Víctor Gómez

Multivariate Time Series With Linear State Space Structure

Springer

Multivariate Time Series With Linear State Space Structure

Víctor Gómez

Multivariate Time Series With Linear State Space Structure

Springer

Víctor Gómez
Ministerio de Hacienda y Administraciones
Públicas
Dirección Gral. de Presupuestos
Subdirección Gral. de Análisis y P.E.
Madrid, Spain

ISBN 978-3-319-28598-6 ISBN 978-3-319-28599-3 (eBook)
DOI 10.1007/978-3-319-28599-3

Library of Congress Control Number: 2016938930

Mathematics Subject Classification (2010): 37M10, 62-XX, 62M10, 93E11, 62M20, 60Gxx, 65Fxx

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG Switzerland

To my wife María C.
and
my daughter Berta

Preface

The subject of this book is the estimation of random vectors given observations of a related random process assuming that there is a linear relation between them. Since the class of linear models is very rich, we restrict our attention to those having a state space structure. The origin of this topic can be traced back to illustrious researches such as Laplace, Gauss, and Legendre and, more recently, to H. Wold, A.N. Kolmogorov, and N. Wiener in the late 1930s and earlier 1940s.

The theory received a great impulse with the incorporation of state space models. The main contributor to this development was R.E. Kalman, who also made important related contributions to linear systems, optimal control, stability theory, etc. The subject matter of state space models has expanded a lot in recent years and today includes nonlinear as well as non-Gaussian models. We have limited the scope of the book to linear state space models, however, because otherwise its size would have been excessive.

In this book, the emphasis is on the development of the theory of least-squares estimation for finite-dimensional linear systems with two aims: firstly, that the foundations are solidly laid and, secondly, that efficient algorithms are simultaneously given to perform the necessary computations. For this reason, the theory is presented in all generality without focusing on specific state space models from the beginning, as is the case in, for example, the books by Harvey (1989) or Durbin & Koopman (2012) with regard to structural time series models. The theory developed in this book covers most aspects of what is generally known as the Wiener–Kolmogorov and Kalman filtering theory.

The book is intended for students and researchers in many fields, including statistics, economics and business, engineering, medicine, biology, sociology, etc. It assumes familiarity with basic concepts of linear algebra, matrix theory, and random processes. Some appendices at the end of several chapters provide the reader with background material in some of these or related areas.

Chapters 1 and 2 deal with the definition of orthogonal projection and the introduction of many topics associated with it, including state space and VARMA models. In Chap. 3, stationary processes and their properties are considered. Chapter 4 is dedicated to the general state space model, including many algorithms for filtering and smoothing. General Markov processes are also included. Chapter 5 contains the development of special features associated with time invariant state space models. In Chap. 6, time invariant state space models with inputs are considered. The Wiener–Kolmogorov theory is developed in Chap. 7. First for infinite and then for finite samples. Finally, the SSMMATLAB software package is described in Chap. 8.

Madrid, Spain
November 2015

Víctor Gómez

Computer Software

Many of the algorithms presented in this book have been implemented by the author in MATLAB in a software package called SSMMATLAB (Gómez, 2014, 2015).

A brief description of the SSMMATLAB package is given in Chap. 8. In this chapter, a list of the most important SSMMATLAB functions is given, together with references to the sections of this book with which the functions are connected.

As described in the SSMMATLAB manual, there are many examples and case studies taken from time series books, such as Box & Jenkins (1976), Reinsel (1997), Lütkepohl (2007), and Durbin & Koopman (2012), published time series articles, or simulated series.

Contents

Chapter 1
Orthogonal Projection

1.1 Expectation and Covariance Matrix of Random Vectors

Let $Y = (Y_1', \dots, Y_k')'$ be a k-dimensional random vector. If $E(Y_i) = \mu_i$ is the mean of Y_i, $i = 1, \dots, k$, the **mean vector** of Y is defined by $\mu_y = (\mu_1, \dots, \mu_k)'$.

In a similar way, the expectation of a matrix whose elements are random variables is defined as the matrix in which each element is replaced by the expectation of that element. If Y and $X = (X_1, \dots, X_m)'$ are random vectors, the **covariance matrix** of Y and X is defined as

$$\Sigma_{yx} = \text{Cov}(Y, X) = E[(Y - \mu_y)(X - \mu_x)'].$$

The (i, j) element of Σ_{yx} is the covariance between Y_i and X_j, $\text{Cov}(Y_i, X_j) = E[(Y_i - E(Y_i))(X_j - E(X_j))]$. In particular, the covariance matrix of Y is $\Sigma_{yy} = \text{Cov}(Y) = E[(Y_i - \mu_y)(Y - \mu_y)']$.

Given two random variables, Y and X, the **correlation coefficient** of Y and X is defined by $\text{Corr}(Y, X) = \text{Cov}(Y, X)/(\sqrt{\text{Var}(Y)}\sqrt{\text{Var}(X)})$. By the Cauchy–Schwarz inequality, the correlation coefficient satisfies $-1 \le \text{Corr}(Y, X) \le 1$. If Y and X are random vectors, the **correlation matrix** of Y and X is defined by

$$\text{Corr}(Y, X) = V_y \text{Cov}(Y, X) V_x,$$

where $V_y = \text{Diag}(1/\sqrt{\text{Var}(Y_1)}, \dots, 1/\sqrt{\text{Var}(Y_k)})$ and $V_x = \text{Diag}((1/\sqrt{\text{Var}(X_1)}, \dots, 1/\sqrt{\text{Var}(X_m)}))$.

If $g(X)$ is an estimator of the random vector X, the **mean squared error** (MSE) of $g(X)$ is defined as $\text{MSE}(g(X)) = E(X - g(X))(X - g(X))'$.

Given two random vectors, X and Y, the **conditional expectation** of X based on Y is the measurable random vector $E(X|Y)$ such that minimizes $\text{trMSE}(g(X))$ for all measurable random vectors $g(X)$, where tr denotes the trace of a matrix.

V. Gómez, *Multivariate Time Series With Linear State Space Structure*,
DOI 10.1007/978-3-319-28599-3_1

If c is a k-dimensional vector of constants and A is an $n \times k$ nonrandom matrix, the random vector $Z = AY + c$ has mean $E(Z) = AE(Y) + c$ and covariance matrix $\mathrm{Var}(Z) = A\mathrm{Var}(Y)A'$. In a similar way, the covariance matrix of the random vectors $Z = AY + c$ and $W = BX + d$ is $\mathrm{Cov}(Z, W) = A\mathrm{Cov}(Y, X)B' = A\Sigma_{yx}B'$.

The covariance matrix Σ_{yy} of a random vector Y is symmetric and nonnegative definite, since $\mathrm{Var}(b'Y) = b'\Sigma_{yy}b$ for all k-dimensional vector b . Thus, Σ_{yy} can be represented by means of the so-called **Cholesky decomposition** as $\Sigma_{yy} = LL'$, where L is a lower triangular matrix that is called Cholesky factor. Also, $\Sigma_{yy} = PDP'$, where P is orthogonal and D is a diagonal with the eigenvalues of Σ_{yy} in the diagonal. The Cholesky decomposition of Σ_{yy} can also be expressed as $\Sigma_{yy} = LDL'$, where L is lower triangular with ones in the diagonal and D is diagonal with nonnegative elements. Note that $\mathrm{Var}(L^{-1}Y) = D$. As we will see later, when we define orthogonality, this means that the elements of $L^{-1}Y$ are orthogonal.

Sometimes, we will be interested in the matrix of **noncentered second moments**, defined by $S_{yx} = E(XY')$.

A random variable X satisfying $E|X|^p < \infty$ is said to be of **class** p, denoted $X \in L_p$. A random vector X is said to be of **class** p if $X_i \in L_p$ for all components X_i of X.

Theorem 1.1 *If a random vector $X \in L_p$, then X has finite noncentered moments of order q for all $q \leq p$.*

Proof Let $X = (X_1', \ldots, X_k')'$ and let $a_i \geq 0, p_i \geq 0, i = 1, \ldots, k$, and $\sum_{i=1}^{k} p_i = 1$. Then, by the inequality between the geometric and the arithmetic mean, we can write

$$\prod_{i=1}^{k} a_i^{p_i} \leq \sum_{i=1}^{k} p_i a_i.$$

This implies

$$\prod_{i=1}^{k} |X_i|^{j_i} = \prod_{i=1}^{k} |X_i|^{q\frac{j_i}{q}} \leq \sum_{i=1}^{k} \frac{j_i}{q} |X_i|^q,$$

where $q = \sum_{i=1}^{k} j_i$. By Jensen's inequality, $Ef(X) \leq f(EX)$, which is true for all convex downward functions f, we can write

$$E \prod_{i=1}^{k} |X_i|^{j_i} \leq E \sum_{i=1}^{k} \frac{j_i}{q} |X_i|^q \leq \sum_{i=1}^{k} \frac{j_i}{q} (E|X_i|^p)^{\frac{q}{p}}.$$

□

1.2 Orthogonality

Two random variables, X and Y, with finite second moments are defined as **orthogonal** if $E(XY) = 0$. It can be verified that the product of random variables defined by $\langle X, Y\rangle = E(XY)$ has all the properties of an inner product. Strictly speaking, the condition $\langle X, X\rangle = 0$ implies $X = 0$ except on a set of zero measure. We can circumvent this difficulty by considering equivalence classes of random variables instead of random variables themselves, where X and Y are defined as equivalent if they differ on at most a set of zero measure. It can be shown that the set of equivalence classes of random variables with finite second moments is a Hilbert space. In what follows we will not distinguish between random variables and their equivalence classes.

Given two random vectors, X and Y, belonging to L_2, we define the product $\langle X, Y\rangle = E(XY')$, where the vectors can have different dimensions. This product also has all the properties of an inner product except that the values taken by the product are matrices and not real numbers. More specifically, the product $\langle , \rangle$ satisfies the following properties:

1. $\langle A_1X_1 + A_2X_2, Y\rangle = A_1\langle X_1, Y\rangle + A_2\langle X_2, Y\rangle$
2. $\langle X, Y\rangle = \langle Y, X\rangle^*$
3. $\langle X, X\rangle = 0$ implies $X = 0$,

where A_1 and A_2 are nonstochastic matrices of appropriate dimensions and X_1, X_2, X and Y are random vectors. Here if A is a matrix with complex elements, A^* denotes its conjugate transpose. Thus, the definition is also valid for complex random vectors. As in the scalar case, condition (3) is valid up to a set of zero measure and two random vectors are defined as equivalent if they differ on at most a set of zero measure. In what follows we will not distinguish between random vectors and their equivalence classes. Two random vectors, X and Y, are defined as **orthogonal** if $\langle X, Y\rangle = 0$. The orthogonality of X and Y is usually denoted as $X \perp Y$.

The notation

$$||X||^2 = \langle X, X\rangle$$

is often used. If X is a nonrandom scalar, then $||X||$ is its absolute value. Note that if X is a zero mean random vector, then

$$||X||^2 = \text{Var}(X).$$

The product $\langle , \rangle$, either for random vectors or for random variables, will be very useful in the sequel and we will refer to it as an inner product.

1.3 Best Linear Predictor and Orthogonal Projection

Given two random vectors, X and Y, with finite second moments, the **best linear predictor** $E^*(Y|X)$ of Y based on X is defined as the **linear estimator** AX such that $\mathrm{tr}E(Y-AX)(Y-AX)' = E(Y-AX)'(Y-AX)$ is minimum with respect to A, where A is a nonrandom matrix.

Proposition 1.1 *Suppose that X and Y are two random vectors with finite second moments,*

$$E\left\{\begin{bmatrix} X \\ Y \end{bmatrix}[X', Y']\right\} = \begin{bmatrix} S_{xx} & S_{xy} \\ S_{yx} & S_{yy} \end{bmatrix}, \tag{1.1}$$

and assume that S_{xx} is nonsingular. Then, the best linear predictor of Y based on X is $E^*(Y|X) = B'X$, where $B' = S_{yx}S_{xx}^{-1}$, and $MSE(E^*(Y|X)) = E(Y-B'X)(Y-B'X)' = S_{yy} - B'S_{xy} = S_{yy} - S_{yx}S_{xx}^{-1}S_{xy}$.

Proof It is not difficult to verify that

$$\begin{aligned} E(Y-AX)'(Y-AX) &= tr\{S_{yy} - S_{yx}A' - AS_{xy} + AS_{xx}A'\} \\ &= tr\{S_{yy} - B'S_{xx}A' - AS_{xx}B + AS_{xx}A'\} \\ &= tr\{S_{yy} - B'S_{xx}B + B'S_{xx}B - B'S_{xx}A' - AS_{xx}B + AS_{xx}A'\} \\ &= tr\{S_{yy} - B'S_{xx}B + (A-B')S_{xx}(A-B')'\}. \end{aligned}$$

Because the third term to the right of the last equality is nonnegative, the expression $E(Y-AX)'(Y-AX)$ is minimized if $A = B'$. This shows that the best linear predictor of Y based on X is $E^*(Y|X) = S_{yx}S_{xx}^{-1}X$.

In addition, because $\mathrm{tr}(Y-AX)(Y-AX)' = (Y-AX)'(Y-AX)$, the previous argument also shows that $\mathrm{MSE}(E^*(Y|X)) = S_{yy} - B'S_{xy} = S_{yy} - S_{yx}S_{xx}^{-1}S_{xy}$. □

Example 1.1 Let $Y = HX + V$, where H is a fixed matrix and X and V are random vectors such that $X \perp V$. Then, $S_{yx} = E(HX+V)X' = HS_{xx}$ and $E^*(Y|X) = HS_{xx}S_{xx}^{-1}X = HX$. On the other hand, $S_{yy} = HS_{xx}H' + S_{vv}$, so that $\mathrm{MSE}(E^*(Y|X)) = S_{yy} - S_{yx}S_{xx}^{-1}S_{xy} = S_{vv}$. ◇

Remark 1.1 The matrix B' of the best linear predictor of Y based on X, $E^*(Y|X) = B'X$, satisfies by Proposition 1.1 the so-called **normal equations**

$$B'S_{xx} = S_{yx}. \tag{1.2}$$

◇

Proposition 1.2 (The Case of Singular $\mathbf{S_{xx}}$) *Suppose that X and Y are as in Proposition 1.1 but that S_{xx} is singular. Then, the normal equations (1.2) are consistent and the solution is not unique. No matter which solution B' is used,* *the corresponding best linear predictor,* $E^*(Y|X) = B'X$*, is unique and so is* $MSE(E^*(Y|X)) = S_{yy} - B'S_{xy}$*. Moreover,* $B' = S_{yx}S_{xx}^-$*, where* S_{xx}^- *is a generalized inverse of* S_{xx}*, is a solution of the normal equations.*

Proof The normal equations can be written as $S_{xx}B = S_{xy}$. It is clear that for a solution to exist it is necessary and sufficient that S_{xy} be in the space generated by the columns of S_{xx}. Thus, the normal equations are consistent if, and only if, $\mathcal{R}(S_{xy}) \subset \mathcal{R}(S_{xx})$. By Corollary 1A.2 in the Appendix to this chapter, we can write

$$\mathcal{R}(S_{xy}) \oplus \mathcal{N}(S_{yx}) = \mathcal{R}(S_{xx}) \oplus \mathcal{N}(S_{xx}). \tag{1.3}$$

By Corollary 1A.1 in the Appendix to this chapter, the condition $\mathcal{R}(S_{xy}) \subset \mathcal{R}(S_{xx})$ can be expressed as $\mathcal{R}(S_{xx}) = \mathcal{R}(S_{xy}) \oplus U$, where U is the orthogonal complement of $\mathcal{R}(S_{xy})$ in $\mathcal{R}(S_{xx})$. It follows from this and (1.3) that the normal equations are consistent if, and only if,

$$\mathcal{R}(S_{xy}) \oplus \mathcal{N}(S_{yx}) = \mathcal{R}(S_{xy}) \oplus U \oplus \mathcal{N}(S_{xx}).$$

It is easily seen that this condition is equivalent to $\mathcal{N}(S_{xx}) \subset \mathcal{N}(S_{yx})$. On the other hand, if $a \in \mathcal{N}(S_{xx})$, then $a'S_{xx}a = 0$ and $a'X$ has zero variance. Assuming for simplicity that X and Y have zero means, this implies $a'X = 0$ and, therefore, $\text{Cov}(Y, a'X) = S_{yx}a = 0$ and $a \in \mathcal{N}(S_{yx})$. Thus, the condition $\mathcal{N}(S_{xx}) \subset \mathcal{N}(S_{yx})$ always holds and the normal equations are consistent.

To see that $E^*(Y|X)$ and $\text{MSE}(E^*(Y|X))$ are unique no matter what solution of the normal equations we use, let B_1' and B_2' be two solutions of the normal equations. Then, $(B_1' - B_2')S_{xx} = 0$ and, assuming for simplicity that X and Y have zero mean, this implies that $(B_1' - B_2')X$ has zero covariance matrix and, therefore, $(B_1' - B_2')X = 0$. In addition, since $\mathcal{N}(S_{xx}) \subset \mathcal{N}(S_{yx})$, we have $(B_1' - B_2')S_{xy} = 0$ and the equality of the MSE of the two best linear predictors follows.

If S_{xx}^- is a generalized inverse of S_{xx} and B' is any solution of the normal equations, then, using the properties of a generalized inverse, we can write

$$\left(S_{yx}S_{xx}^-\right) S_{xx} = B' \left(S_{xx}S_{xx}^- S_{xx}\right) = B'S_{xx} = S_{yx}.$$

Therefore, $S_{yx}S_{xx}^-$ is a solution of the normal equations. □

To define the best linear predictor we have only considered noncentered second moments. To take centered second moments into account, suppose that we want to obtain the best linear predictor of Y based on $(1, X')'$. Then, $E^*(Y|1, X) = A(1, X')' = A_1 + A_2X$, where $A = (A_1, A_2)$ is partitioned conformally to $(1, X')'$. The estimator $A_1 + A_2X$ is called an **affine estimator** instead of a linear estimator. For the rest of this section, let the matrix of noncentered second moments of $(1, X', Y')'$ be

$$E\left\{\begin{bmatrix}1\\X\\Y\end{bmatrix}\left[1, X', Y'\right]\right\} = \begin{bmatrix}1 & m_x' & m_y'\\ m_x & S_{xx} & S_{xy}\\ m_y & S_{yx} & S_{yy}\end{bmatrix}$$

and let

$$\mathrm{Var}\left\{\begin{bmatrix}X\\Y\end{bmatrix}\right\} = \begin{bmatrix}\Sigma_{xx} & \Sigma_{xy}\\ \Sigma_{yx} & \Sigma_{yy}\end{bmatrix}.$$

Proposition 1.3 *Assume that S_{xx} is nonsingular. Then, the best linear predictor of Y based on $(1, X')'$ is $E^*(Y|1, X) = B_1' + B_2'X = m_y + \Sigma_{yx}\Sigma_{xx}^{-1}(X - m_x)$ and $MSE(E^*(Y|1, X)) = \Sigma_{yy} - \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}$.*

Proof Using the result of Proposition 1.1, we can write

$$\left[B_1', B_2'\right] = \left[m_y, S_{yx}\right]\begin{bmatrix}1 & m_x'\\ m_x & S_{xx}\end{bmatrix}^{-1}.$$

Thus,

$$B_1' + B_2'm_x = m_y$$
$$B_1'm_x' + B_2'S_{xx} = S_{yx}.$$

Solving for B_1' in the first equation and substituting in the second one, it is obtained that $B_2'(S_{xx} - m_x m_x') = S_{yx} - m_y m_x'$. From this, we get $B_2' = \Sigma_{yx}\Sigma_{xx}^{-1}$ and $B_1' = m_y - \Sigma_{yx}\Sigma_{xx}^{-1}m_x$. Substituting in $\mathrm{MSE}(E^*(Y|1, X)) = E(Y - B_1' - B_2'X)(Y - B_1' - B_2'X)'$, the expression for the MSE follows easily. □

Remark 1.2 (Dispensing with the Means) The previous result shows that we can work with centered random vectors to compute the best linear predictor. To see this, let $\tilde{X} = X - m_x$ and $\tilde{Y} = Y - m_y$. Then, $E^*(\tilde{Y}|\tilde{X}) = \Sigma_{yx}\Sigma_{xx}^{-1}\tilde{X} = E^*(Y|1, X) - m_y$ and $\mathrm{MSE}(E^*(\tilde{Y}|\tilde{X})) = \mathrm{MSE}(E^*(Y|1, X))$. ◇

In the rest of the section, we will consider some properties of best linear prediction.

Proposition 1.4 *$B'X$ is the best linear predictor if and only if $Y - B'X$ is orthogonal to X.*

Proof If $B'X$ is the best linear predictor, $E(Y - B'X)X' = S_{yx} - B'S_{xx} = 0$. Conversely, if $E(Y - B'X)X' = 0$, then $S_{yx} - B'S_{xx} = 0$ and $B' = S_{yx}S_{xx}^{-1}$. □

The previous proposition shows that the best linear predictor $E^*(Y|X)$ coincides with the **orthogonal projection** of Y on X. This last projection is defined as the vector $B'X$ such that $Y - B'X$ is orthogonal to X.

Remark 1.3 (Projection Theorem for Random Variables) If Y is a random variable and $X = (X_1, \ldots, X_k)'$ is a random vector, the orthogonal projection of Y on X is the linear combination of the elements of X, $B'X = b_1X_1 + \cdots + b_kX_k$, such that $Y - B'X$ is orthogonal to X. Thus, the orthogonal projection is an element of the vector space, $S(X)$, generated by the components of X, that is, the space of all linear combinations of the form $a_1X_1 + \cdots + a_kX_k$, where $a_1, \ldots, a_k \in \mathbb{R}$. By Proposition 1.4, it is in fact the unique element of $S(X)$, $B'X$, such that

$$||Y - B'X|| \leq ||Y - A'X||, \qquad A'X \in S.$$

This is an instance of the so-called Projection Theorem in inner product spaces. ◇

In the rest of the book, whenever we consider an orthogonal projection of a random vector Y on another random vector of the form $(X_1', \ldots, X_k')'$, where the X_i, $i = 1, \ldots, k$ are in turn random vectors, we will write for simplicity and using a slight abuse of notation

$$E^*(Y|X_1, \ldots, X_k),$$

instead of

$$E^*\left[Y|(X_1', \ldots, X_k')'\right].$$

Proposition 1.5 *If $(Y', X_1', \ldots, X_k')'$ is a random vector such that $X_1, \ldots, X_k$ are mutually orthogonal, then*

$$E^*(Y|X_1, \ldots, X_k) = E^*(Y|X_1) + \cdots + E^*(Y|X_k).$$

Proof We can write

$$E^*(Y|X_1, \ldots, X_k) = (S_{yx_1}, \ldots, S_{yx_k})\mathrm{diag}(S_{x_ix_i}^{-1})\begin{bmatrix} X_1 \\ \vdots \\ X_k \end{bmatrix},$$

and the result follows. □

Proposition 1.6 *If $(Y', X_1', \ldots, X_k')'$ is a random vector such that $X_1, \ldots, X_k$ are mutually orthogonal, then*

$$E^*(Y|1, X_1, \ldots, X_k) = E^*(Y|1, X_1) + \cdots + E^*(Y|1, X_k) - (k-1)m_y.$$

Proof Considering centered variables, $\tilde{Y}, \tilde{X}_1, \ldots, \tilde{X}_k$, and applying the previous proposition, it is obtained that

$$\begin{aligned} E^*(\tilde{Y}|\tilde{X}_1, \ldots, \tilde{X}_k) &= E^*(Y|1, X_1, \ldots, X_k) - m_y \\ &= E^*(\tilde{Y}|\tilde{X}_1) + \cdots + E^*(\tilde{Y}|\tilde{X}_k) \\ &= E^*(Y|1, X_1) - m_y + \cdots + E^*(Y|1, X_k) - m_y. \end{aligned}$$

□

Proposition 1.7 *The orthogonal projection is invariant by nonsingular affine transformations. That is,*

$$E^*(Y|1, X) = E^*(Y|1, Z),$$

where

$$\begin{bmatrix} 1 \\ Z \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ n & M \end{bmatrix} \begin{bmatrix} 1 \\ X \end{bmatrix}$$

and $|M| \neq 0$.

Proof We can write

$$\begin{aligned} E^*(Y|1, Z) &= \left[m_y, S_{yx}\right] \begin{bmatrix} 1 & n' \\ 0 & M' \end{bmatrix} \left\{ \begin{bmatrix} 1 & 0 \\ n & M \end{bmatrix} \begin{bmatrix} 1 & m_x' \\ m_x & S_{xx} \end{bmatrix} \begin{bmatrix} 1 & n' \\ 0 & M' \end{bmatrix} \right\}^{-1} \\ &\quad \times \begin{bmatrix} 1 & 0 \\ n & M \end{bmatrix} \begin{bmatrix} 1 \\ X \end{bmatrix} \\ &= \left[m_y, S_{yx}\right] \begin{bmatrix} 1 & m_x' \\ m_x & S_{xx} \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ X \end{bmatrix} \\ &= E^*(Y|1, X). \end{aligned}$$

□

1.4 Orthogonalization of a Sequence of Random Vectors: Innovations

All random variables in this section will be assumed without loss of generality to have zero mean. Given a sequence of random vectors $\{Y_i\}$, $i = 1, \ldots, n$, that may be autocorrelated, we want to address the problem of constructing a new sequence $\{E_i\}$, $i = 1, \ldots, n$, of mutually uncorrelated random vectors such that each E_i depends linearly on Y_j, $j \leq i$. The random vectors E_i of the new sequence will be called **innovations**.

Given two random vectors, X and Y, with finite second moments given by (1.1) and assuming that S_{xx} is nonsingular, we know that, by Proposition 1.4, $Z = Y - E^*(Y|X) = Y - S_{yx}S_{xx}^{-1}X$ is orthogonal to X. Thus, if we make the transformation

$$\begin{bmatrix} X \\ Z \end{bmatrix} = \begin{bmatrix} I & 0 \\ -S_{yx}S_{xx}^{-1} & I \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix}, \tag{1.4}$$

then

$$\begin{aligned} E\left\{ \begin{bmatrix} X \\ Z \end{bmatrix} [X', Z'] \right\} &= \begin{bmatrix} I & 0 \\ -S_{yx}S_{xx}^{-1} & I \end{bmatrix} \begin{bmatrix} S_{xx} & S_{xy} \\ S_{yx} & S_{yy} \end{bmatrix} \begin{bmatrix} I & -S_{xx}^{-1}S_{xy} \\ 0 & I \end{bmatrix} \\ &= \begin{bmatrix} S_{xx} & 0 \\ 0 & S_{yy} - S_{yx}S_{xx}^{-1}S_{xy} \end{bmatrix}. \end{aligned} \tag{1.5}$$

In this case, $\{X, Z\} = \{E_1, E_2\}$ is the desired sequence of innovations and $\{D_1, D_2\} = \{S_{xx}, S_{yy} - S_{yx}S_{xx}^{-1}S_{xy}\}$ is the sequence of their covariance matrices.

Suppose now a sequence of three random vectors $Y = (Y_1', Y_2', Y_3')'$, where $||Y||^2$ is positive definite. Then, to construct three mutually orthogonal random vectors, $E = (E_1', E_2', E_3')'$, where $||E_i||^2 = D_i$, such that each E_i depends linearly on Y_j, $j \leq i$, $i = 1, 2, 3$, we can apply the well-known **Gram–Schmidt** orthogonalization procedure of Euclidean spaces to the sequence Y.

The procedure is recursive. Define first $E_1 = Y_1$ and $D_1 = ||Y_1||^2$. Then, define

$$E_2 = Y_2 - \Theta_{21}E_1$$

such that E_2 is orthogonal to E_1. This implies

$$0 = \langle Y_2, E_1 \rangle - \Theta_{21}||E_1||^2$$

and, thus,

$$\Theta_{21} = \langle Y_2, Y_1 \rangle D_1^{-1}, \qquad D_2 = \langle Y_2, Y_2 \rangle - \Theta_{21}D_1\Theta_{21}'.$$

Note that $\Theta_{21}E_1 = E^*(Y_2|E_1)$ and $E_2 = Y_2 - E^*(Y_2|E_1)$. Finally, define

$$E_3 = Y_3 - \Theta_{32}E_2 - \Theta_{31}E_1$$

such that E_3 is orthogonal to E_2 and E_1. To obtain Θ_{32} and Θ_{31}, we impose the orthogonality conditions. Therefore,

$$\begin{aligned}
0 &= \langle Y_3, E_1\rangle - \Theta_{31}||E_1||^2\\
0 &= \langle Y_3, E_2\rangle - \Theta_{32}||E_2||^2\\
&= \langle Y_3, Y_2 - \Theta_{21}E_1\rangle - \Theta_{32}||E_2||^2\\
&= \langle Y_3, Y_2\rangle - \langle \Theta_{31}E_1, \Theta_{21}E_1\rangle - \Theta_{32}||E_2||^2,
\end{aligned}$$

where the last equality holds because $Y_3 = E_3 + \Theta_{32}E_2 + \Theta_{31}E_1$ and E_3 and E_2 are orthogonal to E_1. From this, it is obtained that

$$\begin{aligned}
\Theta_{31} &= \langle Y_3, Y_1\rangle D_1^{-1}\\
\Theta_{32} &= (\langle Y_3, Y_2\rangle - \Theta_{31}D_1\Theta'_{21})D_2^{-1}\\
D_3 &= \langle Y_3, Y_3\rangle - \Theta_{32}D_2\Theta'_{32} - \Theta_{31}D_1\Theta'_{31}
\end{aligned}$$

Note that $\Theta_{31}E_1 = E^*(Y_3|E_1)$, $\Theta_{32}E_2 = E^*(Y_3|E_2)$ and, by Proposition 1.5, $E_3 = Y_3 - E^*(Y_3|E_1, E_2)$.

The previous algorithm implies the transformation

$$\begin{bmatrix} Y_1\\ Y_2\\ Y_3 \end{bmatrix} = \begin{bmatrix} I & 0 & 0\\ \Theta_{21} & I & 0\\ \Theta_{31} & \Theta_{32} & I \end{bmatrix} \begin{bmatrix} E_1\\ E_2\\ E_3 \end{bmatrix}. \tag{1.6}$$

Given that the variables in E are mutually orthogonal, this transformation produces a block Cholesky decomposition of the matrix $||Y||^2$,

$$||Y||^2 = \begin{bmatrix} I & 0 & 0\\ \Theta_{21} & I & 0\\ \Theta_{31} & \Theta_{32} & I \end{bmatrix} \begin{bmatrix} D_1 & 0 & 0\\ 0 & D_2 & 0\\ 0 & 0 & D_3 \end{bmatrix} \begin{bmatrix} I & 0 & 0\\ \Theta_{21} & I & 0\\ \Theta_{31} & \Theta_{32} & I \end{bmatrix}'.$$

The procedure just described to orthogonalize three random vectors can be generalized to any number of random vectors. We summarize the result in the following theorem.

Theorem 1.2 (Innovations Algorithm) *Let the sequence of random vectors* $Y = (Y'_1, \ldots, Y'_n)'$, *where* $||Y||^2$ *is positive definite and* $S_{ij} = E(Y_iY'_j)$, $i, j = 1, \ldots, n$. *Then, for* $t = 2, \ldots, n$,

$$\begin{aligned}
Y_t &= \Theta_{t1}E_1 + \Theta_{t2}E_2 + \cdots + \Theta_{t,t-1}E_{t-1} + E_t\\
&= E^*(Y_t|E_{t-1}, \ldots, E_1) + E_t,
\end{aligned}$$

where the E_i vectors, $i = 1, \ldots, t$, are mutually orthogonal and the coefficients Θ_{tj}, $j = 1, \ldots, t-1$, and $D_t = ||E_t||^2$, can be obtained from the recursions

$$\Theta_{tj} = \left(S_{tj} - \sum_{i=1}^{j-1} \Theta_{ti} D_i \Theta'_{ji} \right) D_j^{-1}, \quad j = 1, \ldots, t-1,$$

$$D_t = S_{tt} - \sum_{i=1}^{t-1} \Theta_{ti} D_i \Theta'_{ti},$$

initialized with $Y_1 = E_1$ and $D_1 = S_{11}$.

Proof The orthogonality conditions imply

$$\begin{aligned} \Theta_{tj} ||E_j||^2 &= \langle Y_t, E_j \rangle \\ &= \left\langle Y_t, Y_j - \sum_{i=1}^{j-1} \Theta_{ji} E_i \right\rangle \\ &= S_{tj} - \sum_{i=1}^{j-1} \Theta_{ti} D_i \Theta'_{ji}, \end{aligned}$$

where we have made use of the fact that $E(Y_t E'_i) = E(\Theta_{ti} E_i E'_i) = \Theta_{ti} D_i$. The expression for D_t is proved analogously. The equality $E^*(Y_t | E_{t-1}, \ldots, E_1) = \Theta_{t1} E_1 + \cdots + \Theta_{t,t-1} E_{t-1}$ follows from Proposition 1.5. □

The innovations algorithm implies the transformation

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} I & 0 & \cdots & 0 \\ \Theta_{21} & I & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \Theta_{n1} & \Theta_{n2} & \cdots & I \end{bmatrix} \begin{bmatrix} E_1 \\ E_2 \\ \vdots \\ E_n \end{bmatrix}, \tag{1.7}$$

which constitutes a generalization of (1.6), and the corresponding block Cholesky decomposition of $||Y||^2$,

$$||Y||^2 = \begin{bmatrix} I & 0 & \cdots & 0 \\ \Theta_{21} & I & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \Theta_{n1} & \Theta_{n2} & \cdots & I \end{bmatrix} \begin{bmatrix} D_1 & 0 & \cdots & 0 \\ 0 & D_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & D_n \end{bmatrix} \begin{bmatrix} I & 0 & \cdots & 0 \\ \Theta_{21} & I & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \Theta_{n1} & \Theta_{n2} & \cdots & I \end{bmatrix}'. \tag{1.8}$$

It is shown in the proof of the innovations algorithm that the lower triangular matrix in (1.7) can also be written as

$$\begin{bmatrix} I & & & & \\ \langle Y_2, E_1\rangle ||E_1||^{-2} & I & & & \\ \langle Y_3, E_1\rangle ||E_1||^{-2} & \langle Y_3, E_2\rangle ||E_2||^{-2} & I & & \\ \vdots & \vdots & \vdots & \ddots & \\ \langle Y_n, E_1\rangle ||E_1||^{-2} & \langle Y_n, E_2\rangle ||E_2||^{-2} & \langle Y_n, E_3\rangle ||E_3||^{-2} & \cdots & I \end{bmatrix}.$$

Example 1.2 Let Y_1 be a random variable with $\mathrm{E}(Y_1) = 0$ and $\mathrm{Var}(Y_1) = (1+\theta^2)\sigma^2$ and let $Y_t = \theta A_{t-1} + A_t$, $t \geq 2$, where $|\theta| < 1$ and $\{A_t\}$ is an uncorrelated sequence of random variables with zero mean and common variance σ^2 such that $E(Y_1 A_1) = \sigma^2$ and $E(Y_1 A_t) = 0$, $t \geq 2$. Given the sequence $Y = (Y_1, \ldots, Y_n)'$, we can obtain the covariances $S_{ij} = E(Y_i Y_j')$, $i, j = 1, \ldots, n$, as follows. Clearly, $S_{tt} = (1+\theta^2)\sigma^2$, $t \geq 1$. To obtain $S_{t,t-1}$, $t \geq 2$, multiply the equation $Y_t = \theta A_{t-1} + A_t$ by Y_{t-1} to get

$$Y_t Y_{t-1} = \theta A_{t-1} Y_{t-1} + A_t Y_{t-1}.$$

Taking expectations and considering that A_t is uncorrelated with Y_{t-1} and $E(A_{t-1} Y_{t-1}) = \sigma^2$ for $t \geq 2$, we get $S_{t,t-1} = \sigma^2\theta$. Similarly, $S_{t,t-j} = 0$, $j > 1$, $t - j \geq 1$. Thus,

$$\mathrm{Var}(Y) = \sigma^2 \begin{bmatrix} 1+\theta^2 & \theta & & & \\ \theta & 1+\theta^2 & \theta & & \\ & \ddots & \ddots & \ddots & \\ & & \theta & 1+\theta^2 & \theta \\ & & & \theta & 1+\theta^2 \end{bmatrix}.$$

Applying the innovations algorithm, we have $E_1 = Y_1$, $D_1 = (1+\theta^2)\sigma^2$, and for $t \geq 2$

$$\begin{aligned} \Theta_{t,t-1} D_{t-1} &= S_{t,t-1} &&= \theta\sigma^2 \\ D_t &= S_{tt} - \Theta_{t,t-1}^2 D_{t-1} &&= (1+\theta^2)\sigma^2 - \theta^2\sigma^4/D_{t-1}. \end{aligned}$$

This implies for $t \geq 2$

$$Y_t = \Theta_{t,t-1} E_{t-1} + E_t, \quad \Theta_{t,t-1} = \frac{\theta(1+\theta^2+\cdots+\theta^{2(t-2)})}{1+\theta^2+\cdots+\theta^{2(t-1)}},$$

$$D_t = \frac{1+\theta^2+\cdots+\theta^{2t}}{1+\theta^2+\cdots+\theta^{2(t-1)}}\sigma^2.$$

The decomposition (1.8) is in this case $\mathrm{Var}(Y) = LDL'$, with

$$LD = \begin{bmatrix} 1 & & & \\ \Theta_{21} & 1 & & \\ & \ddots & \ddots & \\ & & \Theta_{n,n-1} & 1 \end{bmatrix} \begin{bmatrix} D_1 & & & \\ & D_2 & & \\ & & \ddots & \\ & & & D_n \end{bmatrix}.$$

◇

Example 1.3 Let Y_1 be a random variable with $\mathrm{E}(Y_1) = 0$ and $\mathrm{Var}(Y_1) = \sigma^2/(1-\phi^2)$ and let $Y_t = \phi Y_{t-1} + A_t$, $t \geq 2$, where $|\phi| < 1$ and $\{A_t\}$ is an uncorrelated sequence of random variables with zero mean and common variance σ^2, uncorrelated with Y_1. Given the sequence $Y = (Y_1, \ldots, Y_n)'$, we can obtain the covariances $S_{ij} = E(Y_i Y_j')$, $i, j = 1, \ldots, n$, as follows. Clearly, $S_{11} = \sigma^2/(1-\phi^2)$. To obtain $S_{t,t-1}$, multiply the equation $Y_t = \phi Y_{t-1} + A_t$ by Y_{t-1} to get

$$Y_t Y_{t-1} = \phi Y_{t-1}^2 + A_t Y_{t-1}.$$

Taking expectations and considering that Y_{t-1} is uncorrelated with A_t yields $S_{t,t-1} = \phi S_{tt}$. To obtain $S_{t,t-2}$, replace first Y_{t-1} in $Y_t = \phi Y_{t-1} + A_t$ with $Y_{t-1} = \phi Y_{t-2} + A_{t-1}$ to get

$$Y_t = \phi^2 Y_{t-2} + \phi A_{t-1} + A_t.$$

Then, multiplying by Y_{t-2} and taking expectations, we have $S_{t,t-2} = \phi^2 S_{tt}$. Reiterating this procedure we get $S_{t,t-k} = \phi^k S_{tt}$. Finally, to obtain S_{tt}, we proceed by induction on the hypothesis that $S_{tt} = \sigma^2/(1-\phi^2)$. It is evidently true for $t = 1$. Assuming it is true for $t-1$, taking expectations in $Y_t^2 = (\phi Y_{t-1} + A_t)^2$ and considering that Y_{t-1} is uncorrelated with A_t, we get $S_{tt} = \phi^2 S_{t-1,t-1} + \sigma^2 = \sigma^2/(1-\phi^2)$. Thus,

$$\mathrm{Var}(Y) = \frac{\sigma^2}{1-\phi^2} \begin{bmatrix} 1 & \phi & \phi^2 & \cdots & \phi^{n-1} \\ \phi & 1 & \phi & \cdots & \phi^{n-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \phi^{n-1} & \phi^{n-2} & \phi^{n-3} & \cdots & 1 \end{bmatrix}.$$

The innovations algorithm yields $E_1 = Y_1$, $D_1 = \sigma^2/(1-\phi^2)$,

$$Y_2 = \Theta_{21} E_1 + E_2, \qquad \Theta_{21} = \phi, \qquad D_2 = \sigma^2,$$

and for $t \geq 2$,

$$Y_t = \phi^{t-1} E_1 + \phi^{t-2} E_2 + \cdots + \phi E_{t-1} + E_t, \qquad D_t = \sigma^2.$$

Therefore, the decomposition (1.8) is in this case $\text{Var}(Y) = LDL'$, with

$$LD = \begin{bmatrix} 1 & & & & \\ \phi & 1 & & & \\ \phi^2 & \phi & 1 & & \\ \vdots & \vdots & \vdots & \ddots & \\ \phi^{n-1} & \phi^{n-2} & \cdots & \phi & 1 \end{bmatrix} \begin{bmatrix} \sigma^2/(1-\phi^2) & & & & \\ & \sigma^2 & & & \\ & & \sigma^2 & & \\ & & & \ddots & \\ & & & & \sigma^2 \end{bmatrix}.$$

◇

From (1.6), it is easy to recursively derive the following expressions that give the innovations E_t in terms of the observations Y_s, $s \leq t$, $t = 1, 2, 3$,

$$\begin{aligned} E_1 &= Y_1 \\ E_2 &= -\Theta_{21} Y_1 + Y_2 \\ E_3 &= -(\Theta_{31} - \Theta_{32}\Theta_{21}) Y_1 - \Theta_{32} Y_2 + Y_3. \end{aligned} \tag{1.9}$$

A similar transformation holds in the general setting. To see this, consider that the innovations algorithm allows us to express the random vector Y_t in terms of the orthogonal vectors E_s, $s = 1, \ldots, t-1$. Since these last vectors are linear combinations of the Y_j vectors, $j \leq s$, we can write $Y_t = \Pi_{t1} Y_1 + \cdots + \Pi_{t,t-1} Y_{t-1} + E_t$, where the Π_i depend on the Θ_i. This last transformation amounts to premultiply (1.7) by a suitable block lower triangular matrix to get

$$\begin{bmatrix} I & 0 & \cdots & 0 \\ -\Pi_{21} & I & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -\Pi_{n1} & -\Pi_{n2} & \cdots & I \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} E_1 \\ E_2 \\ \vdots \\ E_n \end{bmatrix}. \tag{1.10}$$

Comparing (1.10) with (1.7), the following relation is obtained

$$\begin{bmatrix} I & 0 & \cdots & 0 \\ -\Pi_{21} & I & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -\Pi_{n1} & -\Pi_{n2} & \cdots & I \end{bmatrix} \begin{bmatrix} I & 0 & \cdots & 0 \\ \Theta_{21} & I & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \Theta_{n1} & \Theta_{n2} & \cdots & I \end{bmatrix} = \begin{bmatrix} I & 0 & \cdots & 0 \\ 0 & I & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & I \end{bmatrix}. \tag{1.11}$$

Thus, the Π_{ti} weights can be obtained recursively from the Θ_{ti} ones. We will state this result in a corollary.

Corollary 1.1 *Let the sequence of random vectors* $Y = (Y'_1, \dots, Y'_n)'$, *where* $||Y||^2$ *is positive definite, and assume we apply the innovations algorithm. Then, for* $t = 2, \dots, n$,

$$\begin{aligned} Y_t &= E^*(Y_t|Y_1, \dots, Y_{t-1}) + E_t \\ &= \Pi_{t1}Y_1 + \Pi_{t2}Y_2 + \cdots + \Pi_{t,t-1}Y_{t-1} + E_t, \end{aligned}$$

where the E_t *vectors,* $t = 2, \dots, n$, *are the innovations and the coefficients* Π_{ti}, $i = 1, \dots, t-1$, *can be obtained recursively by solving*

$$\begin{bmatrix} -\Pi_{t1} & -\Pi_{t2} & \cdots & I \end{bmatrix} \begin{bmatrix} I & 0 & \cdots & 0 \\ \Theta_{21} & I & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \Theta_{t1} & \Theta_{t2} & \cdots & I \end{bmatrix} = \begin{bmatrix} 0 & 0 & \cdots & I \end{bmatrix}.$$

Proof The only thing we have to prove is $Y_t = E^*(Y_t|Y^{t-1}) + E_t$. Let $Y^{t-1} = (Y'_1, \dots, Y'_{t-1})'$ and $E^{t-1} = (E'_1, \dots, E'_{t-1})'$. Since the transformation (1.7), where $n = t - 1$, is nonsingular, by Propositions 1.7 and 1.5, the equality $E^*(Y_t|Y^{t-1}) = E^*(Y_t|E^{t-1})$ holds. By the innovations algorithm, $E^*(Y_t|E^{t-1}) = \sum_{i=1}^{t-1} \Theta_{ti}E_i$ and $E_t = Y_t - E^*(Y_t|E^{t-1})$ is orthogonal to $E^*(Y_t|E^{t-1}) = E^*(Y_t|Y^{t-1})$. Thus, $Y_t = E^*(Y_t|Y^{t-1}) + E_t = \sum_{i=1}^{t-1} \Pi_{ti}Y_i + E_t$. □

Inverting both sides of (1.8) and considering (1.11), it is obtained that

$$||Y||^{-2} = \begin{bmatrix} I & 0 & \cdots & 0 \\ -\Pi_{21} & I & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -\Pi_{n1} & -\Pi_{n2} & \cdots & I \end{bmatrix}' \begin{bmatrix} D_1^{-1} & 0 & \cdots & 0 \\ 0 & D_2^{-1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & D_n^{-1} \end{bmatrix} \begin{bmatrix} I & 0 & \cdots & 0 \\ -\Pi_{21} & I & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -\Pi_{n1} & -\Pi_{n2} & \cdots & I \end{bmatrix}. \tag{1.12}$$

The expression (1.7) can be written more compactly as

$$Y = LE,$$

where $Y = (Y'_1, \dots, Y'_n)'$ is the vector of observations, $E = (E'_1, \dots, E'_n)'$ is the vector of innovations and

$$L = \begin{bmatrix} I & 0 & \cdots & 0 \\ \Theta_{21} & I & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \Theta_{n1} & \Theta_{n2} & \cdots & I \end{bmatrix}. \tag{1.13}$$

Letting $S = ||Y||^2$ and $D = \text{diag}(D_1, \ldots, D_n)$, the decomposition (1.8) of S can be written as

$$S = LDL', \tag{1.14}$$

and letting

$$W = L^{-1} = \begin{bmatrix} I & 0 & \cdots & 0 \\ -\Pi_{21} & I & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -\Pi_{n1} & -\Pi_{n2} & \cdots & I \end{bmatrix}, \tag{1.15}$$

the decomposition (1.12) becomes

$$S^{-1} = W'D^{-1}W. \tag{1.16}$$

The decomposition $S = LDL'$ is similar to the Cholesky decomposition of a symmetric matrix. Note that if $D^{1/2}$ denotes a matrix such that $D = D^{1/2}D^{1/2'}$ and we define $\bar{L} = LD^{1/2}$, the matrix $\bar{L}$ is lower triangular and

$$S = \bar{L}\bar{L}',$$

so that $\bar{L}$ is a Cholesky factor of S. If we define $\bar{E} = D^{-1/2}E$, we have $||\bar{E}||^2 = I$, so that this vector constitutes a vector of **standardized innovations** and $Y = \bar{L}\bar{E}$.

The decompositions (1.14) and (1.16) can be referred to as "**lower–upper**" and "**upper–lower**" triangularizations of symmetric, positive definite matrices. Every symmetric, positive definite matrix admits both a unique lower–upper and a unique upper–lower triangularization.

Letting

$$Y^i = W'D^{-1}E,$$

it is easy to see that the covariance matrix of Y^i is the inverse of the covariance matrix of Y, that is $\text{Var}(Y^i) = V^{-1}$. For this reason, $Y^i = (Y_1^{i'}, \ldots, Y_n^{i'})'$ is called the **the inverse process** of Y.

The following theorem shows that the Π_{ti} and D_t matrices in (1.10) and (1.12) can be recursively computed.

Theorem 1.3 (Autoregressive Representation) *Let the sequence of random vectors $Y = (Y_1', \ldots, Y_n')'$, where $||Y||^2$ is positive definite and $S_{ij} = E(Y_iY_j')$, $i, j = 1, \ldots, n$. Then, for $t = 2, \ldots, n$,*

$$\begin{aligned} Y_t &= E^*(Y_t|Y_1, \ldots, Y_{t-1}) + E_t \\ &= \Pi_{t1}Y_1 + \Pi_{t2}Y_2 + \cdots + \Pi_{t,t-1}Y_{t-1} + E_t, \end{aligned}$$

where the E_s vectors, $s = 1, \ldots, t$, are mutually orthogonal and the coefficients Π_{ti}, $i = 1, \ldots, t-1$, and matrix $D_t = ||E_t||^2$, can be obtained from the recursions

$$C_{tj} = S_{tj} - \sum_{i=1}^{j-1} S_{ti}\Pi'_{ji}, \quad j = 1, \ldots, t-1,$$

$$\Pi_{tj} = C_{tj}D_j^{-1} - \sum_{i=j+1}^{t-1} C_{ti}D_i^{-1}\Pi_{ij}, \quad j = 1, \ldots, t-1,$$

$$D_t = S_{tt} - \sum_{i=1}^{t-1} C_{ti}D_i^{-1}C'_{ti},$$

initialized with $Y_1 = E_1$ and $D_1 = S_{11}$.

Proof Let $Y^{t-1} = (Y'_1, \ldots, Y'_{t-1})'$, $E^{t-1} = (E'_1, \ldots, E'_{t-1})'$, $S_t^{t-1} = (S_{t1}, \ldots, S_{t,t-1})$, $\Pi_t^{t-1} = (\Pi_{t1}, \ldots, \Pi_{t,t-1})$ and

$$\Pi_{t-1} = \begin{bmatrix} I & 0 & \cdots & 0 & 0 \\ -\Pi_{21} & I & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ -\Pi_{t-1,1} & -\Pi_{t-1,2} & \cdots & -\Pi_{t-1,t-2} & I \end{bmatrix}.$$

We have already proved in Corollary 1.1 that $Y_t = E^*(Y_t|Y^{t-1}) + E_t = \Pi_t^{t-1}Y^{t-1} + E_t$. Thus, by Proposition 1.1, $\Pi_t^{t-1} = S_t^{t-1}||Y^{t-1}||^{-2}$ and $D_t = S_{tt} - S_t^{t-1}||Y^{t-1}||^{-2}S_t^{t-1\prime}$. Substituting (1.12) in the previous expressions, it is obtained that

$$\Pi_t^{t-1} = S_t^{t-1}\Pi'_{t-1}D^{t-1}\Pi_{t-1}, \qquad D_t = S_{tt} - S_t^{t-1}\Pi'_{t-1}D^{t-1}\Pi_{t-1}S_t^{t-1\prime},$$

where $D^{t-1} = \mathrm{Diag}(D_1^{-1}, \ldots, D_{t-1}^{-1})$.

Define $C_t^{t-1} = (C_{t1}, \ldots, C_{t,t-1}) = S_t^{t-1}\Pi'_{t-1}$. Then, $\Pi_t^{t-1} = C_t^{t-1}D^{t-1}\Pi_{t-1}$ and $D_t = S_{tt} - C_t^{t-1}D^{t-1}C_t^{t-1\prime}$, and the theorem follows. □

The following corollary is sometimes useful. It is an immediate consequence of Corollary 1.1.

Corollary 1.2 *The coefficient $\Pi_{t,t-1}$ of Y_{t-1} given by the autoregressive representation coincides with the coefficient $\Theta_{t,t-1}$ of E_{t-1} given by the innovations algorithm.*

Remark 1.4 The autoregressive representation allows for the computation of the Π_{ni}, $i = 1, \ldots, n-1$, and D_n matrices recursively, without having to invert matrices of big size, as would be the case if we applied the best linear predictor formulae to compute $E^*(Y_n|Y_1, \ldots, Y_{n-1})$ and its MSE. ◇

Example 1.2 (Continued) Given that $S_{tt} = \sigma^2(1+\theta^2)$ if $t \geq 1$ and $S_{t,t-1} = \sigma^2\theta$ if $t \geq 1$, the autoregressive representation yields $D_1 = S_{11}$, $Y_1 = E_1$,

$$\begin{aligned} C_{21} &= S_{21} \\ \Pi_{21} &= C_{21}D_1^{-1} = \theta/(1+\theta^2) \\ D_2 &= S_{22} - C_{21}D_1^{-1}C_{21} = \sigma^2(1+\theta^2+\theta^4)/(1+\theta^2), \end{aligned}$$

and $Y_2 = \Pi_{21}Y_1 + E_2$. Continuing in this way, it is obtained for $t \geq 2$ that $Y_t = \Pi_{t1}Y_1 + \Pi_{t2}Y_2 + \cdots + \Pi_{t,t-1}Y_{t-1} + E_t$, where

$$D_t = \frac{1+\theta^2+\cdots+\theta^{2t}}{1+\theta^2+\cdots+\theta^{2(t-1)}}\sigma^2, \quad \Pi_{tj} = (-1)^{t-j+1}\theta^{t-j}\frac{1+\theta^2+\cdots+\theta^{2(j-1)}}{1+\theta^2+\cdots+\theta^{2(t-1)}},$$
$$j = 1,\ldots,t-1.$$

◇

Example 1.3 (Continued) In this case, $S_{tt} = \sigma^2/(1-\phi^2)$ and $S_{t,t-k} = \phi^k S_{tt}$ if $t \geq 1$, $k = 1,\ldots,t-1$. Thus, the autoregressive representation yields $D_1 = S_{11}$, $Y_1 = E_1$,

$$\begin{aligned} C_{21} &= S_{21} \\ \Pi_{21} &= C_{21}D_1^{-1} = \phi \\ D_2 &= S_{22} - C_{21}D_1^{-1}C_{21} = \sigma^2, \end{aligned}$$

$Y_2 = \Pi_{21}Y_1 + E_2$, and for $t \geq 2$, $Y_t = \Pi_{t,t-1}Y_{t-1} + E_t$, where $\Pi_{t,t-1} = \phi$ and $D_t = \sigma^2$. Thus, the matrices L^{-1} and D^{-1} in (1.15) and (1.16) are

$$L^{-1} = \begin{bmatrix} 1 & & & & \\ -\phi & 1 & & & \\ & -\phi & 1 & & \\ & & \ddots & \ddots & \\ & & & -\phi & 1 \end{bmatrix}, \; D^{-1} = \begin{bmatrix} (1-\phi^2)/\sigma^2 & & & & \\ & 1/\sigma^2 & & & \\ & & 1/\sigma^2 & & \\ & & & \ddots & \\ & & & & 1/\sigma^2 \end{bmatrix}. \tag{1.17}$$

◇

1.5 The Modified Innovations Algorithm

The innovations algorithm uses the Gram–Schmidt procedure to orthogonalize a sequence of random vectors. There is, however, an alternative that is more numerically stable, the so-called modified Gram–Schmidt algorithm. This algorithm is as follows. Suppose the sequence of random vectors $Y = (Y_1' \dots, Y_n')'$ and let $S_{ij} = E(Y_i Y_j')$, $i,j = 1,\dots,n$. Then, if $E = (E_1', E_2', \dots, E_n')'$ denotes the innovations sequence, the modified Gram–Schmidt algorithm consists of the following steps.

(a) Initialize with $E_1 = Y_1$ and $D_1 = ||E_1||^2$.
(b) Form $\widetilde{Y}_{j|1} = Y_j - E^*(Y_j|E_1), j = 2,\dots,n$, and $S_{ij.1} = E(\widetilde{Y}_{i|1}\widetilde{Y}'_{j|1}), i,j = 2,\dots,n$, $i \geq j$. Set $E_2 = \widetilde{Y}_{2|1}$ and $D_2 = ||E_2||^2$.
(c) Form $\widetilde{Y}_{j|2} = \widetilde{Y}_{j|1} - E^*(\widetilde{Y}_{j|1}|E_2)$, $j = 3,\dots,n$, and $S_{ij.2} = E(\widetilde{Y}_{i|2}\widetilde{Y}'_{j|2})$, $i,j = 3,\dots,n$, $i \geq j$. Set $E_3 = \widetilde{Y}_{3|2}$ and $D_3 = ||E_3||^2$,

and so on. We can arrange the Y_t and the generated $\widetilde{Y}_{t|j}$ in a triangular array such that the diagonal entries are the innovations, E_t, in the following way.

$$\begin{array}{lllll} Y_1 \\ Y_2 & \widetilde{Y}_{2|1} \\ Y_3 & \widetilde{Y}_{3|1} & \widetilde{Y}_{3|2} \\ \vdots & \vdots & \vdots & \ddots \\ Y_n & \widetilde{Y}_{n|1} & \widetilde{Y}_{n|2} & \cdots & \widetilde{Y}_{n|n-1}. \end{array} \quad (1.18)$$

The procedure can be seen as a series of transformations applied to the columns of the triangular array (1.18). The first transformation can be described as premultiplying Y with a lower triangular matrix so that the following equality holds

$$\begin{bmatrix} I \\ -\langle Y_2, E_1\rangle ||E_1||^{-2} & I \\ -\langle Y_3, E_1\rangle ||E_1||^{-2} & 0 & I \\ \vdots & \vdots & \vdots & \ddots \\ -\langle Y_n, E_1\rangle ||E_1||^{-2} & 0 & 0 & \cdots & I \end{bmatrix} \begin{bmatrix} E_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} E_1 \\ \widetilde{Y}_{2|1} \\ \widetilde{Y}_{3|1} \\ \vdots \\ \widetilde{Y}_{n|1} \end{bmatrix}. \quad (1.19)$$

Let T_1 be the lower triangular matrix on the left-hand side of (1.19) and $\widetilde{Y}_{|1} = (\widetilde{Y}'_{2|1}, \widetilde{Y}'_{3|1}, \dots, \widetilde{Y}'_{n|1})'$. Partition $||Y||^2$ as

$$||Y||^2 = \begin{bmatrix} ||E_1||^2 & m_1' \\ m_1 & M_1 \end{bmatrix}.$$

Then, it is not difficult to verify (see Problem 1.3) using the properties of orthogonal projection that the following equality holds

$$T_1 ||Y||^2 T_1' = \begin{bmatrix} ||E_1||^2 & 0 \\ 0 & M_1 - m_1 ||E_1||^{-2} m_1' \end{bmatrix}$$

and, therefore, $||\widetilde{Y}_{|1}||^2 = M_1 - m_1||E_1||^{-2}m_1'$. Thus, the effect of the first transformation on $||Y||^2$ is to make zeros both in the first block column and in the first block row after the first block entry. Also, $S_{ij.1} = S_{ij} - S_{i1}D_1^{-1}S_{j1}'$, $i,j = 2,\ldots,n$, and $D_1 = S_{11}$.

The second transformation consists of premultiplying the second column of the triangular array (1.18) with a lower triangular matrix so that the following equality holds

$$\begin{bmatrix} I & & & & \\ -\langle\widetilde{Y}_{3|1},E_2\rangle||E_2||^{-2} & I & & & \\ -\langle\widetilde{Y}_{3|1},E_2\rangle||E_2||^{-2} & 0 & I & & \\ \vdots & \vdots & \vdots & \ddots & \\ -\langle\widetilde{Y}_{n|1},E_2\rangle||E_2||^{-2} & 0 & 0 & \cdots & I \end{bmatrix} \begin{bmatrix} E_2 \\ \widetilde{Y}_{3|1} \\ \widetilde{Y}_{4|1} \\ \vdots \\ \widetilde{Y}_{n|1} \end{bmatrix} = \begin{bmatrix} E_2 \\ \widetilde{Y}_{3|2} \\ \widetilde{Y}_{4|2} \\ \vdots \\ \widetilde{Y}_{n|2} \end{bmatrix}. \tag{1.20}$$

As with the first transformation, let T_2 be the lower triangular matrix on the left-hand side of (1.20) and $\widetilde{Y}_{|2} = (\widetilde{Y}_{3|2}', \widetilde{Y}_{4|2}', \ldots, \widetilde{Y}_{n|2}')'$. Partition $||\widetilde{Y}_{|1}||^2$ as

$$||\widetilde{Y}_{|1}||^2 = \begin{bmatrix} ||E_2||^2 & m_2' \\ m_2 & M_2 \end{bmatrix}.$$

Then,

$$T_2||\widetilde{Y}_{|1}||^2T_2' = \begin{bmatrix} ||E_2||^2 & 0 \\ 0 & M_2 - m_2||E_2||^{-2}m_2' \end{bmatrix}$$

and, therefore, $||\widetilde{Y}_{|2}||^2 = M_2 - m_2||E_2||^{-2}m_2'$ and $S_{ij.2} = S_{ij.1} - S_{i2.1}D_2^{-1}S_{j2.1}'$, $i,j = 3,\ldots,n$, $D_2 = S_{22.1}$.

Proceeding in this way and letting T be the lower triangular matrix

$$T = \begin{bmatrix} I & 0 & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & I & 0 \\ 0 & 0 & 0 & T_{n-1} \end{bmatrix} \cdots \begin{bmatrix} I & 0 \\ 0 & T_2 \end{bmatrix} T_1,$$

the following equality holds

$$T||Y||^2T' = \begin{bmatrix} ||E_1||^2 & 0 & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & ||E_{n-1}||^2 & 0 \\ 0 & 0 & 0 & ||E_n||^2 \end{bmatrix}.$$

Thus, the effect of premultiplying Y with T is to make the covariance matrix of TY block diagonal. This in turn implies that the block entries of TY are indeed the innovations, E_t, and

$$TY = E.$$

Therefore, $T = L^{-1}$, where L is the matrix (1.13). T is also the matrix (1.15) given by the autoregressive representation. It can be shown (see Problem 1.2) that the ith column of L can be obtained by taking the inner product of the ith column in the triangular array (1.18) with $||E_i||^2 E_i$, $i = 1, \ldots, n$.

We summarize the algorithm based on the modified Gram–Schmidt procedure in the following theorem.

Theorem 1.4 (Modified Innovations Algorithm) *Let the sequence of random vectors* $Y = (Y_1', \ldots, Y_n')'$*, where* $||Y||^2$ *is positive definite and* $S_{ij} = E(Y_i Y_j')$*,* $i, j = 1, \ldots, n$*. Define* $\widetilde{Y}_{j|0} = Y_j$*,* $j = 1, \ldots, n$*, and* $S_{ij.0} = S_{ij}$*,* $i, j = 1, \ldots, n$*. Then, the algorithm is*

Set $E_1 = Y_1, \quad D_1 = S_{11}$
for $t = 2, \ldots, n$

for $j = t, \ldots, n$
$\Theta^M_{j,t-1} = S_{j,t-1.t-2} D^{-1}_{t-1}$
$\widetilde{Y}_{j|t-1} = \widetilde{Y}_{j|t-2} - \Theta^M_{j,t-1} E_{t-1}$
for $k = j, \ldots, n$
$S_{kj.t-1} = S_{kj.t-2} - \Theta^M_{k,t-1} D_{t-1} \Theta^{M\prime}_{j,t-1}$
end
end
$E_t = \widetilde{Y}_{t|t-1}, \quad D_t = S_{tt.t-1}$

end

1.6 State Space Approach to the Innovations Algorithm

State space models will be introduced in the next section. In this section, we will obtain the same output of the innovations algorithm using a heuristic argument. It will turn out that the equations that we will use are a special case of state space models. Let the sequence of zero mean random vectors $Y = (Y_1' \ldots, Y_n')'$, where $||Y||^2$ is positive definite, and let $Y_{s|t} = E^*(Y_s | Y_1, \ldots, Y_t)$, $t < s$. Define the equations

$$x_{t+1} = F x_t + K_t E_t \tag{1.21}$$

$$Y_t = H x_t + E_t, \tag{1.22}$$

where $x_{t+1} = (Y'_{t+1|t}, \ldots, Y'_{t+n-1|t})', x_1 = 0,$

$$F = \begin{bmatrix} 0 & I & 0 & \cdots & 0 \\ 0 & 0 & I & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & I \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix}, \quad H = [I, 0, \ldots, 0],$$

and the matrix K_t is obtained from covariance data only by means of the recursions

$$\begin{aligned} D_t &= S_{tt} - H\Sigma_t H' \\ K_t &= (N_t - F\Sigma_t H')D_t^{-1}, \qquad N_t = \left[S'_{t+1,t}, S'_{t+2,t}, \ldots, S'_{t+n-1,t}\right]', \\ \Sigma_{t+1} &= F\Sigma_t F' + K_t D_t K_t', \quad t = 1, 2, \ldots, n, \end{aligned} \tag{1.23}$$

initialized with $\Sigma_1 = 0$. Here it is understood that all the random vectors that are not in the sample are zero. Thus, $Y_{s|t} = 0$ and $S_{st} = 0$ if $s > n$, $1 \le t \le n$.

Theorem 1.5 *The quantities E_t and D_t, generated by Eqs. (1.21) and (1.22) and the recursions (1.23), initialized with $x_1 = 0$ and $\Sigma_1 = 0$, coincide with the innovations and their variances. In addition, $\Sigma_t = Var(x_t)$.*

Proof The recursions are run for $t = 1, \ldots, n$, starting with $x_1 = 0$ and $\Sigma_1 = 0$. If $t = 1$, then $Y_1 = E_1$ and $D_1 = S_{11}$. If $t > 1$, then, by the definition of x_t, it is clear that $E_t = Y_t - Hx_t$ is the tth innovation. From this, it follows that $S_{tt} = D_t + H\Sigma_t H'$. On the other hand, by the properties of linear projection, the equation $x_{t+1} = Fx_t + K_t E_t$ is true if we define the matrix K_t appropriately. To see this, consider that, by Propositions 1.7 and 1.5,

$$\begin{aligned} Y_{t+i|t} &= E^*(Y_{t+i}|Y_1, \ldots, Y_{t-1}, Y_t) \\ &= E^*\left[Y_{t+i}|Y_1, \ldots, Y_{t-1}, E_t + E^*(Y_t|Y_1, \ldots, Y_{t-1})\right] \\ &= E^*(Y_{t+i}|Y_1, \ldots, Y_{t-1}) + E^*(Y_{t+i}|E_t) \\ &= Y_{t+i|t-1} + K_t E_t. \end{aligned}$$

Since $Y_{t+i|t} = Y_{t+i} - (Y_{t+i} - Y_{t+i|t})$ and $Y_{t+i} - Y_{t+i|t}$ is orthogonal to Y_t, it follows that $\text{Cov}(x_{t+1}, Y_t) = N_t$. Thus, $\text{Cov}(x_{t+1}, E_t) = K_t D_t = \text{Cov}(x_{t+1}, Y_t - Hx_t) = N_t - \text{Cov}(Fx_t, Hx_t) = N_t - F\Sigma_t H'$. Since the equation $\Sigma_{t+1} = F\Sigma_t F' + K_t D_t K_t'$ follows from the orthogonality of x_t and E_t, the result is proved. □

If we replace $E_t = Y_t - Hx_t$ in $x_{t+1} = Fx_t + K_tE_t$, the following state space form is obtained

$$x_{t+1} = (F - K_tH)x_t + K_tY_t$$
$$Y_t = Hx_t + E_t,$$

which produces an output equal to that of the autoregressive representation.

The decompositions (1.8) and (1.12) of $||Y||^2$ and $||Y||^{-2}$ can be written more compactly as

$$||Y||^2 = LDL', \qquad ||Y||^{-2} = W'D^{-1}W,$$

where

$$L = \begin{bmatrix} I & 0 & \cdots & 0 \\ \Theta_{21} & I & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \Theta_{n1} & \Theta_{n2} & \cdots & I \end{bmatrix}, \qquad W = L^{-1} = \begin{bmatrix} I & 0 & \cdots & 0 \\ -\Pi_{21} & I & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -\Pi_{n1} & -\Pi_{n2} & \cdots & I \end{bmatrix},$$

and $D = \mathrm{diag}(D_1, \ldots, D_n)$. The elements Θ_{ti} and Π_{ti} in L and W are time variant. In terms of the matrices of Eqs. (1.21) and (1.22) L and W can be written as

$$L = \begin{bmatrix} I & 0 & 0 & \cdots & 0 \\ HK_1 & I & 0 & \cdots & 0 \\ HFK_1 & HK_2 & I & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ HF^{n-2}K_1 & HF^{n-3}K_2 & HF^{n-4}K_3 & \cdots & I \end{bmatrix} \tag{1.24}$$

and

$$W = L^{-1} = \begin{bmatrix} I & 0 & 0 & \cdots & 0 \\ -HK_1 & I & 0 & \cdots & 0 \\ -HF^3_{p,2}K_1 & -HK_2 & I & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -HF^n_{p,2}K_1 & -HF^n_{p,3}K_2 & -HF^n_{p,4}K_3 & \cdots & I \end{bmatrix}, \tag{1.25}$$

where $F^j_{p,i} = F_{p,j-1}F_{p,j-2}\cdots F_{p,i}$ if $j > i$, $F^i_{p,i} = I$, and $F_{p,t} = F - K_tH$.

The dimensions in these last expressions could be greatly reduced if the data $Y = (Y_1', \ldots, Y_n')'$ had some structure. Equations (1.21) and (1.22) are a special case of state space models, as we will see in the next section.

1.7 Introduction to VARMA and State Space Models

On many occasions, the multivariate data under study have a finite structure that allows, among other things, for the generation of the first two moments of the distribution in a recursive fashion. This is the case of the so-called vector autoregressive moving average processes, or VARMA models for short. A vector process $\{Y_t\}$ with $Y_t \in \mathbb{R}^p$ is said to follow a **VARMA model** if it satisfies a stochastic difference equation of the form

$$Y_t + \Phi_{t,1}Y_{t-1} + \cdots + \Phi_{t,p}Y_{t-p} = \Theta_{t,0}A_t + \Theta_{t,1}A_{t-1} + \cdots + \Theta_{t,q}A_{t-q}, \tag{1.26}$$

where $\{A_t \in \mathbb{R}^s\}$ is a zero mean uncorrelated sequence with nonsingular covariance matrices, $\mathrm{Var}(A_t) = \Sigma_t$. The process will usually start at $t = 1$ and a set of initial conditions, I, will be necessary to specify its covariance structure. For example, if $r = \max(p, q)$ and $I = \{Y_1, \ldots, Y_r, A_1, \ldots, A_r\}$, specifying the covariance structure of I and assuming that I is uncorrelated with A_t for $t > r$ will be sufficient to generate the covariance matrices of Y_t for $t > r$ recursively using (1.26).

Another structure that we will often encounter is that of state space models. A vector process $\{Y_t\}$ with $Y_t \in \mathbb{R}^p$ is said to follow a **state space model** if it satisfies a model of the form

$$x_{t+1} = F_t x_t + G_t \epsilon_t \tag{1.27}$$

$$Y_t = H_t x_t + J_t \epsilon_t, \tag{1.28}$$

where $x_t \in \mathbb{R}^r$ is the state vector, $\{\epsilon_t \in \mathbb{R}^s\}$ is an uncorrelated sequence with $E(\epsilon_t) = 0$ and $\mathrm{Var}(\epsilon_t) = I$, the initial state vector, x_1, is orthogonal to ϵ_t for all t, $E(x_1) = 0$ and $\mathrm{Var}(x_1) = \Omega$. Equations (1.27) and (1.28) are called the "transition equation" and the "measurement equation," respectively.

It seems perhaps surprising that VARMA and state space models are equivalent.

Theorem 1.6 *Suppose that the process $\{Y_t\}$ follows a state space model (1.27) and (1.28) and assume that the matrices $O_{t,r} = [H_t', (H_{t+1}F_t^{t+1})', \ldots, (H_{t+r-1}F_t^{t+r-1})']'$, where $r = \dim(x_t)$, $F_i^j = F_{j-1}F_{j-2}\cdots F_i$ if $j > i$ and $F_i^i = I$, have full column rank for all t. Then, $\{Y_t\}$ follows a VARMA model (1.26).*

Proof Iterating in the state space model and stacking the observations, we get

$$Y_{t:t+r} = O_{t,r+1}x_t + P_{t,r+1}U_{t:t+r}, \tag{1.29}$$

where $Y_{t:t+r} = (Y_t', Y_{t+1}', \ldots, Y_{t+r}')'$,

$$P_{t,r+1} = \begin{bmatrix} J_t & & & \\ H_{t+1}G_t & J_{t+1} & & \\ \vdots & \vdots & \ddots & J_{t+n-1} \\ H_{t+r}F_t^{t+r}G_t & \cdots & \cdots\ H_{t+r}G_{t+r-1} & J_{t+r} \end{bmatrix},$$

$O_{t,r+1} = [O_{t,r}', (H_{t+r}F_t^{t+r})']'$, and $U_{t:t+r} = (\epsilon_t', \epsilon_{t+1}', \text{ and } \ldots, \epsilon_{t+r}')'$. Since $O_{t,r}$ has full column rank, there exists a block matrix, $\Phi_t = [\Phi_{t,r}, \ldots, \Phi_{t,1}, I]$, such that $\Phi_t O_{t,r+1} = 0$. Then, premultiplying (1.29) by Φ_t, it is obtained that

$$[\Phi_{t,r}, \ldots, \Phi_{t,1}, I]Y_{t:t+r} = [\Phi_{t,r}, \ldots, \Phi_{t,1}, I]P_{t,r+1}U_{t:t+r}.$$

The right-hand side in the previous equality is a process of the form $\Theta_{t,0}\epsilon_{t+r} + \Theta_{t,1}\epsilon_{t+r-1} + \cdots + \Theta_{t,r}\epsilon_t$. Thus, the theorem follows. □

Theorem 1.7 *Suppose that the process $\{Y_t\}$ follows a VARMA model (1.26). Then, $\{Y_t\}$ follows a state space model (1.27) and (1.28).*

Proof Let $r = \max(p, q+1)$ and define $\Phi_{t,i} = 0$ if $p < i$ and $\Theta_{t,i} = 0$ if $q < i$. We illustrate the procedure to construct a state space model for $\{Y_t\}$ when $p = 2$ and $q = 2$. The general case is similar. Set $\epsilon_t = L_t^{-1}A_{t+1}$, where $\text{Var}(A_t) = L_tL'$ is the Cholesky decomposition of $\text{Var}(A_t)$, and define

$$\begin{aligned} x_{t+1} &= \begin{bmatrix} -\Phi_{t+1,1} & I & 0 \\ -\Phi_{t+2,2} & 0 & I \\ -\Phi_{t+3,3} & 0 & 0 \end{bmatrix} x_t + \begin{bmatrix} \Theta_{t+1,0} \\ \Theta_{t+2,1} \\ \Theta_{t+3,2} \end{bmatrix} L_t\epsilon_t \\ Y_t &= [I, 0, 0]x_t. \end{aligned}$$

Then, if $x_{t+1} = (x_{t+1,1}', x_{t+1,2}', x_{t+1,3}')'$, using the state space model equations, we get

$$\begin{aligned} x_{t+1,1} &= -\Phi_{t+1,1}x_{t,1} + x_{t,2} + \Theta_{t+1,0}A_{t+1} \\ &= -\Phi_{t+1,1}x_{t,1} - \Phi_{t+1,2}x_{t-1,1} + x_{t-1,3} + \Theta_{t+1,1}A_t + \Theta_{t+1,0}A_{t+1} \\ &= -\Phi_{t+1,1}x_{t,1} - \Phi_{t+1,2}x_{t-1,1} + \Theta_{t+1,0}A_{t+1} + \Theta_{t+1,1}A_t + \Theta_{t+1,2}A_{t-1}. \end{aligned}$$

Therefore, $x_{t,1}$ satisfies the same difference equation as Y_t. □

Remark 1.5 The state space representation used in the proof of the previous theorem is not unique. For example, another representation that can be used for the same case, $p = 2$ and $q = 2$, is the following.

$$x_{t+1} = \begin{bmatrix} -\Phi_{t+1,1} & I \\ -\Phi_{t+2,2} & 0 \end{bmatrix} x_t + \begin{bmatrix} \Theta_{t+1,1} - \Phi_{t+1,1}\Theta_{t,0} \\ \Theta_{t+2,2} - \Phi_{t+2,2}\Theta_{t,0} \end{bmatrix} L_t\epsilon_t$$
$$Y_t = [I, 0]x_t + \Theta_{t,0}L_t\epsilon_t,$$

where $\epsilon_t = L_t^{-1}A_t$. See Problem 1.5. ◇

An alternative state space representation to (1.27) and (1.28) is

$$x_{t+1} = F_t x_t + G_t u_t \tag{1.30}$$

$$Y_t = H_t x_t + v_t, \tag{1.31}$$

where $x_t \in \mathbb{R}^r$ is the state vector,

$$E\left\{\begin{bmatrix} u_t \\ v_t \end{bmatrix}\left[u_s', v_s'\right]\right\} = \begin{bmatrix} Q_t & S_t \\ S_t' & R_t \end{bmatrix}\delta_{ts},$$

$u_t \in \mathbb{R}^s$, $v_t \in \mathbb{R}^p$, $E(u_t) = 0$, $E(v_t) = 0$, the initial state vector, x_1, is orthogonal to u_t and v_t for all t, $E(x_1) = 0$ and $\text{Var}(x_1) = \Omega$. Later in this section, in Lemma 1.1, we will describe how the covariance matrices of a process following (1.30) and (1.31) can be obtained.

The fact that the same term ϵ_t appears in (1.27) and (1.28) is general, not restrictive, and facilitates sometimes the development of more elegant formulae for the filtering and smoothing algorithms that we will describe in Chap. 4. Usually, the matrices G_t and J_t are selection matrices formed with zeros and ones. To pass from the state space representation (1.30) and (1.31)–(1.27) and (1.28), consider the decomposition

$$W_t = \begin{bmatrix} G_tQ_tG_t' & G_tS_t \\ S_t'G_t' & R_t \end{bmatrix}$$
$$= W_t^{1/2}W_t^{1/2'},$$

where if M is a symmetric square matrix, $M^{1/2}$ is any matrix satisfying $M = M^{1/2}M^{1/2'}$. For example, let O be an orthogonal matrix such that $O'MO = D$, where D is a diagonal matrix. Then, we can take $M^{1/2} = OD^{1/2}O'$, where $D^{1/2}$ is the matrix obtained from D by replacing its nonzero elements with their square roots. This choice of $M^{1/2}$ has the advantage of being valid and numerically stable even if M is singular. It follows from this that we can take $(G_t', J_t')' = W_t^{1/2}$.

Example 1.2 (Continued) Equation (1.26) is in this case

$$Y_t = A_t + \theta A_{t-1}.$$

One possible set of Eqs. (1.27) and (1.28) is

$$x_{t+1} = \theta\sigma\epsilon_t$$
$$Y_t = x_t + \sigma\epsilon_t, \qquad t = 1, 2, \ldots, n.$$

Thus, in this case, $F_t = 0$, $H_t = 1$, $G_t = \theta\sigma$, $J_t = \sigma$, $\epsilon_t = A_t/\sigma$, and $\Omega = \text{Var}(x_1) = \theta^2\sigma^2$. ◇

Example 1.3 (Continued) Equation (1.26) is for this example

$$Y_t = \phi Y_{t-1} + A_t.$$

One possible set of Eqs. (1.27) and (1.28) is

$$x_{t+1} = \phi x_t + \sigma\epsilon_t$$
$$Y_t = x_t, \qquad t = 1, 2, \ldots, n.$$

Thus, $F_t = \phi$, $H_t = 1$, $G_t = \sigma$, $J_t = 0$, $\epsilon_t = A_{t+1}/\sigma$, and $\Omega = \text{Var}(x_1) = \sigma^2/(1-\phi^2)$. ◇

1.7.1 Innovations Algorithm for VARMA Models

In the case of a VARMA model, a transformation, originally proposed by Ansley (1979) for time invariant models, simplifies the equations of the innovations algorithm considerably. Let $\{Y_t\}$ with $Y_t \in \mathbb{R}^p$ be a process that follows the model

$$Y_t + \Phi_{t,1}Y_{t-1} + \cdots + \Phi_{t,p}Y_{t-p} = \Gamma_{t,0}A_t + \Gamma_{t,1}A_{t-1} + \cdots + \Gamma_{t,q}A_{t-q},$$

where $\{A_t \in \mathbb{R}^s\}$ is an uncorrelated sequence with nonsingular covariance matrices, $\text{Var}(A_t) = \Sigma_t$. Assume that we have a sequence of vectors, $Y = (Y_1', \ldots, Y_n')'$, generated by the previous model and such that $\text{Var}(Y)$ is positive definite. Define

$$X_t = \begin{cases} Y_t, & \text{if } t = 1, \ldots, r, \\ Y_t + \Phi_{t,1}Y_{t-1} + \cdots + \Phi_{t,p}Y_{t-p}, & \text{if } t > r, \end{cases}$$

where $r = \max(p, q)$. Denoting the covariance function of $\{Y_t\}$ by $\gamma_Y(i,j) = E(Y_i Y_j')$, it is not difficult to verify that the covariance function $S_{ij} = E(X_i X_j')$ is given by

$$S_{ij} = \begin{cases} \gamma_Y(i,j) & \text{if } 1 \le i \le j \le r \\ \gamma_Y(i,j) + \sum_{h=1}^{p} \Phi_{i,h}\gamma_Y(i+h,j) & \text{if } 1 \le i \le r < j \le 2r \\ \sum_{h=0}^{q} \Gamma_{i,h}\Sigma_{i-h}\Gamma'_{j,h+j-i} & \text{if } r < i \le j \le i+q \\ 0 & \text{if } r < i \quad \text{and} \quad i+q < j \\ S'_{ij} & \text{if } j < i, \end{cases} \tag{1.32}$$

where $\Gamma_{t,h} = 0$ if $h > q$. The notable feature of the previous transformation is that $S_{ij} = 0$ if $|j-i| > q$, $i,j > r$. This in turn, by the decomposition (1.8), implies that

$$X_t = \Theta_{t,t-q}E_{t-q} + \cdots + \Theta_{t,t-1}E_{t-1} + E_t, \qquad t > r,$$

when the innovations algorithm is applied to $\{X_t : t = 1, 2, \ldots, n\}$. More specifically, the output of this last algorithm in terms of $\{Y_t : t = 1, 2, \ldots, n\}$ is easily shown to be

$$Y_t = \Theta_{t,1}E_1 + \cdots + \Theta_{t,t-1}E_{t-1} + E_t, \qquad t \le r,$$

where

$$\Theta_{tj} = \left[\gamma_Y(t,j) - \sum_{i=1}^{j-1} \Theta_{ti}D_i\Theta'_{ji}\right] D_j^{-1}, \qquad j = 1, \ldots, t-1,$$

$$D_t = \gamma_Y(t,t) - \sum_{i=1}^{t-1} \Theta_{ti}D_i\Theta'_{ti},$$

and

$$Y_t + \sum_{j=1}^{p} \Phi_{t,j}Y_{t-j} = \Theta_{t,t-q}E_{t-q} + \cdots + \Theta_{t,t-1}E_{t-1} + E_t, \qquad t > r,$$

where

$$\Theta_{tj} = \left(S_{tj} - \sum_{i=t-q}^{j-1} \Theta_{ti}D_i\Theta'_{ji}\right) D_j^{-1}, \qquad j = t-q, \ldots, t-1,$$

$$D_t = S_{tt} - \sum_{i=t-q}^{t-1} \Theta_{ti}D_i\Theta'_{ti},$$

and S_{ij} is given by (1.32). In addition, since the matrix of the transformation that gives $(X_1', \ldots, X_t')'$ in terms of $(Y_1', \ldots, Y_t')'$ is easily seen to be nonsingular for $t = 1, 2, \ldots, n$, it follows from Proposition 1.7 that $E^*(Y_t|X_{t-1}, \ldots, X_1) = E^*(Y_t|Y_{t-1}, \ldots, Y_1)$. This implies $E^*(X_t|X_{t-1}, \ldots, X_1) = E^*(Y_t|Y_{t-1}, \ldots, Y_1)$ for $t = 1, \ldots, r$, and $E^*(X_t|X_{t-1}, \ldots, X_1) = E^*(Y_t|Y_{t-1}, \ldots, Y_1) + \sum_{j=1}^{p} \Phi_{t,j} Y_{t-j}$ for $t = r+1, \ldots, n$. Therefore,

$$E_t = X_t - E^*(X_t|X_{t-1}, \ldots, X_1) = Y_t - E^*(Y_t|Y_{t-1}, \ldots, Y_1), \qquad t = 1, 2, \ldots, n,$$

and the E_t and D_t given by the innovations algorithm applied to $\{X_t : t = 1, 2, \ldots, n\}$ can be used for prediction and likelihood evaluation of $\{Y_t : t = 1, 2, \ldots, n\}$.

Example 1.4 Let Y_1 be a random variable with $\mathrm{E}(Y_1) = 0$ and $\mathrm{Var}(Y_1) = [(1 - 2\theta\phi + \theta^2)/(1 - \phi^2)]\sigma^2$ and let $Y_t + \phi Y_{t-1} = A_t + \theta A_{t-1}$, $t \geq 2$, where $|\phi| < 1$, $|\theta| < 1$ and $\{A_t\}$ is an uncorrelated sequence of random variables with zero mean and common variance σ^2 such that $E(Y_1 A_1) = \sigma^2$ and $E(Y_1 A_t) = 0$, $t \geq 2$. Given the sequence $Y = (Y_1, \ldots, Y_n)'$, we can obtain the covariances $\gamma_Y(i,j) = E(Y_i Y_j')$, $i, j = 1, \ldots, n$, as follows. Clearly, $\gamma_Y(1,1) = \mathrm{Var}(Y_1)$. To obtain $\gamma_Y(2,1)$, multiply the equation $Y_2 + \phi Y_1 = A_2 + \theta A_1$ by Y_1 and take expectations to get

$$\gamma_Y(2,1) = -\phi\gamma_Y(1,1) + \theta\sigma^2.$$

Multiplying the same equation this time by Y_2, taking expectations, and using the previous relation, it is obtained that $\gamma_Y(2,2) = \gamma_Y(1,1)$. We could continue in this way, but it will not be necessary because, following Ansley (1979), we transform the $\{Y_t\}$ process before applying the innovations algorithm so that the new process $\{X_t\}$ is

$$X_t = \begin{cases} Y_t, & \text{if } t = 1 \\ Y_t + \phi Y_{t-1}, & \text{if } t > 1, \end{cases}$$

Then, letting $S_{ij} = E(X_i X_j)$, (1.32) yields

$$S_{ij} = \begin{cases} \gamma_Y(1,1) & \text{if } i = j = 1 \\ \theta\sigma^2 & \text{if } i \geq 1, j = i+1 \\ (1+\theta^2)\sigma^2 & \text{if } i = j, \quad i \geq 2 \\ 0 & \text{if } 1 < i \quad \text{and} \quad i+1 < j \\ S_{ij}' & \text{if } j < i, \end{cases}$$

Applying the innovations algorithm to the transformed model, we get $E_1 = Y_1$, $D_1 = [(1 - 2\theta\phi + \theta^2)/(1 - \phi^2)]\sigma^2$, and for $t \geq 2$, $Y_t + \phi Y_{t-1} = E_t + \Theta_{t,t-1} E_{t-1}$,

where

$$\begin{aligned} \Theta_{t,t-1}D_{t-1} &= S_{t,t-1} &&= \theta\sigma^2 \\ D_t &= S_{tt} - \Theta_{t,t-1}^2 D_{t-1} &&= (1+\theta^2)\sigma^2 - \theta^2\sigma^4/D_{t-1}. \end{aligned}$$

It is shown in Problem 1.6 that the previous recursions can be easily solved. Letting

$$L^{-1} = \begin{bmatrix} 1 & & & \\ \phi & 1 & & \\ & \ddots & \ddots & \\ & & \phi & 1 \end{bmatrix},$$

$Y = (Y_1, \ldots, Y_n)'$ and $X = (X_1, \ldots, X_n)'$, we can write $X = L^{-1}Y$. Then, we can express $\mathrm{Cov}(Y)$ as

$$\mathrm{Var}(Y) = LL_xD_xL_x'L',$$

where L_x and D_x are

$$L_x = \begin{bmatrix} 1 & & & \\ \Theta_{21} & 1 & & \\ & \ddots & \ddots & \\ & & \Theta_{n,n-1} & 1 \end{bmatrix}, \qquad D_x = \begin{bmatrix} D_1 & & & \\ & D_2 & & \\ & & \ddots & \\ & & & D_n \end{bmatrix}.$$

◇

1.7.2 Covariance-Based Filter for State Space Models

The calculations (1.23) to obtain the innovations and their covariance matrices using (1.21) and (1.22), given a stack of random vectors, $X = (Y_1', \ldots, Y_n')'$, can be greatly simplified if the data have some structure.

Assume the state space model (1.30) and (1.31). The following lemma gives the covariance function of the output process $\{Y_t\}$. We leave the proof as an exercise. See Problem 1.7.

Lemma 1.1 *Consider the state space model (1.30) and (1.31) and let* $\Pi_t = E(x_tx_t')$. *Then,* Π_t *satisfies* $\Pi_1 = \Omega = E(x_1x_1')$ *and*

$$\Pi_{t+1} = F_t\Pi_tF_t' + G_tQ_tG_t', \qquad t \geq 1. \tag{1.33}$$

The covariance matrices of the state variables can be written as

$$\gamma_X(r,s) = E(x_r x_s') = \begin{cases} F_s^r \Pi_s & r \geq s \\ \Pi_r F_r^{s\prime} & r \leq s, \end{cases}$$

and the covariance matrices of the output process $\{Y_t\}$ *as*

$$\gamma_Y(r,s) = E(Y_r Y_s') = \begin{cases} H_r F_{s+1}^r N_s & r > s \\ R_r + H_r \Pi_r H_r' & r = s \\ N_r' F_{r+1}^{s\prime} H_s' & r < s, \end{cases} \tag{1.34}$$

where $F_i^j = F_{j-1}F_{j-2}\cdots F_i$ *if* $i < j$, $F_i^i = I$, *and* $N_r = F_r \Pi_r H_r' + G_r S_r = Cov(x_{r+1}, Y_r)$.

Theorem 1.8 *Suppose that the process* $\{Y_t\}$ *follows the state space model (1.30) and (1.31) so that the covariance matrices,* $\gamma_Y(r,s) = E(Y_r Y_s')$, *are generated by (1.33) and (1.34). If* $Y = (Y_1', \ldots, Y_n')'$ *is such that* $Var(Y)$ *is positive definite, then* $\{Y_t\}$ *admits the following state space representation*

$$\begin{aligned} \hat{x}_{t+1|t} &= F_t \hat{x}_{t|t-1} + K_t E_t \\ Y_t &= H_t \hat{x}_{t|t-1} + E_t, \qquad t = 1, 2, \ldots, n, \end{aligned}$$

where $\hat{x}_{t|t-1} = E^*(x_t | Y_1, \ldots, Y_{t-1})$, E_t *is the innovation,* $E_t = Y_t - E^*(Y_t | Y_1, \ldots, Y_{t-1})$, *and the quantities* $Var(E_t) = D_t$, $Var(\hat{x}_{t|t-1}) = \Sigma_t$ *and* K_t *are obtained by means of the recursions based on covariance data only*

$$\begin{aligned} D_t &= \gamma_Y(t,t) - H_t \Sigma_t H_t' \\ K_t &= (N_t - F_t \Sigma_t H_t') D_t^{-1} \\ \Sigma_{t+1} &= F_t \Sigma_t F_t' + K_t D_t K_t', \end{aligned} \tag{1.35}$$

initialized with $\hat{x}_{1|0} = 0$ *and* $\Sigma_1 = 0$.

The whitening filter for $\{Y_t\}$ *is further given by*

$$\begin{aligned} \hat{x}_{t+1|t} &= (F_t - K_t H_t)\hat{x}_{t|t-1} + K_t Y_t \\ E_t &= Y_t - H_t \hat{x}_{t|t-1}. \end{aligned}$$

Proof If $t = 1$, then $Y_1 = E_1$ and $D_1 = \gamma_Y(1,1)$ and there is nothing to prove. If $t > 1$, consider the transformation (1.7) implied by the innovations algorithm. If this transformation coincides with that implied by the state space form of the theorem and (1.35), namely $Y = \Psi E$, where $E = (E_1', \ldots, E_n')'$, $Y = (Y_1', \ldots, Y_n')'$, and Ψ is block lower triangular with blocks of unit matrices in the main diagonal, the first part of the theorem will be proved. Iterating in the state space equations of the theorem, it is not difficult to show that Ψ is given by

$$\Psi = \begin{bmatrix} I & 0 & \cdots & 0 & 0 \\ H_2K_1 & I & \cdots & 0 & 0 \\ H_3F_2^3K_1 & H_3K_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ H_nF_2^nK_1 & H_nF_3^nK_2 & \cdots & H_nK_{n-1} & I \end{bmatrix},$$

where $F_i^j = F_{j-1}F_{j-2}\cdots F_i$ if $i < j$, and $F_i^i = I$,

Let the block coefficients of Ψ different from zero or the unit matrix be $\Theta_{tj} = H_tF_{j+1}^tK_j$. Then,

$$\begin{aligned} \Theta_{tj} &= H_tF_{j+1}^tK_j \\ &= H_tF_{j+1}^t(N_j - F_j\Sigma_jH_j')D_j^{-1} \\ &= (\gamma_Y(t,j) - H_tF_j^t\Sigma_jH_j')D_j^{-1}, \quad j = 1,2,\ldots,t-1. \end{aligned}$$

Also, it follows from the recursion of the theorem for Σ_t that

$$\begin{aligned} H_t\Sigma_tH_t' &= \sum_{i=1}^{t-1} H_tF_{i+1}^tK_iD_iK_i'F_{i+1}^{t'}H_t' \\ &= \sum_{i=1}^{t-1} \Theta_{ti}D_i\Theta_{ti}' \end{aligned} \tag{1.36}$$

and

$$\begin{aligned} H_tF_j^t\Sigma_jH_j' &= \sum_{i=1}^{j-1} H_tF_{i+1}^tK_iD_iK_i'F_{i+1}^{j'}H_j' \\ &= \sum_{i=1}^{j-1} \Theta_{ti}D_i\Theta_{ji}'. \end{aligned}$$

Therefore, we obtain the same output as with the innovations algorithm,

$$\begin{aligned} \Theta_{tj} &= \left(\gamma_Y(t,j) - \sum_{i=1}^{j-1} \Theta_{ti}D_i\Theta_{ji}'\right)D_j^{-1}, \quad j = 1,\ldots,t-1, \\ D_t &= \gamma_Y(t,t) - \sum_{i=1}^{t-1} \Theta_{ti}D_i\Theta_{ti}'. \end{aligned}$$

The last recursion follows from the recursion of the theorem for D_t and (1.36). □

The following corollary will be useful later in the book. The proof is omitted because it is a direct consequence of the previous theorem.

Corollary 1.3 *Under the assumptions and with the notation of Theorem 1.8, the following decompositions hold*

$$||Y||^2 = LDL', \qquad ||Y||^{-2} = WD^{-1}W',$$

where

$$L = \begin{bmatrix} I & 0 & \cdots & 0 & 0 \\ H_2K_1 & I & \cdots & 0 & 0 \\ H_3F_2^3K_1 & H_3K_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ H_nF_2^nK_1 & H_nF_3^nK_2 & \cdots & H_nK_{n-1} & I \end{bmatrix},$$

$$W = L^{-1} = \begin{bmatrix} I & 0 & \cdots & 0 & 0 \\ -H_2K_1 & I & \cdots & 0 & 0 \\ -H_3F_{p,2}^3K_1 & -H_3K_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ -H_nF_{p,2}^nK_1 & -H_nF_{p,3}^nK_2 & \cdots & -H_nK_{n-1} & I \end{bmatrix}$$

$D = diag(D_1, D_2, \ldots, D_n)$, $F_i^j = F_{j-1}F_{j-2}\cdots F_i$ if $i < j$, $F_i^i = I$, $F_{p,i}^j = F_{p,j-1}F_{p,j-2}\cdots F_{p,i}$ if $i < j$, $F_{p,i}^i = I$, and $F_{p,t} = F_t - K_tH_t$.

A special case of Theorem 1.8 that is important in practice is that in which the matrices of the state space model (1.30) and (1.31) are time invariant. That is, $\{Y_t\}$ follows the model

$$x_{t+1} = Fx_t + Gu_t \tag{1.37}$$

$$Y_t = Hx_t + v_t, \tag{1.38}$$

where

$$E\left\{\begin{bmatrix} u_t \\ v_t \end{bmatrix}\begin{bmatrix} u_s', v_s' \end{bmatrix}\right\} = \begin{bmatrix} Q & S \\ S' & R \end{bmatrix}\delta_{ts}.$$

If, in addition, $\Pi = E(x_1x_1')$ satisfies

$$\Pi = F\Pi F' + GQG', \tag{1.39}$$

then, by Lemma 1.1, the covariance matrices between Y_r and Y_s depend on $r - s$ only. More specifically, they are given by

$$\gamma_Y(r-s) = E(Y_r Y_s') = \begin{cases} HF^{r-s-1}N & r > s \\ R + H\Pi H' & r = s \\ N'F^{(s-r-1)'}H' & r < s, \end{cases} \tag{1.40}$$

where $F^0 = I$ and $N = F\Pi H' + GS = \mathrm{Cov}(x_{t+1}, Y_t)$.

The processes $\{Y_t\}$ such that the mean of Y_t is constant and the covariance matrices between Y_{t+j} and Y_t depend on j only are called **wide sense stationary** or, simply, stationary and will be studied in more detail in the next chapter. For stationary processes, we have the following corollary that can be proved as Theorem 1.8.

Corollary 1.4 *Suppose a wide sense stationary vector process $\{Y_t\}$ that follows the state space model (1.37) and (1.38) with $\Pi = E(x_1 x_1')$ satisfying (1.39) or, alternatively, suppose that there exist matrices F, H, and N such that the covariance matrices, $\gamma(j) = E(Y_{t+j}Y_t')$, are $\gamma(0)$ and $\gamma(j) = HF^{j-1}N$, $\gamma(-j) = \gamma(j)'$, $j = 1, 2, \ldots$. If $Y = (Y_1', \ldots, Y_n')'$ is such that Var(Y) is positive definite, then $\{Y_t\}$ admits the following state space representation*

$$\hat{x}_{t+1|t} = F\hat{x}_{t|t-1} + K_t E_t \tag{1.41}$$

$$Y_t = H\hat{x}_{t|t-1} + E_t, \qquad t = 1, 2, \ldots, n, \tag{1.42}$$

where the E_t are the innovations, and the quantities Var$(E_t) = D_t$, Var$(\hat{x}_{t|t-1}) = \Sigma_t$, and K_t are obtained by means of the recursions

$$\begin{aligned} D_t &= \gamma(0) - H\Sigma_t H' \\ K_t &= (N - F\Sigma_t H')D_t^{-1} \\ \Sigma_{t+1} &= F\Sigma_t F' + K_t D_t K_t', \end{aligned} \tag{1.43}$$

initialized with $\hat{x}_{1|0} = 0$ and $\Sigma_1 = 0$.

Example 1.2 (Continued) It is clear that the process $\{Y_t\}$ is stationary. In this case, the matrices F, H, and N of (1.40) can be defined as the scalars 0, 1 and $N = \gamma(1)$, so that the state space equations (1.41) and (1.42) become

$$\begin{aligned} \hat{x}_{t+1|t} &= K_t E_t \\ Y_t &= \hat{x}_{t|t-1} + E_t \end{aligned}$$

and the equations for K_t, D_t, and Σ_t are

$$
\begin{aligned}
D_t &= \gamma(0) - \Sigma_t \\
K_t &= \gamma(1)/D_t \\
\Sigma_{t+1} &= K_t D_t K_t,
\end{aligned}
$$

where $\gamma(0) = (1 + \theta^2)\sigma^2 = \mathrm{Var}(Y_t)$ and $\gamma(1) = \theta\sigma^2$. ◇

Example 1.3 (Continued) Clearly, the process $\{Y_t\}$ is stationary and the matrices F, H and N of (1.40) can be defined as the scalars ϕ, 1 and $N = \gamma(1)$. Thus, the state space equations (1.41) and (1.42) become

$$
\begin{aligned}
\hat{x}_{t+1|t} &= \phi x_{t|t-1} + K_t E_t \\
Y_t &= \hat{x}_{t|t-1} + E_t
\end{aligned}
$$

and the equations for K_t, D_t, and Σ_t are

$$
\begin{aligned}
D_t &= \gamma(0) - \Sigma_t \\
K_t &= (\gamma(1) - \phi\Sigma_t)/D_t \\
\Sigma_{t+1} &= \phi^2\Sigma_t + K_t D_t K_t,
\end{aligned}
$$

where $\gamma(0) = \sigma^2/(1-\phi^2) = \mathrm{Var}(Y_t)$ and $\gamma(1) = \phi\gamma(0)$. It is not difficult to verify that $D_1 = \gamma(0)$, $K_t = \phi$ for $t \geq 1$, and $D_t = \sigma^2$ and $\Sigma_t = \phi^2\gamma(0)$ for $t \geq 2$. ◇

Remark 1.6 Note that in Corollary 1.4 the form of the covariance matrices, $\gamma(j)$, given by (1.40), implies that the covariance Hankel matrix of order r,

$$
G_r = \begin{bmatrix}
\gamma(1) & \gamma(2) & \gamma(3) & \cdots & \gamma(r) \\
\gamma(2) & \gamma(3) & \gamma(4) & \cdots & \gamma(r+1) \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\gamma(r) & \gamma(r+1) & \gamma(t+2) & \cdots & \gamma(2r-1)
\end{bmatrix},
$$

can be expressed as

$$
G_r = \begin{bmatrix}
H \\
HF \\
\vdots \\
HF^{r-1}
\end{bmatrix} \left[N, FN, \ldots, F^{r-1}N\right].
$$

◇

Example 1.5 (Moving Average Processes) A stationary process of the form $Y_t = A_t + \Theta_1 A_{t-1} + \cdots + \Theta_q A_{t-q}$, $t > q$, where $\{A_t : t \geq 1\}$ is a sequence of zero mean, uncorrelated random vectors with common nonsingular covariance matrix Ω, is called a moving average of order q. It is easy to verify that all of its covariance matrices, $\gamma(j)$, are zero except for $j = 0, 1, 2, \ldots, q$, and that $\gamma(j) = HF^{j-1}N$, $j = 1, 2, \ldots, q$, where

$$F = \begin{bmatrix} 0 & I & 0 & \cdots & 0 \\ 0 & 0 & I & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & I \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix}, \quad H = [I, 0, \ldots, 0],$$

$N = [\gamma'(1), \gamma'(2), \ldots, \gamma'(q)]'$ and, letting $\Theta_0 = I$,

$$\gamma(j) = \sum_{i=j}^{q} \Theta_i \Omega \Theta'_{i-j}, \quad j = 0, 1, 2, \ldots, q.$$

In order for the process $\{Y_t : t = 1, 2, \ldots\}$ to be well defined, further assumptions are needed. These are that the vector of initial conditions, $(Y'_1, \ldots, Y'_q)'$, be uncorrelated with the sequence $\{A_t : t > q\}$ and that $\mathrm{Cov}(Y_i, Y_j) = \gamma(i-j), i, j = 1, 2, \ldots, q$, and $\mathrm{Cov}(Y_i, A_j) = \Theta_{i-j}\Omega, j \leq i = 1, 2, \ldots, q$. It will be proved later in the book that the quantities Σ_t, K_t and D_t, given by the recursions of Corollary 1.4, satisfy $\Sigma_t \to \Sigma$, $K_t \to K = \left[\Theta'_1, \Theta'_2, \ldots, \Theta'_q\right]'$ and $D_t \to D = \Omega$ as $t \to \infty$. The limit solution can be obtained by solving the following discrete algebraic Riccati equation

$$\Sigma = F\Sigma F' + (N - F\Sigma H')[\gamma(0) - H\Sigma H']^{-1}(N - F\Sigma H')',$$

and setting $K = (N - F\Sigma H')D^{-1}$ and $D = \gamma(0) - H\Sigma H'$. $\diamond$

1.8 Further Topics Associated With Orthogonal Projection

1.8.1 Sequential Update of an Orthogonal Projection

Suppose we want to predict the random vector Y_3 and we observe the random vector Y_1. Then, as we saw in the discussion that led to (1.6), the best linear predictor is $E^*(Y_3|Y_1)$. If a new observation Y_2 becomes available, the new best linear predictor is $E^*(Y_3|Y_1, Y_2)$. Since the transformation (1.7) is nonsingular, by Propositions 1.7

and 1.5 , the following equality holds

$$\begin{aligned} E^*(Y_3|Y_1, Y_2) &= E^*(Y_3|E_1, E_2) \\ &= E^*(Y_3|E_1) + E^*(Y_3|E_2) \\ &= E^*(Y_3|Y_1) + \Pi_{32}[Y_2 - E^*(Y_2|Y_1)], \end{aligned} \tag{1.44}$$

where $E^*(Y_3|E_2) = \Pi_{32}E_2 = \Pi_{32}[Y_2 - E^*(Y_2|Y_1)]$. Thus, the predictor $E^*(Y_3|Y_1)$ can be updated by first computing the new predictor $E^*(Y_2|Y_1)$ and then the coefficient Π_{32}. However, by Corollary 1.2, Π_{32} in (1.44) coincides with the coefficient matrix Θ_{32} in the expression

$$Y_3 = \Theta_{31}E_1 + \Theta_{32}E_2 + E_3,$$

given by the innovations algorithm. The recursions of this last algorithm yield

$$\begin{aligned} &\Theta_{21} = S_{21}D_1^{-1}, \quad \Theta_{31} = S_{31}D_1^{-1}, \quad D_1 = S_{11} \\ &\Theta_{32} = (S_{32} - S_{31}\Theta'_{21})D_2^{-1} = (S_{32} - S_{31}S_{11}^{-1}S_{12})D_2^{-1}, \quad D_2 = S_{22} - S_{21}S_{11}^{-1}S_{12}. \end{aligned} \tag{1.45}$$

and

$$D_3 = S_{33} - \Theta_{32}D_2\Theta'_{32} - \Theta_{31}D_1\Theta'_{31}.$$

On the other hand, it follows from (1.44) and $\Pi_{32} = \Theta_{32}$ that

$$Y_3 - E^*(Y_3|Y_1, Y_2) + \Theta_{32}[Y_2 - E^*(Y_2|Y_1)] = Y_3 - E^*(Y_3|Y_1)$$

and, because $Y_3 - E^*(Y_3|Y_1, Y_2)$ is orthogonal to $Y_2 - E^*(Y_2|Y_1)$,

$$\text{MSE}[E^*(Y_3|Y_1, Y_2)] = \text{MSE}[E^*(Y_3|Y_1)] - \Theta_{32}\text{MSE}[E^*(Y_2|Y_1)]\Theta'_{32}. \tag{1.46}$$

Since $\text{MSE}[E^*(Y_3|Y_1, Y_2)] = D_3$, $\text{MSE}[E^*(Y_3|Y_1)] = S_{33} - \Theta_{31}D_1\Theta'_{31}$, and $\text{MSE}[E^*(Y_2|Y_1)] = D_2$, the following formula for updating the MSE in (1.44) is obtained

$$\text{MSE}[E^*(Y_3|Y_1, Y_2)] = S_{33.1} - S_{32.1}S_{22.1}^{-1}S_{23.1},$$

where $\text{MSE}[E^*(Y_3|Y_1)] = S_{33.1}$, $\text{MSE}[E^*(Y_2|Y_1)] = S_{22.1}$, and

$$S_{33.1} = S_{33} - S_{31}S_{11}^{-1}S_{13}, \quad S_{22.1} = S_{22} - S_{21}S_{11}^{-1}S_{12}, \tag{1.47}$$

$$S_{32.1} = S_{32} - S_{31}S_{11}^{-1}S_{12}, \quad S_{23.1} = S_{23} - S_{21}S_{11}^{-1}S_{13}. \tag{1.48}$$

Using (1.45), (1.47), and (1.48), the coefficient matrix $\Pi_{32} = \Theta_{32}$ in (1.44) can be expressed as

$$\Pi_{32} = S_{32.1}S_{22.1}^{-1}. \tag{1.49}$$

We summarize the previous results in the following theorem.

Theorem 1.9 *Suppose we want to predict the random vector Y_3 and we observe the random vector Y_1. If a new observation, Y_2, becomes available, we can update the best linear predictor of Y_3 based on Y_1, $E^*(Y_3|Y_1)$, and its MSE using the formulae*

$$\begin{aligned} E^*(Y_3|Y_1, Y_2) &= E^*(Y_3|Y_1) + S_{32.1}S_{22.1}^{-1}[Y_2 - E^*(Y_2|Y_1)] \\ &= S_{31}S_{11}^{-1}Y_1 + S_{32.1}S_{22.1}^{-1}(Y_2 - S_{21}S_{11}^{-1}Y_1) \end{aligned}$$

and

$$\begin{aligned} MSE[E^*(Y_3|Y_1, Y_2)] &= MSE[E^*(Y_3|Y_1)] - S_{32.1}\{MSE[E^*(Y_2|Y_1)]\}^{-1}S_{23.1} \\ &= S_{33.1} - S_{32.1}S_{22.1}^{-1}S_{23.1}, \end{aligned}$$

where $MSE[E^*(Y_3|Y_1)] = S_{33.1}$, $MSE[E^*(Y_2|Y_1)] = S_{22.1}$, and $S_{22.1}$, $S_{23.1}$, $S_{32.1}$ and $S_{33.1}$ are given by (1.47) and (1.48).

The coefficient matrix Π_{32} in (1.44) also has the following important interpretation.

Proposition 1.8 *Π_{32} in (1.44) is the coefficient matrix of the orthogonal projection of the residual $Y_3 - E^*(Y_3|Y_1)$ onto the residual $Y_2 - E^*(Y_2|Y_1)$. Therefore, $\Pi_{32} = S_{32.1}S_{22.1}^{-1}$, where $S_{22.1}$ and $S_{32.1}$ are given by (1.47) and (1.48).*

Proof Π_{32} satisfies $\Pi_{32}(Y_2 - E^*(Y_2|Y_1)) = E^*(Y_3|E_2)$ in (1.44). Since this last predictor is also equal to $E^*(Y_3 - E^*(Y_3|Y_1)|E_2)$ because $E^*(Y_3|Y_1) = \Theta_{31}E_1$ is orthogonal to E_2, the first part of the proposition follows. The formula for Π_{32} is (1.49). □

1.8.2 The Law of Iterated Orthogonal Projection

The following theorem is an immediate consequence of the sequential update formula (1.44).

Theorem 1.10 (Law of Iterated Orthogonal Projection) *Given three random vectors, Y_1, Y_2, and Y_3, the following relation holds*

$$E^*[E^*(Y_3|Y_2, Y_1)|Y_1] = E^*(Y_3|Y_1).$$

Proof The result follows immediately from (1.44) by projecting both sides onto Y_1 and considering that $Y_2 - E^*(Y_2|Y_1)$ is orthogonal to Y_1. □

1.8.3 The Forward and Backward Prediction Problems

Suppose the zero-mean vector random process $\{Y_t\}$ and for fixed t and increasing k, consider the forward and backward innovations

$$\begin{aligned} E_{t,k} &= Y_t - E^*(Y_t|Y_{t-1}, \ldots, Y_{t-k}) \qquad (1.50)\\ &= Y_t - \Phi_{t,k1}Y_{t-1} - \cdots - \Phi_{t,kk}Y_{t-k} \\ R_{t-1,k} &= Y_{t-k-1} - E^*(Y_{t-k-1}|Y_{t-1}, \ldots, Y_{t-k}) \qquad (1.51)\\ &= Y_{t-k-1} - \Phi^*_{t,kk}Y_{t-1} - \cdots - \Phi^*_{t,k1}Y_{t-k}. \end{aligned}$$

These are the forward and backward prediction problems, respectively. When the process $\{Y_t\}$ is stationary, these problems are associated with autoregressive model fitting, as we will see later in the following chapter. The following recursive formulae are extremely important in the time series context.

Theorem 1.11 (Order Recursive Prediction) *Suppose the sequence of zero-mean random vectors $(Y_1', \ldots, Y_t')'$ and, for fixed t and increasing $k = 1, 2, \ldots, t-1$, consider the forward and backward innovations (1.50) and (1.51). Denote $D_{t,k} = Var(E_{t,k})$, $Q_{t-1,k} = Var(R_{t-1,k})$ and $\Delta_{t,k} = Cov(E_{t,k}, R_{t-1,k})$ and let $\Phi_{(t,k)} = (\Phi_{t,k1}, \ldots, \Phi_{t,kk})$, $\Phi^*_{(t,k)} = (\Phi^*_{t,kk}, \ldots, \Phi^*_{t,k1})$, $\Gamma_{(t,k)} = (S_{t,t-1}, \ldots, S_{t,t-k})$, and $\Gamma^*_{(t,k)} = (S_{t-k,t}, \ldots, S_{t-1,t})$, where $S_{ij} = Cov(Y_i, Y_j)$. Then, the following recursions hold*

$$\begin{aligned} \Phi_{t,kk} &= \Delta_{t,k-1}Q^{-1}_{t-1,k-1} = \left[S_{t,t-k} - \Gamma_{(t,k-1)}\Phi^{*\prime}_{(t,k-1)}\right]Q^{-1}_{t-1,k-1} \\ &= \left[S_{t,t-k} - \Phi_{(t,k-1)}\Gamma^{*\prime}_{(t,k-1)}\right]Q^{-1}_{t-1,k-1} \\ (\Phi_{t,k1}, \ldots, \Phi_{t,k-1,k}) &= \Phi_{(t,k-1)} - \Phi_{t,kk}\Phi^*_{(t,k-1)} \\ \Phi^*_{t,kk} &= \Delta'_{t,k-1}D^{-1}_{t,k-1} = \left[S_{t-k,t} - \Gamma^*_{(t,k-1)}\Phi'_{(t,k-1)}\right]D^{-1}_{t,k-1} \\ &= \left[S_{t-k,t} - \Phi^*_{(t,k-1)}\Gamma'_{(t,k-1)}\right]D^{-1}_{t,k-1} \\ (\Phi^*_{t,k-1,k}, \ldots, \Phi^*_{t,k1}) &= \Phi^*_{(t,k-1)} - \Phi^*_{t,kk}\Phi_{(t,k-1)} \\ D_{t,k} &= D_{t,k-1} - \Delta_{t,k-1}Q^{-1}_{t-1,k-1}\Delta'_{t,k-1} = (I - \Phi_{t,kk}\Phi^*_{t,kk})D_{t,k-1} \\ Q_{t,k} &= Q_{t-1,k-1} - \Delta'_{t,k-1}D^{-1}_{t,k-1}\Delta_{t,k-1} = (I - \Phi^*_{t,kk}\Phi_{t,kk})Q_{t-1,k-1} \\ E_{t,k} &= E_{t,k-1} - \Delta_{t,k-1}Q^{-1}_{t-1,k-1}R_{t-1,k-1} = E_{t,k-1} - \Phi_{t,kk}R_{t-1,k-1} \\ R_{t,k} &= R_{t-1,k-1} - \Delta'_{t,k-1}D^{-1}_{t,k-1}E_{t,k-1} = R_{t-1,k-1} - \Phi^*_{t,kk}E_{t,k-1}, \end{aligned}$$

initialized with $E_{t,0} = Y_t$, $R_{t-1,0} = Y_{t-1}$, $\Phi_{t,11} = S_{t,t-1}S^{-1}_{t-1,t-1}$, $D_{t,0} = S_{tt}$, $\Phi^*_{11} = S_{t-1,t}S^{-1}_{tt}$, $\Delta_{t,0} = S_{t,t-1}$ *and* $Q_{t-1,0} = S_{t-1,t-1}$.

Proof Let $Z_3 = Y_t$, $Z_2 = Y_{t-k}$ and $Z_1 = (Y'_{t-1}, \ldots, Y'_{t-k+1})'$ and apply formula (1.44). Then,

$$\begin{aligned} E^*(Z_3|Z_1, Z_2) &= E^*(Z_3|Z_1) + \Pi_{32}(Z_2 - E^*(Z_2|Z_1)) \\ &= \Theta_{31}Z_1 + \Pi_{32}(Z_2 - \Theta_{21}Z_1) \\ &= \Pi_{31}Z_1 + \Pi_{32}Z_2, \end{aligned}$$

where $\Pi_{31} = \Theta_{31} - \Theta_{32}\Theta_{21}$, $\Pi_{32} = \Theta_{32}$ and Θ_{21}, Θ_{31} and Θ_{32} are given by (1.45). From this, it is obtained that $\Theta_{31} = \Phi_{(t,k-1)}$, $\Theta_{21} = \Phi^*_{(t,k-1)}$, $(\Pi_{31}, \Pi_{32}) = \Phi_{(k)}$, $\Pi_{31} = (\Phi_{t,k1}, \ldots, \Phi_{t,k-1,k})$, and $\Theta_{32} = \Phi_{t,kk}$. Thus,

$$(\Phi_{t,k1}, \ldots, \Phi_{t,k-1,k}) = \Phi_{(t,k-1)} - \Phi_{t,kk}\Phi^*_{(t,k-1)}.$$

By Proposition 1.8,

$$\Pi_{32} = \Phi_{t,kk} = \Delta_{t,k-1}Q^{-1}_{t-1,k-1}$$

and, from (1.46), we can write

$$D_{t,k} = D_{t,k-1} - \Delta_{t,k-1}Q^{-1}_{t-1,k-1}\Delta'_{t,k-1}.$$

The expression (1.45) for $\Theta_{32} = \Phi_{t,kk}$ yields

$$\Phi_{t,kk} = \left[S_{t,t-k} - (S_{t,t-1}, \ldots, S_{t,t-k+1})\Phi^{*'}_{(t,k-1)}\right] Q^{-1}_{t-1,k-1}.$$

Interchanging the roles of Z_3 and Z_2 in the previous argument yields the other recursions. The only thing that remains to be proved is the formula $D_{t,k} = D_{t,k-1} - \Delta_{t,k-1}Q^{-1}_{t-1,k-1} \times \Delta'_{t,k-1} = (I - \Phi_{t,kk}\Phi^*_{t,kk})D_{t,k-1}$ and its analogue for $Q_{t,k}$. But, because $\Phi_{t,kk} = \Delta_{t,k-1} \times Q^{-1}_{t-1,k-1}$ and $\Phi^*_{t,kk}D_{t,k-1} = \Delta'_{t,k-1}$, we have $D_{t,k} = D_{t,k-1} - \Phi_{t,kk}\Delta'_{t,k-1} = D_{t,k-1} - \Phi_{t,kk}\Phi^*_{t,kk}D_{t,k-1}$ and the theorem is proved. □

Corollary 1.5 *Under the assumptions and with the notation of the previous theorem, the eigenvalues of the matrix* $\Phi_{t,kk}\Phi^*_{t,kk}$ *are the squared canonical correlations between the residuals* $E_{t,k-1}$ *and* $R_{t-1,k-1}$. *Thus, they are the squared partial canonical correlations between* Y_t *and* Y_{t-k} *with respect to* $\{Y_{t-1}, \ldots, Y_{t-k+1}\}$. *Moreover, these partial canonical correlations are the singular values of* $Cov(\overline{E}_{t,k-1}, \overline{R}_{t-1,k-1}) = D^{-1/2}_{t,k-1}\Delta_{t,k-1}Q^{-1/2'}_{t-1,k-1}$, *where* $\overline{E}_{t,k-1}$ *and* $\overline{R}_{t-1,k-1}$ *are the standardized residuals* $\overline{E}_{t,k-1} = D^{-1/2}_{t,k-1}E_{t,k-1}$ *and* $\overline{R}_{t-1,k-1} = Q^{-1/2}_{t-1,k-1}R_{t-1,k-1}$.

Proof The first part of the corollary follows from the formula

$$\begin{aligned}\Phi_{t,kk}\Phi^*_{t,kk} &= \Delta_{t,k-1}Q^{-1}_{t-1,k-1}\Delta'_{t,k-1}D^{-1}_{t,k-1}\\ &= \Sigma_{ER}\Sigma^{-1}_{RR}\Sigma_{RE}\Sigma^{-1}_{EE},\end{aligned}$$

where $\Sigma_{ER} = \text{Cov}(E_{t,k-1}, R_{t-1,k-1})$, $\Sigma_{RE} = \Sigma'_{ER}$, $\Sigma_{RR} = \text{Var}(R_{t-1,k-1})$, and $\Sigma_{EE} = \text{Var}(E_{t,k-1})$. To prove the second part, consider that the eigenvalues of $\det(\Phi_{t,kk}\Phi^*_{t,kk} - \lambda I) = \det(AA' - \lambda I)$, where $A = D^{-1/2}_{t,k-1}\Delta_{t,k-1}Q^{-1/2'}_{t-1,k-1}$. □

1.8.4 Partial and Multiple Correlation Coefficients

To avoid unnecessary complications, in this section we will assume that all random variables have zero mean. By Remark 1.2, there is no loss of generality in doing this because we can always work with centered variables if the means are not zero.

Suppose two random variables, Y_3 and Y_2, and a random vector Y_1. The **partial correlation coefficient** $r_{32.1}$ of Y_3 and Y_2 with respect to Y_1 measures the correlation of these two variables after having eliminated the influence due to Y_1. More specifically, $r_{32.1}$ is defined

$$r_{32.1} = \text{Corr}(Y_3 - E^*(Y_3|Y_1), Y_2 - E^*(Y_2|Y_1)).$$

By Proposition 1.8, the coefficient Π_{32} in (1.44) coincides with the coefficient of the orthogonal projection of the residual $Y_3 - E^*(Y_3|Y_1)$ onto the residual $Y_2 - E^*(Y_2|Y_1)$. Thus, $\Pi_{32} = \text{Cov}[Y_3 - E^*(Y_3|Y_1), Y_2 - E^*(Y_2|Y_1)]D_2^{-1}$ and

$$r_{32.1} = \Pi_{32}\frac{\sqrt{\text{MSE}(E^*(Y_2|Y_1))}}{\sqrt{\text{MSE}(E^*(Y_3|Y_1))}}. \tag{1.52}$$

Theorem 1.9 and Proposition 1.8 yield

$$r_{32.1} = \frac{S_{32.1}}{\sqrt{S_{22.1}}\sqrt{S_{33.1}}}.$$

The coefficient $r_{32.1}$ can be obtained by means of the sequential updating scheme of the Sect. 1.8.1. That is, first compute the predictor $E^*(Y_3|Y_1)$ and its MSE. Then, compute the predictor $E^*(Y_2|Y_1)$, its MSE, and the coefficient Π_{32}. Finally, compute $r_{32.1}$ using (1.52).

It is also possible to compute $r_{32.1}$ as a by-product of the computation of the orthogonal projection $E^*(Y_3|Y_1, Y_2) = \Pi_{31}Y_1 + \Pi_{32}Y_2$ using the autoregressive representation. To see this, note that these recursions give $D_2 = \text{MSE}(E^*(Y_2|Y_1))$, Π_{32} and $D_3 = S_{33} - S_{31}D_1^{-1}S_{13} - \Pi_{32}D_2\Pi'_{32}$. Finally, since $\text{MSE}(E^*(Y_3|Y_1)) = S_{33} - S_{31}D_1^{-1}S_{13}$, it is obtained that $\text{MSE}(E^*(Y_3|Y_1)) = D_3 + \Pi_{32}D_2\Pi'_{32}$ and we have all the necessary quantities to compute $r_{32.1}$ in (1.52). We have thus proved the following proposition.

Proposition 1.9 *Suppose two random variables, Y_3 and Y_2, and a random vector Y_1. Then, using the autoregressive representation, the partial correlation coefficient, $r_{32.1}$, of Y_3 and Y_2 with respect to Y_1 is given by*

$$r_{32.1} = \Pi_{32} \frac{\sqrt{D_2}}{\sqrt{D_3 + \Pi_{32} D_2 \Pi'_{32}}}.$$

An alternative way to compute (Π_{31}, Π_{32}) is to apply directly the formula that gives the best linear prediction, $(\Pi_{31}, \Pi_{32}) = (S_{31}, S_{32})\text{Var}^{-1}[(Y'_1, Y'_2)']$. In this case, it is not difficult to verify using the autoregressive representation that

$$[\Pi_{31}, \Pi_{32}] = [S_{31}, S_{32}] \begin{bmatrix} D_1^{-1} + \Pi'_{21} D_2^{-1} \Pi_{21} & -\Pi'_{21} D_2^{-1} \\ -D_2^{-1} \Pi_{21} & D_2^{-1} \end{bmatrix},$$

where $\Pi_{21} = S_{21} S_{11}^{-1}$ because, by Corollary 1.2, $\Pi_{21} = \Theta_{21}$. Using again the autoregressive representation, it can be verified that computing D_3 directly by the formula $D_3 = S_{33} - (S_{31}, S_{32})\text{Var}^{-1}[(Y'_1, Y'_2)'](S_{31}, S_{32})'$ yields $D_3 = S_{33} - S_{31} D_1^{-1} S_{13} - \Pi_{32} D_2 \Pi'_{32}$. This allows for the computation of $r_{32.1}$ as in the previous proposition.

Example 1.2 (Continued) In this case, we define $\alpha(1) = \text{Corr}(Y_2, Y_1)$ and

$$\alpha(k) = \text{Corr}(Y_{k+1} - E^*(Y_{k+1}|Y_k, \ldots, Y_2), Y_1 - E^*(Y_1|Y_2, \ldots, Y_k)), \quad , k = 1, 2, \ldots$$

In order to use formula (1.52) to compute $\alpha(k)$ when $k > 1$, we further define $X_1 = (Y_2, \ldots, Y_k)'$, $X_2 = Y_1$ and $X_3 = Y_{k+1}$. It is shown in Problem 1.8 that, applying the autoregressive representation to the sequence $\{X_1, X_2, X_3\}$, it is obtained that

$$\alpha(k) = -\frac{(-\theta)^k}{1 + \theta^2 + \cdots + \theta^{2k}}, \quad k = 1, 2, \ldots$$

◇

Given a random variable Y_2 and a random vector Y_1, the **multiple correlation coefficient** $r_{2.1}$ of Y_2 with respect to Y_1 is defined as the correlation between Y_2 and $E^*(Y_2|Y_1)$,

$$r_{2.1} = \text{Corr}(Y_2, E^*(Y_2|Y_1)).$$

Proposition 1.10 *Suppose a random variable Y_2 and a random vector Y_1. Then, using the innovations algorithm, the multiple correlation coefficient, $r_{2.1}$, of Y_2 with* *respect to Y_1 is given by*

$$r_{2.1} = \sqrt{\frac{\Theta_{21}D_1\Theta_{21}'}{\Theta_{21}D_1\Theta_{21}' + D_2}} = \sqrt{\frac{S_{21}S_{11}^{-1}S_{12}}{S_{22}}}$$

In addition, the square of $r_{2.1}$ satisfies

$$r_{2.1}^2 = 1 - \frac{D_2}{\Theta_{21}D_1\Theta_{21}' + D_2} \tag{1.53}$$

$$= 1 - \frac{S_{22} - S_{21}S_{11}^{-1}S_{12}}{S_{22}}$$

$$= S_{22}^{-1}S_{21}S_{11}^{-1}S_{12}. \tag{1.54}$$

Proof Since $Y_2 = E^*(Y_2|Y_1) + E_2$ and $E^*(Y_2|Y_1)$ is orthogonal to E_2, $\mathrm{Cov}(Y_2, E^*(Y_2|Y_1)) = \Theta_{21}D_1\Theta_{21}'$, where $E^*(Y_2|Y_1) = \Theta_{21}Y_1$, $\Theta_{21} = S_{21}D_1^{-1}$ and $D_1 = S_{11}$. Thus,

$$r_{2.1} = \frac{\Theta_{21}D_1\Theta_{21}'}{\sqrt{\Theta_{21}D_1\Theta_{21}'}\sqrt{\Theta_{21}D_1\Theta_{21}' + D_2}} = \sqrt{\frac{\Theta_{21}D_1\Theta_{21}'}{\Theta_{21}D_1\Theta_{21}' + D_2}}$$

$$= \sqrt{\frac{S_{21}S_{11}^{-1}S_{12}}{S_{22}}}.$$

The statement about the square of $r_{2.1}$ is evident. □

The expression (1.53) implies that $0 \le r_{2.1}^2 \le 1$. Two extreme cases can occur. If $r_{2.1}^2 = 1$, then $E_2 = 0$ and $Y_2 = E^*(Y_2|Y_1)$. On the other hand, if $r_{2.1}^2 = 0$, then $Y_2 = E_2$ and $E^*(Y_2|Y_1) = 0$.

The multiple correlation coefficient has the following property.

Proposition 1.11 *Given a random variable Y_2 and a random vector Y_1, the multiple correlation coefficient of Y_2 with respect to Y_1 coincides with the canonical correlation between Y_2 and Y_1. Therefore, the linear combination ΛY_1 that has maximum correlation with Y_2 is $\Theta_{21}Y_1 = E^*(Y_2|Y_1)$.*

Proof The proposition is a consequence of (1.54). □

Remark 1.7 Box & Tiao (1977) proposed in the time series context a measure of predictability that can be considered as a generalization of the multiple correlation coefficient. Letting $\{Y_t\}$ be a zero mean vector random process, suppose that the variables $\{Y_{t-1}, \ldots, Y_{t-k}\}$ contain enough information about the past of the process to predict the future, Y_t. Letting the predictor be $\widehat{Y}_{t,k} = E^*(Y_t|Y_{t-1}, \ldots, Y_{t-k})$, we can write

$$Y_t = \widehat{Y}_{t,k} + E_{t,k},$$

where $E_{t,k}$ coincides with that given by the recursions of Theorem 1.11, $\widehat{Y}_{t,k} = S_{21}S_{11}^{-1}$, $S_{21} = \mathrm{Cov}(Y_t, Y_{t-1:t-k})$, $S_{11} = \mathrm{Var}(Y_{t-1:t-k})$, and $Y_{t-1:t-k} = (Y'_{t-1}, \ldots, Y'_{t-k})'$. Then, the covariance matrices of Y_t, $\widehat{Y}_{t,k}$, and $E_{t,k}$ satisfy

$$S_{22} = S_{21}S_{11}^{-1}S_{12} + D_{t,k},$$

where $\mathrm{Var}(Y_t) = S_{22}$, $\mathrm{Var}(\widehat{Y}_{t,k}) = S_{21}S_{11}^{-1}S_{12}$, and $\mathrm{Var}(E_{t,k}) = D_{t,k}$. The measure of predictability proposed by Box & Tiao (1977) is

$$\Pi_{t,k} = \mathrm{Var}^{-1}(Y_t)\mathrm{Var}(\widehat{Y}_{t,k}) = S_{22}^{-1}S_{21}S_{11}^{-1}S_{12}.$$

Clearly, the eigenvalues of $\Pi_{t,k}$ are the squared canonical correlations between Y_t and $\widehat{Y}_{t,k}$. Thus, if $1 \geq \rho_1^2 \geq \cdots \geq \rho_n^2 \geq 0$ are the ordered eigenvalues of $\Pi_{t,k}$ and a_i, $i = 1, \ldots, n$, are eigenvectors associated with ρ_i^2, the linear combinations $a_i'Y_t$ can be considered to be ordered from most to least predictable because among all possible linear combinations of Y_t their correlations with all possible linear combinations of $\widehat{Y}_{t,k}$ vary from a maximum of ρ_1 to a minimum of ρ_n. ◇

1.9 Introduction to the Kalman Filter

We have seen in the previous sections several algorithms to obtain the innovations given a sequence of random vectors, $\{Y_t\}$. Some algorithms were based on the application of the **Gram–Schmidt** orthogonalization procedure of Euclidean spaces to the sequence $\{Y_t\}$. When the data have state space structure, we saw a covariance based state space algorithm. Another state space algorithm that can be used to compute the innovations is the celebrated **Kalman filter**. This algorithm will be described in detail later in the book. In this section, we will derive the Kalman filter from first principles for the case of a state space model with time invariant system matrices.

Suppose that the process $\{Y_t\}$ follows the state space model (1.37) and (1.38), where

$$E\left\{\begin{bmatrix} u_t \\ v_t \end{bmatrix}\left[u_s', v_s'\right]\right\} = \begin{bmatrix} Q & S \\ S' & R \end{bmatrix}\delta_{ts}$$

and $\text{Var}(x_1) = \Pi$. Then, the Kalman filter is a set of recursions to compute the linear projections $\hat{x}_{t|t-1} = E^*(x_t|Y_{t-1}, \ldots, Y_1)$ and the innovations $E_t = Y_t - E^*(Y_t|Y_{t-1}, \ldots, Y_1)$, as well as the $\text{MSE}(\hat{x}_{t|t-1})$ and the $\text{Var}(E_t) = D_t$. Given the data $Y = (Y_1', \ldots, Y_n')'$, the Kalman filter recursively provides the vector of innovations $E = (E_1', \ldots, E_n')'$ in the form $E = L^{-1}Y$ without explicitly computing the matrix L^{-1}, where L is a lower triangular matrix with ones in the main diagonal such that $\text{Var}(Y) = LDL'$ and $D = \text{diag}(D_1, D_2, \ldots, D_n)$. In addition, the Kalman filter can also be applied when the state space equations that represent the data structure are time variant.

In the following, we derive the Kalman filter corresponding to (1.37) and (1.38) from first principles, where we assume that $\text{Var}(Y)$ is positive definite. As we know, this implies that the covariance matrices of the innovations are nonsingular.

First note the equality $E_t = Y_t - E^*(Y_t|Y_{t-1}, \ldots, Y_1) = Y_t - H\hat{x}_{t|t-1}$, that follows directly from the observation equation (1.38). By Proposition 1.7, $\hat{x}_{t+1|t} = E^*(x_{t+1}|Y_t, \ldots, Y_1) = E^*(x_{t+1}|E_t, Y_{t-1}, \ldots, Y_1)$ and, by Proposition 1.5 , $\hat{x}_{t+1|t} = E^*(x_{t+1}|Y_{t-1}, \ldots, Y_1) + E^*(x_{t+1}|E_t)$ because the innovation $E_t = Y_t - H\hat{x}_{t|t-1}$ is orthogonal to the variables Y_s, $s < t$. Then, from the transition equation (1.37) we have

$$\begin{aligned} E^*(x_{t+1}|Y_{t-1}, \ldots, Y_1) &= FE^*(x_t|Y_{t-1}, \ldots, Y_1) \\ &= F\hat{x}_{t|t-1}, \end{aligned}$$

where $\hat{x}_{t|t-1} = E^*(x_t|Y_{t-1}, \ldots, Y_1)$. Thus,

$$\hat{x}_{t+1|t} = F\hat{x}_{t|t-1} + K_tE_t, \tag{1.55}$$

where $K_t = \text{Cov}(x_{t+1}, E_t)\text{Var}^{-1}(E_t)$. To compute K_t from first principles, consider the error vector $\tilde{x}_t = x_t - \hat{x}_{t|t-1}$. Then, subtracting Eq. (1.55) from (1.37) and using the equality $E_t = Y_t - H\hat{x}_{t|t-1}$, it is obtained that

$$\tilde{x}_{t+1} = (F - K_tH)\tilde{x}_t + Gu_t - K_tv_t. \tag{1.56}$$

We set $\hat{x}_{1|0} = 0$, $\tilde{x}_1 = x_1$ and $\text{Var}(\hat{x}_{1|0}) = \Pi = \text{Var}(x_1)$. This implies the correct value for the first innovation, namely $E_1 = Y_1$. Equation (1.56) shows that $\tilde{x}_t$ can be expressed as a linear combination of the random vectors $\{x_1, u_1, \ldots, u_{t-1}, v_1, \ldots, v_{t-1}\}$. Since both u_t and v_t are uncorrelated with x_1 and with u_s and v_s for $s < t$, it follows that

$$E(\tilde{x}_tu_t') = 0 \quad \text{and} \quad E(\tilde{x}_tv_t') = 0. \tag{1.57}$$

According to the definition of orthogonal projection, our aim is to compute K_t so that $\text{tr}(P_{t+1})$ is minimum, where $P_{t+1} = E(\tilde{x}_{t+1}\tilde{x}_{t+1}')$ and we assume that $\hat{x}_{t|t-1}$ and

$P_t = E(\tilde{x}_t\tilde{x}_t')$ are known. It follows from (1.56) and (1.57) that

$$\begin{aligned} P_{t+1} &= (F - K_tH)P_t(F - K_tH)' + [G \quad -K_t]\begin{bmatrix} Q & S \\ S' & R \end{bmatrix}\begin{bmatrix} G' \\ -K_t' \end{bmatrix} \\ &= [I \quad K_t]\begin{bmatrix} FP_tF' + GQG' & -(FP_tH' + GS) \\ -(H'P_tF' + S'G') & R + HP_tH' \end{bmatrix}\begin{bmatrix} I \\ K_t' \end{bmatrix}. \end{aligned}$$

From $E_t = Y_t - H\hat{x}_{t|t-1}$, using (1.22), it is obtained that

$$E_t = H\tilde{x}_t + v_t \qquad \text{and} \qquad \mathrm{Var}(E_t) = HP_tH' + R.$$

Since, by assumption, the innovations have nonsingular covariance matrices, the previous equation shows that $HP_tH' + R$ is nonsingular. Letting

$$\begin{bmatrix} FP_tF' + GQG' & -(FP_tH' + GS) \\ -(H'P_tF' + S'G') & R + HP_tH' \end{bmatrix} = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix}$$

and using the block triangular factorization (see Problem 1.10)

$$\begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} = \begin{bmatrix} I & S_{12}S_{22}^{-1} \\ 0 & I \end{bmatrix}\begin{bmatrix} S_{11} - S_{12}S_{22}^{-1}S_{21} & 0 \\ 0 & S_{22} \end{bmatrix}\begin{bmatrix} I & 0 \\ S_{22}^{-1}S_{21} & I \end{bmatrix},$$

we get

$$P_{t+1} = S_{11} - S_{12}S_{22}^{-1}S_{21} + (K_t + S_{12}S_{22}^{-1})S_{22}(K_t + S_{12}S_{22}^{-1})'.$$

Clearly, $\mathrm{tr}(P_{t+1})$ is minimized if we set $K_t = -S_{12}S_{22}^{-1}$ or

$$K_t = (FP_tH' + GS)D_t^{-1},$$

where

$$D_t = R + HP_tH'.$$

With this choice, the minimum error covariance matrix P_{t+1} is

$$\begin{aligned} P_{t+1} &= S_{11} - S_{12}S_{22}^{-1}S_{21} \\ &= FP_tF' + GQG' - K_tD_tK_t'. \end{aligned}$$

Summarizing, we have got the following recursions

$$\begin{aligned} E_t &= Y_t - H\hat{x}_{t|t-1}, \quad D_t = HP_tH' + R \\ K_t &= (FP_tH' + GS)D_t^{-1}, \quad \hat{x}_{t+1|t} = F\hat{x}_{t|t-1} + K_tE_t \\ P_{t+1} &= FP_tF' + GQG' - (FP_tH' + GS)D_t^{-1}(FP_tH' + GS)' \\ &= (F - K_tH)P_tF' + (GQ - K_tS')G', \end{aligned}$$

initialized with $\hat{x}_{1|0} = 0$ and $\mathrm{Var}(\hat{x}_{1|0}) = \Pi$ that correspond to the Kalman filter for the case in which the system matrices are time invariant. The alert reader will have noticed that in the previous discussion $S_{11} = \mathrm{Cov}(x_{t+1})$, $S_{12} = -\mathrm{Cov}(x_{t+1}, E_t)$ and $S_{22} = \mathrm{Var}(E_t)$.

Remark 1.8 It is to be noticed that the equations $Y_t = H\hat{x}_{t|t-1} + E_t$ and (1.55) of the Kalman filter coincide with the corresponding equations given by Theorem 1.8. However, the quantities that appear in these last equations are computed using covariance data only. ◇

Example 1.2 (Continued) One possible set of Eqs. (1.37) and (1.38) is

$$\begin{aligned} x_{t+1} &= \theta A_t \\ Y_t &= x_t + A_t, \qquad t = 1, 2, \ldots, n. \end{aligned}$$

Thus, in this case, $F = 0$, $H = 1$, $G = \theta$, $u_t = v_t = A_t$, $\Omega = \mathrm{Var}(x_1) = \theta^2\sigma^2$, and $Q = R = S = \sigma^2$. Letting the initial conditions be the unconditional mean and variance of the initial state vector, $\hat{x}_{1|0} = 0$ and $P_1 = \theta^2\sigma^2$, the Kalman filter recursions are

$$\begin{aligned} E_t &= Y_t - \hat{x}_{t|t-1}, \qquad D_t = P_t + \sigma^2 \\ K_t &= \theta\sigma^2/D_t, \qquad \hat{x}_{t+1|t} = K_tE_t \\ P_{t+1} &= \theta^2\sigma^2 - K_tD_tK_t. \end{aligned}$$

◇

Example 1.3 (Continued) One possible set of Eqs. (1.37) and (1.38) is

$$\begin{aligned} x_{t+1} &= \phi x_t + A_{t+1} \\ Y_t &= x_t, \qquad t = 1, 2, \ldots, n. \end{aligned}$$

Thus, $F = \phi$, $H = 1$, $G = 1$, $u_t = A_{t+1}$, $v_t = 0$, $\Omega = \mathrm{Var}(x_1) = \sigma^2/(1 - \phi^2)$, $Q = \sigma^2$ and $R = S = 0$. Letting the initial conditions be the unconditional mean and variance of the initial state vector, $\hat{x}_{1|0} = 0$ and $P_1 = \sigma^2/(1 - \phi^2)$, the Kalman filter recursions give $E_1 = Y_1$, $D_1 = \sigma^2/(1-\phi^2)$, $K_1 = \phi$, $\hat{x}_{2|1} = \phi Y_1$, $P_2 = \sigma^2$, and

$$\begin{aligned} E_t &= Y_t - \hat{x}_{t|t-1} = Y_t - \phi Y_{t-1}, \qquad D_t = \sigma^2 \\ K_t &= \phi, \qquad \hat{x}_{t+1|t} = \phi \hat{x}_{t|t-1} + K_t E_t = \phi Y_t \\ P_{t+1} &= \sigma^2, \quad t = 2, 3, \ldots, n. \end{aligned}$$

◇

The matrix K_t given by the Kalman filter corresponding to (1.37) and (1.38) is called the **"Kalman gain"** because it represents the improvement in the prediction of x_{t+1} when a new observation Y_t is incorporated into the sample. For the Example 1.2, the Kalman gain coincides with the K_t given by Eqs. (1.21) and (1.22). In fact, when there is structure in the data, the matrices F, H, K_t, E_t, and $D_t = \text{Var}(E_t)$ of these last equations coincide with those of the Kalman filter corresponding to Eqs. (1.37) and (1.38). Thus, in this case the matrices (1.24) and (1.25) given by both procedures also coincide. However, the P_t matrices given by the Kalman filter do not coincide even in this case with the Σ_t matrices given by Eqs. (1.21) and (1.22). The precise relationship is that the x_t of Eqs. (1.21) and (1.22) is equal to the $\hat{x}_{t|t-1}$ given by the Kalman filter and that the Σ_t of Eqs. (1.21) and (1.22) is equal to $\text{Var}(\hat{x}_{t|t-1})$, whereas the P_t given by the Kalman filter is equal to $\text{MSE}(\hat{x}_{t|t-1}) = \text{Var}(x_t - \hat{x}_{t|t-1})$, where x_t is the state vector of Eqs. (1.37) and (1.38).

One final point that should be mentioned is the conditions under which the Kalman filter recursions corresponding to the time invariant state space equations (1.37) and (1.38), and also the recursions (1.21) and (1.22), reach a steady state. It will be shown later in the book that under fairly general conditions these recursions do reach a steady state. In the case of Example 1.2, it can be shown that $D_1 \geq D_2 \geq \ldots \geq D_t \geq 0$, implying $D_t \to \sigma^2$, $\Sigma_t \to 0$, and $K_t \to \theta = G$ as $t \to \infty$.

1.10 Linear Regression and Ordinary Least Squares

Suppose that y is a random variable and that x is a random vector, both with finite second moments, and define $u = y - E^*(y|x)$. Then, u is orthogonal to x and we can write

$$y = \beta' x + u, \tag{1.58}$$

where $\beta' x$ coincides with the orthogonal projection, $E^*(y|x)$, so that $\beta' = S_{yx} S_{xx}^{-1}$. Conversely, if in the model (1.58) we assume that x and u are orthogonal, then $S_{yx} = \beta' S_{xx}$ and $\beta' x = E^*(y|x)$.

The model (1.58), under the assumption that β is constant and u is orthogonal to x, is called **linear regression model**.

Suppose that we have a random sample (y_t, x_t'), $t = 1, \ldots, n$, corresponding to the linear regression model (1.58) and that we want to estimate β. It seems natural that we use the empirical distribution of the sample, that assigns probability $1/n$ to each point (y_t, x_t), and that we proceed to compute the best linear predictor as we have done using population moments. This procedure is what we call **the sample analogue to best linear prediction**.

More specifically, according to this sample analogue, we have a random vector $(\xi, \chi')'$ that takes values $(y_t, x_t')'$, $t = 1, \ldots, n$, with probability $1/n$. The matrix of noncentered second moments of $(\xi, \chi')'$ is

$$E\left\{\begin{bmatrix}\xi\\ \chi\end{bmatrix}[\xi, \chi']\right\} = \begin{bmatrix}S_{\xi\xi} & S_{\xi\chi}\\ S_{\chi\xi} & S_{\chi\chi}\end{bmatrix},$$

where $S_{\xi\xi} = \sum_{t=1}^n y_t^2/n$, $S_{\xi\chi} = \sum_{t=1}^n y_t x_t'/n$, $S_{\chi\xi} = \sum_{t=1}^n x_t y_t/n$, and $S_{\chi\chi} = \sum_{t=1}^n x_t x_t'/n$.

By Proposition 1.1, if $S_{\chi\chi}$ is nonsingular, the best linear predictor of ξ based on χ is

$$E^*(\xi|\chi) = S_{\xi\chi}S_{\chi\chi}^{-1}\chi. \tag{1.59}$$

Letting $\hat{\beta}' = S_{\xi\chi}S_{\chi\chi}^{-1}$, we see that $\hat{\beta}$ satisfies the normal equations

$$S_{\chi\chi}\hat{\beta} = S_{\chi\xi}$$

or, equivalently,

$$\left(\sum_{t=1}^n x_t x_t'\right)\hat{\beta} = \sum_{t=1}^n x_t y_t.$$

Letting $X = (x_1, \ldots, x_n)'$ and $y = (y_1, \ldots, y_n)'$, we can write the sample in a more compact form as (y, X) and the normal equations as

$$(X'X)\hat{\beta} = X'y.$$

By Proposition 1.2, this system of linear equations is always consistent, so that $E^*(\xi|\chi)$ exists even if $S_{\chi\chi}$ is singular. It is to be noted that $S_{\chi\chi}$ is nonsingular if, and only if, X has full column rank. In this case, the solution is

$$\hat{\beta} = S_{\chi\chi}^{-1}S_{\chi\xi} = \left(\sum_{t=1}^n x_t x_t'\right)^{-1}\left(\sum_{t=1}^n x_t y_t\right) = (X'X)^{-1}X'y. \tag{1.60}$$

The random vector $E^*(\xi|\chi)$, given by (1.59), takes values $\hat{\beta}'x_t$, $t = 1, \ldots, n$, with probability $1/n$. Since $\hat{\beta}'x_t$ is scalar, these values satisfy $\hat{\beta}'x_t = x_t'\hat{\beta}$. Thus, (1.59) can be expressed as

$$y_t = x_t'\hat{\beta} + \hat{u}_t, \quad t = 1, \ldots, n,$$

where the variable $\Upsilon = \xi - \chi'\hat{\beta}$, that takes values $\hat{u}_t = y_t - x_t'\hat{\beta}$ with probability $1/n$ is orthogonal in the sample to χ.

The estimator $\hat{\beta}$, given by (1.60), is called the **ordinary least squares** (OLS) estimator of β in the regression model (1.58).

Remark 1.9 According to Proposition 1.1, $\hat{\beta}$ minimizes the sum of squares

$$E(\xi - \chi'\beta)'(\xi - \chi'\beta) = \sum_{t=1}^{n}(y_t - x_t'\beta)^2/n$$

with respect to β. Thus, apart from the factor $1/n$, that does not affect the results, the OLS estimator minimizes the sum of squares $\sum_{t=1}^{n}(y_t - x_t'\beta)^2$. This last interpretation is often used to define the OLS estimator without making any distributional assumption about the variables in the model. ◇

Remark 1.10 The regression model (1.58) often includes a mean term in $x_t' = (x_{1t}, \ldots, x_{kt})$ such that $x_{1t} = 1$, $t = 1, \ldots, n$. In this case, by Remark (1.2), we can work with centered variables without the results being affected by it. ◇

Remark 1.11 The orthogonality conditions implied by the sample analogue are

$$\frac{1}{n}\sum_{t=1}^{n}(y_t - x_t'\hat{\beta})x_t' = 0.$$

From this, it is concluded that if the mean is included so that $x_{1t} = 1$, $t = 1, \ldots, n$, then

$$\frac{1}{n}\sum_{t=1}^{n}(y_t - x_t'\hat{\beta}) = \frac{1}{n}\sum_{t=1}^{n}\hat{u}_t = 0.$$

That is, the mean of the regression residuals is zero. ◇

Remark 1.12 It is to be noted that the covariance in the sample is, apart from the factor $1/n$, the scalar product in $\mathbb{R}^n$, defined by $x \cdot y = \sum_{i=1}^{n} x_i y_i$. Thus, the statements about covariances in the sample can be interpreted as statements about scalar products in $\mathbb{R}^n$. With this interpretation, the results in this section provide another proof of **the Projection Theorem** in $\mathbb{R}^n$. Namely, if $S \subset \mathbb{R}^n$ is a linear subspace generated by the columns of a full column rank $n \times k$ matrix $X = (x_1, \ldots, x_n)'$ and $y \in \mathbb{R}^n$, then

(i) there exists a unique element $\hat{y} \in S$ such that

$$\|y - \hat{y}\| = \inf \{\|y - x\| : x \in S\}$$

(ii) $\hat{y} \in S$ and $\|y - \hat{y}\| = \inf \{\|y - x\| : x \in S\}$ if and only if $y - \hat{y}$ is orthogonal to all the vectors in S.

Here, $\|x\| = \sqrt{x \cdot x}$ denotes the distance of the vector x to the origin and $\hat{y}$ is the vector of the so-called fitted values, $\hat{y} = X\hat{\beta}$. ◇

1.11 Historical Notes

The selection of a squared error criterion goes back to Gauss. There is a fine translation of Stewart (1995) of Gauss's final publications on the topic that is worth reading. In particular, on pp. 9–11 Gauss argues that the size of the loss assigned to a given observation error is well represented by the square function for a number of reasons, among which the fact that the square function is continuous.

The material of this chapter is quite standard and some of the results have been known for a long time. See, for example, Doob (1953).

The idea of regarding random variables as elements of a vector space seems to go back to Fréchet, see Fréchet (1937). According to this formulation, the problem of least squares estimation reduces to a projection onto a linear subspace. This idea was exploited by Wold (1938) and later by Kolmogorov (1939, 1941) to obtain fundamental results in the theory of stationary stochastic processes.

Wold noted that it would be convenient to transform a correlated sequence of random variables into an uncorrelated one using the Gram–Schmidt procedure of Euclidean spaces. The term innovation was perhaps first used by Wiener and Masani in the mid-fifties.

1.12 Problems

1.1 Use the innovations algorithm in Example 1.2 with $\theta = 1$ to show that

$$E^*(Y_{k+1}|Y_k, \ldots, Y_1) = \frac{k}{k+1}E_k$$

and

$$D_{k+1} = \frac{k+2}{k+1}\sigma^2.$$

1.2 Use the autoregressive representation in Example 1.2 with $\theta = 1$ to show that

$$Y_{k+1} = \Pi_{k+1,1}Y_1 + \cdots + \Pi_{k+1,k}Y_k + E_{k+1},$$

where

$$\Pi_{k+1,j} = (-1)^{k-j}\frac{j}{k+1}, \quad j = 1, \ldots, k.$$

By changing the time index appropriately, show that the previous relations can also be written as

$$Y_t = \Pi_{t,t-1}Y_{t-1} + \cdots + \Pi_{t,t-k}Y_{t-k} + E_t, \quad t > k \geq 1,$$

and

$$\Pi_{t,t-j} = (-1)^{j+1}\left(1 - \frac{j}{t}\right), \quad j = 1, \ldots, k.$$

1.3 Prove that the ith column of the L matrix (1.13) can be obtained by taking the inner product of the ith column in Table 1.18 with $||E_i||^2E_i$, $i = 1, \ldots, n$.

1.4 Consider the modified Gram–Schmidt procedure of Sect. 1.5. Let $\widetilde{Y}_{|1} = (\widetilde{Y}'_{2|1}, \ldots, \widetilde{Y}'_{n|1})'$. Prove that $||\widetilde{Y}_{|1}||^2$ is the Schur complement of the top leftmost entry of $||Y||^2$. That is, partitioning $||Y||^2$ as

$$||Y||^2 = \begin{bmatrix} ||Y_1||^2 & m_1' \\ m_1 & M_1 \end{bmatrix},$$

the relation $||\widetilde{Y}_{|1}||^2 = M_1 - m_1||Y_1||^{-2}m_1'$ holds.

1.5 Let $\{Y_t\}$ follow the VARMA model

$$Y_t + \Phi_{t,1}Y_{t-1} + \Phi_{t,2}Y_{t-2} = \Theta_{t,0}A_t + \Theta_{t,1}A_{t-1} + \Theta_{t,2}A_{t-2},$$

where $\mathrm{Var}(A_t) = \Sigma_t$. Prove that this model can be put into state space form as follows:

$$x_{t+1} = \begin{bmatrix} -\Phi_{t+1,1} & I \\ -\Phi_{t+2,2} & 0 \end{bmatrix} x_t + \begin{bmatrix} \Theta_{t+1,1} - \Phi_{t+1,1}\Theta_{t,0} \\ \Theta_{t+2,2} - \Phi_{t+2,2}\Theta_{t,0} \end{bmatrix} L_t\epsilon_t$$
$$Y_t = [I, 0]x_t + \Theta_{t,0}L_t\epsilon_t,$$

where $\epsilon_t = L_t^{-1}A_t$ and $\Sigma_t = L_tL_t'$ is the Cholesky decomposition of Σ_t.

1.6 Solve the recursions of Example 1.4. To this end, show first that if $R_t = D_t/(D_t - \sigma^2)$, then

$$R_t = \theta^{-2} R_{t-1} + 1, \quad t \geq 2.$$

Deduce from this that

$$R_t = \theta^{-2(t-1)} R_1 + 1 + \theta^{-2} + \theta^{-4} + \cdots + \theta^{-2(t-2)},$$

where $R_1 = (1 - 2\theta\phi + \theta^2)/(\phi^2 - 2\theta\phi + \theta^2)$. This allows for the computation of D_t and $\Theta_{t,t-1}$ as

$$D_t = \frac{R_t}{R_t - 1}\sigma^2, \qquad \Theta_{t,t-1} = \theta \frac{R_{t-1} - 1}{R_{t-1}}.$$

1.7 Prove Lemma 1.1.

1.8 Prove the formula for the partial autocorrelation coefficient, $\alpha(k)$, when $k > 1$ in Example 1.2. To this end, apply first the autoregressive representation to the sequence $\{Y_t\}$ to get

$$Y_{k+1} = \Pi_{k+1,1} Y_1 + \Pi_{k+1,2} Y_2 + \cdots + \Pi_{k+1,k} Y_k + E_{k+1}, \quad k > 1.$$

Define $X_1 = (Y_2, \ldots, Y_k)'$, $X_2 = Y_1$ and $X_3 = Y_{k+1}$, and apply the autoregressive representation to $\{X_1, X_2, X_3\}$ to get $X_3 = \Pi^x_{31} X_1 + \Pi^x_{32} X_2 + E^x_3$. Show first that $\Pi^x_{32} = \Pi_{k+1,1}$ and $E^x_3 = E_{k+1}$. Then, show that the process $\{Y_t : t = 1, 2, \ldots\}$ is stationary, that is, the covariances $S_{i,j}$ satisfy $S_{i,j} = S(i-j, 0)$, $i, j = 0, 1, 2, \ldots$, and that this implies that the process $\{Y_t\}$ can be reversed in time and MSE$[E^*(X_3|X_1)]$ = MSE$[E^*(X_2|X_1)] = D_k$. Deduce that $\alpha(k) = \Pi_{k+1,1}$ and obtain the result.

1.9 Consider the formulae (1.7) and (1.8) given by the innovations algorithm. Prove that

$$Y' ||Y||^{-2} Y = \sum_{t=1}^{n} E_t' D_t^{-1} E_t.$$

1.10 Consider the symmetric block matrix

$$\begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix},$$

such that S_{22} is nonsingular. Prove that the following decomposition holds:

$$\begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} = \begin{bmatrix} I & S_{12} S_{22}^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} S_{11} - S_{12} S_{22}^{-1} S_{21} & 0 \\ 0 & S_{22} \end{bmatrix} \begin{bmatrix} I & 0 \\ S_{22}^{-1} S_{21} & I \end{bmatrix}.$$

Appendix

Orthogonal Projection and Orthogonal Subspaces in $\mathbb{R}^n$

A linear operator $f : \mathbb{R}^n \to \mathbb{R}^n$ is said to be a **projection** if $f^2 = f$. The following proposition characterizes projection operators.

Proposition 1A.1 *The linear operator $f : \mathbb{R}^n \to \mathbb{R}^n$ is a projection if and only if, for all $v \in \mathbb{R}^n$, the equality $v = f(v) + w$ holds, where $f(w) = 0$.*

Proof Suppose that f is a projection and let $v \in \mathbb{R}^n$. Then, if $w = v - f(v)$, applying f to both sides of the previous equality yields $f(w) = f(v) - f^2(v) = 0$. Conversely, suppose that $v = f(v) + w$ and $f(w) = 0$. Then, applying f to both sides of $v = f(v) + w$ yields $f(v) = f^2(v)$. □

Remember that the **kernel** and the **image** of a linear operator $f : \mathbb{R}^n \to \mathbb{R}^n$ are the linear subspaces of $\mathbb{R}^n$ defined by $\mathrm{Ker}(f) = \{x \in \mathbb{R}^n : f(x) = 0\}$ and $\mathrm{Img}(f) = \{x \in \mathbb{R}^n : \exists y, x = f(y)\}$.

Proposition 1A.2 *If $f : \mathbb{R}^n \to \mathbb{R}^n$ is a projection, then $\mathbb{R}^n = Img(f) \oplus Ker(f)$. In addition, $I - f$ is also a projection and $Img(I - f) = Ker(f)$, $Ker(I - f) = Img(f)$.*

Proof The decomposition $\mathbb{R}^n = \mathrm{Img}(f) + \mathrm{Ker}(f)$ is a consequence of the previous proposition. Let $v \in \mathrm{Ker}(f) \cap \mathrm{Img}(f)$. Then, $v = f(w)$ for some $w \in \mathbb{R}^n$ and $f(v) = 0 = f^2(w) = f(w) = v$.

Since $(I - f)^2 = I - 2f + f^2 = I - f$, $I - f$ is a projection. If $v \in \mathrm{Ker}(f)$, then $f(v) = 0$ and, since $v = v - f(v)$, $v \in \mathrm{Img}(I - f)$. If $w \in \mathrm{Img}(I - f)$, then $w = v - f(v)$ for some $v \in \mathbb{R}^n$. Thus, by the previous proposition, $w \in Ker(f)$. We have proved that $\mathrm{Ker}(f) = \mathrm{Img}(I - f)$. Since $I - f$ is a projection, changing the roles of f and $I - f$, we get $\mathrm{Ker}(I - f) = \mathrm{Img}(f)$. □

Example 1A.1 Let $f : \mathbb{R}^3 \to \mathbb{R}^3$ be defined by

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}.$$

Then, f is a projection onto the plane generated by $Img(f) = \{(1, 0, 0)', (0, 1, 0)'\}$ along the direction $\mathrm{Ker}(f) = \{(0, 1, 1)'\}$. The projection $I - f$ is

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}.$$

◇

Given a linear subspace $S \subset \mathbb{R}^n$, the **orthogonal complement space**, $S^{\perp}$, of S is defined as

$$S^{\perp} = \{x \in \mathbb{R}^n : x'y = 0 \quad \text{for all} \quad y \in S\}.$$

A projection f $\mathbb{R}^n \to \mathbb{R}^n$ is called an **orthogonal projection** if $\mathrm{Ker}(f) = \mathrm{Img}(f)^{\perp}$.

Proposition 1A.3 *Let P be the matrix that represents the linear operator $f : \mathbb{R}^n \to \mathbb{R}^n$ with respect to the canonical basis of $\mathbb{R}^n$. Then, f is an orthogonal projection if and only if*

(i) $P^2 = P$
(ii) $P' = P$.

Proof Suppose that f is an orthogonal projection. Then, $P^2 = P$ by definition of projection. Let $v \in \mathbb{R}^n$. Then, $w = (I - P)v$ satisfies $Pw = 0$ and w is in the kernel of f. Since f is an orthogonal projection, letting $\cdot$ denote the scalar product in $\mathbb{R}^n$, $Pv \cdot (I - P)v = 0$ holds. This implies $v'P'(I - P)v = 0$ and $P' = P'P$. Transposing the matrices in the previous equality, it is obtained that $P' = P$.

Conversely, suppose that (i) and (ii) are satisfied. Then, f is a projection. To show that f is orthogonal, we have to prove that, for all $v \in \mathbb{R}^n$, Pv is orthogonal to any vector in the kernel. By the previous proposition, $\mathrm{Ker}(f) = \mathrm{Img}(I - f)$, so that we have to prove that, for all $v, w \in \mathbb{R}^n$, Pv is orthogonal to $(I - P)w$. Since $Pv \cdot (I - P)w = v'P'(I - P)w = v'(P - P^2)w = 0$, the claim is proved. □

Proposition 1A.4 (The Matrix of an Orthogonal Projection) *Suppose that the image of the orthogonal projection $f : \mathbb{R}^n \to \mathbb{R}^n$ consists of linear combinations $x_1\beta_1 + \cdots + x_k\beta_k$ of the linearly independent vectors $\{x_1, \ldots, x_k\}$, $k \leq n$, and let $X = (x_1, \ldots, x_k)$. Then, the matrix of f with respect to the canonical basis of $\mathbb{R}^n$ is $P = X(X'X)^{-1}X'$.*

Proof Let $y \in \mathbb{R}^n$ and suppose that $f(y) = X\beta$, where $\beta = (\beta_1, \ldots, \beta_k)'$. Then, by Proposition 1A.1, $y - X\beta$ belongs to the kernel of f and, since f is an orthogonal projection, it is orthogonal to the columns of X. This implies that $X \cdot (y - X\beta) = 0$, and $X'y - X'X\beta = 0$, where $\cdot$ denotes, as in the previous proof, the scalar product in $\mathbb{R}^n$. Thus, $\beta = (X'X)^{-1}X'y$ and $f(y) = Py = X(X'X)^{-1}X'y$. □

Remark 1A.1 Under the assumptions and with the notation of the previous proposition, consider the matrix $I - P = I - X(X'X)^{-1}X'$. By Proposition 1A.2, $I - P$ is the matrix of a projection and, because it satisfies $(I - P)' = I - P$ and $(I - P)^2 = I - P$, it is an orthogonal projection. Also, since f is orthogonal, the columns of $I - P$, that are in $\mathrm{Ker}(f)$ by Proposition 1A.2, are orthogonal to the columns of X, that are in $Img(f)$. Since P is symmetric, this implies $(I - P)X = 0$. ◇

Remark 1A.2 In the case of the regression model (1.58), if the sample is (y, X), the least squares estimator of β is given by (1.60) and the vector of the so-called fitted values is $\hat{y} = X\hat{\beta} = X(X'X)^{-1}X'y = Py$. Thus, we see that $\hat{y}$ is the image of the orthogonal projection of y onto the space generated by the columns of the matrix X.

The vector of estimated residuals, $\hat{u}$, is given by $\hat{u} = (I - P)y$ and is orthogonal to the columns of the matrix X. ◇

Remark 1A.3 Because the matrix P of a projection $f : \mathbb{R}^n \to \mathbb{R}^n$ satisfies $P^2 = P$, the characteristic polynomial is $z(z-1)$. Thus, the eigenvalues of P are equal to zero or one and P is diagonalizable. That is, there exists a nonsingular matrix Q such that $Q^{-1}PQ = D$, where D is a diagonal matrix with zeros and ones in the main diagonal. If f is an orthogonal projection, Q can be taken to be an orthogonal matrix. ◇

Corollary 1A.1 *Given a linear subspace $S \subset \mathbb{R}^n$, the following decomposition holds*

$$\mathbb{R}^n = S \oplus S^{\perp}.$$

Proof Let $X = (x_1, \ldots, x_k)$, $k \leq n$, be a matrix such that its columns are a basis of S and consider the orthogonal projection $f : \mathbb{R}^n \to \mathbb{R}^n$ of $\mathbb{R}^n$ onto S. Then, by Proposition 1A.4, $f(x) = Px$, where $P = X(X'X)^{-1}X'$, and $\mathrm{Img}(f) = S$. By Proposition 1A.2, $I - P$ is the matrix of $\ker(f)$. Since f is an orthogonal projection, $\mathrm{Ker}(f) = \mathrm{Img}(f)^{\perp}$ and the decomposition

$$x = P(x) + (I - P)(x), \qquad x \in \mathbb{R}^n$$

is the required decomposition. □

Given an $m \times n$ matrix A, we define the following vector spaces. The **range space** of A, $\mathcal{R}(A)$, is the vector space spanned by the columns of A, that is,

$$\mathcal{R}(A) = \{Aa : a \in \mathbb{R}^n\}.$$

The **nullspace** of A, $\mathcal{N}(A)$, is defined as

$$\mathcal{N}(A) = \{a \in \mathbb{R}^n : Aa = 0\}.$$

Note that these spaces can be considered as the image and the kernel of the linear map $f : \mathbb{R}^m \to \mathbb{R}^n$ defined by $f(x) = Ax$ for all $x \in \mathbb{R}^m$ with respect to the canonical bases of $\mathbb{R}^m$ and $\mathbb{R}^n$.

Lemma 1A.1 *Given an $m \times n$ matrix A, the following properties hold*

$$\mathcal{N}(A) = \mathcal{R}(A')^{\perp}, \quad \mathcal{N}(A') = \mathcal{R}(A)^{\perp}, \quad \mathcal{R}(A) = \mathcal{N}(A')^{\perp}, \quad \mathcal{R}(A') = \mathcal{N}(A)^{\perp}.$$

Proof To prove the first property, let $a \in \mathcal{N}(A)$. Then, $Aa = 0$ and $a'A'b = 0$ for all $b \in \mathbb{R}^m$. This implies $a \in \mathcal{R}(A')^{\perp}$ and, therefore, $\mathcal{N}(A) \subset \mathcal{R}(A')^{\perp}$. To prove $\mathcal{R}(A')^{\perp} \subset \mathcal{N}(A)$, suppose $a \in \mathcal{R}(A')^{\perp}$. Then, $a'A'b = 0$ for all $b \in \mathbb{R}^m$ and this implies $a'A' = 0$. Therefore, $a \in \mathcal{N}(A)$. The rest of the properties are proved similarly. □

Corollary 1A.2 *Given an $m \times n$ matrix A, the following properties hold*

$$\mathbb{R}^n = \mathcal{N}(A) \oplus \mathcal{R}(A'), \qquad \mathbb{R}^m = \mathcal{N}(A') \oplus \mathcal{R}(A).$$

Proof It is a consequence of Corollary 1A.1 and the previous Lemma. □

The Multivariate Normal Distribution

A random variable X is said to have a **standard normal distribution** if it has the density

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

The characteristic function of a standard normal variable X is

$$\phi_X(t) = E(e^{itX}) = e^{-t^2/2}.$$

This can be proved as follows. Since $\left|e^{itx - x^2/2}\right| \leq e^{-x^2/2}$, we can differentiate under the integral sign to get

$$\phi_X'(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} ixe^{itx - x^2/2} dx.$$

Since the derivative of $e^{-x^2/2}$ is $-xe^{-x^2/2}$, integrating by parts yields

$$\phi_X'(t) = -t\phi_X(t).$$

We can express this equality as $[\ln \phi_X(t)]' = -t$. Thus,

$$\ln \phi_X(t) = -\frac{t^2}{2} + c,$$

where c is a constant that we can determine from the initial condition $\phi_X(0) = 1$. Therefore,

$$\phi_X(t) = e^{-t^2/2}.$$

A random variable Y is said to have a **normal distribution** if it is of the form $Y = \mu + cX$, where μ and c are real numbers. If $Y = \mu + cX$ has a normal distribution, then the mean and variance of Y exist and are given by $E(Y) = \mu$,

$\text{Var}(Y) = c^2$. In addition, the density and the characteristic function of Y can be easily shown to be

$$f_Y(x) = \frac{1}{\sqrt{2\pi}|c|} e^{-\frac{(x-\mu)^2}{2c^2}}$$

and $\phi_Y(t) = e^{it\mu - t^2c^2/2}$. Thus, the random variable Y depends only on the parameters μ and c^2 and is denoted by $Y \sim N(\mu, c^2)$.

A k-dimensional random vector X is said to have a **multivariate normal distribution** if $t'X$ has a univariate normal distribution for all $t \in \mathbb{R}^k$. The following proposition proves that this definition uniquely determines X.

Proposition 1B.1 (The Cramer–Wold Device) *The multivariate normal vector X is completely determined by the one dimensional distributions of linear functions $t'X$, for every $t \in \mathbb{R}^k$.*

Proof Let $\phi(t, s)$ be the characteristic function of $t'X$. Then, $\phi(t, s) = E(e^{ist'X})$ and, letting $s = 1$, it is obtained that $\phi(t, 1) = E(e^{it'X})$. This last function, considered as a function of t, is the characteristic function of X. Thus, the distribution of X is uniquely determined by the inversion theorem of characteristic functions. □

From the definition of multivariate normal distribution it is not evident whether the mean and covariance matrix exist. The following proposition proves that they do exist and gives also the expression for the characteristic function.

Proposition 1B.2 *Let $Y = (Y_1, \ldots, Y_k)'$ have a multivariate normal distribution. Then,*

(i) $E(Y) = \mu$ *and* $\text{Cov}(Y) = \Sigma$ *exist and are finite*
(ii) The characteristic function of Y is $e^{it'\mu - t'\Sigma t/2}$

Proof

(i) By definition, the components Y_i are univariate normal, so that $E(Y_i)$ and $\text{Var}(Y_i)$ exist and are finite. Thus, $\text{Cov}(Y_i, Y_j)$ exists and is finite because $\text{Var}(Y_i)$ and $\text{Var}(Y_j)$ are finite.
(ii) The characteristic function of $t'X$ is by definition that of a univariate normal distribution with mean $t'\mu$ and covariance matrix $t'\Sigma t$, $\phi_{t'X}(s) = e^{ist'\mu - s^2 t'\Sigma t/2}$. Letting $s = 1$, the result is obtained. □

The previous proposition shows that the distribution of a multivariate normal distribution Y depends on the mean μ and covariance matrix Σ only. We denote it by $Y \sim N(\mu, \Sigma)$.

The following proposition proves that affine transformations of multivariate normal distributions are also multivariate normal distributions. It also gives a criterion to know when two multivariate normal distributions are independent.

Proposition 1B.3 *Let X and Y have multivariate normal distributions. Then,*

(i) if c is a vector of constants and A is a nonrandom matrix such that Z is a random vector satisfying $Z = c + AX$, then Z has a multivariate normal distribution with mean $c + A\mu_x$ and covariance matrix $A\Sigma_{xx}A'$
(ii) X and Y are independent if and only if $\Sigma_{xy} = 0$.

Proof

(i) Let $t'Z$ be a linear function of Z. Then, $t'AX$ is a linear function of X and has, by definition, a univariate normal distribution. Thus, $t'c + t'AX$ also has a univariate normal distribution, with mean $t'c + t'A\mu_x$ and variance $t'A\Sigma_{xx}A't$. The characteristic function of $t'Z$ is then $\phi_{t'Z}(s) = e^{ist'(c+A\mu_x)-s^2t'A\Sigma_{xx}At/2}$. Letting $s = 1$, the result is obtained.
(ii) Let $Z = (X', Y')'$. Then, X and Y are independent if and only if the characteristic function ϕ_Z of Z satisfies $\phi_Z = \phi_X\phi_Y$, where ϕ_X and ϕ_Y denote the characteristic functions of X and Y. It is not difficult to verify that

$$\phi_Z(t) = \phi_X(t_x)\phi_Y(t_y)e^{-t_x'\Sigma_{xy}t_y - t_y'\Sigma_{yx}t_x},$$

where the partition $t = (t_x', t_y')'$ is made conformal to $Z = (X', Y')'$. From this, the result follows easily. □

Sometimes, the multivariate normal distribution is defined as a k-dimensional random vector of the form $Y = a + AX$, where a is a vector of constants, A is a nonrandom matrix, and X is a p-dimensional random vector whose elements are independent and have a standard normal distribution. The following proposition proves that this definition and ours are equivalent. In addition, it also gives the density of a multivariate normal distribution.

Proposition 1B.4 *Let Y be a k-dimensional random vector. Then,*

(i) Y has a multivariate normal distribution if and only if it has the form $Y = a + AX$, where a is a vector of constants, A is a nonrandom matrix and X is a p-dimensional random vector whose elements are independent and have a standard normal distribution.
(ii) if Y has a multivariate normal distribution and Σ_{yy} is nonsingular, then the density of Y is

$$f_Y(x) = (2\pi)^{-k/2}|\Sigma_{yy}|^{-1/2}e^{-(x-\mu_y)'\Sigma_{yy}^{-1}(x-\mu_y)/2}.$$

If Σ_{yy} is singular and $\Sigma_{yy} = LL'$, where L is a full column rank matrix with rank equal to that of Σ_{yy}, then the density of Y is

$$f_Y(x) = (2\pi)^{-k/2}|L'\Sigma_{yy}L|^{-1/2}e^{-(x-\mu_y)'L(L'\Sigma_{yy}L)^{-1}L'(x-\mu_y)/2}.$$

Proof

(i) Suppose that X is as in the statement of the proposition and that $Y = a + AX$ holds. Then, it is not difficult to verify that each linear function $t'X$ of X has characteristic function $e^{-t't/2}$ and has, therefore, normal distribution. By the previous proposition, this implies that Y has multivariate normal distribution. Conversely, suppose that Y has a multivariate normal distribution. Then, the covariance matrix Σ_{yy} can be decomposed as $\Sigma_{yy} = LL'$, where L is a full column rank matrix with rank equal to that of Σ_{yy}. We will prove that Y coincides with $Z = \mu_y + LX$, where X is as in the statement of the proposition. Clearly, the mean and covariance matrix of Z are equal to those of Y. In addition, the distribution of Z is normal by the first part of the proof, and the claim is proved.

(ii) If Σ_{yy} is nonsingular, with the notation of the first part of the proof, the equality $Y = \mu_y + LX$ holds, where L is a nonsingular matrix such that $\Sigma_{yy} = LL'$. Then, the Jacobian of the inverse transformation, $X = L^{-1}(Y - \mu_y)$, is $|L|^{-1}$. Considering that the density of X is $f_X(x) = (2\pi)^{-k/2}e^{-x'x/2}$, the result follows.

If Σ_{yy} is singular and $\Sigma_{yy} = LL'$, where L is a full column rank matrix with rank equal to that of Σ_{yy}, then, by the first part of the proof, $Y = \mu_y + LX$, where X is a random vector whose elements are independent and have a standard normal distribution. Thus, the random vector $Z = (L'L)X = L'(y - \mu_y)$ has a normal distribution with zero mean and covariance matrix $(L'L)(L'L) = L'\Sigma_{yy}L$ and the result follows. □

Conditional Distribution

Theorem 1B.1 *Suppose that*

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim N\left\{ \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \right\},$$

where Σ_{22} is nonsingular. Then, the conditional distribution $X_1|X_2 \sim N\left\{\mu_1 + \Sigma_{12}\Sigma_{22}^{-1} \times (X_2 - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right\}$, so that the conditional expectation $E(X_1|X_2)$ coincides with the orthogonal projection $E^*(X_1|X_2)$.

Proof Define $Z = X_1 - \Sigma_{12}\Sigma_{22}^{-1}X_2$. Then, Var(Z) $= \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$ and Z and X_2 are uncorrelated. Thus, by Proposition 1B.3, Z and X_2 are independent and, with an obvious notation, the densities satisfy $f_{ZX_2} = f_Z f_{X_2}$.

The transformation

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} I & \Sigma_{12}\Sigma_{22}^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} Z \\ X_2 \end{bmatrix}$$

has unit determinant and, thus, $f_{X_1X_2} = f_{ZX_2} = f_Z f_{X_2}$. This shows that f_Z is the density of $X_1|X_2$ and the theorem is proved. □

Chapter 2
Linear Models

2.1 Linear Models and Generalized Least Squares

When computing the best linear predictor of X based on Y, an extremely important case occurs when X and Y are linearly related as

$$Y = HX + V, \tag{2.1}$$

where H is a known $n \times m$ matrix, X, Y, and V are zero mean random vectors, and V is uncorrelated with X. The model (2.1) is called **linear model**. Let $\mathrm{Var}(X) = S_x$ and $\mathrm{Var}(V) = S_v$ and assume that $\mathrm{Var}(Y) = S_y = HS_xH' + S_v > 0$. Then, $\mathrm{Cov}(X, Y) = EX(HX + V)' = S_xH'$,

$$\widehat{X}_{|Y} = S_xH' \left(HS_xH' + S_v\right)^{-1} Y \tag{2.2}$$

and

$$P_x = S_x - S_xH' \left(HS_xH' + S_v\right)^{-1} HS_x, \tag{2.3}$$

where $\widehat{X}_{|Y} = E^*(X|Y)$ and $P_x = \mathrm{MSE}\left(\widehat{X}_{|Y}\right)$.

In many problems it happens that $S_x > 0$ and $S_v > 0$, which of course also ensures that $S_y > 0$. Then, we can express (2.2) and (2.3) in the so-called **information form** that involves the inverses of the covariance matrices rather than the covariance matrices themselves. The name is due to the fact that the amount of information obtained by observing a random variable varies inversely as its variance.

Applying the Matrix Inversion Lemma 4.1 yields

$$S_y^{-1} = (HS_xH' + S_v)^{-1} = S_v^{-1} + S_v^{-1}H(S_x^{-1} + H'S_v^{-1}H)^{-1}H'S_v^{-1}$$

V. Gómez, *Multivariate Time Series With Linear State Space Structure*,
DOI 10.1007/978-3-319-28599-3_2

and, therefore,

$$
\begin{aligned}
S_xH'[HS_xH' + S_v]^{-1} &= S_xH'S_v^{-1} - S_xH'S_v^{-1}H(S_x^{-1} + H'S_v^{-1}H)^{-1}H'S_v^{-1} \\
&= S_x\left[(S_x^{-1} + H'S_v^{-1}H) - H'S_v^{-1}H\right](S_x^{-1} + H'S_v^{-1}H)^{-1}H'S_v^{-1} \\
&= (S_x^{-1} + H'S_v^{-1}H)^{-1}H'S_v^{-1}.
\end{aligned}
$$

Thus,

$$\widehat{X}_{|Y} = (S_x^{-1} + H'S_v^{-1}H)^{-1}H'S_v^{-1}Y.$$

In the same way, it can be verified that

$$P_x = S_x - S_xH'\left(HS_xH' + S_v\right)^{-1}HS_x = (S_x^{-1} + H'S_v^{-1}H)^{-1}.$$

An important consequence of the previous formula is the relation

$$P_x^{-1}\widehat{X}_{|Y} = H'S_v^{-1}Y,$$

which shows that the expression $P_x^{-1}\widehat{X}_{|Y}$ is independent of the covariance matrix, S_x, of X. In summary, the information form of (2.2) and (2.3) is

$$P_x^{-1}\widehat{X}_{|Y} = H'S_v^{-1}Y \tag{2.4}$$

$$P_x^{-1} = S_x^{-1} + H'S_v^{-1}H. \tag{2.5}$$

It is to be noticed that the estimator $\widehat{X}_{|Y}$ can be obtained minimizing with respect to X the function

$$X'S_x^{-1}X + (Y - HX)'S_v^{-1}(Y - HX).$$

In certain problems, one assumes very little a priori knowledge of X, which can be modeled by setting $S_x = kI$, with k a very big number. Then, in the limit as $k \to \infty$, $S_x^{-1} \to 0$ and

$$\widehat{X}_{|Y} \to \widehat{X}_{|Y\infty} = (H'S_v^{-1}H)^{-1}H'S_v^{-1}Y, \qquad P_x \to P_{x|\infty} = (H'S_v^{-1}H)^{-1}. \tag{2.6}$$

The linear model (2.1), when X is assumed to be fixed and $\text{Var}(V) = \sigma^2 I$, is called **ordinary least squares regression model** or OLS model. To estimate X in an OLS model, the usual procedure is ordinary least squares. As we saw in the previous section, this procedure can be interpreted as an orthogonal projection in the sample. The results in this section offer another interpretation. If X is fixed in (2.1) and we adopt a Bayesian approach such that X is assumed to have a diffuse prior distribution ($S_x = kI$ with $k \to \infty$) and $S_v = \sigma^2 I$, then we obtain the OLS estimator via (2.6). An important property of the estimator $\widehat{X}_{|Y\infty}$ is given in the following theorem.

Theorem 2.1 (Gauss Markov Theorem) *If in the linear model (2.1) X is considered fixed, H has full column rank and Var$(V) = \sigma^2 I$, then the estimator $\widehat{X}_{|Y\infty} = (H'H)^{-1}H'Y$ is optimum in the sense that its covariance matrix, $P_{x|\infty} = \sigma^2(H'H)^{-1}$, has minimum trace in the class of all unbiased linear estimators of X based on Y.*

Proof Let $\widehat{Z} = AY$, with A a nonrandom matrix of appropriate dimensions, be an unbiased linear estimator of X based on Y. Then, because $\widehat{Z}$ is unbiased, $E(\widehat{Z}) = X = AHX$ and the equality

$$AH = I$$

holds. This implies $\text{Var}(\widehat{Z}) = E\left[(\widehat{Z} - X)(\widehat{Z} - X)'\right] = \sigma^2 AA'$. On the other hand, using $AH = I$, the covariance matrix $P_{x|\infty}$ can be written as

$$P_{x|\infty} = \sigma^2 AH(H'H)^{-1}H'A'.$$

Therefore,

$$\text{tr}\left[\text{Var}(\widehat{Z}) - P_{x|\infty}\right] = \sigma^2 \text{tr}\left\{A\left[I - H(H'H)^{-1}H'\right]A'\right\}.$$

Since the matrix $I - H(H'H)^{-1}H'$ is a projection matrix, it is nonnegative definite. This implies that the matrix $A\left[I - H(H'H)^{-1}H'\right]A'$ is also nonnegative definite. Thus, the trace of this last matrix is a nonnegative number that reaches its minimum value of zero for $A = (H'H)^{-1}H'$, that is, when $\widehat{Z} = \widehat{X}_{|Y\infty}$. □

If X is fixed and Var(V) is general, the model is called **generalized least squares regression model** or GLS model. If (2.1) is a GLS model and $S_v > 0$, to estimate X, one usually reduces first (2.1) to an OLS model by means of a suitable transformation and then applies ordinary least squares. More specifically, let L be any matrix such that $S_v = LL'$ and premultiply (2.1) by L^{-1} to get

$$L^{-1}Y = L^{-1}HX + L^{-1}V.$$

Then, Var$(L^{-1}V) = I$ and the transformed model is an OLS model. Applying ordinary least squares to the transformed model yields

$$\widehat{X} = (H'S_v^{-1}H)^{-1}H'S_v^{-1}Y, \qquad \text{MSE}(\widehat{X}) = (H'S_v^{-1}H)^{-1},$$

which coincide with $\widehat{X}_{|Y\infty}$ and P_∞. This procedure to estimate X is called **generalized least squares**. Note that the GLS estimator, $\widehat{X}$, is the solution to the problem of minimizing with respect to X the weighted sum of squares

$$(Y - HX)'S_v^{-1}(Y - HX) = \left[L^{-1}(Y - HX)\right]'L^{-1}(Y - HX).$$

2.2 Combined Linear Estimators

Sometimes, we have two different sets of independent observations of the linear model (2.1). The following theorem gives the details on how the estimators can be combined in this case to obtain an estimator that uses all of the available information.

Theorem 2.2 *Suppose we have two different sets of observations of X in the linear model (2.1), for example*

$$Y_a = H_a X + V_a$$
$$Y_b = H_b X + V_b,$$

where $\{V_a, V_b, X\}$ *are mutually uncorrelated, zero mean random vectors, with covariance matrices* S_a, S_b*, and* S_x*, respectively. Then, letting* $\widehat{X}_a = E^*(X|Y_a)$*,* $P_a = \mathrm{MSE}\left(\widehat{X}_a\right)$*,* $\widehat{X}_b = E^*(X|Y_b)$*, and* $P_b = \mathrm{MSE}\left(\widehat{X}_b\right)$*, and considering the stacked model*

$$Y = HX + V,$$

where $Y = [Y_a', Y_b']'$*,* $H = [H_a', H_b']'$ *and* $V = [V_a', V_b']'$*, the information form of* $\widehat{X} = E^*(X|Y)$ *and* $P = \mathrm{MSE}\left(\widehat{X}\right)$ *is given by*

$$\begin{aligned} P^{-1}\widehat{X} &= P_a^{-1}\widehat{X}_a + P_b^{-1}\widehat{X}_b \\ P^{-1} &= P_a^{-1} + P_b^{-1} - S_x^{-1}. \end{aligned}$$

Proof By (2.4), it holds that

$$P_a^{-1}\widehat{X}_a = H_a' S_a^{-1} Y_a, \qquad P_b^{-1}\widehat{X}_b = H_b' S_b^{-1} Y_b,$$

and, considering the model $Y = HX + V$, it also holds that

$$\begin{aligned} P^{-1}\widehat{X} &= H' S_v^{-1} \\ &= [H_a', H_b'] \begin{bmatrix} S_a^{-1} & 0 \\ 0 & S_b^{-1} \end{bmatrix} \begin{bmatrix} Y_a \\ Y_b \end{bmatrix} \\ &= H_a' S_a^{-1} Y_a + H_b' S_b^{-1} Y_b \\ &= P_a^{-1}\widehat{X}_a + P_b^{-1}\widehat{X}_b. \end{aligned}$$

In a similar way, by (2.5), it is obtained that

$$\begin{aligned} P &= (S_x^{-1} + H' S_v^{-1} H)^{-1} \\ &= \left(S_x^{-1} + [H'_a, H'_b] \begin{bmatrix} S_a^{-1} & 0 \\ 0 & S_b^{-1} \end{bmatrix} \begin{bmatrix} H_a \\ H_b \end{bmatrix} \right) \\ &= (S_x^{-1} + H'_a S_a^{-1} H_a + H'_b S_b^{-1} H_b)^{-1} \\ &= (S_x^{-1} + P_a^{-1} - S_x^{-1} + P_b^{-1} - S_x^{-1})^{-1}. \end{aligned}$$

□

2.3 Likelihood Function Definitions for Linear Models

To define the likelihood of a linear model, it will be convenient to consider linear models of the form

$$Y = R\delta + \omega, \tag{2.7}$$

where $\delta \sim N(b, \Pi)$, $\text{Cov}(\delta, \omega) = 0$, $\omega \sim N(m_\omega, V_\omega)$, and Π and V_ω are nonsingular. The notation $x \sim N(m, V)$ denotes a random vector x normally distributed with mean m and covariance matrix V.

The log-likelihood of the linear model (2.7) is given by the following theorem.

Theorem 2.3 *Under the previous assumptions, the log-likelihood, denoted by $\lambda(Y)$, of the linear model (2.7) is given by*

$$\begin{aligned} \lambda(Y) = \textit{constant} - \frac{1}{2} \Big\{ & \ln|\Pi| + \ln|V_\omega| + \ln|\Pi^{-1} + R' V_\omega^{-1} R| + (\hat{\delta} - b)' \Pi^{-1} (\hat{\delta} - b) \\ & + (Y - R\hat{\delta} - m_\omega)' V_\omega^{-1} (Y - R\hat{\delta} - m_\omega) \Big\}, \end{aligned} \tag{2.8}$$

where $\hat{\delta} = (\Pi^{-1} + R' V_\omega^{-1} R)^{-1} [\Pi^{-1} b + R' V_\omega^{-1} (Y - m_\omega)]$ and $MSE(\hat{\delta}) = (\Pi^{-1} + R' V_\omega^{-1} R)^{-1}$. In addition, $\hat{\delta}$ and $MSE(\hat{\delta})$ coincide with the conditional expectation $E(\delta|Y)$ and its covariance matrix, $Var(\delta|Y)$.

Proof The density $p(Y)$ satisfies in model (2.7) the relation $p(\delta|Y)p(Y) = p(Y|\delta)p(\delta)$, where the vertical bar denotes conditional distribution. The maximum likelihood estimator, $\hat{\delta}$, of δ in the right-hand side of this equation should coincide with that of the left-hand side. Equating exponents in the previous equality yields

$$\begin{aligned} &[\delta - E(\delta|Y)]' \Omega_{\delta|Y}^{-1} [\delta - E(\delta|Y)] + (Y - Rb - m_\omega)' \Omega_Y^{-1} (Y - Rb - m_\omega) \\ &\quad = (Y - R\delta - m_\omega)' V_\omega^{-1} (Y - R\delta - m_\omega) + (\delta - b)' \Pi^{-1} (\delta - b), \end{aligned}$$

where $\Omega_{\delta|Y}$ and Ω_Y are the covariance matrices of $p(\delta|Y)$ and $p(Y)$. The value of δ that minimizes the left-hand side of the previous equation is $E(\delta|Y)$. To minimize the right-hand side, consider the regression model

$$\begin{bmatrix} b \\ Y - m_\omega \end{bmatrix} = \begin{bmatrix} I \\ R \end{bmatrix} \delta + \nu, \quad \nu \sim N(0, \text{diag}(\Pi, V_\omega)), \tag{2.9}$$

and the theorem follows. □

Remark 2.1 Clearly, the expressions given in the theorem for $\hat{\delta}$ and $\text{MSE}(\hat{\delta})$ coincide with the GLS estimator of δ and its MSE in the linear model (2.7), considering δ fixed. Taking into account that the residual sum of squares in a GLS model $Z = X\beta + a$, where $\text{Var}(a) = \Sigma$, is $Z'\Sigma^{-1}Z - \hat{\beta}'(X'\Sigma^{-1}X)\hat{\beta}$, we get the following alternative expression for the residual sum of squares in model (2.9)

$$\begin{aligned} &(\hat{\delta} - b)'\Pi^{-1}(\hat{\delta} - b) + (Y - R\hat{\delta} - m_\omega)'V_\omega^{-1}(Y - R\hat{\delta} - m_\omega) \\ &\quad = b'\Pi^{-1}b + (Y - m_\omega)'V_\omega^{-1}(Y - m_\omega) - \hat{\delta}'\left(\Pi^{-1} + R'V_\omega^{-1}R\right)\hat{\delta}. \end{aligned}$$

◇

We can consider two vectors of parameters of interest in model (2.7), the vector ϕ, that contains the parameters on which the matrix R depends, and the vector ψ, that contains the parameters on which m_ω and V_ω depend. The vector δ will often be considered as a vector of nuisance random variables. In the case of the state space model (4.82) and (4.83) defined in Chap. 4, the parameters in ϕ and ψ are determined by those of the system matrices and initial conditions through the relations of Theorem 4.34.

In practice, it is very often the case that the matrix R does not depend on any parameters and so the vector ϕ is empty. This happens, for example, in ARIMA models, in which the elements of the matrix R are given by recursions defined in terms of the differencing operators used to induce stationarity (see Example 4.4). This situation will be summarized by the following assumption.

Assumption 2.1 *The matrix R in model (2.7) does not depend on any parameters and is assumed to be known.*

When δ is assumed to be a vector of nuisance random variables, the problem of defining the likelihood of model (2.7) is similar to that considered by Kalbfleisch & Sprott (1970), henceforth KS. These authors consider several possible definitions of the likelihood in the presence of nuisance parameters. The difference with model (2.7) is that in this model we have a vector of nuisance random variables, δ, instead of a vector of nuisance parameters. However, we can still apply the same methods considered by KS to define the likelihood of model (2.7).

2.3.1 The Diffuse Likelihood

The first approach considered by KS is a Bayesian one. According to this approach, a prior distribution is first selected for the nuisance random vector, δ, and then this random vector is integrated out of the likelihood to obtain the marginal likelihood of the data, $p(Y)$. In model (2.7), we should use an improper prior distribution for δ such as $p(\delta) = 1$. The marginal likelihood corresponding to this improper prior is given by the following theorem.

Theorem 2.4 *Suppose the linear model (2.7), where $\omega \sim N(m_\omega, V_\omega)$ with V_ω nonsingular and $p(\delta) = 1$. Then, the marginal log-likelihood, denoted by $\lambda_D(Y)$, is, apart from a constant,*

$$\lambda_D(Y) = -\frac{1}{2}\left\{\ln|V_\omega| + \ln|R'V_\omega^{-1}R| + (Y - R\hat{\delta} - m_\omega)'V_\omega^{-1}(Y - R\hat{\delta} - m_\omega)\right\}, \tag{2.10}$$

where $\hat{\delta} = (R'V_\omega^{-1}R)^{-1}R'V_\omega^{-1}(Y - m_\omega)$ and $MSE(\hat{\delta}) = (R'V_\omega^{-1}R)^{-1}$. In addition, the density of the posterior distribution, $p(\delta|Y)$, is that of a random vector distributed as $N(\hat{\delta}, (R'V_\omega^{-1}R)^{-1})$.

Proof Under the assumptions of the theorem, the marginal likelihood, $p(Y)$, is given by $p(Y) = \int p(Y|\delta)p(\delta)$, where the exponent of $p(Y|\delta)$ is

$$(Y - R\delta - m_\omega)'V_\omega^{-1}(Y - R\delta - m_\omega) = (Y - R\hat{\delta} - m_\omega)'V_\omega^{-1}(Y - R\hat{\delta} - m_\omega) + (\hat{\delta} - \delta)'R'V_\omega^{-1}R(\hat{\delta} - \delta),$$

and the determinant is $|V_\omega|$. Therefore,

$$\begin{aligned} p(Y) = {} & (2\pi)^{-pn/2}|V_\omega|^{-1/2}e^{-(Y - R\hat{\delta} - m_\omega)'V_\omega^{-1}(Y - R\hat{\delta} - m_\omega)/2} \\ & \times \int e^{-(\delta - \hat{\delta})'R'V_\omega^{-1}R(\delta - \hat{\delta})/2}d\delta, \end{aligned}$$

and the integral of the right-hand side of the previous expression is $(2\pi)^{n_\delta/2}$ $|R'V_\omega^{-1}R|^{-1/2}$, where n_δ is dimension of δ. The other statements of the theorem can be proved similarly. □

The marginal likelihood of the previous theorem is called the **diffuse likelihood**.

Remark 2.2 It is easy to see that the diffuse log-likelihood can also be obtained by taking the limit when $\Pi^{-1} \to 0$ of the log-likelihood of Y in the linear model (2.7), suitably normalized to avoid degeneracy. More specifically, $\lambda_D(Y)$ is the limit of $\lambda(Y) + \frac{1}{2}\ln|\Pi|$ in (2.8) when $\Pi^{-1} \to 0$. This is usually the form in which the diffuse likelihood is introduced in the literature. ◇

Remark 2.3 The estimator $\hat{\delta}$ that appears in (2.10) is the generalized least squares (GLS) estimator of δ in the linear model (2.7) that results from considering δ fixed in that model. ◇

2.3.2 The Transformation Approach and the Marginal Likelihood

The second approach considered by KS is to transform the data so that the transformed data no longer depend on the nuisance random vector, δ. This is implemented by selecting a matrix, J, with unit determinant so that, premultiplying (2.7) by J, it is obtained that

$$JY = \begin{bmatrix} R_1 \\ 0 \end{bmatrix} \delta + J\omega, \tag{2.11}$$

where R_1 is a square nonsingular matrix conformable with δ. Letting $J = [J_1', J_2']'$ so that J_1 and J_2 are conformable with (2.11), the likelihood of $J_2 Y = J_2 \omega$ does not depend on δ. This likelihood is called the **marginal likelihood**. The relation between the diffuse log-likelihood, $\lambda_D(Y)$, and the marginal log-likelihood, denoted by $\lambda_M(Y)$, is given by the following theorem.

Theorem 2.5 *Suppose the linear model (2.7), where $\omega \sim N(m_\omega, V_\omega)$ with V_ω nonsingular, and let $J = [J_1', J_2']'$ be a matrix with $|J| = 1$ such that (2.11) holds with R_1 nonsingular. Then, the marginal log-likelihood, $\lambda_M(Y)$, is*

$$\lambda_M(Y) = \lambda_D(Y) - \ln |R'R|, \tag{2.12}$$

where $\lambda_D(Y)$ is the diffuse log-likelihood.

Proof Let J be as specified in the theorem. Then, $p(Y) = p(JY)$ because $|J| = 1$ and $|R'R| = |R_1'R_1|$. Let $V_J = JV_\omega J'$ and partition $V_J = V_{ij}$, $V_J^{-1} = V^{ij}$, $i, j = 1, 2$ conforming to $J = [J_1', J_2']'$. Then, by Theorem 2.4, $\lambda_D(Y)$ is, apart from a constant

$$\lambda_D(Y) = -\frac{1}{2}\left\{\ln |V_{22}(V^{11})^{-1}| + \ln |R_1' V^{11} R_1| + (J_2 Y - J_2 m_\omega)' V_{22}^{-1} (J_2 Y - J_2 m_\omega)\right\},$$

and, because V^{11}, V_{22}, and R_1 are nonsingular, the theorem is proved. □

Remark 2.4 Under Assumption 2.1, the diffuse likelihood coincides with the marginal likelihood. ◇

2.3.3 The Conditional Likelihood

The third approach considered by KS is that of the **conditional likelihood**, denoted by $\lambda_C(Y)$. According to this approach, the vector δ is considered fixed and the likelihood is defined as the likelihood of the observations conditional to the maximum likelihood estimator, $\hat{\delta}$, of δ in the linear model (2.7). That is,

$$\lambda_C(Y) = \frac{p(Y|\delta)}{p(\hat{\delta}|\delta)},$$

where $\hat{\delta} = (R'V_\omega^{-1}R)^{-1}R'V_\omega^{-1}(Y - m_\omega)$ is the GLS estimator of δ in the linear model (2.7), where δ is considered fixed.

It is easy to prove that the conditional likelihood coincides with the diffuse likelihood. We state this in a theorem which summarizes the relation between the conditional, the diffuse, and the marginal log-likelihoods. We omit its proof.

Theorem 2.6 *The diffuse, marginal, and conditional log-likelihoods of the linear model (2.7) are related by the formula*

$$\lambda_D(Y) = \lambda_C(Y) = \lambda_M(Y) + \ln|R'R|.$$

Under Assumption 2.1, the three log-likelihoods coincide.

Remark 2.5 The conditional likelihood has received less attention in the literature than the marginal likelihood or the diffuse likelihood. We present it here merely for completeness. As the previous theorem shows, the conditional likelihood leads to the same definition of the likelihood as the diffuse likelihood. ◇

2.3.4 The Profile Likelihood

An alternative to using any of the likelihoods considered earlier in this section consists of assuming the random vector δ is fixed in the linear model (2.7). Under this assumption, model (2.7) becomes a generalized least squares model and δ can be concentrated out of the likelihood. This concentrated likelihood is called **profile likelihood**. The profile log-likelihood is given by the following theorem. The simple proof is omitted.

Theorem 2.7 *Suppose the linear model (2.7), where $\omega \sim N(m_\omega, V_\omega)$, V_ω is nonsingular, and δ is considered fixed. Then, the profile log-likelihood, denoted by $\lambda_P(Y)$, is, apart from a constant,*

$$\lambda_P(Y) = -\frac{1}{2}\left\{\ln|V_\omega| + (Y - R\hat{\delta} - m_\omega)'V_\omega^{-1}(Y - R\hat{\delta} - m_\omega)\right\}, \qquad (2.13)$$

where $\hat{\delta} = (R'V_\omega^{-1}R)^{-1}R'V_\omega^{-1}(Y - m_\omega)$ and $MSE(\hat{\delta}) = (R'V_\omega^{-1}R)^{-1}$.

Remark 2.6 The diffuse log-likelihood differs from the profile log-likelihood in the term $\ln|R'V_\omega^{-1}R|$. The influence of this term can have important adverse effects, especially in small samples. Therefore, the diffuse log-likelihood is preferable in small samples. ◇

2.4 Introduction to Signal Extraction

Suppose we are dealing with two sequences of zero mean random vectors, $\{S_t\}$ and $\{Y_t\}$, $1 \leq t \leq n$, and assume that we know the cross-covariance and covariance matrix sequences, $\gamma_{SY}(i,j) = E(S_i Y_j')$ and $\gamma_Y(i,j) = E(Y_i Y_j')$. The sequence $\{S_t\}$ will be referred to as the **signal** and is not supposed to be observable, while $\{Y_t\}$ is the observable sequence. The signal extraction problem consists of estimating S_t given all or part of the Y_t. More specifically, we can consider three types of problems.

(i) **Smoothing:** For each t, $1 \leq t \leq n$, given the vector of observations $Y = (Y_1', \ldots, Y_n')'$, obtain the estimator of S_t based on Y, that is

$$\widehat{S}_{t|n} = E^*(S_t|Y).$$

(ii) **Filtering:** For each t, $1 \leq t \leq n$, given the vector of observations $Y_{1:t} = (Y_1', \ldots, Y_t')'$, obtain the estimator of S_t based on $Y_{1:t}$, that is

$$\widehat{S}_{t|t} = E^*(S_t|Y_{1:t}).$$

(iii) **Prediction:** Given the vector of observations $Y = (Y_1', \ldots, Y_n')'$ and given and integer $j > 0$, obtain the estimator of S_{n+j} based on Y, that is

$$\widehat{S}_{n+j|n} = E^*(S_{n+j}|Y).$$

2.4.1 *Smoothing*

Let $\widehat{S} = (\widehat{S}_{1|n}', \ldots, \widehat{S}_{n|n}')'$, where $\widehat{S}_{t|n} = E^*(S_t|Y_1, \ldots, Y_n)$, be the vector of smoothed estimators of S_t, $t = 1, \ldots, n$, and let $Y = (Y_1', \ldots, Y_n')'$, $S = (S_1', \ldots, S_n')'$, $\Gamma_{SY} = \text{Cov}(S, Y)$, $\Gamma_S = \text{Var}(S)$ and $\Gamma_Y = \text{Var}(Y)$. Then, $\widehat{S}$ is the orthogonal projection of S onto Y, that is,

$$\widehat{S} = \Gamma_{SY}\Gamma_Y^{-1}Y,$$

and

$$\text{MSE}(\widehat{S}) = \Gamma_S - \Gamma_{SY}\Gamma_Y^{-1}\Gamma_{YS}.$$

Suppose that we want to estimate the signal, S, in the **signal plus noise** model

$$Y = S + N, \tag{2.14}$$

where Y, S, and N are random vectors such that S and N are orthogonal. If we assume that Y, S, and N have zero mean and covariance matrices Ω_Y, Ω_S, and Ω_N such that Ω_S and Ω_N are nonsingular, the smoothed estimator of the signal and its MSE are, according to (2.2) and (2.3),

$$\widehat{S} = \Omega_S(\Omega_S + \Omega_N)^{-1}Y$$

and

$$P_S = \mathrm{MSE}(\widehat{S}) = \Omega_S - \Omega_S(\Omega_S + \Omega_N)^{-1}\Omega_S.$$

The information form of the previous estimator and its MSE are

$$\begin{aligned} P_S^{-1}\widehat{S} &= \Omega_N^{-1}Y \\ P_S^{-1} &= \Omega_S^{-1} + \Omega_N^{-1}. \end{aligned}$$

In an analogous way, it is obtained that the estimator of N and its MSE are

$$\widehat{N} = \Omega_N(\Omega_S + \Omega_N)^{-1}Y$$

and

$$P_N = \mathrm{MSE}(\widehat{N}) = \Omega_N - \Omega_N(\Omega_S + \Omega_N)^{-1}\Omega_N.$$

Also, the information form of these expressions is

$$\begin{aligned} P_N^{-1}\widehat{N} &= \Omega_S^{-1}Y \\ P_N^{-1} &= \Omega_S^{-1} + \Omega_N^{-1} = P_S^{-1}. \end{aligned}$$

Example 2.1 Let the scalar process $\{Y_t : t \geq 1\}$ follow the signal plus noise model

$$Y_t = S_t + N_t,$$

where S_t satisfies $S_t = \phi S_{t-1} + A_t$, $A_t \sim WN(0, \sigma_A^2)$, $N_t \sim WN(0, \sigma_N^2)$ and the processes $\{A_t\}$ and $\{N_t\}$ are mutually uncorrelated. In addition, $\mathrm{E}(S_1) = 0$, $\mathrm{Var}(S_1) = \sigma_A^2/(1 - \phi^2)$ and $\{A_t : t \geq 2\}$ and $\{N_t : t \geq 1\}$ are uncorrelated with S_1.

Then, the information form of the estimator $\widehat{S}$ and its MSE, P_S, are

$$P_S^{-1}\widehat{S} = \frac{1}{\sigma_N^2}Y$$

$$P_S^{-1} = \frac{1}{\sigma_A^2}\begin{bmatrix} 1 & -\phi & & & & \\ -\phi & 1+\phi^2 & -\phi & & & \\ & \ddots & \ddots & \ddots & & \\ & & & -\phi & & \\ & & & -\phi & 1+\phi^2 & -\phi \\ & & & & -\phi & 1 \end{bmatrix} + \frac{1}{\sigma_N^2}I.$$

See Problem 2.1. ◇

2.4.2 Smoothing with Incompletely Specified Initial Conditions

Sometimes, given a scalar signal plus noise model (2.14), the signal, the noise or both have a distribution that is not completely specified. Consider, for example, the case in which the signal, $S = (S_1, S_2 \ldots, S_n)'$, follows the random walk model

$$S_t = S_{t-1} + A_t, \tag{2.15}$$

where $\{A_t\}$ is an orthogonal sequence of random variables with zero mean and common variance $\text{Var}(A_t) = \sigma_A^2$, and the noise, $N = (N_1, N_2 \ldots, N_n)'$, is a zero mean random vector orthogonal to S with covariance matrix Ω_N. To start the recursion (2.15), it is usually assumed that S_1 is a random variable orthogonal to the A_t and N_t and such that its distribution is unknown. Under this assumption, we can write model (2.14) in the form

$$Y = A\delta + BV + N, \tag{2.16}$$

where $S = A\delta + BV$, $A = (1, 1, \ldots, 1)'$, $\delta = S_1$, $V = (A_2, A_3, \ldots, A_n)'$ and

$$B = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ 1 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \cdots & 1 \end{bmatrix}.$$

2.4.2.1 Only the Signal has Incompletely Specified Initial Conditions

In this section, we will first consider a signal plus noise model (2.14) in which the noise, N, is a zero mean random vector with nonsingular covariance matrix Ω_N and the signal, S, can be transformed in the following way. It is assumed that there exists a lower triangular matrix

$$\Delta = \begin{bmatrix} \Delta_* \\ \Delta_S \end{bmatrix}$$

with ones in the diagonal such that

$$\Delta S = \begin{bmatrix} \delta \\ V \end{bmatrix},$$

where δ and V are mutually orthogonal random vectors, both orthogonal to N, such that V has zero mean and nonsingular covariance matrix Ω_V and the distribution of δ is unknown. In the signal plus noise model (2.14) where S_t follows (2.15), the matrix Δ is

$$\Delta = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{bmatrix} \tag{2.17}$$

and Δ_* and Δ_S are the first row of Δ and the matrix formed with the rest of the rows, respectively. Premultiplying S with Δ is equivalent to "differencing" the data to get rid of the unspecified part of the distribution of S. This is a standard procedure when dealing with time series whose level increases over time.

Let

$$\Delta^{-1} = [A, B] .$$

Then, we can write

$$S = A\delta + BV. \tag{2.18}$$

As in Sect. 2.1, we will model δ as a zero mean random vector with covariance matrix $\Omega_\delta = kI$ such that $k \to \infty$ or, equivalently, $\Omega_\delta^{-1} = 0$. As mentioned in that section, this corresponds to making δ diffuse because we have very little knowledge about it.

We are interested in obtaining the smoothed estimators of S and N under the previous assumptions. Since we are dealing with a diffuse δ, it will be more convenient to work with the information form of the estimators. However, we will also obtain the estimators in other forms, for completeness and to show some highlights about the signal extraction problem.

Letting $\mathrm{Var}(S) = \Omega_S = kAA' + B\Omega_V B'$, where $\Omega_V = \mathrm{Var}(V)$, the estimator, $\widehat{S}$, of S and its MSE, P_S, are

$$\widehat{S} = (\Omega_S^{-1} + \Omega_N^{-1})^{-1}\Omega_N^{-1}Y$$

and

$$P_S = (\Omega_S^{-1} + \Omega_N^{-1})^{-1}.$$

Since

$$S = A\delta + BV = \Delta^{-1}\begin{bmatrix}\delta\\ V\end{bmatrix},$$

we have

$$\begin{aligned}\Omega_S^{-1} &= \left[\Delta^{-1}\mathrm{Var}\begin{pmatrix}\delta\\ V\end{pmatrix}\Delta^{-1\prime}\right]^{-1}\\ &= \Delta'\begin{bmatrix}k^{-1}I & 0\\ 0 & \Omega_V^{-1}\end{bmatrix}\Delta\\ &\to \Delta_S'\Omega_V^{-1}\Delta_S.\end{aligned}$$

Thus,

$$\widehat{S} \to \widehat{S}_{|\infty} = (\Delta_S'\Omega_V^{-1}\Delta_S + \Omega_N^{-1})^{-1}\Omega_N^{-1}Y \tag{2.19}$$

$$P_S \to P_{S|\infty} = (\Delta_S'\Omega_V^{-1}\Delta_S + \Omega_N^{-1})^{-1}. \tag{2.20}$$

In a similar way, it is obtained for the estimator, $\widehat{N}$, of N and its MSE, P_N, that

$$\begin{aligned}\widehat{N} \to \widehat{N}_{|\infty} &= (\Delta_S'\Omega_V^{-1}\Delta_S + \Omega_N^{-1})^{-1}\Delta_S'\Omega_V^{-1}\Delta_S Y\\ P_N \to P_{N|\infty} &= (\Delta_S'\Omega_V^{-1}\Delta_S + \Omega_N^{-1})^{-1}.\end{aligned}$$

In summary, the information form estimators of S and N and their MSE are

$$\begin{aligned}P_{S|\infty}^{-1}\widehat{S}_{|\infty} &= \Omega_N^{-1}Y\\ P_{S|\infty}^{-1} &= \Delta_S'\Omega_V^{-1}\Delta_S + \Omega_N^{-1}\\ P_{N|\infty}^{-1}\widehat{N}_{|\infty} &= \Delta_S'\Omega_V^{-1}\Delta_S Y\\ P_{N|\infty}^{-1} &= \Delta_S'\Omega_V^{-1}\Delta_S + \Omega_N^{-1} = P_{S|\infty}^{-1}.\end{aligned}$$

Similar expressions were obtained in Gómez (1999) for $\widehat{S}_{|\infty}$, together with two equivalent formulae. We will deal with the equivalence of the different approaches considered in that article later in Chap. 7.

It is seen that the information form expressions for $\widehat{S}_{|\infty}$, $P_{S|\infty}$, $\widehat{N}_{|\infty}$, and $P_{N|\infty}$ are rather simple. Only Δ_S, Ω_V, and Ω_N are needed. Thus, to compute the information form estimator of S and its MSE, there is no need to care about the initial conditions. The information form estimator of N is also the information form estimator of N in the transformed linear model $\Delta_S Y = V + \Delta_S N$, that does not depend on δ. In addition, $\widehat{N}_{|\infty} = Y - \widehat{S}_{|\infty}$.

It is to be noticed that the series $V = \Delta_S S$ is the result of differencing the series S, and it is to be further noticed that the estimator $\widehat{S}_{|\infty}$ can be obtained by minimizing the function

$$S'\Delta_S'\Omega_V^{-1}\Delta_S S + (Y - S)'\Omega_N^{-1}(Y - S)$$

with respect to S. In a similar way, the estimator $\widehat{N}_{|\infty}$ can be obtained by minimizing the expression

$$(Y - N)'\Delta_S'\Omega_V^{-1}\Delta_S(Y - N) + N'\Omega_N^{-1}N$$

with respect to N.

Alternative expressions for $\widehat{S}_{|\infty}$, $P_{S|\infty}$, $\widehat{N}_{|\infty}$, and $P_{N|\infty}$ that are not in information form can be derived by applying the Matrix Inversion Lemma 4.1 to the matrix $(\Delta_S'\Omega_V^{-1}\Delta_S + \Omega_N^{-1})^{-1}$. Thus,

$$P_{S|\infty} = P_{N|\infty} = (\Delta_S'\Omega_V^{-1}\Delta_S + \Omega_N^{-1})^{-1} = \Omega_N - \Omega_N\Delta_S'\left(\Omega_V + \Delta_S\Omega_N\Delta_S'\right)^{-1}\Delta_S\Omega_N \tag{2.21}$$

and, using this expression, we further get

$$\begin{aligned}(\Delta_S'\Omega_V^{-1}\Delta_S + \Omega_N^{-1})^{-1}\Delta_S'\Omega_V^{-1}\Delta_S &= (\Delta_S'\Omega_V^{-1}\Delta_S + \Omega_N^{-1})^{-1}\left(\Delta_S'\Omega_V^{-1}\Delta_S + \Omega_N^{-1} - \Omega_N^{-1}\right)\\ &= I - (\Delta_S'\Omega_V^{-1}\Delta_S + \Omega_N^{-1})^{-1}\Omega_N^{-1}\\ &= \Omega_N\Delta_S'\left(\Omega_V + \Delta_S\Omega_N\Delta_S'\right)^{-1}\Delta_S.\end{aligned}$$

This implies

$$\widehat{N}_{|\infty} = \Omega_N\Delta_S'\left(\Omega_V + \Delta_S\Omega_N\Delta_S'\right)^{-1}\Delta_S Y \tag{2.22}$$

$$\begin{aligned}\widehat{S}_{|\infty} &= Y - \widehat{N}_{|\infty}\\ &= \left[I - \Omega_N\Delta_S'\left(\Omega_V + \Delta_S\Omega_N\Delta_S'\right)^{-1}\Delta_S\right]Y.\end{aligned} \tag{2.23}$$

Example 2.2 Let the scalar process $\{Y_t : t \geq 1\}$ follow the signal plus noise model

$$Y_t = S_t + N_t,$$

where S_t satisfies $S_t - S_{t-1} = A_t$, $A_t \sim WN(0, \sigma_A^2)$, $N_t \sim WN(0, \sigma_N^2)$, the processes $\{A_t\}$ and $\{N_t\}$ are mutually uncorrelated and S_1 is a diffuse random variable orthogonal to A_t and N_t for all t. Then, the information form of the estimator $\widehat{S}_{|\infty}$ and its MSE, $P_{S|\infty}$, are

$$\begin{aligned}
P_{S|\infty}^{-1}\widehat{S}_{|\infty} &= \frac{1}{\sigma_N^2}Y \\
P_{S|\infty}^{-1} &= \frac{1}{\sigma_A^2}\Delta_S'\Delta_S + \frac{1}{\sigma_N^2}I \\
&= \frac{1}{\sigma_A^2}\begin{bmatrix} 1 & -1 & & & & \\ -1 & 2 & -1 & & & \\ & \ddots & \ddots & \ddots & & \\ & & & & -1 & \\ & & & -1 & 2 & -1 \\ & & & & -1 & 1 \end{bmatrix} + \frac{1}{\sigma_N^2}I,
\end{aligned}$$

where Δ_S is the submatrix of Δ, given by (2.17), formed with all the rows of Δ except the first. ◇

2.4.2.2 Transformation Approach to Obtain the Estimators

It follows from (2.21) and (2.22) that $\widehat{N}_{|\infty} = E^*(N|\Delta_S Y)$ and $P_{N|\infty} = MSE(\widehat{N}_{|\infty})$. Thus, as far as the estimation of N is concerned, making δ diffuse in the signal plus noise model (2.14) with S given by (2.32) is equivalent to estimate N in the transformed linear model $\Delta_S Y = V + \Delta_S N$, that does not depend on δ. However, the estimator $\widehat{S}_{|\infty}$ cannot be obtained from the transformed model $\Delta_S Y = V + \Delta_S N = \Delta_S S + \Delta_S N$ because S depends on the diffuse part, δ.

Since $S = A\delta + BV = \Delta^{-1}(\delta', V')'$, we can obtain the estimators, $\hat{\delta}_{|\infty}$ and $\widehat{V}_{|\infty}$, of δ and V as

$$\begin{bmatrix} \hat{\delta}_{|\infty} \\ \widehat{V}_{|\infty} \end{bmatrix} = \Delta \widehat{S}_{|\infty} \tag{2.24}$$

and

$$\text{MSE}\left(\begin{bmatrix} \hat{\delta}_{|\infty} \\ \widehat{V}_{|\infty} \end{bmatrix}\right) = \Delta \ \text{MSE}\left(\widehat{S}_{|\infty}\right) \ \Delta'. \tag{2.25}$$

Substituting (2.19) and (2.20) into (2.24) and (2.25) yields

$$\begin{bmatrix} \hat{\delta}_{|\infty} \\ \widehat{V}_{|\infty} \end{bmatrix} = \Delta(\Delta_S'\Omega_V^{-1}\Delta_S + \Omega_N^{-1})^{-1}\Omega_N^{-1}Y$$

and

$$\text{MSE}\left(\begin{bmatrix} \hat{\delta}_{|\infty} \\ \widehat{V}_{|\infty} \end{bmatrix}\right) = \Delta\ (\Delta_S'\Omega_V^{-1}\Delta_S + \Omega_N^{-1})^{-1}\ \Delta'.$$

Therefore, the information form of $\hat{\delta}_{|\infty}$ and $\widehat{V}_{|\infty}$ and their MSE are

$$\text{MSE}\left(\begin{bmatrix} \hat{\delta}_{|\infty} \\ \widehat{V}_{|\infty} \end{bmatrix}\right)^{-1} \begin{bmatrix} \hat{\delta}_{|\infty} \\ \widehat{V}_{|\infty} \end{bmatrix} = \begin{bmatrix} A' \\ B' \end{bmatrix} \Omega_N^{-1}Y$$

and

$$\text{MSE}\left(\begin{bmatrix} \hat{\delta}_{|\infty} \\ \widehat{V}_{|\infty} \end{bmatrix}\right)^{-1} = \begin{bmatrix} A' \\ B' \end{bmatrix} (\Delta_S'\Omega_V^{-1}\Delta_S + \Omega_N^{-1})[A, B].$$

To obtain expressions for $\hat{\delta}_{|\infty}$ and $\widehat{V}_{|\infty}$ and their MSE that are not in information form, we can substitute (2.23) and (2.21) into (2.24) and (2.25) to get

$$\begin{aligned} \begin{bmatrix} \hat{\delta}_{|\infty} \\ \widehat{V}_{|\infty} \end{bmatrix} &= \begin{bmatrix} \Delta_* Y - \Delta_* \widehat{N}_{|\infty} \\ \Delta_S Y - \Delta_S \widehat{N}_{|\infty} \end{bmatrix} \\ &= \Delta \left[I - \Omega_N \Delta_S' \left(\Omega_V + \Delta_S \Omega_N \Delta_S' \right)^{-1} \Delta_S \right] Y \end{aligned}$$

and

$$\text{MSE}\left(\begin{bmatrix} \hat{\delta}_{|\infty} \\ \widehat{V}_{|\infty} \end{bmatrix}\right) = \Delta\ \left[\Omega_N - \Omega_N \Delta_S' \left(\Omega_V + \Delta_S \Omega_N \Delta_S' \right)^{-1} \Delta_S \Omega_N \right]\ \Delta'.$$

Thus,

$$\begin{aligned} \hat{\delta}_{|\infty} &= \Delta_* \left[I - \Omega_N \Delta_S' \left(\Omega_V + \Delta_S \Omega_N \Delta_S' \right)^{-1} \Delta_S \right] Y \qquad (2.26) \\ \text{MSE}(\hat{\delta}_{|\infty}) &= \Delta_* \left[\Omega_N - \Omega_N \Delta_S' \left(\Omega_V + \Delta_S \Omega_N \Delta_S' \right)^{-1} \Delta_S \Omega_N \right]\ \Delta_*' \end{aligned}$$

$$\begin{aligned}\widehat{V}_{|\infty} &= \Delta_S Y - \Delta_S \Omega_N \Delta_S' \left(\Omega_V + \Delta_S \Omega_N \Delta_S'\right)^{-1} \Delta_S Y \\ &= \left[\Omega_V + \Delta_S \Omega_N \Delta_S' - \Delta_S \Omega_N \Delta_S'\right] \left(\Omega_V + \Delta_S \Omega_N \Delta_S'\right)^{-1} \Delta_S Y \\ &= \Omega_V \left(\Omega_V + \Delta_S \Omega_N \Delta_S'\right)^{-1} \Delta_S Y \end{aligned} \tag{2.27}$$

and

$$\begin{aligned}\text{MSE}\left(\widehat{V}_{|\infty}\right) &= \Delta_S \Omega_N \Delta_S' - \Delta_S \Omega_N \Delta_S' \left(\Omega_V + \Delta_S \Omega_N \Delta_S'\right)^{-1} \Delta_S \Omega_N \Delta_S' \\ &= \Delta_S \Omega_N \Delta_S' \left[I - \left(\Omega_V + \Delta_S \Omega_N \Delta_S'\right)^{-1} \left(\Omega_V - \Omega_V + \Delta_S \Omega_N \Delta_S'\right)\right] \\ &= \Delta_S \Omega_N \Delta_S' \left(\Omega_V + \Delta_S \Omega_N \Delta_S'\right)^{-1} \Omega_V \\ &= \left(\Omega_V - \Omega_V + \Delta_S \Omega_N \Delta_S'\right) \left(\Omega_V + \Delta_S \Omega_N \Delta_S'\right)^{-1} \Omega_V \\ &= \Omega_V - \Omega_V \left(\Omega_V + \Delta_S \Omega_N \Delta_S'\right)^{-1} \Omega_V . \end{aligned}$$

It is to be noticed that $\widehat{V}_{|\infty} = E^*(V|\Delta_S Y)$ and $MSE\left(\widehat{V}_{|\infty}\right) = \text{MSE}\left[E^*(V|\Delta_S Y)\right]$. Therefore, $\widehat{V}_{|\infty}$ and $\text{MSE}\left(\widehat{V}_{|\infty}\right)$ are the estimator of V and its MSE in the transformed linear model $\Delta_S Y = V + \Delta_S N$, that, as mentioned earlier, does not depend on δ.

2.4.2.3 Conditional Likelihood Approach to Obtain the Estimators

Another way to estimate δ and V and, therefore, S, is the following. Since, as we saw earlier in this section, $\widehat{S}_{|\infty}$ can be obtained by minimizing with respect to S the function

$$S' \Delta_S' \Omega_V^{-1} \Delta_S S + (Y - S)' \Omega_N^{-1} (Y - S),$$

and $S = A\delta + BV = \Delta^{-1}(\delta', V')'$, we can estimate δ and V by minimizing with respect to these vectors the function

$$V' \Omega_V^{-1} V + (Y - A\delta - BV)' \Omega_N^{-1} (Y - A\delta - BV). \tag{2.28}$$

This minimization can be done in two steps. In the first step, we consider δ fixed and we minimize (2.28) with respect to V. The result can be easily shown to be

$$\widehat{V}_{|\delta} = \left(\Omega_V^{-1} + B' \Omega_N^{-1} B\right)^{-1} B' \Omega_N^{-1} (Y - A\delta).$$

Using the Matrix Inversion Lemma 4.1, we get

$$\left(\Omega_V^{-1} + B'\Omega_N^{-1}B\right)^{-1} = \Omega_V - \Omega_V B'\left(\Omega_N + B\Omega_V B'\right)^{-1} B\Omega_V$$

and

$$\begin{aligned}\left(\Omega_V^{-1} + B'\Omega_N^{-1}B\right)^{-1} B' &= \Omega_V B' - \Omega_V B'\left(\Omega_N + B\Omega_V B'\right)^{-1} B\Omega_V B' \\ &= \Omega_V B' - \Omega_V B'\left(\Omega_N + B\Omega_V B'\right)^{-1}\left(B\Omega_V B' + \Omega_N - \Omega_N\right) \\ &= \Omega_V B'\left(B\Omega_V B' + \Omega_N\right)^{-1}\Omega_N.\end{aligned}$$

Thus, $\widehat{V}_{|\delta}$ can be expressed in noninformation form as

$$\widehat{V}_{|\delta} = \Omega_V B'\left(B\Omega_V B' + \Omega_N\right)^{-1}(Y - A\delta). \tag{2.29}$$

It is seen that $\widehat{V}_{|\delta}$ is the orthogonal projection of V onto Y in the model $Y - A\delta = BV + N$ when δ is considered fixed. Let us denote this orthogonal projection by $\widehat{V}_{|\delta} = E^*(V|Y,\delta)$.

In the second step, to obtain the estimator, $\hat{\delta}_{|\infty}$, of δ, we can first substitute $\widehat{V}_{|\delta}$ into (2.28) and then minimize the resulting expression with respect to δ. However, according to the result (2.6), the estimator $\hat{\delta}_{|\infty}$ is also the GLS estimator of δ when δ is considered fixed in the linear model $Y = A\delta + BV + N$. Thus,

$$\hat{\delta}_{|\infty} = \left[A'\left(B\Omega_V B' + \Omega_N\right)^{-1} A\right]^{-1} A'\left(B\Omega_V B' + \Omega_N\right)^{-1} Y, \tag{2.30}$$

and the estimator $\widehat{V}_{|\infty}$ is obtained plugging this expression into (2.29). That is,

$$\widehat{V}_{|\infty} = \Omega_V B'\left(B\Omega_V B' + \Omega_N\right)^{-1}(Y - A\hat{\delta}_{|\infty}). \tag{2.31}$$

We will denote the estimator obtained in this way by $\widehat{V}_{|\infty} = E^*(V|Y,\hat{\delta}_{|\infty})$, meaning that we have replaced δ with $\hat{\delta}_{|\infty}$ in the orthogonal projection $E^*(V|Y,\delta)$.

Using this approach, the estimator of S is

$$\widehat{S}_{|\infty} = A\hat{\delta}_{|\infty} + B\left[E^*(V|Y,\hat{\delta}_{|\infty})\right],$$

where $\hat{\delta}_{|\infty}$ and $E^*(V|Y,\hat{\delta}_{|\infty}) = \widehat{V}_{|\infty}$ are given by (2.30) and (2.31).

The MSE of $\hat{\delta}_{|\infty}$ is, according to (2.6),

$$\text{MSE}\left(\hat{\delta}_{|\infty}\right) = \left[A'\left(B\Omega_V B' + \Omega_N\right)^{-1} A\right]^{-1}.$$

Letting $\widehat{S}_{|\delta} = A\delta + B\,[E^*(V|Y,\delta)]$, the MSE of $\widehat{S}_{|\infty}$ can be obtained as follows. Since $S - \widehat{S}_{|\delta}$ is orthogonal to $\widehat{S}_{|\delta} - \widehat{S}_{|\infty}$, we have

$$\begin{aligned}
\text{MSE}(\widehat{S}_{|\infty}) &= \text{Var}(S - \widehat{S}_{|\infty}) \\
&= \text{Var}(S - \widehat{S}_{|\delta} + \widehat{S}_{|\delta} - \widehat{S}_{|\infty}) \\
&= \text{Var}(S - \widehat{S}_{|\delta}) + \text{Var}(\widehat{S}_{|\delta} - \widehat{S}_{|\infty}) \\
&= B\left\{\text{MSE}\left[E^*(V|Y,\delta)\right]\right\}B' + (A + BD)\left[\text{MSE}\left(\hat{\delta}_{|\infty}\right)\right](A + BD)',
\end{aligned}$$

where

$$\text{MSE}\left[E^*(V|Y,\delta)\right] = \Omega_V - \Omega_V B'\left(B\Omega_V B' + \Omega_N\right)^{-1} B\Omega_V$$

and

$$D = \Omega_V B'\left(B\Omega_V B' + \Omega_N\right)^{-1} A.$$

Using (2.30) and (2.31) and the fact that $\widehat{N}_{|\infty} = Y - \widehat{S}_{|\infty}$, we can estimate N as

$$\begin{aligned}
\widehat{N}_{|\infty} &= Y - A\hat{\delta}_{|\infty} - B\widehat{V}_{|\infty} \\
&= \left[I - B\Omega_V B'\left(B\Omega_V B' + \Omega_N\right)^{-1}\right]\left(Y - A\hat{\delta}_{|\infty}\right) \\
&= \left(B\Omega_V B' + \Omega_N - B\Omega_V B'\right)\left(B\Omega_V B' + \Omega_N\right)^{-1}\left(Y - A\hat{\delta}_{|\infty}\right) \\
&= \Omega_N\left(B\Omega_V B' + \Omega_N\right)^{-1}\left(Y - A\hat{\delta}_{|\infty}\right).
\end{aligned}$$

As in the case of $\widehat{V}_{|\infty}$, the previous expression shows that $\widehat{N}_{|\infty}$ can be obtained by means of a two-step process. In the first step, the orthogonal projection, $\widehat{N}_{|\delta}$, of N onto Y is obtained in the model $Y - A\delta = BV + N$, where δ is considered fixed. In the second step, δ is replaced with $\hat{\delta}_{|\infty}$, the GLS estimator of δ in the linear model $Y = A\delta + BV + N$ when δ is considered fixed.

Finally, the MSE of $\widehat{N}_{|\infty}$ can be obtained in a way similar to that used to obtain $\text{MSE}(\widehat{S}_{|\infty})$. More specifically, we have

$$\begin{aligned}
\text{MSE}(\widehat{N}_{|\infty}) &= \text{Var}\left(N - \widehat{N}_{|\infty}\right) \\
&= \text{Var}\left(N - \widehat{N}_{|\delta} + \widehat{N}_{|\delta} - \widehat{N}_{|\infty}\right) \\
&= \text{Var}\left(N - \widehat{N}_{|\delta}\right) + \text{Var}\left(\widehat{N}_{|\delta} - \widehat{N}_{|\infty}\right) \\
&= \text{MSE}\left(\widehat{N}_{|\delta}\right) + E\left[\text{MSE}\left(\hat{\delta}_{|\infty}\right)\right]E',
\end{aligned}$$

where

$$\text{MSE}\left(\widehat{N}_{|\delta}\right) = \Omega_N - \Omega_N \left(B\Omega_V B' + \Omega_N\right)^{-1} \Omega_N$$

and

$$E = \Omega_N \left(B\Omega_V B' + \Omega_N\right)^{-1} A.$$

2.4.2.4 Both Signal and Noise Have Incompletely Specified Initial Conditions

In this section, we consider the scalar signal plus noise model

$$Y_t = S_t + N_t,$$

where both the signal and the noise have incompletely specified initial conditions. More specifically, we assume that there exist two lower triangular matrices with ones in the diagonal,

$$\overline{\Delta}_S = \begin{bmatrix} \Delta_{*S} \\ \Delta_S \end{bmatrix}, \qquad \overline{\Delta}_N = \begin{bmatrix} \Delta_{*N} \\ \Delta_N \end{bmatrix},$$

such that

$$\overline{\Delta}_S S = \begin{bmatrix} \delta_S \\ V_S \end{bmatrix}, \qquad \overline{\Delta}_N N = \begin{bmatrix} \delta_N \\ V_N \end{bmatrix},$$

and

$$\overline{\Delta}_S^{-1} = [A_S, B_S], \qquad \overline{\Delta}_N^{-1} = [A_N, B_N].$$

Here, $[\delta_S', \delta_N']$ and $[V_S', V_N']$ are assumed to be orthogonal random vectors and the distribution of $[\delta_S', \delta_N']$ is unknown. In addition, V_S and V_N are orthogonal, zero mean, random vectors with nonsingular covariance matrices, Ω_{VS} and Ω_{VN}. Thus, it holds that

$$S = A_S\delta_S + B_S V_S, \qquad N = A_N\delta_N + B_N V_N.$$

Premultiplying S with $\overline{\Delta}_S$ is equivalent to "differencing" the signal, S_t, to get rid of the unspecified part, δ_S, of the distribution of S_t, and analogously for the other component, N_t. For example, suppose $\{Y_t\}$ is a quarterly series such that

$Y_t = S_t + N_t$, where S_t and N_t follow the models

$$(1 + B + B^2 + B^3)S_t = (1 + \alpha_1 B + \alpha_2 B^2 + \alpha_3 B^3)B_t$$
$$(1 - B)N_t = (1 + \beta B)C_t,$$

B is the backshift operator, $BY_t = Y_{t-1}$, $\{B_t\}$ and $\{C_t\}$ are mutually orthogonal white noise sequences, and all the roots of the polynomials $1 + \alpha_1 z + \alpha_2 z^2 + \alpha_3 z^3$ and $1 + \beta z$ are on or outside the unit circle. In this case, S_t and N_t are the seasonal and the nonseasonal components. The matrices $\overline{\Delta}_S$ and $\overline{\Delta}_N$ can be

$$\overline{\Delta}_S = \begin{bmatrix} 1\,0\,0\,0 \cdots 0\,0\,0\,0 \\ 0\,1\,0\,0 \cdots 0\,0\,0\,0 \\ 0\,0\,1\,0 \cdots 0\,0\,0\,0 \\ 1\,1\,1\,1 \cdots 0\,0\,0\,0 \\ \vdots\,\vdots\,\vdots\,\vdots\,\ddots\,\vdots\,\vdots\,\vdots\,\vdots \\ 0\,0\,0\,0 \cdots 1\,1\,1\,1 \end{bmatrix}, \qquad \overline{\Delta}_N = \begin{bmatrix} 1 & 0 & 0 \cdots & 0 & 0 \\ -1 & 1 & 0 \cdots & 0 & 0 \\ 0 & -1 & 1 \cdots & 0 & 0 \\ \vdots & \vdots & \vdots \ddots & \vdots & \vdots \\ 0 & 0 & 0 \cdots & -1 & 1 \end{bmatrix}.$$

With this choice, $\delta_S = (S_1, S_2, S_3)'$ and $\delta_N = N_1$.

In the rest of this section, it is assumed that the differencing polynomials that are involved in the definition of the matrices $\overline{\Delta}_S$ and $\overline{\Delta}_N$ have no common factors. Then, proceeding similarly as we did previously in this section, we will model δ_S and δ_N as zero mean random vectors with covariance matrices $k_S I_a$ and $k_N I_b$ such that $k_S \to \infty$ and $k_N \to \infty$, where a and b are the dimensions of δ_S and δ_N. As mentioned earlier in this section, this corresponds to making both δ_S and δ_N diffuse because we have very little knowledge about them. Thus,

$$\Omega_S = k_S A_S A_S' + B_S \Omega_{VS} B_S', \qquad \Omega_N = k_N A_N A_N' + B_N \Omega_{VN} B_N',$$

where $\mathrm{Var}(S) = \Omega_S$ and $\mathrm{Var}(N) = \Omega_N$, and the information form estimates of S and N are

$$P_S^{-1}\widehat{S} = \Omega_N^{-1} Y, \qquad P_N^{-1}\widehat{N} = \Omega_S^{-1} Y,$$

with

$$P_S^{-1} = \Omega_S^{-1} + \Omega_N^{-1} = P_N^{-1}.$$

Letting $k_S \to \infty$ and $k_N \to \infty$, it is obtained that

$$\widehat{S} \to \widehat{S}_{|\infty} = (\Delta_S' \Omega_{VS}^{-1} \Delta_S + \Delta_N' \Omega_{VN}^{-1} \Delta_N)^{-1} \Delta_N' \Omega_{VN}^{-1} \Delta_N Y$$
$$P_S \to P_{S|\infty} = (\Delta_S' \Omega_{VS}^{-1} \Delta_S + \Delta_N' \Omega_{VN}^{-1} \Delta_N)^{-1}$$

and

$$\widehat{N} \to \widehat{N}_{|\infty} = (\Delta_S'\Omega_{VS}^{-1}\Delta_S + \Delta_N'\Omega_{VN}^{-1}\Delta_N)^{-1}\Delta_S'\Omega_{VS}^{-1}\Delta_S Y$$
$$P_N \to P_{N|\infty} = (\Delta_S'\Omega_{VS}^{-1}\Delta_S + \Delta_N'\Omega_{VN}^{-1}\Delta_N)^{-1}.$$

To obtain the estimators $\widehat{S}_{|\infty}$ and $\widehat{N}_{|\infty}$, as well as $P_{S|\infty}$ and $P_{N|\infty}$, in noninformation form, we can proceed as in the case in which only the signal has incompletely specified initial conditions. However, the computations are more complicated. See Problems 2.3, 2.4, and 2.5.

2.4.3 Filtering

Let, as before, $\{S_t\}$ and $\{Y_t\}$, $1 \leq t \leq n$, be two sequences of zero mean random vectors and suppose that we know the cross-covariance and covariance matrix sequences, $\gamma_{SY}(i,j) = E(S_i Y_j')$ and $\gamma_Y(i,j) = E(Y_i Y_j')$. If $\widehat{S}_f = (\widehat{S}_{1|1}', \ldots, \widehat{S}_{n|n}')'$, where $\widehat{S}_{t|t} = E^*(S_t|Y_1, \ldots, Y_t)$, is the vector of filtered estimators of S_t, $t = 1, \ldots, n$, then

$$\widehat{S}_f = K_f Y,$$

where K_f is a lower triangular matrix.

To obtain the matrix K_f, we can use the innovations and the Law of iterated orthogonal projection, Theorem 1.10. To this end, let $Y = (Y_1', \ldots, Y_N')'$, $S = (S_1', \ldots, S_N')'$, $\Gamma_{SY} = \mathrm{Cov}(S, Y)$ and $\Gamma_Y = \mathrm{Var}(Y)$. Then, the smoothed estimator, $\widehat{S}$, of S based on Y is the orthogonal projection of S onto Y, that is,

$$\widehat{S} = \Gamma_{SY}\Gamma_Y^{-1}Y.$$

Let $\Gamma_Y = LDL'$ the block Cholesky decomposition of Γ_Y, such that $E = L^{-1}Y$ is the vector of innovations and D is a block diagonal matrix containing the covariance matrices of E, $D = \mathrm{Var}(E)$. Using this decomposition, we can express $\widehat{S}$ as

$$\widehat{S} = \Gamma_{SY}L^{'-1}D^{-1}E.$$

Projecting each element, $\widehat{S}_{t|n}$, in $\widehat{S} = (\widehat{S}_{1|n}', \ldots, \widehat{S}_{n|n}')'$ onto $(E_1', \ldots, E_t')'$, $t = 1, \ldots, n$, we obtain $\widehat{S}_{t|t}$ by the Law of iterated orthogonal projection 1.10 and Proposition 1.7. Thus, we conclude that

$$\widehat{S}_f = \left[\Gamma_{SY}L^{'-1}D^{-1}\right]_+ E,$$

where if A is a square matrix, the notation $B = [A]_+$ means that B has the same lower triangular part as A and all its other elements (the strictly upper triangular part) are zero. Finally, we get

$$\widehat{S}_f = \left[\Gamma_{SY} L'^{-1} D^{-1}\right]_+ L^{-1} Y,$$

and

$$K_f = \left[\Gamma_{SY} L'^{-1} D^{-1}\right]_+ L^{-1}.$$

To see the relation between the smoothed estimator, $\widehat{S}$, and the filtered estimator, $\widehat{S}_f$, consider the decomposition

$$\begin{aligned}\Gamma_{SY}\Gamma_Y^{-1} &= \Gamma_{SY} L'^{-1} D^{-1} L^{-1} \\ &= \left\{\left[\Gamma_{SY} L'^{-1} D^{-1}\right]_+ + \left[\Gamma_{SY} L'^{-1} D^{-1}\right]_-\right\} L^{-1} \\ &= K_f + \left[\Gamma_{SY} L'^{-1} D^{-1}\right]_- L^{-1},\end{aligned}$$

where if A is a square matrix, the notation $B = [A]_-$ means that B has the same strictly upper triangular part as A and all its other elements (the lower triangular part) are zero. Thus, we get the relation

$$\widehat{S} = \widehat{S}_f + \left[\Gamma_{SY} L'^{-1} D^{-1}\right]_- L^{-1} Y.$$

The mean squared error of $\widehat{S}_f$ is

$$\begin{aligned}\mathrm{MSE}(\widehat{S}_f) &= \mathrm{Var}(S - \widehat{S}_f) = \mathrm{Var}(S - \widehat{S} + \widehat{S} - \widehat{S}_f) \\ &= \mathrm{MSE}(\widehat{S}) + \left[\Gamma_{SY} L'^{-1} D^{-1}\right]_- D \left[\Gamma_{SY} L'^{-1} D^{-1}\right]_-',\end{aligned}$$

because $S - \widehat{S}$ and $\widehat{S} - \widehat{S}_f$ are uncorrelated.

Example 2.3 Suppose the signal plus noise model (2.14) and assume that S and N are orthogonal and $\mathrm{Var}(N)$ is block diagonal. Then, if Y, S, and N have zero mean and covariance matrices Γ_Y, Ω_S, and Ω_N such that Ω_S and Ω_N are nonsingular, the relations $\Gamma_{SY} = \Omega_S$ and $\Gamma_Y = \Omega_S + \Omega_N = LDL'$ hold. Thus, the filtered estimator is $\widehat{S}_f = K_f Y$, where

$$\begin{aligned}K_f &= \left[(\Gamma_Y - \Omega_N) L'^{-1} D^{-1}\right]_+ L^{-1} \\ &= \left[L - \Omega_N L'^{-1} D^{-1}\right]_+ L^{-1} \\ &= I - \left[\Omega_N L'^{-1} D^{-1}\right]_+ L^{-1}.\end{aligned}$$

Since $L^{'-1}$ is upper triangular with blocks of ones in the diagonal and Ω_N and D are block diagonal, it is obtained that

$$\left[\Omega_N L^{'-1} D^{-1}\right]_+ = \Omega_N D^{-1}.$$

Therefore,

$$K_f = I - \Omega_N D^{-1} L^{-1}.$$

$\diamond$

2.4.4 Filtering with Incompletely Specified Initial Conditions

We will start by assuming a scalar signal plus noise model (2.14) in which the noise, N, is a zero mean random vector with nonsingular covariance matrix Ω_N and the signal, S, has incompletely specified initial conditions. As we saw earlier in this chapter, this can be modeled by assuming that there exists a lower triangular matrix

$$\Delta = \begin{bmatrix} \Delta_* \\ \Delta_S \end{bmatrix}$$

with ones in the diagonal such that

$$\Delta S = \begin{bmatrix} \delta \\ V \end{bmatrix},$$

where δ and V are mutually orthogonal random vectors, both orthogonal to N, such that V has zero mean and nonsingular covariance matrix Ω_V and the distribution of δ is unknown. Let

$$\Delta^{-1} = [A, B],$$

so that we can write

$$S = A\delta + BV. \tag{2.32}$$

As in Sect. 2.1, we will model δ as a zero mean random vector with covariance matrix $\Omega_\delta = kI$ such that $k \to \infty$ or, equivalently, $\Omega_\delta^{-1} = 0$.

To obtain the filtered estimator of S, we will use the smoothed estimator obtained using the transformation approach. That is,

$$\widehat{S}_{|\infty} = A\hat{\delta}_{|\infty} + B\widehat{V}_{|\infty},$$

where $\hat{\delta}_{|\infty}$ and $\widehat{V}_{|\infty}$ are given by (2.26) and (2.27). The estimator $\hat{\delta}_{|\infty}$ can be rewritten as

$$\hat{\delta}_{|\infty} = \hat{\delta}_{I|\infty} - \Delta_* \Omega_N \Delta_S' \Omega_W^{-1} \Delta_S Y, \tag{2.33}$$

where $\Omega_W = \Omega_V + \Delta_S \Omega_N \Delta_S' = \mathrm{Var}(\Delta_S Y)$ and $\hat{\delta}_{I|\infty}$ is the GLS estimator of δ in the model formed with the first d equations of the model $Y = A\delta + BV + N$ and d is the dimension of δ. This can be seen by first premultiplying the previous equation with the matrix Δ_* to get $\Delta_* Y = \delta + \Delta_* N$ and then considering that $\mathrm{Var}(\Delta_* N)$ is nonsingular. Thus, $\hat{\delta}_{I|\infty} = \Delta_* Y$.

Using (2.27) and (2.33), we can write

$$\widehat{S}_{|\infty} = A\hat{\delta}_{I|\infty} + \left(B\Omega_V - A\Delta_* \Omega_N \Delta_S'\right) \Omega_W^{-1} \Delta_S Y, \tag{2.34}$$

where $\hat{\delta}_{I|\infty} = \Delta_* Y$ and $\Delta_S Y$ are uncorrelated when $k \to \infty$ in $\Omega_\delta = kI$. To prove this, let $X = \Delta_* Y$, $W = \Delta_S Y$ and

$$\mathrm{Var}\begin{bmatrix} X \\ W \end{bmatrix} = \begin{bmatrix} S_{xx} & S_{xw} \\ S_{wx} & S_{ww} \end{bmatrix} = \begin{bmatrix} I & 0 \\ S_{wx}S_{xx}^{-1} & I \end{bmatrix} \begin{bmatrix} S_{xx} & 0 \\ 0 & S_{ww} - S_{wx}S_{xx}^{-1}S_{xw} \end{bmatrix} \begin{bmatrix} I & S_{xx}^{-1}S_{xw} \\ 0 & I \end{bmatrix}.$$

Since $S_{xx} = \mathrm{Var}(\delta + \Delta_* N) = kI + \mathrm{Var}(\Delta_* N)$, it holds that $S_{xx}^{-1} \to 0$ if $k \to \infty$. Thus, if $k \to \infty$, then

$$\mathrm{Var}\begin{bmatrix} X \\ W \end{bmatrix} \to \begin{bmatrix} \infty & 0 \\ 0 & S_{ww} \end{bmatrix}.$$

To obtain the filtered estimator, $\widehat{S}_{f|\infty}$, we can use, as in the previous section, the innovations and the Law of iterated orthogonal projection, Theorem 1.10. Let $\Gamma_W = LDL'$ be the Cholesky decomposition of Γ_W, such that $E = L^{-1}W$ is the vector of innovations and D is a diagonal matrix containing the covariance matrices of E, $D = \mathrm{Var}(E)$. Using this decomposition, we can express $\widehat{S}_{|\infty}$ in (2.34) as

$$\begin{aligned}\widehat{S}_{|\infty} &= \left[A, \left(B\Omega_V - A\Delta_* \Omega_N \Delta_S'\right) L'^{-1} D^{-1}\right] \begin{bmatrix} \hat{\delta}_{I|\infty} \\ E \end{bmatrix} \\ &= K \begin{bmatrix} \hat{\delta}_{I|\infty} \\ E \end{bmatrix}.\end{aligned}$$

From this, it is easy to obtain the filtered estimator as

$$\widehat{S}_{f|\infty} = K_f \begin{bmatrix} \hat{\delta}_{I|\infty} \\ L^{-1}\Delta_S Y \end{bmatrix},$$

where $K_f = \left[A, \left(B\Omega_V - A\Delta_* \Omega_N \Delta_S'\right) L'^{-1} D^{-1}\right]_+$.

It is to be noticed that the first d elements of $\widehat{S}_{f|\infty}$ are more or less arbitrary because in order to estimate δ we need at least d observations. The filtered estimator of the noise, $\widehat{N}_{f|\infty}$ can be obtained in a similar way. The case in which both the signal and the noise have incompletely specified initial conditions in (2.14) is more complicated. See Problem 2.6.

2.4.5 Prediction

The prediction problem can be considered as a special case of filtering. To see this, let $\{Y_t : t = 1, \ldots, n\}$ be a finite sequence of zero mean random vectors and suppose that we know the covariance matrix sequence, $S_{ij} = E(Y_i Y_j')$. If we are interested in predicting Y_{n+h}, $h \geq 1$, and we define $S_N = Y_{n+h}$, then the desired predictor, $\widehat{Y}_{n+h|n}$, is the filtered estimate $\widehat{S}_{n|n}$. Thus,

$$\widehat{Y}_{n+h|n} = \left[\Gamma_{SY} L'^{-1} D^{-1}\right]_+ L^{-1} Y,$$

where $Y = (Y_1', \ldots, Y_n')'$, $\Gamma_{SY} = \mathrm{Cov}(Y_{n+h}, Y)$, $\mathrm{Var}(Y) = LDL'$, L is a block lower triangular matrix with ones in the main diagonal and D is a block diagonal matrix.

Example 2.4 Consider Example 1.3, where $S_{tt} = \sigma^2/(1-\phi^2)$ and $S_{t,t-k} = \phi^k S_{tt}$ if $t \geq 1$, $k = 1, \ldots, t-1$, and the matrices L^{-1} and D^{-1} are given by (1.17). Then, it can be shown (see Problem 2.2) that

$$\widehat{Y}_{n+h|n} = \phi^h Y_t.$$

◇

2.5 Recursive Least Squares

Suppose that we have the linear model

$$Y = H\beta + V, \tag{2.35}$$

where $Y = (Y_1, Y_2, \ldots, Y_n)'$, $V = (V_1, V_2, \ldots, V_n)'$, H is a known $n \times m$ matrix whose rows are denoted by H_t, $t = 1, 2, \ldots, n$, and β is a constant vector that we want to estimate. We will use a Bayesian approach, according to which, β is assumed to have a diffuse prior distribution, that is, $\mathrm{Var}(\beta) = \Pi$ with $\Pi^{-1} \to 0$, and $\mathrm{Var}(V) = \sigma^2 I$. Then, by (2.6), we get

$$\hat{\beta}_n = (H'H)^{-1} H'Y, \qquad \mathrm{MSE}(\hat{\beta}_n) = (H'H)^{-1}\sigma^2,$$

where we use the subscript n to emphasize that the estimation is based on n observations. If a new observation becomes available, we would augment model (2.35) with one row to get

$$\begin{bmatrix} Y \\ Y_{n+1} \end{bmatrix} = \begin{bmatrix} H \\ H_{n+1} \end{bmatrix} \beta + \begin{bmatrix} V \\ V_{n+1} \end{bmatrix}. \tag{2.36}$$

Letting $\hat{\beta}_{n+1}$ be the OLS estimator of (2.36) and $P_{n+1} = \text{MSE}(\hat{\beta}_{n+1})$, by (2.6), we have

$$\hat{\beta}_{n+1} = (H'H + H'_{n+1}H_{n+1})^{-1} \left[H'Y + H'_{n+1}Y_{n+1}\right]$$

and

$$\text{MSE}(\hat{\beta}_{n+1}) = (H'H + H'_{n+1}H_{n+1})^{-1}\sigma^2.$$

These last two equations suggest that we define

$$P_n = (H'H)^{-1}, \quad P_{n+1} = (H'H + H'_{n+1}H_{n+1})^{-1},$$

so that

$$P_{n+1}^{-1} = P_n^{-1} + H'_{n+1}H_{n+1}.$$

This last relation in turn implies

$$P_{n+1}^{-1}\hat{\beta}_{n+1} = P_n^{-1}\hat{\beta}_n + H'_{n+1}Y_{n+1}. \tag{2.37}$$

We have thus obtained the recursive relation

$$\left(P_{n+1}^{-1}, P_{n+1}^{-1}\hat{\beta}_{n+1}\right) = \left(P_n^{-1}, P_n^{-1}\hat{\beta}_n\right) + H'_{n+1}\left(H_{n+1}, Y_{n+1}\right).$$

Evidently, these recursions are valid for $n = 0, 1, 2, \ldots$ and can be initialized with $P_0^{-1} = 0$ and $P_0^{-1}\hat{\beta}_0 = 0$. Note that it is the inverses of the MSE matrices that are propagated, instead of the MSE themselves. We summarize this result in the following theorem.

Theorem 2.8 (Recursive Least Squares in Information Form) *Given the OLS regression model $Y_t = H_t\beta + V_t$ with $Var(V_t) = \sigma^2$, if $\hat{\beta}_n$ and $\sigma^2 P_n$ are the OLS estimator of β and its MSE based on n observations, then the matrices P_{n+1}^{-1} and $P_{n+1}^{-1}\hat{\beta}_{n+1}$ satisfy the recursions*

$$\left(P_{n+1}^{-1}, P_{n+1}^{-1}\hat{\beta}_{n+1}\right) = \left(P_n^{-1}, P_n^{-1}\hat{\beta}_n\right) + H'_{n+1}\left(H_{n+1}, Y_{n+1}\right), \tag{2.38}$$

initialized with $\left(P_0^{-1}, P_0^{-1}\hat{\beta}_0\right) = (0, 0)$.

If we adopt a Bayesian approach and consider that β is a random variable with zero mean and $\mathrm{Var}(\beta) = \sigma^2\Pi$, where Π is a nonsingular matrix such that Π^{-1} expresses the amount of information that we have about β, we can modify the previous recursions so that the MSE instead of the inverses of the MSE are propagated. Applying the Matrix Inversion Lemma 4.1 with $A = P_n^{-1}$, $B = H'_{n+1}$, C= 1 and $D = H_{n+1}$, we obtain a recursive formula for P_n,

$$P_{n+1} = P_n - P_n H'_{n+1}(1 + H_{n+1}P_n H'_{n+1})^{-1}H_{n+1}P_n, \tag{2.39}$$

initialized with $P_0 = \Pi$. Also, it follows from (2.37) and (2.39) that $\hat{\beta}_n$ can be obtained using the following recursion

$$\begin{aligned}
\hat{\beta}_{n+1} &= P_{n+1}\left(P_n^{-1}\hat{\beta}_n + H'_{n+1}Y_{n+1}\right) \\
&= \hat{\beta}_n - P_n H'_{n+1}(1 + H_{n+1}P_n H'_{n+1})^{-1}H_{n+1}\hat{\beta}_n \\
&\quad + \left[P_n H'_{n+1} - P_n H'_{n+1}(1 + H_{n+1}P_n H'_{n+1})^{-1}H_{n+1}P_n H'_{n+1}\right]Y_{n+1} \\
&= \hat{\beta}_n - P_n H'_{n+1}(1 + H_{n+1}P_n H'_{n+1})^{-1}H_{n+1}\hat{\beta}_n \\
&\quad + P_n H'_{n+1}\left[1 - (1 + H_{n+1}P_n H'_{n+1})^{-1}H_{n+1}P_n H'_{n+1}\right]Y_{n+1} \\
&= \hat{\beta}_n + P_n H'_{n+1}(1 + H_{n+1}P_n H'_{n+1})^{-1}\left(Y_{n+1} - H_{n+1}\hat{\beta}_n\right),
\end{aligned} \tag{2.40}$$

initialized with $\hat{\beta}_0 = 0$. Note that

$$E_{n+1} = Y_{n+1} - H_{n+1}\hat{\beta}_n \tag{2.41}$$

is the one step ahead forecast error. To compute its variance, Σ_{n+1}, first consider that $E_{n+1} = V_{n+1} - H_{n+1}(HH')^{-1}HV$. Then, $\mathrm{Var}(E_{n+1}) = \sigma^2\Sigma_{n+1}$, where

$$\Sigma_{n+1} = 1 + H_{n+1}P_n H'_{n+1}. \tag{2.42}$$

We have thus obtained the following theorem.

Theorem 2.9 (Recursive Least Squares) *Suppose the OLS regression model $Y_t = H_t\beta + V_t$ with $Var(V_t) = \sigma^2$ and assume $Var(\beta) = \sigma^2\Pi$ with Π nonsingular. If $\hat{\beta}_n$ and $\sigma^2 P_n$ are the OLS estimator of β and its MSE based on n observations, then $\hat{\beta}_{n+1}$ and P_{n+1} satisfy the recursions*

$$\hat{\beta}_{n+1} = \hat{\beta}_n + K_{n+1}E_{n+1}$$

and

$$P_{n+1} = P_n - P_n H'_{n+1} \Sigma_{n+1}^{-1} H_{n+1} P_n,$$

where $K_{n+1} = P_n H'_{n+1} \Sigma_{n+1}^{-1}$ *and* E_{n+1} *and* Σ_{n+1} *are given by (2.41) and (2.42), initialized with* $\hat{\beta}_0 = 0$ *and* $P_0 = \Pi$.

The recursions of the last theorem are another example of the Kalman filter. They can be obtained by applying the Kalman filter to the model

$$x_{t+1} = x_t \tag{2.43}$$

$$Y_t = H_t x_t + V_t, \tag{2.44}$$

where $x_t = \beta$. The Kalman filter will be described in detail in Chap. 4.

The following lemma will be useful to obtain a formula for the residual sum of squares. It will also be used to obtain the square root information form of recursive least squares (RLS) in the next section.

Lemma 2.1 *Under the assumptions and with the notation of Theorem 2.9, the following relations hold*

$$\Sigma_n^{-1} E_n = Y_n - H_n \hat{\beta}_n \tag{2.45}$$

$$E'_n \Sigma_n^{-1} E_n = Y'_n Y_n + \hat{\beta}'_{n-1} P_{n-1}^{-1} \hat{\beta}_{n-1} - \hat{\beta}'_n P_n^{-1} \hat{\beta}_n. \tag{2.46}$$

Proof By the Matrix Inversion Lemma 4.1 applied to (2.42), it holds that $\Sigma_n^{-1} = 1 - H_n P_n H'_n$ and, using (2.41) and (2.38), it is obtained that

$$\begin{aligned}
\Sigma_n^{-1} E_n &= \left(1 - H_n P_n H'_n\right) E_n = E_n - H_n \hat{\beta}_n + H_n \hat{\beta}_n - H_n P_n H'_n E_n \\
&= E_n - H_n \hat{\beta}_n + H_n \left(\hat{\beta}_n - P_n H'_n E_n\right) = E_n - H_n \hat{\beta}_n + H_n P_n \left(P_n^{-1} \hat{\beta}_n - H'_n E_n\right) \\
&= E_n - H_n \hat{\beta}_n + H_n P_n \left(P_{n-1}^{-1} \hat{\beta}_{n-1} + H'_n Y_n - H'_n Y_n + H'_n H_n \hat{\beta}_{n-1}\right) \\
&= E_n - H_n \hat{\beta}_n + H_n P_n P_n^{-1} \hat{\beta}_{n-1} \\
&= Y_n - H_n \hat{\beta}_n.
\end{aligned}$$

To prove (2.46), use (2.45) and (2.38) to get

$$\begin{aligned}
E'_n \Sigma_n^{-1} E_n &= E'_t \left(Y_n - H_n \hat{\beta}_n\right) = \left(Y_n - H_n \hat{\beta}_{n-1}\right)' Y_n - E'_n H_n \hat{\beta}_n \\
&= Y'_n Y_n - \hat{\beta}'_{n-1} \left(P_n^{-1} \hat{\beta}_n - P_{n-1}^{-1} \hat{\beta}_{n-1}\right) - E'_n H_n \hat{\beta}_n
\end{aligned}$$

$$
\begin{aligned}
&= Y_n'Y_n + \hat{\beta}_{n-1}'P_{n-1}^{-1}\hat{\beta}_{n-1} - \left[\hat{\beta}_{n-1}'P_n^{-1} + \left(Y_n - H_n\hat{\beta}_{n-1}\right)' H_n\right]\hat{\beta}_n \\
&= Y_n'Y_n + \hat{\beta}_{n-1}'P_{n-1}^{-1}\hat{\beta}_{n-1} \\
&\quad - \left(\hat{\beta}_{n-1}'P_n^{-1} + \hat{\beta}_n'P_n^{-1} - \hat{\beta}_{n-1}'P_{n-1}^{-1} - \hat{\beta}_{n-1}'P_n^{-1} + \hat{\beta}_{n-1}'P_{n-1}^{-1}\right)\hat{\beta}_n \\
&= Y_n'Y_n + \hat{\beta}_{n-1}'P_{n-1}^{-1}\hat{\beta}_{n-1} - \hat{\beta}_n'P_n^{-1}\hat{\beta}_n.
\end{aligned}
$$

□

The following theorem gives a formula for the recursive computation of the residual sum of squares. We omit its proof because it is an immediate consequence of the previous lemma.

Theorem 2.10 (Residual Sum of Squares) *Under the assumptions and with the notation of Theorem 2.9, the following formula holds for the residual sum of squares*

$$
\sum_{t=1}^{n} E_t'\Sigma_t^{-1}E_t = \sum_{t=1}^{n} Y_t'Y_t - \hat{\beta}_n'P_n^{-1}\hat{\beta}_n = Y'Y - \hat{\beta}_n'P_n^{-1}\hat{\beta}_n. \tag{2.47}
$$

Remark 2.7 Since $\hat{\beta}_n'P_n^{-1}\hat{\beta}_n = \left(P_n^{-1}\hat{\beta}_n\right)' P_n \left(P_n^{-1}\hat{\beta}_n\right)$, this quantity can be computed using the recursion (2.38). Thus, to compute (2.47), we can add to (2.38) the recursion

$$
SS_t = SS_{t-1} + Y_t'Y_t, \tag{2.48}
$$

initialized with $SS_0 = 0$, that computes the sum of squares of the observations. ◇

Remark 2.8 The residual sum of squares (2.47) is also equal to $(Y - H\hat{\beta}_n)'(Y - H\hat{\beta}_n)$. ◇

2.5.1 Square Root Form of RLS

The square root form of RLS is basically an algorithm to compute the OLS estimator and its MSE in a regression model in a numerically safe way. It is well known that it is important to have a stable numerical procedure to compute these quantities. One of such procedures consists of using Householder transformations and the *QR* decomposition of a matrix, described in the Appendix. This is what the square root form of RLS does.

The square root form of RLS propagates the square roots of the MSE matrices instead of the MSE matrices themselves, where for any square matrix A a square root of A, denoted by $A^{1/2}$, is any matrix such that $A = A^{1/2}A^{1/2'}$. The square root form of RLS is given by the following theorem.

Theorem 2.11 *Suppose that the process $\{Y_t\}$ follows the state space model (2.43) and (2.44) corresponding to the regression model $Y_t = H_t\beta + V_t$ with $Var(V_t) = \sigma^2$ and assume $Var(\beta) = \sigma^2\Pi$ with Π nonsingular. Then, with the notation of Theorem 2.9, the application of the QR algorithm yields an orthogonal matrix U_t such that*

$$U_t'\begin{bmatrix} 1 & 0 \\ P_t^{1/2'}H_t' & P_t^{1/2'} \end{bmatrix} = \begin{bmatrix} \Sigma_t^{1/2'} & \widehat{K}_t' \\ 0 & P_{t+1}^{1/2'} \end{bmatrix}, \tag{2.49}$$

where $\widehat{K}_t = P_t H_t'\Sigma_t^{-1/2'} = K_t\Sigma_t^{1/2}$. Thus, letting $\widehat{E}_t = \Sigma_t^{-1/2}E_t$, $\hat{\beta}_{t+1}$ can be obtained as $\hat{\beta}_{t+1} = \hat{\beta}_t + \widehat{K}_t\widehat{E}_t$. In addition, the same matrix U_t satisfies

$$U_t'\left[\begin{array}{cc|c} 1 & 0 & -Y_t \\ P_t^{1/2'}H_t' & P_t^{1/2'} & P_t^{-1/2}\hat{\beta}_t \end{array}\right] = \left[\begin{array}{cc|c} \Sigma_t^{1/2'} & \widehat{K}_t' & -\widehat{E}_t \\ 0 & P_{t+1}^{1/2'} & P_{t+1}^{-1/2}\hat{\beta}_{t+1} \end{array}\right]. \tag{2.50}$$

In this case, $\hat{\beta}_{t+1}$ is obtained as $\hat{\beta}_{t+1} = P_{t+1}^{1/2}\left[P_{t+1}^{-1/2}\hat{\beta}_{t+1}\right]$.

Proof The matrix U_t satisfies

$$U_t'\begin{bmatrix} 1 & 0 \\ P_t^{1/2'}H_t' & P_t^{1/2'} \end{bmatrix} = \begin{bmatrix} \Sigma' & K' \\ 0 & P' \end{bmatrix}.$$

Premultiplying the matrices in (2.49) by their respective transposes yields

$$\begin{aligned} 1 + H_tP_t^{1/2}P_t^{1/2'}H_t' &= \Sigma\Sigma' \\ H_tP_t^{1/2}P_t^{1/2'} &= \Sigma K', \\ P_t^{1/2}P_t^{1/2'} &= KK' + PP', \end{aligned}$$

and the first part of the theorem follows. To prove the second part, consider the first and the third and the second and the third block columns of (2.50). Then, it is obtained that

$$\begin{aligned} -Y_t + H_tP_t^{1/2}P_t^{-1/2}\hat{\beta}_t &= -\Sigma_t^{1/2}\widehat{E}_t, \\ P_t^{1/2}P_t^{-1/2}\hat{\beta}_t &= P_{t+1}^{1/2}P_{t+1}^{-1/2}\hat{\beta}_{t+1} - \widehat{K}_t\widehat{E}_t. \end{aligned}$$

□

2.5.2 Fast Square Root Algorithms for RLS: The UD *Filter*

Using square root free fast Givens rotations, described in the Appendix to this chapter, it is possible to substantially reduce the amount of computation needed for the square root form of RLS, as shown by Jover & Kailath (1986). The idea is to put first P_t into the form $P_t = L_t D_t L_t'$, where L_t is a lower triangular matrix with ones in the main diagonal and D_t is a diagonal matrix with positive elements in the main diagonal, and then to update L_t and D_t using the *QDU* decomposition, described in the Appendix to this chapter. Using this decomposition, we can write expression (2.49) of Theorem 2.11 as

$$U' \begin{bmatrix} 1 & 0 \\ 0 & D_t^{1/2} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ L_t'H_t' & L_t' \end{bmatrix} = \begin{bmatrix} \Sigma_t^{1/2'} & 0 \\ 0 & D_{t+1}^{1/2} \end{bmatrix} \begin{bmatrix} 1 & K_t' \\ 0 & L_{t+1}' \end{bmatrix}. \tag{2.51}$$

Then, it is clear that we can use fast Givens rotations, as described in the Appendix to this chapter, to obtain the *QDU* decomposition, to update L_t and D_t.

Another fast algorithm that can be used for the square root form of RLS is the so-called *UD* filter, due to Bierman (1977). This algorithm assumes that the covariance matrix P_t is factored in the form $P_t = U_t D_t U_t'$, where U_t is an upper triangular matrix with ones in the main diagonal and D_t is a diagonal matrix with positive elements in the main diagonal. We will describe the algorithm for the decomposition in terms of lower triangular matrices instead, for coherence with the rest of the book. However, this amounts to a small change in the algorithm.

Let $P_t = P = LDL'$ and $P_{t+1} = P^+ = L_+ D_+ L_+'$. Then, we have to factor

$$L_+ D_+ L_+' = L(D - L'h'\Sigma^{-1}hLD)L',$$

where $h = H_t$ and $\Sigma = \Sigma_t$. Since $L'h'\Sigma^{-1}hLD$ has rank one, Bierman uses a rank one downdating formula due to Agee & Turner (1972) for the factorization of the term in parenthesis in the previous expression. The whole procedure is described in Bierman (1977). Letting $hL = f = [f_1, \ldots, f_n]$, $L = [l_1, \ldots, l_n]$, $L_+ = [l_1^+, \ldots, l_n^+]$, $D = \text{diag}(d_i)$ and $D_+ = \text{diag}(d_i^+)$, the algorithm is as follows.

$\alpha_{n+1} = 1$
$k_{n+1} = 0$
for $i = n, n-1, \ldots, 1$

$\alpha_i = \alpha_{i+1} + d_i f_i^2$
$d_i^+ = d_i(\alpha_{i+1}/\alpha_i)$
$[k_i \quad l_i^+] = [k_{i+1} \quad l_i] \begin{bmatrix} 1 & -f_i/\alpha_{i+1} \\ d_i f_i & 1 \end{bmatrix}$

end

On completion of the algorithm, $\alpha_1 = \Sigma = 1 + hPh'$, $k_1 = Ph'$ and $K_t = k_1/\alpha_1$, where we have omitted the time index for simplicity.

As noted by Jover & Kailath (1986), the *UD* filter is equivalent to the updating (2.51), that uses fast Givens rotations.

It is to be noted that Bierman (1977) emphasized that, with careful programming, the number of computations of the *UD* filter is approximately the same as that of ordinary RLS.

2.5.3 Square Root Information Form of RLS

Suppose the regression model (2.35) with H of full column rank. If we apply the QR algorithm to the matrix H, an orthogonal matrix Q is obtained such that

$$Q'H = \begin{bmatrix} R \\ 0 \end{bmatrix}, \tag{2.52}$$

where R is nonsingular upper triangular. If we extend the matrix H to $[H \quad Y]$, then

$$Q'[H \quad Y] = \begin{bmatrix} R & u \\ 0 & v \end{bmatrix}$$

and the model (2.35) is transformed into the model

$$\begin{aligned} u &= R\beta + w_1 \\ v &= \qquad\ w_2, \end{aligned}$$

where $w = [w_1', w_2'] = Q'V$ and the partition is conformal with (2.52). Since $\mathrm{Var}(w) = \sigma^2 Q'Q = \sigma^2 I$, the OLS estimator of β in the previous model is

$$\hat{\beta}_n = (R'R)^{-1}R'u = R^{-1}u$$

and $\mathrm{MSE}(\hat{\beta}_n) = \sigma^2(R'R)^{-1}$. Thus, $P_n^{-1} = R'R$,

$$u = P_n^{-1/2}\hat{\beta}_n \tag{2.53}$$

$$R = P_n^{-1/2}, \tag{2.54}$$

and, since $Y'Y = u'u + v'v = \hat{\beta}_n' P_n^{-1}\hat{\beta}_n + v'v$, by Theorem 2.10,

$$v'v = \sum_{t=1}^{n} E_t' \Sigma_t^{-1} E_t. \tag{2.55}$$

If a new observation becomes available, we can apply the QR algorithm to the augmented matrix of model (2.36) and compute the estimator of β and its MSE corresponding to the new model. However, it is better to proceed recursively and apply the QR algorithm to the matrix

$$\begin{bmatrix} R & u \\ H_{n+1} & Y_{n+1} \end{bmatrix}$$

to get an orthogonal matrix Q_n such that

$$Q'_n \begin{bmatrix} R & u \\ H_{n+1} & Y_{n+1} \end{bmatrix} = \begin{bmatrix} R_{n+1} & u_{n+1} \\ 0 & v_{n+1} \end{bmatrix}.$$

Then, premultiplying the previous matrix by its transpose and considering (2.53), (2.54) and (2.55), it is obtained that

$$\begin{aligned} P_n^{-1/2'} P_n^{-1/2} + H'_{n+1} H_{n+1} &= R'_{n+1} R_{n+1} \\ \hat{\beta}'_n P_n^{-1/2'} P_n^{-1/2} \hat{\beta}_n + Y'_{n+1} Y_{n+1} &= u'_{n+1} u_{n+1} + v'_{n+1} v_{n+1} \\ \hat{\beta}'_n P_n^{-1/2'} P_n^{-1/2} + Y'_{n+1} H_{n+1} &= u'_{n+1} R_{n+1}. \end{aligned}$$

The following theorem gives a precise meaning to R_{n+1}, u_{n+1} and v_{n+1}. It turns out that these last quantities allow for the recursive propagation of the square root of the inverses of the MSE matrices.

Theorem 2.12 (Square Root Information Form of RLS) *Under the assumptions and with the notation of Theorem 2.9, the QR algorithm produces an orthogonal matrix Q_{n-1} such that*

$$Q'_{n-1} \begin{bmatrix} P_{n-1}^{-1/2} \\ H_n \end{bmatrix} = \begin{bmatrix} P_n^{-1/2} \\ 0 \end{bmatrix}, \tag{2.56}$$

where $P_n^{-1/2}$ is upper triangular or upper trapezoidal and $(P_0^{-1/2}, P_0^{-1/2}\hat{\beta}_0) = (0, 0)$. In addition,

$$Q'_{n-1} \begin{bmatrix} P_{n-1}^{-1/2} & P_{n-1}^{-1/2}\hat{\beta}_{n-1} \\ H_n & Y_n \end{bmatrix} = \begin{bmatrix} P_n^{-1/2} & P_n^{-1/2}\hat{\beta}_n \\ 0 & \Sigma_n^{-1/2} E_n \end{bmatrix}, \tag{2.57}$$

and if we add to the second matrix in the left-hand side an extra block column of the form $[0, 1]'$, then the same matrix Q_{n-1} satisfies

$$Q'_{n-1} \begin{bmatrix} P_{n-1}^{-1/2} & P_{n-1}^{-1/2}\hat{\beta}_{n-1} & 0 \\ H_n & Y_n & 1 \end{bmatrix} = \begin{bmatrix} P_n^{-1/2} & P_n^{-1/2}\hat{\beta}_n & P_n^{1/2'} H'_n \\ 0 & \Sigma_n^{-1/2} E_n & \Sigma_n^{-1/2} \end{bmatrix}. \tag{2.58}$$

Proof Since Q_{n-1} is orthogonal, we get from the left-hand side of (2.57), using (2.46) and (2.38),

$$\begin{bmatrix} P_{n-1}^{-1/2} & P_{n-1}^{-1/2}\hat{\beta}_{n-1} \\ H_n & Y_n \end{bmatrix}' Q_{n-1}Q'_{n-1} \begin{bmatrix} P_{n-1}^{-1/2} & P_{n-1}^{-1/2}\hat{\beta}_{n-1} \\ H_n & Y_n \end{bmatrix}$$
$$= \begin{bmatrix} P_{n-1}^{-1/2'}P_{n-1}^{-1/2} + H'_nH_n & P_n^{-1/2'}P_{n-1}^{-1/2}\hat{\beta}_{n-1} + H'_nY_n \\ \hat{\beta}'_{n-1}P_{n-1}^{-1/2'}P_{n-1}^{-1/2} + Y'_nH_n & \hat{\beta}'_{n-1}P_{n-1}^{-1/2'}P_{n-1}^{-1/2}\hat{\beta}_{n-1} + Y'_nY_n \end{bmatrix}$$
$$= \begin{bmatrix} P_n^{-1/2'}P_n^{-1/2} & P_n^{-1/2'}P_n^{-1/2}\hat{\beta}_n \\ \hat{\beta}'_nP_n^{-1/2'}P_n^{-1/2} & \hat{\beta}'_nP_n^{-1/2'}P_n^{-1/2}\hat{\beta}_n + E'_n\Sigma_n^{-1/2'}\Sigma_n^{-1/2}E_n \end{bmatrix}$$
$$= \begin{bmatrix} P_n^{-1/2} & P_n^{-1/2}\hat{\beta}_n \\ 0 & \Sigma_n^{-1/2}E_n \end{bmatrix}' \begin{bmatrix} P_n^{-1/2} & P_n^{-1/2}\hat{\beta}_n \\ 0 & \Sigma_n^{-1/2}E_n \end{bmatrix}.$$

The rest of the theorem can be proved similarly. □

Remark 2.9 The vectors $\Sigma_t^{-1/2}E_t$ have zero mean and unit variance. Thus, they constitute a sequence of "standardized residuals" and can be used for inference. ◇

Remark 2.10 To compute $\hat{\beta}_n$ in the square root algorithm of the previous theorem, we have to solve the system $\left(P_n^{-1/2}\right)\hat{\beta}_n = P_n^{-1/2}\hat{\beta}_n$, where $P_n^{-1/2}$ is an upper triangular matrix. This computation, based on back substitution, can be avoided by including in the second matrix in the left-hand side of (2.58) an extra block column of the form $[0, P_{n-1}^{1/2}]'$ because

$$Q'_{n-1}\begin{bmatrix} 0 \\ P_{n-1}^{1/2'} \end{bmatrix} = \begin{bmatrix} P_n^{1/2'} \\ \Sigma_n^{-1/2}E_nP_{n-1} \end{bmatrix}.$$

The validity of this formula can be verified using (2.39) and (2.42). ◇

2.5.4 Fast Square Root Information Algorithm for RLS

As in the case of the square root form of RLS, it is possible to use square root free fast Givens rotations, described in the Appendix, to substantially reduce the amount of computation needed for the square root information form of RLS. To this end, assume first that P_t^{-1} is nonsingular and put P_t^{-1} into the form $P_t^{-1} = L_tD_tL'_t = P_t^{-1/2'}P_t^{-1/2}$, where L_t is a lower triangular matrix with ones in the main diagonal and D_t is a diagonal matrix with positive elements in the main diagonal. Then, we can write expression (2.56) of Theorem 2.12 as

$$Q' \begin{bmatrix} D_{n-1}^{1/2} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} L'_{n-1} \\ H_n \end{bmatrix} = \begin{bmatrix} D_n^{1/2} L'_n \\ 0 \end{bmatrix}.$$

We can thus clearly use fast Givens rotations, as described in the Appendix to obtain the QDU decomposition, to update L_t and D_t.

2.6 Historical Notes

In a seminal paper, Kalbfleisch & Sprott (1970) proposed several definitions of the likelihood in the presence of nuisance parameters. A similar approach has been used in this chapter to deal with the problem of defining the likelihood of a linear model when a part of this model has an unspecified distribution. In the statistical literature, this subject has also been considered in the context of state space models with incompletely specified initial conditions. See, for example, Ansley & Kohn (1985) and Francke, Koopman, & Vos (2010).

The classical reference for signal extraction with scalar variables is Whittle (1963b). For the doubly infinite sample in the nonstationary scalar case, Bell (1984) proposed two Assumptions, that he called A and B. He proved that the usual Wiener–Kolmogorov formulae (see Chap. 7) are valid under Assumption A but not under Assumption B. However, it is interesting to note that both assumptions lead to the same result in the finite sample case. See Gómez (1999, p. 9). Assumption B is assumed in all the examples of finite nonstationary series considered in Sect. 2.4. For a derivation of some of these formulae under Assumption A instead of Assumption B, see McElroy (2008).

The origin of the QR decomposition goes back to Housholder (1953, pp. 72–73), who is considered to be one of the pioneers of numerical analysis. It seems that this decomposition first gained attention through a paper by Golub (1965).

2.7 Problems

2.1 Use Example 1.3 and in particular the lower triangular matrix L and diagonal matrix D obtained in that example such that $\mathrm{Var}(Y) = LDL'$ to compute Ω_S^{-1} as $\Omega_S^{-1} = L_S^{-1\prime} D_S^{-1} L_S^{-1}$ in Example 2.1. Since L and D in example 1.3 satisfy

$$L^{-1} = \begin{bmatrix} 1 & & & & & \\ -\phi & 1 & & & & \\ 0 & -\phi & 1 & & & \\ \vdots & \vdots & \vdots & \ddots & & \\ 0 & 0 & \vdots & -\phi & 1 & \\ 0 & 0 & \vdots & 0 & -\phi & 1 \end{bmatrix}, \quad D^{-1} = \begin{bmatrix} (1-\phi^2)/\sigma^2 & & & & \\ & 1/\sigma^2 & & & \\ & & 1/\sigma^2 & & \\ & & & \ddots & \\ & & & & 1/\sigma^2 \end{bmatrix},$$

show that

$$\Omega_S^{-1} = \frac{1}{\sigma_a^2} \begin{bmatrix} 1 & -\phi & & & & \\ -\phi & 1+\phi^2 & -\phi & & & \\ & \ddots & \ddots & \ddots & & \\ & & & & -\phi & \\ & & & -\phi & 1+\phi^2 & -\phi \\ & & & & -\phi & 1 \end{bmatrix}.$$

2.2 Prove the result of Example 2.4. That is, given Example 1.3, where $S_{tt} = \sigma^2/(1-\phi^2)$, $S_{t,t-k} = \phi^k S_{tt}$, $t \geq 1$, $k = 1, \ldots, t-1$, and the matrices L^{-1} and D^{-1} are given by (1.17), show that

$$\widehat{Y}_{n+h|n} = \phi^h Y_t.$$

2.3 Suppose the scalar signal plus noise model

$$Y_t = S_t + N_t,$$

where both the signal and the noise have incompletely specified initial conditions, and let $Y = (Y_1, \ldots, Y_n)'$, $S = (S_1, \ldots, S_n)'$ and $N = (N_1, \ldots, N_n)'$. Assume further that there exist two lower triangular matrices with ones in the diagonal,

$$\overline{\Delta}_S = \begin{bmatrix} \Delta_{*S} \\ \Delta_S \end{bmatrix}, \qquad \overline{\Delta}_N = \begin{bmatrix} \Delta_{*N} \\ \Delta_N \end{bmatrix},$$

such that

$$\overline{\Delta}_S S = \begin{bmatrix} \delta_S \\ V_S \end{bmatrix}, \qquad \overline{\Delta}_N N = \begin{bmatrix} \delta_N \\ V_N \end{bmatrix},$$

where $[\delta_S', \delta_N']$ and $[V_S', V_N']$ are orthogonal random vectors and the distribution of $[\delta_S', \delta_N']$ is unknown. Then, multiplying the two "differencing" polynomials involved in the definition of Δ_S and Δ_N, that are assumed to have no common factors, construct in a way similar to that used to construct these last two matrices a new lower triangular matrix with ones in the diagonal,

$$\overline{\Delta} = \begin{bmatrix} \Delta_* \\ \Delta \end{bmatrix},$$

such that

$$\overline{\Delta} Y = \begin{bmatrix} \delta \\ W \end{bmatrix},$$

where δ has an unspecified and W has a known distribution. For example, if S_t and N_t follow the models $S_t + S_{t-1} = B_t$ and $N_t - N_{t-1} = C_t$ and there are four observations, the differencing polynomials are $p_S(z) = 1 + z$ and $p_N(z) = 1 - z$, the product is $p(z) = 1 - z^2$, and the previous matrices could be

$$\overline{\Delta}_S = \left[\begin{array}{cccc} 1 & 0 & 0 & 0 \\ \hline 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{array}\right], \quad \overline{\Delta}_N = \left[\begin{array}{cccc} 1 & 0 & 0 & 0 \\ \hline -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{array}\right], \quad \overline{\Delta} = \left[\begin{array}{cccc} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \hline -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 \end{array}\right].$$

Let d_S and d_N be the dimensions of δ_S and δ_N, respectively, and let $d = d_S + d_N$ be the dimension of δ. Define $\tilde{\Delta}_S$ as the $(n-d) \times (n-d_N)$ matrix formed with the last $n-d$ rows and the last $n-d_N$ columns of Δ_S and define $\tilde{\Delta}_N$ as the $(n-d) \times (n-d_S)$ matrix formed similarly using Δ_N. For the previous example, these matrices are

$$\tilde{\Delta}_S = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}, \quad \tilde{\Delta}_N = \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}.$$

Prove that

$$\Delta = \tilde{\Delta}_S \Delta_N = \tilde{\Delta}_N \Delta_S. \tag{2.59}$$

2.4 With the assumptions and notation of Problem 2.3, let

$$\overline{\Delta}_S^{-1} = [A_S, B_S], \qquad \overline{\Delta}_N^{-1} = [A_N, B_N],$$

so that

$$S = A_S \delta_S + B_S V_S, \qquad N = A_N \delta_N + B_N V_N.$$

Estimate V_S and V_N as follows. Use first (2.59) in Problem 2.3 to get

$$\Delta Y = W = \tilde{\Delta}_N V_S + \tilde{\Delta}_S V_N.$$

Then, show that

$$\widehat{V}_{S|\infty} = \Omega_{VS}\tilde{\Delta}'_N\Omega_W^{-1}\Delta Y, \quad \widehat{V}_{N|\infty} = \Omega_{VN}\widetilde{\Delta}'_S\Omega_W^{-1}\Delta Y$$

and

$$P_{S|\infty} = \Omega_{VS} - \Omega_{VS}\widetilde{\Delta}'_N\Omega_W^{-1}\widetilde{\Delta}_N\Omega_{VS}, \quad P_{N|\infty} = \Omega_{VN} - \Omega_{VN}\widetilde{\Delta}'_S\Omega_W^{-1}\widetilde{\Delta}_S\Omega_{VN},$$

where $\mathrm{Var}(W) = \Omega_W = \widetilde{\Delta}_N\Omega_{VS}\widetilde{\Delta}'_N + \widetilde{\Delta}_S\Omega_{VN}\widetilde{\Delta}'_S$, $\mathrm{Var}(V_S) = \Omega_{VS}$ and $\mathrm{Var}(V_N) = \Omega_{VN}$.

Inserting $\widehat{V}_{S|\infty}$, $\widehat{V}_{N|\infty}$ in the model

$$\begin{aligned}\Delta_N Y &= \Delta_N S + V_N \\ &= \Delta_N A_S\delta_S + \Delta_N B_S V_S + V_N,\end{aligned} \tag{2.60}$$

obtain first the equation

$$\Delta_N A_S\hat{\delta}_{S|\infty} = \Delta_N Y - \Delta_N B_S\widehat{V}_{S|\infty} - \widehat{V}_{N|\infty}.$$

Then, letting $\Delta_{N,dS}$ be the submatrix of Δ_N formed with the first d_S rows, prove that under the assumption that the two differencing polynomials have no common factors $\Delta_{N,dS}A_S$ is invertible. Letting $M_S = [\Delta_{N,dS}A_S]^{-1}$, conclude that

$$\begin{aligned}\hat{\delta}_{S|\infty} &= M_S\Delta_{N,dS}\left(Y - B_S\widehat{V}_{S|\infty} - B_N\widehat{V}_{N|\infty}\right) \\ &= M_S\Delta_{N,dS}\left[I - \left(B_S\Omega_{VS}\widetilde{\Delta}'_N + B_N\Omega_{VN}\widetilde{\Delta}'_S\right)\Omega_W^{-1}\Delta\right]Y.\end{aligned}$$

Obtain $\mathrm{MSE}(\hat{\delta}_{S|\infty})$.

2.5 With the assumptions and notation of Problem 2.3, obtain $\hat{\delta}_{S|\infty}$ and $\widehat{V}_{S|\infty}$ in model (2.60) in two steps. First, considering δ_S fixed, estimate V_S. Then, in the second step, estimate δ_S by GLS in (2.60). Finally, plug this last estimator in the estimator of V_S obtained in the first step.

2.6 With the assumptions and notation of Problem 2.3, obtain the filtered estimator, $\widehat{S}_{f|\infty}$, of S as follows. Use first Problem 2.4 to get the smoothed estimator, $\widehat{S}_{|\infty}$, of S as

$$\widehat{S}_{|\infty} = A_S\hat{\delta}_{S|\infty} + B_S\widehat{V}_{S|\infty}.$$

Prove that $\hat{\delta}_{S|\infty}$ can be rewritten as

$$\hat{\delta}_{S|\infty} = \hat{\delta}_{S,I|\infty} - M_S\Delta_{N,dS}\left(B_S\Omega_{VS}\widetilde{\Delta}'_N + B_N\Omega_{VN}\widetilde{\Delta}'_S\right)\Omega_W^{-1}\Delta Y,$$

where $\hat{\delta}_{S,I|\infty} = M_S \Delta_{N,dS} Y$ is the GLS estimator of δ_S in the first d_S equations of the model $\Delta_N Y = \Delta_N A_S \delta_S + \Delta_N B_S V_S + V_N$, so that $\widehat{S}_{|\infty}$ can be expressed as

$$\widehat{S}_{|\infty} = A_S \hat{\delta}_{S,I|\infty} + \left[B_S \Omega_{VS} \widetilde{\Delta}_N' - M_S \Delta_{N,dS} \left(B_S \Omega_{VS} \widetilde{\Delta}_N' + B_N \Omega_{VN} \widetilde{\Delta}_S' \right) \right] \Omega_W^{-1} \Delta Y.$$

Prove as in Sect. 2.4.4 that $\hat{\delta}_{S,I|\infty}$ and ΔY are uncorrelated, let $\Omega_W = LDL'$ be the Cholesky decomposition of Ω_W, define the innovations of $W = \Delta Y$ as $E = L^{-1} \Delta Y$ and conclude that

$$\widehat{S}_{f|\infty} = K_{S,f} \begin{bmatrix} \hat{\delta}_{S,I|\infty} \\ L^{-1} \Delta Y \end{bmatrix},$$

where

$$K_{S,f} = \left[A_S, \left\{ B_S \Omega_{VS} \widetilde{\Delta}_N' - M_S \Delta_{N,dS} \left(B_S \Omega_{VS} \widetilde{\Delta}_N' + B_N \Omega_{VN} \widetilde{\Delta}_S' \right) \right\} L^{-1'} D^{-1} \right]_+ .$$

2.7 Given a square matrix A, we can always write for $m \geq 0$

$$A = [A]_{+m} + [A]_{-m},$$

where the (i,j)-th elements of $[A]_{+m}$ and $[A]_{-m}$ are

$$[A]_{+m}(ij) = \begin{cases} A_{ij} & i \leq j + m \\ 0 & i > j + m, \end{cases} \qquad [A]_{-m}(ij) = \begin{cases} 0 & i \leq j + m \\ A_{ij} & i > j + m. \end{cases}$$

Note that if $m = 0$, we obtain $[A]_{+0} = [A]_+$ and $[A]_{-0} = [A]_-$, where $[A]_+$ and $[A]_-$ are the matrices defined in Sect. 2.4.3. Let $\{S_t\}$ and $\{Y_t\}$, $1 \leq t \leq n$, be two sequences of zero mean random vectors and let Γ_{SY} and Γ_Y be the cross-covariance and covariance matrices, that is $\Gamma_{SY}(i,j) = E(S_i Y_j')$ and $\Gamma_Y(i,j) = E(Y_i Y_j')$.

(a) Prove that if $\widehat{S}_m = (\widehat{S}_{1|1-m}', \ldots, \widehat{S}_{n|n-m}')'$ and $m > 0$, where $\widehat{S}_{i|j} = E^*(S_i | Y_1, \ldots, Y_j)$ and it is understood that $Y_j = 0$ if $j < 0$, then $\widehat{S}_m = K_m Y$, where K_m is a lower triangular matrix with zeros on its diagonal and first $(m-1)$ subdiagonals. If $m = 0$, then $K_0 = K_f$ is just a lower triangular matrix, as we know from Sect. 2.4.3.
(b) Show that $K_m = \left[\Gamma_{SY} L^{-1'} D^{-1} \right]_{+m} L^{-1}$, where $\Gamma_Y = LDL'$ is the block Cholesky decomposition of Γ_Y.
(c) Define $U = K_m \Gamma_Y - \Gamma_{SY}$. Show that $U = -\left[\Gamma_{SY} L^{-1'} D^{-1} \right]_{-m} DL'$.
(d) Show that $K_m = \Gamma_{SY} \Gamma_Y^{-1} + U \Gamma_Y^{-1} = (\Gamma_{SY} + U) \Gamma_Y^{-1}$.
(e) Prove that $\text{MSE}(\widehat{S}_m) = \Gamma_S - \Gamma_{SY} \Gamma_Y^{-1} \Gamma_{YS} + U \Gamma_Y^{-1} U'$.

2.8 Let $Y = S + N$, where S and N are uncorrelated random vectors with covariance matrices Ω_S and Ω_N such that Ω_N is block diagonal.

(a) Prove using Problem 2.7 that $U = \left[\Omega_N L^{-1'} D^{-1} - L\right]_{-m} DL'$.
(b) When $m = 1$ (prediction), show that $K_1 = I - L^{-1}$ and $U = \Omega_N - DL'$.
(c) Relate $\widehat{S}_1$ with $\widehat{S}_0 = \widehat{S}_f$. Relate also $\widehat{S}_{t|t}$ with $\widehat{S}_{t|t-1}$.

2.9 Consider Example 2.2 with $\{Y_t : t = 1, 2, 3\}$, $\sigma_A^2 = 1$ and $\sigma_N^2 = 2$.

(a) Compute $\widehat{S}_{|\infty}$ and $P_{|\infty}$ using the information form formulas (2.19) and (2.20).
(b) Compute $\widehat{S}_{|\infty}$ and $P_{|\infty}$ using the noninformation form formulas (2.23) and (2.21).
(c) Compute $\left[\hat{\delta}_{|\infty}, \widehat{V}'_{|\infty}\right]'$ and its MSE using formulas (2.24) and (2.25).

2.10 Consider Example 2.1 with $\{Y_t : t = 1, 2, 3\}$, $\phi = \sqrt{2}/2$, $\sigma_A^2 = 1$ and $\sigma_N^2 = 2$. Compute the filtered estimator, $\widehat{S}_f$, as in Example 2.3. To this end, compute the matrices L^{-1} and D^{-1} corresponding to the Cholesky decomposition $\Gamma_Y = LDL'$ using first the covariance based filter (1.43) corresponding to the state space model

$$x_{t+1} = \phi x_t + A_t$$
$$Y_t = x_t + N_t,$$

where $\gamma(0) = \sigma_N^2 + \Pi$, $\Pi = \sigma_A^2/(1 - \phi^2)$ and $N = \phi\Pi$, and then formula

$$L^{-1} = \begin{bmatrix} 1 & & \\ -K_1 & 1 & \\ -F_{p,2}^3 K_1 & -K_2 & 1 \end{bmatrix},$$

where $F_{p,2} = \phi - K_2$, based on Corollary 1.3.

Appendix

Orthogonal Transformations and the QR and QDU Decompositions

2.A.1 Householder Transformations

Let $x = [x_1, \ldots, x_n]'$ be an n-dimensional real valued vector and suppose that we wish to simultaneously annihilate several entries in it by using an orthogonal involutory matrix Θ (i.e., a transformation Θ that satisfies $\Theta\Theta' = I = \Theta'\Theta$ and $\Theta^2 = I$). More specifically, let $e_1 = [1, 0, \ldots, 0]'$ and suppose that we want Θ to satisfy

$$\Theta[x_1, x_2, \ldots, x_n]' = \alpha e_1 = \alpha[1, 0, \ldots, 0]' \tag{2A.1}$$

for some real scalar α. Then,

$$x'\Theta'\Theta x = \alpha^2 e_1' e_1 = \alpha^2,$$

$x'x = \alpha^2$, and $\alpha = \pm\sqrt{x'x} = \pm||x||$. Define

$$\alpha = \begin{cases} ||x|| & \text{if} \quad x_1 > 0 \\ -||x|| & \text{if} \quad x_1 < 0, \end{cases}$$

$$v = \frac{1}{\alpha}x + e_1,$$

and

$$\Theta = I - \frac{1}{v_1}vv', \tag{2A.2}$$

where $v = [v_1, \ldots, v_n]'$. Then, $\Theta = \Theta'$ and

$$\begin{aligned} \Theta'\Theta &= I - \frac{2}{v_1}vv' + \frac{1}{v_1^2}vv'vv' \\ &= I - \frac{2}{v_1}vv' + \frac{1}{v_1^2}vv'\left(\frac{1}{\alpha}x' + e_1'\right)\left(\frac{1}{\alpha}x + e_1\right) \\ &= I - \frac{2}{v_1}vv' + \frac{1}{v_1^2}vv'2v_1 \\ &= I. \end{aligned}$$

Thus, Θ is orthogonal and involutory. In addition,

$$\begin{aligned} \Theta x &= x - \frac{1}{v_1}vv'x \\ &= x - \frac{1}{v_1}v(\alpha + x_1) \\ &= x - \frac{1}{v_1}v\alpha v_1 \\ &= -\alpha e_1, \end{aligned}$$

and Θ, as defined in (2A.2), satisfies (2A.1) with $\alpha = \mp\sqrt{x'x}$.

Note that $v'v = 2v_1$ and, therefore, Θ can also be written as

$$\Theta = I - \beta vv',$$

where $\beta = 2/v'v$.

2.A.2 The QR Decomposition

Let A be an $m \times n$ matrix with $m \geq n$ and full column rank, and suppose that we want to triangularize A in the sense that we look for an orthogonal matrix Q such that

$$Q'A = \begin{bmatrix} R \\ 0 \end{bmatrix},$$

where R is a nonsingular upper triangular matrix.

To prove that this is always possible, we will use Householder transformations, defined in the previous section. Let Q_1 be a Householder transformation such that if a_1 is the first column of A, then

$$Q_1 a_1 = r_1 [1, 0, \ldots, 0]'.$$

From this, it is obtained that

$$Q_1 A = \begin{bmatrix} r_1 & s_1 \\ 0 & A_2 \end{bmatrix},$$

where A_2 is an $(n-1) \times (n-1)$ matrix and s_1 is a $1 \times (n-1)$ vector. Applying the same procedure to the matrix A_2, let Q_2 be a Householder transformation such that

$$Q_2 A_2 = \begin{bmatrix} r_2 & s_2 \\ 0 & A_3 \end{bmatrix},$$

where r_2 is a number, A_3 is an $(n-2) \times (n-2)$ matrix and s_2 is a $1 \times (n-2)$ vector. Proceeding in this way, it is clear that, after n steps we will have obtained an orthogonal matrix Q_n such that

$$Q_n A_n = \begin{bmatrix} r_n \\ 0 \end{bmatrix},$$

where r_n is a scalar and 0 is an $(m-n) \times 1$ dimensional vector. If we set

$$Q' = \begin{bmatrix} I_{n-1} & 0 \\ 0 & Q_n \end{bmatrix} \cdots \begin{bmatrix} I_1 & 0 \\ 0 & Q_2 \end{bmatrix} Q_1,$$

then,

$$Q'A = \left[\begin{array}{cccc} r_1 & * & * & * \\ 0 & r_1 & * & * \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & r_n \\ \hline 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 \end{array}\right] = \begin{bmatrix} R \\ 0 \end{bmatrix},$$

where R is a nonsingular upper triangular matrix. Thus,

$$A = Q \begin{bmatrix} R \\ 0 \end{bmatrix}$$

and that is why the previous procedure is called the QR algorithm.

We want to emphasize that we can also use other transformations instead of Householder transformations to triangularize a matrix in the previous QR algorithm. Two popular alternative choices are Givens rotations and fast Givens rotations.

2.A.3 Givens Rotations

A Givens rotation matrix is an orthogonal matrix of the form

$$Q = \begin{bmatrix} c & s \\ -s & c \end{bmatrix},$$

where $c = \cos(\theta)$ and $s = \sin(\theta)$. Clearly, this matrix corresponds to a rotation in the plane of an angle θ. To understand how these matrices can be used to make certain zeros in a given matrix, suppose that we have a two-row matrix

$$A = \begin{bmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_n \\ \beta_1 & \beta_2 & \cdots & \beta_n \end{bmatrix}$$

and we want to annihilate the $(2, 1)$ element. This can be achieved if we choose $c = \alpha_1/\sqrt{\alpha_1^2 + \beta_1^2}$ and $s = \beta_1/\sqrt{\alpha_1^2 + \beta_1^2}$, since with these values for c and s, we get

$$QA = \begin{bmatrix} \alpha_1' & \alpha_2' & \cdots & \alpha_n' \\ 0 & \beta_2' & \cdots & \beta_n' \end{bmatrix},$$

where $\alpha_1' = \sqrt{\alpha_1^2 + \beta_1^2}$ and

$$\alpha_i' = c\alpha_i + s\beta_i$$
$$\beta_i' = -s\alpha_i + c\beta_i, \quad i = 2, \ldots, n.$$

We say that we have annihilated the $(2, 1)$ element using as pivot the $(1, 1)$ element.

In order to triangularize an $m \times n$ matrix, A, with $m \geq n$ and full column rank, we can use $m \times m$ orthogonal matrices that correspond to Givens rotations of the form

$$Qij = \begin{bmatrix} 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & & \vdots & & \vdots \\ 0 & \cdots & c & \cdots & s & \cdots & 0 \\ \vdots & & \vdots & \ddots & \vdots & & \vdots \\ 0 & \cdots & -s & \cdots & c & \cdots & 0 \\ \vdots & & \vdots & & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{bmatrix},$$

where c and s appear at the intersections of the i-th and j-th rows and columns, $i < j$, $j = 2, \ldots, m$. Clearly, if we premultiply an m-dimensional vector, a, by Q_{ij}, the j-th element of a is annihilated using as pivot the i-th element.

Using a sequence of Givens rotations, Q_i, $i = 1, 2, \ldots, p$, as we used Householder transformations in the previous section to make zeros first in the first column, then in the second column, etc., it is obtained that

$$A = Q \begin{bmatrix} R \\ 0 \end{bmatrix},$$

where $Q = Q_p Q_{p-1} \cdots, Q_1$ and R is a nonsingular upper triangular matrix.

2.A.4 Fast Givens Rotations

Householder and Givens rotations use square roots. One way to avoid the calculation of square roots in Givens rotations that leads to a substantial reduction in the computational burden is as follows. Suppose that we have a two-row matrix

$$A = \begin{bmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_n \\ \beta_1 & \beta_2 & \cdots & \beta_n \end{bmatrix}$$

and we apply a Givens rotation,

$$Q = \begin{bmatrix} c & s \\ -s & c \end{bmatrix},$$

with $c = \alpha_1/\sqrt{\alpha_1^2 + \beta_1^2}$ and $s = \beta_1/\sqrt{\alpha_1^2 + \beta_1^2}$, to get

$$QA = \begin{bmatrix} \alpha_1' & \alpha_2' & \cdots & \alpha_n' \\ 0 & \beta_2' & \cdots & \beta_n' \end{bmatrix},$$

where $\alpha_1' = \sqrt{\alpha_1^2 + \beta_1^2}$ and

$$\alpha_i' = c\alpha_i + s\beta_i \tag{2A.3}$$

$$\beta_i' = -s\alpha_i + c\beta_i, \quad i = 2, \ldots, n. \tag{2A.4}$$

If we take out a positive scaling factor from each row in both matrices, A and QA, so that

$$\begin{bmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_n \\ \beta_1 & \beta_2 & \cdots & \beta_n \end{bmatrix} = \begin{bmatrix} \sqrt{\alpha} & 0 \\ 0 & \sqrt{\beta} \end{bmatrix} \begin{bmatrix} a_1 & a_2 & \cdots & a_n \\ b_1 & b_2 & \cdots & b_n \end{bmatrix}$$

and

$$\begin{bmatrix} \alpha_1' & \alpha_2' & \cdots & \alpha_n' \\ 0 & \beta_2' & \cdots & \beta_n' \end{bmatrix} = \begin{bmatrix} \sqrt{\alpha'} & 0 \\ 0 & \sqrt{\beta'} \end{bmatrix} \begin{bmatrix} a_1' & \alpha_2' & \cdots & a_n' \\ 0 & b_2' & \cdots & b_n' \end{bmatrix},$$

and we replace $\alpha_i = \sqrt{\alpha}a_i$, $\beta_i = \sqrt{\beta}b_i$, $\alpha_i' = \sqrt{\alpha'}a_i'$, $\beta_i' = \sqrt{\beta'}b_i'$, $i = 1, \ldots, n$, in (2A.3) and (2A.4), it is obtained that

$$c = \frac{\sqrt{\alpha}a_1}{\sqrt{\alpha a_1^2 + \beta b_1^2}}, \qquad s = \frac{\sqrt{\beta}b_1}{\sqrt{\alpha a_1^2 + \beta b_1^2}},$$

$$a_1' = \sqrt{(\alpha a_1^2 + \beta b_1^2)/\alpha'}$$

$$a_i' = \frac{1}{\sqrt{\alpha'}\sqrt{\alpha a_1^2 + \beta b_1^2}}[\alpha a_1 a_i + \beta b_1 b_i]$$

$$b_i' = \frac{\sqrt{\alpha\beta}}{\sqrt{\beta'}\sqrt{\alpha a_1^2 + \beta b_1^2}}[-b_1 a_i + a_1 b_i], \quad i = 2, \ldots, n.$$

If we want to avoid square roots, we should select α' and β' such that a_1', a_i', and b_i' do not require any square root calculations. This can be achieved in a number of ways. One general formulation due to Hsieh, Liu, & Yao (1993) is

$$\alpha' = \frac{\alpha a_1^2 + \beta b_1^2}{\mu^2}$$
$$\beta' = \frac{\alpha\beta}{\nu^2(\alpha a_1^2 + \beta b_1^2)},$$

where μ and ν are parameters to be determined. For example, if $\mu = 1$ and $\nu = 1$, we get the algorithm proposed by Gentleman (1973), given by

$$\begin{aligned}
\alpha' &= \alpha a_1^2 + \beta b_1^2 \\
\beta' &= \alpha\beta/\alpha' \\
a_1' &= 1 \\
a_i' &= (\alpha a_1 a_i + \beta b_1 b_i)/\alpha' \\
b_i' &= -b_1 a_i + a_1 b_i, \quad i = 2, \ldots, n.
\end{aligned}$$

If we define the generalized rotational parameters

$$\bar{c} = \alpha a_1/\alpha', \quad \bar{s} = \beta b_1/\alpha',$$

the previous recursions can be written as

$$\begin{aligned}
a_1' &= 1 \\
a_i' &= \bar{c} a_i + \bar{s} b_i \\
b_i' &= -b_1 a_i + a_1 b_i, \quad i = 2, \ldots, n.
\end{aligned}$$

If $a_1 = 1$, the recursions are simplified to

$$\begin{aligned}
\alpha' &= \alpha + \beta b_1^2, \qquad \beta' = \alpha\beta/\alpha' \\
a_1' &= 1 \\
a_i' &= a_i + \bar{s} b_i' \\
b_i' &= -b_1 a_i + b_i, \quad i = 2, \ldots, n.
\end{aligned}$$

To see this, note that

$$\begin{aligned}
a_i' &= \bar{c} a_i + \bar{s} b_i \\
&= \frac{\alpha}{\alpha'} a_i + \bar{s}(b_i' + b_1 a_i)
\end{aligned}$$

$$= \frac{\alpha + \beta b_1^2}{\alpha'} a_i + \bar{s} b_i'$$
$$= a_i + \bar{s} b_i'.$$

It is not difficult to show that the transformation $[a_1', b_1']' = T[a_1, b_1]'$ corresponding to the previous algorithm is such that

$$T = \begin{bmatrix} \bar{c} & \bar{s} \\ -b_1 & a_1 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{\alpha'} & 0 \\ 0 & 1/\sqrt{\beta'} \end{bmatrix} \Theta \begin{bmatrix} \sqrt{\alpha} & 0 \\ 0 & \sqrt{\beta} \end{bmatrix},$$

where Θ is the orthogonal matrix

$$\Theta = \begin{bmatrix} c & s \\ -s & c \end{bmatrix},$$

such that

$$c = \frac{\sqrt{\alpha}}{\sqrt{\alpha'}} a_1, \qquad s = \frac{\sqrt{\beta'}}{\sqrt{\alpha}} b_1 = \frac{\sqrt{\beta}}{\sqrt{\alpha'}} b_1.$$

2.A.5 The *QDU* Decomposition

Given an $m \times n$ matrix A with $m \geq n$ and full column rank, sometimes it is advantageous to look for an orthogonal matrix Q such that

$$Q'A = \begin{bmatrix} DU \\ 0 \end{bmatrix},$$

where U is a nonsingular upper triangular matrix with ones in the main diagonal and D is a diagonal matrix with positive elements in the main diagonal. Note that, multiplying by its transpose both matrices in the previous equality, it is obtained that

$$A'A = U'D^2U.$$

This decomposition can be achieved using fast Givens rotations, described in the previous section. To see this, suppose that we take out a positive scaling factor from each row of A, so that we can write

$$A = \Delta X$$

where Δ is diagonal matrix of order m with positive elements in the main diagonal. Then, we can apply a sequence of fast Givens rotations to the product ΔX similarly to the application of a sequence of Givens rotations to obtain the *QR* decomposition of A. In this way, we would end up with the desired *QDU* decomposition.

If the rows of X are denoted by x_i, $i = 1,\ldots,m$, and $\Delta = \text{diag}(\sqrt{d_i})$, $i = 1,\ldots,m$, the following algorithm can be used to obtain the QDU decomposition of A. When the algorithm stops, the U and D^2 matrices are stored in X and Δ, respectively. On entry, $\text{diag}(d_i)$, $i = 1,\ldots,m$, is stored in Δ.

```
for j = 1, 2, ..., n

    for i = j + 1, ..., m
        d = d_j x_j^2(j) + d_i x_i^2(j)
        s̄ = d_i x_i(j)/d
        d_i = d_i d_j/d
        if i = j + 1
            c̄ = d_j x_j(j)/d
        end
        v = x_i
        x_i = x_j(j) x_i − x_i(j) x_j
        if i = j + 1
            x_j = c̄ x_j + s̄ v
        else
            x_j = x_j + s̄ x_i
        end
        d_j = d
    end

end

if m = n

    d = x_n(n)
    d_n = d_n d d
    x_n(n) = 1

end
```

The previous algorithm is useful when we have a symmetric $n \times n$ matrix, A, such that

$$A = B\Delta B',$$

where B' is an $m \times n$ matrix with $m \geq n$ and Δ is an $m \times m$ diagonal matrix with positive elements in the main diagonal, and we want to obtain a decomposition of the form $A = U'DU$, where U is an upper triangular matrix with ones in the main diagonal and D is a diagonal matrix with positive elements in the main diagonal. Applying the previous algorithm to $\Delta^{1/2}B'$, we obtain an orthogonal matrix Q such that

$$Q'\Delta^{1/2}B' = \begin{bmatrix} D^{1/2}U \\ 0 \end{bmatrix},$$

where D and U have the required characteristics and the following unique decomposition holds

$$A = U'DU.$$

We finally mention that the previous algorithm tends to suffer from possible overflow or underflow problems. For this reason, self-scaling fast Givens transformations have been proposed by Anda & Park (1994).

Chapter 3
Stationarity and Linear Time Series Models

3.1 Stochastic Processes

In time series analysis, the first task is to select a mathematical model that suits the data adequately. Models are used, among other things, to interpret the data and to forecast future observations. Given that the phenomena that we usually observe in the form of time series data cannot be completely predictable, it is natural to suppose that the observation y_t at time t is the value taken by a certain random vector Y_t. A **time series** y is a finite set of values $\{y_1 \ldots, y_n\}$ taken by certain random vectors $\{Y_1 \ldots, Y_n\}$. The proper framework in which to study time series is that of stochastic processes.

A **stochastic process** is a family of random vectors $\{Y_t, t \in T\}$ defined on a probability space $(\Omega, \mathcal{S}, P)$, where the index set T is usually a discrete set, like the set of positive integers $T = \{1, 2, \ldots\}$ or the set of integer numbers $T = \{0, \pm 1, \pm 2, \ldots\}$, or a continuous set, like the set of nonnegative real numbers $T = [0, \infty)$ or the set of real numbers $T = (-\infty, \infty)$. Notice that Y_t is in fact a function of two arguments $Y_t(\omega)$. The first one t is the time index, whereas the second is the event $\omega \in \Omega$.

A **realization** of a stochastic process $\{Y_t, t \in T\}$ is the collection of values $\{Y_t(\omega), t \in T\}$, where $\omega \in \Omega$ is considered fixed. Thus, a time series is part of a realization of a discrete stochastic process. We usually write y_t or Y_t instead of $Y_t(\omega)$ when the context is clear.

The main difference between time series analysis and classical statistical inference is that in the latter case we have a sample of independent observations of the same variable, whereas in time series analysis the sample (the time series) is formed by observations of different variables which are usually not independent. This is

V. Gómez, *Multivariate Time Series With Linear State Space Structure*,
DOI 10.1007/978-3-319-28599-3_3

illustrated by the following table.

$$
\begin{array}{cccc}
Y_1(\omega_1) & Y_2(\omega_1) & Y_3(\omega_1) & \dots \\
Y_1(\omega_2) & Y_2(\omega_2) & Y_3(\omega_2) & \dots \\
\vdots & \vdots & \vdots & \vdots \\
Y_1(\omega_n) & Y_2(\omega_n) & Y_3(\omega_n) & \dots
\end{array}
\tag{3.1}
$$

The rows in the table represent n independent realizations of the process $\{Y_t, t = 1, 2, \dots\}$. In classical inference, we work with one column of the table, whereas a time series is part of a row of the table. We can think of the rows of the table as the output of n different computers which generate the observations at times $t = 1, 2, \dots$ according to the same random mechanism. In this context, if the time series have length M, the ith time series $\{Y_1(\omega_i), \dots, Y_M(\omega_i)\}$ would be part of the output $\{Y_1(\omega_i), Y_2(\omega_i), \dots\}$ of the ith computer, $i = 1, \dots, n$.

In practice, it is not possible to observe several independent realizations and we only have at our disposal part of one realization, which is the time series. For this reason, to simplify the notation, we simply write $\{y_1, \dots, y_M\}$ instead of $\{Y_1(\omega), \dots, Y_M(\omega)\}$

3.2 Stationary Time Series

Given a time series $\{y_1, \dots, y_n\}$, it is useful to compute the covariance matrix of the random variables $\{Y_1, \dots, Y_n\}$ to gain some insight into the dependence between them.

The **autocovariance function of a stochastic process** $\{Y_t, t \in T\}$ such that all components of Y_t have finite variance for each $t \in T$ is defined by

$$\gamma_Y(r, s) = \text{Cov}(Y_r, Y_s) = E[(Y_r - EY_r)(Y_s - EY_s)'], \qquad r, s, \in T. \tag{3.2}$$

Definition 3.1 (Stationarity) The stochastic process $\{Y_t, t \in \mathbb{Z}\}$, where $\mathbb{Z} = \{0, \pm 1, \pm 2, \dots\}$, is said to be stationary if

i) $EY_t = m$ for all $t \in \mathbb{Z}$,
ii) $\text{Var}(Y_t) = \Sigma$ for all $t \in \mathbb{Z}$, where Σ is a nonnegative-definite matrix,
iii) $\gamma_Y(r, s) = \gamma_Y(r + t, s + t)$ for all $t \in \mathbb{Z}$.

Remark 3.1 Stationarity as just defined is frequently referred to in the literature as weak stationarity, covariance stationarity, stationarity in the wide sense, or second order stationarity. Unless otherwise stated, the term stationarity will always refer to the previous definition in what follows. ◇

Remark 3.2 If $\{Y_t, t \in \mathbb{Z}\}$ is stationary, then $\gamma_Y(r, s) = \gamma_Y(r - s, 0)$ for all $r, s \in \mathbb{Z}$. It is therefore convenient to redefine the autocovariance function of a stationary process as the function of just one variable

$$\gamma_Y(h) = \gamma_Y(h, 0) = \mathrm{Cov}(Y_{t+h}, Y_t) \text{ for all } t, h \in \mathbb{Z}.$$

The function γ_Y will be referred to as the autocovariance function of the process $\{Y_t\}$ and $\gamma_Y(h)$ as its value at "lag" h. ◇

Many of the stationary processes that we will consider in the sequel will be defined in terms of the following simple process.

Definition 3.2 (White Noise) The stochastic process $\{A_t, t \in \mathbb{Z}\}$ is said to be white noise if

i) $E(A_t) = 0$ for all $t \in \mathbb{Z}$,
ii) $\mathrm{Cov}(A_t, A_s) = 0$ for all $t, s \in \mathbb{Z}, t \neq s$,
iii) $E(A_t A_t') = \Sigma$ for all $t \in \mathbb{Z}$, where Σ is a positive-definite matrix

A white noise process is usually denoted by $\{A_t\} \sim \mathrm{WN}(0, \Sigma)$.

It is clear from the definition that a white noise $\mathrm{WN}(0, \Sigma)$ is a stationary process with zero mean and covariance matrix Σ.

Another important and frequently used notion of stationarity is introduced in the following definition.

Definition 3.3 (Strict Stationarity) The stochastic process $\{Y_t, t \in \mathbb{Z}\}$ is said to be strictly stationary if the joint distribution of $(Y_{t_1}', \ldots, Y_{t_k}')'$ and $(Y_{t_1+h}', \ldots, Y_{t_k+h}')'$ are the same for all positive integers k and for all $t_1, \ldots, t_k, h \in \mathbb{Z}$.

Strict stationarity means intuitively that the graphs over two equal-length time intervals of a realization of the time series should exhibit similar statistical characteristics. For example, the proportion of ordinates not exceeding a given level z should be roughly the same for both intervals.

If $\{Y_t\}$ is strictly stationary it immediately follows, on taking $k = 1$ in the previous definition, that Y_t has the same distribution for each $t \in \mathbb{Z}$. If the covariance matrix of Y_t is finite, this implies in particular that EY_t and $\mathrm{Var}(Y_t)$ are both constant. Moreover, taking $k = 2$ in the previous definition, we find that Y_{t+h} and Y_t have the same joint distribution and hence the same covariance for all $h \in \mathbb{Z}$. Thus a strictly stationary process with finite second moments is stationary.

The converse of the previous statement is not true. For example, if $\{Y_t\}$ is a sequence of independent random variables such that Y_t is exponentially distributed with mean one when t is odd and normally distributed with mean one and variance one when t is even, then $\{Y_t\}$ is stationary with $\gamma_Y(0) = 1$ and $\gamma_Y(h) = 0$ for $h \neq 0$. However since Y_1 and Y_2 have different distributions, $\{Y_t\}$ cannot be strictly stationary.

In the important case of normality, however, stationarity implies strict stationarity. This is so because the distribution of a normal variable is completely determined by its first two moments.

Example 3.1 Let $\{A_t : t = 0, 1, \ldots\} \sim \text{WN}(0, \sigma^2)$. Define the process $\{Y_t : t = 1, 2, \ldots\}$ by $Y_t = A_t + \theta A_{t-1}$. Then the autocovariance function of Y_t is given by

$$\begin{aligned}\text{Cov}(Y_{t+h}, Y_t) &= \text{Cov}(A_{t+h} + \theta A_{t+h-1}, A_t + \theta A_{t-1}) \\ &= \begin{cases} (1+\theta^2)\sigma^2 & \text{if } h = 0 \\ \theta\sigma^2 & \text{if } h = \pm 1 \\ 0 & \text{if } |h| > 1, \end{cases}\end{aligned}$$

and hence $\{Y_t\}$ is stationary. In fact it can be shown that $\{Y_t\}$ is strictly stationary. ◇

Example 3.2 Let $\{A_t : t = 1, 2, \ldots\} \sim \text{WN}(0, \sigma^2)$. Define the process $\{Y_t : t = 1, 2, \ldots\}$ by $Y_t = A_1 + A_2 + \cdots + A_t$. Then $\text{Cov}(Y_{t+h}, Y_t) = \sigma^2 t$ if $h > 0$ and thus Y_t is not stationary. ◇

3.2.1 Ergodicity

When analyzing a stationary process $\{Y_t\}$, one is often interested in estimating the first two moments of the distribution of the variables Y_t. Because each column $(Y_t(\omega_1), Y_t(\omega_2), \ldots, Y_t(\omega_n))'$ in table (3.1) constitutes a random sample for the variable Y_t, one can use the sample mean $\bar{y}_t = \sum_{i=1}^{n} Y_t(\omega_i)/n = \sum_{i=1}^{n} y_t^i/n$ to estimate EY_t consistently.

In a similar way, we can consider two columns in table (3.1), $(Y_t(\omega_1), Y_t(\omega_2), \ldots, Y_t(\omega_n))'$ and $(Y_{t-j}(\omega_1), Y_{t-j}(\omega_2), \ldots, Y_{t-j}(\omega_n))'$. Then $\sum_{i=1}^{n}(y_t^i - \bar{y}_t)(y_{t-j}^i - \bar{y}_t)'/n$ is a consistent estimator of $\text{Cov}(Y_t, Y_{t-j})$.

In practice we only have at our disposal one realization and we cannot use the sample mean $\bar{y}_t = \sum_{i=1}^{n} y_t^i/n$, that is an ensemble average, to estimate EY_t. We consider instead a time average $\bar{z} = \sum_{t=1}^{M} Y_t(\omega_1)/M = \sum_{t=1}^{M} y_t/M$. A stationary process $\{Y_t\}$ is said to be **ergodic for the mean** if $\bar{z}$ converges in probability to EY_t when $M \to \infty$.

It can be shown that a stationary process $\{Y_t\}$ is ergodic for the mean if the autocovariance matrices $\gamma(j)$ tend to zero quickly enough when j increases, that is if $\gamma(j) \to 0$ when $j \to \infty$. A sufficient condition for that is $\sum_{j=0}^{\infty} |\gamma_{lm}(j)| < \infty$ for all $l, m = 1, \ldots, k$, where k is the dimension of Y_t and $\gamma(j) = [\gamma_{lm}(j)]_{l,m=1}^{k}$.

A stationary process $\{Y_t\}$ is said to be **ergodic for second moments** if $\sum_{i=1}^{M}(y_t - m)(y_{t-j} - m)'/M$ converges in probability to $\gamma(j)$ for all j when $M \to \infty$, where $m = EY_t$.

The processes usually encountered in practice are both stationary and ergodic and we will not have to worry about ergodicity in the rest of the book. To clarify the concepts of stationarity and ergodicity, however, we present an example of a process that is stationary but not ergodic.

Example 3.3 Define $Y_t = X + A_t$, where $\{A_t : t = 1, 2, \ldots\}$ is a sequence of independent $N(0, \sigma^2)$ variables and X is an $N(0, \lambda^2)$ variable that is independent of the A_t. Then, $EY_t = E(X + A_t) = 0$, $\text{Var}(Y_t) = E(X + A_t)^2 = \lambda^2 + \sigma^2$ and $\text{Cov}(Y_t, Y_{t-j}) = E(X + A_t)(X + A_{t-j}) = \lambda^2$ if $j \neq 0$. Thus, $\{Y_t\}$ is stationary. However, $\sum_{t=1}^{M} y_t/M = \sum_{t=1}^{M} Y_t(\omega)/M = \sum_{t=1}^{M}(x+a_t)/M = x+(\sum_{t=1}^{M} a_t/M) \rightarrow x$. ◇

The definitions of ergodicity given previously in this section are for wide sense stationary processes. A strict sense stationary vector process $\{Y_t\}$ is **ergodic** if the so-called **shift transformation**

$$T : Y_t \rightarrow Y_{t+1}, \qquad t \in \mathbb{Z},$$

is such that every Borel set, S, in the sigma-algebra generated by the cylinder sets, for which $S = T(S)$ almost surely, has probability zero or one.

Theorem 3.1 (Ergodic Theorem) *Let $\{Y_t\}$ be a vector process that is strict sense stationary and ergodic with $E(Y_t) = \mu$. Then,*

$$\frac{1}{n}\sum_{t=1}^{n} Y_t \xrightarrow{a.s.} \mu.$$

Proof See, for example, Rozanov (1967). □

The ergodic theorem constitutes a substantial generalization of Kolmogorov's Law of Large Numbers. In the ergodic theorem, serial dependence is allowed, albeit one that vanishes in the long term. However, the i.i.d. hypothesis in Kolmogorov's Law of Large Numbers excludes any form of serial dependence.

Remark 3.3 The property of ergodicity is not observationally verifiable from a single history. An example is given in (Hannan, 1970, p. 201). For this reason, ergodicity is just assumed when needed. ◇

3.2.2 The Autocovariance and Autocorrelation Functions and Their Properties

The autocovariance function $\{\gamma_Y(h) : h \in \mathbb{Z}\}$ of a stationary process $\{Y_t : t \in \mathbb{Z}\}$ was defined in Remark 3.2. The **autocorrelation function** of a scalar stationary process $\{Y_t : t \in \mathbb{Z}\}$ is defined as the function whose value at lag h is

$$\rho_Y(h) = \gamma_Y(h)/\gamma_Y(0) = \text{Corr}(Y_{t+h}, Y_t) \text{ for all } t, h \in \mathbb{Z}.$$

If $\{Y_t\} = \{(Y'_{t1}, \ldots, Y'_{tk})'\}$ is a vector stationary process, then the autocorrelation function is defined as $\rho(h) = [\rho_{ij}(h)]_{i,j=1}^{k}$, where $\rho_{ij}(h)$ is the autocorrelation function of Y_{ti} and Y_{tj}, $\rho_{ij}(h) = \gamma_{ij}(h)/\sqrt{\gamma_{ii}(0)\gamma_{jj}(0)}$. Note that $\rho(h) = V\gamma(h)V$, where $V = \text{diag}(\gamma_{11}(0)^{-1/2}, \ldots, \gamma_{kk}(0)^{-1/2})$.

Remark 3.4 It will be noticed that we have defined stationarity only in the case when $T = \mathbb{Z}$. It is not difficult to define stationarity using a more general index set, but for our purposes this will not be necessary. If we wish to model a set of data $\{y_t, t \in T \subset \mathbb{Z}\}$ as a realization of a stationary process, we can always consider it to be part of a realization of a stationary process $\{Y_t\}$. ◇

In this section we will see some properties of the autocovariance function of a stationary process.

Proposition 3.1 *The autocovariance function $\gamma(h)$ of a scalar stationary process $\{Y_t : t \in \mathbb{Z}\}$ satisfies*

$$\begin{aligned} \gamma(0) &\geq 0 \\ |\gamma(h)| &\leq \gamma(0) \quad \textit{for all } h \in \mathbb{Z} \\ \gamma(h) &= \gamma(-h) \; \textit{for all } h \in \mathbb{Z} \end{aligned}$$

Proof The first property follows from the fact that $\text{Var}(Y_t) \geq 0$. The second is an immediate consequence of the Cauchy–Schwarz inequality

$$|\text{Cov}(Y_{t+h}, Y_t)| \leq \sqrt{\text{Var}(Y_{t+h})\text{Var}(Y_t)}.$$

The third follows by observing that

$$\gamma(-h) = \text{Cov}(Y_{t-h}, Y_t) = \text{Cov}(Y_t, Y_{t+h}) = \gamma(h).$$

The fourth is a statement of the fact that $\text{E}(\sum_{j=1}^{n} a_j \left(Y_j - \mu\right))^2 \geq 0$. □

Proposition 3.2 *The autocovariance function $\gamma(h)$ of a scalar stationary process $\{Y_t : t \in \mathbb{Z}\}$ is nonnegative definite, that is,*

$$\sum_{i,j=1}^{n} a_i a_j \gamma(i-j) \geq 0 \quad \textit{for all } n \in \{1, 2, \ldots\} \textit{ and } (a_1, \ldots, a_n)' \in \mathbb{R}^n.$$

Proof This is simply a statement of the fact that $\text{Var}(\sum_{j=1}^{n} a_j Y_j)^2 \geq 0$. □

Remark 3.5 According to Herglotz's Theorem (Brockwell & Davis, 1991, pp. 117–119), a complex valued function $\gamma(\cdot)$ defined on the integers is nonnegative definite if, and only if,

$$\gamma(h) = \int_{(-\pi,\pi]} e^{ihx} dF(x),$$

where $F(\cdot)$ is a right-continuous, non-decreasing, bounded function on $[-\pi, \pi]$ and $F(-\pi) = 0$. (The function F is called the **spectral distribution function** of γ) ◇

Example 3.4 Define the process $\{Y_t : t = -1, 0, 1, 2, \dots\}$ by $Y_t + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} = A_t$, where $\{A_t : t = 1, 2, \dots\} \sim \text{WN}(0, \sigma^2)$ and the roots of the polynomial $1 + \phi_1 z + \phi_2 z^2$ are all outside the unit circle. Assuming that the process is stationary and A_t is uncorrelated with Y_s for $s < t$, we get

$$\begin{aligned} \gamma_Y(0) + \phi_1 \gamma_Y(1) + \phi_2 \gamma_Y(2) &= \sigma^2 \\ \gamma_Y(l) + \phi_1 \gamma_Y(l-1) + \phi_2 \gamma_Y(l-2) &= 0, \quad l > 0. \end{aligned} \tag{3.3}$$

Thus, if the initial conditions, $\{Y_{-1}, Y_0\}$, are chosen such that they are uncorrelated with $\{A_t : t \geq 1\}$, have zero mean and covariance matrix

$$E\left\{ \begin{bmatrix} Y_{-1} \\ Y_0 \end{bmatrix} [Y_{-1}, Y_0] \right\} = \begin{bmatrix} \gamma_Y(0) & \gamma_Y(1) \\ \gamma_Y(1) & \gamma_Y(0) \end{bmatrix},$$

where $\gamma_Y(0)$ and $\gamma_Y(1)$ satisfy (3.3), then the process $\{Y_t\}$ is stationary. Notice that the covariances can be obtained recursively for $l \geq 2$ and that we can solve the following system of equations

$$\begin{bmatrix} 1 + \phi_2^2 & \phi_1(1 + \phi_2) \\ 1 + \phi_2 & \phi_1 \end{bmatrix} \begin{bmatrix} \gamma_Y(0) \\ \gamma_Y(1) \end{bmatrix} = \begin{bmatrix} \sigma^2 \\ 0 \end{bmatrix}$$

to compute $\gamma_Y(0)$ and $\gamma_Y(1)$ in terms of ϕ_1, ϕ_2 and σ^2.

The previous system is obtained by replacing $\gamma_Y(2)$ in $\gamma_Y(0) + \phi_1 \gamma_Y(1) + \phi_2 \gamma_Y(2) = \sigma^2$ by $\gamma_Y(2) = -\phi_1 \gamma_Y(1) - \phi_2 \gamma_Y(0)$ and using the resulting equation plus the equation $\gamma_Y(1) + \phi_1 \gamma_Y(0) + \phi_2 \gamma_Y(1) = 0$.

The autocorrelation function is the solution of the following linear difference equation of order two

$$\rho_Y(l) + \phi_1 \rho_Y(l-1) + \phi_2 \rho_Y(l-2) = 0, \qquad l \geq 2.$$

According to the results of Sect. 3.13, the pattern followed by the autocorrelation function depends on the roots of the polynomial $1 + \phi_1 z + \phi_2 z^2$. For example, if $1 + \phi_1 z + \phi_2 z^2 = (1 - \lambda z)(1 - \bar{\lambda} z)$, where $\lambda = re^{i\theta}$ and $\bar{\lambda} = re^{-i\theta}$, the solution has the form

$$\rho_Y(h) = a_1 r^h \cos(h\theta) + a_2 r^h \sin(h\theta), \qquad h \geq 2,$$

where a_1 and a_2 are determined by the initial conditions, $\rho_Y(0) = 1$ and $\rho_Y(1)$. ◇

For vector stationary processes, we can mention the following properties.

Proposition 3.3 *The autocovariance function* $\gamma(h)$ *of a vector stationary process* $\{Y_t : t \in \mathbb{Z}\}$, $\{Y_t\} = \{(Y'_{t1}, \ldots, Y'_{tk})'\}$, *satisfies*

i) $\gamma(0)$ *is a nonnegative-definite matrix*
ii) $|\gamma_{ij}(h)| \leq \sqrt{\gamma_{ii}(0)\gamma_{jj}(0)}$ *for all i,j=1,...,k*
iii) $\gamma(h) = \gamma'(-h)$
iv) $\gamma_{ii}(h)$ *is a scalar autocovariance function,* $i = 1, \ldots, k$
v) $\sum_{i,j=1}^{n} a_i'\gamma(i-j)a_j \geq 0$ *for all* $n \in \{1, 2, \ldots\}$ *and* $a_1, \ldots, a_n \in \mathbb{R}^k$.

Proof We will only prove i) and v). The rest will be left as an exercise. Let $a = (a_1, \ldots, a_k)'$ be an arbitrary real vector. Then, $\text{Var}(a'Y_t) = a'\gamma(0)a \geq 0$. In a similar way, v) is a statement of the fact that $\text{E}(\sum_{j=1}^{n} a_j' (Y_j - \mu))^2 \geq 0$, where $\text{E}(Y_t) = \mu$. □

3.3 Linear Time Invariant Filters

Many stochastic processes found in practice can be rationalized as the output of a linear filter applied to some, often simpler, stochastic process. This is, for example, the case of VARMA models, that will be considered later in this chapter.

Definition 3.4 A **linear time invariant filter** relating an r-dimensional input process $\{X_t\}$ to a k-dimensional output process $\{Y_t\}$ is given by the expression

$$Y_t = \sum_{i=-\infty}^{\infty} \Psi_i X_{t-i}, \tag{3.4}$$

where the Ψ_j are $k \times r$ matrices. The function $\Psi(z) = \sum_{i=-\infty}^{\infty} \Psi_i z^i$ is called the **transfer function** of the filter. The filter is **symmetric** if $\Psi_j = \Psi_{-j}$ for $j \neq 0$. The filter is **physically realizable** or **causal** when $\Psi_j = 0$ for $j < 0$, so that $Y_t = \sum_{i=0}^{\infty} \Psi_i X_{t-i}$. The filter is said to be **stable** if $\sum_{i=-\infty}^{\infty} \|\Psi_i\| < \infty$, where $\|A\|$ denotes a norm for the matrix A such as the Frobenius norm, $\|A\| = \sqrt{\text{tr}(A'A)}$.

Remark 3.6 For the process $\{Y_t\}$ in (3.4) to be well defined the sum must converge. The convergence, however, can be of several types. For example, it can converge almost surely, in probability, in mean square, etc. We will give criteria for some types of convergence later in this chapter.

Remark 3.7 Given a matrix A of dimension $m \times n$, by the singular value decomposition, there exist orthogonal matrices U and V of dimensions $m \times m$ and $n \times n$ such that $U'AV = \text{diag}(\sigma_1, \ldots, \sigma_p)$, where $p = \min\{m, n\}$ and $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_p \geq 0$ are the singular values. This implies that the Frobenius norm satisfies $\|A\|^2 = \sigma_1^2 + \cdots + \sigma_p^2 = \|A'\|$. It also satisfies the submultiplicative property, $\|AB\| \leq \|A\|\|B\|$.

Example 3.5 The filter $Y_t = aX_{-t-1}, t = 0, \pm 1, \pm 2, \ldots$, is not linear time invariant because the weights, Ψ_j, in $Y_t = \sum_{i=-\infty}^{\infty} \Psi_i X_{t-i}$ are not time invariant. ◇

Example 3.6 A univariate causal filter is called a **unilateral ARMA filter** if its transfer function is given by a rational function

$$\Psi(z) = \frac{a_0 + a_1 z + \cdots + a_p z^p}{1 + b_1 z + \cdots + b_q z^q},$$

where the roots $1 + b_1 z + \cdots + b_q z^q$ are all outside of the unit circle. These filters are causal and stable but not symmetric. ◇

Example 3.7 A univariate filter is called a **bilateral ARMA filter** if its transfer function is given by

$$\Psi(z) = k\frac{a(z)a(z^{-1})}{b(z)b(z^{-1})},$$

where $k > 0$, $a(z) = 1 + a_1 z + \cdots + a_p z^p$, $b(z) = 1 + b_1 z + \cdots + b_q z^q$ and the roots of $b(z)$ are all outside of the unit circle. These filters are stable and symmetric but not causal. ◇

Example 3.8 The filter $Y_t = X_{t-1}/2 + X_t + X_{t+1}$ is not physically realizable. ◇

Example 3.9 The filter $Y_t = \sum_{i=0}^{\infty} \Psi_i X_{t-i}$, where $\Psi_j = 1.5^j$, is causal but not stable. ◇

3.4 Frequency Domain

In the **frequency domain** approach to time series analysis, each process $\{Y_t\}$, scalar or multivariate, is considered as the sum of certain oscillatory stochastic components associated with some frequency band in the interval $[-\pi, \pi]$. For example, it is customary to consider that an economic series is the sum of a **trend**, a **seasonal**, and an **irregular** component. The trend is associated with a band of low frequencies of the form $[0, \omega]$, the seasonal component is associated with the **seasonal frequencies**, defined as $2\pi k/s$, $k = 1, 2, \ldots, [s/2]$, where s is the number of seasons and $[s/2]$ denotes the integer part of $s/2$, and the irregular component is associated with a band of high frequencies of the form $[\omega, \pi]$.

To gain some insight into the properties of stationary processes in the frequency domain, consider first a deterministic function of the form

$$f(t) = A\cos(\omega t + \varphi).$$

This function is periodic with **period** $\tau = 2\pi/\omega$. That is, $f(t + \tau) = f(t)$. Also, A is the **amplitude**, ω is the **frequency**, and φ is the **phase**. The function $f(t)$ can alternatively be expressed as

$$f(t) = a\cos(\omega t) + b\sin(\omega t).$$

To see this, use the formula $\cos(a+b) = \cos(a)\cos(b) - \sin(a)\sin(b)$ in $f(t)$ to get

$$\begin{aligned} f(t) &= A\cos(\omega t + \varphi) \\ &= A\left[\cos(\omega t)\cos(\varphi) - \sin(\omega t)\sin(\varphi)\right] \\ &= A\cos(\varphi)\cos(\omega t) - A\sin(\varphi)\sin(\omega t) \\ &= a\cos(\omega t) + b\sin(\omega t), \end{aligned}$$

where $a = A\cos(\varphi)$ and $b = -A\sin(\varphi)$.

It is well known in Mathematical Analysis that every well-behaved real function can be developed as a convergent series of functions each one having the form of $f(t)$. In fact, this result, known as Fourier series expansion, is at the center of the part of Mathematical Analysis known as Harmonic Analysis.

Instead of using functions like $f(t)$, it is often convenient to use complex-valued functions of the form

$$g(t) = Ae^{it\omega},$$

where A is a complex number, $A = a + bi$. Substituting in $g(t)$, we get

$$\begin{aligned} g(t) &= Ae^{it\omega} \\ &= (a+bi)(\cos(\omega t) + i\sin(\omega t)) \\ &= \left[a\cos(\omega t) - b\sin(\omega t)\right] + i\left[a\sin(\omega t) + b\cos(\omega t)\right], \end{aligned}$$

and it is seen that $f(t)$ can be considered as the real part of a certain function having the same form as $g(t)$.

Suppose now that we have two functions, $g_1(t) = A_1e^{it\omega_1}$ and $g_2(t) = A_2e^{it\omega_2}$, such that A_1 and A_2 are zero mean, uncorrelated, complex-valued random variables with $\mathrm{Var}(A_k) = \sigma_k^2$, $k = 1, 2$. Here, it is understood that we extend the definition of mean and scalar product of real-valued random variables to complex-valued random variables in the following way. If $A_k = a_k + b_k i$, $k = 1, 2$, then $E(A_k) = E(a_k) + iE(b_k)$ and $< A_k, A_j >= E(A_k\overline{A_j})$, $k, j = 1, 2$, where the bar denotes complex conjugation. With these definitions, the process $Y_t = g_1(t) + g_2(t)$ is stationary. This can be seen as follows. Clearly, the mean of $g(t)$ is zero. In addition,

$$\begin{aligned} \mathrm{Cov}(Y_t, Y_{t-h}) &= E\left[Y_t\overline{Y}_{t-h}\right] \\ &= \sum_{k=1}^{2}\sigma_k^2 e^{it\omega_k}e^{-i(t-h)\omega_k} \\ &= \sum_{k=1}^{2}\sigma_k^2 e^{ih\omega_k}, \end{aligned}$$

and the covariance depends on h only. We can generalize this to a finite number of components. Let

$$Y_t = \sum_{k=1}^{n} A_k e^{it\omega_k}, \tag{3.5}$$

where the A_k, $k = 1, \ldots, n$, are zero mean uncorrelated complex-valued random variables with $E(A_k \overline{A_k}) = \sigma_k^2$ and $-\pi < \omega_1 < \omega_2 < \cdots < \omega_n \leq \pi$. In this case, it is not difficult to verify that $E(Y_t)= 0$ and $\mathrm{Cov}(Y_t, Y_{t-h}) = \sum_{k=1}^{n} \sigma_k^2 e^{ih\omega_k}$ and, therefore, the process is stationary. If we define the distribution function

$$F(x) = \sum_{\{k:\, \omega_k \leq x\}} \sigma_k^2, \qquad x \in (-\pi, \pi],$$

we can express the covariances, $\gamma(h)$, as the Stieltjes integral

$$\gamma(h) = \int_{(-\pi,\pi]} e^{ihx} dF(x).$$

It is to be noticed that, since $\gamma(0) = \sum_{k=1}^{n} \sigma_k^2$, each frequency ω_k contributes with σ_k^2 to the variance. The function $F(x)$ is known as the **spectral distribution function** of $\{Y_t\}$.

It is remarkable that every zero-mean stationary process can be interpreted as the limit in mean square of processes of the form (3.5). In fact, it can be shown (Brockwell & Davis, 1991, pp. 145–146) that, for every scalar stationary process $\{Y_t\}$, the so-called **spectral representation of the process**

$$Y_t = \int_{-\pi}^{\pi} e^{itx} dZ(x) \tag{3.6}$$

holds, where $\{Z(x) : x \in [-\pi, \pi]\}$ is a stochastic process with independent increments, continuous on the right. In addition, the autocovariance function, $\gamma(h)$, can be expressed as

$$\gamma(h) = \int_{(-\pi,\pi]} e^{ihx} dF(x), \tag{3.7}$$

where F is a distribution function with $F(-\pi) = 0$, $F(\pi) = E(Y_t Y_t')$ and

$$F(x_2) - F(x_1) = E|Z(x_2) - Z(x_1)|^2, \qquad -\pi \leq x_1 < x_2 \leq \pi.$$

Equation (3.7) is called the **spectral representation of the autocovariance function**. For example, for the process (3.5), the process $\{Z(x) : x \in [-\pi, \pi]\}$ in (3.6) is

$$Z(x) = \sum_{\{k:\, \omega_k \leq x\}} A_k, \qquad x \in [-\pi, \pi],$$

and satisfies

$$F(x_2) - F(x_1) = \sum_{\{k:\, x_1 < \omega_k \leq x_2\}} \sigma_k^2, \qquad -\pi \leq x_1 < x_2 \leq \pi.$$

It is also shown in (Brockwell & Davis, 1991, p. 144), that the so-called **Kolmogorov's isomorphism theorem** holds. Let H be the Hilbert space spanned by $\{Y_t\}$ and let $L^2(F)$ be the Hilbert space of square integrable complex functions with respect to the measure induced by the spectral distribution function F. Then, the theorem states that there is a unique Hilbert space isomorphism, T, of H onto $L^2(F)$ such that

$$T(Y_t) = e^{it\cdot}.$$

This isomorphism establishes the link between the time domain and the frequency domain. So, for example,

$$\begin{aligned}\langle T\,(a_1 Y_{t_1} + a_2 Y_{t_2})\,, T\,(bY_s)\rangle_{L^2(F)} &= \sum_{i=1}^{2} a_i \overline{b} \left\langle e^{it_i\cdot}, e^{is\cdot}\right\rangle_{L^2(F)} \\ &= \sum_{i=1}^{2} a_i \overline{b} \int_{(-\pi,\pi]} e^{i(t_i - s)x} dF(x) \\ &= \sum_{i=1}^{2} a_i \overline{b}\, \langle Y_{t_i}, Y_s\rangle \\ &= \langle a_1 Y_{t_1} + a_2 Y_{t_2}, bY_s\rangle\,,\end{aligned}$$

where, as usual, $\langle Y_{t_i}, Y_s\rangle = E\left[Y_{t_i}\overline{Y}_s\right]$ and the bar denotes complex conjugation.

The integral (3.6) is a stochastic integral with respect to an orthogonal-increment process. Heuristically, this integral can be interpreted as the limit in mean square of finite sums of complex exponentials with stochastic coefficients.

To see the effect of a linear filter $\Psi(z) = \sum_{i=-\infty}^{\infty} \Psi_i z^i$ on these components, let us suppose that the Ψ_j weights are scalar and the elementary complex exponential function e^{itx} is passed through the filter. Then, the output $\{Y_t\}$ of the filter is

$$Y_t = \sum_{j=-\infty}^{\infty} \Psi_j e^{i(t-j)x} = \left(\sum_{j=-\infty}^{\infty} \Psi_j e^{-ijx}\right) e^{itx}.$$

Thus, the effects of the filter are summarized by the complex-valued function

$$\hat{\Psi}(x) = \sum_{j=-\infty}^{\infty} \Psi_j e^{-ijx},$$

which is the Fourier transform of the sequence $\{\ldots, \Psi_{-1}, \Psi_0, \Psi_1, \ldots\}$.

The spectral representation theorem and Kolmogorov's isomorphism are also valid for multivariate series. See Sect. 11.8 in Brockwell & Davis (1991).

Definition 3.5 Given a linear time invariant filter with transfer function $\Psi(z) = \sum_{i=-\infty}^{\infty} \Psi_i z^i$, the function

$$\hat{\Psi}(x) = \Psi(e^{-ix}) = \sum_{j=-\infty}^{\infty} \Psi_j e^{-ijx},$$

is called the **frequency response function** of the filter. For a univariate filter, the **gain**, $G(x)$, and **phase**, $\phi(x)$, functions of the filter are defined as

$$G(x) = |\hat{\Psi}(x)|, \qquad \phi(x) = \arg[\hat{\Psi}(x)], \qquad x \in [-\pi, \pi].$$

If a scalar or multivariate stationary process $\{Y_t\}$ is passed through a stable filter $\Psi(z) = \sum_{i=-\infty}^{\infty} \Psi_i z^i$ and the spectral representation of $\{Y_t\}$ is given by (3.6), it can be shown (Brockwell & Davis, 1991) that the spectral representation of the transformed process $\{X_t\}$ is

$$X_t = \int_{-\pi}^{\pi} \hat{\Psi}(x) e^{itx} dZ(x).$$

The previous equation shows intuitively the effect of the filter $\Psi(z)$ on a scalar input process $\{Y_t\}$. The effect is twofold on the sinusoids at frequency $x \in [-\pi, \pi]$. First, the amplitudes are multiplied by the modulus $G(x) = |\hat{\Psi}(x)|$ of the complex number $\hat{\Psi}(x)$ and second, there is a shift effect measured by the argument $\phi(x)$ of $\hat{\Psi}(x)$. Note that if the filter is symmetric, there is no phase effect because the number $\hat{\Psi}(x)$ is a real number. Thus, a unilateral ARMA filter has a shift effect but a bilateral ARMA filter has not.

Example 3.10 (Low Pass Filters) In electrical engineering, a low pass filter is a unilateral filter that passes signals with a frequency lower than a certain cutoff frequency and attenuates signals with frequencies higher than the cutoff frequency. An important example of low pass filters are **Butterworth filters**, first proposed by Butterworth (1930). These filters are usually referred to as **maximally flat magnitude filters**. The squared gain of Butterworth filters is given by

$$|G(x)|^2 = \frac{1}{1 + \left(\frac{\sin(x/2)}{\sin(x_c/2)}\right)^{2d}},$$

and $d = 1, 2, \ldots$, or by the same function but replacing sin with tan.

When one is working with economic series, low pass filters are appropriate to estimate trends, for example. However, in economics one is often more interested in bilateral than in unilateral filters. The gain function of bilateral Butterworth filters based on the sin function is given by

$$G(x) = \frac{1}{1 + \left(\frac{\sin(x/2)}{\sin(x_c/2)}\right)^{2d}},$$

and $d = 1, 2, \ldots$. The transfer function of these filters is

$$\Psi(z) = \frac{1}{1 + \lambda[(1-z)(1-z^{-1})]^d},$$

where $\lambda = [2\sin(x_c/2)]^{-2d}$ and thus the filter depends on two parameters only, d and λ. Increasing d, the filter gain becomes flatter in the origin and its fall towards zero gets steeper. Increasing λ, the pass band, which is the part of the gain function different from zero, gets narrower and the output gets less volatile because only very low frequency components can pass through the filter. More details about bilateral Butterworth filters are given in Gómez (2001) and Gómez & Maravall (2001b). ◇

In the following, we will study some properties of linear time invariant filters.

Proposition 3.4 *Let* $\{X_t = (X_{t1}, \ldots, X_{tk})' : t \in \mathbb{Z}\}$ *be a sequence of random vectors such that* $\sup\{E|X_{ti}|^2 : t \in \mathbb{Z}, i = 1, \ldots, k\} < \infty$ *and* $\sum_{i=-\infty}^{\infty} \|\Psi_i\| < \infty$*. Then, the series*

$$\sum_{i=-\infty}^{\infty} \Psi_i X_{t-i}$$

converges in mean square, that is $E[(Y_t - \sum_{i=-m}^{n} \Psi_i X_{t-i})(Y_t - \sum_{i=-m}^{n} \Psi_i X_{t-i})'] \to 0$ *as* $m, n \to \infty$.

Proof Let $N = \sup\{E|X_{ti}|^2 : t \in \mathbb{Z}, i = 1, \ldots, k\} < \infty$ and $n > m > 0$. Then, we can write

$$\begin{aligned}
\left\| E\Big(\sum_{m<|j|\le n} \Psi_j X_{t-j}\Big)\Big(\sum_{m<|p|\le n} \Psi_p X_{t-p}\Big)' \right\|^2 &\le \sum_{m<|j|\le n}\sum_{m<|p|\le n} \|\Psi_j E(X_{t-j}X'_{t-p})\Psi'_p\|^2 \\
&\le \left(\sum_{m<|j|\le n} \|\Psi_j\|^2\right)\left(\sum_{m<|p|\le n} \|\Psi_p\|^2\right) Nkr \\
&\to 0 \quad \text{as} \quad m, n \to \infty.
\end{aligned}$$

By the Cauchy criterion, the series $\sum_{i=-\infty}^{\infty} \Psi_i X_{t-i}$ converges in mean square. □

Corollary 3.1 *Under the assumptions of Proposition 3.4, if X_t is a stationary process with covariance matrices $\gamma_X(h)$, then the output, $Y_t = \sum_{i=-\infty}^{\infty} \Psi_i X_{t-i}$, is a stationary process with covariance matrices*

$$\gamma_Y(h) = \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} \Psi_i \gamma_X(h-i+j)\Psi_j'.$$

Proof By the mean square convergence and continuity of the inner product $< X, Y >= E(XY')$, we can write

$$< Y_t, 1 >= \lim_{n\to\infty} < \sum_{j=-n}^{n} \Psi_j X_{t-j}, 1 > = E(Y_t)$$

$$= \lim_{n\to\infty} \sum_{j=-n}^{n} \Psi_j E(X_{t-j}) = \left(\sum_{i=-\infty}^{\infty} \Psi_i \right) E(X_t),$$

and

$$< Y_{t+h}, Y_t >= E(Y_{t+h}Y_t') = \lim_{n\to\infty} E\left[\left(\sum_{j=-n}^{n} \Psi_j X_{t+h-j} \right) \left(\sum_{k=-n}^{n} \Psi_k X_{t-k} \right)' \right]$$

$$= \sum_{j,k=-\infty}^{\infty} \Psi_j \left[\gamma_X(h-j+k) + E(X_t)E(X_t') \right] \Psi_k'.$$

Thus, $\{Y_t\}$ is stationary because EY_t and $E(Y_{t+h}Y_t)$ are both finite and independent of t. The autocovariance matrix $\gamma_Y(h)$ is given by

$$\gamma_Y(h) = E(Y_{t+h}Y_t) - E(Y_{t+h})E(Y_t) = \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} \Psi_i \gamma_X(h-i+j)\Psi_j'.$$

□

Proposition 3.5 *Let $\{X_t = (X_{t1}, \ldots, X_{tk})' : t \in \mathbb{Z}\}$ be a sequence of random vectors such that* $\sup\{E|X_{ti}| : t \in \mathbb{Z}, i = 1, \ldots, k\} < \infty$ *and* $\sum_{i=-\infty}^{\infty} \|\Psi_i\| < \infty$. *Then, the series*

$$\sum_{i=-\infty}^{\infty} \Psi_i X_{t-i}$$

converges absolutely with probability one and thus the output random vector $Y_t = \sum_{i=-\infty}^{\infty} \Psi_i X_{t-i}$ exists uniquely. If in addition $\sup\{E|X_{ti}|^2 : t \in \mathbb{Z}, i = 1, \ldots, k\} < \infty$, *then the series converges in mean square to the same limit.*

Proof Let $M = \sup\{E|X_{ti}| : t \in \mathbb{Z}, i = 1, \ldots, k\} < \infty$. Then, by the monotone convergence theorem we can write

$$\begin{aligned} E\left(\sum_{i=-\infty}^{\infty} \|\Psi_i X_{t-i}\|\right) &= \lim_{n\to\infty} E\left(\sum_{i=-n}^{n} \|\Psi_i X_{t-i}\|\right) \\ &\leq \lim_{n\to\infty}\left(\sum_{i=-n}^{n} \|\Psi_i\|\right) M\sqrt{kr} \\ &< \infty. \end{aligned}$$

From this it follows that $\sum_{i=-\infty}^{\infty} \|\Psi_i X_{t-i}\|$ and $\sum_{i=-\infty}^{\infty} \Psi_i X_{t-i}$ are both finite with probability one. Thus, $Y_t = \sum_{i=-\infty}^{\infty} \Psi_i X_{t-i}$ exists and is uniquely determined.

By Proposition 3.4 the series $\sum_{i=-\infty}^{\infty} \Psi_i X_{t-i}$ converges in mean square. Let S_t denote the mean square limit. Then, by Fatou's lemma,

$$\begin{aligned} E(S_t - Y_t)(S_t - Y_t)' &= E \liminf_{n\to\infty}\left(S_t - \sum_{j=-n}^{n} \Psi_j X_{t-j}\right)\left(S_t - \sum_{j=-n}^{n} \Psi_j X_{t-j}\right)' \\ &\leq \liminf_{n\to\infty} E\left(S_t - \sum_{j=-n}^{n} \Psi_j X_{t-j}\right)\left(S_t - \sum_{j=-n}^{n} \Psi_j X_{t-j}\right)' \\ &= 0, \end{aligned}$$

and the limits S_t and Y_t are equal with probability one. □

3.5 Linear Time Series Model Representation for a Stationary Process

An important class of models in time series analysis is that of linear time series models. A famous theorem due to Wold states that every stationary model $\{Y_t\}$ that is nondeterministic (i.e., Y_t cannot be perfectly predicted from past values) can be expressed as the sum of a linear time series model plus a deterministic model.

Definition 3.6 (Linear Time Series Model) The k-dimensional stochastic process $\{Y_t\}$ is said to follow a linear time series model if

$$Y_t = \sum_{j=0}^{\infty} \Psi_j A_{t-j},$$

where $\{A_t\} \sim \text{WN}(0, \Sigma)$, the Ψ_j matrices have dimension $k \times r$, and $\sum_{j=0}^{\infty} \|\Psi_j\|^2 < \infty$. The function $\Psi(z) = \sum_{j=0}^{\infty} \Psi_j z^j$ is called the **transfer function** of the process.

Remark 3.8 In a linear time series model $Y_t = \sum_{j=0}^{\infty} \Psi_j A_{t-j}$, $\{Y_t\}$ is the output of a causal filter applied to a white noise. The filter is not necessarily stable but does satisfy the weaker condition $\sum_{j=0}^{\infty} \|\Psi_j\|^2 < \infty$. Note that $\sum_{j=0}^{\infty} \|\Psi_j\|^2 < \left(\sum_{j=0}^{\infty} \|\Psi_j\|\right)^2$. ◇

Example 3.11 Suppose the scalar stochastic difference equation $Y_t + \phi Y_{t-1} = A_t$ with $|-\phi| < 1$ and $\{A_t\} \sim \text{WN}(0, \sigma^2)$. By Corollary 3.1, the process $Y_t = \sum_{j=0}^{\infty} \Psi_j A_{t-j}$, where $\Psi_j = (-\phi)^j$, is stationary and it is easy to verify that it satisfies the previous difference equation. To see that it is the unique stationary solution, assume that $\{Y_t^*\}$ is another stationary solution. Then, $Z_t = Y_t - Y_t^*$ is stationary and satisfies the homogenous equation $Z_t = -\phi Z_{t-1}$. Letting $V = \text{Var}(Z_t)$, we have $(1-\phi^2)V = 0$ and, since $|-\phi| < 1$, this implies $V = 0$. Thus the difference equation has a unique stationary solution that follows a linear time series model. ◇

The following proposition shows that every process following a linear time series model is stationary and gives its covariance function.

Proposition 3.6 *Let $\{A_t\} \sim \text{WN}(0, \Sigma)$. Then, the series $\sum_{j=0}^{\infty} \Psi_j A_{t-j}$ converges in mean square if, and only if, $\sum_{j=0}^{\infty} \|\Psi_j\|^2 < \infty$. If Y_t is the mean square limit of $\sum_{j=0}^{\infty} \Psi_j A_{t-j}$, then the process $\{Y_t\}$ is stationary with zero mean and covariance function $\gamma_Y(h)$ given by*

$$\gamma_Y(h) = \sum_{j=0}^{\infty} \Psi_{j+h} \Sigma \Psi_j', \quad h \in \mathbb{Z}.$$

Proof Suppose that the series $\sum_{j=0}^{\infty} \Psi_j A_{t-j}$ converges in mean square to Y_t and let $LL' = \Sigma$, where L is lower triangular, be the Cholesky decomposition of Σ. Then, by the continuity of the inner product $< X, Y >= E(XY')$, $E(Y_t) = 0$ and $\text{Cov}(Y_{t+h}, Y_t) = \sum_{j=0}^{\infty} \Psi_{j+h} \Sigma \Psi_j'$, showing that Y_t is stationary and the covariance function is as asserted. In addition,

$$\begin{aligned}
\sum_{j=0}^{n} \|\Psi_j\|^2 &= \sum_{j=0}^{n} \|\Psi_j L L^{-1}\|^2 \\
&\leq \|L^{-1}\|^2 \sum_{j=0}^{n} \|\Psi_j L\|^2 = \|L^{-1}\|^2 \sum_{j=0}^{n} tr(L'\Psi_j'\Psi_j L) \\
&= \|L^{-1}\|^2 \sum_{j=0}^{n} tr(\Psi_j L L' \Psi_j') \\
&\leq \|L^{-1}\|^2 tr\left(Var(Y_t)\right) \\
&< \infty.
\end{aligned}$$

Conversely, suppose that $\sum_{j=0}^{\infty} \|\Psi_j\|^2 < \infty$. Then, if $n > m > 0$, we can write

$$\left\| E(\sum_{m<|j|\le n} \Psi_j A_{t-j})(\sum_{m<|k|\le n} \Psi_k A_{t-k})' \right\|^2 = \sum_{m<|j|\le n} \|\Psi_j \Sigma \Psi_j'\|^2$$
$$\le \|\Sigma\|^2 \sum_{m<|j|\le n} \|\Psi_j\|^2$$
$$\to 0 \quad \text{as} \quad m, n \to \infty,$$

and, by the Cauchy criterion, the series $\sum_{j=0}^{\infty} \Psi_j A_{t-j}$ converges in mean square. □

Definition 3.7 The stationary process $\{Y_t : t \in \mathbb{Z}\}$ is **nondeterministic** if $\Sigma = \text{Var}(Y_t - E^*(Y_t|Y_{t-1}, Y_{t-2}, \ldots)) \neq 0$, and it is **purely deterministic** if $\Sigma = 0$. If Σ is positive definite, we say that the process $\{Y_t\}$ is **nondeterministic of full rank**.

Remark 3.9 The proper framework to deal with projections on infinite $\{\ldots, Y_{t-1}, Y_t, Y_{t+1}, \ldots\}$ or semi-infinite samples $\{\ldots, Y_{t-1}, Y_t\}$ is that of Hilbert spaces. We will deal mainly with finite projections in this book, so that we will use Hilbert spaces only occasionally. We refer the reader to Rozanov (1967), Hannan (1970), Gikhman & Skorokhod (1969), Brockwell & Davis (1991), and Pourahmadi (2001) for the study of stationary processes in Hilbert spaces. ◇

Remark 3.10 Every white noise process is nondeterministic. A process $\{Y_t\}$ following a linear time series model, $Y_t = \sum_{j=0}^{\infty} \Psi_j A_{t-j}$, is nondeterministic if and only if $\Psi_0 \Sigma \Psi_0' \neq 0$ and this in turn happens if and only if Ψ_0 has full column rank because Σ is positive definite. ◇

Theorem 3.2 (The Wold Decomposition) *Let the k-dimensional stationary process $\{Y_t : t \in \mathbb{Z}\}$ be zero-mean and nondeterministic. Then it can be expressed as*

$$Y_t = \sum_{j=0}^{\infty} \Psi_j A_{t-j} + V_t,$$

where the Ψ_j are $k \times r$ matrices with $r \le k$, Ψ_0 has rank r, $\sum_{j=0}^{\infty} \|\Psi_j\|^2 < \infty$, $\{A_t\} \sim \text{WN}(0, \Sigma)$, $E(A_t V_s') = 0$ for all $t, s \in \mathbb{Z}$, and V_t is purely deterministic.

Proof See Rozanov (1967) or Hannan (1970). □

Remark 3.11 There is a dual form of the Wold decomposition in which the time runs backwards instead of forwards. See, for example, Corollary 4.5.9, p. 126, in Lindquist & Picci (2015). ◇

Example 3.12 An example of a deterministic process is the solution $\{Y_t : t = -1, 0, 1, \ldots\}$ of the scalar stochastic difference equation

$$Y_t - 2\rho \cos(\varphi) Y_{t-1} + \rho^2 Y_{t-2} = 0$$

where $|\rho| < 1$ and $\varphi \in [-\pi, \pi]$. This difference equation can be written in vector form as

$$\begin{bmatrix} Y_{t-1} \\ Y_t \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -\rho^2 & 2\rho\cos(\varphi) \end{bmatrix} \begin{bmatrix} Y_{t-2} \\ Y_{t-1} \end{bmatrix}$$

and, by the results of Sect. 3.13, the solution has the form

$$Y_t = X_1\rho^t \cos(t\varphi) + X_2\rho^t \sin(t\varphi),$$

where X_1 and X_2 are random variables that depend linearly on the initial conditions, Y_{-1} and Y_0. Thus, for $t > 0$, $E^*(Y_t|Y_{t-1}, Y_{t-2}, \ldots, Y_{-1}) = X_1\rho^t\cos(t\varphi) + X_2\rho^t\sin(t\varphi) = Y_t$, the innovations are $A_t = Y_t - E^*(Y_t|Y_{t-1}, Y_{t-2}, \ldots, Y_{-1}) = 0$, and the process $\{Y_t\}$ is purely deterministic. ◇

3.6 The Backshift Operator

Suppose a stationary vector process $\{Y_t\} = \{(Y'_{t1}, \ldots, Y'_{tk})'\}$ defined on a probability space $(\Omega, \mathcal{S}, P)$ such that $Y_t \in L_2$ and let H_Y be the closed linear hull generated by all the random variables Y_{ti}, $t = 0, \pm 1, \pm 2, \ldots$, $i = 1, \ldots, k$. Given an element of H_Y of the form $X = \sum_{i=1}^n c_i Y_{t_i j_i}$, the **backshift operator** B is defined by

$$BX = \sum_{i=1}^{n} c_i Y_{t_i - 1, j_i}.$$

This operator has an inverse $F = B^{-1}$, the **forward operator**,

$$FX = \sum_{i=1}^{n} c_i Y_{t_i + 1, j_i}$$

and it preserves the scalar product,

$$\begin{aligned} E\left\{\left[B\left(\sum_{i=1}^{n} c_i Y_{t_i, j_i}\right)\right]\left[B\left(\sum_{i=1}^{m} d_i Y_{s_i, h_i}\right)\right]\right\} &= \sum_{i=1}^{n}\sum_{h=1}^{n} c_i d_h E(Y_{t_i-1} Y_{s_h-1}) \\ &= \sum_{i=1}^{n}\sum_{h=1}^{n} c_i d_h E(Y_{t_i} Y_{s_h}) \\ &= E\left[\left(\sum_{i=1}^{n} c_i Y_{t_i, j_i}\right)\left(\sum_{i=1}^{m} d_i Y_{s_i, h_i}\right)\right]. \end{aligned}$$

By Theorem 1, p. 179, of Gikhman & Skorokhod (1969), the operator B can be extended as a continuous operator to H_Y. It then becomes a unitary operator in H_Y. This implies

$$\|B\| = \sup\left\{\frac{\|BY_{ti}\|}{\|Y_{ti}\|} : \|Y_{ti}\| \neq 0\right\} = 1,$$

where $\|B\|$ is the norm of the operator B and $\|Y_{ti}\|^2 = E|Y_{ti}|^2$.

Remark 3.12 The backshift operator is very useful to express many linear relations in time series analysis. For example, if $\{Y_t\}$ has a linear time series model representation $Y_t = \sum_{j=0}^{\infty} \Psi_j A_{t-j}$, then we can write $Y_t = \Psi(B)A_t$, where $\Psi(B) = \sum_{j=0}^{\infty} \Psi_j B^j$ is an operator defined as the sum of an operator series. From the previous discussion, the norm of this operator satisfies $\|\Psi(B)\| \leq \|\sum_{j=0}^{\infty} \Psi_j B^j\| \leq \sum_{j=0}^{\infty} \|\Psi_j\|$ and thus the operator series converges if $\sum_{j=0}^{\infty} \|\Psi_j\| < \infty$. ◇

Remark 3.13 Recall that, according to Kolmogorov's isomorphism theorem, letting H be the Hilbert space spanned by a scalar stationary process $\{Y_t\}$, there is a unique Hilbert space isomorphism, T, of H onto $L^2(F_Y)$ such that

$$T(Y_t) = e^{it\cdot},$$

where F_Y is the spectral distribution function of the process. Multiplication by the backshift operator in the time domain corresponds to multiplication by the function e^{-ix} in the frequency domain. For example, if $X = aY_t + bY_{t-1}$, then

$$T(BX) = T(aY_{t-1} + bY_{t-2}) = e^{-ix}\left(ae^{itx} + be^{i(t-1)x}\right) = e^{-ix}T(X).$$

◇

3.7 VARMA Models and Innovations State Space Models

Suppose a stationary vector process $\{Y_t\}$ that has the linear time series model representation $Y_t = \sum_{j=0}^{\infty} \Psi_j A_{t-j} = \Psi(B)A_t$, where $\Psi(B) = \sum_{j=0}^{\infty} \Psi_j B^j$ and B is the backshift operator, $BA_t = A_{t-1}$. If one is interested in estimating this model using an observed sample, $\{Y_1, \ldots, Y_n\}$, the situation seems hopeless because there is an infinite number of parameters in the model. This consideration and the fact that there are theorems in functional analysis, like the Stone–Weierstrass theorem, that asserts that every continuous function can be uniformly approximated by polynomials, motivates the search for an approximation to $\Psi(z)$ of the form $\Phi(z)^{-1}\Theta(z)$, where $\Phi(z)$ and $\Theta(z)$ are polynomial matrices in the variable z. Thus, we are led to consider the following definition.

Definition 3.8 A vector processes $\{Y_t\}$ is said to follow a **vector ARMA**(p, q) or **VARMA**(p, q) model if it satisfies a linear stochastic difference equation of the form

$$Y_t + \Phi_1 Y_{t-1} + \cdots + \Phi_p Y_{t-p} = A_t + \Theta_1 A_{t-1} + \cdots + \Theta_q A_{t-q}, \tag{3.8}$$

or, more compactly,

$$\Phi(B)Y_t = \Theta(B)A_t,$$

where $\Phi(B) = I + \Phi_1 B + \cdots + \Phi_p B^p$, $\Theta(B) = I + \Theta_1 B + \cdots + \Theta_q B^q$ and $\{A_t\} \sim \text{WN}(0, \Sigma)$. The function $\Psi(z) = \sum_{j=0}^{\infty} \Psi_j z^j = \Phi^{-1}(z)\Theta(z)$ is called the **transfer function** of the process $\{Y_t\}$.

The acronym ARMA stands for Auto–Regressive–Moving–Average and the polynomial matrices $\Phi(B)$ and $\Theta(B)$ are called the autoregressive and moving average polynomial matrices. To gain some insights into the properties of VARMA models, let us consider the simple scalar process

$$Y_t = \rho Y_{t-1} + A_t, \tag{3.9}$$

where $\{A_t\} \sim \text{WN}(0, \sigma^2)$ and ρ is a real number. If $|\rho| < 1$, we can iterate in (3.9) to get

$$Y_t = \rho^k Y_{t-k} + A_t + \rho A_{t-1} + \cdots + \rho^{k-1} A_{t-k+1}, \qquad k = 2, 3, \ldots$$

From this, it is clear that, letting $k \to \infty$, the linear time series model representation

$$Y_t = \sum_{j=0}^{\infty} \rho^j A_{t-j}$$

is obtained. By Propositions 3.4 and 3.5 and Corollary 3.1, the series $\sum_{j=0}^{\infty} \rho^j A_{t-j}$ converges in mean square and with probability one and the process $Y_t = \sum_{j=0}^{\infty} \rho^j A_{t-j}$ is stationary. In addition, it is the only solution of (3.9) for, if $\{Z_t\}$ is another solution, then $X_t = Y_t - Z_t$ is stationary and satisfies $X_t = \rho X_{t-1}$. Thus, if $V = \text{Var}(X_t)$, we have $V = 0$.

If $\rho = 1$ in (3.9) and we assume that the process starts at some finite point in the past, for example at $t = 1$, we can write

$$Y_t = Y_1 + A_t + A_{t-1} + \cdots + A_2.$$

For any distribution of Y_1 that we may select, it is easy to verify that the process is not stationary. Thus, if $\rho = 1$, we cannot eliminate the effect of the initial condition and the process is not stationary.

Finally, suppose that $|\rho| > 1$ in (3.9). Then, we can divide both terms of (3.9) by ρ to get

$$Y_{t-1} = \frac{1}{\rho} Y_t + W_t, \tag{3.10}$$

where $W_t = -A_t/\rho$ and $\{W_t\} \sim \mathrm{WN}(0, \sigma^2/\rho^2)$. Iterating in this equation, we get

$$Y_t = \frac{1}{\rho^k} Y_{t+k} + W_{t+1} + \frac{1}{\rho} W_{t+2} + \cdots + \frac{1}{\rho^{k-1}} W_{t+k}, \qquad k = 2, 3, \ldots$$

and, letting $k \to \infty$,

$$Y_t = \sum_{k=0}^{\infty} \frac{1}{\rho^k} W_{t+1+k}.$$

As in the case $|\rho| < 1$, the series $\sum_{k=0}^{\infty} \frac{1}{\rho^k} W_{t+1+k}$ converges in mean square and with probability one and is the unique stationary solution of (3.10). However, this kind of solution is undesirable because it depends on the future, which is unknown. For this reason, the following definition is important.

Definition 3.9 A VARMA(p, q) model (3.8) is said to be **causal** if $\{Y_t\}$ can be represented as a linear time series model, $Y_t = \sum_{j=0}^{\infty} \Psi_j A_{t-j}$, for $t = 0, \pm 1, \pm 2, \ldots$, with $\sum_{j=0}^{\infty} \|\Psi_j\| < \infty$, where the transfer function of the model, $\Psi(z) = \Phi^{-1}(z)\Theta(z)$, satisfies $\Psi(z) = \sum_{j=0}^{\infty} \Psi_j z^j$.

We will see later in Theorem 3.7 that if the roots of $\det[\Phi(z)]$ in (3.8) are all greater than one in modulus, then there exists a unique stationary solution Y_t that is causal.

When dealing with VARMA models, the question naturally arises as to whether there exists an infinite autoregressive representation of the process.

Definition 3.10 A VARMA(p, q) model (3.8) is said to be **invertible** if $\{Y_t\}$ can be represented as $Y_t + \sum_{j=1}^{\infty} \Pi_j Y_{t-j} = A_t$, for $t = 0, \pm 1, \pm 2, \ldots$, with $\sum_{j=0}^{\infty} \|\Pi_j\| < \infty$, $\Pi_0 = I$.

We will see later in Proposition 3.11 that if the model (3.8) admits a causal stationary solution $\{Y_t\}$ and the roots of $\det[\Theta(z)]$ in (3.8) are all greater than one in modulus, then Y_t is invertible.

There is a close link between VARMA models and the time invariant state space models (1.37) and (1.38) introduced in Sect. 1.7.2, as we will see later in Theorem 3.6.

A state space form that is very useful when one is working with VARMA models is the one proposed by Akaike (1974a) that we now describe. We will call this representation **Akaike's state space form**. Let $\{Y_t\}$ be a stationary vector process that follows a causal VARMA(p, q) model (3.8) such that $Y_t = \sum_{j=0}^{\infty} \Psi_j A_{t-j}$, where

$\Psi(z) = \Phi^{-1}(z)\Theta(z) = \sum_{j=0}^{\infty} \Psi_j z^j$ is the transfer function of the process $\{Y_t\}$. Define $r = \max(p, q)$ and $x_{t,1} = Y_t - A_t$, $x_{t,2} = Y_{t+1} - A_{t+1} - \Psi_1 A_t, \ldots, x_{t,r} = Y_{t+r-1} - \sum_{j=0}^{r-1} \Psi_j A_{t+r-1-j}$. Then, taking into account the relation $\Phi(z)\Psi(z) = \Theta(z)$, it is not difficult to verify that the following equations hold

$$x_{t+1} = Fx_t + KA_t$$
$$Y_t = Hx_t + A_t,$$

where

$$F = \begin{bmatrix} 0 & I & 0 & \cdots & 0 \\ 0 & 0 & I & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & I \\ -\Phi_r & -\Phi_{r-1} & -\Phi_{r-2} & \cdots & -\Phi_1 \end{bmatrix}, \quad K = \begin{bmatrix} \Psi_1 \\ \Psi_2 \\ \vdots \\ \Psi_{r-1} \\ \Psi r \end{bmatrix}, \tag{3.11}$$

$\Phi_i = 0$ if $i > p$, $x_t = (x'_{t,1}, \ldots, x'_{t,r})'$ and $H = [I, 0, \ldots, 0]$. By the results in Sect. 3.A.2, we can express Y_t in the previous state space representation as

$$Y_t = (A_t + \Psi_1 A_{t-1} + \cdots + \Psi_{t-1} A_1) + h_t x_1, \tag{3.12}$$

where $\Psi_i = HF^{i-1}K$, $i = 1, \ldots, t-1$, and $h_t = HF^{t-1}$. This implies that the transfer function can be expressed in terms of the matrices of the state space model as $\Psi(z) = I + zH(I - Fz)^{-1}K$ and $h_t \to 0$ as $t \to \infty$. In fact, we will see later in Proposition 3.10 that the matrix F has as eigenvalues the inverses of the roots of $\det[\Phi(z)]$ and, because the VARMA process (3.8) is stationary, by Theorem 3.7, the roots of $\det[\Phi(z)]$ are all outside the unit circle. Thus, $F^t \to 0$ as $t \to \infty$.

Remark 3.14 If the process $\{Y_t\}$ is stationary and follows a causal VARMA(p, q) model, the elements of the state vector, x_t, are the first r-step ahead predictors of Y_t. This will be seen in more detail in Chaps. 5 and 7. This interpretation is also valid in the nonstationary case provided the initial conditions are handled appropriately. ◇

Example 3.13 Let $\{Y_t\}$ follow the scalar ARMA$(2, 1)$ process

$$Y_t + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} = A_t + \theta_1 A_{t-1},$$

where $\{A_t\} \sim \mathrm{WN}(0, \sigma^2)$. Then, Akaike's state space form is

$$x_{t+1} = Fx_t + KA_t$$
$$Y_t = Hx_t + A_t,$$

where

$$F = \begin{bmatrix} 0 & 1 \\ -\phi_2 & -\phi_1 \end{bmatrix}, \qquad K = \begin{bmatrix} \theta_1 - \phi_1 \\ \phi_2 - \theta_1\phi_1 + \phi_1^2 \end{bmatrix},$$

$x_t = (x_{t,1}, x_{t,2})'$ and $H = (1, 0)$. ◇

It is to be emphasized that Akaike's representation is by no means the only state space form that can be used with VARMA models. In fact, there are many ways to represent a VARMA model in state space form. For example, another valid representation for the VARMA(p, q) model (3.8) is

$$x_{t+1} = Fx_t + KA_t$$
$$Y_t = Hx_t + A_t,$$

where

$$F = \begin{bmatrix} -\Phi_1 & I & 0 & \cdots & 0 \\ -\Phi_2 & 0 & I & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\Phi_{r-1} & 0 & 0 & \cdots & I \\ -\Phi_r & 0 & 0 & \cdots & 0 \end{bmatrix}, \qquad K = \begin{bmatrix} \Theta_1 - \Phi_1 \\ \Theta_2 - \Phi_2 \\ \vdots \\ \Theta_{r-1} - \Phi_{r-1} \\ \Theta_r - \Phi_r \end{bmatrix},$$

$r = \max(p, q)$, $\Phi_i = 0$ if $i > p$, $\Theta_i = 0$ if $i > q$, and $H = [I, 0, \ldots, 0]$. The previous considerations motivate the following definition.

Definition 3.11 A vector processes $\{Y_t\}$ is said to be in **innovations state space form** if it follows a state space model of the form

$$x_{t+1} = Fx_t + KA_t \tag{3.13}$$
$$Y_t = Hx_t + A_t, \tag{3.14}$$

where $\{A_t\} \sim \mathrm{WN}(0, \Sigma)$. The function $\Psi(z) = \sum_{j=0}^{\infty} \Psi_j z^j = I + zH(I - Fz)^{-1}K$ is called the **transfer function** of the process $\{Y_t\}$.

An innovations state space model is characterized by the system matrices (F, K, Z, I). It is said to be in innovations form because the innovations, A_t, are on both equations of the state space model.

Remark 3.15 It is to be noticed that the expression (3.12), where $\Psi_i = HF^{i-1}K$, $i = 1, \ldots, t-1$, and $h_t = HF^{t-1}$, is satisfied for every innovations state space model. Whether or not $h_t \to 0$ depends on the eigenvalues of F. It will be shown in Chap. 5 that, for stationary processes, the eigenvalues of F are always inside the unit circle and, therefore, $h_t \to 0$. ◇

3.8 Minimality, Observability, and Controllability

As we will see later in Example 3.15, Akaike's state space form has minimal state vector dimension when the observations are scalar. However, in the multivariate case, it is usually possible to obtain an innovations state space form with a state vector of dimension smaller than rk, where $r = \max(p, q)$ and k is the dimension of Y_t. The reason for this is that, given a VARMA(p, q) model (3.8) or its Akaike's state space form (3.13) and (3.14), where $H = [I, 0, \dots, 0]$, and F and K are given by (3.11), the **Hankel matrix**,

$$H_t = \begin{bmatrix} \Psi_1 & \Psi_2 & \Psi_3 & \cdots & \Psi_t \\ \Psi_2 & \Psi_3 & \Psi_4 & \cdots & \Psi_{t+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \Psi_t & \Psi_{t+1} & \Psi_{t+2} & \cdots & \Psi_{2t-1} \end{bmatrix},$$

corresponding to the transfer function $\Psi(z) = \Phi^{-1}(z)\Theta(z) = \sum_{j=0}^{\infty} \Psi_j z^j = I + zH(I - Fz)^{-1}K$ has constant rank r_H such that $r_H \leq rk$ for $t \geq r$, where $r = \max(p, q)$ and $k = \dim(Y_t)$. This is due to the fact that the relation $\Phi(z)\Psi(z) = \Theta(z)$ implies

$$[0, \dots, 0, \Phi_r, \Phi_{r-1}, \dots, \Phi_1, I]H_t = 0, \quad t > r,$$

and, therefore, all the block rows of H_t beyond the rth block row are linearly dependent on the preceding block rows. If $k = 1$, the relation $r_H = r$ holds when $\Phi(z)$ and $\Theta(z)$ have no common factors, but if $k > 1$, r_H can be much smaller than rk. This issue has to do with the concept of **minimality** for state space systems.

Definition 3.12 An innovations state space form (3.13) and (3.14) is said to be **minimal** if the state vector has minimal dimension among all innovation forms that have the same transfer function $\Psi(z) = I + zH(I - Fz)^{-1}K$.

There is a connection between minimality and the concepts of **controllability** and **observability**.

Definition 3.13 The pair of matrices $[A, B]$ such that A and B have dimensions $a \times a$ and $a \times b$ is said to be **controllable** if the controllability matrix, $C_a = [B, AB, \dots, A^{a-1}B]$, has full row rank. The pair $[A, C]$ such that A and C have dimensions $a \times a$ and $c \times a$ is said to be **observable** if the observability matrix, $O_a = [C', A'C', \dots, (A^{a-1})'C']'$, has full column rank.

Given an innovations state space form (3.13) and (3.14), if $[F, K]$ is not controllable and $K \neq 0$, by the results in Chap. 5, there exists a nonsingular matrix P such that

$$P^{-1}FP = \begin{bmatrix} F_c & F_{12} \\ 0 & F_{\bar{c}} \end{bmatrix}, \qquad P^{-1}K = \begin{bmatrix} K_c \\ 0 \end{bmatrix} \tag{3.15}$$

and the pair $[F_c, K_c]$ is controllable. If we premultiply (3.13) by P^{-1}, a new state space form is obtained where the state vector is $\bar{x}_t = P^{-1}x_t$. If we partition $\bar{x}_t = (x'_{c,t}, x'_{\bar{c},t})'$ conformally to (3.15), it is not difficult to verify that $x_{\bar{c},t}$ evolves according to $x_{\bar{c},t+1} = F_{\bar{c}}x_{\bar{c},t}$. Thus, the influence of the initial part, $x_{\bar{c},1}$, will persist unless it is zero. If the eigenvalues of $F_{\bar{c}}$ are all smaller than one in modulus, the influence of $x_{\bar{c},1}$ will diminish until it eventually disappears. Partitioning $P = [P_c, P_2]$ conforming to (3.15) and defining $H_c = HP_c$, the state space form

$$x_{c,t+1} = F_c x_{c,t} + K_c A_t \tag{3.16}$$

$$\overline{Y}_t = H_c x_{c,t} + A_t \tag{3.17}$$

is obtained. It will be shown later in this section that if the pair $[F_c, H_c]$ is observable, then (3.16) and (3.17) is minimal. Note that the output process, $\{\overline{Y}_t\}$, can be different from the original output process, $\{Y_t\}$. Conditions for the equality of the two processes will also be given later in this section.

In a similar way, if $[F, H]$ is not observable in (3.13) and (3.14), by the results in Chap. 5, there exists a nonsingular matrix P such that

$$P^{-1}FP = \begin{bmatrix} F_o & 0 \\ F_{21} & F_{\bar{o}} \end{bmatrix}, \quad P^{-1}K = \begin{bmatrix} K_o \\ K_2 \end{bmatrix}, \quad HP = [H_o, 0], \tag{3.18}$$

and the pair $[F_o, H_o]$ is observable. As in the noncontrollable case, using P^{-1} it is possible to transform (3.13) and (3.14) into an observable state space form. Letting $\bar{x}_t = P^{-1}x_t$ and partitioning $\bar{x}_t = (x'_{o,t}, x'_{\bar{o},t})'$ conformally to (3.18), the following state space form is obtained

$$x_{o,t+1} = F_o x_{o,t} + K_o A_t$$

$$Y_t = H_o x_{o,t} + A_t.$$

Here, the same output process is generated if the initial state satisfies $P^{-1}x_1 = (x'_{o,1}, x'_{\bar{o},1})'$.

To see the connection of minimality to controllability and observability, note that for the state space form (3.13) and (3.14) the following equality holds

$$H_t = \begin{bmatrix} \Psi_1 & \Psi_2 & \Psi_3 & \cdots & \Psi_t \\ \Psi_2 & \Psi_3 & \Psi_4 & \cdots & \Psi_{t+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \Psi_t & \Psi_{t+1} & \Psi_{t+2} & \cdots & \Psi_{2t-1} \end{bmatrix}$$

$$= \begin{bmatrix} HK & HFK & HF^2K & \cdots & HF^{t-1}K \\ HFK & HF^2K & HF^3K & \cdots & HF^tK \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ HF^{t-1}K & HF^tK & HF^{t+1}K & \cdots & HF^{2t-2}K \end{bmatrix}$$

$$= \begin{bmatrix} H \\ HF \\ \vdots \\ HF^{t-1} \end{bmatrix} \left[K, FK, \ldots, F^{t-1}K\right]. \tag{3.19}$$

The precise relation is given by the following theorem.

Theorem 3.3 *Given an innovations state space form (3.13) and (3.14), a necessary and sufficient condition for minimality is that the pairs* $[F, K]$ *and* $[F, H]$ *be controllable and observable.*

Proof Let $n = \dim(x_t)$. Then, the Hankel matrix of order n, H_n, is

$$\begin{aligned} H_n &= \begin{bmatrix} \Psi_1 & \Psi_2 & \Psi_3 & \cdots & \Psi_n \\ \Psi_2 & \Psi_3 & \Psi_4 & \cdots & \Psi_{n+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \Psi_n & \Psi_{n+1} & \Psi_{n+2} & \cdots & \Psi_{2n-1} \end{bmatrix} \\ &= \begin{bmatrix} H \\ HF \\ \vdots \\ HF^{n-1} \end{bmatrix} \left[K, FK, \ldots, F^{n-1}K\right] \\ &= O_n C_n. \end{aligned}$$

If $[F, K]$ and $[F, H]$ are controllable and observable, then the rank of H_n is n and cannot be reduced. Therefore, the state space form is minimal. Conversely, if the state space form is minimal and, for example, $[F, K]$ is not controllable, there is a nonsingular matrix P such that (3.15) holds and $[F_c, K_c]$ is controllable. Then, the Hankel matrix H_n can be expressed as

$$H_n = \begin{bmatrix} H \\ HF \\ \vdots \\ HF^{n-1} \end{bmatrix} PP^{-1} \left[K, FK, \ldots, F^{n-1}K\right]$$

$$= \begin{bmatrix} H_c & * \\ H_cF_c & * \\ \vdots & \vdots \\ H_c(F_c)^{n-1} & * \end{bmatrix} \begin{bmatrix} K_c & F_cK_c & \cdots & (F_c)^{n-1}K_c \\ 0 & 0 & \cdots & 0 \end{bmatrix}$$

$$= \begin{bmatrix} H_c \\ H_cF_c \\ \vdots \\ H_c(F_c)^{n-1} \end{bmatrix} \begin{bmatrix} K_c & F_cK_c & \cdots & (F_c)^{n-1}K_c \end{bmatrix},$$

where the asterisks indicate elements that are not relevant to our discussion. Therefore, we have found an innovations form, (3.16) and (3.17), with the same transfer function, $\Psi(z) = I + zH(I - Fz)^{-1}K = I + zH_c(I - F_cz)^{-1}K_c$, and $\dim(x_{c,t}) < n$, a contradiction. The case in which $[F, H]$ is not observable can be proved analogously. □

Example 3.14 Let the innovations state space model

$$x_{t+1} = ax_t$$
$$Y_t = x_t + A_t,$$

where Y_t is scalar and a is a number different from zero. Thus, $F = a$, $K = 0$ and $H = 1$. It is easy to see that the pair $[F, H]$ is observable and the pair $[F, K]$ is not controllable. Therefore, the state space model is not minimal. In fact, according to (3.19), the Hankel matrices, H_t, are all zero and this means that the transfer function is $\Psi(z) = 1$. A minimal state space form is

$$\overline{Y}_t = A_t,$$

where the state vector is zero. If $|a| < 1$ and the process is assumed to start in the infinitely remote past, the processes $\{Y_t\}$ and $\{\overline{Y}_t\}$ coincide. But if $|a| \geq 1$ and the processes are assumed to start at time $t = 1$, then $Y_t = a^{t-1}x_1 + A_t$ and $\overline{Y}_t = A_t$, $t \geq 1$. We can eliminate the state in the equations for $\{Y_t\}$ as follows. Stack first the observations Y_t and Y_{t+1} to get

$$\begin{bmatrix} Y_t \\ Y_{t+1} \end{bmatrix} = \begin{bmatrix} 1 \\ a \end{bmatrix} x_t + \begin{bmatrix} At \\ A_{t+1} \end{bmatrix}.$$

Then, premultiply by the vector $[-a, 1]$. In this way, the following ARMA model is obtained

$$(1 - aB)Y_t = (1 - aB)A_t.$$

The factor $1 - az$ can be canceled when $|a| < 1$ and the process is assumed to start in the infinitely remote past. However, the initial conditions should be taken into account when the process is assumed to start at some finite time in the past. ◇

It follows from (3.19) and the Cayley–Hamilton theorem that for an innovations state space model (3.13) and (3.14) the Hankel matrices, H_t, have finite rank for $t > r$, where r is the state dimension. To see this, suppose that $P(\lambda) = \det(\lambda I - F) = \lambda^r + a_1\lambda^{r-1} + \cdots + a_r$ is the characteristic polynomial of F. Then, by the Cayley–Hamilton theorem, we get $F^r + a_1F^{r-1} + \cdots + a_rI = 0$ and this implies that the i-block of rows of the matrix $[H', F'H', \ldots, F^{t-1'}H']'$ depends linearly on the previous block rows for $t > r$. Since, as mentioned earlier, the Hankel matrices corresponding to VARMA models (3.8) have finite rank, we have proved half of the following proposition.

Proposition 3.7 *The Hankel matrices, H_t, corresponding to VARMA models (3.8) as well as those corresponding to innovations state space forms (3.13) and (3.14) have finite rank for $t > r$, where r is a fixed positive integer. Conversely, let the linear time series model $Y_t = \sum_{j=0}^{\infty} \Psi_j A_{t-j} = \Psi(B)A_t$, where $\Psi(B) = \sum_{j=0}^{\infty} \Psi_j B^j$ and B is the backshift operator, $BA_t = A_{t-1}$. If there exists a positive integer, r, such that the Hankel matrices, H_t, corresponding to this model have finite rank for $t > r$, then there exist a VARMA model (3.8) and an innovations state space form (3.13) and (3.14) with the same transfer function, $\Psi(z) = \sum_{j=0}^{\infty} \Psi_j z^j$.*

Proof Suppose a linear time series model, $Y_t = \Psi(B)A_t$, such that the Hankel matrices H_t have finite rank for $t > r$, where r is a fixed positive integer. Since all the block rows of H_t beyond the rth block row are linearly dependent on the preceding block rows, by the structure of the H_t matrices, there exist matrices Φ_i, $i = 1, \ldots, r$ satisfying

$$[0, \ldots, 0, \Phi_r, \Phi_{r-1}, \ldots, \Phi_1, I]H_t = 0, \quad t > r.$$

Define $\sum_{j=0}^{i} \Phi_j \Psi_{i-j} = \Theta_i$, $i = 0, 1, \ldots, r$, where $\Phi_0 = I$. Then,

$$\begin{aligned}\Phi(B)Y_t &= \Phi(B)\left[I + \Psi_1 B + \cdots + \Psi_{t-1}B^{t-1} + \Psi_t B^t + \cdots\right]A_t \\ &= \Theta(B)A_t, \quad t > r,\end{aligned}$$

and thus $\{Y_t\}$ satisfies the VARMA model $\Phi(B)Y_t = \Theta(B)A_t$. Since a VARMA model can be put into Akaike's state space form, the proof is complete. □

Since we are interested in linear time series models with some structure, we will make the following assumption.

Assumption 3.1 *Given a transfer function, $\Psi(z) = I + \sum_{i=1}^{\infty} \Psi_i z^i$, we assume that the Hankel matrices, H_t, have finite rank for $t > r$, where r is a fixed positive integer.*

For Akaike's state space form the matrix $[H', F'H', \ldots, F^{r-1'}H']'$ is the unit matrix and, therefore, the observability matrix, $O_{rk} = [H', F'H', \ldots, F^{rk-1'}H']'$, has full rank. Thus, Akaike's state space form is always observable. It is minimal if, and only if, $[F, K]$ is controllable. It will be seen in Example 3.15 that if the observations are scalar ($k = 1$) and $\Phi(z)$ and $\Theta(z)$ have no common factors, the pair $[F, K]$ is controllable, so that in that case Akaike's state space form is minimal. But in the multivariate case ($k > 1$), even if $\Phi(z)$ and $\Theta(z)$ are left coprime (see later in this section), $[F, K]$ is often not controllable. This leads to the study of **echelon forms**, that will be considered later in the book and are defined in terms of certain algebraic invariants called **Kronecker indices** and **McMillan degree**.

Definition 3.14 Under Assumption 3.1, the McMillan degree of the transfer function $\Psi(z) = I + \sum_{i=1}^{\infty} \Psi_i z^i$ is defined as the rank, r_H, of the Hankel matrices H_t for all $t > r$.

Remark 3.16 Note that the structure of the Hankel matrices H_t and the fact that they have finite rank for $t > r$ implies that if $h_t(i,j)$, $t > r$, is the jth row in the ith block or rows of H_t and $h_t(i,j)$ depends linearly on $h_t(i_1,j_1), \ldots, h_t(i_m,j_m)$ with $i_s < i$, $s = 1, \ldots, m$, then the row $h_{t+1}(i+1,j)$ depends linearly on $h_{t+1}(i_1 + 1,j_1), \ldots, h_{t+1}(i_m + 1,j_m)$. ◇

The definition of Kronecker indices has to do with the selection of a basis for the rows of H_t when H_t has finite rank for $t > r$.

Definition 3.15 Under Assumption 3.1, the ith Kronecker index, n_i, $i = 1, 2, \ldots, k$, of the transfer function $\Psi(z) = I + \sum_{i=1}^{\infty} \Psi_i z^i$ is the smallest number such that the ith row in the $(n_i + 1)$th block of rows in the Hankel matrix H_t, $t > r$, is linearly dependent on the previous rows of H_t.

Proposition 3.8 *If, for $i = 1, 2, \ldots, k$, we select the n_i linearly independent ith rows of the first n_i blocks of rows in H_t, $t > r$, we obtain a basis for the space of rows of H_t. Therefore, the sum of the Kronecker indices is equal to the McMillan degree,*

$$\sum_{i=1}^{k} n_i = r_H.$$

Proof Under the assumption of the proposition, due to the structure of H_t, $t > r$, all the ith rows in the jth blocks of rows such that $j > n_i$ will also be linearly dependent on the rows preceding the ith row in the $(n_i + 1)$th block of rows. □

Proposition 3.9 *The dimension of the state vector of any minimal state space form is equal to the McMillan degree.*

Proof It is an immediate consequence of Theorem 3.3. □

The issue of echelon forms, Kronecker indices and McMillan degree will be further addressed in Chap. 5.

Example 3.15 When the observations $\{Y_t\}$ are scalar and the polynomials $\Phi(z)$ and $\Theta(z)$ have no common factors, Akaike's state space representation is minimal. To see this, we have to check that $[F, H]$ is observable and $[F, K]$ is controllable. The pair $[F, H]$ is observable because $[H', F'H', \ldots, F^{r-1'}H']'$ is the unit matrix. To see that $[F, K]$ is controllable, by the results in Chap. 5, it suffices to verify that there does not exist a left eigenvector of F that is orthogonal to K. Assume to the contrary, that is, there is a nonzero vector $v = (v_1, \ldots, v_r)'$ such that $v'F = \lambda v'$ and $v'K = 0$ for some constant λ, where $r = \max(p, q)$. Then, $0 = \Phi_r v_r + \lambda v_1$, $v_1 = \Phi_{r-1} v_r + \lambda v_2$, ..., $v_{r-1} = \Phi_1 v_r + \lambda v_r$. If $v_r = 0$, then $v = 0$ and we have a contradiction. Assume then $v_r \neq 0$. By back substituting in the previous equalities, it is obtained that $v_i = (\Phi_{r-i} + \lambda\Phi_{r-i-1} + \cdots + \lambda^{r-i})v_r$, $i = 1, \ldots, r-1$. Taking these expressions into $v'K = 0$ and considering the relation $\Phi(z)\Psi(z) = \Theta(z)$ yields

$$\begin{aligned} v'K &= v_1\Psi_1 + \cdots + v_r\Psi_r \\ &= \{\Psi_1\left(\Phi_{r-1} + \lambda\Phi_{r-2} + \cdots + \lambda^{r-1}\right) + \cdots + \Psi_{r-1}\left(\Phi_1 + \lambda\right) + \Psi_r\}v_r \\ &= \{\Theta_r - \Phi_r + \lambda\left(\Theta_{r-1} - \Phi_{r-1}\right) + \cdots + \lambda^{r-1}\left(\Theta_1 - \Phi_1\right)\}v_r \qquad (3.20) \\ &= 0, \end{aligned}$$

where $\Theta_i = 0$ if $i > q$. Because λ is an eigenvalue of F and the eigenvalues of F are the inverse roots of $\Phi(z)$, the equality $\Phi_r + \Phi_{r-1}\lambda + \cdots + \Phi_1\lambda^{r-1} + \lambda^r = 0$ holds. This, together with $v_r \neq 0$ and (3.20), implies $\lambda^r + \Theta_1\lambda^{r-1} + \cdots + \Theta_r = 0$, but then the polynomials $\Phi(z)$ and $\Theta(z)$ would have a common factor and we have found a contradiction. ◇

Example 3.16 Let $\{Y_t\}$ follow the ARMA(2, 2) process

$$Y_t + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} = A_t + \theta_1 A_{t-1} + \theta_2 A_{t-2},$$

where $\{A_t\} \sim \text{WN}(0, \sigma^2)$. This model has the minimal state space representation

$$\begin{aligned} x_{t+1} &= Fx_t + KA_t \\ Y_t &= Hx_t + A_t, \end{aligned}$$

where

$$F = \begin{bmatrix} 0 & 1 \\ -\phi_2 & -\phi_1 \end{bmatrix}, \qquad K = \begin{bmatrix} \psi_1 \\ \psi_2 \end{bmatrix},$$

$\psi(z) = \sum_{j=0}^{\infty} \psi_j z^j = \theta(z)/\phi(z)$, $\theta(z) = 1 + \theta_1 z + \theta_2 z^2$, $\phi(z) = 1 + \phi_1 z + \phi_2 z^2$, $x_t = (x_{t,1}, x_{t,2})'$ and $H = (1, 0)$. ◇

The next theorem gives a necessary and sufficient condition for two innovations state space forms to be minimal.

Theorem 3.4 *Two minimal state space forms, (3.13) and (3.14) and*

$$\bar{x}_{t+1} = \overline{F}\bar{x}_t + \overline{K}A_t \tag{3.21}$$

$$\overline{Y}_t = \overline{H}\bar{x}_t + A_t, \tag{3.22}$$

have the same transfer function if, and only if, there exists a nonsingular matrix, P, such that $\overline{F} = P^{-1}FP$, $\overline{K} = P^{-1}K$ and $\overline{H} = HP$.

Proof If a nonsingular matrix P exists with $\overline{F} = P^{-1}FP$, $\overline{K} = P^{-1}K$ and $\overline{H} = HP$, then the Hankel matrix, H_t, can be expressed as

$$\begin{aligned} H_t &= \begin{bmatrix} H \\ HF \\ \vdots \\ HF^{t-1} \end{bmatrix} PP^{-1}\left[K, FK, \ldots, F^{t-1}K\right] \\ &= \begin{bmatrix} \overline{H} \\ \overline{HF} \\ \vdots \\ \overline{HF}^{t-1} \end{bmatrix} \left[\overline{K}\ \overline{FK} \cdots \overline{F}^{t-1}\overline{K}\right] \end{aligned}$$

and both state space forms have the same transfer function. Conversely, suppose that both state space forms have the same transfer function. Let $n = \dim(x_t)$ and let O_n, C_n, $\overline{O}_n$, and $\overline{C}_n$ be the observability and controllability matrices of (3.13) and (3.14) and (3.21) and (3.22), respectively. Then,

$$H_n = O_nC_n = \overline{O}_n\overline{C}_n$$

and thus

$$C_n = (O_n'O_n)^{-1}O_n'\overline{O}_n\overline{C}_n \quad \text{and} \quad \overline{C}_n = (\overline{O}_n'\overline{O}_n)^{-1}\overline{O}_n'O_nC_n.$$

Letting $P = (O_n'O_n)^{-1}O_n'\overline{O}_n$ and $\overline{P} = (\overline{O}_n'\overline{O}_n)^{-1}\overline{O}_n'O_n$, it follows that $P\overline{P} = I$. Then, from the equalities

$$O_nFC_n = \overline{O}_n\overline{FC}_n,$$

$\overline{C}_n = P^{-1}C_n$ and $\overline{O}_n = O_nP$, we get $\overline{F} = P^{-1}FP$, $\overline{K} = P^{-1}K$ and $\overline{H} = HP$. □

We now turn to the question of finding solutions to the stochastic difference equation (3.8). First we note that the polynomial matrices $\Phi(z)$ and $\Theta(z)$ in (3.8) should have no left factors other than **unimodular** polynomial matrices, that is, polynomial matrices such that their determinants are nonzero constants. Two polynomial matrices satisfying this property are called **left coprime**. The property of left coprimeness is an extension to the multivariate case of the property of having no common factors other than constants when considering two scalar polynomials.

The following lemma will be useful later.

Lemma 3.1 *Let the matrices A and B have dimensions $p \times k$ and $k \times p$. Then,*

$$\det(I_p + AB) = \det(I_k + BA).$$

Proof Consider the equality

$$\begin{bmatrix} I_p & A \\ 0 & I_k \end{bmatrix} \begin{bmatrix} I_p & -A \\ B & I_k \end{bmatrix} = \begin{bmatrix} I_p + AB & 0 \\ B & I_k \end{bmatrix}.$$

Then, the lemma follows from the identity

$$\det \begin{bmatrix} A & C \\ B & D \end{bmatrix} = \det(A) \det(D - BA^{-1}C).$$

□

Suppose the VARMA model (3.8) and consider Akaike's state space representation (3.13) and (3.14), where F and K are given by (3.11) and $H = [I, 0, \ldots, 0]$. The following proposition gives expressions for $\det[\Phi(z)]$ and $\det[\Theta(z)]$ in terms of the matrices in (3.13) and (3.14).

Proposition 3.10 *Suppose the VARMA model (3.8). Then,* $\det[\Phi(z)]$ *and* $\det[\Theta(z)]$ *can be expressed as*

$$\det[\Phi(z)] = \det(I - Fz) \qquad \text{and} \qquad \det[\Theta(z)] = \det[I - (F - KH)z]$$

in terms of the matrices of Akaike's innovations state space representation (3.13) and (3.14), where F and K are as in (3.11) and $H = [I, 0, \ldots, 0]$. *Thus, the roots of* $\Phi(z)$ *and* $\Theta(z)$ *coincide with the inverses of the eigenvalues of F and* $F - KH$.

Proof It is a standard exercise in linear algebra to show that $\det[\Phi(z)] = \det(I - Fz)$. To prove the formula for $\det[\Theta(z)]$, consider first the transfer functions $x_t = z(I - Fz)^{-1}KA_t$ and $Y_t = [I + zH(I - Fz)^{-1}K]A_t$ corresponding to (3.13) and (3.14). Then, $\Phi^{-1}(z)\Theta(z) = I + zH(I - Fz)^{-1}K$ and, using Lemma 3.1, it is obtained that

$$\begin{aligned} \det[\Theta(z)] &= \det[\Phi(z)] \det[\Phi^{-1}(z)\Theta(z)] \\ &= \det(I - Fz) \det[I + zH(I - Fz)^{-1}K] \\ &= \det(I - Fz) \det[I + z(I - Fz)^{-1}KH] \\ &= \det[I - (F - KH)z]. \end{aligned}$$

□

Remark 3.17 For any other innovations form obtained from Akaike's innovations state space representation (3.13) and (3.14), where F and K are as in (3.11) and $H = [I, 0, \dots, 0]$, by means of a transformation,

$$x^-_{t+1} = F^- x^-_t + K^- A_t$$
$$Y_t = H^- x^-_t + A_t,$$

where $F^- = Q^{-1}FQ$, $K^- = Q^{-1}K$, $H^- = HQ$ and Q is nonsingular, it holds that

$$\det[\Phi(z)] = \det(I - F^- z) \qquad \text{and} \qquad \det[\Theta(z)] = \det[I - (F^- - K^- H^-)z].$$

To see this, consider, for example, that $\det(I - F^- z) = \det[Q^{-1}(I - Fz)Q]$. ◇

Remark 3.18 As mentioned earlier, if Akaike's state space representation is not minimal, the pair $[F, K]$ is not controllable. In that case, there is a nonsingular matrix P such that (3.15) holds and $[F_c, K_c]$ is controllable. Then, proceeding as in the proof of Theorem 3.3, the Hankel matrix H_t, $t \geq r$, can be expressed as

$$H_t = \begin{bmatrix} \Psi_1 & \Psi_2 & \Psi_3 & \cdots & \Psi_t \\ \Psi_2 & \Psi_3 & \Psi_4 & \cdots & \Psi_{t+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \Psi_t & \Psi_{t+1} & \Psi_{t+2} & \cdots & \Psi_{2t-1} \end{bmatrix}$$
$$= \begin{bmatrix} H_c \\ H_c F_c \\ \vdots \\ H_c (F_c)^{n-1} \end{bmatrix} \begin{bmatrix} K_c & F_c K_c & \cdots & (F_c)^{n-1} K_c \end{bmatrix}, \quad t \geq r. \tag{3.23}$$

Because the matrix $[H', F'H', \dots, F^{r-1'}H']'$ is the unit matrix, I_{rk}, where $\dim(Y_t) = k$, the matrix $[H_c', F_c'H_c', \dots, (F_c)^{n-1'}H_c']'$ has full rank and the pair $[F_c, K_c]$ is observable. Thus, the representation (3.16) and (3.17) so constructed is minimal. Note that both Akaike's representation and (3.16) and (3.17) produce exactly the same Ψ_i weights of the transfer function $\Psi(z) = I + \sum_{i=1}^{\infty} \Psi_i z^i$. They generate the same process $\{Y_t\}$ only if the initial conditions are appropriately selected, as we will see in the following theorem. ◇

Theorem 3.5 *Assume that the process $\{Y_t\}$ follows a VARMA model (3.8) and Akaike's state space model is not minimal. Then, there exists a nonsingular matrix P such that (3.15) holds with $[F_c, K_c]$ controllable and (3.16) and (3.17) minimal,*

$$\det[\Phi(z)] = \det(I - F_c z)\det(I - F_{\bar{c}} z) \tag{3.24}$$

and

$$\det[\Theta(z)] = \det[I - (F_c - K_cH_c)z]\det(I - F_{\bar{c}}z), \tag{3.25}$$

where $\det(I - F_{\bar{c}}z) = 1$ *if, and only if,* $\Phi(z)$ *and* $\Theta(z)$ *are left coprime. In addition, if the innovations process,* $\{A_t\}$*, is given and* $Y_{1:r} = (Y_1', \ldots, Y_r')'$ *are initial conditions such that* $P^{-1}x_1 = (x_{c,1}', 0)'$*, where the partition is conformal to (3.15), then (3.16) and (3.17), Akaike's representation, and (3.8) all generate the same process* $\{Y_t\}$*.*

Proof Since Akaike's state space model is not minimal, we proceed as in Remark 3.18 to construct the minimal state space model (3.16) and (3.17). If we take $Y_{1:r}$ as initial conditions such that $P^{-1}x_1 = (x_{c,1}', x_{\bar{c},1}')' = (x_{c,1}', 0)'$ and assume that the innovations process, $\{A_t\}$, is given, we can write

$$\begin{aligned} Y_t &= HF^{t-1}x_1 + A_t + HKA_{t-1} + \cdots + HF^{t-2}KA_1 \\ &= H(PP^{-1})F^{t-1}(PP^{-1})x_1 + A_t + \Psi_1 A_{t-1} + \cdots + \Psi_{t-1}A_1 \\ &= H_c(F_c)^{t-1}x_{c,1} + A_t + H_cK_cA_{t-1} + \cdots + H_c(F_c)^{t-2}K_cA_1, \quad t = 1, 2, \ldots, \end{aligned}$$

and we see that the same process, $\{Y_t\}$, is generated by Akaike's representation, (3.16) and (3.17), and, obviously, (3.8).

Letting $P^{-1}x_t = (x_{c,t}', x_{,t}')'$, we can write

$$\begin{bmatrix} x_{c,t+1} \\ x_{\bar{c},t+1} \end{bmatrix} = \begin{bmatrix} F_c & F_{12} \\ 0 & F_{\bar{c}} \end{bmatrix} \begin{bmatrix} x_{c,t} \\ x_{\bar{c},t} \end{bmatrix} + \begin{bmatrix} K_c \\ 0 \end{bmatrix} A_t$$

$$Y_t = [H_c, HP_2] \begin{bmatrix} x_{c,t} \\ x_{\bar{c},t} \end{bmatrix} + A_t, \quad t = 1, 2, \ldots$$

Formulas (3.24) and (3.25) follow from this, Proposition 3.10 and Remark 3.17. To prove the statement about $\det(I - F_{\bar{c}}z)$, express first (3.16) as $(I - F_cB)x_{c,t} = K_cA_{t-1}$ and premultiply then this expression by $\Lambda(B) = \mathrm{adj}(I - F_cB)$, where $\mathrm{adj}(M)$ denotes the adjoint matrix of M, to get $\det(I - F_cB)x_{c,t} = \Lambda(B)K_cA_{t-1}$. Substituting in (3.17), it is obtained that

$$\det(I - F_cB)Y_t = H_c\Lambda(B)K_cA_{t-1} + \det(I - F_cB)A_t. \tag{3.26}$$

Since the previous expression is a VARMA model for $\{Y_t\}$, if $\Phi(z)$ and $\Theta(z)$ are left coprime in (3.8), we conclude that the autoregressive polynomial matrix in (3.26), $\det(I - F_cz)I$, must contain all the roots of $\det[\Phi(z)]$ and, therefore, $\det(I - F_{\bar{c}}z) = 1$. Conversely, if $\det(I - F_{\bar{c}}z) = 1$ and the polynomial matrices $\Phi(z)$ and $\Theta(z)$ are not left coprime, we can cancel some nonunimodular common left factor to get polynomial matrices $\tilde{\Phi}(z)$ and $\tilde{\Theta}(z)$ with degree of $\det[\tilde{\Phi}(z)]$ strictly less than degree of $\det[\Phi(z)]$. By the first part of the proof, we would get another minimal representation,

$$x_{t+1}^- = F^- x_t^- + K^- A_t$$
$$Y_t = H^- x_t^- + A_t,$$

where $F^- = Q^{-1}F_cQ$ for some matrix Q and, because $\tilde{\Phi}(z)$ and $\tilde{\Theta}(z)$ are left coprime, $\det(I - F^-z) = \det[\tilde{\Phi}(z)]$. This would imply $\det(I - F^-z) = \det(I - F_cz)$ and, therefore, $\det[\tilde{\Phi}(z)] = \det[\Phi(z)]$, a contradiction. □

Remark 3.19 If the process $\{Y_t\}$ is stationary and we assume that the origin is in the infinitely remote past, then the condition $P^{-1}x_1 = (x'_{c,1}, 0)'$ in the last theorem is automatically satisfied. This is a consequence of the definition of x_1 and the relation

$$\begin{bmatrix} Y_{t+1|t} \\ Y_{t+2|t} \\ \vdots \\ Y_{t+i|t} \\ \vdots \end{bmatrix} = \begin{bmatrix} \Psi_1 & \Psi_2 & \cdots\cdots \\ \Psi_2 & \Psi_3 & \cdots\cdots \\ \vdots & \vdots & \vdots \quad \vdots \\ \Psi_i & \Psi_{i+1} & \cdots\cdots \\ \vdots & \vdots & \vdots \quad \vdots \end{bmatrix} \begin{bmatrix} A_t \\ A_{t-1} \\ \vdots \\ A_{t-i} \\ \vdots \end{bmatrix},$$

where $Y_{t+i|t} = Y_{t+i} - \sum_{j=0}^{i-1} \Psi_j A_{t+i-j}$ is the predictor of Y_{t+i} based on $\{Y_s : s \leq t\}$. Thus, relations among the rows of the Hankel matrices are equivalent to relations among predictors. The expression $P^{-1}x_1 = (x'_{c,1}, 0)'$ simply means that there are relations among the predictors in the initial state vector, x_1. More details about these relations will be given in Chap. 5. ◇

The concept of minimality is similar to left coprimeness for VARMA models (Hannan & Deistler, 1988; Kailath, 1980). The following theorem states the precise relationship between both concepts and establishes also the equivalence between VARMA models and innovations state space forms.

Theorem 3.6 *The transfer function,* $\Psi(z) = \Phi^{-1}(z)\Theta(z)$ *, of a VARMA model can always be represented by a minimal innovations state space model with state vector dimension equal to the McMillan degree of* $\Psi(z)$*. Conversely, the transfer function,* $\Psi(z) = I + zH(I - Fz)^{-1}K$*, of an innovations state space model can always be represented by a VARMA model (3.8) with* $\Phi(z)$ *and* $\Theta(z)$ *that are left coprime and McMillan degree of* $\Psi(z) = \Phi^{-1}(z)\Theta(z)$ *equal to* $\max\{\delta \det[\Phi(z)], \delta \det[\Theta(z)]\}$*, where, given a polynomial* $p(z)$*,* $\delta p(z)$ *denotes the degree of* $p(z)$*.*

Proof Suppose a VARMA model (3.8), let $r = \max\{p, q\}$, and let $\Phi(z)$ and $\Theta(z)$ be the autoregressive and moving average polynomial matrices. Consider Akaike's state space form. Then, proceeding as in Remark 3.18, we construct a minimal and, therefore, controllable and observable state space model (3.16) and (3.17) with transfer function $\Psi(z) = I + zH^*(I - F^*z)^{-1}K^* = \Phi^{-1}(z)\Theta(z)$. By (3.23), it is clear that the rank of the Hankel matrix H_t, $t \geq r$, is equal to the dimension of the state vector x_t^*.

Conversely, suppose that $\{Y_t\}$ follows an innovations state space model (3.13) and (3.14) with transfer function $I + zH(I - Fz)^{-1}K = \Psi(z)$ and let the Kronecker indices be n_i, $i = 1, \ldots, k$. The definition of the Kronecker indices and the structure of the Hankel matrices H_t imply that there exists a unique matrix, $\Phi = [\Phi_r, \ldots, \Phi_1, I]$, where $r = \max\{n_i : i = 1, \ldots, k\}$, such that $\Phi H_{r+1} = 0$. The rows of Φ simply express each of the rows of the $(r+1)$th block of rows of H_{r+1} as a linear combination of the basis of rows of Proposition 3.8. By (3.19), this implies $\Phi O_{r+1} = 0$ where $O_{r+1} = [H', F'H', \ldots, F'^{r}H']'$. Thus, if we stack the observations to get

$$Y_{t:t+r} = O_{r+1}x_t + \hat{\Psi}_{r+1}A_{t:t+r} \tag{3.27}$$

where $Y_{t:t+r} = (Y_t', \ldots, Y_{t+r}')'$, $A_{t:t+r} = (A_t', \ldots, A_{t+r}')'$ and

$$\hat{\Psi}_{r+1} = \begin{bmatrix} I & & & \\ HK & I & & \\ \vdots & \vdots & \ddots & I \\ HF^{r-1}K & \cdots & \cdots & HK\ I \end{bmatrix} = \begin{bmatrix} I & & & \\ \Psi_1 & I & & \\ \vdots & \vdots & \ddots & I \\ \Psi_r & \cdots & \cdots & \Psi_1\ I \end{bmatrix}, \tag{3.28}$$

and we premultiply (3.27) by Φ, the following VARMA model is obtained

$$\Phi(B)Y_t = \Theta(B)A_t,$$

where $\Phi(z) = I + \Phi_1 z + \cdots + \Phi_r z^r$, $\Theta(z) = I + \Theta_1 z + \cdots + \Theta_r z^r$, and the Θ_i, $i = 1, \ldots, r$, are given by the product $\Phi\hat{\Psi}_{r+1}$. This VARMA model must be left coprime because otherwise we can cancel some common nonunimodular left factor to get left coprime polynomial matrices, $\hat{\Phi}(z)$ and $\hat{\Theta}(z)$, with degrees of $\det[\hat{\Phi}(z)]$ and $\det[\hat{\Theta}(z)]$ strictly less than the degrees of $\det[\Phi(z)]$ and $\det[\Theta(z)]$, respectively, $\hat{\Phi}(0) = I$ and $\hat{\Theta}(0) = I$. Premultiplying (3.27) by the matrix $\hat{\Phi} = [\hat{\Phi}_r, \ldots, \hat{\Phi}_1, I]$, where $\hat{\Phi}_j = 0$ if $j > s$ and $\hat{\Phi}(z) = \sum_{i=0}^{s} \hat{\Phi}_i z^i$, we obtain $\hat{\Phi} O_{r+1} = 0$ and, therefore, $\hat{\Phi} H_{r+1} = 0$. Since the degree of $\det[\hat{\Phi}(z)]$ is less than the degree of $\det[\Phi(z)]$, the expression $\hat{\Phi} H_{r+1} = 0$ implies a simplification in the unique representation, $\Phi H_{r+1} = 0$, of the rows of the $(r+1)$th block of rows of H_{r+1} as a linear combination of the basis of rows of Proposition 3.8, a contradiction.

To prove the last statement of the theorem, assume without loss of generality that the polynomial matrices $\Phi(z)$ and $\Theta(z)$ are in echelon form, defined in Sect. 5.9.1. Thus, $\Phi(z)$ and $\Theta(z)$ are left coprime and are of the form

$$\Phi(z) = \Phi_0 + \Phi_1 z + \cdots + \Phi_r z^r, \qquad \Theta(z) = \Phi_0 + \Theta_1 z + \cdots + \Theta_r z^r,$$

where Φ_0 is a lower triangular matrix with ones in the main diagonal and the degree of each entry of both the ith rows of $\Phi(z)$ and $\Theta(z)$ is less than or equal to n_i, $i = 1, \ldots, k$.

Next, let us express the polynomial matrix $\Phi(z)$ in the form

$$\Phi(z) = \text{diag}(z^{n_1}, \ldots, z^{n_k})\Phi_c + \Phi_{rp}(z),$$

where Φ_c is a constant matrix and $\Phi_{rp}(z)$ is a polynomial matrix such that the degree of each term of its ith row is less than n_i, $i = 1, \ldots, k$. For example, if

$$\Phi(z) = \begin{bmatrix} 3z^2 - 2z + 1 & -5z^2 \\ 3z + 2 & -z + 1 \end{bmatrix},$$

then

$$\Phi(z) = \begin{bmatrix} z^2 & 0 \\ 0 & z \end{bmatrix} \begin{bmatrix} 3 & -5 \\ 3 & -1 \end{bmatrix} + \begin{bmatrix} -2z + 1 & 0 \\ 2 & 1 \end{bmatrix}.$$

Similarly, let

$$\Theta(z) = \text{diag}(z^{n_1}, \ldots, z^{n_k})\Theta_c + \Theta_{rp}(z),$$

where Θ_c and $\Theta_{rp}(z)$ are defined analogously to Φ_c and $\Phi_{rp}(z)$.

With these definitions, we will now prove that $\delta\det[\Phi(z)] = n$ if, and only if, Φ_c is nonsingular, where $n = \sum_{i=1}^{k} n_i$ is the McMillan degree. Once this is proved, we will also have proved that $\delta\det[\Theta(z)]= n$ if, and only if, Θ_c is nonsingular because $\Theta(z)$ is also in echelon form. To prove the first proposition, let $\Phi_c = LQ$ be the LQ decomposition of Φ_c, where L and Q are a lower triangular and an orthogonal matrix, respectively. Then,

$$\begin{aligned} \det[\Phi(z)] &= \det\left[\text{diag}(z^{n_1}, \ldots, z^{n_k})L + \Theta_{rp}(z)Q'\right]\det(Q) \\ &= \det(\Phi_c)z^{\sum_{i=1}^{k} n_i} + \text{terms with degree less than } n. \end{aligned}$$

Returning to the last statement of the theorem and assuming without loss of generality that $\Phi(z)$ and $\Theta(z)$ are in echelon form, if $\delta\det[\Phi(z)] < n$ and $\delta\det[\Theta(z)] < n$, then, by the previous propositions, both Φ_c and Θ_c are singular matrices. But this implies that some Kronecker index could be made smaller, a contradiction. □

Remark 3.20 It will be shown in Theorem 5A.7 of the Appendix to Chap. 5 that if the transfer function of a VARMA model (3.8) in echelon form, defined in Sect. 5.9.1, is expressed in terms of the forward operator, then the McMillan degree coincides with the degree of the determinant of the denominator polynomial matrix. This is one of the advantages of working with the forward instead of the backshift operator. More details about VARMA models expressed in terms of the forward operator will be given in the Appendix to Chap. 5. ◇

Theorem 3.7 *Assume that* $\Phi(z)$ *and* $\Theta(z)$ *are left coprime in (3.8). Then, the stochastic difference equation (3.8) has a unique causal stationary nondeterministic solution of full rank if, and only if, the roots of the polynomial det*$[\Phi(z)]$ *are all greater than one in modulus.*

Proof Consider Akaike's state space representation corresponding to $\Phi(z)$ and $\Theta(z)$. If it is not minimal, we use the procedure described in Remark 3.18 to obtain a minimal state space representation from it. Thus, by Theorem 3.5, we can assume without loss of generality that there is a minimal innovations state space representation (3.13) and (3.14) such that $I + zH(I - Fz)^{-1}K = \Phi^{-1}(z)\Theta(z)$ and $\det(I - Fz) = \det[\Phi(z)]$.

To prove sufficiency, assume that all the nonzero roots of the polynomial $\det[\Phi(z)]$ are greater than one in modulus. Then, the eigenvalues of F have all modulus less than one. Let the Schur decomposition of F be $F = PUP'$, where P is an orthogonal matrix and U is an upper triangular matrix. Then, $\|F^j\| = \|U^j\|$ and letting $U = D + N$, where D is a diagonal matrix with the eigenvalues of F in the main diagonal and N is upper triangular with zeros in the main diagonal, we can write $(D+N)^j = D^j(I + F_j)$, where F_j is a matrix such that all of its elements tend to zero as $j \to \infty$. Thus, for j sufficiently large $\|U^j\| \leq \rho^j K$, where $\rho = |\lambda_M| < 1$, λ_M is the eigenvalue of F with largest modulus and K is a positive constant. From this it follows that the series $\sum_{j=0}^{\infty} \|F^j\|$ converges and, by Propositions 3.4 and 3.5 and Corollary 3.1, the series $\sum_{j=0}^{\infty} F^j K A_{t-1-j}$ converges in mean square and with probability one to a unique causal stationary process $\{x_t\}$. It is easy to see that this process satisfies the stochastic difference equation (3.13). In addition, $\{x_t\}$ is the unique causal stationary solution of (3.13). To see this, assume there is another solution $\{x_t^*\}$. Then, the difference $H_t = x_t - x_t^*$ would be stationary and would satisfy the homogeneous equation $H_{t+1} = FH_t$. Letting $V = \text{Var}(H_t)$, we would have $(I - F \otimes F)\text{vec}(V) = 0$ and since F has eigenvalues with modulus less than one, the only solution would be $V = 0$. Thus, the process $Y_t = Hx_t + A_t$ is a stationary solution of (3.8). To see that it is the unique causal stationary solution, assume that there is another solution $\{Y_t^*\}$. Then, we would get another solution x_t^* of (3.13) and, because it is unique, it would coincide with x_t. The solution $\{Y_t\}$ is nondeterministic of full rank because $\text{Var}(Y_t) = HVH' + \Sigma$, where $V = \text{Var}(x_t)$ and $\Sigma = \text{Var}(A_t)$ is positive definite.

To prove necessity, assume that a unique causal stationary nondeterministic solution of full rank of (3.8) exists. Then, because $Hx_t = Y_t - A_t$ is stationary in (3.14) and the pair $[F, H]$ is observable, (3.13) has a causal stationary solution $\{x_t\}$ such that $V = FVF' + K\Sigma K'$, where $V = \text{Var}(x_t)$. To see this, stack first the observations,

$$Y_{t:t+r-1} = O_r x_t + \hat{\Psi}_r A_{t:t+r-1},$$

where r is the dimension of x_t, $Y_{t:t+r-1} = (Y_t', Y_{t+1}', \ldots, Y_{t+r-1}')'$, $A_{t:t+r-1} = (A_t', A_{t+1}', \ldots, A_{t+r-1}')'$, $O_r = [H', F'H', \ldots, F'^{r-1}H']'$ and $\hat{\Psi}_r$ is that of (3.28) but with

$r+1$ replaced with r. Then, premultiply by O_r' and $(O_r'O_r)^{-1}$ in succession to get

$$x_t = (O_r'O_r)^{-1}O_r'(Y_{t:t+r-1} - \hat{\Psi}_r A_{t:t+r-1}),$$

where it is to be noted that $O_r'O_r$ is nonsingular because $[F, H]$ is observable and O_r has full rank. If there is an eigenvalue of F with $|\lambda| \geq 1$, let v be a left eigenvector of F associated with λ. Then, we have $(1 - |\lambda|^2)v'V\bar{v} = v'K\Sigma K'\bar{v}$, which implies $v'K = 0$ and the pair $[F, K]$ would not be controllable, contradicting the minimality of (3.13) and (3.14). □

The following proposition gives a sufficient condition for invertibility of a VARMA(p, q) model.

Proposition 3.11 *Assume that* $\Phi(z)$ *and* $\Theta(z)$ *are left coprime in (3.8) and this stochastic difference equation has a unique stationary causal solution* $\{Y_t\}$. *Then,* $\{Y_t\}$ *is invertible if, and only if, the roots of the polynomial* $\det[\Theta(z)]$ *are all greater than one in modulus.*

Proof The proof is similar to that of Theorem 3.7. □

Example 3.17 Let $\{Y_t\}$ follow the MA(2) process

$$Y_t = A_t + \theta_1 A_{t-1} + \theta_2 A_{t-2},$$

where $\{A_t\} \sim \text{WN}(0, \sigma^2)$. Assuming that the two roots of the polynomial $\Theta(z) = 1 + \theta_1 z + \theta_2 z^2$ are greater than one in modulus, the invertible model is $A_t = \sum_{j=0}^{\infty} \pi_j Y_{t-j}$. The π_j can be recursively obtained from the equality $\pi(z)\Theta(z) = 1$, where $\pi(z) = \sum_{j=0}^{\infty} \pi_j z^j$. Thus,

$$\begin{aligned} \pi_0 &= 1 \\ \pi_1 + \theta_1 &= 0 \\ \pi_2 + \pi_1\theta_1 + \theta_2 &= 0 \\ \pi_k + \pi_{k-1}\theta_1 + \theta_2 &= 0, \qquad k > 2. \end{aligned}$$

◇

The state space form (3.13) and (3.14) will be very useful to obtain the forecasts of Y_t based on the finite sample $\{Y_{t-1}, \ldots, Y_1\}$. This will be seen in more detail when we describe forecasting using the Kalman filter.

3.9 Finite Linear Time Series Models

Sometimes, we will consider linear time series models that are not stationary. A simple example of that kind of model is the univariate random walk, defined by $Y_t = Y_{t-1} + A_t$, $\{A_t\} \sim \text{WN}(0, \sigma^2)$. If we assume that Y_1 is a random variable uncorrelated with the A_t, $t = 1, 2, \ldots$, we can express Y_t as $Y_t = (Y_1 - A_1) + \sum_{j=0}^{t-1} A_{t-j}$. Then, $\text{Var}(Y_t) = \text{Var}(Y_1) + (t-1)\sigma^2$ and the process is not stationary. The processes of this kind all need to start at some finite time in the past because otherwise the variance would become infinite.

Definition 3.16 (Finite Linear Time Series Model) The k-dimensional stochastic process $\{Y_t\}$ is said to follow a finite linear time series model if

$$Y_t = \sum_{j=0}^{t-1} \Psi_j A_{t-j} + h_t \alpha_1, \quad t = 1, 2, \ldots, \tag{3.29}$$

where $\{A_t\} \sim \text{WN}(0, \Sigma)$, the Ψ_j matrices have dimension $k \times l$, h_t is a deterministic $k \times s$ matrix, and α_1 is an s-dimensional stochastic vector that defines the initial conditions. The vector α_1 is specified as

$$\alpha_1 = A\delta + x,$$

where A is a nonstochastic matrix, x is a stochastic vector with a known distribution, and δ is a stochastic vector with an undefined distribution that models the uncertainty with respect to the initial conditions.

In a finite linear time series model, it is usually assumed that the vectors x and δ are orthogonal and that α_1 is orthogonal to the $\{A_t\}$ sequence. However, we will not make any distributional assumptions on the initial vector, α_1, in this section. The vector δ comes into play when there is some unspecified part in α_1, otherwise it is zero.

In almost all practical applications the innovations, A_t, have the same dimension as the observations, Y_t, and thus the Ψ_j matrices have dimension $k \times k$. The fact that nonstationary linear time series models may need many parameters to be identified leads to the search for some structure. Thus, in the previous example of a random walk the model can be expressed as the stochastic difference equation $Y_t = Y_{t-1} + A_t$, $\{A_t\} \sim \text{WN}(0, \sigma^2)$, and the representation is $Y_t = \sum_{j=0}^{t-1} A_{t-j} + (Y_1 - A_1)$, where $\Psi_j = 1$, $h_t = 1$ and $\alpha_1 = \delta + x$, $\delta = Y_1$ and $x = -A_1$. The difference between this model and the VARMA models described in the previous section lies in the presence of unit roots in the autoregressive part of the difference equation. One generalization of VARMA models to the nonstationary situation are **VARMA models with unit roots**. These models are known in the univariate case as **ARIMA models**, where the I stands for integrated. The random walk model is said to be integrated because Y_t is obtained by integrating (i.e., summing) the differenced process $A_t = Y_t - Y_{t-1}$.

In VARMA models with unit roots, the autoregressive operator has to eliminate the dependence on the initial conditions and, at the same time, reveal the structure in the Ψ_j coefficients. That is, the following relations have to take place

$$(I + \Phi_1 z + \cdots + \Phi_p z^p)(I + \Psi_1 z + \Psi_2 z^2 + \cdots) = I + \Theta_1 z + \cdots + \Theta_q z^q \quad (3.30)$$

$$(I + \Phi_1 z + \cdots + \Phi_p z^p)h_t = 0, \; t > r = \max\{p, q\}, \quad (3.31)$$

for some polynomial matrices $\Phi(z) = I + \Phi_1 z + \cdots + \Phi_p z^p$ and $\Theta(z) = I + \Theta_1 z + \cdots + \Theta_q z^q$, where the polynomial $\det[\Phi(z)]$ has some roots of unit modulus. VARMA models with unit roots will be described in more detail later in the book.

In the rest of this section, all VARMA models and innovations state space models will be assumed to start at some finite time in the past.

The concepts of McMillan degree and Kronecker indices can be extended to finite linear time series models. To this end, suppose that $\{Y_t\}$ follows the model (3.29) and define the **augmented Hankel matrix** of order t, H_t^a, by

$$H_t^a = \begin{bmatrix} K_1 & K_2 & K_3 & \cdots & K_t \\ K_2 & K_3 & K_4 & \cdots & K_{t+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ K_t & K_{t+1} & K_{t+2} & \cdots & K_{2t-1} \end{bmatrix}, \quad (3.32)$$

where $K_t = [\Psi_t, h_t]$. We will make the following assumption.

Assumption 3.2 *There exists a positive integer, r, such that the augmented Hankel matrices, H_t^a, of the stochastic process (3.29) have finite rank for all $t > r$.*

Remark 3.21 The structure of the augmented Hankel matrices, H_t^a, and the fact that they have finite rank for $t > r$ implies that if $k_t(i,j)$, $t > r$, is the jth row in the ith block or rows of H_t^a and $k_t(i,j)$ depends linearly on $k_t(i_1,j_1), \ldots, k_t(i_m,j_m)$ with $i_s < i$, $s = 1, \ldots, m$, then the row $k_{t+1}(i+1,j)$ depends linearly on $k_{t+1}(i_1+1,j_1), \ldots, k_{t+1}(i_m+1,j_m)$. Therefore, the McMillan degree and the Kronecker indices of finite linear time series models can be defined as in the case of VARMA and innovations state space models. In fact, it will be shown in Chap. 5 that the initial vector, α_1, can be chosen to have minimal dimension equal to the McMillan degree. ◇

Example 3.18 Let the scalar ARMA model

$$(1 + \phi_1 B + \phi_2 B^2)Y_t = (1 + \theta_1 B + \theta_2 B^2)A_t,$$

where the roots of the autoregressive polynomial can be anywhere in the complex plane. One way to put this model into finite linear time series model form is as follows. Define

$$\alpha_1 = \begin{bmatrix} Y_1 - A_1 \\ Y_2 - A_2 - \Psi_1 A_1 \end{bmatrix},$$

where $\Psi(z) = \sum_{j=0}^{\infty} \Psi_j z^j = (1+\phi_1 z+\phi_2 z^2)^{-1}(1+\theta_1 z+\theta_2 z^2)$, and let $h_1 = (1, 0)$, $h_2 = (0, 1)$, and $h_t = -\phi_1 h_{t-1} - \phi_2 h_{t-2}$ for $t > 2$. Then, it is not difficult to verify that $Y_t = \sum_{j=0}^{t-1} \Psi_j A_{t-j} + h_t \alpha_1$, $t = 1, 2, \ldots$.

If $1+\phi_1 z+\phi_2 z^2 = (1+\phi z)(1-z)$ with $|\phi| < 1$, then the process $U_t = (1-B)Y_t$ is stationary. Assuming that Y_1 has an unspecified distribution and is orthogonal to $\{A_t\}$, and letting $\delta = Y_1$, the vector α_1 can be written as $\alpha_1 = A\delta + x$ with

$$A = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad x = \begin{bmatrix} -A_1 \\ U_2 - A_2 - \Psi_1 A_1 \end{bmatrix}.$$

◇

The following theorem states the equivalence between VARMA models and finite linear time series models.

Theorem 3.8 *Given the innovations process* $\{A_t\} \sim (0, \Sigma)$, $\Sigma > 0$, *the k-dimensional process* $\{Y_t\}$ *follows a VARMA model (3.8) with initial conditions* $\{Y_1, \ldots, Y_r\}$, $r = \max\{p, q\}$, *if, and only if, it follows a finite linear time series model (3.29) with the same initial conditions and* $\dim(\alpha_1) = kr$ *that satisfies*

$$\sum_{j=0}^{r} \Phi_j [\Psi_{t-j}, h_{t-j}] = [0, 0], \quad t > r.$$

Proof We can assume in the proof without loss of generality that $p = q = r$. Suppose that $\{Y_t\}$ follows the VARMA model (3.8) with initial conditions $\{Y_1, \ldots, Y_r\}$ and define

$$Y_{t,P} = \sum_{j=0}^{t-1} \Psi_j A_{t-j}, \quad t = 1, 2, \ldots,$$

where the Ψ_j are given by $\sum_{i=0}^{\infty} \Psi_i z^i = \Phi^{-1}(z)\Theta(z)$. Then, $\{Y_{t,P}\}$ is a solution of (3.8) with initial conditions $\{Y_{1,P}, \ldots, Y_{r,P}\}$. Let $\alpha_1 = [(Y_1 - Y_{1,P})', \ldots, (Y_r - Y_{r,P})']'$, $h_t = (0, \ldots, I, \ldots, 0)$, where the I is in the tth position for $t = 1, \ldots, r$, and $h_t = -\Phi_1 h_{t-1} - \cdots - \Phi_r h_{t-r}$ for $t > r$. Then, $\Phi(B)h_t = 0$ for $t > r$ by definition of h_t and $\{Y_t\} = \{Y_{t,P} + h_t \alpha_1\}$ is the solution of (3.8) with initial conditions $\{Y_1, \ldots, Y_r\}$ because

$$\Phi(B)(Y_{t,P} + h_t \alpha_1) = \Phi(B) Y_{t,P} = \Theta(B) A_t, \quad t > r.$$

Conversely, define $\sum_{j=0}^{i} \Phi_j \Psi_{i-j} = \Theta_i$, $i = 0, 1, \ldots, r$. Then,

$$\Phi(B) Y_t = \Phi(B) \left[\sum_{j=0}^{t-1} \Psi_j A_{t-j} + h_t \alpha_1 \right]$$

$$= \Phi(B)\left[I + \Psi_1 B + \cdots + \Psi_{t-1}B^{t-1}\right]A_t$$
$$= \Theta(B)A_t, \quad t > r,$$

and thus $\{Y_t\}$ is a solution of (3.8) if the same initial conditions, $\{Y_1, \ldots, Y_r\}$, are selected. □

The next theorem establishes the equivalence between finite linear time series models and innovations state space models.

Theorem 3.9 *Given the innovations process $\{A_t\} \sim (0, \Sigma)$, $\Sigma > 0$, the k-dimensional process $\{Y_t\}$ follows an Akaike's innovations state space model, where F and K are as in (3.11) and $H = [I, 0, \ldots, 0]$, with initial conditions $\{Y_1, \ldots, Y_r\}$ if, and only if, it follows a finite linear time series model (3.29) with the same initial conditions and dim$(\alpha_1) = kr$ that satisfies*

$$\sum_{j=0}^{r} \Phi_j[\Psi_{t-j}, h_{t-j}] = [0, 0], \quad t > r.$$

Proof Suppose that $\{Y_t\}$ follows an innovations state space model (3.13) and (3.14), where F and K are as in (3.11) and $H = [I, 0, \ldots, 0]$, with initial conditions $\{Y_1, \ldots, Y_r\}$ and stack the observations to get

$$Y_{t:t+r} = O_{r+1}x_t + \hat{\Psi}_{r+1}A_{t:t+r}, \tag{3.33}$$

where $Y_{t:t+r} = (Y'_t, \ldots, Y'_{t+r})'$, $A_{t:t+r} = (A'_t, \ldots, A'_{t+r})'$, $O_{r+1} = [H', F'H', \ldots, F'^{r}H']'$, and

$$\hat{\Psi}_{r+1} = \begin{bmatrix} I & & & \\ HK & I & & \\ \vdots & \vdots & \ddots & I \\ HF^{r-1}K & \cdots & \cdots\, HK & I \end{bmatrix}. \tag{3.34}$$

It is not difficult to verify that $HF^i = (0, \ldots, I, \ldots, 0)$, $i = 0, 1, \ldots, r-1$, where the I is in the $(i+1)$th position, and this in turn implies that $HF^iK = \Psi_{i+1}$, $i = 0, 1, \ldots, r-1$, in (3.34). Thus, if we premultiply (3.33) by $\hat{\Phi} = [\Phi_r, \Phi_{r-1}, \ldots, \Phi_1, I]$, we get

$$Y_t + \Phi_1 Y_{t-1} + \cdots + \Phi_r Y_{t-r} = A_t + \Theta_1 A_{t-1} + \cdots + \Theta_r A_{t-r}, \quad t > r,$$

where the Θ_i, $i = 1, \ldots, r$, are given by the product $\hat{\Phi}\hat{\Psi}_{r+1}$ and $\hat{\Phi}O_{r+1} = 0$. To see this last equality, consider that, by Proposition 3.10, $\det[\Phi(z)] = \det(I - Fz)$, where $\Phi(z) = I + \Phi_1 z + \cdots + \Phi_r z^r$. Then, the equality follows from the Cayley–Hamilton theorem. Applying Theorem 3.8, we see that $\{Y_t\}$ follows a finite linear time series model with the desired properties.

Conversely, if $\{Y_t\}$ follows a finite linear time series model with the properties stated in the theorem, then, by Theorem 3.8, it follows a VARMA model. Since this VARMA model can be put into Akaike's state space form with F, K, and H as stated in the theorem and initial conditions $\{Y_1, \ldots, Y_r\}$, the theorem is proved. □

Remark 3.22 The previous two theorems have shown the equivalence between finite linear time series models under Assumption 3.2, innovations state space models and VARMA models. However, the dimensions of all these models are not always minimal. In Chap. 5, we will prove similar theorems for the minimal case. ◇

3.10 Covariance Generating Function and Spectrum

In Sect. 1.4 we showed that, given a sequence of observations $\{Y_1, \ldots, Y_n\}$, the evaluation of the sequence of innovations $\{E_1, \ldots, E_n\}$ is equivalent to the triangular factorization of the covariance matrix $\text{Var}(Y)$, where $Y = (Y_1', \ldots, Y_n')'$. In fact, letting $E = (E_1', \ldots, E_n')'$ be the vector of innovations, $V = \text{Var}(Y)$ and $D = \text{diag}(D_1, \ldots, D_n)$, the factorization (1.8) of V can be written as $V = LDL'$, where L is the lower triangular matrix with unit diagonal entries given by 1.13. Letting $W = L^{-1}$, the decomposition (1.12) becomes $V^{-1} = W'D^{-1}W$. The relation that links Y with E is

$$Y = LE, \qquad E = WY.$$

One important fact to note about the L and D matrices is that their entries are time-variant. However, when $\{Y_t\}$ is a stationary process with origin in the infinitely remote past, the transformation from $(Y_t', Y_{t-1}', \ldots)'$ to $(E_t', E_{t-1}', \ldots,)'$ and vice versa turns out to be time-invariant. This was the setting of the seminal studies of Wold (1938), Kolmogorov (1939, 1941), and Wiener (1949), and some of their results will be introduced in this section.

It turns out that there are important differences in the analysis between univariate and multivariate processes. For this reason, we shall first focus in this section on univariate stationary processes. Later in the section, we will address the difficulties in the multivariate case. Some of these difficulties will be resolved later in the book using state space models.

3.10.1 Covariance Generating Function

Consider a zero-mean univariate stationary process $\{Y_t : t \in \mathbb{Z}\}$, with covariance function $\gamma_Y(k) = (Y_{t+k}Y_t)$, $k = 0, 1, \ldots$, and assume it has the linear time series model representation $Y_t = \sum_{j=0}^{\infty} \Psi_j A_{t-j}$, where $\{A_t\} \sim \text{WN}(0, \sigma^2)$, $\Psi_0 = 1$ and $\sum_{j=0}^{\infty} |\Psi_j|^2 < \infty$. We will first show that the computation of the innovations process $\{A_t\}$ can be reduced to the equivalent problem of computing the so-called canonical factorization of the **covariance generating function** of $\{Y_t\}$, defined by

$$G_Y(z) = \sum_{k=-\infty}^{\infty} \gamma_Y(k) z^k.$$

Then, we will investigate the conditions under which such a factorization exists.

If we consider infinite vectors and matrices, the relation $Y_t = \sum_{j=0}^{\infty} \Psi_j A_{t-j}$ can be expressed as

$$\begin{bmatrix} \vdots \\ Y_{t-1} \\ Y_t \\ Y_{t+1} \\ \vdots \end{bmatrix} = \begin{bmatrix} \ddots & \ddots & \ddots & & & \bigcirc \\ \cdots & \Psi_2 & \Psi_1 & 1 & & \\ & \cdots & \Psi_2 & \Psi_1 & 1 & \\ & & \cdots & \Psi_2 & \Psi_1 & 1 \\ & & & \ldots & \ddots & \ddots & \ddots \end{bmatrix} \begin{bmatrix} \vdots \\ A_{t-1} \\ A_t \\ A_{t+1} \\ \vdots \end{bmatrix}$$

or, more compactly, as

$$Y = \Psi A, \tag{3.35}$$

where $Y = (\ldots, Y_{t-1}, Y_t, Y_{t+1}, \ldots)'$, $A = (\ldots, A_{t-1}, A_t, A_{t+1}, \ldots)'$ and Ψ is the infinite lower triangular matrix with unit diagonal entries

$$\Psi = \begin{bmatrix} \ddots & \ddots & \ddots & & & \bigcirc \\ \cdots & \Psi_2 & \Psi_1 & 1 & & \\ & \cdots & \Psi_2 & \Psi_1 & 1 & \\ & & \cdots & \Psi_2 & \Psi_1 & 1 \\ & & & \ldots & \ddots & \ddots & \ddots \end{bmatrix}. \tag{3.36}$$

It is not difficult to verify that, by Proposition 3.6, the infinite matrix $\Gamma_Y = \text{Var}(Y)$ of covariances $\gamma_Y(h) = E(Y_{t+h}Y_t)$ satisfies

$$\Gamma_Y = \Psi \Sigma \Psi', \tag{3.37}$$

where Ψ is given by (3.36),

$$\Gamma_Y = \begin{bmatrix} \ddots & \ddots & \ddots & \cdots & & & \\ \cdots & \gamma_Y(1) & \gamma_Y(0) & \gamma_Y(1) & \cdots & & \\ & \cdots & \gamma_Y(1) & \gamma_Y(0) & \gamma_Y(1) & \cdots & \\ & & \cdots & \gamma_Y(1) & \gamma_Y(0) & \gamma_Y(1) & \cdots \\ & & & \cdots & \ddots & \ddots & \ddots \end{bmatrix},$$

and

$$\Sigma = \begin{bmatrix} \ddots & & & & \bigcirc \\ & \sigma^2 & & & \\ & & \sigma^2 & & \\ & & & \sigma^2 & \\ \bigcirc & & & & \ddots \end{bmatrix}.$$

It is convenient to describe the infinite relations (3.35) and (3.37) by means of generating functions. Given a doubly infinity vector $l = (\ldots, l_{t-1}, l_t, l_{t+1}, \ldots)'$, its **generating function** is defined as $l(z) = \sum_{i=-\infty}^{\infty} l_i z^i$. The generating function of a Toeplitz matrix

$$T = \begin{bmatrix} \ddots & \ddots & \ddots & \cdots & & & \\ \cdots & t_1 & t_0 & t_{-1} & \cdots & & \\ & \cdots & t_1 & t_0 & t_{-1} & \cdots & \\ & & \cdots & t_1 & t_0 & t_{-1} & \cdots \\ & & & \cdots & \ddots & \ddots & \ddots \end{bmatrix}, \tag{3.38}$$

is defined by $T(z) = \sum_{i=-\infty}^{\infty} t_i z^i$.

To obtain the desired relation between generating functions, we introduce the doubly infinite row vector

$$\lambda(z) = \left[\ldots, z^{-2}, z^{-1}, 1, z, z^1, z^2, \ldots\right].$$

Premultiplying the Toeplitz matrix (3.38) by $\lambda(z)$ yields

$$\lambda(z)T = T(z)\left[\ldots, z^{-2}, z^{-1}, 1, z, z^1, z^2, \ldots\right] = T(z)\lambda(z)$$

and, thus, $\lambda(z)$ is a left eigenvector of T with eigenvalue $T(z)$.

Using this result, it is not difficult to verify that, premultiplying (3.35) by $\lambda(z)$, the following relation is obtained

$$Y(z) = \Psi(z)A(z), \tag{3.39}$$

where $\Psi(z) = \sum_{j=0}^{\infty} \Psi_j z^j$ and $Y(z) = \sum_{j=-\infty}^{\infty} Y_j z^j$ and $A(z) = \sum_{j=-\infty}^{\infty} A_j z^j$ are the generating functions of Y and A. In a similar way, premultiplying (3.37) by $\lambda(z)$ yields

$$G_Y(z) = \Psi(z)\sigma^2\Psi(z^{-1}), \tag{3.40}$$

where $G_Y(z)$ is the covariance generating function of $\{Y_t\}$. Equation (3.40) provides the required factorization of the covariance generating function $G_Y(z)$. The innovations A_t can be obtained by inverting (3.39) to get

$$A(z) = \Psi^{-1}(z)Y(z), \tag{3.41}$$

assuming the inverse function $\Psi^{-1}(z)$ exists.

Suppose a function $G(z)$ of the form $G(z) = \sum_{k=-\infty}^{\infty} \gamma_k z^k$, where the sequence $\{\gamma_k\}$ is a covariance sequence of a stationary process. This means that $\gamma(h) = \gamma(-h)$ for all $h \in \mathbb{Z}$, and that the matrix $(\gamma(i-j))_{i,j=1}^{n}$ is nonnegative definite for all positive integer n. The following theorem gives a necessary and sufficient condition for a factorization of $G(z)$ of the form (3.40) to exist.

Theorem 3.10 *Given $G(z) = \sum_{k=-\infty}^{\infty} \gamma_k z^k$, where the sequence $\{\gamma_k\}$ is a scalar covariance sequence of a stationary process, there exists a unique positive number σ^2 and a unique function $\Psi(z)$ satisfying $G(z) = \Psi(z)\sigma^2\Psi(z^{-1})$ with the following properties:*

i) $\Psi(z)$ *and* $\Psi^{-1}(z)$ *are analytic in* $D = \{z \in \mathbb{C} : |z| < 1\}$
ii) $\sum_{i=0}^{\infty} |\Psi_i|^2 < \infty$, *where* $\Psi(z) = \sum_{i=0}^{\infty} \Psi_i z^i$ *and* $\Psi_0 = 1$,

if, and only if, $f(x) = G(e^{-ix})$, $x \in [-\pi, \pi]$, *is a nonnegative function that is Lebesgue-integrable and satisfies the so-called Paley–Wiener condition*

$$\int_{-\pi}^{\pi} \ln[f(x)] > -\infty. \tag{3.42}$$

In this case, the following Kolmogorov–Szegö formula for σ^2 holds

$$\sigma^2 = e^{\frac{1}{2\pi}\int_{-\pi}^{\pi} \ln[f(x)]dx}. \tag{3.43}$$

Proof See Doob (1953), Grenander & Rosenblatt (1957), or Gikhman & Skorokhod (1969). □

The following theorem shows that the stationary processes $\{Y_t\}$ that have a linear time series model representation $Y_t = \sum_{j=0}^{\infty} \Psi_j A_{t-j}$ with $\{A_t\} \sim \mathrm{WN}(0, \sigma^2)$, $\Psi_0 = 1$ and $\sum_{j=0}^{\infty} |\Psi_j|^2 < \infty$ are precisely the stationary processes whose covariance generating function admits a factorization (3.40) with the two properties of the previous theorem.

Theorem 3.11 *For the scalar stationary process* $\{Y_t\}$ *to have a covariance generating function* $G_Y(z) = \sum_{k=-\infty}^{\infty} \gamma_Y(k)z^k$ *admitting a factorization* $G_Y(z) = \Psi(z)\sigma^2\Psi(z^{-1})$ *with* $\Psi(z)$ *unique and satisfying the two properties of the previous theorem it is necessary and sufficient that* $\{Y_t\}$ *has the linear time series model representation* $Y_t = \sum_{j=0}^{\infty} \Psi_j A_{t-j}$ *with* $\{A_t\} \sim \text{WN}(0, \sigma^2)$, $\Psi(z) = \sum_{j=0}^{\infty} \Psi_j z^j$, $\Psi_0 = 1$ *and* $\sum_{j=0}^{\infty} |\Psi_j|^2 < \infty$.

Proof See Gikhman & Skorokhod (1969). □

Remark 3.23 The previous theorem shows that if the stationary process $\{Y_t\}$ admits a linear time series model representation $Y_t = \sum_{j=0}^{\infty} \Psi_j A_{t-j}$ with $\{A_t\} \sim \text{WN}(0, \sigma^2)$, $\Psi_0 = 1$ and $\sum_{j=0}^{\infty} |\Psi_j|^2 < \infty$, then the function $\Psi(z) = \sum_{j=0}^{\infty} \Psi_j z^j$ is analytic and has an analytic inverse function $\Psi^{-1}(z)$ in the disk $D = \{z \in \mathbb{C} : |z| < 1\}$. However, this does not guarantee that the function $\Psi^{-1}(z)$ is convergent in the unit circle $|z| = 1$, see Problem 3.2. For this reason, it is usually required that the functions $\Psi(z)$ and $\Psi^{-1}(z)$ be analytic in some annulus containing the unit circle, $r^{-1} < |z| < r$ with $r > 1$. For example, these stronger conditions are met if $\Psi(z)$ is a rational function and all the roots of both the numerator and the denominator polynomials are greater than one in modulus. Thus, the stationary and invertible ARMA models satisfy these conditions and, consequently, for these models the innovations $\{A_t\}$ of $\{Y_t\}$ can be obtained using formula (3.41). ◇

Example 3.19 If $\{Y_t\}$ is a univariate ARMA process satisfying

$$\Phi(B)Y_t = \Theta(B)A_t,$$

where $\{A_t\} \sim \text{WN}(0, \sigma^2)$ and the roots of $\Phi(z)$ are all outside of the unit circle, then the covariance generating function is

$$G_Y(z) = \frac{\Theta(z)\Theta(z^{-1})}{\Phi(z)\Phi(z^{-1})}\sigma^2.$$

In this case, the series $G_Y(z) = \sum_{k=-\infty}^{\infty} \gamma_Y(k)z^k$ converges for all z in some annulus containing the unit circle, $r^{-1} < |z| < r$ with $r > 1$, and this implies $\sum_{k=-\infty}^{\infty} |\gamma_Y(k)| < \infty$. This is a consequence of the fact that the series $\sum_{j=0}^{\infty} \Psi_j z^j = \Theta(z)/\Phi(z) = \Psi(z)$ converges in some annulus, $r^{-1} < |z| < r$ with $r > 1$, because the roots of $\Phi(z)$ are all outside the unit circle. Note that it follows from this that $\sum_{j=0}^{\infty} |\Psi_j| < \infty$ and that this condition is stronger than the square summability condition in Theorem 3.11. It is further noted that if the model is invertible, the function $\Psi^{-1}(z) = \Phi(z)/\Theta(z)$ is analytic in some annulus $r^{-1} < |z| < r$ with $r > 1$. ◇

Example 3.20 Let $\{Y_t\}$ follow the MA(2) process

$$Y_t = A_t + \theta_1 A_{t-1} + \theta_2 A_{t-2},$$

where $\{A_t\} \sim \mathrm{WN}(0, \sigma^2)$. Then,

$$\begin{aligned} G_Y(z) &= (1+\theta_1 z+\theta_2 z^2)(1+\theta_1 z^{-1}+\theta_2 z^{-2})\sigma^2 \\ &= \left[(1+\theta_1^2+\theta_2^2)+(\theta_1+\theta_1\theta_2)(z+z^{-1})+\theta_2(z^2+z^{-2})\right]\sigma^2, \end{aligned}$$

from which it follows that

$$\begin{aligned} \gamma_Y(0) &= (1+\theta_1^2+\theta_2^2)\sigma^2 \\ \gamma_Y(\pm 1) &= \theta_1(1+\theta_2)\sigma^2 \\ \gamma_Y(\pm 2) &= \theta_2\sigma^2 \\ \gamma_Y(k) &= 0, \qquad |k|>2. \end{aligned}$$

Example 3.21 Let $\{Y_t\}$ follow the AR(1) process

$$Y_t + \phi Y_{t-1} = A_t,$$

where $\{A_t\} \sim \mathrm{WN}(0, \sigma^2)$. Then,

$$G_Y(z) = \frac{\sigma^2}{(1+\phi z)(1+\phi z^{-1})},$$

and to find the autocovariances we will perform a partial fraction expansion. However, we have to be careful with the powers of z^{-1}. One way to proceed is as follows (see Lemma 7.1 in Sect. 7.1.2 for more details).

$$\begin{aligned} G_Y(z) &= \frac{z^{-1}}{z^{-1}}\left[\frac{z\sigma^2}{(1+\phi z)(z+\phi)}\right] = \frac{z^{-1}}{z^{-1}}\left[\frac{A}{1+\phi z}+\frac{B}{z+\phi}\right] \\ &= \frac{z^{-1}}{z^{-1}}\left[\frac{\sigma^2}{(1-\phi^2)(1+\phi z)}+\frac{-\phi\sigma^2}{(1-\phi^2)(z+\phi)}\right] \\ &= \frac{\sigma^2}{(1-\phi^2)(1+\phi z)}+z^{-1}\frac{-\phi\sigma^2}{(1-\phi^2)(1+\phi z^{-1})}. \end{aligned}$$

The first term to the right of the last equality is equal to $\gamma(0)+\sum_{j=1}^{\infty}\gamma(j)z^j$. Since $1/(1+\phi z) = 1+\sum_{j=1}^{\infty}(-\phi)^j z^j$, we obtain

$$\begin{aligned} \gamma(0) &= \frac{\sigma^2}{1-\phi^2} \\ \gamma(1) &= \frac{-\phi\sigma^2}{1-\phi^2} \\ \gamma(k) &= (-\phi)^k\gamma(0), \qquad k=\pm 1, \pm 2, \ldots \end{aligned}$$

◇

3.10.2 Spectrum

As in the previous section, consider a zero-mean univariate stationary process $\{Y_t : t \in \mathbb{Z}\}$ with covariance function $\gamma_Y(k) = (Y_{t+k}Y_t)$, $k \in \mathbb{Z}$, and assume it has the linear time series model representation $Y_t = \sum_{j=0}^{\infty} \Psi_j A_{t-j}$, where $\{A_t\} \sim \text{WN}(0, \sigma^2)$, $\Psi_0 = 1$ and $\sum_{j=0}^{\infty} |\Psi_j|^2 < \infty$. Then, by Theorem 3.11, the covariance generating function $G_Y(z)$ admits the factorization $G_Y(z) = \Psi(z)\sigma^2\Psi(z^{-1})$, where $\Psi(z) = \sum_{j=0}^{\infty} \Psi_j z^j$, and the function $G_Y(e^{-ix})$ is nonnegative in $x \in [\pi, \pi]$.

The function $f_Y(x) = G_Y(e^{-ix})/(2\pi)$ plays a fundamental role in the frequency domain analysis of time series.

Proposition 3.12 *The function $f_Y(x)$ is nonnegative and integrable in $x \in [-\pi, \pi]$.*

Proof Consider the space $\bar{L}_2$ of square integrable functions defined in $[-\pi, \pi]$. As is well known, in this space the set $\{e^{ikx}/\sqrt{2\pi} : k \in \mathbb{Z}\}$ constitutes an orthonormal basis. If we define the functions $g(x) = \sum_{k=0}^{\infty} \Psi_k e^{-ikx} = \Psi(e^{-ix})$ and $g_h(x) = e^{ihx}g(x)$ and consider in $\bar{L}_2$ the inner product defined by $< F, G >= \int_{-\pi}^{\pi} F\bar{G}$, where $F, G \in \bar{L}_2$ and the bar denotes complex conjugation, using an argument similar to that in the proof of Proposition 3.6, it is not difficult to verify that $g, g_h \in L_2$ and

$$\gamma_Y(h) = \sum_{j=0}^{\infty} \Psi_{j+h}\sigma^2\Psi_j = \frac{\sigma^2}{2\pi}\int_{-\pi}^{\pi} e^{ihx}|g(x)|^2 dx. \tag{3.44}$$

Thus, the function

$$f_Y(x) = \frac{1}{2\pi}G_Y(e^{-ix}) = \frac{\sigma^2}{2\pi}|\Psi(e^{-ix})|^2$$

is nonnegative and integrable in $[-\pi, \pi]$. □

The function $f_Y(x)$ is called the **spectrum** or the **spectral density function** of the process $\{Y_t : t \in \mathbb{Z}\}$. From (3.44), it follows that

$$\gamma_Y(h) = \int_{-\pi}^{\pi} e^{ihx} f_Y(x)dx. \tag{3.45}$$

In particular, replacing h by 0 in (3.45), we get

$$\gamma_Y(0) = \int_{-\pi}^{\pi} f_Y(x)dx, \tag{3.46}$$

and thus the area corresponding to the spectrum in a small frequency interval can be interpreted as the proportion of the variance that can be attributable to those frequencies, considering that the series can be seen as a sum of random oscillatory components associated with different frequencies in the interval $[-\pi, \pi]$. This interpretation is basic in the frequency domain approach to time series analysis.

Example 3.22 If $\{Y_t\}$ is a univariate ARMA process satisfying

$$\Phi(B)Y_t = \Theta(B)A_t,$$

where $\{A_t\} \sim \text{WN}(0, \sigma^2)$ and the roots of $\Phi(z)$ are all outside the unit circle, then the spectrum is

$$f_Y(x) = \frac{\sigma^2}{2\pi} \frac{|\Theta(e^{-ix})|^2}{|\Phi(e^{-ix})|^2}.$$

If the process is invertible, $f_Y(x) > 0$ for all $x \in [-\pi, \pi]$. ◇

Example 3.23 Let $\{Y_t\}$ follow the MA(1) process

$$Y_t = A_t + \theta A_{t-1},$$

where $\{A_t\} \sim \text{WN}(0, \sigma^2)$. Then,

$$f_Y(x) = \frac{\sigma^2}{2\pi}|1 + \theta e^{-ix}|^2 = \frac{\sigma^2}{2\pi}(1 + 2\theta\cos(x) + \theta^2), \qquad x \in [-\pi, \pi].$$

◇

Example 3.24 Let $\{Y_t\}$ follow the AR(1) process

$$Y_t + \phi Y_{t-1} = A_t,$$

where $\{A_t\} \sim \text{WN}(0, \sigma^2)$. Then,

$$f_Y(x) = \frac{\sigma^2}{2\pi}|1 + \phi e^{-ix}|^{-2} = \frac{\sigma^2}{2\pi}(1 + 2\phi\cos(x) + \phi^2)^{-1}, \qquad x \in [-\pi, \pi].$$

◇

3.10.3 Multivariate Processes

Consider a zero-mean multivariate stationary process $\{Y_t : t \in \mathbb{Z}\}$ with covariance function $\gamma_Y(k) = (Y_{t+k}Y_t'), k = 0, 1, \ldots$, and assume it has the linear time series model representation $Y_t = \sum_{j=0}^{\infty} \Psi_j A_{t-j}$, where $\{A_t\} \sim \text{WN}(0, \Sigma)$ and $\sum_{j=0}^{\infty} \|\Psi_j\|^2 < \infty$. As in the univariate case, the covariance generating function of $\{Y_t\}$ is defined by

$$G_Y(z) = \sum_{k=-\infty}^{\infty} \gamma_Y(k) z^k.$$

Using an argument similar to that of the scalar case, it is obtained that

$$Y(z) = \Psi(z)A(z), \qquad G_Y(z) = \Psi(z)\Sigma\Psi'(z^{-1}), \tag{3.47}$$

where $\Psi(z) = \sum_{j=0}^{\infty} \Psi_j z^j$.

Suppose a function $G(z)$ of the form $G(z) = \sum_{k=-\infty}^{\infty} \gamma_k z^k$ such that the sequence $\{\gamma_k\}$ is a covariance sequence of a multivariate stationary process, defined by

i) $\gamma(-h) = \gamma(h)', \qquad h \in \mathbb{Z}$
ii) the matrix $(\gamma(i-j))_{i,j=1}^{n}$ is nonnegative definite for all positive integer n.

As in the univariate case, the question arises as to the existence of a factorization $G(z) = L(z)SL'(z^{-1})$, where S is a positive definite matrix and $L(z) = \sum_{j=0}^{\infty} L_j z^j$ is an analytic matrix function in $D = \{z \in \mathbb{C} : |z| < 1\}$ such that $\sum_{j=0}^{\infty} \|L_j\|^2 < \infty$. As mentioned earlier, this problem is considerably more difficult for general processes $\{Y_t\}$ with $\dim(Y_t) = k > 1$ because, as shown by Rozanov (1967), the matrices $L(z)$ and S can be of dimensions $k \times r$ and $r \times r$ with $r < k$.

In the full rank case, $k = r$, the matrices $L(z)$ and S of the factorization $G(z) = L(z)SL'(z^{-1})$ are square and the theory is relatively easier than for general multivariate processes. The following theorem is a generalization of Theorem 3.10 to the multivariate full rank case.

Theorem 3.12 *Given $G(z) = \sum_{k=-\infty}^{\infty} \gamma_k z^k$, where the sequence $\{\gamma_k\}$ is a multivariate covariance sequence of a stationary process, there exist a unique positive definite matrix Σ and a unique square matrix function $\Psi(z)$ satisfying $G(z) = \Psi(z)\Sigma\Psi'(z^{-1})$ with the following properties:*

i) $\Psi(z)$ and $\Psi^{-1}(z)$ are analytic in $D = \{z \in \mathbb{C} : |z| < 1\}$
ii) $\sum_{i=0}^{\infty} \|\Psi_i\|^2 < \infty$, where $\Psi(z) = \sum_{i=0}^{\infty} \Psi_i z^i$ and $\Psi_0 = I$,

if, and only if, $f(x) = G(e^{-ix})$, $x \in [-\pi, \pi]$, is an almost everywhere positive definite matrix function with Lebesgue-integrable components that satisfies the generalized Paley–Wiener condition

$$\int_{-\pi}^{\pi} \ln \det [f(x)] > -\infty.$$

In this case, the following Kolmogorov–Szegö formula holds

$$\det(\Sigma) = e^{\frac{1}{2\pi}\int_{-\pi}^{\pi} \ln \det[f(x)]}.$$

Proof See Rozanov (1967). □

It turns out that the stationary processes $\{Y_t\}$ that have a linear time series model representation $Y_t = \sum_{j=0}^{\infty} \Psi_j A_{t-j}$ with $\{A_t\} \sim \text{WN}(0, \Sigma)$, $\Psi_0 = I$ and $\sum_{j=0}^{\infty} \|\Psi_j\|^2 < \infty$ are precisely the stationary nondeterministic processes of full rank whose covariance generating function admit a factorization with the two properties of the previous theorem.

Theorem 3.13 *For the stationary nondeterministic process of full rank $\{Y_t\}$ to have a covariance generating function $G_Y(z) = \sum_{k=-\infty}^{\infty} \gamma_Y(k) z^k$ admitting a factorization $G_Y(z) = \Psi(z)\Sigma\Psi'(z^{-1})$ with Σ unique positive definite and $\Psi(z)$ square, unique and satisfying the two properties of the previous theorem it is necessary and sufficient that $\{Y_t\}$ has the linear time series model representation $Y_t = \sum_{j=0}^{\infty} \Psi_j A_{t-j}$ with $\{A_t\} \sim$ WN$(0, \Sigma)$, $\Psi(z) = \sum_{j=0}^{\infty} \Psi_j z^j$, $\Psi_0 = I$ and $\sum_{j=0}^{\infty} \|\Psi_j\|^2 < \infty$.*

Proof See Rozanov (1967). □

Remark 3.24 For the less than full rank case, Theorem 3.12 still holds, but replacing the factorization $G(z) = \Psi(z)\Sigma\Psi'(z^{-1})$ with $G(z) = \overline{\Psi}(z)\overline{\Psi}'(z^{-1})$, where $\overline{\Psi}(z) = \sum_{j=0}^{\infty} \overline{\Psi}_j z^j$ is a non unique $k \times r$ matrix with $0 < r < k$ and $\overline{\Psi}_0$ has rank r, and replacing the condition $\int_{-\pi}^{\pi} \ln \det [f(x)] > -\infty$ with $\int_{-\pi}^{\pi} \ln \det [F(x)] > -\infty$, where $\det F(x)$ is a principal minor of $f(x)$ of order r which is different from zero almost everywhere.

For a vector stationary nondeterministic process $\{Y_t\}$ that is not of full rank, Theorem 3.13 also holds, but replacing the factorization $G(z) = \Psi(z)\Sigma\Psi'(z^{-1})$ with $G(z) = \overline{\Psi}(z)\overline{\Psi}'(z^{-1})$, where $\overline{\Psi}(z) = \sum_{j=0}^{\infty} \overline{\Psi}_j z^j$ is a not unique $k \times r$ matrix with $0 < r < k$ and $\overline{\Psi}_0$ has rank r, and replacing the linear time series model representation of that theorem with the linear time series model representation $Y_t = \sum_{j=0}^{\infty} \overline{\Psi}_j \overline{A}_{t-j}$, where $\{\overline{A}_t\} \sim \text{WN}(0, I)$, $\overline{\Psi}(z) = \sum_{j=0}^{\infty} \overline{\Psi}_j z^j$, and $\sum_{j=0}^{\infty} \|\overline{\Psi}_j\|^2 < \infty$. Proofs of these results can be found in Rozanov (1967). ◇

Remark 3.25 It is to be noted that if Y_t is a vector stationary process that admits the linear time series model representation $Y_t = \sum_{j=0}^{\infty} \overline{\Psi}_j \overline{A}_{t-j}$, where $\{\overline{A}_t\} \sim \text{WN}(0, I)$, $\overline{\Psi}(z) = \sum_{j=0}^{\infty} \overline{\Psi}_j z^j$ is a $k \times r$ matrix with $0 < r < k$, $\overline{\Psi}_0$ has rank r, and $\sum_{j=0}^{\infty} \|\overline{\Psi}_j\|^2 < \infty$, then it is possible to put $\overline{\Psi}_0 = \Psi_0 L$, where L is a submatrix of $\overline{\Psi}_0$ of rank r and $\Psi_0 = \overline{\Psi}_0 L^{-1}$. Thus, defining $\Psi_j = \overline{\Psi}_j L^{-1}, j = 1, 2, \ldots$ and $A_t = L\overline{A}_t$, $t \in \mathbb{Z}$, the linear time series model representation $Y_t = \sum_{j=0}^{\infty} \Psi_j A_{t-j}$ is obtained, where $\{A_t\} \sim \text{WN}(0, \Sigma)$, $\Sigma = LL'$, and $\sum_{j=0}^{\infty} \|\Psi_j\|^2 < \infty$. This implies that the covariance generating function $G_Y(z)$ of $\{Y_t\}$ can be factorized as $G_Y(z) = \Psi(z)\Sigma\Psi'(z^{-1})$, where $\Psi(z) = \sum_{j=0}^{\infty} \Psi_j z^j$. For this reason, we will always assume in what follows that all factorizations of covariance generating functions of stationary nondeterministic processes $\{Y_t\}$ are of the form $G_Y(z) = \Psi(z)\Sigma\Psi'(z^{-1})$.

When Σ has dimension $r < k$, it is always possible to have a linear time series model representation of the form $Y_t = \sum_{j=0}^{\infty} \Theta_j E_{t-j}$ in which E_t has dimension k and $\Theta_0 \Omega = \Omega = \Omega \Theta_0'$, where $\text{Var}(E_t) = \Omega$. To see this, define $\Theta_i = \Psi_i \left(\Psi_0' \Psi_0\right)^{-1} \Psi_0'$, $i = 0, 1, 2, \ldots$, $E_t = \Psi_0 A_t$ and $\Omega = \text{Var}(E_t) = \Psi_0 \Sigma \Psi_0'$. Then, $\Psi_i A_{t-i} = \Theta_i E_{t-i}$

and $\Theta_0\Omega = \Omega = \Omega\Theta_0'$. Note that in this case $G_Y(z) = \Theta(z)\Omega\Theta'(z^{-1})$, where $\Theta(z) = \Theta_0 + \Theta_1 z + \Theta_2 z^2 + \cdots$. ◇

Example 3.25 If $\{Y_t\}$ is a multivariate ARMA process satisfying

$$\Phi(B)Y_t = \Theta(B)A_t,$$

where $\{A_t\} \sim \text{WN}(0, \Sigma)$ and the roots of $\det[\Phi(z)]$ are all outside of the unit circle, by Theorem 3.7, the process $\{Y_t\}$ is causal. Thus, the covariance generating function is

$$G_Y(z) = \Phi^{-1}(z)\Theta(z)\Sigma\Theta'(z^{-1})\Phi'^{-1}(z^{-1}).$$

As in the univariate case, the matrix series $G_Y(z) = \sum_{k=-\infty}^{\infty} \gamma_Y(k)z^k$ converges for all z in some annulus containing the unit circle, $r^{-1} < |z| < r$ with $r > 1$, and this implies $\sum_{k=-\infty}^{\infty} \|\gamma_Y(k)\| < \infty$. This is a consequence of the series $\sum_{j=0}^{\infty} \Psi_j z^j = \Phi^{-1}(z)\Theta(z) = \Psi(z)$ being convergent in some annulus, $r^{-1} < |z| < r$ with $r > 1$, because the roots of $\det[\Phi(z)]$ are all outside the unit circle. Note that, as in the univariate case, it follows from this that $\sum_{j=0}^{\infty} \|\Psi_j\| < \infty$ and that, if the model is invertible, the function $\Psi^{-1}(z) = \Theta^{-1}(z)\Phi(z)$ is analytic in some annulus $r^{-1} < |z| < r$ with $r > 1$. ◇

Example 3.26 Suppose a vector process $\{Y_t\}$ that follows a VARMA(p, q) model (3.8) such that the roots of $\det[\Phi(z)]$ are all greater than one in modulus. Then, considering Akaike's innovations state space representation, where F and K are as in (3.11) and $H = [I, 0, \ldots, 0]$, the eigenvalues of F are all of modulus less than one and, as shown in the proof of Theorem 3.7, Eq. (3.13) has a unique stationary solution given by $x_t = \sum_{j=0}^{\infty} F^j K A_{t-1-j}$, where $\{A_t\} \sim \text{WN}(0, \Sigma)$. Defining $\Psi_j = F^j K$, $j = 0, 1, 2, \ldots$, we can write $x_t = \Psi(B)A_{t-1}$, where $\Psi(z) = \sum_{j=0}^{\infty} \Psi_j z^j$ and B is the backshift operator, $BA_t = A_{t-1}$. Since the series $\Psi(z)$ is convergent in an annulus, $r^{-1} < |z| < r$ with $r > 1$, the filter $(I - BF)^{-1}K$ is well defined and stable and we have $\Psi(z) = (I - zF)^{-1}K$. Thus, the factorization of the covariance generating function, $G_x(z)$, of $\{x_t\}$ is

$$G_x(z) = (I - zF)^{-1}K\Sigma K'(I - z^{-1}F')^{-1}.$$

Taking $x_t = \Psi(B)A_{t-1}$ into (3.13) and considering the decomposition (3.35), if $Y(z)$ and $A(z)$ are the generating functions of $\{Y_t\}$ and $\{A_t\}$, it is obtained that $Y(z) = (I + zH\Psi(z))A(z)$. Since $\{A_t\}$ are the innovations, according to (3.47), the covariance generating function $G_Y(z)$ of $\{Y_t\}$ is

$$G_Y(z) = \left[I + zH(I - zF)^{-1}K\right]\Sigma\left[I + z^{-1}K'(I - z^{-1}F')^{-1}H'\right].$$

Note that the factorization of $G_Y(z)$ is of full rank whereas that of $G_x(z)$ is not. ◇

For a zero-mean vector stationary nondeterministic process $\{Y_t : t \in \mathbb{Z}\}$, with covariance function $\gamma_Y(k) = (Y_{t+k}Y_t)$, $k \in \mathbb{Z}$, and admitting the linear time series model representation $Y_t = \sum_{j=0}^{\infty} \Psi_j A_{t-j}$, where $\{A_t\} \sim \text{WN}(0, \Sigma)$ and $\sum_{j=0}^{\infty} \|\Psi_j\|^2 < \infty$, the **spectral density matrix**, or spectrum, is defined by

$$f_Y(x) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma_Y(k) e^{-ikx} = \frac{1}{2\pi} \Psi(e^{-ikx}) \Sigma \Psi'(e^{ikx}), \qquad x \in [-\pi, \pi],$$

where $\Psi(z) = \sum_{j=0}^{\infty} \Psi_j z^j$.

Using arguments similar to those of the univariate case, it can be shown that the matrix function $f_Y(x)$ is nonnegative-definite for almost all $x \in [-\pi, \pi]$ and that the components of $f_Y(x)$ are integrable in $[-\pi, \pi]$. Given $f_Y(x)$, the covariances $\gamma_Y(h)$ of $\{Y_t\}$ can be obtained from the formula

$$\gamma_Y(h) = \int_{-\pi}^{\pi} e^{ihx} f_Y(x) dx, \qquad h \in \mathbb{Z}. \tag{3.48}$$

If Y_t has dimension two, $Y_t = (Y_{t1}, Y_{t2})'$, then

$$f_Y(x) = \begin{bmatrix} f_{11}(x) & f_{12}(x) \\ f_{21}(x) & f_{22}(x) \end{bmatrix},$$

and $f_{11}(x)$ and $f_{22}(x)$ are the spectrums of the univariate series Y_{t1} and Y_{t2}. The functions $f_{ij}(x)$, $i \neq j$, $i,j = 1, 2$, are referred to as the **cross spectrums** of Y_{ti} and Y_{tj}. Since the autocovariance function $\gamma_Y(h)$ is in general not symmetric, the cross spectrum $f_{ij}(x)$ is typically complex valued and the following relation

$$f_{21}(x) = \overline{f_{12}(x)}$$

holds. The **coherence** function $k_{ij}(x)$ is defined as

$$k_{ij}(x) = \frac{f_{ij}(x)}{\sqrt{f_{11}(x) f_{22}(x)}}.$$

By the Cauchy–Schwarz inequality, the following inequality

$$0 \leq |k_{ij}(x)|^2 \leq 1, \qquad x \in [-\pi, \pi],$$

holds. A value of $|k_{ij}(x)|^2$ close to one indicates a strong linear relationship between the components of Y_{ti} and Y_{tj} associated with the frequency x. The cross spectrum and the coherence can be easily generalized to a multivariate process of dimension greater than two.

Example 3.27 If $\{Y_t\}$ is a VARMA process satisfying

$$\Phi(B)Y_t = \Theta(B)A_t,$$

where $\{A_t\} \sim \text{WN}(0, \Sigma)$ and the roots of $\det[\Phi(z)]$ are all outside of the unit circle, by Theorem 3.7, the process $\{Y_t\}$ is causal and the spectrum is

$$f_Y(z) = \frac{1}{2\pi}\Phi^{-1}(e^{-ix})\Theta(e^{-ix})\Sigma\Theta'(e^{ix})\Phi^{-1'}(e^{ix}).$$

◇

Proposition 3.13 *Suppose that the scalar stationary process $\{Y_t\}$ can be represented in state space innovations form*

$$x_{t+1} = Fx_t + KA_t$$
$$Y_t = Hx_t + A_t,$$

where $\{A_t\} \sim \text{WN}(0, \Sigma)$. Then, the following conditions are equivalent

i) the ARMA process followed by $\{Y_t\}$ is invertible
ii) the matrix $F - KH$ has eigenvalues with modulus less than one
iii) the spectrum density of $\{Y_t\}$, $f_Y(z)$, satisfies $f_Y(z) > 0$ almost surely for $x \in [-\pi, \pi]$.

Proof By Propositions 3.11 and 3.10, i) and ii) are equivalent. Given that the spectrum of $\{Y_t\}$ is

$$f_Y(z) = \frac{1}{2\pi}\Phi^{-1}(e^{-ix})\Theta(e^{-ix})\Sigma\Theta'(e^{ix})\Phi^{-1'}(e^{ix}).$$

in terms of an equivalent ARMA model, it is clear that the spectrum satisfies $f_Y(z) > 0$ if, and only if, the polynomial $\det[\Theta(z)]$ has no roots equal to one. □

3.10.4 Linear Operations on Stationary Processes

In many applications, stochastic processes are considered that are the output of stable linear time invariant filters applied to some input stationary processes. By Corollary 3.1, these output processes are stationary and the question arises as to the relation between the covariance generating functions of the input and output processes. First, however, given two vector stationary processes, $\{Y_t\}$ and $\{X_t\}$, we define their **cross covariance generating function** as

$$G_{YX}(z) = \sum_{k=-\infty}^{\infty} \gamma_{YX}(k)z^k,$$

where $\gamma_{YX}(k) = \text{Cov}(Y_{t+k}, X_t)$.

Proposition 3.14 *Let $\{Y_t\}$ be the vector stationary process that is obtained by passing a zero-mean vector stationary process $\{X_t\}$ through a stable linear time invariant filter with transfer function $\Psi(z)$. Then the following relations hold*

$$G_Y(z) = \Psi(z)G_X(z)\Psi'(z^{-1}), \qquad G_{YX}(z) = \Psi(z)G_X(z),$$

and

$$f_Y(x) = \Psi(e^{-ix})f_X(x)\Psi'(e^{ix}), \qquad f_{YX}(x) = \Psi(e^{-ix})f_X(e^{-ix}).$$

Finally, if $\{Z_t\}$ is jointly stationary with $\{Y_t, X_t\}$ as just defined, then

$$G_{ZY}(z) = G_{ZX}(z)\Psi'(z^{-1}).$$

Proof By Corollary 3.1, the process $\{Y_t\}$ is stationary and its covariance function is given by

$$\gamma_Y(h) = \sum_{i=-\infty}^{\infty}\sum_{j=-\infty}^{\infty} \Psi_i \gamma_X(h-i+j)\Psi_j'.$$

From this it follows that

$$\begin{aligned} G_Y(z) &= \sum_{h=-\infty}^{\infty}\sum_{i=-\infty}^{\infty}\sum_{j=-\infty}^{\infty} \Psi_i \gamma_X(h-i+j)\Psi_j' z^h \\ &= \sum_{l=-\infty}^{\infty}\sum_{i=-\infty}^{\infty}\sum_{j=-\infty}^{\infty} \Psi_i \gamma_X(l)\Psi_j' z^{l+i-j} \\ &= \left(\sum_{i=-\infty}^{\infty} \Psi_i z^i\right)\left(\sum_{l=-\infty}^{\infty} \gamma_X(l) z^l\right)\left(\sum_{j=-\infty}^{\infty} \Psi_j' z^{-j}\right) \\ &= \Psi(z)G_X(z)\Psi'(z^{-1}). \end{aligned}$$

The other formulae can be obtained similarly. □

3.10.5 Computation of the Autocovariance Function of a Stationary VARMA Model

There are several methods to compute the autocovariance function, $\gamma_Y(h)$, of a stationary VARMA model, $\{Y_t\}$. Since VARMA models can be put into state space form, we can consider two types of methods, polynomial and state space methods.

The computation of covariance matrices when the model is in state space form will be considered in Chap. 5. In this section, we will present a polynomial method based on the decomposition of the autocovariance generating function $G_Y(z)$ into the sum of two rational functions from which the computation of the covariance matrices $\gamma_Y(h)$ is easy. First, we will consider the univariate case and later we will consider the multivariate, more demanding, case.

Let $\{Y_t\}$ be a scalar ARMA(p,q) process satisfying

$$\Phi(B)Y_t = \Theta(B)A_t,$$

where $\Phi(z) = 1+\phi_1 z+\cdots+\phi_p z^p$, $\Theta(z) = 1+\theta_1 z+\cdots+\theta_q z^q$, $\{A_t\} \sim \text{WN}(0,\sigma^2)$ and the roots of $\det[\Phi(z)]$ are all outside of the unit circle. By Theorem 3.7, the process $\{Y_t\}$ is causal and the covariance generating function $G_Y(z) = \sum_{j=-\infty}^{\infty} \gamma_Y(j)z^j$ is well defined and equal to

$$G_Y(z) = \frac{\Theta(z)\Theta(z^{-1})}{\Phi(z)\Phi(z^{-1})}\sigma^2.$$

Considering the Wold decomposition of $\{Y_t\}$, it is not difficult to check that the covariances satisfy

$$\begin{array}{ll} \gamma_Y(l) + \sum_{j=1}^{p} \phi_j\gamma_Y(l-j) = \sum_{j=l}^{q} \theta_j\sigma^2\psi_{j-l}, & 0 \le l \le q \\ \gamma_Y(l) + \sum_{j=1}^{p} \phi_j\gamma_Y(l-j) = 0, & l > q, \end{array} \tag{3.49}$$

where $\theta_0 = 1$ and the ψ_j weights can be recursively computed from $\Phi(z)(\sum_{j=0}^{\infty} \psi_j z^j) = \Theta(z)$. This constitutes the first method to compute the covariances.

Another, more efficient, method is the following. It can be called the polynomial method. The covariance generating function, $G_Y(z)$, can be written as

$$G_Y(z) = A(z) + A'(z^{-1}), \tag{3.50}$$

where

$$A(z) = \frac{1}{2}\gamma_Y(0) + \gamma_Y(1)z + \gamma_Y(2)z^2 + \cdots .$$

If we know the function $A(z)$, the covariances can be easily computed. If we premultiply and postmultiply (3.50) by $\Phi(z)$ and $\Phi(z^{-1})$, it is obtained that

$$\Theta(z)\sigma^2\Theta(z^{-1}) = B(z)\Phi(z^{-1}) + \Phi(z)B(z^{-1}), \tag{3.51}$$

where $B(z) = \Phi(z)A(z) = \sum_{j=0}^{\infty} b_j z^j$ satisfies

$$\begin{aligned}
&\tfrac{1}{2}\gamma_Y(0) &&= b_0\\
&\gamma_Y(1) + \tfrac{1}{2}\phi_1\gamma_Y(0) &&= b_1\\
&\gamma_Y(2) + \phi_1\gamma_Y(1) + \tfrac{1}{2}\phi_2\gamma_Y(0) &&= b_2\\
&\cdots &&= \cdots\\
&\gamma_Y(p) + \phi_1\gamma_Y(p-1) + \cdots + \tfrac{1}{2}\phi_p\gamma_Y(0) &&= b_p\\
&\gamma_Y(l) + \phi_1\gamma_Y(l-1) + \cdots + \phi_p\gamma_Y(l-p) &&= b_l,\ l > p\ .
\end{aligned} \tag{3.52}$$

Since the covariances satisfy (3.49), the relation $b_l = 0$ holds for $l > r = \max(p, q)$, and it turns out that $B(z)$ is a polynomial, $B(z) = b_0 + b_1 z + \cdots + b_r z^r$.

Given the coefficients $b_0, b_1, \ldots, b_r$, the covariances can be recursively computed from (3.52). However, if $p < q$, by (3.49) and (3.52), only $b_0, b_1, \ldots, b_p$ need to be computed because $b_{p+1}, \ldots, b_r$ are given by

$$b_l = \sum_{j=l}^{q} \theta_j \sigma^2 \psi_{j-l}, \qquad l = p+1, \ldots, q. \tag{3.53}$$

To compute $b_0, b_1, \ldots, b_p$ from Eq. (3.51), first let

$$\Theta(z)\sigma^2\Theta(z^{-1}) = \alpha_0 + \alpha_1(z + z^{-1}) + \cdots + \alpha_q(z^q + z^{-q}).$$

Then, equating coefficients of z^j for $j = 0, 1, \ldots, r$ in (3.51), the following system of linear equations in $b_0, b_1, \ldots, b_r$, is obtained

$$(T_r + H_r)X_r = A_r, \tag{3.54}$$

where

$$T_r = \begin{bmatrix} 1 & & & \bigcirc \\ \phi_1 & 1 & & \\ \vdots & \vdots & \ddots & \\ \phi_r & \phi_{r-1} & \cdots & 1 \end{bmatrix}, \qquad H_r = \begin{bmatrix} \bigcirc & & & \phi_r \\ & & \phi_r & \phi_{r-1} \\ & & \vdots & \vdots \\ \phi_r & \cdots & \phi_1 & 1 \end{bmatrix},$$

$X_r = (b_r, b_{r-1}, \ldots, b_0)'$, $A_r = (\alpha_r, \alpha_{r-1}, \ldots, \alpha_0)'$, and $\alpha_j = 0$ if $j > q$. If $p < q = r$, we can compute $b_{p+1}, \ldots, b_r$ according to (3.53) and reduce the number of equations in (3.54) so that the following linear system is obtained

$$(T_p + H_p)X_p = A_p, \tag{3.55}$$

where

$$T_p = \begin{bmatrix} 1 & & & \bigcirc \\ \phi_1 & 1 & & \\ \vdots & \vdots & \ddots & \\ \phi_p & \phi_{p-1} & \cdots & 1 \end{bmatrix}, \qquad H_p = \begin{bmatrix} \bigcirc & & & \phi_p \\ & & \phi_p & \phi_{p-1} \\ & & \vdots & \vdots \\ \phi_p & \cdots & \phi_1 & 1 \end{bmatrix},$$

$X_p = (b_p, b_{p-1}, \ldots, b_0)'$, $A_p = (a_p, a_{p-1}, \ldots, a_0)'$, $a_0 = \alpha_0$, $a_i = \alpha_i - \sum_{j=p+1}^{r} \phi_{j-i} b_j$, $i = 1, \ldots, p$, and $\phi_i = 0$ if $i > p$. Note that the system (3.55) corresponds to the polynomial equation

$$\overline{B}(z)\Phi(z^{-1}) + \Phi(z)\overline{B}(z^{-1}) = a_p z^{-p} + \cdots + a_1 z^{-1} + a_0 + a_1 z + \cdots + a_p z^p,$$

where $\overline{B}(z) = b_0 + b_1 z + \cdots + b_p z^p$ is the polynomial $B(z)$ truncated to degree p. Thus, we have reduced the original polynomial equation (3.51) by reducing the degree from $q > p$ to p.

To solve (3.55), first note that T_p is a Toeplitz matrix, H_p is a Hankel matrix and

$$T_p^{-1} = \begin{bmatrix} 1 & & & \bigcirc \\ \xi_1 & 1 & & \\ \vdots & \vdots & \ddots & \\ \xi_p & \xi_{p-1} & \cdots & 1 \end{bmatrix},$$

where the ξ_j weights can be obtained from $\xi(z)\phi(z) = 1$, $\xi(z) = \sum_{j=0}^{\infty} \xi_j z^j$. Then, partition $X_p = (X_1', X_2')'$ and $A_p = (A_1', A_2')'$, where $X_1 = (b_p, \ldots, b_{n+1})'$, $X_2 = (b_n, \ldots, b_0)'$, $A_1 = (a_p, \ldots, a_{n+1})'$, $A_2 = (a_n, \ldots, a_0)'$, $n = [p/2]$, and $[m]$ denotes the integer part of $m \in \mathbb{R}$. If we partition T_p and H_p conforming to the previous partition, we can write (3.55) as

$$\left(\begin{bmatrix} T_{11} & \\ T_{21} & T_{22} \end{bmatrix} + \begin{bmatrix} & H_{12} \\ H_{21} & H_{22} \end{bmatrix} \right) \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix}. \tag{3.56}$$

Partitioning also T_p^{-1} conforming to $X_p = (X_1', X_2')'$,

$$T_p^{-1} = \begin{bmatrix} T^{11} & \\ T^{21} & T^{22} \end{bmatrix},$$

and premultiplying (3.56) by T_p^{-1} yields

$$\left(\begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix} + \begin{bmatrix} 0 & T^{11} H_{12} \\ T^{22} H_{21} & T^{21} H_{12} + T^{22} H_{22} \end{bmatrix} \right) \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} T^{11} A_1 \\ T^{21} A_1 + T^{22} A_2 \end{bmatrix}. \tag{3.57}$$

Letting $D_{12} = T^{11}H_{12}$, $D_{21} = T^{22}H_{21}$, $D_{22} = T^{21}H_{12} + T^{22}H_{22}$, $C_1 = T^{11}A_1$ and $C_2 = T^{21}A_1 + T^{22}A_2$, we can write (3.57) as

$$\left(\begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix} + \begin{bmatrix} 0 & D_{12} \\ D_{21} & D_{22} \end{bmatrix}\right)\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} C_1 \\ C_2 \end{bmatrix}. \tag{3.58}$$

From the first equation in (3.58), we get $X_1 = C_1 - D_{12}X_2$. Replacing X_1 in the second equation of (3.58) with this expression yields

$$(I + D_{22} - D_{21}D_{12})X_2 = C_2 - D_{21}C_1. \tag{3.59}$$

We first compute X_2 in (3.59) and then compute X_1 as $X_1 = C_1 - D_{12}X_2$. This, together with (3.53), allows us to recursively compute the covariances using (3.52).

Example 3.28 (The Autocovariance Function of an MA(q) Process) Assume $\{Y_t\}$ is the MA(q) process

$$Y_t = \Theta(B)A_t,$$

where $\{A_t\} \sim \text{WN}(0, \sigma^2)$ and $\Theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q$ has all its roots outside the unit circle. Then, by (3.49),

$$\begin{aligned} \gamma_Y(l) &= \sigma^2 \textstyle\sum_{j=l}^{q} \theta_j\theta_{j-l}, \quad 0 \le l \le q \\ \gamma_Y(l) &= 0, \qquad\qquad\qquad l > q, \end{aligned}$$

where $\theta_0 = 1$. ◇

Let $\{Y_t\}$ be a k-dimensional stationary VARMA(p, q) process satisfying

$$\Phi(B)Y_t = \Theta(B)A_t,$$

where $\Phi(z) = \Phi_0 + \Phi_1 z + \cdots + \Phi_p z^p$, $\Theta(z) = \Theta_0 + \Theta_1 z + \cdots + \Theta_q z^q$, $\Phi_0 = \Theta_0$ is a lower triangular matrix with ones in the main diagonal, and $\{A_t\} \sim \text{WN}(0, \Sigma)$. Then, the autocovariance generating function $G_Y(z) = \sum_{j=-\infty}^{\infty} \gamma_Y(j)z^j$ is

$$G_Y(z) = \Phi^{-1}(z)\Theta(z)\Sigma\Theta'(z^{-1})\Phi'^{-1}(z^{-1}).$$

We consider this more general form of VARMA(p, q) models to allow for the so-called VARMA echelon forms, that will be described in detail in Chap. 5.

Similarly to the univariate case, the autocovariance matrices satisfy

$$\begin{aligned} \Phi_0\gamma_Y(l) + \textstyle\sum_{j=1}^{p} \Phi_j\gamma_Y(l-j) &= \textstyle\sum_{j=l}^{q} \Theta_j\Sigma\Psi'_{j-l} \quad 0 \le l \le q \\ \Phi_0\gamma_Y(l) + \textstyle\sum_{j=1}^{p} \Phi_j\gamma_Y(l-j) &= 0 \qquad\qquad\qquad l > q, \end{aligned} \tag{3.60}$$

where the Ψ_j weights can be recursively computed from $\Phi(z)(\sum_{j=0}^{\infty}\Psi_j z^j) = \Theta(z)$. Solving these equations constitutes the first method to compute the autocovariance matrices.

We can use the polynomial method as in the univariate case to compute the autocovariance matrices. This method is more efficient than the previous one and is as follows. The autocovariance generating function, $G_Y(z)$, can be written as

$$G_Y(z) = A(z) + A'(z^{-1}) \tag{3.61}$$

where

$$A(z) = \frac{1}{2}A_0 + \gamma_Y(1)z + \gamma_Y(2)z^2 + \cdots,$$

and

$$\gamma_Y(0) = \frac{1}{2}\left(A_0 + A_0'\right). \tag{3.62}$$

The matrix A_0 can be any matrix depending on $k(k+1)/2$ parameters because $\gamma_Y(0)$ is symmetric. If we premultiply and postmultiply (3.61) by $\Phi(z)$ and $\Phi'(z^{-1})$, it is obtained that

$$\Theta(z)\Sigma\Theta'(z^{-1}) = B(z)\Phi'(z^{-1}) + \Phi(z)B'(z^{-1}), \tag{3.63}$$

where $B(z) = \Phi(z)A(z) = \sum_{j=0}^{\infty} B_j z^j$ satisfies

$$\begin{array}{ll}
\frac{1}{2}\Phi_0 A_0 & = B_0 \\
\Phi_0\gamma_Y(1) + \frac{1}{2}\Phi_1 A_0 & = B_1 \\
\Phi_0\gamma_Y(2) + \Phi_1\gamma_Y(1) + \frac{1}{2}\Phi_2 A_0 & = B_2 \\
\cdots & = \cdots \\
\Phi_0\gamma_Y(p) + \Phi_1\gamma_Y(p-1) + \cdots + \frac{1}{2}\Phi_p A_0 & = B_p \\
\Phi_0\gamma_Y(l) + \Phi_1\gamma_Y(l-1) + \cdots + \Phi_p\gamma_Y(l-p) & = B_l,\ l > p\,.
\end{array} \tag{3.64}$$

As in the univariate case, since the autocovariance matrices satisfy (3.60), the relation $B_l = 0$ holds for $l > r = \max(p,q)$, and $B(z)$ is a polynomial matrix, $B(z) = B_0 + B_1 z + \cdots + B_r z^r$, where the B_j coefficients are now $k \times k$ matrices. Since the matrix B_0 depends on $k(k+1)/2$ parameters, we can take B_0 to be symmetric and this will be assumed in the following.

Given the matrix coefficients $B_0, B_1, \ldots, B_r$, first the matrix A_0 and the autocovariance matrices $\gamma_Y(i)$, $i = 1, 2, \ldots$, can be recursively computed from (3.64). Then, $\gamma_Y(0)$ can be obtained using (3.62). However, if $p < q$, by (3.60) and (3.64), only $B_0, B_1, \ldots, B_p$ need to be computed because $B_{p+1}, \ldots, B_r$ are given by

$$B_l = \sum_{j=l}^{q} \Theta_j \Sigma \Psi_{j-l}', \qquad l = p+1, \ldots, q. \tag{3.65}$$

To compute $B_0, B_1, \dots, B_p$ from Eq. (3.63), first let

$$\Theta(z)\Sigma\Theta(z^{-1}) = \Gamma_0 + (\Gamma_{-1}z^{-1} + \Gamma_1 z) + \cdots + (\Gamma_{-q}z^{-q} + \Gamma_q z^q),$$

where $\Gamma_{-j} = \Gamma_j'$. Then, equating coefficients of z^j for $j = 0, 1, \dots, r$ in (3.63) and letting $b_j = \mathrm{vec}(B_j)$, $\mathrm{vec}(B_j') = Wb_j$, $\alpha_j = \mathrm{vec}(\Gamma_j)$, $j = 0, 1, \dots, r$, where $W = (I_k \otimes e_1, I_k \otimes e_2, \dots, I_k \otimes e_k)$ and e_i is a column vector with a one in the ith position and zeros elsewhere, the following system of linear equations in $b_0, b_1, \dots, b_r$, is obtained

$$(T_r + H_r)X_r = A_r, \tag{3.66}$$

where

$$T_r = \begin{bmatrix} (\Phi_0 \otimes I_k) & & & \bigcirc \\ (\Phi_1 \otimes I_k) & (\Phi_0 \otimes I_k) & & \\ \vdots & \vdots & \ddots & \\ (\Phi_r \otimes I_k) & (\Phi_{r-1} \otimes I_k) & \cdots & (\Phi_0 \otimes I_k) \end{bmatrix},$$

$$H_r = \begin{bmatrix} & \bigcirc & & (I_k \otimes \Phi_r)W \\ & & (I_k \otimes \Phi_r)W & (I_k \otimes \Phi_{r-1})W \\ & & \vdots & \vdots \\ (I_k \otimes \Phi_r)W & \cdots & (I_k \otimes \Phi_1)W & (I_k \otimes \Phi_0)W \end{bmatrix},$$

$X_r = (b_r', b_{r-1}', \dots, b_0')'$, $A_r = (\alpha_r', \alpha_{r-1}', \dots, \alpha_0')'$, and $\alpha_j = 0$ if $j > q$. As in the univariate case, if $p < q = r$, we compute $B_{p+1}, \dots, B_r$ according to (3.65) and reduce the number of equations in (3.66) so that the following linear system is obtained

$$(T_p + H_p)X_p = A_p, \tag{3.67}$$

where T_p and H_p are like T_r and H_r but with r replaced with p, $X_p = (b_p', b_{p-1}', \dots, b_0')'$, $A_p = (a_p', a_{p-1}', \dots, a_0')'$, $a_0 = \alpha_0$, $a_i = \alpha_i - \sum_{j=p+1}^{r} \left(I \otimes \Phi_{j-i}\right) b_j$, $i = 1, \dots, p$, and $\Phi_i = 0$ if $i > p$. As in the univariate case, we have reduced the original symmetric polynomial matrix equation (3.63) of degree $q > p$ to the equation

$$\overline{B}(z)\Phi'(z^{-1}) + \Phi(z)\overline{B}'(z^{-1}) = \Omega_p' z^{-p} + \cdots + \Omega_1' z^{-1} + \Omega_0 + \Omega_1 z + \cdots + \Omega_p z^p,$$

of degree p, where $\mathrm{vec}(\Omega_i) = a_i$, $i = 0, 1, \dots, p$, and $\overline{B}(z) = B_0 + B_1 z + \cdots + B_p z^p$.

To solve (3.67), note that T_p is a block Toeplitz matrix, H_p is a block Hankel matrix,

$$T_p^{-1} = \begin{bmatrix} (\Xi_0 \otimes I_k) & & & \bigcirc \\ (\Xi_1 \otimes I_k) & (\Xi_0 \otimes I_k) & & \\ \vdots & \vdots & \ddots & \\ (\Xi_p \otimes I_k) & (\Xi_{p-1} \otimes I_k) & \cdots & (\Xi_0 \otimes I_k) \end{bmatrix},$$

where the Ξ_j weights can be obtained from $\left(\sum_{j=0}^{\infty} \Xi_j z^j\right) \Phi(z) = I$, and partition $X_p = (X_1', X_2')'$ and $A_p = (A_1', A_2')'$, where $X_1 = (b_p', \ldots, b_{n+1}')'$, $X_2 = (b_n', \ldots, b_0')'$, $A_1 = (a_p', \ldots, a_{n+1}')'$, $A_2 = (a_n', \ldots, a_0')'$, and $n = [p/2]$. If we partition T_p and H_p conforming to the previous partition, we can write (3.67) as

$$\left(\begin{bmatrix} T_{11} & \\ T_{21} & T_{22} \end{bmatrix} + \begin{bmatrix} & H_{12} \\ H_{21} & H_{22} \end{bmatrix} \right) \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix},$$

which is analogous to (3.56). We can proceed then as in the univariate case, partitioning also T_p^{-1} conforming to $X_p = (X_1', X_2')'$,

$$T_p^{-1} = \begin{bmatrix} T^{11} & \\ T^{21} & T^{22} \end{bmatrix},$$

and solving for X_1 and X_2 to get $B_0, B_1, \ldots, B_p$. As mentioned earlier, if $r = q > p$, $B_{p+1}, \ldots, B_r$ are obtained using (3.65). The autocovariance matrices can be recursively computed using (3.64) and (3.62). It is to be noted that, since B_0 is symmetric, $b_0 = \text{vec}(B_0)$ can be expressed in terms of $\text{vec}(B_0)$ in the previous equations.

Finally, it is to be further noted that it is possible to proceed as suggested by Mittnik (1993), converting first (3.67) into a block Toeplitz system and applying then the method of Akaike (1973). However, the gain seems marginal because the matrix T_p^{-1} is available without further computation.

3.10.6 Algorithms for the Factorization of a Scalar Covariance Generating Function

In this section, we will address the problem of finding the factorization of a scalar covariance generating function $G(z) = \sum_{k=-\infty}^{\infty} \gamma(k) z^k$ of a stationary process $\{Y_t\}$. That is, given the sequence $\{\gamma_k\}$ of covariances of $\{Y_t\}$, we are interested in finding a unique positive number σ^2 and a unique function $\Psi(z) = \sum_{i=0}^{\infty} \Psi_i z^i$, $\Psi_0 = 1$, satisfying $G(z) = \Psi(z)\sigma^2\Psi(z^{-1})$. According to Theorems 3.10 and 3.11, the factorization exists and is unique for a linear time series model.

The first method we will consider is the **cepstral method**, which goes back to Kolmogorov (1939). This method is based on the development of the function

$$g(z) = \ln[G(z)] = \sum_{j=-\infty}^{\infty} c_j z^j, \qquad z \in \mathbb{C},$$

as a Laurent series in an annulus containing the unit circle. Because the Paley–Wiener condition of Theorem 3.10 is satisfied, the function $G(z)$ cannot be zero in a set of positive Lebesgue measure in the unit circle. That is, it cannot be zero too often in the unit circle. In addition, it can be shown (Gikhman & Skorokhod, 1969, pp. 219–222; Pourahmadi 2001, p. 68) that

$$\sigma\Psi(z) = \exp\left(\frac{c_0}{2} + \sum_{j=1}^{\infty} c_j z^j\right)$$

and

$$\sigma^2 = e^{c_0} = 2\pi \exp\left(\frac{1}{2\pi}\int_{-\pi}^{\pi} \ln[f(x)]dx\right), \qquad \Psi(z) = \exp\left(\sum_{j=1}^{\infty} c_j z^j\right), \tag{3.68}$$

where $f(x) = G(e^{-ix})/(2\pi)$, $x \in [-\pi, \pi]$, is the spectrum of $\{Y_t\}$, and $\Psi(z) = \sum_{j=0}^{\infty} \Psi_j z^j$ is such that $\sum_{j=0}^{\infty} |\Psi_j|^2 < \infty$. The Ψ_j coefficients can be obtained recursively in terms of the c_j as follows. Differentiating in (3.68) with respect to z gives

$$\Psi'(z) = (c_1 + 2c_2 z + 3c_3 z^2 + \cdots)\Psi(z)$$

and equating coefficients on both sides yields

$$k\Psi_k = c_1\Psi_{k-1} + 2c_2\Psi_{k-2} + \cdots + kc_k\Psi_0 = \sum_{j=1}^{k} jc_j\Psi_{k-j}, \quad k = 1, 2, \ldots \tag{3.69}$$

In the cepstral method, first, the coefficients c_j are approximately computed using the discrete Fourier transform and its inverse, described later in this section. Then, the relations (3.69) are used to compute the Ψ_j weights. The procedure will be illustrated at the end of this section for the particular case of a moving average process.

Other algorithms that can be used to compute the factor $\Psi(z)$ and σ^2 are the **Bauer algorithm**, the **Schur algorithm**, and the **Levinson–Durbin algorithm**. The Bauer algorithm consists of iterating in the innovations algorithm until $\Theta_{t,t-k} \to \Psi_k$ and $D_t \to \sigma^2$ as $t \to \infty$ in the innovations representation $Y_t = \Theta_{t,1}E_1 + \cdots +$

$\Theta_{t,t-1}E_{t-1} + E_t$. This can be shown to happen if the process $\{Y_t\}$ follows a linear time series model, $Y_t = A_t + \sum_{k=1}^{\infty} \Psi_k A_{t-k}$, $\{A_t\} \sim \mathrm{WN}(0, \sigma^2)$. Assuming that the autoregressive representation $Y_t = Y_t + \sum_{k=1}^{\infty} \Pi_k Y_{t-k}$ of $\{Y_t\}$ exists and computing the coefficients ϕ_{jk} in $Y_t = \phi_{j1}Y_{t-1} + \cdots + \phi_{jj}Y_{t-j} + E_{t,j}$ and $D_j = \mathrm{Var}(E_{t,j})$ for $j = 1, 2, \ldots$ using the Levinson–Durbin algorithm, it can be shown that $\phi_{j,k} \to \Pi_k$ and $D_j \to \sigma^2$ as $j \to \infty$. Thus, the Levinson–Durbin algorithm gives in fact the inverse factor $\Pi(z) = \Psi^{-1}(z)$. The Levinson–Durbin algorithm will be described in the next section. The Schur algorithm is more involved and will not be described in this book. The interested reader can consult Sayed & Kailath (2001) for more details on covariance function factorization methods.

In the rest of the section we will consider ARMA models because, as mentioned earlier, every linear model can be approximated by an ARMA model with any degree of accuracy. Let $\{Y_t\}$ be a univariate ARMA process satisfying

$$\Phi(B)Y_t = \Theta(B)A_t,$$

where $\{A_t\} \sim \mathrm{WN}(0, \sigma^2)$ and the roots of $\Phi(z)$ are all outside of the unit circle. Then the covariance generating function is

$$G(z) = \Psi(z)\sigma^2\Psi(z^{-1}) = \frac{\Theta(z)\Theta(z^{-1})}{\Phi(z)\Phi(z^{-1})}\sigma^2$$

and the series $\sum_{j=0}^{\infty} \Psi_j z^j = \Theta(z)/\Phi(z) = \Psi(z)$ converges in some annulus, $r^{-1} < |z| < r$ with $r > 1$, because the roots of $\Phi(z)$ are all outside the unit circle. This implies that the series $G(z) = \sum_{k=-\infty}^{\infty} \gamma(k)z^k$ converges for all z in some annulus containing the unit circle, $r^{-1} < |z| < r$ with $r > 1$, and $\sum_{k=-\infty}^{\infty} |\gamma(k)| < \infty$.

Assume that we only know the covariance generating function, $G(z)$, given by the ratio of two symmetric Laurent polynomials, say

$$G(z) = \frac{Q(z)}{P(z)},$$

where $P(z)$ and $Q(z)$ have the special form

$$Q(z) = \sum_{j=-q}^{q} Q_j z^j, \qquad P(z) = \sum_{j=-p}^{p} P_j z^j, \qquad Q_j = Q_{-j}, \qquad P_j = P_{-j},$$

p is the degree of $\Phi(z)$, and q is the degree of $\Theta(z)$, and we want to compute the factor $\Psi(z) = \Theta(z)/\Phi(z)$ and σ^2. By stationarity, $P(e^{-ix}) > 0$ for all $x \in [-\pi, \pi]$. If we assume invertibility, we also have that $Q(e^{-ix}) > 0$ for all $x \in [-\pi, \pi]$. So, for stationary and invertible ARMA models, if we can factor $P(z)$ and $Q(z)$ into

$$P(z) = \Phi(z)\Phi(z^{-1}), \qquad Q(z) = \Theta(z)\sigma^2\Theta(z^{-1}),$$

we have the desired factorization of $G(z)$. For this reason, we shall focus in the rest of the section on the factorization of symmetric Laurent polynomials $Q(z)$ that are strictly positive on the unit circle and that correspond to invertible moving average processes.

We will start be describing some of the existing methods to factorize a symmetric Laurent polynomial

$$Q(z) = \sum_{j=-q}^{q} Q_j z^j, \qquad Q_j = Q_{-j}.$$

One obvious procedure consists of transforming $Q(z)$, which is a polynomial in z and z^{-1}, into a polynomial in z only by defining $Q^*(z) = z^q Q(z)$. The new polynomial, $Q^*(z)$, has degree $2q$ and its roots can be grouped in pairs that are symmetric with respect to the unit circle. Thus, if we find the roots of $Q^*(z)$, r_i, $i = 1, \ldots, 2q$, and, without loss of generality, we assume $|r_i| \leq |r_{i+1}|$, $i = 1, \ldots, 2q-1$, we can compute the factor $\Theta(z)$ and the variance σ^2 as

$$\Theta(z) = (1 - r_1 z)(1 - r_2 z)\cdots(1 - r_q z), \qquad \sigma^2 = (-1)^q Q_q/(r_1 \cdots r_q),$$

where $|r_i| < 1, i = 1, \ldots, q$.

A more sophisticated procedure can be obtained if we take the symmetry of the roots into account. Define the new variable $y = z + z^{-1}$. Then, due to the symmetry, the polynomial $Q(z)$ can be rewritten in terms of y as

$$Q(z) = R(y) = R_0 + R_1 y + \cdots R_q y^q,$$

which is an ordinary polynomial of degree q. Once we have found the roots of $R(y)$, s_i, $i = 1, \ldots, q$, the roots of $Q(z)$ are obtained from $s_i = z + z^{-1}$. This is a second degree equation that gives two roots of $Q(z)$ that are symmetric with respect to the unit circle for each root of $R(y)$. The coefficients of $Q(z)$ and $R(y)$ are linearly related by the equation

$$(R_0, R_1, \ldots, R_q) = (Q_0, Q_1, \ldots, Q_q)S,$$

where S is a $q \times q$ lower triangular sparse matrix whose entries are given by the recursion

$$S_{i,j} = S_{i-1,j-1} - S_{i-2,j}, \qquad i = 4, \ldots, q, \qquad j = 1, 2, \ldots, i,$$

$S_{i,j} = 0$ if $i \leq 0$ or $j \leq 0$, initialized with $S_{1,1} = 1$, $S_{2,2} = 1$, $S_{3,1} = -2$, and $S_{3,3} = 1$. Note that S has ones in the main diagonal and that the odd lower subdiagonals are zero. This method may be subject to numerical difficulties if there are multiple roots or there is a great number of roots.

As mentioned earlier in this section, another method that can be used is the cepstral method, that we now describe. But first, we need a definition.

Definition 3.17 Given a vector of real or complex numbers $y = (y_1, y_2, \dots, y_n)'$, the **discrete Fourier transform** (DFT) of y is the vector Y defined by

$$Y = \phi_n y,$$

where

$$\phi_n = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & \xi_n & \xi_n^2 & \cdots & \xi_n^{n-1} \\ 1 & \xi_n^2 & \xi_n^4 & \cdots & \xi_n^{n-2} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & \xi_n^{n-1} & \xi_n^{n-2} & \cdots & \xi_n \end{bmatrix}, \qquad \phi_{jk,n} = \xi_n^{jk}, \qquad \xi_n = e^{-\frac{2\pi i}{n}}.$$

The following lemma shows how to invert the matrix ϕ_n. The proof is simple and is omitted.

Lemma 3.2 *The matrix* $\frac{1}{\sqrt{n}}\phi_n$ *is symmetric and unitary. Thus,*

$$\left(\frac{1}{\sqrt{n}}\phi_n\right)^{-1} = \frac{1}{\sqrt{n}}\phi_n^* = \frac{1}{\sqrt{n}}\overline{\phi}_n, \qquad \phi_n^{-1} = \frac{1}{n}\phi^* = \frac{1}{n}\overline{\phi}_n.$$

By the previous lemma, the **inverse of the DFT** (IDFT) is given by

$$y = \phi_n^{-1} Y = \frac{1}{n}\overline{\phi}_n Y.$$

For numerical computation of the DFT, the efficient recursive **fast Fourier transform** (FFT) algorithm was developed by Cooley & Tukey (1965).

In the cepstral method, given $Q(z) = \sum_{j=-q}^{q} Q_j z^j$, $Q_j = Q_{-j}$, we want to compute the coefficients c_j in $\ln[Q(z)] = \sum_{j=-\infty}^{\infty} c_j z^j$. To this end, we can use the FFT algorithm as follows. Assume that we know the values of $\ln[Q(z)]$ in a finite number of Fourier interpolating points, ω^{-k}, $k = -M, \dots, 0, \dots, M$, $\omega = \exp[2\pi i/(2M+1)]$, in the unit circle and we make the approximation

$$C_k = \sum_{j=-M}^{M} c_j \omega^{-kj} \simeq \ln[Q(\omega^{-k})].$$

Then, $C = (C_0, C_1, \dots, C_{2M})'$ is the DFT of $c = (c_0, c_1, \dots, c_M, c_M, \dots, c_1)'$,

$$C = \phi_{2M+1} c,$$

and we can recover c by computing the IDFT of C. That is,

$$c = \phi_{2M+1}^{-1} C.$$

The previous considerations suggest that we may compute an approximation to the c_j coefficients using the following algorithm.

Step 1 Using the FFT algorithm, compute the DFT of the vector

$$\begin{aligned} b &= (Q_0, Q_1, \ldots, Q_q, 0, 0, \ldots, 0, Q_q, \ldots, Q_1)' \\ &= (b_0, b_1, \ldots, b_{2M})', \end{aligned}$$

where M is approximately 50–100 times larger than q. That is, find B such that

$$B = \phi_{2M+1} b = (B_0, B_1, \ldots, B_{2M})'$$

and

$$B_k = Q(\omega^{-k}) = \sum_{j=-q}^{q} Q_j \omega^{-kj}, \qquad \omega = \exp[2\pi i/(2M+1)].$$

Step 2 Compute the vector

$$C = (\ln(B_0), \ln(B_1), \ldots, \ln(B_{2M}))'.$$

Step 3 Using the FFT algorithm, compute the IDFT of the vector C. That is, find $c = (c_0, c_1, \ldots, c_M, c_M, \ldots, c_1)'$ such that

$$c = \phi_{2M+1}^{-1} C.$$

Step 4 Select the first n components of c, $(c_0, c_1, \ldots, c_n)'$, where $n \geq q$ and use (3.68) and the recursions (3.69) to compute σ^2 and the Ψ_j coefficients, $j = 1, 2, \ldots, n$. The coefficients Ψ_j with $j > q$ can be used to check whether the approximation is good by comparing them to a small number since they should be zero.

Remark 3.26 In practice it is convenient to select $2M + 1 = 2^R$ in the previous algorithm so that the FFT transform can be applied more efficiently because the number of points to which it is applied is a power of two. The accuracy of the approximation in the computation of the c_j coefficients depends on the number 2^R of interpolation points. ◇

3.10.7 Algorithms for the Factorization of a Multivariate Covariance Generating Function

There are basically two methods for the factorization of a covariance generating function in the vector case. One is the method proposed by Tunnicliffe-Wilson (1972), that is valid in principle for general stationary processes. The other is the method based on solving the DARE, that is valid only for stationary processes with time invariant state space structure. This last method will be described in full detail in Chap. 5, where it will be shown that time invariant state space models and VARMA models are equivalent. Therefore, the method based on the DARE can also be used with VARMA models. In this section, we will describe Tunnicliffe-Wilson's method.

Let $\{Y_t\}$ be a stationary nondeterministic vector process and let $G_Y(z) = \sum_{k=-\infty}^{\infty} \gamma_Y(k)z^k$ be its covariance generating function. Assuming that the generalized Paley–Wiener condition of Theorem 3.12 is satisfied, $G_Y(z)$ admits a factorization

$$G_Y(z) = \Psi(z)\Sigma\Psi'(z^{-1}),$$

with Σ unique positive definite and $\Psi(z)$ square, unique and satisfying the two properties of Theorem 3.12. Let $\Sigma = LL'$, with L a lower triangular matrix, be the Cholesky factorization of Σ and define $\overline{\Psi}(z) = \Psi(z)L$. Then, $G_Y(z)$ can be expressed as

$$G_Y(z) = \overline{\Psi}(z)\overline{\Psi}'(z^{-1}),$$

and we seek an iterative sequence, $\{\overline{\Psi}_n(z)\}$, such that each function of the sequence satisfies the two properties of Theorem 3.12 and $\lim_{n\to\infty} \overline{\Psi}_n(z) = \overline{\Psi}(z)$.

To derive the method, assume the approximation $\overline{\Psi}_n(z)$ at iteration n is known. Then, $\overline{\Psi}_{n+1}(z)$ is found by imposing the requirements:

(i) $\overline{\Psi}_{n+1}(z) = \overline{\Psi}_n(z) + \delta_n(z)$, for some small correction term $\delta_n(z)$.
(ii) $\overline{\Psi}_{n+1}(z)\overline{\Psi}'_{n+1}(z^{-1}) \approx G_Y(z)$.

Writing (ii) above as

$$\left[\overline{\Psi}_n(z) + \delta_n(z)\right]\left[\overline{\Psi}'_n(z^{-1}) + \delta'_n(z^{-1})\right] \approx G_Y(z),$$

and neglecting the second order term $\delta_n(z)\delta'_n(z^{-1})$ leads to a recursion relating $\overline{\Psi}_n(z)$, $\overline{\Psi}_{n+1}(z)$ and $G_Y(z)$,

$$\overline{\Psi}_{n+1}(z)\overline{\Psi}'_n(z^{-1}) + \overline{\Psi}_n(z)\overline{\Psi}'_{n+1}(z^{-1}) = G_Y(z) + \overline{\Psi}_n(z)\overline{\Psi}'_n(z^{-1}). \tag{3.70}$$

Tunnicliffe-Wilson (1972) proved that if one starts the recursion (3.70) with an initial polynomial matrix $\overline{\Psi}_0(z)$ that is nonsingular for $|z| \leq 1$, then all successive $\overline{\Psi}_n(z)$ will also be nonsingular for $|z| \leq 1$ and the algorithm will converge to the unique invertible factor $\overline{\Psi}(z)$ such that $G_Y(z) = \overline{\Psi}(z)\overline{\Psi}'(z^{-1})$. Moreover, convergence is quadratic in nature. The first term, $\overline{\Psi}_{n0}$, in the development of $\overline{\Psi}_n(z) = \sum_{i=0}^{\infty} \overline{\Psi}_{ni}z^i$ is constrained to be nonsingular lower triangular in all iterations.

One efficient way to solve (3.70) is to first define $X_n(z) = \overline{\Psi}_n(z) + 2\delta_n(z)$ so that $\overline{\Psi}_{n+1}(z)$ becomes

$$\overline{\Psi}_{n+1}(z) = \frac{1}{2}\left[\overline{\Psi}_n(z) + X_n(z)\right],$$

and then transform (3.70) into

$$X_n(z)\overline{\Psi}_n'(z^{-1}) + \overline{\Psi}_n(z)X_n'(z^{-1}) = 2G_Y(z). \tag{3.71}$$

This polynomial matrix equation is similar to (3.63) and, when $\overline{\Psi}(z)$ is a finite polynomial matrix, that is, $G_Y(z)$ is the covariance generating function of a moving average model, the same technique as that used to compute the autocovariances of a VARMA model can be applied. In this case, a convenient choice for the initial value is $\overline{\Psi}_0 = C$, where C is a constant lower triangular matrix such that $CC' = \gamma_Y(0)$. Note that if $\overline{\Psi}_n(z) = \sum_{i=0}^{q} \overline{\Psi}_{ni}z^i$, then $\overline{\Psi}_{00} = C$ and $\overline{\Psi}_{0i} = 0$, $i = 1, \ldots, q$, and that $\lim_{n\to\infty} X_n(z) = \overline{\Psi}_n(z)$ in (3.70). If $G_Y(z) = \Theta(z)\Sigma\Theta'(z^{-1})$ with $\Theta(z) = I + \Theta_1 z + \cdots + \Theta_q z^q$, once the algorithm has converged to $\overline{\Psi}(z) = \sum_{i=0}^{q} \overline{\Psi}_i z^i$, we set $\Sigma = \overline{\Psi}_0\overline{\Psi}_0'$ and $\Theta_i = \overline{\Psi}_i\overline{\Psi}_0^{-1}$, $i = 1, \ldots, q$.

3.11 Recursive Autoregressive Fitting for Stationary Processes: Partial Autocorrelations

Suppose a zero-mean random vector process $\{Y_t\}$ and consider, as in Sect. 1.8.3, for fixed t and increasing k, the forward and backward innovations (1.50) and (1.51). As mentioned earlier, these are the forward and backward prediction problems, respectively. If the process is stationary, the autocovariance matrices are $E(Y_iY_j') = \gamma_Y(i-j)$ and it turns out that the algorithm of Theorem 1.4 can be simplified because the coefficient matrices in (1.50) and (1.51) do not depend on the index t.

In this section, we will focus on the simplification in the algorithm of Theorem 1.4 when the process $\{Y_t\}$ is stationary. In the univariate case, the algorithm is called **the Levinson–Durbin recursions** Durbin (1960); Levinson (1947) and in the multivariate case it is called **Whittle's algorithm** Whittle (1963a). We will first consider the univariate case and then the multivariate case.

3.11.1 Univariate Processes

Theorem 3.14 (The Levinson–Durbin Recursions) *Suppose a zero-mean scalar stationary process* $\{Y_t\}$ *with covariance generating function* $G_Y(z) = \sum_{j=-\infty}^{\infty} \gamma(j)z^j$ *satisfying the Paley–Wiener condition of Theorem 3.10. For fixed t and increasing k, define the forward innovations* $\{E_{t,k}\}$ *by*

$$\begin{aligned} Y_t &= E^*(Y_t|Y_{t-1},\ldots,Y_{t-k}) + E_{t,k} \\ &= \phi_{k1}Y_{t-1} + \cdots + \phi_{kk}Y_{t-k} + E_{t,k} \end{aligned}$$

and denote $D_k = \text{Var}(E_{t,k})$. *Let* $\phi_{(k)} = (\phi_{k1},\ldots,\phi_{kk})$, $\phi^*_{(k)} = (\phi_{kk},\ldots,\phi_{k1})$ *and* $\Gamma_{(k)} = (\gamma(1),\ldots,\gamma(k))$. *Then, the following recursions hold*

$$\begin{aligned} \phi_{kk} &= \left(\gamma(k) - \Gamma_{(k-1)}\phi^{*\prime}_{(k-1)}\right)/D_{k-1} \\ (\phi_{k1},\ldots,\phi_{k,k-1}) &= \phi_{(k-1)} - \phi_{kk}\phi^*_{(k-1)} \\ D_k &= (1-\phi^2_{kk})D_{k-1}, \end{aligned}$$

initialized with $D_0 = \gamma(0)$ *and* $\phi_{11} = \gamma(1)/\gamma(0)$. *In addition, the coefficient* ϕ_{kk} *is the* ***partial autocorrelation coefficient*** *between* Y_t *and* Y_{t-k-1} *and satisfies* $|\phi_{kk}| < 1$.

Proof First note that, by stationarity, the coefficients of the linear projection $E^*(Y_t|Y_{t-1},\ldots,Y_{t-k})$ are the same as those of $E^*(Y_{t-k-1}|Y_{t-1},\ldots,Y_{t-k})$ but with the order reversed. Then, to obtain the algorithm, let $Z_3 = Y_t$, $Z_2 = Y_{t-k}$ and $Z_1 = (Y'_{t-1},\ldots,Y'_{t-k+1})'$. Applying formula (1.44) yields

$$\begin{aligned} E^*(Z_3|Z_1,Z_2) &= E^*(Z_3|Z_1) + \Pi_{32}(Z_2 - E^*(Z_2|Z_1)) \\ &= \Theta_{31}Z_1 + \Pi_{32}(Z_2 - \Theta_{21}Z_1) \\ &= \Pi_{31}Z_1 + \Pi_{32}Z_2, \end{aligned}$$

where $\Pi_{31} = \Theta_{31} - \Theta_{32}\Theta_{21}$, $\Pi_{32} = \Theta_{32}$ and Θ_{21}, Θ_{31} and Θ_{32} are given by (1.45). The previous relations imply $\Theta_{31} = \phi_{(k-1)}$, $\Theta_{21} = \phi^*_{(k-1)}$, $(\Pi_{31},\Pi_{32}) = \phi_{(k)}$, $\Pi_{31} = (\phi_{k1},\ldots,\phi_{k,k-1})$, and $\Theta_{32} = \phi_{kk}$. Thus,

$$(\phi_{k1},\ldots,\phi_{k,k-1}) = \phi_{(k-1)} - \phi_{kk}\phi^*_{(k-1)}.$$

Since, by stationarity, $D_{k-1} = MSE[E^*(Y_3|Y_1)] = MSE[E^*(Y_2|Y_1)]$, from (1.46), we can write

$$D_k = D_{k-1} - \phi_{kk}D_{k-1}\phi_{kk}.$$

The expression (1.45) for $\Theta_{32} = \phi_{kk}$ yields

$$\phi_{kk} = \left[\gamma(k) - (\gamma(1), \ldots, \gamma(k-1))\phi^{*\prime}_{(k-1)}\right] / D_{k-1}.$$

Finally, formula (1.52) applied in the present context shows that ϕ_{kk} is the partial correlation coefficient between Y_t and Y_{t-k-1}. If $|\phi_{kk}| = 1$ for some k, then $D_k = 0$ and Y_t would be perfectly predictable using $\{Y_{t-1}, \ldots, Y_{t-k}\}$, in contradiction with Theorems 3.2 and 3.11. □

Remark 3.27 Iterating in the Levinson–Durbin recursions, we get the formula

$$D_k = (1 - \phi^2_{kk})(1 - \phi^2_{k-1,k-1}) \cdots (1 - \phi^2_{11})\gamma(0).$$

If we replace Y_2 and Y_1 with Y_t and $(Y_{t-1}, \ldots, Y_{t-k})$ in the definition of the multiple correlation coefficient $r_{2.1}$, given by (1.53), we get $1 - r^2_{2.1} = D_k/\gamma(0)$. Thus, letting R_k be the **multiple correlation coefficient** of Y_t and $(Y_{t-1}, \ldots, Y_{t-k})$, we get the formula

$$1 - R^2_k = (1 - \phi^2_{kk})(1 - \phi^2_{k-1,k-1}) \cdots (1 - \phi^2_{11}).$$

◇

Example 3.29 Suppose the MA(1) model

$$Y_t = A_t + \theta A_{t-1},$$

where $\{A_t\} \sim \mathrm{WN}(0, \sigma^2)$. Then, the Levinson–Durbin recursions are initialized with $D_0 = \sigma^2(1 + \theta^2)$ and $\phi_{11} = \theta/(1 + \theta^2)$. Given that $\Gamma_k = (\theta, 0, \ldots, 0)$ for $k > 1$, the recursions yield $D_1 = \sigma^2\left(1 + \theta^2 + \theta^4\right) / \left(1 + \theta^2\right)$,

$$\phi_{22} = -\frac{\theta^2}{1 + \theta^2 + \theta^4}$$
$$\phi_{21} = \frac{\theta\left(1 + \theta^2\right)}{1 + \theta^2 + \theta^4}$$
$$D_2 = \sigma^2(1 + \theta^2 + \theta^4 + \theta^6)/(1 + \theta^2 + \theta^4).$$

Continuing in this way, it is obtained for $t \geq 2$ that $Y_t = \phi_{k1}Y_{t-1} + \cdots + \phi_{kk}Y_{t-k} + E_{t,k}$, where

$$D_k = \frac{1 + \theta^2 + \cdots + \theta^{2(k+1)}}{1 + \theta^2 + \cdots + \theta^{2k}}\sigma^2, \quad \phi_{kj} = (-1)^{j+1}\theta^j\frac{1 + \theta^2 + \cdots + \theta^{2(k-j)}}{1 + \theta^2 + \cdots + \theta^{2k}}, \quad j = 1, \ldots, k.$$

◇

The Levinson–Durbin recursions allow us to establish the following theorem.

Theorem 3.15 (Equivalence Between Autocorrelations and Partial Autocorrelations) *Suppose a zero-mean scalar stationary process* $\{Y_t\}$ *with covariance sequence* $\{\gamma(i)\}$ *satisfying the Paley–Wiener condition of Theorem 3.10 and let* $\{\rho_i\}$ *and* $\{\phi_{ii}\}$ *be the autocorrelation and partial autocorrelation sequences of* $\{Y_t\}$*. Then, there is a bijection between the two sequences and, therefore, the distribution of* $\{Y_t\}$ *is completely determined by either of the two sequences plus* $\gamma(0)$.

Proof Since $\rho_k = \gamma(k)/\gamma(0)$, given $\gamma(0)$, there is a bijection between the autocorrelations and the autocovariances. That the partial autocorrelations are univocally determined by the autocovariances is clear by the Levinson–Durbin recursions. To see the converse, first use again these recursions to get

$$\gamma(k) = \phi_{kk}D_{k-1} + (\gamma(1), \dots, \gamma(k-1))\phi^{*'}_{(k-1)}.$$

Then, the result follows by induction on k. □

Remark 3.28 The previous theorem shows that we can parameterize a stationary process using $\gamma(0)$ and the partial autocorrelation sequence, $\{\phi_{ii}\}$. This constitutes an unconstrained parametrization simpler than using the autocovariance sequence, $\{\gamma(i)\}$. ◇

3.11.1.1 The Yule–Walker Estimators

For each k, the coefficients $(\phi_{k1}, \dots, \phi_{kk})$ given by the Levinson–Durbin recursions are equal to the so-called **Yule–Walker estimators**. Assuming that the process $\{Y_t\}$ follows an AR(k) model and the autocovariances are known, these estimators are computed by solving the system of linear equations in the unknowns ϕ_{kj}, $j = 1, 2, \dots, k$,

$$\begin{bmatrix} \gamma(0) & \gamma(1) & \cdots & \gamma(k-1) \\ \gamma(1) & \gamma(0) & \cdots & \gamma(k-2) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma(k-1) & \gamma(k-2) & \cdots & \gamma(0) \end{bmatrix} \begin{bmatrix} \phi_{k1} \\ \phi_{k2} \\ \vdots \\ \phi_{kk} \end{bmatrix} = \begin{bmatrix} \gamma(1) \\ \gamma(2) \\ \vdots \\ \gamma(k) \end{bmatrix}, \tag{3.72}$$

$$\gamma(0) - \gamma(1)\phi_{k1} - \cdots - \gamma(k)\phi_{kk} = D_k, \tag{3.73}$$

which are given by (3.49). The Levinson–Durbin recursions constitute a far more efficient approach to computing these coefficients than solving the previous system of equations for each $k = 1, 2, \dots$.

3.11.1.2 Reversing the Levinson–Durbin Recursions

Given the scalar autoregressive model

$$Y_t = a_1Y_{t-1} + a_2Y_{t-2} + \cdots + a_kY_{t-k} + A_t,$$

it may be of interest to compute the first k partial autocorrelation coefficients of the process $\{Y_t\}$. To this end, we can apply the Levinson–Durbin recursions in reversed order, starting with $(\phi_{k1}, \ldots, \phi_{kk}) = (a_1, \ldots, a_k)$. Note first that we can write the Levinson–Durbin recursions in matrix form as

$$\begin{bmatrix} \phi_{k1} \\ \phi_{k2} \\ \vdots \\ \phi_{k,k-1} \end{bmatrix} = (I_{k-1} - \phi_{kk} J_{k-1}) \begin{bmatrix} \phi_{k-1,1} \\ \phi_{k-1,2} \\ \vdots \\ \phi_{k-1,k-1} \end{bmatrix},$$

where J_{k-1} is the $(k-1) \times (k-1)$ reversing matrix defined by

$$J_{k-1} = \begin{bmatrix} 0 & 0 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 1 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \end{bmatrix}.$$

The matrix J_{k-1} is orthogonal and satisfies $J_{k-1}^2 = I_{k-1}$. In addition, it is not difficult to verify that $(I_{k-1} - \phi_{kk} J_{k-1})^{-1} = [1/(1-\phi_{kk}^2)](I_{k-1} + \phi_{kk} J_{k-1})$. Then, we can write

$$\begin{bmatrix} \phi_{k-1,1} \\ \phi_{k-1,2} \\ \vdots \\ \phi_{k-1,k-1} \end{bmatrix} = \frac{1}{1-\phi_{kk}^2}(I_{k-1} + \phi_{kk} J_{k-1}) \begin{bmatrix} \phi_{k1} \\ \phi_{k2} \\ \vdots \\ \phi_{k,k-1} \end{bmatrix}$$

and we can recursively obtain the coefficients of the autoregressive fittings of order $k-1, \ldots, 1$ and thus the partial autocorrelation coefficients, $\phi_{ii}, i = k, \ldots, 1$.

3.11.1.3 Checking the Stability of a Polynomial

An immediate application of the algorithm of the previous section is to checking the stability of a polynomial, $p(z) = 1 + a_1 z + \cdots + a_k z^k$. If the polynomial is stable, we can consider the autoregressive process $Y_t + a_1 Y_{t-1} + \cdots + a_k Y_{t-k} = A_t$, where $\{A_t\}$ is a white noise process, and apply the reversed Levinson–Durbin recursions to obtain the partial autocorrelation coefficients, $\phi_{ii}, i = k, \ldots, 1$. The test then consists of checking that $|\phi_{ii}| < 1$ for all $i = k, k-1, \ldots, 1$. This follows from the easily established equivalences: (1) $\{Y_t\}$ stationary if, and only if, $p(z)$ is stable, and (2) $\{Y_t\}$ stationary if, and only if, the partial autocorrelation coefficients, $\phi_{ii}, i = k, \ldots, 1$, satisfy $|\phi_{ii}| < 1$.

3.11.1.4 Backward Prediction Problem and Lattice Representation

For fixed t and increasing k, define the forward and backward innovations

$$\begin{aligned} E_{t,k} &= Y_t - E^*(Y_t|Y_{t-1},\ldots,Y_{t-k}) \\ &= Y_t - \phi_{k1}Y_{t-1} - \cdots - \phi_{kk}Y_{t-k} \end{aligned} \tag{3.74}$$

$$\begin{aligned} R_{t-1,k} &= Y_{t-k-1} - E^*(Y_{t-k-1}|Y_{t-1},\ldots,Y_{t-k}) \\ &= Y_{t-k-1} - \phi^*_{kk}Y_{t-1} - \cdots - \phi^*_{k1}Y_{t-k}. \end{aligned} \tag{3.75}$$

These are the so-called forward and backward prediction problems in the stationary case. As mentioned in the proof of Theorem 3.14, the coefficients of both problems satisfy the relation $\phi^*_{kj} = \phi_{kj}, j = 1,\ldots,k$, due to stationarity.

Theorem 3.16 (Lattice Representation) *Under the assumptions and with the notation of Theorem 3.14, the forward and backward innovations (3.74) and (3.75) satisfy the following recursions*

$$\begin{bmatrix} E_{t,k} \\ R_{t,k} \end{bmatrix} = \begin{bmatrix} 1 & -\phi_{kk} \\ -\phi_{kk} & 1 \end{bmatrix} \begin{bmatrix} E_{t,k-1} \\ R_{t-1,k-1} \end{bmatrix} \tag{3.76}$$

$$D_k = (1-\phi^2_{kk})D_{k-1}$$

$$Q_k = D_k,$$

initialized with $E_{t,0} = Y_t$, $R_{t-1,0} = Y_{t-1}$, $\phi_{11} = \gamma(1)\gamma^{-1}(0)$, $D_0 = \gamma(0)$ *and* $Q_0 = \gamma(0)$, *where* $D_k = \mathrm{Var}(E_{t,k})$ *and* $Q_k = \mathrm{Var}(R_{t-1,k})$.

Proof As in the proof of Theorem 3.14, let $Z_3 = Y_t$, $Z_2 = Y_{t-k}$ and $Z_1 = (Y'_{t-1},\ldots, Y'_{t-k+1})'$. Applying formula (1.44) yields

$$\begin{aligned} E^*(Z_3|Z_1,Z_2) &= E^*(Z_3|Z_1) + \Pi_{32}(Z_2 - E^*(Z_2|Z_1)) \\ &= \Theta_{31}Z_1 + \Pi_{32}(Z_2 - \Theta_{21}Z_1), \end{aligned}$$

where $\Pi_{32} = \Theta_{32}$, $\Theta_{31} = \phi_{(k-1)}$, $\Theta_{21} = \phi^*_{(k-1)}$ and $\Theta_{32} = \phi_{kk}$. Subtracting from $Y_t = Z_3$ the previous expression yields the recursion for $E_{t,k}$. The recursion for D_k was proved in Theorem 3.14. The recursions for $R_{t,k}$ and Q_k are proved interchanging the roles of Z_3 and Z_2 in the previous argument. □

3.11.1.5 Recursive Autoregressive Fitting Based on the Levinson–Durbin Algorithm

Given a time series, $\{Y_1,\ldots,Y_n\}$, that is assumed to have zero mean, an obvious way to fit autoregressive models of increasing order to the data is to compute first the sample covariances, $c(j) = (1/n)\sum_{t=j+1}^{n} Y_t Y_{t-j}, j = 0, 1, 2, \ldots$, and then to use the Levinson–Durbin algorithm with the population covariances replaced with the sample ones.

3.11.1.6 Recursive Autoregressive Fitting. Burg's Algorithm

In the early 1960s a geophysicist named John Parker Burg developed a method for spectral estimation based on autoregressive modeling that he called the "maximum entropy method." As a part of this method Burg (1968) developed an algorithm to estimate the autoregressive parameters "directly from the data" without the intermediate step of computing a covariance matrix and solving the Yule–Walker equations. Burg's method is based on the Levinson–Durbin algorithm (Theorem 3.14) and the lattice representation (Theorem 3.16).

Since the forward and backward prediction problems are statistically identical, Burg suggested to use the recursions (3.76) but estimating at each step the partial autocorrelation coefficient, ϕ_{kk}, by minimizing the sum of squares

$$\begin{aligned} S_{f,p}(k) &= \sum_{t=k+1}^{n} (E_{t,k}^2 + R_{t,k}^2) \\ &= \sum_{t=k+1}^{n} [E_{t,k-1}, R_{t-1,k-1}] \begin{bmatrix} 1 & -\phi_{kk} \\ -\phi_{kk} & 1 \end{bmatrix}^2 \begin{bmatrix} E_{t,k-1} \\ R_{t-1,k-1} \end{bmatrix} \\ &= \sum_{t=k+1}^{n} [E_{t,k-1}, R_{t-1,k-1}] \begin{bmatrix} 1+\phi_{kk}^2 & -2\phi_{kk} \\ -2\phi_{kk} & 1+\phi_{kk}^2 \end{bmatrix} \begin{bmatrix} E_{t,k-1} \\ R_{t-1,k-1} \end{bmatrix}. \end{aligned}$$

Using standard calculus, it is not difficult to verify that the solution to this minimization problem is

$$\hat{\phi}_{kk} = -\frac{2\sum_{t=k+1}^{n} E_{t,k-1}R_{t-1,k-1}}{\sum_{t=k+1}^{n} E_{t,k-1}^2 + \sum_{t=k+1}^{n} R_{t-1,k-1}^2}.$$

A notable feature of this solution is that $|\hat{\phi}_{kk}| \leq 1$. To see this, consider $x, y \in \mathbb{R}^n$, where $x = (x_1, \ldots, x_n)'$, $y = (y_1, \ldots, y_n)'$ and $\mathbb{R}^n$ is the standard Euclidean space with scalar product and norm defined by $x \cdot y = \sum_{i=1}^{n} x_i y_i$ and $||x|| = \sqrt{x \cdot x}$. Then,

$$(x+y)\cdot(x+y) = ||x||^2 + ||y||^2 + 2x\cdot y \geq 0$$

and

$$\frac{-2x\cdot y}{||x||^2 + ||y||^2} \leq 1.$$

We summarize Burg's algorithm in the following theorem.

Theorem 3.17 (Burg's Algorithm) *Suppose a time series* $\{Y_1, \ldots, Y_n\}$ *that is assumed to have zero mean. Then, the following recursions*

$$\phi_{kk} = -\frac{2\sum_{t=k+1}^{n} E_{t,k-1}R_{t-1,k-1}}{\sum_{t=k+1}^{n} E_{t,k-1}^2 + \sum_{t=k+1}^{n} R_{t-1,k-1}^2}$$

$$\begin{bmatrix} E_{t,k} \\ R_{t,k} \end{bmatrix} = \begin{bmatrix} 1 & -\phi_{kk} \\ -\phi_{kk} & 1 \end{bmatrix} \begin{bmatrix} E_{t,k-1} \\ R_{t-1,k-1} \end{bmatrix},$$

initialized with $E_{t,0} = Y_t$ *and* $R_{t-1,0} = Y_{t-1}$*, estimate the partial autocorrelation coefficients,* ϕ_{kk}*,* $k = 1, 2, \ldots$*, where the procedure ensures that* $|\phi_{kk}| \leq 1$ *for all* k*. In addition, the autoregressive coefficients,* $(\phi_{k1}, \ldots, \phi_{kk})$*, and the variances* $D_k = \mathrm{Var}(E_{t,k})$ *and* $Q_k = \mathrm{Var}(R_{t-1,k})$ *can be computed with the supplementary recursions*

$$(\phi_{k1}, \ldots, \phi_{k,k-1}) = \phi_{(k-1)} - \phi_{kk}\phi_{(k-1)}^*$$

$$D_k = (1 - \phi_{kk}^2)D_{k-1}$$

$$Q_k = D_k,$$

initialized with $D_0 = Q_0 = (1/n)\sum_{t=1}^{n} Y_t^2$*, where* $\phi_{(k)} = (\phi_{k1}, \ldots, \phi_{kk})$ *and* $\phi_{(k)}^* = (\phi_{kk}, \ldots, \phi_{k1})$*.*

Proof The only thing that requires proof are the supplementary recursions. But these are an immediate consequence of Theorems 3.14 and 3.16. □

3.11.1.7 Recursive Autoregressive Fitting. Modified Burg's Algorithm

An alternative procedure to Burg's algorithm that, as we will see, can be easily extended to the multivariate case consists of using the recursions (3.76) but estimating the partial autocorrelation coefficients, ϕ_{kk}, in a different way. By Theorem 3.16, ϕ_{kk} is the partial autocorrelation coefficient between Y_t and Y_{t-k}. Thus,

$$\phi_{kk} = \frac{\mathrm{Cov}(E_{t,k-1}, R_{t-1,k-1})}{\sqrt{D_{k-1}}\sqrt{Q_{k-1}}},$$

where $E_{t,k-1}$ and $R_{t-1,k-1}$ are given by (3.74) and (3.75), $D_{k-1} = \mathrm{Var}(E_{t,k-1})$ and $Q_{k-1} = \mathrm{Var}(R_{t-1,k-1})$, and, by stationarity, ϕ_{kk} minimizes both

$$E(E_{t,k-1} - \phi_{kk}R_{t-1,k-1})^2$$

and

$$E(R_{t-1,k-1} - \phi_{kk}E_{t,k-1})^2.$$

Therefore, we can estimate ϕ_{kk} at each step in the Levinson–Durbin algorithm by

$$\hat{\phi}_{kk} = \frac{\sum_{t=k+1}^{n} E_{t,k-1} R_{t-1,k-1}}{\sqrt{\sum_{t=k+1}^{n} E_{t,k-1}^2}\sqrt{\sum_{t=k+1}^{n} R_{t-1,k-1}^2}},$$

where $E_{t,k-1}$ and $R_{t-1,k-1}$ can be computed using the recursions (3.76), initialized with $E_{t,0} = Y_t$ and $R_{t-1,0} = Y_{t-1}$. Note that $|\hat{\phi}_{kk}| \leq 1$ because $\hat{\phi}_{kk}$ is the sample autocorrelation coefficient between $E_{t,k-1}$ and $R_{t-1,k-1}$. We summarize this result in the following theorem.

Theorem 3.18 (Modified Burg's Algorithm) *Suppose a time series* $\{Y_1, \ldots, Y_n\}$ *that is assumed to have zero mean. Then, the following recursions*

$$\phi_{kk} = \frac{\sum_{t=k+1}^{n} E_{t,k-1} R_{t-1,k-1}}{\sqrt{\sum_{t=k+1}^{n} E_{t,k-1}^2}\sqrt{\sum_{t=k+1}^{n} R_{t-1,k-1}^2}}$$

$$\begin{bmatrix} E_{t,k} \\ R_{t,k} \end{bmatrix} = \begin{bmatrix} 1 & -\phi_{kk} \\ -\phi_{kk} & 1 \end{bmatrix} \begin{bmatrix} E_{t,k-1} \\ R_{t-1,k-1} \end{bmatrix},$$

initialized with $E_{t,0} = Y_t$ *and* $R_{t-1,0} = Y_{t-1}$*, estimate the partial autocorrelation coefficients,* ϕ_{kk}*,* $k = 1, 2, \ldots$*, where the procedure ensures that* $|\phi_{kk}| \leq 1$ *for all* k*. In addition, the autoregressive coefficients,* $(\phi_{k1}, \ldots, \phi_{kk})$*, and the variances* $D_k = \mathrm{Var}(E_{t,k})$ *and* $Q_k = \mathrm{Var}(R_{t-1,k})$ *can be computed with the supplementary recursions*

$$\begin{aligned} (\phi_{k1}, \ldots, \phi_{k,k-1}) &= \phi_{(k-1)} - \phi_{kk}\phi^{*}_{(k-1)} \\ D_k &= (1 - \phi_{kk}^2) D_{k-1} \\ Q_k &= D_k, \end{aligned}$$

initialized with $D_0 = Q_0 = (1/n)\sum_{t=1}^{n} Y_t^2$*, where* $\phi_{(k)} = (\phi_{k1}, \ldots, \phi_{kk})$ *and* $\phi^{*}_{(k)} = (\phi_{kk}, \ldots, \phi_{k1})$*.*

3.11.2 Multivariate Processes

In the multivariate case, two recursions are necessary in the generalization of the Levinson–Durbin algorithm because of the lack of symmetry of the autocovariance sequence, that is, $\gamma(-h) = \gamma'(h)$. The generalization is called Whittle's algorithm and is described in the next theorem.

Theorem 3.19 (Whittle's Algorithm Whittle 1963a) *Suppose a zero-mean multivariate stationary process* $\{Y_t\}$ *with covariance generating function* $G_Y(z) = \sum_{j=-\infty}^{\infty} \gamma(j) z^j$ *satisfying the Paley–Wiener condition of Theorem 3.12. For fixed* t

and increasing k, define the forward and backward innovations

$$\begin{aligned}E_{t,k} &= Y_t - E^*(Y_t|Y_{t-1},\ldots,Y_{t-k})\\ &= Y_t - \Phi_{k1}Y_{t-1} - \cdots - \Phi_{kk}Y_{t-k}\\ R_{t-1,k} &= Y_{t-k-1} - E^*(Y_{t-k-1}|Y_{t-1},\ldots,Y_{t-k})\\ &= Y_{t-k-1} - \Phi^*_{kk}Y_{t-1} - \cdots - \Phi^*_{k1}Y_{t-k},\end{aligned}$$

and denote $D_k = \text{Var}(E_{t,k})$*,* $Q_k = \text{Var}(R_{t-1,k})$ *and* $\Delta_k = \text{Cov}(E_{t,k}, R_{t-1,k})$*. Let* $\Phi_{(k)} = (\Phi_{k1},\ldots,\Phi_{kk})$*,* $\Phi^*_{(k)} = (\Phi^*_{kk},\ldots,\Phi^*_{k1})$*,* $\Gamma_{(k)} = (\gamma(1),\ldots,\gamma(k))$*, and* $\Gamma^*_{(k)} = (\gamma'(k),\ldots,\gamma'(1))$*. Then, the following recursions hold*

$$\begin{aligned}\Phi_{kk} &= \Delta_{k-1}Q_{k-1}^{-1} = \left(\gamma(k) - \Gamma_{(k-1)}\Phi^{*'}_{(k-1)}\right)Q_{k-1}^{-1}\\ &= \left(\gamma(k) - \Phi_{(k-1)}\Gamma^{*'}_{(k-1)}\right)Q_{k-1}^{-1}\\ (\Phi_{k1},\ldots,\Phi_{k,k-1}) &= \Phi_{(k-1)} - \Phi_{kk}\Phi^*_{(k-1)}\\ \Phi^*_{kk} &= \Delta'_{k-1}D_{k-1}^{-1} = \left(\gamma'(k) - \Gamma^*_{(k-1)}\Phi'_{(k-1)}\right)D_{k-1}^{-1}\\ &= \left(\gamma'(k) - \Phi^*_{(k-1)}\Gamma'_{(k-1)}\right)D_{k-1}^{-1}\\ (\Phi^*_{k,k-1},\ldots,\Phi^*_{k1}) &= \Phi^*_{(k-1)} - \Phi^*_{kk}\Phi_{(k-1)}\\ D_k &= D_{k-1} - \Delta_{k-1}Q_{k-1}^{-1}\Delta'_{k-1} = (I - \Phi_{kk}\Phi^*_{kk})D_{k-1}\\ Q_k &= Q_{k-1} - \Delta'_{k-1}D_{k-1}^{-1}\Delta_{k-1} = (I - \Phi^*_{kk}\Phi_{kk})Q_{k-1},\end{aligned}$$

initialized with $E_{t,0} = Y_t$*,* $R_{t-1,0} = Y_{t-1}$*,* $\Phi_{11} = \gamma(1)\gamma^{-1}(0)$*,* $D_0 = \gamma(0)$*,* $\Phi^*_{11} = \gamma'(1)\gamma^{-1}(0)$*,* $\Delta_0 = \gamma(1)$ *and* $Q_0 = \gamma(0)$*. In addition,* Φ_{kk} *is the* ***partial autoregression matrix*** *of order k,*

$$\rho_k = D_{k-1}^{-1/2}\Delta_{k-1}Q_{k-1}^{-1/2'} = D_{k-1}^{-1/2}\Phi_{kk}Q_{k-1}^{1/2} = D_{k-1}^{1/2'}\Phi^{*'}_{kk}Q_{k-1}^{-1/2'}$$

is the ***partial cross-correlation matrix*** *of order k, and the following lattice representation holds*

$$\begin{bmatrix}E_{t,k}\\ R_{t,k}\end{bmatrix} = \begin{bmatrix}I & -\Phi_{kk}\\ -\Phi^*_{kk} & I\end{bmatrix}\begin{bmatrix}E_{t,k-1}\\ R_{t-1,k-1}\end{bmatrix}.$$

Proof Let $Z_3 = Y_t$, $Z_2 = Y_{t-k}$ and $Z_1 = (Y'_{t-1}, \ldots, Y'_{t-k+1})'$ and apply formula (1.44). Then,

$$\begin{aligned} E^*(Z_3|Z_1, Z_2) &= E^*(Z_3|Z_1) + \Pi_{32}(Z_2 - E^*(Z_2|Z_1)) \\ &= \Theta_{31}Z_1 + \Pi_{32}(Z_2 - \Theta_{21}Z_1) \\ &= \Pi_{31}Z_1 + \Pi_{32}Z_2, \end{aligned}$$

where $\Pi_{31} = \Theta_{31} - \Theta_{32}\Theta_{21}$, $\Pi_{32} = \Theta_{32}$ and Θ_{21}, Θ_{31} and Θ_{32} are given by (1.45). From this, it is obtained that $\Theta_{31} = \Phi_{(k-1)}$, $\Theta_{21} = \Phi^*_{(k-1)}$, $(\Pi_{31}, \Pi_{32}) = \Phi_{(k)}$, $\Pi_{31} = (\Phi_{k1}, \ldots, \Phi_{k,k-1})$, and $\Theta_{32} = \Phi_{kk}$. Thus,

$$(\Phi_{k1}, \ldots, \Phi_{k,k-1}) = \Phi_{(k-1)} - \Phi_{kk}\Phi^*_{(k-1)}.$$

By Proposition 1.8,

$$\Pi_{32} = \Phi_{kk} = \Delta_{k-1}Q^{-1}_{k-1}$$

and, from (1.46), we can write

$$D_k = D_{k-1} - \Delta_{k-1}Q^{-1}_{k-1}\Delta'_{k-1}.$$

The expression (1.45) for $\Theta_{32} = \Phi_{kk}$ yields

$$\Phi_{kk} = \left[\gamma(k) - (\gamma(1), \ldots, \gamma(k-1))\Phi^{*'}_{(k-1)}\right] Q^{-1}_{k-1}.$$

Interchanging the roles of Z_3 and Z_2 in the previous argument, by stationarity, the other recursions follow. The only thing that remains to be proved is the formula $D_k = D_{k-1} - \Delta_{k-1}Q^{-1}_{k-1}\Delta'_{k-1} = (I - \Phi_{kk}\Phi^*_{kk})D_{k-1}$ and its analogue for Q_k. But, because $\Phi_{kk} = \Delta_{k-1}Q^{-1}_{k-1}$ and $\Phi^*_{kk}D_{k-1} = \Delta'_{k-1}$, we have $D_k = D_{k-1} - \Phi_{kk}\Delta'_{k-1} = D_{k-1} - \Phi_{kk}\Phi^*_{kk}D_{k-1}$ and the theorem is proved. □

Corollary 3.2 *Under the assumptions and with the notation of the previous theorem, the eigenvalues of the matrix $\Phi_{kk}\Phi^*_{kk}$ are the squared canonical correlations between the residuals $E_{t,k-1}$ and $R_{t-1,k-1}$. Thus, they are the squared partial canonical correlations between Y_t and Y_{t-k} with respect to $\{Y_{t-1}, \ldots, Y_{t-k+1}\}$. Moreover, these partial canonical correlations are the singular values of $\rho_k = \mathrm{Cov}(\overline{E}_{t,k-1}, \overline{R}_{t-1,k-1}) = D^{-1/2}_{k-1}\Delta_{k-1}Q^{-1/2'}_{k-1}$, where $\overline{E}_{t,k-1}$ and $\overline{R}_{t-1,k-1}$ are the standardized residuals $\overline{E}_{t,k-1} = D^{-1/2}_{k-1}E_{t,k-1}$ and $\overline{R}_{t-1,k-1} = Q^{-1/2}_{k-1}R_{t-1,k-1}$.*

Proof The first part of the corollary follows from the formula

$$\begin{aligned} \Phi_{kk}\Phi^*_{kk} &= \Delta_{k-1}Q^{-1}_{k-1}\Delta'_{k-1}D^{-1}_{k-1} \\ &= \Sigma_{ER}\Sigma^{-1}_{RR}\Sigma_{RE}\Sigma^{-1}_{EE}, \end{aligned}$$

where $\Sigma_{ER} = \text{Cov}(E_{t,k-1}, R_{t-1,k-1})$, $\Sigma_{RE} = \Sigma'_{ER}$, $\Sigma_{RR} = \text{Var}(R_{t-1,k-1})$, and $\Sigma_{EE} = \text{Var}(E_{t,k-1})$. To prove the second part, consider that $\det(\Phi_{kk}\Phi^*_{kk} - \lambda I) = \det(\rho_k\rho'_k - \lambda I)$, where $\rho_k = D_{k-1}^{-1/2}\Delta_{k-1}Q_{k-1}^{-1/2'}$. □

Theorem 3.20 (Normalized Whittle's Algorithm Morf, Vieira, & Kailath 1978) *Suppose a zero-mean multivariate stationary process $\{Y_t\}$ with covariance generating function $G_Y(z) = \sum_{j=-\infty}^{\infty} \gamma(j)z^j$. With the notation of Theorem 3.19, for fixed t and increasing k define the forward and backward standardized innovations*

$$\overline{E}_{t,k} = D_k^{-1/2}E_{t,k}$$
$$\overline{R}_{t-1,k} = Q_k^{-1/2}R_{t-1,k}$$

and denote $\rho_k = \text{Cov}(\overline{E}_{t,k-1}, \overline{R}_{t-1,k-1}) = D_{k-1}^{-1/2}\Delta_{k-1}Q_{k-1}^{-1/2'}$ and $\tilde{\Delta}_k = \Delta_k Q_k^{-1/2'}$. Let $\tilde{\Phi}_{(k)} = D_k^{-1/2}\left[I, -\Phi_{(k)}\right]$ and $\tilde{\Phi}^*_{(k)} = Q_k^{-1/2}\left[-\Phi^*_{(k)}, I\right]$. Then, the following recursions hold

$$\begin{aligned}
\rho_k &= D_{k-1}^{-1/2}\tilde{\Delta}_{k-1} \\
\tilde{\Phi}_{(k)} &= (I - \rho_k\rho'_k)^{-1/2}\left\{\left[\tilde{\Phi}_{(k-1)}, 0\right] - \rho_k\left[0, \tilde{\Phi}^*_{(k-1)}\right]\right\} \\
\tilde{\Phi}^*_{(k)} &= (I - \rho'_k\rho_k)^{-1/2}\left\{\left[0, \tilde{\Phi}^*_{(k-1)}\right] - \rho'_k\left[\tilde{\Phi}_{(k-1)}, 0\right]\right\} \\
D_k^{-1/2} &= (I - \rho_k\rho'_k)^{-1/2}D_{k-1}^{-1/2} \\
Q_k^{-1/2} &= (I - \rho'_k\rho_k)^{-1/2}Q_{k-1}^{-1/2} \\
\tilde{\Delta}_k &= \Gamma_{(k+1)}\tilde{\Phi}^{*'}_{(k)} \\
\overline{E}_{t,k} &= (I - \rho_k\rho'_k)^{-1/2}\left(\overline{E}_{t,k-1} - \rho_k\overline{R}_{t-1,k-1}\right) \\
\overline{R}_{t,k} &= (I - \rho'_k\rho_k)^{-1/2}\left(\overline{R}_{t-1,k-1} - \rho'_k\overline{E}_{t,k-1}\right),
\end{aligned}$$

initialized with $(I - \rho_0\rho'_0)^{1/2} = \gamma(0)^{1/2} = (I - \rho'_0\rho_0)^{1/2}$, $D_0^{-1/2} = Q_0^{-1/2} = \tilde{\Phi}_{(0)} = \tilde{\Phi}^*_{(0)} = \gamma(0)^{-1/2}$, $\tilde{\Delta}_0 = \gamma(1)\gamma(0)^{-1/2'}$, $\overline{E}_{t,0} = \gamma(0)^{-1/2}Y_t$ and $\overline{R}_{t-1,0} = \gamma(0)^{-1/2}Y_{t-1}$.

Proof One can express the recursion of Theorem 3.19

$$D_k = D_{k-1} - \Delta_{k-1}Q_{k-1}^{-1}\Delta'_{k-1}$$

in the form

$$D_k^{1/2}D_k^{1/2'} = D_{k-1}^{1/2}D_{k-1}^{1/2'} - \Delta_{k-1}Q_{k-1}^{-1/2'}Q_{k-1}^{-1/2}\Delta'_{k-1}.$$

It follows from this that

$$\begin{aligned} D_{k-1}^{-1/2} D_k^{1/2} D_k^{1/2'} D_{k-1}^{-1/2'} &= I - D_{k-1}^{-1/2} \Delta_{k-1} Q_{k-1}^{-1/2'} Q_{k-1}^{-1/2} \Delta'_{k-1} D_{k-1}^{-1/2'} \\ &= I - \rho_k \rho'_k. \end{aligned}$$

Therefore,

$$(I - \rho_k \rho'_k)^{1/2} = D_{k-1}^{-1/2} D_k^{1/2}. \tag{3.77}$$

Premultiplying the recursion of Theorem 3.19,

$$E_{t,k} = E_{t,k-1} - \Delta_{k-1} Q_{k-1}^{-1} R_{t-1,k-1},$$

by $D_k^{-1/2}$, it is obtained that

$$\begin{aligned} \overline{E}_{t,k} &= D_k^{-1/2} D_{k-1}^{1/2} \overline{E}_{t,k-1} - D_k^{-1/2} D_{k-1}^{1/2} D_{k-1}^{-1/2} \Delta_{k-1} Q_{k-1}^{-1/2'} Q_{k-1}^{-1/2} R_{t-1,k-1} \\ &= D_k^{-1/2} D_{k-1}^{1/2} \left(\overline{E}_{t,k-1} - \rho_k \overline{R}_{t-1,k-1}\right) \\ &= (I - \rho_k \rho'_k)^{-1/2} \left(\overline{E}_{t,k-1} - \rho_k \overline{R}_{t-1,k-1}\right). \end{aligned}$$

The formula

$$(I - \rho'_k \rho_k)^{1/2} = Q_{k-1}^{-1/2} Q_k^{1/2}$$

and the recursion for $\overline{R}_{t,k}$ are proved similarly.

It follows from (3.77) that

$$\begin{aligned} D_k^{1/2} &= D_{k-1}^{1/2} (I - \rho_k \rho'_k)^{1/2} \\ &= (I - \rho_0 \rho'_0)^{1/2} \cdots (I - \rho_k \rho'_k)^{1/2} \end{aligned}$$

and, given the definitions of ρ_k and $\tilde{\Delta}_k$, that

$$\rho_k = (I - \rho_{k-1} \rho'_{k-1})^{-1/2} \cdots (I - \rho_0 \rho'_0)^{-1/2} \tilde{\Delta}_{k-1}.$$

Also, $D_k^{-1/2} = (I - \rho_k \rho'_k)^{-1/2} D_{k-1}^{-1/2}$ and

$$\begin{aligned} \tilde{\Phi}_{(k)} &= (I - \rho_k \rho'_k)^{-1/2} D_{k-1}^{-1/2} \left[I, -\Phi_{(k)}\right] \\ &= (I - \rho_k \rho'_k)^{-1/2} \left[D_{k-1}^{-1/2}, -D_{k-1}^{-1/2} \left(\Phi_{(k-1)} - \Phi_{kk} \Phi^*_{(k-1)}, \Phi_{kk}\right)\right] \\ &= (I - \rho_k \rho'_k)^{-1/2} \left\{D_{k-1}^{-1/2} \left[I, -\Phi_{(k-1)}, 0\right] - D_{k-1}^{-1/2} \left[0, -\Phi_{kk} \Phi^*_{(k-1)}, \Phi_{kk}\right]\right\} \\ &= (I - \rho_k \rho'_k)^{-1/2} \left\{\left[\tilde{\Phi}_{(k-1)}, 0\right] - \rho_k \left[0, \tilde{\Phi}^*_{(k-1)}\right]\right\}, \end{aligned}$$

where we have used the formula $\rho_k = D_{k-1}^{-1/2}\Phi_{kk}Q_{k-1}^{1/2}$ from Theorem 3.19. The recursion for $\tilde{\Phi}^*_{(k)}$ can be proved similarly.

Finally, using the first recursion of Theorem 3.19, we can write

$$\begin{aligned}\tilde{\Delta}_k &= \Delta_k Q_k^{-1/2'} \\ &= \left[\gamma(k+1) - \Gamma_{(k)}\Phi^{*'}_{(k)}\right] Q_k^{-1/2'} \\ &= \Gamma_{(k+1)}\tilde{\Phi}^{*'}_{(k)}.\end{aligned}$$

□

Theorem 3.21 (Modified Burg's Algorithm Morf, Vieira, Lee, & Kailath 1978) *Suppose a vector time series $\{Y_1, \ldots, Y_n\}$ that is assumed to have zero mean. With the notation of Theorem 3.19, define*

$$\Sigma_{EE,k-1} = \sum_{t=k+1}^{n} E_{t,k-1}E'_{t,k-1}, \qquad \Sigma_{ER,k-1} = \sum_{t=k+1}^{n} E_{t,k-1}R'_{t-1,k-1}$$

and $\Sigma_{RR,k-1} = \sum_{t=k+1}^{n} R_{t-1,k-1}R'_{t-1,k-1}$. Then, the following recursions

$$\begin{aligned}\rho_k &= \Sigma_{EE,k-1}^{-1/2}\Sigma_{ER,k-1}\Sigma_{RR,k-1}^{-1/2'} \\ \Phi_{kk} &= D_{k-1}^{1/2}\rho_k Q_{k-1}^{-1/2} \\ \Phi^*_{kk} &= Q_{k-1}^{1/2}\rho'_k D_{k-1}^{-1/2} \\ \begin{bmatrix} E_{t,k} \\ R_{t,k} \end{bmatrix} &= \begin{bmatrix} I & -\Phi_{kk} \\ -\Phi^*_{kk} & I \end{bmatrix}\begin{bmatrix} E_{t,k-1} \\ R_{t-1,k-1} \end{bmatrix} \\ (\Phi_{k1}, \ldots, \Phi_{k,k-1}) &= \Phi_{(k-1)} - \Phi_{kk}\Phi^*_{(k-1)} \\ (\Phi^*_{k,k-1}, \ldots, \Phi^*_{k1}) &= \Phi^*_{(k-1)} - \Phi^*_{kk}\Phi_{(k-1)} \\ D_k &= (I - \Phi_{kk}\Phi^*_{kk})D_{k-1} \\ Q_k &= (I - \Phi^*_{kk}\Phi_{kk})Q_{k-1},\end{aligned}$$

initialized with $E_{t,0} = Y_t$, $R_{t-1,0} = Y_{t-1}$, and $D_0 = Q_0 = (1/n)\sum_{t=1}^{n} Y_t^2$, estimate the partial cross-correlation matrices, ρ_k, $k = 1, 2, \ldots$, the autoregressive coefficient matrices, $(\Phi_{k1}, \ldots, \Phi_{kk})$ and $(\Phi^*_{kk}, \ldots, \Phi^*_{k1})$, and the covariance matrices $D_k = \mathrm{Var}(E_{t,k})$ and $Q_k = \mathrm{Var}(R_{t-1,k})$.

3.12 Historical Notes

The problem of finding innovations for stationary discrete time scalar valued processes was first studied by Wold (1938) in a famous dissertation and then in greater generality by Kolmogorov (1939, 1941). The theory in all generality was presented by Doob (1953) and Grenander & Rosenblatt (1957) for the scalar and by Rozanov (1967) and Hannan (1970) for the vector case. More recent monographs are Caines (1988) and Lindquist & Picci (2015).

The spectral representation theorem is due to Cramér (1940). See also Kolmogorov (1939).

The use of scalar ARMA and ARIMA models was popularized by Box and Jenkins in the 1970s. See Box & Jenkins (1976) and also Granger & Newbold (1977). VARMA models were considered early on by Quenouille (1957). Two standard references on VARMA models are Reinsel (1997) and Lütkepohl (2007). See also the references therein.

The concepts of observability and controllability were introduced in the 1960s by R. E. Kalman in deterministic realization theory in relation to minimality. The standard reference for deterministic realization theory is Kalman, Falb, & Arbib (1969). Another classical reference is Brockett (1970).

VARMA models and innovations state space forms are studied from a statistical point of view in Hannan & Deistler (1988). See also Kailath (1980) and Chen (1984) for an engineering perspective.

The algorithms for autoregressive fitting for stationary series were proposed by Levinson (1947) and Durbin (1960) in the scalar and by Whittle (1963a) in the multivariate case. The forward and backward predictor spaces were introduced by Akaike (1974a) in the context of canonical correlation analysis.

3.13 Problems

3.1 Consider the covariance generating function $G(z) = (1 + z)(1 + z^{-1})\sigma^2$.

(a) Prove that $f(x) = G(e^{-ix}) = 2\sigma^2 (1 + \cos(x)) = 4\sigma^2 \cos^2(x/2)$ and

$$\int_{-\pi}^{\pi} \ln f(x) = 2\pi \ln(4\sigma^2) + 8 \int_0^{\pi/2} \ln\cos(t).$$

(b) Let $\phi(x) = -\int_0^x \ln\cos(t)$, $x \in [0, \pi/2]$. Prove that

$$\phi(x) = 2\phi\left(\frac{\pi}{4} + \frac{x}{2}\right) - 2\phi\left(\frac{\pi}{4} - \frac{x}{2}\right) - x\ln(2).$$

Hint: First make the change $t = \pi/2 - w$ to get $\phi(x) = -\int_0^x \ln\cos(t) = \int_{\pi/2}^{\pi/2-x} \ln\sin(w)$. Then, substitute $\sin(w) = 2\sin(w/2)\cos(w/2)$ in the last integral and express it as the sum of three integrals. One of them is easy to compute and the sum of the other two can be put in terms of the function $\phi(x)$ by means of a change of variable in each one.

(c) Conclude that

$$\phi\left(\frac{\pi}{2}\right) = -\int_0^{\pi/2} \ln\cos(t) = \frac{\pi}{2}\ln(2),$$

so that the Paley–Wiener condition (3.42) is satisfied and

$$\int_{-\pi}^{\pi} \ln f(x) = 2\pi \ln(\sigma^2).$$

(d) Prove that (3.46) and the Kolmogorov–Szegö formula (3.43) hold, that is,

$$\gamma(0) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x) = 2\sigma^2, \qquad \sigma^2 = e^{\frac{1}{2\pi}\int_{-\pi}^{\pi}\ln f(x)}.$$

3.2 Let $Y_t = A_t + A_{t-1}$, where $A_t \sim \text{WN}(0, \sigma^2)$, be a stationary process.

(a) Prove that its covariance generating function, $G(z)$, coincides with that of Problem 3.1.

(b) Letting $\Psi(z) = 1 + z$ such that $G(z) = \Psi(z)\Psi(z^{-1})\sigma^2$, verify that $\Psi(z)$ and $\Psi^{-1}(z)$ are analytic for $|z| > 1$. In addition, show that the power series expansions $\Psi(z) = \sum_{i=0}^{\infty} \Psi_i z^i$ and $\Psi^{-1}(z) = \sum_{i=0}^{\infty} \overline{\Psi}_i z^i$ are such that $\sum_{i=0}^{\infty} \Psi_i z^i$ is square summable while $\sum_{i=0}^{\infty} \overline{\Psi}_i z^i$ is not.

(c) Use Example 3.29 to show that $E^*(Y_t|Y_{t-1}, \ldots, Y_{t-k}) = \phi_{k1}Y_{t-1} + \cdots + \phi_{kk}Y_{t-k}$ with

$$\phi_{kj} = (-1)^{j+1}\frac{k-j+1}{k+1}, \quad D_k = \frac{k+2}{k+1}\sigma^2, \quad j = 1, \ldots, k,$$

where $D_k = \text{Var}\,[Y_t - E^*(Y_t|Y_{t-1}, \ldots, Y_{t-k})]$. Obtain $\lim_{k\to\infty} D_k$ and conclude that

$$\lim_{k\to\infty}\left[E^*(Y_t|Y_{t-1}, \ldots, Y_{t-k})\right] = \sum_{i=1}^{\infty} \phi_i Y_{t-i}$$

converges in mean square and $A_t = Y_t - \sum_{i=1}^{\infty} \phi_i Y_{t-i}$. However, $\sum_{i=0}^{\infty} \overline{\Psi}_i z^i$ in c) does not converge in mean square and cannot be used to describe the innovations as $A_t = \sum_{i=0}^{\infty} \overline{\Psi}_i Y_{t-i}$.

3.3 Let $a \in \mathbb{C}$ such that $|a| < 1$. Prove that

$$\int_{-\pi}^{\pi} \ln \left|1 - ae^{-ix}\right|^2 dx = 0.$$

Hint: Use the Taylor series expansion of $\ln(1-z)$ for $|z| < 1$ and the fact that $\int_{-\pi}^{\pi} e^{ikx}dx = 0$ for all $k \neq 0$.

3.4 Let

$$G(z) = \frac{\Theta(z)\Theta(z^{-1})}{\Phi(z)\Phi(z^{-1})}\sigma^2$$

be the covariance generating function of a univariate ARMA process, $\{Y_t\}$, satisfying

$$\Phi(B)Y_t = \Theta(B)A_t,$$

where $\{A_t\} \sim \text{WN}(0, \sigma^2)$ and the roots of $\Phi(z)$ and $\Theta(z)$ are all outside of the unit circle.

(a) Using Problem 3.3, prove that

$$\int_{-\pi}^{\pi} \ln \left|\Phi(e^{-ix})\right|^2 dx = \int_{-\pi}^{\pi} \ln \left|\Theta(e^{-ix})\right|^2 dx = 0$$

(b) Conclude that the Kolmogorov–Szegö formula (3.43) holds, that is,

$$\sigma^2 = e^{\frac{1}{2\pi}\int_{-\pi}^{\pi} \ln G(e^{-ix})}.$$

3.5 Let a zero-mean scalar stationary process $\{Y_t\}$ with covariance generating function $G(z) = \sum_{j=-\infty}^{\infty} \gamma(j)z^j$. For fixed t and increasing k, define the forward innovations $\{E_{t,k}\}$ by

$$\begin{aligned} Y_t &= E^*(Y_t|Y_{t-1}, \ldots, Y_{t-k}) + E_{t,k} \\ &= \phi_{k1}Y_{t-1} + \cdots + \phi_{kk}Y_{t-k} + E_{t,k} \end{aligned}$$

and denote $D_k = \text{Var}(E_{t,k})$.

(a) Prove that

$$D_1 \geq D_2 \geq \cdots \geq 0$$

and conclude that $\lim_{k\to\infty} D_k = D \geq 0$.

(b) Prove that if $D_k = 0$ for some $k \geq 0$, then $D_n = 0$ for all $n \geq k$.

(c) Prove that $D_k > 0$ if and only if Γ_{k+1} is invertible, where

$$\Gamma_{k+1} = \begin{vmatrix} \gamma(0) & \gamma(1) & \cdots & \gamma_{k-1} \\ \gamma(1) & \gamma(0) & \cdots & \gamma(k-2) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma(k-1) & \gamma(k-2) & \cdots & \gamma(0) \end{vmatrix}.$$

In this case,

$$D_k = \gamma(0) - g_k' \Gamma_k^{-1} g_k = \frac{|\Gamma_{k+1}|}{|\Gamma_k|},$$

where $g_k = (\gamma(1), \ldots, \gamma(k))$.
Hint: Use the Yule–Walker equations (3.72) and (3.73).

(d) Prove that if $\{Y_t\}$ is nondeterministic (that is, $D > 0$), then $D_k > 0$ and Γ_k is invertible for all $k \geq 1$. Moreover,

$$D = \exp\left(\lim_{k\to\infty} \frac{\ln |\Gamma_k|}{k} \right).$$

Hint: To prove the last equality, consider that $\ln(x)$ is a continuous function and that if $\{a_k\}$ is a sequence of positive numbers such that $\lim_{k\to\infty} a_k = a$, then

$$\lim_{k\to\infty} \frac{a_1 + a_2 + \cdots + a_k}{k} = a.$$

Remark: According to a limit theorem by Szegö , if $D > 0$, then

$$\lim_{k\to\infty} \frac{\ln |\Gamma_k|}{k} = \int_{-\pi}^{\pi} \ln G(e^{-ix}).$$

(e) Prove that the following inequalities hold for $k \geq 1$

$$\lambda_{\max} \Gamma_{k+1} \geq \lambda_{\max} \Gamma_k, \qquad \lambda_{\min} \Gamma_{k+1} \leq \lambda_{\min} \Gamma_k,$$

where, given a symmetric nonnegative matrix A, $\lambda_{\max}(A)$ and $\lambda_{\min}(A)$ refer to the maximum and the minimum eigenvalue, respectively.
Hint: For any $n \times n$ symmetric nonnegative matrix A, the following equalities hold

$$\lambda_{\max}(A) = \sup \left\{ \frac{x'Ax}{x'x} : x \in \mathbb{R}^n, x \neq 0 \right\}, \ \lambda_{\min}(A) = \inf \left\{ \frac{x'Ax}{x'x} : x \in \mathbb{R}^n, x \neq 0 \right\}.$$

3.6 Let $\{a_n\}$ a sequence of real numbers such that $0 \leq a_n < 1$, n 0 $1, 2, \ldots$. Prove that $\prod_{n=1}^{\infty}(1 - a_n) < \infty$ if and only if $\sum_{n=1}^{\infty} a_n < \infty$.

Hint: First, note that we can assume that $0 < a_n < 1$. Then, letting $f(x) = -[\ln(1-x)]/x$, $x > 0$, obtain $\lim_{x\to 0} f(x)$ and use it to compare the series $\sum a_n$ and $-\sum \ln(1-a_n)$.

3.7 Under the assumptions and with the notation of Problem 3.5, use Problem 3.6 to prove that $D > 0$ if and only if $\sum_{k=1}^{\infty} \phi_{kk}^2 < \infty$.
Hint: Note that by iterating in the Levinson–Durbin recursions, the following formula is obtained

$$D_k = (1-\phi_{kk}^2)(1-\phi_{k-1,k-1}^2)\cdots(1-\phi_{11}^2)\gamma(0).$$

3.8 Let $\{Y_t\}$ follow the ARMA$(2,1)$ model

$$Y_t - Y_{t-1} + (1/4)Y_{t-2} = A_t + A_{t-1},$$

where $A_t \sim \text{WN}(0,\sigma^2)$. Prove, using the method described in Sect. 3.10.5, that the autocovariances are given by the formula

$$\gamma(k) = \sigma^2 2^{-k}\left[\frac{32}{3} + 8k\right], \quad k \geq 0.$$

3.9 Consider a zero mean stationary scalar random process $\{Y_t\}$ that follows the ARMA (2,1) model

$$Y_t + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} = A_t + \theta_1 A_{t-1},$$

where the polynomials $\phi(z) = 1 + \phi_1 z + \phi_2 z^2$ and $\theta(z) = 1 + \theta_1 z$ are coprime and have all their roots outside the unit circle and $A_t \sim \text{WN}(0,\sigma^2)$. Verify the validity of the following two dimensional state space model for the process $\{Y_t\}$:

$$\begin{aligned} x_{t+1} &= Fx_t + GA_{t+1} \\ Y_t &= Hx_t, \end{aligned}$$

where

$$F = \begin{bmatrix} -\phi_1 & -\phi_2 \\ 1 & 0 \end{bmatrix}, \qquad G = \begin{bmatrix} 1 \\ 0 \end{bmatrix},$$

and $H = [1, \theta_1]$. Show that $[F, G]$ is controllable and $[F, H]$ is observable.

3.10 Let $\{Y_t\}$ follow the ARMA$(1,1)$ model

$$(1+\phi B)Y_t = (1+\theta B)A_t,$$

where $|\phi| < 1$, $|\theta| < 1$, $\phi \neq \theta$, and $A_t \sim \text{WN}(0, \sigma^2)$. Determine the coefficients $\{\Psi_i\}$ of the transfer function $\Psi(z) = \sum_{i=0}^{\infty} \Psi_i z^i = (1 + \theta z)/(1 + \phi z)$ and, using the method described in Sect. 3.10.5, prove that the autocovariances are given by

$$\gamma(k) = \begin{cases} \frac{1+\theta^2-2\phi\theta}{1-\phi^2}, & k = 0 \\ \frac{(1-\phi\theta)(\theta-\phi)}{1-\phi^2}(-\phi)^{k-1}, & k \geq 1. \end{cases}$$

Show that the autocorrelation coefficients of $\{Y_t\}$ are $\rho(1) = (1-\phi\theta)(\theta-\phi)/(1+\theta^2 - 2\phi\theta)$, $\rho(k) = (-\phi)^{k-1}\rho(1)$ for $k \geq 1$.

Appendix

Difference Equations

In this section, we will consider difference equations with constant coefficients of the form

$$Y_t + \Phi_1 Y_{t-1} + \cdots \Phi_p Y_{t-p} = A_t, \tag{3A.1}$$

where the Φ_i are $k \times k$ matrices and $\{A_t\}$ is a known sequence of $k \times 1$ vectors. We are interested in finding a sequence of $k \times 1$ vectors $\{Y_t : t = 1, 2, \ldots\}$ that satisfies (3A.1). It is convenient to transform (3A.1) into state space form

$$x_{t+1} = Fx_t + GE_t$$
$$Y_t = Hx_t,$$

where $x_t = (Y_t', Y_{t+1}', \ldots, Y_{t+p-1}')'$, $G = (0, 0, \ldots, 0, I)'$, $E_t = A_{t+p-1}$, $H = (I, 0, \ldots, 0, 0)$, and

$$F = \begin{bmatrix} 0 & I & 0 & \cdots & 0 \\ 0 & 0 & I & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & I \\ -A_p & -A_{p-1} & -A_{p-2} & \cdots & -A_1 \end{bmatrix}.$$

Notice that the nonzero eigenvalues of F coincide with the inverse of the roots of the polynomial $\det[A(z)]$, where $A(z) = I + A_1 z + \cdots + A_p z^p$.

First, we will consider the **homogeneous equation**, which is Eq. (3A.1) with $A_t = 0$. Given an initial condition, $x_1 = c$, the sequence of solution vectors $\{x_t : t = 1, 2, \ldots\}$ of the homogeneous equation satisfies

$$x_{t+1} = Fx_t \tag{3A.2}$$

$$x_1 = c, \tag{3A.3}$$

and, therefore, this sequence is of the form $\{F^{t-1}c : t = 1, 2, \ldots\}$.

In the rest of the section, unless otherwise specified, we will consider the case in which Y_t is scalar. Most of the results we will obtain can be extended to the multivariate case in a straight forward manner.

Proposition 3A.1 *The set of solutions $\{x_t : t = 1, 2, \ldots\}$ of (3A.2) and (3A.3) is a vector space of dimension p.*

Proof Let $\{x_{1t} : t = 1, 2, \ldots\}$ and $\{x_{2t} : t = 1, 2, \ldots\}$ be two solutions of (3A.2) and (3A.3). Then, it is not difficult to verify that $\{x_{1t} + rx_{2t} : t = 1, 2, \ldots\}$, where r is a real number, is also a solution. Thus, the set of solutions is a vector space. To see that this vector space has dimension p, define the mapping

$$f : x \in \mathbb{R}^p \longmapsto \{F^{t-1}x : t = 1, 2, \ldots\}.$$

Then, it is not hard to prove that f is linear and that it is in fact a bijection. Thus, the dimension of the image of f is p. □

Corollary 3A.1 *A base of the solution space of (3A.2) and (3A.3) is given by the sequences $\{F^{t-1}e_j : j = 1, \ldots, p,\ t = 1, 2, \ldots\}$, where e_j is a unit $p \times 1$ vector with a one in the jth position. Such a base is called a* ***fundamental system of solutions****.*

Proof Letting f be the mapping defined in the proof of Proposition 3A.1, it is clear that a base of the space of solutions will be the image under f of the canonical base $\{e_1, \ldots, e_p\}$ of $\mathbb{R}^p$. But $f(e_j) = \{F^{t-1}e_j : s = 1, 2, \ldots\}$. □

Remark 3A.1 It is easy to see that the vector $F^{t-1}e_j$ is the jth column of F^{t-1}. Thus, the fundamental system of solutions of Corollary 3A.1 can be compactly represented as $\{F^{t-1} : t = 1, 2, \ldots\}$ and any solution of (3A.2) and (3A.2) is of the form $\{F^{t-1}c : t = 1, 2, \ldots\}$ for some real number c. This number is determined by the initial condition $x_1 = c$. ◇

According to Remark 3A.1, the solutions of (3A.2) and (3A.3) depend on the powers of the matrix F. Let P be a nonsingular matrix such that

$$F = PJP^{-1},$$

and J is the Jordan form of P. The matrix J is given by

$$J = \begin{bmatrix} J_1 & & \\ & \ddots & \\ & & J_M \end{bmatrix},$$

where each block J_j has the form

$$J_j = \begin{bmatrix} \lambda_j & \delta_j & & \\ & \ddots & \ddots & \\ & & \ddots & \delta_j \\ & & & \lambda_j \end{bmatrix} \tag{3A.4}$$

with all $\delta_j = 1$ or all $\delta_j = 0$. Then, the powers of F can be expressed in terms of those of J as follows:

$$F^t = PJ^tP^{-1}. \tag{3A.5}$$

If all $\delta_j = 0$ in (3A.4), the powers of J_j are $J_j^t = \text{diag}(\lambda_j^t, \ldots, \lambda_j^t)$. If, on the contrary, all $\delta_j = 1$, then

$$J_j^t = \begin{bmatrix} \lambda_j^t & t\lambda_j^{t-1} & \cdots & \binom{t}{k_j - 1} \lambda_j^{t-k_j+1} \\ & \ddots & \ddots & \vdots \\ & & \ddots & t\lambda_j^{t-1} \\ & & & \lambda_j^t \end{bmatrix},$$

where k_j is the dimension of J_j. This can be seen by expressing J_j as $J_j = D_j + N_j$, where D_j is a diagonal matrix and N_j is an upper triangular matrix with ones in the first superdiagonal. Then,

$$J_j^t = (D_j + N_j)^t = \sum_{k=0}^{k_j-1} \binom{t}{k} D_j^{t-k} N_j^k,$$

because the matrices D_j and N_j commute, that is, $D_jN_j = N_jD_j$, and $N_j^{k_j} = 0$. Taking these relations into consideration, by (3A.5), we get

$$\begin{aligned} F^t &= \sum_{j=1}^{M} \sum_{k=0}^{k_j-1} B_{jk} \binom{t}{k} \lambda_j^{t-k} \\ &= \sum_{j=1}^{M} \sum_{k=0}^{k_j-1} C_{jk} t^k \lambda_j^t \end{aligned} \tag{3A.6}$$

where the coefficient $p \times p$ matrices C_{jk} depend on P and λ_j. Since F is a matrix of real numbers, if $\lambda_j = r_j e^{i\theta_j} = r_j(\cos\theta_j + i\sin\theta_j)$ is a complex eigenvalue of F, then there exists another eigenvalue that is the complex conjugate of λ_j, $\bar{\lambda}_j = r_j e^{-i\theta_j} = r_j(\cos\theta_j - i\sin\theta_j)$. In addition, for each matrix C_{jk} associated with λ_j there exists another matrix associated with $\bar{\lambda}_j$ equal to the complex conjugate of C_{jk}, $\overline{C}_{jk}$. Grouping the terms in λ_j and $\bar{\lambda}_j$ yields

$$\sum_{k=0}^{k_j-1} C_{jk} t^k \lambda_j^t + \sum_{k=0}^{k_j-1} \overline{C}_{jk} t^k \bar{\lambda}_j^t = \sum_{k=0}^{k_j-1} \left[D_{jk}\cos(t\theta_j) + E_{jk}\sin(t\theta_j)\right] t^k r_j^t,$$

where D_{jk} and E_{jk} are real coefficient matrices that depend on C_{jk}. Thus, (3A.6) can be expressed as

$$F^t = \sum_{j=1}^{M_1}\sum_{k=0}^{k_j-1} C_{jk} t^k \lambda_j^t + \sum_{j=1}^{M_2}\sum_{k=0}^{k_j-1} \left[D_{jk}\cos(t\theta_j) + E_{jk}\sin(t\theta_j)\right] t^k r_j^{tk},$$

where $M_1 + M_2 = M$ and the λ_j in the first term on the right-hand side of the previous equality are real. By Remark 3A.1, the solutions of (3A.2) and (3A.3) are of the form $\{F^{t-1}c : t = 1, 2, \ldots\}$ for some real number c. Since $Y_t = Hx_t$, the solutions of (3A.1) with $A_t = 0$ (the homogeneous difference equation) are of the form $\{HF^{t-1}c : t = 1, 2, \ldots\}$, where $c = (Y_1, \ldots, Y_p)'$. We summarize these results in the following theorem.

Theorem 3A.1 *The solutions $\{Y_t : t = 1, 2, \ldots\}$ of the homogenous equation*

$$Y_t + \Phi_1 Y_{t-1} + \cdots \Phi_p Y_{t-p} = 0,$$

are of the form

$$\begin{aligned} Y_t &= \sum_{j=1}^{M_1}\sum_{k=0}^{k_j-1} c_{jk} t^k \lambda_j^t + \sum_{j=1}^{M_2}\sum_{k=0}^{k_j-1} \left[d_{jk}\cos(t\theta_j) + e_{jk}\sin(t\theta_j)\right] t^k r_j^t \\ &= \sum_{j=1}^{M_1}\sum_{k=0}^{k_j-1} c_{jk} t^k \lambda_j^t + \sum_{j=1}^{M_2}\sum_{k=0}^{k_j-1} a_{jk}\cos\left[t\theta_j + b_{jk}\right] t^k r_j^t, \end{aligned}$$

where the c_{jk}, d_{jk}, e_{jk}, a_{jk}, and b_{jk} are real numbers that are determined by the initial conditions $\{Y_1, \ldots, Y_p\}$.

Proof The only thing that has not been proved is the formula

$$d_{jk}\cos(t\theta_j) + e_{jk}\sin(t\theta_j) = a_{jk}\cos\left[t\theta_j + b_{jk}\right].$$

But it is an immediate consequence of standard trigonometric identities that

$$a_{jk} = \sqrt{d_{jk}^2 + e_{jk}^2}, \qquad b_{jk} = \arctan(-d_{jk}/e_{jk}).$$

□

Example 3A.1 Let $\{Y_t\}$ be the solution of

$$\nabla_4 Y_t = 0$$

where $\nabla_4 = 1 - B^4$ and B is the backshift operator, $BY_t = Y_{t-1}$. Then, the equation in state space form is $x_{t+1} = Fx_t$, where

$$F = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}.$$

The eigenvalues of F are the inverses of the roots of the polynomial $1 - z^4$. Since $1 - z^4 = (1 - z)(1 + z)(1 + z^2)$, the eigenvalues are $1, -1, e^{i\pi/2}, e^{-i\pi/2}$. According to Theorem 3A.1, the general solution is

$$Y_t = c_1 + c_2(-1)^t + c_3 \cos(t\pi/2) + c_4 \sin(t\pi/2),$$

where c_1, c_2, c_3, and c_4 are determined by the initial conditions.

◇

Example 3A.2 (The Autocorrelation Function of a Scalar ARMA Model) Let the scalar process $\{Y_t\}$ follow the ARMA(p, q) model

$$\Phi(B)Y_t = \Theta(B)A_t,$$

where $\Phi(z) = 1 + \phi_1 z + \cdots + \phi_p z^p$, $\Theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q$, $\{A_t\} \sim \text{WN}(0, \sigma^2)$ and the roots of $\Phi(z)$ are all outside of the unit circle. Then, the autocovariances satisfy

$$\begin{array}{ll} \gamma_Y(l) + \sum_{j=1}^{p} \phi_j \gamma_Y(l-j) = \sum_{j=l}^{q} \theta_j \sigma^2 \psi_{j-l}, & 0 \le l \le q \\ \gamma_Y(l) + \sum_{j=1}^{p} \phi_j \gamma_Y(l-j) = 0, & l > q. \end{array}$$

Thus, dividing by $\gamma_Y(0)$ the last equation, we get

$$\rho_Y(h) + \sum_{j=1}^{p} \phi_j \rho_Y(h-j) = 0, \qquad h > \max(p, q),$$

and the autocorrelations $\rho(h)$ satisfy a difference equation of order p for $h >\max(p, q)$. The pattern of the autocorrelations will be determined by the roots of the polynomial $\Phi(z)$. ◇

3.A.1 The Nonhomogeneous Case

Returning to the nonhomogeneous equation (3A.1), consider the equation in vector form, $x_{t+1} = Fx_t + GE_t$. Iterating in this equation, we get

$$x_t = F^{t-1}x_1 + \left(GE_{t-1} + FGE_{t-2} + \cdots + F^{t-2}GE_1\right).$$

and it is seen that the general solution is the sum of the solution of the homogeneous equation plus the term $\left(GE_{t-1} + FGE_{t-2} + \cdots + F^{t-2}GE_1\right)$, which is sometimes called the **particular solution**. Since $Y_t = Hx_t$, the general solution of (3A.1) is

$$Y_t = HF^{t-1}x_1 + \left(HGE_{t-1} + HFGE_{t-2} + \cdots + HF^{t-2}GE_1\right).$$

3.A.2 Stochastic Difference Equations

Until now we have considered nonstochastic difference equations. Let $\{Y_t\}$ be a vector process that follows a VARMA model

$$Y_t + \Phi_1 Y_{t-1} + \cdots + \Phi_p Y_{t-p} = A_t + \Theta_3 A_{t-1} + \cdots + \Theta_q A_{t-q}, \tag{3A.7}$$

or, more compactly, $\Phi(B)Y_t = \Theta(B)A_t$, where B is the backshift operator, $BY_t = Y_{t-1}$. Using, for example, Akaike's representation (3.13) and (3.14), where F and K are given by (3.11) and $H = [I, 0, \ldots, 0]$, this model can be put into innovations state space form

$$x_{t+1} = Fx_t + KA_t$$
$$Y_t = Hx_t + A_t.$$

Then, iterating in the transition equation, it is obtained that

$$x_t = \left(KA_{t-1} + FKA_{t-2} + \cdots + F^{t-2}KA_1\right) + F^{t-1}x_1$$

and

$$\begin{aligned} Y_t &= \left(A_t + HKA_{t-1} + HFKA_{t-2} + \cdots + HF^{t-2}KA_1\right) + HF^{t-1}x_1 \\ &= (A_t + \Psi_1 A_{t-1} + \cdots + \Psi_{t-1}A_1) + h_t x_1, \end{aligned} \tag{3A.8}$$

where $\sum_{i=0}^{\infty} \Psi_i z^i = \Phi^{-1}(z)\Theta(z) = I + zH(I - Fz)^{-1}K$ and $h_t = HF^{t-1}$.

It is to be noticed that $\{x_t\}$ satisfies the difference equation $x_{t+1} = Fx_t + KA_t$ and that, therefore, $\{F^{t-1}x_1\}$ is a solution of the homogenous equation $x_{t+1} = Fx_t$ and $\{KA_{t-1} + FKA_{t-2} + \cdots + F^{t-2}KA_1\}$ is a particular solution. And it is to be further noticed that, because $\Phi(z)\left(\sum_{i=0}^{\infty} \Psi_i z^i\right) = \Theta(z)$, if we apply the operator

$\Phi(B)$ to (3A.8), then

$$\Phi(B)Y_t = \Theta(B)A_t + \Phi(B)h_t x_1, \qquad t > r,$$

and, therefore,

$$\Phi(B)h_t = 0, \qquad t > r,$$

where $r = \max\{p, q\}$. The dimension of x_1 depends on the properties of the state space form. If the state space form is minimal, the dimension of x_1 is equal to the McMillan degree.

If only polynomial methods are used, one possible way to generate the solution, Y_t, of (3A.7) in the form (3A.8) is to set $h_t = (0, \ldots, I, \ldots, 0)$, $t = 1, 2, \ldots, r$, where the I is in the tth position, and $h_t = -\Phi_1 h_{t-1} - \cdots - \Phi_1 h_{t-p}$ for $t > r$, and to obtain the Ψ_i weights from the relation $\Phi(z)\left(\sum_{i=0}^{\infty} \Psi_i z^i\right) = \Theta(z)$. However, in this procedure the dimension of x_1 will not be minimal in general. In fact,

$$x_1 = \begin{bmatrix} Y_1 - A_1 \\ Y_2 - A_2 - \Psi_1 A_1 \\ \vdots \\ Y_r - A_r - \Psi_1 A_{r-1} - \cdots - \Psi_{r-1} A_1 \end{bmatrix}.$$

To obtain a minimal x_1, we need to first put the VARMA model (3A.7) into echelon form, $\Phi_E(B)Y_t = \Theta_E(B)A_t$, that will be described in Chap. 4. Let $\Psi_{i,j}$ denote the ith row of the matrix Ψ_j, $i = 1, 2, \ldots$, $Y_t = (Y_{1t}, \ldots, Y_{kt})'$ and $A_t = (A_{1t}, \ldots, A_{kt})'$, and let n_i be the ith Kronecker index, $i = 1, \ldots, k$. Define

$$Y_{i|0}^p = Y_{pi} - A_{pi} - \Psi_{p,1} A_{i-1} + \cdots + \Psi_{p,i-1} A_1, \quad i = 1, \ldots, n_i, \quad p = 1, \ldots, k,$$

and

$$x_1 = \left[Y_{1|0}^1, \ldots, Y_{n_1|0}^1, Y_{1|0}^2, \ldots, Y_{n_2|0}^2, \ldots, Y_{1|0}^k, \ldots, Y_{n_k|0}^k\right]'.$$

Then, as shown in Chap. 4, the elements of x_1 constitute a basis of the space of predictors and there exists a selector matrix, J_n, where $n = \sum_{i=1}^{k} n_i$ is the McMillan degree, formed with zeros and ones such that

$$x_1 = J_n \begin{bmatrix} Y_1 - A_1 \\ Y_2 - A_2 - \Psi_1 A_1 \\ \vdots \\ Y_r - A_r - \Psi_1 A_{r-1} - \cdots - \Psi_{r-1} A_1 \end{bmatrix},$$

where $r = \max\{n_i : i = 1, \ldots, k\}$. Since the elements of x_1 form a basis, J_n has rank n and we can define h_t, $t = 1, \ldots, r$, as

$$(J_n'J_n)^{-1}J_n'x_1 = \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_r \end{bmatrix} x_1 = \begin{bmatrix} Y_1 - A_1 \\ Y_2 - A_2 - \Psi_1 A_1 \\ \vdots \\ Y_r - A_r - \Psi_1 A_{r-1} - \cdots - \Psi_{r-1}A_1 \end{bmatrix}. \quad (3A.9)$$

Then, we define h_t for $t > r$ recursively so that they satisfy

$$\Phi_E(B)h_t = 0, \qquad t > r.$$

As we will see in Chap. 4, the elements of x_1 are simply the first linearly independent rows of the right-hand side matrix in (3A.9), and these rows are determined by the Kronecker indices, as are the first linearly independent rows of the Hankel matrices, H_t, for $t > r$.

3.A.3 Generating Functions and State Space Models

Consider the state space model

$$\begin{aligned} x_{t+1} &= Fx_t + Gu_t \\ Y_t &= Hx_t + Ju_t, \quad t \geq 0, \end{aligned}$$

where $\mathrm{Var}(u_t) = I$ and we assume that the time starts at zero for simplicity in what follows. Define the generating functions $G_x(z) = \sum_{i=0}^{\infty} x_i z^i$, $G_u(z) = \sum_{i=0}^{\infty} u_i z^i$ and $G_Y(z) = \sum_{i=0}^{\infty} Y_i z^i$. Then, we have

$$x_1 + x_2 z + x_3 z^2 + \cdots = F(x_0 + x_1 z + x_2 z^2 + \cdots) + G(u_0 + u_1 z + u_2 z^2 + \cdots),$$

that can be abbreviated to

$$z^{-1}(G_x(z) - x_0) = FG_x(z) + GG_u(z).$$

It follows from this that

$$(z^{-1}I - F)G_x(z) = z^{-1}x_0 + GG_u(z)$$

and

$$G_x(z) = (I - zF)^{-1}\left[x_0 + zGG_u(z)\right].$$

Using this last expression, we finally get

$$\begin{aligned} G_Y(z) &= HG_x(z) + JG_u(z) \\ &= H(I - zF)^{-1}x_0 + \left[J + zH(I - zF)^{-1}G\right]G_u(z). \end{aligned}$$

If $x_0 = 0$, then

$$\begin{aligned} G_Y(z) &= \left[J + zH(I - zF)^{-1}G\right]G_u(z) \\ &= \Psi(z)G_u(z), \end{aligned}$$

where $\Psi(z) = J + zH(I - zF)^{-1}G$ is the transfer function of the state space model.

Chapter 4
The State Space Model

4.1 The State Space Model

Suppose that $\{Y_t\}$ is a multivariate process with $Y_t \in \mathbb{R}^p$ that admits the state space representation

$$x_{t+1} = F_t x_t + G_t u_t \tag{4.1}$$

$$Y_t = H_t x_t + v_t, \qquad t = 1, 2, 3, \ldots, \tag{4.2}$$

where $x_t \in \mathbb{R}^r$ is the state vector,

$$E\left\{ \begin{bmatrix} u_t \\ v_t \end{bmatrix} \begin{bmatrix} u_s', v_s' \end{bmatrix} \right\} = \sigma^2 \begin{bmatrix} Q_t & S_t \\ S_t' & R_t \end{bmatrix} \delta_{ts},$$

$u_t \in \mathbb{R}^s, v_t \in \mathbb{R}^p, E(u_t) = 0, E(v_t) = 0$, the **"initial state vector,"** x_1, is orthogonal to u_t and v_t for all $t, E(x_1) = a$ and $\text{Var}(x_1) = \sigma^2\Omega$. Equations (4.1) and (4.2) are called the **"transition equation"** and the **"measurement equation,"** respectively. As we shall see, the parameter σ^2 can be concentrated out of the likelihood. In the following we will usually assume $\sigma^2 = 1$ unless otherwise specified.

The first two moments of the distribution of $\{Y_t\}$ are completely defined by the state space model equations (4.1) and (4.2). If the covariance matrix of the sample $Y = (Y_1', \ldots, Y_n')'$ is singular, there are linear combinations of Y that are deterministic. For this reason, we will make the assumption that $\text{Var}(Y)$ is nonsingular unless otherwise specified.

V. Gómez, *Multivariate Time Series With Linear State Space Structure*,
DOI 10.1007/978-3-319-28599-3_4

4.2 The Kalman Filter

When the variables u_t, v_t, and x_1 in the state space model (4.1) and (4.2) have a Gaussian distribution, the Kalman filter recursions produce the conditional expectation $\hat{x}_{t+1|t} = E(x_{t+1}|Y_{1:t})$ and the conditional variance $P_{t+1} = \text{Var}(x_{t+1}|Y_{1:t})$ for $t = 1, \ldots, n$, where $Y_{1:t} = \{Y_1, \ldots, Y_t\}$. If the Gaussian assumption is dropped, the Kalman filter can still be applied, but then the $\hat{x}_{t|t-1}$ are only minimum mean squared estimators of x_t based on $Y_{1:t}$ in the class of linear estimators.

Theorem 4.1 (The Kalman Filter (Kalman, 1960b)) *The Kalman filter corresponding to the sample $\{Y_t : 1 \leq t \leq n\}$ of the state space model (4.1) and (4.2) is given for $t = 1, \ldots, n$ by the recursions*

$$\begin{aligned}
E_t &= Y_t - H_t\hat{x}_{t|t-1}, \quad \Sigma_t = H_tP_tH_t' + R_t \\
K_t &= (F_tP_tH_t' + G_tS_t)\Sigma_t^{-1}, \quad \hat{x}_{t+1|t} = F_t\hat{x}_{t|t-1} + K_tE_t \qquad (4.3)\\
P_{t+1} &= F_tP_tF_t' + G_tQ_tG_t' - (F_tP_tH_t' + G_tS_t)\Sigma_t^{-1}(F_tP_tH_t' + G_tS_t)' \\
&= (F_t - K_tH_t)P_tF_t' + (G_tQ_t - K_tS_t')G_t',
\end{aligned}$$

initialized with $\hat{x}_{1|0} = a$ and $P_1 = \Omega$. In the previous recursions, the E_t are the innovations, Var$(E_t) = \Sigma_t$, the quantity K_t is called the Kalman gain, $\hat{x}_{t|t-1}$ is the orthogonal projection of x_t on $\{Y_1, \ldots, Y_{t-1}\}$, and MSE$(\hat{x}_{t|t-1}) = P_t$.

Proof By the properties of orthogonal projection, the estimator of x_{t+1} based on $Y_{1:t}$ can be obtained by projecting x_{t+1} onto $\{Y_1, \ldots, Y_{t-1}, E_t\} = \{Y_{1:t-1}, E_t\}$, where E_t is the innovation, $E_t = Y_t - E^*(Y_t|Y_{1:t-1})$. Since $Y_{1:t-1}$ and $\{E_t\}$ are orthogonal and $\{E_t\}$ and $\{u_t\}$ are orthogonal to $Y_{1:t-1}$, it is obtained that

$$\begin{aligned}
\hat{x}_{t+1|t} &= E^*(x_{t+1}|Y_{1:t-1}) + \text{Cov}(x_{t+1}, E_t)\text{Var}^{-1}(E_t)E_t \\
&= E^*(F_tx_t + G_tu_t|Y_{1:t-1}) + K_tE_t \\
&= F_t\hat{x}_{t|t-1} + K_tE_t,
\end{aligned}$$

where $K_t = \text{Cov}(x_{t+1}, E_t)\text{Var}^{-1}(E_t)$.

Since $\{v_t\}$ is orthogonal to $Y_{1:t-1}$, we have that $E^*(Y_t|Y_{1:t-1}) = E^*(H_tx_t + v_t|Y_{1:t-1}) = H_t\hat{x}_{t|t-1}$ and $E_t = Y_t - H_t\hat{x}_{t|t-1}$. Since E_t and $\hat{x}_{t|t-1}$ are orthogonal, letting $\text{Var}(E_t) = \Sigma_t$, it follows from this that $E_t = H_t\tilde{x}_t + v_t$, where $\tilde{x}_t = x_t - \hat{x}_{t|t-1}$, and $\Sigma_t = H_tP_tH_t' + R_t$. In addition, since v_t is orthogonal to $\tilde{x}_t$ and $\hat{x}_{t|t-1}$ and $\tilde{x}_t$ is orthogonal to u_t and $\hat{x}_{t|t-1}$,we get

$$\begin{aligned}
\text{Cov}(x_{t+1}, E_t) &= \text{Cov}(x_{t+1}, H_t\tilde{x}_t + v_t) \\
&= \text{Cov}(F_tx_t + G_tu_t, H_t\tilde{x}_t + v_t) \\
&= \text{Cov}(F_t\tilde{x}_t + G_tu_t + F_t\hat{x}_{t|t-1}, H_t\tilde{x}_t + v_t) \\
&= F_tP_tH_t' + G_tS_t,
\end{aligned}$$

and the formula for K_t follows.

To obtain the formula for $\mathrm{MSE}(\hat{x}_{t+1|t}) = P_{t+1}$, consider that

$$\begin{aligned}\tilde{x}_{t+1} + K_tE_t &= x_{t+1} - F_t\hat{x}_{t|t-1}\\ &= F_t\tilde{x}_t + Gu_t\end{aligned}$$

and that $\tilde{x}_{t+1}$ and $\tilde{x}_t$ are orthogonal to E_t and u_t, respectively. □

In the rest of the chapter, unless otherwise stated, we will assume for simplicity and without loss of generality that $E(x_1) = 0$. Therefore, the initial condition for the Kalman filter will be $\hat{x}_{1|0} = 0$.

4.2.1 Innovations Model for the Output Process

If $\{Y_t\}$ follows the state space model (4.1) and (4.2), an immediate consequence of the Kalman filter is the following causal and causally invertible state space model for the output process $\{Y_t\}$

$$\begin{aligned}\hat{x}_{t+1|t} &= F_t\hat{x}_{t|t-1} + K_tE_t\\ Y_t &= H_t\hat{x}_{t|t-1} + E_t,\end{aligned}$$

where $\mathrm{Var}(E_t) = \Sigma_t$ and $\hat{x}_{1|0} = 0$. Note that the model is invertible because we can obtain E_t from Y_t using the recursions

$$\begin{aligned}\hat{x}_{t+1|t} &= (F_t - K_tH_t)\hat{x}_{t|t-1} + K_tY_t\\ E_t &= -H_t\hat{x}_{t|t-1} + Y_t,\end{aligned}$$

where $\hat{x}_{1|0} = 0$.

4.2.2 Triangular Factorizations of $Var(\mathbf{Y_t})$ and $Var^{-1}(\mathbf{Y_t})$

If $\{Y_t\}$ follows the state space model (4.1) and (4.2) and the Kalman filter is applied, the innovations, $E_t, t = 1, \ldots, n$, are serially orthogonal by the properties of orthogonal projection discussed in Chap. 1. Also, if $Y = (Y_1', \ldots, Y_n')', E = (E_1', \ldots, E_n')', \Sigma = \mathrm{diag}(\Sigma_1, \ldots, \Sigma_n)$, and $\mathrm{Var}(Y) = \Sigma_Y$, then $E = L^{-1}Y$, $\mathrm{Var}(E) = \Sigma$ and $\Sigma_Y = L\Sigma L'$, where L is a lower block triangular matrix with blocks of unit matrices in the main diagonal.

The elements of the L and L^{-1} matrices can be obtained by iterating in the Kalman filter. Using $E_t = Y_t - H_t\hat{x}_{t|t-1}$ and $\hat{x}_{t+1|t} = (F_t - K_tH_t)\hat{x}_{t|t-1} + K_tY_t$, it is not difficult to show that

$$L = \begin{bmatrix} I & 0 & \cdots & 0 & 0 \\ H_2K_1 & I & \cdots & 0 & 0 \\ H_3F_2^3K_1 & H_3K_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ H_nF_2^nK_1 & H_nF_3^nK_2 & \cdots & H_nK_{n-1} & I \end{bmatrix}, \tag{4.4}$$

where $F_i^j = F_{j-1}F_{j-2}\cdots F_i$ if $i < j$, $F_i^i = I$, and

$$L^{-1} = \begin{bmatrix} I & 0 & \cdots & 0 & 0 \\ -H_2K_1 & I & \cdots & 0 & 0 \\ -H_3F_{p,2}^3K_1 & -H_3K_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ -H_nF_{p,2}^nK_1 & -H_nF_{p,3}^nK_2 & \cdots & -H_nK_{n-1} & I \end{bmatrix}, \tag{4.5}$$

where $F_{p,i}^j = F_{p,j-1}F_{p,j-2}\cdots F_{p,i}$ if $i < j$, $F_{p,i}^i = I$, and $F_{p,t} = F_t - K_tH_t$. Clearly, $\Sigma_Y^{-1} = L^{'-1}\Sigma^{-1}L^{-1}$.

Example 4.1 Assuming the set up of Example 1.3, one possible set of Eqs. (4.1) and (4.2) is

$$x_{t+1} = \phi x_t + \sigma u_t$$
$$Y_t = x_t, \qquad t = 1, 2, \ldots, n.$$

Thus, $F_t = \phi, H_t = 1, G_t = \sigma, v_t = 0, u_t = A_{t+1}/\sigma, Q_t = 1, R_t = 0, S_t = 0$, and $\Omega = \text{Var}(x_1) = \sigma^2/(1-\phi^2)$. The Kalman filter is initialized with

$$\hat{x}_{1|0} = 0, \qquad P_1 = \Omega,$$

and produces the output $E_1 = Y_1, \Sigma_1 = P_1, K_1 = \phi, \hat{x}_{2|1} = \phi Y_1, P_2 = \sigma^2$, and, for $t > 1$,

$$E_t = Y_t - \phi Y_{t-1}, \quad \Sigma_t = \sigma^2, \quad K_t = \phi, \quad \hat{x}_{t+1|t} = \phi Y_t, \quad P_{t+1} = \sigma^2.$$

◇

4.2.3 Measurement and Time Updates

There are applications where the measurements are made at irregular intervals, for example in tracking satellites using data from stations around the world. To address this issue, Dr. S.F. Schmidt at the NASA Ames Research Center developed in the late 1960s a decomposition of the original Kalman filter problem into a measurement-update problem of going from the predicted estimator, $\hat{x}_{t|t-1}$, to the so-called **filtered estimator**, $\hat{x}_{t|t} = E^*(Y_t|Y_{1:t})$, and a separate time-update problem of going from $\hat{x}_{t|t}$ to $\hat{x}_{t+1|t}$.

Theorem 4.2 (Measurement Update) *Consider the state space model (4.1) and (4.2) and suppose that we have computed the estimator of x_t based on $Y_{1:t-1}, \hat{x}_{t|t-1}$, and its MSE, P_t, and a new measurement, Y_t, becomes available. Then, we can update the estimator $\hat{x}_{t|t-1}$ and P_t, using the formulae*

$$\hat{x}_{t|t} = \hat{x}_{t|t-1} + K_{f,t}E_t \tag{4.6}$$
$$MSE(\hat{x}_{t|t}) = P_{t|t} = P_t - K_{f,t}\Sigma_t K'_{f,t} = P_t - P_t H'_t \Sigma_t^{-1} H_t P_t,$$

where $K_{f,t} = P_t H'_t \Sigma_t^{-1}$.

Proof The estimator of x_t based on $Y_{1:t}$ can be obtained by projecting x_t onto $\{Y_1, \ldots, Y_{t-1}, E_t\} = \{Y_{1:t-1}, E_t\}$. Since $Y_{1:t-1}$ and $\{E_t\}$ are orthogonal, it is obtained that

$$\begin{aligned}\hat{x}_{t|t} &= \hat{x}_{t|t-1} + \mathrm{Cov}(x_t, E_t)\mathrm{Var}^{-1}(E_t)E_t \\ &= \hat{x}_{t|t-1} + K_{f,t}E_t,\end{aligned}$$

where $K_{f,t} = \mathrm{Cov}(x_t, E_t)\mathrm{Var}^{-1}(E_t)$. Let $\tilde{x}_t = x_t - \hat{x}_{t|t-1}$. Then, $E_t = Y_t - H_t\hat{x}_{t|t-1} = H_t\tilde{x}_t + v_t$, v_t is orthogonal to $\tilde{x}_t$ and $\hat{x}_{t|t-1}$, and $\tilde{x}_t$ is orthogonal to $\hat{x}_{t|t-1}$. Thus, we get

$$\mathrm{Cov}(x_t, E_t) = \mathrm{Cov}(x_t, H_t\tilde{x}_t + v_t) = \mathrm{Cov}(\tilde{x}_t + \hat{x}_{t|t-1}, H_t\tilde{x}_t + v_t) = P_t H'_t$$

and the formula for $\hat{x}_{t|t}$ follows because $\mathrm{Var}(E_t) = \Sigma_t$.

To obtain the formula for $\mathrm{MSE}(\hat{x}_{t|t}) = P_{t|t}$, consider that $x_t - \hat{x}_{t|t} + K_{f,t}E_t = x_t - \hat{x}_{t|t-1}$ and that $x_t - \hat{x}_{t|t}$ and E_t are orthogonal. □

Theorem 4.3 (Time Update) *Consider the state space model (4.1) and (4.2) and suppose that we have computed the estimator of x_t based on $Y_{1:t}, \hat{x}_{t|t}$, and its MSE, $P_{t|t}$, and without any further measurements wish to find $\hat{x}_{t+1|t}$ and P_{t+1}. This can be done using the formulae*

$$\hat{x}_{t+1|t} = F_t\hat{x}_{t|t} + G_t\hat{u}_{t|t}, \qquad \hat{u}_{t|t} = S_t\Sigma_t^{-1}E_t,$$
$$P_{t+1} = F_t P_{t|t} F'_t + G_t(Q_t - S_t\Sigma_t^{-1}S'_t)G'_t - F_t K_{f,t} S'_t G'_t - G_t S_t K'_{f,t} F'_t,$$

where $K_{f,t} = P_t H'_t \Sigma_t^{-1}$.

Proof The state equation (4.1) allows us to write $\hat{x}_{t+1|t} = F_t\hat{x}_{t|t} + G_t\hat{u}_{t|t}$, where $\hat{u}_{t|t} = E^*(u_t|Y_{1:t})$ and the estimator of u_t based on $Y_{1:t}$ can be obtained by projecting

u_t onto $\{E_1, \ldots, E_{t-1}, E_t\}$. Since u_t is orthogonal to E_s if $s \leq t-1$, it is obtained that

$$\hat{u}_{t|t} = \mathrm{Cov}(u_t, E_t)\mathrm{Var}^{-1}(E_t)E_t.$$

Let $\tilde{x}_t = x_t - \hat{x}_{t|t-1}$. Then, $E_t = Y_t - H_t\hat{x}_{t|t-1} = H_t\tilde{x}_t + v_t$ and u_t is orthogonal to $\tilde{x}_t$. From this, we get

$$\mathrm{Cov}(u_t, E_t) = \mathrm{Cov}(u_t, H_t\tilde{x}_t + v_t) = S_t$$

and the formula for $\hat{u}_{t|t}$ follows because $\mathrm{Var}(E_t) = \Sigma_t$. The formula for P_{t+1} follows by a straightforward calculation. □

4.2.4 Updating of the Filtered Estimator

An immediate consequence of the measurement and time updates is the following formula to update the filtered estimator

$$\begin{aligned} \hat{x}_{t+1|t+1} &= \hat{x}_{t+1|t} + K_{f,t+1}E_{t+1} \\ &= F_t\hat{x}_{t|t} + G_tS_t\Sigma_t^{-1}E_t + K_{f,t+1}E_{t+1}. \end{aligned}$$

4.2.5 Sequential Processing

An important application of the measurement and time updates of the Kalman filter is the so-called sequential processing to handle the case of multivariate observations. Assume the state space model (4.1) and (4.2) in which $p > 1$ and let $Y_t = (Y_{1t}, \ldots, Y_{pt})', H_t = (h'_{1t}, \ldots, h'_{pt})'$ and $v_t = (v_{1t}, \ldots, v_{pt})'$. Suppose that the matrix R_t is diagonal, $R_t = \mathrm{diag}(r_{1t}, \ldots, r_{pt})$, and that we have obtained $\hat{x}_{t|t-1}$ and its MSE, P_t. To update these estimators when the new measurement, Y_t, is available, the idea is to proceed sequentially, replacing the Eqs. (4.1) and (4.2) with

$$\begin{aligned} x_t^{i+1} &= x_t^i \\ Y_{it} &= h_{it}x_t^i + v_{it}, \qquad i = 1, \ldots, p, \end{aligned} \tag{4.7}$$

where $x_t^1 = x_t$. This is possible because the variables v_{it} in the state space model (4.7) are orthogonal. If we apply the measurement update corresponding to $i = 1$ in the state space model (4.7), starting with $\hat{x}_t^{1|0} = \hat{x}_{t|t-1}$ and $P_t^{1|0} = P_t$, we get

$$\hat{x}_t^{1|1} = \hat{x}_t^{1|0} + K_{f,t}^1 E_{1t}, \qquad P_t^{1|1} = P_t^{1|0} - P_t^{1|0}h'_{1t}\Sigma_{1t}^{-1}h_{1t}P_t^{1|0}, \tag{4.8}$$

where $E_{1t} = Y_{1t} - h_{1t}\hat{x}_t^{1|0}$, $K_{f,t}^1 = P_t^{1|0}h_{1t}'\Sigma_{1t}^{-1}$ and $\Sigma_{1t} = h_{1t}P_t^{1|0}h_{1t}' + r_{1t}$. Applying now the time update to the state space model (4.7), it is obtained that

$$\hat{x}_t^{2|1} = \hat{x}_t^{1|1}, \qquad P_t^{2|1} = P_t^{1|1}.$$

Continuing in this way, we see that we can process Y_t using the state space model (4.7) with the measurement updates

$$\hat{x}_t^{i|i} = \hat{x}_t^{i-1|i-1} + K_{f,t}^i E_{it}, \quad P_t^{i|i} = P_t^{i-1|i-1} - P_t^{i-1|i-1}h_{it}'\Sigma_{it}^{-1}h_{it}P_t^{i-1|i-1}, \quad i = 2, \ldots, p,$$

where

$$E_{it} = Y_{it} - h_{it}\hat{x}_t^{i-1|i-1}, \qquad K_{f,t}^i = P_t^{i-1|i-1}h_{it}'\Sigma_{it}^{-1}, \qquad \Sigma_{it} = h_{it}P_t^{i-1|i-1}h_{it}' + r_{it},$$

initialized with (4.8). At the end of this process, clearly $\hat{x}_{t|t} = \hat{x}_t^{p|p}$ and $P_{t|t} = P_t^{p|p}$. That is, we have performed the multivariate measurement update of $\hat{x}_{t|t-1}$ and its MSE, P_t, by means of p univariate measurement updates.

If the matrix R_t is not diagonal, we can first perform the decomposition $U_tR_tU_t' = D_t$, where U_t is an orthogonal and D_t is a diagonal matrix, and then premultiply the observation equation (4.2) by U_t. After this, the previous procedure can be applied to the transformed observation equation, $Y_t^* = H_t^*x_t + v_t^*$, where $Y_t^* = U_tY_t$, $H_t^* = U_tH_t$ and $v_t^* = U_tv_t$, because $\mathrm{Var}(v_t^*) = D_t$.

4.3 Single Disturbance State Space Representation

In Chap. 1 we introduced the state space model representation

$$x_{t+1} = F_tx_t + G_t\epsilon_t \tag{4.9}$$

$$Y_t = H_tx_t + J_t\epsilon_t, \tag{4.10}$$

where $\{\epsilon_t\}$ is an uncorrelated sequence, $\epsilon_t \sim (0, \sigma^2 I)$, $x_1 \sim (0, \sigma^2\Omega)$, and the notation $c \sim (m, V)$ denotes a random vector c with mean m and covariance matrix V. It is assumed that x_1 is orthogonal to the $\{\epsilon_t\}$ sequence and it is further assumed that $\sigma^2 = 1$ unless otherwise stated. As mentioned in Sect. 1.7, the fact that the same error term appears in both (4.9) and (4.10) does not imply loss of generality in the state space model.

It is not difficult to verify that the Kalman filter corresponding to (4.9) and (4.10) is given for $t = 1, \ldots, n$ by the recursions

$$\begin{aligned} E_t &= Y_t - H_t\hat{x}_{t|t-1}, \qquad \Sigma_t = H_tP_tH_t' + J_tJ_t', \\ K_t &= (F_tP_tH_t' + G_tJ_t')\Sigma_t^{-1}, \qquad \hat{x}_{t+1|t} = F_t\hat{x}_{t|t-1} + K_tE_t, \\ P_{t+1} &= (F_t - K_tH_t)P_tF_t' + (G_t - K_tJ_t)G_t', \end{aligned} \tag{4.11}$$

initialized with $\hat{x}_{1|0} = 0$ and $P_1 = \Omega$.

4.4 Square Root Covariance Filter

Sometimes, the propagation of the covariance matrices P_t in the Kalman filter is not numerically stable. For this reason, it is desirable to propagate square roots of these matrices. Given a nonnegative definite matrix, A, a square root of A is any matrix, $A^{1/2}$, such that

$$A = A^{1/2}A^{1/2'}.$$

The Kalman filter obtained by propagating the square roots of P_t is called the square root covariance filter. It turns out that the square root covariance filter is simpler with the single disturbance state space model (4.9) and (4.10) than if we use (4.1) and (4.2).

4.4.1 Square Root Filter for the Single Disturbance State Space Model

Theorem 4.4 *Suppose that the process $\{Y_t\}$ follows the state space model (4.9) and (4.10). Then, the application of the QR algorithm produces an orthogonal matrix U_t such that*

$$U_t' \begin{bmatrix} P_t^{1/2'}H_t' & P_t^{1/2'}F_t' \\ J_t' & G_t' \end{bmatrix} = \begin{bmatrix} \Sigma_t^{1/2'} & \widehat{K}_t' \\ 0 & P_{t+1}^{1/2'} \end{bmatrix}, \tag{4.12}$$

where $\widehat{K}_t = (F_tP_tH_t' + G_tJ_t')\Sigma_t^{-1/2'} = K_t\Sigma_t^{1/2}$. Thus, letting $\widehat{E}_t = \Sigma_t^{-1/2}E_t, \hat{x}_{t+1|t}$ can be obtained as $\hat{x}_{t+1|t} = F_t\hat{x}_{t|t-1} + \widehat{K}_t\widehat{E}_t$.

Proof The matrix U_t satisfies

$$U_t' \begin{bmatrix} P_t^{1/2'}H_t' & P_t^{1/2'}F_t' \\ J_t' & G_t' \end{bmatrix} = \begin{bmatrix} \Sigma' & K' \\ 0 & P' \end{bmatrix}.$$

Premultiplying the matrices in (4.12) by their respective transposes yields

$$H_tP_t^{1/2}P_t^{1/2'}H_t' + J_tJ_t' = \Sigma\Sigma', \quad H_tP_t^{1/2}P_t^{1/2'}F_t' + J_tG_t' = \Sigma K'$$
$$F_tP_t^{1/2}P_t^{1/2'}F_t' + G_tG_t' = KK' + PP',$$

and the theorem follows. □

4.4.2 Fast Square Root Filter for the Single Disturbance State Space Model

Letting $P_t = L_t D_t L_t'$ and $\Sigma_t = l_t d_t l_t'$, where D_t and d_t are diagonal matrices with non negative elements in the main diagonal and L_t and l_t are lower triangular matrices with ones in the main diagonal, we can use fast Givens rotations in Theorem 4.4 to get

$$U_t' \begin{bmatrix} D_t^{1/2} & \\ & I \end{bmatrix} \begin{bmatrix} L_t'H_t' & L_t'F_t' \\ J_t' & G_t' \end{bmatrix} = \begin{bmatrix} d_t^{1/2} & \\ & D_{t+1}^{1/2} \end{bmatrix} \begin{bmatrix} l_t' & d_t^{-1/2}\widehat{K}_t' \\ 0 & L_{t+1}' \end{bmatrix}.$$

4.4.3 Square Root Filter for the Several Sources of Error State Space Model

It is possible to derive a square root covariance filter when the process $\{Y_t\}$ follows the state space model (4.1) and (4.2). We will derive this filter under the assumption that Q_t is nonsingular. This does not imply any loss of generality however because of the presence of the matrix G_t. To see this, assume that Q_t is singular and let O_t be an orthogonal matrix such that $O_t'Q_tO_t = D_t$, where D_t is a diagonal matrix with the diagonal elements in decreasing order of absolute value so that the last elements are zero. Let $D_t = \text{diag}(\overline{D}_t, 0)$, where $\overline{D}_t$ contains all the nonzero elements of D, and define $\bar{u}_t = U_t u_t$, where $\text{Var}(\bar{u}_t) = \overline{D}_t$ and U_t is the submatrix of O_t that contains the first $\text{rank}(D_t)$ columns of this matrix. Then, $u_t = U_t'\bar{u}_t$ and we can redefine u_t, G_t, Q_t and S_t as $\bar{u}_t, G_tU_t', \overline{D}_t$ and U_tS_t, respectively. In the following, for a symmetric matrix, S, the notation $S > 0$ means that this matrix is positive definite.

Theorem 4.5 *Suppose that the process $\{Y_t\}$ follows the state space model (4.1) and (4.2) and that $Q_t > 0$, and define $R_t^s = R_t - S_t'Q_t^{-1}S_t$. Then, the application of the QR algorithm yields an orthogonal matrix U_t such that*

$$U_t' \begin{bmatrix} (R_t^s)^{1/2'} & 0 \\ P_t^{1/2'}H_t' & P_t^{1/2'}F_t' \\ Q_t^{-1/2}S_t & Q_t^{1/2'}G_t' \end{bmatrix} = \begin{bmatrix} \Sigma_t^{1/2'} & \widehat{K}_t' \\ 0 & P_{t+1}^{1/2'} \\ 0 & 0 \end{bmatrix}, \tag{4.13}$$

where $\widehat{K}_t = (F_tP_tH_t' + G_tS_t)\Sigma_t^{-1/2'} = K_t\Sigma_t^{1/2}$. Thus, letting $\widehat{E}_t = \Sigma_t^{-1/2}E_t, \hat{x}_{t+1|t}$ can be obtained as $\hat{x}_{t+1|t} = F_t\hat{x}_{t|t-1} + \widehat{K}_t\widehat{E}_t$. In addition, if R_t^s and P_t are nonsingular, the same matrix U_t satisfies

$$U_t' \left[\begin{array}{cc|c} (R_t^s)^{1/2'} & 0 & -(R_t^s)^{-1/2}Y_t \\ P_t^{1/2'}H_t' & P_t^{1/2'}F_t' & P_t^{-1/2}\hat{x}_{t|t-1} \\ Q_t^{-1/2}S_t & Q_t^{1/2'}G_t' & 0 \end{array}\right] = \left[\begin{array}{cc|c} \Sigma_t^{1/2'} & \widehat{K}_t' & -\widehat{E}_t \\ 0 & P_{t+1}^{1/2'} & P_{t+1}^{-1/2}\hat{x}_{t+1|t} \\ 0 & 0 & * \end{array}\right], \tag{4.14}$$

where the asterisk indicates an element that is not relevant to our purposes. In this case, $\hat{x}_{t+1|t}$ *is obtained as* $\hat{x}_{t+1|t} = P_{t+1}^{1/2}\left[P_{t+1}^{-1/2}\hat{x}_{t+1|t}\right]$.

Proof The matrix U_t satisfies

$$U_t'\begin{bmatrix} (R_t^s)^{1/2'} & 0 \\ P_t^{1/2'}H_t' & P_t^{1/2'}F_t' \\ Q_t^{-1/2}S_t & Q_t^{1/2'}G_t' \end{bmatrix} = \begin{bmatrix} \Sigma' & K' \\ 0 & P' \\ 0 & 0 \end{bmatrix}.$$

Premultiplying the matrices in (4.13) by their respective transposes yields

$$\begin{aligned} (R_t^s)^{1/2}(R_t^s)^{1/2'} + H_tP_t^{1/2}P_t^{1/2'}H_t' + S_t'Q_t^{-1/2'}Q_t^{-1/2}S_t &= \Sigma\Sigma' \\ H_tP_t^{1/2}P_t^{1/2'}F_t' + S_t'Q_t^{-1/2}Q_t^{1/2'}G_t' &= \Sigma K', \\ F_tP_t^{1/2}P_t^{1/2'}F_t' + G_tQ_t^{1/2}Q_t^{1/2'}G_t' &= KK' + PP', \end{aligned}$$

and the first part of the theorem follows. To prove the second part, consider the first and the third and the second and the third block columns of (4.14). Then, it is obtained that

$$\begin{aligned} -(R_t^s)^{1/2}(R_t^s)^{-1/2}Y_t + H_tP_t^{1/2}P_t^{-1/2}\hat{x}_{t|t-1} &= -\Sigma_t^{1/2}\widehat{E}_t, \\ F_tP_t^{1/2}P_t^{-1/2}\hat{x}_{t|t-1} &= P_{t+1}^{1/2}P_{t+1}^{-1/2}\hat{x}_{t+1|t} - \widehat{K}_t\widehat{E}_t. \end{aligned}$$

□

4.4.4 Measurement Update in Square Root Form

We will first develop the equations for the several sources of error state space model (4.1) and (4.2). The equations for the single disturbance state space model (4.9) and (4.10) follow easily from these.

4.4.4.1 Several Sources of Error State Space Model

Theorem 4.6 *Suppose that the process* $\{Y_t\}$ *follows the state space model (4.1) and (4.2). Then, the application of the QR algorithm produces an orthogonal matrix* U_t *such that*

$$U_t'\begin{bmatrix} P_t^{1/2'}H_t' & P_t^{1/2'} \\ R_t^{1/2'} & 0 \end{bmatrix} = \begin{bmatrix} \Sigma_t^{1/2'} & \widehat{K}_{f,t}' \\ 0 & P_{t|t}^{1/2'} \end{bmatrix}, \tag{4.15}$$

where $\widehat{K}_{f,t} = P_t H_t' \Sigma_t^{-1/2'} = K_{f,t}\Sigma_t^{1/2}$*. Thus, letting* $\widehat{E}_t = \Sigma_t^{-1/2} E_t$*,* $\hat{x}_{t|t}$ *can be obtained as* $\hat{x}_{t|t} = \hat{x}_{t|t-1} + \widehat{K}_{f,t}\widehat{E}_t$*. In addition, if* R_t *and* P_t *are nonsingular, the same matrix* U_t *satisfies*

$$U_t' \left[\begin{array}{cc|c} P_t^{1/2'} H_t' & P_t^{1/2'} & P_t^{-1/2}\hat{x}_{t|t-1} \\ R_t^{1/2'} & 0 & -R_t^{-1/2} Y_t \end{array} \right] = \left[\begin{array}{cc|c} \Sigma_t^{1/2'} & \widehat{K}_{f,t}' & -\widehat{E}_t \\ 0 & P_{t|t}^{1/2'} & P_{t|t}^{-1/2}\hat{x}_{t|t} \end{array} \right].$$

In this case, $\hat{x}_{t|t}$ *is obtained as* $\hat{x}_{t|t} = P_{t|t}^{1/2}\left[P_{t|t}^{-1/2}\hat{x}_{t|t}\right]$*.*

Proof The matrix U_t satisfies

$$U_t' \begin{bmatrix} P_t^{1/2'} H_t' & P_t^{1/2'} \\ R_t^{1/2'} & 0 \end{bmatrix} = \begin{bmatrix} \Sigma' & K' \\ 0 & P' \end{bmatrix}.$$

Premultiplying the matrices in the previous equality by their respective transposes yields

$$R_t^{1/2} R_t^{1/2'} + H_t P_t^{1/2} P_t^{1/2'} H_t' = \Sigma\Sigma', \quad H_t P_t^{1/2} P_t^{1/2'} = \Sigma K'$$
$$P_t^{1/2} P_t^{1/2'} = KK' + PP',$$

and the first part of the theorem follows. The rest is proved similarly. □

4.4.4.2 Single Disturbance State Space Model

The following theorem is an immediate consequence of Theorem 4.6.

Theorem 4.7 *Suppose that the process* $\{Y_t\}$ *follows the state space model (4.9) and (4.10). Then, the application of the QR algorithm produces an orthogonal matrix* U_t *such that*

$$U_t' \begin{bmatrix} P_t^{1/2'} H_t' & P_t^{1/2'} \\ J_t' & 0 \end{bmatrix} = \begin{bmatrix} \Sigma_t^{1/2'} & \widehat{K}_{f,t}' \\ 0 & P_{t|t}^{1/2'} \end{bmatrix}, \tag{4.16}$$

where $\widehat{K}_{f,t} = P_t H_t' \Sigma_t^{-1/2'} = K_{f,t}\Sigma_t^{1/2}$*. Thus, letting* $\widehat{E}_t = \Sigma_t^{-1/2} E_t$*,* $\hat{x}_{t|t}$ *can be obtained as* $\hat{x}_{t|t} = \hat{x}_{t|t-1} + \widehat{K}_{f,t}\widehat{E}_t$*. In addition, if* R_t *and* P_t *are nonsingular, the same matrix* U_t *satisfies*

$$U_t' \left[\begin{array}{cc|c} P_t^{1/2'} H_t' & P_t^{1/2'} & P_t^{-1/2}\hat{x}_{t|t-1} \\ J_t^{1/2} & 0 & -R_t^{-1/2} Y_t \end{array} \right] = \left[\begin{array}{cc|c} \Sigma_t^{1/2'} & \widehat{K}_{f,t}' & -\widehat{E}_t \\ 0 & P_{t|t}^{1/2'} & P_{t|t}^{-1/2}\hat{x}_{t|t} \end{array} \right].$$

In this case, $\hat{x}_{t|t}$ *is obtained as* $\hat{x}_{t|t} = P_{t|t}^{1/2}\left[P_{t|t}^{-1/2}\hat{x}_{t|t}\right]$*.*

4.4.5 Fast Square Root Algorithms for Measurement Update: The UD Filter

We will consider first the case in which the observations are scalar. The general case can be reduced to this by using sequential processing.

The equations for the measurement update,

$$\Sigma_t = R_t + H_t P_t H_t'$$
$$P_{t|t} = P_t - P_t H_t' \Sigma_t^{-1} H_t P_t,$$

are practically identical to the update of the P_t matrix in RLS. Thus, we can proceed as in Chap. 2 to obtain fast square root algorithms for measurement update. Let $P_t = L_t D_t L_t'$, where L_t is a lower triangular matrix with ones in the main diagonal and D_t is a diagonal matrix with positive elements in the main diagonal and let $P_{t|t} = L_{t|t} D_{t|t} L_{t|t}'$, where $L_{t|t}$ and $D_{t|t}$ are defined similarly. Then, we can write expression (4.16) of Theorem 4.6 as

$$U_t' \begin{bmatrix} 0 & D_t^{1/2} \\ R_t^{1/2'} & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ L_t' H_t' & L_t' \end{bmatrix} = \begin{bmatrix} \Sigma_t^{1/2'} & 0 \\ 0 & D_{t|t}^{1/2} \end{bmatrix} \begin{bmatrix} 1 & K_{f,t}' \\ 0 & L_{t|t}' \end{bmatrix}. \tag{4.17}$$

Then, it is clear that we can use square root free fast Givens rotations, as described in the Appendix of Chap. 2 to obtain the *QDU* decomposition, to perform the measurement update.

Another fast algorithm that can be used for the square root form of measurement update is the so-called *UD* filter, due to Bierman (1977). As mentioned in Chap. 2, this algorithm assumes that the covariance matrix P_t is factored in the form $P_t = U_t D_t U_t'$, where U_t is an upper triangular matrix with ones in the main diagonal and D_t is a diagonal matrix with positive elements in the main diagonal. We will describe the algorithm for the decomposition in terms of lower triangular matrices instead, for coherence with the rest of the book. However, this amounts to a small change in the algorithm.

Let $P_t = P = LDL'$ and $P_{t|t} = P^+ = L_+ D_+ L_+'$. Then, as in the case of RLS, we have to factor

$$L_+ D_+ L_+' = L(D - L'h'\Sigma^{-1}hLD)L',$$

where $h = H_t$ and $\Sigma = \Sigma_t$. Since $L'h'\Sigma^{-1}hLD$ has rank one, Bierman uses a rank one downdating formula due to Agee & Turner (1972) for the factorization of the term in parenthesis in the previous expression. The whole procedure is described in Bierman (1977). Letting $hL = f = [f_1, \ldots, f_n], L = [l_1, \ldots, l_n], L_+ =$

$[l_1^+, \ldots, l_n^+]$, $R = R_t$, $D = \text{diag}(d_i)$ and $D_+ = \text{diag}(d_i^+)$, the algorithm is as follows.

$$\begin{aligned}
&\alpha_{n+1} = R \\
&k_{n+1} = 0 \\
&\text{for } i = n, n-1, \ldots, 1 \\
&\qquad \alpha_i = \alpha_{i+1} + d_i f_i^2 \\
&\qquad d_i^+ = d_i(\alpha_{i+1}/\alpha_i) \\
&\qquad [k_i \quad l_i^+] = [k_{i+1} \quad l_i] \begin{bmatrix} 1 & -f_i/\alpha_{i+1} \\ d_i f_i & 1 \end{bmatrix} \\
&\text{end}
\end{aligned}$$

On completion of the algorithm, $\alpha_1 = \Sigma = R + hPh'$, $k_1 = Ph'$ and $K_t = k_1/\alpha_1$.

As noted by Jover & Kailath (1986), the *UD* filter is equivalent to the updating (4.17), that uses square root free fast Givens rotations.

It is to be noted that Bierman (1977) emphasized that, with careful programming, the number of computations of the *UD* filter is approximately the same as that of the measurement update.

In the case of vector observations, we can still use fast Givens rotations as follows. Let $P_t = L_t D_t L_t'$, where L_t and D_t are as described earlier, and let $R_t = L_{R,t} D_{R,t} L_{R,t}'$ and $\Sigma_t = L_{\Sigma,t} D_{\Sigma,t} L_{\Sigma,t}'$, where $L_{R,t}$, $D_{R,t}$, $L_{\Sigma,t}$, and $D_{\Sigma,t}$ are defined similarly to L_t and D_t. Then, we can write expression (4.16) of Theorem 4.6 as

$$U_t' \begin{bmatrix} 0 & D_t^{1/2} \\ D_{R,t}^{1/2} & 0 \end{bmatrix} \begin{bmatrix} L_{R,t}' & 0 \\ L_t' H_t' & L_t' \end{bmatrix} = \begin{bmatrix} D_{\Sigma,t}^{1/2} & 0 \\ 0 & D_{t|t}^{1/2} \end{bmatrix} \begin{bmatrix} L_{\Sigma,t}' & K_{f,t}' \\ 0 & L_{t|t}' \end{bmatrix}.$$

It is clear from this that we can use square root free fast Givens rotations, as described in the Appendix of Chap. 2 to obtain the *QDU* decomposition, to perform the measurement update.

4.4.6 Time Update in Square Root Form

Time updates in square root form are available only in the special case $S_t = 0$. However, this implies no loss of generality as we shall see later in Sect. 4.5.

Theorem 4.8 *Suppose that the process $\{Y_t\}$ follows the state space model (4.1) and (4.2) and that $S_t = 0$. Then, the application of the QR algorithm produces an orthogonal matrix U_t such that*

$$U_t' \begin{bmatrix} P_{t|t}^{1/2'} F_t' \\ Q_t^{1/2'} G_t' \end{bmatrix} = \begin{bmatrix} P_{t+1|t}^{1/2'} \\ 0 \end{bmatrix} \tag{4.18}$$

and $\hat{x}_{t+1|t}$ *can be obtained as* $\hat{x}_{t+1|t} = F_t\hat{x}_{t|t}$*. In addition, if* P_t *is nonsingular, the same matrix* U_t *satisfies*

$$U_t'\left[\begin{array}{c|c} P_{t|t}^{1/2'}F_t' & P_{t|t}^{-1/2}\hat{x}_{t|t} \\ Q_t^{1/2'}G_t' & 0 \end{array}\right] = \left[\begin{array}{c|c} P_{t+1}^{1/2'} & P_{t+1}^{-1/2}\hat{x}_{t+1|t} \\ 0 & * \end{array}\right],$$

where the asterisk indicates an element that is not relevant to our purposes. In this case, $\hat{x}_{t+1|t}$ *is obtained as* $\hat{x}_{t+1|t} = P_{t+1}^{1/2}\left[P_{t+1}^{-1/2}\hat{x}_{t+1|t}\right]$.

Proof The matrix U_t satisfies

$$U_t'\begin{bmatrix} P_{t|t}^{1/2'}F_t' \\ Q_t^{1/2'}G_t' \end{bmatrix} = \begin{bmatrix} P' \\ 0 \end{bmatrix}.$$

Premultiplying the matrices in the previous equality by their respective transposes yields

$$F_tP_t^{1/2}P_t^{1/2'}F_t' + G_tQ_t^{1/2}Q_t^{1/2'}G_t' = PP',$$

and the first part of the theorem follows. The rest is proved similarly. □

4.4.7 Fast Square Root Algorithms for Time Update

As in the previous section, suppose that $S_t = 0$ in the state space model (4.1) and (4.2). Then, the time update formula for P_t is

$$P_{t+1} = F_tP_{t|t}F_t' + G_tQ_tG_t'.$$

Without loss of generality, suppose also that Q_t is nonsingular. Let $P_t = L_tD_tL_t'$, where L_t is a lower triangular matrix with ones in the main diagonal and D_t is a diagonal matrix with positive elements in the main diagonal, and let $P_{t|t} = L_{t|t}D_{t|t}L_{t|t}'$ and $Q_t = L_{Q,t}D_{Q,t}L_{Q,t}'$, where $L_{t|t}$, $D_{t|t}$, $L_{Q,t}$, and $D_{Q,t}$ are defined similarly to L_t and D_t. Then, we can write expression (4.18) of Theorem 4.8 as

$$U_t'\begin{bmatrix} D_{t|t}^{1/2} & 0 \\ 0 & D_{Q,t}^{1/2} \end{bmatrix}\begin{bmatrix} L_{t|t}'F_t' \\ L_{Q,t}'G_t' \end{bmatrix} = \begin{bmatrix} D_{t+1}^{1/2}L_{t+1}' \\ 0 \end{bmatrix}.$$

It is clear from this that we can use square root free fast Givens rotations, as described in the Appendix of Chap. 2 to obtain the *QDU* decomposition, to perform the measurement update.

Another fast algorithm that can be used for the square root form of time update is the fast weighted Gram–Schmidt procedure due to C. Thornton, described in, for example, Bierman (1977).

4.5 A Transformation to get $S_t = 0$

In this section, we will see that it is possible to make a transformation in the state space model (4.1) and (4.2) so that we can work with an equivalent state space model in which the disturbance variables are uncorrelated. More specifically, we can pass from the variables $\{v_t, u_t\}$ to the variables $\{v_t, u_t^s\}$ such that v_t is uncorrelated to u_t^s and such that the corresponding state space model is equivalent to the original one. If $R_t > 0$, this can be achieved by applying the Gram–Schmidt procedure. Thus, if we define

$$u_t^s = u_t - \text{Cov}(u_t, v_t)\text{Var}^{-1}(v_t)v_t = u_t - S_t R_t^{-1} v_t,$$

then

$$\text{Cov}(u_t^s, v_t) = 0, \qquad \text{Var}(u_t^s) = Q_t - S_t R_t^{-1} S_t' = Q_t^s.$$

If R_t is singular, by Proposition 1.2, the orthogonal projection $E^*(u_t|v_t)$ exists, is unique, and is given by $E^*(u_t|v_t) = S_t R_t^- v_t$, where R_t^- is a generalized inverse of R_t. Thus, in this case the transformation is

$$u_t^s = u_t - S_t R_t^- v_t,$$

and u_t^s satisfies

$$\text{Cov}(u_t^s, v_t) = 0, \qquad \text{Var}(u_t^s) = Q_t - S_t R_t^- S_t' = Q_t^s. \tag{4.19}$$

Using u_t^s, we can rewrite the state space model (4.1) and (4.2) as

$$\begin{aligned} x_{t+1} &= F_t x_t + G_t(u_t^s + S_t R_t^- v_t), \\ Y_t &= H_t x_t + v_t. \end{aligned}$$

Since $v_t = Y_t - H_t x_t$, we can further transform it into

$$x_{t+1} = F_t^s x_t + G_t u_t^s + G_t S_t R_t^- Y_t, \tag{4.20}$$

$$Y_t = H_t x_t + v_t, \tag{4.21}$$

where

$$F_t^s = F_t - G_t S_t R_t^- H_t. \tag{4.22}$$

The notable feature of the state space model (4.20) and (4.21) is that the disturbances, u_t^s and v_t, are uncorrelated. The presence of the term $G_tS_tR_t^{-}Y_t$ in the transition equation (4.20) causes no problem for prediction and filtering of the state vector because this term is a known function of the observations. In fact, it can be easily shown that the Kalman filter corresponding to (4.20) and (4.21) is given by the recursions

$$\begin{aligned} E_t &= Y_t - H_t\hat{x}_{t|t-1}, \quad \Sigma_t = H_tP_tH_t' + R_t \\ K_t^s &= F_t^sP_tH_t'\Sigma_t^{-1}, \quad \hat{x}_{t+1|t} = F_t^s\hat{x}_{t|t-1} + K_t^sE_t + G_tS_tR_t^{-}Y_t \\ P_{t+1} &= F_t^sP_tF_t^{s'} + G_tQ_t^sG_t' - F_t^sP_tH_t'\Sigma_t^{-1}H_tP_tF_t^{s'}, \end{aligned} \tag{4.23}$$

where Q_t^s and F_t^s are given by (4.19) and (4.22). In addition, it is not difficult to show that the measurement update corresponding to (4.20) and (4.21) coincides with that of (4.1) and (4.2), whereas the time update is

$$\hat{x}_{t+1|t} = F_t^s\hat{x}_{t|t} + G_tS_tR_t^{-}Y_t \tag{4.24}$$

$$P_{t+1} = F_t^sP_{t|t}F_t^{s'} + G_tQ_t^sG_t'. \tag{4.25}$$

4.5.1 Another Expression for the Square Root Covariance Filter

Theorem 4.9 *Suppose that the process $\{Y_t\}$ follows the state space model (4.1) and (4.2). Then, the application of the QR algorithm produces an orthogonal matrix U_t such that*

$$U_t'\begin{bmatrix} R_t^{1/2'} & 0 \\ P_t^{1/2'}H_t' & P_t^{1/2'}F_t^{s'} \\ 0 & (Q_t^s)^{1/2'}G_t' \end{bmatrix} = \begin{bmatrix} \Sigma_t^{1/2'} & \widehat{K}_t^{s'} \\ 0 & P_{t+1}^{1/2'} \\ 0 & 0 \end{bmatrix}, \tag{4.26}$$

where $\widehat{K}_t^s = F_t^sP_tH_t'\Sigma_t^{-1/2'} = K_t^s\Sigma_t^{1/2}, K_t^s = F_t^sP_tH_t'$, and Q_t^s and F_t^s are given by (4.19) and (4.22). Thus, letting $\widehat{E}_t = \Sigma_t^{-1/2}E_t, \hat{x}_{t+1|t}$ can be obtained as $\hat{x}_{t+1|t} = F_t^s\hat{x}_{t|t-1} + \widehat{K}_t^s\widehat{E}_t + G_tS_tR_t^{-1}Y_t$. In addition, if R_t, Q_t^s, and P_t are nonsingular, the same matrix U_t satisfies

$$U_t'\left[\begin{array}{cc|c} R_t^{1/2'} & 0 & -R_t^{-1/2}Y_t \\ P_t^{1/2'}H_t' & P_t^{1/2'}F_t^{s'} & P_t^{-1/2}\hat{x}_{t|t-1} \\ 0 & (Q_t^s)^{1/2'}G_t' & (Q_t^s)^{-1/2}S_tR_t^{-1}Y_t \end{array}\right] = \left[\begin{array}{cc|c} \Sigma_t^{1/2'} & \widehat{K}_t^{s'} & -\widehat{E}_t \\ 0 & P_{t+1}^{1/2'} & P_{t+1}^{-1/2}\hat{x}_{t+1|t} \\ 0 & 0 & * \end{array}\right].$$

In this case, $\hat{x}_{t+1|t}$ can be obtained as $\hat{x}_{t+1|t} = P_{t+1}^{1/2}\left[P_{t+1}^{-1/2}\hat{x}_{t+1|t}\right]$.

Proof The matrix U_t satisfies

$$U_t' \begin{bmatrix} R_t^{1/2'} & 0 \\ P_t^{1/2'} H_t' & P_t^{1/2'} F_t' \\ 0 & Q_t^{1/2'} G_t' \end{bmatrix} = \begin{bmatrix} \Sigma' & K' \\ 0 & P' \\ 0 & 0 \end{bmatrix}.$$

Premultiplying the matrices in (4.26) by their respective transposes yields

$$R_t^{1/2} R_t^{1/2'} + H_t P_t^{1/2} P_t^{1/2'} H_t' = \Sigma \Sigma', \quad H_t P_t^{1/2} P_t^{1/2'} F_t^{s'} = \Sigma K',$$
$$F_t P_t^{1/2} P_t^{1/2'} F_t' + G_t Q_t^{1/2} Q_t^{1/2'} G_t' = KK' + PP',$$

and the first part of the theorem follows. The rest is proved similarly. □

4.5.2 Measurement and Time Updates in Square Root Form When $S_t \neq 0$

When $S_t = 0$, it is possible to separate the square root covariance filter into measurement and time update steps. However, as noted earlier, the general case in which $S_t \neq 0$ can be reduced to this case by using the transformation that passes from the state space model (4.1) and (4.2)–(4.20) and (4.21). When $S_t \neq 0$, the measurement update is as in Sect. 4.4.4 and the time update is given by the following theorem.

Theorem 4.10 *Suppose that the process $\{Y_t\}$ follows the state space model (4.1) and (4.2). Then, the application of the QR algorithm produces an orthogonal matrix U_t such that*

$$U_t' \begin{bmatrix} P_{t|t}^{1/2'} F_t^{s'} \\ (Q_t^s)^{1/2'} G_t' \end{bmatrix} = \begin{bmatrix} P_{t+1|t}^{1/2'} \\ 0 \end{bmatrix}$$

and $\hat{x}_{t+1|t}$ can be obtained as $\hat{x}_{t+1|t} = F_t^s \hat{x}_{t|t} + G_t S_t R_t^- Y_t$. In addition, if Q_t^s and P_t are nonsingular, the same matrix U_t satisfies

$$U_t' \left[\begin{array}{c|c} P_{t|t}^{1/2'} F_t^{s'} & P_{t|t}^{-1/2} \hat{x}_{t|t} \\ (Q_t^s)^{1/2'} G_t' & (Q_t^s)^{-1/2} S_t R_t^- Y_t \end{array} \right] = \left[\begin{array}{c|c} P_{t+1}^{1/2'} & P_{t+1}^{-1/2} \hat{x}_{t+1|t} \\ 0 & * \end{array} \right],$$

where the asterisk indicates an element that is not relevant to our purposes. In this case, $\hat{x}_{t+1|t}$ can be obtained as $\hat{x}_{t+1|t} = P_{t+1}^{1/2} \left[P_{t+1}^{-1/2} \hat{x}_{t+1|t} \right]$.

Proof The matrix U_t satisfies

$$U_t'\begin{bmatrix} P_{t|t}^{1/2'}F_t^{s'} \\ (Q_t^s)^{1/2'}G_t' \end{bmatrix} = \begin{bmatrix} P' \\ 0 \end{bmatrix}.$$

Premultiplying the matrices in the previous equality by their respective transposes yields

$$F_t^s P_t^{1/2} P_t^{1/2'} F_t^{s'} + G_t (Q_t^s)^{1/2} (Q_t^s)^{1/2'} G_t' = PP',$$

and the first part of the theorem follows. The rest is proved similarly. □

4.6 Information Filter

When the initial covariance matrix, Ω, of the Kalman filter has very large entries, it is often preferable to propagate P_t^{-1}, which can be done if $\Omega > 0, R_t > 0$, and the F_t are invertible as we shall see in this section. Since the inverse of the variance of a parameter is usually considered as a measure of the information contained in the parameter because a large variance means high uncertainty, the filter that is obtained by propagating P_t^{-1} is usually called the information filter. The information filter formulae are rather complicated in the general case, so they are generally represented separately in terms of measurement and time updates.

To obtain the information filter, the following lemma will be needed. We leave the proof as an exercise. See Problem 4.1.

Lemma 4.1 (The Matrix Inversion Lemma) *If A and C are nonsingular matrices and A, B, C, and D are matrices of appropriate dimensions, then*

$$(A + BCD)^{-1} = A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1}.$$

The following lemma gives a sufficient condition for the existence of P_t^{-1}.

Lemma 4.2 (Sufficient Condition for the Existence of $\mathbf{P_t^{-1}}$) *Consider the state space model (4.1) and (4.2) and suppose that $\Omega > 0, S_t = 0, R_t > 0$ and the F_t are nonsingular. Then, $P_t > 0$ and, hence, it is invertible.*

Proof Using the Matrix Inversion Lemma 4.1, we can rearrange the recursion for P_t in the Kalman filter as follows

$$P_{t+1} = G_t Q_t G_t' + F_t(P_t^{-1} + H_t' R_t^{-1} H_t)^{-1} F_t',$$

which shows that $P_{t+1} \geq F_t(P_t^{-1} + H_t' R_t^{-1} H_t)^{-1} F_t' > 0$, where the last inequality depends on the nonsingularity of F_t. Since $P_1 = \Omega$, we lemma is proved by induction. □

4.6.1 Measurement Update in Information Form

The measurement update equations are based on the propagation of the inverse of $P_{t|t}$.

Theorem 4.11 (Measurement Update in Information Form) *Consider the state space model (4.1) and (4.2) and suppose that we have computed the estimator of x_t based on $Y_{1:t-1}$, $\hat{x}_{t|t-1}$, and its MSE, P_t, and a new measurement, Y_t, becomes available. Then, if $R_t > 0$ and P_t^{-1} exists, we can write*

$$P_{t|t}^{-1}\hat{x}_{t|t} = P_t^{-1}\hat{x}_{t|t-1} + H_t'R_t^{-1}Y_t \tag{4.27}$$

$$P_{t|t}^{-1} = P_t^{-1} + H_t'R_t^{-1}H_t. \tag{4.28}$$

Proof By applying the Matrix Inversion Lemma 4.1 to the measurement update formula for $P_{t|t}$, it is obtained that

$$P_{t|t} = P_t - P_tH_t'(R_t + H_tP_tH_t')^{-1}H_tP_t = (P_t^{-1} + H_t'R_t^{-1}H_t)^{-1}.$$

Using this formula and the Matrix Inversion Lemma again, we can write

$$\begin{aligned} K_{f,t} &= P_tH_t'(R_t + H_tP_tH_t')^{-1} \\ &= P_tH_t'\left[R_t^{-1} - R_t^{-1}H_t(P_t^{-1} + H_t'R_t^{-1}H_t)^{-1}H_t'R_t^{-1}\right] \\ &= P_t\left[I - H_t'R_t^{-1}H_tP_{t|t}\right]H_t'R_t^{-1} \\ &= P_t\left[I - (P_{t|t}^{-1} - P_t^{-1})P_{t|t}\right]H_t'R_t^{-1} = P_{t|t}H_t'R_t^{-1}. \end{aligned} \tag{4.29}$$

It follows from this that

$$P_{t|t}^{-1}\hat{x}_{t|t} = P_{t|t}^{-1}\left[(I - K_{f,t}H_t)\hat{x}_{t|t-1} + K_{f,t}Y_t\right] = P_t^{-1}\hat{x}_{t|t-1} + H_t'R_t^{-1}Y_t.$$

□

4.6.2 Time Update in Information Form

Theorem 4.12 (Time Update in Information Form) *Consider the state space model (4.1) and (4.2) and suppose that we have computed the estimator of x_t based on $Y_{1:t}$, $\hat{x}_{t|t}$, and its MSE, $P_{t|t}$, and without any further measurements wish to find $\hat{x}_{t+1|t}$ and P_{t+1}. Then, if $R_t > 0$ and $(F_t^s)^{-1}$, $(Q_t^s)^{-1}$ and $P_{t|t}^{-1}$ exist, the following relations hold*

$$P_{t+1}^{-1}\hat{x}_{t+1|t} = \left(I - A_tG_tQ_t^rG_t'\right)\left[(F_t^s)^{-1'}P_{t|t}^{-1}\hat{x}_{t|t} + A_tG_tS_tR_t^{-1}Y_t\right] \tag{4.30}$$

$$P_{t+1}^{-1} = A_t - A_tG_tQ_t^rG_t'A_t, \tag{4.31}$$

where

$$A_t = (F_t^s)^{-1'} P_{t|t}^{-1} (F_t^s)^{-1}, \quad Q_t^r = \left[(Q_t^s)^{-1} + G_t' A_t G_t \right]^{-1},$$

and Q_t^s and F_t^s are given by (4.19) and (4.22).

Proof By applying the Matrix Inversion Lemma 4.1 to the measurement update formula (4.25), it is obtained that

$$P_{t+1}^{-1} = (F_t^s)^{-1'} P_{t|t}^{-1} (F_t^s)^{-1} - (F_t^s)^{-1'} P_{t|t}^{-1} (F_t^s)^{-1} G_t Q_t^r G_t' (F_t^s)^{-1'} P_{t|t}^{-1} (F_t^s)^{-1}.$$

Using this formula, the Matrix Inversion Lemma 4.1 and (4.24), we can write

$$\begin{aligned} P_{t+1}^{-1} \hat{x}_{t+1|t} &= \left(I - A_t G_t Q_t^r G_t' \right) A_t F_t^s \hat{x}_{t|t} + \left(A_t - A_t G_t Q_t^r G_t' A_t \right) G_t S_t R_t^{-1} Y_t \\ &= \left(I - A_t G_t Q_t^r G_t' \right) (F_t^s)^{-1'} P_{t|t}^{-1} \hat{x}_{t|t} + \left(A_t - A_t G_t Q_t^r G_t' A_t \right) G_t S_t R_t^{-1} Y_t. \end{aligned}$$

□

4.6.3 Further Results in Information Form

The following theorems give some further results that follow from the measurement and time updates in information form.

Theorem 4.13 (Information Filter) *Consider the state space model (4.1) and (4.2). Then, if $R_t > 0$ and $(F_t^s)^{-1}$, $(Q_t^s)^{-1}$ and P_t^{-1} exist, the following recursions hold*

$$\begin{aligned} P_{t+1}^{-1} \hat{x}_{t+1|t} &= \left(I - A_t G_t Q_t^r G_t' \right) (F_t^s)^{-1'} P_t^{-1} \hat{x}_{t|t-1} \\ &\quad + \left(I - A_t G_t Q_t^r G_t' \right) \left[(F_t^s)^{-1'} H_t' + A_t G_t S_t \right] R_t^{-1} Y_t \\ P_{t+1}^{-1} &= A_t - A_t G_t Q_t^r G_t' A_t, \end{aligned}$$

where

$$A_t = (F_t^s)^{-1'} (P_t^{-1} + H_t' R_t^{-1} H_t)(F_t^s)^{-1}, \quad Q_t^r = \left[(Q_t^s)^{-1} + G_t' A_t G_t \right]^{-1},$$

and Q_t^s and F_t^s are given by (4.19) and (4.22).

Proof The first recursion follows from (4.27) and (4.30) and the second from (4.28) and (4.31). □

Theorem 4.14 (Information Form for the Filtered Estimators) *Consider the state space model (4.1) and (4.2) and suppose that we have computed the estimator* *of x_t based on $Y_{1:t}$, $\hat{x}_{t|t}$, and its MSE, $P_{t|t}$. Then, if $R_t > 0$ and $(F_t^s)^{-1}, (Q_t^s)^{-1}$ and $P_{t|t}^{-1}$ exist, the following recursions hold*

$$P_{t+1|t+1}^{-1}\hat{x}_{t+1|t+1} = \left(I - A_t G_t Q_t^r G_t'\right)\left[(F_t^s)^{-1'} P_{t|t}^{-1}\hat{x}_{t|t} + A_t G_t S_t R_t^{-1} Y_t\right] + H_{t+1}' R_{t+1}^{-1} Y_{t+1}$$
$$P_{t+1|t+1}^{-1} = A_t - A_t G_t Q_t^r G_t' A_t + H_{t+1}' R_{t+1}^{-1} H_{t+1},$$

where

$$A_t = (F_t^s)^{-1'} P_{t|t}^{-1} (F_t^s)^{-1}, \quad Q_t^r = \left[(Q_t^s)^{-1} + G_t' A_t G_t\right]^{-1},$$

and Q_t^s and F_t^s are given by (4.19) and (4.22).

Proof The first recursion follows from (4.27) and (4.30) and the second from (4.28) and (4.31). □

4.7 Square Root Covariance and Information Filter

As mentioned earlier, there are situations in which it is desirable to propagate the inverse quantities P_t^{-1}. To avoid the accumulation of numerical errors, it is sometimes preferable to propagate the square root factors, $P_t^{-1/2}$. The resulting algorithms are said to be in square root information form. One convenient way of obtaining such algorithms is to augment the square root covariance filters to construct nonsingular matrices that can be inverted. For example, take algorithm (4.13) and form the augmented matrices

$$U_t' \begin{bmatrix} (R_t^s)^{1/2'} & 0 & 0 \\ P_t^{1/2'} H_t' & P_t^{1/2'} F_t' & 0 \\ Q_t^{-1/2} S_t & Q_t^{1/2'} G_t' & Q_t^{1/2'} \end{bmatrix} = \begin{bmatrix} \Sigma_t^{1/2'} & \widehat{K}_t' & X_1 \\ 0 & P_{t+1}^{1/2'} & X_2 \\ 0 & 0 & X_3 \end{bmatrix},$$

where the matrices X_1, X_2, and X_3 can be obtained as usual by premultiplying the previous matrices by their transposes and equating entries. Assuming that $R_t^s > 0, Q_t > 0, \Omega > 0$ and the F_t are nonsingular, by Lemma 4.2, we can invert the matrices in the previous equality to get

$$\begin{bmatrix} (R_t^s)^{-1/2'} & 0 & 0 \\ -F_t^{-1'} H_t' (R_t^s)^{-1/2'} & F_t^{-1'} P_t^{-1/2'} & 0 \\ L_t' (R_t^s)^{-1/2'} & -G_t' F_t^{-1'} P_t^{-1/2'} & Q_t^{-1/2'} \end{bmatrix} U_t = \begin{bmatrix} \Sigma_t^{-1/2'} & -K_t' P_{t+1}^{-1/2'} & * \\ 0 & P_{t+1}^{-1/2'} & * \\ 0 & 0 & * \end{bmatrix},$$

where $L_t = H_t F_t^{-1} G_t - S_t' Q_t^{-1}$ and the asterisks denote elements that are not relevant to our purposes. Transposing the matrices in the previous equality, we see that the same orthogonal matrix, U_t, that is used in the square root covariance filter also satisfies

$$U_t'\begin{bmatrix} (R_t^s)^{-1/2} & -(R_t^s)^{-1/2}H_tF_t^{-1} & (R_t^s)^{-1/2}L_t \\ 0 & P_t^{-1/2}F_t^{-1} & -P_t^{-1/2}F_t^{-1}G_t \\ 0 & 0 & Q_t^{-1/2} \end{bmatrix} = \begin{bmatrix} \Sigma_t^{-1/2} & 0 & 0 \\ -P_{t+1}^{-1/2}K_t & P_{t+1}^{-1/2} & 0 \\ * & * & * \end{bmatrix}.$$

We have thus proved most of the following theorem. We leave the rest of the proof to the reader.

Theorem 4.15 (Square Root Covariance and Information Filter When $\mathbf{Q_t > 0}$) *Suppose that the process $\{Y_t\}$ follows the state space model (4.1) and (4.2) and that $Q_t > 0$, and define $R_t^s = R_t - S_t'Q_t^{-1}S_t$. Then, if $R_t^s > 0, \Omega > 0$ and the F_t are nonsingular, the application of the QR algorithm yields an orthogonal matrix U_t such that*

$$U_t'\left[\begin{array}{cc|cc|c} (R_t^s)^{1/2'} & 0 & -(R_t^s)^{-1/2}H_tF_t^{-1} & (R_t^s)^{-1/2}L_t & -(R_t^s)^{-1/2}Y_t \\ P_t^{1/2'}H_t' & P_t^{1/2'}F_t' & P_t^{-1/2}F_t^{-1} & -P_t^{-1/2}F_t^{-1}G_t & P_t^{-1/2}\hat{x}_{t|t-1} \\ Q_t^{-1/2}S_t & Q_t^{1/2'}G_t' & 0 & Q_t^{-1/2} & 0 \end{array}\right]$$
$$= \left[\begin{array}{cc|cc|c} \Sigma_t^{1/2'} & \widehat{K}_t' & 0 & 0 & -\widehat{E}_t \\ 0 & P_{t+1}^{1/2'} & P_{t+1}^{-1/2} & 0 & P_{t+1}^{-1/2}\hat{x}_{t+1|t} \\ 0 & 0 & * & * & * \end{array}\right],$$

where $L_t = H_tF_t^{-1}G_t - S_t'Q_t^{-1}, \widehat{K}_t = (F_tP_tH_t' + G_tS_t)\Sigma_t^{-1/2'} = K_t\Sigma_t^{1/2}, \widehat{E}_t = \Sigma_t^{-1/2}E_t$ and the asterisks denote elements that are not relevant to our purposes. Here it is understood that U_t is an orthogonal matrix that either upper-triangularizes the first two block columns or lower-triangularizes the middle two block columns of the left-hand side matrix. The predicted estimates, $\hat{x}_{t+1|t}$, can be obtained as either $\hat{x}_{t+1|t} = F_t\hat{x}_{t|t-1} + \widehat{K}_t\widehat{E}_t$ or $\hat{x}_{t+1|t} = P_{t+1}^{1/2}\left[P_{t+1}^{-1/2}\hat{x}_{t+1|t}\right]$.

The following theorem is analogous to Theorem 4.15. We omit its proof.

Theorem 4.16 (Square Root Covariance and Information Filter When $\mathbf{R_t > 0}$) *Suppose that the process $\{Y_t\}$ follows the state space model (4.1) and (4.2) and that $R_t > 0$. Then, if $Q_t^s > 0, \Omega > 0$ and the F_t^s are nonsingular, where Q_t^s and F_t^s are given by (4.19) and (4.22), the application of the QR algorithm yields an orthogonal matrix U_t such that*

$$U_t'\left[\begin{array}{cc|cc|c} R_t^{1/2'} & 0 & -R_t^{-1/2}H_t(F_t^s)^{-1} & R_t^{-1/2}H_t(F_t^s)^{-1}G_t & -R_t^{-1/2}Y_t \\ P_t^{1/2'}H_t' & P_t^{1/2'}F_t^{s'} & P_t^{-1/2}(F_t^s)^{-1} & -P_t^{-1/2}(F_t^s)^{-1}G_t & P_t^{-1/2}\hat{x}_{t|t-1} \\ 0 & (Q_t^s)^{1/2'}G_t' & 0 & (Q_t^s)^{-1/2} & (Q_t^s)^{-1/2}S_tR_t^{-1}Y_t \end{array}\right]$$
$$= \left[\begin{array}{cc|cc|c} \Sigma_t^{1/2'} & \widehat{K}_t' & 0 & 0 & -\widehat{E}_t \\ 0 & P_{t+1}^{1/2'} & P_{t+1}^{-1/2} & 0 & P_{t+1}^{-1/2}\hat{x}_{t+1|t} \\ 0 & 0 & * & * & * \end{array}\right],$$

where $\widehat{K}_t = (F_tP_tH_t' + G_tS_t)\Sigma_t^{-1/2'} = K_t\Sigma_t^{1/2}, \widehat{E}_t = \Sigma_t^{-1/2}E_t$ *and the asterisks denote elements that are not relevant to our purposes. Here it is understood that* U_t *is an orthogonal matrix that either upper-triangularizes the first two block columns or lower-triangularizes the middle two block columns of the left-hand side matrix. The predicted estimates,* $\hat{x}_{t+1|t}$*, can be obtained as either* $\hat{x}_{t+1|t} = F_t\hat{x}_{t|t-1} + \widehat{K}_t\widehat{E}_t + G_tS_tR_t^{-1}Y_t$ *or* $\hat{x}_{t+1|t} = P_{t+1}^{1/2}\left[P_{t+1}^{-1/2}\hat{x}_{t+1|t}\right]$.

In the following theorem, more details are given of the square root information filter when $R_t > 0$. This result will be useful later in the development of some smoothing results.

Theorem 4.17 (Square Root Information Filter When $\mathbf{R_t} > 0$) *Suppose that the process* $\{Y_t\}$ *follows the state space model (4.1) and (4.2) and that* $R_t > 0$*. Then, if* $Q_t^s > 0, \Omega > 0$ *and the* F_t^s *are nonsingular, where* Q_t^s *and* F_t^s *are given by (4.19) and (4.22), the application of the QR algorithm yields an orthogonal matrix* U_t *such that*

$$U_t'\left[\begin{array}{cc|c} (Q_t^s)^{-1/2} & 0 & (Q_t^s)^{-1/2}S_tR_t^{-1}Y_t \\ -P_t^{-1/2}(F_t^s)^{-1}G_t & P_t^{-1/2}(F_t^s)^{-1} & P_t^{-1/2}\hat{x}_{t|t-1} \\ R_t^{-1/2}H_t(F_t^s)^{-1}G_t & -R_t^{-1/2}H_t(F_t^s)^{-1} & -R_t^{-1/2}Y_t \end{array}\right]$$
$$= \left[\begin{array}{cc|c} (Q_t^r)^{-1/2} & -\widehat{K}_{b,t} & -\widehat{K}_{b,t}\hat{x}_{t+1|t} + (Q_t^r)^{-1/2}S_tR_t^{-1}Y_t \\ 0 & P_{t+1}^{-1/2} & P_{t+1}^{-1/2}\hat{x}_{t+1|t} \\ 0 & 0 & -\widehat{E}_t \end{array}\right],$$

where

$$Q_t^r = \left[(Q_t^s)^{-1} + G_t'A_tG_t\right]^{-1}, \qquad A_t = (F_t^s)^{-1'}(P_t^{-1} + H_t'R_t^{-1}H_t)(F_t^s)^{-1},$$

$\widehat{E}_t = \Sigma_t^{-1/2}E_t$ *and* $\widehat{K}_{b,t} = (Q_t^r)^{1/2'}G_t'A_t$.

Proof The theorem can be proved multiplying first both terms in the equality of the theorem by their respective transposes and equating then the respective matrix entries. Thus, letting L and R be the (1,2) elements in the left- and right-hand sides, respectively, it is obtained that

$$\begin{aligned} L &= -G_t'(F_t^s)^{-1'}P_t^{-1/2'}P_t^{-1/2}(F_t^s)^{-1} - G_t'(F_t^s)^{-1'}H_t'R_t^{-1/2'}R_t^{-1/2}H_t(F_t^s)^{-1} \\ &= -G_t'(F_t^s)^{-1'}\left[P_t^{-1} + H_t'R_t^{-1}H_t\right](F_t^s)^{-1} \\ &= -G_t'A_t, \end{aligned}$$

and

$$R = -(Q_t^r)^{-1/2'}\widehat{K}_{b,t}.$$

From this, it follows that $L = R$. In a similar way, letting now L and R be the (1,3) elements in the left- and right-hand sides, respectively, we get

$$\begin{aligned} L &= (Q_t^s)^{-1/2'}(Q_t^s)^{-1/2}S_tR_t^{-1}Y_t - G_t'(F_t^s)^{-1'}\left[P_t^{-1/2'}P_t^{-1/2}\hat{x}_{t|t-1} + H_t'R_t^{-1/2'}R_t^{-1/2}Y_t\right] \\ &= (Q_t^s)^{-1}S_tR_t^{-1}Y_t - G_t'(F_t^s)^{-1'}\left[P_t^{-1}\hat{x}_{t|t-1} + H_t'R_t^{-1}Y_t\right] \\ &= \left[(Q_t^r)^{-1} - G_t'A_tG_t\right]S_tR_t^{-1}Y_t - G_t'(F_t^s)^{-1'}P_{t|t}^{-1}\hat{x}_{t|t} \\ &= (Q_t^r)^{-1}S_tR_t^{-1}Y_t - G_t'A_t\left[G_tS_tR_t^{-1}Y_t + F_t^s\hat{x}_{t|t}\right] \\ &= (Q_t^r)^{-1}S_tR_t^{-1}Y_t - G_t'A_t\hat{x}_{t+1|t} \\ &= (Q_t^r)^{-1/2'}\left[(Q_t^r)^{-1/2}S_tR_t^{-1}Y_t - (Q_t^r)^{1/2'}G_t'A_t\hat{x}_{t+1|t}\right], \end{aligned}$$

where we have used the equality $(Q_t^s)^{-1} = (Q_t^r)^{-1} - G_t'A_tG_t$, the measurement update formula (4.27), the equality $A_t = (F_t^s)^{-1'}P_{t|t}^{-1}(F_t^s)^{-1}$, that follows from (4.28), and the time update formula (4.24). On the other hand,

$$R = (Q_t^r)^{-1/2'}\left[-\widehat{K}_{b,t}\hat{x}_{t+1|t} + (Q_t^r)^{-1/2}S_tR_t^{-1}Y_t\right],$$

and, clearly, $L = R$. The rest of the theorem follows from Theorem 4.16. □

An interesting distinction between the algorithms of Theorems 4.15 and 4.16 when $Q_t > 0$ and $R_t > 0$ is that only the second one can be separated into measurement and time update steps.

4.7.1 Square Root Covariance and Information Form for Measurement Update

Theorem 4.18 (Square Root Covariance and Information Filter for Measurement Update When $\mathbf{R_t} > 0$) *Consider the state space model (4.1) and (4.2) and suppose that we have computed the estimator of x_t based on $Y_{1:t-1}, \hat{x}_{t|t-1}$, and its MSE, P_t, and a new measurement, Y_t, becomes available. Then, if $R_t > 0$ and P_t^{-1} exists, the QR algorithm produces an orthogonal matrix U_t such that*

$$U_t'\left[\begin{array}{cc|cc|c} R_t^{1/2'} & 0 & R_t^{-1/2} & -R_t^{-1/2}H_t & -R_t^{-1/2}Y_t \\ P_t^{1/2'}H_t' & P_t^{1/2'} & 0 & P_t^{-1/2} & P_t^{-1/2}\hat{x}_{t|t-1} \end{array}\right] = \left[\begin{array}{cc|cc|c} \Sigma_t^{1/2'} & \widehat{K}_{f,t}' & \Sigma_t^{-1/2} & 0 & -\widehat{E}_t \\ 0 & P_{t|t}^{1/2'} & -P_{t|t}^{-1/2}K_{f,t} & P_{t|t}^{-1/2} & P_{t|t}^{-1/2}\hat{x}_{t|t} \end{array}\right],$$

where $\widehat{K}_{f,t} = P_tH_t'\Sigma_t^{-1/2'} = K_{f,t}\Sigma_t^{1/2}$ and $\widehat{E}_t = \Sigma_t^{-1/2}E_t$. Here it is understood that U_t is an orthogonal matrix that either upper-triangularizes the first block column or lower-triangularizes the fourth block column of the left-hand side matrix. The *filtered estimates,* $\hat{x}_{t|t}$*, can be obtained as either* $\hat{x}_{t|t} = \hat{x}_{t|t-1} + \widehat{K}_{f,t}\widehat{E}_t$ *or* $\hat{x}_{t|t} = P_{t|t}^{1/2}\left[P_{t|t}^{-1/2}\hat{x}_{t|t}\right]$.

Proof Consider the algorithm for measurement update of Theorem 4.6

$$U_t' \begin{bmatrix} R_t^{1/2'} & 0 \\ P_t^{1/2'}H_t' & P_t^{1/2'} \end{bmatrix} = \begin{bmatrix} \Sigma_t^{1/2'} & \widehat{K}_{f,t}' \\ 0 & P_{t|t}^{1/2'} \end{bmatrix},$$

Inverting the previous matrices, it is obtained that

$$\begin{bmatrix} R_t^{-1/2'} & 0 \\ -H_t'R_t^{-1/2'} & P_t^{-1/2'} \end{bmatrix} U_t = \begin{bmatrix} \Sigma_t^{-1/2'} & -K_{f,t}'P_{t|t}^{-1/2'} \\ 0 & P_{t|t}^{-1/2'} \end{bmatrix}.$$

Transposing the previous matrices and using again Theorem 4.4.4, we can write

$$U_t' \begin{bmatrix} R_t^{-1/2} & -R_t^{-1/2}H_t \\ 0 & P_t^{-1/2} \end{bmatrix} = \begin{bmatrix} \Sigma_t^{-1/2} & 0 \\ -P_{t|t}^{-1/2}K_{f,t} & P_{t|t}^{-1/2} \end{bmatrix}.$$

□

Remark 4.1 The vectors $\widehat{E}_t$ have zero mean and unit covariance matrix. Thus, they constitute a sequence of "standardized residuals" and can be used for inference. ◇

We state the square root information form for measurement update alone as a corollary.

Corollary 4.1 (Square Root Information Filter for Measurement Update When $\mathbf{R_t > 0}$) *Consider the state space model (4.1) and (4.2) and suppose that we have computed the estimator of* x_t *based on* $Y_{1:t-1}, \hat{x}_{t|t-1}$*, and its MSE,* P_t*, and a new measurement,* Y_t*, becomes available. Then, if* $R_t > 0$ *and* P_t^{-1} *exists, the QR algorithm produces an orthogonal matrix* U_t *such that*

$$U_t' \left[\begin{array}{cc|c} P_t^{-1/2} & P_t^{-1/2}\hat{x}_{t|t-1} & 0 \\ -R_t^{-1/2}H_t & -R_t^{-1/2}Y_t & R_t^{-1/2} \end{array}\right]$$
$$= \left[\begin{array}{cc|c} P_{t|t}^{-1/2} & P_{t|t}^{-1/2}\hat{x}_{t|t} & -P_{t|t}^{-1/2}K_{f,t} \\ 0 & -\widehat{E}_t & \Sigma_t^{-1/2} \end{array}\right],$$

where $\widehat{E}_t = \Sigma_t^{-1/2}E_t$*. The filtered estimates,* $\hat{x}_{t|t}$*, can be obtained as* $\hat{x}_{t|t} = P_{t|t}^{1/2}\left[P_{t|t}^{-1/2}\hat{x}_{t|t}\right]$.

4.7.2 Square Root Covariance and Information Form for Time Update

Theorem 4.19 (Square Root Covariance and Information Filter for Time Update When $\mathbf{R_t} > 0$) *Consider the state space model (4.1) and (4.2) and suppose that we have computed the estimator of x_t based on $Y_{1:t}$, $\hat{x}_{t|t}$, and its MSE, $P_{t|t}$, and without any further measurements wish to find $\hat{x}_{t+1|t}$ and P_{t+1}. Then, if $R_t > 0$ and $(F_t^s)^{-1}$, $(Q_t^s)^{-1}$ and P_t^{-1} exist, the QR algorithm produces an orthogonal matrix U_t such that*

$$U_t' \left[\begin{array}{cc|cc|c} P_{t|t}^{1/2'} F_t^{s'} & 0 & P_{t|t}^{-1/2}(F_t^s)^{-1} & -P_{t|t}^{-1/2}(F_t^s)^{-1}G_t & P_{t|t}^{-1/2}\hat{x}_{t|t} \\ (Q_t^s)^{1/2'}G_t' & (Q_t^s)^{1/2'} & 0 & (Q_t^s)^{-1/2} & (Q_t^s)^{-1/2}S_tR_t^{-1}Y_t \end{array} \right]$$
$$= \left[\begin{array}{cc|cc|c} P_{t+1}^{1/2'} & P_{t+1}^{1/2'}A_tG_tQ_t^r & P_{t+1}^{-1/2} & 0 & P_{t+1}^{-1/2}\hat{x}_{t+1|t} \\ 0 & (Q_t^r)^{1/2'} & -\widehat{K}_{b,t} & (Q_t^r)^{-1/2} & -\widehat{K}_{b,t}\hat{x}_{t+1|t} + (Q_t^r)^{-1/2}S_tR_t^{-1}Y_t \end{array} \right],$$

where

$$A_t = (F_t^s)^{-1'}P_{t|t}^{-1}(F_t^s)^{-1}, \quad Q_t^r = \left[(Q_t^s)^{-1} + G_t'A_tG_t\right]^{-1},$$

$\widehat{K}_{b,t} = (Q_t^r)^{1/2'}G_t'A_t$, and Q_t^s and F_t^s are given by (4.19) and (4.22). Here it is understood that U_t is an orthogonal matrix that either upper-triangularizes the first block column or lower-triangularizes the fourth block column of the left-hand side matrix. The predicted estimates, $\hat{x}_{t+1|t}$, can be obtained as either $\hat{x}_{t+1|t} = F_t^s\hat{x}_{t|t-1} + G_tS_tR_t^{-1}Y_t$ or $\hat{x}_{t+1|t} = P_{t+1}^{1/2}\left[P_{t+1}^{-1/2}\hat{x}_{t+1|t}\right]$

Proof We proceed as in the proof of Theorem 4.15, augmenting the square root covariance filter to construct nonsingular matrices that can be inverted. Since Q_t^s is nonsingular, this can be achieved in the algorithm of Theorem 4.10 by forming the augmented matrices

$$U_t' \begin{bmatrix} P_{t|t}^{1/2'}F_t^{s'} & 0 \\ (Q_t^s)^{1/2'}G_t' & (Q_t^s)^{1/2'} \end{bmatrix} = \begin{bmatrix} P_{t+1}^{1/2'} & X_1 \\ 0 & X_2 \end{bmatrix},$$

where X_1 and X_2 are to be determined. By inverting both sides of the equality, it is obtained that

$$\begin{bmatrix} (F_t^s)^{-1'}P_{t|t}^{-1/2'} & 0 \\ -G_t'(F_t^s)^{-1'}P_{t|t}^{-1/2'} & (Q_t^s)^{-1/2'} \end{bmatrix} U_t = \begin{bmatrix} P_{t+1}^{-1/2'} & Y_1 \\ 0 & Y_2 \end{bmatrix},$$

Postmultiplying the matrices in the previous equality by their respective transposes yields

$$\begin{bmatrix} (F_t^s)^{-1'} P_{t|t}^{-1/2'} & 0 \\ -G_t'(F_t^s)^{-1'} P_{t|t}^{-1/2'} & (Q_t^s)^{-1/2'} \end{bmatrix} U_t U_t' \begin{bmatrix} P_{t|t}^{-1/2}(F_t^s)^{-1} & -P_{t|t}^{-1/2}(F_t^s)^{-1}G_t \\ 0 & (Q_t^s)^{-1/2} \end{bmatrix}$$
$$= \begin{bmatrix} A_t & -A_tG_t \\ -G_t'A_t & G_t'A_tG_t + (Q_t^s)^{-1} \end{bmatrix} = \begin{bmatrix} P_{t+1}^{-1} + Y_1Y_2' & Y_1Y_2' \\ Y_2Y_1' & Y_2Y_2' \end{bmatrix}.$$

Thus, $Y_1 = -A_tG_t(Q_t^r)^{1/2}$, $Y_2 = (Q_t^r)^{-1/2'}$ and X_1 and X_1 can be obtained from

$$\begin{bmatrix} P_{t+1}^{1/2'} & X_1 \\ 0 & X_2 \end{bmatrix} = \begin{bmatrix} P_{t+1}^{-1/2'} & -(Q_t^r)^{-1/2}A_tG_t \\ 0 & (Q_t^r)^{-1/2'} \end{bmatrix}^{-1}$$
$$= \begin{bmatrix} P_{t+1}^{-1/2'} & P_{t+1}^{1/2'}A_tG_tQ_t^r \\ 0 & (Q_t^r)^{1/2'} \end{bmatrix}.$$

The rest of the theorem can be proved as in the proof of Theorem 4.17. More specifically, by Theorem 4.10 and what has just been proved,

$$U_t' \left[\begin{array}{cc|c} P_{t|t}^{-1/2}(F_t^s)^{-1} & -P_{t|t}^{-1/2}(F_t^s)^{-1}G_t & P_{t|t}^{-1/2}\hat{x}_{t|t} \\ 0 & (Q_t^s)^{-1/2} & (Q_t^s)^{-1/2}S_tR_t^{-1}Y_t \end{array}\right]$$
$$= \left[\begin{array}{cc|c} P_{t+1}^{-1/2} & 0 & P_{t+1}^{-1/2}\hat{x}_{t+1|t} \\ -\widehat{K}_{b,t} & (Q_t^r)^{-1/2} & X \end{array}\right],$$

where X has to be determined. Multiplying both terms of the previous equality by their transposes and equating the (2,3) element of each of the resulting matrices yields the result after using an argument similar to that in the proof of Theorem 4.17.

□

We state the square root information form alone for time update as a corollary.

Corollary 4.2 (Square Root Information Filter for Time Update When $R_t > 0$) *Consider the state space model (4.1) and (4.2) and suppose that we have computed the estimator of x_t based on $Y_{1:t}$, $\hat{x}_{t|t}$, and its MSE, $P_{t|t}$, and without any further measurements wish to find $\hat{x}_{t+1|t}$ and P_{t+1}. Then, if $R_t > 0$ and $(F_t^s)^{-1}$, $(Q_t^s)^{-1}$ and P_t^{-1} exist, the QR algorithm produces an orthogonal matrix U_t such that*

$$U_t' \left[\begin{array}{cc|c} (Q_t^s)^{-1/2} & 0 & (Q_t^s)^{-1/2}S_tR_t^{-1}Y_t \\ -P_{t|t}^{-1/2}(F_t^s)^{-1}G_t & P_{t|t}^{-1/2}(F_t^s)^{-1} & P_{t|t}^{-1/2}\hat{x}_{t|t} \end{array}\right]$$
$$= \left[\begin{array}{cc|c} (Q_t^r)^{-1/2} & -\widehat{K}_{b,t} & -\widehat{K}_{b,t}\hat{x}_{t+1|t} + (Q_t^r)^{-1/2}S_tR_t^{-1}Y_t \\ 0 & P_{t+1}^{-1/2} & P_{t+1}^{-1/2}\hat{x}_{t+1|t} \end{array}\right],$$

where

$$A_t = (F_t^s)^{-1'} P_{t|t}^{-1} (F_t^s)^{-1}, \quad Q_t^r = \left[(Q_t^s)^{-1} + G_t' A_t G_t \right]^{-1},$$

$\widehat{K}_{b,t} = (Q_t^r)^{1/2'} G_t' A_t$*, and* Q_t^s *and* F_t^s *are given by (4.19) and (4.22). The predicted estimates,* $\hat{x}_{t+1|t}$*, can be obtained as* $\hat{x}_{t+1|t} = P_{t+1}^{1/2} \left[P_{t+1}^{-1/2} \hat{x}_{t+1|t} \right]$.

4.8 Likelihood Evaluation

Assuming $\sigma^2 \neq 1$ and using the output of the Kalman filter, the log-likelihood of $Y = (Y_1', \ldots, Y_n')'$, where $\{Y_t\}$ follows the state space model (4.1) and (4.2), is

$$l(Y) = \text{constant} - \frac{1}{2} \left\{ \frac{1}{\sigma^2} \sum_{t=1}^{n} E_t' \Sigma_t^{-1} E_t + \sum_{t=1}^{n} \ln |\sigma^2 \Sigma_t| \right\}.$$

Differentiating with respect to σ^2 in the previous expression and equating to zero yields the maximum likelihood estimator of $\sigma^2, \hat{\sigma}^2 = \sum_{t=1}^{n} E_t' \Sigma_t^{-1} E_t / (np)$. Thus, σ^2 can be concentrated out of the likelihood and the σ^2-maximized log-likelihood is

$$\begin{aligned} c(Y) &= \text{constant} - \frac{1}{2} \left\{ (np) \ln \left(\sum_{t=1}^{n} E_t' \Sigma_t^{-1} E_t \right) + \sum_{t=1}^{n} \ln |\Sigma_t| \right\} \\ &= \text{constant} - \frac{np}{2} \ln(S), \end{aligned} \tag{4.32}$$

where $S = \left(\sum_{t=1}^{n} E_t' \Sigma_t^{-1} E_t \right) \prod_{t=1}^{n} |\Sigma_t|^{1/(np)}$. Using the Cholesky decomposition $\Sigma_t = \Sigma_t^{1/2} \Sigma_t^{1/2'}$, where $\Sigma_t^{1/2}$ is a lower triangular matrix, and defining $e_t = \Sigma_t^{-1/2} E_t$, S can be written as the nonlinear sum of squares

$$S = \left(\prod_{t=1}^{n} |\Sigma_t|^{1/(2np)} \right) \left(\sum_{t=1}^{n} e_t' e_t \right) \left(\prod_{t=1}^{n} |\Sigma_t|^{1/(2np)} \right).$$

Thus, maximizing (4.32) is equivalent to minimizing S, for which specialized software exists that takes advantage of the special form of S.

4.9 Forecasting

Suppose that $\{Y_t\}$ follows the state space model (4.1) and (4.2) and that the Kalman filter is applied. Then, one-period-ahead forecasts are given by the Kalman filter recursions.

Denoting by $\hat{x}_{n+h|n}$, where $h > 1$, the orthogonal projection of x_{n+h} onto the sample $Y_{1:n}$ and assuming for simplicity $\sigma^2 = 1$, it can be easily shown that h-period-ahead forecasts and their mean squared error P_{n+h} are obtained recursively by

$$\begin{aligned}\hat{x}_{n+h|n} &= F_{n+h-1}\hat{x}_{n+h-1|n}\\ P_{n+h} &= F_{n+h-1}P_{n+h-1}F'_{n+h-1} + G_{n+h-1}Q_{n+h-1}G'_{n+h-1}.\end{aligned}$$

The forecasts for Y_{n+h}, where $h \geq 1$, and the corresponding mean squared error matrices are given by

$$\begin{aligned}\widehat{Y}_{n+h|n} &= H_{n+h}\hat{x}_{n+h|n}\\ MSE(\widehat{Y}_{n+h|n}) &= H_{n+h}P_{n+h}H'_{n+h} + R_{n+h}.\end{aligned}$$

4.10 Smoothing

Suppose that $\{Y_t\}$ follows the state space model (4.1) and (4.2) and assume for simplicity that $\sigma^2 = 1$. In this section, we will consider the smoothing problem, that is related to the Kalman filter corresponding to the sample $Y_{1:n}$. The smoothing problem consists of computing the estimators of random vectors, such as the state or the disturbance vectors, and their MSE based on the whole sample $Y_{1:n}$.

4.10.1 Smoothing Based on the Bryson–Frazier Formulae

Theorem 4.20 (The Bryson–Frazier Formulae (Bryson & Frazier, 1963) for the Fixed-Interval Smoother) *Consider the state space model (4.1) and (4.2). For* $t = n, n-1, \ldots, 1$, *define the so-called* ***adjoint variable***, λ_t, *and its covariance matrix,* Λ_t, *by the recursions*

$$\lambda_t = F'_{p,t}\lambda_{t+1} + H'_t\Sigma_t^{-1}E_t, \quad \Lambda_t = F'_{p,t}\Lambda_{t+1}F_{p,t} + H'_t\Sigma_t^{-1}H_t, \tag{4.33}$$

initialized with $\lambda_{n+1} = 0$ *and* $\Lambda_{n+1} = 0$. *Then, for* $t = n, n-1, \ldots, 1$, *the projection,* $\hat{x}_{t|n}$, *of* x_t *onto the whole sample* $\{Y_t : 1 \leq t \leq n\}$ *and its MSE,* $P_{t|n}$, *satisfy the recursions*

$$\hat{x}_{t|n} = \hat{x}_{t|t-1} + P_t\lambda_t, \quad P_{t|n} = P_t - P_t\Lambda_tP_t. \tag{4.34}$$

Proof With the notation of the Kalman filter, by Proposition 1.5, we can write

$$\begin{aligned}\hat{x}_{t|n} &= E^*(x_t|Y_{1:t-1}, E_t, \ldots, E_n) = E^*(x_t|Y_{1:t-1}) + \sum_{s=t}^{n} E^*(x_t|E_s)\\ &= \hat{x}_{t|t-1} + \sum_{s=t}^{n} \mathrm{Cov}(x_t|E_s)\Sigma_s^{-1}E_s.\end{aligned}$$

Letting $\tilde{x}_s = x_s - \hat{x}_{s|s-1}, s \geq t$, since v_s is orthogonal to x_t and $E_s = H_s\tilde{x}_s + v_s$, we get

$$\begin{aligned}\mathrm{Cov}(x_t, E_s) &= \mathrm{Cov}(x_t, H_s\tilde{x}_s + v_s)\\ &= \mathrm{Cov}(x_t, H_s\tilde{x}_s)\\ &= \mathrm{Cov}(\tilde{x}_t, \tilde{x}_s)H_s', \quad s \geq t.\end{aligned}$$

From the Kalman filter equation $\hat{x}_{t+1|t} = F_t\hat{x}_{t|t-1} + K_tE_t$, it is obtained after a little algebra that $\tilde{x}_{t+1} = (F_t - K_tH_t)\tilde{x}_t - K_tv_t + G_tu_t = F_{p,t}\tilde{x}_t - K_tv_t + G_tu_t$. Since $\tilde{x}_t$ is uncorrelated with v_t and u_t, it follows that

$$\mathrm{Cov}(x_t, E_s) = \begin{cases} P_tH_t' & \text{for } s = t\\ P_tF_{p,t}' \cdots F_{p,s-1}'H_s' & \text{for } s > t, \end{cases} \tag{4.35}$$

and we get the formula

$$\begin{aligned}\hat{x}_{t|n} &= \hat{x}_{t|t-1} + P_t\sum_{s=t}^{n} F_{p,s}^{t'}H_s'\Sigma_s^{-1}E_s\\ &= \hat{x}_{t|t-1} + P_t\lambda_t,\end{aligned}$$

where $\lambda_t = \sum_{s=t}^{n} F_{p,s}^{t'}H_s'\Sigma_s^{-1}E_s$, $F_{p,i}^{j} = F_{p,j-1}F_{p,j-2}\cdots F_{p,i}$ if $i < j$, and $F_{p,i}^{i} = I$. From the previous formula, it is obtained that

$$x_t - \hat{x}_{t|t-1} = x_t - \hat{x}_{t|n} + P_t\lambda_t,$$

and, since $x_t - \hat{x}_{t|n}$ is orthogonal to λ_t, also that

$$P_t = P_{t|n} + P_t\Lambda_tP_t,$$

where $\Lambda_t = \mathrm{Var}(\lambda_t)$. Finally, the formula (4.35) immediately suggests that λ_t and Λ_t can be computed with the recursion (4.33). □

The previous theorem was proved by Bryson & Frazier (1963) for the continuous case as a two-point boundary value problem using calculus of variations. Derivation of the Bryson–Frazier formulae for the discrete case was given by Bryson & Ho (1969).

Corollary 4.3 *The fixed interval smoother can be alternatively expressed in terms of filtered estimates as*

$$\hat{x}_{t|n} = \hat{x}_{t|t} + P_t F'_{p,t}\lambda_{t+1}, \quad P_{t|n} = P_{t|t} - P_t F'_{p,t}\Lambda_{t+1}F_{p,t}P_t. \tag{4.36}$$

Proof The corollary follows from Theorem 4.20 and formulae (4.6). □

The following corollary is a direct consequence of Theorem 4.20. It can be used to prove other smoothing formulae.

Corollary 4.4 *Let t and m be positive integers such that* $1 \leq t < t+m-1 \leq n$ *and define for* $s = t+m-1, t+m-2, \ldots, t$ *the auxiliary variable,* $\lambda_{s|t+m-1}$*, and its MSE,* $\Lambda_{s|t+m-1}$*, by the recursions*

$$\lambda_{s|t+m-1} = F'_{p,s}\lambda_{s+1|t+m-1} + H'_s\Sigma_s^{-1}E_s, \tag{4.37}$$

$$\Lambda_{s|t+m-1} = F'_{p,s}\Lambda_{s+1|t+m-1}F_{p,s} + H'_s\Sigma_s^{-1}H_s, \tag{4.38}$$

initialized with $\lambda_{t+m|t+m-1} = 0$ *and* $\Lambda_{t+m|t+m-1} = 0$*. Then, the projection,* $\hat{x}_{t|t+m-1}$*, of* x_t *onto the sample* $\{Y_s : 1 \leq s \leq t+m-1\}$ *and its MSE,* $P_{t|t+m-1}$*, satisfy the relations*

$$\hat{x}_{t|t+m-1} = \hat{x}_{t|t-1} + P_t\lambda_{t|t+m-1}, \quad P_{t|t+m-1} = P_t - P_t\Lambda_{t|t+m-1}P_t. \tag{4.39}$$

Using the previous corollary, since $\lambda_{t|t+m} - \lambda_{t|t+m-1} = F'^{t+m-1}_{p,t}H'_{t+m}\Sigma^{-1}_{t+m}E_{t+m}$, where $F^j_{p,i} = F_{p,j-1}F_{p,j-2}\cdots F_{p,i}$ if $i < j$, and $F^i_{p,i} = I$, we get the following corollary.

Corollary 4.5 (The Fixed-Point Smoother) *Under the assumptions of the previous corollary, if we fix t and let m increase in time, the projection,* $\hat{x}_{t|t+m}$*, of* x_t *onto the sample* $\{Y_s : 1 \leq s \leq t+m\}$ *and its MSE,* $P_{t|t+m}$*, satisfy the recursions*

$$\hat{x}_{t|t+m} = \hat{x}_{t|t+m-1} + P_t F'^{t+m-1}_{p,t}H'_{t+m}\Sigma^{-1}_{t+m}E_{t+m},$$

$$P_{t|t+m} = P_{t|t+m-1} - P_t F'^{t+m-1}_{p,t}H'_{t+m}\Sigma^{-1}_{t+m}H_{t+m}F^{t+m-1}_{p,t}P_t.$$

These recursions can be expressed in a more compact form as

$$\hat{x}_{t|j} = \hat{x}_{t|j-1} + K^j_t E_j$$

$$P_{t|j} = P_{t|j-1} - K^j_t H_j P^{j'}_t, \qquad j = t, t+1, \ldots, n,$$

where $K^j_t = P^j_t H'_j \Sigma^{-1}_j$ *and* $P^j_t = P^{j-1}_t F'_{p,j-1}$*, initialized with* $P^t_t = P_t$ *and* $P_{t|t-1} = P_t$*.*

Combining Theorem 4.20 and the previous Corollary, we obtain the following result. The proof is omitted.

Corollary 4.6 (Combining the Fixed-Point with the Fixed-Interval Smoother) *Under the assumptions and with the notation of Theorem 4.20 and the previous corollary, the following recursions hold for fixed $k \le n$ and $t = k, k-1, \ldots, 1$.*

$$\hat{x}_{t|n} = \hat{x}_{t|k} + P_t^{k+1}\lambda_{k+1}$$
$$P_{t|n} = P_{t|k} - P_t^{k+1}\Lambda_{k+1}P_t^{k+1'}.$$

From Corollary 4.4, we get the following corollary.

Corollary 4.7 (The Fixed-Lag Smoother) *Under the assumptions and with the notation of Corollary 4.4, if we fix m and let t increase in time, the projection, $\hat{x}_{t|t+m-1}$, of x_t onto the sample $\{Y_s : 1 \le s \le t+m-1\}$ and its MSE, $P_{t|t+m-1}$, can be recursively obtained using (4.39), where $\lambda_{t|t+m-1}$ and $\Lambda_{t|t+m-1}$ satisfy the recursions*

$$\lambda_{t|t+m-1} = F'_{p,t}\lambda_{t+1|t+m} + H'_t\Sigma_t^{-1}E_t - F'^{t+m-1}_{p,t}H'_{t+m}\Sigma_{t+m}^{-1}E_{t+m}, \tag{4.40}$$
$$\Lambda_{t|t+m-1} = F'_{p,t}\Lambda_{t+1|t+m}F_{p,t} + H'_t\Sigma_t^{-1}H_t - F'^{t+m-1}_{p,t}H'_{t+m}\Sigma_{t+m}^{-1}H_{t+m}F^{t+m-1}_{p,t}. \tag{4.41}$$

The recursions (4.40) and (4.41) are run backwards for $t = n-m+1, n-m, \ldots, 1$, initialized with $\lambda_{n-m+1|n}$ computed using (4.37) and (4.38).

Proof The recursions (4.40) and (4.41) are deduced from (4.37) and (4.38) and the relation $\lambda_{t|t+m} - \lambda_{t|t+m-1} = F'^{t+m-1}_{p,t}H'_{t+m}\Sigma_{t+m}^{-1}E_{t+m}$. □

The following theorem introduces the so-called inverse process. This process has numerous applications in smoothing. First, we will need a lemma.

Lemma 4.3 *Let $Y = (Y'_1 \ldots, Y'_n)'$ be a sequence of zero mean random vectors such that $Var(Y) = \Sigma_Y$ is positive definite and suppose that the subvector $Y^m = JY$ is missing, where J is a selector matrix formed with zeros and ones. Suppose that we replace the missing values with tentative ones and let $\overline{Y}^m$ and $\overline{Y}$ be the vector of tentative values and the filled-in sequence, respectively. Then, if Y^o is the subvector of observed values, the following relations hold*

$$E^*(Y^m|Y^o) = \overline{Y}^m - (J\Sigma_Y^{-1}J')^{-1}J\Sigma_Y^{-1}\overline{Y}$$
$$MSE\left[E^*(Y^m|Y^o)\right] = (J\Sigma_Y^{-1}J')^{-1}.$$

Proof Let $P = [J', K']'$ be an orthogonal selection matrix formed with zeros and ones such that $[Y^{m'}, Y^{o'}]' = PY$. Then,

$$Y'\Sigma_Y^{-1}Y = [Y^{m'}, Y^{o'}]P\Sigma_Y^{-1}P'[Y^{m'}, Y^{o'}]' = [Y^{m'}, Y^{o'}]\left(P\Sigma_YP'\right)^{-1}[Y^{m'}, Y^{o'}]',$$

where $P\Sigma_Y P'$ is the covariance matrix of $[Y^{m'}, Y^{o'}]'$. Thus, using (1.4) and (1.5), we can write

$$Y'\Sigma_Y^{-1}Y = \left[Y^m - E^*(Y^m|Y^o)\right]' \Sigma_{Y^m|Y^o}^{-1} \left[Y^m - E^*(Y^m|Y^o)\right] + Y^{o'}\Sigma_{Y^o}^{-1}Y^o, \quad (4.42)$$

where $\Sigma_{Y^m|Y^o} = MSE\,[E^*(Y^m|Y^o)]$ and $\mathrm{Var}(Y^o) = \Sigma_{Y^o}$. Let $\omega = \overline{Y}^m - Y^m$. Then, $\overline{Y} = J'\omega + Y$ and, substituting this expression in (4.42), it is obtained that

$$(\overline{Y} - J'\omega)'\Sigma_Y^{-1}(\overline{Y} - J'\omega) = \left[\overline{Y}^m - E^*(Y^m|Y^o) - \omega\right]'$$
$$\times \Sigma_{Y^m|Y^o}^{-1} \left[\overline{Y}^m - E^*(Y^m|Y^o) - \omega\right] + Y^{o'}\Sigma_{Y^o}^{-1}Y^o.$$

Minimizing both sides of the previous equality with respect to ω and equating the estimators, the lemma is proved. ☐

Theorem 4.21 (The Inverse Process and Interpolation) *For $t = n, n-1, \ldots, 1$, define the **inverse process** $\{Y_t^i\}$ and its MSE, M_t, by the recursions*

$$Y_t^i = \Sigma_t^{-1}E_t - K_t'\lambda_{t+1}, \quad M_t = \Sigma_t^{-1} + K_t'\Lambda_{t+1}K_t. \quad (4.43)$$

*Then, letting $Y^i = (Y_1^{i'}, \ldots, Y_n^{i'})', Y = (Y_1', \ldots, Y_n')', E = (E_1', \ldots, E_n')', \Sigma =$ $\mathrm{diag}(\Sigma_1, \ldots, \Sigma_n)$, $Var(Y) = \Sigma_Y$, and Ψ be the matrix defined by (4.4), the equalities $Y^i = \Psi'^{-1}\Sigma^{-1}E$ and $Var(Y^i) = \Sigma_Y^{-1}$ hold. In addition, the inverse process is related to **interpolation** as follows. Suppose that there is a missing observation Y_t in the sample, that we replace it with a tentative value Y_t^m, and that we apply the Kalman filter to the filled-in series. Then, the projection $Y_{t|s\neq t}$ of Y_t onto $\{Y_s : 1 \leq s \leq n, s \neq t\}$ and its MSE are given by the formulae*

$$Y_{t|s\neq t} = Y_t^m - M_t^{-1}Y_t^i, \quad MSE(Y_{t|s\neq t}) = M_t^{-1},$$

where $\{Y_t^i\}$ is the inverse process corresponding also to the filled-in series.

Proof Transposing (4.5) and considering the backwards recursions (4.33) and (4.43), it follows that $Y^i = \Psi'^{-1}\Sigma^{-1}E$ and $\mathrm{Var}(Y^i) = \Sigma_Y^{-1}$. Also, it is evident that $M_t = \mathrm{Var}(Y_t^i) = \Sigma_t^{-1} + K_t'\Lambda_{t+1}K_t$ because E_t and λ_{t+1} are uncorrelated. The rest of the theorem can be proved easily using Lemma 4.3. ☐

Note that the covariance matrix of Y^i is the inverse of the covariance matrix of the vector of observations, Y, thus justifying the name of inverse process for $\{Y_t^i\}$. However, the quantities Y_t^i are sometimes called "smoothations" in the statistical literature instead of inverse process. See, for example, Jong & Penzer (1998). Using the inverse process, it is possible to define the so-called **inverse state space model** in which the time runs backwards

$$\lambda_t = F'_{p,t}\lambda_{t+1} + H'_t\Sigma_t^{-1}E_t \tag{4.44}$$

$$Y_t^i = -K'_t\lambda_{t+1} + \Sigma_t^{-1}E_t, \qquad t = n, n-1, \ldots, 1, \tag{4.45}$$

initialized with $\lambda_{n+1} = 0$ and $\mathrm{Var}(\lambda_{n+1}) = \Lambda_{n+1} = 0$.

Using (4.44) and (4.45), it is possible to easily obtain the covariances of the inverse process and, therefore, the matrix $\mathrm{Var}(Y^i) = \Sigma_Y^{-1}$. We do this in the following lemma. The proof is left as an exercise. See Problem 4.2.

Lemma 4.4 (Covariances of the Inverse Process) *Consider the inverse state space model (4.44) and (4.45). Then, the covariances of the adjoint variables, λ_t, can be written as*

$$\gamma_\lambda(r,s) = E(\lambda_r\lambda'_s) = \begin{cases} \Lambda_r F^r_{p,s} & r \geq s \\ F^{s'}_{p,r}\Lambda_s & r \leq s, \end{cases}$$

and the covariances of the inverse process $\{Y_t^i\}$ as

$$\gamma_{Y^i}(r,s) = E(Y_r^i Y_s^{i'}) = \begin{cases} -N_r^{i'}F^r_{p,s+1}K_s & r > s \\ \Sigma_r^{-1} + K_r\Lambda_{r+1}K'_r & r = s \\ -K'_r F^{s'}_{p,r+1}N_s^i & r < s, \end{cases} \tag{4.46}$$

where $F^j_{p,k} = F_{p,j-1}F_{p,j-2}\cdots F_{p,k}$ if $k < j, F^k_{p,k} = I$, and $N_r^i = -F'_{p,r}\Lambda_{r+1}K_r + H'_r\Sigma_r^{-1} = Cov(\lambda_r, Y_r^i)$.

Remark 4.2 Using the previous lemma and Lemma 4.3, it is possible to interpolate a vector of missing values, Y^m, and compute the MSE of this interpolator. More specifically, we can first compute the necessary elements of $\mathrm{Var}(Y^i) = \Sigma_Y^{-1}$ using (4.46). Then, using the notation and the formulas of Lemma 4.3, we can compute $E^*(Y^m|Y^o)$ and its MSE. This procedure was used in Gómez, Maravall, & Peña (1999). ◇

The following theorem is a direct consequence of the Bryson–Frazier formulae.

Theorem 4.22 (The Disturbance Smoothers) *For $t = n, n-1, \ldots, 1$, the projection $v_{t|n}$ of v_t onto the whole sample $\{Y_t : 1 \leq t \leq n\}$ and its MSE satisfy the recursions*

$$\hat{v}_{t|n} = R_t\Sigma_t^{-1}E_t + (S'_tG'_t - R_tK'_t)\lambda_{t+1} = R_tY_t^i + S'_tG'_t\lambda_{t+1}.$$

$$MSE(\hat{v}_{t|n}) = R_t - \left[R_t\Sigma_t^{-1}R'_t + (S'_tG'_t - R_tK'_t)\Lambda_{t+1}(S'_tG'_t - R_tK'_t)'\right].$$

Similarly, the projection $u_{t|n}$ *of* u_t *onto the whole sample* $\{Y_t : 1 \leq t \leq n\}$ *and its MSE satisfy for* $t = n, n-1, \ldots, 1$*, the recursions*

$$\hat{u}_{t|n} = S_t \Sigma_t^{-1} E_t + (Q_t G_t' - S_t K_t')\lambda_{t+1} = S_t Y_t^i + Q_t G_t' \lambda_{t+1}.$$
$$MSE(\hat{u}_{t|n}) = Q_t - \left[S_t \Sigma_t^{-1} S_t' + (Q_t G_t' - S_t K_t')\Lambda_{t+1}(Q_t G_t' - S_t K_t')'\right].$$

Proof Projecting onto the whole sample in $v_t = Y_t - Hx_t$ yields

$$\begin{aligned}
\hat{v}_{t|n} &= Y_t - H_t \hat{x}_{t|n} = Y_t - H_t(\hat{x}_{t|t-1} + P_t \lambda_t) \\
&= E_t - H_t P_t \lambda_t = E_t - H_t P_t (H_t \Sigma_t^{-1} E_t + F_{p,t}' \lambda_{t+1}) \\
&= E_t - (\Sigma_t - R_t)\Sigma_t^{-1} E_t - H_t P_t F_{p,t}' \lambda_{t+1} \\
&= R_t \Sigma_t^{-1} E_t - H_t P_t (F_t' - H_t' K_t')\lambda_{t+1} \\
&= R_t \Sigma_t^{-1} E_t - \left[H_t P_t F_t' - (\Sigma_t - R_t) K_t'\right] \lambda_{t+1} \\
&= R_t \Sigma_t^{-1} E_t - \left[H_t P_t F_t' - H_t P_t F_t' - S_t' G_t' + R_t K_t'\right] \lambda_{t+1} \\
&= R_t \Sigma_t^{-1} E_t + (S_t' G_t' - R_t K_t')\lambda_{t+1}.
\end{aligned}$$

To prove the formula for the MSE, consider that $(v_t - \hat{v}_{t|n}) + \hat{v}_{t|n} = v_t$ and that the two terms to the left of the equality are uncorrelated. Then, taking expectations and considering that E_t and λ_{t+1} are uncorrelated, the formula for the MSE follows.

To prove the formula for $\hat{u}_{t|n}$, note that, by Proposition 1.5, we can write

$$\begin{aligned}
\hat{u}_{t|n} &= E^*(u_t | E_1, E_2, \ldots, E_n) = \sum_{s=1}^{n} E^*(u_t | E_s) \\
&= \sum_{s=1}^{n} \text{Cov}(u_t | E_s) \Sigma_s^{-1} E_s.
\end{aligned}$$

Let $\tilde{x}_s = x_s - \hat{x}_{s|s-1}$ and note, as in the proof of Theorem 4.20, that $E_s = H_s \tilde{x}_s + v_s$ and $\tilde{x}_{t+1} = (F_t - K_t H_t)\tilde{x}_t - K_t v_t + G_t u_t = F_{p,t}\tilde{x}_t - K_t v_t + G_t u_t$. Then, u_t is uncorrelated with E_s for $s < t$,

$$\hat{u}_{t|n} = \sum_{s=t}^{n} \text{Cov}(u_t | E_s) \Sigma_s^{-1} E_s,$$

and

$$\begin{aligned}
\text{Cov}(u_t, E_s) &= \text{Cov}(u_t, H_s \tilde{x}_s + v_s) \\
&= \begin{cases} S_t & \text{for } s = t \\ (Q_t G_t' - S_t K_t')\left(F_{p,t+1}^{s}\right)' H_s' & \text{for } s > t, \end{cases}
\end{aligned}$$

where $F^{j}_{p,i} = F_{p,j-1}F_{p,j-2}\cdots F_{p,i}$ if $i < j$, $F^{i}_{p,i} = I$, and $F_{p,t} = F_t - K_tH_t$. Thus, we get the formula

$$\begin{aligned}\hat{u}_{t|n} &= S_t\Sigma_t^{-1}E_t + (Q_tG_t' - S_tK_t')\sum_{s=t+1}^{n}\left(F^{s}_{p,t+1}\right)' H_s'\Sigma_s^{-1}E_s \\ &= S_t\Sigma_t^{-1}E_t + (Q_tG_t' - S_tK_t')\lambda_{t+1}.\end{aligned}$$

As for its MSE, consider again that $(u_t - \hat{u}_{t|n}) + \hat{u}_{t|n} = u_t$ and that the two terms to the left of the equality are uncorrelated. □

4.10.2 Smoothing With the Single Disturbance State Space Model

As regards smoothing with the representation (4.9) and (4.10), the recursions for the adjoint variable, λ_t, and its MSE, Λ_t, coincide with (4.33) for $t = n, \ldots, 1$, initialized with $\lambda_{n+1} = 0$ and $\Lambda_{n+1} = 0$. The formulae for smoothing the state vector also coincide. For example, the fixed interval smoother is given for $t = n, \ldots, 1$ by the recursions (4.34).

To smooth the disturbances ϵ_t, the following recursions can be used for $t = n, \ldots, 1$

$$\hat{\epsilon}_{t|n} = J_t'\Sigma_t^{-1}E_t + M_t'\lambda_{t+1}, \qquad MSE(\hat{\epsilon}_{t|n}) = I - (J_t'\Sigma_t^{-1}J_t + M_t'\Lambda_{t+1}M_t),$$

where $M_t = G_t - K_tJ_t$ and $\hat{\epsilon}_{t|n} = E(\epsilon_t|Y_{1:n})$ is the orthogonal projection of ϵ onto $Y_{1:n}$.

Sometimes, if the MSE in the fixed interval smoother are not desired, it is advantageous to use the following fast forward recursion to compute $\hat{x}_{t|n}$

$$\hat{x}_{t+1|n} = F_t\hat{x}_{t|n} + G_t\hat{\epsilon}_{t|n},$$

initialized with $\hat{x}_{1|n} = \hat{x}_{1|0} + P_1\lambda_1$. If the state space model is (4.1) and (4.2), the recursion is

$$\hat{x}_{t+1|n} = F_t\hat{x}_{t|n} + G_t\hat{u}_{t|n},$$

with the same initialization.

4.10.3 The Rauch–Tung–Striebel Recursions

There are two versions of the Rauch–Tung–Striebel formulae (Rauch, Tung, & Striebel, 1965) for smoothing. Letting F_t^s be defined as in (4.22) for the state space model (4.1) and (4.2), the first one assumes that F_t^s is nonsingular, whereas in the second one F_t can be singular. The following theorem gives the recursions for the first case.

Theorem 4.23 *Suppose that the process* $\{Y_t\}$ *follows the state space model (4.1) and (4.2) and assume that*

1. F_t^s *is nonsingular*
2. $P_t > 0, R_t > 0$ *and* $Q_t^s > 0$,

where Q_t^s *and* F_t^s *are given by (4.19) and (4.22). Then,*

$$\hat{x}_{t|n} = F_{s,t}\hat{x}_{t+1|n} + \left(F_t^s\right)^{-1} G_t Q_t^s G_t' P_{t+1}^{-1}\hat{x}_{t+1|t} + \left(F_t^s\right)^{-1} G_t S_t R_t^{-1} Y_t$$
$$P_{t|n} = F_{s,t}P_{t+1|n}F_{s,t}' + \left(F_t^s\right)^{-1} G_t Q_t^r G_t' \left(F_t^s\right)^{-1'},$$

where

$$F_{s,t} = \left(F_t^s\right)^{-1}\left(I - G_t Q_t^s G_t' P_{t+1}^{-1}\right) = P_t F_{p,t}^{s'} P_{t+1}^{-1} = P_{t|t} F_t^{s'} P_{t+1}^{-1}, \tag{4.47}$$

$F_{p,t}^s = F_t^s - K_t^s H_t, K_t^s$ *is given by (4.23), and* $Q_t^r = Q_t^s - Q_t^s G_t' P_{t+1}^{-1} G_t Q_t^s$.

Proof Using (4.20), it is obtained that

$$x_{t+1} = F_t^s x_t + G_t u_t^s + G_t S_t R_t^{-1} Y_t. \tag{4.48}$$

Projecting onto the whole sample, we get

$$\hat{x}_{t+1|n} = F_t^s \hat{x}_{t|n} + G_t \hat{u}_{t|n}^s + G_t S_t R_t^{-1} Y_t,$$

where, according to the definition of $u_{t|n}^s$ and Theorem 4.22, the following equality holds

$$\begin{aligned}\hat{u}_{t|n}^s &= \hat{u}_{t|n} - S_t R_t^{-1}\hat{v}_{t|n}\\ &= S_t Y_t^i + Q_t G_t' \lambda_{t+1} - S_t R_t^{-1}\left(R_t Y_t^i + S_t' G_t' \lambda_{t+1}\right)\\ &= Q_t^s G_t' \lambda_{t+1}.\end{aligned}$$

Since P_{t+1} is nonsingular, we get from (4.36) the following equality

$$\lambda_{t+1} = P_{t+1}^{-1}\left(\hat{x}_{t+1|n} - \hat{x}_{t+1|t}\right).$$

Thus,

$$\hat{x}_{t+1|n} = F_t^s \hat{x}_{t|n} + G_t Q_t^s G_t' P_{t+1}^{-1} \left(\hat{x}_{t+1|n} - \hat{x}_{t+1|t} \right) + G_t S_t R_t^{-1} Y_t, \tag{4.49}$$

Multiplying (4.48) and (4.49) by the inverse of F_t^s and reordering terms, it is obtained that

$$\begin{aligned} x_t &= \left(F_t^s\right)^{-1} x_{t+1} - \left(F_t^s\right)^{-1} G_t u_t^s - \left(F_t^s\right)^{-1} G_t S_t R_t^{-1} Y_t \\ &= F_{s,t} x_{t+1} - \left(F_t^s\right)^{-1} G_t \left[u_t^s - Q_t^s G_t' P_{t+1}^{-1} x_{t+1} + S_t R_t^{-1} Y_t \right] \end{aligned} \tag{4.50}$$

and

$$\begin{aligned} \hat{x}_{t|n} &= \left(F_t^s\right)^{-1} \left(I - G_t Q_t^s G_t' P_{t+1}^{-1} \right) \hat{x}_{t+1|n} + \left(F_t^s\right)^{-1} G_t \left[Q_t^s G_t' P_{t+1}^{-1} \hat{x}_{t+1|t} - S_t R_t^{-1} Y_t \right] \\ &= F_{s,t} \hat{x}_{t+1|n} + \left(F_t^s\right)^{-1} G_t \left[Q_t^s G_t' P_{t+1}^{-1} \hat{x}_{t+1|t} - S_t R_t^{-1} Y_t \right]. \end{aligned} \tag{4.51}$$

Thus, the recursion for $\hat{x}_{t|n}$ is proved. To prove the recursion for $P_{t|n}$, subtract (4.51) from (4.50) to get

$$\tilde{x}_{t|n} = F_{s,t} \tilde{x}_{t+1|n} - \left(F_t^s\right)^{-1} G_t \tilde{u}_t^s, \tag{4.52}$$

where $\tilde{x}_{t|n} = x_t - \hat{x}_{t|n}$, $\tilde{u}_t^s = u_t^s - Q_t^s G_t' P_{t+1}^{-1} \tilde{x}_{t+1|t}$ and $\tilde{x}_{t+1|t} = x_{t+1} - \hat{x}_{t+1|t}$. The recursion for $P_{t|n}$ will be proved if we prove that $\tilde{x}_{t+1|n}$ and $\tilde{u}_t^s$ in (4.52) are uncorrelated and $\mathrm{Var}(\tilde{u}_t^s) = Q_t^r$. To prove the last equality, consider first that

$$\begin{aligned} \tilde{x}_{t+1|t} &= G_t u_t^s + F_t^s \tilde{x}_{t|t-1} - K_t^s \left(v_t + H_t \tilde{x}_{t|t-1} \right) \\ &= F_{p,t}^s \tilde{x}_{t|t-1} + G_t u_t^s - K_t^s v_t, \end{aligned} \tag{4.53}$$

and, therefore, $\mathrm{Cov}(u_t^s, \tilde{x}_{t+1|t}) = Q_t^s G_t'$ since the three terms in the right-hand side of (4.53) are mutually orthogonal. Then, by the definition of $\tilde{u}_t^s$, it holds that $\mathrm{Var}(\tilde{u}_t^s) = Q_t^s - Q_t^s G_t' P_{t+1}^{-1} G_t Q_t^s = Q_t^r$.

To prove that $\tilde{x}_{t+1|n}$ and $\tilde{u}_t^s$ in (4.52) are uncorrelated, use first (4.36) to get

$$\tilde{x}_{t+1|n} = \tilde{x}_{t+1|t} - P_{t+1} \lambda_{t+1} \tag{4.54}$$

$$\tilde{u}_t^s = u_t^s - Q_t^s G_t' P_{t+1}^{-1} \tilde{x}_{t+1|t}. \tag{4.55}$$

Then, by repeated application of (4.53), it is not difficult to verify that for $j \geq t+1$ it holds that

$$\mathrm{Cov}(\tilde{x}_{j|j-1}, \tilde{x}_{t+1|t}) = F_{p,t+1}^{j} \tag{4.56}$$

and

$$\mathrm{Cov}(\tilde{x}_{j|j-1}, u_t^s) = F_{p,t+1}^{j} G_t Q_t^s, \tag{4.57}$$

where $F_{p,i}^{j} = F_{p,j-1}F_{p,j-2}\cdots F_{p,i}$ if $i < j$, and $F_{p,i}^{i} = I$. Also, since

$$\begin{aligned} E_t &= Y_t - H_t\hat{x}_{t|t-1} \\ &= H_t\tilde{x}_{t|t-1} + v_t, \end{aligned} \tag{4.58}$$

clearly

$$\mathrm{Cov}(E_j, u_t^s) = H_j F_{p,t+1}^{j} G_t Q_t^s. \tag{4.59}$$

Given that $\lambda_{t+1} = \sum_{j=t+1}^{n} F_{p,j}^{t+1'} H_j' \Sigma_j^{-1} E_j$ in (4.54), if follows from (4.56)–(4.59) that

$$\mathrm{Cov}(\tilde{x}_{t+1|n}, \tilde{x}_{t+1|t}) = P_{t+1} - P_{t+1} \sum_{j=t+1}^{n} F_{p,j}^{t+1'} H_j' \Sigma_j^{-1} F_{p,j}^{t+1} P_{t+1} \tag{4.60}$$

and

$$\mathrm{Cov}(\tilde{x}_{t+1|n}, u_t^s) = G_t Q_t^s - P_{t+1} \sum_{j=t+1}^{n} F_{p,j}^{t+1'} H_j' \Sigma_j^{-1} F_{p,j}^{t+1} G_t Q_t^s. \tag{4.61}$$

Letting $B_t = \sum_{j=t+1}^{n} F_{p,j}^{t+1'} H_j' \Sigma_j^{-1} F_{p,j}^{t+1}$, it follows from (4.60) and (4.61) that the covariance between $\tilde{x}_{t+1|n}$ and $\tilde{u}_t^s$, given by (4.54) and (4.55), is

$$\begin{aligned} \mathrm{Cov}(\tilde{x}_{t+1|n}, \tilde{u}_t^s) &= \mathrm{Cov}(\tilde{x}_{t+1|n}, u_t^s) - \mathrm{Cov}(\tilde{x}_{t+1|n}, \tilde{x}_{t+1|t}) P_{t+1}^{-1} G_t Q_t^s \\ &= G_t Q_t^s - P_{t+1} B_t G_t Q_t^s - (P_{t+1} - P_{t+1} B_t P_{t+1})\, P_{t+1}^{-1} G_t Q_t^s \\ &= 0. \end{aligned}$$

Finally, to prove the formulae (4.47), first take into account the equality

$$P_{t+1} = F_t^s P_t F_{p,t}^{s'} + G_t Q_t^s G_t'.$$

Then, it follows that

$$\left(F_t^s\right)^{-1} \left(P_{t+1} - G_t Q_t^s G_t'\right) = P_t F_{p,t}^{s'},$$

and the first equality in (4.47) is proved. The second equality in (4.47) can be proved similarly considering the formula $P_{t+1} = F_t^s P_{t|t} F_t^{s'} + G_t Q_t^s G_t'$. □

The following theorem gives the original formulation of the Rauch–Tung–Striebel formulae, where the matrix F_t can be singular.

Theorem 4.24 (Original Formulation of the Rauch–Tung–Striebel Formulae) *Suppose that the process* $\{Y_t\}$ *follows the state space model (4.1) and (4.2) and assume that* $P_t > 0$. *Then,*

$$\hat{x}_{t|n} = \hat{x}_{t|t-1} + P_t H_t' \Sigma_t^{-1} E_t + F_{o,t} \left(\hat{x}_{t+1|n} - \hat{x}_{t+1|t} \right) \tag{4.62}$$

$$P_{t|n} = P_t - P_t H_t' \Sigma_t^{-1} H_t P_t - F_{o,t} \left(P_{t+1} - P_{t+1|n} \right) F_{o,t}', \tag{4.63}$$

where $F_{o,t} = P_t' F_{p,t} P_{t+1}^{-1}$ *and* $F_{p,t} = F_t - K_t H_t$.

Proof Using (4.33) and (4.34), it is obtained that

$$\begin{aligned} \hat{x}_{t|n} &= \hat{x}_{t|t-1} + P_t \left(F_{p,t}' \lambda_{t+1} + H_t' \Sigma_t^{-1} E_t \right) \\ &= \hat{x}_{t|t-1} + P_t H_t' \Sigma_t^{-1} E_t + P_t F_{p,t}' \lambda_{t+1}. \end{aligned} \tag{4.64}$$

Since P_t is nonsingular, it follows from (4.34) that

$$\lambda_{t+1} = P_{t+1}^{-1} \left(\hat{x}_{t+1|n} - \hat{x}_{t+1|t} \right).$$

Substituting the previous equality into (4.64), we get (4.62). To prove (4.63), use first (4.64) to obtain

$$x_t - \hat{x}_{t|t-n} + P_t H_t' \Sigma_t^{-1} E_t + P_t F_{p,t}' \lambda_{t+1} = x_t - \hat{x}_{t|t-1}.$$

Then, since the terms on the left-hand side of the previous equality are uncorrelated, it holds that

$$P_{t|n} + P_t H_t' \Sigma_t^{-1} H_t P_t + P_t F_{p,t}' \Lambda_{t+1} F_{p,t} P_t = P_t. \tag{4.65}$$

Since P_t is nonsingular, it follows from (4.34) that

$$\Lambda_{t+1} = P_{t+1}^{-1} \left(P_{t+1} - P_{t+1|n} \right) P_{t+1}^{-1}.$$

Substituting the previous equality into (4.65) yields (4.63). □

4.10.4 Square Root Smoothing

We will consider in the following the square root form of the Bryson–Frazier formulae.

Theorem 4.25 *Suppose that the process $\{Y_t\}$ follows the state space model (4.9) and (4.10) or the state space model (4.1) and (4.2) and let $\widehat{E}_t = \Sigma_t^{-1/2}E_t, \widehat{H}_t = \Sigma_t^{-1/2}H_t, \widehat{K}_t = K_t\Sigma_t^{1/2}$ and $F_{p,t} = F_t - \widehat{K}_t\widehat{H}_t$. Then, the QR algorithm produces an orthogonal matrix U_t such that*

$$U_t' \left[\begin{array}{c|c} \widehat{H}_t & \widehat{E}_t \\ \Lambda_{t+1}^{1/2'} F_{p,t} & \Lambda_{t+1}^{-1/2}\lambda_{t+1} \end{array} \right] = \left[\begin{array}{c|c} \Lambda_t^{1/2'} & \Lambda_t^{-1/2}\lambda_t \\ 0 & * \end{array} \right],$$

where $\Lambda_{n+1}^{1/2} = 0$ and $\Lambda_{n+1}^{-1/2}\lambda_{n+1} = 0$.

Proof Let

$$U_t' \left[\begin{array}{c|c} \widehat{H}_t & \widehat{E}_t \\ \Lambda_{t+1}^{1/2'} F_{p,t} & \Lambda_{t+1}^{-1/2}\lambda_{t+1} \end{array} \right] = \left[\begin{array}{c|c} X & Y \\ 0 & * \end{array} \right]$$

Then, premultiplying the previous matrices by their respective transposes and equating entries, it is obtained that

$$\widehat{E}_t'\widehat{H}_t + \left(\Lambda_{t+1}^{-1/2}\lambda_{t+1}\right)' \Lambda_{t+1}^{1/2'} F_{p,t} = Y'X, \qquad \widehat{H}_t'\widehat{H}_t + \left(\Lambda_{t+1}^{1/2'} F_{p,t}\right)' \Lambda_{t+1}^{1/2'} F_{p,t} = X'X.$$

By (4.33), the theorem follows. □

If the process $\{Y_t\}$ follows the state space model (4.9) and (4.10) or the state space model (4.1) and (4.2), using the previous theorem it is possible to obtain a square root form of the fixed interval smoother as follows.

Step 1 In the forward pass, use Theorem 4.4 or Theorem 4.5 to compute and store the quantities $\widehat{E}_t, \widehat{K}_t, \widehat{H}_t, F_{p,t} = F_t - \widehat{K}_t\widehat{H}_t, \hat{x}_{t|t-1}$, and $P_t^{1/2}$.

Step 2 In the backward pass, use Theorem 4.25 to compute $\Lambda_t^{1/2'}P_t^{1/2}$ and $\Lambda_t^{-1/2}\lambda_t$, where $\Lambda_{n+1}^{1/2} = 0$ and $\Lambda_{n+1}^{-1/2}\lambda_{n+1} = 0$. Finally, at the same time, compute recursively the fixed interval smoothing quantities

$$\hat{x}_{t|n} = \hat{x}_{t|t-1} + P_t^{1/2}\left(\Lambda_t^{1/2'}P_t^{1/2}\right)'\left(\Lambda_t^{-1/2}\lambda_t\right)$$

$$P_{t|n} = P_t^{1/2}\left[I - \left(\Lambda_t^{1/2'}P_t^{1/2}\right)'\left(\Lambda_t^{1/2'}P_t^{1/2}\right)\right]P_t^{1/2'}.$$

The following theorem gives another version of the square root form of the Bryson–Frazier formulae corresponding to the situation in which the process $\{Y_t\}$ follows the state space model (4.1) and (4.2) under additional assumptions.

Theorem 4.26 *Suppose that the process $\{Y_t\}$ follows the state space model (4.1) and (4.2), where it is assumed that $Q_t > 0$, and let $R_t^s = R_t - S_t'Q_t^{-1}S_t$. Then, if R_t^s and P_t are nonsingular, the application of the QR algorithm in a forward pass yields* *an orthogonal matrix U_t such that*

$$U_t' \left[\begin{array}{cc|cc} (R_t^s)^{1/2'} & 0 & 0 & -(R_t^s)^{-1/2} Y_t \\ P_t^{1/2'} H_t' & P_t^{1/2'} F_t' & I & P_t^{-1/2} \hat{x}_{t|t-1} \\ Q_t^{-1/2} S_t & Q_t^{1/2'} G_t' & 0 & 0 \end{array} \right] = \left[\begin{array}{cc|cc} \Sigma_t^{1/2'} & \widehat{K}_t' & X_t & -\widehat{E}_t \\ 0 & P_{t+1}^{1/2'} & Z_t & P_{t+1}^{-1/2} \hat{x}_{t+1|t} \\ 0 & 0 & * & * \end{array} \right],$$

where $\widehat{K}_t = (F_t P_t H_t' + G_t S_t)\Sigma_t^{-1/2'} = K_t \Sigma_t^{1/2}, \widehat{E}_t = \Sigma_t^{-1/2} E_t,$

$$X_t = \Sigma_t^{-1/2} H_t P_t^{1/2}, \qquad Z_t = P_{t+1}^{-1/2} F_{p,t} P_t^{1/2},$$

$F_{p,t} = F_t - K_t H_t$, *and the asterisk indicates an element that is not relevant to our purposes. Here it is understood that U_t upper-triangularizes the first two block columns of the left-hand side matrix. Using X_t and Z_t, the application of the QR algorithm in a backward pass yields an orthogonal matrix V_t such that*

$$V_t' \left[\begin{array}{c|c} X_t & \widehat{E}_t \\ \left(\Lambda_{t+1}^{1/2'} P_{t+1}^{1/2}\right) Z_t & \Lambda_{t+1}^{-1/2} \lambda_{t+1} \end{array} \right] = \left[\begin{array}{c|c} \Lambda_t^{1/2'} P_t^{1/2} & \Lambda_t^{-1/2} \lambda_t \\ 0 & * \end{array} \right],$$

where V_t upper-triangularizes the first two block columns of the left-hand side matrix and the asterisk indicates an element that is not relevant to our purposes.

Proof Most of the results for the forward pass have been proved in Theorem 4.5. To complete the proof for the forward pass, premultiply the matrices on both sides by their respective transposes and equate entries to get

$$H_t P_t^{1/2} = \Sigma_t^{1/2} X_t, \qquad F_t P_t^{1/2} = \widehat{K}_t X_t + P_{t+1}^{1/2} Y_t.$$

The result for the backward pass is proved using an argument similar to that of the proof of Theorem 4.25. □

Using Theorem 4.26, the square root form of the fixed interval smoother is as follows.

Step 1 In the forward pass, compute and store the quantities $\widehat{E}_t, X_t, Z_t, P_t^{-1/2} \hat{x}_{t|t-1}$ and $P_t^{1/2}$.

Step 2 In the backward pass, compute $\Lambda_t^{1/2'} P_t^{1/2}$ and $\Lambda_t^{-1/2} \lambda_t$, where $\Lambda_{n+1}^{1/2'} P_{n+1}^{1/2} = 0$ and $\Lambda_{n+1}^{-1/2} \lambda_{n+1} = 0$. Finally, at the same time, compute recursively the fixed interval smoothing quantities

$$\hat{x}_{t|n} = P_t^{1/2} \left[\left(P_t^{-1/2} \hat{x}_{t|t-1} \right) + \left(\Lambda_t^{1/2'} P_t^{1/2} \right)' \left(\Lambda_t^{-1/2} \lambda_t \right) \right]$$

$$P_{t|n} = P_t^{1/2} \left[I - \left(\Lambda_t^{1/2'} P_t^{1/2} \right)' \left(\Lambda_t^{1/2'} P_t^{1/2} \right) \right] P_t^{1/2'}.$$

4.10.5 Square Root Information Smoothing

It is possible to obtain a square root information smoother by first using the square root information filter of Theorem 4.17 in a forward pass and then putting the Rauch–Tung–Striebel smoothing formulae into square root form. The next theorem gives the details, but first we need a lemma.

Lemma 4.5 *Under the assumptions and with the notation of Theorem 4.17, the following equality holds*

$$\widehat{K}_{b,t} = (Q_t^r)^{-1/2'} Q_t^s G_t' P_{t+1}^{-1}.$$

Proof We will prove that $P_{t+1}^{-1} G_t Q_t^s \left(Q_t^r\right)^{-1} = A_t G_t$. Using that $\left(Q_t^r\right)^{-1} = \left(Q_t^s\right)^{-1} + G_t' A_t G_t$, it is obtained that

$$\begin{aligned}
P_{t+1}^{-1} G_t Q_t^s \left(Q_t^r\right)^{-1} &= P_{t+1}^{-1} G_t Q_t^s \left[\left(Q_t^s\right)^{-1} + G_t' A_t G_t\right] \\
&= P_{t+1}^{-1} G_t \left(I + Q_t^s G_t' A_t G_t\right) \\
&= \left(A_t - A_t G_t Q_t^r G_t' A_t\right)\left(I + G_t Q_t^s G_t' A_t\right) G_t \\
&= \left[A_t - A_t G_t \left(-Q_t^r + Q_t^s - Q_t^r G_t' A_t G_t Q_t^s\right) G_t' A_t\right] G_t \\
&= \left(A_t - A_t G_t \left\{-Q_t^r + Q_t^s - Q_t^r \left[\left(Q_t^r\right)^{-1} - \left(Q_t^s\right)^{-1}\right] Q_t^s\right\} G_t' A_t\right) G_t \\
&= \left(A_t - A_t G_t \left\{-Q_t^r + Q_t^s - Q_t^s + Q_t^r\right\} G_t' A_t\right) G_t \\
&= A_t G_t.
\end{aligned}$$

□

Theorem 4.27 *Under the assumptions and with the notation of Theorem 4.17, suppose we apply the algorithm of that theorem in a forward pass. Then, the application of the QR algorithm in a backward pass yields an orthogonal matrix* V_t *such that*

$$V_t' \left[\begin{array}{c|c} P_{t+1|n}^{1/2'}\left[I - G_t (Q_t^r)^{1/2} \widehat{K}_{b,t}\right](F_t^s)^{-1'} & P_{t+1|n}^{-1/2} \hat{x}_{t+1|n} \\ (Q_t^r)^{1/2'} G_t' (F_t^s)^{-1'} & \widehat{K}_{b,t} \hat{x}_{t+1|t} + (Q_t^r)^{-1/2} S_t R_t^{-1} Y_t \end{array}\right] = \left[\begin{array}{c|c} P_{t|n}^{1/2'} & P_{t|n}^{-1/2} \hat{x}_{t|n} \\ 0 & * \end{array}\right],$$

initialized with $P_{n+1|n}^{1/2} = P_{n+1}^{1/2}$ *and* $P_{n+1|n}^{-1/2} \hat{x}_{n+1|n} = \left(P_{n+1}^{-1/2}\right) \hat{x}_{n+1|n}$. *In addition, the computation of* $(Q_t^r)^{1/2}$ *in the backward pass by inverting* $(Q_t^r)^{-1/2}$, *given by the forward pass, can be avoided if we incorporate a block column in the algorithm of*

Theorem 4.17 as follows

$$U_t' \left[\begin{array}{cc|cc} (Q_t^s)^{-1/2} & 0 & (Q_t^s)^{1/2'} & (Q_t^s)^{-1/2} S_t R_t^{-1} Y_t \\ -P_t^{-1/2}(F_t^s)^{-1} G_t & P_t^{-1/2}(F_t^s)^{-1} & 0 & P_t^{-1/2} \hat{x}_{t|t-1} \\ R_t^{-1/2} H_t (F_t^s)^{-1} G_t & -R_t^{-1/2} H_t (F_t^s)^{-1} & 0 & -R_t^{-1/2} Y_t \end{array} \right]$$

$$= \left[\begin{array}{cc|cc} (Q_t^r)^{-1/2} & -\widehat{K}_{b,t} & (Q_t^r)^{1/2'} & -\widehat{K}_{b,t}\hat{x}_{t+1|t} + (Q_t^r)^{-1/2} S_t R_t^{-1} Y_t \\ 0 & P_{t+1}^{-1/2} & P_{t+1}^{1/2'} A_t G_t Q_t^r & P_{t+1}^{-1/2} \hat{x}_{t+1|t} \\ 0 & 0 & * & -\widehat{E}_t \end{array} \right].$$

Proof Let

$$V_t' \left[\begin{array}{c|c} P_{t+1|n}^{1/2'} \left[I - G_t (Q_t^r)^{1/2} \widehat{K}_{b,t} \right] (F_t^s)^{-1'} & P_{t+1|n}^{-1/2} \hat{x}_{t+1|n} \\ (Q_t^r)^{1/2'} G_t' (F_t^s)^{-1'} & \widehat{K}_{b,t}\hat{x}_{t+1|t} + (Q_t^r)^{-1/2} S_t R_t^{-1} Y_t \end{array} \right]$$

$$= \left[\begin{array}{c|c} X & \alpha \\ 0 & \beta \end{array} \right].$$

Then, premultiplying both sides of the previous equality by their respective transposes, using the Rauch–Tung–Striebel recursions and Lemma 4.5, and equating entries, it is obtained that

$$\begin{aligned} &(F_t^s)^{-1} \left[I - G_t (Q_t^r)^{1/2} \widehat{K}_{b,t} \right]' P_{t+1|n} \left[I - G_t (Q_t^r)^{1/2} \widehat{K}_{b,t} \right] (F_t^s)^{-1'} + (F_t^s)^{-1} G_t Q_t^r G_t' (F_t^s)^{-1'} \\ &= F_{s,t} P_{t+1|n} F_{s,t}' + (F_t^s)^{-1} G_t Q_t^r G_t' (F_t^s)^{-1'} \\ &= P_{t|n} \\ &= X'X \end{aligned}$$

and

$$\begin{aligned} &(F_t^s)^{-1} \left\{ \left[I - G_t (Q_t^r)^{1/2} \widehat{K}_{b,t} \right]' P_{t+1|n} \hat{x}_{t+1|t} + G_t (Q_t^r)^{1/2} \left[\widehat{K}_{b,t} \hat{x}_{t+1|t} + (Q_t^r)^{-1/2} S_t R_t^{-1} Y_t \right] \right\} \\ &\qquad = F_{s,t} \hat{x}_{t+1|t} + (F_t^s)^{-1} G_t \left(Q_t^r \right)^{1/2} \left[\widehat{K}_{b,t} \hat{x}_{t+1|t} + \left(Q_t^r \right)^{-1/2} S_t R_t^{-1} Y_t \right] \\ &\qquad = x_{t|n} \\ &\qquad = X'\alpha. \end{aligned}$$

The last part of the theorem can be proved by inverting the equality

$$U_t' \begin{bmatrix} (Q_t^s)^{-1/2} & 0 & 0 \\ -P_t^{-1/2}(F_t^s)^{-1} G_t & P_t^{-1/2}(F_t^s)^{-1} & 0 \\ R_t^{-1/2} H_t (F_t^s)^{-1} G_t & -R_t^{-1/2} H_t (F_t^s)^{-1} & I \end{bmatrix} = \begin{bmatrix} (Q_t^r)^{-1/2} & -\widehat{K}_{b,t} & * \\ 0 & P_{t+1}^{-1/2} & * \\ 0 & 0 & * \end{bmatrix}.$$

□

Using Theorem 4.27, the square root form of the fixed interval smoother is as follows.

Step 1 In the forward pass, use the algorithm of Theorem 4.17, modified as in Theorem 4.27, to compute and store the quantities $(Q_t^r)^{1/2}$, $\widehat{K}_{b,t}$, $(Q_t^r)^{-1/2}S_tR_t^{-1}Y_t$, and $\widehat{K}_{b,t}\hat{x}_{t+1|t} = -\left(-\widehat{K}_{b,t}\hat{x}_{t+1|t} + (Q_t^r)^{-1/2}S_tR_t^{-1}Y_t\right) + (Q_t^r)^{-1/2}S_tR_t^{-1}Y_t$.

Step 2 In the backward pass, compute $P_{t|n}^{1/2'}$ and $P_{t|n}^{-1/2}\hat{x}_{t|n}$ by means of the algorithm of Theorem 4.27. Finally, at the same time, compute recursively the fixed interval smoothing quantities

$$\hat{x}_{t|n} = P_{t|n}^{1/2}\left(P_{t|n}^{-1/2}\hat{x}_{t|n}\right)$$
$$P_{t|n} = P_{t|n}^{1/2}P_{t|n}^{1/2'}.$$

4.11 Covariance-Based Filters

Assume the state space model (4.1) and (4.2). The following theorem was proved in Chap. 1 using results connected with the Innovations Algorithm. Here we provide a more direct proof.

Theorem 4.28 *Suppose that the process $\{Y_t\}$ follows the state space model (4.1) and (4.2) so that the covariances, $\gamma_Y(r,s) = \mathrm{E}(Y_rY_s')$, are generated by (1.34). Then, $\{Y_t\}$ admits the following innovations state space representation*

$$\hat{x}_{t+1|t} = F_t\hat{x}_{t|t-1} + K_tE_t$$
$$Y_t = H_t\hat{x}_{t|t-1} + E_t,$$

where $\hat{x}_{t|t-1}$, K_t and E_t have the same interpretation as in the Kalman filter and the quantities $Var(E_t) = \Sigma_t$, $Var(\hat{x}_{t|t-1}) = \widehat{\Sigma}_t$ and K_t are obtained by means of the recursions based on covariance data only

$$\begin{aligned}\Sigma_t &= \gamma_Y(t,t) - H_t\widehat{\Sigma}_tH_t' \\ K_t &= (N_t - F_t\widehat{\Sigma}_tH_t')\Sigma_t^{-1} \\ \widehat{\Sigma}_{t+1} &= F_t\widehat{\Sigma}_tF_t' + K_t\Sigma_tK_t',\end{aligned} \tag{4.66}$$

initialized with $\hat{x}_{1|0} = 0$ and $\widehat{\Sigma}_1 = 0$.

The whitening filter for $\{Y_t\}$ is further given by

$$\hat{x}_{t+1|t} = (F_t - K_tH_t)\hat{x}_{t|t-1} + K_tY_t$$
$$E_t = H_t\hat{x}_{t|t-1} - Y_t.$$

Proof The problem is to determine the innovations, $\{E_t\}$, using only knowledge of the covariance parameters, $\{F_t, H_t, N_t, \gamma_Y(t,t)\}$, where $N_t = F_t\Pi_t H_t' + G_t S_t$. The problem will be solved if we can re-express $\{K_t, \Sigma_t\}$ in terms of the covariance data. To this end, define first $\text{Var}(\hat{x}_{t|t-1}) = \widehat{\Sigma}_t$. Then, use the following equality

$$x_t = x_t - \hat{x}_{t|t-1} + \hat{x}_{t|t-1}$$

to get

$$\Pi_t = P_t + \widehat{\Sigma}_t,$$

because $x_t - \hat{x}_{t|t-1}$ is orthogonal to $\hat{x}_{t|t-1}$. In addition, using the Kalman filter equation

$$\hat{x}_{t+1|t} = F_t\hat{x}_{t|t-1} + K_t E_t,$$

we get the covariance equation

$$\widehat{\Sigma}_{t+1} = F_t\widehat{\Sigma}_t F_t' + K_t\Sigma_t K_t',$$

because $\{E_t\}$ is a white noise process. Hence, we have

$$\begin{aligned}
\Sigma_t &= R_t + H_t P_t H_t' = R_t + H_t(\Pi_t - \widehat{\Sigma}_t)H_t' \\
&= \gamma_Y(t,t) - H_t\widehat{\Sigma}_t H_t' \\
K_t &= (F_t P_t H_t' + G_t S_t)\Sigma_t^{-1} = (F_t\Pi_t H_t' + G_t S_t - F_t\widehat{\Sigma}_t H_t')\Sigma_t^{-1} \\
&= (N_t - F_t\widehat{\Sigma}_t H_t')\Sigma_t^{-1}.
\end{aligned}$$

□

4.12 Markov Processes

A stochastic vector process $\{Y_t : t \in \mathbb{Z}\}$ is said to be **strictly Markovian** or a **strict sense Markov process** if the conditional distribution functions satisfy

$$F(Y_{i_k}|Y_{i_{k-1}}, \ldots, Y_{i_1}) = F(Y_{i_k}|Y_{i_{k-1}}), \quad i_k > i_{k-1} > \ldots > i_1, \quad i_j \in \mathbb{Z}, j = 1, 2, \ldots, k.$$

An example of these processes is the scalar normal AR(1) process $\{Y_t : t \in \mathbb{Z}\}$, defined by

$$Y_t = \rho Y_{t-1} + A_t, \tag{4.67}$$

where $0 < \rho < 1$ and $\{A_t\}$ is a sequence of independent normal variables with zero mean and common variance σ^2.

As it happens with stationarity, if we are only using first and second order statistical information, we may define, following Doob (1953), a **wide sense Markov process** as a process $\{Y_t : t \in \mathbb{Z}\}$ such that

$$E^*(Y_{i_k}|Y_{i_{k-1}},\ldots,Y_{i_1}) = E^*(Y_{i_k}|Y_{i_{k-1}}), \quad i_k > i_{k-1} > \ldots > i_1,$$
$$i_j \in \mathbb{Z}, j = 1,2,\ldots,k.$$

The following lemma gives a characterization of wide sense Markov processes in terms of an equivalent but simpler definition.

Lemma 4.6 *A stochastic process $\{Y_t : t \in \mathbb{Z}\}$ is wide sense Markov if, and only if,*

$$E^*(Y_i|Y_j,Y_k) = E^*(Y_i|Y_j), \quad i > j > k, \quad i,j,k \in \mathbb{Z}.$$

Proof If $\{Y_t : t \in \mathbb{Z}\}$ is wide sense Markov, the condition clearly holds. To prove the converse, suppose the condition applies and let $i_j \in \mathbb{Z}, j = 1,2,\ldots,k$, such that $i_k > i_{k-1} > \ldots > i_1$. Define $Z_2 = (Y_{i_{k-1}},\ldots,Y_{i_3})$ and $Z_1 = (Y_{i_2},Y_{i_1})$. Then, by formula (1.44), that corresponds to the sequential update of an orthogonal projection, it is obtained that

$$\begin{aligned} E^*(Y_{i_k}|Y_{i_{k-1}},\ldots,Y_{i_1}) &= E^*(Y_{i_k}|Z_2,Z_1) \\ &= E^*(Y_{i_k}|Z_1) + \Pi\left[Z_2 - E^*(Z_2|Z_1)\right] \\ &= E^*(Y_{i_k}|Y_{i_2}) + \Pi\left[Z_2 - E^*(Z_2|Y_{i_2})\right] \\ &= E^*(Y_{i_k}|Y_{i_{k-1}},\ldots,Y_{i_2}), \end{aligned}$$

where $E^*(Y_{i_k}|E_2) = \Pi E_2$ and $E_2 = Z_2 - E^*(Z_2|Z_1) = Z_2 - E^*(Z_2|Y_{i_2})$. Repeating this procedure, we finally get $E^*(Y_{i_k}|Y_{i_{k-1}},\ldots,Y_{i_1}) = E^*(Y_{i_k}|Y_{i_{k-1}})$. □

Sometimes the following lemma can be useful when checking whether a process is wide sense Markov.

Lemma 4.7 (Covariance Test for Wide Sense Markov Processes) *Assume that $Var(x_j) > 0, j \geq 1$. Then, the process $\{x_j : j = 1,2,\ldots\}$ is wide sense Markov if, and only if, for any $i > j > k$*

$$Cov(x_i,x_k) = Cov(x_i,x_j)Var^{-1}(x_j)Cov(x_j,x_k).$$

Proof We will first prove the proposition that $\{x_j\}$ is wide sense Markov if, and only if,

$$x_i - E^*(x_i|x_j) \perp x_k, \qquad i > j > k.$$

The condition is evident if $\{x_j\}$ is wide sense Markov. Conversely, if the condition is true, then, by the properties of orthogonal projection,

$$\begin{aligned} 0 &= E^*\left[x_i - E^*(x_i|x_j)|x_j, x_k\right] \\ &= E^*(x_i|x_j, x_k) - E^*(x_i|x_j), \end{aligned}$$

and the process is wide sense Markov by Lemma 4.6. Using the previous proposition, we get

$$\begin{aligned} 0 &= \mathrm{Cov}(x_i - E^*(x_i|x_j), x_k) \\ &= \mathrm{Cov}(x_i, x_k) - \mathrm{Cov}(x_i, x_j)\mathrm{Var}^{-1}(x_j)\mathrm{Cov}(x_j, x_k). \end{aligned}$$

□

Remark 4.3 The previous lemma is also valid if the covariance matrices $\mathrm{Var}(x_j), j \geq 1$, are singular. See Problem 4.3. ◇

4.12.1 Forwards Markovian Models

Given that the previous AR(1) example (4.67) is strict and wide sense Markov, the question arises as to whether the process $\{x_t\}$ is also wide sense Markov, where x_t is the state vector of the state space model (4.1) and (4.2). This is motivated by the fact that the transition equation (4.1) resembles the equation of the AR(1) model (4.67). The answer is affirmative, but what is perhaps surprising is that the converse also holds, as shown by the following theorem.

Theorem 4.29 *A stochastic process $\{x_t : t = 1, 2, \ldots\}$ is wide sense Markov if, and only if, it can be represented in the form*

$$x_{t+1} = F_t x_t + G_t u_t, \qquad t = 1, 2, \ldots, \tag{4.68}$$

where

$$E\left\{\begin{bmatrix} u_t \\ x_1 \end{bmatrix}\left[u_s', x_1'\right]\right\} = \begin{bmatrix} Q_t\delta_{ts} & 0 \\ 0 & \Pi_1 \end{bmatrix}.$$

Proof Suppose $\{x_t\}$ follows the model (4.68). Then, using the properties of orthogonal projection, it is obtained that

$$\begin{aligned} E^*(x_i|x_j, x_k) &= F_{i-1}E^*(x_{i-1}|x_j, x_k) + G_{i-1}E^*(u_{i-1}|x_j, x_k) \\ &= F_{i-1}\cdots F_j E^*(x_j|x_j, x_k) \\ &= F_{i-1}\cdots F_j x_j \\ &= E^*(x_i|x_j), \qquad i > j > k, \end{aligned}$$

and $\{x_t\}$ is wide sense Markov by Lemma 4.6. Here, we have used the facts that, by (4.68), x_t is a linear combination of the vectors $\{x_1, u_1, \ldots, u_{t-1}\}$ and u_j is orthogonal to x_k for all $j \geq k \geq 1$.

To prove the converse, suppose that $\{x_t : t = 1, 2, \ldots\}$ is wide sense Markov and define

$$\begin{aligned} e_1 &= x_1 \\ e_{t+1} &= x_{t+1} - E^*(x_{t+1}|x_t), \qquad t \geq 1. \end{aligned}$$

Then, by definition, $E^*(x_{t+1}|x_t) = F_t x_t$, where F_t is any solution of the normal equations $F_t \mathrm{Var}(x_t) = \mathrm{Cov}(x_{t+1}, x_t)$. Thus, we can write

$$x_{t+1} = F_t x_t + u_t,$$

where $u_t = e_{t+1}$ and $\mathrm{Cov}(u_t, u_s) = \mathrm{Cov}(e_{t+1}, e_{s+1}) = 0, t \neq s$, by the orthogonality of the innovations since, by the wide sense Markov property, $E^*(x_{t+1}|x_t) = E^*(x_{t+1}|x_t, \ldots, x_1)$. Letting $\Pi_t = \mathrm{Var}(x_t)$ and $Q_t = \mathrm{Var}(u_t)$, we have, by the properties of orthogonal projection, the relation $Q_t = \mathrm{Var}(e_{t+1}) = \Pi_{t+1} - F_t \Pi_t F_t'$. Finally, notice that $u_t = e_{t+1}$ is orthogonal to $e_1 = x_1$ and, therefore,

$$E\left\{\begin{bmatrix} u_t \\ x_1 \end{bmatrix} \begin{bmatrix} u_s', x_1' \end{bmatrix}\right\} = \begin{bmatrix} Q_t \delta_{ts} & 0 \\ 0 & \Pi_1 \end{bmatrix}.$$

□

4.12.2 Backwards Markovian Models

We have seen in the previous section that a wide sense Markov process, $\{x_t : t = 1, 2, \ldots\}$, can always be represented by a forwards time model (4.68). Suppose that the matrices F_t in (4.68) are invertible. Then, we can reverse time and write

$$x_t = F_t^{-1} x_{t+1} - F_t^{-1} G_t u_t, \qquad t \leq n, \tag{4.69}$$

and the question arises as to whether this model is a backwards time Markovian model. As it turns out, the assumption in (4.68) that the innovations, $\{u_t : t = 1, 2, \ldots\}$, are uncorrelated with the initial state, x_1, is a crucial one. It is this assumption that fails in (4.69). That is, it is not true that $\mathrm{Cov}(u_t, x_{n+1}) = 0$ for $t \leq n$. For this reason, the model (4.69) is not a backwards time Markovian model, although the sequence $\{u_t : t = 1, 2, \ldots\}$ continues to be uncorrelated.

The following lemma shows that the properties of a wide sense Markov process do not change when the time is reversed.

Lemma 4.8 (Independence from Time Direction) *Let* $\{x_t : t = 1, 2, \ldots\}$ *be a wide sense Markov process. Then,*

$$E^*(x_k|x_i, x_j) = E^*(x_k|x_j), \qquad i > j > k, \qquad i, j, k \in \mathbb{Z}.$$

Proof Suppose first that $\mathrm{Var}(x_j) > 0, j \geq 1$. Then, by Lemma 4.7,

$$\mathrm{Cov}(x_i, x_k) = \mathrm{Cov}(x_i, x_j)\mathrm{Var}^{-1}(x_j)\mathrm{Cov}(x_j, x_k).$$

Transposing in the previous equality, considering the process $\{x_t : t = 1, 2, \ldots\}$ with the time reversed, and applying Lemma 4.7 again, the lemma is proved. In the general case, we can consider generalized inverses instead of inverses of $\mathrm{Var}(x_j)$ in Lemma 4.7. See Remark 4.3. □

It is in fact true that any wide sense Markov process, $\{x_t : t = 1, 2, \ldots\}$, can always be represented by a backwards time Markovian model. Here, a **backwards Markovian model** for a stochastic process, $\{x_t : t = 1, 2, \ldots\}$, is defined as

$$x_t = F^b_{t+1}x_{t+1} + G^b_{t+1}u^b_{t+1}, \qquad t = n, n-1, \ldots, 1, \tag{4.70}$$

where

$$E\left\{\begin{bmatrix} u^b_{t+1} \\ x_{n+1} \end{bmatrix}\begin{bmatrix} u^{b\prime}_{s+1}, x'_{n+1} \end{bmatrix}\right\} = \begin{bmatrix} Q^b_{t+1}\delta_{ts} & 0 \\ 0 & \Pi_{n+1} \end{bmatrix}.$$

Using Lemma 4.8 and proceeding as in the proof of Theorem 4.29 with the time reversed, it is not difficult to prove the following analogous theorem. We omit its proof.

Theorem 4.30 *A stochastic process* $\{x_t : t = n+1, n, \ldots\}$ *is wide sense Markov if, and only if, it can be represented in the form*

$$x_t = F^b_{t+1}x_{t+1} + G^b_{t+1}u^b_{t+1}, \qquad t = n, n-1, \ldots,$$

where

$$E\left\{\begin{bmatrix} u^b_{t+1} \\ x_{n+1} \end{bmatrix}\begin{bmatrix} u^{b\prime}_{s+1}, x'_{n+1} \end{bmatrix}\right\} = \begin{bmatrix} Q^b_{t+1}\delta_{ts} & 0 \\ 0 & \Pi_{n+1} \end{bmatrix}.$$

4.12.3 Backwards Models From Forwards State Space Models

Our aim in this section is to show that it is possible to construct a backwards state space model given a forwards state space model. Suppose the forwards state space model

$$x_{t+1} = F_t x_t + G_t u_t \tag{4.71}$$

$$Y_t = H_t x_t + v_t, \qquad t = 1, \ldots, n, \tag{4.72}$$

where

$$E\left\{\begin{bmatrix} u_t \\ v_t \end{bmatrix}\begin{bmatrix} u_s', v_s' \end{bmatrix}\right\} = \begin{bmatrix} Q_t & S_t \\ S_t' & R_t \end{bmatrix}\delta_{ts},$$

$E(u_t) = 0, E(v_t) = 0$, the initial state vector, x_1, is orthogonal to u_t and v_t for all $t, E(x_1) = 0$ and $\mathrm{Var}(x_1) = \Pi_1$.

To obtain a backwards state space model, we first need a lemma.

Lemma 4.9 *Given a forwards Markovian model*

$$x_{t+1} = F_t x_t + G_t u_t, \qquad t = 1, 2, \ldots, n, \tag{4.73}$$

where

$$E\left\{\begin{bmatrix} u_t \\ x_1 \end{bmatrix}\begin{bmatrix} u_s', x_1' \end{bmatrix}\right\} = \begin{bmatrix} Q_t\delta_{ts} & 0 \\ 0 & \Pi_1 \end{bmatrix},$$

there exists a backwards Markovian model

$$x_t = F^b_{t+1} x_{t+1} + u^b_{t+1}, \qquad t = n, n-1, \ldots, 1,$$

with

$$E\left\{\begin{bmatrix} u^b_{t+1} \\ x_{n+1} \end{bmatrix}\begin{bmatrix} u^{b\prime}_{s+1}, x_{n+1}' \end{bmatrix}\right\} = \begin{bmatrix} Q^b_{t+1}\delta_{ts} & 0 \\ 0 & \Pi_{n+1} \end{bmatrix},$$

where F^b_{t+1} is any solution of the equation $F^b_{t+1}\Pi_{t+1} = \Pi_t F_t', Q^b_{t+1} = \Pi_t - F^b_{t+1}\Pi_{t+1}F^{b\prime}_{t+1}$, and $Var(x_t) = \Pi_t$ satisfies $\Pi_{t+1} = F_t\Pi_t F_t' + G_t Q_t G_t'$.

Proof We could use Theorem 4.30, but we will give a proof from scratch. Define the backwards innovations as $e^b_{n+1} = x_{n+1}$, and, for $t = n, \ldots, 1$,

$$\begin{aligned} e^b_t &= x_t - E^*(x_t|x_{t+1}, \ldots, x_{n+1}) \\ &= x_t - E^*(x_t|x_{t+1}) \\ &= x_t - K^b_{t+1}x_{t+1}, \end{aligned}$$

where we have used the independence from time direction (Lemma 4.8) and K^b_{t+1} is any solution of the normal equations $K^b_{t+1}\Pi_{t+1} = \mathrm{Cov}(x_t, x_{t+1})$. From (4.73), it is obtained that $\mathrm{Cov}(x_t, x_{t+1}) = \Pi_t F'_t$, and the formula $F^b_{t+1}\Pi_{t+1} = \Pi_t F'_t$ follows. Also from (4.73), the relation $\Pi_{t+1} = F_t\Pi_t F'_t + G_t Q_t G'_t$ holds.

Defining $F^b_{t+1} = K^b_{t+1}, G^b_{t+1} = I$ and $u^b_{t+1} = e^b_t$, we can write

$$x_t = F^b_{t+1}x_{t+1} + u^b_{t+1}, \qquad t = n, n-1, \ldots, 1.$$

Finally, using $\mathrm{Cov}(x_t, x_{t+1}) = F^b_{t+1}\Pi_{t+1}$, it is obtained that

$$\begin{aligned}
Q^b_{t+1} &= \mathrm{Var}(e^b_t)\\
&= \mathrm{Var}(x_t - F^b_{t+1}x_{t+1})\\
&= \Pi_t - \mathrm{Cov}(x_t, x_{t+1})F^{b\prime}_{t+1} - F^b_{t+1}\mathrm{Cov}(x_{t+1}, x_t) + F^b_{t+1}\Pi_{t+1}F^{b\prime}_{t+1}\\
&= \Pi_t - F^b_{t+1}\Pi_{t+1}F^{b\prime}_{t+1}.
\end{aligned}$$

□

It is to be noticed that the backwards Markovian model of the previous lemma can be made more explicit if the matrices $\{\Pi_t : t = 1, 2, \ldots, n\}$ are nonsingular. In this case, the formulae

$$F^b_{t+1} = \Pi_t F'_t \Pi^{-1}_{t+1}, \qquad Q^b_{t+1} = \Pi_t - \Pi_t F'_t \Pi^{-1}_{t+1} F_t \Pi_t$$

hold. Sufficient conditions for $\{\Pi_t : t = 1, 2, \ldots, n\}$ to be nonsingular are

1. $\Pi_1 > 0$ and $\{F_t\}$ nonsingular, $t = 1, \ldots, n-1$.
2. $\Pi_1 > 0$ and $(F_t, G_t Q_t^{1/2})$ controllable, $t = 1, \ldots, n-1$.

Lemma 4.9 provides the backwards model corresponding to the forwards model (4.71). To obtain the backwards version of (4.72), we need two more lemmas.

Lemma 4.10 *The process* $\left\{\left[x'_t, Y'_t\right]'\right\}$ *is wide sense Markov, where* x_t *and* Y_t *are the state and the observation vector, respectively, of the state space model (4.71) and (4.72).*

Proof Using (4.71) and (4.72), we can write

$$\begin{bmatrix} x_{t+1} \\ Y_{t+1} \end{bmatrix} = \begin{bmatrix} F_t & 0 \\ H_{t+1}F_t & 0 \end{bmatrix}\begin{bmatrix} x_t \\ Y_t \end{bmatrix} + \begin{bmatrix} G_t & 0 \\ H_{t+1}G_t & I \end{bmatrix}\begin{bmatrix} u_t \\ v_{t+1} \end{bmatrix}.$$

Projecting onto $\left\{\left[x_t', Y_t'\right]', \ldots, \left[x_1', Y_1'\right]'\right\}$ and using the properties of orthogonal projection, it is obtained that

$$\begin{aligned}
E^*\left\{\begin{bmatrix} x_{t+1} \\ Y_{t+1} \end{bmatrix} \middle| \begin{bmatrix} x_t \\ Y_t \end{bmatrix}, \ldots, \begin{bmatrix} x_1 \\ Y_1 \end{bmatrix}\right\} &= \begin{bmatrix} F_t & 0 \\ H_{t+1}F_t & 0 \end{bmatrix}\begin{bmatrix} x_t \\ Y_t \end{bmatrix} + \begin{bmatrix} G_t & 0 \\ H_{t+1}G_t & I \end{bmatrix} \\
&\quad \times E^*\left\{\begin{bmatrix} u_t \\ v_{t+1} \end{bmatrix} \middle| \begin{bmatrix} x_t \\ Y_t \end{bmatrix}, \ldots, \begin{bmatrix} x_1 \\ Y_1 \end{bmatrix}\right\} \\
&= \begin{bmatrix} F_t & 0 \\ H_{t+1}F_t & 0 \end{bmatrix}\begin{bmatrix} x_t \\ Y_t \end{bmatrix} + \begin{bmatrix} G_t & 0 \\ H_{t+1}G_t & I \end{bmatrix} \\
&\quad \times E^*\left\{\begin{bmatrix} u_t \\ v_{t+1} \end{bmatrix} \middle| \begin{bmatrix} x_t \\ Y_t \end{bmatrix}\right\} \\
&= E^*\left\{\begin{bmatrix} x_{t+1} \\ Y_{t+1} \end{bmatrix} \middle| \begin{bmatrix} x_t \\ Y_t \end{bmatrix}\right\},
\end{aligned}$$

where we have used the fact that $\text{Cov}(u_t, y_t) = S_t$, $\text{Cov}(u_t, y_s) = 0, s < t$, and $\text{Cov}(v_{t+1}, y_s) = 0$, $\text{Cov}(v_{t+1}, x_s) = 0, s \leq t$.

Finally, letting $Z_j = [x_j', Y_j']'$ and using the law of iterated orthogonal projection, if $i > j > k$, then

$$\begin{aligned}
E^*(Z_i|Z_j, Z_k) &= E^*\left[E^*(Z_i|Z_{i-1}, Z_{i-2}, \ldots, Z_1)\,\middle|Z_j, Z_k\right] \\
&= E^*\left[E^*(Z_i|Z_{i-1})\,\middle|Z_j, Z_k\right] \\
&= \Pi_{i-1}E^*(Z_{i-1}|Z_j, Z_k) \\
&= \Pi_{i-1}\cdots\Pi_j E^*(Z_j|Z_j, Z_k) \\
&= \Pi_{i-1}\cdots\Pi_j Z_j \\
&= E^*(Z_i|Z_j),
\end{aligned}$$

where $E^*(Z_r|Z_{r-1}) = \Pi_{r-1}Z_{r-1}, r = i, i-1, \ldots, j+1$. Thus, $\left\{\left[x_t', Y_t'\right]'\right\}$ is wide sense Markov by Lemma 4.6. □

Lemma 4.11 *Let $\left\{\left[x_t', Y_t'\right]'\right\}$ be the process such that x_t and Y_t are the state and the observation vector, respectively, of the state space model (4.71) and (4.72). Then,*

$$E^*\left\{\begin{bmatrix} x_t \\ Y_t \end{bmatrix} \middle| \begin{bmatrix} x_{t+1} \\ Y_{t+1} \end{bmatrix}\right\} = \begin{bmatrix} F_{t+1}^b \\ H_{t+1}^b \end{bmatrix} x_{t+1},$$

where F_{t+1}^b is given by Lemma 4.9 and H_{t+1}^b is any solution of the normal equations $H_{t+1}^b Var(x_{t+1}) = Cov(Y_t, x_{t+1})$.

Proof Given that $Y_{t+1} = H_{t+1}x_{t+1} + v_{t+1}$ and considering that v_{t+1} is orthogonal to both x_t and Y_t, we can write

$$E^*\left\{\begin{bmatrix} x_t \\ Y_t \end{bmatrix} \middle| \begin{bmatrix} x_{t+1} \\ Y_{t+1} \end{bmatrix}\right\} = E^*\left\{\begin{bmatrix} x_t \\ Y_t \end{bmatrix} \middle| \begin{bmatrix} x_{t+1} \\ H_{t+1}x_{t+1} \end{bmatrix}\right\}$$
$$= E^*\left\{\begin{bmatrix} x_t \\ Y_t \end{bmatrix} |x_{t+1}\right\},$$

where the last equality follows from Proposition 1.7 if we take a nonsingular matrix M such that

$$M\begin{bmatrix} x_{t+1} \\ H_{t+1}x_{t+1} \end{bmatrix} = \begin{bmatrix} x_{t+1} \\ 0 \end{bmatrix}.$$

This can always be achieved by annihilating the rows corresponding to $H_{t+1}x_{t+1}$ in the matrix

$$\begin{bmatrix} x_{t+1} \\ H_{t+1}x_{t+1} \end{bmatrix}$$

using elementary row operations. The matrix M is then the matrix obtained from the unit matrix by performing the same row operations. □

We are now in a position to obtain the backwards state space model corresponding to the forwards model (4.71) and (4.72).

Theorem 4.31 *Given the forwards state space model (4.71) and (4.72), there exists a backwards model*

$$x_t = F^b_{t+1}x_{t+1} + u^b_{t+1} \tag{4.74}$$
$$Y_t = H^b_{t+1}x_{t+1} + v^b_{t+1}, \qquad t = n, n-1, \ldots, 1, \tag{4.75}$$

where

$$E\left\{\begin{bmatrix} u^b_{t+1} \\ v^b_{t+1} \end{bmatrix}\left[u^{b\prime}_{s+1}, v^{b\prime}_{s+1}\right]\right\} = \begin{bmatrix} Q^b_{t+1} & S^b_{t+1} \\ S^{b\prime}_{t+1} & R^b_{t+1} \end{bmatrix}\delta_{ts},$$

$E(u^b_{t+1}) = 0, E(v^b_{t+1}) = 0$*, the initial state vector,* x_{n+1}*, is orthogonal to* u^b_{t+1} *and* v^b_{t+1} *for all* $t, E(x_{n+1}) = 0$ *and, letting* $Var(x_t) = \Pi_t, t = n+1, n, \ldots, 1, F^b_{t+1}$ *and* H^b_{t+1} *are any solutions of the equations*

$$F^b_{t+1}\Pi_{t+1} = \Pi_t F'_t, \qquad H^b_{t+1}\Pi_{t+1} = N'_t,$$

and Q^b_{t+1}, S^b_{t+1} *and* R^b_{t+1} *are given by*

$$Q^b_{t+1} = \Pi_t - F^b_{t+1}\Pi_{t+1}F^{b\prime}_{t+1}, \qquad S^b_{t+1} = \Pi_t H'_t - F^b_{t+1}N_t,$$
$$R^b_{t+1} = D_t - H^b_{t+1}\Pi_{t+1}H^{b\prime}_{t+1},$$

where $N_t = F_t\Pi_t H'_t + G_t S_t$ *and* $D_t = H_t\Pi_t H'_t + R_t$*. In addition, the following alternative expressions hold*

$$S^b_{t+1} = Q^b_{t+1}H'_t - F^b_{t+1}G_t S_t \tag{4.76}$$
$$R^b_{t+1} = R_t + H_t S^b_{t+1} - S'_t G'_t H^{b\prime}_{t+1}. \tag{4.77}$$

Proof By Lemma 4.10, the process $\left\{[x'_t, Y'_t]'\right\}$ is wide sense Markov. Then, by Theorem 4.30, this process can be represented as a backwards Markovian model

$$\begin{bmatrix} x_t \\ Y_t \end{bmatrix} = \begin{bmatrix} F^{11}_{t+1} & F^{12}_{t+1} \\ F^{21}_{t+1} & F^{22}_{t+1} \end{bmatrix} \begin{bmatrix} x_{t+1} \\ Y_{t+1} \end{bmatrix} + \begin{bmatrix} G^{11}_{t+1} & G^{12}_{t+1} \\ G^{21}_{t+1} & G^{22}_{t+1} \end{bmatrix} \begin{bmatrix} u^b_{t+1} \\ v^b_{t+1} \end{bmatrix},$$

where, by the wide sense Markov property and Lemma 4.11,

$$\begin{aligned} E^*\left\{\begin{bmatrix} x_t \\ Y_t \end{bmatrix} \middle| \begin{bmatrix} x_{t+1} \\ Y_{t+1} \end{bmatrix}, \dots, \begin{bmatrix} x_{n+1} \\ Y_{n+1} \end{bmatrix}\right\} &= E^*\left\{\begin{bmatrix} x_t \\ Y_t \end{bmatrix} \middle| \begin{bmatrix} x_{t+1} \\ Y_{t+1} \end{bmatrix}\right\} \\ &= \begin{bmatrix} F^{11}_{t+1} & F^{12}_{t+1} \\ F^{21}_{t+1} & F^{22}_{t+1} \end{bmatrix} \begin{bmatrix} x_{t+1} \\ Y_{t+1} \end{bmatrix} \\ &= \begin{bmatrix} F^b_{t+1} & 0 \\ H^b_{t+1} & 0 \end{bmatrix} \begin{bmatrix} x_{t+1} \\ Y_{t+1} \end{bmatrix}, \end{aligned}$$

with F^b_{t+1} and H^b_{t+1} being the solutions of the normal equations

$$F^b_{t+1}\mathrm{Var}(x_{t+1}) = \mathrm{Cov}(x_t, x_{t+1}), \qquad H^b_{t+1}\mathrm{Var}(x_{t+1}) = \mathrm{Cov}(Y_t, x_{t+1}),$$

that give the formulae of the theorem. Letting

$$\begin{bmatrix} G^{11}_{t+1} & G^{12}_{t+1} \\ G^{21}_{t+1} & G^{22}_{t+1} \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix},$$

we see that

$$\begin{bmatrix} u^b_{t+1} \\ v^b_{t+1} \end{bmatrix} = \begin{bmatrix} x_t \\ Y_t \end{bmatrix} - E^*\left\{\begin{bmatrix} x_t \\ Y_t \end{bmatrix} \middle| \begin{bmatrix} x_{t+1} \\ Y_{t+1} \end{bmatrix}, \dots, \begin{bmatrix} x_{n+1} \\ Y_{n+1} \end{bmatrix}\right\},$$

and $\left\{[u^{b'}_{t+1}, v^{b'}_{t+1}]'\right\}$ is an innovation sequence with the required orthogonality properties. By Lemma 4.9, Q^b_{t+1} is given by the formula of the theorem. On the other hand, $v^b_{t+1} = Y_t - H^b_{t+1}x_{t+1}$ and, therefore,

$$\begin{aligned}\mathrm{Var}(v^b_{t+1}) &= \mathrm{Var}(Y_t) - H^b_{t+1}\Pi_{t+1}H^{b'}_{t+1}\\ &= H_t\Pi_t H'_t + R_t - H^b_{t+1}\Pi_{t+1}H^{b'}_{t+1}.\end{aligned}$$

Finally,

$$\begin{aligned}\mathrm{Cov}(u^b_{t+1}, v^b_{t+1}) &= \mathrm{Cov}(x_t - F^b_{t+1}x_{t+1}, Y_t - H^b_{t+1}x_{t+1})\\ &= \mathrm{Cov}(x_t, Y_t) - \mathrm{Cov}(x_t, x_{t+1})H^{b'}_{t+1}\\ &\quad -F^b_{t+1}\mathrm{Cov}(x_{t+1}, Y_t) + F^b_{t+1}\Pi_{t+1}F^{b'}_{t+1}\\ &= \mathrm{Cov}(x_t, H_t x_t + v_t) - \Pi_t F'_t H^{b'}_{t+1} - F^b_{t+1}N_t + F^b_{t+1}\Pi_{t+1}H^{b'}_{t+1}\\ &= \Pi_t H'_t - F^b_{t+1}N_t + \left[-\Pi_t F'_t + F^b_{t+1}\Pi_{t+1}\right]H^{b'}_{t+1}\\ &= \Pi_t H'_t - F^b_{t+1}N_t.\end{aligned}$$

To prove (4.76), substitute $N_t = F_t\Pi_t H'_t + G_t S_t$ and use $F_t\Pi_t = \Pi_{t+1}F^{b'}_{t+1}$ in the formula for S^b_{t+1} to give

$$\begin{aligned}S^b_{t+1} &= \Pi_t H'_t - F^b_{t+1}\left(\Pi_{t+1}F^{b'}_{t+1}H'_t + G_t S_t\right)\\ &= Q^b_{t+1}H'_t - F^b_{t+1}G_t S_t.\end{aligned}$$

To prove (4.76), use first $H^b_{t+1}\Pi_{t+1} = N'_t$, $F_t\Pi_t = \Pi_{t+1}F^{b'}_{t+1}$ and $N_t = F_t\Pi_t H'_t + G_t S_t$ to get

$$\begin{aligned}H^b_{t+1}\Pi_{t+1}H^{b'}_{t+1} &= \left(F_t\Pi_t H'_t + G_t S_t\right)' H^{b'}_{t+1}\\ &= H_t F^b_{t+1}\Pi_{t+1}H^{b'}_{t+1} + S'_t G'_t H^{b'}_{t+1}\\ &= H_t F^b_{t+1}\left(F_t\Pi_t H'_t + G_t S_t\right) + S'_t G'_t H^{b'}_{t+1}\\ &= H_t F^b_{t+1}\Pi_{t+1}F^{b'}_{t+1}H'_t + H_t F^b_{t+1}G_t S_t + S'_t G'_t H^{b'}_{t+1}.\end{aligned}$$

Then, substituting the previous expression in the formula for R^b_{t+1}, it is obtained that

$$\begin{aligned}R^b_{t+1} &= R_t + H_t\left(\Pi_t - F^b_{t+1}\Pi_{t+1}F^b_{t+1}\right)H'_t - H_t F^b_{t+1}G_t S_t - S'_t G'_t H^{b'}_{t+1}\\ &= R_t + H_t Q^b_{t+1}H'_t - H_t F^b_{t+1}G_t S_t - S'_t G'_t H^{b'}_{t+1}.\end{aligned}$$

□

4.12.4 Backwards State Space Model When the Π_t are Nonsingular

If the Π_t matrices are nonsingular, the backwards model (4.74) and (4.75) can be simplified. Since

$$F^b_{t+1} = \Pi_t F'_t \Pi^{-1}_{t+1}, \qquad Q^b_{t+1} = \Pi_t - \Pi_t F'_t \Pi^{-1}_{t+1} F_t \Pi_t,$$

we can define

$$\bar{x}_t = \Pi^{-1}_t x_t, \qquad \bar{u}^b_{t+1} = \Pi^{-1}_t u^b_{t+1}$$

to get the backwards state space model

$$\bar{x}_t = F'_t \bar{x}_{t+1} + \bar{u}^b_{t+1} \tag{4.78}$$

$$Y_t = N'_t \bar{x}_{t+1} + v^b_{t+1}, \qquad t = n, n-1, \ldots, 1, \tag{4.79}$$

where

$$E\left\{ \begin{bmatrix} \bar{u}^b_{t+1} \\ v^b_{t+1} \end{bmatrix} \begin{bmatrix} \bar{u}^{b\prime}_{s+1}, v^{b\prime}_{s+1} \end{bmatrix} \right\} = \begin{bmatrix} \overline{Q}^b_{t+1} & \overline{S}^b_{t+1} \\ \overline{S}^{b\prime}_{t+1} & R^b_{t+1} \end{bmatrix} \delta_{ts},$$

$E(\bar{u}^b_{t+1}) = 0, E(v^b_{t+1}) = 0, N_t = F_t \Pi_t H'_t + G_t S_t, \overline{Q}^b_{t+1} = \Pi^{-1}_t - F'_t \Pi^{-1}_{t+1} F_t, \overline{S}^b_{t+1} = H'_t - F'_t \Pi^{-1}_{t+1} N_t$, and $R^b_{t+1} = R_t + H_t \Pi_t H'_t - N'_t \Pi^{-1}_{t+1} N_t$. The initial state satisfies $E(\bar{x}_1) = 0$ and $\mathrm{Var}(\bar{x}_1) = \Pi^{-1}_1$. The matrices Π_t are recursively generated according to the formula $\Pi_{t+1} = F_t \Pi_t F'_t + G_t Q_t G'_t$ with $\Pi_1 > 0$.

Example 4.2 Suppose that the model (4.71) and (4.72) is time invariant and stationary (see Sect. 5.2), that is,

$$x_{t+1} = F x_t + G u_t$$
$$Y_t = H x_t + v_t, \qquad t = 1, \ldots, n,$$

where

$$E\left\{ \begin{bmatrix} u_t \\ v_t \end{bmatrix} \begin{bmatrix} u'_s, v'_s \end{bmatrix} \right\} = \begin{bmatrix} Q & S \\ S' & R \end{bmatrix} \delta_{ts},$$

$E(u_t) = 0, E(v_t) = 0$, the initial state vector, x_1, is orthogonal to u_t and v_t for all $t, E(x_1) = 0$ and $\mathrm{Var}(x_1) = \Pi > 0$ satisfies the Lyapunov equation $\Pi = F\Pi F' + GQG'$. Then, because the model is stationary (see again Sect. 5.2), $\mathrm{Var}(x_t) = \Pi$ for

$t = 1, 2, \ldots, n$, and it follows from this that the backwards model

$$\bar{x}_t = F'\bar{x}_{t+1} + \bar{u}^b_{t+1}$$
$$Y_t = N'\bar{x}_{t+1} + v^b_{t+1}, \qquad t = n, n-1, \ldots, 1,$$

where $\bar{x}_t = \Pi^{-1}x_t, N = F\Pi H' + GS$, $\mathrm{Var}(\bar{u}^b_{t+1}) = \Pi^{-1} - F'\Pi^{-1}F, E(\bar{u}^b_{t+1}v^{b'}_{t+1}) = H' - F'\Pi^{-1}N$ and $\mathrm{Var}(v^b_{t+1}) = R + H\Pi H' - N'\Pi^{-1}N$, is also time invariant and stationary. ◇

4.12.5 The Backwards Kalman Filter

Given the forwards state space model (4.71) and (4.72), we have seen in Theorem 4.31 that there exists a backwards state space model (4.74) and (4.75) with the properties specified in that theorem. It is straightforward to derive the Kalman filter corresponding to this backwards model. One only needs to let the time run backwards in the Kalman filter recursions. We give the result in the following theorem. The proof is left as an exercise. See Problem 4.4.

Theorem 4.32 *The backwards Kalman filter is given by the following recursions*

$$E^b_t = Y_t - H^b_{t+1}\hat{x}^b_{t+1|t+1}, \quad \Sigma^b_t = H^b_{t+1}P^b_{t+1|t+1}H^{b'}_{t+1} + R^b_{t+1}$$
$$K^b_t = (F^b_{t+1}P^b_{t+1|t+1}H^{b'}_{t+1} + S^b_{t+1})\Sigma^{-b}_t, \quad \hat{x}^b_{t|t} = F^b_{t+1}\hat{x}^b_{t+1|t+1} + K^b_t E^b_t$$
$$P^b_{t|t} = F^b_{t+1}P^b_{t+1|t+1}F^{b'}_{t+1} + Q^b_{t+1} - K^b_t\Sigma^b_t K^{b'}_t$$
$$= (F^b_{t+1} - K^b_t H^b_{t+1})P^b_{t+1|t+1}F^{b'}_{t+1} + Q^b_{t+1} - K^b_t S^{b'}_{t+1},$$

initialized with $\hat{x}^b_{n+1|n+1} = 0$ *and* $P^b_{n+1|n+1} = \Pi_{n+1}$, *where* $\hat{x}^b_{t|t} = E^*(x_t|Y_{t:n})$ *and* $P^b_{t|t}$ *is its MSE.*

It is to be noted that the backwards Kalman recursions are in fact recursions for the filtered estimator of the state.

It is also easy to derive measurement and time updates for the backwards model. We leave the development of the corresponding equations as an exercise for the reader.

4.13 Application of Backwards State Space Models to Smoothing

In Sect. 4.10, we have used the forwards state space model (4.71) and (4.72) to derive the formulae to compute the estimator $\hat{x}_{t|n}$ based on the sample $\{Y_1, \ldots, Y_n\}$ and its MSE. However, for fixed-interval smoothing problems the direction of time is not important, and it should be possible to also obtain the estimator $\hat{x}_{t|n}$ processing the data backwards starting with Y_n and ending with Y_1. In this section, we will present some results in this connection.

4.13.1 Two-Filter Formulae

Theorem 4.33 (The General Two-Filter Smoothing Formulae) *Given the state space model (4.71) and (4.72), where the Π_t are nonsingular, we can write*

$$\hat{x}_{t|n} = P_{t|n}\left(P_t^{-1}\hat{x}_{t|t-1} + P_{t|t}^{-b}\hat{x}_{t|t}^{b}\right) \tag{4.80}$$

$$P_{t|n} = \left(P_t^{-1} + P_{t|t}^{-b} + \Pi_t^{-1}\right)^{-1}, \tag{4.81}$$

where $P_{t|t}^{-b}$ is the inverse of $P_{t|t}^{b}$ and $\hat{x}_{t|t}^{b}$ and $P_{t|t}^{b}$ are given by the backwards Kalman filter.

Proof We will prove the theorem using the combined linear estimators formula given by Theorem 2.2. To this end, we use the forwards model (4.71) and (4.72) to generate $Y_b = (Y_t', Y_{t+1}', \ldots, Y_n')'$ and the backwards model (4.74) and (4.75) to generate $Y_a = (Y_1', \ldots, Y_{t-2}', Y_{t-1}')'$, both models using x_t as initial condition. Then, we can write

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_{t-2} \\ Y_{t-1} \end{bmatrix} = \begin{bmatrix} H_2^b F^b(2,t) \\ \vdots \\ H_{t-1}^b F^b(t-1,t) \\ H_t^b \end{bmatrix} x_t + \begin{bmatrix} \sum_{i=2}^{t-1} H_2^b F^b(1,i) u_i^b + v_2^b \\ \vdots \\ H_{t-1}^b u_t^b + v_{t+1}^b \\ v_t^b \end{bmatrix}$$

and

$$\begin{bmatrix} Y_t \\ Y_{t+1} \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} H_t \\ H_{t+1}F(t+1,t) \\ \vdots \\ H_n F(n,t) \end{bmatrix} x_t + \begin{bmatrix} v_t \\ H_{t+1}G_t u_t + v_{t+1} \\ \vdots \\ \sum_{i=t}^{n-1} H_n F(n,i+1) G_i u_i + v_n \end{bmatrix},$$

where $F^b(i,j) = F_{i+1}^b \cdots F_j^b$ if $i < j$ and $F^b(i,i) = I$, and $F(i,j) = F_{i-1}F_{i-2}\cdots F_j$ if $i > j$ and $F(i,i) = I$. Using an obvious notation, we can write the previous expression in a more concise way as

$$y_a = H_a x_t + v_a, \qquad y_b = H_b x_t + v_b.$$

By the properties of the forwards and the backwards models, it is easy to verify that $\{v_a, v_b, x_t\}$ are mutually uncorrelated random variables with zero mean. The result now follows if we identify $\hat{x}_a = \hat{x}_{t|t-1}, \hat{x}_b = \hat{x}^b_{t|t}$ and apply Theorem 2.2. □

4.13.2 Backwards Model When $\Pi^{-1}_{n+1} = 0$ and the F_t are Nonsingular

In the special case in which $\Pi^{-1}_{n+1} = 0$ and the F_t are nonsingular, it follows from $\Pi_{t+1} = F_t\Pi_t F'_t + G_t Q_t G'_t$ that

$$\Pi_t = F_t^{-1}\Pi_{t+1}F_t^{-1'} - F_t^{-1}G_tQ_tG'_tF_t^{-1'}.$$

Letting $\Pi_{n+1} = kI$ and $k \to \infty$ and proceeding recursively, it follows from the previous equality and the fact that the matrices F_t^{-1} are nonsingular that $\Pi_t^{-1} = 0$ for $1 \leq t \leq n+1$. Based on this, we get

$$\begin{aligned} F^b_{t+1} &= \Pi_t F'_t \Pi^{-1}_{t+1} \\ &= \left(F_t^{-1}\Pi_{t+1} - F_t^{-1}G_tQ_tG'_t\right)\Pi^{-1}_{t+1} \\ &= F_t^{-1} - F_t^{-1}G_tQ_tG'_t\Pi^{-1}_{t+1} \\ &= F_t^{-1}. \end{aligned}$$

Using the equality $F_t\Pi_t = \Pi_{t+1}F_t^{-1'} - G_tQ_tG'_tF_t^{-1'}$, it is obtained that

$$\begin{aligned} H^b_{t+1} &= N'_t\Pi^{-1}_{t+1} \\ &= \left(H_tF_t^{-1}\Pi_{t+1} - H_tF_t^{-1}G_tQ_tG'_t + S'_tG'_t\right)\Pi^{-1}_{t+1} \\ &= H_tF_t^{-1}. \end{aligned}$$

Thus, the transition equation is in this case

$$x_t = F_t^{-1}x_{t+1} + u^b_{t+1}.$$

From this, we get

$$\begin{aligned} x_{t+1} &= F_tx_t - F_tu^b_{t+1} \\ &= F_tx_t + G_tu_t \end{aligned}$$

and, therefore,

$$u^b_{t+1} = -F_t^{-1}G_tu_t.$$

In a similar way, it is obtained that

$$\begin{aligned} Y_t &= H_t F_t^{-1}\left(F_t x_t - F_t u_{t+1}^b\right) + v_{t+1}^b \\ &= H_t x_t + v_{t+1}^b - H_t u_{t+1}^b \\ &= H_t x_t + v_t \end{aligned}$$

and

$$v_{t+1}^b = v_t - H_t F_t^{-1} G_t u_t.$$

It easily follows from this that

$$\begin{aligned} Q_{t+1}^b &= F_t^{-1} G_t Q_t G_t' F_t^{-1'}, \\ S_{t+1}^b &= F_t^{-1}\left(G_t Q_t G_t' F_t^{-1'} H_t' - G_t S_t\right), \end{aligned}$$

and

$$R_{t+1}^b = R_t + H_t F_t^{-1} G_t Q_t G_t' F_t^{-1'} H_t' - H_t F_t^{-1} G_t S_t - S_t' G_t' F_t^{-1'} H_t'.$$

However, instead of the backwards model (4.74) and (4.75), it is easier in this case to use the backwards model

$$\begin{aligned} x_t &= F_t^{-1} x_{t+1} + u_{t+1}^b \\ Y_t &= H_t x_t + v_t, \qquad t = n, n-1, \ldots, 1, \end{aligned}$$

where

$$E\left\{\begin{bmatrix} u_{t+1}^b \\ v_t \end{bmatrix}\begin{bmatrix} u_{s+1}^{b'}, v_s \end{bmatrix}\right\} = \begin{bmatrix} F_t^{-1} G_t Q_t G_t' F_t^{-1'} & -F_t^{-1} G_t S_t \\ -S_t' G_t' F_t^{-1'} & R_t \end{bmatrix}\delta_{ts}.$$

Note that the assumption $\Pi_{n+1}^{-1} = 0$ is crucial for the previous backwards state space model to be valid because it ensures that x_{n+1} is orthogonal to u_{t+1}^b and v_{t+1}^b for all t. Otherwise, it would suffice to assume only that F_t is nonsingular to obtain the same state space model.

The assumption that $\Pi_{n+1}^{-1} = 0$ and the F_t are nonsingular simplifies the two-filter formulae of Theorem 4.33 because $\Pi_t^{-1} = 0$ for all t and the backwards Kalman filter recursions are simpler.

4.14 The State Space Model With Constant Bias and Incompletely Specified Initial Conditions

Given a time series $Y = (Y_1', \dots, Y_n')'$ with $Y_t \in \mathbb{R}^p$, the state space model with constant bias and incompletely specified initial conditions is defined by

$$x_{t+1} = W_t\beta + F_t x_t + G_t u_t, \tag{4.82}$$

$$Y_t = V_t\beta + H_t x_t + v_t, \qquad t = 1, \dots, n, \tag{4.83}$$

where W_t, F_t, G_t, V_t, and H_t are time-varying deterministic matrices, $\beta \in \mathbb{R}^k$ is the constant bias vector, $x_t \in \mathbb{R}^r$ is the state vector,

$$E\left\{\begin{bmatrix} u_t \\ v_t \end{bmatrix}\begin{bmatrix} u_s', v_s' \end{bmatrix}\right\} = \sigma^2 \begin{bmatrix} Q_t & S_t \\ S_t' & R_t \end{bmatrix}\delta_{ts},$$

$u_t \in \mathbb{R}^s$, $v_t \in \mathbb{R}^p$, $E(u_t) = 0$ and $E(v_t) = 0$. The initial state vector x_1 is specified as

$$x_1 = W\beta + A\delta + x, \tag{4.84}$$

where $x \sim (a, \sigma^2\Omega)$, the matrices W, A, and Ω are fixed and known, and $\delta \in \mathbb{R}^d$, $\delta \sim (b, \sigma^2\Pi)$ is a random vector that models the unknown initial conditions. Here, the notation $v \sim (m, \Sigma)$ means that the vector v has mean m and covariance matrix Σ.

It is assumed that the vectors x and δ are mutually orthogonal and that x_1 is orthogonal to the $\{u_t\}$ and $\{v_t\}$ sequences. As in Sect. 4.1, we will usually assume $\sigma^2 = 1$ unless otherwise specified. Finally, we will further assume that if δ and β are zero in the state space model (4.82) and (4.83), then the generated data have a covariance matrix that is nonsingular. That is, if the model reduces to (4.1) and (4.2) with $x_1 = x$ and if $Y = (Y_1', \dots, Y_n')'$, then $\mathrm{Var}(Y)$ is nonsingular.

Equations (4.82) and (4.83) are called the "transition equation" and the "measurement equation," respectively.

Instead of the model (4.82) and (4.83), the following alternative single disturbance state space model can be used

$$x_{t+1} = W_t\beta + F_t x_t + G_t\epsilon_t, \tag{4.85}$$

$$Y_t = V_t\beta + H_t x_t + J_t\epsilon_t, \qquad t = 1, \dots, n, \tag{4.86}$$

where $\epsilon_t \sim (0, \sigma^2 I)$ and the $\{\epsilon_t\}$ sequence is serially uncorrelated and uncorrelated with x_1.

There are several approaches to handle the problem of unspecified initial conditions. The most popular one is the **Bayesian approach**, in which δ is assumed to be **"diffuse"** with $\Pi^{-1} = 0$. The second approach, that we will call the **"transformation approach,"** consists of making a transformation of the data so that the transformed data does not depend on δ. The third approach, that we will call the **"conditional likelihood approach,"** consists of considering δ fixed and conditioning on the maximum likelihood estimator of δ. Finally, in the fourth approach, called the **"profile likelihood approach,"** δ is considered fixed instead of a random variable. These four approaches were described in detail in Sect. 2.3 for linear models. In the next theorem, we will see that the data generated by the state space model (4.82) and (4.83) are in fact also generated by a linear model. Thus, the previous approaches apply to (4.82) and (4.83).

Define the vector $\gamma = (\delta', \beta')'$, the stack of the vectors that model the diffuse and bias effects. Then, it can be shown by iterating in the state space model equations (4.82) and (4.83) that the stack of the observations, Y, depends linearly on γ. That is,

$$Y = X\gamma + \varepsilon, \tag{4.87}$$

where the matrix X depends on the system matrices and the error term ε is a linear combination of the u_t, v_t, and x with coefficients that depend in turn on the system matrices. The following theorem gives the details. The simple proof is omitted.

Theorem 4.34 *If $Y = (Y_1', \ldots, Y_n')'$ is generated by the state space model (4.82) and (4.83), where we assume without loss of generality that $x \sim (0, \sigma^2\Omega)$, then (4.87) holds, where $X = (R, S)$ and the rows, S_t, R_t and ϵ_t, of S, R and ϵ can be obtained from the recursions*

$$(R_t, S_t) = (0, V_t) + H_t A_t$$
$$A_{t+1} = (0, W_t) + F_t A_t,$$

with the initial condition $A_1 = (A, W)$, and

$$\epsilon_t = v_t + H_t \eta_t$$
$$\eta_{t+1} = F_t \eta_t + G_t u_t,$$

with the initial condition $\eta_1 = x$. In addition, $\varepsilon \sim (0, \sigma^2\Sigma)$ with Σ nonsingular and $Cov(\delta, \varepsilon) = 0$.

4.14.1 Examples

Example 4.3 Consider the signal plus noise model $Y_t = S_t + N_t$, where the signal, S_t, follows the model $\nabla S_t = b_t$, $\nabla = 1 - B$, and B is the backshift operator $B^j Y_t = Y_{t-j}$. The processes $\{b_t\}$ and $\{N_t\}$ are mutually and serially uncorrelated with zero mean, $\text{Var}(b_t) = 1$ and $\text{Var}(N_t) = 2$.

The model can be easily put into state space form (4.85) and (4.86) by defining $x_t = S_t, \epsilon_t = (b_{t+1}, N_t/\sqrt{2})', V_t = 0, H_t = 1, J_t = (0, \sqrt{2}), G_t = (1, 0), W_t = 0,$

and $F_t = 1$. The initial state vector is $x_1 = x + A\delta$, where $x = b_1, A = 1$, and $\delta = S_0$. ◇

Example 4.4 Suppose the signal plus noise model of Example 4.3, but with the signal following the model $\nabla^2 S_t = b_t$ and $\mathrm{Var}(N_t) = \lambda$. This is the model underlying the famous Hodrick–Prescott filter (Gómez, 2001) used by econometricians.

The model can be cast into state space form (4.85) and (4.86) as follows. Define $\epsilon_t = (b_{t+1}, N_t/\sqrt{\lambda})', x_t = (S_t, S_{t+1|t})', S_{t+1|t} = S_{t+1} - b_{t+1}, V_t = 0, W_t = 0,$

$$F_t = \begin{bmatrix} 0 & 1 \\ -1 & 2 \end{bmatrix}, \quad G_t = \begin{bmatrix} 1 & 0 \\ 2 & 0 \end{bmatrix},$$

$H_t = (1, 0)$ and $J_t = (0, \sqrt{\lambda})$. Then, the representation is correct because $S_{t+2|t+1} + S_t - 2S_{t+1|t} = S_{t+2} - 2S_{t+1} + S_t - b_{t+2} + 2b_{t+1} = 2b_{t+1}$. Note that the first column of G_t is formed with the first two coefficients obtained from the expansion of $1/(1-B)^2$ and that $S_{t+1|t}$ can be interpreted as the forecast of S_t based on $\{S_s : s = 1, \ldots, t\}$ and the starting values S_{-1} and S_0.

Since the process $\{S_t\}$ follows an ARIMA model, it can be generated as linear combinations of some starting values and elements of the differenced process, $b_t = \nabla^2 S_t$. Let the starting values be $\delta = (S_{-1}, S_0)'$. Then, following Bell (1984), the S_t can be generated from $S_t = D_t'\delta + \sum_{i=0}^{t-1} \xi_i b_{t-i}$, where $t > 0$, $1/(1-B)^2 = \sum_{i=0}^{\infty} \xi_i B^i$ and the $D_t = (D_{1t}, D_{2t})'$ can be recursively generated from

$$D_{-1} = (1, 0), \quad D_0 = (0, 1)$$
$$D_t = 2D_{t-1} - D_{t-2}, \quad t > 0.$$

The previous recursions imply $S_1 = (-1, 2)\delta + b_1$ and $S_{2|1} = (-2, 3)\delta + 2b_1$. Thus, the initial state is $x_1 = x + A\delta$, where $x = (b_1, 2b_1)'$ and $A = (A_{11}', A_{12}')', A_{11} = (-1, 2), A_{12} = (-2, 3)$. ◇

Example 4.5 Consider the model $(1 + \phi B)(\nabla Y_t - \mu) = A_t$, where B and ∇ are as in Example 1, $|\phi| < 1, \mu$ is a constant and $\{A_t\}$ is an uncorrelated sequence with zero mean and $\mathrm{Var}(A_t) = 1$. Define the variable $Z_t = t - 1, t = 1, \ldots, n$. Then, we can write $Y_t = Z_t\mu + U_t$, where U_t follows the model $(1 + \phi B)\nabla U_t = A_t$.

To put the model into state space form (4.85) and (4.86), we can use the representation of Example 4.4 with one component and no observational noise. Define $\epsilon_t = A_{t+1}, x_t = (U_t, U_{t+1|t})', U_{t+1|t} = U_{t+1} - A_{t+1}, V_t = Z_t, \beta = \mu, H_t = (1, 0), W_t = 0, J_t = 0, G_t = (1, 1 - \phi)'$, and

$$F_t = \begin{bmatrix} 0 & 1 \\ \phi & 1 - \phi \end{bmatrix}.$$

Let $u_t = U_t - U_{t-1}$ and $\delta = U_0$. Then, it is not difficult to verify that $U_t = U_0 + \sum_{s=1}^{t} u_s, U_1 = \delta + u_1$, and $U_{2|1} = \delta + (1 - \phi)u_1$. Thus, the initial state is $x_1 = x + A\delta$, where $x = (u_1, (1 - \phi)u_1)'$ and $A = (1, 1)'$. ◇

4.14.2 Initial Conditions in the Time Invariant Case

In the previous section, we have considered several examples of state space models in which the system matrices did not depend on time and the initial conditions were obtained in more or less ad-hoc manner. In this section, we will address the issue of constructing initial conditions for time invariant state space models in an automatic way. In these models, the distribution of the initial state vector often depends on the eigenvalues of the matrix F in the transition equation. More specifically, if the eigenvalues of F are inside the unit circle, the distribution of x_1 can usually be completely specified. But if F has some eigenvalues on the unit circle, this will not be the case and x_1 will depend on the unspecified part, δ. We will make the derivation for the state space model with a single disturbance (4.85) and (4.86). The case of the state space model (4.82) and (4.83) can be reduced to this easily.

We will assume first that there is no bias. Thus, suppose the single disturbance time invariant state space model

$$x_{t+1} = Fx_t + G\epsilon_t \tag{4.88}$$

$$Y_t = Hx_t + J\epsilon_t, \tag{4.89}$$

where $\text{Var}(\epsilon) = \sigma^2 I$. Then, the initial state vector can be specified as follows. According to the Schur decomposition (Golub & Van Loan, 1996), there exists an orthogonal matrix, P, such that $P'FP = U$, where U is an upper block triangular matrix. By reordering the eigenvalues if necessary, we can assume without loss of generality that

$$U = \begin{bmatrix} U_N & U_{12} \\ 0 & U_S \end{bmatrix}, \tag{4.90}$$

where the eigenvalues of U_N have unit modulus and the eigenvalues of U_S have modulus strictly less than one. If U_{12} is not zero, we can make an additional transformation in (4.88) and (4.89) to eliminate U_{12}. The details are given by the following lemma.

Lemma 4.12 *Let U be as in (4.90), where U_N, U_{12}, and U_S are arbitrary matrices. If the Lyapunov equation*

$$U_N X - X U_S = U_{12} \tag{4.91}$$

has a solution and we define $Q = \begin{bmatrix} I & X \\ 0 & I \end{bmatrix}$, *then* $QUQ^{-1} = \begin{bmatrix} U_N & 0 \\ 0 & U_S \end{bmatrix}$. *Moreover, Eq. (4.91) has a unique solution if, and only if, U_N and U_S have no common eigenvalues.*

Proof This is Lemma 7.1.4, pp. 314–315 of Golub & Van Loan (1996). □

If we define the transformation $[x'_{N,t}, x'_{S,t}]' = Rx_t$, where the partition is conformal to (4.90), $R = QP'$ and Q is the unique matrix given by Lemma 4.12, then

$$RFR^{-1} = \begin{bmatrix} U_N & 0 \\ 0 & U_S \end{bmatrix}, \quad RG = \begin{bmatrix} G_N \\ G_S \end{bmatrix}, \quad HR^{-1} = [H_N, H_S], \tag{4.92}$$

and all partitions are conformal to (4.90). We see that the state space model (4.88) and (4.89) is transformed into

$$\begin{aligned} x_{N,t+1} &= U_N x_{N,t} + G_N \epsilon_t, \qquad x_{S,t+1} = U_S x_{S,t} + G_S \epsilon_t \\ Y_t &= H_N x_{N,t} + H_S x_{S,t} + J\epsilon_t, \qquad t = 1, \ldots n, \end{aligned}$$

where $\{x_{N,t}\}$ is purely nonstationary and $\{x_{S,t}\}$ is purely stationary. This motivates that we define $x_{N,1} = \delta$, where δ is a diffuse vector, and $x_{S,1} = c_S$, where c_S is a zero-mean stationary random vector with $\mathrm{Var}(c_S) = V$ satisfying the discrete-time Lyapunov equation

$$V = U_S V U'_S + G_S G'_S. \tag{4.93}$$

Letting $R^{-1} = (R_N, R_S)$ and $P = (P_N, P_S)$, where the partitions are conformal to (4.90), we finally obtain the initial state in the form (4.84), $x_1 = A\delta + x$, where $A = R_N, x = R_S c_S$ and $(R_N, R_S) = P \begin{bmatrix} I & -X \\ 0 & I \end{bmatrix} = (P_N, P_S - P_N X)$.

If the state space model is (4.85) and (4.86), we can consider the same transformation, $[x'_{N,t}, x'_{S,t}]' = Rx_t$, and the transformed model is

$$\begin{aligned} x_{N,t+1} &= W_{N,t}\beta + U_N x_{N,t} + G_N \epsilon_t, \qquad x_{S,t+1} = W_{S,t}\beta + U_S x_{S,t} + G_S \epsilon_t \\ Y_t &= V_t\beta + H_N x_{N,t} + H_S x_{S,t} + J\epsilon_t, \qquad t = 1, \ldots n, \end{aligned}$$

where $RW_t = [W'_{N,t}, W'_{S,t}]'$ and the partition is conformal to (4.90). Clearly, $\{x_{N,t}\}$ is again purely nonstationary. As for $\{x_{S,t}\}$, we can write

$$\begin{aligned} x_{S,t} &= (I - U_S B)^{-1} W_{S,t-1}\beta + (I - U_S B)^{-1} G_S \epsilon_{t-1} \\ &= m_{S,t} + v_{S,t}, \end{aligned}$$

where B is the backshift operator, $BY_t = Y_{t-1}$, $v_{S,t}$ is a zero mean stationary random vector, $\mathrm{Var}(v_{S,t}) = V$ satisfies (4.93) and $m_{S,t}$ is the mean of $x_{S,t}$, usually unknown because in the majority of cases there is no information about the infinite past of $W_{S,t}$. However, it may happen that $\{W_{S,t}\}$ is a stationary process, in which case we can take $m_{S,t} = (I - U_S)^{-1}\mu_W\beta$, where μ_W is the mean of $\{W_{S,t}\}$.

If we have some guess about $m_{S,1}$ of the form $m_{S,1} = W_S\beta$, where W_S is a known fixed matrix, then we can set $x_1 = W\beta + A\delta + x$, where $W = R_S W_S, A = R_N, x = R_S c_S$, δ is diffuse and c_S is a zero mean stationary random vector such that $\mathrm{Var}(c_S) =$

V satisfies (4.93). If $m_{S,1}$ is unknown, it has to be estimated along with the other unknown parameters in the model.

Another possibility, less rigorous from a statistical point of view and similar to the procedure used in subspace methods, is to estimate first x_1 and β using the observation equation, $Y_t = V_t\beta + Hx_t + J\epsilon_t$, and then to estimate $m_{S,1}$ from $[\hat{x}'_{N,1}, \hat{m}'_{S,1}]' = R\hat{x}_1$, where $\hat{x}_1$ is the estimate of x_1. In this case, we set $x_1 = A\delta + x$, where δ is diffuse, $A = R_N, x = R_S(\hat{m}_{S,1} + c_S)$ and c_S is as before.

The vectors β and x_1 can be estimated by regression using the equation

$$Y_t = \left[H(z^{-1}I - F)_t^{-1}W_t + V_t\right]\beta + HF^{t-1}x_1 + v_t,$$

where z^{-1} is the forward operator, $z^{-1}Y_t = Y_{t+1}, (z^{-1}I - F)_t^{-1} = \sum_{j=1}^{t-1} z^j F^{j-1}$, and $v_t = \left[H(z^{-1}I - F)_t^{-1}G + J\right]\epsilon_t$.

4.14.3 The Diffuse Likelihood

The linear model corresponding to the state space model (4.82) and (4.83) is, according to Theorem 4.34,

$$Y = R\delta + \omega, \tag{4.94}$$

where, assuming without loss of generality $x \sim (0, \sigma^2\Omega)$ in (4.84), $\omega \sim (S\beta, \sigma^2\Sigma)$ and $\mathrm{Cov}(\delta, \omega) = 0$. Thus, by Theorem 2.4, the diffuse log-likelihood is, apart from a constant,

$$\lambda_D(Y) = -\frac{1}{2}\Big\{\ln|\sigma^2\Sigma| + \ln|R'(\sigma^2\Sigma)^{-1}R| + (Y - R\hat{\delta} - S\beta)' \\ \times\Sigma^{-1}(Y - R\hat{\delta} - S\beta)/\sigma^2\Big\},$$

where $\hat{\delta} = (R'V_\omega^{-1}R)^{-1}R'V_\omega^{-1}(Y - m_\omega), m_\omega = S\beta$ and $V_\omega = \sigma^2\Sigma$. It turns out that we can concentrate β and σ^2 out of the diffuse log-likelihood. The next proposition gives the details. The proof is left as an exercise. See Problem 4.5.

Proposition 4.1 *The parameters β and σ^2 can be concentrated out of the diffuse log-likelihood and the (β, σ^2)-maximized diffuse log-likelihood, denoted by $\lambda_D(Y; \hat{\beta}, \hat{\sigma}^2)$, is*

$$\lambda_D(Y; \hat{\beta}, \hat{\sigma}^2) = constant - \frac{1}{2}\Big\{(np - n_\delta)\ln\left[(Y - X\hat{\gamma})'\Sigma^{-1}(Y - X\hat{\gamma})\right] \\ + \ln|\Sigma| + \ln|R'\Sigma^{-1}R|\Big\}, \tag{4.95}$$

where n_δ is the dimension of δ, $X = [R, S]$, $\gamma = (\delta', \beta')'$ and $\hat{\gamma}$ and $\hat{\sigma}^2$ are the GLS estimators of γ and σ^2 in the linear model (4.87), assuming γ fixed and, without loss of generality, $x \sim (0, \sigma^2\Omega)$ in (4.84).

Remark 4.4 In the case of the state space model (4.82) and (4.83) the parameters in β are parameters of interest and are treated differently to the elements of δ, that are nuisance random variables. In addition, Assumption 2.1 in Chap. 2 is usually satisfied in the state space model (4.82) and (4.83). ◇

Remark 4.5 If in the linear model (4.87) we assume that γ is fixed and all of its elements are parameters of interest, it is not difficult to verify that the (β, σ^2)-maximized log-likelihood of this model is equal to constant $-\left\{(np)\ln\left[(Y - X\hat{\gamma})'\Sigma^{-1}(Y - X\hat{\gamma})\right] + \ln|\Sigma|\right\}/2$ and thus differs from the (β, σ^2)-maximized diffuse log-likelihood, given by the previous proposition, in the term $\ln|R'\Sigma^{-1}R|$ and in the number of degrees of freedom. The log-likelihood obtained in this way is called the (β, σ^2)-maximized profile log-likelihood and will be considered in the next section. As mentioned earlier, in the diffuse likelihood approach the parameters in β are parameters of interest and are treated differently to the elements of δ, that are nuisance random variables. ◇

4.14.4 The Profile Likelihood

Consider the linear model (4.94) corresponding to the state space model (4.82) and (4.83), where, without loss of generality, $x \sim (0, \sigma^2\Omega)$ in (4.84), $\omega \sim (S\beta, \sigma^2\Sigma)$ and δ is assumed fixed. Then, by Theorem 2.7, the profile log-likelihood is, apart from a constant,

$$\lambda_P(Y) = -\frac{1}{2}\left\{\ln|\sigma^2\Sigma| + (Y - R\hat{\delta} - S\beta)'\Sigma^{-1}(Y - R\hat{\delta} - S\beta)/\sigma^2\right\},$$

where $\hat{\delta} = (R'V_\omega^{-1}R)^{-1}R'V_\omega^{-1}(Y - m_\omega)$, $m_\omega = S\beta$ and $V_\omega = \sigma^2\Sigma$. As with the diffuse log-likelihood, we can concentrate β and σ^2 out of the profile log-likelihood. The next proposition gives the details. The simple proof is omitted.

Proposition 4.2 *The parameters β and σ^2 can be concentrated out of the profile log-likelihood and the (β, σ^2)-maximized profile log-likelihood, denoted by $\lambda_P(Y; \hat{\beta}, \hat{\sigma}^2)$, is*

$$\lambda_P(Y; \hat{\beta}, \hat{\sigma}^2) = constant - \frac{1}{2}\left\{(np)\ln\left[(Y - X\hat{\gamma})'\Sigma^{-1}(Y - X\hat{\gamma})\right] + \ln|\Sigma|\right\}, \tag{4.96}$$

where $X = [R, S], \gamma = (\delta', \beta')'$ *and* $\hat{\gamma}$ *and* $\hat{\sigma}^2$ *are the GLS estimators of* γ *and* σ^2 *in the linear model (4.87), assuming* γ *fixed and, without loss of generality,* $x \sim (0, \sigma^2\Omega)$ *in (4.84).*

Remark 4.6 The (β, σ^2)-maximized diffuse log-likelihood differs from the (β, σ^2)-maximized profile log-likelihood in the term $\ln|R'\Sigma^{-1}R|$ and in the number of degrees of freedom. The influence of the term $\ln|R'\Sigma^{-1}R|$ can be important in small samples and, therefore, the diffuse log-likelihood is preferable in this case. ◇

Remark 4.7 The profile log-likelihood usually arises when the whole initial state, x_1, is considered fixed. In this case, it is often assumed that $x_1 = \delta$ with $\mathrm{Var}(\delta) = 0$. This simplification can have important adverse effects in small samples. ◇

4.14.5 The Marginal and Conditional Likelihoods

Apart from the diffuse and the profile likelihoods, described in the previous two sections for the linear model (4.94) corresponding to the state space model (4.82) and (4.83), we can consider the marginal and the conditional likelihoods, introduced in Sect. 2.3. The precise relationship between the different likelihoods is given by Theorem 2.6.

An example of the use of the transformation approach and the marginal likelihood is provided by the likelihood function defined for ARIMA models in Box & Jenkins (1976). In this book, the likelihood function of an ARIMA model is defined as the likelihood of the differenced series. Thus, the transformation is given by the differencing operator and the marginal likelihood coincides with the likelihood of the differenced series.

The conditional likelihood was used in Gómez & Maravall (1994a) to define the likelihood of a nonstationary ARIMA model. These authors also proved that the conditional likelihood was in this case equal to the likelihood of the differenced series, that, as mentioned in the previous paragraph, is equal to the marginal likelihood.

In Gómez & Maravall (1994b), the previously mentioned different likelihoods defined for the linear model (4.94) corresponding to the state space model (4.82) and (4.83) are discussed and some examples are provided of their computation.

4.15 The Augmented-State Kalman Filter and the Two-Stage Kalman Filter

For algorithmic reasons, the approach used in this book to evaluate the log-likelihood and to compute the state estimators for prediction and smoothing is that of the diffuse likelihood. The reason for this is that the Kalman filter is well suited for a Bayesian approach, as the following example illustrates.

Example 4.6 Consider the regression model $Y_t = H_t\beta + v_t, t = 1, \ldots, n$, where $\mathrm{Var}(v_t) \sim N(0, 1)$ and the k-dimensional vector β is assumed to have the prior distribution $N(0, \Pi)$ with Π nonsingular. This model can be put into state space form (4.1) and (4.2) by defining $x_t = \beta, F_t = I_k, G_t = 0$ and $u_t = 0$. That is,

$$x_{t+1} = x_t$$
$$Y_t = H_t x_t + v_t.$$

The usual Kalman filter can be applied to obtain the state estimator, which in this case is the posterior mean of β. Note that if we assume β diffuse, as is usually the case when we want to obtain the classical OLS estimator of β, then, by Remarks 2.2 and 2.3, $\Pi^{-1} \to 0$ and the ordinary Kalman filter cannot be applied. In this case, we need the information form of the Kalman filter described in Sect. 4.6.

Since $G_t = 0$ and $F_t^s = 1$, by Theorem 4.13, we get the recursions

$$P_{t+1}^{-1}\hat{x}_{t+1|t} = P_t^{-1}\hat{x}_{t|t-1} + H_t' Y_t$$
$$P_{t+1}^{-1} = P_t^{-1} + H_t' H_t,$$

where $\hat{x}_{t+1|t} = \hat{\beta}_{t+1|t}$ is the OLS estimator of β based on $(Y_1', \ldots, Y_t')'$ and P_{t+1} is its MSE. Assuming $\Pi^{-1} \to 0$, we can initialize the recursions with $\hat{\beta}_{1|0} = 0$ and $P_1^{-1} = 0$ to get the ordinary OLS estimator of β and its MSE recursively. ◇

The **augmented-state Kalman filter** is a device that is usually applied to evaluate the log-likelihood and to compute the state estimators for prediction and smoothing when dealing with the state space model (4.82) and (4.83). It consists simply of first appending the vector $\gamma = (\delta', \beta')'$ to the state vector, x_t, to get the augmented state space model

$$x_{t+1}^a = F_t^a x_t^a + G_t^a u_t, \tag{4.97}$$
$$Y_t = H_t^a x_t^a + v_t, \qquad t = 1, \ldots, n, \tag{4.98}$$

where

$$F_t^a = \begin{bmatrix} F_t & (0, W_t) \\ 0 & I \end{bmatrix}, \quad G_t^a = \begin{bmatrix} G_t \\ 0 \end{bmatrix},$$

$H_t^a = [H_t, (0, V_t)]$ and $x_t^a = (x_t', \gamma')'$. Then, some extra (Bayesian) assumption is made about γ, for example $\gamma \sim [(b', c')', \Pi_1]$ with Π_1 nonsingular and γ uncorrelated with x and the $\{u_t\}$ and $\{v_t\}$ sequences, and the ordinary Kalman filter is applied to (4.97) and (4.98). The output of this Kalman filter will be denoted by

$$\widehat{E}_t = Y_t - H_t^a \hat{x}_{t|t-1}^a, \quad \widehat{\Sigma}_t = H_t^a \widehat{P}_t^a H_t^{a'} + R_t \tag{4.99}$$

$$\widehat{K}_t^a = (F_t^a \widehat{P}_t^a H_t^{a'} + G_t^a S_t)\widehat{\Sigma}_t^{-1}, \quad \hat{x}_{t+1|t}^a = F_t^a \hat{x}_{t|t-1}^a + \widehat{K}_t^a \widehat{E}_t \tag{4.100}$$

$$\widehat{P}_{t+1}^a = (F_t^a - \widehat{K}_t^a H_t^a)\widehat{P}_t^a F_t^{a'} + (G_t^a Q_t - \widehat{K}_t^a S_t')G_t^{a'}, \tag{4.101}$$

where

$$\widehat{P}_t^a = \begin{bmatrix} \widehat{P}_t & P_t^\gamma \\ P_t^{\gamma'} & \Pi_t \end{bmatrix}, \quad \hat{x}_{t|t-1}^a = \begin{bmatrix} \hat{x}_{t|t-1} \\ \hat{\gamma}_t \end{bmatrix}, \quad \widehat{K}_t^a = \begin{bmatrix} \widehat{K}_t \\ K_t^\gamma \end{bmatrix}, \tag{4.102}$$

and the initial conditions are $\hat{x}_{1|0}^a = (b'A' + c'W' + a', b', c')'$ and

$$\widehat{P}_1^a = \begin{bmatrix} \widehat{P}_1 & P_1^\gamma \\ P_1^{\gamma'} & \Pi_1 \end{bmatrix} = \left[\begin{array}{c|c} (A, W)\Pi_1(A, W)' + \Omega & (A, W)\Pi_1 \\ \hline \Pi_1(A, W)' & \Pi_1 \end{array}\right]. \tag{4.103}$$

The **two-stage Kalman filter (TSKF)** was originally proposed by Friedland (1969) and later developed by other authors, see Ignagni (1981) and the references therein, to address the problem of handling γ in the state space model (4.82) and (4.83) with the aim of simplifying the augmented-state Kalman filter. These results seem to have passed unnoticed in the statistical literature and, in fact, some of them have been reinvented by several authors, like Rosenberg (1973) and Jong (1991).

The TSKF consists of two decoupled Kalman filters that produce the same output as the augmented-state Kalman filter but with less computational effort as we will see later in this chapter. The first filter is called the **modified bias-free filter** and is given by the recursions

$$(E_t, e_t) = (0, V_t, Y_t) - H_t(-U_t, x_{t|t-1}) \tag{4.104}$$

$$\Sigma_t = H_t P_t H_t' + R_t, \quad K_t = (F_t P_t H_t' + G_t S_t)\Sigma_t^{-1} \tag{4.105}$$

$$(-U_{t+1}, x_{t+1|t}) = (0, -W_t, 0) + F_t(-U_t, x_{t|t-1}) + K_t(E_t, e_t) \tag{4.106}$$

$$P_{t+1} = (F_t - K_t H_t)P_t F_t' + (G_t Q_t - K_t S_t')G_t', \tag{4.107}$$

with initial conditions $(-U_1, x_{1|0}) = (-A, -W, a)$ and $P_1 = \Omega$. The second filter is called the **bias filter** and is given by the recursions

$$\widehat{E}_t = e_t - E_t \hat{\gamma}_t, \qquad \widehat{\Sigma}_t = \Sigma_t + E_t \Pi_t E_t' \tag{4.108}$$

$$K_t^\gamma = \Pi_t E_t' \widehat{\Sigma}_t^{-1}, \quad \hat{\gamma}_{t+1} = \hat{\gamma}_t + K_t^\gamma \widehat{E}_t \tag{4.109}$$

$$\Pi_{t+1} = \Pi_t - K_t^\gamma \widehat{\Sigma}_t K_t^{\gamma'} = \Pi_t - K_t^\gamma E_t \Pi_t, \tag{4.110}$$

initialized with $\hat{\gamma}_1 = (b', c')'$ and $\text{MSE}(\hat{\gamma}_1) = \Pi_1$. Note that this filter is the Kalman filter corresponding to the regression model $e_t = E_t\gamma + a_t$, where $\{a_t\}$ is an uncorrelated sequence of zero mean random vectors with $\text{Var}(a_t) = \Sigma_t$ and the state space form is

$$x_{t+1}^{\gamma} = x_t^{\gamma} \tag{4.111}$$

$$e_t = E_t x_t^{\gamma} + a_t, \tag{4.112}$$

where $x_t^{\gamma} = \gamma$.

In the proof of the next theorem we will show that the modified bias-free filter is equal to the so-called **bias-free filter** with the addition of two recursions for the so-called **blending matrices**, E_t and U_t. The bias-free filter is the Kalman filter for the state space model (4.82) and (4.83) under the assumption $\gamma = 0$, and is given by the recursions

$$e_t = Y_t - H_t x_{t|t-1}, \quad \Sigma_t = H_t P_t H_t' + R_t, \quad K_t = (F_t P_t H_t' + G_t S_t)\Sigma_t^{-1} \tag{4.113}$$

$$x_{t+1|t} = F_t x_{t|t-1} + K_t e_t, \quad P_{t+1} = (F_t - K_t H_t)P_t F_t' + (G_t Q_t - K_t S_t')G_t', \tag{4.114}$$

with the initialization $x_{1|0} = a$ and $P_1 = \Omega$. If in (4.104)–(4.107) we postmultiply Eqs. (4.104) and (4.106) by $(-\gamma', 1)'$, we obtain the so-called **perfectly known bias filter**,

$$\begin{aligned}
e_t - E_t\gamma &= Y_t - (0, V_t)\gamma - H_t(x_{t|t-1} + U_t\gamma) \\
\Sigma_t &= H_t P_t H_t' + R_t, \quad K_t = (F_t P_t H_t' + G_t S_t)\Sigma_t^{-1} \\
x_{t+1|t} + U_{t+1}\gamma &= (0, W_t)\gamma + F_t(x_{t|t-1} + U_t\gamma) + K_t(e_t - E_t\gamma) \\
P_{t+1} &= (F_t - K_t H_t)P_t F_t' + (G_t Q_t - K_t S_t')G_t',
\end{aligned}$$

initialized with $x_{1|0} + U_1\gamma = a + (A, W)\gamma$ and $P_1 = \Omega$. In the proof of the next theorem we will show that this filter is the Kalman filter for (4.82) and (4.83) when γ is assumed to be fixed and known.

The intuition behind the TSKF is that if we assume γ to be fixed and known in (4.82) and (4.83) and we apply the modified bias-free Kalman filter to this model, we transform the GLS model (4.87) into the model $e = E\gamma + a$, where $e = (e_1', \ldots, e_n')', E = (E_1', \ldots, E_n')'$ and $\text{Var}(a) = \text{diag}(\Sigma_1, \ldots, \Sigma_n)$. We can, at the same time, apply the Kalman filter to this transformed model to recursively obtain the estimator of γ and its mean squared error and this is precisely what the bias-filter does.

Note that, if we assume $\epsilon \sim (0, \Sigma)$ in the model (4.87), the application of the modified bias-free Kalman filter to this model produces a triangular factorization of Σ of the form $\Sigma = LDL'$, where L is a lower triangular matrix with ones in the main diagonal and D is a block diagonal matrix, such that $e = L^{-1}Y, E = L^{-1}X$ and $D = \mathrm{Var}(a)$. In the following examples, we will illustrate how the application of the modified bias-free Kalman filter to some state space models (4.82) and (4.83) produces automatically the matrices (e, E) and D corresponding to the regression model $e = E\gamma + a$, without explicitly constructing the matrix L^{-1} such that $e = L^{-1}Y$ and $E = L^{-1}X$.

Example 4.7 Suppose the regression model $Y_t = V_t\beta + S_t$, where S_t follows the model $(1 - \rho B)S_t = A_t, B$ is the backshift operator, $B^jY_t = Y_{t-j}$, and the process $\{A_t\}$ is serially uncorrelated with zero mean and $\mathrm{Var}(A_t) = \sigma^2$.

The model can be easily put into state space form (4.85) and (4.86) by defining $x_t = S_t, \epsilon_t = A_{t+1}, H_t = 1, J_t = 0, G_t = 1, W_t = 0$, and $F_t = \rho$. The initial state vector is $x_1 = x$, where $x = S_1 \sim (0, \sigma^2\Omega)$ and $\Omega = 1/(1 - \rho^2)$.

The modified bias-free filter is initialized with $(-U_1, x_{1|0}) = (0, 0)$ and $P_1 = \Omega$. The recursions yield $(E_1, e_1) = (V_1, Y_1), \Sigma_1 = 1/(1 - \rho^2), K_1 = \rho, (-U_2, x_{2|1}) = \rho(V_1, Y_1), P_2 = 1$ and

$$
\begin{aligned}
(E_t, e_t) &= (V_t - \rho V_{t-1}, Y_t - \rho Y_{t-1}), \quad \Sigma_t = 1 \\
K_t &= \rho, \quad (-U_{t+1}, \hat{x}_{t+1|t}) = \rho(-U_t, \hat{x}_{t|t-1}) + K_t(E_t, e_t) = (0, V_t, Y_t) \\
P_{t+1} &= 1, \quad t = 2, 3, \ldots, n.
\end{aligned}
$$

It is easy to see that in this case the matrices L^{-1} and D are

$$
L^{-1} = \begin{bmatrix} 1 & & & & \\ -\rho & 1 & & & \\ & -\rho & 1 & & \\ & & \ddots & \ddots & \\ & & & -\rho & 1 \end{bmatrix}, \quad D = \begin{bmatrix} 1/(1-\phi^2) & & & & \\ & 1 & & & \\ & & 1 & & \\ & & & \ddots & \\ & & & & 1 \end{bmatrix}.
$$

◇

Example 4.8 Assume the regression model $Y_t = V_t\beta + S_t$, where S_t follows the model $(1-B)S_t = A_t, B$ is the backshift operator, $B^jY_t = Y_{t-j}$, and the process $\{A_t\}$ is serially uncorrelated with zero mean and $\mathrm{Var}(A_t) = \sigma^2$.

The model can be cast into state space form (4.85) and (4.86) by defining $x_t = S_t, \epsilon_t = A_{t+1}, H_t = 1, J_t = 0, G_t = 1, W_t = 0$, and $F_t = 1$. The initial state vector is $x_1 = x + A\delta$, where $x = A_1, A = 1$ and $\delta = S_0$.

The modified bias-free filter is initialized with $(-U_1, x_{1|0}) = (-1, 0, 0)$ and $P_1 = 1$. The recursions yield $(E_1, e_1) = (1, V_1, Y_1), \Sigma_1 = 1, K_1 = 1, (-U_2, x_{2|1}) = (0, V_1, Y_1), P_2 = 1$ and

$$
\begin{aligned}
(E_t, e_t) &= (0, V_t - V_{t-1}, Y_t - Y_{t-1}), \quad \Sigma_t = 1 \\
K_t &= 1, \quad (-U_{t+1}, \hat{x}_{t+1|t}) = (-U_t, \hat{x}_{t|t-1}) + (E_t, e_t) = (0, V_t, Y_t) \\
P_{t+1} &= 1, \quad t = 2, 3, \ldots, n.
\end{aligned}
$$

It is not difficult to verify that in this case the matrices L^{-1} and D are

$$L^{-1} = \begin{bmatrix} 1 & & & & \\ -1 & 1 & & & \\ & -1 & 1 & & \\ & & \ddots & \ddots & \\ & & & -1 & 1 \end{bmatrix}, \; D = \begin{bmatrix} 1 & & & & \\ & 1 & & & \\ & & 1 & & \\ & & & \ddots & \\ & & & & 1 \end{bmatrix}.$$

◇

Example 4.9 Consider the model $(1+\phi B)(1-B)Y_t = A_t$, where B is the backshift operator, $B^j Y_t = Y_{t-j}$, and the process $\{A_t\}$ is serially uncorrelated with zero mean and $\mathrm{Var}(A_t) = \sigma^2$.

The model can be put into state space form (4.85) and (4.86) by defining $x_t = (Y_t, Y_{t+1|t})', Y_{t+1|t} = Y_{t+1} - A_{t+1}, \epsilon_t = A_{t+1}, V_t = 0, H_t = (1,0), J_t = 0, G_t = (1, 1-\phi)', W_t = 0$, and

$$F_t = \begin{bmatrix} 0 & 1 \\ \phi & 1-\phi \end{bmatrix}.$$

Let $U_t = Y_t - Y_{t-1}$ and $\delta = Y_0$. Then, $Y_1 = U_1 + \delta$ and $Y_{2|1} = U_1 + \delta + U_2 - A_2$. Since U_t follows the model $(1+\phi B)U_t = A_t$, the initial state vector is $x_1 = x + A\delta$, where $A = (1,1)', x = (U_1, (1-\phi)U_1)' \sim (0, \sigma^2\Omega)$ and

$$\Omega = \frac{1}{1-\phi^2}\begin{bmatrix} 1 & 1-\phi \\ 1-\phi & (1-\phi)^2 \end{bmatrix}.$$

It is shown in Problem 4.6 that the matrices L^{-1} and D are in this case

$$L^{-1} = \begin{bmatrix} 1 & & & & \\ \phi & 1 & & & \\ & \phi & 1 & & \\ & & \ddots & \ddots & \\ & & & \phi & 1 \end{bmatrix} \times \begin{bmatrix} 1 & & & & \\ -1 & 1 & & & \\ & -1 & 1 & & \\ & & \ddots & \ddots & \\ & & & -1 & 1 \end{bmatrix}$$

and

$$D = \begin{bmatrix} 1/(1-\phi^2) & & & & \\ & 1 & & & \\ & & 1 & & \\ & & & \ddots & \\ & & & & 1 \end{bmatrix}.$$

◇

The modified bias-free filter coincides with the first part of the diffuse Kalman filter of Jong (1991). We will see later in this chapter that the second part of de Jong's filter, the so-called Q_t recursion, is an information form of the bias-filter adapted to the diffuse situation, in which the covariance matrices of the estimator $\hat{\gamma}_t$ are infinite at the start of filtering.

The relation of the TSKF with the augmented-state Kalman filter is given by the following theorem. This result does not seem to have appeared in the statistical literature before. We will prove the theorem in a more general setting than that assumed by Ignagni (1981).

Theorem 4.35 *The quantities $\widehat{E}_t$, $\widehat{\Sigma}_t$, K_t^{γ}, $\hat{\gamma}_t$, and Π_t given by the TSKF coincide with the corresponding ones of the augmented-state Kalman filter, given by (4.99)–(4.101). In addition, the following relations between the two filters hold.*

$$\hat{x}_{t+1|t} = x_{t+1|t} + U_{t+1}\hat{\gamma}_{t+1}, \quad \widehat{P}_{t+1} = P_{t+1} + U_{t+1}\Pi_{t+1}U'_{t+1} \tag{4.115}$$

$$\widehat{K}_t = K_t + U_{t+1}K_t^{\gamma}, \quad P_t^{\gamma} = U_t\Pi_t. \tag{4.116}$$

Proof If $\gamma = 0$ in the state space model (4.82) and (4.83), this state space model is equal to (4.1) and (4.2) and its Kalman filter is (4.3), that coincides with the bias-free Kalman filter (4.113) and (4.114). If we assume γ to be fixed and known in (4.94), we should modify (4.113) and (4.114) to get the perfectly known bias filter

$$\bar{e}_t = Y_t - (0, V_t)\gamma - H_t\bar{x}_t, \quad \Sigma_t = H_tP_tH_t' + R_t, \quad K_t = (F_tP_tH_t' + G_tS_t)\Sigma_t^{-1}$$
$$\bar{x}_{t+1} = (0, W_t)\gamma + F_t\bar{x}_t + K_t\bar{e}_t, \quad P_{t+1} = (F_t - K_tH_t)P_tF_t' + (G_tQ_t - K_tS_t')G_t',$$

with initial conditions $\bar{x}_1 = a + (A, W)\gamma$ and $P_1 = \Omega$.

It is clear from the previous recursions that $\bar{e}_t$ and $\bar{x}_t$ are of the form $\bar{e}_t = e_t - E_t\gamma$ and $\bar{x}_t = x_{t|t-1} + U_t\gamma$, where E_t and U_t are matrices to be determined. A simple inspection of the previous recursions shows that U_t is given by the recursion $U_{t+1} = (0, W_t) + F_tU_t - K_tE_t$, initialized with $U_1 = (A, W)$, and that $E_t = (0, V_t) + H_tU_t$, $t = 1, \ldots, n$. Therefore, the bias-free filter with the addition of these last two recursions coincides with the modified bias-free filter (4.104)–(4.107).

To see the relation between the augmented-state Kalman filter and the two-stage Kalman filter, we will make a transformation of the augmented-state Kalman filter equations (4.99)–(4.101) so that the MSE matrices of the transformed state estimators are block diagonal. To this end, consider the Schur decomposition of $\widehat{P}_t^a$,

$$\widehat{P}_t^a = \begin{bmatrix} \widehat{P}_t & P_t^{\gamma} \\ P_t^{\gamma'} & \Pi_t \end{bmatrix} = \begin{bmatrix} I & P_t^{\gamma}\Pi_t^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} \widehat{P}_t - P_t^{\gamma}\Pi_t^{-1}P_t^{\gamma'} & 0 \\ 0 & \Pi_t \end{bmatrix} \begin{bmatrix} I & 0 \\ \Pi_t^{-1}P_t^{\gamma'} & I \end{bmatrix}.$$

This relation can be written as $\widehat{P}^a_t = T(U_t)\text{diag}(P_t, \Pi_t)T(U_t)'$, where $U_t = P^\gamma_t \Pi^{-1}_t, P_t = \widehat{P}_t - P^\gamma_t \Pi^{-1}_t P^{\gamma'}_t$ and

$$T(U_t) = \begin{bmatrix} I & U_t \\ 0 & I \end{bmatrix}.$$

Since $T^{-1}(U_t) = T(-U_t)$, if we define $\overline{x}^a_{t|t-1} = (x'_{t|t-1}, \hat{\gamma}'_t)' = T(-U_t)\hat{x}^a_{t|t-1}$, Eqs. (4.99)–(4.101) are transformed into

$$\widehat{E}_t = Y_t - \overline{H}^a_t \overline{x}^a_{t|t-1}, \quad \widehat{\Sigma}_t = \overline{H}^a_t \overline{P}^a_t \overline{H}^{a'}_t + R_t \tag{4.117}$$

$$\overline{K}^a_t = (\overline{F}^a_t \overline{P}^a_t \overline{H}^{a'}_t + \overline{G}^a_t S_t)\widehat{\Sigma}^{-1}_t, \quad \overline{x}^a_{t+1|t} = \overline{F}^a_t \overline{x}^a_{t|t-1} + \overline{K}^a_t \widehat{E}_t \tag{4.118}$$

$$\overline{P}^a_{t+1} = (\overline{F}^a_t - \overline{K}^a_t \overline{H}^a_t)\overline{P}^a_t \overline{F}^{a'}_t + (\overline{G}^a_t Q_t - \overline{K}^a_t S'_t)\overline{G}^{a'}_t, \tag{4.119}$$

where $\overline{P}^a_t = T(-U_t)\widehat{P}^a_t T(-U_t)' = \text{diag}(P_t, \Pi_t), \overline{F}^a_t = T(-U_{t+1})F^a_t T(U_t), \overline{H}^a_t = H^a_t T(U_t), \overline{G}^a_t = T(-U_{t+1})G^a_t$ and $\overline{K}^a_t = T(-U_{t+1})\widehat{K}^a_t$.

Letting $\overline{K}^a_t = \left[K'_t, K^{\gamma'}_t\right]'$ and $\overline{H}^a_t = [H_t, E_t]$, our aim is to show that the quantities $x_{t|t-1}, U_t, E_t, P_t, K_t$, and K^γ_t, defined in the transformed filter, coincide with those of the modified bias-free filter. We will prove this by verifying that these quantities satisfy the same recursions and initial conditions as those of the modified bias-free filter. By definition of $\overline{H}^a_t$,

$$E_t = (0, V_t) - H_t U_t, \tag{4.120}$$

and substituting $\overline{H}^a_t = [H_t, E_t]$ into (4.117), it is obtained that

$$\widehat{E}_t = e_t - E_t \hat{\gamma}_t, \quad \widehat{\Sigma}_t = \Sigma_t + E_t \Pi_t E'_t, \tag{4.121}$$

where

$$e_t = Y_t - H_t x_{t|t-1}, \quad \Sigma_t = H_t P_t H'_t + R_t. \tag{4.122}$$

Thus, assuming, as we will show later, that $x_{t|t-1}$ and Σ_t are those of the bias-free filter, Eq. (4.121) is Eq. (4.108) of the bias filter.

From the definitions of $\overline{F}^a_t, \overline{K}^a_t, \overline{x}^a_{t+1|t}$ and $\overline{P}^a_t$, it is obtained that

$$\overline{F}^a_t = \begin{bmatrix} F_t & A_t \\ 0 & I \end{bmatrix}, \quad \overline{K}^a_t = \begin{bmatrix} K_t \\ K^\gamma_t \end{bmatrix} = \begin{bmatrix} F_t P_t H'_t + A_t \Pi_t E'_t + G_t S_t \\ \Pi_t E'_t \end{bmatrix} \widehat{\Sigma}^{-1}_t, \tag{4.123}$$

$$\overline{x}^a_{t+1|t} = \begin{bmatrix} x_{t+1|t} \\ \hat{\gamma}_{t+1} \end{bmatrix} = \begin{bmatrix} F_t x_{t|t-1} + A_t \hat{\gamma}_t + K_t \widehat{E}_t \\ \hat{\gamma}_t + K^\gamma_t \widehat{E}_t \end{bmatrix}, \tag{4.124}$$

and

$$\overline{P}^a_{t+1} = \begin{bmatrix} (F_t - K_tH_t)P_tF_t' + \overline{U}_t\Pi_tA_t' & \overline{U}_t\Pi_t \\ * & \Pi_t - K_t^\gamma E_t\Pi_t \end{bmatrix} + \begin{bmatrix} (G_tQ_t - K_tS_t')G_t' & 0 \\ * & 0 \end{bmatrix}$$
$$= \begin{bmatrix} P_{t+1} & 0 \\ 0 & \Pi_{t+1} \end{bmatrix}, \quad (4.125)$$

where $A_t = -U_{t+1} + W_t + F_tU_t$, $\overline{U}_t = A_t - K_tE_t$ and the asterisk indicates elements that are not relevant to our purposes. Thus, by (4.123)–(4.125), (4.109) and (4.110) of the bias filter are satisfied and, since $\overline{P}^a_{t+1}$ is block diagonal, (4.125) implies $\overline{U}_t\Pi_t = 0$. Because Π_t is nonsingular, this in turn implies $\overline{U}_t = 0$. Thus, $A_t = K_tE_t$ and, by (4.125),

$$P_{t+1} = (F_t - K_tH_t)P_tF_t' + (G_tQ_t - K_tS_t')G_t'. \quad (4.126)$$

From the equality $A_t = K_tE_t$ and the definition of A_t, we obtain the recursion

$$U_{t+1} = W_t + F_tU_t - K_tE_t. \quad (4.127)$$

In addition, substituting $A_t = K_tE_t$ in (4.124) and (4.123) yields

$$x_{t+1|t} = F_tx_{t|t-1} + K_te_t \quad (4.128)$$

and $K_t = \left(F_tP_tH_t' + K_tE_t\Pi_tE_t' + G_tS_t\right)\widehat{\Sigma}_t^{-1}$. This last expression implies $K_t(\widehat{\Sigma}_t - E_t\Pi_tE_t') = F_tP_tH_t' + G_tS_t$ and, because $\widehat{\Sigma}_t = \Sigma_t + E_t\Pi_tE_t'$,

$$K_t = \left(F_tP_tH_t' + G_tS_t\right)\Sigma_t^{-1}. \quad (4.129)$$

The recursions (4.120), (4.122), (4.126)–(4.129) coincide with those of the modified bias-free filter, (4.104)–(4.107). To check that the initial conditions also coincide, consider first that $(x_{1|0}', \hat{\gamma}_1')' = T(-U_1)\hat{x}_1^a$ and $\mathrm{diag}(P_1, \Pi_1) = T(-U_1)\widehat{P}_1^aT(-U_1)'$. Then, by (4.103), $U_1 = P_1^\gamma\Pi_1^{-1} = (A, W)$, $x_{1|0} = a$, $\hat{\gamma}_1 = (b', c')'$ and $P_1 = \Omega$.

It remains to prove the relations (4.115) and (4.116). From the definitions of U_t and P_t it follows that $\widehat{P}_{t+1} = P_{t+1} + U_{t+1}\Pi_{t+1}U_{t+1}'$ and $P_t^\gamma = U_t\Pi_t$. The expressions for $\hat{x}_{t+1|t}$ and $\widehat{K}_t$ follow from $\hat{x}_t^a = T(U_t)\bar{x}_t^a$ and $\widehat{K}_t^a = T(U_{t+1})\overline{K}_t^a$. □

Remark 4.8 From the previous theorem, we see that the extra computations of the augmented-state Kalman filter with respect to the TSKF are the recursions (4.115) and (4.116). These recursions are not needed for likelihood evaluation and the recursions (4.115) are only needed for prediction and smoothing. The recursions (4.116) are never needed in normal applications. We conclude that the augmented-state Kalman filter should not be used in current applications. One should use the two-stage Kalman filter because it requires less computational effort, and if one needs any of the recursions (4.115) and (4.116), one can obtain them using the output of this filter. ◇

Remark 4.9 We can give a statistical meaning to the recursions (4.115). They are the result of replacing the unknown γ in the perfectly known bias filter with its estimator $\hat{\gamma}_t$, based on the sample $Y_{1:t-1} = \{Y_1, \ldots, Y_{t-1}\}$. More specifically, if $\bar{x}_t = x_{t|t-1} + U_t\gamma$ is the estimator of x_t based on $\{Y_{1:t-1}, \gamma\}$, then the estimator, $\hat{x}_{t|t-1}$, of x_t based on $Y_{1:t-1}$ is obtained by replacing γ with $\hat{\gamma}_t$ in $\bar{x}_t$. This gives $\hat{x}_{t|t-1} = x_{t|t-1} + U_t\hat{\gamma}_t$. To obtain its MSE, consider that $\text{MSE}(\hat{x}_{t|t-1}) = \text{MSE}(\bar{x}_t) + Var\,[U_t(\hat{\gamma}_t - \gamma)]$ because $x_t - \bar{x}_t$ and $\hat{\gamma}_t$ are uncorrelated. ◇

Remark 4.10 Instead of using the bias filter to recursively compute the estimator of γ and its mean square error in the model $e = E\gamma + a$, where $e = (e_1', \ldots, e_n')'$, $E = (E_1', \ldots, E_n')'$ and $\text{Var}(a) = \text{diag}(\Sigma_1, \ldots, \Sigma_n)$, it is possible not to use this filter and to proceed in two steps. In the first step, the modified bias-free filter is run and the quantities e_t, E_t, and Σ_t are stored. In the second step, γ is estimated in the regression model $e = E\gamma + a$. This procedure was used in, for example, Doménech & Gómez (2006) when dealing with a state space model for the U.S. economy. ◇

Remark 4.11 The TSKF cannot handle diffuse situations, in which the covariance matrices of the bias filter are infinite. For this reason, the information form of this filter should be used in this case. ◇

4.16 Information Form and Square Root Information Form of the Bias Filter

As mentioned in the previous section, the bias filter (4.108)–(4.110), initialized with $\hat{\gamma}_1 = (b', c')'$ and $\text{MSE}(\hat{\gamma}_1) = \Pi_1$, is the Kalman filter corresponding to the regression model $e_t = E_t\gamma + a_t, t = 1, \ldots, n$, where $\{a_t\}$ is an uncorrelated sequence with zero mean and $\text{Var}(a_t) = \Sigma_t$, the state space form is given by (4.111) and (4.112), and e_t, E_t, and Σ_t are given by the modified bias-free filter.

The bias filter cannot be used when Π_1 is infinite and, therefore, γ is diffuse, which is what is usually assumed when one wants to recursively compute the estimator of γ and its MSE. For this reason, it is important to use an information form of the Kalman filter for the previous regression model. In the information form, it is the inverses of the covariance matrices, Π_t^{-1}, that are propagated instead of the covariance matrices themselves, Π_t.

Using the Matrix Inversion Lemma 4.1, we can prove the following theorem.

Theorem 4.36 (Information Form of the Bias Filter) *Under the assumptions and with the notation of the bias filter (4.108)–(4.110) for the regression model $e_t = E_t\gamma + a_t$ and assuming γ diffuse, the matrices Π_{t+1}^{-1} and $\Pi_{t+1}^{-1}\hat{\gamma}_{t+1}$ satisfy the recursions*

$$\left(\Pi_{t+1}^{-1}, \Pi_{t+1}^{-1}\hat{\gamma}_{t+1}\right) = \left(\Pi_t^{-1}, \Pi_t^{-1}\hat{\gamma}_t\right) + E_t'\Sigma_t^{-1}\left(E_t, e_t\right), \tag{4.130}$$

initialized with $\left(\Pi_1^{-1}, \Pi_1^{-1}\hat{\gamma}_1\right) = (0, 0)$.

Proof The Theorem follows from a direct application of Theorem 4.13. However, we give a proof from scratch. Applying Lemma 4.1 to $\Pi_{t+1} = \Pi_t - \Pi_t E_t'(\Sigma_t + E_t \Pi_t E_t')^{-1} E_t \Pi_t$, we get

$$\Pi_{t+1}^{-1} = \Pi_t^{-1} + E_t' \Sigma_t^{-1} E_t. \tag{4.131}$$

On the other hand,

$$\begin{aligned} K_t^{\gamma} &= \Pi_t E_t' \widehat{\Sigma}_t^{-1} = \Pi_t E_t' \Sigma_t^{-1} - \Pi_t E_t' \widehat{\Sigma}_t^{-1} \left(\widehat{\Sigma}_t - \Sigma_t \right) \Sigma_t^{-1} \\ &= \left(\Pi_t - \Pi_t E_t' \widehat{\Sigma}_t^{-1} E_t \Pi_t \right) E_t' \Sigma_t^{-1} = \Pi_{t+1} E_t' \Sigma_t^{-1}, \end{aligned} \tag{4.132}$$

and thus, using (4.131), it is obtained that

$$\begin{aligned} \Pi_{t+1}^{-1} \hat{\gamma}_{t+1} &= \Pi_{t+1}^{-1} \left[\hat{\gamma}_t + K_t^{\gamma} \widehat{E}_t \right] = \Pi_{t+1}^{-1} \left[\left(I - \Pi_{t+1} E_t' \Sigma_t^{-1} E_t \right) \hat{\gamma}_t + \Pi_{t+1} E_t' \Sigma_t^{-1} e_t \right] \\ &= \Pi_{t+1}^{-1} \left[\Pi_{t+1} \Pi_t^{-1} \hat{\gamma}_t + \Pi_{t+1} E_t' \Sigma_t^{-1} e_t \right] \\ &= \Pi_t^{-1} \hat{\gamma}_t + E_t' \Sigma_t^{-1} e_t. \end{aligned} \tag{4.133}$$

If we put (4.131) and (4.133) together, we get (4.130). □

Remark 4.12 Using (4.130) to get $\hat{\gamma}_{t+1}$ and Π_{t+1} requires less computational effort than (4.108)–(4.110) because (4.108) and $K_t^{\gamma} = \Pi_t E_t' \widehat{\Sigma}_t^{-1}$ are not needed. If the orthogonal residuals, $\widehat{E}_t$, and their MSE, $\widehat{\Sigma}_t$, are needed for inference or for other purposes, they can be computed along with (4.130). ◇

Remark 4.13 The recursion (4.130) allows for the computation of $\ln |R' \Sigma^{-1} R|$ in the concentrated diffuse log-likelihood (4.95) because $R' \Sigma^{-1} R$ is the submatrix of Π_{n+1}^{-1} formed with the first n_δ rows and the first n_δ columns. This can be seen first by considering that, by Theorem 4.34, the first n_δ columns of the X matrix constitute the R. Then, the modified bias-free filter transforms the GLS model (4.87), where $X = [R, S]$ and $\epsilon \sim (0, \Sigma)$, into the model $e = E\gamma + a$, where $e = (e_1', \ldots, e_n')'$, $E = (E_1', \ldots, E_n')'$ and $\mathrm{Var}(a) = \mathrm{diag}(\Sigma_1, \ldots, \Sigma_n)$. Finally, since $\Pi_{n+1} = \mathrm{MSE}(\gamma_{n+1}) = (X' \Sigma^{-1} X)^{-1}$, we can write

$$\Pi_{n+1}^{-1} = \begin{bmatrix} R' \Sigma^{-1} R & R' \Sigma^{-1} S \\ S' \Sigma^{-1} R & S' \Sigma^{-1} S \end{bmatrix} = \begin{bmatrix} \Pi_{n+1}^{11} & \Pi_{n+1}^{12} \\ \Pi_{n+1}^{21} & \Pi_{n+1}^{22} \end{bmatrix}. \tag{4.134}$$

◇

The following lemma will be useful to obtain the square root information form of the bias filter.

Lemma 4.13 *Under the assumptions and with the notation of the bias filter (4.108)–(4.110) for the regression model* $e_t = E_t \gamma + a_t$, *the following relations* *hold*

$$\widehat{\Sigma}_t^{-1}\widehat{E}_t = \Sigma_t^{-1}\left(e_t - E_t\hat{\gamma}_{t+1}\right) \tag{4.135}$$

$$\widehat{E}_t'\widehat{\Sigma}_t^{-1}\widehat{E}_t = e_t'\Sigma_t^{-1}e_t + \hat{\gamma}_t'\Pi_t^{-1}\hat{\gamma}_t - \hat{\gamma}_{t+1}'\Pi_{t+1}^{-1}\hat{\gamma}_{t+1}. \tag{4.136}$$

Proof By the Matrix Inversion Lemma 4.1, it holds that $\widehat{\Sigma}_t^{-1} = \Sigma_t^{-1} - \Sigma_t^{-1}E_t\Pi_{t+1}E_t'\Sigma_t^{-1}$ and, using (4.108)–(4.110) and (4.132), it is obtained that

$$\begin{aligned}
\widehat{\Sigma}_t^{-1}\widehat{E}_t &= \left(\Sigma_t^{-1} - \Sigma_t^{-1}E_t\Pi_{t+1}E_t'\Sigma_t^{-1}\right)\widehat{E}_t = \Sigma_t^{-1}\left(e_t - E_t\hat{\gamma}_t\right) - \Sigma_t^{-1}E_t\Pi_{t+1}E_t'\Sigma_t^{-1}\widehat{E}_t \\
&= \Sigma_t^{-1}\left[e_t - E_t\left(\hat{\gamma}_t + \Pi_{t+1}E_t'\Sigma_t^{-1}\widehat{E}_t\right)\right] = \Sigma_t^{-1}\left[e_t - E_t\left(\hat{\gamma}_t + K_t^{\gamma}\widehat{E}_t\right)\right] \\
&= \Sigma_t^{-1}\left(e_t - E_t\hat{\gamma}_{t+1}\right).
\end{aligned}$$

Using (4.135) and (4.130), it is obtained that

$$\begin{aligned}
\widehat{E}_t'\widehat{\Sigma}_t^{-1}\widehat{E}_t &= \widehat{E}_t'\Sigma_t^{-1}\left(e_t - E_t\hat{\gamma}_{t+1}\right) = \left(e_t - E_t\hat{\gamma}_t\right)'\Sigma_t^{-1}e_t - \widehat{E}_t'\Sigma_t^{-1}E_t\hat{\gamma}_{t+1} \\
&= e_t'\Sigma_t^{-1}e_t - \hat{\gamma}_t'\left(\Pi_{t+1}^{-1}\hat{\gamma}_{t+1} - \Pi_t^{-1}\hat{\gamma}_t\right) - \widehat{E}_t'\Sigma_t^{-1}E_t\hat{\gamma}_{t+1} \\
&= e_t'\Sigma_t^{-1}e_t + \hat{\gamma}_t'\Pi_t^{-1}\hat{\gamma}_t - \left[\hat{\gamma}_t'\Pi_{t+1}^{-1} + \left(e_t - E_t\hat{\gamma}_t\right)'\Sigma_t^{-1}E_t\right]\hat{\gamma}_{t+1} \\
&= e_t'\Sigma_t^{-1}e_t + \hat{\gamma}_t'\Pi_t^{-1}\hat{\gamma}_t - \left(\hat{\gamma}_t'\Pi_{t+1}^{-1} + \hat{\gamma}_{t+1}'\Pi_{t+1}^{-1} - \hat{\gamma}_t'\Pi_t^{-1}\right. \\
&\quad \left. -\hat{\gamma}_t'\Pi_{t+1}^{-1} + \hat{\gamma}_t'\Pi_t^{-1}\right)\hat{\gamma}_{t+1} \\
&= e_t'\Sigma_t^{-1}e_t + \hat{\gamma}_t'\Pi_t^{-1}\hat{\gamma}_t - \hat{\gamma}_{t+1}'\Pi_{t+1}^{-1}\hat{\gamma}_{t+1}.
\end{aligned}$$

□

The following theorem gives a formula for the recursive computation of the residual sum of squares of the regression model $e_t = E_t\gamma + a_t$. We omit its proof because it is an immediate consequence of the previous lemma.

Theorem 4.37 (Residual Sum of Squares) *Under the assumptions and with the notation of the bias filter (4.108)–(4.110) for the regression model* $e_t = E_t\gamma + a_t$ *and assuming* γ *diffuse, the following formula holds for the residual sum of squares*

$$\sum_{t=1}^{n}\widehat{E}_t'\widehat{\Sigma}_t^{-1}\widehat{E}_t = \sum_{t=1}^{n}e_t'\Sigma_t^{-1}e_t - \hat{\gamma}_{n+1}'\Pi_{n+1}^{-1}\hat{\gamma}_{n+1}. \tag{4.137}$$

Remark 4.14 Since $\hat{\gamma}_{n+1}'\Pi_{n+1}^{-1}\hat{\gamma}_{n+1} = \left(\Pi_{n+1}^{-1}\hat{\gamma}_{n+1}\right)'\Pi_{n+1}\left(\Pi_{n+1}^{-1}\hat{\gamma}_{n+1}\right)$, this quantity can be computed with the recursion (4.130). Thus, to compute (4.137), we can add to (4.130) the recursion

$$\mathrm{RSS}_{t+1}^{\mathrm{BFF}} = \mathrm{RSS}_t^{\mathrm{BFF}} + e_t'\Sigma_t^{-1}e_t, \tag{4.138}$$

initialized with $\text{RSS}_1^{\text{BFF}} = 0$, that computes the residual sum of squares of the bias-free filter. ◇

Remark 4.15 The residual sum of squares (4.137) is also equal to $(Y - X\hat{\gamma})'\Sigma^{-1}(Y - X\hat{\gamma})$ in the concentrated diffuse log-likelihood (4.95). This is so because, as mentioned in the previous section, the modified bias-free filter transforms the GLS model (4.87), where $\epsilon \sim (0, \Sigma)$, into the model $e = E\gamma + a$, where $e = (e_1', \ldots, e_n')', E = (E_1', \ldots, E_n')'$, $\text{Var}(a) = \text{diag}(\Sigma_1, \ldots, \Sigma_n), e = L^{-1}Y, E = L^{-1}X$ and L is a lower triangular matrix with ones in the main diagonal such that $\Sigma = L\text{Var(a)}L'$. Thus,

$$\begin{aligned}(Y - X\hat{\gamma}_{n+1})'\Sigma^{-1}(Y - X\hat{\gamma}_{n+1}) &= \sum_{t=1}^{n} (e_t - E_t\hat{\gamma}_{n+1})' \, \Sigma_t^{-1} \, (e_t - E_t\hat{\gamma}_{n+1}) \\ &= \sum_{t=1}^{n} e_t'\Sigma_t^{-1}e_t - \hat{\gamma}_{n+1}'\Pi_{n+1}^{-1}\hat{\gamma}_{n+1}\end{aligned}$$

and the concentrated diffuse log-likelihood (4.95) can be computed using the TSKF and the information form of the bias filter as

$$\begin{aligned}\lambda_D(Y; \hat{\beta}, \hat{\sigma}^2) = \text{constant} - \frac{1}{2} \Bigg\{ (np - n_\delta) \ln \Big[\text{RSS}_{n+1}^{\text{BFF}} - \left(\Pi_{n+1}^{-1}\hat{\gamma}_{n+1}\right)' \\ \times \Pi_{n+1} \left(\Pi_{n+1}^{-1}\hat{\gamma}_{n+1}\right) \Big] + \sum_{t=1}^{n} \ln |\Sigma_t| + \ln |\Pi_{n+1}^{11}| \Bigg\} ,\end{aligned} \tag{4.139}$$

where $\text{RSS}_{n+1}^{\text{BFF}}$ can be obtained with the recursion (4.138), n_δ is the dimension of δ, and Π_{n+1}^{11} is the $n_\delta \times n_\delta$ submatrix of Π_{n+1}^{-1} given in (4.134). ◇

The information form of the bias filter is basically an algorithm to compute the estimator and its MSE of the vector of parameters of a regression model. It is well known that it is important to have a stable numerical procedure to compute these quantities in a regression model. One of such procedures consists of using Householder transformations and the *QR* decomposition of a matrix. This is what the square root information form of the bias filter does.

In general, the information form of the Kalman filter propagates the square roots of the inverses of the covariance matrices of the filter instead of the covariance matrices themselves, where for any square matrix A a square root of A, denoted by $A^{1/2}$, is any matrix such that $A = A^{1/2}A^{1/2'}$. The square root information form of the bias filter is given by the following theorem.

Theorem 4.38 (Square Root Information Form of the Bias Filter) *Under the assumptions and with the notation of the bias filter (4.108)–(4.110) for the regression model $e_t = E_t\gamma + a_t$ and assuming γ diffuse, the QR algorithm produces* *an orthogonal matrix Q_t such that*

$$Q_t' \begin{bmatrix} \Pi_t^{-1/2} \\ \Sigma_t^{-1/2} E_t \end{bmatrix} = \begin{bmatrix} \Pi_{t+1}^{-1/2} \\ 0 \end{bmatrix}, \tag{4.140}$$

where $\Pi_{t+1}^{-1/2}$ is upper triangular or upper trapezoidal and $(\Pi_1^{-1/2}, \Pi_1^{-1/2}\hat{\gamma}_1) = (0, 0)$. In addition,

$$Q_t' \begin{bmatrix} \Pi_t^{-1/2} & \Pi_t^{-1/2}\hat{\gamma}_t \\ \Sigma_t^{-1/2} E_t & \Sigma_t^{-1/2} e_t \end{bmatrix} = \begin{bmatrix} \Pi_{t+1}^{-1/2} & \Pi_{t+1}^{-1/2}\hat{\gamma}_{t+1} \\ 0 & \widehat{\Sigma}_t^{-1/2}\widehat{E}_t \end{bmatrix}, \tag{4.141}$$

and if we add to the second matrix in the left-hand side an extra block column of the form $[0, \Sigma_t^{-1/2'}]'$, then the same matrix Q_t satisfies

$$Q_t' \begin{bmatrix} \Pi_t^{-1/2} & \Pi_t^{-1/2}\hat{\gamma}_t & 0 \\ \Sigma_t^{-1/2} E_t & \Sigma_t^{-1/2} e_t & \Sigma_t^{-1/2} \end{bmatrix} = \begin{bmatrix} \Pi_{t+1}^{-1/2} & \Pi_{t+1}^{-1/2}\hat{\gamma}_{t+1} & \Pi_{t+1}^{1/2'} E_t' \Sigma_t^{-1} \\ 0 & \widehat{\Sigma}_t^{-1/2}\widehat{E}_t & \widehat{\Sigma}_t^{-1/2} \end{bmatrix}. \tag{4.142}$$

Proof The theorem follows from Theorem 4.15 if we consider the model

$$\begin{aligned} x_{t+1} &= x_t \\ e_t &= E_t x_t + a_t, \end{aligned}$$

where $x_t = \gamma$. However, we give a proof from scratch. Since Q_t is orthogonal, we get from the left-hand side of (4.141), using (4.136) and (4.130),

$$\begin{aligned} &\begin{bmatrix} \Pi_t^{-1/2} & \Pi_t^{-1/2}\hat{\gamma}_t \\ \Sigma_t^{-1/2} E_t & \Sigma_t^{-1/2} e_t \end{bmatrix}' Q_t Q_t' \begin{bmatrix} \Pi_t^{-1/2} & \Pi_t^{-1/2}\hat{\gamma}_t \\ \Sigma_t^{-1/2} E_t & \Sigma_t^{-1/2} e_t \end{bmatrix} \\ &= \begin{bmatrix} \Pi_t^{-1/2'}\Pi_t^{-1/2} + E_t'\Sigma_t^{-1/2'}\Sigma_t^{-1/2}E_t & \Pi_t^{-1/2'}\Pi_t^{-1/2}\hat{\gamma}_t + E_t'\Sigma_t^{-1/2'}\Sigma_t^{-1/2}e_t \\ \hat{\gamma}_t'\Pi_t^{-1/2'}\Pi_t^{-1/2} + e_t'\Sigma_t^{-1/2'}\Sigma_t^{-1/2}E_t & \hat{\gamma}_t'\Pi_t^{-1/2'}\Pi_t^{-1/2}\hat{\gamma}_t + e_t'\Sigma_t^{-1/2'}\Sigma_t^{-1/2}e_t \end{bmatrix} \\ &= \begin{bmatrix} \Pi_{t+1}^{-1/2'}\Pi_{t+1}^{-1/2} & \Pi_{t+1}^{-1/2'}\Pi_{t+1}^{-1/2}\hat{\gamma}_{t+1} \\ \hat{\gamma}_{t+1}'\Pi_{t+1}^{-1/2'}\Pi_{t+1}^{-1/2} & \hat{\gamma}_{t+1}'\Pi_{t+1}^{-1/2'}\Pi_{t+1}^{-1/2}\hat{\gamma}_{t+1} + \widehat{E}_t'\widehat{\Sigma}_t^{-1/2'}\widehat{\Sigma}_t^{-1/2}\widehat{E}_t \end{bmatrix} \\ &= \begin{bmatrix} \Pi_{t+1}^{-1/2} & \Pi_{t+1}^{-1/2}\hat{\gamma}_{t+1} \\ 0 & \widehat{\Sigma}_t^{-1/2}\widehat{E}_t \end{bmatrix}' \begin{bmatrix} \Pi_{t+1}^{-1/2} & \Pi_{t+1}^{-1/2}\hat{\gamma}_{t+1} \\ 0 & \widehat{\Sigma}_t^{-1/2}\widehat{E}_t \end{bmatrix}. \end{aligned}$$

The rest of the theorem is proved similarly. □

Remark 4.16 The vectors $\widehat{\Sigma}_t^{-1/2}\widehat{E}_t$ have zero mean and unit covariance matrix. Thus, they constitute a sequence of "standardized residuals" and can be used for inference. ◇

Remark 4.17 To compute $\hat{\gamma}_{t+1}$ in the square root algorithm of the previous theorem, we have to solve the system $\left(\Pi_{t+1}^{-1/2}\right)\hat{\gamma}_{t+1} = \Pi_{t+1}^{-1/2}\hat{\gamma}_{t+1}$, where $\Pi_{t+1}^{-1/2}$ is an upper triangular matrix. This computation, based on back substitution, can be avoided by including in the second matrix in the left-hand side of (4.142) an extra block column of the form $[0, \Pi_t^{1/2}]'$ because

$$Q_t'\begin{bmatrix} 0 \\ \Pi_t^{1/2'} \end{bmatrix} = \begin{bmatrix} \Pi_{t+1}^{1/2'} \\ \widehat{\Sigma}_t^{-1/2}E_t\Pi_t \end{bmatrix}.$$

The validity of this formula can be verified using (4.110). ◇

4.17 Fast Square Root Information Form of the Bias Filter

As in the case of the information square root form of RLS, it is possible to use square root free fast Givens rotations, described in the Appendix of Chap. 2, to substantially reduce the amount of computation needed for the square root information form of the bias filter. To see this, assume first that Π_t^{-1} is nonsingular and put Π_t^{-1} into the form $\Pi_t^{-1} = L_tD_tL_t' = \Pi_t^{-1/2'}\Pi_t^{-1/2}$, where L_t is a lower triangular matrix with ones in the main diagonal and D_t is a diagonal matrix with positive elements in the main diagonal. Then, we can write expression (4.140) of Theorem 4.38 as

$$Q_t'\begin{bmatrix} D_t^{1/2} & 0 \\ 0 & I_p \end{bmatrix}\begin{bmatrix} L_t' \\ \Sigma_t^{-1/2}E_t \end{bmatrix} = \begin{bmatrix} D_{t+1}^{1/2}L_{t+1}' \\ 0 \end{bmatrix}. \tag{4.143}$$

We can thus clearly use fast Givens rotations, as described in the Appendix of Chap. 2 to obtain the *QDU* decomposition, to update L_t and D_t.

The previous update is equivalent, although more numerically stable, to use sequential processing in the model $e_t = E_t\gamma + a_t$ of the bias filter. To verify this, premultiply the previous equation by $\Sigma_t^{-1/2}$ to get $\Sigma_t^{-1/2}e_t = \Sigma_t^{-1/2}E_t\gamma + \Sigma_t^{-1/2}a_t$. Then, the new error terms have covariance matrices equal to the identity matrix and one can apply sequential processing and perform the update (4.141) using scalar observations one by one instead of one update using a vector of observations.

4.18 Evaluation of the Concentrated Diffuse Log-likelihood with the TSKF and the Information Form Bias Filter

As described in Remark 4.15, the TSKF with the information form of the bias filter can be used to evaluate the concentrated diffuse log-likelihood (4.95). This follows first from $(Y - X\hat{\gamma})'\Sigma^{-1}(Y - X\hat{\gamma}) = \sum_{t=1}^{n}\widehat{E}_t'\widehat{\Sigma}_t^{-1}\widehat{E}_t$. Then, by (4.134) in

Remark 4.13, $R'\Sigma^{-1}R$ is the submatrix of Π_{n+1}^{-1} formed with the first n_δ rows and the first n_δ columns. Putting it all together, we get the following expression for the concentrated diffuse log-likelihood (4.95),

$$\lambda_D(Y;\hat{\beta},\widehat{\sigma}^2) = \text{constant} - \frac{1}{2}\left\{(np-n_\delta)\ln\left(\sum_{t=1}^{n}\widehat{E}_t'\widehat{\Sigma}_t^{-1}\widehat{E}_t\right)\right.$$
$$\left.+\sum_{t=1}^{n}\ln|\Sigma_t| + \ln|\Pi_{n+1}^{11}|\right\},$$

where Π_{n+1}^{11} is the $n_\delta \times n_\delta$ submatrix of Π_{n+1}^{-1} given in (4.134).

4.19 Square Root Information Form of the Modified Bias-free Filter

The modified bias-free filter (4.104)–(4.107) uses the blending matrices E_t and U_t to handle the vector $\gamma = (\delta', \beta')'$, where δ is a diffuse random vector with $\text{Var}(\delta) = \sigma^2\Pi$ that models the unknown initial conditions and β is the bias vector. Assuming without loss of generality $\sigma^2 = 1$, the covariance matrix, P_1, of the initial state vector, $x_1 = W\beta + A\delta + x$, is

$$P_1 = A\Pi A' + \Omega. \tag{4.144}$$

If A has full column rank and the matrices Π and Ω are nonsingular, we can apply the matrix inversion lemma to get

$$P_1^{-1} = \Omega^{-1} - \Omega^{-1}A\left(\Pi^{-1} + A'\Omega^{-1}A\right)^{-1}A'\Omega^{-1}.$$

Also, if

$$\begin{aligned} P_1 &= A\Pi A' + B\Omega_s B' \\ &= R\begin{bmatrix}\Pi & 0\\ 0 & \Omega_s\end{bmatrix}R', \end{aligned} \tag{4.145}$$

where the matrices $R = [A, B]$ and Ω_s are nonsingular, then

$$P_1^{-1} = R^{-1'}\begin{bmatrix}\Pi^{-1} & 0\\ 0 & \Omega_s^{-1}\end{bmatrix}R^{-1}.$$

If δ is diffuse, then $\Pi^{-1} = 0$ and the previous formulae simplify. If, in addition, the information filter or the square root information filter can be applied, there is no need for the blending matrices in the modified bias-free filter to accommodate for δ. Only the regression vector, β, has to be taken care of. This means that there will be no need for collapsing, a term that will be defined later in Sect. 4.21. Collapsing refers to getting rid of the diffuse part, δ, in the modified bias-free filter recursions. It turns out that, after some iterations, collapsing is usually possible. That is, we can re-initialize the modified bias-free filter so that the new equations do not depend on δ.

In this section, we will derive the square root information filter of the modified bias-free filter (4.104)–(4.107). The following theorems give the details. We omit the proofs because they are similar to the corresponding ones in Sect. 4.7.

Theorem 4.39 (Square Root Information Filter for Measurement Update) *Consider the state space model (4.82) and (4.83), where P_1^{-1} exists as in (4.144) or in (4.145). Suppose that we have computed the modified bias-free filter quantities $\left(-U_t, x_{t|t-1}\right)$ and P_t. If a new measurement, Y_t, becomes available, then, if $R_t > 0$ and P_t^{-1} exists, the QR algorithm produces an orthogonal matrix Q_t such that*

$$Q_t' \left[\begin{array}{cc|c} P_t^{-1/2} & P_t^{-1/2}\left(-U_t, x_{t|t-1}\right) & 0 \\ -R_t^{-1/2} H_t & -R_t^{-1/2}\left(V_t, Y_t\right) & R_t^{-1/2} \end{array} \right] = \left[\begin{array}{cc|c} P_{t|t}^{-1/2} & P_{t|t}^{-1/2}\left(-U_{t|t}, x_{t|t}\right) & -P_{t|t}^{-1/2} K_{f,t} \\ 0 & -\Sigma_t^{-1/2}\left(E_t, e_t\right) & \Sigma_t^{-1/2} \end{array} \right],$$

where $K_{f,t} = P_t H_t' \Sigma_t^{-1}$. Here, the blending matrix U_t refers to β only, not to δ. That is, $\gamma = \beta$. Also, $\left(-U_{t|t}, x_{t|t}\right)$ satisfies the modified bias-free filter recursion $\left(-U_{t|t}, x_{t|t}\right) = \left(-U_t, x_{t|t-1}\right) + K_{f,t}\left(E_t, e_t\right)$ and can be obtained as $\left(-U_{t|t}, x_{t|t}\right) = P_{t|t}^{1/2}\left[P_{t|t}^{-1/2}\left(-U_{t|t}, x_{t|t}\right)\right]$.

Theorem 4.40 (Square Root Information Filter for Time Update) *Consider the state space model (4.82) and (4.83), where P_1^{-1} exists as in (4.144) or in (4.145). Suppose that we have computed the modified bias-free filter quantities $\left(-U_{t|t}, x_{t|t}\right)$ and $P_{t|t}$. If, without any further measurements, we wish to find $\left(-U_{t+1}, x_{t+1|t}\right)$ and P_{t+1}, then, if $R_t > 0$ and $(F_t^s)^{-1}$, $(Q_t^s)^{-1}$ and P_t^{-1} exist, the QR algorithm produces an orthogonal matrix Q_t such that*

$$Q_t' \left[\begin{array}{cc|c} (Q_t^s)^{-1/2} & 0 & (Q_t^s)^{-1/2} S_t R_t^{-1}\left(V_t, Y_t\right) \\ -P_{t|t}^{-1/2}(F_t^s)^{-1} G_t & P_{t|t}^{-1/2}(F_t^s)^{-1} & P_{t|t}^{-1/2}\left\{-U_{t|t} - (F_t^s)^{-1} W_t, x_{t|t}\right\} \end{array} \right]$$
$$= \left[\begin{array}{cc|c} (Q_t^r)^{-1/2} & -\widehat{K}_{b,t} & -\widehat{K}_{b,t}\left(-U_{t+1}, x_{t+1|t}\right) + (Q_t^r)^{-1/2} S_t R_t^{-1}\left(V_t, Y_t\right) \\ 0 & P_{t+1}^{-1/2} & P_{t+1}^{-1/2}\left(-U_{t+1}, x_{t+1|t}\right) \end{array} \right],$$

where

$$A_t = (F_t^s)^{-1'} P_{t|t}^{-1} (F_t^s)^{-1}, \quad Q_t^r = \left[(Q_t^s)^{-1} + G_t' A_t G_t\right]^{-1},$$

$\widehat{K}_{b,t} = (Q_t^r)^{1/2'} G_t' A_t$, *and* Q_t^s *and* F_t^s *are given by (4.19) and (4.22). Here, the blending matrix* U_t *refers to* β *only, not to* δ. *That is,* $\gamma = \beta$. *Also,* $\left(-U_{t+1}, x_{t+1|t}\right)$ *satisfies the modified bias-free filter recursion* $\left(-U_{t+1}, x_{t+1|t}\right) = (W_t, 0) + F_t \left(-U_{t|t}, x_{t|t}\right) + G_t S_t \Sigma_t^{-1} (E_t, e_t)$ *and can be obtained as* $\left(-U_{t+1}, x_{t+1|t}\right) = P_{t+1}^{1/2} \left[P_{t+1}^{-1/2} \left(-U_{t+1}, x_{t+1|t}\right)\right]$.

The following theorem gives the combined measurement and time updates.

Theorem 4.41 (Square Root Information Filter for the Combined Measurement and Time Updates) *Suppose that the process* $\{Y_t\}$ *follows the state space model (4.82) and (4.83), where* P_1^{-1} *exists as in (4.144) or in (4.145). Then, if* $R_t > 0, Q_t^s > 0$ *and the* F_t^s *are nonsingular, where* Q_t^s *and* F_t^s *are given by (4.19) and (4.22), the application of the QR algorithm yields an orthogonal matrix* O_t *such that*

$$O_t' \left[\begin{array}{cc|c} (Q_t^s)^{-1/2} & 0 & (Q_t^s)^{-1/2} S_t R_t^{-1} (V_t, Y_t) \\ -P_t^{-1/2} (F_t^s)^{-1} G_t & P_t^{-1/2} (F_t^s)^{-1} & P_t^{-1/2} \left\{-U_t - (F_t^s)^{-1} W_t, x_{t|t-1}\right\} \\ R_t^{-1/2} H_t (F_t^s)^{-1} G_t & -R_t^{-1/2} H_t (F_t^s)^{-1} & -R_t^{-1/2} (V_t, Y_t) \end{array}\right]$$
$$= \left[\begin{array}{cc|c} (Q_t^r)^{-1/2} & -\widehat{K}_{b,t} & -\widehat{K}_{b,t} \left(-U_{t+1}, x_{t+1|t}\right) + (Q_t^r)^{-1/2} S_t R_t^{-1} (V_t, Y_t) \\ 0 & P_{t+1}^{-1/2} & P_{t+1}^{-1/2} \left(-U_{t+1}, x_{t+1|t}\right) \\ 0 & 0 & -\Sigma_t^{-1/2} (E_t, e_t) \end{array}\right],$$

where

$$Q_t^r = \left[(Q_t^s)^{-1} + G_t' A_t G_t\right]^{-1}, \qquad A_t = (F_t^s)^{-1'} (P_t^{-1} + H_t' R_t^{-1} H_t)(F_t^s)^{-1},$$

and $\widehat{K}_{b,t} = (Q_t^r)^{1/2'} G_t' A_t$. *Here, the blending matrix* U_t *refers to* β *only, not to* δ. *That is,* $\gamma = \beta$.

4.20 The Two-stage Kalman Filter With Square Root Information Bias Filter

The TSKF consists of two decoupled filters, the modified bias-free filter (4.104)–(4.107) and the bias filter (4.108)–(4.110). The bias filter is the Kalman filter associated with the regression model $e = E\gamma + a$, where $e = (e_1', \ldots, e_n')', E = (E_1', \ldots, E_n')'$, $\mathrm{Var}(a) = \mathrm{diag}(\Sigma_1, \ldots, \Sigma_n)$, the state space form is given by (4.111) and (4.112), and the e_t, E_t, and Σ_t are given by the modified bias-free filter.

There are two aspects of the bias filter that deserve mention. The first one is that it cannot handle diffuse situations, in which it is assumed that $\Pi_1^{-1} = 0$, where Π_1 is the initial covariance matrix of γ. To incorporate this possibility into the bias filter, one should consider some information form for the filter. The second aspect has to do with numerical stability. It is standard practice to solve regression problems by using some numerically stable procedure, such as Householder transformations or Givens rotations (Kailath, Sayed, and Hassibi, 2000, pp. 53–54). For these reasons, we propose the use of the **TSKF with square root information bias filter** (hereafter TSKF–SRIBF), given by (4.104)–(4.107) and (4.141), that is,

$$\begin{aligned}(E_t, e_t) &= (0, V_t, Y_t) - H_t(-U_t, x_{t|t-1})\\ \Sigma_t &= H_tP_tH_t' + R_t, \quad K_t = (F_tP_tH_t' + G_tS_t)\Sigma_t^{-1}\\ (-U_{t+1}, x_{t+1|t}) &= (0, -W_t, 0) + F_t(-U_t, x_{t|t-1}) + K_t(E_t, e_t)\\ P_{t+1} &= (F_t - K_tH_t)P_tF_t' + (G_tQ_t - K_tS_t')G_t'\\ Q_t'\begin{bmatrix} \Pi_t^{-1/2} & \Pi_t^{-1/2}\hat{\gamma}_t \\ \Sigma_t^{-1/2}E_t & \Sigma_t^{-1/2}e_t \end{bmatrix} &= \begin{bmatrix} \Pi_{t+1}^{-1/2} & \Pi_{t+1}^{-1/2}\hat{\gamma}_{t+1} \\ 0 & \widehat{\Sigma}_t^{-1/2}\widehat{E}_t & \cdot \end{bmatrix},\end{aligned}$$

initialized with $(-U_1, x_{1|0}) = (-A, -W, a), P_1 = \Omega$ and $(\Pi_1^{-1/2}, \Pi_1^{-1/2}\hat{\gamma}_1) = (0, 0)$. Of course, we can use the fast recursions (4.143) instead of (4.141) to speed up the computations.

If there are reasons to believe that the covariances matrices of the modified bias-free filter are ill-conditioned, then a square root covariance implementation of this filter should be used. But this is something that depends on the nature of the problem at hand.

The TSKF with information form bias filter, given by (4.104)–(4.107) and (4.130), with the additional recursion (4.138), that is,

$$\begin{aligned}(E_t, e_t) &= (0, V_t, Y_t) - H_t(-U_t, x_{t|t-1})\\ \Sigma_t &= H_tP_tH_t' + R_t, \quad K_t = (F_tP_tH_t' + G_tS_t)\Sigma_t^{-1}\\ (-U_{t+1}, x_{t+1|t}) &= (0, -W_t, 0) + F_t(-U_t, x_{t|t-1}) + K_t(E_t, e_t)\\ P_{t+1} &= (F_t - K_tH_t)P_tF_t' + (G_tQ_t - K_tS_t')G_t'\\ \left(\Pi_{t+1}^{-1}, \Pi_{t+1}^{-1}\hat{\gamma}_{t+1}\right) &= \left(\Pi_t^{-1}, \Pi_t^{-1}\hat{\gamma}_t\right) + E_t'\Sigma_t^{-1}(E_t, e_t)\\ \mathrm{RSS}_{t+1}^{\mathrm{BFF}} &= \mathrm{RSS}_t^{\mathrm{BFF}} + e_t'\Sigma_t^{-1}e_t,\end{aligned}$$

initialized with $(-U_1, x_{1|0}) = (-A, -W, a), P_1 = \Omega, \left(\Pi_1^{-1}, \Pi_1^{-1}\hat{\gamma}_1\right) = (0, 0)$ and $\mathrm{RSS}_1^{\mathrm{BFF}} = 0$, was proposed by Jong (1991) and called by him the "Diffuse Kalman Filter." The concentrated diffuse log-likelihood (4.95) can be computed using the Diffuse Kalman Filter as (4.139) in Remark 4.15.

Apparently, in Jong (1991) and other subsequent publications by this and other authors (de Jong & Chu Chun Lin, 1994; Durbin & Koopman, 2012; Harvey, 1989; Jong & Chu Chun Lin, 2003) on Kalman filtering and smoothing under diffuse situations in the statistical literature these authors were not aware of Friedland (1969) and Ignagni (1981), or other articles about the same subject appeared in the engineering literature. In particular, nothing is mentioned in the previous references about the equivalence about the augmented-state Kalman filter and the TSKF and the precise relation between the two filters, something that, as we have seen, is vital for a full understanding of the issue of Kalman filtering under diffuse situations.

A square root information form for the TSKF was proposed by Bierman (1975). However, the state space model considered by this author is less general than (4.82) and (4.83) and the square root information form is used for both filters, the modified bias-free filter and the bias filter. Moreover, there is no mention in the previous article about the fact that these two filters can be decoupled.

4.20.1 Evaluation of the Concentrated Diffuse Log-likelihood with the TSKF–SRIBF

The TSKF–SRIBF can be used to evaluate the concentrated diffuse log-likelihood (4.95). By Remark 4.15, $(Y - X\hat{\gamma})'\Sigma^{-1}(Y - X\hat{\gamma}) = \sum_{t=1}^{n} \widehat{E}_t'\widehat{\Sigma}_t^{-1}\widehat{E}_t$. This last expression can be computed by the TSKF–SRIBF as $\sum_{t=1}^{n}\left(\widehat{\Sigma}_t^{-1/2}\widehat{E}_t\right)'$ $\left(\widehat{\Sigma}_t^{-1/2}\widehat{E}_t\right)$.

On the other hand, by (4.134) in Remark 4.13, $R'\Sigma^{-1}R$ is the submatrix of Π_{n+1}^{-1} formed with the first n_δ rows and the first n_δ columns. Letting

$$\Pi_{n+1}^{-1/2} = \begin{bmatrix} R_{dd,n+1} & R_{db,n+1} \\ & R_{bb,n+1} \end{bmatrix},$$

be the matrix given by (4.141) at the end of filtering ($t = n$), we see that $R'\Sigma^{-1}R = R_{dd,n+1}'R_{dd,n+1}$, where $R_{dd,n+1}$ is a square upper triangular matrix with dimension n_δ.

Putting it all together, we get the following expression for the concentrated diffuse log-likelihood (4.95),

$$\lambda_D(Y;\hat{\beta},\widehat{\sigma}^2) = \text{constant} - \frac{1}{2}\left\{(np - n_\delta)\ln\left[\sum_{t=1}^{n}\left(\widehat{\Sigma}_t^{-1/2}\widehat{E}_t\right)'\left(\widehat{\Sigma}_t^{-1/2}\widehat{E}_t\right)\right]\right.$$
$$\left. + \sum_{t=1}^{n}\ln|\Sigma_t| + \ln|R_{dd,n+1}|^2\right\}.$$

4.20.2 The Diffuse Likelihood When the Square Root Information Form of the Modified Bias-free Filter is Used

When the square root information form of the modified bias-free filter is used, the effect of the diffuse part, δ, is absorbed into the state vector from the beginning. Therefore, the square root information filter depends on β only and the concentrated diffuse log-likelihood (4.95) is

$$\lambda_D(Y;\hat{\beta},\widehat{\sigma}^2) = \text{constant} - \frac{1}{2}\left\{(np-n_\delta)\ln\left[\sum_{t=1}^{n}\left(\widehat{\Sigma}_t^{-1/2}\widehat{E}_t\right)'\left(\widehat{\Sigma}_t^{-1/2}\widehat{E}_t\right)\right]\right.$$
$$\left. + \sum_{t=1}^{n}\ln|\Sigma_t|\right\},$$

where, by remark (4.15), $(Y - X\hat{\beta})'\Sigma^{-1}(Y - X\hat{\beta}) = \sum_{t=1}^{n}\widehat{E}_t'\widehat{\Sigma}_t^{-1}\widehat{E}_t$. This last expression can be computed by the square root information filter as $\sum_{t=1}^{n}\left(\widehat{\Sigma}_t^{-1/2}\widehat{E}_t\right)'\left(\widehat{\Sigma}_t^{-1/2}\widehat{E}_t\right)$.

4.20.3 Forecasting With the TSKF–SRIBF

Denoting by $\hat{x}_{n+h|n}$, where $h \geq 1$, the orthogonal projection of x_{n+h} onto the sample $Y_{1:n}$, it is not difficult to show that h-period-ahead forecasts and their mean squared error, $\widehat{P}_{n+h}$, can be recursively obtained by

$$\begin{aligned}\hat{x}_{n+h|n} &= x_{n+h|n} + U_{n+h}\hat{\gamma}_{n+1}\\ \widehat{P}_{n+h} &= P_{n+h} + U_{n+h}\Pi_{n+1}U'_{n+h},\end{aligned}$$

where $\hat{\gamma}_{n+1}$ and Π_{n+1} are the GLS estimator of γ based on $Y_{1:n}$ and its MSE and for $h > 1$

$$\begin{aligned}(-U_{n+h}, x_{n+h|n}) &= (0, -W_{n+h-1}, 0) + F_{n+h-1}(-U_{n+h-1}, x_{n+h-1|n})\\ P_{n+h} &= F_{n+h-1}P_{n+h-1}F'_{n+h-1} + G_{n+h-1}Q_{n+h-1}G'_{n+h-1},\end{aligned}$$

initialized with $\hat{x}_{n+1|n} = x_{n+1|n} + U_{n+1}\hat{\gamma}_{n+1}$ and $\widehat{P}_{n+1} = P_{n+1} + U_{n+1}\Pi_{n+1}U'_{n+1}$.

The forecasts for Y_{n+h}, where $h \geq 1$, and the corresponding mean squared error matrices are given by

$$\begin{aligned} \widehat{Y}_{n+h|n} &= X_{n+h}\hat{\gamma}_{n+1} + H_{n+h}x_{n+h|n} \\ MSE(\widehat{Y}_{n+h|n}) &= S_{n+h} + X_{n+h}\Pi_{n+1}X'_{n+h}, \end{aligned}$$

where for $h > 1$

$$\begin{aligned} X_{n+h} &= (0, V_{n+h}) + H_{n+h}U_{n+h} \\ S_{n+h} &= H_{n+h}P_{n+h}H'_{n+h} + R_{n+h}. \end{aligned}$$

4.20.4 Smoothing With the TSKF–SRIBF Without Collapsing

When the TSKF–SRIBF is used without collapsing, the adjoint variable, λ_t, of the Bryson–Frazier formulae is augmented with a matrix, L_t, to get for $t = n, \ldots, 1$,

$$(L_t, \lambda_t) = H'_t\Sigma_t^{-1}(E_t, e_t) + F'_{p,t}(L_{t+1}, \lambda_{t+1}), \quad \Lambda_t = H'_t\Sigma_t^{-1}H_t + F'_{p,t}\Lambda_{t+1}F_{p,t},$$

where $F_{p,t} = F_t - K_tH_t$, initialized with $(L_{n+1}, \lambda_{n+1}) = (0, 0)$ and $\Lambda_{n+1} = 0$. Letting $\hat{x}_{t|n} = E(x_t|Y_{1:n})$ be the orthogonal projection of x_t onto $Y_{1:n}$ and $\widehat{P}_{t|n} = \mathrm{MSE}(\hat{x}_{t|n})$ for $t = n, \ldots, 1$, the fixed interval smoother is given by

$$\begin{aligned} \hat{x}_{t|n} &= \left[(-U_t, x_{t|t-1}) + P_t(L_t, \lambda_t)\right]\begin{pmatrix} -\hat{\gamma}_{n+1} \\ 1 \end{pmatrix} \\ \widehat{P}_{t|n} &= P_{t|n} + (P_tL_t - U_t)\Pi_{n+1}(P_tL_t - U_t)' \\ P_{t|n} &= P_t - P_t\Lambda_tP_t, \end{aligned} \tag{4.146}$$

where $\hat{\gamma}_{n+1}$ and Π_{n+1} are the GLS estimator of γ based on $Y_{1:n}$ and its MSE.

To smooth the disturbances of the state space model (4.85) and (4.86), the following recursions are used.

$$\begin{aligned} E(\epsilon_t|Y_{1:n}) &= \left[J'_t\Sigma_t^{-1}(E_t, e_t) + M'_t(L_{t+1}, \lambda_{t+1})\right]\begin{pmatrix} -\hat{\gamma}_{n+1} \\ 1 \end{pmatrix} \\ \mathrm{Var}(\epsilon_t|Y_{1:n}) &= \left[I - (J'_t\Sigma_t^{-1}J_t + M'_t\Lambda_{t+1}M_t)\right] \\ &\quad + (J'_t\Sigma_t^{-1}E_t + M'_tL_{t+1})\Pi_{n+1}(J'_t\Sigma_t^{-1}E_t + M'_tL_{t+1})', \end{aligned}$$

where $M_t = G_t - K_tJ_t$.

4.20.5 Square Root Information Smoothing With the Modified Bias-Free Filter

The following theorem gives the details of square root information smoothing for the modified bias-free filter. The proof is similar to that of Theorem 4.27 and is left as an exercise. See Problem 4.7.

Theorem 4.42 *Under the assumptions and with the notation of Theorem 4.41, suppose we apply the algorithm of that theorem in a forward pass. Then, the application of the QR algorithm in a backward pass yields an orthogonal matrix O_t such that*

$$O_t' \left[\begin{array}{c|c} P_{t+1|n}^{1/2'} \left[I - G_t (Q_t^r)^{1/2} \widehat{K}_{b,t} \right] (F_t^s)^{-1'} & P_{t+1|n}^{-1/2} \left(-U_{t+1|n}, x_{t+1|n} \right) \\ (Q_t^r)^{1/2'} G_t' (F_t^s)^{-1'} & \widehat{K}_{b,t} \left(-U_{t+1|n}, x_{t+1|n} \right) + (Q_t^r)^{-1/2} S_t R_t^{-1} (V_t, Y_t) \end{array} \right]$$
$$= \left[\begin{array}{c|c} P_{t|n}^{1/2'} & P_{t|n}^{-1/2} \left(-U_{t|n}, \hat{x}_{t|n} \right) \\ 0 & * \end{array} \right],$$

initialized with $P_{n+1|n}^{1/2} = P_{n+1}^{1/2}$ and $P_{n+1|n}^{-1/2} \left(-U_{n+1|n}, x_{n+1|n} \right) = \left(P_{n+1}^{-1/2} \right) \left(-U_{n+1}, x_{n+1|n} \right)$. In addition, the computation of $(Q_t^r)^{1/2}$ in the backward pass by inverting $(Q_t^r)^{-1/2}$, given by the forward pass, can be avoided if we incorporate a block column in the algorithm of Theorem 4.41 as follows

$$O_t' \left[\begin{array}{cc|cc} (Q_t^s)^{-1/2} & 0 & (Q_t^s)^{1/2'} & (Q_t^s)^{-1/2} S_t R_t^{-1} (V_t, Y_t) \\ -P_t^{-1/2} (F_t^s)^{-1} G_t & P_t^{-1/2} (F_t^s)^{-1} & 0 & P_t^{-1/2} \left\{ -U_t - (F_t^s)^{-1} W_t, x_{t|t-1} \right\} \\ R_t^{-1/2} H_t (F_t^s)^{-1} G_t & -R_t^{-1/2} H_t (F_t^s)^{-1} & 0 & -R_t^{-1/2} (V_t, Y_t) \end{array} \right]$$
$$= \left[\begin{array}{cc|cc} (Q_t^r)^{-1/2} & -\widehat{K}_{b,t} & (Q_t^r)^{1/2'} & -\widehat{K}_{b,t} \left(-U_{t+1}, x_{t+1|t} \right) + (Q_t^r)^{-1/2} S_t R_t^{-1} (V_t, Y_t) \\ 0 & P_{t+1}^{-1/2} & P_{t+1}^{1/2'} A_t G_t Q_t^r & P_{t+1}^{-1/2} \left(-U_{t+1}, x_{t+1|t} \right) \\ 0 & 0 & * & -\Sigma_t^{-1/2} (E_t, e_t) \end{array} \right].$$

Using Theorem 4.42, the square root form of the fixed interval smoother is as follows.

Step 1 In the forward pass, use the algorithm of Theorem 4.41, modified as in Theorem 4.42, to compute and store the quantities $(Q_t^r)^{1/2}, \widehat{K}_{b,t}$, $(Q_t^r)^{-1/2} S_t R_t^{-1} (V_t, Y_t)$, and $\widehat{K}_{b,t} \left(-U_{t+1}, x_{t+1|t} \right) = -\Big[-\widehat{K}_{b,t} \left(-U_{t+1}, x_{t+1|t} \right) + (Q_t^r)_t^{-1/2} R_t^{-1} (V_t, Y_t) \Big] + (Q_t^r)^{-1/2} S_t R_t^{-1} (V_t, Y_t)$.

Step 2 In the backward pass, compute $P_{t|n}^{1/2'}$ and $P_{t|n}^{-1/2} \left(-U_{t|n}, x_{t|n} \right)$ by means of the algorithm of Theorem 4.42. Finally, at the same time, compute recursively the fixed interval smoothing quantities

$$\left(-U_{t|n}, x_{t|n}\right) = P_{t|n}^{1/2}\left[P_{t|n}^{-1/2}\left(-U_{t|n}, x_{t|n}\right)\right]$$
$$P_{t|n} = P_{t|n}^{1/2}P_{t|n}^{1/2'}$$

and

$$\begin{aligned}\hat{x}_{t|n} &= (-U_{t|n}, x_{t|n})\begin{pmatrix}-\hat{\beta}_{n+1}\\ 1\end{pmatrix}\\ \widehat{P}_{t|n} &= P_{t|n} + U_t\Pi_{n+1}U_t'.\end{aligned}$$

4.21 Collapsing in the TSKF–SRIBF to Get Rid of the Nuisance Random Variables

The vector δ in (4.84) models the uncertainty in the initial conditions and can be considered as a vector of nuisance random variables. For this reason, it is convenient to get rid of it in the most efficient manner. It is to be stressed that the elements of β are parameters of interest and, therefore, they should be treated differently to those of δ.

Assume that β is fixed and known, that is $\mathrm{Var}(\beta) = 0$, in the state space model (4.82) and (4.83) and suppose that we apply the TSKF filter. Then, since β is not stochastic, it is easy to see that the Π_t matrices are of the form

$$\Pi_t = \begin{bmatrix}\Pi_{d,t} & 0\\ 0 & 0\end{bmatrix},$$

where $\Pi_{d,t}$ is a square matrix of dimension n_δ that will depend on the initial matrix, $\Pi_{d,1}$. Since δ is assumed to be diffuse, $\Pi_{d,1}$ is infinite and we should apply the TSKF–SRIBF. This filter uses the same modified bias-free filter as in the case in which both δ and β are unknown, but a simplified square root information bias filter that propagates $\Pi_{d,t}^{-1/2}$ only because β is not stochastic.

Assume that after a few iterations, $t = 1, 2, \ldots, k$, in this TSKF–SRIBF we get $\left(\Pi_{d,k+1}^{-1/2}, \Pi_{d,k+1}^{-1/2}\tilde{\delta}_{k+1}\right) = (R_{dd,k+1}, r_{d,k+1})$, where $\tilde{\delta}_{k+1}$ denotes the estimator of δ assuming β fixed and known, $\mathrm{MSE}(\tilde{\delta}_{k+1}) = \Pi_{d,k+1}$ and $R_{dd,k+1}$ is nonsingular upper triangular. Then, according to (4.115) and (4.116) of Theorem 4.35, we can write

$$\hat{x}_{k+1|t} = x_{k+1|t} + U_{d,k+1}\tilde{\delta}_{k+1} + U_{b,k+1}\beta, \quad \widehat{P}_{k+1} = P_{k+1} + U_{d,k+1}\Pi_{d,k+1}U_{d,k+1}', \tag{4.147}$$

where $\tilde{\delta}_{k+1} = R_{dd,k+1}^{-1} r_{d,k+1}$ and $\Pi_{d,k+1} = R_{dd,k+1}^{-1} R_{dd,k+1}^{-1'}$ are given by the algorithm of Theorem 4.38, $U_{k+1} = (U_{d,k+1}, U_{b,k+1})$ is given by the modified bias-free filter, and the partition is made according to $\gamma = (\delta', \beta')'$. It follows from this that we can use a perfectly known bias filter for $t = k+1, \ldots, n$,

$$E_t = Y_t - V_t\beta - H_t\hat{x}_{t|t-1}, \quad \Sigma_t = H_t\widehat{P}_tH_t' + R_t, \quad K_t = (F_t\widehat{P}_tH_t' + G_tS_t)\Sigma_t^{-1}$$
$$\hat{x}_{t+1|t} = W_t\beta + F_t\hat{x}_{t|t-1} + K_tE_t, \quad \widehat{P}_{t+1} = (F_t - K_tH_t)\widehat{P}_tF_t' + (G_tQ_t - K_tS_t')G_t',$$

with (4.147) as initial conditions because there is no unspecified part in (4.147). That is, we have eliminated the nuisance random vector δ in the new filter. This process is known as a **collapse of the TSKF–SRIBF**. We will not need to update $\tilde{\delta}_{k+1}$ in the new filter because the equations (4.82) and (4.83) do not depend on δ when the initial conditions are completely specified. We can see the application of the TSKF–SRIBF for $t = 1, 2, \ldots, k$ and the subsequent collapse as a way to construct initial conditions for a Kalman filter that does not depend on the unspecified part δ.

If β is unknown in the previous situation, we can proceed in a similar manner. However, the collapsed filter will not be a perfectly known bias filter in this case, but a TSKF–SRIBF of reduced dimension. More specifically, suppose that we apply the TSKF–SRIBF for $t = 1, \ldots, k$ assuming that both δ and β are unknown and let

$$\Pi_{k+1}^{-1/2} = \begin{bmatrix} R_{dd,k+1} & R_{db,k+1} \\ & R_{bb,k+1} \end{bmatrix}, \quad \Pi_{k+1}^{-1/2}\hat{\gamma}_{k+1} = \begin{bmatrix} r_{d,k+1} \\ r_{b,k+1} \end{bmatrix}, \tag{4.148}$$

where, defining $E_t = (E_{d,t}, E_{b,t})$ conforming to $\gamma = (\delta', \beta')'$ and considering (4.141), we see that $R_{dd,k+1}$ and $r_{d,k+1}$ coincide with those considered earlier in this section when β was assumed to be fixed and known. Then,

$$\hat{\gamma}_{k+1} = \begin{bmatrix} \hat{\delta}_{k+1} \\ \hat{\beta}_{k+1} \end{bmatrix} = \begin{bmatrix} R_{dd,k+1}^{-1} & -R_{dd,k+1}^{-1}R_{db,k+1}R_{bb,k+1}^{-} \\ & R_{bb,k+1}^{-} \end{bmatrix} \begin{bmatrix} r_{d,k+1} \\ r_{b,k+1} \end{bmatrix},$$
$$\Pi_{k+1} = \begin{bmatrix} R_{dd,k+1}^{-1} & -R_{dd,k+1}^{-1}R_{db,k+1}R_{bb,k+1}^{-} \\ & R_{bb,k+1}^{-} \end{bmatrix} \begin{bmatrix} R_{dd,k+1}^{-1'} & \\ -R_{bb,k+1}^{-'}R_{db,k+1}'R_{dd,k+1}^{-1'} & R_{bb,k+1}^{-'} \end{bmatrix}$$

and

$$\hat{\delta}_{k+1} = \tilde{\delta}_{k+1} - S_{db,k+1}\hat{\beta}_{k+1}, \quad \hat{\beta}_{k+1} = R_{bb,k+1}^{-}r_{b,k+1} \tag{4.149}$$
$$MSE(\hat{\delta}_{k+1}) = \Pi_{d,k+1} + S_{db,k+1}MSE(\hat{\beta}_{k+1})S_{db,k+1}', \tag{4.150}$$

where $\tilde{\delta}_{k+1}$ and $\Pi_{d,k+1}$ are those of (4.147), $S_{db,k+1} = R_{dd,k+1}^{-1}R_{db,k+1}$, $MSE(\hat{\beta}_{k+1}) = R_{bb,k+1}^{-}R_{bb,k+1}^{-'}$ and $R_{bb,k+1}^{-}$ is a generalized inverse of $R_{bb,k+1}$. By (4.115) and (4.116) of Theorem 4.35, if we use (4.149) and (4.150), it is obtained that

$$\hat{x}_{k+1|k} = x_{k+1|k} + U_{d,k+1}\tilde{\delta}_{k+1} + U_{db,k+1}\hat{\beta}_{k+1}$$
$$\widehat{P}_{k+1} = P_{k+1} + U_{d,k+1}\Pi_{d,k+1}U'_{d,k+1} + U_{db,k+1}MSE(\hat{\beta}_{k+1})U'_{db,k+1},$$

where $U_{db,k+1} = U_{b,k+1} - U_{d,k+1}S_{db,k+1}$, $U_{k+1} = (U_{d,k+1}, U_{b,k+1})$ and the partition is as in (4.147). This will allow us to define for $t = k+1, \ldots, n$ a new TSKF–SRIBF of reduced dimension. We summarize the result in the following theorem.

Theorem 4.43 (Collapsing of the TSKF–SRIBF) *Assume that after applying the TSKF–SRIBF for $t = 1, \ldots, k, k < n$, the matrix $R_{dd,k+1}$ in (4.148) is nonsingular. Then, a new TSKF–SRIBF of reduced dimension can be used for $t = k+1, \ldots, n$ in which γ is redefined as $\gamma = \beta$, the modified bias-free Kalman filter is initialized with $(-U_{k+1}, x_{k+1|k})$ and P_{k+1} redefined to*

$$(-U_{k+1}, x_{k+1|k}) = (-U_{db,k+1}, x_{k+1|k} + U_{d,k+1}\tilde{\delta}_{k+1}) \tag{4.151}$$
$$P_{k+1} = P_{k+1} + U_{d,k+1}\Pi_{d,k+1}U'_{d,k+1}, \tag{4.152}$$

and the square root information bias filter is initialized with $\left(\Pi_{k+1}^{-1/2}, \Pi_{k+1}^{-1/2}\hat{\gamma}_{k+1}\right)$ redefined to

$$\left(\Pi_{k+1}^{-1/2}, \Pi_{k+1}^{-1/2}\hat{\beta}_{k+1}\right) = (R_{bb,k+1}, r_{b,k+1}), \tag{4.153}$$

where $\tilde{\delta}_{k+1} = R^{-1}_{dd,k+1}r_{d,k+1}$ is the estimator of δ assuming β fixed and known, $MSE(\tilde{\delta}_{k+1}) = \Pi_{d,k+1} = R^{-1}_{dd,k+1}R^{-1'}_{dd,k+1}$, $U_{db,k+1} = U_{b,k+1} - U_{d,k+1}S_{db,k+1}$, $S_{db,k+1} = R^{-1}_{dd,k+1}R_{db,k+1}$, $U_{k+1} = (U_{d,k+1}, U_{b,k+1})$ and the partition is made conforming to $\gamma = (\delta', \beta')'$.

Remark 4.18 Instead of collapsing δ completely as in Theorem 4.43, we can apply this theorem sequentially, collapsing some part of δ each time, until δ disappears. In this way, the calculations involved in collapsing are simplified. This procedure is illustrated in Example 4.2 in the next section. ◇

4.21.1 Examples of Collapsing

Example 4.1 (Continued) Suppose the observed series is $Y = (Y_1, Y_2, Y_3)'$. The TSKF–SRIBF is initialized with $(-U_1, x_{1|0}) = (-1, 0)$ and $P_1 = 1$. The first iteration yields

$$(E_1, e_1) = (1, Y_1), \quad \Sigma_1 = 3, \quad K_1 = 1/3$$
$$(-U_2, x_{2|1}) = (-2/3, Y_1/3), \quad P_2 = 5/3,$$

and $\left(\Pi_2^{-1/2}, \Pi_2^{-1/2}\hat{\delta}_2\right) = \left(1/\sqrt{3}, Y_1/\sqrt{3}\right)$. Then, $\hat{\delta}_2 = Y_1$ and $\Pi_2 = \text{MSE}(\hat{\delta}_2) = 3$, and there is a collapse to the Kalman filter at $t = 2$ with $\hat{x}_{2|1} = Y_1$ and $P_2 = 3$. The second iteration yields

$$e_2 = Y_2 - Y_1, \quad \Sigma_2 = 5, \quad K_2 = 3/5$$

$$\hat{x}_{3|2} = 3Y_2/5 + 2Y_1/5, \quad P_3 = 11/5.$$

Finally, the third iteration produces

$$e_3 = Y_3 - 3Y_2/5 - 2Y_1/5, \quad \Sigma_3 = 21/5, \quad K_3 = 11/21.$$

◇

Example 4.2 (Continued) The TSKF–SRIBF is initialized with

$$(-U_1, x_{1|0}) = \begin{bmatrix} 1 & -2 & 0 \\ 2 & -3 & 0 \end{bmatrix}, \quad P_1 = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}.$$

The first update yields

$$(E_1, e_1) = (-1, 2, Y_1), \quad \Sigma_1 = 1 + \lambda, \quad K_1 = (2, 3)'\Sigma_1^{-1},$$

and

$$\left(\Pi_2^{-1/2}, \Pi_2^{-1/2}\hat{\delta}_2\right) = \begin{bmatrix} -1 & 2 & Y_1 \\ 0 & 0 & 0 \end{bmatrix} \frac{1}{\sqrt{1+\lambda}}. \tag{4.154}$$

We can proceed sequentially in collapsing. That is, instead of collapsing the whole of $\delta = (\delta_1, \delta_2)'$, we collapse first δ_1 and then δ_2. To this end, we first redefine $\gamma = (\delta, \beta)' = (\delta_1, \delta_2)'$ and apply Theorem 4.43 to collapse $\delta = \delta_1$. Then, we redefine $\delta = \beta$ and apply Theorem 4.43 again to collapse $\delta = \delta_2$. According to this approach, it follows from (4.154) that $\tilde{\delta}_2 = -Y_1$, $\Pi_{d,2} = \text{MSE}(\tilde{\delta}_2) = \Sigma_1$, and there is a collapse at $t = 2$ with

$$(-U_2, x_{2|1}) = \begin{bmatrix} 1 & 2Y_1 \\ 2 & 3Y_1 \end{bmatrix}, \quad P_2 = \begin{bmatrix} 1 + 4\lambda & 2 + 6\lambda \\ 2 + 6\lambda & 4 + 9\lambda \end{bmatrix}.$$

We can consider that a new TSKF–SRIBF is initialized by these equations. Thus, the second update of the former filter coincides with the first update of the latter and gives

$$(E_2, e_2) = (-1, Y_2 - 2Y_1), \quad \Sigma_2 = 1 + 5\lambda, \quad K_2 = (6\lambda + 2, 8\lambda + 3)'\Sigma_2^{-1},$$

and $\left(\Pi_3^{-1/2}, \Pi_3^{-1/2}\hat{\delta}_3\right) = (-1, Y_2 - 2Y_1)/\sqrt{1+5\lambda}$. Then, applying again Theorem 4.43 yields $\hat{\delta}_3 = -Y_2 + 2Y_1$ and $\Pi_3 = \text{MSE}(\hat{\delta}_3) = \Sigma_2$, and a complete collapse to the Kalman filter is possible at $t = 3$. The Kalman filter is initialized with

$$\hat{x}_{3|2} = \begin{bmatrix} 2Y_2 - Y_1 \\ 3Y_2 - 2Y_1 \end{bmatrix}, \quad P_3 = \begin{bmatrix} 1+5\lambda & 2+8\lambda \\ 2+8\lambda & 4+13\lambda \end{bmatrix}.$$

◇

Example 4.3 (Continued) The TSKF–SRIBF is initialized with

$$(-U_1, x_{1|0}) = \begin{bmatrix} -1 & 0 & 0 \\ -1 & 0 & 0 \end{bmatrix}, \quad P_1 = \frac{1}{1-\phi^2}\begin{bmatrix} 1 & 1-\phi \\ 1-\phi & (1-\phi)^2 \end{bmatrix}.$$

The first update yields

$$(E_1, e_1) = (1, 0, Y_1), \quad \Sigma_1 = 1/(1-\phi^2), \quad K_1 = \left(1-\phi, \phi + (1-\phi)^2\right)',$$

and

$$\left(\Pi_2^{-1/2}, \Pi_2^{-1/2}\hat{\delta}_2\right) = \begin{bmatrix} 1 & 0 & Y_1 \\ 0 & 0 & 0 \end{bmatrix}\sqrt{1-\phi^2}.$$

Then, $\left(\Pi_{d,2}^{-1/2}, \Pi_{d,2}^{-1/2}\tilde{\delta}_2\right) = (1, Y_1)\sqrt{1-\phi^2}$, $\tilde{\delta}_2 = Y_1$, $\Pi_{d,2} = \text{MSE}(\tilde{\delta}_2) = \Sigma_1$, and there is a collapse at $t = 2$ with

$$(-U_2, x_{2|1}) = \begin{bmatrix} 0 & Y_1 \\ 0 & Y_1 \end{bmatrix}, \quad P_2 = \frac{1}{1-\phi^2}\begin{bmatrix} 1 & 1-\phi \\ 1-\phi & (1-\phi)^2 \end{bmatrix}.$$

The TSKF–SRIBF does not depend on δ for $t \geq 2$. A total collapse to the Kalman filter cannot be done because there is a β part in the model. It is not difficult to verify that

$$(E_2, e_2) = (1, Y_2 - Y_1), \quad \Sigma_2 = 1/(1-\phi^2), \quad K_2 = \left(1-\phi, \phi + (1-\phi)^2\right)',$$

and, for $t > 2$,

$$(E_t, e_t) = (1+\phi, \nabla Y_t + \phi \nabla Y_{t-1}), \quad \Sigma_t = 1, \quad K_t = \left(1-\phi, \phi + (1-\phi)^2\right)'.$$

◇

4.21.2 Evaluation of the Concentrated Diffuse Log-likelihood With the TSKF–SRIBF Under Collapsing

Under the assumptions and with the notation of the previous theorem, if the TSKF–SRIBF is applied and a collapse takes place at $t = k$, the concentrated diffuse log-likelihood (4.95) is

$$\lambda_D(Y; \hat{\beta}, \widehat{\sigma}^2) = \text{constant} - \frac{1}{2}\left\{ (np - n_\delta) \ln\left[\sum_{t=1}^{n} \left(\widehat{\Sigma}_t^{-1/2} \widehat{E}_t \right)' \left(\widehat{\Sigma}_t^{-1/2} \widehat{E}_t \right) \right] + \sum_{t=1}^{n} \ln|\Sigma_t| + \ln|R_{dd,k+1}|^2 \right\}.$$

To see this, consider that the concentrated log-likelihood can be expressed as

$$\lambda_D(Y; \hat{\beta}, \widehat{\sigma}^2) = \lambda_D(Y_1, Y_2, \ldots, Y_k; \hat{\beta}_{k+1}, \widehat{\sigma}^2_{k+1}) + \lambda_D(Y_{k+1}, \ldots, Y_n | Y_1, \ldots, Y_k; \hat{\beta}_{n+1}, \widehat{\sigma}^2_{n+1}).$$

4.21.3 Smoothing with the TSKF–SRIBF Under Collapsing

If the TSKF–SRIBF is used for $t = 1, \ldots, n$ and a collapse takes place at $t = k, k < n$, the augmented part is reduced or eliminated at $t = k + 1$. If $\beta \neq 0$, for $t = n, \ldots, k + 1$ the augmented Bryson–Frazier recursions can be used

$$(L_{b,t}, \lambda_t) = H_t' \Sigma_t^{-1}(E_t, e_t) + F_{p,t}'(L_{b,t+1}, \lambda_{t+1}), \quad \Lambda_t = H_t' \Sigma_t^{-1} H_t + F_{p,t}' \Lambda_{t+1} F_{p,t}, \tag{4.155}$$

where $F_{p,t} = F_t - K_t H_t$, initialized with $(L_{b,n+1}, \lambda_{n+1}) = (0, 0)$ and $\Lambda_{n+1} = 0$. If $\beta = 0$, the augmented part, $L_{b,t}$ disappears completely after collapsing and we would obtain the usual Bryson–Frazier formulae for smoothing for $t = n, \ldots, k+1$.

4.21.3.1 The Fixed-Point Smoother Under Collapsing

The details of the fixed point smoother under collapsing are given by the following theorem.

Theorem 4.44 *Suppose that the TSKF–SRIBF is applied for $t = 1, \ldots, n$ and let $j \leq n$. Then, the recursion for $x_{j|t}$ in the fixed point smoother is augmented with a* *matrix, C_t, for $t = j, j+1, \ldots, n$ to get*

$$K_j^t = P_j^t H_t' \Sigma_t^{-1}, \qquad (-C_t, x_{j|t}) = (-C_{t-1}, x_{j|t-1}) + K_j^t (E_t, e_t)$$
$$P_{j|t} = P_{j|t-1} - K_j^t H_t P_j^{t\prime}, \qquad P_j^{t+1} = P_j^t F_{p,t}',$$

initialized with $P_j^j = P_j$, $P_{j|j-1} = P_j$ and $(-C_{j-1}, x_{j|j-1}) = (-U_j, x_{j|j-1})$.

If there is a collapse at time k and $j \leq k$, then, with the notation of Theorem 4.43, $(-C_k, x_{j|k})$, $P_{j|k}$, and P_j^{k+1} are redefined to

$$(-C_k, x_{j|k}) = (-C_{db,k}, x_{j|k} + C_{d,k}\tilde{\delta}_{k+1}), \quad P_{j|k} = P_{j|k} + C_{d,k}\Pi_{d,k+1}C_{d,k}'$$
$$P_j^{k+1} = P_j^{k+1} + C_{d,k}\Pi_{d,k+1}U_{d,k+1}'$$
$$= P_j F_{p,j}^{k\prime} + C_{d,k}\Pi_{d,k+1}U_{d,k+1}',$$

respectively, where $F_{p,j}^{k\prime} = F_{p,j}' \cdots F_{p,k}'$, $C_{db,k} = C_{b,k} - C_{d,k}S_{db,k+1}$, $C_k = (C_{d,k}, C_{b,k})$ before collapsing and the partition is made conforming to $\gamma = (\delta', \beta')'$.

If $\beta \neq 0$, the estimator $\hat{x}_{j|n}$ of x_j, $j \leq n$, $n > k$, based on $Y_{1:n}$ and its mean squared error, $\widehat{P}_{j|n}$, are given by

$$\hat{x}_{j|n} = x_{j|n} + C_n \hat{\beta}_{n+1}, \quad \widehat{P}_{j|n} = P_{j|n} + C_n \Pi_{n+1} C_n',$$

where $\hat{\beta}_{n+1}$ and Π_{n+1} are the GLS estimator of β based on $Y_{1:n}$ and its MSE. If $\beta = 0$, the augmented part disappears after collapsing and $\hat{x}_{j|n}$ and $\widehat{P}_{j|n}$ are given by the ordinary fixed point smoother, where $P_{j|n} = \widehat{P}_{j|n}$.

Proof We proceed as Anderson & Moore (2012) do to prove the fixed point smoother. To smooth x_j, $1 \leq j \leq k \leq n$, define for $t \geq j$ an augmented state vector $x_t^a = (x_t', x_j')'$ so that the associated state space model is

$$x_{t+1}^a = \begin{bmatrix} F_t & \\ & I \end{bmatrix} x_t^a + \begin{bmatrix} G_t \\ 0 \end{bmatrix} \epsilon_t$$
$$Y_t = (H_t, 0)x_t^a + J_t \epsilon_t,$$

where we assume $\beta = 0$ for simplicity. The proof for the general case is similar. Instead of using the notation $(-U_t, x_{t|t-1})$ in the TSKF–SRIBF, we use the simpler notation X_t. Applying the TSKF–SRIBF to the previous augmented state space model for $t = j, j+1, \ldots, n$ with starting conditions

$$X_j^a = \begin{bmatrix} X_j \\ X_j \end{bmatrix}, \qquad P_j^a = \begin{bmatrix} P_j & P_j \\ P_j & P_j \end{bmatrix}$$

yields in the first iteration (E_j, e_j), Σ_j and

$$K_j^a = \begin{bmatrix} K_j \\ K_j^j \end{bmatrix}, \quad X_{j+1}^a = \begin{bmatrix} X_{j+1} \\ X_{j|j} \end{bmatrix}, \quad P_{j+1}^a = \begin{bmatrix} P_{j+1} & P_j^{j+1'} \\ P_j^{j+1} & P_{j|j} \end{bmatrix},$$

where $K_j^j = P_j H_j' \Sigma_j^{-1}, X_{j|j} = X_j + K_j^j(E_j, e_j) = (-C_j, x_{j|j}), P_j^{j+1} = P_j F_{p,j}', P_{j|j} = P_j - K_j^j H_j P_j$. For $t = j+1, \ldots, n$, the TSKF–SRIBF produces (E_t, e_t), Σ_t and

$$K_t^a = \begin{bmatrix} K_t \\ K_j^t \end{bmatrix}, \quad X_{t+1}^a = \begin{bmatrix} X_{t+1} \\ X_{j|t} \end{bmatrix}, \quad P_{t+1}^a = \begin{bmatrix} P_{t+1} & P_j^{t+1'} \\ P_j^{t+1} & P_{j|t} \end{bmatrix},$$

where $K_j^t = P_j^t H_t' \Sigma_t^{-1}, X_{j|t} = X_{j|t-1} + K_j^t(E_t, e_t), P_{j|t} = P_{j|t-1} - K_j^t H_t P_j^{t'}$, and $P_j^{t+1} = P_j^t F_{p,t}'$. If there is a collapse at $t = k$ and $j \leq k \leq n$, then $X_{j|k}, P_{j|k}$ and P_j^{k+1} are redefined as $X_{j|k} = (-C_{db,k}, x_{j|k} + C_{d,k}\tilde{\delta}_{k+1}), P_{j|k} = P_{j|k} + C_{d,k} MSE(\tilde{\delta}_{k+1}) C_{d,k}'$ and $P_j^{k+1} = P_j^{k+1} + C_{d,k} MSE(\tilde{\delta}_{k+1}) U_{d,k+1}'$, where $C_{db,k} = C_{b,k} - C_{d,k} S_{db,k}$ and $C_k = (C_{d,k}, C_{b,k})$. The result now follows from Theorem 4.43 by letting $j \leq k$ fixed and iterating for $t = j, \ldots, n$. □

Remark 4.19 It is to be noted that in the fixed point smoother only the matrix P_j^t has to be stored in addition to the usual quantities required for fixed interval smoothing. If we are interested in estimating lx_t, where l is some $r \times p$ matrix with $r < p$, then only lP_j^t has to be stored. ◇

The following theorem gives the relation between the fixed point and the fixed interval smoother. We omit its proof because it is similar to that of the previous theorem.

Theorem 4.45 *Under the assumptions and with the notation of Theorem 4.44, suppose that $\beta \neq 0$ and there is a collapse at time $t = k, k < n$. Then, the estimator, $\hat{x}_{j|n}$, of $x_j, j \leq k$, based on $Y_{1:n}$ and its mean squared error, $\widehat{P}_{j|n}$, can be computed as*

$$\hat{x}_{j|n} = \left[(-C_k, x_{j|k}) + P_j^{k+1}(L_{b,k+1}, \lambda_{k+1}) \right] \begin{bmatrix} -\hat{\beta}_{n+1} \\ 1 \end{bmatrix} \tag{4.156}$$

$$\widehat{P}_{j|n} = P_{j|k} - P_j^{k+1} \Lambda_{k+1} P_j^{k+1'} + \left[P_j^{k+1} L_{b,k+1} - C_k \right] \Pi_{n+1} \left[P_j^{k+1} L_{b,k+1} - C_k \right]', \tag{4.157}$$

where $P_j^{k+1}, (-C_k, x_{j|k})$ and $P_{j|k}$ are given by the fixed point smoother applied for $t = j, j+1, \ldots, k$, Λ_{k+1} are given by (4.155) and $\hat{\beta}_{n+1}$ and Π_{n+1} are the GLS estimator of β based on $Y_{1:n}$ and its MSE. If $\beta = 0$, the augmented part disappears *after collapsing and the recursions simplify to*

$$\hat{x}_{j|n} = \hat{x}_{j|k} + P_j^{k+1}\lambda_{k+1}, \tag{4.158}$$

$$\widehat{P}_{j|n} = P_{j|k} - P_j^{k+1}\Lambda_{k+1}P_j^{k+1'}, \tag{4.159}$$

4.21.3.2 The Fixed-Interval Smoother Under Collapsing

The following theorem gives the details of the fixed interval smoothing under collapsing.

Theorem 4.46 *Suppose that the TSKF–SRIBF is applied for $t = 1, \dots, n$ and that a collapse takes place at $t = k, k < n$. If $\beta \neq 0$, for $t = n, \dots, k+1$ the augmented Bryson–Frazier recursions (4.155) are used and the fixed interval smoother is given by*

$$\begin{aligned}
\hat{x}_{t|n} &= \left[(-U_{b,t}, x_{t|t-1}) + P_t(L_{b,t}, \lambda_t)\right]\begin{bmatrix} -\hat{\beta}_{n+1} \\ 1 \end{bmatrix} \\
\widehat{P}_{t|n} &= P_t - P_t\Lambda_t P_t + (P_t L_{b,t} - U_{b,t})\Pi_{n+1}(P_t L_{b,t} - U_{b,t})',
\end{aligned}$$

where $\hat{\beta}_{n+1}$ and Π_{n+1} are the GLS estimator of β based on $Y_{1:n}$ and its MSE. If $\beta = 0$, the Bryson–Frazier recursions (4.33) and the fixed interval smoother (4.34) are used for $t = n, \dots, k+1$.

For $t = k, \dots, 1$, the recursions for Λ_t are continued and the recursions for $(L_{b,t}, \lambda_t)$ if $\beta \neq 0$ or λ_t if $\beta = 0$ are also continued, augmented with $L_{d,t}$, to get

$$(L_t, \lambda_t) = H_t'\Sigma_t^{-1}(E_t, e_t) + F_{p,t}'(L_{t+1}, \lambda_{t+1}), \tag{4.160}$$

where $L_t = (L_{d,t}, L_{b,t})$ or $L_t = L_{d,t}$, respectively, and $L_{d,k+1} = 0$. Let $A_t = P_t L_t - U_t$ and $B_t = L_t - \Lambda_t U_t$, where (L_t, λ_t) and Λ_t are given by (4.160) and U_t is given by the TSKF–SRIBF before collapsing, and let $\hat{\gamma}_{n+1} = (\hat{\delta}_{n+1}', \hat{\beta}_{n+1}')'$ and Ψ_{n+1} be the GLS estimator of γ based on $Y_{1:n}$ and its MSE. If $\beta \neq 0$, the fixed interval smoothing is given for $t = k, \dots, 1$ by the recursions

$$\begin{aligned}
\hat{x}_{t|n} &= x_{t|t-1} + P_t\lambda_t - A_t\hat{\gamma}_{n+1} \\
\widehat{P}_{t|n} &= P_t - P_t\Lambda_t P_t - A_t\Phi_{k+1}B_t'P_t - P_t B_t\Phi_{k+1}'A_t' + A_t\Psi_{n+1}A_t',
\end{aligned}$$

where, with the notation of Theorem 4.43,

$$\hat{\delta}_{n+1} = \tilde{\delta}_{k+1} + \Pi_{d,k+1}U_{d,k+1}'\lambda_{k+1} - T_{db,k+1}\hat{\beta}_{n+1}$$

$$T_{db,k+1} = S_{db,k+1} + \Pi_{d,k+1}U_{d,k+1}'L_{b,k+1}$$

$$\Phi_{k+1} = \begin{bmatrix} \Pi_{d,k+1} & 0 \\ 0 & 0 \end{bmatrix}$$

$$\Psi_{n+1} = \begin{bmatrix} Z_{k+1} + T_{db,k+1}\Pi_{n+1}T'_{db,k+1} & -T_{db,kl+1}\Pi_{n+1} \\ -\Pi_{n+1}T'_{db,kl+1} & \Pi_{n+1} \end{bmatrix}$$

and $Z_{k+1} = \Pi_{d,k+1} - \Pi_{d,k+1}U'_{d,k+1}\Lambda_{k+1}U_{d,k+1}\Pi_{d,k+1}$. *If* $\beta = 0$*, the fixed interval smoothing is given for* $t = k, \ldots, 1$ *by the recursions*

$$\begin{aligned} \hat{x}_{t|n} &= x_{t|t-1} + P_t\lambda_t - A_t\hat{\delta}_{n+1} \\ \widehat{P}_{t|n} &= P_t - P_t\Lambda_tP_t - A_t\Pi_{k+1}B'_tP_t - P_tB_t\Pi_{k+1}A'_t + A_t\Psi_{n+1}A'_t, \end{aligned}$$

where

$$\hat{\delta}_{n+1} = \tilde{\delta}_{k+1} + \Pi_{k+1}U'_{k+1}\lambda_{k+1}$$

$$\Psi_{n+1} = \Pi_{k+1} - \Pi_{k+1}U'_{k+1}\Lambda_{k+1}U_{k+1}\Pi_{k+1}.$$

Proof Suppose first that $\beta = 0$. To prove the formula for $\hat{\delta}_{n+1}$ and Ψ_{n+1}, define an augmented state vector $x^a_t = (x'_t, \delta')'$ so that the associated state space model is

$$x^a_{t+1} = \begin{bmatrix} F_t & \\ & I \end{bmatrix} x^a_t + \begin{bmatrix} G_t \\ 0 \end{bmatrix} \epsilon_t$$

$$Y_t = (H_t, 0)x^a_t + J_t\epsilon_t,$$

where we assume that $a = 0$ for simplicity. As in the proof of Theorem 4.44, instead of using the notation $(-U_t, x_{t|t-1})$ in the TSKF–SRIBF, we use the simpler notation X_t. Applying the TSKF–SRIBF to the previous augmented state space model for $t = 1, 2, \ldots, n$ with starting conditions

$$X^a_1 = \begin{bmatrix} -A & 0 \\ -I & 0 \end{bmatrix}, \qquad P^a_1 = \begin{bmatrix} \Omega & 0 \\ 0 & 0 \end{bmatrix}$$

yields (E_t, e_t), Σ_t and

$$K^a_t = \begin{bmatrix} K_t \\ K^t_1 \end{bmatrix}, \quad X^a_{t+1} = \begin{bmatrix} X_{t+1} \\ X_{1|t} \end{bmatrix}, \quad P^a_{t+1} = \begin{bmatrix} P_{t+1} & P^{t+1'}_1 \\ P^{t+1}_1 & P_{1|t} \end{bmatrix},$$

where, proceeding as in the proof of Theorem 4.44, it is not difficult to show that the following fixed point smoothing algorithm holds for $t = 1, 2, \ldots, n$

$$K^t_1 = P^t_1H'_t\Sigma^{-1}_t, \qquad X_{1|t} = X_{1|t-1} + K^t_1(E_t, e_t)$$

$$P_{1|t} = P_{1|t-1} - K^t_1H_tP^{t'}_1, \qquad P^{t+1}_1 = P^t_1F'_{p,t},$$

initialized with $P_1^1 = 0, P_{1|0} = 0$ and $X_{1|0} = (-I, 0)$. Since $P_1^t = 0, P_{1|t} = 0$ and $X_{1|t} = (-I, 0)$ for $t = 1, \ldots, n$, if there is a collapse at time $k, X_{1|k}, P_{1|k}$ and P_1^{k+1} are redefined to

$$X_{1|k} = \tilde{\delta}_{k+1}, \quad P_{1|k} = \Pi_{k+1}$$
$$P_1^{k+1} = \Pi_{k+1} U'_{k+1}.$$

Iterating and using (4.158) and (4.159), the result for $\hat{\delta}_{n+1}$ and Ψ_{n+1} is obtained.

To prove the formula for the fixed interval smoother, consider that, by (4.158) and (4.159), we can write

$$\hat{x}_{t|n} = \hat{x}_{t|k} + P_t^{k+1} \lambda_{k+1},$$
$$\widehat{P}_{t|n} = P_{t|k} - P_t^{k+1} \Lambda_{k+1} P_t^{k+1'},$$

where $P_t^{k+1} = P_t F_{p,t}^{k'} + C_k \Pi_{k+1} U'_{k+1}$. If we define for $t = k, \ldots, 1$ the backwards recursion

$$(L_t, \bar{\lambda}_t) = H'_t \Sigma_t^{-1} (E_t, e_t) + F'_{p,t} (L_{t+1}, \bar{\lambda}_{t+1}), \quad \bar{\Lambda}_t = H'_t \Sigma_t^{-1} H_t + F'_{p,t} \bar{\Lambda}_{t+1} F_{p,t}, \tag{4.161}$$

where L_t coincides with that of (4.160), initialized with $(L_{k+1}, \bar{\lambda}_{k+1}) = (0, 0)$ and $\bar{\Lambda}_{k+1} = 0$, then, according to the formula for the fixed interval smoother without collapsing (4.146), it holds that

$$\hat{x}_{t|k} = x_{t|t-1} + P_t \bar{\lambda}_t + (U_t - P_t L_t) \tilde{\delta}_{k+1}$$
$$\widehat{P}_{t|k} = P_t - P_t \bar{\Lambda}_t P_t + (U_t - P_t L_t) \Pi_{k+1} (U_t - P_t L_t)'$$

It is not difficult to verify that λ_t in (4.160) satisfies $\lambda_t = \bar{\lambda}_t + F_{p,t}^{k'} \lambda_{k+1}$. Using the recursion (4.161), the formula for the fixed interval smoother without collapsing (4.146) and Theorem 4.44, it holds that

$$\left(-C_k, x_{t|k}\right) = (-U_t, x_{t|t-1}) + P_t (L_t, \bar{\lambda}_t)$$

and thus $C_k = U_t - P_t L_t = -A_t$. Then, putting it all together we have

$$\begin{aligned}
\hat{x}_{t|n} &= x_{t|t-1} + P_t \lambda_t - A_t \left(\tilde{\delta}_{k+1} + \Pi_{k+1} U'_{k+1} \lambda_{k+1} \right) \\
\widehat{P}_{t|n} &= P_t - P_t \bar{\Lambda}_t P_t + A_t \Pi_{k+1} A'_t - (P_t F_{p,t}^{k'} - A_t \Pi_{k+1} U'_{k+1}) \\
&\quad \times \Lambda_{k+1} (P_t F_{p,t}^{k'} - A_t \Pi_{k+1} U'_{k+1})' \\
&= P_t - P_t \left(\bar{\Lambda}_t + F_{p,t}^{k} \Lambda_{k+1} F_{p,t}^{k} \right) P_t + A_t \left(\Pi_{k+1} - \Pi_{k+1} U'_{k+1} \Lambda_{k+1} U_{k+1} \Pi_{k+1} \right) A'_t \\
&\quad + A_t \Pi_{k+1} U'_{k+1} \Lambda_{k+1} F_{p,t}^{k} P_t + P_t F_{p,t}^{k'} \Lambda_{k+1} U_{k+1} \Pi_{k+1} A'_t.
\end{aligned}$$

It is easy to show that $\bar{\Lambda}_t + F^k_{p,t}\Lambda_{k+1}F^k_{p,t} = \Lambda_t$. It remains to prove that $B_t = -F^{k'}_{p,t}\Lambda_{k+1}U_{k+1}$. To this end, consider first that U_{t+1} is given for $t \leq k$ by the modified bias-free filter as $U_{t+1} = F_tU_t - K_tE_t = F_{p,t}U_t$. Then, we can write

$$\begin{aligned} -F'_{p,t}\Lambda_{t+1}U_{t+1} &= -F'_{p,t}\Lambda_{t+1}F_{p,t}U_t \\ &= -(\Lambda_t - H'_t\Sigma_t^{-1}H_t)U_t \\ &= H'_t\Sigma_t^{-1}E_t - \Lambda_tU_t. \end{aligned}$$

It follows from this that $-F'_{p,k}\Lambda_{k+1}U_{k+1} = L_k - \Lambda_kU_k$ and

$$\begin{aligned} -F^{k'}_{p,t}\Lambda_{k+1}U_{k+1} &= -F'_{p,t}\cdots F'_{p,k}\Lambda_{k+1}U_{k+1} \\ &= F'_{p,t}\cdots F'_{p,k-1}\left(L_k - \Lambda_kU_k\right) \\ &= F'_{p,t}\cdots F'_{p,k-2}\left(F'_{p,k-1}L_k + H'_{k-1}\Sigma_{k-1}^{-1}E_{k-1} - \Lambda_{k-1}U_{k-1}\right) \\ &= F'_{p,t}\cdots F'_{p,k-2}\left(L_{k-1} - \Lambda_{k-1}U_{k-1}\right) \\ &= L_t - \Lambda_tU_t. \end{aligned}$$

If $\beta \neq 0$, we can prove the formula for $\hat{\gamma}_{n+1}$ and Ψ_{n+1} as in the first part of the proof, by defining an augmented state vector $x^a_t = (x'_t, \gamma')'$ and considering its associated state space model. In this way, we obtain the same fixed point smoother as before, with the appropriate initial conditions. However, if there is a collapse at time k, this time $X_{1|k}, P_{1|k}$ and P^{k+1}_1 are redefined to

$$X_{1|k} = \begin{bmatrix} S_{db,k+1} & \tilde{\delta}_{d,k+1} \\ -I & 0 \end{bmatrix}, \quad P_{1|k} = \begin{bmatrix} \Pi_{d,k+1} & 0 \\ 0 & 0 \end{bmatrix}, \quad P^{k+1}_1 = \begin{bmatrix} \Pi_{d,k+1}U'_{d,k+1} \\ 0 \end{bmatrix}.$$

Iterating and using (4.156) and (4.157), the result for $\hat{\gamma}_{n+1}$ and Ψ_{n+1} is obtained. More specifically,

$$\begin{aligned} \hat{\gamma}_{n+1} &= \left[X_{1|k} + P^{k+1}_1\left(L_{b,k+1}, \lambda_{k+1}\right)\right]\begin{bmatrix} -\hat{\beta}_{n+1} \\ 1 \end{bmatrix} \\ \Psi_{n+1} &= P_{1|k} - P^{k+1}_1\Lambda_{k+1}P^{k+1'}_1 + \begin{bmatrix} T_{db,k+1} \\ -I \end{bmatrix}\Pi_{n+1}\left[T'_{db,k+1}\ -I\right]. \end{aligned}$$

The proof of the fixed interval smoother formula runs parallel to the previous one for the case $\beta = 0$. By (4.156) and (4.157), we can write

$$\begin{aligned} \hat{x}_{t|n} &= \left[(-C_k, x_{t|k}) + P^{k+1}_t(L_{b,k+1}, \lambda_{k+1})\right]\begin{bmatrix} -\hat{\beta}_{n+1} \\ 1 \end{bmatrix} \\ \widehat{P}_{t|n} &= P_{t|k} - P^{k+1}_t\Lambda_{k+1}P^{k+1'}_t + \left[P^{k+1}_tL_{b,k+1} - C_k\right]\Pi_{n+1}\left[P^{k+1}_tL_{b,k+1} - C_k\right]', \end{aligned}$$

where $P_t^{k+1} = P_t F_{p,t}^{k\prime} + C_{d,k}\Pi_{d,k+1}U'_{d,k+1}$. Define for $t = k\dots,1$ the backwards recursion

$$(\bar{L}_t, \bar{\lambda}_t) = H_t'\Sigma_t^{-1}(E_t, e_t) + F_{p,t}'(\bar{L}_{t+1}, \bar{\lambda}_{t+1}), \quad \bar{\Lambda}_t = H_t'\Sigma_t^{-1}H_t + F_{p,t}'\bar{\Lambda}_{t+1}F_{p,t}, \tag{4.162}$$

where $\bar{L}_t = (L_{d,t}, \bar{L}_{b,t})$, initialized with $(\bar{L}_{k+1}, \bar{\lambda}_{k+1}) = (0,0)$ and $\bar{\Lambda}_{k+1} = 0$. Then, using (4.162), the formula for the fixed interval smoother without collapsing (4.146) and Theorem 4.44, it holds that

$$\left[(-C_{d,k}, -C_{b,k}), x_{t|k}\right] = (-U_t, x_{t|t-1}) + P_t(\bar{L}_t, \bar{\lambda}_t)$$
$$P_{t|k} = P_t - P_t\bar{\Lambda}_t P_t.$$

It follows from this that, before collapsing in the fixed point smoother,

$$C_k = (C_{d,k}, C_{b,k}), \quad C_{b,k} = P_t\bar{L}_{b,t} - U_{b,t}, \quad C_{d,k} = -A_t\begin{bmatrix} I \\ 0 \end{bmatrix}, \quad x_{t|k} = x_{t|t-1} + P_t\bar{\lambda}_t.$$

According to Theorem 4.44, after collapsing, C_k, $x_{t|k}$ and $P_{t|k}$ are redefined to

$$C_k = A_t\begin{bmatrix} S_{db,k+1} \\ 0 \end{bmatrix} + P_t\bar{L}_{b,t} - U_{b,t}, \quad x_{t|k} = x_{t|t-1} + P_t\bar{\lambda}_t - A_t\begin{bmatrix} \tilde{\delta}_{d,k+1}) \\ 0 \end{bmatrix},$$

and

$$P_{t|k} = P_t - P_t\bar{\Lambda}_t P_t + A_t\begin{bmatrix} \Pi_{d,k+1} & 0 \\ 0 & 0 \end{bmatrix} A_t'.$$

On the other hand,

$$P_t^{k+1}(L_{b,k+1}, \lambda_{k+1}) = P_t F_{p,t}^{k\prime}(L_{b,k+1}, \lambda_{k+1}) - A_t\begin{bmatrix} \Pi_{d,k+1}U'_{d,k+1} \\ 0 \end{bmatrix}(L_{b,k+1}, \lambda_{k+1}).$$

It is not difficult to verify that $(L_{b,t}, \lambda_t)$ in (4.160) satisfies

$$(L_{b,t}, \lambda_t) = (\bar{L}_{b,t}, \bar{\lambda}_t) + F_{p,t}^{k\prime}(L_{b,t}, \lambda_t)$$

and thus

$$P_t(\bar{L}_{b,t}, \bar{\lambda}_t) + P_t F_{p,t}^{k\prime}(L_{b,k+1}, \lambda_{k+1}) = P_t(L_{b,t}, \lambda_t).$$

Putting it all together, we get

$$\hat{x}_{t|n} = x_{t|t-1} + P_t\lambda_t - A_t \begin{bmatrix} \tilde{\delta}_{k+1} + \Pi_{d,k+1}U'_{d,k+1}\lambda_{k+1} - T_{db,k+1}\hat{\beta}_{n+1} \\ \hat{\beta}_{n+1} \end{bmatrix}$$

$$\begin{aligned} \widehat{P}_{t|n} &= P_{t|k} - P_t^{k+1}\Lambda_{k+1}P_t^{k+1'} + \left[P_t^{k+1}L_{b,k+1} - C_k\right]\Pi_{n+1}\left[P_t^{k+1}L_{b,k+1} - C_k\right]' \\ &= P_t - P_t\Lambda_tP_t - A_t\Phi_{k+1}B'_tP_t - P_tB_t\Phi'_{k+1}A'_t + A_t\begin{bmatrix} Z_{k+1} & 0 \\ 0 & 0 \end{bmatrix}A'_t \\ &\quad + \left[P_t^{k+1}L_{b,k+1} - C_k\right]\Pi_{n+1}\left[P_t^{k+1}L_{b,k+1} - C_k\right]' \\ &= P_t - P_t\Lambda_tP_t - A_t\Phi_{k+1}B'_tP_t - P_tB_t\Phi'_{k+1}A'_t + A_t\begin{bmatrix} Z_{k+1} & 0 \\ 0 & 0 \end{bmatrix}A'_t \\ &\quad + A_t\begin{bmatrix} T_{db,k+1} \\ -I \end{bmatrix}\Pi_{n+1}\begin{bmatrix} T'_{db,k+1} & -I \end{bmatrix}A'_t. \end{aligned}$$

□

4.21.3.3 Interpolation and Disturbance Smoothing Under Collapsing

Suppose that in the state space model (4.85) and (4.86) we want to smooth the vector $Z_t = X_t\beta + C_tx_t + D_t\varepsilon_t$. If there are missing observations and only part of the vector Y_t is observed for $t = 1, 2, \ldots, n$, then setting $X_t = V_t, C_t = H_t$ and $D_t = J_t$ in the expression for Z_t will allow us to interpolate Y_t. Setting $X_t = 0, C_t = 0$ and $D_t = I$ will give us the disturbance smoother and, finally, setting $X_t = 0, C_t = I$, and $D_t = 0$ will provide the state smoother. Thus, smoothing the vector Z_t is important.

As before, suppose that the TSKF–SRIBF is used for $t = 1, \ldots, n$ and a collapse takes place at $t = k, k < n$, after which the augmented part is reduced or eliminated at $t = k + 1$. If $\beta = 0$ in (4.85) and (4.86), then $X_t = 0$ in Z_t.

The following theorem gives the details for the smoothing of Z_t when the TSKF–SRIBF is used under collapsing. The proof is similar to that of Theorem 4.46 and is omitted.

Theorem 4.47 *Suppose that we want to smooth $Z_t = X_t\beta + C_tx_t + D_t\varepsilon_t$ with $t \leq n$ in the state space model (4.85) and (4.86) when the TSKF–SRIBF is used and a collapse takes place at $t = k, k < n$. Let $\widehat{Z}_{t|n}$ and $\widehat{N}_{t|n}$ be the estimator of Z_t based on $Y_{1:n}$ and its mean squared error. If $\beta \neq 0$, for $t = n, \ldots, k + 1$ the augmented Bryson–Frazier recursions (4.155) are used and the following recursions hold*

$$\begin{aligned} \widehat{Z}_{t|n} &= \left[C_t(-U_{b,t}, x_{t|t-1}) + K_t^Z(E_{b,t}, e_t) + N_t(L_{b,t+1}, \lambda_{t+1}) - (X_t, 0)\right]\begin{bmatrix} -\hat{\beta}_{n+1} \\ 1 \end{bmatrix} \\ \widehat{N}_{t|n} &= N_{t|t} - N_t\Lambda_{t+1}N'_t + A_{b,t}\Pi_{n+1}A'_{b,t}, \end{aligned}$$

where $K^Z_t = (C_tP_tH'_t + D_tJ'_t)\Sigma_t^{-1}, N_t = (C_t - K^Z_tH_t)P_tF'_t + (D_t - K^Z_tJ_t)G'_t, N_{t|t} = (C_t - K^Z_tH_t)P_tC'_t + (D_t - K^Z_tJ_t)D'_t, A_{b,t} = -C_tU_{b,t} + K^Z_tE_{b,t} + N_tL_{b,t+1} - X_t$ *and* $\hat{\beta}_{n+1}$ *and* Π_{n+1} *are the GLS estimator of* β *based on* $Y_{1:n}$ *and its MSE. If* $\beta = 0$*, the Bryson–Frazier recursions (4.33) and the following recursions are used for* $t = n, \ldots, k+1$

$$\widehat{Z}_{t|n} = C_t\hat{x}_{t|t-1} + K^Z_te_t + N_t\lambda_{t+1}$$
$$\widehat{N}_{t|n} = N_{t|t} - N_t\Lambda_{t+1}N'_t.$$

For $t = k, \ldots, 1$*, the recursions for* Λ_t *are continued and the recursions for* $(L_{b,t}, \lambda_t)$ *if* $\beta \neq 0$ *or* λ_t *if* $\beta = 0$ *are also continued, augmented with* $L_{d,t}$*, to get (4.160), where* $L_t = (L_{d,t}, L_{b,t})$ *or* $L_t = L_{d,t}$*, respectively, and* $L_{d,k+1} = 0$*. Let* $A_t = -C_tU_t + K^Z_tE_t + N_tL_{t+1} - (0, X_t)$ *and* $B_t = L_t - \Lambda_tU_t$*, where* (L_t, λ_t) *and* Λ_t *are given by (4.160) and* U_t *and* E_t *are given by the TSKF–SRIBF before collapsing, and let* $\hat{\gamma}_{n+1} = (\hat{\delta}'_{n+1}, \hat{\beta}'_{n+1})'$ *and* Ψ_{n+1} *be the GLS estimator of* γ *based on* $Y_{1:n}$ *and its MSE. For* $t = k, \ldots, 1$*, the following recursions hold*

$$\widehat{Z}_{t|n} = C_t\hat{x}_{t|t-1} + K^Z_te_t + N_t\lambda_{t+1} - A_t\hat{\gamma}_{n+1}$$
$$\widehat{N}_{t|n} = N_{t|t} - N_t\Lambda_{t+1}N'_t - A_t\Gamma_{k+1}B'_{t+1}N'_t - N_tB_{t+1}\Gamma'_{k+1}A'_t + A_t\Psi_{n+1}A'_t,$$

where $\Gamma_{k+1} = \Phi_{k+1}$ *if* $\beta \neq 0$ *or* $\Gamma_{k+1} = \Pi_{k+1}$ *if* $\beta = 0$*, and* $\Phi_{k+1}, \Pi_{k+1}, \hat{\gamma}_{n+1}$ *and* Ψ_{n+1} *are as in Theorem 4.46.*

The following theorem gives the formulae for smoothing the disturbances in the state space model (4.85) and (4.86). We omit its proof because, as mentioned earlier in this section, it is a direct consequence of the previous theorem.

Theorem 4.48 (Disturbance Smoother) *The disturbance smoother is obtained by setting in Theorem 4.47* $X_t = 0, C_t = 0$ *and* $D_t = I$*. This yields* $K^Z_t = J'\Sigma_t^{-1}, N_t = (G_t - K_tJ_t)', N_{t|t} = I - J'_t\Sigma_t^{-1}J_t, A_{b,t} = J'_t\Sigma_t^{-1}E_{b,t} + N_tL_{b,t+1}$ *and* $A_t = J'_t\Sigma_t^{-1}E_t + N_tL_{t+1}$*. Thus, if* $\beta \neq 0$*, for* $t = n, \ldots, k+1$ *the augmented Bryson–Frazier recursions (4.155) are used and the following recursions hold*

$$E(\epsilon_t|y) = \left[J'\Sigma_t^{-1}(E_{b,t}, e_t) + (G_t - K_tJ_t)'(L_{b,t+1}, \lambda_{t+1})\right]\begin{bmatrix} -\hat{\beta}_{n+1} \\ 1 \end{bmatrix}$$
$$MSE(\epsilon_t|y) = I - J'_t\Sigma_t^{-1}J_t - (G_t - K_tJ_t)'\Lambda_{t+1}(G_t - K_tJ_t) + A_{b,t}\Pi_{n+1}A'_{b,t}.$$

If $\beta = 0$*, the Bryson–Frazier recursions (4.33) and the following recursions are used for* $t = n, \ldots, k+1$

$$E(\epsilon_t|y) = J'_t\Sigma_t^{-1}e_t + (G_t - K_tJ_t)'\lambda_{t+1}$$
$$MSE(\epsilon_t|y) = I - (J'_t\Sigma_t^{-1}J_t + (G_t - K_tJ_t)'\Lambda_{t+1}(G_t - K_tJ_t).$$

For $t = k, \ldots, 1$*, the following recursions hold*

$$\begin{aligned} E(\epsilon_t|y) &= J_t'\Sigma_t^{-1}e_t + N_t\lambda_{t+1} - A_t\hat{\gamma}_{n+1} \\ MSE(\epsilon_t|y) &= I - (J_t'\Sigma_t^{-1}J_t + N_t\Lambda_{t+1}N_t') - A_t\Gamma_{k+1}B_{t+1}'N_t' \\ &\quad -N_tB_{t+1}\Gamma_{k+1}'A_t' + A_t\Psi_{n+1}A_t', \end{aligned}$$

where (L_t, λ_t) *and* Λ_t *are given by (4.160) and* $\Gamma_{k+1}, B_t, \hat{\gamma}_{n+1}$ *and* Ψ_{n+1} *are as in Theorem 4.47.*

4.22 Historical Notes

The use of what engineers call state space models or descriptions of systems described by high-order differential or difference equations is well known in books on differential equations. These representations were used by physicists in studying Markov processes, see, for example, Wang & Uhlenbeck (1945). A detailed study of stationary Gaussian processes having the Markov property was given by Doob (1944), but this paper was not cited in his famous 1953 textbook (Doob, 1953). Had he done so, development of state space estimation results might have occurred much earlier.

During the mid-fifties, some scientists and engineers in both the USA and the USSR began to reemphasize the value of state space descriptions in control problems, switching circuits and automata theory and even classical circuit theory. However, it was undoubtedly Kalman, through his outstanding research who brought the state space point of view into the center stage in system theory. See, for example, Kalman (1960a, 1960c) and Kalman et al. (1969).

The fact that the state space model (4.1) and (4.2) has the properties that make it wide sense Markov is critical in the derivations. If, for example, x_1 were correlated with $\{u_t\}, t \geq 1$, or if $\{v_t\}$ were not a white process, then the Kalman filter arguments would break down. The Markov property is not mentioned explicitly in Kalman (1960b). However, it is made explicit in Kalman (1963). Verghese & Kailath (1979) gave a physical argument to show that, given a forwards Markovian model, this model can also be represented by a backwards Markovian model.

The idea of propagating square root factors, rather than the matrices themselves, was first introduced by Potter for the measurement update step of the Kalman filter recursions. See Potter & Stern (1963). However, Potter's method was not applicable to the time update problem, which requires an array formulation. The time update formulation was proposed by Schmidt (1970), who combined it with Potter's measurement update for application in an airborne navigation system for precision approach and landing.

Independently of these developments, Golub (1965) and Businger & Golub (1965) proposed the method of orthogonal triangularization suggested by Housholder (1953) to solve the recursive least square problem.

The fact that smoothing problems were left unsolved in the original papers of Kalman Kalman (1960b, 1963) triggered a cascade of efforts to solve the problem. See Meditch (1973) for a nice survey of the literature up to 1973.

It is ironic that smoothing problems are easier to address in the Wiener–Kolmogorov theory. See Chap. 7. One reason is that recursive implementations of the Wiener–Kolmogorov smoother were not pursued. Doing so would have required not only the need for properly defining backwards Markovian models but also the introduction of state space models.

Nerlove (1967) proposed for economic series the unobserved components hypothesis, according to which the underlying economic variables are not directly observable by agents. It is assumed that the observed economic variables are the sum or product of unobserved components, such as trend, cycle, seasonal component, etc. The unobserved components hypothesis gave rise to the so-called structural time series models, in which it is assumed from the beginning that the observed series is the sum of several unobserved components. These models are easily put into state space form. See, for example, Harrison & Stevens (1976), Harvey (1989), and Kitagawa & Gersch (1996) and the references therein.

Standard texts in the statistical and econometric literature regarding state space models and the Kalman filter are Harvey (1989) and Durbin & Koopman (2012). However, these books do not treat the theory of linear state space models in all generality, but focus from the beginning on structural time series models.

The TSKF filter was originally proposed by Friedland (1969). See also Ignagni (1981) and the references therein. These results seem to have passed unnoticed in the statistical literature and, in fact, some of them have been reinvented by several authors, like Rosenberg (1973) and Jong (1991). A different approach was taken by Ansley & Kohn (1985) to tackle the problem of defining the likelihood of a state space problem with incompletely specified initial conditions. This approach has also been used by other authors, like Koopman (1997) and Durbin & Koopman (2012).

4.23 Problems

4.1 With the notation of Theorem 4.17, prove that

$$Q_t^r = Q_t^s - Q_t^s G_t' P_{t+1}^{-1} G_t Q_t^s$$

and

$$\left(I + A_t G_t Q_t^s G_t'\right)^{-1} = I - A_t G_t Q_t^r G_t'.$$

Hint: Use the Matrix Inversion Lemma 4.1 and the relation $P_{t+1} = F_t^s P_{t|t} F_t^{s'} + G_t Q_t^s G_t'$ to prove the first equality. To prove the second one, use Lemma 4.1 again and the definition of Q_t^r in Theorem 4.17.

4.2 Prove Lemma 4.4.

4.3 Using Proposition 1.2, prove Lemma 4.7.

4.4 Prove Theorem 4.32.

4.5 Prove Proposition 4.1.

4.6 Prove that the modified bias-free filter corresponding to Example 4.9 produces the L^{-1} and D matrices stated in that example.

4.7 Prove Theorem 4.42.

4.8 After running the TSKF–SRIBF with collapsing using the state space model of Example 4.3 and the sample $Y = (Y_1, Y_2, Y_3)'$, apply the fixed interval smoother of Theorem 4.45 and show that

$$\begin{bmatrix} \hat{x}_{1|3} \\ \hat{x}_{2|3} \\ \hat{x}_{3|3} \end{bmatrix} = \begin{bmatrix} 11/21 & 6/21 & 4/21 \\ 2/7 & 3/7 & 2/7 \\ 4/21 & 2/7 & 11/21 \end{bmatrix} Y, \qquad \begin{bmatrix} P_{1|3} \\ P_{2|3} \\ P_{3|3} \end{bmatrix} = \begin{bmatrix} 22/21 \\ 6/7 \\ 22/21 \end{bmatrix}.$$

4.9 Suppose that the process $\{Y_t : t = 1, \ldots, n\}$ follows the model

$$Y_t - Y_{t-4} = A_t + \theta A_{t-1},$$

where the A_t are uncorrelated random variables with zero mean and common variance σ^2. Let $\Phi(z) = 1 - z^4, \Theta(z) = 1 + \theta z$ and $\Psi(z) = \Phi(z)^{-1}\Theta(z) = \sum_{j=0}^{\infty} \Psi_j z^j$ and define $x_{t,i} = Y_{t+i-1} - \sum_{j=0}^{i-1} \Psi_j A_{t+i-1-j}, j = 1, 2, 3, 4$, and $x_t = (x_{t,1}, \ldots, x_{t,4})'$. Show that Akaike's representation

$$\begin{aligned} x_{t+1} &= F x_t + K A_t \\ Y_t &= H x_t + A_t, \qquad t = 1, 2, \ldots, n, \end{aligned}$$

where

$$F = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}, \qquad K = \begin{bmatrix} \theta \\ 0 \\ 0 \\ 1 \end{bmatrix}, \qquad H = [1, 0, 0, 0],$$

is correct also in this case in spite of the fact that the process $\{Y_t : t = 1, \ldots, n\}$ is not stationary. Justify, using the results of Sect. 4.14.2, why we should take $x_1 = \delta$, where δ is diffuse. Show that, if this initial state vector is used and no collapsing is performed, the equations of the modified bias-free filter reduce to

$$(E_t, e_t) = (0, Y_t) - H(-U_t, x_{t|t-1})$$
$$(-U_{t+1}, x_{t+1|t}) = F(-U_t, x_{t|t-1}) + K(E_t, e_t)$$

with initial conditions $(-U_1, x_{1|0}) = (-I, 0)$ and $P_1 = 0$, because for all $t, P_t = 0$, $\Sigma_t = \sigma^2$, and $K_t = K$.

4.10 Using the results of Sect. 4.14.2, show that we can also take in Example 4.4 $x_1 = \delta$, where δ is diffuse, as initial state vector. Prove that, using this initial state vector, we obtain the same Kalman filter initialization after collapsing the TSKF–SRIBF at $t = 3$ as with the original initial state vector.

Chapter 5
Time Invariant State Space Models

Given a time series $Y = (Y'_1, \ldots, Y'_n)'$ with k-dimensional observations Y_t, we say that it follows a time invariant state space model if we can write

$$x_{t+1} = Fx_t + G\epsilon_t, \tag{5.1}$$

$$Y_t = Hx_t + J\epsilon_t, \quad t = 1, \ldots, n, \tag{5.2}$$

where $\{\epsilon_t\}$ is a zero mean serially uncorrelated sequence of dimension q, $\text{Var}(\epsilon_t) = \sigma^2 I_q$, and x_t has dimension r. The initial state vector x_1 is specified as

$$x_1 = A\delta + x, \tag{5.3}$$

where x is a zero mean stochastic vector with $\text{Var}(x) = \sigma^2\Omega$, the matrix A is fixed and known, and δ is a zero mean stochastic vector with $\text{Var}(\delta) = \sigma^2\Pi$ that models the uncertainty about the initial conditions. If $\Pi^{-1} = 0$, δ is a "diffuse" random vector, see Examples 4.3, 4.4 and 4.9. If there are no diffuse effects, then δ is zero in (5.3) and $x_1 = x$. If $\Pi = 0$, δ is fixed and x_1 has mean $A\delta$ and variance $\sigma^2\Omega$. The notable feature of Eqs. (5.1) and (5.2) is that the system matrices F, G, H, and J do not depend on the time index t.

In most of this chapter, it will be more convenient to use instead of (5.1) and (5.2) the following state space form

$$x_{t+1} = Fx_t + Gu_t \tag{5.4}$$

$$Y_t = Hx_t + v_t, \quad t = 1, \ldots, n, \tag{5.5}$$

V. Gómez, *Multivariate Time Series With Linear State Space Structure*,
DOI 10.1007/978-3-319-28599-3_5

where

$$E\left\{\begin{bmatrix} u_t \\ v_t \end{bmatrix}\begin{bmatrix} u_s', v_s' \end{bmatrix}\right\} = \sigma^2 \begin{bmatrix} Q & S \\ S' & R \end{bmatrix} \delta_{ts},$$

x_1 is as in (5.3), and x_1 is orthogonal to the zero mean $\{u_t\}$ and $\{v_t\}$ sequences.

5.1 Covariance Function of a Time Invariant Model

In this section we assume that δ in (5.3) is zero and thus there is no diffuse part in the initial state vector. Without loss of generality, we also assume that $\sigma^2 = 1$.

The covariances of the state vectors and the observations in (5.4) and (5.5) are given by the following lemma.

Lemma 5.1 *Consider the time invariant state space model (5.4) and (5.5) and let* $\Pi_t = E(x_t x_t')$. *Then,* Π_t *satisfies* $\Pi_1 = \Omega = E(x_1 x_1')$ *and*

$$\Pi_{t+1} = F\Pi_t F' + GQG', \qquad t \geq 1. \tag{5.6}$$

The covariances of the state variables can be written as

$$\gamma_X(r, s) = E(x_r x_s') = \begin{cases} F^{r-s}\Pi_s & r \geq s \\ \Pi_r F^{(s-r)'} & r \leq s, \end{cases}$$

and the covariances of the output process $\{Y_t\}$ *as*

$$\gamma_Y(r, s) = E(Y_r Y_s') = \begin{cases} HF^{r-s-1}N_s & r > s \\ R + H\Pi_r H' & r = s \\ N_r' F^{(s-r-1)'} H' & r < s, \end{cases}$$

where $N_r = F\Pi_r H' + GS = \text{Cov}(x_{r+1}, Y_r)$.

Proof A straightforward direct calculation gives the results. □

5.2 Stationary State Space Models

An important consequence of Lemma 5.1 is that although the underlying state space model is time invariant, and although the disturbance processes $\{u_t\}$ and $\{v_t\}$ are stationary, when we start at $t = 1$ neither the state process $\{x_t : t \geq 1\}$ nor the output process $\{Y_t : t \geq 1\}$ is in general stationary. The reason is that the state covariance matrix Π_t is in general time dependent. However, the processes $\{x_t : t \geq 1\}$ and

$\{Y_t : t \geq 1\}$ can be made stationary if Π_t is time invariant, as the following lemma shows.

Lemma 5.2 *The processes* $\{x_t : t \geq 1\}$ *and* $\{Y_t : t \geq 1\}$ *in (5.4) and (5.5) are stationary if, and only if,* $E(x_r x_r') = \Pi_r$ *is time invariant,* $\Pi_r = \Pi$, *for* $r \geq 1$.

Proof The condition is obviously necessary. By Lemma 5.1, if Π_r is time invariant, then N_r is also time invariant, $N_r = N$, and the condition is also sufficient. □

By Lemma 5.2 and (5.6), if the process $\{x_t : t \geq 1\}$ is stationary, then $E(x_r x_r') = \Pi$ for $r \geq 1$ and Π satisfies

$$\Pi = F\Pi F' + GQG'. \tag{5.7}$$

This equation is the celebrated discrete-time Lyapunov equation. We will see later in this chapter that if F has all its eigenvalues inside the unit circle and Q is positive semi-definite, there is a unique solution Π of (5.7) that is positive semi-definite. Assuming that this condition on F is satisfied, the following theorem gives a necessary and sufficient condition on the initial state vector x_1 for the processes $\{x_t : t \geq 1\}$ and $\{Y_t : t \geq 1\}$ to be stationary. It also gives the covariances of these last two processes when they are stationary.

Theorem 5.1 *Consider the time invariant state space model (5.4) and (5.5) and suppose that* F *has all its eigenvalues inside the unit circle and that* Π *is the unique solution of the Lyapunov equation (5.7). Then, the processes* $\{x_t : t \geq 1\}$ *and* $\{Y_t : t \geq 1\}$ *are stationary with covariances given by*

$$\gamma_X(r-s) = E(x_r x_s') = \begin{cases} F^{r-s}\Pi & r \geq s \\ \Pi F^{(s-r)'} & r \leq s, \end{cases}$$

and

$$\gamma_Y(r-s) = E(Y_r Y_s') = \begin{cases} HF^{r-s-1}N & r > s \\ R + H\Pi H' & r = s \\ N'F^{(s-r-1)'}H' & r < s, \end{cases}$$

where $N = F\Pi H' + GS = \mathrm{Cov}(x_{t+1}, Y_t)$, *if, and only if,* $x_1 = x$ *with* $x \sim (0, \Pi)$.

Proof By Lemma 5.2 and (5.6), the condition is necessary. To see that it is also sufficient, let $x_1 = x$ with $x \sim (0, \Pi)$, where Π is a solution of (5.7). Then, it follows from (5.6) that $\Pi_t = \Pi$ for $t \geq 1$. This in turn implies that $N_t = F\Pi H' + GS$ is time invariant, $N_t = N$, and the processes $\{x_t : t \geq 1\}$ and $\{Y_t : t \geq 1\}$ are stationary. The formulae for the covariances are a straightforward consequence of those of Lemma 5.1. □

Example 5.1 Consider the process $\{Y_t\}$ following the autoregressive model $Y_t = \phi Y_{t-1} + A_t$, where $|\phi| < 1$ and $\{A_t\}$ is an uncorrelated sequence of random variables with zero mean and common variance σ^2. This model can be put into state space form (5.4) and (5.5) as

$$x_{t+1} = Fx_t + Gu_t$$
$$Y_t = Hx_t, \qquad t = 1, 2, \ldots, n,$$

where $F = \phi$, $H = 1$, $G = 1$, $v_t = 0$, $u_t = A_{t+1}$, $Q = \sigma^2$, and $x_1 \sim (0, \sigma^2\Pi)$. For the process to be stationary, the matrix $\sigma^2\Pi$ should satisfy (5.7) or

$$\Pi = \phi^2\Pi + 1.$$

Thus, $\Pi = 1/(1-\phi^2)$ and, using Theorem 5.1, it is obtained that

$$\text{Var}(Y) = \frac{\sigma^2}{1-\phi^2}\begin{bmatrix} 1 & \phi & \phi^2 & \cdots & \phi^{n-1} \\ \phi & 1 & \phi & \cdots & \phi^{n-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \phi^{n-1} & \phi^{n-2} & \phi^{n-3} & \cdots & 1 \end{bmatrix}.$$

◇

Example 5.2 Let the process $\{Y_t\}$ follow the model $Y_t + \phi Y_{t-1} = A_t + \theta A_{t-1}$, where $|\phi| < 1$, $|\theta| < 1$ and $\{A_t\}$ is an uncorrelated sequence of random variables with zero mean and common variance σ^2. This model can be put into state space form (5.4) and (5.5) by defining $F = -\phi$, $G = \theta - \phi$, $H = 1$, $u_t = v_t = A_t$, $Q = R = S = \sigma^2$, and $x_1 \sim (0, \sigma^2\Pi)$. According to (5.7), the process is stationary if, and only if,

$$\Pi = \phi^2\Pi + (\theta - \phi)^2,$$

or $\Pi = (\theta-\phi)^2/(1-\phi^2)$. The covariances of x_t and Y_t can be obtained easily using Theorem 5.1. See Problem 5.3. ◇

5.3 The Lyapunov Equation

Given the $n \times n$ matrices F, Π and Q, the Lyapunov equation is defined as

$$\Pi = F\Pi F' + Q, \tag{5.8}$$

where we assume that the matrix Q is symmetric and positive semi-definite. Let λ_i, $i = 1, \ldots, n$, be the eigenvalues of F. The following lemma gives a necessary and sufficient condition for (5.8) to have a symmetric solution Π.

Lemma 5.3 *The Lyapunov equation (5.8) has a unique symmetric solution* Π *if, and only if,*

$$\lambda_i \lambda_j \neq 1, \qquad i,j = 1, \ldots, n. \tag{5.9}$$

Proof Using the properties of the Kronecker product, if we apply the vec operator to (5.8), we get $(I - F \otimes F)\text{vec}(\Pi) = \text{vec}(Q)$. Thus, a unique solution $\text{vec}(\Pi)$ exists if, and only if, the matrix $(I - F \otimes F)$ is nonsingular. Since the eigenvalues of this last matrix are given by $1 - \lambda_i \lambda_j$, $i,j = 1, \ldots, n$, the nonsingularity of $(I - F \otimes F)$ is equivalent to (5.9). To see that the solution Π of (5.8) is symmetric, consider that Π' also satisfies (5.8). □

It may be the case that the Lyapunov equation (5.8) has a unique solution Π that is not positive semi-definite. The following lemma gives a necessary and sufficient condition for a solution to be positive semi-definite.

Lemma 5.4 *Equation (5.8) has a positive semi-definite solution* Π *if, and only if, the series* $\sum_{j=0}^{\infty} F^j Q F'^j$ *converges and its sum is equal to* Π*, where* $F^0 = I$.

Proof To see that the condition is necessary, iterate in $\Pi = F\Pi F' + Q$ to get

$$\begin{aligned}\Pi &= F(F\Pi F' + Q)F' + Q \\ &= \sum_{j=0}^{n-1} F^j Q F'^j + F^n \Pi F'^n, \qquad n = 1, 2, \ldots .\end{aligned} \tag{5.10}$$

Let $\Pi_n = \sum_{j=0}^{n} F^j Q F'^j$. Then, (5.10) shows that, for any vector v, the sequence $\{v'\Pi_n v\}$ is nondecreasing and bounded by $v'\Pi v$. Thus, it has a finite limit. Letting $v = e_{(i)}$, where $e_{(i)}$ is a vector with a one in the i-th position and zeros otherwise, it is obtained that the i-th element of Π_n converges. The convergence of $(e_{(i)} + e_{(j)})'\Pi_n(e_{(i)} + e_{(j)})$ shows in turn that the (i,j) element of Π_n converges.

To verify the sufficiency of the condition, consider that if the limit Π of $\{\Pi_n\}$ exists, then

$$\begin{aligned}\Pi &= \left(Q + FQF' + \cdots\right) \\ &= Q + F\left(Q + FQF' + \cdots\right)F' \\ &= Q + F\Pi F'.\end{aligned}$$

□

The following lemma gives a necessary and sufficient condition for the convergence of the series $\sum_{j=0}^{\infty} F^j Q F'^j$.

Lemma 5.5 *The series* $\sum_{j=0}^{\infty} F^j Q F'^j$ *converges if, and only if,* $v'Q^{1/2} = 0$ *for all the vectors* v *in a Jordan base of* F' *corresponding to the eigenvalues* λ *with* $|\lambda| \geq 1$*, where* $Q^{1/2}$ *is any matrix satisfying* $Q = Q^{1/2}Q^{1/2'}$.

Proof To show that the condition is necessary, let $\Pi_n = \sum_{j=0}^{n} F^j Q F'^j$ and let R be a nonsingular matrix such that $R^{-1}FR = J = D + N$, where J is a Jordan canonical matrix corresponding to F, D is a diagonal matrix, and N is a matrix with zeros except possibly in the first superdiagonal where it may have ones. Defining $P = R'^{-1}$, we get $P^{-1}F'P = J' = D + N'$ and the columns of P constitute a Jordan base for F', but ordered in such a way that all of the possible ones are located in the first diagonal under the main diagonal instead of on the first superdiagonal. In addition, it is easy to verify that $F = P'^{-1}JP'$, $F' = PJ'P^{-1} = \bar{P}\bar{J}'\bar{P}^{-1}$, where the bar denotes complex conjugation, and $P'\Pi_n\bar{P} = \sum_{j=0}^{n} J^j P' Q \bar{P} \bar{J}'^j$. We can also assume without loss of generality that

$$J = \begin{bmatrix} J_N & 0 \\ 0 & J_S \end{bmatrix}, \tag{5.11}$$

where J_N contains the eigenvalues λ of F' with $|\lambda| \geq 1$ and J_S contains all the other eigenvalues. Partitioning $P = [P_N, P_S]$ conforming to (5.11) and defining $G_N = P_N' Q^{1/2}$ and $G_S = P_S' Q^{1/2}$, we obtain that the submatrix $W_n = \sum_{j=0}^{n} J_N^j G_N \bar{G}_N' \bar{J}_N'^j$ of $P'\Pi_n\bar{P}$ converges when $n \to \infty$ if, and only if, $\mathrm{vec}(W_n)$ converges. Since

$$\begin{aligned} \mathrm{vec}(W_n) &= \sum_{j=0}^{n} (\bar{J}_N^j \otimes J_N^j)\mathrm{vec}(G_N \bar{G}_N') \\ &= \sum_{j=0}^{n} (\bar{J}_N \otimes J_N)^j \mathrm{vec}(G_N \bar{G}_N') \end{aligned} \tag{5.12}$$

and $\sum_{j=0}^{\infty} (\bar{J}_N \otimes J_N)^j$ diverges, for the convergence of $\mathrm{vec}(W_n)$ one must have $\mathrm{vec}(G_N \bar{G}_N') = 0$, which implies $G_N = 0$. To see this, consider first that $\bar{J}_N \otimes J_N$ is an upper triangular matrix whose eigenvalues, by the properties of the Kronecker product, have all modulus greater than or equal to one. Then, proceeding from bottom to top in (5.12) we get that $P_N' Q^{1/2} = 0$.

To prove the sufficiency of the condition, suppose that $P_N' Q^{1/2} = 0$. Then,

$$\Pi_n = P'^{-1} \begin{bmatrix} 0 & 0 \\ 0 & U_n \end{bmatrix} \bar{P}^{-1},$$

where $U_n = \sum_{j=0}^{n} J_S^j G_S \bar{G}_S' \bar{J}_S'^j$. By an argument similar to that used in the first part of the proof, it is easy to verify that $\mathrm{vec}(U_n)$ converges to $(I - \bar{J}_S \otimes J_S)^{-1}\mathrm{vec}(G_S \bar{G}_S')$ when $n \to \infty$ because the eigenvalues of J_S, and thus those of $\bar{J}_S \otimes J_S$, have all modulus less than 1. □

The previous lemmas allow us to give a necessary and sufficient condition for the existence of a unique, positive semi-definite, solution of the Lyapunov equation (5.8). This is summarized in the following theorem.

Theorem 5.2 *The Lyapunov equation (5.8) has a unique positive semi-definite solution* Π *if, and only if, the condition of Lemma 5.5 is satisfied and* $\lambda_i\lambda_j \neq 1$, $i, j = 1, \ldots, n$. *The unique solution* Π, *if it exists, is given by* $\text{vec}(\Pi) = (I - F \otimes F)^{-1}\text{vec}(Q)$.

Proof Applying the vec operator in $\Pi = F\Pi F' + Q$, it is obtained that $(I - F \otimes F)\text{vec}(\Pi) = \text{vec}(Q)$. If a unique, positive, semi-definite, solution Π exists, then the condition of Lemma 5.5 is satisfied and the system $(I - F \otimes F)\text{vec}(\Pi) = \text{vec}(Q)$ has a unique solution. Thus, the matrix $(I - F \otimes F)$ is nonsingular and $F \otimes F$ cannot have eigenvalues equal to 1.

Conversely, if the condition of Lemma 5.5 is satisfied and the matrix $F \otimes F$ has eigenvalues different from 1, then the system $(I - F \otimes F)\text{vec}(\Pi) = \text{vec}(Q)$ has a unique solution and the Lemmas 5.4 and 5.5 show that Π is positive semi-definite. □

Corollary 5.1 *Suppose that F in the Lyapunov equation (5.8) has all its eigenvalues inside the unit circle. Then, (5.8) has a unique positive semi-definite solution* Π *given by* $\text{vec}(\Pi) = (I - F \otimes F)^{-1}\text{vec}(Q)$.

Proof The conditions of Theorem 5.2 are trivially satisfied. □

When the pair $[F, Q^{1/2}]$ is stabilizable (see Sect. 5.12), where $Q^{1/2}$ is any matrix satisfying $Q = Q^{1/2}Q^{1/2'}$, the following theorem gives a necessary and sufficient condition for the convergence of the series $\sum_{j=0}^{\infty} F^j Q F'^j$.

Theorem 5.3 *If* $[F, Q^{1/2}]$ *is stabilizable, then the series* $\sum_{j=0}^{\infty} F^j Q F'^j$ *converges if, and only if, F is stable.*

Proof Suppose that the series converges and F is not stable. Then, because the pair $[F, Q^{1/2}]$ is stabilizable, there exists a left eigenvector, v, corresponding to an eigenvalue, λ, with $|\lambda| \geq 1$ such that $v'Q^{1/2} \neq 0$. But this contradicts Lemma 5.5. Conversely, if F is stable, by Lemma 5.5, the series converges. □

Remark 5.1 Consider the state space model (5.4) and (5.5) and replace Q by GQG' in the Lyapunov equation (5.8). Then, we can replace in Lemma 5.1 the condition that F has all its eigenvalues inside the unit circle by the two conditions of Theorem 5.2 and the lemma still holds. However, with the notation in the proof of Lemma 5.5, if $V_N = \text{Var}(P'_N x_t)$ and $\{P'_N x_t\}$ is the process corresponding to the partitioning (5.11), we have $V_N = 0$. ◇

The most important solutions Π of the Lyapunov equation (5.8) are those that are positive definite, to avoid situations like those of the previous remark, in which there are linear combinations of the state vector that are perfectly linearly predictable. The following theorem gives a necessary and sufficient condition for the solution Π to be unique and positive definite.

Theorem 5.4 *The Lyapunov equation (5.8) has a unique, positive definite, solution* Π *if, and only if, the pair* $(F, Q^{1/2})$ *is controllable and all the eigenvalues of F are inside the unit circle, where* $Q^{1/2}$ *is any matrix satisfying* $Q = Q^{1/2}Q^{1/2'}$.

Proof We show first that the condition is necessary. By Lemma 5.4, $\sum_{j=0}^{n} F^j Q F'^j$ converges and $\sum_{j=0}^{\infty} F^j Q F'^j = \Pi$. If F has eigenvalues λ with $|\lambda| \geq 1$, we saw in the proof of Lemma 5.5 that $\det(\Pi) = 0$, a contradiction. This shows that all the eigenvalues of F are inside the unit circle.

By the properties of the Kronecker product, $F \otimes F$ also has its eigenvalues inside the unit circle. If the pair $(F, Q^{1/2})$ is not controllable, there is a vector $v \neq 0$ orthogonal to $Q^{1/2}, FQ^{1/2}, \ldots, F^{r-1}Q^{1/2}$. Then, by the Cayley–Hamilton theorem, v is also orthogonal to any matrix of the form $F^n Q^{1/2}$, where $n \geq r$. Because $F \otimes F$ has all its eigenvalues inside the unit circle, the series $\sum_{j=0}^{\infty}(F \otimes F)^j$ converges to $(I - F \otimes F)^{-1}$ and there exists a unique solution given by $\mathrm{vec}(\Pi) = (I - F \otimes F)^{-1}\mathrm{vec}(Q)$. Thus, $\Pi = \sum_{j=0}^{\infty} F^j Q F'^j$, v satisfies $v'\Pi v = 0$ and Π is not positive definite, a contradiction.

To prove sufficiency, consider first that there exists a unique solution because F has all its eigenvalues inside the unit circle. Then, as we saw in the first part of the proof, this solution is $\Pi = \sum_{j=0}^{\infty} F^j Q F'^j$. In addition, if v is any vector, then $v'\Pi v \geq v'\Pi_{r-1}v$, where $\Pi_{r-1} = \sum_{j=0}^{r-1} F^j Q F'^j$ is positive definite because the pair $(F, G^{1/2})$ is controllable. □

The following corollary is an immediate consequence of Theorem 5.4.

Corollary 5.2 *There exists a unique causal nondeterministic solution $\{x_t\}$ of (5.5) with $Var(x_t) > 0$ if, and only if, $(F, GQ^{1/2})$ is controllable and all the eigenvalues of F are inside the unit circle.*

5.4 Covariance Generating Function

Consider the time invariant state space model (5.4) and (5.5) and suppose that F has all its eigenvalues inside the unit circle and $x_1 \sim (0, \Pi)$, where Π is the unique solution of the Lyapunov equation (5.7). Then, by Lemma 5.1, the processes $\{x_t : t \geq 1\}$ and $\{Y_t : t \geq 1\}$ are stationary with covariances given by

$$\gamma_X(r-s) = E(x_r x_s') = \begin{cases} F^{r-s}\Pi & r \geq s \\ \Pi F^{(s-r)'} & r \leq s, \end{cases}$$

and

$$\gamma_Y(r-s) = E(Y_r Y_s') = \begin{cases} HF^{r-s-1}N & r > s \\ R + H\Pi H' & r = s \\ N'F^{(s-r-1)'}H' & r < s, \end{cases} \tag{5.13}$$

where $N = F\Pi H' + GS = \mathrm{Cov}(x_{t+1}, Y_t)$. Thus, the covariance generating function of $\{Y_t\}$, $G_Y(z) = \sum_{k=-\infty}^{\infty} \gamma_Y(k) z^k$, is given by

$$\begin{aligned} G_Y(z) &= (R + H\Pi H') + \sum_{k=1}^{\infty} HF^{k-1}Nz^k + \sum_{k=1}^{\infty} N'(F')^{-k-1}H'z^k \\ &= (R + H\Pi H') + zH(I - Fz)^{-1}N + z^{-1}N'(I - F'z^{-1})^{-1}H'. \end{aligned} \quad (5.14)$$

This can be written in matrix form as

$$G_Y(z) = [zH(I - Fz)^{-1} \quad I] \begin{bmatrix} 0 & N \\ N' & R + H\Pi H' \end{bmatrix} \begin{bmatrix} z^{-1}(I - F'z^{-1})^{-1}H' \\ I \end{bmatrix}. \quad (5.15)$$

Note that if we use generating functions in (5.4) and (5.5), we get

$$G_Y(z) = [zH(I - Fz)^{-1} \quad I] \begin{bmatrix} GQG' & GS \\ S'G' & R \end{bmatrix} \begin{bmatrix} z^{-1}(I - F'z^{-1})^{-1}H' \\ I \end{bmatrix}. \quad (5.16)$$

Remark 5.2 The formula (5.14) for $G_Y(z)$ implies that the covariance Hankel matrix of order r,

$$G_r = \begin{bmatrix} \gamma(1) & \gamma(2) & \gamma(3) & \cdots & \gamma(r) \\ \gamma(2) & \gamma(3) & \gamma(4) & \cdots & \gamma(r+1) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \gamma(r) & \gamma(r+1) & \gamma(t+2) & \cdots & \gamma(2r-1) \end{bmatrix},$$

can be expressed as

$$G_r = \begin{bmatrix} H \\ HF \\ \vdots \\ HF^{r-1} \end{bmatrix} \left[N, FN, \ldots, F^{r-1}N\right]. \quad (5.17)$$

Consider the VARMA(p, q) model (3.8) and its Akaike's state space form (3.13) and (3.14), where F and K are given by (3.11) and $H = [I, 0, \ldots, 0]$. Then, it is not difficult to verify that the observability matrix, $[H', F'H', \ldots, (F')^{r-1}H']'$, where $r = \max\{p, q\}$, is the unit matrix and thus from the definition of G_r and (5.17) it follows that

$$\begin{aligned} N &= \mathrm{Cov}(x_{t+1}, Y_t) \\ &= F\Pi H' + GS \\ &= \begin{bmatrix} \gamma(1) \\ \vdots \\ \gamma(r) \end{bmatrix} \end{aligned}$$

in this case. ◇

5.5 Computation of the Covariance Function

To compute the covariance function of (5.4) and (5.5), assuming that the processes $\{x_t : t \geq 1\}$ and $\{Y_t : t \geq 1\}$ are stationary, according to Lemma 5.2, we must first solve the Lyapunov equation

$$\Pi = F\Pi F' + GQG'.$$

Then, we compute $N = F\Pi H' + GS$ and the covariances according to (5.13).

A numerically efficient procedure to solve any Lyapunov equation of the form $\Pi = F\Pi F' + Q$ consists of first finding the Schur form of F, i.e. an orthogonal matrix O such that $OFO' = T$, where T is a block upper triangular matrix, and then solving $M = TMT' + OQO'$ for $M = O\Pi O'$ using a procedure similar to back substitution. See, e.g., Hammarling (1991).

5.6 Factorization of the Covariance Generating Function

Consider again the time invariant state space model (5.4) and (5.5) and suppose that F has all its eigenvalues inside the unit circle (F is stable) and $x_1 \sim (0, \Pi)$, where Π is the unique solution of the Lyapunov equation (5.7). Then, by Lemma 5.1, the processes $\{x_t : t \geq 1\}$ and $\{Y_t : t \geq 1\}$ are stationary.

The covariance generating function of $\{Y_t\}$, $G_Y(z) = \sum_{k=-\infty}^{\infty} \gamma_Y(k) z^k$, is given by (5.16). The covariance factorization problem consists of finding a unique positive definite matrix Σ and a unique square matrix function $\Psi(z)$ satisfying $G(z) = \Psi(z)\Sigma\Psi'(z^{-1})$ and the two properties of Theorem 3.12.

To gain some insight into this problem, suppose that we can find matrices Σ and K such that (5.16) factorizes as

$$\begin{aligned} G_Y(z) &= [zH(I-Fz)^{-1} \quad I]\begin{bmatrix} K \\ I \end{bmatrix}\Sigma\,[K' \quad I]\begin{bmatrix} z^{-1}(I-F'z^{-1})^{-1}H' \\ I \end{bmatrix} \\ &= [I+zH(I-Fz)^{-1}K]\Sigma\,[I+z^{-1}K'(I-F'z^{-1})^{-1}H']. \end{aligned} \tag{5.18}$$

Then, we can obtain the desired covariance factorization by defining $\Psi(z) = I + zH(I-Fz)^{-1}K$. By the matrix inversion lemma, the inverse of $\Psi(z)$, if it exists, is given by

$$\Psi^{-1}(z) = I - zH(I-F_pz)^{-1}K, \tag{5.19}$$

where $F_p = F - KH$. It is not difficult to see that the inverse exists if, and only if, F_p is stable.

The following lemma will be useful to obtain the factorization.

Lemma 5.6 *For any symmetric matrix V, the generating function (5.16) is invariant under transformations of the form*

$$\begin{bmatrix} GQG' & GS \\ S'G' & R \end{bmatrix} \longrightarrow \begin{bmatrix} GQG'-V+FVF' & GS+FVH' \\ S'G'+HVF' & R+HVH' \end{bmatrix}.$$

Proof Given a symmetric matrix V, the lemma will be proved if we prove

$$[H(z^{-1}I-F)^{-1} \quad I]\begin{bmatrix} -V+FVF' & FVH' \\ HVF' & HVH' \end{bmatrix}\begin{bmatrix} (zI-F')^{-1}H' \\ I \end{bmatrix} = 0.$$

The left-hand side of the previous equality can be written as

$$\begin{aligned} &\left[H(z^{-1}I-F)^{-1}(-V+FVF')+HVF', H(z^{-1}I-F)^{-1}FVH'+HVH'\right] \\ &\times\begin{bmatrix} (zI-F')^{-1}H' \\ I \end{bmatrix} \\ &= \left[H(z^{-1}I-F)^{-1}(-V+FVF')+HVF'\right](zI-F')^{-1}H' + H(z^{-1}I-F)^{-1}FVH' \\ &\quad +HVH' \\ &= H(z^{-1}I-F)^{-1}\left[(-V+FVF')(zI-F')^{-1}H'\right. \\ &\quad \left.+(z^{-1}I-F)VF'(zI-F')^{-1}H'+FVH'\right] + H(z^{-1}I-F)^{-1}(z^{-1}I-F)VH' \\ &= H(z^{-1}I-F)^{-1}\left[-V(zI-F')^{-1}H' + FVF'(zI-F')^{-1}H'\right. \\ &\quad \left.+(z^{-1}I-F)VF'(zI-F')^{-1}H' + FVH' + (z^{-1}I-F)VH'\right] \end{aligned}$$

$$= H(z^{-1}I - F)^{-1}\left[-V(zI - F')^{-1}H' + FVF'(zI - F')^{-1}H' + z^{-1}VF'(zI - F')^{-1}H'\right.$$
$$\left.-FVF'(zI - F')^{-1}H' + FVH' + z^{-1}VH' - FVH'\right]$$
$$= H(z^{-1}I - F)^{-1}\left[-V + z^{-1}VF' + z^{-1}V(zI - F')\right](zI - F')^{-1}H'$$
$$= 0.$$

□

By Lemma 5.6, it seems natural to try to select a matrix P such that

$$\begin{bmatrix} -P + FPF' + GQG' & FPH' + GS \\ HPF' + S'G' & R + HPH' \end{bmatrix} = \begin{bmatrix} K \\ I \end{bmatrix} \Sigma \, [K' \quad I], \tag{5.20}$$

and $F_p = F - KH$ is stable. Equating terms in (5.20), it is seen that the matrices K and Σ should satisfy

$$\begin{aligned} \Sigma &= R + HPH' \\ K &= (FPH' + GS)\Sigma^{-1} \\ P &= FPF' + GQG' - K\Sigma K' \end{aligned} \tag{5.21}$$

and that the selected matrix, P, should satisfy the discrete algebraic Riccati equation (DARE)

$$P = FPF' + GQG' - (FPH' + GS)(R + HPH')^{-1}(FPH' + GS)'. \tag{5.22}$$

It turns out that, under some rather general assumptions specified in Sect. 5.12.1, there exists a unique positive semidefinite solution, P, of the DARE (5.22) such that Σ is nonsingular and F_p is stable. Such a solution is called a **stabilizing solution**. If M is a real symmetric matrix, we will sometimes use in the following the notation $M \geq 0$ to indicate that M is positive semidefinite.

As we will see later in Theorem 7.4, the matrix P in (5.22) is the MSE matrix of the estimated state in the steady state Kalman filter of (5.4) and (5.5). That is, if $\hat{x}_{t+1|t}$ is the one-period-ahead predictor of x_{t+1} based on the semi-infinite sample $\{Y_s : s \leq t\}$, the following recursion holds

$$\hat{x}_{t+1|t} = F\hat{x}_{t|t-1} + KA_t, \tag{5.23}$$
$$Y_t = H\hat{x}_{t|t-1} + A_t, \qquad t > -\infty, \tag{5.24}$$

and the solution P of the DARE (5.22) satisfies $P = \mathrm{Var}(x_t - \hat{x}_{t|t-1})$. Note the relation

$$E(x_t x_t') = E[(x_t - \hat{x}_{t|t-1})(x_t - \hat{x}_{t|t-1})'] + E(\hat{x}_{t|t-1}\hat{x}_{t|t-1}')$$

that in turn follows from the fact that $x_t - \hat{x}_{t|t-1}$ is orthogonal to $\hat{x}_{t|t-1}$. Then, letting $\mathrm{Var}(\hat{x}_{t|t-1}) = \overline{\Sigma}$, it holds that

$$\Pi = P + \overline{\Sigma}. \tag{5.25}$$

If we select the matrix $V = \Pi - \overline{\Sigma} = P$ in Lemma 5.6, we get, by (5.20), the relation

$$\begin{bmatrix} \overline{\Sigma} - F\overline{\Sigma}F' & -F\overline{\Sigma}H' + N \\ -H\overline{\Sigma}F' + N' & R + H\Pi H' - H\overline{\Sigma}H' \end{bmatrix} = \begin{bmatrix} K \\ I \end{bmatrix} \Sigma \, [K' \quad I]. \tag{5.26}$$

Equating terms and considering that $\gamma(0) = R + H\Pi H'$ yields

$$\begin{aligned} \Sigma &= \gamma(0) - H\overline{\Sigma}H' \\ K &= (N - F\overline{\Sigma}H')\Sigma^{-1} \\ \overline{\Sigma} &= F\overline{\Sigma}F' + K\Sigma K'. \end{aligned} \tag{5.27}$$

Thus, the matrix $\overline{\Sigma}$ satisfies the DARE

$$\overline{\Sigma} = F\overline{\Sigma}F' + (N - F\overline{\Sigma}H')[\gamma(0) - H\overline{\Sigma}H']^{-1}(N - F\overline{\Sigma}H')'. \tag{5.28}$$

It follows from (5.25), (5.21), and (5.27) that the DARE (5.22) has a solution P if, and only if, the DARE (5.28) has a solution $\overline{\Sigma}$.

Finally, if we set $V = \Pi$ in Lemma 5.6, we get the center in (5.15), that is,

$$\begin{bmatrix} -\Pi + F\Pi F' + GQG' & F\Pi H' + GS \\ H\Pi F' + S'G' & R + HPH' \end{bmatrix} = \begin{bmatrix} 0 & N \\ N' & R + H\Pi H' \end{bmatrix}, \tag{5.29}$$

because Π satisfies the Lyapunov equation (5.7).

Conditions for the convergence of the Kalman filter recursions to the steady state recursions (5.23) and (5.24) will be given in Sect. 5.12. By the previous results, these conditions will also apply for the convergence to the steady state of the recursions (1.41), (1.42) and (1.43) of Corollary 1.4 when the covariances are those of the state space model (5.4) and (5.5), given by (5.13).

Consider the VARMA(p, q) model (3.8) and its Akaike's state space form (3.13) and (3.14), where F and K are given by (3.11) and $H = [I, 0, \ldots, 0]$. This corresponds to setting $u_t = v_t = A_t$, $K = G$, and $R = S = Q = \mathrm{Var}(A_t)$ in (5.4) and (5.5). Therefore, if we define $\Sigma = \mathrm{Var}(A_t)$, we see that (5.20) holds with $P = 0$ and the factorization (5.18) obtains. Since $\Psi(z) = I + zH(I - Fz)^{-1}K = \Phi^{-1}(z)\Theta(z) = \sum_{i=0}^{\infty} \Psi_i z^i$, $G_Y(z)$ can be written as $G_Y(z) = \Psi(z)\Sigma\Psi'(z^{-1})$. In addition, by Remark 5.2, it holds that $\gamma(j) = HF^{j-1}N$ and

$$N = \mathrm{Cov}(x_{t+1}, Y_t) = \begin{bmatrix} \gamma(1) \\ \vdots \\ \gamma(r) \end{bmatrix}.$$

If we only knew the covariance sequence, $\{\gamma(j)\}$, and the autoregressive polynomial, $\Phi(z)$, of Y_t we could obtain the vector K and $\Sigma = \mathrm{Var}(A_t)$ by first solving the DARE (5.28) and then using (5.27).

Example 5.3 Suppose that $\{Y_t\}$ follows the MA(q) process

$$Y_t = \Theta(B)A_t,$$

where $\Theta(z) = I + \Theta_1 z + \cdots + \Theta_q z^q$ and $\mathrm{Var}(A_t) = \Sigma$, and we are interested in obtaining the moving average polynomial matrix, $\Theta(z)$, and the covariance matrix of the innovations, Σ, using the covariances, $\gamma(0)$,$\gamma(1), \ldots, \gamma(q)$, only. In this case, the system matrices are given by

$$F = \begin{bmatrix} 0 & I & 0 & \cdots & 0 \\ 0 & 0 & I & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & I \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix}, \qquad N = \begin{bmatrix} \gamma(1) \\ \vdots \\ \gamma(q) \end{bmatrix},$$

and $H = [I, 0, \ldots, 0]$. Solving the DARE (5.28), we get $\Sigma = \gamma(0) - F\overline{\Sigma}F'$ and

$$K = (N - F\overline{\Sigma}H')\Sigma^{-1} = \begin{bmatrix} \Theta_1 \\ \Theta_2 \\ \vdots \\ \Theta q \end{bmatrix}.$$

From this, we obtain the covariance factorization

$$G_Y(z) = \Theta(z)\Sigma\Theta'(z^{-1}).$$

By the properties of the DARE, that we will see in Sect. 5.12, the procedure guarantees that all the roots of the polynomial $\det[\Theta(z)]$ are on or outside the unit circle and Σ is positive definite. ◇

Remark 5.3 There has been a lot of research in connection with the DARE (5.22) and efficient and reliable software exists to solve this equation. As just mentioned, the solution to the factorization problem of an MA(q) process can be obtained by solving the DARE (5.28). It is easy to verify that this can be achieved by reducing the Eq. (5.28) to the form (5.22) by defining $Q = 0$, $S = -I$, $R = \gamma(0)$ and $G = [\gamma(1)', \ldots, \gamma(r)']'$. ◇

5.7 Cointegrated VARMA Models

Let the k-dimensional process $\{Y_t\}$ follow the VARMA model

$$\Phi(B)Y_t = \Theta(B)A_t, \tag{5.30}$$

where $\Phi(z) = \Phi_0 + \Phi_1 z + \cdots + \Phi_l z^l$, $\Theta(z) = \Theta_0 + \Theta_1 z + \cdots + \Theta_l z^l$, $\Theta_0 = \Phi_0$, Φ_0 is lower triangular with ones in the main diagonal, the roots of $\det[\Theta(z)]$ are all outside the unit circle, and $\det[\Phi(z)] = 0$ implies $|z| > 1$ or $z = 1$. The fact that both $\Phi(z)$ and $\Theta(z)$ are supposed to be matrix polynomials with the same degree does not imply loss of generality because we can always complete one of the matrix polynomials with zero matrices if that is not the case. The model (5.30) can be in echelon form, to be described later in this chapter, or not. The following discussion is not affected by this.

We assume that the matrix Π, defined by

$$\Pi = -\Phi(1),$$

has rank r such that $0 < r < k$ and that there are exactly $k - r$ roots in the model equal to one. When the model (5.30) satisfies these two conditions, it is called a **cointegrated VARMA model** with **cointegration rank** equal to r.

Under the previous assumptions, the Π matrix can be expressed (non uniquely) as

$$\Pi = \alpha\beta',$$

where α and β are $k \times r$ of rank r. Let $\beta_\perp$ be a $k \times (k-r)$ matrix of rank $k-r$ such that

$$\beta'\beta_\perp = 0_{r\times(k-r)}$$

and define the matrix P as

$$P = [P_1, P_2] = [\beta_\perp, \beta].$$

Then, it is not difficult to verify that $Q = P^{-1}$ is given by

$$Q = \begin{bmatrix} Q_1 \\ Q_2 \end{bmatrix} = \begin{bmatrix} (\beta'_\perp \beta_\perp)^{-1}\beta'_\perp \\ (\beta'\beta)^{-1}\beta' \end{bmatrix}$$

and that if we further define $U_1 = P_1 Q_1$ and $U_2 = P_2 Q_2$, the following relations hold

$$U_1 + U_2 = I_k, \qquad U_1 U_2 = U_2 U_1 = 0. \tag{5.31}$$

Thus, we can write

$$I_k - zI_k = (I_k - U_1 z)(I_k - U_2 z) = (I_k - U_2 z)(I_k - U_1 z). \tag{5.32}$$

The so-called **error correction form** corresponding to model (5.30) is

$$\Gamma(B)\nabla Y_t = \Pi Y_{t-1} + \Theta(B)A_t, \tag{5.33}$$

where $\nabla = I_k - BI_k$, $\Gamma(z) = \Gamma_0 + \sum_{i=1}^{l-1} \Gamma_i z^i$, and the Γ_i matrices are defined by $\Gamma_0 = \Phi_0$ and

$$\Gamma_i = -\sum_{j=i+1}^{l} \Phi_j, \qquad i = 1, \ldots, l-1.$$

It follows from (5.33) that $\beta' Y_{t-1}$ is stationary because all the terms in this equation different from $\Pi Y_{t-1} = \alpha\beta' Y_{t-1}$ are stationary. Therefore, there are r so-called **cointegration relations** in the model given by $\beta' Y_t$.

Considering (5.31) and (5.32), the following relation between the autoregressive polynomials in (5.30) and (5.33) holds

$$\begin{aligned} \Phi(z) &= \Gamma(z)(I_k - zI_k) - \Pi z \\ &= [\Gamma(z)(I_k - U_2 z) - \Pi z]\,(I_k - U_1 z) \end{aligned}$$

because $\Pi U_1 = 0$. Thus, defining $\Phi^*(z) = \Gamma(z)(I_k - U_2 z) - \Pi z$ and $D(z) = I_k - U_1 z$, we can write $\Phi(z)$ as

$$\Phi(z) = \Phi^*(z)D(z) \tag{5.34}$$

and the model (5.30) as

$$\Phi^*(B)D(B)Y_t = \Theta(B)A_t. \tag{5.35}$$

Since both $U_1 = \beta_\perp(\beta'_\perp\beta_\perp)^{-1}\beta'_\perp$ and $U_2 = \beta(\beta'\beta)^{-1}\beta'$ are idempotent and symmetric matrices of rank $k-r$ and r, respectively, the eigenvalues of these two matrices are all equal to one or zero. In particular,

$$\det(I_k - U_1 z) = (1-z)^{k-r}$$

and therefore, the matrix polynomial $D(z) = I_k - U_1 z$ in (5.34) is a **"differencing" matrix polynomial** because it contains all the unit roots in the model. This implies in turn that the matrix polynomial $\Phi^*(z)$ in (5.34) has all its roots outside the unit circle and the series $\{D(B)Y_t\}$ in (5.35) is stationary. Thus, the matrix polynomial $\Phi^*(z)$ can be inverted so that the following relation holds

$$D(B)Y_t = \left[\Phi^*(B)\right]^{-1}\Theta(B)A_t.$$

Premultiplying the previous expression by $\beta'_\perp$, we can see that there are $k-r$ linear combinations of Y_t that are $I(1)$ given by $\beta'_\perp Y_t$. In a similar way, premultiplying by β', it follows as before that there are r linear combinations of Y_t that are $I(0)$ given by $\beta' Y_t$.

The series $\{D(B)Y_t\}$ can be considered as the **"differenced series,"** and the notable feature of (5.35) is that the model followed by $\{D(B)Y_t\}$ is stationary. Therefore, we can specify and estimate a stationary VARMA model if we know the series $\{D(B)Y_t\}$.

The matrix U_1 can be parameterized as follows. Let

$$\beta_\perp = \begin{bmatrix}\beta_1\\ \beta_2\end{bmatrix},$$

where β_1 is a $(k-r)\times(k-r)$ matrix and suppose that β_1 is nonsingular. Then, we can write

$$\beta_\perp = \begin{bmatrix} I_{k-r}\\ \beta_2\beta_1^{-1}\end{bmatrix}\beta_1$$

and $U_1 = \overline{\beta}_\perp(\overline{\beta}'_\perp\overline{\beta}_\perp)^{-1}\overline{\beta}'_\perp$, where

$$\overline{\beta}_\perp = \begin{bmatrix} I_{k-r}\\ \beta_2\beta_1^{-1}\end{bmatrix}.$$

This means that U_1 can be parameterized in terms of the $k\times(k-r)$ matrix $\beta_2\beta_1^{-1}$. In practice, it may happen that the matrix β_1 is singular. However, since $\beta_\perp$ has full column rank, there is always a nonsingular $(k-r)\times(k-r)$ submatrix of $\beta_\perp$ and we would proceed with this submatrix in a way similar to the one just described in terms of β_1 to parameterize U_1.

5.7.1 Parametrizations and State Space Forms

We can consider two ways to handle cointegrated VARMA models. The first one parameterizes model (5.33) in terms of the matrix polynomials $\Gamma(z)$ and $\Theta(z)$ and the matrices α and $\beta_{\perp}$, where this latter matrix is parameterized as described in the previous paragraph. The second one parameterizes model (5.35) in terms of the matrix polynomials $\Phi^{*}(z)$ and $\Theta(z)$ and the matrix $\beta_{\perp}$. The advantage of the latter parametrization is that we can specify a stationary VARMA model in echelon form for the "differenced" series by directly specifying $\Phi^{*}(z)$ and $\Theta(z)$. There is no need for a reverse echelon form considered by some authors (Lütkepohl, 2007).

If a model is parameterized in terms of the matrix polynomials $\Gamma(z)$ and $\Theta(z)$ and the matrices α and $\beta_{\perp}$ corresponding to the error correction model (5.33), one can first obtain the matrix polynomial $\Phi(z)$ as $\Phi(z) = \Gamma(z)(I_k - zI_k) - \Pi z$ and then set up a state space model corresponding to (5.30). The initial conditions can be obtained following the procedure described in Sect. 4.14.2 and taking into account that the number of unit roots is known.

On the other hand, if the model is parameterized in terms of the matrix polynomials $\Phi^{*}(z)$ and $\Theta(z)$ and the matrix $\beta_{\perp}$ corresponding to the model for the "differenced" series (5.35), the state space model can be set up directly in terms of (5.35). That is, the data are $\{D(B)Y_t\}$ in the observation equation and the system matrices correspond to the VARMA model (5.35).

5.7.2 Forecasting

Once a cointegrated VARMA model has been estimated, one can obtain forecasts with this model as described in Sect. 5.18. To this end, it is advantageous to use the state space model corresponding to the model (5.33), where the matrix polynomial $\Phi(z)$ is given by $\Phi(z) = \Gamma(z)(I_k - zI_k) - \Pi z$ or (5.34), depending on the parametrization used. As mentioned earlier, the initial conditions can be obtained following the procedure described in Sect. 4.14.2 and taking into account that the number of unit roots is known.

5.8 The Likelihood of a Time Invariant State Space Model

Suppose the state space model (5.4) and (5.5), where the processes $\{x_t\}$ and $\{Y_t\}$ can be stationary or not. As described in Sect. 4.14.2, whether these processes are stationary or not depends on the eigenvalues of the F matrix. If the eigenvalues of F are all inside the unit circle, by Lemma 5.2, the processes are stationary if, and only if, the Lyapunov equation (5.7) is satisfied.

Letting x_1 be given by (5.3), we can use the TSKF-SRIBF to evaluate the likelihood. According to Sect. 4.20.1 and with the notation of that section, the concentrated diffuse log-likelihood (4.95) is given by

$$\lambda_D(Y;\widehat{\sigma}^2) = \text{constant} - \frac{1}{2}\left\{(np - n_\delta)\ln\left[\sum_{t=1}^{n}\left(\widehat{\Sigma}_t^{-1/2}\widehat{E}_t\right)'\left(\widehat{\Sigma}_t^{-1/2}\widehat{E}_t\right)\right]\right.$$
$$\left. + \sum_{t=1}^{n}\ln|\Sigma_t| + \ln|R_{dd,n+1}|^2\right\}.$$

5.9 Canonical Forms for VARMA and State Space Models

In this section we will suppose a process $\{Y_t\}$ that follows a finite linear time series model (3.29) such that $\Psi_0 = I$ and the rank of the augmented Hankel matrices, H_t^a, given by (3.32), is constant for $t > r$, where r is a fixed positive integer. The process can be stationary or not. The following development will not be affected by this. Let $\{n_i : i = 1, \ldots, k\}$ be the Kronecker indices and $n = \sum_{i=1}^{k} n_i$ the McMillan degree corresponding to this model.

5.9.1 VARMA Echelon Form

By the definition of the i-th Kronecker index and the structure of the augmented Hankel matrices, there exists a vector $\phi_i = [\phi_{i,n_i}, \ldots, \phi_{i,1}, \phi_{i,0}]$, where the $\phi_{i,j}$ have dimension $1 \times k$, $j = 0, 1, \ldots, n_i$, and $\phi_{i,0}$ has a one in the i-th position and zeros thereafter, such that

$$[0_{1\times k}, \ldots, 0_{1\times k}, \phi_{i,n_i}, \ldots, \phi_{i,1}, \phi_{i,0}, 0_{1\times k}, \ldots, 0_{1\times k}]H_t^a = 0, \quad t > n_i. \tag{5.36}$$

Note that the vector ϕ_i can be moved either to the left or to right in the previous expression without altering the relation due to the structure of H_t^a, $t > n_i$. This implies that if $l = \max\{n_i : i = 1, \ldots, k\}$, there exists a block vector $\Phi = [\Phi_l, \ldots, \Phi_1, \Phi_0]$ with Φ_0 a lower triangular matrix with ones in the main diagonal such that

$$\Phi H_{l+1}^a = 0. \tag{5.37}$$

In fact, the i-th row of Φ is $[0, \phi_i]$ if $n_i < l$ and ϕ_i if $n_i = l$, $i = 1 \ldots, k$. It follows from (5.37) that if we stack the observations to get

$$Y_{t:t+l} = \widehat{H}_{l+1}\alpha_1 + \widehat{\Psi}_{l+1}A_{t:t+l}, \tag{5.38}$$

where $Y_{t:t+l} = (Y_t', \ldots, Y_{t+l}')'$, $A_{t:t+l} = (A_1', \ldots, A_{t+l}')'$,

$$\widehat{H}_{l+1} = \begin{bmatrix} h_t \\ h_{t+1} \\ \vdots \\ h_{t+l} \end{bmatrix}, \quad \text{and} \quad \widehat{\Psi}_{l+1} = \begin{bmatrix} \Psi_{t-1} & \cdots & \Psi_1 & I & & & & \\ \Psi_t & \cdots & \Psi_2 & \Psi_1 & I & & & \\ \vdots & \cdots & \vdots & \vdots & \vdots & \ddots & I & \\ \Psi_{t+l-1} & \cdots & \Psi_{l+1} & \Psi_l & \Psi_{l-1} & \cdots & \Psi_1 & I \end{bmatrix},$$

and we premultiply (5.38) by Φ, the following VARMA model is obtained

$$\Phi(B)Y_t = \Theta(B)A_t, \tag{5.39}$$

where $\Phi(z) = \Phi_0 + \Phi_1 z + \cdots + \Phi_l z^l$, $\Theta(z) = \Theta_0 + \Theta_1 z + \cdots + \Theta_l z^l$, $\Theta_0 = \Phi_0$, and the Θ_i, $i = 0, 1, \ldots, l$, are given by the product of Φ and the last $l + 1$ blocks of columns of $\widehat{\Psi}_{l+1}$. More specifically,

$$\Theta_j = \Phi_j + \Phi_{j-1}\Psi_1 + \cdots + \Phi_0\Psi_j, \quad j = 0, 1, \ldots, l. \tag{5.40}$$

The VARMA model (5.39) is called the VARMA canonical form or the **VARMA echelon form**.

Given the structure of the rows of the matrix Φ, if $j > n_i$, the i-th row of Φ_j is zero, and if $j \leq n_i$, the i-th row of Φ_j is $\phi_{i,j}$. Since, by (5.36), the i-th row of $\Phi_{j-1}\Psi_1 + \cdots + \Phi_0\Psi_j$ is zero if $j > n_i$, it follows that, by (5.40), the i-th row of Θ_j is zero if $j > n_i$.

The relation (5.36) expresses the fact that the i-th row of the $(n_i + 1)$-th block of rows depends linearly on the rows of the basis of Proposition 3.8. If any of the rows preceding the i-th row of the $(n_i + 1)$-th block of rows depends linearly on the rows of the basis, this row can be eliminated and the corresponding coefficient in ϕ_i can be made zero. Thus, if $p < i$ and $n_p \leq n_i$, all of the p-th elements of $\phi_{i,0}$, $\phi_{i,1}, \ldots, \phi_{i,n_i-n_p}$ are zero. If $p > i$ and $n_p < n_i$, all of the p-th elements of $\phi_{i,1}, \ldots, \phi_{i,n_i-n_p}$ are zero. Also, by the structure of $\phi_{i,0}$, if $p > i$, the p-th element of $\phi_{i,0}$ is zero.

Letting $\phi_{ip,j}$ and $\theta_{ip,j}$ be the (i,p)-th elements in the matrices Φ_j and Θ_j, $i, p = 1, \ldots, k$, $j = 0, 1, \ldots, l$, and taking into account all the previous relations among the coefficients in Φ_j and Θ_j, we can express the matrix polynomials

$$\Phi(z) = \left[\begin{array}{c|c|c|c|c} \phi_{11}(z) & \cdots & \phi_{1i}(z) & \cdots & \phi_{1k}(z) \\ \hline \vdots & \ddots & \vdots & \ddots & \vdots \\ \hline \phi_{i1}(z) & \cdots & \phi_{ii}(z) & \cdots & \phi_{ik}(z) \\ \hline \vdots & \ddots & \vdots & \ddots & \vdots \\ \hline \phi_{k1}(z) & \cdots & \phi_{ki}(z) & \cdots & \phi_{kk}(z) \end{array}\right] = \Phi_0 + \Phi_1 z + \cdots + \Phi_l z^l$$

and

$$\Theta(z) = \left[\begin{array}{c|c|c|c|c} \theta_{11}(z) & \cdots & \theta_{1i}(z) & \cdots & \theta_{1k}(z) \\ \hline \vdots & \ddots & \vdots & \ddots & \vdots \\ \hline \theta_{i1}(z) & \cdots & \theta_{ii}(z) & \cdots & \theta_{ik}(z) \\ \hline \vdots & \ddots & \vdots & \ddots & \vdots \\ \hline \theta_{k1}(z) & \cdots & \theta_{ki}(z) & \cdots & \theta_{kk}(z) \end{array}\right] = \Theta_0 + \Theta_1 z + \cdots + \Theta_l z^l$$

as follows

$$\phi_{ii}(z) = 1 + \sum_{j=1}^{n_i} \phi_{ii,j} z^j, \quad i = 1, \ldots, k, \tag{5.41}$$

$$\phi_{ip}(z) = \sum_{j=n_i-n_{ip}+1}^{n_i} \phi_{ip,j} z^j, \quad i \neq p \tag{5.42}$$

$$\theta_{ip}(z) = \sum_{j=0}^{n_i} \theta_{ip,j} z^j, \quad i, p = 1, \ldots, k, \tag{5.43}$$

where $\Theta_0 = \Phi_0$ and

$$n_{ip} = \begin{cases} \min\{n_i + 1, n_p\} & \text{for } i > p \\ \min\{n_i, n_p\} & \text{for } i < p \end{cases} \quad i, p = 1, \ldots, k.$$

Note that n_{ip} specifies the number of free coefficients in the polynomial $\phi_{ip}(z)$ for $i \neq p$.

Proposition 5.1 *The VARMA echelon form (5.39) is left coprime.*

Proof Assume that (5.39) is not left coprime and premultiply (5.39) by Φ_0^{-1} to get

$$\widetilde{\Phi}(B)Y_t = \widetilde{\Theta}(B)A_t,$$

where $\widetilde{\Phi}(0) = I$ and $\widetilde{\Theta}(0) = I$. Then, we can cancel some common nonunimodular left factor to get left coprime matrix polynomials $\widehat{\Phi}(z)$ and $\widehat{\Theta}(z)$ with degrees of $\det[\widehat{\Phi}(z)]$ and $\det[\widehat{\Theta}(z)]$ strictly less than degrees of $\det[\Phi(z)]$ and $\det[\Theta(z)]$, respectively, $\widehat{\Phi}(0) = I$ and $\widehat{\Theta}(0) = I$. Premultiplying (5.38) by the matrix $\widehat{\Phi} = [\Phi_0\widehat{\Phi}_l, \ldots, \Phi_0\widehat{\Phi}_1, \Phi_0]$, where $\widehat{\Phi}_j = 0$ if $j > s$ and $\widehat{\Phi}(z) = \sum_{i=0}^{s} \widehat{\Phi}_i z^i$, we obtain $\widehat{\Phi}H_{l+1} = 0$ and $\widehat{\Phi}H_{l+1}^a = 0$. Since the degree of $\det[\widehat{\Phi}(z)]$ is less than the degree of $\det[\Phi(z)]$, the expression $\widehat{\Phi}H_{l+1}^a = 0$ implies a simplification in the unique representation, $\Phi H_{l+1}^a = 0$, of the rows of the $(l+1)$-th block of rows of H_{l+1}^a as a linear combination of the previous rows implied by the Kronecker indices, and this is a contradiction. □

Remark 5.4 It has become customary in the statistical and econometric literature to use the term VARMA echelon form (Lütkepohl, 2007; Reinsel, 1997) when referring to the canonical form of a VARMA model expressed in terms of the backshift operator. This canonical form has been sometimes called reversed echelon form in the engineering literature. See, for example, Hannan & Deistler (1988). However, these forms are best suited for forecasting purposes and have not been used predominantly in the engineering literature. Instead, engineers often prefer to work with the forward operator and the corresponding so-called canonical matrix fraction descriptions (MFDs), see the Appendix to this chapter. These canonical MFDs have some advantages with respect to VARMA echelon forms. For example, as shown in the Appendix to this chapter, the McMillan degree is always the determinantal degree of the denominator matrix of the canonical MFD. See also Kailath (1980) and Hannan & Deistler (1988). ◇

5.9.2 State Space Echelon Form

Suppose that $\{Y_t\}$ follows the finite linear time series model (3.29) and define the forecasts as

$$Y_{t+i|t} = Y_{t+i} - A_{t+i} - \Psi_1 A_{t+i-1} + \cdots + \Psi_{i-1} A_{t+1}, \quad i = 1, 2, \ldots . \tag{5.44}$$

Then, we can write

$$\begin{bmatrix} Y_{t+1|t} \\ Y_{t+2|t} \\ \vdots \\ Y_{t+i|t} \end{bmatrix} = \left[\begin{array}{cccc|c} \Psi_1 & \Psi_2 & \cdots & \Psi_t & h_{t+1} \\ \Psi_2 & \Psi_3 & \cdots & \Psi_{t+1} & h_{t+2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \Psi_i & \Psi_{i+1} & \cdots & \Psi_{t+i-1} & h_{t+i} \end{array}\right] \left[\begin{array}{c} A_t \\ A_{t-1} \\ \vdots \\ A_1 \\ \hline \alpha_1 \end{array}\right], \tag{5.45}$$

and, in particular, for $i = t$

$$\begin{bmatrix} Y_{t+1|t} \\ \vdots \\ Y_{2t|t} \end{bmatrix} = \left[\begin{array}{c|c} H_t & \begin{matrix} h_{t+1} \\ \vdots \\ h_{2t} \end{matrix} \end{array}\right] \left[\begin{array}{c} A_t \\ \vdots \\ A_1 \\ \hline \alpha_1 \end{array}\right].$$

Thus, we see that relations among rows of the augmented Hankel matrices, H_t^a, are equivalent to relations among forecasts. The following expression, that is a direct consequence of (5.44) and (5.45), is also useful.

$$\begin{bmatrix} Y_t \\ Y_{t+1|t} \\ Y_{t+2|t} \\ \vdots \\ Y_{t+i|t} \end{bmatrix} = \begin{bmatrix} I \\ \Psi_1 \\ \Psi_2 \\ \vdots \\ \Psi_i \end{bmatrix} A_t + \begin{bmatrix} Y_{t|t-1} \\ Y_{t+1|t-1} \\ Y_{t+2|t-1} \\ \vdots \\ Y_{t+i|t-1} \end{bmatrix}. \tag{5.46}$$

Let us assume first that the Kronecker indices satisfy $n_i \geq 1$, $i = 1, \ldots, k$. Then, it follows from (5.36) and (5.45) that

$$\phi_{i,0} Y_{t+n_i+1|t} + \phi_{i,1} Y_{t+n_i|t} + \cdots + \phi_{i,n_i} Y_{t+1|t} = 0.$$

This in turn implies, by (5.41) and (5.42), that

$$Y^i_{t+n_i+1|t} + \sum_{j=1}^{n_i} \phi_{ii,j} Y^i_{t+n_i+1-j|t} + \sum_{i \neq p} \sum_{j=n_i-n_{ip}+1}^{n_i} \phi_{ip,j} Y^p_{t+n_i+1-j|t} = 0, \tag{5.47}$$

where $Y^p_{t+j|t}$ denotes the p-th element of $Y_{t+j|t}$, $p = 1, \ldots, k$, $j = 1, 2, \ldots$. From (5.46) and (5.47), the following relations are obtained

$$Y^i_{t+j|t} = Y^i_{t+j|t-1} + \Psi_{i,j} A_t, \qquad j = 1, 2 \ldots, n_i - 1 \tag{5.48}$$

$$Y^i_{t+n_i|t} = -\sum_{j=1}^{n_i} \phi_{ii,j} Y^i_{t+n_i-j|t-1} - \sum_{i \neq p} \sum_{j=n_i-n_{ip}+1}^{n_i} \phi_{ip,j} Y^p_{t+n_i-j|t-1} + \Psi_{i,n_i} A_t, \tag{5.49}$$

where $\Psi_{i,j}$ denotes the i-th row of the matrix Ψ_j, $i = 1, 2, \ldots$.

By (5.45), to the basis of rows of the augmented Hankel matrices implied by the Kronecker indices and specified in Proposition 3.8 corresponds a basis of the space of forecasts. If we stack the elements of this basis of forecasts in the vector

$$x_{t+1} = \left[Y^1_{t+1|t}, \ldots, Y^1_{t+n_1|t}, Y^2_{t+1|t}, \ldots, Y^2_{t+n_2|t}, \ldots, Y^k_{t+1|t}, \ldots, Y^k_{t+n_k|t} \right]', \tag{5.50}$$

where $\dim(x_{t+1}) = \sum_{i=1}^k n_i$, by (5.46), (5.48) and (5.49), it is obtained that

$$x_{t+1} = F x_t + K A_t \tag{5.51}$$

$$Y_t = H x_t + A_t, \tag{5.52}$$

where

$$F = \begin{bmatrix} F_{11} & \cdots & F_{1i} & \cdots & F_{1k} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ F_{i1} & \cdots & F_{ii} & \cdots & F_{ik} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ F_{k1} & \cdots & F_{ki} & \cdots & F_{kk} \end{bmatrix}, \quad K = \begin{bmatrix} K_1 \\ \vdots \\ K_i \\ \vdots \\ K_k \end{bmatrix}, \quad K_i = \begin{bmatrix} \Psi_{i,1} \\ \vdots \\ \Psi_{i,n_i} \end{bmatrix},$$

$$F_{ii} = \begin{bmatrix} 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \\ -\phi_{ii,n_i} & \cdots & \cdots & -\phi_{ii,1} \end{bmatrix}, \quad F_{ip} = \begin{bmatrix} 0 & \cdots & 0 & 0 \cdots 0 \\ \vdots & \ddots & \vdots & \vdots \ddots \vdots \\ 0 & \cdots & 0 & 0 \cdots 0 \\ -\phi_{ip,n_i} & \cdots & -\phi_{ip,n_i-n_{ip}+1} & 0 \cdots 0 \end{bmatrix},$$

$$H = \left[\begin{array}{cccc|cccc|c|cccc} 1 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 \\ 0 & \cdots & 0 & 0 & 1 & \cdots & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 1 & \cdots & 0 & 0 \end{array}\right],$$

F_{ii} is $n_i \times n_i$, F_{ip} is $n_i \times n_p$, and H is $k \times (n_1 + \cdots + n_k)$. The state space form (5.51) and (5.52) is called the **state space echelon form**.

Proposition 5.2 *The state space echelon form (5.51) and (5.52) is minimal.*

Proof The result is a consequence of Proposition 3.9, since the dimension of x_t is equal to the McMillan degree. □

When some of the Kronecker indices are zero, the echelon form is different from the one we have just described. The situation is best illustrated with an example. Let $[3, 0, 2, 0]$ be the vector of Kronecker indices. Thus, $k = 4$ and the McMillan degree is $n = 5$. Letting $Y_t = (Y_{1t}, \ldots, Y_{4t})'$ and considering the second and fourth rows of the VARMA echelon form given by (5.41), (5.42), and (5.43), we get the equations

$$Y_{2t} + \phi_{21,0}Y_{1t} = A_{2t} + \phi_{21,0}A_{1t}$$
$$Y_{4t} + \phi_{41,0}Y_{1t} + \phi_{43,0}Y_{3t} = A_{4t} + \phi_{41,0}A_{1t} + \phi_{43,0}A_{3t}.$$

It follows from this that

$$\begin{aligned} Y_{2t} &= A_{2t} - \phi_{21,0}(Y_{1t} - A_{1t}) \\ &= A_{2t} - \phi_{21,0}Y^1_{t|t-1} \\ Y_{4t} &= A_{4t} - \phi_{41,0}(Y_{1t} - A_{1t}) - \phi_{43,0}(Y_{3t} - A_{3t}) \\ &= A_{4t} - \phi_{41,0}Y^1_{t|t-1} - \phi_{43,0}Y^3_{t|t-1}. \end{aligned}$$

Since the state vector is

$$x_{t+1} = \left[Y^1_{t+1|t}, Y^1_{t+2|t}, , Y^1_{t+3|t}, Y^3_{t+1|t}, Y^3_{t+2|t}\right]',$$

we have to modify the H matrix of the echelon form so that it becomes

$$H = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ -\phi_{21,0} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ -\phi_{41,0} & 0 & 0 & -\phi_{43,0} & 0 \end{bmatrix}.$$

The F and K matrices of the echelon form are

$$F = \left[\begin{array}{c|c} F_{11} & F_{13} \\ \hline F_{31} & F_{33} \end{array}\right], \quad K = \left[\begin{array}{c} K_1 \\ \hline K_3 \end{array}\right], \quad K_i = \begin{bmatrix} \Psi_{i,1} \\ \vdots \\ \Psi_{i,n_i} \end{bmatrix},$$

$$F_{ii} = \begin{bmatrix} 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \\ -\phi_{ii,n_i} & \cdots & \cdots & -\phi_{ii,1} \end{bmatrix}, \quad F_{ip} = \begin{bmatrix} 0 & \cdots & 0 & 0 \cdots 0 \\ \vdots & \ddots & \vdots & \vdots \ddots \vdots \\ 0 & \cdots & 0 & 0 \cdots 0 \\ -\phi_{ip,n_i} & \cdots & -\phi_{ip,n_i-n_{ip}+1} & 0 \cdots 0 \end{bmatrix},$$

$$i, p = 1, 3; \qquad n_1 = 3, \ n_2 = 0, \ n_3 = 2, \ n_4 = 0.$$

5.9.3 Relation Between VARMA and State Space Echelon Forms

Theorem 5.5 *Given the innovations process* $\{A_t\} \sim (0, \Sigma)$*,* $\Sigma > 0$*, and the initial conditions,* $I = \{Y_{11}, \ldots, Y_{1n_1}, \ldots, Y_{k1}, \ldots, Y_{kn_k}\}$*, of the process* $\{Y_t\}$*,* $Y_t = (Y_{1t}, \ldots, Y_{kt})'$*, the following statements are equivalent*

i) $\{Y_t\}$ *follows a finite linear time series model (3.29) such that the rank of the augmented Hankel matrices,* H^a_t*, is constant for* $t > r$*, the Kronecker indices are* $\{n_i : i = 1, \ldots, k\}$*,* $dim(\alpha_1) = \sum_{i=1}^k n_i$*, and the initial conditions are I.*
ii) $\{Y_t\}$ *follows a VARMA echelon form (5.39) such that the Kronecker indices are* $\{n_i : i = 1, \ldots, k\}$*, and the initial conditions are I.*
iii) $\{Y_t\}$ *follows a state space echelon form (5.51) and (5.52) such that the Kronecker indices are* $\{n_i : i = 1, \ldots, k\}$*, and the initial conditions are I.*

Proof Let, as before, $\Psi_{i,j}$ denote the i-th row of the matrix Ψ_j, $i = 1, 2, \ldots$. We will first prove that i) implies iii). It has been proved in Sect. 5.9.2 that $\{Y_t\}$

satisfies (5.51) and (5.52). Given the definition (5.50), if we select

$$x_1 = \left[Y^1_{1|0}, \dots, Y^1_{n_1|0}, Y^2_{1|0}, \dots, Y^2_{n_2|0}, \dots, Y^k_{1|0}, \dots, Y^k_{n_k|0}\right]',$$

where

$$Y^p_{i|0} = Y_{pi} - A_{pi} - \Psi_{p,1}A_{i-1} + \cdots + \Psi_{p,i-1}A_1, \quad i = 1, \dots, n_i, \quad p = 1, \dots, k,$$

and $A_t = (A_{1t}, \dots, A_{kt})'$, it is clear that $\{Y_t\}$ is a solution of (5.51) and (5.52) with initial conditions I.

To prove that iii) implies ii), iterate first in (5.51) and (5.52) to get

$$\begin{aligned} Y_t &= HF^{t-1}x_1 + A_t + HKA_{t-1} + \cdots + HF^{t-2}KA_1 \\ &= h_t x_1 + A_t + \Psi_1 A_{t-1} + \cdots + \Psi_{t-1}A_1, \quad t = 1, 2, \dots . \end{aligned}$$

Thus, $\{Y_t\}$ follows a finite linear time series model (3.29) with $h_t = HF^{t-1}$, $\Psi_t = HF^{t-1}K$, Kronecker indices $\{n_i : i = 1, \dots, k\}$, $\alpha_1 = x_1$, $\dim(x_1) = \sum_{i=1}^k n_i$, and initial conditions I. Then, it has been proved in Sect. 5.9.1 that $\{Y_t\}$ satisfies (5.39). Given the structure (5.41)–(5.43), it is easy to verify that $\{Y_t\}$ is a solution of (5.39) if the same initial conditions, I, are selected.

We finally prove that ii) implies i). Let $l = \max\{n_i : i = 1, \dots, k\}$, $\Psi(z) = \sum_{j=0}^{\infty} \Psi_j z^j = \Phi^{-1}(z)\Theta(z)$ and $\alpha_1 = x_1$, where x_1 has been defined in the first part of the proof. Define $h_t = HF^{t-1}$, $t = 1, \dots, l$, where F and H are as in Sect. 5.9.2, F_{ii} is $n_i \times n_i$, F_{ip} is $n_i \times n_p$, H is $k \times (n_1 + \cdots + n_k)$ and the $\phi_{ip,j}$ parameters of F are equal to the corresponding ones in $\Phi(z)$. Define further $\Phi_0 h_t = -\Phi_1 h_{t-1} - \cdots - -\Phi_l h_{t-l}$ for $t > l$. We will prove that the process $\{Y_{t,P}\}$, where

$$Y_{t,P} = h_t\alpha_1 + A_t + \Psi_1 A_{t-1} + \cdots + \Psi_{t-1}A_1, \quad t = 1, 2, \dots,$$

has the required properties and thus coincides with $\{Y_t\}$. It is not difficult but somewhat tedious to verify that the initial conditions of $\{Y_{t,P}\}$ are I and that its Kronecker indices are $\{n_i : i = 1, \dots, k\}$. On the other hand, by the definition of h_t, it follows that

$$\begin{aligned} \Phi(B)Y_{t,P} &= \Phi(B)\left[\sum_{j=0}^{t-1} \Psi_j A_{t-j} + h_t\alpha_1\right] \\ &= \Phi(B)\left[I + \Psi_1 B + \cdots + \Psi_{t-1}B^{t-1}\right]A_t \\ &= \Theta(B)A_t, \quad t > l, \end{aligned}$$

and the theorem is proved. □

5.9.4 Echelon and Overlapping Parametrizations

Suppose we have a time series that is a sample of a process, $\{Y_t\}$, following a finite linear time series model (3.29) under the assumptions in i) of Theorem 5.5. According to this theorem, we can use the state space or the VARMA echelon form to represent the process. These echelon forms are unique and we can use the parameters different from one or zero in them to parameterize the process. These parameters constitute the **echelon parametrization**. Once we have identified the state space or the VARMA echelon form, we can estimate the parameters using the observed time series. However, if the echelon form is incorrectly identified, the estimation algorithm will encounter numerical difficulties.

The previous consideration is at the root of the so-called overlapping parametrizations for state space models that have a fixed McMillan degree. The idea behind the overlapping parametrizations is to use a finite set of parametrizations in the estimation algorithm that we know can represent the process in a neighborhood of the true parameter values. We also know that the process can be represented by some subset of these parametrizations at the true parameter values. Thus, when we are approaching the true values, any parametrization in the set can be used, but as we get closer and closer to the true values those parametrizations that do not represent the process at these values will turn numerically unstable. We can detect using some test when this situation occurs and in this case we change the parametrizations until we find some alternative one that is numerically stable and thus represents the process at the true values. The optimization algorithm is not affected by the changes in the parametrizations because, as mentioned earlier, all parametrizations can represent the process when the parameter values are not sufficiently close to the true ones. Therefore, the gains in the optimization process are maintained when we change the parametrization.

To make the previous idea mathematically rigorous, we start with the following definition.

Definition 5.1 Suppose a finite linear time series model (3.29) such that the rank of the augmented Hankel matrices, H_t^a, is constant for $t > r$, the Kronecker indices are $\{n_i : i = 1, \ldots, k\}$ and the McMillan degree is $n = \sum_{i=1}^{k} n_i$. A selection, $(i_1, \ldots, i_n)$, of n rows of H_t^a, $t > r$, is called "**nice**" if the following two conditions are satisfied.

i) $1, 2, \ldots, k \in (i_1, \ldots, i_n)$
ii) if $j \in (i_1, \ldots, i_n)$ and $j > k$, then $j - k \in (i_1, \ldots, i_n)$

Given the structure of the nice selections, to any nice selection there corresponds a set of integers, $\{m_1, \ldots, m_k\}$, such that

$$n = m_1 + \cdots + m_k, \quad m_i \geq 1, \quad i = 1, \ldots, k,$$

called the set of **intrinsic invariants**. To each number m_i, $i = 1, \ldots, k$, the rows $kp + i, p = 0, 1, \ldots, m_i - 1$, are in the nice selection.

It can be shown that the number of nice selections is equal to $\binom{n-1}{k-1}$. It is to be noted that the rows of the basis implied by the Kronecker indices, n_i, $i = 1, \ldots, k$, of Proposition 3.8 constitute a nice selection when $n_i \geq 1$ for all i. Another selection that is also nice is that formed with the first n rows. This selection is called the **generic selection**. If the McMillan degree is $n = \sum_{i=1}^{k} m_i$, the intrinsic invariants of the generic selection are given by

$$m_1 = m_2 = \cdots = m_s = \left[\frac{n}{k}\right] + 1$$

and

$$m_{s+1} = \cdots = m_k = \left[\frac{n}{k}\right]$$

for some s, where $[x]$ denotes the integer part of x. In the case in which the first n rows are a basis of the space of rows of the augmented Hankel matrix, H_t^a, $t > r$, the intrinsic invariants of the generic selection coincide with the Kronecker indices.

Suppose that a nice selection with intrinsic invariants $\{m_1, \ldots, m_k\}$ and McMillan degree $n = \sum_{i=1}^{k} m_i$ is a basis of the space of rows of the augmented Hankel matrix, H_t^a, $t > r$. Then, we can derive a canonical form similar to the echelon form. This canonical form has $2nk$ parameters and can be obtained as follows.

As mentioned earlier, to the basis of the rows of the augmented Hankel matrices corresponds a basis of forecasts. Given the structure of the nice selection, this basis is $\{Y^1_{t+1|t}, \ldots, Y^1_{t+m_1|t}, Y^2_{t+1|t}, \ldots, Y^2_{t+m_2|t}, \ldots, Y^k_{t+1|t}, \ldots, Y^k_{t+m_k|t}\}$. By the definition of nice selection, the $(m_i k + i)$-th row of the augmented Hankel matrices depends linearly on all the rows in the basis, $i = 1, \ldots, k$. This in turn implies that there exist unique numbers $\phi_{ip,j}$, $i, p = 1, \ldots, k, j = 1, \ldots, m_p$, such that

$$Y^i_{t+m_i+1|t} + \sum_{p=1}^{k} \sum_{j=1}^{m_p} \phi_{ip,j} Y^p_{t+m_p+1-j|t} = 0. \tag{5.53}$$

From (5.46) and (5.53), the following relations are obtained

$$Y^i_{t+j|t} = Y^i_{t+j|t-1} + \Psi_{i,j} A_t, \qquad j = 1, 2 \ldots, m_i - 1 \tag{5.54}$$

$$Y^i_{t+m_i|t} = -\sum_{p=1}^{k} \sum_{j=1}^{m_p} \phi_{ip,j} Y^p_{t+m_p-j|t-1} + \Psi_{i,m_i} A_t. \tag{5.55}$$

If we stack the elements of the basis of forecasts in the vector

$$x_{t+1} = \left[Y^1_{t+1|t}, \ldots, Y^1_{t+m_1|t}, Y^2_{t+1|t}, \ldots, Y^2_{t+m_2|t}, \ldots, Y^k_{t+1|t}, \ldots, Y^k_{t+m_k|t}\right]', \tag{5.56}$$

where $\dim(x_{t+1}) = \sum_{i=1}^{k} m_i$, by (5.46), (5.54), and (5.55), it is obtained that

$$x_{t+1} = Fx_t + KA_t \tag{5.57}$$

$$Y_t = Hx_t + A_t, \tag{5.58}$$

where

$$F = \left[\begin{array}{c|c|c|c|c} F_{11} & \cdots & F_{1i} & \cdots & F_{1k} \\ \hline \vdots & \ddots & \vdots & \ddots & \vdots \\ \hline F_{i1} & \cdots & F_{ii} & \cdots & F_{ik} \\ \hline \vdots & \ddots & \vdots & \ddots & \vdots \\ \hline F_{k1} & \cdots & F_{ki} & \cdots & F_{kk} \end{array}\right], \quad K = \left[\begin{array}{c} K_{11} \\ \hline \vdots \\ \hline K_{i1} \\ \hline \vdots \\ \hline K_{k1} \end{array}\right], \quad K_{i1} = \begin{bmatrix} \Psi_{i,1} \\ \vdots \\ \Psi_{i,m_i} \end{bmatrix},$$

$$F_{ii} = \begin{bmatrix} 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \\ -\phi_{ii,m_i} & \cdots & \cdots & -\phi_{ii,1} \end{bmatrix}, \quad F_{ip} = \begin{bmatrix} 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \\ -\phi_{ip,m_p} & \cdots & -\phi_{ip,1} \end{bmatrix},$$

$$H = \left[\begin{array}{cccc|cccc|c|cccc} 1 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 \\ 0 & \cdots & 0 & 0 & 1 & \cdots & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 1 & \cdots & 0 & 0 \end{array}\right],$$

F_{ii} is $m_i \times m_i$, F_{ip} is $m_i \times m_p$, and H is $k \times (m_1 + \cdots + m_k)$. The state space form (5.57) and (5.58) is called the **state space canonical form** corresponding to the intrinsic invariants $\{m_i : i = 1, \ldots, k\}$ and McMillan degree $n = \sum_{i=1}^{k} m_i$. To each canonical form corresponding to the intrinsic invariants $\{m_i : i = 1, \ldots, k\}$ and McMillan degree $n = \sum_{i=1}^{k} m_i$ we can associate a **parameter vector**, $\theta_{n:m_1,\ldots,m_k} = [\phi_{11,1}, \ldots, \phi_{kk,m_k}, \psi_{11,1}, \ldots, \psi_{kk,m_k}]' \in \mathbb{R}^{2nk}$, where $\psi_{ip,j}$ is the element (i,p) in Ψ_j.

Proposition 5.3 *The canonical form (5.57) and (5.58) corresponding to a nice selection with intrinsic invariants $\{m_1, \ldots, m_k\}$ and McMillan degree $n = \sum_{i=1}^{k} m_i$ that is a basis of the space of rows of the augmented Hankel matrix, H_t^a, $t > r$, is minimal.*

Proof The result is a consequence of Proposition 3.9, since the dimension of x_t is equal to the McMillan degree. □

Proposition 5.4 *The canonical form (5.57) and (5.58) corresponding to a nice selection with intrinsic invariants $\{m_1, \ldots, m_k\}$ and McMillan degree $n = \sum_{i=1}^{k} m_i$ is observable but not necessarily controllable.*

Proof Let $O_{n:m_1,\ldots,m_k} = [H', F'H', \ldots, F'^{n-1}H']'$ be the observability matrix. Then, stack the observations to get

$$Y_{t+1:t+n} = O_{n:m_1,\ldots,m_k} F x_t + \widehat{\Psi}_{n+1} A_{t:t+n} \tag{5.59}$$

where $Y_{t+1:t+n} = (Y'_{t+1}, \ldots, Y'_{t+n})'$, $A_{t:t+n} = (A'_t, \ldots, A'_{t+n})'$,

$$\widehat{\Psi}_{n+1} = \begin{bmatrix} \Psi_1 & I & & & \\ \Psi_2 & \Psi_1 & I & & \\ \vdots & \vdots & \vdots & \ddots & I \\ \Psi_n & \Psi_{n-1} & \cdots & \cdots & \Psi_1 & I \end{bmatrix},$$

and $\Psi_i = HF^{i-1}K$, $i = 1, \ldots, n$. Considering the definition of the forecasts (5.44), it follows from (5.59) that

$$\begin{bmatrix} Y_{t+1|t} \\ Y_{t+2|t} \\ \vdots \\ Y_{t+n|t} \end{bmatrix} = O_{n:m_1,\ldots,m_k} F x_t + \begin{bmatrix} \Psi_1 \\ \Psi_2 \\ \vdots \\ \Psi_n \end{bmatrix} A_t. \tag{5.60}$$

Let $J_{n:m_1,\ldots,m_k}$ be a selection matrix formed with zeros and ones such that

$$x_{t+1} = J_{n:m_1,\ldots,m_k} [Y'_{t+1|t}, \ldots, Y'_{t+n|t}]'.$$

Then, premultiplying (5.60) by $J_{n:m_1,\ldots,m_k}$, it is obtained that $J_{n:m_1,\ldots,m_k} O_{n:m_1,\ldots,m_k}$ is the unit matrix, I_n, and, therefore, (5.57) and (5.58) are observable. □

Remark 5.5 It follows from Proposition 5.4 that the echelon form (5.51) and (5.52) is observable because the Kronecker indices are a special case of intrinsic invariants. ◇

Proposition 5.5 *The set, $S_{n:m_1,\ldots,m_k} \subset \mathbb{R}^{2nk}$, of parameter vectors of canonical forms (5.57) and (5.58) corresponding to a nice selection with intrinsic invariants $\{m_i : i = 1, \ldots, k\}$ and McMillan degree $n = \sum_{i=1}^{k} m_i$ that is a basis of the space of rows of the augmented Hankel matrix, H^a_t, $t > r$, is a dense open subset of $\mathbb{R}^{2nk}$.*

Proof We will first prove that $S_{n:m_1,\ldots,m_k}$ is open. By Proposition 5.4, (5.57) and (5.58) are always observable. By Proposition 5.3, (5.57) and (5.58) are minimal and, therefore, controllable, for parameter vectors $\theta \in S_{n:m_1,\ldots,m_k}$. Let $C_{n:m_1,\ldots,m_k}(\theta)$ be the controllability matrix, defined in Sect. 5.12, for $\theta \in S_{n:m_1,\ldots,m_k}$ and let

$$A(\theta) = \max\{|A_i(\theta)| : A_i(\theta) \text{ is a minor of order } n \text{ of } C_{n:m_1,\ldots,m_k}(\theta)\}.$$

Then,

$$S_{n:m_1,\ldots,m_k} = \{\alpha \in \mathbb{R}^{2nk} : A(\alpha) > 0\}$$

and $S_{n:m_1,\ldots,m_k}$ is open because it is the inverse image of an open set under a continuous function.

To prove that $S_{n:m_1,\ldots,m_k}$ is dense, we have to prove that for all $\alpha \in \mathbb{R}^{2nk}$ there is a neighborhood of α that contains a point in $S_{n:m_1,\ldots,m_k}$. If $\alpha \in S_{n:m_1,\ldots,m_k}$, the theorem is proved. Suppose otherwise. Then, the rows of the nice selection do not constitute a basis of the space of rows of the augmented Hankel matrix, $H_t^a(\alpha)$, $t > r$, and, therefore, (5.57) and (5.58) are not minimal. By Theorem 3.3 and Proposition 5.4, the controllability matrix does not have rank n. However, in any neighborhood of α it is easy to see that it is possible to find a state space model (5.57) and (5.58) in which the controllability matrix has rank n. This model is thus minimal and its parameter vector is in $S_{n:m_1,\ldots,m_k}$ because otherwise the model would not be minimal. □

The **overlapping parametrizations** are defined for the augmented Hankel matrices, H_t^a, of McMillan degree n as the parameter vectors, $\theta_{n:m_1,\ldots,m_k} \in \mathbb{R}^{2nk}$, of the canonical forms (5.57) and (5.58) that correspond to the different intrinsic invariants $\{m_i : i = 1, \ldots, k\}$ such that $n = \sum_{i=1}^{k} m_i$. Note that it is not assumed that the rows corresponding to the different intrinsic invariants form a basis of the space of rows of the augmented Hankel matrices.

To estimate a state space model of McMillan degree n using overlapping parametrizations, we start with a set of intrinsic invariants, $\{m_i : i = 1, \ldots, k\}$, and a parameter vector, $\theta_{n:m_1,\ldots,m_k}$, corresponding to a canonical form (5.57) and (5.58). We will iterate in the optimization algorithm using this set of intrinsic invariants unless we encounter numerical difficulties that we measure with some kind of test. If this is the case, we change the set of intrinsic invariants and the corresponding vector of parameters until we find some suitable parametrization to continue with the optimization. This process is repeated until convergence has been achieved. Proposition 5.5 guarantees that any overlapping parametrization that we use in the neighborhood of the true parameter values will be a valid one. It is only when we get close to these values that some parametrizations may become unstable. In this case, we replace the unstable parametrization with a stable one and proceed with the optimization. Note that the intrinsic invariants are not part of the parameter vector when using this estimation method. This is in contrast with the method that uses the echelon parametrizations, where the Kronecker indices are assumed to be known at the start of the optimization process.

To find the matrix that transforms a state vector, x_t, corresponding to the intrinsic invariants $\{m_i : i = 1, \ldots, k\}$ to another state vector, $\bar{x}_t$, corresponding to the intrinsic invariants $\{\overline{m}_i : i = 1, \ldots, k\}$, consider (5.60). Letting $J_{n:\overline{m}_1,\ldots,\overline{m}_k}$ a selection matrix formed with zeros and ones such that $\bar{x}_{t+1} = J_{n:\overline{m}_1,\ldots,\overline{m}_k}[Y'_{t+1|t}, \ldots, Y'_{t+n|t}]'$, we see that the matrix T such that $\bar{x}_{t+1} = Tx_{t+1}$ is

$$T = J_{n:\overline{m}_1,\ldots,\overline{m}_k} O_{n:m_1,\ldots,m_k}.$$

5.10 Covariance Factorization for State Space Echelon Forms

Consider a multivariate stationary process $\{Y_t\}$ that follows the VARMA(p, q) model (5.39) in echelon form and the state space echelon form (5.51) and (5.52). As mentioned in Sect. 5.6, this corresponds to setting $u_t = v_t = A_t$, $R = S = Q = \mathrm{Var}(A_t)$ and $K = G$ in (5.4) and (5.5). Defining $\Sigma = \mathrm{Var}(A_t)$, we see that (5.20) holds with $P = 0$ and the factorization (5.18) holds. If we further define $\Psi(z) = I + zH(I - Fz)^{-1}K$, then $\Psi(z) = \Phi^{-1}(z)\Theta(z) = \sum_{i=0}^{\infty} \Psi_i z^i$ and $G_Y(z)$ can be written as $G_Y(z) = \Psi(z)\Sigma\Psi'(z^{-1})$. In addition,

$$K = \begin{bmatrix} K_1 \\ \hline \vdots \\ \hline K_i \\ \hline \vdots \\ \hline K_k \end{bmatrix}, \quad K_i = \begin{bmatrix} \Psi_{i,1} \\ \vdots \\ \Psi_{i,n_i} \end{bmatrix},$$

where $\Psi_{i,j}$ denotes the i-th row of the matrix Ψ_j, $i = 1, 2, \ldots$, and $\{n_i : i = 1, \ldots, k\}$ are the Kronecker indices, and, by Remark 5.2, $\gamma(j) = HF^{j-1}N$ and

$$\begin{aligned} \begin{bmatrix} \gamma(1) \\ \vdots \\ \gamma(n) \end{bmatrix} &= \begin{bmatrix} H \\ HF \\ \vdots \\ HF^{n-1} \end{bmatrix} N \\ &= O_n N, \end{aligned} \tag{5.61}$$

where $N = \mathrm{Cov}(x_{t+1}, Y_t)$, $n = \sum_{i=1}^{k} n_i$ is the McMillan degree and O_n is the observability matrix.

Stacking the observations in (5.51) and (5.52), we get

$$Y_{t+1:t+n} = O_n F x_t + \widehat{\Psi}_{n+1} A_{t:t+n} \tag{5.62}$$

where $Y_{t+1:t+n} = (Y'_{t+1}, \ldots, Y'_{t+n})'$, $A_{t:t+n} = (A'_t, \ldots, A'_{t+n})'$,

$$\widehat{\Psi}_{n+1} = \begin{bmatrix} \Psi_1 & I & & & \\ \Psi_2 & \Psi_1 & I & & \\ \vdots & \vdots & \vdots & \ddots & I \\ \Psi_n & \Psi_{n-1} & \cdots & \cdots & \Psi_1 & I \end{bmatrix},$$

and $\Psi_i = HF^{i-1}K$, $i = 1, \ldots, n$. Considering the definition of the forecasts (5.44), it follows from (5.62) that

$$\begin{bmatrix} Y_{t+1|t} \\ Y_{t+2|t} \\ \vdots \\ Y_{t+n|t} \end{bmatrix} = O_n F x_t + \begin{bmatrix} \Psi_1 \\ \Psi_2 \\ \vdots \\ \Psi_n \end{bmatrix} A_t.$$

Let J_n be a selection matrix formed with zeros and ones such that

$$\begin{aligned} x_{t+1} &= J_n[Y'_{t+1|t}, \ldots, Y'_{t+n|t}]' \\ &= \left[Y^1_{t+1|t}, \ldots, Y^1_{t+n_1|t}, Y^2_{t+1|t}, \ldots, Y^2_{t+n_2|t}, \ldots, Y^k_{t+1|t}, \ldots, Y^k_{t+n_k|t}\right]'. \end{aligned}$$

Then, premultiplying (5.62) by J_n, it is obtained that $J_n O_n$ is the unit matrix, I_n, and, by (5.61),

$$N = J_n \begin{bmatrix} \gamma(1) \\ \vdots \\ \gamma(n) \end{bmatrix}. \tag{5.63}$$

In fact, it can be verified that the rows $\{(r-1)p + i : 1 \le r \le n_i; 1 \le i \le k\}$ of O_n are equal to I_n. Therefore, these same rows of the matrices

$$\begin{bmatrix} \gamma(1) \\ \vdots \\ \gamma(n) \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} \Psi_1 \\ \Psi_2 \\ \vdots \\ \Psi_n \end{bmatrix}$$

are equal to the matrices N and K, respectively.

If we only knew the covariance sequence, $\{\gamma(j)\}$, and the autoregressive polynomial, $\Phi(z)$, of Y_t, we could estimate the vector K and $\Sigma = \mathrm{Var}(A_t)$ by first solving the DARE (5.28) and then using (5.27). The moving average polynomial, $\Theta(B)A_t$, could be obtained using the relations (5.40). Note that, because $\Psi_i = HF^{i-1}K$,

$i = 1, \ldots, n$, only the rows of the Ψ_i matrices contained in K are needed for the calculations in (5.40).

5.11 Observability and Controllability

Let A and B be matrices with dimensions $a \times a$ and $a \times b$. The following definitions are important regarding (5.4) and (5.5) or (5.67) in the next section and (5.5).

- The pair of matrices (A, B) is **controllable** if there does not exist a left eigenvector of A that is orthogonal to B. That is, for all vector $v \neq 0$ and scalar λ such that $vA = \lambda v$, the relation $vB \neq 0$ holds. A necessary and sufficient condition for the pair (A, B) to be controllable is that the **controllability matrix**, $[B, AB, \ldots, A^{a-1}B]$, has full row rank.
- The pair of matrices (A, B) is **stabilizable** if there does not exist a left eigenvector of A corresponding to an unstable eigenvalue that is orthogonal to B. That is, for all vector $v \neq 0$ and scalar λ such that $vA = \lambda v$ and $|\lambda| \geq 1$, the relation $vB \neq 0$ holds.
- The pair of matrices (A, B) is **observable** if and only if the pair (A', B') is controllable. A necessary and sufficient condition for the pair (A, B) to be observable is that the **observability matrix**, $[B', A'B', \ldots, (A')^{a-1}B']'$, has full column rank.
- The pair of matrices (A, B) is **detectable** if and only if the pair (A', B') is stabilizable.
- A real symmetric positive semidefinite solution P of the DARE (5.22) is said to be a **strong solution** if the corresponding matrix $F_p = F - KH$, where $K = (FPH' + GS)(R + HPH')^{-1}$, has all its eigenvalues inside or on the unit circle. If F_p has all its eigenvalues inside the unit circle, the solution is called a **stabilizing solution**.

The following two lemmas are important in connection with Observability and Controllability.

Lemma 5.7 (Controllability Staircase Form) *Given a pair of matrices, (F, G), with $F \in \mathbb{R}^{(r,r)}$ and $G \in \mathbb{R}^{(r,q)}$, $G \neq 0$, there always exists an orthogonal matrix $U \in \mathbb{R}^{(r,r)}$ such that*

$$\left[U'G \parallel U'FU\right] = \left[\begin{array}{c||cccc|c} X_1 & F_{1,1} & F_{1,2} & \cdots & F_{1,l} & F_{1,l+1} \\ 0 & X_2 & F_{2,2} & & \vdots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots & \vdots \\ \vdots & & \ddots & X_l & F_{l,l} & F_{l,l+1} \\ \hline 0 & \cdots & \cdots & 0 & 0 & F_{l+1,l+1} \end{array}\right], \tag{5.64}$$

where the $F_{i,i}$, $i = 1, \dots, l$, *are* $\rho_i \times \rho_i$ *matrices and the* X_i, $i = 1, \dots, l$, *are* $\rho_i \times \rho_{i-1}$ *matrices of full row rank (with* $\rho_0 = q$*). As a consequence,* $q = \rho_0 \geq \rho_i \geq \cdots \geq \rho_l > 0$ *and* $F_{l+1,l+1}$ *is a square matrix of dimension* $(r - \sigma_l) \times (r - \sigma_l)$ *with* $\sigma_l = \sum_{i=1}^{l} \rho_i$.

Proof We will illustrate the process first for $q = 1$. The general case will be proved later. It is easy to see that one can find an orthogonal matrix, U_1, such that

$$U_1' G = \begin{bmatrix} X_1 \\ \hline 0 \\ \vdots \\ 0 \end{bmatrix}, \quad U_1' F U_1 = \left[\begin{array}{c|ccc} \times & \times & \cdots & \times \\ \hline \times & \times & \cdots & \times \\ \vdots & \vdots & \ddots & \vdots \\ \times & \times & \cdots & \times \end{array}\right],$$

where X_1 is a nonzero number and the symbol $\times$ indicates a number that can be different from zero. In fact U_1 is the matrix of a Householder transformation that puts G in the required form. If one now redefines the matrices G and F as G_1 and F_1 and the bottom two blocks of $U_1' F U_1$ as G_2 and F_2, then the additional transformation $U_2' = \left[\begin{array}{c|c} 1 & \\ \hline & H_2' \end{array}\right]$, where H_2' is the matrix of a Householder transformation, is such that

$$U_2' U_1' G_1 = \begin{bmatrix} X_1 \\ \hline 0 \\ \vdots \\ 0 \end{bmatrix}, \quad U_2' U_1' F_1 U_1 U_2 = \left[\begin{array}{c|c} \times & \times \ \ldots \ \times \\ \hline H_2' G_2 & H_2' F_2 H_2 \end{array}\right].$$

Thus, U_2' in fact applies H_2' to G_2 and F_2 in a similar manner as U_1' was applied to G_1 and F_1, but leaving the previously created zeros in $U_1' G_1$ unaffected. So, we can choose H_2 such that

$$U_2' U_1' G_1 = \begin{bmatrix} X_1 \\ \hline 0 \\ \vdots \\ 0 \end{bmatrix}, \quad U_2' U_1' F_1 U_1 U_2 = \left[\begin{array}{c|c|c} \times & \times & \cdots\cdots \times \\ \hline X_2 & \times & \times \ \cdots \ \times \\ 0 & & \\ \vdots & G_3 & F_3 \\ 0 & & \end{array}\right],$$

where X_2 is a nonzero number. We can continue the process with G_3 and F_3 using $U_3' = \left[\begin{array}{c|c} I_2 & \\ \hline & H_3' \end{array}\right]$, and so on until we get the desired form. Note that $\rho_0 = \rho_i = 1$, $i = 1, \dots, l$, in this case.

The general case is a block version of the previous procedure, stopping as soon as a "zero" X_i is encountered, at step $l + 1$ in this case. The proof is thus constructive.

At each step $i = 1, \dots, l$ a QR factorization is performed of the current G_i matrix yielding X_i of full row rank. If a zero rank matrix G_i is encountered (at step $l + 1$), then we obtain the form (5.64). If not, then the method terminates with $r = \sigma_l$ and the bottom matrix $F_{l+1,l+1}$ is empty. □

Lemma 5.8 (Observability Staircase Form) *Given a pair of matrices,* (F, H)*, with* $F \in \mathbb{R}^{(r,r)}$ *and* $H \in \mathbb{R}^{(k,r)}$*,* $H \neq 0$*, there always exist an orthogonal matrix* $U \in \mathbb{R}^{(r,r)}$ *such that*

$$\left[\begin{array}{c} U'FU \\ \hline\hline HU \end{array}\right] = \left[\begin{array}{c|cccc} F_{l+1,l+1} & F_{l,l+1} & \cdots & F_{1,l+1} \\ \hline 0 & F_{l,l} & \cdots & F_{1,l} \\ 0 & Y_l & \cdots & F_{1,l-1} \\ \vdots & & \ddots & Y_2 & F_{1,1} \\ \hline\hline 0 & \cdots & 0 & Y_1 \end{array}\right], \tag{5.65}$$

where the $F_{i,i}$*,* $i = 1, \dots, l$*, are* $\rho_i \times \rho_i$ *matrices and the* Y_i*,* $i = 1, \dots, l$*, are* $\rho_{i-1} \times \rho_i$ *matrices of full column rank (with* $\rho_0 = k$*). As a consequence,* $k = \rho_0 \geq \rho_i \geq \cdots \geq \rho_l > 0$ *and* $F_{l+1,l+1}$ *is a square matrix of dimension* $(r - \sigma_l) \times (r - \sigma_l)$ *with* $\sigma_l = \sum_{i=1}^{l} \rho_i$*.*

Proof The lemma follows by applying Lemma 5.7 to the pair (F', H'), transposing, and then interchanging the columns in HU and interchanging the rows and columns in $U^t FU$. Note that interchanging the rows or columns in a matrix is equivalent to multiplying by an orthogonal matrix. □

Remark 5.6 The proof of Lemma 5.7 is based on a result in Van Dooren (1979) and follows Van Dooren (2003), where an algorithm in MATLAB code can also be found to compute the orthogonal matrix U. ◇

The following two theorems give procedures to obtain the rank of the controllability and the observability matrices based on Lemmas 5.7 and 5.8.

Theorem 5.6 *Under the assumptions and with the notation of Lemma 5.7, the rank of the controllability matrix,* $[G, FG, \dots, F^{r-1}G]$*, is* σ_l*.*

Theorem 5.7 *Under the assumptions and with the notation of Lemma 5.8, the rank of the observability matrix,* $[H', F'H', \dots, (F')^{r-1}H']'$*, is* σ_l*.*

Proof of Theorems 5.6 and 5.7 To prove Theorem 5.6, consider first that the rank, r_c, of the matrix $[G, FG, \dots, F^{r-1}G]$ is the same as that of the matrix $[U'G, U'FG, \dots, U'F^{r-1}G] = [U'G, (U'FU)U'G, \dots, (U'FU)^{r-1}U'G]$ for any orthogonal matrix U. Thus, if U is the transformation of Lemma 5.7, then r_c is the rank of the matrix

$$\left[\begin{array}{c|c|c|c|c|c|c} X_{1,1} & \times & \ddots & \times & \times & \cdots & \times \\ 0 & X_{2,1} & \ddots & \vdots & \vdots & & \vdots \\ \vdots & \ddots & X_{l,1} & \times & \times & \cdots & \times \\ 0 & \cdots & 0 & 0 & \cdots & \cdots & 0 \end{array}\right] \begin{array}{l} \}\rho_1 \\ \vdots \\ \}\rho_l \\ \end{array} \tag{5.66}$$

where the matrices $X_{i,1} = X_i \cdot X_{i-1} \cdots X_1$ are of full row rank since they are the product of full row rank matrices. But the form of the matrix (5.66) indicates that $r_c = \sigma_l = \sum_{i=1}^{l} \rho_i$. The proof of Theorem 5.7 is similar. □

The following corollary is a direct consequence of Lemma 5.7 and Theorem 5.6. The proof is left to the reader.

Corollary 5.3 *Let the state space model (5.1) and (5.2), where* $F \in \mathbb{R}^{(r,r)}$, $G \in \mathbb{R}^{(r,q)}$, $H \in \mathbb{R}^{(p,r)}$ *and* $J \in \mathbb{R}^{(p,q)}$. *If the rank of the controllability matrix,* $[G, FG, \ldots, F^{r-1}G]$, *is* $r_1 < r$, *then there exists an orthogonal matrix* U *such that*

$$U'FU = \begin{bmatrix} F_c & F_{12} \\ 0 & F_{\bar{c}} \end{bmatrix}, \qquad U'G = \begin{bmatrix} G_c \\ 0 \end{bmatrix}$$

F_c *is* $r_1 \times r_1$, $F_{\bar{c}}$ *is* $(r-r_1) \times (r-r_1)$ *and the pair* $[F_c, G_c]$ *is controllable. In addition, the transfer function,* $\Psi(z)$, *of the system*

$$\begin{aligned} \bar{x}_{t+1} &= F_c \bar{x}_t + G_c \epsilon_t, \\ \overline{Y}_t &= H_c \bar{x}_t + J \epsilon_t, \end{aligned}$$

where $HU = [H_c, H_{12}]$ *and* H_c *is* $p \times r_1$, *coincides with that of (5.1) and (5.2). That is,* $\Psi(z) = J + zH(I - zF)^{-1}G = J + zH_c(I - zF_c)^{-1}G_c$.

Corollary 5.4 *Let the state space model (5.1) and (5.2), where* $F \in \mathbb{R}^{(r,r)}$, $G \in \mathbb{R}^{(r,q)}$, $H \in \mathbb{R}^{(p,r)}$ *and* $J \in \mathbb{R}^{(p,q)}$. *If the rank of the observability matrix,* $[H', F'H', \ldots, (F')^{r-1}H']'$, *is* $r_2 < r$, *then there exists an orthogonal matrix* U *such that*

$$U'FU = \begin{bmatrix} F_o & 0 \\ F_{21} & F_{\bar{o}} \end{bmatrix}, \qquad HU = [H_o, 0]$$

F_o *is* $r_2 \times r_2$, $F_{\bar{o}}$ *is* $(r-r_2) \times (r-r_2)$ *and the pair* $[F_o, H_o]$ *is observable. In addition, the transfer function,* $\Psi(z)$, *of the system*

$$\begin{aligned} \bar{x}_{t+1} &= F_o \bar{x}_t + G_o \epsilon_t, \\ \overline{Y}_t &= H_o \bar{x}_t + J \epsilon_t, \end{aligned}$$

where $G'U = [G'_o, G'_2]'$ *and* H_o *is* $p \times r_2$*, coincides with that of (5.1) and (5.2). That is,* $\Psi(z) = J + zH(I - zF)^{-1}G = J + zH_o(I - zF_o)^{-1}G_o$*.*

Proof Letting U be the orthogonal matrix of Lemma 5.8, by Theorem 5.7, we can write

$$U'FU = \begin{bmatrix} F_{\bar{o}} & F_{21} \\ 0 & F_o \end{bmatrix}, \qquad HU = [0, H_o],$$

where F_o is $r_2 \times r_2$, $F_{\bar{o}}$ is $(r-r_2) \times (r-r_2)$ and the pair $[F_o, H_o]$ is observable. Define the orthogonal matrix

$$P' = \begin{bmatrix} 0 & I_{r-r_2} \\ I_{r_2} & 0 \end{bmatrix}.$$

Then, it holds that

$$(P'U')F(UP) = \begin{bmatrix} F_o & 0 \\ F_{21} & F_{\bar{o}} \end{bmatrix}, \qquad H(UP) = [H_o, 0].$$

The rest of the proof is left to the reader. □

Combining the two previous corollaries, we have the following Kalman decomposition theorem.

Theorem 5.8 (Kalman Decomposition Theorem) *Let the state space model (5.1) and (5.2), where* $F \in \mathbb{R}^{(r,r)}$*,* $G \in \mathbb{R}^{(r,q)}$*,* $H \in \mathbb{R}^{(p,r)}$ *and* $J \in \mathbb{R}^{(p,q)}$*. Then, there exists an orthogonal matrix* U *such that the transformed state* $[x'_{co,t}, x'_{c\bar{o},t}, x'_{\bar{c}o,t}, x'_{\bar{c}\bar{o},t}]' = U'x_t$ *satisfies the following state space model*

$$\begin{bmatrix} x_{co,t+1} \\ x_{c\bar{o},t+1} \\ x_{\bar{c}o,t+1} \\ x_{\bar{c}\bar{o},t+1} \end{bmatrix} = \begin{bmatrix} F_{co} & 0 & F_{13} & 0 \\ F_{21} & F_{c\bar{o}} & F_{23} & F_{24} \\ 0 & 0 & F_{\bar{c}o} & 0 \\ 0 & 0 & F_{43} & F_{\bar{c}\bar{o}} \end{bmatrix} \begin{bmatrix} x_{co,t} \\ x_{c\bar{o},t} \\ x_{\bar{c}o,t} \\ x_{\bar{c}\bar{o},t} \end{bmatrix} + \begin{bmatrix} G_{co} \\ G_{c\bar{o}} \\ 0 \\ 0 \end{bmatrix} \epsilon_t$$

$$Y_t = \begin{bmatrix} H_{co} & 0 & H_{\bar{c}o} & 0 \end{bmatrix} + J\epsilon_t,$$

where the pair $[F_{co}, G_{co}]$ *is controllable and the pair* $[F_{co}, H_{co}]$ *observable. Furthermore, the transfer function,* $\Psi(z)$*, of the system*

$$x_{co,t+1} = F_{co}x_{co,t} + G_{co}\epsilon_t,$$
$$\overline{Y}_t = H_{co}x_{co,t} + J\epsilon_t$$

coincides with that of (5.1) and (5.2). That is, $\Psi(z) = J + zH(I - zF)^{-1}G = J + zH_{co}(I - zF_{co})^{-1}G_{co}$*.*

5.12 Limit Theorems for the Kalman Filter and the Smoothing Recursions

To study the asymptotic behavior of the Kalman filtering and smoothing recursions, it is convenient to transform the state space equations (5.4) and (5.5) so that in the new equations the covariance between the disturbances u_t and v_t is zero. The transformation was introduced in Sect. 4.5. Assume $R > 0$ and define $F^s = F - GSR^{-1}H$ and $Q^s = Q - SR^{-1}S'$. Then,

$$\begin{aligned} x_{t+1} &= F^s x_t + GSR^{-1}Hx_t + Gu_t = F^s x_t + G\left[SR^{-1}(Y_t - v_t) + u_t\right] \\ &= GSR^{-1}Y_t + F^s x_t + Gu_t^s, \end{aligned} \tag{5.67}$$

where $u_t^s = u_t - SR^{-1}v_t$ is orthogonal to v_t and $\text{Var}(u_t^s) = Q^s$. Since at time $t+1$, Y_t is known, it is possible to replace (5.4) and (5.5) with (5.67) and (5.5), where u_t^s and v_t are orthogonal, to study the asymptotic properties of the Kalman filtering and smoothing recursions. If R is singular, it is possible to replace R^{-1} in the previous formulae with R^-, where R^- is a generalized inverse of R.

When running the Kalman filter (4.3) corresponding to (5.4) and (5.5), if P_t converges as $t \to \infty$, then the limiting solution P will satisfy the DARE (5.22), obtained from (4.3) by putting $P_{t+1} = P_t = P$. In this case, K_t in (4.3) and $F_{p,t} = F - K_t H$ converge to the steady state quantities $K = (FPH' + GS)(R + HPH')^{-1}$ and $F_p = F - KH$.

5.12.1 Solutions of the DARE

There is an extensive literature on the solutions of the DARE. Since in this chapter we are concerned with the stabilizing solution, we will only consider those results that are relevant in this respect. The following two theorems give the details.

Theorem 5.9 (Theorem 3.1 of Chan, Goodwin, & Sin (1984)) *If* $R > 0$, $\begin{bmatrix} Q & S \\ S' & R \end{bmatrix} \geq 0$, *and* (F, H) *is detectable, then*

- *i) the strong solution of the DARE exists and is unique.*
- *ii) if* $(F^s, GQ^{s/2})$ *is stabilizable, then the strong solution is the only positive semidefinite solution of the DARE.*
- *iii) if* $(F^s, GQ^{s/2})$ *has no uncontrollable eigenvalue on the unit circle, then the strong solution coincides with the stabilizing solution.*
- *iv) if* $(F^s, GQ^{s/2})$ *has an uncontrollable eigenvalue on the unit circle, then, although the strong solution exists, there is no stabilizing solution.*
- *v) if* $(F^s, GQ^{s/2})$ *has an uncontrollable eigenvalue inside or on the unit circle, then the strong solution is not positive definite.*

vi) if $(F^s, GQ^{s/2})$ *has an uncontrollable eigenvalue outside the unit circle, then as well as the strong solution, there is at least one other positive semidefinite solution of the DARE.*

Theorem 5.10 (Theorem 3.2 of de Souza, Gevers, & Goodwin (1986)) *If* $R > 0$ *and* $\begin{bmatrix} Q & S \\ S' & R \end{bmatrix} \geq 0$, *then*

A) the strong solution of the DARE exists and is unique if and only if (F, H) *is detectable.*

B) the strong solution is the only positive semidefinite solution of the DARE if and only if (F, H) *is detectable and* $(F^s, GQ^{s/2})$ *has no uncontrollable eigenvalue outside the unit circle.*

C) the strong solution coincides with the stabilizing solution if and only if (F, H) *is detectable and* $(F^s, GQ^{s/2})$ *has no uncontrollable eigenvalue on the unit circle.*

D) the stabilizing solution is positive definite if and only if (F, H) *is detectable and* $(F^s, GQ^{s/2})$ *has no uncontrollable eigenvalue inside or on the unit circle.*

Remark 5 According to the previous two theorems, sufficient conditions for the existence of a unique stabilizing solution of the DARE are that (F, H) be detectable and $(F^s, GQ^{s/2})$ be stabilizable. Under these conditions, it is possible to obtain the covariance factorization $G_Y(z) = \Psi(z)\Sigma\Psi'(z^{-1})$, where $\Psi(z) = I + zH(I - Fz)^{-1}K$, $K = (FPH' + GS)\Sigma^{-1}$ and $\Sigma = R + HPH'$, described in Sect. 5.6 and needed for the Wiener–Kolmogorov formulae in Chap. 7. Note that, as mentioned earlier, $\Psi^{-1}(z) = I - zH(I - F_p z)^{-1}K$, where $F_p = F - KH$ is stable. ◇

Remark 6 The proofs of the previous two theorems are algebraic in nature and do not require that F be stable. This will allow for the extension of the Kalman filtering and smoothing results to the nonstationary case in Chap. 7. ◇

Remark 7 Assuming the strong solution exists and is unique, $(F^s, GQ^{s/2})$ has an uncontrollable eigenvalue on the unit circle if and only if the moving average part in the VARMA representation has a root of unit modulus. To see this, consider that, because a strong solution exists and is unique, there exists an innovations representation (3.13) and (3.14). Then, $F^s = F_p$ and $Q^s = 0$ for this representation and, as mentioned earlier, the process $\{Y_t\}$ in (3.13) and (3.14) admits a VARMA representation, $\Phi(B)Y_t = \Theta(B)A_t$, and $\det(I - F_p z) = \det[\Theta(z)]$. ◇

Remark 8 The assumption $R > 0$ in the previous two theorems is made to ensure that $R + HPH' > 0$ for any P and that, therefore, $R + HPH'$ can be inverted. Thus, this assumption can be replaced with $R + HPH' > 0$, where R may be singular. A numerically efficient procedure to compute the unique stabilizing or strong solution of the DARE, assuming it exists under the assumption $R + HPH' > 0$, is based on solving the generalized eigenvalue problem corresponding to the matrix pencil

$A - \lambda B$, where

$$A = \begin{bmatrix} F' & 0 & H' \\ GQG' & -I & GS \\ S'G' & 0 & \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} I & 0 & 0 \\ 0 & -F & 0 \\ 0 & -H & 0 \end{bmatrix}.$$

The interested reader can consult Ionescu, Oară, & Weiss (1997) for details. ◇

The following two lemmas give criteria for the pair (F,H) to be detectable and the pair $(F^s, GQ^{s/2})$ to be stabilizable in the important practical case of multivariate unobserved VARMA components models. The proofs are straightforward generalizations to the multivariate case of the ones given by Burridge & Wallis (1988) for the univariate case.

Lemma 5.9 *Let $Y_t = S_t + N_t + U_t$, where $\{S_t\}$ and $\{N_t\}$ follow multivariate, possibly nonstationary, VARMA models, $\Phi_S(B)S_t = \Theta_S(B)A_{st}$ and $\Phi_N(B)N_t = \Theta_N(B)A_{nt}$, and $\{U_t\}$ is a white noise sequence such that $\{A_{st}\}$, $\{A_{nt}\}$, and $\{U_t\}$ are mutually and serially uncorrelated, and let (5.4) and (5.5) be a state space representation of Y_t with $v_t = U_t$. Then, the pair (F,H) is detectable if and only if the polynomials det$[\Phi_S(z)]$ and det$[\Phi_N(z)]$ do not contain common factors of the form $1-\lambda z$ with $|\lambda| \geq 1$.*

Lemma 5.10 *Under the same assumptions of Lemma 5.9, the pair $(F, GQ^{1/2})$ is controllable if and only if the polynomials det$[\Phi_S(z)]$ and det$[\Theta_S(z)]$, as well as the polynomials det$[\Phi_N(z)]$ and det$[\Theta_N(z)]$, do not contain common factors of the form $1-\lambda z$ with $\lambda \neq 0$.*

5.12.2 Convergence of the DARE

The following theorem gives sufficient conditions for the convergence of the sequence of covariance matrices $\{P_t\}$ given by the Kalman filter (4.3) to the stabilizing solution. A proof can be found in, for example, Anderson & Moore (2012).

Theorem 5.11 *If $R > 0$, $\begin{bmatrix} Q & S \\ S' & R \end{bmatrix} \geq 0$, $(F^s, GQ^{s/2})$ is stabilizable and (F,H) is detectable, then $\{P_t\}$ converges (exponentially fast) to the unique stabilizing solution P of the DARE from all initial conditions $P_1 \geq 0$.*

The following theorem gives sufficient conditions for the sequence $\{P_t\}$ to converge to the strong solution of the DARE, even when there is no stabilizing solution.

Theorem 5.12 (Theorem 4.2 of de Souza et al. (1986)) *If* $R > 0$, $\begin{bmatrix} Q & S \\ S' & R \end{bmatrix} \geq 0$, *and* $P_1 - P \geq 0$, *then* $\{P_t\}$ *converges to the unique strong solution* P *of the DARE if and only if* (F, H) *is detectable.*

5.13 Fast Kalman Filter Algorithm: The CKMS Recursions

If the process $\{Y_t\}$ follows the state space model (5.4) and (5.5), with initial conditions $\mathrm{E}(x_1) = 0$ and $\mathrm{Var}(x_1) = \Pi$, the Kalman filter corresponding to the sample $\{Y_t : 1 \leq t \leq n\}$ is given for $t = 1, \ldots, n$ by the recursions

$$\begin{aligned} E_t &= Y_t - H\hat{x}_{t|t-1}, \quad \Sigma_t = HP_tH' + R \\ K_t &= (FP_tH' + GS)\Sigma_t^{-1}, \quad \hat{x}_{t+1|t} = F\hat{x}_{t|t-1} + K_tE_t \\ P_{t+1} &= FP_tF' + GQG' - (FP_tH' + GS)\Sigma_t^{-1}(FP_tH' + GS)' \\ &= (F - K_tH)P_tF' + (GQ - K_tS')G', \end{aligned} \tag{5.68}$$

initialized with $\hat{x}_{1|0} = 0$ and $P_1 = \Pi$. In the Kalman filter recursions, even if the system matrices, $\{F, G, H, R, Q, S\}$, are time-varying, the updating of the MSE matrix P_t requires the greatest computational effort. More specifically, it takes $O(n^3)$ operations (additions and multiplications of real numbers), where n is the dimension of the state vector, to update P_t to P_{t+1}, whether the matrices $\{F_t, G_t, H_t, R_t, S_t, Q_t\}$ are constant or not.

One would expect that constant-parameter problems should be easier to handle than similar time-varying problems. In fact, it turns out that the Kalman filter recursions for a constant-parameter state space model can be simplified by replacing the updating of P_t with a different recursion. As we shall see, the updating of the new recursion requires $O(n^2)$ rather than $O(n^3)$ flops for each update, a great reduction if n is large.

The new recursion is based on the increments

$$\delta P_t = P_{t+1} - P_t, \qquad t \geq 1, \tag{5.69}$$

and the matrices $F_{p,t} = F - K_tH$ and

$$\overline{K}_t = FP_tH' + GS. \tag{5.70}$$

First we need a lemma.

Lemma 5.11 *The increments of* Σ_t, $\overline{K}_t$, K_t, *and* $F_{p,t}$ *can be written as*

$$\Sigma_{t+1} - \Sigma_t = H\delta P_t H' \tag{5.71}$$

$$\overline{K}_{t+1} - \overline{K}_t = F\delta P_t H' \tag{5.72}$$

$$K_{t+1} - K_t = F_{p,t}\delta P_t H' \Sigma_{t+1}^{-1} \tag{5.73}$$

$$F_{p,t+1} = F_{p,t}\left(I - \delta P_t H' \Sigma_{t+1}^{-1} H\right). \tag{5.74}$$

Proof The proof of the first two equalities is straightforward.

$$\Sigma_{t+1} = R + HP_{t+1}H' = R + HP_tH' + H\delta P_t H' = \Sigma_t + H\delta P_t H'$$
$$\overline{K}_{t+1} = FP_{t+1}H' + GS = FP_tH' + GS + F\delta P_t H' = \overline{K}_t + F\delta P_t H'.$$

The third requires a little more effort.

$$\begin{aligned} K_{t+1} &= \left(FP_{t+1}H' + GS\right)\Sigma_{t+1}^{-1} = \left(FP_tH' + GS + F\delta P_t H'\right)\Sigma_{t+1}^{-1} \\ &= K_t\Sigma_t\Sigma_{t+1}^{-1} + F\delta P_t H'\Sigma_{t+1}^{-1} \\ &= K_t\left(\Sigma_{t+1} - H\delta P_t H'\right)\Sigma_{t+1}^{-1} + F\delta P_t H'\Sigma_{t+1}^{-1} \\ &= K_t + (F - K_tH)\,\delta P_t H'\Sigma_{t+1}^{-1} \\ &= K_t + F_{p,t}\delta P_t H'\Sigma_{t+1}^{-1}. \end{aligned}$$

Finally, the increment for $F_{p,t}$ follows from the following identities.

$$\begin{aligned} F_{p,t+1} &= F - K_{t+1}H = F - K_tH - F_{p,t}\delta P_t H'\Sigma_{t+1}^{-1}H \\ &= F_{p,t}\left(I - \delta P_t H'\Sigma_{t+1}^{-1}H\right). \end{aligned}$$

□

With the help of this lemma, it is not difficult to prove the following theorem.

Theorem 5.13 (A Generalized Stokes Identity)

$$\delta P_{t+1} = F_{p,t}\left[\delta P_t - \delta P_t H'\Sigma_{t+1}^{-1}H\delta P_t\right]F'_{p,t}. \tag{5.75}$$

Proof Using Lemma 5.11, we can write

$$\begin{aligned}\delta P_{t+1} = P_{t+2} - P_{t+1} &= F\delta P_t F' - K_{t+1}\Sigma_{t+1}K'_{t+1} + K_t\Sigma_t K'_t \\ &= \left(F_{p,t} + K_t H\right)\delta P_t \left(F'_{p,t} + H'K'_t\right) \\ &\quad - \left(K_t + F_{p,t}\delta P_t H'\Sigma^{-1}_{t+1}\right)\Sigma_{t+1}\left(K_t + F_{p,t}\delta P_t H'\Sigma^{-1}_{t+1}\right)' \\ &\quad + K_t\left(\Sigma_{t+1} - H\delta P_t H'\right)K'_t \\ &= F_{p,t}\left[\delta P_t - \delta P_t H'\Sigma^{-1}_{t+1}H\delta P_t\right]F'_{p,t}.\end{aligned}$$

□

Remark 5.7 Applying the matrix inversion lemma to the matrix $I + H'\Sigma_t^{-1}H\delta P_t$, we obtain the following alternative formula

$$\delta P_{t+1} = F_{p,t}P_t\left(I + H'\Sigma_t^{-1}H\delta P_t\right)F'_{p,t}.$$

◇

An immediate consequence of Eq. (5.75) is that the rank of δP_t will never exceed the rank of δP_1. Hence, although the matrices P_t may have full rank, the matrices δP_t can have low rank. This is the key to developing the fast recursions.

Since P_t is symmetric, so are δP_t, $t = 1, 2, \ldots, n$. Therefore, we can always express δP_1 as

$$\delta P_1 = L_1 M_1 L'_1,$$

where M_1 has dimension $a \times a$, L_1 is $k \times a$ and a is the rank of δP_1. More specifically, a is the rank of the matrix

$$F\Pi F' + GQG' - \overline{K}_1\Sigma_1^{-1}\overline{K}'_1 - \Pi.$$

Lemma 5.12 (Factorization of $\boldsymbol{\delta P_t}$) *Assume that $\delta P_1 = L_1 M_1 L'_1$, where M_1 is symmetric, nonsingular of dimension $a \times a$. Then,*

$$\delta P_t = L_t M_t L'_t, \qquad t \geq 1, \tag{5.76}$$

where M_t is symmetric, nonsingular of size $a \times a$. It can be defined, along with L_t, by the recursions

$$L_{t+1} = F_{p,t}L_t \tag{5.77}$$

$$M_{t+1} = M_t - M_t L'_t H'\Sigma^{-1}_{t+1}HL_t M_t. \tag{5.78}$$

Proof By induction. If $\delta P_t = L_t M_t L_t'$ holds, then substituting into (5.75) yields

$$\begin{aligned}\delta P_t &= F_{p,t}\left[L_t M_t L_t' - L_t M_t L_t' H' \Sigma_{t+1}^{-1} H L_t M_t L_t'\right] F_{p,t}' \\ &= F_{p,t} L_t \left[M_t - M_t L_t' H' \Sigma_{t+1}^{-1} H L_t M_t\right] L_t' F_{p,t}',\end{aligned}$$

which can be rewritten as $\delta P_{t+1} = L_{t+1} M_{t+1} L_{t+1}'$ by defining L_{t+1} and M_{t+1} as in (5.77) and (5.78). The matrix M_t is clearly symmetric. For the nonsingularity of M_{t+1} note that if M_t is nonsingular, then so is M_{t+1} since, by the matrix inversion lemma,

$$\begin{aligned}M_{t+1}^{-1} &= M_t^{-1} + L_t' H' \left[\Sigma_{t+1} - H L_t M_t L_t' H'\right]^{-1} H L_t \\ &= M_t^{-1} + L_t' H' \left[R + H P_{t+1} H' - H(P_{t+1} - P_t) H'\right]^{-1} H L_t \\ &= M_t^{-1} + L_t' H' \Sigma_t^{-1} H L_t, \qquad (5.79)\end{aligned}$$

which shows that the inverse of M_{t+1} is well defined in terms of the inverses of $\{M_t, \Sigma_t\}$. □

The factorization (5.76) leads immediately to the following fast algorithm, which we will call the Chandrasekhar–Kailath–Morf–Sidhu (CKMS) recursions (Chandrasekhar 1947a, 1947b; Morf, Sidhu, & Kailath 1974).

Theorem 5.14 (The CKMS Recursions) *The quantities $\overline{K}_t$ and Σ_t, given by (5.68) and (5.70), can be recursively computed by the following set of coupled recursions*

$$\overline{K}_{t+1} = \overline{K}_t - F L_t R_{r,t}^{-1} L_t' H' \qquad (5.80)$$

$$L_{t+1} = F L_t - \overline{K}_t \Sigma_t^{-1} H L_t = F_{p,t} L_t \qquad (5.81)$$

$$\begin{aligned}\Sigma_{t+1} &= \Sigma_t - H L_t R_{r,t}^{-1} L_t' H' \\ R_{r,t+1} &= R_{r,t} - L_t' H' \Sigma_t^{-1} H L_t, \qquad (5.82)\end{aligned}$$

where $F_{p,t} = F - K_t H$. If desired, the P_t matrices of the Kalman filter can be computed as

$$P_{t+1} = \Pi - \sum_{j=1}^{t} L_j R_{r,j}^{-1} L_j'.$$

The initial conditions are computed as follows. $\overline{K}_1 = F\Pi H' + GS$ and $\Sigma_1 = R + H\Pi H'$. Factor (nonuniquely)

$$\delta P_1 = F\Pi F' + GQG' - \overline{K}_1 \Sigma_1^{-1} \overline{K}_1' - \Pi$$

as $\delta P_1 = -L_1 R_{r,1}^{-1} L_1'$, *where* L_1 *has rank* $k \times a$ *and* $R_{r,1}$ *is symmetric, nonsingular and of dimension* $a \times a$, *to obtain the initial conditions* $\{L_1, R_{r,1}\}$.

Proof First, define $R_{r,t} = -M_t^{-1}$. Then, (5.79) is exactly the recursion (5.82). Moreover, combining (5.72) and (5.76) yields

$$\overline{K}_{t+1} = \overline{K}_t - FL_t M_t L_t' H' = \overline{K}_t - FL_t R_{r,t}^{-1} L_t' H',$$

which is just (5.80). Next note that (5.77) is just (5.81). To calculate the error covariance matrices, just consider that, by (5.76),

$$P_{t+1} = P_1 - \sum_{j=1}^{t} \delta P_j = \Pi - \sum_{j=1}^{t} L_j R_{r,j}^{-1} L_j'.$$

□

Remark 5.8 (Stationary Processes) If the process $\{Y_t\}$ is stationary, then

$$\Pi = F \Pi F' + GQG'$$

and $\delta P_1 = -\overline{K}_1 \Sigma_1^{-1} \overline{K}_1'$. Therefore, the initial conditions are

$$L_1 = \overline{K}_1 = N, \qquad R_{r,1} = \Sigma_1 = \gamma_Y(0).$$

Note that, by formula (5.17), N can be obtained directly from covariance data using the equation

$$\begin{bmatrix} \gamma_Y(1) \\ \gamma_Y(2) \\ \vdots \\ \gamma_Y(n) \end{bmatrix} = \begin{bmatrix} H \\ HF \\ \vdots \\ HF^{n-1} \end{bmatrix} N.$$

In the case in which the model is in echelon form, N can be obtained as in (5.63). ◇

Lemma 5.13 (The $\mathbf{R_{r,t}}$ Matrices Have Constant Inertia) *Assume* $R > 0$ *and* $\Pi = P_1 \geq 0$. *Then, the matrices* $R_{r,t}$ *have the same inertia for all* t. *More specifically,*

$$In\{R_{r,t}\} = In\{\delta P_1\} \tag{5.83}$$

where $\delta P_1 = P_2 - \Pi = F \Pi F' + GQG' - K_1 \Sigma_1 K_1' - \Pi$.

Proof Consider the block matrix

$$\begin{bmatrix} \Sigma_t & HL_t \\ L_t'H' & R_{r,t} \end{bmatrix}$$

and note that $R_{r,t+1}$ is its Schur complement with respect to Σ_t, while Σ_{t+1} is its Schur complement with respect to $R_{r,t}$. Hence, the lower–upper and upper–lower block triangular factorizations of the above matrix would lead to the following equalities

$$\begin{aligned} \begin{bmatrix} \Sigma_t & HL_t \\ L_t'H' & R_{r,t} \end{bmatrix} &= \begin{bmatrix} I & \\ L_t'H'\Sigma_t & I \end{bmatrix} \begin{bmatrix} \Sigma_t & \\ & R_{r,t+1} \end{bmatrix} \begin{bmatrix} I & \\ L_t'H'\Sigma_t & I \end{bmatrix}' \\ &= \begin{bmatrix} I & HL_tR_{r,t}^{-1} \\ & I \end{bmatrix} \begin{bmatrix} \Sigma_{t+1} & \\ & R_{r,t} \end{bmatrix} \begin{bmatrix} I & HL_tR_{r,t}^{-1} \\ & I \end{bmatrix}'. \end{aligned}$$

It then follows from Sylverster's law of inertia that the matrices $\mathrm{diag}(\Sigma_t, R_{r,t+1})$ and $\mathrm{diag}(\Sigma_{t+1}, R_{r,t})$ have the same inertia. But since both Σ_t and Σ_{t+1} are positive definite (in view of the assumptions $R > 0$ and $P_1 \geq 0$), we conclude that $R_{r,t+1}$ and $R_{r,t}$ should have the same inertia. □

Remark 5.9 ($\{\mathbf{P_t}\}$ is Monotone Nonincreasing in the Stationary Case) Since $L_1 = \overline{K}_1$ and $R_{r,1} = \Sigma_1 > 0$, by Lemma 5.13, $R_{r,t} > 0$ for all t. Using this fact and

$$P_t = \Pi - \sum_{j=1}^{t-1} L_j R_{r,j}^{-1} L_j',$$

we note that in the stationary case $P_{t+1} \leq P_t$ when $P_1 = \Pi \geq 0$ and $R > 0$. Also, $\Sigma_{t+1} \leq \Sigma_t$. ◇

5.14 CKMS Recursions Given Covariance Data

The process $\{Y_t\}$ is defined by covariance data given by the matrices $\{F, H, N_t\}$ and $\gamma_Y(t,t)$, where $N_t = F\Pi_t H' + GS = \mathrm{Cov}(x_{t+1}, Y_t)$, $\Pi_t = \mathrm{E}(x_t x_t')$ and $\gamma_Y(t,s) = \mathrm{E}(Y_t Y_s')$ is given by the formulae of Theorem 5.1. The initial conditions that are needed for the CKMS recursions are

$$\Sigma_1 = \gamma_Y(1,1), \qquad \overline{K}_1 = N_1.$$

To obtain $\{L_1, R_{r,1}\}$ we need to factor

$$\begin{aligned} P_2 - P_1 &= FP_1F' + GQG' - \overline{K}_1\Sigma_1^{-1}\overline{K}_1' - P_1 \\ &= FP_1F' + GQG' - N_1\gamma_Y^{-1}(1,1)N_1' - P_1 \\ &= \Pi_2 - N_1\gamma_Y^{-1}(1,1)N_1' - \Pi_1. \end{aligned}$$

This means that we still need to identify the difference matrix $\Pi_2 - \Pi_1$ from the given covariance data. To this end, consider the matrices

$$R_1 = \begin{bmatrix} \gamma_Y(1,1) & \gamma_Y(1,2) & \cdots & \gamma_Y(1,n) \\ \gamma_Y(2,1) & \gamma_Y(2,2) & \cdots & \gamma_Y(1,n) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_Y(n,1) & \gamma_Y(n,2) & \cdots & \gamma_Y(n,n) \end{bmatrix}$$

and

$$R_2 = \begin{bmatrix} \gamma_Y(2,2) & \gamma_Y(2,3) & \cdots & \gamma_Y(2,n+1) \\ \gamma_Y(3,2) & \gamma_Y(3,3) & \cdots & \gamma_Y(3,n+1) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_Y(n+1,2) & \gamma_Y(n+1,3) & \cdots & \gamma_Y(n+1,n+1) \end{bmatrix}.$$

Then, it is not difficult to verify that

$$\begin{aligned} R_2 - R_1 &= \begin{bmatrix} H \\ HF \\ \vdots \\ HF^{n-1} \end{bmatrix} (\Pi_2 - \Pi_1) \left[H' \; F'H' \cdots (HF^{n-1})' \right] \\ &= O_n(\Pi_2 - \Pi_1)O_n', \end{aligned}$$

where O_n is the observability matrix. For example,

$$\begin{aligned} \gamma_Y(3,2) - \gamma_Y(2,1) &= HN_2 - HN_1 = H(F\Pi_2H' + GS) - H(F\Pi_1H' + GS) \\ &= HF(\Pi_2 - \Pi_1)H'. \end{aligned}$$

If $\{F, H\}$ is observable, O_n is of full rank and we can obtain $\Pi_2 - \Pi_1$ as

$$\Pi_2 - \Pi_1 = O_n^{\#}(R_2 - R_1)O_n^{\#'},$$

where $O_n^{\#} = (O_n' O_n)^{-1} O_n'$. Once $\Pi_2 - \Pi_1$ is known, $P_2 - P_1 = -L_1 R_{r,1}^{-1} L_1'$. With the initial conditions $\{\overline{K}_1, L_1, \Sigma_1, R_{r,t}\}$ so determined, we can proceed with the CKMS recursions.

Remark 5.10 (Stationary Processes) If the process $\{Y_t\}$ is stationary, then

$$P_2 - P_1 = -N_1 \gamma_Y^{-1}(0) N_1'$$

because $\Pi_2 = \Pi_1 = \Pi$. Then, $\overline{K}_1 = N_1 = N$ and $R_{r,1} = \gamma_Y(0) = \Sigma_1$. Note that, by formula (5.17), N can be obtained directly from covariance data using the equation

$$\begin{bmatrix} \gamma_Y(1) \\ \gamma_Y(2) \\ \vdots \\ \gamma_Y(n) \end{bmatrix} = \begin{bmatrix} H \\ HF \\ \vdots \\ HF^{n-1} \end{bmatrix} N = O_n N.$$

In the case in which the model is in echelon form, N can be obtained as in (5.63). ◇

5.15 Fast Covariance Square Root Filter

Define a **signature matrix** J as a matrix of the form

$$J = \begin{bmatrix} I_p & \\ & -I_q \end{bmatrix},$$

and, given a signature matrix J, define a **J-orthogonal matrix** Θ as a matrix that satisfies

$$\Theta J \Theta' = \Theta' J \Theta = J.$$

We will often write J as $J = (I_p \oplus -I_q)$.

At any time instant, t, we introduce a (nonunique) factorization of the form

$$P_{t+1} - P_t = \overline{L}_t J_t \overline{L}_t',$$

where $\overline{L}_t$ is an $n \times \alpha_t$ matrix, J_t is an $\alpha_t \times \alpha_t$ signature matrix with as many ± 1's as $(P_{t+1} - P_t)$ has positive and negative eigenvalues, and $\alpha_t = \text{rank}(P_{t+1} - P_t)$. The time subscript, t, is used in both J_t and α_t for generality but in fact it turns out that they are constant. To show this, form the pre-array

$$A = \begin{bmatrix} \Sigma_t^{1/2} & H\overline{L}_t \\ \overline{K}_{p,t} & F\overline{L}_t \end{bmatrix}, \qquad \overline{K}_{p,t} = \overline{K}_t \Sigma_t^{-1/2'}, \qquad \overline{K}_t = FP_tH' + GS,$$

and triangularize it via an $(I \oplus J_t)$-orthogonal matrix Θ, that is,

$$A\Theta = \begin{bmatrix} \Sigma_t^{1/2} & H\overline{L}_t \\ \overline{K}_{p,t} & F\overline{L}_t \end{bmatrix} \Theta = \begin{bmatrix} X & \\ Y & Z \end{bmatrix} \tag{5.84}$$

for some Θ such that

$$\Theta \begin{bmatrix} I & \\ & J_t \end{bmatrix} \Theta' = \begin{bmatrix} I & \\ & J_t \end{bmatrix} = \Theta' \begin{bmatrix} I & \\ & J_t \end{bmatrix} \Theta.$$

To prove that such a Θ exists, consider first the equality

$$\Sigma_{t+1} = R + HP_{t+1}H' = \Sigma_t + H\overline{L}_t J_t \overline{L}_t' H'$$

or, equivalently,

$$\begin{bmatrix} \Sigma_t^{1/2} & H\overline{L}_t \end{bmatrix} \begin{bmatrix} I & \\ & J_t \end{bmatrix} \begin{bmatrix} \Sigma_t^{1/2'} \\ \overline{L}_t' H' \end{bmatrix} = \begin{bmatrix} \Sigma_{t+1}^{1/2} & 0 \end{bmatrix} \begin{bmatrix} I & \\ & J_t \end{bmatrix} \begin{bmatrix} \Sigma_{t+1}^{1/2'} \\ 0 \end{bmatrix}.$$

To proceed further, we need a lemma.

Lemma 5.14 *Let A and B be $n \times m$ matrices ($n \leq m$) and let $J = (I_p \oplus -I_q)$ be a signature matrix with $p + q = m$. If $AJA' = BJB'$ is full rank, then there exists a J-orthogonal matrix Θ such that $A = B\Theta$.*

Proof Since AJA' is symmetric and invertible, we can factor it as $AJA' = RSR'$, where R is invertible and $S = (I_\alpha \oplus -I_\beta)$ is a signature matrix (with $\alpha + \beta = n$). We normalize A and B by defining $\overline{A} = R^{-1}A$ and $\overline{B} = R^{-1}B$. Then, $\overline{A}J\overline{A}' = \overline{B}J\overline{B}' = S$.

Now consider the block triangular factorizations

$$\begin{aligned} \begin{bmatrix} S & \overline{A} \\ \overline{A}' & J \end{bmatrix} &= \begin{bmatrix} I & \\ \overline{A}'S & I \end{bmatrix} \begin{bmatrix} S & \\ & J - \overline{A}'S\overline{A} \end{bmatrix} \begin{bmatrix} I & \\ \overline{A}'S & I \end{bmatrix}' \\ &= \begin{bmatrix} I & \overline{A}J \\ & I \end{bmatrix} \begin{bmatrix} S - \overline{A}J\overline{A}' & \\ & J \end{bmatrix} \begin{bmatrix} I & \overline{A}J \\ & I \end{bmatrix}', \end{aligned}$$

where $S - \overline{A}J\overline{A}' = 0$. Using the fact that the central matrices must have the same inertia we conclude that $In(J - \overline{A}'S\overline{A}) = In(J) - In(S) = \{p-\alpha, q-\beta, n\}$. Similarly, we can show that $In(J - \overline{B}'S\overline{B}) = \{p-\alpha, q-\beta, n\}$.

Define the signature matrix $J_1 = (I_{p-\alpha} \oplus -I_{q-\beta})$. The above inertia conditions then mean that we can factor $J - \overline{A}'S\overline{A}$ and $J - \overline{B}'S\overline{B}$ as

$$J - \overline{A}'S\overline{A} = XJ_1X', \qquad J - \overline{B}'S\overline{B} = YJ_1Y'.$$

Finally, introduce the square matrices

$$\Sigma_1 = \begin{bmatrix} \overline{A} \\ X' \end{bmatrix}, \qquad \Sigma_2 = \begin{bmatrix} \overline{B} \\ Y' \end{bmatrix}.$$

It is easy to verify that these matrices satisfy $\Sigma_1'(S \oplus J_1)\Sigma_1 = J$ and $\Sigma_2'(S \oplus J_1)\Sigma_2 = J$, which further shows that Σ_1 and Σ_2 are invertible. Therefore, we also obtain that $\Sigma_1 J \Sigma_1' = S \oplus J_1$ and $\Sigma_2 J \Sigma_2' = S \oplus J_1$. These relations allow us to relate Σ_1 and Σ_2 as $\Sigma_1 = \Sigma_2 \left[J\Sigma_2'(S \oplus J_1)\Sigma_1\right]$. If we set $\Theta = \left[J\Sigma_2'(S \oplus J_1)\Sigma_1\right]$, then it is immediate to check that Θ is J-orthogonal and, from the equality of the first block row of $\Sigma_1 = \Sigma_2\Theta$, that $\overline{A} = \overline{B}\Omega$. Hence, $A = B\Theta$. □

By the previous lemma, there exists a $(I \oplus J_t)$-orthogonal rotation Θ relating the following arrays

$$\begin{bmatrix} \Sigma_t^{1/2} & H\overline{L}_t \end{bmatrix} \Theta = \begin{bmatrix} \Sigma_{t+1}^{1/2} & 0 \end{bmatrix}.$$

This proves (5.84). To identify X, Y, Z, form

$$\begin{bmatrix} \Sigma_t^{1/2} & H\overline{L}_t \\ \overline{K}_{p,t} & F\overline{L}_t \end{bmatrix} \Theta \begin{bmatrix} I & \\ & J_t \end{bmatrix} \Theta' \begin{bmatrix} \Sigma_t^{1/2} & H\overline{L}_t \\ \overline{K}_{p,t} & F\overline{L}_t \end{bmatrix}' = \begin{bmatrix} X & \\ Y & Z \end{bmatrix} \begin{bmatrix} I & \\ & J_t \end{bmatrix} \begin{bmatrix} X & \\ Y & Z \end{bmatrix}'.$$

Then,

$$\begin{aligned} XX' &= \Sigma_t + H\overline{L}_t J_t \overline{L}_t' H' = \Sigma_t + H(P_{t+1} - P_t)H' \\ &= R + HP_tH' + HP_{t+1}H' - HP_tH' = R + HP_{t+1}H' \\ &= \Sigma_{t+1}, \end{aligned}$$

$$\begin{aligned} XY' &= \overline{K}_t + F\overline{L}_t J_t \overline{L}_t' H' = \overline{K}_t + F(P_{t+1} - P_t)H' \\ &= (FP_tH' + GS) + FP_{t+1}H' - FP_tH' = GS + FP_{t+1}H' \\ &= K_{t+1}, \end{aligned}$$

and

$$
\begin{aligned}
YY' + ZJ_tZ' &= \overline{K}_t\Sigma_t^{-1}\overline{K}_t' + F\overline{L}_tJ_t\overline{L}_t'H' \\
&= \overline{K}_t\Sigma_t^{-1}\overline{K}_t' + FP_{t+1}F' - FP_tF'.
\end{aligned}
$$

Therefore, $X = \Sigma_{t+1}^{1/2}$, $Y = \overline{K}_{t+1}\Sigma_{t+1}^{-1/2'} = \overline{K}_{p,t+1}$, $ZJ_tZ' = P_{t+2} - P_{t+1}$, and $Z = \overline{L}_{t+1}$.

We can set $J_{t+1} = J_t$ since, by definition, $P_{t+2} - P_{t+1} = \overline{L}_{t+1}J_{t+1}\overline{L}_{t+1}'$. So, in fact, we can choose J_{t+1} to be equal to the first inertia matrix J, defined by the factorization

$$
P_2 - P_1 = (F\Pi F' + GQG' - \overline{K}_1\Sigma_1^{-1}\overline{K}_1') - \Pi = \overline{L}_1J\overline{L}_1'. \tag{5.85}
$$

Likewise, α_i can be chosen to be equal to the size, α, of J. In summary, the above derivation shows that $\{\overline{K}_{p,t}, \Sigma_t^{1/2}\}$ can be recursively updated via the array algorithm

$$
\Theta \left[\begin{array}{c|c} \Sigma_t^{1/2'} & \overline{K}_{p,t}' \\ (H\overline{L}_t)' & (F\overline{L}_t)' \end{array} \right] = \left[\begin{array}{c|c} \Sigma_{t+1}^{1/2'} & \overline{K}_{p,t+1}' \\ 0 & \overline{L}_{t+1}' \end{array} \right],
$$

where Θ is any $(I \oplus J)$-orthogonal matrix that produces the block zero entry in the post-array. Moreover, the initial conditions are $\Sigma_1 = R + H\Pi H'$ and $\overline{K}_1 = F\Pi H' + GS$, with $(\overline{L}_1, J)$ obtained via the factorization (5.85).

We should note that the $(I \oplus J)$-orthogonal transformation Θ can be implemented in several ways, especially by a sequence of elementary J-orthogonal Householder transformations.

Remark 5.11 (Stationary Processes) If the process $\{Y_t\}$ is stationary, from (5.85) we get

$$
P_2 - P_1 = -\overline{K}_1\Sigma_1^{-1}\overline{K}_1' = \overline{L}_1J\overline{L}_1',
$$

where $\Sigma_1 = R + H\Pi H'$ and $\overline{K}_1 = F\Pi H' + GS$. Then, $\overline{L}_1 = \overline{K}_1\Sigma_1^{-1/2'} = \overline{K}_{p,1}$ and $J = -I$. Note that, in terms of the autocovariance matrices of the process, $\overline{K}_1 = N$ and $\Sigma_1 = \gamma_Y(0)$. Thus, by formula (5.17), N can be obtained directly from covariance data using the equation

$$
\begin{bmatrix} \gamma_Y(1) \\ \gamma_Y(2) \\ \vdots \\ \gamma_Y(n) \end{bmatrix} = \begin{bmatrix} H \\ HF \\ \vdots \\ HF^{n-1} \end{bmatrix} N = O_nN.
$$

In the case in which the model is in echelon form, N can be obtained as in (5.63).

The $(I \oplus -I)$-orthogonal matrix Θ has dimension $2k \times 2k$, where $k = \dim(Y_t)$, so that if $\{Y_t\}$ is univariate, Θ is 2×2. It seems that the fast square root covariance filter described in this section is easier and stabler than the CKMS recursions. ◇

5.15.1 J-*Orthogonal Householder Transformations*

Assume a signature matrix $J = (I_p \oplus -I_q)$ with $p, q \geq 1$. A J-orthogonal transformation Θ is a matrix that satisfies

$$\Theta J \Theta' = \Theta' J \Theta = J.$$

If $J = I$, Θ is orthogonal.

Given two column vectors, x and y, define the "J-product" as

$$\langle x, y \rangle_J = x' J y.$$

This is not a scalar product. However, for our purposes we can define the squared "J-norm" of a column vector x as

$$||x||_J^2 = x' J x.$$

Note that we use the term J-norm loosely because, strictly speaking, the quantity $x'Jx$ does not define a norm.

J-orthogonal transformations preserve the squared J-norm of a vector. That is, if $y = \Theta x$, then

$$||y||_J^2 = y' J y = x' \Theta' J \Theta x = x' J x = ||x||_J^2.$$

Consider an n-dimensional real valued vector $x = [x_1, \ldots, x_n]'$ and suppose that we wish to simultaneously annihilate several entries in it by using a J-orthogonal involutory matrix Θ (i.e., a transformation Θ that satisfies $\Theta J \Theta' = J = \Theta' J \Theta$ and $\Theta^2 = I$), say

$$\Theta [x_1, x_2, \ldots, x_n]' = \alpha e_1 = \alpha [1, 0, \ldots, 0]' \tag{5.86}$$

for some real scalar α. Clearly, the transformation (5.86) is only possible if the vector x has positive squared J-norm because $||e_1||_J^2 = 1$ and $||\Theta x||_J^2 = ||x||_J^2 = |\alpha|^2$. Otherwise, when x has negative squared J-norm, we should seek a transformation Θ that reduces x to the form

$$\Theta [x_1, x_2, \ldots, x_n]' = \alpha e_n = \alpha [0, 0, \ldots, 1]'.$$

This is possible because the squared J-norm of αe_n is equal to $-|\alpha|^2$, which is negative. We thus consider two cases: $||x||_J^2 > 0$ and $||x||_J^2 < 0$.

(a) Suppose $||x||_J^2 > 0$, choose $\alpha = \pm\sqrt{||x||_J^2}$ and let $y = x - \alpha e_1$. Using a geometric argument and proceeding as if the product $<,>_J$ replaced the ordinary scalar product in a Householder transformation, the "projection" of x onto y is $\langle x, y\rangle_J ||y||_J^{-2} y$. Then, imposing the symmetry condition of the Householder transformation, we get

$$\begin{aligned}\alpha e_1 &= x - 2 < y, x >_J ||y||_J^{-2} y \\ &= x - 2y'Jx(y'Jy)^{-1}y \\ &= \left[I - \frac{2}{y'Jy} yy'J\right] x.\end{aligned}$$

The matrix $\Theta = I - (2/y'Jy)\, yy'J$ is called a hyperbolic Householder matrix. It is both J-orthogonal and involutory. Thus, if $y = x \pm \sqrt{||x||_J^2} e_1$, then

$$\Theta x = \mp\sqrt{||x||_J^2} e_1, \quad \Theta = I - \big(2/y'Jy\big)\, yy'J.$$

(b) Suppose $||x||_J^2 < 0$, choose $\alpha = \pm\sqrt{-||x||_J^2}$ and let $y = x - \alpha e_n$. Proceeding as in the previous case, we are led to the equality

$$\begin{aligned}\alpha e_n &= x - 2y'Jx(y'Jy)^{-1}y \\ &= \left[I - \frac{2}{y'Jy} yy'J\right] x.\end{aligned}$$

Thus, if $y = x \pm \sqrt{-||x||_J^2} e_n$, then $\Theta x = \mp\sqrt{-||x||_J^2} e_n$, $\Theta = I - (2/y'Jy)\, yy'J$.

Another possibility is to define

$$\alpha = \begin{cases} \pm\sqrt{||x||_J^2} & \text{if } ||x||_J^2 > 0 \\ \pm\sqrt{-||x||_J^2} & \text{if } ||x||_J^2 < 0, \end{cases}$$

$$v = \pm\left(\frac{1}{\alpha} x + e_i\right),$$

and

$$\Theta = I - \frac{1}{v_i} v v' J,$$

where $v = [v_1, \ldots, v_n]'$, $i = 1$ if $||x||_J^2 > 0$, and $i = n$ if $||x||_J^2 < 0$. It can be verified that, with the previous definitions, it holds that $\Theta' J \Theta = \Theta J \Theta' = J$ and $\Theta x = -\alpha e_i$.

5.16 The Likelihood of a Stationary State Space Model

Let Y $= (Y_1', \ldots, Y_n')'$ be a k-dimensional time series generated by a stationary state space model under the normality assumption. Then, letting $\text{Var}(Y) = \sigma^2 \Omega$, the density of Y is

$$f_Y(x) = (2\pi)^{-k/2} |\sigma^2 \Omega|^{-1/2} e^{-x' \Omega^{-1} x/(2\sigma^2)}.$$

The innovations E_t generated by the Kalman filter are uncorrelated with $\text{Var}(E_t) = \Sigma_t \sigma^2$ (notice that in the Kalman filter it is assumed that $\sigma^2 = 1$). Letting $E = (E_1', \ldots, E_n')'$, by the results of Sect. 1.4, we have $E = WY$ and $\text{diag}(\Sigma_t) = W\text{Var}(Y)W' / \sigma^2$, where W is a lower triangular matrix with ones in the diagonal. Thus, using the prediction error decomposition

$$f_Y = f_{n|n-1} \cdots f_{2|1} f_1,$$

where $f_{j|j-1}$ is the density of the conditional distribution $Y_j | Y_{j-1}, \ldots, Y_1$, the log-likelihood is

$$l(Y) = \text{constant} - \frac{1}{2} \left\{ \frac{1}{\sigma^2} \sum_{t=1}^{n} E_t' \Sigma_t^{-1} E_t + \sum_{t=1}^{n} \ln |\sigma^2 \Sigma_t| \right\}.$$

Differentiating with respect to σ^2 in the previous expression and equating to zero yields the maximum likelihood estimator $\hat{\sigma}^2 = \sum_{t=1}^{n} E_t' \Sigma_t^{-1} E_t / (nk)$ of σ^2. Thus, σ^2 can be concentrated out of the likelihood and the σ^2-maximized log-likelihood is

$$c(Y) = \text{constant} - \frac{1}{2} \left\{ (nk) \ln\left(\sum_{t=1}^{n} E_t' \Sigma_t^{-1} E_t\right) + \sum_{t=1}^{n} \ln |\Sigma_t| \right\}. \quad (5.87)$$

5.17 The Innovations Algorithm Approach for Likelihood Evaluation

The innovations algorithm can be used as an alternative to the Kalman filter for likelihood evaluation because it also gives the innovations E_t and their covariance matrices D_t recursively. In the case of a VARMA(p, q) model, a transformation due to Ansley (1979) simplifies the equations considerably. Let $\{Y_t\}$ be a k-dimensional stationary process that follows the model

$$\Phi(B)Y_t = \Theta(B)A_t,$$

where $\Phi(B) = I + \Phi_1 B + \cdots + \Phi_p B^p$, $\Theta(B) = I + \Theta_1 B + \cdots + \Theta_q B^q$, and $\{A_t\} \sim \text{WN}(0, \sigma^2\Sigma)$. Assume that we have a sequence of vectors, $Y = (Y_1', \ldots, Y_n')'$, generated by the previous model and such that $\text{Var}(Y)$ is positive definite. Define

$$X_t = \begin{cases} Y_t, & \text{if } t = 1, \ldots, r, \\ \Phi(B)Y_t, & \text{if } t > r, \end{cases}$$

where $r = \max(p, q)$. Then, the Jacobian of the transformation is one. Assuming that $\sigma^2 = 1$ because this parameter can be concentrated out of the likelihood as we saw in the derivation of (5.87), and denoting the covariance function of $\{Y_t\}$ by $\gamma_Y(h) = E(Y_{t+h}Y_t')$, it is not difficult to verify that the covariance function $S_{ij} = E(X_i X_j')$ is given by

$$S_{ij} = \begin{cases} \gamma_Y(i-j) & \text{if } 1 \le i \le j \le r \\ \gamma_Y(i-j) + \sum_{h=1}^{p} \Phi_h \gamma_Y(i+h-j) & \text{if } 1 \le i \le r < j \le 2r \\ \sum_{h=0}^{q} \Theta_h \Sigma \Theta_{h+j-i}' & \text{if } r < i \le j \le i+q \\ 0 & \text{if } r < i \quad \text{and} \quad i+q < j \\ S_{ij}' & \text{if } j < i, \end{cases} \tag{5.88}$$

where $\Theta_0 = I$ and $\Theta_h = 0$ if $h > q$. The notable feature of the previous transformation is that $S_{ij} = 0$ if $|j - i| > q$, $i, j > r$. This in turn implies that

$$X_t = \Theta_{t,t-q}E_{t-q} + \cdots + \Theta_{t,t-1}E_{t-1} + E_t, \qquad t > r,$$

when the innovations algorithm is applied to $\{X_t : t = 1, 2, \ldots, n\}$. More specifically, the output of this last algorithm in terms of $\{Y_t : t = 1, 2, \ldots, n\}$ is easily shown to be

$$Y_t = \Theta_{t,1}E_1 + \cdots + \Theta_{t,t-1}E_{t-1} + E_t, \qquad t \le r,$$

where

$$\Theta_{tj} = \left[\gamma_Y(t-j) - \sum_{i=1}^{j-1} \Theta_{ti} D_i \Theta_{ji}'\right] D_j^{-1}, \qquad j = 1, \ldots, t-1,$$

$$D_t = \gamma_Y(0) - \sum_{i=1}^{t-1} \Theta_{ti} D_i \Theta_{ti}',$$

and

$$Y_t + \sum_{j=1}^{p} \Phi_j Y_{t-j} = \Theta_{t,t-q} E_{t-q} + \cdots + \Theta_{t,t-1} E_{t-1} + E_t, \qquad t > r, \tag{5.89}$$

where

$$\Theta_{tj} = \left(S_{tj} - \sum_{i=t-q}^{j-1} \Theta_{ti} D_i \Theta'_{ji} \right) D_j^{-1}, \qquad j = t-q, \ldots, t-1,$$

$$D_t = S_{tt} - \sum_{i=t-q}^{t-1} \Theta_{ti} D_i \Theta'_{ti},$$

and S_{ij} is given by (5.88). In addition, since the matrix of the transformation that gives $(X'_1, \ldots, X'_t)'$ in terms of $(Y'_1, \ldots, Y'_t)'$ is easily seen to be nonsingular for $t = 1, 2, \ldots, n$, it follows from Proposition 1.7 that $E^*(Y_t|X_{t-1}, \ldots, X_1) = E^*(Y_t|Y_{t-1}, \ldots, Y_1)$. This implies $E^*(X_t|X_{t-1}, \ldots, X_1) = E^*(Y_t|Y_{t-1}, \ldots, Y_1)$ for $t = 1, \ldots, r$, and $E^*(X_t|X_{t-1}, \ldots, X_1) = E^*(Y_t|Y_{t-1}, \ldots, Y_1) + \sum_{j=1}^{p} \Phi_j Y_{t-j}$ for $t = r+1, \ldots, n$. Therefore,

$$E_t = X_t - E^*(X_t|X_{t-1}, \ldots, X_1) = Y_t - E^*(Y_t|Y_{t-1}, \ldots, Y_1), \qquad t = 1, 2, \ldots,$$

and the E_t and D_t given by the innovations algorithm applied to $\{X_t : t = 1, 2, \ldots, n\}$ can be used for likelihood evaluation directly in (5.87).

5.18 Finite Forecasting

One-period-ahead forecasts are given by the Kalman filter. Denoting by $x_{n+h|n}$, where $h > 1$, the orthogonal projection of x_{n+h} onto the sample $Y = (Y'_1, \ldots, Y'_n)'$, if the one source of error state space model (5.1) and (5.2) is used, it can be shown that h-period-ahead forecasts and their mean squared error $\Sigma_{n+h|n}$ are obtained recursively by

$$\begin{aligned} x_{n+h|n} &= F_{n+h-1} x_{n+h-1|n} \\ \Sigma_{n+h|n} &= (F_{n+h-1} \Sigma_{n+h-1|n} F'_{n+h-1} + G_{n+h-1} G'_{n+h-1}) \sigma^2, \end{aligned}$$

where $x_{n+1|n} = x_{n+1}$ and $\Sigma_{n+1|n} = \Sigma_{n+1}$. The forecasts for $Y_{n+h} = E^*(Y_{n+h}|Y_n, \ldots, Y_1)$, where $h \geq 1$, and the corresponding mean squared error matrices are given by

$$\begin{aligned} Y_{n+h|n} &= H_{n+h-1} x_{n+h|n} \\ \mathrm{MSE}(Y_{n+h|n}) &= (H_{n+h-1} \Sigma_{n+h-1|n} H'_{n+h-1} + J_{n+h-1} J'_{n+h-1}) \sigma^2. \end{aligned}$$

5.19 Finite Forecasting Using the Innovations Algorithm

If the innovations algorithm is applied and the sample is $Y = (Y_1', \ldots, Y_n')'$, the forecasts $Y_{n+h|n} = E^*(Y_{n+h}|Y_n, \ldots, Y_1)$, $h = 1, 2, \ldots$, and their mean square error matrices are

$$Y_{n+h|n} = \Theta_{n+h,n}E_n + \cdots + \Theta_{n+h,1}E_1,$$

$$\text{MSE}(Y_{n+h|n}) = D_{n+h} + \sum_{j=n+1}^{n+h-1} \Theta_{n+h,j}D_j\Theta_{n+h,j}',$$

where the E_t, $t = 1, \ldots, n$, are the innovations and $\text{Var}(E_t) = D_t$.

In the case of a VARMA(p, q) model, the equations for the forecasts can be simplified by using the transformation suggested by Ansley (1979) and described in Sect. 5.17. Let $\{Y_t\}$ be a p-dimensional stationary process that follows the model

$$\Phi(B)Y_t = \Theta(B)A_t,$$

where $\Phi(B) = I + \Phi_1 B + \cdots + \Phi_p B^p$, $\Theta(B) = I + \Theta_1 B + \cdots + \Theta_q B^q$ and $\{A_t\}$ $\sim$WN$(0, \Sigma)$. Then, assuming $t > r = \max(p, q)$, as is always the case for practical prediction, and using (5.89), the forecasts $Y_{n+h|n}$ can be obtained recursively from

$$Y_{n+h|n} + \sum_{j=1}^{p} \Phi_j Y_{n+h-j|n} = \Theta_{n+h,n+h-q}E_{n+h-q} + \cdots + \Theta_{n+h,n}E_n, \tag{5.90}$$

where $Y_{n+h-j|n} = Y_{n+h-j}$ if $n + h - j \leq n$. Notice that

$$Y_{n+h|n} + \sum_{j=1}^{p} \Phi_j Y_{n+h-j|n} = 0, \qquad h > r, \tag{5.91}$$

because $\Theta_{n+h,n+h-j} = 0$ for $j > q$. Equation (5.91) is usually called **the eventual forecast function**. To obtain the mean square error, consider that

$$Y_{n+h} + \sum_{j=1}^{p} \Phi_j Y_{n+h-j} = \Theta_{n+h,n+h-q}E_{n+h-q} + \cdots + \Theta_{n+h,n+h-1}E_{n+h-1} + E_{n+h}. \tag{5.92}$$

Then, subtracting (5.90) from (5.92), we get

$$Y_{n+h} - Y_{n+h|n} + \sum_{j=1}^{p} \Phi_j(Y_{n+h-j} - Y_{n+h-j|n}) = \sum_{j=1}^{h-1} \Theta_{n+h,n+j}E_{n+j} + E_{n+h}, \quad h = 1, 2, \ldots,$$

where $\Theta_{n+h,n+j} = 0$ if $j < h - q$. The previous equations can be written in matrix form as

$$\Phi_{(n,h)} Y_{(n,h)} = \Theta_{(n,h)} E_{(n,h)},$$

where

$$\Phi_{(n,h)} = \begin{bmatrix} I & & & & & \\ \Phi_1 & I & & & & \\ \vdots & \ddots & \ddots & & & \\ \Phi_p & \cdots & \Phi_1 & I & & \\ & \ddots & & \ddots & \ddots & \\ & & \Phi_p & \cdots & \Phi_1 & I \end{bmatrix}, \qquad Y_{(n,h)} = \begin{bmatrix} Y_{n+1} - Y_{n+1|n} \\ Y_{n+2} - Y_{n+2|n} \\ \vdots \\ Y_{n+p} - Y_{n+p|n} \\ \vdots \\ Y_{n+h} - Y_{n+h|n} \end{bmatrix},$$

$$\Theta_{(n,h)} = \begin{bmatrix} I & & & & & \\ \Theta_{n+2,n+1} & I & & & & \\ \vdots & \ddots & \ddots & & & \\ \Theta_{n+q+1,n+1} & \cdots & \Theta_{n+q+1,n+q} & I & & \\ & \ddots & & \ddots & \ddots & \\ & & \Theta_{n+h,n+h-q} & \cdots & \Theta_{n+h,n+h-1} & I \end{bmatrix}, \; E_{(n,h)} = \begin{bmatrix} E_{n+1} \\ E_{n+2} \\ \vdots \\ E_{n+q+1} \\ \vdots \\ E_{n+h} \end{bmatrix},$$

and $\Theta_{(n,h)}$ is a band matrix because $\Theta_{n+i,n+j} = 0$ if $j < i - q$. Letting $\mathrm{Var}(Y_{(n,h)}) = V$ and $\mathrm{Var}(E_{(n,h)}) = D_{(n,h)} = \mathrm{diag}(D_{n+1}, \ldots, D_{n+h})$, we can write

$$\Phi_{(n,h)} V \Phi'_{(n,h)} = \Theta_{(n,h)} D_{(n,h)} \Theta'_{(n,h)},$$

and

$$V = \Phi^{-1}_{(n,h)} \Theta_{(n,h)} D_{(n,h)} \Theta'_{(n,h)} \Phi'^{-1}_{(n,h)}.$$

It is not difficult to verify that

$$\Phi^{-1}_{(n,h)} = \begin{bmatrix} I & & & \\ \Xi_1 & I & & \\ \vdots & \ddots & \ddots & \\ \Xi_{h-1} & \cdots & \cdots \Xi_1 & I \end{bmatrix},$$

where the Ξ_j weights are given by $\Phi(z)(\sum_{j=0}^{\infty} \Xi_j z^j) = I$, and thus the matrix $\Psi_{(n,h)} = \Phi^{-1}_{(n,h)} \Theta_{(n,h)}$ can be easily computed. Since the block diagonal elements of V are the MSE$(Y_{n+j|n})$, $j = 1, 2, \ldots, h$, if we define

$$\Psi_{(n,h)} = \begin{bmatrix} I & & & \\ \Psi_{n+2,1} & I & & \\ \vdots & \ddots & \ddots & \\ \Psi_{n+h,1} & \cdots & \Psi_{n+h,h-1} & I \end{bmatrix},$$

we get

$$\text{MSE}(Y_{n+h|n}) = D_{n+h} + \sum_{j=1}^{h-1} \Psi_{n+h,j} D_{n+j} \Psi'_{n+h,j}.$$

5.20 Inverse Process

The **inverse process**, $\{Y_t^i\}$, of a stationary process, $\{Y_t\}$, that follows the linear time series model $Y_t = \sum_{j=0}^{\infty} \Psi_j A_{t-j}$, where $\{A_t\} \sim$WN$(0, \Sigma)$, is defined by

$$Y_t^i = \Psi^{-1'}(B^{-1})\Sigma^{-1}A_t, \tag{5.93}$$

where B is the backshift operator, $BY_t = Y_{t-1}$, $\Psi(z) = \sum_{j=0}^{\infty} \Psi_j z^j$ and the inverse of $\Psi(z)$ is assumed to exist. Note that the time runs backwards in (5.93). The name is justified because the covariance generating function of $\{Y_t^i\}$, $G_{Y^i}(z)$, is the inverse of the covariance generating function of $\{Y_t\}$, $G_Y(z)$. That is, $G_{Y^i}(z) = G_Y^{-1}(z)$. To see this, consider first that $G_Y(z) = \Psi(z)\Sigma\Psi'(z^{-1})$. Then,

$$\begin{aligned} G_{Y^i}(z) &= \Psi^{-1'}(z^{-1})\Sigma^{-1}\Psi^{-1}(z) \\ &= G_Y^{-1}(z). \end{aligned}$$

If the process has some structure, in particular, if $\{Y_t\}$ follows the innovations state space model in echelon form

$$x_{t+1} = Fx_t + KA_t \tag{5.94}$$

$$Y_t = Hx_t + A_t, \tag{5.95}$$

where $\{A_t\} \sim$WN$(0, \Sigma)$, then

$$Y_t^i = \left[I - B^{-1}K'(I - F_p'B^{-1})^{-1}H'\right]\Sigma^{-1}A_t, \tag{5.96}$$

where $F_p = F - KH$. To prove this, let

$$\begin{aligned} G_Y(z) &= [zH(I - Fz)^{-1} \quad I] \begin{bmatrix} K \\ I \end{bmatrix} \Sigma \, [K' \quad I] \begin{bmatrix} z^{-1}(I - F'z^{-1})^{-1}H' \\ I \end{bmatrix} \\ &= [I + zH(I - Fz)^{-1}K]\Sigma \, [I + z^{-1}K'(I - F'z^{-1})^{-1}H'] \\ &= \Psi(z)\Sigma\Psi'(z^{-1}), \end{aligned}$$

where $\Psi(z) = I + zH(I - Fz)^{-1}K$, be the covariance generating function of $\{Y_t\}$. Then, $\Psi(z)$ satisfies

$$\begin{aligned} Y_t &= [I + BH(I - FB)^{-1}K]A_t \\ &= \Psi(B)A_t, \end{aligned}$$

and is therefore the transfer function of $\{Y_t\}$. By the matrix inversion Lemma (4.1), the inverse of $\Psi(z)$, if it exists, is given by

$$\Psi^{-1}(z) = I - zH(I - F_p z)^{-1}K.$$

The inverse exists if, and only if, F_p is stable. It follows from this and (5.96) that $\Psi^{-1'}(z^{-1})$ is the transfer function of the inverse process, that is

$$Y_t^i = \Psi^{-1'}(B^{-1})\Sigma^{-1}A_t$$

as claimed. It follows from (5.96) that a state space model that realizes the inverse process is

$$\lambda_t = F_p'\lambda_{t+1} + H'A_t^i \tag{5.97}$$

$$Y_t^i = -K'\lambda_{t+1} + A_t^i, \tag{5.98}$$

where $\mathrm{Var}(A_t^i) = \Sigma^{-1}$ and the time runs backwards.

Since state space models and VARMA models are equivalent, if the stationary process $\{Y_t\}$ follows (5.94) and (5.95), it also follows a VARMA process

$$\Phi(B)Y_t = \Theta(B)A_t, \tag{5.99}$$

and the question arises as to what is the VARMA model followed by the inverse process. To answer this question, note first that the transfer function of $\{Y_t\}$ can be expressed in terms of (5.99) as

$$\Psi(z) = \Phi^{-1}(z)\Theta(z).$$

Then, taking inverses, we get

$$\Psi^{-1}(z) = \Theta^{-1}(z)\Phi(z),$$

and this implies that the inverse process can be expressed as

$$Y_t^i = \Phi'(B^{-1})\Theta^{-1'}(B^{-1})\Sigma^{-1}A_t. \tag{5.100}$$

By the properties of MFDs (Kailath, 1980), there exist polynomial matrices, $\widetilde{\Phi}'(z^{-1})$ and $\widetilde{\Theta}'(z^{-1})$, with the same eigenvalues as $\Phi'(z^{-1})$ and $\Theta'(z^{-1})$ such that

$$Y_t^i = \widetilde{\Theta}^{-1'}(B^{-1})\widetilde{\Phi}'(B^{-1})A_t^i, \tag{5.101}$$

where $\mathrm{Var}(A_t^i) = \Sigma^{-1}$. Therefore, the inverse process follows the VARMA model (5.101), that runs backwards in time and corresponds to the state space model (5.97) and (5.98).

Another question that naturally arises in connection with the inverse process is the relation between the McMillan degree and the Kronecker indices of the transfer function $\Psi(z) = I + zH(I - Fz)^{-1}K$, corresponding to a process $\{Y_t\}$ that follows the innovations state space model (5.94) and (5.95), and those of the transfer function of its inverse process, $\Psi^{-1'}(z^{-1}) = I - z^{-1}K'(I - F_p'z^{-1})^{-1}H'$. The following proposition addresses this point.

Proposition 5.6 *Let $\{Y_t\}$ be a process, not necessarily stationary, following the innovations state space model in echelon form (5.94) and (5.95). Then, the McMillan degree of the transfer function $\Psi(z) = I + zH(I - Fz)^{-1}K$ coincides with that of the transfer function of the inverse process, $\{Y_t^i\}$, namely $\Psi^{-1'}(z) = I - zK'(I - F_p'z)^{-1}H'$. In addition, the Kronecker indices of $\Psi^{-1'}(z^{-1})$ and $\Psi'(z)$ are the same.*

Proof A state space model corresponding to the inverse process, $\{Y_t^i\}$, is (5.97) and (5.98). If we could find another state space model,

$$\lambda_t = \overline{F}_p'\lambda_{t+1} + \overline{H}'A_t^i$$
$$Y_t^i = -\overline{K}'\lambda_{t+1} + A_t^i,$$

of smaller dimension, then, applying the matrix inversion Lemma (4.1), we could express the transfer function $\Psi(z)$ in the form

$$\Psi(z) = I + z\overline{H}(I - \overline{F}z)^{-1}\overline{K},$$

where $\overline{F} = \overline{F}_p + \overline{KH}$, a contradiction because (5.94) and (5.95) have minimal dimension. Thus, the McMillan degrees of $\Psi(z)$ and $\Psi^{-1'}(z)$ coincide.

Letting (5.99) be the VARMA model in echelon form followed by $\{Y_t\}$, since $\Psi^{-1'}(z^{-1}) = \Phi'(z^{-1})\Theta^{-1'}(z^{-1})$, the following relation holds

$$\Theta(z)\Psi^{-1}(z) = \Phi(z).$$

Therefore, the Kronecker indices of $\Psi(z)$ and $\Psi^{-1}(z)$ coincide and so do those of $\Psi'(z)$ and $\Psi^{-1'}(z)$. □

5.21 Method of Moments Estimation of VARMA Models

Suppose that the stationary process $\{Y_t\}$ follows the identified VARMA(p, q) model

$$Y_t + \Phi_1 Y_{t-1} + \cdots + \Phi_p Y_{t-p} = A_t + \Theta_1 A_{t-1} + \cdots + \Theta_q A_{t-q}, \tag{5.102}$$

and we want to estimate the matrices Φ_i, $i = 1, \ldots, p$, and Θ_i, $i = 1, \ldots, q$, using the method of moments. Then, postmultiplying (5.102) by Y'_{t-m}, $m = q+1, \ldots, q+p$, and taking expectations, we get

$$\begin{aligned}
\Phi_1\gamma_q + \cdots + \Phi_p\gamma_{q-p+1} &= -\gamma_{q+1} \\
\Phi_1\gamma_{q+1} + \cdots + \Phi_p\gamma_{q-p+2} &= -\gamma_{q+2} \\
\cdots &= \cdots \\
\Phi_1\gamma_{q+p-1} + \cdots + \Phi_p\gamma_q &= -\gamma_{q+p},
\end{aligned}$$

that are the so-called **extended Yule–Walker equations**. From these equations, it is possible to estimate the autoregressive coefficient matrices as

$$[\Phi_1, \ldots, \Phi_p] = -[\gamma_{q+1}, \ldots, \gamma_{q+p}] \begin{bmatrix} \gamma_q & \gamma_{q+1} & \cdots & \gamma_{q+p-1} \\ \gamma_{q-1} & \gamma_q & \cdots & \gamma_{q+p-2} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{q-p+1} & \gamma_{q-p+2} & \cdots & \gamma_q \end{bmatrix}^{-1}.$$

To estimate the moving average coefficient matrices, we can first set the VARMA model into Akaike's state space form (3.13) and (3.14), where F and K are given by (3.11) and $H = [I, 0, \ldots, 0]$, and then solve the DARE (5.28), where, as mentioned in Sect. 5.6,

$$\begin{aligned}
N &= \mathrm{Cov}(x_{t+1}, Y_t) \\
&= \begin{bmatrix} \gamma(1) \\ \vdots \\ \gamma(r) \end{bmatrix},
\end{aligned}$$

and $r = \max\{p, q\}$. The coefficient matrices Θ_i, $i = 1, \ldots, q$, can be obtained from the relations

$$\Theta_j = \Phi_j + \Phi_{j-1}\Psi_1 + \cdots + \Psi_j, \quad j = 0, 1, \ldots, q,$$

where

$$K = \begin{bmatrix} \Psi_1 \\ \Psi_2 \\ \vdots \\ \Psi r \end{bmatrix}.$$

Instead of using the DARE, it is possible to use polynomial methods to solve an equivalent spectral factorization problem. In this procedure, first we obtain the autocovariances of the process $W_t = Y_t + \Phi_1 Y_{t-1} + \cdots + \Phi_p Y_{t-p}$, and then we use the method of Sect. 3.10.7 to compute the stable matrix polynomial $\Theta(z)$ and the positive definite matrix Σ such that

$$G_W(z) = \Theta(z)\Sigma\Theta'(z^{-1}),$$

where $G_W(z)$ is the covariance generating function of $\{W_t\}$.

Instead of solving the extended Yule–Walker equations to estimate the autoregressive coefficient matrices, it is possible to use the recursive procedure proposed by Akaike (1974a, p. 371). According to this procedure, we recursively compute

$$Y^i_{t+i} = Y_{t+i} + B_1(i)Y_{t+i-1} + \cdots + B_i(i)Y_t, \quad i = 1, \ldots, K,$$

where $K = \max(p, q + 1)$, such that Y^i_{t+i} is orthogonal to $\{Y_t, Y_{t-1}, Y_{t-2}, \ldots, Y_{t-i+1}\}$. When $i = K$, if we premultiply

$$Y^K_{t+K} = Y_{t+K} + B_1(K)Y_{t+K-1} + \cdots + B_K(K)Y_t$$

by Y_{t+K-m}, $m = q + 1, \ldots, q + p$, and take expectations, by the orthogonality condition, we get the extended Yule–Walker equations. Therefore,

$$[\Phi_1, \ldots, \Phi_p] = [B_1(K), \ldots, B_p(K)]$$

and $B_j(K) = 0$ if $j > p$.

The algorithm is as follows.

$i = 1)$ $Y^1_{t+1} = Y_{t+1} + B_1(1)Y_t$, where Y^1_{t+1} is orthogonal to $\{Y_t\}$.
Imposing the condition that Y^1_{t+1} is orthogonal to Y_t yields $B_1(1) = -\gamma_1\gamma_0^{-1}$.

$i = 2)$ $Y^2_{t+2} = Y_{t+2} + B_1(2)Y_{t+1} + B_2(2)Y_t$, where Y^2_{t+2} is orthogonal to $\{Y_t, Y_{t-1}\}$.
Define first $Z^1_{t+2} = Y^1_{t+2} + B_{22}Y_t$, such that Z^1_{t+2} is orthogonal to $\{Y_t\}$. This yields $B_{22} = -(\gamma_2 + B_1(1)\gamma_1)\,\gamma_0^{-1}$. Then, define $Y^2_{t+2} = Z^1_{t+2} + B_{21}Y^1_{t+1}$ such that Y^2_{t+2} is orthogonal to $\{Y_{t-1}\}$. Imposing the orthogonality condition, we get $B_{21} = -(\gamma_3 + B_1(1)\gamma_2 + B_{22}\gamma_1)\,(\gamma_2 + B_1(1)\gamma_1)^{-1}$.
Summing up,

$$\begin{aligned} Y^2_{t+2} &= Y^1_{t+2} + B_{22}Y_t + B_{21}Y^1_{t+1} \\ &= Y_{t+2} + [B_1(1) + B_{21}]\,Y_{t+1} + [B_{21}B_1(1) + B_{22}]\,Y_t. \end{aligned}$$

$i = k)$ $Y^k_{t+k} = Y_{t+k} + B_1(k)Y_{t+k-1} + \cdots + B_k(k)Y_t$, where Y^k_{t+k} is orthogonal to $\{Y_t, Y_{t-1}, \ldots, Y_{t-k+1}\}$.
Define

$$\begin{aligned} Z^{k-1}_{t+k} &= Y^{k-1}_{t+k} + B_{k2}Y^{k-2}_{t+k-2} \\ &= Y_{t+k} + B_1(k-1)Y_{t+k-1} + \cdots + B_{k-1}(k-1)Y_{t+1} \\ &\quad + B_{k2}\left[Y_{t+k-2} + B_1(k-2)Y_{t+k-3} + \cdots + B_{k-2}(k-2)Y_t\right], \end{aligned}$$

orthogonal to $\{Y_{t-k+2}\}$, and

$$\begin{aligned} Y^k_{t+k} &= Z^{k-1}_{t+k} + B_{k1}Y^{k-1}_{t+k-1} \\ &= Z^{k-1}_{t+k} + B_{k1}\left[Y_{t+k-1} + B_1(k-1)Y_{t+k-2} + \cdots + B_{k-1}(k-1)Y_t\right], \end{aligned}$$

orthogonal to $\{Y_{t-k+1}\}$. Then, imposing the orthogonality conditions yields the equations

$$\begin{aligned} &\gamma_{2k-2} + B_1(k-1)\gamma_{2k-3} + \cdots + B_{k-1}(k-1)\gamma_{k-1} \\ &\quad = -B_{k2}\left[\gamma_{2k-4} + B_1(k-2)\gamma_{2k-5} + \cdots + B_{k-2}(k-2)\gamma_{k-2}\right] \end{aligned}$$

and

$$\begin{aligned} &\gamma_{2k-1} + B_1(k-1)\gamma_{2k-2} + \cdots + B_{k-1}(k-1)\gamma_k \\ &\quad + B_{k2}\left[\gamma_{2k-3} + B_1(k-2)\gamma_{2k-4} + \cdots + B_{k-2}(k-2)\gamma_{k-1}\right] \\ &\quad = -B_{k1}\left[\gamma_{2k-2} + B_1(k-1)\gamma_{2k-3} + \cdots + B_{k-1}(k-1)\gamma_{k-1}\right], \end{aligned}$$

from which it is possible to compute first B_{k2} and then B_{k1}. Finally,

$$Y^k_{t+k} = Y^{k-1}_{t+k} + B_{k2}Y^{k-2}_{t+k-2} + B_{k1}Y^{k-1}_{t+k-1}.$$

If the stationary process $\{Y_t\}$ follows the VARMA model in echelon form

$$\Phi_0 Y_t + \Phi_1 Y_{t-1} + \cdots + \Phi_p Y_{t-r} = \Theta_0 A_t + \Theta_1 A_{t-1} + \cdots + \Theta_q A_{t-r}, \qquad (5.103)$$

and we want to estimate the matrices Φ_i, $i = 0, \ldots, r$, and Θ_i, $i = 1, \ldots, r$, using the method of moments, we can proceed as follows. First, postmultiply (5.103) by Y'_{t-m}, $m = r+1, \ldots, 2r+1$, and take expectations to get

$$\begin{aligned}
\Phi_0 \gamma_{r+1} + \Phi_1 \gamma_r + \cdots + \Phi_r \gamma_1 &= 0 \\
\Phi_0 \gamma_{r+2} + \Phi_1 \gamma_{r+1} + \cdots + \Phi_r \gamma_2 &= 0 \\
\cdots &= \cdots \\
\Phi_0 \gamma_{2r+1} + \Phi_1 \gamma_{2r} + \cdots + \Phi_r \gamma_{r+1} &= 0.
\end{aligned}$$

Then, setting $\Phi_0 = \Phi_0 - I_k + I_k$, we can rewrite the previous equations as

$$\begin{aligned}
(\Phi_0 - I_k)\, \gamma_{r+1} + \Phi_1 \gamma_r + \cdots + \Phi_r \gamma_1 &= -\gamma_{r+1} \\
(\Phi_0 - I_k)\, \gamma_{r+2} + \Phi_1 \gamma_{r+1} + \cdots + \Phi_r \gamma_2 &= -\gamma_{r+2} \\
\cdots &= \cdots \\
(\Phi_0 - I_k)\, \gamma_{2r+1} + \Phi_1 \gamma_{2r} + \cdots + \Phi_r \gamma_{r+1} &= -\gamma_{2r+1},
\end{aligned}$$

and applying the vec operator, we have

$$\begin{aligned}
(\gamma'_{r+1} \otimes I_k)\mathrm{vec}\,(\Phi_0 - I_k) + (\gamma'_r \otimes I_k)\mathrm{vec}(\Phi_1) + \cdots + (\gamma'_1 \otimes I_k)\mathrm{vec}(\Phi_r) &= -\mathrm{vec}(\gamma_{r+1}) \\
(\gamma'_{r+2} \otimes I_k)\mathrm{vec}\,(\Phi_0 - I_k) + (\gamma'_{r+1} \otimes I_k)\mathrm{vec}(\Phi_1) + \cdots + (\gamma'_2 \otimes I_k)\mathrm{vec}(\Phi_r) &= -\mathrm{vec}(\gamma_{r+2}) \\
\cdots &= \cdots \\
(\gamma'_{2r+1} \otimes I_k)\mathrm{vec}\,(\Phi_0 - I_k) + (\gamma'_{2r} \otimes I_k)\mathrm{vec}(\Phi_1) + \cdots + (\gamma'_{r+1} \otimes I_k)\mathrm{vec}(\Phi_r) &= -\mathrm{vec}(\gamma_{2r+1}).
\end{aligned}$$

Letting $\alpha = (\mathrm{vec}'\,(\Phi_0 - I_k), \ldots, \mathrm{vec}'(\Phi_r))'$ and $W_j = \left(-\gamma'_j \otimes I_k, \ldots, -\gamma'_{j-r} \otimes I_k\right)$, $j = r+1, \ldots, 2r+1$, we can rewrite the previous equations as

$$\mathrm{vec}(\gamma_j) = W_j \alpha, \qquad j = r+1, \ldots, 2r+1.$$

Finally, we can incorporate the echelon form parameter restrictions in the form

$$\alpha = R\beta,$$

where β is the vector of nonzero parameters and R is a selection matrix formed with zeros and ones. Thus, letting $X_j = W_j R$, $j = r+1, \ldots, 2r+1$, we get

$$\operatorname{vec}(\gamma_j) = X_j \beta, \qquad j = r+1, \ldots, 2r+1. \tag{5.104}$$

We can estimate β in (5.104) using OLS.

To estimate the moving average coefficient matrices, we first put the model into echelon state space form and solve the DARE (5.28). To obtain the matrix N, we can proceed as in Sect. 5.10. That is,

$$N = J_n \begin{bmatrix} \gamma_1 \\ \vdots \\ \gamma_n \end{bmatrix},$$

where $\{n_i : i = 1, \ldots, k\}$ are the Kronecker indices, n is the McMillan degree, and J_n is a selection matrix that selects the rows $\{(r-1)p + i : 1 \leq r \leq n_i; 1 \leq i \leq k\}$ of $\left[\gamma_1', \ldots, \gamma_n'\right]'$. Then, the moving average polynomial, $\Theta(B)A_t$, can be obtained using the relations (5.40). As in Sect. 5.10, note that, because $\Psi_i = HF^{i-1}K$, $i = 1, \ldots, n$, only the rows of the Ψ_i matrices contained in K are needed for the calculations in (5.40).

5.22 Historical Notes

Canonical forms for state space and VARMA models were proposed by Luenberger (1967), Denham (1974), Rissanen (1974), and Guidorzi (1981). The concept of overlapping parametrizations was first suggested by Glover & Willems (1974).

Standard references for state space and VARMA echelon forms are Kailath (1980) and Hannan & Deistler (1988).

In the statistical literature, Tsay (1989) gave a lucid account on the use of McMillan degree and Kronecker indices for identification in VARMA models. In addition, the article discusses the relationship between echelon forms and the so-called scalar component models proposed by Tiao & Tsay (1989).

Box & Tiao (1977) proposed a canonical transformation of multivariate time series so that the components of the transformed process can be ordered from least to most predictable. The least predictable components can be considered as stable contemporaneous relationships among the original variables. This type of relationship is what econometricians would later call cointegration. The concept of cointegration was introduced by Granger (1981) and Engle & Granger (1987). A standard reference is Johansen (1995).

The CKMS recursions were first presented in Kailath, Morf, & Sidhu (1973), as a rather difficult extension to discrete time of results first obtained in continuous time.

The fast covariance square root filter was originally derived by Morf & Kailath (1975).

5.23 Problems

5.1 Given the VARMA process

$$\Phi_0 Y_t + \Phi_1 Y_{t-1} = \Theta_0 A_t + \Theta_1 A_{t-1},$$

where

$$\Phi_0 = \Theta_0 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \phi_{31,0} & \phi_{32,0} & 1 \end{bmatrix}, \quad \Phi_1 = \begin{bmatrix} \phi_{11,1} & \phi_{12,1} & 0 \\ \phi_{21,1} & \phi_{22,1} & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$\Theta_1 = \begin{bmatrix} \theta_{11,1} & \theta_{12,1} & \theta_{13,1} \\ \theta_{21,1} & \theta_{22,1} & \theta_{23,1} \\ 0 & 0 & 0 \end{bmatrix},$$

and the parameters that are not zero or one vary freely, what are the Kronecker indices?

5.2 Suppose a 3-dimensional VARMA process $\{Y_t\}$ has Kronecker indices $(3, 1, 2)$. Specify the echelon form matrices corresponding to this model.

5.3 Compute the covariances of x_t and Y_t in Example 5.2.

5.4 Suppose that the process $\{Y_t\}$ follows the model $Y_t + \phi Y_{t-1} = A_t$, where $|\phi| < 1$ and $\{A_t\}$ is an uncorrelated sequence with zero mean and common variance σ^2. Put this model into Akaike state space form (3.13) and (3.14), where F and K are given by (3.11) and $H = [1, 0, \ldots, 0]$, and obtain the state space form of the inverse model, (5.97) and (5.98).

5.5 Suppose $\{Y_t\}$ follows the signal plus noise model

$$Y_t = S_t + N_t,$$

where S_t satisfies $S_t + \phi S_{t-1} = A_t$, $|\phi| < 1$, $A_t \sim$WN$(0, \sigma_A^2)$, $N_t \sim$WN$(0, \sigma_N^2)$ and the processes $\{A_t\}$ and $\{N_t\}$ are mutually uncorrelated.

(i) Show that $\{Y_t\}$ can be put into state space form (5.4) and (5.5) by defining $F = -\phi$, $G = 1$, $H = 1$, $u_t = A_{t+1}$, $v_t = N_t$, $Q = \sigma_A^2$, $R = \sigma_N^2$, $S = 0$, and $x_1 \sim (0, \Omega)$, $\Omega = \sigma_A^2/(1 - \phi^2)$.

(ii) Prove that the DARE corresponding to the previous state space model is

$$P^2 + P(\sigma_N^2 - \phi^2\sigma_N^2 - \sigma_A^2) - \sigma_A^2\sigma_N^2 = 0$$

and that this equation has real solutions of opposite sign such that the positive solution satisfies $P > \sigma_A^2$.

(iii) Obtain K and Σ as functions of the positive solution, P, of the DARE and the other parameters of the model so that the generating function, $G_Y(z)$, of $\{Y_t\}$ factorizes as (5.18). Show that $\{Y_t\}$ follows a model of the form $Y_t + \phi Y_{t-1} = U_t + \theta U_{t-1}$, where $U_t \sim$WN$(0, \Sigma)$ and find θ in terms of P and the other parameters in the model.

5.6 Consider the following state space model (5.4) and (5.5), where

$$F = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}, \quad G = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad H = \begin{bmatrix} 1 & 1 \end{bmatrix},$$

where $Q > 0$, $S = 0$ and $R > 0$.

(i) If $0 < \lambda_1 < 1$, $\lambda_1 \neq 0.5$, and $\lambda_2 = 2$, show that the processes $\{x_t : t \geq 1\}$ and $\{Y_t : t \geq 1\}$ are stationary. What is the solution of the Lyapunov equation (5.7) in this case? Is it of full rank?

(ii) Show that the pair $[F, GQ^{1/2}]$ is not controllable for all λ_1 and λ_2. For what values of these parameters, if any, is the pair $[F, GQ^{1/2}]$ stabilizable, the pair $[F, H]$ observable, or the pair $[F, H]$ detectable?

(iii) If $\lambda_1 = 1$ and $|\lambda_2| < 1$, show that the strong solution of the DARE exists, is unique, and coincides with the stabilizing solution.

5.7 Write down the likelihood function of the process of Problem 5.4 for the observations $Y_1, \ldots, Y_n$ in two ways. First, using the covariance matrix of the data and then using the Kalman filter and the prediction error decomposition. What is the σ^2-maximized log-likelihood (5.87)?

5.8 For the scalar process following the model $Y_t = A_t + \theta_1 A_{t-1} + \cdots + \theta_q A_{t-q}$, where $A_t \sim$ WN$(0, \sigma^2)$, $\sigma^2 > 0$, write down its Akaike's state space form (3.13) and (3.14), where F and K are given by (3.11) and $H = [1, 0, \ldots, 0]$. Show that the pair $[F, H]$ is always observable and the pair $[F^s, GQ^{s/2}]$ is controllable if $\theta_q \neq 0$. Show that $[F^s, GQ^{s/2}]$ is stabilizable if the roots of the polynomial $1+\theta_1 z+\cdots+\theta_q z^q$ are all outside the unit circle.

5.9 Let the VARMA model

$$Y_t - \begin{bmatrix} 2 & 2 \\ 3 & 7 \end{bmatrix} Y_{t-1} = A_t + \begin{bmatrix} -0.4 & 0 \\ 5 & -0.6 \end{bmatrix} A_{t-1}.$$

(i) Show that the autoregressive matrix polynomial has one unit root and that there is exactly one cointegration relation. Prove that $\Pi = \alpha\beta'$, where

$$\alpha = [1, 3], \quad \beta = [1, 2].$$

(ii) Obtain $\beta_\perp$ and the matrices U_1 and U_2 in (5.31).
(iii) Obtain the differencing matrix polynomial $D(z)$ and the autoregressive polynomial Φ^* in (5.34).

5.10 For the model of Example 5.2, obtain $N = F\Pi H' + GS$ and the covariance Hankel matrix of order r,

$$G_r = \begin{bmatrix} \gamma(1) & \gamma(2) & \gamma(3) & \cdots & \gamma(r) \\ \gamma(2) & \gamma(3) & \gamma(4) & \cdots & \gamma(r+1) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \gamma(r) & \gamma(r+1) & \gamma(t+2) & \cdots & \gamma(2r-1) \end{bmatrix},$$

using (5.17).

Appendix

Matrix Fraction Descriptions

Suppose the state space model

$$x_{t+1} = Fx_t + Gu_t \tag{5A.1}$$

$$Y_t = Hx_t + Ju_t, \tag{5A.2}$$

where $\text{Var}(u_t) = I$. In Sect. 3.A.3, we used generating functions to obtain the transfer function, $\Psi(z) = J + zH(I - zF)^{-1}G$, corresponding to this state space model. This transfer function is a matrix whose elements are rational functions. All transfer functions considered in this Appendix will be matrices of this type. Many of the results in this Appendix are borrowed from Chen (1984). See also Kailath (1980) and Hannan & Deistler (1988).

Given a transfer function, $\Psi(z)$, expressions of the form

$$\Psi(z) = D_L^{-1}(z)N_L(z) \quad \text{or} \quad \Psi(z) = N_R(z)D_R^{-1}(z),$$

where N_L, D_L, N_R, D_R are polynomial matrices, are called **matrix fraction descriptions (MFD)**. In general, left and right MFDs do not coincide. For example, the transfer function

$$\Psi(z) = \left(\frac{z}{1+z}, \frac{2z}{1+3z}\right)$$

has the left MFD

$$\Psi(z) = [(1+z)(1+3z)]^{-1}\,[z(1+3z), 2z(1+z)]$$

and the right MFD

$$\Psi(z) = [z, 2z]\begin{bmatrix} 1+z & 0 \\ 0 & 1+3z \end{bmatrix}^{-1}.$$

A left MFD is called **irreducible** if both the numerator and the denominator are left coprime. In a similar way, a right MFD is called irreducible if both the numerator and the denominator are right coprime. Recall that two polynomial matrices are left coprime if they have no left factors other than **unimodular** matrices, that is, matrices whose determinant is a nonzero number. A similar definition holds for right coprime polynomial matrices. The **degree** of a polynomial matrix, $D(z)$, is defined as the degree of the polynomial $\det[D(z)]$. Every entry of an MFD is a ratio of polynomials. An MFD is called **proper** if for every entry the degree of the numerator is less than or equal to the degree of the denominator, and it is called **strictly proper** if for every entry the degree of the numerator is less than the degree of the denominator. The definitions of proper and strictly proper MFDs can also be stated in terms of limits of the MFD when $z \to \infty$. Thus, an MFD $\Psi(z)$ is proper if $\Psi(\infty)$ is a finite matrix and it is strictly proper if $\Psi(\infty) = 0$.

Elementary Operations and Hermite Form

Given a polynomial matrix $M(z)$, we introduce the following **elementary operations** on $M(z)$:

1. Multiplication of a row or column by a nonzero real or complex number.
2. Interchange any two rows or columns.
3. Addition of the product of one row or column and a polynomial to another row or column.

A matrix obtained from the unit matrix by performing an elementary operation is called an **elementary matrix**. It is easy to see that an elementary row or column operation on $M(z)$ can be carried out by premultiplying or postmultiplying $M(z)$ by the corresponding elementary matrix.

In the next theorem, we show that every polynomial matrix $M(z)$ with independent columns can be transformed using exclusively elementary row operations into a triangular polynomial matrix of the form

$$H(z) = \begin{bmatrix} \times & \times & \cdots & \times \\ 0 & \times & \cdots & \times \\ 0 & 0 & \ddots & \vdots \\ \vdots & \vdots & & \times \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix},$$

where the diagonal elements are nonzero monic polynomials of higher degree than the elements above. This is called **Hermite form**.

Theorem 5A.1 *Every polynomial matrix $M(z)$ with independent columns can be transformed into Hermite form by a sequence of elementary row operations. Furthermore, there exists a unimodular matrix, $U(z)$, such that*

$$U(z)M(z) = H(z),$$

where $H(z)$ is in Hermite form.

Proof Using row interchanges, bring to the (1,1) position the element of lowest degree in the first column and call this element $\bar{m}_{11}(z)$. By the Euclidean algorithm, every other element in the first column can be written as a multiple of $\bar{m}_{11}(z)$ plus a remainder of lower degree than $\bar{m}_{11}(z)$. Thus, using elementary row operations, we can subtract from every other entry the appropriate multiple of $\bar{m}_{11}(z)$ so that only remainders of lower degree than $\bar{m}_{11}(z)$ are left. Repeat the operation with a new (1,1) element of lower degree than $\bar{m}_{11}(z)$ and continue until all the elements in the first column except the (1,1) element are zero.

Consider the second column of the resulting matrix and repeat the previous procedure with all the elements of this column except the first until all the entries below the (2,2) element are zero. If the degree of the (1,2) element is not lower than that of the (2,2) entry, use the Euclidean algorithm and the corresponding elementary row operation to replace the (1,2) entry with a polynomial of lower degree than the (2,2) entry. Continuing in this way with the remaining columns, it is clear that after a finite number of steps the Hermite form is obtained.

Every elementary row operation applied to a polynomial matrix, $\overline{M}(z)$, in the previous procedure can be carried over by premultiplying $\overline{M}(z)$ with an appropriate elementary matrix, $\overline{U}(z)$. Since elementary matrices are unimodular and the product of unimodular matrices is again a unimodular matrix, the previous algorithm yields a unimodular matrix, $U(z)$, such that

$$U(z)M(z) = H(z),$$

where $H(z)$ is in Hermite form. □

Column and Row Reduced Polynomial Matrices

Given a polynomial matrix, $M(z)$, we define the **degree of the *i*-th row** of $M(z)$, $\delta_{ri}M(z)$, as the highest degree of all polynomials in that row. The **degree of the *i*-th column** of $M(z)$, $\delta_{ci}M(z)$, is defined similarly. For example, for

$$M(z) = \begin{bmatrix} z^2+1 & z^3-2z+3 & z-1 \\ 0 & z^3-7 & z+1 \end{bmatrix}$$

we have $\delta_{r1} = 3$, $\delta_{r2} = 3$, $\delta_{c1} = 2$, $\delta_{c2} = 3$ and $\delta_{c3} = 1$.

Theorem 5A.2 *If $\Psi(z)$ is a $q \times p$ proper MFD such that $\Psi(z) = D_L^{-1}(z)N_L(z) = N_R(z)D_R^{-1}(z)$, then*

$$\delta_{ri}N_L(z) \le \delta_{ri}D_L(z), \quad i = 1, 2, \ldots, q$$

and

$$\delta_{cj}N_R(z) \le \delta_{cj}D_R(z), \quad j = 1, 2, \ldots, p.$$

If $\Psi(z)$ is a strictly proper MFD, then the previous inequalities should be replaced by strict inequalities.

Proof Since $D_L(z)\Psi(z) = N_L(z)$, letting $N_{ij}(z)$, $D_{ij}(z)$ and $\Psi_{ij}(z)$ denote the ij-th element of $N_L(z)$, $D_L(z)$ and $\Psi(z)$, we can write the elements of the i-th row of $N_L(z)$ as

$$N_{ij}(z) = \sum_{k=1}^{q} D_{ik}(z)\Psi_{kj}(z), \quad j = 1, 2, \ldots, p.$$

For each $N_{ij}(z)$, the summation in the previous formula is carried over the i-th row of $D_L(z)$. It is not difficult to show that if $\Psi(z)$ is proper, the degree of $N_{ij}(z)$, $j = 1, 2, \ldots, p$, is smaller than or equal to the highest degree in D_{ik}, $k = 1, 2, \ldots, q$. Thus,

$$\delta_{ri}N_L(z) \le \delta_{ri}D_L(z), \quad i = 1, 2, \ldots, q.$$

The rest of the theorem can be proved similarly. □

A nonsingular $p \times p$ polynomial matrix $M(z)$ is called **row reduced** if

$$\delta \det[M(z)] = \sum_{i=1}^{p} \delta_{ri} M(z),$$

where $\delta \det[M(z)]$ denotes the degree of $\det[M(z)]$. It is called **column reduced** if

$$\delta \det[M(z)] = \sum_{j=1}^{p} \delta_{cj} M(z).$$

Let $\delta_{ri} M(z) = n_i$. Then, the polynomial matrix $M(z)$ can be written in the form

$$M(z) = \text{diag}(z^{n_1}, \ldots, z^{n_p}) M_{rc} + M_{rp}(z), \tag{5A.3}$$

where M_{rc} is a constant matrix and $M_{rp}(z)$ is a polynomial matrix such that the degree of each term of its i-th row is less than n_i, $i = 1, \ldots, p$. For example, if

$$M(z) = \begin{bmatrix} 3z^2 - 2z + 1 & -5z^2 \\ 3z + 2 & -z + 1 \end{bmatrix},$$

then

$$M(z) = \begin{bmatrix} z^2 & 0 \\ 0 & z \end{bmatrix} \begin{bmatrix} 3 & -5 \\ 3 & -1 \end{bmatrix} + \begin{bmatrix} -2z + 1 & 0 \\ 2 & 1 \end{bmatrix}.$$

The constant matrix H_{rc} is called the **row-degree coefficient matrix**.

Lemma 5A.1 *The polynomial matrix $M(z)$ is row reduced if, and only if, its row-degree coefficient matrix, M_{rc}, is nonsingular.*

Proof Let $\delta_{ri} M(z) = n_i$, $n = \sum_{i=1}^{p} n_i$, and let $M_{rc} = LQ$ be the LQ decomposition of M_{rc}, where L and Q are a lower triangular and an orthogonal matrix, respectively. Then, using (5A.3), it is obtained that

$$\begin{aligned} \det[M(z)] &= \det\left[\text{diag}(z^{n_1}, \ldots, z^{n_p}) L + M_{rp}(z) Q'\right] \det(Q) \\ &= \det(M_{rc}) z^{\sum_{i=1}^{p} n_i} + \text{terms with degree less than } n. \end{aligned}$$

□

Similar to (5A.3), we can also write $M(z)$ as

$$M(z) = M_{cc} \text{diag}(z^{m_1}, \ldots, z^{m_p}) + M_{cp}(z), \tag{5A.4}$$

where $\delta_{cj}M(z) = m_j$ is the degree of the j-th column, $j = 1, 2, \ldots, p$. The constant matrix H_{cc} is called the **column-degree coefficient matrix**. The following lemma follows from (5A.4). The proof is omitted because it is similar to that of the previous lemma.

Lemma 5A.2 *The polynomial matrix $M(z)$ is column reduced if, and only if, its column-degree coefficient matrix, M_{cc}, is nonsingular.*

We can now generalize Theorem 5A.2 to the following theorem.

Theorem 5A.3 *Let $\Psi(z)$ be a $q \times p$ MFD such that $\Psi(z) = D_L^{-1}(z)N_L(z) = N_R(z)D_R^{-1}(z)$ with $D_L(z)$ row reduced and $D_R(z)$ column reduced. Then, $\Psi(z)$ is proper if, and only if,*

$$\delta_{ri}N_L(z) \leq \delta_{ri}D_L(z), \quad i = 1, 2, \ldots, q$$

and

$$\delta_{cj}N_R(z) \leq \delta_{cj}D_R(z), \quad j = 1, 2, \ldots, p.$$

$\Psi(z)$ is strictly proper if, and only if, the previous inequalities hold replaced with strict inequalities.

Proof The necessity part has been proved in Theorem 5A.2. To prove sufficiency, use (5A.3) to get

$$D_L(z) = \mathrm{diag}(z^{n_1}, \ldots, z^{n_p})D_{rc} + D_{rp}(z)$$
$$N_L(z) = \mathrm{diag}(z^{n_1}, \ldots, z^{n_p})N_{rc} + N_{rp}(z),$$

where $\delta_{ri}D_L(z) = n_i > \delta_{ri}D_{rp}(z)$ and, by assumption, $\delta_{ri}N_{rp}(z) < n_i$, $i = 1, 2, \ldots, p$. Letting $H(z) = \mathrm{diag}(z^{n_1}, \ldots, z^{n_p})$, we can write

$$\Psi(z) = \left[D_{rc} + H^{-1}(z)D_{rp}(z)\right]^{-1}\left[N_{rc} + H^{-1}(z)N_{rp}(z)\right].$$

□

The Hermite form, $H(z)$, of a polynomial matrix $M(z)$, introduced in Sect. 5.23, is column reduced. By Theorem 5A.1, there exists a unimodular matrix, $U(z)$, such that

$$U(z)M(z) = H(z).$$

Thus, using elementary row operations, it is possible to obtain a column reduced matrix from a polynomial matrix that is not column reduced. It turns out that it is possible to obtain also a row reduced matrix using elementary row operations. In fact, the following theorem holds.

Theorem 5A.4 *For every nonsingular polynomial matrix $M(z)$, there exist unimodular matrices $U(z)$ and $V(z)$ such that $U(z)M(z)$ and $M(z)V(z)$ are row reduced and column reduced, respectively.*

Proof The best way to proceed is by using an example. Let

$$M(z) = \begin{bmatrix} z + z^3 & z^2 + z + 1 \\ z & 1 \end{bmatrix}.$$

The row-degree coefficient matrix is

$$H_{rc} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix},$$

and is singular. Hence, $M(z)$ is not row reduced. We can subtract the second row of $M(z)$ multiplied by z^2 from the first row to get

$$\begin{bmatrix} 1 & -z^2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} z + z^3 & z^2 + z + 1 \\ z & 1 \end{bmatrix} = \begin{bmatrix} z & 1 + z \\ z & 1 \end{bmatrix},$$

which is row reduced. It is clear that we can use this procedure with more complicated matrices until they are row reduced. The case of polynomial matrices that are not column reduced is similar. □

Matrix Fraction Descriptions Expressed in Terms of the Forward Operator: Canonical MFDs

Using z as the backshift operator in (5A.1) and (5A.2), $zx_t = x_{t-1}$, we can write

$$(I - zF)x_{t+1} = Gu_t$$

and it follows that

$$Y_t = \left[J + zH(I - zF)^{-1}G\right] u_t.$$

Thus, the transfer function $\Psi(z) = J + zH(I - zF)^{-1}G$ is expressed in terms of the backshift operator. However, using the forward operator, z^{-1}, it is obtained that

$$\left(z^{-1}I - F\right) x_t = Gu_t, \quad Y_t = \left[J + H\left(z^{-1}I - F\right)^{-1} G\right] u_t,$$

and $\Psi(z) = J + H\left(z^{-1}I - F\right)^{-1} G$. For example, the transfer function

$$\Psi(z) = \left(\frac{1}{1+z}, \frac{2}{3+z}\right)$$

can be expressed in terms of the forward operator as

$$\Psi(z) = \left(\frac{z^{-1}}{z^{-1}+1}, \frac{2z^{-1}}{3z^{-1}+1}\right).$$

It turns out that it is sometimes more convenient to use the forward than the backshift operator when handling MFDs. To see this, consider, for example, the transfer function of the ARMA model $(1+\phi z)Y_t = (1+\theta_1 z + \theta_2 z^2)A_t$, where $zY_t = Y_{t-1}$ is the backshift operator,

$$\Psi(z) = \frac{1+\theta_1 z + \theta_2 z^2}{1+\phi z}.$$

The McMillan degree of this model is 2, the degree of the numerator. If we pass to the forward operator, we obtain

$$\Psi(z) = \frac{z^{-2}+\theta_1 z^{-1}+\theta_2}{z^{-2}+\phi z^{-1}}$$

and now the McMillan degree is the degree of the denominator. It can be shown that this property also holds for proper and irreducible MFDs of transfer functions expressed in terms of the forward operator. We will prove this for VARMA models in echelon form later in this section.

Theorem 5A.5 *A transfer function, $\Psi(z)$, expressed in terms of the forward operator is the transfer function of a state space model (5A.1) and (5A.2) if, and only if, it is a proper MFD.*

Proof Suppose $\Psi(z) = J + H(z^{-1}I - F)^{-1}G$ is the transfer function of a state space model (5A.1) and (5A.2), where $k = \dim(Y_t)$. Then, we can write

$$\Psi(z) = J + \frac{1}{\det(z^{-1}I - F)} H\left[\operatorname{adj}\left(z^{-1}I - F\right)\right] G,$$

where $\operatorname{adj}(M)$ denotes the adjoint matrix of M. If F is an $r \times r$ matrix, then $\det\left(z^{-1}I - F\right)$ has degree r. Since every entry of $\operatorname{adj}\left(z^{-1}I - F\right)$ is the determinant of an $(r-1)\times(r-1)$ submatrix of $\left(z^{-1}I - F\right)$, all the entries of $\operatorname{adj}\left(z^{-1}I - F\right)$ have degree at most $r-1$. Their linear combinations again have at most degree $r-1$. Thus, $H(z^{-1}I - F)^{-1}G$ is a strictly proper and $\Psi(z) = J + H(z^{-1}I - F)^{-1}G$ is a proper MFD.

Conversely, suppose that $\Psi(z)$, written in terms of the forward operator, is a $k \times k$ proper MFD. Since $\Psi(z)$ is proper, we can decompose it as

$$\Psi(z) = J + \Psi_{\rm sp}(z),$$

where J is a constant matrix that can be zero and $\Psi_{\rm sp}(z)$ is the strictly proper part of $\Psi(z)$. For example, if $\Psi(z) = (2z^{-1} + 1)/(z^{-1} + 3)$, then $\Psi(z) = 2 - 5/(z^{-1} + 3)$, $J = 2$, and $\Psi_{\rm sp}(z) = -5/(z^{-1} + 3)$. Let

$$d(z^{-1}) = z^{-n} + d_1 z^{-n+1} + \cdots + d_{n-1} z^{-1} + d_n$$

be the unique monic least common denominator (lcd) of all entries of $\Psi_{\rm sp}(z)$. That is, among all lcd we select the one with leading coefficient equal to one. Then, we can write

$$\Psi_{\rm sp}(z) = \frac{1}{d(z^{-1})} N(z^{-1}) = \frac{1}{d(z^{-1})} \left[N_1 z^{-n+1} + N_2 z^{-n+2} + \cdots + N_n\right],$$

where N_i, $i = 1, \ldots, n$, are constant matrices. Passing to the backward operator, it is obtained that

$$\Psi(z) = J + \frac{1}{1 + d_1 z + \cdots + d_{n-1} z^{n-1} + d_n z^n} \left[N_1 z + N_2 z^2 + \cdots + N_n z^n\right].$$

This implies that $\{Y_t\}$ follows the VARMA model

$$Y_t + d_1 Y_{t-1} + \cdots + d_n Y_{t-n} = J u_t + (N_1 + d_1 J) u_{t-1} + \cdots + (N_n + d_n J) u_{t-n}.$$

By Theorem 1.7, $\{Y_t\}$ follows a state space model (5A.1) and (5A.2). □

Theorem 5A.6 *Let $\Psi(z) = \Phi^{-1}(z)\Theta(z)$ be a $p \times q$ left MFD, where $\Phi(z) = \Phi_0 + \Phi_1 z + \cdots + \Phi_p z^p$, $\Theta(z) = \Theta_0 + \Theta_1 z + \cdots + \Theta_q z^q$, and z is the backshift operator. If $\Psi(z) = \widetilde{\Phi}^{-1}(z^{-1})\widetilde{\Theta}(z^{-1})$ is the MFD expressed in terms of the forward operator, then $\widetilde{\Phi}^{-1}(z^{-1})\widetilde{\Theta}(z^{-1})$ is proper if, and only if, Φ_0 is nonsingular. A similar statement holds for right MFDs.*

Proof Suppose Φ_0 is nonsingular and let $n_i = \max\{\delta_{ri}\Phi(z), \delta_{ri}\Theta(z)\}$, $i = 1, 2, \ldots, p$. Then,

$$\widetilde{\Phi}(z^{-1}) = \mathrm{diag}(z^{-n_1}, \ldots, z^{-n_p})\Phi_0 + \Phi_{rp}(z^{-1})$$
$$\widetilde{\Theta}(z^{-1}) = \mathrm{diag}(z^{-n_1}, \ldots, z^{-n_p})\Theta_0 + \Theta_{rp}(z^{-1}),$$

where $\delta_{ri}\Phi_{rp}(z^{-1}) < n_i$ and $\delta_{ri}\Theta_{rp}(z^{-1}) < n_i$, $i = 1, 2, \ldots, p$. By Lemma 5A.1 and Theorem 5A.3, $\widetilde{\Phi}^{-1}(z^{-1})\widetilde{\Theta}(z^{-1})$ is proper.

Conversely, suppose that $\widetilde{\Phi}^{-1}(z^{-1})\widetilde{\Theta}(z^{-1})$ is proper. If $\widetilde{\Phi}^{-1}(z^{-1})$ is not row reduced, there exists by Theorem 5A.4 a unimodular matrix $U(z)$ such that

$$\overline{\Phi}(z^{-1}) = U(z)\widetilde{\Phi}(z^{-1})$$

is row reduced. Letting $\overline{\Theta}(z^{-1}) = U(z)\widetilde{\Theta}(z^{-1})$, we can express $\Psi(z)$ as

$$\Psi(z) = \overline{\Phi}^{-1}(z^{-1})\overline{\Theta}(z^{-1}).$$

Because $\overline{\Phi}(z^{-1})$ is row reduced, we can write

$$\overline{\Phi}(z^{-1}) = \text{diag}(z^{-n_1}, \ldots, z^{-n_p})\Phi_{rc} + \Phi_{rp}(z^{-1}),$$

where the row-degree coefficient matrix, Φ_{rc}, is nonsingular. Premultiplying $\overline{\Phi}(z^{-1})$ and $\overline{\Theta}(z^{-1})$ by $\text{diag}(z^{n_1}, \ldots, z^{n_p})$, it is obtained that $\Psi(z) = \Phi^{-1}(z)\Theta(z)$, where $\Phi(z) = \text{diag}(z^{n_1}, \ldots, z^{n_p})\overline{\Phi}(z^{-1})$ and $\Theta(z) = \text{diag}(z^{n_1}, \ldots, z^{n_p})\overline{\Theta}(z^{-1})$. By Theorem 5A.3, $\Psi(z) = \Phi^{-1}(z)\Theta(z)$ is the left MFD expressed in terms of the backshift operator. Since $\Phi_0 = \Phi_{rc}$ is nonsingular, the theorem is proved. □

Corollary 5A.1 *Let $\Psi(z) = \Phi^{-1}(z)\Theta(z)$ be a $p \times q$ left MFD, where $\Phi(z) = \Phi_0 + \Phi_1 z + \cdots + \Phi_p z^p$, $\Theta(z) = \Theta_0 + \Theta_1 z + \cdots + \Theta_q z^q$, and z is the backshift operator. Then, $\Psi(z)$ is the transfer function of a state space model (5A.1) and (5A.2) if, and only if, Φ_0 is nonsingular. A similar statement holds for right MFDs.*

Proof By the previous theorem and using its notation, $\Psi(z) = \widetilde{\Phi}^{-1}(z^{-1})\widetilde{\Theta}(z^{-1})$ is proper if, and only if, Φ_0 is nonsingular. By Theorem 5A.5, $\widetilde{\Phi}^{-1}(z^{-1})\widetilde{\Theta}(z^{-1})$ is proper if, and only if, $\Psi(z)$ is the transfer function of a state space model (5A.1) and (5A.2). □

Theorem 5A.7 *Suppose a VARMA model in echelon form*

$$\Phi_0 Y_t + \Phi_1 Y_{t-1} + \cdots + \Phi_r Y_{t-r} = \Theta_0 A_t + \Theta_1 A_{t-1} + \cdots + \Theta_r A_{t-r},$$

where $\Phi_0 = \Theta_0$ is a lower triangular matrix with ones in the main diagonal. Let the transfer function be

$$\Psi(z) = \Phi^{-1}(z)\Theta(z),$$

where $\Phi(z) = \Phi_0 + \Phi_1 z + \cdots + \Phi_r z^r$ and $\Theta(z) = \Theta_0 + \Theta_1 z + \cdots + \Theta_r z^r$. If $\Psi(z) = \widetilde{\Phi}^{-1}(z^{-1})\widetilde{\Theta}(z^{-1})$ is the transfer function expressed in terms of the forward operator, then $\widetilde{\Phi}^{-1}(z^{-1})\widetilde{\Theta}(z^{-1})$ is a proper and irreducible left MFD and the degree of the denominator, $\widetilde{\Phi}(z^{-1})$, coincides with the McMillan degree.

Proof Since Φ_0 is nonsingular, by Theorem 5A.6, $\widetilde{\Phi}^{-1}(z^{-1})\widetilde{\Theta}(z^{-1})$ is proper. By Proposition 5.1, $\Phi(z)$ and $\Theta(z)$ are left coprime and $\Psi(z)$ is irreducible. To prove the statement about the McMillan degree, let $k = \dim(Y_t)$ and $D(z^{-1}) = \mathrm{diag}\,(z^{-n_1}, z^{-n_2}, \ldots, z^{-n_k})$, where $n_1, n_2, \ldots, n_k$ are the Kronecker indices. Then, $\widetilde{\Phi}(z^{-1}) = D(z^{-1})\Phi(z)$ and $\widetilde{\Theta}(z^{-1}) = D(z^{-1})\Theta(z)$. Since Φ_0 is lower triangular with ones in the main diagonal, proceeding as in the proof of Lemma 5A.1, we can prove that the degree of $\det\left[\widetilde{\Phi}(z^{-1})\right]$ is $\sum_{i=1}^{k} n_i$, the McMillan degree. Finally, if $\widetilde{\Phi}(z^{-1})$ and $\widetilde{\Theta}(z^{-1})$ were not left coprime, we could cancel a nonunimodular common factor and the determinantal degree of the resulting denominator would be less than the McMillan degree, a contradiction. □

Under the assumptions and with the notation of the previous theorem, the MFD in the forward operator $\widetilde{\Phi}^{-1}(z^{-1})\widetilde{\Theta}(z^{-1})$ is called in the engineering literature **canonical MFD**. It is shown in the proof of the theorem that the canonical MFD is obtained from the VARMA echelon form, $\Phi^{-1}(z)\Theta(z)$, as $\widetilde{\Phi}(z^{-1}) = D(z^{-1})\Phi(z)$ and $\widetilde{\Theta}(z^{-1}) = D(z^{-1})\Theta(z)$, where $D(z^{-1}) = \mathrm{diag}\,(z^{-n_1}, z^{-n_2}, \ldots, z^{-n_k})$ and $n_1, n_2, \ldots, n_k$ are the Kronecker indices. See, for example, Gevers (1986) and Hannan & Deistler (1988) on the relation between canonical MFDs and VARMA echelon forms.

Passing Right Matrix Fraction Descriptions to Left Coprime Matrix Fraction Descriptions

Suppose a $q \times p$ proper right MFD expressed in terms of the backshift operator,

$$\Psi(z) = N(z)D^{-1}(z),$$

where $D(z) = D_0 + D_1 z + \cdots + D_m z^m$ and $N(z) = N_0 + N_1 z + \cdots + N_m z^m$. By Theorem 5A.6, D_0 is nonsingular. But we are not assuming that $N(z)$ and $D(z)$ are right coprime. In this section, we look for a left MFD of $\Psi(z)$ such that

$$\Psi(z) = \Phi^{-1}(z)\Theta(z) = N(z)D^{-1}(z)$$

and $\Phi(z)$ and $\Theta(z)$ are left coprime. The previous equality can be written as $\Theta(z)D(z) = \Phi(z)N(z)$ or

$$\begin{bmatrix} -\Theta(z) & \Phi(z) \end{bmatrix} \begin{bmatrix} D(z) \\ N(z) \end{bmatrix} = 0. \tag{5A.5}$$

Letting $\Phi(z) = \Phi_0 + \Phi_1 z + \cdots + \Phi_r z^r$ and $\Theta(z) = \Theta_0 + \Theta_1 z + \cdots + \Theta_r z^r$, if we equate coefficient matrices in (5A.5), it is obtained that

$$\left[-\Theta_r\ \Phi_r \ \vdots\ -\Theta_{r-1}\ \Phi_{r-1}\ \vdots \cdots \vdots\ -\Theta_0\ \Phi_0\right]$$

$$\times \left[\begin{array}{cccccccc} D_m & D_{m-1} & \cdots & D_0 & 0 & 0 & \cdots & 0 \\ N_m & N_{m-1} & \cdots & N_0 & 0 & 0 & \cdots & 0 \\ \hline 0 & D_m & \cdots & D_1 & D_0 & 0 & \cdots & 0 \\ 0 & N_m & \cdots & N_1 & N_0 & 0 & \cdots & 0 \\ \hline \vdots & \vdots & & & & & \vdots & \vdots \\ \hline 0 & 0 & \cdots & \cdots & D_m & \cdots & D_1 & D_0 \\ 0 & 0 & \cdots & \cdots & N_m & \cdots & N_1 & N_0 \end{array}\right] = 0. \tag{5A.6}$$

The matrix in the previous equality formed from D_i and N_i is called the **generalized resultant** of $N(z)$ and $D(z)$. Since, by assumption, D_0 is nonsingular, it is clear by inspection that all D block rows in the generalized resultant are linearly independent of the previous rows.

Searching for $r = 0, 1, \ldots$, the first linearly dependent rows in the corresponding generalized resultant, we can obtain the left coprime polynomial matrices $\Phi(z)$ and $\Theta(z)$. Furthermore, in this way we can obtain the Kronecker indices and the echelon form of $\Psi(z)$.

As an application of the previous procedure, we can obtain a left MFD corresponding to the state space form (5A.1) and (5A.2). Since the transfer function is $\Psi(z) = J + zH(I - zF)^{-1}G$, we can obtain left coprime polynomial matrices $\Phi(z)$ and $\Theta(z)$ such that

$$(zH)(I - zF)^{-1} = \Phi^{-1}(z)\Theta(z).$$

Then,

$$\Psi(z) = \Phi^{-1}(z)\left[\Phi(z)J + \Theta(z)G\right].$$

Chapter 6
Time Invariant State Space Models with Inputs

Given a k-dimensional stochastic process $\{Y_t\}$, we say that it follows a state space model with **strongly exogenous** inputs, $\{Z_t\}$, if we can write

$$x_{t+1} = Fx_t + WZ_t + Gu_t \tag{6.1}$$

$$Y_t = Hx_t + VZ_t + Ju_t, \qquad t = 1, 2, \ldots, n, \tag{6.2}$$

where the $\{Z_t\}$ and the $\{u_v\}$ are orthogonal for all $v \leq t$, Z_t has dimension s, $\{u_t\}$ is a zero mean serially uncorrelated sequence of dimension q, $\mathrm{Var}(u_t) = \sigma^2 I_q$, and x_t has dimension r. This state space representation is general, and is more convenient to our purposes than other possible state space forms. The initial state vector is

$$x_1 = M\beta + A\delta + x, \tag{6.3}$$

where M and A are nonstochastic matrices, β is a constant bias vector, x is a zero mean stochastic vector with $\mathrm{Var}(x) = \sigma^2\Omega$, and δ is a stochastic vector with an undefined distribution (diffuse) that models the uncertainty with respect to the initial conditions.

The definition of strong exogeneity (Harvey, 1989, pp. 374–375) is as follows. Let $X_t = (Y_t, Z_t)$ be a two-dimensional random vector. Then, Z_t is **weakly exogenous** with respect to Y_t if the density functions satisfy

i) $f(Y_t, Z_t|X_{t-1}, \ldots, X_1; \lambda) = f(Y_t|Z_t, X_{t-1}, \ldots, X_1; \lambda_1) f(Z_t|X_{t-1}, \ldots, X_1; \lambda_2)$,

where $\lambda = (\lambda_1, \lambda_2)$ is the vector of parameters and λ_1 and λ_2 vary freely. Here, λ_1 is supposed to be the vector of parameters of interest. Thus, condition i) tells us that we can estimate λ_1 using the distribution of Y_t conditional on $(Z_t, X_{t-1}, \ldots, X_1)$ only. For forecasting and other purposes, condition i) is not enough and this leads to the definition of strong exogeneity. Z_t is **strongly exogenous** with respect to Y_t if, in addition to i), the following condition holds.

V. Gómez, *Multivariate Time Series With Linear State Space Structure*,
DOI 10.1007/978-3-319-28599-3_6

ii) $f(Z_t|X_{t-1},\ldots,X_1;\lambda_2)=f(Z_t|Z_{t-1},\ldots,Z_1;\lambda_2)$.

Condition ii) is equivalent to saying that Y_t does not **Granger-cause** Z_t (Granger, 1969). This means that there is no feedback from Y_t to Z_t and we can estimate λ_2 using the variable Z_t only.

By the strong exogeneity assumption, the variable Z_t in (6.1) and (6.2) can be treated as fixed and the parameters in (6.1) and (6.2) can be estimated independently of the parameters in the model followed by $\{Z_t\}$ if $\{Z_t\}$ is stochastic. Thus, even if $\{Z_t\}$ is stochastic and follows a well-specified model, the unknown parameters contained in the initial state vector, x_1, must be estimated using the model (6.1) and (6.2) and not the model followed by $\{Z_t\}$.

The state space model with inputs (6.1) and (6.2) is said to be in **innovations form** if

$$x_{t+1} = Fx_t + WZ_t + KA_t \tag{6.4}$$

$$Y_t = Hx_t + VZ_t + A_t, \tag{6.5}$$

where $\{A_t\}$ ~WN$(0,\Sigma)$ and x_1 is given by (6.3). The transfer function of the state space model (6.4) and (6.5) can be defined as follows. Let z be the backshift operator, $zY_t = Y_{t-1}$. Then, we get from (6.4)

$$x_{t+1} = (I - zF)^{-1}[W,K]\begin{bmatrix} Z_t \\ A_t \end{bmatrix}.$$

Substituting in (6.5), it is obtained that

$$Y_t = \left\{[V,I] + zH(I-zF)^{-1}[W,K]\right\}\begin{bmatrix} Z_t \\ A_t \end{bmatrix}.$$

Thus, the **transfer function** is defined as

$$\Delta(z) = \sum_{j=0}^{\infty}\Delta_j z^j = \left\{[V,I] + zH(I-zF)^{-1}[W,K]\right\} = \sum_{j=0}^{\infty}[\Xi_j,\Psi_j]z^j.$$

Since $WZ_t = \text{vec}(WZ_t) = (Z_t'\otimes I_r)\text{vec}(W)$ and $VZ_t = \text{vec}(VZ_t) = (Z_t'\otimes I_k)\text{vec}(V)$, if we define $\beta_v = \text{vec}(V)$, $\beta_w = \text{vec}(W)$, $\beta = (\beta_v', \beta_w')'$, $W_t = (0, Z_t'\otimes I_r)$ and $V_t = (Z_t'\otimes I_k, 0)$, we can write (6.1) and (6.2) in the form

$$\begin{aligned} x_{t+1} &= Fx_t + W_t\beta + Gu_t \\ Y_t &= Hx_t + V_t\beta + Ju_t, \qquad t = 1,2,\ldots,n. \end{aligned}$$

It follows from this that, under the strong exogeneity assumption, the model (6.1) and (6.2) can be considered as a state space model with constant bias of the form (4.85) and (4.86).

Conversely, given a state space model with constant bias, (4.85) and (4.86), since $W_t\beta = \mathrm{vec}(W_t\beta) = (\beta' \otimes I_r)\mathrm{vec}(W_t)$ and $V_t\beta = \mathrm{vec}(V_t\beta) = (\beta' \otimes I_k)\mathrm{vec}(V_t)$, if we define $Z_{v,t} = \mathrm{vec}(V_t)$, $Z_{w,t} = \mathrm{vec}(W_t)$, $Z_t = (Z'_{v,t}, Z'_{w,t})'$, $W = (0, \beta' \otimes I_r)$ and $V = (\beta' \otimes I_k, 0)$, we can write (4.82) and (4.83) in the form (6.1) and (6.2).

We have thus proved the following theorem.

Theorem 6.1 *The k-dimensional process $\{Y_t\}$ can be represented by a state space model with strongly exogenous inputs (6.1) and (6.2) if, and only if, it can be represented by a state space model with constant bias (4.85) and (4.86).*

Remark 6.1 By Theorem 6.1, the initial state vector, x_1, of (6.1) and (6.2) can be modeled using the procedure outlined in Sect. 4.14.2. As mentioned earlier, the unknown parameters contained in x_1 must be estimated using the model (6.1) and (6.2) and not the model followed by $\{Z_t\}$, assuming that such a model exists. ◇

6.1 Stationary State Space Models with Inputs

Under the strong exogeneity assumption, letting $E(x_t) = \mu_t$ and $E(Y_t) = \alpha_t$, $t = 1, 2, \ldots, n$, and taking expectations in (6.1) and (6.2), it is obtained that

$$\mu_{t+1} = F\mu_t + WZ_t \tag{6.6}$$

$$\alpha_t = H\mu_t + VZ_t, \qquad t = 1, 2, \ldots, n. \tag{6.7}$$

Subtracting (6.6) and (6.7) from (6.1) and (6.2) yields

$$x_{t+1} - \mu_{t+1} = F(x_t - \mu_t) + Gu_t$$
$$Y_t - \alpha_t = H(x_t - \mu_t) + Ju_t, \qquad t = 1, 2, \ldots, n.$$

Thus, the covariance matrices of $\{x_t\}$ and $\{Y_t\}$ are given by the formulae in Lemma 5.1 and these processes are covariance stationary if, and only if, the matrix $\Pi = \mathrm{Var}(x_1)$ satisfies the Lyapunov equation (5.7).

A sufficient condition for $\{x_t\}$ and $\{Y_t\}$ in (6.1) and (6.2) to be covariance stationary is that all the eigenvalues of F are inside the unit circle, that there is no diffuse part ($\delta = 0$) in the initial state vector (6.3), and that $\mathrm{Var}(x_1)$ satisfies the Lyapunov equation (5.7).

If the input process, $\{Z_t\}$, is stationary with mean $E(Z_t) = \mu_Z$, then it follows from (6.6) and (6.7) that $\{x_t\}$ and $\{Y_t\}$ are mean stationary if $E(x_t) = (I - F)^{-1}W\mu_Z$ and $E(Y_t) = \left[H(I - F)^{-1}W + V\right]\mu_Z$. A sufficient condition for this is that the matrix $I - F$ be invertible.

From what we have just seen, the processes $\{x_t\}$ and $\{Y_t\}$ are stationary if the input process, $\{Z_t\}$, is stationary with mean $E(Z_t) = \mu_Z$, all the eigenvalues of F are inside the unit circle, and the initial state vector, x_1, has mean $E(x_t) = (I-F)^{-1}W\mu_Z$ and covariance matrix $\Pi = \mathrm{Var}(x_1)$ that satisfies the Lyapunov equation (5.7).

It is to be noted that when the mean of the initial state vector, x_1, is not known, we can proceed as in Sect. 4.14.2.

6.2 VARMAX and Finite Linear Models with Inputs

As we will see later in this chapter, state space models with inputs of the form (6.1) and (6.2) are strongly connected with vector autoregressive moving average models with exogenous variables or **VARMAX** models as they are usually called. The vector random process $\{Y_t\}$ is said to follow a VARMAX model with input process $\{Z_t\}$ if it satisfies an equation of the form

$$\Phi(B)Y_t = \Omega(B)Z_t + \Theta(B)A_t, \tag{6.8}$$

where $\Phi(B) = I + \Phi_1 B + \cdots + \Phi_p B^p$, $\Omega(B) = \Omega_0 + \Omega_1 B + \cdots + \Omega_r B^r$, $\Theta(B) = I + \Theta_1 B + \cdots + \Theta_q B^q$, $\{A_t\}$ is a multivariate white noise process and the process $\{Z_t\}$ is strongly exogenous with respect to $\{Y_t\}$. It is assumed that Z_t and A_v are orthogonal for all $v \leq t$.

The **transfer function** of the VARMAX model (6.8) is defined as $\Delta(z) = \sum_{j=0}^{\infty} \Delta_j z^j = \Phi^{-1}(z)[\Omega(z), \Theta(z)] = \sum_{j=0}^{\infty}[\Xi_j, \Psi_j]z^j$. The concepts of **McMillan degree** and **Kronecker indices** are defined as in the VARMA case but using as Hankel matrices the previous H_t matrices with the Ψ_j coefficients replaced with Δ_j. That is,

$$H_t = \begin{bmatrix} \Delta_1 & \Delta_2 & \Delta_3 & \cdots & \Delta_t \\ \Delta_2 & \Delta_3 & \Delta_4 & \cdots & \Delta_{t+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \Delta_t & \Delta_{t+1} & \Delta_{t+2} & \cdots & \Delta_{2t-1} \end{bmatrix},$$

where $\Delta_j = [\Xi_j, \Psi_j]$.

The k-dimensional process $\{Y_t\}$ follows a **finite linear model with strongly exogenous inputs** if

$$Y_t = \sum_{j=0}^{t-1}[\Xi_j, \Psi_j]\begin{bmatrix} Z_{t-j} \\ A_{t-j} \end{bmatrix} + h_t\alpha_1, \quad t = 1, 2, \ldots, \tag{6.9}$$

where $\{A_t\}$ is a multivariate white noise process, the process $\{Z_t\}$ has dimension s and is strongly exogenous with respect to $\{Y_t\}$, the Ψ_j matrices have dimension $k \times l$,

h_t is a deterministic $k \times p$ matrix, and α_1 is a p-dimensional stochastic vector that defines the initial conditions. The vector α_1 is usually decomposed as

$$\alpha_1 = M\beta + A\delta + x,$$

where M and A are nonstochastic matrices, β is a constant bias vector, x is a zero mean stochastic vector with a known distribution, and δ is a stochastic vector with a diffuse distribution that models the uncertainty with respect to the initial conditions.

The **transfer function** of the model (6.9) is defined as $\Delta(z) = \sum_{j=0}^{\infty} \Delta_j z^j = \sum_{j=0}^{\infty} [\Xi_j, \ \Psi_j] z^j$. The concepts of **McMillan degree** and **Kronecker indices** can be extended to finite linear models with exogenous inputs defining the **augmented Hankel matrix** of order t, H_t^a, by

$$H_t^a = \begin{bmatrix} K_1 & K_2 & K_3 & \cdots & K_t \\ K_2 & K_3 & K_4 & \cdots & K_{t+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ K_t & K_{t+1} & K_{t+2} & \cdots & K_{2t-1} \end{bmatrix}, \tag{6.10}$$

where $K_t = [\Xi_j, \Psi_t, h_t] = [\Delta_j, h_t]$. We will make the following assumption.

Assumption 6.1 *The augmented Hankel matrices, H_t^a, of the linear process (6.9) have finite rank for all $t > r$, where r is a fixed positive integer.*

Remark 6.2 Theorems 3.8 and 3.9 can be easily extended to processes with exogenous inputs because the proofs of those theorems are algebraic in nature. Therefore, there is an equivalence between finite linear models with inputs, innovations state space models with inputs and VARMAX models. However, in the previous theorems the dimensions of all these models were not necessarily minimal. ◇

The state space models and the VARMAX models that we will consider in this chapter can be nonstationary. In this case, we will make the following assumption.

Assumption 6.2 *If a process $\{Y_t\}$ is nonstationary and its transfer function can be represented either by a state space model (6.1) and (6.2) or a VARMAX model (6.8), we will assume that the initial conditions start at some finite point in the past and that they are selected in such a way that the same process is generated in both cases.*

The following theorem establishes the equivalence, under Assumption 6.2, between state space models (6.1) and (6.2) and left coprime VARMAX models. The result is important in its own right and will be often used in the rest of the book.

Theorem 6.2 *Under Assumption 6.2, a process $\{Y_t\}$ can be represented by an identified, observable, and controllable state space model with exogenous inputs (6.1) and (6.2) if, and only if, it can be represented by an identified left coprime VARMAX model of the form*

$$\Phi(B)Y_t = \Gamma(B)Z_t + \Theta(B)A_t, \tag{6.11}$$

where B is the backshift operator, $BY_t = Y_{t-1}$, $\Phi(z) = I + \Phi_1 z + \cdots + \Phi_b z^b$, $\Gamma(z) = \Gamma_0 + \Gamma_1 z + \cdots + \Gamma_c z^c$, $\Theta(z) = \Theta_0 + \Theta_1 z + \cdots + \Theta_d z^d$, *the McMillan degree is equal to* $r = \dim(x_t)$, $Var(A_t) = \Sigma$ *may be singular and satisfies* $\Sigma = \Theta_0\Sigma = \Sigma\Theta_0'$, *and the* Θ_i, $i = 0, \ldots, d$, *may be nonunique. In addition, if* Σ *is nonsingular, then* $\Theta_0 = I$ *and the* Θ_i, $i = 1, \ldots, d$, *are unique.*

Proof Assume first that $\{Y_t\}$ follows (6.1) and (6.2) under Assumption 6.2. We will first prove that there exists an innovations representation. We will do this without using standard results about the discrete algebraic Riccati equation (DARE) because all of them suppose that the covariance matrix of the innovations is positive definite and we do not make this assumption. Stacking the observations in the usual way (Harvey 1989, pp. 319–320; Van Dooren 2003, p. 81), it is obtained that

$$Y_{t:t+r} = H_{r+1}x_t + V_{r+1}Z_{t:t+r} + J_{r+1}U_{t:t+r}, \tag{6.12}$$

where $Y_{t:t+r} = (Y_t', Y_{t+1}', \ldots, Y_{t+r}')'$, $U_{t:t+r} = (u_t', u_{t+1}', \ldots, u_{t+r}')'$, $Z_{t:t+r} = (Z_t', Z_{t+1}', \ldots, Z_{t+r}')'$, $H_{r+1} = [H', (HF)', \ldots, (HF^r)']'$,

$$J_{r+1} = \begin{bmatrix} J & & & \\ HG & J & & \\ \vdots & \vdots & \ddots & J \\ HF^{r-1}G & \cdots & \cdots HG & J \end{bmatrix} \tag{6.13}$$

and

$$V_{r+1} = \begin{bmatrix} V & & & \\ HW & V & & \\ \vdots & \vdots & \ddots & V \\ HF^{r-1}W & \cdots & \cdots HW & V \end{bmatrix}.$$

Since the model (6.1) and (6.2) is observable, the submatrix of H_{r+1} formed with its first r blocks (the observability matrix) has rank r and the rows of HF^r in H_{r+1} depend linearly on the previous rows. Thus, there exists a block matrix, $\Phi = [\Phi_r, \Phi_{r-1}, \ldots, \Phi_1, I]$, such that $\Phi H_{r+1} = 0$. Premultiplying (6.12) by Φ, it is obtained that

$$\Phi_r Y_t + \cdots + \Phi_1 Y_{t+r-1} + Y_{t+r} = \Gamma_r Z_t + \cdots + \Gamma_0 Z_{t+r} + \Lambda_r u_t + \cdots + \Lambda_0 u_{t+r}, \tag{6.14}$$

where the matrices Γ_i and Λ_i are given by the products ΦV_{r+1} and ΦJ_{r+1}. The sum $\Lambda_r u_t + \cdots + \Lambda_0 u_{t+r}$ in (6.14) is a moving average process that has a Wold decomposition of the form $\Theta(B)A_t$ with the properties stated in the theorem (Caines, 1988, pp. 28–29), and thus $\{Y_t\}$ follows a VARMAX model of the form (6.11). Letting $V_t = (Z_t', A_t')'$ and expressing the right-hand side of (6.11) in terms of V_t, we obtain $Y_t = \Phi^{-1}(B)[\Gamma(B), \Theta(B)]V_t$. This in turn implies that the following innovations form

$$x_{t+1} = Fx_t + [W, K]V_t \tag{6.15}$$

$$Y_t = Hx_t + [V, \Theta_0]V_t \tag{6.16}$$

corresponding to (6.1) and (6.2) exists, where K is given as the solution of $[H', (HF)', \ldots, (HF^{r-1})']'K = [\Psi_1', \Psi_2', \ldots, \Psi_r']'$, $\sum_{i=0}^{\infty} \Psi_i z^i = \Phi^{-1}(z)\Theta(z)$. Let $l = \max\{n_i : i = 1, \ldots, k\}$, where the n_i are the Kronecker indices corresponding to (6.15) and (6.16), and stack the observations to get

$$Y_{t:t+l} = H_{l+1}x_t + K_{l+1}V_{t:t+l}, \tag{6.17}$$

where $Y_{t:t+l} = (Y_t', Y_{t+1}', \ldots, Y_{t+l}')'$, $V_{t:t+l} = (V_t', V_{t+1}', \ldots, V_{t+l}')'$, $H_{l+1} = [H', (HF)', \ldots, (HF^l)']'$ and K_{l+1} is like J_{r+1} in (6.13) but with r, J and G replaced with l, $[V, \Theta_0]$ and $[W, K]$. The definition of Kronecker indices implies that there exists a unique block matrix $\Phi = [\Phi_l, \Phi_{l-1}, \ldots, \Phi_1, I]$ such that $\Phi H_{l+1} = 0$. The rows of Φ simply express each of the rows of the $(l + 1)$-th block of rows of H_{l+1} as a linear combination of the basis of rows given by the Kronecker indices. Premultiplying (6.17) by Φ, it is obtained that

$$\Phi_l Y_t + \cdots + \Phi_1 Y_{t+l-1} + Y_{t+l} = \Delta_l V_t + \cdots + \Delta_0 V_{t+l}, \tag{6.18}$$

where the matrices Δ_i, $i = 0, \ldots, l$, are given by the product ΦK_{l+1}. The VARMAX model (6.18) has McMillan degree equal to $r = \sum_{i=1}^{k} n_i$ and is left coprime because otherwise we can cancel some nonunimodular left factor and this implies that we can get a simplified block matrix $\hat{\Phi}$ such that $\hat{\Phi}H_{l+1} = 0$, in contradiction to the fact that Φ is unique.

Conversely, if $\{Y_t\}$ follows the identified and left coprime VARMAX model (6.11) with McMillan degree r, letting again $V_t = (Z_t', A_t')'$, we can put this model into state space form (6.15) and (6.16) with $V = \Gamma_0$ and appropriate matrices W and K using, for example, Akaike's (1974) representation for multivariate series. This representation is observable but may be not controllable. If it is not controllable, we can apply the algorithm of Lemma 5.7 to obtain an observable and controllable state space form with $\dim(x_t) = r$.

In the nonstationary case, Assumption 6.2 guarantees that the different processes considered in the proof are well defined. □

Note that the theorem is algebraic in nature and that, therefore, it is also valid for nonstationary processes. In most of the cases found in practice the matrix Σ of Theorem 6.2 is nonsingular and, therefore, the matrix polynomial $\Theta(z)$ is unique. A sufficient condition for this, under the conditions of the theorem, is that JJ' be nonsingular, where J is that of (6.2). Note also that when Σ is singular, the model is degenerate, in the sense that there exist linear combinations of Y_t that can be perfectly predicted.

In the rest of the chapter, unless otherwise specified, we will assume that Σ is nonsingular.

6.3 Kalman Filter and Likelihood Evaluation for the State Space Model with Inputs

If the process $\{Y_t\}$ follows the state space model (6.1) and (6.2) with strongly exogenous inputs, $\{Z_t\}$, and there is no diffuse part ($\delta = 0$) in the initial state vector (6.3), the following Kalman filter can be used

$$E_t = Y_t - VZ_t - H\hat{x}_{t|t-1}, \qquad \Sigma_t = HP_tH' + JJ', \tag{6.19}$$

$$K_t = (FP_tH' + GJ')\Sigma_t^{-1}, \qquad \hat{x}_{t+1|t} = WZ_t + F\hat{x}_{t|t-1} + K_tE_t, \tag{6.20}$$

$$P_{t+1} = (F - K_tH)P_tF' + (G - K_tJ)G', \tag{6.21}$$

initialized with $\hat{x}_{1|0} = M\beta$ and $P_1 = \Omega$, where M is a nonstochastic matrix, β is a constant bias vector, and $\hat{x}_{t|t-1} = E^*(x_t|Y_{t-1},\ldots,Y_1)$. It is assumed that both M and β are known and it is to be noticed that the stochastic vectors x_t and Y_t have nonzero means given by the recursions (6.6) and (6.7). The Kalman filter recursions (6.19)–(6.21) can be proved using the properties of orthogonal projection with nonzero means. See Problem 6.1.

If there is a diffuse part ($\delta \neq 0$) in x_1 or β is not known, a two-stage Kalman filter corresponding to (6.19)–(6.21) should be used. More specifically, the modified bias-free filter is given by

$$(E_t, e_t) = (0, 0, Y_t - VZ_t) - H(-U_t, x_{t|t-1}) \tag{6.22}$$

$$\Sigma_t = HP_tH' + JJ', \quad K_t = (FP_tH' + GJ')\Sigma_t^{-1} \tag{6.23}$$

$$(-U_{t+1}, x_{t+1|t}) = (0, 0, WZ_t) + F(-U_t, x_{t|t-1}) + K_t(E_t, e_t) \tag{6.24}$$

$$P_{t+1} = (F - K_tH)P_tF' + (G - K_tJ)G', \tag{6.25}$$

with initial conditions $(-U_1, x_{1|0}) = (-A, -M, 0)$ and $P_1 = \Omega$, and the recursions of the information form bias filter are

$$\left(\Pi_{t+1}^{-1}, \Pi_{t+1}^{-1}\hat{\gamma}_{t+1}\right) = \left(\Pi_t^{-1}, \Pi_t^{-1}\hat{\gamma}_t\right) + E_t'\Sigma_t^{-1}(E_t, e_t), \tag{6.26}$$

initialized with $\left(\Pi_1^{-1}, \Pi_1^{-1}\hat{\gamma}_1\right) = (0, 0)$, where $\gamma = [\delta', \beta']'$.

In the case of an innovations model (6.4) and (6.5) such that $x = 0$ in x_1, that is $x_1 = M\beta + A\delta$, the previous modified bias-free filter can be simplified to (6.22) and (6.24), where $G = K\Sigma^{1/2}$, $J = \Sigma^{1/2}$, $K_t = K$, and $\mathrm{Var}(A_t) = \Sigma = \Sigma^{1/2}\Sigma^{1/2'}$. This can be proved by checking that $P_t = 0$ for $t = 1, 2, \ldots, n$. See Problem 6.2.

As mentioned earlier, instead of using the previous Kalman filters, one can first transform the state space model (6.1) and (6.2) into a state space model with constant bias (4.85) and (4.86) and then use the two-stage Kalman filter described for this model in Sect. 4.15. This last filter has the advantage of allowing for more parameters to be concentrated out of the likelihood than with the previously mentioned filters. But of course, any of these filters can be used for likelihood evaluation.

For example, when there is no diffuse part ($\delta = 0$) and the initial conditions are $\hat{x}_{1|0} = M\beta$ and $P_1 = \Omega$, where M and β are known, the log-likelihood is

$$l(Y) = \text{constant} - \frac{1}{2}\left\{\frac{1}{\sigma^2}\sum_{t=1}^{n} E_t'\Sigma_t^{-1}E_t + \sum_{t=1}^{n}\ln|\sigma^2\Sigma_t|\right\},$$

where E_t and Σ_t are given by the Kalman filter (6.19)–(6.21).

As another example, one way to evaluate the log-likelihood in the general case, in which x_1 is given by (6.3) with β unknown and $\delta \neq 0$ diffuse, is to use the previous modified bias-free filter (6.22)–(6.25) and information bias filter (6.26) as follows. Suppose for simplicity and without loss of generality that $\sigma^2 = 1$. Then, it is not difficult to show (see Problem 6.3) that the log-likelihood is, apart from a constant,

$$l(Y) = -\frac{1}{2}\left\{\left[RSS_{n+1}^{BFF} - \left(\Pi_{n+1}^{-1}\hat{\gamma}_{n+1}\right)'\Pi_{n+1}\left(\Pi_{n+1}^{-1}\hat{\gamma}_{n+1}\right)\right] + \sum_{t=1}^{n}\ln|\Sigma_t| + \ln|\Pi_{n+1}^{11}|\right\},$$

where RSS_{n+1}^{BFF} can be obtained with the recursion (4.138),

$$\Pi_{n+1}^{-1} = \sum_{t=1}^{n} E_t'\Sigma_t^{-1}E_t = \begin{bmatrix}\Pi_{n+1}^{11} & \Pi_{n+1}^{12}\\ \Pi_{n+1}^{21} & \Pi_{n+1}^{22}\end{bmatrix},$$

and the partition is conformal with $\gamma = [\delta', \beta']'$.

An alternative to the previous procedures is to use a decoupled VARMAX model. To illustrate the method, suppose that $\{Y_t\}$ follows the left coprime VARMAX model

$$\Phi(B)Y_t = \Gamma(B)Z_t + \Theta(B)A_t,$$

where $BY_t = Y_{t-1}$, $\Phi(z) = I + \Phi_1 z + \cdots + \Phi_b z^b$, $\Gamma(z) = \Gamma_0 + \Gamma_1 z + \cdots + \Gamma_c z^c$ and $\Theta(z) = I + \Theta_1 z + \cdots + \Theta_d z^d$, whose existence is guaranteed by Theorem 6.2. Then, premultiplying by $\Phi^{-1}(B)$, it is obtained that

$$\begin{aligned} Y_t &= \Phi^{-1}(B)\Gamma(B)Z_t + \Phi^{-1}(B)\Theta(B)A_t \\ &= (\Xi_0 + \Xi_1 B + +\Xi_2 B^2\cdots)Z_t + (I + \Psi_1 B + \Psi_2 B^2 + \cdots)A_t, \end{aligned}$$

where the coefficient matrices, $[\Xi_j, \Psi_j]$ satisfy the relations

$$\begin{aligned} (I + \Phi_1 z + \cdots + \Phi_b z^b)(\Xi_0 + \Xi_1 z + \cdots) &= \Gamma_0 + \Gamma_1 z + \cdots + \Gamma_c z^c \\ (I + \Phi_1 z + \cdots + \Phi_b z^b)(I + \Psi_1 z + \cdots) &= I + \Theta_1 z + \cdots + \Theta_d z^d. \end{aligned}$$

The previous relations imply that if we define the processes $\{W_t\}$ and $\{V_t\}$ by $W_t = \sum_{j=0}^{t-1} \Xi_j Z_{t-j}$ and $V_t = \sum_{j=0}^{t-1} \Psi_j A_{t-j}$, then

$$Y_t = W_t + h_{W,t}\alpha_{W,1} + V_t + h_{V,t}\alpha_{V,1}, \tag{6.27}$$

where $\Phi(B)h_{V,t} = 0$ and $\Phi(B)h_{W,t} = 0$ for $t > \max\{b, c, d\}$, and the initial conditions are $\alpha_{W,1}$ and $\alpha_{V,1}$. Letting $\tilde{V}_t = V_t + h_{V,t}\alpha_{V,1}$, we can write

$$Y_t - W_t = h_{W,t}\alpha_{W,1} + \tilde{V}_t, \tag{6.28}$$

and, since $\tilde{V}_t$ follows the VARMA model $\Phi(B)\tilde{V}_t = \Theta(B)A_t$, we can perform likelihood evaluation by applying the two-stage Kalman filter to the VARMA model with constant bias (6.28) in which the data are $\{Y_t - W_t\}$ and the bias is $\alpha_{W,1}$.

To implement this procedure, the weights Ξ_i can be obtained from the relation

$$(I + \Phi_1 z + \cdots + \Phi_b z^b)(\Xi_0 + \Xi_1 z + \cdots) = \Gamma_0 + \Gamma_1 z + \cdots + \Gamma_c z^c,$$

and the variables $h_{W,t}$ such that $\Phi(B)h_{W,t} = 0$ for $t > \max\{b, c, d\}$ can be computed using the procedure of Sect. 3.A.1.

It is to be noticed that the vectors $\alpha_{W,1}$ and $\alpha_{V,1}$ in (6.27) will not have in general minimal dimension. This is because the two pairs of polynomial matrices $[\Phi(z), \Theta(z)]$ and $[\Phi(z), \Gamma(z)]$ may have common factors. In this case, they can be replaced with two pairs of left coprime polynomial matrices, $[\Phi_V(z), \Theta_V(z)]$ and $[\Phi_W(z), \Gamma_W(z)]$, such that

$$\Phi_W(B)W_t = \Gamma_W(B)Z_t, \qquad \Phi_V(B)V_t = \Theta_V(B)A_t.$$

In terms of the state space model (6.1) and (6.2), the previous procedure can be described as follows. Let z^{-1} be the forward operator, $z^{-1}Y_t = Y_{t+1}$. Then, rewriting (6.1) as

$$x_t = \left(z^{-1}I - F\right)^{-1}(WZ_t + KA_t),$$

and substituting in (6.2), it is obtained that

$$\begin{aligned} Y_t &= H\left(z^{-1}I - F\right)^{-1}(WZ_t + KA_t) + DZ_t + A_t \\ &= \left[DZ_t + H\left(z^{-1}I - F\right)^{-1} WZ_t\right] + \left[A_t + H\left(z^{-1}I - F\right)^{-1} KA_t\right] \\ &= \tilde{W}_t + \tilde{V}_t. \end{aligned}$$

It is not difficult to verify that $\{\tilde{W}_t\}$ and $\{\tilde{V}_t\}$ can be realized with the state space models

$$\begin{aligned} \alpha_{t+1} &= F\alpha_t + WZ_t \\ \tilde{W}_t &= H\alpha_t + DZ_t \end{aligned}$$

and

$$\beta_{t+1} = F\beta_t + KA_t \tag{6.29}$$

$$\tilde{V}_t = H\beta_t + A_t. \tag{6.30}$$

Defining $W_t = \left[DZ_t + H\left(z^{-1}I - F\right)_t^{-1} WZ_t\right]$ and $h_{W,t} = HF^{t-1}$, where

$$\left(z^{-1}I - F\right)_t^{-1} = \sum_{j=1}^{t-1} z^j F^{j-1},$$

we can write $\tilde{W}_t = W_t + h_{W,t}\alpha_1$ and the following innovations state space model with constant bias is obtained

$$Y_t - W_t = h_{W,t}\alpha_1 + \tilde{V}_t,$$

in which the data are $\{Y_t - W_t\}$, the bias is α_1 and $\{\tilde{V}_t\}$ follows the model (6.29) and (6.30).

As it happened with VARMAX models, the state space models corresponding to $\{\tilde{W}_t\}$ and $\{\tilde{V}_t\}$ may not have minimal dimension. In this case, we can replace these models with state space echelon forms, as described in Sect. 6.5.5.

6.4 The Case of Stochastic Inputs

When working with VARMAX models (6.8) sometimes the input process, $\{Z_t\}$, is stochastic and follows a VARMA model

$$\Phi_Z(B)Z_t = \Theta_Z(B)U_t, \tag{6.31}$$

where $\Phi_Z(B) = I + \Phi_{Z1}B + \cdots + \Phi_{Za}B^a$, $\Theta_Z(B) = I + \Theta_{Z1}B + \cdots + \Theta_{Zb}B^b$ and $\{U_t\}$ is a multivariate white noise process uncorrelated with $\{A_t\}$. In this case, the joint process $\{(Y_t', Z_t')'\}$ follows the model

$$\begin{bmatrix} I & -\Gamma_0 \\ & I \end{bmatrix}\begin{bmatrix} Y_t \\ Z_t \end{bmatrix} + \begin{bmatrix} \Phi_1 & -\Gamma_1 \\ & \Phi_{Z1} \end{bmatrix}\begin{bmatrix} Y_{t-1} \\ Z_{t-1} \end{bmatrix} + \cdots + \begin{bmatrix} \Phi_m & -\Gamma_m \\ & \Phi_{Zm} \end{bmatrix}\begin{bmatrix} Y_{t-m} \\ Z_{t-m} \end{bmatrix} =$$

$$\begin{bmatrix} I & \\ & I \end{bmatrix}\begin{bmatrix} A_t \\ U_t \end{bmatrix} + \begin{bmatrix} \Theta_1 & \\ & \Theta_{Z1} \end{bmatrix}\begin{bmatrix} A_{t-1} \\ U_{t-1} \end{bmatrix} + \cdots + \begin{bmatrix} \Theta_m & \\ & \Theta_{Zm} \end{bmatrix}\begin{bmatrix} A_{t-m} \\ U_{t-m} \end{bmatrix},$$

where $m = \max\{p, q, r, a, b\}$ and the matrices Φ_i, Φ_{Zi}, Γ_i, Θ_i, Θ_{Zi} are zero for $i > p$, $i > a$, $i > r$, $i > q$ and $i > b$, respectively. If $\Gamma_0 \neq 0$, then, premultiplying the previous expression by

$$\begin{bmatrix} I & -\Gamma_0 \\ & I \end{bmatrix}^{-1},$$

a VARMA process is obtained in which the covariance matrix of the innovations is not block diagonal. On the other hand, if $\Gamma_0 = 0$, the process $\{(Y_t', Z_t')'\}$ follows a VARMA model such that the covariance matrix of the innovations is block diagonal. We have thus proved part of the following theorem.

Theorem 6.3 *The process $(Y_t', Z_t')'$ follows a VARMAX model (6.8) in which the input process $\{Z_t\}$ follows a VARMA model (6.31) if, and only if, $(Y_t', Z_t')'$ follows a VARMA model in which all the matrices in the autoregressive and moving average matrix polynomials are block upper triangular and both first matrices in these matrix polynomials are the unit matrix. In this case, $\Gamma_0 \neq 0$ if, and only if, the covariance matrix of the innovations corresponding to the VARMA model followed by $(Y_t', Z_t')'$ is not block diagonal.*

Proof We have already proved one implication. To prove the other one, suppose that $\{(Y_t', Z_t')'\}$ follows the VARMA model

$$\begin{bmatrix} \Phi(B) & -\Gamma(B) \\ & \Phi_Z(B) \end{bmatrix} \begin{bmatrix} Y_t \\ Z_t \end{bmatrix} = \begin{bmatrix} \Theta(B) & \Theta_1(B) \\ & \Theta_Z(B) \end{bmatrix} \begin{bmatrix} C_t \\ D_t \end{bmatrix}. \tag{6.32}$$

Then, it follows from the second equation that

$$D_t = \Theta_Z^{-1}(B)\Phi_Z(B)Z_t$$

and substituting this expression into the first equation, it is obtained that

$$\Phi(B)Y_t = \left[\Gamma(B) + \Theta_1(B)\Theta_Z^{-1}(B)\Phi_Z(B)\right]Z_t + \Theta(B)C_t. \tag{6.33}$$

The expression $\Theta_1(z)\Theta_Z^{-1}(z)$ is a right matrix fraction description (MFD) that can be transformed into a left MFD (Chen, 1984; Kailath, 1980), $\Theta_1(z)\Theta_Z^{-1}(z) = \tilde{\Theta}_Z^{-1}(z)\tilde{\Theta}_1(z)$. By changing the matrix polynomials if necessary, we may assume without loss of generality that the first matrix in $\tilde{\Theta}_Z$ is the identity matrix. Then, (6.33) can be rewritten as

$$\Phi(B)Y_t = \left[\Gamma(B) + \tilde{\Theta}_Z^{-1}(B)\tilde{\Theta}_1(B)\Phi_Z(B)\right]Z_t + \Theta(B)C_t,$$

and, premultiplying by $\tilde{\Theta}_Z(B)$, it is obtained that

$$\begin{aligned} \tilde{\Theta}_Z(B)\Phi(B)Y_t &= \left[\tilde{\Theta}_Z(B)\Gamma(B) + \tilde{\Theta}_1(B)\Phi_Z(B)\right]Z_t + \tilde{\Theta}_Z(B)\Theta(B)C_t \\ &= \tilde{\Gamma}(B)Z_t + \tilde{\Theta}_Z(B)\Theta(B)C_t, \end{aligned}$$

where $\tilde{\Gamma}(z) = \tilde{\Theta}_Z(z)\Gamma(z) + \tilde{\Theta}_1(z)\Phi_Z(z)$. If C_t and D_t are uncorrelated, we are under the same assumptions of the first part of the proof and we are done. If not, let

$$\text{Var}\left(\begin{bmatrix} C_t \\ D_t \end{bmatrix}\right) = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

and set

$$\begin{bmatrix} \tilde{C}_t \\ D_t \end{bmatrix} = \begin{bmatrix} I & -\Sigma_{12}\Sigma_{22}^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} C_t \\ D_t \end{bmatrix}.$$

Then, $(\tilde{C}_t', D_t')'$ has a block diagonal covariance matrix and

$$\begin{bmatrix} \tilde{\Theta}_Z(B)\Phi(B) & -\tilde{\Gamma}(B) \\ & \Phi_Z(B) \end{bmatrix} \begin{bmatrix} Y_t \\ Z_t \end{bmatrix} = \begin{bmatrix} \tilde{\Theta}_Z(B)\Theta(B) & \\ & \Theta_Z(B) \end{bmatrix} \begin{bmatrix} I & \Sigma_{12}\Sigma_{22}^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} \tilde{C}_t \\ D_t \end{bmatrix}$$

is a VARMA with uncorrelated innovations such that the autoregressive and moving average matrix polynomials are block upper triangular and both first matrices in these matrix polynomials are the unit matrix. Thus, we have reduced this case to the previous one and this completes the proof. □

Theorem 6.4 *The process $(Y_t', Z_t')'$ follows a VARMAX model (6.8) in which the input process $\{Z_t\}$ follows a VARMA model (6.31) if, and only if, $(Y_t', Z_t')'$ follows an innovations state space model*

$$x_{t+1}^* = F^* x_t^* + K^* A_t^* \tag{6.34}$$

$$Y_t^* = H^* x_t^* + A_t^*, \qquad t = 1, 2, \ldots, m. \tag{6.35}$$

in which all the matrices F^*, K^*, and H^* are block upper triangular. In this case, $\Gamma_0 \neq 0$ if, and only if, the covariance matrix of the innovations, $\{A_t^*\}$, is not block diagonal.

Proof Suppose $(Y_t', Z_t')'$ follows a VARMAX model (6.8) in which the input process $\{Z_t\}$ follows a VARMA model (6.31). Then, proceeding as in Sect. 6.3, we get the representation

$$\begin{aligned} Y_t &= \Phi^{-1}(B)\Gamma(B)Z_t + \Phi^{-1}(B)\Theta(B)A_t \\ &= (\Xi_0 + \Xi_1 B + +\Xi_2 B^2 \cdots)Z_t + (I + \Psi_1 B + \Psi_2 B^2 + \cdots)A_t. \end{aligned}$$

Based on Akaike (1974) or the material later in this chapter regarding canonical forms, we can find innovations representations for $V_t = \Phi^{-1}(B)\Gamma(B)Z_t$ and $W_t = \Phi^{-1}(B)\Theta(B)A_t$. Let these representations be

$$\begin{aligned} \alpha_{t+1} &= F_Z \alpha_t + K_Z Z_t \\ V_t &= H_Z \alpha_t + \Xi_0 Z_t, \qquad t = 1, 2, \ldots, m. \end{aligned}$$

and

$$\begin{aligned}\beta_{t+1} &= F_A\beta_t + K_A A_t\\ W_t &= H_A\beta_t + A_t, \qquad t = 1, 2, \ldots, m.\end{aligned}$$

Letting $x_t = (\alpha_t', \beta_t')'$, we can write

$$\begin{aligned}x_{t+1} &= \begin{bmatrix} F_Z & \\ & F_A \end{bmatrix} x_t + \begin{bmatrix} K_Z \\ 0 \end{bmatrix} Z_t + \begin{bmatrix} 0 \\ K_A \end{bmatrix} A_t\\ Y_t &= [H_Z, H_A]x_t + \Xi_0 Z_t + A_t.\end{aligned}$$

Also, $\{Z_t\}$ has an innovations representation

$$\begin{aligned}\gamma_{t+1} &= F_U\gamma_t + K_U U_t\\ Z_t &= H_U\gamma_t + U_t, \qquad t = 1, 2, \ldots, m.\end{aligned}$$

Putting all together, we can write

$$\begin{aligned}\begin{bmatrix} x_{t+1} \\ \gamma_{t+1} \end{bmatrix} &= \begin{bmatrix} F & WH_U \\ & F_U \end{bmatrix} \begin{bmatrix} x_t \\ \gamma_t \end{bmatrix} + \begin{bmatrix} K & W \\ & K_U \end{bmatrix} \begin{bmatrix} A_t \\ U_t \end{bmatrix}\\ \begin{bmatrix} Y_t \\ Z_t \end{bmatrix} &= \begin{bmatrix} H & VH_U \\ & H_U \end{bmatrix} \begin{bmatrix} x_t \\ \gamma_t \end{bmatrix} + \begin{bmatrix} I & V \\ & I \end{bmatrix} \begin{bmatrix} A_t \\ U_t \end{bmatrix},\end{aligned}$$

where

$$F = \begin{bmatrix} F_Z & \\ & F_A \end{bmatrix}, \quad W = \begin{bmatrix} K_Z \\ 0 \end{bmatrix}, \quad K = \begin{bmatrix} 0 \\ K_A \end{bmatrix} \quad , H = [H_Z, H_A] \quad \text{and} \quad V = \Xi_0.$$

Since $\Gamma_0 = \Xi_0 = V$ and A_t and U_t are uncorrelated, letting

$$A_t^* = \begin{bmatrix} I & V \\ & I \end{bmatrix} \begin{bmatrix} A_t \\ U_t \end{bmatrix},$$

the first part of the theorem follows.

To prove the converse, assume the state space model in innovations form

$$\begin{aligned}x_{t+1} &= F^* x_t + K^* A_t^*\\ Y_t^* &= H^* x_t + A_t^*, \qquad t = 1, 2, \ldots, m,\end{aligned}$$

where $Y_t^* = (Y_t', Z_t')'$,

$$F^* = \begin{bmatrix} F & F_{1z} \\ & F_z \end{bmatrix}, \quad K^* = \begin{bmatrix} K & K_{1z} \\ & K_z \end{bmatrix}, \quad \text{and} \quad H^* = \begin{bmatrix} H & H_{1z} \\ & H_z \end{bmatrix}.$$

Letting r be the dimension of x_t, $Y_{t:t+r}^* = (Y_t^{*'}, Y_{t+1}^{*'}, \ldots, Y_{t+r}^{*'})'$, $A_{t:t+r}^* = (A_t^{*'}, A_{t+1}^{*'}, \ldots, A_{t+r}^{*'})'$ and $O_{r+1} = [H^{*'}, (H^*F^*)', \ldots, (H^*(F^*)^r)']'$, and letting $\hat{\Psi}_{r+1}$ be like in (3.28) but with H, K, and F replaced with H^*, K^*, and F^*, respectively, we can write

$$Y_{t:t+r}^* = O_{r+1}x_t + \hat{\Psi}_{r+1}A_{t:t+r}^*.$$

By the Cayley–Hamilton theorem, if $\det(\lambda I - F^*) = \lambda^r + f_1\lambda^{r-1} + \cdots + f_r$, then $(F^*)^r + f_1(F^*)^{r-1} + \cdots + f_rI = 0$. Thus, if we premultiply the previous equation by $[f_rI, \ldots, f_1I, I]$, it is obtained that

$$Y_{t+r}^* + f_1Y_{t+r-1}^* + \cdots + f_rY_t^* = A_{t+r}^* + \Theta_1A_{t+r-1}^* + \cdots + \Theta_rA_t^*,$$

where the Θ_i, $i = 1\ldots, r$, are block upper triangular matrices. Therefore, the previous expression corresponds to a VARMAX model in which all of the matrices in the autoregressive and moving average matrix polynomials are block upper triangular and both first matrices in these matrix polynomials are the unit matrix. We can apply Theorem 6.3 to complete the proof. □

6.5 Canonical Forms for VARMAX and State Space Models with Inputs

In this section, we will suppose a k-dimensional process $\{Y_t\}$ that follows a finite linear model with strongly exogenous inputs (6.9) such that $\Psi_0 = I$, $\{Z_t\}$ has dimension s and the rank of the augmented Hankel matrices, H_t^a, given by (6.10), is constant for $t > r$, where r is a fixed positive integer. Let $\{n_i : i = 1, \ldots, k\}$ be the Kronecker indices and $n = \sum_{i=1}^k n_i$ the McMillan degree corresponding to this linear model.

6.5.1 VARMAX Echelon Form

Like in the VARMA case, by the definition of the i-th Kronecker index and the structure of the augmented Hankel matrices, there exists a vector $\phi_i = [\phi_{i,n_i}, \ldots, \phi_{i,1}, \phi_{i,0}]$, where the $\phi_{i,j}$ have dimension $1 \times k$, $j = 0, 1, \ldots, n_i$, and $\phi_{i,0}$

has a one in the i-th position and zeros thereafter, such that

$$[0_{1\times k}, \ldots, 0_{1\times k}, \phi_{i,n_i}, \ldots, \phi_{i,1}, \phi_{i,0}, 0_{1\times k}, \ldots, 0_{1\times k}]H_t^a = 0, \quad t > n_i. \tag{6.36}$$

where the vector ϕ_i can be moved either to the left or to right in the previous expression without altering the relation due to the structure of H_t^a, $t > n_i$. This implies that if $l = \max\{n_i : i = 1, \ldots, k\}$, there exists a block vector $\Phi = [\Phi_l, \ldots, \Phi_1, \Phi_0]$ with Φ_0 a lower triangular matrix with ones in the main diagonal such that

$$\Phi H_{l+1}^a = 0. \tag{6.37}$$

In fact, the i-th row of Φ is $[0, \phi_i]$ if $n_i < l$ and ϕ_i if $n_i = l$, $i = 1 \ldots, k$. It follows from (6.37) that if we stack the observations to get

$$Y_{t:t+l} = \widehat{H}_{l+1}\alpha_1 + \hat{\Delta}_{l+1}U_{t:t+l} \tag{6.38}$$

where $Y_{t:t+l} = (Y_t', \ldots, Y_{t+l}')'$, $U_{t:t+l} = (U_1', \ldots, U_{t+l}')'$, $U_t = [Z_t', A_t']'$,

$$\widehat{H}_{l+1} = \begin{bmatrix} h_t \\ h_{t+1} \\ \vdots \\ h_{t+l} \end{bmatrix}, \text{ and } \hat{\Delta}_{l+1} = \begin{bmatrix} \Delta_{t-1} & \cdots & \Delta_1 & \Delta_0 & & & & \\ \Delta_t & \cdots & \Delta_2 & \Delta_1 & \Delta_0 & & & \\ \vdots & \cdots & \vdots & \vdots & \vdots & \ddots & \Delta_0 & \\ \Delta_{t+l-1} & \cdots & \Delta_{l+1} & \Delta_l & \Delta_{l-1} & \cdots & \Delta_1 & \Delta_0 \end{bmatrix},$$

and we premultiply (6.38) by Φ, the following VARMAX model is obtained

$$\Phi(B)Y_t = \Gamma(B)Z_t + \Theta(B)A_t, \tag{6.39}$$

where $\Phi(z) = \Phi_0 + \Phi_1 z + \cdots + \Phi_l z^l$, $\Gamma(z) = \Gamma_0 + \Gamma_1 z + \cdots + \Gamma_l z^l$, $\Theta(z) = \Theta_0 + \Theta_1 z + \cdots + \Theta_l z^l$, $\Theta_0 = \Phi_0$, and the Γ_i, Θ_i, $i = 0, 1, \ldots, l$, are given by the product of Φ and the last $l + 1$ blocks of columns of $\hat{\Delta}_{l+1}$. More specifically,

$$[\Gamma_j, \Theta_j] = \Phi_j\Delta_0 + \Phi_{j-1}\Delta_1 + \cdots + \Phi_0\Delta_j, \quad j = 0, 1, \ldots, l. \tag{6.40}$$

The VARMAX model (6.39) is called the VARMAX canonical form or the **VARMAX echelon form**.

Proceeding as in the VARMA case to obtain the restrictions in the coefficients of the echelon form, letting $\phi_{ip,j}$ and $\theta_{ip,j}$ be the (i,p)-th elements in the matrices Φ_j and Θ_j, $i, p = 1, \ldots, k$, $j = 0, 1, \ldots, l$, and letting $\gamma_{ip,j}$ be the (i,p)-th elements in the matrix Γ_j, $i = 1, \ldots, k$, $p = 1, \ldots, s$, $j = 0, 1, \ldots, l$, we can express the matrix polynomials

$$\Phi(z) = \left[\begin{array}{c|c|c|c|c} \phi_{11}(z) & \cdots & \phi_{1i}(z) & \cdots & \phi_{1k}(z) \\ \hline \vdots & \ddots & \vdots & \ddots & \vdots \\ \hline \phi_{i1}(z) & \cdots & \phi_{ii}(z) & \cdots & \phi_{ik}(z) \\ \hline \vdots & \ddots & \vdots & \ddots & \vdots \\ \hline \phi_{k1}(z) & \cdots & \phi_{ki}(z) & \cdots & \phi_{kk}(z) \end{array}\right] = \Phi_0 + \Phi_1 z + \cdots + \Phi_l z^l,$$

$$\Gamma(z) = \left[\begin{array}{c|c|c|c|c} \gamma_{11}(z) & \cdots & \gamma_{1i}(z) & \cdots & \gamma_{1s}(z) \\ \hline \vdots & \ddots & \vdots & \ddots & \vdots \\ \hline \gamma_{i1}(z) & \cdots & \gamma_{ii}(z) & \cdots & \gamma_{is}(z) \\ \hline \vdots & \ddots & \vdots & \ddots & \vdots \\ \hline \gamma_{k1}(z) & \cdots & \gamma_{ki}(z) & \cdots & \gamma_{ks}(z) \end{array}\right] = \Gamma_0 + \Gamma_1 z + \cdots + \Gamma_l z^l$$

and

$$\Theta(z) = \left[\begin{array}{c|c|c|c|c} \theta_{11}(z) & \cdots & \theta_{1i}(z) & \cdots & \theta_{1k}(z) \\ \hline \vdots & \ddots & \vdots & \ddots & \vdots \\ \hline \theta_{i1}(z) & \cdots & \theta_{ii}(z) & \cdots & \theta_{ik}(z) \\ \hline \vdots & \ddots & \vdots & \ddots & \vdots \\ \hline \theta_{k1}(z) & \cdots & \theta_{ki}(z) & \cdots & \theta_{kk}(z) \end{array}\right] = \Theta_0 + \Theta_1 z + \cdots + \Theta_l z^l$$

as follows

$$\phi_{ii}(z) = 1 + \sum_{j=1}^{n_i} \phi_{ii,j} z^j, \quad i = 1, \ldots, k, \tag{6.41}$$

$$\phi_{ip}(z) = \sum_{j=n_i-n_{ip}+1}^{n_i} \phi_{ip,j} z^j, \quad i \neq p \tag{6.42}$$

$$\gamma_{ip}(z) = \sum_{j=0}^{n_i} \gamma_{ip,j} z^j, \quad i = 1, \ldots, k,\ p = 1, \ldots, s, \tag{6.43}$$

$$\theta_{ip}(z) = \sum_{j=0}^{n_i} \theta_{ip,j} z^j, \quad i, p = 1, \ldots, k, \tag{6.44}$$

where $\Theta_0 = \Phi_0$ and

$$n_{ip} = \begin{cases} \min\{n_i + 1, n_p\} & \text{for } i > p \\ \min\{n_i, n_p\} & \text{for } i < p \end{cases} \quad i, p = 1, \ldots, k.$$

Note that n_{ip} specifies the number of free coefficients in the polynomial $\phi_{ip}(z)$ for $i \neq p$.

The following proposition is analogous to Proposition 5.1 and can be proved similarly.

Proposition 6.1 *The VARMAX echelon form (6.39) is left coprime.*

6.5.2 State Space Echelon Form

Letting $U_t = [Z_t', A_t']'$, define the forecasts of $\{Y_t\}$ as

$$\begin{aligned} Y_{t+i|t} &= Y_{t+i} - \Delta_0 U_{t+i} - \cdots - \Delta_{i-1} U_{t+1} \\ &= Y_{t+i} - (A_{t+i} + \Xi_0 Z_{t+i}) - \cdots - (\Psi_{i-1} A_{t+1} + \Xi_{i-1} Z_{t+1}) \end{aligned} \tag{6.45}$$

Then, we can write

$$\begin{bmatrix} Y_{t+1|t} \\ Y_{t+2|t} \\ \vdots \\ Y_{t+i|t} \end{bmatrix} = \left[\begin{array}{cccc|c} \Delta_1 & \Delta_2 & \cdots & \Delta_t & h_{t+1} \\ \Delta_2 & \Delta_3 & \cdots & \Delta_{t+1} & h_{t+2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \Delta_i & \Delta_{i+1} & \cdots & \Delta_{t+i-1} & h_{t+i} \end{array}\right] \left[\begin{array}{c} U_t \\ U_{t-1} \\ \vdots \\ U_1 \\ \hline \alpha_1 \end{array}\right], \tag{6.46}$$

and, in particular, for $i = t$

$$\begin{bmatrix} Y_{t+1|t} \\ \vdots \\ Y_{2t|t} \end{bmatrix} = \left[\begin{array}{c|c} H_t & \begin{matrix} h_{t+1} \\ \vdots \\ h_{2t} \end{matrix} \end{array}\right] \left[\begin{array}{c} U_t \\ \vdots \\ U_1 \\ \hline \alpha_1 \end{array}\right].$$

Thus, like in the VARMA case, we see that relations among rows of the augmented Hankel matrices, H_t^a, are equivalent to relations among forecasts. Also, it follows from (6.45) and (6.46) that

$$\begin{bmatrix} Y_t \\ Y_{t+1|t} \\ Y_{t+2|t} \\ \vdots \\ Y_{t+i|t} \end{bmatrix} = \begin{bmatrix} \Delta_0 \\ \Delta_1 \\ \Delta_2 \\ \vdots \\ \Delta_i \end{bmatrix} U_t + \begin{bmatrix} Y_{t|t-1} \\ Y_{t+1|t-1} \\ Y_{t+2|t-1} \\ \vdots \\ Y_{t+i|t-1} \end{bmatrix}. \tag{6.47}$$

Let us assume first that the Kronecker indices satisfy $n_i \geq 1$, $i = 1, \ldots, k$. Then, it follows from (6.36) and (6.46) that

$$\phi_{i,0} Y_{t+n_i+1|t} + \phi_{i,1} Y_{t+n_i|t} + \cdots + \phi_{i,n_i} Y_{t+1|t} = 0$$

and this in turn implies, by (6.41) and (6.42), that

$$Y^i_{t+n_i+1|t} + \sum_{j=1}^{n_i} \phi_{ii,j} Y^i_{t+n_i+1-j|t} + \sum_{i \neq p} \sum_{j=n_i-n_{ip}+1}^{n_i} \phi_{ip,j} Y^p_{t+n_i+1-j|t} = 0, \tag{6.48}$$

where $Y^p_{t+j|t}$ denotes the p-th element of $Y_{t+j|t}$, $p = 1, \ldots, k$, $j = 1, 2, \ldots$. From (6.47) and (6.48), the following relations are obtained

$$Y^i_{t+j|t} = Y^i_{t+j|t-1} + \Delta_{i,j} U_t, \qquad j = 1, 2 \ldots, n_i - 1 \tag{6.49}$$

$$Y^i_{t+n_i|t} = -\sum_{j=1}^{n_i} \phi_{ii,j} Y^i_{t+n_i-j|t-1} - \sum_{i \neq p} \sum_{j=n_i-n_{ip}+1}^{n_i} \phi_{ip,j} Y^p_{t+n_i-j|t-1} + \Delta_{i,n_i} U_t, \tag{6.50}$$

where $\Delta_{i,j}$ denotes the i-th row of the matrix Δ_j, $i = 1, 2, \ldots$.

By (6.46), to the basis of rows of the augmented Hankel matrices implied by the Kronecker indices and specified in Proposition 3.8 corresponds a basis of the space of forecasts. If we stack the elements of this basis of forecasts in the vector

$$x_{t+1} = \left[Y^1_{t+1|t}, \ldots, Y^1_{t+n_1|t}, Y^2_{t+1|t}, \ldots, Y^2_{t+n_2|t}, \ldots, Y^k_{t+1|t}, \ldots, Y^k_{t+n_k|t} \right]', \tag{6.51}$$

where $\dim(x_{t+1}) = \sum_{i=1}^k n_i$, by (6.47), (6.49), and (6.50), it is obtained that

$$x_{t+1} = F x_t + W Z_t + K A_t \tag{6.52}$$

$$Y_t = H x_t + V Z_t + A_t, \tag{6.53}$$

where

$$F = \left[\begin{array}{c|c|c|c|c} F_{11} & \cdots & F_{1i} & \cdots & F_{1k} \\ \hline \vdots & \ddots & \vdots & \ddots & \vdots \\ \hline F_{i1} & \cdots & F_{ii} & \cdots & F_{ik} \\ \hline \vdots & \ddots & \vdots & \ddots & \vdots \\ \hline F_{k1} & \cdots & F_{ki} & \cdots & F_{kk} \end{array} \right], \quad W = \left[\begin{array}{c} W_1 \\ \hline \vdots \\ \hline W_i \\ \hline \vdots \\ \hline W_k \end{array} \right], \quad K = \left[\begin{array}{c} K_1 \\ \hline \vdots \\ \hline K_i \\ \hline \vdots \\ \hline K_k \end{array} \right], \quad [W_i, K_i] = \left[\begin{array}{c} \Delta_{i,1} \\ \vdots \\ \Delta_{i,n_i} \end{array} \right],$$

$$F_{ii} = \begin{bmatrix} 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \\ -\phi_{ii,n_i} & \cdots & \cdots & -\phi_{ii,1} \end{bmatrix}, \quad F_{ip} = \begin{bmatrix} 0 & \cdots & 0 & 0 \cdots 0 \\ \vdots & \ddots & \vdots & \vdots \ddots \vdots \\ 0 & \cdots & 0 & 0 \cdots 0 \\ -\phi_{ip,n_i} & \cdots & -\phi_{ip,n_i-n_{ip}+1} & 0 \cdots 0 \end{bmatrix},$$

$$H = \left[\begin{array}{cccc|cccc|c|cccc} 1 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 \\ 0 & \cdots & 0 & 0 & 1 & \cdots & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 1 & \cdots & 0 & 0 \end{array}\right], \quad V = \Gamma_0,$$

F_{ii} is $n_i \times n_i$, F_{ip} is $n_i \times n_p$, and H is $k \times (n_1 + \cdots + n_k)$. The state space form (6.52) and (6.53) is called the **state space echelon form**.

The following proposition is analogous to Proposition 5.2. The proof is omitted.

Proposition 6.2 *The state space echelon form (6.52) and (6.53) is minimal.*

As in the VARMA case, when some of the Kronecker indices are zero, the echelon form is different from the one we have just described. To illustrate the situation, suppose the same example as in Sect. 5.9.2 but with two inputs and let $[3, 0, 2, 0]$ be the vector of Kronecker indices. Thus, $k = 4$, $s = 2$ and the McMillan degree is $n = 5$. Letting $Y_t = (Y_{1t}, \ldots, Y_{4t})'$ and considering the second and fourth rows of the VARMAX echelon form (6.41), (6.42), (6.43), and (6.44), we get the equations

$$Y_{2t} + \phi_{21,0}Y_{1t} = A_{2t} + \phi_{21,0}A_{1t} + \sum_{j=1}^{2} \gamma_{2j,0}Z_{jt}$$

$$Y_{4t} + \phi_{41,0}Y_{1t} + \phi_{43,0}Y_{3t} = A_{4t} + \phi_{41,0}A_{1t} + \phi_{43,0}A_{3t} + \sum_{j=1}^{2} \gamma_{4j,0}Z_{jt}.$$

It follows from this and the definition of the forecasts that

$$\begin{aligned} Y_{2t} &= A_{2t} - \phi_{21,0}(Y_{1t} - A_{1t}) + \sum_{j=1}^{2} \gamma_{2j,0}Z_{jt} \\ &= A_{2t} - \phi_{21,0}Y^{1}_{t|t-1} + \sum_{j=1}^{2}(\gamma_{2j,0} - \phi_{21,0}\gamma_{1j,0})Z_{jt} \\ Y_{4t} &= A_{4t} - \phi_{41,0}(Y_{1t} - A_{1t}) - \phi_{43,0}(Y_{3t} - A_{3t}) + \sum_{j=1}^{2} \gamma_{4j,0}Z_{jt} \\ &= A_{4t} - \phi_{41,0}Y^{1}_{t|t-1} - \phi_{43,0}Y^{3}_{t|t-1} + \sum_{j=1}^{2}(\gamma_{4j,0} - \phi_{41,0}\gamma_{1j,0} - \phi_{43,0}\gamma_{3j,0})Z_{jt}. \end{aligned}$$

Since the state vector is

$$x_{t+1} = \left[Y^1_{t+1|t}, Y^1_{t+2|t}, Y^1_{t+3|t}, Y^3_{t+1|t}, Y^3_{t+2|t}\right]',$$

we have to modify the H and V matrices of the echelon form so that they become

$$H = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ -\phi_{21,0} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ -\phi_{41,0} & 0 & 0 & -\phi_{43,0} & 0 \end{bmatrix},$$

$$V = \begin{bmatrix} \gamma_{11,0} & \gamma_{12,0} \\ \gamma_{21,0} - \phi_{21,0}\gamma_{11,0} & \gamma_{22,0} - \phi_{21,0}\gamma_{12,0} \\ \gamma_{31,0} & \gamma_{32,0} \\ \gamma_{41,0} - \phi_{41,0}\gamma_{11,0} - \phi_{43,0}\gamma_{31,0} & \gamma_{42,0} - \phi_{41,0}\gamma_{12,0} - \phi_{43,0}\gamma_{32,0} \end{bmatrix}.$$

The F, W, and K matrices of the echelon form are

$$F = \left[\begin{array}{c|c} F_{11} & F_{13} \\ \hline F_{31} & F_{33} \end{array}\right], \quad W = \left[\begin{array}{c} W_1 \\ \hline W_3 \end{array}\right], \quad K = \left[\begin{array}{c} K_1 \\ \hline K_3 \end{array}\right], \quad [W_i, K_i] = \begin{bmatrix} \Delta_{i,1} \\ \vdots \\ \Delta_{i,n_i} \end{bmatrix},$$

$$F_{ii} = \begin{bmatrix} 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \\ -\phi_{ii,n_i} & \cdots & \cdots & -\phi_{ii,1} \end{bmatrix}, \quad F_{ip} = \begin{bmatrix} 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 \\ -\phi_{ip,n_i} & \cdots & -\phi_{ip,n_i-n_{ip}+1} & 0 & \cdots & 0 \end{bmatrix},$$

$$i, p = 1, 3; \qquad n_1 = 3, n_2 = 0, n_3 = 2, n_4 = 0.$$

6.5.3 Relation Between the VARMAX and the State Space Echelon Forms

The following theorem is analogous to Theorem 5.5 and can be proved similarly.

Theorem 6.5 *Given the innovations process* $\{A_t\} \sim (0, \Sigma)$, $\Sigma > 0$, *the input process* $\{Z_t\}$, *and the initial conditions,* $I = \{Y_{11}, \ldots, Y_{1n_1}, \ldots, Y_{k1}, \ldots, Y_{kn_k}\}$, *of the process* $\{Y_t\}$, $Y_t = (Y_{1t}, \ldots, Y_{kt})'$, *the following statements are equivalent*

i) $\{Y_t\}$ *follows a finite linear model (6.9) with strongly exogenous inputs such that the rank of the augmented Hankel matrices,* H^a_t, *is constant for* $t > r$, *the Kronecker indices are* $\{n_i : i = 1, \ldots, k\}$, $dim(\alpha_1) = \sum_{i=1}^k n_i$, *and the initial conditions are* I.

ii) *$\{Y_t\}$ follows a VARMAX echelon form (6.39) such that the Kronecker indices are $\{n_i : i = 1, \ldots, k\}$, and the initial conditions are I.*

iii) *$\{Y_t\}$ follows a state space echelon form (6.52) and (6.53) such that the Kronecker indices are $\{n_i : i = 1, \ldots, k\}$, and the initial conditions are I.*

6.5.4 Decoupled VARMAX Echelon Form

Suppose that the echelon form of the process $\{Y_t\}$ is (6.39). Then, premultiplying by $\Phi^{-1}(B)$, it is obtained that

$$Y_t = \Phi^{-1}(B)\Gamma(B)Z_t + \Phi^{-1}(B)\Theta(B)A_t,$$

and we see that the coefficient matrices, $[\Xi_j, \Psi_j]$, satisfy the relations

$$(\Phi_0 + \Phi_1 z + \cdots + \Phi_l z^l)(\Xi_0 + \Xi_1 z + \cdots) = \Gamma_0 + \Gamma_1 z + \cdots + \Gamma_l z^l \quad (6.54)$$

$$(\Phi_0 + \Phi_1 z + \cdots + \Phi_l z^l)(I + \Psi_1 z + \cdots) = \Theta_0 + \Theta_1 z + \cdots + \Theta_l z^l. \quad (6.55)$$

In the relations (6.54) and (6.55) there may be common factors, but it is clear that if we define the processes $\{W_t\}$ and $\{V_t\}$ by $W_t = \sum_{j=0}^{t-1} \Xi_j Z_{t-j}$ and $V_t = \sum_{j=0}^{t-1} \Psi_j A_{t-j}$, (6.54) and (6.55) imply the existence of VARMA echelon forms for $\{W_t\}$ and $\{V_t\}$. Denote the echelon form of $\{V_t\}$ by

$$\Phi_V(B)V_t = \Theta_V(B)A_t. \quad (6.56)$$

This echelon form is as described in Sect. 5.9.1. However, the echelon form of $\{W_t\}$ has to be defined because the $k \times s$ matrix Ξ_0 is not in general the unit matrix. To this end, consider the Hankel matrix corresponding to $\{W_t\}$,

$$H_{W,t} = \begin{bmatrix} \Xi_1 & \Xi_2 & \Xi_3 & \cdots & \Xi_t \\ \Xi_2 & \Xi_3 & \Xi_4 & \cdots & \Xi_{t+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \Xi_t & \Xi_{t+1} & \Xi_{t+2} & \cdots & \Xi_{2t-1} \end{bmatrix}.$$

The relation (6.54) implies that there exists a fixed positive integer, u, such that $H_{W,t}$ has constant rank equal to u for $t > u$. We can thus proceed as in Sect. 5.9.1 to find matrix polynomials, $\Phi_W(z) = \Phi_{W,0} + \Phi_{W,1}z + \cdots + \Phi_{W,a}z^a$ and $\Gamma_W(z) = \Gamma_{W,0} + \Gamma_{W,1}z + \cdots + \Gamma_{W,a}z^a$, where $\Phi_{W,0}$ is a lower triangular matrix with ones in the main diagonal, such that

$$\Phi_W(B)W_t = \Gamma_W(B)Z_t. \quad (6.57)$$

and this is the echelon form of $\{W_t\}$. The matrix $\Gamma_{W,0}$ is not, in general, lower triangular. It can be singular, even zero.

Note that, $\Phi_{W,0}$ is different in general from $\Gamma_{W,0}$ and that the two pairs of left coprime polynomial matrices, $[\Phi_W(z), \Gamma_W(z)]$ and $[\Phi_V(z), \Theta_V(z)]$, satisfy

$$\Phi_W(z)(\Xi_0 + \Xi_1 z + \cdots) = \Gamma_W(z)$$
$$\Phi_V(z)(I + \Psi_1 z + \cdots) = \Theta_V(z).$$

Since

$$Y_t = W_t + V_t + h_t\alpha_1,$$

if we choose the initial conditions of (6.56) and (6.57) appropriately, the sum of the new two processes, $\{\tilde{W}_t\}$ and $\{\tilde{V}_t\}$, satisfies

$$Y_t = \tilde{W}_t + \tilde{V}_t. \tag{6.58}$$

More specifically, the processes $\{W_t\}$ and $\{V_t\}$ have been defined with zero initial conditions. Let $\tilde{W}_t = W_t + h_{W,t}\alpha_{W,1}$ be a solution of (6.57), where $\Phi_W(B)h_{W,t} = 0$ for $t > a$, and let $\tilde{V}_t = V_t + h_{V,t}\alpha_{V,1}$ be a solution of (6.56), where $\Phi_V(B)h_{V,t} = 0$ for $t > b$ and b is the degree of $\Phi_V(z)$. By an argument similar to that of Theorem 5.5, we can choose the initial conditions of $\{Y_t\}$, $\{\tilde{W}_t\}$ and $\{\tilde{V}_t\}$ so that (6.58) holds.

6.5.5 Decoupled State Space Echelon Form

Defining as in Sect. 6.5.4 the processes $\{W_t\}$ and $\{V_t\}$ by $W_t = \sum_{j=0}^{t-1} \Xi_j Z_{t-j}$ and $V_t = \sum_{j=0}^{t-1} \Psi_j A_{t-j}$, we saw in that section that both $\{W_t\}$ and $\{V_t\}$ can be expressed in VARMA echelon form, (6.56) and (6.57). Then, we saw in Sect. 5.9.2 that we can express $\{V_t\}$ in state space echelon form,

$$x_{V,t+1} = F_V x_{V,t} + K_V A_t \tag{6.59}$$
$$V_t = H_V x_{V,t} + A_t. \tag{6.60}$$

As for $\{W_t\}$, we can proceed as in Sect. 5.9.2 to define the echelon form. Define first

$$W_{t+i|t} = W_{t+i} - \Xi_0 Z_{t+i} - \Xi_1 Z_{t+i-1} + \cdots + \Xi_{i-1} Z_{t+1}, \quad i = 1, 2, \ldots.$$

Then, define the matrices F_W, K_W, and H_W and the state vector $x_{W,t}$ as in Sect. 5.9.2 with $\Phi(z)$, $\Theta(z)$ and $Y_{t+i|t}$ replaced with $\Phi_W(z)$, $\Theta_W(z)$, and $W_{t+i|t}$ to get

$$x_{W,t+1} = F_W x_{W,t} + K_W Z_t \tag{6.61}$$
$$W_t = H_W x_{W,t} + \Xi_0 Z_t. \tag{6.62}$$

This is the state space echelon form of $\{W_t\}$.

As in Sect. 6.5.4, if we choose the initial conditions of (6.59) and (6.60) and (6.61) and (6.62) appropriately, the sum of the new two processes, $\{\tilde{W}_t\}$ and $\{\tilde{V}_t\}$, satisfies

$$Y_t = \tilde{W}_t + \tilde{V}_t. \tag{6.63}$$

Using the state space forms (6.59) and (6.60) and (6.61) and (6.62) and the initial conditions in these forms so that (6.63) is satisfied, we can represent $\{Y_t\}$ in the following decoupled state space echelon form

$$\begin{aligned} x_{t+1} &= \begin{bmatrix} F_W & 0 \\ 0 & F_V \end{bmatrix} x_t + \begin{bmatrix} K_W \\ 0 \end{bmatrix} Z_t + \begin{bmatrix} 0 \\ K_V \end{bmatrix} A_t \\ Y_t &= [H_W, H_V] x_t + \Xi_0 Z_t + A_t, \end{aligned}$$

where $x_t = [x'_{W,t}, x'_{V,t}]'$.

6.6 Estimation of VARMAX Models Using the Hannan–Rissanen Method

Suppose that the process $\{Y_t\}$ follows the VARMAX model in echelon form

$$\Phi_0 Y_t + \cdots + \Phi_r Y_{t-r} = \Omega_0 Z_t + \cdots + \Omega_r Z_{t-r} + \Theta_0 A_t + \cdots + \Theta_r A_{t-r}, \tag{6.64}$$

where $\Phi_0 = \Theta_0$ is a lower triangular matrix with ones in the main diagonal. Equation (6.64) can be rewritten as

$$Y_t = (I_k - \Phi_0)V_t - \sum_{j=1}^{r} \Phi_j Y_{t-j} + \sum_{j=0}^{r} \Omega_j Z_{t-j} + \sum_{j=1}^{r} \Theta_j A_{t-j} + A_t, \tag{6.65}$$

where $V_t = Y_t - A_t$ and A_t in (6.65) is uncorrelated with Z_s, $s \leq t$, Y_u, A_u, $u \leq t-1$, and

$$V_t = \Phi_0^{-1} \left(-\sum_{j=1}^{r} \Phi_j Y_{t-j} + \sum_{j=0}^{r} \Omega_j Z_{t-j} + \sum_{j=1}^{r} \Theta_j A_{t-j} \right).$$

Applying the vec operator to (6.65), it is obtained that

$$\begin{aligned} Y_t = &-\sum_{j=1}^{r} (Y'_{t-j} \otimes I_k)\mathrm{vec}(\Phi_j) + \sum_{j=0}^{r} (Z'_{t-j} \otimes I_k)\mathrm{vec}(\Omega_j) - (V'_t \otimes I_k)\mathrm{vec}(\Theta_0 - I_k) \\ &+ \sum_{j=1}^{r} (A'_{t-j} \otimes I_k)\mathrm{vec}(\Theta_j) + A_t \end{aligned}$$

$$= [W_{1,t}, W_{2,t}, W_{3,t}] \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} + A_t$$

$$= W_t\alpha + A_t, \tag{6.66}$$

where $W_{1,t} = [-Y'_{t-1} \otimes I_k, \ldots, -Y'_{t-r} \otimes I_k]$, $W_{2,t} = [Z'_t \otimes I_k, \ldots, Z'_{t-r} \otimes I_k]$, $W_{3,t} = [-V'_t \otimes I_k, A'_{t-1} \otimes I_k, \ldots, A'_{t-r} \otimes I_k]$, $\alpha_1 = [\text{vec}'(\Phi_1), \ldots, \text{vec}'(\Phi_r)]'$, $\alpha_2 = [\text{vec}'(\Omega_0), \ldots, \text{vec}'(\Omega_r)]'$, $\alpha_3 = [\text{vec}'(\Theta_0 - I_k), \text{vec}'(\Theta_1), \ldots, \text{vec}'(\Theta_r)]'$, $W_t = [W_{1,t}, W_{2,t}, W_{3,t}]$ and $\alpha = [\alpha'_1, \alpha'_2, \alpha'_3]'$.

The restrictions in the parameters of the echelon form (6.64) can be incorporated into Eq. (6.66) by defining a selection matrix, R, containing zeros and ones such that

$$\alpha = R\beta, \tag{6.67}$$

where β is the vector of parameters that are not restricted in the matrices Φ_i, Ω_i or Θ_i, $i = 0, 1, \ldots, r$. Using (6.67), Eq. (6.66) can be rewritten as

$$Y_t = W_tR\beta + A_t$$

$$= X_t\beta + A_t, \tag{6.68}$$

where $X_t = W_tR$. Notice that, as mentioned earlier, X_t is uncorrelated with A_t in (6.68) and that if we knew X_t, we could estimate β by GLS. To this end, let $\text{Var}(A_t) = \Sigma$ and let L be a lower triangular matrix such that $\Sigma = LL'$ is the Cholesky decomposition of Σ. Then, The GLS estimator of β in (6.68) can be obtained as the OLS estimator of β in

$$L^{-1}Y_t = L^{-1}X_t\beta + L^{-1}A_t. \tag{6.69}$$

The idea behind the Hannan–Rissanen method (Hannan & Kavalieris, 1986; Hannan and Rissanen, 1982) is to estimate β in (6.69) after we have replaced the unknown innovations in X_t and $\text{Var}(A_t) = \Sigma$ with those estimated using a long vector autoregressive model with exogenous inputs (VARX). A VARX model is a model of the form

$$Y_t = \sum_{j=1}^{p} \Pi_j Y_{t-j} + \sum_{j=0}^{p} \Gamma_j Z_{t-j} + A_t, \tag{6.70}$$

and these models are important because every VARMAX model can be approximated to any degree of accuracy by a VARX model with a sufficiently big order.

VARX models are usually estimated using OLS, but this method of estimation can give nonstable models. To remedy this, Whittle's or the modified Burg's algorithm can be used instead.

To apply any of the two previously mentioned algorithms to estimate a VARX model of order p, define first $X_{0t} = Y_t$, $X_{1t} = (Y'_{t-1}, \ldots, Y'_{t-p})'$, $X_{2t} = (Z'_t, Z'_{t-1}, \ldots, Z'_{t-p})'$, $t = p+1, \ldots, n$, $\Pi = (\Pi_1, \ldots, \Pi_p)$, $\Gamma = (\Gamma_0, \Gamma_1, \ldots, \Gamma_p)$ and

$$M_{ij} = \frac{1}{n-p} \sum_{t=p+1}^{n} X_{it} X'_{jt}, \quad i,j = 0,1,2.$$

Then, model (6.70) can be written as

$$X_{0t} = \Pi X_{1t} + \Gamma X_{2t} + A_t$$

and, given Π and $\mathrm{Var}(A_t)$, the conditional maximum likelihood estimator of Γ in the previous model is

$$\hat{\Gamma} = M_{02} M_{22}^{-1} - \Pi M_{12} M_{22}^{-1}.$$

Therefore, Γ can be concentrated out of the likelihood and, to estimate Π, we can apply Whittle's or the modified Burg's algorithm to the model

$$R_{0t} = \Pi R_{1t} + U_t,$$

where $\mathrm{Var}(U_t) = \mathrm{Var}(A_t)$ and R_{0t} and R_{1t} are the residuals

$$R_{0t} = X_{0t} - M_{02} M_{22}^{-1} X_{2t}$$
$$R_{1t} = X_{1t} - M_{12} M_{22}^{-1} X_{2t}.$$

Once we have the estimate, $\hat{\Pi}$, of Π, the estimate of Γ is

$$\hat{\Gamma} = M_{02} M_{22}^{-1} - \hat{\Pi} M_{12} M_{22}^{-1}.$$

After the two steps of the Hannan–Rissanen method described so far, namely first estimation of the innovations using a VARX model and then estimation of β by GLS in (6.68) after having replaced the unknown innovations in X_t and $\mathrm{Var}(A_t) = \Sigma$ with the estimated ones, these authors proposed to perform a third step to correct the bias of the second step estimator.

This third step is as follows. Using the parameters estimated in the second step, we compute first new residuals from model equation (6.64),

$$\tilde{A}_t = Y_t + \Phi_0^{-1} \left(\sum_{j=1}^{r} \Phi_j Y_{t-j} - \sum_{j=0}^{r} \Omega_j Z_{t-j} - \sum_{j=1}^{r} \Theta_j \tilde{A}_{t-j} \right), \quad t = r+1, \ldots, n.$$

To initialize the recursion, either $\tilde{A}_t = 0$ or, preferably, $\tilde{A}_t = \widehat{A}_t$, $t = 1, \ldots, r$ can be used, where $\widehat{A}_t$ are the residuals obtained with the initial VARX model. Then, the third step consists of performing one Gauss–Newton iteration to minimize

$$f(\beta) = \sum_{t=r+1}^{n} \tilde{A}_t' \tilde{\Sigma}^{-1} \tilde{A}_t, \tag{6.71}$$

where $\tilde{\Sigma}$ has to be estimated and its parameters do not depend on β. Let $\overline{A}_t = \tilde{L}^{-1}\tilde{A}_t$ be the standardized residuals, where $\tilde{L}$ is a lower triangular matrix such that $\tilde{\Sigma} = \tilde{L}\tilde{L}'$ is the Cholesky decomposition of $\tilde{\Sigma}$, and consider the first order Taylor expansion,

$$\overline{A}_t(\beta) \simeq \left(\frac{\partial \overline{A}_t}{\partial \beta'} \right) \Bigg|_{\beta=\hat{\beta}} (\beta - \hat{\beta}) + \overline{A}_t(\hat{\beta}), \tag{6.72}$$

corresponding to the sum of squares function $f(\beta) = \sum_{t=r+1}^{n} \overline{A}_t'(\beta)\overline{A}_t(\beta)$, where $\hat{\beta}$ is the parameter vector estimated in the second step and we write $\overline{A}_t(\beta)$ to emphasize the dependence of $\overline{A}_t$ on β. Then, substituting the linear approximation (6.72) into $f(\beta)$, it is obtained that

$$f(\beta) \simeq \sum_{t=r+1}^{n} \left[\overline{A}_t(\hat{\beta}) - \left(\frac{\partial \overline{A}_t}{\partial \beta'} \right) \Bigg|_{\beta=\hat{\beta}} (\hat{\beta} - \beta) \right]' \left[\overline{A}_t(\hat{\beta}) - \left(\frac{\partial \overline{A}_t}{\partial \beta'} \right) \Bigg|_{\beta=\hat{\beta}} (\hat{\beta} - \beta) \right],$$

and it is clear that the value of β that minimizes $f(\beta)$ is the OLS estimator in the linear model

$$\overline{A}_t(\hat{\beta}) - \left(\frac{\partial \overline{A}_t}{\partial \beta'} \right) \Bigg|_{\beta=\hat{\beta}} \hat{\beta} = - \left(\frac{\partial \overline{A}_t}{\partial \beta'} \right) \Bigg|_{\beta=\hat{\beta}} \beta + \overline{A}_t(\beta), \tag{6.73}$$

where $\mathrm{Var}[\overline{A}_t(\beta)] = I_k$. To compute the partial derivative $\partial \overline{A}_t / \partial \beta'$, consider first the following rule (Lütkepohl, 2007, p. 666) to derive the vec of the product of two matrices, $A(\beta) \in \mathbb{R}^{n \times p}$ and $B(\beta) \in \mathbb{R}^{p \times q}$, with respect to β',

$$\frac{\partial \mathrm{vec}[A(\beta)B(\beta)]}{\partial \beta'} = [I_q \otimes A(\beta)] \frac{\partial \mathrm{vec}[B(\beta)]}{\partial \beta'} + [B'(\beta) \otimes I_n] \frac{\partial \mathrm{vec}[A(\beta)]}{\partial \beta'}. \tag{6.74}$$

Then, since $\tilde{L}^{-1}$ does not depend on β, by (6.74) applied to $\overline{A}_t = \tilde{L}^{-1}\tilde{A}_t$, we get

$$\frac{\partial \overline{A}_t}{\partial \beta'} = \tilde{L}^{-1} \frac{\partial \tilde{A}_t}{\partial \beta'},$$

and it suffices to compute $\partial\tilde{A}_t/\partial\beta'$. Since $\Theta_j\tilde{A}_{t-j} = \text{vec}(\Theta_j\tilde{A}_{t-j})$, by (6.74), we obtain

$$\frac{\partial(\Theta_j\tilde{A}_{t-j})}{\partial\beta'} = \Theta_j\frac{\partial\tilde{A}_{t-j}}{\partial\beta'} + (\tilde{A}'_{t-j}\otimes I_k)\frac{\partial\text{vec}(\Theta_j)}{\partial\beta'}$$

and, analogously,

$$\frac{\partial(\Phi_jY_{t-j})}{\partial\beta'} = (Y'_{t-j}\otimes I_k)\frac{\partial\text{vec}(\Phi_j)}{\partial\beta'} \qquad \text{and} \qquad \frac{\partial(\Gamma_jZ_{t-j})}{\partial\beta'} = (Z'_{t-j}\otimes I_k)\frac{\partial\text{vec}(\Gamma_j)}{\partial\beta'}.$$

Using these results in

$$\tilde{A}_t = Y_t+(\Theta_0-I_k)(Y_t-\tilde{A}_t)+\sum_{j=1}^{r}\Phi_jY_{t-j}-\sum_{j=0}^{r}\Omega_jZ_{t-j}-\sum_{j=1}^{r}\Theta_j\tilde{A}_{t-j}, \quad t=r+1,\ldots,n,$$

it follows that $C_t = \partial\tilde{A}_t/\partial\beta'$ can be computed using the recursion

$$C_t = -\Theta_0^{-1}\sum_{j=1}^{r}\Theta_jC_{t-j} - \Theta_0^{-1}X_t, \quad t=r+1,\ldots,n, \tag{6.75}$$

initialized with $C_j = 0, j = 1,2,\ldots,r$, where X_t is that of (6.68).

The Gauss–Newton iteration of the third step of the Hannan–Rissanen method consists of, using the estimation $\tilde{\Sigma} = \tilde{L}\tilde{L}' = \sum_{t=r+1}^{n}\tilde{A}_t\tilde{A}'_t/(n-r)$ in (6.71), computing the OLS estimator of β in the regression model (6.73), where $C_t = \partial\tilde{A}_t/\partial\beta'$ is computed using the recursion (6.75) with the parameters evaluated at $\hat{\beta}$.

In summary, the Hannan–Rissanen method consists of the following steps.

Step 1 Estimate a long VARX,

$$Y_t = \sum_{j=1}^{p}\Pi_jY_{t-j} + \sum_{j=0}^{p}\Gamma_jZ_{t-j} + A_t,$$

using OLS, Whittle's algorithm, or the modified Burg's algorithm, and obtain estimates of the residuals, $\widehat{A}_t$, $t = p+1,\ldots,n$, and of the residual covariance matrix, $\hat{\Sigma} = \sum_{t=p+1}^{n}\widehat{A}_t\widehat{A}'_t/(n-p)$.

The order of the VARX, p, can be obtained using an information criterion like AIC, the sequential likelihood ratio procedure, or using a formula like $p = [\ln(n)]^a$, where n is the sample size and $1.5 \le a \le 2$.

To obtain estimated residuals for $t = 1,\ldots,p$, one possibility is to fit VARX models of orders $0,1,\ldots,p-1$ and store the first residual of each of them, $\widehat{A}_1,\widehat{A}_2,\ldots,\widehat{A}_p$, before fitting the VARX model of order p.

Step 2 Using the residuals estimated in the first step, $\widehat{A}_t$, and the lower triangular matrix $\widehat{L}$ such that $\hat{\Sigma} = \widehat{L}\widehat{L}'$, estimate by OLS the parameter vector, β, in the model

$$\widehat{L}^{-1}Y_t = \widehat{L}^{-1}\widehat{X}_t\beta + E_t, \tag{6.76}$$

where $\widehat{X}_t$ is the matrix obtained from X_t in (6.68) by replacing the unknown A_t with $\widehat{A}_t$, $E_t = \widehat{L}^{-1}\left[A_t + (W_{3,t} - \widehat{W}_{3,t})R\beta\right]$ and $\widehat{W}_{3,t} = \left[(\widehat{A}_t - Y_t)' \otimes I_k, \widehat{A}'_{t-1} \otimes I_k, \ldots, \widehat{A}'_{t-r} \otimes I_k\right]$.

Step 3 Using the parameters estimated in Step 2, compute first new residuals from model equation (6.64),

$$\tilde{A}_t = Y_t + \Phi_0^{-1}\left(\sum_{j=1}^{r} \Phi_j Y_{t-j} - \sum_{j=0}^{r} \Omega_j Z_{t-j} - \sum_{j=1}^{r} \Theta_j \tilde{A}_{t-j}\right), \quad t = r+1, \ldots, n.$$

To initialize the recursion, either $\tilde{A}_t = 0$ or, preferably, $\tilde{A}_t = \widehat{A}_t$, $t = 1, \ldots, r$, can be used. Then, perform one Gauss–Newton iteration to minimize

$$f(\beta) = \sum_{t=r+1}^{n} \tilde{A}'_t \tilde{\Sigma}^{-1} \tilde{A}_t,$$

where the estimator $\tilde{\Sigma} = \sum_{t=r+1}^{n} \tilde{A}_t\tilde{A}'_t/(n-p)$ is used. To this end, compute the OLS estimator of β in the regression model

$$\tilde{L}^{-1}\left(\tilde{A}_t - C_t\hat{\beta}\right) = \left(-\tilde{L}^{-1}C_t\right)\beta + U_t,$$

where $\tilde{L}$ is a lower triangular matrix such that $\tilde{\Sigma} = \tilde{L}\tilde{L}'$, $\mathrm{Var}(U_t) = I_k$ and C_t is obtained using the recursion (6.75) with the parameters evaluated at $\hat{\beta}$. The third step can be iterated until the estimator of β stabilizes.

The Hannan–Rissanen method was successfully used in Gómez & Maravall (2001a) for the automatic model identification of 35 series. Of the 35 series, 13 were series which had appeared in published articles and for which an ARIMA model had been identified by some expert in time series analysis. The rest were simulated series.

6.7 Estimation of State Space Models with Inputs Using Subspace Methods

Using a notation borrowed from the engineering literature, let us assume that the output process $\{Y_t : t = 1, 2, \ldots, N\}$ has dimension k and follows the minimal innovations state space model

$$x_{t+1} = Ax_t + BZ_t + KA_t, \tag{6.77}$$

$$Y_t = Cx_t + DZ_t + A_t, \qquad t = 1, 2, \ldots, N \tag{6.78}$$

where $\{A_t\}$ is the sequence of k-dimensional innovations, which for simplicity are assumed to be i.i.d. Gaussian random variables with zero mean and covariance matrix $\Omega > 0$, $\{Z_t\}$ is the m-dimensional input process, which is assumed to be strongly exogenous, and $\{x_t\}$ is the sequence of n-dimensional unobserved states. The system matrices, $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, $C \in \mathbb{R}^{k\times n}$, $D \in \mathbb{R}^{k\times m}$, and $K \in \mathbb{R}^{n\times k}$ are to be estimated.

Choosing two integers, f and p, that stand for "future" and "past," we can define the following vectors

$$Y_{t,f}^{+} = \begin{bmatrix} Y_t \\ Y_{t+1} \\ \vdots \\ Y_{t+f-1} \end{bmatrix} \in \mathbb{R}^{fk}, \; Y_{t,p}^{-} = \begin{bmatrix} Y_{t-1} \\ Y_{t-2} \\ \vdots \\ Y_{t-p} \end{bmatrix} \in \mathbb{R}^{pk} \quad \text{and} \quad U_{t,p}^{-} = \begin{bmatrix} Z_{t,p}^{-} \\ Y_{t,p}^{-} \end{bmatrix} \in \mathbb{R}^{p(m+k)},$$

where $Z_{t,p}^{-}$ is defined analogously to $Y_{t,p}^{-}$ using the input vector Z_t instead of Y_t. In a similar way, we define $Z_{t,f}^{+}$ and $A_{t,f}^{+}$ using the input vector Z_t and the innovations A_t instead of Y_t.

By repeated application of Eqs. (6.77) and (6.78), it is obtained that

$$Y_{t+i} = CA^i x_t + \sum_{j=0}^{i} L_j Z_{t+i-j} + \sum_{j=0}^{i} K_j A_{t+i-j},$$

where $\{L_j\}$ and $\{K_j\}$ are the impulse response sequences, $L_j \in \mathbb{R}^{s\times m}$, $K_j \in \mathbb{R}^{s\times s}$, given by $L_0 = D$, $K_0 = I$, $L_j = CA^{j-1}B$, $K_j = CA^{j-1}K$, $j > 0$. From this, we get

$$Y_{t,f}^{+} = O_f x_t + Z_f Z_{t,f}^{+} + A_f A_{t,f}^{+}, \tag{6.79}$$

where $O_f = [C', A'C', \ldots, (A^{f-1})'C']'$ denotes the extended observability matrix,

$$Z_f = \begin{bmatrix} D & & & \\ CB & D & & \\ \vdots & \vdots & \ddots & D \\ CA^{f-2}B & \cdots & \cdots & CB\ D \end{bmatrix}, \quad \text{and} \quad A_f = \begin{bmatrix} I & & & \\ CK & I & & \\ \vdots & \vdots & \ddots & I \\ CA^{f-2}K & \cdots & \cdots & CK\ I \end{bmatrix}.$$

Note that Z_f and A_f are the Toeplitz matrices of the impulse responses $\{L_j\}$ and $\{K_j\}$.

From (6.78), we can write $A_t = Y_t - Cx_t - DZ_t$. Substituting this expression in (6.77) and iterating, it is obtained that

$$x_t = K_p U_{t,p}^- + (A - KC)^p x_{t-p}, \tag{6.80}$$

where $K_p = [K_{z,p}, K_{y,p}]$, $K_{z,p} = [B_K, A_K B_K, \ldots, A_K^{p-1} B_K]$, $K_{y,p} = [K, A_K K, \ldots, A_K^{p-1} K]$, $A_K = A - KC$ and $B_K = B - KD$.

We can rewrite Eq. (6.79) using (6.80) as

$$Y_{t,f}^+ = O_f K_p U_{t,p}^- + Z_f Z_{t,f}^+ + N_t, \tag{6.81}$$

where $N_t = O_f (A - KC)^p x_{t-p} + A_f A_{t,f}^+$. Defining $\beta_p = O_f K_p$ and $\beta_f = Z_f$ in Eq. (6.81), it is obtained that

$$Y_{t,f}^+ = \beta_p U_{t,p}^- + \beta_f Z_{t,f}^+ + N_t, \quad t = p+1, \ldots, T-f. \tag{6.82}$$

If the process $\{Y_t\}$ satisfies the minimum phase assumption, that is, the eigenvalues of the matrix $A - KC$ have modulus less than one, then the process admits an infinite autoregressive representation. Under this assumption, the central equation (6.82) has the following features:

- The vector N_t is asymptotically uncorrelated with the remaining terms on the right-hand side of Eq. (6.82) as $p \to \infty$ because $(A - KC)^p \to 0$.
- The matrix β_p has rank n, the system order.
- $N_t \to A_f A_{t,f}^+$ as $p \to \infty$.
- $A_f A_{t,f}^+$ is an MA(f) process.

The following subspace algorithm is based on the previous observations. It is known in the literature as the CCA method, proposed in Larimore (1983).

1. Since the rank of β_p in the regression equation (6.82) is equal to the system order n, use the results of Anderson (1951) based on partial canonical correlations to test H_0: rank$(\beta_p) = n$ for suitable values of n. After the rank of β_p has been determined, estimate β_p by $\hat{\beta}_p = \widehat{O}_f \widehat{K}_p$, where $\widehat{O}_f$ and $\widehat{K}_p$ both have rank n. From this and Eqs. (6.79)–(6.81), an estimate of the state based on past information is obtained as $\hat{x}_{t|t-1} = \widehat{K}_p Z_{t,p}^-$. Also, the matrices C and A can be estimated from the

observability matrix $O_f = [C', A'C', \ldots, (A^{f-1})'C']'$. The estimate of C is simply the first s rows of $\widehat{O_f}$ and the estimate of A is obtained by solving the following overdetermined system in the least squares sense,

$$\widehat{\begin{pmatrix} C \\ CA \\ \vdots \\ CA^{f-2} \end{pmatrix}} A = \widehat{\begin{pmatrix} CA \\ CA^2 \\ \vdots \\ CA^{f-1} \end{pmatrix}}.$$

It may be possible at this stage that the estimated A matrix is not stable (not all of its eigenvalues have modulus less than one). In this case, we should transform it into a stable matrix.

2. After having estimated the regression equation (6.82), the residuals, $\hat{N}_t$, are an estimator of $A_f A^{+}_{t,f}$. If $M_t = M_t^{1/2}(M_t^{1/2})'$ is the Cholesky decomposition of the covariance matrix of $\hat{N}_t$, where $M_t^{1/2}$ is a lower triangular matrix, then $M_t^{1/2}$ is an estimator of

$$A_f \operatorname{diag}(\Omega^{1/2}, \ldots, \Omega^{1/2}) = \begin{bmatrix} \Omega^{1/2} & 0 & \cdots & 0 \\ CK\Omega^{1/2} & \Omega^{1/2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ CA^{f-2}K\Omega^{1/2} & CA^{f-3}K\Omega^{1/2} & \cdots & \Omega^{1/2} \end{bmatrix}$$

and the first block column of the previous matrix is used, together with the estimated observability matrix, to estimate Ω and K. If the estimated $A - KC$ matrix is not stable, we would solve the following DARE

$$P = APA' - (APC' + S)(CPC' + R)^{-1}(APC' + S)' + Q$$

to get the stabilizing solution $P \geq 0$ and the corresponding Kalman gain $K = (APC' + S)(CPC' + R)^{-1}$ making $A - KC$ stable, where

$$\widehat{W} = \begin{bmatrix} Q & S \\ S' & R \end{bmatrix}$$

is an estimate of $W = \operatorname{Var}(A_t'K', A_t')'$ in (6.77) and (6.78).

3. There are two methods to estimate the matrices B and D. In the first method, the matrices B and D and the initial state vector, x_1, are estimated by regression using the estimated matrices A, C and, possibly, K. If K is not used, the regression equation is

$$Y_t = \left[C(z^{-1}I - A)_t^{-1}B + D\right]Z_t + CA^{t-1}x_1 + v_t,$$

where z^{-1} is the forward operator, $z^{-1}Y_t = Y_{t+1}$, $(z^{-1}I - A)_t^{-1} = \sum_{j=1}^{t-1} z^j A^{j-1}$, and $v_t = \left[C(z^{-1}I - A)_t^{-1}K + I\right]A_t$. If, on the contrary, K is used, then the regression equation is

$$\left[I - C(z^{-1}I - A_K)_t^{-1}K\right]Y_t = \left[C(z^{-1}I - A_K)_t^{-1}B_K + D\right]Z_t + CA_K^{t-1}x_1 + A_t, \qquad (6.83)$$

where $(z^{-1}I - A_K)_t^{-1} = \sum_{j=1}^{t-1} z^j A_K^{j-1}$ and, as before, $A_K = A - KC$ and $B_K = B - KD$. Note that in this last equation the residuals are white. The regression in (6.83) can be efficiently performed using the augmented Kalman filter corresponding to (6.77) and (6.78) in which the initial state, x_1, is considered fixed and the matrices D and B are to be estimated along with x_1. As mentioned earlier, (6.77) and (6.78) can be first transformed into a state space model with constant bias. Then, since $P_1 = 0$, the augmented Kalman filter is reduced to a couple of equations. See Problem 6.2.

Let

$$\hat{\beta}_f = \Sigma_{fz|p}\Sigma_{zz|p}^{-1},$$

be the estimator of β_f in (6.82), where $\Sigma_{ff|p}$, $\Sigma_{zz|p}$, and $\Sigma_{fz|p}$ are the covariance matrices of the residuals of the regression of $Y_{t,f}^{+}$ and $Z_{t,f}^{+}$ onto $U_{t,p}^{-}$. Then, according to the second method, the estimators of B and D are obtained by least squares from

$$\begin{bmatrix} D & & & \\ CB & D & & \\ \vdots & \vdots & \ddots & D \\ CA^{f-2}B & \cdots & \cdots\ CB & D \end{bmatrix} = \hat{\beta}_f,$$

considering that once we know C and A the previous equation is a linear system in B and D. See Katayama (2005, p. 289) for details. A problem with the previous estimator is that $\hat{\beta}_f$ may be not block lower triangular. However, this can be circumvented by using a constrained least squares method to estimate β_p and β_f. See Katayama & Picci (1999, pp. 1643–1644) for details.

4. The matrix $\Omega = \mathrm{Var}(A_t)$ can be estimated in several ways besides the one described earlier when estimating K. For example, one can use the regression equation (6.83) to estimate the residuals A_t. Another possibility is to use the observation equation (6.78) with the state x_t replaced with the estimated state $\hat{x}_{t|t-1}$ to estimate the residuals.

It is to be noted that the test statistic for H_0: $\mathrm{rank}(\beta_p) = n$ in (6.82), based on partial canonical correlations (Anderson, 1951), used in step 1 of the previous algorithm has an asymptotic chi-squared distribution under the assumption that the errors are white noise. This is not the case in Eq. (6.82), however, because the errors are asymptotically a moving average process. For this reason, Tsay (1989) has proposed a modification of this statistic.

An alternative to use partial canonical correlations to estimate the system order is to use an information criterion like AIC or BIC. To this end, we would use the augmented Kalman filter to compute the likelihood corresponding to (6.77) and (6.78) in which the initial state, x_1, is considered fixed and the matrices D and B are to be estimated along with x_1.

As mentioned earlier, the procedure in step 1 of the previous algorithm is based on the computation of an SVD. More specifically, let $\Sigma_{ff|z}$, $\Sigma_{pp|z}$ and $\Sigma_{fp|z}$ be the covariance matrices of the residuals of the regression of $Y_{t,f}^{+}$ and $U_{t,p}^{-}$ onto $Z_{t,f}^{+}$ and let

$$\hat{\beta}_p = \Sigma_{fp|z}\Sigma_{pp|z}^{-1}$$

be the estimator of β_p in (6.82). Then, computing the SVD

$$\Sigma_{ff|z}^{-1/2}\Sigma_{fp|z}(\Sigma_{pp|z}^{-1/2})' = USV', \tag{6.84}$$

one estimates the partial canonical correlations between $Y_{t,f}^{+}$ and $U_{t,p}^{-}$, given $Z_{t,f}^{+}$, as the elements in the diagonal of S, see Problem 6.9. After having tested for the number of canonical correlations that are nonzero, we have the approximation

$$\Sigma_{ff|z}^{-1/2}\Sigma_{fp|z}(\Sigma_{pp|z}^{-1/2})' \simeq \widehat{U}\widehat{S}\widehat{V}',$$

where $\widehat{S}$ has rank n, the estimated system order. The matrices O_f and K_p are estimated as

$$\widehat{O}_f = \Sigma_{ff|z}^{1/2}\widehat{U}\widehat{S}^{1/2} \quad \text{and} \quad \widehat{K}_p = \widehat{S}^{1/2}\widehat{V}'\Sigma_{pp|z}^{-1/2}.$$

As regards the orders f and p in the previous algorithm, there is no general consensus. One possibility is to fit a VARX approximation to the data and make p equal to the number of lags in the fitted autoregression. The number of future lags, f, should be at least equal to the maximum system order, n, considered. Some authors simply put $f = p$. Another possibility is to make f grow with the sample size according to a formula of the type $f = \log^a(N)$ with $1.5 \leq a \leq 2$.

The previous algorithm was used in Gómez, Aparicio-Pérez, & Sánchez-Ávila (2010) for forecasting two time series as an alternative to forecasting with transfer function methods.

6.8 Fast Estimation of State Space Models with Inputs

Suppose the following state space representation of the k-dimensional process $\{Y_t\}$ with strongly exogenous inputs, $\{Z_t\}$,

$$x_{t+1} = Fx_t + WZ_t + Gu_t \tag{6.85}$$

$$Y_t = Hx_t + VZ_t + Ju_t, \qquad t = 1, 2, \ldots, n, \tag{6.86}$$

where $\dim(x_t) = r$, $\dim(Z_t) = s$, $\dim(u_t) = q$, $\mathrm{Var}(u_t) = I_q$ and the processes $\{x_t\}$ and $\{Y_t\}$ can be stationary or not. Solving the DARE (A.3) in the Appendix to this chapter, we can put (6.85) and (6.86) into innovations form

$$x_{t+1} = Fx_t + WZ_t + KA_t$$

$$Y_t = Hx_t + VZ_t + A_t, \qquad t = 1, 2, \ldots, n,$$

where $K = (FPH' + GJ')\Sigma^{-1}$, $\Sigma = JJ' + HPH'$ and $\mathrm{Var}(A_t) = \Sigma$.

Assuming that the initial state, $x_1 = \delta$, is fixed and unknown in the previous state space form, we can evaluate the profile likelihood. In order to do that, we can use the fast TSKF described in Sect. 6.3,

$$(E_t, e_t) = (0, 0, Y_t - VZ_t) - H(-U_t, x_{t|t-1})$$

$$(-U_{t+1}, x_{t+1|t}) = (0, 0, WZ_t) + F(-U_t, x_{t|t-1}) + K(E_t, e_t),$$

with initial conditions $(-U_1, x_{1|0}) = (-I, 0)$ and $P_1 = 0$, where $P_t = P$, $K_t = K$, and $\Sigma_t = \Sigma$ for all $t = 1, 2, \ldots, n$, and the corresponding square root information form of the bias filter. It is to be noticed that the recursions of the information form bias filter are given by

$$\left(\Pi_{t+1}^{-1}, \Pi_{t+1}^{-1}\hat{\delta}_{t+1}\right) = \left(\Pi_t^{-1}, \Pi_t^{-1}\hat{\delta}_t\right) + E_t'\Sigma^{-1}\left(E_t, e_t\right),$$

initialized with $\left(\Pi_1^{-1}, \Pi_1^{-1}\hat{\delta}_1\right) = (0, 0)$, and that the profile log-likelihood can be expressed in terms of the fast TSKF and the information bias filter as

$$l(Y) = \text{constant} - \frac{1}{2}\left\{\left[RSS_{n+1}^{BFF} - \left(\Pi_{n+1}^{-1}\hat{\delta}_{n+1}\right)' \Pi_{n+1}\left(\Pi_{n+1}^{-1}\hat{\delta}_{n+1}\right)\right] + n\ln|\Sigma|\right\},$$

where RSS_{n+1}^{BFF} can be obtained with the recursion (4.138).

By Theorem 6.1, the k-dimensional process $\{Y_t\}$ can be represented by a state space model with constant bias. Since $WZ_t = \mathrm{vec}(WZ_t) = (Z_t' \otimes I_r)\mathrm{vec}(W)$ and $VZ_t = \mathrm{vec}(VZ_t) = (Z_t' \otimes I_k)\mathrm{vec}(V)$, if we define $\beta_v = \mathrm{vec}(V)$, $\beta_w = \mathrm{vec}(W)$, $\beta = (\beta_v', \beta_w')'$, $W_t = (0, Z_t' \otimes I_r)$ and $V_t = (Z_t' \otimes I_k, 0)$, we can write (6.85) and (6.86) as

$$x_{t+1} = Fx_t + W_t\beta + Gu_t \tag{6.87}$$

$$Y_t = Hx_t + V_t\beta + Ju_t, \qquad t = 1, 2, \ldots, n. \tag{6.88}$$

If we continue to assume that the initial state, $x_1 = \delta$, is fixed and unknown in the previous state space form, we can now concentrate out of the profile likelihood the vector β, something that we could not do earlier. To this end, we first solve the DARE and transform (6.87) and (6.88) into innovations form

$$x_{t+1} = Fx_t + W_t\beta + KA_t \tag{6.89}$$

$$Y_t = Hx_t + V_t\beta + A_t, \qquad t = 1, 2, \ldots, n, \tag{6.90}$$

where $K = (FPH' + GJ')\Sigma^{-1}$, $\Sigma = JJ' + HPH'$ and $\mathrm{Var}(A_t) = \Sigma$. Then, we can use the corresponding fast TSKF and information form bias filter

$$(E_t, e_t) = (0, V_t, Y_t) - H(-U_t, x_{t|t-1})$$

$$(-U_{t+1}, x_{t+1|t}) = (0, -W_t, 0) + F(-U_t, x_{t|t-1}) + K(E_t, e_t),$$

with initial conditions $(-U_1, x_{1|0}) = (-I, 0, 0)$ and $P_1 = 0$, where $P_t = P$, $K_t = K$ and $\Sigma_t = \Sigma$ for all $t = 1, 2, \ldots, n$, and

$$\left(\Pi_{t+1}^{-1}, \Pi_{t+1}^{-1}\hat{\gamma}_{t+1}\right) = \left(\Pi_t^{-1}, \Pi_t^{-1}\hat{\gamma}_t\right) + E_t'\Sigma^{-1}\left(E_t, e_t\right),$$

initialized with $\left(\Pi_1^{-1}, \Pi_1^{-1}\hat{\gamma}_1\right) = (0, 0)$, where $\gamma = [\delta', \beta']'$. The profile likelihood is now

$$l(Y) = \text{constant} - \frac{1}{2}\left\{\left[RSS_{n+1}^{BFF} - \left(\Pi_{n+1}^{-1}\hat{\gamma}_{n+1}\right)'\Pi_{n+1}\left(\Pi_{n+1}^{-1}\hat{\gamma}_{n+1}\right)\right] + n\ln|\Sigma|\right\},$$

where RSS_{n+1}^{BFF} can be obtained with the recursion (4.138).

Summing up, the fast estimation procedure to estimate the parameters in the models (6.85) and (6.86) or (6.87) and (6.88) consists of first passing the model to innovations form solving the DARE and then evaluating the profile likelihood using the corresponding fast TSKF and square root information bias filter described earlier in this section.

If instead of the profile likelihood we want to evaluate the diffuse likelihood, we can modify the previous procedure to accommodate the diffuse part in the initial state vector as we now outline. Suppose that the k-dimensional process $\{Y_t\}$, stationary or not, is represented by the state space model (6.85) and (6.86), where $\mathrm{Var}(u_t) = \sigma^2 I_q$. Suppose further that this state space model is first transformed into the state space model with constant bias (6.87) and (6.88) and then into innovations form (6.89) and (6.90), as described earlier in this section. Let $x_1 = W_0\beta + A\delta + Bc_s$, $\mathrm{Var}(\delta) = \sigma^2\Pi$ and $\mathrm{Var}(c_s) = \sigma^2\Omega_s$, and suppose that δ and c_s are uncorrelated with

$\Pi^{-1} \to 0$. Define $\bar{\delta} = A\delta + Bc_s$ and $\mathrm{Var}(\bar{\delta}) = \sigma^2\overline{\Pi}$, where

$$\overline{\Pi} = A\Pi A' + B\Omega_s B' \\ = M \begin{bmatrix} \Pi & 0 \\ 0 & \Omega_s \end{bmatrix} M',$$

and suppose that the matrices $M = [A, B]$ and Ω_s are nonsingular. This specification of the initial state vector can always be obtained following the procedure of Sect. 4.14.2, but the matrix Ω_s can be singular. We will consider the case of singular Ω_s at the end of this section.

Replacing x_1 in Theorem 4.34 with $x_1 = W_0\beta + \bar{\delta}$ and proceeding as in the proof of that theorem, we get the following linear model corresponding to the state space model (6.89) and (6.90)

$$Y = R\bar{\delta} + v, \tag{6.91}$$

where $v \sim (S\beta, \sigma^2 V)$ and $\mathrm{Cov}(\bar{\delta}, v) = 0$. Under the assumption of normality, by Theorem 2.3, the log-likelihood of the linear model (6.91), denoted by $\lambda(Y)$, is given by

$$\lambda(Y) = \text{constant} - \frac{1}{2}\Big\{\ln|\sigma^2\overline{\Pi}| + \ln|\sigma^2 V| + \ln|\sigma^{-2}\left(\overline{\Pi}^{-1} + R'V^{-1}R\right)| + \hat{\bar{\delta}}'\overline{\Pi}^{-1}\hat{\bar{\delta}}/\sigma^2 \\ + (Y - R\hat{\bar{\delta}} - S\beta)'V^{-1}(Y - R\hat{\bar{\delta}} - S\beta)/\sigma^2\Big\},$$

where $\hat{\bar{\delta}} = (\overline{\Pi}^{-1} + R'V^{-1}R)^{-1}[R'V^{-1}(Y - S\beta)]$ and $\mathrm{MSE}(\hat{\bar{\delta}}) = \sigma^2(\overline{\Pi}^{-1} + R'V^{-1}R)^{-1}$.

Since $|\overline{\Pi}| = |\Pi||\Omega_s||M|^2$ and, by Remark 2.2, the diffuse log-likelihood, $\lambda_D(Y)$, can be obtained by taking the limit when $\Pi^{-1} \to 0$ of $\lambda(Y) + \frac{1}{2}\ln|\sigma^2\Pi|$, the diffuse log-likelihood is, apart from a constant,

$$\lambda_D(Y) = -\frac{1}{2}\Big\{\ln|\sigma^2\Omega_s| + \ln|M|^2 + \ln|\sigma^2 V| \\ + \ln|\sigma^{-2}\left(\overline{\Pi}^{-1} + R'V^{-1}R\right)| + \hat{\bar{\delta}}'\overline{\Pi}^{-1}\hat{\bar{\delta}}/\sigma^2 \\ + (Y - R\hat{\bar{\delta}} - S\beta)'V^{-1}(Y - R\hat{\bar{\delta}} - S\beta)/\sigma^2\Big\}.$$

Concentrating out of the diffuse log-likelihood β and σ^2 yields the (β, σ^2)-maximized diffuse log-likelihood, denoted by $\lambda_D(Y; \hat{\beta}, \hat{\sigma}^2)$. This log-likelihood is, apart from a constant,

$$\lambda_D(Y;\hat{\beta},\hat{\sigma}^2) = -\frac{1}{2}\left\{(nk-n_\delta)\ln\left[\hat{\hat{\delta}}'\overline{\Pi}^{-1}\hat{\hat{\delta}} + (Y-R\hat{\hat{\delta}}-S\hat{\beta})'V^{-1}(Y-R\hat{\hat{\delta}}-S\hat{\beta})\right]\right.$$
$$\left. +\ \ln|\Omega_s| + \ln|M|^2 + \ln|V| + \ln|\overline{\Pi}^{-1} + R'V^{-1}R|\right\}, \qquad (6.92)$$

where n_δ is the dimension of δ. To evaluate the concentrated diffuse log-likelihood (6.92), we can use the following fast TSKF and information form bias filter

$$(E_t, e_t) = (0, V_t, Y_t) - H(-U_t, x_{t|t-1})$$
$$(-U_{t+1}, x_{t+1|t}) = (0, -W_t, 0) + F(-U_t, x_{t|t-1}) + K(E_t, e_t),$$

with initial conditions $(-U_1, x_{1|0}) = (-I, -W_0, 0)$ and $P_1 = 0$, where $P_t = P$, $K_t = K$ and $\Sigma_t = \Sigma$ for all $t = 1, 2, \ldots, n$, and

$$\left(\Pi_{t+1}^{-1}, \Pi_{t+1}^{-1}\hat{\gamma}_{t+1}\right) = \left(\Pi_t^{-1}, \Pi_t^{-1}\hat{\gamma}_t\right) + E_t'\Sigma^{-1}(E_t, e_t),$$

initialized with $\left(\Pi_1^{-1}, \Pi_1^{-1}\hat{\gamma}_1\right) = \left[\mathrm{diag}\left(\overline{\Pi}^{-1}, 0\right), 0\right]$, where

$$\overline{\Pi}^{-1} = (M')^{-1}\begin{bmatrix} 0 & \\ & \Omega_s^{-1}\end{bmatrix}M^{-1}$$

and $\gamma = [\bar{\delta}', \beta']'$. It is to be noticed that in the previous procedure the parameter σ^2 is one parameter that has been concentrated out of the covariance matrix of the innovations, $\mathrm{Var}(A_t)$, in the state space model (6.89) and (6.90). Thus, $\mathrm{Var}(A_t) = \sigma^2\Sigma$. To see this, first consider that, by assumption,

$$E\left\{\begin{bmatrix} Gu_t \\ Ju_t \end{bmatrix}\left[u_s'G', u_s'J'\right]\right\} = \sigma^2\begin{bmatrix} GG' & GJ' \\ JG' & JJ' \end{bmatrix}\delta_{ts}.$$

Then, by the results in Sect. 5.6 concerning the DARE, the following equality holds

$$\sigma^2\begin{bmatrix} -P + FPF' + GG' & FPH' + GJ' \\ HPF' + JG' & JJ' + HPH' \end{bmatrix} = \sigma^2\begin{bmatrix} K \\ I \end{bmatrix}\Sigma\,[K' \quad I],$$

proving that $\mathrm{Var}(A_t) = \sigma^2\Sigma$.

Using the previous fast TSKF and information form bias filter, we can evaluate $\lambda_D(Y;\hat{\beta},\hat{\sigma}^2)$, apart from a constant, as

$$\lambda_D(Y;\hat{\beta},\hat{\sigma}^2) = -\frac{1}{2}\left\{(nk-n_\delta)\ln\left[RSS_{n+1}^{BFF} - \left(\Pi_{n+1}^{-1}\hat{\gamma}_{n+1}\right)'\Pi_{n+1}\left(\Pi_{n+1}^{-1}\hat{\gamma}_{n+1}\right)\right]\right.$$
$$\left. +\ \ln|\Omega_s| + \ln|M|^2 + n\ln|\Sigma| + \ln|\Pi_{n+1}|,\right\},$$

where RSS_{n+1}^{BFF} can be obtained with the recursion (4.138).

If Ω_s is singular, let $\Omega_s = USU'$, where U is an orthogonal and S is a nonnegative diagonal matrix, be the singular value decomposition of Ω_s. Assuming that the singular values are in the diagonal of S in descending order, let S_1 be the submatrix of S containing all the nonzero diagonal elements of S and let U_1 be the submatrix of U such that $\Omega_s = U_1 S_1 U_1'$. Then, the previous procedure remains valid (see Problem 6.6) if we replace Ω_s^{-1} with $\Omega_s^{-} = U_1 S_1^{-1} U_1^{-1}$ and $|\Omega_s|$ with $|S_1|$ in the previous expressions for $\overline{\Pi}^{-1}$ and $\lambda_D(Y; \hat{\beta}, \hat{\sigma}^2)$, respectively.

6.9 The Information Matrix

Let the process $\{Y_t\}$ follow the state space model (6.1) and (6.2) with strongly exogenous inputs, $\{Z_t\}$. Suppose that $\sigma^2 = 1$ and the initial state vector is $x_1 = M\beta + x$, so that there is no diffuse part. Then, applying the Kalman filter (6.19)–(6.21), initialized with $\hat{x}_{1|0} = M\beta$ and $P_1 = \Omega$, to the sample $Y = (Y_1', \ldots, Y_n')'$ to evaluate the log-likelihood, $l(Y)$, yields

$$l(Y) = -\frac{1}{2}\sum_{t=1}^{n}\left[k\ln(2\pi) + E_t'\Sigma_t^{-1}E_t + \ln|\Sigma_t|\right].$$

Letting γ be the vector of parameters in the model (6.1) and (6.2) and applying the rules for vector and matrix differentiation, it is obtained that the gradient of the log-likelihood is

$$\begin{aligned}
\frac{\partial l(Y)}{\partial \gamma'} &= -\frac{1}{2}\sum_{t=1}^{n}\left[\mathrm{vec}\left(\frac{\partial \ln|\Sigma_t|}{\partial \Sigma_t}\right)'\frac{\partial \mathrm{vec}(\Sigma_t)}{\partial \gamma'} + \frac{\partial \mathrm{tr}\left(E_t'\Sigma_t^{-1}E_t\right)}{\partial \gamma'}\right] \\
&= -\frac{1}{2}\sum_{t=1}^{n}\left[\mathrm{vec}\left(\Sigma_t^{-1}\right)'\frac{\partial \mathrm{vec}(\Sigma_t)}{\partial \gamma'} + 2E_t'\Sigma_t^{-1}\frac{\partial E_t}{\partial \gamma'} - \mathrm{vec}\left(\Sigma_t^{-1}E_tE_t'\Sigma_t^{-1}\right)'\frac{\partial \mathrm{vec}(\Sigma_t)}{\partial \gamma'}\right] \\
&= -\frac{1}{2}\sum_{t=1}^{n}\left\{\mathrm{vec}\left[\Sigma_t^{-1}\left(I_k - E_tE_t'\Sigma_t^{-1}\right)\right]'\frac{\partial \mathrm{vec}(\Sigma_t)}{\partial \gamma'} + 2E_t'\Sigma_t^{-1}\frac{\partial E_t}{\partial \gamma'}\right\}.
\end{aligned}$$

The information matrix, $\mathcal{I}(\gamma)$, is defined as

$$\mathcal{I}(\gamma) = E\left[-\frac{\partial^2 l(Y)}{\partial\delta\partial\delta'}\right].$$

Using $E(E_t) = 0$, $E(E_tE_t') = \Sigma_t$ and again the rules for vector and matrix differentiation yields

$$\mathcal{I}(\gamma) = \frac{1}{2}\sum_{t=1}^{n}\left[\frac{\partial \mathrm{vec}(\Sigma_t)'}{\partial \gamma}\left(\Sigma_t^{-1}\otimes\Sigma_t^{-1}\right)\frac{\partial \mathrm{vec}(\Sigma_t)}{\partial \gamma'} + 2E\left(\frac{\partial E_t'}{\partial \gamma}\Sigma_t^{-1}\frac{\partial E_t}{\partial \gamma'}\right)\right].$$

6.10 Historical Notes

Granger (1963) is a fundamental article on the relation between feedback and causality that generated a large literature on the study of feedback in stochastic processes. See, for example, Caines & Chan (1975) and Anderson & Gevers (1982).

Canonical forms for state space models with inputs and VARMAX models can be obtained in a way similar to that used with VARMA models. Since the argument is algebraic in nature, it is only the dimensions of the moving average part that change. The rest of the argument remains the same. Thus, all references given in Sect. 5.22 continue to be valid in this chapter.

The Hannan–Rissanen method was proposed by Hannan & Rissanen (1982) for the univariate case and extended to the multivariate case by Hannan & Kavalieris (1984, 1986). The method is also described in Hannan & Deistler (1988).

What is now called the "subspace" approach to time series identification, based on the predictor space construction and canonical correlation analysis, was proposed by Akaike (1974a) and Akaike (1974b). Subspace identification has been a very active area of research in the two decades between 1990 and 2010. See, for example, Larimore (1983), Van Overschee & De Moor (1994), and Van Overschee & De Moor (1996).

6.11 Problems

6.1 Prove the Kalman filter recursions (6.19)–(6.21) using the properties of orthogonal projection with nonzero means.

6.2 Suppose the innovations model (6.4) and (6.5) with $x_1 = M\beta + A\delta$, where β is unknown and δ is diffuse. Prove that the modified bias-free filter corresponding to this model,

$$(E_t, e_t) = (0, 0, Y_t - VZ_t) - H(-U_t, x_{t|t-1}) \tag{6.93}$$

$$\Sigma_t = HP_tH' + JJ', \quad K_t = (FP_tH' + GJ')\Sigma_t^{-1} \tag{6.94}$$

$$(-U_{t+1}, x_{t+1|t}) = (0, 0, WZ_t) + F(-U_t, x_{t|t-1}) + K_t(E_t, e_t) \tag{6.95}$$

$$P_{t+1} = (F - K_tH)P_tF' + (G - K_tJ)G', \tag{6.96}$$

with initial conditions $(-U_1, x_{1|0}) = (-A, -M, 0)$ and $P_1 = 0$, can be simplified to (6.93) and (6.95), where $G = K\Sigma^{1/2}$, $J = \Sigma^{1/2}$, $K_t = K$, and $\mathrm{Var}(A_t) = \Sigma = \Sigma^{1/2}\Sigma^{1/2'}$. Hint: check that $P_t = 0$ for $t = 1, 2, \ldots, n$.

6.3 Suppose the state space model (6.1) and (6.2), in which $\sigma^2 = 1$ and x_1 is given by (6.3) with β unknown and $\delta \neq 0$ diffuse. Prove that using the modified bias-free filter (6.22)–(6.25) and information bias filter (6.26) the log-likelihood is, apart from a constant,

$$l(Y) = -\frac{1}{2}\left\{\left[RSS_{n+1}^{BFF} - \left(\Pi_{n+1}^{-1}\hat{\gamma}_{n+1}\right)'\Pi_{n+1}\left(\Pi_{n+1}^{-1}\hat{\gamma}_{n+1}\right)\right] + \sum_{t=1}^{n}\ln|\Sigma_t| + \ln|\Pi_{n+1}^{11}|\right\},$$

where RSS_{n+1}^{BFF} can be obtained with the recursion (4.138),

$$\Pi_{n+1}^{-1} = \sum_{t=1}^{n} E_t'\Sigma_t^{-1}E_t = \begin{bmatrix}\Pi_{n+1}^{11} & \Pi_{n+1}^{12}\\ \Pi_{n+1}^{21} & \Pi_{n+1}^{22}\end{bmatrix},$$

and the partition is conformal with $\gamma = [\delta', \beta']'$.

6.4 Put the following univariate VARMAX model with one input

$$Y_t = 3Z_{t-1} - 2Z_{t-2} + A_t - .7A_{t-1}$$

into innovations state space form (6.4) and (6.5).

6.5 Suppose $\{Y_t\}$ follows the signal plus noise model

$$Y_t = \mu + S_t + N_t,$$

where μ is a constant, S_t satisfies $S_t - S_{t-1} = A_t$, $A_t \sim WN(0, \sigma_A^2)$, $N_t \sim WN(0, \sigma_N^2)$ and the processes $\{A_t\}$ and $\{N_t\}$ are mutually uncorrelated.

i) Show that $\{Y_t\}$ can be put into state space form (6.1) and (6.2) by defining $F = 1$, $G = (\sigma_A, 0)$, $H = 1$, $J = (0, \sigma_N)$, $V = \mu$, $W = 0$, $Z_t = 1$, $u_t = (A_{t+1}/\sigma_A, N_t/\sigma_N)'$, and $x_1 = \delta$ with δ diffuse.
ii) Let $\sigma_A^2 = 1$ and $\sigma_N^2 = 2$. Prove that the DARE corresponding to the previous state space model is

$$P^2 - P - 2 = 0$$

and that this equation has a positive solution $P = 2$. Obtain K and Σ as functions of the positive solution, P, of the DARE and the other parameters of the model.
iii) Under the assumptions of ii), obtain the fast TSKF and information form bias filter corresponding to the innovations state space model.

6.6 Prove the formula for $\lambda_D(Y; \hat{\beta}, \hat{\sigma}^2)$ outlined at the end of Sect. 6.8 corresponding to the case in which Ω_s is singular.

6.7 Given the univariate VARMAX model with two inputs

$$Y_t = (2B^3 + 4B^4)Z_{1t} + \frac{1.5B^2 + 3B^3}{1 - B + .24B^2}Z_{2t} + \frac{1}{1 - 1.3B + .4B^2}A_t,$$

where B is the backshift operator, $BY_t = Y_{t-1}$, put this model into a decoupled state space echelon form, as described in Sect. 6.5.5.

6.8 Suppose $\{Y_t\}$ follows the signal plus noise model

$$Y_t = \mu + S_t + N_t,$$

where μ is a constant, S_t satisfies $S_t + \phi S_{t-1} = A_t$, $|\phi| < 1$, $A_t \sim WN(0,1)$, $N_t \sim WN(0,2)$ and the processes $\{A_t\}$ and $\{N_t\}$ are mutually uncorrelated.

i) Show that $\{Y_t\}$ can be put into state space form (6.1) and (6.2) by defining $F = -\phi$, $G = (1,0)$, $H = 1$, $J = (0,\sqrt{2})$, $V = \mu$, $W = 0$, $Z_t = 1$, $u_t = (A_{t+1}, N_t/\sqrt{2})'$, and $x_1 \sim (0,\Omega)$, $\Omega = 1/(1-\phi^2)$.
ii) Assuming μ is known, write down the Kalman filter recursions (6.19)–(6.21) corresponding to the state space model defined in i).
iii) Assuming μ is not known, write down the modified bias-free filter (6.22)–(6.25) and information bias filter (6.26) corresponding to the state space model defined in i).

6.9 Prove that the elements in the diagonal of the matrix S in (6.84) are the canonical correlations between $Y^+_{t,f}$ and $U^-_{t,p}$, given $Z^+_{t,f}$.

6.10 With the notation of Sect. 6.7, let

$$\frac{1}{N}\begin{bmatrix} Z^+_{t,f} \\ U^-_{t,p} \\ Y^+_{t,f} \end{bmatrix} [Y^{+'}_{t,f}, U^{-'}_{t,p}, Z^{+'}_{t,f}] = \begin{bmatrix} \Sigma_{zz} & \Sigma_{zp} & \Sigma_{zf} \\ \Sigma_{pz} & \Sigma_{pp} & \Sigma_{pf} \\ \Sigma_{fz} & \Sigma_{fp} & \Sigma_{ff} \end{bmatrix}.$$

Suppose we perform the following QR decomposition

$$Q'\left\{\frac{1}{\sqrt{N}}[Y^{+'}_{t,f}, U^{-'}_{t,p}, Z^{+'}_{t,f}]\right\} = R,$$

so that the lower triangular matrix $L = R'$ satisfies, with an obvious notation,

$$\frac{1}{\sqrt{N}}\begin{bmatrix} Z^+_{t,f} \\ U^-_{t,p} \\ Y^+_{t,f} \end{bmatrix} = LQ' = \begin{bmatrix} L_{11} & 0 & 0 \\ L_{21} & L_{22} & 0 \\ L_{31} & L_{32} & L_{33} \end{bmatrix} Q'.$$

Prove that

$$\Sigma_{ff|z} = \Sigma_{ff} - \Sigma_{fz}\Sigma_{zz}^{-1}\Sigma_{zf} = L_{32}L'_{32} - L_{33}L'_{33}$$

and

$$\Sigma_{pp|z} = L_{22}L'_{22}, \qquad \Sigma_{fp|z} = L_{32}L'_{22}.$$

Appendix

Observability, Controllability, and the DARE

Let us suppose the following state space representation of the k-dimensional process $\{Y_t\}$ with strongly exogenous inputs, $\{Z_t\}$,

$$x_{t+1} = Fx_t + WZ_t + Gu_t \tag{A.1}$$

$$Y_t = Hx_t + VZ_t + Ju_t, \qquad t = 1, 2, \ldots, \tag{A.2}$$

where $\dim(x_t) = r$, $\dim(Z_t) = s$, $\dim(u_t) = q$ and $\mathrm{Var}(u_t) = I_q$.

The system (A.1) and (A.2) is controllable if the pair $(F, \tilde{G})$ is controllable, where $\tilde{G} = (W, G)$, and it is observable if the pair (F, H) is observable.

If the mean squared error matrix, P_t, of the orthogonal projection of the state, x_t, onto $\{Z_1, Y_1, \ldots, Z_{t-1}, Y_{t-1}, Z_t\}$ converges as $t \to \infty$ in the Kalman recursions

$$E_t = Y_t - VZ_t - H\hat{x}_{t|t-1}, \qquad \Sigma_t = HP_tH' + JJ',$$
$$K_t = (FP_tH' + GJ')\Sigma_t^{-1}, \qquad \hat{x}_{t+1|t} = WZ_t + F\hat{x}_{t|t-1} + K_tE_t,$$
$$P_{t+1} = (F - K_tH)P_tF' + (G - K_tJ)G',$$

initialized with $\hat{x}_{1|0} = a$ and $P_1 = \Omega$, then the limiting solution, P, will satisfy the DARE

$$P = FPF' + GG' - (FPH' + GJ')(JJ' + HPH')^{-1}(FPH' + GJ')'. \tag{A.3}$$

In this case, the Kalman gain, K_t, of the Kalman filter converges to the steady state quantity $K = (FPH' + GJ')\Sigma^{-1}$, where $\Sigma = JJ' + HPH'$. The Kalman filter recursions corresponding to this steady state yield the so-called **innovations representation** of the model (A.1) and (A.2),

$$\hat{x}_{t+1} = F\hat{x}_t + WZ_t + KA_t$$
$$Y_t = H\hat{x}_t + VZ_t + A_t,$$

where $\hat{x}_t$ is the predictor of x_t based on the semi-infinite sample, $\{Z_t, Z_s, Y_s : s \leq t-1\}$, $A_t = Y_t - H\hat{x}_{t|t-1} - VZ_t$ is the innovation and $\mathrm{Var}(A_t) = \Sigma$. It is usually assumed that Σ is nonsingular. It may happen that $\mathrm{Var}(\hat{x}_{t|t-1}) \to \infty$ as $t \to \infty$ while P_t has a finite limit. In this case, $\hat{x}_t$ can be considered as the limit of $\hat{x}_{t|t-1}$ as $t \to \infty$.

Chapter 7
Wiener–Kolmogorov Filtering and Smoothing

Thus far in the book, we have mainly considered projection problems involving a finite number of random vectors. In this chapter, we shall study problems involving an infinite collection of random vectors.

Projection problems for stochastic processes observed over doubly infinite and semi-infinite intervals were first considered by N. Wiener and A. N. Kolmogorov around 1940 in some now celebrated studies that are often regarded as the foundation of some disciplines such as **Statistical Prediction Theory** and **Signal Extraction**. It is interesting to mention that Wiener argued in the frequency domain while Kolmogorov did it in the time domain.

These authors developed some closed formulae for filtering and smoothing that are based on the covariance generating functions of the processes involved.

Bell (1984) extended the classical Wiener–Kolmogorov formulae for univariate processes with ARMA structure to the nonstationary ARIMA case.

Although the Wiener–Kolmogorov formulae are valid for multivariate stationary processes without any particular structure, it is important to know what happens when some structure is assumed. In particular, if the processes under study are stationary processes with a Hankel matrix of finite rank, they can be represented by a time invariant state space model or, equivalently, a VARMA model.

7.1 The Classical Wiener–Kolmogorov Formulae

Suppose a zero-mean stationary vector random process $\{(S_t', Y_t')' : t \in \mathbb{Z}\}$ with covariance function

$$\gamma(h) = \begin{bmatrix} \gamma_S(h) & \gamma_{SY}(h) \\ \gamma_{YS}(h) & \gamma_Y(h) \end{bmatrix},$$

V. Gómez, *Multivariate Time Series With Linear State Space Structure*,
DOI 10.1007/978-3-319-28599-3_7

where $\gamma(h) = E[(S'_t, Y'_t)'(S'_{t-h}, Y'_{t-h})]$, and such that the covariance generating function

$$G(z) = \sum_{h=-\infty}^{\infty} \gamma(h)z^h = \begin{bmatrix} G_S(z) & G_{SY}(z) \\ G_{YS}(z) & G_Y(z) \end{bmatrix}, \tag{7.1}$$

converges for all z in some annulus containing the unit circle, $r^{-1} < |z| < r$ with $r > 1$. It follows from this that $\sum_{h=-\infty}^{\infty} \|\gamma(h)\| < \infty$, where $\|A\|$ denotes a norm for the matrix A such as $\|A\| = \sqrt{\text{tr}(A'A)}$, and that the covariance factorization $G_Y(z) = \Psi(z)\Sigma\Psi'(z^{-1})$ exists, where $\Psi(z) = \sum_{h=0}^{\infty} \Psi_h z^h$ and $\Psi^{-1}(z)$ are analytic in $D = \{z \in \mathbb{C} : |z| < 1\}$.

In addition, we will make the assumption that $G_Y(z) > 0$ for all z in $U = \{z : |z| = 1\}$. This implies that $\Psi(z)$ has an inverse that is analytic in some annulus that contains the unit circle. Also, the equality $Y_t = \Psi(B)A_t$ holds, where $\{A_t\}$ are the innovations and B is the backshift operator, $BA_t = A_{t-1}$.

We want to find the best linear predictor $E^*(S_t|Y_s : s \in T)$ of S_t based on the sample $\{Y_s : s \in T\}$, where T can be a doubly infinite set, $T = \mathbb{Z}$, or a semi-infinite set, $T = \{s \leq t + m : s \in \mathbb{Z}\}$, such that t and $m \geq 0$ are fixed and $t, m \in \mathbb{Z}$, where $\mathbb{Z}$ is the set of all integers. We will denote the estimator $E^*(S_t|Y_s : s \in T)$ by $\hat{S}_{t|\infty}$ if $T = \mathbb{Z}$, and by $\hat{S}_{t|t+m}$ if $T = \{s \leq t + m : s \in \mathbb{Z}\}$. In the special case in which $m = 0$, we will write $\hat{S}_{t|t}$ instead of $\hat{S}_{t|t+0}$. In the literature, the problem of finding $\hat{S}_{t|t}$ is known as **Wiener–Kolmogorov filtering** and that of finding $\hat{S}_{t|t+m}$ or $\hat{S}_{t|\infty}$ is known as **Wiener–Kolmogorov smoothing**. Assuming the series have state space structure, the problem of finite filtering and smoothing, that is, when $T = \{t_1, t_2, \ldots, t_n\}, t_1 < t_2 < \cdots < t_n, t_n \geq t$, can be solved using the state space framework and **Kalman filtering and smoothing**.

7.1.1 Wiener–Kolmogorov Smoothing

We are interested in finding a formula for $\hat{S}_{t|\infty}$, the projection of S_t onto the Hilbert space generated by the finite linear combinations of elements of $\{Y_t\}$, and their limits in mean square. As is well known, this projection is unique and has the form

$$\hat{S}_{t|\infty} = \sum_{j=-\infty}^{\infty} \Psi_{tj} Y_j,$$

for some set of filter weights, $\{\Psi_{tj}\}$, that need to be determined, where the series is mean square convergent. Assuming that the filter $\Psi_t(z) = \sum_{j=-\infty}^{\infty} \Psi_{tj} z^j$ is stable, the orthogonality condition yields

$$E\left[\left(S_t - \sum_{j=-\infty}^{\infty} \Psi_{tj} Y_j\right) Y_l'\right] = 0, \qquad l \in \mathbb{Z},$$

or

$$\gamma_{SY}(t-l) = \sum_{j=-\infty}^{\infty} \Psi_{tj} \gamma_Y(j-l), \qquad l \in \mathbb{Z}, \tag{7.2}$$

Note that, by the continuity of the inner product $< X, Y >= E(XY')$, we can write

$$E\left(\hat{S}_{t|\infty} Y_l'\right) = \lim_{n \to \infty} \sum_{j=-n}^{n} \Psi_{tj} E(Y_j Y_l') = \sum_{j=-\infty}^{\infty} \Psi_{tj} \gamma_Y(j-l),$$

and (7.2) is justified. If we apply in (7.2) the change of variables $t - l = h$ and $j - l = k$, we can write

$$\gamma_{SY}(h) = \sum_{k=-\infty}^{\infty} \Psi_{h+l,k+l} \gamma_Y(k), \qquad l \in \mathbb{Z}. \tag{7.3}$$

But since the left-hand side of (7.12) does not depend on l, putting $l = -h$ yields

$$\gamma_{SY}(h) = \sum_{k=-\infty}^{\infty} \Psi_{0,k-h} \gamma_Y(k), \qquad h \geq 0.$$

Since the set of weights $\{\Psi_{tj}\}$ in the orthogonal projection $\hat{S}_{t|\infty} = \sum_{j=-\infty}^{\infty} \Psi_{tj} Y_j$ is unique, we conclude that

$$\Psi_{0,k-h} = \Pi_{h-k},$$

for some sequence $\{\Pi_j\}$. Thus, we obtain

$$\gamma_{SY}(h) = \sum_{k=-\infty}^{\infty} \Pi_{h-k} \gamma_Y(k) = \sum_{j=-\infty}^{\infty} \Pi_j \gamma_Y(h-j), \tag{7.4}$$

and, therefore,

$$\hat{S}_{t|\infty} = \sum_{j=-\infty}^{\infty} \Pi_j Y_{t-j}.$$

Letting $\Pi(z) = \sum_{j=-\infty}^{\infty} \Pi_j z^j$ be the generating function of the sequence $\{\Pi_j : j \in \mathbb{Z}\}$, (7.4) implies

$$G_{SY}(z) = \Pi(z) G_Y(z).$$

Since, by assumption, $G_{SY}(z)$ and $G_Y(z)$ are convergent for all z in some annulus containing the unit circle, $r^{-1} < |z| < r$ with $r > 1$, and $G_Y(z)$ has no unit circle zeros, we can write

$$\Pi(z) = G_{SY}(z) G_Y^{-1}(z),$$

where $\Pi(z)$ is well defined and stable because it is convergent in some annulus containing the unit circle. We summarize this result in the following theorem.

Theorem 7.1 *Given the zero-mean stationary vector random process* $\{(S_t', Y_t')' : t \in \mathbb{Z}\}$ *with covariance generating function (7.1) convergent for all* z *in some annulus containing the unit circle,* $r^{-1} < |z| < r$ *with* $r > 1$*, and such that* $G_Y(z)$ *has no unit circle zeros, the estimator* $\hat{S}_{t|\infty}$ *of* S_t *based on the doubly infinite sample* $\{Y_s : s = 0, \pm 1, \pm 2, \ldots\}$ *is given by* $\hat{S}_{t|\infty} = \sum_{j=-\infty}^{\infty} \Pi_j Y_{t-j}$*, where the filter is stable (*$\sum_{j=-\infty}^{\infty} \|\Pi_j\| < \infty$*) and its weights,* Π_j*, are given by*

$$\Pi(z) = \sum_{j=-\infty}^{\infty} \Pi_j z^j = G_{SY}(z) G_Y^{-1}(z). \tag{7.5}$$

In addition, the error $E_{t|\infty} = S_t - \hat{S}_{t|\infty}$ *is stationary, has a covariance generating function,* $G_{E\infty}(z)$*, given by* $G_{E\infty}(z) = G_S(z) - G_{SY}(z) G_Y^{-1}(z) G_{YS}(z^{-1})$ *and*

$$\mathrm{MSE}(\hat{S}_{t|\infty}) = \mathrm{Var}(E_{t|\infty}) = \frac{1}{2\pi} \int_{-\pi}^{\pi} G_{E\infty}(e^{-ix}) dx.$$

Proof The only things that remain to be proved are the statements about the error $E_{t|\infty}$ and the formula for the mean squared error. Since $\hat{S}_{t|\infty}$ is the result of passing Y_t through the stable linear time invariant filter $\Pi(z)$, by Corollary 3.1, $\hat{S}_{t|\infty}$ is stationary. In addition, by Proposition 3.14, the covariance generating function, $G_{\hat{S}}(z)$, of $\hat{S}_{t|\infty}$ is

$$G_{\hat{S}}(z) = \Pi(z) G_Y(z) \Pi'(z^{-1}) = G_{SY}(z) G_Y^{-1}(z) G_{YS}(z^{-1}).$$

Since $E_{t|\infty}$ is orthogonal to S_t, we have the decomposition into orthogonal components

$$S_t = E_{t|\infty} + \hat{S}_{t|\infty},$$

from which the formula for the covariance generating function of the error follows. Then, by definition of spectrum and (3.46), we get the desired formula for the mean square error. □

Remark 7.1 The assumptions we are making to derive the formula for $\hat{S}_{t|\infty}$ and its mean square error are stronger than needed. See Gohbergh & Fel´dman (1974) for the general case. ◇

Example 7.1 Suppose the signal-plus-noise model

$$Y_t = S_t + N_t,$$

where S_t follows the model

$$(1 - \rho B)S_t = b_t,$$

$|\rho| < 1$, $\{b_t\} \sim \mathrm{WN}(0, \sigma_b^2)$, $\{N_t\} \sim \mathrm{WN}(0, \sigma_n^2)$ and $\{b_t\}$ and $\{N_t\}$ are mutually uncorrelated. Then,

$$\hat{S}_{t|\infty} = \frac{1}{1 + k(1 - \rho B)(1 - \rho F)} Y_t,$$

where F is the forward operator, $FY_t = Y_{t+1}$, and $k = \sigma_n^2/\sigma_b^2$.

Applying $1 - \rho B$ to Y_t gives

$$(1 - \rho B)Y_t = b_t + (1 - \rho B)N_t. \tag{7.6}$$

If $\{\gamma(h) : h \in \mathbb{Z}\}$ are the covariances of the right-hand side of (7.6), then $\gamma(0) = \sigma_b^2 + (1 + \rho^2)\sigma_n^2$, $\gamma(1) = -\rho\sigma_n^2$, and $\gamma(k) = 0$ for $k \neq 0, 1$. It is thus an MA(1) model and $\{Y_t\}$ follows the ARMA$(1, 1)$ model

$$(1 - \rho B)Y_t = (1 - \theta B)A_t,$$

where θ and $\sigma_a^2 = \mathrm{Var}(A_t)$ are determined from the covariance function factorization

$$\gamma(0) + \gamma(1)(z + z^{-1}) = (1 - \theta z)(1 - \theta z^{-1})\sigma_a^2.$$

Equating coefficients in the previous equation yields

$$\begin{aligned}(1 + \theta^2)\sigma_a^2 &= \sigma_b^2 + (1 + \rho^2)\sigma_n^2\\ -\theta\sigma_a^2 &= -\rho\sigma_n^2.\end{aligned}$$

Notice from these two equations that

$$\begin{aligned}\frac{\sigma_b^2}{\sigma_a^2} &= 1+\theta^2-(1+\rho^2)\frac{\sigma_n^2}{\sigma_a^2}\\ &= 1+\theta^2-(1+\rho^2)\frac{\theta}{\rho}\\ &= \frac{(\rho-\theta)(1-\rho\theta)}{\rho}. \qquad (7.7)\end{aligned}$$

The covariance generating function of the error $E_{t|\infty} = S_t - \hat{S}_{t|\infty}$ is $G_{E\infty}(z) = G_S(z) - G_{SY}(z)G_Y^{-1}(z)G_{YS}(z^{-1})$. Since S_t and N_t are uncorrelated, $G_{SY}(z) = G_S(z)$ and $G_Y(z) = G_S(z) + G_N(z)$, where $G_N(z)$ is the covariance generating function of N_t.Thus,

$$\begin{aligned}G_{E\infty}(z) &= G_S(z)\left[1 - G_Y^{-1}(z)G_S(z^{-1})\right] \qquad (7.8)\\ &= G_S(z)G_N(z)G_Y^{-1}(z)\\ &= \frac{\sigma_b^2}{(1-\rho z)(1-\rho z^{-1})} \times \frac{\sigma_n^2(1-\rho z)(1-\rho z^{-1})}{\sigma_a^2(1-\theta z)(1-\theta z^{-1})}\\ &= \frac{(\rho-\theta)(1-\rho\theta)\sigma_n^2}{\rho} \times \frac{1}{(1-\theta z)(1-\theta z^{-1})}. \qquad (7.9)\end{aligned}$$

◇

7.1.2 Wiener–Kolmogorov Filtering

We are interested in finding a formula for $\hat{S}_{t|t}$, the projection of S_t onto the Hilbert space generated by the finite linear combinations of elements of $\{Y_s : s \le t\}$ and their limits in mean square. As is well known, this projection is unique and has the form

$$\hat{S}_{t|t} = \sum_{j=-\infty}^{t} \Psi_{tj} Y_j,$$

for some set of filter weights, $\{\Psi_{tj}\}$, that are to be determined, where the series is mean square convergent. Assuming that the filter $\Psi_t(z) = \sum_{j=-\infty}^{t} \Psi_{tj} z^j$ is stable, the orthogonality condition yields

$$E\left[\left(S_t - \sum_{j=-\infty}^{t} \Psi_{tj} Y_j\right) Y_l'\right] = 0, \qquad -\infty < l \le t,$$

or

$$\gamma_{SY}(t-l) = \sum_{j=-\infty}^{t} \Psi_{tj}\gamma_Y(j-l), \qquad -\infty < l \leq t. \tag{7.10}$$

Note that, as mentioned in the previous section, (7.10) is justified by the continuity of the inner product $< X, Y >= E(XY')$. Letting $t-l=h$ in (7.10) yields

$$\gamma_{SY}(h) = \sum_{j=-\infty}^{h+l} \Psi_{h+l,j}\gamma_Y(j-l), \qquad h \geq 0. \tag{7.11}$$

Letting in turn $k=j-l$ in (7.11) leads to the equation

$$\gamma_{SY}(h) = \sum_{k=-\infty}^{h} \Psi_{h+l,k+l}\gamma_Y(k), \qquad h \geq 0, \tag{7.12}$$

which should hold for all l. But since the left-hand side of (7.12) does not depend on l, putting $l=-h$ yields

$$\gamma_{SY}(h) = \sum_{k=-\infty}^{h} \Psi_{0,k-h}\gamma_Y(k) = \sum_{j=-\infty}^{0} \Psi_{0,j}\gamma_Y(h+j), \qquad h \geq 0.$$

Since the set of weights $\{\Psi_{tj}\}$ in the orthogonal projection $\hat{S}_{t|t} = \sum_{j=-\infty}^{t} \Psi_{tj}Y_j$ is unique, we conclude that

$$\Psi_{0,j} = \Pi_{-j},$$

for some sequence $\{\Pi_j : j = 0, 1, \ldots\}$, which implies

$$\gamma_{SY}(h) = \sum_{j=0}^{\infty} \Pi_j\gamma_Y(h-j), \qquad h \geq 0, \tag{7.13}$$

and, therefore,

$$\hat{S}_{t|t} = \sum_{j=0}^{\infty} \Pi_j Y_{t-j}.$$

Equation (7.13) is hard to solve because the equality holds only for $h \geq 0$. Otherwise, taking generating functions would give the solution in a simple way. To overcome this difficulty, Wiener & Hopf (1931) used the following clever technique.

Define the sequence $\{g_h\}$ by

$$g_h = \gamma_{SY}(h) - \sum_{j=0}^{\infty} \Pi_j \gamma_Y(h-j), \qquad h \in \mathbb{Z}, \tag{7.14}$$

which by (7.13) is strictly anticausal, that is,

$$g_h = 0, \qquad h \geq 0.$$

Since (7.14) is defined for all $h \in \mathbb{Z}$, we can take generating functions to get

$$G_g(z) = G_{SY}(z) - \Pi(z)G_Y(z), \tag{7.15}$$

where $\Pi(z) = \sum_{j=0}^{\infty} \Pi_j z^j$. Let the covariance function factorization of $G_Y(z)$ be

$$G_Y(z) = \Psi(z)\Sigma\Psi'(z^{-1}).$$

Then, by our assumptions, $\Psi(z)$ and its inverse are analytic in some annulus that contains the unit circle and $\Psi(0) = I$. Postmultiplying both sides of (7.15) by $\Psi'^{-1}(z^{-1})\Sigma^{-1}$, it is obtained that

$$G_g(z)\Psi'^{-1}(z^{-1})\Sigma^{-1} = G_{SY}(z)\Psi'^{-1}(z^{-1})\Sigma^{-1} - \Pi(z)\Psi(z). \tag{7.16}$$

Since $G_g(z)$ and $\Psi'^{-1}(z^{-1})\Sigma^{-1}$ are generating functions of anticausal sequences and $G_g(z) = \sum_{j=-\infty}^{-1} g_j z^j$, its product is the generating function of a strictly anticausal sequence. On the other hand, $\Pi(z)\Psi(z)$ is the generating function of a causal sequence. Before we proceed, we need to introduce some notation.

Let $F(z)$ be an analytic function in some annulus that contains the unit circle and such that $F(z) = \sum_{j=-\infty}^{\infty} f_j z^j$ is its Laurent expansion. We introduce the operators $[.]_+$ and $[.]_-$ that yield the "causal part" and the "strictly anticausal part" of the function to which it is applied. That is,

$$[F]_+ = \sum_{j=0}^{\infty} f_j z^j \quad \text{and} \quad [F(z)]_- = F(z) - [F(z)]_+ \,.$$

Applying the $[.]_+$ operator to both sides of (7.16) yields

$$\left[G_g(z)\Psi'^{-1}(z^{-1})\Sigma^{-1}\right]_+ = \left[G_{SY}(z)\Psi'^{-1}(z^{-1})\Sigma^{-1}\right]_+ - [\Pi(z)\Psi(z)]_+ \,. \tag{7.17}$$

Since the generating function in the left-hand side of (7.17) corresponds to a strictly anticausal sequence, we have

$$\left[G_g(z)\Psi'^{-1}(z^{-1})\Sigma^{-1}\right]_+ = 0.$$

On the other hand, the generating function in the second term on the right-hand side of (7.17) corresponds to a causal sequence. Thus,

$$[\Pi(z)\Psi(z)]_+ = \Pi(z)\Psi(z),$$

and we finally obtain

$$\Pi(z) = \left[G_{SY}(z)\Psi'^{-1}(z^{-1})\Sigma^{-1}\right]_+ \Psi^{-1}(z),$$

where the filter $\Pi(z)$ is stable because it is convergent in some annulus containing the unit circle. We summarize the results obtained so far in the following theorem.

Theorem 7.2 *Under the same assumptions of Theorem 7.1, if* $G_Y(z) = \Psi(z)\Sigma\Psi'(z^{-1})$*, then the estimator* $\hat{S}_{t|t}$ *of* S_t *based on the semi-infinite sample* $\{Y_s : s \leq t\}$ *is given by* $\hat{S}_{t|t} = \sum_{j=0}^{\infty} \Pi_j Y_{t-j}$*, where the filter is stable and its weights,* Π_j*, are given by*

$$\Pi(z) = \sum_{j=0}^{\infty} \Pi_j z^j = \left[G_{SY}(z)\Psi'^{-1}(z^{-1})\right]_+ \Sigma^{-1}\Psi^{-1}(z). \tag{7.18}$$

In addition, the error $E_{t|t} = S_t - \hat{S}_{t|t}$ *is stationary, has a covariance generating function,* $G_{E0}(z)$*, given by* $G_{E0}(z) = G_{E\infty}(z) + \Omega_0(z)\Sigma^{-1}\Omega_0'(z^{-1})$*, where* $G_{E\infty}(z)$ *is that of Theorem 7.1 and* $\Omega_0(z) = \left[G_{SY}(z)\Psi'^{-1}(z^{-1})\right]_-$*, and*

$$\mathrm{MSE}(\hat{S}_{t|t}) = \mathrm{Var}(E_{t|t}) = \frac{1}{2\pi}\int_{-\pi}^{\pi} G_{E0}(e^{-ix})dx.$$

Proof The only things that remain to be proved are the statements about the error and the formula for the mean square error. Since $\hat{S}_{t|t}$ is the result of passing Y_t through the stable linear time invariant filter $\Pi(z)$, by corollary 3.1, $\hat{S}_{t|t}$ is stationary. We can write $E_{t|t}$ as

$$\begin{aligned} E_{t|t} &= (S_t - \hat{S}_{t|\infty}) + (\hat{S}_{t|\infty} - \hat{S}_{t|t}) \\ &= E_{t|\infty} + Z_t, \end{aligned}$$

where $E_{t|\infty} = S_t - \hat{S}_{t|\infty}$, $Z_t = \hat{S}_{t|\infty} - \hat{S}_{t|t}$ and $E_{t|\infty}$ is orthogonal to Z_t. By Theorem 7.1,

$$\hat{S}_{t|\infty} = G_{SY}(z)G_Y^{-1}(z)Y_t.$$

On the other hand, $Z_t = \Omega(B)Y_t$, where the filter $\Omega(z)$ is given by

$$\begin{aligned}\Omega(z) &= G_{SY}(z)G_Y^{-1}(z) - \left[G_{SY}(z)\Psi'^{-1}(z^{-1})\right]_+ \Sigma^{-1}\Psi^{-1}(z) \\ &= \left[G_{SY}(z)\Psi'^{-1}(z^{-1})\right]_- \Sigma^{-1}\Psi^{-1}(z).\end{aligned}$$

By Corollary 3.1, $E_{t|\infty}$ and Z_t are stationary. In addition, by Proposition 3.14, the covariance generating function, $G_Z(z)$, of Z_t is

$$G_Z(z) = \Omega(z)G_Y(z)\Omega(z^{-1}) = \Omega_0(z)\Sigma^{-1}\Omega_0'(z^{-1}),$$

where $\Omega_0(z) = \left[G_{SY}(z)\Psi'^{-1}(z^{-1})\right]_-$, and that of $E_{t|\infty}$ is given in Theorem 7.1. Then, the formula for the covariance generating function of the error follows. By definition of spectrum and (3.46), we get the desired formula for the mean square error. □

Remark 7.2 The assumptions we are making to derive the formula for $\hat{S}_{t|t}$ and its mean square error are stronger than needed. See Gohbergh & Fel´dman (1974) for the general case. ◇

Remark 7.3 So far we have been concerned with finding the estimator $\hat{S}_{t|t}$ of S_t based on $\{Y_s : s \leq t\}$ and its mean square error. Suppose we would like to compute the estimator $\hat{S}_{t+k|t}$ of S_{t+k}, where k is a fixed positive or negative integer, based on the same set $\{Y_s : s \leq t\}$. Then, using the same arguments that led to prove Theorem 7.2, it can be shown that $\hat{S}_{t+k|t} = \sum_{j=0}^{\infty} \Pi_j Y_{t-j}$, where the filter is stable and its weights, Π_j, are given by

$$\Pi(z) = \sum_{j=0}^{\infty} \Pi_j z^j = \left[z^{-k}G_{SY}(z)\Psi'^{-1}(z^{-1})\right]_+ \Sigma^{-1}\Psi^{-1}(z). \tag{7.19}$$

In addition, $E_{t+k|t} = S_t - \hat{S}_{t+k|t}$ is stationary and the covariance generating function, $G_{E,k}(z)$, of $\hat{S}_{t+k|t}$ is given by

$$G_{Ek}(z) = G_{E\infty}(z) + \Omega_k(z)\Sigma^{-1}\Omega_k'(z^{-1}), \tag{7.20}$$

where $\Omega_k(z) = \left[z^{-k}G_{SY}(z)\Psi'^{-1}(z^{-1})\right]_-$. ◇

7.1.3 Polynomial Methods

When the covariance generating functions of $\{S_t\}$ and $\{Y_t\}$ are rational, the use of **polynomial methods** can be very advantageous. The following lemma will be needed in the sequel.

Lemma 7.1 *Let*

$$H(z, z^{-1}) = \frac{\alpha(z)\beta(z^{-1})}{\varphi(z)\theta(z^{-1})},$$

where α, β, φ, *and* θ *are polynomials of degree* a, b, c, *and* d, *respectively, and the roots of* φ *are different from the inverses of the roots of* θ. *Then,*

$$H(z, z^{-1}) = \frac{d(z)}{\varphi(z)} + z^{-1}\frac{c(z^{-1})}{\theta(z^{-1})},$$

where $c(z^{-1}) = c_0 + c_1 z^{-1} + \cdots + c_h z^{-h}$, $h = \max\{b, d\} - 1$, *and* $d(z) = d_0 + d_1 z + \cdots + d_k z^k$, $k = \max\{a, c - 1\}$. *In addition, the decomposition is unique and the coefficients of* $d(z)$ *and* $c(z)$ *can be obtained by equating coefficients in*

$$\alpha(z)\beta(z^{-1}) = d(z)\theta(z^{-1}) + z^{-1}c(z^{-1})\varphi(z).$$

Proof

1. Case $b = d$. If $p(z)$ is a polynomial in z of degree g, let $p'(z)$ denote the polynomial $p(z^{-1}) = z^{-g}p'(z)$. Note that the coefficients of $p'(z)$ are those of $p(z)$ in reversed order and that the roots of $p'(z)$ are the inverses of the roots of $p(z)$. Put

$$H(z, z^{-1}) = \frac{z^{-b}}{z^{-d}}\left[\frac{\alpha(z)\beta'(z)}{\varphi(z)\theta'(z)}\right] = \frac{z^{-b}}{z^{-d}}\left[\rho_1(z) + \frac{\rho_2(z)}{\varphi(z)} + \frac{\rho_3'(z)}{\theta'(z)}\right],$$

where $\deg\{\rho_2\} = c - 1$, $\deg\{\rho_3'\} = d - 1$, and $\rho_1(z)$ is not zero only if $\deg\{\alpha\beta'\} = a + b \geq c + d = \deg\{\varphi\theta'\}$, that is, if $a \geq c$. Then,

$$H(z, z^{-1}) = \frac{d(z)}{\varphi(z)} + z^{-1}\frac{c(z^{-1})}{\theta(z^{-1})},$$

where $d(z) = \rho_1(z)\varphi(z) + \rho_2(z)$ has degree $\max\{a, c - 1\}$ and $c(z^{-1}) = \rho_3(z^{-1})$ has degree $d - 1$.

2. Case $b < d$. We reduce this case to the previous one by defining $\bar{\beta}'(z) = z^{d-b}\beta'(z)$. Then,

$$H(z, z^{-1}) = \frac{z^{-d}}{z^{-d}} \left[\frac{\alpha(z)\bar{\beta}'(z)}{\varphi(z)\theta'(z)} \right] = \frac{z^{-d}}{z^{-d}} \left[\rho_1(z) + \frac{\rho_2(z)}{\varphi(z)} + \frac{\rho_3'(z)}{\theta'(z)} \right],$$

   where $\rho_3(z^{-1})$ has degree $d-1$.
3. Case $b > d$. We reduce this case to the first by defining $\bar{\theta}'(z) = z^{b-d}\theta'(z)$. Then,

$$H(z, z^{-1}) = \frac{z^{-b}}{z^{-b}} \left[\frac{\alpha(z)\bar{\beta}'(z)}{\varphi(z)\bar{\theta}'(z)} \right] = \frac{z^{-b}}{z^{-b}} \left[\rho_1(z) + \frac{\rho_2(z)}{\varphi(z)} + \frac{\rho_3'(z)}{\bar{\theta}'(z)} \right],$$

   $\rho_3(z^{-1})$ has degree $b-1$.

The uniqueness of the decomposition is a consequence of the uniqueness of the partial fraction expansion. □

Example 7.1 (Continued) Suppose the signal-plus-noise model of Example 7.1, $Y_t = S_t + N_t$. The filter to obtain $\hat{S}_{t|t}$ is

$$\Pi(z) = \left[\frac{G_{SY}(z)}{\sigma_a^2 \Psi(z^{-1})} \right]_+ \frac{1}{\Psi(z)},$$

where $G_{SY}(z) = G_S(z)$ and $\Psi(z) = (1-\theta z)/(1-\rho z)$. Letting $\theta(z) = 1 - \theta z$ and $\phi(z) = 1 - \rho z$, we have

$$\begin{aligned} \Pi(z) &= \left[\frac{\sigma_b^2 \phi(z^{-1})}{\sigma_a^2 \phi(z)\phi(z^{-1})\theta(z^{-1})} \right]_+ \frac{\phi(z)}{\theta(z)} \\ &= \frac{\sigma_b^2}{\sigma_a^2} \left[\frac{1}{\phi(z)\theta(z^{-1})} \right]_+ \frac{\phi(z)}{\theta(z)}. \end{aligned} \tag{7.21}$$

Using Lemma 7.1, we decompose the term inside the brackets in (7.21) as

$$\frac{1}{\phi(z)\theta(z^{-1})} = \frac{d}{\phi(z)} + z^{-1} \frac{c}{\theta(z^{-1})},$$

where $d = 1/(1-\rho\theta)$ and $c = \theta/(1-\rho\theta)$. Then, using (7.7), we get

$$\begin{aligned} \Pi(z) &= \frac{\sigma_b^2}{\sigma_a^2} \frac{d\phi(z)}{\phi(z)\theta(z)} \\ &= \frac{\rho - \theta}{\rho} \times \frac{1}{1 - \theta z}. \end{aligned}$$

The error $E_{t|t} = S_t - \hat{S}_{t|t}$ has a covariance generating function, $G_{E0}(z)$, given by $G_{E0}(z) = G_{E\infty}(z) + \Omega_0(z)\sigma_a^{-2}\Omega_0'(z^{-1})$, where $G_{E\infty}(z)$ is given by (7.9) and

$$\Omega_0(z) = \left[G_{SY}(z)\Psi'^{-1}(z^{-1})\right]_-$$
$$= z^{-1}\frac{\sigma_b^2 c}{\theta(z^{-1})}.$$

Thus,

$$\Omega_0(z)\sigma_a^{-2}\Omega_0'(z^{-1}) = \frac{\sigma_b^2}{\sigma_a^2}z^{-1}\frac{c}{\theta(z^{-1})}z\frac{c}{\theta(z)}$$
$$= \frac{\theta^2(\rho-\theta)}{\rho(1-\rho\theta)} \times \frac{1}{(1-\theta z)(1-\theta z^{-1})}.$$

Finally, using this last expression and (7.9) we get

$$G_{E0}(z) = \frac{(\rho-\theta)(1-\rho\theta)\sigma_n^2}{\rho} \times \frac{1}{(1-\theta z)(1-\theta z^{-1})} + \frac{\theta^2(\rho-\theta)}{\rho(1-\rho\theta)}$$
$$\times \frac{1}{(1-\theta z)(1-\theta z^{-1})}.$$

◇

7.1.4 Prediction Based on the Semi-infinite Sample

Consider the special case

$$S_t = Y_{t+k},$$

where k is a positive integer, and assume that $\{Y_t\}$ follows the linear model

$$Y_t = \sum_{j=0}^{\infty} \Psi_j A_{t-j},$$

where $\{A_t\}$ $\sim$WN$(0, \Sigma)$ and $\Psi(z) = \sum_{j=0}^{\infty} \Psi_j z^j$ is analytic in some annulus containing the unit circle, $r^{-1} < |z| < r$ with $r > 1$. In this case, the Wiener–Kolmogorov estimator $\hat{S}_{t|t}$ coincides with the predictor, $\hat{Y}_{t+k|t}$, of Y_{t+k} based on the semi-infinite sample $\{Y_s : s \leq t\}$. Notice that $G_{SY}(z) = z^{-k}G_Y(z) =$

$z^{-k}\Psi(z)\Sigma\Psi'(z^{-1})$. Then, the filter, $\Pi(z)$, for the predictor becomes

$$\Pi(z) = \left[G_{SY}(z)\Psi'^{-1}(z^{-1})\right]_+ \Sigma^{-1}\Psi^{-1}(z) = \left[z^{-k}\Psi(z)\right]_+ \Psi^{-1}(z).$$

Since $\Psi(z) = I + \sum_{j=1}^{\infty} \Psi_j z^j$, we have

$$\left[z^{-k}\Psi(z)\right]_+ = \sum_{j=k}^{\infty} \Psi_j z^{j-k} = z^{-k}\Psi(z) - z^{-k}\left[I + \sum_{j=1}^{k-1} \Psi_j z^j\right],$$

and

$$\Pi(z) = z^{-k}\left[I - (I + \sum_{j=1}^{k-1} \Psi_j z^j)\Psi^{-1}(z)\right]. \tag{7.22}$$

Thus, the predictor is

$$\hat{Y}_{t+k|t} = Y_{t+k} - A_{t+k} - \sum_{j=1}^{k-1} \Psi_j A_{t+k-j}.$$

By Theorem 7.2, the error $E_t = Y_{t+k} - \hat{Y}_{t+k|t}$ is stationary and has a covariance generating function, $G_E(z)$, given by

$$G_E(z) = G_S(z) - G_{SY}(z)G_Y^{-1}(z)G_{YS}(z^{-1}) + \Omega_0(z)\Sigma^{-1}\Omega_0'(z^{-1})$$

where $\Omega_0(z) = \left[G_{SY}(z)\Psi'^{-1}(z^{-1})\right]_-$, $G_S(z) = G_Y(z), G_{SY}(z) = z^{-k}G_Y(z)$ and

$$\left[G_{SY}(z)\Psi'^{-1}(z^{-1})\right]_- = \left[z^{-k}\Psi(z)\Sigma\right]_- = z^{-k}\left[I + \sum_{j=1}^{k-1} \Psi_j z^j\right]\Sigma.$$

Thus,

$$G_E(z) = \left[I + \sum_{j=1}^{k-1} \Psi_j z^j\right]\Sigma\left[I + \sum_{j=1}^{k-1} \Psi_j z^{-j}\right]'$$

and

$$\mathrm{MSE}(\hat{Y}_{t|t+k}) = \Sigma + \sum_{j=1}^{k-1} \Psi_j \Sigma \Psi_j'. \tag{7.23}$$

Remark 7.4 As in the scalar case, the elements $x_{t+1,j}$ of the state space representation (6.77) and (6.78) of a vector ARMA(p, q) process coincide with the predictors $\hat{Y}_{t+j|t}$, $j = 1, 2, \ldots, r$. To see this, use the state space equations to deduce that $x_{t+1,j} = Y_{t+j} - A_{t+j} - \sum_{h=1}^{j-1} \Psi_h A_{t+j-h}$. ◇

Example 7.2 Suppose that $\{Y_t\}$ follows the AR(1) model

$$Y_t + \phi Y_{t-1} = A_t,$$

where $\{A_t\} \sim$WN$(0, \sigma^2)$ and ϕ is a real number such that $|\phi| < 1$. Since $\Psi(z) = 1/(1+\phi z) = 1 + \sum_{j=1}^{\infty}(-\phi)^j z^j$, the filter (7.22) for the predictor $\hat{Y}_{t+k|t}$ of Y_{t+k} is

$$\begin{aligned}\Pi(z) &= z^{-k}\left[1 - \left(1 + \sum_{j=1}^{k-1}(-\phi)^j z^j\right)(1+\phi z)\right] \\ &= z^{-k}\left[1 - \frac{1-(-\phi)^k z^k}{1+\phi z}(1+\phi z)\right] \\ &= (-\phi)^k.\end{aligned}$$

Thus, the predictor is

$$\hat{Y}_{t+k|t} = (-\phi)^k Y_t.$$

From (7.23), the mean square error is

$$\text{MSE}(\hat{Y}_{t+k|t}) = \sigma_a^2\left[1 + \sum_{j=1}^{k-1}|\phi|^{2j}\right]$$

◇

7.1.5 *Innovations Approach*

In this approach, the estimation problem is broken into two parts: (1) finding the innovations $\{A_t\}$ from the observations $\{Y_t\}$, and (2) finding the estimator, $\hat{S}_{t|\infty}$ or $\hat{S}_{t|t}$, of S_t from the innovations. Since the innovations are uncorrelated, we may expect that the estimation problem will be easy.

The innovations approach was originally used by Kolmogorov (1939), as opposed to Wiener (1949) who worked in the frequency domain.

To gain some insight into the problem, suppose, for example, that we are interested in finding

$$\hat{S}_{t|t} = \sum_{j=0}^{\infty} \Delta_j A_{t-j}. \tag{7.24}$$

Then, the orthogonality condition gives $E\left[(S_t - \hat{S}_{t|t})A_j\right] = 0$, $j \leq t$, and, since $\{S_t\}$ is jointly stationary with $\{A_t\}$, we can define the cross-autocovariances $\gamma_{SA}(h)$ as $\gamma_{SA}(h) = E(S_{t+h}A_t)$. Using these autocovariances, the orthogonality condition becomes $\gamma_{SA}(h) = \Delta_h \Sigma$. It follows from this, that the weights Δ_j in (7.24) are given by

$$\Delta_h = \begin{cases} \gamma_{SA}(h)\Sigma^{-1} & \text{for } h \geq 0 \\ 0 & \text{for } h < 0 \end{cases} \tag{7.25}$$

Of course, we could have anticipated this result, since the expression (7.24), by Proposition 3.5 (assuming the filter $\Delta(z) = \sum_{j=0}^{\infty} \Delta_j z^j$ is stable), is equal to the following limit with probability one

$$\begin{aligned} \hat{S}_{t|t} &= \lim_{n\to\infty} E^*(S_t|A_t, A_{t-1}, \ldots, A_{t-n}) \\ &= \lim_{n\to\infty} \sum_{j=0}^{n} \mathrm{Cov}(S_t, A_{t-j})\mathrm{Var}^{-1}(A_{t-j})A_{t-j} \\ &= \lim_{n\to\infty} \sum_{j=0}^{n} \gamma_{SA}(j)\Sigma^{-1}A_{t-j}. \end{aligned}$$

Given our assumptions on the process $\{Y_t\}$, the filter $\Psi(z)$ in the factorization $G_Y(z) = \Psi(z)\Sigma\Psi'(z^{-1})$ has an analytic inverse $\Psi^{-1}(z)$. If $A(z) = \sum_{j=-\infty}^{\infty} A_j z^j$ is the generating function of the innovations, $\{A_t\}$, and $Y(z) = \sum_{j=-\infty}^{\infty} Y_j z^j$ is the generating function of the observations, $\{Y_t\}$, then $Y(z) = \Psi(z)A(z)$ and $A(z) = \Psi^{-1}(z)Y(z)$. This, together with Lemma 3.14, implies that the cross-covariance generating function $G_{SA}(z) = \sum_{j=-\infty}^{\infty} \gamma_{SA}(j)z^j$ can be expressed as

$$G_{SA}(z) = G_{SY}(z)\Psi'^{-1}(z^{-1}). \tag{7.26}$$

Letting $\Delta(z) = \sum_{j=0}^{\infty} \Delta_j z^j$ and using (7.25), we get $\Delta(z) = [G_{SA}(z)]_+ \Sigma^{-1}$. Comparing this expression with (7.26), it is obtained that $\Delta(z) = \left[G_{SY}(z)\Psi'^{-1}(z^{-1})\right]_+ \Sigma^{-1}$. Letting $\hat{S}_{t|t} = \Omega(z)Y_t$, it follows that

$$\Omega(z) = \Delta(z)\Psi^{-1}(z) = \left[G_{SY}(z)\Psi'^{-1}(z^{-1})\right]_+ \Sigma^{-1}\Psi^{-1}(z),$$

which coincides with (7.18) in Theorem 7.2.

7.2 Wiener–Kolmogorov Filtering and Smoothing for Stationary State Space Models

The material in this section follows Gómez (2006) closely. Some of the results are new in the literature, like the ones regarding the mean squared errors for smoothing based on the doubly infinite sample, its extension to the nonstationary case, or the computation of the filter weights.

Let $\{Y_t : t \in \mathbb{Z}\}$ be a zero-mean multivariate stationary process with Wold decomposition $Y_t = \sum_{j=0}^{\infty} \Psi_j A_{t-j}$ and covariance factorization $G_Y(z) = \Psi(z)\Sigma\Psi'(z^{-1})$, where $\mathrm{Var}(A_t) = \Sigma$ is positive definite, denoted $\Sigma > 0$. Suppose that $\{Y_t\}$ admits a state space representation

$$x_{t+1} = Fx_t + Gu_t, \tag{7.27}$$

$$Y_t = Hx_t + v_t, \qquad t > -\infty \tag{7.28}$$

where

$$E\left\{\begin{bmatrix} u_t \\ v_t \end{bmatrix}\begin{bmatrix} u_s', & v_s' \end{bmatrix}\right\} = \begin{bmatrix} Q & S \\ S' & R \end{bmatrix}\delta_{ts}. \tag{7.29}$$

By Lemma 5.2, a necessary and sufficient condition for x_t and Y_t in (7.27) and (7.28) to be jointly stationary is that $\mathrm{Var}(x_t) = \Pi_t$ be time invariant, $\Pi_t = \Pi$. In this case, Π satisfies the Lyapunov equation

$$\Pi = F\Pi F' + GQG'. \tag{7.30}$$

By Corollary 5.1, a sufficient condition for a unique positive semidefinite Π to exist satisfying (7.30) is that F be stable (i.e., has all its eigenvalues inside the unit circle). We assume from now on until further notice that F is stable.

Let $\hat{x}_{t+1|t}$ and $\hat{x}_{t|\infty}$ be the estimators of x_{t+1} and x_t based on $\{Y_s : s \leq t\}$ and $\{Y_s : s = 0, \pm 1, \pm 2, \ldots\}$, respectively, and let $\hat{x}_{+1|}(z)$ and $\hat{x}_{|\infty}(z)$ be their generating functions. Then, from (7.5) and (7.19), we can write

$$\hat{x}_{+1|}(z) = \left[z^{-1}G_{XY}(z)\Psi'^{-1}(z^{-1})\right]_+ \Sigma^{-1}\Psi^{-1}(z)Y(z), \tag{7.31}$$

and

$$\hat{x}_{|\infty}(z) = G_{XY}(z)\Psi'^{-1}(z^{-1})\Sigma^{-1}\Psi^{-1}(z)Y(z), \tag{7.32}$$

where $G_{XY}(z)$ is the cross-covariance function of $\{x_t\}$ and $\{Y_t\}$, $Y(z)$ is the generating function of $\{Y_t\}$, and $G_Y(z) = \Psi(z)\Sigma\Psi'(z^{-1})$ is the covariance factorization of $\{Y_t\}$.

Formulae (7.31) and (7.32) show that the covariance factorization of $\{Y_t\}$ plays a central role in Wiener–Kolmogorov prediction and smoothing. To make some progress, we need to express the covariance factorization of $\{Y_t\}$ in terms of the matrices in (7.27), (7.28), and (7.29).

To this end, we proceed as in Sect. 5.6 and apply generating functions in (7.27) and (7.28) to get

$$G_Y(z) = [zH(I - Fz)^{-1} \quad I] \begin{bmatrix} GQG' & GS \\ S'G' & R \end{bmatrix} \begin{bmatrix} z^{-1}(I - F'z^{-1})^{-1}H' \\ I \end{bmatrix}. \tag{7.33}$$

Clearly, if we can find matrices Σ and K such that (7.33) factorizes as

$$\begin{aligned} G_Y(z) &= [zH(I - Fz)^{-1} \quad I] \begin{bmatrix} K \\ I \end{bmatrix} \Sigma\, [K' \quad I] \begin{bmatrix} z^{-1}(I - F'z^{-1})^{-1}H' \\ I \end{bmatrix} \\ &= [I + zH(I - Fz)^{-1}K]\Sigma\, [I + z^{-1}K'(I - F'z^{-1})^{-1}H'], \end{aligned} \tag{7.34}$$

then we can obtain the desired covariance factorization by defining $\Psi(z) = I + zH(I - Fz)^{-1}K$. By Lemma 4.1 (matrix inversion lemma), the inverse of $\Psi(z)$, if it exists, is given by

$$\Psi^{-1}(z) = I - zH(I - F_p z)^{-1}K, \tag{7.35}$$

where $F_p = F - KH$. It is not difficult to see that the inverse exists if, and only if, F_p is stable.

By Lemma 5.6, it seems natural to try to select a matrix P such that

$$\begin{bmatrix} -P + FPF' + GQG' & FPH' + GS \\ HPF' + S'G' & R + HPH' \end{bmatrix} = \begin{bmatrix} K \\ I \end{bmatrix} \Sigma\, [K' \quad I],$$

and $F_p = F - KH$ is stable. Equating terms in the previous expression, it is seen that the matrices K and Σ should satisfy $K = (FPH' + GS)\Sigma^{-1}$ and $\Sigma = R + HPH'$, and that the desired matrix P should satisfy the DARE

$$-P + FPF' + GQG' - (FPH' + GS)(R + HPH')^{-1}(FPH' + GS)' = 0. \tag{7.36}$$

Under some rather general assumptions specified in Sect. 5.12.1, there exists a unique positive semidefinite solution, P, of the DARE (7.36) such that Σ is nonsingular and F_p is stable. As noted in Sect. 5.6, such a solution is called a **stabilizing solution**. Also as in Sect. 5.6, if M is a real symmetric matrix, we will sometimes use in the following the notation $M \geq 0$ to indicate that M is positive semidefinite.

In the rest of the section, we will make the following basic assumption.

Assumption 7.1 *We assume in the state space model (7.27) and (7.28) that* (F, H) *is detectable and* $(F^s,\ GQ^{s/2})$ *is stabilizable, where these terms are defined in*

Sect. 5.12. Thus, by the results of Sect. 5.12.1, the DARE (7.36) has a unique stabilizing solution $P \geq 0$.

The following theorem is a direct consequence of Lemma 5.6 and Assumption 7.1.

Theorem 7.3 *Let (7.33) be the covariance generating function of a process* $\{Y_t\}$ *generated by the state space model (7.27) and (7.28). Then, under Assumption 7.1, the covariance factorization* $G_Y(z) = \Psi(z)\Sigma\Psi'(z^{-1})$ *obtains with* $\Psi(z) = I + zH(I - Fz)^{-1}K$*, where* $K = (FPH' + GS)\Sigma^{-1}$, $\Sigma = R + HPH' > 0$ *and P is the unique stabilizing solution of the DARE (7.36).*

The previous theorem is known in the literature (Anderson & Moore, 2012). However, the proof is usually based on iterating in the Kalman filter recursions, whereas the proof we have presented is algebraic in nature. Our procedure will allow for the development of recursive formulae for Wiener–Kolmogorov filtering and smoothing in the next section without using the Kalman filter. As we will see in Sect. 7.2.7, these last recursions coincide with those obtained by iterating in the Kalman filtering and smoothing recursions. This will prove the equivalence of both approaches.

By Theorem 7.3, if we let $Y(z)$ and $A(z)$ be the generating functions of the observations $\{Y_t\}$ and the innovations $\{A_t\}$ and we define

$$\begin{aligned}\theta(z) &= \sum_{j=-\infty}^{\infty} \theta_j z^j \\ &= z(I - zF)^{-1}KA(z),\end{aligned}$$

we get $Y(z) = \Psi(z)A(z) = H\theta(z) + A(z)$ and $z^{-1}\theta(z) = F\theta(z) + KA(z)$. These relations imply the following state space representation

$$\begin{aligned}\theta_{t+1} &= F\theta_t + KA_t, \\ Y_t &= H\theta_t + A_t, \qquad t > -\infty,\end{aligned} \tag{7.37}$$

where $\text{Var}(A_t) = \Sigma$. It is to be noticed that the expression

$$Y_t = \Psi(B)A_t = (I + BH(I - FB)^{-1}K)A_t = \sum_{j=0}^{\infty} \Psi_j A_{t-j}$$

where B is the backshift operator, $BY_t = Y_{t-1}$, constitutes the Wold representation of $\{Y_t\}$. For this reason, the representation (7.37) is called the **innovations representation**.

By Proposition 3.7, the process $\{Y_t\}$ in (7.37) also admits a VARMA representation, $\Phi(B)Y_t = \Theta(B)A_t$.

7.2.1 Recursive Wiener–Kolmogorov Filtering and Smoothing

In this section, we will first prove that the state space model (7.37) we obtained from (7.27) and (7.28) under Assumption 7.1 constitutes a recursive formula for the Wiener–Kolmogorov one-period-ahead forecast based on the infinitely remote past. Then, we will obtain some recursions for other Wiener–Kolmogorov prediction and smoothing problems. All this will be achieved by purely algebraic means, without iterating in the Kalman filter recursions. In fact, we will see in Sect. 7.2.7 that, under Assumption 7.1, the same recursions can be obtained by iterating in the Kalman filter and smoothing recursions. This will establish the equivalence between the two approaches.

Using (7.26), we can write (7.31) as

$$\hat{x}_{+1|}(z) = \left[z^{-1}G_{XA}(z)\right]_{+} \Sigma^{-1}A(z), \tag{7.38}$$

where $G_{XA}(z)$ is the cross-covariance function of $\{x_t\}$ and the innovations $\{A_t\}$.

To proceed further, we need to decompose $G_{XA}(z)$ in (7.38) in such a way that it becomes easy to select the terms with nonnegative powers of z. The following lemma provides this decomposition.

Lemma 7.2 *The covariance generating function $G_{XA}(z)$ can be decomposed as*

$$G_{XA}(z) = z(I - Fz)^{-1}(FPH' + GS) + PH' + z^{-1}PF_p'(I - F_p'z^{-1})^{-1}H',$$

where P is the unique positive semidefinite solution of the DARE (7.36), $F_p = F - KH$, $K = (FPH' + GS)\Sigma^{-1}$ and $\Sigma = R + HPH'$.

Proof Using (7.35), we can write $A(z) = \Psi^{-1}(z)Y(z)$ as

$$\begin{aligned}
A(z) &= \left[I - zH(I - F_pz)^{-1}K\right]Y(z) \\
&= \left[I - zH(I - F_pz)^{-1}K\right][HX(z) + V(z)] \\
&= H\left[I - z(I - F_pz)^{-1}KH\right]X(z) + \left[I - zH(I - F_pz)^{-1}K\right]V(z) \\
&= H(I - F_pz)^{-1}\left[I - F_pz - zKH\right]X(z) + \left[I - zH(I - F_pz)^{-1}K\right]V(z) \\
&= H(I - F_pz)^{-1}[I - Fz]\,z(I - Fz)^{-1}GU(z) + \left[I - zH(I - F_pz)^{-1}K\right]V(z) \\
&= zH(I - F_pz)^{-1}GU(z) + \left[I - zH(I - F_pz)^{-1}K\right]V(z),
\end{aligned}$$

where $U(z)$ and $V(z)$ are the generating functions of $\{u_t\}$ and $\{v_t\}$, respectively. Since $X(z) = z(I - Fz)^{-1}GU(z)$, by Lemma 3.14, the cross-covariance generating function $G_{XA}(z)$ can be expressed as $G_{XA}(z) = z(I - Fz)^{-1}GG_{UA}(z)$. It follows from this, using Lemma 3.14 again, that

$$\begin{aligned} G_{XA}(z) &= (I - Fz)^{-1}G\left[QG'(I - F_p'z^{-1})^{-1}H' - SK'(I - F_p'z^{-1})^{-1}H'\right] \\ &\quad + z(I - Fz)^{-1}GS \\ &= (I - Fz)^{-1}(GQG' - GSK')(I - F_p'z^{-1})^{-1}H' + z(I - Fz)^{-1}GS. \end{aligned}$$

Since the DARE can be expressed as $P = FPF_p' + GQG' - GSK'$, substituting in the previous expression, it is obtained that

$$\begin{aligned} G_{XA}(z) &= z(I - Fz)^{-1}(P - FPF_p')(I - F_p'z^{-1})^{-1}H'z^{-1} + z(I - Fz)^{-1}GS \\ &= z(I - Fz)^{-1}\left[P(I - F_p'z^{-1}) + (I - zF)P - (I - zF)P(I - F_p'z^{-1})\right] \\ &\quad \times(I - F_p'z^{-1})^{-1}H'z^{-1} + z(I - Fz)^{-1}GS \\ &= PH' + z(I - Fz)^{-1}FPH' + PF_p'(I - F_p'z^{-1})H'z^{-1} + z(I - Fz)^{-1}GS \\ &= z(I - Fz)^{-1}(FPH' + GS) + PH' + z^{-1}PF_p'(I - F_p'z^{-1})H'. \end{aligned}$$

□

It follows from Lemma 7.2 that $\left[z^{-1}G_{XA}(z)\right]_+ = (I - Fz)^{-1}(FPH' + GS)$, which together with (7.38) yields

$$\hat{x}_{+1|}(z) = (I - Fz)^{-1}KA(z). \tag{7.39}$$

With the help of (7.39), it is possible to derive a recursion for the one-period-ahead predictor $\hat{x}_{t+1|t}$. The following theorem gives the details.

Theorem 7.4 *The one-period-ahead predictor $\hat{x}_{t+1|t}$ based on the semi-infinite sample $\{Y_s : s \leq t\}$ satisfies the recursions*

$$\begin{aligned} \hat{x}_{t+1|t} &= F\hat{x}_{t|t-1} + KA_t, \\ Y_t &= H\hat{x}_{t|t-1} + A_t. \qquad t > -\infty. \end{aligned} \tag{7.40}$$

Moreover, the unique positive semidefinite solution P of the DARE (7.36) is the MSE of $\hat{x}_{t|t-1}$, that is, $P = \mathrm{Var}(x_t - \hat{x}_{t|t-1})$, the process $\{\hat{x}_{t|t-1}\}$ is stationary, and if $\overline{\Sigma} = \mathrm{Var}(\hat{x}_{t|t-1})$, then the matrix $\overline{\Sigma}$ is the unique positive semidefinite solution of the Lyapunov equation

$$\overline{\Sigma} = F\overline{\Sigma}F' + K\Sigma K'. \tag{7.41}$$

Proof The generating function of $\{\theta_t\}$ in (7.37) coincides with (7.39). Given that the predictor $\hat{x}_{t+1|t}$ is unique, this shows that $\hat{x}_{t+1|t} = \theta_{t+1}$. Since the process $\{\hat{x}_{t|t-1}\}$

satisfies (7.40) and the matrix F is stable, it is stationary. It follows from this that the matrix $\overline{\Sigma} = \mathrm{Var}(\hat{x}_{t|t-1})$ satisfies the Lyapunov equation (7.41).

Let $P \geq 0$ be the unique solution of the DARE (7.36). To prove that P is the MSE of $\hat{x}_{t|t-1}$, note that, by (7.20), the error $x_t - \hat{x}_{t|t-1}$ is stationary. Letting $\overline{P} = \mathrm{MSE}(\hat{x}_{t|t-1})$, replacing $\overline{\Sigma}$ with $\Pi - \overline{P}$ in (7.41) and considering that P satisfies the DARE (7.36), it is obtained that

$$\begin{aligned}\Pi - \overline{P} &= F(\Pi - \overline{P})F' + K\Sigma K' = \Pi - GQG' - F\overline{P}F' + K\Sigma K' \\ &= \Pi - F\overline{P}F' - P + FPF'.\end{aligned}$$

From this we get $P - \overline{P} = F(P - \overline{P})F'$. Since F is stable, the lemma is proved. □

Using the second equation in (7.40), we get $A_t = Y_t - H\hat{x}_{t|t-1}$. If we substitute this expression in the first equation in (7.40), we obtain the following corollary.

Corollary 7.1 *The filter for the one-period-ahead predictor* $\hat{x}_{t+1|t}$ *is*

$$\hat{x}_{t+1|t} = (F - KH)\hat{x}_{t|t-1} + KY_t = F_p\hat{x}_{t|t-1} + KY_t. \tag{7.42}$$

The following theorem gives a recursion for the fixed interval smoother. More specifically, it provides a recursion for the estimator $\hat{x}_{t|\infty}$ of x_t based on the doubly infinite sample $\{Y_s : s = 0, \pm 1, \pm 2, \ldots\}$.

Theorem 7.5 *The estimator* $\hat{x}_{t|\infty}$ *of* x_t *based on the doubly infinite sample* $\{Y_s : s = 0, \pm 1, \pm 2, \ldots\}$ *satisfies the recursions*

$$\hat{x}_{t|\infty} = \hat{x}_{t|t-1} + P\lambda_{t|\infty}, \tag{7.43}$$

where the adjoint process $\{\lambda_{t|\infty}\}$ *is stationary, satisfies the backward recursion*

$$\lambda_{t|\infty} = F_p'\lambda_{t+1|\infty} + H'\Sigma^{-1}A_t, \tag{7.44}$$

and its covariance matrix, $\Lambda_{|\infty} = \mathrm{Var}(\lambda_{t|\infty})$*, is the unique positive semidefinite solution of the Lyapunov equation*

$$\Lambda_{|\infty} = F_p'\Lambda_{|\infty}F_p + H'\Sigma^{-1}H. \tag{7.45}$$

Moreover, $\mathrm{MSE}(\hat{x}_{t|\infty}) = P - P\Lambda_{|\infty}P'$.

Proof By (7.5) and (7.26), $\hat{x}_{|\infty}(z) = G_{XA}(z)\Sigma^{-1}A(z)$. Then, using Lemma 7.2, we get

$$\hat{x}_{|\infty}(z) = z\hat{x}_{+1|}(z) + P\lambda_{|\infty}(z), \tag{7.46}$$

where $\lambda_{|\infty}(z) = (I - F_p'z^{-1})^{-1}H'\Sigma^{-1}A(z)$. Let $\{\lambda_{t|\infty}\}$ be the so-called adjoint process associated with $\lambda_{|\infty}(z)$. Then, it follows from the definition of $\{\lambda_{t|\infty}\}$

that the backward recursion (7.44) holds and that $\{\lambda_{t|\infty}\}$ is stationary because F_p' is stable. Also, its covariance matrix, $\Lambda_{|\infty} = \text{Var}(\lambda_{t|\infty})$, is the unique positive semidefinite solution of the Lyapunov equation (7.45).

The recursion (7.43) follows from (7.46) and (7.44). To prove the formula for the MSE, note that $x_t - \hat{x}_{t|t-1} = x_t - \hat{x}_{t|\infty} + P\lambda_{t|\infty}$, where $x_t - \hat{x}_{t|\infty}$ and $\lambda_{t|\infty}$ are uncorrelated. Then, taking expectations yields the result. □

The recursions (7.44) and (7.43) are known as the **Bryson–Frazier formulae** (Bryson & Frazier, 1963) for the **fixed–interval smoother** based on the doubly infinite sample.

For any integer k, let $\hat{x}_{t+k|t}$ be the estimator of x_{t+k} based on $\{Y_s : s \leq t\}$ and let $\hat{x}_{+k|}(z)$ be its generating functions. If $k = 0$, we get the filtering problem. If $k > 0$, it is a prediction, and if $k < 0$, it is a smoothing problem, both based on the semi-infinite sample.

As regards prediction, the following theorem holds.

Theorem 7.6 *The predictor $\hat{x}_{t+k|t}$ of x_{t+k} based on the semi-infinite sample $\{Y_s : s \leq t\}$, where $k \geq 2$, satisfies the recursion $\hat{x}_{t+k|t} = F\hat{x}_{t+k-1|t}$. Moreover, its MSE is given by the recursion $\text{MSE}(\hat{x}_{t+k|t}) = F\left[\text{MSE}(\hat{x}_{t+k-1|t})\right]F' + GQG'$, initialized with $\text{MSE}(\hat{x}_{t+2|t}) = FPF' + GQG'$.*

Proof Suppose $k > 0$. Then, by (7.19) and (7.26), $\hat{x}_{+k|}(z) = \left[z^{-k}G_{XA}(z)\right]_+ \Sigma^{-1}A(z)$. Then, using Lemma 7.2, we can write

$$
\begin{aligned}
\hat{x}_{+k|}(z) &= \left[z^{-k+1}(I - Fz)^{-1}(FPH' + GS)\right]_+ \Sigma^{-1}A(z) \\
&= \left[z^{-k+1}(I + Fz + F^2z^2 + \cdots + F^{k-1}z^{k-1} + F^kz^k + \cdots)\right]_+ KA(z) \\
&= F^{k-1}(I - Fz)^{-1}KA(z) \\
&= F^{k-1}\hat{x}_{+1|}(z).
\end{aligned}
$$

To prove the recursion for the MSE, note that $x_{t+k} - \hat{x}_{t+k|t} = F(x_{t+k-1} - \hat{x}_{t+k-1|t}) + Gu_{t+k-1}$ and that the last two terms are uncorrelated. In addition, $\text{MSE}(\hat{x}_{t+1|t}) = P$. □

As far as smoothing is concerned, we have the following theorem.

Theorem 7.7 *The estimator $\hat{x}_{t|t+m}$ of x_t based on the semi-infinite sample $\{Y_s : s \leq t + m\}$, where $m \geq 0$, satisfies the relation*

$$\hat{x}_{t|t+m} = \hat{x}_{t|t-1} + P\lambda_{t|t+m}, \tag{7.47}$$

where $\lambda_{t|t+m}$ is given by the recursion

$$\lambda_{s|t+m} = F_p'\lambda_{s+1|t+m} + H'\Sigma^{-1}A_s, \quad s = t+m, \ldots, t, \tag{7.48}$$

initialized with $\lambda_{t+m+1|t+m} = 0$. *Moreover,* $\text{MSE}(\hat{x}_{t|t+m}) = P - P\Lambda_{t|t+m}P'$, *where* $\Lambda_{t|t+m} = \text{Var}(\lambda_{t|t+m})$, *is obtained from the recursion*

$$\Lambda_{s|t+m} = F_p'\Lambda_{s+1|t+m}F_p + H'\Sigma^{-1}H, \quad s = t+m,\ldots,t, \tag{7.49}$$

initialized with $\Lambda_{t+m+1|t+m} = 0$.

Proof Suppose $k \leq 0$, $k = -m$ with $m \geq 0$. Then, by Lemma 7.2, it is obtained that $\hat{x}_{-m|}(z) = z^{m+1}(I - Fz)^{-1}KA(z) + z^mP\left[\sum_{j=0}^{m}(F_p'z^{-1})^j\right]H'\Sigma^{-1}A(z)$. Multiplying the previous expression by z^{-m} yields

$$\hat{x}_{t|t+m} = \hat{x}_{t|t-1} + P\left[\sum_{j=0}^{m}(F_p')^jz^{-j}\right]H'\Sigma^{-1}A_t = \hat{x}_{t|t-1} + P\left[\sum_{j=0}^{m}(F_p')^jH'\Sigma^{-1}A_{t+j}\right]. \tag{7.50}$$

To compute the sum in (7.50), the backward recursion (7.48) is defined. From this, the recursion (7.49) is straightforward. To prove that $\text{MSE}(\hat{x}_{t|t+m}) = P - P\Lambda_{t|t+m}P'$, note that $x_t - \hat{x}_{t|t-1} = x_t - \hat{x}_{t|t+m} + P\lambda_{t|t+m}$, where $x_t - \hat{x}_{t|t+m}$ and $\lambda_{t|t+m}$ are uncorrelated. Then, taking expectations yields the result. □

Since $\lambda_{t|t+m} - \lambda_{t|t+m-1} = F_p'^{m}H'\Sigma^{-1}A_{t+m}$, the following so-called **fixed-point smoother** recursions are an immediate consequence of Theorem 7.7.

Corollary 7.2 *For t fixed and increasing* $m \geq 0$, *the estimator* $\hat{x}_{t|t+m}$ *of* x_t *based on the semi-infinite sample* $\{Y_s : s \leq t+m\}$ *satisfies the recursions*

$$\hat{x}_{t|t+m} = \hat{x}_{t|t+m-1} + PF_p'^{m}H'\Sigma^{-1}A_{t+m},$$

$$\text{MSE}(\hat{x}_{t|t+m}) = \text{MSE}(\hat{x}_{t|t+m-1}) - PF_p'^{m}H'\Sigma^{-1}HF_p^mP.$$

Remark 7.5 It follows from Theorems 7.5 and 7.7 that $\hat{x}_{t|\infty} = \hat{x}_{t|t+m} + P(\lambda_{t|\infty} - \lambda_{t|t+m})$. This proves that $P(\lambda_{t|\infty} - \lambda_{t|t+m})$ is the **revision** from $\hat{x}_{t|t+m}$ to $\hat{x}_{t|\infty}$. ◇

7.2.2 Covariance Generating Function of the Process

Under Assumption 7.1, the covariance generating function, $G_Y(z) = \sum_{h=-\infty}^{\infty} \gamma_Y(h)z^h$, of the stationary process $\{Y_t\}$ generated by the state space model (7.27) and (7.28) is given by (7.34). The following two theorems give alternative expressions. They can be proved using the relations (5.26) and (5.29) obtained in Sect. 5.6.

Theorem 7.8 *Under Assumption 7.1, the covariance generating function, $G_Y(z)$, of $\{Y_t\}$ is given by*

$$G_Y(z) = \Sigma + H\overline{\Sigma}H' + zH(I - Fz)^{-1}(K\Sigma + F\overline{\Sigma}H') + z^{-1}(\Sigma K' + H\overline{\Sigma}F')(I - F'z^{-1})^{-1}H'.$$

Theorem 7.9 *The covariance generating function of $\{Y_t\}$ in terms of the system matrices of (7.27) and (7.28) is given by*

$$G_Y(z) = R + H\Pi H' + zH(I - Fz)^{-1}(GS + F\Pi H') + z^{-1}(S'G' + H\Pi F')(I - F'z^{-1})^{-1}H'.$$

7.2.3 Covariance Generating Functions of the State Errors

The errors $E_{t|t}$, $E_{t|\infty}$, and $E_{t+k|t}$ were shown to be stationary in Sect. 7.1. However, the expressions given by Theorem 7.1 or Theorem 7.2 for their generating functions are not very operational in the case of stationary state space models. In this section, we will provide operational expressions for the errors $\epsilon_{t|t-1} = x_t - \hat{x}_{t|t-1}$, $\epsilon_{t|\infty} = x_t - \hat{x}_{t|\infty}$ and $\epsilon_{t|t+m} = x_t - \hat{x}_{t|t+m}$, where $m \geq 0$.

As regards the error $\epsilon_{t|t-1}$, we first need a lemma.

Lemma 7.3 *The error $\epsilon_{t|t-1}$ is stationary and is given by $\epsilon_{t|t-1} = (I - F_pB)^{-1}(Gu_{t-1} - Kv_{t-1})$. Moreover, its covariance generating function, $G_{\epsilon,-1}(z)$, is*

$$G_{\epsilon,-1}(z) = (I - F_pz)^{-1}[G \;\; -K]\begin{bmatrix} Q & S \\ S' & R \end{bmatrix}\begin{bmatrix} G' \\ -K' \end{bmatrix}(I - F_p'z^{-1})^{-1}. \tag{7.51}$$

Proof Using Corollary 7.1, it is obtained that

$$\begin{aligned}
\epsilon_{t|t-1} &= x_t - (I - F_pB)^{-1}KB(Hx_t + v_t) \\
&= \left[I - (I - F_pB)^{-1}KHB\right]x_t - (I - F_pB)^{-1}KBv_t \\
&= (I - F_pB)^{-1}\left[I - F_pB - (F - F_p)B\right]x_t - (I - F_pB)^{-1}KBv_t \\
&= (I - F_pB)^{-1}(I - FB)(I - FB)^{-1}BGu_t - (I - F_pB)^{-1}KBv_t \\
&= (I - F_pB)^{-1}B(Gu_t - Kv_t).
\end{aligned}$$

Equation (7.51) follows easily from this. In addition, $\epsilon_{t|t-1}$ is stationary because F_p is stable. □

The following theorem gives an expression for $G_{\epsilon,-1}(z)$ simpler than (7.51).

Theorem 7.10 *The covariance generating function $G_{\epsilon,-1}(z)$ can be written as*

$$G_{\epsilon,-1}(z) = (I - F_p z)^{-1} P + P F_p' z^{-1} (I - F_p' z^{-1})^{-1}.$$

Proof To obtain the result, we will make use of the following alternative expression for the DARE (7.36)

$$P = F_p P F_p' + [G \ -K] \begin{bmatrix} Q & S \\ S' & R \end{bmatrix} \begin{bmatrix} G' \\ -K' \end{bmatrix}. \tag{7.52}$$

This can be proved by noting the recursion

$$\epsilon_{t+1|t} = F\epsilon_{t|t-1} + Gu_t - K(H\epsilon_{t|t-1} + v_t) = F_p \epsilon_{t|t-1} + [G \ -K] \begin{bmatrix} u_t \\ v_t \end{bmatrix},$$

that follows by subtracting (7.40) from (7.27). Using (7.52), it is obtained that

$$\begin{aligned} G_{\epsilon,-1}(z) &= (I - F_p z)^{-1} (P - F_p P F_p') (I - F_p' z^{-1})^{-1} \\ &= (I - F_p z)^{-1} \left[P(I - F_p' z^{-1}) + (I - F_p z) P \right. \\ &\quad \left. -(I - F_p z) P (I - F_p' z^{-1}) \right] (I - F_p' z^{-1})^{-1} \\ &= (I - F_p z)^{-1} P + P (I - F_p' z^{-1})^{-1} - P. \end{aligned}$$

□

Before obtaining the covariance generating function of the error $\epsilon_{t|\infty}$, we need the covariance generating function, $G_{\lambda,\infty}(z)$, of $\lambda_{t|\infty}$. It follows from (7.44) that

$$G_{\lambda,\infty}(z) = (I - F_p' z^{-1})^{-1} H' \Sigma^{-1} H (I - F_p z)^{-1}. \tag{7.53}$$

Lemma 7.4 gives a simpler expression for $G_{\lambda,\infty}(z)$ than (7.53). The proof is similar to that of Theorem 7.10 and is omitted.

Lemma 7.4 *The covariance generating function, $G_{\lambda,\infty}(z)$, of $\lambda_{t|\infty}$ can be written as*

$$G_{\lambda,\infty}(z) = \Lambda_{|\infty} (I - F_p z)^{-1} + (I - F_p' z^{-1})^{-1} z^{-1} F_p' \Lambda_{|\infty},$$

where $\Lambda_{|\infty}$ is given by (7.45).

As regards the error $\epsilon_{t|\infty}$, the following theorem holds.

Theorem 7.11 *The error $\epsilon_{t|\infty}$ is stationary and is given by $\epsilon_{t|\infty} = \epsilon_{t|t-1} - P\lambda_{t|\infty}$. Its covariance generating function, $G_{\epsilon,\infty}(z)$, satisfies*

$$G_{\epsilon,\infty}(z) = P(I - \Lambda_{|\infty}P) + z(I - P\Lambda_{|\infty})(I - F_p z)^{-1}F_p P$$
$$+z^{-1}PF_p'(I - F_p'z^{-1})^{-1}(I - \Lambda_{|\infty}P).$$

Proof First, note the relation

$$\epsilon_{t|\infty} = x_t - \hat{x}_{t|\infty} = x_t - \hat{x}_{t|t-1} - P\lambda_{t|\infty}$$
$$= \epsilon_{t|t-1} - P\lambda_{t|\infty}.$$

Since $\epsilon_{t|t-1}$ and $\lambda_{t|\infty}$ are stationary, it follows that $\epsilon_{t|\infty}$ is stationary. Then, $\epsilon_{t|t-1} = \epsilon_{t|\infty} + P\lambda_{t|\infty}$, where $\epsilon_{t|\infty}$ and $\lambda_{t|\infty}$ are uncorrelated, and it follows that $G_{\epsilon,\infty}(z) = G_{\epsilon,-1}(z) - PG_{\lambda,\infty}(z)P$. Using Lemmas 7.10 and 7.4, after some manipulation, it is obtained that

$$G_{\epsilon,\infty}(z) = G_{\epsilon,-1}(z) - PG_{\lambda,\infty}(z)P$$
$$= (I - P\Lambda_{|\infty})(I - F_p z)^{-1}P + PF_p'(I - F_p'z^{-1})^{-1}z^{-1}(I - \Lambda_{|\infty}P).$$

□

As far as the error $\epsilon_{t|t+m}$ is concerned, the following theorem holds.

Theorem 7.12 *The error $\epsilon_{t|t+m}$ is stationary and is given by $\epsilon_{t|t+m} = \epsilon_{t|\infty} + P(\lambda_{t|\infty} - \lambda_{t|t+m})$, where $\epsilon_{t|\infty}$ is the* ***final estimation error*** *and $P(\lambda_{t|\infty} - \lambda_{t|t+m})$ is the* ***revision*** *from $\hat{x}_{t|t+m}$ to $\hat{x}_{t|\infty}$. Its covariance generating function, $G_{\epsilon,+m}(z)$, satisfies*

$$G_{\epsilon,+m}(z) = G_{\epsilon,\infty}(z) + PF_p'^{m+1}G_{\lambda,\infty}(z)F_p^{m+1}P$$
$$= (I - P\Lambda_{|\infty} + PF_p'^{m+1}\Lambda_{|\infty}F_p^{m+1})(I - F_p z)^{-1}P$$
$$+PF_p'(I - F_p'z^{-1})^{-1}z^{-1}(I - \Lambda_{|\infty}P + F_p'^{m+1}\Lambda_{|\infty}F_p^{m+1}P).$$

where $G_{\epsilon,\infty}(z)$ and $G_{\lambda,\infty}(z)$ are given in Theorem 7.11 and Lemma 7.4.

Proof First, note that

$$\epsilon_{t|t+m} = x_t - \hat{x}_{t|t+m} = (x_t - \hat{x}_{t|\infty}) + (x_{t|\infty} - \hat{x}_{t|t+m})$$
$$= \epsilon_{t|\infty} + P(\lambda_{t|\infty} - \lambda_{t|t+m}),$$

where the last two terms in the right-hand side of the previous expression are uncorrelated. Then, letting $\omega_t = \lambda_{t|\infty} - \lambda_{t|t+m}$ and using (7.44) and (7.48), it can be easily verified that the generating function, $\omega(z)$, of ω_t is

$$\omega(z) = z^{-m-1}F_p'^{m+1}(I - F_p'z^{-1})^{-1}H'\Sigma^{-1}A(z) = z^{-m-1}F_p'^{m+1}\lambda_{|\infty}(z),$$

and its covariance generating function, $G_\omega(z)$, is $G_\omega(z) = F_p'^{m+1}G_{\lambda,\infty}(z)F_p^{m+1}$. □

Remark 7.6 It is to be noticed that, to compute the covariance matrices of $\epsilon_{t|\infty}$ and $\epsilon_{t|t+m}$, it is necessary to solve both, the DARE (7.36) and the Lyapunov equation (7.45). ◇

7.2.4 Computing the Filter Weights

In this section, we will derive some recursions to compute some of the filter weights for Wiener–Kolmogorov filtering and smoothing.

The weights of the predictor $\hat{x}_{t|t-1}$ can be obtained from Corollary 7.1. Let $\hat{x}_{t|t-1} = \sum_{j=0}^{\infty} \Pi_i Y_{t-1-j} = \Pi(B)Y_{t-1}$. Then, using the equality $(I - F_p z)\Pi(z) = K$, it is obtained that

$$\hat{x}_{t|t-1} = (I - F_pB)^{-1}KY_{t-1} = \sum_{j=0}^{\infty} F_p^j KY_{t-1-j}. \quad (7.54)$$

Note that, because F_p is stable, the weights will decrease to zero.

The innovations A_t can be expressed in terms of the observations Y_t by inverting the equation $Y_t = \Psi(B)A_t = \left[I + BH(I - FB)^{-1}K\right]A_t$. Using (7.35), we can write

$$A_t = \left[I - BH(I - F_pB)^{-1}K\right]Y_t = Y_t + \sum_{j=0}^{\infty} HF_p^j KY_{t-1-j}. \quad (7.55)$$

By (7.44) and (7.55), we can express $\lambda_{t|\infty}$ in terms of the observations Y_t as

$$\lambda_{t|\infty} = (I - F_p'B^{-1})^{-1}H'\Sigma^{-1}\left[I - BH(I - F_pB)^{-1}K\right]Y_t. \quad (7.56)$$

The following theorem gives the filter weights for the smoother $\hat{x}_{t|\infty}$.

Theorem 7.13 *Letting the estimator* $\hat{x}_{t|\infty} = \sum_{j=-\infty}^{\infty} \Omega_j Y_{t+j}$, *its weights* Ω_j *are given by*

$$\begin{aligned}
\Omega_0 &= P(H'\Sigma^{-1} - F_p'\Lambda_{|\infty}K), \\
\Omega_j &= (I - P\Lambda_{|\infty})F_p^{-j-1}K, \qquad j < 0 \\
\Omega_j &= PF_p'^j(H'\Sigma^{-1} - F_p'\Lambda_{|\infty}K), \; j > 0,
\end{aligned}$$

where $\Lambda_{|\infty}$ *can be obtained by solving the Lyapunov equation (7.45).*

Proof By Corollary 7.1 and (7.56), we can rewrite the recursion (7.43) as

$$\begin{aligned}\hat{x}_{t|\infty} &= (I - F_pB)^{-1}KY_{t-1} + P(I - F_p'B^{-1})^{-1}H'\Sigma^{-1}\left[I - BH(I - F_pB)^{-1}K\right]Y_t \\ &= (I - F_pB)^{-1}KY_{t-1} + P(I - F_p'B^{-1})^{-1}H'\Sigma^{-1}Y_t \\ &\quad -P(I - F_p'B^{-1})^{-1}H'\Sigma^{-1}H(I - F_pB)^{-1}KY_{t-1}.\end{aligned}$$

Since $(I - F_p'B^{-1})^{-1}H'\Sigma^{-1}H(I - F_pB)^{-1} = G_{\lambda,\infty}(z)$, by Lemma 7.4, it is obtained that

$$\begin{aligned}\hat{x}_{t|\infty} &= (I - P\Lambda_{|\infty})(I - F_pB)^{-1}KY_{t-1} \\ &\quad +P(I - F_p'B^{-1})^{-1}(H'\Sigma^{-1} - F_p'\Lambda_{|\infty}K)Y_t.\end{aligned}$$

Rearranging terms in the previous expression, the theorem is proved. □

Before computing the weights of the estimator $\hat{x}_{t|t+m}$, where $m \geq 0$, we first need a lemma.

Lemma 7.5 *The following equality holds*

$$\sum_{j=0}^{m} F_p'^j z^{-j} H'\Sigma^{-1}H(I - F_pz)^{-1} = \sum_{j=0}^{m} F_p'^j z^{-j}\Lambda_{t+j|t+m} + \Lambda_{t|t+m}(I - F_pz)^{-1} - \Lambda_{t|t+m},$$

where the $\Lambda_{s|t+m}$ *are given by (7.49).*

Proof By the recursion (7.49), we can write

$$\begin{aligned}&\sum_{j=0}^{m} F_p'^j z^{-j} H'\Sigma^{-1}H(I - F_pz)^{-1} \\ &\quad = \sum_{j=0}^{m} F_p'^j z^{-j}(\Lambda_{t+j|t+m} - F_p'\Lambda_{t+j+1|t+m}F_p)(I - F_pz)^{-1} \\ &\quad = \sum_{j=0}^{m} F_p'^j z^{-j}\left[\Lambda_{t+j|t+m}(I - F_pz) + \Lambda_{t+j|t+m}F_pz - F_p'\Lambda_{t+j+1|t+m}F_p\right](I - F_pz)^{-1} \\ &\quad = \sum_{j=0}^{m} F_p'^j z^{-j}\Lambda_{t+j|t+m} + \sum_{j=0}^{m} F_p'^j z^{-j}\left[\Lambda_{t+j|t+m}F_pz - F_p'\Lambda_{t+j+1|t+m}F_p\right](I - F_pz)^{-1}.\end{aligned}$$

The lemma will follow if we prove that $\sum_{j=0}^{m} F_p'^{j} z^{-j}\left[\Lambda_{t+j|t+m}F_p z - F_p'\Lambda_{t+j+1|t+m}F_p\right] = \Lambda_{t|t+m}F_p z$. To see this, note the telescoping sum

$$\begin{aligned}
&\sum_{j=0}^{m} F_p'^{j} z^{-j}\left[\Lambda_{t+j|t+m}F_p z - F_p'\Lambda_{t+j+1|t+m}F_p\right] \\
&\quad = z\Lambda_{t|t+m}F_p \qquad\qquad\quad -F_p'\Lambda_{t+1|t+m}F_p \\
&\qquad +F_p'\Lambda_{t+1|t+m}F_p \qquad\quad -F_p'^{2}z^{-1}\Lambda_{t+2|t+m}F_p \\
&\qquad +F_p'^{2}z^{-1}\Lambda_{t+2|t+m}F_p \quad -F_p'^{3}z^{-2}\Lambda_{t+2|t+m}F_p \\
&\qquad \ldots \qquad\qquad\qquad\qquad\quad \ldots \\
&\qquad +F_p'^{m}z^{-m+1}\Lambda_{t+m|t+m}F_p \; -F_p'^{m+1}z^{-m}\Lambda_{t+m+1|t+m}F_p \\
&\quad = \Lambda_{t|t+m}F_p z.
\end{aligned}$$

□

The weights of the estimator $\hat{x}_{t|t+m}$ are given by the following theorem.

Theorem 7.14 *Letting the estimator* $\hat{x}_{t|t+m} = \sum_{j=-\infty}^{m} \Pi_j Y_{t+j}$, *its weights* Π_j *are given by*

$$\begin{aligned}
\Pi_0 &= P(H'\Sigma^{-1} - F_p'\Lambda_{t+1|t+m}K), \\
\Pi_j &= (I - P\Lambda_{t|t+m})F_p^{-j-1}K, \qquad\qquad j < 0 \\
\Pi_j &= PF_p'^{j}(H'\Sigma^{-1} - F_p'\Lambda_{t+1+j|t+m}K), \; j = 1, 2, \ldots, m,
\end{aligned}$$

where the $\Lambda_{s|t+m}$ *can be computed using the recursions (7.49).*

Proof By Corollary 7.1 and (7.55), we can rewrite the recursion (7.47) as

$$\begin{aligned}
\hat{x}_{t|t+m} &= (I - F_pB)^{-1}KY_{t-1} + P\sum_{j=0}^{m}\{F_p'^{j}B^{-j}H'\Sigma^{-1}\left[I - BH(I - F_pB)^{-1}K\right]Y_t\} \\
&= (I - F_pB)^{-1}KY_{t-1} + P\sum_{j=0}^{m}\{F_p'^{j}B^{-j}H'\Sigma^{-1}Y_t\} \\
&\quad -P\sum_{j=0}^{m}\{F_p'^{j}B^{-j}H'\Sigma^{-1}H(I - F_pB)^{-1}KY_{t-1}\}.
\end{aligned}$$

Then, using Lemma 7.5, it is obtained that

$$\begin{aligned}
\hat{x}_{t|t+m} &= (I - P\Lambda_{t|t+m})(I - F_pB)^{-1}KY_{t-1} \\
&\qquad + P\sum_{j=0}^{m} F_p'^{j}B^{-j}(H'\Sigma^{-1} - F_p'\Lambda_{t+1+j|t+m}K)Y_t.
\end{aligned}$$

Rearranging terms in the previous expression, the theorem is proved. □

7.2.5 Disturbance Smoothing and Interpolation

Define the **inverse process**, $\{Y_t^i\}$, of $\{Y_t\}$ by

$$Y_t^i = \Sigma^{-1}A_t - K'\lambda_{t+1|\infty} = \left[I - B^{-1}K'(I - F_p'B^{-1})^{-1}H'\right]\Sigma^{-1}A_t.$$

Then, by (7.35), $\{Y_t^i\} = \Psi'^{-1}(B^{-1})\Sigma^{-1}A_t$ and the name is justified because the covariance generating function of $\{Y_t^i\}$ is the inverse of the covariance generating function of $\{Y_t\}$, that is, $G_{Y^i}(z) = G_Y^{-1}(z) = \Psi'^{-1}(z^{-1})\Sigma^{-1}\Psi^{-1}(z)$. Note the relation

$$Y_t^i = \Psi'^{-1}(B^{-1})\Sigma^{-1}\Psi^{-1}(B)Y_t = G_{Y^i}(B)Y_t. \tag{7.57}$$

As we will see, the inverse process plays a fundamental role in interpolation and is related to the estimators, $\hat{u}_{t|\infty}$ and $\hat{v}_{t|\infty}$, of u_t and v_t in (7.27) and (7.28) based on the doubly infinite sample $\{Y_s : s = 0, \pm 1, \pm 2, \ldots\}$. Also, the so-called **inverse model**, that goes backward in time, is defined in terms of the inverse process,

$$\begin{aligned}\lambda_{t|\infty} &= F_p'\lambda_{t+1|\infty} + H'\Sigma^{-1}A_t \\ Y_t^i &= -K'\lambda_{t+1|\infty} + \Sigma^{-1}A_t,\end{aligned}$$

where $\lambda_{t|\infty}$ is given by the recursions (7.44).

Using (7.26), it is not difficult to see that the generating functions, $\hat{u}_{|\infty}(z)$ and $\hat{v}_{|\infty}(z)$, of $\hat{u}_{t|\infty}$ and $\hat{v}_{t|\infty}$ are $\hat{u}_{|\infty}(z) = G_{UA}(z)\Sigma^{-1}A(z)$ and $\hat{v}_{|\infty}(z) = G_{VA}(z)\Sigma^{-1}A(z)$, where $G_{UA}(z)$ and $G_{VA}(z)$ are the cross-covariance functions of $\{u_t\}$ and $\{v_t\}$ and the innovations $\{A_t\}$. Proceeding as in the proof of Lemma 7.2, we get the following lemma. The proof is omitted.

Lemma 7.6 *The covariance generating functions, $G_{VA}(z)$ and $G_{UA}(z)$, can be decomposed as*

$$\begin{aligned}G_{VA}(z) &= R + (S'G' - RK')(I - F_p'z^{-1})^{-1}H'z^{-1}, \\ G_{UA}(z) &= S + (QG' - SK')(I - F_p'z^{-1})^{-1}H'z^{-1}.\end{aligned}$$

where $F_p = F - KH$, $K = (FPH' + GS)\Sigma^{-1}$, $\Sigma = R + HPH'$, and P is the unique positive semidefinite solution of the DARE (7.36).

The following theorem gives expressions for $\hat{u}_{t|\infty}$ and $\hat{v}_{t|\infty}$ and their MSE.

Theorem 7.15 *The estimators, $\hat{u}_{t|\infty}$ and $\hat{v}_{t|\infty}$, of u_t and v_t based on the doubly infinite sample $\{Y_s : s = 0, \pm 1, \pm 2, \ldots\}$ are given by the recursions*

$$\begin{aligned}\hat{v}_{t|\infty} &= R\Sigma^{-1}A_t + (S'G' - RK')\lambda_{t+1|\infty} = RY_t^i + S'G'\lambda_{t+1|\infty}, \\ \hat{u}_{t|\infty} &= S\Sigma^{-1}A_t + (QG' - SK')\lambda_{t+1|\infty} = SY_t^i + QG'\lambda_{t+1|\infty},\end{aligned}$$

where Y_t^i is the inverse model and $\lambda_{t|\infty}$ is given by the recursions (7.44). In addition, the MSE of $\hat{u}_{t|\infty}$ and $\hat{v}_{t|\infty}$ are given by the formulae

$$\mathrm{MSE}(\hat{v}_{t|\infty}) = R - \left[R\Sigma^{-1}R' + (S'G' - RK')\Lambda_{|\infty}(S'G' - RK')'\right],$$
$$\mathrm{MSE}(\hat{u}_{t|\infty}) = Q - \left[S\Sigma^{-1}S' + (QG' - SK')\Lambda_{|\infty}(QG' - SK')'\right].$$

Proof Using Lemma 7.6 and $\lambda_{|\infty}(z) = (I - F_p'z^{-1})^{-1}H'\Sigma^{-1}A(z)$, we can express $\hat{u}_{|\infty}(z)$ and $\hat{v}_{|\infty}(z)$ as $\hat{u}_{|\infty}(z) = S\Sigma^{-1}A(z) + (QG' - SK')z^{-1}\lambda_{|\infty}(z)$ and $\hat{v}_{|\infty}(z) = R\Sigma^{-1}A(z) + (S'G' - RK')z^{-1}\lambda_{|\infty}(z)$. To obtain the MSE, consider that $u_t - \hat{u}_{t|\infty} + S\Sigma^{-1}A_t + (QG' - SK')\lambda_{t+1|\infty} = u_t$ and that $u_t - \hat{u}_{t|\infty}$ and $S\Sigma^{-1}A_t + (QG' - SK')\lambda_{t+1|\infty}$ are uncorrelated. Since A_t and $\lambda_{t+1|\infty}$ are also uncorrelated, it follows that

$$\mathrm{MSE}(\hat{u}_{t|\infty}) = Q - \left[S\Sigma^{-1}S' + (QG' - SK')\Lambda_{|\infty}(QG' - SK')'\right].$$

Proceeding similarly with $\hat{v}_{t|\infty}$, the theorem is proved. □

Using (7.55), (7.56) and Lemma 7.4, we can express $\hat{u}_{t|\infty}$ and $\hat{v}_{t|\infty}$ in terms of the observations $\{Y_s : s = 0, \pm 1, \pm 2, \ldots\}$. The following theorem gives the details. The proof is omitted.

Theorem 7.16 *Letting $J' = S'G' - RK'$ and $L' = QG' - SK'$, the weights for $\hat{v}_{t|\infty}$ and $\hat{u}_{t|\infty}$ are given by*

$$\begin{aligned}\hat{v}_{t|\infty} = &(R\Sigma^{-1} - J'\Lambda_{|\infty}K)Y_t - (J'\Lambda_{|\infty}F_p + R\Sigma^{-1}H)(I - F_pB)^{-1}Y_{t-1}\\ &-J'(I - F_p'B^{-1})^{-1}(F_p'\Lambda_{|\infty}K - H'\Sigma^{-1})Y_{t+1}\end{aligned}$$

$$\begin{aligned}\hat{u}_{t|\infty} = &(S\Sigma^{-1} - L'\Lambda_{|\infty}K)Y_t + -(L'\Lambda_{|\infty}F_p + S\Sigma^{-1}H)(I - F_pB)^{-1}Y_{t-1}\\ &-L'(I - F_p'B^{-1})^{-1}(F_p'\Lambda_{|\infty}K - H'\Sigma^{-1})Y_{t+1}.\end{aligned}$$

The inverse process $\{Y_t^i\}$ plays an important role in the interpolation of missing observations. Let $G_Y^{-1}(z) = G_{Y^i}(z) = \sum_{j=-\infty}^{\infty}\gamma_i(j)z^j$. The following lemma gives an expression for $G_{Y^i}(z)$ in terms of the system matrices.

Lemma 7.7 *The covariance generating function, $G_{Y^i}(z) = \sum_{j=-\infty}^{\infty}\gamma_i(j)z^j$, of the inverse model can be expressed in terms of the system matrices as*

$$\begin{aligned}G_{Y^i}(z) = &\Sigma^{-1} + K'\Lambda_{|\infty}K + z(K'\Lambda_{|\infty}F_p - \Sigma^{-1}H)(I - F_pz)^{-1}K\\ &+z^{-1}K'(I - F_p'z^{-1})^{-1}(F_p'\Lambda_{|\infty}K - H'\Sigma^{-1}).\end{aligned}$$

Proof Consider first that

$$\begin{aligned} G_{Y^i}(z) &= \Psi'^{-1}(z^{-1})\Sigma^{-1}\Psi^{-1}(z) \\ &= \Sigma^{-1} - z\Sigma^{-1}H(I-F_pz)^{-1}K - z^{-1}K'(I-F_p'z^{-1})^{-1}H'\Sigma^{-1} \\ &\quad + K'(I-F_p'z^{-1})^{-1}H'\Sigma^{-1}H(I-F_pz)^{-1}K. \end{aligned}$$

Then, apply Lemma 7.4. □

Suppose that we want to estimate Y_t based on $\{Y_s : s \neq t\}$ and let the desired estimator be $\hat{Y}_{t|s\neq t} = \sum_{j\neq 0} \Pi_j Y_{t-j}$. The following theorem gives the weights of $\hat{Y}_{t|s\neq t}$ and its MSE.

Theorem 7.17 *The interpolator, $\hat{Y}_{t|s\neq t}$, of Y_t based on $\{Y_s : s \neq t\}$ is given by*

$$\begin{aligned} \hat{Y}_{t|s\neq t} &= Y_t - \gamma_i^{-1}(0)Y_t^i = -\gamma_i^{-1}(0)\sum_{j\neq 0}\gamma_i(j)Y_{t-j} \\ &= -(\Sigma^{-1} + K'\Lambda_{|\infty}K)^{-1}\left[B(K'\Lambda_{|\infty}F_p - \Sigma^{-1}H)(I-F_pB)^{-1}K \right. \\ &\quad \left. + B^{-1}K'(I-F_p'B^{-1})^{-1}(F_p'\Lambda_{|\infty}K - H'\Sigma^{-1})\right]Y_t. \end{aligned}$$

where $\gamma_i(l) = E(Y_t^iY_{t-l}^{i'})$. In addition, $\mathrm{MSE}(\hat{Y}_{t|s\neq t}) = \gamma_i(0) = \Sigma^{-1} + K'\Lambda_{|\infty}K$.

Proof The orthogonality conditions to obtain $\hat{Y}_{t|s\neq t} = \sum_{j\neq 0}\Pi_jY_{t-j}$ are

$$E\left[(Y_t - \sum_{j\neq 0}\Pi_jY_{t-j})Y_{t-l}'\right] = 0, \quad l \neq 0,$$

or, equivalently, $\gamma(l) - \sum_{j\neq 0}\Pi_j\gamma(l-j) = 0$, $l \neq 0$, where $\gamma(l) = E(Y_tY_{t-l}')$. To use the Wiener–Hopf technique, define $G(z) = \sum_{j=-\infty}^{\infty} G_iz^i$, where $G_i = \gamma(l) - \sum_{j\neq 0}\Pi_j\gamma(l-j)$, $j = 0, \pm 1, \pm 2, \ldots$ Then, if $\Pi(z)$ is the generating function of the Π_j weights, it holds that

$$G(z) = G_Y(z) - \Pi(z)G_Y(z) = G_0,$$

because $G_l = 0$ if $l \neq 0$. It follows from this that $\Pi(z) = I - G_0G_Y^{-1}(z)$ and, therefore, $I - G_0\gamma^i(0) = 0$, where $G_Y^{-1}(z) = G_{Y^i}(z) = \sum_{j=-\infty}^{\infty}\gamma_i(j)z^j$. That is, $G_0 = \gamma_i^{-1}(0)$.

Using Lemma 7.7, we can write

$$\begin{aligned}\Pi(z) &= I - \gamma_i^{-1}(0)\sum_{j=-\infty}^{\infty}\gamma_i(j)z^j = -\sum_{j\neq 0}\gamma_i^{-1}(0)\gamma_i(j)z^j\\ &= -(\Sigma^{-1}+K'\Lambda_{|\infty}K)^{-1}\left[z(K'\Lambda_{|\infty}F_p-\Sigma^{-1}H)(I-F_pz)^{-1}K\right.\\ &\quad \left.+z^{-1}K'(I-F_p'z^{-1})^{-1}(F_p'\Lambda_{|\infty}K-H'\Sigma^{-1})\right].\end{aligned}$$

From this and (7.57), the expression for $\hat{Y}_{t|s\neq t}$ follows. To obtain the MSE, consider the error $E_{t|s\neq t} = Y_t - \hat{Y}_{t|s\neq t}$. Then, $E_{t|s\neq t} = Y_t - \left[Y_t - \gamma_i^{-1}(0)G_Y^{-1}(B)Y_t\right] = \gamma_i^{-1}(0)Y_t^i$, and the theorem is proved. □

7.2.6 Covariance Generating Functions of the Disturbance Errors

In this section, we will complement the results of the previous section by giving some results on covariance generating functions of the disturbance errors.

Define the disturbance errors $\eta_{t|\infty} = v_t - \hat{v}_{t|\infty}$ and $\delta_{t|\infty} = u_t - \hat{u}_{t|\infty}$ and let $G_{\eta,\infty}(z)$, $G_{\delta,\infty}(z)$, $G_{\eta\delta,\infty}(z)$, $G_{\epsilon\eta,\infty}(z)$, and $G_{\epsilon\delta,\infty}(z)$ be the covariance generating functions of $\eta_{t|\infty}$ and $\delta_{t|\infty}$ and the cross-covariance generating functions of $\eta_{t|\infty}$ and $\delta_{t|\infty}$, $\epsilon_{t|\infty}$ and $\eta_{t|\infty}$, and $\epsilon_{t|\infty}$ and $\delta_{t|\infty}$.

The previous covariance and cross covariance generating functions are given by the following theorem.

Theorem 7.18 *The covariance generating functions, $G_{\eta,\infty}(z)$ and $G_{\delta,\infty}(z)$, of $\eta_{t|\infty}$ and $\delta_{t|\infty}$ and the cross-covariance generating functions, $G_{\eta\delta,\infty}(z)$, $G_{\epsilon\eta,\infty}(z)$ and $G_{\epsilon\delta,\infty}(z)$, of $\eta_{t|\infty}$ and $\delta_{t|\infty}$, $\epsilon_{t|\infty}$ and $\eta_{t|\infty}$, and $\epsilon_{t|\infty}$ and $\delta_{t|\infty}$ are given by*

$$\begin{aligned}G_{\eta,\infty}(z) &= R - R\Sigma^{-1}R - J'\Lambda_{|\infty}J\\ &\quad -(R\Sigma^{-1}H + J'\Lambda_{|\infty}F_p)(I-F_pz)^{-1}Jz\\ &\quad -J'(I-F_p'z^{-1})^{-1}(H'\Sigma^{-1}R + F_p'\Lambda_{|\infty}J)z^{-1},\end{aligned}$$

$$\begin{aligned}G_{\delta,\infty}(z) &= Q - S\Sigma^{-1}S' - L'\Lambda_{|\infty}L\\ &\quad -(S\Sigma^{-1}H + L'\Lambda_{|\infty}F_p)(I-F_pz)^{-1}Lz\\ &\quad -L'(I-F_p'z^{-1})^{-1}(H'\Sigma^{-1}S' + F_p'\Lambda_{|\infty}L)z^{-1},\end{aligned}$$

$$\begin{aligned}G_{\eta\delta,\infty}(z) &= S' - R\Sigma^{-1}S' - J'\Lambda_{|\infty}L\\ &\quad -(R\Sigma^{-1}H + J'\Lambda_{|\infty}F_p)(I-F_pz)^{-1}Lz\\ &\quad -J'(I-F_p'z^{-1})^{-1}(H'\Sigma^{-1}S' + F_p'\Lambda_{|\infty}L)z^{-1},\end{aligned}$$

$$
\begin{aligned}
G_{\epsilon\eta,\infty}(z) = & -P(H'\Sigma^{-1}R + F_p'\Lambda_{|\infty}J) \\
& +(I - P\Lambda_{|\infty})(I - F_p z)^{-1}Jz \\
& -P(I - F_p'z^{-1})^{-1}F_p'(H'\Sigma^{-1}R + F_p'\Lambda_{|\infty}J)z^{-1},
\end{aligned}
$$

and

$$
\begin{aligned}
G_{\epsilon\delta,\infty}(z) = & -P(H'\Sigma^{-1}S' + F_p'\Lambda_{|\infty}L) \\
& +(I - P\Lambda_{|\infty})(I - F_p z)^{-1}Lz \\
& -P(I - F_p'z^{-1})^{-1}F_p'(H'\Sigma^{-1}S' + F_p'\Lambda_{|\infty}L)z^{-1},
\end{aligned}
$$

where $L' = QG' - SK'$ *and* $J' = S'G' - RK'$.

Proof To obtain the covariance and cross covariance generating functions, we use the following equality

$$
\begin{pmatrix} u_t \\ v_t \\ \epsilon_{t|t-1} \end{pmatrix} = \begin{pmatrix} \delta_{t|\infty} \\ \eta_{t|\infty} \\ \epsilon_{t|\infty} \end{pmatrix} + \begin{pmatrix} S\Sigma^{-1}A_t + (QG' - SK')\lambda_{t+1|\infty} \\ R\Sigma^{-1}A_t + (S'G' - RK')\lambda_{t+1|\infty} \\ P\lambda_{t|\infty} \end{pmatrix},
$$

that follows directly from Theorems 7.11 and 7.15. By Lemma 7.3 and (7.44), we can write

$$
\begin{pmatrix} I & 0 \\ 0 & I \\ (I - F_p z)^{-1}zG & -(I - F_p z)^{-1}zK \end{pmatrix} \begin{pmatrix} u_t \\ v_t \end{pmatrix}
$$

$$
= \begin{pmatrix} \delta_{t|\infty} \\ \eta_{t|\infty} \\ \epsilon_{t|\infty} \end{pmatrix} + \left\{ \begin{pmatrix} S \\ R \\ 0 \end{pmatrix} + \begin{bmatrix} (QG' - SK')z^{-1} \\ (S'G' - RK')z^{-1} \\ P \end{bmatrix} (I - F_p'z^{-1})^{-1}H' \right\} \Sigma^{-1}A_t,
$$

where the last two terms are uncorrelated. It follows from this that the covariance generating function of the left-hand side is equal to the sum of the covariance generating functions of the two terms in the right-hand side. Letting

$$
G(z) = \begin{pmatrix} G_{\delta,\infty}(z) & G_{\delta\eta,\infty}(z) & G_{\delta\epsilon,\infty}(z) \\ G_{\eta\delta,\infty}(z) & G_{\eta,\infty}(z) & G_{\eta\epsilon,\infty}(z) \\ G_{\epsilon\delta,\infty}(z) & G_{\epsilon\eta,\infty}(z) & G_{\epsilon,\infty}(z) \end{pmatrix}
$$

be the covariance generating function of $(\delta_{t|\infty}, \eta_{t|\infty}, \epsilon_{t|\infty})'$, $L = GQ - KS'$ and $J = GS - KR$, we get

$$\begin{pmatrix} Q & S & L'(I-F_p'z^{-1})^{-1}z^{-1} \\ S' & R & J'(I-F_p'z^{-1})^{-1}z^{-1} \\ (I-F_pz)^{-1}zL & (I-F_pz)^{-1}zJ & G_{\epsilon,-1}(z) \end{pmatrix}$$
$$= G(z) + \begin{pmatrix} S\Sigma^{-1}S' & S\Sigma^{-1}R & 0 \\ R\Sigma^{-1}S' & R\Sigma^{-1}R & 0 \\ 0 & 0 & 0 \end{pmatrix}$$
$$+ \begin{pmatrix} L'z^{-1}(I-F_p'z^{-1})^{-1}H'\Sigma^{-1}S' & L'z^{-1}(I-F_p'z^{-1})^{-1}H'\Sigma^{-1}R & 0 \\ J'z^{-1}(I-F_p'z^{-1})^{-1}H'\Sigma^{-1}S' & J'z^{-1}(I-F_p'z^{-1})^{-1}H'\Sigma^{-1}R & 0 \\ P(I-F_p'z^{-1})^{-1}H'\Sigma^{-1}S' & P(I-F_p'z^{-1})^{-1}H'\Sigma^{-1}R & 0 \end{pmatrix}$$
$$+ \begin{pmatrix} S\Sigma^{-1}H(I-F_pz)^{-1}Lz & S\Sigma^{-1}H(I-F_pz)^{-1}Jz & S\Sigma^{-1}H(I-F_pz)^{-1}P \\ R\Sigma^{-1}H(I-F_pz)^{-1}Lz & R\Sigma^{-1}H(I-F_pz)^{-1}Jz & R\Sigma^{-1}H(I-F_pz)^{-1}P \\ 0 & 0 & 0 \end{pmatrix}$$
$$+ \begin{pmatrix} L'G_{\lambda,\infty}(z)L & L'G_{\lambda,\infty}(z)J & L'G_{\lambda,\infty}(z)Pz^{-1} \\ J'G_{\lambda,\infty}(z)L & J'G_{\lambda,\infty}(z)J & J'G_{\lambda,\infty}(z)Pz^{-1} \\ PG_{\lambda,\infty}(z)Lz & PG_{\lambda,\infty}(z)Jz & PG_{\lambda,\infty}(z)P \end{pmatrix}.$$

Equating the entries in the left- and right-hand sides and using Lemma 7.4, the theorem is proved. □

7.2.7 Equivalence Between Wiener–Kolmogorov and Kalman Filtering

Under the assumptions of Theorem 5.11, the Kalman filtering and smoothing recursions converge to a steady state that coincides with the recursive Wiener–Kolmogorov filtering and smoothing recursions of Sect. 7.2. This establishes the equivalence between the two approaches.

7.2.8 Nonstationary Time Invariant State Space Models

When the processes of interest are nonstationary, for example because the matrix F in (7.27) is not stable, the Wiener–Kolmogorov formulae do not apply. However, by the results of Sect. 5.11, under the assumptions that (F, H) be detectable and $(F^s, GQ^{s/2})$ be stabilizable, the Kalman filtering and smoothing recursions continue to have a steady state limit in which the resulting recursive formulae are again the Wiener–Kolmogorov formulae. The reason for this is that, although the state vector and its estimator have infinite variance, the difference between the state vector and the estimator is stationary and thus has finite variance. The implication of the previous results is that the Wiener–Kolmogorov formulae can also be used in the nonstationary case provided the estimators are interpreted not as minimum mean squared error estimators but as limits of these.

Remark 7.7 Koopman & Harvey (2003) give an algorithm for computing finite signal extraction weights for general (possibly time variant) state space models. Earlier in Sect. 5.11, sufficient conditions were given for the Kalman filtering and smoothing recursions to converge to the steady state recursions of Sects. 7.2.1 and 7.2.5. Under these conditions, the finite sample weights will converge to the steady state weights given in Sect. 7.2.4. ◇

7.3 Wiener–Kolmogorov Filtering and Smoothing in Finite Samples

In the previous sections, we have developed Wiener–Kolmogorov filtering and smoothing by purely algebraic methods assuming that the sample was either semi-infinite or doubly infinite. In this section, our purpose will be to develop Wiener–Kolmogorov filtering and smoothing when the sample is finite.

The approach in this section is new in the literature. It is based on finite generating functions and certain algebraic rules defined to manipulate them.

7.3.1 Finite Generating Functions

When dealing with finite samples, matrices replace the generating functions of the time invariant infinite case. The idea to define finite generating functions is to associate to each row in a matrix a generating function. Suppose that we have two matrices, M and N, of dimensions $m \times p$ and $p \times n$, respectively, and we want to define the generating functions $G_{M,t}(z)$ and $G_{N,t}(z)$ corresponding to the tth rows of M and N, $t = 1, 2, \ldots, m$. If

$$M = \begin{bmatrix} M(1,1) & M(1,2) & \cdots & M(1,p) \\ M(2,1) & M(2,2) & \cdots & M(2,p) \\ \vdots & \vdots & \ddots & \vdots \\ M(m,1) & M(m,2) & \cdots & M(m,p) \end{bmatrix},$$

then the tth row is $[M(t,1), M(t,2), \ldots, M(t,p)]$ and we define $G_{M,t}(z)$ as

$$G_{M,t}(z) = M(t,1)z^{t-1} + M(t,2)z^{t-2} + \cdots + M(t,t)z^0 + \cdots + M(t,p)z^{t-p}.$$

Similarly, we define

$$G_{N,t}(z) = N(t,1)z^{t-1} + N(t,2)z^{t-2} + \cdots + N(t,t)z^0 + \cdots + N(t,n)z^{t-n}$$

for the tth row of N. Given the previous definitions, the question arises as to how to define the algebraic operations with these generating functions so that they conform to matrix multiplication. That is, if $P = MN$, then

$$G_{P,t}(z) = G_{M,t}(z)G_{N,t}(z)$$

should hold, where $G_{P,t} = \sum_{i=1}^{n} P(t,i)z^{t-i}$. Let

$$\begin{aligned}[P(t,1), \ldots, P(t,n)] = {} & M(t,1)\,[N(1,1), N(1,2), \ldots, N(1,n)] \\ & + M(t,2)\,[N(2,1), N(2,2), \ldots, N(2,n)] \\ & + \quad \cdots \\ & + M(t,p)\,[N(p,1), N(p,2), \ldots, N(p,n)]\end{aligned}$$

be the tth row of $P = MN$. If, given a matrix H, $H(i,:)$ and $H(:,j)$ denote its ith row and its jth column, respectively, the previous equality can be written in terms of the generating function $G_{P,t}(z)$ as

$$G_{P,t}(z) = [M(t,:)N(:,1)]\,z^{t-1} + [M(t,:)N(:,2)]\,z^{t-2} + \cdots + [M(t,:)N(:,n)]\,z^{t-n},$$

or

$$\begin{aligned}G_{P,t}(z) = {} & M(t,1)\left[N(1,1)z^{t-1} + N(1,2)z^{t-2} + \cdots + N(1,n)z^{t-n}\right] \\ & + M(t,2)\left[N(2,1)z^{t-1} + N(2,2)z^{t-2} + \cdots + N(2,n)z^{t-n}\right] \\ & + \quad \cdots \\ & + M(t,p)\left[N(p,1)z^{t-1} + N(p,2)z^{t-2} + \cdots + N(p,n)z^{t-n}\right].\end{aligned}$$

If we define $z^l z^k = z^{l+k}$, then

$$\begin{aligned}G_{P,t}(z) = {} & M(t,1)\left[N(1,1)z^0 + N(1,2)z^{-1} + \cdots + N(1,n)z^{1-n}\right]z^{t-1} \\ & + M(t,2)\left[N(2,1)z^1 + N(2,2)z^0 + \cdots + N(2,n)z^{2-n}\right]z^{t-2} \\ & + \quad \cdots \\ & + M(t,p)\left[N(p,1)z^{p-1} + N(p,2)z^{p-2} + \cdots + N(p,n)z^{p-n}\right]z^{t-p},\end{aligned}$$

or, equivalently,

$$G_{P,t}(z) = M(t,1)G_{N,1}(z)z^{t-1} + M(t,2)G_{N,2}(z)z^{t-2} + \cdots + M(t,p)G_{N,p}(z)z^{t-p}$$

holds. In addition, if the equality

$$G_{P,t}(z) = G_{M,t}(z)G_{N,t}(z) = \left(\sum_{i=1}^{p} M(t,i)z^{t-i}\right) G_{N,t}(z)$$

is to be satisfied, we must also define

$$z^k G_{N,t}(z) = G_{N,t-k}(z)z^k$$

for all integer k.

The same definitions are valid if M and N are block matrices. Suppose that we have two stacked vectors of random vectors, $X = (X_1', X_2', \ldots, X_n')'$ and $Y = (Y_1', Y_2', \ldots, Y_n')'$, and that the following relation exists between these two vectors

$$Y = AX,$$

where

$$A = \begin{bmatrix} A(1,1) & A(1,2) & \cdots & A(1,n) \\ A(2,1) & A(2,2) & \cdots & A(2,n) \\ \vdots & \vdots & \ddots & \vdots \\ A(n,1) & A(n,2) & \cdots & A(n,n) \end{bmatrix}$$

is a block nonstochastic matrix. If we want to express the equality $Y_t = A(t,:)X$, where $A(t,:)$ is the tth block row of A, in terms of generating functions, we can apply the previous rules. Letting $G_{Y,t}(z)$, $G_{A,t}(z)$, and $G_{X,t}(z)$ be the generating functions of the tth block rows of Y, A, and X, respectively, it follows from the definition of generating function that

$$G_{Y,t}(z) = Y_t z^{t-1}, \qquad G_{X,t}(z) = X_t z^{t-1},$$

and

$$G_{A,t}(z) = A(t,1)z^{t-1} + A(t,2)z^{t-2} + \cdots + A(t,n)z^{t-n}.$$

Since $G_{Y,t}(z) = G_{A,t}(z)G_{X,t}(z)$, we can write

$$\begin{aligned} Y_t z^{t-1} &= \left[A(t,1)z^{t-1} + A(t,2)z^{t-2} + \cdots + A(t,n)z^{t-n}\right] G_{X,t}(z) \\ &= A(t,1)G_{X,1}(z)z^{t-1} + A(t,2)G_{X,2}(z)z^{t-2} + \cdots + A(t,n)G_{X,n}(z)z^{t-n} \\ &= \left[A(t,1)X_1 + A(t,2)X_2 + \cdots + A(t,n)X_n\right] z^{t-1} \end{aligned}$$

because $G_{X,i}(z) = X_i z^{i-1}$, $i = 1, 2, \ldots, n$. Postmultiplying the previous equality by z^{1-t} and defining $z^0 = I$, we can write

$$\begin{aligned} Y_t &= A(t,1)X_1 + A(t,2)X_2 + \cdots + A(t,t)X_t + \cdots + A(t,n)X_n \\ &= \left[A(t,1)z^{t-1} + A(t,2)z^{t-2} + \cdots + A(t,t)z^0 + \cdots + A(t,n)z^{t-n}\right] X_t. \end{aligned}$$

If $G_{Y_t}(z)$ and $G_{X_t}(z)$ are the generating functions of Y_t and X_t, then $G_{Y_t}(z) = Y_t$ and $G_{X_t}(z) = X_t$ because Y_t and X_t are 1×1 block matrices and the following equality holds

$$G_{Y_t}(z) = G_{A,t}(z)G_{X_t}(z).$$

We summarize the previous definitions in the following algebraic rule for manipulating finite generating matrices.

Basic Algebraic Rule *The equality $z^l z^k = z^{l+k}$ holds for all integers l and k and $z^0 = I$. Sums of generating functions and multiplication of a generating function by a constant are again generating functions. In addition, if $G_{M,t}(z) = M(t,1)z^{t-1} + \cdots + M(t,n)z^{t-n}$ is a generating function that represents the tth block row of an $m \times n$ block matrix M and X_t is a random vector, then*

$$\begin{aligned} z^k G_{M,t}(z) &= G_{M,t-k}(z)z^k \\ &= M(t-k,1)z^{t-1} + \cdots + M(t-k,n)z^{t-n} \\ z^k X_t &= X_{t-k} \end{aligned}$$

for all integer k, where it is understood that $z^k = 0$ if $k < t-n$ or $k > t-1$. Moreover, the previous operations should always be performed from left to right.

We finally note that, instead of writing $G_{M,t}(z)$ to denote the generating function corresponding to the tth row of a block matrix M, we will often write $M_t(z)$ to abbreviate the notation. Also, we will write $M(t,k)$ instead of $M(t,k)z^0$, where $M(t,k)$ is the (t,k)th entry of a block matrix M.

7.3.2 Innovations Representation

Suppose that we have a finite sample, $\{Y_t : 1 \leq t \leq n\}$, where Y_t is a zero mean random vector, and define $Y = (Y_1', \dots, Y_n')'$. Then, if $\mathrm{Var}(Y) = \Sigma_Y$, let $\Sigma_Y = \Psi\Sigma\Psi'$ be the block Cholesky decomposition of Σ_Y, where

$$\Psi = \begin{pmatrix} I & & & & \\ \Psi_{12} & I & & & \\ \Psi_{23} & \Psi_{13} & I & & \\ \vdots & \vdots & \vdots & \ddots & \\ \Psi_{n-1,n} & \Psi_{n-2,n} & \cdots & \Psi_{1,n} & I \end{pmatrix}$$

and $\Sigma = \mathrm{diag}(\Sigma_1, \dots, \Sigma_n)$. In finite samples, the Cholesky decomposition plays the role of the covariance generating function in infinite samples. It can also be used to obtain a representation of the process in terms of the innovations. To see this, define $E = \Psi^{-1}Y$, where

$$\Psi^{-1} = \begin{pmatrix} I & & & & \\ \overline{\Psi}_{12} & I & & & \\ \overline{\Psi}_{23} & \overline{\Psi}_{13} & I & & \\ \vdots & \vdots & \vdots & \ddots & \\ \overline{\Psi}_{n-1,n} & \overline{\Psi}_{n-2,n} & \cdots & \overline{\Psi}_{1,n} & I \end{pmatrix}$$

is the inverse of Ψ. Then, $E = (E_1', \dots, E_n')'$ is the vector of innovations because $\mathrm{Var}(E) = \Psi^{-1}\Sigma_Y\Psi^{-1'} = \Sigma$, $\mathrm{Var}(E_t) = \Sigma_t, t = 1, \dots, n$, the vectors E_t are serially uncorrelated, and we can write $Y = \Psi E, Y_t = E_t + \Psi_{1t}E_{t-1} + \cdots + \Psi_{t-1,t}E_1$, $t = 1, \dots, n$.

The innovations representation of Y motivates that we define for each observation, Y_t, the generating functions

$$\Psi_t(z) = I + \Psi_{1t}z + \cdots + \Psi_{t-1,t}z^{t-1} \tag{7.58}$$

and

$$\Psi_t^{-1}(z) = I + \overline{\Psi}_{1t}z + \cdots + \overline{\Psi}_{t-1,t}z^{t-1}, \tag{7.59}$$

that correspond to the tth rows of Ψ and Ψ^{-1}. Then, passing to generating functions in $Y = \Psi E$ and $E = \Psi^{-1}Y$, we can write

$$Y_t = \Psi_t(z)E_t$$

and

$$E_t = \Psi_t^{-1}(z)Y_t.$$

7.3.3 Covariance Generating Function

Let us return to the Cholesky decomposition, $\Sigma_Y = \Psi\Sigma\Psi'$, of the previous section and define $\gamma(t,s) = E(Y_tY_s')$ and the generating functions

$$\Psi_t'(z^{-1}) = I + \Psi_{1,t+1}'z^{-1} + \cdots + \Psi_{n-t,n}'z^{t-n} \tag{7.60}$$

and

$$G_{Y_t}(z) = \gamma(t,1)z^{t-1} + \gamma(t,2)z^{t-2} + \cdots + \gamma(t,t) + \cdots + \gamma(t,n)z^{t-n}, \tag{7.61}$$

that correspond to the tth rows of Ψ' and Σ_Y. Then, passing to generating functions in $\Sigma_Y = \Psi\Sigma\Psi'$ and applying the basic rule, it is obtained that

$$G_{Y_t}(z) = \Psi_t(z)\Sigma_t\Psi_t'(z^{-1}). \tag{7.62}$$

Example 7.3 Suppose $n = 3$. Then, to obtain $G_{Y_2}(z)$, we can perform the following calculation.

$$\begin{aligned}
G_{Y_2}(z) &= \Psi_2(z)\Sigma_2\Psi_2'(z^{-1}) \\
&= (I + \Psi_{12}z)\Sigma_2\Psi_2'(z^{-1}) \\
&= \Sigma_2\Psi_2'(z^{-1}) + \Psi_{12}\Sigma_1\Psi_1'(z^{-1})z \\
&= \Sigma_2(I + \Psi_{13}'z^{-1}) + \Psi_{12}\Sigma_1(I + \Psi_{12}'z^{-1} + \Psi_{23}'z^{-2})z \\
&= \Psi_{12}\Sigma_1 z + (\Sigma_2 + \Psi_{12}\Sigma_1\Psi_{12}') + (\Sigma_2\Psi_{13}' + \Psi_{12}\Sigma_1\Psi_{23}')z^{-1}.
\end{aligned}$$

Thus, $\gamma(2,1) = \Psi_{12}\Sigma_1$, $\gamma(2,2) = \Sigma_2 + \Psi_{12}\Sigma_1\Psi_{12}'$ and $\gamma(2,3) = \Sigma_2\Psi_{13}' + \Psi_{12}\Sigma_1\Psi_{23}'$. ◇

7.3.4 Inverse Process

The inverse process is defined as $Y^i = (\Psi')^{-1}\Sigma^{-1}\Psi^{-1}Y$. It is easy to see that $\mathrm{Var}(Y^i) = \Sigma_Y^{-1}$, which justifies the name. The inverse process plays a crucial role in Wiener–Kolmogorov filtering and smoothing as we will see.

We can associate with the tth rows of $(\Psi')^{-1}$ and Σ_Y^{-1} the generating functions

$$(\Psi')_t^{-1}(z^{-1}) = I + \overline{\Psi}'_{1,t+1}z^{-1} + \cdots + \overline{\Psi}'_{n-t,n}z^{t-n}$$

and

$$G_{Y_t^i}(z) = \gamma^i(t,1)z^{t-1} + \gamma^i(t,2)z^{t-2} + \cdots + \gamma^i(t,t) + \cdots + \gamma^i(t,n)z^{t-n},$$

where $\gamma^i(t,s) = E(Y_t^i Y_s^{i'})$. Then, applying the basic rule, we get

$$G_{Y_t^i}(z) = (\Psi')_t^{-1}(z^{-1})\Sigma_t^{-1}\Psi_t^{-1}(z). \tag{7.63}$$

Note that $G_{Y_t^i}(z) = [G_{Y_t}(z)]^{-1}$ and that $Y_t^i = G_{Y_t^i}(z)Y_t$.

7.3.5 Finite Wiener–Kolmogorov Filtering and Smoothing

Suppose that we have a finite sample, $\{(S_t', Y_t')' : 1 \leq t \leq n\}$, where S_t and Y_t are zero mean random vectors, and define $S = (S_1', \ldots, S_n')'$ and $Y = (Y_1', \ldots, Y_n')'$. Then, letting

$$\text{Var}\begin{pmatrix} S \\ Y \end{pmatrix} = \begin{pmatrix} \Sigma_S & \Sigma_{SY} \\ \Sigma_{YS} & \Sigma_Y \end{pmatrix},$$

if Σ_Y is nonsingular, the orthogonal projection, $E^*(S|Y)$, of S onto Y is given by $E^*(S|Y) = \Sigma_{SY}\Sigma_Y^{-1}Y$. In addition, $\text{MSE}[E^*(S|Y)] = \Sigma_S - \Sigma_{SY}\Sigma_Y^{-1}\Sigma_{YS}$.

Suppose we are interested in the weights of the matrix $W = \Sigma_{SY}\Sigma_Y^{-1}$. If $[\gamma_W(t,1), \ldots, \gamma_W(t,n)]$ is the tth row of W, that corresponds to the estimator $\hat{S}_{t|n} = E^*(S_t|Y)$, we can associate with it the generating function

$$G_{W,t}(z) = \gamma_W(t,1)z^{t-1} + \cdots + \gamma_W(t,n)z^{t-n}.$$

Also, if $[\gamma_{SY}(t,1), \ldots, \gamma_{SY}(t,n)]$ is the tth row of Σ_{SY}, we can associate with it the generating function

$$G_{SY,t}(z) = \gamma_{SY}(t,1)z^{t-1} + \cdots + \gamma_{SY}(t,n)z^{t-n}.$$

On the other hand, as we saw in the previous section, the generating function of the tth row of Σ_Y^{-1} is given by $[G_{Y_t}(z)]^{-1} = G_{Y_t^i}(z) = (\Psi')_t^{-1}(z^{-1})\Sigma_t^{-1}\Psi_t^{-1}(z)$.

Passing to generating functions in $W = \Sigma_{SY}\Sigma_Y^{-1}$ and applying the basic rule, we get

$$G_{W,t}(z) = G_{SY,t}(z)G_{Y_t^i}(z)$$
$$= G_{SY,t}(z)\left[G_{Y_t}(z)\right]^{-1}.$$

Thus,

$$\hat{S}_{t|n} = G_{SY,t}(z)\left[G_{Y_t}(z)\right]^{-1} Y_t. \tag{7.64}$$

Let $\Sigma_Y = \Psi\Sigma\Psi'$, as before, be the Cholesky decomposition of Σ_Y and suppose that we are interested in the estimator $\hat{S}_{t|t} = E^*(S_t|Y_1,\ldots,Y_t)$, $t = 1,\ldots,n$. Then, since the vector of innovations, $E = (E_1',\ldots,E_n')'$, is given by $E = \Psi^{-1}Y$, we can write $E^*(S|Y) = \Sigma_{SY}\Psi'^{-1}\Sigma^{-1}\Psi^{-1}Y = \Sigma_{SY}\Psi'^{-1}\Sigma^{-1}E$. Using generating functions, it follows from this and (7.64) that $\hat{S}_{t|n} = G_{SY,t}(z)(\Psi')_t^{-1}(z^{-1})\Sigma_t^{-1}E_t$. Since the innovations are orthogonal, by the law of iterated orthogonal projections, we get

$$\hat{S}_{t|t} = \left[G_{SY,t}(z)(\Psi')_t^{-1}(z^{-1})\right]_+ \Sigma_t^{-1}E_t$$
$$= \left[G_{SY,t}(z)(\Psi')_t^{-1}(z^{-1})\right]_+ \Sigma_t^{-1}\Psi_t^{-1}(z)Y_t,$$

where if $F(z) = \sum_{i=t-n}^{t-1} f_j z^j$, then $[F(z)]_+ = \sum_{i\geq 0} f_j z^j$.

Using similar arguments, it can be shown that the covariance generating functions, $G_{En}(z)$ and $G_{Et}(z)$, of the errors $E_{t|n} = S_t - \hat{S}_{t|n}$ and $E_{t|t} = S_t - \hat{S}_{t|t}$ are given by

$$G_{En}(z) = G_{S,t}(z) - G_{SY,t}(z)\left[G_{Y_t}(z)\right]^{-1} G_{YS,t}(z^{-1})$$

and

$$G_{Et}(z) = G_{En}(z) + \Omega_t(z)\Sigma_t^{-1}\Omega_t'(z^{-1}),$$

where $G_{S,t}(z)$ and $G_{YS,t}(z)$ are the generating functions corresponding to the tth rows of Σ_S and Σ_{YS} and $\Omega_t(z) = \left[G_{SY,t}(z)(\Psi')_t^{-1}(z^{-1})\right]_+$. Here, by the basic rule, $z^k G_{YS,t}(z^{-1}) = G_{YS,t-k}(z^{-1})z^k$ and $z^k\Omega_t(z^{-1}) = \Omega_{t-k}(z^{-1})z^k$ for all integer k and $z^k = 0$ if $k < t-n$ or $k > t-1$.

7.3.6 *Finite Wiener–Kolmogorov Filtering for Multivariate Processes with State Space Structure*

To proceed further, we need to assume that the data have some structure. In particular, we will assume that the data have been generated by a state space model to be specified later.

The celebrated Kalman filter (Kalman, 1960b) has become the standard tool to apply when the sample is finite and the data have state space structure. As we have shown earlier in this section, the structure of the covariance matrix of the data can be studied by means of generating functions that, contrary to the infinite sample case, are time dependent. We will show that we can obtain Wiener–Kolmogorov formulae for filtering and smoothing in terms of these generating functions and that the Wiener–Kolmogorov formulae for the infinite sample are limits of the finite sample formulae. We will also show the equivalence between Wiener–Kolmogorov filtering and smoothing and Kalman filtering and smoothing.

The Wiener–Kolmogorov approach allows for the possibility of obtaining the filter weights. It is also more amenable than the Kalman filter approach to analytic investigation.

As we saw earlier in the book, the Kalman filter can be applied even when the system matrices are time variant. Suppose that $\{Y_t\}$ is a multivariate process that admits the state space representation (4.1) and (4.2) and assume that $E(x_1) = 0$ and $\mathrm{Var}(x_1) = \Pi$.

7.3.6.1 Covariance Generating Function of the Process

Before computing the covariance generating function $G_{Y_t}(z)$, given by (7.61), when $\{Y_t\}$ follows the state space model (4.1) and (4.2), we need several definitions. Iterating in (4.1), we get $x_t = \bar{u}_{t-1} + F^t_{t-1}\bar{u}_{t-2} + \cdots + F^t_2\bar{u}_1 + F^t_1 x_1$, where $\bar{u}_t = G_t u_t$ for $t = 1, \ldots, n$. Thus, if we stack the vectors x_t, we get

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} I & & & & \\ F^2_1 & I & & & \\ F^3_1 & F^3_2 & I & & \\ \vdots & \vdots & \vdots & \ddots & \\ F^n_1 & F^n_2 & \cdots & F^n_{n-1} & I \end{bmatrix} \begin{bmatrix} x_1 \\ \bar{u}_1 \\ \bar{u}_2 \\ \vdots \\ \bar{u}_{n-1} \end{bmatrix} \tag{7.65}$$

$$= \begin{bmatrix} I \\ F^2_1 \\ F^3_1 \\ \vdots \\ F^{n-1}_1 \\ F^n_1 \end{bmatrix} x_1 + \mathcal{Z} \begin{bmatrix} I & & & & & \\ F^3_2 & I & & & & \\ F^4_2 & F^4_3 & I & & & \\ \vdots & \vdots & \vdots & \ddots & & \\ F^n_2 & F^n_3 & \cdots & F^n_{n-1} & I & \\ 0 & 0 & \cdots & 0 & 0 & I \end{bmatrix} \begin{bmatrix} \bar{u}_1 \\ \bar{u}_2 \\ \bar{u}_3 \\ \vdots \\ \bar{u}_{n-1} \\ \bar{u}_n \end{bmatrix}, \tag{7.66}$$

where $\mathcal{Z}$ is the shift matrix

$$\mathcal{Z} = \begin{bmatrix} 0 & & & & \\ I & 0 & & & \\ 0 & I & 0 & & \\ \vdots & \vdots & \vdots & \ddots & \\ 0 & 0 & \cdots & I & 0 \end{bmatrix},$$

such that $\mathcal{Z}Y = (0, Y_1', \ldots, Y_{n-1}')'$. Define

$$\overline{F} = \begin{bmatrix} I & & & & & \\ F_2^3 & I & & & & \\ F_2^4 & F_3^4 & I & & & \\ \vdots & \vdots & \vdots & \ddots & & \\ F_2^n & F_3^n & \cdots & F_{n-1}^n & I & \\ 0 & 0 & \cdots & 0 & 0 & I \end{bmatrix}$$

and let $F_t(z) = I + F_t^{t+1}z + \cdots + F_2^{t+1}z^{t-1}$ be the generating function that represents the tth row of $\overline{F}$, $t = 1, 2, \ldots, n$. Then, it is not difficult to verify that

$$\overline{F}^{-1} = \begin{bmatrix} I & & & & & \\ -F_2^3 & I & & & & \\ & -F_3^4 & I & & & \\ & & \ddots & \ddots & & \\ & & & -F_{n-1}^n & I & \\ & & & & 0 & I \end{bmatrix}$$

and, therefore, $F_t^{-1}(z) = I - F_t^{t+1}z = I - F_t z$ and

$$(I - F_t z)^{-1} = I + F_t^{t+1}z + \cdots + F_2^{t+1}z^{t-1}, \tag{7.67}$$

where $(I - F_1 z)^{-1} = I$. Passing to generating functions in (7.66) and using (7.67), we can write

$$x_t = \begin{bmatrix} (I - F_{t-1}z)^{-1}z & F_1^t \end{bmatrix} \begin{bmatrix} G_t u_t \\ x_1 \end{bmatrix}, \tag{7.68}$$

where we define $(I - F_0 z)^{-1} = 0$. From this, it is obtained that

$$\begin{aligned} Y_t &= H_t F_1^t x_1 + H_t (I - F_{t-1}z)^{-1} z \bar{u}_t + v_t \\ &= H_t F_1^t x_1 + \begin{bmatrix} H_t (I - F_{t-1}z)^{-1} z & I \end{bmatrix} \begin{bmatrix} G_t u_t \\ v_t \end{bmatrix}. \end{aligned}$$

Equivalently, stack the observations Y_t to get

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} H_1 \\ H_2F_1^2 \\ H_3F_1^3 \\ \vdots \\ H_nF_1^n \end{bmatrix} x_1 + \begin{bmatrix} 0 & I \\ H_2 & 0 \\ H_3F_2^3 & 0 \\ \vdots & \vdots \\ H_nF_2^n & 0 \end{bmatrix} \begin{bmatrix} G_1u_1 \\ v_1 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & I \\ H_3 & 0 \\ \vdots & \vdots \\ H_nF_3^n & 0 \end{bmatrix} \begin{bmatrix} G_2u_2 \\ v_2 \end{bmatrix}$$

$$+ \cdots + \begin{bmatrix} 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \\ 0 & I \\ H_n & 0 \end{bmatrix} \begin{bmatrix} G_{n-1}u_{n-1} \\ v_{n-1} \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \\ 0 & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} G_nu_n \\ v_n \end{bmatrix}$$

$$= \begin{bmatrix} H_1 \\ H_2F_1^2 \\ H_3F_1^3 \\ \vdots \\ H_nF_1^n \end{bmatrix} x_1 + \mathcal{Z}^0 \begin{bmatrix} 0 & I \\ H_2 & 0 \\ H_3F_2^3 & 0 \\ \vdots & \vdots \\ H_nF_2^n & 0 \end{bmatrix} \begin{bmatrix} G_1u_1 \\ v_1 \end{bmatrix} + \mathcal{Z}^1 \begin{bmatrix} 0 & I \\ H_3 & 0 \\ \vdots & \vdots \\ H_nF_3^n & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} G_2u_2 \\ v_2 \end{bmatrix}$$

$$+ \cdots + \mathcal{Z}^{n-2} \begin{bmatrix} 0 & I \\ H_n & 0 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \end{bmatrix} \begin{bmatrix} G_{n-1}u_{n-1} \\ v_{n-1} \end{bmatrix} + \mathcal{Z}^{n-1} \begin{bmatrix} 0 & I \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} G_nu_n \\ v_n \end{bmatrix}.$$

Letting $F_t'(z^{-1}) = I + F_{t+1}'^{t+2} z^{-1} + \cdots + F_{t+1}'^{n} z^{t-n+1}$ represent the tth row of $\overline{F}'$ and proceeding as we did earlier in this section to obtain (7.67), we get $F_t'^{-1}(z^{-1}) = I - F_{t+1}'^{t+2} z^{-1} = I - F_{t+1}' z^{-1}$ and

$$(I - F_{t+1}' z^{-1})^{-1} = I + F_{t+1}'^{t+2} z^{-1} + \cdots + F_{t+1}'^{n} z^{t-n+1}, \tag{7.69}$$

where $(I - F_n' z^{-1})^{-1} = I$.

To compute a first expression for $G_{Y_t}(z)$, define, as in Sect. 1.7.2, $\Pi_t = \mathrm{Var}(x_t)$. Then, by Lemma 1.1, Π_t satisfies the recursion $\Pi_{t+1} = F_t\Pi_tF_t' + G_tQ_tG_t'$, initialized with $\Pi_1 = \Pi$. Using this recursion, the following theorem is obtained.

Theorem 7.19 *The covariance generating function, $G_{Y_t}(z)$, of Y_t, where $\{Y_t\}$ follows the state space model (4.1) and (4.2), is given by*

$$\begin{aligned} G_{Y_t}(z) &= R_t + H_t\Pi_t H_t' + H_t(I - F_{t-1}z)^{-1}N_{t-1}z + N_t'(I - F_{t+1}'z^{-1})^{-1}H_{t+1}'z^{-1} \\ &= \left[H_t(I - F_{t-1}z)^{-1}z \quad I\right]\begin{bmatrix} 0 & N_t \\ N_t' & R_t + H_t\Pi_t H_t' \end{bmatrix}\begin{bmatrix} (I - F_{t+1}'z^{-1})^{-1}H_{t+1}'z^{-1} \\ I \end{bmatrix}, \end{aligned}$$

where $\Pi_t = \mathrm{Var}(x_t)$ satisfies the recursion $\Pi_{t+1} = F_t\Pi_t F_t' + G_tQ_tG_t'$, initialized with $\Pi_1 = \Pi$, $N_t = F_t\Pi_t H_t' + G_tS_t$, $(I - F_{t-1}z)^{-1} = I + F_{t-1}^t z + \cdots + F_2^t z^{t-2}$ and $(I - F_{t+1}'z^{-1})^{-1} = I + F_{t+1}^{\prime t+2}z^{-1} + \cdots + F_{t+1}^{\prime n}z^{t+1-n}$.

Proof Substituting (1.34) in (7.61) yields

$$\begin{aligned} G_{Y_t}(z) &= H_tF_2^tN_1z^{t-1} + H_tF_3^tN_2z^{t-2} + \cdots + H_tF_t^tN_{t-1}z + R_t + H_t\Pi_tH_t' \\ &\quad + N_t'F_{t+1}^{t+1'}H_{t+1}'z^{-1} + \cdots + N_t'F_{t+1}^{n'}H_n'z^{t-n} \\ &= R_t + H_t\Pi_tH_t' + H_t\left[F_t^tzN_t + \cdots + F_3^tz^{t-2}N_t + F_2^tz^{t-1}N_t\right] \\ &\quad + N_t'\left[F_{t+1}^{t+1'}z^{-1}H_t' + \cdots + F_{t+1}^{n'}z^{t-n}H_t'\right] \\ &= R_t + H_t\Pi_tH_t' + H_t(I - F_{t-1}z)^{-1}N_{t-1}z + N_t'(I - F_{t+1}'z^{-1})^{-1}H_{t+1}'z^{-1}. \end{aligned}$$

□

To obtain an alternative expression for $G_{Y_t}(z)$, we can pass to generating functions in (7.65). First, note that

$$\overline{F}_1 = \begin{bmatrix} I & & & & \\ F_1^2 & I & & & \\ F_1^3 & F_2^3 & I & & \\ \vdots & \vdots & \vdots & \ddots & \\ F_1^n & F_2^n & \cdots & F_{n-1}^n & I \end{bmatrix} = \begin{bmatrix} I & & & & \\ -F_1^2 & I & & & \\ & -F_2^3 & I & & \\ & & \ddots & \ddots & \\ & & & -F_{n-1}^n & I \end{bmatrix}^{-1}.$$

Then, if we define $u_0 = x_1$ and $G_0 = I$, we can write

$$\begin{aligned} x_t &= \bar{u}_{t-1} + F_{t-1}^t\bar{u}_{t-2} + \cdots + F_2^t\bar{u}_1 + F_1^t\bar{u}_0 \\ &= \left(I + F_{t-1}^tz + \cdots + F_2^tz^{t-2} + F_1^tz^{t-1}\right)\bar{u}_{t-1}, \end{aligned}$$

where $\bar{u}_t = G_tu_t$ for $t = 1, \ldots, n$. Letting $F_{t,1}(z) = I + F_{t-1}^tz + \cdots + F_2^tz^{t-2} + F_1^tz^{t-1}$ be the generating function that represents the tth row of $\overline{F}_1$, $t = 1, 2, \ldots, n$, we get $F_{t,1}^{-1}(z) = I - F_{t-1}^tz = I - F_{t-1}z$ and

$$(I - F_{t-1}z)_1^{-1} = I + F_{t-1}^tz + \cdots + F_2^tz^{t-2} + F_1^tz^{t-1}, \tag{7.70}$$

where we define $(I - F_0 z)_1^{-1} = I$ and we use the subindex 1 to differentiate this generating function from (7.67). Note that

$$(I - F_{t-1}z)_1^{-1} = (I - F_{t-1}z)^{-1} + F_1^t z^{t-1}. \tag{7.71}$$

Using (7.70), we can write

$$\begin{aligned} Y_t &= H_t F_1^t x_1 + H_t (I - F_{t-1}z)^{-1} z\bar{u}_t + v_t \\ &= \begin{bmatrix} H_t (I - F_{t-1}z)_1^{-1} z & I \end{bmatrix} \begin{bmatrix} G_t u_t \\ v_t \end{bmatrix}. \end{aligned}$$

Letting $F'_{t,1}(z^{-1}) = I + F_t'^{t+1} z^{-1} + \cdots + F_t'^{n} z^{t-n}$ represent the tth row of $\overline{F}_1'$ and proceeding as we did earlier in this section to obtain (7.70), we get $F'^{-1}_{t,1}(z^{-1}) = I - F_t'^{t+1} z^{-1} = I - F'_t z^{-1}$ for $t = 1, 2, \ldots, n$. It follows from this that

$$(I - F'_t z^{-1})^{-1} = I + F_t'^{t+1} z^{-1} + \cdots + F_t'^{n} z^{t-n}, \qquad t = 2, 3, \ldots, n,$$

coincides with (7.69) and we get the new expression

$$(I - F'_1 z^{-1})^{-1} = I + F_1'^{2} z^{-1} + \cdots + F_1'^{n} z^{1-n}. \tag{7.72}$$

The following theorem gives an expression for $G_{Y_t}(z)$ different from that of Theorem 7.19.

Theorem 7.20 *The covariance generating function, $G_{Y_t}(z)$, of Y_t, where $\{Y_t\}$ follows the state space model (4.1) and (4.2), is given by*

$$\begin{aligned} G_{Y_t}(z) &= H_t F_1^t \Pi \left[H'_1 z^{t-1} + F_1^{2'} H'_2 z^{t-2} + \cdots + F_1^{n'} H'_n z^{t-n} \right] \\ &\quad + \begin{bmatrix} H_t (I - F_{t-1}z)^{-1} z & I \end{bmatrix} \begin{bmatrix} G_t Q_t G'_t & G_t S_t \\ S'_t G'_t & R_t \end{bmatrix} \begin{bmatrix} (I - F'_{t+1} z^{-1})^{-1} H'_{t+1} z^{-1} \\ I \end{bmatrix}, \end{aligned} \tag{7.73}$$

where $\Pi_t = \mathrm{Var}(x_t)$ satisfies the recursion $\Pi_{t+1} = F_t \Pi_t F'_t + G_t Q_t G'_t$, initialized with $\Pi_1 = \Pi$, $(I - F_{t-1}z)^{-1} = I + F_{t-1}^t z + \cdots + F_2^t z^{t-2}$ and $(I - F'_{t+1} z^{-1})^{-1} = I + F_{t+1}'^{t+2} z^{-1} + \cdots + F_{t+1}'^{n} z^{t+1-n}$. Using (7.70) and (7.72), $G_{Y_t}(z)$ can be expressed as

$$\begin{aligned} G_{Y_t}(z) &= H_t F_1^t \Pi \left(I - F'_1 z^{-1} \right)^{-1} H'_1 z^{t-1} \\ &\quad + \begin{bmatrix} H_t (I - F_{t-1}z)^{-1} z & I \end{bmatrix} \begin{bmatrix} G_t Q_t G'_t & G_t S_t \\ S'_t G'_t & R_t \end{bmatrix} \begin{bmatrix} (I - F'_{t+1} z^{-1})^{-1} H'_{t+1} z^{-1} \\ I \end{bmatrix} \\ &= \begin{bmatrix} H_t (I - F_{t-1}z)_1^{-1} z & I \end{bmatrix} \begin{bmatrix} G_t Q_t G'_t & G_t S_t \\ S'_t G'_t & R_t \end{bmatrix} \begin{bmatrix} (I - F'_{t+1} z^{-1})^{-1} H'_{t+1} z^{-1} \\ I \end{bmatrix}, \end{aligned}$$

where $Q_0 = \Pi$, $R_0 = 0$, $S_0 = 0$, $F_0 = 0$, $H_0 = 0$ and $\Pi_0 = 0$.

Proof Using the recursion $\Pi_{t+1} = F_t\Pi_t F_t' + G_t Q_t G_t'$, initialized with $\Pi_1 = \Pi$, we get

$$\begin{aligned}
&(I - F_{t-1}z)_1^{-1} G_{t-1}Q_{t-1}G_{t-1}' \times (I - F_t'z^{-1})^{-1} \\
&= (I - F_{t-1}z)_1^{-1}(\Pi_t - F_{t-1}\Pi_{t-1}F_{t-1}')(I - F_t'z^{-1})^{-1} \\
&= (I - F_{t-1}z)_1^{-1}\left[\Pi_t(I - F_t'z^{-1}) + (I - F_{t-1}z)\Pi_t\right. \\
&\quad \left. -(I - F_{t-1}z)\Pi_t(I - F_t'z^{-1})\right](I - F_t'z^{-1})^{-1} \\
&= (I - F_{t-1}z)_1^{-1}\Pi_t + \Pi_t(I - F_t'z^{-1})^{-1} - \Pi_t \\
&= (I - F_{t-1}z)_1^{-1}\Pi_t + \Pi_t F_t'(I - F_{t+1}'z^{-1})^{-1}z^{-1}.
\end{aligned}$$

On the other hand,

$$\begin{aligned}
H_t(I - F_{t-1}z)_1^{-1}G_{t-1}S_{t-1}z &= H_t(I - F_{t-1}z)_1^{-1}(\Pi_t H_t' + G_{t-1}S_{t-1}z) \\
&\quad -H_t(I - F_{t-1}z)_1^{-1}\Pi_t H_t' \\
&= H_t(I - F_{t-1}z)_1^{-1}\left[(I - F_{t-1}z)\Pi_t H_t' + F_{t-1}\Pi_{t-1}H_{t-1}'z\right. \\
&\quad \left. +G_{t-1}S_{t-1}z\right] - H_t(I - F_{t-1}z)_1^{-1}\Pi_t H_t' \\
&= H_t\Pi_t H_t' + H_t(I - F_{t-1}z)_1^{-1}(F_{t-1}\Pi_{t-1}H_{t-1}' + G_{t-1}S_{t-1})z \\
&\quad -H_t(I - F_{t-1}z)_1^{-1}\Pi_t H_t'.
\end{aligned}$$

Letting $N_t = F_t\Pi_t H_t' + G_t S_t$, it follows from this, (7.71), and Theorem 7.19 that

$$\begin{aligned}
&\left[H_t(I - F_{t-1}z)_1^{-1}z \quad I\right]\begin{bmatrix} G_tQ_tG_t' & G_tS_t \\ S_t'G_t' & R_t \end{bmatrix}\begin{bmatrix} (I - F_{t+1}'z^{-1})^{-1}H_{t+1}'z^{-1} \\ I \end{bmatrix} \\
&= H_t\Pi_t H_t' + R_t + H_t(I - F_{t-1}z)_1^{-1}N_{t-1}z + N_t'(I - F_{t+1}'z^{-1})^{-1}H_{t+1}'z^{-1} \\
&= H_t\Pi_t H_t' + R_t + H_t(I - F_{t-1}z)^{-1}N_{t-1}z + H_tF_1^tz^tN_t + N_t'(I - F_{t+1}'z^{-1})^{-1}H_{t+1}'z^{-1} \\
&= H_t\Pi_t H_t' + R_t + H_t(I - F_{t-1}z)^{-1}N_{t-1}z + N_t'(I - F_{t+1}'z^{-1})^{-1}H_{t+1}'z^{-1} \\
&= G_{Y_t}(z),
\end{aligned}$$

where we have used that $z^tN_t = N_0z^t = 0$. □

The expressions for the covariance generating function, $G_{Y_t}(z)$, of Y_t in Theorems 7.19 and 7.20 correspond to two decompositions of the covariance matrix Σ_Y, namely,

$$\Sigma_Y = \sum_{t=1}^{n} \mathcal{Z}^{t-1} \begin{bmatrix} 0 & I \\ H_{t+1} & 0 \\ H_{t+2}F_{t+1}^{t+2} & 0 \\ \vdots & \vdots \\ H_n F_{t+1}^{n} & 0 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & N_t \\ N_t' & R_t + H_t\Pi_t H_t' \end{bmatrix} \begin{bmatrix} 0 & I \\ H_{t+1} & 0 \\ H_{t+2}F_{t+1}^{t+2} & 0 \\ \vdots & \vdots \\ H_n F_{t+1}^{n} & 0 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \end{bmatrix}' \mathcal{Z}^{t-1\prime}$$

and

$$\Sigma_Y = O\Pi O' + \sum_{t=1}^{n} \mathcal{Z}^{t-1} \begin{bmatrix} 0 & I \\ H_{t+1} & 0 \\ H_{t+2}F_{t+1}^{t+2} & 0 \\ \vdots & \vdots \\ H_n F_{t+1}^{n} & 0 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \end{bmatrix} \begin{bmatrix} G_t Q_t G_t' & G_t S_t \\ S_t' G_t' & R_t \end{bmatrix} \begin{bmatrix} 0 & I \\ H_{t+1} & 0 \\ H_{t+2}F_{t+1}^{t+2} & 0 \\ \vdots & \vdots \\ H_n F_{t+1}^{n} & 0 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \end{bmatrix}' \mathcal{Z}^{t-1\prime},$$

where

$$O = \begin{bmatrix} H_1 \\ H_2 F_1^2 \\ H_3 F_1^3 \\ \vdots \\ H_n F_1^n \end{bmatrix}.$$

7.3.6.2 Covariance Factorization

Let $\{Y_t\}$ follow the state space model (4.1) and (4.2). The covariance factorization of the time invariant case has its counterpart in the Cholesky decomposition of the covariance matrix, $\Sigma_Y = \Psi\Sigma\Psi'$. As we saw earlier in Sect. 7.3.3, the factorization of the covariance generating function (7.62) corresponds to this Cholesky decomposition.

As in the time invariant case, if we can find matrices Σ_t and K_t such that $G_{Y_t}(z)$ factorizes as

$$\begin{aligned} G_{Y_t}(z) &= \left[H_t(I - F_{t-1}z)^{-1}z \quad I\right] \begin{bmatrix} K_t \\ I \end{bmatrix} \Sigma_t \, [K_t' \quad I] \begin{bmatrix} (I - F_{t+1}'z^{-1})^{-1}H_{t+1}'z^{-1} \\ I \end{bmatrix} \\ &= [I + H_t(I - F_{t-1}z)^{-1}K_{t-1}z]\Sigma_t \, [I + K_t'(I - F_{t+1}'z^{-1})^{-1}H_{t+1}'z^{-1}], \end{aligned} \tag{7.74}$$

then we can obtain the desired factorization by defining $\Psi_t(z) = I + H_t(I - F_{t-1}z)^{-1}K_{t-1}z$ and $\Psi_t'(z^{-1}) = I + K_t'(I - F_{t+1}'z^{-1})^{-1}H_{t+1}'z^{-1}$.

The following lemma, that is the exact analog to Lemma 5.6, will be useful to obtain the factorization.

Lemma 7.8 *The covariance matrix, Σ_Y, of Y_t, where $\{Y_t\}$ follows the state space model (4.1) and (4.2) is invariant under transformations of the form*

$$\Pi \longrightarrow \Pi - Z_1$$

and

$$\begin{bmatrix} G_tQ_tG_t' & G_tS_t \\ S_t'G_t' & R_t \end{bmatrix} \longrightarrow \begin{bmatrix} -Z_{t+1} + F_tZ_tF_t' + G_tQ_tG_t' & F_tZ_tH_t' + G_tS_t \\ H_tZ_tF_t' + S_t'G_t' & R_t + H_tZ_tH_t' \end{bmatrix},$$

for any sequence of symmetric matrices $\{Z_t\}$.

Proof Given a sequence of symmetric matrices, $\{Z_t\}$, the lemma will be proved if we prove that the difference, $\Delta G_{Y_t}(z)$, between the covariance generating function (7.73) and the covariance generating function obtained after replacing in (7.73) $G_tQ_tG_t'$, G_tS_t, R_t and Π_1 by $-Z_{t+1} + F_tZ_tF_t' + G_tQ_tG_t'$, $F_tZ_tH_t' + G_tS_t$, $H_tZ_tH_t' + R_t$ and $\Pi_1 - Z_1$, respectively, is zero. That is, if we prove that

$$\begin{aligned} &\Delta G_{Y_t}(z) \\ &= \left[H_t(I - F_{t-1}z)_1^{-1}z \quad I\right] \begin{bmatrix} -Z_{t+1} + F_tZ_tF_t' & F_tZ_tH_t' \\ H_tZ_tF_t' & H_tZ_tH_t' \end{bmatrix} \begin{bmatrix} (I - F_{t+1}'z^{-1})^{-1}H_{t+1}'z^{-1} \\ I \end{bmatrix} \\ &= 0, \end{aligned}$$

where $F_0 = 0$, $H_0 = 0$ and $Z_0 = 0$, for $t = 1, 2, \ldots, n$. This expression for $\Delta G_{Y_t}(z)$ corresponds to the covariance generating function, as given by Theorem 7.19, of a state space model of the form

$$\begin{aligned} x_{t+1} &= F_tx_t \\ Y_t &= H_tx_t, \end{aligned}$$

where $\mathrm{Var}(x_t) = Z_t$, $t = 1, 2, \ldots, n$, because in this model $Z_{t+1} = F_t Z_t F_t'$, $N_t = F_t Z_t H_t'$ and, since $N_0 = 0$, $(I - F_{t-1}z)_1^{-1} z N_t = (I - F_{t-1}z)^{-1} z N_t$. Thus, applying Theorem 7.20 to the previous state space model, $\Delta G_{Y_t}(z)$ can be written as

$$\Delta G_{Y_t}(z) = \begin{bmatrix} H_t(I - F_{t-1}z)_1^{-1}z & I \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} (I - F_{t+1}'z^{-1})^{-1} H_{t+1}' z^{-1} \\ I \end{bmatrix} = 0$$

for $t = 1, 2, \ldots, n$. □

We would like to choose $\{Z_t\} = \{P_t\}$ in such a manner that the center matrices

$$C_t = \begin{bmatrix} -P_{t+1} + F_t P_t F_t' + G_t Q_t G_t' & F_t P_t H_t' + G_t S_t \\ H_t P_t F_t' + S_t' G_t' & R_t + H_t P_t H_t' \end{bmatrix}$$

factorize as

$$C_t = \begin{bmatrix} C_{1,t} \\ C_{2,t} \end{bmatrix} \begin{bmatrix} C_{1,t}' & C_{2,t}' \end{bmatrix}$$

for some matrices $C_{1,t}$ and $C_{2,t}$ of appropriate dimensions, $t = 1, 2, \ldots, n$.

Assuming that $R_t + H_t P_t H_t'$ is nonsingular, we can consider the Schur decomposition of C_t,

$$C_t = \begin{bmatrix} I & X_t \\ 0 & I \end{bmatrix} \begin{bmatrix} \Delta_t & 0 \\ 0 & R_t + H_t P_t H_t' \end{bmatrix} \begin{bmatrix} I & 0 \\ X_t' & I \end{bmatrix}, \tag{7.75}$$

where

$$X_t = (F_t P_t H_t' + G_t S_t)(R_t + H_t P_t H_t')^{-1}$$

and Δ_t is the Schur complement

$$\Delta_t = -P_{t+1} + F_t P_t F_t' + G_t Q_t G_t' - X_t(R_t + H_t P_t H_t')X_t'.$$

Equation (7.75) shows that if we choose P_t so as to make Δ_t zero, that is,

$$-P_{t+1} + F_t P_t F_t' + G_t Q_t G_t' - X_t(R_t + H_t P_t H_t')X_t' = 0,$$

then C_t can be expressed as

$$C_t = \begin{bmatrix} F_t P_t H_t' + G_t S_t \\ R_t + H_t P_t H_t' \end{bmatrix} (R_t + H_t P_t H_t')^{-1} \begin{bmatrix} F_t P_t H_t' + G_t S_t \\ R_t + H_t P_t H_t' \end{bmatrix}',$$

obtaining the required factorization. Note that the recursion for P_t is simply the Kalman filter recursion for the MSE of the predictor, $\hat{x}_{t|t-1}$, of x_t. We can therefore identify

$$C_{1,t} = \Sigma_t^{1/2}, \qquad C_{2,t} = K_t\Sigma_t^{1/2},$$

where $K_t = (F_tP_tH_t' + G_tS_t)\Sigma_t^{-1}$ and $\Sigma_t = R_t + H_tP_tH_t'$ are the Kalman gain and the innovations covariance matrix of the Kalman filter. We have thus the following result.

Theorem 7.21 *If $\{Y_t\}$ follows the state space model (4.1) and (4.2) and the covariance matrix Σ_Y is nonsingular, the covariance generating function, $G_{Y_t}(z)$, of Y_t factorizes as*

$$G_{Y_t}(z) = [I + H_t(I - F_{t-1}z)^{-1}K_{t-1}z]\Sigma_t\, [I + K_t'(I - F_{t+1}'z^{-1})^{-1}H_{t+1}'z^{-1}],$$

where P_t, $K_t = (F_tP_tH_t' + G_tS_t)\Sigma_t^{-1}$ and $\Sigma_t = R_t + H_tP_tH_t'$ are obtained with the same recursions as those of the Kalman filter.

In terms of the covariance matrix Σ_Y, the factorization of the previous theorem can be expressed as

$$\Sigma_Y = \sum_{t=1}^{n} \mathcal{Z}^{t-1} \begin{bmatrix} I \\ H_{t+1}K_t \\ H_{t+2}F_{t+1}^{t+2}K_t \\ \vdots \\ H_nF_{t+1}^{n}K_t \\ 0 \\ \vdots \\ 0 \end{bmatrix} \Sigma_t \begin{bmatrix} I \\ H_{t+1}K_t \\ H_{t+2}F_{t+1}^{t+2}K_t \\ \vdots \\ H_nF_{t+1}^{n}K_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}' \mathcal{Z}^{t-1\prime}.$$

If in Lemma 7.8 we choose $Z_1 = \Pi$ and $Z_t = \Pi_t = \mathrm{Var}(x_t)$, so that $Z_t = P_t + \overline{\Sigma}_t$, where P_t is as earlier in this section, $P_t = \mathrm{Var}(x_t - \hat{x}_{t|t-1})$, and, therefore, $\overline{\Sigma}_t = \mathrm{Var}(\hat{x}_{t|t-1})$, after a little algebra, the following theorem is obtained. The proof is omitted.

Theorem 7.22 *The covariance generating function, $G_{Y_t}(z)$, of Y_t, where $\{Y_t\}$ follows the state space model (4.1) and (4.2), is given by*

$$\begin{aligned} G_{Y_t}(z) &= \Sigma_t + H_t\overline{\Sigma}_tH_t' + H_t(I - F_{t-1}z)^{-1}(F_{t-1}\overline{\Sigma}_{t-1}H_{t-1}' + K_{t-1}\Sigma_{t-1})z \\ &\quad + (H_t\overline{\Sigma}_tF_t' + \Sigma_tK_t')(I - F_{t+1}'z^{-1})^{-1}H_{t+1}'z^{-1} \\ &= \left[H_t(I - F_{t-1}z)^{-1}z \quad I\right] \begin{bmatrix} 0 & N_t \\ N_t' & \Sigma_t + H_t\overline{\Sigma}_tH_t' \end{bmatrix} \begin{bmatrix} (I - F_{t+1}'z^{-1})^{-1}H_{t+1}'z^{-1} \\ I \end{bmatrix} \end{aligned}$$

where $\overline{\Sigma}_t = \mathrm{Var}(\hat{x}_{t|t-1})$ *satisfies the recursion* $\overline{\Sigma}_{t+1} = F_t\overline{\Sigma}_t F_t' + K_t\Sigma_t K_t'$*, initialized with* $\overline{\Sigma}_1 = 0$*,* $N_t = F_t\overline{\Sigma}_t H_t' + K_t\Sigma_t$ *and* $(I - F_{t-1}z)^{-1}$ *and* $(I - F_{t+1}'z^{-1})^{-1}$ *are given by (7.67) and (7.69).*

7.3.6.3 Innovations Representation

Consider the decomposition (7.62) of the generating function $G_{Y_t}(z)$ of Y_t, where $\{Y_t\}$ follows the state space model (4.1) and (4.2). By Theorem 7.21, it is clear that, when there is state space structure, the function (7.58) is given by $\Psi_t(z) = I + H_t(I - F_{t-1}z)^{-1}K_{t-1}z$ and that this generating function corresponds to the tth row of the Cholesky factor, Ψ, in the Cholesky decomposition $\Sigma_Y = \Psi\Sigma\Psi'$. Also, the function $\Psi_t'(z^{-1}) = I + K_t'(I - F_{t+1}'z^{-1})^{-1}H_{t+1}'z^{-1}$ corresponds to the tth row of Ψ'.

Letting $Y = (Y_1', \ldots, Y_n')'$ and $E = (E_1', \ldots, E_n')'$ be the stacked vectors of observations and innovations, respectively, from the block Cholesky decomposition of Σ_Y, $\Sigma_Y = \Psi\Sigma\Psi'$, we have the relation $Y = \Psi E$. Passing to generating functions, this equation implies

$$\begin{aligned} Y_t &= \Psi_t(z)E_t \\ &= \left[I + H_t(I - F_{t-1}z)^{-1}K_{t-1}z\right]E_t \\ &= H_t(I - F_{t-1}z)^{-1}K_{t-1}E_{t-1} + E_t. \end{aligned}$$

Letting $\theta_t = (I - F_{t-1}z)^{-1}K_{t-1}E_{t-1}$ and using the definition of $(I - F_t z)^{-1}$ and the basic algebraic rule, it is easy to verify that θ_t satisfies the recursion $\theta_{t+1} = F_t\theta_t + K_tE_t$. Thus, by the covariance factorization of Theorem 7.21, we see that $\{Y_t\}$ is generated by the state space model

$$\theta_{t+1} = F_t\theta_t + K_tE_t \tag{7.76}$$

$$Y_t = H_t\theta_t + E_t, \tag{7.77}$$

where E_t is the tth innovation, $\theta_1 = 0$ and P_t, $K_t = (F_tP_tH_t' + G_tS_t)\Sigma_t^{-1}$ and $\Sigma_t = R_t + H_tP_tH_t'$ are obtained with the same recursions as those of the Kalman filter. Therefore, (7.76) and (7.77) is the innovations representation of the Kalman filter!

Note that, from (7.76) and (7.77), the generating function of θ_t is $G_{\theta_t}(z) = z(I - F_tz)^{-1}K_t$ and, therefore, we can write $\Psi_t(z) = I + H_tG_{\theta_t}(z)$. Also, from (7.76) and (7.77) we can obtain the inverse filter

$$\begin{aligned} \theta_{t+1} &= F_{p,t}\theta_t + K_tY_t \\ E_t &= -H_t\theta_t + Y_t, \end{aligned}$$

where $F_{p,t} = F_t - K_t H_t$. Using this inverse filter, it is not difficult to show that

$$\Psi_t^{-1}(z) = I - H_t(I - F_{p,t-1}z)^{-1}K_{t-1}z, \tag{7.78}$$

where $(I - F_{p,t-1}z)^{-1} = I + F_{p,t-1}^{t}z + \cdots + F_{p,2}^{t}z^{t-2}$ corresponds to the tth row of Ψ^{-1}, and

$$(\Psi')_t^{-1}(z^{-1}) = I - K_t'(I - F_{p,t+1}'z^{-1})^{-1}H_{t+1}'z^{-1}, \tag{7.79}$$

where $(I - F_{p,t+1}'z^{-1})^{-1} = I + F_{p,t+1}'^{t+2}z^{-1} + \cdots + F_{p,t+1}'^{n}z^{t-n+1}$ corresponds to the tth row of $\Psi^{-1'}$.

7.3.6.4 Recursive Wiener–Kolmogorov Filtering

In this section, we will show that the Kalman filter can be obtained from the finite Wiener–Kolmogorov filter and that, therefore, the state vector, θ_t, in the state space model (7.76) and (7.77) coincides with the estimator, $\hat{x}_{t|t-1}$, of x_t in (4.1) and (4.2) based on $(Y_{t-1}', \ldots, Y_1')'$.

To this end, suppose that $\{Y_t\}$ follows the state space model (4.1) and (4.2) and let, as before, $Y = (Y_1', \ldots, Y_n')'$ and $E = (E_1', \ldots, E_n')'$ be the stacked vectors of observations and innovations, respectively. Then, from the block Cholesky decomposition of Σ_Y, $\Sigma_Y = \Psi\Sigma\Psi'$, we have the relations $Y = \Psi E$ and $E = \Psi^{-1}Y$. If $X = (x_1', \ldots, x_{n+1}')'$ is the stack of the state vectors and Σ_{XE} and Σ_{XY} are the cross covariance matrices between X and E and X and Y, respectively, it holds that

$$\Sigma_{XE} = \Sigma_{XY}\Psi^{-1'}.$$

To see this, consider that

$$\Sigma_{XE} = E(XE') = E\left[XY'\Psi^{-1'}\right] = \Sigma_{XY}\Psi^{-1'}.$$

Letting $G_{XE,t}(z)$ and $G_{XY,t}(z)$ be the generating functions that represent the tth rows of Σ_{XE} and Σ_{XY}, respectively, it can be proved as earlier in this chapter that the following relation holds

$$G_{XE,t}(z) = G_{XY,t}(z)(\Psi')_t^{-1}(z^{-1}), \tag{7.80}$$

where the usual rules for multiplication and addition apply (the basic rule) and $(\Psi')_t^{-1}(z^{-1})$ is given by (7.79).

As in Sect. 7.3.5, it is not difficult to verify that

$$\hat{x}_{t|n} = G_{XY,t}(z)(\Psi')_t^{-1}(z^{-1})\Sigma_t^{-1}\Psi_t^{-1}(z)Y_t \tag{7.81}$$

and

$$\hat{x}_{t+1|t} = \left[z^{-1}G_{XY,t}(z)(\Psi')_t^{-1}(z^{-1})\right]_+ \Sigma_t^{-1}\Psi_t^{-1}(z)Y_t.$$

Using (7.80), we get

$$\hat{x}_{t+1|t} = \left[z^{-1}G_{XE,t}(z)\right]_+ \Sigma_t^{-1}E_t.$$

As in the time invariant case, we need a lemma to appropriately decompose $G_{XE,t}(z)$. The following lemma is the analog to Lemma 7.2.

Lemma 7.9 *The covariance generating function $G_{XE,t}(z)$ can be decomposed as*

$$\begin{aligned} G_{XE,t}(z) &= (I - F_{t-1}z)^{-1}N_{t-1}z + P_tH_t' + P_tF_{p,t}'(I - F_{p,t+1}'z^{-1})^{-1}H_{t+1}'z^{-1} \\ &= (I - F_{t-1}z)^{-1}N_{t-1}z + P_t(I - F_{p,t}'z^{-1})^{-1}H_t', \end{aligned}$$

where $N_t = F_tP_tH_t' + G_tS_t$, $F_{p,t} = F_t - K_tH_t$, and P_t, $K_t = (F_tP_tH_t' + G_tS_t)\Sigma_t^{-1}$ and $\Sigma_t = R_t + H_tP_tH_t'$ are obtained with the same recursions as those of the Kalman filter.

Proof Let

$$G_{XE,t}(z) = \gamma_{XE}(t, 1)z^{t-1} + \cdots + \gamma_{XE}(t, t) + \cdots + \gamma_{XE}(t, n)z^{t-n}, \tag{7.82}$$

where $\gamma_{XE}(t, s) = \mathrm{Cov}(x_t, E_s)$. Then, subtracting from the Eqs. (4.1) and (4.2) the Eqs. (7.76) and (7.77), respectively, and letting $\tilde{x}_t = x_t - \theta_t$, it is obtained that

$$\begin{aligned} \tilde{x}_{t+1} &= F_{p,t}\tilde{x}_t + G_tu_t - K_tv_t \\ E_t &= H_t\tilde{x}_t + v_t, \end{aligned}$$

where $F_{p,t} = F_t - K_tH_t$. Letting $Z_t = \mathrm{Var}(\tilde{x}_t)$, it is not difficult to see that Z_t satisfies the recursion

$$Z_{t+1} = F_{p,t}Z_tF_{p,t}' + [G_t \;\; -K_t]\begin{bmatrix} Q_t & S_t \\ S_t' & R_t \end{bmatrix}\begin{bmatrix} G_t' \\ -K_t' \end{bmatrix},$$

initialized with $Z_1 = \Pi$ and that this recursion coincides with that of P_t. Therefore, $Z_t = P_t$.

Since θ_t in (7.76) and (7.77) is a linear combination of innovations E_s, $s = t - 1, \ldots, 1$, we have $\mathrm{Cov}(\theta_t, E_s) = 0$ for $s \geq t$. Thus, if $s \geq t$,

$$\begin{aligned} \gamma_{XE}(t, s) &= \mathrm{Cov}(x_t, E_s) = \mathrm{Cov}(\tilde{x}_t, E_s) \\ &= \mathrm{Cov}(\tilde{x}_t, H_s\tilde{x}_s + v_s) \end{aligned}$$

$$= \mathrm{Cov}(\tilde{x}_t, \tilde{x}_s)H_s'$$

$$= \begin{cases} P_tH_t' & s = t \\ P_tF_{p,t}' \cdots F_{p,s-1}'H_s' & s > t. \end{cases}$$

Since E_s is a linear combination of x_1, u_t, $t = s-1, \ldots, 1$, and v_t, $t = s, \ldots, 1$, if $s < t$,

$$\begin{aligned} \gamma_{XE}(t,s) &= \mathrm{Cov}(x_t, E_s) = \mathrm{Cov}(F_{t-1}x_{t-1} + G_{t-1}u_{t-1}, E_s) \\ &= \mathrm{Cov}(F_{t-1}F_{t-2}\cdots F_sx_s + F_{t-1}F_{t-2}\cdots F_{s+1}G_su_s, E_s) \\ &= \mathrm{Cov}(F_s^t\tilde{x}_s + F_{s+1}^tG_su_s, E_s) \\ &= \mathrm{Cov}(F_s^t\tilde{x}_s + F_{s+1}^tG_su_s, H_s\tilde{x}_s + v_s) \\ &= F_s^tP_sH_s' + F_{s+1}^tG_sS_s \\ &= F_{s+1}^t\left(F_sP_sH_s' + G_sS_s\right) \\ &= F_{s+1}^tN_s, \end{aligned}$$

where $N_s = F_sP_sH_s' + G_sS_s$. Substituting these expressions for $\gamma_{XE}(t,s)$ in (7.82), it is obtained that

$$\begin{aligned} G_{XE,t}(z) &= F_2^tN_1z^{t-1} + \cdots + F_t^tN_{t-1}z + P_tH_t' + P_tF_{p,t}^{t+1'}H_{t+1}'z^{t+1} + \cdots + P_tF_{p,t}^{n'}H_n'z^{t-n} \\ &= (I - F_{t-1}z)^{-1}N_{t-1}z + P_tH_t' + P_tF_{p,t}'(I - F_{p,t+1}'z^{-1})^{-1}H_{t+1}z^{-1}. \end{aligned}$$

□

From Lemma 7.9, we get

$$\hat{x}_{t+1|t} = (I - F_tz)^{-1}N_t\Sigma_t^{-1}E_t = (I - F_tz)^{-1}K_tE_t,$$

where $(I - F_tz)^{-1} = I + F_t^{t+1}z + \cdots + F_2^{t+1}z^{t-1}$ and, therefore, $\hat{x}_{t+1|t} = \theta_{t+1}$, where θ_{t+1} is the state vector of (7.76) and (7.77).

7.3.6.5 Prediction

As we saw in the previous section, $E_t = Y_t - H_t\hat{x}_{t|t-1}$ and $\hat{x}_{t+1|t} = (F_t - K_tH_t)\hat{x}_{t|t-1} + K_tY_t$. From this, it is obtained that

$$\hat{x}_{t+1|t} = (I - F_tz)^{-1}K_tE_t = (I - F_{p,t}z)^{-1}K_tY_t, \tag{7.83}$$

where $(I - F_tz)^{-1} = I + F_t^{t+1}z + \cdots + F_2^{t+1}z^{t-1}$ and $(I - F_{p,t}z)^{-1} = I + F_{p,t}^{t+1}z + \cdots + F_{p,2}^{t+1}z^{t-1}$. In addition, it is not difficult to show that the predictor $\hat{x}_{t+k|t}$ of x_{t+k}

based on $\{Y_s : 1 \leq s \leq n\}$, where $k \geq 2$, and its MSE satisfy the recursions $\hat{x}_{t+k|t} = F_t\hat{x}_{t+k-1|t}$ and $\mathrm{MSE}(\hat{x}_{t+k|t}) = F_t\mathrm{MSE}(\hat{x}_{t+k-1|t})F_t' + G_tQ_tG_t'$, where $\mathrm{MSE}(\hat{x}_{t+2|t}) = F_tP_{t+1}F_t' + G_tQ_tG_t'$.

7.3.6.6 Smoothing

From (7.81), (7.80), (7.78), (7.83), Lemma 7.9 and the basic rule, we get

$$\begin{aligned}\hat{x}_{t|n} &= G_{XY,t}(z)(\Psi')_t^{-1}(z^{-1})\Sigma_t^{-1}\Psi_t^{-1}(z)Y_t \\ &= G_{XE,t}(z)\Sigma_t^{-1}E_t \\ &= (I - F_{t-1}z)^{-1}K_{t-1}E_{t-1} + P_t\lambda_t \\ &= \hat{x}_{t|t-1} + P_t\lambda_t, \end{aligned} \tag{7.84}$$

where we define

$$\begin{aligned}\lambda_t &= (I - F_{p,t}'z^{-1})^{-1}H_t'\Sigma_t^{-1}E_t \qquad (7.85)\\ &= (I - F_{p,t}'z^{-1})^{-1}H_t'\Sigma_t^{-1}\left[I - H_t(I - F_{p,t-1}z)^{-1}K_{t-1}z\right]Y_t\end{aligned}$$

$(I - F_{p,t}'z^{-1})^{-1} = I + F_{p,t}'^{t+1}z^{-1} + \cdots + F_{p,t}'^{n}z^{t-n}$ and $(I - F_{p,t-1}z)^{-1} = I + F_{p,t-1}'^{t}z + \cdots + F_{p,2}^{t}z^{t-2}$. The definition of λ_t implies that λ_t follows the backwards recursion

$$\lambda_t = F_{p,t}'\lambda_{t+1} + H_t'\Sigma_t^{-1}E_t, \tag{7.86}$$

initialized with $\lambda_{n+1} = 0$. Therefore, λ_t is the adjoint variable and (7.84) is the formula for the fixed interval smoother of Theorem 4.20. If $\mathrm{Var}(\lambda_t) = \Lambda_t$, it follows from (7.86) that Λ_t satisfies the recursion

$$\Lambda_t = F_{p,t}'\Lambda_{t+1}F_{p,t} + H_t'\Sigma_t^{-1}H_t.$$

By (7.84), $x_t - \hat{x}_{t|t-1} = x_t - \hat{x}_{t|n} + P_t\lambda_t$, where $x_t - \hat{x}_{t|t-1}$ and λ_t are uncorrelated because, by (7.86), λ_t is a linear combination of E_s, $s = t, t+1, \ldots, n$. Thus, if $\mathrm{MSE}(\hat{x}_{t|n}) = P_{t|n}$, $P_{t|n}$ satisfies the recursion

$$P_{t|n} = P_t - P_t\Lambda_tP_t.$$

To simplify the computation of (7.85), we need the following lemma.

Lemma 7.10 *With the previous notation, the following equality holds.*

$$(I-F_{p,t}'z^{-1})^{-1}H_t'\Sigma_t^{-1}H_t(I-F_{p,t-1}z)^{-1} = \Lambda_t(I-F_{p,t-1}z)^{-1} + (I-F_{p,t}'z^{-1})^{-1}F_{p,t}'\Lambda_{t+1}z^{-1}.$$

Proof Using $H_t'\Sigma_t^{-1}H_t = \Lambda_t - F_{p,t}'\Lambda_{t+1}F_{p,t}$ and the basic rule, we can write

$$
\begin{aligned}
&(I - F_{p,t}'z^{-1})^{-1}H_t'\Sigma_t^{-1}H_t \times (I - F_{p,t-1}z)^{-1} \\
&\quad = (I - F_{p,t}'z^{-1})^{-1}(\Lambda_t - F_{p,t}'\Lambda_{t+1}F_{p,t})(I - F_{p,t-1}z)^{-1} \\
&\quad = (I - F_{p,t}'z^{-1})^{-1}\left[(I - F_{p,t}'z^{-1})\Lambda_t + \Lambda_t(I - F_{p,t-1}z)\right. \\
&\qquad \left. -(I - F_{p,t}')\Lambda_t(I - F_{p,t-1}z)\right](I - F_{p,t-1}z)^{-1} \\
&\quad = \Lambda_t(I - F_{p,t-1}z)^{-1} + (I - F_{p,t}'z^{-1})^{-1}\Lambda_t - \Lambda_t \\
&\quad = \Lambda_t(I - F_{p,t-1}z)^{-1} + (I - F_{p,t}'z^{-1})^{-1}\Lambda_t \\
&\qquad -(I - F_{p,t}'z^{-1})^{-1}(I - F_{p,t}z^{-1})\Lambda_t \\
&\quad = \Lambda_t(I - F_{p,t-1}z)^{-1} \\
&\qquad +(I - F_{p,t}'z^{-1})^{-1}(I - I + F_{p,t}'z^{-1})\Lambda_t \\
&\quad = \Lambda_t(I - F_{p,t-1}z)^{-1} + (I - F_{p,t}'z^{-1})^{-1}F_{p,t}'\Lambda_{t+1}z^{-1}.
\end{aligned}
$$

□

Using Lemma 7.10, we get the following lemma.

Lemma 7.11 *The weights of the adjoint variable, λ_t, in Theorem 4.20 are given by*

$$\lambda_t = (I - F_{p,t}'z^{-1})^{-1}(H_t'\Sigma_t^{-1} - F_{p,t}'\Lambda_{t+1}K_t)Y_t - \Lambda_t(I - F_{p,t-1}z)^{-1}K_{t-1}Y_{t-1}.$$

By Theorem 4.20 or, equivalently, (7.84), the fixed-interval smoother satisfies the recursion $\hat{x}_{t|n} = \hat{x}_{t|t-1} + P_t\lambda_t$. Thus, using the previous lemma and (7.83), we can obtain the filter weights for the fixed-interval smoother. We summarize the result in the following theorem. We omit its proof.

Theorem 7.23 *The weights for the fixed-interval smoother of Theorem 4.20 are given by*

$$\hat{x}_{t|n} = (I - P_t\Lambda_t)(I - F_{p,t-1}z)^{-1}K_{t-1}Y_{t-1} + P_t(I - F_{p,t}'z^{-1})^{-1}(H_t'\Sigma_t^{-1} - F_{p,t}'\Lambda_{t+1}K_t)Y_t.$$

More specifically, the weights Ω_j of the estimator $\hat{x}_{t|n} = \sum_{j=t-n}^{t-1}\Omega_jY_{t-j}$ are given by

$$
\begin{aligned}
\Omega_0 &= P_t(H_t'\Sigma_t^{-1} - F_{p,t}'\Lambda_{t+1}K_t), \quad \Omega_1 = (I - P_t\Lambda_t)K_{t-1}, \\
\Omega_j &= (I - P_t\Lambda_t)F_{p,t-j+1}^t K_{t-j}, && j > 1 \\
\Omega_j &= P_tF_{p,t}^{\prime t-j}(H_{t-j}'\Sigma_{t-j}^{-1} - F_{p,t-j}'\Lambda_{t+1-j}K_{t-j}), && j < 0.
\end{aligned}
$$

As noted earlier, Koopman & Harvey (2003) give an algorithm for computing finite signal extraction weights for state space models. However, their computations are based on first principles and not on finite covariance generating functions like ours. This last approach offers the advantage of immediately giving the weights when, in the time invariant case, the Kalman filter converges to the steady state recursions and the sample becomes infinite.

In Sect. 7.2.7, sufficient conditions were given for the convergence of the Kalman filtering and smoothing recursions in the time invariant case to the steady state recursions.

7.3.6.7 Inverse Process and Interpolation

By Theorem 4.21, the inverse process, Y_t^i, satisfies the recursion $Y_t^i = \Sigma_t^{-1}E_t - K_t'\lambda_{t+1}$. This result can also be obtained using covariance generating functions. To see this, consider first that $Y_t^i = G_{Y_t^i}(z)Y_t$ and, by (7.63), $G_{Y_t^i}(z) = (\Psi')_t^{-1}(z^{-1})\Sigma_t^{-1}\Psi_t^{-1}(z)$. Then, by (7.79), (7.78), (7.85) and the basic rule, we can write

$$\begin{aligned}
Y_t^i &= (\Psi')_t^{-1}(z^{-1})\Sigma_t^{-1}\Psi_t^{-1}(z)Y_t \\
&= \left[I - K_t'(I - F_{p,t+1}'z^{-1})^{-1}H_{t+1}'z^{-1}\right]\Sigma_t^{-1}E_t \\
&= \Sigma_t^{-1}E_t - K_t'\lambda_{t+1}.
\end{aligned}$$

Using (7.78) and Lemma 7.11, the following theorem is obtained.

Theorem 7.24 *The weights of the inverse process Y_t^i are given by*

$$\begin{aligned}
Y_t^i = {} & (\Sigma_t^{-1} + K_t'\Lambda_{t+1}K_t)Y_t + (K_t'\Lambda_{t+1}F_{p,t} - \Sigma_t^{-1}H_t)(I - F_{p,t-1}z)^{-1}Y_{t-1} \\
& + K_t'(I - F_{p,t+1}'z^{-1})^{-1}(F_{p,t+1}'\Lambda_{t+2}K_{t+1} - H_{t+1}'\Sigma_{t+1}^{-1})Y_{t+1}.
\end{aligned}$$

This theorem, together with Theorem 4.21, gives the following expression for the interpolator $Y_{t|s\neq t}$.

$$\begin{aligned}
Y_{t|s\neq t} = {} & -(K_t'\Lambda_{t+1}F_{p,t} - \Sigma_t^{-1}H_t)(I - F_{p,t-1}z)^{-1}Y_{t-1} \\
& - K_t'(I - F_{p,t+1}'z^{-1})^{-1}(F_{p,t+1}'\Lambda_{t+2}K_{t+1} - H_{t+1}'\Sigma_{t+1}^{-1})Y_{t+1}.
\end{aligned}$$

7.3.6.8 Disturbance Smoothers

By Theorem 4.22, the disturbance smoothers satisfy the recursions

$$\hat{v}_{t|n} = R_t\Sigma_t^{-1}E_t + (S_t'G_t' - R_tK_t')\lambda_{t+1} \tag{7.87}$$

$$\hat{u}_{t|n} = S_t\Sigma_t^{-1}E_t + (Q_tG_t' - S_tK_t')\lambda_{t+1}. \tag{7.88}$$

The recursions (7.87) and (7.88) can also be obtained using generating functions. We omit the details. Using (7.78) and Lemma 7.11, the following theorem is obtained.

Theorem 7.25 *Letting $J_t' = S_t'G_t' - R_tK_t'$ and $L_t' = Q_tG_t' - S_tK_t'$, the weights for $\hat{v}_{t|n}$ and $\hat{u}_{t|n}$ are given by*

$$\begin{aligned}\hat{v}_{t|n} &= (R_t\Sigma_t^{-1} - J_t'\Lambda_{t+1}K_t)Y_t - (J_t'\Lambda_{t+1}F_{p,t} + R_t\Sigma_t^{-1}H_t)(I - F_{p,t-1}z)^{-1}Y_{t-1}\\ &\quad -J_t'(I - F_{p,t+1}'z^{-1})^{-1}(F_{p,t+1}'\Lambda_{t+2}K_{t+1} - H_{t+1}'\Sigma_{t+1}^{-1})Y_{t+1}\end{aligned}$$

$$\begin{aligned}\hat{u}_{t|n} &= (S_t\Sigma_t^{-1} - L_t'\Lambda_{t+1}K_t)Y_t - (L_t'\Lambda_{t+1}F_{p,t} + S_t\Sigma_t^{-1}H_t)(I - F_{p,t-1}z)^{-1}Y_{t-1}\\ &\quad -L_t'(I - F_{p,t+1}'z^{-1})^{-1}(F_{p,t+1}'\Lambda_{t+2}K_{t+1} - H_{t+1}'\Sigma_{t+1}^{-1})Y_{t+1}.\end{aligned}$$

7.3.6.9 Covariance Generating Functions of the State Errors

In this section, we will provide operational expressions for the errors $\epsilon_{t|t-1} = x_t - \hat{x}_{t|t-1}$ and $\epsilon_{t|n} = x_t - \hat{x}_{t|n}$ in the Kalman filter and smoother.

Subtracting the expression for $\hat{x}_{t+1|t}$ given by (4.3) from (4.1), and using $E_t = Y_t - H_t\hat{x}_{t|t-1}$ and (4.2), the following recursions are obtained

$$\epsilon_{t+1|t} = F_{p,t}\epsilon_{t|t-1} + [G_t, -K_t]\begin{bmatrix} u_t \\ v_t \end{bmatrix}, \tag{7.89}$$

$$P_{t+1} = F_{p,t}P_tF_{p,t}' + [G_t, -K_t]\begin{bmatrix} Q_t & S_t \\ S_t' & R_t \end{bmatrix}\begin{bmatrix} G_t' \\ -K_t' \end{bmatrix}, \tag{7.90}$$

initialized with $\epsilon_{1|0} = x_1$ and $P_1 = \Pi$. Iterating, we get $\epsilon_{t|t-1} = \bar{u}_{t-1} + F_{p,t-1}^t\bar{u}_{t-2} + \cdots + F_{p,1}^t\bar{u}_0$, where $\bar{u}_0 = x_1$ and $\bar{u}_t = [G_t, -K_t][u_t', v_t']'$ for $t = 1, \ldots, n$. Thus, the generating function of $\epsilon_{t|t-1}$ is $\Psi_{\epsilon,t}(z) = (I + F_{p,t-1}^tz + \cdots + F_{p,1}^tz^{t-1})z$. Since it is not difficult to verify that

$$\begin{pmatrix} I & & & & & \\ F_{p,1}^2 & I & & & & \\ F_{p,1}^3 & F_{p,2}^3 & I & & & \\ \vdots & \vdots & \vdots & \ddots & & \\ F_{p,1}^n & F_{p,2}^n & \cdots & F_{p,n-1}^n & I & \\ 0 & 0 & 0 & 0 & 0 & I \end{pmatrix} = \begin{pmatrix} I & & & & & \\ -F_{p,1}^2 & I & & & & \\ & -F_{p,2}^3 & I & & & \\ & & \ddots & \ddots & & \\ & & & -F_{p,n-1}^n & I & \\ & & & & 0 & I \end{pmatrix}^{-1},$$

we can write $\Psi_{\epsilon,t}(z) = (I - F_{p,t-1}z)_1^{-1}z$ and $\epsilon_{t|t-1} = (I - F_{p,t-1}z)_1^{-1}z\bar{u}_t$, where $(I - F_{p,t-1}z)_1^{-1} = I + F_{p,t-1}^tz + \cdots + F_{p,1}^tz^{t-1}$. It follows from this and (7.90) that the covariance generating function, $G_{\epsilon_{t|t-1}}(z)$, of $\epsilon_{t|t-1}$ is given by

$$\begin{aligned} G_{\epsilon_{t|t-1}}(z) &= (I - F_{p,t-1}z)_1^{-1}(P_t - F_{p,t-1}P_{t-1}F'_{p,t-1})(I - F'_{p,t}z^{-1})^{-1} \\ &= (I - F_{p,t-1}z)_1^{-1}\left[P_t(I - F'_{p,t}z^{-1}) + (I - F_{p,t-1}z)P_t \right. \\ &\quad \left. -(I - F_{p,t-1}z)P_t(I - F'_{p,t}z^{-1})\right](I - F'_{p,t}z^{-1})^{-1} \\ &= (I - F_{p,t-1}z)_1^{-1}P_t + P_t(I - F'_{p,t}z^{-1})^{-1} - P_t \\ &= (I - F_{p,t-1}z)_1^{-1}P_t + P_tF'_{p,t}(I - F'_{p,t+1}z^{-1})^{-1}z^{-1}. \end{aligned}$$

This proves the following theorem.

Theorem 7.26 *The covariance generating function, $G_{\epsilon_{t|t-1}}(z)$, of the error $\epsilon_{t|t-1}$ can be expressed as*

$$G_{\epsilon_{t|t-1}}(z) = (I - F_{p,t-1}z)_1^{-1}P_t + P_tF'_{p,t}(I - F'_{p,t+1}z^{-1})^{-1}z^{-1},$$

where $(I - F'_{p,t}z^{-1})^{-1} = I + F'^{t+1}_{p,t}z^{-1} + \cdots + F'^{n}_{p,t}z^{t-n}$ and $(I - F_{p,t-1}z)_1^{-1} = I + F^t_{p,t-1}z + \cdots + F^t_{p,1}z^{t-1}$.

To obtain the covariance generating function of the error $\epsilon_{t|n}$, first note that

$$\begin{aligned} \epsilon_{t|n} &= x_t - \hat{x}_{t|n} = x_t - \hat{x}_{t|t-1} - P_t\lambda_t \\ &= \epsilon_{t|t-1} - P_t\lambda_t. \end{aligned} \tag{7.91}$$

Then, $\epsilon_{t|t-1} = \epsilon_{t|n} + P_t\lambda_t$, where $\epsilon_{t|n}$ and λ_t are uncorrelated, and it holds that the covariance generating function, $G_{\epsilon_{t|n}}(z)$, of the error $\epsilon_{t|n}$ satisfies $G_{\epsilon_{t|n}}(z) = G_{\epsilon_{t|t-1}}(z) - P_tG_{\lambda_t}(z)P_t$, where $G_{\lambda_t}(z)$ is the covariance generating function of λ_t. It follows from (7.85) that $G_{\lambda_t}(z) = (I - F'_{p,t}z^{-1})^{-1}H'_t\Sigma_t^{-1}H_t(I - F_{p,t-1}z)_1^{-1}$, where $(I - F'_{p,t}z^{-1})^{-1} = I + F'^{t+1}_{p,t}z^{-1} + \cdots + F'^{n}_{p,t}z^{t-n}$ and $(I - F_{p,t-1}z)_1^{-1} = I + F^t_{p,t-1}z + \cdots + F^t_{p,1}z^{t-1}$. Analogously to Lemma 7.10, we can prove the following lemma.

Lemma 7.12 *The covariance generating function, $G_{\lambda_t}(z)$, of λ_t is given by the following expression*

$$\begin{aligned} G_{\lambda_t}(z) &= (I - F'_{p,t}z^{-1})^{-1}H'_t\Sigma_t^{-1}H_t(I - F_{p,t-1}z)_1^{-1} \\ &= \Lambda_t(I - F_{p,t-1}z)_1^{-1} + (I - F'_{p,t}z^{-1})^{-1}F'_{p,t}\Lambda_{t+1}z^{-1}, \end{aligned}$$

where $(I - F'_{p,t}z^{-1})^{-1} = I + F'^{t+1}_{p,t}z^{-1} + \cdots + F'^{n}_{p,t}z^{t-n}$ and $(I - F_{p,t-1}z)_1^{-1} = I + F^t_{p,t-1}z + \cdots + F^t_{p,1}z^{t-1}$.

By Lemma 7.12 and Theorem 7.26, we can write

$$
\begin{aligned}
G_{\epsilon_{t|n}}(z) &= (I - F_{p,t-1}z)_1^{-1}P_t + P_tF'_{p,t}(I - F'_{p,t+1}z^{-1})^{-1}z^{-1} \\
&\quad - P_t\left[\Lambda_t(I - F_{p,t-1}z)_1^{-1} + (I - F'_{p,t}z^{-1})^{-1}F'_{p,t}\Lambda_{t+1}z^{-1}\right]P_t \\
&= P_t(I - \Lambda_tP_t) + (I - P_t\Lambda_t)(I - F_{p,t-1}z)_1^{-1}F_{p,t-1}P_{t-1}z \\
&\quad + P_tF'_{p,t}(I - F'_{p,t+1}z^{-1})^{-1}(I - \Lambda_{t+1}P_{t+1})z^{-1}.
\end{aligned}
$$

We summarize this result in the following theorem.

Theorem 7.27 *The covariance generating function, $G_{\epsilon_{t|n}}(z)$, of the error $\epsilon_{t|n}$ is given by*

$$
\begin{aligned}
G_{\epsilon_{t|n}}(z) &= P_t(I - \Lambda_tP_t) + (I - P_t\Lambda_t)(I - F_{p,t-1}z)_1^{-1}F_{p,t-1}P_{t-1}z \\
&\quad + P_tF'_{p,t}(I - F'_{p,t+1}z^{-1})^{-1}(I - \Lambda_{t+1}P_{t+1})z^{-1},
\end{aligned}
$$

where $(I - F'_{p,t}z^{-1})^{-1} = I + F'^{t+1}_{p,t}z^{-1} + \cdots + F'^{n}_{p,t}z^{t-n}$ and $(I - F_{p,t-1}z)_1^{-1} = I + F^t_{p,t-1}z + \cdots + F^t_{p,1}z^{t-1}$.

7.3.6.10 Covariance Generating Functions of the Disturbance Errors

Define the disturbance errors $\eta_{t|n} = v_t - \hat{v}_{t|n}$ and $\delta_{t|n} = u_t - \hat{u}_{t|n}$ and let $G_{\eta_{t|n}}(z)$, $G_{\delta_{t|n}}(z)$, $G_{\eta_{t|n},\delta_{t|n}}(z)$, $G_{\epsilon_{t|n},\eta_{t|n}}(z)$, and $G_{\epsilon_{t|n},\delta_{t|n}}(z)$ be the covariance generating functions of $\eta_{t|n}$ and $\delta_{t|n}$ and the cross-covariance generating functions of $\eta_{t|n}$ and $\delta_{t|n}$, $\epsilon_{t|n}$ and $\eta_{t|n}$, and $\epsilon_{t|n}$ and $\delta_{t|n}$.

To obtain the previous covariance and cross covariance generating functions, we use the following equality

$$
\begin{pmatrix} u_t \\ v_t \\ \epsilon_{t|t-1} \end{pmatrix} = \begin{pmatrix} \delta_{t|n} \\ \eta_{t|n} \\ \epsilon_{t|n} \end{pmatrix} + \begin{pmatrix} S_t\Sigma_t^{-1}E_t + (Q_tG'_t - S_tK'_t)\lambda_{t+1}. \\ R_t\Sigma_t^{-1}E_t + (S'_tG'_t - R_tK'_t)\lambda_{t+1} \\ P_t\lambda_t \end{pmatrix},
$$

that follows directly from (7.87), (7.88), and (7.91). By (7.89) and (7.85), we can write

$$
\begin{aligned}
&\begin{pmatrix} I & 0 \\ 0 & I \\ (I - F_{p,t-1}z)_1^{-1}zG_t & -(I - F_{p,t-1}z)_1^{-1}zK_t \end{pmatrix}\begin{pmatrix} u_t \\ v_t \end{pmatrix} \\
&= \begin{pmatrix} \delta_{t|n} \\ \eta_{t|n} \\ \epsilon_{t|n} \end{pmatrix} + \left\{\begin{pmatrix} S_t \\ R_t \\ 0 \end{pmatrix} + \begin{bmatrix} (Q_tG'_t - S_tK'_t)z^{-1} \\ (S'_tG'_t - R_tK'_t)z^{-1} \\ P_t \end{bmatrix}(I - F'_{p,t}z^{-1})^{-1}H'_t\right\}\Sigma_t^{-1}E_t,
\end{aligned}
$$

where the last two terms are uncorrelated. It follows from this that the covariance generating function of the left-hand side is equal to the sum of the covariance generating functions of the two terms in the right-hand side. Letting

$$G(z) = \begin{pmatrix} G_{\delta_{t|n}}(z) & G_{\delta_{t|n}\eta_{t|n}}(z) & G_{\delta_{t|n}\epsilon_{t|n}}(z) \\ G_{\eta_{t|n}\delta_{t|n}}(z) & G_{\eta_{t|n}}(z) & G_{\eta_{t|n}\epsilon_{t|n}}(z) \\ G_{\epsilon_{t|n}\delta_{t|n}}(z) & G_{\epsilon_{t|n}\eta_{t|n}}(z) & G_{\epsilon_{t|n}}(z) \end{pmatrix}$$

be the covariance generating function of $(\delta_{t|n}, \eta_{t|n}, \epsilon_{t|n})'$, $L_t' = Q_t G_t' - S_t K_t'$ and $J_t' = S_t' G_t' - R_t K_t'$, we get

$$\begin{pmatrix} Q_t & S_t & L_t' z^{-1}(I - F_{p,t}' z^{-1})^{-1} \\ S_t' & R_t & J_t' z^{-1}(I - F_{p,t}' z^{-1})^{-1} \\ (I - F_{p,t-1} z)_1^{-1} z L_t & (I - F_{p,t-1} z)_1^{-1} z J_t & G_{\epsilon_{t|t-1}}(z) \end{pmatrix}$$
$$= G(z) + \begin{pmatrix} S_t \Sigma_t^{-1} S_t' & S_t \Sigma_t^{-1} R_t & 0 \\ R_t \Sigma_t^{-1} S_t' & R_t \Sigma_t^{-1} R_t & 0 \\ 0 & 0 & 0 \end{pmatrix}$$
$$+ \begin{pmatrix} L_t' z^{-1}(I - F_{p,t}' z^{-1})^{-1} H_t' \Sigma_t^{-1} S_t' & L_t' z^{-1}(I - F_{p,t}' z^{-1})^{-1} H_t' \Sigma_t^{-1} R_t & 0 \\ J_t' z^{-1}(I - F_{p,t}' z^{-1})^{-1} H_t' \Sigma_t^{-1} S_t' & J_t' z^{-1}(I - F_{p,t}' z^{-1})^{-1} H_t' \Sigma_t^{-1} R_t & 0 \\ P_t(I - F_{p,t}' z^{-1})^{-1} H_t' \Sigma_t^{-1} S_t' & P_t(I - F_{p,t}' z^{-1})^{-1} H_t' \Sigma_t^{-1} R_t & 0 \end{pmatrix}$$
$$+ \begin{pmatrix} S_t \Sigma_t^{-1} H_t z (I - F_{p,t} z)_1^{-1} L_t & S_t \Sigma_t^{-1} H_t z (I - F_{p,t} z)_1^{-1} J_t & S_t \Sigma_t^{-1} H_t (I - F_{p,t-1} z)_1^{-1} P_t \\ R_t \Sigma_t^{-1} H_t z (I - F_{p,t} z)_1^{-1} L_t & R_t \Sigma_t^{-1} H_t z (I - F_{p,t} z)_1^{-1} J_t & R_t \Sigma_t^{-1} H_t (I - F_{p,t-1} z)_1^{-1} P_t \\ 0 & 0 & 0 \end{pmatrix}$$
$$+ \begin{pmatrix} L_t' G_{\lambda_{t+1}}(z) L_t & L_t' G_{\lambda_{t+1}}(z) J_t & L_t' z^{-1} G_{\lambda_t}(z) P_t \\ J_t' G_{\lambda_{t+1}}(z) L_t & J_t' G_{\lambda_{t+1}}(z) J_t & J_t' z^{-1} G_{\lambda_t}(z) P_t \\ P_t G_{\lambda_t}(z) z L_t & P_t G_{\lambda_t}(z) z J_t & P_t G_{\lambda_t}(z) P_t \end{pmatrix},$$

where $G_{\lambda_t}(z)$ is given by Lemma 7.12. Equating the entries in the left- and right-hand sides and using Lemma 7.4, the following theorem is obtained.

Theorem 7.28 *The covariance generating functions, $G_{\eta_{t|n}}(z)$ and $G_{\delta_{t|n}}(z)$, of $\eta_{t|n}$ and $\delta_{t|n}$ and the cross-covariance generating functions, $G_{\eta_{t|n}\delta_{t|n}}(z)$, $G_{\epsilon_{t|n}\eta_{t|n}}(z)$ and $G_{\epsilon_{t|n}\delta_{t|n}}(z)$, of $\eta_{t|n}$ and $\delta_{t|n}$, $\epsilon_{t|n}$ and $\eta_{t|n}$, and $\epsilon_{t|n}$ and $\delta_{t|n}$ are given by*

$$\begin{aligned} G_{\eta_{t|n}}(z) = R_t - R_t \Sigma_t^{-1} R_t - J_t' \Lambda_{t+1} J_t \\ &- \left[R_t \Sigma_t^{-1} H_t + J_t' \Lambda_{t+1} F_{p,t} \right] (I - F_{p,t-1} z)_1^{-1} J_{t-1} z \\ &- J_t' (I - F_{p,t+1}' z^{-1})^{-1} \left[H_{t+1}' \Sigma_{t+1}^{-1} R_{t+1} + F_{p,t+1}' \Lambda_{t+2} J_{t+1} \right] z^{-1}, \end{aligned}$$

$$
\begin{aligned}
G_{\delta_{t|n}}(z) &= Q_t - S_t\Sigma_t^{-1}S_t' - L_t'\Lambda_{t+1}L_t \\
&\quad - \left[S_t\Sigma_t^{-1}H_t + L_t'\Lambda_{t+1}F_{p,t}\right](I - F_{p,t-1}z)_1^{-1}L_{t-1}z \\
&\quad - L_t'(I - F_{p,t+1}'z^{-1})^{-1}\left[H_{t+1}'\Sigma_{t+1}^{-1}S_{t+1}' + F_{p,t+1}'\Lambda_{t+2}L_{t+1}\right]z^{-1},
\end{aligned}
$$

$$
\begin{aligned}
G_{\eta_{t|n}\delta_{t|n}}(z) &= S_t' - R_t\Sigma_t^{-1}S_t' - J_t'\Lambda_{t+1}L_t \\
&\quad - \left[R_t\Sigma_t^{-1}H_t + J_t'\Lambda_{t+1}F_{p,t}\right](I - F_{p,t-1}z)_1^{-1}L_{t-1}z \\
&\quad - J_t'(I - F_{p,t+1}'z^{-1})^{-1}\left[H_{t+1}'\Sigma_{t+1}^{-1}S_{t+1}' + F_{p,t+1}'\Lambda_{t+2}L_{t+1}\right]z^{-1},
\end{aligned}
$$

$$
\begin{aligned}
G_{\epsilon_{t|n}\eta_{t|n}}(z) &= -P_t(H_t'\Sigma_t^{-1}R_t + F_{p,t}'\Lambda_{t+1}J_t) \\
&\quad + (I - P_t\Lambda_t)(I - F_{p,t-1}z)_1^{-1}J_{t-1}z \\
&\quad - P_t(I - F_{p,t}'z^{-1})^{-1}F_{p,t}'\left[H_{t+1}'\Sigma_{t+1}^{-1}R_{t+1} + F_{p,t+1}'\Lambda_{t+1}J_{t+1}\right]z^{-1},
\end{aligned}
$$

and

$$
\begin{aligned}
G_{\epsilon_{t|n}\delta_{t|n}}(z) &= -P_t(H_t'\Sigma_t^{-1}S_t' + F_{p,t}'\Lambda_{t+1}L_t) \\
&\quad + (I - P_t\Lambda_t)(I - F_{p,t-1}z)_1^{-1}L_{t-1}z \\
&\quad - P_t(I - F_{p,t}'z^{-1})^{-1}F_{p,t}'\left[H_{t+1}'\Sigma_{t+1}^{-1}S_{t+1}' + F_{p,t+1}'\Lambda_{t+1}L_{t+1}\right]z^{-1},
\end{aligned}
$$

where $L_t' = Q_tG_t' - S_tK_t'$ *and* $J_t' = S_t'G_t' - R_tK_t'$.

7.4 Historical Notes

Wiener's solution to the problems of filtering and prediction had a huge impact on many fields of engineering and mathematics, although it is ironic that it fell short of its original goal of solving the anti-aircraft-gun control problem. According to Kalman (1963), "There is no doubt that Wiener's theory of statistical prediction and filtering is one of the great contributions to engineering science." Doob (1953) devoted the last chapter of his celebrated textbook to Wiener's theory.

As regards the mathematical side, Wiener's development even in the scalar case was not as general as the independent work of Kolmogorov (1939, 1941). A footnote in Wiener (1949, p. 59) describes well the relationship between Wiener's work and that of Kolmogorov's.

A standard reference in discrete time Wiener–Hopf theory is Whittle (1963b) and its updated second edition Whittle (1983).

Wiener–Kolmogorov filtering and smoothing has received a lot of attention in the statistical literature on time series. See, for example, Cleveland & Tiao (1976), Burman (1980), Burridge & Wallis (1988), Gómez (2001), Gómez & Maravall (2001b), Harvey & Trimbur (2003), and the references therein.

In the engineering literature, the monograph by Kailath et al. (2000) contains two chapters on Wiener–Kolmogorov theory, Chaps. 7 and 8.

Some of the results on recursive Wiener–Kolmogorov filtering and smoothing presented in Gómez (2006) are new in the literature, like the ones regarding the mean squared errors for smoothing based on the doubly infinite sample, its extension to the nonstationary case, or the computation of the filter weights.

For the doubly infinite sample in the nonstationary univariate case, Bell (1984) proposed two Assumptions, that he called A and B. He proved that the usual Wiener–Kolmogorov formulae are valid under Assumption A but not under Assumption B. By the results of this chapter, if we put the ARIMA unobserved components model considered by Bell (1984) into state space form, make any of the two Assumptions, and iterate in the Kalman filter and smoother, we obtain again the previous Wiener–Kolmogorov formulae. The reason for this is that both Assumptions lead to the same result in the finite sample case (Gómez, 1999, p. 9). The only difference is that with the limiting method the estimators are interpreted not as minimum mean squared error estimators but as limits of these. However, this seems to make more sense because in Assumptions A and B the initial conditions are in the middle of the infinite sample, which is a little odd, whereas in the Kalman filter limiting approach they tend to be in the infinitely remote past. If we adopt this position, we get to the conclusion that the two Assumptions are equivalent both in the finite and in the infinite sample. This would end a controversy that has been going on for quite a long time in the statistical literature. See, for example, Burridge & Wallis (1988), Gómez (1999), and McElroy (2008).

7.5 Problems

7.1 By the results in Sect. 7.2.7, the Wiener–Kolmogorov formulae for smoothing are also valid for nonstationary series. Consider the signal-plus-noise model

$$Y_t = S_t + N_t,$$

where S_t follows the nonstationary model

$$(1 - B)S_t = b_t,$$

$\{b_t\}$ $\sim$WN$(0, \sigma_b^2)$, $\{N_t\}$ $\sim$WN$(0, \sigma_n^2)$ and $\{b_t\}$ and $\{N_t\}$ are mutually uncorrelated. Prove that

$$\hat{S}_{t|\infty} = \frac{1}{1 + k(1 - B)(1 - F)} Y_t, \tag{7.92}$$

where F is the forward operator, $FY_t = Y_{t+1}$, and $k = \sigma_n^2/\sigma_b^2$. Obtain θ and $\sigma_a^2 = \mathrm{Var}(A_t)$ in the model

$$(1-B)Y_t = (1-\theta B)A_t,$$

followed by $\{Y_t\}$.

7.2 Obtain the covariance generating function, $G_{E\infty}(z) = G_S(z) - G_{SY}(z)G_Y^{-1}(z)G_{YS}(z^{-1})$, of the error $E_{t|\infty} = S_t - \hat{S}_{t|\infty}$ in Problem 7.1.

7.3 Use Lemma 7.1 in Problem 7.1 to obtain the filter for $\hat{S}_{t|t}$, that is,

$$\Pi(z) = \left[\frac{G_{SY}(z)}{\sigma_a^2\Psi(z^{-1})}\right]_+ \frac{1}{\Psi(z)},$$

where $G_{SY}(z) = G_S(z)$ and $\Psi(z) = (1-\theta z)/(1-z)$.

7.4 In Problem 7.1, obtain the covariance generating function, $G_{E0}(z)$, of the error $E_{t|t} = S_t - \hat{S}_{t|t}$, where $G_{E0}(z) = G_{E\infty}(z) + \Omega_0(z)\sigma_a^{-2}\Omega_0'(z^{-1})$, $G_{E\infty}(z)$ is that of Problem 7.2, and

$$\Omega_0(z) = \left[G_{SY}(z)\Psi'^{-1}(z^{-1})\right]_- .$$

7.5 Suppose the setting of Problem 7.1.

1. Show that $\{Y_t\}$ can be put into state space form (7.27) and (7.28) by defining $x_t = S_t$, $F = 1$, $G = 1$, $H = 1$, $u_t = b_{t+1}$, $v_t = N_t$, $Q = \sigma_b^2$, $R = \sigma_n^2$, and $S = 0$.
2. Prove that the DARE (7.36) corresponding to the previous state space model is

$$P^2 - P\sigma_b^2 - \sigma_b^2\sigma_n^2 = 0$$

   and that this equation has real solutions of opposite sign such that the positive solution satisfies $P > \sigma_b^2$.
3. Obtain K and Σ as functions of the positive solution, P, of the DARE and the other parameters of the model so that the generating function, $G_Y(z)$, of $\{Y_t\}$ factorizes as (7.34). Show that $\{Y_t\}$ follows a model of the form $Y_t - Y_{t-1} = A_t - \theta A_{t-1}$, where $A_t \sim WN(0, \Sigma)$ and find θ in terms of P and the other parameters in the model.
4. Since $S_t = x_t$ in the state space form (7.27) and (7.28) for this model, obtain the recursions (7.42) and (7.43) corresponding to $\hat{S}_{t|t-1}$ and $\hat{S}_{t|\infty}$, respectively.

7.6 In Problem 7.5, use Theorem 7.13 to compute the weights Ω_j of the estimator $\hat{S}_{t|\infty} = \sum_{j=-\infty}^{\infty}\Omega_j Y_{t+j}$ for $j = \{0, \pm 1, \pm 2\}$.

7.7 Under the assumptions and with the notation of Problem 7.1, prove that (7.92) can be expressed as

$$\hat{S}_{t|\infty} = \frac{\sigma_b^2/\sigma_a^2}{(1-\theta B)(1-\theta F)} Y_t.$$

Use Lemma 7.1 to decompose the previous filter in the following way

$$\hat{S}_{t|\infty} = \left[\frac{k_0}{(1-B)} + \frac{k_0}{(1-F)} \right] Y_t,$$

and compute the weights Ω_j of the estimator $\hat{S}_{t|\infty} = \sum_{j=-\infty}^{\infty} \Omega_j Y_{t+j}$ for $j = \{0, \pm 1, \pm 2\}$.

7.8 Suppose the nonstationary bivariate series $Y_t = (Y_{1t}, Y_{2t})'$ that follows the trend plus seasonal model

$$Y_t = N_t + S_t, \tag{7.93}$$

where $N_{t+1} = \Phi_N N_t + A_{N,t+1}$ and $S_{t+1} = \Phi_S S_t + A_{S,t+1}$, $\Phi_N = Q_N[\text{diag}(1, .8)]Q_N'$, $\Phi_S = Q_S[\text{diag}(-1, -.8)]Q_S'$,

$$Q_N = \begin{bmatrix} 1 & .5 \\ .8 & 2 \end{bmatrix}, \quad Q_S = \begin{bmatrix} 2 & -.7 \\ 1 & .8 \end{bmatrix},$$

$\{A_{N,t}\}$ and $\{A_{S,t}\}$ are mutually and serially uncorrelated sequences, $\text{Var}(A_{N,t}) = L_N L_N'$, $\text{Var}(A_{S,t}) = L_S L_S'$,

$$L_N = \begin{bmatrix} .5 & 0 \\ -1.2 & .7 \end{bmatrix} \quad \text{and} \quad L_S = \begin{bmatrix} .8 & 0 \\ 1.1 & .4 \end{bmatrix}.$$

The model (7.93) can be cast into state space form (7.27) and (7.28) by defining $F = \text{diag}(\Phi_N, \Phi_S)$, $G = I_4$, $H = (I_2, I_2)$, $u_t = (A'_{N,t+1}, A'_{S,t+1})'$, $v_t = 0$, $Q = \text{diag}(L_N L_N', L_S L_S')$, $S = 0$ and $R = 0$.

1. Prove that the assumptions in Lemmas 5.9 and 5.10 are satisfied and, therefore, the pair (F, H) is detectable and the pair $(F, GQ^{1/2})$ is stabilizable.
2. Compute the covariance factorization of $\{Y_t\}$, $G_Y(z) = \Psi(z)\Sigma\Psi'(z^{-1})$, where $\Psi(z) = I + zH(I - Fz)^{-1}K$, and $\Psi^{-1}(z) = I - zH(I - F_p z)^{-1}K$. To this end, solve the DARE (7.36) to get

$$P = \begin{bmatrix} .3178 & -.5599 & 0.0670 & 0.0354 \\ & 2.1004 & 0.0738 & 0.1875 \\ & & 0.7141 & 0.9538 \\ & & & 1.5775 \end{bmatrix}, \quad K = \begin{bmatrix} 0.4717 & -0.2297 \\ -0.4419 & 0.4523 \\ -0.6412 & -0.2115 \\ -0.6420 & -0.3128 \end{bmatrix},$$

$$\Sigma = \begin{bmatrix} 1.1659 & 0.5031 \\ & 4.0529 \end{bmatrix} \quad \text{and} \quad F_p = \begin{bmatrix} 0.5783 & 0.1672 & -0.4717 & 0.2297 \\ 0.6419 & 0.2977 & 0.4419 & -0.4523 \\ 0.6412 & 0.2115 & -0.2979 & 0.0898 \\ 0.6420 & 0.3128 & 0.5724 & -0.5480 \end{bmatrix}.$$

Show that the eigenvalues of F_p are 0.6163, −0.5863, 0 and 0, confirming that P is a stabilizing solution.

7.9 Consider the setting of Problem 7.7.

1. Obtain the MSE of the estimators $\hat{x}_{t+1|t}$ and $\hat{x}_{t|\infty}$. To this end, compute $\Lambda_{|\infty}$ by solving the Lyapunov equation (7.45) to get

$$\Lambda_{|\infty} = \begin{bmatrix} 3.2314 & 0.6712 & -0.0309 & 0.1082 \\ & 0.5407 & -0.2919 & 0.2351 \\ & & 2.4596 & -1.0628 \\ & & & 0.9127 \end{bmatrix}.$$

2. Use Theorem 7.13 to compute the weights Ω_j of the estimator $\hat{x}_{t|\infty} = \sum_{j=-\infty}^{\infty} \Omega_j Y_{t+j}$ for $j = \{0, \pm 1, \pm 2\}$. Prove that

$$\Omega_{-2} = \begin{bmatrix} 0.0650 & 0.0041 \\ 0.3053 & -0.0371 \\ -0.0650 & -0.0041 \\ -0.3053 & 0.0371 \end{bmatrix}, \quad \Omega_{-1} = \begin{bmatrix} 0.2119 & -0.0174 \\ 0.0280 & 0.0456 \\ -0.2119 & 0.0174 \\ -0.0280 & -0.0456 \end{bmatrix},$$

$$\Omega_0 = \begin{bmatrix} 0.3748 & -0.1325 \\ -0.7118 & 0.6711 \\ 0.6252 & 0.1325 \\ 0.7118 & 0.3289 \end{bmatrix}, \Omega_1 = \begin{bmatrix} 0.0997 & 0.0275 \\ 0.0174 & 0.1578 \\ -0.0997 & -0.0275 \\ -0.0174 & -0.1578 \end{bmatrix},$$

$$\Omega_2 = \begin{bmatrix} 0.0294 & 0.0467 \\ 0.1171 & -0.0015 \\ -0.0294 & -0.0467 \\ -0.1171 & 0.0015 \end{bmatrix}.$$

7.10 Under the assumptions and with the notation of Problem 7.8, the aim of this problem is to obtain the model

$$\Phi(B)Y_t = \Theta(B)A_t$$

followed by $\{Y_t\}$, where B is the backshift operator, $BY_t = Y_{t-1}$ and $\{A_t\} \sim \mathrm{WN}(0, \Sigma)$, using both polynomial and state space methods.

1. First, consider the equality

$$Y_t = \Phi^{-1}(B)\Theta(B)A_t = (I_2 - \Phi_N B)^{-1} A_{N,t} + (I_2 - \Phi_S B)^{-1} A_{S,t}.$$

Then, pass to covariance generating functions to get

$$\begin{aligned} G_Y(z) &= \Phi^{-1}(z)\Theta(z)\Sigma\Theta'(z^{-1})\Phi^{-1'}(z^{-1}) \\ &= \Phi_N^{-1}(z)\Sigma_N\Phi_N^{-1'}(z^{-1}) + \Phi_S^{-1}(z)\Sigma_S\Phi_S^{-1'}(z^{-1}) \\ &= \Phi_S^{-1}(z)\left[\Phi_S(z)\Phi_N^{-1}(z)\Sigma_N\Phi_N^{-1'}(z^{-1})\Phi_S'(z^{-1}) + \Sigma_S\right]\Phi_S^{-1'}(z^{-1}), \end{aligned}$$

where $\Phi_N(z) = I_2 - \Phi_N z$, $\Phi_S(z) = I_2 - \Phi_S z$, $\Sigma_N = L_N L_N'$ and $\Sigma_S = L_S L_S'$. In addition, using the results in Sect. 5.23, obtain a left coprime MFD, $\hat{\Phi}_N^{-1}(z)\hat{\Phi}_S(z)$, such that $\hat{\Phi}_N^{-1}(z)\hat{\Phi}_S(z) = \Phi_S(z)\Phi_N^{-1}(z)$ and

$$\begin{aligned} G_Y(z) &= \Phi^{-1}(z)\Theta(z)\Sigma\Theta'(z^{-1})\Phi^{-1'}(z^{-1}) \\ &= \Phi_S^{-1}(z)\hat{\Phi}_N^{-1}(z)G_{N,S}(z,z^{-1})\hat{\Phi}_N^{-1'}(z^{-1})\Phi_S^{-1'}(z^{-1}), \end{aligned}$$

where $\Phi(z) = \hat{\Phi}_N(z)\Phi_S(z)$ and

$$G_{N,S}(z,z^{-1}) = \hat{\Phi}_S(z)\Sigma_N\hat{\Phi}_S'(z^{-1}) + \hat{\Phi}_N(z)\Sigma_S\hat{\Phi}_N'(z^{-1}).$$

To conclude, perform the covariance factorization $G_{N,S}(z,z^{-1}) = \Theta(z)\Sigma\Theta'(z^{-1})$ using Wilson's method described in Sect. 3.10.7.

2. Using the results in Sect. 5.23, obtain a left coprime MFD, $\Phi^{-1}(z)\hat{\Theta}(z)$, such that $\Phi^{-1}(z)\hat{\Theta}(z) = zH(I - Fz)^{-1}$ and

$$\begin{aligned} \Psi(z) &= I + zH(I - Fz)^{-1}K \\ &= I + \Phi^{-1}(z)\hat{\Theta}(z)K \\ &= \Phi^{-1}(z)\left[\Phi(z) + \hat{\Theta}(z)K\right]. \end{aligned}$$

Conclude by setting $\Theta(z) = \Phi(z) + \hat{\Theta}(z)K$.

Chapter 8
SSMMATLAB

8.1 Introduction

SSMMATLAB is a set of programs written by the author in MATLAB for the statistical analysis of state space models. The state space model considered is very general. It may have univariate or multivariate observations, time-varying system matrices, exogenous inputs, regression effects, incompletely specified initial conditions, such as those that arise with nonstationary VARMA models, and missing values. It has the form

$$x_{t+1} = W_t\beta + F_t x_t + G_t\epsilon_t,$$
$$Y_t = V_t\beta + H_t x_t + J_t\epsilon_t, \qquad t = 1, \ldots, n,$$

where $\epsilon_t \sim (0, \sigma^2 I)$ and the $\{\epsilon_t\}$ sequence is serially uncorrelated and uncorrelated with x_1, and thus coincides with (4.85) and (4.86). The initial state vector is as in (4.84), that is,

$$x_1 = W\beta + A\delta + x,$$

where $x \sim (a, \sigma^2\Omega)$, the matrices W, A, and Ω are fixed and known, and $\delta \in \mathbb{R}^d$, $\delta \sim (b, \sigma^2\Pi)$ is a random vector that models the unknown initial conditions. As mentioned previously in Sect. 4.14, the notation $v \sim (m, \Sigma)$ means that the vector v has mean m and covariance matrix Σ. It is further assumed that the vectors x and δ are mutually orthogonal.

There are functions to put frequently used models, such as multiplicative ARIMA or VARMA models, cointegrated VARMA models, VARMAX models in echelon form, transfer function models, and univariate structural or ARIMA model-based unobserved components models, into state space form. Once the model is in state space form, other functions can be used for likelihood evaluation, model estimation,

V. Gómez, *Multivariate Time Series With Linear State Space Structure*,
DOI 10.1007/978-3-319-28599-3_8

forecasting, and smoothing. Functions for automatic ARIMA and transfer function identification and automatic outlier detection are also provided. A set of examples illustrates the use of these functions. SSMMATLAB is described in Gómez (2015).

In this chapter, we will group the more important SSMMATLAB functions into several categories with regard to the different sections of this book. In this way, the reader can easily ascertain, for example, whether a specific algorithm described in this book is implemented in SSMMATLAB.

8.2 Kalman Filter and Likelihood Evaluation

In SSMMATLAB, there are functions to implement the two-stage Kalman filter. These functions can be used for likelihood evaluation and are described in detail in the SSMMATLAB documentation. A list of these functions is given in Table 8.1, together with a short description and reference to the section or sections in this book where the relevant algorithms are described.

Table 8.1 SSMMATLAB functions

Function	Sections	Remarks
scakfle2	4.15, 4.21	Two-stage Kalman filter (TSKF), collapsing
scakflesqrt	4.15, 4.4.1, 4.21	TSKF, square root covariance filter, collapsing
scakff	4.15, 4.2.3, 4.16, 4.21	TSKF, state filtering, recursive residuals, collapsing
scakfff	4.15, 4.2.3, 4.16, 4.21	TSKF, state filtering, recursive residuals, fixed regression parameters, collapsing

8.3 Estimation and Residual Diagnostics

To estimate a state space model in SSMMATLAB, the Levenberg–Marquardt Levenberg (1944); Marquardt (1963) method is used. This method minimizes a nonlinear sum of squares. There is also a function to compute sample autocovariance and autocorrelation matrices that can be used to test model adequacy. These functions are listed in Table 8.2, together with a short description and reference to the section in this book where the implemented algorithms are described.

Table 8.2 SSMMATLAB functions

Function	Sections	Remarks
marqdt	4.8	Method is that of Levenberg–Marquardt (Levenberg, 1944; Marquardt, 1963)
mautcov	1.1	Computes sample autocovariance and autocorrelation matrices of residuals

8.4 Smoothing

In SSMMATLAB, there are functions to implement the two-stage Kalman smoother. These functions are described in detail in the SSMMATLAB documentation. A list of these functions is provided in Table 8.3, together with a short description and reference to the sections in this book where the algorithms used are described.

Table 8.3 SSMMATLAB functions

Function	Sections	Remarks
scakfs	4.15, 4.10, 4.21	TSKF, smoothing, collapsing
scakfssqrt	4.15, 4.10, 4.4.1, 4.21	TSKF, smoothing, square root covariance filter, collapsing
smoothgen	4.15, 4.10, 4.21.3, 4.21	TSKF, general vector smoothing, collapsing

8.5 Forecasting

In SSMMATLAB, there is a function for forecasting using the two-stage Kalman filter. It is described in detail in the SSMMATLAB documentation. This function is given in Table 8.4, together with a short description and reference to the section in this book where forecasting is described.

Table 8.4 SSMMATLAB functions

Function	Sections	Remarks
ssmpred	4.9	TSKF, forecasting

8.6 Time Invariant State Space Models

There are functions in SSMMATLAB to handle some specific questions related to time invariant state space models. These functions are described in detail in the SSMMATLAB documentation. A list of these functions is provided in Table 8.5, together with a short description and reference to the section or sections in this book where the relevant algorithms are described.

Table 8.5 SSMMATLAB functions

Function	Sections	Remarks
incossm	4.14.2, 5.8	Initial conditions for the TSKF, time invariant case
dlyapsq	5.3	Solves the discrete time Lyapunov equation
mclyapunov	4.14.2	Solves the continuous time Lyapunov equation
sqrt_ckms	4.15, 5.13, 5.15, 5.8	TSKF, fast CKMS recursions, likelihood evaluation
stair	5.11	Performs the staircase reduction of the pair (A,B)

8.7 ARIMA and Transfer Function Models

Although ARIMA and transfer function models are not specifically discussed in this book, all these models can be put into state space form and can be handled using many of the algorithms described in the previous chapters.

In SSMMATLAB, there are quite a few functions that handle many problems associated with ARIMA and transfer function models. These include automatic model identification, model estimation, forecasting, smoothing, signal extraction, outlier detection, residual diagnostics, etc. These functions are listed in Table 8.6, together with a short description and reference to the section or sections in this book where information related to them is given.

Table 8.6 SSMMATLAB functions

Function	Sections	Remarks
arimam	4.15, 3.7	TSKF, sets up state space model
diferm	5.7	Applies differencing operator to the series
sacspacdif	1.1, 1.8.4	Computes sample autocovariances and autocorrelations
crcreg	5.7	Estimates number of regular and seasonal unit roots in an ARIMA model (Gómez, 2013)
arimaestos	3.7	Automatic identification, estimation and forecasting of ARIMA or transfer function models for one or several series (Gómez, 2009; Gómez & Maravall, 2001a)
cinest	6.6	Estimates parameters in ARIMA model using the Hannan–Rissanen method
lkhev	4.15, 5.8	TSKF, likelihood evaluation
fstlkhev	4.15, 5.13, 5.8	TSKF, CKMS recursions, likelihood evaluation
arimaopt	4.8	Exact maximum likelihood estimation using the Levenberg–Marquardt method (Levenberg, 1944; Marquardt, 1963)
residual2x	4.15, 5.13	TSKF, computes residuals of an ARIMA model using the CKMS recursions
predt	4.15, 4.9	TSKF, forecasting
dsinbut	3.4	Design of a Butterworth filter based on the sine function
dtanbut	3.4	Design of a Butterworth filter based on the tangent function
fasttf	4.15, 5.13, 5.8, 6.3	TSKF, CKMS recursions, likelihood evaluation
pu2ma	3.10.6	Computes the spectral factorization of a scalar covariance generating function
durlev	3.11.1	Applies the Durbin–Levinson algorithm to fit an AR model, computes partial autocorrelations
akaikessm1	3.7	Sets up Akaike's state space representation of minimal dimension
autcov	1.1	Computes sample autocovariances and autocorrelations

8.8 Structural Models

As is the case with ARIMA and transfer function models, structural models are not specifically discussed in this book. However, these models can be easily cast into state space form and can be handled using many of the algorithms described previously in the book.

In SSMMATLAB, there are some functions to handle structural models. A list of these functions is given in Table 8.7, together with a short description and reference to the section in this book where the algorithms used are described.

Table 8.7 SSMMATLAB functions

Function	Sections	Remarks
suusm	4.15	TSKF, sets up state space form corresponding to a structural model
usmestim	4.8	Exact maximum likelihood estimation using the Levenberg–Marquardt method (Levenberg, 1944; Marquardt, 1963)
pr2usm	4.15	TSKF, sets up state space form given the parameters of a structural model

8.9 VARMAX Models

In SSMMATLAB, there are functions to handle many aspects of VARMAX models. These functions are described in detail in the SSMMATLAB documentation. A list of these functions is provided in Table 8.8, together with a short description and reference to the section or sections in this book where the implemented algorithms are described.

Table 8.8 SSMMATLAB functions

Function	Sections	Remarks
suvarmapqPQ	4.15	TSKF, sets up state space model
varmapqPQestim	4.8	Exact maximum likelihood estimation using the Levenberg–Marquardt method (Levenberg, 1944; Marquardt, 1963)
macgf	3.10.5	Computes the autocovariances of a VARMA model
pmspectfac	3.10.7	Computes the spectral factorization of a multivariate covariance generating function
pright2leftcmfd	5.23	Computes a left coprime MFD given a right MFD
estvarmaxkro	6.6, 6.5.1	Estimates a VARMAX model in echelon form using the Hannan–Rissanen method
mhanris	6.6, 6.5.1	Estimates a VARMAX model in echelon form using the Hannan–Rissanen method after fixing some parameters
estvarmaxpqrPQR	6.6, 6.5.1	Estimates a multiplicative VARMAX model using the Hannan–Rissanen method
mexactestim	6.6, 5.13, 5.8	Estimates a VARMAX model in echelon form using the fast CKMS recursions
armaxe2sse	5.9	Passes a VARMAX model in echelon form to state space echelon form
qarmax2ss2	3.7	Sets up a VARMA model into state space form

8.10 Cointegrated VARMA Models

In SSMMATLAB, there are functions to handle several questions regarding cointegrated VARMAX models. These functions are described in detail in the SSMMATLAB documentation and are listed in Table 8.9, together with a short description and reference to the section or sections in this book where the relevant information is given.

Table 8.9 SSMMATLAB functions

Function	Sections	Remarks
mdfestim1r	5.7	Computes differencing polynomial matrix and differenced series
mid2mecf	5.7, 5.7.1	Computes the error correction form
mecf2mid	5.7, 5.7.1	Given the error correction form, computes the differencing polynomial matrix
suvarmapqPQe	5.7, 5.7.1	Sets up state space model
varmapqPQestime	5.7, 5.7.1	Estimates a cointegrated VARMA model

Bibliography

Agee, W. S., & Turner, R. H. (1972). *Triangular decomposition of a positive definite matrix plus a symmetric dyad with aplication to kalman filtering* (Technical Report 38). White Sands Missile Range, New Mexico.

Akaike, H. (1974a). Markovian representation of stochastic processes and its application to the analysis of autoregressive moving average processes. *Annals of the Institute of Statistical Mathematics*, *26*, 363–387.

Akaike, H. (1974b). A new look at the statisitical model identification. *IEEE Transactions on Automatic Control*, *19*, 716–722.

Anda, A. A., & Park, H. (1994). Fast plane rotations with dynamic scaling. *SIAM Journal on Matrix Analysis and Applications*, *15*, 162–174.

Anderson, B. D. O., & Gevers, M. R. (1982). Identificability of linear stochastic systems operating under linear feedback. *Automatica*, *18*, 195–213.

Anderson, B. D. O., & Moore, J. B. (2012). *Optimal filtering*. New York: Dover Publications, Inc.

Anderson, T. W. (1951). Estimating linear restrictions on regression coefficients for multivariate normal distributions. *Annals of Mathematical Statistics*, *22*, 327–351.

Ansley, C. F. (1979). An algorithm for the exact likelihood of a mixed autoregressive-moving average process. *Biometrika*, *66*, 59–65.

Ansley, C. F., & Kohn, R. (1985). Estimation, filtering, and smoothing in state space models with incompletely specified initial conditions. *The Annals of Statistics*, *13*, 1286–1316.

Bell, W. (1984). Signal extraction for nonstationary series. *The Annals of Statistics*, *12*, 646–664.

Bierman, G. J. (1975). The treatment of bias in the square-root information filter/smoother. *Journal of Optimization Theory and Applications*, *16*, 165–178.

Bierman, G. J. (1977). *Factorization methods for discrete sequential estimation*. New York: Academic.

Box, G. E. P., & Jenkins, G. M. (1976). *Time series analysis: Forecasting and control* (revised ed.). San Francisco: Holden-Day.

Box, G. E. P., & Tiao, G. C. (1977). A canonical analysis of multiple time series. *Biometrica*, *64*, 355–365.

Brockett, R. W. (1970). *Finite dimensional linear systems*. New York: Wiley.

Brockwell, P. J., & Davis, R. A. (1991). *Time series: Theory and methods*. New York: Springer.

Bryson, A. E., & Frazier, M. (1963). Smoothing for linear and nonlinear dynamic systems. In *Proceedings of the optimum systems synthesis conference* (U. S. Air Force Technical Report ASD-TDR, pp. 63–119).

Bryson, A. E., & Ho, Y. C. (1969). *Applied optimal control*. Wattham: Ginn (Blaisdell).

V. Gómez, *Multivariate Time Series With Linear State Space Structure*,
DOI 10.1007/978-3-319-28599-3

Burg, J. P. (1968). A new analysis technique for time series data. In *Modern spectrum analysis (edited by d. g. childers), nato advanced study institute of signal processing with emphasis on underwater acoustics.* New York: IEEE.

Burman, J. P. (1980). Seasonal adjustment by signal extraction. *Journal of the Royal Statistical Society Series A*, *143*, 321–337.

Burridge, P., & Wallis, K. (1988). Prediction theory for autoregressive moving average processes. *Econometric Reviews*, *7*, 65–95.

Businger, P., & Golub, G. H. (1965). Linear least-squares solution by housholder transformations. *Numerische Mathematik*, *7*, 269–276.

Butterworth, S. (1930). On the theory of filter amplifiers. *Experimental Wireless and the Wireless Engineer*, *7*, 536–541.

Caines, P. E. (1988). *Linear stochastic systems*. New York: Wiley.

Caines, P. E., & Chan, C. (1975). Feedback between stationary stochastic processes. *IEEE Transactions on Automatic Control*, *20*, 498–508.

Chan, S. W., Goodwin, G. C., & Sin, K. S. (1984). Convergence properties of the riccati difference equation in optimal filtering of nonstabilizable systems. *IEEE Transactions on Automatic Control*, *29*, 110–118.

Chandrasekhar, S. (1947a). On the radiative equilibrium of a stellar atmosphere. *Astrophysical Journal*, *106*, 152–216.

Chandrasekhar, S. (1947b). On the radiative equilibrium of a stellar atmosphere. *Astrophysical Journal*, *107*, 48–72.

Chen, C. T. (1984). *Linear system theory and design* (3rd ed.). New York: Holt, Rinehart and Winston, Inc..

Cleveland, W. P., & Tiao, G. C. (1976). Decomposition of seasonal time series: A model for the X-11 Program. *Journal of the American Statistical Association*, *71*, 581–587.

Cooley, J. W., & Tukey, J. W. (1965). An algorithm for the machine calculation of complex fourier series. *Mathematics of Computations*, *19*, 297–301.

Cramér, H. (1940). On the theory of stationary random processes. *Annals of Mathematics*, *41*, 215–230.

de Jong, P. (1991). The diffuse kalman filter. *The Annals of Statistics*, *19*, 1073–1083.

de Jong, P., & Chu Chun Lin, S. F. (1994). Fast likelihood evaluation and prediction for nonstationary state space models. *Biometrika*, *81*, 133–142.

de Jong, P., & Chu Chun Lin, S. F. (2003). Smoothing with and unknown initial condition. *Journal of Time Series Analysis*, *24*, 141–148.

de Jong, P., & Penzer, J. (1998). Diagnosing schocks in time series. *Journal of the American Statistical Association*, *93*, 796–806.

de Souza, C. E., Gevers, M. R., & Goodwin, G. C. (1986). Riccati equations in optimal filtering of nonstabilizable systems having singular state transition matrices. *IEEE Transactions on Automatic Control*, *31*, 831–838.

Denham, M. J. (1974). Canonical forms for the identification of linear multivariable systems. *IEEE Transactions on Automatic Control*, *AC-19*, 646–656.

Doménech, R., & Gómez, V. (2006). Estimating potential output, core inflation, and the nairu as latent variables. *Journal of Business and Economic Statistics*, *24*, 354–365.

Doob, J. L. (1944). The elementary gaussian processes. *Annals of Mathematical Statistics*, *15*, 229–282.

Doob, J. L. (1953). *Stochastic processes*. New York: Wiley.

Durbin, J. (1960). The fitting of time series models. *Review of the International Statistical Instutute*, *28*, 233–243.

Durbin, J., & Koopman, S. J. (2012). *Time series analysis by state space methods* (2nd ed.). Oxford: Oxford University Press.

Engle, R. F., & Granger, C. W. J. (1987). Cointegration and error correction: Representation, estimation and testing. *Econometrica*, *55*, 251–276.

Francke, M., Koopman, S. J., & de Vos, A. F. (2010). Likelihood functions for state space models wih diffuse initial conditions. *Time Series Analysis*, *31*, 407–414.

Fréchet, M. (1937). *Recherches théoriques modernes sur le calcul des probalitités* (Vol. 1). Paris: Gauthier-Villars.

Friedland, B. (1969). Treatment of bias in recursive filtering. *IEEE Transactions on Automatic Control*, *AC-14*, 387–390.

Gentleman, M. (1973). Least squares computations by givens transformations. *Journal of the Institute of Mathematics and its Applications*, *12*, 329–336.

Gevers, M. R. (1986). Arma models, their kronecker indices and their mcmillan degree. *International Journal of Control*, *43*, 1745–1761.

Gikhman, I. I., & Skorokhod, A. V. (1969). *Introduction to the theory of random processes*. New York: Dover Publications, Inc.

Glover, K., & Willems, J. C. (1974). Parmetrization of linear dynamical systems, canonical forms and identifiability. *IEEE Transactions on Automatic Control*, *AC-19*, 640–646.

Gohbergh, I. C., & Fel´dman, I. A. (1974). *Convolution equations and projection methods for their solution* (Vol. 41). Providence: American Mathematical Society.

Golub, G. H. (1965). Numerical methods for solving linear least squares problems. *Numerische Mathematik*, *7*, 206–216.

Golub, G. H., & Van Loan, C. F. (1996). *Matrix computations* (3rd ed.). Baltimore: The John Hopkins University Press.

Gómez, V. (1999). Three equivalent methods for filtering finite nonstationary time series. *Journal of Business and Economic Statistics*, *17*, 109–116.

Gómez, V. (2001). The use of butterworth filters for trend and cycle estimation in economic time series. *Journal of Business and Economic Statistics*, *19*, 365–373.

Gómez, V. (2006). Wiener-kolmogorov filtering and smoothing for multivariate series with state-space structure. *Journal of Time Series Analysis*, *28*, 361–385.

Gómez, V. (2009). Transfer function model identification. *Boletín de Estadística e Investigación Operativa*, *25*, 99–105.

Gómez, V. (2013). A strongly consistent criterion to decide between i(1) and i(0) processes based on different convergence rates. *Communications in Statistics-Simulation and Computation*, *42*, 1848–1864.

Gómez, V. (2014). *Ssmmatlab.* Retrieved from http://www.sepg.pap.minhap.gob.es/sitios/sepg/en-GB/Presupuestos/Documentacion/Paginas/SSMMATLAB.aspx.

Gómez, V. (2015). Ssmmatlab: A set of matlab programs for the statistical analysis of state space models. *Journal of Statistical Software*, *66*(9), 1–37. Retrieved from http://www.jstatsoft.org/v66/i09.

Gómez, V., Aparicio-Pérez, F., & Sánchez-Ávila, A. (2010). An alternative to transfer function forecasting based on subspace methods. *Communications in Statistics-Simulation and Computation*, *39*, 1855–1867.

Gómez, V., & Maravall, A. (1994a). Estimation, prediction, and interpolation for nonstationary series with the kalman filter. *Journal of the American Statistical Association*, *89*, 611–624.

Gómez, V., & Maravall, A. (1994b). Initializing the kalman filter with incompletely specified initial conditions. In G. R. Chen (Ed.), *Approximate kalman filtering (series on approximation and decomposition).* London: World Scientific Publishing Company.

Gómez, V., & Maravall, A. (2001a). Automatic modeling methods for univariate time series. In D. Peña, G. C. Tiao, & R. S. Tsay (Eds.), *A course in time series analysis* (pp. 171–201). New York: Wiley.

Gómez, V., & Maravall, A. (2001b). Seasonal adjustment and signal extraction in economic time series. In D. Peña, G. C. Tiao, & R. S. Tsay (Eds.), *A course in time series analysis* (pp. 202–246). New York: Wiley.

Gómez, V., Maravall, A., & Peña, D. (1999). Missing observations in arima models: Skipping approach versus additive outlier approach. *Journal of Econometrics*, *88*, 341–363.

Granger, C. W. J. (1963). Economic processes involving feedback. *Information and Control*, *6*, 28–48.

Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, *37*, 428–438.

Granger, C. W. J. (1981). Some properties of time series data and their use in econometric model specification. *Journal of Econometrics*, *16*, 121–130.

Granger, C. W. J., & Newbold, P. (1977). *Forecasting economic time series*. New York: Academic.

Grenander, U., & Rosenblatt, M. (1957). *Statistical analysis of time series*. New York: Wiley.

Guidorzi, R. P. (1981). Invariants and canonical forms for systems structural and parametric identification. *Automatica*, *17*, 117–133.

Hammarling, S. (1991). Numerical solution of the discrete-time, convergent, non-negative definite lyapunov equation. *Systems and Control Letters*, *17*, 137–139.

Hannan, E. J. (1970). *Multiple time series*. New York: Wiley.

Hannan, E. J., & Deistler, M. (1988). *The statistical theory of linear systems*. New York: Wiley.

Hannan, E. J., & Kavalieris, L. (1984). Multivariate linear time series models. *Advances in Applied Probability*, *16*, 492–561.

Hannan, E. J., & Kavalieris, L. (1986). Regression, autoregression models. *Journal of Time Series Analysis*, *7*, 27–49.

Hannan, E. J., & Rissanen, J. (1982). Recursive estimation of mixed autoregressive-moving average order. *Biometrika*, *69*, 81–94.

Harrison, P. J., & Stevens, C. F. (1976). Bayesian forecasting. *Journal of the Royal Statistical Society Series B*, *38*, 205–247.

Harvey, A. C. (1989). *Forecasting, structural time series models and the kalman filter*. Cambridge: Cambridge University Press.

Harvey, A. C., & Trimbur, T. M. (2003). General model-based filters for extracting cycles and trends in economic time series. *Review of Economics and Statistics*, *85*(2), 244–255.

Housholder, A. S. (1953). *Principles of numerical analysis*. New York: Mc-Graw Hill.

Hsieh, S. F., Liu, K. J. R., & Yao, K. (1993). A unified square-root-free approach for qrd-based recursive least-squares estimation. *IEEE Transaction on Signal Processing*, *41*, 1405–1409.

Ignagni, M. B. (1981). An alternate derivation and extension of friedland's two-stage kalman estimator. *IEEE Transactions on Automatic Control*, *AC-26*, 746–750.

Ionescu, V., Oară, C., & Weiss, M. (1997). General matrix pencil techniques for the solution of algebraic riccati equations: A unified approach. *IEEE Transactions on Automatic Control*, *42*, 1085–1097.

Johansen, S. (1995). *Likelihood-based inference in cointegrated vector autoregressive models*. Oxford: Oxford University Press.

Jover, J. M., & Kailath, T. (1986). A parallel architecture for kalman filter measurement update and parameter estimation. *Automatica*, *22*, 43–57.

Kailath, T. (1980). *Linear systems*. Englewood Cliffs: Prentice Hall.

Kailath, T., Morf, M., & Sidhu, G. S. (1973). Some new algorithms for recursive estimation in constan linear discrete-time systems. In *Proceedings of the seventh princeton conference on information science and systems* (pp. 344–352). Princeton: Princeton University.

Kailath, T., Sayed, A. H., & Hassibi, B. (2000). *Linear estimation*. Upper Saddle River: Prentice Hall.

Kalbfleisch, J. D., & Sprott, D. A. (1970). Application of likelihood methods to models involving large numbers of parameters. *Journal of the Royal Statistical Society Series B*, *32*, 175–208.

Kalman, R. E. (1960a). Contributions to the theory of optimal control. *Boletín de la Sociedad Matemática Mejicana*, *5*, 102–119.

Kalman, R. E. (1960b). A new approach to linear filtering and prediction problems. *Transactions of ASME, Journal of Basic Engineering*, *82*, 34–45.

Kalman, R. E. (1960c). On the general theory of control systems. In *Proceedings of the first IFAC congress, London* (Vol. 1, pp. 481–491).

Kalman, R. E. (1963). New methods of wiener filtering theory. In J. L. Bogdanoff & F. Kozin (Eds.), *First symposium of engineering applications of random function theory and probability* (pp. 279–388). New York: Wiley.

Kalman, R. E., Falb, P. L., & Arbib, M. A. (1969). *Topics in mathematical system theory*. New York: Mc-Graw Hill.

Katayama, T. (2005). *Subspace methods for system indentification*. Leipziz: Springer.

Katayama, T., & Picci, G. (1999). Realization of stochastic systems with exogenous inputs and subspace identification methods. *Automatica*, *35*, 1635–1652.

Kitagawa, G., & Gersch, W. (1996). *Smoothness priors analysis of time series*. New York: Springer.

Kolmogorov, A. N. (1939). Sur l'interpolation et extrapolation des suites stationaries. *Comptes Rendus de l'Académie des Sciences Série I Mathématique*, *208*, 2043–2045.

Kolmogorov, A. N. (1941). *Stationary sequences in hilbert space (in russian). Bulletin of Mathematics*. (Vol. 2). University of Moscow [A translation by N. Artin is available in many libraries].

Koopman, S. J. (1997). Exact initial kalman filtering and smoothing for nonstationary time series models. *Journal of the American Statistical Association*, *92*, 1630–1638.

Koopman, S. J., & Harvey, A. C. (2003). Computing observation weights for signal extraction and filtering. *Journal of Economic Dynamics and Control*, *27*, 1317–1333.

Larimore, W. E. (1983). System identification, reduced-order filtering and modeling via canonical variate analysis. In *Proceedings of 1983 American control conference* (pp. 445–451).

Levenberg, K. (1944). A method for the solution of certain non-linear problems in least squares. *The Quarterly of Applied Mathematics*, *2*, 164–168.

Levinson, N. (1947). The wiener rms error criterion in filter design and prediction. *Journal of Mathematical Physics*, *25*, 261–278.

Lindquist, A., & Picci, G. (2015). *Linear stochastic systems a geometric approach to modeling, estimations and identification*. Berlin: Springer.

Luenberger, D. G. (1967). Canonical forms for linear multivariable systems. *IEEE Transactions on Automatic Control*, *AC-12*, 290–293.

Lütkepohl, H. (2007). *New introduction to multiple time series analysis*. Berlin: Springer.

Marquardt, D. (1963). An algorithm for least-squares estimation of nonlinear parameters. *SIAM Journal on Applied Mathematics*, *11*, 431–441.

McElroy, T. S. (2008). Matrix formulas for non-stationary signal extraction. *Econometric Theory*, *24*, 1–22.

Meditch, J. S. (1973). A survey of data smoothing for linear and nonlinear systems. *Automatica*, *9*, 151–162.

Mittnik, S. (1993). Computing theoretical autocovariances of multivariate autoregressive moving average models by using a block levinson algorithm. *Journal of the Royal Statistical Society Series B*, *55*, 435–440.

Morf, M., & Kailath, T. (1975). Square root algorithms for least squares estimation. *IEEE Transactions for Automatic Control*, *AC-20*, 487–497.

Morf, M., Sidhu, G. S., & Kailath, T. (1974). Some new algorithms for recursive estimation in constant, linear, discrete-time systems. *IEEE Transactions on Atomatic Control*, *19*, 315–323.

Morf, M., Vieira, A., & Kailath, T. (1978). Covariance characterization by partial autocorrelation matrices. *The Annals of Statistics*, *6*, 643–648.

Morf, M., Vieira, A., Lee, D. T. L., & Kailath, T. (1978). Recursive multichannel maximum entropy spectral estimation. *IEEE Transactions on Geoscience Electronics*, *16*, 85–94.

Nerlove, M. (1967). In W. Fellner, et al. (Eds.), *Ten economic studies in the tradition of Irving fisher* (pp. 128–169). New York: Wiley.

Potter, J. E., & Stern, R. G. (1963). Statistical filtering of space navigation measurements. In E. A. L. Massachusetts Institute of Technology (Ed.), *The AIAA guidance, navigation, and control conference.*

Pourahmadi, M. (2001). *Foundations of time series analysis and prediction theory*. New York: Wiley.

Quenouille, M. H. (1957). *The analysis of multiple time series*. London: Griffin.

Rauch, H. E., Tung, F., & Striebel, C. T. (1965). Maximum likelihood estimates of linear dynamic systems. *AIAA Journal*, *3*(8), 1445–1450.

Reinsel, G. C. (1997). *Elements of multivariate time series analysis*. New York: Springer.

Rissanen, J. (1974). Basis of invariants and canonical forms for linear dynamic systems. *Automatica*, *70*, 175–182.

Rosenberg, B. (1973). Random coefficients models: The analysis of a cross-section of time series by stochasticaly convergent parameter regression. *Annals of Economic and Social Measurement*, *2*, 399–428.

Rozanov, Y. A. (1967). *Stationary random processes*. San Francisco: Holden Day.

Sayed, A. H., & Kailath, T. (2001). A survey of spectral factorization methods. *Numerical Linear Algebra With Applications*, *8*, 467–496.

Schmidt, S. F. (1970). Computational techniques in kalman filtering. In *Theory and applications of kalman filtering* (Agardograph 139, Technical Report).

Stewart, G. W. (1995). *Theory of the combination of observations least subject to errors [Translation of original works by C. F. Gauss under the title: Theoria combinationis observationum erroribus minimis obnoxiae]. Classics in Applied Mathematics*. Philadelphia: SIAM.

Tiao, G. C., & Tsay, R. S. (1989). Model specification in multivariate time series. *Journal of the Royal Statistical Society Series B*, *51(2)*, 157–213.

Tsay, R. S. (1989). Parsimonious parameterization of vector autoregressive moving average models. *Journal of Business and Economic Statistics*, *7*, 327–341.

Tunnicliffe-Wilson, G. (1972). The factorization of matricial spectral densities. *SIAM Journal on Applied Mathematics*, *23*, 420–426.

Van Dooren, P. M. (1979). The computation of kronecker's canonical form of a singular pencil. *Linear Algebra and Its Applications*, *27*, 103–141.

Van Dooren, P. M. (2003). *Numerical linear algebra for linear systems and control.* Retrieved from http://perso.uclouvain.be/paul.vandooren/PVDnotes.pdf (Course Notes).

Van Overschee, P., & De Moor, B. (1994). N4sid: Subspace algorithms for the identification of combined deterministic-stochastic systems. *Automatica*, *30*, 75–93.

Van Overschee, P., & De Moor, B. (1996). *Subspace identification for linear systems*. Boston: Kluwer Academic.

Verghese, G., & Kailath, T. (1979). A further note on backwards markovian models. *IEEE Transactions on Information Theory*, *IT-25*, 121–124.

Wang, M. C., & Uhlenbeck, G. E. (1945). On the theory of brownian motion, ii. *Review of Modern Physics*, *17*, 323–342.

Whittle, P. (1963a). On the fitting of multivariate autoregressions, and the approximate canonical factorization of a spectral density matrix. *Biometrika*, *50*, 129–134.

Whittle, P. (1963b). *Prediction and regulation by linear least-square methods*. Princeton: Van Nostrand.

Whittle, P. (1983). *Prediction and regulation by linear least-square methods* (2nd ed.). Minneapolis: University of Minnesota Press.

Wiener, N. (1949). *The extrapolation, interpolation and smoothing of stationary time series with engineering applications*. New York: Wiley. [Originally published in 1942 as a classified Nat. Defense Res. Council Report].

Wiener, N., & Hopf, E. (1931). Über eine klasse singulärer integralgleichungen. *Sitzungsberichte der Preussischen Akademie der Wissenschaften zu Berlin, Physikalisch-Mathematische Klasse*, *31*, 696–706.

Wold, H. O. A. (1938). *A study in the analysis of stationary time series*. Uppsala: Almqvist & Wiksell.

Author Index

V. Gómez, *Multivariate Time Series With Linear State Space Structure*,
DOI 10.1007/978-3-319-28599-3

Subject Index

V. Gómez, *Multivariate Time Series With Linear State Space Structure*,
DOI 10.1007/978-3-319-28599-3