problem, that is estimating $X(t)$ based on observations $Y(s)$ for $-\infty < s < \infty$. 

52
Assumption 6.1. We assume that $X$ and $N$ are uncorrelated and mean square continuous. Further, we shall assume that $X$ and $N$ assume spectral densities $f_X$ and $f_N$, respectively. However, we shall assume that these spectral densities are not known despite their existence.

The assumption of not knowing the spectral densities is the main difference to the setting in the classical non-causal filtering problem considered in Section 5.2.

Spectral Classes

Definition 6.2. A subset $\mathcal{X}$ of the space of integrable functions on the real line, that is $\mathcal{X} \subset L^1(\mathbb{R})$, is called a spectral class if, for all $f \in \mathcal{X}$, the inequality $f(\lambda) \geq 0$ holds for all $\lambda$.

Remark 6.3. We shall interpret a spectral class as a set of potential spectral densities for the respective true but unknown spectral density. Such spectral class may be obtained through statistical estimation based on data. It could be a confidence band, for instance. That said, a practical interpretation of spectral classes is not necessary at this point.

Minimax Filtering Problem

In light of the practical motivation given in Section 6.1 on the preceding page, we are now ready to state the main problem which we refer to as the minimax filtering problem. Its formulation is due to Poor [13], Section II.

Definition 6.4. Let $\mathcal{C}$ denote the space of bounded complex-valued Lebesgue measurable functions on the real line.

Problem 6.5. Recall the non-causal filtering error (48) in Definition 5.11 on page 38. Given spectral classes $\mathcal{X}$ and $\mathcal{N}$, we wish to find a solution $\Phi_M \in \mathcal{C}$ to the minimax game

$$\inf_{\Phi \in \mathcal{C}} \sup_{(f,g) \in \mathcal{X} \times \mathcal{N}} e(f,g;\Phi). \tag{72}$$

We say that $\Phi_M$ is a solution to (72) if

$$\inf_{\Phi \in \mathcal{C}} \sup_{(f,g) \in \mathcal{X} \times \mathcal{N}} e(f,g;\Phi) = \sup_{(f,g) \in \mathcal{X} \times \mathcal{N}} e(f,g;\Phi_M). \tag{73}$$

Definition 6.6. A solution to the minimax game (72) is called minimax filter.

Remark 6.7. Let us draw a connection between Problem 6.5 and the classical problem of non-causal filtering of stationary processes as presented in Section 5.2 on page 38, where we assumed
that the spectral densities \( f_X \) and \( f_N \) are known. In the classical problem, we aim to find the Wiener filter \( \Phi^* \), see (49) in Definition 5.12 on page 38. As we saw in Proposition 5.13 on page 39, the Wiener filter is given by \( \Phi^*(\lambda) = \frac{f_X(\lambda)}{f_X(\lambda) + f_N(\lambda)} \) which belongs to \( C \). Therefore, the classical non-causal filtering problem is a special case of Problem 6.5 on the preceding page when the spectral classes \( \mathcal{X} \) and \( \mathcal{N} \) only contain the true spectral densities, that is when \( \mathcal{X} = \{ f_X \} \) and \( \mathcal{N} = \{ f_N \} \).

### 6.3 Least Favorable Spectral Densities

In this section, we shall be interested in when a solution to the minimax game (72) can be obtained, and how such a solution, that is a minimax filter, may look. In particular, we will see that the minimax game (72) can be straightforwardly solved provided that there exists a least favorable pair of spectral densities, which we shall define in Definition 6.8. At this point, the reader should be reminded of the non-causal filtering error (48) in Definition 5.11 on page 38.

**Definition 6.8.** ([13], Definition 1) A pair of spectral densities \((f_L, g_L)\) in \( \mathcal{X} \times \mathcal{N} \) is called **least favorable** for Wiener filtering for spectral classes \( \mathcal{X} \) and \( \mathcal{N} \) if

\[
e(f; g; \Phi^*_L) \leq e(f_L, g_L; \Phi^*_L) \quad \text{for all } (f, g) \in \mathcal{X} \times \mathcal{N}, \tag{74}
\]

where \( \Phi^*_L \) is the Wiener filter associated with \( f_L \) and \( g_L \). (see Definition 5.12 on page 38)

Henceforth, when using the term **least favorable**, we shall mean the above definition. The respective spectral classes \( \mathcal{X} \) and \( \mathcal{N} \) should then be apparent from the context.

**Proposition 6.9.** If \((f_L, g_L)\) is least favorable, then the pair \(((f_L, g_L), \Phi^*_L)\) will be a saddle point solution to the minimax game (72), that means

\[
\inf_{\Phi \in C} e(f_L, g_L; \Phi) = e(f_L, g_L; \Phi^*_L) = \sup_{(f, g) \in \mathcal{X} \times \mathcal{N}} e(f, g; \Phi^*_L). \tag{75}
\]

**Proof.** The first equality follows from definition of \( \Phi^*_L \) because it is by definition the minimizing function for \( e(f_L, g_L; \Phi) \), and moreover, it is measurable and bounded by 1, hence belongs to \( C \). The second equality follows from definition of \( (f_L, g_L) \), see (74).

Consequently, in order to solve the minimax game (72) it is sufficient to find a least favorable pair of spectral densities: based on such a pair, the corresponding Wiener filter will be a minimax filter. The problem is thus reduced to finding such a pair. The following Theorem gives an approach to finding least favorable spectral densities.
Theorem 6.10. ([13], Theorem 1) Suppose \( X \) and \( N \) be convex spectral classes. Then, the pair \((f_L, g_L)\) is least favorable in \( X \times N \) if and only if

\[
e^*(f_L, g_L) = \max_{(f, g) \in X \times N} e^*(f, g).
\]  

(76)

Proof. Let us first show necessity. If \((f_L, g_L)\) is least favorable in \( X \times N \), we observe

\[
e^*(f_L, g_L) = e(f_L, g_L; \Phi_L^*) \geq e(f, g; \Phi_L^*) \geq \min_{\Phi \in \mathcal{C}} e(f, g; \Phi) = e^*(f, g),
\]

where the first inequality follows from choice of \((f_L, g_L)\). This yields (76).

Next, we show sufficiency. Note that \( T \) defined by \( T(a, b) = \frac{ab}{a+b} \) is concave in \( a \) and \( b \), separately, for \( a \geq 0 \) and \( b \geq 0 \). So, the error functional \( e^* \), which is simply given by

\[
e^*(f, g) = \int_{-\infty}^{\infty} \frac{f(\lambda)g(\lambda)}{f(\lambda) + g(\lambda)} d\lambda = \int_{-\infty}^{\infty} T(f(\lambda), g(\lambda)) d\lambda,
\]

is concave in \( f \) and \( g \), separately, as well. Suppose a pair \((f_L, g_L) \in X \times N\) maximizes \( e^*(f, g) \). Since \( X \) and \( N \) are convex, it follows that the function \( J_f \) given by

\[
J_f(\epsilon) = e^*((1 - \epsilon)f_L + \epsilon f, g_L) = \int_{-\infty}^{\infty} \frac{[(1 - \epsilon)f_L(\lambda) + \epsilon f(\lambda)]g_L(\lambda)}{(1 - \epsilon)f_L(\lambda) + \epsilon f(\lambda) + g_L(\lambda)} d\lambda
\]

(77)

has a maximum at \( \epsilon = 0 \) for all \( f \in X \). Let the integrand in (77) be denoted by \( I_f(\epsilon, \lambda) \). Since \( f \mapsto e^*(f, g_L) \) is concave, this in turn implies that the right-side derivative of \( J \) at \( \epsilon = 0 \) is non-positive, that is

\[
\lim_{\epsilon \downarrow 0} \frac{J_f(\epsilon) - J_f(0)}{\epsilon} \leq 0 \quad \text{for all } f \in X.
\]

An application of Fatou’s Lemma yields

\[
\int_{-\infty}^{\infty} \lim_{\epsilon \downarrow 0} \frac{I_f(\epsilon, \lambda) - I_f(0, \lambda)}{\epsilon} d\lambda \leq \lim_{\epsilon \downarrow 0} \frac{J_f(\epsilon) - J_f(0)}{\epsilon} \leq 0 \quad \text{for all } f \in X.
\]

(78)

After doing the algebra and applying the limit, we obtain

\[
\lim_{\epsilon \downarrow 0} \frac{I_f(\epsilon, \lambda) - I_f(0, \lambda)}{\epsilon} = \left[ \frac{g_L(\lambda)}{f_L(\lambda) + g_L(\lambda)} \right]^2 (f(\lambda) - f_L(\lambda)),
\]

and thus (78) becomes

\[
\int_{-\infty}^{\infty} \left[ \frac{g_L(\lambda)}{f_L(\lambda) + g_L(\lambda)} \right]^2 f(\lambda) d\lambda \leq \int_{-\infty}^{\infty} \left[ \frac{g_L(\lambda)}{f_L(\lambda) + g_L(\lambda)} \right]^2 f_L(\lambda) d\lambda
\]

(79)
6.4 Examples

6.4.1 The p-Point Model

Let us illustrate the concept of least favorability with the p-point model. In this model, the spectral classes \( \mathcal{X} \) and \( \mathcal{N} \) are given as follows: for a partition \( A_1, \ldots, A_n \) of the real line and some positive numbers \( b_j \) and \( c_j \) consider

\[
\mathcal{X} = \left\{ f \mid \int_{A_j} f(\lambda) d\lambda = b_j \text{ for } j \leq n \right\} \quad \text{and} \quad \mathcal{N} = \left\{ g \mid \int_{A_j} g(\lambda) d\lambda = c_j \text{ for } j \leq n \right\} .
\]  

(81)

The minimax filtering problem with spectral classes (81) has been studied by Kassam and Cimini in [5], and we shall summarize its solution.
Let \( a_j = (\int_{A_j} d\lambda)^{-1} \), and consider the pair \((f_L, g_L)\) in \(\mathcal{X} \times \mathcal{N}\) given by

\[
f_L(\lambda) = \sum_{j=1}^{n} I_{A_j}(\lambda) a_j b_j \quad \text{and} \quad g_L(\lambda) = \sum_{j=1}^{n} I_{A_j}(\lambda) a_j c_j.
\]

We shall show that \((f_L, g_L)\) is a least favorable pair in \(\mathcal{X} \times \mathcal{N}\), see Definition 6.8 on page 54. The Wiener filter \(\Phi_L^*\) associated with \(f_L\) and \(g_L\) assumes the form

\[
\Phi_L^*(\lambda) = \frac{f_L(\lambda)}{f_L(\lambda) + g_L(\lambda)} = \sum_{j=1}^{n} I_{A_j}(\lambda) \frac{b_j}{b_j + c_j}.
\]  

Let \(d_j = \frac{b_j}{b_j + c_j}\). The error \(e(f, g; \Phi_L^*)\) for any \((f, g) \in \mathcal{X} \times \mathcal{N}\) is given by

\[
e(f, g; \Phi_L^*) = \int_{-\infty}^{\infty} |1 - \Phi_L^*(\lambda)|^2 f(\lambda) + |\Phi_L^*(\lambda)|^2 g(\lambda) \, d\lambda
\]

\[
= \sum_{j=1}^{n} \int_{A_j} (1 - d_j)^2 f(\lambda) + d_j^2 g(\lambda) \, d\lambda
\]

\[
= \sum_{j=1}^{n} (1 - d_j)^2 b_j + d_j^2 c_j
\]

\[
= \sum_{j=1}^{n} \int_{A_j} |1 - \Phi_L^*(\lambda)|^2 f_L(\lambda) + |\Phi_L^*(\lambda)|^2 g_L(\lambda) \, d\lambda.
\]  

The last term (83) is just the Wiener filtering error of \(f_L\) and \(g_L\), hence \(e(f, g; \Phi_L^*) = e(f_L, g_L; \Phi_L^*)\). This shows that \((f_L, g_L)\) is indeed a least favorable pair. Thus, \(\Phi_L^*\) is a solution to the minimax game (72). Moreover, the error \(e(f, g; \Phi_L^*)\) is constant over all \((f, g) \in \mathcal{X} \times \mathcal{N}\) because of \(e(f, g; \Phi_L^*) = e(f_L, g_L; \Phi_L^*)\) for all \((f, g) \in \mathcal{X} \times \mathcal{N}\) and

\[
e(f_L, g_L; \Phi_L^*) = \int_{-\infty}^{\infty} \frac{f_L(\lambda) g_L(\lambda)}{f_L(\lambda) + g_L(\lambda)} \, d\lambda = \sum_{j=1}^{n} \int_{A_j} \frac{a_j b_j a_j c_j}{a_j b_j + a_j c_j} \, d\lambda = \sum_{j=1}^{n} \frac{b_j c_j}{b_j + c_j}.
\]  

### 6.4.2 The Band Model and \(\epsilon\)-Contaminated Model

In this section, we briefly present the band model as well as the \(\epsilon\)-contaminated model. These two models for spectral classes have been studied by Kassam and Lim in [6]. The authors assume that there exist positive constants \(\nu_X^2\) and \(\nu_N^2\), such that \(\int f(\lambda) d\lambda = \nu_X^2\) and \(\int g(\lambda) d\lambda = \nu_N^2\) for all \((f, g) \in \mathcal{X} \times \mathcal{N}\).
6.4 Examples

The Band Model Suppose there are known functions $0 \leq L_X \leq U_X$ and $0 \leq L_N \leq U_N$ such that, for all $(f, g) \in X \times N$ and all $\lambda$,

$$L_X(\lambda) \leq f(\lambda) \leq U_X(\lambda) \quad \text{and} \quad L_N(\lambda) \leq g(\lambda) \leq U_N(\lambda).$$

A pair of least favorable spectral densities is given in Theorem 1 of the paper [6].

The $\epsilon$-Contaminated Model Suppose there exist non-negative constants $\epsilon_X, \epsilon_N$ and non-negative integrable functions $h_X, h_N$ such that for all $(f, g) \in X \times N$ there is pair of non-negative integrable functions $(\sigma_f, \sigma_g)$ such that, for all $\lambda$,

$$f(\lambda) = (1 - \epsilon_X)h_X(\lambda) + \epsilon_X\sigma_f(\lambda) \quad \text{and} \quad g(\lambda) = (1 - \epsilon_N)h_N(\lambda) + \epsilon_N\sigma_g(\lambda).$$

A pair of least favorable spectral densities is given in Theorem 2 of the paper [6].

We omit a closed form solution for both models due to their bulky formulation.
7 Analogy to Robust Hypotheses Testing

In what follows, we shall briefly introduce some notation from the theory of robust hypotheses testing which we will then apply to minimax filtering. This theory is largely due to Huber, see for instance [4]. The material presented here is, however, drawn from Poor’s paper [13].

7.1 Robust Hypotheses Testing

Let $\mathcal{P}_0$ and $\mathcal{P}_1$ be disjoint classes of probability density functions (pdf’s).

**Definition 7.1.** ([13], Definition 2) A pair of pdf’s $(q_0, q_1)$ in $\mathcal{P}_0 \times \mathcal{P}_1$ is said to be least favorable in terms of risk for $\mathcal{P}_0$ versus $\mathcal{P}_1$ if, for any $\pi \in (0, 1)$ and any $(p_0, p_1) \in \mathcal{P}_0 \times \mathcal{P}_1$, the inequalities

$$\int_{\pi_0 q_0 < \pi_1 q_1} q_0(x) \, dx \geq \int_{\pi_0 p_0 < \pi_1 p_1} p_0(x) \, dx,$$

$$\int_{\pi_0 q_0 \geq \pi_1 q_1} q_1(x) \, dx \geq \int_{\pi_0 p_0 \geq \pi_1 p_1} p_1(x) \, dx,$$

hold, where $\pi_0 = \pi$ and $\pi_1 = 1 - \pi$. We shall also use the short notation least favorable.

**Connection to Least Favorability in Terms of Bayes Risk**

Let us illustrate a connection between Definition 7.1 and least favorability in terms of Bayes risk. Let $p_0$ and $p_1$ be members of $\mathcal{P}_0$ and $\mathcal{P}_1$. Consider the simple hypotheses test $V \sim p_0$ versus $V \sim p_1$. Suppose prior probabilities $\pi_0 = \pi$ and $\pi_1 = 1 - \pi$, for some $\pi \in (0, 1)$, of $p_0$ and $p_1$ are given. Let $\phi$ be any given non-randomized test. The risk with zero-one-loss of testing two simple hypotheses using test $\phi$ is

$$B(\phi; p_0, p_1, \pi) = \pi_0 P(\phi(V) = 1|p = p_0) + \pi_1 P(\phi(V) = 0|p = p_1).$$

Let $\tilde{\phi}$ denote the Bayes test, that is the test minimizing the risk $B(\phi; p_0, p_1, \pi)$; it is given by

$$\tilde{\phi}(x) = \begin{cases} 0 & \text{if } \frac{p_0(x)}{p_1(x)} \geq \frac{\pi_1}{\pi_0} \\ 1 & \text{if } \frac{p_0(x)}{p_1(x)} < \frac{\pi_1}{\pi_0}. \end{cases}$$

The Bayes risk, that is the smallest Bayes risk among all non-randomized tests, is

$$B(\tilde{\phi}; p_0, p_1, \pi) = \pi_0 P(\tilde{\phi}(V) = 1|p = p_0) + \pi_1 P(\tilde{\phi}(V) = 0|p = p_1)$$

$$= \pi_0 \int_{\pi_0 p_0 < \pi_1 p_1} p_0(x) \, dx + \pi_1 \int_{\pi_0 p_0 \geq \pi_1 p_1} p_1(x) \, dx.$$
Note that if \((q_0, q_1)\) is a least favorable pair, then for any other pair \((p_0, p_1)\) and any \(\pi \in (0, 1)\),

\[
\pi_0 \int_{\pi_0 q_0 < \pi_1 q_1} q_0(x) \, dx + \pi_1 \int_{\pi_0 q_0 \geq \pi_1 q_1} q_1(x) \, dx \geq \pi_0 \int_{\pi_0 p_0 < \pi_1 p_1} p_0(x) \, dx + \pi_1 \int_{\pi_0 p_0 \geq \pi_1 p_1} p_1(x) \, dx.
\]

Thus, Definition 7.1 on the preceding page also implies that for any other pair \((p_0, p_1)\) and any \(\pi \in (0, 1)\),

\[
B(\hat{\phi}; q_0, q_1, \pi) \geq B(\hat{\phi}; p_0, p_1, \pi).
\]

In other words, the pair \((q_0, q_1)\) is also least favorable in terms of the Bayes risk for any prior probability \(\pi\).

### Characterization of Least Favorable Probability Densities

The following Lemma gives a characterization of least favorable pdf’s. A proof to this rather technical result can be found in [13].

**Lemma 7.2.** ([13], Lemma 1) Suppose \(P_0\) and \(P_1\) are classes of pdf’s such that all members of \(P_0 \cup P_1\) have the same support. If \((q_0, q_1) \in P_0 \times P_1\) is a least favorable pair then, for all continuous concave functions \(\psi\) and all \((p_0, p_1) \in P_0 \times P_1\),

\[
\int_{-\infty}^{\infty} \psi \left( \frac{q_1(x)}{q_0(x)} \right) q_0(x) \, dx \geq \int_{-\infty}^{\infty} \psi \left( \frac{p_1(x)}{p_0(x)} \right) p_0(x) \, dx.
\]

### 7.2 Connection to Minimax Filtering

Let us assume the second moments of the signal \(X\) and noise \(N\) are known. That is, consider spectral classes \(\mathcal{X}\) and \(\mathcal{N}\) such that, for known positive constants \(\nu_{\mathcal{X}}^2\) and \(\nu_{\mathcal{N}}^2\), they satisfy

\[
\int_{-\infty}^{\infty} f(\lambda) d\lambda = \nu_{\mathcal{X}}^2 \quad \text{and} \quad \int_{-\infty}^{\infty} g(\lambda) d\lambda = \nu_{\mathcal{N}}^2 \quad \text{for all } (f, g) \in \mathcal{X} \times \mathcal{N}.
\]

For the sake of readability, we change the notation to \(P_0 = P_{\mathcal{N}}\) and \(P_1 = P_{\mathcal{X}}\) where

\[
P_{\mathcal{N}} = \{ p \mid p = \nu_{\mathcal{N}}^{-2} g \text{ for some } g \in \mathcal{N} \} \quad \text{and} \quad P_{\mathcal{X}} = \{ p \mid p = \nu_{\mathcal{X}}^{-2} f \text{ for some } f \in \mathcal{X} \}.
\]

Note that these are simply the normalized versions of \(\mathcal{N}\) and \(\mathcal{X}\), respectively.

**Assumption 7.3.** We shall assume that \(\mathcal{X}\) and \(\mathcal{N}\) are such that \(P_{\mathcal{X}}\) and \(P_{\mathcal{N}}\) are disjoint.

The following Theorem draws a connection between the notions of least favorability in terms of risk and least favorability for Wiener filtering.
Theorem 7.4. ([13], Theorem 2) Suppose all members of $\mathcal{X} \cup \mathcal{N}$ have the same support. If $(q_X, q_N) \in \mathcal{P}_X \times \mathcal{P}_N$ is a pair of least favorable pdf’s in terms of risk for $\mathcal{P}_N$ versus $\mathcal{P}_X$ then

$$e^*(f_Q, g_Q) \geq e^*(f, g) \quad \text{for all } (f, g) \in \mathcal{X} \times \mathcal{N},$$

where $f_Q$ and $g_Q$ are defined as

$$f_Q(\lambda) = \nu^2_X q_X(\lambda) \quad \text{and} \quad g_Q(\lambda) = \nu^2_N q_N(\lambda).$$

Proof. Let us define the function $\psi$ as follows

$$\psi(x) = \frac{\nu^2_X \nu^2_N x}{(\nu^2_N + \nu^2_X x)}.$$

Note that $\psi$ is continuous and concave. Now, let $(f, g) \in \mathcal{X} \times \mathcal{N}$ be arbitrary, and let $(p_X, p_N) \in \mathcal{P}_X \times \mathcal{P}_N$ be defined by $p_X(\lambda) = \nu^{-2}_X f(\lambda)$ and $p_N(\lambda) = \nu^{-2}_N g(\lambda)$. Then, the Wiener filtering error of $(f, g)$ is given by

$$e^*(f, g) = \int_{-\infty}^{\infty} \frac{\nu^2_X p_X(\lambda) \nu^2_N p_N(\lambda)}{\nu^2_N p_X(\lambda) + \nu^2_X p_N(\lambda)} \, d\lambda.$$

Using $\psi$ we may rewrite this error as

$$e^*(f, g) = \int_{-\infty}^{\infty} \psi \left( \frac{p_X(x)}{p_N(x)} \right) p_N(x) \, d\lambda.$$

An application of Lemma 7.2 on the previous page yields the inequality

$$\int_{-\infty}^{\infty} \psi \left( \frac{q_X(x)}{q_N(x)} \right) q_N(x) \, d\lambda \geq \int_{-\infty}^{\infty} \psi \left( \frac{p_X(x)}{p_N(x)} \right) p_N(x) \, d\lambda,$$

which is equivalent to $e^*(f_Q, g_Q) \geq e^*(f, g)$. \qed

An application of Theorem 6.10 on page 55 yields the following Corollary.

Corollary 7.5. ([13], Corollary 1) Suppose $\mathcal{X}$ and $\mathcal{N}$ are convex spectral classes such that all members of $\mathcal{X} \cup \mathcal{N}$ have the same support. If $(q_X, q_N) \in \mathcal{P}_X \times \mathcal{P}_N$ is a pair of least favorable pdf’s in terms of risk for $\mathcal{P}_N$ versus $\mathcal{P}_X$, then the pair $(f_Q, g_Q)$ defined by

$$f_Q(\lambda) = \nu^2_X q_X(\lambda) \quad \text{and} \quad g_Q(\lambda) = \nu^2_N q_N(\lambda)$$

is a pair of least favorable spectral densities for $\mathcal{X}$ and $\mathcal{N}$. 61
A general method for finding such a pair of least favorable pdf’s in terms of risk is due to Huber [4], where the author gives an explicit solution. In the context of minimax filtering, this method has been successfully demonstrated by Poor [13] for total-variation spectral classes, which we shall present in the next section.

### 7.3 Example – Total-Variation Model

In this section, which is drawn from Poor’s paper [13], Section III B, we consider spectral classes for which the analogy to robust hypotheses testing yields a minimax filter based on least favorable spectral densities. Let $\mathcal{X}$ and $\mathcal{N}$ be such that for all $(f, g)$ in $\mathcal{X} \times \mathcal{N}$ we have

$$\int_{-\infty}^{\infty} |f(\lambda) - f_0(\lambda)|d\lambda \leq \alpha_{\mathcal{X}} \quad \text{and} \quad \int_{-\infty}^{\infty} |g(\lambda) - g_0(\lambda)|d\lambda \leq \alpha_{\mathcal{N}},$$

where $\alpha_{\mathcal{X}}$ and $\alpha_{\mathcal{N}}$ are fixed constants, and $(f_0, g_0)$ is a known pair of non-negative integrable functions satisfying the constraints (87), respectively. In other words, $\mathcal{X}$ and $\mathcal{N}$ are $L^1(\mathbb{R})$ neighborhoods of $f_0$ and $g_0$, respectively. It should be pointed out, however, that we do not claim that $f_0$ and $g_0$ are the actual true spectral densities belonging to $\mathcal{X}$ and $\mathcal{N}$!

Based on $\mathcal{X}$ and $\mathcal{N}$, we define classes of pdf’s by

$$\mathcal{P}_\mathcal{X} = \left\{ p \mid \int_{-\infty}^{\infty} |p(\lambda) - p_X(\lambda)|d\lambda \leq \beta_X \right\} \quad \text{and} \quad \mathcal{P}_\mathcal{N} = \left\{ p \mid \int_{-\infty}^{\infty} |p(\lambda) - p_N(\lambda)|d\lambda \leq \beta_N \right\},$$

where $p_X(\lambda) = \nu_{\mathcal{X}}^{-2}f_0(\lambda)$ and $p_N(\lambda) = \nu_{\mathcal{N}}^{-2}g_0(\lambda)$ are known pdf’s, and $\beta_X = \nu_{\mathcal{X}}^{-2}\alpha_{\mathcal{X}}$ and $\beta_N = \nu_{\mathcal{N}}^{-2}\alpha_{\mathcal{N}}$ are fixed constants.

The author particularly considers the case $0 < \beta_X = \beta_N = \beta$, which we shall pursue as well. Further, the author assumes that $\beta$ is small enough so that $\mathcal{P}_\mathcal{X}$ and $\mathcal{P}_\mathcal{N}$ do not intersect. Let $q_X$ and $q_N$ be pdf’s defined by

$$q_X(\lambda) = \begin{cases} \frac{c'}{1+c'} (p_N(\lambda) + p_X(\lambda)) & \text{if } p_X(\lambda) \leq c' p_N(\lambda) \\ p_X(\lambda) & \text{if } c' p_N(\lambda) < p_X(\lambda) < c'' p_N(\lambda) \\ \frac{c''}{1+c''} (p_N(\lambda) + p_X(\lambda)) & \text{if } p_X(\lambda) \geq c'' p_N(\lambda), \end{cases}$$

$$q_N(\lambda) = \begin{cases} \frac{1}{1+c'} (p_N(\lambda) + p_X(\lambda)) & \text{if } p_X(\lambda) \leq c' p_N(\lambda) \\ p_N(\lambda) & \text{if } c' p_N(\lambda) < p_X(\lambda) < c'' p_N(\lambda) \\ \frac{1}{1+c''} (p_N(\lambda) + p_X(\lambda)) & \text{if } p_X(\lambda) \geq c'' p_N(\lambda), \end{cases}$$
where \( c' \) and \( c'' \) are numbers such that

\[
\int_{-\infty}^{\infty} |q_X(\lambda) - p_X(\lambda)|d\lambda = \int_{-\infty}^{\infty} |q_N(\lambda) - p_N(\lambda)|d\lambda = \beta. \tag{94}
\]

Now, as was shown by Huber [4], \( c' \) and \( c'' \) can in fact be chosen that way with \( 0 < c' < c'' < \infty \) since \( \mathcal{P}_X \) and \( \mathcal{P}_N \) do not intersect by assumption. Further, Huber shows in [4] that \((q_X, q_N)\) is least favorable in terms of risk for \( \mathcal{P}_N \) versus \( \mathcal{P}_X \). Thus, invoking Corollary 7.5 on page 61, we deduce that

\[
(f_L, g_L) = (\nu^2_X q_X, \nu^2_N q_N) \tag{95}
\]

is a pair of least favorable spectral densities in \( X \times N \) provided that \( \nu^{-2}_X \alpha_X = \nu^{-2}_N \alpha_N \) and that \( \beta = \nu^{-2}_X \alpha_X \) is such that (94) holds with \( 0 < c' < c'' < \infty \).

Therefore, the Wiener filter \( \Phi^*_L \) for the pair \((f_L, g_L)\), that is

\[
\Phi^*_L(\lambda) = \frac{\nu^2_X q_X(\lambda)}{\nu^2_X q_X(\lambda) + \nu^2_N q_N(\lambda)}, \tag{96}
\]

is a solution to the minimax game (72). It can be rewritten as

\[
\Phi^*_L(\lambda) = \min \left\{ k'', \max \left\{ k', \Phi^*_0(\lambda) \right\} \right\}, \tag{97}
\]

where \( \Phi^*_0 \) is the Wiener filter associated to the known spectral densities \((f_0, g_0)\), that is \( \Phi^*_0(\lambda) = \frac{f_0(\lambda)}{f_0(\lambda) + g_0(\lambda)} \), and \( k' = \frac{c' \nu^2_X}{\nu^2_X + c' \nu^2_X} \) and \( k'' = \frac{c'' \nu^2_X}{\nu^2_X + c'' \nu^2_X} \) are constants.
Conclusion and Further Research

To this end, we have considered the non-causal minimax filtering problem. However, similar investigations have been undertaken for causal minimax filtering as well. The interested reader may be referred to the survey paper [7] by Kassam and Poor in which the authors provide a general overview of minimax techniques in prediction and filtering of stationary processes.

Further, there are some well known and mathematically engaging filters such as the Chebyshev filter or the Cauer-Zolotarev filter, see for instance [9]. These filters possess certain optimality properties, but not of the type attributed to minimax as presented in this work. Regarding further research, it would be interesting to investigate whether there is a meaningful connection between those classical filters and the type of optimality associated with the minimax filters. In other words, whether there are some classes of spectral densities for which those filters are minimax, or approximately minimax.
References


