Chapter 3
Spectral Representation of Stationary Processes

In this chapter we review the *spectral representation* of stationary processes. This representation theory is useful for at least two reasons. First it leads to concrete representation results of stationary processes in terms of white noise. These representations are basic for filtering and prediction and also for state-space modeling of random signals. Second, spectral representation theory provides a functional calculus for random variables and processes in terms of functions of a complex variable, much in the same spirit as the Fourier transform for deterministic signals. Unfortunately the Fourier transform of a stationary process cannot be defined in a deterministic pathwise sense. For it is well-known that the sample paths of a discrete-time stationary Gaussian process of, say, independent random variables (discrete time white noise) are neither in ℓ^2 nor uniformly bounded with probability one, and hence as functions of time they do not admit a Fourier transform [129].

The Fourier transform of a stationary process can however be defined in a (global) mean-square sense, but this transform will not provide a stochastic process in the ordinary sense but rather an equivalence class of processes with orthogonal increments, or an *orthogonal random measure*, as these objects are commonly called in the literature.

3.1 Orthogonal-Increments Processes and the Wiener Integral

Let $\mathbb{T}$ be a subinterval (possibly infinite) of the real line $\mathbb{R}$. A scalar continuous-time process $x = \{x(t);\ t \in \mathbb{T}\}$, is said to have *orthogonal increments* if whenever $s_1 < t_1 \leq s_2 < t_2$ we have

$$\mathrm{E}\{(x(t_2) - x(s_2))\overline{(x(t_1) - x(s_1))}\} = 0, \tag{3.1}$$

A. Lindquist, G. Picci, *Linear Stochastic Systems*, Series in Contemporary Mathematics 1, DOI 10.1007/978-3-662-45750-4_3

where the overline denotes complex conjugation. To this requirement we shall also add the zero mean condition,

$$\mathrm{E}(x(t) - x(s)) = 0 \quad t, s \in \mathbb{T}. \tag{3.2}$$

We alert the reader to the fact that *complex* orthogonal increment processes defined on the imaginary axis will play an important role in spectral representation, discussed in Sect. 3.3.

Proposition 3.1.1. *Let x be a process with orthogonal increments, then there is a real monotone nondecreasing function F, uniquely determined by x up to an additive constant, such that,*

$$\mathrm{E}\{|x(t) - x(s)|^2\} = F(t) - F(s), \quad t \geq s. \tag{3.3}$$

Proof. Let us fix an arbitrary t_0 and define,

$$F_0(t) := \begin{cases} \mathrm{E}\{|x(t) - x(t_0)|^2\}, & t \geq t_0 \\ -\mathrm{E}\{|x(t) - x(t_0)|^2\}, & t < t_0 \end{cases}$$

Then, by using the property (3.1), it is immediate to check that F_0 is monotone and satisfies (3.3). The function F_0 is clearly the unique function satisfying (3.3) normalized at t_0 so as $F_0(t_0) = 0$. Hence any function $F(t) := F_0(t) +$ an arbitrary constant, also satisfies (3.3) and is independent of t_0. □

The relation (3.3) is often written symbolically as

$$\mathrm{E}\{|dx(t)|^2\} = dF(t).$$

It follows from (3.3) that an orthogonal increments process has the same continuity properties (in mean square) as the monotone function F. In particular x has right and left limits at every point t and an at most countable set of points of discontinuity which can only be jumps. Without much loss of generality, x can be modified at the jump points to have $x(t+) = x(t)$, and hence also $F(t+) = F(t)$, for all $t \in \mathbb{T}$. If $\mathbb{T} = (a, b]$, then in this way the process is automatically extended to the closure $[a, b]$.

A mean-square continuous process $w := \{w(t), \ t \in \mathbb{R}\}$, with *stationary* orthogonal increments will be called a (wide-sense) *Wiener process*. Note that, by stationarity of the increments, $F(t + h) - F(t) = F(h) - F(0)$ for all t, so that for a Wiener process the derivative $F'(t)$ (which a priori exists almost everywhere) is independent of t. By continuity, one finds a unique monotone nondecreasing solution of the form

$$F(t) = \sigma^2 t + \text{constant}$$

where σ^2 is a positive constant. Hence for a Wiener process, we have $E\{|dw(t)|^2\} = \sigma^2 dt$. In other words, the variance of the process grows linearly in time. If $\sigma^2 = 1$ the Wiener process is said to be *normalized*.

The Wiener process is a mathematically tractable version of the concept of "continuous-time stationary *white noise*" which, intuitively, is a process with completely uncorrelated variables and should correspond to the derivative

$$n(t) = \frac{dw(t)}{dt}.$$

It is easy to see that this derivative cannot exist in mean square. It has been shown in many ways that it is actually impossible to give n a precise interpretation as a stochastic process in the sense we understand this term in probability theory, see e.g. [315]. On the other hand, white noise and representations of various random variables as functionals of white noise constitute an extremely useful tool in the analysis of stationary processes. For this reason there is a need for a rigorous theory of white-noise representation involving in particular integrals with respect to the Wiener process, which we shall now proceed to define.

Definition 3.1.2. Let $\{\Omega, \mathcal{A}, \mu\}$ be a probability space and let $\mathcal{R}$ be the family of bounded semi-open intervals $(a, b]$ of the real line.[1] An *orthogonal stochastic measure* on $\mathbb{R}$ is a family of random variables $\{\zeta(\Delta);\ \Delta \in \mathcal{R}\}$, where $\zeta(\Delta) : \{\Omega, \mathcal{A}, \mu\} \to \mathbb{C}$ is such that

(i) For each interval $\Delta \in \mathcal{R}$, $\zeta(\Delta)$ is a random variable with zero mean and finite variance

$$m(\Delta) = \mathrm{E}\{|\zeta(\Delta)|^2\} < \infty, \quad \Delta \in \mathcal{R}. \tag{3.4}$$

(ii) For any pair of disjoint intervals Δ_1, Δ_2 with $\Delta_1 \cap \Delta_2 = \emptyset$,

$$\mathrm{E}\{\zeta(\Delta_1)\overline{\zeta(\Delta_2)}\} = 0. \tag{3.5}$$

(iii) ζ is σ-additive, i.e. for any $\Delta \in \mathcal{R}$ which is the disjoint union of countably many sets $\Delta_k \in \mathcal{R}$,

$$\zeta(\Delta) = \sum_{k=1}^{\infty} \zeta(\Delta_k), \quad a.s. \tag{3.6}$$

where the series in the second member converges in mean square.

[1] The family $\mathcal{R}$ is a *semi-ring* of sets, see [130, p. 22]. A semi-ring is sometimes also called a *decomposable class* of sets. More generally, a stochastic orthogonal measure could be defined on an arbitrary semi-ring of sets.

Note that, by Lemma B.2.1 in the appendix, the series of orthogonal random variables (3.6) converges if and only if

$$m(\Delta) = \sum_{k=1}^{\infty} \mathrm{E}\{|\zeta(\Delta_k)|^2\} = \sum_{k=1}^{\infty} m(\Delta_k) < \infty$$

so that m is a nonnegative σ-additive set function which can be extended as a σ-finite measure on the Borel σ-algebra of sets generated by $\mathcal{R}$ see e.g. [117, 130]. Conversely, m being σ-additive on $\mathcal{R}$ implies that ζ is σ-additive in the sense of (3.6) above. In this sense, it is then possible to extend ζ to the σ-ring generated by $\mathcal{R}$, where $m(\Delta) < \infty$, see also [270, p. 5]. Note that ζ may not be extendable to unbounded sets.

The measure ζ is called *finite* if $\mathrm{E}\,|\zeta(\mathbb{R})|^2 < \infty$. This is clearly the case if and only if m is a finite Borel measure.

The notion of orthogonal stochastic measure is the natural starting point to discuss stochastic integration. Before embarking on this, we remark that any orthogonal increments process x defines a stochastic orthogonal measure, which we shall denote dx, by the assignment

$$dx((a,b]) := x(b) - x(a), \quad a < b.$$

The variance measure m associated to dx is uniquely determined by the variance function F of the process as

$$m((a,b]) := F(b) - F(a), \quad a < b.$$

Conversely, any orthogonal random measure ζ determines an orthogonal increments process z by

$$z(t) := \begin{cases} \zeta((t_0,t]), & t \geq t_0, \\ -\zeta((t,t_0]), & t < t_0, \end{cases}$$

where t_0 is an arbitrary fixed time instant. The orthogonal increments process z is normalized so that $z(t_0) = 0$; in fact, ζ determines a whole equivalence class of orthogonal increments processes, all differing from the one just defined by an arbitrary additive random variable.

In particular, for the stochastic orthogonal measure corresponding to the normalized Wiener process w, the variance measure m is the Lebesgue measure. Since the increments of w are the only thing that matters in this book, it will be convenient to identify a Wiener process with the corresponding orthogonal stochastic measure dw. Therefore, in the following, whenever we talk about a Wiener process we will always refer to a whole equivalence class of processes defined modulo an arbitrary additive random variable. Note that the stochastic measure dw is not finite.

We shall now proceed to define the stochastic integral with respect to an orthogonal random measure ζ. Let I_Δ denote the indicator function of the set Δ, i.e. $I_\Delta(t) = 1$ if $t \in \Delta$ and zero otherwise. For a scalar *simple function*

$$f(t) = \sum_{k=1}^{N} c_k I_{\Delta_k}(t), \quad \Delta_k \in \mathcal{R},\ \Delta_k \cap \Delta_j = \emptyset,\ k \neq j,$$

the integral of f with respect to ζ is defined as follows,

$$\int_{\mathbb{R}} f(t) d\zeta(t) := \sum_{k=1}^{N} c_k \zeta(\Delta_k). \tag{3.7}$$

Note that the integrals of simple functions are just the (zero-mean) random variables in the linear vector space

$$\mathbf{L}(\zeta) := \text{span}\{\zeta(\Delta) \mid \Delta \in \mathcal{R}\} = \text{span}\{\zeta((a,b]) \mid -\infty < a < b < +\infty\}, \tag{3.8}$$

generated by the increments of ζ.

The fundamental property of the stochastic integral of simple functions is

$$\mathrm{E}\left\{\left|\int_{\mathbb{R}} f(t) d\zeta(t)\right|^2\right\} = \sum_{k=1}^{N} |c_k|^2 m(\Delta_k) = \int_{\mathbb{R}} |f(t)|^2 dm, \tag{3.9}$$

showing that the integral is an *isometric map* mapping the dense linear manifold of simple functions in the Lebesgue space $L^2(\mathbb{R}, dm)$ onto $\mathbf{L}(\zeta)$. We denote this map by the symbol $\mathfrak{I}_\zeta$. Using this compact notation the formula (3.9) reads

$$\|\mathfrak{I}_\zeta(f)\| = \|f\|_{L^2(\mathbb{R},dm)},$$

where the norm in the first member is the variance norm in the linear manifold $\mathbf{L}(\zeta)$.

Let us now take an arbitrary function $f \in L^2(\mathbb{R}, dm)$. Then f is the limit in mean square of a sequence of simple square-integrable functions f_n,

$$\int_{\mathbb{R}} |f(t) - f_n(t)|^2 dm \to 0, \quad n \to \infty,$$

so that by the isometric property of the integral

$$\|\mathfrak{I}_\zeta(f_n) - \mathfrak{I}_\zeta(f_k)\| = \|f_n - f_k\|_{L^2(\mathbb{R},dm)} \to 0$$

as $n, k \to \infty$. Therefore the sequence $\{\mathfrak{I}_\zeta(f_n)\}$ is a fundamental sequence in $L^2(\Omega, \mathcal{A}, \mu)$ and converges to a random variable with finite variance which we shall

define to be the integral of f with respect to the stochastic measure ζ. In other words, for an arbitrary $f \in L^2(\mathbb{R}, dm)$, the stochastic integral of f with respect to ζ is the mean square limit

$$\mathcal{I}_\zeta(f) = \int_{\mathbb{R}} f(t)d\zeta(t) := \lim_{n\to\infty} \int_{\mathbb{R}} f_n(t)\, d\zeta(t). \tag{3.10}$$

It is easy to check that the limit is indeed independent of the particular sequence of simple functions. The fundamental property of the integral is recorded in the following theorem. The proof is straightforward invoking Theorem B.2.7.

Theorem 3.1.3. *The stochastic integral $\mathcal{I}_\zeta$ is a linear bijective map from $L^2(\mathbb{R}, dm)$ onto the Hilbert space $\mathbf{H}(\zeta) = \text{closure}\, \mathbf{L}(\zeta)$ which preserves inner product*

$$\mathrm{E}\left\{ \int_{\mathbb{R}} f(t)d\zeta(t) \overline{\int_{\mathbb{R}} g(t)d\zeta(t)} \right\} = \int_{\mathbb{R}} f(t)\bar{g}(t)\, dm. \tag{3.11}$$

In other words, $\mathcal{I}_\zeta$ is a unitary map $L^2(\mathbb{R}, dm) \to \mathbf{H}(\zeta)$.

We omit the proof of the following immediate corollary of Theorem 3.1.3.

Corollary 3.1.4. *The map assigning to any Borel set $\Delta \subset \mathbb{R}$ the random variable*

$$\eta(\Delta) := \int_\Delta f(t)d\zeta(t) = \int_{\mathbb{R}} I_\Delta(t) f(t)d\zeta(t) \tag{3.12}$$

is a finite stochastic orthogonal measure if and only if $f \in L^2(\mathbb{R}, dm)$.

This measure we shall denote $d\eta = f d\zeta$.

3.2 Harmonic Analysis of Stationary Processes

There is a fundamental result in analysis which provides a harmonic representation of the covariance function of a stationary process. This result, here reported without proof, is known as the *Herglotz Theorem* in the discrete-time case and as the *Bochner Theorem* in continuous time.

Let $\tau \to \Lambda(\tau)$ be the covariance function of a scalar stationary random process[2] y, where $\tau \in \mathbb{Z}$ in the discrete time case and $\tau \in \mathbb{R}$ in the continuous time case. In continuous time Λ will be assumed to be a continuous function[3] of $\tau \in \mathbb{R}$.

[2]Recall that all stationary processes considered in this book will have finite second-order moments.

[3]This is equivalent to assuming y mean-square continuous.

Theorem 3.2.1 (Herglotz, Bochner). *Let $\mathcal{I}$ be the interval $[-\pi, \pi]$ in the discrete time case and $(-\infty, +\infty)$ in the continuous time case. Then, given a covariance function Λ, there is a finite positive measure dF on the Borel subsets of the interval $\mathcal{I}$ such that*

$$\Lambda(\tau) = \int_{\mathcal{I}} e^{i\theta\tau} dF(\theta)\,. \tag{3.13}$$

The measure dF is uniquely determined by Λ.

An equivalent (although a bit more cumbersome) way of formulating the result is to say that there is a real right-continuous monotone non-decreasing function F defined on the interval $[-\pi, \pi]$ (discrete time) or $(-\infty, +\infty)$ (continuous time) such that (3.13) holds. The monotone function F, uniquely determined by Λ modulo an arbitrary additive constant, is called the *spectral distribution function* of the process y. One can make F unique by imposing say $F(-\pi) = 0$ (in this case dF has no mass at $\theta = -\pi$). Since

$$\infty > \mathrm{E}\{|y(t)|^2\} = \Lambda(0) = \int_{-\pi}^{\pi} dF(\theta) = F(\pi),$$

the function F must actually be *bounded*. This spectral distribution function describes how the "statistical power" $E\{|y(t)|^2\} = \Lambda(0)$, of the process y is distributed in frequency. For this reason it is called *power spectral distribution function* in the engineering literature.

Example 3.2.2. Consider a random sum of simple harmonic oscillations

$$y(t) = \sum_{k=-N}^{N} y_k e^{i\theta_k t},$$

where $-\pi < \theta_k \leq \pi$ are deterministic frequencies and y_k are mutually uncorrelated zero-mean random variables with variance σ_k^2. This process is stationary with a quasi-periodic covariance function

$$\Lambda(\tau) = \sum_{k=-N}^{N} \sigma_k^2 e^{i\theta_k \tau}.$$

Since we can formally rewrite $\Lambda(\tau)$ in the form (3.13) with F the monotone function

$$F(\theta) := \sum_{k=-N}^{N} \sigma_k^2 1(\theta - \theta_k) \quad -\pi \leq \theta \leq \pi$$

where $1(\theta)$ is the indicator function of the half line $\{\theta \geq 0\}$, it follows that F is the distribution function of the process. In this simple example the power spectral distribution function increases only at the jumps of F and the statistical power of the process $\Lambda(0) = \sum_{k=-N}^{N} \sigma_k^2$ is all concentrated at the discrete frequencies θ_k. In more general situations the power of the process will also be distributed continuously on the interval $-\pi < \theta \leq \pi$.

Like every real monotone function, the spectral distribution function F can be split in two components

$$F = F_1 + F_2 \tag{3.14}$$

where F_1 is the *absolutely continuous component*,

$$F_1(\theta) = \int_{-\pi}^{\theta} \Phi(\lambda) \frac{d\lambda}{2\pi}$$

and F_2 is the *singular component* of F, whose points of increase are a set of Lebesgue measure zero. The singular part F_2 carries all discontinuities (finite jumps) of F. The non-negative function Φ is called the *spectral density function* of the process.

If Λ is a summable function, i.e., $\sum_{\tau=-\infty}^{+\infty} |\Lambda(\tau)| < \infty$, then the series

$$\sum_{\tau=-\infty}^{+\infty} e^{-i\theta\tau} \Lambda(\tau) \tag{3.15}$$

converges pointwise uniformly in the interval $[-\pi, \pi]$ to a periodic function $\hat{\Lambda}(\theta)$, and then the coefficients $\{\Lambda(\tau)\}$ must necessarily be the Fourier-series coefficients of $\hat{\Lambda}(\theta)$; i.e.,

$$\Lambda(\tau) = \int_{-\pi}^{+\pi} e^{i\theta\tau} \hat{\Lambda}(\theta) \frac{d\theta}{2\pi}. \tag{3.16}$$

It follows that in this case the distribution function is absolutely continuous and the spectral density function is just $\hat{\Lambda}(\theta)$, namely

$$\Phi(\theta) = \hat{\Lambda}(\theta).$$

Remark 3.2.3. To make contact with the Fourier transform of ordinary functions (which we shall need to do later on), it turns out to be convenient to extend the distribution function F in the Herglotz representation as a periodic function to the whole real axis. Equivalently, one can always think of F as being a function defined on the unit circle, $\mathbb{T} := \{z = e^{i\theta};\ -\pi < \theta \leq \pi\}$, of the complex plane. Therefore

it is more natural to define the density Φ as a function defined on the unit circle and hence as a function of $e^{i\theta}$. In view of this, with a slight misuse of notation, we write $F(e^{i\theta})$ or $\Phi(e^{i\theta})$ instead of $F(\theta)$ or $\Phi(\theta)$ whenever convenient, without further notice. Similarly in continuous time, it turns out to be convenient to regard the spectral distribution F or Φ as a function on the imaginary axis $\mathbb{I}$; i.e., as a function of $i\omega$.

3.3 The Spectral Representation Theorem

The Fourier-like representation of the covariance function of a stationary process provided by Herglotz's Theorem is the basis of a stochastic Fourier-like representation for the process y itself. This representation theorem is important as it provides very precise information about the structure of the elements of the space $\mathbf{H}(y)$.

We shall define a linear map, which for the moment we denote $\mathcal{I}$ (a more descriptive notation will be introduced in the following), mapping the functions $\hat{f} \in L^2\{[-\pi,\pi], dF\}$, square integrable with respect to the spectral distribution dF, into random variables belonging to $\mathbf{H}(y)$. This map will be first defined on a dense set of functions and then extended by continuity.

Let $\mathcal{I}$ map the elementary trigonometric functions $\theta \to e_k(\theta) := e^{i\theta k}$ into the random variables $y(k)$; $k \in \mathbb{Z}$. We extend $\mathcal{I}$ by linearity so that

$$\mathcal{I}\left(\sum_k c_k e_k\right) := \sum_k c_k y(k), \quad k \in \boldsymbol{\zeta},\ c_k \in \mathbb{C}, \tag{3.17}$$

for all finite linear combinations $\sum_k c_k e_k$, called *trigonometric polynomials*. In this way $\mathcal{I}$ maps the linear manifold $\mathcal{E} \subset L^2\{[-\pi,\pi],\ dF\}$ of all trigonometric polynomials onto the dense linear manifold $\mathbf{L}(y) \subset \mathbf{H}(y)$ spanned by the random variables of the process

$$\mathbf{L}(y) := \text{span}\{y(t)\,;\, t \in \mathbb{Z}\}. \tag{3.18}$$

Now, it follows from Weierstrass approximation theorem that the manifold $\mathcal{E}$ *is dense in* $L^2\{[-\pi,\pi], dF\}$; a proof of this fact can for example be found in [231, 232]. Then, by a simple application of Herglotz's Theorem one can see that the map $\mathcal{I}$ is *isometric*, as

$$\langle e_k, e_j \rangle_{L^2\{[-\pi,\pi],dF\}} = \Lambda(k-j) = \langle y(k), y(j) \rangle_{\mathbf{H}(y)}, \tag{3.19}$$

and hence, since any $\hat{f} \in L^2\{[-\pi,\pi], dF\}$ is the mean square limit of a sequence of trigonometric polynomials $(\hat{f}_k)$, $\mathcal{I}$ can be extended by continuity to the whole

of $L^2\{[-\pi,\pi], dF\}$. In fact, by (3.19), $\mathcal{I}(\hat{f}_k)$ also converges in mean square to some random variable in $\mathbf{H}(y)$. We just define $\mathcal{I}(\hat{f})$ to be this limit

$$\mathcal{I}(\hat{f}) := \lim_{k\to\infty} \mathcal{I}(\hat{f}_k)$$

in $L^2(\Omega, \mathcal{A}, \mu)$. In this way the extended map (still denoted by) $\mathcal{I}$, becomes a *unitary map* from $L^2\{[-\pi,\pi], dF\}$ onto $\mathbf{H}(y)$ (Theorem B.2.7). This leads to the following fundamental result.

Theorem 3.3.1. *There is a finite orthogonal stochastic measure $d\hat{y}$ on the (Borel sets of the) interval $-\pi < \theta \le \pi$, such that*

$$\mathcal{I}(\hat{f}) = \int_{-\pi}^{+\pi} \hat{f}(\theta) d\hat{y}(\theta), \quad \hat{f} \in L^2\{[-\pi,\pi], dF\}, \tag{3.20}$$

so that, in particular,

$$y(t) = \int_{-\pi}^{\pi} e^{i\theta t} d\hat{y}(\theta), \quad t \in \mathbb{Z}. \tag{3.21}$$

The orthogonal stochastic measure is uniquely determined by the process y and satisfies

$$\mathrm{E}\{d\hat{y}(\theta)\} = 0, \quad \mathrm{E}\{|d\hat{y}(\theta)|^2\} = dF(\theta), \tag{3.22}$$

where F is the spectral distribution function of y.

It is implicit in the statement of the theorem that every discrete-time stationary process admits an integral representation of the form (3.21). Formula (3.21) is normally called the *spectral representation* of the discrete-time stationary process y. The stochastic measure $d\hat{y}$ will be referred to as the *Fourier transform* of the process y in this book. The map $\mathcal{I}$ corresponding to a specific process y will hereafter be denoted by $\mathcal{I}_{\hat{y}}$.

Proof. Let $\Delta := (\theta_1, \theta_2]$ be a subinterval of $[-\pi,\pi]$, let I_Δ be the indicator function of Δ, and define

$$\hat{y}(\Delta) := \mathcal{I}(I_\Delta) \tag{3.23}$$

so that by the isometric character of $\mathcal{I}$ we have $\mathrm{E}\{|\hat{y}(\Delta)|^2\} = \|I_\Delta\|^2_{L^2\{[-\pi,\pi],dF\}} = F(\Delta)$. Here we have denoted by F also the Borel measure induced by the spectral distribution function F. Also, for an arbitrary pair of intervals Δ_1, Δ_2 we have

$$\mathrm{E}\{\hat{y}(\Delta_1)\overline{\hat{y}(\Delta_2)}\} = \langle I_{\Delta_1}, I_{\Delta_2}\rangle_{L^2\{[-\pi,\pi],dF\}} = F(\Delta_1 \cap \Delta_2),$$

from which, taking $\Delta_1 \cap \Delta_2 = \emptyset$, it is easily seen that $\hat{y}$ is a stochastic orthogonal measure defined on the semi-open intervals of $[-\pi, \pi]$ satisfying (3.23). Obviously this measure is finite as $\mathrm{E}\{|\hat{y}((-\pi, \pi])|^2\} = F((-\pi, \pi]) < \infty$ and can then be extended to the Borel sets of the interval $[-\pi, \pi]$.

We now proceed to show that (3.20) holds for all $\hat{f} \in L^2\{[-\pi, \pi], dF\}$. This is certainly true for simple functions since in this case

$$\mathcal{I}(\hat{f}) = \sum_{k=1}^{N} c_k \mathcal{I}(I_{\Delta_k}) = \sum_{k=1}^{N} c_k \hat{y}(\Delta_k) = \int_{-\pi}^{\pi} \hat{f}(\theta) d\,\hat{y}(\theta)$$

by the very definition of the stochastic integral. Now, simple functions are dense in $L^2\{[-\pi, \pi], dF\}$, and, by the isometry described above, the family of random variables $\{\mathcal{I}(\hat{f}) |\ \hat{f}$ simple$\}$ is dense in $\mathbf{H}(y)$. Hence any random variable $\xi \in \mathbf{H}(y)$, being the limit in mean square of a sequence $\mathcal{I}(\hat{f}_k)$ with $\hat{f}_k$ simple functions, is at the same time the limit of a sequence of stochastic integrals of simple functions $\mathcal{I}_{\hat{y}}(\hat{f}_k)$. Therefore every random variable of $\mathbf{H}(y)$ is a stochastic integral of some function $\hat{f} \in L^2\{[-\pi, \pi], dF\}$ with respect to the stochastic measure $\hat{y}$.

Note that the converse of this statement is obviously also true as all $\hat{y}(\Delta)$ are random variables in $\mathbf{H}(y)$ by definition and the stochastic integral of all functions $\hat{f} \in L^2\{[-\pi, \pi], dF\}$ are then also in $\mathbf{H}(y)$.

□

3.3.1 *Connections to the Classical Definition of Stochastic Fourier Transform*

It is instructive to examine the relation of the spectral representation, as it has been introduced in this section, with the classical early definition of stochastic Fourier transform. This is done below, in a series of conceptual steps. The details of the procedure can be found in the early literature or, in condensed form, in [270, pp. 26–27].

1. Let t be a discrete time parameter. One may first try to formally define the Fourier transform of a stationary second-order process y as the limit (in mean square)

$$Y(\theta) = \lim_{N \to \infty} \sum_{t=-N}^{+N} e^{-i\theta t} y(t), \tag{3.24}$$

but for a stationary process this mean square limit cannot exist. (The case that y is white noise is quite obvious.)

2. Then one formally integrates (3.24) with respect to θ on an interval $\Delta := [\theta_1, \theta_2] \subset [-\pi, \pi]$. Setting

$$\chi_t(\Delta) = \begin{cases} \frac{e^{-i\theta_2 t} - e^{-i\theta_1 t}}{-2\pi i t}, & t \neq 0 \\ \frac{\theta_2 - \theta_1}{2\pi}, & t = 0 \end{cases}$$

the integrated Fourier series

$$\lim_{N\to\infty} \sum_{t=-N}^{+N} \chi_t(\Delta) y(t) \tag{3.25}$$

now *converges in mean square* and converges to the stochastic orthogonal measure (which we defined as the Fourier transform of y) $\hat{y}(\Delta)$. Hence $\hat{y}(\Delta)$ is an integrated version of the formal Fourier transform and we may write

$$\hat{y}(\Delta) := \int_{\theta_1}^{\theta_2} Y(\lambda) \frac{d\lambda}{2\pi}.$$

One can show convergence by working out the following steps

(a) The *deterministic* Fourier series

$$S_N(\theta) := \sum_{t=-N}^{+N} \chi_t(\Delta) e^{i\theta t} \tag{3.26}$$

converges pointwise as $N \to \infty$ to the indicator function $I_\Delta(\theta)$ of the interval $\Delta := [\theta_1, \theta_2]$. Actually, for this to be literally true one needs to modify slightly the definition of I_Δ at the extreme points of the interval, in order to have pointwise convergence also at θ_1, θ_2.

(b) Since $S_N(\theta)$ converges boundedly pointwise to $I_\Delta(\theta)$, we also have

$$S_N \to I_\Delta \quad \text{in } L^2([-\pi, \pi], dF),$$

where F is the spectral distribution of the process y. Hence, by the well-known isometric property of the stochastic integral,

$$\hat{y}(\Delta) = \int_{-\pi}^{\pi} I_\Delta(\theta) d\hat{y}(\theta) = \lim_{N\to\infty} \int_{-\pi}^{\pi} S_N(\theta) d\hat{y}(\theta).$$

(c) The last integral in the equation is just the integrated Fourier series (3.25).

3. In this sense one may say that the formal Fourier series (3.24) converges to the white noise $Y(\theta)$ on $[-\pi, \pi]$ as $N \to \infty$.

3.3.2 Continuous-Time Spectral Representation

The continuous-time analog of Theorem 3.3.1 is as follows.

Theorem 3.3.2. *Every stationary process* $y := \{y(t)\,;\, t \in \mathbb{R}\}$*, continuous in mean-square, admits a representation*

$$y(t) = \int_{-\infty}^{+\infty} e^{i\omega t}\, d\hat{y}(i\omega), \quad t \in \mathbb{R}, \tag{3.27}$$

where $d\hat{y}$ *is a finite orthogonal stochastic measure uniquely determined by the process, which satisfies*

$$\mathrm{E}\{d\hat{y}(i\omega)\} = 0, \qquad \mathrm{E}\{|d\hat{y}(i\omega)|^2\} = dF(i\omega), \tag{3.28}$$

where F *is the spectral distribution function of* y*. The map* $\mathfrak{I}_{\hat{y}}$ *defined by the stochastic integral*

$$\mathfrak{I}_{\hat{y}}(\hat{f}) = \int_{-\infty}^{+\infty} \hat{f}(i\omega) d\hat{y}(i\omega), \quad \hat{f} \in L^2\{(-\infty, +\infty), dF\}, \tag{3.29}$$

is an isometry from $L^2\{(-\infty, +\infty), dF\}$ *onto* $\mathbf{H}(y)$*.*

The orthogonal stochastic measure $\hat{y}$ (more commonly denoted $d\hat{y}$ in the following) is called *the Fourier transform of the stationary process* y*.*

The following corollary describes explicitly the fundamental isomorphism by which random elements in $\mathbf{H}(y)$ correspond to elements of the space $L^2\{[-\pi, \pi], dF\}$ and the corresponding action of the shift group.

Corollary 3.3.3 (Spectral Isomorphism Theorem). *Let* y *be a stationary discrete-time process. Then, every random element* $\xi \in \mathbf{H}(y)$ *can be written in a unique way as a stochastic integral* $\mathfrak{I}_{\hat{y}}(\hat{f})$*, with respect to the Fourier transform* $\hat{y}$ *of the process* y*, of some function* $\hat{f} \in L^2\{[-\pi, \pi], dF\}$*. In fact the map* $\mathfrak{I}_{\hat{y}} : L^2\{[-\pi, \pi], dF\} \to \mathbf{H}(y)$ *is isometric and bijective, i.e., unitary. It transforms the shift operator* $\mathcal{U}$ *into the operator of multiplication by the exponential function* $e(\theta) : \theta \to e^{i\theta}$*, acting on* $L^2\{[-\pi, \pi], dF\}$*, i.e.,*

$$\mathcal{U}\xi = \mathfrak{I}_{\hat{y}}(e\hat{f}), \quad \xi = \mathfrak{I}_{\hat{y}}(\hat{f}). \tag{3.30}$$

A totally analogous statement holds for continuous-time processes provided one substitutes $[-\pi, \pi]$ *for* $(-\infty, +\infty)$*, the unitary operator* $\mathcal{U}$ *for the shift group* $\{\mathcal{U}_t\,;t \in \mathbb{R}\}$*, and* $e^{i\theta}$ *for* $e^{i\omega t}$*,* $t \in \mathbb{R}$*.*

A generalization of this result to vector-valued processes will be given in the next sections.

3.3.3 Remark on Discrete-Time White Noise

A very simple but important kind of discrete-time stationary process is *(wide-sense) stationary white noise*. This is a stationary process $w = \{w(t), t \in \mathbb{Z}\}$ with uncorrelated (i.e., orthogonal) components. The covariance function of this process is a scalar multiple of the delta function, say $\Lambda(\tau) = \sigma^2\delta(\tau)$ where $\delta(\tau) = 1$ for $\tau = 0$ and zero otherwise. Since Λ is trivially a summable function, this process has an absolutely continuous spectral distribution function with a (spectral) density, which is just a constant function $\Phi(\theta) = \sigma^2$, $\theta \in [-\pi, \pi]$. The "flat" spectral density is the reason why this process is called white.

It follows that the spectral measure $d\hat{w}$, of a white noise process has the following property

$$\mathrm{E}\{d\hat{w}(\theta)d\hat{w}(\theta)^*\} = \sigma^2\frac{d\theta}{2\pi},$$

i.e., *$\hat{w}$ is a Wiener process on* $[-\pi, \pi]$. It is easy to see that, conversely, every process w with a spectral measure of the Wiener type,

$$w(t) = \int_{-\pi}^{\pi} e^{i\theta t}d\hat{w}(\theta), \quad t \in \mathbb{Z},$$

is white noise.

3.3.4 Real Processes

If the process y is real, its spectral measure has some special symmetry properties.

Proposition 3.3.4. *If y is a real stationary process, its spectral measure $\hat{y}$ is such that*

$$\overline{\hat{y}(\Delta)} = \hat{y}(-\Delta) \tag{3.31}$$

for every Borel set Δ of the interval $[-\pi, \pi]$, where $-\Delta = \{\theta \mid -\theta \in \Delta\}$. Moreover, the real and imaginary parts of $\hat{y}(\Delta) = \hat{r}(\Delta) + i\hat{s}(\Delta)$ are mutually orthogonal stochastic measures, i.e.,

$$\mathrm{E}\{\hat{r}(\Delta_1)\hat{s}(\Delta_2)\} = 0 \tag{3.32}$$

for all Borel sets Δ_1, Δ_2.

Proof. Notwithstanding the fact that the $y(t)$ are real random variables, we shall keep on working in the *complex* Hilbert space $\mathbf{H}(y)$. It is easy to see

that if $\hat{f}(\theta)$ corresponds under $\mathcal{I}_{\hat{y}}$ to the random variable η, then the complex conjugate $\bar{\eta}$ must be associated to the function $\overline{\hat{f}}(-\theta)$. This fact is true for all trigonometric polynomials $\hat{f}(\theta) = \sum_k c_k e_k(\theta)$ which correspond under $\mathcal{I}_{\hat{y}}$ to finite linear combinations $\eta := \sum_k c_k y(k)$, $c_k \in \mathbb{C}$, since clearly the complex conjugate $\bar{\eta} = \sum_k \bar{c}_k y(k)$ is associated to the function $\sum_k \bar{c}_k e_k(\theta) = \overline{\hat{f}}(-\theta)$. Then, since $\mathcal{I}_{\hat{y}} : I_\Delta \to \hat{y}(\Delta)$ we also have $\mathcal{I}_{\hat{y}} : \bar{I}_{-\Delta} \to \overline{\hat{y}}(\Delta)$, but $\bar{I}_{-\Delta} = I_{-\Delta}$, since the indicator is a real function and therefore (3.31) follows.

To prove the remaining statement first note that $\hat{r}$ and $\hat{s}$ are both σ-additive real stochastic measures and that from (3.31) we get

$$\hat{r}(\Delta) = \hat{r}(-\Delta), \quad \hat{s}(\Delta) = -\hat{s}(-\Delta) \tag{3.33}$$

for all Borel sets Δ. Moreover, since $\mathrm{E}\{\hat{y}(\Delta_1)\overline{\hat{y}}(\Delta_2)\} = \mathrm{E}\,|\hat{y}(\Delta_1 \cap \Delta_2)|^2 \geq 0$ it follows that $\mathbb{I}\mathrm{m}\ \mathrm{E}\{\hat{y}(\Delta_1)\overline{\hat{y}}(\Delta_2)\} = 0$, i.e.,

$$\mathrm{E}[\hat{s}(\Delta_1)\hat{r}(\Delta_2) - \hat{r}(\Delta_1)\hat{s}(\Delta_2)] = 0.$$

Combining this relation with the analogous one obtained by substituting $-\Delta_1$ in place of Δ_1 and using (3.33) one gets the orthogonality relation (3.32). Hence $\mathrm{E}\{\hat{y}(\Delta_1)\overline{\hat{y}(\Delta_2)}\} = \mathrm{E}\{\hat{r}(\Delta_1 \cap \Delta_2)^2 + \hat{s}(\Delta_1 \cap \Delta_2)^2\}$. However, $\Delta_1 \cap \Delta_2 = \emptyset$ implies that $\mathrm{E}\{\hat{r}(\Delta_1 \cap \Delta_2)\} = \mathrm{E}\{\hat{s}(\Delta_1 \cap \Delta_2)\} = 0$. This shows that $\hat{r}$ and $\hat{s}$ are also orthogonal measures and concludes the proof.

□

For real processes the spectral representation (3.21) can be written completely in terms of real quantities. From (3.33) one easily obtains

$$y(t) = \int_{-\pi}^{\pi} \cos\theta t \, d\hat{r}(\theta) - \int_{-\pi}^{\pi} \sin\theta t \, d\hat{s}(\theta), \quad t \in \mathbb{Z}.$$

3.4 Vector-Valued Processes

If we denote by $d\hat{y}_k$, $k = 1, 2, \ldots, m$, the spectral measure corresponding to the k-th component of an m-dimensional stationary process y, we can write the spectral representation of an m-dimensional process in vector form as

$$y(t) = \int e^{i\theta t} d\hat{y}(\theta), \quad t \in \mathbb{Z},$$

where $\hat{y}$ is now a *vector* stochastic orthogonal measure

$$\hat{y}(\Delta) = \begin{bmatrix} \hat{y}_1(\Delta) \\ \hat{y}_2(\Delta) \\ \vdots \\ \hat{y}_m(\Delta) \end{bmatrix}. \tag{3.34}$$

The limits of integration are $(-\pi, \pi)$ in discrete time and $(-\infty, +\infty)$ in continuous time. It is convenient to use matrix notations. Introduce the $m \times m$ matrix

$$F(\Delta) := \left[\mathrm{E}\{\hat{y}_k(\Delta)\overline{\hat{y}_j(\Delta)}\} \right]_{k,j=1}^{m}, \tag{3.35}$$

where Δ is a Borel set in $[-\pi, \pi]$. Then $F(\Delta)^* = F(\Delta)$, i.e., $F(\Delta)$ is *Hermitian*. Moreover, by Schwartz' inequality,

$$|F_{kj}(\Delta)| \le \|\hat{y}_k(\Delta)\| \|\hat{y}_j(\Delta)\| = \Lambda_{kk}(0)^{1/2} \Lambda_{jj}(0)^{1/2}$$

so that $F(\Delta)$ is bounded for all Borel subsets Δ.

Since for any $a \in \mathbb{C}^m$, $a^* F\, a$ is the spectral measure of the scalar process $a^* y(t)$, it follows also immediately that F is a positive semidefinite, σ-additive function of Δ, i.e., a *matrix measure*. We shall call F (or dF) *the spectral matrix measure of the process* y. Naturally, to the matrix measure F we may associate an equivalence class of Hermitian matrix valued functions $\theta \to F(\theta)$, each defined modulo an additive arbitrary constant matrix, which are monotonic nondecreasing in the sense that $F(\theta_2) - F(\theta_1) \ge 0$ (positive semidefinite) for $\theta_2 \ge \theta_1$. The vector-valued generalization provides readily the representation of the covariance matrix of the process as a Fourier-like integral of the form

$$\Lambda(\tau) = \int_{-\pi}^{\pi} e^{i\theta\tau} dF(\theta), \quad \tau \in \mathbb{Z}; \qquad \Lambda(\tau) = \int_{-\infty}^{\infty} e^{i\omega\tau} dF(\omega), \quad \tau \in \mathbb{R},$$

where we have taken the liberty of denoting by the same symbol dF the two (obviously different) matrix measures of discrete-time and of continuous time processes. These are the matrix versions of the Herglotz and Bochner Theorems.

As in the scalar case we have the canonical decomposition

$$F = F_1 + F_2,$$

where F_1 is the absolutely continuous component and F_2 the singular part of F.

The absolutely continuous part is the indefinite integral of a *spectral density matrix* Φ which is Hermitian and positive semidefinite ($\Phi(\theta) \ge 0$, $\theta \in [-\pi, \pi]$). For processes taking values in $\mathbb{R}^m$, which will be also called *real* for short, the symmetry relation (3.31) translates into $F_{kj}(\Delta) = F_{jk}(-\Delta)$, $k, j = 1, 2, \ldots, m$, which, for the spectral density matrix reads $\Phi(\theta)^* = \Phi(-\theta)'$ or, equivalently,

$\Phi(-\theta)' = \Phi(\theta)$. With the notational convention described in Remark 3.2.3 this can be rewritten as

$$\Phi(e^{-i\theta})' = \Phi(e^{i\theta}). \tag{3.36}$$

This property is sometimes called *parahermitian symmetry*.

The vector analogue of the spectral isomorphism theorem requires a preliminary brief digression on integration with respect to the matrix measure F. Deterministic vector-valued functions will be written as *row vectors* hereafter. As in the scalar case, the integral of f with respect to F is first defined for simple m-dimensional functions

$$f(\theta) = \sum_{k=1}^{N} c_k I_{\Delta_k}(\theta), \quad \Delta_k \subset [-\pi, \pi],\ \Delta_k \cap \Delta_j = \emptyset\ k \neq j,$$

where c_k are *row* vectors in $\mathbb{C}^m$, as

$$\int_{-\pi}^{\pi} f(\theta) dF(\theta) := \sum_{k=1}^{N} c_k F(\Delta_k)$$

and is then extended to all measurable m-dimensional functions by the usual limiting procedure. This clearly applies to matrix-valued simple functions as well. The integral of bilinear (or quadratic) forms of the type

$$\int_{-\pi}^{\pi} f(\theta) dF(\theta) g(\theta)^*$$

may also be defined in terms of sequences of vector-valued simple functions (f_k) and (g_j) approximating f and g (so that $(f_k g_j^*)$ is a sequence of simple matrix functions approximating fg^*) as the limit

$$\int_{-\pi}^{\pi} f(\theta) dF(\theta) g(\theta)^* := \lim_{k,j\to\infty} \operatorname{trace} \int_{-\pi}^{\pi} g_j(\theta)^* f_k(\theta) dF(\theta).$$

The space of m-dimensional square integrable functions with respect to the matrix measure F is denoted by $L_m^2([-\pi, \pi], dF)$. It has been shown [79, p. 1349] that this space is complete and hence a Hilbert space with respect to the scalar product

$$\langle f, g \rangle := \int_{-\pi}^{\pi} f(\theta) dF(\theta) g(\theta)^*, \tag{3.37}$$

provided one agrees to identify vector functions whose difference has norm equal to zero with each other. Functions $f_1,\ f_2$ such that $\|f_1 - f_2\| = 0$ are said to be

equal F-almost everywhere. If F happens to be singular on nontrivial subsets, it may happen that f_1 and f_2 are equal F-almost everywhere but are widely different pointwise.

The fundamental isometric property of the stochastic integral with respect to a vector stochastic measure can now be stated in the following form,

$$\mathrm{E}\{\mathcal{I}_{\hat{y}}(f)\mathcal{I}_{\hat{y}}(g)^*\} = \mathrm{E}\left\{\int_{-\pi}^{\pi} f(\theta)\, d\,\hat{y}(\theta)\left[\int_{-\pi}^{\pi} g(\theta)\, d\,\hat{y}(\theta)\right]^*\right\} = \int_{-\pi}^{\pi} f(\theta) dF(\theta) g(\theta)^* = \langle f, g\rangle_{L^2_m([-\pi,\pi],dF)}, \tag{3.38}$$

where f and g are functions in $L^2_m([-\pi,\pi], dF)$ and F is the spectral matrix measure of $\hat{y}$.

The vector version of the spectral isomorphism theorem follows.

Theorem 3.4.1 (Spectral Isomorphism Theorem). *Let y be an m-dimensional stationary process with stochastic Fourier transform $\hat{y}$. Then every $\xi \in \mathbf{H}(y)$ can be written as a stochastic integral $\mathcal{I}_{\hat{y}}(\hat{f})$ of a unique function $\hat{f} \in L^2_m\{[-\pi,\pi], dF\}$. In fact, the map $\mathcal{I}_{\hat{y}} : L^2_m\{[-\pi,\pi], dF\} \to \mathbf{H}(y)$ is unitary. It maps the elementary exponential function $[0,\ldots,e_t,\ldots,0]$ ($e_t(\theta) = e^{i\theta t}$ in the k-th place) into the random variables $y_k(t)$, for $k = 1,2,\ldots,m$, and transforms the shift operator $\mathcal{U}$ of the process y into the operator M_e, the multiplication by the exponential function $e : \theta \to e^{i\theta}$, acting on $L^2_m\{[-\pi,\pi], dF\}$. In other words, the diagram*

$$\begin{array}{ccc} \mathbf{H}(y) & \xrightarrow{\mathcal{U}} & \mathbf{H}(y) \\ {\scriptstyle \mathcal{I}_{\hat{y}}}\uparrow & & \uparrow{\scriptstyle \mathcal{I}_{\hat{y}}} \\ L^2_m\{[-\pi,\pi], dF\} & \xrightarrow[M_e]{} & L^2_m\{[-\pi,\pi], dF\} \end{array}$$

commutes. A totally analogous statement holds for continuous-time processes provided one substitutes $[-\pi,\pi]$ for $(-\infty,+\infty)$, the unitary operator $\mathcal{U}$ for the shift group $\{\mathcal{U}_t\,; t \in \mathbb{R}\}$, and $e^{i\theta}$ for $e^{i\omega t}$, $t \in \mathbb{R}$.

3.5 Functionals of White Noise

Let $\ell^2_m \equiv \ell^2_m(\mathbb{Z})$ be the Hilbert space of square summable m-dimensional functions (sequences) $f : \mathbb{Z} \to \mathbb{C}^m$, endowed with the inner product

$$\langle f, g \rangle := \sum_{-\infty}^{+\infty} f(t) g(t)^*.$$

In the engineering literature ℓ_m^2 is sometimes referred to as the *space of signals with finite energy*, the energy being just the squared norm

$$\| f \|^2 = \sum_{-\infty}^{+\infty} |f(t)|^2,$$

where $|\cdot|$ denotes the Euclidean norm. For $m = 1$ (scalar square-summable sequences) the subscript will be dropped.

Functions which are zero for negative [positive] values of the argument, $f(t) = 0$, $t < 0$, $[t > 0]$ are called *causal* [*anticausal*]. If $f(t) = 0$, $t \leq 0$, $[t \geq 0]$, f is called *strictly causal* [*strictly anticausal*]. The subspaces of causal and anticausal functions in $\ell_m^2(\mathbb{Z})$ will be denoted by the symbols ℓ_m^{2+} and by ℓ_m^{2-} respectively. They are clearly isomorphic to $\ell_m^2(\mathbb{Z}_+)$ and to $\ell_m^2(\mathbb{Z}_-)$.

An m-dimensional *white noise process* w, is just a stationary vector process whose components are pairwise uncorrelated so that

$$E\{w(t)w(s)^*\} = Q\delta(t-s), \tag{3.39}$$

where the *variance matrix* Q is a Hermitian positive-semidefinite matrix. In the following we shall assume that Q is nonsingular and denote by $Q^{1/2}$ an arbitrary square root of Q, i.e., a square matrix A satisfying $AA' = Q$. Therefore we may as well consider the *normalized white noise* process $\tilde{w} := Q^{-1/2}w$ which has variance matrix equal to the identity and obviously generates the same Hilbert space $\mathbf{H}(w)$.

Note that if w has a singular covariance matrix, there are matrices A, which are rectangular but with linearly independent columns, such that $AA' = Q$ is a rank factorization. In this case define $u := A^{-L}w$ where $^{-L}$ denotes left-inverse and set $\tilde{w} := Au$ where the dimension of u is equal to the rank of Q. Since $(I - AA^{-L})Q = (I - AA^{-L})AA' = 0$, the difference $w - \tilde{w} = (I - AA^{-L})w$ has covariance zero and hence $\tilde{w} = w = Au$ almost surely. It follows that $\mathbf{H}(w) = \mathbf{H}(u)$, i.e., the space can also be generated by a normalized white noise u of a smaller dimension.

The elements (linear functionals) in the Hilbert space $\mathbf{H}(w)$ of a white noise process have an explicit and particularly simple form. The following representation theorem will describe their structure. Although rather elementary, this result will turn out to be extremely useful.

Theorem 3.5.1. *Let w be an m-dimensional normalized white noise process. The linear functionals $\eta \in \mathbf{H}(w)$ have the form*

$$\eta = \sum_{s=-\infty}^{+\infty} f(-s)w(s), \quad f \in \ell^2_m, \tag{3.40}$$

where the function f is uniquely determined by η. The linear map $\mathcal{I}_w : \ell^2_m \to \mathbf{H}(w)$ defined by Eq. (3.40) is unitary and transforms the translation operator T in ℓ^2_m into the shift $\mathcal{U}$ acting on random variables of $\mathbf{H}(w)$, namely, if $[T^t f](s) = f(t+s)$, then

$$\eta(t) := \mathcal{U}^t \eta = \sum_{s=-\infty}^{+\infty} f(t-s)w(s) = \mathcal{I}_w(T^t f). \tag{3.41}$$

Note that we have been abusing notations as the symbol $\mathcal{I}_w$ denotes a transformation which strictly speaking is not a stochastic integral (but is the discrete-time analog of one).

Proof. The proof is particularly simple in the scalar case. Then the representation formula (3.40) follows readily from the fact that the random variables $\{w(s) \mid s \in \mathbb{Z}\}$ form an orthonormal basis for the Hilbert space $\mathbf{H}(w)$. In fact,

$$f(-s) = E\{\eta \overline{w(s)}\}$$

is just the s-th Fourier coefficient of η with respect to that basis. It is well-known that these coefficients are unique and form a square summable sequence. The last part of the statement also follows since $\mathcal{U}^{-t} w(s) = w(s-t)$ and

$$E\{\eta(t)\overline{w(s)}\} = \langle \mathcal{U}^t \eta, w(s) \rangle = \langle \eta, \mathcal{U}^{-t} w(s) \rangle = f(t-s).$$

We shall leave the details of the generalization of this argument to the vector case to the reader. □

Note that the continuous-time analog of Theorem 3.5.1 is contained as a particular case in Theorem 3.1.3: we just need to take ζ to be (the orthogonal stochastic measure defined by) an m-dimensional normalized Wiener process w. Then the following is just an immediate corollary of that result.

Corollary 3.5.2. *Let w be an m-dimensional normalized Wiener process. The linear functionals $\eta \in \mathbf{H}(dw)$ have the form*

$$\eta = \int_{-\infty}^{+\infty} f(-s)\, dw(s), \quad f \in L^2_m(\mathbb{R}), \tag{3.42}$$

where the function f is uniquely determined by η. The linear map $\mathcal{I}_w : L^2_m(\mathbb{R}) \to \mathbf{H}(dw)$ defined by Eq. (3.42) is unitary and transforms the translation operator T_t in L^2_m into the shift $\mathcal{U}_t$ acting on random variables of $\mathbf{H}(dw)$, namely, if $[T_t f](s) = f(t+s)$, then

$$\eta(t) := \mathcal{U}^t \eta = \int_{-\infty}^{+\infty} f(t-s)dw(s) = \mathfrak{I}_w(T_t f). \tag{3.43}$$

For white noise processes we have two representation theorems of $\mathbf{H}(w)$: the general spectral representation Theorem 3.4.1 and the time-domain representation that we have just seen. These two representations in the frequency and in the time domain are related by the Fourier transform.

3.5.1 The Fourier Transform

Related to the well-known fact that the trigonometric functions

$$e_t(\theta) := e^{i\theta t}, \quad t \in \mathbb{Z},$$

form a complete orthonormal system (an orthonormal basis) in $L^2([-\pi, \pi], \frac{d\theta}{2\pi})$, is the following basic result in harmonic analysis (the so-called *Fourier-Plancherel Theorem*).

Theorem 3.5.3. *The Fourier transform*

$$\mathfrak{F} : \ell^2_m \to L^2_m([-\pi, \pi], \frac{d\theta}{2\pi}), \quad \mathfrak{F}(f) := \sum_{t=-\infty}^{+\infty} e^{-i\theta t} f(t),$$

where the sum is convergent for all $f \in \ell^2_m$ in the topology of the space $L^2_m([-\pi, \pi], \frac{d\theta}{2\pi})$, is a norm preserving and surjective map, i.e., a unitary map.

The norm preserving property

$$\sum_{t=-\infty}^{+\infty} |f(t)|^2 = \int_{-\pi}^{\pi} |\hat{f}(\theta)|^2 \frac{d\theta}{2\pi}, \quad \text{where } \hat{f} = \mathfrak{F}(f),$$

is known as *Parseval's identity*. It is easy to check that this property holds for functions (sequences) with compact support and, since these sequences are obviously dense in ℓ^2_m, by invoking Theorem B.2.7 the theorem can be proved by the same isometric extension argument used for the definition of the stochastic integral.

One reason for the importance of the Fourier transform in the study of dynamical models of time sequences, is the fact that the *translation operator* T in ℓ^2_m,

$$T(f)(t) := f(t+1)$$

corresponds, in the frequency domain, to the algebraic operation of multiplication by the scalar exponential function $e : \theta \to e^{i\theta}$, acting on $L^2_m([-\pi, \pi], \frac{d\theta}{2\pi})$. In other

words $\mathfrak{F}(Tf) = M_e\mathfrak{F}(f)$ where M_e is the multiplication operator by the function e; i.e. $(M_e\hat{f})(\theta) = e^{i\theta}\hat{f}(\theta)$. The importance of this property and its numerous consequences in the study of deterministic signals and systems are well established.

In the continuous time case there is a perfectly analogous version of Theorem 3.5.3 which is also known as the *Fourier-Plancherel Theorem*.

Theorem 3.5.4. *Let $\mathbb{I}$ denote the imaginary axis. The Fourier transform*

$$\mathfrak{F} : L^2_m(\mathbb{R}) \to L^2_m(\mathbb{I}, \frac{d\omega}{2\pi}), \quad \mathfrak{F}(f) := \int_{-\infty}^{+\infty} e^{-i\omega t} f(t)\, dt,$$

where the integral is defined as a limit in the topology of the space $L^2_m(\mathbb{I}, \frac{d\omega}{2\pi})$, is well-defined for all $f \in L^2_m(\mathbb{R})$ and is a norm preserving and surjective map, i.e., a unitary map.

Again the norm preserving property

$$\int_{-\infty}^{+\infty} |f(t)|^2\, dt = \int_{-\infty}^{+\infty} |\hat{f}(i\omega)|^2 \frac{d\omega}{2\pi}, \quad \text{where } \hat{f} = \mathfrak{F}(f),$$

is known as *Parseval's identity*. The *translation operator* T_t, $t \in \mathbb{R}$, acting in $L^2_m(\mathbb{R})$, is defined as

$$T_t(f)(s) := f(t+s), \quad s \in \mathbb{R},$$

and corresponds, in the frequency domain, to the algebraic operation of multiplication by the scalar exponential function $e_t : i\omega \to e^{i\omega t}$, acting on $L^2_m(\mathbb{I}, \frac{d\omega}{2\pi})$. In other words, $\mathfrak{F}(T_t f) = M_{e_t}\mathfrak{F}(f)$ where M_{e_t} is multiplication by the function e_t; i.e., $(M_{e_t}\hat{f})(i\omega) = e^{i\omega t}\hat{f}(i\omega)$. The family of translations $\{T_t, t \in \mathbb{R}\}$ forms a group of unitary operators in $L^2_m(\mathbb{R})$ which, via the Fourier transform, corresponds (in fact is unitarily equivalent) to the unitary group of multiplication operators by $e^{i\omega t}$ acting in $L^2_m(\mathbb{I}, \frac{d\omega}{2\pi})$.

The following fundamental representation theorem relates the spectral representation of random functionals of white noise in $\mathbf{H}(w)$ to the Fourier-Plancherel transform.

Theorem 3.5.5. *Let w be an m-dimensional normalized white noise process. The unitary representation map $\mathcal{I}_w : \ell^2_m \to \mathbf{H}(w)$ defined by Eq. (3.40) admits a factorization as the composite map*

$$\mathcal{I}_w = \mathcal{I}_{\hat{w}}\mathfrak{F} \tag{3.44}$$

i.e., the frequency-domain representative function of any linear functional in $\mathbf{H}(w)$ is just the Fourier transform of the time-domain function f in (3.40). In other words $\eta = \mathcal{I}_{\hat{w}}(\hat{f}) = \mathcal{I}_w(f)$ if and only if $\hat{f} = \mathfrak{F}f$. In fact, the two unitary representation maps $\mathcal{I}_{\hat{w}}$ and $\mathcal{I}_w$ are related as in the commutative diagram

$$\begin{array}{ccc}
\mathbf{H}(w) & \xrightarrow{\mathcal{U}} & \mathbf{H}(w) \\
{\scriptstyle \mathcal{I}_{\hat{w}}}\uparrow & & \uparrow{\scriptstyle \mathcal{I}_{\hat{w}}} \\
L^2_m([-\pi,\pi],\, \frac{d\theta}{2\pi}) & \xrightarrow[M_{e^{i\theta}}]{} & L^2_m([-\pi,\pi],\, \frac{d\theta}{2\pi}) \\
{\scriptstyle \mathfrak{F}}\uparrow & & \uparrow{\scriptstyle \mathfrak{F}} \\
\ell^2_m(\mathbb{Z}) & \xrightarrow[T]{} & \ell^2_m(\mathbb{Z})
\end{array}$$

Proof. The frequency-domain isomorphism $\mathcal{I}_{\hat{w}}$ maps trigonometric polynomials $p(\theta) = \sum_{-N}^{M} f(-k)e^{i\theta k}$ into finite linear combinations $\eta = \sum_{-N}^{M} f(-k)w(k) = \mathcal{I}_w(f)$, where f is an ℓ^2 function of bounded support. It is obvious that $p(\theta) = \sum_{-M}^{N} f(k)e^{-i\theta k}$ is the Fourier transform of f, i.e., $p = \hat{f}$. Hence it follows that

$$\mathcal{I}_w(f) = \mathcal{I}_{\hat{w}}(\hat{f}) = \mathcal{I}_{\hat{w}}(\mathfrak{F}f)$$

for the dense linear manifold of finite support functions f. Since both maps $\mathcal{I}_w$ and $\mathcal{I}_{\hat{w}}\mathfrak{F}$ are unitary, (3.44) follows. The rest follows by well-know properties of the Fourier transform.

□

The continuous-time analog is immediate and will be stated without proof.

Theorem 3.5.6. *Let w be an m-dimensional normalized Wiener process. The unitary representation map $\mathcal{I}_w : L^2_m(\mathbb{R}) \to \mathbf{H}(dw)$ defined in Corollary 3.5.2 factors exactly as the composite map (3.44) in Theorem 3.5.5. In other words, the representative function in the frequency-domain of any linear functional in $\mathbf{H}(dw)$ is just the Fourier transform of the time-domain function f in (3.42). Hence, $\eta = \mathcal{I}_{\hat{w}}(\hat{f}) = \mathcal{I}_w(f)$ if and only if $\hat{f} = \mathfrak{F}f$. In fact, the two representation maps $\mathcal{I}_{\hat{w}}$ and $\mathcal{I}_w$ are related as in the commutative diagram*

$$\begin{array}{ccc}
\mathbf{H}(dw) & \xrightarrow{\mathcal{U}_t} & \mathbf{H}(dw) \\
{\scriptstyle \mathcal{I}_{\hat{w}}}\uparrow & & \uparrow{\scriptstyle \mathcal{I}_{\hat{w}}} \\
L^2_m(\mathbb{I},\, \frac{d\omega}{2\pi}) & \xrightarrow[M_{e^{i\omega t}}]{} & L^2_m(\mathbb{I},\, \frac{d\omega}{2\pi}) \\
{\scriptstyle \mathfrak{F}}\uparrow & & \uparrow{\scriptstyle \mathfrak{F}} \\
L^2_m(\mathbb{R}) & \xrightarrow[T_t]{} & L^2_m(\mathbb{R})
\end{array}$$

3.6 Spectral Representation of Stationary Increment Processes

Let $I_{[\omega_1,\omega_2]}(i\omega)$ be the indicator function of a finite subinterval $[i\omega_1, i\omega_2]$ of the imaginary axis (equal to one for $\omega \in [\omega_1, \omega_2]$ and zero otherwise) and consider the elementary identity

$$\frac{e^{-i\omega_2 t} - e^{-i\omega_1 t}}{-2\pi i t} = (\mathfrak{F}^{-1} I_{[\omega_1,\omega_2]})(-t)\,. \tag{3.45}$$

Since trivially these are square integrable functions, given a p-dimensional Wiener process dw we can define a process $\hat{w}$ on the imaginary axis $\mathbb{I}$ with increments

$$\hat{w}(i\omega_2) - \hat{w}(i\omega_1) = \int_{-\infty}^{\infty} \frac{e^{-i\omega_2 t} - e^{-i\omega_1 t}}{-2\pi i t} dw(t). \tag{3.46}$$

Then, since $\mathfrak{F}^{-1}$ is unitary (Theorem 3.5.4), and hence $\langle \mathfrak{F}^{-1}\hat{f}, \mathfrak{F}^{-1}\hat{g}\rangle_1 = \langle \hat{f}, \hat{g}\rangle_2$, with $\langle\cdot,\cdot\rangle_1$ and $\langle\cdot,\cdot\rangle_2$ the inner products in $L^2_p(\mathbb{R})$ and $L^2_p(\mathbb{I}, \frac{d\omega}{2\pi})$ respectively, (3.45) and (3.46) imply that

$$\mathrm{E}\{[\hat{w}(i\omega_2) - \hat{w}(i\omega_1)][\hat{w}(i\omega_4) - \hat{w}(i\omega_3)]^*\} = I_p \int_{-\infty}^{\infty} I_{[\omega_1,\omega_2]}(i\omega) I_{[\omega_3,\omega_4]}(i\omega) \frac{d\omega}{2\pi},$$

and hence it follows that the process $\hat{w}$ has orthogonal increments. In fact,

$$\mathrm{E}\{d\hat{w} d\hat{w}^*\} = I_p \frac{d\omega}{2\pi}. \tag{3.47}$$

Therefore, $d\hat{w}$ is a p-dimensional Wiener process on the imaginary axis. Now, (3.46) may be written

$$\int_{-\infty}^{\infty} I_{[\omega_1,\omega_2]}(i\omega) d\hat{w}(i\omega) = \int_{-\infty}^{\infty} (\mathfrak{F}^{-1} I_{[\omega_1,\omega_2]})(-t) dw(t),$$

and, since the indicator functions are dense in L^2, one has, for all $f \in L^2(\mathbb{R})$,

$$\int_{-\infty}^{\infty} \hat{f}(i\omega) d\hat{w} = \int_{-\infty}^{\infty} f(-t) dw\,, \tag{3.48}$$

the function $\hat{f}$ being the Fourier-Plancherel transform of f. Incidentally, we have just proven that the spectral representation map $\mathfrak{I}_{\hat{w}} : L^2_p(\mathbb{I}, \frac{d\omega}{2\pi}) \to \mathbf{H}(dw)$ defined by

$$\mathfrak{I}_{\hat{w}} \hat{f} = \int_{-\infty}^{\infty} \hat{f}(i\omega) d\hat{w}(i\omega)$$

factorizes as in (3.44) in Theorem 3.5.6. Dually, by choosing f to be the indicator function of the interval $[t_1, t_2]$, (3.48) yields

$$w(t_2) - w(t_1) = \int_{-\infty}^{\infty} \frac{e^{i\omega t_2} - e^{i\omega t_1}}{i\omega} d\hat{w}(i\omega). \tag{3.49}$$

This is a particular instance of *spectral representation of a stationary increments process*; in fact, of the stationary increments (Wiener) process dw [77]. Note that the *spectral measure of* dw is also of the Wiener type, being precisely the orthogonal random measure $d\hat{w}$ defined in (3.46).

More generally, one can prove the following result.

Theorem 3.6.1. *Every $\mathbb{R}^m$-valued process with finite second moments and continuous stationary increments dz admits a* spectral representation

$$z(t) - z(s) = \int_{-\infty}^{+\infty} \frac{e^{i\omega t} - e^{i\omega s}}{i\omega} d\hat{z}(i\omega)\ , \quad t, s \in \mathbb{R}, \tag{3.50}$$

where $d\hat{z}$ is an m-dimensional orthogonal random measure (or an orthogonal increments process) on the imaginary axis $\mathbb{I}$ uniquely determined by dz. The matrix spectral distribution of dz, defined by

$$\mathrm{E}\{d\hat{z}(i\omega) d\hat{z}(i\omega)^*\} = dZ(i\omega), \tag{3.51}$$

is a (not necessarily finite) nonnegative definite Hermitian matrix measure on the Borel sets of the imaginary axis.

The orthogonal stochastic measure $d\hat{z}$ will also be called the *Fourier transform of* dz.

Example 3.6.2. As an example consider the process dz defined as the output of the linear stochastic system

$$\begin{cases} dx = Axdt + Bdw \\ dz = Cxdt + Ddw \end{cases}, \tag{3.52}$$

where all eigenvalues of the matrix A have negative real parts. In the time domain (3.52) has the following solution

$$x(t) = \int_{-\infty}^{t} e^{A(t-\tau)} B\, dw, \tag{3.53a}$$

$$z(t) - z(s) = \int_{s}^{t} Cx(\tau) d\tau + D[w(t) - w(s)]. \tag{3.53b}$$

Applying (3.48) to the first of these equations, we obtain

$$x(t) = \int_{-\infty}^{\infty} e^{i\omega t}(i\omega I - A)^{-1} B\, d\hat{w}, \tag{3.54}$$

which then inserted into (3.53b) together with (3.49) yields the spectral representation

$$z(t) - z(s) = \int_{-\infty}^{\infty} \frac{e^{i\omega t} - e^{i\omega s}}{i\omega} d\hat{z}(i\omega), \tag{3.55}$$

where $d\hat{z} = W(i\omega)d\hat{w}(i\omega)$, the matrix function W being the transfer function of the system (3.52) given by

$$W(s) = C(sI - A)^{-1}B + D, \tag{3.56}$$

which is the Laplace transform of the impulse response of the system (3.52). In this example dz has an absolutely continuous spectral distribution

$$\mathrm{E}\{d\hat{z}d\hat{z}^*\} = \Phi(i\omega)\frac{d\omega}{2\pi},$$

where the spectral density Φ is given by $\Phi(s) = W(s)W(-s)'$. Note that if $D \neq 0$ the spectral distribution is not a finite measure and hence an expression like $\int_{-\infty}^{\infty} e^{i\omega t} d\hat{z}$ does not make sense.

Proposition 3.6.3. *If the spectral measure $d\hat{z}$ in Theorem 3.6.1 is finite, the process dz has a (stationary) derivative in mean square, i.e. $dz(t) = y(t)dt$, with*

$$y(t) = \int_{-\infty}^{+\infty} e^{i\omega t} d\hat{z}, \tag{3.57}$$

in which case $d\hat{y} = d\hat{z}$.

Proof. Let $y(t)$ be defined as in (3.57), then

$$\frac{z(t+h) - z(t)}{h} - y(t) = \int_{-\infty}^{+\infty} \Delta_h(i\omega) e^{i\omega t}\, d\hat{z}(i\omega)\ ,$$

where the function

$$\Delta_h(i\omega) := \frac{e^{i\omega h} - 1}{i\omega h} - 1 = e^{i\omega h/2}\frac{\sin(\omega h/2)}{\omega h/2} - 1$$

converges boundedly pointwise to zero for $h \to 0$. □

3.7 Multiplicity and the Module Structure of $\mathbf{H}(y)$

Since the early 1960s there has been considerable interest in representations of a process y as a linear function of a random process of the simplest possible structure, by which term we shall here typically mean *white noise*, i.e., representations of the form

$$y(t) = \sum_{k=1}^{N} \sum_{s=-\infty}^{+\infty} h_k(t,s) w_k(s), \tag{3.58}$$

where $h_k(t,\cdot)$ are deterministic functions, the w_k are generally non-stationary, uncorrelated white noise processes (i.e. $\mathrm{E}\{w_k(t)w_j(s)\} = 0$ for all k, j and $t \neq s$) and the series is convergent in mean square. Special representations of this kind where the kernel function is *causal* (i.e. $h_k(t,s) = 0$ for $t > s$) have a special importance and motivated the work of H. Wold [314], who discovered the so called *Wold decomposition*, a causal representation of the type (3.58), which is valid for the special class of *purely non deterministic* stationary processes. We shall define and study these representations in detail in Chap. 4.

The smallest integer N (i.e. the smallest number of independent white noises) for which a representation of the type (3.58) holds is commonly called the *multiplicity* of the process y. As we shall demonstrate in this section, this is in harmony with the formal definition of multiplicity to be given on page 92. It can be shown, using spectral theory of linear operators in Hilbert space, that under very general conditions representations of the type (3.58) for second-order processes indeed exist. In general however the integer N in the sum may be infinite even for a scalar process y. Moreover, the support $T_k \subset \mathbb{Z}$ of each white noise process w_k (i.e. the subset of $\mathbb{Z}$ where the variance function $\lambda_k(t) = \mathrm{E}\, w_k(t)^2$ is nonzero) is in general not the whole line and in fact may depend on k so the number of terms w_k in the sum in general varies with t. However one can always arrange things in such a way that $T_1 \supseteq T_2 \supseteq \ldots \supseteq T_N$. With this proviso, we may rewrite (3.58) in matrix form

$$y(t) = \sum_{s=-\infty}^{+\infty} H(t,s) w(s) \tag{3.59}$$

in terms of an N-dimensional nonstationary white noise process w. If

$$\mathbf{H}(y) = \bigoplus_{k=1}^{N} \mathbf{H}(w_k) = \mathbf{H}(w), \tag{3.60}$$

the process will be called *orthogonalizable*. This means that the components $w_k(t)$, $k = 1, 2, \ldots, N$, $t \in \mathbb{Z}$, form an orthogonal basis in $\mathbf{H}(y)$.

It turns out that, when y is wide-sense *stationary* and admits a representation of the type (3.58), then the multiplicity N is always finite and smaller or equal to the dimension m of $y(t)$. The white noises w_k can also be taken to be stationary, and all supports T_k coincide with the entire time axis $\mathbb{Z}$. This is so due to the fact that $y(t)$ propagates in time by the action of the unitary operator $\mathcal{U}$, called the *shift* of the process, which was defined in Sect. 2.5.1. This fact can be justified via traditional spectral theory of unitary operators as done in, e.g., [270], but has more to do with a basic algebraic structure induced on $\mathbf{H}(y)$ by the action of the operator $\mathcal{U}$. This algebraic structure is essentially a *module structure* which, as pointed out by Kalman, underlies many basic constructions of linear systems theory.

A *module* is an algebraic structure which generalizes vector spaces. It is a set with two binary operations: addition, with respect to which it is an ordinary Abelian Group, and multiplication by scalars. The fundamental difference is that the scalars live in a *ring* R, instead of a field like $\mathbb{R}$ or $\mathbb{C}$. A good survey of module theory with a view to applications to system theory can be found in the first chapter of Fuhrmann's book [104].

In this section we shall discuss the notion of multiplicity of a stationary process from a module theoretic point of view. This approach will reveal the true essence of the concept and hopefully clarify some misconceptions that are rather common in the literature, in particular in relation to the notions of *rank* and spectral factorization of stationary processes.

3.7.1 Definition of Multiplicity and the Module Structure of $\mathbf{H}(y)$

An important property of the Hilbert space $\mathbf{H}(y)$ generated by a stationary stochastic process y is to be *finitely generated* by the shift $\mathcal{U}$ in the following sense: there is a *finite* number of *generators* namely elements $y_1, y_2, \ldots, y_m \in \mathbf{H}(y)$ which are "cyclic" for the shift, i.e., have the property

$$\overline{\text{span}}\{\mathcal{U}^t y_k \mid k = 1, 2, \ldots, m,\ t \in \mathbb{Z}\} = \mathbf{H}(y), \tag{3.61}$$

where $\overline{\text{span}}$ means closed linear hull. The cardinality of the smallest set of generators is called the *multiplicity* of the shift $\mathcal{U}$ on the Hilbert space $\mathbf{H}(y)$; see, e.g., [133], [104, p. 105]. We shall call this number the *multiplicity of the process* y. Note that there are m natural generators in $\mathbf{H}(y)$, namely $y_k = y_k(0)$, $k = 1, 2, \ldots, m$, and hence a process of dimension m has a finite multiplicity, less than or equal to m.

The main fact here is that the shift operator acting on the Hilbert space $\mathbf{H}(y)$ induces a natural *module structure* on this space. The concept of multiplicity has to do with the algebraic concept of *basis* in module theory.

The starting point to see this is the observation that there is a natural multiplication $p \cdot \eta$ between trigonometric polynomials

$$p(e^{i\theta}) := \sum_{k=k_0}^{k_1} p_k e^{ik\theta}, \quad k_0 \le k_1, \in \mathbb{Z},$$

and elements $\eta \in \mathbf{H}(y)$, defined as

$$p \cdot \eta := p(\mathcal{U}) \cdot \eta = \left[\sum_{k=k_0}^{k_1} p_k \mathcal{U}^k \right] \eta. \tag{3.62}$$

It is trivial to check that trigonometric polynomials form a ring and that the algebraic module axioms are satisfied. Naturally, the ring of trigonometric polynomials should be extended in order to make multiplication by scalars a continuous operation in $\mathbf{H}(y)$, which thereby becomes a *Hilbert module*. This is accomplished as follows.

Every element $\eta \in \mathbf{H}(y)$ has a spectral representation

$$\eta = \int_{-\pi}^{\pi} \hat{f}(e^{i\theta})\, d\hat{y}(e^{i\theta}),$$

where $\hat{f} \in L^2_m\{[-\pi, \pi], dF\}$ is the (unique dF-almost everywhere) spectral representative of η with respect to $d\hat{y}$ (Theorem 3.3.3). Using this representation one can write (3.62) in the spectral domain as

$$p \cdot \eta = \int_{-\pi}^{\pi} p(e^{i\theta}) \hat{f}(e^{i\theta})\, d\hat{y}(e^{i\theta}).$$

The spectral representation map $\mathfrak{I}_{\hat{y}} : L^2_m\{[-\pi, \pi], dF\} \to \mathbf{H}(y)$, by which each function $\hat{f} \in L^2_m\{[-\pi, \pi], dF\}$ is mapped into the Wiener integral $\int_{-\pi}^{\pi} \hat{f}(e^{i\theta})\, d\hat{y}(e^{i\theta})$, is unitary and satisfies the intertwining relation

$$\mathfrak{I}_{\hat{y}} M_{e^{i\theta}} = \mathcal{U} \mathfrak{I}_{\hat{y}},$$

where $M_{e^{i\theta}}$ is the multiplication operator by the function $\theta \to e^{i\theta}$. Hence $\mathfrak{I}_{\hat{y}}$ is an algebraic unitary isomorphism of the modules $\mathbf{H}(y)$ and $L^2_m\{[-\pi, \pi], dF\}$ having the ring of trigonometric polynomials as a ring of scalars.

Now, it is a well-known consequence of Weierstrass' Approximation Theorem that the trigonometric polynomials are dense in the sup norm in the space of continuous functions on the interval $[-\pi, \pi]$. Consequently, it follows by Lusin's theorem (see e.g. [280, pp. 56–57]) that any function $\varphi \in L^\infty[-\pi, \pi]$ is the limit in L^∞ of sequences of trigonometric polynomials (p_k), and, since

$$\left\| \int_{-\pi}^{\pi} [\varphi(e^{i\theta}) - p_k(e^{i\theta})] \hat{f}(e^{i\theta})\, d\hat{y}(e^{i\theta}) \right\| \le \|\varphi - p_k\|_{L^\infty} \|\hat{f}\|_{L^2_m\{[-\pi,\pi],dF\}} \to 0$$

as $k \to \infty$, we may define the product

$$\varphi(\mathcal{U}) \cdot \eta := \lim_{k\to\infty} p_k(\mathcal{U}) \cdot \eta, \quad \eta \in \mathbf{H}(y)$$

for every φ in $L^\infty[-\pi, \pi]$ as a limit in $\mathbf{H}(y)$. Thereby $L^\infty[-\pi, \pi]$ becomes the extended ring of scalars by which we may multiply elements of $\mathbf{H}(y)$. It is then immediate to check that multiplication by scalars is continuous and thereby $\mathbf{H}(y)$ becomes a *bona fide* Hilbert module.

Proposition 3.7.1. *Endowed with the multiplication* (3.62), $\mathbf{H}(y)$ *becomes a Hilbert module, unitarily isomorphic, via the spectral representation map* $\mathcal{I}_{\hat{y}}$, *to* $L^2_m\{[-\pi, \pi], dF\}$ *as a* $L^\infty[-\pi, \pi]$*-module.*

It follows readily from (3.61) that the module $\mathbf{H}(y)$ is in fact *free*, as it admits the m generators $y_1(0), y_2(0), \ldots, y_m(0)$. These generators correspond under the isomorphism to the m unit vector functions $e_1, e_2, \ldots, e_m$ in $L^2_m\{[-\pi, \pi], dF\}$, where the k-th component of e_k is identically equal to one while the others are zero almost everywhere.

A *submodule* of an R-module $\mathbf{M}$ is a subset $\mathbf{M}' \subset \mathbf{M}$ which remains invariant with respect to multiplication by elements of R, i.e., $\mathbf{M}' = R\mathbf{M}'$. Accordingly a subspace $\mathbf{H} \subset \mathbf{H}(y)$ is a submodule if it is invariant with respect to multiplication by all elements of $L^\infty[-\pi, \pi]$. By continuity, this happens if and only if

$$\overline{\text{span}}\,\{\mathcal{U}^k \eta,\ \eta \in \mathbf{H},\ k \in \mathbb{Z}\} = \mathbf{H}\,.$$

For this reason, submodules of $\mathbf{H}(y)$ are called *doubly invariant subspaces* (for the shift operator $\mathcal{U}$). The module-theoretic concept of *basis* corresponds to a set of generators of minimal cardinality. Hence the multiplicity of a stationary process is just the dimension of a basis for the Hilbert module $\mathbf{H}(y)$.

One would like to have a test to check if the generators $y_1(0), y_2(0), \ldots, y_m(0)$ form a basis. It should of course be appreciated that this question is more subtle than for vector spaces, since, for example, a module of dimension one can have infinitely many proper submodules (still of dimension one). Consider for example a scalar stationary white noise process w with spectral measure $d\hat{w}$. It is obvious that $w(0)$ is a generator of $\mathbf{H}(w)$; i.e., a basis for the Hilbert module $\mathbf{H}(w)$. Let us then define a stationary process y with spectral measure

$$d\hat{y} := I_\Delta d\hat{w},$$

where I_Δ is the indicator of some proper Borel subset $\Delta \subset [-\pi, \pi]$ of normalized Lebesgue measure $\frac{|\Delta|}{2\pi} \leq 1$. The question is if $y(0)$ is also a basis of $\mathbf{H}(w)$. In a vector space setting the answer would obviously be yes, but in the present setting the answer is generally negative.

Proposition 3.7.2. *Unless* Δ *has full Lebesgue measure,* $\mathbf{H}(y)$ *is a doubly invariant subspace properly contained in* $\mathbf{H}(w)$. *In fact, for any* $\varphi \in L^\infty[-\pi, \pi]$, *the*

stationary process y with spectral measure $d\hat{y} := \varphi d\hat{w}$ generates the whole space, i.e., $\mathbf{H}(y) = \mathbf{H}(w)$, if and only if φ is nonzero almost everywhere in $[-\pi, \pi]$.

Proof. The statement follows from a classical characterization of doubly invariant subspaces of $L^2[-\pi, \pi]$ due to Wiener, which can be found in, e.g., Helson's book [138, Theorem 2, p. 7], according to which all doubly invariant subspaces are of the form $I_\Delta L^2[-\pi, \pi]$. Hence a doubly invariant subspace is the whole of $L^2[-\pi, \pi]$ if and only if Δ has full Lebesgue measure (equivalently, is nonzero almost everywhere). Since every $\varphi \in L^\infty[-\pi, \pi]$ can be written as a product $\varphi I_{\Delta(\varphi)}$ where $\Delta(\varphi)$ is the essential support of φ, the result carries over to an arbitrary φ. □

Put in systems-theoretic language, a stochastic process generated by filtering a white noise w with a filter φ, cannot generate the whole space $\mathbf{H}(w)$ unless φ has constant rank (equal to one) almost everywhere on the unit circle $[-\pi, \pi]$.

3.7.2 Bases and Spectral Factorization

We shall say that two jointly stationary vector processes u and y, which generate the same Hilbert space are *equivalent*. In order to characterize equivalent processes we shall introduce a concept which generalizes absolute continuity to matrix measures.

Definition 3.7.3. Let dF_1 and dF_2 be $m \times m$ respectively $p \times p$ positive matrix measures on $[-\pi, \pi]$. We say that dF_1 is *absolutely continuous* with respect to dF_2 (notation: $dF_1 \ll dF_2$) if there is a measurable $m \times p$ matrix function M with rows M_k, $k = 1, 2, \ldots, m$, belonging to $L_p^2\{[-\pi, \pi], dF_2\}$ such that

$$dF_1 = M(e^{i\theta}) dF_2 M(e^{i\theta})^* . \tag{3.63}$$

If $dF_1 \ll dF_2$ and $dF_2 \ll dF_1$, we say that the two measures are *equivalent* and write $dF_1 \simeq dF_2$. In this case, there also exist a measurable matrix function N with rows N_j, $j = 1, 2, \ldots, p$, belonging to $L_m^2\{[-\pi, \pi], dF_1\}$, such that

$$dF_2 = N(e^{i\theta}) dF_1 N(e^{i\theta})^* . \tag{3.64}$$

Note that there may be nontrivial matrix functions Q such that, for example, $dF_2 = Q(e^{i\theta}) dF_2 Q(e^{i\theta})^*$, in which case $\hat{M} := MQ$ would also satisfy the factorization relation (3.63). Hence the functions M and N in (3.63) and (3.64), which hereafter will be called *spectral factors* (of dF_1 with respect to dF_2 and conversely), need not be unique.

Lemma 3.7.4. *Let u be a p-dimensional stationary processes with spectral distribution measures dF_u. Then, if $\mathbf{H}(y)$ is a submodule (doubly invariant subspace) of $\mathbf{H}(u)$ with generators $y_1(0), y_2(0), \ldots, y_m(0) \in \mathbf{H}(u)$, the matrix spectral distribution measure dF_y of the process $y(t) = \mathfrak{U}^t y(0)$ is absolutely continuous*

with respect to dF_u. Conversely, if some $m \times m$ spectral distribution matrix dF satisfies $dF \ll dF_u$, then there is an m-dimensional process y jointly stationary with u such that $dF_y = dF$, and $\mathbf{H}(y) \subset \mathbf{H}(u)$ is a doubly invariant subspace. If two stationary processes y and u are equivalent, i.e., $\mathbf{H}(y) = \mathbf{H}(u)$, then their spectral distribution measures dF_y and dF_u are also equivalent.

Proof. If $\mathbf{H}(y) = \overline{\text{span}}\{y_k(t)\,;\, k = 1, \dots, m,\, t \in \mathbb{Z}\} \subset \mathbf{H}(u)$ is an invariant subspace, then the random vector of generators $y(0) = [\, y_1(0)\ y_2(0), \dots, y_m(0)\,]'$ for $\mathbf{H}(y)$ can be written as $y(0) = \int_{-\pi}^{\pi} M(e^{i\theta}) d\hat{u}$ for some matrix function M whose rows belong to $L_p^2\{[-\pi, \pi], dF_u\}$. Hence $d\hat{y} = M(e^{i\theta}) d\hat{u}$ and $dF_y \ll dF_u$ readily follows. Similarly, if $\mathbf{H}(u) = \overline{\text{span}}\{u_k(t)\,;\, k = 1, \dots, p,\, t \in \mathbb{Z}\} \subset \mathbf{H}(y)$ then each random vector $u(t)$ has the spectral representation $u(t) = \int_{-\pi}^{\pi} e^{i\theta t} N(e^{i\theta}) d\hat{y}$ for some matrix function N whose rows belong to $L_p^2\{[-\pi, \pi], dF_y\}$ and hence $dF_u \ll dF_y$. Therefore $\mathbf{H}(y) = \mathbf{H}(u)$ implies $dF_y \simeq dF_u$. Conversely, assume there is an $m \times p$ matrix function M with rows M_k, $k = 1, 2, \dots, m$, belonging to $L_p^2\{[-\pi, \pi], dF_u\}$ such that

$$dF = M(e^{i\theta}) dF_u M(e^{i\theta})^*,$$

and define the random spectral measure $d\hat{y} := M(e^{i\theta}) d\hat{u}$. Then the corresponding stationary process y generates an invariant subspace of $\mathbf{H}(u)$ and has spectral distribution measure equal to dF. □

The following theorem states that left-invertible spectral factors produce equivalent processes.

Theorem 3.7.5. *Let dF_y be the spectral distribution measure of the process y, and assume that dF_u is a $p \times p$ positive matrix measure on $[-\pi, \pi]$ such that $dF_y \ll dF_u$, i.e., there is a $m \times p$ matrix function M with rows M_k, $k = 1, 2, \dots, m$, belonging to $L_p^2\{[-\pi, \pi], dF_u\}$ such that*

$$dF_y = M(e^{i\theta}) dF_u M(e^{i\theta})^*. \tag{3.65}$$

Assume M is left-invertible*, i.e., there is a $p \times m$ matrix function N with rows N_k, $k = 1, 2, \dots, p$, belonging to $L_p^2\{[-\pi, \pi], dF_y\}$ such that*

$$N(e^{i\theta}) M(e^{i\theta}) = I_p, \quad dF_u\text{-a.e.} \tag{3.66}$$

Then the stationary process $u(t) = \int_{-\pi}^{\pi} e^{i\theta t} d\hat{u}$ with random spectral measure $d\hat{u} := N(e^{i\theta}) d\hat{y}$ is jointly stationary with y, has spectral distribution measure dF_u, and is equivalent to y, i.e., $\mathbf{H}(y) = \mathbf{H}(u)$. The above holds for all functions $\hat{N}$ in the same equivalence class (mod dF_y)[4] of the function N.

[4] In other words such that $\int_{-\pi}^{\pi} [\hat{N}(e^{i\theta}) - N(e^{i\theta})] dF_y(e^{i\theta}) [\hat{N}(e^{i\theta}) - N(e^{i\theta})]^* = 0$.

Proof. Since $\mathrm{E}\{d\hat{u}d\hat{u}^*\} = NdF_yN^* = NMdF_uM^*N^* = dF_u$, the spectral distribution measure of u is exactly dF_u. Since each random vector $u(t)$ has the spectral representation $u(t) = \int_{-\pi}^{\pi} e^{i\theta t}N(e^{i\theta})d\hat{y}$, its components belong to $\mathbf{H}(y)$ and obviously the shift of the process y acts also on the process u. This also implies that $\mathbf{H}(u) = \overline{\mathrm{span}}\{u_k(t)\,;\, k = 1,\ldots,p,\, t \in \mathbb{Z}\} \subset \mathbf{H}(y)$. Hence we just need to prove that the converse inclusion also holds. To this end, we shall show that, for any function N satisfying (3.66), the difference $I_m - MN$ (I_m being the $m \times m$ identity matrix function) is equal to zero dF_y-almost everywhere. If this is true, then $d\hat{y} = MNd\hat{y} = Md\hat{u}$, and, by the dual argument to the one used above, we can conclude that $\mathbf{H}(y) \subset \mathbf{H}(u)$. Notice now that from (3.65) and (3.66) we have

$$(I_m - MN)dF_y(I_m - MN)^* = (I_m - MN)MdF_uM^*(I_m - N^*M^*) = 0$$

dF_u-almost everywhere. Hence $MN = I_m$ dF_y-almost everywhere. This concludes the proof. □

In linear algebra left-invertibility of a matrix is associated to a condition of full column rank of the matrix. In order to make contact with this notion we shall need to choose special dominating measures.

Lemma 3.7.6. *Let the spectral distribution measure dF_u of the process u be of the* diagonal type, *i.e., $dF_u = \mathrm{diag}\,\{d\mu_1, d\mu_2, \ldots, d\mu_p\}$ with μ_k, $k = 1, 2, \ldots, p$, positive Borel measures on $[-\pi, \pi]$. Then $\{u_1(0), u_2(0), \ldots, u_p(0)\}$ is a set of generators of $\mathbf{H}(u)$ of smallest cardinality, i.e., a basis for the module $\mathbf{H}(u)$.*

Proof. By assumption, $\mathrm{E}\{d\hat{u}d\hat{u}^*\} = dF_u$ satisfies

$$\mathrm{E}\{d\hat{u}_k d\hat{u}_j^*\} = 0 \qquad \text{for } k \neq j,$$

and hence it follows that, for $k \neq j$, $u_k(t) = \mathcal{U}^t u_k(0)$ and $u_j(s) = \mathcal{U}^s u_j(0)$ are orthogonal for all $t, s \in \mathbb{Z}$. Clearly the module generated by any proper subset of $\{u_k(0),\, k = 1, 2, \ldots, p\}$ has a nonzero orthogonal complement and must then be a proper submodule of $\mathbf{H}(u)$. Hence the random variables $u_1(0), u_2(0), \ldots, u_p(0)$ are a minimal set of generators. □

The lemma obviously holds also for the special case of *scalar type* measures, which are of the form $dF_u = I_p d\mu$.

Recall that the elements of every matrix measure dF_y are absolutely continuous with respect to some scalar Borel measure. There are many such measures, the sum of the elements or the trace of dF_y being simple examples. The special case when one takes μ to be the Lebesgue measure will be examined in the next subsection. In any case, for any such scalar dominating measure, one can show that $dF_y \ll dF_u = I_m d\mu$, in the sense defined earlier and hence we shall have $dF_y = M(e^{i\theta})M(e^{i\theta})^* d\mu$ for some measurable matrix function M; compare (3.65). The matrix function $\Phi(e^{i\theta}) := M(e^{i\theta})M(e^{i\theta})^*$ for which

$$dF_y = \Phi d\mu \tag{3.67}$$

is called the *spectral density matrix* of dF_y with respect to the scalar measure μ. It is a measurable μ-a.e. Hermitian positive semidefinite $m \times m$ matrix function on $[-\pi, \pi]$.

The following is a finite-dimensional version of a fundamental result of spectral theory of linear operators in Hilbert space, known as the *Hellinger-Hahn Theorem*. The proof can be found in [104, Chapter 6].

Theorem 3.7.7. *Let μ be a scalar Borel measure such that $dF_y \ll I_m d\mu$. Then there exists a diagonal matrix measure dM with nonzero diagonal entries $\mu_1, \ldots, \mu_p$ such that $d\mu_k = m_k(e^{i\theta})\, d\mu$, and the following statements hold:*

(i) $\mu_1 \gg \mu_2 \gg \ldots \gg \mu_p$
(ii) *There exists a measurable $m \times p$ matrix function $H(e^{i\theta})$ such that*

$$H(e^{i\theta})^* H(e^{i\theta}) = I_p, \quad \mu\text{-a.e. on } [-\pi, \pi], \tag{3.68}$$

for which $dF_y = H(e^{i\theta}) dM H(e^{i\theta})^*$.

The diagonal matrix measure $dM = \operatorname{diag}\{m_1(e^{i\theta}), \ldots, m_p(e^{i\theta})\} d\mu$ with the properties (i) *and* (ii) *is unique modulo equivalence with respect to scalar measures. In particular, the integer p is uniquely determined by dF_y.*

Clearly, in view of (3.68), the matrix $H(e^{i\theta})$ must be left-invertible μ-a.e. Let H^{-L} be any left inverse and define the stochastic p-dimensional vector measure $d\hat{u} := H^{-L} d\hat{y}$. It follows that $d\hat{u}$ has the diagonal spectral distribution dM, and, by Theorem 3.7.5 and Lemma 3.7.6, the components of the corresponding random vector $u(0) := \int H^{-L}(e^{i\theta}) d\hat{y}$ form a minimal set of generators for $\mathbf{H}(y)$. Hence p is the multiplicity of y.

Definition 3.7.8. A process y has *uniform multiplicity* p if the measures $\mu_1, \ldots, \mu_p$ are all mutually absolutely continuous. Equivalently, the scalar densities $m_k(e^{i\theta})$, $k = 1, \ldots, p$, all have the same support.

The following theorem provides a linear algebra characterization of uniform multiplicity.

Theorem 3.7.9. *Let Φ be the spectral density of dF_y with respect to a scalar dominating measure μ. Then the stationary process y has uniform multiplicity p if and only if*

$$\operatorname{rank} \Phi(e^{i\theta}) = p, \quad \mu\text{-a.e. on } [-\pi, \pi]. \tag{3.69}$$

In particular, the components $y_1(0), y_2(0), \ldots, y_m(0)$ form a basis for the module $\mathbf{H}(y)$ if and only if the spectral density of F_y with respect to any dominating scalar measure μ has constant rank m μ-almost everywhere.

Proof. (only if): Assume the process has uniform multiplicity and let $\Delta(e^{i\theta}) := \operatorname{diag}\{m_1(e^{i\theta}), \ldots, m_p(e^{i\theta})\}$. Then $\Phi(e^{i\theta}) = H(e^{i\theta}) \Delta(e^{i\theta}) H(e^{i\theta})^*$ has μ-a.e.

constant rank equal to the integer p in (3.68). (if): Just note that, by (3.68), the rank of $H(e^{i\theta})$ must be constant and equal to p μ-a.e. Hence $\Delta(e^{i\theta})$ has pointwise the same rank as $\Phi(e^{i\theta})$, which, by assumption, is μ-a.e. constant and equal to p. Therefore the multiplicity is uniform. □

3.7.3 *Processes with an Absolutely Continuous Distribution Matrix*

An important special case in the analysis above occurs when μ is the normalized Lebesgue measure on $[-\pi, \pi]$. Recall that a p-dimensional (orthonormal) white noise process w is one for which $dF_w(\theta) = I_p \frac{d\theta}{2\pi}$.

Definition 3.7.10. We shall say that y is a *(stationarily) orthonormalizable* process if there is a white noise process w, jointly stationary with y, such that $\mathbf{H}(y) = \mathbf{H}(w)$.

Clearly we can express each scalar component of an orthonormalizable process y in terms of the orthonormal basis $\{w(t);\ t \in \mathbb{Z}\}$, thereby obtaining a representation of the form

$$y(t) = \sum_{k=1}^{p} \sum_{s=-\infty}^{+\infty} h_k(t-s) w_k(s), \tag{3.70}$$

where the dependence of h on $t-s$ is a consequence of stationarity. In fact, since the p scalar components of w are orthogonal, y has exactly multiplicity $p = \dim[w(t)]$. Orthonormalizable processes are just the class of second-order processes for which the multiplicity can be computed as the rank, Lebesgue-almost everywhere, of a certain matrix function. Hence we have the following characterization.

Corollary 3.7.11. *A stationary process is orthonormalizable if and only if its spectral distribution dF_y is absolutely continuous with respect to the Lebesgue measure, i.e., $dF_y = \Phi\, d\theta/2\pi$ where the spectral density Φ has constant rank, say p, almost everywhere on $[-\pi, \pi]$. The multiplicity of y is equal to the rank of $\Phi(e^{i\theta})$ a.e. on $[-\pi, \pi]$. The spectral density Φ of an orthonormalizable process of rank p admits $m \times p$* spectral factors *W such that*

$$\Phi(e^{i\theta}) = W(e^{i\theta}) W(e^{i\theta})^*, \tag{3.71}$$

which are (left-invertible) of constant rank p almost everywhere.

Whenever a stationary process admits a spectral density (in particular for orthonormalizable processes) with respect to the Lebesgue measure, the rank a.e. of Φ is commonly referred to as the *rank* of the process. As we have seen, in this case the rank of y is just the same as the multiplicity of the process.

A special case of orthonormalizable processes are processes which are *causally orthonormalizable* in the sense that they are causally equivalent to a white noise; i.e.,

$$\mathbf{H}_t^-(y) = \mathbf{H}_t^-(w), \quad t \in \mathbb{Z}, \tag{3.72}$$

where $\mathbf{H}_t^-(y)$ denotes the Hilbert space spanned by the past random variables of the process at time t, $\{y_k(s)\,;\, s < t, k = 1, \ldots, m\}$. These processes and the associated spectral factorization problem will be studied in much detail in the context of prediction theory and are normally called *purely non-deterministic*, or *linearly regular* in the Russian literature. We anticipate that a fundamental result of Paley and Wiener implies that the spectral density of a purely non-deterministic process y must admit *analytic* spectral factors (in H^2). We shall not enter into this subject now. We just remark that pure non-deterministic property of y has (among other things) to do with the existence of analytic spectral factors and has essentially nothing to do with rank or multiplicity.

In particular, the concept of a "full-rank process" and the notion of a purely non-deterministic (or regular) process, often confused in the literature, have little to do with each other. In fact, the notion of rank and the conditions of Theorem 3.7.9 apply to a much wider class of stationary processes. For example a full-rank process may be purely deterministic and a purely non-deterministic process may well be rank-deficient.

3.8 Bibliographical Notes

Most of the material in this chapter is classical. The spectral representation theorem is due to Cramèr [64–66]; see also the work of his former student K. Karhunen [164, 165] and of Kolmogorov [170]. The proof given here is adapted from [117, p. 203]. A more direct approach in [270] uses the full power of the spectral representation of unitary groups in Hilbert space. In relation to this, it has been remarked by J.L. Doob [77, pp. 635–636] that the stochastic integral, first introduced by Wiener in [304], was defined in exactly the same way as the spectral integrals known in the spectral theory of self-adjoint operators in Hilbert spaces. The spectral representation of stationary processes could then be seen as a chapter of spectral theory in Hilbert spaces. For this reason most of the abstract properties of the pair $(\mathbf{H}(y), \mathcal{U})$ which are used in this book are also valid for any pair $(\mathbf{H}, \mathcal{U})$ where, instead of a stationary process on a probability space, one has just a (separable) Hilbert space $\mathbf{H}$ and a unitary operator $\mathcal{U}$ on $\mathbf{H}$ of finite multiplicity.

The concept of multiplicity can be defined for more general classes of processes than stationary, see e.g. [69, 142]. Multiplicity theory of random processes is an old subject going back to the work of Levy [184], Cramèr [67–70], Hida [142] and others. These references consider continuous-time processes. Here we have restricted the analysis to discrete time for simplicity, but all what we have said can be translated to continuous time without difficulty. The module theory of Sect. 3.7

appears in [253] and is inspired by Fuhrmann's book [104], where it is introduced for self-adjoint operators; see Chapter II, especially pp. 101–102. The rank condition of Theorem 3.7.6 explains in particular why spectral factorization of purely non-deterministic stationary processes must be of "constant rank", a fact which may appear rather mysterious from the way it is normally introduced in the literature.