Stationary Gaussian Processes and the Kolmogorov–Wiener Prediction Problem

Ziran Liu

Courant Institute of Mathematical Sciences

New York University

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

May 2016

Acknowledgements

I am deeply indebted to Professor Henry P. McKean for his constant encouragement, insightful guidance, and remarkable generosity with his time. His explanations, intuition, and mathematical vision have been an enduring source of inspiration throughout this work.

I would also like to thank the faculty and students at the Courant Institute for providing a stimulating and supportive mathematical environment.

Abstract

This thesis studies the classical prediction problem for centered stationary Gaussian processes, with particular emphasis on the Kolmogorov–Wiener problem of predicting the future from the whole past.

The exposition is organized around three main ingredients. First, we review the Hardy spaces on the upper and lower half-planes, including the Paley–Wiener theorem, the orthogonal splitting of $L^2(\mathbb{R})$, outer-inner factorization, and the F. & M. Riesz theorem. Second, we develop the spectral representation of a mean-square continuous stationary Gaussian process via its covariance function and spectral measure, and identify the Gaussian Hilbert space generated by the process with the spectral space $L^2(\mathbb{R}, \mu)$. Third, we formulate the prediction problem in this spectral setting, explain the role of Szegő's alternative and the remote past, and derive an explicit prediction formula in the purely nondeterministic case by means of Hardy-space factorization.

In particular, when the spectral measure is absolutely continuous and its density satisfies the Szegő condition, the best linear predictor and its error variance can be written explicitly in terms of an outer spectral factor. We also describe the role of the singular spectral component as the deterministic part that survives in the remote past, and identify the finite-horizon innovation space with $L^2(0, T)$.

Contents

Introduction

The purpose of this thesis is to study the classical prediction problem for centered stationary Gaussian processes, especially the problem of predicting $X(T)$ from the whole past $\{X(t) : t \leq 0\}$. The guiding idea is that this prediction problem, although probabilistic in origin, becomes a problem in Hilbert-space geometry after passing to the Gaussian Hilbert space generated by the process.

A second fundamental step is the spectral representation of the process. Under this representation, the process is transported to the spectral space $L^2(\mathbb{R}, \mu)$, and the past becomes a closed span of exponentials. At this stage, methods from Hardy-space theory become available. In particular, outer-inner factorization and Szegő's alternative provide the key tools for distinguishing deterministic and nondeterministic behavior and for obtaining explicit prediction formulas in the purely nondeterministic case.

The thesis is organized as follows. In Chapter 1, we review the Hardy spaces on the upper and lower half-planes, together with the Paley–Wiener theorem, outer functions, and the F. & M. Riesz theorem. In Chapter 2, we develop the spectral representation of mean-square continuous stationary Gaussian processes and identify the corresponding Gaussian Hilbert space with $L^2(\mathbb{R}, \mu)$. In Chapter 3, we formulate the Kolmogorov–Wiener prediction problem in the spectral space, state Szegő's alternative, derive explicit formulas for the best linear predictor and its error variance in the purely nondeterministic case, and describe the associated finite-horizon innovation space.

Chapter 1

Hardy Spaces on the Half-Plane

This chapter collects the function-theoretic tools needed later for the prediction problem of Kolmogorov and Wiener. The main objects are the Hardy spaces on the upper and lower half-planes, their boundary-value theory, the orthogonal splitting of $L^2(\mathbb{R})$, and the outer-inner factorization. Standard references are Paley–Wiener [8], Beurling [1], and Dym–McKean [3, 2].

1.1 Fourier Transform and Hardy Spaces

We begin by fixing the Fourier-transform convention used throughout the thesis. For $f \in L^1(\mathbb{R})$, define

$$\widehat{f}(\gamma) = \int_{\mathbb{R}} e^{i\gamma x} f(x)\ dx.$$

If in addition $\widehat{f} \in L^1(\mathbb{R})$, then the inversion formula reads

$$f(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-ix\gamma} \widehat{f}(\gamma)\ d\gamma.$$

For functions for which the inverse Fourier transform is meaningful, we write

$$f^{\vee}(x) := \frac{1}{2\pi} \int_{\mathbb{R}} e^{-ix\gamma} f(\gamma)\ d\gamma.$$

Whenever $f \in L^2(\mathbb{R})$, the inverse transform is understood in the L^2-sense via Plancherel's theorem.

By density of the Schwartz class in $L^2(\mathbb{R})$, the Fourier transform extends uniquely to $L^2(\mathbb{R})$, and Plancherel's identity holds:

$$\|\widehat{f}\|_{L^2(\mathbb{R})} = (2\pi)^{1/2} \|f\|_{L^2(\mathbb{R})}.$$

Equivalently,

$$(\widehat{f}, \widehat{g})_{L^2(\mathbb{R})} = 2\pi (f, g)_{L^2(\mathbb{R})}.$$

We write

$$\mathbb{C}_+ := \{z \in \mathbb{C} : \Im z > 0\}, \qquad \mathbb{C}_- := \{z \in \mathbb{C} : \Im z < 0\}.$$

For a function h on $\mathbb{C}_+$ and $b > 0$, define the horizontal translate

$$h_b(a) := h(a + ib), \qquad a \in \mathbb{R}.$$

Definition 1.1. A holomorphic function h on $\mathbb{C}_+$ belongs to the Hardy space $H^2(\mathbb{C}_+)$ if

$$\|h\|^2_{H^2(\mathbb{C}_+)} := \sup_{b>0} \int_{\mathbb{R}} |h(a + ib)|^2 \, da < \infty.$$

Similarly, a holomorphic function h on $\mathbb{C}_-$ belongs to $H^2(\mathbb{C}_-)$ if

$$\|h\|^2_{H^2(\mathbb{C}_-)} := \sup_{b<0} \int_{\mathbb{R}} |h(a + ib)|^2 \, da < \infty.$$

The corresponding H^1-spaces are defined in the same way with the L^1-norm replacing the L^2-norm.

Definition 1.2. A holomorphic function h on $\mathbb{C}_+$ belongs to $H^1(\mathbb{C}_+)$ if

$$\|h\|_{H^1(\mathbb{C}_+)} := \sup_{b>0} \int_{\mathbb{R}} |h(a + ib)| \, da < \infty.$$

The space $H^1(\mathbb{C}_-)$ is defined analogously.

1.2 Paley–Wiener Representation for $H^2(\mathbb{C}_+)$

The fundamental structure theorem for $H^2(\mathbb{C}_+)$ is the Paley–Wiener theorem.

Theorem 1.3 (Paley–Wiener theorem for $H^2(\mathbb{C}_+)$)**.** *A holomorphic function h belongs to $H^2(\mathbb{C}_+)$ if and only if there exists a function $g \in L^2(0, \infty)$ such that*

$$h(z) = \int_0^\infty e^{izx} g(x) \, dx, \qquad z \in \mathbb{C}_+.$$

Moreover,

$$\sup_{b>0} \int_{\mathbb{R}} |h(a + ib)|^2 \, da = 2\pi \int_0^\infty |g(x)|^2 \, dx.$$

Proof. See [8, Chapter X] or [3, Chapter 2]. For later use, we note the easy direction: if $g \in L^2(0,\infty)$ and

$$h(a+ib) = \int_0^\infty e^{iax} e^{-bx} g(x)\, dx,$$

then h is holomorphic on $\mathbb{C}_+$, and

$$h_b = (\widehat{e^{-bx} g(x) \mathbf{1}_{(0,\infty)}}(x)).$$

Hence Plancherel gives

$$\int_{\mathbb{R}} |h(a+ib)|^2\, da = 2\pi \int_0^\infty e^{-2bx} |g(x)|^2\, dx \le 2\pi \int_0^\infty |g(x)|^2\, dx.$$

Taking the supremum over $b > 0$ yields the required estimate. The converse is the classical Paley–Wiener theorem. $\square$

The lower-half-plane version is completely analogous.

Theorem 1.4 (Paley–Wiener theorem for $H^2(\mathbb{C}_-)$)**.** *A holomorphic function h belongs to $H^2(\mathbb{C}_-)$ if and only if there exists a function $g \in L^2(-\infty, 0)$ such that*

$$h(z) = \int_{-\infty}^0 e^{izx} g(x)\, dx, \qquad z \in \mathbb{C}_-.$$

Moreover,

$$\sup_{b<0} \int_{\mathbb{R}} |h(a+ib)|^2\, da = 2\pi \int_{-\infty}^0 |g(x)|^2\, dx.$$

As a consequence, every $h \in H^2(\mathbb{C}_+)$ admits nontangential boundary values $h_{0+} \in L^2(\mathbb{R})$, and the map

$$H^2(\mathbb{C}_+) \ni h \longmapsto h_{0+} \in L^2(\mathbb{R})$$

identifies $H^2(\mathbb{C}_+)$ with the closed subspace of boundary functions whose inverse Fourier transform is supported in $[0,\infty)$. Likewise, $H^2(\mathbb{C}_-)$ is identified with the subspace of boundary functions whose inverse Fourier transform is supported in $(-\infty, 0]$.

1.3 Orthogonal Splitting of $L^2(\mathbb{R})$

The Paley–Wiener theorem yields the basic orthogonal decomposition of $L^2(\mathbb{R})$.

Proposition 1.5. *There is an orthogonal direct-sum decomposition*

$$L^2(\mathbb{R}) = H^2(\mathbb{C}_-) \oplus H^2(\mathbb{C}_+),$$

where the Hardy spaces are identified with their L^2-boundary values on $\mathbb{R}$.

Proof. By Plancherel,

$$L^2(\mathbb{R}) \cong L^2((-\infty,0)) \oplus L^2((0,\infty)),$$

via inverse Fourier transform. Theorems 1.3 and 1.4 identify these two summands with $H^2(\mathbb{C}_+)$ and $H^2(\mathbb{C}_-)$, respectively. Orthogonality is inherited from the orthogonality of $L^2((-\infty,0))$ and $L^2((0,\infty))$. ☐

Let P_+ and P_- denote the orthogonal projections of $L^2(\mathbb{R})$ onto $H^2(\mathbb{C}_+)$ and $H^2(\mathbb{C}_-)$, respectively.

Proposition 1.6 (Cauchy projection)**.** *For $f \in L^2(\mathbb{R})$, the Cauchy transform*

$$(Cf)(z) := \frac{1}{2\pi i}\int_{\mathbb{R}} \frac{f(\gamma)}{\gamma - z}\, d\gamma, \qquad z \in \mathbb{C}_+,$$

belongs to $H^2(\mathbb{C}_+)$, and its boundary value on $\mathbb{R}$ is P_+f. Similarly, for $z \in \mathbb{C}_-$,

$$\frac{-1}{2\pi i}\int_{\mathbb{R}} \frac{f(\gamma)}{\gamma - z}\, d\gamma$$

has boundary value P_-f.

Proof. This is the standard Cauchy–Szegő projection formula for the half-plane; see [3, Chapter 2] or [2, Chapter 17]. ☐

1.4 Poisson Kernel and Outer Functions

For $b > 0$, define the Poisson kernel on the upper half-plane by

$$p_b(a) := \frac{1}{\pi}\frac{b}{a^2 + b^2}.$$

If $u \in L^1_{\mathrm{loc}}(\mathbb{R})$ and

$$\int_{\mathbb{R}} \frac{|u(\gamma)|}{1+\gamma^2}\, d\gamma < \infty,$$

then its Poisson integral is

$$(Pu)(a+ib) := (p_b * u)(a) = \frac{1}{\pi}\int_{\mathbb{R}} \frac{b\, u(\gamma)}{(a-\gamma)^2 + b^2}\, d\gamma.$$

We next record the outer-function theorem in the form needed later.

Theorem 1.7 (Outer function associated with a spectral density)**.** *Let $\Delta_0 \geq 0$ be measurable on $\mathbb{R}$, with $\Delta_0 \in L^1(\mathbb{R})$. Then*

$$\int_{\mathbb{R}} \frac{\log^+ \Delta_0(\gamma)}{1+\gamma^2}\, d\gamma < \infty.$$

Moreover, there exists an outer function $h \in H^2(\mathbb{C}_+)$ such that

$$|h_{0+}(\gamma)|^2 = \Delta_0(\gamma) \qquad \text{for a.e. } \gamma \in \mathbb{R}$$

if and only if

$$\int_{\mathbb{R}} \frac{\log \Delta_0(\gamma)}{1+\gamma^2}\, d\gamma > -\infty.$$

In that case one may take

$$h(z) = \exp\left\{ \frac{1}{2\pi i} \int_{\mathbb{R}} \frac{1+\gamma z}{\gamma - z} \, \frac{\log \Delta_0(\gamma)}{1+\gamma^2}\, d\gamma \right\}, \qquad z \in \mathbb{C}_+,$$

where $\log \Delta_0$ *is any measurable representative.*

Proof. This is the standard outer-function construction on the half-plane; see [3, Chapter 2] or [8, Chapter X]. The integrability of $\log^+ \Delta_0$ follows from the elementary bound

$$\log^+ x \leq x, \qquad x \geq 0,$$

hence

$$\int_{\mathbb{R}} \frac{\log^+ \Delta_0(\gamma)}{1+\gamma^2}\, d\gamma \leq \int_{\mathbb{R}} \Delta_0(\gamma)\, d\gamma < \infty.$$

□

Definition 1.8. A nonzero function $h \in H^2(\mathbb{C}_+)$ is called *outer* if

$$\log |h(z)| = (P \log |h_{0+}|)(z), \qquad z \in \mathbb{C}_+.$$

A bounded holomorphic function j on $\mathbb{C}_+$ is called *inner* if its boundary value satisfies

$$|j_{0+}(\gamma)| = 1 \qquad \text{for a.e. } \gamma \in \mathbb{R}.$$

Theorem 1.9 (Inner-outer factorization)**.** *Every nonzero function $h \in H^2(\mathbb{C}_+)$ can be written in the form*

$$h = j\, h_{\text{out}},$$

where j is inner and h_{out} is outer. This factorization is unique up to multiplication of j and h_{out} by unimodular constants.

Proof. See [3, Chapter 2] or [2, Chapter 17]. □

The following criterion explains why outer functions appear naturally in prediction theory.

Theorem 1.10 (Test for outer functions)**.** *Let $h \in H^2(\mathbb{C}_+)$ be nonzero, and write $h = jh_{\mathrm{out}}$ for its inner-outer factorization. Then the closed linear span of*

$$\{e^{i\gamma x}h(\gamma) : x \geq 0\}$$

in $L^2(\mathbb{R})$ is precisely $jH^2(\mathbb{C}_+)$. In particular, this span is all of $H^2(\mathbb{C}_+)$ if and only if h is outer.

Proof. See [3, Chapter 2]. The key point is that multiplication by an inner function is an isometry on $H^2(\mathbb{C}_+)$, while the translates $e^{i\gamma x}$ correspond, on the inverse Fourier side, to shifts of functions supported on $[0, \infty)$. □

Remark 1.11. All the preceding statements have completely analogous versions on the lower half-plane $\mathbb{C}_-$. In particular, if $h \in H^2(\mathbb{C}_-)$ is outer, then the closed linear span of

$$\{e^{i\gamma x}h(\gamma) : x \leq 0\}$$

is all of $H^2(\mathbb{C}_-)$.

1.5 H^1 and the F. & M. Riesz Theorem

The final function-theoretic ingredient needed later is the classical theorem of F. and M. Riesz.

Theorem 1.12 (F. and M. Riesz)**.** *Let F be a complex function of bounded variation on $\mathbb{R}$, and let dF be its associated finite signed measure. Assume that*

$$\int_{\mathbb{R}} e^{-i\gamma x}\, dF(\gamma) = 0 \qquad \textit{for every } x < 0.$$

Then dF is absolutely continuous with respect to Lebesgue measure, and

$$dF(\gamma) = h_{0+}(\gamma)\, d\gamma$$

for some boundary function h_{0+} of a function $h \in H^1(\mathbb{C}_+)$. Equivalently,

$$F(\gamma) = C + \int_{-\infty}^{\gamma} h_{0+}(\gamma')\, d\gamma'$$

for some constant C.

Proof. This is the classical F. & M. Riesz theorem for the half-plane; see [3, Chapter 2] and [2, Chapter 17]. The point relevant for us is that a finite measure whose negative Fourier frequencies vanish must be absolutely continuous, with density given by an H^1-boundary function. □

We shall use Theorem 1.12 later in the proof of Szegő's alternative, where it serves as the bridge from orthogonality relations in the spectral space to analytic structure in the Hardy space.

Chapter 2

Stationary Gaussian Processes

In this chapter we review the basic structure theory of centered stationary Gaussian processes. The main goal is to pass from the process itself to its spectral representation, which will be the natural setting for the prediction problem studied in the next chapter.

The central points are the following:

(1) a centered Gaussian process is completely determined by its covariance function;

(2) for a stationary Gaussian process, the covariance depends only on the time lag and is a continuous positive-definite function;

(3) by the spectral theorem for strongly continuous unitary groups, such a covariance function admits a spectral representation by a finite positive measure;

(4) the closed linear span of the process is naturally identified with an L^2-space over that spectral measure.

2.1 Gaussian Families and Gaussian Hilbert Spaces

We begin with the basic Gaussian framework.

Definition 2.1 (Centered Gaussian family)**.** A family of real random variables $\{X_\alpha\}_{\alpha\in I}$, defined on a common probability space $(\Omega, \mathcal{F}, P)$, is called a *centered Gaussian family* if for every $n \geq 1$, every choice $\alpha_1, \dots, \alpha_n \in I$, and every real numbers $c_1, \dots, c_n$, the random variable

$$c_1 X_{\alpha_1} + \cdots + c_n X_{\alpha_n}$$

has a (possibly degenerate) Gaussian distribution with mean 0.

For a centered Gaussian family, all finite-dimensional distributions are determined by the covariance matrix

$$E[X_{\alpha_j} X_{\alpha_k}], \qquad 1 \le j, k \le n.$$

Indeed, if

$$Y = \sum_{j=1}^{n} c_j X_{\alpha_j},$$

then

$$E[Y] = 0, \qquad \operatorname{Var}(Y) = E[Y^2] = \sum_{j,k=1}^{n} c_j c_k \, E[X_{\alpha_j} X_{\alpha_k}].$$

Definition 2.2 (Gaussian Hilbert space)**.** Let $\{X_\alpha\}_{\alpha \in I}$ be a centered Gaussian family. Its associated *Gaussian Hilbert space* is the closed linear span

$$G := \overline{\operatorname{span}_{\mathbb{R}}\{X_\alpha : \alpha \in I\}} \subset L^2(\Omega, \mathcal{F}, P),$$

equipped with the inner product

$$\langle Y, Z \rangle_G := E[YZ].$$

Thus G is a real Hilbert space. Later, when we pass to spectral representations, we shall also use its complexification

$$G_{\mathbb{C}} := \overline{\operatorname{span}_{\mathbb{C}}\{X_\alpha : \alpha \in I\}} \subset L^2(\Omega, \mathcal{F}, P; \mathbb{C}),$$

equipped with the complex inner product

$$\langle Y, Z \rangle_{G_{\mathbb{C}}} := E\Big[Y \overline{Z}\Big].$$

2.2 Stationary Gaussian Processes and the Covariance Function

Definition 2.3 (Centered stationary Gaussian process)**.** A *centered stationary Gaussian process* is a centered Gaussian family

$$\{X(t) : t \in \mathbb{R}\}$$

such that for every $n \geq 1$, every $t_1, \dots, t_n \in \mathbb{R}$, and every shift $h \in \mathbb{R}$, the vectors

$$\big(X(t_1), \dots, X(t_n)\big) \qquad \text{and} \qquad \big(X(t_1 + h), \dots, X(t_n + h)\big)$$

have the same distribution.

Because the process is Gaussian and centered, stationarity is equivalent to invariance of the covariance under shifts.

Proposition 2.4. *A centered Gaussian process $\{X(t)\}_{t\in\mathbb{R}}$ is stationary if and only if there exists a function $Q : \mathbb{R} \to \mathbb{R}$ such that*

$$E[X(t)X(s)] = Q(t - s), \qquad s, t \in \mathbb{R}.$$

In that case,

$$Q(\tau) = E[X(\tau)X(0)]$$

is called the covariance function of the process.

Proof. If the process is stationary, then

$$E[X(t)X(s)] = E[X(t - s)X(0)],$$

so the covariance depends only on $t - s$.

Conversely, for a centered Gaussian process, every finite-dimensional distribution is determined by its covariance matrix. If the covariance is of the form

$$E[X(t_j)X(t_k)] = Q(t_j - t_k),$$

then for every shift h,

$$E[X(t_j + h)X(t_k + h)] = Q((t_j + h) - (t_k + h)) = Q(t_j - t_k),$$

so the covariance matrix is unchanged, and hence the finite-dimensional distribution is unchanged. Therefore the process is stationary. □

We now impose the natural regularity assumption that the process be continuous in mean square.

Definition 2.5 (Mean-square continuity)**.** A process $\{X(t)\}_{t\in\mathbb{R}} \subset L^2(\Omega)$ is called *mean-square continuous* if

$$\lim_{h\to 0} \|X(t + h) - X(t)\|_{L^2(\Omega)} = 0 \qquad \text{for every } t \in \mathbb{R}.$$

For a centered stationary Gaussian process, mean-square continuity is equivalent to continuity of the covariance function at the origin.

Proposition 2.6. *Let $\{X(t)\}_{t\in\mathbb{R}}$ be a centered stationary Gaussian process with covariance function Q. Then*

$$E\Big[(X(t)-X(s))^2\Big] = 2\Big(Q(0)-Q(t-s)\Big), \qquad s,t\in\mathbb{R}.$$

Consequently, the process is mean-square continuous if and only if Q is continuous at 0*.*

Proof. By stationarity,

$$E[X(t)^2] = E[X(s)^2] = Q(0), \qquad E[X(t)X(s)] = Q(t-s).$$

Hence

$$E\Big[(X(t)-X(s))^2\Big] = E[X(t)^2] + E[X(s)^2] - 2E[X(t)X(s)] = 2\Big(Q(0)-Q(t-s)\Big).$$

Therefore

$$\|X(t+h)-X(t)\|_{L^2}^2 = 2\Big(Q(0)-Q(h)\Big),$$

and the claim follows immediately. □

Proposition 2.7. *Let Q be the covariance function of a centered stationary Gaussian process. Then:*

(1) $Q(0)\geq 0$*;*

(2) Q *is even:*

$$Q(-t) = Q(t), \qquad t\in\mathbb{R};$$

(3) Q *is positive definite, i.e.*

$$\sum_{j,k=1}^{n} c_j\overline{c_k}\,Q(t_j-t_k)\geq 0$$

for every $n\geq 1$, every $t_1,\dots,t_n\in\mathbb{R}$, and every $c_1,\dots,c_n\in\mathbb{C}$;

(4) Q *is bounded and satisfies*

$$|Q(t)|\leq Q(0), \qquad t\in\mathbb{R}.$$

Proof. The first point is immediate since

$$Q(0) = E[X(0)^2] \geq 0.$$

For the second point,

$$Q(-t) = E[X(-t)X(0)].$$

By stationarity,

$$E[X(-t)X(0)] = E[X(0)X(t)] = E[X(t)X(0)] = Q(t),$$

so Q is even.

For the third point, take arbitrary $c_1, \ldots, c_n \in \mathbb{C}$ and $t_1, \ldots, t_n \in \mathbb{R}$. Then

$$Y := \sum_{j=1}^{n} c_j X(t_j) \in G_{\mathbb{C}},$$

and therefore

$$0 \leq E[|Y|^2] = E\left[\left(\sum_{j=1}^{n} c_j X(t_j)\right) \overline{\left(\sum_{k=1}^{n} c_k X(t_k)\right)}\right] = \sum_{j,k=1}^{n} c_j \overline{c_k}\, E[X(t_j)X(t_k)].$$

Since $E[X(t_j)X(t_k)] = Q(t_j - t_k)$, the claimed positive-definiteness follows.

For the fourth point, Cauchy–Schwarz gives

$$|Q(t)| = |E[X(t)X(0)]| \leq \Big(E[X(t)^2]\Big)^{1/2} \Big(E[X(0)^2]\Big)^{1/2} = Q(0),$$

again by stationarity. □

The previous proposition gives the correct structural condition on Q: it is *positive definite*, not merely pointwise nonnegative. This distinction is essential.

Proposition 2.8. *If a centered stationary Gaussian process is mean-square continuous, then the Gaussian Hilbert space*

$$G = \overline{\operatorname{span}_{\mathbb{R}}\{X(t) : t \in \mathbb{R}\}}$$

is separable. More precisely,

$$G = \overline{\operatorname{span}_{\mathbb{R}}\{X(q) : q \in \mathbb{Q}\}}.$$

Proof. Fix $t \in \mathbb{R}$, and choose rationals $q_n \to t$. By mean-square continuity,

$$\|X(q_n) - X(t)\|_{L^2} \to 0.$$

Thus every $X(t)$ lies in the closure of $\operatorname{span}_{\mathbb{R}}\{X(q) : q \in \mathbb{Q}\}$, and the claim follows. $\square$

2.3 Gaussian Subspaces, Independence, and Conditional Expectation

We next recall two basic facts about Gaussian Hilbert spaces which will be used later in the prediction problem.

Proposition 2.9. *Let A and B be closed subspaces of a Gaussian Hilbert space G. Then A and B are independent if and only if they are orthogonal, i.e.*

$$E[ab] = 0 \qquad \text{for every } a \in A,\ b \in B.$$

Proof. Assume first that $A \perp B$. Take arbitrary finite families

$$a_1, \dots, a_m \in A, \qquad b_1, \dots, b_n \in B.$$

Since all these random variables belong to the same Gaussian Hilbert space, the vector

$$(a_1, \dots, a_m, b_1, \dots, b_n)$$

is jointly Gaussian and centered. Its characteristic function is therefore

$$E \exp\left(i \sum_{j=1}^{m} u_j a_j + i \sum_{k=1}^{n} v_k b_k \right) = \exp\left(-\frac{1}{2} \operatorname{Var}\left(\sum_{j=1}^{m} u_j a_j + \sum_{k=1}^{n} v_k b_k \right) \right).$$

Now orthogonality implies

$$E[a_j b_k] = 0 \qquad \text{for all } j, k.$$

Hence the variance splits:

$$\operatorname{Var}\left(\sum_{j=1}^{m} u_j a_j + \sum_{k=1}^{n} v_k b_k \right) = \operatorname{Var}\left(\sum_{j=1}^{m} u_j a_j \right) + \operatorname{Var}\left(\sum_{k=1}^{n} v_k b_k \right).$$

Therefore the characteristic function factors:

$$E \exp\left(i \sum_{j=1}^{m} u_j a_j + i \sum_{k=1}^{n} v_k b_k \right) = E \exp\left(i \sum_{j=1}^{m} u_j a_j \right) \, E \exp\left(i \sum_{k=1}^{n} v_k b_k \right).$$

Thus $(a_1, \dots, a_m)$ and $(b_1, \dots, b_n)$ are independent. Since the finite families were arbitrary, the σ-fields generated by A and B are independent, so A and B are independent.

Conversely, if A and B are independent, then for every $a \in A$ and $b \in B$,

$$E[ab] = E[a]E[b] = 0,$$

because all elements of G are centered. Hence $A \perp B$. □

Proposition 2.10. *Let G be the Gaussian Hilbert space generated by a mean-square continuous stationary Gaussian process, let $A \subset G$ be a closed subspace, and let $\mathcal{A} := \sigma(A)$ be the σ-field generated by A. Then for every $Y \in G$,*

$$P_A Y = E[Y \mid \mathcal{A}],$$

where P_A denotes the orthogonal projection from G onto A.

Proof. Since G is separable, we may choose a countable dense family $\{a_n\}_{n \geq 1} \subset A$. Then

$$\mathcal{A} = \sigma(a_1, a_2, \dots).$$

Fix $Y \in G$, and write

$$Y = P_A Y + (Y - P_A Y).$$

The term $P_A Y$ belongs to A, hence is $\mathcal{A}$-measurable.

Now $Y - P_A Y \perp A$, so in particular

$$E[(Y - P_A Y) a_n] = 0 \qquad \text{for every } n \geq 1.$$

By the previous proposition, $Y - P_A Y$ is independent of every finite collection $(a_1, \dots, a_n)$, hence independent of $\mathcal{A}$. Since it is centered,

$$E[Y - P_A Y \mid \mathcal{A}] = 0.$$

Therefore

$$E[Y \mid \mathcal{A}] = E[P_A Y \mid \mathcal{A}] + E[Y - P_A Y \mid \mathcal{A}] = P_A Y.$$

This proves the claim. $\square$

2.4 The Shift Group and the Spectral Representation

From now on, let $\{X(t)\}_{t\in\mathbb{R}}$ be a centered mean-square continuous stationary Gaussian process, and let

$$G = \overline{\operatorname{span}_{\mathbb{R}}\{X(t) : t \in \mathbb{R}\}}, \qquad G_{\mathbb{C}} = \overline{\operatorname{span}_{\mathbb{C}}\{X(t) : t \in \mathbb{R}\}}.$$

Stationarity gives a natural unitary action of time shifts on G.

Proposition 2.11. *For each $h \in \mathbb{R}$, the rule*

$$U_h X(t) := X(t+h), \qquad t \in \mathbb{R},$$

extends uniquely to a unitary operator $U_h : G \to G$. Moreover:

(1) $U_{h_1+h_2} = U_{h_1}U_{h_2}$ for all $h_1, h_2 \in \mathbb{R}$;

(2) $U_0 = \mathrm{Id}$;

(3) the family $\{U_h\}_{h\in\mathbb{R}}$ is strongly continuous on G, i.e.

$$\lim_{h\to 0} \|U_h Y - Y\|_G = 0 \qquad \text{for every } Y \in G.$$

The same statements hold on the complexification $G_{\mathbb{C}}$.

Proof. Define U_h first on finite linear combinations

$$Y = \sum_{j=1}^{n} c_j X(t_j)$$

by

$$U_h Y := \sum_{j=1}^{n} c_j X(t_j + h).$$

To see that U_h is isometric, take

$$Y = \sum_{j=1}^{n} c_j X(t_j), \qquad Z = \sum_{k=1}^{m} d_k X(s_k).$$

Then

$$\langle U_h Y, U_h Z\rangle_G = \sum_{j,k} c_j d_k \, E[X(t_j + h)X(s_k + h)].$$

By stationarity,

$$E[X(t_j + h)X(s_k + h)] = E[X(t_j)X(s_k)],$$

hence

$$\langle U_h Y, U_h Z\rangle_G = \langle Y, Z\rangle_G.$$

So U_h extends uniquely to an isometry on G. Since U_{-h} is its inverse, U_h is unitary.

The group identities are immediate on the generating family $\{X(t)\}$, hence on all of G.

It remains to prove strong continuity. First take a finite linear combination

$$Y = \sum_{j=1}^{n} c_j X(t_j).$$

Then

$$\|U_h Y - Y\|_G = \left\| \sum_{j=1}^{n} c_j \Big(X(t_j + h) - X(t_j) \Big) \right\|_G \le \sum_{j=1}^{n} |c_j| \, \|X(t_j + h) - X(t_j)\|_G.$$

Each term tends to 0 as $h \to 0$ by mean-square continuity, hence

$$\|U_h Y - Y\|_G \to 0.$$

Now let $Y \in G$ be arbitrary, and choose finite linear combinations $Y_n \to Y$ in G. Since U_h is unitary,

$$\|U_h Y - Y\|_G \le \|U_h(Y - Y_n)\|_G + \|U_h Y_n - Y_n\|_G + \|Y_n - Y\|_G = 2\|Y - Y_n\|_G + \|U_h Y_n - Y_n\|_G.$$

First choose n so that the first term is small, then let $h \to 0$. This proves strong continuity on G. The complexified statement is identical. □

We may now apply the spectral theorem for strongly continuous unitary groups.

Theorem 2.12 (Spectral representation / Bochner theorem)**.** *There exists a unique finite positive Borel measure μ on $\mathbb{R}$ such that*

$$Q(t) = \int_{\mathbb{R}} e^{it\lambda} \, \mu(\, d\lambda), \qquad t \in \mathbb{R}.$$

Moreover,

$$\mu(\mathbb{R}) = Q(0).$$

If the process is real-valued, then μ is symmetric:

$$\mu(B) = \mu(-B) \qquad \textit{for every Borel set } B \subset \mathbb{R}.$$

Proof. By the previous proposition, $\{U_h\}_{h\in\mathbb{R}}$ is a strongly continuous unitary group on the complex Hilbert space $G_{\mathbb{C}}$. By the spectral theorem for strongly continuous unitary groups, there exists a projection-valued measure $E(\cdot)$ on $\mathbb{R}$ such that

$$U_h = \int_{\mathbb{R}} e^{ih\lambda}\, E(\,d\lambda), \qquad h \in \mathbb{R}.$$

Set

$$X_0 := X(0) \in G \subset G_{\mathbb{C}},$$

and define a scalar measure μ by

$$\mu(B) := \langle E(B)X_0, X_0\rangle_{G_{\mathbb{C}}}, \qquad B \subset \mathbb{R} \text{ Borel}.$$

Since $E(B)$ is an orthogonal projection, μ is a finite positive measure, and

$$\mu(\mathbb{R}) = \langle E(\mathbb{R})X_0, X_0\rangle = \langle X_0, X_0\rangle = E[X(0)^2] = Q(0).$$

Now

$$Q(h) = E[X(h)X(0)] = \langle X(h), X(0)\rangle = \langle U_h X_0, X_0\rangle.$$

Using the spectral resolution of U_h,

$$\langle U_h X_0, X_0\rangle = \left\langle \int_{\mathbb{R}} e^{ih\lambda} E(\,d\lambda)X_0,\, X_0 \right\rangle = \int_{\mathbb{R}} e^{ih\lambda}\, \mu(\,d\lambda).$$

This gives the required representation.

To prove uniqueness, suppose also

$$Q(h) = \int_{\mathbb{R}} e^{ih\lambda}\, \nu(\,d\lambda) \qquad \text{for all } h \in \mathbb{R},$$

where ν is another finite positive measure. Then the finite signed measure $\sigma := \mu - \nu$ satisfies

$$\int_{\mathbb{R}} e^{ih\lambda}\, \sigma(\,d\lambda) = 0 \qquad \text{for all } h \in \mathbb{R}.$$

By the uniqueness theorem for Fourier transforms of finite measures, $\sigma = 0$. Hence $\mu = \nu$.

Finally, if the process is real-valued, then Q is real and even. Let μ^- be the reflected measure

$$\mu^-(B) := \mu(-B).$$

Then

$$\int_{\mathbb{R}} e^{ih\lambda}\, \mu^-(\,d\lambda) = \int_{\mathbb{R}} e^{-ih\lambda}\, \mu(\,d\lambda) = Q(-h) = Q(h) = \int_{\mathbb{R}} e^{ih\lambda}\, \mu(\,d\lambda).$$

By uniqueness, $\mu^- = \mu$. Thus μ is symmetric. $\square$

The finite positive measure μ is called the *spectral measure* of the process.

Remark 2.13. The spectral measure is the most natural spectral datum. If one prefers the classical Stieltjes notation, one may choose a right-continuous nondecreasing function Δ such that

$$\mu((a, b]) = \Delta(b) - \Delta(a),$$

and then write

$$\mu(\,d\lambda) = \,d\Delta(\lambda), \qquad Q(t) = \int_{\mathbb{R}} e^{it\lambda}\,\, d\Delta(\lambda).$$

What is essential is that Δ be a nondecreasing distribution function for a finite positive measure. There is no need to impose oddness on Δ.

We now identify the Gaussian Hilbert space generated by the process with an L^2-space over the spectral measure.

Theorem 2.14 (Spectral-space identification)**.** *The correspondence*

$$X(t) \longmapsto e^{it\lambda}$$

extends uniquely to a unitary isomorphism

$$W : G_{\mathbb{C}} \longrightarrow L^2(\mathbb{R}, \mu).$$

In particular,

$$\langle X(t), X(s)\rangle_{G_{\mathbb{C}}} = \int_{\mathbb{R}} e^{i(t-s)\lambda}\, \mu(\,d\lambda) = \langle e^{it\lambda}, e^{is\lambda}\rangle_{L^2(\mu)}.$$

Proof. Define W first on finite linear combinations by

$$W\left(\sum_{j=1}^{n} c_j X(t_j)\right) := \sum_{j=1}^{n} c_j e^{it_j\lambda}.$$

We claim that W is isometric. Indeed,

$$\begin{aligned}
\left\| \sum_{j=1}^{n} c_j X(t_j) \right\|_{G_{\mathbb{C}}}^2 &= \sum_{j,k=1}^{n} c_j \overline{c_k} \, \langle X(t_j), X(t_k) \rangle \\
&= \sum_{j,k=1}^{n} c_j \overline{c_k} \, Q(t_j - t_k) \\
&= \sum_{j,k=1}^{n} c_j \overline{c_k} \int_{\mathbb{R}} e^{i(t_j - t_k)\lambda} \, \mu(\, d\lambda) \\
&= \int_{\mathbb{R}} \left| \sum_{j=1}^{n} c_j e^{it_j\lambda} \right|^2 \, \mu(\, d\lambda) \\
&= \left\| \sum_{j=1}^{n} c_j e^{it_j\lambda} \right\|_{L^2(\mu)}^2 .
\end{aligned}$$

Hence W extends uniquely to an isometric embedding

$$W : G_{\mathbb{C}} \to L^2(\mathbb{R}, \mu).$$

It remains to prove surjectivity. Let $f \in L^2(\mu)$ be orthogonal to $W(G_{\mathbb{C}})$. Then in particular

$$\int_{\mathbb{R}} f(\lambda) e^{-it\lambda} \, \mu(\, d\lambda) = 0 \qquad \text{for every } t \in \mathbb{R},$$

because $e^{it\lambda} \in W(G_{\mathbb{C}})$ for every t.

Now define a finite complex measure ν by

$$\nu(B) := \int_B f(\lambda) \, \mu(\, d\lambda).$$

Since μ is finite and $f \in L^2(\mu)$, Cauchy–Schwarz yields

$$|\nu|(\mathbb{R}) \leq \|f\|_{L^2(\mu)} \, \mu(\mathbb{R})^{1/2} < \infty.$$

The Fourier transform of ν is

$$\widehat{\nu}(t) := \int_{\mathbb{R}} e^{-it\lambda} \, \nu(\, d\lambda) = \int_{\mathbb{R}} f(\lambda) e^{-it\lambda} \, \mu(\, d\lambda) = 0 \qquad \text{for all } t \in \mathbb{R}.$$

By uniqueness of Fourier transforms of finite measures, $\nu = 0$. Hence $f = 0$ μ-almost everywhere. Therefore $W(G_{\mathbb{C}})^{\perp} = \{0\}$, so the range of W is dense. Since the range is also closed (because W is an isometry), it is all of $L^2(\mathbb{R}, \mu)$. □

Proposition 2.15 (Intertwining relation)**.** *For every $t \in \mathbb{R}$ and every $Y \in G_{\mathbb{C}}$,*

$$W(U_t Y)(\lambda) = e^{it\lambda}(WY)(\lambda) \qquad \text{in } L^2(\mathbb{R}, \mu).$$

Proof. It is enough to verify the identity on the dense subspace of finite linear combinations of the variables $X(s)$. For a generator $X(s)$,

$$W(U_t X(s)) = W(X(s+t)) = e^{i(s+t)\lambda} = e^{it\lambda}e^{is\lambda} = e^{it\lambda}W(X(s)).$$

By linearity and density, the identity extends to all $Y \in G_{\mathbb{C}}$. □

2.5 Lebesgue Decomposition of the Spectral Measure

The decomposition of the spectral measure into absolutely continuous and singular parts is central in prediction theory.

Proposition 2.16. *Let μ be the spectral measure of the process. Then μ admits a unique Lebesgue decomposition*

$$\mu(\,d\lambda) = \mu_{\mathrm{ac}}(\,d\lambda) + \mu_{\mathrm{s}}(\,d\lambda) = \Delta_0(\lambda)\;d\lambda + \mu_{\mathrm{s}}(\,d\lambda),$$

where $\Delta_0 \in L^1(\mathbb{R})$, $\Delta_0 \geq 0$, and $\mu_{\mathrm{s}} \perp d\lambda$.

Consequently,

$$L^2(\mathbb{R}, \mu) = L^2(\mathbb{R}, \Delta_0(\lambda)\;d\lambda) \oplus L^2(\mathbb{R}, \mu_{\mathrm{s}}),$$

orthogonally.

Proof. The Lebesgue decomposition theorem gives

$$\mu = \mu_{\mathrm{ac}} + \mu_{\mathrm{s}}, \qquad \mu_{\mathrm{ac}} \ll d\lambda, \qquad \mu_{\mathrm{s}} \perp d\lambda.$$

Since μ is finite and positive, there exists $\Delta_0 \in L^1(\mathbb{R})$, $\Delta_0 \geq 0$, such that

$$\mu_{\mathrm{ac}}(\,d\lambda) = \Delta_0(\lambda)\;d\lambda.$$

The orthogonal decomposition of $L^2(\mu)$ then follows immediately from the fact that μ_{ac} and μ_{s} are mutually singular. □

Under the unitary map W, the previous decomposition induces an orthogonal decomposition of the complexified Gaussian Hilbert space:

$$G_{\mathbb{C}} = G_{\mathrm{ac}} \oplus G_{\mathrm{s}},$$

where

$$W(G_{\mathrm{ac}}) = L^2(\mathbb{R}, \Delta_0(\lambda)\ d\lambda), \qquad W(G_{\mathrm{s}}) = L^2(\mathbb{R}, \mu_{\mathrm{s}}).$$

This decomposition will play a decisive role in the prediction problem. Roughly speaking, the absolutely continuous part is the part on which Hardy-space factorization acts directly, while the singular part is responsible for the deterministic component of the process.

Remark 2.17. For later use, it is worth keeping in mind the chain of correspondences

$$\text{process } \{X(t)\} \quad \longleftrightarrow \quad \text{covariance } Q \quad \longleftrightarrow \quad \text{spectral measure } \mu \quad \longleftrightarrow \quad L^2(\mathbb{R}, \mu).$$

The prediction problem will be solved not directly in the original Gaussian Hilbert space, but after transporting it to the spectral space $L^2(\mathbb{R}, \mu)$, where the past becomes a closed span of exponentials and Hardy-space methods become available.

Chapter 3

The Kolmogorov–Wiener Prediction Problem

In this chapter we formulate the prediction problem of Kolmogorov and Wiener in the Gaussian Hilbert space generated by the process, then transport it to the spectral space $L^2(\mathbb{R}, \mu)$, where Hardy-space methods become available.

Throughout the chapter, $\{X(t)\}_{t\in\mathbb{R}}$ denotes a centered mean-square continuous stationary Gaussian process, Q its covariance function, μ its spectral measure, and

$$W : G_{\mathbb{C}} \longrightarrow L^2(\mathbb{R}, \mu)$$

the unitary spectral identification from Chapter 2, characterized by

$$W(X(t)) = e^{it\lambda}.$$

Remark 3.1. From now on, whenever a real closed subspace of G is viewed inside $G_{\mathbb{C}}$, we use the same notation for its complexification.

3.1 Formulation of the Prediction Problem

For $a \in \mathbb{R}$, define the closed past subspace in the Gaussian Hilbert space by

$$G_{(-\infty,a]} := \overline{\operatorname{span}_{\mathbb{R}}\{X(t) : t \leq a\}} \subset G.$$

The prediction problem at time $T > 0$ from the whole past $(-\infty, 0]$ is to determine the orthogonal projection of $X(T)$ onto $G_{(-\infty,0]}$.

Definition 3.2. For $T > 0$, the *best linear predictor* of $X(T)$ from the past $(-\infty, 0]$ is

$$\widehat{X}(T) := P_{G_{(-\infty,0]}} X(T).$$

The corresponding *prediction error* is

$$\varepsilon(T) := X(T) - \widehat{X}(T).$$

Since G is a Gaussian Hilbert space, orthogonal projection and conditional expectation coincide.

Proposition 3.3. *For every $T > 0$,*

$$\widehat{X}(T) = E[X(T) \mid \sigma\{X(t) : t \leq 0\}] .$$

Moreover, $\varepsilon(T)$ is orthogonal to $G_{(-\infty,0]}$, hence independent of the past σ-field $\sigma\{X(t) : t \leq 0\}$.

Proof. This is an immediate consequence of the Gaussian-Hilbert-space projection theorem proved in Chapter 2. The orthogonality of $\varepsilon(T)$ to $G_{(-\infty,0]}$ follows from the defining property of orthogonal projection, and independence follows because orthogonality is equivalent to independence for Gaussian subspaces. □

Consequently, the conditional law of $X(T)$ given the past is Gaussian with mean $\widehat{X}(T)$ and variance $\|\varepsilon(T)\|_G^2$:

$$P(X(T) \leq x \mid \sigma\{X(t) : t \leq 0\}) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi \|\varepsilon(T)\|_G^2}} \exp\left(-\frac{(y - \widehat{X}(T))^2}{2\|\varepsilon(T)\|_G^2} \right) \, dy.$$

Thus the whole prediction problem reduces to computing $\widehat{X}(T)$ and $\|\varepsilon(T)\|_G^2$.

3.2 Transport to the Spectral Space

Under the unitary map W, the past subspaces become closed spans of exponentials.

Definition 3.4. For $a \in \mathbb{R}$, define

$$Z_{(-\infty,a]} := \overline{\operatorname{span}_{\mathbb{C}}\{e^{it\lambda} : t \leq a\}} \subset L^2(\mathbb{R}, \mu).$$

Define also the *remote past* by

$$Z_{-\infty} := \bigcap_{a<0} Z_{(-\infty,a]}.$$

Proposition 3.5. *For every $a \in \mathbb{R}$,*

$$W\Big(G_{(-\infty,a]}\Big) = Z_{(-\infty,a]}.$$

In particular, if $\Pi_{(-\infty,a]}$ denotes the orthogonal projection in $L^2(\mathbb{R},\mu)$ onto $Z_{(-\infty,a]}$, then

$$W\Big(P_{G_{(-\infty,a]}} Y\Big) = \Pi_{(-\infty,a]}(WY), \qquad Y \in G_{\mathbb{C}}.$$

Proof. By definition of W,

$$W(X(t)) = e^{it\lambda}.$$

Therefore

$$W(\operatorname{span}_{\mathbb{C}}\{X(t) : t \leq a\}) = \operatorname{span}_{\mathbb{C}}\{e^{it\lambda} : t \leq a\}.$$

Since W is unitary, it preserves closures, so

$$W\Big(G_{(-\infty,a]}\Big) = Z_{(-\infty,a]}.$$

The projection identity follows immediately from unitarity. □

Corollary 3.6.

$$W(G_{-\infty}) = Z_{-\infty}, \qquad G_{-\infty} := \bigcap_{a<0} G_{(-\infty,a]}.$$

Proof. Since W is bijective and linear,

$$W\left(\bigcap_{a<0} G_{(-\infty,a]}\right) = \bigcap_{a<0} W(G_{(-\infty,a]}) = \bigcap_{a<0} Z_{(-\infty,a]} = Z_{-\infty}.$$

□

Hence the best linear predictor of $X(T)$ corresponds to the orthogonal projection of $e^{iT\lambda}$ onto $Z_{(-\infty,0]}$:

$$W(\widehat{X}(T)) = \Pi_{(-\infty,0]}(e^{iT\lambda}),$$

and

$$\|\varepsilon(T)\|_G^2 = \left\|e^{iT\lambda} - \Pi_{(-\infty,0]}(e^{iT\lambda})\right\|_{L^2(\mu)}^2.$$

Remark 3.7. The prediction problem is therefore equivalent to the following geometric problem in $L^2(\mathbb{R},\mu)$:

Given $T > 0$, project $e^{iT\lambda}$ onto the closed span of $\{e^{it\lambda} : t \leq 0\}$.

This is precisely the form in which Hardy-space factorization can be applied.

3.3 Determinism, the Remote Past, and Szegő's Alternative

We next record the classical dichotomy governing whether the future is already determined by the past.

Definition 3.8. The process is called *deterministic* if

$$X(0) \in G_{(-\infty,0]},$$

equivalently

$$e^{i0\lambda} = 1 \in Z_{(-\infty,0]}.$$

Otherwise it is called *nondeterministic.*

Definition 3.9. The process is called *purely nondeterministic* if its remote past is trivial:

$$G_{-\infty} := \bigcap_{a<0} G_{(-\infty,a]} = \{0\}.$$

Equivalently,

$$Z_{-\infty} = \{0\}.$$

Remark 3.10. In probabilistic terms, the remote past σ-field is

$$\mathcal{F}_{-\infty} := \bigcap_{a<0} \sigma\{X(t) : t \leq a\}.$$

The Gaussian Hilbert space generated by $\mathcal{F}_{-\infty}$ is precisely $G_{-\infty}$. Thus pure nondeterminism means that the remote past contains no nontrivial Gaussian information.

By Chapter 2, the spectral measure admits the Lebesgue decomposition

$$\mu(\,d\lambda) = \Delta_0(\lambda)\ d\lambda + \mu_{\mathrm{s}}(\,d\lambda),$$

where $\Delta_0 \in L^1(\mathbb{R})$, $\Delta_0 \geq 0$, and $\mu_{\mathrm{s}} \perp d\lambda$.

The fundamental theorem is the following continuous-time version of Szegő's alternative.

Theorem 3.11 (Szegő's alternative; classical form)**.** *Exactly one of the following two cases occurs:*

(I) *If*

$$\int_{\mathbb{R}} \frac{\log \Delta_0(\lambda)}{1+\lambda^2}\ d\lambda > -\infty,$$

then the process is nondeterministic and

$$Z_{-\infty} = L^2(\mathbb{R}, \mu_{\mathrm{s}}).$$

(II) *If*

$$\int_{\mathbb{R}} \frac{\log \Delta_0(\lambda)}{1+\lambda^2}\, d\lambda = -\infty,$$

then the process is deterministic; equivalently,

$$Z_{(-\infty,0]} = L^2(\mathbb{R}, \mu).$$

Proof. This is a classical theorem; see Dym–McKean [3, Chapter 3], based on the F. & M. Riesz theorem and outer-function theory developed in Chapter 1. □

The theorem has the following immediate corollary, which is the form most relevant for prediction theory.

Corollary 3.12. *The process is purely nondeterministic if and only if*

$$\mu_{\mathrm{s}} = 0 \qquad \textit{and} \qquad \int_{\mathbb{R}} \frac{\log \Delta_0(\lambda)}{1+\lambda^2}\, d\lambda > -\infty.$$

Proof. If the process is purely nondeterministic, then $Z_{-\infty} = \{0\}$. By Theorem 3.11, case (II) is impossible, so case (I) must hold. Hence

$$Z_{-\infty} = L^2(\mathbb{R}, \mu_{\mathrm{s}}) = \{0\},$$

which forces $\mu_{\mathrm{s}} = 0$, and the logarithmic integral is finite.

Conversely, if $\mu_{\mathrm{s}} = 0$ and the logarithmic integral is finite, then by Theorem 3.11,

$$Z_{-\infty} = L^2(\mathbb{R}, \mu_{\mathrm{s}}) = \{0\}.$$

Thus the remote past is trivial, so the process is purely nondeterministic. □

Remark 3.13. Theorem 3.11 separates two different obstructions to fresh innovation:

(1) the absolutely continuous part may itself be too degenerate, as detected by the divergence of

$$\int \frac{\log \Delta_0}{1+\lambda^2};$$

(2) even when the above integral is finite, the singular part survives unchanged inside the remote past.

Thus, for continuous-time Gaussian prediction, the singular spectral component is exactly the deterministic component that remains in the remote past.

3.4 Prediction in the Purely Nondeterministic Case

We now assume that the process is purely nondeterministic. By Corollary 3.12,

$$\mu(\,d\lambda) = \Delta_0(\lambda)\; d\lambda, \qquad \int_{\mathbb{R}} \frac{\log \Delta_0(\lambda)}{1+\lambda^2}\; d\lambda > -\infty.$$

By the outer-function theorem from Chapter 1, there exists an outer function $h \in H^2(\mathbb{C}_+)$ such that

$$|h_{0+}(\lambda)|^2 = \Delta_0(\lambda) \qquad \text{for a.e. } \lambda \in \mathbb{R}.$$

To simplify notation, we write $h(\lambda)$ for the L^2-boundary value $h_{0+}(\lambda)$.

Let

$$h(\lambda) = \int_0^\infty e^{i\lambda t} g(t)\; dt, \qquad g \in L^2(0,\infty),$$

as given by the Paley–Wiener theorem.

Lemma 3.14. *Multiplication by $\overline{h(\lambda)}$ defines a unitary operator*

$$M_{\overline{h}} : L^2(\mathbb{R},\mu) \to L^2(\mathbb{R},\, d\lambda), \qquad (M_{\overline{h}}f)(\lambda) := \overline{h(\lambda)}\, f(\lambda).$$

Moreover,

$$M_{\overline{h}}\Big(Z_{(-\infty,0]}\Big) = H^2(\mathbb{C}_-)$$

as a closed subspace of $L^2(\mathbb{R},\, d\lambda)$.

Proof. Since $\mu(\,d\lambda) = |h(\lambda)|^2\; d\lambda$, we have

$$\|M_{\overline{h}}f\|^2_{L^2(\,d\lambda)} = \int_{\mathbb{R}} |f(\lambda)|^2 |h(\lambda)|^2\; d\lambda = \|f\|^2_{L^2(\mu)}.$$

Thus $M_{\overline{h}}$ is an isometry.

Since

$$\int_{\mathbb{R}} \frac{\log \Delta_0(\lambda)}{1+\lambda^2}\; d\lambda > -\infty,$$

we necessarily have $\Delta_0(\lambda) > 0$ for almost every λ. Hence $h(\lambda) \neq 0$ for almost every λ, and division by $\overline{h(\lambda)}$ is legitimate almost everywhere. Therefore $M_{\overline{h}}$ is onto because any

$u \in L^2(\,d\lambda)$ can be written as $u = \overline{h}\,f$ with $f = u/\overline{h} \in L^2(\mu)$.

Now

$$M_{\overline{h}}(e^{it\lambda}) = e^{it\lambda}\overline{h(\lambda)}.$$

Since $h \in H^2(\mathbb{C}_+)$ is outer, its reflected boundary value $\overline{h}$ is outer in $H^2(\mathbb{C}_-)$. Therefore, by the lower-half-plane version of the test for outer functions from Chapter 1, the closed linear span of

$$\{e^{it\lambda}\overline{h(\lambda)} : t \leq 0\}$$

is all of $H^2(\mathbb{C}_-)$. Taking closures gives

$$M_{\overline{h}}\Big(Z_{(-\infty,0]}\Big) = H^2(\mathbb{C}_-).$$

□

Let P_- and P_+ denote the orthogonal projections of $L^2(\mathbb{R},\,d\lambda)$ onto $H^2(\mathbb{C}_-)$ and $H^2(\mathbb{C}_+)$, respectively.

Theorem 3.15 (Explicit prediction formula in the purely nondeterministic case)**.** *Assume the process is purely nondeterministic. Then for every $T > 0$,*

$$\Pi_{(-\infty,0]}(e^{iT\lambda}) = \frac{1}{\overline{h(\lambda)}}\,P_-\Big(e^{iT\lambda}\overline{h(\lambda)}\Big) \qquad \text{in } L^2(\mathbb{R},\mu),$$

and

$$e^{iT\lambda} - \Pi_{(-\infty,0]}(e^{iT\lambda}) = \frac{1}{\overline{h(\lambda)}}\,P_+\Big(e^{iT\lambda}\overline{h(\lambda)}\Big).$$

More explicitly,

$$\Pi_{(-\infty,0]}(e^{iT\lambda}) = \frac{e^{iT\lambda}}{\overline{h(\lambda)}} \int_T^\infty e^{-i\lambda t}\,\overline{g(t)}\,dt, \tag{3.1}$$

and

$$e^{iT\lambda} - \Pi_{(-\infty,0]}(e^{iT\lambda}) = \frac{e^{iT\lambda}}{\overline{h(\lambda)}} \int_0^T e^{-i\lambda t}\,\overline{g(t)}\,dt. \tag{3.2}$$

Consequently,

$$\Big\|e^{iT\lambda} - \Pi_{(-\infty,0]}(e^{iT\lambda})\Big\|^2_{L^2(\mu)} = 2\pi \int_0^T |g(t)|^2\,dt. \tag{3.3}$$

Proof. By Lemma 3.14, $M_{\overline{h}}$ is unitary and maps $Z_{(-\infty,0]}$ onto $H^2(\mathbb{C}_-)$. Therefore the orthogonal projection in $L^2(\mu)$ onto $Z_{(-\infty,0]}$ is transported by $M_{\overline{h}}$ to the orthogonal projection in $L^2(\,d\lambda)$ onto $H^2(\mathbb{C}_-)$. Hence

$$M_{\overline{h}}\Big(\Pi_{(-\infty,0]}(e^{iT\lambda})\Big) = P_-\Big(e^{iT\lambda}\overline{h(\lambda)}\Big),$$

which yields

$$\Pi_{(-\infty,0]}(e^{iT\lambda}) = \frac{1}{\overline{h(\lambda)}} P_-\Big(e^{iT\lambda}\overline{h(\lambda)}\Big).$$

Since $P_+ + P_- = \mathrm{Id}$ on $L^2(\,d\lambda)$, the second displayed identity follows as well.

It remains to compute $P_-(e^{iT\lambda}\overline{h})$ and $P_+(e^{iT\lambda}\overline{h})$ explicitly. By Paley–Wiener,

$$h(\lambda) = \int_0^\infty e^{i\lambda t} g(t)\, dt,$$

hence

$$\overline{h(\lambda)} = \int_0^\infty e^{-i\lambda t}\, \overline{g(t)}\, dt = \int_{-\infty}^0 e^{i\lambda s}\, \overline{g(-s)}\, ds.$$

Therefore

$$e^{iT\lambda}\overline{h(\lambda)} = \int_{-\infty}^0 e^{i\lambda(s+T)}\, \overline{g(-s)}\, ds.$$

Make the change of variables $y = s + T$. Since $s \le 0$, we obtain

$$e^{iT\lambda}\overline{h(\lambda)} = \int_{-\infty}^T e^{i\lambda y}\, \overline{g(T-y)}\, dy.$$

Split the integral at $y = 0$:

$$e^{iT\lambda}\overline{h(\lambda)} = \int_{-\infty}^0 e^{i\lambda y}\, \overline{g(T-y)}\, dy + \int_0^T e^{i\lambda y}\, \overline{g(T-y)}\, dy.$$

The first term is the boundary value of a function in $H^2(\mathbb{C}_-)$, because its inverse Fourier transform is supported on $(-\infty, 0]$. The second term is the boundary value of a function in $H^2(\mathbb{C}_+)$, because its inverse Fourier transform is supported on $[0, \infty)$. Hence

$$P_-\Big(e^{iT\lambda}\overline{h(\lambda)}\Big) = \int_{-\infty}^0 e^{i\lambda y}\, \overline{g(T-y)}\, dy,$$

$$P_+\Big(e^{iT\lambda}\overline{h(\lambda)}\Big) = \int_0^T e^{i\lambda y}\, \overline{g(T-y)}\, dy.$$

Now change variables $t = T - y$. For the P_--term, $y \in (-\infty, 0]$ corresponds to $t \in [T, \infty)$, so

$$P_-\Big(e^{iT\lambda}\overline{h(\lambda)}\Big) = e^{iT\lambda} \int_T^\infty e^{-i\lambda t}\, \overline{g(t)}\, dt.$$

Likewise, for the P_+-term, $y \in [0, T]$ corresponds to $t \in [0, T]$, so

$$P_+\Big(e^{iT\lambda}\overline{h(\lambda)}\Big) = e^{iT\lambda} \int_0^T e^{-i\lambda t}\, \overline{g(t)}\, dt.$$

Substituting these into the two projection formulas yields (3.1) and (3.2).

Finally, since $M_{\overline{h}}$ is unitary,

$$\left\| e^{iT\lambda} - \Pi_{(-\infty,0]}(e^{iT\lambda}) \right\|^2_{L^2(\mu)} = \left\| P_+\Big(e^{iT\lambda}\overline{h(\lambda)}\Big) \right\|^2_{L^2(\,d\lambda)}.$$

Using the explicit formula for P_+ and Plancherel,

$$\left\| P_+\Big(e^{iT\lambda}\overline{h(\lambda)}\Big) \right\|^2_{L^2(\,d\lambda)} = 2\pi \int_0^T |g(t)|^2 \, dt.$$

This proves (3.3). □

Transporting the result back to the Gaussian Hilbert space gives the desired prediction theorem for the process itself.

Corollary 3.16 (Prediction formula for the process)**.** *Under the assumptions of Theorem 3.15, the best linear predictor of $X(T)$ from the past $(-\infty, 0]$ is*

$$\widehat{X}(T) = W^{-1}\left(\frac{e^{iT\lambda}}{\overline{h(\lambda)}} \int_T^\infty e^{-i\lambda t}\, \overline{g(t)} \, dt \right),$$

and the prediction error variance is

$$E\Big[(X(T) - \widehat{X}(T))^2\Big] = 2\pi \int_0^T |g(t)|^2 \, dt.$$

Proof. This follows immediately from Theorem 3.15 and the unitarity of W. □

Remark 3.17. If in addition Δ_0 is even, one may choose the outer factor h so that

$$h(-\lambda) = \overline{h(\lambda)} \qquad \text{for a.e. } \lambda,$$

in which case the inverse Fourier transform g may be chosen real-valued. Then the formulas above simplify by removing the complex conjugates on g.

3.5 Finite-Horizon Innovation Space

The previous theorem has a useful geometric consequence: under pure nondeterminism, the new information created between times 0 and T is modeled exactly by $L^2(0, T)$.

Definition 3.18. For $T > 0$, define the finite-horizon innovation space

$$\mathcal{I}_T := Z_{(-\infty,T]} \ominus Z_{(-\infty,0]} \subset L^2(\mathbb{R}, \mu).$$

Proposition 3.19. *Assume the process is purely nondeterministic. Then the map*

$$\mathcal{I}_T \ni f \longmapsto \left(M_{\overline{h}} f\right)^{\vee}\Big|_{(0,T)}$$

is a unitary isomorphism from $\mathcal{I}_T$ onto $L^2(0,T)$. Equivalently,

$$\mathcal{I}_T \cong L^2(0,T).$$

Proof. By Lemma 3.14, $M_{\overline{h}}$ is unitary from $L^2(\mu)$ onto $L^2(\,d\lambda)$, and

$$M_{\overline{h}}\Big(Z_{(-\infty,0]}\Big) = H^2(\mathbb{C}_-).$$

Likewise,

$$M_{\overline{h}}\Big(Z_{(-\infty,T]}\Big) = e^{iT\lambda} H^2(\mathbb{C}_-),$$

because multiplying the generators $e^{it\lambda}$ with $t \le T$ by $\overline{h}$ gives the closed span of

$$\{e^{it\lambda}\overline{h}(\lambda) : t \le T\} = e^{iT\lambda}\{e^{is\lambda}\overline{h}(\lambda) : s \le 0\}.$$

Hence

$$M_{\overline{h}}(\mathcal{I}_T) = e^{iT\lambda} H^2(\mathbb{C}_-) \ominus H^2(\mathbb{C}_-).$$

Now pass to inverse Fourier transforms. Under inverse Fourier transform,

$$H^2(\mathbb{C}_-) \longleftrightarrow L^2((-\infty,0]),$$

and multiplication by $e^{iT\lambda}$ corresponds to translation of the inverse Fourier transform by T. Therefore

$$e^{iT\lambda} H^2(\mathbb{C}_-) \longleftrightarrow L^2((-\infty,T]).$$

Thus

$$e^{iT\lambda} H^2(\mathbb{C}_-) \ominus H^2(\mathbb{C}_-) \longleftrightarrow L^2(0,T).$$

Since all identifications involved are unitary, the conclusion follows. □

Remark 3.20. Proposition 3.19 is the continuous-time innovation picture behind the prediction formula. It says that, after spectral factorization by the outer function h, the genuinely new information created over the interval $(0,T]$ is represented exactly by the standard Hilbert space $L^2(0,T)$.

3.6 The General Finite-Log Case and the Role of the Singular Component

We finally return to the case where

$$\int_{\mathbb{R}} \frac{\log \Delta_0(\lambda)}{1+\lambda^2}\, d\lambda > -\infty,$$

but μ_{s} need not vanish. Then Theorem 3.11 gives

$$Z_{-\infty} = L^2(\mathbb{R}, \mu_{\mathrm{s}}).$$

Thus the singular spectral subspace lies entirely in the remote past.

Let

$$L^2(\mathbb{R}, \mu) = L^2(\mathbb{R}, \Delta_0(\lambda)\, d\lambda) \oplus L^2(\mathbb{R}, \mu_{\mathrm{s}}),$$

and let Π_{ac}, Π_{s} denote the corresponding orthogonal projections.

Proposition 3.21. *Assume*

$$\int_{\mathbb{R}} \frac{\log \Delta_0(\lambda)}{1+\lambda^2}\, d\lambda > -\infty.$$

Then for every $T > 0$,

$$\Pi_{\mathrm{s}}\Pi_{(-\infty,0]}(e^{iT\lambda}) = \Pi_{\mathrm{s}}(e^{iT\lambda}),$$

and

$$\Pi_{\mathrm{ac}}\Pi_{(-\infty,0]}(e^{iT\lambda}) = \frac{1}{\overline{h(\lambda)}}\, P_-\Big(e^{iT\lambda}\overline{h(\lambda)}\Big) \qquad \text{in } L^2(\mathbb{R}, \Delta_0(\lambda)\, d\lambda),$$

where h is any outer factor satisfying $|h|^2 = \Delta_0$. Consequently,

$$\Big\|e^{iT\lambda} - \Pi_{(-\infty,0]}(e^{iT\lambda})\Big\|^2_{L^2(\mu)} = 2\pi \int_0^T |g(t)|^2\, dt,$$

with $h(\lambda) = \int_0^\infty e^{i\lambda t} g(t)\, dt$.

Proof. Since $Z_{-\infty} = L^2(\mathbb{R}, \mu_{\mathrm{s}}) \subset Z_{(-\infty,0]}$, every vector in the singular spectral subspace already belongs to the past. Therefore

$$\Pi_{\mathrm{s}}\Pi_{(-\infty,0]}(e^{iT\lambda}) = \Pi_{\mathrm{s}}(e^{iT\lambda}).$$

On the absolutely continuous subspace $L^2(\mathbb{R}, \Delta_0\, d\lambda)$, the same argument as in the purely nondeterministic case applies verbatim, because there the measure is exactly $\Delta_0(\lambda)\ d\lambda$. Hence

$$\Pi_{\mathrm{ac}}\Pi_{(-\infty,0]}(e^{iT\lambda}) = \frac{1}{\overline{h(\lambda)}}\, P_{-}\Big(e^{iT\lambda}\overline{h(\lambda)}\Big).$$

Finally, the prediction error belongs entirely to the absolutely continuous component, because the singular component has already been projected away exactly. Therefore its norm is the same as in the purely nondeterministic case:

$$\Big\| e^{iT\lambda} - \Pi_{(-\infty,0]}(e^{iT\lambda}) \Big\|^2_{L^2(\mu)} = 2\pi \int_0^T |g(t)|^2 \ dt.$$

□

Remark 3.22. Proposition 3.21 shows that the singular spectral component contributes no finite-horizon innovation variance. It belongs entirely to the deterministic/remote-past part of the process. The nontrivial prediction error is governed exactly by the absolutely continuous spectral density through the outer factor h.

Chapter 4

Research Frontiers and Open Problems

This chapter places the present thesis within the broader research line of Kolmogorov–Wiener prediction theory and proposes several directions for further work. The point of departure is an honest one: the core scalar theory developed in the previous chapters belongs to the classical part of the subject. Its value lies in clarifying the analytic, spectral, and probabilistic structure of the theory; the natural question is therefore how one should move beyond it.

The discussion below has two goals. First, it reviews the current status of the research line and identifies where the frontier has actually moved. Second, it formulates several concentrated open problems that remain close in spirit to the present thesis but go materially beyond the classical scalar framework.

4.1 Where the Classical Theory Stands

The scalar continuous-time prediction problem for centered stationary Gaussian processes, in its basic whole-past and finite-past forms, is classical. In particular, the problems of predicting from the whole past and from a finite section of the past have long been solved, and recent work has often focused not on discovering new formulas in the classical scalar setting, but rather on giving shorter proofs, clearer probabilistic interpretations, or more efficient computational viewpoints; see Bingham [?].

Accordingly, the present thesis should be understood as a study of the classical architecture:

$$\text{spectral representation} \longrightarrow \text{Hardy-space factorization} \longrightarrow \text{orthogonal projection and prediction.}$$

This architecture remains fundamental, but the genuinely open and promising directions now arise when one modifies at least one of its ingredients: the state space, the spectral object, the geometry of the available observations, or the asymptotic regime.

4.2 Current Frontiers Along the Prediction-Theory Line

The recent literature suggests that the frontier has moved mainly along the following axes.

4.2.1 Infinite-dimensional and operator-valued prediction

A first major development is the move from scalar processes to Hilbert-space-valued processes and operator-valued spectral theory. Bingham's survey on stationary functional time series [?] makes clear that prediction in infinitely many dimensions is far from closed, and explicitly points to a range of open questions surrounding infinite-dimensional analogues of classical objects such as Verblunsky's theorem, Szegő's theorem, the Szegő alternative, and related factorization theory.

A particularly important recent step is the work of Missaoui and Bingham [?], which extends parts of classical Szegő theory to the infinite-dimensional operator-valued setting. This does not close the subject; rather, it shows that the operator-valued route is both mathematically meaningful and technically viable. From the perspective of the present thesis, this is perhaps the most natural upgrade of the scalar theory.

4.2.2 Multivariate prediction, matrix factorization, and causality

A second frontier is multivariate prediction and matrix-valued spectral factorization. In this setting, the scalar outer factor is replaced by a matrix spectral factor, and one is led naturally to problems involving matrix Hardy spaces, matrix Szegő conditions, and matrix-valued prediction error formulas.

Two recent developments are especially relevant. First, Ephremidze [?] extended the Janashia–Lagvilava matrix spectral factorization algorithm from the unit circle to the real line and explicitly observed that the algorithm can be used directly for continuous-time models. Second, in [?], stationary processes, Wiener–Granger causality, and matrix spectral factorization are treated together, with explicit attention to prediction errors and future large-scale applications. These works indicate that the multivariate line is not only open, but also computationally and conceptually active.

4.2.3 Prediction from incomplete past and geometric observation sets

A third frontier arises when one changes the geometry of the observed past. In the classical one-parameter theory, the available information is the whole half-line $(-\infty, 0]$. Once one replaces this by a more complicated observation set—for example by deleting finitely many observations, by allowing gaps, or by moving to a random field indexed by $\mathbb{Z}^2$ or $\mathbb{R}^2$—the problem quickly leaves the standard textbook framework.

A representative result is Cheng's work on stationary Gaussian random fields with incomplete quarterplane past [**?**]. There, the solution already requires strongly outer factorization and a refined duality argument, and the theory extends beyond the quarterplane to a broader class of parameter sets. This suggests that the next genuinely new layer of prediction theory is not merely "more formulas", but a systematic understanding of how the geometry of the available observations interacts with factorization and projection.

4.2.4 Finite predictor asymptotics, long memory, and critical spectra

A fourth active line concerns the asymptotic behavior of finite predictors, especially in the presence of long memory or near-singular spectral behavior. This remains much closer to the classical core, but it is not finished. Recent work of Chigansky and Kleptsyna [**?**] develops a new analytic approach to the asymptotic analysis of the finite predictor for fractional Gaussian noise and obtains exact asymptotics for the relative prediction error and partial correlation coefficients. This is significant because the scalar "closed-form theory" for whole-past prediction does not automatically settle the asymptotic behavior of finite-window prediction in critical regimes.

In a broader modern direction, one also sees the emergence of locally stationary functional prediction theory; see Cui and Zhou [**?**]. That direction is important, but it is technically broader and less tightly tied to the current thesis than the four concentrated lines above. For that reason, it is better viewed here as a second-wave extension rather than as the most immediate sequel to the present work.

4.3 Open Problem I: Operator-Valued Continuous-Time Kolmogorov–Wiener Theory

Motivation and current status

Among all possible continuations of the present thesis, the most structurally natural one is to replace scalar-valued processes by processes taking values in a Hilbert space H, or equivalently to replace scalar spectral measures by operator-valued ones. The scalar architecture developed in the previous chapters strongly suggests what the operator-valued analogue should look like:

$$\text{operator-valued spectral representation} \longrightarrow \text{operator-valued Hardy / Szegő theory} \longrightarrow \text{operator-valued}$$

At present, however, the operator-valued theory is only partially in place. The work of Missaoui and Bingham [?] shows that some pieces of infinite-dimensional Szegő theory can indeed be pushed through, but a complete continuous-time Kolmogorov–Wiener package in this setting does not seem to exist in the literature in a form parallel to the scalar theory developed in this thesis.

Problem statement

Construct a continuous-time operator-valued prediction theory for centered mean-square continuous stationary Gaussian processes $X(t)$ with values in a separable Hilbert space H. The goal is to prove analogues of the following scalar results:

(a) a spectral-space identification with an L^2-space over an operator-valued spectral measure;

(b) an operator-valued Szegő criterion distinguishing nondeterministic and deterministic behavior;

(c) a description of the remote past and pure nondeterminism;

(d) an explicit whole-past predictor formula under an appropriate operator-valued outer-factor assumption;

(e) an explicit formula for the prediction error covariance operator.

Why this problem is meaningful

This problem is not an artificial generalization. It is the canonical infinite-dimensional continuation of the scalar theory, and it connects prediction theory with functional time series, operator-valued Hardy theory, and infinite-dimensional spectral analysis. Solving even a restricted version of it would produce genuinely new mathematics while staying very close to the core architecture of the present thesis.

A concentrated first target

A realistic first target is not the fully general case, but the following restricted theorem program:

> Assume that the operator-valued spectral measure is absolutely continuous with density $F(\lambda)$, where $F(\lambda)$ is a bounded positive invertible operator on H for almost every λ, and where an operator-valued Szegő condition holds in a form strong enough to guarantee operator-valued outer factorization. Under these assumptions, derive an explicit whole-past prediction formula and identify the error covariance operator.

This would already amount to a publishable step, because it would create the direct operator-valued analogue of the classical explicit prediction formula proved in Chapter 3.

Possible route

The scalar thesis suggests a concrete route:

(1) build the operator-valued spectral representation and the analogue of the unitary map W;

(2) define the correct operator-valued Hardy spaces on the half-plane;

(3) prove a version of the multiplication-isometry lemma with an operator-valued outer factor;

(4) reduce prediction to an operator-valued Hardy projection problem;

(5) recover the predictor and error covariance operator by transporting back to the Gaussian Hilbert space.

The most delicate steps are likely to be the outer-factor theory and the operator-valued Szegő condition. Those are precisely the points where the recent infinite-dimensional literature becomes relevant.

4.4 Open Problem II: Multivariate Continuous-Time Prediction and Real-Line Matrix Spectral Factorization

Motivation and current status

The finite-dimensional vector-valued case lies between the scalar theory and the full operator-valued theory. It is therefore both mathematically natural and more accessible technically. Recent matrix spectral factorization work on the real line [**?**] and the renewed connection with Wiener–Granger causality and prediction errors [**?**] make this an especially promising line for a first research paper.

What is attractive here is that the problem is both theoretically substantial and computationally meaningful. One is not only asking for a matrix-valued analogue of the classical predictor formula, but also for a theory robust enough to support perturbation bounds and numerical factorization.

Problem statement

Let $X(t)$ be a d-dimensional centered mean-square continuous stationary Gaussian process with matrix-valued spectral density $S(\lambda)$. Develop a continuous-time real-line version of the Kolmogorov–Wiener whole-past prediction theory in which:

(a) the scalar outer factor h is replaced by a matrix spectral factor H satisfying

$$S(\lambda) = H(\lambda)^* H(\lambda);$$

(b) the projection formula from Chapter 3 is replaced by an explicit matrix formula for the best linear predictor;

(c) the scalar error variance becomes a prediction error covariance matrix;

(d) one proves perturbation or stability bounds for the predictor and the error covariance under perturbations of S.

Why this problem is meaningful

This problem sits at an ideal intersection of classical theory, modern computation, and applications. It is close enough to the scalar thesis to be attacked systematically, but far enough from the completed classical line to contain genuine new work. It also interfaces naturally with causality, control, and high-dimensional time series.

A concentrated first target

The most focused first theorem package would be the following:

(i) assume $S \in L^1(\mathbb{R}; \mathbb{C}^{d\times d})$ is positive definite almost everywhere and satisfies a matrix Szegő condition strong enough to guarantee a matrix outer factor on the real line;

(ii) derive a continuous-time whole-past prediction formula on $L^2(\mathbb{R}, S(\lambda)\, d\lambda)$;

(iii) prove a quantitative bound of the schematic form

$$\|\widehat{X}_S(T) - \widehat{X}_{\widetilde{S}}(T)\| \le C\, \Phi\Big(\|S - \widetilde{S}\|\Big)$$

for a natural matrix norm.

Even a robust theorem of this sort in the finite-dimensional case would already be interesting and publishable.

Possible route

The scalar proof in Chapter 3 suggests the following route:

(1) formulate the spectral space as $L^2(\mathbb{R}, S(\lambda)\, d\lambda)$;

(2) transport the past subspace to a matrix Hardy subspace by multiplication with a matrix spectral factor;

(3) identify the predictor as a matrix Hardy projection;

(4) derive the predictor covariance and prediction error covariance from that factorization;

(5) use recent real-line factorization algorithms and estimates as the input for perturbation theory.

This problem is especially well suited for a first concentrated paper because it combines a clean theorem target with a realistic path to proof.

4.5 Open Problem III: Prediction from Incomplete Past and Geometric Observation Sets

Motivation and current status

The geometry of the observation set is one of the most natural ways to leave the classical setting. In one dimension, the present thesis treats the idealized case in which the entire half-line $(-\infty, 0]$ is observed. In practice and in many related mathematical problems, one instead knows the process only on a more complicated set: a half-line with gaps, a union of intervals, a quarterplane with missing sites, or a more general parameter set.

Cheng's work on incomplete quarterplane past [**?**] already shows that as soon as one moves away from the full half-space geometry, the structure of the problem changes sharply: strong outer factorization and modified duality arguments become necessary, and the theory expands beyond a single canonical observation set.

Problem statement

Develop a Hardy-space and spectral-factorization approach to prediction from incomplete or geometrically nontrivial pasts. The most immediate one-parameter version is:

> Given a centered stationary Gaussian process $X(t)$, predict $X(T)$ when the available observations are
> $$(-\infty, 0] \setminus \bigcup_{j=1}^{m} I_j,$$
> where $I_1, \ldots, I_m$ are finitely many bounded intervals removed from the past.

A second stage would be to extend the same philosophy to random fields and product-type pasts.

Why this problem is meaningful

This is not merely a technical variant. It probes how prediction depends on the *geometry* of information. That is a conceptually deeper question than the classical full-past problem, and one that naturally forces the introduction of new tools: product Hardy spaces, several-complex-variables factorization, de Branges-type methods, or refined duality.

A concentrated first target

The best short-term target is the continuous-time one-parameter case with finitely many missing intervals in the past. That problem remains close enough to the current thesis that one can still hope to transport it into a spectral space and to identify the obstruction created by the missing intervals in terms of a modified projection problem. One can think of this as the simplest genuinely nontrivial deformation of the classical geometry.

Possible route

A reasonable route is:

(1) express the incomplete past as a closed subspace in the spectral space;

(2) identify the orthogonal complement created by the missing intervals;

(3) reformulate the predictor as a constrained Hardy projection problem;

(4) isolate the additional correction term caused by the missing observation intervals;

(5) only after the one-parameter case is understood, move to random fields or quarterplane-type geometries.

The attraction of this problem is that it stays conceptually close to the thesis while opening the door to genuinely new geometry.

4.6 Open Problem IV: Finite-Window Asymptotics Under Critical or Long-Memory Spectra

Motivation and current status

Among all the possible continuations of the present thesis, this is perhaps the one that stays closest to the scalar classical line while still looking genuinely active. The whole-past theory is classical, but the asymptotic behavior of *finite-window* predictors in critical regimes remains delicate. Recent work of Chigansky and Kleptsyna [?] gives a new approach to the asymptotic analysis of the finite predictor for fractional Gaussian noise and derives exact asymptotics for relative prediction errors and partial correlations. This shows that the frontier is not exhausted even in prediction theory close to its classical core.

Problem statement

Develop a continuous-time asymptotic theory for prediction from a finite past window $[-L, 0]$ as $L \to \infty$, in regimes where the spectral density is critical or nearly singular. Typical examples include densities behaving like

$$f(\lambda) \sim c|\lambda|^{-2d} \qquad (0 < d < 1/2)$$

near $\lambda = 0$, or continuous-time analogues of fractional and Matérn-type models.

The aim is to derive exact or sharp asymptotic formulas for:

(a) the finite-window prediction error;

(b) the relative gain of using the whole past instead of a finite past;

(c) the asymptotic form of the predictor kernel itself.

Why this problem is meaningful

This problem is especially attractive because it is both mathematically substantive and tightly connected to the current thesis. It stays within the architecture of spectral factorization and projection, but it forces one to study how the factorization degenerates near singular frequencies and how truncated prediction differs from ideal whole-past prediction. It is therefore a natural bridge between classical theory and modern asymptotic analysis.

A concentrated first target

A very promising first paper would be:

> Take a continuous-time stationary Gaussian process with absolutely continuous spectral measure and long-memory singularity near the origin. Obtain a sharp asymptotic formula, as $L \to \infty$, for the error of predicting $X(T)$ from $[-L, 0]$, and compare it explicitly with the whole-past error from Chapter 3.

This would remain close to the current thesis while still producing new mathematics.

Possible route

The likely ingredients are:

(1) the whole-past outer-factor formula from Chapter 3 as a reference solution;

(2) a Wiener–Hopf or truncated Toeplitz reformulation of the finite-window problem;

(3) asymptotic analysis of the corresponding factor or symbol near the critical frequency;

(4) comparison between whole-past and finite-past predictors through the tail behavior of the inverse Fourier transform of the outer factor.

This problem is especially well suited to the "short and concentrated" strategy because it does not require rebuilding the entire theory in a new category; it asks instead for a refined asymptotic analysis of a very classical object.

4.7 A Focused Research Strategy

The four problems above are not equally suitable as first projects. If the criterion is to produce a concentrated and realistic first paper while staying very close to the present thesis, then the following order is the most natural.

First wave

(1) **Open Problem II: multivariate continuous-time prediction and real-line matrix factorization;**
this is probably the best balance between depth, tractability, and publishability;

(2) **Open Problem IV: finite-window asymptotics under long-memory or critical spectra;**
this remains particularly close to the current thesis and is likely the most focused scalar sequel.

Second wave

(3) **Open Problem I: operator-valued continuous-time theory;**
this is arguably the most conceptually important continuation, but it is more ambitious and technically heavier;

(4) **Open Problem III: incomplete past and geometric observation sets.**
this has very high mathematical value, but the geometry can quickly enlarge the technical burden.

A broader horizon

Beyond these four focused problems, the larger modern horizon includes locally stationary and time-varying prediction theory; see, for instance, Cui and Zhou [**?**]. That direction is undoubtedly important, but relative to the present thesis it is a broader second-generation program rather than the most concentrated immediate sequel.

Remark 4.1. The most important conclusion of this chapter is that the present thesis should not be viewed as an endpoint, but as a base architecture. What has been built in the scalar continuous-time setting is not itself the frontier; rather, it is the platform from which one can move toward operator-valued prediction, matrix factorization, geometric observation sets, and asymptotic finite-past theory. In that sense, the thesis is not merely a summary of classical theory, but a usable springboard into several active and promising research programs.

Bibliography

[1] A. Beurling. *On two problems concerning linear transformations in Hilbert space*. Acta Math. **81** (1949), 239–255.

[2] H. Dym and H. P. McKean. *Fourier Series and Integrals*. Academic Press, New York, 1972.

[3] H. Dym and H. P. McKean. *Gaussian Processes, Function Theory, and the Inverse Spectral Problem*. Dover Publications, Mineola, NY, 2008.

[4] A. N. Kolmogorov. *Sur l'interpolation et l'extrapolation des suites stationnaires*. C. R. Acad. Sci. Paris **208** (1939), 2043–2045.

[5] A. N. Kolmogorov. *Interpolation und Extrapolation von stationären zufälligen Folgen*. Izv. Akad. Nauk SSSR Ser. Mat. **5** (1941), 3–14.

[6] A. N. Kolmogorov. *Stationary sequences in Hilbert space*. Bull. Math. Univ. Moscow **2** (1941), 1–40.

[7] M. G. Krein. *On a problem of extrapolation of A. N. Kolmogorov*. Dokl. Akad. Nauk SSSR **46** (1945), 306–309.

[8] R. E. A. C. Paley and N. Wiener. *Fourier Transforms in the Complex Domain*. American Mathematical Society Colloquium Publications, Vol. 19, Amer. Math. Soc., New York, 1934.

[9] G. Szegő. *Beiträge zur Theorie der Toeplitzschen Formen*. Math. Z. **6** (1920), 167–202.

[10] S. R. S. Varadhan. *Probability Theory*. Courant Lecture Notes in Mathematics, Vol. 7, American Mathematical Society, Providence, RI, 2001.

[11] N. Wiener. *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*. MIT Press / Wiley, Cambridge, MA / New York, 1949.