7

CONTINUOUS STATE HMM

7.1 CONTINUOUS AND DISCRETE STATE HMM

Continuous state HMMs are very popular in many fields that include control theory, signal processing, speech and image recognition, finance and many others. The application of a general continuous time HMM is much more difficult than of the discrete state HMM because it usually requires to compute multidimensional integrals rather than multiply matrices. There is one class of the continuous state HMM — hidden Gauss-Markov processes (HGMM) for which there are closed form expressions for the multidimensional integrals. This class has been studied intensively by many researchers who developed a rich theory related to the so-called state space linear systems. There are many textbooks and monographs devoted to this theory.

In contrast with the traditional approach, we present this theory as a generalization of the theory of the discrete state HMM developed in the previous chapters. We derive the major facts and algorithms of the theory on the basis of computing certain probability density functions similarly to Refs. 1,4 while the traditional approach is more "statistical" and is based on the estimation theory. [2,17] The probabilistic approach allowed us not only to simplify the presentation, but also to explain the relationship between the discrete state and continuous state HMM theories.

We consider first general continuous state HMM and show that the introduction of the operator probability allows us to generalize the results of the matrix probability theory presented in Chapter 2. This general theory is than applied to the HGMM in which case the calculation of the multidimensional integrals is reduced to calculations of the corresponding moments (means and covariance matrices). As an example of this approach, we show that the forward algorithm can be realized using the Kalman filter. The Baum-Welch algorithm considered in Chapter 3 is realized by computing the corresponding moments.

In the case of non-Gaussian continuous state HMMs, the integrals can be computed numerically by one of the numerical methods which can be interpreted as an approximation of the continuous state HMM by the discrete state HMM which allows us to use all the results of the previous chapters directly. We call this discrete state HMM the skeleton of the continuous state HMM. The main reason for using discrete state HMMs is the simplicity of representing various probabilities using matrix algebra. The HGMM play a similar role allowing us to use matrix algebra for computing the corresponding covariance matrices, but at the cost of narrowing its application only to Gauss-Markov processes. It can also be applied to the non-Gaussian HMMs whose state transition PDFs are Gaussian. The theory

W. Turin, *Performance Analysis and Modeling of Digital Transmission Systems*

can be also extended to the combination discrete-continuous state space.

As we mentioned before, an HMM can be defined as a Markov process † whose states are partially observable. Consider a Markov process whose state at the moment t has the form ‡ $\mathbf{x}_t = (\boldsymbol{s}_t, \mathbf{a}_t)$ where $\boldsymbol{s}_t \in S$ and $\mathbf{a}_t \in A$ represent the unobservable and observable portion of the state, respectively; $\boldsymbol{s}_t$ is called the *hidden* state, $\mathbf{a}_t$ is called the *observation*, and $\mathbf{x}_t$ is called the *complete* state. The process is called an HMM if its state transition PDF has the following special form

$$p_{t_1 t_2}(\mathbf{x}_{t_2} \mid \mathbf{x}_{t_1}) = p_{t_1 t_2}(\mathbf{x}_{t_2} \mid \boldsymbol{s}_{t_1}) = p_{t_1 t_2}(\boldsymbol{s}_{t_2}, \mathbf{a}_{t_2} \mid \boldsymbol{s}_{t_1}), \quad t_1 < t_2 \tag{7.1.1}$$

This means that for any $t_1 < t_2 < \ldots < t_n$ and any sequences of states

$$p(\mathbf{x}_{t_n} \mid \mathbf{x}_{t_1}, \mathbf{x}_{t_2}, \ldots, \mathbf{x}_{t_{n-1}}) = p(\boldsymbol{s}_{t_n}, \mathbf{a}_{t_n} \mid \boldsymbol{s}_{t_{n-1}})$$

and

$$p(\mathbf{x}_{t_1}, \mathbf{x}_{t_2}, \ldots, \mathbf{x}_{t_n} \mid \boldsymbol{s}_{t_0}) = p(\boldsymbol{s}_{t_1}, \mathbf{a}_{t_1} \mid \boldsymbol{s}_{t_0}) p(\boldsymbol{s}_{t_2}, \mathbf{a}_{t_2} \mid \boldsymbol{s}_{t_1}) \cdots p(\boldsymbol{s}_{t_n}, \mathbf{a}_{t_n} \mid \boldsymbol{s}_{t_{n-1}}) \tag{7.1.2}$$

The hidden portion of the state is a Markov process with the transition PDF

$$p_{t_1 t_2}(\boldsymbol{s}_{t_2} \mid \boldsymbol{s}_{t_1}) = \int_A p_{t_1 t_2}(s_{t_2}, \mathbf{a}_{t_2} \mid \boldsymbol{s}_{t_1}) d\mathbf{a}_{t_2}$$

where $\int_A$ is a shorthand for a multiple integral over the region A of the observation space of vectors $\mathbf{a}_t = (a_{1,t}, a_{2,t}, \ldots, a_{k,t})$ and $d\mathbf{a}_t = da_{1,t} da_{2,t} \cdots da_{k,t}$ denotes the element of volume in this space.

The observation sequence $\mathbf{a}_t$ is not a Markov process in the general case. Note, however, that any Markov process whose states are partially observable can be presented as an HMM by increasing the state space. Indeed, suppose that the state transition PDF does not satisfy Eq. (7.1.1). In this case, we can treat the Markov process $\mathbf{x}_t = (\boldsymbol{s}_t, \mathbf{a}_t)$ as nonobservable and the observation $\mathbf{b}_t$ in state $(\boldsymbol{s}_t, \mathbf{a}_t)$ is $\mathbf{a}_t$. In other words, we consider a new process $\mathbf{y}_t = (\mathbf{x}_t, \mathbf{b}_t)$ whose state transition PDF has the form

$$p_{t_1 t_2}(\mathbf{y}_{t_2} \mid \mathbf{y}_{t_1}) = p_{t_1 t_2}(\mathbf{x}_{t_2}, \mathbf{b}_{t_2} \mid \mathbf{x}_{t_1}) = p_{t_1 t_2}(\boldsymbol{s}_{t_2}, \mathbf{a}_{t_2} \mid \mathbf{x}_{t_1}) \delta(\mathbf{b}_2 - \mathbf{a}_2)$$

In the most popular in applications case, $p_{t_1 t_2}(\mathbf{a}_{t_2} \mid \boldsymbol{s}_{t_1}, \boldsymbol{s}_{t_2}) = p_{t_1 t_2}(\mathbf{a}_{t_2} \mid \boldsymbol{s}_{t_2})$ the observation depends only on the current state. In this case, we can write

$$p_{t_1 t_2}(\boldsymbol{s}_{t_2}, \mathbf{a}_{t_2} \mid \boldsymbol{s}_{t_1}) = p_{t_1 t_2}(\boldsymbol{s}_{t_2} \mid \boldsymbol{s}_{t_1}) p_{t_1 t_2}(\mathbf{a}_{t_2} \mid \boldsymbol{s}_{t_2})$$

For a discrete-time HMM, we assume, without loss of generality, that $t_i = i$ and denote the one-step transition PDF as

$$p_{t-1,t}(\boldsymbol{s}_t, \mathbf{a}_t \mid \boldsymbol{s}_{t-1}) = p_t(\boldsymbol{s}_t, \mathbf{a}_t \mid \boldsymbol{s}_{t-1})$$

We consider only the discrete-time HMMs in the sequel. For the discrete-time HMM, the PDF of the observation sequence $\mathbf{a}_1^T$ can be written as

---

† A brief description of the Markov processes properties used in this chapter is given in Appendix 7.

‡ In this chapter, $\mathbf{a} = (a_1, a_2, \ldots, a_k)$ denotes a column vector while $\mathbf{b} = [a_1, a_2, \ldots, a_k] = \mathbf{a}'$ is a row vector.

$$p(\mathbf{a}_1^T) = \int_{S^{T+1}} p_0(\boldsymbol{s}_0) \prod_{i=1}^{T} p_i(\boldsymbol{s}_i, \mathbf{a}_i \mid \boldsymbol{s}_{i-1}) d\boldsymbol{s}_0^T \tag{7.1.3}$$

The process is called *homogeneous* if its state transition PDF depends only on time difference $t_2 - t_1$. A discrete-time continuous-state homogeneous HMM is described by the state initial PDF $p_0(s_0)$ and the one-step transition PDF which does not depend on t: $p_t(\boldsymbol{s}_t, \mathbf{a}_t \mid \boldsymbol{s}_{t-1}) = p(\boldsymbol{s}_t, \mathbf{a}_t \mid \boldsymbol{s}_{t-1})$.

We can define, similarly to Sec. 1.1, the input-output HMM whose observation consists of a pair $(\mathbf{b}_t, \mathbf{a}_t)$ of an input (source) and corresponding output. The output is always observable, while the input is observable during the system identification (training) and unobservable otherwise.

A discrete space HMM is a special case of the continuous space HMM when the transition PDF has the form

$$p(\boldsymbol{s}_t, \mathbf{a}_t \mid \boldsymbol{s}_{t-1} = i) = \sum_j p_{ij} \delta(\boldsymbol{s}_t - j)$$

On the other hand, the continuous state HMM can be viewed as a limiting case of the discrete state HMM. Thus, a problem related to a continuous state HMM can be solved numerically by approximating it with the discrete state HMM. (This approximation is usually called state quantization.) The goal of this chapter is to generalize the methods developed in the previous chapters for the discrete state HMMs to the continuous state HMMs.

7.2 OPERATOR PROBABILITY

In the previous chapters, we demonstrated the usefulness of the matrix probabilities in applications of discrete state HMMs. In this chapter, we generalize the notion of the matrix probability and define the operator probabilities which are applicable to continuous state HMMs.

Let us start by developing a formula for the PDF of the observation sequence $\mathbf{a}_1^T$. Since the sequence of hidden states is a Markov process, the two-step transition PDF can be evaluated using the Chapman-Kolmogorov equation (see Appendix 7).

$$p(\boldsymbol{s}_2, \mathbf{a}_1, \mathbf{a}_2 \mid \boldsymbol{s}_0) = \int_S p_2(\boldsymbol{s}_2, \mathbf{a}_2 \mid \boldsymbol{s}_1) p_1(\boldsymbol{s}_1, \mathbf{a}_1 \mid \boldsymbol{s}_0) d\boldsymbol{s}_1 \tag{7.2.1}$$

The right hand side of this equation is a generalization of the product of matrices (see Appendix 7) and can be formalized using the integral (Fredholm) operators.

Indeed, let us define the operator $\mathbf{P}_t(\mathbf{a})$ as mapping to itself of the set of functions $\{f(x)\}$ for which the following integral exists

$$f\mathbf{P}_t(\mathbf{a}) = \int_S f(\boldsymbol{s}) p_t(\boldsymbol{x}, \mathbf{a} \mid \boldsymbol{s}) d\boldsymbol{s}$$

This definition generalizes the multiplication of a row-vector by the matrix. The mapping

$$\mathbf{P}_t(\mathbf{a}) f = \int_S f(\boldsymbol{x}) p_t(\boldsymbol{x}, \mathbf{a} \mid \boldsymbol{s}) d\boldsymbol{x}$$

generalizes the multiplication of a matrix by a column-vector. The function $p_t(\boldsymbol{s}, \mathbf{a} \mid \boldsymbol{x})$ is called a *kernel* of the operator $\mathbf{P}(\mathbf{a})$. Thus, for the same kernel, we defined two different operators depending on the selection of the variable of integration ($\boldsymbol{s}$ or $\boldsymbol{x}$) which is denoted

symbolically by the order of the operands.

Equation (7.2.1) defines the kernel of the the two-step operator $\mathbf{P}(\mathbf{a}_1,\mathbf{a}_2) = \mathbf{P}_1(\mathbf{a}_1)\mathbf{P}_2(\mathbf{a}_2)$ which we call the product of operators. The kernel of the operator

$$\mathbf{P}(\mathbf{a}_1^T) = \mathbf{P}_1(\mathbf{a}_1)\mathbf{P}_2(\mathbf{a}_2)\cdots\mathbf{P}_T(\mathbf{a}_T) = \prod_{i=1}^{T}\mathbf{P}_i(\mathbf{a}_i) \tag{7.2.2}$$

is the transition PDF

$$p(\boldsymbol{s}_T,\mathbf{a}_1^T \mid \boldsymbol{s}_0) = \int_{S^{T-1}}\prod_{i=1}^{T} p_i(\boldsymbol{s}_i,\mathbf{a}_i \mid \boldsymbol{s}_{i-1})\,d\boldsymbol{s}_1^{T-1}$$

This kernel can be evaluated using the following forward algorithm

$$p(\boldsymbol{s}_k,\mathbf{a}_1^k \mid \boldsymbol{s}_0) = \int_S p_k(\boldsymbol{s}_k,\mathbf{a}_k \mid \boldsymbol{s}_{k-1})p(\boldsymbol{s}_{k-1},\mathbf{a}_1^{k-1} \mid \boldsymbol{s}_0)\,d\boldsymbol{s}_{k-1}, \quad \text{for } k=2,3,\ldots,T$$

which can be written in the operator form as

$$\mathbf{P}(\mathbf{a}_1^k) = \mathbf{P}(\mathbf{a}_1^{k-1})\mathbf{P}(\mathbf{a}_k), \quad k=2,3,\ldots,T$$

It can also be evaluated by the backward algorithm

$$\mathbf{P}(\mathbf{a}_k^T) = \mathbf{P}(\mathbf{a}_k)\mathbf{P}(\mathbf{a}_{k+1}^T), \quad k=T-1,T-2,\ldots,1$$

or by the combination of both to obtain $\mathbf{P}(\mathbf{a}_1^T) = \mathbf{P}(\mathbf{a}_1^r)\mathbf{P}(\mathbf{a}_{r+1}^T)$ where $1<r<T$.

The PDF of the observation sequence $\mathbf{a}_1^T$ can be expressed using these operators as

$$p(\mathbf{a}_1^T) = \mathbf{p}_0\mathbf{P}(\mathbf{a}_1)\mathbf{P}(\mathbf{a}_2)\cdots\mathbf{P}(\mathbf{a}_T)\mathbf{1} = \mathbf{p}_0\prod_{i=1}^{T}\mathbf{P}(\mathbf{a}_i)\mathbf{1} \tag{7.2.3}$$

where $\mathbf{p}_0 = p_0(\boldsymbol{s}_0)$ is the initial state PDF and $\mathbf{1}$ denotes the function $f(x)\equiv 1$. Note that this equation represents the operator form of Eq. (7.1.3). We can rewrite Eq. (7.2.3) as $p(\mathbf{a}_1^T) = \boldsymbol{\alpha}(\mathbf{a}_1^T)\mathbf{1}$ where

$$\boldsymbol{\alpha}(\mathbf{a}_1^T) = \mathbf{p}_0\mathbf{P}(\mathbf{a}_1^T) = p(\boldsymbol{s}_T,\mathbf{a}_1^T) \tag{7.2.4}$$

is the forward PDF which can be evaluated using the forward algorithm

$$\boldsymbol{\alpha}(\mathbf{a}_1^0) = \mathbf{p}_0, \quad \boldsymbol{\alpha}(\mathbf{a}_1^k) = \boldsymbol{\alpha}(\mathbf{a}_1^{k-1})\mathbf{P}(\mathbf{a}_k) \tag{7.2.5a}$$

$k=1,2,\ldots,T$. This operator equation is just a shorthand for the following scalar equation

$$\alpha(\boldsymbol{s}_k,\mathbf{a}_1^k) = \int_S \alpha(\boldsymbol{s}_{k-1},\mathbf{a}_1^{k-1})p(\boldsymbol{s}_k,\mathbf{a}_k \mid \boldsymbol{s}_{k-1})\,d\boldsymbol{s}_{k-1} \tag{7.2.5b}$$

We can also rewrite Eq. (7.2.3) as $p(\mathbf{a}_1^T) = \mathbf{p}_0\boldsymbol{\beta}(\mathbf{a}_1^T)$ where the backward PDF

$$\boldsymbol{\beta}(\mathbf{a}_1^T) = \mathbf{P}(\mathbf{a}_1^T)\mathbf{1} = p(\mathbf{a}_1^T \mid \boldsymbol{s}_0) \tag{7.2.6}$$

can be evaluated by the following backward algorithm

$$\boldsymbol{\beta}(\mathbf{a}_T^{T+1}) = \mathbf{1}, \quad \boldsymbol{\beta}(\mathbf{a}_k^T) = \mathbf{P}(\mathbf{a}_k)\boldsymbol{\beta}(\mathbf{a}_{k+1}^T) \tag{7.2.7a}$$

for $k=m,T-1,\ldots,1$. The scalar form of this equation is

$$\beta(\boldsymbol{s}_{k-1},\mathbf{a}_k^T) = \int_S p(\boldsymbol{s}_k,\mathbf{a}_k \mid \boldsymbol{s}_{k-1})\beta(\boldsymbol{s}_k,\mathbf{a}_{k+1}^T)\,d\boldsymbol{s}_k \tag{7.2.7b}$$

Combining both algorithms we obtain

$$p(\mathbf{a}_1^T) = \boldsymbol{\alpha}(\mathbf{a}_1^k)\cdot\boldsymbol{\beta}(\mathbf{a}_{k+1}^T) = \int_S \alpha(\boldsymbol{s}_k,\mathbf{a}_1^k)\beta(\boldsymbol{s}_k,\mathbf{a}_{k+1}^T)\,d\boldsymbol{s}_k$$

As we can see, Eq. (7.2.3) formally coincides with Eq. (2.1.2) which formed the basis of the matrix probability theory. Thus, we can extend this theory to the *operator probability theory* by defining the operator probability of an event Ξ corresponding to a subset of the observation sequences $\mathbf{a}_1^T$ as

$$\mathbf{P}(\Xi) = \int_{\Xi} \mathbf{P}(\mathbf{a}_1^T)\,d\mathbf{a}_1^T$$

The (scalar) probability of this event is given by

$$Pr(\Xi) = \mathbf{p}_0\mathbf{P}(\Xi)\mathbf{1} \tag{7.2.8}$$

Example 7.1. Let the transition PDF have a form

$$p(s_t,a_t \mid s_{t-1}) = \mathrm{N}(s_t - fs_{t-1},\sigma^2)\mathrm{N}(a_t - hs_t,\rho^2)$$

Find the conditional PDF $p(a_t \mid s_{t-1})$.

This conditional PDF is obtained by the marginalization of $p(s_t,a_t \mid s_{t-1})$:

$$p(a_t \mid s_{t-1}) = \int_{-\infty}^{\infty} \mathrm{N}(s_t - fs_{t-1},\sigma^2)\mathrm{N}(a_t - hs_t,\rho^2)\,ds_t$$

The integral in this equation can be evaluated using Eq. (A.7.22) of Appendix 7 which brings us

$$p(a_t \mid s_{t-1}) = \mathrm{N}(a_t - hfs_{t-1},\sigma_1^2) = \frac{\exp[-(a_t - hfs_{t-1})^2/2\sigma_1^2]}{\sqrt{2\pi}\,\sigma_1}$$

where $\sigma_1^2 = \rho^2 + h^2\sigma^2$.

Equations (7.2.2) and (7.2.8) formally coincide with Eq. (2.1.1) and (2.1.5). Therefore, all the results of the matrix probability theory developed in Chapter 2 can be formally extended to the operator probabilities. However, this formalism is not very useful unless we show how to evaluate effectively these operators probabilities.

7.3 FILTERING, PREDICTION, AND SMOOTHING

In many applications (see, for example, Sec. 4.3) it is necessary to compute the conditional PDF $p(\boldsymbol{s}_t \mid \mathbf{a}_1^T)$ of the state $\boldsymbol{s}_t$ given the observation sequence $\mathbf{a}_1^T$. In some applications (such as decoding and channel equalization) it is sufficient to find the most probable state

$$\hat{\boldsymbol{s}}_t = \underset{\boldsymbol{s}_t}{arg\,max}\, p(\boldsymbol{s}_t \mid \mathbf{a}_1^T) \tag{7.3.1}$$

If $T = t$, the process of finding this state is called *filtering*, if $T < t$, it is called *prediction*, and, if $T > t$, it is called *smoothing*.

The filtering PDF is called the normalized forward PDF and can be obtained using the forward algorithm since

$$p(\boldsymbol{s}_t \mid \mathbf{a}_1^t) = \bar{\alpha}(\boldsymbol{s}_t, \mathbf{a}_1^t) = \alpha(\boldsymbol{s}_t, \mathbf{a}_1^T)/p(\mathbf{a}_1^t)$$

where

$$p(\mathbf{a}_1^t) = \boldsymbol{\alpha}(\mathbf{a}_1^t)\mathbf{1} = \int_S \alpha(\boldsymbol{s}_t, \mathbf{a}_1^T) d\boldsymbol{s}_t$$

The prediction PDF can be also obtained using the modified forward algorithm which consists of the forward algorithm (7.2.5) for the first T steps ($k \leq T$) and continues with $p(\boldsymbol{s}_k, \mathbf{a}_k \mid \boldsymbol{s}_{k-1}) = p(\boldsymbol{s}_k \mid \boldsymbol{s}_{k-1})$ for $k > m$. The prediction PDF can be written in the following operator form

$$p(\boldsymbol{s}_t \mid \mathbf{a}_1^T) = \bar{\boldsymbol{\alpha}}(\mathbf{a}_1^T)\mathbf{P}^{t-T}$$

In the fixed-lag prediction $t - T$ is constant and operator $\mathbf{P}^{t-T}$ can be precomputed.

The smoothing PDF of state $\boldsymbol{s}_t$ is usually denoted as $\gamma_t(\boldsymbol{s}_t, \mathbf{a}_1^T) = p(\boldsymbol{s}_t \mid \mathbf{a}_1^T)$ and can be expressed using the Bayes' theorem as

$$\gamma_t(\boldsymbol{s}_t, \mathbf{a}_1^T) = p(\boldsymbol{s}_t, \mathbf{a}_1^t) p(\mathbf{a}_{t+1}^T \mid \boldsymbol{s}_t)/p(\mathbf{a}_1^T)$$

or, using the notations of the previous section, as

$$\gamma_t(\boldsymbol{s}_t, \mathbf{a}_1^T) = \alpha(\boldsymbol{s}_t, \mathbf{a}_1^t)\beta(\boldsymbol{s}_t, \mathbf{a}_{t+1}^T)/p(\mathbf{a}_1^T) \tag{7.3.2}$$

Thus, the smoothing PDF can be obtained using both the forward and backward algorithms.

The *sliding window* smoothing PDF is defined as $\gamma_{t-k}(\boldsymbol{s}_{t-k}^t, \mathbf{a}_1^T) = p(\boldsymbol{s}_{t-k}^t \mid \mathbf{a}_1^T)$ where k is the size of the window. Using the Markovian property, we can write

$$\gamma_{t-k}(\boldsymbol{s}_{t-k}^t, \mathbf{a}_1^T) = p(\boldsymbol{s}_{t-k}, \mathbf{a}_1^{t-k}) p(\boldsymbol{s}_{t-k+1}^t, \mathbf{a}_{t-k+1}^t \mid \boldsymbol{s}_{t-k}) p(\mathbf{a}_{t+1}^T \mid \boldsymbol{s}_t)/p(\mathbf{a}_1^T)$$

This equation can be rewritten using the forward and backward PDFs as

$$\gamma_{t-k}(\boldsymbol{s}_{t-k}^t, \mathbf{a}_1^T) = \alpha(\boldsymbol{s}_{t-k}, \mathbf{a}_1^{t-k}) \prod_{i=t-k+1}^{t} p(\boldsymbol{s}_i, \mathbf{a}_i \mid \boldsymbol{s}_{i-1}) \beta(\boldsymbol{s}_t, \mathbf{a}_{t+1}^T)/p(\mathbf{a}_1^T)$$

If we replace k in this equation by $k - 1$, we obtain

$$\gamma_{t-k+1}(\boldsymbol{s}_{t-k+1}^t, \mathbf{a}_1^T) = \alpha(\boldsymbol{s}_{t-k+1}, \mathbf{a}_1^{t-k+1}) \prod_{i=t-k+2}^{t} p(\boldsymbol{s}_i, \mathbf{a}_i \mid \boldsymbol{s}_{i-1}) \beta(\boldsymbol{s}_t, \mathbf{a}_{t+1}^T)/p(\mathbf{a}_1^T)$$

Comparing these equations, we obtain the following relation between $\gamma_{t-k}(\boldsymbol{s}_{t-k}^t, \mathbf{a}_1^T)$ and $\gamma_{t-k+1}(\boldsymbol{s}_{t-k+1}^t, \mathbf{a}_1^T)$

$$\gamma_{t-k}(\boldsymbol{s}_{t-k}^t, \mathbf{a}_1^T) = \gamma_{t-k+1}(\boldsymbol{s}_{t-k+1}^t, \mathbf{a}_1^T) \frac{\alpha(\boldsymbol{s}_{t-k}, \mathbf{a}_1^{t-k}) p(\boldsymbol{s}_{t-k+1}, \mathbf{a}_{t-k+1} \mid \boldsymbol{s}_{t-k})}{\alpha(\boldsymbol{s}_{t-k+1}, \mathbf{a}_1^{t-k+1})}$$

This equation allows us to compute the smoothing PDF for the larger window using that of the smaller one and only the filtering PDFs.

$$\gamma_{t-k}(\boldsymbol{s}_{t-k}^{t-1}, \mathbf{a}_1^T) = \gamma_{t-k+1}(\boldsymbol{s}_{t-k+1}^t, \mathbf{a}_1^T) \frac{\alpha(\boldsymbol{s}_{t-k}, \mathbf{a}_1^{t-k}) p(\boldsymbol{s}_{t-k+1}, \mathbf{a}_{t-k+1} \mid \boldsymbol{s}_{t-k})}{\alpha(\boldsymbol{s}_{t-k+1}, \mathbf{a}_1^{t-k+1})}$$

In particular, for $k = 1$ we have

$$\gamma_{t-1}(\boldsymbol{s}_{t-1}, \boldsymbol{s}_t, \mathbf{a}_1^T) = \gamma_t(\boldsymbol{s}_t, \mathbf{a}_1^T) \alpha(\boldsymbol{s}_{t-1}, \mathbf{a}_1^{t-1}) p(\boldsymbol{s}_t, \mathbf{a}_t \mid \boldsymbol{s}_{t-1})/\alpha(\boldsymbol{s}_t, \mathbf{a}_1^t) \tag{7.3.3}$$

By integrating both sides of the previous equation with respect to $\boldsymbol{s}_t$ (or, in other words, by marginalizing the conditional PDF) we obtain

$$\gamma_{t-1}(\boldsymbol{s}_{t-1},\mathbf{a}_1^T) = \alpha(\boldsymbol{s}_{t-1},\mathbf{a}_1^{t-1})\int_S \gamma_t(\boldsymbol{s}_t,\mathbf{a}_1^T)p(\boldsymbol{s}_t,\mathbf{a}_t \mid \boldsymbol{s}_{t-1})/\alpha(\boldsymbol{s}_t,\mathbf{a}_1^t)d\boldsymbol{s}_t$$

This equation allows us to compute the smoothing PDFs using the following forward-backward algorithm:

Algorithm 7.1.

`Initialize` (*Forward part*)

$$\alpha(\boldsymbol{s}_0,\mathbf{a}_0) = p_0(\boldsymbol{s}_0) \tag{7.3.4a}$$

`For` $t=0,1,...,T-1$
`Begin`
`Compute and save`

$$\alpha(\boldsymbol{s}_{t+1},\mathbf{a}_1^{t+1}) = \int_S p(\boldsymbol{s}_{t+1},\mathbf{a}_{t+1} \mid \boldsymbol{s}_t)\alpha(\boldsymbol{s}_t,\mathbf{a}_1^t)d\boldsymbol{s}_t \tag{7.3.4b}$$

`End`
`Initialize:` (*Backward Part*)

$$\gamma_T(\boldsymbol{s}_T,\mathbf{a}_1^T) = \overline{\alpha}(\boldsymbol{s}_T,\mathbf{a}_T) = \alpha(\boldsymbol{s}_T,\mathbf{a}_T)/\int_S \alpha(\boldsymbol{s}_T,\mathbf{a}_T)d\boldsymbol{s}_T \tag{7.3.4c}$$

`For` $t=T-1,T-2,...,1$
`Begin`
`Compute`

$$\gamma_{t-1}(\boldsymbol{s}_{t-1},\mathbf{a}_1^T) = \alpha(\boldsymbol{s}_{t-1},\mathbf{a}_1^{t-1})\int_S \gamma_t(\boldsymbol{s}_t,\mathbf{a}_1^T)p(\boldsymbol{s}_t,\mathbf{a}_t \mid \boldsymbol{s}_{t-1})/\alpha(\boldsymbol{s}_t,\mathbf{a}_1^t)d\boldsymbol{s}_t \tag{7.3.4d}$$

`End`

The *fixed – interval state* smoothing is defined by the solution of Eq. (7.3.1) for $t=1,2,...,T$ and can be achieved by adding the maximization step to the backward part of the previous algorithm. In the case of finite discrete-state HMM, the forward-backward algorithm required to save all the forward probabilities for $t=1,2,...,T$. However, in the case of continuous-state HMM it is, generally, impossible to save the forward PDFs. Nevertheless, if these PDFs depend on the finite number of parameters, it is possible to save the parameters (which is equivalent to saving the PDFs) and, thus, apply the forward-backward algorithm in the parametric form. We will use this approach for the HMM with the Gaussian transition PDF in which case the recursive algorithm for computing the parameters of the forward PDFs is called the Kalman filter while the backward part is the Rauch-Tung-Striebel (RTS) smoother. [14]

Note that thus defined smoothing might deliver a sequence of states $\hat{\boldsymbol{s}}_1^T$ that is not possible for a particular HMM while the *fixed – interval state – sequence* smoothing which is defined as the solution of the following equation

$$\hat{\boldsymbol{s}}_1^T = arg\max_{\boldsymbol{s}_1^T} p(\boldsymbol{s}_1^T \mid \mathbf{a}_1^T) = arg\max_{\boldsymbol{s}_1^T} p(\mathbf{x}_1^T) \tag{7.3.5}$$

does not have this problem. Because of the Markovian property, presented by Eq. (7.1.2),

this equation can be solved using the Viterbi algorithm (see Sec. 4.3.1).

In the forward path, the algorithm is initialized as $\psi_0(\boldsymbol{s}_0) = p_0(\boldsymbol{s}_0)$ and then computes recursively

$$\begin{aligned} \psi_t(\boldsymbol{s}_t) &= \max_{\boldsymbol{s}_{t-1}} p(\boldsymbol{s}_t, \mathbf{a}_t \mid \boldsymbol{s}_{t-1}) \psi_{t-1}(\boldsymbol{s}_{t-1}) \\ \eta_{t-1}(\boldsymbol{s}_t) &= \arg\max_{\boldsymbol{s}_{t-1}} p(\boldsymbol{s}_t, \mathbf{a}_t \mid \boldsymbol{s}_{t-1}) \psi_{t-1}(\boldsymbol{s}_{t-1}), \quad t = 1,2,...,T \end{aligned} \tag{7.3.6}$$

Then it finds $\hat{\boldsymbol{s}}_T = \arg\max_{\boldsymbol{s}} \psi_T(\boldsymbol{s}_T)$ and performs backtracking:

$$\hat{\boldsymbol{s}}_{t-1} = \eta_{t-1}(\hat{\boldsymbol{s}}_t), \quad t = T, T-1,...,2 \tag{7.3.7}$$

As with the forward-backward algorithm, the main problem with this approach is that it is impossible to remember the functions $\psi_t(s)$ and $\eta_t(s)$ unless they depend on the finite number of parameters. Note, however, that instead of evaluating and remembering $\eta_t(s)$ and backtracking, we can can use the following backward algorithm:

$$\hat{\boldsymbol{s}}_{t-1} = \arg\max_{\boldsymbol{s}_{t-1}} p(\hat{\boldsymbol{s}}_t, \mathbf{a}_t \mid \boldsymbol{s}_{t-1}) \psi_{t-1}(\boldsymbol{s}_{t-1}) \tag{7.3.8}$$

In general, Eq. (7.3.1) and (7.3.5) have different solutions. However, if the PDF $p(\boldsymbol{s}_1^T \mid \mathbf{a}_1^T) = \mathrm{N}(\boldsymbol{s}_1^T - \boldsymbol{\mu}_1^T, \boldsymbol{\Sigma}_T)$ is Gaussian, then the solutions of these equations are identical. Indeed, any Gaussian PDF (see Appendix 7) achieves its global maximum at its mean. Therefore, the solution of Eq. (7.3.5) is $E\{\boldsymbol{s}_1^T \mid \mathbf{a}_1^T\} = \boldsymbol{\mu}_1^T$ while the solution of Eq. (7.3.1) is $E\{\boldsymbol{s}_t \mid \mathbf{a}_1^T\} = \boldsymbol{\mu}_t$. In other words, the solutions of (7.3.1) delivered on the basis of the forward-backward algorithm and solutions of (7.3.5) delivered by the Viterbi algorithm are identical for the HMM with the Gaussian PDFs. These HMMs are considered in the next section.

7.4 LINEAR SYSTEMS

Any discrete time HMM can be represented by a nonlinear system of equations. Indeed, let $\mathbf{x}_t = (\boldsymbol{s}_t, \mathbf{a}_t)$ be an HMM with the transition PDF $p(\mathbf{x}_t \mid \boldsymbol{s}_{t-1})$. Denote as $F_t(\mathbf{X}_t \mid \boldsymbol{s}_{t-1}) = Pr(\mathbf{x}_t < \mathbf{X}_t \mid \boldsymbol{s}_{t-1})$ its cumulative probability distribution function and $F_t^{-1}(\mathbf{u}_t \mid \boldsymbol{s}_{t-1}) = inf\{\mathbf{x}_t : F_t(\mathbf{x}_t \mid \boldsymbol{s}_{t-1}) \geq \mathbf{u}_t\}$ its inverse function. Then it is easy to show that the random variable $\mathbf{x}_t$ can be simulated by using the inverse function method: [15]

$$\mathbf{x}_t = F_t^{-1}(\mathbf{u}_t \mid \boldsymbol{s}_{t-1}) \tag{7.4.1}$$

where $\mathbf{u}_t$ is a random vector whose components are independent random variables and are uniformly distributed in the interval [0,1].

The inverse statement is also true: A stochastic process defined by Eq. (7.4.1), in which $\mathbf{u}_t$ are independent vectors, is an HMM. Thus, we can use stochastic recursive equations to describe HMMs. A special class of these HMMs are represented by the linear stochastic equations:

$$\begin{aligned} \boldsymbol{s}_{t+1} &= \mathbf{F}_t \boldsymbol{s}_t + \mathbf{G}_t \boldsymbol{w}_t \\ \mathbf{a}_{t+1} &= \mathbf{H}_{1,t} \boldsymbol{s}_{t+1} + \mathbf{H}_{2,t} \boldsymbol{s}_t + \mathbf{M}_t \boldsymbol{v}_t, \quad t = 0,1,... \end{aligned} \tag{7.4.2}$$

where $(\mathbf{w}_t, \mathbf{v}_t)$ are independent random variables. If we substitute $\boldsymbol{s}_{t+1}$ from the first equation of (7.4.2) into its second equation, then (7.4.2) can be rewritten as

$$\begin{aligned} \boldsymbol{s}_{t+1} &= \mathbf{F}_t \boldsymbol{s}_t + \mathbf{u}_{s,t} \\ \mathbf{a}_{t+1} &= \mathbf{H}_t \boldsymbol{s}_t + \mathbf{u}_{a,t} \end{aligned}$$

where $\mathbf{H}_t = \mathbf{H}_{1,t}\mathbf{F}_t + \mathbf{H}_{2,t}$ and

$$\mathbf{u}_{s,t} = \mathbf{G}_t \mathbf{w}_t, \quad \mathbf{u}_{a,t} = \mathbf{M}_t \boldsymbol{v}_t + \mathbf{H}_{1,t}\mathbf{G}_t \boldsymbol{w}_t$$

In other words, a general linear system describing an HMM can be written in terms of its complete states as

$$\mathbf{x}_{t+1} = \mathbf{C}_t \boldsymbol{s}_t + \mathbf{u}_t, \quad t = 0,1,... \tag{7.4.3}$$

where $\mathbf{C}_t' = [\mathbf{F}_t', \mathbf{H}_t']$ and $\mathbf{u}_t$ are independent variables.

If we further assume that $\mathbf{u}_t$ are independent Gaussian variables and the initial state $\boldsymbol{s}_0$ is also Gaussian or fixed, then the complete sequence $\mathbf{x}_t$ is a special case of the Gauss-Markov process (see Appendix 7). We call this process a hidden Gauss-Markov model (HGMM). In order to apply the results developed for the Gauss-Markov processes to HGMMs, we rewrite equation (7.4.3) in the standard form presented by Eq. (A.7.38) of Appendix 7 †

$$\mathbf{x}_{t+1} = \mathbf{A}_t \mathbf{x}_t + \mathbf{u}_t \tag{7.4.4}$$

where we define $\mathbf{x}_0' = (\boldsymbol{s}_0, 0)$ and

$$\mathbf{A}_t = \begin{bmatrix} \mathbf{F}_t & 0 \\ \mathbf{H}_t & 0 \end{bmatrix} = \begin{bmatrix} \mathbf{C}_t & 0 \end{bmatrix} \tag{7.4.5}$$

We assume that $\mathbf{u}_t \sim N(0, \mathbf{S}_t)$ is zero-mean Gaussian variable whose variance matrix has the following block form corresponding to the hidden states $\boldsymbol{s}_t$ and observations $\mathbf{a}_t$

$$\mathbf{S}_t = \begin{bmatrix} \mathbf{S}_{ss,t} & \mathbf{S}_{sa,t} \\ \mathbf{S}_{as,t} & \mathbf{S}_{aa,t} \end{bmatrix}$$

As we pointed out in Sec. 7.1, any partially observable Markov process can be presented as an HMM by increasing the number of hidden states. In particular, suppose that the process is described by Eq. (7.4.4) where the matrix $\mathbf{A}_t$ has the form

$$\mathbf{A}_t = \begin{bmatrix} \mathbf{A}_{ss,t} & \mathbf{A}_{sa,t} \\ \mathbf{A}_{as,t} & \mathbf{A}_{aa,t} \end{bmatrix}$$

---

† Typically, the linear system is presented in the state space form as

$$\begin{aligned} \boldsymbol{s}_{t+1} &= \mathbf{F}_t \boldsymbol{s}_t + \mathbf{G}_t \boldsymbol{w}_t \\ \mathbf{a}_t &= \mathbf{H}_t \boldsymbol{s}_t + \mathbf{M}_t \boldsymbol{v}_t, \quad t = 0,1,2,... \end{aligned}$$

which assumes that there is an observation $\mathbf{a}_0$ in the original state while Eq. (7.4.2) assumes that there is no observations at $t = 0$. As we have seen in the previous chapters, the latter assumption is more convenient allowing us to develop the matrix probability theory. Moreover, this concept leads to a single Eq. (7.4.3) representing HGMM as a special case of Gauss-Markov processes. Nevertheless, it is easy to see that by the simple substitution $\mathbf{b}_{t+1} = \mathbf{a}_t$ we convert the system presented in this footnote into (7.4.2) and we obtain equation (7.4.3) with $\mathbf{x}_t = (\boldsymbol{s}_t, \mathbf{a}_{t-1})$

and the states $\boldsymbol{s}_t$ are not observable. To present this process as an HMM, we treat $\mathbf{x}_t$ as not observable and introduce a new process $\mathbf{y}_t = (\mathbf{x}_t, \mathbf{a}_t)$. This new process is described by the equations

$$\mathbf{y}_{t+1} = \mathbf{A}_t^{(1)} \mathbf{x}_t + \mathbf{u}_t^{(1)} \tag{7.4.6}$$

where

$$\mathbf{A}_t^{(1)} = \begin{bmatrix} \mathbf{A}_{ss,t} & \mathbf{A}_{sa,t} \\ \mathbf{A}_{as,t} & \mathbf{A}_{aa,t} \\ \mathbf{A}_{as,t} & \mathbf{A}_{aa,t} \end{bmatrix}$$

which has the form of Eq. (7.4.3). We call this method of constructing the HMM a *trivial augmentation* of the hidden states.

According to Eq. (7.4.3) the one-step transition PDF has the form

$$p(\mathbf{x}_t \mid \boldsymbol{s}_{t-1}) = \mathsf{N}(\mathbf{x}_t - \mathbf{C}_{t-1}\boldsymbol{s}_{t-1}, \mathbf{S}_{t-1}) \tag{7.4.7}$$

This PDF can be factored as $p(\mathbf{x}_t \mid \boldsymbol{s}_{t-1}) = p(\mathbf{a}_t \mid \boldsymbol{s}_{t-1}) p(\boldsymbol{s}_t \mid \mathbf{a}_t, \boldsymbol{s}_{t-1})$ which (according to Eq. (A.7.26) of Appendix 7) can be written as

$$p(\mathbf{x}_t \mid \boldsymbol{s}_{t-1}) = \mathsf{N}(\mathbf{a}_t - \mathbf{H}_{t-1}\boldsymbol{s}_{t-1}, \mathbf{S}_{aa,t-1}) \mathsf{N}(\boldsymbol{s}_t - \boldsymbol{\mu}_{t,t-1}, \boldsymbol{\Sigma}_{t,t-1}) \tag{7.4.8}$$

where the conditional mean $\boldsymbol{\mu}_{t,t-1} = E\{\boldsymbol{s}_t \mid \mathbf{a}_t, \boldsymbol{s}_{t-1}\}$ has the form

$$\boldsymbol{\mu}_{t,t-1} = \mathbf{F}_{t-1}\boldsymbol{s}_{t-1} + \mathbf{S}_{sa,t-1}\mathbf{S}_{aa,t-1}^{-1}(\mathbf{a}_t - \mathbf{H}_{t-1}\boldsymbol{s}_{t-1}) = \mathbf{D}_{t-1}\boldsymbol{s}_{t-1} + \mathbf{K}_{t-1}\mathbf{a}_t$$

and conditional variance matrix

$$\boldsymbol{\Sigma}_{t,t-1} = \mathbf{S}_{ss,t-1} - \mathbf{S}_{sa,t-1}\mathbf{S}_{aa,t-1}^{-1}\mathbf{S}_{as,t-1} = \mathbf{S}_{ss,t-1} - \mathbf{K}_{t-1}\mathbf{S}_{as,t-1}$$

where

$$\mathbf{D}_{t-1} = \mathbf{F}_{t-1} - \mathbf{K}_{t-1}\mathbf{H}_{t-1}, \quad \mathbf{K}_{t-1} = \mathbf{S}_{sa,t-1}\mathbf{S}_{aa,t-1}^{-1}$$

We assume here and in the sequel that the inverse matrices $\mathbf{S}_{aa,t-1}^{-1}$ exist for $t = 1, 2, \ldots$.

Example 7.2. Let

$$s_{t+1} = fs_t + w_t, \quad a_{t+1} = bs_{t+1} + v_t$$

where $w_t \sim N(0, \sigma^2)$ and $v_t \sim N(0, \rho^2)$. After the substitution of s_{t+1}, the system takes the form

$$s_{t+1} = fs_t + w_t, \quad a_{t+1} = hs_t + u_t \tag{7.4.9}$$

where $h = bf$ and $u_t = v_t + bw_t$. Thus, we can write Eq. (7.4.4) with

$$\mathbf{A} = \begin{bmatrix} f & 0 \\ h & 0 \end{bmatrix}, \quad \mathbf{x}_t = \begin{bmatrix} s_t \\ a_t \end{bmatrix}$$

$$\mathbf{S}_t = \begin{bmatrix} \sigma^2 & b\sigma^2 \\ b\sigma^2 & b^2\sigma^2 + \rho^2 \end{bmatrix} = \sigma^2 \begin{bmatrix} 1 & b \\ b & b^2 + \lambda^2 \end{bmatrix}$$

where we denoted as $\lambda = \rho/\sigma$ to simplify notation.

According to Eq. (7.4.8), the one step transition PDF takes the following form:

$$p(\mathbf{x}_t \mid \boldsymbol{s}_{t-1}) = \mathsf{N}[a_t - hs_{t-1}, \sigma^2(b^2 + \lambda^2)] \mathsf{N}(s_t - \mu_{t,t-1}, \Sigma_{t,t-1}) \tag{7.4.10}$$

where

$$\mu_{t,t-1} = fs_{t-1} + (a_t - hs_{t-1})b^2/(b^2+\lambda^2)$$
$$\Sigma_{t,t-1} = \sigma^2 - b^2\sigma^2/(b^2+\lambda^2) = \lambda^2\sigma^2/(b^2+\lambda^2)$$

We will use these equations frequently in the future examples.

7.4.1 Autocovariance Function

Since the complete sequence $\{\mathbf{x}_t\}$ is a Gauss-Markov process its autocovariance function $\mathbf{R}_x(t,T) = \mathbf{E}\{(\mathbf{x}_t - \boldsymbol{\mu}_t)(\mathbf{x}_T - \boldsymbol{\mu}_T)'\}$ has the form [see Appendix 7, Eq. (A.7.48)]:

$$\mathbf{R}_x(t,T) = \begin{cases} \prod_{i=t-1}^{T} \mathbf{A}_i \boldsymbol{\Sigma}_T & \text{for } t > T \\ \boldsymbol{\Sigma}_T & \text{for } t = T \\ \boldsymbol{\Sigma}_t \prod_{i=t}^{T-1} \mathbf{A}_i' & \text{for } t < T \end{cases} \tag{7.4.11}$$

where

$$\boldsymbol{\mu}_T = \mathbf{A}_{T-1}\boldsymbol{\mu}_{T-1} = \prod_{i=T-1}^{0} \mathbf{A}_i \boldsymbol{\mu}_0 \tag{7.4.12}$$

and $\boldsymbol{\Sigma}_t$ are found using the Riccati equations (see Appendix 7):

$$\boldsymbol{\Sigma}_t = \mathbf{A}_{t-1}\boldsymbol{\Sigma}_{t-1}\mathbf{A}_{t-1}' + \mathbf{S}_{t-1} \tag{7.4.13}$$

the initial conditions are $\mu_0 = (\boldsymbol{s}_0, 0)$, $\boldsymbol{\Sigma}_0 = 0$ if the initial state is fixed. If the initial condition is Gaussian $\boldsymbol{s}_0 \sim N(\boldsymbol{\nu}_{s,0}, \boldsymbol{\Sigma}_{ss,0})$, then $\boldsymbol{\mu}_0 = (\boldsymbol{\nu}_{s,0}, 0)$ and

$$\boldsymbol{\Sigma}_0 = \begin{bmatrix} \boldsymbol{\Sigma}_{ss,0} & 0 \\ 0 & 0 \end{bmatrix}$$

The autocovariance function of the hidden states $\mathbf{R}_{ss}(t,T)$, of observations $\mathbf{R}_{aa}(t,T)$ and their cross-covariance functions $\mathbf{R}_{sa}(t,T)$ represent the subblocks of

$$\mathbf{R}_x(t,T) = \begin{bmatrix} \mathbf{R}_{ss}(t,T) & \mathbf{R}_{sa}(t,T) \\ \mathbf{R}_{as}(t,T) & \mathbf{R}_{aa}(t,T) \end{bmatrix}, \quad \mathbf{R}_x(t,t) = \boldsymbol{\Sigma}_t = \begin{bmatrix} \boldsymbol{\Sigma}_{ss,t} & \boldsymbol{\Sigma}_{sa,t} \\ \boldsymbol{\Sigma}_{as,t} & \boldsymbol{\Sigma}_{aa,t} \end{bmatrix}$$

For $t \neq T$, these subblocks can be evaluated using the recursive Eq. (A.7.47) of Appendix 7 which in our case can be written in the following block form

$$\mathbf{R}_{ss}(t,T) = \mathbf{F}_{t-1}\mathbf{R}_{ss}(t-1,T) = \mathbf{R}_{ss}(t,T-1)\mathbf{F}_{t-1}'$$
$$\mathbf{R}_{as}(t,T) = \mathbf{R}'_{sa}(T,t) = \mathbf{H}_{t-1}\mathbf{R}_{ss}(t-1,T) = \mathbf{R}_{as}(t,T-1)\mathbf{F}_{t-1}'$$
$$\mathbf{R}_{aa}(t,T) = \mathbf{H}_{t-1}\mathbf{R}_{sa}(t-1,T) = \mathbf{H}_{t-1}\mathbf{R}_{ss}(t-1,T-1)\mathbf{H}_{T-1}' \tag{7.4.14}$$

For $t = T$ we need to use the Riccati Eq. (7.4.13) to evaluate the blocks of $\mathbf{R}_x(t,t) = \boldsymbol{\Sigma}_t$ which can be rewritten as

$$\boldsymbol{\Sigma}_{ss,t} = \mathbf{F}_{t-1}\boldsymbol{\Sigma}_{ss,t-1}\mathbf{F}_{t-1}' + \mathbf{S}_{ss,t-1}$$
$$\boldsymbol{\Sigma}_{sa,t} = \mathbf{F}_{t-1}\boldsymbol{\Sigma}_{ss,t-1}\mathbf{H}_{t-1}' + \mathbf{S}_{sa,t-1}, \quad \boldsymbol{\Sigma}_{aa,t} = \mathbf{H}_{t-1}\boldsymbol{\Sigma}_{ss,t-1}\mathbf{H}_{t-1}' + \mathbf{S}_{aa,t-1}$$

The corresponding means are given by

$$\begin{aligned} \mathbf{v}_{s,T} &= \mathbf{F}_{T-1}\mathbf{v}_{s,T-1} = \prod_{i=T-1}^{0} \mathbf{F}_i \mathbf{v}_{s,0} \\ \mathbf{v}_{a,T} &= \mathbf{H}_{T-1}\mathbf{v}_{s,T-1} = \mathbf{H}_T \prod_{i=T-2}^{0} \mathbf{F}_i \mathbf{v}_{s,0} \end{aligned} \tag{7.4.15}$$

Thus, all the block covariance matrices and means can be expressed through the covariance matrices and mean of the hidden states. It follows from Eq. (7.4.11) that

$$\mathbf{R}_{ss}(t,T) = \begin{cases} \prod_{i=t-1}^{T} \mathbf{F}_i \boldsymbol{\Sigma}_{ss,T} & \text{for } t>T \\ \boldsymbol{\Sigma}_{ss,t} & \text{for } t=T \\ \boldsymbol{\Sigma}_{ss,t} \prod_{i=t}^{T-1} \mathbf{F}_i' & \text{for } t<T \end{cases} \tag{7.4.16}$$

The process is homogeneous if the coefficients $\mathbf{C}_t = \mathbf{C}$ and the variance matrices $\mathbf{S}_t = \mathbf{S}$ are constant. If the process is stationary, then $\boldsymbol{\mu}_t = 0$ and $\boldsymbol{\Sigma}_t = \boldsymbol{\Sigma}_\infty$ satisfies the Lyapunov equation $\boldsymbol{\Sigma}_\infty = \mathbf{A}\boldsymbol{\Sigma}_\infty\mathbf{A}' + \mathbf{S}$ (Eq. (A.7.46) of Appendix 7). This equation is equivalent to the following equations for the matrix blocks

$$\boldsymbol{\Sigma}_{ss,\infty} = \mathbf{F}\boldsymbol{\Sigma}_{ss,\infty}\mathbf{F}' + \mathbf{S}_{ss} \tag{7.4.17}$$

$$\boldsymbol{\Sigma}_{aa,\infty} = \mathbf{H}\boldsymbol{\Sigma}_{ss,\infty}\mathbf{H}' + \mathbf{S}_{aa}, \quad \boldsymbol{\Sigma}_{sa,\infty} = \mathbf{F}\boldsymbol{\Sigma}_{ss,\infty}\mathbf{H}' + \mathbf{S}_{sa} \tag{7.4.18}$$

Thus, we need to solve only the Lyapunov Eq. (7.4.17) corresponding to the hidden states the rest of the subblocks are expressed through this solution.

For the stationary process, $\mathbf{R}_x(t,t+\tau) = \mathbf{R}_x(\tau)$ does not depend on t and Eq. (7.4.11) takes the form

$$\mathbf{R}_x(\tau) = \begin{cases} \mathbf{A}^{-\tau}\boldsymbol{\Sigma}_\infty & \text{for } \tau\le 0 \\ \boldsymbol{\Sigma}_\infty(\mathbf{A}')^{\tau} & \text{for } \tau\ge 0 \end{cases}$$

The subblocks of this function can be expressed using the solution of Eq. (7.4.17). The solution is unique if all the eigenvalues of $\mathbf{F}$ lie inside the unit circle $|z|=1$. For example, the autocovariance function of the observation sequence can be written as

$$\mathbf{R}_{aa}(0) = \boldsymbol{\Sigma}_{aa,\infty}, \quad \mathbf{R}_{aa}(\tau) = \begin{cases} \mathbf{H}\mathbf{F}^{-\tau-1}\boldsymbol{\Sigma}_{sa,\infty} & \text{for } \tau<0 \\ \boldsymbol{\Sigma}_{as,\infty}(\mathbf{F}')^{\tau-1}\mathbf{H}' & \text{for } \tau>0 \end{cases} \tag{7.4.19}$$

The power spectral density (PSD) of the complete sequence $\{\mathbf{x}_t\}$ has the form (Eq. (A.7.50) of Appendix 7):

$$\boldsymbol{\Phi}_x(z) = \sum_{\tau=-\infty}^{\infty} \mathbf{R}_x(\tau)z^{\tau} = (\mathbf{I}-\mathbf{A}z^{-1})^{-1}\mathbf{S}(\mathbf{I}-\mathbf{A}'z)^{-1}$$

The PSD of the hidden states and observations are the corresponding subblocks of this

matrix. For example, substituting **A** from Eq. (7.4.5) and multiplying the block matrices in the previous equation we obtain the observation sequence PSD as

$$\Phi_{aa}(z) = \mathbf{M}(z)\mathbf{S}_{ss}\mathbf{M}'(z^{-1})+\mathbf{M}(z)\mathbf{S}_{sa}+\mathbf{S}_{as}\mathbf{M}'(z^{-1})+\mathbf{S}_{aa} \tag{7.4.20}$$

where $\mathbf{M}(z) = \mathbf{H}z^{-1}(\mathbf{I}-\mathbf{F}z^{-1})^{-1}$. $\Sigma_{ss,\infty}$ is the coefficient of z^0 in the expansion of $\mathbf{\Phi}_{ss}(z) = (\mathbf{I}-\mathbf{F}z^{-1})^{-1}\mathbf{S}_{ss}(\mathbf{I}-\mathbf{F}'z)^{-1}$ into the power (Laurent's) series and, therefore, can be found using the residue theorem (instead of solving the Lyapunov Eq. (7.4.17)):

$$\Sigma_{ss,\infty} = \frac{1}{2\pi j}\int_{|z|=1}\frac{\Phi_{ss}(z)}{z}dz = \sum Res_{\lambda_i}\frac{\Phi_{ss}(z)}{z}$$

where the sum of residues is taken over all the eigenvalues of **F**.

Example 7.3. For the system in example 7.2, the Lyapunov equation is $\Sigma_{ss,\infty} = f^2\Sigma_{ss,\infty}+\sigma^2$. Its solution has the form $\Sigma_{ss,\infty} = \sigma^2(1-f^2)^{-1}$.

The same result can be obtained using the residue theorem. Indeed,

$$\Phi_{ss}(z) = (1-fz^{-1})^{-1}\sigma^2(1-fz)^{-1} = \frac{\sigma^2 z}{(z-f)(1-fz)}$$

and

$$\Sigma_{ss,\infty} = Res_f\frac{\sigma^2}{(z-f)(1-zf)} = \frac{\sigma^2}{1-f^2}$$

The autocorrelation function of the observation sequence is given by Eq. (7.4.19) which in our case has the form

$$R_{aa}(0) = h^2\Sigma_{ss,\infty}+\sigma^2(b^2+\lambda^2), \quad R_{aa}(\tau) = hf^{|\tau|-1}(fh\Sigma_{ss,\infty}+b\sigma^2)$$

7.4.2 Observation Sequence PDF

Since all the considered sequences are Gaussian, their multidimensional PDFs can be expressed using the corresponding means and autocovariance functions presented in the previous section. For example, $\mathbf{x}_0^T \sim \mathsf{N}$ where $\boldsymbol{\mu}_0^T$ is obtained using Eq. (7.4.12) and the blocks $\mathbf{R}_x(i,j)$ of the covariance matrix

$$\boldsymbol{D}_T = \left[\mathbf{R}_x(i,j)\right]_{T+1,T+1}. \tag{7.4.21}$$

are obtained using Eq. (7.4.11) for $i,j=0,1,\ldots,T$. If the original state $\boldsymbol{s}_0$ is fixed, we obtain the conditional PDF $p(\mathbf{x}_1^T \mid s_0)i \sim \mathsf{N}(\mathbf{x}_1^T-\boldsymbol{\mu}_1^T,\mathbf{D}_{1,T})$ by removing from $\mathbf{D}_T$ the submatrices $\mathbf{R}(0,j)$ and $\mathbf{R}(i,0)$).

Example 7.4. For the system of Example 7.2, find the PDF $p(\mathbf{x}_1^2 \mid \boldsymbol{s}_0)$.

In this case, the original state $\boldsymbol{s}_0$ is fixed. Using the matrices **A** and **S** defined in Example 7.2 and Eq. (7.4.12), we can write

$$\boldsymbol{\mu}_1^2 = [fs_0 \;\; hs_0 \;\; f^2s_0 \;\; hfs_0]'$$

Equation (7.4.21) becomes

$$D_2 = \begin{bmatrix} \mathbf{S} & \mathbf{S}\mathbf{A}' \\ \mathbf{A}\mathbf{S} & \mathbf{A}\mathbf{S}\mathbf{A}' + \mathbf{S} \end{bmatrix} = \sigma^2 \begin{bmatrix} 1 & b & f & h \\ b & b^2+\lambda^2 & bf & bh \\ f & bf & f^2+1 & fh+b \\ h & bh & fh+b & b^2+h^2+\lambda^2 \end{bmatrix}$$

Thus, $p(\mathbf{x}_1^2 \mid s_0) = \mathrm{N}(\mathbf{x}_1^2 - \boldsymbol{\mu}_1^2, \mathbf{D}_2)$.

The PDF of the observation sequence $\mathbf{a}_1^T$ is Gaussian with the mean $\mathbf{v}_{a,1}^T$ and covariance matrix given by

$$\mathbf{v}_{a,1}^T = \left[\mathbf{v}_{a,i}\right]_{1,T}, \quad \boldsymbol{\Lambda}_T = \left[\mathbf{R}_{aa}(i,j)\right]_{T,T}$$

where $\mathbf{v}_{a,t}$ is given by Eq. (7.4.15) and the blocks $\mathbf{R}_{aa}(i,j)$ of $\boldsymbol{\Lambda}_T$ are evaluated using Eq. (7.4.14).

Note that if the initial state s_0 is fixed, the observation sequence PDF represent the backward PDF $\boldsymbol{\beta}(\mathbf{a}_1^T) = p(\mathbf{a}_1^T \mid s_0)$ defined by equation (7.2.6).

The forward PDF $\boldsymbol{\alpha}(\mathbf{a}_1^T) = p(\mathbf{a}_1^T, \boldsymbol{s}_T)$ defined by Eq. (7.2.4) is a Gaussian PDF whose mean $\boldsymbol{\xi}_T = E\{\mathbf{a}_1^T, \boldsymbol{s}_T\}$ is obtained using Eq. (7.4.15) as

$$\boldsymbol{\xi}_T = (\mathbf{v}_{a,1}^T, \mathbf{v}_{s,T}) \tag{7.4.22}$$

and the covariance matrix $\boldsymbol{\Xi}_T$ is obtained by adding to $\boldsymbol{\Lambda}_T$ the rows and columns corresponding to $\boldsymbol{s}_T$:

$$\boldsymbol{\Xi}_T = \begin{bmatrix} \boldsymbol{\Lambda}_T & \mathbf{R}(\mathbf{a}_1^T, \boldsymbol{s}_T) \\ \mathbf{R}(\boldsymbol{s}_T, \mathbf{a}_1^T) & \mathbf{R}_{ss}(T,T) \end{bmatrix} \tag{7.4.23}$$

where

$$\mathbf{R}(\boldsymbol{s}_T, \mathbf{a}_1^T) = [\mathbf{R}_{sa}(T,1) \quad \mathbf{R}_{sa}(T,2) \quad \ldots \quad \mathbf{R}_{sa}(T,T)]$$

and $\mathbf{R}(\mathbf{a}_1^T, \boldsymbol{s}_T) = \mathbf{R}'(\boldsymbol{s}_T, \mathbf{a}_1^T)$.

Example 7.5. Let us find the vector of means and the covariance matrix for the PDF $p(s_2, \mathbf{a}_1, \mathbf{a}_2 \mid s_0)$ using the results of Example 7.4.

According to Eq. (7.4.22) and (7.4.23), we have

$$\boldsymbol{\xi}_2 = \begin{bmatrix} hs_0 \\ hfs_0 \\ f^2 s_0 \end{bmatrix}, \quad \boldsymbol{\Xi}_2 = \sigma^2 \begin{bmatrix} b^2+\lambda^2 & bh & bf \\ bh & b^2+h^2+\lambda^2 & fh+b \\ bf & fh+b & f^2+1 \end{bmatrix}$$

To obtain the mean $\mathbf{v}_{a,1}^2$ and covariance matrix $\boldsymbol{\Lambda}_2$ corresponding to $p(\mathbf{a}_1, \mathbf{a}_2 \mid s_0) \sim N(\mathbf{v}_{a,1}^2, \boldsymbol{\Lambda}_2)$ we need to remove the entries corresponding to s_2 (that is, the last row and column):

$$\mathbf{v}_{a,1}^2 = \begin{bmatrix} hs_0 \\ hfs_0 \end{bmatrix}, \quad \boldsymbol{\Lambda}_2 = \sigma^2 \begin{bmatrix} b^2+\lambda^2 & bh \\ bh & b^2+h^2+\lambda^2 \end{bmatrix}$$

7.4.3 Kalman Filter

Equations (7.4.22) and (7.4.23) allow us to evaluate the forward PDFs $\boldsymbol{\alpha}(\boldsymbol{s}_t, \mathbf{a}_1^t)$. The sizes of the covariance matrices in these equations might be large which makes the direct inversion of $\boldsymbol{\Xi}_T$ a difficult numerical problem. It turns out that the forward algorithm in Eq. (7.2.5) allows us to evaluate this PDF recursively. In general, there is no closed-form expression for

the integrals required by this algorithm. However, for the Gauss-Markov processes, we can find the closed-form expression.

Suppose that the initial state PDF is $\mathsf{N}(\boldsymbol{s}_0 - \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$. According to Eq. (7.4.7) the transition PDF has the form

$$p(\mathbf{a}_1, \boldsymbol{s}_1 \mid \boldsymbol{s}_0) = \mathsf{N}(\mathbf{x}_1 - \mathbf{C}_0 \boldsymbol{s}_0, \mathbf{S}_0)$$

Then Eq. (7.1.5) gives us

$$\alpha(\mathbf{a}_1, \boldsymbol{s}_1) = \int_S \mathsf{N}(\mathbf{x}_1 - \mathbf{C}_0 \boldsymbol{s}_0, \mathbf{S}_0)\mathsf{N}(\boldsymbol{s}_0 - \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)\, d\boldsymbol{s}_0 \tag{7.4.24}$$

which, according to Eq. (A.7.22) of Appendix 7, is equal to

$$\alpha(\mathbf{a}_1, \boldsymbol{s}_1) = \mathsf{N}(\mathbf{x}_1 - \mathbf{C}_0 \boldsymbol{\mu}_0, \mathbf{P}_0) \tag{7.4.25}$$

where $\mathbf{P}_0 = \mathbf{S}_0 + \mathbf{C}_0 \boldsymbol{\Sigma}_0 \mathbf{C}_0'$. The next step of the forward algorithm (7.1.5) gives us

$$\alpha(\mathbf{a}_1, \mathbf{a}_2, \boldsymbol{s}_2) = \int_S \mathsf{N}(\mathbf{x}_2 - \mathbf{C}_1 \boldsymbol{s}_1, \mathbf{S}_1)\mathsf{N}(\mathbf{x}_1 - \mathbf{C}_0 \boldsymbol{\mu}_0, \mathbf{P}_0)\, d\boldsymbol{s}_1 \tag{7.4.26}$$

To evaluate this integral using Eq. (A.7.22) of Appendix 7, we factor the right hand side of the previous equation using the conditional PDF [see Eq. (A.7.26) of Appendix 7]

$$\mathsf{N}(\mathbf{x}_1 - \mathbf{C}_0 \boldsymbol{\mu}_0, \mathbf{P}_0) = p(\mathbf{a}_1)\mathsf{N}(\boldsymbol{s}_1 - \boldsymbol{s}_{1|1}, \mathbf{Q}_{1|1})$$

where

$$p(\mathbf{a}_1) = \mathsf{N}(\mathbf{a}_1 - \mathbf{H}_0 \boldsymbol{\mu}_0, \mathbf{P}_{aa,0}) \tag{7.4.27}$$

is the PDF of the observation $\mathbf{a}_1$

$$\boldsymbol{s}_{1|1} = \mathbf{E}(\boldsymbol{s}_1 \mid \mathbf{a}_1) = \mathbf{F}_0 \boldsymbol{\mu}_0 + \mathbf{P}_{sa,0}\mathbf{P}_{aa,0}^{-1}(\mathbf{a}_1 - \mathbf{H}_0 \boldsymbol{\mu}_0)$$

$$\mathbf{Q}_{1|1} = \mathbf{P}_{ss,0} - \mathbf{P}_{sa,0}\mathbf{P}_{aa,0}^{-1}\mathbf{P}_{as,0}$$

In these equations, we use the standard notations $\boldsymbol{s}_{i|j} = \mathbf{E}\{\boldsymbol{s}_i \mid \mathbf{a}_1^j\}$ and for the state $\boldsymbol{s}_i$ conditional mean and $\mathbf{Q}_{i|j} = \mathbf{E}\{(\boldsymbol{s}_i - \boldsymbol{s}_{i|j})(\boldsymbol{s}_i - \boldsymbol{s}_{i|j})' \mid \mathbf{a}_1^j\}$ for the variance matrix given $\mathbf{a}_1^j$.

Now we can rewrite Eq. (7.4.26) as

$$\alpha(\mathbf{a}_1, \mathbf{a}_2, \boldsymbol{s}_2) = p(\mathbf{a}_1) \int_S \mathsf{N}(\mathbf{x}_2 - \mathbf{C}_1 \boldsymbol{s}_1, \mathbf{S}_1)\mathsf{N}(\boldsymbol{s}_1 - \boldsymbol{s}_{1|1}, \mathbf{Q}_{1|1})\, d\boldsymbol{s}_1 \tag{7.4.28}$$

Using this representation and formula (A.7.22) of Appendix 7, we compute the integral and obtain

$$\alpha(\mathbf{a}_1, \mathbf{a}_2, \boldsymbol{s}_2) = \mathsf{N}(\mathbf{x}_2 - \mathbf{C}_1 \boldsymbol{s}_{1|1}, \mathbf{P}_1)p(\mathbf{a}_1) \tag{7.4.29}$$

where $\mathbf{P}_1 = \mathbf{C}_1 \mathbf{Q}_{1|1} \mathbf{C}_1' + \mathbf{S}_1$,

Applying Eq. (A.7.26) of Appendix 7, we can present Eq. (7.4.29) as

$$\alpha(\mathbf{a}_1^2, \boldsymbol{s}_2) = \mathsf{N}(\boldsymbol{s}_2 - \boldsymbol{s}_{2|2}, \mathbf{Q}_{2|2})p(\mathbf{a}_1^2)$$

where

$$p(\mathbf{a}_1^2) = \mathsf{N}(\mathbf{a}_2 - \mathbf{H}_1 \boldsymbol{s}_{1|1}, \mathbf{P}_{aa,1})p(\mathbf{a}_1) \tag{7.4.30}$$

$$\boldsymbol{s}_{2|2} = \mathbf{F}_1\boldsymbol{s}_{1|1}+\mathbf{P}_{sa,1}\mathbf{P}_{aa,1}^{-1}(\mathbf{a}_2-\mathbf{H}_1\boldsymbol{s}_{1|1}) = \mathbf{F}_1\boldsymbol{s}_{1|1}+\mathbf{K}_1(\mathbf{a}_2-\mathbf{H}_1\boldsymbol{s}_{1|1})$$
$$\mathbf{Q}_{2|2} = \mathbf{P}_{ss,1}-\mathbf{P}_{sa,1}\mathbf{P}_{aa,1}^{-1}\mathbf{P}_{as,1} = \mathbf{P}_{ss,1}-\mathbf{K}_1\mathbf{P}_{as,1}$$

We can assume now that the forward PDF has the form

$$\alpha(\mathbf{a}_1^t,\boldsymbol{s}_t) = \mathrm{N}(\boldsymbol{s}_t-\boldsymbol{s}_{t|t},\mathbf{Q}_{t|t})p(\mathbf{a}_1^t)$$

and prove it using the method of mathematical induction by computing $\alpha(\mathbf{a}_1^{t+1},\boldsymbol{s}_{t+1})$.

According to forward algorithm (7.1.5) we have

$$\alpha(\mathbf{a}_1^{t+1},\boldsymbol{s}_{t+1}) = p(\mathbf{a}_t)\int_S \mathrm{N}(\mathbf{x}_{t+1}-\mathbf{C}_t\boldsymbol{s}_t,\mathbf{S}_t)\mathrm{N}(\boldsymbol{s}_t-\boldsymbol{s}_{t|t},\mathbf{Q}_{t|t})d\boldsymbol{s}_t$$

Similarly to computing the integral in (7.4.28) we obtain

$$\begin{aligned}\alpha(\mathbf{a}_1^{t+1},\boldsymbol{s}_{t+1}) &= \mathrm{N}(\mathbf{x}_{t+1}-\mathbf{C}_t\boldsymbol{s}_{t|t},\mathbf{P}_t)p(\mathbf{a}_1^t)\\ &= \mathrm{N}(\boldsymbol{s}_{t+1}-\boldsymbol{s}_{t+1|t+1},\mathbf{Q}_{t+1|t+1})p(\mathbf{a}_1^{t+1})\end{aligned} \tag{7.4.31}$$

where

$$\boldsymbol{s}_{t+1|t+1} = \mathbf{F}_t\boldsymbol{s}_{t|t}+\mathbf{K}_t(\mathbf{a}_{t+1}-\mathbf{H}_t\boldsymbol{s}_{t|t}), \quad \mathbf{K}_t = \mathbf{P}_{sa,t}\mathbf{P}_{aa,t}^{-1}$$
$$\mathbf{Q}_{t+1|t+1} = \mathbf{P}_{ss,t}-\mathbf{K}_t\mathbf{P}_{as,t}$$
$$\mathbf{P}_t = \mathbf{C}_t\mathbf{Q}_{t|t}\mathbf{C}_t' + \mathbf{S}_t$$
$$p(\mathbf{a}_1^{t+1}) = p(\mathbf{a}_1^t)\mathrm{N}(\mathbf{a}_{t+1} - \mathbf{H}_t\boldsymbol{s}_{t|t},\mathbf{Q}_{t|t})$$

This algorithm for computing the forward PDFs for the HGMM is called the Kalman filter [10] and the coefficient $\mathbf{K}_t = \mathbf{P}_{sa,t}\mathbf{P}_{aa,t}^{-1}$ is called the Kalman gain. The forward algorithm can be summarized as follows:

Algorithm 7.2.

Initialize:

$$\boldsymbol{s}_{0|0} = \boldsymbol{\mu}_0, \quad \mathbf{Q}_{0|0} = \boldsymbol{\Sigma}_0, \quad p(\mathbf{a}_0) = 1 \tag{7.4.32a}$$

For $t=0,1,\ldots,T-1$

Begin

$$\mathbf{P}_t = \mathbf{C}_t\mathbf{Q}_{t|t}\mathbf{C}_t' + \mathbf{S}_t \tag{7.4.32b}$$
$$p(\mathbf{a}_1^{t+1}) = p(\mathbf{a}_1^t)\mathrm{N}(\mathbf{a}_{t+1} - \mathbf{H}_t\boldsymbol{s}_{t|t},\mathbf{P}_{aa,t}) \tag{7.4.32c}$$
$$\mathbf{K}_t = \mathbf{P}_{sa,t}\mathbf{P}_{aa,t}^{-1}, \quad \boldsymbol{s}_{t+1|t+1} = \mathbf{F}_t\boldsymbol{s}_{t|t}+\mathbf{K}_t(\mathbf{a}_{t+1}-\mathbf{H}_t\boldsymbol{s}_{t|t}) \tag{7.4.32d}$$
$$\mathbf{Q}_{t+1|t+1} = \mathbf{P}_{ss,t}-\mathbf{K}_t\mathbf{P}_{as,t} \tag{7.4.32e}$$

End

A special structure of Eq. (7.4.32c) is usually explored to express matrix blocks of $\mathbf{P}_i$ more explicitly. Indeed, since $\mathbf{C}_i' = [\mathbf{F}_i' \quad \mathbf{H}_i']$, we can express (7.4.32c) as

$$\begin{aligned} \mathbf{P}_{ss,i} &= \mathbf{F}_i \mathbf{Q}_{i|i} \mathbf{F}_i' + \mathbf{S}_{ss,i} \quad \mathbf{P}_{sa,i} = \mathbf{F}_i \mathbf{Q}_{i|i} \mathbf{H}_i' + \mathbf{S}_{sa,i} \\ \mathbf{P}_{as,i} &= \mathbf{H}_i \mathbf{Q}_{i|i} \mathbf{F}_i' + \mathbf{S}_{as,i} \quad \mathbf{P}_{aa,i} = \mathbf{H}_i \mathbf{Q}_{i|i} \mathbf{H}_i' + \mathbf{S}_{aa,i} \end{aligned} \tag{7.4.33}$$

Since the matrix $\mathbf{P}_i$ is symmetric, $\mathbf{P}_{sa,i} = \mathbf{P}_{as,i}'$ so we need to compute only $\mathbf{P}_{sa,i}$ (or $\mathbf{P}_{as,i}$). Substituting Eq. (7.4.33) into Algorithm 7.2 we can rewrite it more explicitly as

Algorithm 7.3.

Initialize:

$$\boldsymbol{s}_{0|0} = \boldsymbol{\mu}_0, \quad \mathbf{Q}_{0|0} = \boldsymbol{\Sigma}_0, \quad p(\mathbf{a}_0) = 1 \tag{7.4.34a}$$

For $t = 0, 1, \ldots, T-1$

Begin

$$p(\mathbf{a}_1^{t+1}) = p(\mathbf{a}_1^t) \mathsf{N}(\mathbf{a}_{t+1} - \mathbf{H}_t \boldsymbol{s}_{t|t}, \mathbf{H}_t \mathbf{Q}_{t|t} \mathbf{H}_t' + \mathbf{S}_{aa,t}) \tag{7.4.34b}$$

$$\mathbf{K}_t = (\mathbf{F}_t \mathbf{Q}_{t|t} \mathbf{H}_t' + \mathbf{S}_{sa,t})(\mathbf{H}_t \mathbf{Q}_{t|t} \mathbf{H}_t' + \mathbf{S}_{aa,t})^{-1} \tag{7.4.34c}$$

$$\boldsymbol{s}_{t+1|t+1} = \mathbf{F}_{t-1} \boldsymbol{s}_{t|t} + \mathbf{K}_t (\mathbf{a}_{t+1} - \mathbf{H}_t \boldsymbol{s}_{t|t}) \tag{7.4.34d}$$

$$\mathbf{Q}_{t+1|t+1} = \mathbf{F}_t \mathbf{Q}_{t|t} \mathbf{F}_t' + \mathbf{S}_{ss,t} - \mathbf{K}_t (\mathbf{F}_t \mathbf{Q}_{t|t} \mathbf{H}_t' + \mathbf{S}_{sa,t})' \tag{7.4.34e}$$

End

The forward PDF is obtained at the end of the algorithm using the following formula

$$\alpha(\mathbf{a}_1^T, \boldsymbol{s}_T) = p(\mathbf{a}_1^T, \boldsymbol{s}_T) = \mathsf{N}(\boldsymbol{s}_T - \boldsymbol{s}_{T|T}, \mathbf{Q}_{T|T}) p(\mathbf{a}_1^T) \tag{7.4.35}$$

The normalized forward density

$$\bar{\alpha}(\boldsymbol{s}_T, \mathbf{a}_1^T) = p(\boldsymbol{s}_T \mid \mathbf{a}_1^T) \tag{7.4.36}$$

according to Eq. (7.4.35) can be written as

$$\bar{\alpha}(\boldsymbol{s}_T, \mathbf{a}_1^T) = \mathsf{N}(\boldsymbol{s}_T - \boldsymbol{s}_{T|T}, \mathbf{Q}_{T|T}) \tag{7.4.37}$$

Thus, the Kalman filter can be viewed as a parametric realization of the forward algorithm (7.2.5) for the HGMM. The observation sequence PDF is given by Eq. (7.4.34b).

If the initial state PDF is a mixture of Gaussian PDFs

$$p_0(\boldsymbol{s}_0) = \sum_{i=1}^{M} w_i \mathsf{N}(\boldsymbol{s}_0 - \boldsymbol{\mu}_{i,0}, \boldsymbol{\Sigma}_{i,0}) \tag{7.4.38}$$

then it is easy to see that

$$\alpha(\mathbf{a}_1^T, \boldsymbol{s}_T) = \sum_{i=1}^{M} w_i \alpha_i(\mathbf{a}_1^T, \boldsymbol{s}_T)$$

where $\alpha_i(\boldsymbol{s}_T, \mathbf{a}_1^T)$ is obtained by the Kalman filter with the initial conditions $\boldsymbol{\mu}_{i,0}$ and $\boldsymbol{\Sigma}_{i,0}$.

It follows from Eq. (7.4.32) that the covariance matrices $\mathbf{P}_t$ and $\mathbf{Q}_t$ of the Kalman filter do not depend on the observation sequence. Therefore, if we need to compute the observation PDF values for different observation sequences, these covariance matrices need to be computed and saved only one time.

On the other hand, it also follows from Eq. (7.4.32) that the state conditional means $\boldsymbol{s}_{t|t}$

are linear functions of $\boldsymbol{s}_{t-1|t-1}$ and, therefore, of the initial mean $\boldsymbol{\mu}_0$. To find the coefficients of these linear functions, we need to modify the Kalman filter in the following way.

Equation (7.4.32d) can be written as

$$\boldsymbol{s}_{t|t} = \mathbf{D}_{t-1}\boldsymbol{s}_{t-1|t-1} + \mathbf{d}_{t-1}$$

where

$$\mathbf{D}_{t-1} = \mathbf{F}_{t-1} - \mathbf{K}_{t-1}\mathbf{H}_{t-1}$$

$$\mathbf{d}_{t-1} = \mathbf{K}_{t-1}\mathbf{a}_t$$

Starting with $\boldsymbol{s}_{0|0} = \boldsymbol{\mu}_0$ and $\mathbf{P}_0 = \mathbf{S}_0$, these equations allow us to find $\boldsymbol{s}_{T|T}$ as a function of $\boldsymbol{\mu}_0$ recursively:

$$\boldsymbol{s}_{1|1} = \mathbf{D}_0\boldsymbol{\mu}_0 + \mathbf{d}_0$$

$$\boldsymbol{s}_{2|2} = \mathbf{D}_1\mathbf{D}_0\boldsymbol{\mu}_0 + \mathbf{D}_1\mathbf{d}_0 + \mathbf{d}_1$$

and so on. Finally, we obtain

$$\boldsymbol{s}_{T|T} = \mathbf{U}_T\boldsymbol{\mu}_0 + \mathbf{u}_T \tag{7.4.39}$$

where

$$\mathbf{U}_T = \mathbf{D}_{T-1}\mathbf{D}_{T-2}\cdots\mathbf{D}_0$$

and

$$\mathbf{u}_T = \mathbf{d}_{T-1} + \mathbf{D}_{T-1}\mathbf{d}_{T-2} + \ldots + \mathbf{D}_{T-1}\mathbf{D}_{T-2}\cdots\mathbf{D}_1\mathbf{d}_0$$

Substituting Eq. (7.4.39) into (7.4.35) we obtain the explicit expression for the normalized forward PDF as a function of $\boldsymbol{\mu}_0$:

$$\bar{\alpha}(\mathbf{a}_1^T, \boldsymbol{s}_T; \boldsymbol{\mu}_0) = \mathrm{N}(\boldsymbol{s}_T - \mathbf{U}_T\boldsymbol{\mu}_0 - \mathbf{u}_T, \mathbf{Q}_{T|T}) \tag{7.4.40}$$

It follows from the previous equations that $\mathbf{U}_{t-1}$ and $\mathbf{u}_{t-1}$ can be found recursively as

$$\mathbf{U}_t = \mathbf{D}_{t-1}\mathbf{U}_{t-1} \tag{7.4.41}$$

$$\mathbf{u}_t = \mathbf{d}_{t-1} + \mathbf{D}_{t-1}\mathbf{u}_{t-1}$$

Thus, the coefficients of the linear function can be obtained by the Kalman filter in which Eq. (7.4.32d) is replaced by Eq. (7.4.41):

Algorithm 7.4.

`Initialize:`

$$\mathbf{u}_0 = 0, \quad \mathbf{U}_0 = \mathbf{I}, \quad \mathbf{Q}_{0|0} = \boldsymbol{\Sigma}_0 \tag{7.4.42a}$$

`For` $t = 0, 1, \ldots, T-1$
`Begin`

$$\mathbf{P}_t = \mathbf{C}_t\mathbf{Q}_{t|t}\mathbf{C}_t' + \mathbf{S}_t \quad \mathbf{K}_t = \mathbf{P}_{sa,t}\mathbf{P}_{aa,t}^{-1} \tag{7.4.42b}$$

$$\mathbf{D}_t = \mathbf{F}_t - \mathbf{K}_t\mathbf{H}_t \tag{7.4.42c}$$

$$\mathbf{U}_{t+1} = \mathbf{D}_t\mathbf{U}_t, \quad \mathbf{u}_{t+1} = \mathbf{D}_t\mathbf{u}_t + \mathbf{K}_t\mathbf{a}_{t+1} \tag{7.4.42d}$$

$$\mathbf{Q}_{t+1|t+1} = \mathbf{P}_{ss,t} - \mathbf{K}_t\mathbf{P}_{as,t} \tag{7.4.42e}$$

End

We obtained the Kalman filter using the forward algorithm in its integral form. The same results can be obtained algebraically by Gaussian elimination of the rows and columns corresponding to $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_T$ from the matrix $[\Xi_T \mid (\mathbf{a}_1^T, \boldsymbol{s}_T)' - \xi_T]$ where ξ_T and Ξ_T are given by Eq. (7.4.22) and (7.4.23), respectively (see Appendix 7). Thus, the k-th step of the Kalman filter can be viewed as a Gaussian elimination of the $\mathbf{a}_k$ from the matrix that has a special structure (see Problem 1).

Example 7.6. Let us illustrate the Kalman filter application to the system of Example 7.2 assuming that the initial state is fixed.
The first step of the Kalman filter gives us

$$s_{1|1} = fs_0 + \frac{b}{b^2+\lambda^2}(a_1 - hs_0)$$

$$Q_{1|1} = \sigma^2 - \frac{b^2\sigma^2}{b^2+\lambda^2} = \frac{\lambda^2\sigma^2}{b^2+\lambda^2}$$

$$\mathbf{P}_1 = \mathbf{C}\mathbf{C}'\frac{\lambda^2\sigma^2}{b^2+\lambda^2} + \mathbf{S} = Q_{1|1}\begin{bmatrix} f^2\lambda^2+b^2+\lambda^2 & fh\lambda^2+b^3+b\lambda^2 \\ fh\lambda^2+b^3+b\lambda^2 & h^2\lambda^2+(b^2+\lambda^2)^2 \end{bmatrix}$$

Next step of the Kalman filter gives us

$$s_{2|2} = fs_{1|1} + \frac{fh\lambda^2+b^3+b\lambda^2}{h^2\lambda^2+(b^2+\lambda^2)^2}(a_2 - hs_{1|1})$$

$$Q_{2|2} = \frac{\sigma^2}{b^2+\lambda^2}\left[f^2\lambda^2+b^2+\lambda^2 - \frac{(fh\lambda^2+b^3+b\lambda^2)^2}{h^2\lambda^2+(b^2+\lambda^2)^2}\right]$$

$$\mathbf{P}_2 = \mathbf{C}\mathbf{C}'Q_{2|2} + \mathbf{S}$$

Alternatively, we can find the mean and the variance of s_2 by Gaussian elimination of the columns and rows corresponding to a_1 and a_2 from the following matrix (see Example 7.5)

$$[\Xi_2 \mid (\mathbf{a}_1^2, \boldsymbol{s}_2)' - \xi_2] = \begin{bmatrix} \sigma^2(b^2+\lambda^2) & \sigma^2(bh) & \sigma^2(bf) & a_1 - hs_0 \\ \sigma^2(bh) & \sigma^2(b^2+h^2+\lambda^2) & \sigma^2(fh+b) & a_2 - hfs_0 \\ \sigma^2(bf) & \sigma^2(fh+b) & \sigma^2(f^2+1) & s_2 - f^2 s_0 \end{bmatrix} \tag{7.4.43}$$

In the first step of the forward algorithm, we apply the Gaussian elimination of the rows and columns corresponding to a_1 (that is we eliminate the first row and then remove the first column):

$$\begin{bmatrix} \sigma^2\left(b^2+h^2+\lambda^2-\frac{b^2h^2}{b^2+\lambda^2}\right) & \sigma^2\left(fh+b-\frac{b^2fh}{b^2+\lambda^2}\right) & a_2 - hs_{1|1} \\ \sigma^2\left(fh+b-\frac{b^2fh}{b^2+\lambda^2}\right) & \sigma^2\left(f^2+1-\frac{b^2f^2}{b^2+\lambda^2}\right) & s_2 - fs_{1|1} \end{bmatrix}$$

Note that it follows from this expression that the conditional mean $s_{2|1} = \mathbf{E}(s_2 \mid a_1, s_0) = fs_{1|1}$.

In the next step of the forward algorithm we eliminate the rows and the columns corresponding to a_2 (that is the second row an column of the previous matrix) to obtain the desired result

$$\begin{bmatrix} Q_{2|2} & s_2 - s_{2|2} \end{bmatrix}$$

In the backward algorithm, we eliminate the columns and rows from the matrix (7.4.43) corresponding to a_2 first and then the columns and rows corresponding to a_1.

7.4.4 The Innovation Representation

Equation (7.4.32d) has the form

$$\boldsymbol{s}_{t+1|t+1} = \mathbf{F}_k \boldsymbol{s}_{t|t} + \mathbf{K}_t(\mathbf{a}_{t+1} - \mathbf{H}_t \boldsymbol{s}_{t|t})$$

It computes the state conditional (filtered) expectations given the observation sequence. If the observations are random variables, $\boldsymbol{s}_{t|t}$ become random and can be treated as new states. If we denote $\mathbf{e}_t = \mathbf{a}_{t+1} - \mathbf{H}_t \boldsymbol{s}_{t|t}$, then we can write

$$\begin{aligned} \boldsymbol{s}_{t+1|t+1} &= \mathbf{F}_t \boldsymbol{s}_{t|t} + \mathbf{K}_t \mathbf{e}_t \\ \mathbf{a}_{t+1} &= \mathbf{H}_t \boldsymbol{s}_{t|t} + \mathbf{e}_t \end{aligned} \tag{7.4.44}$$

This equation represents an HGMM. To prove it we need to show that $\mathbf{e}_t$ are independent zero mean Gaussian variables.

It follows from Eq. (7.4.27) that $\mathbf{e}_0 \sim \mathrm{N}(\mathbf{e}_0, \mathbf{P}_{aa,0})$. Equation (7.4.30) tells us that $p(\mathbf{e}_0, \mathbf{e}_1) = \mathrm{N}(\mathbf{e}_0, \mathbf{P}_{aa,0})\mathrm{N}(\mathbf{e}_1, \mathbf{P}_{aa,1})$ which means that $\mathbf{e}_0$ and $\mathbf{e}_1$ are independent zero mean Gaussian variables with the variance matrices $\mathbf{P}_{aa,0}$ and $\mathbf{P}_{aa,1}$, respectively. Equation (7.4.32c) allows us to conclude that $\mathbf{e}_t$ are indeed independent zero mean Gaussian variables whose variance matrices are $\mathbf{P}_{aa,t}$. And, therefore, Eq. (7.4.44) represents an HGMM. Equation (7.4.44) is called the *innovation form* of the HGMM since $\mathbf{e}_t$ is indeed the innovation: $\mathbf{e}_t = \mathbf{a}_{t+1} - E\{\mathbf{a}_{t+1} \mid \mathbf{a}_1^t\}$ (see Appendix 7). If $\mathbf{e}_t$ is random, we will denote the model's states as $\boldsymbol{s}_t$.

The complete state $\mathbf{x}_t = (\boldsymbol{s}_t, \mathbf{a}_t)$ is described by Eq. (7.4.3) which we rewrite here for the convenience of the reader

$$\mathbf{x}_{t+1} = \mathbf{C}_t \boldsymbol{s}_t + \mathbf{u}_t, \quad t = 0, 1, \ldots \tag{7.4.45a}$$

where $\mathbf{C}_t' = [\mathbf{F}_t', \mathbf{H}_t']$ and, according to Eq. (7.4.44),

$$\boldsymbol{s}_0 = \boldsymbol{\mu}_0, \quad \mathbf{u}_t \sim \mathrm{N}(0, \mathbf{S}_t^{(I)}), \quad \mathbf{S}_t^{(I)} = \begin{bmatrix} \mathbf{K}_t \mathbf{P}_{aa,t} \mathbf{K}_t' & \mathbf{K}_t \mathbf{P}_{aa,t} \\ \mathbf{P}_{aa,t} \mathbf{K}_t' & \mathbf{P}_{aa,t} \end{bmatrix} \tag{7.4.45b}$$

The original HGMM and its innovation form are equivalent if they produce the same observation sequence PDF $p(\mathbf{a}_1^T)$, $\forall\ \mathbf{a}_1^T$ and T. This means that the variance matrices $\mathbf{P}_{aa,t}$, the matrices $\mathbf{K}_t$, and initial conditions should be the same for both models. To satisfy this requirement, we use Eq. (7.4.33) and (7.4.34) to have

$$\begin{aligned} \mathbf{P}_{aa,t} &= \mathbf{H}_t \mathbf{Q}_{t|t} \mathbf{H}_t' + \mathbf{S}_{aa,t} \\ \mathbf{K}_t &= (\mathbf{F}_{t-1} \mathbf{Q}_{t-1|t-1} \mathbf{H}_{t-1}' + \mathbf{S}_{sa,t-1})(\mathbf{H}_{t-1} \mathbf{Q}_{t-1|t-1} \mathbf{H}_{t-1}' + \mathbf{S}_{aa,t-1})^{-1} \\ \mathbf{Q}_{t|t} &= \mathbf{F}_{t-1} \mathbf{Q}_{t-1|t-1} \mathbf{F}_{t-1}' + \mathbf{S}_{ss,t-1} - \mathbf{K}_t(\mathbf{F}_{t-1} \mathbf{Q}_{t-1|t-1} \mathbf{H}_{t-1}' + \mathbf{S}_{sa,t-1})' \end{aligned} \tag{7.4.46}$$

Thus, the innovation model is equivalent to the original model if the innovation variance matrix is obtained using Eq. (7.4.46) with the initial condition $\mathbf{Q}_{0|0} = \boldsymbol{\Sigma}_0$ and $\boldsymbol{s}_{0|0} = \boldsymbol{\mu}_0$. Obviously, the Kalman filter for both model is the same and is given by Eq. (7.4.34).

Suppose now that the original HGMM is defined by Eq. (7.4.45) is homogeneous (time invariant): $\mathbf{C}_t' = \mathbf{C} = [\mathbf{F}' \quad \mathbf{H}']$ and $\mathbf{S}_t = \mathbf{S}$ for $t = 0, 1, \ldots$. However, in this case according

to Eq. (7.4.46), the corresponding innovation model is not homogeneous since $\mathbf{S}_t^{(I)}$ is time varying. Under certain conditions [2] the innovation model is asymptotically homogeneous in which case there exists a a nonnegative definite matrix $\mathbf{Q} = \lim_{t\to\infty} \mathbf{Q}_{t|t}$. Both models are homogeneous if the process existed for infinite time and reached the steady state. In this case, according to Eq. (7.4.46), $\mathbf{Q}$ is a nonnegative definite solution of the following equation

$$\mathbf{Q} = \mathbf{F}\mathbf{Q}\mathbf{F}' + \mathbf{S}_{ss} - (\mathbf{F}\mathbf{Q}\mathbf{H}' + \mathbf{S}_{sa})(\mathbf{H}\mathbf{Q}\mathbf{H}' + \mathbf{S}_{aa})^{-1}(\mathbf{F}\mathbf{Q}\mathbf{H}' + \mathbf{S}_{sa})'$$

which is called the *steady state Riccati equation.* Using this solution, we find

$$\mathbf{P}_{aa} = \mathbf{H}\mathbf{Q}\mathbf{H}' + \mathbf{S}_{aa}$$
$$\mathbf{K} = (\mathbf{F}\mathbf{Q}\mathbf{H}' + \mathbf{S}_{sa})(\mathbf{H}\mathbf{Q}\mathbf{H}' + \mathbf{S}_{aa})^{-1}$$

Suppose that the original model (7.4.3) has the form of (7.4.45) where $\mathbf{P}_{aa,t}$ and $\mathbf{K}_t$ are the model independent parameters. It is not difficult to prove that this model is also its innovation form if we assume that the variance matrices $\mathbf{P}_{aa,t}$ are invertible and the initial state $\boldsymbol{s}_0 = \boldsymbol{s}_{0|0}$ is fixed. To prove it, we need to derive the Kalman filter for the model structure given by equations (7.4.45) using Algorithm 7.2 and verify that the model parameter $\mathbf{K}_t$ is indeed the Kalman gain. We leave the proof as an exercise for the reader.

If the HGMM is presented by equations of the form (7.4.45), it is called the *direct parametrization* of the HGMM. Usually, this structure has less parameters and is simpler than the general HGMM structure. If the original state is fixed, the Kalman filter is obtained by the trivial substitution, otherwise the model parameter $\mathbf{K}_t$ is not the Kalman gain. If the initial state is fixed, then, according to Eq. (7.4.46), $\mathbf{Q}_{t|t} = 0$ and Algorithm 7.4 is significantly simplified:

Algorithm 7.5.

Initialize:

$$\mathbf{u}_0 = 0, \quad \mathbf{U}_0 = \mathbf{I}$$

For $t=0,1,\ldots,T-1$
Begin

$$\mathbf{D}_t = \mathbf{F}_t - \mathbf{K}_t\mathbf{H}_t$$

$$\mathbf{U}_{t+1} = \mathbf{D}_t\mathbf{U}_t, \quad \mathbf{u}_{t+1} = \mathbf{D}_t\mathbf{u}_t + \mathbf{K}_t\mathbf{a}_{t+1}$$

End

For the fixed initial state, Eq. (7.4.40) takes the form

$$\bar{\alpha}(\mathbf{a}_1^T, \boldsymbol{s}_T; \boldsymbol{s}_0) = \mathrm{N}(\boldsymbol{s}_T - \mathbf{U}_T\boldsymbol{s}_0 - \mathbf{u}_T, 0) = \delta(\boldsymbol{s}_T - \mathbf{U}_T\boldsymbol{s}_0 - \mathbf{u}_T)$$

For the homogeneous HGMM, $\mathbf{U}_t = \mathbf{D}^t = (\mathbf{F}-\mathbf{K}\mathbf{H})^t$. If the eigenvalues of the matrix $\mathbf{F}-\mathbf{K}\mathbf{H}$ are inside the unit circle on the complex plane, then $\lim_{t\to\infty} \mathbf{U}_t = 0$ and the stationary Kalman filter exists. Since

$$\boldsymbol{s}_{t|t}(\boldsymbol{s}_0) = \mathbf{U}_t\boldsymbol{s}_0 + \mathbf{u}_t$$

for the stationary Kalman filter, the forward PDF does not depend on the initial state.

The state transition PDF, according to Eq. (7.4.8), has the form

$$p(\mathbf{x}_t \mid \boldsymbol{s}_{t-1}) = \mathsf{N}(\mathbf{a}_t - \mathbf{H}_{t-1}\boldsymbol{s}_{t-1}, \mathbf{S}_{aa,t-1})\mathsf{N}(\boldsymbol{s}_t - \boldsymbol{\mu}_{t,t-1}, 0)$$

where

$$\boldsymbol{\mu}_{t,t-1} = \mathbf{F}_{t-1}\boldsymbol{s}_{t-1} + \mathbf{K}_{t-1}(\mathbf{a}_t - \mathbf{H}_{t-1}\boldsymbol{s}_{t-1}) = \mathbf{D}_{t-1}\boldsymbol{s}_{t-1} + \mathbf{K}_{t-1}\mathbf{a}_t$$

and $\mathsf{N}(\boldsymbol{s}_t - \boldsymbol{\mu}_{t,t-1}, 0) = \delta(\boldsymbol{s}_t - \boldsymbol{\mu}_{t,t-1})$. We can prove formally that $\boldsymbol{\Sigma}_{t,t-1} = \mathbf{Q}_{0|0} = 0$ using equation (7.4.45b), but it also follows from Eq. (7.4.44) that $\boldsymbol{s}_{t+1}$ is uniquely defined, given $\mathbf{a}_{t+1}$ and $\boldsymbol{s}_t$ and, therefore,

$$p(\boldsymbol{s}_{t+1} \mid \boldsymbol{s}_t, \mathbf{a}_{t+1}) = \delta(\boldsymbol{s}_{t+1} - \boldsymbol{\mu}_{t+1,t})$$

The observation PSD is given by Eq. (7.4.20). For the homogeneous innovation model, according to (7.4.45b), this equation takes the form

$$\Phi_{aa}(z) = \mathbf{M}(z)\mathbf{K}\mathbf{P}_{aa}\mathbf{K}'\mathbf{M}'(z^{-1}) + \mathbf{M}(z)\mathbf{K}\mathbf{P}_{aa} + \mathbf{P}_{aa}\mathbf{K}'\mathbf{M}'(z^{-1}) + \mathbf{P}_{aa}$$

where

$$\mathbf{M}(z) = \mathbf{H}z^{-1}(\mathbf{I} - \mathbf{F}z^{-1})^{-1} = \mathbf{H}(\mathbf{I}z - \mathbf{F})^{-1}$$

It is easy to verify by direct multiplication that $\Phi_{aa}(z)$ can be expressed as

$$\Phi_{aa}(z) = \mathbf{W}_{aa}(z)\mathbf{P}_{aa}\mathbf{W}'_{aa}(z^{-1}) \tag{7.4.47}$$

where

$$\mathbf{W}_{aa}^{-1}(z) = [\mathbf{I} + \mathbf{M}(z)\mathbf{K}]^{-1}$$

Using the matrix inversion lemma we can express

$$\mathbf{W}_{aa}^{-1}(z) = \mathbf{I} - \mathbf{H}(\mathbf{I}z - \mathbf{F} + \mathbf{K}\mathbf{H})^{-1}\mathbf{K}$$

Equation (7.4.47) represents the so called spectral factorization.

7.4.5 The Backward Algorithm

Consider now the computation of the backward PDF $\beta(\boldsymbol{s}_{t-1}, \mathbf{a}_t^T) = p(\mathbf{a}_t^T \mid \boldsymbol{s}_{t-1})$. This PDF can be computed using the backward algorithm in Eq. (7.2.7). In this section, we show that there is a closed form expression for the integrals in this algorithm in the case of HGMM.

Indeed, if $t = T$, the closed form expression follows from Eq. (7.4.8):

$$\beta(\boldsymbol{s}_{T-1}, \mathbf{a}_T) = p(\mathbf{a}_T \mid \boldsymbol{s}_{T-1}) = \mathsf{N}(\mathbf{a}_T - \mathbf{H}_{T-1}\boldsymbol{s}_{T-1}, \mathbf{S}_{aa,T-1})$$

Using the notation $\mathsf{G}(\mathbf{x}, \mathbf{B}) = \exp\{-0.5\mathbf{x}'\mathbf{B}\mathbf{x}\}$ we can rewrite this equation as

$$\beta(\boldsymbol{s}_{T-1}, \mathbf{a}_T) = g_T\mathsf{G}(\mathbf{G}_T\boldsymbol{s}_{T-1} - \mathbf{a}_T, \mathbf{M}_T)$$

with $\mathbf{G}_T = \mathbf{H}_{T-1}$, $\mathbf{b}_T = \mathbf{a}_T$, $\mathbf{M}_T = \mathbf{S}_{aa,T-1}^{-1}$, and $g_T = (2\pi)^{-m/2}|\mathbf{S}_{aa,T-1}|^{-1/2}$ (see Eq. (A.7.10) of Appendix 7). We can assume that the other backward PDFs also have this form:

$$\beta(\boldsymbol{s}_t, \mathbf{a}_{t+1}^T) = g_{t+1}\mathsf{G}(\mathbf{G}_{t+1}\boldsymbol{s}_t - \mathbf{b}_{t+1}, \mathbf{M}_{t+1}) \tag{7.4.48}$$

We use this form because (as we will see later) the matrix $\mathbf{M}_{t+1}$ is might not be positive definite for other backward PDFs. In general, we can only assume that $\mathbf{M}_{t+1}$ is a

nonnegative definite matrix, and prove it using the method of mathematical induction by showing that $\beta(\boldsymbol{s}_{t-1}, \mathbf{a}_t^T)$ also has this form.

To prove this statement, we use Eq. (7.2.7) which in our case takes the form

$$\beta(\boldsymbol{s}_{t-1}, \mathbf{a}_t^T) = g_{t+1} \int_S \mathsf{N}(\mathbf{x}_t - \mathbf{C}_{t-1}\boldsymbol{s}_{t-1}, \mathbf{S}_{t-1}) \mathsf{G}(\mathbf{G}_{t+1}\boldsymbol{s}_t - \mathbf{b}_{t+1}, \mathbf{M}_{t+1}) d\boldsymbol{s}_t \quad (7.4.49)$$

To evaluate this integral, we use Eq. (7.4.8) to factor the integrand:

$$\mathsf{N}(\mathbf{x}_t - \mathbf{C}_{t-1}\boldsymbol{s}_{t-1}, \mathbf{S}_{t-1}) = \mathsf{N}(\mathbf{a}_t - \mathbf{H}_{t-1}\boldsymbol{s}_{t-1}, \mathbf{S}_{aa,t-1}) \mathsf{N}(\boldsymbol{s}_t - \boldsymbol{\mu}_{t,t-1}, \boldsymbol{\Sigma}_{t,t-1}) \quad (7.4.50)$$

where

$$\begin{aligned} \boldsymbol{\mu}_{t,t-1} &= \mathbf{D}_{t-1}\boldsymbol{s}_{t-1} + \mathbf{K}_{t-1}\mathbf{a}_t \\ \boldsymbol{\Sigma}_{t,t-1} &= \mathbf{S}_{ss,t-1} - \mathbf{K}_{t-1}\mathbf{S}_{as,t-1} \end{aligned} \quad (7.4.51)$$

Substituting (7.4.50) into (7.4.49) we obtain

$$\beta(\boldsymbol{s}_{t-1}, \mathbf{a}_t^T) = \mathsf{N}(\mathbf{a}_t - \mathbf{H}_{t-1}\boldsymbol{s}_{t-1}, \mathbf{S}_{aa,t-1}) I(\boldsymbol{s}_{t-1}) \quad (7.4.52)$$

where

$$I(\boldsymbol{s}_{t-1}) = g_{t+1} \int_S \mathsf{N}(\boldsymbol{s}_t - \boldsymbol{\mu}_{t,t-1}, \boldsymbol{\Sigma}_{t,t-1}) \mathsf{G}(\mathbf{G}_{t+1}\boldsymbol{s}_t - \mathbf{b}_{t+1}, \mathbf{M}_{t+1}) d\boldsymbol{s}_t$$

Using Eq. (A.7.19) of Appendix 7, we can write

$$I(\boldsymbol{s}_{t-1}) = g_{t+1} = h_t \mathsf{G}(\mathbf{G}_{t+1}\boldsymbol{\mu}_{t,t-1} - \mathbf{b}_{t+1}, \mathbf{V}_t) \quad (7.4.53)$$

where

$$\begin{aligned} \mathbf{V}_t &= \mathbf{M}_{t+1} - \mathbf{M}_{t+1}\mathbf{G}_{t+1}(\mathbf{G}_{t+1}'\mathbf{M}_{t+1}\mathbf{G}_{t+1} + \boldsymbol{\Sigma}_{t,t-1}^{-1})^{-1}\mathbf{G}_{t+1}'\mathbf{M}_{t+1} \\ &= \mathbf{M}_{t+1}(\mathbf{G}_{t+1}\boldsymbol{\Sigma}_{t,t-1}\mathbf{G}_{t+1}'\mathbf{M}_{t+1} + \mathbf{I})^{-1} \\ h_t &= g_{t+1} |\mathbf{G}_{t+1}'\mathbf{M}_{t+1}\mathbf{G}_{t+1}\boldsymbol{\Sigma}_{t,t-1}^{-1} + \mathbf{I}|^{-1/2} \end{aligned}$$

After the substitution of $\boldsymbol{\mu}_{t,t-1}$ from Eq. (7.4.51) into (7.4.53) the latter takes the form

$$I(\boldsymbol{s}_{t-1}) = h_t \mathsf{G}(\mathbf{R}_t\boldsymbol{s}_{t-1} - \mathbf{r}_t, \mathbf{V}_t)$$

where

$$\mathbf{R}_t = \mathbf{G}_{t+1}\mathbf{D}_{t-1}, \quad \mathbf{r}_t = \mathbf{b}_{t+1} - \mathbf{G}_{t+1}\mathbf{S}_{sa,t-1}\mathbf{S}_{aa,t-1}^{-1}\mathbf{a}_t$$

Using this result, we can rewrite Eq. (7.4.52) as

$$\beta(\boldsymbol{s}_{t-1}, \mathbf{a}_t^T) = h_t \mathsf{N}(\mathbf{a}_t - \mathbf{H}_{t-1}\boldsymbol{s}_{t-1}, \mathbf{S}_{aa,t-1}) \mathsf{G}(\mathbf{R}_t\boldsymbol{s}_{t-1} - \mathbf{r}_t, \mathbf{V}_t)$$

To complete the proof, we need to show that this function can be expressed in the form of Eq. (7.4.48). This can be achieved by completing the square with respect to $\boldsymbol{s}_{t-1}$ in the exponent of the previous expression:

$$(\mathbf{a}_t - \mathbf{H}_{t-1}\boldsymbol{s}_{t-1})' \mathbf{S}_{aa,t-1}^{-1} (\mathbf{a}_t - \mathbf{H}_{t-1}\boldsymbol{s}_{t-1}) + (\mathbf{R}_t\boldsymbol{s}_{t-1} - \mathbf{r}_t)' \mathbf{V}_t (\mathbf{R}_t\boldsymbol{s}_{t-1} - \mathbf{r}_t) \quad (7.4.54)$$

We use Eq. (A.5.59) of Appendix 5 to complete the square in Eq. (7.4.54) and obtain

$$\beta(\boldsymbol{s}_{t-1}, \mathbf{a}_t^T) = g_t \mathsf{G}(\mathbf{G}_t\boldsymbol{s}_{t-1} - \mathbf{b}_t, \mathbf{M}_t) \quad (7.4.55)$$

where $\mathbf{M}_t = \mathbf{G}_t^{\dagger}$

$$\mathbf{G}_t = \mathbf{H}_{t-1}'\mathbf{S}_{aa,t-1}^{-1}\mathbf{H}_{t-1} + \mathbf{R}_t'\mathbf{V}_t\mathbf{R}_{t+1} \tag{7.4.56}$$

$$\mathbf{b}_t = \mathbf{H}_{t-1}'\mathbf{S}_{aa,t-1}^{-1}\mathbf{a}_t + \mathbf{R}_t'\mathbf{V}_t\mathbf{r}_t$$

$$g_t = h_t \exp\{-0.5(\mathbf{a}_t'\mathbf{S}_{aa,t-1}^{-1}\mathbf{a}_t + \mathbf{r}_t'\mathbf{V}_t\mathbf{r}_t - \mathbf{b}_t'\mathbf{G}_t^{\dagger}\mathbf{b}_t)\}$$

As we can see, $\mathbf{G}_t$ defined by Eq. (7.4.56) might not be invertible. Therefore, we use the pseudoinverse matrix $\mathbf{M}_t = \mathbf{G}_t^{\dagger}$ in Eq. (7.4.55).

Putting it all together we can describe the backward algorithm as follows:

Algorithm 7.6.

Initialize:

$$\mathbf{G}_T = \mathbf{H}_{T-1}, \quad \mathbf{b}_T = \mathbf{a}_T, \quad \mathbf{M}_T = \mathbf{S}_{aa,T-1}^{-1}, \quad g_T = (2\pi)^{-m/2}|\mathbf{S}_{aa,T-1}|^{-1/2} \tag{7.4.57a}$$

For $t = T-1, T-2, \ldots, 1$

Begin

$$\mathbf{K}_{t-1} = \mathbf{S}_{sa,t-1}\mathbf{S}_{aa,t-1}^{-1} \tag{7.4.57b}$$

$$\mathbf{\Sigma}_{t,t-1} = \mathbf{S}_{ss,t-1} - \mathbf{K}_{t-1}\mathbf{S}_{as,t-1} \tag{7.4.57b}$$

$$\mathbf{V}_t = \mathbf{M}_{t+1}(\mathbf{G}_{t+1}\mathbf{\Sigma}_{t,t-1}\mathbf{G}_{t+1}'\mathbf{M}_{t+1} + \mathbf{I})^{-1} \tag{7.4.57b}$$

$$h_t = g_{t+1}\,|\mathbf{G}_{t+1}'\mathbf{M}_{t+1}\mathbf{G}_{t+1}\mathbf{\Sigma}_{t,t-1} + \mathbf{I}|^{-1/2} \tag{7.4.57c}$$

$$\mathbf{R}_t = \mathbf{G}_{t+1}(\mathbf{F}_{t-1} - \mathbf{K}_{t-1}\mathbf{H}_{t-1}) \tag{7.4.57d}$$

$$\mathbf{r}_t = \mathbf{b}_{t+1} - \mathbf{G}_{t+1}\mathbf{K}_{t-1}\mathbf{a}_t \tag{7.4.57e}$$

$$\mathbf{b}_t = \mathbf{H}_{t-1}'\mathbf{S}_{aa,t-1}^{-1}\mathbf{a}_t + \mathbf{R}_t'\mathbf{V}_t\mathbf{r}_t \tag{7.4.57f}$$

$$\mathbf{G}_t = \mathbf{H}_{t-1}'\mathbf{S}_{aa,t-1}^{-1}\mathbf{H}_{t-1} + \mathbf{R}_t'\mathbf{V}_t\mathbf{R}_{t+1} \tag{7.4.57g}$$

$$\mathbf{M}_t = \mathbf{G}_t^{\dagger} \tag{7.4.57h}$$

$$g_t = h_t \exp\{-0.5(\mathbf{a}_t'\mathbf{S}_{aa,t-1}^{-1}\mathbf{a}_t + \mathbf{r}_t'\mathbf{V}_t\mathbf{r}_t - \mathbf{b}_t'\mathbf{M}_t)\} \tag{7.4.57i}$$

End

This algorithm computes the parameters of the backward PDF; the PDF is presented by Eq. (7.4.48). Note that since $\mathbf{G}_{t+1}' = \mathbf{G}_{t+1}$ and $\mathbf{M}_{t+1} = \mathbf{G}_{t+1}^{\dagger}$ for $t < T-1$, we can simplify these equations by using the identity $\mathbf{G}_{t+1}\mathbf{G}_{t+1}^{\dagger}\mathbf{G}_{t+1} = \mathbf{G}_{t+1}$.

Example 7.7. Let us illustrate the backward algorithm for the system (7.4.9) of Example 7.2. According to Eq. (7.4.10), the one step transition PDF can be written in the following form

$$p(\mathbf{x}_t \mid \mathbf{s}_{t-1}) = \mathrm{N}(a_t - hs_{t-1}, S)\mathrm{N}(s_t - Ds_{t-1} - Ka_t, \Sigma) \tag{7.4.58}$$

where

$$K = b^2/(b^2+\lambda^2), \quad D = f - hK$$

$$\Sigma = \lambda^2\sigma^2/(b^2+\lambda^2)$$
$$S = \sigma^2(b^2+\lambda^2)$$

Since the transition PDF in Eq. (7.4.58) has a simple form, we rederive the backward algorithm and compare the results with that of the general case.

We have

$$\beta(s_{T-1},a_t) = \mathsf{N}(a_T - hs_{T-1},S)$$

$$\beta(s_{T-2},a_{T-1},a_T) = \int_{-\infty}^{\infty} \mathsf{N}(a_{T-1}-hs_{T-2},S)\mathsf{N}(s_{T-1}-Ds_{T-2}-Ka_{T-1},\Sigma)\beta(s_{T-1},a_t)ds_{T-1}$$
$$= \mathsf{N}(a_{T-1}-hs_{T-2},S)\mathsf{N}(a_T-hDs_{T-2}-hKa_{T-1},S+h^2\Sigma)$$

where we used Eq. (A.7.22) of Appendix 7 to evaluate the integral. In the next step, we complete the square in the $\log\beta(s_{T-2},a_{T-1},a_T)$.

It is easy to verify the following identity (which is a special case of Eq. (A.7.16) of Appendix 7)

$$\frac{(a_1x-b_1)^2}{\Sigma_1}+\frac{(a_2x-b_2)^2}{\Sigma_2} = \frac{(Ax-B)^2}{A\Sigma_1\Sigma_2}+\frac{(a_1b_2-a_2b_1)^2}{A}$$

where

$$A = a_1^2\Sigma_2 + a_2^2\Sigma_1, \quad B = a_1b_1\Sigma_2 + a_2b_2\Sigma_1$$

Using this identity, we can write

$$\mathsf{N}(a_1x-b_1,\Sigma_1)\mathsf{N}(a_2x-b_2,\Sigma_2) = A^2\mathsf{N}(Ax-B,A\Sigma_1\Sigma_2)\mathsf{N}(a_1b_2-a_2b_1,A) \tag{7.4.59}$$

(This identity is a special case of (A.7.20) of Appendix 7.) Applying this identity to $\beta(s_{T-2},a_{T-1},a_T)$ we can write:

$$\beta(s_{T-2},a_{T-1},a_T) = g_{T-1}\mathsf{N}(A_{T-1}s_{T-2}-B_{T-1},\Sigma_{T-1})$$

where

$$g_{T-1} = A_{T-1}^2\mathsf{N}(ha_T-Ua_{T-1},A_{T-1}), \quad \Sigma_{T-1} = A_{T-1}\Delta_{T-1}S$$

where $U = h(hK+D)$

$$\Delta_{T-1} = S+h^2\Sigma, \quad A_{T-1} = h^2(D^2S+\Delta_{T-1}), \quad B_{T-1} = ha_{T-1}\Delta_{T-1}+SDh(a_T-hKa_{T-1})$$

Thus, we can assume that

$$\beta(s_t,a_{t+1}^T) = g_{t+1}\mathsf{N}(A_{t+1}s_t-B_{t+1},\Sigma_{t+1})$$

The backward algorithm is given by Eq. (7.4.49) which in our case takes the form

$$\beta(s_{t-1},a_t^T) = g_{t+1}\mathsf{N}(a_t-hs_{t-1},S)\int_{-\infty}^{\infty}\mathsf{N}(s_t-Ds_{t-1}-Ka_t,\Sigma)\beta(s_t,a_{t+1}^T)ds_t$$
$$= g_{t+1}\mathsf{N}(a_t-hs_{t-1},S)\mathsf{N}(A_{t+1}Ds_{t-1}+A_{t+1}Ka_t-B_{t+1},\Delta_t)$$

where

$$\Delta_t = \Sigma_{t+1}+A_{t+1}^2\Sigma$$

Applying Eq. (7.4.59) we obtain

$$\beta(s_{t-1},a_t^T) = g_t\mathsf{N}(A_ts_{t-1}-B_t,\Sigma_t)$$

where

$$A_t = h^2\Delta_t+D^2SA_{t+1}^2 \; B_t = ha_t\Delta_t+DSA_{t+1}(B_{t+1}-KA_{t+1}a_t), \quad \Sigma_t = SA_{t+1}(\Delta_t+A_{t+1}^2\Sigma)$$

and

$$g_t = g_{t+1}A_t^2\mathsf{N}(UA_{t+1}a_t-hB_{t+1},A_t)$$

We leave it as an exercise to the reader to verify that these equations agree with Eq. (7.4.57).

We would like to note in conclusion that the backward algorithm can be developed using the expression for the kernel of the operator $\mathbf{P}(a_t^T)$ presented in Sec. 7.4.2 and applying the Gaussian elimination to the extended covariance matrix. We have illustrated this approach in Example 7.6 for the forward PDF where we performed the elimination from the "top down."

For the backward PDF, we need to perform the elimination from the "bottom up."

7.4.6 The Forward-Backward Algorithm

Let us now address the problem of evaluating the state smoothing PDF $\gamma_t(\boldsymbol{s}_t,\mathbf{a}_1^T) = p(\boldsymbol{s}_t \mid \mathbf{a}_1^T)$ given by Eq. (7.3.2). One way to present this PDF is by writing the PDF $p(\boldsymbol{s}_t,\mathbf{a}_1^T) = \mathrm{N}[(\boldsymbol{s}_t,\mathbf{a}_1^T)-\boldsymbol{\mu}(t,T),\boldsymbol{\Sigma}(t,T)]$ where $\boldsymbol{\mu}(t,T)$ and $\boldsymbol{\Sigma}(t,T)$ are computed using equations of Sec. 7.3.1. To obtain the smoothing PDF, we need to perform Gaussian elimination of the rows and columns corresponding to $\mathbf{a}_1^T$ from the matrix $[\boldsymbol{\Sigma}(t,T),\boldsymbol{\mu}(t,T)]$. This elimination can be complicated if m is large. Besides, in many applications we need to find the smoothing PDF for all $t=1,2,\ldots,T$. Therefore, it would be desirable to develop a recursive procedure for computing the PDF.

It follows from Eq. (7.3.2) that the forward-backward algorithm delivers us the recursive procedure. In the forward part, we compute and save in the computer memory $\boldsymbol{s}_{t|t}$, $\mathbf{Q}_{t|t}$, and $p(\mathbf{a}_1^t)$ using the Kalman filter described by Eq. (7.4.32). In the backward, part we compute $\beta(\boldsymbol{s}_t,\mathbf{a}_{t+1}^T)$ using equation (7.4.48) and (7.4.57) and then use equation (7.3.2) which can be written as

$$\gamma_t(\boldsymbol{s}_t,\mathbf{a}_1^T) = \frac{g_{t+1}p(\mathbf{a}_1^t)}{p(\mathbf{a}_1^T)}\,\mathrm{N}(\boldsymbol{s}_t-\boldsymbol{s}_{t|t},\mathbf{Q}_{t|t})\,\mathrm{G}(\mathbf{G}_{t+1}\boldsymbol{s}_t-\mathbf{b}_{t+1},\mathbf{M}_{t+1})$$

After completing the square in the exponent of this function, we can rewrite it using Eq. (A.7.18) of Appendix 7 as

$$\gamma_t(\boldsymbol{s}_t,\mathbf{a}_1^T) = \mathrm{N}(\boldsymbol{s}_t-\boldsymbol{s}_{t|T},\mathbf{Q}_{t|T}) \tag{7.4.60}$$

where

$$\mathbf{Q}_{t|T} = (\mathbf{G}_{t+1}'\mathbf{M}_{t+1}\mathbf{G}_{t+1} + \mathbf{Q}_{t|t}^{-1})^{-1}$$

$$\boldsymbol{s}_{t|T} = \mathbf{Q}_{t|T}(\mathbf{G}_{t+1}'\mathbf{M}_{t+1}\mathbf{b}_{t+1} + \mathbf{Q}_{t|t}^{-1}\boldsymbol{s}_{t|t}) = \boldsymbol{s}_{t|t}+\mathbf{W}_t(b_{t+1}-\mathbf{G}_{t+1}\boldsymbol{s}_{t|t})$$

$$\mathbf{W}_t = \mathbf{Q}_{t|T}\mathbf{G}_{t+1}'\mathbf{M}_{t+1}$$

Equation (7.4.60) shows that we need to compute only the mean $\boldsymbol{s}_{t|T}$ and the variance matrix $\mathbf{Q}_{t|T}$ to evaluate the smoothing PDF $\gamma_t(\boldsymbol{s}_t,\mathbf{a}_1^T)$. Therefore, there is no need to compute $p(\mathbf{a}_1^t)$ in the forward algorithm and the coefficients g_{t+1} in the backward algorithm. Since $\mathbf{G}_{t+1}' = \mathbf{G}_{t+1}$ and $\mathbf{M}_{t+1} = \mathbf{G}_{t+1}^{\dagger}$ for $t<T-1$, we can simplify these equations by using the identity $\mathbf{G}_{t+1}\mathbf{G}_{t+1}^{\dagger}\mathbf{G}_{t+1} = \mathbf{G}_{t+1}$.

7.4.7 The RTS Smoother

In Sec. 7.3, we introduced the alternative forward-backward algorithm for computing the conditional PDF $\gamma_t(\boldsymbol{s}_t,\mathbf{a}_1^T)$ without computing the backward PDFs. The backward part of this algorithm is given by equation (7.3.4d) which can be rewritten as

$$\gamma_t(\boldsymbol{s}_t,\mathbf{a}_1^T) = \int_S \gamma_{t+1}(\boldsymbol{s}_{t+1},\mathbf{a}_1^T)\,\phi(\boldsymbol{s}_t,\boldsymbol{s}_{t+1},\mathbf{a}_1^T)\,d\boldsymbol{s}_{t+1} \tag{7.4.61}$$

where

$$\phi(\boldsymbol{s}_t,\boldsymbol{s}_{t+1},\mathbf{a}_1^T) = \alpha(\boldsymbol{s}_t,\mathbf{a}_1^t)p(\boldsymbol{s}_{t+1},\mathbf{a}_{t+1} \mid \boldsymbol{s}_t)/\alpha(\boldsymbol{s}_{t+1},\mathbf{a}_1^{t+1}) \tag{7.4.62}$$

To simplify this expression, we use Eq. (7.4.31) to write

$$\alpha(\mathbf{a}_1^t,\boldsymbol{s}_t)p(\boldsymbol{s}_{t+1},\mathbf{a}_{t+1} \mid \boldsymbol{s}_t) = \mathsf{N}(\boldsymbol{s}_t-\boldsymbol{s}_{t|t},\mathbf{Q}_{t|t})p(\mathbf{a}_1^t)\mathsf{N}(\mathbf{x}_{t+1}-\mathbf{C}_t\boldsymbol{s}_t,\mathbf{S}_t)$$

This expression can be factored according to Eq. (A.7.20) of Appendix 7 as

$$\alpha(\mathbf{a}_1^t,\boldsymbol{s}_t)p(\boldsymbol{s}_{t+1},\mathbf{a}_{t+1} \mid \boldsymbol{s}_t) = \mathsf{N}(\boldsymbol{s}_t-\mathbf{q}_t,\mathbf{R}_t)\mathsf{N}(\mathbf{x}_{t+1}-\mathbf{C}_t\boldsymbol{s}_{t|t},\mathbf{P}_t)p(\mathbf{a}_1^t) \tag{7.4.63a}$$

where

$$\mathbf{q}_t = \boldsymbol{s}_{t|t} + \mathbf{W}_t(\mathbf{x}_{t+1}-\mathbf{C}_t\boldsymbol{s}_{t|t}) \tag{7.4.63b}$$

$$\mathbf{W}_t = \mathbf{Q}_{t|t}\mathbf{C}_t'\mathbf{P}_t^{-1} \tag{7.4.63c}$$

$$\mathbf{R}_t = \mathbf{Q}_{t|t} - \mathbf{W}_t\mathbf{C}_t\mathbf{Q}_{t|t} \tag{7.4.63d}$$

According to Eq. (7.4.31), $\mathsf{N}(\mathbf{x}_{t+1}-\mathbf{C}_t\boldsymbol{s}_{t|t},\mathbf{P}_t)p(\mathbf{a}_1^t) = \alpha(\boldsymbol{s}_{t+1},\mathbf{a}_1^{t+1})$ and Eq. (7.4.63) can be written as

$$\alpha(\boldsymbol{s}_t,\mathbf{a}_1^t)p(\boldsymbol{s}_{t+1},\mathbf{a}_{t+1} \mid \boldsymbol{s}_t) = \alpha(\boldsymbol{s}_{t+1},\mathbf{a}_1^{t+1})\mathsf{N}(\boldsymbol{s}_t-\mathbf{q}_t,\mathbf{R}_t)$$

This expression allows us to simplify Eq. (7.4.62) to

$$\phi(\boldsymbol{s}_t,\boldsymbol{s}_{t+1},\mathbf{a}_1^T) = \mathsf{N}(\boldsymbol{s}_t-\mathbf{q}_t,\mathbf{R}_t) \tag{7.4.64}$$

With this simplification, the backward recursive Eq. (7.4.61) becomes

$$\gamma_t(\boldsymbol{s}_t,\mathbf{a}_1^T) = \int_S \gamma_{t+1}(\boldsymbol{s}_{t+1},\mathbf{a}_1^T)\mathsf{N}(\boldsymbol{s}_t-\mathbf{q}_t,\mathbf{R}_t)d\boldsymbol{s}_{t+1} \tag{7.4.65}$$

starting with the initial value

$$\gamma_T(\boldsymbol{s}_T,\mathbf{a}_1^T) = \alpha(\boldsymbol{s}_T,\mathbf{a}_1^T) = \mathsf{N}(\boldsymbol{s}_T-\boldsymbol{s}_{T|T},\mathbf{Q}_{T|T})$$

obtained at the end of the forward algorithm.

Substituting

$$\gamma_{t+1}(\boldsymbol{s}_{t+1},\mathbf{a}_1^T) = \mathsf{N}(\boldsymbol{s}_{t+1}-\boldsymbol{s}_{t+1|T},\mathbf{Q}_{t+1|T})$$

into Eq. (7.4.65) and computing the integral according to Eq. (A.7.22) of Appendix 7 we obtain

$$\gamma_t(\boldsymbol{s}_t,\mathbf{a}_1^T) = \int_S \mathsf{N}(\boldsymbol{s}_{t+1}-\boldsymbol{s}_{t+1|T},\mathbf{Q}_{t+1|T})\mathsf{N}(\boldsymbol{s}_t-\mathbf{q}_t,\mathbf{R}_t)d\boldsymbol{s}_{t+1} = \mathsf{N}(\boldsymbol{s}_t-\boldsymbol{s}_{t|T},\mathbf{Q}_{t|T})$$

where

$$\boldsymbol{s}_{t|T} = \boldsymbol{s}_{t|t} + \mathbf{W}_t(\mathbf{x}_{t+1|T}-\mathbf{C}_t\boldsymbol{s}_{t|t}) \tag{7.4.66}$$

$$\mathbf{x}_{t+1|T} = (\boldsymbol{s}_{t+1|T},\mathbf{a}_{t+1})$$

$$\mathbf{Q}_{t|T} = \mathbf{R}_t+\mathbf{W}_t\mathbf{Q}_{t+1|T}\mathbf{W}_t' = \mathbf{Q}_{t|t}+\mathbf{W}_t(\mathbf{Q}_{t+1|T}\mathbf{W}_t' - \mathbf{C}_t\mathbf{Q}_{t|t})$$

The forward-backward algorithm that uses the Kalman filter in the forward path and the RTS smoother in the backward part can be summarized as follows.

Algorithm 7.7.

(Forward Part)
Initialize:

$$\boldsymbol{s}_{0|0} = \boldsymbol{\mu}_0, \quad \mathbf{Q}_{0|0} = \boldsymbol{\Sigma}_0 \tag{7.4.67a}$$

For $t=0,1,\ldots,T-1$
Begin

$$\mathbf{P}_t = \mathbf{C}_t \mathbf{Q}_{t|t} \mathbf{C}_t' + \mathbf{S}_t \tag{7.4.67b}$$

$$\mathbf{K}_t = \mathbf{P}_{sa,t} \mathbf{P}_{aa,t}^{-1} \tag{7.4.67c}$$

$$\boldsymbol{s}_{t+1|t+1} = \mathbf{F}_t \boldsymbol{s}_{t|t} + \mathbf{K}_t (\mathbf{a}_{t+1} - \mathbf{H}_t \boldsymbol{s}_{t|t}) \tag{7.4.67d}$$

$$\mathbf{Q}_{t+1|t+1} = \mathbf{P}_{ss,t} - \mathbf{K}_t \mathbf{P}_{as,t} \tag{7.4.67e}$$

Store $\boldsymbol{s}_{t+1|t+1}$, $\mathbf{Q}_{t+1|t+1}$, and $\mathbf{P}_t$
End

(Backward Part)
For $t=T-1,T-2,\ldots,1$
Begin

$$\mathbf{W}_t = \mathbf{Q}_{t|t} \mathbf{C}_t' \mathbf{P}_t^{-1} \tag{7.4.67f}$$

$$\boldsymbol{s}_{t|T} = \boldsymbol{s}_{t|t} + \mathbf{W}_t (\mathbf{x}_{t+1|T} - \mathbf{C}_t \boldsymbol{s}_{t|t}) \tag{7.4.67g}$$

$$\mathbf{Q}_{t|T} = \mathbf{Q}_{t|t} + \mathbf{W}_t (\mathbf{Q}_{t+1|T} \mathbf{W}_t' - \mathbf{C}_t \mathbf{Q}_{t|t}) \tag{7.4.67h}$$

End

We assumed in our derivations that the initial state PDF is Gaussian. If the initial PDF is a mixture of Gaussian PDFs as in Eq. (7.4.38), then the smoothing PDF is a mixture of smoothing PDFs which are obtained by running M forward-backward algorithms.

$$\gamma_t(\boldsymbol{s}_t, \mathbf{a}_1^T) = \sum_{i=1}^{M} w_i \gamma_{i,t}(\boldsymbol{s}_T, \mathbf{a}_1^T) \tag{7.4.68}$$

where $\gamma_{i,t}(\boldsymbol{s}_t, \mathbf{a}_1^T)$ is obtained by the forward-backward algorithm with the initial conditions $\boldsymbol{\mu}_{i,0}$ and $\boldsymbol{\Sigma}_{i,0}$.

As we can see, the conditional covariance matrices $\mathbf{P}_{t|T}$ and $\mathbf{Q}_{t|T}$ do not depend on the observation sequence and the conditional means are linear functions of the original mean whose coefficients can be found using the following algorithm.

Algorithm 7.8.
(Forward Part)
Initialize:

$$\mathbf{u}_0 = 0, \quad \mathbf{U}_0 = \mathbf{I}, \quad \mathbf{Q}_{0|0} = \boldsymbol{\Sigma}_0 \tag{7.4.69a}$$

For $t=0,1,\ldots,T-1$
Begin

$$\mathbf{P}_t = \mathbf{C}_t\mathbf{Q}_{t|t}\mathbf{C}_t' + \mathbf{S}_t \tag{7.4.69b}$$

$$\mathbf{K}_t = \mathbf{P}_{sa,t}\mathbf{P}_{aa,t}^{-1} \tag{7.4.69c}$$

$$\mathbf{D}_t = \mathbf{F}_t - \mathbf{K}_t\mathbf{H}_t \tag{7.4.69d}$$

$$\mathbf{d}_t = \mathbf{K}_t\mathbf{a}_{t+1} \tag{7.4.69e}$$

$$\mathbf{U}_{t+1|t+1} = \mathbf{D}_t\mathbf{U}_{t|t} \tag{7.4.69f}$$

$$\mathbf{u}_{t+1|t+1} = \mathbf{D}_t\mathbf{u}_{t|t} + \mathbf{d}_{t|t} \tag{7.4.69g}$$

$$\mathbf{Q}_{t+1|t+1} = \mathbf{P}_{ss,t} - \mathbf{K}_t\mathbf{P}_{as,t} \tag{7.4.69h}$$

Store $\mathbf{u}_{t+1|t+1}$, $\mathbf{U}_{t+1|t+1}$, $\mathbf{Q}_{t+1|t+1}$, and $\mathbf{P}_t$
End
(Backward Part)
For $t=T-1,T-2,\ldots,1$
Begin

$$\mathbf{W}_t = \mathbf{Q}_{t|t}\mathbf{C}_t'\mathbf{P}_t^{-1} \tag{7.4.69i}$$

$$\mathbf{y}_{t|T} = \begin{bmatrix} \mathbf{u}_{t+1|T} \\ \mathbf{a}_{t+1} \end{bmatrix} \tag{7.4.69j}$$

$$\mathbf{u}_{t|T} = \mathbf{u}_{t|t} + \mathbf{W}_t(\mathbf{y}_{t+1|T} - \mathbf{C}_t\mathbf{u}_{t|t}) \tag{7.4.69k}$$

$$\mathbf{U}_{t|T} = (\mathbf{I} - \mathbf{W}_t\mathbf{C}_t)\mathbf{U}_{t|t} + \mathbf{W}_{s,t}\mathbf{U}_{t+1|T} \tag{7.4.69l}$$

$$\mathbf{Q}_{t|T} = \mathbf{Q}_{t|t} + \mathbf{W}_t(\mathbf{Q}_{t+1|T}\mathbf{W}_t' - \mathbf{C}_t\mathbf{Q}_{t|t}) \tag{7.4.69m}$$

End

It is not difficult to verify that

$$\boldsymbol{s}_{t|T} = \mathbf{U}_{t|T}\boldsymbol{s}_0 + \mathbf{u}_{t|T}$$

is the solution of Eq. (7.4.66) and, therefore,

$$\gamma_t(\boldsymbol{s}_t, \mathbf{a}_1^T; \boldsymbol{s}_0) = \mathsf{N}(\boldsymbol{s}_t - \mathbf{U}_{t|T}\boldsymbol{s}_0 - \mathbf{u}_{t|T}, \mathbf{Q}_{t|T})$$

Consider now computing the sliding-window smoothing PDF for the size 2 window. This PDF can be found using equation (7.3.3) which in our case, according to Eq. (7.4.62) and (7.4.64) can be written as

$$\gamma_t(\boldsymbol{s}_t, \boldsymbol{s}_{t+1}, \mathbf{a}_1^T) = \mathsf{N}(\boldsymbol{s}_{t+1} - \boldsymbol{s}_{t+1|T}, \mathbf{Q}_{t|T})\mathsf{N}(\boldsymbol{s}_t - \mathbf{q}_t, \mathbf{R}_t) \tag{7.4.70}$$

Obviously, this is a Gaussian PDF. To present it explicitly, we need to find the covariance matrix for $\boldsymbol{s}_t$ and $\boldsymbol{s}_{t+1}$.

It follows from the right hand side of the previous equation that we can write

$$\boldsymbol{s}_{t+1} = \boldsymbol{s}_{t+1|T} + \varepsilon_1$$
$$\boldsymbol{s}_t = \mathbf{q}_t + \varepsilon_2 = \boldsymbol{s}_{t|t} + \mathbf{W}_t(\mathbf{x}_{t+1} - \mathbf{C}_t\boldsymbol{s}_{t|t}) + \varepsilon_2$$

where $\varepsilon_1 \sim \mathsf{N}(0, \mathbf{Q}_{t|T})$ and $\varepsilon_2 \sim \mathsf{N}(0, \mathbf{R}_t)$. Using this equation, we obtain the mean

$$E\{\boldsymbol{s}_t\} = \boldsymbol{s}_{t|t} + \mathbf{W}_t(\mathbf{x}_{t+1|T} - \mathbf{C}_t\boldsymbol{s}_{t|t}) = \boldsymbol{s}_{t|T}$$

which obviously coincides with (7.4.66). To find the covariance matrix, we use the previous equation to write

$$\boldsymbol{s}_t' - \boldsymbol{s}_{t|t}' = (\mathbf{x}_{t+1}' - \mathbf{x}_{t+1|T}')\mathbf{W}_t' = [(\boldsymbol{s}_{t+1} - \boldsymbol{s}_{t+1|t+1})' \quad 0]\mathbf{W}_t' = \varepsilon_1'\mathbf{W}_{s,t}' \tag{7.4.71}$$

where $\mathbf{W}_{s,t}$ is the subblock of the matrix $\mathbf{W}_t$. Using this expression we obtain the variance matrix

$$E\{(\boldsymbol{s}_{t+1} - \boldsymbol{s}_{t+1|T})(\boldsymbol{s}_t - \boldsymbol{s}_{t|T})'\} = E\{\varepsilon_1\varepsilon_1'\}\mathbf{W}_{s,t}' = \mathbf{Q}_{t+1|T}\mathbf{W}_{s,t}'$$

Therefore, we can rewrite Eq. (7.4.70) as

$$\gamma_t(\boldsymbol{s}_t, \boldsymbol{s}_{t+1}, \mathbf{a}_1^T) = \mathsf{N}(\xi_t - \xi_{t|T}, \Xi_{t|T})$$

where

$$\xi_t = \begin{bmatrix} \boldsymbol{s}_{t+1} \\ \boldsymbol{s}_t \end{bmatrix}, \quad \xi_{t|T} = \begin{bmatrix} \boldsymbol{s}_{t+1|T} \\ \boldsymbol{s}_{t|T} \end{bmatrix}, \quad \Xi_{t|T} = \begin{bmatrix} \mathbf{Q}_{t+1|T} & \mathbf{Q}_{t+1|T}\mathbf{W}_{s,t}' \\ \mathbf{W}_{s,t}\mathbf{Q}_{t+1|T} & \mathbf{Q}_{t|T} \end{bmatrix}$$

7.4.8 The Viterbi Algorithm

The Viterbi algorithm, given by Eq. (7.3.6) and (7.3.7) or (7.3.8), finds the sequence of conditional means $\boldsymbol{s}_{1|T}, \boldsymbol{s}_{2|T}, \ldots, \boldsymbol{s}_{T|T}$. For the HGMM, the first equation in (7.3.6) is

$$\psi_1(\boldsymbol{s}_1) = \max_{\boldsymbol{s}_0} p(\boldsymbol{s}_1, \mathbf{a}_1 \mid \boldsymbol{s}_0)p_0(\boldsymbol{s}_0) = \max_{\boldsymbol{s}_0} \mathsf{N}(\mathbf{x}_1 - \mathbf{C}_0\boldsymbol{s}_0, \mathbf{S}_0)\mathsf{N}(\boldsymbol{s}_0 - \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$$

Since the PDF in the right hand side of this equation is Gaussian, its maximum is proportional to the marginal PDF (see Eq. (A.7.30) of Appendix 7). Therefore,

$$\psi_1(\boldsymbol{s}_1) = Cp(\mathbf{x}_1) = C\alpha(\boldsymbol{s}_1, \mathbf{a}_1)$$

We see that $\psi_1(\boldsymbol{s}_1)$ is proportional to the forward PDF given by equation (7.4.25) which in its turn is proportional to the normalized forward PDF given by equation (7.4.37). Therefore, we can write

$$\psi_1(\boldsymbol{s}_1) = C_1\mathsf{N}(\boldsymbol{s}_1 - \boldsymbol{s}_{1|1}, \mathbf{Q}_{1|1})$$

Repeating the Viterbi algorithm, we find that

$$\psi_k(\boldsymbol{s}_k) = C_k\mathsf{N}(\boldsymbol{s}_k - \boldsymbol{s}_{k|k}, \mathbf{Q}_{k|k})$$

$k = 1, 2, \ldots, T$. This equation shows that, for the HGMM, the forward part of the Viterbi algorithm coincides with the forward algorithm and is performed using the Kalman filter.

At the end of the forward part we have $\psi(\boldsymbol{s}_T) = C_T\mathsf{N}(\boldsymbol{s}_T - \boldsymbol{s}_{T|T}, \mathbf{Q}_{T|T})$. In the first step of the backward part of the Viterbi algorithm we compute

$$\boldsymbol{s}_{T-1|T} = \underset{\boldsymbol{s}_{T-1}}{arg\max} \mathrm{N}(\mathbf{x}_{T|T} - \mathbf{C}_{T-1}, \mathbf{S}_{T-1}) \mathrm{N}(\boldsymbol{s}_{T-1} - \boldsymbol{\mu}_{T-1}, \boldsymbol{\Sigma}_0)$$

The solution to this type of equation is given by Eq. (A.7.21) of Appendix 7 which in our case has the form

$$\boldsymbol{s}_{T-1|T} = \boldsymbol{s}_{T-1|T-1} + \mathbf{W}_{T-1}(\mathbf{x}_T - \mathbf{C}_{T-1}\boldsymbol{s}_{T-1|T-1})$$

where

$$\mathbf{W}_{T-1} = \mathbf{Q}_{T-1|T-1}\mathbf{C}'_{T-1}(\mathbf{C}_{T-1}\mathbf{Q}_{T-1|T-1}\mathbf{C}'_{T-1} + \mathbf{S}_{T-1})^{-1}$$

which coincides with the first equation of the RTS smoother. We can prove similarly that this is true for the rest of the backward steps of the Viterbi algorithm:

$$\boldsymbol{s}_{k|T} = \boldsymbol{s}_{k|k} + \mathbf{W}_k(\mathbf{x}_{k+1|T} - \mathbf{C}_k\boldsymbol{s}_{k|k}), \quad k = T-1, T-2, \ldots, 0$$

Therefore, we conclude that the forward part of the Viterbi algorithm coincides with the forward algorithm and is performed using the Kalman filter while the backward part of the Viterbi algorithm is performed using the RTS smoother. The latter, however, computes not only the means $\boldsymbol{s}_{k|T}$, but also the variance matrices $\mathbf{Q}_{k|T}$.

7.5 AUTOREGRESSIVE MOVING AVERAGE PROCESSES

Scalar autoregressive moving average (ARMA) process is defined by the following equation

$$\sum_{i=0}^{p} a_i y_{t-i} = \sum_{i=0}^{q} b_i w_{t-i}, \quad t = 1, 2, \ldots \tag{7.5.1}$$

where $a_0 \neq 0$, $b_0 \neq 0$, $w_t \sim \mathrm{N}(w_t, \sigma^2)$ and are independent. If $q = 0$, this equation represents the *autoregressive* (AR) process (see Appendix 7). If $p = 0$, it represents the *moving average* process.

Without loss of generality, we can assume that $a_0 = 1$ (otherwise we divide both sides of the equation by a_0).

Introducing the delay operator z^{-m} defined as $z^{-m}y_t = y_{t-m}$, we can present Eq. (7.5.1) as

$$a(z^{-1})y_t = b(z^{-1})w_t$$

where

$$a(z) = \sum_{i=0}^{p} a_i z^i, \quad b(z) = \sum_{i=0}^{q} b_i z^i$$

Thus, we can write

$$y_t + \sum_{i=1}^{p} a_i y_{t-i} = \sum_{i=0}^{q} b_i w_{t-i}, \quad t = 1, 2, \ldots$$

or

$$y_t = -\mathbf{a}'\mathbf{y}_{t-1} + \mathbf{b}'\mathbf{w}_t + b_0 w_t, \quad t = 1, 2, \ldots$$

where $\mathbf{a} = (a_1, a_2, \ldots, a_p)$, $\mathbf{b} = (b_1, b_2, \ldots, b_q)$, $\mathbf{y}_t = (y_t, y_{t-1}, \ldots, y_{t-p+1})$, and $\mathbf{w}_t = (w_t, w_{t-1}, \ldots, w_{t-q+1})$.

7.5.1 State Space Representation

To apply the previously developed theory, we need to represent the ARMA process in the state space form. Obviously, the ARMA process represents a partially observable Gauss-Markov process with states $\mathbf{x}_t = (\mathbf{w}_t, \mathbf{y}_t)$ where $\mathbf{y}_t = (y_t, y_{t-1}, \ldots, y_{t-p+1})$ represents the observable part of the state while $\mathbf{w}_t = (w_t, w_{t-1}, \ldots, w_{t-q+1})$ is not observable. As we know, any partially observable Markov process can be represented as an HGMM using the trivial augmentation of the hidden states described by Eq. (7.4.6). This representation is called the direct type I structure in the signal processing theory. It treats the complete state $\mathbf{x}_t$ as not observable and, therefore, the state dimension is $p+q$ which requires $p+q$ shift registers to store the state variable.

There are many other structures describing ARMA processes which require less shift registers than the direct type I structure. We consider here the most popular structure (which is used by the MATLAB function `filter`) the so-called transposed direct type II structure requiring the minimum number of the shift registers for its implementation. Without loss of generality and to simplify notation we can assume that vectors **a** and **b** have the same length (we can equalize the lengths by setting $a_{p+1}=a_{p+2}=\cdots=a_q=0$ if $p<q$ or $b_{q+1}=b_{q+2}=\ldots=b_p=0$ if $p>q$; the MATLAB function `eqtflength` performs this equalization).

To derive the state space equation, we rewrite Eq. (7.5.1) as

$$y_t = b_0 w_t + \sum_{i=1}^{p} (b_i w_{t-i} - a_i y_{t-i})$$

and introduce the state variables recursively as partial sums of the sum in the previous equation:

$$\eta_p^{(t-p)} = b_p w_{t-p} - a_p y_{t-p}, \quad \eta_i^{(t-i)} = \eta_{i+1}^{(t-i-1)} + b_i w_{t-i} - a_i y_{t-i} \tag{7.5.2}$$

$i=p-1, p-2, \ldots, 1$. It follows from these equations that

$$\eta_1^{(t-1)} = \sum_{i=1}^{p} (b_i w_{t-i} - a_i y_{t-i})$$

And, therefore,

$$y_t = b_0 w_t + \eta_1^{(t-1)}$$

The structure implementing these equations is depicted in Fig. 7.1. The state of the system at the moment t is defined by the contents of the shift registers $\boldsymbol{s}_t = (\eta_1^{(t)}, \eta_2^{(t)}, \ldots, \eta_{p-1}^{(t)})$. According to Eq. (7.5.2) and Fig. 7.1 we can write

$$\begin{aligned} \boldsymbol{s}_t &= \boldsymbol{\Gamma}_p \boldsymbol{s}_{t-1} + \mathbf{b} w_t - \mathbf{a} y_t \\ y_t &= \mathbf{H} \boldsymbol{s}_{t-1} + b_0 w_t \end{aligned} \tag{7.5.3}$$

where

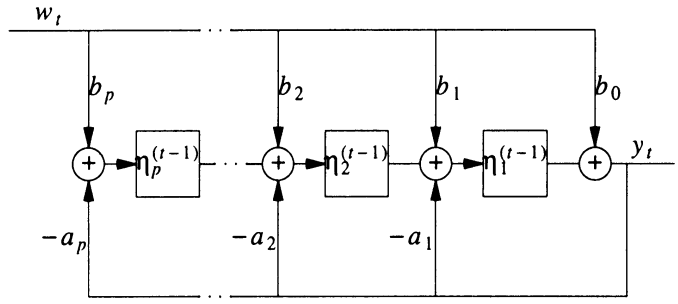

Figure 7.1. Transposed direct form II structure of the ARMA model.

$$\mathbf{\Gamma}_p = \begin{bmatrix} 0 & 1 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 1 \\ 0 & 0 & \cdots & 0 \end{bmatrix}, \quad \mathbf{H} = [1 \ 0 \ 0 \ \ldots \ 0]$$

As we can see, Eq. (7.5.3) describes a partially observable Markov process. We can convert it into the state space form of the HGMM by eliminating y_t from the first equation using the second equation to obtain

$$\begin{aligned} \boldsymbol{s}_t &= \mathbf{F}\boldsymbol{s}_{t-1} + \mathbf{G}w_t \\ y_t &= \mathbf{H}\boldsymbol{s}_{t-1} + b_0 w_t \end{aligned} \tag{7.5.4}$$

where

$$\mathbf{F} = \mathbf{\Gamma}_p - \mathbf{a}\mathbf{H} = \begin{bmatrix} -a_1 & 1 & \cdots & 0 \\ -a_2 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ -a_{p-1} & 0 & \cdots & 1 \\ -a_p & 0 & \cdots & 0 \end{bmatrix} \tag{7.5.5}$$

$$\mathbf{G} = \mathbf{b} - b_0\mathbf{a}$$

The MATLAB function `tf2ss` computes the matrices in Eq. (7.5.4).

The initial conditions are defined as the initial contents of the shift registers $\boldsymbol{s}_0 = (\eta_1^{(0)}, \eta_2^{(0)}, \ldots, \eta_p^{(0)})$. If we know first p initial observations $\boldsymbol{y}_1^p$, we can find the initial conditions producing these observations by solving first p equations in (7.5.4) for $\boldsymbol{s}_0$ (see example 7.8). The solution can be found using the MATLAB function `filtic`.

Equation (7.4.3) for the complete state $\mathbf{x}_k = (\boldsymbol{s}_k, y_k)$ has the form

$$\mathbf{x}_k = \mathbf{C}\boldsymbol{s}_{k-1} + \mathbf{u}_k \tag{7.5.6a}$$

where

$$\mathbf{C} = \begin{bmatrix} \mathbf{F} \\ \mathbf{H} \end{bmatrix} = \begin{bmatrix} -a_1 & 1 & 0 & \cdots & 0 \\ -a_2 & 0 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ -a_{p-1} & 0 & 0 & \cdots & 1 \\ -a_p & 0 & 0 & \cdots & 0 \\ 1 & 0 & 0 & \cdots & 0 \end{bmatrix} \tag{7.5.6b}$$

$$\mathbf{u}_k = \sigma w_t \begin{bmatrix} \mathbf{G} \\ b_0 \end{bmatrix} \sim \mathrm{N}(\mathbf{u}_k, \mathbf{S}), \quad \mathbf{S} = \begin{bmatrix} \mathbf{S}_{ss} & \mathbf{S}_{sa} \\ \mathbf{S}_{as} & \mathbf{S}_{aa} \end{bmatrix} = \sigma^2 \begin{bmatrix} \mathbf{G}\mathbf{G}' & b_0\mathbf{G} \\ b_0\mathbf{G}' & b_0^2 \end{bmatrix} \tag{7.5.6c}$$

As we can see, $\mathbf{u}_t$ is a singular Gaussian variable (see Appendix 7). Since all the columns (and rows) of the variance matrix $\mathbf{S}$ are proportional, $rank\mathbf{S} = 1$.

Example 7.8. Consider the ARMA process

$$y_t + a_1 y_{t-1} + a_2 y_{t-2} = w_t + b_1 w_{t-1}$$

In this case, $p = 2$ and $q = 1$. Therefore, we assume that $b_2 = 0$ to equalize the lengths of the vectors $\mathbf{a} = (a_1, a_2)$ and $\mathbf{b} = (b_1, 0)$.

Equation (7.5.4) has the form

$$\boldsymbol{s}_t = \begin{bmatrix} -a_1 & 1 \\ -a_2 & 0 \end{bmatrix} \boldsymbol{s}_{t-1} + \begin{bmatrix} b_1 - a_1 \\ -a_2 \end{bmatrix} w_t, \quad y_t = \begin{bmatrix} 1 & 0 \end{bmatrix} \boldsymbol{s}_{t-1} + w_t$$

If we want the two initial observations to have particular values y_1 and y_2, we can find the corresponding initial conditions $\eta_1^{(0)}, \eta_2^{(0)}$ by solving the system

$$y_1 = \eta_1^{(0)}, \quad y_2 = -a_1 y_1 + \eta_2^{(0)}$$

assuming that $w_1 = w_2 = 0$. We recommend the reader to verify the correctness of this statement using Fig. 7.1 in which shift register contents were originally $\eta_1^{(0)} = y_1$ and $\eta_2^{(0)} = y_2 + a_1 y_1$.

The complete state $\mathbf{x}_t = (\boldsymbol{s}_t, y_t)$ is described by

$$\mathbf{x}_t = \begin{bmatrix} -a_1 & 1 \\ -a_2 & 0 \\ 1 & 0 \end{bmatrix} \boldsymbol{s}_{t-1} + \mathbf{u}_{t-1}$$

where

$$\mathbf{u}_{t-1} \sim \mathrm{N}(\mathbf{u}_{t-1}, \mathbf{S}), \quad \mathbf{S} = \sigma^2 \begin{bmatrix} (b_1 - a_1)^2 & -a_2(b_1 - a_1) & b_1 - a_1 \\ -a_2(b_1 - a_1) & a_2^2 & -a_2 \\ b_1 - a_1 & -a_2 & 1 \end{bmatrix}$$

7.5.2 Autocovariance Function

The autocovariance function for the ARMA process can be obtained using equations developed in Sec. 7.3. In particular, for the stationary process the autocovariance function is given by Eq. (7.4.19). To use this equation, we need to find first the variance matrix $\boldsymbol{\Sigma}_{ss,\infty}$ (which exists only if all the eigenvalues of $\mathbf{F}$ are inside the unit circle of the complex plane). $\boldsymbol{\Sigma}_{ss,\infty}$ can be found by solving the Lyapunov Eq. (7.4.17). The solution can be found using the iterative algorithm A.6.2 or the the Yule-Walker Eq. (A.7.55) of Appendix 7. MATLAB function `dlyap` finds the solution. Then we find $\boldsymbol{\Sigma}_{aa,\infty}$ and $\boldsymbol{\Sigma}_{sa,\infty}$ using equation (7.4.18) and then substitute them into Eq. (7.4.19) to find the covariance function which in our case takes the form

$$R_{aa}(0) = \Sigma_{aa,\infty}, \quad R_{aa}(\tau) = \mathbf{H}\mathbf{F}^{-\tau-1}\Sigma_{sa,\infty} \quad \text{for} \quad \tau<0 \tag{7.5.7}$$

$R_{aa}(\tau) = R_{aa}(-\tau)$ for $\tau>0$.

Example 7.9. Let us find the autocovariance function for the ARMA process

$$y_k - ay_{k-1} = w_k + bw_{k-1}$$

The state space form for this process is given by Eq. (7.5.4) which in our case takes the following form

$$s_t = as_{t-1} + (a+b)w_t, \quad y_t = s_{t-1} + w_t \tag{7.5.8}$$

In other words,

$$\mathbf{F} = a, \quad \mathbf{H} = 1, \quad \mathbf{G} = a+b, \quad \mathbf{S} = \sigma^2 \begin{bmatrix} (a+b)^2 & a+b \\ a+b & 1 \end{bmatrix}$$

We find $\Sigma_{ss,\infty} = \sigma^2(a+b)^2/(1-a^2)$ (which exists only if $|a|<1$) by solving the Lyapunov Eq. (7.4.17) $\Sigma_{ss,\infty} = a^2\Sigma_{ss,\infty} + \sigma^2(a+b)^2$. Equations (7.4.18) take the form

$$\Sigma_{aa,\infty} = \frac{(a+b)^2\sigma^2}{(1-a^2)} + \sigma^2, \quad \Sigma_{sa,\infty} = \frac{a(a+b)^2\sigma^2}{(1-a^2)} + (a+b)\sigma^2$$

and the autocovariance function is given by (7.5.7):

$$R_{aa}(0) = \Sigma_{aa,\infty}, \quad R_{aa}(\tau) = \Sigma_{sa,\infty}a^{|\tau|-1} \quad \text{for } \tau\neq 0$$

7.5.3 The Forward Algorithm

The forward algorithm for the ARMA process is realized by the Kalman filter. As we can see, Eq. (7.5.6) have the form of Eq. (7.4.45). Therefore, Eq. (7.5.6) represent the direct parametrization of the innovation model. Consequently, if the initial state $\boldsymbol{s}_0 = \boldsymbol{s}_{0|0}$, the Kalman filter is obtained by substituting w_t from the second equation of (7.5.4) into its first equation:

$$\boldsymbol{s}_{t|t} = \mathbf{F}\boldsymbol{s}_{t-1|t-1} + \mathbf{K}(y_t - \mathbf{H}\boldsymbol{s}_{t-1|t-1})$$

where the Kalman gain $\mathbf{K} = b_0^{-1}\mathbf{G} = b_0^{-1}\mathbf{b} - \mathbf{a}$.

Example 7.10. Let us consider the forward algorithm for the ARMA process given in Example 7.11 assuming that the initial state $s_0 \sim \mathrm{N}(s_0-\mu_0,\sigma_0^2)$.

Using Algorithm 7.3, we obtain the following relations

$$Q_{t+1|t+1} = Q_{t|t}a^2 + \sigma^2(a+b)^2 - K_t[Q_{t|t}a + \sigma^2(a+b)]$$
$$K_t = [Q_{t|t}a + \sigma^2(a+b)](Q_{t|t}+\sigma^2)^{-1}$$

After simplifications, we obtain

$$Q_{t+1|t+1} = \sigma^2 b^2 Q_{t|t}(Q_{t|t}+\sigma^2)^{-1}, \quad K_t = [Q_{t|t}a + \sigma^2(a+b)](Q_{t|t}+\sigma^2)^{-1} \tag{7.5.9}$$

which gives us $K_0 = [\sigma_0^2 a + \sigma^2(a+b)](\sigma_0^2+\sigma^2)^{-1}$, $s_{1|1} = a\mu_0 + K_0(y_1-\mu_0)$, $Q_{1|1} = \sigma^2 b^2\sigma_0^2(\sigma_0^2+\sigma^2)^{-1}$, $K_1 = [Q_{1|1}a + \sigma^2(a+b)](Q_{1|1}+\sigma^2)^{-1}$, and so on. As we can see, the Kalman filter is time varying even if the HGMM is stationary (in which case $\sigma_0^2 = \Sigma_{ss,\infty}$).

It follows from Eq. (7.5.9) that $\lim_{t\to\infty} Q_{t|t} = 0$ if $|b|<1$. In this case, $\lim_{t\to\infty} K_t = a+b$. Thus, the *stationary Kalman filter* takes the form

$$s_{t+1|t+1} = as_{t|t} + (a+b)(y_{t+1} - s_{t|t})$$

This equation can be obtained by substituting w_t from the second equation of (7.5.8) into its first equation. If the initial state is fixed ($s_{0|0} = s_0$ and $\sigma_0 = 0$), then Eq. (7.5.8) coincides with the Kalman filter. Using this equation we can write

$$s_{t+1|t+1} = -bs_{t|t} + (a+b)y_{t+1} = b^2 s_{t-1|t-1} - b(a+b)y_t + (a+b)y_{t+1} = \cdots$$
$$= (a+b)\sum_{k=0}^{\infty}(-b)^k y_{t-k+1}$$

Thus, the stationary Kalman filter can be viewed as the weighted sum of all the previous observations.

7.5.4 Power Spectrum

Let us compute the PSD for the ARMA model using Eq. (7.4.47) which in our case has the form $\mathbf{\Phi}_{yy} = W(z)W(z^{-1})\sigma^2$ where

$$W(z) = 1 + \mathbf{H}z^{-1}(\mathbf{I} - \mathbf{F}z^{-1})^{-1}\mathbf{K} \tag{7.5.10}$$

Without loss of generality and to simplify notation we assume that $b_0 = 1$. (Otherwise, we replace $\mathbf{b}$ with $\mathbf{b}_1 = b_0^{-1}\mathbf{b}$ and replace σ^2 with $\sigma_1^2 = \sigma^2 b_0^2$.) In this case, $\mathbf{K} = \mathbf{b} - \mathbf{a}$ and It is not difficult to show (see Problem 2) that

$$\mathbf{H}(\mathbf{I} - \mathbf{F}z)^{-1} = \frac{1}{a(z)}[1 \;\; z \;\; \ldots \;\; z^{p-1}]$$

where $a(z) = 1 + a_1 z + \ldots + a_p z^p$. Also

$$[1 \;\; z \;\; \ldots \;\; z^{p-1}]\mathbf{K} = [1 \;\; z \;\; \ldots \;\; z^{p-1}](\mathbf{b} - \mathbf{a}) = z^{-1}[b(z) - a(z)]$$

where $b(z) = 1 + b_1 z + \ldots + b_p z^p$. Consequently, Eq. (7.5.10) takes the form

$$W(z) = 1 + [b(z) - a(z)]/a(z) = a(z)/b(z)$$

and Eq. (7.4.47) gives us the PSD

$$\Phi_{yy}(z) = \frac{b(z)b(z^{-1})}{a(z)a(z^{-1})}\sigma^2$$

The PSD on the unit circle represents the Fourier transform of the autocorrelation function and can be written as

$$\Phi_{yy}(e^{j\omega}) = \frac{|b(e^{j\omega})|^2}{|a(e^{j\omega})|^2}\sigma^2$$

7.6 PARAMETER ESTIMATION

7.6.1 HMM Approximation

As in Sec. 3.2.2, consider a problem of approximating a process with multidimensional PDF $f(\mathbf{a}_1^T)$ with an HMM by minimizing the KLD between their PDFs which is equivalent to maximizing the likelihood function (see Sec. 3.1.3):

$$L(\tau) = \int_{A^T} f(\mathbf{a}_1^T)\log p(\mathbf{a}_1^T;\tau)\,d\mathbf{a}_1^T = \mathbf{E}\{\log p(\mathbf{a}_1^T;\tau) \mid f(\mathbf{a}_1^T)\} \tag{7.6.1}$$

where τ is the unknown parameter and $p(\mathbf{a}_1^T;\tau)$ is defined by Eq. (7.1.3) or (7.2.3).

This problem can be solved using the EM algorithm. The EM algorithm for the

continuous state HMM can be derived similarly to its discrete counterpart of Sec. 3.2.2:

$$\tau_{p+1} = \underset{\tau}{arg\max}\, Q(\tau,\tau_p) \tag{7.6.2}$$

where

$$Q(\tau,\tau_p) = E\{q(\mathbf{a}_1^T,\tau,\tau_p) \mid f(\mathbf{a}_1^T)\}, \quad p=0,1,... \tag{7.6.3}$$

$$q(\mathbf{a}_1^T,\tau,\tau_p) = E\{\log\psi(\boldsymbol{s}_0^T;\mathbf{a}_1^T,\tau) \mid \kappa(\boldsymbol{i}_0^T;\mathbf{a}_1^T,\tau_p)\} \tag{7.6.4}$$

As in Sec. 3.2.2, we define the auxiliary function $\kappa(\boldsymbol{s}_0^T;\mathbf{a}_1^T,\tau) = p(\boldsymbol{s}_0^T \mid \mathbf{a}_1^T,\tau)$. It follows from this definition that

$$\psi(\boldsymbol{s}_0^T;\mathbf{a}_1^T,\tau) = \kappa(\boldsymbol{s}_0^T;\mathbf{a}_1^T,\tau)p(\mathbf{a}_1^t;\tau) = p(\boldsymbol{s}_0^T,\mathbf{a}_1^T;\tau)$$

is the PDF of the complete state sequence $\mathbf{x}_t = (\boldsymbol{s}_t,\mathbf{a}_t)$. Since $\mathbf{x}_t$ is a Markov process, then

$$\psi(\boldsymbol{s}_0^T;\mathbf{a}_1^T,\tau) = p_0(\boldsymbol{s}_0,\tau)\prod_{t=1}^{T} p(\mathbf{x}_t \mid \boldsymbol{s}_{t-1};\tau)$$

and, therefore,

$$\log\psi(\boldsymbol{s}_0^T;\mathbf{a}_1^T,\tau) = \log p_0(\boldsymbol{s}_0,\tau) + \sum_{t=1}^{T} \log p(\mathbf{x}_t \mid \boldsymbol{s}_{t-1};\tau)$$

Substituting this expression into Eq. (7.6.4) we obtain

$$q(\mathbf{a}_1^T,\tau,\tau_p) = q_0(\mathbf{a}_1^T,\tau,\tau_p) + q_1(\mathbf{a}_1^T,\tau,\tau_p) \tag{7.6.5}$$

where

$$q_0(\mathbf{a}_1^T,\tau,\tau_p) = E\{\log p_0(\boldsymbol{s},\tau) \mid \kappa(\boldsymbol{s}_0^T;\mathbf{a}_1^T,\tau_p)\}$$

$$q_1(\mathbf{a}_1^T,\tau,\tau_p) = \sum_{t=1}^{T} E\{\log p(\mathbf{x}_t \mid \boldsymbol{s}_{t-1};\tau) \mid \kappa(\boldsymbol{s}_0^T;\mathbf{a}_1^T,\tau_p)\}$$

Since $\kappa(\boldsymbol{s}_0^T;\mathbf{a}_1^T,\tau) = p(\boldsymbol{s}_0^T \mid \mathbf{a}_1^T,\tau)$ we can simplify these expressions:

$$q_0(\mathbf{a}_1^T,\tau,\tau_p) = E\{\log p_0(\boldsymbol{s},\tau) \mid \gamma_0(\boldsymbol{s},\mathbf{a}_1^T,\tau_p)\}$$

$$q_1(\mathbf{a}_1^T,\tau,\tau_p) = \sum_{t=1}^{T} E\{\log p(\mathbf{x}_t \mid \boldsymbol{s}_{t-1};\tau) \mid \gamma_{t-1}(\boldsymbol{s}_{t-1},\boldsymbol{s}_t,\mathbf{a}_1^T,\tau_p)\}$$

where $\gamma_{t-1}(\boldsymbol{s}_{t-1},\boldsymbol{s}_t,\mathbf{a}_1^T,\tau) = p(\boldsymbol{s}_{t-1},\boldsymbol{s}_t \mid \mathbf{a}_1^T,\tau)$ can be evaluated using the forward-backward algorithm described in Sec. 7.2 and 7.3. Using these equations, we can express Eq. (7.6.3) as

$$Q(\tau,\tau_p) = Q_0(\tau,\tau_p) + Q_1(\tau,\tau_p) \tag{7.6.6}$$

where

$$\begin{aligned} Q_0(\tau,\tau_p) &= E\{\log p_0(\boldsymbol{s},\tau) \mid \Gamma_0(s,\tau,\tau_p)\} \\ \Gamma_0(\boldsymbol{s},\tau,\tau_p) &= E\{\gamma_0(\boldsymbol{s},\mathbf{a}_1^T,\tau_p) \mid f(\mathbf{a}_1^T)\} \end{aligned} \tag{7.6.7}$$

$$Q_1(\tau,\tau_p) = \sum_{t=1}^{T} E\{\log p(\mathbf{x}_t \mid \boldsymbol{s}_{t-1};\tau) \mid \rho_{t-1}(\boldsymbol{s}_{t-1},\boldsymbol{s}_t,\mathbf{a}_1^T,\tau_p)\}$$
$$\rho_{t-1}(\boldsymbol{s}_{t-1},\boldsymbol{s}_t,\mathbf{a}_1^T,\tau_p)\} = \gamma_{t-1}(\boldsymbol{s}_{t-1},\boldsymbol{s}_t,\mathbf{a}_1^T,\tau) f(\mathbf{a}_1^T) \tag{7.6.8}$$

and then apply the EM algorithm given by Eq. (7.6.2).

7.6.1.1 Fitting HMM to Experimental Data

Suppose that we have multiple independent observation sequences $\{\mathbf{a}_1\}_1^{T_1} = (\mathbf{a}_{11},\mathbf{a}_{12},\ldots,\mathbf{a}_{1,T_1})$, $\{\mathbf{a}_2\}_1^{T_2} = (\mathbf{a}_{21},\mathbf{a}_{22},\ldots,\mathbf{a}_{2,T_2}),\ldots,$ $\{\mathbf{a}_K\}_1^{T_K} = (\mathbf{a}_{K1},\mathbf{a}_{K2},\ldots,\mathbf{a}_{K,T_2})$ of the same HMM. In this case, the maximum likelihood estimate of the model parameters is

$$\hat{\tau} = \arg\max_{\tau} \sum_{k=1}^{K} \log p(\{\mathbf{a}_k\}_1^{T_k},\tau)$$

As in Sec 3.2.3 we can show that this problem is a special case of the HMM approximation in which the expectation $E\{(\cdot) \mid f(\boldsymbol{x}_1^T)\}$ with respect to $f(\boldsymbol{x}_1^T)$ is replaced by $(1/K)\sum_1^K(\cdot)$. In this case, the EM algorithm described by Eq. (7.6.2) takes the form

$$\tau_{p+1} = \arg\max_{\tau} \sum_{k=1}^{K} q(\{\mathbf{a}_k\}_1^{T_k},\tau,\tau_p), \quad p=0,1,\ldots$$

As we can see, the EM algorithm for the HMM approximation is similar to its discrete counterpart of Sec. 3.2.2 and can be considered as a generalization of the latter (if we use δ-functions). Note that the continuous state algorithm is, generally, more complex than the finite state algorithm, because integration is a more complex operation than multiplication of matrices. However, as we will show in the next sections, it is possible to find closed form expressions for the integrals in the case of HGMM.

7.6.2 HGMM Approximation

We consider a case of approximating a continuous state process with the multidimensional PDF $f(\mathbf{a}_1^T)$ by the homogeneous HGMM which is described by equation (7.4.3) which in our case takes the form

$$\boldsymbol{s}_{t+1} = \mathbf{F}\boldsymbol{s}_t + \mathbf{u}_{s,t}$$
$$\mathbf{a}_{t+1} = \mathbf{H}\boldsymbol{s}_t + \mathbf{u}_{a,t}$$

If we denote as $\mathbf{C}' = [\mathbf{F}',\mathbf{H}']$ and $\mathbf{u}_t = (\mathbf{u}_{s,t},\mathbf{u}_{a,t})$, this equation takes a more compact form:

$$\mathbf{x}_{t+1} = \mathbf{C}\boldsymbol{s}_t + \mathbf{u}_t$$

We assume that $\mathbf{u}_t \sim N(0,\mathbf{S})$ and the initial state PDF is also Gaussian $\boldsymbol{s}_0 \sim N(\boldsymbol{\mu}_0,\boldsymbol{\Sigma}_0)$. Our goal is to find the parameter $\tau = \{\mathbf{C},\mathbf{S},\boldsymbol{\mu}_0,\boldsymbol{\Sigma}_0\}$ maximizing the likelihood function (7.6.1).

To solve this problem using the EM algorithm (7.6.2), we need to find the auxiliary function $q(\mathbf{a}_1^T,\tau,\tau_p)$ given by Eq. (7.6.5). Since $p_0(\boldsymbol{s},\tau) = \mathrm{N}(\boldsymbol{s}-\boldsymbol{\mu}_0,\boldsymbol{\Sigma}_0)$ and

$p(\mathbf{x}_t \mid \boldsymbol{s}_{t-1}) = \mathsf{N}(\mathbf{x}_t - \mathbf{C}_{t-1}\boldsymbol{s}_{t-1}, \mathbf{S})$, Eq. (7.6.7) and (7.6.8) with some abuse of notation can be written as

$$Q_0(\tau,\tau_p) = -0.5E\{[\ (\boldsymbol{s}-\boldsymbol{\mu}_0)'\boldsymbol{\Sigma}_0^{-1}(\boldsymbol{s}-\boldsymbol{\mu}_0) + \log|\boldsymbol{\Sigma}_0|\ ] \mid \Gamma_0(\boldsymbol{s},\tau_p)\}$$

$$Q_1(\tau,\tau_p) = -0.5\sum_{t=1}^{T} E\{[(\mathbf{x}_t - \mathbf{C}\boldsymbol{s}_{t-1})'\mathbf{S}^{-1}(\mathbf{x}_t - \mathbf{C}\boldsymbol{s}_{t-1}) + \log|\mathbf{S}|] \mid \rho_{t-1}(\boldsymbol{s}_{t-1},\boldsymbol{s}_t,\mathbf{a}_1^T,\tau_p)\}$$

where we neglected the additive constants that do not depend on τ.

Since the model parameters are independent, we can maximize $Q_0(\tau,\tau_p)$ and $Q(\tau,\tau_p)$ separately. It is not difficult to show (see Example 3.1.2 of Chapter 3) that the maximum of $Q_0(\tau,\tau_p)$ is attained at

$$\boldsymbol{\mu}_{0,p+1} = E\{\boldsymbol{s} \mid \Gamma_0(\boldsymbol{s},\tau_p)\}$$

and

$$\boldsymbol{\Sigma}_{0,p+1} = E\{(\boldsymbol{s}-\boldsymbol{\mu}_{0,p+1})(\boldsymbol{s}-\boldsymbol{\mu}_{0,p+1})' \mid \Gamma_0(\boldsymbol{s},\tau_p)\}$$

To find a maximum of $Q_1(\tau,\tau_p)$ we use the identity

$$(\mathbf{x}_t - \mathbf{C}\boldsymbol{s}_{t-1})'\mathbf{S}^{-1}(\mathbf{x}_t - \mathbf{C}\boldsymbol{s}_{t-1}) = tr[\mathbf{S}^{-1}(\mathbf{x}_t - \mathbf{C}\boldsymbol{s}_{t-1})(\mathbf{x}_t - \mathbf{C}\boldsymbol{s}_{t-1})']$$

and the properties of the trace operator (see Appendix 5) to express $Q_1(\tau,\tau_p)$ as

$$Q_1(\tau,\tau_p) = -0.5tr[\mathbf{S}^{-1}(\mathbf{B}_{1,p} - \mathbf{C}\mathbf{B}_{2,p}' - \mathbf{B}_{2,p}\mathbf{C}' + \mathbf{C}\mathbf{B}_{3,p}\mathbf{C}')] - 0.5T\log|\mathbf{S}| \tag{7.6.9}$$

where

$$\mathbf{B}_{1,p} = \sum_{t=1}^{T} E\{\mathbf{x}_t\mathbf{x}_t' \mid \rho_{t-1}(\boldsymbol{s}_{t-1},\boldsymbol{s}_t,\mathbf{a}_1^T,\tau_p)\} \tag{7.6.10a}$$

$$\mathbf{B}_{2,p} = \sum_{t=1}^{T} \mathbf{E}\{\mathbf{x}_t\boldsymbol{s}_{t-1}' \mid \rho_{t-1}(\boldsymbol{s}_{t-1},\boldsymbol{s}_t,\mathbf{a}_1^T,\tau_p)\} \tag{7.6.10b}$$

$$\mathbf{B}_{3,p} = \sum_{t=1}^{T} \mathbf{E}\{\boldsymbol{s}_{t-1}\boldsymbol{s}_{t-1}' \mid \rho_{t-1}(\boldsymbol{s}_{t-1},\boldsymbol{s}_t,\mathbf{a}_1^T,\tau_p)\} \tag{7.6.10c}$$

It is not difficult to show (see Eq. (7.6.9), (A.7.60), and (A.7.62) of Appendix 7) that the maximization step for $Q_1(\tau,\tau_p)$ can be expressed by equations

$$\mathbf{C}_{p+1} = \mathbf{B}_{2,p}\mathbf{B}_{3,p}^{-1} \tag{7.6.11a}$$

$$\mathbf{S}_{p+1} = \frac{1}{T}(\mathbf{B}_{1,p} - \mathbf{B}_{2,p}\mathbf{B}_{3,p}^{-1}\mathbf{B}_{2,p}') = \frac{1}{T}(\mathbf{B}_{1,p} - \mathbf{C}_{p+1}\mathbf{B}_{2,p}') \tag{7.6.11b}$$

provided that $\mathbf{B}_{3,p}^{-1}$ exists.

7.6.2.1 Fitting HGMM to Experimental Data

Since fitting the a HGMM to experimental data $\{\mathbf{a}_1\}_1^{T_1} = (\mathbf{a}_{11},\mathbf{a}_{12},\ldots,\mathbf{a}_{1,T_1})$, $\{\mathbf{a}_2\}_1^{T_2} = (\mathbf{a}_{21},\mathbf{a}_{22},\ldots,\mathbf{a}_{2,T_2}),\ldots,$ $\{\mathbf{a}_K\}_1^{T_K} = (\mathbf{a}_{K1},\mathbf{a}_{K2},\ldots,\mathbf{a}_{K,T_2})$ is a special case of approximation when $\Gamma_0(\boldsymbol{s},\tau) = \sum_{k=1}^{K}\gamma_0(\boldsymbol{s},\{\mathbf{a}_k\}_1^{T^*},\tau)/K$ and

$$\rho_{t-1}(\boldsymbol{s}_{t-1},\boldsymbol{s}_t,\{\mathbf{a}_k\}_1^{T_t},\tau_p) = \gamma_{t-1}(\boldsymbol{s}_{t-1},\boldsymbol{s}_t,\{\mathbf{a}_k\}_1^{T_t},\tau_p)/K$$

we can use the EM algorithm given by Eq. (7.6.11). To simplify notation, we consider only the case of one experimental sequence $\mathbf{a}_1^T$. (As we pointed out previously, the general case is obtained by simple averaging with respect to the multiple sequences.) For the one experimental sequence ($K=1$), $\Gamma_0(\boldsymbol{s},\tau) = \gamma_0(\boldsymbol{s},\mathbf{a}_1^T,\tau)$, and $\rho_{t-1}(\boldsymbol{s}_{t-1},\boldsymbol{s}_t,\mathbf{a}_1^T,\tau_p) = \gamma_{t-1}(\boldsymbol{s}_{t-1},\boldsymbol{s}_t,\mathbf{a}_1^T,\tau_p)$. In this case we have $\boldsymbol{\mu}_{0,p+1} = \boldsymbol{s}_{0|T,p}$ and $\boldsymbol{\Sigma}_{0,p+1} = \mathbf{Q}_{0|T,p}$ where index p means that the smoothing has been performed using the HGMM parameters obtained at p-th iteration (reestimation) of the EM algorithm.

$$\mathbf{B}_{1,p} = \sum_{t=1}^{T} E\{\mathbf{x}_t\mathbf{x}_t' \mid \gamma_{t-1}(\boldsymbol{s}_{t-1},\boldsymbol{s}_t,\mathbf{a}_1^T,\tau_p)\}, \quad \text{where} \quad \begin{bmatrix} \boldsymbol{s}_t\boldsymbol{s}_t' & \boldsymbol{s}_t\mathbf{a}_t' \\ \mathbf{a}_t\boldsymbol{s}_t' & \mathbf{a}_t\mathbf{a}_t' \end{bmatrix}$$

The variance matrix $E\{(\boldsymbol{s}_t-\boldsymbol{s}_{t|T})(\boldsymbol{s}_t-\boldsymbol{s}_{t|T})'\} = \mathbf{Q}_{t|T}$ can be found using the forward-backward algorithm (the Kalman filtering followed by the RTS smoother). From this equation, we find $E\{\boldsymbol{s}_t\boldsymbol{s}_t'\} = \mathbf{Q}_{t|T} + \boldsymbol{s}_{t|T}\boldsymbol{s}_{t|T}'$. Therefore,

$$\mathbf{B}_{1,p} = \sum_{t=1}^{T} \begin{bmatrix} \mathbf{Q}_{t|T,p} + \boldsymbol{s}_{t|T,p}\boldsymbol{s}_{t|T,p}' & \boldsymbol{s}_{t|T,p}\mathbf{a}_t' \\ \mathbf{a}_t\boldsymbol{s}_{t|T,p}' & \mathbf{a}_t\mathbf{a}_t' \end{bmatrix} \tag{7.6.12a}$$

The rest of the expectations in Eq. (7.6.10) can be found similarly

$$\mathbf{B}_{2,p} = \sum_{t=1}^{T} \begin{bmatrix} \mathbf{Q}_{t|T,p}\mathbf{W}_{s,t}' + \boldsymbol{s}_{t|T,p}\boldsymbol{s}_{t-1|T,p}' \\ \mathbf{a}_t\boldsymbol{s}_{t-1|T,p}' \end{bmatrix} \tag{7.6.12b}$$

where $\mathbf{W}_{s,t}$ is defined by Eq. (7.4.71), and

$$\mathbf{B}_{3,p} = \sum_{t=1}^{T} \mathbf{Q}_{t-1|T,p} + \boldsymbol{s}_{t-1|T,p}\boldsymbol{s}_{t-1|T,p}' \tag{7.6.12c}$$

7.7 ARMA CHANNEL MODELING

7.7.1 Fading Channel

In this section, we describe the ARMA modeling of the wireless fading channels based on the real measured data which was collected at 1900 Mhz along several drive routes in a residential area and on a highway, with vehicle speeds of 30 and 65 mph, and downrange distances between 2 to 5 miles. Further details on the test system and measurements are presented in Ref. 11. We considered two cases. [18] In case R (for "residential"), vehicle moves at 30 mph while in case H (for "highway") it moves at 65 mph. Our results show that ARMA modeling of the fading process is quite adequate while the broadly used theoretical model [3,8] does not agree with the experimental data. [18]

Let $x(t)$ be a low-pass equivalent of the transmitted signal with the inphase component $x_I(t) = \mathrm{Re}\{x(t)\}$ and quadrature component $x_Q(t) = \mathrm{Im}\{x(t)\}$. Consider a frequency-nonselective (flat) fading channel with the additive white Gaussian noise (AWGN) $n(t)$. This channel can be modeled as [12]

$$y(t) = c(t)x(t) + n(t) \tag{7.7.1}$$

where $y(t)$ is the received signal and fading is modeled by the complex random process $c(t)$, channel noise and fading are assumed to be independent. In the case of Rayleigh fading, $c(t)$ is a stationary zero-mean Gaussian processes with independent and identically distributed real and imaginary parts. It is called the Rayleigh fading, because its envelope $\alpha(t) = |c(t)|$ is Rayleigh distributed:

$$Pr\{\alpha(t) < a\} = 1 - \exp(-0.5a^2/\mu)$$

If fading mean is non-zero, the fading is called Rician which accounts for the presence of a line-of-sight (LOS) component.

Different models of fading channels are based on different assumptions about the fading PSD $S(f)$ (or autocorrelation function $R(\tau)$). The Clarke's model [3,8] assumes that

$$S(f) = \mu / \pi\sqrt{f_D^2 - f^2}, \qquad R(\tau) = \mu J_0(2\pi f_D|\tau|)$$

where $J_0(\cdot)$ is the Bessel function of the first kind and f_D is the maximal Doppler frequency. The normalized autocorrelation function $J_0(2\pi f_D|\tau|)$ has only one independent parameter and, therefore, fits poorly to the experimental data that are not necessarily agree with the theoretical model. The ARMA model, on the other hand, is more versatile because its PSD is a rational function and, therefore, can approximate reasonably well any measured PSD. It is also more convenient to use the ARMA model because of the rich theory of HGMM and availability of MATLAB tools.

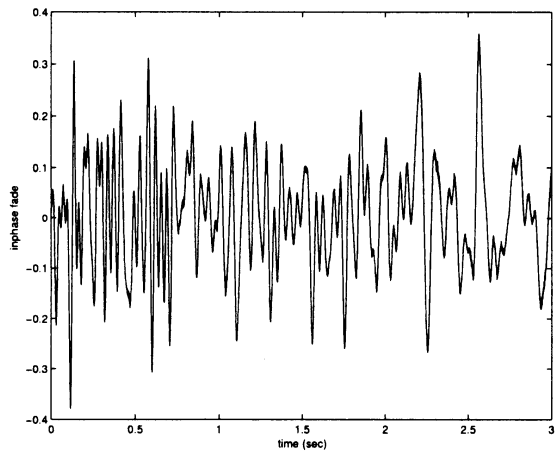

Figure 7.2. Fading inphase component (case R).

Based on the measured data, we found that the inphase and quadrature components of the fading are independent and can be studied independently. A sample of the inphase component (for the case R) is depicted in Figure 7.2. The model's power spectral density of the ARMA model for this component is compared with that of measured data in Figure 7.3. As we can see in this figure, they agree quite well and do not agree with the PSD of the

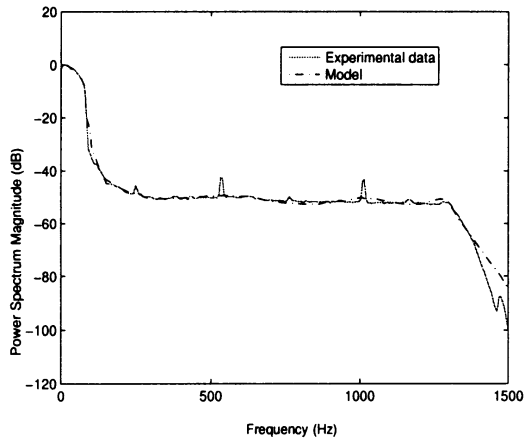

Figure 7.3. Comparison of power spectral densities (case R).

Clarke's model which has a characteristic U-shape. [3,8]

We would like to note in conclusion that, for the mobile applications, a better choice would be to use a switched HGMM model in which the observation sequence is described by the HGMM

$$\boldsymbol{s}_t = \mathbf{F}(\boldsymbol{r}_{t-1})\boldsymbol{s}_{t-1} + \mathbf{G}(\boldsymbol{r}_{t-1})w_t$$

$$y_t = \mathbf{H}(\boldsymbol{r}_{t-1})\boldsymbol{s}_{t-1} + b_0(\boldsymbol{r}_{t-1})w_t$$

whose parameters depend on the switching process $\boldsymbol{r}_t$. The switching process describes the changing environment (such as velocity, shadowing, etc) in mobile applications while the HGMM models the fading for these various environments. The swithcing process is usually described by a discrete state Markov chain whose parameters are estimated by aggregating the observation process.

7.7.2 Ultra-Wide Bandwidth Radio Channel Modeling

Typically, the ARMA model is applied to describing correlated stochastic processes in the time domain (as we demonstrated in the previous section). In this section, we demonstrate the model application in the frequency domain to describe the correlation between the frequency responses at different frequencies.

Based on the experimental data for the in-home environment [19] we found that the ultra-wide bandwidth (UWB) channel frequency response does not exhibit significant variability in time and can be modeled by the AR model of second order:

$$H(f_i) + a_1 H(f_{i-1}) + a_2 H(f_{i-2}) = w_i$$

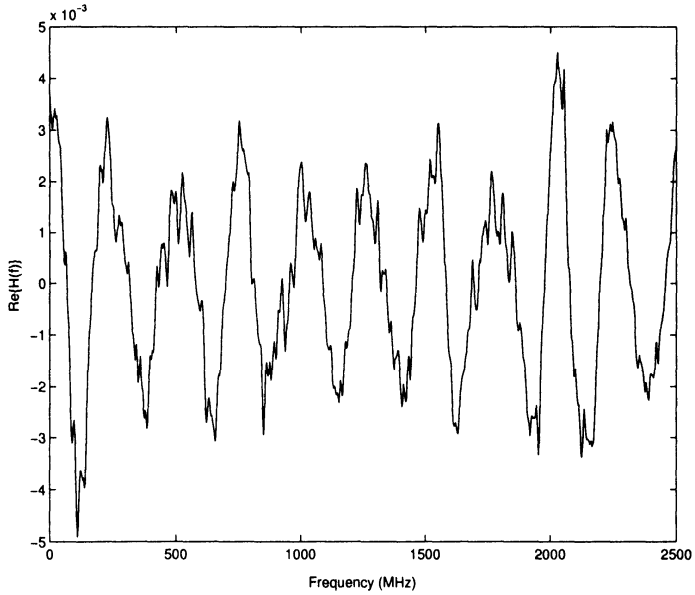

Figure 7.4. Measured real part of frequency response.

where $w_i \sim N(w_i, \sigma^2)$. The coefficients of this equation, the initial conditions, and σ^2 are changing from house to house and also depend on the distance between the transmitter and receiver. A similar approach was used in Ref. 6 and 7.

Typically, for the time domain processes we have one long observation sequence so that we do not have enough statistical data to estimate the PDF of the initial conditions. For the frequency response data, we have many short observation sequences allowing us to estimate this PDF. We applied the MATLAB function `filtic` to obtain the initial conditions $\eta_1^{(0)}$ and $\eta_2^{(0)}$ (see Fig 7.3) from the experimental data and then found they are strongly correlated with each other but independent from the rest of the model parameters. The model coefficients a_1 and a_2 are represented by the poles p_1 and p_2 which are the roots of the polynomial $z^2 + a_1 z + a_2$. We found that the poles are correlated but independent from the rest of the parameters. The white noise standard deviation does not depend on the rest of the model parameters. The results of our study are reported in Ref. 19.

As an illustration, consider one of the models corresponding to 2 ft transmitter-receiver separation. A sample of measured data is depicted in Fig. 7.4 We fitted the AR model to the frequency response and obtained the following parameters

$$a_1 = -1.5025 + j0.8017, \quad a_2 = 0.4427 + j0.7538,$$
$$\eta_1 = (3.168 - j0.003)10^{-3}, \quad \eta_2 = (-2.18 + j2.86)10^{-3}, \quad \sigma = 1.36\ 10^{-4}$$

We found [19] that this model agrees reasonably well with the experimental data. The model based computations of various time domain characteristics also agree with that obtained using the experimental data. For example, we obtained good agreement between the estimated and theoretically predicted delay profiles. [13] The popular time domain models of wireless channels assume that the delay profile samples at different delays are independent

Gaussian variables. We found that these models use more parameters and are less accurate than the frequency domain model for the UWB channel.

Some models assume that the delay profile can be represented by the exponential function with the AWGN. According to our model, however, the mean delay profile is represented by the rational function which agrees better with the experimental data than the exponential function.

7.8 CONCLUSION

Continuous state HMMs are very popular in many applications. Traditionally, their subset presented by the stochastic linear equations was studied thoroughly without referencing the HMMs. The theory of linear stochastic systems was devoted predominantly to estimating of their states and parameters. The theory of discrete state HMMs on the other hand, was developed independently using a more "probabilistic" approach consisting of computing of various probabilities. In this chapter, we applied this approach to studying the continuous state HMMs. This allowed us to consider all the HMMs using the same framework, clearly present the results related general (not necessarily Gaussian) HMMs, and obtain basic equations related to the Gaussian HMMs much simpler than it is done using the traditional approach.

Traditionally, the non-Gaussian HMMs are approximated by the HGMM to be able to apply their theory. By considering them in the framework of the general HMMs it is possible to use the alternative approaches using the quantized state approximation, numerical methods for computing integrals, and mixed discrete-continuous state HMMs.

We show that the continuous state HMMs can be defined as partially observable Markov processes. This definition is more natural and clearer than the traditional one. It covers not only discrete and continuous state HMMs, but also the mixed discrete-continuous state HMMs. The introduction of the operator probabilities allowed us to generalize methods of the matrix probability theory to the operator probability theory. The matrix probability theory is a special case of the operator probability theory.

We do not consider the case of continuous time and space HMM. The theory of these processes is beyond the scope of this book primarily oriented on applications. The theory can be developed as a generalization of the discrete time HMM theory similarly to Chapter 6. We just would like to point out that this theory in some cases is simpler. For example, the backward differential equations are obtained from the forward equations by a simple time variable substitution. Any implementation of a system of differential equations on a *digital* computer is a *discrete time* finite difference system. Therefore, from the practical point of view, it is sufficient to consider only discrete time models.

Problems

1. Derive the Kalman filter by writing the mean and covariance matrix for $(\boldsymbol{s}_t, \mathbf{a}_1^t)$ (see Example 7.6).
2. Let $\mathbf{Y}_t = [y_t \;\; y_{t-1} \;\; \cdots \;\; y_{t-p+1}]$
   — Show that

$$\sum_{i=p-1} \mathbf{Y}_i z^i = [\psi_{p-1}(z) \;\; \psi_{p-2}(z) \;\; \ldots \;\; \psi_0(z)]$$

where

$$\psi_m(z) = z^{-m}(\phi(z) - \sum_{k=0}^{m-1} y_k z^k) \tag{7.P.1}$$

and $\phi(z) = \sum_{i=0}^{\infty} y_i z^i$ is the generating function for the sequence y_0^{∞}.

— Show that the generating function for the solution of the linear homogeneous equations

$$y_t + \sum_{i=1}^{p} a_i y_{t-i} = 0, \quad t = p, p+1, \ldots \tag{7.P.2}$$

with the initial condition $(y_{p-1}, y_{p-2}, \ldots, y_0)$ has the form

$$\phi(z) = [y_0 a_1(z) + z y_1 a_2(z) + \ldots + z^{p-1} y_{p-1} a_p(z)]/a_0(z) \tag{7.P.3}$$

where $a_m(z) = 1 + a_1 z + a_2 z^2 + \ldots + a_{p-m} z^{p-m}$

— Show that, for the matrix $\mathbf{F}$ given by Eq. (7.5.5),

$$\mathbf{Y}_{p-1}(\mathbf{I} - \mathbf{F}z)^{-1} = [\psi_{p-1}(z) \;\; \psi_{p-2}(z) \;\; \ldots \;\; \psi_0(z)]$$

where $\psi_m(z)$ and $\phi(z)$ are given by Eq. (7.P.1) and (7.P.3), respectively.
(Hint: the left hand side of this equation is the generating function for the solution of equation $\mathbf{Y}_t = \mathbf{Y}_{t-1}\mathbf{F}$ which can be written as (7.P.2).)

3. Show that

$$p(\boldsymbol{s}_{t-1} \mid \boldsymbol{s}_t, \mathbf{a}_1^t) = p(\boldsymbol{s}_{t-1} \mid \boldsymbol{s}_t, \mathbf{a}_1^T)$$

[Hint: use Eq. (7.4.62)]

References

1 P. L. Ainsleigh, N. Kehtarnavaz, and R. L. Streit, "Hidden Gauss-Markov models for signal classification," *IEEE Trans. Signal Proc.*, **50**(6), 1355-1367, June (2002).

2 B. D. O. Anderson and J. B. Moore, *Optimal Filtering*, Prentice-Hall, Inglewood Cliffs, N.J., 1979

3 R. H. Clarke, "A statistical theory of mobile radio reception," *Bell Syst. Tech. J.*, **47** 957-1000, 1968.

4 B. Delyon, "Remarks on linear and nonlinear filtering," *IEEE Trans. Inf. Theory,* **IT-41**(1), 317-322 (1995).

5 V. Digalakis, J. R. Rohlicek, and M. Ostendorf, "ML estimation of a stochastic linear system with the EM algorithm and its application to speech recognition," *IEEE Trans. Speech Audio Processing*, **1**, 431-442, Oct. (1993).

6 S.J. Howard, K. Pahlavan, " Measurement and Analysis of the indoor radio channel in the frequency domain", *IEEE Trans. Instrum. Measure.,*, **39** 751-755, Oct. (1990).

7 S.J. Howard, K. Pahlavan, "Autoregressive Modeling of Wide-Band Indoor Radio Propagation",*IEEE Transaction on Commun.,* **40** 1540-1552, September (1992).

8 W.C. Jakes, *Microwave Mobile Communications*, (Wiley, 1974, New York).

9 M. Morf and T. Kailath, Square-root algorithms for least-squares estimation, *IEEE Trans. AC*, **20(4)**, 487-496, (1975).

10 R. E. Kalman, "A new approach to linear filtering and prediction problems," *Trans. ASME,* **82D** 34-45, Mar. (1960).

11 C. C. Martin, J. H. Winters, and N. R. Sollenberger, "Multiple-input multiple-output (MIMO) radio channel measurements," *Proc. IEEE Veh. Technol. Conf.,* 774-779, Sept. (2000).

12 J.G. Proakis, *Digital Communications,* (McGraw-Hill, New York, 1989).

13 T. S. Rappaport, *Wireless Communications: Principles and Practice.* (Prenice-Hall, New Jersey, 1996).

14 H. E. Rauch, F. Tung, and C. T. Streibel, "Maximum likelihood estimates of linear dynamic systems," *AIAA Journ.*, **3**(8), 1445-1450, Aug. (1965).

15 R. Y. Rubinstein, *Simulation and Monte-Carlo Method,* (Wiley, New York, 1981).

16 R. H. Shumway and D. S. Stoffer, "An approach to time series smoothing and forecasting using the EM algorithm," *J. Time Series Anal.*, **3**(4), 253-264, (1982).

17 T. Soderstrom and P. Stoica, *System Identification*, (Prentice Hall, New York, 1989).

18 W. Turin, R. Jana, S. S. Ghassemzadeh, C. Rice, and V. Tarokh, "Autoregressive modeling of an indoor UWB channel,"*IEEE Conference on Ultra Wideband Systems and Technologies*, 71-74, May (2002).