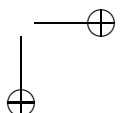
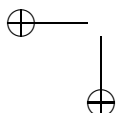
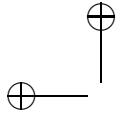
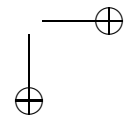
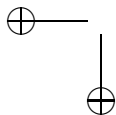
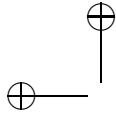


# Linear Stochastic Systems: A Geometric Approach to Modeling, Estimation and Identification

Anders Lindquist and Giorgio Picci

September 28, 2006





# Contents

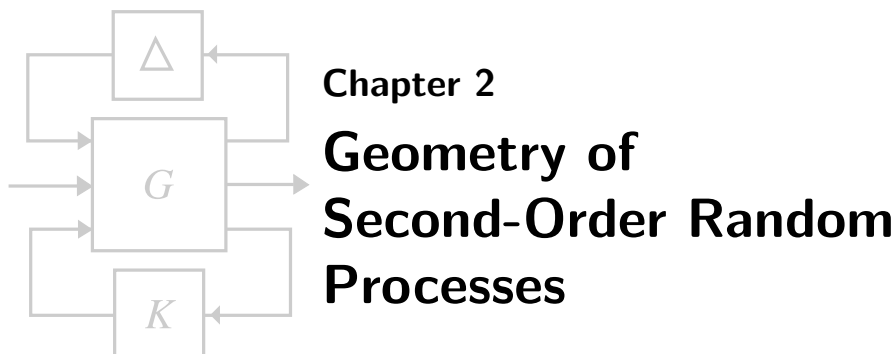
<b>Preface</b>	<b>vii</b>
<b>1 Introduction</b>	<b>1</b>
<b>2 Geometry of Second-Order Random Processes</b>	<b>3</b>
2.1 Hilbert space of second-order random variables . . . . .	3
2.2 Orthogonal projections . . . . .	4
2.3 Angles and singular values . . . . .	10
2.4 Conditional orthogonality . . . . .	12
2.5 Second-order processes and the shift operator . . . . .	15
2.6 Conditional orthogonality and modeling . . . . .	18
2.7 Stationary increments processes in continuous-time . . . . .	30
2.8 Bibliographical notes . . . . .	31
<b>3 Spectral Representation of Stationary Processes</b>	<b>33</b>
3.1 Orthogonal-increments processes and the Wiener integral . . . . .	33
3.2 Harmonic analysis of a stationary process . . . . .	38
3.3 The spectral representation theorem . . . . .	40
3.4 Vector-valued processes . . . . .	46
3.5 Functionals of white noise . . . . .	48
3.6 Spectral representation of stationary increment processes . . . . .	53
3.7 Multiplicity and the module structure of $\mathbf{H}(y)$ . . . . .	56
<b>4 Innovations, Wold Decomposition, and Spectral Factorization</b>	<b>61</b>
4.1 The Wiener-Kolmogorov theory of filtering and prediction . . . . .	61
4.2 Orthonormalizable processes and spectral factorization . . . . .	68
4.3 Hardy spaces . . . . .	72
4.4 Analytic spectral factorization . . . . .	75
4.5 The Wold decomposition . . . . .	77
4.6 The outer spectral factor . . . . .	85
4.7 Bibliographical notes . . . . .	94
<b>5 Wold Decomposition, and Spectral Factorization in Continuous Time</b>	<b>97</b>

5.1	Stationary increments processes and the continuous-time Wold decomposition . . . . .	97
5.2	Hardy spaces of the half-plane . . . . .	98
5.3	Analytic spectral factorization in continuous time . . . . .	102
5.4	Wide sense semimartingales . . . . .	106
5.5	Stationary increments semimartingales in the spectral domain . . . . .	110
5.6	Bibliographical notes . . . . .	115
<b>6</b>	<b>Linear Stochastic Systems</b>	<b>117</b>
6.1	Basic principles of deterministic realization theory . . . . .	117
6.2	Stochastic state-space models . . . . .	124
6.3	Anticausal state-space models . . . . .	127
6.4	The forward and backward generating processes and the structural function . . . . .	130
6.5	The idea of state space and coordinate-free representation . . . . .	133
6.6	Observability, constructibility and minimality . . . . .	135
6.7	The forward and the backward predictor space . . . . .	138
6.8	Rational spectral factorization . . . . .	142
6.9	The Riccati inequality and Kalman filtering . . . . .	149
6.10	Bibliographic notes . . . . .	154
<b>7</b>	<b>The Geometry of Splitting Subspaces</b>	<b>155</b>
7.1	Deterministic realization theory revisited: The abstract idea of state space construction . . . . .	155
7.2	Perpendicular intersection . . . . .	157
7.3	Splitting subspaces . . . . .	160
7.4	Markovian splitting subspaces . . . . .	164
7.5	The Markov semigroup . . . . .	171
7.6	Minimality and dimension . . . . .	172
7.7	Partial ordering of minimal splitting subspaces . . . . .	176
7.8	Bibliographic notes . . . . .	183
<b>8</b>	<b>Markovian Representations</b>	<b>185</b>
8.1	The fundamental representation theorems . . . . .	186
8.2	Normality, properness and the Markov semigroup . . . . .	191
8.3	The forward and backward systems (the finite-dimensional case) . . . . .	195
8.4	Reachability, controllability and the deterministic subspace . . . . .	198
8.5	Minimality and nonminimality of finite-dimensional models . . . . .	207
8.6	Parameterization of finite-dimensional minimal Markovian representations . . . . .	209
8.7	The forward and backward systems (the general case) . . . . .	215
8.8	Regularity of Markovian representations . . . . .	224
8.9	Models without observation noise . . . . .	228
8.10	Bibliographical notes . . . . .	230
<b>9</b>	<b>Proper Markovian Representations in Hardy Space</b>	<b>233</b>

Contents	v
9.1	Functional representations of Markovian representations . . . . . 233
9.2	Minimality of Markovian representations . . . . . 241
9.3	Degeneracy . . . . . 251
9.4	Models without observation noise . . . . . 260
9.5	Bibliographical notes . . . . . 262
<b>10</b>	<b>Stochastic Realization Theory in Continuous Time</b> <span style="float: right;"><b>263</b></span>
10.1	Continuous-time stochastic models . . . . . 263
10.2	Markovian representations . . . . . 268
10.3	Forward and backward realizations: the finite-dimensional case . 282
10.4	Spectral factorization and Kalman filtering . . . . . 291
10.5	Forward and backward stochastic realizations (the general case) 301
10.6	Bibliographical notes . . . . . 314
<b>11</b>	<b>Stochastic Balancing and Model Reduction</b> <span style="float: right;"><b>315</b></span>
11.1	Balancing of stochastic systems . . . . . 315
11.2	Canonical correlation analysis . . . . . 322
11.3	Basic principles of stochastic model reduction . . . . . 331
11.4	Prediction error and relative error criteria . . . . . 334
11.5	Relative error minimization in $H^\infty$ . . . . . 339
11.6	Stochastically balanced truncation . . . . . 342
11.7	Bibliographical notes . . . . . 347
<b>12</b>	<b>Finite-interval Stochastic Realization</b> <span style="float: right;"><b>349</b></span>
12.1	Finite-interval realization . . . . . 350
12.2	Finite-interval realization by canonical correlation analysis . . . 355
12.3	Partial realization of covariance sequences . . . . . 358
12.4	The question of positive rational extension . . . . . 364
12.5	Bibliographical notes . . . . . 366
<b>13</b>	<b>Subspace Identification for Time Series</b> <span style="float: right;"><b>367</b></span>
13.1	The Hilbert Space of a second-order ergodic time series . . . . . 367
13.2	Subspace identification of time series . . . . . 371
13.3	The question of positivity . . . . . 374
13.4	Bibliographical notes . . . . . 377
<b>14</b>	<b>Zero Dynamics and the Geometry of the Riccati Inequality</b> <span style="float: right;"><b>379</b></span>
14.1	The tightest internal bounds . . . . . 379
14.2	Zero dynamics . . . . . 380
14.3	Output-induced subspaces . . . . . 380
14.4	Connections to geometric control theory . . . . . 380
14.5	Zeros and ordering . . . . . 380
14.6	Introducing coordinates . . . . . 380
14.7	The local structure of $\mathcal{P}$ . . . . . 380
14.8	Invariant subspaces and the algebraic Riccati equation . . . . . 380
14.9	Bibliographical notes . . . . . 380

---

<b>15</b>	<b>The Structure of Discrete-Time Models</b>	<b>381</b>
15.1	Invariant directions . . . . .	381
15.2	The maximal output-induced subspace . . . . .	381
15.3	The zero dynamics in the general case . . . . .	381
15.4	Bibliographical notes . . . . .	381
<b>16</b>	<b>Smoothing and Interpolation</b>	<b>383</b>
16.1	Stationary smoothing: Geometric principles of noncausal estimation . . . . .	383
16.2	Minimal stationary smoothing estimators . . . . .	383
16.3	Finite-interval stochastic realization theory (continuous time) . . . . .	383
16.4	Smoothing on a finite interval (continuous time) . . . . .	383
16.5	Finite-interval stochastic realization theory in discrete time, revisited . . . . .	383
16.6	Smoothing in discrete time . . . . .	383
16.7	Interpolating missing state data . . . . .	383
16.8	Bibliographical notes . . . . .	383
<b>17</b>	<b>Stochastic Realization with Inputs</b>	<b>385</b>
17.1	More on feedback-free processes and causality . . . . .	385
17.2	Oblique splitting subspaces . . . . .	385
17.3	Modeling of stationary processes with inputs . . . . .	385
17.4	Issues in identification . . . . .	385
17.5	Bibliographical notes . . . . .	385
<b>A</b>	<b>Appendix</b>	<b>387</b>
A.1	Hilbert spaces . . . . .	387
A.2	Subspace algebra . . . . .	392
A.3	Some facts from linear algebra . . . . .	395
	<b>Bibliography</b>	<b>399</b>
	<b>Index</b>	<b>409</b>



In this book, modeling and estimation problems of random processes are treated in a unified geometric framework. For this, we need some basic facts about the Hilbert space theory of stochastic vector processes that have finite second order moments and are stationary in the wide sense. Such a process  $\{y(t)\}_{t \in \mathbb{Z}}$  is a collection of random variables  $y_k(t), k = 1, 2, \dots, m, t \in \mathbb{Z}$ , which generate a Hilbert space  $\mathbf{H}$  with inner product

$$\langle \xi, \eta \rangle = E\{\xi\eta\},$$

where  $E$  denotes mathematical expectation. This Hilbert space is endowed with a *shift*, i.e., a unitary operator  $U : \mathbf{H} \rightarrow \mathbf{H}$  with the property that

$$y_k(t + 1) = Uy_k(t), \quad k = 1, 2, \dots, t \in \mathbb{Z}.$$

In this chapter we introduce some basic geometric facts for such Hilbert spaces. Although we shall assume that the reader has some knowledge of elementary Hilbert space theory, for the benefit of the reader, some relevant facts are collected in Appendix A.1.

## 2.1 Hilbert space of second-order random variables

A real random variable,  $\xi$ , is just a real-valued measurable function defined on some underlying probability space  $\{\Omega, \mathcal{A}, P\}$  ( $P$  is the probability measure on  $\Omega$  and  $\mathcal{A}$  the  $\sigma$ -algebra<sup>1</sup> of events). The symbol  $E\{\xi\} := \int_{\Omega} \xi dP$  denotes mathematical expectation of the random variable  $\xi$ . Random variables which have finite second moment,  $E\{|\xi|^2\} < \infty$ , are commonly called *second order* random variables.

The set of real or complex-valued second-order random variables  $f$  defined on the same probability space  $\{\Omega, \mathcal{A}, P\}$  is obviously a linear vector space under the usual operations of sum and multiplication by real (or complex ) numbers. This

<sup>1</sup>Knowledge of general measure-theoretic probability theory (and concepts such as  $\sigma$ -algebra) will not be needed for reading this book.

vector space comes naturally equipped with an inner product

$$\langle \xi, \eta \rangle = E \xi \bar{\eta},$$

which is just the correlation of the random variables  $\xi, \eta$ . Note that the norm  $\|\xi\| = \langle \xi, \xi \rangle^{1/2}$  induced by this inner product (the square root of the second moment of  $\xi$ ) is positive, i.e.  $\|\xi\| = 0 \Leftrightarrow \xi = 0$ , only if we agree to identify random variables which are equal almost surely, i.e. differ on a set of probability zero. Consider the set of equivalence classes of second-order random variables  $f$  with respect to almost sure equality. This set, once equipped with the inner product  $\langle \cdot, \cdot \rangle$ , becomes an inner product space, denoted  $L^2(\Omega, \mathcal{A}, P)$ . Convergence with respect to the norm of this space is called *convergence in mean square*. It is a very well-known fact that  $L^2(\Omega, \mathcal{A}, P)$  is actually closed with respect to convergence in mean square and is therefore a Hilbert space.

### Notations and conventions

In this book the term *subspace* of a Hilbert space  $\mathbf{H}$ , will in general mean *closed* subspace. For finite-dimensional vector quantities,  $|v|$  will denote Euclidean length (or absolute value in the scalar case).

The *sum* of two linear vector spaces  $\mathbf{X} + \mathbf{Y}$ , is, by definition, the linear vector space  $\{x + y \mid x \in \mathbf{X}, y \in \mathbf{Y}\}$ . Even when  $\mathbf{X}$  and  $\mathbf{Y}$  are both (closed) subspaces, this linear manifold may fail to be closed. The (*closed*) *vector sum* of  $\mathbf{X}$  and  $\mathbf{Y}$ , denoted  $\mathbf{X} \vee \mathbf{Y}$ , is the closure of  $\mathbf{X} + \mathbf{Y}$ .

In this book, the symbols  $+$ ,  $\vee$ ,  $\dot{+}$  and  $\oplus$  will denote sum, (closed) vector sum, direct sum, and *orthogonal* direct sum of subspaces. The symbol  $\mathbf{X}^\perp$  denotes the orthogonal complement of the subspace  $\mathbf{X}$  with respect to some predefined ambient space. The *linear vector space generated by a family of elements*  $\{x_\alpha\}_{\alpha \in \mathbb{A}} \subset \mathbf{H}$ , denoted  $\text{span}\{x_\alpha \mid \alpha \in \mathbb{A}\}$ , is the vector space whose elements are all finite linear combinations of the *generators*  $\{x_\alpha\}$ . The *subspace generated by the family*  $\{x_\alpha\}_{\alpha \in \mathbb{A}}$  is the closure of this linear vector space and is denoted by  $\overline{\text{span}}\{x_\alpha \mid \alpha \in \mathbb{A}\}$ .

In Appendix A.1 one can find more details and comments on these concepts.

## 2.2 Orthogonal projections

Consider now the following problem: a second order random variable  $x$ , whose values are not accessible to direct observation, can be measured indirectly by some measurement (or observation) device. This device produces a sequence of real valued-observations which we model as sample values of a family of random variables (a stochastic process)  $y = \{y(t) \mid t \in \mathbb{T}\}$  defined in the same probability space of  $x$ . From an observed trajectory of  $y$  we want to get the best possible reconstruction of the value, say  $\bar{x}$ , of  $x$ , which occurred during the measurement. This means that we want to find a function of the observed data (an “estimator”),  $\varphi(y)$ , which best approximates  $x$ , i.e. produces, “in the average”, the smallest possible estimation error  $x - \varphi(y)$  (note that this quantity is itself a random variable). Both  $x$  and the scalar components of  $y$  are assumed with finite second order moments, so we



may regard them as elements of the Hilbert space  $L^2(\Omega, \mathcal{A}, P)$  of the underlying probability space. It is then natural to require that  $\varphi(y)$  should also have finite second-order moment.

Now, the second-order functions of the process  $y$  form a closed subspace of  $L^2(\Omega, \mathcal{A}, P)$  which can be identified with  $L^2(\Omega, \mathcal{Y}, P)$ ,  $\mathcal{Y} \subset \mathcal{A}$  being the  $\sigma$ -algebra generated by the process  $y$ . In other words, any admissible  $\varphi(y)$  is just an element of the subspace  $L^2(\Omega, \mathcal{Y}, P)$  of the Hilbert space  $L^2(\Omega, \mathcal{A}, P)$ .

It is then natural to formulate the problem in the following way: find a random variable  $z$  in  $L^2(\Omega, \mathcal{Y}, P)$  for which the estimation error  $x - z$  has the smallest possible  $L^2$  norm, namely solve the following optimization problem

$$\min_{z \in L^2(\Omega, \mathcal{Y}, P)} \|x - z\| \tag{2.2.1}$$

where  $\|x - z\|^2 = E\{|x - z|^2\}$ .

It is well-known that this minimum distance problem has a unique solution and that this solution is the *orthogonal projection* of  $x$  onto the subspace  $L^2(\Omega, \mathcal{Y}, P)$ . For future reference we shall recall here without proof the following basic result which is sometimes referred to as the *Orthogonal Projection Lemma*.

**Lemma 2.2.1.** *Let  $\mathbf{Y}$  be a closed subspace of a Hilbert space  $\mathbf{H}$ . Given  $x \in \mathbf{H}$ , the element  $z \in \mathbf{Y}$  which has shortest distance from  $x$ , i.e. minimizes  $\|x - z\|$  is unique and is the orthogonal projection of  $x$  onto  $\mathbf{Y}$ .*

*A necessary and sufficient condition for  $z$  to be equal to the orthogonal projection of  $x$  onto  $\mathbf{Y}$  is that  $x - z \perp \mathbf{Y}$ , or, equivalently, for any system of generators  $\{y_\alpha; \alpha \in A\}$  of  $\mathbf{Y}$  it should hold that*

$$\langle x - z, y_\alpha \rangle = 0, \quad \alpha \in A \tag{2.2.2}$$

*(orthogonality principle).*

If we take as generators of  $L^2(\Omega, \mathcal{Y}, P)$  the indicator functions  $\{I_A, A \in \mathcal{Y}\}$  we may rewrite the orthogonality relation (2.2.2) as

$$E\{xI_A\} = E\{zI_A\}, \quad A \in \mathcal{Y}$$

which is the well-known defining relation of the *conditional expectation*

$$z = E[x | \mathcal{Y}] \equiv E[x | y]$$

Hence the best estimator of the random variable  $x$ , based on the observed data  $y = \{y(t) | t \in \mathbb{T}\}$  in the sense of the smallest “mean square error” (the distance in  $L^2$ ), is just the *conditional expectation* of  $x$  given the data.

Unfortunately this insight is not of much use since, except in the notable case when  $x$  and  $y$  have a jointly Gaussian distribution, the conditional expectation is most of the times practically impossible to compute.

In a non-Gaussian setting, or more realistically, when there is not enough information about the probability law of the variables involved, one needs to restrict

“a priori” the class of functions  $\varphi$  of the data which constitute admissible estimators. We shall henceforth restrict ourselves to estimators which are *linear* functions of the data. As we shall see in a moment, minimum mean-square error linear estimators are completely determined by the second order statistics of the variables of the problem. Moreover, since in the Gaussian case the conditional expectation turns out to be a *linear function of the data*, the best linear estimator coincides in this case with the best (non linear) function of the data.

### Linear estimation and orthogonal projections

We shall first consider the “static” finite-dimensional case where the observable  $y$  is a random vector with  $m$  components. Let  $x$  be an  $n$ -dimensional inaccessible random vector of dimension  $n$ . Assume that the joint covariance matrix of  $x$  and  $y$

$$\Sigma = \begin{bmatrix} \Sigma_x & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_y \end{bmatrix} \tag{2.2.3}$$

is given. It will be convenient to subtract off the expected values from all random quantities involved (which will henceforth assumed to have zero-mean). Let

$$\mathbf{H}(y) := \text{span} \{y_k \mid k = 1, \dots, m\}$$

be the (finite dimensional) subspace of  $L^2(\Omega, \mathcal{A}, P)$  linearly generated by the components of  $y$ . The *best linear estimator* of  $x$  based on (or given)  $y$ , is the  $n$ -dimensional random vector  $\hat{x}$ , whose components  $\hat{x}_k \in \mathbf{H}(y)$ ,  $k = 1, \dots, n$ , individually solve the minimum problems

$$\min_{z_k \in \mathbf{H}(y)} \|x_k - z_k\| \quad k = 1, \dots, n, \tag{2.2.4}$$

In view of Lemma 2.2.1,  $\hat{x}_k$  is just the orthogonal projection of  $x_k$  onto  $\mathbf{H}(y)$ . According to our previous conventions, we shall denote this projection by the symbols

$$E[x_k \mid \mathbf{H}(y)] \quad \text{or} \quad E^{\mathbf{H}(y)} x_k$$

More generally, the orthogonal projection onto an arbitrary (closed) subspace  $\mathbf{Y} \subset L^2(\Omega, \mathcal{A}, P)$  will be denoted by  $E[\cdot \mid \mathbf{Y}]$ , or by the shorthand  $E^{\mathbf{Y}}$ . The abuse of notation will be harmless since in this book we shall have no occasion of using the conditional expectation operator other than for Gaussian variables. The notation  $E[x \mid \mathbf{Y}]$  will be used also when  $x$  is vector-valued. The symbol will then just denote the vector with components  $E[x_k \mid \mathbf{Y}]$ ,  $k = 1, \dots, n$ . When the projection is expressed in terms of some specific set of generators say  $y = \{y_\alpha\}$  (i.e.  $\mathbf{Y} = \overline{\text{span}}\{y_\alpha\}$ ), we shall denote it  $E[x \mid y]$ .

**Remark 2.2.2.** Since each  $x_k$  has zero mean, searching for the optimum in the seemingly more general class of *affine* functions of the vector  $y$  of the form  $\varphi(y) = a'y + b$  would be futile, as  $b = 0$  is trivially a necessary condition for minimizing the distance<sup>2</sup>. This is the reason why it is enough to project onto the subspace of

<sup>2</sup>For the same reason the squared norm of the difference  $\|x_k - \hat{x}_k\|^2$  is a variance.

## 2.2. Orthogonal projections

7

linear functionals  $\mathbf{H}(y)$ , which is generated by the observation variables, centered to mean zero.

Clearly the  $n$  scalar optimization problems (2.2.4) can be reformulated as one single equivalent problem where one seeks to minimize

$$\text{var}(x - z) := \sum_{k=1}^n \|x_k - z_k\|^2, \quad z_k \in \mathbf{H}(y)$$

which is the *scalar variance* of the error vector  $x - z$ . The scalar variance is the trace of the matrix

$$\text{Var}(x - z) := \mathbf{E}\{(x - z)(x - z)'\}.$$

**Proposition 2.2.3.** *Let  $x$  and  $y$  be zero-mean second-order random vectors of dimensions  $n$  and  $m$  respectively with covariance matrix (2.2.3). Then the orthogonal projection (minimum variance linear estimator) of  $x$  onto the linear subspace spanned by the components of  $y$  is given by*

$$\mathbf{E}[x | y] = \Sigma_{xy} \Sigma_y^\dagger y \tag{2.2.5}$$

where  $\dagger$  denotes the Moore-Penrose pseudoinverse.<sup>3</sup> The (residual) error vector has covariance matrix,

$$\Lambda := \text{Var}(x - \mathbf{E}[x | y]) = \Sigma_x - \Sigma_{xy} \Sigma_y^\dagger \Sigma_{yx}. \tag{2.2.6}$$

This is the smallest error covariance matrix obtainable in the class of all linear functions of the data, i.e.  $\Lambda \leq \text{Var}(x - Ay)$  for any matrix  $A \in \mathbb{R}^{n \times m}$ , where the inequality is understood in the sense of the positive semidefinite ordering among symmetric matrices.

**Proof.** Writing the vector  $z$  as  $z = Ay$ , and invoking the orthogonality condition (2.2.2) for each component  $x_k$ , we obtain

$$\mathbf{E}\{(x - Ay)y'\} = 0 \quad (n \times m)$$

which is equivalent to  $\Sigma_{xy} - A\Sigma_y = 0$ . If  $\Sigma_y$  is non-singular the pseudoinverse is a true inverse, and (2.2.5) is proven. The case when  $\Sigma_y$  is singular is discussed in the following two Lemmas.

**Lemma 2.2.4.** *Let  $\text{rank } \Sigma_y = r \leq m$  and let the columns of  $U \in \mathbb{R}^{m \times r}$  form a basis for the range space  $\text{Im } \Sigma_y$ . Then the components of the  $r$ -dimensional random vector  $v$  defined by*

$$v = U'y \tag{2.2.7}$$

form a basis for  $\mathbf{H}(y)$ . The Moore-Penrose pseudoinverse of  $\Sigma_y$  can be written as

$$\Sigma_y^\dagger = U(U'\Sigma_y U)^{-1}U' \tag{2.2.8}$$

<sup>3</sup>See Section A.3 in the appendix for the definition of the Moore-Penrose pseudoinverse.

and does not depend on the particular choice of  $U$ .

**Proof.** Let the columns of  $U$  and  $V$  form an orthonormal basis for  $\text{Im } \Sigma_y$  and  $\text{Ker } \Sigma_y$  respectively, and define  $y_U := U' y$  and  $y_V := V' y$ . Note that  $T := \begin{bmatrix} U & V \end{bmatrix}$  is an orthogonal (in particular nonsingular)  $m \times m$  matrix so that the components of  $y_U$  and  $y_V$  together span  $\mathbf{H}(y)$ . Observe however that

$$\Sigma_{y_V} = \text{E}[V' y y' V] = V' \text{E}[y y'] V = V' \Sigma_y V = 0$$

i.e. the variance of  $y_V$  is zero, so that the random vector  $y_V$  is also zero. It follows that

$$T' y = \begin{bmatrix} U' \\ V' \end{bmatrix} y = \begin{bmatrix} y_U \\ 0 \end{bmatrix}. \quad (2.2.9)$$

Since  $T'$  is non singular,  $\mathbf{H}(y) = \mathbf{H}(T' y) = \mathbf{H}(y_U)$ . Moreover  $\Sigma_{y_U} = U' \Sigma_y U$  is non singular. In fact,  $w \in \text{Ker } \Sigma_{y_U}$  i.e.  $\Sigma_y U w = 0$  can hold true only when  $U w = 0$  since the columns of  $U$  are a basis for the orthogonal complement of  $\text{ker } \Sigma_y$ . But the columns of  $U$  are linearly independent so that  $w = 0$ .

Now we check that  $U(U' \Sigma_y U)^{-1} U'$  is the Moore-Penrose pseudoinverse of  $\Sigma_y$ . Using the property of the pseudo-inverse,  $[T^{-1} A T^{-1}]^\dagger = T A^\dagger T'$ , we see that

$$\begin{aligned} \Sigma_y^\dagger &= \left[ T^{-1} \begin{bmatrix} U' \\ V' \end{bmatrix} \Sigma_y \underbrace{\begin{bmatrix} U & V \end{bmatrix}}_T T^{-1} \right]^\dagger = T \left[ \begin{bmatrix} U' \\ V' \end{bmatrix} \Sigma_y \begin{bmatrix} U & V \end{bmatrix} \right]^\dagger T' = \\ &= T \begin{bmatrix} U' \Sigma_y U & 0 \\ 0 & 0 \end{bmatrix}^\dagger T' = \begin{bmatrix} U & V \end{bmatrix} \begin{bmatrix} (U' \Sigma_y U)^{-1} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} U' \\ V' \end{bmatrix} = \\ &= U(U' \Sigma_y U)^{-1} U' \end{aligned}$$

and this concludes the proof of the lemma.  $\square$

Using the generators (2.2.7) we can now reduce the general singular covariance case to the nonsingular one. In this case (2.2.5) yields

$$\text{E}[x | y] = \text{E}[x | v] = \Sigma_{xv} \Sigma_v^{-1} U' y = \Sigma_{xy} U \Sigma_v^{-1} U' y$$

and by (2.2.8) the general formula (2.2.5) follows.

The formula for the error covariance follows easily from the orthogonality condition. For what concerns the minimum matrix variance property of the estimator, we readily see that for arbitrary  $A \in \mathbb{R}^{n \times m}$ , one has

$$\text{Var}(x - Ay) = \text{Var}(x - \text{E}[x | y] + \text{E}[x | y] - Ay) = \Lambda + \text{Var}(\text{E}[x | y] - Ay)$$

as  $\text{E}[x | y] - Ay$  has components in  $\mathbf{H}(y)$  and hence is (componentwise) orthogonal to  $x - \text{E}[x | y]$ . From this the minimum property is obvious.  $\square$

Since in the Gaussian (zero-mean) case the conditional expectation  $\text{E}[x | y]$  is a linear function of  $y$ , for Gaussian vectors the orthogonal projection (2.2.5)

coincides with the true conditional expectation. Note that in this case, among all square integrable functions of the data, the best, for the purpose of mean-square approximation of  $x$ , turns out to be linear.

### Facts about orthogonal projections

Below we shall list some useful technical facts about orthogonal projection operators. Henceforth the symbol  $E^{\mathbf{X}} \mathbf{Y}$  will denote the *closure* of  $\{E^{\mathbf{X}} \eta \mid \eta \in \mathbf{Y}\}$ .

**Lemma 2.2.5.** *Let  $\mathbf{A}$  and  $\mathbf{B}$  be orthogonal subspaces of a Hilbert space  $\mathbf{H}$ . Then*

$$E^{\mathbf{A} \oplus \mathbf{B}} \lambda = E^{\mathbf{A}} \lambda + E^{\mathbf{B}} \lambda, \quad \lambda \in \mathbf{H} \tag{2.2.10}$$

**Lemma 2.2.6.** *Let  $\mathbf{A}$  and  $\mathbf{B}$  be subspaces of a Hilbert space  $\mathbf{H}$ . Then*

$$\mathbf{A} = E^{\mathbf{A}} \mathbf{B} \oplus \mathbf{A} \cap \mathbf{B}^{\perp} \tag{2.2.11}$$

where  $\mathbf{B}^{\perp}$  is the orthogonal complement of  $\mathbf{B}$  in any space containing  $\mathbf{A} \vee \mathbf{B}$ .

**Proof.** Set  $\mathbf{C} := \mathbf{A} \ominus E^{\mathbf{A}} \mathbf{B}$ . We want to show that  $\mathbf{C} = \mathbf{A} \cap \mathbf{B}^{\perp}$ . Let  $\alpha \in \mathbf{A}$  and  $\beta \in \mathbf{B}$ . Then, since  $(\beta - E^{\mathbf{A}} \beta) \perp \mathbf{A}$ ,

$$\langle \alpha, E^{\mathbf{A}} \beta \rangle = \langle \alpha, \beta \rangle$$

so that  $\alpha \perp E^{\mathbf{A}} \mathbf{B}$  if and only if  $\alpha \perp \mathbf{B}$ . Consequently  $\mathbf{C} = \mathbf{A} \cap \mathbf{B}^{\perp}$  as claimed.  $\square$

Let  $\mathbf{A}$  and  $\mathbf{B}$  be subspaces of  $\mathbf{H}$ . Consider the restricted orthogonal projection

$$E^{\mathbf{A}} |_{\mathbf{B}} : \mathbf{B} \rightarrow \mathbf{A}, \tag{2.2.12}$$

mapping a random variable  $\xi \in \mathbf{B}$  into its orthogonal projection on the subspace  $\mathbf{A}$ .

**Lemma 2.2.7.** *Let  $\mathbf{A}$  and  $\mathbf{B}$  be subspaces of  $\mathbf{H}$ . Then the adjoint of  $E^{\mathbf{A}} |_{\mathbf{B}}$  is  $E^{\mathbf{B}} |_{\mathbf{A}}$ ; i.e.,*

$$(E^{\mathbf{A}} |_{\mathbf{B}})^* = E^{\mathbf{B}} |_{\mathbf{A}}. \tag{2.2.13}$$

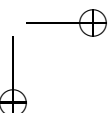
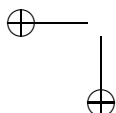
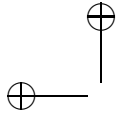
**Proof.** Follows since

$$\langle \alpha, E^{\mathbf{A}} \beta \rangle = \langle \alpha, \beta \rangle = \langle E^{\mathbf{B}} \alpha, \beta \rangle$$

for all  $\alpha \in \mathbf{A}$  and  $\beta \in \mathbf{B}$ .  $\square$

**Lemma 2.2.8.** *Let  $\mathbf{A}$  and  $\mathbf{B}$  be subspaces of  $\mathbf{H}$ . Then*

$$\|E^{\mathbf{A}} |_{\mathbf{B}}\| = \|E^{\mathbf{A}} E^{\mathbf{B}}\|. \tag{2.2.14}$$



**Proof.** In fact, for  $x \in \mathbf{H}$  we have  $\|E^{\mathbf{A}} E^{\mathbf{B}} x\| \leq \|E^{\mathbf{A}}|_{\mathbf{B}}\| \|E^{\mathbf{B}} x\|$  where we actually have equality for  $x \in \mathbf{B}$ . Since for  $x \notin \mathbf{B}$ ,  $\|x\| > \|E^{\mathbf{B}} x\|$ , the supremum of the ratio  $\|E^{\mathbf{A}} E^{\mathbf{B}} x\|/\|x\|$  is attained for  $x \in \mathbf{B}$ . Hence (2.2.14) holds.  $\square$

Let the Hilbert space  $\mathbf{H}$  be equipped with a unitary operator  $U : \mathbf{H} \rightarrow \mathbf{H}$ . An important special case of the restricted projection operator  $E^{\mathbf{A}}|_{\mathbf{B}}$  occurs when  $\mathbf{A}$  is a  $U$ -invariant subspace and  $\mathbf{B}$  is a  $U^*$ -invariant subspace of  $\mathbf{H}$ ; for example, the future and past subspaces at time zero of a stationary process  $y$ ; see (2.5.4). In this case the operator  $E^{\mathbf{A}}|_{\mathbf{B}}$  is called a *Hankel operator*. This kind of operator will play a major role in this book.

**Lemma 2.2.9.** *Let  $U$  be a unitary operator on a Hilbert space  $\mathbf{H}$ . We have*

$$U E^{\mathbf{Y}} \xi = E^{U\mathbf{Y}} U \xi, \quad \xi \in \mathbf{H} \tag{2.2.15}$$

for any subspace  $\mathbf{Y} \subset \mathbf{H}$ .

**Proof.** By the orthogonal projection Lemma,  $h := E[U\xi | U\mathbf{Y}]$  is the unique solution of

$$\langle U\xi - h, Uy \rangle = 0 \quad y \in \mathbf{Y}$$

which is equivalent to saying that  $U^*h$  is the unique solution of  $\langle \xi - U^*h, y \rangle = 0$  for all  $y \in \mathbf{Y}$ . Therefore  $U^*h = E[\xi | \mathbf{Y}]$  and this proves the Lemma.  $\square$

### 2.3 Angles and singular values

Let  $\mathbf{A}$  and  $\mathbf{B}$  be two subspaces of a Hilbert space  $\mathbf{H}$ . Since the number

$$\rho := \sup\{\langle x, y \rangle \mid x \in \mathbf{A}, y \in \mathbf{B}, \|x\| = 1, \|y\| = 1\} = \sup\left\{\frac{|\langle x, y \rangle|}{\|x\| \|y\|} \mid x \in \mathbf{A}, y \in \mathbf{B}\right\}$$

is always between zero (which happens if and only if  $\mathbf{A}$  and  $\mathbf{B}$  are orthogonal) and one (which is true only when  $\mathbf{A}$  and  $\mathbf{B}$  have a nonzero vector in common), there is a unique  $\alpha := \alpha(\mathbf{A}, \mathbf{B})$ ,  $0 \leq \alpha \leq \pi/2$ , such that  $\rho = \cos \alpha$ . The number  $\alpha(\mathbf{A}, \mathbf{B})$ , is called the *angle* between the two subspaces  $\mathbf{A}$  and  $\mathbf{B}$ .

Now, since for arbitrary  $x \in \mathbf{A}$ ,  $y \in \mathbf{B}$  there are  $f, g \in \mathbf{H}$  such that  $x = E^{\mathbf{A}} f$  and  $y = E^{\mathbf{B}} g$ , we have

$$\langle x, y \rangle = \langle E^{\mathbf{A}} f, E^{\mathbf{B}} g \rangle = \langle E^{\mathbf{B}} E^{\mathbf{A}} f, g \rangle$$

and it follows by Proposition A.1.2 and Lemma 2.2.8 that  $\rho = \|E^{\mathbf{B}}|_{\mathbf{A}}\| = \|E^{\mathbf{A}}|_{\mathbf{B}}\|$  (Lemma 2.2.7) and hence

$$\rho = \cos \alpha(\mathbf{A}, \mathbf{B}) = \|E^{\mathbf{A}}|_{\mathbf{B}}\|. \tag{2.3.1}$$

This is just a simple instance of a more general circle of ideas involving the *principal angles* between the subspaces  $\mathbf{A}, \mathbf{B}$  and the *singular value decomposition* of the

operator  $E^A|_{\mathbf{B}}$ . Here we shall need to assume that this operator is *compact* (or *completely continuous* as it is called in the Russian literature), see e.g. [2, 22] for a definition. The compactness of the operator  $E^A|_{\mathbf{B}}$  is a condition which is verified in many cases of interest. In fact, in most cases we shall merely be interested in the case in which  $E^A|_{\mathbf{B}}$  has a finite dimensional range space, whence compactness is trivially guaranteed.

Below we shall report a general theorem dealing with the *singular value decomposition* (SVD) of a compact operator, which is in particular valid also for operators defined on finite dimensional Euclidean spaces (matrices). It is a very important device which will be used in many circumstances.

**Theorem 2.3.1 (Singular value decomposition).** *Let  $T : \mathbf{H}_1 \rightarrow \mathbf{H}_2$  be a compact operator from the Hilbert space  $\mathbf{H}_1$  to the Hilbert space  $\mathbf{H}_2$ . The self-adjoint (compact) operators  $T^*T$  and  $TT^*$  have identical eigenvalues  $\{\sigma_k^2 \geq 0, k = 1, 2, \dots\}$  with the same (finite) geometric multiplicity, except perhaps for the eigenvalue zero. Assume that the eigenvalues are arranged in decreasing order, repeated according to multiplicity. Let  $\{u_k, k = 1, 2, \dots\} \subset \mathbf{H}_2$  be a complete orthonormal system of eigenvectors<sup>4</sup> of  $TT^*$  and let  $\{v_k, k = 1, 2, \dots\} \subset \mathbf{H}_1$  be a complete orthonormal system of eigenvectors of  $T^*T$ .*

*As  $k \rightarrow \infty, \sigma_k^2 \rightarrow 0$  and*

$$Tx = \sum_{k=1}^{+\infty} \sigma_k \langle x, v_k \rangle u_k, \quad x \in \mathbf{H}_1 \tag{2.3.2}$$

*This expression is to be interpreted in the sense that the finite rank approximations*

$$T_n = \sum_{k=1}^n \sigma_k \langle \cdot, v_k \rangle u_k \tag{2.3.3}$$

*converge to  $T$  for  $n \rightarrow \infty$ , both in the strong and in the uniform operator topology. Moreover the approximation error is given by*

$$\|T - T_n\| = \min \left\{ \begin{array}{l} R : \mathbf{H}_1 \rightarrow \mathbf{H}_2, \\ \text{rank } R \leq n \end{array} \right\} \|T - R\| = \sigma_{n+1} \tag{2.3.4}$$

*for  $n = 0, 1, 2, \dots$ , the zero rank approximant being just the operator mapping  $\mathbf{H}_1$  into the zero vector of  $\mathbf{H}_2$  (i.e.  $T_0 \equiv 0$ ).*

The numbers  $\{\sigma_k, k = 1, 2, \dots\}$  are called the *singular values* of the operator  $T$ . Equation (2.3.4) (for  $n = 0$ ) implies that the maximal singular value,  $\sigma_1$ , is just the norm of  $T$ ; i.e.,

$$\sigma_1 = \|T\|. \tag{2.3.5}$$

---

<sup>4</sup>It is well known from functional analysis that any self-adjoint compact operator on a Hilbert space has a complete system of eigenvectors, provided of course we include in the system any orthonormal basis for the nullspace. Recall that all Hilbert spaces that will be considered in this book are *separable*.

Moreover,

$$\overline{\text{span}}\{u_1, u_2, \dots\} = \text{range}\{T\}, \quad \overline{\text{span}}\{v_1, v_2, \dots\} = \ker\{T\}^\perp. \quad (2.3.6)$$

It follows in particular, that all the singular values of the operator  $T = E^{\mathbf{A}}|_{\mathbf{B}}$  are bounded between zero and one,  $1 \geq \sigma_k \geq 0$ , and have (as does  $\rho = \sigma_1$ ) also an interpretation as cosines of angles between subspaces. In this setting, the normalized eigenvectors  $\{u_k, k = 1, 2, \dots\}$  and  $\{v_k, k = 1, 2, \dots\}$  are called the *principal directions* of the subspaces  $\mathbf{A}$  and  $\mathbf{B}$ , and  $\alpha_k := \arccos \sigma_k$  the *k-th principal angle*. In the statistical literature, where  $\mathbf{A}$  and  $\mathbf{B}$  are spaces of random variables, the  $\sigma_k$ 's are also called *canonical correlation coefficients* and  $\{u_k, k = 1, 2, \dots\}$  and  $\{v_k, k = 1, 2, \dots\}$ , the *canonical variables*. As it will be seen in the following, these notions play an important role in various problems of model reduction, approximation and identification of stochastic systems. We shall embark in a detailed study of these concepts in Chapter 11.

Using the approximation property (2.3.4), it can be shown that  $\sigma_{n+1} = \cos \alpha_{n+1}$  is the solution of the following minimization problem

$$\sigma_{n+1} = \langle u_{n+1}, v_{n+1} \rangle = \max_{u \in \mathbf{A}, v \in \mathbf{B}} \{\langle u, v \rangle\} \quad (2.3.7)$$

subject to:

$$\begin{aligned} \langle u, u_k \rangle &= 0 & k = 1, \dots, n \\ \langle v, v_k \rangle &= 0 & k = 1, \dots, n \\ \|u_k\| &= \|v_k\| = 1 \end{aligned} \quad (2.3.8)$$

which is a generalization of the well-known Rayleigh quotient iteration of linear algebra. A proof for the finite-dimensional case can also be found in [40, p.584].

## 2.4 Conditional orthogonality

We say that two subspaces  $\mathbf{A}$  and  $\mathbf{B}$  of a Hilbert space  $\mathbf{H}$  are *conditionally orthogonal* given a third subspace  $\mathbf{X}$  if

$$\langle \alpha - E^{\mathbf{X}} \alpha, \beta - E^{\mathbf{X}} \beta \rangle = 0 \quad \text{for all } \alpha \in \mathbf{A}, \beta \in \mathbf{B} \quad (2.4.1)$$

and we shall denote this  $\mathbf{A} \perp \mathbf{B} | \mathbf{X}$ . When  $\mathbf{X} = 0$ , this reduces to the usual orthogonality  $\mathbf{A} \perp \mathbf{B}$ . Conditional orthogonality is orthogonality after subtracting the projections on  $\mathbf{X}$ . Using the definition of the projection operator  $E^{\mathbf{X}}$ , it is straightforward to see that (2.4.1) may also be written

$$\langle E^{\mathbf{X}} \alpha, E^{\mathbf{X}} \beta \rangle = \langle \alpha, \beta \rangle \quad \text{for all } \alpha \in \mathbf{A}, \beta \in \mathbf{B}. \quad (2.4.2)$$

The following lemma is a trivial consequence of the definition.

**Lemma 2.4.1.** *If  $\mathbf{A} \perp \mathbf{B} | \mathbf{X}$ , then  $\mathbf{A}_0 \perp \mathbf{B}_0 | \mathbf{X}$  for all  $\mathbf{A}_0 \subset \mathbf{A}$  and  $\mathbf{B}_0 \subset \mathbf{B}$ .*

Let  $\mathbf{A} \oplus \mathbf{B}$  denote the *orthogonal* direct sum of  $\mathbf{A}$  and  $\mathbf{B}$ . If  $\mathbf{C} = \mathbf{A} \oplus \mathbf{B}$ , then  $\mathbf{B} = \mathbf{C} \ominus \mathbf{A}$  is the orthogonal complement of  $\mathbf{A}$  in  $\mathbf{C}$ . There are several useful alternative characterizations of conditional orthogonality.



**Proposition 2.4.2.** *The following statements are equivalent.*

- (i)  $\mathbf{A} \perp \mathbf{B} \mid \mathbf{X}$
- (ii)  $\mathbf{B} \perp \mathbf{A} \mid \mathbf{X}$
- (iii)  $(\mathbf{A} \vee \mathbf{X}) \perp \mathbf{B} \mid \mathbf{X}$
- (iv)  $E^{\mathbf{A} \vee \mathbf{X}} \beta = E^{\mathbf{X}} \beta$  for all  $\beta \in \mathbf{B}$
- (v)  $(\mathbf{A} \vee \mathbf{X}) \ominus \mathbf{X} \perp \mathbf{B}$
- (vi)  $E^{\mathbf{A}} \beta = E^{\mathbf{A}} E^{\mathbf{X}} \beta$  for all  $\beta \in \mathbf{B}$

**Proof.** The equivalence of (i), (ii) and (iii) follows directly from the definition. Since  $(\beta - E^{\mathbf{X}} \beta) \perp \mathbf{X}$ , the conditional orthogonality (2.4.1) may be written

$$\langle \alpha, \beta - E^{\mathbf{X}} \beta \rangle = 0 \quad \text{for all } \alpha \in \mathbf{A}, \beta \in \mathbf{B}. \quad (2.4.3)$$

Hence (iii) is equivalent to  $(\beta - E^{\mathbf{X}} \beta) \perp \mathbf{A} \vee \mathbf{X}$ , i.e.

$$E^{\mathbf{A} \vee \mathbf{X}} (\beta - E^{\mathbf{X}} \beta) = 0 \quad \text{for all } \beta \in \mathbf{B},$$

and hence to (iv). Moreover, (2.4.3) is equivalent to

$$E^{\mathbf{A}} (\beta - E^{\mathbf{X}} \beta) = 0 \quad \text{for all } \beta \in \mathbf{B},$$

i.e. to (vi). Finally, set  $\mathbf{Z} := (\mathbf{A} \vee \mathbf{X}) \ominus \mathbf{X}$ . Then  $\mathbf{A} \vee \mathbf{X} = \mathbf{X} \oplus \mathbf{Z}$ , i.e.

$$E^{\mathbf{A} \vee \mathbf{X}} \beta = E^{\mathbf{X}} \beta + E^{\mathbf{Z}} \beta \quad \text{for all } \beta \in \mathbf{B}.$$

Hence (iv) is equivalent to  $E^{\mathbf{Z}} \beta = 0$  for all  $\beta \in \mathbf{B}$ , i.e.  $\mathbf{Z} \perp \mathbf{B}$ , which is the same as (v).  $\square$

Next we give an important example of conditional orthogonality.

**Proposition 2.4.3.** *For any subspaces  $\mathbf{A}$  and  $\mathbf{B}$ ,*

$$\mathbf{A} \perp \mathbf{B} \mid E^{\mathbf{A}} \mathbf{B}. \quad (2.4.4)$$

*Moreover, any  $\mathbf{X} \subset \mathbf{A}$  such that  $\mathbf{A} \perp \mathbf{B} \mid \mathbf{X}$  contains  $E^{\mathbf{A}} \mathbf{B}$ .*

**Proof.** If  $\mathbf{X} \subset \mathbf{A}$ , by Proposition 2.4.2 (v),  $\mathbf{A} \perp \mathbf{B} \mid \mathbf{X}$  is equivalent to  $\mathbf{A} \ominus \mathbf{X} \perp \mathbf{B}$  or, which is the same, to

$$\mathbf{A} \ominus \mathbf{X} \subset \mathbf{A} \cap \mathbf{B}^{\perp}.$$

But, in view of Lemma 2.2.6,

$$\mathbf{A} \ominus E^{\mathbf{A}} \mathbf{B} = \mathbf{A} \cap \mathbf{B}^{\perp}.$$

Consequently (2.4.4) holds, and  $E^{\mathbf{A}} \mathbf{B} \subset \mathbf{X}$  for all  $\mathbf{X}$  such that  $\mathbf{A} \perp \mathbf{B} | \mathbf{X}$ .  $\square$

The distributive law

$$\mathbf{X} \cap (\mathbf{A} + \mathbf{B}) = (\mathbf{X} \cap \mathbf{A}) + (\mathbf{X} \cap \mathbf{B}) \quad (2.4.5)$$

is of course not valid for arbitrary subspaces  $\mathbf{X}$ ,  $\mathbf{A}$ , and  $\mathbf{B}$  but holds only in very special situations. See for example Proposition A.2.1 in Appendix A.2. However, (2.4.5) is always true when  $\mathbf{A} \perp \mathbf{B} | \mathbf{X}$ . In fact, this is a corollary of a more general result.

**Proposition 2.4.4.** *Let  $\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_n$  and  $\mathbf{X}$  be subspaces such that*

$$\mathbf{A}_i \perp \mathbf{A}_j | \mathbf{X} \quad \text{for all } (i, j) \text{ such that } i \neq j. \quad (2.4.6)$$

*Then*

$$\mathbf{X} \cap (\mathbf{A}_1 + \mathbf{A}_2 + \dots + \mathbf{A}_n) = (\mathbf{X} \cap \mathbf{A}_1) + (\mathbf{X} \cap \mathbf{A}_2) + \dots + (\mathbf{X} \cap \mathbf{A}_n). \quad (2.4.7)$$

**Proof.** First note that

$$\mathbf{X} \cap (\mathbf{A}_1 + \mathbf{A}_2 + \dots + \mathbf{A}_n) \supset (\mathbf{X} \cap \mathbf{A}_1) + (\mathbf{X} \cap \mathbf{A}_2) + \dots + (\mathbf{X} \cap \mathbf{A}_n)$$

holds trivially. To prove the reversed inclusion, note that any

$$\xi \in \mathbf{X} \cap (\mathbf{A}_1 + \mathbf{A}_2 + \dots + \mathbf{A}_n)$$

can be written as

$$\xi = \xi_1 + \xi_2 + \dots + \xi_n,$$

where  $\xi_k \in \mathbf{A}_k$ ,  $k = 1, 2, \dots, n$ . We need to show that  $\xi_k \in \mathbf{X}$ ,  $k = 1, 2, \dots, n$ . To this end, note that

$$\sum_{k=1}^n (\xi_k - E^{\mathbf{X}} \xi_k) = \xi - E^{\mathbf{X}} \xi = \xi - \xi = 0.$$

But, in view of (2.4.6),

$$(\xi_1 - E^{\mathbf{X}} \xi_1) \perp (\xi_2 - E^{\mathbf{X}} \xi_2) \perp \dots \perp (\xi_n - E^{\mathbf{X}} \xi_n),$$

and hence  $\xi_k - E^{\mathbf{X}} \xi_k = 0$  for  $k = 1, 2, \dots, n$ , implying that  $\xi_k \in \mathbf{X}$  for  $k = 1, 2, \dots, n$ .  $\square$

Generalizing the definition given at the beginning of this section, shall say that  $\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_n$  are *conditionally orthogonal given  $\mathbf{X}$*  if (2.4.6) holds, and we shall denote this

$$\mathbf{A}_1 \perp \mathbf{A}_2 \perp \dots \perp \mathbf{A}_n | \mathbf{X}.$$

## 2.5 Second-order processes and the shift operator

A *stochastic process*  $y$  is an ordered collection of random variables  $y := \{y(t)\}$ , all defined in the same probability space. The *time* variable  $t$  will in general be discrete ( $t \in \mathbb{Z}$ ), but occasionally we shall discuss also continuous-time processes ( $t \in \mathbb{R}$ ). A generic symbol used for the time set will be  $\mathbb{T}$ . Typically we shall deal with discrete-time processes in this book, modifications to continuous time usually being straightforward. Whenever the modifications needed to carry the results over to continuous time case are nontrivial, we shall discuss this case separately.

The random variables  $y(t)$  could take values in either  $\mathbb{R}$  or  $\mathbb{C}$ , in which case we shall talk about a *scalar* (real or complex) process, or, more generally,  $y(t)$  could be vector-valued, taking values in  $\mathbb{R}^m$  or  $\mathbb{C}^m$ . In this case we shall write  $y(t)$  as a column vector. In this book we shall normally consider real vector valued, say  $\mathbb{R}^m$ -valued, processes, just referring to them as “ $m$ -dimensional processes”.

The objective of this book is to study dynamical descriptions of stochastic processes by means of *linear models* and to understand the statistical problems of process estimation and identification in this framework. For this, it will normally suffice to assume that the only available statistical description of the process consists of the *mean*  $m(t) := \mathbb{E} y(t)$ ,  $t \in \mathbb{T}$  and the *covariance* function

$$\Lambda(t, s) := \mathbb{E}\{[y(t) - m(t)][y(s) - m(s)]^*\}, \quad t, s \in \mathbb{T}. \quad (2.5.1)$$

where  $*$  denotes conjugate transpose. A covariance function is a *function of positive type* in the sense that

$$\sum_{k,j=1}^N a_k^* \Lambda(t_k, t_j) a_j \geq 0 \quad (2.5.2)$$

for arbitrary vector coefficients  $a_k \in \mathbb{C}^m$  and all finite choices of  $t_1, \dots, t_N$ .

On a given space  $\{\Omega, \mathcal{A}\}$  one can define an equivalence class of random processes having prescribed first and second-order moments. This class is normally called a *second-order process*. A second order process contains in particular a Gaussian process, whose probability law is uniquely determined by the given moments. As the mean  $m(t)$  is known for all  $t$ , it can be subtracted from the  $y(t)$ , so without loss of generality, second order processes may be assumed to have *zero mean*. In what follows we shall adhere to this convention. Also the attribute “second-order” will be omitted hereafter.

Consider the vector space linearly generated by the scalar components of an  $m$ -dimensional processes  $y := \{y(t); t \in \mathbb{T}\}$ , i.e., the vector space of all real random variables which are finite linear combinations (with real coefficients) of these scalar components. This space is a vector space contained in  $L^2(\Omega, \mathcal{A}, P)$ , which we shall denote  $\text{span}\{y(t); t \in \mathbb{T}\}$ . At first sight, this notation may be a bit misleading; note that the intended meaning is

$$\text{span}\{y(t); t \in \mathbb{T}\} := \left\{ \sum a'_t y(t) \mid t \in \mathbb{T}, a_t \in \mathbb{R}^m \right\}. \quad (2.5.3)$$

where, in the sum, all but a finite number of vector coefficients  $a_t$  are zero. Closing this vector space in  $L^2(\Omega, \mathcal{A}, P)$ , i.e., adding all limits in mean square of fundamental

sequences, we obtain a Hilbert space denoted

$$\mathbf{H}(y) = \overline{\text{span}}\{y(t) \mid t \in \mathbb{T}\}.$$

This space contains all scalar random variables which depend linearly on the random variables of  $y$ . The space  $\mathbf{H}(y)$  is called the *Hilbert space (linearly) generated by the process  $y$* .

In discrete time, the Hilbert space  $\mathbf{H}(y)$  is separable, as it admits a countable dense set by construction. In continuous time,  $\mathbf{H}(y)$  is separable if the process  $y$  is, say, continuous in mean square, in which case  $\{y_k(r) \mid k = 1, \dots, m, r \text{ rational}\}$  is a countable dense set of random variables..

The *past space* and the *future space* at time  $t$ ,  $\mathbf{H}_t^-(y)$  and  $\mathbf{H}_t^+(y)$ , respectively, are the subspaces of  $\mathbf{H}(y)$  constructed from the past and the future histories of the process respectively, i.e.,

$$\mathbf{H}_t^-(y) := \overline{\text{span}}\{y(s) \mid s < t\} \quad \mathbf{H}_t^+(y) := \overline{\text{span}}\{y(s) \mid s \geq t\} \quad (2.5.4)$$

According to a widely accepted convention, in discrete time the present is included in the future only and not in the past. Other choices are of course possible, but the condition

$$\mathbf{H}_t^-(y) \vee \mathbf{H}_t^+(y) = \mathbf{H}(y), \quad \text{for all } t$$

must be respected.

For a continuous-time process that is continuous in mean square, it makes no difference whether or not the present random variable  $y(t)$  is included in the definitions of past and future, since  $y(t)$  is a limit of past (or future) values ( $\lim_{s \rightarrow t} y(s) = y(t)$ ). Indeed,  $\{y(s) \mid s < t\}$  and  $\{y(s) \mid s \leq t\}$  generate the same subspace, as do  $\{y(s) \mid s > t\}$  and  $\{y(s) \mid s \geq t\}$ .

Although  $\mathbf{H}_t^-(y)$  is monotonically increasing in  $t$ , and  $\mathbf{H}_t^+(y)$  is monotonically decreasing, the past and future subspaces of an arbitrary process can vary abruptly with time. In fact, there are scalar mean-square continuous processes  $y$  for which  $\mathbf{H}_t^-(y) = 0$  and for which the *germ space*

$$\lim_{s \downarrow t} \mathbf{H}_s^-(y) := \bigcap_{\epsilon > 0} \mathbf{H}_{t+\epsilon}^-(y).$$

is infinite-dimensional. (See [?] for an example.) This cannot happen for *stationary* processes, which we shall consider next.

### Stationarity

The process  $y$  is said to be *wide-sense stationary* if the covariance matrix (2.5.1) is a function of the *difference*,  $t - s$ , of the arguments. With a slight abuse of notations, we shall write

$$\Lambda(t, s) = \Lambda(t - s).$$

In the following we shall write simply “stationarity”, omitting the attribute “wide sense”. If  $y$  is a stationary discrete-time process, one can define a linear isometric operator  $U : \mathbf{H}(y) \rightarrow \mathbf{H}(y)$ , called the *shift* of the process  $y$ , such that

$$U y_k(t) = y_k(t + 1), \quad t \in \mathbb{Z}, k = 1, \dots, m.$$

More precisely, for both discrete-time and continuous-time processes, the *forward shift* of  $t$  units of time,  $U_t$ ,  $t \in \mathbb{T}$ , is first defined on the subset  $Y := \{y_k(t); k = 1, 2, \dots, m, t \in \mathbb{T}\}$  by setting

$$U_t y_k(s) := y_k(s + t), \quad k = 1, \dots, m. \tag{2.5.5}$$

Then, by stationarity, we have

$$\langle U_t \xi, U_t \eta \rangle = \langle \xi, \eta \rangle \quad \xi, \eta \in Y, \tag{2.5.6}$$

i.e.,  $U_t$  is isometric. Hence,  $U_t$  can be extended to the closure  $\mathbf{H}(y)$  by continuity (see Theorem A.1.5 of Appendix A.1). In fact, it is clear that the extended operator  $U_t$  maps  $\mathbf{H}(y)$  onto itself. Hence  $U_t$  is a unitary operator on  $\mathbf{H}(y)$  and the adjoint,  $U_t^*$ , satisfies the relations

$$U_t^* U_t = U_t U_t^* = I.$$

That is,  $U_t$  is invertible, and  $U_t^* = U_t^{-1}$ . In particular,

$$U_t^* y_k(s) = y_k(s - t), \quad t, s \in \mathbb{T}, k = 1, \dots, m.$$

i.e.,  $U_t^*$  is the *backward shift*.

In conclusion, the family  $\{U_t \mid t \in \mathbb{T}\}$  is a *group of unitary operators on the Hilbert space  $\mathbf{H}(y)$* . In discrete time this just means that  $U_t = U^t$  for all  $t \in \mathbb{Z}$ ,  $U$  being the shift of one step ahead in time. In continuous-time, to avoid pathologies it is normally assumed that the stationary process  $\{y(t); t \in \mathbb{R}\}$  is mean-square continuous. Then, it is easy to show that

$$\lim_{t \rightarrow s} U_t \xi = U_s \xi, \quad \text{for all } \xi \in \mathbf{H}(y).$$

This property is just *strong continuity* in  $\mathbf{H}(y)$ . The shift of a continuous-time mean-square continuous process, is therefore a *strongly continuous* group of unitary operators on the Hilbert space  $\mathbf{H}(y)$ .

Estimation and modeling of stationary processes on infinite or semi-infinite time intervals, naturally involve various linear operations on the random variables of the process which are *time-invariant*, i.e. independent of the particular instant of time chosen as the “present”. In this context, one may fix the present instant of time to an arbitrary value, say  $t = 0$ . Whenever needed, one can shift to any other time instant by merely applying the unitary operator  $U_t$  to the data.

In particular, the future and past subspaces of the process will normally be considered at time  $t = 0$  and denoted simply  $\mathbf{H}^+(y)$  and  $\mathbf{H}^-(y)$ . Then, for an arbitrary present instant  $t$ , we have

$$\mathbf{H}_t^+(y) = U_t \mathbf{H}^+(y), \quad \mathbf{H}_t^-(y) = U_t \mathbf{H}^-(y).$$

As we have noted,  $\mathbf{H}_t^+(y)$  is decreasing with  $t$  while  $\mathbf{H}_t^-(y)$  is increasing in time. This, together with stationarity, is an important property of the future and past subspaces that can be expressed as an *invariance* relation for the shift and its adjoint, namely

$$U \mathbf{H}^+(y) \subset \mathbf{H}^+(y), \quad U^* \mathbf{H}^-(y) \subset \mathbf{H}^-(y). \tag{2.5.7}$$

Shift-invariant subspaces have been studied intensively in operator theory and have nice analytic characterizations. We shall return to this in more detail in Chapter 4. For the moment, we just mention some elementary general facts.

Let  $U$  be a unitary operator on a Hilbert space  $\mathbf{H}$ . A subspace  $\mathbf{X}$  which is invariant for both  $U$  and  $U^*$  is said to be *doubly invariant* for  $U$ . Trivial examples of doubly invariant subspaces are the zero space and the whole space  $\mathbf{H}(y)$ . An invariant subspace  $\mathbf{X}$  is called *reducing* if there is a complementary subspace  $\mathbf{Y}$ , i.e.,  $\mathbf{H} = \mathbf{X} + \mathbf{Y}$ , which is also invariant. The following result follows from Lemma A.1.6 of Appendix A.1.

**Lemma 2.5.1.** *Let  $U$  be a unitary operator on a Hilbert space  $\mathbf{H}$ . Then  $\mathbf{X}$  is doubly invariant if and only if so is  $\mathbf{X}^\perp$  and the orthogonal sum  $\mathbf{X} \oplus \mathbf{X}^\perp$  is reducing for both  $U$  and  $U^*$ .*

As a simple example, the subspaces  $\mathbf{H}(y_k)$ ,  $k = 1, \dots, m$  generated by the scalar components  $y_k$  of a stationary process, are doubly invariant for the shift of the process.

## 2.6 Conditional orthogonality and modeling

Conditional orthogonality is a concept related to that of *sufficient statistic* and has to do with modeling and data reduction. Below we shall discuss a few examples.

### The Markov property

The Markov property is just the mathematical formalization of the idea of *state*, in a stochastic setting. It will play a crucial role later, in the study of stochastic systems.

Suppose we have a time-indexed family of subspaces  $\{\mathbf{X}_t; t \in \mathbb{T}\}$  of a common ambient Hilbert space  $\mathbf{H}$ , and define the *past* and the *future* of the family at time  $t$  as

$$\mathbf{X}_t^- := \overline{\text{span}}\{\mathbf{X}_s; s \leq t\}, \quad \mathbf{X}_t^+ := \overline{\text{span}}\{\mathbf{X}_s; s \geq t\} \quad (2.6.1)$$

We shall say that the family  $\{\mathbf{X}_t; t \in \mathbb{T}\}$  is *Markovian*<sup>5</sup> if, for every  $t \in \mathbb{T}$ , the future and the past are conditionally orthogonal given the present, i.e.

$$\mathbf{X}_t^- \perp \mathbf{X}_t^+ \mid \mathbf{X}_t \quad (2.6.2)$$

which can be written also in the two equivalent ways

$$E^{\mathbf{X}_t^-} \lambda = E^{\mathbf{X}_t} \lambda \quad \text{for all } \lambda \in \mathbf{X}_t^+, \quad (2.6.3a)$$

$$E^{\mathbf{X}_t^+} \mu = E^{\mathbf{X}_t} \mu \quad \text{for all } \mu \in \mathbf{X}_t^-. \quad (2.6.3b)$$

<sup>5</sup>Here, to conform with the standard terminology in the literature, we should add the attribute “wide sense”, but since we shall never talk about “strict sense” properties in this book, we shall refrain to do so.

Note that in this notion past and future play a completely symmetric role.

A *stationary* Markovian family propagates in time by a unitary shift,  $\{U_t\}$ , i.e.

$$\mathbf{X}_{t+s} = U_s \mathbf{X}_t, \quad t, s \in \mathbb{T}$$

In this case we may simplify notations denoting  $\mathbf{X}_0$  as  $\mathbf{X}$  and denoting past and future of the family at time zero by  $\mathbf{X}^-$ ,  $\mathbf{X}^+$  respectively. We shall also let

$$\mathbf{H} := \vee_t \mathbf{X}_t$$

play the role of ambient Hilbert space.

As for deterministic dynamical models, one expects that the state property should lead to some kind of “local description” of the time evolution of the family (like for example a differential equation for deterministic systems in continuous time). In the following we shall study the local description of Markov processes in discrete time, and hence  $U_t = U^t$ , where  $U$  is the unit shift.

It follows from property (v) of Proposition 2.4.2 that for a Markovian family

$$\mathbf{X}^- = \mathbf{X} \oplus (\mathbf{X}^+)^{\perp} \quad \mathbf{X}^+ = \mathbf{X} \oplus (\mathbf{X}^-)^{\perp} \quad (2.6.4)$$

so that we have the orthogonal decomposition

$$\mathbf{H} = (\mathbf{X}^+)^{\perp} \oplus \mathbf{X} \oplus (\mathbf{X}^-)^{\perp} \quad (2.6.5)$$

Now,  $\mathbf{X}^-$  is invariant for  $\{U_t^* \mid t \geq 0\}$ , a unitary semigroup which we shall call the *left (or backward) shift*, and, dually,  $\mathbf{X}^+$  is invariant for the *right (or forward) shift* semigroup  $\{U_t \mid t \geq 0\}$ . It follows from Lemma A.1.6 that  $(\mathbf{X}^+)^{\perp}$  is an  $U_t^*$ -invariant subspace of  $\mathbf{X}^-$  and hence its orthogonal complement,  $\mathbf{X}$  in  $\mathbf{X}^-$  is invariant for the adjoint,  $\{T_t \mid t \geq 0\}$ , of  $\{U_t^* \mid t \geq 0\}$  on  $\mathbf{X}^-$ . This adjoint is a semigroup which is no longer unitary, called the *compressed right shift*, which can be easily seen to admit the representation

$$T_t = E^{\mathbf{X}^-} U_t : \mathbf{X}^- \rightarrow \mathbf{X}^-, \quad t \geq 0$$

We can now give the following characterization of the Markov property in terms of semigroups and invariant subspaces.

**Proposition 2.6.1.** *The family of subspaces  $\{\mathbf{X}_t, t \in \mathbb{Z}\}$  generated by a unitary group*

$$\mathbf{X}_t = U_t \mathbf{X}, \quad t \in \mathbb{Z}$$

*is Markovian if and only if  $\mathbf{X}$  is an invariant subspace for the compressed right shift, namely*

$$E^{\mathbf{X}^-} U|_{\mathbf{X}} = E^{\mathbf{X}} U|_{\mathbf{X}} \quad (2.6.6)$$

*Dually,  $\{\mathbf{X}_t\}$  is Markovian if and only if  $\mathbf{X}$  is an invariant subspace for the left shift compressed to  $\mathbf{X}^+$ , namely*

$$E^{\mathbf{X}^+} U^*|_{\mathbf{X}} = E^{\mathbf{X}} U^*|_{\mathbf{X}} \quad (2.6.7)$$

The characterizations (2.6.6), (2.6.7) are rather obvious from a statistical point of view and are an equivalent formulation of properties (2.6.3a) (2.6.3b). Stated in the semigroup language, they will be useful later on to obtain explicit functional representations of a stationary Markov process.

The map  $U(\mathbf{X}) := E^{\mathbf{X}} U|_{\mathbf{X}}$  is called the *generator* of the Markovian family. Using the relation (2.2.15), it is easy to check that

$$U_t(\mathbf{X}) := E^{\mathbf{X}^-} U^t|_{\mathbf{X}} = E^{\mathbf{X}^-} U^{t-1} U(\mathbf{X}) = \dots = U(\mathbf{X})^t, \quad t \geq 0 \quad (2.6.8)$$

so that  $\{U_t(\mathbf{X}) \mid t \geq 0\}$  is a semigroup with generator  $U(\mathbf{X})$ . An analogous property holds for the adjoint.

There is a difference equation representation of a stationary Markovian family. This evolution equation will be further discussed and generalized in Chapter 8, so we shall not go into too much detail here.

Define the subspaces

$$\mathbf{V}_t = U\mathbf{X}_t^- \ominus \mathbf{X}_t^-$$

representing the “new information” carried by  $\mathbf{X}_{t+1}$  which was not already contained in  $\mathbf{X}_t^-$ . The subspaces  $\{\mathbf{V}_t\}$  are stationary and by construction

$$\mathbf{V}_s \perp \mathbf{V}_t \quad s \neq t. \quad (2.6.9)$$

**Theorem 2.6.2.** *For any random variable  $\xi \in \mathbf{X}$ , the stationary translate  $\xi(t) = U_t \xi$  evolves in time according to a linear equation of the type*

$$\xi(t+1) = U(\mathbf{X}_t)\xi(t) + v_\xi(t), \quad t \in \mathbb{Z} \quad (2.6.10)$$

where  $\{v_\xi(t) \in \mathbf{V}_t, t \in \mathbb{Z}\}$  is a stationary sequence of orthogonal random variables (white noise).

**Proof.** By Lemma 2.2.5,

$$\xi(t+1) = E^{\mathbf{X}^t} U\xi(t) + E^{\mathbf{V}^t} \xi(t+1) = U(\mathbf{X}_t)\xi(t) + v_\xi(t).$$

By (2.6.9),  $\{v_\xi(t)\}$  is a stationary sequence of orthogonal random variables.  $\square$

This geometric theory accommodates the study of infinite-dimensional Markov processes. In fact, given a Markov process  $\{x(t); t \in \mathbb{T}\}$  that takes values in a separable Hilbert space  $\mathcal{X}$ , the subspaces

$$\mathbf{X}_t := \overline{\text{span}}\{(a, x(t))_{\mathcal{X}} \mid a \in \mathcal{X}\}, \quad t \in \mathbb{T} \quad (2.6.11)$$

is a Markovian family. This leads naturally to the next example.

### Stochastic dynamical systems

A fundamental concept in this book is the notion of *stochastic system*.



**Definition 2.6.3.** A stochastic system on  $\mathbf{H}$  is a pair  $(x, y)$  of centered stochastic processes  $\{x(t); t \in \mathbb{T}\}$  and  $\{y(t); t \in \mathbb{T}\}$ , taking values in a real separable Hilbert space  $\mathcal{X}$  and  $\mathbb{R}^m$ , respectively, such that  $\mathbf{X}_t$ ,  $t \in \mathbb{T}$ , defined by (2.6.11), and  $\mathbf{H}(y)$  are contained in  $\mathbf{H}$  and

$$(\mathbf{H}_t^-(y) \vee \mathbf{X}_t^-) \perp (\mathbf{H}_t^+(y) \vee \mathbf{X}_t^+) \mid \mathbf{X}_t, \quad t \in \mathbb{T}, \quad (2.6.12)$$

where  $\mathbf{X}_t^-$  and  $\mathbf{X}_t^+$  are defined by (2.6.1). The processes  $x$  and  $y$  are called the state process and output process, respectively, and  $\mathbf{X}_t$  is the state space at time  $t$ . The stochastic system is finite-dimensional if  $\dim \mathcal{X} < \infty$ .

In particular, (2.6.12) implies that, for each  $t \in \mathbb{T}$ ,

$$\mathbf{H}_t^-(y) \perp \mathbf{H}_t^+(y) \mid \mathbf{X}_t. \quad (2.6.13)$$

That is,  $\mathbf{X}_t$  is a *splitting subspace* with respect to the past space  $\mathbf{H}_t^-(y)$  and the future space  $\mathbf{H}_t^+(y)$ . Moreover,

$$\mathbf{X}_t^- \perp \mathbf{X}_t^+ \mid \mathbf{X}_t. \quad (2.6.14)$$

In other words,  $\{\mathbf{X}_t; t \in \mathbb{T}\}$  is a Markovian family and, equivalently,  $\{x(t); t \in \mathbb{T}\}$  is a *Markov process*.

We shall say that two stochastic systems are *equivalent* if, for each  $t \in \mathbb{T}$ , their output processes agree a.s. and their state spaces are the same.

As an example, we consider discrete-time stochastic systems with  $\mathbb{T} = \mathbb{Z}_+$ . To this end recall that a *normalized white noise*  $w$  is a sequence of orthogonal random vectors with a unit variance; i.e.,

$$\mathbb{E}\{w(t)w(s)'\} = I\delta_{ts} := \begin{cases} I & \text{if } s = t \\ 0 & \text{if } s \neq t. \end{cases} \quad (2.6.15)$$

**Theorem 2.6.4.** Suppose that  $\mathbb{T} = \mathbb{Z}_+$ . Then, all finite-dimensional stochastic have a representation of the type

$$\begin{cases} x(t+1) = A(t)x(t) + B(t)w(t), & x(0) = x_0, \\ y(t) = C(t)x(t) + D(t)w(t), \end{cases} \quad (2.6.16)$$

where  $\{A(t), B(t), C(t), D(t); t \in \mathbb{T}\}$  are matrices of appropriate dimensions,  $x_0$  a zero-mean random vector, and  $w$  a normalized white noise that is orthogonal to  $x_0$ . Conversely, any pair  $(x, y)$  of stochastic processes satisfying (2.6.16) is a stochastic system.

**Proof.** Let  $(x, y)$  be a stochastic system with  $\mathbb{T} = \mathbb{Z}_+$  and the state process  $x$  taking values in  $\mathbb{R}^n$ . We shall first prove that  $(x, y)$  has a representation (2.6.16). To this end, first note that

$$\begin{bmatrix} x(t+1) \\ y(t) \end{bmatrix} = \mathbb{E}^{\mathbf{H}_t^-(y) \vee \mathbf{X}_t^-} \begin{bmatrix} x(t+1) \\ y(t) \end{bmatrix} + \mathbb{E}^{(\mathbf{H}_t^-(y) \vee \mathbf{X}_t^-)^\perp} \begin{bmatrix} x(t+1) \\ y(t) \end{bmatrix}. \quad (2.6.17)$$

Now, in view of (2.6.12) and Proposition 2.4.2,

$$\mathbb{E}^{\mathbf{H}_t^-(y) \vee \mathbf{X}_t^-} \lambda = \mathbb{E}^{\mathbf{X}_t} \lambda \quad \text{for all } \lambda \in \mathbf{H}_t^+(y) \vee \mathbf{X}_t^+,$$

and consequently there are matrices  $A(t)$  and  $C(t)$  such that

$$\mathbb{E}^{\mathbf{H}_t^-(y) \vee \mathbf{X}_t^-} \begin{bmatrix} x(t+1) \\ y(t) \end{bmatrix} = \begin{bmatrix} A(t) \\ C(t) \end{bmatrix} x(t).$$

The second term in (2.6.17) is an orthogonal sequence and can be normalized to a normalized white noise  $w$  so that

$$\mathbb{E}^{(\mathbf{H}_t^-(y) \vee \mathbf{X}_t^-)^\perp} \begin{bmatrix} x(t+1) \\ y(t) \end{bmatrix} = \begin{bmatrix} B(t) \\ B(t) \end{bmatrix} w(t),$$

where  $B(t)$  and  $D(t)$  are matrices such that  $\begin{bmatrix} B(t) \\ B(t) \end{bmatrix}$  has full rank. Hence  $(x, y)$  satisfies (2.6.16), as claimed. It remains to prove that  $x_0 \perp \mathbf{H}(w)$ . Since however  $\begin{bmatrix} B(t) \\ B(t) \end{bmatrix}$  has full rank,  $w(t) \in (\mathbf{H}_t^-(y) \vee \mathbf{X}_t^-)^\perp$  for all  $t \in \mathbb{Z}_+$ , and hence  $x_0 \perp \mathbf{H}(w)$ .

Conversely, suppose that  $(x, y)$  satisfies (2.6.16). For each  $t \in \mathbb{Z}_+$ , set  $\mathbf{X}_t := \text{span} \{a'x(t) \mid a \in \mathbb{R}\}$ . Since  $w$  is a white noise process and  $x_0 \perp \mathbf{H}(w)$ , the space  $\mathbf{X}_0 \oplus \mathbf{H}^-(w)$  is orthogonal to  $\mathbf{H}^+(w)$ , which, in view of Proposition 2.4.2(v), is equivalent to

$$(\mathbf{X}_0 \oplus \mathbf{H}^-(w)) \perp (\mathbf{H}^+(w) \oplus \mathbf{X}_t) \mid \mathbf{X}_t.$$

However, from (2.6.16) it is easy to see that  $\mathbf{H}^-(y) \vee \mathbf{X}^- \subset \mathbf{X}_0 \oplus \mathbf{H}^-(w)$  and that  $\mathbf{H}^+(y) \vee \mathbf{X}_t^+ \subset \mathbf{H}^+(w) \oplus \mathbf{X}_t$ , and hence (2.6.12) holds (Lemma 2.4.1). Therefore  $(x, y)$  is a stochastic systems.  $\square$

Similar results hold for continuous-time processes and for stationary processes defined on the whole real line. This is a major topic in Chapters 6, 8 and 10.

### Factor analysis

A (static) *factor analysis* model is a representation

$$y = Ax + e, \tag{2.6.18}$$

of  $m$  observable variables  $y = [y_1 \dots y_m]'$ , having zero-mean and finite variance, as linear combinations of  $n$  *common factors*  $x = [x_1 \dots x_n]'$ , plus uncorrelated “noise” or “error” terms  $e = [e_1 \dots e_m]'$ . The  $m$  components of the error  $e$  should be zero-mean and mutually uncorrelated random variables, i.e.

$$\Sigma_{xe} := \mathbb{E}\{xe'\} = 0, \tag{2.6.19a}$$

$$\Delta := \mathbb{E}ee' = \text{diag} \{\sigma_1^2, \dots, \sigma_m^2\}. \tag{2.6.19b}$$

The purpose of these models is to provide an explanation of the mutual interrelation between the observable variables  $y$  in terms of a (hopefully) small number of common factors. In this respect, setting

$$\hat{y}_i := a'_i x, \tag{2.6.20}$$

where  $a'_i$  is the  $i$ -th row of the matrix  $A$ , one has exactly

$$E\{y_i y_j\} = E\{\hat{y}_i \hat{y}_j\}, \tag{2.6.21}$$

for all  $i \neq j$ . In other words, the predictions of the  $y_i$ 's based on the factor vector have the same mutual correlation as the observed variables  $y_i$ . Clearly this property is equivalent to

$$\langle e_i, e_j \rangle = \langle y_i - \hat{y}_i, y_j - \hat{y}_j \rangle = 0, \quad i \neq j$$

which, in view of (2.4.1) just *conditional orthogonality* of  $\{y_1, \dots, y_m\}$ , given  $x$ . We define this concept formally below.

**Definition 2.6.5.** *The random variables  $\{y_1, \dots, y_m\}$  are conditionally orthogonal given  $x$  if for all  $i \neq j$  we have  $y_i \perp y_j \mid x$ .*

It is rather easy to see that  $y$  admits a representation of the type (2.6.18) if and only if  $\{y_1, \dots, y_m\}$  are conditionally orthogonal given  $x$ .

The property of making  $\{y_1, \dots, y_m\}$  conditionally orthogonal, is really a property of the subspace of random variables linearly generated by the scalar components of the vector  $x$

$$\mathbf{X} := \text{span} \{x_k, k = 1, \dots, n\} \tag{2.6.22}$$

which we shall call the *factor subspace* of the model. A factor subspace  $\mathbf{X}$  makes the components of  $y$  conditionally orthogonal given  $\mathbf{X}$ . The variables  $\hat{y}_i$  are then just the orthogonal projections  $\hat{y}_i = E[y_i \mid \mathbf{X}]$ .

Introducing a matrix  $A^\perp$  such that  $A^\perp A = 0$  one can eliminate the factors from the model (2.6.18), obtaining an “external” description in terms of the “true” (unmeasurable) variables  $\hat{y}$  and the additive errors  $\epsilon$ , of the following type

$$A^\perp \hat{y} = 0, \quad y = \hat{y} + \epsilon \tag{2.6.23}$$

which is called an *errors in variables* (EIV) model. The study of models of this type in the statistical literature goes back to the early years of the twentieth century.

A factor subspace may be unnecessarily large just because it carries unnecessary random variables which are uncorrelated (i.e. orthogonal) to the variables  $y$  to be represented. This redundancy can be eliminated by imposing that  $\mathbf{X}$  satisfies the “non-redundancy” condition  $\mathbf{X} = \hat{\mathbf{X}}$  where

$$\hat{\mathbf{X}} = \text{span} \{E[y_i \mid \mathbf{X}]; i = 1, \dots, m\} = E^{\mathbf{X}} \mathbf{Y} \tag{2.6.24}$$

or, equivalently  $\hat{\mathbf{X}} = \text{span} \{[Ax]_i; i = 1, \dots, m\}$ . Since, by Lemma 2.2.6,  $\mathbf{X} = \hat{\mathbf{X}} \oplus (\mathbf{X} \cap \mathbf{Y}^\perp)$ , we have  $E[y_i \mid \mathbf{X}] = E[y_i \mid \hat{\mathbf{X}}]$  and hence an arbitrary factor

space  $\mathbf{X}$  can always be substituted by its non-redundant subspace  $\hat{\mathbf{X}}$  preserving the conditional orthogonality property. From now on, we shall assume that condition (2.6.24) is satisfied.

Any set of generating variables for  $\mathbf{X}$  can serve as a common factor vector. In particular it is no loss of generality to choose the generating vector  $x$  of minimal cardinality (a basis) and normalized, i.e.

$$E\{xx'\} = I, \quad (2.6.25)$$

which we shall do in the following. We may then call  $n = \dim x = \dim \mathbf{X}$  the *rank* of the model. Obviously by virtue of condition (2.6.24), for a model of rank  $n$  we automatically have rank  $A = n$  as well (i.e.  $A$  will always be left invertible).

Two factor analysis models for the same observable  $y$ , whose factors span the same subspace  $\mathbf{X}$  will be regarded as *equivalent*. The convention for the choice of generators implies that two equivalent factor analysis models will have factor vectors related by a real orthogonal transformation matrix.

The common factors are *nonobservable* quantities (also called *latent* variables in the econometric literature) which, although representing the same output variables  $y$ , could in principle be chosen in many different ways giving rise to representations (i.e. models) with different properties and of a different complexity. In the applications one would like to have models with  $n \ll m$  and possibly have some idea about the minimal possible number of factors necessary to represent  $y$ . Models with a minimal number of factors correspond to factor subspaces  $\mathbf{X}$  of minimal dimension. These models will be called *minimal* henceforth.

It is known that there are in general many (in fact infinitely many) minimal factor subspaces for a given family of observables  $\{y_1, \dots, y_m\}$ . Hence there are in general many nonequivalent minimal factor analysis models (with normalized factors) representing a fixed  $m$ -tuple of random variables  $y$ .

For example, by choosing for each  $k \in \{1, \dots, m\}$ , as a factor the  $m - 1$ -dimensional vector  $x := [y_1 \dots y_{k-1} y_{k+1} \dots y_m]'$ , one obtains  $m$  "extremal" models called *elementary regressions*, of the form

$$\begin{cases} y_1 = [1 \dots 0]x + 0 \\ \vdots \\ y_k = \hat{a}'_k x + e_k \\ \vdots \\ y_m = [0 \dots 1]x + 0 \end{cases} \quad (2.6.26)$$

where  $\hat{a}'_k = E y_k x' (E x x')^{-1}$ . Note that in each elementary regression model there is just one nonzero element in the error variance matrix  $\Delta$ . Clearly, the elementary regression (2.6.26) corresponds to EIV models with errors affecting only the  $k$ -th true variable.

In this example the factor subspaces are spanned by  $m - 1$  observable variables. A subspace  $\mathbf{X}$  contained in the *data space*  $\mathbf{Y} := \text{span}\{y_1 \dots y_m\}$  (i.e. generated by linear functionals of  $y$ ) is called *internal*. Accordingly, factor analysis models whose factor  $x$  is made of linear functionals of  $y$ , are called *internal* models.

**Identifiability.** The inherent nonuniqueness of factor analysis models brings up the question of which model one should use in identification. This is called “factor indeterminacy” (or unidentifiability) in the literature and the term is usually referred to parameter unidentifiability as it may appear that in these models there are always “too many” parameters to be estimated. It may be argued that once a model (in essence, a factor subspace) is selected, it can always be parametrized in a one-to-one (and hence identifiable) way. The difficulty seems more a question of understanding the properties of the different possible models, i.e. a question of *classification*. Unfortunately, the classification of all possible (minimal) factor subspaces and an explicit characterization of minimality are, to a large extent, still an open problem.

We shall address here only very superficially the question of identifiability. To this end, we need to consider the additive decomposition of the covariance matrix  $\Lambda := E\{yy'\}$  of the observables induced by a factor analysis model, namely

$$\Lambda = AA' + \Delta. \tag{2.6.27}$$

This is called a *factor analysis decomposition* of  $\Lambda$ . The rank of the model is also called the rank of the decomposition.

Note that for any fixed  $\Lambda \in \mathbb{S}_+^{m \times m}$  (the space of  $m \times m$  symmetric positive definite matrices), once a diagonal matrix  $\Delta$  such that  $\text{rank}\{\Lambda - \Delta\} = n$  has been found, the matrix  $A$  in the decomposition (2.6.27) is just a full rank factor of  $\Lambda - \Delta$ , i.e. an  $m \times n$  matrix satisfying  $AA' = \Lambda - \Delta$ . Such a factor can be rendered unique by choosing an appropriate canonical form in the equivalence class<sup>6</sup> of  $A \in \mathbb{R}_*^{m \times n}$  defined modulo right multiplication by  $n \times n$  orthogonal matrices. The following is the main question concerning identifiability.

**Hidden rank.** What is the minimal  $n$  for which a given  $\Lambda$  admits a factor analysis decomposition of rank  $n$ . This number,  $n_*(\Lambda)$ , (often denoted  $mr(\Lambda)$  in the literature) will be called the *hidden rank* of  $\Lambda$ . Clearly  $n_*(\Lambda) \leq m - 1$  for all  $\Lambda$ . Note that diagonal  $\Lambda$ 's admit a (unique) trivial factor analysis decomposition of rank zero. Conditions for  $\Lambda$  to admit a factor analysis decomposition of rank one ( $n_*(\Lambda) = 1$ ) have been known since the beginning of the 20th century, (in the literature a positive definite covariance matrix admitting a factor analysis decomposition of rank one is called a *Spearman matrix*). The hidden rank question is unsolved.

The decompositions of rank  $m - 1$  are particularly simple to describe. In fact, the solutions are described in terms of the coordinates  $\{\sigma_1^2, \dots, \sigma_m^2\}$  in the space of nonnegative definite diagonal matrices  $\Delta = \text{diag}\{\sigma_1^2, \dots, \sigma_m^2\}$  by the polynomial equation  $\det(\Lambda - \Delta) = 0$ , subject to the constraint that all principal minors of order  $m - 1$  of  $\Lambda - \Delta$  be nonnegative with at least one being nonzero. These algebraic conditions define a smooth hypersurface (an hyperboloid with concavity facing the origin) in the positive orthant of  $\mathbb{R}^m$ . Moreover, this hypersurface intersects the  $k$ -th coordinate axis exactly at the value  $\sigma_k^2$  equal to the error variance of the  $k$ -th elementary regressor.

**Factor estimation.** In identification of a factor analysis model from observed data there is yet another source of difficulty which has to do with the factor vector. How

<sup>6</sup>We denote by  $\mathbb{R}_*^{m \times n}$  the space of *full rank*  $m \times n$  real matrices.

does one obtain an estimate of  $x$ ? This problem is often overlooked in textbooks because the models most often used are of the “elementary regression” type, and hence it is known that  $x$  is a function of the observed data. For example, the ARMAX models used in dynamic system identification are of this type, since a component of the observed variables (the input usually denoted by the letter  $u$ ) is treated as a “true” variables with a tacit assumption of no “observation noise” superimposed. In all internal models the auxiliary variables (say the state variables in a state-space model, or the white noise input variable in ARMAX models) are deterministic functions of the observed data (in fact causal functions for innovations-like models). These functions have known structure depending only on the unknown parameters of the model. Estimation of the auxiliary variables is then automatically accomplished after model-parameter estimation. In more general situations however this does not happen. When the model is noninternal, the estimation of the auxiliary variables eventually has to be handled by an appropriate methodology. For most factor analysis models the factor variables are indeed noninternal.

**Proposition 2.6.6.** *All internal factor analysis models are regressions. All non-trivial factor analysis models with  $\Delta > 0$  are noninternal.*

*Proof.* To prove the first statement notice that a model is internal if and only if

$$\mathbf{X} \cap (\mathbf{Y}_1 + \mathbf{Y}_2 + \cdots + \mathbf{Y}_n) = \mathbf{X}$$

but by virtue of splitting, the first member is equal to the vector sum  $\mathbf{X} \cap \mathbf{Y}_1 + \cdots + \mathbf{X} \cap \mathbf{Y}_n$  (Proposition 2.4.4). Since  $\dim \mathbf{Y}_k = 1$ , either  $\mathbf{X} \cap \mathbf{Y}_k$  is equal to  $\mathbf{Y}_k$  or is the zero subspace. Hence

$$\mathbf{X} = \bigvee_{\mathbf{Y}_k \subset \mathbf{X}} \mathbf{Y}_k$$

i.e.  $\mathbf{X}$  is spanned by a finite number of  $y_k$ 's, and the model is a regression.

Next, assume  $x \neq 0$  and internal. Then there is some  $n \times m$  matrix  $B$  such that  $x = By$ . Imposing the orthogonality  $x \perp e$  we get

$$B\Lambda(I - AB)' = 0 \tag{2.6.28}$$

Moreover by definition of  $\Delta$  the matrix  $B$  satisfies also

$$(I - AB)\Lambda(I - AB)' = \Delta.$$

Now, by (2.6.28) this last equation can be rewritten as

$$\Lambda(I - AB)' = \Delta$$

which combined again with (2.6.28) yields  $B\Delta = 0$ . Since  $\Delta > 0$  it follows that  $B$  must be zero and hence  $x = 0$ , the degenerate situation which has just been excluded. It follows that  $x$  cannot be a linear function of  $y$ .  $\square$

Estimation of auxiliary variables in factor analysis models may be approached from the point of view of stochastic realization theory. The following Theorem describes how to construct the auxiliary variable  $x$  starting from the observable random quantities and from the knowledge of the parameters  $(A, \Delta)$  of the structured covariance matrix of the data.

**Theorem 2.6.7.** *Every normalized common-factors vector for the factor analysis model  $y = Ax + e$ ,  $E\{ee'\} = \Delta$ , has the form*

$$x = A'\Lambda^{-1}y + z, \tag{2.6.29}$$

where  $z$  is an  $n$ -dimensional zero-mean random vector orthogonal to  $\mathbf{Y}$  with covariance  $I - A'\Lambda^{-1}A$ .

**Proof.** Necessity: let  $x$  and  $e$  be as in the statement of the theorem. The factor vector  $x$  can be written as the orthogonal sum

$$x = E[x | \mathbf{Y}] + z$$

where  $E[x | \mathbf{Y}] = A'\Lambda^{-1}y$  and  $z$  is the “estimation error” of the vector  $x$  based on  $y$ . It is immediate to check that the covariance matrix of  $z$  has the required form.

Sufficiency: let  $x$  be as stated. Define  $B := A'\Lambda^{-1}$  and  $w := y - Ax = (I - AB)y - Az$  (so that  $y = Ax + w$  by construction). We proceed to check that  $x \perp w$  and that the covariance matrix of  $w$  is exactly  $\Delta$ . For

$$E\{xw'\} = E\{(By + z)((I - AB)y - Az)'\} \tag{2.6.30}$$

$$= B\Lambda(I - B'A') - (I - B\Lambda B')A' = 0. \tag{2.6.31}$$

Moreover,

$$E\{ww'\} = (I - AB)\Lambda(I - AB)' + A(I - B\Lambda B')A' = \Lambda - AA' = \Delta,$$

as claimed.  $\square$

In practice we must estimate both the parameters (identification) and the auxiliary variable  $x$  (factor estimation) of the model, starting from the observed data. In this problem we have both deterministic and random parameters which must be treated differently. It is actually not obvious how to carry out this program from first principles.

### Causality and feedback-free processes

In the 1960's there was a long debate in the econometric literature regarding the notion of *causality* of time series. In mathematical terms, one would like an “intrinsic” (and testable) definition of when one stochastic process “causes” another.

Let  $y$  and  $u$  be two vector stochastic processes, which we shall assume jointly stationary. This property is assumed here for reasons of simplicity but is not needed

in the definition, as the concept applies to much more general situations. In general one may express both  $y$  and  $u$  as a sum of the best linear estimate based on the past and present of the other variable, plus error terms:

$$y(t) = E[y(t) | \mathbf{H}_{t+1}^-(u)] + v(t) \tag{2.6.32a}$$

$$u(t) = E[u(t) | \mathbf{H}_{t+1}^-(y)] + r(t) \tag{2.6.32b}$$

so that each variable  $y(t)$  and  $u(t)$  can be expressed as a sum of a causal linear transformation of the past of the other, plus noise. Here the noise terms are uncorrelated with the past of  $u$  and  $y$  respectively but may in general be mutually correlated.

To carry on the discussion we shall need to anticipate some of the notions which will be introduced later on. As it will be explained in Chapter 4, each of the linear estimators above can be expressed as the output of a linear filter, represented by a causal transfer function,  $F(z)$  and  $H(z)$  respectively, so that the joint model (2.6.32) corresponds to a block diagram of the type

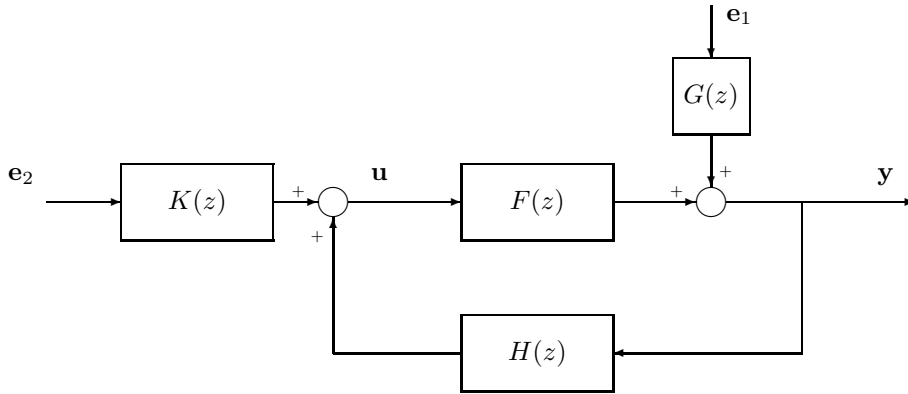


Figure 2.6.1. Joint model of the signals  $y$  and  $u$ .

This diagram shows that there is an *intrinsic feedback* mechanism relating the two processes. The concept of causality will have to be related to this mechanism.

Following Granger [41], we say that *there is no feedback from  $y$  to  $u$*  if the future of  $u$  is conditionally uncorrelated with the past of  $y$ , given the past of  $u$  itself. In our Hilbert space framework this is written as

$$\mathbf{H}_t^+(u) \perp \mathbf{H}_t^-(y) | \mathbf{H}_t^-(u). \tag{2.6.33}$$

Equation (2.6.33) expresses the fact that the future time evolution of the process  $u$  is not influenced by the past of  $y$  once the past of  $u$  is known. This captures in a coordinate free way the absence of feedback (from  $y$  to  $u$ ). Taking  $\mathbf{A} = \mathbf{H}_t^+(u)$  and  $\mathbf{B} = \mathbf{H}_t^-(y)$ , condition (iii) of Proposition 2.4.2 shows that the feedback-free condition is equivalent to  $\mathbf{H}_t^-(y) \perp \mathbf{H}_t^+(u) | \mathbf{H}_t^-(u)$  and hence, by (iv), to  $E[\mathbf{H}_t^-(y) | \mathbf{H}_t^+(u)] = E[\mathbf{H}_t^-(y) | \mathbf{H}_t^-(u)]$  for all  $t \in \mathbb{Z}$ , so that, using  $t + 1$  instead of  $t$ , we get in particular

$$E[y(t) | \mathbf{H}(u)] = E[y(t) | \mathbf{H}_{t+1}^-(u)] \quad \text{for all } t \in \mathbb{Z}. \tag{2.6.34}$$



namely, the noncausal estimate of  $y(t)$ ,  $E[y(t) | \mathbf{H}(u)]$ , given the whole history of  $u$ , depends only on the past and present values of the process  $u$  and not on its future history. Equation (2.6.34) can be taken as a definition of *causality*. In this case, it is appropriate to call  $u$  an *input* variable as one can interpret  $u$  as an exogenous cause of the evolution of  $y$  but not conversely. One says that *there is causality from  $u$  to  $y$*  (or that  $u$  “causes”  $y$ ). If this condition holds, it can be shown that, if there are no cancellations in forming the transfer function,  $H = 0$  in the feedback loop of Figure 2.6.1. This is discussed, for example, in [37].

It also follows from 2.6.34 that

$$y_s(t) := y(t) - E[y(t) | \mathbf{H}_{t+1}^-(u)] = y(t) - E[y(t) | \mathbf{H}(u)] = E[y(t) | \mathbf{H}(u)^\perp] \quad (2.6.35)$$

so that  $y_s(t) \perp \mathbf{H}(u)$  for all  $t$ , i.e. the “causal estimation error” is uncorrelated with the whole history of the input process  $u$ . We shall call the process  $y_s$  the *stochastic component of  $y$* . (In the feedback-free case, this notation is used for the “estimation error”  $v$  of (2.6.32)). Similarly, the stochastic process  $y_d$  defined by the complementary projection

$$y_d(t) := E[y(t) | \mathbf{H}(u)], \quad t \in \mathbb{Z} \quad (2.6.36)$$

is called the *deterministic component of  $y$* .

For convenience, let  $\tilde{\mathbf{H}}(y)$  be the Hilbert subspace of  $\mathbf{H}(u) \vee \mathbf{H}(y)$  linearly generated by  $\{y_s(t) | t \in \mathbb{Z}\}$  and let  $\hat{\mathbf{H}}(y)$  be the Hilbert subspace of  $\mathbf{H}(u)$  linearly generated by  $\{y_d(t) | t \in \mathbb{Z}\}$ . Note that not only do we have  $\mathbf{H}(y) \vee \mathbf{H}(u) = \tilde{\mathbf{H}}(y) \oplus \mathbf{H}(u)$ , but in virtue of causality,

$$\mathbf{H}_t^-(y) \vee \mathbf{H}(u) = \tilde{\mathbf{H}}_t^-(y) \oplus \mathbf{H}(u) \quad (2.6.37)$$

for all  $t$ , where  $\tilde{\mathbf{H}}_t^-(y)$  is the past space of the process  $y_s$  at time  $t$ . We stress that the stochastic and deterministic components in the decomposition

$$y(t) = y_s(t) + y_d(t) \quad (2.6.38)$$

are completely uncorrelated, i.e.,  $E\{y_s(t)y_d(\tau)'\} = 0$  for all  $t, \tau \in \mathbb{Z}$ .

If there is no causality, or, equivalently, if there is feedback from  $y$  to  $u$ , the very notion of input loses its meaning. In fact, as shown in (2.6.32), the variable  $u(t)$  is then also determined by a dynamical relation involving the past output process  $y$ , which in turn is now playing the role of an exogenous variable determining  $u$ .

Identification in the presence of feedback (and, of course, in the absence of any other specific information on the feedback loop) is in general equivalent to identification of the *joint process*  $(y, u)$ , in the sense of time-series identification. Let  $e_s(t)$  be the one-step prediction error of the process  $y_s$  based on its own past  $\tilde{\mathbf{H}}_t^-(y)$ , i.e.,

$$e_s(t) = y_s(t) - E[y_s(t) | \tilde{\mathbf{H}}_t^-(y)] \quad (2.6.39)$$

The process  $e_s$  is the (*forward*) *innovation process* of  $y_s$ .

**Proposition 2.6.8.** *In the feedback-free case, the innovation of the process  $y_s$  is the conditional innovation of  $y$  given observations of  $u$  up to the present time. More*

precisely, if (2.6.34) holds, then

$$\begin{aligned} e_s(t) &= y(t) - \mathbf{E}\{y(t) \mid \mathbf{H}_{t+1}^-(u) \vee \mathbf{H}_t^-(y)\} = \\ &= y(t) - \mathbf{E}\{y(t) \mid \mathbf{H}_t^-(y) \vee \mathbf{H}(u)\} \end{aligned}$$

**Proof.** From the first of equations (2.6.35) and the fact that  $y_d(t) \perp \tilde{\mathbf{H}}_t^-(y)$ , we have that

$$e_s(t) = y(t) - \mathbf{E}[y(t) \mid \mathbf{H}_{t+1}^-(u)] - \mathbf{E}[y(t) \mid \tilde{\mathbf{H}}_t^-(y)].$$

However,  $\mathbf{H}_{t+1}^-(u) \vee \mathbf{H}_t^-(y) = \mathbf{H}_{t+1}^-(u) \oplus \tilde{\mathbf{H}}_t^-(y)$ , and hence, in view of Lemma 2.2.5, the first of the equations in the statement of the proposition follows. The second statement follows in the same way from the second equation in (2.6.35) by observing (2.6.37).  $\square$

## 2.7 Stationary increments processes in continuous-time

All the concepts introduced in the previous sections for stationary discrete-time processes have obvious continuous-time counterparts. However, in continuous time, the notion of stationary process may not be the most interesting concept from the point of view of applications. Most of the interesting continuous-time signals in engineering are modeled as “wideband” signals, and often the mathematically simplest description is a process with a superimposed “white noise” component. For this reason we shall now introduce the notion of a *stationary increments* process, which will allow us to deal rather naturally with this class of objects.

Let  $z := \{z(t); t \in \mathbb{R}\}$  be an  $m$ -dimensional continuous-time process defined on some probability space  $\{\Omega, \mathcal{F}, P\}$ . We shall assume that the increments  $\{z_k(t) - z_k(s); t, s \in \mathbb{R}, k = 1, 2, \dots, m\}$  have zero mean and finite second moment. If all covariances

$$\mathbf{E}\{(z_k(t+h) - z_k(s+h))(z_j(t) - z_j(s))\}; \quad t, s \in \mathbb{R}, k, j = 1, 2, \dots, m \quad (2.7.1)$$

are independent of  $h$ , we say that  $z$  has *stationary increments*.

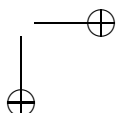
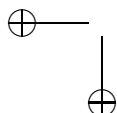
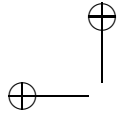
Consider the Hilbert subspace of  $L^2\{\Omega, \mathcal{F}, P\}$  linearly generated by the increments of  $z$

$$\mathbf{H}(dz) := \overline{\text{span}}\{z_k(t) - z_k(s); t, s \in \mathbb{R}, k = 1, 2, \dots, m\} \quad (2.7.2)$$

It is clear that, if  $z$  has stationary increments, the operators  $U_h$ , defined, for any  $h \in \mathbb{R}$ , on a dense subset of  $\mathbf{H}(dz)$ , by

$$U_h(z_k(t) - z_k(s)) = z_k(t+h) - z_k(s+h); \quad t, s \in \mathbb{R}, k = 1, 2, \dots, m \quad (2.7.3)$$

are isometric, and can be extended to the whole of  $\mathbf{H}(dz)$  to form a one-parameter *unitary group*  $\{U_t; t \in \mathbb{R}\}$ .



In what follows all processes with stationary increments that we shall encounter will have *increments which are continuous in mean square*; i.e.  $z_k(t+h) - z_k(s+h) \rightarrow z_k(t) - z_k(s)$  as  $h \rightarrow 0$  for all  $t, s \in \mathbb{R}$ ,  $k = 1, 2, \dots, m$ . In this case the unitary group  $\{U_t; t \in \mathbb{R}\}$  will be strongly continuous.

Generally speaking, processes with stationary increments are integrated versions of the random signals which are being modelled, and the only thing of interest are the increments. For this reason each such process  $\{z(t)\}$  is viewed as an equivalence class defined up to an additive constant random vector  $z_0$ . This equivalence class is denoted by the symbol  $dz$ . Obviously, in case  $\{z(t)\}$  is differentiable in mean square, there exists a (mean-square) derivative process  $\{s(t)\}$  for which we can write  $z(t) - z(s) = \int_s^t s(\tau) d\tau$ , or symbolically,  $dz(t) = s(t)dt$ . It is easy to check that the derivative must be stationary; i.e.  $s(t+h) = U_h s(t)$ . However in general this will not be the case in many applications of interest. In general, under a very mild *conditional Lipschitz condition*, which is discussed in detail in [84], a stationary increments process admits semimartingale representations of the type,

$$dz(t) = s(t)dt + Ddw(t) \quad (2.7.4)$$

where  $\{s(t)\}$  is stationary,  $D$  is a constant  $m \times p$  matrix and  $dw$  is a  $p$ -dimensional normalized *Wiener process*, that is a process with stationary orthogonal increments, which plays the role of integrated white noise. These processes will be studied in detail in the next chapter.

## 2.8 Bibliographical notes

The material of Section 2.2 is standard. Proofs of the orthogonal projection Lemma can be found in the textbooks [44, 124]. The modern definition of conditional expectation was given by Kolmogorov in [66], see also [23]. The interpretation as orthogonal projection operator in  $L^2$  can be found in the first chapter of [98]. The role of the Moore-Penrose pseudoinverse (see e.g. [40, p.139]) in the expression of the conditional expectation of (conditionally) Gaussian random vectors has been emphasized by [88].

Section 2.3. The singular value decomposition (SVD) for compact operators is discussed for example in [22, p. 333]; conditions for compactness of the *Hankel operator*  $E_{\mathbf{B}}^{\mathbf{A}}$  are discussed in [101]. The optimization characterization of the singular values, which generalizes the so-called Rayleigh quotient iteration in Euclidean spaces, is elegantly discussed in [129, p.204]. The SVD for finite dimensional operators is now a standard device in linear algebra. See e.g. [40] and the references therein.

Section 2.4. Conditional uncorrelation and conditional independence are standard notions in probability theory. These concepts play a very important role in modeling and realization of stochastic systems. For this reason they have been deepened and reformulated in various equivalent ways in the stochastic realization literature in view of answering basic system-theoretic questions like stochastic minimality etc. Proposition 2.4.2 is as formulated in [85].

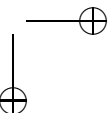
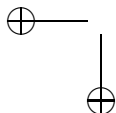
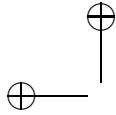
Section 2.5. Besides Komogorov's original papers, [65, 63], the classical ref-

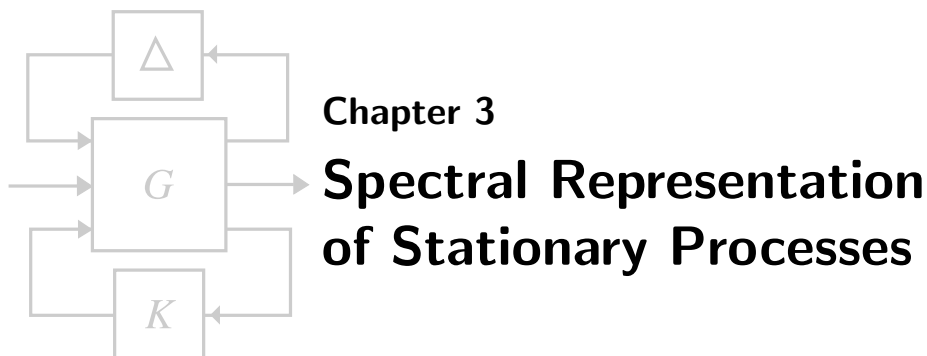
erences for the material discussed here are the papers by Cramèr [17, 19], who in particular stressed the notion of multiplicity and its relation with stationarity [20, 21], Karhunen, [62], Wold [123]. A basic reference for the linear theory of stationary stochastic processes is Rozanov's book [106].

Section 2.6. The operator-theoretic formulation of the Markov property discussed in this section seems to be originally due to [79]. It plays an important role in the rest of this book. Theorem 2.6.4 and its proof is taken from [87]. Factor analysis (and EIV) modeling is an old problem in statistics and econometrics which has been revitalized in recent years by Kalman [57, 58, 59]. Our discussion here is based on [9, 8, 70, 71, 103].

The concepts of causality and its relation to (absence of) feedback between stochastic processes has been introduced by Granger in [41]. The study of feedback between stochastic processes has generated a large literature. See for example [15, 37, 3]. That Granger definition of absence of feedback is a conditional orthogonality (or, more generally, a conditional independence) condition is nearly obvious but does not seem to have been appreciated in the literature.

Stationary increments processes in continuous time are discussed in Chapter 1 of [38].





In this chapter we review the the so-called *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 of 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  $\ell^2$  nor uniformly bounded with probability one, and hence as functions of time they do not admit a Fourier transform.

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

$$E\{(x(t_2) - x(s_2))\overline{(x(t_1) - x(s_1))}\} = 0 \tag{3.1.1}$$

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

$$E(x(t) - x(s)) = 0 \quad t, s \in \mathbb{I}. \tag{3.1.2}$$

We alert the reader to the fact that *complex* orthogonal increment processes defined on a imaginary “time” axis will play an important role in spectral representation, discussed in Section 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,*

$$\mathbb{E}\{|x(t) - x(s)|^2\} = F(t) - F(s), \quad t \geq s \quad (3.1.3)$$

**Proof.** Let us fix an arbitrary  $t_0$  and define,

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

Then by using the property (3.1.1), it is immediate to check that  $F_0$  is monotone and satisfies (3.1.3). The function  $F_0$  is clearly the unique function satisfying (3.1.3) normalized at  $t_0$  so as  $F_0(t_0) = 0$ . Hence any function  $F(t) := F_0(t) +$  an arbitrary constant, also satisfies (3.1.3) and is independent of  $t_0$ .  $\square$

The relation (3.1.3) is often written symbolically as

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

It follows from (3.1.3) 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 so as to have  $x(t+) = x(t)$  (and hence also  $F(t+) = F(t)$ ) for all  $t \in \mathbb{T}$ . If say  $\mathbb{T} = (a, b]$  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 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  $\sigma^2$  is a positive constant. Hence for a Wiener process, we have  $\mathbb{E}\{|dw(t)|^2\} = \sigma^2 dt$ . 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. [124]. 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<sup>7</sup>. An orthogonal stochastic measure on  $\mathbb{R}$  is a family of random variables  $\zeta : \{\zeta(\Delta); \Delta \in \mathcal{R}\}$  where  $\zeta(\Delta) : \{\Omega, \mathcal{A}, \mu\} \rightarrow \mathbb{C}$  such that

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

$$m(\Delta) = E\{|\zeta(\Delta)|^2\} < \infty, \quad \Delta \in \mathcal{R} \tag{3.1.4}$$

2. For any pair of disjoint intervals  $\Delta_1, \Delta_2$  with  $\Delta_1 \cap \Delta_2 = \emptyset$ ,

$$E\{\zeta(\Delta_1)\overline{\zeta(\Delta_2)}\} = 0 \tag{3.1.5}$$

3.  $\zeta$  is  $\sigma$ -additive, i.e. for any  $\Delta \in \mathcal{R}$  which is the disjoint union of countably many sets  $\Delta_k \in \mathcal{R}$ ,

$$\zeta(\Delta) = \sum_{k=1}^{\infty} \zeta(\Delta_k), \quad a.s. \tag{3.1.6}$$

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

Note that by Lemma A.1.1 in Appendix A.1 the series of orthogonal random variables (3.1.6) converges if and only if

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

so that  $m$  is a nonnegative  $\sigma$ -additive set function which can be extended as a  $\sigma$ -finite measure on the the Borel  $\sigma$ -algebra of sets generated by  $\mathcal{R}$  see e.g. [42, 38]. Conversely,  $m$  being  $\sigma$ -additive on  $\mathcal{R}$  implies that  $\zeta$  is  $\sigma$ -additive in the sense of (3.1.6) above. In this sense, it is then possible to extend  $\zeta$  to the  $\sigma$ -ring generated by  $\mathcal{R}$ , where  $m(\Delta) < \infty$ , see also [106, p. 5]. Note that  $\zeta$  may not be extendable to unbounded sets.

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

<sup>7</sup>The family  $\mathcal{R}$  is a *semi-ring* of sets, see [42, 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.

The notion of orthogonal stochastic measure is the natural starting point to discuss stochastic integration. Before embarking into 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  $\zeta$  determines an orthogonal increments process  $z$  by the position

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

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

In particular, the stochastic orthogonal measure corresponding to the normalized Wiener process  $w$ , has  $m =$  Lebesgue measure. Since in this book the only thing that will matter will be the increments of  $w$ , it will be convenient to identify a Wiener process with the corresponding orthogonal stochastic measure  $dw$ , so in the future, whenever we shall 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  $\zeta$ . Let  $I_\Delta$  denote the indicator function of the set  $\Delta$ , i.e.  $I_\Delta(t) = 1$  if  $t \in \Delta$  and zero otherwise. For a scalar *simple function*

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

the integral of  $f$  with respect to  $\zeta$  is defined as follows,

$$\int_{\mathbb{R}} f(t) d\zeta(t) := \sum_{k=1}^N c_k \zeta(\Delta_k). \tag{3.1.7}$$

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

$$\mathbf{L}(\zeta) := \text{span}\{\zeta(\Delta) \mid \Delta \in \mathcal{R}\} = \text{span}\{\zeta((a, b]) \mid -\infty < a < b < +\infty\} \tag{3.1.8}$$

generated by the increments of  $\zeta$ .



The fundamental property of the stochastic integral of simple functions is

$$\mathbb{E}\left\{\left|\int_{\mathbb{R}} f(t)d\zeta(t)\right|^2\right\} = \sum_{k=1}^N |c_k|^2 m(\Delta_k) = \int_{\mathbb{R}} |f(t)|^2 dm, \quad (3.1.9)$$

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  $\mathcal{J}_\zeta$ . Using this compact notation the formula (3.1.9) reads

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

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  $f_n$ ,

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

so that by the isometric property of the integral

$$\|\mathcal{J}_\zeta(f_n) - \mathcal{J}_\zeta(f_k)\| = \|f_n - f_k\|_{L^2(\mathbb{R}, dm)} \rightarrow 0$$

as  $n, k \rightarrow \infty$ . Therefore the sequence  $\{\mathcal{J}_\zeta(f_n)\}$  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  $\zeta$ . In other words, for an arbitrary  $f \in L^2(\mathbb{R}, dm)$ , the stochastic integral of  $f$  with respect to  $\zeta$  is the mean square limit

$$\mathcal{J}_\zeta(f) = \int_{\mathbb{R}} f(t)d\zeta(t) := \lim_{n \rightarrow \infty} \int_{\mathbb{R}} f_n(t) d\zeta(t). \quad (3.1.10)$$

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 and will be omitted.

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

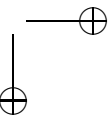
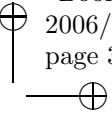
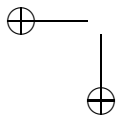
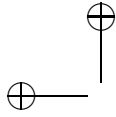
$$\mathbb{E}\left\{\int_{\mathbb{R}} f(t)d\zeta(t) \overline{\int_{\mathbb{R}} g(t)d\zeta(t)}\right\} = \int_{\mathbb{R}} f(t)\bar{g}(t) dm \quad (3.1.11)$$

*in other words,  $\mathcal{J}_\zeta$  is a unitary map  $L^2(\mathbb{R}, dm) \rightarrow \mathbf{H}(\zeta)$ .*

We shall also 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*

$$\eta(\Delta) := \int_{\Delta} f(t)d\zeta(t) = \int_{\mathbb{R}} I_{\Delta}(t)f(t)d\zeta(t) \quad (3.1.12)$$



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

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

### 3.2 Harmonic analysis of a stationary process

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

Let  $\tau \rightarrow \Lambda(\tau)$  be the covariance function of a scalar stationary random process<sup>8</sup>  $y$ , where  $\tau \in \mathbb{Z}$  if the process is discrete time and  $\tau \in \mathbb{R}$  in the continuous time case. In continuous time  $\Lambda$  will be assumed to be a continuous function<sup>9</sup> of  $\tau \in \mathbb{R}$ .

**Theorem 3.2.1 (Herglotz, Bochner).** *There is a finite positive measure  $dF$  on the Borel subsets of the interval  $[-\pi, \pi]$  (discrete time) or  $(-\infty, +\infty)$  (continuous time), such that*

$$\Lambda(\tau) = \int e^{i\theta\tau} dF(\theta). \tag{3.2.1}$$

*the limits of the integral being  $(-\pi, \pi)$  (discrete time) or  $(-\infty, +\infty)$  (continuous time). The measure  $dF$  is uniquely determined by  $\Lambda$ .*

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 (3.2.1) holds. The monotone function  $F$ , uniquely determined by  $\Lambda$  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

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

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 process sum of simple harmonic oscillations

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

<sup>8</sup>Recall that all stationary processes considered in this book will have finite second-order moments.

<sup>9</sup>This is equivalent to assuming  $y$  mean-square continuous.

where  $-\pi < \theta_k \leq \pi$  are deterministic frequencies and  $y_k$  are mutually uncorrelated zero-mean random variables with variance  $\sigma_k^2$ . 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 (3.2.1) 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  $\theta_k$ . 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

$$F = F_1 + F_2 \tag{3.2.2}$$

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.  $F_2$  carries all discontinuities (finite jumps) of  $F$ . The non-negative function  $\Phi$  is called the *spectral density function* of the process.

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

$$\sum_{\tau=-\infty}^{+\infty} e^{-i\theta\tau} \Lambda(\tau) \tag{3.2.3}$$

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

$$\Lambda(\tau) = \int_{-\pi}^{+\pi} e^{i\theta\tau} \hat{\Lambda}(\theta) \frac{d\theta}{2\pi}. \tag{3.2.4}$$

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  $\Phi$  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  $\Phi$  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 lies at the grounds of a stochastic Fourier-like representation for the process  $y$  itself. This representation theorem is important as it provides very precise informations about the structure of the elements of the space  $\mathbf{H}(y)$ .

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

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

$$\mathcal{J}\left(\sum_k c_k e_k\right) := \sum_k c_k y(k), \quad k \in \mathbb{Z}, c_k \in \mathbb{C} \tag{3.3.1}$$

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

$$\mathbf{L}(y) := \text{span}\{y(t); t \in \mathbb{Z}\}. \tag{3.3.2}$$

Now, it follows from Weierstarss 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 Natanson book [96, 97]. Then, by a simple application of Herglotz's Theorem one can see that the map  $\mathcal{J}$  is *isometric*, as

$$\langle e_k, e_j \rangle_{L^2\{[-\pi, \pi], dF\}} = \Lambda(k - j) = \langle y(k), y(j) \rangle_{\mathbf{H}(y)} \tag{3.3.3}$$

and hence, since any  $\hat{f} \in L^2\{[-\pi, \pi], dF\}$  is the mean square limit of a sequence of trigonometric polynomials  $\hat{f}_k$ ,  $\mathcal{J}$  can be extended by continuity to the whole of  $L^2\{[-\pi, \pi], dF\}$ . In fact, by (3.3.3),  $\mathcal{J}(\hat{f}_k)$  also converges in mean square to some

random variable in  $\mathbf{H}(y)$ . We just define  $\mathcal{J}(\hat{f})$  to be this limit

$$\mathcal{J}(\hat{f}) := \lim_{k \rightarrow \infty} \mathcal{J}(\hat{f}_k)$$

in  $L^2(\Omega, \mathcal{A}, \mu)$ . In this way the extended map (still denoted by)  $\mathcal{J}$ , becomes a *unitary map* from  $L^2\{[-\pi, \pi], dF\}$  onto  $\mathbf{H}(y)$ . 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  $-\pi < \theta \leq \pi$ , such that*

$$\mathcal{J}(\hat{f}) = \int_{-\pi}^{+\pi} \hat{f}(\theta) d\hat{y}(\theta), \quad \hat{f} \in L^2\{[-\pi, \pi], dF\}, \quad (3.3.4)$$

so that, in particular

$$y(t) = \int_{-\pi}^{\pi} e^{i\theta t} d\hat{y}(\theta), \quad t \in \mathbb{Z}. \quad (3.3.5)$$

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

$$\mathbf{E}\{d\hat{y}(\theta)\} = 0, \quad \mathbf{E}\{|d\hat{y}(\theta)|^2\} = dF(\theta) \quad (3.3.6)$$

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 (3.3.5). Formula (3.3.5) 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  $\mathcal{J}$  corresponding to a specific process  $y$  will hereafter be denoted by  $\mathcal{J}_{\hat{y}}$ .

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

$$\hat{y}(\Delta) := \mathcal{J}(I_{\Delta}) \quad (3.3.7)$$

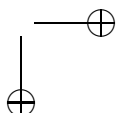
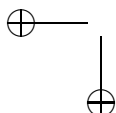
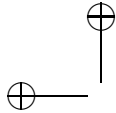
so that by the isometric character of  $\mathcal{J}$  we have  $\mathbf{E}\{|\hat{y}(\Delta)|^2\} = \|I_{\Delta}\|_{L^2\{[-\pi, \pi], dF\}}^2 = F(\Delta)$ . Here we have denoted by  $F$  also the Borel measure induced by the spectral distribution function  $F$ . Also, for an arbitrary pair of intervals  $\Delta_1, \Delta_2$  we have

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

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 (3.3.7). Obviously this measure is finite as  $\mathbf{E}\{|\hat{y}((-\pi, \pi])|^2\} = dF((-\pi, \pi]) = F(\pi) < \infty$  and can then be extended to the Borel sets of the interval  $[-\pi, \pi]$ .

We now proceed to show that (3.3.4) holds for all  $\hat{f} \in L^2\{[-\pi, \pi], dF\}$ . This is certainly true for simple functions since in this case

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



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  $\{\mathcal{J}(\hat{f}) \mid \hat{f} \text{ simple}\}$  is dense in  $\mathbf{H}(y)$ . Hence any random variable  $\xi \in \mathbf{H}(y)$  being the limit in mean square of a sequence  $\mathcal{J}(\hat{f}_k)$  with  $\hat{f}_k$  simple functions, is at the same time the limit of a sequence of stochastic integrals of simple functions  $\mathcal{J}_{\hat{y}}(\hat{f}_k)$ . 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  $\hat{y}(\Delta)$ 's 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  $\mathbf{H}(y)$ .  $\square$

### Connections with 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 Rozanov's book [106, p. 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)

$$Y(\theta) = \lim_{N \rightarrow \infty} \sum_{t=-N}^{+N} e^{-i\theta t} y(t). \tag{3.3.8}$$

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 (3.3.8) with respect to  $\theta$  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

$$\lim_{N \rightarrow \infty} \sum_{t=-N}^{+N} \chi_t(\Delta) y(t) \tag{3.3.9}$$

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

$$S_N(\theta) := \sum_{t=-N}^{+N} \chi_t(\Delta) e^{i\theta t} \tag{3.3.10}$$

converges pointwise as  $N \rightarrow \infty$  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  $I_\Delta$  at the extreme points of the interval, in order to have pointwise convergence also at  $\theta_1, \theta_2$ .

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

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

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

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

(c) The last integral in the equation is just the integrated Fourier series (3.3.9).

3. In this sense one may say that the formal Fourier series (3.3.8) converges to the white noise  $Y(\theta)$  on  $[-\pi, \pi]$  as  $N \rightarrow \infty$ .

### 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*

$$y(t) = \int_{-\infty}^{+\infty} e^{i\omega t} d\hat{y}(i\omega), \quad t \in \mathbb{R}. \tag{3.3.11}$$

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

$$E\{d\hat{y}(i\omega)\} = 0, \quad E\{|d\hat{y}(i\omega)|^2\} = dF(i\omega) \tag{3.3.12}$$

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

$$\mathcal{J}_{\hat{y}}(\hat{f}) = \int_{-\infty}^{+\infty} \hat{f}(i\omega) d\hat{y}(i\omega), \quad \hat{f} \in L^2\{(-\infty, +\infty), dF\}, \tag{3.3.13}$$

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 of  $\mathbf{H}(y)$  correspond to elements of the function 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. Every random element  $\xi \in \mathbf{H}(y)$  can be written in a unique way as a stochastic integral  $\mathcal{J}_{\hat{y}}(\hat{f})$  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  $\mathcal{J}_{\hat{y}} : L^2\{[-\pi, \pi], dF\} \rightarrow \mathbf{H}(y)$  is isometric and bijective, i.e. unitary. It transforms the shift operator  $U$  into the operator of multiplication by the exponential function  $e^{i\theta} : \theta \rightarrow e^{i\theta}$ , acting on  $L^2\{[-\pi, \pi], dF\}$ , i.e.*

$$U\xi = \mathcal{J}_{\hat{y}}(e\hat{f}), \quad \xi = \mathcal{J}_{\hat{y}}(\hat{f}). \tag{3.3.14}$$

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

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

**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) variables. 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  $\Lambda$  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  $\hat{w}$ , of a white noise process has the following property

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

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.



### 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*

$$\overline{\hat{y}(\Delta)} = \hat{y}(-\Delta) \tag{3.3.15}$$

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

$$E\{r(\Delta_1)s(\Delta_2)\} = 0 \tag{3.3.16}$$

for all Borel sets  $\Delta_1, \Delta_2$ .

**Proof.** Notwithstanding the fact that the  $y(t)$ 's 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  $\mathcal{J}_{\hat{y}}$  to the random variable  $\eta$ , 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  $\mathcal{J}_{\hat{y}}$  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  $\mathcal{J}_{\hat{y}} : I_{\Delta} \rightarrow \hat{y}(\Delta)$  we also have  $\mathcal{J}_{\hat{y}} : \bar{I}_{-\Delta} \rightarrow \overline{\hat{y}(\Delta)}$ , but  $\bar{I}_{-\Delta} = I_{-\Delta}$ , since the indicator is a real function and therefore (3.3.15) follows.

To prove the remaining statement first notice that  $r$  and  $s$  are both  $\sigma$ -additive real stochastic measures and that from (3.3.15) we get,

$$r(\Delta) = r(-\Delta), \quad s(\Delta) = -s(-\Delta) \tag{3.3.17}$$

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

$$E[s(\Delta_1)r(\Delta_2) - r(\Delta_1)s(\Delta_2)] = 0.$$

Combining this relation with the analogous one obtained by substituting  $-\Delta_1$  in place of  $\Delta_1$  and using (3.3.17) one gets the orthogonality relation (3.3.16). Hence  $E\{\hat{y}(\Delta_1)\overline{\hat{y}(\Delta_2)}\} = E\{r(\Delta_1 \cap \Delta_2)^2 + s(\Delta_1 \cap \Delta_2)^2\}$  and  $\Delta_1 \cap \Delta_2 = \emptyset$  yields  $E\{r(\Delta_1 \cap \Delta_2)\} = E\{s(\Delta_1 \cap \Delta_2)\} = 0$ . This shows that  $r$  and  $s$  are also orthogonal measures and concludes the proof.  $\square$

For real processes the spectral representation (3.3.5) can be written completely in terms of real quantities. From (3.3.17) one easily obtains

$$y(t) = \int_{-\pi}^{\pi} \cos \theta t \, dr(\theta) - \int_{-\pi}^{\pi} \sin \theta t \, ds(\theta), \quad t \in \mathbb{Z}.$$

### 3.4 Vector-valued processes

If we denote by  $d\hat{y}_k$ ,  $k = 1, \dots, m$  the spectral measure corresponding to the  $k$ -th component of a  $m$ -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

$$\hat{y}(\Delta) = \begin{bmatrix} \hat{y}_1(\Delta) \\ \hat{y}_2(\Delta) \\ \vdots \\ \hat{y}_m(\Delta) \end{bmatrix} \tag{3.4.1}$$

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

$$F(\Delta) := [E\{\hat{y}_k(\Delta)\overline{\hat{y}_j(\Delta)}\}]_{kj=1,\dots,m} \tag{3.4.2}$$

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

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

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

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,  $\sigma$ -additive function of  $\Delta$ , 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  $\theta \rightarrow F(\theta)$ , each defined modulo an additive arbitrary constant matrix, which are monotonic nondecreasing in the sense that  $F(\theta_2) - F(\theta_1) \geq 0$  (positive semidefinite) for  $\theta_2 \geq \theta_1$ . The vector-valued generalization provides readily the representation of the covariance matrix of the process as a ‘‘Fourier-like’’ integral of the form,

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

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

As in the scalar case we have the canonical decomposition

$$F = F_1 + F_2$$

where  $F_1$  is the 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*  $\Phi$  which is Hermitian and positive semidefinite ( $\Phi(\theta) \geq 0$ ,  $\theta \in [-\pi, \pi]$ ). For processes taking values in  $\mathbb{R}^m$ , which will be also called *real* for short, the symmetry relation (3.3.15) translates into  $F_{kj}(\Delta) = F_{jk}(-\Delta)$ ,  $k, j = 1, \dots, m$  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,

$$\Phi(e^{-i\theta})' = \Phi(e^{i\theta}). \tag{3.4.3}$$

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

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

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 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  $f g^*$ ) as the limit

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

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 [25, p. 1349], that this space is complete and hence a Hilbert space with respect to the scalar product

$$\langle f, g \rangle := \int_{-\pi}^{\pi} f(\theta) dF(\theta) g(\theta)^* \tag{3.4.4}$$

provided one agrees to identify vector functions whose difference has norm equal to zero. Functions  $f_1, f_2$  such that  $\|f_1 - f_2\| = 0$  (the norm being induced by the inner product defined above) are said to be equal  $F$ -almost everywhere. If  $F$  happens to

be singular on “large” 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,

$$\begin{aligned} \mathbb{E}\{\mathcal{J}_{\hat{y}}(f)\mathcal{J}_{\hat{y}}(g)^*\} &= \mathbb{E}\left\{\int_{-\pi}^{\pi} f(\theta) d\hat{y}(\theta)\left[\int_{-\pi}^{\pi} g(\theta) d\hat{y}(\theta)\right]^*\right\} = \\ &= \int_{-\pi}^{\pi} f(\theta)dF(\theta)g(\theta)^* = \langle f, g \rangle_{L_m^2([-\pi, \pi], dF)} \end{aligned} \quad (3.4.5)$$

where  $f$  and  $g$  are functions in  $L_m^2([-\pi, \pi], dF)$  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 Fourier transform  $\hat{y}$ . Every random functional  $\xi \in \mathbf{H}(y)$  can be written as a stochastic integral  $\mathcal{J}_{\hat{y}}(\hat{f})$  for a unique function  $\hat{f} \in L_m^2\{[-\pi, \pi], dF\}$ . In fact the map  $\mathcal{J}_{\hat{y}} : L_m^2\{[-\pi, \pi], dF\} \rightarrow \mathbf{H}(y)$  is unitary. It maps the elementary exponential functions  $[0, \dots, e_t, \dots, 0]$  ( $e_t(\theta) = e^{i\theta t}$  in the  $k$ -th place) into the random variable  $y_k(t)$  and transforms the shift operator  $U$  of the process  $y$  into the operator  $M_e$ , of multiplication by the exponential function  $e(\theta) : \theta \rightarrow e^{i\theta}$ , acting on  $L_m^2\{[-\pi, \pi], dF\}$ . In other words, the following diagram commutes,*

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

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

### 3.5 Functionals of white noise

Recall the well-known Hilbert space  $\ell_m^2 \equiv \ell_m^2(\mathbb{Z})$ , of square summable  $m$ -dimensional functions (sequences)  $f : \mathbb{Z} \rightarrow \mathbb{C}^m$ , endowed with the inner product

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

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

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

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  $\ell_m^{2+}$  and by  $\ell_m^{2-}$  respectively. They are clearly isomorphic to  $\ell_m^2(\mathbb{Z}_+)$  and to  $\ell_m^2(\mathbb{Z}_-)$ .

A vector-valued, say  $m$ -dimensional *white noise process*  $w$ , is just a stationary vector process whose random variables are pairwise uncorrelated,

$$E\{w(t)w(s)^*\} = Q\delta(t - s). \tag{3.5.1}$$

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  $AA' = Q$ . In this case 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  $\mathbf{H}(w)$ , of  $w$ .

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*

$$\eta = \sum_{s=-\infty}^{+\infty} f(-s)w(s), \quad f \in \ell_m^2 \tag{3.5.2}$$

where the function  $f$  is uniquely determined by  $\eta$ . The linear map  $\mathcal{J}_w : \ell_m^2 \rightarrow \mathbf{H}(w)$  defined by equation (3.5.2) is unitary and transforms the translation operator  $T$  in  $\ell_m^2$  into the shift  $U$  acting on random variables of  $\mathbf{H}(w)$ , namely, if  $[T^t f](s) = f(t + s)$ ; i.e.,

$$\eta(t) := U^t \eta = \sum_{s=-\infty}^{+\infty} f(t - s)w(s) = \mathcal{J}_w(T^t f) \tag{3.5.3}$$

Note that we have been abusing notations as the symbol  $\mathcal{J}_w$  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 (3.5.2) 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  $\eta$  with respect to that basis. It is very well-known that these coefficients are unique and form a square summable sequence. The last part of the statement also follows since  $U^{-t}w(s) = w(t-s)$  and

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

We shall leave the details of the generalization of this argument to the vector case to the reader.  $\square$

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  $\zeta$  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*

$$\eta = \int_{-\infty}^{+\infty} f(-s) dw(s), \quad f \in L_m^2(\mathbb{R}) \quad (3.5.4)$$

where the function  $f$  is uniquely determined by  $\eta$ . The linear map  $\mathcal{J}_w : L_m^2(\mathbb{R}) \rightarrow \mathbf{H}(dw)$  defined by equation (3.5.4) is unitary and transforms the translation operator  $T_t$  in  $L_m^2$  into the shift  $U_t$  acting on random variables of  $\mathbf{H}(dw)$ , namely, if  $[T_t f](s) = f(t+s)$ , then

$$\eta(t) := U^t \eta = \int_{-\infty}^{+\infty} f(t-s) dw(s) = \mathcal{J}_w(T_t f) \quad (3.5.5)$$

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.

## 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,*

$$\mathfrak{F} : \ell_m^2 \rightarrow L_m^2\left([-\pi, \pi], \frac{d\theta}{2\pi}\right), \quad \mathfrak{F}(f) := \sum_{t=-\infty}^{+\infty} e^{-i\theta t} f(t)$$

the sum being convergent (for all  $f \in \ell_m^2$ ) in the metric of the space  $L_m^2\left([-\pi, \pi], \frac{d\theta}{2\pi}\right)$ , is a norm preserving and onto, i.e. a unitary, map from  $\ell_m^2$  onto  $L_m^2\left([-\pi, \pi], \frac{d\theta}{2\pi}\right)$ .

**Proof.** The norm preserving property, i.e. the equality of the energy norm of the time signal  $f$  to the integral with respect to frequency of the square norm of its Fourier transform  $\hat{f}(\theta) = \mathfrak{F}(f)(\theta)$ ,

$$\sum_{t=-\infty}^{+\infty} |f(t)|^2 = \int_{-\pi}^{\pi} |\hat{f}(\theta)|^2 \frac{d\theta}{2\pi}$$

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  $\ell_m^2$ , the theorem can be proved by the same isometric extension argument which was used for the definition of the stochastic integral earlier on.  $\square$

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  $\ell_m^2$ ,

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

corresponds, in the frequency domain, to the algebraic operation of multiplication by the scalar exponential function  $e(\theta) : \theta \rightarrow e^{i\theta}$ , acting on  $L_m^2\left([-\pi, \pi], \frac{d\theta}{2\pi}\right)$ . 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 assumed to be known.

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

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

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

the integral being defined as a limit in the metric of the space  $L_m^2\left(\mathbb{I}, \frac{d\omega}{2\pi}\right)$ , is well-defined for all  $f \in L_m^2(\mathbb{R})$  and is a norm preserving and onto, i.e. a unitary, map from  $L_m^2(\mathbb{R})$  onto  $L_m^2\left(\mathbb{I}, \frac{d\omega}{2\pi}\right)$ .

The norm preserving property, i.e. the equality of the energy norm of the continuous-time signal  $f$  to the integral with respect to frequency of the square

norm of its Fourier transform  $\hat{f}(i\omega) = \mathfrak{F}(f)(i\omega)$ ,

$$\int_{-\infty}^{+\infty} |f(t)|^2 dt = \int_{-\infty}^{+\infty} |\hat{f}(i\omega)|^2 \frac{d\omega}{2\pi}$$

is also known as *Parseval's identity*.

In continuous time the *translation operator*  $T_t$ ;  $t \in \mathbb{R}$ , acting in  $L_m^2(\mathbb{R})$ , 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  $e_t : i\omega \rightarrow e^{i\omega t}$ , acting on  $L_m^2(\mathbb{I}, \frac{d\omega}{2\pi})$ . In other words  $\mathfrak{F}(T_t f) = M_{e_t} \mathfrak{F}(f)$  where  $M_{e_t}$  is the multiplication operator 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  $L_m^2(\mathbb{R})$  which, via Fourier transform, corresponds (in fact is unitarily equivalent) to the unitary group of multiplication operators by  $e^{i\omega t}$  acting in  $L_m^2(\mathbb{I}, \frac{d\omega}{2\pi})$ .

The following fundamental representation theorem relates the 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  $\mathcal{J}_w : \ell_m^2 \rightarrow \mathbf{H}(w)$  defined by equation (3.5.2) admits a factorization as the composite map*

$$\mathcal{J}_w = \mathcal{J}_{\hat{w}} \mathfrak{F} \quad (3.5.6)$$

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 (3.5.2). In other words  $\eta = \mathcal{J}_{\hat{w}}(\hat{f}) = \mathcal{J}_w(f)$  if and only if  $\hat{f} = \mathfrak{F}f$ . In fact, the two unitary representation maps  $\mathcal{J}_{\hat{w}}$  and  $\mathcal{J}_w$  are related as in the following commutative diagram

$$\begin{array}{ccc} \mathbf{H}(w) & \xrightarrow{U} & \mathbf{H}(w) \\ \mathcal{J}_{\hat{w}} \uparrow & & \uparrow \mathcal{J}_{\hat{w}} \\ L_m^2([-\pi, \pi], \frac{d\theta}{2\pi}) & \xrightarrow{M_e} & L_m^2([-\pi, \pi], \frac{d\theta}{2\pi}) \\ \mathfrak{F} \uparrow & & \uparrow \mathfrak{F} \\ \ell_m^2(\mathbb{Z}) & \xrightarrow{T} & \ell_m^2(\mathbb{Z}) \end{array}$$

**Proof.** The frequency-domain isomorphism  $\mathcal{J}_{\hat{w}}$  maps trigonometric polynomials  $p(\theta) = \sum_{-N}^M f(-k) e^{i\theta k}$  into finite linear combinations  $\eta = \sum_{-N}^M f(-k) w(k) = \mathcal{J}_w(f)$ , where  $f$  is an  $\ell^2$  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

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



for the dense linear manifold of finite support functions  $f$ . Since both maps  $J_w$  and  $J_{\hat{w}}\mathfrak{F}$  are unitary, (3.5.6) follows. The rest follows by well-know properties of the Fourier transform.  $\square$

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  $J_w : L_m^2(\mathbb{R}) \rightarrow \mathbf{H}(dw)$  defined in Corollary 3.5.2 factorizes exactly as the composite map (3.5.6) 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 (3.5.4). Hence,  $\eta = J_{\hat{w}}(\hat{f}) = J_w(f)$  if and only if  $\hat{f} = \mathfrak{F}f$ . In fact, the two representation maps  $J_{\hat{w}}$  and  $J_w$  are related as in the following commutative diagram*

$$\begin{array}{ccc}
 \mathbf{H}(dw) & \xrightarrow{U_t} & \mathbf{H}(dw) \\
 J_{\hat{w}} \uparrow & & \uparrow J_{\hat{w}} \\
 L_m^2(\mathbb{I}, \frac{d\omega}{2\pi}) & \xrightarrow{M_{e_t}} & L_m^2(\mathbb{I}, \frac{d\omega}{2\pi}) \\
 \mathfrak{F} \uparrow & & \uparrow \mathfrak{F} \\
 L_m^2(\mathbb{R}) & \xrightarrow{T_t} & L_m^2(\mathbb{R})
 \end{array}$$

### 3.6 Spectral representation of stationary increment processes

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

$$\frac{e^{-i\omega_2 t} - e^{-i\omega_1 t}}{2\pi i t} = (\mathfrak{F}^{-1}I_{[\omega_1, \omega_2]})(-t). \tag{3.6.1}$$

Since these are trivially 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,

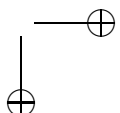
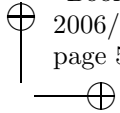
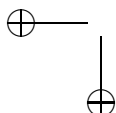
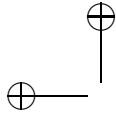
$$\hat{w}(i\omega_2) - \hat{w}(i\omega_1) = \int_{-\infty}^{\infty} \frac{e^{-i\omega_2 t} - e^{-i\omega_1 t}}{2\pi i t} dw(t) \tag{3.6.2}$$

Now, by the isometric property of the Fourier-Plancherel transform, one readily sees that the process  $\hat{w}$  has orthogonal increments. In fact,

$$E\{d\hat{w}d\hat{w}^*\} = I \frac{d\omega}{2\pi} \tag{3.6.3}$$

where  $*$  denotes transpose and conjugation. Hence,  $d\hat{w}$  is a  $p$ -dimensional Wiener process on the imaginary axis. Now, (3.6.2) 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  $f \in L^2(\mathbb{R})$

$$\int_{-\infty}^{\infty} \hat{f}(i\omega) d\hat{w} = \int_{-\infty}^{\infty} f(-t) dt, \quad (3.6.4)$$

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

$$\mathcal{J}_w \hat{f} = \int_{-\infty}^{\infty} \hat{f}(i\omega) d\hat{w}(i\omega)$$

factorizes as in (3.5.6) in Theorem 3.5.6. Dually, by choosing  $f$  to be the indicator function of the interval  $[t_1, t_2]$ , (3.6.4) yields

$$w(t_2) - w(t_1) = \int_{-\infty}^{\infty} \frac{e^{i\omega t_2} - e^{i\omega t_1}}{i\omega} d\hat{w}(i\omega) \quad (3.6.5)$$

This is a particular instance of *spectral representation of a stationary increments process*; in fact, of the stationary increments (Wiener) process  $dw$ , [23]. Note that the *spectral measure of  $dw$*  is also of the Wiener type, being precisely the orthogonal random measure  $d\hat{w}$  defined in (3.6.2).

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*

$$z(t) - z(s) = \int_{-\infty}^{+\infty} \frac{e^{i\omega t} - e^{i\omega s}}{i\omega} d\hat{z}(i\omega), \quad t, s \in \mathbb{R} \quad (3.6.6)$$

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,

$$\mathbf{E}\{d\hat{z}(i\omega)d\hat{z}(i\omega)^*\} = dZ(i\omega) \quad (3.6.7)$$

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

$$\begin{cases} dx = Axdt + Bdw \\ dz = Cxdt + Ddw \end{cases}, \quad (3.6.8)$$

where all eigenvalues of the matrix  $A$  have negative real parts. In the time domain (3.6.8) has the following solution

$$x(t) = \int_{-\infty}^t e^{A(t-\tau)} B dw \tag{3.6.9}$$

$$z(t) - z(s) = \int_s^t Cx(\tau) d\tau + D[w(t) - w(s)] \tag{3.6.10}$$

Applying (3.6.4) to the first of these equations, we obtain

$$x(t) = \int_{-\infty}^{\infty} e^{i\omega t} (i\omega I - A)^{-1} B d\hat{w} \tag{3.6.11}$$

which then inserted into (3.6.10) together with (3.6.5) yields the spectral representation

$$z(t) - z(s) = \int_{-\infty}^{\infty} \frac{e^{i\omega t} - e^{i\omega s}}{i\omega} d\hat{z}(i\omega) \tag{3.6.12}$$

where  $d\hat{z} = W(i\omega)d\hat{w}(i\omega)$ , the matrix function  $W$  being the transfer function of the system (3.6.8) given by

$$W(s) = C(sI - A)^{-1} B + D. \tag{3.6.13}$$

In this example  $dz$  has an absolutely continuous spectral distribution

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

where the spectral density  $\Phi$  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*

$$y(t) = \int_{-\infty}^{+\infty} e^{i\omega t} d\hat{z} \tag{3.6.14}$$

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

**Proof.** Let  $y(t)$  be defined as in (3.6.14), 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  $h \rightarrow 0$ .  $\square$

### 3.7 Multiplicity and the module structure of $\mathbf{H}(y)$

An important property of the Hilbert space  $\mathbf{H}(y)$  generated by a stochastic process  $y$ , is to be *finitely generated* by the shift  $U$ , in the following sense: there is a *finite* number of *generators* namely elements  $y_1, \dots, y_m \in \mathbf{H}(y)$  that are “cyclic” for the shift, i.e., having the property

$$\overline{\text{span}}\{U^t y_k \mid k = 1, \dots, m, t \in \mathbb{Z}\} = \mathbf{H}(y). \tag{3.7.1}$$

The cardinality of the smallest set of generators is called the *multiplicity* of the shift  $U$  on the Hilbert space  $\mathbf{H}(y)$ ; see [44], [34, p.105]. Here, for short, this number will be called the multiplicity of  $y$  or of  $\mathbf{H}(y)$ . In  $\mathbf{H}(y)$  we trivially have  $m$  generators,  $y_k = y_k(0)$ ,  $k = 1, \dots, m$ , and hence  $\mathbf{H}(y)$  has a finite multiplicity, less or equal to  $m$ .

The concept of multiplicity has to do with the algebraic concept of *basis* in module theory. This technical observation (which may sound a bit extraneous to the general spirit of this book), may help in understanding certain striking similarities between stochastic system theory and certain concepts which are commonly introduced in the deterministic system theoretic setting. A basic reference for this is Fuhrmann’s book [34]. 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 starting point to see this is the observation that there is a natural multiplication  $p \cdot \eta$  between trigonometric polynomials

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

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

$$p \cdot \eta := p(U) \cdot \eta = \left[ \sum_{k=k_0}^{k_1} p_k U^k \right] \eta = \int_{-\pi}^{\pi} p(e^{i\theta}) \hat{f}(e^{i\theta}) d\hat{y}(e^{i\theta}), \tag{3.7.2}$$

where  $\hat{f} \in L_m^2\{[-\pi, \pi], dF\}$  is the spectral representative of  $\eta$  with respect to the Fourier transform  $d\hat{y}$ , of the process  $y$ ; compare Theorem 3.3.3.

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, any function  $\varphi \in L^\infty[-\pi, \pi]$  is the (almost everywhere uniform) limit of sequences of trigonometric polynomials  $\{p_k\}$ . Hence we may define the product

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

for every  $\varphi$  in the ring  $L^\infty[-\pi, \pi]$ , of scalar essentially bounded functions on  $[-\pi, \pi]$ , which thereby becomes a *ring of scalars* by which we may multiply elements of  $\mathbf{H}(y)$ . It is immediate to check that, endowed with this multiplication,  $\mathbf{H}(y)$  is a (Hilbert) module unitarily isomorphic to  $L_m^2\{[-\pi, \pi], dF\}$  as a  $L^\infty[-\pi, \pi]$ -module.

It follows readily from (3.7.1) that the module  $\mathbf{H}(y)$  is in fact *free*, as it admits the  $m$  generators  $\{y_k(0), k = 1, \dots, m\}$ . These generators correspond in the isomorphism to the  $m$  unit vector functions  $\{e_k, k = 1, 2, \dots, m\}$  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 a.e..

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  $\mathbf{H}(y)$ . How do we check if the generators  $\{y_k(0), k = 1, \dots, m\}$  form a basis? In what follows we shall try to give an answer to this question.

First we would like the reader to appreciate that this question is more subtle than it may appear at first sight. Consider for example a scalar normalized white noise process with spectral measure  $d\hat{w}$  and let us define a stationary process  $y$  with spectral measure

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

where  $I_\Delta$  is the indicator of some proper Borel subset  $\Delta \subsetneq [-\pi, \pi]$  of normalized Lebesgue measure  $\frac{|\Delta|}{2\pi} \leq 1$ . It is nearly obvious that  $w(0)$  is a generator of  $\mathbf{H}(w)$  (for proving this we just need to notice that  $L^\infty[-\pi, \pi]$  is dense in  $L^2[-\pi, \pi]$ ); i.e. a basis for the Hilbert module  $\mathbf{H}(w)$ . 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 this case the answer is generally negative.

**Proposition 3.7.1.** *Unless  $\Delta$  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  $\varphi$  is nonzero almost everywhere in  $[-\pi, \pi]$ .*

**Proof.** We shall rely on a characterization of doubly invariant subspaces of  $L^2[-\pi, \pi]$  which can be found e.g. in Helson's book [45, 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  $\Delta$  has full Lebesgue measure (equivalently, is nonzero almost everywhere). Note that 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  $\varphi$ .  $\square$

We start with the following lemma.

**Lemma 3.7.2.** *Let  $y$  and  $u$  be  $m$ - and  $p$ -dimensional jointly stationary stochastic processes with spectral distribution measures  $dF_y$  and  $dF_u$ , such that  $\mathbf{H}(y) = \mathbf{H}(u)$ . Then there exist matrix functions  $M$  and  $N$  with rows  $M_k, k = 1, 2, \dots, m$  belonging to  $L^2_p\{[-\pi, \pi], dF_u\}$  and  $N_j, j = 1, 2, \dots, p$  belonging to  $L^2_m\{[-\pi, \pi], dF_y\}$ , such that,*

$$d\hat{y}(e^{i\theta}) = M(e^{i\theta})d\hat{u}(e^{i\theta}), \quad d\hat{u}(e^{i\theta}) = N(e^{i\theta})d\hat{y}(e^{i\theta}), \quad (3.7.3)$$

*which can be expressed by saying that the two vector stochastic measures  $d\hat{y}$  and  $d\hat{u}$  are equivalent. Moreover*

$$M(e^{i\theta})N(e^{i\theta}) = I_m, \quad N(e^{i\theta})M(e^{i\theta}) = I_p \quad (3.7.4)$$

$dF_y$  (and also  $dF_u$ )-almost everywhere.

**Proof.** Since all  $y_k(t)$ 's,  $k = 1, 2, \dots, m$ , belongs to  $\mathbf{H}(u)$ , there are  $m$  row vector functions  $M_k \in L_p^2\{[-\pi, \pi], dF_u\}$  such that,

$$y_k(t) = \int_{-\pi}^{+\pi} e^{i\theta t} M_k(e^{i\theta}) d\hat{u}(e^{i\theta}) \quad k = 1, 2, \dots, m$$

and, by uniqueness of spectral representation, it then follows that  $d\hat{y}_k(e^{i\theta}) = M_k(e^{i\theta})d\hat{u}(e^{i\theta})$ ,  $k = 1, 2, \dots, m$ . This proves the first relation in (3.7.3). A symmetric argument yields the second equality. Finally, (3.7.4) are obtained by substituting one of the (3.7.3) into the other, getting for example

$$d\hat{y}(e^{i\theta}) = M(e^{i\theta})N(e^{i\theta})d\hat{y}(e^{i\theta})$$

which, again by uniqueness of the spectral measure, implies the first equality in (3.7.4),  $dF_y$ -almost everywhere.  $\square$

**Theorem 3.7.3.** Assume as before that  $\mathbf{H}(y) = \mathbf{H}(u)$  and let  $dF_u$  be of the scalar type; i.e.,  $dF_u = I_p d\mu$  with  $\mu$  a positive Borel measure on  $[-\pi, \pi]$ . Then  $\{u_k(0), k = 1, \dots, p\}$  are a set of generators of  $\mathbf{H}(y)$  of smallest cardinality, i.e. a minimal set of generators. In this case

$$\text{rank } M(e^{i\theta}) = p, \quad \mu - \text{almost everywhere.} \tag{3.7.5}$$

In particular for  $p = 1$ , the  $m \times 1$  function  $M(e^{i\theta})$  must be a nonzero vector  $\mu$ -almost everywhere.

**Proof.** We have

$$E\{d\hat{u}_k(e^{i\theta})d\hat{u}_j(e^{i\theta})^*\} = dF_{u,k,j}(e^{i\theta}) = \delta_{kj}d\mu$$

hence, for  $k \neq j$ ,  $u_k(t) = U^t u_k(0)$  and  $u_j(s) = U^s u_j(0)$  are orthogonal for all  $t, s \in \mathbb{Z}$ . Clearly the module generated by any proper subset of the  $\{u_k(0), k = 1, \dots, p\}$  must then be a proper submodule of  $\mathbf{H}(u)$ . Hence the random variables  $\{u_k(0), k = 1, \dots, p\}$  are a minimal set of generators. Since the  $\{y_k(0), k = 1, \dots, m\}$  are also generators, it follows that  $m \geq p$ . Now the condition (3.7.5) follows from the second of the (3.7.4), which in this case is just saying that  $M(e^{i\theta})$  has  $\mu$ -almost everywhere a left inverse (of dimension  $p \times m$ ).  $\square$

**Remark 3.7.4.** Under the assumptions of Theorem 3.7.3 we have

$$dF_y(e^{i\theta}) = M(e^{i\theta})M(e^{i\theta})^* d\mu \tag{3.7.6}$$

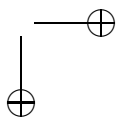
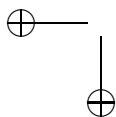
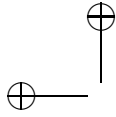
so that  $dF_y$  has a density, equal to the  $m \times m$  matrix  $M(e^{i\theta})M(e^{i\theta})^*$ , with respect to the scalar type measure  $I d\mu$ . It follows from Sylvester inequalities (see [35, p. 66]) that this density has constant rank  $p$ ,  $\mu$ -almost everywhere.

There is a notion of *rank* of a stationary process that is usually defined only for orthonormalizable processes (see Chapter 4). In this setting, the rank of  $y$  is equal to the rank almost everywhere of the spectral density matrix of the process. This notion turns out to coincide with the (more general) notion of multiplicity given above. That the two notions reduce to the same thing (in the particular context of orthonormalizable processes) is a consequence of the spectral factorization theorem which we shall prove in Chapter 4 (Theorem 4.2.1).

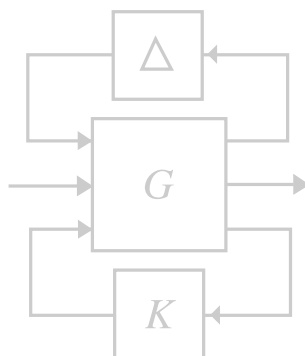
## Bibliographical notes

Most of the material in this section is classical. The spectral representation theorem is due to Cramèr, [17, 18, 19], see also the work of his former student K. Karhunen, [61, 62]. The proof given here is adapted from [38, p. 203]. A more direct approach in [106] 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 [23, p. 635-636] that the stochastic integral, first introduced by Wiener in [118], was defined in exactly the same way as the spectral integrals commonly introduced in spectral theory 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  $(\mathbf{H}(y), U)$  which are used in this book are also valid for any pair  $(\mathbf{H}, U)$  where, instead of a stationary process on a probability space, one has just a (separable) Hilbert space  $\mathbf{H}$  and a unitary operator  $U$  on  $\mathbf{H}$  with finite multiplicity.

The concept of multiplicity can be defined for more general classes of processes than stationary, see e.g. [20], [49]. The module theory of section 3.7 is inspired by Fuhrmann's book, [34] where it is introduced for self-adjoint operators; see chapter II, especially p. 101-102. The rank condition of Theorem 3.7.3 seems to be new; it explains why spectral factorization of stationary processes must be of "constant rank", a fact which may appear rather mysterious from the way it is normally introduced in the literature.







## Chapter 4

# Innovations, Wold Decomposition, and Spectral Factorization

We shall begin this chapter by reviewing some basic concepts of the theory of dynamic estimation in the classical setup of Wiener and Kolmogorov. The theory leads naturally to consider certain white noise representations of the observation process which are prototypes of stochastic dynamical systems described in input-output form. These representations have been first introduced by geometrical means in the seminal work of H. Wold's on stationary processes and prediction theory. Wold's ideas have been generalized in many directions. One such generalization will be discussed in this chapter. It will form the basis of representation theorems which will be used throughout the book. Generalizations of Wold Decomposition theory have become part of functional analysis and have led to a unifying view of certain fundamental problems in operator theory and Hardy spaces. The operator theoretic (and Hardy space) results which stems from this idea can, in a sense, be seen as function-analytic counterparts of results in stationary processes and prediction theory. In Section 4.6 we take advantage of this conceptual connection of the fields, to review, in an economical and essentially self-contained way, some basic parts of Hardy space theory that will be needed in various parts of the book.

### 4.1 The Wiener-Kolmogorov theory of filtering and prediction

From now on we shall deal with *real* (vector-valued) stationary processes only. Let  $x$  be an  $n$ -dimensional inaccessible random signal and  $y$  an  $m$ -dimensional process which is to be interpreted as an *observation* or *measurement* of  $x$ . We want to find the best (in the sense of minimum error variance) linear estimator of the random value  $x(t)$  of the signal  $x$  at some instant of time  $t$ , based on an observed sample trajectory of the random process  $y$  on a certain interval of time  $\mathbb{T}$ . Certain particular classes of linear estimation problems of this kind have been extensively studied in the past decades. In particular the *filtering problem* where one wants to compute the best linear estimate of  $x(t)$ , given the observed data  $y$  up to time  $t$ , and the  $k$ -step ahead *prediction problem*, a special filtering problem where  $x(t) = y(t + k)$ ,  $k > 0$ ,

were first formulated and studied in the 1940s by A.N. Kolmogorov and N. Wiener.

Filtering and prediction are *causal* problems where one is allowed to use only the information contained in the past observations. These problems often occur in applications to communication and control systems. In these applications the estimates need to be computed in *real-time* or “on-line”. This means that special computational schemes should allow to update easily the estimate at time  $t$ , to the new estimate which must be computed at time  $t + 1$  as the new observation  $y(t + 1)$  becomes available.

Instead, the problems of *smoothing* and *interpolation* are estimation problems where one is given a fixed observation record (a finite “chunk” of data) which may be processed off-line to generate the estimate. This class of problems is conceptually easier to solve.

### The role of Fourier transform and spectral representation

The theory of Wiener and Kolmogorov deals with stationary processes and leads to linear minimum-variance estimators which also evolve in time as stationary processes, so it deals, in a certain sense, with *steady-state estimation*. As for deterministic systems, the stationary setup is best dealt with by Fourier analysis methods and for this reason in this chapter we shall make quite an extensive use of the tools of harmonic analysis of stationary processes introduced in the previous chapter.

We stress that the stochastic Fourier transform, which is commonly called *spectral representation* of stationary processes, has the same properties and serves exactly the same purpose as the Fourier transform defined for deterministic signals and systems. In the stochastic setting one replaces the deterministic translation operator by the stochastic shift  $U$ . One has to accept the additional complication that the Fourier transform becomes a random function of frequency, but this random function of frequency is of a very special type and turns out to be easy to work with. As for deterministic signals, linear time-invariant operators on an input stationary process will in general be convolution operators, a typical example being

$$y(t) = \sum_{s=-\infty}^{+\infty} F(t-s)w(s)$$

where  $w$  is a  $p$  dimensional white noise process and the rows  $F_k$ ,  $k = 1, 2, \dots, m$  of  $F$  are square summable sequences with Fourier transforms  $\hat{F}_k$ . This can be represented in the frequency domain as multiplication of the signal Fourier transform by the transfer function  $\hat{F}$ , namely

$$d\hat{y}(e^{i\theta}) = \hat{F}(e^{i\theta})d\hat{w}(e^{i\theta})$$

These operations can be visualized by block diagrams consisting of blocks, representing transfer functions and arrows, representing (Fourier transforms of) stochastic signals. These elements can be composed according to simple algebraic composition rules (e.g. multiplications and sums). In fact, with the above convention, we can apply to the stochastic setting exactly the same rules valid for Fourier transforms of deterministic signals and transfer functions of (linear time-invariant) systems of linear system theory.

### Acausal and causal Wiener filters

There are two main assumptions on the data of the problem which guarantee stationarity of the estimator:

1. The second order processes  $x$  and  $y$  are jointly stationary. Their joint spectral distribution function is absolutely continuous with known spectral density matrix,

$$\Phi(e^{i\theta}) = \begin{bmatrix} \Phi_x(e^{i\theta}) & \Phi_{xy}(e^{i\theta}) \\ \Phi_{yx}(e^{i\theta}) & \Phi_y(e^{i\theta}) \end{bmatrix} \quad -\pi \leq \theta \leq \pi \quad (4.1.1)$$

2. The observation interval is unbounded below, i.e. the measurements have been collected since time  $t_0 = -\infty$ .

Under these assumptions, we shall consider two typical classes of estimation problems, namely

- Compute the *acausal* linear estimator,  $\hat{x}(t) = E[x(t) | \mathbf{H}(y)]$  based on observations of the whole time history of the process  $y$ ;
- Compute the *causal* linear estimator,  $\hat{x}_-(t) := E[x(t) | \mathbf{H}_t^-(y)]$  based on the past history of  $y$  up to time  $t$ . Note that, since  $\mathbf{H}_t^-(y)$  does not include the present,  $\hat{x}_-(t)$  is really a “one-step predictor” of  $x(t)$ .

Having defined the problems, we should declare from the outset that our goal in this chapter will not be to review Wiener-Kolmogorov filtering theory in much detail. We shall discuss these two problems mainly for motivating and introducing some basic concepts in the theory of stationary processes, such as the equivalence to white noise and spectral factorization.

It follows at once from Lemma 2.2.9 that, under the stated assumption, both estimators are stationary with respect to the shift  $U$  of the Hilbert space  $\mathbf{H}(x, y) := \mathbf{H}(x) \vee \mathbf{H}(y)$ , generated by the joint process  $\{x, y\}$ . Note that having started the observation process at  $t_0 = -\infty$  plays a crucial role here, for this guarantees that

$$U^s \mathbf{H}(y) = \mathbf{H}(y), \quad U^s \mathbf{H}_t^-(y) = \mathbf{H}_{t+s}^-(y)$$

and hence

$$\hat{x}(t+s) = U^s \hat{x}(t), \quad \hat{x}_-(t+s) = U^s \hat{x}_-(t).$$

Naturally the infinite observation interval assumption is done for mathematical convenience. Under suitable regularity conditions, the stationary acausal or causal problems with infinite-interval measurements can be seen to be just the (steady-state) limit solution of more realistic *smoothing*, *interpolation* or *filtering* problems with a *finite* data set, when the length of the observation interval goes to infinity.

**Lemma 4.1.1.** *Assume the observation process  $y$  is a normalized white noise process (for convenience we then denote  $y$  as  $w$ ). Then the best linear estimator of  $x(t)$  given the whole history of  $w$ , has the convolution structure*

$$\hat{x}(t) = E[x(t) | \mathbf{H}(w)] := \sum_{-\infty}^{\infty} F(t-s)w(s) \quad (4.1.2)$$

where the matrix function  $F$  is given by

$$F(t) = \Lambda_{xw}(t), \quad t \in \mathbb{Z} \quad (4.1.3)$$

where  $\Lambda_{xw}$  is the cross covariance matrix of the processes  $x$  and  $w$ .

**Proof.** That the estimator has the convolution structure of formula (4.1.2) above follows from stationarity and from Theorem 3.5.1. The orthogonality condition gives

$$\mathbb{E}\left\{\left[x(t) - \sum_{-\infty}^{\infty} F(t-s)w(s)\right]w(\tau)'\right\} = 0 \quad \tau \in \mathbb{Z}$$

which can be written as

$$\Lambda_{xw}(t-\tau) = \sum_{-\infty}^{\infty} F(t-s)I\delta(s-\tau), \quad \tau \in \mathbb{Z}$$

and, after a change of variables, immediately yields (4.1.3). Note that the  $k$ -th row of  $\Lambda_{xw}$  is the sequence of the Fourier coefficients of  $x_k(t)$  with respect to the orthonormal sequence  $\{w(s); -\infty < s < +\infty\}$ , i.e.

$$\Lambda_{x_k w}(t-s) = \mathbb{E}\{x_k(s+t)w(s)'\}, \quad t \in \mathbb{Z}$$

It follows that  $\Lambda_{x_k w}$  is square-summable.  $\square$

From this lemma one sees that acausal Wiener filtering is a rather trivial problem if the observation process is white. In essence, the problem is then converted into the problem of transforming  $y$  into white noise. We shall see that this operation, although not always possible, is feasible for a wide class of stationary processes. We shall call a stationary process  $y$  *orthonormalizable*, when it is possible to find a normalized vector white noise process  $w$ , jointly stationary with  $y$ , such that

$$\mathbf{H}(y) = \mathbf{H}(w). \quad (4.1.4)$$

where  $w$  (in general also vector valued) is of, say, dimension  $p$ . Note that, if this is possible, then we have a representation formula for the elements of  $\mathbf{H}(y)$  as random linear functionals of  $w$  (see Theorem 3.5.1). In particular, if (4.1.4) holds, the random variables  $y_k(0)$ ,  $k = 1, \dots, m$  have a representation,

$$y_k(0) = \sum_{s=-\infty}^{+\infty} \check{W}_k(-s)w(s) \quad k = 1, \dots, m \quad (4.1.5)$$

where  $\check{W}_k$ ,  $k = 1, \dots, m$ , are square summable row vector functions in  $\ell_p^2$ . Recall that the shift  $U$  acting on the random variables of  $\mathbf{H}(w)$  corresponds to the translation operator  $T$  acting on representative functions (Theorem 3.5.1), hence, by letting  $\eta = y_k(0)$  in formula (3.5.3), one obtains corresponding representations for

the family  $y_k(t) = U^t y_k(0)$ ,  $t \in \mathbb{Z}$ . Using vector notations and denoting by  $\check{W}$  the  $m \times p$  matrix function with rows  $\check{W}_k$ ,  $k = 1, \dots, m$ , one readily obtains

$$y(t) = \sum_{s=-\infty}^{+\infty} \check{W}(t-s) w(s) \tag{4.1.6}$$

which can be seen as a representation of the process  $y$  as the *output of a linear time-invariant filter driven by white noise*. This is also called a *shaping filter* representation of  $y$ , in the engineering literature. Note that this filter has a square-summable impulse response matrix  $\check{W}$ .

As we shall see, transforming an orthonormalizable process into white noise requires the solution of a *spectral factorization* problem.

Anticipating a little what will be seen in the next section of this chapter, the acausal Wiener filter which computes the estimate  $\hat{x}(t)$  starting from an orthonormalizable observation process  $y$ , can be decomposed in the cascade of two operators as schematically shown in the block diagram of Fig. 4.1.1 below,

1. A *whitening filter* which performs the orthonormalization of  $y$ ; i.e., a linear time-invariant operator which transforms  $y$  into a normalized white noise process  $w$ . Determining the transfer function of this filter requires the computation of a (full-rank) spectral factor  $W$  of the spectral density matrix,  $\Phi_y$ , of  $y$ . The Fourier transform of the noise process  $w$  is obtained by a simple normalization of the Fourier transform of  $y$ , using a left-inverse of  $W$ . See (4.2.7) below.
2. A linear filter (estimator) which operates on the whitened process  $w$ . This linear filter can be realized in the time-domain by a convolution operation, as described in Lemma 4.1.1. Both operations are in general non-causal.

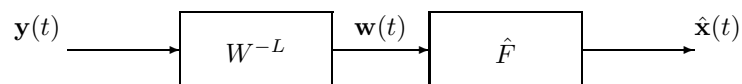


Figure 4.1.1. Cascade structure of the Wiener filter.

### Causal Wiener filtering

The *causal* Wiener filtering problem consists in computing the best estimate of  $x(t)$  given a past trajectory at time  $t$  of the observations  $y$  and hence involves the computation of the orthogonal projection  $\hat{x}_-(t) = E[x(t) | \mathbf{H}_t^-(y)]$  onto the past space  $\mathbf{H}_t^-(y)$ . To this end, we shall follow a procedure very close to that used to solve the acausal problem. The key idea is still whitening, but now (4.1.4) must be substituted by a *causal equivalence condition*. It is now necessary to find a

normalized white noise process (jointly stationary with  $y$ ) for which

$$\mathbf{H}_t^-(w) = \mathbf{H}_t^-(y), \quad t \in \mathbb{Z}. \tag{4.1.7}$$

A process  $w$ , jointly stationary with  $y$ , for which (4.1.7) holds, is said to be *causally equivalent* to  $y$ .

**Definition 4.1.2.** *A process  $y$  admitting a causally equivalent white noise process is called a forward purely nondeterministic processes.*<sup>10</sup>

In the following we shall abbreviate “purely nondeterministic” to *p.n.d.*. Clearly, being p.n.d. is a stronger condition than just being orthonormalizable.

In Sections 4.5 and 4.6 we shall study the characterization of p.n.d. processes and shall see that, under certain regularity conditions on the spectral density matrix, there is a normalized white noise process which is causally equivalent to  $y$  and that this white noise process, denoted  $w_-$ , is essentially *unique*. Following Wiener and Masani [119], it will be called the *forward innovation* process of  $y$ .

The following specialization of Theorem 3.5.1, describes the structure of causal functionals of white noise.

**Lemma 4.1.3.** *Let  $w$  be a  $p$ -dimensional normalized white noise process. All linear functionals of the past and present of  $w$  up to time  $t = 0$  (included); i.e. all random variables  $\eta \in \mathbf{H}_1^-(w)$ , admit a representation  $\eta = \mathcal{J}_w(f)$  of the type (3.5.2), where the function  $f$  is causal, i.e. belongs to  $\ell_p^{2+}$ . In fact, the linear map  $\mathcal{J}_w$  maps  $\ell_p^{2+}$  unitarily onto  $\mathbf{H}_1^-(w)$ .*

*Equivalently, all random functionals  $\eta(t)$  in  $\mathbf{H}_{t+1}^-(w)$  admit a causal convolution representation of the type*

$$\eta(t) = \sum_{s=-\infty}^t f(t-s)w(s) \tag{4.1.8}$$

for a unique  $f \in \ell_p^{2+}$ .

**Proof.** The result is an immediate corollary of Theorem 3.5.1. For,  $\eta = \mathcal{J}_w(f)$  is in  $\mathbf{H}_1^-(w)$  if and only if  $\eta \perp \{w(t), t > 0\}$ , and this in turn is true if and only if the function  $f$  is causal, since  $f(-t) = \langle \eta, w(t) \rangle = 0$  for all  $t > 0$ . The last statement follows since all random functionals  $\eta(t) \in \mathbf{H}_{t+1}^-(w)$  are shifts of elements  $\eta \in \mathbf{H}_1^-(w)$ , i.e.  $\eta(t) := U^t \eta = \sum_{s=-\infty}^{+\infty} f(t-s)w(s)$ .  $\square$

The causal analog of Lemma 4.1.1 is as follows.

**Lemma 4.1.4.** *Assume the observation process is white normalized (and for convenience denote  $y$  by  $w$ ). Then the matrix function  $F$  defining the best linear causal*

<sup>10</sup>This should be called *causally orthonormalizable* at this point. The reason for using this new terminology will become clear later on. In the Russian literature, purely nondeterministic processes are called *linearly regular*.

estimator of  $x(t + 1)$  given the past history of  $w$  up to time  $t$  (included),

$$\hat{x}_-(t + 1) = \mathbf{E}[x(t + 1) | \mathbf{H}_{t+1}^-(w)] = \sum_{s=-\infty}^{\infty} F(t - s)w(s) \quad (4.1.9)$$

is given by

$$F(t) = \begin{cases} \Lambda_{xw}(t), & t \geq 0 \\ 0, & t < 0 \end{cases} \quad (4.1.10)$$

where  $\Lambda_{xw}$  is the cross covariance matrix of the processes  $x$  and  $w$ .

**Proof.** The estimator has the same convolution structure of formula (4.1.2) since the components of  $\hat{x}_-(t + 1)$  are in  $\mathbf{H}(w)$ . The orthogonality condition provides

$$\mathbf{E}\{[x(t) - \sum_{s=-\infty}^{\infty} F(t - s)w(s)]w(\tau)'\} = 0 \quad \tau \leq t$$

which can be written as

$$\Lambda_{xw}(t - \tau) = \sum_{s=-\infty}^{\infty} F(t - s)I\delta(s - \tau), \quad \tau \leq t$$

and, after a change of variables, yields

$$\Lambda_{xw}(t) = F(t), \quad t \geq 0$$

which identifies the matrix weighting function of the filter,  $F$ , on the positive time axis  $\{t \geq 0\}$ . Note however that each row of  $F$  corresponds to a component of  $\hat{x}_-(t + 1)$  which is actually a causal random functional in  $\mathbf{H}_{t+1}^-(w)$ . Hence by Lemma 4.1.3,  $F_k(t) = 0$  for  $t < 0$ ,  $k = 1, \dots, n$ , and (4.1.10) is proven.  $\square$

The causal filtering problem for a general nonwhite observation process  $y$  can be solved under the assumption that  $y$  is forward p.n.d.. In this case, assuming  $w_-$  satisfies (4.1.7),  $y$  admits a convolution representation of the type

$$y(t) = \sum_{s=-\infty}^t \check{W}_-(t - s)w_-(s) \quad (4.1.11)$$

where the  $m \times p$  matrix function  $\check{W}_-$  has rows in  $\ell_p^{2+}$ . Then, exactly as found for the acausal problem, the causal “Wiener filter”, which computes the estimate  $\hat{x}_-(t)$  from the past of the observation process  $y$ , can be decomposed in the cascade of two operations:

1. A *whitening filter* which does the orthonormalization of  $y$  to produce the innovation process  $w_-$ . This is a special normalized white noise process causally equivalent to  $y$ . The whitening filter in certain cases can be realized as a convolution operator. It requires the computation of the special full-rank spectral factor  $W_-$  of the spectral density  $\Phi_y$ , which is associated to  $w_-$ . The special properties of  $W_-$  will be elucidated in Section 4.2.

2. A causal linear estimator which operates on the “whitened” process  $w_-$ . This linear filter can be realized in the time-domain by a causal convolution operation with a matrix function  $F$  whose rows belong to  $\ell_p^{2+}$ , as described in Lemma 4.1.4 above.

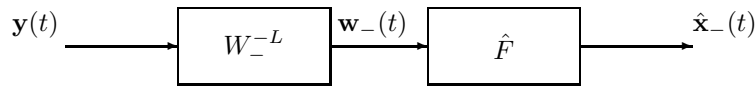


Figure 4.1.2. Cascade structure of the causal Wiener filter.

## 4.2 Orthonormalizable processes and spectral factorization

Following the program sketched in the previous section, we shall now study the problem of finding a normalized vector white noise process  $w$ , jointly stationary with  $y$ , such that (4.1.4) holds. The basic result about orthonormalizable processes is in the following theorem.

**Theorem 4.2.1 (Spectral Factorization).** *A stationary  $m$ -dimensional process is orthonormalizable; i.e. there is a  $p$ -dimensional normalized white noise process  $w$ , jointly stationary with  $y$ , satisfying (4.1.4), if and only if,*

- its spectral distribution function is absolutely continuous and the relative spectral density matrix  $\Phi$  has constant rank  $p$  almost everywhere on  $[-\pi, \pi]$ ,
- there are  $m \times p$  matrix functions  $W$  satisfying the spectral factorization equation,

$$\Phi(e^{i\theta}) = W(e^{i\theta})W(e^{i\theta})^* \tag{4.2.1}$$

almost everywhere on  $[-\pi, \pi]$ .

**Definition 4.2.2.** *A matrix function  $W$  satisfying (4.2.1) is called a Spectral Factor of  $\Phi$ . A spectral density matrix satisfying the two conditions of the theorem is said to be factorizable.*

**Remark 4.2.3.** It follows from Sylvester’s inequality (Section A.3 in the appendix) that  $\text{rank } \Phi(e^{i\theta}) = \text{rank } W(e^{i\theta})$  almost everywhere. Hence the number of columns ( $p$ ) of the spectral factors mentioned in the theorem is equal to their rank a.e.. In other words, these  $W$ ’s have a.e. linearly independent columns<sup>11</sup>. They will be called *full rank spectral factors*. Of course full rank spectral factors are rather

<sup>11</sup>A matrix with this property is commonly said to be of *full column rank*.



special as not all spectral factors need to have full (or even constant) rank; for take any measurable  $p \times r$  ( $r \geq p$ ) matrix function  $Q(e^{i\theta})$  on  $[-\pi, \pi]$ , obeying the condition,

$$Q(e^{i\theta})Q(e^{i\theta})^* = I_d, \quad \text{a.e.}, \quad (4.2.2)$$

where  $I_p$  is the  $p \times p$  identity matrix. It is immediate to verify that for any  $m \times p$   $W$  satisfying (4.2.1),  $W(e^{i\theta})Q(e^{i\theta})$  is also a  $(m \times r)$  spectral factor, generally not of full rank.

The necessity proof of Theorem 4.2.1 will be given below in form of a lemma. The proof of sufficiency will result from a series of constructive steps which we shall discuss after the lemma.

**Lemma 4.2.4.** *Assume there is a  $p$ -dimensional normalized white noise process  $w$ , jointly stationary with  $y$ , satisfying (4.1.4). Then the spectral distribution of  $y$  is absolutely continuous with a factorizable spectral density matrix  $\Phi$  of rank  $p$  a.e. and there is a full rank spectral factor  $W$  such that  $d\dot{y} = Wd\dot{w}$ . In other words, any normalized white noise process  $w$  satisfying the equality (4.1.4) must come from a full-rank spectral factor; in particular, all such  $w$ 's must have the same dimension  $p$ .*

**Proof.** The lemma is an immediate corollary of Theorem 3.7.3 and of Remark 3.7.4 once we take  $\mu =$  Lebesgue measure.  $\square$

The rank (a.e.) of  $\Phi$  is commonly called the *rank of the process  $y$*  and a process for which  $p = m$  is called a *full rank process*. One sees, as anticipated in Remark 3.7.4 of the previous chapter, that the rank of a stationary orthonormalizable process coincides with its multiplicity.

**Remark 4.2.5.** If, instead of (4.1.4), we merely assume that  $\mathbf{H}(y) \subseteq \mathbf{H}(w)$  (a doubly invariant subspace), where  $w$  is a  $r$ -dimensional normalized white noise process, then it is easy to see that  $dF_y$  is still absolutely continuous and the spectral density admits factorizations of the form (4.2.1).

In fact, if  $\mathbf{H}(y) \subseteq \mathbf{H}(w)$ , then  $y$  admits a representation of the type (4.1.6) with respect to  $w$ . Denoting by  $W_k$  the Fourier transform of  $\check{W}_k \in \ell_r^2$ , by the isomorphism theorem 3.5.5, for random functionals of a white noise, we obtain

$$\begin{aligned} \Lambda_{kj}(\tau) &= \mathbf{E}\{y_k(\tau)\overline{y_j(0)}\} = \langle \mathcal{J}_w(\check{W}_k(\tau + \cdot)), \mathcal{J}_w(\check{W}_j) \rangle = \\ &= \langle \mathcal{J}_{\dot{w}}(e^{i\theta\tau}W_k), \mathcal{J}_{\dot{w}}(W_j) \rangle = \langle e^{i\theta\tau}W_k, W_j \rangle_{L^2_{\mathbb{C}}([- \pi, \pi], \frac{d\theta}{2\pi})} = \\ &= \int_{-\pi}^{\pi} e^{i\theta\tau}W_k(e^{i\theta})W_j(e^{i\theta})^* \frac{d\theta}{2\pi} \quad (4.2.3) \end{aligned}$$

From the uniqueness of the spectral measure in Herglotz's Theorem, it is clear that the spectral distribution function (spectral matrix measure) of  $y$  must be absolutely continuous with a density of the form  $\Phi(e^{i\theta}) = W(e^{i\theta})W(e^{i\theta})^*$ . In this case however we can no longer in general say that the rank of  $W$  is constant a.e..

Assume now that  $y$  has an absolutely continuous spectrum and that the density is factorizable as in (4.2.1). We shall pick a full rank spectral factor  $W$  so that  $W(e^{i\theta})$  has (almost everywhere) a  $p \times m$  left inverse<sup>12</sup>  $W^{-L}(e^{i\theta})$ , whereby

$$W^{-L}(e^{i\theta})W(e^{i\theta}) = I_p \quad \theta \in [-\pi, \pi] \tag{4.2.4}$$

(a.e.).

The basic idea in the construction which follows is to “normalize” the stochastic measure  $d\hat{y}$  so as to make it into the spectral measure of a white noise. Formally, for each Borel set  $\Delta \subset [-\pi, \pi]$ , define a  $p$ -dimensional random vector

$$\hat{w}(\Delta) := \int_{\Delta} W^{-L}(e^{i\theta}) d\hat{y}(e^{i\theta}). \tag{4.2.5}$$

The integral is well-defined since the rows of  $W^{-L}$  are square integrable functions with respect to the matrix measure  $dF = \Phi \frac{d\theta}{2\pi}$  of the process  $y$ . In fact, from (4.2.4), (4.2.1) we get,

$$\begin{aligned} E\{\hat{w}(\Delta_1)\hat{w}(\Delta_2)^*\} &= \int_{\Delta_1 \cap \Delta_2} W^{-L}(e^{i\theta})\Phi(e^{i\theta})(W^{-L}(e^{i\theta}))^* \frac{d\theta}{2\pi} = \\ &= \int_{\Delta_1 \cap \Delta_2} I_p \frac{d\theta}{2\pi} = \frac{I_p}{2\pi} |\Delta_1 \cap \Delta_2| \end{aligned} \tag{4.2.6}$$

where  $\frac{1}{2\pi}|\Delta|$  is the normalized Lebesgue measure of the set  $\Delta \subset [-\pi, \pi]$ . We see that  $\hat{w}$  defined in (4.2.5) is not only a bona-fide spectral measure but is, indeed, the spectral measure of a  $p$ -dimensional white noise process  $w$ . Equation (4.2.5) is rewritten symbolically as

$$d\hat{w}(e^{i\theta}) := W^{-L}(e^{i\theta}) d\hat{y}(e^{i\theta}). \tag{4.2.7}$$

If  $\Phi$  has rank  $m$ ,  $W$  is square and hence has a unique left-inverse equal to  $W^{-1}$ . In this case the formula above can be rewritten in the form

$$d\hat{y}(e^{i\theta}) = W(e^{i\theta})d\hat{w}(e^{i\theta})$$

and from this one immediately obtains the following spectral representation of  $y(t)$

$$y(t) = \int_{-\pi}^{\pi} e^{i\theta t} W(e^{i\theta})d\hat{w}(e^{i\theta}). \tag{4.2.8}$$

This is the frequency domain version of (4.1.6). In fact, by Theorem 3.5.5,  $W$  is now just the *Fourier transform* of the time-domain impulse response function of the filter (denoted  $W$  in (4.1.6)).

If  $\Phi$  has (a.e.) rank  $p < m$  the same result holds, but we need to worry about some technicalities having to do with the non-uniqueness of the left-inverse.

<sup>12</sup>Note that the left-inverse is in general non-unique.

**Lemma 4.2.6.** *Let  $W$  be a full-rank spectral factor. No matter how the left-inverse  $W^{-L}$  is chosen, one has*

$$d\hat{y}(e^{i\theta}) = W(e^{i\theta})W^{-L}(e^{i\theta})d\hat{y}(e^{i\theta}) = W(e^{i\theta})d\hat{w}(e^{i\theta})$$

*with probability one, where  $d\hat{w}$  is the corresponding stochastic measure. In fact, for any other, not necessarily full-rank, spectral factor  $\hat{G}$  it holds that*

$$\hat{G}(e^{i\theta}) = W(e^{i\theta})W^{-L}(e^{i\theta})\hat{G}(e^{i\theta})$$

*almost everywhere.*

**Proof.** We shall show that the matrix functions  $I_m$  and  $WW^{-L}$  are equal  $F$ -almost everywhere, where  $dF = \Phi d\theta/2\pi$  is the spectral distribution of the process  $y$ . For, from  $W^{-L}(e^{i\theta})W(e^{i\theta}) = I_p$  it follows that

$$(I - WW^{-L})\Phi(I - WW^{-L})^* = (I - WW^{-L})WW^*(I - WW^{-L})^* = 0$$

almost everywhere on  $[-\pi, \pi]$ . Clearly we may instead substitute  $\hat{G}\hat{G}^*$  in place of  $\Phi$  in the identity above, without changing the conclusion. This implies that  $(I - WW^{-L})\hat{G}$  must be zero a.e. yielding the second relation of the lemma.  $\square$

In conclusion, corresponding to any full rank spectral factor  $W$  there is a white noise process  $w$  which yields a shaping filter representation of  $y$  of the type (4.1.6) (or equivalently (4.2.8)). This white noise process  $w$  is generated by “whitening” the process  $y$  as specified by (4.2.5) or more explicitly

$$w(t) = \int_{-\pi}^{\pi} e^{i\theta t} W^{-L}(e^{i\theta}) d\hat{y}(e^{i\theta}) \tag{4.2.9}$$

It is clear from this representation that  $w$  is jointly stationary with  $y$  and  $w(t) \in \mathbf{H}(y)$  so that  $\mathbf{H}(w) \subset \mathbf{H}(y)$ . On the other hand, from (4.2.8) it follows that  $y(t) \in \mathbf{H}(w)$  and the converse inclusion also holds. Therefore  $y$  is orthonormalizable. This concludes the proof of Theorem 4.2.1.  $\square$

One last observation is in order. As we have just seen, to each white noise process  $w$  which generates the same Hilbert space as  $y$  we have been able to associate a full rank spectral factor (Lemma 4.2.4) and, conversely, to each full rank spectral factor  $W$  we have been able to attach a white noise process  $w$  which generates the same Hilbert space as  $y$ . In the following we shall examine more closely the question of uniqueness in the correspondence  $w \leftrightarrow W$ .

Let us denote by  $\mathcal{O}(p)$  the orthogonal group of  $p \times p$  orthogonal matrices (i.e.  $T \in \mathcal{O}(p)$  if and only if  $TT' = T'T = I_p$ ). Let  $w_1$  and  $w_2$  be two  $p$ -dimensional normalized white noise processes defined in the same probability space. We shall say that  $w_1$  and  $w_2$  are *equal modulo*  $\mathcal{O}(p)$ , or simply *mod*  $\mathcal{O}$  if there is an orthogonal matrix  $T \in \mathcal{O}(p)$  such that  $w_1(t) = Tw_2(t)$  for all  $t \in \mathbb{Z}$ . Obviously two normalized white noises equal mod  $\mathcal{O}$  have the same covariance function and generate the same family of subspaces, namely  $\mathbf{H}_t^\pm(w_1) = \mathbf{H}_t^\pm(w_2)$  (in particular  $\mathbf{H}(w_1) = \mathbf{H}(w_2)$ ) and hence cannot be distinguished on the basis of second order statistics nor on the

basis of linear functionals. For example, in the scalar case one cannot distinguish between a stationary normalized white noise  $w$  and its opposite,  $-w$ . For this reason, hereafter it will be convenient not to distinguish between normalized white processes equal modulo  $\mathcal{O}$  and we shall consider them as being the *same process*.

**Definition 4.2.7.** Let  $W_1(e^{i\theta})$  and  $W_2(e^{i\theta})$  be two  $m \times p$  matrix functions defined on  $[-\pi, \pi]$ . We shall say that  $W_1$  and  $W_2$  are equal modulo  $\mathcal{O}$ , if there is an orthogonal matrix  $T \in \mathcal{O}$  such that  $W_1(e^{i\theta}) = W_2(e^{i\theta})T$  for almost all  $\theta \in [-\pi, \pi]$ .

Note that if  $W_1$  is a full rank spectral factor of the spectral density  $\Phi$ , then all  $W$ 's which are equal to  $W_1 \bmod \mathcal{O}$  are also full rank spectral factors. Obviously however  $W_1(e^{i\theta})$  and  $W_2(e^{i\theta})$  being both full rank spectral factors does not necessarily imply their equality mod  $\mathcal{O}$ .

**Proposition 4.2.8.** To any normalized white noise processes  $w$  jointly stationary with  $y$ , satisfying (4.1.4), there corresponds a unique equivalence class mod  $\mathcal{O}$ , of full-rank spectral factors  $W$  for which the representation (4.2.8) holds. Hence there is a one-to-one correspondence between normalized white noise processes generating  $\mathbf{H}(y)$  and equivalence classes of full rank spectral factors equal modulo  $\mathcal{O}$ .

**Proof.** In the spectral domain, the equivalence class of full-rank  $W$ 's corresponding to a fixed  $p$  dimensional white noise  $d\hat{w}$  is defined by the equality  $d\hat{y} = Wd\hat{w}$ . But for any  $T \in \mathcal{O}$  we have  $Wd\hat{w} = WTT'd\hat{w}$  and  $T'd\hat{w}$  is the same as  $d\hat{w}$ .  $\square$

### 4.3 Hardy spaces

In Chapter 3, the spectral representatives of random variables in  $\mathbf{H}(w)$  are characterized as Fourier transforms, i.e. as (possibly vector or matrix-valued) functions in  $L^2([-\pi, \pi], d\theta/2\pi)$ . In certain cases these functions admit an extension as analytic functions on the exterior (or on the interior) of the unit circle. The extension could formally be accomplished by substituting a complex variable  $z$  for  $e^{i\theta}$  in the Fourier series expansion

$$\hat{f}(\theta) = \sum_{t=-\infty}^{+\infty} f(t)e^{-i\theta t} \Rightarrow F(z) = \sum_{t=-\infty}^{+\infty} f(t)z^{-t}$$

so that the second member assumes the aspect of a Laurent expansion of an “analytic function”  $F(z)$  (this is called the (double-sided)  $Z$ -transform of the signal  $f$  in the engineering literature).

However, even if the coefficients  $\{f(t)\}$  form a square summable sequence, the symbol  $F(z)$  may very well be meaningless, as the Laurent series in the second member is in general guaranteed to converge (in the  $L^2$ -sense!) only on the unit circle and may not converge pointwise for any value of  $z$ . As we shall see below, it is the property of causality which brings in the analytic structure. To make this

statement precise, we shall need to recall some facts from the theory of analytic functions.

**Definition 4.3.1.** *The space of  $p$ -dimensional vector-valued functions  $F(z)$  which are analytic in the region  $\{|z| > 1\}$  of the complex plane and whose coefficients  $f(k)$  in the Laurent expansion about infinity are square-summable, i.e. of functions such that,*

$$F(z) = \sum_{k=0}^{+\infty} f(k)z^{-k}, \quad |z| > 1, \quad \sum_{k=0}^{+\infty} |f(k)|^2 < \infty \quad (4.3.1)$$

is called the  $p$ -dimensional vectorial Hardy space of the unit disk, denoted  $H_p^2(\mathbb{D})$  or  $H_p^2$  for short, when there is no danger of confusion. The functions in this space will be called analytic.

The space of  $p$ -dimensional vector-valued functions  $F(z)$  which are analytic in the region  $\{|z| < 1\}$  of the complex plane and whose coefficients  $f(k)$  in the Taylor expansion about zero are square-summable, i.e. of functions such that,

$$F(z) = \sum_{k=0}^{+\infty} f(k)z^k, \quad |z| < 1, \quad \sum_{k=0}^{+\infty} |f(k)|^2 < \infty \quad (4.3.2)$$

is called the  $p$ -dimensional vectorial conjugate Hardy space of the unit disk, denoted  $\overline{H}_p^2(\mathbb{D})$  or  $\overline{H}_p^2$  for short, when there is no danger of confusion. The functions in this space will be called conjugate-analytic or co-analytic for short.

Functions in the spaces  $H_p^2$  and  $\overline{H}_p^2$  are called *real* if their Laurent coefficients are  $\mathbb{R}^p$ -valued sequences.

Note that  $H_p^2$  and  $\overline{H}_p^2$  are linear vector spaces. One can introduce a norm in  $H_p^2$  in the following way. Let  $z = \rho e^{i\theta}$  and consider, for  $\rho > 1$ , the  $L^2$ -norm of the functions  $F_\rho : \theta \mapsto F(\rho e^{i\theta})$ . We have,

$$\begin{aligned} \|F_\rho\|^2 &:= \int_{-\pi}^{\pi} |F(\rho e^{i\theta})|^2 d\theta / 2\pi = \int_{-\pi}^{\pi} F(\rho e^{i\theta}) F(\rho e^{i\theta})^* d\theta / 2\pi = \\ &= \sum_{k=0}^{+\infty} |f(k)|^2 \rho^{-2k} \end{aligned} \quad (4.3.3)$$

and the first member is clearly a monotonically non-increasing function of  $\rho$ , bounded above by the squared  $\ell^2$ -norm of the coefficient sequence  $f = \{f(k)\}$ . It follows that

$$\|F\|^2 := \lim_{\rho \downarrow 1} \|F_\rho\|^2$$

exists and it is easy to see that the limit is actually equal to  $\|f\|_{\ell_p^2}^2$ . Using this definition one can endow  $H_p^2$  with an inner product, which makes this space isometric to the space of Laurent coefficients  $f = \{f(k)\}$  of the relative functions, in fact to the space  $\ell_p^{2+}$  of causal square-summable sequences. In this way  $H_p^2$  is

given a Hilbert space structure and the map associating  $F$  to its Laurent coefficients,  $F \mapsto f$ , is an isometry of  $H_p^2$  onto  $\ell_p^{2+}$ .

Note that a completely symmetric argument works for the conjugate space  $\overline{H_p^2}$  and the anticausal sequence space  $\ell_p^{2-}$ .

Hardy functions have remarkable analytic properties and have been extensively studied. Their importance in many applied fields, especially in systems and control engineering, comes from a fundamental theorem due to Paley and Wiener which relates them to Fourier transforms of causal functions. The following is a discrete-time version of the Paley-Wiener Theorem <sup>13</sup>.

**Theorem 4.3.2.** *Let  $F \in H_p^2$  have Laurent coefficients  $f = \{f(k)\}$  and let  $\hat{f}$  be the Fourier transform of  $f$ . Then  $\hat{f}$  is the boundary value of  $F$  on the unit circle, in the sense that,*

$$\lim_{\rho \downarrow 1} F_\rho = \hat{f} \tag{4.3.4}$$

both in  $L_p^2([-\pi, \pi], d\theta/2\pi)$  and pointwise almost everywhere in  $\theta$ . In fact,  $F(z) \rightarrow \hat{f}(\theta)$  almost everywhere, when  $z \rightarrow e^{i\theta}$  along any nontangential curve in  $\{|z| > 1\}$ . Conversely, the Fourier transform  $\hat{f}$  of a causal sequence in  $\ell^2$  can be extended by the Cauchy integral formula to an analytic function in  $H_p^2$ , preserving its norm, i.e.  $\|F\| = \|\hat{f}\|$ . The correspondence  $F \leftrightarrow \hat{f}$  is thus unitary. In this sense the two function spaces can be identified and one may write,

$$\mathfrak{F}(\ell_p^{2+}) = H_p^2(\mathbb{D}). \tag{4.3.5}$$

Symmetrically, every function  $G$  in  $\overline{H_p^2}$  admits as boundary value on the unit circle (both in  $L^2$  and almost everywhere along non-tangential paths internal to the unit disc) the Fourier transform  $\hat{g}$  of its Taylor coefficients  $g = \{g(-k)\} \in \ell_p^{2-}$ . The function  $G$  is uniquely determined in  $\{|z| < 1\}$  by its boundary values  $\hat{g}$  and the correspondence is norm-preserving, so that one can make the identification

$$\mathfrak{F}(\ell_p^{2-}) = \overline{H_p^2}(\mathbb{D}). \tag{4.3.6}$$

The proof of this theorem can be found in [51, p. 131] or in [34, p. 172].

Since the Fourier transform of a causal  $\ell^2$  signal can be identified with its analytic extensions of class  $H^2$  (and symmetrically for anticausal sequences), it is common practice to treat the two concepts (i.e the Fourier transform and the analytic extension) as if they were the same thing. In this spirit the space  $H_p^2$  is often identified with the subspace of  $L_p^2([-\pi, \pi], d\theta/2\pi)$  of functions whose Fourier coefficients of negative index vanish, i.e.

$$H_p^2 = \{ \hat{f} \in L_p^2([-\pi, \pi], d\theta/2\pi) \mid \int_{-\pi}^{\pi} \hat{f}(e^{i\theta}) e^{i\theta t} d\theta/2\pi = 0, t < 0 \}.$$

<sup>13</sup>To be sure, this is just a corollary of the original Paley-Wiener Theorem, a result of much wider scope than what interests us here.

Of course the conjugate space  $\overline{H_p^2}$  can be described in an analogous way.

*Notation:* In the following it will be more natural to think about the Fourier transforms of  $\ell^2$  signals as functions defined on the unit circle of the complex plane and denote them by  $\hat{f}(e^{i\theta})$ . The analytic extension (whenever existing) is then the function  $\hat{f}(z)$ ,  $z \in \mathbb{C}$ .

We denote by  $H_{m \times p}^\infty(\mathbb{D})$  (shorthand:  $H_{m \times p}^\infty$ ) the space of matrix functions which are uniformly bounded and analytic in  $\{|z| > 1\}$  and by  $\overline{H}_{m \times p}^\infty(\mathbb{D})$  (shorthand:  $\overline{H}_{m \times p}^\infty$ ) the space of matrix functions which are uniformly bounded and analytic in  $\{|z| < 1\}$ . Since the  $L^2$ -norm of a vector function which is bounded on  $[-\pi, \pi]$  is obviously finite, the rows of a function in  $H_{m \times p}^\infty$  belongs in particular to  $H_p^2$ . Hence any  $F \in H_{m \times p}^\infty$  (or, respectively in the conjugate space) admits limits almost everywhere as  $z \rightarrow e^{i\theta}$  along any nontangential curve in  $\{|z| > 1\}$ , (or in  $\{|z| < 1\}$  respectively). The *boundary value* limits form two closed subspaces of the (Banach) space of essentially bounded functions  $L_{m \times p}^\infty([-\pi, \pi], d\theta/2\pi)$ , whose Fourier coefficients of negative (respectively, positive) index vanish.

If  $\hat{f} \in H_m^2$  and  $\hat{A} \in H_{m \times p}^\infty$ , then it is obvious that  $\hat{f}\hat{A} \in H_p^2$ . The space  $H_{m \times p}^\infty$  plays very naturally the role of *space of linear causal operators*  $A : H_m^2 \rightarrow H_p^2$ . Note that the operator of multiplication by the function  $e : \theta \mapsto e^{-i\theta}$  (the Fourier transform of the backward shift) maps the space  $H_p^2$  into itself. The following theorem, whose proof is not particularly difficult but is skipped for reasons of space, establishes the converse.

**Theorem 4.3.3 (Bochner-Chandrasekharan).** *A linear bounded map  $A : H_m^2 \rightarrow H_p^2$  which commutes with the operator of multiplication by the function  $e : \theta \mapsto e^{-i\theta}$  is the operator of multiplication by an  $m \times p$  matrix function  $\hat{A}$  in  $H_{m \times p}^\infty$ .*

Note that the property of commuting with  $e^{-i\theta}$  is just *time-invariance*. In the time-domain this result is saying that the most general linear time-invariant operation on causal sequences is convolution with a causal matrix kernel whose Fourier transform is in  $H_{m \times p}^\infty$ .

## 4.4 Analytic spectral factorization

The main question that we shall address in this section is under what conditions a  $m$ -dimensional stationary process admits a causally equivalent normalized white noise. In formulas, under what conditions on the second order description of the process does there exist a normalized white noise process  $w_-$  such that (4.1.7) holds true. Equivalently, when can a stationary process  $y$  be represented as in (4.1.11), i.e. can be generated as the output of a *causal* time-invariant linear filter having as input a normalized white noise processes  $w_-$ . Such a process was called a (forward) p.n.d. processes in the previous section.

Since (4.1.7) obviously implies that  $\mathbf{H}(w_-) = \mathbf{H}(y)$ , a p.n.d process must be orthonormalizable. For this reason we shall henceforth *assume that  $y$  has an absolutely continuous spectrum with a factorizable spectral density matrix of constant*

rank  $p \leq m$ . As we shall see below, the p.n.d. property implies the existence of solutions  $W$  of the spectral factorization problem with strong analytical properties.

Hereafter it will be convenient to write the spectral density matrix of a stationary process as a function of the complex variable  $z$ , defined on the unit circle of the complex plane. Since  $\Phi(e^{i\theta})$  is a real function,  $\Phi(e^{i\theta})^* = \Phi(e^{-i\theta})'$ , so that using the variable  $z = e^{i\theta}$ , the Hermitian symmetry property of the spectrum (compare (3.4.3)) becomes

$$\Phi(1/z)' = \Phi(z). \tag{4.4.1}$$

which will naturally also be called *parahermitian symmetry*.

Let  $H_r^2$  be the Hardy space of  $r$ -dimensional row-vector functions that are analytic in the unit disc  $\mathbb{D}$ . We refer the reader to Section 4.3 for the basics of Hardy space theory needed in the sequel.

**Theorem 4.4.1.** *A stationary  $m$ -dimensional process  $y$  can be represented as a causal functional of a normalized  $r$ -dimensional ( $r \geq p$ ) white noise process  $w$ , or, equivalently, there is a normalized  $r$ -dimensional white noise  $w$  such that*

$$\mathbf{H}_t^-(y) \subset \mathbf{H}_t^-(w), \quad t \in \mathbb{Z}, \tag{4.4.2}$$

*only if there are  $m \times r$  analytic spectral factors of  $\Phi$ , i.e. only if there are solutions  $W$ , of the spectral factorization equation,*

$$\Phi(z) = W(z)W(1/z)'. \tag{4.4.3}$$

*with rows in the Hardy space  $H_r^2$ .*

*Conversely, if  $\Phi$  admits analytic spectral factors, the process  $y$  is p.n.d.; i.e. there exist in particular an analytic spectral factor  $W_-$  and a normalized white process  $w_-$  such that  $dy = W_-dw_-$ , for which the inclusion (4.4.2) holds with the equality sign. The analytic spectral factor  $W_-$  is a.e. of full rank  $p$ .*

The proof of “only if” part is relatively straightforward and will be given presently.

**Proof.** That (4.4.2) implies the existence of an analytic  $m \times r$  spectral factors of  $\Phi$ , follows from Lemma 4.1.3, a computation similar to that in (4.2.3) and the Paley-Wiener theorem 4.3.2 in Section 4.3 which states that the transfer function of any causal shaping filter must have rows in  $H^2$  (i.e. be analytic). The particular form of the spectral factorization equation (4.4.3) results from the fact that we are actually looking for *real* spectral factors here, for which  $W(e^{i\theta})^* = W(e^{-i\theta})'$ .  $\square$

The proof of sufficiency will be given later by showing that the existence of analytic spectral factors entails that of a special full-rank analytic factor  $W_-$ , called the *outer spectral factor*, which in a sense (to be made precise later on) has an analytic left-inverse and hence may serve as the transfer function of a causal whitening filter generating a normalized white noise  $w_-$  for which the inclusion (4.4.2) is actually an equality. In order to characterize the special analytic spectral factor  $W_-$  and



the special whitening process  $w_-$ , which satisfies the causal equivalence property (4.1.7) we shall have to go through a rather long digression. The first step is to introduce the Wold decomposition.

## 4.5 The Wold decomposition

A fundamental classification of second order processes is made on the basis of certain properties of their one-step ahead predictor. We first define a class of processes which are exactly predictable given their infinite past.

**Definition 4.5.1.** *A second order process  $y$  is purely deterministic (abbreviated to p.d. in the following), if the one-step prediction error,*

$$e(t) := y(t) - \hat{y}_-(t) = y(t) - \mathbb{E}[y(t) | \mathbf{H}_t^-(y)] \tag{4.5.1}$$

*is zero (a.s.).*

Obviously  $y$  is p.d. if and only if the components of  $y(t)$  belong to the past space at time  $t$ , i.e.  $\mathbf{H}_{t+1}^-(y) = \mathbf{H}_t^-(y)$ . In fact, since  $y$  is stationary, this will happen at every instant of time, so that the p.d. property is equivalent to

$$\mathbf{H}_t^-(y) = \mathbf{H}(y), \quad t \in \mathbb{Z} \tag{4.5.2}$$

**Example 4.5.2.** Consider the process

$$y(t) = \sum_{-N}^{+N} y_k e^{i\theta_k t}, \quad t \in \mathbb{Z}$$

where  $\theta_{-k} = -\theta_k$ , ( $\theta_0 = 0$ ) are deterministic frequencies and the  $\{y_k\}$  are zero mean-uncorrelated random variables with finite variances,  $\text{var } y_{-k} = \text{var } y_k = \sigma_k^2$ . This is a real stationary process, which is the sum of a fixed random variable  $y_0$  plus  $N$  uncorrelated harmonic oscillations with random amplitude. It is the simplest example of a purely deterministic stationary process.

To check this last statement we use a system theoretic argument, as follows. Note that  $y(t)$  can be formally written as the output of a linear system whose state variable  $x(t)$  satisfies a  $2N + 1$ -dimensional vector difference equation,

$$x(t + 1) = \Omega x(t), \quad \Omega = \text{diag}\{e^{-i\theta_N}, e^{-i\theta_{N-1}}, \dots, e^{i\theta_{N-1}}, e^{i\theta_N}\}$$

with initial conditions  $x_k(0) = y_k, k = -N, \dots, N$ . In fact  $y(t) = \sum_{k=-N}^N x_k(t)$  or, in vector notation,

$$y(t) = cx(t), \quad c = [1, \dots, 1] \in \mathbb{R}^{2N+1}.$$

Assuming that the  $\theta_k$  are all distinct (if not, the system would be non-minimal and could be reduced to a similar one of lower dimension), it is easy to check that this linear system is *completely observable* (see e.g. [60]). Hence, given that the matrix

$\Omega$ , describing the dynamics of the system is trivially invertible (as the eigenvalues are exactly  $e^{\pm i\theta_k}$ ), looking at the inverse system  $x(t-1) = \Omega^{-1}x(t)$ ,  $y(t) = cx(t)$ , one can reconstruct the state,  $x(t_0)$ , as a linear function of the  $2N+1$  preceding output samples, say  $y(t_0-1), \dots, y(t_0-2N-2)$ . It follows that

$$y(t_0) = cx(t_0) \in \text{span} \{y(t_0-1), \dots, y(t_0-2N-2)\} \subset \mathbf{H}_{t_0}^-(y)$$

for all  $t_0$ . Hence the prediction of  $y(t_0)$  can be done exactly given the past, and  $y$  is p.d..

The same reasoning can be applied, using a limiting argument, to the infinite sum

$$y(t) = \sum_{-\infty}^{+\infty} y_k e^{i\theta_k t}, \quad y_k \perp y_j, \quad k \neq j$$

under the additional assumptions that the series  $\sum_{-\infty}^{+\infty} \sigma_k^2$  converges, which guarantees convergence of the infinite sum of harmonic oscillations.

Let us define the *remote past* and the *remote future* of the process  $y$  to be the subspaces

$$\mathbf{H}_{-\infty}(y) = \bigcap_{t \leq k} \mathbf{H}_t^-(y) \quad \mathbf{H}_{+\infty}(y) = \bigcap_{t \geq k} \mathbf{H}_t^+(y), \quad (4.5.3)$$

respectively. Since  $\mathbf{H}_t^-(y)$  is increasing with  $t$  and  $\mathbf{H}_t^+(y)$  is decreasing with  $t$  the two intersections are actually independent of which index  $k$  is chosen as a starting time. Clearly  $y$  is p.d. if and only if

$$\mathbf{H}_{-\infty}(y) = \mathbf{H}_t^-(y) = \mathbf{H}(y) \quad t \in \mathbb{Z}. \quad (4.5.4)$$

This notion relates to the degeneracy of the *causal* prediction problem for the process  $y$  and is asymmetric in time. We shall introduce also the notion of a *backward* p.d. process, for which the *backward prediction error*

$$\bar{e}(t) := y(t) - \hat{y}_+(t) = y(t) - \mathbf{E}[y(t) | \mathbf{H}_{t+1}^+(y)] \quad (4.5.5)$$

is zero (a.s.). For a backward p.d. process one has, dually,

$$\mathbf{H}_{+\infty}(y) = \mathbf{H}_t^+(y) = \mathbf{H}(y) \quad t \in \mathbb{Z}. \quad (4.5.6)$$

Later in this section it will be shown that, for a large class of so-called *reversible processes*, the remote past and the remote future *are the same*; i.e.,

$$\mathbf{H}_{-\infty}(y) = \mathbf{H}_{+\infty}(y), \quad (4.5.7)$$

and hence in particular the p.d. property in the forward and in the backward direction are actually the same thing.

From the definition (4.5.3) it is immediate to check that the property

$$U^t \mathbf{H}_{-\infty}(y) = \mathbf{H}_{-\infty}(y) \quad U^t \mathbf{H}_{+\infty}(y) = \mathbf{H}_{+\infty}(y) \quad t \in \mathbb{Z}.$$

holds for both positive and negative times  $t$ . This is equivalent to saying that the subspaces  $\mathbf{H}_{-\infty}(y)$ ,  $\mathbf{H}_{+\infty}(y)$  are invariant under the action both of the shift  $U$  and its adjoint  $U^*$  and are hence *doubly invariant* in the sense defined in Appendix A.1.

The “simple” invariance property can be of two types. Recall that  $U$  is called the *forward-* or *right-shift* operator, while the adjoint  $U^* = U^{-1}$  is called the *backward-* or *left-shift* operator on  $\mathbf{H}(y)$ .

**Definition 4.5.3.** *A subspace  $\mathbf{Y} \subset \mathbf{H}$  is called backward-shift (or left-shift) invariant if*

$$U^*\mathbf{Y} \subset \mathbf{Y} \tag{4.5.8}$$

*while  $\bar{\mathbf{Y}}$  is forward-shift (or right-shift) invariant if*

$$U\bar{\mathbf{Y}} \subset \bar{\mathbf{Y}}. \tag{4.5.9}$$

Examples of backward-shift invariant and forward-shift invariant subspaces are, respectively, the past  $\mathbf{H}^-(y)$  and the future,  $\mathbf{H}^+(y)$ . We have already noted that the invariance conditions of the definition are an equivalent way to express the fact that the stationary family of subspaces  $\{\mathbf{Y}_t := U^t\mathbf{Y} \mid t \in \mathbb{Z}\}$  generated by  $\mathbf{Y}$ , is *increasing* and, respectively, that  $\{\bar{\mathbf{Y}}_t := U^t\bar{\mathbf{Y}} \mid t \in \mathbb{Z}\}$  is *decreasing* in time.

Generalizing the definitions in (4.5.3), the *remote past* and the *remote future* of an arbitrary increasing or decreasing stationary family of subspaces, respectively, are defined as

$$\mathbf{Y}_{-\infty} = \bigcap_t \mathbf{Y}_t \quad \bar{\mathbf{Y}}_{+\infty} = \bigcap_t \bar{\mathbf{Y}}_t. \tag{4.5.10}$$

A backward-shift invariant subspace  $\mathbf{Y}$  is called *purely nondeterministic* (abbreviated to p.n.d. in the following) if  $\mathbf{Y}_{-\infty} = \{0\}$  (the subspace containing only the zero element). Dually, we shall call a forward-shift invariant subspace  $\bar{\mathbf{Y}}$  *purely nondeterministic*, if  $\bar{\mathbf{Y}}_{+\infty} = \{0\}$ . It is easy to see that for purely nondeterministic subspaces the inclusions in (4.5.8) or in (4.5.9) are *strict*.

The *multiplicity of a simply invariant subspace  $\mathbf{Y}$* , can also be defined as the smallest number of generators of  $\mathbf{Y}$ , i.e. the smallest number,  $m$ , of random variables,  $y_1, \dots, y_m$  in  $\mathbf{Y}$ , such that

$$\overline{\text{span}}\{U^t y_k \mid k = 1, \dots, m, t \leq 0\} = \mathbf{Y}$$

Let  $\mathbf{H}(\mathbf{Y})$  be the smallest doubly invariant subspace containing  $\mathbf{Y}$ . Since

$$\mathbf{H}(\mathbf{Y}) := \bigvee_{t \in \mathbb{Z}} U^t \mathbf{Y} = \overline{\text{span}}\{U^t y_k \mid k = 1, \dots, m, t \in \mathbb{Z}\}$$

the multiplicity of  $\mathbf{H}(\mathbf{Y})$  is seen to be the same of that of  $\mathbf{Y}$  and is in fact the same concept defined previously in Section 2.5. Naturally, whenever  $\mathbf{Y}$  is a subspace of a larger Hilbert space of finite multiplicity, it must also have finite multiplicity.

The theorems and the corollaries below are slight generalizations of facts discovered by H. Wold and published in his seminal 1938 doctoral thesis.

**Theorem 4.5.4 (Wold).** *A left-invariant subspace  $\mathbf{Y}$  is p.n.d if and only if it is the past space at time zero of some vector-valued stationary white noise process.*

There is in fact a unique (modulo multiplication by a constant orthogonal matrix) normalized white noise  $w$  such that

$$\mathbf{Y} = \mathbf{H}^-(w) \tag{4.5.11}$$

and the dimension of  $w$  is equal to the multiplicity of  $\mathbf{Y}$ .

Dually, a right-invariant subspace  $\bar{\mathbf{Y}}$  is p.n.d. if and only if it is the future space at time zero of some stationary vector white noise process. There is a unique (modulo multiplication by a constant orthogonal matrix) normalized white noise  $\bar{w}$  such that

$$\bar{\mathbf{Y}} = \mathbf{H}^+(\bar{w}) \tag{4.5.12}$$

and the dimension of  $\bar{w}$  is equal to the multiplicity of  $\bar{\mathbf{Y}}$ . The white noises  $w$  and  $\bar{w}$  are called the generating processes of the invariant subspaces  $\mathbf{Y}$  and  $\bar{\mathbf{Y}}$ .

**Proof.** It will be enough to prove the first part of the statement only. To this purpose, consider the orthogonal complement  $\mathbf{W}_{t-1}$  of  $\mathbf{Y}_{t-1}$  in  $\mathbf{Y}_t$ , so that

$$\mathbf{Y}_t = \mathbf{W}_{t-1} \oplus \mathbf{Y}_{t-1}.$$

Note that  $\mathbf{W}_{t-1}$  is nontrivial, as the sequence of subspaces  $\{\mathbf{Y}_t\}$  must be strictly increasing by the p.n.d. property and so, by iterating the orthogonal decomposition for  $t-1, t-2, \dots, s$  we obtain

$$\mathbf{Y}_t = \mathbf{W}_{t-1} \oplus \mathbf{W}_{t-2} \dots \oplus \mathbf{W}_s \oplus \mathbf{Y}_s, \quad s < t. \tag{4.5.13}$$

It is obvious that the subspaces  $\{\mathbf{W}_t\}$  are pairwise orthogonal (by construction), moreover, by Proposition A.2.5 in Appendix A.2, it is clear that  $\{\mathbf{W}_t\}$  is a stationary sequence of subspaces, i.e.  $\mathbf{W}_t = U^t \mathbf{W}$ ,  $t \in \mathbb{Z}$  where

$$\mathbf{W} := U\mathbf{Y} \ominus \mathbf{Y}.$$

A subspace  $\mathbf{W}$  with these two properties, is called a *wandering subspace* for the shift  $U$ .

Now, for  $t$  fixed, the two projections of any element  $\eta(t) \in \mathbf{Y}_t$  onto  $\bigoplus_{s=-\infty}^{t-1} \mathbf{W}_s$  and onto  $\mathbf{Y}_s$ , say

$$\eta(t) = \hat{\eta}(s) + \tilde{\eta}(s),$$

both must converge as  $s \rightarrow -\infty$ . That  $\hat{\eta}(s)$  converges follows by the orthogonal series Lemma A.1.1 in the appendix, since  $\hat{\eta}(s)$  is a sum of orthogonal terms and its norm is bounded above by  $\|\eta(t)\|$ ; therefore the other addend in the sum also converges as the left hand member is independent of  $s$ . However,  $\lim_{s \rightarrow -\infty} \tilde{\eta}(s)$  must belong to  $\mathbf{H}_{-\infty}$  and by the p.n.d. property this subspace may contain only the zero random variable. Hence  $\tilde{\eta}(s)$  tends to zero as  $s \rightarrow -\infty$  and so any  $\eta(t) \in \mathbf{Y}_t$  is equal to the infinite orthogonal sum of its projections onto the subspaces  $\{\mathbf{W}_s, -\infty < s < t\}$ .

This property may be written as

$$\mathbf{Y}_t = \bigoplus_{s=-\infty}^{t-1} \mathbf{W}_s \tag{4.5.14}$$

It remains only to notice that, if  $\mathbf{Y}$  has finite multiplicity,  $\mathbf{W}$  is finite-dimensional. For, it is the projection of a space with a finite number, say  $p$ , of generators onto the orthogonal complement  $\mathbf{Y}^\perp$ . Then, any orthonormal basis  $\{w_1, \dots, w_p\}$  of  $\mathbf{W}$ , shifted in time, becomes a generating normalized white noise,  $\{w_k(t) := U^t w_k, t \in \mathbb{Z}\}$ , yielding the representation (4.5.11).

The proof of the converse statement follows from the next lemma.  $\square$

**Lemma 4.5.5.** *The remote past (and the remote future) of a white noise process is trivial.*

**Proof.** Let  $u = \{u(t)\}$  be a (not necessarily stationary) second order process with orthogonal variables, i.e.  $u(t) \perp u(s), t, s \in \mathbb{Z}$ . Every random variable  $h \in \mathbf{H}_t(u)$  is orthogonal to the future values  $u(s), s \geq t$ . In particular, if  $h \in \mathbf{H}_{-\infty}(u)$  it follows that  $h$  must be orthogonal to all  $u(t), t \in \mathbb{Z}$ . Therefore, by continuity of the scalar product,  $h$  must be orthogonal to the whole of  $\mathbf{H}(u)$ . Hence it must be zero.  $\square$

The definition of a (forward) p.n.d. process given in the previous section in terms of causal equivalence to white noise has now a geometric counterpart in terms of remote past of the process.

**Corollary 4.5.6.** *A stationary process  $y$  is p.n.d. in the forward sense if and only if  $\mathbf{H}^-(y)$  is a p.n.d. subspace i.e.  $\cap_t \mathbf{H}_t^-(y) = 0$ . The normalized innovation  $w_-(t)$  is an orthonormal basis for the wandering subspace at time  $t$ ,  $\mathbf{H}_{t+1}^-(y) \ominus \mathbf{H}_t^-(y)$  of  $\mathbf{H}^-(y)$ .*

**Proof.** Necessity is an immediate application of the lemma that we have just proven above. Conversely, if  $\mathbf{H}^-(y)$  is a p.n.d. subspace, then by Wold's Theorem 4.5.4 there is a normalized white noise  $w_-$  which is causally equivalent to  $y$ .  $\square$

Note that for full-rank processes the one-step prediction error  $e(t) = y(t) - \hat{y}_-(t)$  is a basis for the wandering (or innovation) subspace  $\mathbf{W}_t$ . Hence we have a causal representation of  $y(t)$  in terms of past prediction errors,

$$y(t) = \sum_{-\infty}^t V_-(t-s)e(s) \tag{4.5.15}$$

where

$$V_-(t-s)e(s) := \mathbb{E}[y(t) | e(s)] = \mathbb{E}[y(t) e(s)'] \{ \mathbb{E}[e(s) e(s)'] \}^{-1} e(s).$$

Since for  $s > t$ ,  $e(s)$  is uncorrelated with (i.e. orthogonal to)  $y(t)$ , it is readily seen from the expression above that  $V_-$  is a causal function with rows in  $\ell_m^2$ . Note also that

$$V_-(0) = I.$$

The prediction error process  $\{e(t)\}$  is sometimes called the forward *unnormalized innovation process* of  $y$  and (4.5.15) the *unnormalized innovations representation* of

$y$ . It is a natural representation for full-rank processes. In general, for processes of rank  $p < m$  we can write

$$e(t) = D_- w_-(t) \tag{4.5.16}$$

where  $D_-$  is some (non-unique)  $m \times p$  matrix factor of the *innovation variance*  $\Lambda := E\{e(t)e(t)'\}$ , i.e.

$$\Lambda = D_- D_-'$$

We collect the above observations in the following statement.

**Proposition 4.5.7.** *For a p.n.d. process the rank of the variance matrix of the unnormalized innovation coincides with the rank (and multiplicity) of the process itself; i.e.*

$$p = \text{rank } y = \text{rank } E\{e(t)e(t)'\}. \tag{4.5.17}$$

Naturally, everything that has been said so far, can be repeated *mutatis mutandis* for the backward prediction error  $\bar{e}(t)$  leading to *anticausal* (or backward) unnormalized innovations representations of the process  $y$  with symmetric properties.

The p.n.d. property is in a sense a special case. How things go in general, is described by the following theorem.

**Theorem 4.5.8 (Wold).** *Every backward-shift invariant subspace  $\mathbf{Y}$  admits a decomposition in the orthogonal direct sum of a doubly invariant and a purely non-deterministic subspace. In fact,*

$$\mathbf{Y} = \mathbf{Y}_{-\infty} \oplus \mathbf{Z} \tag{4.5.18}$$

where  $\mathbf{Z}$  is backward-shift invariant and p.n.d.. Dually, every forward shift invariant subspace  $\bar{\mathbf{Y}}$  admits the orthogonal decomposition

$$\bar{\mathbf{Y}} = \bar{\mathbf{Y}}_{+\infty} \oplus \bar{\mathbf{Z}} \tag{4.5.19}$$

where  $\bar{\mathbf{Z}}$  is forward-shift invariant and p.n.d.. The two decompositions with the stated properties are unique.

**Proof.** Since  $\mathbf{Y}_{-\infty} \subseteq \mathbf{Y}$ , we may define  $\mathbf{Z}$  to be the orthogonal complement,  $\mathbf{Y}_{-\infty}^\perp$ , of  $\mathbf{Y}_{-\infty}$  in  $\mathbf{Y}$ . Since  $\mathbf{Y}_{-\infty}$  is invariant also for the adjoint,  $U^*$ , of the backward shift ( $U^*$ ), it follows from Lemma A.1.6 that

$$U^* \mathbf{Y}_{-\infty}^\perp \subset \mathbf{Y}_{-\infty}^\perp$$

from which it is seen that  $\mathbf{Y}_{-\infty}$  is a reducing subspace for  $U^*$  (i.e. it is invariant together with its orthogonal complement). At the same time,  $\mathbf{Y}_{-\infty}^\perp = \mathbf{Z}$  is also left-shift invariant, i.e. the stationary family  $\mathbf{Z}_t = U_t \mathbf{Z}$  is increasing with  $t$ . There cannot be nonzero elements  $\xi$  belonging to the intersection  $\cap_t \mathbf{Z}_t$ , since any such element  $\xi$  should then be in  $\mathbf{Z}_t \subset \mathbf{Y}_t$  for all  $t$  and therefore should belong to the intersection,  $\mathbf{Y}_{-\infty}$ . This would however imply that  $\xi \in \mathbf{Y}_{-\infty}$  is orthogonal to itself

and hence equal to zero. Uniqueness of a decomposition with the stated properties is obvious since by Proposition A.2.5 of Appendix A.2,  $\mathbf{Y} = \mathbf{Y}^{(1)} \oplus \mathbf{Y}^{(2)}$  with  $\mathbf{Y}^{(1)}$  p.d. and  $\mathbf{Y}^{(2)}$  p.n.d. implies that  $\cap_t U_t \mathbf{Y} = \cap_t U_t \mathbf{Y}^{(1)} \oplus \cap_t U_t \mathbf{Y}^{(2)}$  and the p.d. subspace must necessarily coincide with  $\mathbf{Y}_{-\infty}$ .

The dual statement is proven in exactly the same way.  $\square$

**Corollary 4.5.9.** *Every stationary vector process  $y$  admits a decomposition*

$$y(t) = v(t) + z(t), \quad t \in \mathbb{Z} \tag{4.5.20}$$

where the processes  $v = \{v(t)\}$  and  $z = \{z(t)\}$  are completely uncorrelated, i.e.  $E\{v(t)z(s)'\} = 0$ ,  $t, s \in \mathbb{Z}$ ,  $v = \{v(t)\}$  is forward p.d. and  $z = \{z(t)\}$  is forward p.n.d..

There is just one decomposition (4.5.20), satisfying the conditions above, such that  $\mathbf{H}_t^-(z) \subset \mathbf{H}_t^-(y)$ ,  $t \in \mathbb{Z}$ . In this case  $v$  generates the same remote past as  $y$ , i.e.

$$\mathbf{H}(v) = \mathbf{H}_{-\infty}(v) = \mathbf{H}_{-\infty}(y), \quad \mathbf{H}_{-\infty}(z) = 0 \tag{4.5.21}$$

The processes  $v$  and  $z$  are called the (forward) p.d. and p.n.d. components of  $y$ .

Dually, an analogous orthogonal decomposition exists

$$y(t) = \bar{v}(t) + \bar{z}(t), \quad t \in \mathbb{Z} \tag{4.5.22}$$

where the properties of  $\bar{v}(t)$  of being p.d. and of  $\bar{z}(t)$  of being p.n.d. hold in the backward sense. There is just one (backward) p.n.d. process  $\bar{z}$  such that  $\mathbf{H}_t^+(\bar{z}) \subset \mathbf{H}_t^+(y)$ ,  $t \in \mathbb{Z}$ , satisfying the conditions above. It is called the (backward) p.n.d. component of  $y$ .

**Proof.** Consider the decomposition of the invariant subspace  $\mathbf{H}_t^-(y)$  into its doubly invariant and (forward) p.n.d. components

$$\mathbf{H}_t^-(y) = \mathbf{H}_{-\infty} \oplus \tilde{\mathbf{H}}_t(y) \quad t \in \mathbb{Z}$$

A (forward) decomposition of  $y(t)$  with all the stated properties, can be obtained by projecting  $y(t)$  onto the above orthogonal sum of subspaces. In fact, define, for  $k = 1, \dots, m$ ,

$$v_k(t) := E[y_k(t) | \mathbf{H}_{-\infty}], \quad z_k(t) := E[y_k(t) | \tilde{\mathbf{H}}_{t+1}(y)]$$

Since  $\mathbf{H}_{-\infty} \subset \mathbf{H}_t(y)$  and hence  $\mathbf{H}_{-\infty} \cap \mathbf{H}_t(y)^\perp = 0$ , it follows from Lemma 2.2.11 in Chapter 2, that

$$\mathbf{H}_{-\infty} = E^{\mathbf{H}_{-\infty}} \mathbf{H}_t(y) \oplus (\mathbf{H}_{-\infty} \cap \mathbf{H}_t(y)^\perp) = E^{\mathbf{H}_{-\infty}} \mathbf{H}_t(y) = \overline{\text{span}}\{v(s) | s \leq t\}$$

and similarly, since  $\tilde{\mathbf{H}}_t(y) \subset \mathbf{H}_t(y)$ ,

$$\tilde{\mathbf{H}}_t(y) = \overline{\text{span}}\{u(s) | s \leq t\}$$

so that the processes  $v$  and  $z$  have the stated properties. Uniqueness now follows by the same argument given in the proof of Theorem 4.5.8.

In the same way one can prove the statement relative to the backward decomposition.  $\square$

Note that the multiplicity of the p.n.d. component  $z$  will in general be smaller than that of  $y$ .

### Reversibility

Now it follows from Corollary 4.5.9 that the past and future spaces of  $y$  admit the decompositions,

$$\mathbf{H}_t^-(y) = \mathbf{H}_{-\infty}(y) \oplus \mathbf{H}_t^-(u) \tag{4.5.23}$$

$$\mathbf{H}_t^+(y) = \mathbf{H}_{+\infty}(y) \oplus \mathbf{H}_t^+(\bar{u}) \tag{4.5.24}$$

and, letting  $t \rightarrow -\infty$  in (4.5.23) and  $t \rightarrow +\infty$  in (4.5.24) we get

$$\mathbf{H}(y) = \mathbf{H}_{-\infty} \oplus \mathbf{H}(u) \tag{4.5.25}$$

$$\mathbf{H}(y) = \mathbf{H}_{+\infty} \oplus \mathbf{H}(\bar{u}). \tag{4.5.26}$$

It is natural to ask when the two decompositions are the same. To this end, we shall say that a stationary process is *reversible* if the remote past and the remote future coincide; i.

$$\mathbf{H}_{-\infty}(y) = \mathbf{H}_{+\infty}(y). \tag{4.5.27}$$

Consequently, if  $y$  is reversible, the decompositions (4.5.25) and (4.5.26) are the same. In particular,

$$\mathbf{H}(u) = \mathbf{H}(\bar{u}). \tag{4.5.28}$$

Therefore, a purely nondeterministic process that is reversible is also purely nondeterministic in the backward direction.

To understand reversibility we introduce the “time-reversed” process  $\bar{y}(t) := y(-t)$  whose covariance function we denote by  $\bar{\Lambda}(\tau)$ . Since

$$\bar{\Lambda}(\tau) = \mathbb{E} \bar{y}(t + \tau) \bar{y}(t)' = (\mathbb{E} y(-t) y(-t - \tau)')' = \Lambda(\tau)'$$

the spectral distribution of  $\bar{y}$  is the transpose of that of  $y$ . Hence, the spectral densities of the two processes can exist only simultaneously and are necessarily the transpose of each other, i.e.  $\bar{\Phi}(e^{i\theta}) = \Phi(e^{i\theta})'$ . Now since

$$\begin{aligned} \mathbf{H}_t^-(\bar{y}) &= \overline{\text{span}}\{\bar{y}_k(s); k = 1, \dots, m, s < t\} = \overline{\text{span}}\{y_k(-s); k = 1, \dots, m, s < t\} \\ &= \overline{\text{span}}\{\bar{y}_k(\tau); k = 1, \dots, m, \tau \geq -t\} = \mathbf{H}_{-t}^+(y) \end{aligned}$$

we have

$$\mathbf{H}_{-\infty}(\bar{y}) = \bigcap_{t \leq k} \mathbf{H}_{-t}^+(y) = \bigcap_{t' \geq -k} \mathbf{H}_{t'}^+(y) = \mathbf{H}_{+\infty}(y)$$

and  $\bar{y}$  is forward p.n.d. if and only if  $y$  is backward p.n.d. We may summarize this in the following proposition



**Proposition 4.5.10.** *A forward p.n.d. process is reversible if and only if the transpose of its spectral density matrix also admits analytic spectral factors.*

This can be seen also from the following argument. Let  $\Phi(z)'$  have the factorization

$$\Phi(z)' = G(z)G(z^{-1})'$$

where  $G(z)$  is an analytic spectral factor, which without loss of generality we can assume of full column rank. Since then  $\Phi(z)$  admits also the factorization  $\Phi(z) = \bar{W}(z)\bar{W}(1/z)'$  where  $\bar{W}(z) := G(1/z)$  is a *coanalytic* spectral factor, i.e. a matrix function whose rows belong to the conjugate Hardy space  $\bar{H}_p^2$ , we can, by dualizing the argument used in the proof of Theorem 4.4.1, represent  $y$  by an *anticausal* linear time-invariant convolution operator

$$y(t) = \sum_t^{+\infty} \bar{W}(t-s)\bar{w}(s)$$

where  $\bar{w}$  is a normalized white noise. From this representation we have  $\mathbf{H}_t^+(y) \subset \mathbf{H}_t^+(\bar{w})$  and since  $\cap \mathbf{H}_t^+(\bar{w}) = \{0\}$ , the backward p.n.d. property of  $y$  follows.

**Proposition 4.5.11.** *Every full rank p.n.d. process is reversible.*

**Proof.** The proof for vector processes depends on a function-theoretic criterion for pure nondeterminism which can be found in [106, p. 85]. The criterion states that a full rank stationary process is p.n.d. if and only if

$$\int_{-\pi}^{\pi} \log \det \Phi(e^{i\theta}) d\theta > -\infty$$

which, since  $\det \Phi(e^{i\theta}) = \det \Phi(e^{i\theta})'$ , can only hold simultaneously both for  $\Phi$  and its transpose.

A very simple “geometric” proof for scalar processes can be given as follows. Define the *reflection operator*  $R$  on the set  $Y := \{y(t) \mid t \in \mathbb{Z}\}$  by setting  $Ry(t) := y(-t)$ . Similar to what was done for extending the shift operator  $U$  in Section 2.5, we may check that  $R$  is isometric and linear on  $Y$  and can be extended to the vector space generated by  $Y$  so that  $R \sum \alpha_k y(t_k) := \sum \alpha_k y(-t_k)$ . Since  $R$  is isometric, it can be extended as a unitary operator to the closure  $\mathbf{H}(y)$ . On this space  $R$  is actually a unitary *involution* i.e.  $R^2 = I$ . By continuity we have  $R\mathbf{H}_{-\infty}(y) = \mathbf{H}_{+\infty}(y)$  and hence one of these two spaces is zero if and only if the other is. Unfortunately, in the vector case the operator  $R$  is not isometric (the reader is invited to check this).  $\square$

## 4.6 The outer spectral factor

We shall now return to the missing half of the proof of Theorem 4.4.2. Our first objective will be to investigate what special properties should be enjoyed by a pair

$(W, w)$ , with  $W$  a spectral factor and  $w$  a normalized white noise process satisfying  $d\hat{y} = Wd\hat{w}$ , in order to satisfy the causal equivalence condition (4.1.7). Because of stationarity, the causal equivalence condition is equivalent to equality of the two subspaces at time zero,

$$\mathbf{H}^-(w) = \mathbf{H}^-(y),$$

and this, in view of the representation Theorem 3.5.1, is equivalent to asking that the spectral factor  $W$  should be such that<sup>14</sup>

$$\overline{\text{span}}\{e^{i\theta t}e_k \mid k = 1, \dots, p, t \leq 0\} = \overline{\text{span}}\{e^{i\theta t}W_k \mid k = 1, \dots, m, t \leq 0\}$$

where  $e_k$  is the function identically equal to the  $k$ -th unit vector

$$e_k(e^{i\theta}) = [0, \dots, 1, \dots, 0]$$

(1 in the  $k$ -th place). Clearly the first member is just the Hardy space  $H_p^2$ , so we may rewrite this equality as

$$\overline{\text{span}}\{e^{i\theta t}W_k \mid k = 1, \dots, m, t \leq 0\} = H_p^2 \tag{4.6.1}$$

This is a function-theoretic characterization of the spectral factors  $W$  for which the causal equivalence condition holds. The equation above states in fact that any such spectral factor must be an *outer function in  $H_p^2$*  (also called *minimum phase*). The formal definition is given below.

**Definition 4.6.1.** An  $m \times p$  matrix-valued function  $F$  with rows in  $H_p^2$  is called *outer* if<sup>15</sup>  $\overline{\text{span}}\{z^t F_k \mid k = 1, \dots, m, t \leq 0\} = H_p^2$ . Symmetrically, an  $m \times p$  matrix-valued function  $G$  with rows in  $\overline{H}_p^2$  is called *conjugate outer* if  $\overline{\text{span}}\{z^t G_k \mid k = 1, \dots, m, t \geq 0\} = \overline{H}_p^2$ .

A  $p \times p$  matrix-valued function  $Q \in H_{p \times p}^\infty$  with unitary boundary values on the unit circle,

$$Q(e^{i\theta})Q(e^{i\theta})^* = I \tag{4.6.2}$$

is called *inner*. A conjugate inner function still satisfies (4.6.2) but is instead bounded analytic in  $\{|z| < 1\}$ .

Note that the left-hand member in (4.6.1) is just the smallest *invariant subspace* for the backward (or right-) translation operator  $z^{-1}$  acting on  $H_p^2$ , containing the rows of the matrix  $W$ . Hence a matrix function is outer if and only if the smallest invariant subspace containing its rows is the largest possible, i.e. the whole of  $H_p^2$ .

**Remark 4.6.2.** For any function  $F \in H_p^2$  the invariant subspace  $\overline{\text{span}}\{z^t F \mid t \leq 0\}$  is the closure in  $L_p^2$  of products of  $F$  times scalar *analytic trigonometric polynomials*

<sup>14</sup>In the formulas below the notations are a little inconsistent. To get things to look right one should introduce the (unitary) operator  $M_e$ , of multiplication by the function  $e(e^{i\theta}) = e^{i\theta}$  and instead of  $e^{i\theta t}W_k$  write  $M_e^t W_k$ .

<sup>15</sup>Here again we should write  $M_e^t F_k$  instead of  $z^t F_k$ . Note that  $M_e^t F_k \in H_p^2$  for  $t \leq 0$ .

$$p(z^{-1}) := \sum_{k=0}^N p_k z^{-k}.$$

Now, by the Weierstrass approximation Theorem, these polynomials are dense (in the sup norm) in the subspace of continuous functions on the unit circle  $\{|z| = 1\}$  which have vanishing negative Fourier coefficients. It follows that analytic trigonometric polynomials are dense in the scalar  $H^\infty$  space. Consider a sequence of analytic polynomials  $\{p_n\}$  such that  $p_n \rightarrow \varphi \in H^\infty$  in the sup (denoted  $\infty$ )-norm. Since

$$\|p_n F - \varphi F\|_{L_p^2} \leq \|p_n - \varphi\|_\infty \|F\|_{L_p^2} \rightarrow 0$$

as  $n \rightarrow \infty$ , the linear manifold  $\{\varphi F \mid \varphi \in H^\infty\}$  is a dense vector subspace of the smallest invariant subspace containing  $F$ . In other words,

$$\overline{\text{span}}\{z^t F \mid t \leq 0\} = \overline{\text{span}}\{\varphi F \mid \varphi \in H^\infty\}. \tag{4.6.3}$$

This in particular implies that,

**Proposition 4.6.3.** *An outer matrix function  $F$  must be almost everywhere of full column rank.*

**Proof.** For, by limits of linear combinations of its rows  $\sum_1^m \varphi_k F_k$ ,  $\varphi_k \in H^\infty$ , it must be possible to generate the unit vector functions  $\{e_k, k = 1, \dots, p\}$  of  $H_p^2$ . Hence there exists a sequence of matrices  $H_k \in H_{p \times m}^\infty$  such that

$$H_k F \rightarrow I$$

in  $L^2$  as  $k \rightarrow \infty$ . But then there must be a subsequence  $\{H_{n_k}\}$  such that  $H_{n_k} F$  converges to the identity matrix almost everywhere. This couldn't possibly be true if  $\text{rank } F < p$  on a set of positive measure.  $\square$

Hence an outer spectral factor of a density  $\Phi$  of rank  $p$  must be a full rank spectral factor, of dimension  $m \times p$ <sup>16</sup>. The white noise process for which the causal equivalence condition (4.1.7) holds, is thereby uniquely determined by the outer spectral factor as  $d\hat{w} = W^{-L} d\hat{y}$ .

### Invariant subspaces and the factorization theorem

The subspace in the left member of formula (4.6.1) is just a particular example of invariant subspace for the operator of right translation (i.e. multiplication by  $z^{-1} \equiv e^{-i\theta}$ ) on the space  $L_p^2$ . In the 1960s there has been a great amount of work on the structure of general translation-invariant subspaces and the representation

<sup>16</sup>Soon we shall prove that this factor is essentially unique and denote it by the symbol  $W_-$ .

of invariant subspaces. These results, some of which will be used many times in this book, form now a cornerstone of operator theory.

In general, an invariant subspace  $\mathcal{Y}$  of  $L_p^2$  can be seen as the frequency domain representative of a space of second order random variables in the Hilbert space  $\mathbf{H}(w)$ , generated by some normalized  $p$ -dimensional white noise process  $w$ . This is so, since for any such white noise process the representation operator  $\mathcal{J}_w$  is unitary and we have (see (A.2.5) in Appendix A.2),

$$\mathbf{Y} := \mathcal{J}_w(\mathcal{Y}) \subset \mathcal{J}_w(L_p^2) = \mathbf{H}(w)$$

so that the translation invariance of  $\mathcal{Y}$  is the same as invariance of  $\mathbf{Y}$  with respect to the backward (or left) shift of the process  $w$ , i.e.

$$U^*\mathbf{Y} \subset \mathbf{Y}.$$

It follows then that studying the structure of invariant subspaces of  $L_p^2$  is the same as studying the shift-invariant subspaces of the Hilbert space  $\mathbf{H}(w)$ . In this way we make contact with the problem area discussed in Section 4.5. In fact, the problem of discovering the structure of translation-invariant subspaces is completely answered by the two theorems of Wold which were presented in the previous section.

Following the terminology introduced in Section 4.5, an invariant subspace  $\mathcal{Y} \subset L_p^2$  will be called *doubly invariant* (or *purely deterministic*) if  $z^{-1}\mathcal{Y} = \mathcal{Y}$ . In this case  $\mathcal{Y}_t := z^{-t}\mathcal{Y}$  is actually constant in time. The subspace  $\mathcal{Y}$  is called *purely-non-deterministic* (p.n.d. for short) if  $\cap_t \mathcal{Y}_t = 0$ . An invariant subspace is called of *full range* if

$$\vee_{t \in \mathbb{Z}} z^t \mathcal{Y} = L_p^2,$$

equivalently, a subspace  $\mathcal{Y}$  is full range if its orthogonal complement is p.n.d..

The following result, called the Beurling-Lax Theorem, is a direct corollary of Theorem 4.5.4.

**Theorem 4.6.4.** *Every full-range p.n.d. invariant subspace  $\mathcal{Y} \subset L_p^2$  has the form,*

$$\mathcal{Y} = \{fQ \mid f \in H_p^2\} := H_p^2 Q \tag{4.6.4}$$

where  $Q$  is a  $p \times p$  matrix function with unitary values on the unit circle, i.e.

$$Q(e^{i\theta})Q(e^{i\theta})^* = I. \tag{4.6.5}$$

If  $\mathcal{Y} \subset H_p^2$  then  $Q$  is actually inner. In any case  $Q$  is uniquely determined by  $\mathcal{Y}$ , modulo a constant unitary factor.

**Proof.** By Theorem 4.5.4, every p.n.d. left-invariant subspace  $\mathbf{Y}$  of  $\mathbf{H}(w)$  has the form  $\mathbf{H}^-(u)$  where  $u$  is normalized white noise. If  $\mathbf{Y} \equiv \mathbf{H}^-(u)$  is full-range then the dimensions of  $w$  and  $u$  must be the same (equal to the multiplicity  $p$ ) and, in fact, we must have

$$\mathbf{H}(u) = \mathbf{H}(w).$$

Since  $\mathbf{H}_t^-(u) \subset \mathbf{H}(w)$ , for all  $t$ ,  $u(t)$  can be expressed as a linear (not necessarily causal) functional of the white noise  $w$ . It is an immediate consequence of Theorem 3.5.1 and Lemma 4.2.4 that the Fourier transforms of  $u$  and  $w$  are related to each other by an invertible (a.e.)  $p \times p$  matrix function  $Q = Q(e^{i\theta})$  with rows in  $L_p^2$ , i.e.

$$d\hat{u} = Qd\hat{w},$$

and from the spectral factorization condition 4.2.1 (recall that the spectral density matrix of  $w$  is equal to the Identity matrix), it follows immediately that  $Q$  must be unitary on the unit circle. Therefore, by changing stochastic measure in the integrals, it follows that  $\mathbf{H}^-(u) = \mathcal{J}_{\hat{u}}(H_p^2) = \mathcal{J}_{\hat{w}}(H_p^2Q)$ , that is  $\mathcal{Y} = H_p^2Q$ . This proves the representation formula (4.6.4).

Next, we have  $\mathcal{Y} \subset H_p^2$  if and only if  $\mathbf{H}^-(u) \subset \mathbf{H}^-(w)$  and, by the proven part of Theorem 4.4.1 above, this is true only if  $Q$  is an analytic spectral factor of the identity, i.e. an inner function. The uniqueness of  $Q$  only modulo multiplication by constant unitary (i.e. orthogonal) matrices is a natural consequence of the fact that normalized white noise processes are only distinguishable modulo this kind equivalence.  $\square$

The *inner-outer factorization theorem* for full-rank functions in  $H_p^2$  follows easily from the invariant subspace theorem.

**Theorem 4.6.5.** *Every matrix function  $F \in H_{m \times p}^2$  of full column rank a.e., has a factorization  $F = F_-Q$  where  $F_-$  is outer and  $Q$  inner  $p \times p$ . In this factorization  $F_-$  and  $Q$  are unique up to  $p \times p$  constant orthogonal factors.*

**Proof.** Let  $\mathcal{Y}_F$  be the invariant subspace in  $H_p^2$  generated by the rows of  $F$ . Since  $F$  is of full rank  $\mathcal{Y}_F$  has  $p$  linearly independent generators, and therefore it is full range. Hence  $\mathcal{Y}_F = H_p^2Q$  where  $Q$  is a uniquely determined inner function. Since the rows of  $F$  belong to  $\mathcal{Y}_F$  we have  $F = F_-Q$  where  $F_-$  is an  $m \times p$  matrix function in  $H^2$ . Since multiplication by an inner matrix function is a unitary operator, it is easy to see, by this factorization, that the invariant subspace generated by  $F$  has the form

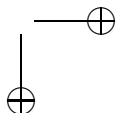
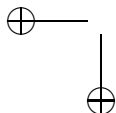
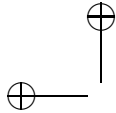
$$\mathcal{Y}_F = \mathcal{Y}_{F_-}Q.$$

By uniqueness of the representation of the invariant subspace  $\mathcal{Y}_F$  it must be that  $\mathcal{Y}_{F_-} = H_p^2$ , i.e.  $F_-$  must be outer. If  $F = GQ_1$  is another such factorization, then  $\mathcal{Y}_F = H_p^2Q_1$  so  $Q$  and  $Q_1$  are equal up to a constant unitary matrix factor. It follows that  $F_-$  and  $G$  must also be equal up to a constant unitary right matrix factor.  $\square$

We shall need also a generalization of Theorem 4.6.5 to non full-rank matrix functions.

**Definition 4.6.6.** *A function  $R \in H_{p \times r}^\infty$  where  $r \geq p$ , such that*

$$R(e^{i\theta})R(e^{i\theta})^* = I_p \tag{4.6.6}$$



is called a unilateral inner function<sup>17</sup>.

**Theorem 4.6.7.** *Every matrix function  $F \in H^2_{m \times r}$  of rank  $p \leq r$  a.e., has a factorization  $F = F_- R$  where  $F_-$  is outer  $m \times p$  and  $R$  is a unilateral inner function of dimension  $p \times r$ . In this factorization  $F_-$  is unique up to  $p \times p$  right constant unitary factors. The factor  $R$  is unique modulo multiplication by orthogonal matrices only if  $p = m$  (in which case  $F_-$  is square).*

**Proof.** Pick an  $r$ -dimensional normalized white noise process  $w$  and consider the stationary  $m$ -dimensional process  $y$ , defined in the same probability space of  $w$ , by the relation  $dy = Fd\hat{w}$ . Since  $F$  is analytic, by the Paley-Wiener Theorem (4.3.2),  $y(t)$  is a causal functional of the process  $w$  and hence

$$\mathbf{H}^-(y) \subset \mathbf{H}^-(w)$$

i.e. the past of  $y$  is a shift-invariant subspace of  $\mathbf{H}^-(w)$  and therefore a p.n.d. subspace of multiplicity  $p$ . This, in view of the Theorem 4.5.4, is equivalent to the existence of a  $p$ -dimensional normalized white noise process  $u$  such that  $\mathbf{H}^-(y) = \mathbf{H}^-(u)$ . It follows that there is an analytic  $p \times r$  spectral factor  $R$  of  $\Phi_w = I_r$  (i.e. a  $p \times r$  unilateral inner function) such that  $d\hat{u} = Rd\hat{w}$ . Hence from the subspace inclusion above we get,

$$\mathbf{H}^-(y) = \mathcal{J}_{\hat{w}}(\overline{\text{span}}\{e^{i\theta t} F_k \mid k = 1, \dots, m, t \leq 0\}) = \mathcal{J}_{\hat{w}}(H^2_p R)$$

and since the rows of  $F$  are contained in the subspace  $H^2_p R$  there must be  $m$  row-functions  $\{G_k \mid k = 1, \dots, m\}$ , in  $H^2_p$  such that  $F = GR$ . On the other hand we must have  $\overline{\text{span}}\{e^{i\theta t} G_k \mid k = 1, \dots, m, t \leq 0\} = H^2_p$  and hence  $G$  is outer. The rest is obvious.  $\square$

At this point we have available the techniques to discuss the solutions of the analytic spectral factorization problem (4.4.3). The main result is stated in the following theorem.

**Theorem 4.6.8.** *Assume  $\Phi(z)$  is an  $m \times m$  spectral density matrix of rank  $p$  a.e. admitting analytic spectral factors. Then  $\Phi(z)$  admits an outer spectral factor  $W_-$ , of dimension  $m \times p$ . This is the unique outer factor of  $\Phi(z)$ , modulo right multiplication by a constant  $p \times p$  unitary matrix.*

*Every full-rank analytic spectral factor  $W$  can be written*

$$W(z) = W_-(z)Q(z), \tag{4.6.7}$$

where  $Q(z)$  is an inner function uniquely determined by  $W \bmod \mathcal{O}$ .

*All other analytic spectral factors of dimension  $m \times r$ ,  $r \geq p$  are of the form*

$$W(z) = W_-(z)R(z), \tag{4.6.8}$$

<sup>17</sup>This denomination is non-standard. Such functions are called *rigid* in Fuhrmann's [34].

where  $R(z)$  is a  $p \times r$  unilateral inner function.

A completely symmetric result holds for the coanalytic spectral factorization problem  $\Phi(z) = \overline{W}(z)\overline{W}(1/z)'$  with  $\overline{W}_k \in H_p^2$ ,  $k = 1, \dots, m$ .

**Proof.** We only need to prove the uniqueness of the outer factor, since the factorizations (4.6.7) and (4.6.8) follow immediately from the outer-inner factorization Theorems 4.6.5 and 4.6.7. To this purpose, let  $W_1$  and  $W_2$  be both outer. Then, by the spectral factorization equation,  $W_1W_1^* = W_2W_2^*$ , and the function,

$$Q := W_1^{-L}W_2 = W_1^*(W_2^*)^{-R} = (W_2^{-L}W_1)^*,$$

is a  $p \times p$  unitary matrix function on the unit circle, irrespective of the version chosen for the left inverses. This follows by the second identity in Lemma 4.2.6 which implies that  $QW_2^{-L}W_1 = W_1^{-L}W_1 = I$  and, likewise,  $W_2^{-L}W_1Q = W_2^{-L}W_2 = I$ , showing that  $W_2^{-L}W_1$  is actually the inverse of  $Q$ .

Now, again by Lemma 4.2.6 we have  $W_1Q := W_1W_1^{-L}W_2 = W_2$  and since both  $W_1$  and  $W_2$  are outer  $Q$  must be a constant unitary matrix. This concludes the proof.  $\square$

**Remark 4.6.9.** Theorem 4.6.8 provides the missing “if” part of the proof of Theorem 4.4.2. For, assume there is an arbitrary analytic spectral factor; then from (4.6.8) we see that  $W_-$  is also necessarily a spectral factor and in fact the (unique) outer factor. Then the left-inverse of this factor provides the whitening filter which generates the innovation process  $w_-$  causally equivalent to  $y$ .

How do we recognize outer functions? in other words what are their distinctive analytic properties? There are very precise characterizations of outer functions in the scalar case, see e.g. [51] but these formulas do not generalize in a simple way to the vector case and we shall not report them here. As we shall see below some important analytic properties of outer functions can be derived directly from the geometric definition without dwelling too much into complex variable theory.

### Inner functions

Scalar inner functions have been described and classified completely in the literature [11, 51]. It can be shown that a real scalar function  $Q$  ( $\overline{Q(z)} = Q(z^{-1})$ ) is inner if and only if it is of the form  $Q(z) = cB(z)S(z)$  where  $c$  is a constant of modulus one; i.e.  $c = \pm 1$ ,  $B(z)$  is a Blaschke product, namely <sup>18</sup>

$$B(z) = \prod_{k=1}^{+\infty} \frac{1 - \alpha_k z}{z - \bar{\alpha}_k}, \quad |\alpha_k| < 1. \tag{4.6.9}$$

<sup>18</sup>Recall that a real analytic function has poles and zeros which come in conjugate pairs; i.e.  $\alpha_k$  is a pole (or a zero) if and only if the conjugate  $\bar{\alpha}_k$  is also a pole (or a zero). For this reason, either at the numerator or at the denominator of the expression (4.6.9)  $\alpha_k$  may be replaced by the conjugate  $\bar{\alpha}_k$ . Also, in this case there is no need to introduce the convergence factors  $\bar{\alpha}_k/|\alpha_k|$  which are needed in the general case [51, p. 64].

and  $S(z)$  is a *singular inner function*, which has the general expression

$$S(z) = \exp\left\{-\int_{-\pi}^{\pi} \frac{z + e^{i\theta}}{z - e^{i\theta}} d\mu(e^{i\theta})\right\} \tag{4.6.10}$$

where  $\mu$  is a finite positive measure on the unit circle whose support has Lebesgue measure zero, in other words, a finite positive *singular* measure .

In the expression (4.6.9) the zeros,  $\{1/\alpha_k\}$ , are all in the region of analyticity  $\{|z| > 1\}$ , including possibly the point at infinity and are assumed to appear repeatedly according to their multiplicity. The poles are at the reciprocal (and reciprocal-conjugate) locations in the interior of the unit circle. The conjugate function,  $B(z^{-1})$ , has symmetric properties which actually correspond to the standard way Blaschke functions are introduced in the literature, where the interior of the unit circle plays (contrary to what is done in this book) the role of region of analyticity.

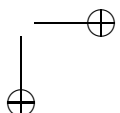
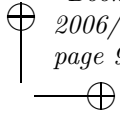
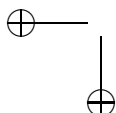
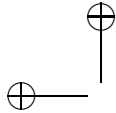
It can be shown that a necessary and sufficient condition for the convergence of the infinite product (4.6.9) in  $\{|z| > 1\}$ , is that the product  $\prod_{k=1}^{+\infty} |\alpha_k|$  (or equivalently  $\prod_{k=1}^{+\infty} 1/|\alpha_k|$ ) converges. This in turn is equivalent to  $\sum_{k=1}^{\infty} (1 - |\alpha_k|) < \infty$  (or  $\sum_{k=1}^{\infty} (1 - 1/|\alpha_k|) < \infty$ ), see e.g. [50, p. 223], which prescribes the possible rate of accumulation of the zeros (equivalently, the poles) of the function  $B(z)$ . In fact, since convergence of the first series occurs only if  $1 - |\alpha_k| \rightarrow 0$ , the accumulation points may only be on the unit circle.

The finite Blaschke products are just the scalar rational inner functions. It should be noted that the analytic expressions given above for Blaschke functions make generally (i.e. except for singularities located in the unit circle) sense also in the region  $\{|z| < 1\}$  of the complex plane. It can be shown (but we shall not go into this here) that all inner functions actually have an analytic continuation across the unit circle, except at the points of the unit circle where the singular component  $S(z)$  is supported or at accumulation points of zeros. In particular, when there is no singular part, one may think of a scalar inner function as an analytic function which is defined and analytic almost everywhere on the complex plane and is determined (modulo a constant unitary factor) by its *zeros*, which are all located in the region  $\{|z| > 1\}$ . For it is clear from the expressions above that assigning a countable set of points in  $\{|z| > 1\}$  obeying the convergence constraint for Blaschke products, to be the zeros of the function, specifies the inner function uniquely up to the constant factor of modulus one.

In the matrix case there are no simple general expressions of this type. However, it is easy to see using Binet Theorem, that the determinant of a matrix inner function must be inner and then one can introduce a classification of matrix inner functions based on the structure of their determinant. For details we refer the reader to Helson's book [45] p. 80-89.

We now proceed to state a criterion for outer functions in terms of zeros.

**Definition 4.6.10.** *Let the matrix function  $F \in H_{m \times p}^2$  have full column rank a.e.. A complex number  $\alpha$  in the region of analyticity  $\{|z| > 1\}$  is a (right) zero of  $F$ , if*





there is a nonzero vector  $v \in \mathbb{C}^p$ , called an associated zero direction to  $\alpha$ , such that,

$$F(\alpha)v = 0. \tag{4.6.11}$$

Dually, in case  $F \in H^2_{m \times p}$  has full row rank a.e., we shall call a complex number  $\alpha$  in the region of analyticity  $\{|z| > 1\}$  is a (left) zero of  $F$ , if there is a nonzero vector  $w \in \mathbb{C}^m$ , called an associated zero direction to  $\alpha$ , such that,

$$v'F(\alpha) = 0. \tag{4.6.12}$$

We shall elaborate further on the notion of zeros in Chapter ?? where we shall restrict to rational matrix functions. Here we shall only mention that the rational of the definition is to consider as zeros only points of the region  $\{|z| > 1\}$  where the rank of  $F$  drops below its generic value (which is  $p$  or  $m$  in the two situations considered). The left- or right- zeros defined above are just the *invariant zeros* which will be introduced in Chapter ??. Note that the zeros of an  $H^2$  function can in general be defined only in the region of analyticity  $\{|z| > 1\}$ , as  $F$  may not have an analytic continuation to the interior of the unit circle,  $\{|z| \leq 1\}$ . However for special subclasses (e.g. rational functions) admitting analytic continuation, the same definition applies to arbitrary complex numbers as well.

Outer functions have full column rank. They can be characterized as analytic functions without (right) zeros.

**Theorem 4.6.11.** *An outer function cannot have zeros in  $\{|z| > 1\}$ , including the point at infinity. In particular, a rational function in  $H^2$  is outer if and only if it has no poles in  $\{|z| \geq 1\}$ , and no zeros in  $\{|z| > 1\}$ , including the point at infinity.*

**Proof.** We shall show that if  $F$  admits a zero; i.e. there is  $\alpha$  outside the unit disk for which (4.6.11) holds, then the invariant subspace  $\mathcal{F} := \overline{\text{span}}\{e^{i\theta t}F_k \mid k = 1, \dots, m, t \leq 0\}$  cannot be the whole of  $H^2_p$ , that is,  $F$  cannot be outer.

Since for any nonsingular  $m \times m$  and  $p \times p$  constant matrices  $T, S$ , the function  $TFS$  spans the same invariant subspace  $\mathcal{F}$ , there is no loss of generality to assume that  $v$  is just the first vector  $e_1 := [1, 0, \dots, 0]'$ , in the canonical basis of  $\mathbb{C}^p$ . Hence we may and shall assume that the elements  $F_{1,j}, j = 1, \dots, p$  all have a zero at  $\alpha$ ; i.e.  $F_{1,j}(\alpha) = 0$ . Since the zeros of a function in  $H^2$  must be isolated and of finite multiplicity [51], all  $F_{1,j}$  can be written as

$$F_{1,j}(z) = \hat{F}_{1,j}(z) \left( \frac{z - \alpha}{1 - \bar{\alpha}z} \right)^{k_j},$$

where  $\hat{F}_{1,j}(z) \in H^2$ ,  $k_j \geq 1$  are the multiplicities of the zero, and  $\hat{F}_{1,j}(z)(\alpha) \neq 0$ . It follows that the first component  $f_1$ , of any function in  $\mathcal{F}$  must have a zero of (at least) multiplicity  $k := M.C.D.\{k_j\}$  at  $\alpha$ . Letting  $Q(z)$  be the maximum common divisor of all elementary Blaschke factors on the right, all such first components can be written as

$$f_1(z) = h(z)Q(z), \quad h \in H^2.$$

Now, for any nontrivial inner function  $Q$ , there are functions  $g \in H^2$  such that  $\bar{g}(z) := Q^*(z)g(z)$  is in  $\bar{H}^2$ . These functions fill in fact the orthogonal complement  $(H^2Q)^\perp$  in  $H^2$ . For instance,

$$g(z) = \frac{1}{(1 - \bar{a}z)^k}$$

is one such function. Consider now the subspace  $\mathcal{G}$  of  $H_p^2$  made of vector functions of the form  $[g \ 0 \ \dots \ 0]$ ;  $g \in (H^2Q)^\perp$ . Since for any  $f = [f_1 \ f_2 \ \dots \ f_p] \in \mathcal{F}$  we have  $\langle f_1, g \rangle_{H^2} = \langle h, Q^*g \rangle_{H^2} = 0, j = 1, \dots, m$  we have  $\mathcal{G} \perp \mathcal{F}$ .  $\square$

Hence, a scalar rational outer function has no poles in  $\{|z| \geq 1\}$ , no zeros in  $\{|z| > 1\}$  and equal degrees of the numerator and denominator polynomials. These functions are called *minimum phase* in the engineering literature.

### 4.7 Bibliographical notes

The early references on prediction theory are [64] [121]. The idea of whitening filter and the cascade structure of the filter appear in the paper [12]. Spectral factorization as a tool for solving filtering and prediction problems was introduced by Wiener [121, 119, 120].

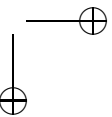
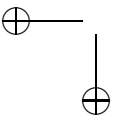
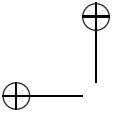
Orthonormalizable processes are called *processes of constant rank* in [106]. Since every Hermitian positive semidefinite matrix  $H$  admits square roots, i.e. matrices  $W$  that satisfy  $H = WW^*$ , it is sometimes stated in the literature that  $y$  is orthonormalizable (in our terminology) if and only if its spectral distribution function is absolutely continuous and the relative spectral density matrix  $\Phi$  has constant rank a.e. on  $[-\pi, \pi]$ . This somewhat simpler statement requires however a proof that the square-root matrices of  $\Phi$ , (which unfortunately *cannot be defined pointwise*) can be suitably chosen for each  $\theta$  and patched together so as to form (at least) a measurable matrix function  $W$ . The seemingly more restrictive factorizability condition in our Theorem 4.2.1 avoids these annoying technicalities.

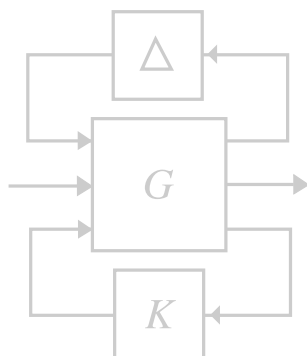
Wold's representation theorem in terms of wandering subspaces was of course first introduced in the seminal work of H. Wold [123] on stationary processes and prediction theory. Wold's ideas have been generalized in many directions. Starting with [43, 89], generalizations of "Wold Decomposition" theory have become part of functional analysis and have led to a unifying view of certain fundamental problems in operator theory and Hardy spaces [45]. The basic operator-theoretic (and Hardy space) results which have stemmed from this idea can, as it is shown in Section 3.5, simply be seen as isomorphic function-analytic counterparts of the geometric Hilbert space results exposed in Section 3.4. Several excellent books have been written on Hardy spaces, some of which are classical reference works. We shall just mention here [51, 45, 26, 36]. The Paley-Wiener Theorem is in [99].

The Beurling-Lax Theorem on invariant subspaces first appeared in [11] (for the scalar case) and was then generalized to vectorial functions in [68].

Much research has been devoted in the 1950s and 1960s to understand the function-theoretic properties which characterize the spectral density matrix of vec-

torial p.n.d. processes, and the related “causal” spectral-factorization problem. Among the basic references we quote [67, 119, 46, 106].





## Chapter 5

# Wold Decomposition, and Spectral Factorization in Continuous Time

In this chapter we shall describe the continuous-time analogs of the ideas and representation results of the previous chapter. As discussed before, the interesting generalization of the discrete-time setting is to continuous-time *stationary increments processes*. For this reason we shall be mostly concerned with this class.

### 5.1 Stationary increments processes and the continuous-time Wold decomposition

Let  $\{y(t); t \in \mathbb{R}\}$  be an  $m$ -dimensional mean square continuous process with stationary increments,  $\mathbf{H}(dy)$  the Hilbert space generated by its increments and let  $\{U_t\}$  be the associated strongly continuous unitary group in  $\mathbf{H}(dy)$ .

In general, given any subspace  $\mathbf{K}$  of  $\mathbf{H}(dy)$ , we define the stationary family of translates  $\{\mathbf{K}_t\}$ , of  $\mathbf{K}$ , by setting  $\mathbf{K}_t := U_t\mathbf{K}$ ,  $t \in \mathbb{R}$  and introduce the *past* and *future* (at the time zero) of the family  $\{\mathbf{K}_t\}$  by

$$\mathbf{K}^- := \vee_{t \leq 0} \mathbf{K}_t, \quad \mathbf{K}^+ := \vee_{t \geq 0} \mathbf{K}_t \tag{5.1.1}$$

where the symbol  $\vee$  denotes closed vector sum. Clearly,  $\mathbf{K}_t^- := U_t\mathbf{K}^-$  and  $\mathbf{K}_t^+ := U_t\mathbf{K}^+$  form an increasing, respectively, a decreasing family of subspaces of  $\mathbf{H}(dy)$ .

Subspaces  $\mathbf{K}$  for which  $\mathbf{K}_t = \mathbf{K}_t^-$  or  $\mathbf{K}_t = \mathbf{K}_t^+$  can be characterized in the following way. Introduce the *forward* and *backward shift semigroups*  $\{U_t; t \geq 0\}$  and  $\{U_t^*; t \geq 0\}$  acting on  $\mathbf{H}(dy)$ , where  $U_t$  is the shift induced by  $dy$ , defined in (2.7.3). It is then easy to check that a subspace  $\mathbf{K}$  generates an increasing stationary family of translates  $\{\mathbf{K}_t\}$  if and only if it is *backward-shift invariant*; i.e.,

$$U_t^* \mathbf{K} \subset \mathbf{K} \quad \text{for all } t \geq 0. \tag{5.1.2}$$

Similarly,  $\mathbf{K}$  generates a decreasing family of translates  $\{\mathbf{K}_t\}$  if and only if,

$$U_t \mathbf{K} \subset \mathbf{K} \quad \text{for all } t \geq 0 \tag{5.1.3}$$

i.e.  $\mathbf{K}$  is a *forward shift invariant* subspaces. A subspace satisfying both conditions (5.1.2), (5.1.3) is called *doubly invariant*.

In analogy with the discrete-time setting, we shall say that an increasing family  $\{\mathbf{K}_t\}$  is *purely nondeterministic* (p.n.d) if the “remote past”  $\mathbf{K}_{-\infty} := \bigcap_{t \in \mathbb{R}} \mathbf{K}_t$  contains only the zero random variable. The property of being p.n.d. depends on the structure of the backward shift invariant subspace  $\mathbf{K}$  alone. Dually, for a decreasing family  $\{\mathbf{K}_t\}$  in  $\mathbf{H}(dy)$ , define the “remote future”  $\bar{\mathbf{K}}_{\infty} := \bigcap_{t \in \mathbb{R}} \bar{\mathbf{K}}_t$ . If  $\bar{\mathbf{K}}_{\infty}$  is trivial we say that  $\{\bar{\mathbf{K}}_t\}$  is p.n.d. or that  $\bar{\mathbf{K}}$  is a p.n.d. (forward shift) invariant subspace. A stationary-increment process  $dy$  will be called (*forward*) *p.n.d.* whenever  $\mathbf{H}^-(dy)$  is p.n.d. and *backward p.n.d.* when  $\mathbf{H}^+(dy)$  is p.n.d..

The following representation theorem is the continuous-time version of the Wold representation theorem (Theorem 4.5.4).

**Theorem 5.1.1.** *A necessary and sufficient condition for a subspace  $\mathbf{S} \subset \mathbf{H}(dy)$  to be backward shift-invariant and p.n.d. is that there is a vector Wiener process  $dw$  such that*

$$\mathbf{S} = \mathbf{H}^-(dw) \tag{5.1.4}$$

*Similarly, a necessary and sufficient condition for a subspace  $\bar{\mathbf{S}} \subset \mathbf{H}(dy)$  to be forward shift-invariant and p.n.d. is that there is a vector Wiener process  $d\bar{w}$  such that*

$$\bar{\mathbf{S}} = \mathbf{H}^+(d\bar{w}) \tag{5.1.5}$$

*Both  $dw$  and  $d\bar{w}$  are uniquely determined by  $\mathbf{S}$  and  $\bar{\mathbf{S}}$  modulo multiplication by a constant orthogonal matrix. The dimension of  $dw$  is called the multiplicity of  $\mathbf{S}$  or  $\mathbf{H}(dw)$  and the dimension of  $d\bar{w}$  the multiplicity of  $\bar{\mathbf{S}}$  or of  $\mathbf{H}(d\bar{w})$*

A proof can be obtained from the discrete-time result by the so-called *Cayley transform*. For details see [69, 90].

Note that whenever  $\bigvee_{t \in \mathbb{R}} \mathbf{S}_t = \mathbf{H}(dy)$ , in which case  $\mathbf{S}$  is said to be of *full range*, we have a representation of the space  $\mathbf{H}(dy)$  as

$$\mathbf{H}(dy) = \mathbf{H}(dw) \tag{5.1.6}$$

An analogous representation of  $\mathbf{H}(dy)$  is obtained in the case  $\bar{\mathbf{S}}$  is full range.

## 5.2 Hardy spaces of the half-plane

A similar construction to the one described in Section 4.3, but starting from the familiar Lebesgue space  $L_p^2(\mathbb{R})$  of (equivalence classes of) square integrable functions on the real line leads to the *Hardy spaces of the half-plane*.

**Definition 5.2.1.** *The Hardy space of the half-plane, denoted  $H_p^2(\mathbb{C}_+)$  (or  $H_p^2$  for short, when there is no danger of confusion), consists of  $p$ -dimensional vector functions analytic on the right half of the complex plane, having the property that the family of maps  $\{i\omega \rightarrow f(\sigma + i\omega); \sigma > 0\}$  is uniformly bounded in the  $L_p^2(\mathbb{I})$  norm. Dually, the conjugate Hardy space of the half-plane, denoted  $\bar{H}_p^2(\mathbb{C}_+)$  (or*

$\bar{H}_p^2$  for short) consists of  $p$ -dimensional vector functions analytic on the left half of the complex plane, having the property that  $\{i\omega \rightarrow f(\sigma + i\omega); \sigma < 0\}$ , is uniformly bounded in the  $L_p^2(\mathbb{I})$  norm.

It can be shown, see e.g. [51, p.128], that the functions in  $H_p^2(\mathbb{C}_+)$  ( $\bar{H}_p^2(\mathbb{C}_+)$ ) have nontangential boundary values on the imaginary axis of the complex plane, convergence taking place both in the  $L_p^2(\mathbb{I})$  norm and almost everywhere. Moreover they can be uniquely recovered from their boundary values (belonging to  $L_p^2(\mathbb{I})$ ). By introducing a proper definition of norm, the correspondence between analytic functions and their boundary values can actually be made unitary so one does not need to distinguish between the two classes. This convention we shall follow also in this book, so functions on the imaginary axis which are boundary values of functions in the Hardy space  $H_p^2(\mathbb{C}_+)$  (resp.  $\bar{H}_p^2(\mathbb{C}_+)$ ) will be called *analytic* (resp. *co-analytic*).

The (continuous-time) Paley-Wiener theorem describes exactly which functions on the imaginary axis are boundary values of functions in the Hardy space  $H_p^2(\mathbb{C}_+)$  (resp.  $\bar{H}_p^2(\mathbb{C}_+)$ ). Recall that functions  $f \in L_p^2(\mathbb{R})$  which vanish a.e. on the negative (positive) axis are called *causal* (*anticausal*). The causal and anticausal functions form complementary Hilbert subspaces of  $L_p^2(\mathbb{R})$  which are denoted  $L_p^{2+}(\mathbb{R})$  and  $L_p^{2-}(\mathbb{R})$  respectively. As it has been recalled in Chapter ??, the Fourier operator  $\mathfrak{F}$  maps  $L_p^2(\mathbb{R})$  unitarily onto  $L_p^2(\mathbb{I}, \frac{d\omega}{2\pi})$  (written as  $L_p^2(\mathbb{I})$  for short).

**Theorem 5.2.2 (Paley-Wiener).** *With the conventions established above, the Hardy space  $H_p^2(\mathbb{C}_+)$  is the image of the subspace  $L_p^{2+}(\mathbb{R})$  of causal  $p$ -dimensional functions under the Fourier map,*

$$\mathfrak{F}(L_p^{2+}) = H_p^2(\mathbb{C}_+). \tag{5.2.1}$$

*Dually, the conjugate Hardy space of the half-plane,  $\bar{H}_p^2(\mathbb{C}_+)$ , is the image of the subspace  $L_p^{2-}(\mathbb{R})$  of anticausal  $p$ -dimensional functions under the Fourier map,*

$$\mathfrak{F}(L_p^{2-}) = \bar{H}_p^2(\mathbb{C}_+). \tag{5.2.2}$$

$H_p^2(\mathbb{C}_+)$  and  $\bar{H}_p^2(\mathbb{C}_+)$  are orthogonal complementary subspaces of  $L_p^2(\mathbb{I})$ .

Hardy spaces of the disk and of the half-plane are commonly regarded, in a sense, as isomorphic objects. This however is not quite correct. There is a wider class of analytic functions on the half plane which (in a very precise sense to be defined later) is the continuous-time analog of the Hardy class of the disk. These functions arise in connection with the spectral representation of stationary increments processes and for this reason we shall have to study their properties in some detail. They are constructed in the following way.

If we map the exterior of the unit disk  $\mathbb{D}$  onto the right half plane  $\mathbb{C}_+$  (and hence the interior of the unit disk onto the left-half plane) by the conformal transformation,

$$s = \rho := \frac{z + 1}{z - 1} \tag{5.2.3}$$

and define the corresponding mapping  $T_\rho$  acting on functions, by

$$T_\rho f(s) = f(z)|_{z=\rho^{-1}(s)} = f\left(\frac{s+1}{s-1}\right), \quad (5.2.4)$$

then  $H_p^2(\mathbb{D})$  is mapped onto a space of analytic functions in right half plane  $\mathbb{C}_+$  which we shall name  $\mathcal{W}_p^2$ . Similarly, the transformation  $T_\rho$  takes any function  $f \in \bar{H}_p^2(\mathbb{D})$  into a “conjugate space” of functions,  $\bar{\mathcal{W}}_p^2$  analytic in  $\mathbb{C}_- = \{\Re s < 0\}$ . The following characterization of  $\mathcal{W}_p^2$  and  $\bar{\mathcal{W}}_p^2$  can be obtained by an easy generalization of similar results presented for the scalar case in [51, p.128-130].

**Theorem 5.2.3.** *The spaces  $\mathcal{W}_p^2$  and  $\bar{\mathcal{W}}_p^2$ , are described by*

$$\mathcal{W}_p^2 = \{f : f = (1+s)\hat{f} \mid \hat{f} \in H_p^2(\mathbb{C}_+)\} \equiv (1+s)H_p^2(\mathbb{C}_+) \quad (5.2.5)$$

$$\bar{\mathcal{W}}_p^2 = \{f : f = (1-s)\hat{f} \mid \hat{f} \in \bar{H}_p^2(\mathbb{C}_+)\} \equiv (1-s)\bar{H}_p^2(\mathbb{C}_+) \quad (5.2.6)$$

Every function in  $\mathcal{W}_p^2$  ( $\bar{\mathcal{W}}_p^2$ ) has nontangential boundary values (a.e.) on the imaginary axis belonging to the space

$$\mathcal{L}_p^2 := L_p^2 \left[ \mathbb{I}, \frac{d\omega}{\pi(1+\omega^2)} \right] \quad (5.2.7)$$

from which it can be uniquely recovered. The map  $T_\rho$  defined in (5.2.4) is an isometry of  $L_p^2$  of the unit circle onto  $\mathcal{L}_p^2$  under which  $H_p^2(\mathbb{D})$  is mapped onto  $\mathcal{W}_p^2$  and  $\bar{H}_p^2(\mathbb{D})$  onto  $\bar{\mathcal{W}}_p^2$ .

The Hardy spaces of the half plane  $H_p^2(\mathbb{C}_+)$  and  $\bar{H}_p^2(\mathbb{C}_+)$  are properly contained in  $\mathcal{W}_p^2$  and  $\bar{\mathcal{W}}_p^2$  (respectively) and correspond, under  $T_\rho^{-1}$  to the subspaces,

$$T_\rho^{-1}H_p^2(\mathbb{C}_+) = \{f : f \in H_p^2(\mathbb{D}) \mid \frac{z}{z-1}f(z) \in H_p^2(\mathbb{D})\} \quad (5.2.8)$$

$$T_\rho^{-1}\bar{H}_p^2(\mathbb{C}_+) = \{f : f \in \bar{H}_p^2(\mathbb{D}) \mid \frac{1}{z-1}f(z) \in \bar{H}_p^2(\mathbb{D})\}. \quad (5.2.9)$$

Let us introduce the forward and backward difference operators of width  $h > 0$  in  $\mathcal{L}_p^2$ , as the multiplication operators by the functions  $\chi_h(i\omega) := \frac{e^{i\omega h}-1}{i\omega}$  and  $\bar{\chi}_h(i\omega) := \chi_h(-i\omega)$  respectively. Note that, since  $\chi_h$  (resp.  $\bar{\chi}_h$ ) is the Fourier transform of the anticausal indicator function  $I_{[-h,0]}$  (respectively,  $\bar{\chi}_h$  of the causal indicator function  $I_{[0,h]}$ ), it has a bounded analytic continuation to  $\Re s < 0$  (to  $\Re s > 0$ ). In fact,  $\bar{\chi}_h$  belongs to the scalar Hardy space  $H^\infty$  of uniformly bounded analytic functions in  $\Re s > 0$  and  $\chi_h$  belongs to the conjugate Hardy space  $\bar{H}^\infty$ .

**Lemma 5.2.4.** *The following alternative characterizations of  $\mathcal{W}_p^2$  and  $\bar{\mathcal{W}}_p^2$  hold.*

- An element  $f \in \mathcal{L}_p^2$  is in  $\mathcal{W}_p^2$  if and only if  $\bar{\chi}_h f$  belongs to  $H_p^2(\mathbb{C}_+)$  for all  $h > 0$ .



- Dually,  $f \in \mathcal{L}_p^2$  is in  $\bar{W}_p^2$  if and only if  $\chi_h f$  belongs to the conjugate Hardy space  $\bar{H}_p^2(\mathbb{C}_+)$  for all  $h > 0$ .

**Proof.**

We shall prove only the first statement as the dual follows by symmetry. Let  $f \in \mathcal{W}_p^2$  be written as  $f(i\omega) = (1 + i\omega)g(i\omega)$  with  $g \in H_p^2(\mathbb{C}_+)$  (see (5.2.5) above). Since for all  $h > 0$ ,  $\bar{\chi}_h \in H^\infty$ , and all terms in the right hand side of

$$\bar{\chi}_h f = \bar{\chi}_h g - e^{-i\omega h} g + g$$

are in  $H_p^2(\mathbb{C}_+)$ , we have  $\bar{\chi}_h f \in H_p^2(\mathbb{C}_+)$  for all  $h > 0$ .

Conversely, assume that  $\bar{\chi}_h f \in H_p^2(\mathbb{C}_+)$  for all  $h > 0$ . We shall show that this implies  $f(s)/(1 + s) \in H_p^2$  (equivalently  $f(s) = (1 + s)g(s)$  for some  $g \in H_p^2$ ) and this will lead to the desired conclusion by (5.2.5), Theorem 5.2.3. To this end we shall use the Laplace transform formula:

$$\frac{1}{1 + s} = - \int_0^\infty \frac{e^{-st} - 1}{s} e^{-t} dt, \quad \Re s > -1.$$

which, multiplying both sides by  $f$  and introducing the Borel measure  $dm(t) := e^{-t} dt$  on  $\mathbb{R}_+$ , yields,

$$\frac{1}{1 + i\omega} f(i\omega) = - \int_0^\infty \bar{\chi}_t(i\omega) f(i\omega) dm(t).$$

By assumption the map  $\gamma_t : t \rightarrow \bar{\chi}_t f$  takes values in  $H_p^2$ . If we can make sense of the integral in the right side as a vector-valued integral in  $H_p^2$  (see [127, p. 132]) it will automatically follow that the first member is also in  $H_p^2$  and the lemma will be proven. A sufficient condition for this to be the case is that  $\gamma_t$  be (strongly) continuous and that

$$\int_0^\infty \|\gamma_t\|^2 dm(t) < \infty$$

the norm being that of  $H_p^2$ . Now, let  $d\mu(\omega) := d\omega/(1 + \omega^2)$ , and consider the identity

$$\|\gamma_t\|^2 = \int_{-\infty}^{+\infty} |\bar{\chi}_t(i\omega)|^2 |f(i\omega)|^2 d\omega = \int_{-\infty}^{+\infty} \frac{\sin^2(\omega t/2)}{(\omega/2)^2} (1 + \omega^2) |f(i\omega)|^2 d\mu(\omega)$$

where  $|f(i\omega)|^2$  is the Euclidean norm of the vector  $f(i\omega)$ . In view of the inequality

$$\frac{\sin^2(\omega t/2)}{(\omega/2)^2} (1 + \omega^2) = \frac{\sin^2(\omega t/2)}{(\omega t/2)^2} t^2 + 4 \sin^2(\omega t/2) \leq t^2 + 4,$$

and the fact that  $f \in \mathcal{L}_p^2$ , we see, by dominated convergence, that  $\|\gamma_t\|^2 \rightarrow 0$  as  $t \downarrow 0$ . Note however that for  $t_1 \geq t_2$  we have  $\|\gamma_{t_1} - \gamma_{t_2}\|^2 = \|\bar{\chi}_{t_1 - t_2} f\|^2 = \|\gamma_{t_1 - t_2}\|^2$ . It is easy to check that the same is actually true also for  $0 \leq t_1 < t_2$ , so that  $\gamma_{t_1} \rightarrow \gamma_{t_2}$  strongly as  $t_1 \rightarrow t_2$ . This proves continuity. To see that the integral of the squared norm is finite, just notice that we have the bound

$$\|\gamma_t\|^2 \leq (t^2 + 4) \|f\|_{\mathcal{L}_p^2}.$$

This concludes the proof.  $\square$

Using this lemma it is easy to derive a generalization of the Paley-Wiener criterion to the spaces  $\mathcal{W}_p^2$  and  $\bar{\mathcal{W}}_p^2$ .

**Theorem 5.2.5.** *The space  $\mathcal{W}_p^2$  consists precisely of those functions in  $\mathcal{L}_p^2$  for which,*

$$\int_{-\infty}^{+\infty} \frac{e^{i\omega t} - e^{i\omega s}}{i\omega} f(i\omega) d\omega = 0 \quad \text{for all } t, s < 0. \quad (5.2.10)$$

Dually,  $\bar{\mathcal{W}}_p^2$  consists of those functions in  $\mathcal{L}_p^2$  for which,

$$\int_{-\infty}^{+\infty} \frac{e^{i\omega t} - e^{i\omega s}}{i\omega} f(i\omega) d\omega = 0 \quad \text{for all } t, s > 0. \quad (5.2.11)$$

The intersection  $\mathcal{W}_p^2 \cap \bar{\mathcal{W}}_p^2$  contains only the constant vector functions and is isomorphic to  $\mathbb{R}^p$ .

**Proof.** As  $f \in \mathcal{W}_p^2$  if and only if  $\bar{\chi}_h f$  belongs to  $H_p^2(\mathbb{C}_+)$  for all  $h > 0$  (Lemma 5.2.4) and this in turn happens (by Paley Wiener), if and only if

$$\int_{-\infty}^{+\infty} e^{i\omega t} \bar{\chi}_h(i\omega) f(i\omega) d\omega = 0 \quad \text{for all } t < 0,$$

i.e., if and only if

$$\int_{-\infty}^{+\infty} \frac{e^{i\omega(t-h)} - e^{i\omega t}}{i\omega} f(i\omega) d\omega = 0 \quad \text{for all } t < 0, h > 0,$$

we see that (5.2.10) is indeed equivalent to  $f \in \mathcal{W}_p^2$ . That  $\mathcal{W}_p^2 \cap \bar{\mathcal{W}}_p^2 \equiv \mathbb{R}^p$  follows from the isomorphism with  $H^2$  spaces of the unit disk stated in Theorem 5.2.3.  $\square$

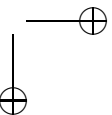
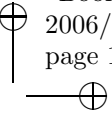
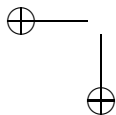
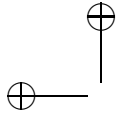
### 5.3 Analytic spectral factorization in continuous time

The subspaces  $\mathbf{S}$  and  $\bar{\mathbf{S}}$ , defined by (5.1.4) and (5.1.5), consist of random variables with stochastic-integral representations of the type (3.5.4) in which, in the case of  $\mathbf{S}$ ,  $f$  is a *casual* function in  $L_p^2(\mathbb{R})$ , i.e.  $f(t) = 0$  a.e. for  $t < 0$  or, in case of  $\bar{\mathbf{S}}$ , an *anticausal* function, for which  $f(t) = 0$  a.e. for  $t > 0$ . Causal and anticausal functions form orthogonal complementary subspaces of  $L_p^2$ . From this it follows that the subspaces  $\mathbf{S}$  and  $\bar{\mathbf{S}}$  in (5.1.4), (5.1.5) naturally correspond to the Hardy spaces  $H_p^2$  and  $\bar{H}_p^2$  of the half plane, under the appropriate representation maps, namely,

$$\mathbf{S} = \mathbf{H}^-(d\omega) = \mathcal{J}_{\bar{\omega}} H_p^2, \quad \bar{\mathbf{S}} = \mathbf{H}^+(d\bar{\omega}) = \mathcal{J}_{\omega} H_{\bar{p}}^2 \quad (5.3.1)$$

where  $p$  and  $\bar{p}$  are the respective multiplicities.

Assume now that the stationary-increment process  $dy$  is purely non-deterministic in both the forward and the backward direction; see Section 5.1. Then, by



Theorem 5.1.1 applied to the subspaces  $\mathbf{S} = \mathbf{H}^-(dy)$  and  $\bar{\mathbf{S}} = \mathbf{H}^+(dy)$ , there are two Wiener processes, which throughout this book are denoted  $dw_-$  and  $d\bar{w}_+$ , called the *forward* and, respectively, *backward innovation processes* of  $dy$ , such that  $\mathbf{H}^-(dy) = \mathbf{H}^-(dw_-)$  and  $\mathbf{H}^+(dy) = \mathbf{H}^+(d\bar{w}_+)$ . Note that this implies that  $\mathbf{H}(dw_-) = \mathbf{H}(dy)$  so that the two Wiener processes have the same dimensions  $p$ , which is called the *multiplicity*, or *rank*, of the process  $dy$ . (A stationary increments process is *full rank* if its multiplicity equals its dimensions).

Now, for any  $h > 0$ ,  $y(-h) - y(0) \in H^-(dw_-)$ , and  $y(h) - y(0) \in \mathbf{H}^+(d\bar{w}_+)$ , so that there are  $m \times p$  analytic and coanalytic matrix functions,  $W_h$ , respectively,  $\bar{W}_h$  (with rows in  $H_p^2$  and, respectively,  $\bar{H}_p^2$ ) such that,

$$y(-h) - y(0) = \int_{-\infty}^{+\infty} W_h(i\omega) d\hat{w}_-(i\omega) \tag{5.3.2}$$

and

$$y(h) - y(0) = \int_{-\infty}^{+\infty} \bar{W}_h(i\omega) d\hat{w}_+(i\omega) \tag{5.3.3}$$

where  $d\hat{w}_-$ ,  $d\hat{w}_+$  are the spectral measures of  $dw_-$ ,  $d\bar{w}_+$  [compare (3.6.2)]. Using the difference operators  $\chi_h$ ,  $\bar{\chi}_h$  we can rewrite (5.3.2), (5.3.3) in terms of the new functions

$$W_- := \bar{\chi}_h^{-1} W_h \tag{5.3.4}$$

$$\bar{W}_+ := \chi_h^{-1} \bar{W}_h \tag{5.3.5}$$

Once we rewrite the integral representations (5.3.2), (5.3.3) in terms of (5.3.4), (5.3.5), it becomes evident, by comparison with the spectral representation (3.6.6), that

$$d\hat{y} = W_- d\hat{w}_- = \bar{W}_+ d\hat{w}_+ \tag{5.3.6}$$

the relations holding by uniqueness of the spectral measure  $d\hat{y}$ . From this it is easily seen that  $W_-$  and  $\bar{W}_+$  do not depend on  $h$  and, by Lemma 5.2.4,

$$W_- \in \mathcal{W}_p^2, \quad \bar{W}_+ \in \bar{\mathcal{W}}_p^2 \tag{5.3.7}$$

We have therefore proven the following statement.

**Proposition 5.3.1.** *The spectral distribution  $dF$  of a purely nondeterministic (both in the forward and in the backward sense) stationary increments processes must be absolutely continuous with a (matrix) spectral density  $\Phi := dF/d(\omega/2\pi)$  satisfying*

$$\Phi(i\omega) = W_-(i\omega)W_-(i\omega)^* = \bar{W}_+(i\omega)\bar{W}_+(i\omega)^* \tag{5.3.8}$$

*almost everywhere on the imaginary axis. The matrix functions,  $W_-$  and  $\bar{W}_+$ , belong to the spaces  $\mathcal{W}_p^2$  and  $\bar{\mathcal{W}}_p^2$ , and are analytic and co-analytic full-rank spectral factors of  $\Phi$ .*

### Outer spectral factors in $\mathcal{W}^2$

The  $m \times p$  functions  $W_-$  and  $\bar{W}_+$  are rather special solutions of the spectral factorization equation,

$$\Phi(i\omega) = W(i\omega)W(i\omega)^*. \tag{5.3.9}$$

In fact, we shall see that  $W_-$  and  $\bar{W}_+$  are the unique (mod  $\mathcal{O}$ ) *outer* and *conjugate outer* spectral factors of  $\Phi$ . That  $W_- \in \mathcal{W}_p^2$  and deserves to be called *outer* and, similarly,  $\bar{W}_+ \in \bar{\mathcal{W}}_p^2$  *conjugate outer*, follows readily from the identities  $\mathbf{H}^-(dy) = \mathbf{H}^-(dw_-)$  and  $\mathbf{H}^+(dy) = \mathbf{H}^+(d\bar{w}_+)$ , the spectral representations (5.3.2), (5.3.3) and the definitions (5.3.4), (5.3.5), which imply that,

$$\overline{\text{span}}\{\bar{\chi}_h W_- ; h > 0\} = H_p^2, \tag{5.3.10a}$$

$$\overline{\text{span}}\{\chi_h \bar{W}_+ ; h > 0\} = \bar{H}_p^2. \tag{5.3.10b}$$

We shall take (5.3.10a) and (5.3.10b) as the defining properties of outer and conjugate outer functions in  $\mathcal{W}_p^2$  and  $\bar{\mathcal{W}}_p^2$ .

The following theorem generalizes the inner-outer factorization theorem ?? to  $\mathcal{W}^2$  spaces.

**Theorem 5.3.2.** *Every matrix function  $F \in \mathcal{W}_{m \times p}^2$  of full column rank a.e., has a factorization  $F = F_- Q$  where  $F_-$  is an outer function in  $\mathcal{W}_{m \times p}^2$  and  $Q$  is inner  $p \times p$ . In this factorization  $F_-$  and  $Q$  are unique up to  $p \times p$  constant orthogonal factors.*

**Proof.** It is immediate to check that  $\overline{\text{span}}\{\bar{\chi}_h F_k ; h > 0, k = 1, 2, \dots, m\}$  is a subspace of  $H_p^2$  which is invariant for the operators of multiplication by  $i\omega \mapsto e^{i\omega t}; t \leq 0$ . Hence there is an (essentially unique) inner function  $Q$  such that

$$\overline{\text{span}}\{\bar{\chi}_h F_k ; h > 0, k = 1, 2, \dots, m\} = H_p^2 Q.$$

Each function  $\bar{\chi}_h F_k, h > 0$ , has therefore a representation  $\bar{\chi}_h F_k = G_{k,h} Q$  with  $G_{k,h} \in H_p^2$ . Now  $\bar{\chi}_h^{-1} G_{k,h}$  is in  $\mathcal{W}_p^2$  and is clearly independent of  $h$  (for it is equal to  $F_k Q^*$ ) so we can rename it just  $G_k$ . Now we claim that the matrix  $G$  with rows  $G_k$  constructed above, must be outer; i.e. it must be true that,

$$\overline{\text{span}}\{\bar{\chi}_h G_k ; h > 0, k = 1, 2, \dots, m\} = H_p^2.$$

For otherwise this would instead be a proper invariant subspace of the form  $H_p^2 R$  for some nontrivial inner function  $R$ . In this circumstance however we would get

$$\overline{\text{span}}\{\bar{\chi}_h F_k ; h > 0, k = 1, 2, \dots, m\} = \overline{\text{span}}\{\bar{\chi}_h G_k Q ; h > 0, k = 1, 2, \dots, m\} = H_p^2 R Q,$$

which contradicts uniqueness of the representation of the invariant subspace  $\overline{\text{span}}\{\bar{\chi}_h F_k ; h > 0, k = 1, 2, \dots, m\}$ .  $\square$

There is a generalization to  $\mathcal{W}^2$  spaces of the factorization theorem in case of non full rank functions. We shall report it below without proof.

**Theorem 5.3.3.** *Every matrix function  $F \in \mathcal{W}_{m \times r}^2$  of rank  $p \leq r$  a.e., has a factorization  $F = F_- R$  where  $F_-$  is outer  $m \times p$  and  $R$  is a unilateral inner function of dimension  $p \times r$ . In this factorization  $F_-$  is unique up to  $p \times p$  right constant unitary factors. The factor  $R$  is unique modulo multiplication by orthogonal matrices only if  $p = m$  (in which case  $F_-$  is square).*

*A totally analogous factorization holds for coanalytic matrix functions  $\bar{F} \in \bar{\mathcal{W}}_{m \times r}^2$ .*

In perfect analogy to what we saw in section 4.6, these factorization theorems lead to a complete classification of the solutions of the spectral factorization problem for stationary increments processes.

**Theorem 5.3.4.** *Assume  $\Phi$  is the  $m \times m$  spectral density matrix of a stationary increments process of rank  $p$ , admitting analytic spectral factors. Then  $\Phi$  admits an outer spectral factor  $W_-$  in  $\mathcal{W}_{m \times p}^2$ . This is the unique outer factor of  $\Phi$ , modulo right multiplication by a constant  $p \times p$  unitary matrix.*

*Every full-rank analytic spectral factor  $W$  can be written*

$$W = W_- Q, \tag{5.3.11}$$

where  $Q$  is an inner function uniquely determined by  $W \bmod \mathcal{O}$ .

*All other analytic spectral factors of dimension  $m \times r$ ,  $r \geq p$  are of the form*

$$W = W_- R, \tag{5.3.12}$$

where  $R$  is a  $p \times r$  unilateral inner function.

*A completely symmetric result holds for the coanalytic spectral factorization problem  $\Phi(z) = \bar{W}(z)\bar{W}(1/z)'$  with  $\bar{W}_k \in \bar{\mathcal{W}}_p^2$ ,  $k = 1, \dots, m$ .*

The proof of this theorem is the exact analog of the proof of Theorem 4.6.8 and is therefore omitted. We are finally in the position of stating and proving the continuous-time version of the fundamental representation theorem 4.4.1.

**Theorem 5.3.5.** *Let  $dy$  be a mean-square continuous  $m$ -dimensional stationary increments process. Then  $dy$  can be represented as a causal functional of a normalized  $r$ -dimensional ( $r \geq p$ ) Wiener process  $dw$ , or, equivalently, there is a normalized  $r$ -dimensional Wiener process  $dw$  such that*

$$\mathbf{H}_t^-(dy) \subset \mathbf{H}_t^-(dw), \quad t \in \mathbb{Z}, \tag{5.3.13}$$

*only if the spectral distribution function of the process is absolutely continuous and the spectral density  $\Phi$  has  $m \times r$  analytic spectral factors; i.e. only if there are solutions  $W \in \mathcal{W}_r^2$ , of the spectral factorization equation,*

$$\Phi(z) = W(z)W(1/z)'. \tag{5.3.14}$$

*Conversely, if  $\Phi$  admits analytic spectral factors, the process  $dy$  is p.n.d.; i.e. there exist in particular an analytic spectral factor  $W_-$  and a normalized Wiener process*

$dw_-$  (the forward innovation process) such that  $d\hat{y} = W_-d\hat{w}_-$ , for which the inclusion (5.3.13) holds with the equality sign. The analytic spectral factor  $W_-$  is the (unique mod  $\mathcal{O}$ ) outer factor of  $\Phi$ .

Totally symmetric statements hold regarding anticausal representations of  $dy$ . In particular the inclusion,

$$\mathbf{H}_t^+(dy) \subset \mathbf{H}_t^+(d\bar{w}), \quad t \in \mathbb{Z}, \tag{5.3.15}$$

can hold only if the spectral distribution function of the process is absolutely continuous and the spectral density  $\Phi$  has  $m \times r$  coanalytic spectral factors; i.e. only if there are solutions  $\bar{W} \in \bar{W}_r^2$ , of the spectral factorization equation (5.3.14). If  $\Phi$  admits coanalytic spectral factors, the process  $dy$  is p.n.d. in the backward direction; i.e. there exist in particular a coanalytic spectral factor  $\bar{W}_+$  and a normalized Wiener process  $d\bar{w}_+$  (the backward innovation process) such that  $d\hat{y} = \bar{W}_+d\hat{w}_+$ , for which the inclusion (5.3.15) holds with the equality sign. The coanalytic spectral factor  $\bar{W}_+$  is the (unique mod  $\mathcal{O}$ ) conjugate outer factor of  $\Phi$ .

**Proof.** The “only if” part follows by a slight generalization of the argument leading to Proposition 5.3.1. Just replace the past space  $\mathbf{H}^-(dy)$  with any p.n.d.  $\mathbf{S} \supset \mathbf{H}^-(dy)$  and  $\mathbf{H}^+(dy)$  by any  $\bar{\mathbf{S}} \supset \mathbf{H}^+(dy)$ .

The “if” part follows directly from the spectral factorization theorem 5.3.4.  $\square$

In case  $dy$  has a rational spectral density  $\Phi$ , the factorizability condition is automatically satisfied [128].

The theorem applies in particular to mean-square differentiable processes and hence we obtain as a corollary the classical spectral factorization theorem for stationary processes,

**Corollary 5.3.6.** *A continuous stationary process  $y = \{y(t); t \in \mathbb{R}\}$  is purely nondeterministic (in the forward direction) if and only if its spectral distribution  $dF$  is absolutely continuous with a spectral density matrix  $\Phi$  which admits analytic (in  $H_p^2$ ) spectral factors. Likewise, it is purely nondeterministic in the backward direction if and only if its spectral distribution  $dF$  is absolutely continuous with a spectral density matrix  $\Phi$  which admits coanalytic (in  $\bar{H}_p^2$ ) spectral factors.*

## 5.4 Wide sense semimartingales

In this section we shall study the structure of stationary increments processes. As we have seen earlier, a particular instance of a stationary increments process is the indefinite integral of a stationary process; on the opposite extreme there are processes with stationary orthogonal increments which are very irregular and cannot be the integral of anything. We shall show that under a mild regularity condition all m.s. continuous stationary increments processes can be decomposed in the sum of an integrated stationary process plus a process with stationary orthogonal increments. Such decomposition is a particular instance of a *semimartingale decomposition*.

Semimartingales have been studied in depth in the probabilistic literature, see e.g. [53], here however we shall need only simple mean-square versions of the pathwise concepts of the general theory. As these concepts are not related to stationarity, initially we shall not invoke stationarity and deal with general processes which will only assumed to have finite second-order moments.

An  $m$ -dimensional continuous-time process  $\{a(t)\}$  will be called of *finite mean variation* (or simply of finite variation) if, for all bounded intervals  $I$  and finite subdivisions  $\pi := \{t_0 \leq t_1 \leq \dots \leq t_N; t_k \in I\}$ , the supremum

$$\mu(I) := \sup_k \left\{ \sum_k \|a(t_{k+1}) - a(t_k)\| \right\} \tag{5.4.1}$$

is finite. Note that this condition relates to the increments  $a(t) - a(s)$  only and is not affected by adding an arbitrary fixed random vector to  $\{a(t)\}$ . If  $\{a(t)\}$  is of finite variation, then the supremum (5.4.1) on intervals of the form  $(a, b]$  is a finitely additive set function which can be extended to a unique Borel measure  $\mu$  on the real line. Exactly as it happens to real functions, it can be shown that the measure  $\mu$  is non atomic; i.e. has no point masses, if and only if  $\{a(t)\}$  is mean square continuous. The proof is essentially the same as that of [114, Theorem 8.14(c), p. 173] and will not be reported here.

An important fact which characterizes m.s. continuous processes of finite variation is stated in the following lemma.

**Lemma 5.4.1.** *Let  $\{a(t)\}$  be m.s. continuous and of finite variation,  $I$  be any bounded interval of the real line and  $\{\pi_n\}$  any sequence of finite subdivisions of  $I$  such that the mesh  $\Delta(\pi_n) := \max_k |t_{k+1}^n - t_k^n|$  tends to zero as  $n \rightarrow \infty$ . Then*

$$\lim_{\Delta(\pi_n) \rightarrow 0} \sum_k \|a(t_{k+1}) - a(t_k)\|^2 = 0. \tag{5.4.2}$$

**Proof.** Since  $\|a(t_{k+1}) - a(t_k)\| \leq \mu((t_k, t_{k+1}])$  we have

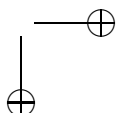
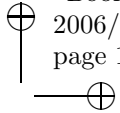
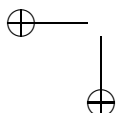
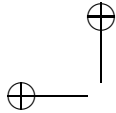
$$\sum_k \|a(t_{k+1}) - a(t_k)\|^2 \leq \sum_k \mu((t_k, t_{k+1}])^2 = \sum_{k=1}^N (\mu \otimes \mu)((t_k, t_{k+1}] \times (t_k, t_{k+1}])$$

where  $\mu \otimes \mu$  is the product measure on  $I \times I$ . As  $\Delta(\pi_n) \rightarrow 0$  the last sum converges to the product measure of the diagonal  $D$  of the square  $I \times I$ . But since  $\mu$  has no point masses  $(\mu \otimes \mu)(D) = 0$ .  $\square$

Let  $\{\mathbf{S}_t\}$  be an increasing family of real zero-mean random variables with the usual inner product. Suppose the  $m$ -dimensional process  $\{y(t)\}$  can be written, for all  $t, s$  in the form,

$$y(t) - y(s) = a(t) - a(s) + m(t) - m(s) \tag{5.4.3}$$

where,



108 Chapter 5. Wold Decomposition, and Spectral Factorization in Continuous Time

1.  $a(t) - a(s) \in \mathbf{S}_t$  for all  $t \geq s$  and  $\{a(t)\}$  is a process of finite mean variation,
2.  $m(t) - m(s) \in \mathbf{S}_t$  for all  $t \geq s$  and  $m(t+h) - m(t) \perp \mathbf{S}_t$  for all  $h \geq 0$ ; i.e.  $\{m(t)\}$  is an  $\mathbf{S}_t$ -martingale,

then we say that  $\{y(t)\}$  has a semimartingale representation relative to the family of subspaces  $\{\mathbf{S}_t\}$ .

**Proposition 5.4.2.** *A representation of the type (5.4.3), relative to a given increasing family of subspaces  $\{\mathbf{S}_t\}$ , is unique.*

**Proof.** Assume that  $\{a_1(t)\}$  and  $\{a_2(t)\}$  both satisfy (i) and  $\{m_1(t)\}$  and  $\{m_2(t)\}$  are  $\mathbf{S}_t$ -martingales for which

$$y(t) - y(s) = a_i(t) - a_i(s) + m_i(t) - m_i(s), \quad i = 1, 2$$

then, setting  $\tilde{a}(t) := a_1(t) - a_2(t)$  and  $\tilde{m}(t) := m_1(t) - m_2(t)$ , we would have  $\tilde{a}(t) - \tilde{a}(s) = -[\tilde{m}(t) - \tilde{m}(s)]$ , with  $\tilde{a}(t)$   $\mathbf{S}_t$ -adapted, continuous and of finite variation and  $\tilde{m}(t)$  an  $\mathbf{S}_t$ -martingale. It follows from Lemma 5.4.1 that for any interval  $[a, b]$  and for any subdivision  $\pi = \{a = t_0 < t_1 < \dots < t_N = b\}$ , the sum

$$\sum_k \|\tilde{a}(t_{k+1}) - \tilde{a}(t_k)\|^2 = \sum_k \|\tilde{m}(t_{k+1}) - \tilde{m}(t_k)\|^2$$

tends to zero as  $\Delta(\pi) \rightarrow 0$ . But, since any martingale has orthogonal increments, the sum on the right is actually equal to  $\|\tilde{m}(b) - \tilde{m}(a)\|^2$  so that  $\tilde{m}(b) = \tilde{m}(a)$  for all  $a, b \in \mathbb{R}$ . This implies that the differences  $m_1(t) - m_1(s)$  and  $m_2(t) - m_2(s)$  are the the same for all  $t, s$ . Hence  $a_1(t) - a_1(s)$  and  $a_2(t) - a_2(s)$  also coincide.  $\square$

The mean quadratic variation of an  $m$ -dimensional process  $\{y(t)\}$ , on the interval  $[s, t]$ , is the  $m \times m$  matrix  $Q(t, s)$  defined by

$$Q_{i,j}(t, s) := \lim_{\Delta(\pi_n) \rightarrow 0} \sum_k \langle y_i(t_{k+1}) - y_i(t_k), y_j(t_{k+1}) - y_j(t_k) \rangle \quad (5.4.4)$$

where  $\{\pi_n\}$  is a sequence of finite subdivisions of the interval  $[s, t]$ . Every martingale has finite quadratic variation on a bounded interval. In fact, since each component  $\{m_i(t)\}$  has orthogonal increments,

$$\begin{aligned} & \sum_k \langle m_i(t_{k+1}) - m_i(t_k), m_j(t_{k+1}) - m_j(t_k) \rangle \\ &= \sum_k \sum_l \langle m_i(t_{k+1}) - m_i(t_k), m_j(t_{l+1}) - m_j(t_l) \rangle \\ &= \langle m_i(t) - m_i(s), m_j(t) - m_j(s) \rangle. \end{aligned}$$

Hence for a martingale we have

$$Q(t, s) = E\{[m(t) - m(s)][m(t) - m(s)]'\}. \quad (5.4.5)$$

**Proposition 5.4.3.** *The mean quadratic variation of a  $\mathbf{S}_t$ -semimartingale coincides with the mean quadratic variation of its martingale part.*



**Proof.** For brevity we shall write differences such as  $z_i(t_{k+1}) - z_i(t_k)$  as  $\Delta z_i(k)$ . The mean quadratic variation of the semimartingale (5.4.3) is

$$\begin{aligned} \sum_k \langle \Delta y_i(k), \Delta y_j(k) \rangle &= \sum_k \langle \Delta a_i(k), \Delta a_j(k) \rangle + \sum_k \langle \Delta a_i(k), \Delta m_j(k) \rangle \\ &+ \sum_k \langle \Delta m_i(k), \Delta a_j(k) \rangle + \sum_k \langle \Delta m_i(k), \Delta m_j(k) \rangle. \end{aligned}$$

Note first that,

$$\begin{aligned} \sum_k \langle \Delta a_i(k), \Delta a_j(k) \rangle &\leq \sum_k \|\Delta a_i(k)\| \|\Delta a_j(k)\| \\ &\leq \frac{1}{2} \sum_k (\|\Delta a_i(k)\|^2 + \|\Delta a_j(k)\|^2) \end{aligned}$$

and the last member tends to zero as  $\Delta(\pi_n) \rightarrow 0$ , by Lemma 5.4.1. Further, since all sums are finite,

$$\begin{aligned} \sum_k \langle \Delta a_i(k), \Delta m_j(k) \rangle &\leq \sum_k \langle \Delta a_i(k), \Delta m_j(k) \rangle \\ &\leq (\sum_k \|\Delta a_i(k)\|^2)^{1/2} (\sum_k \|\Delta m_j(k)\|^2)^{1/2} \\ &= (\sum_k \|\Delta a_i(k)\|^2)^{1/2} \|m_j(t) - m_j(s)\|. \end{aligned}$$

This term also tends to zero as  $\Delta(\pi_n) \rightarrow 0$  and since the same thing happen if the indices  $i$  and  $j$  are interchanged, the result follows.  $\square$

This proposition can be interpreted in the following way: The quadratic variation of a  $\mathbf{S}_t$ -semimartingale is independent of the increasing family of subspaces  $\mathbf{S}_t$ . This is so since, from the way it is defined, the mean quadratic variation of a martingale does not depend on  $\mathbf{S}_t$ . In fact If  $\{y(t)\}$  also admits a semmartingale representation with respect to some *decreasing* family of subspaces  $\bar{\mathbf{S}}_t$  then the mean quadratic variation of  $\{y(t)\}$  is the same as that of any (forward) martingale component of  $\{y(t)\}$ .

### Stationary increments semimartingales

We shall henceforth assume that  $y \equiv \{y(t)\}$  is a process with continuous stationary increments defined on the real line. In the following we shall be concerned with the following question.

**Question 1:** Let  $\{\mathbf{S}_t\}$  be a stationary p.n.d. increasing family of subspaces with  $\mathbf{S}_t \supset \mathbf{H}_t(dy)$ . Under what conditions does  $dy$  admit a semimartingale representation of the form

$$y(t) - y(s) = \int_s^t z(\sigma) d\sigma + m(t) - m(s), \tag{5.4.6}$$

where  $\{z(t)\}$  is a process adapted to  $\{\mathbf{S}_t\}$  (i.e.  $z(t) \in \mathbf{S}_t, t \in \mathbb{R}$ ) and  $\{m(t)\}$  is a  $\mathbf{S}_t$ -martingale?

Dually, let  $\{\bar{\mathbf{S}}_t\}$  be a stationary p.n.d. decreasing family of subspaces with  $\bar{\mathbf{S}}_t \supset \mathbf{H}_t^+(dy)$ . Under what conditions does  $dy$  admit a backward semimartingale representation of the form

$$y(t) - y(s) = \int_s^t \bar{z}(\sigma) d\sigma + \bar{m}(t) - \bar{m}(s), \tag{5.4.7}$$

where  $\{\bar{z}(t)\}$  is now a process adapted to  $\{\bar{\mathbf{S}}_t\}$  (i.e.  $\bar{z}(t) \in \bar{\mathbf{S}}_t, t \in \mathbb{R}$ ) and  $\{\bar{m}(t)\}$  is a backward  $\bar{\mathbf{S}}_t$ -martingale, namely  $\bar{m}(t) - \bar{m}(s) \in \bar{\mathbf{S}}_t$  for  $s \geq t$  and  $\bar{m}(t-h) - \bar{m}(t) \perp \bar{\mathbf{S}}_t, \forall h \geq 0$  ?

Question 1 is answered by the following theorem.

**Theorem 5.4.4.** *Let  $dy$  and  $\{\mathbf{S}_t\}$  be as stated above. Then a necessary and sufficient condition for  $dy$  to admit a semimartingale representation with respect to  $\{\mathbf{S}_t\}$  of the type (5.4.6) is that there exists a constant  $k$  independent of  $h$ , such that*

$$\|E^{\mathbf{S}}[y(h) - y(0)]\| \leq kh \quad \forall h \geq 0. \tag{5.4.8}$$

In the representation (5.4.6),  $\{z(t)\}$  can be chosen stationary and mean-square continuous and  $\{m(t)\}$  with stationary (orthogonal) increments. The integral can be interpreted as a mean square Riemann integral.

Dually, let  $\{\bar{\mathbf{S}}_t\}$  be as specified in Question 1. Then a necessary and sufficient condition for  $dy$  to admit a a representation of the form (5.4.7) is that

$$\|E^{\bar{\mathbf{S}}}[y(-h) - y(0)]\| \leq \bar{k}h \quad \forall h \geq 0 \tag{5.4.9}$$

where the constant  $\bar{k}$  is independent of  $h$ . Here again  $\{\bar{z}(t)\}$  can be chosen stationary and mean-square continuous,  $\{\bar{m}(t)\}$  with stationary (orthogonal) increments and the integral can be interpreted as a mean square Riemann integral.

A process satisfying condition (5.4.8) (or (5.4.9)) will be called *conditionally Lipschitz* with respect to  $\{\mathbf{S}_t\}$  (or  $\{\bar{\mathbf{S}}_t\}$ ). Note that a  $m$ -dimensional martingale with stationary (orthogonal) increments must be a constant (matrix) multiple of a vector Wiener processes (see Chapter 3). It then follows that a stationary increments process which satisfies the conditional Lipschitz condition with respect to some increasing family  $\{\mathbf{S}_t\}$  has a unique decomposition into a m.s. differentiable stationary component adapted to  $\{\mathbf{S}_t\}$ , plus a matrix multiple of a vector Wiener process. It will be seen shortly that the Wiener process is in fact the generating process of  $\{\mathbf{S}_t\}$ . The stationary process  $\{z(t)\}$  is called the *conditional derivative of  $dy$  with respect to the increasing family  $\{\mathbf{S}_t\}$* .

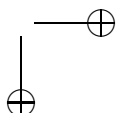
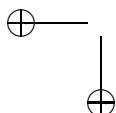
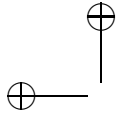
A completely dual picture holds for the backward setting.

## 5.5 Stationary increments semimartingales in the spectral domain

In this section we shall prove Theorem 5.4.4, our main representation result, by spectral domain techniques. We shall also establish a number of spectral domain characterizations of stationary increments processes which are believed to be of interest in themselves.

**Lemma 5.5.1.** *Let  $f \in \mathcal{L}_p^2$  and let  $P^{H_p^2}$  denote the orthogonal projection from  $\mathcal{L}_p^2$  onto  $H_p^2$ . Then the condition*

$$\|P^{H_p^2} \chi_h f\|_{\mathcal{L}_p^2} \leq kh, \quad h \geq 0, \tag{5.5.1}$$



is necessary and sufficient for  $f$  to admit a decomposition

$$f = g + \bar{g}, \tag{5.5.2}$$

where  $g \in H_p^2$  and  $\bar{g} \in \bar{W}_p^2$ . Dually

$$\|P^{\bar{H}_p^2} \bar{\chi}_h f\|_{\mathcal{L}_p^2} \leq \bar{k}h, \quad h \geq 0, \tag{5.5.3}$$

is necessary and sufficient for  $f$  to admit a decomposition of the form (5.5.2) but now with  $g \in W_p^2$  and  $\bar{g} \in \bar{H}_p^2$ . The decomposition (5.5.2) is unique.

**Proof.** (Necessity) Let (5.5.2) hold with  $g \in H_p^2$  and  $\bar{g} \in \bar{W}_p^2$ . Then  $\chi_h \bar{g}$  is orthogonal to  $H_p^2$ , so

$$\|P^{H_p^2} \chi_h f\|_{\mathcal{L}_p^2} = \|P^{H_p^2} \chi_h g\|_{\mathcal{L}_p^2} \leq \|\chi_h g\|_{\mathcal{L}_p^2} \leq \sup_{\omega} |\chi_h(\omega)| \|g\|_{\mathcal{L}_p^2}$$

and since  $\sup_{\omega} |\chi_h(\omega)| = h$ , we obtain (5.5.1).

(Sufficiency) Define the  $H_p^2$ -valued map  $h \mapsto z_h$ , where

$$z_h := P^{H_p^2} \chi_h f, \quad h \geq 0.$$

Clearly  $z_0 = 0$ . We shall show that if (5.5.1) holds, then the limit

$$\lim_{h \downarrow 0} \frac{1}{h} (z_h - z_0) = \lim_{h \downarrow 0} \frac{1}{h} P^{H_p^2} \chi_h f = g$$

exists weakly in  $H_p^2$ . To this end we shall introduce the *restricted right-shift semigroup*  $\Sigma_t : f \mapsto P^{H_p^2} e^{i\omega t} f ; t \geq 0$ , in  $H_p^2$ , see e.g. [45]. Note that  $\Sigma_t$  annihilates the anticausal part (in  $\bar{H}_p^2$ ) of any  $f \in L^2(\mathbb{I})$ , so that  $\Sigma_t f = \Sigma_t P^{H_p^2} f$  for all  $t \geq 0$ . It is then clear that

$$z_{t+h} - z_t = P^{H_p^2} e^{i\omega t} \frac{e^{i\omega h} - 1}{i\omega} f = \Sigma_t (P^{H_p^2} \chi_h f) = \Sigma_t (z_h - z_0)$$

for all  $t \geq 0$  and  $h \geq 0$  and hence, choosing an arbitrary  $\varphi \in H_p^2$ , we have from (5.5.1),

$$|\langle \varphi, z_{t+h} - z_t \rangle| \leq \|\Sigma_t^* \varphi\| \|z_h - z_0\| \leq \|\varphi\| kh,$$

where  $\Sigma_t^*$  is the operator of multiplication by  $e^{i\omega t}$ , the adjoint in  $H_p^2$  of  $\Sigma_t$ . By this inequality, we see that  $f_{\varphi}(t) := \langle \varphi, f \rangle$  is a Lipschitzian function of  $t$  in  $\mathbb{R}_+$  and hence admits a derivative there, except perhaps on set  $N_{\varphi}$  of Lebesgue measure zero. In other words, the limit

$$\lim_{h \downarrow 0} \langle \varphi, \frac{1}{h} (z_{t+h} - z_t) \rangle = \lim_{h \downarrow 0} \langle \Sigma_t^* \varphi, \frac{1}{h} (z_h - z_0) \rangle$$

exists for all  $t \in \mathbb{R}_+ - N_{\varphi}$ . Now  $\Sigma_t^*$  is a strongly continuous semigroup and  $\{\Sigma_t^* \varphi ; t \in \mathbb{R}_+ - N_{\varphi}, \varphi \in H_p^2\}$  certainly contains a dense set in  $H_p^2$ . On the other hand  $\frac{1}{h} (z_h - z_0)$

112 Chapter 5. Wold Decomposition, and Spectral Factorization in Continuous Time

is bounded in norm for all  $h \geq 0$  by virtue of condition (5.5.1). Hence, by a well know characterization of weak convergence (see e.g. [2, p. 47]),  $\frac{1}{h}(z_h - z_0)$  converges weakly to an element  $g$  of  $H_p^2$ . But then the limit

$$\lim_{h \downarrow 0} \frac{1}{h}(z_{t+h} - z_t) = \lim_{h \downarrow 0} \Sigma_t \frac{1}{h}(z_h - z_0)$$

exists also weakly for all  $t \geq 0$  and is equal to  $\Sigma_t g$ . The left (weak) derivative also exists at any  $t > 0$  since  $\frac{1}{h}(z_t - z_{t-h}) = \Sigma_{t-h} \frac{1}{h}(z_h - z_0)$  and for all  $h > 0$  such that  $t - h > 0$  we have

$$\begin{aligned} \langle \varphi, \frac{1}{h}(z_t - z_{t-h}) \rangle &= \langle \Sigma_{t-h}^* \varphi, \frac{1}{h}(z_h - z_0) \rangle \\ &= \langle (\Sigma_{t-h}^* - \Sigma_t^*) \varphi, \frac{1}{h}(z_h - z_0) \rangle + \langle \Sigma_t^* \varphi, \frac{1}{h}(z_h - z_0) \rangle. \end{aligned}$$

The first term in the last member tends to zero as  $h \downarrow 0$ , since  $\Sigma_t^*$  is strongly continuous and  $\frac{1}{h}(z_h - z_0)$  is bounded in norm. Hence  $\langle \varphi, 1/h(z_t - z_{t-h}) \rangle \rightarrow \langle \Sigma_t^* \varphi, g \rangle = \langle \varphi, \Sigma_t g \rangle$  for all  $\varphi \in H_p^2$ , so we have shown that  $f_\varphi(t) = \langle \varphi, z_t \rangle$  is differentiable with a continuous derivative  $\dot{f}_\varphi(t) = \langle \varphi, \Sigma_t g \rangle$  at every point  $t \geq 0$ . Thus

$$\langle \varphi, z_h - z_0 \rangle = \int_0^h \langle \varphi, \Sigma_t g \rangle dt, \quad h \geq 0$$

for any  $\varphi \in H_p^2$ . Now just note that the integral  $\int_0^h \Sigma_t g dt$  exists in the strong sense in  $H_p^2$  since  $t \mapsto \Sigma_t g$  is continuous. Therefore we can write (see e.g. [126])

$$\int_0^h \langle \varphi, \Sigma_t g \rangle dt = \langle \varphi, \int_0^h \Sigma_t g dt \rangle$$

which, by the previous equality implies that

$$z_h - z_0 = \int_0^h \Sigma_t g dt, \quad h \geq 0.$$

Incidentally, we have just shown that  $z_t$  is *strongly differentiable*. Recalling the definitions of  $z_h$  and of  $\Sigma_t$  we obtain

$$P^{H_p^2} \chi_h f = P^{H_p^2} \int_0^h e^{i\omega t} g dt = P^{H_p^2} \chi_h g.$$

Define now  $\bar{g} := f - g$ . Then  $\bar{g}$  is an  $\mathcal{L}_p^2$  function which in virtue of the above equality, satisfies

$$P^{H_p^2} \chi_h \bar{g} = 0, \quad \forall h \geq 0$$

that is,  $\chi_h \bar{g} \in \bar{H}_p^2, \forall h \geq 0$ . It follows by Lemma 5.2.4 that  $\bar{g} \in \bar{W}_p^2$ .

The dual statement follows by the same arguments.  $\square$

An important special case of the lemma is obtained by considering functions  $f$  belonging to the subspaces  $W_p^2$  or  $\bar{W}_p^2$  of  $\mathcal{L}_p^2$ .

Let  $f \in \mathcal{W}_p^2$ . Let us agree to call such an  $f$  *decomposable*, if it can be written in the form  $f = g + c$  where  $g \in H_p^2$  and  $c$  is a constant vector. Decomposability can be defined *mutatis mutandis* also for functions  $\bar{f} \in \bar{\mathcal{W}}_p^2$ . Matrix valued functions with rows in  $\mathcal{W}_p^2$  (or in  $\bar{\mathcal{W}}_p^2$ ) are decomposable if they can be split as a sum of a matrix function with rows in  $H_p^2$  ( $\bar{H}_p^2$ ) plus a constant matrix. Note that these decompositions are unique<sup>19</sup>.

**Corollary 5.5.2.** *Let  $\mathbf{S}$  be a backward shift invariant p.n.d. subspace containig  $\mathbf{H}^-(dy)$  and  $W \in \mathcal{W}_{m \times p}^2$  the corresponding causal full-rank spectral factor. Then  $dy$  is conditionally Lipschitz with respect to  $\{\mathbf{S}_t\}$  if and only if  $W$  is decomposable; i.e. there exists a constant  $m \times p$  matrix  $D$  and a matrix function  $G \in H_{m \times p}^2$  such that*

$$W(i\omega) = G(i\omega) + D \tag{5.5.4}$$

*Dually, let  $\bar{\mathbf{S}}$  be a forward shift invariant p.n.d. subspace containig  $\mathbf{H}^+(dy)$  and  $\bar{W} \in \bar{\mathcal{W}}_{m \times p}^2$  the corresponding anticausal full-rank spectral factor. Then  $dy$  is conditionally Lipschitz with respect to  $\{\bar{\mathbf{S}}_t\}$  if and only if  $\bar{W}$  is decomposable; i.e. there exists a constant  $m \times p$  matrix  $\bar{D}$  and a matrix function  $\bar{G} \in \bar{H}_{m \times p}^2$  such that*

$$\bar{W}(i\omega) = \bar{G}(i\omega) + \bar{D} \tag{5.5.5}$$

**Proof.** Let  $dw$  be the generating Wiener process of  $\mathbf{S}$  and  $d\hat{w}$  its Fourier transform. From the spectral representation of  $dy$  it follows that the random variables  $E^{\mathbf{S}}[y_k(h) - y_k(0)], k = 1, \dots, m$  correspond, under the isomorphism  $J_{\hat{w}}$ , to  $P^{H_p^2} \chi_h W_k, k = 1, \dots, m, W_k$  being the  $k$ -th row of  $W$ . Since  $J_{\hat{w}}$  is a unitary map,  $dy$  is conditionally Lipschitz if and only if  $\|P^{H_p^2} \chi_h W_k\| = O(h)$  for all  $k$ . Then the decomposition (5.5.5) follows from (5.5.2) since in this case all  $\bar{g}$ 's must be constant.  $\square$

We have now available the instruments to provide a proof of Theorem 5.4.4

### Proof of Theorem 5.4.4

Sufficiency is almost immediate. For let  $dw$  be the generating Wiener process of  $\mathbf{S}$  and  $W$  be the corresponding causal spectral factor. Since the conditional Lipschitz condition is equivalent to the decomposability (5.5.5), we have

$$\begin{aligned} y(h) - y(0) &= \int_{-\infty}^{+\infty} \chi_h(i\omega) W(i\omega) d\hat{w} = \int_{-\infty}^{+\infty} \chi_h(i\omega) G(i\omega) d\hat{w} + \int_{-\infty}^{+\infty} \chi_h(i\omega) D d\hat{w} \\ &= \int_0^h z(t) dt + D[w(h) - w(0)] \end{aligned}$$

where  $\{z(t)\}$  is the stationary process

$$z(t) = \int_{-\infty}^{+\infty} e^{i\omega t} G(i\omega) d\hat{w} \tag{5.5.6}$$

<sup>19</sup>This also follows from the fact that  $H_p^2$  functions of the half plane tend uniformly to zero as  $s \rightarrow \infty$  within the region of analiticity, see [51].

114 Chapter 5. Wold Decomposition, and Spectral Factorization in Continuous Time

which is clearly adapted to  $\{\mathbf{S}_t\}$  (as  $G \in H^2_{m \times p}$ ) and mean square continuous.

To prove necessity we shall first show that the process  $\{z(t)\}$  in the representation (5.4.6) can always be chosen stationary and mean-square continuous.

For (5.4.6) to make sense we have at least to assume  $\{z(t)\}$  measurable and with a locally square integrable norm. Note then that  $y(t+h) - y(t) = U_t[y(h) - y(0)]$  can be written either as

$$y(t+h) - y(t) = \int_t^{t+h} z(s) ds + m(t+h) - m(t)$$

or as

$$y(t+h) - y(t) = \int_0^h U_t z(s) ds + U_t[m(h) - m(0)].$$

Keeping  $t$  fixed and letting  $h$  vary in  $\mathbb{R}_+$ , the second term in the last expression defines a martingale with respect to the increasing family  $\tilde{\mathbf{S}}_h := \mathbf{S}_{t+h}$ . The same is of course true for  $m(t+h) - m(t)$ . On the other hand both first terms in the two expressions above are in  $\mathbf{S}_h$  and of bounded mean variation, as functions of  $h$ . By the uniqueness proposition 5.4.2, they must be equal; i.e.

$$\int_0^h [z(t+s) - U_t z(s)] ds = 0, \quad \forall h \geq 0$$

almost surely, and hence  $\{z(t+s)\}$  and  $\{U_t z(s)\}$  are equivalent processes for all  $t$ . We can therefore choose  $\{z(t)\}$  to be generated by the shift; i.e.  $z(t) = U_t z(0)$  and thereby stationary and mean square continuous (for  $U_t$  is a strongly continuous semigroup). The integral in (5.4.6) can therefore be understood as a mean square Riemann integral.

We shall now show that the conditional Lipschitz condition is necessary for  $dy$  to admit a representation of the form (5.4.6). Since  $z(t) = U_t z(0)$  we have  $\|z(t)\| = \|z(0)\|$  and it follows from (5.4.6) that

$$\|E^{\mathbf{S}}[y(h) - y(0)]\| \leq \int_0^h \|E^{\mathbf{S}} z(t)\| dt \leq \|z(0)\| h$$

and hence condition (5.4.8) is implied by (5.4.6). This concludes the proof.  $\square$

Note that the decomposition (5.5.5) implies that the semimartingale representation of  $dy$  with respect to a stationary family  $\mathbf{S}_t = \mathbf{H}_t^-(dw)$ , can be written in the frequency domain as

$$d\hat{y} = Gd\hat{w} + Dd\hat{w} \tag{5.5.7}$$

the first term on the right (namely  $d\hat{z} := Gd\hat{w}$ ) being the stationary component of  $dy$ . It should be pointed out that the matrix  $DD'$  is invariant over all semimartingale representations of  $dy$ . In fact  $DD' h$  is the quadratic variation of the process on the interval  $[0, h]$  and, as explained before, this quantity is independent of the particular  $\mathbf{S}$  (or  $\tilde{\mathbf{S}}$ ) with respect to which the process has a semimartingale representation. As we shall see, in the rational case we have  $DD' = \lim_{s \rightarrow \infty} \Phi(s)$ .

We shall call a stationary increments process *nondegenerate* if  $\text{rank } DD' = \text{rank } \Phi(i\omega)$  a.e.; in other words, in case the rank of  $DD'$  is equal to the multiplicity  $p$  of the process, and *degenerate* otherwise.

**Proposition 5.5.3.** *Assume that  $dy$  has the semimartingale representation (5.5.7), with respect to some increasing family  $\mathbf{S}_t = \mathbf{H}_t^-(dw)$ , and let  $\text{rank } DD' = r < p$  (i.e. the process is degenerate). There is a constant orthogonal transformation of the generating process  $dw$  which permits to decompose  $dy$  in the sum of two uncorrelated semimartingales  $dy_1$  and  $dy_2$ , the first nondegenerate of multiplicity  $r$  and the second without martingale part (i.e. completely degenerate).*

**Proof.** The matrix  $DD'$  can be factored as  $\tilde{D}\tilde{D}'$  with  $\tilde{D}$  of dimension  $m \times r$  and of full column rank ( $r$ ). Hence  $\tilde{D}$  has a left-inverse, e.g.  $\tilde{D}^{-L} = (\tilde{D}'\tilde{D})^{-1}\tilde{D}'$ . Define the  $r$ -dimensional Wiener process  $du$  by setting

$$du := \tilde{D}^{-L}Ddw := Ndw.$$

Note that  $N$  is an orthogonal  $r \times p$  matrix; i.e.  $NN' = I_r$ , so that  $\mathbb{E} du du' = I_r dt$ . Also, because of orthogonality, both  $NN'$  and  $I_p - NN'$  are projection matrices, in fact, complementary projections in  $\mathbb{R}^p$ . Since  $\text{rank}(I_p - NN') = p - r$ , we can find a factorization:  $I_p - NN' = MM'$  with  $M$  a full rank matrix of dimension  $(p - r) \times p$ , which is necessarily orthogonal; i.e.  $MM' = I_{p-r}$ . Define now the normalized Wiener process  $dv := Mdw$ . It is easy to check that the increments of  $du$  and  $dv$  are orthogonal, since

$$\mathbb{E} du dv' = NM' = NN'NM'MM' = N[N'N(I - N'N)]M' = 0.$$

Hence it follows from the orthogonal decomposition

$$dw = N'Ndw + (I_p - N'N)dw = [N' \ M'] \begin{bmatrix} du \\ dv \end{bmatrix} \tag{5.5.8}$$

that  $\mathbf{S}_t = \mathbf{H}_t^-(dw)$  splits in the orthogonal direct sum  $\mathbf{S}_t = \mathbf{H}_t^-(du) \oplus \mathbf{H}_t^-(dv)$ . In fact, by setting  $G [N' \ M'] := [G_1 \ G_2]$ , substituting (the Fourier transform of) (5.5.8) into (5.5.7) and recalling that  $DN' = \tilde{D}$  and  $DM' = \tilde{D}NM' = 0$ , we obtain

$$d\hat{y} = (G_1 d\hat{u} + \tilde{D}d\hat{u}) + G_2 d\hat{v} := d\hat{y}_1 + d\hat{y}_2.$$

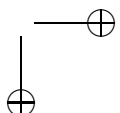
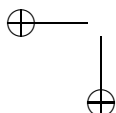
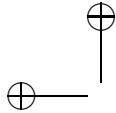
Here  $d\hat{y}_1$  and  $d\hat{y}_2$  are clearly uncorrelated and  $\tilde{D}$  has rank  $r$  so  $dy_1$  is nondegenerate.  $\square$

## 5.6 Bibliographical notes

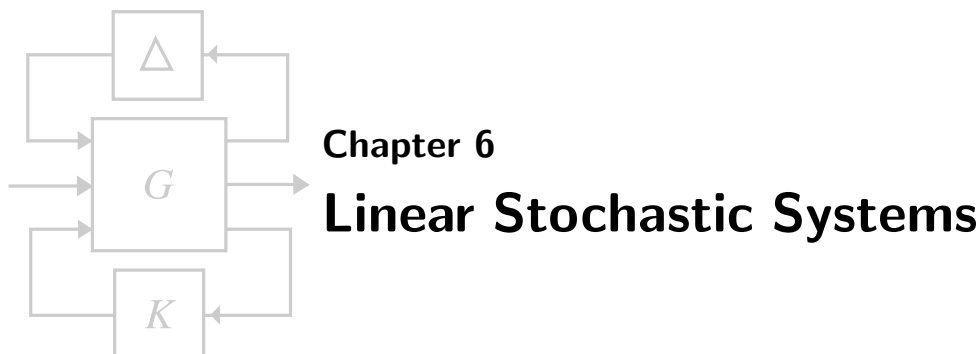
The extension of the spectral factorization theory to stationary increments process described in this chapter is due to the authors, [84]. Semimartingales (called quasi-martingales in the early literature) were introduced by Fisk, [29] and have since then played a prominent role in the theory of continuous time stochastic processes. The wide-sense version discussed in this chapter requires only a minimum of measure-theoretic technicalities. It does not seem to have been treated before [84] in the literature. Conditions of the type (5.4.8) have to do with the characterization of

116 Chapter 5. Wold Decomposition, and Spectral Factorization in Continuous Time

the domain of the *conditional shift semigroup*  $E^S U_t; t \geq 0$  and appear in a very general context in the work of Rishel [105], see also [115] and others. Our condition is of course quite weaker than what is needed in the strict-sense theory.







This chapter is an introduction to linear state-space modeling of second-order, stationary, purely nondeterministic, stochastic vector processes with a rational spectral density. Such processes can be modeled as the output  $\{y(t)\}$  of a finite-dimensional linear system

$$\begin{cases} x(t+1) = Ax(t) + Bw(t) \\ y(t) = Cx(t) + Dw(t) \end{cases} \quad (6.0.1)$$

driven by white noise input  $\{w(t)\}$ , where  $A, B, C$  and  $D$  are matrices of appropriate dimensions, and where  $A$  has all its eigenvalues in the open unit disc. Stochastic realization theory consists in characterizing and determining any such representation. This leads to spectral factorization. The geometric structure of stochastic models is described in terms of coordinate-free representations based on elementary Hilbert space concepts.

### 6.1 Basic principles of deterministic realization theory

Before introducing stochastic realization theory, we briefly review the basic principles of state space construction in deterministic realization theory. To this end, consider the constant, linear system  $\Sigma$  described by

$$(\Sigma) \quad \begin{cases} x(t+1) = Ax(t) + Bu(t) \\ y(t) = Cx(t) + Du(t) \end{cases} \quad (6.1.1)$$

where  $x$  takes values in the state space  $X$ ,  $u$  in the input space and  $y$  in the output space  $Y$ . The spaces  $X, U$  and  $Y$  will be identified with  $\mathbb{R}^n, \mathbb{R}^\ell$  and  $\mathbb{R}^m$ , respectively, and  $A \in \mathbb{R}^{n \times n}, B \in \mathbb{R}^{n \times \ell}, C \in \mathbb{R}^{m \times n}$  and  $D \in \mathbb{R}^{m \times \ell}$  are matrices. The *dimension* is defined to be the dimension the state space  $X$ , i.e.,  $\dim \Sigma := n$ .

Assuming that the system  $\Sigma$  is at rest at time  $t = 0$ , i.e.,  $x(0) = 0$ , an input signal  $u(t) = u(0)\delta_{t0}$ , i.e., an impulse<sup>20</sup> at time  $t = 0$ , yields an output signal

<sup>20</sup> $\delta_{st}$  is the Kronecker symbol that is one when  $s = t$  and zero otherwise.

$y(t) = R_t u(0)$  for  $t \geq 0$ . The sequence

$$R_0, R_1, R_2, R_3, \dots \tag{6.1.2}$$

of  $m \times \ell$  matrices is called the *impulse response* of  $\Sigma$ . Clearly,

$$R_0 = D, \quad R_k = CA^{k-1}B, \quad k = 1, 2, 3, \dots, \tag{6.1.3}$$

so the the *transfer function* of  $\Sigma$ ,

$$R(z) = \sum_{k=0}^{\infty} R_k z^{-k}, \tag{6.1.4}$$

converges in the neighborhood of infinity (outside a disc of radius equal to the maximum eigen value of  $A$ ) to the rational  $m \times \ell$  matrix function

$$R(z) = C(zI - A)^{-1}B + D. \tag{6.1.5}$$

The realization problem is the inverse problem of determining a system  $\Sigma$  with a given transfer function  $R$ . Such a  $\Sigma$  is called a *realization* of  $R$ . A realization  $\Sigma$  is *minimal* if there is no other realization of  $R$  of smaller dimension. The *McMillan degree*  $\delta(R)$  of  $R$  is the dimension of a minimal realization of  $R$ .

In other words, given a linear time-invariant input/output system

$$\xrightarrow{u} \boxed{R(z)} \xrightarrow{y}$$

with an impulse response (6.1.2), the realization problem amounts to determining matrices  $(A, B, C, D)$  such that the corresponding system  $\Sigma$  has this impulse response. The matrix  $D$  can immediately be identified as  $R_0$ , so it really remains to determine  $(A, B, C)$ . By time-invariance, the input  $u(t) = u(s)\delta_{ts}$  will yield the output  $y(t) = R_{t-s}u(s)$ , so, by superposition, one obtains the output

$$y(t) = \sum_{s=-N}^{-1} R_{t-s}u(s), \quad t \geq 0 \tag{6.1.6}$$

from an input string  $\{u(-N), u(-N+1), \dots, u(-1)\}$ .

An important tool in realization theory is the Hankel map obtained by passing such a finite string  $\{u(-N), u(-N+1), \dots, u(-1)\}$  of inputs through the system  $\Sigma$  that is originally at rest ( $x(N) = 0$ ) and then observing the output sequence  $\{y(0), y(1), y(2), \dots\}$ . This yields precisely (6.1.6), or, equivalently, the block Hankel system

$$\begin{bmatrix} y(0) \\ y(1) \\ y(2) \\ \vdots \end{bmatrix} = \begin{bmatrix} R_1 & R_2 & R_3 & \dots \\ R_2 & R_3 & R_4 & \dots \\ R_3 & R_4 & R_5 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} \begin{bmatrix} u(-1) \\ u(-2) \\ \vdots \\ u(-N) \\ 0 \\ \vdots \end{bmatrix}.$$

### The Hankel factorization

The basic idea in realization theory is that, if  $R$  has a finite-dimensional realization  $(A, B, C, D)$ , the Hankel matrix

$$\mathcal{H} := \begin{bmatrix} R_1 & R_2 & R_3 & \cdots \\ R_2 & R_3 & R_4 & \cdots \\ R_3 & R_4 & R_5 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} \tag{6.1.7}$$

has finite rank and admits the factorization

$$\mathcal{H} = \mathcal{O}\mathcal{R} \tag{6.1.8}$$

where  $\mathcal{R}$  is the reachability matrix

$$\mathcal{R} = [B \quad AB \quad A^2B \quad \cdots] \tag{6.1.9}$$

and  $\mathcal{O}$  is the observability matrix

$$\mathcal{O} = \begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \end{bmatrix}. \tag{6.1.10}$$

In fact, it follows from (6.1.3) that

$$\mathcal{H} := \begin{bmatrix} CB & CAB & CA^2B & \cdots \\ CAB & CA^2B & CA^3B & \cdots \\ CA^2B & CA^3B & CA^4B & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} = \begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \end{bmatrix} [B \quad AB \quad A^2B \quad \cdots].$$

Abstractly this factorization may be illustrated by the commutative diagram

$$\begin{array}{ccc} \mathcal{U} & \xrightarrow{\mathcal{H}} & \mathcal{Y} \\ \mathcal{R} \searrow & & \nearrow \mathcal{O} \\ & \mathcal{X} & \end{array}$$

where  $\mathcal{Y}$  is the class of output sequences such that  $y(t) = 0$  for  $t < 0$  and  $\mathcal{U}$  is the class of finite input sequences such that  $u(t) = 0$  for  $t \geq 0$  and  $t < -N$  for some finite  $N$ . In fact,

$$\mathcal{R} \begin{bmatrix} u(-1) \\ u(-2) \\ u(-3) \\ \vdots \end{bmatrix} = Bu(-1) + ABu(-2) + A^2Bu(-3) + \cdots = x(0)$$

and

$$\begin{bmatrix} y(0) \\ y(1) \\ y(2) \\ \vdots \end{bmatrix} = \begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \end{bmatrix} x(0) = \mathcal{O}x(0).$$

The system  $\Sigma$  is said to be *completely reachable* if

$$\text{Im } \mathcal{R} = X, \tag{6.1.11}$$

i.e.,  $\mathcal{R}$  is surjective (onto) and *completely observable* if

$$\ker \mathcal{O} = 0, \tag{6.1.12}$$

i.e.,  $\mathcal{O}$  is injective (one-to-one). For simplicity, we shall say that the pair  $(A, B)$  is *reachable* if and only if (6.1.11) holds and that  $(C, A)$  is *observable* if and only if (6.1.12) holds.

### Solving the realization problem

To determine  $(A, B, C)$  from such a factorization it is better to deal with finite matrices. The assumption that  $R$  is rational and proper lets us do precisely this. In fact, let

$$\rho(z) = z^r + a_1 z^{r-1} + \cdots + a_r \tag{6.1.13}$$

be the least common denominator of the elements of  $R(z)$ . Then  $\rho(z)R(z)$  is a polynomial, so identifying coefficients of negative powers in

$$\rho(z)R(z) = (z^r + a_1 z^{r-1} + \cdots + a_r)(R_0 + R_1 z^{-1} + R_2 z^{-2} + R_3 z^{-3} + \cdots),$$

we see that the impulse response (6.1.2) must satisfy the *finiteness condition*

$$R_{r+k} = -a_1 R_{r+k-1} - a_2 R_{r+k-2} - \cdots - a_r R_k, \quad k = 1, 2, 3, \dots \tag{6.1.14}$$

Consequently, for  $\nu \geq r$ , successively adding block rows and block columns in

$$\mathcal{H}_\nu := \begin{bmatrix} R_1 & R_2 & \cdots & R_\nu \\ R_2 & R_3 & \cdots & R_{\nu+1} \\ \vdots & \vdots & \vdots & \ddots \\ R_\nu & R_{\nu+1} & \cdots & R_{2\nu-1} \end{bmatrix}, \quad \nu = 0, 1, 2, \dots \tag{6.1.15}$$

does not increase the rank. Hence we have the following lemma.

**Lemma 6.1.1.** *Let  $\rho(z)$  the least common denominator (6.1.13) of the elements of  $R(z)$ . Then*

$$\text{rank } \mathcal{H}_\nu = \text{rank } \mathcal{H} \quad \text{for all } \nu \geq r,$$

where  $r := \deg \rho$ .

Therefore, instead of (6.1.8), we can consider the finite-dimensional factorization problem

$$\mathcal{H}_\nu = \mathcal{O}_\nu \mathcal{R}_\nu \tag{6.1.16}$$

where

$$\mathcal{O}_\nu = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{\nu-1} \end{bmatrix}, \quad \mathcal{R}_\nu = [B \quad AB \quad \cdots \quad A^{\nu-1}B]. \tag{6.1.17}$$

By (6.1.14),  $\mathcal{O}_\nu$  and  $\mathcal{R}_\nu$  have the same ranks as  $\mathcal{O}$  and  $\mathcal{R}$ , respectively, and hence the following holds.

**Lemma 6.1.2.** *Let  $\Sigma$  be a system (6.1.1) of dimension  $n$ , and let  $\nu \geq r := \deg \rho$ . Then the system  $\Sigma$  is completely observable if and only if  $\text{rank } \mathcal{O}_\nu = n$  and completely reachable if and only if  $\text{rank } \mathcal{R}_\nu = n$ .*

From the factorization (6.1.16) it follows that

$$\text{rank } \mathcal{H} \leq n := \dim \Sigma, \tag{6.1.18}$$

and therefore  $\text{rank } \mathcal{H}$  is a lower bound for the McMillan degree  $\delta(R)$ . We shall demonstrate that in fact  $\delta(R) = \text{rank } \mathcal{H}$  by constructing a realization of dimension precisely  $\text{rank } \mathcal{H}$ . To this end, we perform a *minimal factorization* of  $\mathcal{H}_\nu$  for some  $\nu > r := \deg \rho$ . More precisely, given  $p := \text{rank } \mathcal{H}_\nu$ , we determine two matrices  $\Omega_\nu$  and  $\Gamma_\nu$  of dimensions  $m\nu \times p$  and  $p \times k\nu$ , respectively, such that

$$\mathcal{H}_\nu = \Omega_\nu \Gamma_\nu. \tag{6.1.19}$$

This factorization can be illustrated by the commutative diagram

$$\begin{array}{ccc} \mathcal{U} & \xrightarrow{\mathcal{H}_\nu} & \mathcal{Y} \\ \Gamma_\nu \searrow & & \nearrow \Omega_\nu \\ & \mathbf{X} & \end{array}$$

where  $\dim \mathbf{X} = p := \text{rank } \mathcal{H}_\nu$ . The idea is to determine a minimal realization  $(A, B, C, D)$  from these factors. For this, we need some notation. Given an  $m\nu \times p$  matrix  $\Omega_\nu$  with  $\nu > r$ , let  $\sigma(\Omega_r)$  be the shifted  $mr \times p$  matrix obtained by removing the first  $m \times p$  block row and the last  $\nu + 1 - r$  block rows. Moreover, let  $\sigma(\Gamma_r)$  be the  $p \times kr$  matrix obtained by an analogous operation on the block columns of  $\Gamma_\nu$ . Finally, let  $Q^\dagger$  denote the Moore-Penrose pseudo-inverse of the matrix  $Q$ , and let

$$E_k := [I_k \quad 0 \quad \cdots \quad 0]'$$

be the  $rk \times k$  matrix consisting of  $r$  blocks of dimensions  $k \times k$ , the first being the identity and all others zero.

**Theorem 6.1.3.** *Given a proper rational  $m \times \ell$  matrix function  $R(z)$ , let  $\rho(z)$  the least common denominator (6.1.13) of the elements of  $R$ . Moreover, for some*

$\nu > r := \deg \rho$ , let (6.1.19) be a minimal factorization of the Hankel matrix (6.1.15). Then

$$A = (\Omega_r)^\dagger \sigma(\Omega_r), \quad B = \Gamma_r E_\ell, \quad C = E'_m \Omega_r, \quad D = R_0 \quad (6.1.20)$$

is a minimal realization of  $R$ , and its dimension is  $\text{rank } \mathcal{H}$ . Symmetrically,  $A$  is also given by

$$A = \sigma(\Gamma_r)(\Gamma_r)^\dagger. \quad (6.1.21)$$

Moreover, the corresponding observability and reachability matrices (6.1.17) are given by

$$\mathcal{O}_r = \Omega_r, \quad \mathcal{R}_r = \Gamma_r. \quad (6.1.22)$$

**Proof.** First note that factorizations (6.1.19) can be performed consistently for different choices of  $\nu$  so that  $\Omega_\nu$  and  $\Gamma_\nu$  are submatrices of  $\Omega_\mu$  and  $\Gamma_\mu$ , respectively, whenever  $\nu \leq \mu$ . Now, choosing  $\nu$  sufficiently large, we can form the multiple shifts  $\sigma^k(\Omega_r)$  and  $\sigma^k(\Gamma_r)$  by deleting  $k$  blocks at the beginning and  $\nu - k - r$  blocks at the end of  $\Omega_\mu$  and  $\Gamma_\mu$ , respectively. Then, by inspection, we see that

$$\sigma^j(\Omega_r)\sigma^k(\Gamma_r) = \sigma^{j+k}(\mathcal{H}_r) \quad (6.1.23)$$

where

$$\sigma^k(\mathcal{H}_r) = \begin{bmatrix} R_{k+1} & R_{k+2} & \cdots & R_{k+r} \\ R_{k+2} & R_{k+3} & \cdots & R_{k+r+1} \\ \vdots & \vdots & \ddots & \vdots \\ R_{k+r} & R_{k+r+1} & \cdots & R_{k+2r-1} \end{bmatrix}.$$

Therefore, taking  $A := (\Omega_r)^\dagger \sigma(\Omega_r)$ ,

$$A\sigma^k(\Gamma_r) = (\Omega_r)^\dagger \sigma(\Omega_r)\sigma^k(\Gamma_r) = (\Omega_r)^\dagger \Omega_r \sigma^{k+1}(\Gamma_r) = \sigma^{k+1}(\Gamma_r),$$

since  $\Omega_r$  has full column rank and thus  $(\Omega_r)^\dagger \Omega_r = I$ . This immediately yields

$$A^k \Gamma_r = \sigma^k(\Gamma_r), \quad k = 0, 1, 2, \dots \quad (6.1.24)$$

In particular, choosing  $k = 1$ , we obtain  $A\Gamma_r = \sigma(\Gamma_r)$ , from which (6.1.21) follows. In the same way, we also have

$$\Omega_r A^k = \sigma^k(\Omega_r), \quad k = 0, 1, 2, \dots \quad (6.1.25)$$

Given (6.1.20), we obtain

$$C A^{k-1} B = E'_m \Omega_r A^{k-1} \Gamma_r E_\ell,$$

which, in view of (6.1.24) and (6.1.23), yields

$$C A^{k-1} B = E'_m \Omega_r \sigma^{k-1}(\Gamma_r) E_\ell = E'_m \sigma^{k-1}(\mathcal{H}_r) E_\ell = R_k$$

for  $k = 1, 2, \dots$ . Since trivially  $D = R_0$ , this establishes (6.1.3). From (6.1.24) we have  $A^k B = \sigma^k(\Gamma_r) E_\ell$ , which yields

$$[B \quad AB \quad \cdots \quad A^{r-1}B] = \Gamma_r,$$

i.e.,  $\mathcal{R}_r = \Gamma_r$ , as claimed. In same way,  $\mathcal{O}_r = \Omega_r$  is derived from (6.1.25).  $\square$

In view of (6.1.18) and Lemma 6.1.1, we immediately have the following corollary.

**Corollary 6.1.4.** *The McMillan degree  $\delta(R)$  of  $R$  equals the rank of the Hankel matrix  $\mathcal{H}$ .*

As another corollary we have the following fundamental fact in deterministic realization theory.

**Theorem 6.1.5.** *A realization  $\Sigma$  of  $R$  is minimal if and only if it is both completely reachable and completely observable.*

**Proof.** Let  $n := \dim \Sigma$ . Then, by Corollary 6.1.4 and Lemma 6.1.1,  $\Sigma$  is a minimal realization of  $R$  if and only if

$$\text{rank } \mathcal{H}_r = n. \tag{6.1.26}$$

In this holds, by (6.1.16),

$$n = \text{rank } \mathcal{H}_r \leq \min(\text{rank } \mathcal{O}_r, \text{rank } \mathcal{R}_r) \leq n,$$

and hence

$$\text{rank } \mathcal{O}_r = \text{rank } \mathcal{R}_r = n, \tag{6.1.27}$$

which, by Lemma 6.1.2, is equivalent to  $\Sigma$  being completely observable and completely reachable. Conversely, if (6.1.27) holds, the  $n \times n$  matrices  $\mathcal{O}'_r \mathcal{O}_r$  and  $\mathcal{R}_r \mathcal{R}'_r$  both have rank  $n$ , and hence so does

$$\mathcal{O}'_r \mathcal{O}_r \mathcal{R}_r \mathcal{R}'_r = \mathcal{O}'_r \mathcal{H}_r \mathcal{R}'_r.$$

But then  $\mathcal{H}_r$  too must have rank  $n$ .  $\square$

An  $n \times n$  matrix  $A$  is said to be a *stability matrix* if all its eigenvalues are less than one in modulus.

**Corollary 6.1.6.** *Let  $\Sigma$  be a minimal realization of  $R$ . Then  $A$  is a stability matrix if and only if  $R_k \rightarrow 0$  as  $k \rightarrow \infty$ .*

**Proof.** In view of (6.1.3), trivially  $R_k \rightarrow 0$  if  $A$  is a stability matrix. Conversely, if  $R_k \rightarrow 0$  as  $k \rightarrow \infty$ ,  $\mathcal{O}_r A^k \mathcal{R}_r \rightarrow 0$ . But then  $\mathcal{O}'_r \mathcal{O}_r A^k \mathcal{R}_r \mathcal{R}'_r \rightarrow 0$ , and consequently  $A^k \rightarrow 0$ , establishing stability.  $\square$

**Corollary 6.1.7.** *Let  $(\hat{A}, \hat{B}, \hat{C}, \hat{D})$  be any minimal realization of  $R$ . Then there is a nonsingular matrix  $T$  such that*

$$(\hat{A}, \hat{B}, \hat{C}, \hat{D}) = (TAT^{-1}, TB, CT^{-1}, D), \tag{6.1.28}$$

where  $(A, B, C, D)$  is as defined in Theorem 6.1.3. Conversely, for any nonsingular  $T$ , (6.1.28) is a minimal realization of  $R$ .

**Proof.** If  $(A, B, C, D)$  and  $(\hat{A}, \hat{B}, \hat{C}, \hat{D})$  are minimal realizations of  $R$ , then the corresponding observability and reachability matrices  $\mathcal{O}_r, \mathcal{R}_r$  and  $\hat{\mathcal{O}}_r, \hat{\mathcal{R}}_r$ , respectively, have full rank. Moreover,

$$\mathcal{O}_r \mathcal{R}_r = \mathcal{H}_r = \hat{\mathcal{O}}_r \hat{\mathcal{R}}_r \tag{6.1.29}$$

$$\mathcal{O}_r A \mathcal{R}_r = \sigma(\mathcal{H}_r) = \hat{\mathcal{O}}_r A \hat{\mathcal{R}}_r \tag{6.1.30}$$

From (6.1.29) it follows that

$$\hat{\mathcal{R}}_r = T \mathcal{R}_r, \quad \hat{\mathcal{O}}_r = \mathcal{O}_r T^{-1}$$

where

$$T = (\hat{\mathcal{O}}_r)^\dagger \mathcal{O}_r = \hat{\mathcal{R}}_r (\mathcal{R}_r)^\dagger.$$

In particular, this implies that  $\hat{B} = TB$  and  $\hat{C} = CT^{-1}$ . Then  $\hat{A} = TAT^{-1}$  follows from (6.1.30). Trivially, we have  $\hat{D} = R_0 = D$ .  $\square$

## 6.2 Stochastic state-space models

Wide-sense stationary processes with a rational spectral density matrix provide a natural and useful class of finitely-parametrized stochastic models leading to simple recursive estimation algorithms. It turns out that these processes are precisely those admitting finite-dimensional state-space descriptions (or stochastic realizations) with constant parameters. The essential structural property that leads to finite-dimensional recursive filtering (and identification) is in fact the representability of the process as a linear function of a finite-dimensional Markov process. This Markov process, called the *state* of the process represented, has a “sufficient statistic” property, which generalizes the “dynamic memory” property of the state variables in deterministic system theory. Much of modern statistical signal processing is based on this property.

More precisely, such a linear state-space model takes the form

$$(\Sigma) \quad \begin{cases} x(t+1) = Ax(t) + Bw(t) \\ y(t) = Cx(t) + Dw(t) \end{cases} \tag{6.2.1}$$

where  $A$  is a stability matrix (i.e., all the eigenvalues of  $A$  lie in the open unit disc), and where  $\{w(t)\}$  is a  $p$ -dimensional normalized white noise; i.e.,

$$E\{w(t)w(s)'\} = I\delta_{ts} \quad E\{w(t)\} = 0,$$

$\{x(t)\}_{t \in \mathbb{Z}}$  is the  $n$ -dimensional state process, and  $\{y(t)\}_{t \in \mathbb{Z}}$  is the  $m$ -dimensional (wide sense) stationary process to be represented. In this model both  $x$  and  $w$  are part of the representation and can be chosen in different ways. The situation is



thus drastically different than in the deterministic setting described in the previous section. This point will be clarified as we proceed.

We begin by analyzing the structure of *causal* linear state-space models, i.e., models (6.2.1) evolving forward in time, and describe them in a coordinate-free way. This description is symmetric with respect to the past and the future, and therefore we can derive symmetric state-space models, evolving backward in time. This leads to forward and backward Kalman filtering, the steady-state versions of which are state space models of particular importance.

Given a causal model (6.2.1), the process  $y$  can be thought as the output obtained by passing a white noise signal  $w$  through a linear time-invariant filter

$$\text{white noise } \xrightarrow{w} \boxed{W(z)} \xrightarrow{y}$$

with a stable transfer function

$$W(z) = C(zI - A)^{-1}B + D \tag{6.2.2}$$

for an infinitely long time so that the system is in statistical steady state. Then, since  $A$  is a stability matrix,

$$x(t) = \sum_{j=-\infty}^{t-1} A^{t-1-j} Bw(j)$$

and

$$y(t) = \sum_{j=-\infty}^{t-1} CA^{t-1-j} Bw(j) + Dw(t).$$

In particular,  $x$  and  $y$  are jointly stationary.

The system (6.2.1) can be regarded as a linear map defining  $x$  and  $y$  as linear functionals of the input noise  $w$ . In fact, since the matrix  $A$  is stable, this map is a *causal* map. In order to describe this property in a precise way it is convenient to think of the (components of)  $x(t)$  and  $y(t)$ , for all  $t$ , as elements of the infinite dimensional Hilbert space of second order random variables

$$\mathbf{H}(w) = \overline{\text{span}}\{w_i(t) \mid t \in \mathbb{Z}; i = 1, 2, \dots, p\} \tag{6.2.3}$$

with inner product  $(\xi, \eta) = E\{\xi\eta\}$ . In order to avoid too large a Hilbert space  $\mathbf{H}(w)$  we assume that the matrix

$$\begin{bmatrix} B \\ D \end{bmatrix}$$

is full column rank.

By causality, the *past subspaces* of  $x$  and  $y$

$$\mathbf{H}_t^-(x) = \overline{\text{span}}\{x_i(s) \mid s < t; i = 1, 2, \dots, n\} \tag{6.2.4}$$

$$\mathbf{H}_t^-(y) = \overline{\text{span}}\{y_i(s) \mid s < t; i = 1, 2, \dots, m\} \tag{6.2.5}$$

are both contained in  $\mathbf{H}_t^-(w)$  and hence the *future space* of  $w$

$$\mathbf{H}_t^+(w) = \overline{\text{span}}\{w_i(s) \mid s \geq t; i = 1, 2, \dots, p\} \quad (6.2.6)$$

will be orthogonal to both  $\mathbf{H}_{t+1}^-(x)$  and  $\mathbf{H}_t^-(y)$ .

Causality can thus be characterized by the orthogonality relation

$$\mathbf{H}_t^+(w) \perp (\mathbf{H}_{t+1}^-(x) \vee \mathbf{H}_t^-(y)) \quad \text{for all } t \in \mathbb{Z}. \quad (6.2.7)$$

This is equivalent to saying that  $\Sigma$  is a *forward* representation or that it evolves *forward* in time. In particular,  $E\{x(t)w(t)'\} = 0$  for all  $t \in \mathbb{Z}$ .

The family  $\{\mathbf{X}_t\}$  of finite-dimensional subspaces, defined by

$$\mathbf{X}_t = \text{span}\{x_1(t), x_2(t), \dots, x_n(t)\} \subset \mathbf{H}(w) \quad t \in \mathbb{Z}, \quad (6.2.8)$$

will play a fundamental role in this book. The subspace  $\mathbf{X}_t$  is called the *state space* of the system (6.2.1) at the instant  $t$ . Clearly  $\dim \mathbf{X}_t$  is constant by stationarity, and  $\dim \mathbf{X}_t \leq n$  with equality if and only if  $\{x_1(t), x_2(t), \dots, x_n(t)\}$  is a basis in  $\mathbf{X}_t$ . This can be characterized in terms of the *state covariance*

$$P = E\{x(0)x(0)'\}. \quad (6.2.9)$$

It follows from stationarity that  $P = E\{x(t)x(t)'\}$  for all  $t$ , and hence the first equation in (6.2.1) yields

$$P = APA' + BB', \quad (6.2.10)$$

which is a Lyapunov equation. (See Appendix A.3.) Since the matrix  $A$  has all its eigenvalues in the open unit disc, by Proposition A.3.2, we can form

$$P = \sum_{j=0}^{\infty} A^{j-1} BB'(A')^{j-1} = \mathcal{R}\mathcal{R}',$$

where  $\mathcal{R}$  is defined by (6.1.9); i.e.,  $P$  is the reachability grammian of  $\Sigma$ , and hence positive definite if and only if  $(A, B)$  is reachable.

**Proposition 6.2.1.** *The  $n$  stochastic variables  $\{x_1(t), x_2(t), \dots, x_n(t)\}$  form a basis in  $\mathbf{X}_t$  if and only if  $P > 0$ ; i.e., if and only if  $(A, B)$  is reachable.*

*Proof.* Since

$$\|a'x(t)\|^2 = a'Pa,$$

for all  $a \in \mathbb{R}^n$ , there fails to be a nonzero  $a$  such that  $a'x(t) = 0$  precisely when  $P > 0$ .  $\square$

In this chapter, we shall assume that  $(A, B)$  is reachable unless otherwise stated.

**Assumption 6.2.2.** *The pair  $(A, B)$  in  $(\Sigma)$  is reachable.*

The subspace characterizations given above suggest that the linear state-space description can be done entirely in terms of Hilbert space geometry. In fact, the property that the state-space representation (6.2.1) evolves forward in time is characterized by (6.2.7), i.e., by

$$\mathbf{S}_t \perp \mathbf{H}_t^+(w) \quad \text{where } \mathbf{S}_t := \mathbf{H}_{t+1}^-(x) \vee \mathbf{H}_t^-(y). \quad (6.2.11)$$

Therefore, since

$$\begin{aligned} x(s) &= A^{s-t}x(t) + \sum_{j=t}^{s-1} A^{s-1-j}Bw(j) \\ y(s) &= CA^{s-t}x(t) + \sum_{j=t}^{s-1} CA^{s-1-j}Bw(j) + Dw(s) \end{aligned}$$

for all  $s \geq t$ ,

$$\mathbf{E}^{\mathbf{S}_t} b' \begin{bmatrix} x(s) \\ y(s) \end{bmatrix} = b' \begin{bmatrix} A^{s-t} \\ CA^{s-t} \end{bmatrix} x(t) = \mathbf{E}^{\mathbf{X}_t} b' \begin{bmatrix} x(s) \\ y(s) \end{bmatrix}$$

for all  $b \in \mathbb{R}^{n+m}$ , from which we can deduce that

$$\mathbf{E}^{\mathbf{S}_t} \lambda = \mathbf{E}^{\mathbf{X}_t} \lambda \quad \text{for all } t \in \mathbb{Z} \text{ and } \lambda \in \bar{\mathbf{S}}_t := \mathbf{H}_t^+(x) \vee \mathbf{H}_t^+(y), \quad (6.2.12)$$

where  $\mathbf{X}_t$  is the state space (6.2.8).

Next, comparing conditions (i) and (iv) of Proposition 2.4.2 in Chapter 2 and noting that  $\mathbf{X}_t \subset \mathbf{H}_{t+1}^-(x)$ , we see that (6.2.12) is equivalent to

$$\mathbf{S}_t \perp \bar{\mathbf{S}}_t \mid \mathbf{X}_t \quad \text{for all } t \in \mathbb{Z}, \quad (6.2.13)$$

i.e., we have the following important observation.

**Proposition 6.2.3.** *The spaces  $\mathbf{S}_t := \mathbf{H}_{t+1}^-(x) \vee \mathbf{H}_t^-(y)$  and  $\bar{\mathbf{S}}_t := \mathbf{H}_t^+(x) \vee \mathbf{H}_t^+(y)$  are conditionally orthogonal given  $\mathbf{X}_t$  for each  $t \in \mathbb{Z}$ .*

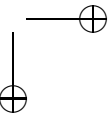
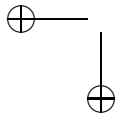
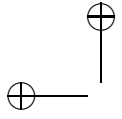
### 6.3 Anticausal state-space models

It is important to note that the conditional orthogonality condition (6.2.13) is completely symmetric with respect to reversal of time. Therefore, as can be seen from Proposition 2.4.2 in Chapter 2, (6.2.13) is equivalent not only to (6.2.12) but also to

$$\mathbf{E}^{\bar{\mathbf{S}}_t} \lambda = \mathbf{E}^{\mathbf{X}_t} \lambda \quad \text{for all } \lambda \in \mathbf{S}_t \text{ and } t \in \mathbb{Z}. \quad (6.3.1)$$

From this observation we shall now derive a linear stochastic system which, unlike (6.2.1), evolves *backwards* in time. To this end, first note that

$$\mathbf{S}_t = \mathbf{H}_t^-(z), \quad \text{where } z(t) := \begin{bmatrix} x(t+1) \\ y(t) \end{bmatrix}, \quad (6.3.2)$$



and that (6.2.1) is the same as

$$z(t) = \hat{z}(t) + v(t), \quad (6.3.3)$$

where  $\hat{z}(t)$  is the one-step predictor with components

$$\hat{z}_i(t) := \mathbf{E}^{\mathbf{H}_{t-1}^-} z_i(t), \quad i = 1, 2, \dots, n + m,$$

and  $v(t) := z(t) - \hat{z}(t)$  is the corresponding innovation process.

Next we shall use a symmetric argument to derive a backward system with state process

$$\bar{x}(t) := P^{-1}x(t+1). \quad (6.3.4)$$

In fact, it is immediately seen that

$$\bar{\mathbf{S}}_t = \mathbf{H}_t^+(\bar{z}), \quad \text{where } \bar{z}(t) := \begin{bmatrix} \bar{x}(t-1) \\ y(t) \end{bmatrix}. \quad (6.3.5)$$

Moreover, analogously to (6.3.3), we have the orthogonal decomposition

$$\bar{z}(t) = \hat{\bar{z}}(t) + \bar{v}(t), \quad (6.3.6)$$

where  $\hat{\bar{z}}(t)$  is the backward one-step predictor with components

$$\hat{\bar{z}}_i(t) := \mathbf{E}^{\mathbf{H}_{t+1}^+} \bar{z}_i(t), \quad i = 1, 2, \dots, n + m,$$

and  $\bar{v}(t) := \bar{z}(t) - \hat{\bar{z}}(t)$  is the backward innovation process, which clearly must be a white noise, i.e.,

$$\mathbf{E}\{\bar{v}(t)\bar{v}(s)'\} = \bar{V}\delta_{ts},$$

where the  $(n+m) \times (n+m)$  matrix weight  $\bar{V}$  remains to be determined.

We begin by determining  $\hat{\bar{z}}$ . To this end, observe that  $b'\bar{z}(t) \in \mathbf{S}_{t+1}$  for all  $b \in \mathbb{R}^{n+m}$ , and consequently (6.3.1) yields

$$\begin{aligned} b'\hat{\bar{z}} &= \mathbf{E}^{\bar{\mathbf{S}}_{t+1}} b'\bar{z}(t) = \mathbf{E}^{\mathbf{X}_{t+1}} b'\bar{z}(t) \\ &= b'\mathbf{E}\{\bar{z}(t)x(t+1)'\}\mathbf{E}\{x(t+1)x(t+1)'\}^{-1}x(t+1) \\ &= b' \begin{bmatrix} A' \\ CPA' + DB' \end{bmatrix} P^{-1}x(t+1), \end{aligned}$$

where we have used the projection formula of Proposition 2.2.3 and the fact that  $\mathbf{E}\{x(t)x(t+1)'\} = PA'$  and  $\mathbf{E}\{y(t)x(t+1)'\} = CPA' + DB'$ . Consequently,

$$\hat{\bar{z}}(t) = \begin{bmatrix} A' \\ \bar{C} \end{bmatrix} \bar{x}(t), \quad (6.3.7)$$

where

$$\bar{C} := CPA' + DB'. \quad (6.3.8)$$

**Theorem 6.3.1.** Consider the forward state-space model (6.2.1) with state covariance matrix

$$P := E\{x(t)x(t)'\}, \quad (6.3.9)$$

and set  $\Lambda_0 := E\{y(t)y(t)'\}$ . Then (6.3.4) is the state process of the backward system

$$(\bar{\Sigma}) \quad \begin{cases} \bar{x}(t-1) = A'\bar{x}(t) + \bar{B}\bar{w}(t) \\ y(t) = \bar{C}\bar{x}(t) + \bar{D}\bar{w}(t), \end{cases} \quad (6.3.10)$$

with state covariance

$$\bar{P} := P^{-1} = E\{\bar{x}(t)\bar{x}(t)'\}. \quad (6.3.11)$$

Here  $\bar{C}$  is given by (6.3.8),  $\bar{B}$  and  $\bar{D}$  are matrices, defined, uniquely modulo an orthogonal transformation, via a minimum-rank factorization

$$\begin{bmatrix} \bar{B} \\ \bar{D} \end{bmatrix} \begin{bmatrix} \bar{B} \\ \bar{D} \end{bmatrix}' = \begin{bmatrix} \bar{P} - A'\bar{P}A & C' - A'\bar{P}\bar{C}' \\ C - \bar{C}\bar{P}A & \Lambda_0 - \bar{C}\bar{P}C' \end{bmatrix}, \quad (6.3.12)$$

and  $\bar{w}$  is a centered, normalized white noise. The linear stochastic system (6.3.10) is a backward state-space model in the sense that

$$\mathbf{H}_t^-(\bar{w}) \perp (\mathbf{H}_t^+(x) \vee \mathbf{H}_t^+(y)) \quad \text{for all } t \in \mathbb{Z}, \quad (6.3.13)$$

which should be compared with the corresponding forward property (6.2.7).

**Proof.** In view of (6.3.7), the orthogonal decomposition (6.3.6) can be written

$$\begin{bmatrix} \bar{x}(t-1) \\ y(t) \end{bmatrix} = \begin{bmatrix} A' \\ \bar{C} \end{bmatrix} \bar{x}(t) + \bar{v}(t), \quad (6.3.14)$$

so to obtain (6.3.10) it remains to show that there are matrices  $\bar{B}$  and  $\bar{D}$  satisfying (6.3.12) such that

$$\bar{v}(t) = \begin{bmatrix} \bar{B} \\ \bar{D} \end{bmatrix} \bar{w}(t) \quad (6.3.15)$$

for some normalized white noise  $\bar{w}$ . This will be done next. In fact, from the orthogonal decomposition (6.3.14) we have

$$E\left\{ \begin{bmatrix} \bar{x}(t-1) \\ y(t) \end{bmatrix} \begin{bmatrix} \bar{x}(t-1)' & y(t)' \end{bmatrix} \right\} = \begin{bmatrix} A' \\ \bar{C} \end{bmatrix} \bar{P} \begin{bmatrix} A & \bar{C}' \end{bmatrix} + E\{\bar{v}(t)\bar{v}(t)'\},$$

and consequently, in view of (6.3.12),

$$E\{\bar{v}(t)\bar{v}(s)'\} = \begin{bmatrix} \bar{B} \\ \bar{D} \end{bmatrix} \begin{bmatrix} \bar{B} \\ \bar{D} \end{bmatrix}' \delta_{ts}. \quad (6.3.16)$$

Since, by assumption, the matrix factor has full rank, we can solve (6.3.15) uniquely for  $\bar{w}$ . In fact,

$$\bar{B}'\bar{B} + \bar{D}'\bar{D} = \begin{bmatrix} \bar{B} \\ \bar{D} \end{bmatrix}' \begin{bmatrix} \bar{B} \\ \bar{D} \end{bmatrix}$$

is nonsingular, and hence

$$\bar{w}(t) = (\bar{B}'\bar{B} + \bar{D}'\bar{D})^{-1} \begin{bmatrix} \bar{B}' \\ \bar{D}' \end{bmatrix}' \bar{v}(t), \quad (6.3.17)$$

which clearly satisfies  $E\{\bar{w}(t)\bar{w}(s)'\} = I\delta_{ts}$ . Also  $\mathbf{H}_t^-(\bar{w}) = \mathbf{H}_t^-(\bar{v})$ , which in turn is orthogonal to  $\bar{\mathbf{S}}_t = \mathbf{H}_t^+(\bar{v})$ . This establishes the backward property of (6.3.10).  $\square$

Theorem 6.3.1 shows that the process  $y$  can be regarded as the output obtained by passing a white noise signal  $\bar{w}$  backwards in time through a linear time-invariant filter<sup>21</sup>

$$\xleftarrow{y} \boxed{\bar{W}(z)} \xleftarrow{\bar{w}} \text{white noise}$$

with antistable transfer function

$$\bar{W}(z) = \bar{C}(z^{-1}I - A')^{-1}\bar{B} + \bar{D} \quad (6.3.18)$$

since  $t = +\infty$ . To say that  $\bar{W}$  is *antistable* is to say that all poles of  $\bar{W}$  lie strictly *outside* the unit circle.

We shall call the white noises  $w$  and  $\bar{w}$  the forward and backward *generating processes* corresponding to the state spaces  $\{\mathbf{X}_t\}_{t \in \mathbb{Z}}$ .

## 6.4 The forward and backward generating processes and the structural function

We have seen that to each forward model  $\Sigma$ , defined by (6.2.1), with a generating process  $w$ , there corresponds an backward model  $\bar{\Sigma}$ , defined by (6.3.10), with generating process  $\bar{w}$ , via the state transformation (6.3.4). Conversely, by a symmetric argument, we can see that to each backward model  $\bar{\Sigma}$  there is a forward model  $\Sigma$  connected via the state transformation

$$x(t) = \bar{P}^{-1}\bar{x}(t-1), \quad (6.4.1)$$

where  $\bar{P} = P^{-1}$ . We shall now investigate the relation between the generating processes  $w$  and  $\bar{w}$ .

**Theorem 6.4.1.** *Let  $(w, \bar{w})$  be the pair of generating processes of  $\Sigma$  and  $\bar{\Sigma}$ , respectively. Then, the correlation matrix*

$$V := E\{\bar{w}(t)w(t)'\}. \quad (6.4.2)$$

*satisfies the relations*

$$VV' = I - \bar{B}'P\bar{B}, \quad (6.4.3a)$$

$$V'V = I - B'\bar{P}B \quad (6.4.3b)$$

<sup>21</sup>The direction of the arrows reflects anticausality; i.e., the fact that the future of  $\bar{w}$  is mapped into the past of  $y$ .

6.4. The forward and backward generating processes and the structural function 131

Moreover,

$$\bar{w}(t) = \bar{B}'x(t) + Vw(t), \quad (6.4.4a)$$

$$w(t) = B'\bar{x}(t) + V'\bar{w}(t). \quad (6.4.4b)$$

Finally,

$$\mathbf{H}(\bar{w}) = \mathbf{H}(w). \quad (6.4.5)$$

In view of the last statement, we may define *ambient space*  $\mathbb{H}$  of the pair of models  $(\Sigma, \bar{\Sigma})$ , via

$$\mathbf{H}(\bar{w}) = \mathbb{H} = \mathbf{H}(w). \quad (6.4.6)$$

**Proof.** In view of (6.3.4) and (6.2.1), (6.3.14) can be written

$$\begin{bmatrix} \bar{P}x(t) \\ y(t) \end{bmatrix} = \begin{bmatrix} A' \\ \bar{C} \end{bmatrix} \bar{P}(Ax(t) + Bw(t)) + \bar{v}(t),$$

from which we have

$$\bar{v}(t) = \begin{bmatrix} \bar{P} - A'\bar{P}A \\ C - \bar{C}\bar{P}A \end{bmatrix} x(t) + \begin{bmatrix} -A'\bar{P}B \\ D - \bar{C}\bar{P}B \end{bmatrix} w(t).$$

This in turn equals

$$\bar{v}(t) = \begin{bmatrix} \bar{B} \\ \bar{D} \end{bmatrix} \bar{B}'x(t) + \begin{bmatrix} -A'\bar{P}B \\ D - \bar{C}\bar{P}B \end{bmatrix} w(t) \quad (6.4.7)$$

by (6.3.12). Therefore, inserting (6.4.7) into (6.3.17) yields (6.4.4a) for some matrix  $V$ . Forming  $E\{\bar{w}(t)w(t)'\}$  from (6.4.4a) and noting that  $E\{x(t)w(t)'\} = 0$ , we obtain precisely (6.4.2). In the same way, forming  $E\{\bar{w}(t)\bar{w}(t)'\}$ , we have

$$I = \bar{B}'P\bar{B} + VV',$$

which is the same as (6.4.3a). By symmetry between the systems  $\Sigma$  and  $\bar{\Sigma}$ , (6.4.3b) and (6.4.4b) follow from (6.4.2). Therefore,  $\mathbf{H}(\bar{w}) = \mathbf{H}(w)$ , in view of (6.3.4), (6.2.1) and (6.3.10).  $\square$

**Corollary 6.4.2.** *The system matrices of  $\Sigma$  are related to those of  $\bar{\Sigma}$  via*

$$AP\bar{B} + BV' = 0, \quad (6.4.8a)$$

$$CP\bar{B} + DV' = \bar{D}. \quad (6.4.8b)$$

**Proof.** Inserting (6.4.4b) into (6.2.1) we obtain

$$\begin{aligned} 0 &= AP[\bar{x}(t-1) - A'\bar{x}(t)] + BV'\bar{w}(t), \\ y(t) &= CP[\bar{x}(t-1) - A'\bar{x}(t)] + [CPA' + DB']\bar{x}(t) + DV'\bar{w}(t), \end{aligned}$$

where we have used (6.2.10) and (6.3.4) to make the substitutions  $BB' = P - APA'$ ,  $x(t + 1) = P\bar{x}(t)$ , and  $x(t) = P\bar{x}(t - 1)$ . In view of (6.3.10) and (6.3.8), we may exchange  $\bar{x}(t - 1) - A'\bar{x}(t)$  for  $\bar{B}\bar{w}(t)$  and  $CPA' + DB$  for  $\bar{C}$  to obtain

$$0 = [AP\bar{B} + BV']\bar{w}(t),$$

$$y(t) = \bar{C}\bar{x}(t) + [CP\bar{B} + DV']\bar{w}(t).$$

Postmultiplying the first equation by  $\bar{w}(t)'$  and taking expectation, we obtain (6.4.8a). Comparing the second equation to (6.3.10) yields (6.4.8b).  $\square$

From (6.4.4a) and (6.2.1), we see that  $\bar{w}$  is the output of a stable linear system

$$\text{white noise} \xrightarrow{w} \boxed{K(z)} \xrightarrow{\bar{w}} \text{white noise}$$

driven by  $w$  and with transfer function

$$K(z) = \bar{B}'(zI - A)^{-1}B + V, \quad (6.4.11)$$

which will be referred to as the *structural function* of  $(\Sigma, \bar{\Sigma})$ . Such a system, transforming white noise to white noise, is called an *all pass filter*. In particular,  $K$  is an inner function. In fact, from (6.4.4b) we also have the transfer function

$$K(z)^* = B'(z^{-1}I - A')^{-1}\bar{B} + V', \quad (6.4.12)$$

transforming  $\bar{w}$  to  $w$ , and hence  $K^{-1} = K^*$ .

**Theorem 6.4.3.** *Let  $K$  be the structural function of  $(\Sigma, \bar{\Sigma})$ , and let  $W$  and  $\bar{W}$  be the corresponding transfer functions, given by (6.2.2) and (6.3.18), respectively. Then*

$$W = \bar{W}K. \quad (6.4.13)$$

**Proof.** In view of (6.2.2) and (6.4.12),

$$\begin{aligned} (zI - A)^{-1}BK(z)^* &= (zI - A)^{-1}BB'(z^{-1}I - A')^{-1}\bar{B} + (zI - A)^{-1}BV' \\ &= P\bar{B} + PA'(z^{-1}I - A')^{-1}\bar{B} + (zI - A)^{-1}(AP\bar{B} + BV'), \end{aligned}$$

where we have made the substitutions  $BB' = P - APA'$ , in harmony with (6.2.10), and used the identity

$$P - APA' = (zI - A)P(z^{-1}I - A') + (zI - A)PA' + AP(z^{-1}I - A'), \quad (6.4.14)$$

valid for all symmetric  $P$ . Consequently, in view of (6.4.8a),

$$(zI - A)^{-1}BK(z)^* = P\bar{B} + PA'(z^{-1}I - A')^{-1}\bar{B} \quad (6.4.15)$$

$$= z^{-1}P(z^{-1}I - A')^{-1}\bar{B}, \quad (6.4.16)$$



where the second equation will be used later in Chapter 9. Therefore, in view of (6.4.12),

$$\begin{aligned} W(z)K(z)^* &= CP\bar{B} + CPA'(z^{-1}I - A')^{-1}\bar{B} + DB'(z^{-1}I - A')^{-1}\bar{B} + DV' \\ &= (CPA' + DB')(z^{-1}I - A')^{-1}\bar{B} + CP\bar{B} + DV', \end{aligned}$$

which by (6.3.8) and (6.4.8b), is the same as  $\bar{C}(z^{-1}I - A')^{-1}\bar{B} + \bar{D}$ . Then, in view of (6.3.18) and  $K^* = K^{-1}$ ,  $W = \bar{W}K$ , as claimed.  $\square$

Since the structural function  $K$  is rational and inner (all-pass), it has a matrix fraction description

$$K(z) = \bar{M}(z)M(z)^{-1}, \tag{6.4.17}$$

where  $M$  and  $\bar{M}$  are  $p \times p$  matrix polynomials with  $\det M$  having all its roots in the open unit disc and  $\det \bar{M}$  having all its roots in the complement of the closed unit disc. Since  $K^* = K^{-1}$ ,

$$M(z^{-1})'M(z) = \bar{M}(z^{-1})'\bar{M}(z). \tag{6.4.18}$$

**Corollary 6.4.4.** *Let  $K$ ,  $W$  and  $\bar{W}$  be as defined in Theorem 6.4.3, and let  $K$  have the matrix fraction description (6.4.17). Then there is an  $m \times p$  matrix polynomial  $N$  such that*

$$W(z) = N(z)M(z)^{-1}, \tag{6.4.19a}$$

$$\bar{W}(z) = N(z)\bar{M}(z)^{-1}. \tag{6.4.19b}$$

**Proof.** From (6.4.13) and (6.4.17) we obtain

$$WM = \bar{W}\bar{M},$$

which is a rational  $m \times p$  matrix function that we call  $N$ . However,  $\bar{W}\bar{M}$  is analytic in the closed unit disc and  $WM$  is analytic in the complement of the open unit disc. Hence  $N$  must be a matrix polynomial, and (6.4.19) holds.  $\square$

## 6.5 The idea of state space and coordinate-free representation

In view of (6.4.6) both the forward system (6.2.1) and the corresponding backward system (6.3.10) can be represented in the same basic Hilbert space  $\mathbb{H}$ , called the *ambient space*. Moreover,

$$\{a'x(t) \mid a \in \mathbb{R}^n\} = \mathbf{X}_t = \{a'\bar{x}(t-1) \mid a \in \mathbb{R}^n\}, \tag{6.5.1}$$

so the two systems have the same families of state spaces.

Since all random processes involved are jointly stationary, we only need to consider one instance of time, say  $t = 0$ . In fact, as explained in Chapter 2, the Hilbert space  $\mathbb{H} := \mathbf{H}(w)$  is endowed with a shift  $U$  such that

$$Uw_i(t) = w_i(t + 1), \tag{6.5.2}$$

which is inherited by the other processes. Obviously, since  $\mathbf{H}(x) = \mathbf{H}(\bar{x}) \subset \mathbb{H}$  and  $\mathbf{H}(y) \subset \mathbb{H}$ , the processes  $x$ ,  $\bar{x}$  and  $y$  are shifted by  $U$  in the same manner. Thus, for example,

$$\mathbf{X}_t = U^t \mathbf{X} \quad \text{where } \mathbf{X} = \mathbf{X}_0. \tag{6.5.3}$$

What is given in stochastic realization theory is the output process  $y$ . Therefore we shall introduce a particularly simple notation for  $\mathbf{H}(y)$  and its past and future spaces, namely

$$\mathbf{H} := \mathbf{H}(y), \quad \mathbf{H}^- := \mathbf{H}_0^-(y), \quad \mathbf{H}^+ := \mathbf{H}_0^+(y), \tag{6.5.4}$$

in terms of which

$$\mathbf{H}_t^-(y) = U^t \mathbf{H}^- \quad \text{and} \quad \mathbf{H}_t^+(y) = U^t \mathbf{H}^+. \tag{6.5.5}$$

Clearly,

$$\mathbf{H} = \mathbf{H}^- \vee \mathbf{H}^+ \subset \mathbb{H}, \tag{6.5.6}$$

and

$$U^{-1} \mathbf{H}^- \subset \mathbf{H}^- \quad \text{and} \quad U \mathbf{H}^+ \subset \mathbf{H}^+. \tag{6.5.7}$$

Likewise, given the state space  $\mathbf{X} = \mathbf{X}_0$  at  $t = 0$ , we may form

$$\mathbf{X}^- := \bigvee_{t=-\infty}^0 \mathbf{X}_t = \bigvee_{t=-\infty}^0 U^t \mathbf{X} = U \mathbf{H}_0^-(x), \tag{6.5.8a}$$

$$\mathbf{X}^+ := \bigvee_{t=0}^{\infty} \mathbf{X}_t = \bigvee_{t=0}^{\infty} U^t \mathbf{X} = \mathbf{H}_0^+(x). \tag{6.5.8b}$$

Therefore, the conditional orthogonality condition (6.2.13) can be written in the following equivalent form:

$$(\mathbf{H}^- \vee \mathbf{X}^-) \perp (\mathbf{H}^+ \vee \mathbf{X}^+) \mid \mathbf{X}. \tag{6.5.9}$$

From Lemma 2.4.1 we see that this implies that the past and the future spaces of the process  $y$  are conditionally orthogonal to the state space  $\mathbf{X}$  at  $t = 0$ , i.e.,

$$\mathbf{H}^- \perp \mathbf{H}^+ \mid \mathbf{X}. \tag{6.5.10}$$

Any subspace  $\mathbf{X}$  satisfying (6.5.10) is called a *splitting subspace* for  $y$ , and one satisfying (6.5.9) is called a *Markovian splitting subspace*. Hence, the state space  $\mathbf{X}$  of any linear stochastic system with output  $y$  is a Markovian splitting subspace for  $y$ , a concept which will be studied in depth in the next chapter, where we will prove

that determining all models (6.2.1) with output  $y$  is equivalent to determining all Markovian splitting subspaces  $\mathbf{X}$  of  $y$ .

In view of Proposition 2.4.2 in Chapter 2, a state space  $\mathbf{X}$  of a stochastic model is a subspace of  $\mathbb{H}$  with the property that

$$E^{\mathbf{H}^- \vee \mathbf{X}} \lambda = E^{\mathbf{X}} \lambda \quad \text{for all } \lambda \in \mathbf{H}^+;$$

i.e.,  $\mathbf{X}$  serves as a “memory” or “sufficient statistics” which contains everything from the past which is needed in predicting the future. Therefore, to obtain real data reduction, we shall be interested in models whose state spaces  $\mathbf{X}$  are *minimal* in the sense that they have minimal dimension.

We have thus shown that many important properties of a linear stochastic system (6.2.1) are captured in a coordinate-free manner by the family of state spaces

$$U^t \mathbf{X} = \{a'x(t) \mid a \in \mathbb{R}^n\}. \tag{6.5.11}$$

The state space  $\mathbf{X}$  is said to be *internal* if  $\mathbb{H} = \mathbf{H}$ , i.e., if  $\mathbf{X} \subset \mathbf{H}$ , the Hilbert space generated by the output process.

## 6.6 Observability, constructibility and minimality

Kalman introduced four basic systems-theoretic concepts in deterministic realization theory related to minimality: reachability, observability, controllability and (re)constructibility. Reachability and observability have been defined in Section 6.1. *Controllability and constructibility* are the corresponding concepts under time reversal, and here they refer to the backward dynamics (6.3.10).

The linear systems  $\Sigma$  and  $\bar{\Sigma}$ , given by (6.2.1) and (6.3.10), can be regarded as representations of the process  $y$  having the same state space  $\mathbf{X}$ , one evolving forward and the other backward in time. As pointed out in Section 6.2,  $\Sigma$  is reachable if and only if  $P > 0$ ; i.e., if and only if  $x(0)$  is a basis in  $\mathbf{X}$  (Proposition 6.2.1). Under our present assumptions,  $x(0)$  will always be a basis, and hence reachability will always be satisfied. For the same reasons,  $\bar{\Sigma}$  will always be controllable, i.e., reachable in the backward sense. In Chapter 8 these assumptions will be relaxed to allow for a purely deterministic state component. For now, only observability and constructibility will be needed.

We begin with a geometric characterization. Let  $\mathbf{X}$  be the state space of a linear stochastic system. An element  $\xi \in \mathbf{X}$  is said to be *unobservable* if it cannot be distinguished from zero by observing the future of  $y$ , or more precisely, if  $\xi \perp \mathbf{H}^+$ . Analogously,  $\xi \in \mathbf{X}$  is *unconstructible* if  $\xi \perp \mathbf{H}^-$ , i.e., it cannot be distinguished from zero by observing the past of  $y$ . Hence  $\mathbf{X} \cap (\mathbf{H}^+)^{\perp}$  is the unobservable and  $\mathbf{X} \cap (\mathbf{H}^-)^{\perp}$  the unconstructible subspace of  $\mathbf{X}$ .

**Definition 6.6.1.** *The state space  $\mathbf{X}$  of a linear stochastic system is observable if  $\mathbf{X} \cap (\mathbf{H}^+)^{\perp} = 0$  and constructible if  $\mathbf{X} \cap (\mathbf{H}^-)^{\perp} = 0$ .*

**Theorem 6.6.2.** *Let (6.2.1) and (6.3.10) be a pair of state-space models, one evolving forward and the other backward in time, and let  $\mathbf{X}$  be the corresponding*

state space. Then  $\mathbf{X}$  is observable if and only if

$$\bigcap_{t=0}^{\infty} \ker CA^t = 0, \tag{6.6.1}$$

and constructible if and only if

$$\bigcap_{t=0}^{\infty} \ker \bar{C}(A')^t = 0. \tag{6.6.2}$$

**Proof.** First observe that to each  $\xi \in \mathbf{X}$  there corresponds an  $a \in \mathbb{R}^n$  such that  $\xi = a'x(0)$ . Under this correspondence,  $\xi \in \mathbf{X} \cap (\mathbf{H}^+)^{\perp}$  if and only if

$$a'x(0) \perp b'y(t) \quad \text{for all } b \in \mathbb{R}^m \text{ and } t = 0, 1, 2, \dots,$$

i.e.,

$$E\{y(t)x(0)'\}a = 0 \quad \text{for } t = 0, 1, 2, \dots$$

But, since  $E\{y(t)x(0)'\}a = CA^tPa$ , this is equivalent to

$$Pa \in \bigcap_{t=0}^{\infty} \ker CA^t,$$

and consequently, since  $P$  is nonsingular,

$$\mathbf{X} \cap (\mathbf{H}^+)^{\perp} = 0 \iff \bigcap_{t=0}^{\infty} \ker CA^t = 0.$$

The proof that  $\mathbf{X}$  is constructible if and only if (6.6.2) holds is completely analogous.  $\square$

**Corollary 6.6.3.** *Let (6.2.1) and (6.3.10) be a pair of state-space models with state space  $\mathbf{X}$ , and let  $W$  and  $\bar{W}$  be the corresponding transfer functions with matrix fraction representations as in Corollary 6.4.4. Then  $\mathbf{X}$  is observable if and only if the representation*

$$W(z) = N(z)M(z)^{-1}$$

*is coprime and constructible if and only if the representation*

$$\bar{W}(z) = N(z)\bar{M}(z)^{-1}$$

*is coprime.*

The first statement in Corollary 6.6.3 is an immediate consequence of Theorem 6.6.2 and [34, p. 41] or [54, p. 439]. The second statement follows by symmetry.

The linear stochastic system (6.2.1) is called a (forward) *stochastic realization* of  $y$ . A stochastic realization is not an input-output map like a deterministic realization (see Section 6.1), but a representation of a stochastic process. Similarly, the backward linear stochastic system (6.3.10) is a backward stochastic realization. This pair of stochastic realizations of  $y$ , corresponding to the state space  $\mathbf{X}$ , is unique modulo the choice of basis in  $\mathbf{X}$  and tied together via (6.3.11),

We say that a stochastic realization is *minimal* if it has the smallest dimension among all realizations of  $y$ . The forward stochastic realization (6.2.1) is minimal if and only if the backward stochastic realization (6.3.10) is minimal.

To see how minimality relates to observability and constructibility, we form the Hankel matrix if the covariance sequence

$$\Lambda_t := E\{y(t+k)y(k)'\} = E\{y(t)y(0)'\},$$

In fact, given (6.2.1) and/or (6.3.10), it is a straight-forward calculation to show that the covariance sequence of  $y$  is given by

$$\Lambda_t = \begin{cases} CA^{t-1}\bar{C}' & \text{for } t > 0; \\ CPC' + DD' & \text{for } t = 0; \\ \bar{C}(A')^{|t|-1}C' & \text{for } t < 0. \end{cases} \quad (6.6.3)$$

Consequently, the infinite block Hankel matrix

$$\mathbf{\Gamma} := \begin{bmatrix} \Lambda_1 & \Lambda_2 & \Lambda_3 & \cdots \\ \Lambda_2 & \Lambda_3 & \Lambda_4 & \cdots \\ \Lambda_3 & \Lambda_4 & \Lambda_5 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

admits the factorization

$$\mathbf{\Gamma} = \begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \end{bmatrix} \begin{bmatrix} \bar{C} \\ \bar{C}A' \\ \bar{C}(A')^2 \\ \vdots \end{bmatrix}'. \quad (6.6.4)$$

Clearly, the dimension  $n$  of the system  $\Sigma$ , and of  $\bar{\Sigma}$ , satisfies

$$n \geq \text{rank } \mathbf{\Gamma} \quad (6.6.5)$$

with equality if and only if both (6.6.1) and (6.6.2) hold, i.e., if and only if  $\mathbf{X}$  is both observable and constructible. Since both  $(A, B)$  and  $(A', \bar{B})$  are reachable, this happens if and only if (6.2.2) is a minimal realization of  $W$  and (6.3.18) is a minimal realization of  $\bar{W}$ , both in the deterministic sense. Note that we need both these conditions. Consequently, we have proved the following theorem.

**Theorem 6.6.4.** *A stochastic realization (6.2.1) is minimal if and only if its state space  $\mathbf{X}$  is both observable and constructible.*

In view of Theorem 6.6.2 and (6.3.8), we have the following useful corollary, which shows that observability and reachability is not enough for a stochastic realization to be minimal.

**Corollary 6.6.5.** *A stochastic realization (6.2.1) is minimal if and only if*

- (i)  $(C, A)$  is observable,
- (ii)  $(A, B)$  is reachable,
- (iii)  $(CPA' + DB', A)$  is observable, where  $P$  is the solution of the Lyapunov equation  $P = APA' + BB'$ .

Obviously a stochastic realization is minimal if and only if its state space has minimal dimension. In Chapter 7 we show that this concept of minimality is equivalent to that of subspace inclusion. We say that the stochastic realization is *internal* if its state space  $\mathbf{X} \subset \mathbf{H}$ , the space generated by  $y$ .

## 6.7 The forward and the backward predictor space

Next we provide two important examples of minimal stochastic realizations that also happen to be internal. Let  $y$  be a purely nondeterministic stationary vector process with second-order statistics (6.6.3), where  $(C, A)$  and  $(\bar{C}, A')$  are observable and  $A$  is a stability matrix.

**Theorem 6.7.1.** *Let  $y$  and  $(A, C, \bar{C})$  be given as above. The predictor space*

$$\mathbf{X}_- = E^{\mathbf{H}^-} \mathbf{H}^+ \tag{6.7.1}$$

*and the backward predictor space*

$$\mathbf{X}_+ = E^{\mathbf{H}^+} \mathbf{H}^- \tag{6.7.2}$$

*are both state spaces of minimal stochastic realizations of  $y$ . In fact,  $y$  has a stochastic realization*

$$(\Sigma_-) \quad \begin{cases} x_-(t+1) = Ax_-(t) + B_-w_-(t) \\ y(t) = Cx_-(t) + D_-w_-(t) \end{cases} \tag{6.7.3}$$

*with state space  $\mathbf{X}_-$ , where the normalized white noise  $w_-$  is the forward innovation process of  $y$ ; i.e.,*

$$\mathbf{H}^-(w_-) = \mathbf{H}^-. \tag{6.7.4}$$

*Likewise,  $y$  has a backward stochastic realization*

$$(\bar{\Sigma}_+) \quad \begin{cases} \bar{x}_+(t-1) = A'\bar{x}_+(t) + \bar{B}_+\bar{w}_+(t) \\ y(t) = \bar{C}\bar{x}_+(t) + \bar{D}_+\bar{w}_+(t) \end{cases} \tag{6.7.5}$$

with state space  $\mathbf{X}_+$ , where the normalized white noise  $\bar{w}_+$  is the backward innovation process of  $y$ ; i.e.,

$$\mathbf{H}^+(\bar{w}_+) = \mathbf{H}^+. \quad (6.7.6)$$

If  $p$  is the rank of the process  $y$ ,  $D_-$  and  $\bar{D}_+$  are  $m \times p$  matrix of full column rank. In particular, if  $y$  is a full-rank process, they are square and nonsingular.

A proof could be constructed by choosing a basis  $x(t)$  in  $U^t \mathbf{X}_-$  and showing that this process is Markov. This will be done in Chapter 8. Here we shall provide a different proof.

**Proof.** Let

$$\hat{y}_k(t) := E^{\mathbf{H}_t^-} y_k(t), \quad k = 1, 2, \dots, m, \quad (6.7.7)$$

and let  $w_-$  be the normalized forward innovation process defined by (4.1.7), i.e., by (6.7.4). Then, by (4.5.16),

$$D_- w_-(t) = y(t) - \hat{y}(t), \quad (6.7.8)$$

where  $D_-$  is a full-rank matrix factor of the innovation variance

$$D_- D_-' = E\{[y(0) - \hat{y}(0)][y(0) - \hat{y}(0)]'\}.$$

By Proposition 4.5.7, the number of columns,  $p$  of  $D_-$  equals the rank of the process  $y$ , and the components of  $w_-(t)$  span  $\mathbf{H}_{t+1}^- \ominus \mathbf{H}_t^-$ .

Now, suppose that  $A$ , given by (6.6.3), is  $n \times n$ . We begin by proving that there is a stochastic vector  $\xi := (\xi_1, \xi_2, \dots, \xi_n)'$  such that

$$E\{\xi y(-t)'\} = A^{t-1} \bar{C}', \quad t = 1, 2, 3, \dots,$$

or, equivalently, that, for  $k = 1, 2, \dots, n$

$$\langle \xi_k, \eta \rangle = c_k(\eta), \quad \text{for all } \eta \in \mathbf{H}^-,$$

where  $c_k(\eta)$  is a real number formed from the  $k$ :th rows of  $A^{t-1} \bar{C}'$ ,  $t = 1, 2, 3, \dots$ , in the manner that  $\eta$  is formed from  $y(-t)$ ,  $t = 1, 2, 3, \dots$ . Now, consider the bounded linear functional  $L_k : \mathbf{H}^- \rightarrow \mathbb{R}$  that sends  $\eta$  to  $c_k(\eta)$ . Then, by Riesz' representation theorem, there is a  $\xi_k \in \mathbf{H}^-$  such that  $L_k(\eta) = \langle \xi_k, \eta \rangle$ , and hence there is a  $\xi$  with prescribed property. Defining  $x_-(t)$  to be the vector process with components  $U^t \xi_k$ ,  $k = 1, 2, \dots, n$ , we then have

$$E\{x_-(0)y(-t)'\} = A^{t-1} \bar{C}', \quad t = 1, 2, 3, \dots \quad (6.7.9)$$

In view of (6.6.3), (6.7.9) yields

$$E\{[y(0) - Cx_-(0)]y(-t)'\} = 0, \quad t = 1, 2, 3, \dots,$$

i.e.,

$$\hat{y}_k(0) = Cx_-(0), \quad (6.7.10)$$

and consequently it follows from (6.7.8) that

$$y(t) = Cx_-(t) + D_-w_-(t).$$

Analogously,

$$\mathbb{E}\{[x_-(1) - Ax_-(0)]y(-t)'\} = 0, \quad t = 1, 2, 3, \dots,$$

and hence the components of  $x_-(1) - Ax_-(0)$  belong to  $U\mathbf{H}^- \ominus \mathbf{H}^-$ . Therefore,

$$x_-(1) - Ax_-(0) = B_-w_-(0)$$

for some  $n \times m$  matrix  $B_-$ , which yield the first equation in (6.7.3) after applying the shift  $U^t$  componentwise. More generally, it follows from (6.6.3) and (6.7.9) that

$$\mathbb{E}\{[y(\tau) - CA^\tau x_-(0)]y(-t)'\} = 0, \quad t = 1, 2, 3, \dots,$$

for  $\tau = 0, 1, 2, \dots$ , and consequently

$$[CA^\tau x_-(0)]_k = \mathbb{E}^{\mathbf{H}^-} y_k(\tau) \in \mathbb{E}^{\mathbf{H}^-} \mathbf{H}^+ = \mathbf{X}_-. \quad (6.7.11)$$

Since  $(C, A)$  is observable, this implies that the components of  $x_-(0)$  belong to  $\mathbf{X}_-$ . Consequently,

$$\mathbf{X}_1 := \{a'x_-(0) \mid a \in \mathbb{R}\} \subset \mathbf{X}_-.$$

However, since  $\mathbf{X}_1$  is the state space of the stochastic realization (6.7.3), it follows from Section 6.5 and (6.5.10) that  $\mathbf{X}_-$  is a splitting subspace; i.e.,  $\mathbf{H}^- \perp \mathbf{H}^+ \mid \mathbf{X}_1$ . But  $\mathbf{X}_1 \subset \mathbf{H}^-$ , and therefore, by Theorem 2.4.3,  $\mathbf{X}_1 \supset \mathbf{X}_-$ . Hence,  $\mathbf{X}_1 = \mathbf{X}_-$ , as claimed.

In view of Theorem 6.6.2,  $\mathbf{X}_-$  is observable. Hence, by Theorem 6.6.4, it just remains to show that  $\mathbf{X}_-$  is constructible, to establish that (6.7.3) is a minimal realization of  $y$ . To this end, just note that

$$\mathbf{X}_- \cap (\mathbf{H}^+)^\perp = \mathbf{X}_- \cap \mathbf{H}^- \cap (\mathbf{H}^+)^\perp = 0,$$

since  $\mathbf{X}_-$  and  $\mathbf{H}^- \cap (\mathbf{H}^+)^\perp$  are orthogonal by Lemma 2.2.6.

The second part of the theorem follows by symmetry.  $\square$

To avoid (uninteresting) technical complications, from now on we shall make the following blanket assumption.

**Assumption 6.7.2.** *The output process  $y$  has full rank.*

**Corollary 6.7.3.** *The transfer function of  $\Sigma_-$ ,*

$$W_-(z) = C(zI - A)^{-1}B_- + D_-, \quad (6.7.12)$$

*is minimum-phase; i.e.,  $W_-(z)$  has all its poles in the open unit disc and all its zeros in the closed unit disc. Symmetrically, the transfer function of  $\bar{\Sigma}_+$ ,*

$$\bar{W}_+(z) = \bar{C}(z^{-1}I - A)^{-1}\bar{B}_+ + \bar{D}_+, \quad (6.7.13)$$



is conjugate minimum-phase; i.e.,  $\bar{W}_+(z)$  has all its poles in the complement of the closed unit disc and all its zeros in the complement of the open unit disc.

**Proof.** The stochastic system  $\Sigma_-$  is merely a state-space realization of the (normalized) innovation representation (4.1.11). Hence  $W_-(z)$  is the outer spectral factor that has no zeros outside of the closed unit disc. The statement about  $\bar{W}_+(z)$  follows by a completely symmetric argument.  $\square$

The stochastic realization  $\Sigma_-$  can also be written

$$x_-(t+1) = Ax_-(t) + B_-D_-^{-1}[y(t) - Cx_-(t)]. \quad (6.7.14)$$

As we shall see in Section 6.9,  $\Sigma_-$  can be interpreted as a steady-state Kalman filter. From (6.7.14) and the second of equations (6.7.3) we readily obtain the inverse of  $\Sigma_-$ , namely

$$\begin{cases} x_-(t+1) = \Gamma_-x_-(t) + B_-D_-^{-1}y(t) \\ w_-(t) = -D_-^{-1}Cx_-(t) + D_-^{-1}y(t), \end{cases} \quad (6.7.15)$$

where

$$\Gamma_- = A - B_-D_-^{-1}C. \quad (6.7.16)$$

By Corollary 6.7.3,  $\Gamma_-$  has all its eigenvalues in the closed unit disc.

Likewise,  $\bar{\Sigma}_+$  can be written as a backward steady-state Kalman filter:

$$\bar{x}_+(t-1) = A'\bar{x}_+(t) + \bar{B}_+\bar{D}_+^{-1}[y(t) - \bar{C}\bar{x}_+(t)], \quad (6.7.17)$$

and we obtain the inverse of  $\bar{\Sigma}_+$  as

$$\begin{cases} \bar{x}_+(t-1) = \bar{\Gamma}_+\bar{x}_+(t) + \bar{B}_+\bar{D}_+^{-1}y(t) \\ \bar{w}_+(t) = -\bar{D}_+^{-1}\bar{C}\bar{x}_+(t) + \bar{D}_+^{-1}y(t), \end{cases} \quad (6.7.18)$$

where

$$\bar{\Gamma}_+ = A' - B_-\bar{B}_+\bar{D}_+^{-1}\bar{C} \quad (6.7.19)$$

has all its eigenvalues in the complement of the open unit disc.

The forward stochastic realization  $\Sigma_-$  and the backward stochastic realization  $\bar{\Sigma}_+$  are both minimal stochastic realizations, but they have different state spaces. The predictor space  $\mathbf{X}_-$  also has a backward realization

$$(\bar{\Sigma}_-) \quad \begin{cases} \bar{x}_-(t-1) = A'\bar{x}_-(t) + \bar{B}_-\bar{w}_-(t) \\ y(t) = \bar{C}\bar{x}_-(t) + \bar{D}_-\bar{w}_-(t), \end{cases} \quad (6.7.20)$$

whose transfer function

$$\bar{W}_-(z) = \bar{C}(z^{-1}I - A')^{-1}\bar{B}_- + \bar{D}_- \quad (6.7.21)$$

has all its poles outside the closed unit disc but still its zeros inside. In fact, in view of Corollary 6.4.4,  $\bar{W}_-$  has the same zeros as  $W_-$ . In the same way, the backward

predictor space  $\mathbf{X}_+$  has a forward realization

$$(\Sigma_+) \quad \begin{cases} x_+(t+1) = Ax_+(t) + B_+w_+(t) \\ y(t) = Cx_+(t) + D_+w_+(t), \end{cases} \quad (6.7.22)$$

with a transfer function

$$W_+(z) = C(zI - A)^{-1}B_+ + D_+, \quad (6.7.23)$$

**Corollary 6.7.4.** *The transfer function (6.7.23) of  $\Sigma_+$  is maximum-phase; i.e.,  $W_+(z)$  has all its poles in the open unit disc and all its zeros in the complement of the open unit disc. Symmetrically, the transfer function of  $\bar{\Sigma}_-$ , is conjugate maximum-phase; i.e.,  $\bar{W}_-(z)$  has all its poles in the complement of the closed unit disc and all its zeros in the closed unit disc.*

**Proof.** The statements concerning the poles are trivial, since the location of the poles is determined by the eigenvalues of  $A$ . By Corollary 6.4.4,  $W_+(z)$  has the same zeros as  $\bar{W}_+(z)$ , and  $\bar{W}_-(z)$  has the same zeros as  $W_-(z)$ . Hence the statements about the zeros follow from Corollary 6.7.3.  $\square$

We have thus constructed the stochastic realizations corresponding to the forward and backward predictor spaces. In Chapter 8 we shall construct the stochastic realizations of arbitrary Markovian splitting subspaces in a more systematic fashion.

## 6.8 Rational spectral factorization

Obviously, (6.6.5) is fulfilled with equality if and only if

$$\Phi_+(z) = C(zI - A)^{-1}\bar{C}' + \frac{1}{2}\Lambda_0 \quad (6.8.1)$$

is a minimal (deterministic) realization of the rational function  $\Phi_+$  defined by the Laurent expansion

$$\Phi_+(z) = \frac{1}{2}\Lambda_0 + \Lambda_1z^{-1} + \Lambda_2z^{-2} + \dots \quad (6.8.2)$$

in an open region containing the complement of the closed unit disc. This function is the “negative tail” of the *spectral density*

$$\Phi(z) = \Phi_+(z) + \Phi_+(z^{-1})', \quad (6.8.3)$$

defined by the Laurent series

$$\Phi(z) = \sum_{t=-\infty}^{\infty} \Lambda_t z^{-t} \quad (6.8.4)$$

in an open annulus containing the unit circle. Since the spectral density must be nonnegative definite on the unit circle,  $\Phi_+$  must satisfy the positivity condition

$$\Phi_+(e^{i\theta}) + \Phi_+(e^{-i\theta})' \geq 0 \quad \theta \in [-\pi, \pi], \quad (6.8.5)$$

and, since  $A$  is stability matrix,  $\Phi_+$  has all its poles in the open unit disc. A function with these properties is called *positive real*. Therefore we shall call  $\Phi_+$  the *positive-real part* of  $\Phi$ .

**Proposition 6.8.1.** *The transfer functions (6.2.2) and (6.3.18) of  $\Sigma$  and  $\bar{\Sigma}$ , respectively, are spectral factors of  $\Phi$ , i.e.*

$$W(z)W(z^{-1})' = \Phi(z) \tag{6.8.6}$$

and

$$\bar{W}(z)\bar{W}(z^{-1})' = \Phi(z). \tag{6.8.7}$$

Since  $W$  has all its poles in the open unit disc and is finite at infinity, we say that it is a *stable* or *analytic* spectral factor, while  $\bar{W}$ , which has its poles strictly outside the unit circle, is said to be *coanalytic*. There is complete symmetry between the forward and the backward settings, and therefore we only consider  $W$ . We provide a purely algebraic derivation that does not require stability of  $A$  and stationarity of the processes involved. (Of course, in this case the “spectrum”  $\Phi(z)$  does not need to have a probabilistic meaning). It only requires existence of a solution to the Lyapunov equation  $P = APA' + BB'$ , which in our case is guaranteed, since  $A$  is stable.

To this end, we shall use an algebraic decomposition based on the useful identity (6.4.14), i.e.,

$$P - APA' = (zI - A)P(z^{-1}I - A') + (zI - A)PA' + AP(z^{-1}I - A'),$$

to the product  $W(z)W(z^{-1})'$ , a well-known trick from Kalman-Yakubovich theory. In fact, a straightforward calculation shows that

$$\begin{aligned} W(z)W(z^{-1})' &= [C(zI - A)^{-1}B + D][B'(zI - A')^{-1}C' + D'] \\ &= C(zI - A)^{-1}BB'(z^{-1}I - A')^{-1}C' \\ &\quad + C(zI - A)^{-1}BD' + DB'(z^{-1}I - A)^{-1}C' + DD' \end{aligned}$$

so, in view of (6.2.10) and (6.4.14),

$$\begin{aligned} W(z)W(z^{-1})' &= CPC' + DD' + \\ &\quad + C(zI - A)^{-1}(APC' + BD') + (CPA' + DB')(z^{-1}I - A')^{-1}C' \\ &= \Phi_+(z) + \Phi_+(z^{-1})', \end{aligned} \tag{6.8.8}$$

where the last equality follows from (6.3.8) and the expression for  $\Lambda_0$  in (6.6.3), thus establishing (6.8.6).

**Proposition 6.8.2.** *Let  $W$  be an arbitrary rational analytic spectral factor of  $\Phi$ . Then*

$$\deg W \geq \deg \Phi_+, \tag{6.8.9}$$

where  $\deg$  denotes McMillan degree.

**Proof.** Suppose  $(A, B, C, D)$  is a minimal realization (6.2.2) of  $W$ . Then, if  $A$  is  $n \times n$ ,  $\deg W = n$ . Moreover, as shown in (6.8.8), the positive real part of  $\Phi$  takes the form

$$\Phi_+(z) = C(zI - A)^{-1}\bar{C}' + \frac{1}{2}\Lambda_0,$$

which clearly has degree less or equal to  $n$ .  $\square$

Since, in this proof,  $(C, A)$  is supposed to be observable, there can be strict inequality in (6.8.9) only if the pair  $(A, \bar{C}')$  is not reachable.

**Definition 6.8.3.** The spectral factor  $W$  of  $\Phi$  is *minimal* if

$$\deg W = \deg \Phi_+.$$

As we have seen, minimal spectral factors always do exist. Well-known examples of minimal spectral factor are the *minimum phase* and the *maximum phase* spectral factors, denoted  $W_-$  and  $W_+$  respectively. Both  $W_-$  and  $W_+$  are analytic, but, as we have already seen, the former has no zeros outside of the closed unit disk while the second has instead no zeros inside the open unit disk. It follows from Theorem 4.6.8 that all analytic (stable) rational spectral factors can be constructed by postmultiplying the minimum phase factor  $W_-$  by a inner rational matrix function  $Q(z)$ ; i.e., an analytic rational matrix function such that

$$Q(z)Q(z^{-1})' = I.$$

More generally, if  $W(z)$  is a spectral factor and  $Q(z)$  is inner, then

$$W_1(z) = W(z)Q(z)$$

is also a spectral factor. If there are no cancellation of poles and zeros,  $\deg W_1 > \deg W$ ; i.e.,  $W_1$  is nonminimal.

### The converse problem

Let us now consider the converse problem. Given a rational spectral density  $\Phi$ , i.e. an  $m \times m$  rational matrix function that is *parahermitian*; i.e., satisfies

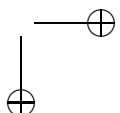
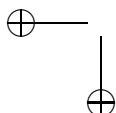
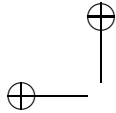
$$\Phi(z^{-1}) = \Phi(z)',$$

that is positive semidefinite on the unit circle, consider the problem of finding all minimal analytic spectral factors  $W$  and the corresponding (minimal) realizations

$$W(z) = H(zI - F)^{-1}B + D. \tag{6.8.10}$$

To solve this problem, first make the decomposition

$$\Phi(z) = \Phi_+(z) + \Phi_+(z^{-1})', \tag{6.8.11}$$



where  $\Phi_+$  has all its poles in the open unit disk (so that it is the positive-real part of  $\Phi$ ), and then compute a minimal realization

$$\Phi_+(z) = C(zI - A)^{-1}\bar{C}' + \frac{1}{2}\Lambda_0. \quad (6.8.12)$$

Then  $A$  is a stability matrix, and, if  $A$  is  $n \times n$ ,  $\deg \Phi_+ = n$ . We shall solve the spectral factorization equation (6.8.6), providing a procedure for determining  $(F, H, B, D)$  from the given matrices  $(A, C, \bar{C}, \Lambda_0)$ .

To this end first note that, in view of (6.8.11) and (6.8.12), the spectral density  $\Phi$  may be written

$$\Phi(z) = [C(zI - A)^{-1} \quad I] \begin{bmatrix} 0 & \bar{C}' \\ \bar{C} & \Lambda_0 \end{bmatrix} \begin{bmatrix} (z^{-1}I - A')^{-1}C' \\ I \end{bmatrix}. \quad (6.8.13)$$

However, in view of the identity (6.4.14), which holds for all symmetric  $P$  and all  $z \in \mathbb{C}$ ,

$$[C(zI - A)^{-1} \quad I] \begin{bmatrix} P - APA' & -APC' \\ -CPA' & -CPC' \end{bmatrix} \begin{bmatrix} (z^{-1}I - A')^{-1}C' \\ I \end{bmatrix} \equiv 0, \quad (6.8.14)$$

which added to (6.8.13) yields

$$\Phi(z) = [C(zI - A)^{-1} \quad I] M(P) \begin{bmatrix} (z^{-1}I - A')^{-1}C' \\ I \end{bmatrix}, \quad (6.8.15)$$

where

$$M(P) = \begin{bmatrix} P - APA' & \bar{C}' - APC' \\ \bar{C} - CPA' & \Lambda_0 - CPC' \end{bmatrix}. \quad (6.8.16)$$

Therefore, if there is a  $P$  satisfying the *linear matrix inequality*

$$M(P) \geq 0, \quad (6.8.17)$$

$M(P)$  can be factored as

$$M(P) = \begin{bmatrix} B \\ D \end{bmatrix} [B' \ D'], \quad (6.8.18)$$

which, inserted into (6.8.15), implies that the rational function

$$W(z) := C(zI - A)^{-1}B + D \quad (6.8.19)$$

satisfies the spectral factorization equation

$$W(z)W(z^{-1})' = \Phi(z). \quad (6.8.20)$$

Hence, in particular, we can choose  $F = A$  and  $H = C$  in (6.8.10). Since  $\deg W \leq n$ , being equal to  $n$  if and only if  $(A, B)$  is reachable,  $\deg W \leq \deg \Phi_+$ . But, in view of (6.8.9), this implies that  $\deg W = \deg \Phi_+$ . Consequently,  $W$  is a minimal spectral factor.

Let  $\mathcal{P}$  be the set of all symmetric matrices  $P$  such that (6.8.17) holds. The existence of minimal spectral factors is connected to the question of whether  $\mathcal{P}$  is nonempty. The following fundamental result, which is a corollary of the important Kalman-Yakubovich-Popov Lemma, clarifies this point.

**Theorem 6.1 (The Positive Real Lemma).** *Let  $\Phi_+$  be a stable  $m \times m$  transfer function with a minimal realization*

$$\Phi_+(z) = C(zI - A)^{-1}\bar{C}' + \frac{1}{2}\Lambda_0. \tag{6.8.21}$$

*More precisely, let  $A$  be a stable  $n \times n$  matrix, and suppose that  $(C, A)$  and  $(\bar{C}, A')$  are both observable. Moreover, let  $M : \mathbb{R}^{n \times n} \rightarrow \mathbb{R}^{(n+m) \times (n+m)}$  be the linear map defined by (6.8.16). Then, the set  $\mathcal{P}$  of all symmetric matrices  $P$  such that the linear matrix inequality*

$$M(P) \geq 0 \tag{6.8.22}$$

*holds is nonempty if and only if  $\Phi_+$  is positive real. Finally, any  $P \in \mathcal{P}$  is positive definite.*

Therefore (6.8.18), i.e.,

$$P = APA' + BB' \tag{6.8.23a}$$

$$\bar{C} = CPA' + DB' \tag{6.8.23b}$$

$$\Lambda_0 = CPC' + DD' \tag{6.8.23c}$$

are often called the *positive-real-lemma equations*. Note that (6.8.23a) is the Lyapunov equation (6.2.10) and (6.8.23b) is the definition (6.3.8) of  $\bar{C}$ .

**Proof.** If  $\mathcal{P}$  is nonempty, there is a  $P$  such that  $M(P) \geq 0$ , and then it follows from (6.8.11) and (6.8.15) that

$$\Phi_+(e^{i\theta}) + \Phi_+(e^{-i\theta}) \geq 0, \quad \text{for all } \theta,$$

and consequently that  $\Phi_+$  is positive real. Conversely, suppose  $\Phi_+$  is positive real. Then, there is a stochastic realization (6.7.3) having the covariance structure (6.6.3). The corresponding covariance matrix

$$P_- := E\{x_-(0)x_-(0)'\}$$

clearly satisfies (6.8.17), and therefore  $\mathcal{P}$  is nonempty.

As we have established above, any  $P \in \mathcal{P}$  must satisfy the Lyapunov equation (6.8.23a) for some  $B$  such that  $(A, B)$  is reachable, and therefore, since  $A$  is a stability matrix, any  $P \in \mathcal{P}$  must be positive definite.  $\square$

Since the state-space basis can be chosen so that  $H = C$  and  $F = A$ , the problem of determining the minimal spectral factors can thus be reduced to finding

$P \in \mathcal{P}$  and then factoring  $M(P)$  as in (6.8.18) to obtain  $B$  and  $D$ . To avoid redundancy we shall require that  $\begin{bmatrix} B \\ D \end{bmatrix}$  is full column rank. Then, for each  $P \in \mathcal{P}$ , the factorization problem (6.8.18) yields a pair  $(B, D)$ , which is unique modulo an orthogonal transformation.

**Theorem 6.8.4.** *Let  $\Phi$  be a full-rank spectral density, and let  $A, C, \bar{C}, \Lambda_0$  be matrices such that (6.8.12) is a minimal realization of  $\Phi_+$ , the positive real part of  $\Phi$ . Then there is a one-to-one correspondence between minimal analytic spectral factors of  $\Phi$  and symmetric  $n \times n$  matrices  $P$  solving the Linear Matrix Inequality (6.8.17) in the following sense: Corresponding to each solution  $P = P'$  of (6.8.17), necessarily positive definite, there corresponds a minimal analytic spectral factor (6.8.19) where  $A$  and  $C$  are as defined above and  $\begin{bmatrix} B \\ D \end{bmatrix}$  is the unique (modulo orthogonal transformations) full-rank factor (6.8.18) of  $M(P)$ . Conversely, to each analytic minimal spectral factor (6.8.10) of  $\Phi(z)$  there is a  $P \in \mathcal{P}$  so that  $(B, D)$  is obtained from (6.8.18) and  $F = A, H = C$ .*

**Proof.** It remains to prove the converse statement that to each minimal analytic spectral factor  $W$  there corresponds a  $P \in \mathcal{P}$  with the stated properties. Let  $W$  have a minimal realization (6.8.10), and let  $P$  be the unique symmetric solution to the Lyapunov equation

$$P = FPF' + BB'$$

Since  $W$  is minimal,  $A$  and  $F$  have the same dimensions. A calculation such as in (6.8.8) then shows that

$$\Phi_+(z) = H(zI - F)^{-1}G + \frac{1}{2}\Lambda_0,$$

where  $G = FPH' + BD'$  and  $\Lambda_0 = HPH' + DD'$ . Hence, in view of (6.8.12), there is a nonsingular  $n \times n$  matrix  $T$  such that

$$(H, F, G) = (CT^{-1}, TAT^{-1}, T\bar{C}')$$

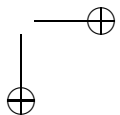
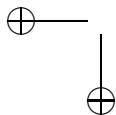
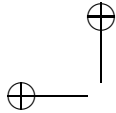
Here we may clearly choose basis in state space so that  $T = I$ , thereby obtaining the required result.  $\square$

**Example 6.8.5.** Consider a spectral density  $\Phi(z)$  with the positive real part

$$\Phi_+(z) = \frac{\frac{5}{3}}{z - \frac{1}{2}} + \frac{7}{6}.$$

Then  $A = \frac{1}{2}, \bar{C} = \frac{5}{3}, C = 1$ , and  $\Lambda_0 = \frac{7}{3}$ , and therefore the linear matrix inequality (6.8.17) becomes

$$M(P) = \begin{bmatrix} \frac{3}{4}P & \frac{5}{3} - \frac{1}{2}P \\ \frac{5}{3} - \frac{1}{2}P & \frac{7}{3} - P \end{bmatrix} \geq 0,$$



which holds if and only if  $P > 0$ ,  $\frac{7}{3} - P > 0$  and

$$\det M(P) = -P^2 + \frac{41}{12}P - \frac{25}{9} = -(P - \frac{4}{3})(P - \frac{25}{12}) \geq 0.$$

These inequalities hold precisely for  $P \in [\frac{4}{3}, \frac{25}{12}]$ , and hence  $\mathcal{P}$  is the interval  $[P_-, P_+]$ , where  $P_- = \frac{4}{3}$  and  $P_+ = \frac{25}{12}$ .

Since

$$M(P_-) = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix},$$

$P = P_-$  yields  $B = 1$  and  $D = 1$  and the minimal spectral factor

$$W_-(z) = \frac{1}{z - \frac{1}{2}} + 1 = \frac{z + \frac{1}{2}}{z - \frac{1}{2}},$$

which clearly is minimum phase. On the other hand,

$$M(P_+) = \begin{bmatrix} 25/16 & 5/8 \\ 5/8 & 1/4 \end{bmatrix}$$

yielding  $B = \frac{5}{4}$  and  $D = \frac{1}{2}$  and the maximum phase spectral factor

$$W_+(z) = \frac{\frac{5}{4}}{z - \frac{1}{2}} + \frac{1}{2} = \frac{1 + \frac{1}{2}z}{z - \frac{1}{2}}.$$

Finally, let us take a  $P$  in the interior of  $\mathcal{P}$ . With  $P = 2 \in [\frac{4}{3}, \frac{25}{12}]$  we obtain

$$M(P) = \begin{bmatrix} 3/2 & 2/3 \\ 2/3 & 1/3 \end{bmatrix}.$$

Without restriction we may take

$$\begin{bmatrix} B \\ D \end{bmatrix} = \begin{bmatrix} b_1 & b_2 \\ d & 0 \end{bmatrix}$$

and then

$$\begin{bmatrix} B \\ D \end{bmatrix} \begin{bmatrix} B \\ D \end{bmatrix}' = \begin{bmatrix} b_1^2 + b_2^2 & b_1 d \\ b_1 d & d^2 \end{bmatrix} = \begin{bmatrix} 3/2 & 2/3 \\ 2/3 & 1/3 \end{bmatrix}$$

which may be solved to yield  $d = \frac{1}{\sqrt{3}}$ ,  $b_1 = \frac{2}{\sqrt{3}}$  and, choosing one root,  $b_2 = \frac{1}{\sqrt{6}}$ , thus defining a rectangular spectral factor

$$W(z) = \left( \frac{\frac{2}{\sqrt{3}}}{z - \frac{1}{2}} + \frac{1}{\sqrt{3}}, \frac{\frac{1}{\sqrt{6}}}{z - \frac{1}{2}} \right).$$

In this example all minimal spectral factors, except  $W_-$  and  $W_+$  which are scalar, are  $1 \times 2$  matrix valued.



Obviously there is a completely symmetric factorization theory for coanalytic spectral factors  $\bar{W}$ , corresponding to backward stochastic realizations. This symmetry can be highlighted by writing the linear matrix inequality (6.8.17) as

$$\begin{bmatrix} P & \bar{C}' \\ \bar{C} & \Lambda_0 \end{bmatrix} - \begin{bmatrix} A \\ C \end{bmatrix} P \begin{bmatrix} A' & C' \end{bmatrix} \geq 0, \tag{6.8.24}$$

which, since  $P$  is positive definite, is equivalent to

$$\begin{bmatrix} P & \bar{C}' & A \\ \bar{C} & \Lambda_0 & C \\ A' & C' & P^{-1} \end{bmatrix} \geq 0. \tag{6.8.25}$$

In fact, (6.8.24) is the Schur complement (Appendix A.3) of (6.8.25). Taking the lower Schur complement instead and observing that  $\bar{P} := P^{-1}$ , we see that (6.8.25) is also equivalent to

$$\bar{M}(\bar{P}) := \begin{bmatrix} \bar{P} - A' \bar{P} A & C' - A' \bar{P} \bar{C}' \\ C - \bar{C}' \bar{P} A & \Lambda_0 - \bar{C}' \bar{P} C' \end{bmatrix} \geq 0. \tag{6.8.26}$$

From this linear matrix inequality,  $\bar{B}$  and  $\bar{D}$  are determined via factorization as in (6.3.12), yielding the coanalytic spectral factor

$$\bar{W}(z) = \bar{C}(z^{-1}I - A')^{-1} \bar{B} + \bar{D}. \tag{6.8.27}$$

## 6.9 The Riccati inequality and Kalman filtering

We have shown that the family of minimal spectral factors can be parameterized by the solutions of the linear matrix inequality (6.8.17). Next, we want to show that there is a more compact characterization of the set  $\mathcal{P}$ , namely by means of the Riccati inequality, which is of dimension  $n$  instead of  $n + m$  as with the linear matrix inequality.

Recall that in characterizing the *analytic* spectral factors, we keep the poles fixed. Roughly speaking, therefore, the spectral factors will differ by the zeros, which may be flipped to the reciprocal locations in the complex plane to create new spectral factors. While zeros located either at zero or at  $z = \infty$  can be handled by the linear matrix inequality, this will not be possible if we want to do parametrization by the Riccati inequality. As we will see in Chapter ??, the Riccati inequality is a device capable of flipping only *finite* zeros to their *finite* reciprocals. Hence to proceed in this direction we need to exclude the presence of zeros either at  $z = 0$  or at  $z = \infty$  in the spectral factors. The condition that precludes this is a condition on the spectral density matrix of the process, called *regularity*, (Corollary 9.3.10) and is studied in detail in Sections 8.8 and 9.3. A process  $y$  having a spectral density  $\Phi$  with this property is called a *regular process* and is described by the condition

$$\Delta(P) := \Lambda_0 - CPC' > 0 \quad \text{for all } P \in \mathcal{P}. \tag{6.9.1}$$

Clearly, a regular process must be full rank.

Example 6.8.5 illustrates the fact that the number of columns of the spectral factor  $W(z)$  varies with  $P \in \mathcal{P}$ . In fact, if we agree to keep the rank of  $\begin{bmatrix} B' & D' \\ 0 & I \end{bmatrix}$  full,  $W(z)$  is  $m \times p$ , where  $p := \text{rank } M(P)$ . Then, if  $T := -(\bar{C}' - APC')\Delta(P)^{-1}$ , a straight-forward calculation yields

$$\begin{bmatrix} I & T \\ 0 & I \end{bmatrix} M(P) \begin{bmatrix} I & 0 \\ T' & I \end{bmatrix} = \begin{bmatrix} -R(P) & 0 \\ 0 & \Delta(P) \end{bmatrix},$$

where

$$R(P) = APA' - P + (\bar{C}' - APC')\Delta(P)^{-1}(\bar{C}' - APC')'. \quad (6.9.2)$$

Hence,  $P \in \mathcal{P}$  if and only if it satisfies the *Riccati inequality*

$$R(P) \leq 0. \quad (6.9.3)$$

Moreover,

$$p = \text{rank } M(P) = m + \text{rank } R(P) \geq m.$$

If  $P$  satisfies the *algebraic Riccati equation*  $R(P) = 0$ , i.e.,

$$P = APA' + (\bar{C}' - APC')\Delta(P)^{-1}(\bar{C}' - APC')', \quad (6.9.4)$$

$\text{rank } M(P) = m$ . Then the spectral factor  $W$  is  $m \times m$ . The family of  $P$  corresponding to square spectral factors form a subfamily  $\mathcal{P}_0$  of  $\mathcal{P}$ . If  $P \notin \mathcal{P}_0$ ,  $W$  is rectangular. In Chapter 8 we show that  $\mathcal{P}_0$  corresponds to internal state spaces. It is obvious that  $\mathcal{P}$  is closed and convex. Next we shall demonstrate that it is also bounded with a minimum and maximum element.

We shall now demonstrate that the steady-state Kalman filter for an arbitrary minimal (forward) stochastic realization (6.2.1) is in fact itself a minimal realization, namely the forward system corresponding to the predictor space  $\mathbf{X}_-$ . To this end, given a minimal stochastic realization  $\Sigma$  defined by (6.2.1), let us consider a Kalman filter initiated at time  $\tau$  (rather than time 0) and define

$$\mathbf{H}_{[\tau,t]}(y) = \text{span}\{a'y(k) \mid a \in \mathbb{R}^m, k = \tau, \tau + 1, \dots, t\}$$

and the state estimate

$$\hat{x}_k(t) = \mathbf{E}^{\mathbf{H}_{[\tau,t-1]}} x_k(t), \quad k = 1, 2, \dots, n.$$

Then, by Lemma 2.2.4, we obtain the orthogonal decomposition

$$\hat{x}(t+1) = A\hat{x}(t) + \mathbf{E}\{x(t+1)\tilde{y}(t)'\}(\mathbf{E}\{\tilde{y}(t)\tilde{y}(t)'\})^{-1}\tilde{y}(t),$$

where  $\tilde{y}$  is the innovation process

$$\tilde{y}(t) = y(t) - C\hat{x}(t).$$

Consequently, recalling that the filter is initiated at  $t = \tau$ , we have the *Kalman filter*

$$\hat{x}(t+1) = A\hat{x}(t) + K(t-\tau)[y(t) - C\hat{x}(t)]; \quad \hat{x}(\tau) = 0, \quad (6.9.5)$$

where

$$K(t - \tau) = E\{x(t + 1)\tilde{y}(t)'\} (E\{\tilde{y}(t)\tilde{y}(t)'\})^{-1}. \quad (6.9.6)$$

**Proposition 6.9.1.** *The gain function  $K$  in the Kalman filter (6.9.5) is given by*

$$K(t) = (\bar{C}' - A\Pi(t)C')\Delta(\Pi(t))^{-1}, \quad (6.9.7)$$

where  $P \mapsto \Delta(P)$  is the matrix function (6.9.1), and  $\Pi(t)$  is the solution of the matrix Riccati equation

$$\Pi(t + 1) = \Pi(t) + R(\Pi(t)) \quad \Pi(0) = 0, \quad (6.9.8)$$

where the matrix function  $P \mapsto R(P)$  is given by (6.9.2).

**Proof.** The usual orthogonality arguments yield

$$E\{\tilde{y}(t)\tilde{y}(t)'\} = E\{y(t)[y(t) - \hat{y}(t)]'\} = \Lambda_0 - C\Pi(t - \tau)C',$$

where

$$\Pi(t) := E\{\hat{x}(t + \tau)\hat{x}(t + \tau)'\}. \quad (6.9.9)$$

In the same way,

$$E\{x(t + 1)\tilde{y}(t)'\} = AE\{x(t)[x(t) - \hat{x}(t)]'\}C' + BD = \bar{C}' - A\Pi(t - \tau)C',$$

where we have also used (6.3.8). Therefore (6.9.7) is a direct consequence of (6.9.6). Moreover, from (6.9.5), we have

$$\Pi(t + 1) = A\Pi(t)A' + K(t)\Delta(\Pi(t))K(t)',$$

which, in view of (6.9.7), yields (6.9.8).  $\square$

It is interesting and important to observe that the filter equations depend only on quantities which pertain to the positive real part (6.8.2) of the spectral density of  $y$  and not in any way to the particular choice of system (6.2.1). In fact, all  $\Pi(t)$  are lower bounds of any  $P \in \mathcal{P}$ .

**Lemma 6.9.2.** *Let  $\{\Pi(t)\}_{t \in \mathbb{Z}_+}$  be the solution of the matrix Riccati equation (6.9.8). Then*

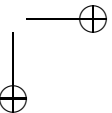
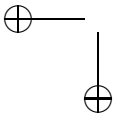
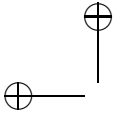
$$P \geq \Pi(t + 1) \geq \Pi(t) \geq 0 \quad \text{for all } P \in \mathcal{P} \text{ and } t = 0, 1, 2, \dots$$

**Proof.** A straightforward calculation shows that

$$P - \Pi(t) = E\{[x(t + \tau) - \hat{x}(t + \tau)][x(t + \tau) - \hat{x}(t + \tau)]'\} \geq 0,$$

which proves that  $P \geq \Pi(t)$  for all  $t \geq 0$ . To see that  $\Pi(t + 1) \geq \Pi(t)$ , first observe that, by joint stationarity, the stochastic vector  $z(t)$  with components

$$z_k(t) := E^{\mathbf{H}_{[\tau-1, t-1]}(y)} x_k(t + \tau), \quad k = 1, 2, \dots, n$$



has the same covariance matrix as  $\hat{x}(t+\tau+1)$ , i.e.,  $E\{z(t)z(t)'\} = \Pi(t+1)$ . However, since  $\mathbf{H}_{[\tau,t-1]} \subset \mathbf{H}_{[\tau-1,t-1]}$ ,

$$\hat{x}_k(t+\tau) = E^{\mathbf{H}_{[\tau-1,t-1]}(y)} z_k(t), \quad k = 1, 2, \dots, n$$

and consequently,  $\Pi(t) \leq \Pi(t+1)$ .  $\square$

This lemma shows that  $\{\Pi(t)\}_{t \in \mathbb{Z}_+}$  is monotonely nondecreasing and bounded from above, and consequently  $\Pi(t)$  tends to a limit  $P_-$  as  $t \rightarrow \infty$ . In view of (6.9.8),  $P_-$  is a solution of the algebraic Riccati equation (6.9.4), i.e.,  $R(P_-) = 0$ . Hence  $P_- \in \mathcal{P}_0$ , and therefore it must correspond to a realization (6.2.1) of  $y$ .

**Theorem 6.9.3.** *The solution  $\Pi(t)$  of the matrix Riccati equation (6.9.8) tends monotonely to a limit  $P_- \in \mathcal{P}_0$  as  $t \rightarrow \infty$ , which is the state covariance*

$$P_- := E\{x_-(0)x_-(0)'\} \quad (6.9.10)$$

of the stochastic realization (6.7.3) whose state space is the predictor space  $\mathbf{X}_-$ , defined by (6.7.1). The matrices  $B_-$  and  $D_-$  in (6.7.3) are given by

$$B_- = (\bar{C}' - AP_-C')\Delta(P_-)^{-\frac{1}{2}} \quad \text{and} \quad D_- = \Delta(P_-)^{\frac{1}{2}}. \quad (6.9.11)$$

Moreover,

$$a'x_-(t) = E^{\mathbf{H}_t^-} a'x(t) \quad \text{for all } a \in \mathbb{R}^n, \quad (6.9.12)$$

and  $P_-$  is the minimum element of the family  $\mathcal{P}$  in the sense that

$$P \geq P_- \quad \text{for all } P \in \mathcal{P}. \quad (6.9.13)$$

**Proof.** Let  $t$  be fixed and let  $\tau$  tend to  $-\infty$  in the Kalman filter

$$\hat{x}(t+1) = A\hat{x}(t) + K(t-\tau)[y(t) - C\hat{x}(t)]$$

We have already shown above that  $\Pi(t-\tau) \rightarrow P_- \in \mathcal{P}_0$  as  $\tau \rightarrow -\infty$ , and hence, in view of (6.9.7),  $K(t-\tau) \rightarrow B_-D_-^{-\frac{1}{2}}$  with  $B_-, D_-$  given by (6.9.11). Therefore, if we can show that, for all  $a \in \mathbb{R}^n$ ,

$$E^{\mathbf{H}_{[\tau,t-1]}(y)} a'x(t) \rightarrow a'x_-(t) := E^{\mathbf{H}_t^-} a'x(t) \quad (6.9.14)$$

as  $\tau \rightarrow \infty$ , then we obtain in the limit the steady state Kalman filter

$$x_-(t+1) = Ax_-(t) + B_-D_-^{-1}[y(t) - Cx_-(t)].$$

To prove (6.9.14), we shall need the following lemma.

**Lemma 6.2.** *Let  $\mathbf{A}_1 \subset \mathbf{A}_2 \subset \mathbf{A}_3 \subset \dots$  be an infinite sequence of subspaces in a Hilbert space  $\mathbf{H}$ , and let  $\xi \in \mathbf{H}$ . Then, setting  $\mathbf{A}_\infty := \bigvee_{j=0}^n \mathbf{A}_j$ , the sequence*

$$\xi_j := E^{\mathbf{A}_j} \xi \rightarrow E^{\mathbf{A}_\infty} \xi$$

strongly as  $j \rightarrow \infty$ .

**Proof.** Since

$$\|\xi_1\| \leq \|\xi_2\| \leq \|\xi_3\| \leq \dots \leq \|\xi\|$$

$\|\xi_j\|$  tends to a limit as  $j \rightarrow \infty$ . Now, for  $i < j$ ,  $\xi_i = E^{A^i} \xi_j$ , and hence

$$\|\xi_j - \xi_i\|^2 = \|\xi_j\|^2 - \|\xi_i\|^2 \rightarrow 0 \quad \text{as } i, j \rightarrow \infty.$$

Consequently,  $\xi_j$  is a Cauchy sequence and hence tends strongly to a limit in  $\mathbf{H}$ , which then must be  $E^{A^\infty} \xi$ .  $\square$

Now, (6.9.14) follows immediately by taking  $\mathbf{A}_j := \mathbf{H}_{[-j, t-1]}(y)$  in this lemma, in which case  $\mathbf{A}_\infty = \mathbf{H}_t^-$ . Moreover,  $\tilde{y}(t) \rightarrow \nu_\infty(t)$ , where  $\nu_\infty := y - Cx_-$  is a white noise such that

$$E\{\nu_\infty(t)\nu_\infty(s)'\} = \Delta(P_-)\delta_{ts} = D_-D'_-\delta_{ts},$$

so defining  $w_-(t) := D_-^{-\frac{1}{2}}\nu_\infty(t)$ , we obtain precisely (6.7.3), whose state space is  $\mathbf{X}_-$  by Theorem 6.7.1.

Finally, by Lemma 6.9.2,  $P \geq P_-$ , which establishes (6.9.13).  $\square$

We can also show that  $\mathcal{P}$  has a maximum element  $P_+ \in \mathcal{P}_0$  connected to the backward predictor space  $\mathbf{X}_+$ . To this end, construct a backward Kalman filter based on the backward model (6.3.10). An analysis that is completely symmetric to the one presented above, projecting over the future, yields a backward steady state Kalman filter (6.7.17), which can be written as a backward stochastic realization (6.7.5) with Markovian splitting subspace  $\mathbf{X}_+$  and with state covariance  $\bar{P}_+$  such that

$$\bar{P} \geq \bar{P}_+.$$

In terms of the corresponding forward systems (6.2.1) and (6.7.22), this may be written

$$P^{-1} \geq P_+^{-1}$$

(see Theorem 6.3.1). In other words, there is a  $P_+ \in \mathcal{P}_0$  such that

$$P \leq P_+ \quad \text{for all } P \in \mathcal{P}. \tag{6.9.15}$$

In fact,

$$P_+ := E\{x_+(0)x_+(0)'\}, \tag{6.9.16}$$

where  $x_+$  is the state process of (6.7.22), the forward stochastic realization of  $\mathbf{X}_+$ .

Consequently we have established the existence of two elements in  $\mathcal{P}_0$ , namely  $P_-$  and  $P_+$ , such that

$$P_- \leq P \leq P_+ \quad \text{for all } P \in \mathcal{P}. \tag{6.9.17}$$

This also establishes the boundedness of the set  $\mathcal{P}$ .

By Lemma 6.9.2,  $\Pi(t)$  approaches  $P_-$  from outside  $\mathcal{P}$  starting at  $\Pi(0) = 0$ . It can be shown that all the elements in  $\mathcal{P}_0$  are extreme points of  $\mathcal{P}$ . The converse is often, but not always, true. In Example 6.8.5, we now have  $P_- = \frac{4}{3}$ ,  $P_+ = \frac{25}{12}$  and  $\mathcal{P}_0 = \{\frac{4}{3}, \frac{25}{12}\}$ .

## 6.10 Bibliographic notes

The material in Section 6.1 is standard. The books [60, 13] are excellent early references.

Constructing (strong sense) anticausal models in the manner of Section 6.3 was first done in [78] in continuous time and subsequently in [100] in discrete time.

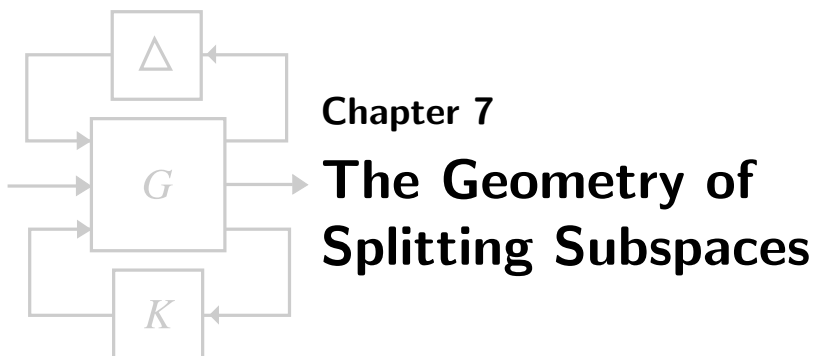
The notion of minimal splitting subspace, a generalization of a concept introduced in [91], was first applied to the stochastic realization problem in [102], where the predictor space  $\mathbf{X}_-$  was considered, and then in general in [75, 77, 76]. Independently, G. Ruckebusch developed a geometric theory of Markovian [109, 107, 108]. This research led to the joint paper [87].

The definitions of observability and constructibility in Section 6.6 were introduced in the context of Markovian representations by Ruckebusch [108].

The forward and backward predictor spaces were introduced by Akaike [1] in the context of canonical correlation analysis. The idea of using Riesz' representation theorem in the proof of Theorem 6.7.1 was suggested by Gy. Michaletzky. This establishes the existence of a minimal spectral factor. The existence of a unique minimum phase spectral factor of a rational spectral density was one of the main results in Youla's classical 1961 paper [128].

There is an extensive literature on rational spectral factorization; In the present context of stochastic processes, see, in particular, the excellent book [28]. The Positive Real Lemma is a version of the Kalman-Yakubovich-Popov Lemma [55, 125, 104]. Theorem 10.4.4 is due to B.D.O. Anderson [4].

The set  $\mathcal{P}$  is studied extensively in [28], where the solvability of  $M(P) \geq 0$  is characterized using algebraic methods (Theorem 3.1), and where an algorithm for determining  $P_+$  is provided. Establishing the partial ordering of  $\mathcal{P}$  via Kalman filtering was done in [78] in continuous time and subsequently in [100] in discrete time.



## Chapter 7

# The Geometry of Splitting Subspaces

The purpose of this chapter is to introduce coordinate-free representations of a stationary process  $y$  by constructing state spaces from basic principles. This will in particular accommodate both finite- and infinite-dimensional stochastic systems.

More precisely, we introduce the geometry underlying linear stochastic models in a more abstract Hilbert space setting which can also be applied to a wider class of problems. The basic setting in this chapter is a fixed real Hilbert space  $\mathbb{H}$  with inner product  $\langle \cdot, \cdot \rangle$ , a unitary shift  $U : \mathbb{H} \rightarrow \mathbb{H}$  acting on it and two subspaces  $\mathbf{H}^-$  and  $\mathbf{H}^+$ , representing the *past* and the *future* respectively, which enjoy the invariance properties

$$U^* \mathbf{H}^- \subset \mathbf{H}^- \quad \text{and} \quad U \mathbf{H}^+ \subset \mathbf{H}^+$$

and the property that the subspace

$$\mathbf{H} := \mathbf{H}^- \vee \mathbf{H}^+$$

is *doubly invariant*, i.e., invariant under both  $U$  and the adjoint shift  $U^*$ . The orthogonal projection of  $\eta \in \mathbb{H}$  onto the subspace  $X$  will be denoted  $E^X \eta$ , and  $E^X \mathbf{Y}$  will denote the *closure* of  $\{E^X \eta \mid \eta \in \mathbf{Y}\}$ .

## 7.1 Deterministic realization theory revisited: The abstract idea of state space construction

This deterministic state space construction of Section 6.1 follows an abstract pattern, which, to a certain extent to be explained below, also applies to the stochastic setting. Given the Hankel map  $\mathcal{H} : \mathcal{U} \rightarrow \mathcal{Y}$ , defined in Section 6.1, one constructs a factorization

$$\begin{array}{ccc} \mathcal{U} & \xrightarrow{\mathcal{H}} & \mathcal{Y} \\ \text{onto} \searrow & & \nearrow 1-1 \\ & X & \end{array}$$

which is *canonical* in the sense that the map  $\mathcal{U} \rightarrow X$  is onto, the map  $X \rightarrow \mathcal{Y}$  is one-to-one, and the dimension of  $X$  is equal to  $\text{rank } \mathcal{H}$ . This amounts to factoring out the kernel of  $\mathcal{H}$ .

In fact, two inputs  $u_1, u_2 \in \mathcal{U}$  are said to be (Nerode) *equivalent* if  $\mathcal{H}u_1 = \mathcal{H}u_2$ , i.e.,  $u_1 - u_2 \in \ker \mathcal{H}$ . Next, define the *canonical projection*

$$\pi_H u = \{v \in \mathcal{U} \mid v \sim u\}$$

which assigns to each  $u \in \mathcal{U}$  the equivalence class to which it belongs, and let

$$\mathcal{U}/\ker H := \{\pi_H u \mid u \in \mathcal{U}\}$$

be the quotient space of all equivalence classes. Setting  $X = \mathcal{U}/\ker H$  yields the factorization

$$\begin{array}{ccc} \mathcal{U} & \xrightarrow{\mathcal{H}} & \mathcal{Y} \\ \pi_H \searrow & & \nearrow \varphi \\ & X & \end{array}$$

where  $\varphi$  assigns the common  $H$ -value to the equivalence class, i.e.,  $\varphi(\pi_H u) = Hu$ . Clearly  $\pi_H$  is onto and  $\varphi$  is one-to-one so that the factorization is canonical.

Next, we observe that the space  $\mathcal{Y}$ , defined on page 119, is invariant under the shift  $\sigma_t y(\tau) = y(\tau + t)$ ,  $t \geq 0$ ;

$$\sigma_t \mathcal{Y} \subset \mathcal{Y}, \quad t = 0, 1, 2, \dots$$

We seek a restricted shift on  $\mathbf{X}$ ; i.e., an operator  $\sigma_t(\mathbf{X}) : \mathbf{X} \rightarrow \mathbf{X}$  that makes the following diagram commute:

$$\begin{array}{ccc} \mathbf{X} & \xrightarrow{\mathcal{O}} & \mathcal{Y} \\ \sigma_t(\mathbf{X}) \downarrow & & \downarrow \sigma_t \\ \mathbf{X} & \xrightarrow{\mathcal{O}} & \mathcal{Y} \end{array}$$

Here  $\mathcal{O}$  is the observability operator, defined in Section 6.1. Comparing with Theorem 6.1.3 and its proof, we see that  $A$  is a matrix representation of the one-step restricted shift  $\sigma(\mathbf{X}) := \sigma_1(\mathbf{X})$  and that the semigroup property

$$\sigma_s(\mathbf{X})\sigma_t(\mathbf{X}) = \sigma_{s+t}(\mathbf{X})$$

holds. In fact,  $\sigma_t(\mathbf{X}) := \sigma(\mathbf{X})^t$  for  $t = 0, 1, 2, \dots$

When modeling a stochastic process  $y$  there are no external inputs, and the construction of the state space will have to be based on somewhat different principles. The main ideas here are the concepts of Markovian splitting subspace and scattering pair representation, which will bring into play certain (white noise) generating processes that serve as inputs for a pair of causal and anticausal representations of  $y$ , as described in Chapter 6. In analyzing these input/output maps, the abstract deterministic realization theory will be used.



## 7.2 Perpendicular intersection

Let  $\mathbf{A}$ ,  $\mathbf{B}$  and  $\mathbf{X}$  be subspaces of the real Hilbert space  $\mathbb{H}$ . We recall from Chapter 2 that  $\mathbf{A}$  and  $\mathbf{B}$  are *conditionally orthogonal* given  $\mathbf{X}$  if

$$\langle \alpha - E^{\mathbf{X}} \alpha, \beta - E^{\mathbf{X}} \beta \rangle = 0 \quad \text{for } \alpha \in \mathbf{A}, \beta \in \mathbf{B} \quad (7.2.1)$$

and that this is denoted  $\mathbf{A} \perp \mathbf{B} \mid \mathbf{X}$ . When  $\mathbf{X} = 0$ , this reduces to the usual orthogonality  $\mathbf{A} \perp \mathbf{B}$ . (See Section 2.4 for alternative characterizations of conditional orthogonality.)

It is trivial that the conditional orthogonality condition  $\mathbf{A} \perp \mathbf{B} \mid \mathbf{X}$  remains true if  $\mathbf{A}$  and  $\mathbf{B}$  are replaced by arbitrary subspaces of  $\mathbf{A}$  and  $\mathbf{B}$  respectively. The converse question of how much  $\mathbf{A}$  and  $\mathbf{B}$  can be enlarged is less trivial and fundamental for what follows.

**Lemma 7.2.1.** *Suppose  $\mathbf{A} \perp \mathbf{B} \mid \mathbf{X}$ . Then*

- (i)  $\mathbf{A} \cap \mathbf{B} \subset \mathbf{X}$
- (ii)  $(\mathbf{A} \vee \mathbf{X}) \perp (\mathbf{B} \vee \mathbf{X}) \mid \mathbf{X}$
- (iii)  $\mathbf{X} = (\mathbf{A} \vee \mathbf{X}) \cap (\mathbf{B} \vee \mathbf{X})$

**Proof.** To prove (i), let  $\lambda \in \mathbf{A} \cap \mathbf{B}$ . Then, since  $\mathbf{A} \perp \mathbf{B} \mid \mathbf{X}$ ,

$$\langle \lambda - E^{\mathbf{X}} \lambda, \lambda - E^{\mathbf{X}} \lambda \rangle = 0,$$

i.e.,  $\|\lambda - E^{\mathbf{X}} \lambda\| = 0$ , and hence  $E^{\mathbf{X}} \lambda = \lambda$ , i.e.  $\lambda \in \mathbf{X}$ . Statement (ii) follows from Proposition 2.4.2 (i)–(iii) in Chapter 2. Finally, to prove (iii), note that property (i) applied to (ii) yields  $(\mathbf{A} \vee \mathbf{X}) \cap (\mathbf{B} \vee \mathbf{X}) \subset \mathbf{X}$ . But  $\mathbf{X} \subset (\mathbf{A} \vee \mathbf{X}) \cap (\mathbf{B} \vee \mathbf{X})$  is trivial.  $\square$

Setting  $\mathbf{S} := \mathbf{A} \vee \mathbf{X}$  and  $\bar{\mathbf{S}} := \mathbf{B} \vee \mathbf{X}$ , Lemma 7.2.1 implies that  $\mathbf{A} \perp \mathbf{B} \mid \mathbf{X}$  is equivalent to

$$\mathbf{S} \perp \bar{\mathbf{S}} \mid \mathbf{S} \cap \bar{\mathbf{S}}.$$

We shall provide some alternative characterizations of this property.

**Proposition 7.2.2.** *The following conditions are equivalent*

- (i)  $\mathbf{S} \perp \bar{\mathbf{S}} \mid \mathbf{S} \cap \bar{\mathbf{S}}$
- (ii)  $E^{\mathbf{S}} \bar{\mathbf{S}} = \mathbf{S} \cap \bar{\mathbf{S}}$
- (iii)  $E^{\bar{\mathbf{S}}} \mathbf{S} = \mathbf{S} \cap \bar{\mathbf{S}}$
- (vi)  $E^{\bar{\mathbf{S}}} \mathbf{S} = E^{\mathbf{S}} \bar{\mathbf{S}}$

**Proof.** By Proposition 2.4.3,

$$\mathbf{S} \perp \bar{\mathbf{S}} \mid E^{\mathbf{S}} \bar{\mathbf{S}}, \quad (7.2.2)$$

and therefore Lemma 7.2.1 (i) implies that

$$\mathbf{S} \cap \bar{\mathbf{S}} \subset E^{\mathbf{S}} \bar{\mathbf{S}}. \quad (7.2.3)$$

Also, by Proposition 2.4.3,  $E^{\mathbf{S}} \bar{\mathbf{S}} \subset \mathbf{X}$  for any  $\mathbf{X} \subset \mathbf{S}$  such that  $\mathbf{S} \perp \bar{\mathbf{S}} \mid \mathbf{X}$ . Hence (i) implies (ii). A symmetric argument shows that (i) implies (iii) as well, and therefore (i) also implies (iv). Consequently, if (iv) holds, then  $E^{\mathbf{S}} \bar{\mathbf{S}} \subset \mathbf{S} \cap \bar{\mathbf{S}}$ , which together with (7.2.3) and (7.2.2) yields (i).  $\square$

**Definition 7.2.3.** A pair  $(\mathbf{S}, \bar{\mathbf{S}})$  of subspaces satisfying the conditions of Proposition 7.2.2 are called *perpendicularly intersecting*.

The property of perpendicular intersection is depicted in Figure 5.1.

**Theorem 7.2.4.** Let  $\mathbf{S}$  and  $\bar{\mathbf{S}}$  be subspaces such that  $\mathbf{S} \vee \bar{\mathbf{S}} = \mathbb{H}$ . Then the following conditions are equivalent.

- (i)  $\mathbf{S}$  and  $\bar{\mathbf{S}}$  intersect perpendicularly
- (ii)  $\bar{\mathbf{S}}^\perp \subset \mathbf{S}$  or, equivalently,  $\mathbf{S}^\perp \subset \bar{\mathbf{S}}$
- (iii)  $\mathbb{H} = \bar{\mathbf{S}}^\perp \oplus (\mathbf{S} \cap \bar{\mathbf{S}}) \oplus \mathbf{S}^\perp$
- (iv)  $E^{\mathbf{S}}$  and  $E^{\bar{\mathbf{S}}}$  commute

**Proof.** Set  $\mathbf{X} = \mathbf{S} \cap \bar{\mathbf{S}}$ . If (i) holds,  $\mathbf{X} = E^{\mathbf{S}} \bar{\mathbf{S}}$ , and hence  $\mathbf{S} \ominus \mathbf{X} \perp \bar{\mathbf{S}}$  (Lemma 2.2.11). But, since  $\mathbf{X} \subset \bar{\mathbf{S}}$  and  $\mathbf{S} \vee \bar{\mathbf{S}} = \mathbb{H}$ , we have  $(\mathbf{S} \ominus \mathbf{X}) \oplus \bar{\mathbf{S}} = \mathbb{H}$ , and therefore  $\mathbf{S} \ominus \mathbf{X} = \bar{\mathbf{S}}^\perp$ , i.e.  $\mathbf{S} = \mathbf{X} \oplus \bar{\mathbf{S}}^\perp$ . Hence both (ii) and (iii) follow. Each of conditions (ii) and (iii) implies the existence of a subspace  $\mathbf{X}$  with the property  $\mathbb{H} = \bar{\mathbf{S}}^\perp \oplus \mathbf{X} \oplus \mathbf{S}^\perp$  so that, if  $\lambda \in \mathbb{H}$ ,

$$E^{\mathbf{S}} E^{\bar{\mathbf{S}}} \lambda = E^{\mathbf{X}} E^{\bar{\mathbf{S}}} \lambda + E^{\bar{\mathbf{S}}^\perp} E^{\bar{\mathbf{S}}} \lambda = E^{\mathbf{X}} \lambda$$

and

$$E^{\bar{\mathbf{S}}} E^{\mathbf{S}} \lambda = E^{\mathbf{X}} \lambda = E^{\mathbf{X}} E^{\mathbf{S}} \lambda + E^{\mathbf{S}^\perp} E^{\mathbf{S}} \lambda = E^{\mathbf{X}} \lambda,$$

and therefore (iv) follows. It just remains to prove that (iv) implies (i). But,  $E^{\mathbf{S}} E^{\bar{\mathbf{S}}} \mathbb{H} = E^{\bar{\mathbf{S}}} E^{\mathbf{S}} \mathbb{H}$  yields  $E^{\mathbf{S}} \bar{\mathbf{S}} = E^{\bar{\mathbf{S}}} \mathbf{S}$ , i.e.,  $\mathbf{S}$  and  $\bar{\mathbf{S}}$  intersect perpendicularly (Proposition 7.2.2).  $\square$

**Corollary 7.2.5.** Let  $\mathbf{S}$  and  $\bar{\mathbf{S}}$  be any subspaces of  $\mathbb{H}$ . Then, if  $\bar{\mathbf{S}}^\perp \subset \mathbf{S}$  or, equivalently,  $\mathbf{S}^\perp \subset \bar{\mathbf{S}}$ ,  $\mathbf{S}$  and  $\bar{\mathbf{S}}$  intersect perpendicularly, and conditions (iii) and (iv) in Theorem 7.2.4 hold.

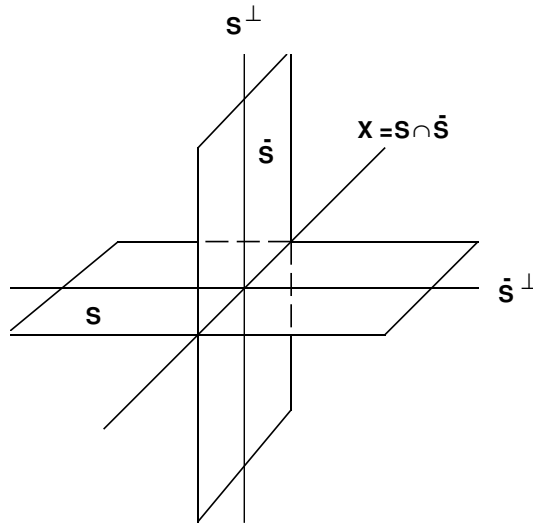


Figure 5.1: The splitting geometry.

**Proof.** By Lemma 2.2.11 and  $\bar{S}^\perp \subset S$ , we have

$$S = E^S \bar{S} \oplus \bar{S}^\perp$$

which implies that  $E^S \bar{S} \subset \bar{S}$ . Hence

$$S \cap \bar{S} = E^S \bar{S},$$

and therefore  $S$  and  $\bar{S}$  intersect perpendicularly (Proposition 7.2.2) and (iii) holds. Consequently, (iv) also holds.  $\square$

We are now in a position to answer the question of how much the subspaces  $A$  and  $B$  may be enlarged while retaining the conditional orthogonality  $A \perp B \mid X$  in the special, but important, special case that  $A \vee B = \mathbb{H}$ .

**Theorem 7.2.6.** *Let  $A$  and  $B$  be subspaces such that  $A \vee B = \mathbb{H}$ , and suppose that  $A \perp B \mid X$ . Let  $S \supset A$  and  $\bar{S} \supset B$ . Then  $S \perp \bar{S} \mid X$  if and only if*

$$S \subset A \vee X \quad \text{and} \quad \bar{S} \subset B \vee X. \tag{7.2.4}$$

*If the upper bounds in (7.2.4) are attained, i.e.,  $S = A \vee X$  and  $\bar{S} \subset B \vee X$ , then  $X = S \cap \bar{S}$  and  $S$  and  $\bar{S}$  intersect perpendicularly.*

**Proof.** It is clear from Lemmas 7.2.1 and 2.4.1 that  $S \perp \bar{S}$  if (7.2.4) holds. Conversely, suppose that  $S \perp \bar{S}$ . Then, by Proposition 2.4.2 (iii) in Chapter 2,  $(S \vee X) \perp \bar{S} \mid X$ , which, in view of Lemma 2.4.1, implies that

$$Z \perp B \mid X, \tag{7.2.5}$$

where  $\mathbf{Z} := (\mathbf{S} \vee \mathbf{X}) \ominus (\mathbf{A} \vee \mathbf{X})$ . But (7.2.5) is equivalent to

$$\mathbf{Z} \perp (\mathbf{B} \vee \mathbf{X}) \ominus \mathbf{X} \quad (7.2.6)$$

by Proposition 2.4.2 (v) in Chapter 2. Since, by definition,  $\mathbf{Z} \perp (\mathbf{A} \vee \mathbf{X})$ , we have  $\mathbf{Z} \perp \mathbf{A} \vee \mathbf{B} = \mathbb{H}$ , and hence  $\mathbf{Z} = 0$ , proving the first of inclusions (7.2.4). A symmetric argument shows that the second inclusion must also hold. The last statement follows from Lemma 7.2.1 (iii) and Proposition 7.2.2.  $\square$

### 7.3 Splitting subspaces

A *splitting subspace* is a subspace  $\mathbf{X} \subset \mathbb{H}$  with the property

$$\mathbf{H}^- \perp \mathbf{H}^+ \mid \mathbf{X}, \quad (7.3.1)$$

i.e., the past and the future spaces are conditionally orthogonal given  $\mathbf{X}$ . From Proposition 2.4.2 in Chapter 2 it follows that  $\mathbf{X}$  is a splitting subspace if and only if

$$E^{\mathbf{H}^- \vee \mathbf{X}} \lambda = E^{\mathbf{X}} \lambda \quad \text{for all } \lambda \in \mathbf{H}^+ \quad (7.3.2)$$

or, equivalently,

$$E^{\mathbf{H}^+ \vee \mathbf{X}} \lambda = E^{\mathbf{X}} \lambda \quad \text{for all } \lambda \in \mathbf{H}^-. \quad (7.3.3)$$

Consequently,  $\mathbf{X}$  serves as a “memory” or “sufficient statistics” which contains everything from the past which is needed in predicting the future and everything from the which is needed in predicting the past. Clearly  $\mathbb{H}$ ,  $\mathbf{H}$ ,  $\mathbf{H}^-$  and  $\mathbf{H}^+$  are splitting subspaces. Therefore, to obtain real data reduction, we shall be interested in splitting subspaces  $\mathbf{X}$  which are *minimal* in the sense that if  $\mathbf{X}_1$  is also a splitting subspace and  $\mathbf{X}_1 \subset \mathbf{X}$ , then  $\mathbf{X}_1 = \mathbf{X}$ .

The following result, which is a corollary of Theorem 2.4.3, provides us with two examples of minimal splitting subspaces.

**Proposition 7.3.1.** *The predictor spaces*

$$\mathbf{X}_- := E^{\mathbf{H}^-} \mathbf{H}^+ \quad \text{and} \quad \mathbf{X}_+ := E^{\mathbf{H}^+} \mathbf{H}^-$$

are minimal splitting subspaces. In fact,  $\mathbf{X}_-$  is the only minimal splitting subspace contained in  $\mathbf{H}^-$  and  $\mathbf{X}_+$  is the only minimal splitting subspace contained in  $\mathbf{H}^+$ .

To shed some light on the splitting property, observe that, in view of Proposition 2.4.2 (vi) in Chapter 2,

$$E^{\mathbf{H}^+} \lambda = E^{\mathbf{H}^+} E^{\mathbf{X}} \lambda \quad \text{for } \lambda \in \mathbf{H}^- \quad (7.3.4)$$

is an equivalent characterization of a splitting subspace  $\mathbf{X}$ . To understand better this characterization, we introduce the *observability operator*

$$\mathcal{O} := E^{\mathbf{H}^+} \mid_{\mathbf{X}} \quad (7.3.5)$$

and the *constructibility operator*

$$\mathcal{C} := E^{\mathbf{H}^-} |_{\mathbf{X}}. \tag{7.3.6}$$

Then, since

$$\mathcal{O}^* := E^{\mathbf{X}} |_{\mathbf{H}^+} \quad \text{and} \quad \mathcal{C}^* := E^{\mathbf{X}} |_{\mathbf{H}^-} \tag{7.3.7}$$

are the adjoints of  $\mathcal{O}$  and  $\mathcal{C}$  respectively (Lemma 2.2.7 in Chapter 2), (7.3.4) may be written

$$\mathcal{H} = \mathcal{O}\mathcal{C}^*, \tag{7.3.8}$$

where  $\mathcal{H}$  is the Hankel operator

$$\mathcal{H} := E^{\mathbf{H}^+} |_{\mathbf{H}^-}. \tag{7.3.9}$$

Equivalently, we have

$$\mathcal{H}^* = \mathcal{C}\mathcal{O}^*, \tag{7.3.10}$$

where

$$\mathcal{H}^* = E^{\mathbf{H}^-} |_{\mathbf{H}^+}. \tag{7.3.11}$$

Consequently, the splitting property can be characterized as a factorization of a Hankel operator over the splitting subspace  $\mathbf{X}$  so that the following diagram commutes:

$$\begin{array}{ccc} \mathbf{H}^- & \xrightarrow{\mathcal{H}} & \mathbf{H}^+ \\ \mathcal{C}^* \searrow & & \nearrow \mathcal{O} \\ & \mathbf{X} & \end{array} \quad \mathcal{H} = \mathcal{O}\mathcal{C}^*$$

Such a factorization is said to be *canonical* if  $\mathcal{C}^*$  maps onto a dense subset of  $\mathbf{X}$  and  $\mathcal{O}$  is injective, i.e.,  $\ker \mathcal{O} = 0$ . Equivalently, the same splitting property can be illustrated by factoring the adjoint Hankel operator (7.3.11) over  $\mathbf{X}$  so that the dual diagram

$$\begin{array}{ccc} \mathbf{H}^+ & \xrightarrow{\mathcal{H}^*} & \mathbf{H}^- \\ \mathcal{O}^* \searrow & & \nearrow \mathcal{C} \\ & \mathbf{X} & \end{array} \quad \mathcal{H}^* = \mathcal{C}\mathcal{O}^*,$$

corresponding to the factorization (7.3.10), commutes. Again the factorization is canonical if the range  $\text{Im } \mathcal{O}^*$  is dense in  $\mathbf{X}$  and  $\mathcal{C}$  is injective, i.e.,  $\ker \mathcal{C} = 0$ .

The equivalence between these formulations of canonicity is a simple consequence of the fact that  $\overline{\text{Im } \mathcal{O}^*} = \mathbf{X}$  if and only if  $\ker \mathcal{O} = 0$  and  $\overline{\text{Im } \mathcal{C}^*} = \mathbf{X}$  if and only if  $\ker \mathcal{C} = 0$ . This property, which holds for all bounded linear operators (Theorem A.1.3 in the appendix), can in the present setting be illustrated by applying Lemma 2.2.11 to obtain the orthogonal decompositions

$$\mathbf{X} = E^{\mathbf{X}} \mathbf{H}^+ \oplus \mathbf{X} \cap (\mathbf{H}^+)^{\perp} \tag{7.3.12a}$$

$$\mathbf{X} = E^{\mathbf{X}} \mathbf{H}^- \oplus \mathbf{X} \cap (\mathbf{H}^-)^{\perp} \tag{7.3.12b}$$

Noting that

$$\ker \mathcal{O} = \mathbf{X} \cap (\mathbf{H}^+)^{\perp} \quad \text{and} \quad \ker \mathcal{C} = \mathbf{X} \cap (\mathbf{H}^-)^{\perp}. \tag{7.3.13}$$

and

$$\overline{\text{Im } \mathcal{O}^*} = E^{\mathbf{X}} \mathbf{H}^+ \quad \text{and} \quad \overline{\text{Im } \mathcal{C}^*} = E^{\mathbf{X}} \mathbf{H}^-, \quad (7.3.14)$$

it is seen that (7.3.12) is the well-known decomposition of Theorem A.1.3 in the appendix.

We shall call  $E^{\mathbf{X}} \mathbf{H}^+$  the *observable* and  $\mathbf{X} \cap (\mathbf{H}^+)^{\perp}$  the *unobservable subspace* of  $\mathbf{X}$ . This is in harmony with Kalman's nomenclature since any  $\xi \in \mathbf{X} \cap (\mathbf{H}^+)^{\perp}$  is unobservable in the sense that it cannot be distinguished from zero by observing elements in the space  $\mathbf{H}^+$  of future outputs. Similarly  $E^{\mathbf{X}} \mathbf{H}^-$  is called the *constructible* and  $\mathbf{X} \cap (\mathbf{H}^-)^{\perp}$  the *unconstructible subspace* of  $\mathbf{X}$ .

We restate Definition 6.6.1 in terms of splitting subspaces.

**Definition 7.3.2.** *The splitting subspace is said to be observable if  $\mathbf{X} \cap (\mathbf{H}^+)^{\perp} = 0$  and constructible if  $\mathbf{X} \cap (\mathbf{H}^-)^{\perp} = 0$ .*

Consequently, the factorizations (7.3.8) and (7.3.10) are canonical if and only if  $\mathbf{X}$  is both observable and constructible. Next we show that this canonicity is equivalent to minimality of  $\mathbf{X}$ . To this end, we need the following lemma.

**Lemma 7.3.3.** *Let  $\mathbf{X}$  be a splitting subspace and suppose it has the orthogonal decomposition*

$$\mathbf{X} = \mathbf{X}_1 \oplus \mathbf{X}_2, \quad (7.3.15)$$

where  $\mathbf{X}_1$  and  $\mathbf{X}_2$  are subspaces of  $\mathbf{X}$ . Then  $\mathbf{X}_1$  is a splitting subspace if and only if

$$E^{\mathbf{X}_2} \mathbf{H}^- \perp E^{\mathbf{X}_2} \mathbf{H}^+. \quad (7.3.16)$$

**Proof.** Using the alternative definition (2.4.2) of conditional orthogonality, the splitting property  $\mathbf{H}^- \perp \mathbf{H}^+ | \mathbf{X}$  may be written

$$\langle E^{\mathbf{X}} \lambda, E^{\mathbf{X}} \mu \rangle = \langle \lambda, \mu \rangle \quad \text{for all } \lambda \in \mathbf{H}^- \text{ and } \mu \in \mathbf{H}^+.$$

Therefore, since

$$\langle E^{\mathbf{X}} \lambda, E^{\mathbf{X}} \mu \rangle = \langle E^{\mathbf{X}_1} \lambda, E^{\mathbf{X}_1} \mu \rangle + \langle E^{\mathbf{X}_2} \lambda, E^{\mathbf{X}_2} \mu \rangle,$$

the proof of the lemma is immediate.  $\square$

**Lemma 7.3.4.** *If  $\mathbf{X}$  is a splitting subspace, then so are  $E^{\mathbf{X}} \mathbf{H}^+$  and  $E^{\mathbf{X}} \mathbf{H}^-$ .*

**Proof.** This follows readily from Lemma 7.3.3 and the orthogonal decompositions (7.3.12). With  $\mathbf{X}_2 := \mathbf{X} \cap (\mathbf{H}^+)^{\perp}$ , we have  $E^{\mathbf{X}_2} \mathbf{H}^+ = 0$ , and consequently  $\mathbf{X}_1 := E^{\mathbf{X}} \mathbf{H}^+$  is a splitting subspace. Similarly, setting  $\mathbf{X}_2 := \mathbf{X} \cap (\mathbf{H}^-)^{\perp}$ , it is seen that  $\mathbf{X}_1 := E^{\mathbf{X}} \mathbf{H}^-$  is a splitting subspace.  $\square$

**Theorem 7.3.5.** *A splitting subspace is minimal if and only if it is both observable and constructible.*

**Proof.** Suppose that  $\mathbf{X}$  is a minimal splitting subspace. Then, since  $E^{\mathbf{X}} \mathbf{H}^+$  and  $E^{\mathbf{X}} \mathbf{H}^-$  are also splitting subspaces (Lemma 7.3.4), it follows from decompositions (7.3.12) that  $\mathbf{X} \cap (\mathbf{H}^+)^{\perp} = 0$  and  $\mathbf{X} \cap (\mathbf{H}^-)^{\perp} = 0$ , i.e.,  $\mathbf{X}$  is both observable and constructible.

Conversely, suppose that  $\mathbf{X} \cap (\mathbf{H}^-)^{\perp} = 0$  and  $\mathbf{X} \cap (\mathbf{H}^+)^{\perp} = 0$  and that  $\mathbf{X}_1 \subset \mathbf{X}$  is a splitting subspace. We want to show that  $\mathbf{X}_2 := \mathbf{X} \ominus \mathbf{X}_1$  is the zero space so that  $\mathbf{X}_1 = \mathbf{X}$ . It follows from (7.3.12a) that  $\mathbf{X} = E^{\mathbf{X}} \mathbf{H}^+$ . Applying  $E^{\mathbf{X}_2}$  to this and observing that  $E^{\mathbf{X}_2} E^{\mathbf{X}} = E^{\mathbf{X}_2}$ , we obtain that

$$E^{\mathbf{X}_2} \mathbf{H}^+ = \mathbf{X}_2. \quad (7.3.17)$$

A symmetric argument using (7.3.12b) yields

$$E^{\mathbf{X}_2} \mathbf{H}^- = \mathbf{X}_2. \quad (7.3.18)$$

Since  $\mathbf{X}_1$  is a splitting subspace, Lemma 7.3.3 implies that (7.3.17) and (7.3.18) are orthogonal, which can only happen if  $\mathbf{X}_2 = 0$  as claimed.  $\square$

The splitting property can also be characterized in terms of perpendicularly intersecting subspaces.

**Theorem 7.3.6.** *A subspace  $\mathbf{X} \subset \mathbb{H}$  is a splitting subspace if and only if*

$$\mathbf{X} = \mathbf{S} \cap \bar{\mathbf{S}} \quad (7.3.19)$$

for some pair  $(\mathbf{S}, \bar{\mathbf{S}})$  of perpendicularly intersecting subspaces such that  $\mathbf{S} \supset \mathbf{H}^-$  and  $\bar{\mathbf{S}} \supset \mathbf{H}^+$ . Then

$$\mathbf{X} = E^{\mathbf{S}} \bar{\mathbf{S}} = E^{\bar{\mathbf{S}}} \mathbf{S}. \quad (7.3.20)$$

In particular,

$$E^{\mathbf{S}} \lambda = E^{\mathbf{X}} \lambda, \quad \text{for all } \lambda \in \bar{\mathbf{S}} \quad (7.3.21)$$

and

$$E^{\bar{\mathbf{S}}} \lambda = E^{\mathbf{X}} \lambda, \quad \text{for all } \lambda \in \mathbf{S}. \quad (7.3.22)$$

**Proof.** (if): Suppose  $\mathbf{S}$  and  $\bar{\mathbf{S}}$  intersect perpendicularly. Then, by Proposition 7.2.2,  $\mathbf{S} \perp \bar{\mathbf{S}} \mid \mathbf{X}$  with  $\mathbf{X} = \mathbf{S} \cap \bar{\mathbf{S}}$ . But, since  $\mathbf{S} \supset \mathbf{H}^-$  and  $\bar{\mathbf{S}} \supset \mathbf{H}^+$ , this implies that  $\mathbf{H}^- \perp \mathbf{H}^+ \mid \mathbf{X}$ , i.e.  $\mathbf{X}$  is a splitting subspace. Then (7.3.20) follows directly from Proposition 7.2.2, and, by Proposition 2.4.2, (7.3.21) and (7.3.22) are equivalent to  $\mathbf{S} \perp \bar{\mathbf{S}} \mid \mathbf{X}$ .

(only if): Suppose that  $\mathbf{H}^- \perp \mathbf{H}^+ \mid \mathbf{X}$  and set  $\mathbf{S} := \mathbf{H}^- \vee \mathbf{X}$  and  $\bar{\mathbf{S}} := \mathbf{H}^+ \vee \mathbf{X}$ . Then, by Theorem 7.2.6,  $\mathbf{S} \perp \bar{\mathbf{S}} \mid \mathbf{X}$  where  $\mathbf{X} = \mathbf{S} \cap \bar{\mathbf{S}}$ . Hence  $\mathbf{S}$  and  $\bar{\mathbf{S}}$  intersect perpendicularly (Proposition 7.2.2) and  $\mathbf{S} \supset \mathbf{H}^-$  and  $\bar{\mathbf{S}} \supset \mathbf{H}^+$ .  $\square$

We shall call a pair  $(\mathbf{S}, \bar{\mathbf{S}})$  satisfying the conditions of Theorem 7.3.6 a *scattering pair* of  $\mathbf{X}$ , due to certain similarities to incoming and outgoing subspaces in Lax-Philips scattering theory [69]. The correspondence to the scattering framework

of Lax and Philips will become complete once we introduce invariance with respect to the unitary operator  $U$ , as we shall do in the next section.

In general, an  $\mathbf{X}$  may have more than one scattering pair. In the next section further conditions will be imposed which will allow us to assign a unique scattering pair to each  $\mathbf{X}$ . However, if we take  $\mathbb{H} = \mathbf{H} := \mathbf{H}^- \vee \mathbf{H}^+$ , each splitting subspace has a unique scattering pair  $(\mathbf{S}, \bar{\mathbf{S}})$ . This choice of space  $\mathbb{H}$  amounts to only considering *internal* splitting subspaces, i.e., splitting subspaces such that  $\mathbf{X} \subset \mathbf{H}$ .

**Proposition 7.3.7.** *Suppose that  $\mathbb{H} = \mathbf{H}$ . Then each splitting subspace  $\mathbf{X}$  has a unique scattering pair  $(\mathbf{S}, \bar{\mathbf{S}})$ , namely*

$$\mathbf{S} = \mathbf{H}^- \vee \mathbf{X}, \quad \bar{\mathbf{S}} = \mathbf{H}^+ \vee \mathbf{X}. \quad (7.3.23)$$

**Proof.** It follows from Theorem 7.3.6 that  $\mathbf{S} \supset \mathbf{H}^- \vee \mathbf{X}$  and  $\bar{\mathbf{S}} \supset \mathbf{H}^+ \vee \mathbf{X}$ . But, since  $\mathbf{H}^- \vee \mathbf{H}^+ = \mathbb{H}$ , Theorem 7.2.6 implies that  $\mathbf{S} \subset \mathbf{H}^- \vee \mathbf{X}$  and  $\bar{\mathbf{S}} \subset \mathbf{H}^+ \vee \mathbf{X}$ , and consequently (7.3.23) holds.  $\square$

Proposition 7.3.7 makes the study of internal splitting subspaces much simpler than in the general case.

## 7.4 Markovian splitting subspaces

The splitting property insures that  $\mathbf{X}$  contains the information needed to act as a state space but says nothing about how this dynamical memory evolves in time. Therefore, we shall need to assume that  $\mathbf{X}$  has the additional property that it splits the combined past and the combined future of  $\{y(t)\}$  and  $\mathbf{X}$ , i.e. that

$$(\mathbf{H}^- \vee \mathbf{X}^-) \perp (\mathbf{H}^+ \vee \mathbf{X}^+) | \mathbf{X}, \quad (7.4.1)$$

where  $\mathbf{X}^- := \overline{\text{span}}\{U^t \mathbf{X} \mid t \leq 0\}$  and  $\mathbf{X}^+ := \overline{\text{span}}\{U^t \mathbf{X} \mid t \geq 0\}$ . Clearly the splitting property  $\mathbf{H}^- \perp \mathbf{H}^+ | \mathbf{X}$  is also implied by (7.4.1) as is the Markov property

$$\mathbf{X}^- \perp \mathbf{X}^+ | \mathbf{X}. \quad (7.4.2)$$

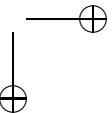
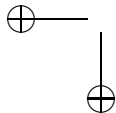
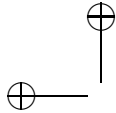
Moreover, define the *ambient space*  $\mathbb{H}_{\mathbf{X}}$  of  $\mathbf{X}$  as the smallest subspace of  $\mathbb{H}$  which contains both  $\mathbf{H}$  and  $\mathbf{X}$  and which is invariant under both forward shift  $U$  and backward shift  $U^*$ . More precisely,

$$\mathbb{H}_{\mathbf{X}} = \mathbf{H} \vee \overline{\text{span}}\{U^t \mathbf{X} \mid t \in \mathbb{Z}\} \quad (7.4.3)$$

We say that  $\mathbf{X}$  is a *Markovian splitting subspace* if it satisfies (7.4.1), and the triplet  $(\mathbb{H}_{\mathbf{X}}, U, \mathbf{X})$  is called a *Markovian representation*. If  $\mathbb{H}_{\mathbf{X}} = \mathbf{H}$ , i.e.  $\mathbf{X} \subset \mathbf{H}$ , we say that the Markovian representation is *internal*.

The subspaces  $\mathbf{S}$  and  $\bar{\mathbf{S}}$  of Theorem 7.3.6 may be regarded as extension of the past space  $\mathbf{H}^-$  and the future space  $\mathbf{H}^+$ . Since  $\mathbf{H}^-$  and  $\mathbf{H}^+$  satisfy the invariance properties

$$U^* \mathbf{H}^- \subset \mathbf{H}^- \quad \text{and} \quad U \mathbf{H}^+ \subset \mathbf{H}^+,$$





the following theorem shows that  $\mathbf{S}$  and  $\bar{\mathbf{S}}$  indeed play the roles of extended past and future spaces.

**Theorem 7.4.1.** *A splitting subspace  $\mathbf{X}$  is a Markovian splitting subspace if and only if it has a scattering pair  $(\mathbf{S}, \bar{\mathbf{S}})$  such that*

$$U^*\mathbf{S} \subset \mathbf{S} \quad \text{and} \quad U\bar{\mathbf{S}} \subset \bar{\mathbf{S}} \quad (7.4.4)$$

For each  $\mathbf{X}$  there is a unique such scattering pair contained in the ambient space  $\mathbb{H}_{\mathbf{X}}$ , and it is given by

$$\mathbf{S} = \mathbf{H}^- \vee \mathbf{X}^- \quad \text{and} \quad \bar{\mathbf{S}} = \mathbf{H}^+ \vee \mathbf{X}^+. \quad (7.4.5)$$

Moreover,  $\mathbf{S} \vee \bar{\mathbf{S}} = \mathbb{H}_{\mathbf{X}}$ .

**Proof.** To prove the (only if) part, suppose that (7.4.1) holds, i.e. that  $\mathbf{S} \perp \bar{\mathbf{S}} \mid \mathbf{X}$  where  $\mathbf{S} := \mathbf{H}^- \vee \mathbf{X}^-$  and  $\bar{\mathbf{S}} := \mathbf{H}^+ \vee \mathbf{X}^+$ . Then, by Lemma 7.2.1,  $\mathbf{X} = \mathbf{S} \cap \bar{\mathbf{S}}$  so that  $\mathbf{S} \perp \bar{\mathbf{S}} \mid \mathbf{S} \cap \bar{\mathbf{S}}$ , and hence  $\mathbf{S}$  and  $\bar{\mathbf{S}}$  intersect perpendicularly. Hence, since  $\mathbf{S} \supset \mathbf{H}^-$  and  $\bar{\mathbf{S}} \supset \mathbf{H}^+$ ,  $(\mathbf{S}, \bar{\mathbf{S}})$  is a scattering pair of  $\mathbf{X}$ .

To prove the (if) part, suppose  $\mathbf{X}$  is a splitting subspace with a scattering pair  $(\mathbf{S}, \bar{\mathbf{S}})$  satisfying the invariance property (7.4.4). Since  $\mathbf{X} \subset \mathbf{S}$ , (7.4.4) implies that  $U^{-1}\mathbf{X} \subset \mathbf{S}$ , and hence

$$U^t\mathbf{X} \subset \mathbf{S} \quad \text{for } t \leq 0.$$

Therefore  $\mathbf{X}^- \subset \mathbf{S}$ . But  $\mathbf{H}^- \subset \mathbf{S}$ , and consequently

$$\mathbf{H}^- \vee \mathbf{X}^- \subset \mathbf{S}. \quad (7.4.6)$$

A symmetric argument yields

$$\mathbf{H}^+ \vee \mathbf{X}^+ \subset \bar{\mathbf{S}}. \quad (7.4.7)$$

Hence (7.4.1) follows from  $\mathbf{S} \perp \bar{\mathbf{S}} \mid \mathbf{X}$ .

Finally, we prove uniqueness. If  $(\mathbf{S}, \bar{\mathbf{S}})$  is a scattering pair for  $\mathbf{X}$ , we have  $\mathbf{S} \perp \bar{\mathbf{S}} \mid \mathbf{X}$  where  $\mathbf{X} = \mathbf{S} \cap \bar{\mathbf{S}}$ . Therefore, it follows from (7.4.1) and Theorem 7.2.6 that

$$\mathbf{S} \subset \mathbf{H}^- \vee \mathbf{X}^- \quad \text{and} \quad \bar{\mathbf{S}} \subset \mathbf{H}^+ \vee \mathbf{X}^+. \quad (7.4.8)$$

In fact, set  $\mathbf{A} := \mathbf{H}^- \vee \mathbf{X}^-$  and  $\mathbf{B} := \mathbf{H}^+ \vee \mathbf{X}^+$  and observe that  $\mathbf{A} \vee \mathbf{B} = \mathbb{H}_{\mathbf{X}}$  and that  $\mathbf{A} \vee \mathbf{X} = \mathbf{H}^- \vee \mathbf{X}^-$  and  $\mathbf{B} \vee \mathbf{X} = \mathbf{H}^+ \vee \mathbf{X}^+$ . Then (7.4.6) (7.4.7) and (7.4.8) yield the required uniqueness, and  $(\mathbf{S}, \bar{\mathbf{S}})$  is given by (7.4.5).  $\square$

For any Markovian splitting subspace we write  $\mathbf{X} \sim (\mathbf{S}, \bar{\mathbf{S}})$  to refer to the one-one correspondence between  $\mathbf{X} = \mathbf{S} \cap \bar{\mathbf{S}}$  and the unique scattering pair  $(\mathbf{S}, \bar{\mathbf{S}})$  contained in  $\mathbb{H}_{\mathbf{X}}$ . In view of Proposition 7.3.7, we have  $\mathbf{H}^- \vee \mathbf{X}^- = \mathbf{H}^- \vee \mathbf{X}$  and  $\mathbf{H}^+ \vee \mathbf{X}^+ = \mathbf{H}^+ \vee \mathbf{X}$  for internal splitting subspaces, but in general these equations do not hold.

In view of Theorem 7.2.4, we may characterize the splitting property of  $\mathbf{X} \sim (\mathbf{S}, \bar{\mathbf{S}})$  by means of the orthogonal decomposition

$$\mathbb{H}_{\mathbf{X}} = \mathbf{S}^{\perp} \oplus \mathbf{X} \oplus \bar{\mathbf{S}}^{\perp}, \tag{7.4.9}$$

where  $\mathbf{S}^{\perp}$  and  $\bar{\mathbf{S}}^{\perp}$  are the orthogonal components of  $\mathbf{S}$  and  $\bar{\mathbf{S}}$  in  $\mathbb{H}_{\mathbf{X}}$ . (This will be the notational convention for the rest of this section.) The decomposition (7.4.9) is illustrated in Figure 5.1, which also illustrates that  $\mathbf{S}$  and  $\bar{\mathbf{S}}$  intersect perpendicularly if and only if  $\bar{\mathbf{S}}^{\perp} \subset \mathbf{S}$  or, equivalently,  $\mathbf{S}^{\perp} \subset \bar{\mathbf{S}}$  (Theorem 7.2.4). Since, in addition,  $\mathbf{S} \supset \mathbf{H}^{-}$  and  $\bar{\mathbf{S}} \supset \mathbf{H}^{+}$  (Theorem 7.3.6), the splitting geometry requires that

$$\bar{\mathbf{S}} \supset \mathbf{H}^{+} \vee \mathbf{S}^{\perp} \quad \text{and} \quad \mathbf{S} \supset \mathbf{H}^{-} \vee \bar{\mathbf{S}}^{\perp}. \tag{7.4.10}$$

A Markovian splitting subspace is said to be *minimal* if it contains no proper subspace which is also a Markovian splitting subspace. We turn now to the question of how minimality can be characterized in terms of the scattering pair  $(\mathbf{S}, \bar{\mathbf{S}})$ . Since  $\mathbf{X} = \mathbf{S} \cap \bar{\mathbf{S}}$ , minimality of  $\mathbf{X}$  should be expected to be connected to some minimality conditions on  $\mathbf{S}$  and  $\bar{\mathbf{S}}$ .

**Lemma 7.4.2.** *Let  $\mathbf{X}_1 \sim (\mathbf{S}_1, \bar{\mathbf{S}}_1)$  and  $\mathbf{X}_2 \sim (\mathbf{S}_2, \bar{\mathbf{S}}_2)$  be Markovian splitting subspaces. Then  $\mathbf{X}_1 \subset \mathbf{X}_2$  if and only if  $\mathbf{S}_1 \subset \mathbf{S}_2$  and  $\bar{\mathbf{S}}_1 \subset \bar{\mathbf{S}}_2$ .*

**Proof.** The (if) part follows from the fact that  $\mathbf{X} = \mathbf{S} \cap \bar{\mathbf{S}}$  for all  $\mathbf{X} \sim (\mathbf{S}, \bar{\mathbf{S}})$ . The (only if) part follows from (7.4.5)  $\square$

Given an arbitrary Markovian splitting subspace  $\mathbf{X} \sim (\mathbf{S}, \bar{\mathbf{S}})$ , how do we find a minimal one contained in it? If this is at all possible, Lemma 7.4.2 suggests that we would need to reduce  $\mathbf{S}$  and  $\bar{\mathbf{S}}$  as far as possible while preserving the splitting geometry; i.e., satisfying the constraints (7.4.10) and the invariance conditions (7.4.4).

**Theorem 7.4.3.** *Let  $\mathbf{X} \sim (\mathbf{S}, \bar{\mathbf{S}})$  be a Markovian splitting subspace with ambient space  $\mathbb{H}_{\mathbf{X}}$ , and set  $\bar{\mathbf{S}}_1 := \mathbf{H}^{+} \vee \mathbf{S}^{\perp}$  and  $\mathbf{S}_1 = \mathbf{H}^{-} \vee \bar{\mathbf{S}}_1^{\perp}$ , where  $^{\perp}$  denotes orthogonal complement in  $\mathbb{H}_{\mathbf{X}}$ . Then  $\mathbf{X}_1 \sim (\mathbf{S}_1, \bar{\mathbf{S}}_1)$  is a minimal Markovian splitting subspace such that  $\mathbf{X}_1 \subset \mathbf{X}$ .*

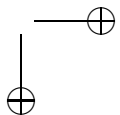
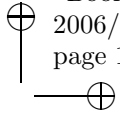
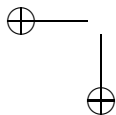
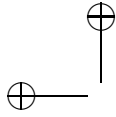
**Proof.** The subspaces  $\mathbf{S}_1$  and  $\bar{\mathbf{S}}_1$  intersect perpendicularly by virtue of the fact that  $\bar{\mathbf{S}}_1^{\perp} \subset \mathbf{S}_1$  (Corollary 7.2.5). Therefore, since  $\mathbf{S}_1 \supset \mathbf{H}^{-}$  and  $\bar{\mathbf{S}}_1 \supset \mathbf{H}^{+}$ ,  $\mathbf{X}_1 = \mathbf{S}_1 \cap \bar{\mathbf{S}}_1$  is a splitting subspace (Theorem 7.3.6). We need to show that it is Markovian. Since  $U^*\mathbf{S} \subset \mathbf{S}$ , we have  $U\mathbf{S}^{\perp} \subset \mathbf{S}^{\perp}$  (Lemma A.1.6), which together with the invariance property  $U\mathbf{H}^{+} \subset \mathbf{H}^{+}$  yields

$$U\bar{\mathbf{S}}_1 \subset \bar{\mathbf{S}}_1.$$

Consequently,  $\mathbf{X}_1 \sim (\mathbf{S}_1, \bar{\mathbf{S}}_1)$  is a Markovian splitting subspace (Theorem 7.4.1).

Next we show that  $\mathbf{X}_1 \subset \mathbf{X}$ . To this end, first note that  $\mathbf{S}^{\perp} \subset \bar{\mathbf{S}}_1$ , or, equivalently,  $\bar{\mathbf{S}}_1^{\perp} \subset \mathbf{S}$ , which together with  $\mathbf{H}^{-} \subset \mathbf{S}$  yields

$$\mathbf{S}_1 \subset \mathbf{S}.$$



Also, since  $\bar{\mathbf{S}}_1 = \mathbf{H}^+ \vee \mathbf{S}^\perp$ , the first of equations (7.4.10) yields

$$\bar{\mathbf{S}}_1 \subset \bar{\bar{\mathbf{S}}}_1.$$

Consequently, by Lemma 7.4.2,  $\mathbf{X}_1 \subset \mathbf{X}$ .

Finally, to prove that  $\mathbf{X}_1$  is minimal, we assume that there is a Markovian splitting subspace  $\mathbf{X}_2 \sim (\mathbf{S}_2, \bar{\mathbf{S}}_2)$  such that  $\mathbf{X}_2 \subset \mathbf{X}_1$ . Then, by Lemma 7.4.2 and the fact that  $\mathbf{S}_1 \subset \mathbf{S}$ , we have  $\mathbf{S}_2 \subset \mathbf{S}$  and  $\bar{\mathbf{S}}_2 \subset \bar{\mathbf{S}}$  so that  $\mathbf{S}^\perp \subset \mathbf{S}_2^\perp$  and  $\bar{\mathbf{S}}^\perp \subset \bar{\bar{\mathbf{S}}}_2^\perp$ . Therefore, in view of the splitting conditions (7.4.10) for  $\mathbf{X}_2$ ,

$$\bar{\mathbf{S}}_2 \supset \mathbf{H}^+ \vee \mathbf{S}_2^\perp \supset \mathbf{H}^+ \vee \mathbf{S}^\perp = \bar{\mathbf{S}}_1$$

and

$$\mathbf{S}_2 \supset \mathbf{H}^- \vee \bar{\mathbf{S}}_2^\perp \supset \mathbf{H}^- \vee \bar{\mathbf{S}}_1^\perp = \mathbf{S}_1,$$

and consequently, by Lemma 7.4.2,  $\mathbf{X}_2 \supset \mathbf{X}_1$ . Therefore, we must have  $\mathbf{X}_2 = \mathbf{X}_1$ , proving minimality of  $\mathbf{X}_1 \sim (\mathbf{S}_1, \bar{\mathbf{S}}_1)$ .  $\square$

Later we shall need the following corollary, the proof of which follows along similar lines as for Theorem 7.4.3.

**Corollary 7.4.4.** *Let  $\mathbf{X} \sim (\mathbf{S}, \bar{\mathbf{S}})$  be a Markovian splitting subspace, and let  $\mathbf{S}_1$  and  $\bar{\mathbf{S}}_1$  be defined as in Theorem 7.4.3. Then  $\mathbf{X}'_1 \sim (\mathbf{S}, \bar{\mathbf{S}}_1)$  and  $\mathbf{X}''_1 \sim (\mathbf{S}_1, \bar{\mathbf{S}})$  are Markovian splitting subspace contained in  $\mathbf{X}$ .*

Let us illustrate Theorem 7.4.3 by giving some examples. It is immediately seen from Theorems 7.3.6 and 7.4.1 that  $\mathbf{H}^- \sim (\mathbf{H}^-, \mathbf{H})$  is a Markovian splitting subspace with ambient space  $\mathbf{H}$ . Applying Theorem 7.4.3 we obtain  $\bar{\mathbf{S}}_1 = \mathbf{H}^+ \vee (\mathbf{H}^-)^\perp$  and consequently  $\bar{\mathbf{S}}_1^\perp = \mathbf{N}^-$ , where

$$\mathbf{N}^- = \mathbf{H}^- \cap (\mathbf{H}^+)^\perp, \tag{7.4.11}$$

so that  $\mathbf{S}_1 = \mathbf{H}^-$ . Therefore, in view of (7.3.20), the minimal Markovian splitting subspace  $\mathbf{X}_1 \sim (\mathbf{S}_1, \bar{\mathbf{S}}_1)$  contained in  $\mathbf{H}^-$  is given by

$$\mathbf{X}_1 = E^{\mathbf{H}^-} [\mathbf{H}^+ \vee (\mathbf{H}^-)^\perp] = E^{\mathbf{H}^-} \mathbf{H}^+.$$

**Proposition 7.4.5.** *The predictor space*

$$\mathbf{X}_- := E^{\mathbf{H}^-} \mathbf{H}^+ \tag{7.4.12}$$

*is a minimal Markovian splitting subspace and  $\mathbf{X}_- \sim (\mathbf{H}^-, (\mathbf{N}^-)^\perp)$ , where  $\mathbf{N}^-$  is given by (7.4.11).*

The subspace  $\mathbf{N}^-$ , which contains everything in the past which is orthogonal to the future, and hence, loosely speaking, gives no information about the future, will play an important role in what follows.

Likewise, applying Theorem 7.4.3 to  $\mathbf{H}^+ \sim (\mathbf{H}, \mathbf{H}^+)$  we obtain  $\bar{\mathbf{S}}_1 = \mathbf{H}^+$  and  $\mathbf{S}_1 = \mathbf{H}^- \vee (\mathbf{H}^+)^{\perp} = (\mathbf{N}^+)^{\perp}$ , where

$$\mathbf{N}^+ = \mathbf{H}^+ \cap (\mathbf{H}^-)^{\perp}, \quad (7.4.13)$$

and consequently, in view of (7.3.20), the minimal Markovian splitting subspace  $\mathbf{X}_1 \sim (\mathbf{S}_1, \bar{\mathbf{S}}_1)$  now is given by

$$\mathbf{X}_1 = \mathbf{E}^{\mathbf{H}^+} [\mathbf{H}^- \vee (\mathbf{H}^+)^{\perp}] = \mathbf{E}^{\mathbf{H}^+} \mathbf{H}^-.$$

**Proposition 7.4.6.** *The backward predictor space*

$$\mathbf{X}_+ := \mathbf{E}^{\mathbf{H}^+} \mathbf{H}^- \quad (7.4.14)$$

*is a minimal Markovian splitting subspace and  $\mathbf{X}_+ \sim ((\mathbf{N}^+)^{\perp}, \mathbf{H}^+)$ , where  $\mathbf{N}^+$  is given by (7.4.13).*

Theorem 7.4.3 has some important corollaries. The first, which is nontrivial only if  $\mathbf{X}$  is infinite-dimensional, concerns the existence of minimal Markovian splitting subspaces.

**Corollary 7.4.7.** *Every Markovian splitting subspace contains a minimal Markovian splitting subspace.*

**Corollary 7.4.8.** *A Markovian splitting subspace  $\mathbf{X} \sim (\mathbf{S}, \bar{\mathbf{S}})$  is a minimal Markovian splitting subspace if and only if*

$$\bar{\mathbf{S}} = \mathbf{H}^+ \vee \mathbf{S}^{\perp} \quad \text{and} \quad \mathbf{S} = \mathbf{H}^- \vee \bar{\mathbf{S}}^{\perp}. \quad (7.4.15)$$

This corollary shows that  $\mathbf{X} \sim (\mathbf{S}, \bar{\mathbf{S}})$  is minimal if and only if there is equality in both the inclusions (7.4.8). Next, we show that these minimality conditions on  $\bar{\mathbf{S}}$  and  $\mathbf{S}$  correspond to observability and constructibility respectively.

**Theorem 7.4.9.** *A Markovian splitting subspace  $\mathbf{X} \sim (\mathbf{S}, \bar{\mathbf{S}})$  is observable if and only if*

$$\bar{\mathbf{S}} = \mathbf{H}^+ \vee \mathbf{S}^{\perp} \quad (7.4.16)$$

*and constructible if and only if*

$$\mathbf{S} = \mathbf{H}^- \vee \bar{\mathbf{S}}^{\perp}. \quad (7.4.17)$$

**Proof.** First note that if  $\mathbf{A}$  and  $\mathbf{B}$  are any subspaces, then  $(\mathbf{A} \vee \mathbf{B})^{\perp} = \mathbf{A}^{\perp} \cap \mathbf{B}^{\perp}$ . Now, condition (7.4.16) is equivalent to

$$[\mathbf{H}^+ \vee \mathbf{S}^{\perp}] \oplus \bar{\mathbf{S}}^{\perp} = \mathbb{H}_{\mathbf{X}}, \quad (7.4.18)$$

which can also be written

$$\mathbf{H}^+ \vee \mathbf{S}^\perp \vee \bar{\mathbf{S}}^\perp = \mathbb{H}_{\mathbf{X}} \quad (7.4.19)$$

Clearly (7.4.18) implies (7.4.19). To see that the converse is also true, note that the first of equations (7.4.15) implies  $(\mathbf{H}^+ \vee \mathbf{S}^\perp) \perp \bar{\mathbf{S}}^\perp$ . But (7.4.19) is equivalent to

$$(\mathbf{H}^+)^\perp \cap \mathbf{S} \cap \bar{\mathbf{S}} = 0$$

which in view of the fact that  $\mathbf{X} = \mathbf{S} \cap \bar{\mathbf{S}}$  is precisely the observability condition  $\mathbf{X} \cap (\mathbf{H}^+)^\perp = 0$ . The statement about constructibility follows by a symmetric argument.  $\square$

From Corollary 7.4.8 and Theorem 7.4.9 we have the following two corollaries, the second of which shows that the minimality property and the Markov property can be studied separately.

**Corollary 7.4.10.** *The Markovian splitting subspace  $\mathbf{X} \sim (\mathbf{S}, \bar{\mathbf{S}})$  is minimal if and only if it is both observable and constructible.*

**Corollary 7.4.11.** *A minimal Markovian splitting subspace is a minimal splitting subspace.*

**Corollary 7.4.12.** *A subspace  $\mathbf{X}$  is an observable Markovian splitting subspace if and only if there is a subspace  $\mathbf{S} \supset \mathbf{H}^-$ , satisfying  $U^*\mathbf{S} \subset \mathbf{S}$ , such that*

$$\mathbf{X} = E^{\mathbf{S}} \mathbf{H}^+. \quad (7.4.20)$$

*It is a constructible Markovian splitting subspace if and only if there is a subspace  $\bar{\mathbf{S}} \supset \mathbf{H}^+$ , satisfying  $U\bar{\mathbf{S}} \subset \bar{\mathbf{S}}$ , such that*

$$\mathbf{X} = E^{\bar{\mathbf{S}}} \mathbf{H}^-. \quad (7.4.21)$$

*The subspaces  $\mathbf{S}$  and  $\bar{\mathbf{S}}$  are those of Theorem 7.4.9; i.e.,  $\mathbf{X} \sim (\mathbf{S}, \bar{\mathbf{S}})$ .*

**Proof.** Suppose that  $\mathbf{X} \sim (\mathbf{S}, \bar{\mathbf{S}})$  is an observable Markovian splitting subspace. Then  $\mathbf{X} = E^{\mathbf{S}} \bar{\mathbf{S}}$  (Theorem 7.3.6), which together with the observability condition (7.4.16) yields (7.4.20). Conversely, suppose that there is an invariant  $\mathbf{S} \supset \mathbf{H}^-$  such that (7.4.20) holds. Define  $\bar{\mathbf{S}} := \mathbf{H}^+ \vee \mathbf{S}^\perp$ , which is clearly invariant under  $U$ . Then  $\mathbf{X} = E^{\mathbf{S}} \bar{\mathbf{S}}$ , and  $\mathbf{S}$  and  $\bar{\mathbf{S}}$  intersect perpendicularly (Theorem 7.2.4), hence satisfying the equivalent conditions of Proposition 7.2.2. Therefore,  $\mathbf{X}$  is a splitting subspace (Theorem 7.3.6), which satisfies (7.4.4) and therefore is Markovian. Consequently,  $\mathbf{X} \sim (\mathbf{S}, \bar{\mathbf{S}})$ . By construction,  $\mathbf{X}$  satisfies (7.4.16) and is thus observable. The rest follows by a symmetric argument.  $\square$

It follows from either Lemma 2.2.11 or from Propositions 7.4.5 and 7.4.6 that

$$\mathbf{H}^- = \mathbf{X}_- \oplus \mathbf{N}^- \quad \text{and} \quad \mathbf{H}^+ = \mathbf{X}_+ \oplus \mathbf{N}^+ \quad (7.4.22)$$

and, therefore, since  $\mathbf{H} = \mathbf{H}^- \vee \mathbf{H}^+$  and since  $\mathbf{X}_- \subset \mathbf{H}^- \perp \mathbf{N}^+$  and  $\mathbf{X}_+ \subset \mathbf{H}^+ \perp \mathbf{N}^-$ , we have the orthogonal decomposition

$$\mathbf{H} = \mathbf{N}^- \oplus \mathbf{H}^\square \oplus \mathbf{N}^+, \quad (7.4.23)$$

where  $\mathbf{H}^\square$  is the *frame space*

$$\mathbf{H}^\square = \mathbf{X}_- \vee \mathbf{X}_+. \quad (7.4.24)$$

The following result, which holds for splitting subspaces in general and not only for Markovian splitting subspaces, describes the role played by the predictor spaces  $\mathbf{X}_-$  and  $\mathbf{X}_+$  in Kalman filtering (see Section 6.9).

**Proposition 7.4.13.** *Let  $\mathbf{X}$  be a splitting subspace, and let  $\mathbf{N}^-$  and  $\mathbf{N}^+$  be defined by (7.4.11) and (7.4.13) respectively. Then*

$$E^{\mathbf{H}^-} \mathbf{X} = \mathbf{X}_- \quad (7.4.25)$$

*if and only if  $\mathbf{X} \perp \mathbf{N}^-$ , and*

$$E^{\mathbf{H}^+} \mathbf{X} = \mathbf{X}_+ \quad (7.4.26)$$

*if and only if  $\mathbf{X} \perp \mathbf{N}^+$ .*

**Proof.** Applying the projection  $E^{\mathbf{H}^-}$  to  $\mathbf{X} = E^{\mathbf{S}} \bar{\mathbf{S}}$  (Theorem 7.3.6) and noting that  $\mathbf{H}^- \subset \mathbf{S}$ , we obtain  $E^{\mathbf{H}^-} \mathbf{X} = E^{\mathbf{H}^-} \bar{\mathbf{S}}$ . But  $\bar{\mathbf{S}} \supset \mathbf{H}^+$ , and hence  $E^{\mathbf{H}^-} \mathbf{X} \supset \mathbf{X}_-$ . Conversely, suppose that  $\xi \in \mathbf{X}$ . Then, since  $\mathbf{H}^- = \mathbf{X}_- \oplus \mathbf{N}^-$ ,  $E^{\mathbf{H}^-} \xi = E^{\mathbf{X}_-} \xi + E^{\mathbf{N}^-} \xi$ , showing that  $E^{\mathbf{H}^-} \mathbf{X} \subset \mathbf{X}_-$  if and only if  $\mathbf{X} \perp \mathbf{N}^-$ . This establishes the first part. The second follows by symmetry.  $\square$

In particular, the conditions  $\mathbf{X} \perp \mathbf{N}^-$  and  $\mathbf{X} \perp \mathbf{N}^+$  can be replaced by the stronger conditions that  $\mathbf{X}$  be observable and constructible respectively.

**Corollary 7.4.14.** *Let  $\mathbf{X}$  be a splitting subspace. Then  $\mathbf{X} \perp \mathbf{N}^-$  if  $\mathbf{X}$  is observable and  $\mathbf{X} \perp \mathbf{N}^+$  if  $\mathbf{X}$  is constructible. If  $\mathbf{X}$  is minimal, it is orthogonal to both  $\mathbf{N}^-$  and  $\mathbf{N}^+$ .*

**Proof.** If  $\mathbf{X}$  is observable, then, by (7.3.12a),  $\mathbf{X} = E^{\mathbf{X}} \mathbf{H}^+$ . But  $\mathbf{X}$  is a splitting subspace satisfying (7.3.2) so that  $\mathbf{X} = E^{\mathbf{H}^- \vee \mathbf{X}} \mathbf{H}^+$ , and hence

$$E^{\mathbf{H}^-} \mathbf{X} = E^{\mathbf{H}^-} E^{\mathbf{H}^- \vee \mathbf{X}} \mathbf{H}^+ = E^{\mathbf{H}^-} \mathbf{H}^+,$$

which yields (7.4.25). Therefore it follows from Proposition 7.4.13 that  $\mathbf{X} \perp \mathbf{N}^-$ . In the same way we show that  $\mathbf{X} \perp \mathbf{N}^+$  is a consequence of  $\mathbf{X}$  being constructible. Then, the last statement follows from Theorem 7.3.5.  $\square$

We are now in a position to show that the frame space  $\mathbf{H}^\square$  is actually the closed linear hull of the interior parts  $\mathbf{X} \cap \mathbf{H}$  of all minimal splitting subspaces.

**Proposition 7.4.15.** *The frame space  $\mathbf{H}^\square$  is a Markovian splitting subspace, and  $\mathbf{H}^\square \sim ((\mathbf{N}^+)^\perp, (\mathbf{N}^-)^\perp)$ . Moreover,*

$$\mathbf{X} \cap \mathbf{H} \subset \mathbf{H}^\square \tag{7.4.27}$$

for all minimal splitting subspaces  $\mathbf{X}$ .

**Proof.** Since  $\mathbf{S} := (\mathbf{N}^+)^\perp = \mathbf{H}^- \vee (\mathbf{H}^+)^\perp \supset \mathbf{H}^-$  and  $\bar{\mathbf{S}} := (\mathbf{N}^-)^\perp = \mathbf{H}^+ \vee (\mathbf{H}^-)^\perp \supset \mathbf{H}^+$ , the first statement follows by comparing the decomposition (7.4.23) to (7.4.9) and noting that the invariance conditions (7.4.4) hold. The inclusion (7.4.27) follows from (7.4.23) and Corollary 7.4.14.  $\square$

Decomposition (7.4.23) partitions the output space  $\mathbf{H}$  into three parts. The subspace  $\mathbf{N}^-$  is the part of the past  $\mathbf{H}^-$  which is orthogonal to the future  $\mathbf{H}^+$ , and  $\mathbf{N}^+$  is the part of the future which is orthogonal to the past. Consequently, the inclusion (7.4.27) reflects the fact that the spaces  $\mathbf{N}^-$  and  $\mathbf{N}^+$  play no role in the interaction between past and future and hence in minimal state space construction. As we shall see in Chapter 16, decomposition (7.4.23) also provides an important conceptual paradigm for smoothing. In fact, it follows immediately from Corollary 7.4.14 that

$$\mathbf{E}^{\mathbf{H}} \mathbf{X} \subset \mathbf{H}^\square \tag{7.4.28}$$

for any minimal  $\mathbf{X}$ , relating the smoothing estimate to the forward and backward predictor estimates.

## 7.5 The Markov semigroup

Defining a semigroup on a splitting subspace  $\mathbf{X}$  in the style of Section 7.1 requires that  $\mathbf{X}$  has a scattering pair  $(\mathbf{S}, \bar{\mathbf{S}})$  satisfying the invariance properties  $U^* \mathbf{S} \subset \mathbf{S}$  and  $U \bar{\mathbf{S}} \subset \bar{\mathbf{S}}$ ; i.e.  $\mathbf{X} \sim (\mathbf{S}, \bar{\mathbf{S}})$  must be a Markovian splitting subspace. Then, defining the restricted shift on  $\mathbf{X}$ ,

$$U(\mathbf{X}) = \mathbf{E}^{\mathbf{X}} U|_{\mathbf{X}}, \tag{7.5.1}$$

or more generally

$$U_t(\mathbf{X}) = \mathbf{E}^{\mathbf{X}} U^t|_{\mathbf{X}}, \quad t = 0, 1, 2, \dots, \tag{7.5.2}$$

we have the following theorem.

**Theorem 7.5.1.** *Let  $\mathbf{X} \sim (\mathbf{S}, \bar{\mathbf{S}})$  be a Markovian splitting subspace. Then, for  $t = 0, 1, 2, \dots$ , the diagrams*

$$\begin{array}{ccc} \mathbf{H}^+ & \xrightarrow{\mathcal{O}^*} & \mathbf{X} & & \mathbf{H}^- & \xrightarrow{\mathcal{C}^*} & \mathbf{X} \\ U^t \downarrow & & \downarrow U_t(\mathbf{X}) & & (U^*)^t \downarrow & & \downarrow U_t(\mathbf{X})^* \\ \mathbf{H}^+ & \xrightarrow{\mathcal{O}^*} & \mathbf{X} & & \mathbf{H}^- & \xrightarrow{\mathcal{C}^*} & \mathbf{X} \end{array}$$

commute, where  $\mathcal{O}$  is the observability operator  $\mathbf{E}^{\mathbf{H}^+}|_{\mathbf{X}}$  and  $\mathcal{C}$  is the constructibility operator  $\mathbf{E}^{\mathbf{H}^-}|_{\mathbf{X}}$ . Moreover, the restricted shift satisfies the semigroup property

$$U_s(\mathbf{X})U_t(\mathbf{X}) = U_{s+t}(\mathbf{X}); \tag{7.5.3}$$

i.e., in particular,

$$U_t(\mathbf{X}) = U(\mathbf{X})^t. \tag{7.5.4}$$

For each  $\xi \in \mathbf{X}$  and  $t = 0, 1, 2, \dots$ ,

$$E^{\mathbf{S}} U^t \xi = U_t(\mathbf{X}) \xi \tag{7.5.5}$$

$$E^{\bar{\mathbf{S}}} U^{-t} \xi = U_t(\mathbf{X})^* \xi. \tag{7.5.6}$$

A Markovian splitting subspace  $\mathbf{X} \sim (\mathbf{S}, \bar{\mathbf{S}})$  will be called *proper* if it satisfies both (10.2.6) and (10.2.7).

**Proof.** Let  $\lambda \in \bar{\mathbf{S}}$  and take  $t = 0, 1, 2, \dots$ . Then, since  $\bar{\mathbf{S}} = \mathbf{X} \oplus \mathbf{S}^\perp$  (Theorem 7.2.4),

$$E^{\mathbf{X}} U^t \lambda = E^{\mathbf{X}} U^t E^{\mathbf{X}} \lambda + E^{\mathbf{X}} U^t E^{\mathbf{S}^\perp} \lambda$$

However, the last term is zero, since  $U^t \mathbf{S}^\perp \subset \mathbf{S}^\perp \perp \mathbf{X}$ . Therefore,

$$E^{\mathbf{X}} U^t \lambda = E^{\mathbf{X}} U^t E^{\mathbf{X}} \lambda. \tag{7.5.7}$$

Consequently, for any  $\lambda \in \mathbf{H}^+ \subset \bar{\mathbf{S}}$ ,

$$U_t(\mathbf{X}) \mathcal{O}^* \lambda = E^{\mathbf{X}} U^t E^{\mathbf{X}} \lambda = E^{\mathbf{X}} U^t \lambda = \mathcal{O}^* U^t \lambda,$$

and thus the first diagram commutes. A completely symmetric argument show that also the second diagram commutes. From (7.5.7) we also immediately see that (7.5.3) holds. Moreover, since  $\mathbf{S} \perp \bar{\mathbf{X}} \mid \mathbf{X}$  and  $U \lambda \in \bar{\mathbf{S}}$ , the left member of (7.5.7) can be exchanged for  $E^{\bar{\mathbf{S}}} U^t \lambda$ . Therefore, since  $\mathbf{X} \subset \bar{\mathbf{S}}$ , (7.5.5) follows. Then (7.5.6) follows by symmetry.  $\square$

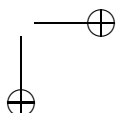
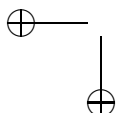
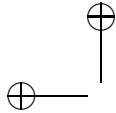
## 7.6 Minimality and dimension

In the geometric theory of splitting subspaces minimality is defined in terms of subspace inclusion. This is natural since this concept of minimality is meaningful also for infinite-dimensional splitting subspaces. The question of whether minimal splitting subspaces all have the same (finite or infinite) dimension is therefore natural.

**Theorem 7.6.1.** *All minimal (Markovian or non-Markovian) splitting subspaces have the same dimension.*

As a preliminary for proving this theorem let us again consider the splitting factorization

$$\begin{array}{ccc} \mathbf{H}^- & \xrightarrow{\mathcal{H}} & \mathbf{H}^+ \\ \mathcal{C}^* \searrow & & \nearrow \mathcal{O} \\ & \mathbf{X} & \end{array} \quad \mathcal{H} = \mathcal{O} \mathcal{C}^*$$





described in Section 7.3. Recall that  $\mathbf{X}$  is observable if and only if  $\text{Im } \mathcal{O}^*$  is dense in  $\mathbf{X}$  and constructible if and only if  $\text{Im } \mathcal{C}^*$  is dense in  $\mathbf{X}$ . We shall say that  $\mathbf{X}$  is *exactly observable* if  $\mathcal{O}^*$  is surjective,  $\text{Im } \mathcal{O}^* = \mathbf{X}$  and *exactly constructible* if  $\mathcal{C}^*$  is surjective, i.e.,  $\text{Im } \mathcal{C}^* = \mathbf{X}$ . If  $\mathbf{X}$  is both exactly observable and exactly constructible, we say that the factorization, and hence also  $\mathbf{X}$ , is *exactly canonical*.

**Lemma 7.6.2.** *If the Hankel operator  $\mathcal{H} := E^{\mathbf{H}^+} |_{\mathbf{H}^-}$  has a closed range, then all minimal splitting subspaces are exactly canonical. If one splitting subspace is exactly canonical, then  $\mathcal{H}$  has a closed range.*

**Proof.** Recall that if a map has a closed range, then so does its adjoint [127, p.205]; this will be used several times in the proof. Let  $\mathbf{X}$  be a minimal splitting subspace. Then  $\mathcal{H} = \mathcal{O}\mathcal{C}^*$  and  $\mathcal{C}^*\mathbf{H}^-$  is dense in  $\mathbf{X}$ . Clearly  $\mathcal{H}\mathbf{H}^- = \mathcal{O}\mathcal{C}^*\mathbf{H}^- \subset \mathcal{O}\mathbf{X}$ . We want to show that, if  $\mathcal{H}\mathbf{H}^-$  is closed, then  $\mathcal{H}\mathbf{H}^- = \mathcal{O}\mathbf{X}$  so that  $\mathcal{O}$ , and hence  $\mathcal{O}^*$ , has a closed range, i.e.,  $\mathbf{X}$  is exactly observable. To this end, let  $\xi \in \mathbf{X}$  be arbitrary. Then there is a sequence  $\{\xi_k\}$  in  $\mathcal{C}^*\mathbf{H}^-$  such that  $\xi_k \rightarrow \xi$  as  $k \rightarrow \infty$ . But  $\mathcal{O}\xi_k \in \mathcal{H}\mathbf{H}^-$ , and, since  $\mathcal{O}$  is continuous,  $\mathcal{O}\xi_k \rightarrow \mathcal{O}\xi \in \mathcal{H}\mathbf{H}^-$ , and consequently  $\mathcal{O}\mathbf{X} \subset \mathcal{H}\mathbf{H}^-$ . Hence, since  $\mathcal{O}\mathbf{X} \supset \mathcal{H}\mathbf{H}^-$  trivially,  $\mathcal{O}\mathbf{X} = \mathcal{H}\mathbf{H}^-$  as required. In the same way, we use the adjoint factorization  $\mathcal{H}^* = \mathcal{C}\mathcal{O}^*$ , which is also canonical, to prove that  $\mathbf{X}$  is exactly constructible. Conversely, assume that  $\mathbf{X}$  is exactly canonical. Then  $\mathcal{C}^*\mathbf{H}^- = \mathbf{X}$ , and therefore, since  $\mathcal{O}\mathbf{X}$  is closed,  $\mathcal{H}\mathbf{H}^- = \mathcal{O}\mathcal{C}^*\mathbf{H}^-$  is closed.  $\square$

Certain results in the geometric theory of splitting subspaces are much easier to prove in the finite-dimensional case. The reason for this is that the ranges of the operators  $\mathcal{O}^*$  and  $\mathcal{C}^*$  are always closed in this case. Hence it is the fact that observability and constructibility is always exact in the finite-dimensional case which implies that certain technical difficulties do not occur.

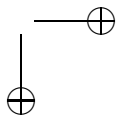
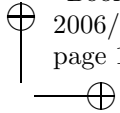
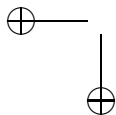
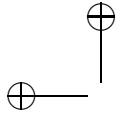
**Proof of Theorem 7.6.1.** Let us first assume that the Hankel operator  $\mathcal{H}$  has closed range. Then, for any minimal  $\mathbf{X}$ ,  $\mathcal{C}^*$  is surjective and  $\mathcal{O}$  injective. Now, suppose  $\mathbf{X}_1$  and  $\mathbf{X}_2$  are two minimal splitting subspaces. Then, if, for  $i=1,2$ ,  $\mathcal{O}_i$  is the observability operator and  $\mathcal{C}_i$  the constructability operator of  $\mathbf{X}_i$ , the diagram

$$\begin{array}{ccccc}
 & & \mathbf{X}_1 & & \\
 \mathcal{C}_1^* \nearrow & & | & \searrow \mathcal{O}_1 & \\
 \mathbf{H}^- & & T & & \mathbf{H}^+ \\
 \mathcal{C}_2^* \searrow & & | & \nearrow \mathcal{O}_2 & \\
 & & \mathbf{X}_2 & & 
 \end{array}$$

commutes. We want to show that there is a bijective linear operator  $T : \mathbf{X}_1 \rightarrow \mathbf{X}_2$  so that the diagram amended with the dashed arrow still commutes.

Since  $\mathcal{C}_1^*$  is surjective, for each  $\xi_1 \in \mathbf{X}_1$  there is a  $\lambda \in \mathbf{H}^-$  such that  $\mathcal{C}_1^*\lambda = \xi_1$ . For any such  $\lambda \in \mathbf{H}^-$  commutativity yields

$$\mathcal{O}_1\mathcal{C}_1^*\lambda = \mu = \mathcal{O}_2\mathcal{C}_2^*\lambda.$$



Moreover, since  $\mathcal{O}_2$  is injective, there is a unique  $\xi_2 \in \mathbf{X}$  such that  $\mu = \mathcal{O}_2\xi_2$ . Define  $T : \mathbf{X}_1 \rightarrow \mathbf{X}_2$  to be the linear map sending  $\xi_1$  to  $\xi_2$ . Then,

$$\mathcal{O}_2T\mathcal{C}_1^*\lambda = \mathcal{O}_2T\xi_1 = \mathcal{O}_2\xi_2 = \mu = \mathcal{O}_2\mathcal{C}_2^*\lambda.$$

Since  $\mathcal{O}_2$  is injective, this implies that  $T\mathcal{C}_1^* = \mathcal{C}_2^*$ , so the left triangle in the diagram commutes. To see that the right triangle in the diagram also commutes, note that  $\mu = \mathcal{O}_1\xi_1$  and  $\mu = \mathcal{O}_2\xi_2 = \mathcal{O}_2T\xi_1$ , which implies that  $\mathcal{O}_1 = \mathcal{O}_2T$ .

Next, since  $\mathcal{C}_2^*$  is surjective and  $\mathcal{O}_1$  is injective, a completely symmetric argument shows that there is a map  $\tilde{T} : \mathbf{X}_2 \rightarrow \mathbf{X}_1$  such that the diagram amended with this map also commutes. But, then  $\tilde{T}T$  must be identity in  $\mathbf{X}_1$  and  $T\tilde{T}$  the identity in  $\mathbf{X}_2$ , and hence  $\tilde{T} = T^{-1}$ . Consequently,  $\mathbf{X}_1$  and  $\mathbf{X}_2$  are isomorphic as vector spaces and hence they have the same dimension. It remains to consider the case in which  $\mathcal{H}$  does not have a closed range. But then, by Lemma 7.6.2, no minimal splitting subspace is exactly canonical, and consequently all are infinite-dimensional. Therefore, since  $\mathbb{H}$  is a separable Hilbert space, all  $\mathbf{X}$  have dimension  $\aleph_0$ .  $\square$

**Corollary 7.6.3.** *A finite-dimensional splitting subspace is minimal if and only if its dimension is minimal.*

**Proof.** Let  $\mathbf{X}$  be a finite-dimensional splitting subspace. First assume that there is a splitting subspace  $\mathbf{X}_1$  of smaller dimension than  $\mathbf{X}$ . By Corollary 7.4.7,  $\mathbf{X}_1$  contains a minimal splitting subspace  $\mathbf{X}_2$ . Since  $\dim \mathbf{X}_2 \leq \dim \mathbf{X}_1 < \dim \mathbf{X}$ , Theorem 7.6.1 implies that  $\mathbf{X}$  is nonminimal. Conversely, suppose  $\mathbf{X}$  is not minimal. Then it contains a minimal splitting subspace as a proper subspace (Corollary 7.4.7), and thus  $\mathbf{X}$  cannot have minimal dimension.  $\square$

Recall that the conditions  $\mathbf{X} \perp \mathbf{N}^-$  and  $\mathbf{X} \perp \mathbf{N}^+$  are weaker than observability and constructibility respectively (Corollary 7.4.14). Nevertheless, we have the following alternative characterization of minimality in the case that  $\mathcal{H}$  has a closed range, and, in particular, when  $\mathbf{X}$  is finite dimensional.

**Theorem 7.6.4.** *Suppose that the Hankel operator  $\mathcal{H} := E^{\mathbf{H}^+}|_{\mathbf{H}^-}$  has a closed range. Then, for any splitting subspace  $\mathbf{X}$ , the following conditions are equivalent.*

- (i)  $\mathbf{X}$  is minimal
- (ii)  $\mathbf{X}$  is observable and  $\mathbf{X} \perp \mathbf{N}^+$
- (iii)  $\mathbf{X}$  is constructible and  $\mathbf{X} \perp \mathbf{N}^-$

**Proof.** It follows from Corollaries 7.4.10 and 7.4.14 that (i) implies (ii) and (iii). To prove the converse, first assume that (ii) holds. Then, in view of Proposition 7.4.13, we have  $E^{\mathbf{H}^+} \mathbf{X} = \mathbf{X}_+$ , and therefore  $\text{Im } \mathcal{O} = \mathbf{X}_+$ . Hence we can restrict the range

of  $\mathcal{O}$  to  $\mathbf{X}_+$  to obtain

$$\begin{array}{ccc} \mathbf{H}^- & \xrightarrow{\mathcal{G}} & \mathbf{X}_+ \\ \mathcal{C}^* \searrow & & \nearrow \hat{\mathcal{O}} \\ & \mathbf{X} & \end{array}$$

where  $\hat{\mathcal{O}} := E^{\mathbf{X}_+} |_{\mathbf{X}}$  and  $\mathcal{G} := E^{\mathbf{X}_+} |_{\mathbf{H}^-}$ . The restricted observability operator  $\hat{\mathcal{O}}$  is both injective (observability) and surjective, i.e.  $\hat{\mathcal{O}}$  is bijective so that the inverse  $\hat{\mathcal{O}}^{-1} : \mathbf{X}_+ \rightarrow \mathbf{X}$  is well-defined and onto. Consequently,  $\mathcal{C}^* = \hat{\mathcal{O}}^{-1}\mathcal{G}$  is onto, i.e.,  $\mathbf{X}$  is constructible. Hence  $\mathbf{X}$  is minimal (Corollary 7.4.10), as claimed. A symmetric argument shows that (iii) implies (i).  $\square$

Another version of Theorem 7.6.4 for Markovian splitting subspaces, which does not require the condition that  $\mathcal{H}$  has closed range, will be given in Chapter 8 (Theorem 9.2.17).

In the next section, we shall need the following corollary. Since any minimal splitting subspace is orthogonal to both  $\mathbf{N}^-$  and to  $\mathbf{N}^+$  (Corollary 7.4.14), the splitting condition  $\mathbf{H}^- \perp \mathbf{H}^+ | \mathbf{X}$  is equivalent to

$$\mathbf{X}_- \perp \mathbf{X}_+ | \mathbf{X}, \tag{7.6.1}$$

where  $\mathbf{N}^-$  and  $\mathbf{N}^+$  have been removed from the past and the future. We shall restrict the observability and constructability operators accordingly.

**Corollary 7.6.5.** *Let  $\mathbf{X}$  be a minimal splitting subspace. Then, the restricted observability and constructability operators,  $\hat{\mathcal{O}} : \mathbf{X} \rightarrow \mathbf{X}_-$  and  $\hat{\mathcal{C}} : \mathbf{X} \rightarrow \mathbf{X}_+$  respectively, of  $\mathbf{X}$ , defined by*

$$\hat{\mathcal{O}} := E^{\mathbf{X}_+} |_{\mathbf{X}} \quad \text{and} \quad \hat{\mathcal{C}} := E^{\mathbf{X}_-} |_{\mathbf{X}},$$

as well as their adjoints

$$\hat{\mathcal{O}}^* := E^{\mathbf{X}} |_{\mathbf{X}_+} \quad \text{and} \quad \hat{\mathcal{C}}^* := E^{\mathbf{X}} |_{\mathbf{X}_-},$$

are quasi-invertible; i.e., one-one and densely onto. Moreover,

$$\hat{\mathcal{O}}\hat{\mathcal{C}}^* = \hat{\mathcal{O}}_-, \tag{7.6.2}$$

where  $\hat{\mathcal{O}}_-$  is the restricted observability operator of  $\mathbf{X}_-$ .

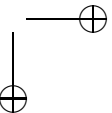
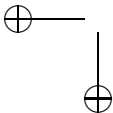
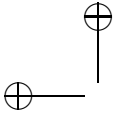
**Proof.** It follows from Proposition 2.4.2(vi) that (7.6.2) is equivalent to the restricted splitting condition (7.6.1). This establishes the last statement.

In view of Corollary 7.4.14 and Proposition 7.4.13, (7.4.25) holds, and hence, since  $\mathbf{X}_+ \subset \mathbf{H}^+$ ,

$$E^{\mathbf{X}_+} \mathbf{X} = E^{\mathbf{X}_+} E^{\mathbf{H}^+} \mathbf{X} = \mathbf{X}_+,$$

from which it follows that  $\text{Im } \hat{\mathcal{O}}$  is dense in  $\mathbf{X}_+$ . Moreover, in view of Proposition 7.4.6,

$$\ker \hat{\mathcal{O}} = \mathbf{X} \cap (\mathbf{X}_+)^{\perp} = \mathbf{X} \cap (\mathbf{N}^+ \oplus (\mathbf{H}^+)^{\perp}) = \mathbf{X} \cap (\mathbf{H}^+)^{\perp} = 0,$$



by the observability of  $\mathbf{X}$ . This establishes that  $\hat{\mathcal{O}}$  is one-one and densely onto. A symmetric argument shows that  $\hat{\mathcal{C}}$  has the same property. The statements about  $\hat{\mathcal{O}}^*$  and  $\hat{\mathcal{C}}^*$  then follow by Theorem A.1.3.  $\square$

As readily seen from the proof, this corollary can be strengthened in a form that will be useful in Chapter 8.

**Corollary 7.6.6.** *The operator  $\hat{\mathcal{O}}$  ( $\hat{\mathcal{O}}^*$ ), defined in Corollary 7.6.5, is quasi-invertible if and only if  $\mathbf{X}$  is observable and  $\mathbf{X} \perp \mathbf{N}^+$ . Moreover, for all  $t \geq 0$ ,*

$$U_t(\mathbf{X})\hat{\mathcal{O}}^* = \hat{\mathcal{O}}^*U_t(\mathbf{X}_+). \tag{7.6.3}$$

*Likewise,  $\hat{\mathcal{C}}$  ( $\hat{\mathcal{C}}^*$ ) is quasi-invertible if and only if  $\mathbf{X}$  is constructible and  $\mathbf{X} \perp \mathbf{N}^+$ . Moreover,*

$$U_t(\mathbf{X})\hat{\mathcal{C}}^* = \hat{\mathcal{C}}^*U_t(\mathbf{X}_-) \tag{7.6.4}$$

for all  $t \geq 0$ .

**Proof.** The statements concerning the quasi-invertibility of  $\hat{\mathcal{O}}$ ,  $\hat{\mathcal{O}}^*$ ,  $\hat{\mathcal{C}}$  and  $\hat{\mathcal{C}}^*$  follow from the proof of Corollary 7.6.5. To prove (7.6.3), consider the commutative diagrams of Theorem 7.5.1. First take  $\xi \in \mathbf{X}_+ \subset \mathbf{H}^+$ . Then, the first commutative diagram yields

$$U_t(\mathbf{X})\hat{\mathcal{O}}^*\xi = E^{\mathbf{X}}U^t\xi = E^{\mathbf{X}}E^{(\mathbf{H}^+)^{\perp}}U^t\xi + E^{\mathbf{X}}E^{\mathbf{X}_+}U^t\xi + E^{\mathbf{X}}E^{\mathbf{N}^+}U^t\xi,$$

since  $\mathbf{X}_+ \sim ((\mathbf{N}^+)^{\perp}, \mathbf{H}^+)$  and therefore  $\mathbf{H} = (\mathbf{H}^+)^{\perp} \oplus \mathbf{X}_+ \oplus \mathbf{N}^+$ . However, since  $U^t\xi \in \mathbf{H}^+$ , the first term is zero. Moreover, since  $\mathbf{X} \perp \mathbf{N}^+$ , the last term is also zero. This proves (7.6.3). The equation (7.6.4) follows by a symmetric argument.  $\square$

## 7.7 Partial ordering of minimal splitting subspaces

To investigate the structure of the family of minimal splitting subspaces, we introduce a partial ordering on this set.

**Definition 7.7.1.** *Given two minimal splitting subspaces,  $\mathbf{X}_1$  and  $\mathbf{X}_2$ , let  $\mathbf{X}_1 \prec \mathbf{X}_2$  denote the ordering*

$$\|E^{\mathbf{X}_1}\lambda\| \leq \|E^{\mathbf{X}_2}\lambda\| \quad \text{for all } \lambda \in \mathbf{H}^+, \tag{7.7.1}$$

where  $\|\cdot\|$  is the norm in the Hilbert space  $\mathbb{H}$ .

This partial ordering has the following interpretation. If  $\mathbf{X}_1 \prec \mathbf{X}_2$ , then  $\mathbf{X}_2$  is closer to the future  $\mathbf{H}^+$  than  $\mathbf{X}_1$  (or, loosely speaking, contains more information about the future than  $\mathbf{X}_1$ ) in the sense that for every subspace  $\mathbf{A}$  of  $\mathbf{H}^+$  we have

$$\alpha(\mathbf{X}_1, \mathbf{A}) \geq \alpha(\mathbf{X}_2, \mathbf{A}) \tag{7.7.2}$$

where  $\alpha(\mathbf{X}, \mathbf{A})$  is the *angle* between the subspaces  $\mathbf{X}$  and  $\mathbf{A}$ , defined in Chapter 2, Section 2.3.

The partial ordering (7.7.1) has actually a symmetric interpretation with respect to the past.

**Lemma 7.7.2.** *The relation  $\mathbf{X}_1 \prec \mathbf{X}_2$  holds if and only if*

$$\|E^{\mathbf{X}_2} \lambda\| \leq \|E^{\mathbf{X}_1} \lambda\| \quad \text{for all } \lambda \in \mathbf{H}^-. \quad (7.7.3)$$

**Proof.** Since  $\mathbf{X}_1$  and  $\mathbf{X}_2$  are minimal, they are orthogonal to  $\mathbf{N}^-$  and to  $\mathbf{N}^+$  (Corollary 7.4.14), and therefore, in view of (7.4.22), the condition (7.7.1) is equivalent to

$$\|E^{\mathbf{X}_1} \lambda\| \leq \|E^{\mathbf{X}_2} \lambda\| \quad \text{for all } \lambda \in \mathbf{X}_+ \quad (7.7.4)$$

and the condition (7.7.3) to

$$\|E^{\mathbf{X}_2} \lambda\| \leq \|E^{\mathbf{X}_1} \lambda\| \quad \text{for all } \lambda \in \mathbf{X}_-. \quad (7.7.5)$$

Now, for  $i = 1, 2$ , let  $\hat{O}_i$  and  $\hat{C}_i$  be the restricted observability and constructibility operator respectively of  $\mathbf{X}_i$ , as defined in Corollary 7.6.5, and let  $\hat{O}_i^*$  and  $\hat{C}_i^*$  be their adjoints. By Corollary 7.6.5, these operators are injective with dense range. In this notation, it thus remains to prove that

$$\|\hat{O}_1^* \lambda\| \leq \|\hat{O}_2^* \lambda\| \quad \text{for all } \lambda \in \mathbf{X}_+ \quad (7.7.6)$$

implies

$$\|\hat{C}_2^* \lambda\| \leq \|\hat{C}_1^* \lambda\| \quad \text{for all } \lambda \in \mathbf{X}_-. \quad (7.7.7)$$

Then, the converse will follow by symmetry.

From (7.7.6) it follows that

$$\|\hat{O}_1^*(\hat{O}_2^*)^{-1} \xi\| \leq \|\xi\|$$

for all  $\xi$  in a dense subset of  $\mathbf{X}_2$ . The operator  $T := \hat{O}_1^*(\hat{O}_2^*)^{-1}$  can be continuously extended to the rest of  $\mathbf{X}_2$  as a bounded operator with norm  $\|T\| \leq 1$ . In fact, for any  $\xi \in \mathbf{X}_2$ , there is a Cauchy sequence  $\{\xi_k\}$  such that

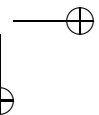
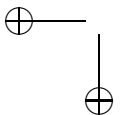
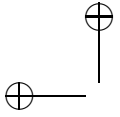
$$\|T\xi_k - T\xi_j\| \leq \|\xi_k - \xi_j\|,$$

implying that  $\{T\xi_k\}$  converges. Then define  $T\xi := \lim_{k \rightarrow \infty} T\xi_k$ . Since

$$\langle \eta, \hat{O}_1^*(\hat{O}_2^*)^{-1} \xi \rangle = \langle (\hat{O}_2)^{-1} \hat{O}_1 \eta, \xi \rangle$$

for all  $\xi$  in the range of  $\hat{O}_2^*$  and all  $\eta$ , the operator  $T^* := (\hat{O}_2)^{-1} \hat{O}_1$  is the adjoint of  $T$ . Now, in view of (7.6.2), we have  $\hat{O}_2 \hat{C}_2^* = \hat{O}_1 \hat{C}_1^*$ , and consequently, since  $\|T^*\| = \|T\| \leq 1$ ,

$$\|\hat{C}_2^* \lambda\| = \|T^* \hat{C}_1^* \lambda\| \leq \|\hat{C}_1^* \lambda\|$$



for all  $\lambda \in \mathbf{X}_+$ , which yields (7.7.7), as required.  $\square$

**Theorem 7.7.3.** *The family of minimal splitting subspaces has a unique minimal element  $\mathbf{X}_-$  and a unique maximal element  $\mathbf{X}_+$ , i.e.*

$$\mathbf{X}_- \prec \mathbf{X} \prec \mathbf{X}_+ \tag{7.7.8}$$

for all minimal  $\mathbf{X}$ , and these are precisely the predictor spaces

$$\mathbf{X}_- := \mathbf{E}^{\mathbf{H}^-} \mathbf{H}^+ \tag{7.7.9a}$$

$$\mathbf{X}_+ := \mathbf{E}^{\mathbf{H}^+} \mathbf{H}^- \tag{7.7.9b}$$

defined in Proposition 7.3.1.

**Proof.** Since  $\mathbf{E}^{\mathbf{X}}$  is a projector,

$$\|\mathbf{E}^{\mathbf{X}} \lambda\| \leq \|\lambda\| \quad \text{for all } \lambda \in \mathbf{X}_+ \tag{7.7.10}$$

However,  $\|\mathbf{E}^{\mathbf{X}_+} \lambda\| = \|\lambda\|$  for all  $\lambda \in \mathbf{X}_+$ , and consequently, in view of (7.7.4),  $\mathbf{X} \prec \mathbf{X}_+$ . Moreover, for each  $\mathbf{X} \neq \mathbf{X}_+$ , there is a  $\lambda$  in  $\mathbf{X}_+$  for which strict inequality holds in (7.7.10), which proves uniqueness. A symmetric argument using Lemma 7.7.2 gives the rest.  $\square$

Whenever both  $\mathbf{X}_1 \prec \mathbf{X}_2$  and  $\mathbf{X}_2 \prec \mathbf{X}_1$  hold, we say that  $\mathbf{X}_1$  and  $\mathbf{X}_2$  are *equivalent*, writing  $\mathbf{X}_1 \sim \mathbf{X}_2$ . We shall see below (Corollary 7.7.10) that, if at least one of  $\mathbf{X}_1$  and  $\mathbf{X}_2$  is internal,  $\mathbf{X}_1 \sim \mathbf{X}_2$  implies  $\mathbf{X}_1 = \mathbf{X}_2$ . Let us define  $\mathcal{X}$  to be the family of all equivalence classes of minimal splitting subspaces, and let  $\mathcal{X}_0$  be the subset of those  $\mathbf{X}$  which are internal ( $\mathbf{X} \subset \mathbf{H}$ ). Then the order relation (7.7.1) makes  $\mathcal{X}$  into a partially ordered set with a maximal and minimal element, namely  $\mathbf{X}_+$  and  $\mathbf{X}_-$ , respectively. Note that each equivalence class in  $\mathcal{X}_0$  is a singleton, and consequently  $\mathcal{X}_0$  is just a family of minimal  $\mathbf{X}$ .

### Uniform choices of bases

Next, for the finite-dimensional case, we shall illustrate the meaning of the partial ordering defined above in terms of covariance matrices. More precisely, we shall parametrize  $\mathcal{X}$  by a certain family of positive definite matrices. To this end, we introduce a *uniform choice of bases on  $\mathcal{X}$* . By Theorem 7.6.1, all  $\mathbf{X} \in \mathcal{X}$  have the same dimension, which we denote by  $n$ . Let  $(\xi_{+,1}, \xi_{+,2}, \dots, \xi_{+,n})$  be an arbitrary basis in  $\mathbf{X}_+$  and define

$$\xi_k = \mathbf{E}^{\mathbf{X}} \xi_{+,k}, \quad k = 1, 2, \dots, n \tag{7.7.11}$$

for every minimal splitting subspace  $\mathbf{X}$ .

**Lemma 7.7.4.** *The random variables  $(\xi_1, \xi_2, \dots, \xi_n)$  form a basis in  $\mathbf{X}$ .*

**Proof.** Since  $\hat{O}^* := E^{\mathbf{X}}|_{\mathbf{X}_+}$  is a bijection (Corollary 7.6.5), it sends a basis into a basis.  $\square$

For simplicity of notation, we introduce the vector notation

$$x = \begin{bmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_n \end{bmatrix}, \tag{7.7.12}$$

and define  $x_+$  accordingly in terms of  $(\xi_{+,1}, \xi_{+,2}, \dots, \xi_{+,n})$ .

Now, to each basis  $(\xi_1, \xi_2, \dots, \xi_n)$  we associate the covariance matrix

$$P = E\{xx'\}, \tag{7.7.13}$$

which is symmetric and positive definite. For a fixed choice of  $(\xi_{+,1}, \xi_{+,2}, \dots, \xi_{+,n})$ , let  $\mathcal{P}$  be the family of all covariance matrices (7.7.13) obtained as  $\mathbf{X}$  varies over all minimal splitting subspaces, and let  $\mathcal{P}_0$  be the subfamily generated by the internal  $\mathbf{X}$ . Note that  $\mathcal{P}$  is equipped with the natural ordering:  $P_1 \leq P_2$  if and only if  $P_2 - P_1$  is nonnegative definite.

**Proposition 7.7.5.** *There is a one-one correspondence between  $\mathcal{X}$  and  $\mathcal{P}$  which is order-preserving in the sense that  $P_1 \leq P_2$  if and only if  $\mathbf{X}_1 \prec \mathbf{X}_2$ .*

**Proof.** To each  $\lambda \in \mathbf{X}_+$ , there corresponds a unique  $a \in \mathbb{R}^n$  such that  $\lambda = a'x_+$ . By (7.7.11),  $E^{\mathbf{X}}\lambda = a'x$ , and hence

$$\|E^{\mathbf{X}}\lambda\|^2 = a'Pa \tag{7.7.14}$$

Therefore, in view of the ordering condition (7.7.4),  $\mathbf{X}_1 \prec \mathbf{X}_2$  if and only if  $P_1 \leq P_2$ . Moreover, from (7.7.14) we see that two  $\mathbf{X}$  have the same  $P$  if and only if they are equivalent, establishing the one-one correspondence between  $\mathcal{X}$  and  $\mathcal{P}$ .  $\square$

**Remark 7.7.6.** All the results on splitting subspaces in this section may be formulated instead for Markovian splitting subspaces, as we shall do in detail Section 8.6 for the discrete-time setting and in Section 10.4 for the continuous-time setting. If we take  $\mathcal{X}$  to be the family of minimal Markovian splitting subspaces, then  $\mathcal{P}$  will be precisely the set of covariance matrices introduced in Chapter 6; i.e., the set of all symmetric matrices  $P$  such that (6.8.17) holds. Here  $P_-$  corresponds to  $\mathbf{X}_-$  and  $P_+$  to  $\mathbf{X}_+$ .

The uniform choice of bases allows us to state some useful alternative characterizations of ordering in terms of splitting.

**Proposition 7.7.7.** *Let  $\mathbf{X}_1$  and  $\mathbf{X}_2$  be finite-dimensional minimal splitting subspaces, at least one of which is internal. Then,  $\mathbf{X}_1 \prec \mathbf{X}_2$  if and only if*

$$a'x_1 = E^{\mathbf{X}_1}a'x_2, \quad \text{for all } a \in \mathbb{R}^n, \tag{7.7.15}$$

for any uniform choice of basis  $x_1$  and  $x_2$  in  $\mathbf{X}_1$  and  $\mathbf{X}_2$ , respectively.

**Proof.** From (7.7.11) we see that (7.7.19) is equivalent to

$$E^{\mathbf{X}_1} \lambda = E^{\mathbf{X}_1} E^{\mathbf{X}_2} \lambda \quad \text{for all } \lambda \in \mathbf{X}_+, \quad (7.7.16)$$

which, due to the fact that  $\mathbf{X}_1$  and  $\mathbf{X}_2$  are orthogonal to  $\mathbf{N}_+ := \mathbf{H}^+ \ominus \mathbf{X}_+$  (Corollary 7.4.14), can be extended to all  $\lambda \in \mathbf{H}^+$ . This in turn is equivalent to

$$E^{\mathbf{X}_1} \lambda = E^{\mathbf{X}_1} E^{\mathbf{S}_2} \lambda \quad \text{for all } \lambda \in \mathbf{H}^+, \quad (7.7.17)$$

because of the splitting property of  $\mathbf{X}_2$ , i.e., to  $\mathbf{X}_1 \perp \mathbf{H}^+ \mid \mathbf{S}_2$ , or equivalently, to  $\mathbf{S}_1 \perp \bar{\mathbf{S}}_2 \mid \mathbf{S}_2$ , which holds if and only if

$$\mathbf{S}_1 \perp \mathbf{H}_2 \ominus \mathbf{S}_2 \quad (7.7.18)$$

(Proposition 2.4.2), where  $\mathbf{H}_2$  is the ambient space of  $\mathbf{X}_2$ . Now, first assume that  $\mathbf{X}_1$  is internal. Then, (7.7.18) is equivalent to  $\mathbf{S}_1 \subset \mathbf{S}_2$ , i.e.,  $\mathbf{X}_1 \prec \mathbf{X}_2$  (Theorem 7.7.11). Next, assume that  $\mathbf{X}_2$  is internal. The (7.7.18) is equivalent to  $\mathbf{S}_1 \subset \mathbf{S}_2 \oplus \mathbf{H}^\perp$ , or, equivalently,  $E^{\mathbf{H}} \mathbf{S}_1 \subset \mathbf{S}_2$ , i.e.  $\mathbf{X}_1 \prec \mathbf{X}_2$  (Theorem 7.7.11).  $\square$

**Proposition 7.7.8.** *Let  $\mathbf{X}$ ,  $\mathbf{X}_1$  and  $\mathbf{X}_2$  be finite-dimensional minimal Markovian splitting subspaces with  $\mathbf{X}_1$  and  $\mathbf{X}_2$  internal. Then, if  $\mathbf{X}_1 \prec \mathbf{X} \prec \mathbf{X}_2$ ,*

$$\mathbf{X}_1 \perp \mathbf{X}_2 \mid \mathbf{X}.$$

**Proof.** Let  $x, x_1$  and  $x_2$  be a uniform choice of bases in  $\mathbf{X}$ ,  $\mathbf{X}_1$  and  $\mathbf{X}_2$ , respectively. Then, applying Proposition 7.7.7 first to  $\mathbf{X}_1 \prec \mathbf{X}_2$  and then to  $\mathbf{X}_1 \prec \mathbf{X}$  and  $\mathbf{X} \prec \mathbf{X}_2$  we obtain two representations for  $x_1$  yielding the equation

$$E^{\mathbf{X}_1} a'x_2 = E^{\mathbf{X}_1} E^{\mathbf{X}} a'x_2, \quad \text{for all } a \in \mathbb{R}^n,$$

which is equivalent to  $\mathbf{X}_1 \perp \mathbf{X}_2 \mid \mathbf{X}$ .  $\square$

### Ordering and scattering pairs

It is useful to express the ordering between minimal  $\mathbf{X} \in \mathcal{X}$  in terms of geometric conditions of subspace inclusions. To this end, we need the following lemma.

**Lemma 7.7.9.** *Let  $\mathbf{X}_1$  and  $\mathbf{X}_2$  be two minimal splitting subspaces, and suppose  $(\mathbf{S}_1, \bar{\mathbf{S}}_1)$  is a scattering pair of  $\mathbf{X}_1$  and  $(\mathbf{S}_2, \bar{\mathbf{S}}_2)$  a scattering pairs of  $\mathbf{X}_2$ . Then  $\mathbf{X}_1 \prec \mathbf{X}_2$  if and only if*

$$\|E^{\mathbf{S}_1} \lambda\| \leq \|E^{\mathbf{S}_2} \lambda\|, \quad \text{for all } \lambda \in \mathbf{H}, \quad (7.7.19)$$

or, equivalently,

$$\|E^{\bar{\mathbf{S}}_2} \lambda\| \leq \|E^{\bar{\mathbf{S}}_1} \lambda\|, \quad \text{for all } \lambda \in \mathbf{H}. \quad (7.7.20)$$



**Proof.** In view of the splitting property (7.3.21) and the fact that  $\mathbf{H}^+ \subset \mathbf{S}_i$ ,  $i = 1, 2$ , (7.7.1) is equivalent to

$$\|E^{\mathbf{S}_1} \lambda\| \leq \|E^{\mathbf{S}_2} \lambda\|, \quad \text{for all } \lambda \in \mathbf{H}^+. \quad (7.7.21)$$

Therefore, to show that condition (7.7.19) is equivalent to  $\mathbf{X}_1 \prec \mathbf{X}_2$ , by Definition 7.7.1 we need to prove that (7.7.21) implies (7.7.19); the converse is obvious. Now, for  $i = 1, 2$ , let  $\mathbf{Z}_i$  be the orthogonal complement of  $\mathbf{H}^-$  in  $\mathbf{S}_i$ , i.e.  $\mathbf{S}_i = \mathbf{H}^- \oplus \mathbf{Z}_i$ . Then

$$\|E^{\mathbf{S}_i} \lambda\|^2 = \|E^{\mathbf{H}^-} \lambda\|^2 + \|E^{\mathbf{Z}_i} \lambda\|^2,$$

so it only remains to prove that, if

$$\|E^{\mathbf{Z}_1} \lambda\| \leq \|E^{\mathbf{Z}_2} \lambda\| \quad (7.7.22)$$

holds for all  $\lambda \in \mathbf{H}^+$ , then it also holds for all  $\lambda \in \mathbf{H} := \mathbf{H}^- \vee \mathbf{H}^+$ . To this end, suppose (7.7.22) holds for all  $\lambda \in \mathbf{H}^+$ . Since  $\mathbf{Z}_i \subset (\mathbf{H}^-)^\perp$  for  $i = 1, 2$ , it follows that

$$\|E^{\mathbf{Z}_1} E^{(\mathbf{H}^-)^\perp} \lambda\| \leq \|E^{\mathbf{Z}_2} E^{(\mathbf{H}^-)^\perp} \lambda\|, \quad \text{for all } \lambda \in \mathbf{H}^+. \quad (7.7.23)$$

But, by Lemma 2.2.6, we have

$$E^{(\mathbf{H}^-)^\perp} \mathbf{H}^+ = (\mathbf{H}^-)^\perp \ominus \mathbf{H}^\perp$$

and consequently (7.7.22) holds for all  $\lambda \in \mathbf{Y} := (\mathbf{H}^-)^\perp \ominus \mathbf{H}^\perp$ . The extension from  $\mathbf{Y}$  to all of  $\mathbf{H}$  is then trivial. In fact,  $\mathbf{H} = \mathbf{H}^- \oplus \mathbf{Y}$ , so for any  $\lambda \in \mathbf{H}$ , there is a unique representation  $\lambda = \mu + \eta$ , where  $\mu \in \mathbf{Y}$  and  $\eta \in \mathbf{H}^-$ . Moreover,  $E^{\mathbf{Z}_i} \lambda = E^{\mathbf{Z}_i} \mu$  for  $i = 1, 2$ , so if (7.7.22) holds for all  $\mu \in \mathbf{Y}$ , it also holds for all  $\lambda \in \mathbf{H}$ . This concludes the proof that (7.7.19) is equivalent to (7.7.1). A symmetric argument shows that (7.7.20) is equivalent to (7.7.3). Then the rest follows from Lemma 7.7.2.  $\square$

**Corollary 7.7.10.** *Let  $\mathbf{X}_1$  and  $\mathbf{X}_2$  be equivalent minimal splitting subspaces. Then, if one is internal,  $\mathbf{X}_1 = \mathbf{X}_2$ .*

**Proof.** If both  $\mathbf{X}_1 \prec \mathbf{X}_2$  and  $\mathbf{X}_2 \prec \mathbf{X}_1$  hold, then

$$\|E^{\mathbf{S}_1} \lambda\| = \|E^{\mathbf{S}_2} \lambda\|, \quad \text{for all } \lambda \in \mathbf{H} \quad (7.7.24)$$

$$\|E^{\bar{\mathbf{S}}_2} \lambda\| = \|E^{\bar{\mathbf{S}}_1} \lambda\|, \quad \text{for all } \lambda \in \mathbf{H} \quad (7.7.25)$$

by Lemma 7.7.9. Now, suppose for example, that  $\mathbf{X}_1$  is internal, i.e.,  $\mathbf{X}_1 \subset \mathbf{H}$ . Then, for any  $\lambda \in \mathbf{S}_1$ , (7.7.24) yields  $\|\lambda\| = \|E^{\mathbf{S}_2} \lambda\|$ , which implies that  $\lambda \in \mathbf{S}_2$ . Hence  $\mathbf{S}_1 \subset \mathbf{S}_2$ . In the same way, we show that  $\bar{\mathbf{S}}_1 \subset \bar{\mathbf{S}}_2$ , using (7.7.25). Then, by Theorem 7.3.6,

$$\mathbf{X}_1 = \mathbf{S}_1 \cap \bar{\mathbf{S}}_1 \subset \mathbf{S}_2 \cap \bar{\mathbf{S}}_2 = \mathbf{X}_2.$$

But,  $\mathbf{X}_2$  is minimal, and hence  $\mathbf{X}_1 = \mathbf{X}_2$ , as claimed.  $\square$

Sofar all results in this section have been formulated for general splitting subspaces, since ordering does not require the Markovian property. In fact, Lemma 7.7.9 does not require that each splitting subspace has a unique scattering pair. To avoid this ambiguity, the next theorem will be stated for Markovian splitting subspaces, although, strictly speaking the results would hold in general.

**Theorem 7.7.11.** *Let  $\mathbf{X}_1 \sim (\mathbf{S}_1, \bar{\mathbf{S}}_1)$  and  $\mathbf{X}_2 \sim (\mathbf{S}_2, \bar{\mathbf{S}}_2)$  be minimal Markovian splitting subspaces. Then:*

- (i) *If  $\mathbf{X}_1, \mathbf{X}_2 \in \mathcal{X}_0$ , then  $\mathbf{X}_1 \prec \mathbf{X}_2 \Leftrightarrow \mathbf{S}_1 \subset \mathbf{S}_2 \Leftrightarrow \bar{\mathbf{S}}_2 \subset \bar{\mathbf{S}}_1$ .*
- (ii) *If  $\mathbf{X}_1 \in \mathcal{X}_0$ , then  $\mathbf{X}_1 \prec \mathbf{X}_2 \Leftrightarrow \mathbf{S}_1 \subset \mathbf{S}_2 \Leftrightarrow E^{\mathbf{H}} \bar{\mathbf{S}}_2 \subset \bar{\mathbf{S}}_1$ .*
- (iii) *If  $\mathbf{X}_2 \in \mathcal{X}_0$ , then  $\mathbf{X}_1 \prec \mathbf{X}_2 \Leftrightarrow E^{\mathbf{H}} \mathbf{S}_1 \subset \mathbf{S}_2 \Leftrightarrow \bar{\mathbf{S}}_2 \subset \bar{\mathbf{S}}_1$ .*

**Proof.** First, prove that

$$\text{if } \mathbf{X}_1 \in \mathcal{X}_0, \text{ then } \mathbf{X}_1 \prec \mathbf{X}_2 \Leftrightarrow \mathbf{S}_1 \subset \mathbf{S}_2 \tag{7.7.26}$$

using (7.7.19). It is trivial that  $\mathbf{S}_1 \subset \mathbf{S}_2$  implies  $\mathbf{X}_1 \prec \mathbf{X}_2$ , and to prove the converse, we take  $\lambda \in \mathbf{S}_1 \subset \mathbf{H}$  in (7.7.19), thereby obtaining  $\|\lambda\| \leq \|E^{\mathbf{S}_2} \lambda\|$  which implies that  $\lambda \in \mathbf{S}_2$ , and therefore  $\mathbf{S}_1 \subset \mathbf{S}_2$ . Obviously, by symmetry and (7.7.20), (7.7.26) has a backward version, namely

$$\text{if } \mathbf{X}_2 \in \mathcal{X}_0, \text{ then } \mathbf{X}_1 \prec \mathbf{X}_2 \Leftrightarrow \bar{\mathbf{S}}_2 \subset \bar{\mathbf{S}}_1. \tag{7.7.27}$$

Next, we prove that

$$\text{if } \mathbf{X}_2 \in \mathcal{X}_0, \text{ then } \mathbf{X}_1 \prec \mathbf{X}_2 \Leftrightarrow E^{\mathbf{H}} \mathbf{S}_1 \subset \mathbf{S}_2. \tag{7.7.28}$$

To see this, use (7.7.27), noting that  $\bar{\mathbf{S}}_2 \subset \bar{\mathbf{S}}_1$  if and only if  $\bar{\mathbf{S}}_1^\perp \subset \bar{\mathbf{S}}_2^\perp \oplus (\mathbf{H}_1 \ominus \mathbf{H})$ , where  $\mathbf{H}_1$  is the ambient space of  $\mathbf{X}_1$ . By the constructibility condition (7.4.17), this is equivalent to

$$\mathbf{S}_1 \subset \mathbf{S}_2 \oplus (\mathbf{H}_1 \ominus \mathbf{H}) \tag{7.7.29}$$

from which follows that

$$E^{\mathbf{H}} \mathbf{S}_1 \subset \mathbf{S}_2. \tag{7.7.30}$$

Conversely, if (7.7.30) holds,

$$\mathbf{S}_1 \subset E^{\mathbf{H}} \mathbf{S}_1 \oplus E^{\mathbf{H}_1 \ominus \mathbf{H}} \mathbf{S}_1 \subset \mathbf{S}_2 \oplus (\mathbf{H}_1 \ominus \mathbf{H})$$

which is (7.7.29). The backward version of (7.7.28) reads

$$\text{if } \mathbf{X}_1 \in \mathcal{X}_0, \text{ then } \mathbf{X}_1 \prec \mathbf{X}_2 \Leftrightarrow E^{\mathbf{H}} \bar{\mathbf{S}}_2 \subset \bar{\mathbf{S}}_1.$$

Now, the last statement together with (7.7.26), (7.7.27) and (7.7.28) cover all the cases of the theorem.  $\square$

## 7.8 Bibliographic notes

The concept of perpendicular intersection was introduced in [80]. The treatment in Section 7.2 follows that in [80] and [85]. Theorems 7.2.4 and 7.2.6 are Theorems 2.2 and 2.1, respectively, in [85].

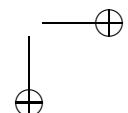
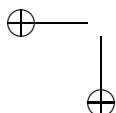
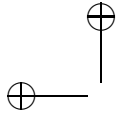
Geometric construction of the forward and backward predictor spaces was introduced simultaneously by Akaike [1] and Picci [102]. The early results toward a more complete geometric theory of stochastic realization were obtained independently by Ruckebusch [109, 107, 108] and Lindquist and Picci [75, 77, 76] and led to a joint paper [87].

Section 7.3 is essentially based on the material in [79, 80, 85]. Observability and constructibility was introduced in the context of Markovian representations in [108], where also Theorem 7.3.5 was proved. Lemma 7.3.3 appears in [87] as Lemma 1. Theorem 7.3.6 and Proposition 7.3.7 can be found in [75, 77]. Theorem 7.4.1 is Theorem 4.1 in [86]. Theorem 7.4.3 is a generalization [86, Theorem 4.2] of a result in [80], there formulated for internal splitting subspaces. Corollary 7.4.8 and Theorem 7.4.9 can be found in [79]. Together, these results imply Corollary 7.4.10, which is due to Ruckebusch [108].

The concept of *frame space* was introduced in [75], and the decomposition (7.4.23) appeared in [80]. Theorem 7.6.1 appeared in [85]; the present proof is based on a technique used in [60]. Lemma 7.6.2 can be found in [111]. Theorem 7.6.1 appeared in [85]. Lemma 7.6.2 can be found in [111] and Theorem 7.6.4 in [79].

Section 7.6 essentially follows [85]. The proof of Theorem 7.6.1 is based on the ideas of [60, Section 10.6].

The material in Section 7.7 appeared in [86]. This partial ordering, which turns out to be the natural one, is much "finer" than that proposed in [111]. The idea of uniform choice of bases was first proposed in [16].





## Chapter 8 Markovian Representations

As we have seen, any  $m$ -dimensional stationary vector process  $\{y(t)\}_{t \in \mathbb{Z}}$  generates a Hilbert space  $\mathbf{H} := \mathbf{H}(y)$  with subspaces  $\mathbf{H}^- := \mathbf{H}^-(y)$  and  $\mathbf{H}^+ := \mathbf{H}^+(y)$ , the *past space* and *future space* of  $y$  respectively, such that

$$\mathbf{H}^- \vee \mathbf{H}^+ = \mathbf{H}.$$

A *Markovian representation* of  $y$  is a triplet  $(\mathbb{H}, U, \mathbf{X})$  consisting of a Markovian splitting subspace  $\mathbf{X}$  in a Hilbert space  $\mathbb{H}$  of random variables with a unitary shift  $U : \mathbb{H} \rightarrow \mathbb{H}$  and having the properties:

- (i)  $\mathbf{H} \subset \mathbb{H}$  is a doubly invariant subspace, and the restricted shift  $U|_{\mathbf{H}}$  is the natural shift on  $\mathbf{H}$ , i.e.,

$$Uy_k(t) = y_k(t + 1) \quad \text{for } k = 1, 2, \dots, m \text{ and } t \in \mathbb{Z} \quad (8.0.1)$$

- (ii)  $\mathbb{H}$  is the *ambient space* of  $\mathbf{X}$  in the sense that

$$\mathbb{H} = \mathbf{H} \vee \overline{\text{span}\{U^t \mathbf{X} \mid t \in \mathbb{Z}\}}$$

and has finite multiplicity under the shift  $U$ .

A Markovian representation is said to be *internal* if  $\mathbb{H} = \mathbf{H}$  and *observable*, *constructible* or *minimal* if the splitting subspace  $\mathbf{X}$  is.

As explained in Chapter 6, this concept of Markovian representation is motivated by the study of linear stochastic systems

$$\begin{cases} x(t + 1) = Ax(t) + Bw(t) \\ y(t) = Cx(t) + Dw(t) \end{cases} \quad (8.0.2)$$

driven by white noise  $\{w(t)\}_{t \in \mathbb{Z}}$  and having the process  $\{y(t)\}_{t \in \mathbb{Z}}$  as its output. In this context,  $\mathbb{H}$  is the Hilbert space spanned by the white noise and possible purely

deterministic components in the state process  $\{x(t)\}_{t \in \mathbb{Z}}$ ,  $U$  is the natural shift on the processes in the model, and  $\mathbf{X}$  is the subspace

$$\mathbf{X} = \{a'x(0) \mid a \in \mathbb{R}^n\}$$

in  $\mathbb{H}$  generated by the components  $x_1(0), x_2(0), \dots, x_n(0)$  of the state  $x(0)$  at time  $t = 0$ .

In most applications we want to study finite-dimensional Markovian representations, i.e., Markovian representations  $(\mathbb{H}, U, \mathbf{X})$  for which  $\dim \mathbf{X} < \infty$ . Nevertheless, the geometric theory accommodates infinite-dimensional Markovian representations as well, but in this case models such as (8.0.2) must be interpreted in some weak sense. Therefore, we shall allow for infinite-dimensional  $\mathbf{X}$  only as long as no further technical difficulties are introduced, as the study of finite-dimensional systems is the main topic of this book.

## 8.1 The fundamental representation theorems

We collect the main results of Chapter 7 concerning Markovian splitting subspaces in a theorem formulated in the context of Markovian representations.

**Theorem 8.1.1.** *Given an  $m$ -dimensional stationary vector process  $\{y(t)\}_{t \in \mathbb{Z}}$ , let  $\mathbb{H} \supset \mathbf{H} := \mathbf{H}(y)$  be a Hilbert space of random variables with a shift  $U$  satisfying (8.0.1), and let  $\mathbf{X}$  be a subspace of  $\mathbb{H}$  such that*

$$\mathbb{H} = \mathbf{H} \vee \overline{\text{span}} \{U^t \mathbf{X} \mid t \in \mathbb{Z}\} \quad (8.1.1)$$

Then  $(\mathbb{H}, U, \mathbf{X})$  is a Markovian representation of  $y$  if and only if

$$\mathbf{X} = \mathbf{S} \cap \bar{\mathbf{S}} \quad (8.1.2)$$

for some pair  $(\mathbf{S}, \bar{\mathbf{S}})$  of subspaces of  $\mathbb{H}$  such that

- (i)  $\mathbf{H}^- \subset \mathbf{S}$  and  $\mathbf{H}^+ \subset \bar{\mathbf{S}}$ ,
- (ii)  $U^* \mathbf{S} \subset \mathbf{S}$  and  $U \bar{\mathbf{S}} \subset \bar{\mathbf{S}}$ , and
- (iii)  $\mathbb{H} = \bar{\mathbf{S}}^\perp \oplus (\mathbf{S} \cap \bar{\mathbf{S}}) \oplus \mathbf{S}^\perp$ ,

where  $\perp$  denotes the orthogonal complement in  $\mathbb{H}$ . Moreover, the correspondence  $\mathbf{X} \leftrightarrow (\mathbf{S}, \bar{\mathbf{S}})$  is one-one. In fact,

$$\mathbf{S} = \mathbf{H}^- \vee \mathbf{X}^- \quad \text{and} \quad \bar{\mathbf{S}} = \mathbf{H}^+ \vee \mathbf{X}^+. \quad (8.1.3)$$

Finally,  $(\mathbb{H}, U, \mathbf{X})$  is observable if and only if

$$\bar{\mathbf{S}} = \mathbf{H}^- \vee \mathbf{S}^\perp, \quad (8.1.4)$$

and constructible if and only if

$$\mathbf{S} = \mathbf{H}^+ \vee \bar{\mathbf{S}}^\perp, \quad (8.1.5)$$

and minimal if and only if both (8.1.4) and (8.1.5) hold.

**Proof.** The theorem follows immediately from Theorems 7.3.6 and 7.4.1, recalling that perpendicular intersection can be characterized by property (iii) (Theorem 7.2.4), and from Corollary 7.4.8 and Theorem 7.4.9.  $\square$

To each Markovian representation  $(\mathbb{H}, U, \mathbf{X})$  we want to associate two dynamical representations, one evolving forward in time and one backwards. The abstract idea behind this construction can be illustrated by two commutative diagrams. Recall that, since  $\mathbf{S} \perp \bar{\mathbf{S}} \mid \mathbf{X}$  for any Markovian splitting subspace  $\mathbf{X} \sim (\mathbf{S}, \bar{\mathbf{S}})$ , by Lemma 2.4.1, we also have

$$\mathbf{S} \perp \mathbf{H}^+ \mid \mathbf{X} \quad \text{and} \quad \bar{\mathbf{S}} \perp \mathbf{H}^- \mid \mathbf{X} \tag{8.1.6}$$

As can be seen from Proposition 2.4.2 (vi), the first of these is equivalent to the factorization

$$\begin{array}{ccc} \mathbf{S} & \xrightarrow{E^{\mathbf{H}^+} \mid \mathbf{S}} & \mathbf{H}^+ \\ E^{\mathbf{X}} \mid \mathbf{S} \searrow & & \nearrow \mathcal{O} \\ & \mathbf{X} & \end{array} \tag{8.1.7}$$

where  $\mathcal{O} := E^{\mathbf{H}^+} \mid \mathbf{X}$  is the observability operator of  $\mathbf{X}$  and  $E^{\mathbf{X}} \mid \mathbf{S}$  is an insertion operator which is always surjective since  $\mathbf{X} \subset \mathbf{S}$ . Note that

$$U^* \mathbf{S} \subset \mathbf{S} \tag{8.1.8}$$

so that  $\mathbf{S}$  can act as a past space in our construction and so that  $E^{\mathbf{H}^+} \mid \mathbf{S}$  is a Hankel operator. The invariance property (8.1.8) allows us also to form the space

$$\mathbf{W} = U\mathbf{S} \ominus \mathbf{S}, \tag{8.1.9}$$

representing the new information carried by the “next input”. In that context, a model of type (8.0.2) would entail representations of  $U\mathbf{X}$  and

$$\mathbf{Y} := \{b'y(0) \mid b \in \mathbb{R}^m\} \tag{8.1.10}$$

in terms of  $\mathbf{X}$  and  $\mathbf{W}$ .

Before tending to this matter, let us also consider the backward setting. The second of statements (8.1.6) is equivalent to the factorization

$$\begin{array}{ccc} \bar{\mathbf{S}} & \xrightarrow{E^{\mathbf{H}^-} \mid \bar{\mathbf{S}}} & \mathbf{H}^- \\ E^{\mathbf{X}} \mid \bar{\mathbf{S}} \searrow & & \nearrow \mathcal{C} \\ & \mathbf{X} & \end{array} \tag{8.1.11}$$

where  $\mathcal{C} := E^{\mathbf{H}^-} \mid \mathbf{X}$  is the constructibility operator of  $\mathbf{X}$ , and  $E^{\mathbf{X}} \mid \bar{\mathbf{S}}$  is an insertion map. Since, by Theorem 7.4.1,

$$U\bar{\mathbf{S}} \subset \bar{\mathbf{S}}, \tag{8.1.12}$$

$\bar{\mathbf{S}}$  can serve as a future space, and we can form the space

$$\bar{\mathbf{W}} := \bar{\mathbf{S}} \ominus U\bar{\mathbf{S}}. \quad (8.1.13)$$

Moreover,  $E^{\mathbf{H}^-} |_{\bar{\mathbf{S}}}$  is a Hankel operator which maps the future space  $\bar{\mathbf{S}}$  backwards into the past  $\mathbf{H}^-$ . The construction of a backward model would therefore involve a representation of  $U^*\mathbf{X}$  and  $U^*\mathbf{Y}$  in terms of  $\mathbf{X}$  and  $U^*\bar{\mathbf{W}}$ .

**Theorem 8.1.2.** *Let  $\mathbf{X} \sim (\mathbf{S}, \bar{\mathbf{S}})$  be a Markovian splitting subspace, and let  $\mathbf{W}, \bar{\mathbf{W}}$  and  $\mathbf{Y}$  be defined by (8.1.9), (8.1.13) and (8.1.10) respectively. Then*

$$\begin{cases} U\mathbf{X} \subset \mathbf{X} \oplus \mathbf{W} \\ \mathbf{Y} \subset \mathbf{X} \oplus \mathbf{W} \end{cases} \quad (8.1.14)$$

and

$$\begin{cases} U^*\mathbf{X} \subset \mathbf{X} \oplus (U^*\bar{\mathbf{W}}) \\ U^*\mathbf{Y} \subset \mathbf{X} \oplus (U^*\bar{\mathbf{W}}) \end{cases} \quad (8.1.15)$$

**Proof.** To prove (8.1.14) first note that, since  $\mathbf{X} = \mathbf{S} \cap \bar{\mathbf{S}}$  and  $U\bar{\mathbf{S}} \subset \bar{\mathbf{S}}$ ,

$$U\mathbf{X} \subset (U\mathbf{S}) \cap \bar{\mathbf{S}}. \quad (8.1.16)$$

Moreover,

$$\mathbf{Y} \subset (U\mathbf{H}^-) \cap \mathbf{H}^+ \subset (U\mathbf{S}) \cap \bar{\mathbf{S}}. \quad (8.1.17)$$

Therefore, (8.1.14) would follow if we could show that

$$(U\mathbf{S}) \cap \bar{\mathbf{S}} = \mathbf{X} \oplus \mathbf{W}. \quad (8.1.18)$$

To this end, note that, since  $U\mathbf{S} = \mathbf{S} \oplus \mathbf{W}$ , we have  $\mathbf{S} \oplus \mathbf{W} \oplus (U\mathbf{S})^\perp = \mathbb{H}$ , and hence

$$\mathbf{S}^\perp = (U\mathbf{S})^\perp \oplus \mathbf{W}, \quad (8.1.19)$$

which inserted into  $\mathbb{H} = \bar{\mathbf{S}}^\perp \oplus \mathbf{X} \oplus \mathbf{S}^\perp$  yields

$$\mathbf{X} \oplus \mathbf{W} = [(U\mathbf{S})^\perp \oplus \bar{\mathbf{S}}^\perp]^\perp = (U\mathbf{S}) \cap \bar{\mathbf{S}}$$

as claimed.

To prove (8.1.15), we note that, since  $U^*\mathbf{S} \subset \mathbf{S}$ ,

$$U^*\mathbf{X} \subset \mathbf{S} \cap (U^*\bar{\mathbf{S}}). \quad (8.1.20)$$

Moreover,

$$U^*\mathbf{Y} \subset \mathbf{H}^- \cap (U^*\mathbf{H}^+) \subset \mathbf{S} \cap (U^*\bar{\mathbf{S}}). \quad (8.1.21)$$

Hence it only remains to prove that

$$\mathbf{S} \cap (U^*\bar{\mathbf{S}}) = \mathbf{X} \oplus U^*\bar{\mathbf{W}}. \quad (8.1.22)$$



However, this follows in the same way as above by first showing that

$$\bar{\mathbf{S}}^\perp = (U^*\bar{\mathbf{S}})^\perp \oplus U^*\bar{\mathbf{W}}, \quad (8.1.23)$$

and then inserting this into  $\mathbb{H} = \bar{\mathbf{S}}^\perp \oplus \mathbf{X} \oplus \mathbf{S}^\perp$ .  $\square$

The subspaces  $\mathbf{W}$  and  $\bar{\mathbf{W}}$ , defined by (8.1.9) and (8.1.13), satisfy the orthogonality relations

$$U^j\mathbf{W} \perp U^k\mathbf{W} \quad \text{and} \quad U^j\bar{\mathbf{W}} \perp U^k\bar{\mathbf{W}} \quad \text{for } j \neq k. \quad (8.1.24)$$

Such subspaces are called *wandering subspaces*. In fact, the obvious decompositions

$$\mathbf{S} = U^{-1}\mathbf{W} \oplus U^{-2}\mathbf{W} \oplus U^{-3}\mathbf{W} \oplus \dots \oplus U^{-N}\mathbf{W} \oplus U^{-N}\mathbf{S} \quad (8.1.25)$$

and

$$\bar{\mathbf{S}} = \bar{\mathbf{W}} \oplus U\bar{\mathbf{W}} \oplus U^2\bar{\mathbf{W}} \oplus \dots \oplus U^{N-1}\bar{\mathbf{W}} \oplus U^N\bar{\mathbf{S}} \quad (8.1.26)$$

are the *Wold decompositions* described in Chapter 2, applied successively to shifted subspaces, and  $U^{-N}\mathbf{S}$  and  $U^N\bar{\mathbf{S}}$  tend to doubly invariant subspaces  $\mathbf{S}_{-\infty}$  and  $\bar{\mathbf{S}}_\infty$  respectively as  $N \rightarrow \infty$  (Theorem 4.5.8).

From Theorem 4.5.4 we know that  $\mathbf{W}$  and  $\bar{\mathbf{W}}$  are finite-dimensional with dimensions which are bounded by the multiplicity of  $(\mathbb{H}, U, \mathbf{X})$ . Therefore, by choosing orthonormal bases  $\{\eta_1, \eta_2, \dots, \eta_p\}$  and  $\{\bar{\eta}_1, \bar{\eta}_2, \dots, \bar{\eta}_{\bar{p}}\}$  respectively, we see from (8.1.24) that

$$w(t) = \begin{bmatrix} U^t\eta_1 \\ U^t\eta_2 \\ \vdots \\ U^t\eta_p \end{bmatrix} \quad \text{and} \quad \bar{w}(t) = \begin{bmatrix} U^t\bar{\eta}_1 \\ U^t\bar{\eta}_2 \\ \vdots \\ U^t\bar{\eta}_{\bar{p}} \end{bmatrix} \quad (8.1.27)$$

are normalized white noise processes, one corresponding to  $\mathbf{S}$  and one corresponding to  $\bar{\mathbf{S}}$ .

**Theorem 8.1.3.** *Let  $(\mathbb{H}, U, \mathbf{X})$  be a Markovian representation of multiplicity  $\mu$ . Then the wandering subspaces  $\mathbf{W}$  and  $\bar{\mathbf{W}}$  have finite dimensions such that  $p := \dim \mathbf{W} \leq \mu$  and  $\bar{p} := \dim \bar{\mathbf{W}} \leq \mu$ . Moreover, if  $\mathbf{X} \sim (\mathbf{S}, \bar{\mathbf{S}})$ ,*

$$\mathbf{S} = \mathbf{H}^-(w) \oplus \mathbf{S}_{-\infty}, \quad (8.1.28)$$

where  $\{w(t)\}_{t \in \mathbb{Z}}$  is a  $p$ -dimensional normalized white noise process, i.e.,

$$\mathbf{E}\{w(t)\} = 0, \quad \mathbf{E}\{w(s)w(t)'\} = I_p\delta_{st}, \quad (8.1.29)$$

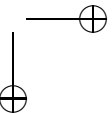
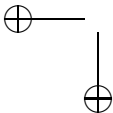
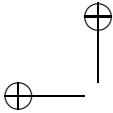
such that

$$\mathbf{W} := \{a'w(0) \mid a \in \mathbb{R}^p\}, \quad (8.1.30)$$

and  $\mathbf{S}_{-\infty}$  is a doubly invariant subspace, i.e., invariant under both  $U$  and  $U^*$ .

Similarly,

$$\bar{\mathbf{S}} = \mathbf{H}^+(\bar{w}) \oplus \bar{\mathbf{S}}_\infty, \quad (8.1.31)$$



where  $\{\bar{w}(t)\}_{t \in \mathbb{Z}}$  is a  $\bar{p}$ -dimensional normalized white noise process such that

$$\bar{\mathbf{W}} := \{a' \bar{w}(0) \mid a \in \mathbb{R}^{\bar{p}}\}, \quad (8.1.32)$$

and  $\bar{\mathbf{S}}_\infty$  is a doubly invariant.

Finally,

$$\mathbb{H} = \mathbf{H}(w) \oplus \mathbf{S}_{-\infty} = \mathbf{H}(\bar{w}) \oplus \bar{\mathbf{S}}_\infty, \quad (8.1.33)$$

i.e., in particular,

$$\mathbf{S}^\perp = \mathbf{H}^+(w) \quad \text{and} \quad \bar{\mathbf{S}}^\perp = \mathbf{H}^-(\bar{w}). \quad (8.1.34)$$

**Proof.** In view of Theorem 4.5.8 and Theorem 4.5.4, it only remains to prove (8.1.34). However, this follows from (8.1.33) and the fact that  $\mathbf{H}^-(w) \perp \mathbf{H}^+(w)$ , the latter of which is a consequence of  $w$  being a white noise.  $\square$

The processes  $w$  and  $\bar{w}$  are called the *forward* respectively the *backward generating processes* of  $(\mathbb{H}, U, \mathbf{X})$ , and clearly they are unique modulo linear coordinate transformations in  $\mathbf{W}$  and  $\bar{\mathbf{W}}$  respectively. The subspaces  $\mathbf{S}_{-\infty}$  and  $\bar{\mathbf{S}}_\infty$  are called the *remote past* and the *remote future spaces*, respectively.

**Definition 8.1.4.** The Markovian representation  $(\mathbb{H}, U, \mathbf{X})$  is normal if  $\mathbf{S}_{-\infty} = \bar{\mathbf{S}}_\infty$  and proper if  $\mathbf{S}_{-\infty} = \bar{\mathbf{S}}_\infty = 0$ .

In view of (8.1.33),  $\bar{p} = p$  if  $(\mathbb{H}, U, \mathbf{X})$  is normal. As we shall see in the next section, all finite-dimensional Markovian representations are normal. However, in the infinite-dimensional case one may even have minimal Markovian representations that are not normal, as the following example shows.

**Example 8.1.5.** Let  $\{y(t)\}_{t \in \mathbb{Z}}$  be a p.n.d. process with spectral density

$$\Phi(e^{i\theta}) = \sqrt{1 + \cos \theta}. \quad (8.1.35)$$

Then  $\mathbf{X}_- = \mathbf{H}^- \sim (\mathbf{H}^-, \mathbf{H})$ , and consequently  $\mathbf{S}_{-\infty} = 0$  and  $\bar{\mathbf{S}}_\infty = \mathbf{H}$ . Therefore the minimal Markovian representation  $(\mathbf{H}, U, \mathbf{X}_-)$  is not normal. (This is a discrete-time version of an example given in [27, p.99]; also see [24, p.43].)

From this example we see that we may indeed have  $\bar{p} \neq p$ .

**Proposition 8.1.6.** The Markovian representation  $(\mathbb{H}, U, \mathbf{X})$  with  $\mathbf{X} \sim (\mathbf{S}, \bar{\mathbf{S}})$  is proper if and only if  $\mathbf{S}^\perp$  and  $\bar{\mathbf{S}}^\perp$  are full range.

**Proof.** Since  $\mathbf{S}_{-\infty} = \bigcap_{t=-\infty}^0 U^t \mathbf{S}$  and  $\bar{\mathbf{S}}_\infty = \bigcap_0^\infty U^t \bar{\mathbf{S}}$ ,  $\mathbf{X} \sim (\mathbf{S}, \bar{\mathbf{S}})$  is proper if and only if

$$\bigvee_{t=-\infty}^0 U^t \mathbf{S}^\perp = \mathbb{H} = \bigvee_0^\infty U^t \bar{\mathbf{S}}^\perp,$$

as claimed.  $\square$

## 8.2 Normality, properness and the Markov semigroup

Given a Markovian representation  $(\mathbb{H}, U, \mathbf{X})$ , as in Section 7.5 we define the restricted shift

$$U(\mathbf{X}) = \mathbf{E}^{\mathbf{X}} U|_{\mathbf{X}} \tag{8.2.1}$$

on the Markovian splitting subspace  $\mathbf{X} \sim (\mathbf{S}, \bar{\mathbf{S}})$  and the Markov semigroup

$$U_t(\mathbf{X}) := \mathbf{E}^{\mathbf{X}} U^t|_{\mathbf{X}} = U(\mathbf{X})^t, \quad t = 0, 1, 2, \dots; \tag{8.2.2}$$

see Theorem 7.5.1.

**Theorem 8.2.1.** *The semigroup  $U_t(\mathbf{X})$  tends strongly to zero as  $t \rightarrow \infty$  if and only if*

$$\mathbf{S}_{-\infty} := \bigcap_{t=-\infty}^0 U^t \mathbf{S} = 0, \tag{8.2.3}$$

and  $U_t(\mathbf{X})^*$  tend strongly to zero as  $t \rightarrow \infty$  if and only if

$$\bar{\mathbf{S}}_{\infty} := \bigcap_{t=0}^{\infty} U^t \bar{\mathbf{S}} = 0. \tag{8.2.4}$$

**Proof.** Let  $\xi \in \mathbf{X}$ . By Lemma 2.2.9, we have  $\mathbf{E}^{\mathbf{S}} U^t = U^t \mathbf{E}^{U^{-t}\mathbf{S}}$ . Therefore, setting

$$\xi_t := U_t(\mathbf{X})\xi, \quad t = 0, 1, 2, \dots, \tag{8.2.5}$$

(7.5.4) and (7.5.5) imply that

$$U^{-t}\xi_t = U^{-t} \mathbf{E}^{\mathbf{S}} U^t \xi = \mathbf{E}^{U^{-t}\mathbf{S}} \xi = \mathbf{E}^{U^{-t}\mathbf{X}} \xi. \tag{8.2.6}$$

Then, if  $\mathbf{S}_{-\infty} = 0$ ,

$$\|\xi_t\| = \|\mathbf{E}^{U^{-t}\mathbf{S}} \xi\| \rightarrow 0 \quad \text{as } t \rightarrow \infty.$$

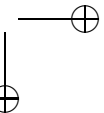
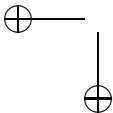
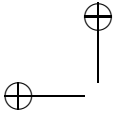
In fact, since  $U^{-t}\mathbf{S} \subset U^{-s}\mathbf{S}$  for  $t > s$ , the sequence  $(\|\xi_t\|)$  is monotonely nonincreasing and bounded from below by zero, and therefore  $(\|\xi_t\|)$  tends to a limit. Consequently,

$$\|\xi_t - \xi_s\| \leq \|\xi_t\| - \|\xi_s\| \rightarrow 0 \quad \text{as } s, t \rightarrow \infty,$$

so  $\xi_t$  tend to a limit  $\xi_{\infty}$ . Clearly,  $\xi_{\infty} \in U^{-t}\mathbf{S}$  for  $t = 0, 1, 2, \dots$  and hence  $\xi_{\infty} \in \mathbf{S}_{-\infty}$ . Therefore,  $\xi_k \rightarrow 0$  as  $t \rightarrow \infty$  if  $\mathbf{S}_{-\infty} = 0$ .

Conversely, suppose that  $\xi_t \rightarrow 0$  as  $t \rightarrow \infty$ . We want to show that  $\mathbf{S}_{-\infty} = 0$ . In view of (7.5.4),

$$\begin{aligned} \xi - U^{-t}\xi_t &= \xi - U^{-t}U(\mathbf{X})^t \xi \\ &= \sum_{k=0}^{t-1} \left[ U^{-k}U(\mathbf{X})^k - U^{-(k+1)}U(\mathbf{X})^{(k+1)} \right] \xi \\ &= \sum_{k=0}^{t-1} U^{-(k+1)} [U - U(\mathbf{X})] \xi_k. \end{aligned}$$



However,  $\xi_k \in \mathbf{X} \subset \mathbf{S}$ , and therefore in view of (8.2.6)

$$[U - U(\mathbf{X})]\xi_k = U\xi - E^{\mathbf{S}}U\xi_k \in U\mathbf{S} \ominus \mathbf{S} = \mathbf{W}, \quad (8.2.7)$$

where we have also used (8.1.9). Consequently,  $\xi - U^{-t}\xi_t \in \mathbf{H}^-(w)$  for  $t = 0, 1, 2, \dots$  and all  $\xi \in \mathbf{X}$ ; i.e.,  $\mathbf{X} \subset \mathbf{H}^-(w)$ . Therefore,

$$\bar{\mathbf{S}} = \mathbf{X} \oplus \mathbf{S}^\perp = \mathbf{X} \oplus \mathbf{H}^+(w) \subset \mathbf{H}(w).$$

Hence,

$$\mathbb{H} = \bigvee_{t=0}^\infty U^{-t}\bar{\mathbf{S}} \subset \mathbf{H}(w),$$

and consequently, by (8.1.33),  $\mathbf{S}_{-\infty} = 0$ , as claimed.

A symmetric argument yields the other half of the theorem.  $\square$

Consequently,  $U_t(\mathbf{X})$  and  $U_t(\mathbf{X})^*$  both tend strongly to zero as  $t \rightarrow \infty$  if and only if  $(\mathbb{H}, U, \mathbf{X})$  is proper.

An inspection of the proof of Theorem 8.2.1 reveals that  $\mathbf{X}$  need not be a splitting subspace of a particular fixed process  $y$ . We reformulate the theorem for later reference.

**Corollary 8.2.2.** *The statement of Theorem 8.2.1 remains true with  $\mathbf{X} = \mathbf{S} \cap \bar{\mathbf{S}}$ , where  $\mathbf{S}$  and  $\bar{\mathbf{S}}$  intersect perpendicularly and  $U^*\mathbf{S} \subset \mathbf{S}$  and  $U\bar{\mathbf{S}} \subset \bar{\mathbf{S}}$ .*

Indeed, if  $\mathbf{S}$  and  $\bar{\mathbf{S}}$  intersect perpendicularly, then

$$\mathbf{S} \vee \bar{\mathbf{S}} =: \mathbb{H} = \bar{\mathbf{S}}^\perp \oplus \mathbf{X} \oplus \mathbf{S}^\perp \quad (8.2.8)$$

(Theorem 7.2.4), which together with the invariance properties of  $\mathbf{S}$  and  $\bar{\mathbf{S}}$  is all that we need in the proof of Theorem 8.2.1.

**Theorem 8.2.3.** *The Markovian splitting subspace  $\mathbf{X}$  admits a unique orthogonal decomposition*

$$\mathbf{X} = \mathbf{X}_0 \oplus \mathbf{X}_\infty \quad (8.2.9)$$

such that  $U(\mathbf{X})|_{\mathbf{X}_\infty}$  is unitary and  $U(\mathbf{X})|_{\mathbf{X}_0}$  is completely nonunitary; i.e., it is not unitary on any nontrivial subspace. Moreover,

$$\mathbf{X}_\infty = \mathbf{S}_{-\infty} \cap \bar{\mathbf{S}}_\infty, \quad (8.2.10)$$

where  $\mathbf{S}_{-\infty}$  and  $\bar{\mathbf{S}}_\infty$  are defined as in Theorem 8.1.3, or, equivalently, as in Theorem 8.2.1.

**Proof.** Since  $U(\mathbf{X}) : \mathbf{X} \rightarrow \mathbf{X}$  is a contraction, the existence of a decomposition (8.2.9) with  $U(\mathbf{X})|_{\mathbf{X}_\infty}$  unitary and  $U(\mathbf{X})|_{\mathbf{X}_0}$  completely nonunitary is ensured by Theorem 3.2 in [116], where it is established that

$$\mathbf{X}_\infty = \{\xi \in \mathbf{X} \mid \|U_t(\mathbf{X})\xi\| = \|\xi\| = \|U_t(\mathbf{X})^*\xi\|, \quad t = 0, 1, 2, \dots\}.$$

Now, if  $\xi \in \mathbf{X}_\infty$ , then, in view of (8.2.6),  $U^{-t}U_t(\mathbf{X})\xi = E^{U^{-t}\mathbf{X}}\xi$ , and hence, since  $\|U_t(\mathbf{X})\xi\| = \|\xi\|$ , we have  $\|E^{U^{-t}\mathbf{X}}\xi\| = \|\xi\|$ . Therefore,  $\xi = E^{U^{-t}\mathbf{X}}\xi \in U^{-t}\mathbf{X}$  for all  $\xi \in \mathbf{X}_\infty$  and  $t = 0, 1, 2, \dots$ . A symmetric argument also shows that  $\xi \in U^t\mathbf{X}$  for all  $\xi \in \mathbf{X}_\infty$  and  $t = 0, 1, 2, \dots$ . Consequently,

$$\mathbf{X}_\infty = \bigcap_{t=-\infty}^{\infty} U^t\mathbf{X} \subset \mathbf{S}_{-\infty} \cap \bar{\mathbf{S}}_\infty. \quad (8.2.11)$$

To show that  $\mathbf{S}_{-\infty} \cap \bar{\mathbf{S}}_\infty \subset \mathbf{X}_\infty$ , and hence that (8.2.10) holds, we first note that  $\mathbf{H}(w) = \mathbf{S}_{-\infty}^\perp$  and  $\mathbf{H}(\bar{w}) = \bar{\mathbf{S}}_\infty^\perp$  in view of (8.1.33), and therefore

$$[\mathbf{H}(w) \vee \mathbf{H}(\bar{w})]^\perp = \mathbf{S}_{-\infty} \cap \bar{\mathbf{S}}_\infty \subset \mathbf{S} \cap \bar{\mathbf{S}} \subset \mathbf{X}. \quad (8.2.12)$$

Now, suppose  $\xi \in \mathbf{S}_{-\infty} \cap \bar{\mathbf{S}}_\infty$ . Then (8.2.5) and (8.2.7) imply that

$$U^{-k}U(\mathbf{X})^k\xi - U^{-(k+1)}U(\mathbf{X})^{k+1}\xi \in U^{-k}\mathbf{W}, \quad k = 0, 1, 2, \dots$$

and consequently, since  $\xi \perp \mathbf{H}(w)$ ,

$$\langle U^k\xi, U(\mathbf{X})^k\xi \rangle = \langle U^{k+1}\xi, U(\mathbf{X})^{k+1}\xi \rangle, \quad k = 0, 1, 2, \dots,$$

from which we have

$$\|U_k(\mathbf{X})\xi\| = \|U_{k+1}(\mathbf{X})\xi\|, \quad k = 0, 1, 2, \dots$$

Likewise, we can see that  $\xi \perp \mathbf{H}(\bar{w})$  implies that

$$\|U_k(\mathbf{X})^*\xi\| = \|U_{k+1}(\mathbf{X})^*\xi\|, \quad k = 0, 1, 2, \dots$$

Consequently,  $\mathbf{S}_{-\infty} \cap \bar{\mathbf{S}}_\infty \subset \mathbf{X}_\infty$ , as required.  $\square$

The following corollary is an immediate consequence of Theorem 8.2.3.

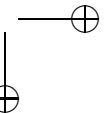
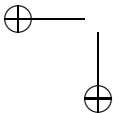
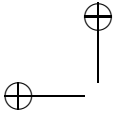
**Corollary 8.2.4.** *If  $(\mathbb{H}, U, \mathbf{X})$  is normal, then*

$$\mathbf{X}_\infty = \mathbf{S}_{-\infty} = \bar{\mathbf{S}}_\infty. \quad (8.2.13)$$

**Corollary 8.2.5.** *The statement of Theorem 8.2.3 remains true with  $\mathbf{X} = \mathbf{S} \cap \bar{\mathbf{S}}$ , where  $\mathbf{S}$  and  $\bar{\mathbf{S}}$  intersect perpendicularly and  $U^*\mathbf{S} \subset \mathbf{S}$  and  $U\bar{\mathbf{S}} \subset \bar{\mathbf{S}}$ . Moreover, if  $\mathbf{S}_0 := \mathbf{S} \ominus \mathbf{X}_\infty$  and  $\bar{\mathbf{S}}_0 := \bar{\mathbf{S}} \ominus \mathbf{X}_\infty$ , then  $\mathbf{X}_0 = \mathbf{S}_0 \cap \bar{\mathbf{S}}_0$ , where  $\mathbf{S}_0$  and  $\bar{\mathbf{S}}_0$  intersect perpendicularly and  $U^*\mathbf{S}_0 \subset \mathbf{S}_0$  and  $U\bar{\mathbf{S}}_0 \subset \bar{\mathbf{S}}_0$ .*

**Proof.** The first statement follows by the same argument as for Corollary 8.2.2. To prove the second statement, first note that

$$\mathbf{X} = \mathbf{S} \cap \bar{\mathbf{S}} = (\mathbf{S}_0 \oplus \mathbf{X}_\infty) \cap (\bar{\mathbf{S}}_0 \oplus \mathbf{X}_\infty) = (\mathbf{S}_0 \cap \bar{\mathbf{S}}_0) \oplus \mathbf{X}_\infty,$$



which implies that  $\mathbf{X}_0 = \mathbf{S}_0 \cap \bar{\mathbf{S}}_0$ . Moreover, if  $\mathbf{S}$  and  $\bar{\mathbf{S}}$  intersect perpendicularly, then  $\mathbf{S} \perp \bar{\mathbf{S}} \mid \mathbf{X}$ ; i.e.,

$$\langle \lambda - E^{\mathbf{X}_0 \oplus \mathbf{X}_\infty} \lambda, \mu - E^{\mathbf{X}_0 \oplus \mathbf{X}_\infty} \mu \rangle = 0$$

for  $\lambda \in \mathbf{S}, \mu \in \bar{\mathbf{S}}$ , and hence in particular for  $\lambda \in \mathbf{S}_0, \mu \in \bar{\mathbf{S}}_0$ . However,  $\mathbf{X}_\infty \perp \mathbf{S}_0$  and  $\mathbf{X}_\infty \perp \bar{\mathbf{S}}_0$ , and therefore

$$\langle \lambda - E^{\mathbf{X}_0} \lambda, \mu - E^{\mathbf{X}_0} \mu \rangle = 0$$

for  $\lambda \in \mathbf{S}_0, \mu \in \bar{\mathbf{S}}_0$ ; i.e.,  $\mathbf{S}_0 \perp \bar{\mathbf{S}}_0 \mid \mathbf{X}_0$ . Hence  $\mathbf{S}_0$  and  $\bar{\mathbf{S}}_0$  intersect perpendicularly. Since  $\mathbf{X}_\infty$  is doubly invariant in view of (8.2.11),  $U^*\mathbf{S} \subset \mathbf{S}$  and  $U\bar{\mathbf{S}} \subset \bar{\mathbf{S}}$  is the same as  $(U^*\mathbf{S}_0) \oplus \mathbf{X}_\infty \subset \mathbf{S}_0 \oplus \mathbf{X}_\infty$  and  $(U\bar{\mathbf{S}}_0) \oplus \mathbf{X}_\infty \subset \bar{\mathbf{S}}_0 \oplus \mathbf{X}_\infty$ , respectively, and hence it follows that  $U^*\mathbf{S}_0 \subset \mathbf{S}_0$  and  $U\bar{\mathbf{S}}_0 \subset \bar{\mathbf{S}}_0$ .  $\square$

**Corollary 8.2.6.** *A finite-dimensional Markovian representation  $(\mathbb{H}, U, \mathbf{X})$  is normal.*

**Proof.** Let  $\mathbf{X} \sim (\mathbf{S}, \bar{\mathbf{S}})$ , and let  $\mathbf{X} = \mathbf{X}_0 \oplus \mathbf{X}_\infty$  be the decomposition of Theorem 8.2.3. Then, by Corollary 8.2.5,  $\mathbf{X}_0 = \mathbf{S}_0 \cap \bar{\mathbf{S}}_0$ , where  $\mathbf{S}_0$  and  $\bar{\mathbf{S}}_0$  intersect perpendicularly and  $U^*\mathbf{S}_0 \subset \mathbf{S}_0, U\bar{\mathbf{S}}_0 \subset \bar{\mathbf{S}}_0$ , and

$$U(\mathbf{X})|_{\mathbf{X}_0} = U(\mathbf{X}_0)$$

is a completely nonunitary contraction. Hence, since  $\mathbf{X}_0$  is finite-dimensional,  $U(\mathbf{X}_0)$  (as well as its adjoint  $U(\mathbf{X}_0)^*$ ) has all its eigenvalues in the open unit disc, and therefore  $U_t(\mathbf{X}_0) := U(\mathbf{X}_0)^t$  and  $U_t(\mathbf{X}_0)^*$  tend strongly to zero, as  $t \rightarrow \infty$ . Consequently, by Corollary 8.2.2, the remote past of  $\mathbf{S}_0$  and the remote future of  $\bar{\mathbf{S}}_0$  are trivial, so  $\mathbf{S}_{-\infty} = \mathbf{S}_\infty = \mathbf{X}_\infty$ .  $\square$

This proof does not work in the case that  $\mathbf{X}$  is infinite-dimensional. However, appealing to [116] (Proposition 6.7 in Chapter II and Proposition 4.2 in Chapter III), we have the following criteria in the general case.

**Theorem 8.2.7.** *Let  $\mathbf{X} = \mathbf{X}_0 \oplus \mathbf{X}_\infty$  be the decomposition of Theorem 8.2.3. Then each of the following conditions is sufficient for  $(\mathbb{H}, U, \mathbf{X})$  to be normal:*

1. *The intersection of the spectrum of  $U(\mathbf{X})|_{\mathbf{X}_0}$  with the unit circle has Lebesgue measure zero.*
2. *There is a nontrivial  $\varphi \in H^\infty$  such that  $\varphi(U(\mathbf{X})|_{\mathbf{X}_0}) = 0$ .*

The second condition in Theorem 8.2.7 can be seen as a generalization of Cayley-Hamilton condition in the finite-dimensional case.

### 8.3 The forward and backward systems (the finite-dimensional case)

In this section we consider a Markovian representation  $(\mathbb{H}, U, \mathbf{X})$  which is finite-dimensional so that  $n := \dim \mathbf{X} < \infty$ . The general case will be considered in Section 8.7.

We want to construct a stochastic system (8.0.2) for which

$$\mathbf{X} = \{a'x(0) \mid a \in \mathbb{R}^n\} \tag{8.3.1}$$

is the Markovian splitting subspace  $\mathbf{X} \sim (\mathbf{S}, \bar{\mathbf{S}})$  of the given Markovian representation  $(\mathbb{H}, U, \mathbf{X})$ . For this we shall use Theorem 8.1.2.

Therefore, let  $\{\xi_1, \xi_2, \dots, \xi_n\}$  be a basis in  $\mathbf{X}$ , and define the vector process  $\{x(t)\}_{t \in \mathbb{Z}}$  to be

$$x(t) = \begin{bmatrix} U^t \xi_1 \\ U^t \xi_2 \\ \vdots \\ U^t \xi_n \end{bmatrix}. \tag{8.3.2}$$

This is a stationary vector process and

$$\mathbb{E}\{x(t)x(t)'\} = P := \mathbb{E} \left\{ \begin{bmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_n \end{bmatrix} [\xi_1 \ \xi_2 \ \cdots \ \xi_n] \right\} \tag{8.3.3}$$

for each  $t \in \mathbb{Z}$ . Since  $\{\xi_1, \xi_2, \dots, \xi_n\}$  is a basis and  $\|a'\xi\|^2 = a'Pa$  for all  $a \in \mathbb{R}^n$ , we must have  $P > 0$ .

Then, by (8.1.14) in Theorem 8.1.2, we have

$$\begin{cases} U\xi_i &= \sum_{j=1}^n a_{ij}\xi_j + \sum_{j=1}^p b_{ij}w_j(0), & i = 1, 2, \dots, n \\ y_i(0) &= \sum_{j=1}^n c_{ij}\xi_j + \sum_{j=1}^p d_{ij}w_j(0), & i = 1, 2, \dots, m \end{cases}$$

for some choice of coefficients  $\{a_{ij}, b_{ij}, c_{ij}, d_{ij}\}$ . After applying the shift  $U^t$ , in view of (8.3.2), this can be written

$$\Sigma \begin{cases} x(t+1) &= Ax(t) + Bw(t) \\ y(t) &= Cx(t) + Dw(t) \end{cases} \tag{8.3.4}$$

with obvious definitions of the matrices  $A, B, C$  and  $D$ . This is a *forward stochastic system* in the sense that

$$\mathbf{H}^- \vee \mathbf{X}^- \perp \mathbf{H}^+(w) \tag{8.3.5}$$

so that future generating noise is orthogonal to past output and past and present state. In fact, by Theorem 7.4.1,  $\mathbf{H}^- \vee \mathbf{X}^- = \mathbf{S}$ , which is orthogonal to  $\mathbf{S}^\perp = \mathbf{H}^+(w)$  (Theorem 8.1.3).

Next, introduce a new basis  $\{\bar{\xi}_1, \bar{\xi}_2, \dots, \bar{\xi}_n\}$  in  $\mathbf{X}$  with the property that

$$\langle \bar{\xi}_i, \xi_j \rangle = \delta_{ij}, \quad (8.3.6)$$

i.e., the dual basis of  $\{\xi_1, \xi_2, \dots, \xi_n\}$ . Defining the stationary vector process

$$\bar{x}(t) = \begin{bmatrix} U^{t+1}\bar{\xi}_1 \\ U^{t+1}\bar{\xi}_2 \\ \vdots \\ U^{t+1}\bar{\xi}_n \end{bmatrix}, \quad (8.3.7)$$

this property may be written

$$\mathbf{E}\{\bar{x}(t-1)x(t)'\} = I. \quad (8.3.8)$$

In particular, since  $\bar{x}(t-1) = Tx(t)$  for some nonsingular  $n \times n$  matrix  $T$ , (8.3.8) yields  $T = P^{-1}$  so that

$$\bar{x}(t-1) = P^{-1}x(t). \quad (8.3.9)$$

Therefore,

$$\bar{P} := \mathbf{E}\{\bar{x}(t)\bar{x}(t)'\} = P^{-1}. \quad (8.3.10)$$

Note that the apparent lack of symmetry between (8.3.2) and (8.3.7) is due to the corresponding lack of symmetry between the past and future spaces.

To construct a stochastic realization of  $y$  evolving backward in time we observe that, in view of (8.1.15) in Theorem 8.1.2, we have a representation

$$\begin{cases} U^* \bar{\xi}_i &= \sum_{j=1}^n \bar{a}_{ij} \bar{\xi}_j + \sum_{j=1}^p \bar{b}_{ij} \bar{w}_j(-1), & i = 1, 2, \dots, n \\ y_i(-1) &= \sum_{j=1}^n \bar{c}_{ij} \bar{\xi}_j + \sum_{j=1}^p \bar{d}_{ij} \bar{w}_j(-1), & i = 1, 2, \dots, m \end{cases}$$

to which we apply the shift  $U^{t+1}$  to obtain the stochastic system

$$\bar{\Sigma} \quad \begin{cases} \bar{x}(t-1) &= \bar{A}\bar{x}(t) + \bar{B}\bar{w}(t) \\ y(t) &= \bar{C}\bar{x}(t) + \bar{D}\bar{w}(t) \end{cases}, \quad (8.3.11)$$

This is a *backward stochastic system* in the sense that

$$\mathbf{H}^+ \vee \mathbf{X}^+ \perp \mathbf{H}^-(\bar{w}), \quad (8.3.12)$$

which amounts to the past generating noise being orthogonal to future and present output and state. Condition (8.3.12) follows from  $\bar{\mathbf{S}} = \mathbf{H}^+ \vee \mathbf{X}^+$  (Theorem 7.4.1) and  $\bar{\mathbf{S}}^\perp = \mathbf{H}^-(\bar{w})$  (Theorem 8.1.3).

A pair of stochastic systems (8.3.4) and (8.3.11) formed as above from the dual bases in  $\mathbf{X}$  will be referred to as a dual pair of stochastic realizations.

**Theorem 8.3.1.** *Let  $(\mathbb{H}, U, \mathbf{X})$  be a finite-dimensional Markovian representation, and let  $n := \dim \mathbf{X}$ . Then to each choice of dual bases in  $\mathbf{X}$  there is a pair of dual stochastic realizations, consisting of a forward system (8.3.4) and a backward*



8.3. The forward and backward systems (the finite-dimensional case) 197

system (8.3.11), which is unique modulo the choice of bases in the wandering subspaces  $\mathbf{W}$  and  $\bar{\mathbf{W}}$ , i.e., modulo multiplications of  $\begin{bmatrix} B \\ D \end{bmatrix}$  and  $\begin{bmatrix} \bar{B} \\ \bar{D} \end{bmatrix}$  from the right by orthogonal transformations, and which has the property

$$\{a'x(0) \mid a \in \mathbb{R}^n\} = \mathbf{X} = \{a'\bar{x}(-1) \mid a \in \mathbb{R}^n\}. \tag{8.3.13}$$

The forward and backward systems are connected via the relations

$$\bar{A} = A', \quad \bar{C} = CPA' + DB' \tag{8.3.14}$$

and

$$\bar{x}(t-1) = \bar{P}^{-1}x(t), \quad \bar{P} = P^{-1}, \tag{8.3.15}$$

where

$$P = E\{x(t)x(t)\} \quad \text{and} \quad \bar{P} = E\{\bar{x}(t)\bar{x}(t)\} \quad \text{for all } t \in \mathbb{Z}. \tag{8.3.16}$$

Moreover the splitting subspace  $\mathbf{X}$  is observable if and only if

$$\bigcap_{t=0}^{\infty} \ker CA^t = 0, \tag{8.3.17}$$

i.e.,  $(C, A)$  is (completely) observable, and constructible if and only if

$$\bigcap_{t=0}^{\infty} \ker \bar{C}(A')^t = 0, \tag{8.3.18}$$

i.e.,  $(\bar{C}, A')$  is (completely) observable. Finally, the Markovian representation is minimal if and only if both  $(C, A)$  and  $(\bar{C}, A')$  are observable.

**Proof.** The first statement has already been proven above. The orthogonal transformations are precisely the changes of coordinates in the wandering subspaces (8.1.9) and (8.1.13) under which  $w$  and  $\bar{w}$  and remain normalized white noises. Relations (8.3.15) have also been proven. To prove (8.3.14), note that

$$A = E\{x(1)x(0)'\}P^{-1} \tag{8.3.19}$$

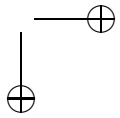
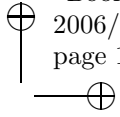
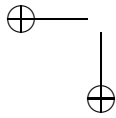
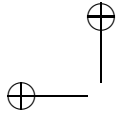
which follows immediately from (8.3.4) by observing that  $E\{w(0)x(0)'\} = 0$ . In the same way, we see from the backward system (8.3.11) that

$$\bar{A} = E\{\bar{x}(-1)\bar{x}(0)'\}\bar{P}^{-1} \quad \text{and} \quad \bar{C} = E\{y(0)\bar{x}(0)'\}\bar{P}^{-1}$$

which, in view of (8.3.15) may be written

$$\bar{A} = P^{-1}E\{x(0)x(1)'\} \quad \text{and} \quad \bar{C} = E\{y(0)x(1)'\}. \tag{8.3.20}$$

From (8.3.19) and (8.3.20) we readily see that  $\bar{A} = A'$ , and by inserting  $y(0) = Cx(0) + Dw(0)$  and  $x(1) = Ax(0) + Bw(0)$  in the second of equations (8.3.20) and



observing that  $E\{x(0)w(0)'\} = 0$ , we see that  $\bar{C} = CPA' + DB'$ . The statements about observability and constructibility follow from Theorem 6.6.2.  $\square$

Consequently, it is not enough that  $(C, A)$  is completely observable for the stochastic realization (8.3.4) to be minimal, even if  $(A, B)$  is completely reachable. In fact, reachability has nothing to do with minimality. As we shall see in the next section, reachability holds if and only if  $\mathbf{X}$  is purely nondeterministic. We may have purely nondeterministic components in the input spaces  $\mathbf{S}$  and  $\bar{\mathbf{S}}$  of (8.3.4) and (8.3.11), respectively, which turn up in the initial condition.

However, before leaving the topic of this section, let us state an important corollary of Theorem 8.3.1 .

**Proposition 8.3.2.** *A purely nondeterministic, stationary, vector process  $\{y(t)\}_{t \in \mathbb{Z}}$  has a rational spectral density if and only if it has a finite-dimensional Markovian representation  $(\mathbb{H}, U, \mathbf{X})$ .*

*Proof.* If  $y$  has a finite-dimensional Markovian representation  $(\mathbb{H}, U, \mathbf{X})$ , by Theorem 8.3.1, it is generated by a forward model (8.3.4) with the rational transfer function (6.2.2). Hence the spectral density  $\Phi(z) := W(z)W(z^{-1})$  is rational. Conversely, if  $\Phi$  is rational, it has a rational, analytic spectral factor, say the outer spectral factor  $W_-$ , and we can construct, along the lines of Chapter 6, a Markovian representation  $(\mathbb{H}, U, \mathbf{X}_-)$  that is finite-dimensional.  $\square$

### 8.4 Reachability, controllability and the deterministic subspace

The dynamics of the forward stochastic system (8.3.4) corresponds to the commutative diagram

$$\begin{array}{ccc}
 \mathbf{H}^-(w) & \xrightarrow{\mathcal{H}_-} & \mathbf{H}^+ \\
 \mathcal{R} \searrow & & \nearrow \mathcal{O} \\
 & \mathbf{X} &
 \end{array}
 \quad \mathcal{H}_- = \mathcal{O}\mathcal{R}
 \tag{8.4.1}$$

for the Hankel map  $\mathcal{H}_- := E^{\mathbf{H}^+} | \mathbf{H}^-(w)$ , where  $\mathcal{O} := E^{\mathbf{H}^+} | \mathbf{X}$  is the observability operator and  $\mathcal{R} := E^{\mathbf{X}} | \mathbf{H}^-(w)$  is the *reachability operator*. In fact, by Proposition 2.4.2 (iv), the factorization (8.4.1) is equivalent to the splitting property  $\mathbf{H}^-(w) \perp \mathbf{H}^+ | \mathbf{X}$ , which in turn follows from  $\mathbf{S} \perp \bar{\mathbf{S}} | \mathbf{X}$ , since  $\mathbf{H}^-(w) \subset \mathbf{S}$  and  $\mathbf{H}^+ \subset \bar{\mathbf{S}}$  (Lemma 2.4.1). Consequently, (8.4.1) holds for all Markovian representations, regardless of whether the dimension is finite or infinite (Theorem 7.4.1 and Theorem 8.1.3).

Likewise, since  $\mathbf{H}^+(\bar{w}) \subset \bar{\mathbf{S}}$  and  $\mathbf{H}^- \subset \mathbf{S}$ ,  $\mathbf{S} \perp \bar{\mathbf{S}} | \mathbf{X}$  implies that  $\mathbf{H}^+(\bar{w}) \perp$

$\mathbf{H}^- \mid \mathbf{X}$ , and consequently the diagram

$$\begin{array}{ccc} \mathbf{H}^+(\bar{w}) & \xrightarrow{\mathcal{H}_+} & \mathbf{H}^- \\ \mathcal{K} \searrow & & \nearrow \mathcal{C} \\ & \mathbf{X} & \end{array} \quad \mathcal{H}_+ = \mathcal{C}\mathcal{K}, \quad (8.4.2)$$

commutes, where  $\mathcal{H}_+ := \mathbf{E}^{\mathbf{H}^-} \mid \mathbf{H}^+(\bar{w})$  is a Hankel operator,  $\mathcal{C}$  is the constructibility operator and  $\mathcal{K} := \mathbf{E}^{\mathbf{X}} \mid \mathbf{H}^+(\bar{w})$  the *controllability operator*. This factorization illustrates the dynamics of the backward stochastic system (8.3.11), but it holds also for infinite-dimensional Markovian representations.

In complete analogy with (7.3.12) we can decompose the splitting subspace  $\mathbf{X}$  in two ways, namely

$$\mathbf{X} = \overline{\text{Im } \mathcal{R}} \oplus \ker \mathcal{R}^* \quad (8.4.3a)$$

$$\mathbf{X} = \overline{\text{Im } \mathcal{K}} \oplus \ker \mathcal{K}^*, \quad (8.4.3b)$$

where  $\overline{\text{Im } \mathcal{R}}$  and  $\overline{\text{Im } \mathcal{K}}$  are the reachable and controllable subspaces respectively. We say that  $\mathbf{X}$  is *reachable* if  $\ker \mathcal{R}^* = 0$  and *controllable* if  $\ker \mathcal{K}^* = 0$ .

**Proposition 8.4.1.** *Let  $(\mathbb{H}, U, \mathbf{X})$  be a Markovian representation with  $\mathbf{X} \sim (\mathbf{S}, \bar{\mathbf{S}})$ , and let  $\mathbf{S}_{-\infty}$  and  $\bar{\mathbf{S}}_{\infty}$  be the remote past of  $\mathbf{S}$  and the remote future of  $\bar{\mathbf{S}}$  respectively. Then  $\mathbf{X}$  is reachable if and only if*

$$\mathbf{X} \cap \mathbf{S}_{-\infty} = 0 \quad (8.4.4)$$

and controllable if and only if

$$\mathbf{X} \cap \bar{\mathbf{S}}_{\infty} = 0. \quad (8.4.5)$$

If  $(\mathbb{H}, U, \mathbf{X})$  is proper,  $\mathbf{X}$  is both reachable and controllable.

Markovian representation  $(\mathbb{H}, U, \mathbf{X})$  will called *purely nondeterministic* if both (8.4.4) and (8.4.5) hold.

**Proof.** From Theorem 8.1.3 it follows that

$$[\mathbf{H}^-(w)]^\perp = \mathbf{H}^+(w) \oplus \mathbf{S}_{-\infty},$$

and consequently, since  $\mathbf{X} \perp \mathbf{S}^\perp = \mathbf{H}^+(w)$ ,

$$\ker \mathcal{R}^* = \mathbf{X} \cap [\mathbf{H}^-(w)]^\perp = \mathbf{X} \cap \mathbf{S}_{-\infty},$$

as claimed; see Lemma A.2.4 in the appendix. The proof that

$$\ker \mathcal{K}^* = \mathbf{X} \cap \bar{\mathbf{S}}_{\infty}$$

is analogous. Then, the last statement is immediate.  $\square$

If  $\dim \mathbf{X} < \infty$ , we can strengthen these statements considerably. To this end, let us first relate reachability and controllability of  $\mathbf{X}$  to the forward and backward systems (8.3.4) and (8.3.11) respectively.

**Proposition 8.4.2.** *Let  $(\mathbb{H}, U, \mathbf{X})$  be a finite-dimensional Markovian representation, and let (8.3.4) and (8.3.11) be a corresponding dual pair of forward and backward realizations. Then  $\mathbf{X}$  is reachable if and only if  $(A, B)$  is reachable and controllable if and only if  $(A', \bar{B})$  is reachable.*

**Proof.** Since  $\ker \mathcal{R}^* = \mathbf{X} \cap [\mathbf{H}^-(w)]^\perp$ ,  $\mathbf{X}$  is reachable if and only if there is no nonzero  $a \in \mathbb{R}^n$  such that  $a'x(0) \perp \mathbf{H}^-(w)$ , i.e.,

$$a' E\{x(0)w(-t)'\}b = 0 \quad \text{for all } b \in \mathbb{R}^p \text{ and } t = 1, 2, 3, \dots \quad (8.4.6)$$

But, in view of (8.3.4),

$$x(0) = A^N x(-N) + A^{N-1} Bw(-N) + \dots + Bw(-1)$$

for all  $N = 1, 2, 3, \dots$ , and hence

$$E\{x(0)w(-t)'\} = A^{t-1} B.$$

Consequently, (8.4.6) is equivalent to

$$a' A^{t-1} B = 0 \quad \text{for } t = 1, 2, 3, \dots, \quad (8.4.7)$$

and hence  $\ker \mathcal{R}^* = 0$  if and only if  $(A, B)$  is reachable, i.e.,

$$[B, AB, A^2 B, \dots]$$

is full rank so that only  $a = 0$  satisfies (8.4.7). A symmetric argument shows that  $\mathbf{X}$  is controllable if and only if  $(A', \bar{B})$  is reachable.  $\square$

**Proposition 8.4.3.** *Let  $(\mathbb{H}, U, \mathbf{X})$  be a finite-dimensional Markovian representation of  $y$ . Then the following conditions are equivalent.*

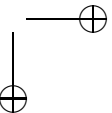
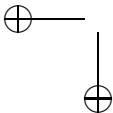
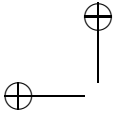
- (i)  $\mathbf{X}$  is proper
- (ii)  $\mathbf{X}$  is reachable
- (iii)  $\mathbf{X}$  is controllable
- (iv)  $A$  is a stability matrix, i.e.,  $|\lambda(A)| < 1$ .

*This can happen only if  $y$  is purely nondeterministic in both directions; i.e.,*

$$\cap_t U^t \mathbf{H}^- = \cap_t U^t \mathbf{H}^+ = 0. \quad (8.4.8)$$

**Proof.** It follows from (8.3.4) that  $P := E\{x(t)x(t)'\}$  satisfies the Lyapunov equation

$$P = APA' + BB' \quad (8.4.9)$$



Likewise, from (8.3.11), we see that  $\bar{P} = E\{\bar{x}(t)\bar{x}(t)'\}$  satisfies

$$\bar{P} = A'\bar{P}A + \bar{B}\bar{B}' \tag{8.4.10}$$

By Proposition 8.4.1, (i) implies (ii) and (iii). If (ii) holds, then  $(A, B)$  is reachable (Proposition 8.4.2), and hence, since  $P > 0$ , it follows from (8.4.9) that (iv) holds (Proposition A.3.2). Similarly, by (8.4.10) and the fact that  $A$  and  $A'$  have the same eigenvalues, we see that (iii) implies (iv).

It therefore remains to show that (iv) implies (i). To this end, note that (8.3.4) implies

$$x(t) = A^{N+t}x(-N) + \sum_{k=-N}^{t-1} A^{t-k-1}Bw(k)$$

for  $N \geq 1-t$ . If  $A$  is stable,  $A^{N+t}$  tends exponentially to zero as  $N \rightarrow \infty$ . Therefore

$$x(t) = \sum_{k=-\infty}^{t-1} A^{t-k-1}Bw(k) \tag{8.4.11}$$

is well-defined and is the unique solution of (8.3.4). Therefore,  $\mathbf{X} \subset \mathbf{H}^-(w)$ . Likewise, the second of equations (8.3.4) shows that

$$y(t) = \sum_{k=-\infty}^{t-1} CA^{t-k-1}Bw(k) + Dw(t) \tag{8.4.12}$$

so that

$$\mathbf{H}^- \subset \mathbf{H}^-(w). \tag{8.4.13}$$

Consequently,  $\mathbf{S} = \mathbf{H}^- \vee \mathbf{X}^- \subset \mathbf{H}^-(w)$  (Theorem 7.4.1). More precisely,  $\mathbf{S} = \mathbf{H}^-(w)$ , so, in view of Theorem 8.1.3,  $\mathbf{S}_{-\infty} = 0$ . A symmetric argument involving the backward system (8.3.11) shows that (iv) also implies that  $\bar{\mathbf{S}}_{\infty} = 0$ . Hence (i) follows as claimed. Moreover,

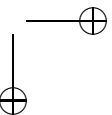
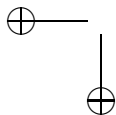
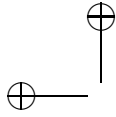
$$\mathbf{H}^+ \subset \mathbf{H}^+(\bar{w}). \tag{8.4.14}$$

Then the last statement of the theorem follows from (8.4.13) and (8.4.14).  $\square$

The last statement of Proposition 8.4.3 raises the question of whether a process  $y$  that admits a finite-dimensional Markovian representation is reversible; i.e., whether it is backward p.n.d. if and only if it is forward p.n.d..

**Proposition 8.4.4.** *A stationary stochastic process  $y$  is reversible if it admits a finite-dimensional Markovian representation. Then  $\cap_t U^t \mathbf{H}^- = \cap_t U^t \mathbf{H}^+$ .*

**Proof.** If  $y$  admits a finite-dimensional Markovian representation, then the predictor spaces  $\mathbf{X}_- \sim (\mathbf{S}_-, \bar{\mathbf{S}}_-)$  and  $\mathbf{X}_+ \sim (\mathbf{S}_+, \bar{\mathbf{S}}_+)$  are finite dimensional (Theorem 7.6.1), and hence so is the frame space  $\mathbf{H}^{\square} \sim (\mathbf{S}_+, \bar{\mathbf{S}}_-)$ . In view of Corollary 8.2.6, these representations are all normal. Therefore, the remote past of  $\mathbf{S}_-$  equals the remote future of  $\bar{\mathbf{S}}_-$ , which in turn equals the remote past of  $\mathbf{S}_+$  (via



$\mathbf{H}^\square$ ), which equals the remote future of  $\bar{\mathbf{S}}_+$ . However  $\cap_t U^t \mathbf{H}^-$  is the remote past of  $\mathbf{S}_-$  and  $\cap_t U^t \mathbf{H}^+$  is the remote future of  $\bar{\mathbf{S}}_+$ .  $\square$

Consequently, if  $y$  has a finite-dimensional Markovian representation, by Corollary 4.5.9, it has a unique decomposition

$$y(t) = y_0(t) + y_\infty(t), \quad t \in \mathbb{Z}, \quad (8.4.15)$$

where  $y_0$  is purely nondeterministic both in the forward and the backward direction, and  $y_\infty$  is purely deterministic in both directions and generates  $\cap_t U^t \mathbf{H}^- = \cap_t U^t \mathbf{H}^+$ .

The next theorem show that, if  $\mathbf{X} \sim (\mathbf{S}, \bar{\mathbf{S}})$  is finite-dimensional, then  $\mathbf{S}$  and  $\bar{\mathbf{S}}$  have the same multiplicity, and hence the forward and backward generating processes have the same dimension.

**Theorem 8.4.5.** *Let  $(\mathbb{H}, U, \mathbf{X})$  be a finite-dimensional Markovian representation of  $y$  with  $\mathbf{X} \sim (\mathbf{S}, \bar{\mathbf{S}})$  and with generating processes  $w$  and  $\bar{w}$ . Then the remote past  $\mathbf{S}_{-\infty}$  of  $\mathbf{S}$  equals the remote future  $\bar{\mathbf{S}}_\infty$  of  $\bar{\mathbf{S}}$ , and  $\mathbf{X}$  and  $\mathbb{H}$  have the orthogonal decompositions*

$$\mathbf{X} = \mathbf{X}_0 \oplus \mathbf{X}_\infty \quad (8.4.16)$$

and

$$\mathbb{H} = \mathbb{H}_0 \oplus \mathbf{X}_\infty \quad (8.4.17)$$

respectively, where  $\mathbf{X}_0 \subset \mathbb{H}_0$ , and where  $\mathbf{X}_\infty$  and  $\mathbb{H}_0$  are the doubly invariant subspaces

$$\mathbf{X}_\infty = \mathbf{S}_{-\infty} = \bar{\mathbf{S}}_\infty \quad (8.4.18)$$

and

$$\mathbb{H}_0 = \mathbf{H}(w) = \mathbf{H}(\bar{w}) \quad (8.4.19)$$

respectively. In particular,  $\bar{p} = p$ , i.e.,  $w$  and  $\bar{w}$  have the same dimension. Moreover, if  $U_0 := U|_{\mathbb{H}_0}$ ,  $(\mathbb{H}_0, U_0, \mathbf{X}_0)$  is a proper Markovian representation for the purely nondeterministic part  $y_0$  of  $y$ , and it has the same generating processes as  $(\mathbb{H}, U, \mathbf{X})$ .

**Proof.** In view of Theorem 8.2.3,  $\mathbf{X}$  has an orthogonal decomposition (8.4.16), where  $\mathbf{X}_\infty$  is doubly invariant. Since  $\mathbf{X}$  is finite-dimensional, it follows from Corollaries 8.2.6 and 8.2.4 that (8.4.18) holds. By Corollary 8.2.5,  $\mathbf{S}_0 := \mathbf{S} \ominus \mathbf{X}_\infty$  and  $\bar{\mathbf{S}}_0 := \bar{\mathbf{S}} \ominus \mathbf{X}_\infty$  intersect perpendicularly,  $\mathbf{X}_0 = \mathbf{S}_0 \cap \bar{\mathbf{S}}_0$ , and  $U^* \mathbf{S}_0 \subset \mathbf{S}_0$  and  $U \bar{\mathbf{S}}_0 \subset \bar{\mathbf{S}}_0$ . Therefore, in view of (8.1.28) and (8.1.31),

$$\mathbf{S}_0 = \mathbf{H}^-(w) \quad \text{and} \quad \bar{\mathbf{S}}_0 = \mathbf{H}^+(\bar{w}). \quad (8.4.20)$$

Moreover,  $\mathbb{H} = \mathbb{H}_0 \oplus \mathbf{X}_\infty$ , where  $\mathbb{H}_0$  is given by (8.4.19). Hence, if  $\mu$  is the multiplicity of  $\mathbb{H}_0$ ,  $\bar{p} = p = \mu$ . Since  $\mathbf{H}_0^- := \mathbf{H}^-(y_0) \subset \mathbf{H}^- \subset \mathbf{S} = \mathbf{S}_0 \oplus \mathbf{X}_\infty$  and  $y_0$  is purely nondeterministic, we must have  $\mathbf{H}_0^- \subset \mathbf{S}_0$ . Likewise,  $\mathbf{H}_0^+ := \mathbf{H}^+(y_0) \subset \bar{\mathbf{S}}_0$ . Consequently,  $(\mathbb{H}_0, U_0, \mathbf{X}_0)$  is a proper Markovian representation for  $y_0$  with generating processes  $w$  and  $\bar{w}$ , as claimed.  $\square$

In the orthogonal decomposition

$$\mathbf{X} = \mathbf{X}_0 \oplus \mathbf{X}_\infty$$

we shall call  $\mathbf{X}_0$  the *proper subspace* of  $\mathbf{X}$  and  $\mathbf{X}_\infty$  the *deterministic subspace* of  $\mathbf{X}$ . In view of Proposition 8.4.4,

$$\mathbf{Y}_\infty := \cap_t U^t \mathbf{H}^- = \cap_t U^t \mathbf{H}^+, \quad (8.4.21)$$

i.e., the remote past and the remote future of  $y$  are the same. The following corollary describes the relation between  $\mathbf{Y}_\infty$  and the deterministic subspace of  $\mathbf{X}$ .

**Corollary 8.4.6.** *If the process  $y$  has a finite-dimensional Markovian representation  $(\mathbb{H}, U, \mathbf{X})$ ,  $\mathbf{Y}_\infty \subset \mathbf{X}_\infty$ , where  $\mathbf{X}_\infty$  is the deterministic subspace of  $\mathbf{X}$ . If  $\mathbf{X}$  is observable or constructible, then  $\mathbf{Y}_\infty = \mathbf{X}_\infty$ .*

**Proof.** Let  $\mathbf{X} \sim (\mathbf{S}, \bar{\mathbf{S}})$  be an arbitrary finite-dimensional Markovian splitting subspace. Then, since  $\mathbf{H}^- \subset \mathbf{S}$ ,

$$\mathbf{Y}_\infty := \cap_t U^t \mathbf{H}^- \subset \cap_t U^t \mathbf{S} = \mathbf{S}_{-\infty} = \mathbf{X}_\infty,$$

which proves the first statement. To prove the second, consider the observability operator  $\mathcal{O} : \mathbf{X} \rightarrow \mathbf{H}^-$  defined by  $\mathcal{O} := \mathbf{E}^{\mathbf{H}^-} | \mathbf{X}$ . By Theorem 8.4.5 and, in particular, (??) and (8.4.20),  $\mathbf{H}_0^+ := \mathbf{H}^+(y_0) \subset \bar{\mathbf{S}}_0 \perp \mathbf{X}_\infty$ , where  $y_0$  is the purely nondeterministic part of  $y$ . Consequently, since  $\mathbf{H}^+ = \mathbf{H}_0^+ \oplus \mathbf{Y}_\infty$ ,

$$\mathcal{O}\lambda = \mathbf{E}^{\mathbf{H}_0^+} \lambda + \mathbf{E}^{\mathbf{Y}_\infty} \lambda = \mathbf{E}^{\mathbf{Y}_\infty} \lambda \quad \text{for } \lambda \in \mathbf{X}_\infty. \quad (8.4.22)$$

Hence, since  $\mathbf{Y}_\infty \subset \mathbf{X}_\infty$ , the operator  $\mathcal{O}$  can be injective only if  $\mathbf{Y}_\infty = \mathbf{X}_\infty$ . Thus observability of  $\mathbf{X}$  implies  $\mathbf{Y}_\infty = \mathbf{X}_\infty$ . In the same way we show that constructibility also implies that  $\mathbf{Y}_\infty = \mathbf{X}_\infty$ .  $\square$

We are now in a position to interpret the decomposition (8.4.3) in terms of the decomposition (8.4.16) of Theorem 8.4.5, in the finite-dimensional case.

**Corollary 8.4.7.** *Let  $(\mathbb{H}, U, \mathbf{X})$  be a finite-dimensional Markovian representation. Then the range spaces of the reachability operator  $\mathcal{R}$  and the controllability operator  $\mathcal{K}$  coincide and are equal to the proper subspace  $\mathbf{X}_0$  of  $\mathbf{X}$ , i.e.,*

$$\text{Im } \mathcal{R} = \mathbf{X}_0 = \text{Im } \mathcal{K}. \quad (8.4.23)$$

Moreover, the purely deterministic part  $\mathbf{X}_\infty$  of  $\mathbf{X}$  is given by

$$\ker \mathcal{R}^* = \mathbf{X}_\infty = \ker \mathcal{K}^*. \quad (8.4.24)$$

**Proof.** As a corollary of Proposition 8.4.1 we have

$$\ker \mathcal{R}^* = \mathbf{X} \cap \mathbf{S}_{-\infty} \quad \text{and} \quad \ker \mathcal{K}^* = \mathbf{X} \cap \bar{\mathbf{S}}_\infty. \quad (8.4.25)$$

But, in the finite-dimensional case,  $(\mathbb{H}, U, \mathbf{X})$  is normal (Corollary 8.2.6), and hence  $\mathbf{X}_\infty = \mathbf{S}_{-\infty} = \bar{\mathbf{S}}_\infty$  (Corollary 8.2.4). Therefore, since  $\mathbf{X}_\infty \subset \mathbf{X}$ , (8.4.24) follows. Consequently, since  $\mathbf{X}_0 = \mathbf{X} \ominus \mathbf{X}_\infty$ , (8.4.23) follows from (8.4.3) and the fact that the range spaces of  $\mathcal{R}$  and  $\mathcal{K}$  are closed in the finite-dimensional case.  $\square$

The orthogonal decomposition (8.4.16) of the Markovian splitting subspace  $\mathbf{X}$  into a proper and a purely nondeterministic part induces a special structure of the corresponding forward and backward stochastic systems, (8.3.4) and (8.3.11) respectively, provided the bases are chosen appropriately. In fact, if  $n := \dim \mathbf{X}$  and  $n_0 := \dim \mathbf{X}_0$ , we take the basis  $\{\xi_1, \xi_2, \dots, \xi_n\}$  to be adapted to the decomposition

$$\mathbf{X} = \mathbf{X}_0 \oplus \mathbf{X}_\infty \quad (8.4.26)$$

in the sense that  $\{\xi_1, \xi_2, \dots, \xi_{n_0}\}$  is a basis in  $\mathbf{X}_0$  and  $\{\xi_{n_0+1}, \dots, \xi_n\}$  is a basis in  $\mathbf{X}_\infty$ . Then the dual basis  $\{\bar{\xi}_1, \bar{\xi}_2, \dots, \bar{\xi}_{n_0}\}$  is also adapted to (8.4.26), and the covariance matrices  $P$  and  $\bar{P}$  take the forms

$$P = \begin{bmatrix} P_0 & 0 \\ 0 & P_\infty \end{bmatrix} \quad \text{and} \quad \bar{P} = \begin{bmatrix} P_0^{-1} & 0 \\ 0 & P_\infty^{-1} \end{bmatrix} \quad (8.4.27)$$

respectively, where  $P_0$  is  $n_0 \times n_0$ .

**Theorem 8.4.8.** *Let  $(\mathbb{H}, U, \mathbf{X})$  be a finite-dimensional Markovian representation of  $y$ , and let (8.3.4) and (8.3.11) be a dual pair of stochastic realizations with bases adapted to the decomposition (8.4.26) in the sense described above. Then the forward system attains the form*

$$\begin{cases} \begin{bmatrix} x_0(t+1) \\ x_\infty(t+1) \end{bmatrix} = \begin{bmatrix} A_0 & 0 \\ 0 & A_\infty \end{bmatrix} \begin{bmatrix} x_0(t) \\ x_\infty(t) \end{bmatrix} + \begin{bmatrix} B_0 \\ 0 \end{bmatrix} w(t) \\ y(t) = \begin{bmatrix} C_0 & C_\infty \end{bmatrix} \begin{bmatrix} x_0(t) \\ x_\infty(t) \end{bmatrix} + Dw(t) \end{cases} \quad (8.4.28)$$

where

$$|\lambda(A_0)| < 1 \quad \text{and} \quad |\lambda(A_\infty)| = 1, \quad (8.4.29)$$

$(A_0, B_0)$  is reachable, and

$$\text{Im } \mathcal{R} = \{a'x_0(0) \mid a \in \mathbb{R}^{n_0}\} \quad \ker \mathcal{R}^* = \{a'x_\infty(0) \mid a \in \mathbb{R}^{n-n_0}\} \quad (8.4.30)$$

Moreover,

$$y_0(t) = C_0x_0(t) + Dw(t) \quad (8.4.31)$$

is the purely nondeterministic part of  $y$  and

$$y_\infty = C_\infty x_\infty(t) \quad (8.4.32)$$

is the purely deterministic part.



Dually, the backward system (8.3.11) takes the form

$$\begin{cases} \begin{bmatrix} \bar{x}_0(t+1) \\ \bar{x}_\infty(t+1) \end{bmatrix} &= \begin{bmatrix} A'_0 & 0 \\ 0 & A'_\infty \end{bmatrix} \begin{bmatrix} \bar{x}_0(t) \\ \bar{x}_\infty(t) \end{bmatrix} + \begin{bmatrix} \bar{B}_0 \\ 0 \end{bmatrix} \bar{w}(t) \\ y(t) &= \begin{bmatrix} \bar{C}_0 & \bar{C}_\infty \end{bmatrix} \begin{bmatrix} \bar{x}_0(t) \\ \bar{x}_\infty(t) \end{bmatrix} + \bar{D}\bar{w}(t) \end{cases} \quad (8.4.33)$$

where  $(A'_0, \bar{B}_0)$  is reachable and

$$\text{Im } \mathcal{K} = \{a'\bar{x}_0(-1) \mid a \in \mathbb{R}^{n_0}\} \quad \ker \mathcal{K}^* = \{a'\bar{x}_\infty(-1) \mid a \in \mathbb{R}^{n-n_0}\} \quad (8.4.34)$$

Finally,

$$\bar{C}_0 = C_0 P_0 A'_0 + D B'_0, \quad \bar{C}_\infty = C_\infty P_\infty A'_\infty, \quad (8.4.35)$$

and  $\mathbf{X}_\infty = \mathbf{Y}_\infty$  if and only if  $(C_\infty, A_\infty)$  is observable, or, equivalently, if and only if  $(\bar{C}_\infty, A'_\infty)$  is observable.

**Proof.** Let

$$y(t) = y_0(t) + y_\infty(t) \quad (8.4.36)$$

be the decomposition (4.5.20) of  $y$ , where  $y_0$  is the purely nondeterministic and  $y_\infty$  the purely deterministic component of  $y$ . Let

$$x(0) = \begin{bmatrix} x_0(0) \\ x_\infty(0) \end{bmatrix} \quad \text{and} \quad \bar{x}(-1) = \begin{bmatrix} \bar{x}_0(-1) \\ \bar{x}_\infty(-1) \end{bmatrix} = P^{-1}x(0)$$

be the dual pair of bases in  $\mathbf{X}$  formed as in Section 8.3 and adapted to the decomposition (8.4.26) so that  $x_0(0)$  and  $\bar{x}_0(-1) := P_0^{-1}x_0(0)$  are bases in  $\mathbf{X}_0$  and  $x_\infty(0)$  and  $\bar{x}_\infty(-1) := P_\infty^{-1}x_\infty(0)$  are bases in  $\mathbf{X}_\infty$ . Then (8.4.30) and (8.4.34) follow from Corollary 8.4.7. By Theorem 8.4.5,  $(\mathbb{H}_0, U_0, \mathbf{X}_0)$  is a proper Markovian representation for  $y_0$  with generating processes  $w$  and  $\bar{w}$ . Consequently, it has a forward system

$$\begin{cases} x_0(t+1) &= A_0 x_0(t) + B_0 w(t) \\ y_0(t) &= C_0 x_0(t) + D w(t) \end{cases} \quad (8.4.37)$$

with  $|\lambda(A_0)| < 1$  and with  $(A_0, B_0)$  reachable (Propositions 8.4.2 and 8.4.3); and a backward system

$$\begin{cases} \bar{x}_0(t+1) &= A'_0 \bar{x}_0(t) + \bar{B}_0 \bar{w}(t) \\ y_0(t) &= \bar{C}_0 \bar{x}_0(t) + \bar{D} \bar{w}(t) \end{cases} \quad (8.4.38)$$

where  $(A'_0, \bar{B}_0)$  is reachable.

Next we derive a representation for the purely deterministic part  $y_\infty$ . Since the components of  $y_\infty(0)$  belong to  $\mathbf{Y}_\infty \subset \mathbf{X}_\infty$ , there is an  $m \times (n - n_0)$  matrix  $C_\infty$  such that

$$y_\infty(0) = C_\infty x_\infty(0). \quad (8.4.39)$$

Moreover, since  $\mathbf{X}_\infty$  is invariant under  $U$ , there is an  $(n - n_0) \times (n - n_0)$  matrix  $A_\infty$  such that

$$x_\infty(1) = A_\infty x_\infty(0). \quad (8.4.40)$$

Applying the shift  $U^t$  to each of the components of (8.4.39) and (8.4.40), we obtain

$$\begin{cases} x_\infty(t+1) &= A_\infty x_\infty(t) \\ y_\infty(t) &= C_\infty x_\infty(t) \end{cases} \quad (8.4.41)$$

which together with (8.4.36) and (8.4.37) yields (8.4.28). Clearly (8.4.28) is a forward stochastic system in the sense of (8.3.5).

A similar analysis in the backward direction yields

$$\begin{cases} \bar{x}_\infty(t-1) &= \bar{A}_\infty \bar{x}_\infty(t) \\ y_\infty(t) &= \bar{C}_\infty \bar{x}_\infty(t) \end{cases} \quad (8.4.42)$$

which together with (8.4.38) and (8.4.36) constitutes the backward counterpart of (8.4.28). Now, in view of the correspondence  $\bar{A} = A'$  between a forward system (8.3.4) and a backward system (8.3.11) constructed from dual bases (Theorem 8.1.3), we must have

$$\bar{A}_\infty = A'_\infty \quad (8.4.43)$$

so that (8.4.33) follows. Moreover, (8.4.35) follows from (8.3.14).

To show that  $|\lambda(A_\infty)| = 1$ , observe from (8.4.41) that  $P_\infty := \mathbf{E}\{x_\infty(t)x_\infty(t)'\}$  satisfies the degenerate Lyapunov equation

$$P_\infty = A_\infty P_\infty A'_\infty, \quad (8.4.44)$$

and hence  $A_\infty$  has all its eigenvalues on the unit circle. In fact,  $A_\infty$  is similar to  $Q^* := P_\infty^{-1/2} A_\infty P_\infty^{1/2}$  and has therefore the same eigenvalues as  $Q$ . But from (8.4.44) it follows that  $Q^*Q = I$  so that, if  $Qv = \lambda v$ , we have  $|\lambda|^2 v^*v = v^*v$ , showing that  $|\lambda| = 1$  as claimed.

Finally, from (8.4.22) we have

$$\ker(\mathcal{O}|_{\mathbf{X}_\infty}) = \mathbf{X}_\infty \ominus \mathbf{Y}_\infty$$

and hence  $\mathbf{Y}_\infty = \mathbf{X}_\infty$  if and only if  $\mathcal{O}|_{\mathbf{X}_\infty}$  is injective. In view of (8.4.41) and Theorem 8.1.2, this is equivalent to  $(C_\infty, A_\infty)$  being observable. Since  $A_\infty$ , having no zero eigenvalues, is nonsingular, this in turn is equivalent to  $(\bar{C}_\infty, A'_\infty)$  being observable.  $\square$

**Remark 8.4.9.** Suppose that the process  $y$  generates a finite-dimensional space  $\mathbf{H}$  of dimension  $n < \infty$ . Then  $y$  must be purely deterministic and  $\mathbf{H}^- = \mathbf{H}^+ = \mathbf{H} = \mathbf{Y}_\infty$ . Clearly  $\mathbf{H}$  can be generated by  $n$  successive variables of the process, which we collect in a  $mn$ -dimensional vector

$$\mathbf{y} := \begin{bmatrix} y(t) \\ \vdots \\ y(t-n+1) \end{bmatrix}.$$

However, the shifted vector

$$\begin{bmatrix} y(t+1) \\ \vdots \\ y(t-n) \end{bmatrix}$$

also generates  $\mathbf{H}$ , and hence there must be a real matrix  $F$  such that

$$Ub'y = b'Fy \tag{8.4.45}$$

for all  $b \in \mathbb{R}^{mn}$ .

We want to construct a minimal stochastic realization of  $y$ . By Corollary 8.4.6,  $y$  has only one minimal Markovian representation, namely  $(\mathbb{H}, U, \mathbf{X})$ , where  $\mathbf{X} = \mathbf{H}$ . Let  $x(0) = (\xi_1, \dots, \xi_n)'$  be a basis in  $\mathbf{X}$ . Then there is a matrix  $\Omega$  with linearly independent columns such that  $y = \Omega x(0)$ , which can be solved for  $x(0)$  to yield  $x(0) = (\Omega'\Omega)^{-1}\Omega'y$ . Together with (8.4.45) this yields

$$Ua'x(0) = Ua'(\Omega'\Omega)^{-1}\Omega'y = a'(\Omega'\Omega)^{-1}\Omega'F\Omega x(0),$$

so setting

$$A := (\Omega'\Omega)^{-1}\Omega'F\Omega, \tag{8.4.46}$$

we have  $Ua'x(0) = a'Ax(0)$  for all  $a \in \mathbb{R}^n$ . Moreover, defining  $C$  to be the first block of  $m$  rows of  $\Omega$ ,  $y(0) = Cx(0)$ . Consequently we have the stochastic system

$$\begin{cases} x(t+1) = Ax(t) \\ y(t) = Cx(t) \end{cases} .$$

In fact,  $\Omega$  is just the observability matrix of this system. As explained on page 206, the matrix  $A$  has all its eigenvalues on the unit circle. A symmetric argument yields the corresponding backward system.

## 8.5 Minimality and nonminimality of finite-dimensional models

Let us sum up what we have learned so far about minimality of an arbitrary finite-dimensional (forward) linear stochastic system

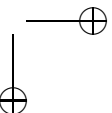
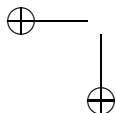
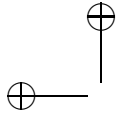
$$(\Sigma) \quad \begin{cases} x(t+1) = Ax(t) + Bw(t) \\ y(t) = Cx(t) + Dw(t) \end{cases} \tag{8.5.1}$$

**Proposition 8.5.1.** *Let  $\Sigma$  be the linear stochastic system (8.5.1), let  $W$  be its transfer function*

$$W(z) = C(zI - A)^{-1}B + D,$$

let  $\Phi(z) := W(z)W(z^{-1})$ , and

$$\mathbf{X} = \{a'x(0) \mid a \in \mathbb{R}^m\} = \mathbf{X}_0 \oplus \mathbf{X}_\infty,$$



where  $\mathbf{X}_\infty$  is the deterministic subspace of  $\mathbf{X}$ . Then  $\Phi(e^{i\theta})$  is the spectral density of the purely nondeterministic part of  $y$  and

$$\frac{1}{2} \deg \Phi \leq \deg W \leq \dim \mathbf{X}_0 \leq \dim \mathbf{X} \leq \dim \Sigma \tag{8.5.2}$$

Moreover,

- (i)  $\frac{1}{2} \deg \Phi = \deg W$  if and only if  $W$  is a minimal,
- (ii)  $\deg W = \dim \mathbf{X}_0$  if and only if  $(C, A)$  is observable,
- (iii)  $\dim \mathbf{X}_0 = \dim \mathbf{X}$  if and only if  $|\lambda(A)| < 1$ ,
- (iv)  $\dim \mathbf{X} = \dim \Sigma$  if and only if  $x(0)$  is a basis in  $\mathbf{X}$ ,
- (v)  $\dim \mathbf{X}_0 = \dim \Sigma$  if and only if  $(A, B)$  is reachable.

In particular, if  $y$  is purely nondeterministic,  $\Sigma$  is a minimal stochastic realization of  $y$  if and only if (i), (ii) and (v) hold. Otherwise,  $\Sigma$  is minimal if and only if (i), (ii) and (iv) hold and

Proposition 8.5.1 is merely a summary of results in Chapter 6 and Section 8.4. In particular, (i) is just a definition (Definition 6.8.3), (ii) follows from Corollary 6.6.3, and (iii) follows from Theorem 6.8.3, where also a number of equivalent conditions are given. We refer the reader to Section 8.4 for the other statements. Note that (ii) implies that (ii) implies that  $\mathbf{X}_\infty = \mathbf{Y}_\infty$ , where  $\mathbf{Y}_\infty$  is defined by (8.4.21) (Corollary 8.4.6).

Next, suppose that  $(A, B)$  is observable and  $A$  is a stability matrix; i.e.,  $|\lambda(A)| < 1$ . Then it is not enough that  $(A, B)$  is reachable to insure that  $\Sigma$  is a minimal realization of  $y$ ; for this we also need that the transfer function  $W$  is a minimal spectral factor. However, it is enough that the steady-state Kalman filter is reachable, as we shall demonstrate next.

To this end, consider the Kalman filter applied to the model  $\Sigma$ ,

$$\hat{x}(t+1) = A\hat{x}(t) + K(t)[y(t) - C\hat{x}(t)], \quad \hat{x}(\tau) = 0,$$

which estimates

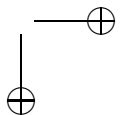
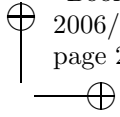
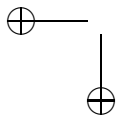
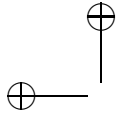
$$\hat{x}_i(t) = \mathbf{E}^{\mathbf{H}_{[\tau, t-1]}} x_i(0), \quad i = 1, 2, \dots, n,$$

where  $\mathbf{H}_{[\tau, t-1]} = \overline{\text{span}}\{a'y(k) \mid a \in \mathbb{R}^m, k = \tau, \tau + 1, \dots, t - 1\}$ . It follows from Lemma 6.2 that, for each  $a \in \mathbb{R}^n$ ,

$$a'\hat{x}(t) \rightarrow a'\hat{x}_\infty(t) := \mathbf{E}^{\mathbf{H}^-} a'x(t)$$

strongly as  $\tau \rightarrow -\infty$ , and hence

$$\{a'\hat{x}_\infty(0) \mid a \in \mathbb{R}\} = \mathbf{E}^{\mathbf{H}^-} \mathbf{X}, \tag{8.5.3}$$



8.6. Parameterization of finite-dimensional minimal Markovian representations 209

where  $\mathbf{X}$  is the splitting subspace of  $\Sigma$ . Moreover,  $y(t) - C\hat{x}(t)$  tends to  $G\nu(t)$ , where  $\nu$  is a normalized white noise and  $G$  is an invertible matrix (cf, Section 6.9). Consequently, we have the steady-state Kalman filter

$$\begin{cases} \hat{x}_\infty(t+1) = A\hat{x}_\infty(t) + K_\infty\nu(t) \\ y(t) = C\hat{x}_\infty(t) + G\nu(t) \end{cases}, \tag{8.5.4}$$

which is itself a stochastic realization of  $y$ . Note that, in general, we  $\hat{x}_\infty \neq x_-$ , the state process of the predictor space  $\mathbf{X}_-$ . Indeed, since  $\Sigma$  is not assumed to be minimal,  $\hat{x}_\infty$  and  $x_-$  may not have the same dimension. However, we have the following minimality criterion.

**Proposition 8.5.2.** *An observable system  $\Sigma$  with  $A$  a stability matrix is a minimal realization of  $y$  if and only if its steady state Kalman filter (8.5.4) is completely reachable in the sense that  $(A, K_\infty)$  is reachable.*

**Proof.** Since  $\mathbf{X}$  is observable, and hence  $\mathbf{X} \perp \mathbf{N}^-$  (Corollary 7.4.14), we have

$$E^{\mathbf{H}^-} \mathbf{X} = \mathbf{X}_-$$

(Proposition 7.4.13), and therefore, by (8.5.3),

$$\{a'\hat{x}_\infty(0) \mid a \in \mathbb{R}\} = \mathbf{X}_-$$

Then, by Proposition 8.5.1 (iii)-(iv),  $\hat{x}_\infty(0)$  is a basis in  $\mathbf{X}_-$  if and only if  $(A, K_\infty)$  is reachable, in which case  $\mathbf{X}$  has the same dimension as the minimal splitting subspace  $\mathbf{X}_-$  and thus is also minimal (Theorem 7.6.1). However,  $\dim \mathbf{X} \leq \dim \mathbf{X}_-$ , and hence the proposition follows.  $\square$

## 8.6 Parameterization of finite-dimensional minimal Markovian representations

Suppose that  $y$  has a finite-dimensional Markovian representation. Then all minimal Markovian representations  $(\mathbb{H}, U, \mathbf{X})$  have the same dimension, say  $n$  (Theorem 7.6.1). In this section we show that the matrices  $(A, C, \tilde{C})$  in (8.3.4) and (8.3.11) can be selected to be the same for all minimal Markovian representations.

To this end, we introduce a partial ordering between minimal Markovian representations along the lines of Section 7.7.

**Definition 8.6.1.** *Given two minimal Markovian representations,  $M_1 := (\mathbb{H}_1, U_1, \mathbf{X}_1)$  and  $M_2 := (\mathbb{H}_2, U_2, \mathbf{X}_2)$  of  $y$ , let  $M_1 \prec M_2$  denote the ordering*

$$\|E^{\mathbf{X}_1} \lambda\| \leq \|E^{\mathbf{X}_2} \lambda\| \quad \text{for all } \lambda \in \mathbf{H}^+, \tag{8.6.1}$$

where the norms are those of the respective ambient spaces  $\mathbb{H}_1$  and  $\mathbb{H}_2$ . If  $M_1 \prec M_2$  and  $M_2 \prec M_1$  both hold, the  $M_1$  and  $M_2$  are said to be equivalent ( $M_1 \sim M_2$ ).

If  $M_1 \sim M_2$  and either  $M_1$  or  $M_2$  are internal, then both are internal, and  $M_1 = M_2$  (Corollary 7.7.10). Let  $\mathcal{M}$  be the family of all equivalence classes of minimal Markovian representations of  $y$ , and let  $\mathcal{M}_0$  be the subclass of all internal minimal Markovian representations. Both  $\mathcal{M}$  and  $\mathcal{M}_0$  are partially ordered sets with a minimum and a maximum element,  $M_- := (\mathbf{H}, U, \mathbf{X}_-)$  and  $M_+ := (\mathbf{H}, U, \mathbf{X}_+)$ , respectively (Theorem 7.7.3).

Let  $(\mathbf{H}, U, \mathbf{X})$  be a minimal Markovian representation. Given any basis  $x_+(0)$  in  $\mathbf{X}_+$ , the random vector  $x(0)$  defined by

$$a'x(0) = E^{\mathbf{X}} a'x_+(0), \quad \text{for all } a \in \mathbb{R}^n. \quad (8.6.2)$$

forms a basis in  $\mathbf{X}$  (Lemma 7.7.4). This selection of bases are referred to as a *uniform choice of bases* in  $\mathcal{M}$ . In particular,

$$a'x_-(0) = E^{\mathbf{X}} a'x_+(0), \quad \text{for all } a \in \mathbb{R}^n \quad (8.6.3)$$

defines a basis  $a$  in  $\mathbf{X}_-$ . Now, as in (8.3.9), define the dual basis

$$\bar{x}_-(-1) = P_-^{-1}x_-(0) \quad (8.6.4)$$

in  $\mathbf{X}$ , where  $P_- := E\{x_-(0)x_-(0)'\}$ . Then, by symmetry

$$a'\bar{x}(-1) = E^{\mathbf{X}} a'\bar{x}_-(-1), \quad \text{for all } a \in \mathbb{R}^n \quad (8.6.5)$$

defines a basis in  $\mathbf{X}$ . This is precisely, the pair of bases introduced in Section 8.3, as seen from the following proposition.

**Proposition 8.6.2.** *Let  $x_+(0)$  be an arbitrary basis in  $\mathbf{X}_+$ . Then the pair of random vectors  $x(0)$  and  $\bar{x}(-1)$ , defined via the construction (8.6.2)–(8.6.5), is a dual pair of bases in  $\mathbf{X}$ ; i.e.,*

$$E\{x(0)\bar{x}(-1)\} = I, \quad (8.6.6)$$

or, equivalently,

$$\bar{x}(-1) = P^{-1}x(0), \quad \text{where } P := E\{x(0)x(0)'\}. \quad (8.6.7)$$

**Proof.** To prove (8.6.6) we form

$$a'E\{x(0)\bar{x}(-1)\}b = \langle a'x(0), E^{\mathbf{X}} b'P^{-1}x_-(0) \rangle = \langle a'x(0), b'P_-^{-1}x_-(0) \rangle.$$

However, by Proposition 7.7.7,

$$a'x_-(0) = E^{\mathbf{X}_-} a'x(0), \quad \text{for all } a \in \mathbb{R}^n,$$

and hence

$$a'(x(0) - x_-(0)) \perp \mathbf{X}_-.$$

Consequently,

$$a' E\{x(0)\bar{x}(-1)\}b = \langle a'x_-(0), b'P_-^{-1}x_-(0) \rangle = a' E\{x_-(0)x_-(0)\}P_-^{-1}b = a'b$$

for all  $a, b \in \mathbb{R}^n$ , which establishes (8.6.6).  $\square$

In Section 8.3 we saw that, given any dual pair of bases  $(x(0), \bar{x}(-1))$  in  $\mathbf{X}$ , there is a corresponding forward system (8.3.4) and backward system (8.3.11). Let us consider the corresponding triplet  $(A, C, \bar{C})$  of systems matrices.

**Theorem 8.6.3.** *For any uniform choice of bases, the triplet  $(A, C, \bar{C})$  is invariant over  $\mathcal{M}$ .*

**Proof.** Let  $U(\mathbf{X})$  be the operator (7.5.1) defined in Section 7.5. From (8.3.4) it readily follows that

$$U(\mathbf{X})a'x(0) = a'Ax(0). \tag{8.6.8}$$

Moreover, if  $\mathcal{O}$  is the observability operator  $E^{\mathbf{H}^+}|_{\mathbf{X}}$ , we see from the first commutative diagram of Theorem 7.5.1 that

$$U(\mathbf{X})\mathcal{O}^*a'x_+(0) = \mathcal{O}^*Ua'x_+(0)$$

for all  $a \in \mathbb{R}^n$ , or, which is the same,

$$U(\mathbf{X})E^{\mathbf{X}}a'x_+(0) = E^{\mathbf{X}}a'x_+(1).$$

In view of (8.6.2) and (8.6.8), this can also be written

$$a'Ax(0) = E^{\mathbf{X}}a'x_+(1). \tag{8.6.9}$$

Now,

$$a'x_+(1) = a'A_+x_+(0) + a'B_+w_+(0).$$

We want to show that the last term is orthogonal to  $\mathbf{X}$ , or, more generally, that

$$b'w_+(0) \perp \mathbf{X} \quad \text{for all } b \in \mathbb{R}^m. \tag{8.6.10}$$

To this end, recall that  $b'w_+(0) \perp \mathbf{S}_+ = \mathbf{H}^-(w_+)$ , and thus  $b'w_+(0) \in \mathbf{N}^+$  (Proposition 7.4.6). However, since  $\mathbf{X}$  is a minimal splitting subspace,  $\mathbf{X} \perp \mathbf{N}^+$  (Theorem 7.6.4), and therefore (8.6.10) holds.

Consequently, (8.6.9) yields

$$a'Ax(0) = E^{\mathbf{X}}a'A_+x_+(0) = a'A_+x(0) \quad \text{for all } a \in \mathbb{R}^n,$$

and hence  $a'AP = a'A_+P$  for all  $a \in \mathbb{R}^n$ . Therefore, since  $P > 0$ ,  $A = A_+$ . Moreover, from (8.3.4) we have

$$E^{\mathbf{X}}b'y(0) = b'Cx(0).$$

However, in view of (8.6.10), we also have

$$E^{\mathbf{X}} b'y(0) = E^{\mathbf{X}} b'C_+x_+(0) = b'C_+x(0)$$

for all  $b \in \mathbb{R}^m$ , and therefore  $C = C_+$ . Finally, a symmetric argument using (8.3.11), yields  $\bar{C} = \bar{C}_-$ . However, taking  $\mathbf{X}$  to be  $\mathbf{X}_+$  in this derivation, we obtain  $\bar{C}_+ = \bar{C}_-$ , and hence  $\bar{C} = \bar{C}_+$ , as claimed.  $\square$

In view of Corollary 8.4.6, any minimal Markovian splitting subspace has the orthogonal decomposition

$$\mathbf{X} = \mathbf{X}_0 \oplus \mathbf{X}_\infty, \quad (8.6.11)$$

where the deterministic subspace  $\mathbf{X}_\infty$  equals  $\mathbf{Y}_\infty$ , defined by (8.4.21), for all  $\mathbf{X}$ , and the proper subspace  $\mathbf{X}_0$  varies. If  $y$  has a nontrivial deterministic part  $y_\infty$ ,  $n_0 := \dim \mathbf{Y}_\infty \neq 0$ . Therefore, to cover this case, we may select the bases uniformly so that they are adapted to the decomposition (8.6.11) as on page 204.

Given such a uniform choice of bases in  $\mathcal{M}$ , the triplet  $(A, C, \bar{C})$  is fixed (Theorem 8.6.3), and has the form

$$A = \begin{bmatrix} A_0 & 0 \\ 0 & A_\infty \end{bmatrix}, \quad [C_0 \quad C_\infty], \quad [\bar{C}_0 \quad \bar{C}_\infty], \quad (8.6.12)$$

where the  $n_0 \times n_0$  matrix  $A_0$  has all its eigenvalues in open unit disc and  $A_\infty$  all eigenvalues on the unit circle (Theorem 8.4.8). Also given

$$\Lambda_0 := E\{x(0)x(0)'\},$$

define, as in Section 6, the map  $M : \mathbb{R}^{n \times n} \rightarrow \mathbb{R}^{(n+m) \times (n+m)}$  by

$$M(P) = \begin{bmatrix} P - APA' & \bar{C}' - APC' \\ \bar{C} - CPA' & \Lambda_0 - CPC' \end{bmatrix}. \quad (8.6.13)$$

We are now in a position to state a more general version of some fundamental results from Chapter 6, now derived from basic geometric principles.

**Theorem 8.6.4.** *Let  $\mathcal{M}$  be a family of (equivalence classes) of  $n$ -dimensional minimal Markovian representations corresponding to a full-rank stationary random process  $y$ . Given a uniform choice of bases for  $\mathcal{M}$ , adapted to (8.6.11), let  $(A, C, \bar{C})$  be the corresponding matrices prescribed by Theorem 8.6.3, and let  $M$  be defined by (8.6.13). Then there is a one-one correspondence between  $\mathcal{M}$  and the set*

$$\mathcal{P} = \{P \in \mathbb{R}^{n \times n} \mid P' = P, M(P) \geq 0\} \quad (8.6.14)$$

that is order-preserving in the sense that  $P_1 \leq P_2$  if and only if  $M_1 \prec M_2$ . Under this correspondence

$$P := E\{x(0)x(0)'\}, \quad (8.6.15)$$

where  $x(0)$  is the uniformly chosen basis in the corresponding minimal Markovian splitting subspace.



8.6. Parameterization of finite-dimensional minimal Markovian representations 213

**Proof.** By Proposition 7.7.5 there is an order-preserving one-one correspondence between  $\mathcal{M}$  and the set  $\widehat{\mathcal{P}}$  of all covariances (8.6.14) corresponding to minimal Markovian splitting subspaces  $\mathbf{X}$ . Clearly,  $\widehat{\mathcal{P}} \subset \mathcal{P}$ . In fact, for any such  $\mathbf{X}$ , by Theorem 8.3.1,  $P := E\{x(0)x(0)'\}$  satisfies

$$M(P) = \begin{bmatrix} B \\ D \end{bmatrix} \begin{bmatrix} B \\ D \end{bmatrix}' \geq 0 \tag{8.6.16}$$

for some  $B, D$ . It remains to show that  $\mathcal{P} \subset \widehat{\mathcal{P}}$ ; i.e., to each  $P \in \mathcal{P}$  there is a minimal Markovian representation  $(\mathbb{H}, U, \mathbf{X})$  such that  $P := E\{x(0)x(0)'\}$  in the given uniform choice of bases.

To this end, first suppose that  $y$  is purely nondeterministic; i.e.,  $n_0 = 0$ . Then, for an arbitrary  $P \in \mathcal{P}$ , determine a pair  $(B, D)$  through a minimal factorization (8.6.16), and form

$$W(z) = C(zI - A)^{-1}B + D.$$

Then, if  $y$  is  $m$ -dimensional,  $W$  is  $m \times p$ , where  $p := \text{rank } M(P) \geq m$ . Moreover, since  $A$  has all its eigenvalues in the open unit disc  $\mathbb{D}$ ,  $W$  is analytic in the complement of  $\mathbb{D}$ . Following the notation of Section 4.2 (and a construction to be discussed in more detail in the next section), we define a  $p$ -dimensional generating process  $w$  via

$$d\hat{w} = W^*\Phi^{-1}d\hat{y} + d\hat{z},$$

where the stationary process  $z(t) := U_z^t z(0)$  is chosen to be orthogonal to  $\mathbf{H} := \mathbf{H}(y)$  and such that  $E\{d\hat{w}d\hat{w}^*\} = d\theta$ . Then we must have  $E\{d\hat{z}d\hat{z}^*\} = (1 - W^*\Phi^{-1}W)d\theta$ . Now, noting that  $WW^* = \Phi$ , we form  $Wd\hat{w} = d\hat{y} + Wd\hat{z}$  to obtain

$$d\hat{y} = Wd\hat{w}.$$

In fact,

$$E\{Wd\hat{z}d\hat{z}^*W^*\} = W(1 - W^*\Phi^{-1}W)W^*d\theta = (\Phi - \Phi)d\theta = 0.$$

Then,

$$y(t) = \int_{-\pi}^{\pi} W(e^{i\theta})d\hat{w}$$

has a realization (8.3.4) such that  $P := E\{x(0)x(0)'\}$  with Markovian splitting subspace  $\mathbf{X} := \{a'x(0) \mid a \in \mathbb{R}^n\}$ . Setting  $\mathbb{H} = \mathbf{H} \oplus \mathbf{H}(z)$  and  $U_w := U \times U_z$ ,  $(\mathbb{H}, U_w, \mathbf{X})$  is the requested Markovian representation of  $y$ .

Next, suppose  $n_0 \neq 0$ . Then, for an arbitrary  $P \in \mathcal{P}$ , determine  $B_0, B_\infty, D$  so that

$$M(P) = \begin{bmatrix} B_0 \\ B_\infty \\ D \end{bmatrix} \begin{bmatrix} B_0 \\ B_\infty \\ D \end{bmatrix}' \tag{8.6.17}$$

is a minimal factorization. Since  $|\lambda(A_\infty)| = 1$ , we have

$$B_\infty = 0, \quad P = \begin{bmatrix} P_0 & 0 \\ 0 & P_\infty \end{bmatrix}, \tag{8.6.18}$$

where  $P_\infty$  is  $n_0 \times n_0$ . In fact, assuming that

$$P = \begin{bmatrix} P_0 & Z' \\ Z & P_\infty \end{bmatrix},$$

(8.6.17) yields (among other things)

$$P_\infty = A_\infty P_\infty A'_\infty + B_\infty B'_\infty, \quad (8.6.19a)$$

$$Z = A_0 Z A'_\infty + B_0 B'_\infty. \quad (8.6.19b)$$

Since  $A_\infty$  has all its eigenvalues on the unit circle, then so does  $G := P_\infty^{1/2} A'_\infty P_\infty^{-1/2}$ . Therefore, for any eigenvector  $v$  of  $G$ , (8.6.19a) yields

$$|B'_\infty P_\infty^{-1/2} v|^2 = |v|^2 - |Gv|^2 = 0,$$

and hence  $B_\infty = 0$ , as claimed. Then, (8.6.17) yields

$$(1 - \lambda A_0) Z v = 0$$

for all eigenvectors  $v$  and corresponding eigenvalues  $\lambda$  of  $A_\infty$ . Since  $|\lambda| = 1$  and  $|\lambda(A_0)| < 1$ , the matrix  $Z$  must be zero. This establishes (8.6.18).

Since  $(C_\infty, A_\infty)$  is observable (Theorem 8.4.8),  $x_\infty(0)$  can be uniquely determined from

$$C_\infty A_\infty^k x_\infty(0) = y_\infty(k), \quad k = 0, 1, \dots, n_0 - 1.$$

Clearly  $P := E\{x_\infty(0)x_\infty(0)'\}$  satisfies (8.6.19a) with  $B_\infty = 0$ , and

$$\begin{cases} x_\infty(t+1) &= A_\infty x_\infty(t) \\ y_\infty(t) &= C_\infty x_\infty(t) \end{cases}$$

holds. From (8.6.17) we also have

$$M_0(P_0) = \begin{bmatrix} B_0 \\ D \end{bmatrix} \begin{bmatrix} B_0 \\ D \end{bmatrix}' \geq 0,$$

where  $M_0$  is defined as  $M$ , just exchanging  $A$ ,  $C$ ,  $\bar{C}$  and  $\Lambda_0$  by  $A_0$ ,  $C_0$ ,  $\bar{C}_0$  and  $E\{y_0 y_0'\}$ , respectively. Then we proceed precisely as in the purely nondeterministic case to define a generating process  $w$  and a stochastic system

$$\begin{cases} x_0(t+1) &= A_0 x_0(t) + B_0 w(t) \\ y_0(t) &= C_0 x_0(t) + D w(t) \end{cases}$$

with  $P_0 = E\{x_0(0)x_0(0)'\}$  and  $\mathbf{X}_0 = \{a'x_0(0) \mid a \in \mathbb{R}^{n-n_0}\}$ . Hence we have constructed a stochastic system (8.4.28) and thus a Markovian representation with the Markovian splitting subspace  $\mathbf{X} = \mathbf{X}_0 \oplus \mathbf{X}_\infty$ .  $\square$

## 8.7 The forward and backward systems (the general case)

Let  $(\mathbb{H}, U, \mathbf{X})$  be an arbitrary proper Markovian representation with splitting subspace  $\mathbf{X} \sim (\mathbf{S}, \bar{\mathbf{S}})$  and generating processes  $(w, \bar{w})$ . By Theorems 8.1.2 and 8.1.3,

$$\begin{cases} U\mathbf{X} \subset \mathbf{X} \oplus \mathbf{W} \\ \mathbf{Y} \subset \mathbf{X} \oplus \mathbf{W}, \end{cases} \quad (8.7.1)$$

where  $\mathbf{W} := U\mathbf{S} \ominus \mathbf{S} = \{a'w(0) \mid a \in \mathbb{R}^p\}$  and  $\mathbf{Y} := \{a'y(0) \mid a \in \mathbb{R}^m\}$ . We also have

$$\begin{cases} \mathbf{X} \subset U\mathbf{X} \oplus \bar{\mathbf{W}} \\ \mathbf{Y} \subset U\mathbf{X} \oplus \bar{\mathbf{W}}, \end{cases} \quad (8.7.2)$$

where  $\bar{\mathbf{W}} := \bar{\mathbf{S}} \ominus U^*\bar{\mathbf{S}} := \{a'\bar{w}(0) \mid a \in \mathbb{R}^{\bar{p}}\}$ .

Let  $\xi \in \mathbf{X} \subset \mathbf{S} = \mathbf{H}^-(w)$ . Then

$$\xi = \sum_{k=-\infty}^{-1} \sum_{j=1}^p a_{kj} w_j(k)$$

for some sequences  $(a_{-1,j}, a_{-2,j}, a_{-3,j}, \dots) \in \ell_2, j = 1, 2, \dots, p$ , and hence, taking  $t \geq 0$ ,

$$U^t \xi = \sum_{k=-\infty}^{-1} \sum_{j=1}^p a_{kj} w_j(k+t) = \sum_{k=-\infty}^{t-1} \sum_{j=1}^p a_{k-t,j} w_j(k). \quad (8.7.3)$$

Therefore, by Theorem 7.5.1,

$$U(\mathbf{X})^t \xi = \mathbf{E}^{\mathbf{H}^-(w)} U^t \xi = \sum_{k=-\infty}^{t-1} \sum_{j=1}^p a_{k-t,j} w_j(k),$$

and consequently

$$a_{-t,j} = \langle U(\mathbf{X})^{t-1} \xi, w_j(-1) \rangle = \langle U(\mathbf{X})^{t-1} \xi, \mathbf{E}^{\mathbf{X}} w_j(-1) \rangle_{\mathbf{X}},$$

or, equivalently,

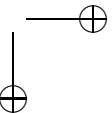
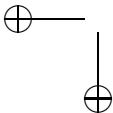
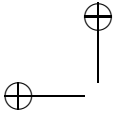
$$a_{-t,j} = \langle \xi, [U(\mathbf{X})^*]^{t-1} \mathbf{E}^{\mathbf{X}} w_j(-1) \rangle_{\mathbf{X}}, \quad (8.7.4)$$

where  $\langle \cdot, \cdot \rangle_{\mathbf{X}}$  is the inner product in the Hilbert space  $\mathbf{X}$ . Then (8.7.3) and (8.7.4) imply that, for any  $\xi \in \mathbf{X}$ ,

$$U^t \xi = \sum_{k=-\infty}^{t-1} \sum_{j=1}^p \langle \xi, [U(\mathbf{X})^*]^{t-k-1} \mathbf{E}^{\mathbf{X}} w_j(-1) \rangle_{\mathbf{X}} w_j(k). \quad (8.7.5)$$

Now, in view of (8.7.1),

$$y_i(0) = \mathbf{E}^{\mathbf{X}} y_i(0) + \mathbf{E}^{\mathbf{W}} y_i(0)$$



for  $i = 1, 2, \dots, m$ , and hence (8.7.5) yields

$$y_i(t) = \sum_{k=-\infty}^{t-1} \sum_{j=1}^p \langle \mathbf{E}^{\mathbf{X}} y_i(0), [U(\mathbf{X})^*]^{t-k-1} \mathbf{E}^{\mathbf{X}} w_j(-1) \rangle_{\mathbf{X}} w_j(k) + \sum_{j=1}^p d_{ij} w_j(t), \quad (8.7.6)$$

where

$$d_{ij} := \langle y_i(0), w_j(0) \rangle. \quad (8.7.7)$$

From (8.7.6) we see that the splitting subspace  $\mathbf{X}$  may also serve as a state space for a (possibly infinite-dimensional) linear stochastic system. To avoid any confusion created by this, we introduce an isomorphic copy of  $\mathbf{X}$ ,  $\mathcal{X}$ , via an isomorphism  $T : \mathbf{X} \rightarrow \mathcal{X}$  such that  $\langle T\xi, T\eta \rangle_{\mathcal{X}} = \langle \xi, \eta \rangle_{\mathbf{X}}$ . In the next chapter we introduce a suitable candidate for  $\mathcal{X}$  in the Hardy space  $H^2$ , namely a coinvariant subspace  $H(K)$  to be defined there. If  $\mathbf{X}$  has a finite dimension  $n$ , we may of course choose  $\mathcal{X} := \mathbb{R}^n$ . Then, in view of (8.7.6),

$$y(t) = \sum_{k=-\infty}^{t-1} CA^{t-k-1}Bw(k) + Dw(t) \quad (8.7.8)$$

$$:= \sum_{k=-\infty}^{t-1} \sum_{j=1}^p CA^{t-k-1}Be_jw_j(k) + Dw(t), \quad (8.7.9)$$

where  $A : \mathcal{X} \rightarrow \mathcal{X}$ ,  $B : \mathbb{R}^p \rightarrow \mathcal{X}$  and  $C : \mathcal{X} \rightarrow \mathbb{R}^m$  are bounded linear operators defined via

$$A = TU(\mathbf{X})^*T^{-1}, \quad (8.7.10a)$$

$$Ba = \sum_{j=1}^p a_j T \mathbf{E}^{\mathbf{X}} w_j(-1) = T \mathbf{E}^{\mathbf{X}} a'w(-1), \quad (8.7.10b)$$

$$(Cx)_i = \langle T \mathbf{E}^{\mathbf{X}} y_i(0), x \rangle_{\mathcal{X}}, \quad (8.7.10c)$$

$D$  is the matrix defined by (8.7.7), and  $e_j$  is the  $j$ :th axis vector in  $\mathbb{R}^p$ . Note that in (8.7.10) we have given two equivalent expressions of the operator  $B$ . The first is meant to emphasize the fact that  $B$  acts on the vector structure of  $w$  in (8.7.12) and not on  $w$  as a random vector. By Theorem 7.5.1,  $A^t$  tends strongly to zero as  $t \rightarrow \infty$  if  $\bigcap_{t=0}^{\infty} U^t \bar{\mathbf{S}} = 0$ ; in particular, if  $\mathbf{X}$  is proper.

Now, from (8.7.6) we also have, for any  $\xi \in \mathbf{X}$ ,

$$U^t \xi = \langle T\xi, x(t) \rangle_{\mathcal{X}} := \sum_{j=1}^p \langle T\xi, \sum_{k=-\infty}^{t-1} A^{t-k-1} Be_j \rangle_{\mathcal{X}} w_j(k), \quad (8.7.11)$$

where

$$x(t) = \sum_{k=-\infty}^{t-1} A^{t-k-1} Bw(k) \quad (8.7.12)$$

is defined in the weak topology of  $\mathcal{X}$ . Thus we regard the object (8.7.12) as an  $\mathcal{X}$ -valued random vector defined in the weak sense of [52] or [39]. It lies outside of

the scope of this book to go deeply into the theory of infinite-dimensional random processes, so we refer the reader to [52, 39] for a full study of Hilbert-space valued random processes. Here it suffices to say that  $x(0)$  is an exact generator of  $\mathbf{X}$  in the sense that  $\langle T\xi, x(0) \rangle_{\mathcal{X}} = \xi$  and

$$\overline{\text{span}}\{\langle f, x(0) \rangle_{\mathcal{X}} \mid f \in \mathcal{X}\} = \mathbf{X}. \tag{8.7.13}$$

The object  $\{x(t) \mid t \in \mathbb{Z}\}$  can be given the meaning of a weak  $\mathcal{X}$ -valued random process [52, 39], in terms of which (8.7.8) may be written

$$\begin{cases} x(t+1) = Ax(t) + Bw(t) \\ y(t) = Cx(t) + Dw(t). \end{cases} \tag{8.7.14}$$

In view of (8.7.11),  $\{x(t) \mid t \in \mathbb{Z}\}$  is stationary in the (weak) sense that

$$\langle T\xi, x(t) \rangle_{\mathcal{X}} = U^t \langle T\xi, x(0) \rangle_{\mathcal{X}}. \tag{8.7.15}$$

Defining the covariance operator  $P : \mathcal{X} \rightarrow \mathcal{X}$  of  $x(0)$  via the bilinear form

$$\langle f, Pg \rangle_{\mathcal{X}} = \mathbb{E} \{ \langle f, x(0) \rangle_{\mathcal{X}} \langle g, x(0) \rangle_{\mathcal{X}} \}, \tag{8.7.16}$$

we obtain the following representation.

**Proposition 8.7.1.** *Then the operator  $P : \mathcal{X} \rightarrow \mathcal{X}$  is given by*

$$P = \sum_{k=0}^{\infty} A^k B B^* (A^*)^k, \tag{8.7.17}$$

where  $A : \mathcal{X} \rightarrow \mathcal{X}$  and  $B : \mathbb{R}^p \rightarrow \mathcal{X}$  are defined by (8.7.10) and the sum is defined in the weak operator topology<sup>22</sup>; i.e.,  $P$  satisfies the Lyapunov equation

$$P = APA^* + BB^*. \tag{8.7.18}$$

Moreover,  $(A, B)$  is exactly reachable.

**Proof.** Since  $x(0) = \sum_{k=0}^{\infty} A^k B w(1-k)$ , we have

$$\begin{aligned} \mathbb{E} \{ \langle f, x(0) \rangle_{\mathcal{X}} \langle g, x(0) \rangle_{\mathcal{X}} \} &= \sum_{k=0}^{\infty} \sum_{j=1}^p \langle B^* (A^*)^k f, e_j \rangle_{\mathcal{X}} \langle B^* (A^*)^k g, e_j \rangle_{\mathcal{X}} \\ &= \sum_{k=0}^{\infty} \langle B^* (A^*)^k f, B^* (A^*)^k g \rangle_{\mathbb{R}^p} \\ &= \sum_{k=0}^{\infty} \langle f, A^k B B^* (A^*)^k g \rangle_{\mathcal{X}}, \end{aligned}$$

<sup>22</sup> $P_k \rightarrow P$  in the weak operator topology if  $\langle f, P_k g \rangle_{\mathcal{X}} \rightarrow \langle f, P g \rangle_{\mathcal{X}}$  for all  $f, g \in \mathcal{X}$ .

which establishes (8.7.17), from which (8.7.18) follow readily. To see that  $(A, B)$  is exactly reachable, note that, in view of (7.5.6) and (7.4.9),

$$\begin{aligned} T^{-1}A^tBa &= E^{\bar{\mathbf{S}}}U^{-t}E^{\mathbf{X}}a'w(-1) \\ &= E^{\bar{\mathbf{S}}}a'w(-t-1) - E^{\bar{\mathbf{S}}}U^{-t}E^{\mathbf{S}^\perp}a'w(-1) - E^{\bar{\mathbf{S}}}U^{-t}E^{\bar{\mathbf{S}}^\perp}a'w(-1), \end{aligned}$$

where the last two terms are zero since  $a'w(-1) \in \mathbf{S}$  and  $U^{-t}\bar{\mathbf{S}}^\perp \subset \bar{\mathbf{S}}^\perp$ . Therefore, since  $\bigvee_{t=0}^{\infty}\{a'w(-1) \mid a \in \mathbb{R}^p\} = \mathbf{H}^-(w) = \mathbf{S}$ , we have

$$\bigvee_{t=0}^{\infty} \text{span}_{a \in \mathbb{R}} T^{-1}A^tBa = E^{\bar{\mathbf{S}}}\mathbf{S} = \mathbf{X}$$

by the splitting property, and hence

$$\bigvee_{t=0}^{\infty} \text{Im}A^tB = \mathcal{X},$$

as claimed.  $\square$

In the same way, we can construct a backward system using (8.7.2). Let  $\xi \in U\mathbf{X} \subset \bar{\mathbf{S}} = U\mathbf{H}^+(\bar{w})$ . Then, in view of (8.7.2), we have a completely symmetric situation to the one in the forward setting, except for the shift of the state. In fact, in analogy with (8.7.5), for any  $\xi \in U\mathbf{X}$ , we have

$$U^t\xi = \sum_{k=t+1}^{\infty} \sum_{j=1}^p \langle \xi, [U(U\mathbf{X})]^{-t-1+k} E^{U\mathbf{X}} \bar{w}_j(1) \rangle_{U\mathbf{X}} \bar{w}_j(k) \quad (8.7.19)$$

$$= \sum_{k=t+1}^{\infty} \sum_{j=1}^p \langle U^*\xi, [U(\mathbf{X})]^{k-t-1} E^{\mathbf{X}} \bar{w}_j(0) \rangle_{\mathbf{X}} \bar{w}_j(k). \quad (8.7.20)$$

Moreover,

$$\begin{aligned} y_i(0) &= E^{U\mathbf{X}}y_i(0) + E^{\bar{\mathbf{W}}}y_i(0) \\ &= UE^{\mathbf{X}}y_i(-1) + \bar{D}\bar{w}(t), \end{aligned}$$

where  $\bar{D}$  is the matrix defined by  $\bar{d}_{ij} := \langle y_i(0), \bar{w}_j(0) \rangle$ , and hence

$$y(t) = \sum_{k=t+1}^{\infty} \bar{C}\bar{A}^{k-t-1}\bar{B}\bar{w}(k) + \bar{D}\bar{w}(t), \quad (8.7.21)$$

where  $\bar{A} : \mathcal{X} \rightarrow \mathcal{X}$ ,  $\bar{B} : \mathbb{R}^p \rightarrow \mathcal{X}$  and  $\bar{C} : \mathcal{X} \rightarrow \mathbb{R}^m$  are bounded linear operators defined via

$$\bar{A} = TU(\mathbf{X})T^{-1}, \quad (8.7.22a)$$

$$\bar{B}a = \sum_{j=1}^{\bar{p}} a_j T E^{\mathbf{X}} \bar{w}_j(0) = T E^{\mathbf{X}} a' \bar{w}(0), \quad (8.7.22b)$$

$$(\bar{C}x)_i = \langle T E^{\mathbf{X}} y_i(-1), x \rangle_{\mathcal{X}}. \tag{8.7.22c}$$

Clearly,

$$\bar{A} = A^*. \tag{8.7.23}$$

Moreover, by Theorem 7.5.1,  $\bar{A}^t$  tends strongly to zero as  $t \rightarrow \infty$  if and only if  $\bigcap_{t=-\infty}^0 U^t \mathbf{S} = 0$ ; in particular, if  $\mathbf{X}$  is proper. Precisely as in Proposition 8.7.1 we can show that  $(A^*, \bar{B})$  is exactly reachable.

Analogously with the forward setting, (8.7.21) can be written

$$\begin{cases} \bar{x}(t-1) &= \bar{A}\bar{x}(t) + \bar{B}\bar{w}(t) \\ y(t) &= \bar{C}\bar{x}(t) + \bar{D}\bar{w}(t) \end{cases}, \tag{8.7.24}$$

where

$$\bar{x}(t) = \sum_{k=t+1}^{\infty} (A^*)^{k-t-1} \bar{B}\bar{w}(k) \tag{8.7.25}$$

is a  $\mathcal{X}$ -valued (weak) random process with the properties

$$\langle T\xi, \bar{x}(-1) \rangle_{\mathcal{X}} = \xi \tag{8.7.26}$$

and

$$\overline{\text{span}}\{\langle f, \bar{x}(-1) \rangle_{\mathcal{X}} \mid f \in \mathcal{X}\} = \mathbf{X}. \tag{8.7.27}$$

Moreover, the covariance operator  $\bar{P} : \mathcal{X} \rightarrow \mathcal{X}$  given by

$$\langle f, \bar{P}g \rangle_{\mathcal{X}} = E \{ \langle f, \bar{x}(-1) \rangle_{\mathcal{X}} \langle g, \bar{x}(-1) \rangle_{\mathcal{X}} \}, \tag{8.7.28}$$

takes the form

$$\bar{P} = \sum_{k=0}^{\infty} (A^*)^k \bar{B}\bar{B}^* A^k;$$

i.e.,  $\bar{P}$  satisfies the Lyapunov equation

$$\bar{P} = A^* \bar{P} A + \bar{B}\bar{B}^*. \tag{8.7.29}$$

**Proposition 8.7.2.** *The covariance operators defined by (8.7.16) and (8.7.28) both equal the identity, i.e.,  $P = \bar{P} = I$ . Moreover,  $\bar{x}(-1) = x(0)$ .*

**Proof.** By (8.7.11), we have  $\langle f, x(0) \rangle_{\mathcal{X}} = T^{-1}f$ , and consequently

$$E \{ \langle f, x(0) \rangle_{\mathcal{X}} \langle g, x(0) \rangle_{\mathcal{X}} \} = E \{ T^{-1}f, T^{-1}g \} = \langle f, g \rangle_{\mathcal{X}},$$

establishing that  $P = I$ . In the same way, we see that  $\bar{P} = I$ . From (8.7.11) and (8.7.26), we have

$$\langle f, x(0) \rangle_{\mathcal{X}} = T^{-1}f = \langle f, \bar{x}(-1) \rangle_{\mathcal{X}}$$

for all  $f \in \mathcal{X}$ , and hence we must have  $\bar{x}(-1) = x(0)$ , as claimed.  $\square$

**Remark 8.7.3.** The coordinate-free results of Proposition 8.7.2 are entirely consistent with the coordinate-dependent results of Sections 8.3 and 6.3. Indeed, when  $\dim \mathbf{X} = n < \infty$ , we may take  $\mathcal{X} = \mathbb{R}^n$  and represent the operators by matrices with respect to the two bases,  $(\xi_1, \xi_2, \dots, \xi_n)$  and  $(\bar{\xi}_1, \bar{\xi}_2, \dots, \bar{\xi}_n)$  introduced in Section 8.3. More precisely, take

$$x(0) = \begin{bmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_n \end{bmatrix} \quad \text{and} \quad \bar{x}(-1) = \begin{bmatrix} \bar{\xi}_1 \\ \bar{\xi}_2 \\ \vdots \\ \bar{\xi}_n \end{bmatrix}.$$

Then, since now  $\langle a, x(0) \rangle_{\mathcal{X}} = a'x(0)$ , we have

$$\mathbf{E}\{\langle a, x(0) \rangle_{\mathcal{X}} \langle b, x(0) \rangle_{\mathcal{X}}\} = a' \mathbf{E}\{x(0)x(0)\}b = a'Pb$$

where, with a slight misuse of notation, the matrix  $P$  is defined by (8.3.3) and is the matrix representation of the operator  $P$  in the appropriate basis. Likewise, we can see that the matrix  $\bar{P}$ , defined by the first of the two equations (8.3.10), is the matrix representation of the operator  $\bar{P}$  in the appropriate basis. Finally, (8.7.11) and (8.7.26) yield

$$a'x(0) = \xi = b'\bar{x}(-1),$$

for all  $\xi \in \mathbf{X}$ , where  $a \in \mathcal{X}$  and  $b \in \mathcal{X}$  correspond to  $\xi$  under the two different bases in  $\mathbf{X}$  of  $x(0)$  and  $\bar{x}(-1)$ , respectively. Then, since the two bases are dual in the sense of (8.3.6),  $b = Pa$ , and hence (8.3.9) and  $\bar{P} = P^{-1}$  follow.

We are now in a position to formulate an operator version of Theorem 8.3.1.

**Theorem 8.7.4.** *To each proper Markovian representation  $(\mathbb{H}, U, \mathbf{X})$  there is a pair of dual stochastic realizations, consisting of a forward system (8.7.14) and a backward system (8.7.24) with  $\mathcal{X}$ -valued state processes having the property*

$$\overline{\text{span}}\{\langle f, x(0) \rangle_{\mathcal{X}} \mid f \in \mathcal{X}\} = \mathbf{X} = \overline{\text{span}}\{\langle f, x(-1) \rangle_{\mathcal{X}} \mid f \in \mathcal{X}\} \quad (8.7.30)$$

The forward and backward systems are connected via the relations

$$\bar{A} = A^*, \quad \bar{C} = CA^* + DB^* \quad (8.7.31)$$

and

$$\bar{x}(t-1) = x(t). \quad (8.7.32)$$

Moreover, the splitting subspace  $\mathbf{X}$  is observable if and only if

$$\bigcap_{t=0}^{\infty} \ker CA^t = 0, \quad (8.7.33)$$

i.e.,  $(C, A)$  is (completely) observable; constructible if and only if

$$\bigcap_{t=0}^{\infty} \ker \bar{C}(A^*)^t = 0, \quad (8.7.34)$$



i.e.,  $(\bar{C}, A^*)$  is (completely) observable; and minimal if and only if both  $(C, A)$  and  $(\bar{C}, A^*)$  are observable. Finally,

$$\Lambda_t := \mathbb{E}\{y(t+k)y(t)'\} = \begin{cases} CA^{t-1}\bar{C}^* & \text{for } t > 0; \\ CC^* + DD^* & \text{for } t = 0; \\ \bar{C}(A')^{|t|-1}C^* & \text{for } t < 0. \end{cases} \quad (8.7.35)$$

**Proof.** It only remains to prove the second of relations (8.7.31), the statements about observability and constructibility, and (8.7.35). To prove that  $C = CA^* + DB^*$ , recall from (8.7.1) that  $\mathbf{Y} \subset \mathbf{X} \oplus \mathbf{W}$ , which, in particular implies that

$$y_i(0) = \mathbb{E}^{\mathbf{X}} y_i(0) + [Dw(0)]_i. \quad (8.7.36)$$

Therefore, for any  $\xi \in \mathbf{X}$ ,

$$\langle y_i(0), U\xi \rangle = \langle \mathbb{E}^{\mathbf{X}} y_i(0), U\xi \rangle + \langle [Dw(0)]_i, U\xi \rangle,$$

or, equivalently,

$$\langle \mathbb{E}^{\mathbf{X}} y_i(-1), \xi \rangle_{\mathbf{X}} = \langle \mathbb{E}^{\mathbf{X}} y_i(0), U(\mathbf{X})\xi \rangle_{\mathbf{X}} + \langle [Dw(-1)]_i, \xi \rangle;$$

i.e., setting  $f := T\xi$ ,

$$\langle T\mathbb{E}^{\mathbf{X}} y_i(-1), f \rangle_{\mathbf{X}} = \langle T\mathbb{E}^{\mathbf{X}} y_i(0), A^*f \rangle_{\mathbf{X}} + \langle [Dw(-1)]_i, \xi \rangle,$$

which is the same as

$$\bar{C}f = CA^*f + DB^*f,$$

valid for all  $f \in \mathcal{X}$ . To see this, observe that

$$\langle f, Ba \rangle_{\mathbf{X}} = \langle \xi, \mathbb{E}^{\mathbf{X}} a'w(-1) \rangle_{\mathbf{X}} = a' \mathbb{E}\{\xi w(-1)\},$$

which implies that  $B^*f = \mathbb{E}\{\xi w(-1)\}$ . This establishes that  $C = CA^* + DB^*$ .

The proof of the observability statement goes much along the lines of the finite-dimensional result. In fact, since  $y_i(0) \in \mathbf{H}^+ \subset \bar{\mathbf{S}}$ ,

$$[CA^t f]_i = \langle \mathbb{E}^{\mathbf{X}} y_i(0), [U(\mathbf{X})^*]^t \xi \rangle_{\mathbf{X}} = \langle y_i(0), \mathbb{E}^{\bar{\mathbf{S}}} U^{-t} \xi \rangle_{\mathbf{X}} = \langle y_i(t), \xi \rangle_{\mathbf{X}},$$

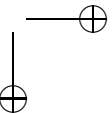
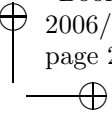
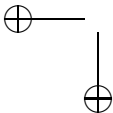
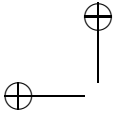
which is zero for  $i = 1, 2, \dots, m$  and all  $t \geq 0$  if and only if  $\mathbf{X} \cap (\mathbf{H}^+)^{\perp} = 0$ ; i.e.,  $\mathbf{X}$  is observable. This proves the statement on observability. The statement on constructibility follows by symmetry.

To prove (8.7.35), first suppose  $t > 0$ . Then, since  $a'y(0) \perp \mathbf{H}^+(w)$ , (8.7.36) yields

$$\Lambda_t a = \langle \mathbb{E}^{\mathbf{X}} y_i(t), \mathbb{E}^{\mathbf{X}} a'y(0) \rangle_{\mathbf{X}} = \langle U(\mathbf{X})^t \mathbb{E}^{\mathbf{X}} y_i(0), U(\mathbf{X}) \mathbb{E}^{\mathbf{X}} a'y(-1) \rangle_{\mathbf{X}}.$$

Here we have used the fact that, in view of Theorem 7.2.4,

$$y_i(t) = U^t \mathbb{E}^{\bar{\mathbf{S}}^{\perp}} y_i(0) + U^t \mathbb{E}^{\mathbf{X}} y_i(0) + U^t \mathbb{E}^{\bar{\mathbf{S}}^{\perp}} y_i(0),$$



where the first term is zero since  $y_i(0) \in \mathbf{H}^+ \perp (\mathbf{H}^+)^{\perp} \supset \bar{\mathbf{S}}^{\perp}$ , and the last term belongs to  $\mathbf{S}^{\perp} \perp \mathbf{X}$ ; and an analogous argument for the second element in the inner product. Hence

$$\Lambda_t a = \langle U(\mathbf{X})^{t-1} \mathbf{E}^{\mathbf{X}} y_i(0), \mathbf{E}^{\mathbf{X}} a'y(-1) \rangle_{\mathbf{X}} = CA^{t-1} \bar{C}^* a,$$

as claimed. In fact,

$$\langle \bar{C}^* a, f \rangle_x = a' \bar{C} f = \langle T \mathbf{E}^{\mathbf{X}} a'y(-1), f \rangle_x.$$

The proof of the case  $t < 0$  is analogous. Finally, by (8.7.36),

$$\begin{aligned} \mathbf{E}\{y_i(0)y(0)'\}a &= \langle \mathbf{E}^{\mathbf{X}} y_i(0), \mathbf{E}^{\mathbf{X}} a'y(0) \rangle_{\mathbf{X}} + \mathbf{E}\{[Dw(0)]_i y(0)'a\} \\ &= [CC^* a + DD^* a]_i, \end{aligned}$$

completing the proof of (8.7.35).  $\square$

In view of Propositions 8.7.1 and 8.7.2, we have the following equations

$$I = AA^* + BB^*, \tag{8.7.37a}$$

$$\bar{C} = CA^* + DB^*, \tag{8.7.37b}$$

$$\Lambda_0 = CC^* + DD^*, \tag{8.7.37c}$$

which, in the case that  $(\mathbb{H}, U, \mathbf{X})$  is minimal, look like an infinite-dimensional version of the positive-real-lemma equations (6.8.23), except that the operators  $A$ ,  $C$  and  $\bar{C}$  clearly depend on the particular choice of  $\mathbf{X}$ , but  $P = I$  does not. This raises the question of what is the relation between the triplets  $(A, C, \bar{C})$  of different minimal Markovian representations.

### State-space isomorphisms and the infinite-dimensional positive-real-lemma equations

The covariance of  $\{y(t)\}_{t \in \mathbb{Z}}$  is given by (8.7.35), where the triplet  $(A, C, \bar{C})$  corresponds to any Markovian representation  $(\mathbb{H}, U, \mathbf{X})$  of  $y$ . In particular, for  $t > 0$ ,

$$\Lambda_t = CA^t \bar{C}^* = C_+ A_+^t \bar{C}_+^*,$$

where  $(A_+, C_+, \bar{C}_+)$  corresponds to  $(\mathbf{X}_+, U, \mathbf{H})$ . More generally, if  $(\mathbb{H}, U, \mathbf{X})$  is minimal, these operators are related in the the following way.

**Theorem 8.7.5.** *Let  $(\mathbb{H}, U, \mathbf{X})$  be a minimal Markovian representation with  $(A, C, \bar{C})$  be defined as in (8.7.10) and (8.7.22) and  $\mathcal{X}$  as defined on page 216. Let  $(A_+, C_+, \bar{C}_+)$ ,  $\mathcal{X}_+$  and  $(A_-, C_-, \bar{C}_-)$ ,  $\mathcal{X}_-$  be corresponding quantities related to  $(\mathbf{X}_+, U, \mathbf{H})$  and  $(\mathbf{X}_-, U, \mathbf{H})$ , respectively.. Then there are quasi-invertible linear operators  $\Omega : \mathcal{X} \rightarrow$*

$\mathcal{X}_+$  and  $\bar{\Omega} : \mathcal{X} \rightarrow \mathcal{X}_-$  such that the diagrams

$$\begin{array}{ccc}
 \mathbb{R}^m & & \mathbb{R}^m \\
 \bar{C}^* \swarrow & & \searrow \bar{C}_+^* \\
 \mathcal{X} & \xrightarrow{\Omega} & \mathcal{X}_+ \\
 A^t \downarrow & & \downarrow A_+^t \\
 \mathcal{X} & \xrightarrow{\Omega} & \mathcal{X}_+ \\
 C \searrow & & \swarrow C_+ \\
 \mathbb{R}^m & & \mathbb{R}^m
 \end{array}
 \qquad
 \begin{array}{ccc}
 \mathbb{R}^m & & \mathbb{R}^m \\
 C^* \swarrow & & \searrow C_-^* \\
 \mathcal{X} & \xrightarrow{\bar{\Omega}} & \mathcal{X}_- \\
 (A^*)^t \downarrow & & \downarrow (A_-^*)^t \\
 \mathcal{X} & \xrightarrow{\bar{\Omega}} & \mathcal{X}_- \\
 \bar{C} \searrow & & \swarrow \bar{C}_- \\
 \mathbb{R}^m & & \mathbb{R}^m
 \end{array}$$

commute.

In Section 9.2 we actually prove that the operators  $A$  of all minimal Markovian representations are actually quasi-equivalent (Corollary 9.2.14).

**Proof.** It follows from Corollary 7.6.6 that

$$\Omega A^t = A_+^t \Omega,$$

where  $\Omega := T_+ \hat{\Omega} T^*$  is quasi-invertible. Here  $T$  is as defined on page 216 and  $T_+$  is the corresponding isomorphism from  $\mathbf{X}_+$  and  $\mathcal{X}_+$ . This establishes the square part of the first commutative diagram. To establish the bottom part form, for any  $f \in \mathcal{X}$  and the corresponding  $\xi := T^{-1}f \in \mathbf{X}$ ,

$$(C_+ \Omega f)_i = \langle T_+ E^{\mathbf{X}^+} y_i(0), T_+ E^{\mathbf{X}^+} \xi \rangle_{\mathcal{X}} = \langle E^{\mathbf{X}^+} y_i(0), E^{\mathbf{X}^+} \xi \rangle_{\mathbf{X}^+} = \langle y_i(0), E^{\mathbf{X}^+} \xi \rangle.$$

Then, since  $\mathbf{H} = (\mathbf{H}^+)^{\perp} \oplus \mathbf{X}_+ \oplus \mathbf{N}^+$  and  $X \perp \mathbf{N}^+$  (Corollary 7.4.14) and  $y_i(0) \in \mathbf{H}^+ \perp (\mathbf{H}^+)^{\perp}$ , we have

$$(C_+ \Omega f)_i = \langle y_i(0), \xi \rangle = \langle E^{\mathbf{X}} y_i(0), \xi \rangle = (Cf)_i.$$

In the same way, with  $\xi \in \mathbf{X}_+$  and  $f = T_+ \xi$ ,

$$(\bar{C} \Omega^* f)_i = \langle E^{\mathbf{X}} y_i(-1), E^{\mathbf{X}} \xi \rangle_{\mathbf{X}} = \langle E^{\mathbf{X}} y_i(-1), \xi \rangle.$$

Hence

$$(\bar{C} \Omega^* f)_i = \langle E^{\mathbf{X}^+} E^{\mathbf{X}} y_i(-1), \xi \rangle_{\mathbf{X}^+} = \langle E^{\mathbf{X}^+} y_i(-1), \xi \rangle_{\mathbf{X}^+} = (\bar{C}_+ f)_i,$$

since  $\mathbf{H}^- \perp \mathbf{X}_+ \mid \mathbf{X}$  (Proposition 2.4.2(vi)); i.e.,  $\bar{C}_+ = \bar{C} \Omega^*$ , or, equivalently,  $\bar{C}_+^* = \Omega \bar{C}^*$ . This establishes the top part of the diagram.

A symmetric argument proves that there is a quasi-invertible  $\bar{\Omega}$  such that the right diagram commutes.  $\square$

In analogy with the finite-dimension setup in Section 8.3 and Chapter 6, we can now reformulate the forward system (8.7.14) of any minimal Markovian representation in terms of  $(A_+, C_+, \bar{C}_+)$ .

**Corollary 8.7.6.** *Let  $(\mathbb{H}, U, \mathbf{X})$  be a minimal Markovian representation. Then there is an  $\mathcal{X}_+$ -valued (weak) stationary random process  $\{x(t)\}_{t \in \mathbb{Z}}$  satisfying*

$$\overline{\text{span}}\{\langle f, x(0) \rangle_{\mathcal{X}_+} \mid f \in \mathcal{X}_+\} = \mathbf{X}.$$

The process  $x$  is generated by the forward system

$$\begin{cases} x(t+1) = A_+x(t) + \hat{B}w(t) \\ y(t) = C_+x(t) + Dw(t) \end{cases}$$

with  $\Omega B$ , where  $\Omega$  is the quasi-invertible map defined in Theorem 8.7.5. The covariance operator  $P : \mathcal{X}_+ \rightarrow \mathcal{X}_+$ , defined via the bilinear form

$$\langle f, Pg \rangle_{\mathcal{X}_+} = \mathbb{E} \{ \langle f, x(0) \rangle_{\mathcal{X}_+} \langle g, x(0) \rangle_{\mathcal{X}_+} \},$$

is given by

$$P = \Omega \Omega^* \leq P_+ = I, \tag{8.7.38}$$

and satisfies the operator positive-real-lemma equations

$$P = A_+ P A_+^* + \hat{B} \hat{B}^*, \tag{8.7.39a}$$

$$\bar{C}_+ = C_+ A_+^* + D \hat{B}^*, \tag{8.7.39b}$$

$$\Lambda_0 = C_+ P C_+^* + D D^*. \tag{8.7.39c}$$

Naturally, there is also a backward version of this corollary, obtained by the obvious substitutions.

**Proof.** The proof is immediate by merely applying the transformations  $\Omega A = A_+ \Omega$ ,  $C = C_+ \Omega$  and  $\bar{C} \Omega^* = \bar{C}_+$  (Theorem 8.7.5) to the corresponding equations obtained earlier in this section. In particular, (8.7.39) is obtain in this way from (8.7.37). Taking  $\xi \in \mathbf{X}_+$  and  $f = T_+^{-1} \xi$ , we have

$$\langle f, Pf \rangle_{\mathcal{X}_+} = \|\Omega^* f\|_{\mathcal{X}_+}^2 = \|\mathbb{E}^{\mathbf{X}} \xi\|_{\mathbf{X}_+}^2 = \|\xi\|_{\mathbf{X}_+}^2 = \|f\|_{\mathcal{X}_+}^2,$$

establishing that  $P \leq I$ , as claimed.  $\square$

Since  $\Omega$  is in general only quasi-invertible, this as far as we can go in generalizing the finite-dimensional results of Section 9.2. If  $\Omega$  is invertible, we can also establish that  $\bar{P} := P^{-1}$  is the covariance operator of the corresponding backward system. We leave this as an exercise for the reader.

## 8.8 Regularity of Markovian representations

In Chapter 6 we introduced the regularity condition (6.9.1), needed in formulating certain results in terms of the Algebraic Riccati Equation, a topic to which we

return in Chapter ???. We are now in a position to formulate general geometric conditions for regularity.

To this end, given a Markovian representation  $(\mathbb{H}, U, \mathbf{X})$  with  $\mathbf{X} \sim (\mathbf{S}, \bar{\mathbf{S}})$ , we introduce the corresponding forward and backward *error space* as

$$\mathbf{Z} := \mathbf{S} \ominus \mathbf{H}^- = \mathbf{S} \cap (\mathbf{H}^-)^\perp, \quad \text{and} \quad \bar{\mathbf{Z}} := \bar{\mathbf{S}} \ominus \mathbf{H}^+ = \bar{\mathbf{S}} \cap (\mathbf{H}^+)^\perp, \quad (8.8.1)$$

respectively, where the orthogonal complements  $(\mathbf{H}^-)^\perp$  and  $(\mathbf{H}^+)^\perp$  are with respect to the ambient space  $\mathbb{H}$ . The subspaces  $\mathbf{Z}$  and  $\bar{\mathbf{Z}}$  are not splitting subspaces, but they are the intersections of perpendicularly intersecting pairs of invariant subspaces, namely  $(\mathbf{S}, (\mathbf{H}^-)^\perp)$  and  $(\bar{\mathbf{S}}, (\mathbf{H}^+)^\perp)$ , respectively (Theorem 7.2.4). In fact,  $\mathbf{Z}^- \perp \mathbf{Z}^+ \mid \mathbf{Z}$  and  $\bar{\mathbf{Z}}^- \perp \bar{\mathbf{Z}}^+ \mid \bar{\mathbf{Z}}$ , so both  $\mathbf{Z}$  and  $\bar{\mathbf{Z}}$  are Markovian. Therefore, we can proceed as in Section 8.7 to define a  $\mathcal{Z}$ -valued weak stationary random process  $\{z(t)\}_{\mathcal{Z}}$  such that

$$\overline{\text{span}}\{\langle f, z(0) \rangle_{\mathcal{Z}} \mid f \in \mathcal{Z}\} = \mathbf{Z},$$

where  $\mathcal{Z}$  is an isomorphic copy of  $\mathbf{Z}$ . Likewise, we can define a  $\mathcal{Z}$ -valued process  $\{\bar{z}(t)\}_{\bar{\mathcal{Z}}}$  such that

$$\overline{\text{span}}\{\langle f, \bar{z}(-1) \rangle_{\bar{\mathcal{Z}}} \mid f \in \bar{\mathcal{Z}}\} = \bar{\mathbf{Z}}.$$

Again along the lines of Section 8.7, we can also form a forward recursion

$$z(t+1) = Fz(t) + Gw(t), \quad (8.8.2)$$

where  $F$  is unitarily equivalent to  $U(\mathbf{Z})^*$  and  $w$  is the forward generating process of  $(\mathbb{H}, U, \mathbf{X})$ ; and a backward recursion

$$\bar{z}(t-1) = \bar{F}\bar{z}(t) + \bar{G}\bar{w}(t), \quad (8.8.3)$$

where  $\bar{F}$  is unitarily equivalent to  $U(\bar{\mathbf{Z}})$  and  $\bar{w}$  is the backward generating processes of  $(\mathbb{H}, U, \mathbf{X})$ . Note that (8.8.2) and (8.8.3) are not a forward/backward pair in the sense of (8.7.14) and (8.7.24), but they represent different spaces  $\mathbf{X}$  and  $\bar{\mathbf{Z}}$ .

**Proposition 8.8.1.** *Let  $(\mathbb{H}, U, \mathbf{X})$  be a Markovian representation with error spaces  $\mathbf{Z}$  and  $\bar{\mathbf{Z}}$ . Then  $\mathbf{X}$  is observable if and only if  $\mathbf{X} \cap \bar{\mathbf{Z}} = 0$  and constructible if and only if  $\mathbf{X} \cap \mathbf{Z} = 0$ .*

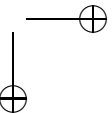
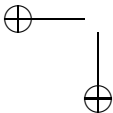
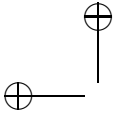
**Proof.** In view of (8.8.1) and the fact that  $\mathbf{X} \subset \mathbf{S}$ ,

$$\mathbf{X} \cap \mathbf{Z} = \mathbf{X} \cap \mathbf{S} \cap (\mathbf{H}^-)^\perp = \mathbf{X} \cap (\mathbf{H}^-)^\perp,$$

which, by definition, equals zero if and only if  $\mathbf{X}$  is constructible. The proof of the observability part is analogous.  $\square$

To formalize the regularity condition (6.9.1) in the present more general setting, consider the forward system (8.7.14) for a minimal Markovian representation  $(\mathbb{H}, U, \mathbf{X})$ . Note that, by (8.7.39c) and (8.7.38),

$$DD^* = \Lambda_0 - CPC^* \geq \Lambda_0 - C_+P_+C_+^* = D_+D_+^*. \quad (8.8.4)$$



In particular, the regularity condition (6.9.1) holds if and only if  $D_+$  has full rank. We shall say that the forward system (8.7.14) is *regular* if  $D$  has full rank, and *singular* otherwise. Likewise, the backward system (8.7.24) is regular if  $\bar{D}$  is full rank.

The rank of  $D$  is determined by the dimension of the null space of  $U(\mathbf{Z})$ , which is finite-dimensional even in the general case when  $\mathbf{Z}$  is infinite-dimensional.

**Proposition 8.8.2.** *Let  $(\mathbb{H}, U, \mathbf{X})$  be a Markovian representation with error spaces  $\mathbf{Z}$  and  $\bar{\mathbf{Z}}$  and generating processes  $(w, \bar{w})$ . Then the  $m \times p$  matrix  $D := E\{y(0)w(0)'\}$  and the  $m \times \bar{p}$  matrix  $\bar{D} := E\{y(0)\bar{w}(0)'\}$  have ranks*

$$\text{rank } D = p - \dim \ker U(\mathbf{Z}) \quad (8.8.5a)$$

$$\text{rank } \bar{D} = \bar{p} - \dim \ker U(\bar{\mathbf{Z}})^* \quad (8.8.5b)$$

where  $U(\mathbf{Z}) := E^{\mathbf{Z}} U|_{\mathbf{Z}}$  and  $U(\bar{\mathbf{Z}})^* := E^{\bar{\mathbf{Z}}} U^*|_{\bar{\mathbf{Z}}}$ .

The proof is based on the following lemma, which, for later reference, we state in more generality than needed here.

**Lemma 8.8.3.** *Let  $(\mathbf{S}, \bar{\mathbf{S}})$  be a pair of subspaces (in some Hilbert space  $\mathbb{H}$ ) with the property that  $U^*\mathbf{S} \subset \mathbf{S}$ ,  $U\bar{\mathbf{S}} \subset \bar{\mathbf{S}}$  and  $\bar{\mathbf{S}}^\perp \subset \mathbf{S}$ , and let  $\mathbf{X} := \mathbf{S} \ominus \bar{\mathbf{S}}^\perp$ . Then*

$$\ker U(\mathbf{X}) = \mathbf{X} \cap (U^*\mathbf{W}) = \bar{\mathbf{S}} \cap (U^*\mathbf{W}) \quad (8.8.6a)$$

$$\ker U(\mathbf{X})^* = \mathbf{X} \cap \bar{\mathbf{W}} = \mathbf{S} \cap \bar{\mathbf{W}}, \quad (8.8.6b)$$

where  $\mathbf{W} := U\mathbf{S} \ominus \mathbf{S}$  and  $\bar{\mathbf{W}} := \bar{\mathbf{S}} \ominus U\bar{\mathbf{S}}$ .

Note that, in this lemma,  $\mathbf{S}$  and  $\bar{\mathbf{S}}$  are perpendicularly intersecting (Corollary 7.2.5) and  $\mathbf{X} := \mathbf{S} \cap \bar{\mathbf{S}}$ , but  $\mathbf{X}$  need not be a splitting subspace since the conditions  $\mathbf{S} \supset \mathbf{H}^-$  and  $\bar{\mathbf{S}} \supset \mathbf{H}^+$  are not required.

**Proof.** Since  $U(\mathbf{X})\xi = 0$  if and only if  $\xi \perp U^*\mathbf{X}$ ,  $\ker U(\mathbf{X}) = \mathbf{X} \cap (U^*\mathbf{X})^\perp$ . Now, by Corollary 7.2.5,  $\mathbb{H} = \bar{\mathbf{S}}^\perp \oplus \mathbf{X} \oplus \mathbf{S}^\perp$ . Hence,  $(U^*\mathbf{X})^\perp = (U^*\mathbf{S})^\perp \oplus (U^*\bar{\mathbf{S}})^\perp$  and  $(U^*\mathbf{S})^\perp = \mathbf{S}^\perp \oplus (U^*\mathbf{W})$ ; i.e.,

$$(U^*\mathbf{X})^\perp = \mathbf{S}^\perp \oplus (U^*\mathbf{W}) \oplus (U^*\bar{\mathbf{S}})^\perp.$$

However,  $\mathbf{S}^\perp \perp \mathbf{X}$  and  $(U^*\bar{\mathbf{S}})^\perp \subset \bar{\mathbf{S}}^\perp \perp \mathbf{X}$ , and hence the first of equations (8.8.6a) follows. Then inserting  $\mathbf{X} = \mathbf{S} \cap \bar{\mathbf{S}}$  into this equation and observing that  $U^*\mathbf{W} \subset \mathbf{S}$ , we obtain the second of equations (8.8.6a). Similarly,  $\ker U(\mathbf{X})^* = \mathbf{X} \cap (U\mathbf{X})^\perp$ , and

$$(U\mathbf{X})^\perp = (U\mathbf{S})^\perp \oplus \bar{\mathbf{S}}^\perp \oplus \bar{\mathbf{W}},$$

where  $(U\mathbf{S})^\perp \subset \mathbf{S}^\perp \perp \mathbf{X}$  and  $\bar{\mathbf{S}}^\perp \perp \mathbf{X}$ , and consequently (8.8.6b) follows.  $\square$

The following corollary is needed for the proof of Proposition 8.8.2.

**Corollary 8.8.4.** *The nullspaces of  $U(\mathbf{Z})$  and  $U(\bar{\mathbf{Z}})$  of Proposition 8.8.2 are*

$$\ker U(\mathbf{Z}) = \mathbf{Z} \cap (U^*\mathbf{W}) \tag{8.8.7a}$$

$$\ker U(\bar{\mathbf{Z}})^* = \bar{\mathbf{Z}} \cap \bar{\mathbf{W}}, \tag{8.8.7b}$$

where  $\mathbf{W}$  and  $\bar{\mathbf{W}}$  are the wandering subspaces (8.1.9) and (8.1.13), respectively.

**Proof.** Equation (8.8.7a) follows directly from (8.8.6a) in Lemma 8.8.3 by taking  $\mathbf{X}$  and  $(\mathbf{S}, \bar{\mathbf{S}})$  to be  $\mathbf{Z}$  and  $(\mathbf{S}, (\mathbf{H}^-)^\perp)$ , respectively, and (8.8.7b) follows from (8.8.6b) by setting  $\mathbf{X}$  and  $(\mathbf{S}, \bar{\mathbf{S}})$  equal to  $\bar{\mathbf{Z}}$  and  $((\mathbf{H}^+)^\perp, \bar{\mathbf{S}})$ , respectively.  $\square$

**Proof of Proposition 8.8.2.** We want to show that  $\ker U(\mathbf{Z})$  and  $\ker D$  have the same dimension. To this end, note that, since  $\mathbf{Z} = \mathbf{S} \cap (\mathbf{H}^-)^\perp$  and  $(U^*\mathbf{W}) \subset \mathbf{S}$ , (8.8.7a) yields  $\ker U(\mathbf{Z}) = (\mathbf{H}^-)^\perp \cap (U^*\mathbf{W})$ . Hence,  $\dim \ker U(\mathbf{Z})$  equals the number  $\nu$  of linear independent  $a_1, a_2, \dots, a_\nu \in \mathbb{R}^p$  such that  $a'_k w(-1) \perp \mathbf{H}^-$  for  $k = 1, 2, \dots, \nu$ , or, equivalently,

$$E\{y(t)w(-1)'\}a_k = 0, \quad t < 0, k = 1, 2, \dots, \nu. \tag{8.8.8}$$

However,  $U^*\mathbf{H}^- \subset U^*\mathbf{S} \perp U^*\mathbf{W}$ , and hence  $E\{y(t)w(-1)'\}a_k$  is always zero for  $t < -1$ . Therefore, (8.8.8) holds if and only if  $Da_k = E\{y(-1)w(-1)'\}a_k = 0$  for  $k = 1, 2, \dots, \nu$ , establishing that  $\ker U(\mathbf{Z})$  and  $\ker D$  have the same dimension. This proves (8.8.5a). The proof of (8.8.5b) is analogous, using instead (8.8.7b).  $\square$

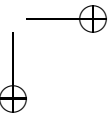
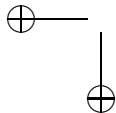
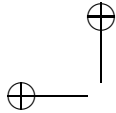
In general, we shall be interested in the situation where  $\{y(t)\}_{t \in \mathbb{Z}}$  is a purely nondeterministic, full-rank process, and  $(\mathbb{H}, U, \mathbf{X})$  is proper. Then,  $\bar{p} = p \geq m$ , so the forward system (8.7.14) is regular if and only if  $\text{rank } D = m$ , and the backward system (8.7.24) is regular if and only if  $\text{rank } \bar{D} = m$ . In particular, by (8.8.4), all forward systems of minimal Markovian representations are regular if  $D_+$  is full rank. Also note that, since the forward error space  $\mathbf{Z}_-$  of the predictor space  $\mathbf{X}_-$  and the backward error space of  $\mathbf{X}_+$  are trivial,  $D_-$  and  $\bar{D}_+$  are always full rank. In other words, the (forward and backward) steady-state Kalman filters are always regular.

Regularity and singularity play an important role in the theory of invariant directions for the matrix Riccati equation, a topic to which we return in Chapter ??, and it is the topic of the next section. We provide Hardy space characterizations of regularity in Section 9.3.

As a preliminary for the next section, we consider a standard procedure for constructing (in general, nonminimal) Markovian representations for which  $D = 0$  and  $\bar{D} = 0$ . This amounts to amending the state process by including the observation noise in the state, which is a standard construction in the finite-dimensional case.

**Proposition 8.8.5.** *Let  $(\mathbb{H}, U, \mathbf{X})$  be a Markovian representation with  $\mathbf{X} \sim (\mathbf{S}, \bar{\mathbf{S}})$ , and set  $\tilde{\mathbf{X}} := (U\mathbf{S}) \cap \bar{\mathbf{S}}$ . Then  $(\tilde{\mathbf{X}}, U, \mathbb{H})$  is a Markovian representation with  $\tilde{\mathbf{X}} \sim (U\mathbf{S}, \bar{\mathbf{S}})$ , and  $y_k(0) \in \tilde{\mathbf{X}}$  for  $k = 1, 2, \dots, m$ . Moreover,*

$$\tilde{\mathbf{X}} = \mathbf{X} \oplus \mathbf{W} = (U\mathbf{X}) \oplus \bar{\mathbf{W}}, \tag{8.8.9}$$



where  $\mathbf{W}$  and  $\bar{\mathbf{W}}$  are the subspaces (8.1.9) and (8.1.13), respectively.

**Proof.** By assumption,  $(\mathbf{S}, \bar{\mathbf{S}})$  satisfies conditions (i), (ii) and (iii) in Theorem 8.1.1. Then so does  $(U\mathbf{S}, \bar{\mathbf{S}})$ . In fact,  $\mathbf{H}^- \subset \mathbf{S} \subset U\mathbf{S}$  and  $U^*(U\mathbf{S}) = \mathbf{S} \subset U\mathbf{S}$  establishes (i) and (ii). From (8.1.9) we have  $U\mathbf{S} = \mathbf{S} \oplus \mathbf{W}$ . Therefore, since  $\mathbf{W} \subset \mathbf{S}^\perp \subset \bar{\mathbf{S}}$ ,

$$(U\mathbf{S}) \cap \bar{\mathbf{S}} = (\mathbf{S} \cap \bar{\mathbf{S}}) \oplus \mathbf{W}. \quad (8.8.10)$$

From (8.1.9) we also have  $(U\mathbf{S})^\perp = \mathbf{S}^\perp \ominus \mathbf{W}$ , and hence

$$((U\mathbf{S}) \cap \bar{\mathbf{S}}) \oplus (U\mathbf{S})^\perp = (\mathbf{S} \cap \bar{\mathbf{S}}) \oplus \mathbf{S}^\perp,$$

establishing that  $(U\mathbf{S}, \bar{\mathbf{S}})$  satisfies (iii). Consequently  $(\tilde{\mathbf{X}}, U, \mathbb{H})$  is a Markovian representation by Theorem 8.1.1. Also, (8.8.10) is precisely the first of equations (8.8.9). To establish the second of equations (8.8.9), observe that (8.1.13) implies that  $\bar{\mathbf{S}} = (U\bar{\mathbf{S}}) \cap \bar{\mathbf{W}}$ , and therefore, since  $\bar{\mathbf{W}} \subset (U\bar{\mathbf{S}})^\perp \subset U\mathbf{S}$ ,

$$\tilde{\mathbf{X}} = (U\mathbf{S}) \cap \bar{\mathbf{S}} = ((U\mathbf{S}) \cap (U\bar{\mathbf{S}})) \oplus \bar{\mathbf{W}} = (U\mathbf{X}) \oplus \bar{\mathbf{W}}.$$

Finally,

$$y_k(0) \in (U\mathbf{H}^-) \cap \mathbf{H}^+ \subset (U\mathbf{S}) \cap \bar{\mathbf{S}} = \tilde{\mathbf{X}}$$

for  $k = 1, 2, \dots, m$ .  $\square$

## 8.9 Models without observation noise

By changing the definition of the past space  $\mathbf{H}^-$  to including  $y(0)$ ; i.e., taking

$$\mathbf{H}^- = \overline{\text{span}}\{a'y(t) \mid t \leq 0; a \in \mathbb{R}^m\} \quad (8.9.1)$$

while retaining the old definition of  $\mathbf{H}^+$ , we obtain complete symmetry between the the past and the future. In this case,

$$a'y(0) \in \mathbf{H}^- \cap \mathbf{H}^+ \subset \mathbf{S} \cap \bar{\mathbf{S}} = \mathbf{X}, \quad a \in \mathbb{R}^m$$

for any Markovian splitting subspace  $\mathbf{X} \sim (\mathbf{S}, \bar{\mathbf{S}})$ , leading to models without observation noise.

Given a Markovian representation  $(\mathbb{H}, U, \mathbf{X})$  in this framework, the construction of the forward and a backward stochastic system follows along similar lines as described earlier in this chapter, but with certain modifications that uphold the symmetry between the forward and the backward model. In particular, the definition (8.1.9) of  $\mathbf{W}$  needs be to changed to  $\mathbf{W} := \mathbf{S} \ominus U^*\mathbf{S}$ , leading to the following modification of (8.1.14) and (8.1.15):

$$\begin{cases} U\mathbf{X} \subset \mathbf{X} \oplus (U\mathbf{W}) \\ \mathbf{Y} \subset \mathbf{X} \end{cases} \quad \begin{cases} U^*\mathbf{X} \subset \mathbf{X} \oplus (U^*\bar{\mathbf{W}}) \\ \mathbf{Y} \subset \mathbf{X} \end{cases}$$



This is the starting point for constructing a forward system

$$\begin{cases} x(t+1) = Ax(t) + Bw(t+1) \\ y(t) = Cx(t), \end{cases} \tag{8.9.2}$$

where the operators  $A : \mathcal{X} \rightarrow \mathcal{X}$  and  $C : \mathcal{X} \rightarrow \mathbb{R}^m$  are as defined in (8.7.10), while  $B : \mathbb{R}^p \rightarrow \mathcal{X}$  now sends  $a \in \mathbb{R}^p$  to  $TE^{\mathbf{X}} a'w(0)$ ; and a backward system

$$\begin{cases} \bar{x}(t-1) = A^* \bar{x}(t) + \bar{B} \bar{w}(t-1) \\ y(t) = \bar{C} \bar{x}(t), \end{cases} \tag{8.9.3}$$

where  $\bar{B} : \mathbb{R}^p \rightarrow \mathcal{X}$  is as defined in (8.7.22), while  $y(-1)$  should be exchanged for  $y(0)$  in the definition of  $\bar{C} : \mathcal{X} \rightarrow \mathbb{R}^m$ . We now have complete symmetry between the forward and backward system, which is also reflected in the fact that now

$$\overline{\text{span}}\{\langle f, x(0) \rangle_{\mathcal{X}} \mid f \in \mathcal{X}\} = \mathbf{X} = \overline{\text{span}}\{\langle f, \bar{x}(0) \rangle_{\mathcal{X}} \mid f \in \mathcal{X}\}. \tag{8.9.4}$$

The observability and constructibility conditions of Theorem 8.7.4 remain the same.

The class of Markovian representations in the new framework with  $\mathbf{H}^-$  given by (8.9.1) is smaller than that in the standard framework. In fact, any Markovian representation in the new framework is clearly a Markovian representation in standard one, but the converse is not true. However, Proposition 8.8.5 suggests a procedure for producing a new Markovian representations from old ones.

For clarity of notation, we return to our standard definition of  $\mathbf{H}^-$ , referring to (8.9.1) as  $U\mathbf{H}^-$ , and we refer to the two types of splitting subspaces as  $(\mathbf{H}^-, \mathbf{H}^+)$ -splitting and  $(U\mathbf{H}^-, \mathbf{H}^+)$ -splitting, respectively.

**Corollary 8.9.1.** *Let  $(\mathbb{H}, U, \mathbf{X})$  be a Markovian representation in the  $(\mathbf{H}^-, \mathbf{H}^+)$  framework, and let  $\tilde{\mathbf{X}}$  be defined as in Proposition 8.8.5. Then,  $(\tilde{\mathbf{X}}, U, \mathbb{H})$  is a Markovian representation in the  $(U\mathbf{H}^-, \mathbf{H}^+)$  framework.*

**Proof.** The corollary follows immediately from Proposition 8.8.5, merely noting that  $U\mathbf{H}^- \subset U\mathbf{S}$ .  $\square$

The Markovian representation  $(\tilde{\mathbf{X}}, U, \mathbb{H})$  of Corollary 8.9.1 is certainly non-minimal in the  $(\mathbf{H}^-, \mathbf{H}^+)$  framework if  $\mathbf{W}$  is nontrivial, but it could very well be minimal in the  $(U\mathbf{H}^-, \mathbf{H}^+)$  framework. The following result shows when this is the case.

**Theorem 8.9.2.** *Let  $(\mathbb{H}, U, \mathbf{X})$  be observable (constructible) in the  $(\mathbf{H}^-, \mathbf{H}^+)$  framework, and let  $\tilde{\mathbf{X}}$  be defined as in Proposition 8.8.5. Then  $(\tilde{\mathbf{X}}, U, \mathbb{H})$  is observable (constructible) in the  $(U\mathbf{H}^-, \mathbf{H}^+)$  framework if and only if  $\ker U(\tilde{\mathbf{Z}}) = 0$  ( $\ker U(\mathbf{Z}) = 0$ ), where  $\mathbf{Z}$  and  $\tilde{\mathbf{Z}}$  are the error spaces (8.8.1) of  $\mathbf{X} \sim (\mathbf{S}, \tilde{\mathbf{S}})$ .*

**Proof.** Let us first prove the statement about observability, which is the simplest part. By Proposition 8.8.1 and Lemma 8.8.4, we need to show that  $\tilde{\mathbf{X}} \cap \tilde{\mathbf{Z}} = 0$  if and

only if  $\bar{\mathbf{Z}} \cap \bar{\mathbf{W}} = 0$ . However, since  $\bar{\mathbf{W}} \subset \tilde{\mathbf{X}}$  by (8.8.9),  $\tilde{\mathbf{X}} \cap \bar{\mathbf{Z}} \supset \bar{\mathbf{Z}} \cap \bar{\mathbf{W}}$ , so the only-if part is trivial. To prove the if part, suppose there is a nonzero  $\lambda \in \tilde{\mathbf{X}} \cap \bar{\mathbf{Z}}$ . Then, in view of (8.8.9),  $\lambda = U\xi + \eta$ , where  $\xi \in \mathbf{X}$  and  $\eta \in \bar{\mathbf{W}}$ . However,  $\lambda \in \bar{\mathbf{Z}} \perp \mathbf{H}^+ \supset U\mathbf{H}^+$  and  $\eta \in \bar{\mathbf{W}} \perp U\bar{\mathbf{S}} \supset U\mathbf{H}^+$ , and consequently  $U\xi \perp U\mathbf{H}^+$ , or, equivalently,  $\xi \perp \mathbf{H}^+$ ; i.e.,  $\xi \in \mathbf{X} \cap (\mathbf{H}^+)^\perp$ , which is zero by observability of  $\mathbf{X}$ . Since therefore  $\lambda \in \bar{\mathbf{W}}$ , we have  $\lambda \in \bar{\mathbf{Z}} \cap \bar{\mathbf{W}}$ . Hence  $\bar{\mathbf{Z}} \cap \bar{\mathbf{W}} = 0$ , proving the if part.

To establish the statement about constructibility, we need to prove that  $\tilde{\mathbf{X}} \cap \tilde{\mathbf{Z}} = 0$  if and only if  $\mathbf{Z} \cap (U^*\mathbf{W}) = 0$ , where  $\tilde{\mathbf{X}} = \mathbf{X} \oplus \mathbf{W}$  and  $\tilde{\mathbf{Z}} = U\mathbf{Z}$ . To this end, first suppose that  $\mathbf{Z} \cap (U^*\mathbf{W}) \neq 0$ . Then, applying the unitary shift  $U$ , we have  $\tilde{\mathbf{Z}} \cap \mathbf{W} \neq 0$ , which, in view of the fact that  $\mathbf{W} \subset \tilde{\mathbf{X}}$ , implies that  $\tilde{\mathbf{X}} \cap \tilde{\mathbf{Z}} \neq 0$ . Conversely, suppose that there is a nonzero  $\lambda \in \tilde{\mathbf{X}} \cap \tilde{\mathbf{Z}}$ . Then,  $\lambda = \xi + \eta$ , where  $\xi \in \mathbf{X}$  and  $\eta \in \mathbf{W}$ . However,  $\lambda \in U\mathbf{Z} \perp U\mathbf{H}^- \supset \mathbf{H}^-$  and  $\eta \in \mathbf{W} \perp \mathbf{S} \supset \mathbf{H}^-$ , and therefore  $\xi \perp \mathbf{H}^-$ ; i.e.,  $\xi \in \mathbf{X} \cap (\mathbf{H}^-)^\perp$ , which is zero by the constructibility of  $\mathbf{X}$ . Hence  $\lambda \in \mathbf{W}$ ; i.e.,  $\lambda \in \tilde{\mathbf{Z}} \cap \mathbf{W}$ , and consequently  $(U\mathbf{Z}) \cap \mathbf{W} \neq 0$ , or, equivalently,  $\mathbf{Z} \cap (U^*\mathbf{W}) \neq 0$ . This proves the constructibility part.  $\square$

We can now relate the the question of the preservation of minimality to the regularity of the forward and backward systems – (8.7.14) and (8.7.24), respectively – of  $(\mathbb{H}, U, \mathbf{X})$  and the size of the ambient space  $\mathbb{H}$ .

**Corollary 8.9.3.** *Let  $(\mathbb{H}, U, \mathbf{X})$  be a minimal proper Markovian representation, in the  $(\mathbf{H}^-, \mathbf{H}^+)$  framework, with generating processes  $(w, \bar{w})$ , and let  $\tilde{\mathbf{X}}$  be defined as in Proposition 8.8.5. Then  $(\tilde{\mathbf{X}}, U, \mathbb{H})$  is minimal in the  $(U\mathbf{H}^-, \mathbf{H}^+)$  framework if and only if all the conditions*

- (i)  $\mathbf{X}$  is internal, i.e.,  $\mathbf{X} \subset \mathbf{H}$ ,
- (ii)  $D := E\{y(0)w(0)'\}$  has full rank,
- (iii)  $\bar{D} := E\{y(0)\bar{w}(0)'\}$  has full rank,

hold.

**Proof.** By Theorem 8.9.2 and Corollary 7.4.10,  $(\tilde{\mathbf{X}}, U, \mathbb{H})$  is minimal if and only if  $\ker U(\mathbf{Z}) = 0$  and  $\ker U(\bar{\mathbf{Z}}) = 0$ . Since  $(\mathbb{H}, U, \mathbf{X})$  is proper,  $p = \bar{p} \geq m$ . Consequently, by Proposition 8.8.2,  $\ker U(\mathbf{Z}) = 0$  and  $\ker U(\bar{\mathbf{Z}}) = 0$  if and only if  $D$  and  $\bar{D}$  have full rank (i.e., rank  $m$ ) and  $p = m$  (i.e.,  $\mathbf{X}$  is internal).  $\square$

In Chapter 9, after having introduced some additional tools, we illustrate this result by a simple example.

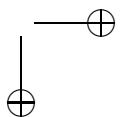
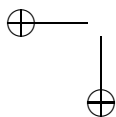
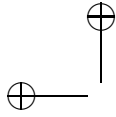
## 8.10 Bibliographical notes

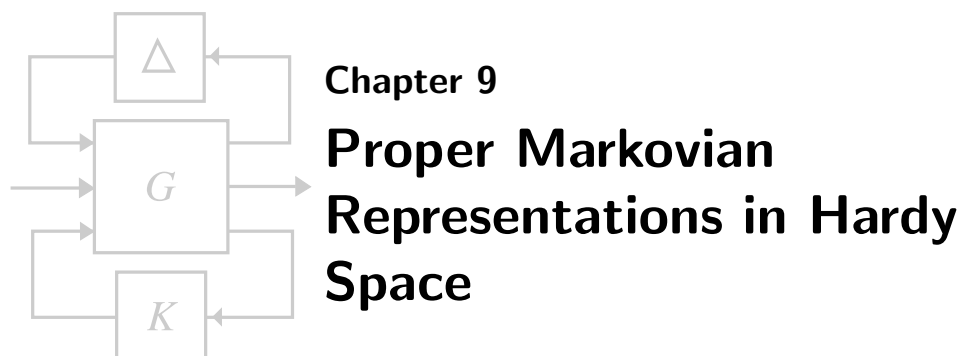
The study of the geometric structure of stochastic models was initiated in [1, 102, 109] and was first developed into a theory of Markovian representation by Ruckebusch [107, 108, 111, 113, 112], Lindquist and Picci [75, 77, 76, 79, 80, 81, 82, 85], Lindquist, Picci and Ruckebusch [87], Lindquist, Pavon and Picci [74] and Lindquist

and Pavon [73]. Early contributions are also due to Caines and Delchamps [16], Frazho [32, 31], Foaş and Frazho [30], and Michaletzky [94, 93].

Sections 8.1 and 8.3 are essentially developed along the lines of [86]. The results in Section 8.2 are basically contained in [116]: Theorem 8.2.1 is Theorem 1.2 in Chapter II of [116] modified to our setting and Theorem 8.2.3 is Proposition 1.4. Theorem 8.2.7 follows from Proposition 6.7 in Chapter II and Proposition 4.2 in Chapter III in [116]. The material in Section 8.4 appears here for the first time. Section 8.5 follows [86], where Proposition 8.5.2 is proved in the continuous-time setting. The material in Section 8.6 generalizes results [86], allowing a purely deterministic component, and connects the geometric theory to classical stochastic realization theory [4, 28, 78]. The development of general dynamical realizations in Section 8.7 is related to the restricted-shift realization of Fuhrmann [33], Helton [47], and Baras and Brockett [6]. It was first developed in this form in [80, 81] in continuous time and in [73, 74] in discrete time. State-space isomorphism results were developed in [85] and is closely related to results in [34]. Sections 8.8 and 8.9, finally, are based on [73] where the theory was developed for internal Markovian representations only.

Finally, we should mention the books by Caines [14] (Chapter 4) and Michaletzky, Bokor and Várlaki [92].





## Chapter 9 Proper Markovian Representations in Hardy Space

In this chapter we reformulate the splitting geometry in terms of functional models in Hardy space. This allows us to use the power of Hardy space theory to prove several additional results and useful characterizations. We shall only deal with proper Markovian representation, and therefore we formulate several functional criteria for properness.

Throughout this chapter, we assume that the process  $\{y(t)\}_{t \in \mathbb{Z}}$  is purely nondeterministic and reversible and hence purely nondeterministic also in the backward direction. Therefore

$$\mathbf{H}^- = \mathbf{H}^-(w_-) \quad \text{and} \quad \mathbf{H}^+ = \mathbf{H}^+(\bar{w}_+), \tag{9.0.1}$$

where the white noise processes  $w_-$  and  $\bar{w}_+$  are the forward and backward innovation processes, respectively.

### 9.1 Functional representations of Markovian representations

In contrast to the finite-dimensional case (Theorems 8.4.3 and 8.6.3), the process  $y$  being purely nondeterministic does not insure that all minimal Markovian representations are proper, as seen from Example 8.1.5. However, properness is insured by the following simple criterion, which will be stated and proved in more generality in Section 9.2 (Proposition 9.2.8 and Theorem 9.2.10).

**Proposition 9.1.1.** *There exist proper Markovian representations of  $y$  if and only if the frame space  $\mathbf{H}^\square$ , defined by (7.4.24), is proper. In this case, all minimal Markovian representations are proper.*

Clearly,  $\mathbf{X}_- \sim (\mathbf{H}^-, \bar{\mathbf{S}}_-)$  is proper only if the remote past of  $\mathbf{H}^-$  is trivial, and  $\mathbf{X}_+ \sim (\mathbf{S}_+, \mathbf{H}^+)$  is proper only if the remote future of  $\mathbf{H}^+$  is trivial. Therefore, for all minimal Markovian representations of  $y$  to be proper it is necessary that  $y$  is

reversible so that it is purely nondeterministic in both the forward and the backward direction, in harmony with our assumption (9.0.1). However, reversibility is not sufficient. Indeed, the process in Example 8.1.5 is full rank, and hence reversible (Proposition 4.5.11), but  $\mathbf{X}_-$  is not proper. A sufficient condition (noncyclicity) will be introduced on page 243.

In Section 8.1 it was shown that, to each proper Markovian representation  $(\mathbb{H}, U, \mathbf{X})$  of multiplicity  $p \geq m$  with  $\mathbf{X} \sim (\mathbf{S}, \bar{\mathbf{S}})$ , there corresponds a pair  $(w, \bar{w})$  of  $p$ -dimensional white noise processes such that  $\mathbf{H}(w) = \mathbf{H}(\bar{w}) = \mathbb{H}$  and

$$\mathbf{S} = \mathbf{H}^-(w) \quad \text{and} \quad \bar{\mathbf{S}} = \mathbf{H}^+(\bar{w}). \quad (9.1.1)$$

These processes are called the *generating processes* of the Markovian representation, and they are uniquely determined modulo multiplication by a constant  $p \times p$  orthogonal matrix.

By Proposition 9.1.1, there will not exist any proper Markovian representations unless the frame space  $\mathbf{H}^\square \sim (\mathbf{S}_+, \bar{\mathbf{S}}_-)$  is proper, i.e., unless there are unique generating processes  $w_+$  and  $\bar{w}_-$  such that

$$\mathbf{S}_+ = \mathbf{H}^-(w_+) \quad \text{and} \quad \bar{\mathbf{S}}_- = \mathbf{H}^+(\bar{w}_-), \quad (9.1.2)$$

where

$$\mathbf{S}_+ := (\mathbf{N}^+)^\perp = \mathbf{H}^+ \vee (\mathbf{H}^-)^\perp \quad \text{and} \quad \bar{\mathbf{S}}_- := (\mathbf{N}^-)^\perp = \mathbf{H}^- \vee (\mathbf{H}^+)^\perp \quad (9.1.3)$$

(Proposition 7.4.15). In this case, the predictor spaces  $\mathbf{X}_-$  and  $\mathbf{X}_+$  are also proper. In fact, by Propositions 7.4.5 and 7.4.6 in Section 7,

$$\mathbf{X}_- \sim (\mathbf{S}_-, \bar{\mathbf{S}}_-) \quad \text{and} \quad \mathbf{X}_+ \sim (\mathbf{S}_+, \bar{\mathbf{S}}_+), \quad (9.1.4)$$

where  $\mathbf{S}_- = \mathbf{H}^-$  and  $\bar{\mathbf{S}}_+ = \mathbf{H}^+$ . Consequently, the predictor spaces  $\mathbf{X}_-$  and  $\mathbf{X}_+$  have generating processes  $(w_-, w_+)$  and  $(\bar{w}_-, \bar{w}_+)$ , respectively, which should explain the notation.

As demonstrated in Section ??, the spectral representations

$$w(t) = \int_{-\pi}^{\pi} e^{i\theta t} d\hat{w} \quad \text{and} \quad \bar{w}(t) = \int_{-\pi}^{\pi} e^{i\theta t} d\hat{\bar{w}} \quad (9.1.5)$$

define two unitary isomorphisms,  $\mathcal{J}_{\hat{w}}, \mathcal{J}_{\hat{\bar{w}}} : L_p^2 \rightarrow \mathbb{H}$ , via

$$\mathcal{J}_{\hat{w}} f = \int_{-\pi}^{\pi} f(e^{i\theta}) d\hat{w} \quad \text{and} \quad \mathcal{J}_{\hat{\bar{w}}} f = \int_{-\pi}^{\pi} f(e^{i\theta}) d\hat{\bar{w}}, \quad (9.1.6)$$

under which  $\mathbf{S}$  and  $\bar{\mathbf{S}}$  are isomorphic to the Hardy spaces  $H_p^2$  and  $\bar{H}_p^2$ , respectively, i.e.,

$$\mathbf{S} := \mathbf{H}^-(w) = \mathcal{J}_{\hat{w}} z^{-1} H_p^2 \quad \text{and} \quad \bar{\mathbf{S}} := \mathbf{H}^+(\bar{w}) = \mathcal{J}_{\hat{\bar{w}}} \bar{H}_p^2, \quad (9.1.7)$$

and the shift  $U$  becomes multiplication by  $z = e^{i\theta}$ , i.e.,

$$U \mathcal{J}_{\hat{w}} = \mathcal{J}_{\hat{w}} M_z. \quad (9.1.8)$$

### Spectral factors and the structural function

We are now in a position to reformulate some basic results of Section 6.4 in a more general setting. Since  $(J_{\hat{w}})^{-1}J_{\hat{w}}$  is a unitary operator which commutes with the shift on  $L_p^2$ , it can be represented by a multiplication operator

$$(J_{\hat{w}})^{-1}J_{\hat{w}} = M_K \tag{9.1.9}$$

where  $M_K f = fK$  and  $K$  is a unitary  $p \times p$  matrix function (Theorem 4.3.3). Generalizing Definition 4.6.1 slightly we shall say that an isometry which sends analytic functions to analytic functions is *inner*. As before, a  $p \times q$  matrix function  $V$  on  $\mathbb{D}$  such that  $H_p^2 V$  is dense in  $H_q^2$  is called *outer*. Functions with the corresponding properties with respect to the conjugate Hardy space  $\bar{H}_p^2$  will be called *conjugate inner* and *conjugate outer* respectively.

**Lemma 9.1.2.** *Let  $(\mathbb{H}, U, \mathbf{X})$  be a proper Markovian representation with generating processes  $w$  and  $\bar{w}$ . Then there is a unique pair  $(W, \bar{W})$  of spectral factors, the first being analytic and the second coanalytic, such that*

$$y(t) = \int_{-\pi}^{\pi} e^{i\theta t} W(e^{i\theta}) d\hat{w} = \int_{-\pi}^{\pi} e^{i\theta t} \bar{W}(e^{i\theta}) d\hat{w}. \tag{9.1.10}$$

Moreover, the matrix function  $K$  defined by (9.1.9) is inner, and satisfies

$$W = \bar{W}K, \tag{9.1.11}$$

i.e., in particular,

$$\bar{w}(t) = \int_{-\pi}^{\pi} e^{i\theta t} K(e^{i\theta}) d\hat{w}. \tag{9.1.12}$$

**Proof.** In view of (9.1.8), conditions (9.1.10) hold if and only if

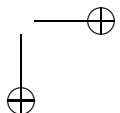
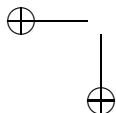
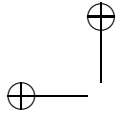
$$W = \begin{bmatrix} J_{\hat{w}}^{-1} y_1(0) \\ J_{\hat{w}}^{-1} y_2(0) \\ \vdots \\ J_{\hat{w}}^{-1} y_m(0) \end{bmatrix} \quad \text{and} \quad \bar{W} = \begin{bmatrix} J_{\bar{w}}^{-1} y_1(0) \\ J_{\bar{w}}^{-1} y_2(0) \\ \vdots \\ J_{\bar{w}}^{-1} y_m(0) \end{bmatrix}. \tag{9.1.13}$$

Clearly  $W$  and  $\bar{W}$  are spectral factors. Moreover, for all  $a \in \mathbb{R}^m$ ,  $a'y(0) \in U\mathbf{H}^- \subset U\mathbf{H}^-(w)$ , and hence  $a'W \in H_p^2$ , implying that  $W$  is analytic. In the same way, since  $a'y(0) \in \mathbf{H}^+ \subset \mathbf{H}^+(\bar{w})$ ,  $a'\bar{W} \in \bar{H}_p^2$  for all  $a \in \mathbb{R}^m$ , and hence  $\bar{W}$  is coanalytic. In view of (9.1.9),

$$\int_{-\pi}^{\pi} f d\hat{w} = \int_{-\pi}^{\pi} fK d\hat{w} \quad \text{for all } f \in H_p^2,$$

and hence (9.1.12) follows. Moreover, since

$$a'y(0) = J_{\hat{w}} a'W = J_{\bar{w}} a'\bar{W} \quad \text{for all } a \in \mathbb{R}^m,$$



(9.1.11) also follows from (9.1.9). Finally, the fact that  $K$  is inner is a consequence of the perpendicular intersection property  $\bar{\mathbf{S}}^\perp \subset \mathbf{S}$  (Theorem 7.2.4), which may be written  $\mathbf{H}^-(\bar{w}) \subset \mathbf{H}^-(w)$ , which is equivalent to  $H_p^2 K \subset H_p^2$  under the isomorphism  $J_{\bar{w}} z^{-1}$ . The subspace  $\mathbf{H}^-(\bar{w})$  is invariant under the backward shift  $U^{-1}$ , and hence  $H_p^2 K$  is invariant under  $z^{-1}$ . Therefore  $K$  is inner (Theorem 4.6.4).  $\square$

It follows from this analysis that the spectral factors  $W$  and  $\bar{W}$  are uniquely determined by the subspaces  $\mathbf{S}$  and  $\bar{\mathbf{S}}$ , once a specific choice of generating process  $w, \bar{w}$  has been made. This amounts to saying that  $W$  and  $\bar{W}$  are determined by  $\mathbf{S}$  and  $\bar{\mathbf{S}}$ , respectively, modulo right multiplication by a constant  $p \times p$  orthogonal matrix. The equivalence class of  $m \times p$  spectral factors

$$[W] := \{WT \mid T \text{ orthogonal } p \times p \text{ matrix}\} \tag{9.1.14}$$

will sometimes be denoted by  $W \bmod \mathcal{O}(p)$ , where  $\mathcal{O}(p)$  is the  $p$ -dimensional orthogonal group, or merely  $W \bmod \mathcal{O}$  if the dimension of  $W$  is clear from context.

Hence, given a proper Markovian representation  $(\mathbb{H}, U, \mathbf{X})$  with  $\mathbf{X} \sim (\mathbf{S}, \bar{\mathbf{S}})$ , we determine a unique (mod  $\mathcal{O}$ ) pair  $(W, \bar{W})$  of  $m \times p$  spectral factors, one being analytic and corresponding to  $\mathbf{S}$ , and the other coanalytic and corresponding to  $\bar{\mathbf{S}}$ . In terms of the splitting geometry the analyticity of  $W$  reflects the condition  $\mathbf{S} \supset \mathbf{H}^-$ , the coanalyticity of  $\bar{W}$  the condition  $\bar{\mathbf{S}} \supset \mathbf{H}^+$ , and  $K$  being inner the perpendicular intersection between  $\mathbf{S}$  and  $\bar{\mathbf{S}}$ . We shall call a triplet  $(W, \bar{W}, K)$  where  $W$  and  $\bar{W}$  are  $m \times p$  spectral factors for some  $p \geq m$  and  $K$  is a  $p \times p$  matrix function satisfying the equation  $W = \bar{W}K$  a *Markovian triplet* if  $W$  is analytic,  $\bar{W}$  coanalytic and  $K$  inner.

In view of (9.1.9),  $K$  is uniquely determined by the Markovian representation  $(\mathbb{H}, U, \mathbf{X})$ , modulo right and left multiplication by orthogonal constant matrices, and we shall call it the *structural function* of  $(\mathbb{H}, U, \mathbf{X})$ . It follows that the Markovian triplets corresponding to a Markovian representation are all related by the equivalence

$$(W, \bar{W}, K) \sim (WT_1, \bar{W}T_2, T_2^{-1}KT_1) \quad T_1, T_2 \in O(p) \tag{9.1.15}$$

The corresponding equivalence class of Markovian triplets will be denoted  $[W, \bar{W}, K]$  or  $(W, \bar{W}, K) \bmod \mathcal{O}$ .

If the Markovian representation  $(\mathbb{H}, U, \mathbf{X})$  is internal, its multiplicity  $p$  equals  $m$ . Then  $W$  and  $\bar{W}$  are square and hence, since  $\Phi$  is full rank, invertible. In this case, (9.1.11) can be solved for  $K$  to yield the structural function

$$K = \bar{W}^{-1}W. \tag{9.1.16}$$

Hence, the predictor spaces  $\mathbf{X}_-$  and  $\mathbf{X}_+$  have the Markovian triplets  $(W_-, \bar{W}_-, K_-)$  and  $(W_+, \bar{W}_+, K_+)$ , respectively, where  $W_-$  is outer,  $\bar{W}_+$  conjugate outer,  $K_- = \bar{W}_-^{-1}W_-$  and  $K_+ = \bar{W}_+^{-1}W_+$ .

A Markovian triplet is called *tight* if  $K$  is uniquely determined by  $W$  and  $\bar{W}$ . As we have just seen, this is always the case for internal Markovian representations,



in which case (9.1.16) holds. Nontightness occurs when the subspace  $\mathbf{X} \cap (\mathbf{H})^\perp$  is nontrivial, in which case some modes of the state process will evolve independently of  $y$ . In fact, tightness is implied by either observability or constructibility, which conditions, as we shall see below, are equivalent to the coprimeness of the factorizations  $W = \bar{W}K$  and  $\bar{W} = WK^*$  respectively (Corollary 9.2.3). Such coprime factorizations are known to be unique mod  $\mathcal{O}$ . Consequently,  $\bar{W}$  and  $K$  are uniquely determined by  $W$  in the observable case, and  $W$  and  $K$  are uniquely determined by  $\bar{W}$  in the constructible case.

### The inner triplet of a Markovian representation

In a Markovian triplet  $(W, \bar{W}, K)$  the spectral factors  $W$  and  $\bar{W}$  have unique outer-inner factorizations

$$W = W_- Q \quad \text{and} \quad \bar{W} = \bar{W}_+ \bar{Q}, \tag{9.1.17}$$

where  $W_-$  is the outer and  $\bar{W}_+$  the conjugate outer spectral factor (Theorem 4.6.5). The  $m \times p$  matrix functions  $Q$  and  $\bar{Q}$  are inner and conjugate inner, respectively, and

$$QQ^* = I_m \quad \text{and} \quad \bar{Q}\bar{Q}^* = I_m. \tag{9.1.18}$$

Thus, to each Markovian representation of  $y$ , there is a unique triplet  $(K, Q, \bar{Q})$  of inner functions, which we shall refer to as its *inner triplet*.

**Lemma 9.1.3.** *Let  $(\mathbb{H}, U, \mathbf{X})$  be a proper Markovian representation with inner triplet  $(K, Q, \bar{Q})$  and generating processes  $w$  and  $\bar{w}$ . Then*

$$w(t) = \int_{-\pi}^{\pi} e^{i\theta t} Q^*(e^{i\theta}) d\hat{w}_- + z(t), \tag{9.1.19}$$

where  $w_-$  is the innovation process, and where  $z$  is a p.n.d. process with the property that  $\mathbf{H}(z) = \mathbf{H}^\perp$  and the spectral density  $\Pi := I - Q^*Q$ . The orthogonal projection  $E^{\mathbf{H}}$  onto the internal subspace  $\mathbf{H}$  corresponds, under the isomorphism  $\mathcal{J}_{\hat{w}}$ , to multiplication with  $Q^*Q$ , i.e.,

$$E^{\mathbf{H}} \mathcal{J}_{\hat{w}} = \mathcal{J}_{\hat{w}} M_{Q^*Q}. \tag{9.1.20}$$

Similarly,

$$\bar{w}(t) = \int_{-\pi}^{\pi} e^{i\theta t} \bar{Q}^*(e^{i\theta}) d\hat{w}_+ + \bar{z}(t), \tag{9.1.21}$$

where  $\bar{w}_+$  is the backward innovation process and

$$\bar{z}(t) = \int_{-\pi}^{\pi} e^{i\theta t} K(e^{i\theta}) d\hat{z}. \tag{9.1.22}$$

Moreover,  $\mathbf{H}(\bar{z}) = \mathbf{H}^\perp$  and

$$E^{\mathbf{H}} \mathcal{J}_{\hat{w}} = \mathcal{J}_{\hat{w}} M_{\bar{Q}^*\bar{Q}}. \tag{9.1.23}$$

Finally,

$$K\Pi = \bar{\Pi}K, \tag{9.1.24}$$

where  $\bar{\Pi} := I - \bar{Q}^* \bar{Q}$  is the spectral density of  $\bar{z}$ .

**Proof.** In view of

$$d\hat{y} = Wd\hat{w} = W_-d\hat{w}_-,$$

(9.1.17) yields

$$d\hat{w}_- = Qd\hat{w}. \quad (9.1.25)$$

Since therefore  $Q^*d\hat{w}_- = Q^*Qd\hat{w}$ ,

$$d\hat{w} = Q^*d\hat{w}_- + d\hat{z}, \quad (9.1.26)$$

where  $d\hat{z} := \Pi d\hat{w}$  with  $\Pi := I - Q^*Q$ . From this we obtain (9.1.19), or, more generally,

$$\int f d\hat{w} = \int f Q^* d\hat{w}_- + \int f d\hat{z} \quad (9.1.27)$$

for all  $f \in L_p^2$ . The last term in (9.1.27) can also be written  $\int f \Pi d\hat{w}$  and is therefore always orthogonal to the internal subspace  $\mathbf{H}$ , or, equivalently, to

$$\int h d\hat{w}_- = \int h Q d\hat{w}$$

for all  $h \in L_m^2$ . In fact, since  $QQ^* = I$ , we have  $Q\Pi = 0$ . From this it also follows that  $\Pi^2 = \Pi$ , and therefore  $\Pi$  is the spectral density of  $z$ . Moreover, the left member of (9.1.27) generates  $\mathbb{H} := \mathbf{H}(w)$  as  $f$  varies over  $L_p^2$ , and, as can be seen by taking  $f = gQ$  with  $g \in L_m^2$ , the closed span of the first term of the right member is  $\mathbf{H}$ . Hence  $\mathbf{H}(z) = \mathbf{H}^\perp := \mathbb{H} \ominus \mathbf{H}$ , as claimed. Consequently, in view of (9.1.25), for any  $\lambda \in \mathbb{H} = \mathbf{H}(w)$ , there is an  $f \in L_p^2$  such that

$$E^{\mathbf{H}} \lambda = \int f Q^* d\hat{w}_- = \int f Q^* Q d\hat{w},$$

which is precisely (9.1.20). The corresponding statements for  $\bar{w}$  are proven in the same way. It remains to prove (9.1.22) and (9.1.24). To this end, note that we have

$$E^{\mathbf{H}} = J_{\hat{w}} M_{Q^*Q} J_{\hat{w}}^{-1} = J_{\hat{w}} M_{\bar{Q}^*\bar{Q}} J_{\hat{w}}^{-1}$$

from (9.1.20) and (9.1.23), which in view of (9.1.9) yields  $KQ^*Q = \bar{Q}^*\bar{Q}K$ . Hence (9.1.24) follows. Finally, by (9.1.12) and (9.1.24),

$$d\hat{z} = \bar{\Pi} d\hat{w} = \bar{\Pi} K d\hat{w} = K \Pi d\hat{w} = K d\hat{z},$$

proving (9.1.22).  $\square$

### State space construction

As we shall see, the Markovian triplets  $(W, \bar{W}, K)$  contain all the systems-theoretic information needed for the construction of state space representations of  $y$ . In particular, the structural function  $K$  determines the state space, while  $W$  and  $\bar{W}$  serve as the transfer functions of two stochastic realizations having the same state space.

The following theorem describes the relation between Markovian representations and Markovian triplets  $(W, \bar{W}, K)$ .

**Theorem 9.1.4.** *There is a one-one correspondence between proper Markovian representations  $(\mathbb{H}, U, \mathbf{X})$  of a purely nondeterministic process  $\{y(t)\}_{t \in \mathbb{Z}}$  with a full-rank spectral factor and pairs  $([W, \bar{W}, K], z)$  where  $[W, \bar{W}, K]$  is an equivalence class of Markovian triplets and  $z$  is a vector process (defined mod  $\mathcal{O}$ ) such that  $\mathbf{H}(z) \perp \mathbf{H}$  and with spectral density  $\Pi := I - W^\sharp W$ , where  $W^\sharp = W^* \Phi^{-1}$ . Under this correspondence*

$$\mathbb{H} = \mathbf{H} \oplus \mathbf{H}(z) \tag{9.1.28}$$

and

$$\mathbf{X} = \mathbf{H}^-(w) \cap \mathbf{H}^+(\bar{w}) \tag{9.1.29}$$

where  $(w, \bar{w})$  are the generating processes given by

$$w(t) = \int_{-\pi}^{\pi} e^{i\theta t} W^\sharp d\hat{y} + z(t) \tag{9.1.30}$$

$$\bar{w}(t) = \int_{-\pi}^{\pi} e^{i\theta t} \bar{W}^\sharp d\hat{y} + \int_{-\pi}^{\pi} e^{i\theta t} K d\hat{z} \tag{9.1.31}$$

where  $\bar{W}^\sharp := \bar{W}^* \Phi^{-1}$ .

**Proof.** Given a Markovian representation  $(\mathbb{H}, U, \mathbf{X})$  with generating processes  $(w, \bar{w})$ , we have shown above that there is a unique equivalence class  $[W, \bar{W}, K]$  of Markovian triplets. Suppose  $\Pi := I - W^\sharp W$ . Since  $\Phi = W_-^* W_-$ , a straightforward calculation shows that  $\Pi$  is as defined in Lemma 9.1.3 and that (9.1.30) is equivalent to (9.1.19) and (9.1.22)–(9.1.23). Since  $d\hat{z} = \Pi d\hat{w}$ , the process  $z$  is uniquely defined mod  $\mathcal{O}$  with the spectral density  $\Pi$ , and, by Lemma 9.1.3, (9.1.28) holds. Since  $(\mathbb{H}, U, \mathbf{X})$  is proper, (9.1.29) is a consequence of the fact that  $\mathbf{X} = \mathbf{S} \cap \bar{\mathbf{S}}$  (Theorem 8.1.1). Conversely, given a triplet  $(W, \bar{W}, K)$  and a process  $z$  with the stated properties, we define  $(w, \bar{w})$  by (9.1.30) and (9.1.31), and we set  $\mathbf{S} := \mathbf{H}^-(w)$  and  $\bar{\mathbf{S}} := \mathbf{H}^+(\bar{w})$ . Then, since  $(W, \bar{W}, K)$  is a Markovian triplet,  $W$  is analytic implying that  $\mathbf{S} \supset \mathbf{H}^-$ ,  $\bar{W}$  is coanalytic implying that  $\bar{\mathbf{S}} \supset \mathbf{H}^+$ , and  $K$  is inner which is equivalent to perpendicular intersection. Hence, by Theorem 8.1.1,  $\mathbf{X} = \mathbf{S} \cap \bar{\mathbf{S}}$  is a Markovian splitting subspace with ambient space  $\mathbb{H} = \mathbf{H} \oplus \mathbf{H}(z)$ , for the invariance condition (ii) is trivially satisfied. The shift is  $U := U_y \times U_z$ , where  $U_y$  and  $U_z$  are the shifts defined by  $y$  and  $z$ , respectively.  $\square$

At this point we have designed a spectral-domain framework, isomorphic to the geometric framework of Markovian representations, in which all random vari-

ables have concrete representations as functions in certain subspaces of  $H_p^2$  or  $\bar{H}_p^2$ . We shall next introduce a general functional model for Markovian splitting subspaces which is of the type studied in [69] and [34] in connection with deterministic scattering theory and linear systems in Hilbert space. Using this representation the characterization of various structural conditions of Markovian splitting subspaces (observability, constructibility and minimality) can be formulated in Hardy space terms. These questions will be studied in the next section.

**Theorem 9.1.5.** *Let  $(\mathbb{H}, U, \mathbf{X})$  be a proper Markovian representation with structural function  $K$  and generating processes  $(w, \bar{w})$ . Then,*

$$\mathbf{X} = \int_{-\pi}^{\pi} z^{-1} H(K) d\hat{w} = \int_{-\pi}^{\pi} \bar{H}(K^*) d\hat{w}, \quad (9.1.32)$$

where

$$H(K) := H_p^2 \ominus H_p^2 K = H_p^2 \cap (z\bar{H}_p^2 K) \quad (9.1.33)$$

and

$$\bar{H}(K^*) := \bar{H}_p^2 \ominus \bar{H}_p^2 K^* = \bar{H}_p^2 \cap (z^{-1}H_p^2 K^*). \quad (9.1.34)$$

Moreover,  $\mathbf{X}$  is finite dimensional if and only if  $K$  is rational, in which case  $\dim \mathbf{X}$  equals the McMillan degree of  $K$ .

**Proof.** By Theorem 8.1.1 and the fact that  $(\mathbb{H}, U, \mathbf{X})$  is proper,

$$\mathbf{X} = \mathbf{S} \ominus \bar{\mathbf{S}}^\perp = \mathbf{H}^-(w) \ominus \mathbf{H}^-(\bar{w})$$

Therefore the first of equations (9.4.7) follows from (9.1.7) and (9.1.9). A symmetric argument yields the second equation (9.4.7). Relations (9.1.33) and (9.1.34) are standard in Hardy space theory, and in our setting of splitting geometry they correspond to  $\mathbf{X} = \mathbf{S} \cap \bar{\mathbf{S}} = \mathbf{H}^-(w) \cap \mathbf{H}^+(\bar{w})$ . In the same way,  $\mathbf{X} = \mathbf{E}^{\mathbf{S}} \bar{\mathbf{S}}$  yields  $H(K) = \overline{\text{Im } H_K}$ , where  $H_K : H_p^2 \rightarrow \bar{H}_p^2$  is the Hankel operator  $f \mapsto P^{H(K)} f K$ . Now, the range of a Hankel operator is finite-dimensional if and only if its symbol is rational [34, Theorem 3-8, p. 256]. Hence,  $H(K)$  is finite-dimensional if and only if  $K$  is rational. It remains to prove that  $\dim \mathbf{X} = \deg K$ . To this end, suppose that  $\dim \mathbf{X} = n < \infty$ . Then the Markovian representation  $(\mathbb{H}, U, \mathbf{X})$  has an  $n$ -dimensional forward system (8.3.4) and an  $n$ -dimensional backward system (8.3.11) (Theorem 8.3.1), since  $(\mathbb{H}, U, \mathbf{X})$  is proper,  $(A, B)$  and  $(A', \bar{B})$  are both reachable (Theorem 8.4.8). Then, it is shown in Section 6.4 that the structural function  $K$  is given by (6.4.11); i.e.,

$$K(z) = \bar{B}'(zI - A)^{-1}B + V. \quad (9.1.35)$$

Since  $(\bar{B}, A)$  is observable and  $(A, B)$  is reachable, (9.1.35) is a minimal realization, and hence  $\deg K = n$ , as claimed.  $\square$

We shall refer to  $H(K)$  as the *coinvariant subspace* of  $H_p^2$  corresponding to the inner function  $K$ .

**Corollary 9.1.6.** *If the inner function  $K$  is given by (9.1.35), then the rows of  $z(zI - A)^{-1}B$  form a basis in  $H(K)$ , and the rows of  $(z^{-1}I - A')^{-1}\bar{B}$  form a basis in  $\bar{H}(K^*)$ .*

**Proof.** This follows immediately from the fact that  $J_{\hat{w}}((zI - A)^{-1}B) = x(0)$ , which is a basis in  $\mathbf{X}$  by construction. The rest follows by symmetry.

Alternatively, a direct Hardy-space proof can be obtained by using (6.4.15), by which

$$a'(I - Az^{-1})^{-1}B = a'P(z^{-1}I - A')^{-1}\bar{B}K(z)$$

for all  $a \in \mathbb{R}^n$ . Here the left member belongs to  $H_p^2$  and the right member belong to  $z\bar{H}_p^2K$ . Thus they belong to  $H(K)$  by (9.1.33). Since  $(A, B)$  is reachable, the  $n$  rows of  $z(zI - A)^{-1}B$  are linearly independent and hence form a basis in the  $n$ -dimensional space  $H(K) = zJ_{\hat{w}}^{-1}\mathbf{X}$ .  $\square$

## 9.2 Minimality of Markovian representations

The interplay between the the past and the future of a purely nondeterministic process  $\{y(t)\}_{t \in \mathbb{Z}}$  with a full-rank spectral density can be described by the ‘‘all-pass filter’’

$$\text{innovation process } \xrightarrow{w_-} \boxed{\Theta} \xrightarrow{\bar{w}_+} \text{backward innovation process}$$

transforming the forward innovation process  $w_-$  to the backward one  $\bar{w}_+$ . The transfer function

$$\Theta := \bar{W}_+^{-1}W_- \tag{9.2.1}$$

is called the *phase function*.

**Lemma 9.2.1.** *Let  $\Theta$  be given by (9.2.1). Then*

$$\bar{w}_+(t) = \int_{-\pi}^{\pi} e^{i\theta t} \Theta(e^{i\theta}) d\hat{w}_-$$

**Proof.** This is an immediate consequence of the fact that

$$d\hat{y} = W_- d\hat{w}_- = \bar{W}_+ d\hat{w}_+,$$

from which it follows that  $d\hat{w}_+ = \Theta d\hat{w}_-$ .  $\square$

If  $(\mathbb{H}, U, \mathbf{X})$  is a proper Markovian representation with inner triplet  $(K, Q, \bar{Q})$  and generating functions  $(w, \bar{w})$ , then, in view of (9.1.16) and (9.1.17),

$$\Theta = \bar{Q}KQ^*, \tag{9.2.2}$$

or, in block-diagram form

$$\xrightarrow{w_-} \boxed{Q^*} \xrightarrow{\bar{w}} \boxed{K} \xrightarrow{\bar{w}} \boxed{\bar{Q}} \xrightarrow{\bar{w}_+}$$

The Markovian representation is minimal if and only if there are no cancellations in this factorization, as can be seen from Theorem 7.4.10 and the following result.

**Theorem 9.2.2.** *Let  $(\mathbb{H}, U, \mathbf{X})$  is a proper Markovian representation with inner triplet  $(K, Q, \bar{Q})$ . Then  $\mathbf{X}$  is constructible if and only if  $K$  and  $Q$  are right coprime, i.e. they have no nontrivial common right inner factor, and  $\mathbf{X}$  is observable if and only if  $K^*$  and  $\bar{Q}$  are right coprime, i.e. they have no nontrivial common right conjugate inner factor.*

**Proof.** By Theorem 7.4.9,  $\mathbf{X}$  is constructible if and only if  $\mathbf{S} = \mathbf{H}^- \vee \bar{\mathbf{S}}^\perp$ , i.e.  $\mathbf{H}^-(w) = \mathbf{H}^-(w_-) \vee \mathbf{H}^-(\bar{w})$ , which under the isomorphism  $zJ_w$  takes the form

$$H_p^2 = (H_m^2 Q) \vee (H_p^2 K) \quad (9.2.3)$$

For (9.2.3) to hold,  $Q$  and  $K$  must clearly be right coprime. Conversely, suppose that  $Q$  and  $K$  are right coprime, and consider the right member of (9.2.3). Clearly it is a full-range invariant subspace of  $H_p^2$ , because  $H_p^2 K$  is, and therefore, by the Beurling-Lax Theorem (Theorem 4.6.4), it has the form  $H_p^2 J$  where  $J$  is inner. But then  $J$  must be a common right inner factor of  $Q$  and  $K$ , and hence  $J = I$ , concluding the proof of the constructibility criterion. A symmetric argument proves the observability part.  $\square$

In particular, the predictor space  $\mathbf{X}_-$  has the inner triplet  $(K_-, I, \bar{Q}_-)$ , where  $K_-$  and  $\bar{Q}_-$  are left coprime and can be determined from

$$\bar{Q}_- K_- = \Theta \quad (9.2.4)$$

by coprime factorization. Likewise,  $\mathbf{X}_+$  has the inner triplet  $(K_+, Q_+, I)$ , where  $K_+$  and  $Q_+$  are right coprime and can be determined from

$$K_+ Q_+^* = \Theta. \quad (9.2.5)$$

Theorem 9.2.2 allows us to interpret minimality in terms of the factorization (9.2.2) of the phase function  $\Theta$ . In fact, by Theorem 7.4.10,  $\mathbf{X}$  is minimal if and only if this factorization is reduced as far as possible in the sense that no further cancellations are possible. The reduction procedure of Theorem 7.4.3 could be interpreted in terms of such cancellations.

**Corollary 9.2.3.** *Let  $(\mathbb{H}, U, \mathbf{X})$  be an observable proper Markovian representation with analytic spectral factor  $W$ . Then its Markovian triplet  $(W, \bar{W}, K)$  is tight and  $\bar{W}$  and  $K$  are the unique (mod  $O$ ) coprime factors of*

$$W = \bar{W} K \quad (9.2.6)$$

*such that  $\bar{W}$  is  $m \times p$  coanalytic and  $K$  is  $p \times p$  inner. Similarly, if  $\mathbf{X}$  is constructible with coanalytic spectral factor  $\bar{W}$ , its Markovian triplet  $(W, \bar{W}, K)$  is tight, and  $W$  and  $K^*$  are the unique (mod  $O$ ) coprime factors of*

$$\bar{W} = W K^*. \quad (9.2.7)$$

Conversely,  $(\mathbb{H}, U, \mathbf{X})$  is observable if the factorization (9.2.6) is coprime and constructible if (9.2.7) is coprime.

The relations (9.2.6) and (9.2.7) are known as the *Douglas-Shapiro-Shields factorization* [24].

### Strictly noncyclic processes and properness

To be able to apply these results we need criteria for determining when Markovian representations are proper. To this end, first recall from Chapters 6 and 7 that the interface between the past and the future of a stationary vector process  $\{y(t)\}_{t \in \mathbb{Z}}$  can be characterized by the Hankel operators

$$\mathcal{H} = E^{\mathbf{H}^+} |_{\mathbf{H}^-} \quad \text{and} \quad \mathcal{H}^* = E^{\mathbf{H}^-} |_{\mathbf{H}^+}. \quad (9.2.8)$$

**Definition 9.2.4.** The process  $\{y(t)\}_{t \in \mathbb{Z}}$  is *strictly noncyclic* if

$$\bigvee_{t=-\infty}^0 U^t \ker \mathcal{H} = \mathbf{H} \quad \text{and} \quad \bigvee_{t=0}^{\infty} U^t \ker \mathcal{H}^* = \mathbf{H}; \quad (9.2.9)$$

i.e., both  $\ker \mathcal{H}$  and  $\ker \mathcal{H}^*$  have full range.

Since  $\ker \mathcal{H} = \mathbf{N}^+$  and  $\ker \mathcal{H}^* = \mathbf{N}^-$ , where

$$\mathbf{N}^- := \mathbf{H}^- \cap (\mathbf{H}^+)^{\perp} \quad \text{and} \quad \mathbf{N}^+ := \mathbf{H}^+ \cap (\mathbf{H}^-)^{\perp}, \quad (9.2.10)$$

and since  $\mathbf{H}^{\square} \sim ((\mathbf{N}^+)^{\perp}, (\mathbf{N}^-)^{\perp})$ , we have the following alternative characterization of noncyclicity.

**Proposition 9.2.5.** *A process  $\{y(t)\}_{t \in \mathbb{Z}}$  is strictly noncyclic if and only if both  $\mathbf{N}^+$  and  $\mathbf{N}^-$ , defined by (9.2.10), have full range; i.e., if and only if the frame space  $\mathbf{H}^{\square}$  is proper.*

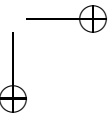
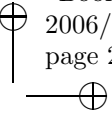
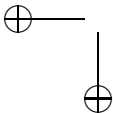
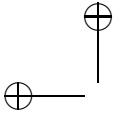
**Lemma 9.2.6.** *Suppose that  $\{y(t)\}_{t \in \mathbb{Z}}$  is purely nondeterministic and full-rank. Then  $\ker \mathcal{H}$  is full range if and only if  $\ker \mathcal{H}^*$  is full range.*

**Proof.** By Proposition 4.5.11,  $y$  is reversible and hence purely nondeterministic in both directions. Setting  $\tilde{\Theta} := z\Theta$ , a simple calculation based on Lemma 9.2.1 shows that, under the isomorphism  $J_{w_-} z^{-1}$ ,  $\mathcal{H}^*$  corresponds to

$$H_{\tilde{\Theta}} := P^{H_p^2} M_{\tilde{\Theta}} |_{\tilde{H}_p^2},$$

where  $P^{H_p^2}$  is the orthogonal projection onto  $H_p^2$ . The operator  $H_{\tilde{\Theta}}$  is called the Hankel operator with symbol  $\tilde{\Theta}$ . In the same way, it is seen that  $\mathcal{H}$  corresponds, under the isomorphism  $J_{\bar{w}_+}$ , to

$$P^{\tilde{H}_p^2} M_{\tilde{\Theta}^*} |_{H_p^2},$$



which in turn is isomorphic to  $H_{\bar{\Theta}}$ , under conjugation. However, by Corollary 3-6(c) in [34, p. 256],  $H_{\bar{\Theta}}$  has a full range kernel if and only if  $H_{\bar{\Theta}}$  does, and hence lemma follows.  $\square$

We say that  $\Theta$  is a *strictly noncyclic function* if the conditions of Lemma 9.2.6 are satisfied [34].

Then the following proposition is immediate.

**Proposition 9.2.7.** *A process  $\{y(t)\}_{t \in \mathbb{Z}}$  with a full-rank spectral density is strictly noncyclic if and only if it is purely nondeterministic and either  $\mathbf{N}^-$  or  $\mathbf{N}^+$  is full range in  $\mathbf{H}$ .*

**Corollary 9.2.8.** *A process  $\{y(t)\}_{t \in \mathbb{Z}}$  with a full-rank spectral density is strictly noncyclic if and only if it is purely nondeterministic and any of the Markovian splitting subspaces  $\mathbf{X}_-$  and  $\mathbf{X}_+$  is proper.*

**Proof.** Since  $\mathbf{X}_- \sim (\mathbf{H}^-, (\mathbf{N}^-)^\perp)$  and  $\mathbf{X}_+ \sim ((\mathbf{N}^+)^\perp, \mathbf{H}^+)$ , the lemma follows from Proposition 9.2.7.  $\square$

**Proposition 9.2.9.** *Let  $\{y(t)\}_{t \in \mathbb{Z}}$  be purely nondeterministic with a full rank spectral density and phase function  $\Theta := \bar{W}_+^{-1}W_-$ . Then  $y$  is strictly noncyclic if and only if one of the following equivalent conditions hold:*

(i) *There are square inner functions  $J_1$  and  $J_2$  such that*

$$\Theta = J_1 J_2^*. \tag{9.2.11}$$

(ii) *There are square inner functions  $J_3$  and  $J_4$  such that*

$$\Theta = J_3^* J_4. \tag{9.2.12}$$

**Proof.** If  $y$  is strictly noncyclic,  $\mathbf{X}_-$  is proper (Corollary 9.2.8). Hence  $\mathbf{X}_-$  has an inner triplet  $(K_-, I, \bar{Q}_-)$ , Therefore (9.2.2) yields  $\Theta = \bar{Q}_- K_-$ , which has the form (9.2.11), as claimed. Conversely, suppose that condition (i) holds. Then  $W := W_- J_2$  is an analytic spectral factor. Define the corresponding white noise process via  $d\hat{w} = W^{-1}d\hat{y}$ . Then  $d\hat{w}_- = J_2 d\hat{w}$ , and hence  $\mathbf{H}(w) = \mathbf{H}$ . Moreover, using the notation (9.1.6),  $\mathcal{J}_{\hat{w}}^{-1} \mathcal{J}_{w_-} = M_{J_2}$ . Since  $J_2$  is inner,  $z^{-1}H_m^2 J_2 \subset z^{-1}H_m^2$ ; i.e.,  $\mathcal{J}_{\hat{w}}^{-1} \mathcal{J}_{w_-} z^{-1}H_m^2 \subset z^{-1}H_m^2$ , which is equivalent to  $\mathbf{H}^- := \mathcal{J}_{w_-} z^{-1}H_m^2 \subset \mathcal{J}_{\hat{w}} z^{-1}H_m^2 =: \mathbf{H}^-(w)$ . Therefore,  $\mathbf{H}^+(w) \subset (\mathbf{H}^-)^\perp$ . Likewise, in view of (9.2.11),  $d\hat{w} = J_1^* d\hat{w}_+$ ; i.e.,  $\mathcal{J}_{\bar{w}_+}^{-1} \mathcal{J}_{\hat{w}} = M_{J_1^*}$ . Since  $J_1^*$  is conjugate inner,  $\bar{H}_m^2 J_1^* \subset \bar{H}_m^2$ ; i.e.,  $\mathcal{J}_{\bar{w}_+}^{-1} \mathcal{J}_{\hat{w}} \bar{H}_m^2 \subset \bar{H}_m^2$ , from which we have  $\mathbf{H}^+(w) = \mathcal{J}_{\hat{w}} \bar{H}_m^2 \subset \mathcal{J}_{\bar{w}_+} \bar{H}_m^2 = \mathbf{H}^+$ . Consequently,  $\mathbf{H}^+(w) \subset \mathbf{H}^+ \cap (\mathbf{H}^-)^\perp =: \mathbf{N}^+$ . However,  $\mathbf{H}^+(w)$  is full range, and therefore so is  $\mathbf{N}^+$ . By a symmetric argument involving  $\mathbf{X}_+$  and the coanalytic spectral factor  $\bar{W} := \bar{W}_+ J_3^*$ , we show that  $y$  being strictly noncyclic is equivalent to (ii).  $\square$



We are now in a position to formulate the following main result.

**Theorem 9.2.10.** *The process  $\{y(t)\}_{t \in \mathbb{Z}}$  is strictly noncyclic if and only if all minimal Markovian representations are proper.*

In view of Corollary 7.4.14, Theorem 9.2.10 is a corollary of the following more general result.

**Theorem 9.2.11.** *Let  $\{y(t)\}_{t \in \mathbb{Z}}$  be strictly noncyclic, and let  $(\mathbb{H}, U, \mathbf{X})$  be a Markovian representation with  $\mathbf{X} \sim (\mathbf{S}, \bar{\mathbf{S}})$  having remote past  $\mathbf{S}_{-\infty}$  and remote future  $\bar{\mathbf{S}}_{\infty}$ . Then  $\mathbf{S}_{-\infty} = 0$  if  $\mathbf{X} \perp \mathbf{N}^+$  and  $\bar{\mathbf{S}}_{\infty} = 0$  if  $\mathbf{X} \perp \mathbf{N}^-$ .*

**Proof.** Since  $y$  is strictly noncyclic, both  $\mathbf{N}^-$  and  $\mathbf{N}^+$  are full range (Proposition 9.2.5). Set  $\mathbf{S}_+ := (\mathbf{N}^+)^{\perp}$ , where orthogonal complement is with respect to the ambient space  $\mathbf{H}$ . If  $\mathbf{X} \perp \mathbf{N}^+$ , we have  $\mathbf{X} \subset \mathbf{S}_+ \oplus \mathbf{H}^{\perp}$ , the right side of which contains  $\mathbf{H}^-$  and is invariant under the backward shift  $U^{-1}$ . Consequently, we have  $\mathbf{S} \subset \mathbf{S}_+ \oplus \mathbf{H}^{\perp}$  so that

$$\mathbf{S}_{-\infty} \subset U^t \mathbf{S} \subset U^t \mathbf{S}_+ \oplus \mathbf{H}^{\perp} \rightarrow \mathbf{H}^{\perp} \quad \text{as } t \rightarrow -\infty.$$

In fact, since  $\mathbf{N}^+$  is full range,  $U^t \mathbf{S}_+ \rightarrow 0$  as  $t \rightarrow -\infty$ . Hence, since  $\mathbf{S}_{-\infty} \subset \mathbf{S}$ ,

$$\mathbf{S}_{-\infty} \subset \mathbf{S} \cap \mathbf{H}^{\perp} \subset \mathbf{S} \cap (\mathbf{H}^+)^{\perp}, \tag{9.2.13}$$

where we have also used the fact that  $\mathbf{H}^{\perp} \subset (\mathbf{H}^+)^{\perp}$ . In the same way we deduce from  $\mathbf{X} \perp \mathbf{N}^-$ ,  $\bar{\mathbf{S}}_{\infty} \subset \bar{\mathbf{S}}$  and  $\mathbf{H}^{\perp} \subset (\mathbf{H}^-)^{\perp}$  that

$$\bar{\mathbf{S}}_{\infty} \subset \bar{\mathbf{S}} \cap \mathbf{H}^{\perp} \subset \bar{\mathbf{S}} \cap (\mathbf{H}^-)^{\perp}. \tag{9.2.14}$$

Now, suppose that, in addition to being orthogonal to  $\mathbf{N}^+$ ,  $\mathbf{X}$  is also observable so that  $\bar{\mathbf{S}}^{\perp} = \mathbf{S} \cap (\mathbf{H}^+)^{\perp}$  (Theorem 8.1.1). Then (9.2.13) implies that  $\mathbf{S}_{-\infty} \subset \bar{\mathbf{S}}^{\perp}$ , and hence

$$\mathbf{S}_{-\infty} = U^t \mathbf{S}_{-\infty} \subset U^t \bar{\mathbf{S}}^{\perp} \rightarrow 0 \quad \text{as } t \rightarrow -\infty.$$

In fact, by Theorem 8.1.3,  $\bar{\mathbf{S}}^{\perp} = \mathbf{H}^-(\bar{w})$ . Consequently,  $\mathbf{S}_{-\infty} = 0$ . By a symmetric argument, we see that  $\mathbf{X} \perp \mathbf{N}^-$  and constructibility imply that  $\bar{\mathbf{S}}_{\infty} = 0$ .

However, observability and constructibility are not needed, as can be seen by applying Theorem 7.4.3. In fact, if  $\mathbf{X} \perp \mathbf{N}^-$ , but  $\mathbf{X}$  is not constructible, define  $\mathbf{S}_1 := \mathbf{H}^- \vee \bar{\mathbf{S}}^{\perp}$ . Then, by Corollary 7.4.4,  $\mathbf{X}_1 \sim (\mathbf{S}_1, \bar{\mathbf{S}})$  is a Markovian splitting subspace. Moreover, by (9.2.14), we have

$$\bar{\mathbf{S}}_{\infty} = U^t \bar{\mathbf{S}}_{\infty} \subset U^t [\bar{\mathbf{S}} \cap (\mathbf{H}^-)^{\perp}] = U^t \mathbf{S}_1^{\perp} \rightarrow 0 \quad \text{as } t \rightarrow \infty,$$

since  $\mathbf{S}_1^{\perp} = \mathbf{H}^+(w_1)$ , where  $w_1$  is the forward generating process of  $\mathbf{X}_1$  (Theorem 8.1.3). Therefore  $\bar{\mathbf{S}}_{\infty} = 0$  if  $\mathbf{X} \perp \mathbf{N}^-$ . A symmetric argument shows that observability is not needed, and  $\mathbf{X} \perp \mathbf{N}^+$  implies that  $\mathbf{S}_{-\infty} = 0$ .  $\square$

### The structural functions of minimal Markovian representations

The structural functions of two minimal proper Markovian splitting subspaces may be quite different (in the multivariate case). In fact, they may not even take values in the same space, being matrices of different sizes. If they are finite dimensional, they have the same degree (Theorems 9.1.5 and 7.6.1). In the general case, there are still some important invariants, namely the nontrivial invariant factors. Recall that the *invariant factors* of a  $p \times p$  inner function  $K$  are  $p$  scalar inner functions  $k_1, k_2, \dots, k_p$  defined in the following way. Set  $\gamma_0 = 1$ , and, for  $i = 1, 2, \dots, p$  define  $\gamma_i$  to be the greatest common inner divisor of all  $i \times i$  minors of  $K$ . Then set  $k_i := \gamma_i / \gamma_{i-1}$  for  $i = 1, 2, \dots, p$ . Clearly, these functions are inner, for  $\gamma_{i-1}$  divides  $\gamma_i$ .

**Theorem 9.2.12.** *Suppose that  $\{y(t)\}_{t \in \mathbb{Z}}$  is strictly noncyclic. Then all internal minimal Markovian splitting subspaces have the same invariant factors; let us denote them*

$$k_1, k_2, k_3, \dots, k_m. \tag{9.2.15}$$

*Moreover, a Markovian splitting subspace of multiplicity  $p$  is minimal if and only if  $m$  invariant factors are given by (10.2.24) and the remaining  $p - m$  are identically one.*

**Proof.** Let  $\mathbf{X}$  be an arbitrary minimal Markovian splitting subspace with structural function  $K$  and multiplicity  $p$ . Since  $\mathbf{X}$  is proper (Theorem 9.2.10), it has generating processes  $(w, \bar{w})$ . Let  $K_+$  denote the structural function of  $\mathbf{X}_+$ , which of course has multiplicity  $m$ , being internal. By Corollaries 7.4.14 and 7.6.6, we have  $U(\mathbf{X})\hat{\mathcal{O}}^* = \hat{\mathcal{O}}^*U(\mathbf{X}_+)$ , where  $\hat{\mathcal{O}}^*$  is a quasi-invertible transformation. Now,  $U_t(X)J_w = J_w S_t(K)$ , where  $S_t(K)$  is the shift  $e^{i\theta}$  in  $H_p^2$  compressed to  $H(K)$ , and therefore  $U_t(\mathbf{X})$  is similar to  $S_t(K)$ . Similarly,  $U_t(\mathbf{X}_+)$  is similar to  $S_t(K_+)$ , but it is a simple calculation to see that it is also similar to

$$\hat{K}_+ = \begin{bmatrix} K_+ & 0 \\ 0 & I_{p-m} \end{bmatrix} \tag{9.2.16}$$

where  $I_k$  is the  $k \times k$  identity. Then the inner functions  $\hat{K}_+$  and  $K$  have the same size,  $p \times p$ , and there is a quasi-invertible transformation  $T$  such that  $S_t(\hat{K}_+)T = TS_t(K)$ . Therefore, we can apply Theorem 4 in [95] to see that  $\hat{K}_+$  and  $K$  are quasi-equivalent, which is equivalent to having the same invariant factors [34]. Conversely, we want to show that any  $\mathbf{X} \sim (\mathbf{S}, \bar{\mathbf{S}})$  whose structural function is quasi-equivalent to  $\hat{K}_+$  is minimal. To this end, apply the two-step reduction algorithm of Theorem 7.4.3 to  $\mathbf{X}$ . First consider the Markovian splitting subspace  $\mathbf{X}_0 \sim (\mathbf{S}, \bar{\mathbf{S}}_1)$  obtained after the first step. Then  $\mathbf{X}_0 \subset \mathbf{X}$ , and hence, since they have the same  $\mathbf{S}$ -space,  $H(K_0) \subset H(K)$ , where  $K_0$  is the structural function of  $\mathbf{X}_0$  (Theorem 9.1.5). Therefore  $H_p^2 K \subset H_p^2 K_0$  so there must be an inner function  $J$  such that  $K = JK_0$ . Next, consider  $\mathbf{X}_1 \sim (\mathbf{S}_1, \bar{\mathbf{S}}_1)$  with structural function  $K_1$ , obtained in the second step. Then  $\mathbf{X}_1$  is minimal and  $\mathbf{X}_1 \subset \mathbf{X}_0$ , and therefore  $\bar{H}(K_1^*) \subset \bar{H}(K_0^*)$ , for  $\mathbf{X}_0$  and  $\mathbf{X}_1$  have the same  $\bar{\mathbf{S}}$ -space. Consequently,  $\bar{H}_p^2 K_0^* \subset \bar{H}_p^2 K_1^*$ , and hence there is

a conjugate inner function  $\bar{J}$  such that  $K_0^* = \bar{J}K_1^*$ , i.e.  $K_0 = K_1\bar{J}^*$ . Combining the two factorizations we obtain

$$K = JK_1\bar{J}^* \tag{9.2.17}$$

where both  $J$  and  $\bar{J}^*$  are inner. In particular,

$$\det K = \det J \cdot \det K_1 \cdot \det \bar{J}^*$$

i.e. a product of scalar inner functions. However,  $\mathbf{X}_1$  is minimal and hence, by the first part of the proof,  $K_1$  has the same invariant factors as  $\hat{K}_+$ , and, by assumption, as  $K$ . Therefore,  $\det K = \det K_1$ , and consequently,  $\det J = \det \bar{J}^* = 1$ , which implies that  $J = \bar{J}^* = I$ . This implies that  $\mathbf{X}_1 = \mathbf{X}_0 = \mathbf{X}$ , proving that  $\mathbf{X}$  is minimal.  $\square$

**Corollary 9.2.13.** *If  $\{y(t)\}_{t \in \mathbb{Z}}$  is a scalar, strictly noncyclic process, all internal minimal Markovian splitting subspaces have the same structural function.*

We are also in a position to state a result that strengthens the state-space isomorphism result of Section 8.7.

**Corollary 9.2.14.** *Let  $\mathbf{X}_1$  and  $\mathbf{X}_2$  be two minimal Markovian splitting subspaces. Then  $U_t(\mathbf{X}_1)$  and  $U_t(\mathbf{X}_2)$  are quasi-similar; i.e., there are quasi-invertible (injective with dense range) linear operators  $P : \mathbf{X}_1 \rightarrow \mathbf{X}_2$  and  $R : \mathbf{X}_2 \rightarrow \mathbf{X}_1$  such that*

$$\begin{cases} PU(\mathbf{X}_1) = U(\mathbf{X}_2)P \\ U(\mathbf{X}_1)R = RU(\mathbf{X}_2). \end{cases}$$

*In the finite-dimensional case, they are similar.*

**Example 9.2.15.** Consider a purely nondeterministic process  $y$  with the spectral density of Example 6.8.5. Clearly the two internal minimal Markovian splitting subspaces  $\mathbf{X}_-$  and  $\mathbf{X}_+$  have the same structural function

$$K_-(z) = K_+(z) = \frac{1 - \frac{1}{2}z}{z - \frac{1}{2}}.$$

The noninternal minimal Markovian representation corresponding to the analytic spectral factor

$$W(z) = \left( \frac{\frac{2}{\sqrt{3}}}{z - \frac{1}{2}} + \frac{1}{\sqrt{3}}, \frac{\frac{1}{\sqrt{6}}}{z - \frac{1}{2}} \right)$$

has the structural function

$$K(z) = \frac{1}{z - \frac{1}{2}} \begin{bmatrix} \sqrt{3}(3 - 2z) & \sqrt{6}z \\ -\sqrt{6} & \sqrt{3}(3z - 2) \end{bmatrix}.$$

This follows from (6.4.11) after determining

$$V = \begin{bmatrix} -1/\sqrt{3} & 1/\sqrt{6} \\ 0 & \sqrt{3}/2 \end{bmatrix} \quad \text{and} \quad \bar{B} = [1/2 \quad -\sqrt{2}/4]$$

from (6.4.3), where the signs in the factorizations are chosen to conform with (6.4.8a). The structural function  $K$  has invariant factors

$$k_1 = 1, \quad k_2 = \det K = \frac{1 - \frac{1}{2}z}{z - \frac{1}{2}}$$

in harmony with Theorem 9.2.12.

**Example 9.2.16.** Suppose  $y$  is a purely nondeterministic vector process with spectral density

$$\Phi(z) = \begin{bmatrix} 1 + \frac{(z-\frac{1}{2})(z-\frac{1}{3})(z^{-1}-\frac{1}{2})(z^{-1}-\frac{1}{3})}{(z-\frac{2}{3})(z^{-1}-\frac{2}{3})} & \frac{1}{1-\frac{1}{4}z} \\ \frac{1}{1-\frac{1}{4}z^{-1}} & \frac{1}{(1-\frac{1}{4}z)(1-\frac{1}{4}z^{-1})} \end{bmatrix}.$$

Then, uniquely defined modulo a right orthogonal transformations,

$$W_-(z) = \begin{bmatrix} \frac{(z-\frac{1}{2})(z-\frac{1}{3})}{z(z-\frac{2}{3})} & 1 \\ 0 & \frac{z}{z-\frac{1}{4}} \end{bmatrix}$$

is the outer spectral factor, and

$$\bar{W}_+(z) = \begin{bmatrix} \frac{-\frac{1}{6}z^3 + \frac{13}{6}z^2 - \frac{33}{6}z + \frac{17}{4}}{17(1-\frac{2}{3}z)(1-\frac{1}{4}z)} & \frac{-\frac{1}{24}z^2 - \frac{55}{24}z + \frac{43}{12}}{17(1-\frac{2}{3}z)(1-\frac{1}{4}z)} \\ \frac{-z}{17(1-\frac{1}{4}z)} & \frac{4}{17(1-\frac{1}{4}z)} \end{bmatrix}$$

is the conjugate outer spectral factor, and consequently the phase function (9.2.1) is given by

$$\Theta = \frac{1}{17} \begin{bmatrix} \frac{4(z-\frac{1}{2})(z-\frac{1}{3})(1-\frac{2}{3}z)}{z(1-\frac{1}{2}z)(1-\frac{1}{3}z)(z-\frac{2}{3})} & \frac{-1-\frac{1}{4}z}{z-\frac{1}{4}} \\ \frac{(z-\frac{1}{2})(z-\frac{1}{3})(1-\frac{2}{3}z)}{(1-\frac{1}{2}z)(1-\frac{1}{3}z)(z-\frac{2}{3})} & \frac{4z(1-\frac{1}{4}z)}{z-\frac{1}{4}} \end{bmatrix}.$$

Now, the coprime factorization  $\bar{Q}_-K_- = \Theta$ , prescribed by (9.2.4), yields

$$\bar{Q}_-(z) = \frac{1}{17} \begin{bmatrix} \frac{4(z-\frac{1}{2})(z-\frac{1}{3})}{(1-\frac{1}{2}z)(1-\frac{1}{3}z)} & -1 \\ \frac{z(z-\frac{1}{2})(z-\frac{1}{3})}{(1-\frac{1}{2}z)(1-\frac{1}{3}z)} & 4z \end{bmatrix}$$

and

$$K_-(z) = \begin{bmatrix} \frac{1-\frac{2}{3}z}{z(z-\frac{2}{3})} & 0 \\ 0 & \frac{1-\frac{1}{4}z}{z-\frac{1}{4}} \end{bmatrix},$$

and therefore the coanalytic spectral factor of the predictor space  $\mathbf{X}_-$  is given by

$$\bar{W}_-(z) := \bar{W}_+\bar{Q}_-(z) = \begin{bmatrix} \frac{(z-\frac{1}{2})(z-\frac{1}{3})}{1-\frac{2}{3}z} & \frac{z-\frac{1}{4}}{z-\frac{1}{4}} \\ 0 & \frac{z}{1-\frac{1}{4}z} \end{bmatrix}.$$

Likewise, the coprime factorization (9.2.5); i.e.,  $K_+Q_+^* = \Theta$ , yields

$$K_+(z) := \bar{W}_+(z)^{-1}W_+(z) = \frac{1}{17} \begin{bmatrix} \frac{4(1-\frac{2}{3}z)}{z(z-\frac{2}{3})} & \frac{-(1-\frac{1}{4}z)}{z(z-\frac{1}{4})} \\ \frac{1-\frac{2}{3}z}{z-\frac{2}{3}} & \frac{4(1-\frac{1}{4}z)}{z-\frac{1}{4}} \end{bmatrix}$$

and

$$Q_+(z) := W_-(z)^{-1}W_+(z) = \begin{bmatrix} \frac{(1-\frac{1}{2}z)(1-\frac{1}{3}z)}{(z-\frac{1}{2})(z-\frac{1}{3})} & 0 \\ 0 & \frac{1}{z} \end{bmatrix},$$

and hence we obtain the analytic spectral factor

$$W_+(z) := W_-(z)Q_+(z) = \begin{bmatrix} \frac{(1-\frac{1}{2}z)(1-\frac{1}{3}z)}{z(z-\frac{2}{3})} & \frac{1}{z} \\ 0 & \frac{1}{z-\frac{1}{4}} \end{bmatrix}$$

of the backward predictor space  $\mathbf{X}_+$ . As prescribed by Theorem 9.2.12, the structural functions  $K_-$  and  $K_+$  have the same invariant factors, namely  $k_1 = 1$  and

$$k_2(z) = \frac{(1-\frac{2}{3}z)(1-\frac{1}{4}z)}{z(z-\frac{2}{3})(z-\frac{1}{4})},$$

which is the determinant  $\det K$ . We shall return to this example in Section 9.3. In Chapter 10, we consider an example with nontrivial invariant factor  $k_1$  (Example 10.2.11) in the continuous-time setting.

### A geometric conditions for minimality

We have the following modification of Theorem 7.6.4 in Chapter 7, where, instead of having a closed range,  $\mathcal{H}$  now has a full-range kernel.

**Theorem 9.2.17.** *Let  $y$  be strictly noncyclic. Then, for any Markovian representation  $(\mathbb{H}, U, \mathbf{X})$ , the following conditions are equivalent.*

- (i)  $\mathbf{X}$  is minimal
- (ii)  $\mathbf{X}$  is observable and  $\mathbf{X} \perp \mathbf{N}^+$
- (iii)  $\mathbf{X}$  is constructible and  $\mathbf{X} \perp \mathbf{N}^-$

**Proof.** It follows from Corollary 7.4.14 that each of conditions (ii) and (iii) implies (i). Therefore it remains to show that (ii) or (iii) implies (i). Suppose that (ii) holds. Then, by Corollary 7.6.6,  $U(\mathbf{X})\hat{\Theta}^* = \hat{\Theta}^*U(\mathbf{X}_+)$ , where  $\hat{\Theta}^*$  is quasi-invertible. Now, since  $\mathbf{X}$  is observable  $\mathbf{X} \perp \mathbf{N}^-$  (Corollary 7.4.14), which together with  $\mathbf{X} \perp \mathbf{N}^+$  implies that  $\mathbf{X}$  is proper (Theorem 9.2.11). Hence, as in the proof of Theorem 9.2.12, there is a quasi-invertible transformation  $T$ , such that

$$S_t(\hat{K}_+)T = TS_t(K).$$

Proceeding as in the proof of Theorem 9.2.12, we see that  $K$  and  $\hat{K}_+$  have the same invariant factors and hence that  $\mathbf{X}$  is minimal. A symmetric argument shows that (iii) implies (i) also.  $\square$

Theorem 9.2.2 provides a Hardy space characterization of observability and constructibility. In view of Theorem 9.2.17, it would be desirable to also characterize the conditions  $\mathbf{X} \perp \mathbf{N}^+$  and  $\mathbf{X} \perp \mathbf{N}^-$  in terms of inner functions.

**Proposition 9.2.18.** *Let  $(\mathbb{H}, U, \mathbf{X})$  be a proper Markovian representation with inner triplet  $(K, Q, \bar{Q})$ . Then  $\mathbf{X} \perp \mathbf{N}^+$  if and only if  $Q^*Q_+$  is analytic, and  $\mathbf{X} \perp \mathbf{N}^-$  if and only if  $\bar{Q}^*Q_-$  is coanalytic.*

**Proof.** If  $\mathbf{X} \sim (\mathbf{S}, \bar{\mathbf{S}})$ , then  $\mathbf{X} \perp \mathbf{N}^+$  if and only if  $\mathbf{S} \perp \mathbf{N}^+$ ; i.e.,  $\mathbf{N}^+ \subset \mathbf{S}^\perp$ , or, equivalently

$$\mathbf{H}^+(w_+) \subset \mathbf{H}^+(w). \tag{9.2.18}$$

However,  $d\hat{y} = W_-Qd\hat{w} = W_-Q_+d\hat{w}_+$ , we have  $d\hat{w}_+ = Q_+^*Qd\hat{w}$ . Consequently, (9.2.18) is equivalent to  $\bar{H}_m^2Q_+^*Q \subset \bar{H}_p^2$  under the isomorphism  $\mathcal{J}_{\hat{w}}$ , which hold if and only if  $Q_+^*Q$  is coanalytic; i.e., if and only if  $Q^*Q_+$  is analytic. A symmetric argument proves the second statement.  $\square$

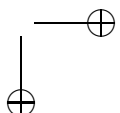
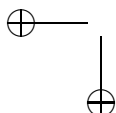
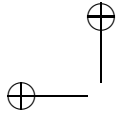
**Corollary 9.2.19.** *Let  $(\mathbb{H}, U, \mathbf{X})$  be a finite-dimensional proper Markovian representation. Then the analytic spectral factor  $W$  of  $\mathbf{X}$  is minimal in the sense of Definition 6.8.3 if and only if  $\mathbf{X} \perp \mathbf{N}^+$ , and the coanalytic spectral factor  $\bar{W}$  is minimal if and only if  $\mathbf{X} \perp \mathbf{N}^-$ .*

**Proof.** Suppose that  $J := Q^*Q_+$  is analytic. Since therefore  $J^*$  is coanalytic and  $W = W_+J^*$  is analytic, all poles of  $J^*$  must be cancelled by zeros of  $W_+$ . Hence  $\deg W \leq \deg W_+$ . But  $W_+$  is minimal, and hence so is  $W$ . Conversely, suppose that  $\deg W = \deg W_+$ . Then  $J^*$  must be coanalytic. Indeed, if  $J^*$  had some pole in the complement of the closed unit disc, then they would have to cancel some zero of  $W_+$  for  $W = W_+J^*$  to be analytic. However,  $W_+$ , being maximum-phase, has all its zeros in the open unit disc. Hence  $J$  must be analytic. Consequently, the corollary follows from Proposition 9.2.18.  $\square$

This result provides a natural infinite-dimensional generalization of minimality of spectral factors.

**Definition 9.2.20.** An analytic spectral factor of a strictly noncyclic process is *minimal* if its inner factor  $Q$  has the property that  $Q^*Q_+$  is analytic. Likewise, a coanalytic spectral factor is *minimal* if its inner factor  $\bar{Q}$  has the property that  $\bar{Q}^*Q_-$  is coanalytic.

We can now state the following corollary of Theorem 9.1.4, which of course has a symmetric “backward” counterpart.



**Corollary 9.2.21.** *Let  $y$  be strictly noncyclic. Then there is a one-one correspondence (mod  $\Theta$ ) between minimal Markovian representations  $(\mathbb{H}, U, \mathbf{X})$  and pairs  $(W, z)$  where  $W$  is a minimal spectral factor and  $z$  is a stationary process with the properties prescribed in Theorem 9.1.4.*

**Proof.** By Theorem 9.2.17,  $\mathbf{X}$  is minimal if and only if  $\mathbf{X}$  is observable and  $\mathbf{S} \perp \mathbf{N}^+$ , i.e.  $W$  is minimal. From the observability condition  $\bar{\mathbf{S}} = \mathbf{H}^+ \vee \mathbf{S}^\perp$  (Theorem 7.4.9) we see that  $\bar{W}$  is determined once  $W$  has been chosen (Corollary 9.2.3).  $\square$

In particular, if  $W$  and  $\bar{W}$  are square,  $W$  is minimal if and only if  $Q$  is a left inner divisor of  $Q_+$ , and  $\bar{W}$  is minimal if and only if  $\bar{Q}$  is a left inner divisor of  $\bar{Q}_-$ .

Consequently, we have a procedure for determining the inner triplets  $(K, Q, \bar{Q})$  of all minimal Markovian representation: First choose  $Q$  so that  $Q^*Q_+$  is analytic. Then form  $T := \Theta Q$ , and determine  $\bar{Q}$  and  $K$  as the coprime factors in

$$\bar{Q}K = T. \tag{9.2.19}$$

### 9.3 Degeneracy

A proper Markovian representation  $(\mathbb{H}, U, \mathbf{X})$  is said to be *degenerate* if its structural function is singular at infinity; i.e., if  $K(\infty)$  is singular. Degeneracy is inherent to the discrete-time setting and, as we shall see in the next chapter, it does not occur for continuous-time Markovian representations. However, in the discrete-time framework, this phenomenon occurs in important subclasses of stochastic systems, a case in point being moving-average processes. As a simple example, take the process  $y$  with spectral density  $\Phi(z) = 5 + 2(z + z^{-1})$ . It has only two internal minimal Markovian splitting subspaces, namely  $\mathbf{X}_-$  corresponding to  $W_-(z) = z^{-1} + 2$  and  $\bar{W}_-(z) = 1 + 2z$  and  $\mathbf{X}_+$  corresponding to  $W_+(z) = 1 + 2z^{-1}$  and  $\bar{W}_+(z) = z + 2$ . As prescribed by Corollary 9.2.13, they have the same structural function, namely  $K(z) = z^{-1}$ , which equals zero at infinity. Hence they are degenerate.

**Proposition 9.3.1.** *If one minimal Markovian representation is degenerate, then all are.*

**Proof.** By Theorem 9.2.12,  $\det K$  is the same for all minimal Markovian representations, and therefore these are degenerate at the same time.  $\square$

Consequently, degeneracy of minimal Markovian representations is a property of the process  $y$ . We shall say that  $y$  is *state-space degenerate* if its minimal Markovian representations are degenerate.

Next we show that  $(\mathbb{H}, U, \mathbf{X})$  is degenerate if and only if  $A$  and  $\bar{A}$  in Theorem 8.7.4 have nontrivial nullspaces. For nondegenerate  $(\mathbb{H}, U, \mathbf{X})$ ,  $A$  and  $\bar{A}$  are quasi-invertible, or, in the finite-dimensional case, invertible.

**Proposition 9.3.2.** *A proper Markovian representation  $(\mathbb{H}, U, \mathbf{X})$  is degenerate if and only if  $\ker U(\mathbf{X}) \neq 0$ , or, equivalently,  $\ker U(\mathbf{X})^* \neq 0$ .*

Since, in view of Lemma 8.8.3,

$$\ker U(\mathbf{X}) = \mathbf{X} \cap (U^* \mathbf{W}) \tag{9.3.1a}$$

$$\ker U(\mathbf{X})^* = \mathbf{X} \cap \bar{\mathbf{W}}, \tag{9.3.1b}$$

Proposition 9.3.2 implies that a degenerate  $\mathbf{X}$  contains some linear combination of the components of  $w(-1)$  and of  $w(0)$ . In particular, some linear functional of the state process  $x(t)$  is white noise. This can also be seen from (8.7.14) by taking  $f \in \ker A^*$  and forming

$$\langle f, x(t+1) \rangle_x = \langle f, Bw(t) \rangle_x,$$

which is white noise, and analogously for  $f \in \ker A$  and (8.7.24).

Proposition 9.3.2 follows immediately by taking  $(W_1, W_2) = (W, \bar{W})$  and  $(w_1, w_2) = (w, \bar{w})$  in the following lemma, which is a variation of [?, Theorem 13].

**Lemma 9.3.3.** *Let  $W_1$  and  $W_2$  be two spectral factors, of dimensions  $m \times p_1$  and  $m \times p_2$ , respectively, and let  $w_1$  and  $w_2$  be vector Wiener processes, of dimensions  $p_1$  and  $p_2$ , such that*

$$y(t) = \int_{-\pi}^{\pi} e^{i\theta t} W_1(e^{i\theta}) d\hat{w}_1 = \int_{-\pi}^{\pi} e^{i\theta t} W_2(e^{i\theta}) d\hat{w}_2. \tag{9.3.2}$$

*Then, if  $\mathbf{H}^-(w_2) \subset \mathbf{H}^-(w_1)$ , there is a  $p_2 \times p_1$  inner function  $R$  such that  $W_1 = W_2 R$ , and*

$$\text{rank } R(\infty) = p_2 - \dim \ker U(\mathbf{Y})^*, \tag{9.3.3}$$

*where  $\mathbf{Y} := \mathbf{H}^-(w_1) \ominus \mathbf{H}^-(w_2)$ . Similarly, if  $\mathbf{H}^+(w_1) \subset \mathbf{H}^+(w_2)$ , there is a  $p_1 \times p_2$  conjugate inner function  $\bar{R}$  such that  $W_2 = W_1 \bar{R}$ , and*

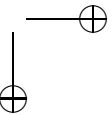
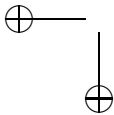
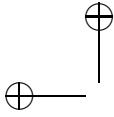
$$\text{rank } \bar{R}^*(\infty) = p_1 - \dim \ker U(\bar{\mathbf{Y}}), \tag{9.3.4}$$

*where  $\bar{\mathbf{Y}} := \mathbf{H}^+(w_2) \ominus \mathbf{H}^+(w_1)$ . If  $p_1 = p_2$ , both conditions hold, and  $\bar{R} = R^*$ .*

**Proof.** Suppose that  $\mathbf{H}^-(w_2) \subset \mathbf{H}^-(w_1)$ . Then  $p_2 \leq p_1$ , and we have  $\mathcal{J}_{w_1}^{-1} \mathcal{J}_{w_2} H_{p_2}^2 \subset H_{p_1}^2$ . However,  $\mathcal{J}_{w_1}^{-1} \mathcal{J}_{w_2}$  is a bounded linear operator that commutes with the shift  $z^{-1}$ , and therefore it must be a multiplication operator  $M_R$  sending  $f \in H_{p_2}^2$  to  $fR \in H_{p_1}^2$  (Theorem 4.3.3). Hence  $H_{p_2}^2 R \subset H_{p_1}^2$ . Since  $\mathbf{H}^-(w_2)$  is invariant under the shift  $U^{-1}$ , the subspace  $H_{p_2}^2 R$  is invariant under  $z^{-1}$ , and therefore  $R$  must be inner (Theorem 4.6.4). In view of (9.3.2), we have  $a'y(0) = \mathcal{J}_{w_1} a'W_1 = \mathcal{J}_{w_2} a'W_2$  for any  $a \in \mathbb{R}$ , and hence  $W_1 = W_2 R$ .

Now, taking orthogonal complements in some common ambient space,  $\mathbf{H}^-(w_1)$  and  $(\mathbf{H}^-(w_2))^\perp$  intersect perpendicularly (Corollary 7.2.5), and hence, in view of Lemma 8.8.3,

$$\ker U(\mathbf{Y})^* = \mathbf{H}^-(w_1) \cap \mathbf{W}_2,$$





where  $\mathbf{W}_2 := \{a'w_2(0) \mid a \in \mathbb{R}^{p_2}\}$ . Then, for any  $\eta \in \ker U(\mathbf{Y})^*$ , there is a vector  $a \in \mathbb{R}^{p_2}$  such that  $\eta = a'w_2(0) \in \mathbf{H}^-(w_1)$ ; i.e.,  $J_{w_1}^{-1}J_{w_2}a' \in z^{-1}H_{p_1}^2$ , or, equivalently,  $a'R \in z^{-1}H_{p_1}^2$ , which holds if and only if  $a'R(\infty) = 0$ . Hence (9.3.3) holds.

Next, suppose that  $\mathbf{H}^+(w_1) \subset \mathbf{H}^+(w_2)$ . Then  $p_1 \leq p_2$ , and we proceed as above to see that there is a conjugate inner function  $\bar{R}$  such that  $\bar{H}_{p_1}^2 \bar{R} \subset \bar{H}_{p_2}^2$  and  $W_2 = W_1 \bar{R}$ . Moreover, from Lemma 8.8.3 we have that

$$\ker U(\bar{\mathbf{Y}}) = \mathbf{H}^+(w_2) \cap (U^* \mathbf{W}_1),$$

where  $\mathbf{W}_1 := \{a'w_1(0) \mid a \in \mathbb{R}^{p_1}\}$ . Therefore, for any  $\eta \in \ker U(\bar{\mathbf{Y}})$ , there is an  $a \in \mathbb{R}^{p_1}$  such that  $\eta = a'w_1(-1) \in \mathbf{H}^+(w_2)$ ; i.e.,  $a'\bar{R} \in z\bar{H}_p^2$ , or, equivalently,  $a'\bar{R}^* \in z^{-1}H_p^2$ , which is the same as  $a'\bar{R}^*(\infty) = 0$ . This establishes (9.3.4). The last statement follows trivially.  $\square$

### Error-space degeneracy

In Section 8.8 we introduced the error spaces

$$\mathbf{Z} := \mathbf{S} \ominus \mathbf{H}^- = \mathbf{H}^-(w) \ominus \mathbf{H}^-(w_-) \tag{9.3.5a}$$

$$\bar{\mathbf{Z}} := \bar{\mathbf{S}} \ominus \mathbf{H}^+ = \mathbf{H}^+(\bar{w}) \ominus \mathbf{H}^-(\bar{w}_+) \tag{9.3.5b}$$

of a proper Markovian splitting subspace  $\mathbf{X} \sim (\mathbf{S}, \bar{\mathbf{S}})$ . Applying Lemma 9.3.3 to these, we obtain the outer-inner factorizations

$$W = W_- Q \quad \text{and} \quad \bar{W} = \bar{W}_+ \bar{Q} \tag{9.3.6}$$

of (9.1.17) and the conditions

$$\text{rank } Q(\infty) = p - \dim \ker U(\mathbf{Z})^*, \tag{9.3.7a}$$

$$\text{rank } \bar{Q}^*(\infty) = p - \dim \ker U(\bar{\mathbf{Z}}). \tag{9.3.7b}$$

In particular, comparing this with Proposition 8.8.2, we have

$$\text{rank } D = \text{rank } \bar{Q}^*(\infty) \quad \text{and} \quad \text{rank } \bar{D} = \text{rank } Q(\infty), \tag{9.3.8}$$

which of course could be obtained directly from (9.3.6) by observing that  $D = W(\infty)$  and  $\bar{D} = \bar{W}^*(\infty)$ . Indeed,

$$D = D_- Q(\infty) \quad \text{and} \quad \bar{D} = \bar{D}_+ \bar{Q}^*(\infty), \tag{9.3.9}$$

where  $D_-$  and  $\bar{D}_+$  have full rank; see page 227.

**Proposition 9.3.4.** *Let  $(\mathbb{H}, U, \mathbf{X})$  be a proper Markovian representation with spectral inner factors  $Q$  and  $\bar{Q}$ . Then the standard forward realization (8.7.14) is regular if and only if  $\bar{Q}^*(\infty)$  has full rank, and the standard backward realization (8.7.24) is regular if and only if  $Q(\infty)$  has full rank.*

We shall say that a Markovian representation (Markovian splitting subspace) is *regular* if both the forward and the backward realization is regular, and *singular* otherwise.

**Proposition 9.3.5.** *If one minimal Markovian representation is singular, then all internal minimal Markovian representation are singular.*

**Proof.** For a singular minimal Markovian representation, either  $DD^*$  or  $\bar{D}\bar{D}^*$  is singular. First suppose that  $DD^*$  is singular. Then, by (8.8.4),  $D_+D_+^*$  is singular, and hence  $Q_+(\infty)$  is singular. Now, consider an arbitrary internal minimal Markovian representation  $(\mathbb{H}, U, \mathbf{X})$  with spectral inner factors  $Q$  and  $\bar{Q}$ . Then, by (9.2.2) and (9.2.5),  $QK^*\bar{Q} = Q_+K_+^*$ , and hence

$$\det Q \det K^* \det \bar{Q}^* = \det Q_+ \det K_+^*.$$

However, by Theorem 9.2.12,  $\det K^* = \det K_+^*$ , and therefore

$$\det Q \det \bar{Q}^* = \det Q_+,$$

so if  $Q_+(\infty)$  is singular, then so is either  $Q(\infty)$  or  $\bar{Q}(\infty)$ , establishing that  $(\mathbb{H}, U, \mathbf{X})$  is singular. By a completely symmetric argument, we can also show that, if  $\bar{D}\bar{D}^*$  is singular, then all internal minimal Markovian representations are singular.  $\square$

The word “internal” in Proposition 9.3.5 is essential. In fact, it is easy to construct a counter example where the internal minimal Markovian representation are singular and the noninternal ones are all regular: Suppose  $y$  has the spectral density  $\Phi(z) = \frac{1}{(z-\frac{1}{2})(z^{-1}-\frac{1}{2})}$ . Then  $Q_-(z) = \bar{Q}_+^*(z) = z^{-1}$ , and hence the only two internal minimal Markovian splitting subspaces,  $\mathbf{X}_-$  and  $\mathbf{X}_+$ , are both singular. However, proceeding along the lines of Example 6.8.5, It is not hard to see that all noninternal minimal Markovian representation are regular.

By Proposition 9.3.5, singularity of minimal Markovian representations is a property of the process  $y$ . We say that  $y$  is *error-space degenerate* if its minimal Markovian representations are singular.

**Corollary 9.3.6.** *A strictly noncyclic process  $y$  is error-space degenerate if and only if  $Q_+(\infty)$  is singular, or, equivalently,  $\bar{Q}_-(\infty)$  is singular.*

### Degenerate processes

Recall from section 7.4 that

$$\mathbf{H} = \mathbf{N}^- \oplus \mathbf{H}^\square \oplus \mathbf{N}^+, \tag{9.3.10}$$

where the frame space  $\mathbf{H}^\square = \mathbf{X}_- \vee \mathbf{X}_+$  is the closed linear hull of all internal minimal splitting subspaces, and  $\mathbf{N}^- := \mathbf{H}^- \cap (\mathbf{H}^+)^\perp$  and  $\mathbf{N}^+ := \mathbf{H}^+ \cap (\mathbf{H}^-)^\perp$  are subspaces generally discarded in minimal state-space construction.

**Definition 9.3.7.** A strictly noncyclic process  $y$  is *degenerate* if its frame space  $\mathbf{H}^\square$  is degenerate.

Since  $\mathbf{H}^\square \sim ((\mathbf{N}^+)^\perp, (\mathbf{N}^-)^\perp)$ , the generating processes of  $\mathbf{H}^\square$  are  $(w_+, \bar{w}_-)$ . In particular, since  $\mathbf{W}_+ \subset \mathbf{N}^+$  and  $U^*\bar{\mathbf{W}}_- \subset \mathbf{N}^-$ , Proposition 9.3.2 and (9.3.1) imply that  $y$  is degenerate if and only if  $(U\mathbf{H}^\square) \cap \mathbf{N}^+ \neq 0$  and  $(U^*\mathbf{H}^\square) \cap \mathbf{N}^- \neq 0$  both hold. Hence degeneracy of  $y$  implies that, as  $\mathbf{H}^\square$  is shifted one step forward or backward in time, some elements of the “discarded” spaces  $\mathbf{N}^-$  and  $\mathbf{N}^+$  become part of the new frame space.

**Theorem 9.3.8.** *A strictly noncyclic process  $y$  is degenerate if and only if it is either state-space degenerate or error-space degenerate or both, If  $y$  is scalar ( $m = 1$ ), both cannot happen at the same time.*

**Proof.** The frame space  $\mathbf{H}^\square$  is proper (Corollary 9.2.8), and its structural function is given by

$$K_\square = \bar{W}_-^{-1}W_+ = \bar{W}_-^{-1}W_-W_-^{-1}W_+ = K_+Q_+.$$

Therefore  $y$  is degenerate if and only either  $\det K_+(\infty) = 0$  or  $\det Q_+(\infty) = 0$  or both conditions hold. However,  $\det K(\infty)$  is the same for all minimal Markovian representations (Theorem 9.2.12), and hence  $\det K_+(\infty) = 0$  if and only if  $y$  is state-space degenerate. Moreover,  $\det Q_+(\infty) = 0$  if and only if  $y$  is error-space degenerate (Corollary 9.3.6). This establishes the first statement.

To prove the second statement, note that  $y$  is state-space degenerate if and only if

$$\ker U(\mathbf{Z}_+) = \mathbf{Z}_+ \cap (U^*\mathbf{W}_+) \neq 0, \tag{9.3.11}$$

by Proposition 9.3.2, and (9.3.1a) and error-space degenerate if and only if

$$\ker U(\mathbf{Z}_+) = \mathbf{Z}_+ \cap (U^*\mathbf{W}_+) \neq 0, \tag{9.3.12}$$

by Corollary 9.3.6, (9.3.7a) and (8.8.7a). However,  $\dim U\mathbf{W}_+ = m = 1$ , and hence (9.3.11) yields  $U\mathbf{W}_+ \subset \mathbf{X}_+$  and (9.3.12) yields  $U\mathbf{W}_+ \subset \mathbf{Z}_+$ . Hence  $\mathbf{X}_+ \cap \mathbf{Z}_+ \neq 0$ . However this contradicts the constructibility of  $\mathbf{X}_+$  (Proposition 8.8.1), and hence the second statement holds.  $\square$

We can illustrate Theorem 9.3.8 by means the forward stochastic realization of the frame space  $\mathbf{H}^\square$ . To this end, first establish some direct-sum decompositions of  $\mathbf{H}^\square$ .

**Proposition 9.3.9.** *Let  $\mathbf{Z}_+$  be the forward error space of the backward predictor space  $\mathbf{X}_+$  and  $\bar{\mathbf{Z}}_-$  be the backward error space of the forward predictor space  $\mathbf{X}_-$ . Then the frame space has the orthogonal decompositions*

$$\mathbf{H}^\square = \mathbf{X}_- \oplus \mathbf{Z}_+ = \mathbf{X}_+ \oplus \bar{\mathbf{Z}}_- \tag{9.3.13}$$

and the direct-sum decompositions

$$\mathbf{H}^\square = \mathbf{X}_+ + \mathbf{Z}_+ = \mathbf{X}_- + \bar{\mathbf{Z}}_-. \tag{9.3.14}$$

**Proof.** By Lemma 2.2.6,  $\mathbf{H}^- = \mathbf{X}_- \oplus \mathbf{N}^-$ . Moreover,  $\mathbf{Z}_+ = (\mathbf{N}^+)^\perp \ominus \mathbf{H}^-$ . Hence,

$$\mathbf{H} = \mathbf{N}^- \oplus \mathbf{X}_- \oplus \mathbf{Z}_+ \oplus \mathbf{N}^+,$$

which compared with (9.3.10) yields the first of equations (9.3.13). The second of equations (9.3.13) is derived by a symmetric argument. Since  $\mathbf{X}_-$  is observable,  $\bar{\mathbf{Z}}_- \cap \mathbf{X}_- = 0$  (Proposition 8.8.1). Taking the orthogonal complement of this in  $\mathbf{H}^\square$  and using (9.3.13), we obtain  $\mathbf{H}^\square = \mathbf{X}_+ \vee \mathbf{Z}_+$ , which must be a direct sum, since  $\mathbf{X}_+$  is constructible, and thus  $\mathbf{X}_- \cap \mathbf{Z}_- = 0$  (Proposition 8.8.1). This establishes the first of decompositions (9.3.14). The proof of the second is analogous.  $\square$

The frame space  $\mathbf{H}^\square$  has generating processes  $(w_+, \bar{w}_-)$ . In particular, it has the same forward generating process as  $\mathbf{X}_+$  and as  $\mathbf{Z}_+ := (\mathbf{N}^+)^\perp \cap (\mathbf{H}^-)^\perp$ , and therefore (9.3.14) yields the forward state equation of  $\mathbf{H}^\square$ ,

$$\begin{bmatrix} x_+(t+1) \\ z_+(t+1) \end{bmatrix} = \begin{bmatrix} A_+ & 0 \\ 0 & F_+ \end{bmatrix} \begin{bmatrix} x_+(t) \\ z_+(t) \end{bmatrix} + \begin{bmatrix} B_+ \\ G_+ \end{bmatrix} w_+(t), \quad (9.3.15)$$

by amending the system (8.8.2) corresponding to  $\mathbf{Z}_+$  to the state equation of (8.7.14) corresponding to  $\mathbf{X}_+$ . Likewise,  $\mathbf{H}^\square$  has the same backward generating process as  $\mathbf{X}_-$  and as  $\bar{\mathbf{Z}}_- := (\mathbf{N}^-)^\perp \cap (\mathbf{H}^+)^\perp$ . Hence the second decomposition (9.3.14) yields backward state equation of  $\mathbf{H}^\square$ ,

$$\begin{bmatrix} \bar{x}_-(t-1) \\ \bar{z}_-(t-1) \end{bmatrix} = \begin{bmatrix} A_-^* & 0 \\ 0 & F_+ \end{bmatrix} \begin{bmatrix} \bar{x}_-(t) \\ \bar{z}_-(t) \end{bmatrix} + \begin{bmatrix} \bar{B}_- \\ \bar{G}_- \end{bmatrix} \bar{w}_-(t). \quad (9.3.16)$$

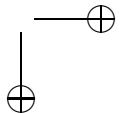
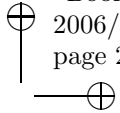
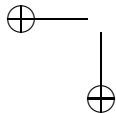
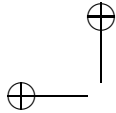
Consequently, for  $\mathbf{H}^\square$ , and hence for  $y$ , to be degenerate, either  $\ker A_+ \neq 0$  or  $\ker F_+ \neq 0$  or both, which is in harmony with Theorem 9.3.8. In fact,  $A_+$  is similar to  $U(\mathbf{X}_+)^*$ , and  $F_+$  is similar to  $U(\mathbf{Z}_+)^*$ , and therefore  $\ker A_+ \neq 0$  if and only if  $K(\infty)$  is singular, and  $\ker F_+ \neq 0$  if and only if  $Q_+(\infty)$  is singular (Lemma 9.3.3), and thus the first statement of Theorem 9.3.8 follows (Corollary 9.3.6). The same conclusion can be drawn from (9.3.16).

Degeneracy can also be described in terms of the behavior of the spectral density at zero and infinity. For simplicity, we shall only consider the rational case, and denote by  $\Phi^{-1}(z)$  the inverse of  $\Phi(z)$  in the field of rational functions.

**Theorem 9.3.10.** *Suppose that  $y$  has a rational,  $m \times m$ , full-rank spectral density  $\Phi$ . (i) If there is a nonzero  $a \in \mathbb{R}^m$  such that  $\Phi(\infty)a = 0$ , then  $y$  is error-space degenerate. If  $y$  is not state-space degenerate, then this condition is also necessary. (ii) If there is a nonzero  $a \in \mathbb{R}^m$  such that  $a'\Phi^{-1}(0) = 0$ , then  $y$  is state-space degenerate. If  $y$  is not error-space degenerate, then this condition is also necessary. In particular, if  $y$  is scalar,  $y$  is error-space degenerate if and only if  $\Phi(\infty) = 0$  and state-space degenerate if and only if  $\Phi^{-1}(0) = 0$ .*

**Proof.** (i) Since  $\Phi(z) = W_+(z)W_+(z^{-1})'$  and  $W_+(z) = W_-(z)Q_+(z)$ , we have  $W_-(z)Q_+(z) = \Phi(z)W_+^{-1}(z^{-1})'$ . Now,  $W_+$  has all its zeros outside the unit disc (Corollary 6.7.4), and hence  $W_+^{-1}(0)'$  is well-defined. Hence

$$D_-Q_+(\infty) = W_+^{-1}(0)'\Phi(\infty),$$



where  $W_-(\infty) = D_-$  is nonsingular (see page 227), and consequently, if there is a nonzero  $a \in \mathbb{R}^m$  such that  $\Phi(\infty)a = 0$ , then  $Q_+(\infty)$  is singular; i.e.,  $y$  is error-space degenerate (Corollary 9.3.6). The converse holds if  $W_+^{-1}(0)$  is nonsingular; that is, if  $W_+$  has no pole at  $z = 0$ . But this happens exactly when  $\ker A = 0$  or, equivalently,  $\ker U(\mathbf{X}) = 0$ ; i.e., when  $y$  is not state-space degenerate.

(ii) Since  $\Phi(z^{-1})' = \Phi(z) = \bar{W}_-(z)\bar{W}_-(z^{-1})'$ , we have

$$K_-(z) := \bar{W}_-^{-1}(z)W_-(z) = \bar{W}_-(z^{-1})'\Phi^{-1}(z^{-1})'W_-(z).$$

However,  $\bar{W}_-$  has no poles in the unit disc, and hence  $\bar{W}_-(0)$  is well-defined, so we can form

$$K_-(\infty)D_-^{-1} = \bar{W}_-(0)'\Phi^{-1}(0)',$$

which singular if there is a nonzero  $a \in \mathbb{R}^m$  such that  $a'\Phi^{-1}(0) = 0$ . However,  $K_-(\infty)$  is singular precisely when  $y$  is state-space degenerate. The converse would also hold if we could show that  $\bar{W}_-(0)$  is nonsingular. However, in view of (9.3.9),  $\bar{W}_-(0) = \bar{D}_+\bar{Q}_-(0)$ , where  $\bar{D}_+$  is nonsingular (see page 227), and hence the converse holds if  $y$  is not error-space degenerate (Corollary 9.3.6).

Finally, by Theorem 9.3.8, a scalar  $y$  cannot be state-space degenerate and error-space degenerate at the same time. Hence the last statement follows.  $\square$

### Some examples

To illustrate the results of this section, we provide two examples taken from [73].

**Example 9.3.11.** Consider a scalar process  $y$  with the rational spectral density

$$\Phi(z) = \frac{(z - \frac{2}{3})(z - \frac{1}{4})(z^{-1} - \frac{2}{3})(z^{-1} - \frac{1}{4})}{(z - \frac{1}{2})^2(z - \frac{1}{3})^2(z^{-1} - \frac{1}{2})^2(z^{-1} - \frac{1}{3})^2}.$$

It is not hard to see that the predictor space  $\mathbf{X}_-$  corresponds to the pair of spectral factors

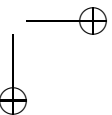
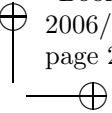
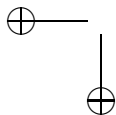
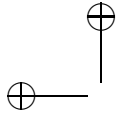
$$W_-(z) = \frac{z^2(z - \frac{2}{3})(z - \frac{1}{4})}{(z - \frac{1}{2})^2(z - \frac{1}{3})^2}, \quad \bar{W}_-(z) = \frac{z^2(z - \frac{2}{3})(z - \frac{1}{4})}{(1 - \frac{1}{2}z)^2(1 - \frac{1}{3}z)^2}.$$

In fact,  $W_-$  has all its poles and zeros in the open unit disc, and  $D_- = W(\infty) \neq 0$ . Moreover,

$$K(z) := \bar{W}_-(z)^{-1}W_-(z) = \frac{(1 - \frac{1}{2}z)^2(1 - \frac{1}{3}z)^2}{(z - \frac{1}{2})^2(z - \frac{1}{3})^2},$$

is inner and coprime with both  $Q_- = I$  and

$$\bar{Q}_-^*(z) = \frac{(1 - \frac{2}{3}z)(1 - \frac{1}{4}z)}{z^2(z - \frac{2}{3})(z - \frac{1}{4})},$$



ensuring minimality (Theorem 9.2.2). Since  $y$  is scalar, all internal minimal Markovian representations have the same structural function  $K$  (Corollary 9.2.13). Similarly, the backward predictor space  $\mathbf{X}_+$  corresponds to the spectral factors

$$W_+(z) = \frac{(1 - \frac{2}{3}z)(1 - \frac{1}{4}z)}{(z - \frac{1}{2})^2(z - \frac{1}{3})^2}, \quad \bar{W}_+(z) = \frac{(1 - \frac{2}{3}z)(1 - \frac{1}{4}z)}{(1 - \frac{1}{2}z)^2(1 - \frac{1}{3}z)^2},$$

because  $\bar{W}_+$  has all its poles and zeros in the complement of the closed unit disc and  $\bar{D}_+ = \bar{W}_+(0) \neq 0$ , and  $\bar{W}_+^{-1}W_+ = K$ .

Next, let us determine all internal minimal Markovian splitting subspaces. To this end, set  $\psi(z) := (z - \frac{1}{2})^2(z - \frac{1}{3})^2$  in terms of which  $K = \bar{\psi}/\psi$ , and note that, by Corollary 9.1.6,

$$H(K) = \left\{ z \frac{\rho}{\psi} \mid \deg \rho < 4 \right\}.$$

Therefore, by Theorem 9.1.5, if  $\mathbf{X}$  is an internal minimal Markovian splitting subspace with corresponding analytic spectral factor  $W = \pi/\psi$ , then

$$\mathbf{X} = \int_{-\pi}^{\pi} z^{-1} H(K) W^{-1} d\hat{y} = \int_{-\pi}^{\pi} \left\{ \frac{\rho}{\pi} \mid \deg \rho < 4 \right\} d\hat{y};$$

i.e.,  $\mathbf{X}$  is uniquely determined by the numerator polynomial  $\pi$  and the  $\deg \psi$ . In particular,  $\pi_-(z) = z^2(z - \frac{2}{3})(z - \frac{1}{4})$ , and consequently, by partial fraction expansion,  $\mathbf{X}$  is spanned by  $\int z^{-1} d\hat{y}$ ,  $\int z^{-2} d\hat{y}$ ,  $\int (z - \frac{2}{3})^{-1} d\hat{y}$ , and  $\int (z - \frac{1}{4})^{-1} d\hat{y}$ , and hence

$$\mathbf{X}_- = \text{span} \{y(-1), y(-2), x_1^-, x_2^-\},$$

where  $x_1^- := \sum_{k=-\infty}^{-1} (2/3)^{-k-1} y(k)$  and  $x_2^- := \sum_{k=-\infty}^{-1} (1/4)^{-k-1} y(k)$ . Likewise,

$$\mathbf{X}_+ = \text{span} \{y(0), y(1), x_1^+, x_2^+\},$$

where  $x_1^+ := \sum_{k=0}^{\infty} (2/3)^k y(k)$  and  $x_2^+ := \sum_{k=0}^{\infty} (1/4)^k y(k)$ . Therefore the frame space, the closed linear hull of all minimal Markovian splitting subspaces, is the eight-dimensional space

$$\mathbf{H}^{\square} = \text{span} \{y(-1), y(-2), y(0), y(1), x_1^-, x_2^-, x_1^+, x_2^+\},$$

and each  $\pi$  of degree at most four satisfying

$$\pi(z)\pi(z^{-1}) = \pi_-(z)\pi_-(z^{-1}), \tag{9.3.17}$$

corresponds a minimal Markovian splitting subspace with spectral factors  $(W, \bar{W}) =$

$(\pi/\psi, \pi/\bar{\psi})$ , and we list them together with the corresponding  $\pi$ .

$\mathbf{X}_- = \text{span} \{y(-1), y(-2), x_1^-, x_2^-\},$	$\pi_-(z) = z^2(z - \frac{2}{3})(z - \frac{1}{4})$
$\mathbf{X}_2 = \text{span} \{y(0), y(-1), x_1^-, x_2^-\},$	$\pi_2(z) = z(z - \frac{2}{3})(z - \frac{1}{4})$
$\mathbf{X}_3 = \text{span} \{y(1), y(0), x_1^-, x_2^-\},$	$\pi_3(z) = (z - \frac{2}{3})(z - \frac{1}{4})$
$\mathbf{X}_4 = \text{span} \{y(-1), y(-2), x_1^-, x_2^+\},$	$\pi_4(z) = z^2(z - \frac{2}{3})(1 - \frac{1}{4}z)$
$\mathbf{X}_5 = \text{span} \{y(0), y(-1), x_1^-, x_2^+\},$	$\pi_5(z) = z(z - \frac{2}{3})(1 - \frac{1}{4}z)$
$\mathbf{X}_6 = \text{span} \{y(1), y(0), x_1^-, x_2^+\},$	$\pi_6(z) = (z - \frac{2}{3})(1 - \frac{1}{4}z)$
$\mathbf{X}_7 = \text{span} \{y(-1), y(-2), x_1^+, x_2^-\},$	$\pi_7(z) = z^2(1 - \frac{2}{3}z)(z - \frac{1}{4})$
$\mathbf{X}_8 = \text{span} \{y(0), y(-1), x_1^+, x_2^-\},$	$\pi_8(z) = z(1 - \frac{2}{3}z)(z - \frac{1}{4})$
$\mathbf{X}_9 = \text{span} \{y(1), y(0), x_1^+, x_2^-\},$	$\pi_9(z) = (1 - \frac{2}{3}z)(z - \frac{1}{4})$
$\mathbf{X}_7 = \text{span} \{y(-1), y(-2), x_1^+, x_2^-\},$	$\pi_7(z) = z^2(1 - \frac{2}{3}z)(z - \frac{1}{4})$
$\mathbf{X}_8 = \text{span} \{y(0), y(-1), x_1^+, x_2^-\},$	$\pi_8(z) = z(1 - \frac{2}{3}z)(z - \frac{1}{4})$
$\mathbf{X}_9 = \text{span} \{y(1), y(0), x_1^+, x_2^-\},$	$\pi_9(z) = (1 - \frac{2}{3}z)(z - \frac{1}{4})$
$\mathbf{X}_{10} = \text{span} \{y(-1), y(-2), x_1^+, x_2^+\},$	$\pi_{10}(z) = z^2(1 - \frac{2}{3}z)(1 - \frac{1}{4}z)$
$\mathbf{X}_{11} = \text{span} \{y(0), y(-1), x_1^+, x_2^+\},$	$\pi_{11}(z) = z(1 - \frac{2}{3}z)(1 - \frac{1}{4}z)$
$\mathbf{X}_+ = \text{span} \{y(1), y(0), x_1^+, x_2^+\},$	$\pi_+(z) = (1 - \frac{2}{3}z)(1 - \frac{1}{4}z)$

We note that eight of these minimal Markovian splitting subspaces contain  $y(0)$ , indicating degeneracy. In fact, since  $\bar{Q}_-(\infty) = 0$ ,  $y$  is error-space degenerate (Corollary 9.3.6). This also follows from Theorem 9.3.10 by observing that  $\Phi(\infty) = 0$ . On the other hand, as expected (Theorem 9.3.8),  $y$  is not state-space degenerate, since  $K(\infty) = 1/36 \neq 0$ . We return to this example in Section 9.4.

**Example 9.3.12.** Let us return to Example 9.2.16, which demonstrates that a process  $y$  can be both state-space degenerate and error-space degenerate. In fact, since  $k_2(\infty) = 0$ ,  $K(\infty)$  is singular for all minimal Markovian splitting subspaces, and hence  $y$  is state-space degenerate. Moreover,

$$Q_+(\infty) = \begin{bmatrix} 1/6 & 0 \\ 0 & 0 \end{bmatrix},$$

so  $y$  is also error-space degenerate (Corollary 9.3.6).

### 9.4 Models without observation noise

Let us now return to the framework of Section 8.9 with  $t = 0$  contained in both the past and the future. More precisely, consider proper Markovian splitting subspaces  $\mathbf{X} \sim (\mathbf{S}, \bar{\mathbf{S}})$  with respect to

$$\begin{cases} \mathbf{H}^- = \overline{\text{span}}\{a'y(t) \mid t \leq 0; a \in \mathbb{R}^m\} \\ \mathbf{H}^+ = \overline{\text{span}}\{a'y(t) \mid t \geq 0; a \in \mathbb{R}^m\} \end{cases} \quad (9.4.1)$$

with generating processes  $(w, \bar{w})$  such that

$$\begin{cases} \mathbf{S} = \mathbf{H}^-(w) := \overline{\text{span}}\{a'w(t) \mid t \leq 0; a \in \mathbb{R}^p\} \\ \bar{\mathbf{S}} = \mathbf{H}^+(\bar{w}) := \overline{\text{span}}\{a'\bar{w}(t) \mid t \geq 0; a \in \mathbb{R}^p\} \end{cases} . \quad (9.4.2)$$

In this setting,

$$\mathbf{X} = \mathbf{S} \ominus \bar{\mathbf{S}}^\perp = \mathbf{H}^-(w) \ominus [U^* \mathbf{H}^-(\bar{w})],$$

that is, applying the isomorphisms (9.1.6),

$$\mathcal{J}_{\hat{w}}^{-1} \mathbf{X} = H_p^2 \ominus [\mathcal{J}_{\hat{w}}^{-1} U^* \mathcal{J}_{\bar{\hat{w}}} H_p^2] = H_p^2 \ominus [z^{-1} \mathcal{J}_{\hat{w}}^{-1} \mathcal{J}_{\bar{\hat{w}}} H_p^2].$$

In the same way as in Section 9.1, we can show that there is a  $p \times p$  inner function  $K$  such that

$$z^{-1} \mathcal{J}_{\hat{w}}^{-1} \mathcal{J}_{\bar{\hat{w}}} = M_K, \quad (9.4.3)$$

and hence

$$\mathcal{J}_{\bar{\hat{w}}}^{-1} \mathbf{X} = H(K) := H_p^2 \ominus H_p^2 K. \quad (9.4.4)$$

Next, introduce the pair  $(W, \bar{W})$  of spectral factors corresponding to  $\mathbf{X}$  via (9.1.13). Then, for all  $a \in \mathbb{R}^m$ ,

$$a'W = \mathcal{J}_{\hat{w}}^{-1} a'y(0) = \mathcal{J}_{\hat{w}}^{-1} \mathcal{J}_{\bar{\hat{w}}} a'\bar{W} = z a'\bar{W} K,$$

and therefore

$$W = z \bar{W} K. \quad (9.4.5)$$

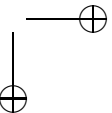
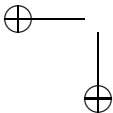
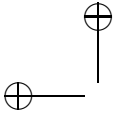
Together with the spectral inner factors defined by (9.1.17), the structural function  $K$  defines an inner triplet  $(K, Q, \bar{Q})$ , where only  $K$  has been altered. Replacing the phase function (9.2.1) by

$$\Theta := z^{-1} \bar{W}_+^{-1} W_-, \quad (9.4.6)$$

the factorization (9.2.2) remains valid. Accordingly, a straightforward modification of Theorems 9.1.5 and 9.2.2 then yields the following theorem.

**Theorem 9.4.1.** *Given the framework (9.4.1), let  $(\mathbb{H}, U, \mathbf{X})$  be a proper Markovian representation with inner triplet  $(K, Q, \bar{Q})$  and generating processes  $(w, \bar{w})$ . Then,*

$$\mathbf{X} = \int_{-\pi}^{\pi} H(K) d\hat{w} = \int_{-\pi}^{\pi} \bar{H}(K^*) d\hat{w}, \quad (9.4.7)$$





and  $\mathbf{X}$  is finite dimensional if and only if  $K$  is rational, in which case  $\dim \mathbf{X}$  equals the McMillan degree of  $K$ . Moreover,  $\mathbf{X}$  is constructible if and only if  $K$  and  $Q$  are right coprime and observable if and only if  $K^*$  and  $\bar{Q}$  are right coprime. Finally, the pair  $(W, \bar{W})$  of spectral factors corresponding to  $\mathbf{X}$  satisfies (9.4.5).

The other Hardy space results of this chapter remain valid for this setting with obvious modifications.

To illustrate the theory of modeling without observation noise, as developed in this section and in Section 8.9, we revisit some previous examples.

**Example 9.4.2.** Consider a process  $y$  with the spectral density analyzed in Examples 6.8.5 and 9.2.15. Clearly, the backward predictor space  $\mathbf{X}_+$  (in the traditional setting) has the standard forward realization

$$\begin{cases} x_+(t+1) = \frac{1}{2}x_+(t) + \frac{5}{2}w_+(t) \\ y(t) = x_+(t) + \frac{1}{2}w_+(t), \end{cases}$$

and  $Q_+(z) = (1 + \frac{1}{2}z)/(z - \frac{1}{2})$ . Since  $D = \frac{1}{2}$ , the forward realization is regular. Moreover, since  $Q_+(\infty) = \frac{1}{2} \neq 0$ , the standard backward realization is regular; i.e.,  $\bar{D} \neq 0$  (Proposition 9.3.4). Consequently, by Corollary 8.9.3,  $\mathbf{X} := \mathbf{X}_+ \oplus \mathbf{W}_+$  is a minimal Markovian splitting subspace in the framework of (9.4.1), and hence, setting  $x = (x_+, w_+)'$ ,

$$\begin{cases} x(t+1) = \begin{bmatrix} \frac{1}{2} & \frac{5}{2} \\ 0 & 0 \end{bmatrix} x(t) + \begin{bmatrix} 0 \\ 1 \end{bmatrix} w_+(t) \\ y(t) = \begin{bmatrix} 1 & \frac{1}{2} \end{bmatrix} x(t) \end{cases}$$

is a minimal realization in the framework of (9.4.1) of precisely the required form (8.9.2).

**Example 9.4.3.** The process  $y$  in Example 9.3.11 is error-space degenerate, since  $\bar{Q}_*(\infty) = 0$  (Corollary 9.3.6), and hence extensions along the lines of Example 9.4.2 will yield nonminimal representations. Using the same indexing of minimal splitting subspaces as in Example 9.3.11, we see that, for example,  $(W_-, \bar{W}_2)$  satisfies (9.4.5) with  $K = \bar{\psi}/\psi$ , and all pairs  $(W_k, \bar{W}_{k+1})$  produce the same  $K$ . Hence the family the minimal Markovian splitting subspaces in the framework (9.4.1) is a subset of the family of minimal Markovian splitting subspaces in the old framework. By Theorem 9.4.1, these splitting subspaces are given by

$$\mathbf{X} = \int_{-\pi}^{\pi} H(K)W^{-1}d\hat{y} = \int_{-\pi}^{\pi} \left\{ z \frac{\rho}{\pi} \mid \deg \rho < 4 \right\} d\hat{y},$$

where the polynomial  $\pi$  is a solution of (9.3.17) such that  $\deg z\pi \leq 4$ . Consequently,  $\mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_5, \mathbf{X}_6, \mathbf{X}_8, \mathbf{X}_9, \mathbf{X}_{11}$  and  $\mathbf{X}_+$  are minimal Markovian splitting subspaces in the framework (9.4.1), as also manifested by the fact that they are precisely the  $\mathbf{X}$  containing  $y(0)$ , and the frame space is the seven-dimensional space

$$\mathbf{H}^{\square} = \text{span} \{y(-1), y(0), y(1), x_1^-, x_2^-, x_1^+, x_2^+\}.$$

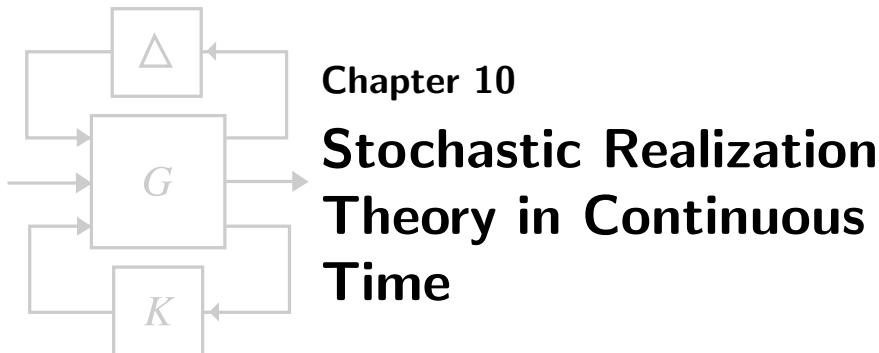
**Example 9.4.4.** In Example 9.2.16/Example 9.3.12 a routine calculation shows that  $K_-$  (and  $K_+$ ) have McMillan degree three, and therefore all minimal realization in the standard framework have dimension three. Extending these  $\mathbf{X}$  along the lines of Corollary 8.9.3 will produce five-dimensional Markovian splitting subspaces. However, due to error space degeneracy, none of these are minimal in the framework (9.4.1). In fact, it can be seen that the minimal ones are four-dimensional, which is consistent with the fact that  $\ker Q_+(\infty)$  is one-dimensional.

## 9.5 Bibliographical notes

Hardy-space representations of Markovian representations in the present form were introduced in [76], inspired by realization theory in Hardy space [34], and was developed further in [110, 79, 80, 81, 82, 32, 31, 30, 74, 112, 73, 85, 86].

In this chapter, Section 9.1 is based on [86] and follows closely the development there. A version of Theorem 9.1.5 appeared in [79]. Section 9.2 is based on [79, 80, 82, 85] and was generalized in [86] to the noninternal setting. Theorem 9.2.2 appeared in [80] in the internal setting, but is closely related to a similar result in [34]. Theorems 9.2.12 and 9.2.17 were presented in [82] for internal Markovian representations and generalized to noninternal representations in [86]. Proposition 9.2.18 is a generalization of [110, 79].

Sections 9.3 follows [73] closely and generalizes result therein to the noninternal setting. The examples are taken from [73]. A version of Theorem 9.3.10 appears in [72]; also see [100]. Finally, Section 9.4 is based on [74].



# Chapter 10

## Stochastic Realization Theory in Continuous Time

This chapter is devoted to continuous-time versions of the basic results in Chapters 6, 8 and 9. In this context, the linear stochastic model (6.0.1) corresponds to a system

$$\begin{cases} dx = Axdt + Bdw \\ dy = Cxdt + Ddw \end{cases}$$

of stochastic differential equations driven by the increments of a vector Wiener process  $w$ . The state process  $x$  will still be a stationary process, but the output process  $y$  has stationary increments. In the case when  $D = 0$ , we may instead consider a model

$$\begin{cases} dx = Axdt + Bdw \\ y = Cx \end{cases}$$

for which the output is a stationary process.

### 10.1 Continuous-time stochastic models

A basic object of our study are linear stochastic systems of the type

$$(\Sigma) \begin{cases} dx = Axdt + Bdw \\ dy = Cxdt + Ddw \end{cases} \tag{10.1.1}$$

defined for all  $t \in \mathbb{R}$ , where  $w$  is a  $p$ -dimensional vector Wiener process, and  $A, B, C, D$  are constant matrices with  $A$  being a *stability matrix*, which, in the present continuous-time setting, amounts to having all its eigenvalues in the open left half-plane. The system is in statistical steady state so that the  $n$ -dimensional *state* process  $x$  and the increments of the  $m$ -dimensional *output* process  $y$  are jointly stationary. We shall think of  $\Sigma$  as a representation of the (increments of the) process  $y$ ; such a representation will be called a (finite-dimensional) *stochastic realization* of  $dy$ . The number of state variables  $n$  will be called the *dimension* of  $\Sigma$ , denoted  $\dim \Sigma$ .

Systems of this type have been used in the engineering literature since the early 1960's as models for random signals. An alternative but, as we shall see below, not entirely equivalent way of representing the signal  $dy$  is obtained by eliminating the state  $x$  from (10.1.1). In this way we obtain a scheme which generates  $dy$  by passing white noise  $dw$  through a shaping filter with rational transfer function

$$W(s) = C(sI - A)^{-1}B + D \tag{10.1.2}$$

as explained in Section 3.6. This produces a stationary increment process  $dy$  with the spectral representation

$$y(t) - y(s) = \int_{-\infty}^{\infty} \frac{e^{i\omega t} - e^{i\omega s}}{i\omega} W(i\omega) d\hat{w} \tag{10.1.3}$$

and hence with the rational spectral density

$$W(s)W(-s)' = \Phi(s). \tag{10.1.4}$$

In other words,  $W$  is a *spectral factor* of  $\Phi$ , which, in view of the fact that  $A$  is a stability matrix, is *analytic*, i.e. has all its poles in the open left halfplane.

However, as in the discrete-time setting, the model  $\Sigma$  is more than just a representation of a stochastic process in terms of white noise. Much more important in applications is that the model (10.1.1) contains a state process  $x$  which serves as a dynamical memory for  $dy$ , which is described in terms of the splitting geometry of Chapter 7. However, let us first present some preliminary observations about stochastic models.

### Minimality and nonminimality of models

As usual, we shall say that  $\Sigma$  is *minimal* if  $dy$  has no other stochastic realization of smaller dimension. Occasionally, as for example in noncausal estimations, we shall also need to consider nonminimal  $\Sigma$ . Therefore, it is important to understand the relation between  $\deg W$ , the McMillan degree of  $W$ , and  $\dim \Sigma$ .

Before turning to this point, we need to recall a few well-known facts about the state process  $x$ . Since  $A$  is a stability matrix, we have

$$x(t) = \int_{-\infty}^t e^{A(t-\tau)} B dw(\tau), \tag{10.1.5}$$

from which it is seen that the state process is a stationary wide-sense Markov process with a constant covariance matrix

$$P := E\{x(t)x(t)'\} = \int_0^{\infty} e^{A\tau} B B' e^{A'\tau} d\tau, \tag{10.1.6}$$

which clearly satisfies the Lyapunov equation

$$AP + PA' + BB' = 0. \tag{10.1.7}$$

From (10.1.6) it is seen that  $P$  is the reachability Grammian for the pair  $(A, B)$ , and therefore the system  $\Sigma$  is reachable if and only if  $P$  is positive definite ( $P > 0$ ), i.e. if and only if  $\{x_1(0), x_2(0), \dots, x_n(0)\}$  is a basis in the space

$$\mathbf{X} = \text{span}\{x_1(0), x_2(0), \dots, x_n(0)\} \tag{10.1.8}$$

consisting of all linear combinations of the components of  $x(0)$ . As we shall see below,  $\mathbf{X}$  is a Markovian splitting subspace, and hence we can appeal to the geometric theory of Chapter 7.

However, as in the discrete-time setting,  $\mathbf{X}$  and  $\Sigma$  are not equivalent representations, as trivially there may be redundancy in  $\Sigma$  due to nonreachability which cannot be seen in  $\mathbf{X}$ . The following proposition makes this point more precise and gives a preview of some facts concerning  $\mathbf{X}$  and  $W$  to be studied in detail in Sections 10.3 and 10.4.

**Proposition 10.1.1.** *Let  $dy$  be a stationary-increment process with a rational spectral density  $\Phi$  having a finite-dimensional stochastic realization  $\Sigma$  of type (10.1.1) with spectral factor  $W$  given by (10.1.2), and let  $\mathbf{X}$  be the state space (10.1.8). Then*

$$\frac{1}{2} \text{deg } \Phi \leq \text{deg } W \leq \dim \mathbf{X} \leq \dim \Sigma \tag{10.1.9}$$

Moreover,  $\text{deg } W = \dim \mathbf{X}$  if and only if  $(C, A)$  is observable, and  $\dim \mathbf{X} = \dim \Sigma$  if and only if  $(A, B)$  is reachable.

The statements concerning the last of inequalities (10.1.9) follows immediately from the preceding discussion, while those concerning the second inequality are a consequence of Theorem 10.3.13 below. The first inequality in the chain follows from Proposition 10.4.2.

From Proposition 10.1.1 we may learn several things about stochastic realizations (10.1.1). First, for  $\Sigma$  to be minimal it is *not* sufficient that  $\Sigma$  is both observable and reachable. For this we must also have

$$\text{deg } W = \frac{1}{2} \text{deg } \Phi \tag{10.1.10}$$

A  $W$  satisfying this condition will be called a *minimal* spectral factor. Secondly, if  $dy$  is generated by a stochastic system (10.1.1) with  $A$  being a stability matrix, reachability plays no role in the geometric theory since the basic object of it is  $\mathbf{X}$  and not  $\Sigma$ .

On the other hand, if we allow  $A$  to have eigenvalues on the unit circle, the geometric concept of reachability introduced in Section 8.4 becomes important. However, in this chapter we shall not dwell on this, as the necessary modifications needed to accommodate a purely deterministic component in  $\mathbf{X}$  can easily be filled in by the reader by following the path in Section 8.4.

### The idea of state space and Markovian representations

There is a trivial equivalence relation between realizations of  $dy$  corresponding to a change of coordinates in the state space and constant orthogonal transformations of

the input Wiener process  $dw$ , which we would like to factor out before undertaking the study of the family of (minimal *and* nonminimal) stochastic realizations. The equivalence classes are defined by

$$(A, B, C, D, dw) \sim (T_1 A T_1^{-1}, T_1 B T_2^{-1}, C T_1^{-1}, D T_2^{-1}, T_2 dw) \quad (10.1.11)$$

where  $T_1$  is an  $n \times n$  nonsingular matrix and  $T_2$  is a  $p \times p$  orthogonal matrix. Clearly, the state space  $\mathbf{X}$ , defined by (10.1.8), is an invariant of this equivalence, and we shall look for conditions under which this invariant is complete in the sense that there is bijective correspondence between equivalence classes  $[\Sigma]$  and spaces  $\mathbf{X}$ . Since realizations  $\Sigma$  and  $\tilde{\Sigma}$  such that  $\begin{bmatrix} B \\ D \end{bmatrix} dw = \begin{bmatrix} \tilde{B} \\ \tilde{D} \end{bmatrix} d\tilde{w}$  give rise to the same  $\mathbf{X}$ , an obvious necessary condition is that

$$\text{rank} \begin{bmatrix} B \\ D \end{bmatrix} = p. \quad (10.1.12)$$

Moreover, as pointed out above, it is necessary to consider only models  $\Sigma$  for which

$$(A, B) \text{ reachable.} \quad (10.1.13)$$

We shall prove that under these two conditions the above one-one correspondence holds.

We proceed to characterize these  $\mathbf{X}$  spaces. Given a realization  $\Sigma$ , first denote by  $\mathbb{H}$  and  $\mathbf{H}$  the spaces of random variables

$$\mathbb{H} := \mathbf{H}(dw) \quad \mathbf{H} := \mathbf{H}(dy), \quad (10.1.14)$$

defined as in Section 2.7, and let  $\{U_t; t \in \mathbb{R}\}$  be the *shift* induced by  $dw$ , i.e. the strongly continuous group of unitary operators on  $\mathbb{H}$  such that

$$U_t[w(\tau) - w(\sigma)] = w(\tau + t) - w(\sigma + t). \quad (10.1.15)$$

Obviously  $\mathbf{X}$  and  $\mathbf{H}$  are subspaces of  $\mathbb{H}$ ,  $\mathbf{H}$  being doubly invariant for the shift so that  $U_t x(\tau) = x(\tau + t)$  and

$$U_t[y(\tau) - y(\sigma)] = y(\tau + t) - y(\sigma + t). \quad (10.1.16)$$

Next define

$$\mathbf{X}^- := \mathbf{H}^-(x), \quad \mathbf{X}^+ := \mathbf{H}^+(x), \quad (10.1.17a)$$

$$\mathbf{H}^- := \mathbf{H}^-(dy), \quad \mathbf{H}^+ := \mathbf{H}^+(dy). \quad (10.1.17b)$$

Now solving (10.1.1) we have

$$x(t) = e^{At} x(0) + \int_0^t e^{A(t-\tau)} B dw(\tau) \quad (10.1.18a)$$

$$y(t) - y(0) = \int_0^t C e^{A\tau} d\tau x(0) + \int_0^t \left[ \int_\tau^t C e^{A(\tau-\sigma)} B d\sigma + D \right] dw(\tau). \quad (10.1.18b)$$

Therefore, since  $\mathbf{H}^- \vee \mathbf{X}^- \subset \mathbf{H}^-(dw) \perp \mathbf{H}^+(dw)$ ,

$$\mathbb{E}^{\mathbf{H}^- \vee \mathbf{X}^-} \lambda = \mathbb{E}^{\mathbf{X}} \lambda \quad \text{for all } \lambda \in \mathbf{H}^+ \vee \mathbf{X}^+ \tag{10.1.19}$$

which, by Proposition 2.4.2, is the conditional orthogonality

$$\mathbf{H}^- \vee \mathbf{X}^- \perp \mathbf{H}^+ \vee \mathbf{X}^+ \mid \mathbf{X}. \tag{10.1.20}$$

Consequently, since

$$\mathbf{X}^- = \vee_{t \leq 0} U_t \mathbf{X} \quad \text{and} \quad \mathbf{X}^+ = \vee_{t \geq 0} U_t \mathbf{X}, \tag{10.1.21}$$

$\mathbf{X} \sim (\mathbf{S}, \bar{\mathbf{S}})$  is a Markovian splitting subspace with  $\mathbf{S} := \mathbf{H}^- \vee \mathbf{X}^-$  and  $\bar{\mathbf{S}} := \mathbf{H}^+ \vee \mathbf{X}^+$  (Section 7.4).

Moreover,

$$\mathbb{H} = \mathbf{H} \vee \overline{\text{span}}\{U_t \mathbf{X} \mid t \in \mathbb{Z}\}. \tag{10.1.22}$$

In fact, if this were not the case, there would be a nonzero  $a \in \mathbb{R}^p$  such that  $a'[w(\tau) - w(\sigma)] \perp \mathbf{H} \vee \overline{\text{span}}\{U^t \mathbf{X} \mid t \in \mathbb{Z}\}$  for some  $\tau, \sigma$  (which actually may be chosen arbitrarily, since the space is doubly invariant), and consequently, using the integral definition of (10.1.1), we have

$$\mathbb{E} \left\{ \begin{bmatrix} x(t) - x(0) \\ y(t) - y(0) \end{bmatrix} [w(\tau) - w(\sigma)]' \right\} a = \begin{bmatrix} B \\ D \end{bmatrix} a = 0,$$

which contradicts the assumption (10.1.12). Therefore, appropriately modifying the discrete-time definition of Chapter 8, as we shall do in the next section,  $(\mathbb{H}, \{U_t\}, \mathbf{X})$  is a Markovian representation of  $dy$ .

Conversely, as we show in this chapter, to each such Markovian representation  $(\mathbb{H}, \{U_t\}, \mathbf{X})$  of  $dy$ , there is an equivalence class  $[\Sigma]$  of realizations (10.1.1). More precisely, in particular we show that there is a one-one correspondence between equivalence classes  $[\Sigma]$  of stochastic realizations of  $dy$  satisfying conditions (10.1.12) and (10.1.13) and proper finite-dimensional Markovian representations  $(\mathbb{H}, \{U_t\}, \mathbf{X})$  of  $dy$  under which  $\mathbf{H}(dw) = \mathbb{H}$  and the state  $x(0) = \{x_1(0), x_2(0), \dots, x_n(0)\}$  of each  $\Sigma \in [\Sigma]$  is a basis in  $\mathbf{X}$ .

### Modeling stationary processes

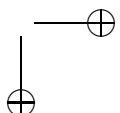
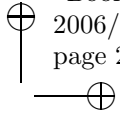
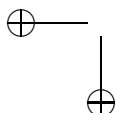
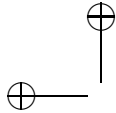
Before proceeding to the geometric state space construction, let us consider realizations

$$\begin{cases} dx = Axdt + Bdw \\ y = Cx \end{cases} \tag{10.1.23}$$

of a stationary process  $y$ . As before, the ambient space  $\mathbb{H}$ , defined as in (10.1.14), is endowed with the shifts (10.1.15), under which

$$U_t y_k(0) = y_k(t), \quad k = 1, 2, \dots, m, \tag{10.1.24a}$$

$$U_t x_k(0) = x_k(t), \quad k = 1, 2, \dots, n. \tag{10.1.24b}$$



Let  $\mathbf{H}^-$  and  $\mathbf{H}^+$  be the spaces

$$\mathbf{H}^- := \mathbf{H}^-(y), \quad \mathbf{H}^+ := \mathbf{H}^+(y), \quad (10.1.25)$$

and let  $\mathbf{H} := \mathbf{H}(y)$ , which as before is a doubly invariant subspace of  $\mathbb{H}$ . Solving (10.1.23) we have

$$x(t) = e^{At}x(0) + \int_0^t e^{A(t-\tau)}Bdw(\tau). \quad (10.1.26)$$

Therefore, since  $\mathbf{X}^- \subset \mathbf{H}^-(dw) \perp \mathbf{H}^+(dw)$ , with  $\mathbf{X}^-$  and  $\mathbf{X}^+$  given by (10.1.21),

$$E^{\mathbf{X}^-} \lambda = E^{\mathbf{X}} \lambda \quad \text{for all } \lambda \in \mathbf{X}^+, \quad (10.1.27)$$

which, by Proposition 2.4.2, is the Markovian property

$$\mathbf{X}^- \perp \mathbf{X}^+ \mid \mathbf{X}. \quad (10.1.28)$$

Moreover,

$$y(0) = \mathbf{X}, \quad (10.1.29)$$

and

$$\mathbb{H} = \overline{\text{span}}\{U_t \mathbf{X} \mid t \in \mathbb{Z}\}. \quad (10.1.30)$$

Consequently,  $\mathbf{X}$  is a Markovian splitting subspace (Section 7.4) with the ambient space  $\mathbb{H}$ .

## 10.2 Markovian representations

Let the past space  $\mathbf{H}^-$  and the future space  $\mathbf{H}^+$  be given by either (10.1.17b) or (10.1.25), depending on whether we want a representation of a stationary increments  $dy$  or of a stationary process  $y$ , and let

$$\mathbf{H}^- \vee \mathbf{H}^+ = \mathbf{H}.$$

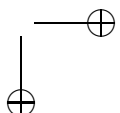
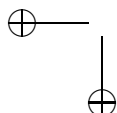
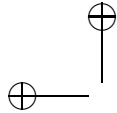
For the rest of this chapter we assume that the underlying process ( $y$  or  $dy$ ) is purely nondeterministic and reversible and hence purely nondeterministic also in the backward direction.

**Assumption 10.2.1.** *The remote past of  $\mathbf{H}^-$  and the remote future of  $\mathbf{H}^+$  are trivial; i.e.,  $\cap_{-\infty}^0 U^t \mathbf{H}^- = 0$  and  $\cap_0^\infty U^t \mathbf{H}^+ = 0$ .*

**Definition 10.2.2.** A *Markovian representation* of  $dy [y]$  is a triplet  $(\mathbb{H}, \{U_t\}, \mathbf{X})$  consisting of a Markovian splitting subspace  $\mathbf{X}$  in a Hilbert space  $\mathbb{H}$  of random variables with a strongly continuous group of unitary operators (shifts) on  $\mathbb{H}$  with the the properties:

- (i)  $\mathbf{H} \subset \mathbb{H}$  is a doubly invariant subspace, and the restricted shifts  $U_t|_{\mathbf{H}}$  are the natural shifts on  $\mathbf{H}$ ; i.e., satisfying (10.1.16) [(10.1.24a)], and

$$U_t \mathbf{H}^- \subset \mathbf{H}^- \text{ for } t \leq 0 \quad \text{and} \quad U_t \mathbf{H}^+ \subset \mathbf{H}^+ \text{ for } t \geq 0.$$





(ii)  $\mathbb{H}$  is the *ambient space* of  $\mathbf{X}$  in the sense that

$$\mathbb{H} = \mathbf{H} \vee \overline{\text{span}}\{U^t \mathbf{X} \mid t \in \mathbb{Z}\}$$

and has finite multiplicity under the shifts  $\{U_t\}$ .

A Markovian representation is said to be *internal* if  $\mathbb{H} = \mathbf{H}$  and *observable*, *constructible* or *minimal* if the splitting subspace  $\mathbf{X}$  is.

Then, with obvious change of notation, Theorem 8.1.1 holds verbatim.

**Theorem 10.2.3.** *Given a stationary-increment vector process  $dy$  (or a stationary vector process  $y$ ),  $(\mathbb{H}, \{U_t\}, \mathbf{X})$  is a Markovian representation of  $dy$  [ $y$ ] if and only if*

$$\mathbf{X} = \mathbf{S} \cap \bar{\mathbf{S}} \tag{10.2.1}$$

for some pair  $(\mathbf{S}, \bar{\mathbf{S}})$  of subspaces of  $\mathbb{H}$  such that

- (i)  $\mathbf{H}^- \subset \mathbf{S}$  and  $\mathbf{H}^+ \subset \bar{\mathbf{S}}$ ,
- (ii)  $U_{-t}\mathbf{S} \subset \mathbf{S}$  and  $U_t\bar{\mathbf{S}} \subset \bar{\mathbf{S}}$  for all  $t \geq 0$ , and
- (iii)  $\mathbb{H} = \bar{\mathbf{S}}^\perp \oplus (\mathbf{S} \cap \bar{\mathbf{S}}) \oplus \mathbf{S}^\perp$ .

Moreover, the correspondence  $\mathbf{X} \leftrightarrow (\mathbf{S}, \bar{\mathbf{S}})$  is one-one. In fact,

$$\mathbf{S} = \mathbf{H}^- \vee \mathbf{X}^- \quad \text{and} \quad \bar{\mathbf{S}} = \mathbf{H}^+ \vee \mathbf{X}^+. \tag{10.2.2}$$

Finally,  $(\mathbb{H}, U, \mathbf{X})$  is observable if and only if

$$\bar{\mathbf{S}} = \mathbf{H}^- \vee \mathbf{S}^\perp, \tag{10.2.3}$$

and constructible if and only if

$$\mathbf{S} = \mathbf{H}^+ \vee \bar{\mathbf{S}}^\perp, \tag{10.2.4}$$

and minimal if and only if both (8.1.4) and (8.1.5) hold.

Given a Markovian representation  $(\mathbb{H}, \{U_t\}, \mathbf{X})$  we introduce the restricted shift on the Markovian splitting subspace  $\mathbf{X} \sim (\mathbf{S}, \bar{\mathbf{S}})$ , namely

$$U_t(\mathbf{X}) = E^{\mathbf{X}} U_t|_{\mathbf{X}}. \tag{10.2.5}$$

Since  $\{U_t\}$  is a strongly continuous group of unitary operators, (10.2.5) defines a strongly continuous contraction semigroup  $\{U_t(\mathbf{X}); t \geq 0\}$  satisfying (7.5.3) (Theorem 7.5.1).

**Theorem 10.2.4.** *The semigroup  $U_t(\mathbf{X})$  tends strongly to zero as  $t \rightarrow \infty$  if and only if*

$$S_{-\infty} := \bigcap_{t \leq 0} U_t \mathbf{S} = 0, \tag{10.2.6}$$

and  $U_t(\mathbf{X})^*$  tend strongly to zero as  $t \rightarrow \infty$  if and only if

$$\bar{S}_\infty := \bigcap_{t \geq 0} U_t \bar{\mathbf{S}} = 0. \tag{10.2.7}$$

**Proof.** The proof that  $U_t(\mathbf{X})$  tends strongly to zero if  $S_{-\infty} = 0$  is a trivial modification of the corresponding part of the proof of Theorem 8.2.1. Conversely, suppose that  $U_t(\mathbf{X})$  tends strongly to zero. Then  $U_1(\mathbf{X})^k$  tends strongly to zero as  $k \rightarrow \infty$  through the natural numbers, and hence  $\bigcap_{k=-\infty}^0 U_k \mathbf{S} = 0$ . However,  $U_t \mathbf{S} \subset U_k \mathbf{S}$  for  $t > k$ , and hence  $S_{-\infty} = 0$ . This proves the first statement. The second statement follows by symmetry.  $\square$

**Definition 10.2.5.** The Markovian representation  $(\mathbb{H}, \{U_t\}, \mathbf{X})$  is normal if  $\mathbf{S}_{-\infty} = \bar{\mathbf{S}}_\infty$  and proper if  $\mathbf{S}_{-\infty} = \bar{\mathbf{S}}_\infty = 0$ .

The remaining results of Section 8.2 can now be carried over to the continuous-time setting with trivial modifications. In particular we have the following counterpart of Corollary 8.2.6.

**Proposition 10.2.6.** A finite-dimensional Markovian representation  $(\mathbb{H}, \{U_t\}W, \mathbf{X})$  is normal.

Applying this proposition to the Markovian representations constructed in Section 10.1, we see that they are proper. In fact,  $\mathbf{S} \subset \mathbf{H}^-(dw)$ , and hence  $\mathbf{S}_{-\infty} = 0$ . Then, by Proposition 10.2.6,  $\bar{S}_\infty = 0$ .

### State space construction

To construct a functional model of this geometry we apply the continuous-time Wold decomposition of Section 5.1. Given a proper Markovian representation  $(\mathbb{H}, \{U_t\}, \mathbf{X})$  of multiplicity  $p \geq m$  with  $\mathbf{X} \sim (\mathbf{S}, \bar{\mathbf{S}})$ , there is a pair  $(dw, d\bar{w})$  of  $p$ -dimensional Wiener processes such that  $\mathbf{H}(dw) = \mathbf{H}(d\bar{w}) = \mathbb{H}$  and

$$\mathbf{S} = \mathbf{H}^-(dw) \quad \text{and} \quad \bar{\mathbf{S}} = \mathbf{H}^+(d\bar{w}) \tag{10.2.8}$$

(Theorem 5.1.1). These processes are called the *generating processes* of the Markovian representation, and they are uniquely determined modulo multiplication by a constant  $p \times p$  orthogonal matrix.

In view of (10.2.8), every random variable  $\eta \in \mathbf{S}$  can be represented by a stochastic integral with respect to  $dw$ ,

$$\eta = \int_{-\infty}^{\infty} f(-t)dw(t) =: \mathcal{J}_w f, \tag{10.2.9}$$

of a function  $f \in L_p^2(\mathbb{R})$  that vanishes on the negative real line. In particular, this naturally leads to representations of  $dy$  (or  $y$ ) by means of causal input-output

map driven by the white noise processes  $dw$ . The most efficient way to study such representations is by spectral-domain techniques. In fact, as explained in Section ??,

$$\eta = \int_{-\infty}^{\infty} \hat{f}(i\omega) d\hat{w} =: \mathcal{J}_{\hat{w}} \hat{f}, \tag{10.2.10}$$

where  $\hat{f} \in H_p^2$ , the Hardy space of  $p$ -dimensional square integrable functions, analytic in the right half of the complex plane. Here  $\hat{f}$  is the Fourier transform of  $f$ , and  $d\hat{w}$  is a (complex) stationary increment process with the property (3.6.3). In the same way, every random variable  $\bar{\eta} \in \bar{\mathbf{S}}$  can be represented by a stochastic integral can be represented by a stochastic integral with respect to  $d\bar{w}$ ,

$$\bar{\eta} = \int_{-\infty}^{\infty} \bar{f}(-t) d\bar{w}(t) =: \mathcal{J}_{\bar{w}} \bar{f}, \tag{10.2.11}$$

of a function  $\bar{f} \in L_p^2(\mathbb{R})$  that vanishes on the positive real line, or, equivalently,

$$\bar{\eta} = \int_{-\infty}^{\infty} \hat{\bar{f}}(i\omega) d\hat{w} =: \mathcal{J}_{\hat{\bar{w}}} \hat{\bar{f}}, \tag{10.2.12}$$

where  $\hat{\bar{f}} \in \bar{H}_p^2$ , the Hardy space of  $p$ -dimensional square integrable functions, analytic in the left half of the complex plane. This defines two unitary maps,  $\mathcal{J}_{\hat{w}}$  and  $\mathcal{J}_{\hat{\bar{w}}}$ , from  $L_p^2(\mathbb{I})$  to  $\mathbb{H}$ , establishing unitary isomorphisms between  $\mathbf{S}$  and  $\bar{\mathbf{S}}$  and the Hardy spaces  $H_p^2$  and  $\bar{H}_p^2$  respectively; i.e.,

$$\mathcal{J}_{\hat{w}} H_p^2 = \mathbf{H}^-(dw) = \mathbf{S} \quad \text{and} \quad \mathcal{J}_{\hat{\bar{w}}} \bar{H}_p^2 = \mathbf{H}^+(d\bar{w}) = \bar{\mathbf{S}}. \tag{10.2.13}$$

(See Section 5.3.) Under each of these isomorphisms the shift  $U_t$  becomes multiplication by  $e^{i\omega t}$ ; i.e.,

$$U_t \mathcal{J}_{\hat{w}} = \mathcal{J}_{\hat{w}} M_{e^{i\omega t}}, \tag{10.2.14}$$

and the orthogonal decomposition

$$\mathbb{H} = \mathbf{H}^-(dw) \oplus \mathbf{H}^+(d\bar{w}) \tag{10.2.15}$$

becomes

$$L_p^2(\mathbb{I}) = H_p^2 \oplus \bar{H}_p^2.$$

In view of Assumption 10.2.1, there are also Wiener processes  $dw_-$  and  $d\bar{w}_+$  such that

$$\mathbf{H}^- = \mathbf{H}^-(dw_-) \quad \text{and} \quad \mathbf{H}^+ = \mathbf{H}^-(d\bar{w}_+) \tag{10.2.16}$$

(Theorem 5.1.1). This is the forward and backward *innovation process*, respectively, and they are unique modulo multiplication by a unitary transformation.

Recall that a function  $Q \in H^\infty$  is called *inner* if the multiplication operator  $M_Q$  is an isometry which sends analytic functions (i.e., functions in  $H_p^2$ ) to analytic functions. A function with the corresponding properties with respect to the conjugate Hardy space  $\bar{H}_p^2$  will be called *conjugate inner*. The phase function is not in general inner, as this happens if and only if  $\mathbf{H}^-$  and  $\mathbf{H}^+$  intersect perpendicularly.

**Lemma 10.2.7.** *Let  $(\mathbb{H}, \{U_t\}, \mathbf{X})$  be a proper Markovian representation with generating processes  $(dw, d\bar{w})$ . Then there is an inner  $p \times p$  matrix functions  $K$ , an inner  $m \times p$  matrix function  $Q$  and a conjugate inner  $m \times p$  matrix function  $\bar{Q}$  such that*

$$(\mathcal{J}_{\hat{w}})^{-1}\mathcal{J}_{\hat{w}^+} = M_K, \quad (\mathcal{J}_{\hat{w}})^{-1}\mathcal{J}_{\hat{w}^-} = M_Q, \quad (\mathcal{J}_{\hat{w}})^{-1}\mathcal{J}_{\hat{w}^+} = M_{\bar{Q}}, \quad (10.2.17)$$

where  $M_V$  denotes the multiplication operator  $M_V f = fV$ . Moreover,

$$(\mathcal{J}_{\hat{w}^-})^{-1}\mathcal{J}_{\hat{w}^+} = M_{\Theta}, \quad (10.2.18)$$

where  $\Theta$  is a unitary  $m \times m$  matrix function on  $\mathbb{I}$  that satisfies

$$\Theta = \bar{Q}KQ^*. \quad (10.2.19)$$

**Proof.** In view of (10.2.14),

$$(\mathcal{J}_{\hat{w}})^{-1}\mathcal{J}_{\hat{w}}M_{e^{i\omega t}} = (\mathcal{J}_{\hat{w}})^{-1}U_t\mathcal{J}_{\hat{w}} = M_{e^{i\omega t}}(\mathcal{J}_{\hat{w}})^{-1}\mathcal{J}_{\hat{w}}.$$

Moreover, from the geometry of  $\mathbf{X} \sim (\mathbf{S}, \bar{\mathbf{S}})$  we have

$$\mathbf{X} = \mathbf{S} \ominus \bar{\mathbf{S}} \quad (10.2.20)$$

(Theorem 10.2.3). This implies that  $\bar{\mathbf{S}}^\perp \subset \mathbf{S}$ , or, equivalently that  $\mathbf{H}^-(d\bar{w}) \subset \mathbf{H}^-(dw)$ ; i.e.,  $\mathcal{J}_{\hat{w}}H_p^2 \subset \mathcal{J}_{\hat{w}}H_m^2$ , which yields

$$(\mathcal{J}_{\hat{w}})^{-1}\mathcal{J}_{\hat{w}}H_p^2 \subset H_p^2.$$

Consequently,  $(\mathcal{J}_{\hat{w}})^{-1}\mathcal{J}_{\hat{w}}$  is a unitary map from  $H_p^2$  to  $H_p^2$  that commutes with the shift. Therefore, by a continuous-time version of Theorem 4.3.3 [34, p. 185], there is an inner  $p \times p$ -matrix function  $K$  such that  $(\mathcal{J}_{\hat{w}})^{-1}\mathcal{J}_{\hat{w}} = M_K$ .

In the same way, we see that  $(\mathcal{J}_{\hat{w}})^{-1}\mathcal{J}_{\hat{w}^-}$  commutes with the shift, and  $\mathbf{H}^- \subset \mathbf{S}$  yields  $\mathcal{J}_{\hat{w}^-}H_m^2 \subset \mathcal{J}_{\hat{w}}H_p^2$ ; i.e.,  $(\mathcal{J}_{\hat{w}})^{-1}\mathcal{J}_{\hat{w}^-}H_m^2 \subset H_p^2$ . Therefore,  $(\mathcal{J}_{\hat{w}})^{-1}\mathcal{J}_{\hat{w}^-} = M_Q$  for some  $Q \in H_{m \times p}^\infty$  [34, p. 185]. However,  $(\mathcal{J}_{\hat{w}})^{-1}\mathcal{J}_{\hat{w}^-}$  is an isometry, and hence  $Q$  is inner. The last of the equations (10.2.17) follows by symmetry.

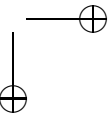
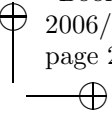
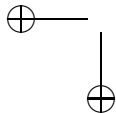
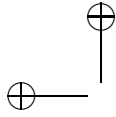
Finally, the last statement of the lemma follows from (10.2.17).  $\square$

The inner function  $K$  is called the *structural function* and  $(K, Q, \bar{Q})$  the *inner triplet* of the Markovian representation  $(\mathbb{H}, \{U_t\}, \mathbf{X})$ . The function  $\Theta$  is called the *phase function*. We return to their connection to spectral factors below.

This leads to a functional model for Markovian splitting subspaces of the type studied in [69], [117] and [34] that allows us to characterize various systems-theoretic properties of Markovian splitting subspaces (observability, constructibility and minimality) in function-theoretic terms.

**Theorem 10.2.8.** *Let  $(\mathbb{H}, \{U_t\}, \mathbf{X})$  be a proper Markovian representation with generating processes  $(dw, d\bar{w})$  and inner triplet  $(K, Q, \bar{Q})$ . Then*

$$\mathbf{X} = \int_{-\infty}^{\infty} H(K)d\hat{w} = \int_{-\infty}^{\infty} \bar{H}(K^*)d\hat{w}, \quad (10.2.21)$$



where  $H(K) := H_p^2 \ominus H_p^2 K$  and  $\bar{H}(K^*) := \bar{H}_p^2 \ominus \bar{H}_p^2 K^*$ . Moreover,  $\mathbf{X}$  is constructible if and only if  $K$  and  $Q$  are right coprime, i.e. they have no nontrivial common right inner factor, and  $\mathbf{X}$  is observable if and only if  $K^*$  and  $\bar{Q}$  are right coprime, i.e. they have no nontrivial common right conjugate inner factor.

**Proof.** From (10.2.20) and Lemma 10.2.7 we have

$$\mathcal{J}_{\hat{w}}^{-1} \mathbf{X} = H_p^2 \ominus (\mathcal{J}_{\hat{w}}^{-1} \mathcal{J}_{\hat{w}} H_p^2) = H_p^2 \ominus H_p^2 K,$$

which yields the first of equations (10.2.21). The second follows in the same way from  $\mathbf{X} = \mathbf{S} \ominus \mathbf{S}^\perp$ .

Likewise, the constructibility condition (10.2.4) yields

$$H_p^2 = \mathcal{J}_{\hat{w}}^{-1} \mathbf{S} = (\mathcal{J}_{\hat{w}}^{-1} \mathcal{J}_{\hat{w}_-} H_m^2) \vee (\mathcal{J}_{\hat{w}}^{-1} \mathcal{J}_{\hat{w}} H_p^2);$$

that is, by Lemma 10.2.7,

$$H_p^2 = (H_m^2 Q) \vee (H_p^2 K). \tag{10.2.22}$$

For (10.2.22) to hold,  $Q$  and  $K$  must clearly be right coprime. Conversely, suppose that  $Q$  and  $K$  are right coprime, and consider the right member of (10.2.22). Clearly it is a full-range invariant subspace of  $H_p^2$ , because  $H_p^2 K$  is, and therefore, by the half-plane version of Beurling-Lax Theorem 4.6.4 [?], it has the form  $H_p^2 J$  where  $J$  is inner. But then  $J$  must be a common right inner factor of  $Q$  and  $K$ , and hence  $J = I$ , concluding the proof of the constructibility criterion. The proof of the observability part is by symmetry.  $\square$

In Section 10.3 we show that  $\mathbf{X}$  is finite dimensional if and only if the structural function  $K$  is rational, in which case  $\dim \mathbf{X}$  equals the McMillan degree of  $K$  (Theorem 10.3.8).

In view of (10.2.19), the inner triplets  $(K, I, \bar{Q}_-)$  and  $(K_+, Q_+, I)$  of the predictor space  $\mathbf{X}_-$  and the backward predictor space  $\mathbf{X}_+$ , respectively, can be determined from the phase function  $\Theta$  via the coprime factorizations

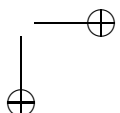
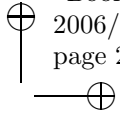
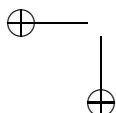
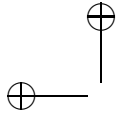
$$\Theta = \bar{Q}_- K_- = K_+ Q_+^*. \tag{10.2.23}$$

To be able to determine the other minimal Markovian splitting subspaces from  $\Theta$ , we need to assume that  $dy(y)$  is *strictly noncyclic*; i.e., the kernels of the operators

$$\mathcal{H} := E^{\mathbf{H}^+} |_{\mathbf{H}^-} \quad \text{and} \quad \mathcal{H}^* := E^{\mathbf{H}^-} |_{\mathbf{H}^+}$$

have full range; cf. Definition 9.2.4. This is equivalent to the frame space  $\mathbf{H}^\square$  being proper; cf. Proposition 9.2.5. Modifying the proof of Lemma 9.2.6 to the continuous-time setting, this condition can be expressed in terms of the phase function  $\Theta$ . More precisely, provided  $y$  is purely nondeterministic and reversible,  $dy(y)$  is strictly noncyclic if and only if the Hankel operator

$$H_\Theta := P^{H_p^2} M_\Theta |_{\bar{H}_p^2}$$



has full range kernel; i.e.,  $\Theta$  is a strictly noncyclic function [34, p. 253]. A processes  $dy(y)$  with a rational spectral density are strictly noncyclic, since the range of  $\mathcal{H}$  and  $\mathcal{H}^*$ ,  $\mathbf{X}_-$  and  $\mathbf{X}_+$ , respectively, are finite-dimensional, and thus proper. However, a scalar process with spectral density  $\Phi(i\omega) = (1 + \omega^2)^{-3/2}$  is not strictly noncyclic, since  $\mathbf{X}_- = \mathbf{H}^-$  and  $\mathbf{X}_+ = \mathbf{H}^+$  [27, p. 99].

Strict noncyclicity is a geometric condition on  $dy(y)$  determined by the splitting geometry, and therefore all of the results of Section 9.2 apply also in the present continuous-time setting with no or obvious modifications. In particular, Theorem 9.2.10 and 9.2.17 have the following counterpart.

**Theorem 10.2.9.** *Let  $dy(y)$  be purely nondeterministic. Then  $dy(y)$  is strictly noncyclic if and only if all minimal Markovian representations are proper. In this case, the following conditions are equivalent.*

- (i)  $\mathbf{X}$  is minimal
- (ii)  $\mathbf{X}$  is observable and  $\mathbf{X} \perp \mathbf{N}^+$
- (iii)  $\mathbf{X}$  is constructible and  $\mathbf{X} \perp \mathbf{N}^-$

Therefore, from Proposition 9.2.18, which is completely geometric and hence remains unaltered, and Theorem 10.2.8, we see that a proper Markovian representation  $(\mathbb{H}, \{U_t\}, \mathbf{X})$  with inner triplet  $(K, Q, \bar{Q})$  is minimal if and only if one of the following conditions hold.

- (ii)'  $K^*$  and  $\bar{Q}$  are right coprime, and  $Q^*Q_+$  is analytic;
- (iii)'  $K$  and  $Q$  are right coprime, and  $\bar{Q}^*Q_-$  is coanalytic.

This leads to a procedure for determining the inner triplets  $(K, Q, \bar{Q})$  of all minimal Markovian representation: Select  $Q$  so that  $Q^*Q_+$  is analytic. Then form  $T := \Theta Q$ , and determine  $\bar{Q}$  and  $K$  as the coprime factors in  $\bar{Q}K = T$ .

Theorem 9.2.12 and Corollary 9.2.13 can also be carried over verbatim.

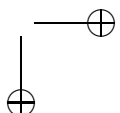
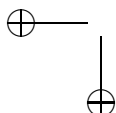
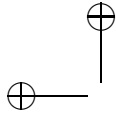
**Theorem 10.2.10.** *Suppose that  $dy(y)$  is strictly noncyclic. Then all internal minimal Markovian splitting subspaces have the same invariant factors; let us denote them*

$$k_1, k_2, k_3, \dots, k_m. \tag{10.2.24}$$

Moreover, a Markovian splitting subspace of multiplicity  $p$  is minimal if and only if  $m$  invariant factors are given by (10.2.24) and the remaining  $p - m$  are identically one. If  $m = 1$ , all internal minimal Markovian splitting subspaces have the same structural function.

**Example 10.2.11.** Let  $y$  be a two-dimensional process with the rational spectral density

$$\Phi(s) := \frac{1}{(s^2 - 1)(s^2 - 4)} \begin{bmatrix} 17 - 2s^2 & -(s + 1)(s - 2) \\ -(s - 1)(s + 2) & 4 - s^2 \end{bmatrix}.$$



Then, using the factorization procedure above, it can be seen that the structural function of  $\mathbf{X}_-$  is

$$K_-(s) = \frac{s-1}{(s+1)(s+2)} \begin{bmatrix} s-1.2 & 1.6 \\ 1.6 & s+1.2 \end{bmatrix}$$

and that the one of  $\mathbf{X}_+$  is

$$K_+(s) = \frac{s-1}{(s+1)(s+2)} \begin{bmatrix} s-70/37 & 24/37 \\ 24/37 & s+70/37 \end{bmatrix}.$$

These functions look quite different, but they have the same invariant factors, namely

$$k_1(s) = \frac{s-1}{s+1} \quad \text{and} \quad k_2(s) = \frac{(s-1)(s-2)}{(s+1)(s+2)},$$

and are therefore quasi-equivalent.

### Spectral factors and the structural function

First consider an  $m$ -dimensional, mean-square continuous, purely nondeterministic, stationary process  $y$  with a full rank spectral density  $\Phi$ . Recall from Section 3.3 that  $y$  has a spectral representation

$$y(t) = \int_{-\infty}^{\infty} e^{i\omega t} d\hat{y}, \tag{10.2.25}$$

(Theorem 3.3.2) where

$$E\{d\hat{y}d\hat{y}^*\} = \frac{1}{2\pi} \Phi(i\omega) d\omega. \tag{10.2.26}$$

Moreover, by Corollary 5.3.6,  $\Phi$  admits a factorization

$$W(s)W(-s)' = \Phi(s), \tag{10.2.27}$$

where  $W$  is  $m \times p$  with  $p \geq m$ . A spectral factor  $W$  is *analytic* if its rows belong to  $H_p^2$  and *coanalytic* if they belong to  $\bar{H}_p^2$ . A spectral  $m \times m$  factor  $W$  is *outer* if  $H_m^2 W$  is dense in  $H_m^2$  and *conjugate outer* if  $\bar{H}_m^2 W$  is dense in  $\bar{H}_m^2$ ; see Section 5.3.

The outer and conjugate outer spectral factors are unique modulo an orthogonal transformation from the right.

**Proposition 10.2.12.** *Let  $y$  be given by (10.2.25). If  $dw$  is a Wiener process such that  $\mathbf{H}^- \subset \mathbf{H}^-(dw)$ , there is a unique analytic spectral factor  $W$  such that*

$$d\hat{y} = Wd\hat{w}, \tag{10.2.28}$$

*and  $\mathbf{H}^- = \mathbf{H}^-(dw)$  if and only if  $W$  is outer. Likewise, if  $d\bar{w}$  is a Wiener process such that  $\mathbf{H}^+ \subset \mathbf{H}^+(d\bar{w})$ , there is a unique coanalytic spectral factor  $\bar{W}$  such that*

$$d\hat{y} = \bar{W}d\hat{w}, \tag{10.2.29}$$

and  $\mathbf{H}^+ = \mathbf{H}^+(d\bar{w})$  if and only if  $\bar{W}$  is conjugate outer.

**Proof.** Let  $W$  be the matrix function with  $m$  rows given by  $J_{\hat{w}}y_k(0)$ ,  $k = 1, 2, \dots, m$ . Then

$$y(t) = \int e^{i\omega t} W d\hat{w},$$

which compared with (10.2.25) yields  $d\hat{y} = W d\hat{w}$ . Now, if  $\mathbf{H}^- \subset \mathbf{H}^-(dw)$  holds, then

$$a'W = J_{\hat{w}}^{-1}y(0) \subset J_{\hat{w}}^{-1}\mathbf{H}^-(dw) = H_p^2$$

for all  $a \in \mathbb{R}^m$ , and hence  $W$  is analytic. Next, since

$$\mathbf{H}^- = \overline{\text{span}}\{a'y(t) \mid t \leq 0, a \in \mathbb{R}^m\} = \int \overline{\text{span}}\{e^{i\omega t} a'W \mid t \leq 0, a \in \mathbb{R}^m\} d\hat{w},$$

$\mathbf{H}^- = \mathbf{H}^-(d\hat{w})$  if and only

$$\overline{\text{span}}\{e^{i\omega t} a'W \mid t \leq 0, a \in \mathbb{R}^m\} = H_m^2;$$

i.e., if and only if  $W$  is outer. A symmetric argument yields the second statement.  $\square$

Next, consider an  $m$ -dimensional, mean-square continuous purely nondeterministic, stationary-increment process  $dy$  with a full rank (incremental) spectral density  $\Phi$ . Then, by Theorem 3.6.1,

$$y(t) - y(s) = \int_{-\infty}^{\infty} \frac{e^{i\omega t} - e^{i\omega s}}{i\omega} d\hat{y}, \quad t, s \in \mathbb{R}, \quad (10.2.30)$$

where the stochastic measure again satisfies (10.2.26) and the spectral density  $\Phi$  admits a spectral factorization (10.2.27) (Theorem 5.3.5). However, the spectral factors  $W$  may not be square-integrable. In fact, in Section 5.3 we introduced the modified Hardy spaces  $\mathcal{W}_p^2$  and  $\bar{\mathcal{W}}_p^2$  consisting of the  $p$ -dimensional row vector functions  $g$  and  $\bar{g}$  respectively such that  $\bar{\chi}_h g \in H_p^2$  and  $\chi_h \bar{g} \in \bar{H}_p^2$ , where

$$\chi_h(i\omega) = \frac{e^{i\omega h} - 1}{i\omega} \quad (10.2.31)$$

and  $\bar{\chi}_h(i\omega) = \chi_h(-i\omega)$ . Note that, for  $h > 0$ ,  $\bar{\chi}_h \in H^\infty$  and  $\chi_h \in \bar{H}^\infty$ . For reasons explained in Section 5.3, a spectral factor  $W$  with rows in  $\mathcal{W}_p^2$  will be called *analytic* and a spectral factor  $\bar{W}$  with rows in  $\bar{\mathcal{W}}_p^2$  *coanalytic*.

**Proposition 10.2.13.** *Let  $dy$  be given by (10.2.30). If  $dw$  is a Wiener process such that  $\mathbf{H}^- \subset \mathbf{H}^-(dw)$ , there is a unique analytic spectral factor  $W$  such that*

$$d\hat{y} = W d\hat{w}, \quad (10.2.32)$$

and  $\mathbf{H}^- = \mathbf{H}^-(d\hat{w})$  if and only if  $\bar{\chi}_h W$  is outer for  $h > 0$ . Likewise, if  $d\bar{w}$  is a Wiener process such that  $\mathbf{H}^+ \subset \mathbf{H}^+(d\bar{w})$ , there is a unique coanalytic spectral factor  $\bar{W}$  such that

$$d\hat{y} = \bar{W} d\hat{w}, \quad (10.2.33)$$



and  $\mathbf{H}^+ = \mathbf{H}^+(d\bar{w})$  if and only if  $\chi_h \bar{W}$  is conjugate outer.

**Proof.** For a fixed  $h > 0$ , let  $W$  be the  $m \times p$  matrix-valued functions with rows  $\bar{\chi}_h^{-1} \mathcal{J}_{\hat{w}}^{-1}[y_k(0) - y_k(-h)]$ ,  $k = 0, 1, \dots, m$ . Then

$$y(-h) - y(0) = \int_{-\infty}^{\infty} \frac{e^{-i\omega h} - 1}{i\omega} W(i\omega) d\hat{w},$$

which compared with (10.2.30) yields  $d\hat{y} = W d\hat{w}$ . In fact, the spectral measure  $d\hat{y}$  is uniquely determined by  $dy$  (Theorem 3.6.1). Clearly,  $W$  is a spectral factor that does not depend on the choice of  $h$ . If  $\mathbf{H}^- \subset \mathbf{H}^-(dw)$ ,

$$\bar{\chi}_h a' W = \mathcal{J}_{\hat{w}}^{-1} a' [y(0) - y(-h)] \subset \mathcal{J}_{\hat{w}}^{-1} \mathbf{H}^-(dw) = H_p^2$$

for all  $a \in \mathbb{R}^m$ , and hence  $W$  is an analytic spectral factor by the definition above. Moreover, since

$$\begin{aligned} \mathbf{H}^- &= \overline{\text{span}}\{a'[y(t) - y(t-h)] \mid t \leq 0, a \in \mathbb{R}^m\} \\ &= \int \overline{\text{span}}\{e^{i\omega t} \bar{\chi}_h(i\omega) a' W(i\omega) \mid t \leq 0, a \in \mathbb{R}^m\} d\hat{w}, \end{aligned}$$

we have  $\mathbf{H}^- = \mathbf{H}^-(dw)$  if and only if  $\bar{\chi}_h(i\omega) a' W \in H_m^2$ , as claimed.

Defining  $\bar{W}$  to be the  $m \times p$  matrix-valued functions with rows  $\chi_h^{-1} \mathcal{J}_{\hat{w}}^{-1}[y_k(h) - y_k(0)]$ ,  $k = 0, 1, \dots, m$ , a symmetric argument yields the second statement.  $\square$

Consequently, although the spectral factor may belong to different spaces, the spectral representations (10.2.26) and (10.2.27) are formally the same in the stationary and the stationary-increment cases, and therefore we shall formulate the spectral results in these quantities. To unify notations, we shall say that a spectral factor  $W_-$  is *minimum-phase* if it is either outer (when  $y$  is given) or  $\bar{\chi}_h W_-$  is outer (when  $dy$  is given). Likewise  $\bar{W}_+$  is *conjugate minimum-phase* if it is either conjugate outer or  $\chi_h \bar{W}_+$  is conjugate outer.

**Corollary 10.2.14.** *Given either (10.2.25) or (10.2.30), let  $dw_-$  and  $d\bar{w}_+$  be the forward and backward innovation, respectively, uniquely defined (modulo an orthogonal transformation) by (10.2.16). Then there is a unique minimum-phase spectral factor  $W_-$  and a unique conjugate minimum-phase spectral factor  $\bar{W}_+$  such that*

$$W_- d\hat{w}_- = d\hat{y} = \bar{W}_+ d\hat{w}_+. \tag{10.2.34}$$

Moreover, the phase function (10.2.18) can be written

$$\Theta = \bar{W}_+^{-1} W_-. \tag{10.2.35}$$

We have thus established that the geometry of a Markovian representation can be described in terms of a pair of spectral factors, one analytic and the other coanalytic.

**Theorem 10.2.15.** *Let  $(\mathbb{H}, \{U_t\}, \mathbf{X})$  be a proper Markovian representation of  $dy$  with inner triplet  $(K, Q, \bar{Q})$  and generating processes  $(dw, d\bar{w})$ . Then there is a unique pair  $(W, \bar{W})$  of spectral factors, the first being analytic and the second coanalytic, such that*

$$d\hat{y} = Wd\hat{w} = \bar{W}d\hat{w}. \quad (10.2.36)$$

Moreover,

$$d\hat{w} = Kd\hat{w}, \quad (10.2.37)$$

and

$$W = \bar{W}K, \quad (10.2.38a)$$

$$W = W_-Q, \quad \bar{W} = \bar{W}_+\bar{Q}, \quad (10.2.38b)$$

where  $W_-$  is the minimum-phase spectral factor and  $\bar{W}_+$  is the conjugate minimum-phase spectral factor.

**Proof.** By Proposition 10.2.12 (in the case that  $y$  is given by (10.2.25)) or Proposition 10.2.13 (in the case that  $dy$  is given by (10.2.30)), it immediately follows that there unique spectral factors  $W$  and  $\bar{W}$  such that (10.2.36) holds. Next, by Lemma 10.2.7, for any  $f \in L^2_p(\mathbb{I})$ ,  $J_{\hat{w}}^{-1}J_{\hat{w}}f = fK$ ; i.e.,

$$\int f d\hat{w} = \int fKd\hat{w},$$

proving (10.2.37). From (10.2.36) and (10.2.37) we have  $d\hat{y} = Wd\hat{w} = \bar{W}Kd\hat{w}$ , from which (10.2.38a) follows by uniqueness (Proposition 10.2.12 or Proposition 10.2.13).

Likewise, proceeding as above, we show that

$$d\hat{w}_- = Qd\hat{w} \quad \text{and} \quad d\hat{w}_+ = \bar{Q}d\hat{w},$$

which together with (10.2.34) yields (10.2.38b).  $\square$

As in Chapter 9, we call  $(K, W, \bar{W})$  of Theorem 10.2.15 the *Markovian triplet* of  $(\mathbb{H}, \{U_t\}, \mathbf{X})$ . In particular,

$$\bar{w}(h) - \bar{w}(0) = \int_{-\infty}^{\infty} \frac{e^{i\omega h} - 1}{i\omega} K(i\omega)d\hat{w}. \quad (10.2.39)$$

Obviously (10.2.38b) are the inner-outer factorizations of the spectral factors  $W$  and  $\bar{W}$ . Since the spectral factors  $W$  and  $\bar{W}$  are uniquely determined by the generating processes  $(dw, d\bar{w})$ , which, in turn, are unique modulo orthogonal transformations, as in Section 9.1, there is an equivalence class  $[K, W, \bar{W}]$  of Markovian triplets described by

$$(W, \bar{W}, K) \sim (WT_1, \bar{W}T_2, T_2^{-1}KT_1), \quad (10.2.40)$$

where  $T_1$  and  $T_2$  are arbitrary orthogonal transformations.

Internal Markovian representation have multiplicity  $p = m$ , and thus  $W$  and  $\bar{W}$  are square and hence invertible, since  $\Phi$  is full rank. In this case, therefore, (10.2.37) may be written

$$K = \bar{W}^{-1}W.$$

In particular, the predictor spaces  $\mathbf{X}_-$  and  $\mathbf{X}_+$  have Markovian triplets  $(W_-, \bar{W}_-, K_-)$  and  $(W_+, \bar{W}_+, K_+)$  respectively, where  $K_- := \bar{W}_-^{-1}W_-$  and  $K_+ := \bar{W}_+^{-1}W_+$ .

### From spectral factors to Markovian representations

Next, we consider the inverse problem of constructing Markovian representations from Markovian triplets. To this end, we need a procedure for constructing the generating processes of  $\mathbf{X} \sim (\mathbf{S}, \bar{\mathbf{S}})$  starting from  $(W, \bar{W}, K)$ . In the internal case this is a simple matter since  $W$  and  $\bar{W}$  can be inverted in (10.2.36) to yield unique  $dw$  and  $d\bar{w}$ . However, in general, the systems (10.2.36) are underdetermined, introducing nonuniqueness in the corresponding generating processes.

**Lemma 10.2.16.** *Let  $W$  be an  $m \times p$  spectral factor with the right inverse  $W^\sharp := W^* \Phi^{-1}$ , and let*

$$\Pi := I - W^\sharp W. \tag{10.2.41}$$

*Then, the  $p$ -dimensional Wiener processes  $d\hat{w}$  satisfying*

$$d\hat{y} = W d\hat{w} \tag{10.2.42}$$

*are given by*

$$d\hat{w} = W^\sharp d\hat{y} + d\hat{z} \tag{10.2.43}$$

*where  $d\hat{z}$  is any  $p$ -dimensional stationary increment process with incremental spectral density  $\frac{1}{2\pi}\Pi$  such that  $\mathbf{H}(d\hat{z}) \perp \mathbf{H}$ . Moreover,  $\Pi(i\omega)$  is an orthogonal projection on  $\mathbf{H}(d\hat{z})$  for almost all  $\omega \in \mathbb{R}$ , and*

$$d\hat{z} = \Pi d\hat{w}. \tag{10.2.44}$$

**Proof.** First note that, since  $\Pi(i\omega)^2 = \Pi(i\omega)$  and  $\Pi(i\omega)^* = \Pi(i\omega)$ ,  $\Pi(i\omega)$  is an orthogonal projection. For any  $d\hat{w}$  satisfying (10.2.42),  $W^\sharp d\hat{y} = (I - \Pi)d\hat{w}$ , and therefore (10.2.43) holds with  $d\hat{z}$  given by (10.2.44). Now,

$$E\{d\hat{z}d\hat{z}^*\} = \frac{1}{2\pi}\Pi^2 d\omega = \frac{1}{2\pi}\Pi d\omega,$$

and hence  $\frac{1}{2\pi}\Pi$  is the incremental spectral density of  $d\hat{z}$ . Moreover,  $E\{d\hat{y}d\hat{z}^*\} = \frac{1}{2\pi}W\Pi d\omega = 0$  establishing the orthogonality  $\mathbf{H}(d\hat{z}) \perp \mathbf{H}$ . Conversely, given a process  $d\hat{z}$  with a spectral density (10.2.44) and with  $\mathbf{H}(d\hat{z}) \perp \mathbf{H}$ , let  $d\hat{w}$  be given by (10.2.43). Then

$$E\{d\hat{w}d\hat{w}^*\} = \frac{1}{2\pi}W^\sharp \Phi (W^\sharp)^* d\omega + \frac{1}{2\pi}\Pi d\omega = \frac{1}{2\pi}I d\omega,$$

so  $d\hat{w}$  is a Wiener process. Moreover  $W d\hat{w} = d\hat{y} + W d\hat{z}$ . However,  $W E\{d\hat{z}d\hat{z}^*\} W^* = 0$ , and hence (10.2.42) holds.  $\square$

Consequently, given a Markovian triplet  $(W, \bar{W}, K)$ , by Lemma 10.2.16 we can construct pairs of generating processes

$$\begin{cases} d\hat{w} = W^\# d\hat{y} + d\hat{z} \\ d\hat{\bar{w}} = \bar{W}^\# d\hat{y} + d\hat{z} \end{cases}, \quad (10.2.45)$$

where the spectral density of  $dz$  is  $\frac{1}{2\pi}\Pi$  with  $\Pi$  given by (10.2.41), and that of  $d\bar{z}$  is  $\frac{1}{2\pi}\bar{\Pi}$ , where

$$\bar{\Pi} := I - \bar{W}^\# \bar{W}. \quad (10.2.46)$$

We now build the space  $\mathbb{H}$  corresponding to the Markovian representation so that  $\mathbb{H} = \mathbf{H}(dw) = \mathbf{H}(d\bar{w})$ . Of course, to this end, we must choose  $dz$  and  $d\bar{z}$  so that

$$\mathbf{H}(d\bar{z}) = \mathbf{H}(dz). \quad (10.2.47)$$

Then the multiplication operators  $M_\Pi$  and  $M_{\bar{\Pi}}$  both represent the projection  $E^{\mathbf{H}^\perp}$  from  $\mathbb{H}$  onto the doubly invariant subspace  $\mathbf{H}^\perp = \mathbf{H}(d\bar{z}) = \mathbf{H}(dz)$ . In fact, if  $\lambda \in \mathbb{H}$  and  $f := J_{\hat{w}}^{-1}\lambda$ ,

$$\lambda = \int fW^\# d\hat{y} + \int f d\hat{z}$$

by (10.2.45), and hence, in view of (10.2.44),

$$E^{\mathbf{H}^\perp} \lambda = \int f d\hat{z} = \int f \Pi d\hat{w} = J_{\hat{w}} f \Pi,$$

showing that  $E^{\mathbf{H}^\perp}$  corresponds to  $M_\Pi$  under the isomorphism  $J_{\hat{w}}$ . A symmetric argument shows that  $E^{\mathbf{H}^\perp}$  corresponds to  $M_{\bar{\Pi}}$  under the isomorphism  $J_{\hat{\bar{w}}}$ . More specifically,  $J_{\hat{w}} M_\Pi J_{\hat{w}}^{-1} = J_{\hat{\bar{w}}} M_{\bar{\Pi}} J_{\hat{\bar{w}}}^{-1}$ ; i.e.,  $M_\Pi J_{\hat{w}} J_{\hat{\bar{w}}} = J_{\hat{\bar{w}}}^{-1} J_{\hat{w}} M_{\bar{\Pi}}$ . Therefore, by Lemma 10.2.7,

$$K \Pi = \bar{\Pi} K, \quad (10.2.48)$$

from which we see that  $\bar{\Pi} d\hat{\bar{w}} = K \Pi d\hat{w}$ ; i.e.,

$$d\hat{\bar{z}} = K d\hat{z}. \quad (10.2.49)$$

The following theorem describes the relation between Markovian representations and Markovian triplets  $(W, \bar{W}, K)$ .

**Theorem 10.2.17.** *There is a one-one correspondence between proper Markovian representations  $(\mathbb{H}, \{U_t\}, \mathbf{X})$  and pairs  $([W, \bar{W}, K], dz)$ , where  $[W, \bar{W}, K]$  is an equivalence class of Markovian triplets and  $dz$  is a vector stationary-increment process (defined mod  $\mathcal{O}$ ) with spectral density  $\Pi := I - W^\# W$  such that  $\mathbf{H}(dz) \perp \mathbf{H}$ . Under this correspondence*

$$\mathbb{H} = \mathbf{H} \oplus \mathbf{H}(dz) \quad (10.2.50)$$

and

$$\mathbf{X} = \mathbf{H}^-(dw) \cap \mathbf{H}^+(d\bar{w}) \quad (10.2.51)$$

where  $(dw, d\bar{w})$  are the generating processes given by (10.2.45).

**Proof.** Given a Markovian representation  $(\mathbb{H}, \{U_t\}, \mathbf{X})$ , we have shown above that there is a unique equivalence class  $[W, \bar{W}, K]$  of Markovian triplets and a corresponding pair of generating processes  $(dw, d\bar{w})$ , defined mod  $\mathcal{O}$  and consequently a unique  $d\hat{z} = \Pi d\hat{w}$  having the required properties. Conversely, given a triplet  $(W, \bar{W}, K)$  and a process  $dz$  with the stated properties, we define  $(d\hat{w}, d\bar{\hat{w}})$  by (10.2.45) and set  $\mathbf{S} := \mathbf{H}^-(dw)$  and  $\bar{\mathbf{S}} := \mathbf{H}^+(d\bar{w})$ . Then since  $(W, \bar{W}, K)$  is a Markovian triplet,  $W$  is analytic implying that  $\mathbf{S} \supset \mathbf{H}^-$ ,  $\bar{W}$  is coanalytic implying that  $\bar{\mathbf{S}} \supset \mathbf{H}^+$ , and  $K$  is inner which is equivalent to perpendicular intersection. Hence, by Theorem 10.2.3,  $\mathbf{X} = \mathbf{S} \cap \bar{\mathbf{S}}$  is a Markovian splitting subspace with ambient space  $\mathbb{H} = \mathbf{H} \oplus \mathbf{H}(dz)$ , for the invariance condition (ii) is trivially satisfied. The shift is induced by  $dy$  and  $dz$ .  $\square$

Let  $(W_-, \bar{W}_-, K_-)$  and  $(W_+, \bar{W}_+, K_+)$  be the Markovian triplets of  $\mathbf{X}_-$  and  $\mathbf{X}_+$ , respectively. In view of (10.2.38b) and the fact that

$$W^\sharp := W^* \Phi^{-1} = Q^* \bar{W}_-^* (W_- W_-^*)^{-1} = Q^* W_-^{-1},$$

we may reformulate Definition 9.2.20 as follows.

**Definition 10.2.18.** An analytic spectral factor  $W$  of a strictly noncyclic process is *minimal* if  $W^\sharp W_+$  is analytic. Likewise, a coanalytic spectral factor  $\bar{W}$  is *minimal* if  $\bar{W}^\sharp \bar{W}_-$  is coanalytic.

In view of Theorem 10.2.9, we have the following counterpart of Corollary 9.2.21.

**Proposition 10.2.19.** *Let  $dy [y]$  be strictly noncyclic. Then there is a one-one correspondence (mod  $\mathcal{O}$ ) between minimal Markovian representations  $(\mathbb{H}, \{U_t\}, \mathbf{X})$  and pairs  $(W, z)$  where  $W$  is a minimal spectral factor and  $dz$  is a stationary process with the properties prescribed in Lemma 10.2.16.*

**Proof.** By Theorem 10.2.9 and condition (ii)' below it,  $\mathbf{X}$  is minimal if and only if  $W^\sharp W_+ = Q^* Q_+$  is analytic, i.e.  $W$  is minimal, and  $K^*$  and  $\bar{Q}$  are coprime. However, by (10.2.19) and (10.2.35),  $\bar{Q}K = \Theta Q = \bar{W}_+^{-1} W_- Q$ , which can be solved for  $\bar{Q}$  (and  $K$ ) uniquely mod  $\mathcal{O}$ . Hence  $\bar{W}$  is determined once  $W$  has been chosen.  $\square$

These result all hold for infinite- as well as finite-dimensional Markovian representations. In the next section we consider the special case that  $\mathbf{X}$  is finite dimensional, and in Section 10.5 we return to the general case.

### 10.3 Forward and backward realizations for finite-dimensional Markovian representations

Given a finite-dimensional Markovian representation  $(\mathbb{H}, \{U_t\}, \mathbf{X})$  determined by its Markovian triplet  $(W, \bar{W}, K)$  and its generating processes  $(dw, d\bar{w})$ , in this section we shall derive two stochastic realizations having the same splitting subspace  $\mathbf{X} \sim (\mathbf{S}, \bar{\mathbf{S}})$ , namely a forward realization  $\Sigma$  corresponding to  $\mathbf{S}$  with transfer function  $W$  and generating noise  $dw$ , and a backward one  $\bar{\Sigma}$  corresponding to  $\bar{\mathbf{S}}$  with transfer function  $\bar{W}$  and generating noise  $d\bar{w}$ . There are several reasons why it is natural and useful to study such pairs  $(\Sigma, \bar{\Sigma})$  of stochastic realizations. There is an intrinsic symmetry between past and future in the geometric theory which naturally carries over to the state-space representation  $\Sigma$  and  $\bar{\Sigma}$ . Recall, for example, that minimality is characterized by the two conditions of observability and constructability which are symmetric with respect to direction of time. As we shall see, observability is a property of  $\Sigma$  and constructibility a property of  $\bar{\Sigma}$ . In applications to noncausal estimation it is natural to consider, not only backward models, but also nonminimal representations which are best understood in terms of pairs  $(\Sigma, \bar{\Sigma})$ .

**Lemma 10.3.1.** *Let  $(\mathbb{H}, \{U_t\}, \mathbf{X})$  be a proper Markovian representation. Then  $\mathbf{X}$  is finite-dimensional if and only if its structural function  $K$  is rational.*

*Proof.* By Theorem 7.3.6 and (10.2.8),

$$\mathbf{X} = \mathbf{E}^{\mathbf{S}} \bar{\mathbf{S}} = \mathbf{E}^{\mathbf{H}^-(dw)} \mathbf{H}^+(\bar{w}) = \mathcal{J}_{\bar{w}} \overline{\text{range}}\{H_K\},$$

where  $H_K : \bar{H}_p^2 \rightarrow H_p^2$  is the Hankel operator with symbol  $K$  sending  $f$  to  $P^{H_p^2} fK$ , the orthogonal projection of  $fK$  onto  $H_p^2$ , which is finite-dimensional if and only if  $K$  is rational [34, Theorem 3.8, p. 256].  $\square$

Suppose that  $\mathbf{X}$  is finite-dimensional and that  $K$  has McMillan degree  $n$ . Then, by Lemma 10.3.1,

$$K(s) = \bar{M}(s)M(s)^{-1}, \quad (10.3.1)$$

for some invertible  $p \times p$  matrix polynomials  $M$  and  $\bar{M}$ , which are right coprime, i.e., any right divisor is unimodular [34]. The matrix polynomials  $M$  and  $\bar{M}$  are unique modulo a common unimodular factor. Since  $K$  is inner,  $\det M$  has all its roots in the open left half of the complex plane, and, we have  $\det \bar{M}(s) = \kappa \det M(-s)$ , where  $\kappa$  is a complex number with modulus one. To maintain the symmetry between the past and the future in our presentation, we also note that

$$K^*(s) = M(s)\bar{M}(s)^{-1}. \quad (10.3.2)$$

The following result shows that  $H(K)$ , the isomorphic image of  $\mathbf{X}$  under  $\mathcal{J}_{\bar{w}}$  (Theorem 10.2.8), consists of rational row-vector functions that are *strictly proper*; i.e., in each component, the numerator polynomial is of lower degree than the denominator polynomial.

**Theorem 10.3.2.** *Let  $K$  be a rational  $p \times p$  inner function with the polynomial matrix-fraction representation (10.3.1). Then*

$$H(K) = \{gM^{-1} \mid g \in \mathbb{R}^p[s]; gM^{-1} \text{ strictly proper}\}, \tag{10.3.3}$$

where  $\mathbb{R}^p[s]$  is the vector space of  $p$ -dimensional row vectors of polynomials.

**Proof.** Let us first show that, if  $K$  is rational, the space  $H(K)$  consists of strictly proper rational functions. To this end, set  $k := \det K$ . Then  $H_p^2 k \subset H_p^2 K$  [34, p. 187], and consequently  $H(K) \subset H(kI)$ . Therefore, it is no restriction to consider the scalar case  $p = 1$ . In fact, if  $K$  is rational, then so is  $k$ . Hence, if we can prove that the space  $H(k)$  of scalar functions consists of strictly proper rational functions, the same holds true for  $H(kI)$  and thus for  $H(K)$ . A scalar rational inner function  $k$  is a finite Blaschke product; i.e., a finite product of coprime functions  $k_i(s) := (s - s_i)^{\nu_i} (s + \bar{s}_i)^{-\nu_i}$ , where for each  $i$ ,  $s_i$  is a complex number,  $\bar{s}_i$  its complex conjugate, and  $\nu_i$  is an integer. Then  $H^2 k = \bigcap_i H^2 k_i$ , and hence  $H(k) = \bigvee_i H(k_i)$ , so it is enough to show that any  $H(k_i)$  consists of strictly proper rational functions. To this end, we quote from [27, p. 34] that

$$e_j(s) = \frac{1}{s + \bar{s}_i} \left[ \frac{s - s_i}{s + \bar{s}_i} \right]^j, \quad j = 0, 1, 2, \dots$$

is an orthogonal basis in  $H^2$ . However,  $e_j k_i = e_{i+j}$ , and hence  $H^2 k_i$  is spanned by  $\{e_{\nu_i}, e_{\nu_i+1}, \dots\}$ . Therefore,  $H(k_i)$  is the span of  $\{e_0, e_1, \dots, e_{\nu_i-1}\}$ , which is a space of strictly proper rational functions. Consequently, the same is true for  $H(K)$ , as required.

Next, recall that

$$H(K) = H_p^2 \cap (\bar{H}_p^2 K) = \{f \in H_p^2 \mid fK^* \in \bar{H}_p^2\}, \tag{10.3.4}$$

and hence, for any  $f \in H(K)$ ,  $\bar{f} := fK^* \in \bar{H}_p^2$ . Now, in view of (10.3.2),  $\bar{f} = fM\bar{M}^{-1}$ , and therefore

$$g := fM = \bar{f}\bar{M}.$$

Since  $fM$  is analytic in the closed right half-plane and  $\bar{f}\bar{M}$  is analytic in the closed left half-plane,  $g$  must be analytic in the whole complex plane. However, since  $f$  is rational, then so is  $g$ , and therefore  $g \in \mathbb{R}^p[s]$ . We have shown above that all  $f := gM^{-1} \in H(K)$  are strictly proper, and hence (10.3.3) follows from (10.3.4).  $\square$

**Corollary 10.3.3.** *Let  $(\mathbb{H}, \{U_t\}, \mathbf{X})$  be a finite-dimensional proper Markovian representation of  $y [dy]$  with Markovian triplet  $(K, W, \bar{W})$ . Then, if  $K$  is given by (10.3.1), there is an  $m \times p$  polynomial matrix  $N$  such that*

$$W(s) = N(s)M(s)^{-1}, \tag{10.3.5a}$$

$$\bar{W}(s) = N(s)\bar{M}(s)^{-1}. \tag{10.3.5b}$$

Moreover,  $W$  and  $\bar{W}$  are proper rational matrix functions.

**Proof.** From (10.2.38a) it follows that  $W = \bar{W}K$ , which together with (10.3.1) yields  $WM = \bar{W}\bar{M}$ , which matrix function we name  $N$ . Since  $WM$  is analytic in the right half plane and  $\bar{W}\bar{M}$  in the left,  $N$  must be an entire  $m \times p$  matrix function. Therefore, if we can show that  $W$  is rational, it follows that  $N$  is a polynomial matrix.

To this end, consider first the stationary case. Then,  $a'W \in H_p^2$  and  $a'\bar{W} \in \bar{H}_p^2$ , and hence since

$$a'\bar{W} = a'WK^*, \quad \text{for all } a \in \mathbb{R}^m, \tag{10.3.6}$$

(10.3.4) implies that  $a'W \in H(K)$  for all  $a \in \mathbb{R}^m$ . Hence  $W$  is rational and strictly proper (Theorem 10.3.2). In the stationary-increment case ( $dy$ ),  $a'W \in \mathcal{W}_p^2 = (1+s)H_p^2$  and  $a'\bar{W} \in \bar{\mathcal{W}}_p^2 = (1-s)\bar{H}_p^2$  (Theorem 5.2.3). Therefore, restating (10.3.6) to read

$$\frac{1}{1-s}a'\bar{W} = \frac{1}{1+s}a'W\tilde{K}^*, \quad \text{for all } a \in \mathbb{R}^m,$$

where  $\tilde{K}(s) := K(s)\frac{1-s}{1+s}$  is inner and rational, we see that  $\frac{1}{1+s}a'W \in H_p^2$ , which consists of (strictly proper) rational functions (Theorem 10.3.2). Hence  $\bar{W}$  is rational and proper.  $\square$

**Corollary 10.3.4.** *The process  $y$  [ $dy$ ] has a rational spectral density if and only if it has a finite-dimensional Markovian representation.*

**Proof.** The Markovian triplet of a finite-dimensional Markovian representation consists of rational matrix functions (Lemma 10.3.1 and Corollary 10.3.3), and hence the spectral density (10.2.27) must also be rational. Conversely, if the spectral density is rational, so are the outer and conjugate outer spectral factors, and hence the same holds for the phase function (10.2.35). Therefore the coprime factor  $K_+$  in (10.2.23) is rational, and hence  $\mathbf{X}_+$  is finite-dimensional (Lemma 10.3.1).  $\square$

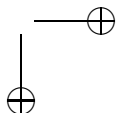
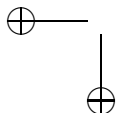
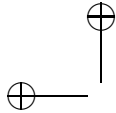
The structural function (10.3.1) can also be expressed as minimal realization

$$K(s) = I - \bar{B}'(sI - A)^{-1}B, \tag{10.3.7}$$

where  $(A, B)$  and  $(A', \bar{B})$  are reachable. Since the Markovian triplet  $(W, \bar{W}, K)$  is defined modulo orthogonal transformations (10.2.40), we can always choose a version of  $K$  such that  $K(\infty) = I$ . Since  $K$  is analytic, the eigenvalues of  $A$  lie in the open left complex halfplane.

**Lemma 10.3.5.** *Let  $K$  be a rational inner function with minimal realization (10.3.7). Then*

$$\bar{B} = P^{-1}B, \tag{10.3.8}$$





where  $P$  is the unique symmetric solution of the Lyapunov equation

$$AP + PA' + BB' = 0. \tag{10.3.9}$$

**Proof.** Given (10.3.7), we have

$$K(s)^{-1} = I + \bar{B}'(sI - A - B\bar{B}')^{-1}B$$

and

$$K^*(s) = K(-s)' = I + B'(sI + A')^{-1}\bar{B}.$$

However, since  $K$  is inner, we must have  $K(s)^{-1} = K^*(s)$ , and consequently there is a regular matrix  $P$  such that

$$(A + B\bar{B}', B, \bar{B}') = (-PA'P^{-1}, P\bar{B}, B'P^{-1}).$$

This implies that  $P$  satisfies the Lyapunov equation (10.3.7) and that  $\bar{B} = P^{-1}B$ , as claimed.  $\square$

**Lemma 10.3.6.** *Let  $K$  be a rational inner function with McMillan degree  $n$  and minimal realization (10.3.7), and let  $H(K)$  and  $\bar{H}(K)$  be the subspaces defined in Theorem 10.2.8. Then, the rows of  $(sI - A)^{-1}B$  form a basis in  $H(K)$  and the rows of  $(sI + A')^{-1}\bar{B}$  form a basis in  $\bar{H}(K^*)$ . In particular,*

$$\dim H(K) = \dim \bar{H}(K) = n. \tag{10.3.10}$$

**Proof.** A straight-forward calculation using (10.3.8) and (10.3.9) yields

$$(sI - A)^{-1}BK^*(s) = P(sI + A')^{-1}\bar{B}. \tag{10.3.11}$$

Therefore, since  $a'(sI - A)^{-1}B \in H_p^2$  and  $a'P(sI + A')^{-1}\bar{B} \in \bar{H}_p^2$  for all  $a \in \mathbb{R}^n$ , it follows from (10.3.4) that

$$a'(sI - A)^{-1}B \in H(K), \quad \text{for all } a \in \mathbb{R}^n;$$

i.e., the rows of  $(sI - A)^{-1}B$  all belong to  $H(K)$ . To see that they actually span  $H(K)$ , first compare (10.3.1) and (10.3.7) to observe that  $\deg \det M = n$ , which together with (10.3.3) implies that  $\dim H(K) = n$ . Moreover, since  $(A, B)$  is reachable, the  $n$  rows of  $(sI - A)^{-1}B$  are linearly independent, and consequently they span  $H(K)$ . The statements about  $\bar{H}(K)$  follow by symmetry.  $\square$

**Remark 10.3.7.** For any  $f, g \in H(K)$  there correspond  $a, b \in \mathbb{R}^n$  such that  $f = a'(sI - A)^{-1}B$  and  $g = b'(sI - A)^{-1}B$ . Then

$$\langle f, g \rangle_{H(K)} = a' \int_{-\infty}^{\infty} (i\omega I - A)^{-1}BB'(-i\omega - A')^{-1}d\omega b = a'Pb =: \langle a, b \rangle_P,$$

where  $P$  is the (positive definite) solution of the Lyapunov equation (10.3.9), thus defining a scalar product with respect to the basis in Lemma 10.3.6.

In view of the fact that  $H(K)$  is the isomorphic image of  $\mathbf{X}$  under  $\mathcal{J}_{\hat{w}}$ , the results of Lemmas 10.3.1 and 10.3.6 may be summarized as follows.

**Theorem 10.3.8.** *Let  $(\mathbb{H}, \{U_t\}, \mathbf{X})$  be a proper Markovian representation. Then  $\mathbf{X}$  is finite-dimensional if and only if its structural function  $K$  is rational, in which case the dimension of  $\mathbf{X}$  equals the McMillan degree of  $K$ .*

**Theorem 10.3.9.** *Let  $(\mathbb{H}, \{U_t\}, \mathbf{X})$  be an  $n$ -dimensional proper Markovian representation with generating processes  $(dw, d\bar{w})$  and structural function  $K$  given by (10.3.7), and consider the vector Markov processes  $x$  and  $\bar{x}$  defined by*

$$x(t) = \int_{-\infty}^t e^{A(t-\tau)} B dw(\tau), \tag{10.3.12a}$$

$$\bar{x}(t) = - \int_t^{\infty} e^{A'(\tau-t)} \bar{B} d\bar{w}(\tau). \tag{10.3.12b}$$

Then  $x(0)$  and  $\bar{x}(0)$  are two bases in  $\mathbf{X}$ . The processes  $x$  and  $\bar{x}$  are related by the linear transformation

$$\bar{x}(t) = P^{-1}x(t), \tag{10.3.13}$$

where  $P := E\{x(t)x(t)'\}$  is the unique symmetric solution of the Lyapunov equation (10.3.9) and  $\bar{B}$  is given (10.3.8). Moreover,

$$d\bar{w} = dw - \bar{B}'xdt. \tag{10.3.14}$$

**Proof.** Since  $A$  is a stability matrix, the integrals (10.3.12) are well-defined. Moreover, from (10.3.15a), we have

$$P := E\{x(t)x(t)'\} = \int_{-\infty}^t e^{A(t-\tau)} BB'e^{A'(t-\tau)} d\sigma = \int_0^{\infty} e^{A\sigma} BB'e^{A'\sigma} d\sigma,$$

which clearly is constant. Therefore differentiation with respect to  $t$  shows that  $P$  is the unique solution of the Lyapunov equation (10.3.9).

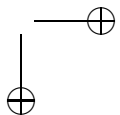
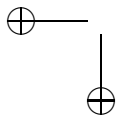
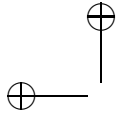
In view of (3.6.4), (10.3.12a) can be written as (3.6.11) and analogously for (10.3.12b); i.e.,

$$x(t) = \int_{-\infty}^{\infty} e^{i\omega t} (i\omega I - A)^{-1} B d\hat{w}, \tag{10.3.15a}$$

$$\bar{x}(t) = \int_{-\infty}^{\infty} e^{i\omega t} (i\omega I + A')^{-1} \bar{B} d\hat{w}. \tag{10.3.15b}$$

Therefore, by Theorem 10.2.8 and Lemma 10.3.5, Lemma 10.3.6 implies that  $x(0)$  and  $\bar{x}(0)$  are bases in  $\mathbf{X}$ , as claimed. Next, applying  $\mathcal{J}_{\hat{w}} M_{e^{i\omega t}}$  to both sides of (10.3.11) and observing that  $M_{K^*} = \mathcal{J}_{\hat{w}}^{-1} \mathcal{J}_{\hat{w}}$  (Lemma 10.2.7), we obtain

$$\int_{-\infty}^{\infty} e^{i\omega t} (i\omega I - A)^{-1} B d\hat{w} = P \int_{-\infty}^{\infty} e^{i\omega t} (i\omega I + A')^{-1} \bar{B} d\hat{w};$$



i.e.,  $x(t) = P\bar{x}(t)$ , which is the same as (10.3.13).

Finally, in view of (10.3.7) and (10.3.15a), (10.2.37) yields

$$\begin{aligned} \bar{w}(h) - \bar{w}(0) &= \int_{-\infty}^{\infty} \chi_h(i\omega)K(i\omega)d\hat{w} \\ &= \int_{-\infty}^{\infty} \chi_h(i\omega)d\hat{w} - \int_0^h \int_{-\infty}^{\infty} e^{i\omega t}\bar{B}'(i\omega I - A)^{-1}Bd\hat{w}dt \\ &= w(h) - w(0) - \int_0^h \bar{B}'x(t)dt, \end{aligned}$$

which establishes (10.3.14).  $\square$

**Corollary 10.3.10.** *Let  $(\mathbb{H}, \{U_t\}, \mathbf{X})$  be a finite-dimensional proper Markovian representation of a stationary process  $y$ , and let the processes  $x$  and  $\bar{x}$  be given by (10.3.12). Then there are unique  $m \times n$  matrices  $C$  and  $\bar{C}$  such that*

$$y(t) = Cx(t) = \bar{C}\bar{x}(t), \quad \bar{C} = CP, \tag{10.3.16}$$

where  $P$  is the unique solution of the Lyapunov equation (10.3.9).

**Proof.** Since  $a'y(0) \in \mathbf{H}^- \cap \mathbf{H}^+ \subset \mathbf{S} \cap \bar{\mathbf{S}} = \mathbf{X}$  for all  $a \in \mathbb{R}^m$  (Theorem 10.2.3) and  $x(0)$  and  $\bar{x}(0)$  are bases in  $\mathbf{X}$  (Theorem 10.3.9), there are matrices  $C$  and  $\bar{C}$  such that  $y(0) = Cx(0) = \bar{C}\bar{x}(0)$ , to which we apply the shift  $U_t$  componentwise to obtain the required representations. Then, (10.3.13) yields  $\bar{C} = CP$ .  $\square$

Consequently given any Markovian representation  $(\mathbb{H}, \{U_t\}, \mathbf{X})$  of a stationary process  $y$ , there two stochastic realizations of  $y$ , namely

$$(\Sigma_0) \begin{cases} dx = Axdt + Bdw \\ y = Cx, \end{cases} \quad (\bar{\Sigma}_0) \begin{cases} d\bar{x} = -A'\bar{x}dt + \bar{B}d\bar{w} \\ y = \bar{C}\bar{x}. \end{cases} \tag{10.3.17}$$

Here

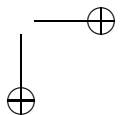
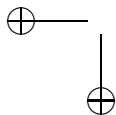
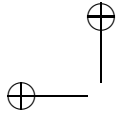
$$\{a'x(0) \mid a \in \mathbb{R}^n\} = \mathbf{X} = \{a'\bar{x}(0) \mid a \in \mathbb{R}^n\}, \tag{10.3.18}$$

and, since, by Theorem 10.2.3,

$$\mathbb{H} = \mathbf{H}^-(d\bar{w}) \oplus \mathbf{X} \oplus \mathbf{H}^+(dw), \tag{10.3.19}$$

$\Sigma_0$  is a forward realization and  $\bar{\Sigma}_0$  is a backward one. In fact, the future input noise in  $\Sigma_0$  is orthogonal to present state  $\mathbf{X}$  and past output  $\mathbf{H}^- \subset \mathbf{H}^-(dw)$  making the system *forward*, and the past input noise of  $\bar{\Sigma}_0$  is orthogonal to present state and future output  $\mathbf{H}^+$  making  $\bar{\Sigma}_0$  a *backward* system.

**Theorem 10.3.11.** *Let  $(\mathbb{H}, \{U_t\}, \mathbf{X})$  be a finite-dimensional proper Markovian representation of a stationary-increment process  $dy$  with incremental spectral density*



$\Phi$ , and let the processes  $x$  and  $\bar{x}$  be given by (10.3.12). Then there are unique matrices  $C$ ,  $\bar{C}$  and  $D$  such that

$$dy = Cxdt + Ddw \quad (10.3.20a)$$

$$dy = \bar{C}\bar{x}dt + Dd\bar{w}. \quad (10.3.20b)$$

Here  $D = W(\infty) = \bar{W}(\infty)$  satisfies

$$DD' = R := \Phi(\infty), \quad (10.3.21)$$

and the  $m \times n$  matrices  $C$  and  $\bar{C}$  satisfy

$$\bar{C} = CP + DB'. \quad (10.3.22)$$

Moreover, the Markovian triplet  $(K, W, \bar{W})$  is given by (10.3.7) and

$$W(s) = C(sI - A)^{-1}B + D \quad (10.3.23a)$$

$$\bar{W}(s) = \bar{C}(sI + A')^{-1}\bar{B} + D. \quad (10.3.23b)$$

**Proof.** By Corollary 10.3.3, the analytic spectral factor  $W$  is a proper rational matrix function. Hence,  $W = W(\infty) + G$ , where  $G$  is strictly proper with rows in  $H_p^2$ . Therefore, since  $\chi(i\omega) = \int_0^h e^{i\omega t} dt$ ,

$$\begin{aligned} y(h) - y(0) &= \int_{-\infty}^{\infty} \chi_h(i\omega)W(i\omega)d\hat{w} \\ &= W(\infty)[w(h) - w(0)] + \int_0^h z(t)dt, \end{aligned}$$

where

$$z_k(t) = \int_{-\infty}^{\infty} e^{i\omega t} G_k(i\omega)d\hat{w} = U_t z_k(0), \quad k = 1, 2, \dots, m$$

with  $G_k$  the  $k$ :th row of  $G$ ; i.e.,  $\{z(t)\}_{t \in \mathbb{R}}$  is the stationary vector process defined by the conditional derivatives

$$z_k(t) = \lim_{h \downarrow 0} \frac{1}{h} \mathbf{E}^{U_t \mathbf{S}}[y_k(t+h) - y_k(t)], \quad k = 1, 2, \dots, m,$$

where  $\mathbf{X} \sim (\mathbf{S}, \bar{\mathbf{S}})$ . Clearly, by Theorem 10.2.3,

$$a'z(0) \in \mathbf{E}^{\mathbf{S}} \mathbf{H}^- \subset \mathbf{E}^{\mathbf{S}} \bar{\mathbf{S}} = \mathbf{S} \cap \bar{\mathbf{S}} = \mathbf{X}$$

for all  $a \in \mathbb{R}^m$ . Therefore, since  $x(0)$  is a basis in  $\mathbf{X}$ , there is an  $m \times n$  matrix  $C$  such that  $z(0) = Cx(0)$ ; i.e.,  $z(t) = Cx(t)$ . Hence, setting  $D := W(\infty)$ , we obtain (10.3.20a). Then (10.3.23a) follows from this and (10.3.12a), and (10.3.21) is then immediate. Next, inserting  $dw = d\bar{w} + B'\bar{x}dt$ , obtained from (10.3.14), (10.3.8) and (10.3.13), into (10.3.20a), we obtain

$$dy = (CP + DB')\bar{x}dt + Dd\bar{w},$$

which is the same as (10.3.20b) if we set  $\bar{C} = CP + DB'$ . Then (10.3.23b) follows from (10.3.12b).  $\square$

Combining the representations of Theorems 10.3.9 and 10.3.11, we have now constructed a forward stochastic realization for  $(\mathbb{H}, \{U_t\}, \mathbf{X})$ ,

$$(\Sigma) \begin{cases} dx = Axdt + Bdw \\ dy = Cxdt + Ddw \end{cases}, \tag{10.3.24}$$

corresponding to the analytic spectral factor  $W$  and the forward generating process  $dw$  and a backward realization for  $(\mathbb{H}, \{U_t\}, \mathbf{X})$ ,

$$(\bar{\Sigma}) \begin{cases} d\bar{x} = -A'\bar{x}dt + \bar{B}d\bar{w} \\ dy = \bar{C}\bar{x}dt + Dd\bar{w} \end{cases}, \tag{10.3.25}$$

corresponding to the coanalytic spectral factor  $\bar{W}$  and the backward generating process  $d\bar{w}$ , such that (10.3.18) hold. As for the models without observation noise,  $\Sigma_0$  and  $\bar{\Sigma}_0$ , the forward and backward character of  $\Sigma$  and  $\bar{\Sigma}$ , respectively, is again a consequence of the splitting property (10.3.19).

**Remark 10.3.12.** We notice here that, by choosing  $K$  to be normalized as in (10.3.7), we can take  $D$  to be the same in the forward representation (10.3.20a) and in the backward representation (10.3.20b). Moreover, we may actually choose the nonzero part of  $D$  to be the same for all Markovian representations by, for example, choosing the arbitrary orthogonal transformation of  $dw$  so that

$$\begin{bmatrix} B \\ D \end{bmatrix} = \begin{bmatrix} B_1 & B_2 \\ R^{1/2} & 0 \end{bmatrix} \tag{10.3.26}$$

where  $R^{1/2}$  is the symmetric positive square root of  $R$ , and  $B_2$  is a full-rank matrix chosen in some canonical way. This shows a drastic difference from the situation in the discrete-time setting, where  $D$  could vary even in rank over the family of minimal Markovian representations. Hence we will not encounter the fine structure of degeneracy of Sections 8.8 and 9.3 in the continuous-time setting.

We close this section with the continuous-time versions of the results in Section 6.6.

**Theorem 10.3.13.** *Let  $(\mathbb{H}, \{U_t\}, \mathbf{X})$  be a Markovian representation with forward realization  $\Sigma [\Sigma_0]$  and backward realization  $\bar{\Sigma} [\bar{\Sigma}_0]$  and Markovian triplet  $(K, W, \bar{W})$ . Then, then given the representations of Corollary 10.3.3, the following statements are equivalent*

- (i)  $\mathbf{X}$  is observable,
- (ii)  $(C, A)$  is observable,
- (iii) the factorization  $W = NM^{-1}$  of (10.3.5a) is coprime.

Symmetrically the following statements are equivalent

- (iv)  $\mathbf{X}$  is constructible,
- (v)  $(\bar{C}, A')$  is observable,
- (vi) the factorization  $\bar{W} = N\bar{M}^{-1}$  of (10.3.5b) is coprime.

In particular,  $\deg W \leq \dim \mathbf{X}$  with equality if and only if  $\mathbf{X}$  is observable, and  $\deg \bar{W} \leq \dim \mathbf{X}$  with equality if and only if  $\mathbf{X}$  is constructible. Moreover,  $W$  [ $\bar{W}$ ] is minimal if and only if its degree is minimal.

**Proof.** We consider only the first part. The second follows by symmetry. The equivalence of (ii) and (iii) follows from, e.g., [34, p. 41], so it only remains to show that (i) and (ii) are equivalent. To this end, let us first consider the stationary case. Setting  $\xi = a'x(0)$ ,  $\xi \in \mathbf{X} \cap (\mathbf{H}^+)^{\perp}$  if and only if

$$a'x(0) \perp b'y(t) \quad \text{for all } b \in \mathbb{R}^m \text{ and } t \geq 0; \tag{10.3.27}$$

i.e.,  $E\{y(t)x(0)'\}a = 0$  all  $t \geq 0$ . However, since  $E\{y(t)x(0)'\}a = Ce^{At}Pa$ , this is equivalent to

$$Pa \in \bigcap_{t=0}^{\infty} \ker Ce^{At}, \tag{10.3.28}$$

and consequently, since  $P$  is nonsingular,  $\mathbf{X} \cap (\mathbf{H}^+)^{\perp} = 0$  if and only if  $\bigcap_{t=0}^{\infty} \ker Ce^{At} = 0$ ; i.e.,  $(C, A)$  is observable [60]. Next, consider the case that  $dy$  is a process with independent increments. Then (10.3.27) needs to be replaced by

$$a'x(0) \perp b'[y(t+h) - y(t)] \quad \text{for all } b \in \mathbb{R}^m \text{ and } t \geq 0,$$

where  $h > 0$ . This is the same as  $E\{[y(t+h) - y(t)]x(0)'\}a = 0$  all  $t \geq 0$ , or, equivalently,

$$\int_t^{t+h} Ce^{At}Padt = 0, \quad \text{for all } t \geq 0 \text{ and all } h > 0,$$

which, in turn is equivalent, to (10.3.28). Hence (i) and (ii) are equivalent also in this case.

Moreover, in view of the representations  $K = \bar{M}M^{-1}$ ,  $W = NM^{-1}$  and  $\bar{W} = N\bar{M}^{-1}$ , where the first is always coprime, the degrees of  $W$  and  $\bar{W}$  do not exceed that of  $K$ . Moreover,  $\deg W = \dim \mathbf{X}$  if and only if  $W = NM^{-1}$  is coprime, and  $\deg \bar{W} = \dim \mathbf{X}$  if and only if  $\bar{W} = N\bar{M}^{-1}$  is coprime. Finally, given a spectral factor  $W$ , let  $\mathbf{X}$  be a corresponding observable Markovian splitting subspace. Then  $\deg W = \dim \mathbf{X}$ . Since  $W$  is minimal if and only if  $\mathbf{X}$  is minimal (Proposition 10.2.19), and  $\mathbf{X}$  is minimal if and only if  $\dim \mathbf{X}$  is minimal (Corollary 7.6.3),  $W$  is minimal if and only if  $\deg W$  is minimal. The proof of the corresponding statement for  $\bar{W}$  is analogous.  $\square$

**Corollary 10.3.14.** Let  $(\mathbb{H}, \{U_t\}, \mathbf{X})$  be a Markovian representation with Markovian triplet  $(K, W, \bar{W})$ . Then the following conditions are equivalent.

- (i)  $\mathbf{X}$  is minimal,
- (ii)  $W$  is minimal and  $\deg W = \dim \mathbf{X}$ ,
- (iii)  $\bar{W}$  is minimal and  $\deg \bar{W} = \dim \mathbf{X}$ .

**Proof.** By Definition 10.2.18,  $W$  is minimal if and only if condition (ii)' on page 274, or, equivalently,  $\mathbf{X} \perp \mathbf{N}^+$  holds. By Theorem 10.3.13,  $\deg W = \dim \mathbf{X}$  if and only if  $\mathbf{X}$  is observable. Hence the equivalence of (i) and (ii) follows from Theorem 10.2.9. The equivalence of (i) and (iii) follows by symmetry.  $\square$

We also have the continuous-time counterpart of Corollary 6.6.5.

**Corollary 10.3.15.** *A stochastic realization  $\Sigma$  is minimal if and only if (i)  $(C, A)$  is observable, (ii)  $(A, B)$  is reachable, and (iii)  $(CP + DB, A)$  is observable, where  $P$  is the unique solution of the Lyapunov equation  $AP + PA' + BB' = 0$ .*

Note that minimality of a stochastic realization is a condition that involves both the forward and the backward realization. Moreover, the minimal realizations are characterized by the numerator polynomial matrix  $N$ ,  $W$  and  $\bar{W}$  having the same zeros.

Theorems 10.3.13 and 10.2.9 suggest a procedure for determining a coprime factorization of  $W = \bar{W}K$  for any analytic rational spectral factor.

**Corollary 10.3.16.** *Let  $W$  be an analytic rational spectral factor, let  $W = NM^{-1}$  be a coprime matrix fraction representation, and let  $\bar{M}$  be the solution of the matrix polynomial factorization problem*

$$\bar{M}(-s)' \bar{M}(s) = M(-s)' M(s) \tag{10.3.29}$$

*with all its zeros in the right half plane. Then the coprime factorization problem  $W = \bar{W}K$  has the solution  $K = \bar{M}M^{-1}$  and  $\bar{W} = N\bar{M}^{-1}$ , where the latter representation is coprime if and only if  $W$  is a minimal spectral factor.*

**Proof.** Since  $W = NM^{-1}$  is coprime, the corresponding  $\mathbf{X}$  is observable (Theorem 10.3.13). Then  $K^*$  and  $\bar{Q}$  are right coprime (Theorem 10.2.8); i.e., the factorization  $W = \bar{W}K$  is coprime. Then  $\bar{W} = NM^{-1}$  is coprime if and only if  $\mathbf{X}$  is minimal (Theorem 10.3.13), which in turn holds if and only if  $W$  is minimal (Theorem 10.2.9).  $\square$

## 10.4 Spectral factorization and Kalman filtering

In Section 7.7 we parametrized the family  $\mathcal{X}$  of equivalence classes of minimal Markovian splitting subspaces by a set  $\mathcal{P}$  of covariance matrices; see Remark 7.7.6. One of the main results of this section identifies the set  $\mathcal{P}$  with the solution set of a certain

linear matrix inequality and connects this with spectral factorization. This establishes a one-one correspondence between  $\mathcal{X}$  and the family (of equivalence classes) of minimal spectral factors, in harmony with Proposition 10.2.19.

### Uniform choice of bases

Let  $(\mathbb{H}, \{U_t\}, \mathbf{X})$  be an  $n$ -dimensional proper Markovian representation with forward and backward realizations,  $\Sigma$  and  $\bar{\Sigma}$ , given by (10.3.24) and (10.3.25), respectively. From (10.3.12a) we have, for  $t \geq 0$ ,

$$U_t(\mathbf{X})a'x(0) = a'e^{At}x(0), \quad \text{for all } a \in \mathbb{R}^n, \quad (10.4.1)$$

where  $U_t(\mathbf{X})$  is the restricted shift (10.2.5).

Now, consider the partial ordering of minimal Markovian splitting subspaces introduced in Section 7.7, and, in particular, the continuous-time version of Definition 8.6.1. Following Section 7.7, and in analogy with the theory of Section 8.6 for the discrete-time setting, we introduce a uniform choice of bases for the family of minimal Markovian splitting subspaces by first fixing a basis  $x_+(0)$  in  $\mathbf{X}_+$  and then choosing the basis  $x(0)$  in any other minimal  $\mathbf{X}$  so that

$$a'x(0) = E^{\mathbf{X}} a'x_+(0), \quad \text{for all } a \in \mathbb{R}^n. \quad (10.4.2)$$

As discussed in Section 7.7, the covariance matrices

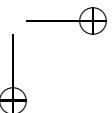
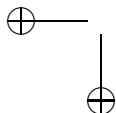
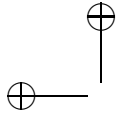
$$P = E\{x(0)x(0)'\} \quad (10.4.3)$$

of the corresponding bases form a set  $\mathcal{P}$  that is partially order so that  $P_1 \leq P_2$  if and only if  $\mathbf{X}_1 \prec \mathbf{X}_2$  (Definition 7.7.1 and Proposition 7.7.7). There is a one-one correspondence between  $\mathcal{P}$  and  $\mathcal{X}$ , the set of equivalence classes of minimal Markovian splitting subspaces. In fact, in this ordering,  $\mathcal{P}$  has a minimal element  $P_-$ , corresponding to the predictor space  $\mathbf{X}_-$ , and a maximal element  $P_+$ , corresponding to the backward predictor space  $\mathbf{X}_+$  (Theorem 7.7.3).

**Theorem 10.4.1.** *Consider the family of minimal Markovian representations of a stationary increment process  $dy$  with rational spectral density. Then all forward-backward pairs  $(\Sigma, \bar{\Sigma})$  of stochastic realizations (10.3.24)–(10.3.25) corresponding to a uniform choice of basis (10.4.2) have the same matrices  $A$ ,  $C$ , and  $\bar{C}$ . Conversely, for any realization (10.3.24) [(10.3.25)] there is a choice of basis  $x_+(0)$  in  $\mathbf{X}_+$  so that (10.4.2) holds.*

**Proof.** Since  $dy$  has a rational spectral density, all finite-dimensional Markovian splitting subspaces are finite-dimensional (Corollary 10.3.4 and Theorem 7.6.1). Let  $\mathbf{X} \sim (\mathbf{S}, \bar{\mathbf{S}})$  be an arbitrary minimal Markovian splitting subspace. We want to prove that  $(A, C, \bar{C})$  corresponding to  $\mathbf{X}$  equals  $(A_+, C_+, \bar{C}_+)$  corresponding to  $\mathbf{X}_+ \sim (\mathbf{S}_+, \bar{\mathbf{S}}_+)$ . First note that (10.4.2) may be written

$$a'x(0) = \hat{O}^* a'x_+(0), \quad \text{for all } a \in \mathbb{R}^n, \quad (10.4.4)$$





where  $\hat{\mathcal{O}}$  is the restricted observability map of Corollary 7.6.5, which, in the present finite-dimensional setting, is invertible (Corollary 7.6.6 and Theorem 10.2.9). Moreover, by Corollary 7.6.6, we have

$$U_t(X)\hat{\mathcal{O}}^*a'x_+(0) = \hat{\mathcal{O}}^*U_t(X_+)a'x_+(0),$$

and, since the left member equals  $U_t(X)a'x(0)$  because of (10.4.4), this is, in view of (10.4.1), equivalent to

$$a'e^{At}x(0) = \hat{\mathcal{O}}^*a'e^{A_+t}x_+(0).$$

Again applying (10.4.4), this is seen to be the same as

$$a'e^{At}x(0) = a'e^{A_+t}x_+(0),$$

yielding  $a'e^{At}P = a'e^{A_+t}P$  for all  $a \in \mathbb{R}^n$  and  $t \geq 0$ , where  $P := E\{x(0)x(0)'\}$  is nonsingular. This proves that  $A = A_+$ .

Next, recall from the proof of Theorem 10.3.11 that

$$b'Cx(0) = \lim_{h \downarrow 0} E^{\mathbf{S}} b'[y(h) - y(0)] = \lim_{h \downarrow 0} E^{\mathbf{X}} b'[y(h) - y(0)]$$

for all  $b \in \mathbb{R}^m$ . Since  $\mathbf{X}$  is minimal,  $\mathbf{X} \perp \mathbf{N}^+$  (Theorem 10.2.9), and hence, since  $\mathbf{S}_+ = (\mathbf{N}^+)^\perp$ , we have  $\mathbf{X} \subset \mathbf{S}_+ \oplus \mathbf{H}^\perp$ . Therefore, since  $b'[y(h) - y(0)] \perp \mathbf{H}$ ,

$$b'Cx(0) = \lim_{h \downarrow 0} E^{\mathbf{S}_+} b'[y(h) - y(0)] = E^{\mathbf{X}} b'C_+x_+(0) = b'C_+x(0),$$

where the last equality follows from (10.4.2). This establishes that  $C = C_+$ . A symmetric argument shows that  $\bar{C} = \bar{C}_-$ . Then, taking  $\mathbf{X} = \mathbf{X}_+$ , it follows that  $\bar{C}_+ = \bar{C}_-$ , and hence  $\bar{C} = \bar{C}_+$  for any minimal  $\mathbf{X}$ . Finally, to prove the last statement of the lemma, note that, since  $\hat{\mathcal{O}}^*$  is invertible,  $x_+(0)$  can be solved uniquely in terms of  $x(0)$  from (10.4.4). If instead  $\bar{x}(0)$  is given,  $x(0)$  can first be determined from (10.3.13).  $\square$

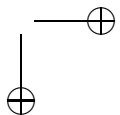
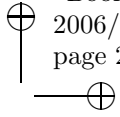
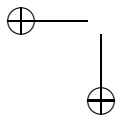
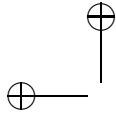
### Spectral factorization, the Linear Matrix Inequality and set $\mathcal{P}$

Since the matrices  $A, C, \bar{C}$  and  $R$  are invariant with the uniform choice of bases, one should be able to determine them from the spectral density  $\Phi$ . To show that this is indeed the case, insert (10.3.23a) into (10.2.27) to obtain

$$\begin{aligned} \Phi(s) &= [C(sI - A)^{-1}B + D][B'(-sI - A')^{-1}C' + D'] \\ &= C(sI - A)^{-1}BB'(-sI - A')^{-1}C' + C(sI - A)^{-1}BD' \\ &\quad + DB'(-sI - A')^{-1}C' + DD'. \end{aligned} \tag{10.4.5}$$

Now, let  $P$  be the unique symmetric solution of the Lyapunov equation  $AP + PA' + BB' = 0$ , given by (10.3.9), which we rewrite the form

$$BB' = (sI - A)P + P(-sI - A'). \tag{10.4.6}$$



Inserting this into (10.4.5) then yields

$$\Phi(s) = \Phi_+(s) + \Phi_+(-s), \tag{10.4.7}$$

where, setting  $\bar{C} := CP + DB'$  as in (10.5.57) and  $DD' = R$  as in (10.3.21),

$$\Phi_+(s) = C(sI - A)^{-1}\bar{C}' + \frac{1}{2}R. \tag{10.4.8}$$

An analytic function  $\Phi_+$  satisfying (10.4.7), where  $\Phi$  is a spectral density, is called *positive real*, and hence we shall refer to (10.4.8) as the *positive real part* of  $\Phi$ . We note that  $\Phi_+$  can be determined from  $\Phi$  by partial fraction expansion.

From this construction we also have the following simple but important observation.

**Proposition 10.4.2.** *For any rational analytic spectral factor  $W$  of  $\Phi$ ,*

$$\deg W \geq \deg \Phi_+ = \frac{1}{2} \deg \Phi, \tag{10.4.9}$$

where  $\Phi_+$  is the positive real part of  $\Phi$ . If  $W$  is a minimal spectral factor,  $\deg W = \deg \Phi_+$ .

**Proof.** Let  $(A, B, C, D)$  be a minimal realization of  $W$ . Then, if  $A$  is  $n \times n$ ,  $\deg W = n$ . From the construction above we have  $\deg \Phi_+ \leq n$ , and hence  $\deg W \geq \deg \Phi_+$ . From (10.4.8) we have  $\deg \Phi = 2 \deg \Phi_+$ . In view of Theorem 10.3.13, to prove the last statement we only need to show that the equality can be attained in (10.4.9). To this end, let now  $(C, A, \bar{C}, \frac{1}{2}R)$  be a minimal realization of  $\Phi_+$ . Then, the Markovian splitting subspace  $\mathbf{X}$  with the basis  $x(0)$  given by (10.3.12a) yields the stochastic realization (10.3.24) with transfer function (10.3.23a). Then, if  $A$  is  $n \times n$ ,  $\dim \mathbf{X} = n = \dim \Phi_+$ , and consequently  $\deg W = n$  by Theorem 10.3.13 and (10.4.9).  $\square$

Conversely, suppose that  $\Phi_+$  is given, and let (10.4.8) be a minimal realization of  $\Phi_+$ ; i.e.,  $(C, A)$  is observable and  $(A, \bar{C}')$  is reachable, or, equivalently,  $(C, A)$  and  $(\bar{C}, A')$  are observable. Then it follows from (10.4.7) that

$$\Phi(s) = \begin{bmatrix} C(sI - A)^{-1} & I \end{bmatrix} \begin{bmatrix} 0 & \bar{C}' \\ \bar{C} & R \end{bmatrix} \begin{bmatrix} (-sI - A')^{-1}C' \\ I \end{bmatrix}. \tag{10.4.10}$$

However, in view of the identity

$$-AP - PA' = (sI - A)P + P(-sI - A'),$$

we have, for all symmetric  $P$

$$0 = \begin{bmatrix} C(sI - A)^{-1} & I \end{bmatrix} \begin{bmatrix} -AP - PA' & -PC' \\ -CP & 0 \end{bmatrix} \begin{bmatrix} (-sI - A')^{-1}C' \\ I \end{bmatrix}, \tag{10.4.11}$$

which added to (10.4.10) yields

$$\Phi(s) = [C(sI - A)^{-1} \quad I] M(P) \begin{bmatrix} (-sI - A')^{-1} C' \\ I \end{bmatrix}, \quad (10.4.12)$$

where  $M : \mathbb{R}^{n \times n} \rightarrow \mathbb{R}^{(n+m) \times (n+m)}$  is the linear map

$$M(P) = \begin{bmatrix} -AP - PA' & \bar{C}' - PC' \\ \bar{C} - CP & R \end{bmatrix}. \quad (10.4.13)$$

If therefore  $P$  satisfies the *Linear Matrix Inequality*

$$M(P) \geq 0, \quad (10.4.14)$$

there is a minimal factorization

$$M(P) = \begin{bmatrix} B \\ D \end{bmatrix} [B' \quad D'], \quad (10.4.15)$$

where the factor  $\begin{bmatrix} B \\ D \end{bmatrix}$  is unique modulo an orthogonal transformation from the left. Inserted into (10.4.12), this yields

$$W(s)W(-s)' = \Phi(s), \quad (10.4.16)$$

where

$$W(s) = C(sI - A)^{-1}B + D. \quad (10.4.17)$$

**Theorem 10.4.3 (Positive Real Lemma).** *The rational matrix function  $\Phi_+$  with minimal realization (10.4.8) is positive real if and only if the Linear Matrix Inequality (10.4.14) has a symmetric solution  $P$ .*

Therefore the equations

$$AP + PA' + BB' = 0 \quad (10.4.18a)$$

$$PC' + BD' = \bar{C}' \quad (10.4.18b)$$

$$DD' = R \quad (10.4.18c)$$

are called the *positive-real-lemma equations*.

**Proof.** Suppose that  $P$  is a symmetric matrix satisfying (10.4.14). Then  $M(P)$  can be factored as in (10.4.15) to yield a spectral factor (10.4.17). Hence

$$[C(sI - A)^{-1} \quad I] M(P) \begin{bmatrix} (-sI - A)^{-1} C' \\ I \end{bmatrix} = W(s)W(-s)',$$

from which we subtract the identity (10.4.11) to yield

$$\Phi_+(s) + \Phi_+(-s) = W(s)W(-s)'. \quad (10.4.19)$$

Hence  $\Phi_+$  is positive real. Conversely, if  $\Phi_+$  is positive real, there is a spectral factor  $W$  that we may take to be minimal so that (10.4.19). Let

$$W(s) = H(sI - F)^{-1}B + D \tag{10.4.20}$$

be a minimal realization. Moreover, let  $P$  be the unique solution of the Lyapunov equation  $FP + PF' + BB' = 0$  and set  $G := PH' + BD'$ . Then, forming  $W(s)W(-s)'$  as in (10.4.5), we obtain (10.4.7), where now the positive part of  $\Phi$  becomes

$$\Phi_+(s) = H(sI - A)^{-1}G + \frac{1}{2}DD'$$

However, then there must be a nonsingular matrix  $T$  such that

$$(H, F, G) = (CT^{-1}, TAT^{-1}, T\bar{C}')$$

Then, choosing  $T = I$ ,  $P$  satisfies the positive-real lemma equations (10.4.18), and hence  $M(P) \geq 0$ .  $\square$

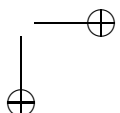
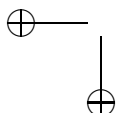
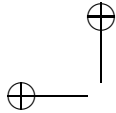
**Theorem 10.4.4.** *Given a minimal realization (10.4.8) of the positive real part  $\Phi_+$  of the spectral density  $\Phi$ , let  $M$  be the linear map given by (10.4.13). Then there is a one-one correspondence between the symmetric solutions of the linear matrix inequality (10.4.14) and the family of equivalence classes of minimal spectral factors of  $\Phi$ . In fact, given a symmetric solution  $P$  of (10.4.14), take  $\begin{bmatrix} B \\ D \end{bmatrix}$  to be the unique (mod  $O$ ) full-rank factor of  $M(P)$  as in (10.4.15) and let  $W(s)$  be given by (10.4.17). Then  $W$  is a minimal spectral factor. Viceversa, given an equivalence class  $[W]$  of  $W$  as in (10.4.17), there is a unique symmetric  $P > 0$  solving (10.4.18) and hence (10.4.14).*

**Proof.** Let  $P$  be a solution of (10.4.14). If the matrix  $A$  is  $n \times n$ ,  $\deg \Phi_+ = n$ . Then the spectral factor (10.4.17), obtained by the construction above, satisfies  $\deg W \leq n$ . Hence, by Proposition 10.4.2,  $W$  is minimal. Conversely, given a minimal spectral factor (10.4.17) with minimal realization (10.4.20), proceed as in the end of the proof of Theorem 10.4.3 to show that there is a unique  $P$  such that  $M(P) \geq 0$ .  $\square$

We are now in a position to prove the following theorem, which establishes every symmetric solution  $P$  of (10.4.14) as a legitimate state covariance.

**Theorem 10.4.5.** *The ordered set  $\mathcal{P}$  of state covariances (10.4.3) is precisely the set of all symmetric solutions of the linear matrix inequality (10.4.14).*

**Proof.** By definition, each  $P \in \mathcal{P}$  corresponds to a minimal stochastic realization (10.3.24). Therefore, Theorems 10.3.9 and 10.3.11 imply that  $P$  satisfies (10.4.18) for some  $(B, D)$ , and hence (10.4.14). Conversely, suppose that  $P$  satisfies the linear matrix inequality (10.4.14). Let  $W$  be a corresponding minimal spectral factor



(10.4.17), prescribed by Theorem 10.4.4. Then, choosing an arbitrary  $dz$  of appropriate dimension, (10.2.43) defines a generating process  $dw$ , which together with  $(A, B, C, D)$  yields a forward realization (10.3.24) with a state process (10.3.12a) such that  $x(0)$  has the covariance matrix  $P$  and is a basis in the corresponding splitting subspace  $\mathbf{X}$  (Theorem 10.3.9). Now,  $\dim \mathbf{X} = \deg W$ , and hence, since  $W$  is minimal,  $\mathbf{X}$  is minimal (Corollary 10.3.14). Then, by Theorem 10.4.1, there is a basis  $x_+(0)$  in  $\mathbf{X}_+$  and a corresponding stochastic realization  $\Sigma_+$  having parameters  $(A, C, \bar{C})$  and a state process  $x_+$  such that (10.4.2) holds. Hence  $P \in \mathcal{P}$ , as claimed.  $\square$

In particular, it follows from this theorem and Theorem 10.4.5 that two finite-dimensional minimal Markovian splitting subspaces are equivalent (in the sense defined in Section 7.7) if and only if they have the same analytic (coanalytic) spectral factor  $W (\bar{W}) \bmod O$ .

**Theorem 10.4.6.** *The family  $\mathcal{P}$  is a closed, bounded, convex set with a maximal element  $P_+$  and a minimal element  $P_-$ . Here  $P_+ = E\{x_+(0)x_+(0)\}$ , where  $x_+(0)$  is the selected basis in the backward predictor space  $\mathbf{X}_+$ , and  $P_- = E\{x_-(0)x_-(0)\}$ , where  $x_-(0)$  is the uniformly chosen basis in the predictor space  $\mathbf{X}_-$ , i.e.,  $a'x_-(0) = E^{\mathbf{X}_-} a'x_+(0)$  for all  $a \in \mathbb{R}^n$ .*

**Proof.** It follows immediately from the linear matrix inequality (10.4.14) that  $\mathcal{P}$  is closed and convex. Proposition 7.7.5 states that the partially ordered set  $\mathcal{P}$  and  $\mathcal{X}$  are isomorphic. Therefore, since  $\mathcal{X}$  has a maximal element,  $X_+$ , and a minimal element,  $X_-$ , given by (7.7.9), there are corresponding  $P_+$  and  $P_-$  with the properties stated. From this it also follows that  $\mathcal{P}$  is bounded.  $\square$

### The algebraic Riccati inequality

Recall that a rational spectral density  $\Phi$  is a rational  $m \times m$  matrix function that is positive semidefinite on the imaginary axis  $\mathbb{I}$  and *parahermitian* in the sense that

$$\Phi(-s) = \Phi(s)'$$

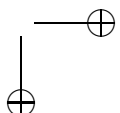
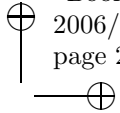
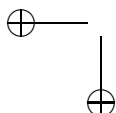
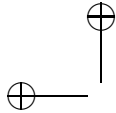
As before, we assume that the positive real part  $\Phi_+$  has the minimal realization

$$\Phi_+(s) = C(sI - A)^{-1}\bar{C}' + \frac{1}{2}R, \tag{10.4.21}$$

where  $R := \Phi(\infty)$ .

From now on we shall also assume that the spectral density  $\Phi$  is *coercive*; i.e.,  $\Phi$  has no zeros on the imaginary axis  $\mathbb{I}$  including the points at infinity. In particular this implies that  $R > 0$ . Then the set  $\mathcal{P}$  can be identified with the symmetric solutions of the *algebraic Riccati inequality*

$$\Lambda(P) \leq 0, \tag{10.4.22}$$



where  $\Lambda : \mathbb{R}^{n \times n} \rightarrow \mathbb{R}^{n \times n}$  is the quadratic matrix function

$$\Lambda(P) = AP + PA' + (\bar{C} - CP)'R^{-1}(\bar{C} - CP) \tag{10.4.23}$$

where  $(A, C, \bar{C})$  are given by (10.4.21). In fact, since  $R > 0$ ,  $M(P)$  can be block diagonalized as

$$\begin{bmatrix} I & T \\ 0 & I \end{bmatrix} M(P) \begin{bmatrix} I & 0 \\ T' & I \end{bmatrix} = \begin{bmatrix} -\Lambda(P) & 0 \\ 0 & R \end{bmatrix}$$

where

$$T = -(\bar{C} - CP)'R^{-1},$$

from which it follows that  $M(P) \geq 0$  if and only if  $\Lambda(P) \leq 0$ . Moreover, for any  $P \in \mathcal{P}$ ,

$$p := \text{rank } M(P) = m + \text{rank } \Lambda(P) \geq m.$$

Consequently, in view of (10.4.12), (10.4.15) and (10.4.17), those  $P \in \mathcal{P}$  that correspond to square  $m \times m$  spectral factors  $W$ , and hence to internal Markovian representations, are precisely those that satisfy the *Algebraic Riccati Equation*

$$\Lambda(P) = 0. \tag{10.4.24}$$

**Proposition 10.4.7.** *Suppose  $R := \Phi(\infty) > 0$ . Then*

$$\mathcal{P} = \{P \mid P' = P; \Lambda(P) \leq 0\}. \tag{10.4.25}$$

*Moreover, the subset  $\mathcal{P}_0 \subset \mathcal{P}$  corresponding to internal Markovian splitting subspaces is given by*

$$\mathcal{P}_0 = \{P \mid P' = P; \Lambda(P) = 0\}. \tag{10.4.26}$$

It is convenient in this situation to fix a representative in each equivalence class of spectral factors by choosing the arbitrary orthogonal transformation in the factorization of (10.4.15) so that (10.3.26) holds. Then (10.4.18b) can be solved for  $B_1$ ; i.e.,

$$B_1 = (\bar{C} - PC)'R^{-1/2}, \tag{10.4.27}$$

which inserted in (10.4.18b) yields

$$\Lambda(P) = -B_2B_2'. \tag{10.4.28}$$

Now, to each  $P \in \mathcal{P}$  there corresponds in a one-to-one fashion an element in  $\mathcal{X}$ ; i.e., an equivalence class of minimal Markovian splitting subspaces with a forward realization

$$\begin{cases} dx = Axdt + B_1du + B_2dv \\ dy = Cxdt + R^{1/2}du \end{cases} \tag{10.4.29}$$

which is uniquely determined except for the arbitrariness of the possible external part of the driving noise  $dw = \begin{bmatrix} du \\ dv \end{bmatrix}$ . Clearly, the internal realizations (10.4.29) are precisely those for which  $B_2 = 0$  so that (10.4.24) holds.

### Kalman filtering

Let  $\Sigma$  be a linear observable (but not necessarily minimal) stochastic system (10.3.24) with state covariance  $P$ , and let  $\mathbf{H}_{[0,t]}^-(dy)$  be the subspace generated by the observed process  $dy$  on the finite interval  $[0, t]$ . Then the linear minimum-variance estimate  $\{\hat{x}(t) \mid t \geq 0\}$  defined by

$$a' \hat{x}(t) = E^{\mathbf{H}_{[0,t]}^-(dy)} a' x(t), \quad \text{for all } a \in \mathbb{R}^n \text{ and } t \geq 0, \quad (10.4.30)$$

is given by the Kalman filter

$$d\hat{x} = A\hat{x}dt + K(t)[dy - C\hat{x}dt], \quad \hat{x}(0) = 0, \quad (10.4.31)$$

where the gain

$$K(t) = [Q(t)C' + BD']R^{-1} \quad (10.4.32)$$

is determined by the error covariance matrix function

$$Q(t) = E\{[x(t) - \hat{x}(t)][x(t) - \hat{x}(t)]'\}, \quad (10.4.33)$$

which satisfies the matrix Riccati equation

$$\begin{cases} \dot{Q} = AQ + QA' - (QC' + BD')R^{-1}(QC' + BD')' + BB' \\ Q(0) = P \end{cases} . \quad (10.4.34)$$

(See, e.g., [60].) It is also well-known and is demonstrated below (Corollary 10.4.10), that, under the present conditions,  $Q(t)$  tends to a limit  $Q_\infty \geq 0$  as  $t \rightarrow \infty$ , thus defining a *steady-state Kalman filter*

$$d\hat{x} = A\hat{x}dt + K_\infty[dy - C\hat{x}dt], \quad (10.4.35)$$

where the gain  $K_\infty$  is constant, the system is defined on the whole real line, and

$$a' \hat{x}_\infty(t) = E^{U_t \mathbf{H}^-} a' x(t), \quad \text{for all } a \in \mathbb{R}^n \text{ and all } t \in \mathbb{R}. \quad (10.4.36)$$

Let the stationary process represented by this system be denoted  $\hat{x}_\infty(t)$ . Then, because the innovation process

$$d\nu = R^{1/2}[dy - C\hat{x}_\infty dt] \quad (10.4.37)$$

is a Wiener process (see, e.g., [78] for details), (10.4.35) defines a stochastic realization

$$\begin{cases} \hat{x}_\infty = A\hat{x}_\infty dt + K_\infty R^{-1/2} d\nu \\ dy = C\hat{x}_\infty dt + R^{1/2} d\nu \end{cases} \quad (10.4.38)$$

of  $dy$  on the real line. By assumption, the Markovian splitting subspace  $\mathbf{X}$  defined by  $\Sigma$  is observable, and hence Proposition 7.4.13 and Corollary 7.4.14 imply that

$$E^{\mathbf{H}^-} \mathbf{X} = \mathbf{X}_-. \quad (10.4.39)$$

Consequently, since

$$E^{H^-} a'x(0) = a'\hat{x}_\infty(0), \quad \text{for all } a \in \mathbb{R}^n, \quad (10.4.40)$$

$\hat{x}_\infty(0)$  is a generator of  $\mathbf{X}_-$ . As explained in Section 10.1,  $\hat{x}_\infty(0)$  is a basis if and only if the model (10.4.38) is reachable. We shall prove that reachability of (10.4.38) is equivalent to minimality of the underlying model  $\Sigma$ .

**Proposition 10.4.8.** *An observable system  $\Sigma$  is a minimal realization of  $dy$  if and only if its steady state Kalman filter (10.4.38) is reachable.*

**Proof.** Let the dimension of  $\mathbf{X}_-$  be  $n$ . Then all minimal  $\mathbf{X}$  have this dimension (Theorem 7.6.1). We have already seen above that (10.4.38) is reachable if and only if the dimension of  $\hat{x}_\infty(0)$  is  $n$ . However,  $\dim \mathbf{X} \leq \dim x(0) = \dim \hat{x}_\infty(0)$ , and consequently (10.4.38) is reachable if and only if  $\dim \mathbf{X} \leq n$ , from which the stated result follows.  $\square$

Now, suppose that the linear stochastic system  $\Sigma$ , regarded as a realization of  $dy$ , is minimal. Then, it follows from what has just been discussed that the steady-state Kalman filtering estimate  $\hat{x}_\infty$  equals  $x_-$ , the (forward) state process corresponding to the predictor space  $\mathbf{X}_-$  in a uniform basis. To see this, compare (10.4.40) with (7.7.15) in Proposition 7.7.7, remembering that, by splitting,  $E^{H^-} \lambda = E^{\mathbf{X}_-} \lambda$  for all  $\lambda \in \bar{\mathbf{S}} \supset \mathbf{X}$ .

With  $\Sigma$  being an arbitrary minimal stochastic realization, we would like to express the Kalman-filtering equations (10.4.32) and (10.4.34) in terms of the invariant parameters  $(A, C, \bar{C}, R)$  of the realization (10.4.21) of  $\Phi_+$ . To this end, introduce a change of variables

$$\Pi(t) := E\{\hat{x}(t)\hat{x}(t)'\} = P - Q(t) \quad (10.4.41)$$

and use the positive real lemma equations (10.4.18) to transform (10.4.32) and (10.4.34) into

$$K(t) = [\bar{C} - C\Pi(t)]R^{-1} \quad (10.4.42)$$

and

$$\dot{\Pi} = \Lambda(\Pi), \quad \Pi(0) = 0, \quad (10.4.43)$$

where  $\Lambda$  is defined by (10.4.23). The matrix Riccati equation (10.4.43) is invariant in the sense that it is independent of the particular choice of model  $\Sigma$ , in harmony with the property (10.4.43).

**Proposition 10.4.9.** *The matrix Riccati equation (10.4.43) has a unique solution for  $t \in (0, \infty)$ . Moreover, for each  $P \in \mathcal{P}$ ,*

$$0 \leq \Pi(\tau) \leq \Pi(t) \leq P, \quad \tau \leq t. \quad (10.4.44)$$

Finally, as  $t \rightarrow \infty$ ,  $\Pi(t) \rightarrow P_- \in \mathcal{P}_0$ .



**Proof.** First note that  $P - \Pi(t) = Q(t) \geq 0$  for all  $t \geq 0$ . This establishes the upper bound. Next, differentiate (10.4.43) to obtain the matrix differential equation

$$\ddot{\Pi} = \Gamma \dot{\Pi} + \dot{\Pi} \Gamma', \quad \dot{\Pi}(0) = \Lambda(0) = \bar{C} R^{-1} \bar{C}',$$

where  $\Gamma := A - (\bar{C} - C\Pi)' R^{-1} C$ . Integrating this, we obtain

$$\dot{\Pi}(t) = \int_0^t \Psi(t, \tau) \bar{C} R^{-1} \bar{C}' \Psi(t, \tau)' d\tau \geq 0,$$

where  $\Psi$  is the Green's function satisfying

$$\frac{\partial}{\partial t} \Psi(t, \tau) = \Gamma(t) \Psi(t, \tau), \quad \Psi(\tau, \tau) = I.$$

Consequently,  $\Pi$  is monotone nondecreasing and bounded from above. Therefore, as  $t \rightarrow \infty$ ,  $\Pi(t)$  tends to a limit  $\Pi_\infty$ , which must satisfy  $\Pi_\infty \leq P$  for all  $P \in \mathcal{P}$ . However,  $\Lambda(\Pi_\infty) = 0$ ; i.e.,  $\Pi_\infty \in \mathcal{P}_0 \subset \mathcal{P}$ . Consequently,  $\mathcal{P} = P_-$ , as claimed.  $\square$

In view (10.4.41), we immediately have the following corollary of Proposition 10.4.9.

**Corollary 10.4.10.** *The matrix Riccati equation (10.4.34) has a unique solution for  $t \in (0, \infty)$ , and  $Q(t)$  tends to a limit  $Q_\infty \geq 0$  as  $t \rightarrow \infty$ .*

Analogously, starting from a minimal backward realization (10.3.25), we can define a backward Kalman filter, the steady-state version of which can be identified with the backward realization of  $\mathbf{X}_+$ . This yields a dual matrix Riccati equation

$$\dot{\bar{\Pi}} = \bar{\Lambda}(\bar{\Pi}), \quad \bar{\Pi}(0) = 0, \tag{10.4.45}$$

where

$$\bar{\Lambda}(\bar{P}) = A' \bar{P} + (\bar{P} A + (C - \bar{C}(\bar{P})' R^{-1} (C - \bar{C}(\bar{P})). \tag{10.4.46}$$

A symmetric argument then shows that

$$\bar{\Pi}(t) \rightarrow \bar{P}_+ = (P_+)^{-1} \tag{10.4.47}$$

and  $\bar{P}_+ \leq \bar{P}$ , or, equivalently,  $P \leq P_+$  or all  $P \in \mathcal{P}$ .

## 10.5 Forward and backward stochastic realizations (the general case)

In Section 10.3, given a Markovian representation  $(\mathbb{H}, \{U_t\}, \mathbf{X})$  of finite dimension  $n$ , we constructed a state process  $\{x(t); t \in \mathbb{R}\}$  taking values in  $\mathbb{R}^n$  and forward and backward differential equation representations for it. If  $(\mathbb{H}, \{U_t\}, \mathbf{X})$  is a Markovian representation of a stationary process  $y$ , this construction leads to pair of forward and backward realizations (10.3.17). Similarly if  $(\mathbb{H}, \{U_t\}, \mathbf{X})$  corresponds to a

stationary increment process  $dy$ , the construction leads to a pair of realizations (10.3.24)–(10.3.25). This corresponds to the situation where  $y$  ( $dy$ ) has a rational spectral density (incremental spectral density).

On the other hand, the geometric theory of Markovian representations is absolutely independent of any restrictions of the dimension of  $\mathbf{X}$ . Moreover, many engineering problems involve random processes with nonrational spectra, e.g. turbulence, wave spectra, gyroscopic noise, etc. The natural question to ask at this point is thus the following. Given a Markovian splitting subspace of an infinite dimension, when is it possible to obtain differential equation representations of the type described above?

This is basically a representation problem in which one seeks a global description in terms of local or infinitesimal data. As such it has no meaningful solution in general. Obtaining differential equation representations for a process with non-rational spectrum necessarily involves restrictions of a technical nature (essentially smoothness conditions) on the underlying spectral factors; such restrictions do not occur in discrete-time setting of Chapter 8. The elucidation of these conditions is one of the goals of this section. Note that there are several possible mathematical frameworks for infinite-dimensional Markov processes as solutions of stochastic differential equations, all of which coincide when specialized to the finite-dimensional case. Here we shall work in a setting which looks most natural to us, but other approaches are possible.

### Forward state representation

Suppose that  $(\mathbb{H}, \{U_t\}, \mathbf{X})$  is an infinite-dimensional, proper Markovian representation with the forward generating process  $dw$ . As in Section 8.7 (p. 216) we want to construct an  $\mathcal{X}$ -valued stochastic process, where  $\mathcal{X}$  is defined from  $\mathbf{X}$  via an isomorphism  $T : \mathbf{X} \rightarrow \mathcal{X}$  such that  $\langle T\xi, T\eta \rangle_{\mathcal{X}} = \langle \xi, \eta \rangle_{\mathbf{X}}$ . To this end, we define  $\mathcal{X}$  as

$$\mathcal{X} = (\mathcal{J}_w)^{-1}\mathbf{X}, \tag{10.5.1}$$

where  $\mathcal{J}_w$  is defined (10.2.9); i.e.,

$$\mathcal{J}_w f = \int_{-\infty}^{\infty} f(-t)dw(t).$$

Now, recalling that  $(\mathcal{J}_{\hat{w}})^{-1}\mathbf{X} = H(K)$  and that  $(\mathcal{J}_{\hat{w}})^{-1}\mathcal{J}_w = \mathfrak{F}$ , the  $(L^2)$  Fourier transform,  $\mathfrak{F}\mathcal{X} = H(K) \subset H_p^2$ ; see (10.2.10) and Theorem 3.5.6. Consequently,  $f$  vanishes on the negative real axis, so, as  $T^* = T^{-1} = \mathcal{J}_w|_{\mathcal{X}}$ , we actually have

$$T^* f = \int_{-\infty}^0 f(-t)dw(t). \tag{10.5.2}$$

We recall from Section 10.2 that

$$U_t(\mathbf{X}) = E^{\mathbf{X}} U_{t|\mathbf{X}} \tag{10.5.3}$$

defines a strongly continuous contraction semigroup  $\{U_t(\mathbf{X}); t \geq 0\}$  satisfying (7.5.3). The infinitesimal generator

$$\Gamma = \lim_{h \rightarrow 0^+} \frac{U_h(\mathbf{X}) - I}{h}$$

is a closed (unbounded) linear operator with a domain  $\mathcal{D}(\Gamma)$  that is dense in  $\mathbf{X}$ ; see, e.g., [126, 69]. Moreover, the adjoint  $\Gamma^*$  is the infinitesimal generator of the adjoint semigroup  $\{U_t(\mathbf{X})^*; t \geq 0\}$  [69, p. 251].

In analogy with the construction in Section 8.7, we define

$$A := T\Gamma^*T^*, \tag{10.5.4}$$

which is then the infinitesimal generator of a contraction semigroup which we denote  $\{e^{At}; t \geq 0\}$ . Likewise,  $A^*$  is the infinitesimal generator of the adjoint semigroup which we denote  $\{e^{A^*t}; t \geq 0\}$ . Consequently,

$$e^{At} = TU_t(\mathbf{X})^*T^* \quad \text{and} \quad e^{A^*t} := TU_t(\mathbf{X})T^*. \tag{10.5.5}$$

The domains  $\mathcal{D}(A)$  and  $\mathcal{D}(A^*)$  are both dense in  $\mathcal{X}$ .

Now, for an arbitrary  $\xi \in \mathbf{X}$ , let  $f \in \mathcal{X}$  be the corresponding point in the state space; i.e.,  $f = T\xi$ . Then, in view of (10.5.2),

$$U_t\xi = \int_{-\infty}^0 f(-\tau)dw(\tau + t) = \int_{-\infty}^t f(t - \tau)dw(\tau). \tag{10.5.6}$$

Then, since  $\mathbf{S} = \mathbf{H}^-(dw)$ , (7.5.5) yields

$$U_t(\mathbf{X})\xi = \int_{-\infty}^0 f(t - \tau)dw(\tau),$$

and consequently

$$(e^{A^*t}f)(\tau) = \begin{cases} f(t + \tau) & \text{for } \tau \geq 0, \\ 0 & \text{for } \tau < 0. \end{cases} \tag{10.5.7}$$

Therefore, whenever defined,  $A^*f$  is the derivative of  $f$  in the  $L^2$  sense.

Now, by a standard construction, define  $\mathcal{Z}$  to be the domain  $\mathcal{D}(A^*)$  of the unbounded operator  $A^*$  equipped with the graph topology

$$\langle f, g \rangle_{\mathcal{Z}} = \langle f, g \rangle_{\mathcal{X}} + \langle A^*f, A^*g \rangle_{\mathcal{X}}. \tag{10.5.8}$$

Since  $A^*$  is a closed operator with a dense domain,  $\mathcal{Z}$  is a Hilbert space that is densely embedded in  $\mathcal{X}$ . The topology of  $\mathcal{Z}$  is stronger than that of  $\mathcal{X}$ , and therefore all continuous linear functionals on  $\mathcal{X}$  are continuous on  $\mathcal{Z}$  as well. Consequently, we can think of the dual space  $\mathcal{X}^*$  as embedded in the dual space  $\mathcal{Z}^*$ . Then, identifying  $\mathcal{X}^*$  with  $\mathcal{X}$  we have

$$\mathcal{Z} \subset \mathcal{X} \subset \mathcal{Z}^*, \tag{10.5.9}$$

where  $\mathcal{Z}$  is dense in  $\mathcal{X}$ , which in turn is dense in  $\mathcal{Z}^*$ . We shall write  $(f, f^*)$  to denote the value of the linear functional  $f^* \in \mathcal{Z}^*$  evaluated at  $f \in \mathcal{Z}$  (or, by reflexivity,

the value at  $f^*$  of  $f$  regarded as a functional on  $\mathcal{Z}^*$ . Clearly, the bilinear form  $(f, f^*)$  coincides with the inner product  $\langle f, f^* \rangle_{\mathcal{X}}$  whenever  $f^* \in \mathcal{X}$ . Since  $A^*f$  is the derivative of  $f$ ,  $\mathcal{Z}$  is a subspace of the Sobolev space  $H^1(\mathbb{R}^+)$ , and  $\mathcal{Z}^*$  is a space of distributions (see, e.g., [5]).

In same way, we define  $\mathbf{Z}$  as the space obtained by equipping  $\mathcal{D}(\Gamma)$  with the graph topology

$$\langle \xi, \eta \rangle_{\mathbf{Z}} = \langle \xi, \eta \rangle_{\mathbf{X}} + \langle \Gamma\xi, \Gamma\eta \rangle_{\mathbf{X}}. \tag{10.5.10}$$

Then  $\mathbf{Z}$  is continuously embedded in the splitting subspace  $\mathbf{X}$ , and (10.5.9) corresponds to

$$\mathbf{Z} \subset \mathbf{X} \subset \mathbf{Z}^*. \tag{10.5.11}$$

Returning to the setting of (10.5.9), define  $Q : \mathcal{Z} \rightarrow \mathcal{X}$  to be the differentiation operator on  $\mathcal{Z}$ . Then  $Qf = A^*f$  for all  $f \in \mathcal{Z}$ , but, since  $\|Qf\|_{\mathcal{X}} \leq \|f\|_{\mathcal{Z}}$ ,  $Q$  is a bounded operator (in the  $\mathcal{Z}$ -topology). Its adjoint  $Q^* : \mathcal{X} \rightarrow \mathcal{Z}^*$  is the extension of  $A$  to  $\mathcal{X}$ , because  $(f, Q^*g) = \langle A^*f, g \rangle_{\mathcal{X}}$ . Since  $\{e^{A^*t}; t \geq 0\}$  is a strongly continuous contraction semigroup,  $Q$  is dissipative; i.e.,  $\langle Qf, f \rangle_{\mathcal{X}} \leq 0$  for all  $f \in \mathcal{Z}$ , and  $I - Q$  maps  $\mathcal{Z}$  onto  $\mathcal{X}$ ; i.e.,

$$(I - Q)\mathcal{Z} = \mathcal{X} \tag{10.5.12}$$

[126, p. 250]. Moreover, in view of the dissipative property,

$$\|(I - Q)f\|_{\mathcal{X}}^2 \geq \|f\|_{\mathcal{X}}^2 + \|Qf\|_{\mathcal{X}}^2, \tag{10.5.13}$$

and therefore  $I - Q$  is injective. Consequently,  $(I - Q)^{-1} : \mathcal{X} \rightarrow \mathcal{Z}$  is defined on all of  $\mathcal{X}$ , and, as can be seen from (10.5.13), it is a bounded operator. Likewise, the adjoint  $(I - Q^*)^{-1}$  is a bounded operator mapping  $\mathcal{Z}^*$  onto  $\mathcal{X}$ . Finally,

$$\|f\|_{\mathcal{Z}}^2 \leq \|(I - Q)f\|_{\mathcal{X}}^2 \leq 2\|f\|_{\mathcal{Z}}^2. \tag{10.5.14}$$

In fact, the first inequality is precisely (10.5.13), whereas the second follows from the inequality  $(a - b) \leq 2(a + b)^2$ .

For later reference, we state the following simple lemma.

**Lemma 10.5.1.** *A subset  $\mathcal{M}$  is dense in  $\mathcal{Z}$  if and only if  $(I - Q)\mathcal{M}$  is dense in  $\mathcal{X}$ .*

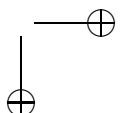
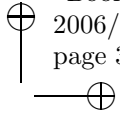
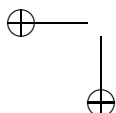
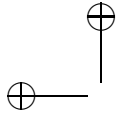
**Proof.** Suppose that  $(I - Q)\mathcal{M}$  is dense in  $\mathcal{X}$ . Then (10.5.12) and the first of the inequalities (10.5.14) imply that  $\mathcal{M}$  is dense in  $\mathcal{Z}$ . Conversely, if  $\mathcal{M}$  is dense in  $\mathcal{Z}$ , (10.5.12) and the second of the inequalities (10.5.14) imply that  $(I - Q)\mathcal{M}$  is dense in  $\mathcal{X}$ .  $\square$

Let  $f \in \mathcal{Z}$ . Since  $\mathcal{Z}$  is a subspace of the Sobolev space  $H^1(\mathbb{R}^+)$ , which is contained in the space of continuous functions (with a stronger topology) [5, p. 195], we can evaluate  $f$  at each point, and consequently, (10.5.7) yields

$$f(t) = (e^{A^*t}f)(0). \tag{10.5.15}$$

However,  $\mathcal{X}$  is our state space, and therefore we use (10.5.12) to reformulate (10.5.15) as

$$f(t) = [(I - Q)^{-1}e^{A^*t}(I - Q)f](0). \tag{10.5.16}$$



In fact, since  $A^*$  commutes with  $e^{A^*t}$ , then so does  $(I - Q)$ . Now, since  $(I - Q)^{-1}$  maps  $\mathcal{X}$  to  $\mathcal{Z}$ ,

$$B^*g = [(I - Q)^{-1}g](0) \tag{10.5.17}$$

defines a bounded map  $B^* : \mathcal{X} \rightarrow \mathbb{R}^p$ . Let  $B : \mathbb{R}^p \rightarrow \mathcal{X}$  be its adjoint. Then (10.5.16) may be written

$$f(t) = B^*e^{A^*t}(I - Q)f, \tag{10.5.18}$$

and therefore, if  $e_k$  is the  $k$ th unit axis vector in  $\mathbb{R}^p$ ,

$$f_k(t) = \langle B^*e^{A^*t}(I - Q)f, e_k \rangle_{\mathbb{R}^p} = \langle (I - Q)f, e^{At}Be_k \rangle_{\mathcal{X}}, \quad k = 1, 2, \dots, p;$$

that is,

$$f_k(t) = \langle g, e^{At}Be_k \rangle_{\mathcal{X}}, \quad k = 1, 2, \dots, p, \tag{10.5.19}$$

where  $g := (I - Q)f$ . This together with (10.5.6) yields, for each  $\xi \in \mathbf{Z}$ , the representation

$$U_t\xi = \sum_{k=1}^p \int_{-\infty}^t \langle g, e^{A(t-\tau)}Be_k \rangle_{\mathcal{X}} dw_k(\tau), \tag{10.5.20}$$

where  $g = (I - Q)T\xi$ .

Now define the  $\mathcal{X}$ -valued stochastic integral

$$x(t) = \int_{-\infty}^t e^{A(t-\tau)}Bdw(\tau) \tag{10.5.21}$$

for each  $t \in \mathbb{R}$  in the weak sense described in Section 8.7 via the recipe

$$\langle g, x(t) \rangle_{\mathcal{X}} := \sum_{k=1}^p \int_{-\infty}^t \langle g, e^{A(t-\tau)}Be_k \rangle_{\mathcal{X}} dw_k(\tau). \tag{10.5.22}$$

In fact, since (10.5.19) is square-integrable by construction, the right member of (10.5.19) is well-defined. Then, for any  $\xi \in \mathbf{Z}$ , (10.5.20) yields

$$U_t\xi = \langle g, x(t) \rangle_{\mathcal{X}}, \tag{10.5.23}$$

where  $g := (I - Q)T\xi$  is the corresponding function in  $\mathcal{X}$ . Consequently, for each  $\xi \in \mathbf{Z}$  there is a  $g \in \mathcal{X}$  such that  $\xi = \langle g, x(0) \rangle_{\mathcal{X}}$ ; i.e.,  $x(0)$  is a (weak) *exact generator* of  $\mathbf{Z}$ . Moreover since  $\mathbf{Z}$  is dense in  $\mathbf{X}$ ,

$$\text{cl} \{ \langle g, x(0) \rangle_{\mathcal{X}} \mid g \in \mathcal{X} \} = \mathbf{X}; \tag{10.5.24}$$

i.e.,  $x(0)$  is a (weak) *generator* of  $\mathbf{X}$  that is not exact. Moreover, if  $f_i := (I - Q)^{-1}g_i$ ,  $i = 1, 2$ , a straight-forward calculation yields

$$\mathbb{E} \{ \langle g_1, x(0) \rangle_{\mathcal{X}} \langle g_2, x(0) \rangle_{\mathcal{X}} \} = \langle f_1, f_2 \rangle_{\mathcal{X}} = \langle g_1, Pg_2 \rangle_{\mathcal{X}}, \tag{10.5.25}$$

where  $P : \mathcal{X} \rightarrow \mathcal{X}$  is the state covariance operator

$$P = (I - A)^{-1}(I - A^*)^{-1}. \tag{10.5.26}$$

**Theorem 10.5.2.** *Let  $(\mathbb{H}, \{U_t\}, \mathbf{X})$  be a proper Markovian representation with the forward generating process  $w$ , and let  $\mathcal{X} := (\mathcal{J}_w)^{-1}\mathbf{X}$ . Let the operators  $A$  and  $B$  be defined by (10.5.4) and (10.5.17), respectively, and let  $\{x(t)\}_{t \in \mathbb{R}}$  be the  $\mathcal{X}$ -valued stochastic process defined in the weak sense via (10.5.22). Then, if  $\Gamma$  is the infinitesimal generator of  $U_t(\mathbf{X})$ ,*

$$\{\langle g, x(0) \rangle_{\mathcal{X}} \mid g \in \mathcal{X}\} = \mathcal{D}(\Gamma) \subset \mathbf{X}, \quad (10.5.27)$$

where  $\mathcal{D}(\Gamma)$  is dense in  $\mathbf{X}$ . In fact, for any  $\xi \in \mathcal{D}(\Gamma)$ , there is a  $g \in \mathcal{X}$  such that

$$U_t \xi = \langle g, x(t) \rangle_{\mathcal{X}}, \quad (10.5.28)$$

and it is given by  $g := (I - A^*)T\xi$ . The pair  $(A, B)$  is reachable in the sense that  $\bigcap_{t \geq 0} \ker B^* e^{A^* t} = 0$ . Finally the covariance operator (10.5.26) satisfies the Lyapunov equation

$$AP + PA^* + BB^* = 0. \quad (10.5.29)$$

**Proof.** It only remains to prove that  $(A, B)$  is reachable and that (10.5.29) holds. In view of (10.5.18),  $g \in \bigcap_{t \geq 0} \ker B^* e^{A^* t}$  holds if and only if  $f = 0$ ; i.e.,  $g = 0$ , which proves reachability. To prove that (10.5.26) satisfies the Lyapunov equation (10.5.29), recall  $A^* f = Qf$  for all  $f \in \mathcal{D}(A^*)$ , where  $Q$  is the differentiation operator. Let  $g_1, g_2 \in \mathcal{X}$ . Then  $f_i = (I - A^*)^{-1}g_i \in \mathcal{Z}$ ,  $i = 1, 2$ , and

$$\langle A^* f_1, f_2 \rangle_{\mathcal{X}} + \langle f_1, A^* f_2 \rangle_{\mathcal{X}} = \int_0^\infty (\dot{f}_1 f_2' + f_1 \dot{f}_2') dt = -f_1(0) f_2(0)'. \quad (10.5.30)$$

Also, in view of (10.5.18),

$$\langle g_1, BB^* g_2 \rangle_{\mathcal{X}} = \langle B^*(I - Q)f_1, B^*(I - Q)f_2 \rangle_{\mathbb{R}^p} = f_1(0) f_2(0)'. \quad (10.5.31)$$

Now adding (10.5.30) and (10.5.31) we obtain

$$\langle g_1, (AP + PA^* + BB^*)g_2 \rangle_{\mathcal{X}} = 0,$$

where  $P := (I - A)^{-1}(I - A^*)^{-1}$ , as claimed.  $\square$

### Backward state representation

To develop a backward state representation, define the operator  $\bar{T}$  to be restriction of  $(\mathcal{J}_w)^{-1}$  to  $\mathbf{X}$ , where  $\mathcal{J}_w$  is given by (10.2.11); i.e.,

$$\bar{T}^* \bar{f} = \int_{-\infty}^\infty \bar{f}(-t) dw(t).$$

Then, setting  $\bar{\mathcal{X}} := \bar{T}\mathbf{X}$ ,  $\bar{T} : \mathbf{X} \rightarrow \bar{\mathcal{X}}$  is an isomorphism such that  $\langle \bar{T}\xi, \bar{T}\eta \rangle_{\bar{\mathcal{X}}} = \langle \xi, \eta \rangle_{\mathbf{X}}$ . Since  $\mathcal{J}_{\bar{w}}$ , defined by (10.2.12), equals  $\mathcal{J}_w \mathfrak{F}^{-1}$ , where  $\mathfrak{F}$  is the  $(L^2)$  Fourier

transform,  $\mathfrak{F}\bar{\mathcal{X}} = \bar{H}(K^*) \subset \bar{H}_p^2$ ,  $\bar{f}$  vanishes on the negative real axis, so we actually have

$$\bar{T}^* \bar{f} = \int_0^\infty \bar{f}(-t) d\bar{w}(t). \tag{10.5.32}$$

Moreover,

$$\bar{A} := \bar{T}\Gamma\bar{T}^* \tag{10.5.33}$$

and its adjoint  $\bar{A}^*$  are the infinitesimal generators of two adjoint semigroups which we denote

$$e^{\bar{A}t} := \bar{T}U_t(\mathbf{X})\bar{T}^* \quad \text{and} \quad e^{\bar{A}^*t} := \bar{T}U_t(\mathbf{X})^*\bar{T}^*, \tag{10.5.34}$$

and the domains  $\mathcal{D}(\bar{A})$  and  $\mathcal{D}(\bar{A}^*)$  are clearly dense in  $\bar{\mathcal{X}}$ .

For an arbitrary  $\xi \in \mathbf{X}$ ,

$$U_t \xi = \int_t^\infty \bar{f}(t - \tau) d\bar{w}(\tau), \tag{10.5.35}$$

where  $\bar{f} = \bar{T}\xi$ , and therefore

$$U_t(\mathbf{X})^* \xi = \mathbf{E}^{\mathbf{S}} \int_{-t}^\infty \bar{f}(-t - \tau) d\bar{w}(\tau) = \int_0^\infty \bar{f}(-t - \tau) d\bar{w}(\tau).$$

Consequently,

$$(e^{\bar{A}^*t} \bar{f})(\tau) = \begin{cases} \bar{f}(-t + \tau) & \text{for } \tau \geq 0, \\ 0 & \text{for } \tau < 0, \end{cases} \tag{10.5.36}$$

so  $\bar{A}^*$  is a differentiation operator. Hence, defining  $\bar{\mathcal{Z}}$  to be  $\mathcal{D}(\bar{A}^*)$  equipped with the corresponding graph topology and  $\bar{\mathbf{Z}} := \bar{T}^*\bar{\mathcal{Z}}$  to be the corresponding space of stochastic variables, we may define an operator  $\bar{Q} : \bar{\mathcal{Z}} \rightarrow \bar{\mathcal{X}}$  such that  $\bar{Q}f = \bar{A}^*f$  for all  $f \in \bar{\mathcal{Z}}$ . Moreover, for all  $f \in \bar{\mathcal{Z}}$ , (10.5.36) yields

$$\bar{f}(-t) = (e^{\bar{A}^*t} \bar{f})(0) = [(I - \bar{Q})^{-1} e^{\bar{A}^*t} (I - \bar{Q}) \bar{f}](0), \quad t \geq 0. \tag{10.5.37}$$

Now, let  $\bar{B} : \mathbb{R}^p \rightarrow \bar{\mathcal{X}}$  be defined via its adjoint

$$\bar{B}^* \bar{g} = [(I - \bar{Q})^{-1} \bar{g}](0). \tag{10.5.38}$$

Then, in view of (10.5.37),

$$\bar{f}(-t) = \bar{B}^* e^{\bar{A}^*t} (I - \bar{Q}) \bar{f}.$$

Therefore, analogously to the forward setting,

$$\bar{f}_k(-t) = \langle g, e^{\bar{A}^*t} \bar{B} e_k \rangle_{\bar{\mathcal{X}}}, \quad k = 1, 2, \dots, p, \tag{10.5.39}$$

where  $\bar{g} := (I - \bar{Q}) \bar{f}$ . Consequently, in view of (10.5.35), for each  $\xi \in \bar{\mathbf{Z}}$ ,

$$U_t \xi = \sum_{k=1}^p \int_t^\infty \langle \bar{g}, e^{\bar{A}^*t} \bar{B} e_k \rangle_{\bar{\mathcal{X}}} d\bar{w}_k(\tau) \tag{10.5.40}$$

with  $\bar{g} := (I - \bar{Q})\bar{T}\xi \in \bar{\mathcal{X}}$ , which may also be expressed in the form

$$U_t\xi = \langle \bar{g}, \bar{x}(t) \rangle_{\bar{\mathcal{X}}} \quad (10.5.41)$$

via

$$\langle \bar{g}, \bar{x}(t) \rangle_{\bar{\mathcal{X}}} := \sum_{k=1}^p \int_t^\infty \langle \bar{g}, e^{\bar{A}t} \bar{B} e_k \rangle_{\bar{\mathcal{X}}} d\bar{w}_k(\tau). \quad (10.5.42)$$

Now the weakly defined  $\bar{\mathcal{X}}$ -valued stochastic process  $\{\bar{x}(t)\}_{t \in \mathbb{R}}$  has the representation

$$\bar{x}(t) = \int_t^\infty e^{\bar{A}(\tau-t)} \bar{B} d\bar{w}(\tau). \quad (10.5.43)$$

Analogously to the forward setting,  $\bar{x}(0)$  is a (weak) exact generator of  $\bar{\mathbf{Z}}$  and a (weak) generator (that is not exact) of  $\mathbf{X}$  in the sense that

$$\text{cl} \{ \langle \bar{g}, \bar{x}(0) \rangle_{\bar{\mathcal{X}}} \mid \bar{g} \in \bar{\mathcal{X}} \} = \mathbf{X}. \quad (10.5.44)$$

Moreover, if  $\bar{f}_i := (I - \bar{Q})^{-1} \bar{g}_i$ ,  $i = 1, 2$ , a straight-forward calculation yields

$$\mathbb{E} \{ \langle \bar{g}_1, x(0) \rangle_{\bar{\mathcal{X}}} \langle \bar{g}_2, x(0) \rangle_{\bar{\mathcal{X}}} \} = \langle \bar{f}_1, \bar{f}_2 \rangle_{\bar{\mathcal{X}}} = \langle \bar{g}_1, \bar{P} \bar{g}_2 \rangle_{\bar{\mathcal{X}}}, \quad (10.5.45)$$

where  $\bar{P} : \bar{\mathcal{X}} \rightarrow \bar{\mathcal{X}}$  is the state covariance operator

$$\bar{P} = (I - \bar{A}^*)^{-1} (I - \bar{A})^{-1}. \quad (10.5.46)$$

We can now establish the backward version of Theorem 10.5.2.

**Theorem 10.5.3.** *Let  $(\mathbb{H}, \{U_t\}, \mathbf{X})$  be a proper Markovian representation with the backward generating process  $\bar{w}$ , and let  $\bar{\mathcal{X}} := (\mathcal{J}_{\bar{w}})^{-1} \mathbf{X}$ . Let the operators  $\bar{A}$  and  $\bar{B}$  be defined by (10.5.33) and (10.5.38), respectively, and let  $\{\bar{x}(t)\}_{t \in \mathbb{R}}$  be the  $\bar{\mathcal{X}}$ -valued stochastic process defined in the weak sense via (10.5.42). Then, if  $\Gamma$  is the infinitesimal generator of  $U_t(\mathbf{X})$ ,*

$$\{ \langle \bar{g}, \bar{x}(0) \rangle_{\bar{\mathcal{X}}} \mid \bar{g} \in \bar{\mathcal{X}} \} = \mathcal{D}(\Gamma^*) \subset \mathbf{X}, \quad (10.5.47)$$

where  $\mathcal{D}(\Gamma^*)$  is dense in  $\mathbf{X}$ . In fact, for any  $\xi \in \mathcal{D}(\Gamma^*)$ , there is a  $\bar{g} \in \bar{\mathcal{X}}$  such that

$$U_t\xi = \langle \bar{g}, \bar{x}(t) \rangle_{\bar{\mathcal{X}}}, \quad (10.5.48)$$

and it is given by  $\bar{g} := (I - \bar{A}^*)\bar{T}\xi$ . The pair  $(\bar{A}, \bar{B})$  is reachable in the sense that  $\bigcap_0^\infty \ker \bar{B}^* e^{\bar{A}^*t} = 0$ . Finally the covariance operator (10.5.46) satisfies the Lyapunov equation

$$\bar{A}\bar{P} + \bar{P}\bar{A}^* + \bar{B}\bar{B}^* = 0. \quad (10.5.49)$$

**Remark 10.5.4.** To establish a connection between the forward and the backward setting, define the isomorphism  $R := T\bar{T}^* : \bar{\mathcal{X}} \rightarrow \mathcal{X}$ . Then, if  $\xi = T^*f = \bar{T}^*\bar{f}$ ,

$$f = R\bar{f}. \quad (10.5.50)$$

Moreover,

$$R e^{\bar{A}t} R^* = e^{A^*t}, \quad A^*R = R\bar{A}. \quad (10.5.51)$$



**Stochastic realizations of a stationary process**

Given an infinite-dimensional, proper Markovian representation  $(\mathbb{H}, \{U_t\}, \mathbf{X})$  of an  $m$ -dimensional, stationary stochastic process  $\{y(t)\}_{t \in \mathbb{R}}$ , is it possible to construct a stochastic realization based on the forward state representation (10.5.21)? From Theorem 10.5.2 it is clear that a necessary and sufficient condition for this is that

$$y_k(0) \in \mathcal{D}(\Gamma), \quad k = 1, 2, \dots, m, \tag{10.5.52}$$

where  $\Gamma$  is the infinitesimal generator of the semigroup  $\{U_t(\mathbf{X})\}$ . In the same way, Theorem 10.5.3 implies that there is a stochastic realization based on the backward state representation (10.5.43) if and only if

$$y_k(0) \in \mathcal{D}(\Gamma^*), \quad k = 1, 2, \dots, m. \tag{10.5.53}$$

**Proposition 10.5.5.** *Let  $(\mathbb{H}, \{U_t\}, \mathbf{X})$  be a proper Markovian representation with analytic spectral factor  $W$  and coanalytic spectral factor  $\bar{W}$ , and let  $\Gamma$  is the infinitesimal generator of the semigroup  $\{U_t(\mathbf{X})\}$ . Then the condition (10.5.52) holds if and only if the rows of  $i\omega W(i\omega) - N$  belong to  $H_p^2$  for some constant  $m \times p$  matrix  $N$ . Similarly, the condition (10.5.53) holds if and only if the rows of  $i\omega \bar{W}(i\omega) - \bar{N}$  belong to  $\bar{H}_p^2$  for some constant  $m \times p$  matrix  $\bar{N}$ .*

**Proof.** Clearly condition (10.5.52) is equivalent to  $\mathfrak{F}^* a' W \in \mathcal{D}(A^*)$  for all  $a \in \mathbb{R}^m$ , which in turn is equivalent to the condition that  $i\omega W(i\omega) - N$  belong to  $H_p^2$  for some constant  $m \times p$  matrix  $N$  [69, Lemma 3.1]. The second statement follows by symmetry.  $\square$

Now, suppose condition (10.5.52) holds. Then it follows from (10.5.22) and (10.5.23) that

$$y(t) = \int_{-\infty}^t C e^{A(t-\tau)} B dw(\tau), \tag{10.5.54}$$

where the bounded operator  $C : \mathcal{X} \rightarrow \mathbb{R}^m$  is given by

$$a' C g = \langle (I - Q) \mathfrak{F}^* a' W, g \rangle_{\mathcal{X}}, \quad \text{for all } a \in \mathbb{R}^m. \tag{10.5.55}$$

Likewise, if condition (10.5.53) holds, it follows from (10.5.40) that

$$y(t) = \int_t^{\infty} \bar{C} e^{\bar{A}(\tau-t)} \bar{B} d\bar{w}(\tau), \tag{10.5.56}$$

where the bounded operator  $\bar{C} : \bar{\mathcal{X}} \rightarrow \mathbb{R}^m$  is given by

$$a' \bar{C} \bar{g} = \langle (I - \bar{Q}) \mathfrak{F}^* a' \bar{W}, \bar{g} \rangle_{\bar{\mathcal{X}}}, \quad \text{for all } a \in \mathbb{R}^m. \tag{10.5.57}$$

**Theorem 10.5.6.** *Let  $(\mathbb{H}, \{U_t\}, \mathbf{X})$  be a proper Markovian representation of an  $m$ -dimensional, stationary stochastic process with spectral factors  $(W, \bar{W})$  and generating processes  $(dw, d\bar{w})$ . If condition (10.5.52) holds,*

$$y(t) = Cx(t), \tag{10.5.58}$$

where  $\{x(t)\}_{t \in \mathbb{R}}$  is the weak  $\mathcal{X}$ -valued stochastic process (10.5.21). The pair  $(C, A)$  is (completely) observable in the sense that  $\bigcap_{t \geq 0} \ker C e^{At} = 0$  if and only if  $\mathbf{X}$  is observable. Moreover,

$$\Lambda(t) := E\{y(t)y(0)'\} = C e^{At} P C^*, \quad (10.5.59)$$

where  $P$  is given by (10.5.26) and satisfies the operator Lyapunov equation (10.5.29). Likewise, if condition (10.5.53) holds,

$$y(t) = \bar{C} \bar{x}(t), \quad (10.5.60)$$

where  $\{\bar{x}(t)\}_{t \in \mathbb{R}}$  is the weak  $\bar{\mathcal{X}}$ -valued stochastic process (10.5.21). The pair  $(\bar{C}, \bar{A})$  is (completely) observable in the sense that  $\bigcap_{t \geq 0} \ker \bar{C} e^{\bar{A}t} = 0$  if and only if  $\mathbf{X}$  is constructible, and

$$\Lambda(t) := E\{y(t)y(0)'\} = \bar{C} e^{\bar{A}t} P \bar{C}^*, \quad (10.5.61)$$

where  $\bar{P}$  is given by (10.5.46) and satisfies the operator Lyapunov equation (10.5.49).

Consequently, for  $\{y(t)\}_{t \in \mathbb{R}}$  to have both a forward and a backward representation with respect to the splitting subspace  $\mathbf{X}$ , we must have

$$y_k(0) \in \mathcal{D}(\Gamma) \cap \mathcal{D}(\Gamma^*), \quad k = 1, 2, \dots, m.$$

The representations (10.5.58) and (10.5.60) follows from what has already been said above. To prove the statements about observability and constructibility we need to introduce a few concepts. Define  $\mathbf{M}$  to be the vector space

$$\mathbf{M} = \text{span}\{E^{\mathbf{X}} y_k(t); t \geq 0, k = 1, 2, \dots, m\}. \quad (10.5.62)$$

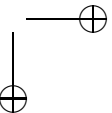
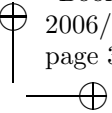
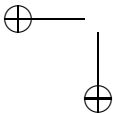
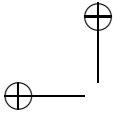
Since  $E^{\mathbf{X}} y_k(t) = U_t(\mathbf{X}) y_k(0)$ ,  $\mathbf{M}$  is invariant under the action of  $U_t(\mathbf{X})$ ; i.e.,  $U_t(\mathbf{X})\mathbf{M} \subset \mathbf{M}$  for all  $t \geq 0$ . Moreover,  $\mathcal{D}(\Gamma)$  is invariant under  $U_t(\mathbf{X})$ ; this is a well-known property of of a semigroup. Hence it follows from (10.5.52) that  $\mathbf{M} \subset \mathbf{Z}$ . Now, if  $\mathbf{X}$  is observable,  $\mathbf{M}$  is dense in  $\mathbf{X}$  (Corollary 7.4.12), but this does not automatically imply that  $\mathbf{M}$  is dense in  $\mathbf{Z}$  (in the graph topology). In the present case, however, this is true, as can be seen from the following lemma, noting that

$$\mathcal{M} := T\mathbf{M} = \text{span}\{e^{A^*t} \mathfrak{F}^* W_k; t \geq 0, k = 1, 2, \dots, m\} \subset \mathcal{M} \quad (10.5.63)$$

and that  $e^{A^*t} = T U_t(\mathbf{X}) T^*$ . In the terminology of [5, p. 101], this means that the Hilbert space  $\mathbf{Z}$ , containing the vector space  $\mathbf{M}$  and continuously embedded in the Hilbert space  $\mathbf{X}$ , is *normal*.

**Lemma 10.5.7.** *Suppose that  $\mathcal{M}$  is a subset of  $\mathcal{Z}$  that is invariant in the sense that  $e^{A^*t} \mathcal{M} \subset \mathcal{M}$  for all  $t \geq 0$ , and suppose that  $\mathcal{M}$  is dense in  $\mathcal{X}$ . Then  $\mathcal{M}$  is dense in  $\mathcal{Z}$  (in the graph topology).*

**Proof.** Suppose that  $\mathcal{M} \subset \mathcal{Z} \subset \mathcal{X}$  and that  $\mathcal{M}$  is dense in  $\mathcal{X}$ . Let  $\overline{\mathcal{M}}$  be the closure of  $\mathcal{M}$  in the graph topology. We know that  $\overline{\mathcal{M}} \subset \mathcal{Z}$ , and we want to show that  $\overline{\mathcal{M}} = \mathcal{Z}$ .



To this end, define  $\overline{Q}$  to be the restriction of  $Q$  to  $\overline{\mathcal{M}}$ . Then  $\overline{Q}$  is an unbounded operator defined on a dense subset of  $\mathcal{X}$ , and, like  $Q$ , it is closed and dissipative. Hence the range of  $(I - \overline{Q})$  is closed [33, Theorem 3.4, p. 79]. Therefore, if we can show that the range of  $(I - \overline{Q})$  is dense in  $\mathcal{X}$ , we know that it is all of  $\mathcal{X}$ . This would mean that  $\overline{Q}$  is maximal dissipative [33, Theorem 3.6, p. 81]. However,  $Q$  is a dissipative extension of  $\overline{Q}$ , and hence  $\overline{Q} = Q$ . Then  $\mathcal{D}(\overline{Q}) = \mathcal{D}(Q)$ ; i.e.,  $\overline{\mathcal{M}} = \mathcal{Z}$ , as required.

Consequently, it remains to prove that  $(I - \overline{Q})\overline{\mathcal{M}}$  is dense in  $\mathcal{X}$ . Since  $\overline{\mathcal{M}}$  is dense in  $\mathcal{X}$ , we only need to show that the equation  $(I - \overline{Q})f = g$ ; i.e.,  $\dot{f} - f = -g$ , has a solution  $f \in \overline{\mathcal{M}}$  for each  $g \in \overline{\mathcal{M}}$ . However, for such a  $g$ , the differential equation  $\dot{f} - f = -g$  has the  $L^2$  solution

$$f(t) = \int_0^\infty e^{-\tau} g(t + \tau) d\tau = \int_0^\infty (e^{A^* \tau} g)(t) dm(\tau), \tag{10.5.64}$$

where  $dm = e^{-\tau} d\tau$ , so it remains to show that this  $f$  belongs to  $\overline{\mathcal{M}}$ . Since  $e^{A^* \tau} \mathcal{M} \subset \mathcal{M}$ , by continuity,  $e^{A^* \tau} g \in \overline{\mathcal{M}}$  for each  $\tau \geq 0$ . The function  $\tau \rightarrow e^{A^* \tau} g$  is therefore mapping  $\mathbb{R}^+$  into  $\overline{\mathcal{M}}$ . It is clearly strongly measurable, and, since  $e^{A^* \tau}$  is a contraction,  $\|e^{A^* \tau} g\|_{\overline{\mathcal{M}}} \leq \|g\|_{\overline{\mathcal{M}}}$ . Hence

$$\int_0^\infty \|e^{A^* \tau} g\|_{\overline{\mathcal{M}}}^2 dm(\tau) < \infty,$$

and consequently (10.5.64) is a Bochner integral [126, p. 133]. Hence, by definition,  $f \in \overline{\mathcal{M}}$  as required.  $\square$

To conclude the proof of Theorem 10.5.6, we first note that, since  $e^{A^* t}$  and  $(I - Q)$  commute,

$$a' C e^{A t} g = \langle (I - Q) e^{A^* t} \mathfrak{F}^*(a' W), g \rangle_{\mathcal{X}}$$

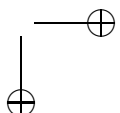
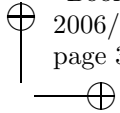
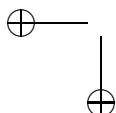
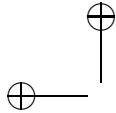
for all  $a \in \mathbb{R}^m$ . Hence, in view of (10.5.63),  $g \in \cap_{t \geq 0} \ker C e^{A t}$  if and only if

$$\langle h, g \rangle = 0, \quad \text{for all } h \in (I - Q)\mathcal{M}. \tag{10.5.65}$$

Now, if  $(C, A)$  is observable; i.e.,  $\cap_{t \geq 0} \ker C e^{A t} = 0$ , only  $g = 0$  satisfies (10.5.65). Hence  $(I - Q)\mathcal{M}$  is dense in  $\mathcal{X}$ . Therefore,  $\mathcal{M}$  is dense in  $\mathcal{Z}$  (Lemma 10.5.1); i.e.,  $\mathbf{X}$  is observable. Conversely, assume that  $\mathbf{X}$  is observable. Then  $\mathcal{M}$  is dense in  $\mathcal{Z}$  (Lemma 10.5.7), and consequently  $(I - Q)\mathcal{M}$  is dense in  $\mathcal{X}$  (Lemma 10.5.1). However, then only  $g = 0$  can satisfy (10.5.65), and therefore  $(C, A)$  is observable. This concludes the observability part of Theorem 10.5.6. The constructibility part follows from a symmetric argument.

### Stochastic realizations of a stationary-increment process

Suppose that  $(\mathbb{H}, \{U_t\}, \mathbf{X})$  is a proper Markovian representation of an  $m$ -dimensional, stationary increment process  $dy$ , and that  $\mathbf{X} \sim (\mathbf{S}, \bar{\mathbf{S}})$  has spectral factors  $(W, \bar{W})$  and generating processes  $(dw, d\bar{w})$ .



In Section 10.3 we were able to construct a forward stochastic realization (10.3.24) for the finite-dimensional case by first observing that  $W$  can be decomposed as

$$W(s) = G(s) + D, \tag{10.5.66}$$

where the rows of  $G$  belong to  $H_p^2$  and  $D$  is an  $m \times p$  matrix. A necessary and sufficient condition for this to hold in the infinite-dimensional case is that  $dy$  is conditionally Lipschitz with respect to  $\mathbf{S}$  (Corollary 5.5.2).

In this case, as explained in Chapter 5 in more detail,

$$y(h) - y(0) = \int_{-\infty}^{\infty} \frac{e^{i\omega h} - 1}{i\omega} W(i\omega) d\hat{w} = \int_0^h z(t) dt + D[w(h) - w(0)],$$

where  $\{z(t)\}_{t \in \mathbb{R}}$  is the stationary process

$$z(t) = \int_{-\infty}^{\infty} e^{i\omega t} G(i\omega) d\hat{w}.$$

By Theorem 10.2.3,

$$a'z(0) \in E^{\mathbf{S}} \mathbf{H}^- \subset E^{\mathbf{S}} \bar{\mathbf{S}} = \mathbf{S} \cap \bar{\mathbf{S}} = \mathbf{X}$$

for all  $a \in \mathbb{R}^m$ . Consequently,

$$dy = zdt + dw, \tag{10.5.67}$$

where the components of  $z(0)$  belong to  $\mathbf{X}$ . Moreover,  $z_k(0)$ ,  $k = 1, 2, \dots, m$  are the conditional derivatives

$$z_k(t) = \lim_{h \downarrow 0} \frac{1}{h} E^{U_t \mathbf{S}} [y_k(t+h) - y_k(t)], \quad k = 1, 2, \dots, m,$$

with respect to  $\mathbf{S}$ .

Similarly,  $dy$  is conditionally Lipschitz with respect to  $\bar{\mathbf{S}}$  if and only if

$$\bar{W}(s) = \bar{G}(s) + \bar{D}, \tag{10.5.68}$$

where the rows of  $\bar{G}$  belong to  $H_p^2$  and  $\bar{D}$  is an  $m \times p$  matrix, in which case there is a stationary process

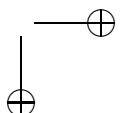
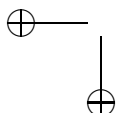
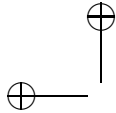
$$\bar{z}(t) = \int_{-\infty}^{\infty} e^{i\omega t} \bar{G}(i\omega) d\hat{w},$$

so that

$$dy = \bar{z}dt + Dd\bar{w}, \tag{10.5.69}$$

where  $\bar{z}_k(0)$ ,  $k = 1, 2, \dots, m$  belong to  $\mathbf{X}$  and are the conditional derivatives of  $dy$  with respect to  $\bar{\mathbf{S}}$ .

**Theorem 10.5.8.** *Let  $(\mathbb{H}, \{U_t\}, \mathbf{X})$  be a proper Markovian representation of an  $m$ -dimensional, stationary increment process  $dy$  with spectral factors  $(W, \bar{W})$  and*



generating processes  $(dw, d\bar{w})$ , let  $\mathbf{X} \sim (\mathbf{S}, \bar{\mathbf{S}})$ , and let  $\Gamma$  be the infinitesimal generator of the semigroup  $\{U_t(\mathbf{X})\}$ .

Then, if  $dy$  is conditionally Lipschitz continuous with respect to  $\mathbf{S}$ , and if the conditional derivatives  $z_k(0)$ ,  $k = 1, 2, \dots, m$ , belong to  $\mathcal{D}(\Gamma)$ , then

$$dy = Cx(t)dt + Ddw, \tag{10.5.70}$$

where  $\{x(t)\}_{t \in \mathbb{R}}$  is the weak  $\mathcal{X}$ -valued stochastic process (10.5.21),  $D := W(\infty)$ , and the bounded operator  $C : \mathcal{X} \rightarrow \mathbb{R}^m$  is given by

$$a' Cg = \langle (I - Q)\mathfrak{F}^* a'(W - D), g \rangle_{\mathcal{X}}, \quad \text{for all } a \in \mathbb{R}^m. \tag{10.5.71}$$

The pair  $(C, A)$  is (completely) observable in the sense that  $\cap_{t \geq 0} \ker Ce^{At} = 0$  if and only if  $\mathbf{X}$  is observable.

Similarly, if  $dy$  is conditionally Lipschitz continuous with respect to  $\bar{\mathbf{S}}$ , and if the conditional derivatives  $\bar{z}_k(0)$ ,  $k = 1, 2, \dots, m$ , belong to  $\mathcal{D}(\Gamma^*)$ , then

$$dy = \bar{C}\bar{x}(t)dt + \bar{D}d\bar{w}, \tag{10.5.72}$$

where  $\{\bar{x}(t)\}_{t \in \mathbb{R}}$  is the weak  $\bar{\mathcal{X}}$ -valued stochastic process (10.5.21),  $\bar{D} := \bar{W}(\infty)$ , and the bounded operator  $\bar{C} : \bar{\mathcal{X}} \rightarrow \mathbb{R}^m$  is given by

$$a' \bar{C}g = \langle (I - \bar{Q})\bar{\mathfrak{F}}^* a'(\bar{W} - \bar{D}), \bar{g} \rangle_{\bar{\mathcal{X}}}, \quad \text{for all } a \in \mathbb{R}^m. \tag{10.5.73}$$

The pair  $(\bar{C}, A^*)$  is (completely) observable in the sense that  $\cap_{t \geq 0} \ker \bar{C}e^{A^*t} = 0$  if and only if  $\mathbf{X}$  is observable.

**Proof.** If  $z_1(0), z_2(0), \dots, z_m(0) \in \mathcal{D}(\Gamma)$ , it follows from (10.5.22) and (10.5.23) that

$$z(t) = \int_{-\infty}^t Ce^{A(t-\tau)} Bdw(\tau), \tag{10.5.74}$$

where the bounded operator  $C : \mathcal{X} \rightarrow \mathbb{R}^m$  is given by (10.5.71), and therefore  $z(t) = Cx(t)$ , which together with (10.5.67) yields (10.5.70), as claimed. The backward representation (10.5.72) follows by a symmetric argument.

The proof of the statement of observability follows the same lines as that in Theorem 10.5.6, now taking  $\mathbf{M}$  to be

$$\mathbf{M} := \text{span}\{E^{\mathbf{X}}[y_k(t+h) - y_k(t)]; t \geq 0, k = 1, 2, \dots, m\}$$

and observing that

$$\mathbf{M} = \text{span}\{E^{\mathbf{X}} z_k(t); t \geq 0, k = 1, 2, \dots, m\}.$$

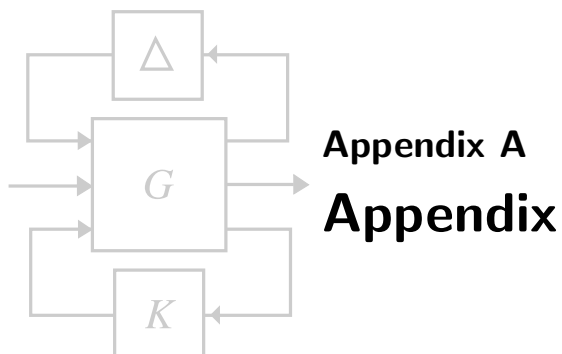
The proof of the constructibility statement follows by symmetry.  $\square$

## 10.6 Bibliographical notes

Basic references for the geometric theory this chapter are [78, 85, 84, 86]. Sections 10.1 and 10.2 follow [86] closely, and Section 10.3 follows [85, 86]. Theorem 10.4.1 in Section 10.4 is from [86].

The classical theory of stochastic realization, primarily dealing with spectral factorization, the linear matrix inequality and the algebraic Riccati equation was initiated by Kalman [56] and mainly developed by by Anderson [4] and Faurre [28]. Theorem 10.4.4 is due to Anderson [4] and solves the so-called “inverse problem of stationary covariance generation” stated by Anderson in [4]. The theory around the Positive Real Lemma is based on classical result due to Yakubovich [125], Kalman [55] and Popov [104]. The geometry of the set  $\mathcal{P}$  has been studied by Faurre et. al. [28] and the geometric part of Theorem 10.4.6 is due to Faurre. The theory of the Linear Matrix Inequality, the Algebraic Riccati Equation and the Algebraic Riccati Inequality was developed to quite some completeness by Willems [122].

Section 10.5 is based on [83, 85] and, as for representations of stationary-increment processes, [84]. The construction of shift realizations is analogous to that in the infinite-dimensional deterministic realization theory of [6, 7, 33, 34, 48] except that the framework is transposed to accommodate the appropriate relation between observability (constructibility) and forward (backward) realizations. The proof of Lemma 10.5.7, which appeared in [83], was originally suggested to us by A. Gombani.



### A.1 Hilbert spaces

The scope of this section is just to set notations and to recall the basic facts.

Recall that an *inner product* or (*scalar product*) on a vector space  $\mathbf{V}$  is a function

$$\langle \cdot, \cdot \rangle : \mathbf{V} \times \mathbf{V} \rightarrow \mathbb{C}$$

which is

1. linear in the first argument

$$\langle \alpha x + \beta y, z \rangle = \alpha \langle x, z \rangle + \beta \langle y, z \rangle \quad x, y, z \in \mathbf{V}$$

2. antisymmetric

$$\langle y, x \rangle = \overline{\langle x, y \rangle}$$

where the overline denotes complex conjugate,

3. positive, in the sense that

$$\|x\|^2 := \langle x, x \rangle > 0 \quad \text{for all } x \neq 0.$$

The quantity  $\|x\|$  is called the *norm* induced by the inner product  $\langle \cdot, \cdot \rangle$ . Every inner product satisfies the *Schwartz inequality*

$$|\langle x, y \rangle| \leq \|x\| \|y\|$$

It is easy to check that  $\|\cdot\|$  satisfies also the axioms of a norm and in particular the triangle inequality

$$\|x + y\| \leq \|x\| + \|y\| \quad x, y \in \mathbf{V}.$$

A *Hilbert space* is an inner product space  $(\mathbf{H}, \langle \cdot, \cdot \rangle)$  which is complete with respect to the metric induced by the inner product. In other words every Cauchy sequence has a limit in  $\mathbf{H}$ . Examples of Hilbert spaces which are used frequently in this book are :

1. *The space of square summable  $m$ -dimensional sequences,  $\ell_m^2$ .* The elements of this space are sequences  $x = \{x(t)\}_{t \in \mathbb{Z}}$  of real (or complex)  $m$ -dimensional vectors  $x(t)$ , which we shall always write as *row vectors*, indexed by the integer-valued parameter  $t$ , satisfying

$$\|x\|^2 := \sum_{t=-\infty}^{+\infty} x(t) x(t)^* < \infty$$

where  $*$  denotes complex conjugate transpose. This norm is sometimes called the “energy” of the signal  $x$ . It is induced by the inner product

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

A simple proof that  $\ell_m^2$  is complete can be found in YOUNG’s book [129].

2. *The Lebesgue space  $L_m^2$ .*

Let  $[a, b]$  be an interval (not necessarily bounded) of the real line. We shall denote by  $L_m^2([a, b])$  the space of functions with values in a  $m$ -dimensional vector space which are square integrable on  $[a, b]$  with respect to the Lebesgue measure. The values,  $f(t)$ , of the functions will also be written as *row vectors*. It is well-known that this space is a Hilbert space under the inner product

$$\langle f, g \rangle := \int_a^b f(t) g(t)^* dt \quad .$$

No subscript will be used to denote the scalar  $\ell^2$  and  $L^2$  spaces. The reason for using a row-vector notation for these spaces is that their elements naturally appear as multipliers in the combination of vector random quantities. Other important examples of Hilbert spaces (e.g. the Hardy spaces  $H_m^2$ ) will be introduced in the next sections.

In this book the term *subspace* of a Hilbert space  $\mathbf{H}$ , will in general mean *closed* subspace. The *sum* of two linear vector spaces  $\mathbf{X} + \mathbf{Y}$ , is, by definition, the linear vector space  $\{x + y \mid x \in \mathbf{X}, y \in \mathbf{Y}\}$ . It may happen that, even when  $\mathbf{X}$  and  $\mathbf{Y}$  are both (closed) subspaces, their sum fails to be closed. This may happen only when both  $\mathbf{X}$ ,  $\mathbf{Y}$  are infinite dimensional. A classical example of sum of two infinite-dimensional subspaces which is not closed, can be found in [44, p. 28.]. The *(closed) vector sum* of  $\mathbf{X}$  and  $\mathbf{Y}$ , denoted  $\mathbf{X} \vee \mathbf{Y}$ , is the closure of  $\mathbf{X} + \mathbf{Y}$ .

The symbols  $+$ ,  $\vee$ ,  $\dot{+}$  and  $\oplus$  will denote sum, (closed) vector sum, direct sum (i.e.  $\mathbf{X} \dot{+} \mathbf{Y} = \mathbf{X} + \mathbf{Y}$  with the extra condition that  $\mathbf{X} \cap \mathbf{Y} = \{0\}$ ), and *orthogonal* direct sum of subspaces. An orthogonal sum of subspaces is always closed. The *linear vector space generated by a family of elements*  $\{x_\alpha\}_{\alpha \in \mathbb{A}} \subset \mathbf{H}$ , denoted  $\text{span}\{x_\alpha \mid \alpha \in \mathbb{A}\}$  is the vector space whose elements are all finite linear combinations of the *generators*  $\{x_\alpha\}$ . The *subspace generated by the family*  $\{x_\alpha\}_{\alpha \in \mathbb{A}}$  is the closure of this linear vector space, and is denoted by  $\overline{\text{span}}\{x_\alpha \mid \alpha \in \mathbb{A}\}$ .



Important examples of subspaces of  $\ell_m^2$  are the subspaces of *causal* signals,  $\ell_m^{2+}$ , which are zero for negative values of  $t$  ( $f(t) = 0, t < 0$ ) and the *anticausal* signals,  $\ell_m^{2-}$ , which are instead zero for positive values of  $t$ , ( $f(t) = 0, t > 0$ ). These two subspaces have a non-empty intersection which is isomorphic to  $\mathbb{R}^m$  (or to  $\mathbb{C}^m$ ). The orthogonal complement,  $\ell_m^{2+\perp}$ , of  $\ell_m^{2+}$  in  $\ell_m^2$  is the subspace of *strictly anticausal* functions which are zero also for  $t = 0$ . Evidently we have the orthogonal decomposition

$$\ell_m^2 = \ell_m^{2+} \oplus \ell_m^{2+\perp}. \tag{A.1.1}$$

We shall often have the occasion of dealing with series of orthogonal random variables. A simple but basic result on convergence of these series is the following.

**Lemma A.1.1.** *A series of orthogonal elements in a Hilbert space,*

$$\sum_{k=0}^{\infty} x_k, \quad x_k \perp x_j, \quad k \neq j,$$

*converges if and only if*

$$\sum_{k=0}^{\infty} \|x_k\|^2 < \infty \tag{A.1.2}$$

*i.e., the series of the square norms of the elements converges.*

**Proof.** In fact the series converges if and only if

$$\left\| \sum_{k=0}^m x_k - \sum_{k=0}^{n-1} x_k \right\| \rightarrow 0$$

as  $n, m \rightarrow \infty$  which is the same as  $\left\| \sum_{k=n}^m x_k \right\|^2 \rightarrow 0$  which in turn is equivalent to  $\sum_{k=n}^m \|x_k\|^2 \rightarrow 0$  as  $n, m \rightarrow \infty$ .  $\square$

Let  $\{e_k\}$  be an orthonormal sequence in a Hilbert space  $\mathbf{H}$ . Since, for an arbitrary  $x \in \mathbf{H}$ , the “approximation error”

$$\left\| x - \sum_{k=0}^N \langle x, e_k \rangle e_k \right\|^2 \leq \|x\|^2 - \sum_{k=0}^N |\langle x, e_k \rangle|^2$$

is non-negative, we have

$$\sum_{k=0}^N |\langle x, e_k \rangle|^2 \leq \|x\|^2 \quad \text{for all } N$$

and hence the series  $\sum_{k=0}^{\infty} \langle x, e_k \rangle e_k$  converges. An immediate consequence of this fact is that the sequence of the Fourier coefficients  $f(k) := \langle x, e_k \rangle, k = 1, \dots$  is in  $\ell^2$ .

### Operators and their adjoints

A *linear operator*  $T$  from a Hilbert space  $\mathbf{H}_1$  to another Hilbert space  $\mathbf{H}_2$ , is a linear map between the two spaces. In general  $T$  may not be defined on the whole of  $\mathbf{H}_1$  (think for example of the differentiation operator in  $L_m^2$ ). If this is instead the case, one says that  $T$  is defined *on*  $\mathbf{H}_1$ . The simplest linear operators to work with are the *continuous*, or *bounded* operators, which are defined on the whole space and satisfy an inequality of the type

$$\|Tx\|_2 \leq k\|x\|_1, \quad x \in \mathbf{H}_1$$

for some constant  $k$ , the subscripts referring to the different norms in the two Hilbert spaces. As one can see, a continuous linear operator is in reality uniformly continuous. The infimum of the  $k$ 's for which the inequality holds is called the *norm* of the operator  $T$  and is denoted by  $\|T\|$ .

**Proposition A.1.2.** *Let  $T : \mathbf{H} \rightarrow \mathbf{H}$  be a bounded operator, then*

$$\|T\| = \sup \frac{\|Tf\|}{\|f\|}, \quad f \in \mathbf{H}$$

or, equivalently

$$\|T\| = \sup \frac{|\langle Tf, g \rangle|}{\|f\|\|g\|}, \quad f, g \in \mathbf{H}.$$

If  $T$  is bounded, it is easy to see that there is a unique bounded linear operator  $T^* : \mathbf{H}_2 \rightarrow \mathbf{H}_1$ , which satisfies

$$\langle Tx, z \rangle_2 = \langle x, T^*z \rangle_1 \quad \forall x \in \mathbf{H}_1, z \in \mathbf{H}_2$$

The operator  $T^*$  is called the *adjoint* of  $T$ . Adjoints of unbounded operators may also exist, under suitable conditions. A linear operator from  $\mathbf{H}$  into itself, for which  $T^* = T$  is called *selfadjoint*. On a finite dimensional space the concept of adjoint corresponds to taking the transpose (or the Hermitian conjugate) of the matrix representing the operator with respect to an orthonormal basis (warning: this is *not true* if the basis is not orthonormal!).

The *image* or *range* of an operator  $T$ , is the linear manifold  $\text{Im } T := \{Tx \mid x \in \mathbf{H}_1\}$ . This manifold needs not be closed, i.e. a subspace of  $\mathbf{H}_2$ . If this is the case  $T$  is said to have *closed range*. The *kernel* or *nullspace* of an operator  $T$ ,  $\text{Ker } T := \{x \mid Tx = 0\}$ , is instead always closed. Operators for which  $\overline{\text{Im } T} = \mathbf{H}_2$  will be called *densely onto*. The following simple but important result, is sometimes called the *Fredholm alternative*

**Theorem A.1.3.** *Let  $T : \mathbf{H}_1 \rightarrow \mathbf{H}_2$  be a bounded operator from the Hilbert space  $\mathbf{H}_1$  to the Hilbert space  $\mathbf{H}_2$ , then*

$$\mathbf{H}_1 = \text{Ker } T \oplus \overline{\text{Im } T^*} \tag{A.1.3}$$

$$\mathbf{H}_2 = \text{Ker } T^* \oplus \overline{\text{Im } T} \tag{A.1.4}$$

A bounded operator  $T$  is *left-invertible* if there exists a bounded operator  $S$  such that  $ST = I_1$  and *right-invertible* if there exists a bounded operator  $R$  such that  $TR = I_2$ . Clearly, right-invertibility implies that  $T$  is onto  $\mathbf{H}_2$  while left-invertibility implies that  $T$  is injective (i.e. one-to-one). In fact it can be shown that *a bounded operator  $T$  is right-invertible if and only if it is onto*. However the dual statement for left-invertibility is in general false.

**Theorem A.1.4.** *A bounded linear operator from one Hilbert space to another is left-invertible if and only if it is injective and has closed range.*

If  $T$  is both left- and right- invertible it is called *invertible* tout-court. Note that left- or right- inverses are in general non-unique. However a two-sided inverse is unique.

A linear map  $T$  between two Hilbert spaces preserving the inner products, i.e. a map for which

$$\langle Tx, Ty \rangle_2 = \langle x, y \rangle_1, \quad x, y \in \mathbf{H}_1$$

is called an *isometry*. An isometry is always an injective map. The following basic result is used repeatedly in this book.

**Theorem A.1.5.** *Every isometry defined on a family of elements  $\{x_\alpha \mid \alpha \in \mathbb{A}\}$  of a Hilbert space  $\mathbf{H}$  can be extended by linearity and continuity to the whole Hilbert space  $\overline{\text{span}}\{x_\alpha \mid \alpha \in \mathbb{A}\}$  linearly generated by the family  $\{x_\alpha\}$ , preserving the property of isometry. The isometric extension is unique.*

One can find a proof of this result in [106, p.14-15].

Note that isometric operators satisfy the relation  $\langle x, T^*Tx \rangle_1 = \langle x, x \rangle_1$ , from which  $T^*T = I_1$  (the identity operator in  $\mathbf{H}_1$ ). If  $T$  is surjective ( $T\mathbf{H}_1 = \mathbf{H}_2$ ) one sees that

$$T^* = T^{-1}.$$

A surjective isometry is called a *unitary operator*. Two linear operators  $A : \mathbf{H}_1 \rightarrow \mathbf{H}_1$  and  $B : \mathbf{H}_2 \rightarrow \mathbf{H}_2$  which are related by

$$A = T^{-1}BT$$

where  $T$  is unitary, are *unitarily equivalent*. Unitary equivalence is a relation which preserves the fundamental characteristics of a linear operator, among which the spectrum. The *Fourier Transform* which will be defined shortly, is an example of a unitary operator.

A subspace  $\mathbf{X}$  is *invariant* for the operator  $T$  if  $T\mathbf{X} \subset \mathbf{X}$ . If a subspace  $\mathbf{X}$  is invariant for  $T$  we denote by  $T|_{\mathbf{X}}$  the *restriction* of  $T$  to the subspace  $\mathbf{X}$ . Recall that a subspace  $\mathbf{X}$  is said to be *reducing* for a linear operator  $T$  if it is invariant for  $T$  and there is a complementary subspace  $\mathbf{Y}$ , i.e.

$$\mathbf{H} = \mathbf{X} + \mathbf{Y}$$

which is also invariant. In this case  $T$  has, with respect to this decomposition, a matrix representation

$$T = \begin{bmatrix} T|_{\mathbf{X}} & 0 \\ 0 & T|_{\mathbf{X}^\perp} \end{bmatrix}$$

**Lemma A.1.6.** *Let  $T$  be a linear operator on a Hilbert space  $\mathbf{H}$ . Then*

$$T\mathbf{X} \subset \mathbf{X} \Leftrightarrow T^*\mathbf{X}^\perp \subset \mathbf{X}^\perp$$

*If  $T$  is self-adjoint, both  $\mathbf{X}$  and  $\mathbf{X}^\perp$  are reducing for  $T$ .*

**Proof.**

For  $\mathbf{X}$  is  $T$ -invariant iff  $\langle Tx, y \rangle = 0$  for all  $x \in \mathbf{X}$  and  $y \in \mathbf{X}^\perp$ . Then just apply the definition of adjoint.  $\square$

## A.2 Subspace algebra

It is well-known that the family of subspaces of a vector space forms a lattice (i.e. a partially ordered set where any pair of elements has an inf and sup) with respect to the operations of intersection ( $\cap$ ) and vector sum ( $\vee$ ). Note that while the Boolean operations on sets are always distributive, i.e.  $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$  for any sets  $A, B, C$ , this does no longer hold for subspace operations when the set-theoretical operation of union is substituted by vector sum.

It is well-known that the lattice of subspaces of a vector space is not distributive. If the vector space is *finite-dimensional*, this lattice is *modular*, in the sense that, if  $\mathbf{A}, \mathbf{B}, \mathbf{C}$  are subspaces and  $\mathbf{A} \supset \mathbf{B}$ , then

$$\mathbf{A} \cap (\mathbf{B} \vee \mathbf{C}) = (\mathbf{A} \cap \mathbf{B}) \vee (\mathbf{A} \cap \mathbf{C}) = \mathbf{B} \vee (\mathbf{A} \cap \mathbf{C}). \quad (\text{A.2.1})$$

The modularity condition can obviously also be stated exchanging  $\mathbf{C}$  and  $\mathbf{B}$  and requiring  $\mathbf{A} \supset \mathbf{C}$ , in which case the last member should be substituted by  $(\mathbf{A} \cap \mathbf{B}) \vee \mathbf{C}$ . For arbitrary subspaces the left member in (A.2.1) contains, but needs not be equal to,  $(\mathbf{A} \cap \mathbf{B}) \vee (\mathbf{A} \cap \mathbf{C})$ . It is easy to construct counterexamples in  $\mathbb{R}^2$ .

The non-modularity in an infinite dimensional space has to do with the fact that the vector sum of two subspaces may not be closed, see [?, p. 175]. By substituting “sum” in place of “vector sum” we can obtain somewhat more general statements.

**Proposition A.2.1.** *Let  $\mathbf{A}, \mathbf{B}, \mathbf{C}$  be vector subspaces. If one of them is contained in any of the others, the distributive law*

$$\mathbf{A} \cap (\mathbf{B} + \mathbf{C}) = (\mathbf{A} \cap \mathbf{B}) + (\mathbf{A} \cap \mathbf{C}) \quad (\text{A.2.2})$$

*holds. In particular an analogous statement holds if  $\mathbf{B}, \mathbf{C}$  are closed orthogonal subspaces when the sum is replaced by orthogonal direct sum.*

**Proof.** First note that if either  $\mathbf{B} \subset \mathbf{C}$  or  $\mathbf{B} \supset \mathbf{C}$ , we clearly have  $\mathbf{A} \cap (\mathbf{B} + \mathbf{C}) = (\mathbf{A} \cap \mathbf{B}) + (\mathbf{A} \cap \mathbf{C})$  since one of the two subspaces in the last sum is contained in the other.

Now, suppose that  $\mathbf{A} \supset \mathbf{B}$ . If  $\alpha \in \mathbf{A} \cap (\mathbf{B} + \mathbf{C})$  then  $\alpha = \beta + \gamma$  for some  $\beta \in \mathbf{B}$  and  $\gamma \in \mathbf{C}$ . Since  $\beta$  must belong to  $\mathbf{A}$ ,  $\gamma = \alpha - \beta \in \mathbf{A}$  as well, whence  $\alpha = \beta + \gamma \in (\mathbf{A} \cap \mathbf{B}) + (\mathbf{A} \cap \mathbf{C})$ . Hence we have shown that the left member of (A.2.2) is contained in the right member and therefore (A.2.2) holds.

Next, suppose that  $\mathbf{A} \subset \mathbf{B}$ . Then

$$\mathbf{A} = \mathbf{A} \cap \mathbf{B} \subset \mathbf{A} \cap (\mathbf{B} + \mathbf{C}) \subset \mathbf{A}$$

which implies that  $\mathbf{A} = \mathbf{A} \cap (\mathbf{B} + \mathbf{C})$ . On the other hand,  $\mathbf{A} = \mathbf{A} + (\mathbf{A} \cap \mathbf{C})$  which is the same as  $\mathbf{A} = (\mathbf{A} \cap \mathbf{B}) + (\mathbf{A} \cap \mathbf{C})$ . Substituting in the previous identity one gets  $(\mathbf{A} \cap \mathbf{B}) + (\mathbf{A} \cap \mathbf{C}) = \mathbf{A} \cap (\mathbf{B} + \mathbf{C})$ . Since, in the two last arguments,  $\mathbf{B}$  and  $\mathbf{C}$  can be interchanged, the proof of the first assertion is complete. The last statement follows since orthogonal direct sum is a particular case of sum of subspaces.  $\square$

**Corollary A.2.2.** *Let  $\mathbf{A}, \mathbf{B}, \mathbf{C}$  be (closed) subspaces. If one of them is contained in any of the others and one, but not necessarily the same, is finite-dimensional, the distributive law*

$$\mathbf{A} \cap (\mathbf{B} \vee \mathbf{C}) = (\mathbf{A} \cap \mathbf{B}) \vee (\mathbf{A} \cap \mathbf{C}) \tag{A.2.3}$$

*holds.*

**Proof.** It only remains to show that (A.2.2) implies (A.2.3) when  $\mathbf{A} \supset \mathbf{B}$ , if one of the subspaces  $\mathbf{A}, \mathbf{B}, \mathbf{C}$  is finite-dimensional. If either  $\mathbf{B}$  or  $\mathbf{C}$  is finite-dimensional, their sum is closed and coincides with  $\mathbf{B} \vee \mathbf{C}$ , and required condition holds. If  $\mathbf{A}$  is finite-dimensional, then *a fortiori*  $\mathbf{B}$  is, and again we see that the required condition holds.  $\square$

The *dual* of a lattice expression is obtained by interchanging  $\cap$  and  $\vee$  and reversing the ordering, i.e., exchanging  $\subset$  for  $\supset$  everywhere. In particular, equality signs are preserved. If a lattice equation is true so is its dual. For example, the following two equations (both stating the unimodularity law) are dual to each other

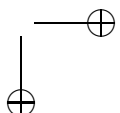
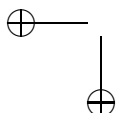
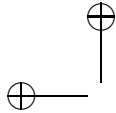
$$\begin{aligned} \mathbf{A} \cap ((\mathbf{A} \cap \mathbf{B}) \vee \mathbf{C}) &= (\mathbf{A} \cap \mathbf{B}) \vee (\mathbf{A} \cap \mathbf{C}) \\ (\mathbf{A} \vee \mathbf{B}) \cap (\mathbf{A} \vee \mathbf{C}) &= \mathbf{A} \vee ((\mathbf{A} \vee \mathbf{B}) \cap \mathbf{C}). \end{aligned}$$

The dual of Proposition A.2.1 can be stated by saying that the inclusion

$$\mathbf{A} \vee (\mathbf{B} \cap \mathbf{C}) \subset (\mathbf{A} \vee \mathbf{B}) \cap (\mathbf{A} \vee \mathbf{C})$$

becomes an equality if (at least) one of the subspaces  $\mathbf{A}, \mathbf{B}, \mathbf{C}$  is contained in any of the others and at least one of them is finite-dimensional.

When the vector space is actually a Hilbert space  $\mathbf{H}$ , then the family of (closed) subspaces of  $\mathbf{H}$  forms a *orthocomplemented lattice*. The duality relations involving orthogonal complement are described in the following Proposition.



**Proposition A.2.3.** *For any family of subspace  $\mathbf{X}_\alpha$  of a Hilbert space, the following identities hold,*

$$(\bigvee_\alpha \mathbf{X}_\alpha)^\perp = \bigcap_\alpha \mathbf{X}_\alpha^\perp, \quad (\bigcap_\alpha \mathbf{X}_\alpha)^\perp = \bigvee_\alpha \mathbf{X}_\alpha^\perp \quad (\text{A.2.4})$$

The following lemma describes other situations where some weak form of distributivity can be proven to hold.

**Lemma A.2.4.** *Let  $\mathbf{A}, \mathbf{B}, \mathbf{C}$  be subspaces of a linear vector space with  $\mathbf{A} \cap \mathbf{C} = 0$  and  $\mathbf{B} \cap \mathbf{C} = 0$ . Then*

$$\mathbf{A} \cap (\mathbf{B} + \mathbf{C}) = \mathbf{A} \cap \mathbf{B}, \quad (\text{A.2.5})$$

$$(\mathbf{A} + \mathbf{C}) \cap (\mathbf{B} + \mathbf{C}) = (\mathbf{A} \cap \mathbf{B}) + \mathbf{C}. \quad (\text{A.2.6})$$

*Assume in addition that the vector space is a Hilbert space and that  $\mathbf{A} \perp \mathbf{C}$  and  $\mathbf{B} \perp \mathbf{C}$ , then (A.2.5) and (A.2.6) hold with orthogonal direct sum  $\oplus$  in place of direct sum.*

**Proof.** To prove the first statement, notice that the first member in (A.2.5) certainly contains the intersection  $\mathbf{A} \cap \mathbf{B}$ . Now all elements of the first subspace are vectors  $\alpha \in \mathbf{A}$  which can be written also as  $\alpha = \beta + \gamma$ ,  $\beta \in \mathbf{B}$ ,  $\gamma \in \mathbf{C}$ , so that  $\alpha - \beta = \gamma$ , i.e.  $\gamma \in \mathbf{A} + \mathbf{B}$ . But this can only happen if  $\gamma = 0$  since the only vector that  $\mathbf{C}$  may have in common with  $\mathbf{A} + \mathbf{B}$  is zero. Therefore all elements of the first member are vectors of  $\mathbf{A}$ ,  $\alpha = \beta$  for some  $\beta \in \mathbf{B}$ . This establishes (A.2.5).

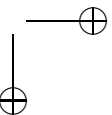
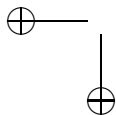
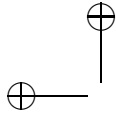
For the second equality, it is again obvious that the left-hand member contains the expression on the right. The opposite inclusion is proved by noting that any  $\lambda \in (\mathbf{A} + \mathbf{C}) \cap (\mathbf{B} + \mathbf{C})$  must be of the form  $\lambda = \alpha + \xi_1 = \beta + \xi_2$ ,  $\xi_1, \xi_2 \in \mathbf{C}$  so that  $\alpha - \beta = \xi_1 - \xi_2 \in \mathbf{C}$ . Since however  $(\mathbf{A} \vee \mathbf{B}) \cap \mathbf{C}$  is zero we must have  $\alpha = \beta$  and  $\xi_1 = \xi_2 = \xi$  so that  $\lambda = \alpha + \xi \in (\mathbf{A} \cap \mathbf{B}) + \mathbf{C}$ .  $\square$

### The shift acting on subspaces

We shall collect below a number of simple but useful technical facts describing how a unitary operator commutes with the most common subspaces operations on a Hilbert space.

**Proposition A.2.5.** *Let  $U$  be a unitary operator acting on a Hilbert space  $\mathbf{H}$  and  $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$  be subspaces of  $\mathbf{H}$ . Then*

1.  $\mathbf{X} \subset \mathbf{Y} \Leftrightarrow U\mathbf{X} \subset U\mathbf{Y}$ .
2.  $\mathbf{X} = \bigcap_{\alpha \in A} \mathbf{X}_\alpha \Leftrightarrow U\mathbf{X} = \bigcap_{\alpha \in A} U\mathbf{X}_\alpha$  for an arbitrary family of subspaces  $\mathbf{X}_\alpha$ .
3.  $\overline{\text{span}}\{Ux_\alpha \mid \alpha \in A\} = U\overline{\text{span}}\{x_\alpha \mid \alpha \in A\}$  for any family of elements  $x_\alpha \in \mathbf{H}$ .
4.  $\mathbf{X} = \bigvee_{\alpha \in A} \mathbf{X}_\alpha \Leftrightarrow U\mathbf{X} = \bigvee_{\alpha \in A} U\mathbf{X}_\alpha$  for an arbitrary family of subspaces  $\mathbf{X}_\alpha$ . The statement also holds if in place of vector sums one has direct sums or orthogonal direct sums.



*Proof.*

1. The direct implication  $\Rightarrow$  is valid for an arbitrary function. The converse follows, for the same reason, by applying  $U^*$  to the subspace inclusion on the right.
2. This statement follows from general properties of the pre-image  $f^{-1}$  of an arbitrary function  $f$ . Precisely,  $f^{-1}$  is a Boolean algebra isomorphism with respect to set-theoretical operations, in particular

$$f^{-1}(\cap_{\alpha \in A} \mathbf{X}_\alpha) = \cap_{\alpha \in A} f^{-1}(\mathbf{X}_\alpha)$$

for an arbitrary family of sets  $\mathbf{X}_\alpha$ .

3. This is just the statement that  $U$  has closed range which is obvious since  $\|Ux\| \geq \|x\|$ .
4. The equality on the right is clearly true for the vector space linearly generated by the  $\mathbf{X}_\alpha$ 's. It extends to the closures in force of the previous result.

□

## A.3 Some facts from linear algebra

### The Moore-Penrose pseudoinverse

Every  $A \in \mathbb{C}^{n \times m}$  is a one-to-one and onto map from the orthogonal complement of its nullspace  $(\text{Ker } A)^\perp = \text{Im } A^*$  to its range,  $\text{Im } A$ . Hence the restriction of  $A$  to  $\text{Im } A^*$  as a map onto  $\text{Im } A$  has an inverse. The Moore-Penrose *generalized- or pseudo- inverse*,  $A^+$ , of  $A$  is introduced by solving the equation  $Ax = y$ ,  $y \in \mathbb{C}^n$  in the least square sense as follows. One first projects  $y$  onto  $\text{Im } A$  obtaining a vector  $\hat{y} \in \text{Im } A$ ; then there is one and only one vector  $x_0 \in (\text{Ker } A)^\perp$  which solves  $Ax_0 = \hat{y}$ . The solution  $x_0 = A^+\hat{y}$ , can be shown to be the unique solution of the least squares problem  $\min_x \|Ax - y\|$ , of minimum norm. The formal characterization of  $A^+$  following from this definition is stated in the following lemma (whose proof can be found in the literature).

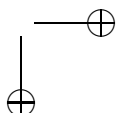
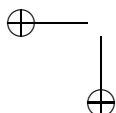
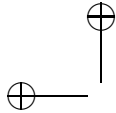
**Lemma A.3.1.** *For any  $A \in \mathbb{C}^{n \times m}$  the matrix  $A^+$  is the unique matrix in  $\mathbb{C}^{m \times n}$  which satisfies the following four properties*

$$A^+AA^+ = A^+, \quad AA^+A = A, \quad (AA^+)^* = AA^+, \quad (A^+A)^* = A^+A. \quad (\text{A.3.1})$$

*This matrix is called the Moore-Penrose pseudoinverse of  $A$ .*

Observe that when  $A$  is square nonsingular  $A^+ = A^{-1}$ .

The following properties follow directly from the four basic relations (A.3.1).



1. Let  $A \in \mathbb{R}^{n \times m}$ ,  $T_1 \in \mathbb{R}^{n \times n}$  and  $T_2 \in \mathbb{R}^{m \times m}$ , with  $T_1$  e  $T_2$  orthogonal matrices. Then

$$(T_1 A T_2)^+ = T_2^{-1} A^+ T_1^{-1} = T_2^* A^+ T_1^*. \quad (\text{A.3.2})$$

2. Let

$$A = \begin{bmatrix} A_1 & 0 \\ 0 & 0 \end{bmatrix} \in \mathbb{R}^{n \times m}$$

with  $A_1 \in \mathbb{R}^{p \times p}$  invertibile. One has,

$$A^+ = \begin{bmatrix} A_1^{-1} & 0 \\ 0 & 0 \end{bmatrix}. \quad (\text{A.3.3})$$

3. Let

$$A = [ U_1 \quad U_2 ] \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_1' \\ V_2' \end{bmatrix}$$

be the singular value decomposition of  $A$  with  $\Sigma_1$  the diagonal matrix of nonzero singular values. Then

$$A^+ = [ V_1 \quad V_2 ] \begin{bmatrix} \Sigma_1^{-1} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} U_1' \\ U_2' \end{bmatrix}. \quad (\text{A.3.4})$$

For a more complete treatment of the subject the reader should consult [10].

### Lyapunov equations

**Proposition A.3.2.** Consider the discrete matrix Lyapunov equation

$$X = AXA' + Q. \quad (\text{A.3.5})$$

Assume  $Q = BB'$ ; then any two of the following statements imply the remaining one

- i)  $(A, B)$  is a reachable pair;
- ii) matrix  $A$  has all eigenvalues strictly inside the unit circle i.e.  $|\lambda\{A\}| < 1$ ;
- iii) (A.3.5) admits a symmetric positive definite solution.

If  $|\lambda\{A\}| < 1$ , the solution of (A.3.5) is unique and is given by

$$P = \sum_0^{\infty} A^k B B' (A')^k. \quad (\text{A.3.6})$$

**Proof.** The last sentence actually follows from a lemma of a slightly wider scope.

**Lemma A.3.3.** The Lyapunov equation (A.3.5) has a unique solution for an arbitrary  $Q \in \mathbb{R}^{n \times n}$ , if and only if the spectrum of  $A$  does not contain reciprocal elements, i.e.  $\lambda_k \in \sigma(A) \Rightarrow 1/\lambda_k \notin \sigma(A)$ .



**Proof.** The linear map from  $\mathbb{R}^{n \times n}$  in to itself

$$X \mapsto X - AXA' \tag{A.3.7}$$

has eigenvalues  $1 - \lambda_k \bar{\lambda}_j$  where the  $\{\lambda_k\}$  are the eigenvalues of  $A$ . In fact, if  $a_k$  is the eigenvector of  $A$  corresponding to  $\lambda_k$ , one readily sees that

$$X_{kj} = a_k a_j^*$$

is the “eigenmatrix” of (A.3.7) corresponding to the eigenvalue  $1 - \lambda_k \bar{\lambda}_j$ . It is obvious that (A.3.7) is injective (and hence also surjective) iff it has no zero eigenvalues, i.e.  $1 - \lambda_k \bar{\lambda}_j \neq 0$ .  $\square$

Therefore, whenever condition ii) holds, there always exist a unique solution of (A.3.5). Since the (convergent) series (A.3.6) is a solution, it is the only solution.

That i) and ii) imply iii) is obvious, since  $P \geq \sum_{k=0}^{n-1} A^k B B' (A')^k$ , which is the reachability Gramian of  $(A, B)$ .

That i) and iii) imply ii) can be seen by contradiction. In fact, assume that  $A$  has an eigenvalue  $\lambda_0$  of modulus greater or equal to one with a corresponding eigenvector  $a$  (in general complex). It follows from iii) that

$$a^* P a = |\lambda_0|^2 a^* P a + a^* B B' a,$$

that is,

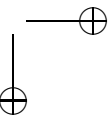
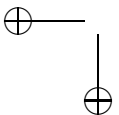
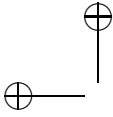
$$(1 - |\lambda_0|^2) a^* P a = a^* B B' a, \tag{A.3.8}$$

where the left member is  $\leq 0$  since  $a^* P a > 0$ , while the term on the right is  $\geq 0$ . It follows that both must be zero. In particular,  $a^* B = 0$  so there is an eigenvector  $a$  of  $A$  that is orthogonal to the columns of  $B$ . This contradicts reachability of  $(A, B)$ . Hence  $|\lambda_0|^2$  must be  $< 1$ .

A similar argument shows that ii) and iii) imply i). In fact, assuming that there is some vector  $a \neq 0$  orthogonal to the columns of  $[B A B \dots A^{n-1} B]$ , one easily shows, using Cayley-Hamilton’s Theorem, that  $a$  must also be orthogonal to the columns of  $A^k B$  for all  $k$ . Hence

$$a^* \sum_0^{+\infty} A^k B B' (A')^k a = a^* P a = 0,$$

which contradicts the strict positivity of  $P$ .  $\square$



# Bibliography

- [1] Hirotugu Akaike. Markovian representation of stochastic processes by canonical variables. *SIAM J. Control*, 13:162–173, 1975.
- [2] N. I. Akhiezer and I. M. Glazman. *Theory of linear operators in Hilbert space. Vol. I*. Translated from the Russian by Merlynd Nestell. Frederick Ungar Publishing Co., New York, 1961.
- [3] B. D. O. Anderson and M. R. Gevers. Identifiability of linear stochastic systems operating under linear feedback. *Automatica—J. IFAC*, 18(2):195–213, 1982.
- [4] Brian D. O. Anderson. The inverse problem of stationary covariance generation. *J. Statist. Phys.*, 1:133–147, 1969.
- [5] Jean-Pierre Aubin. *Applied functional analysis*. John Wiley & Sons, New York-Chichester-Brisbane, 1979. Translated from the French by Carole Labrousse, With exercises by Bernard Cornet and Jean-Michel Lasry.
- [6] J. S. Baras and R. W. Brockett.  $H^2$ -functions and infinite-dimensional realization theory. *SIAM J. Control*, 13:221–241, 1975.
- [7] John S. Baras and Patrick Dewilde. Invariant subspace methods in linear multivariable-distributed systems and lumped-distributed network synthesis. *Proc. IEEE*, 64(1):160–178, 1976. Recent trends in system theory.
- [8] Paul A. Bekker. Identification in restricted factor models and the evaluation of rank conditions. *J. Econometrics*, 41(1):5–16, 1989.
- [9] Paul A. Bekker and Jan de Leeuw. The rank of reduced dispersion matrices. *Psychometrika*, 52(1):125–135, 1987.
- [10] T.N.E. Ben-Israel, A. anf Greville. *Generalized inverses: theory and applications*. Wiley, New York, 1977.
- [11] Arne Beurling. On two problems concerning linear transformations in Hilbert space. *Acta Math.*, 81:17, 1948.

- [12] Henrick W. Bode and Claude E. Shannon. A simplified derivation of linear least-squares smoothing and prediction theory. *Proceedings of the IRE*, 38:417–425, 1950.
- [13] R. W. Brockett. *Finite Dimensional Linear Systems*. John Wiley and Sons, 1970.
- [14] P. Caines. *Linear Stochastic Systems*. John Wiley and Sons, 1988.
- [15] P.E. Caines and C.W. Chan. Estimation, identification and feedback. In R. Mehra and D. Lainiotis, editors, *System Identification: Advances and Case Studies*, pages 349–405. Academic Press, 1976.
- [16] Peter E. Caines and David Delchamps. Splitting subspaces, spectral factorization and the positive real equation: structural features of the stochastic realization problem. In *Proceedings of the 19th IEEE Conference on Decision and Control, Albuquerque, NM*, pages 358–362, 1980.
- [17] Harald Cramér. On the theory of stationary random processes. *Annals of Mathematics*, 41:215–230 (1940), 1940.
- [18] Harald Cramér. On harmonic analysis in certain function spaces. *Ark. Mat. Astr. Fys.*, 28B:17, 1942.
- [19] Harald Cramér. A contribution to the theory of stochastic processes. In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, 1950*, pages 329–339, Berkeley and Los Angeles, 1951. University of California Press.
- [20] Harald Cramér. A contribution to the multiplicity theory of stochastic processes. In *Proc. Fifth Berkeley Sympos. Math. Statist. and Probability (Berkeley, Calif., 1965/66), Vol. II: Contributions to Probability Theory, Part 1*, pages 215–221. Univ. California Press, Berkeley, Calif., 1967.
- [21] Harald Cramér. On the multiplicity of a stochastic vector process. *Ark. Mat.*, 16(1):89–94, 1978.
- [22] J. Dieudonné. *Foundations of modern analysis*. Pure and Applied Mathematics, Vol. X. Academic Press, New York, 1960.
- [23] J. L. Doob. *Stochastic processes*. Wiley Classics Library. John Wiley & Sons Inc., New York, 1990. Reprint of the 1953 original, A Wiley-Interscience Publication.
- [24] R. G. Douglas, H. S. Shapiro, and A. L. Shields. Cyclic vectors and invariant subspaces for the backward shift operator. *Ann. Inst. Fourier (Grenoble)*, 20(fasc. 1):37–76, 1970.

- [25] Nelson Dunford and Jacob T. Schwartz. *Linear operators. Part II*. Wiley Classics Library. John Wiley & Sons Inc., New York, 1988. Spectral theory. Selfadjoint operators in Hilbert space, With the assistance of William G. Bade and Robert G. Bartle, Reprint of the 1963 original, A Wiley-Interscience Publication.
- [26] Peter L. Duren. *Theory of  $H^p$  spaces*. Pure and Applied Mathematics, Vol. 38. Academic Press, New York, 1970.
- [27] H. Dym and H. P. McKean. *Gaussian processes, function theory, and the inverse spectral problem*. Academic Press [Harcourt Brace Jovanovich Publishers], New York, 1976. Probability and Mathematical Statistics, Vol. 31.
- [28] Pierre Faurre, Michel Clerget, and François Germain. *Opérateurs rationnels positifs*, volume 8 of *Méthodes Mathématiques de l'Informatique [Mathematical Methods of Information Science]*. Dunod, Paris, 1979. Application à l'hyperstabilité et aux processus aléatoires.
- [29] D. L. Fisk. Quasi-martingales. *Trans. Amer. Math. Soc.*, 120:369–389, 1965.
- [30] Ciprian Foiaş and Arthur E. Frazho. A note on unitary dilation theory and state spaces. *Acta Sci. Math. (Szeged)*, 45(1-4):165–175, 1983.
- [31] Arthur E. Frazho. Models for noncommuting operators. *J. Funct. Anal.*, 48(1):1–11, 1982.
- [32] Arthur E. Frazho. On minimal splitting subspaces and stochastic realizations. *SIAM J. Control Optim.*, 20(4):553–562, 1982.
- [33] P. A. Fuhrmann. On realization of linear systems and applications to some questions of stability. *Math. Systems Theory*, 8(2):132–141, 1974/75.
- [34] Paul A. Fuhrmann. *Linear systems and operators in Hilbert space*. McGraw-Hill International Book Co., New York, 1981.
- [35] F. R. Gantmacher. *The theory of matrices. Vol. 1*. AMS Chelsea Publishing, Providence, RI, 1998. Translated from the Russian by K. A. Hirsch, Reprint of the 1959 translation.
- [36] John B. Garnett. *Bounded analytic functions*, volume 96 of *Pure and Applied Mathematics*. Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York, 1981.
- [37] M. R. Gevers and B. D. O. Anderson. Representations of jointly stationary stochastic feedback processes. *Internat. J. Control*, 33(5):777–809, 1981.
- [38] I. I. Gikhman and A. V. Skorokhod. *Introduction to the theory of random processes*. Translated from the Russian by Scripta Technica, Inc. W. B. Saunders Co., Philadelphia, Pa., 1969.

- [39] Iosif I. Gikhman and Anatoli V. Skorokhod. *The theory of stochastic processes. I.* Classics in Mathematics. Springer-Verlag, Berlin, 2004. Translated from the Russian by S. Kotz, Reprint of the 1974 edition.
- [40] Gene H. Golub and Charles F. Van Loan. *Matrix computations.* Johns Hopkins Studies in the Mathematical Sciences. Johns Hopkins University Press, Baltimore, MD, third edition, 1996.
- [41] C. W. J. Granger. Economic processes involving feedback. *Information and Control*, 6:28–48, 1963.
- [42] Paul R. Halmos. *Measure Theory.* D. Van Nostrand Company, Inc., New York, N. Y., 1950.
- [43] Paul R. Halmos. Shifts on hilbert spaces. *Journal für die reine un angewandte Mathematik*, 208:102–112, 1961.
- [44] Paul R. Halmos. *Introduction to Hilbert space and the theory of spectral multiplicity.* AMS Chelsea Publishing, Providence, RI, 1998. Reprint of the second (1957) edition.
- [45] Henry Helson. *Lectures on invariant subspaces.* Academic Press, New York, 1964.
- [46] Henry Helson and David Lowdenslager. Prediction theory and Fourier series in several variables. II. *Acta Math.*, 106:175–213, 1961.
- [47] J. William Helton. Discrete time systems, operator models, and scattering theory. *J. Functional Analysis*, 16:15–38, 1974.
- [48] J. William Helton. Systems with infinite-dimensional state space: the Hilbert space approach. *Proc. IEEE*, 64(1):145–160, 1976. Recent trends in system theory.
- [49] T. Hida. Canonical representations of gaussian processes and their applications. *Mem.Coll. Sci. Univ. Kyôto*, 33:109–155, 1960.
- [50] Einar Hille. *Analytic function theory, Vol I.* Ginn and Co, reprinted by AMS Chelsea Pub., A.M.S. Providence, R.I., 1982.
- [51] Kenneth Hoffman. *Banach spaces of analytic functions.* Prentice-Hall Series in Modern Analysis. Prentice-Hall Inc., Englewood Cliffs, N. J., 1962.
- [52] Il'dar Abdulovich Ibragimov and Y. A. Rozanov. *Gaussian random processes*, volume 9 of *Applications of Mathematics.* Springer-Verlag, New York, 1978. Translated from the Russian by A. B. Aries.
- [53] N. Ikeda and S. Watanabe. *Stochastic differential equations and diffusion processes.* North Holland, Dordrecht, 1981.

- [54] Thomas Kailath. *Linear systems*. Prentice-Hall Inc., Englewood Cliffs, N.J., 1980. Prentice-Hall Information and System Sciences Series.
- [55] R. E. Kalman. Lyapunov functions for the problem of Lur'e in automatic control. *Proc. Nat. Acad. Sci. U.S.A.*, 49:201–205, 1963.
- [56] R. E. Kalman. Linear stochastic filtering theory—reappraisal and outlook. In *Proc. Sympos. on System Theory (New York, 1965)*, pages 197–205. Polytechnic Press, Polytechnic Inst. Brooklyn, Brooklyn, N.Y., 1965.
- [57] R. E. Kalman. System identification from noisy data. In A. Bednarek and L. Cesari, editors, *Dynamical Systems II*, pages 135–164. Academic Press, New York, 1982.
- [58] R. E. Kalman. Identifiability and modeling in econometrics. In *Developments in statistics, Vol. 4*, volume 4 of *Develop. in Statist.*, pages 97–136. Academic Press, New York, 1983.
- [59] R. E. Kalman. A theory for the identification of linear relations. In *Frontiers in pure and applied mathematics*, pages 117–132. North-Holland, Amsterdam, 1991.
- [60] R. E. Kalman, P. L. Falb, and M. A. Arbib. *Topics in mathematical system theory*. McGraw-Hill Book Co., New York, 1969.
- [61] Kari Karhunen. Zur Spektraltheorie stochastischer Prozesse. *Ann. Acad. Sci. Fennicae. Ser. A. I. Math.-Phys.*, 1946(34):7, 1946.
- [62] Kari Karhunen. Über lineare Methoden in der Wahrscheinlichkeitsrechnung. *Ann. Acad. Sci. Fennicae. Ser. A. I. Math.-Phys.*, 1947(37):79, 1947.
- [63] A. Kolmogoroff. Interpolation und Extrapolation von stationären zufälligen Folgen. *Bull. Acad. Sci. URSS Sér. Math. [Izvestia Akad. Nauk. SSSR]*, 5:3–14, 1941.
- [64] A. N. Kolmogoroff. Sur l'interpolation et extrapolation des suites stationnaires. *C. R. Acad. Sci Paris*, 208:2043–2045, 1939.
- [65] A. N. Kolmogoroff. Stationary sequences in Hilbert's space. *Bolletín Moskovskogo Gosudarstvenogo Universiteta. Matematika*, 2:40pp, 1941.
- [66] A. N. Kolmogorov. *Foundations of the theory of probability*. Chelsea Publishing Co., New York, 1956. Translation edited by Nathan Morrison, with an added bibliography by A. T. Bharuch-Reid.
- [67] M. G. Kreĭn. On a basic approximation problem of the theory of extrapolation and filtration of stationary random processes. *Doklady Akad. Nauk SSSR (N.S.)*, 94:13–16, 1954.
- [68] Peter D. Lax. Translation-invariant subspaces. *Acta Mathematica*, 101(5):163–178, 1959.

- [69] Peter D. Lax and Ralph S. Phillips. *Scattering theory*, volume 26 of *Pure and Applied Mathematics*. Academic Press Inc., Boston, MA, second edition, 1989. With appendices by Cathleen S. Morawetz and Georg Schmidt.
- [70] W. Ledermann. On the rank of the reduced correlation matrix in multiple factor analysis. *Psychometrika*, 2:85–93, 1937.
- [71] W. Ledermann. On a problem concerning matrices with variable diagonal elements. *Proc. Royal Soc. Edinburgh*, XL:1–17, 1939.
- [72] Anders Lindquist and Michèle Pavon. Markovian representation of discrete-time stationary stochastic vector processes. In *Proceedings of the 20th IEEE Conference on Decision and Control (San Diego, 1981)*, New York, 1981. IEEE.
- [73] Anders Lindquist and Michele Pavon. On the structure of state-space models for discrete-time stochastic vector processes. *IEEE Trans. Automat. Control*, 29(5):418–432, 1984.
- [74] Anders Lindquist, Michèle Pavon, and Giorgio Picci. Recent trends in stochastic realization theory. In *Prediction theory and harmonic analysis*, pages 201–224. North-Holland, Amsterdam, 1983.
- [75] Anders Lindquist and Giorgio Picci. On the structure of minimal splitting subspaces in stochastic realization theory. In *Proceedings of the 1977 IEEE Conference on Decision and Control (New Orleans, La., 1977)*, Vol. 1, pages 42–48. Inst. Electrical Electron. Engrs., New York, 1977.
- [76] Anders Lindquist and Giorgio Picci. A Hardy space approach to the stochastic realization problem. In *Proceedings of the 1978 IEEE Conference on Decision and Control (San Diego, Ca., 1978)*, pages 933–939. Inst. Electrical Electron. Engrs., New York, 1978.
- [77] Anders Lindquist and Giorgio Picci. A state-space theory for stationary stochastic processes. In *Proceedings of the 21st Midwestern Symposium on Circuits and Systems, Ames, Iowa, 1978*, pages 108–113, 1978.
- [78] Anders Lindquist and Giorgio Picci. On the stochastic realization problem. *SIAM J. Control Optim.*, 17(3):365–389, 1979.
- [79] Anders Lindquist and Giorgio Picci. Realization theory for multivariate Gaussian processes I: State space construction. In *Proceedings of the 4th International Symposium on the mathematical Theory of Networks and Systems (Delft, Holland, 1979)*, pages 140–148, 1979.
- [80] Anders Lindquist and Giorgio Picci. Realization theory for multivariate Gaussian processes: II: State space theory revisited and dynamical representations of finite-dimensional state spaces. In *Second International Conference on Information Sciences and Systems (Univ. Patras, Patras, 1979)*, Vol. II, pages 108–129. Reidel, Dordrecht, 1980.



- [81] Anders Lindquist and Giorgio Picci. State space models for stochastic processes. In M. Hazewinkel and Jan Willems, editors, *Stochastic Systems: The Mathematics of Filtering and Identification and Applications*, volume C78 of *NATO Advanced Study Institutes Series*, pages 169–204. Reidel Publ. Co., Dordrecht, Holland, 1981.
- [82] Anders Lindquist and Giorgio Picci. On a condition for minimality of Markovian splitting subspaces. *Systems Control Lett.*, 1(4):264–269, 1981/82.
- [83] Anders Lindquist and Giorgio Picci. Infinite-dimensional stochastic realizations of continuous-time stationary vector processes. In *Topics in operator theory systems and networks (Rehovot, 1983)*, volume 12 of *Oper. Theory Adv. Appl.*, pages 335–350. Birkhäuser, Basel, 1984.
- [84] Anders Lindquist and Giorgio Picci. Forward and backward semimartingale models for Gaussian processes with stationary increments. *Stochastics*, 15(1):1–50, 1985.
- [85] Anders Lindquist and Giorgio Picci. Realization theory for multivariate stationary Gaussian processes. *SIAM J. Control Optim.*, 23(6):809–857, 1985.
- [86] Anders Lindquist and Giorgio Picci. A geometric approach to modelling and estimation of linear stochastic systems. *J. Math. Systems Estim. Control*, 1(3):241–333, 1991.
- [87] Anders Lindquist, Giorgio Picci, and Guy Ruckebusch. On minimal splitting subspaces and Markovian representations. *Math. Systems Theory*, 12(3):271–279, 1978/79.
- [88] R. S. Liptser and A. N. Shiriyayev. *Statistics of random processes. II*. Springer-Verlag, New York, 1978. Applications, Translated from the Russian by A. B. Aries, Applications of Mathematics, Vol. 6.
- [89] P. Masani. Isometric flows on Hilbert space. *Bull Amer. Math. Soc.*, 68:624–632, 1962.
- [90] P. Masani. On the representation theorem of scattering. *Bull. Amer. Math. Soc.*, 74:618–624, 1968.
- [91] H. P. McKean, Jr. Brownian motion with a several-dimensional time. *Teor. Veroyatnost. i Primenen.*, 8:357–378, 1963.
- [92] Gy. Michaletzky, J. Bokor, and P. Várlaki. *Representability of stochastic systems*. Akadémiai Kiadó, Budapest, 1998.
- [93] Gy. Michaletzky and A. Ferante. Splitting subspaces an acausal spectral factors. *J. Math. Systems Estim. Control*, 5(3):1–26, 1995.
- [94] György Michaletzky. Zeros of (non-square) spectral factors and canonical correlations. In *Proc. 11th IFAC World Congress, Tallinn, Estonia*, volume 3, pages 167–172, 1991.

- [95] Berrien Moore, III and Eric A. Nordgren. On quasi-equivalence and quasi-similarity. *Acta Sci. Math. (Szeged)*, 34:311–316, 1973.
- [96] I. P. Natanson. *Theory of functions of a real variable*. Frederick Ungar Publishing Co., New York, 1955. Translated by Leo F. Boron with the collaboration of Edwin Hewitt.
- [97] I. P. Natanson. *Theory of functions of a real variable. Vol. II*. Translated from the Russian by Leo F. Boron. Frederick Ungar Publishing Co., New York, 1961.
- [98] J. Neveu. *Discrete-parameter martingales*. North-Holland Publishing Co., Amsterdam, revised edition, 1975. Translated from the French by T. P. Speed, North-Holland Mathematical Library, Vol. 10.
- [99] Raymond E. A. C. Paley and Norbert Wiener. *Fourier transforms in the complex domain*, volume 19 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI, 1987. Reprint of the 1934 original.
- [100] M. Pavon. Stochastic realization and invariant directions of the matrix Riccati equation. *SIAM J. Control and Optimization*, 28:155–180, 1980.
- [101] V. V. Peller and S. V. Khrushchëv. Hankel operators, best approximations and stationary Gaussian processes. *Uspekhi Mat. Nauk*, 37(1(223)):53–124, 176, 1982.
- [102] Giorgio Picci. Stochastic realization of Gaussian processes. *Proc. IEEE*, 64(1):112–122, 1976. Recent trends in system theory.
- [103] Giorgio Picci. Parametrization of factor-analysis models. *Journal of Econometrics*, 41:17–38, 1989.
- [104] V. M. Popov. Hyperstability and optimality of automatic systems with several control functions. *Rev. Roumaine Sci. Tech. Sér. Électrotech. Énergét.*, 9:629–690, 1964.
- [105] Raymund Rishel. Necessary and sufficient dynamic programming conditions for continuous-time stochastic optimal control. *SIAM J. Control*, 8(4):559–571, 1970.
- [106] Yu. A. Rozanov. *Stationary random processes*. Translated from the Russian by A. Feinstein. Holden-Day Inc., San Francisco, Calif., 1967.
- [107] G. Ruckebusch. On the theory of Markovian representation. In *Measure theory applications to stochastic analysis (Proc. Conf., Res. Inst. Math., Oberwolfach, 1977)*, volume 695 of *Lecture Notes in Math.*, pages 77–87. Springer, Berlin, 1978.

- [108] G. Ruckebusch. A state space approach to the stochastic realization problem. In *Proceedings of the 1978 International Symposium on Circuits and Systems, New York, 1978*.
- [109] Guy Ruckebusch. Représentations markoviennes de processus gaussiens stationnaires. *C. R. Acad. Sc. Paris, Series A*, 282:649–651, 1976.
- [110] Guy Ruckebusch. Factorisations minimales de densités spectrales et représentations markoviennes. In *Proc. Colloque AFCET-SMF, Palaiseau, France, 1978*.
- [111] Guy Ruckebusch. Théorie géométrique de la représentation markovienne. *Ann. Inst. H. Poincaré Sect. B (N.S.)*, 16(3):225–297, 1980.
- [112] Guy Ruckebusch. Markovian representations and spectral factorizations of stationary Gaussian processes. In *Prediction theory and harmonic analysis*, pages 275–307. North-Holland, Amsterdam, 1983.
- [113] Guy Ruckebusch. On the structure of minimal Markovian representations. In *Nonlinear stochastic problems (Algarve, 1982)*, volume 104 of *NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci.*, pages 111–122. Reidel, Dordrecht, 1983.
- [114] Walter Rudin. *Real and complex analysis, 2nd Edition*. McGraw-Hill series in higher mathematics. McGraw-Hill, Inc., New York., 1974.
- [115] C. Stricker. Une caractérisation des quasimartingales. In *Seminaire des Probabilités IX*, Lecture Notes in Mathematics, pages 420–424. Springer, Berlin, 1975.
- [116] Béla Sz.-Nagy and Ciprian Foiaş. *Harmonic analysis of operators on Hilbert space*. Translated from the French and revised. North-Holland Publishing Co., Amsterdam, 1970.
- [117] Béla Sz.-Nagy and Ciprian Foiaş. *Harmonic analysis of operators on Hilbert space*. Translated from the French and revised. North-Holland Publishing Co., Amsterdam, 1970.
- [118] N. Wiener. Differential space. *Journal of Math. Physics*, 2:131–174, 1923.
- [119] N. Wiener and P. Masani. The prediction theory of multivariate stochastic processes. I. The regularity condition. *Acta Math.*, 98:111–150, 1957.
- [120] N. Wiener and P. Masani. The prediction theory of multivariate stochastic processes. II. The linear predictor. *Acta Math.*, 99:93–137, 1958.
- [121] Norbert Wiener. *The extrapolation, interpolation and smoothing of stationary time series*, volume 1942 of *Report of the Services 19, Project DIC-6037, MIT*. John Wiley and Sons, Inc., New York, 1949. Later published in Book form with the same title.

- [122] Jan C. Willems. Least squares stationary optimal control and the algebraic Riccati equation. *IEEE Trans. Automatic Control*, AC-16:621–634, 1971.
- [123] Herman Wold. *A study in the analysis of stationary time series*. Almqvist and Wiksell, Stockholm, 1954. 2d ed, With an appendix by Peter Whittle.
- [124] Eugene Wong and Bruce Hajek. *Stochastic processes in engineering systems*. Springer Texts in Electrical Engineering. Springer-Verlag, New York, 1985.
- [125] V. A. Yakubovich. The solution of some matrix inequalities encountered in automatic control theory. *Dokl. Akad. Nauk SSSR*, 143:1304–1307, 1962.
- [126] Kôsaku Yosida. *Functional analysis*. Die Grundlehren der Mathematischen Wissenschaften, Band 123. Academic Press Inc., New York, 1965.
- [127] Kôsaku Yosida. *Functional analysis*, volume 123 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, sixth edition, 1980.
- [128] D. C. Youla. On the factorization of rational matrices. *IRE Trans.*, IT-7:172–189, 1961.
- [129] Nicholas Young. *An Introduction to Hilbert Space*. Cambridge University Press, Cambridge, 1988.

# Index

- $E^{\mathbf{X}} \mathbf{Y}$ , 9
- $\mathcal{J}_{\zeta}$ , 35
- innovation process; unnormalized (forward), 332
- Kullback Leibler (pseudo)-distance, 334
- mean quadratic variation, 106
- normalized minimum phase spectral factor, 332
- algebraic degree, 361
- algebraic Riccati equation, 148, 296
  - dual, 317
- algebraic Riccati inequality, 295
- all pass filter, 130
- all-pass transfer function, 331
- ambient space, 129, 131, 162, 183, 267
- approximate predictor, 333
- backward predictor space, 166
- backward representation, 127, 128
- backward semimartingale, 107
- backward shift, 17
- balanced
  - positive real, 317
- best linear estimator, 6
- Beurling-Lax Theorem, 86
- Blaschke product, 89
- Canonical Correlation Analysis, 320
- canonical correlation analysis
  - of a stationary process, 321
- canonical correlation coefficients, 12, 320
  - of a stationary process, 322
- canonical factorization, 154, 159
- canonical projection, 154
- canonical variables, 12
- causal equivalence of random processes, 64
- causal/anticausal functions, 47
- causality of random processes, 27
- coherent factorization
  - of Hankel matrices, 357
- coinvariant subspace, 238
- conditional derivative, 108
- conditional expectation, 5
- conditional Lipschitz condition, 29, 108
- conditional orthogonality, 12, 155
- conjugate inner function, 84, 233
- conjugate maximum phase, 140
- conjugate minimum phase, 139
- conjugate outer function, 84, 233
  - in  $\bar{\mathcal{W}}_p^2$ , 102
- constructible, 133
- constructible subspace, 160
- constructibility operator, 159
- constructibility, 133
- controllability, 133
- controllability operator, 197
- decomposable function, 111
- degeneracy, 249
- deterministic component, 27
- deterministic subspace of  $\mathbf{X}$ , 201
- doubly invariant subspace, 18, 187
- dual basis, 194
- Efficiency indices, 313
- error space, 223
- error-space degeneracy, 252
- errors in variables model, 21
- extended observability and constructibility matrices, 326
- factor analysis model, 20

- factor estimation, 23
- factor subspace, 21
- factorizable spectral density, 66
- feedback-free processes, 26
- filtering, 59
- forward innovation, 64
- forward representation, 124
- forward shift, 17
- Fourier transform of a stationary process, 39
- Fourier transform, 48
  - of a continuous-time process, 42
- frame space, 167
- full rank process, 67
- future space, 124, 183
  
- generating process, 78
  - backward, 188
  - forward, 188
- generating processes, 128
  - continuous-time, 268
- generator, 303
  - exact, 215, 303
- generator of the Markovian family, 20
- germ space, 16
- Gramians; Observability and Constructibility, 315
  
- Hankel matrix, 135
- Hankel norm, 330
- Hankel operator, 10, 159
  - factorization of, 159
- Hardy space
  - conjugate, 71
  - conjugate of the half-plane, 97
  - of the half-plane, 97
  - of the unit disk, 71
- Herglotz Theorem, 36
  - for vector processes, 44
- hidden rank, 23
  
- identifiability, 23
- impulse response, 116
- inner function, 84, 233
- inner triplet, 235, 270
- inner-outer factorization, 87
  
- innovations representation
  - unnormalized, 80
- interpolation, 60
- invariant factors, 244
  
- Kalman filter, 148, 297
  - backward
    - steady state, 151
- Kalman-Yakubovich-Popov Lemma, 144
  
- linear matrix inequality, 143, 293
  
- Markov property, 18
- Markovian property, 18
- Markovian representation, 162, 183, 266
  - constructible, 183, 267
  - continuous time, 266
  - controllable, 197
  - degenerate, 249
  - internal, 162, 183, 267
  - minimal, 183, 267
  - normal, 188, 268
  - observable, 183, 267
  - proper, 188, 268
  - purely nondeterministic, 197
  - reachable, 197
  - regular, 252
  - singular, 252
- Markovian splitting subspace, 132
  - proper, 170
  - regular, 252
  - singular, 252
- Markovian triplet, 234, 276
  - tight, 234
- maximum phase, 140
- McMillan degree, 116
- minimal factorization, 119
- minimal spectral factor, 142, 248, 279
- minimum phase, 92, 138
- mod  $\mathcal{O}$ , 69
- mod  $\mathcal{O}$ , 234
- mod  $\mathcal{O}(p)$ , 234
- module structure of  $\mathbf{H}(y)$ , 54
- multiplicity of the shift, 54
  
- Nerode equivalence, 154

- nondegenerate  
     stationary increments process, 112
- observability operator, 158
- observable, 118, 133
- observable subspace, 160
- operator positive-real-lemma-equations, 222
- orthogonal complement, 12
- orthogonal direct sum, 12
- Orthogonal Projection Lemma, 5
- orthogonal projections, 4
- orthogonal stochastic measure, 33
- orthonormalizable process, 62
- outer function, 84, 233  
     in  $\mathcal{W}_p^2$ , 102
- p.n.d., 64, 79
- Paley-Wiener theorem  
     continuous version, 97  
     for  $\mathcal{W}^2$  spaces, 100
- parahermitian symmetry, 45, 74
- Parseval's identity, 49
- partial ordering of  $\mathcal{X}$ , 174
- past and future  
     of a family of subspaces, 18
- past and future spaces, 16
- past space, 123, 183
- perpendicular intersection, 156
- phase function, 239, 258, 270
- positive degree, 361
- positive rational extension, 361
- positive real, 141
- positive real function, 292
- Positive Real Lemma, 144
- positive real part, 292
- positive-real part, 141
- positive-real-lemma equations, 293
- positive-real-lemma-equations, 144, 222
- prediction, 59
- prediction error, 333
- predictor space, 136, 165  
     backward, 136
- Preface, vii
- principal angles, 10, 12
- principal directions, 12
- process  
     with orthogonal increments, 32  
     full rank, 67  
     of finite mean variation, 105  
     orthonormalizable, 62  
     purely nondeterministic, 79  
     reversible, 76, 82
- proper subspace of  $\mathbf{X}$ , 201
- pseudoinverse, 391
- purely deterministic, 75
- purely nondeterministic process, 64
- quasi-invertible, 173
- quasi-similar, 245
- rank  
     of a stationary increments process, 101
- rank of a stationary process, 67
- rank of a stationary process, 57, 67
- rational function  
     strictly proper, 280
- reachability operator, 196
- reachable, 118
- reducing subspace, 18
- regular process, 147
- regular stochastic system, 224
- regularity, 147
- relative error criterion, 337
- remote future, 76, 188
- remote past, 76, 188
- reversibility, 82
- reversible process  
     finite-dimensional case, 199
- Riccati equation, 149
- Riccati inequality, 148
- rigid functions, 87
- scattering pair, 161  
      $\mathbf{X} \sim (\mathbf{S}, \bar{\mathbf{S}})$ , 163
- second-order process, 15
- semimartingale representation, 29, 106
- semimartingales, 105
- shaping filter, 63
- shift, 16, 132
- singular inner function, 90

- singular stochastic system, 224
- singular value decomposition, 10
- singular values; of a transfer function, 330
- smoothing, 60
- spectral density, 140
  - positive-real part, 141
- spectral density matrix, 45
- spectral factor, 66
  - analytic, 74
  - conjugate outer, 84
  - outer, 84
  - analytic, 141, 273, 274
    - continuous time, 262
  - coanalytic, 141, 273, 274
  - conjugate minimum-phase, 275
  - conjugate outer, 273
  - maximum phase, 142
  - maximum-phase, 140
  - minimal, 142, 248, 263, 279
  - minimum phase, 142
  - minimum-phase, 275
  - of full-rank, 66
  - outer, 273
  - stable, 141
- spectral factorization, 66
  - in  $\mathcal{W}^2$ , 103
- spectral isomorphism, 42
- spectral matrix measure of a stationary vector process, 44
- spectral representation, 39
  - of real processes, 43
  - of vector-valued processes, 44
  - continuous time, 41
- splitting subspace, 132, 158
  - constructible, 160
  - Markovian, 132, 162
  - minimal, 158
  - observable, 160
  - partial ordering, 174
  - uniform choice of bases, 176
- stability matrix
  - continuous-time, 261
  - discrete time, 121
- state covariance, 124
- state space, 124
  - constructible, 133
  - internal, 133
  - minimal, 133
  - observable, 133
- state-space degeneracy, 249
- steady state Kalman filter
  - backward, 151
- steady-state Kalman filter, 297
- stochastic balancing, 315
- stochastic component, 27
- stochastic integral, 35
- stochastic model reduction, 329
- stochastic realization, 135
  - backward, 135
  - continuous-time, 261
  - forward, 135
  - interior, 136
  - minimal, 135
- stochastically balanced canonical form, 324
- strictly noncyclic, 271
- strictly noncyclic process, 241
- structural function, 130, 234, 270
- subspace
  - backward-shift-invariant, 77
  - doubly-invariant, 77
  - forward-shift-invariant, 77
  - shift-invariant, 77
  - wandering, 78
- trigonometric polynomial
  - analytic, 85
- trigonometric polynomials, 38
- unconstructible subspace, 160
- uniform choice of bases
  - Markovian representations, 208
  - splitting subspaces, 176
- unobservable subspace, 160
- vector white noise, 47
  - normalized, 47
- wandering subspace, 187
- white noise
  - in discrete time, 42



- white noise processes
  - equal modulo  $\mathcal{O}$ , 69
- whitening filter, 63
  - causal, 65
- Wiener filtering, 61
  - acausal, 61
  - causal, 61
- Wiener process, 32
  - as a stochastic measure, 34
- Wold decomposition, 187
- zeros of a matrix function in  $H^2$ , 90