Università degli Studi di Padova

Dipartimento di Ingegneria dell'Informazione

Dottorato di Ricerca in Automatica e Ricerca Operativa – XVII Ciclo

Ph.D. Thesis

Estimation Problems in Jump Markov Systems

Coordinator: Ch.mo Prof. Augusto FERRANTE

Supervisor: Ch.mo Prof. Giorgio PICCI

Ph.D. candidate: Eugenio CINQUEMANI

Padova, February 28, 2005

Abstract

In this work we address estimation problems in linear systems subject to Markovian jumps. We first introduce the class of jump Markov linear systems. These are discrete-time systems jumping in time among a finite set of known linear modes. Each mode is defined by a stochastic difference equation, describing the evolution of a continuous state vector x_k, and a static equation modeling noisy measurements y_k of x_k. The current operating mode is determined by the value of a discrete state q_k taking values in a finite set $\mathcal{Q}$. The evolution of q_k follows the laws of a discrete-time Markov chain. We illustrate the problem of estimating the values of x_k and q_k based on the measurements y^k collected up to time k. The optimal Bayesian solution is presented and is shown to have complexity exponentially increasing in time. An overview of certain finite-complexity approximate estimation algorithms is presented. Open issues in suboptimal estimation and limitations of the discrete-time model are discussed. A similar model is then considered for systems with continuous-time dynamics. In this model, a stochastic differential equation governs the evolution of a continuous state $\xi(t)$. Measurements y_k of the state are taken at fixed sample times t_k. Switching among different discrete states is determined by the outcome $q(t)$ of a continuous-time Markov chain taking values in $\mathcal{Q}$. We state the problem of estimating $\xi(t)$ and $q(t)$ based on the collection of data y^k and formulate the optimal solution of the problem. Application of conditioned Kalman filtering and a recursive estimation algorithm are discussed for the case of filtering and prediction. State estimation on single-switch models is studied in depth and is shown to be equivalent to a fault detection problem. Algorithms for fault detection are derived and are evaluated on the basis of numerical simulations. Finally, we focus back on state estimation in discrete time, and consider the problem of detecting the whole sequence of discrete states $q_0, \dots q_k$ from the available data y^k. A sequential testing rule for the selection of a set of most probable sequences is considered. We investigate the connection between the structural properties of the system and the ability of the test to discriminate sequences, and determine the laws that regulate the probability of discarding the true sequence.

Keywords: jump Markov linear systems, hybrid systems, Riccati equations, conditioned Kalman filtering, fault detection, hypothesis testing.

Sommario

Il presente lavoro tratta problemi di stima per sistemi lineari soggetti a salti di tipo Markoviano. Si introducono dapprima i cosiddetti *jump Markov linear systems*, ossia sistemi a tempo discreto descritti da una successione temporale non nota di modelli lineari assegnati. Ciascun modello è definito da un'equazione alle differenze stocastica che regola l'evoluzione di uno stato continuo x_k, e da una equazione statica che descrive misure rumorose y_k di x_k. Il modello attivo al tempo k è determinato da uno stato discreto q_k a valori in un insieme finito $\mathcal{Q}$. L'evoluzione di q_k segue le leggi di una catena di Markov a tempo discreto e determina salti fra i diversi modi lineari. Si considera il problema della stima Bayesiana di x_k e q_k a partire dalla collezione di misure y^k disponibili all'istante k. Se ne deriva la soluzione ottima e si dimostra che la complessità di tale soluzione cresce esponenzialmente nel tempo. Si illustrano noti algoritmi di stima approssimata a complessità finita. Si discutono i problemi aperti nella stima subottima ed i limiti del modello a tempo discreto. Si considera quindi un analogo modello per sistemi con dinamica a tempo continuo. In tale modello, un'equazione differenziale stocastica governa l'evoluzione di uno stato continuo $\xi(t)$, mentre una equazione statica descrive misure rumorose y_k effettuate in istanti temporali prefissati. I salti fra i vari stati discreti seguono la dinamica di una catena di Markov a tempo continuo $q(t)$ a valori in $\mathcal{Q}$. Si definisce il problema della stima di $\xi(t)$ e di $q(t)$ basata sulla collezione di misure y^k e se ne formula la soluzione ottima in termini astratti. Nei casi di filtraggio e predizione, si discutono l'applicazione di filtri di Kalman condizionati ed un algoritmo di stima di tipo ricorsivo. Si approfondisce lo studio del problema di stima dello stato nel caso specifico di un sistema con un singolo salto, e se ne dimostra l'equivalenza ad un problema di *fault detection*, ossia di rilevazione di guasti. Si costruiscono algoritmi di *fault detection* e se ne valutano le prestazioni sulla base di simulazioni numeriche. Infine, si riconsiderano i modelli a tempo discreto e si affronta il problema della stima dell'intera traiettoria $q_0, \ldots, q_k$ dello stato discreto basata sulle misure y^k. Si introduce un test di tipo sequenziale per la determinazione di un sottoinsieme di traiettorie più probabili. Si studiano gli effetti delle proprietà strutturali del modello sull'efficacia del test nel distinguere sequenze differenti, e si determinano le leggi che regolano la probabilità di scartare la sequenza corretta.

Parole chiave: jump Markov linear systems, sistemi ibridi, equazioni di Riccati, filtri di Kalman condizionati, fault detection, test di ipotesi.

Contents

Introduction and preview **11**

1 Statistical background **17**
1.1 Discrete-time Markov chains 17
1.2 Continuous-time Markov chains 19
1.3 Poisson processes 22
1.4 Uniform chains 22
1.5 Case study 24
1.6 Bayesian estimation 26
1.7 Kalman filtering 29
1.8 Hypothesis testing 31
1.9 Sequential hypothesis testing 33

2 Jump Markov linear systems **37**
2.1 Model statement 37
2.2 State estimation problem 39
2.3 Recursive solution 42
2.4 Approximate estimation methods 45
2.5 Discussion 50

3 Superposition principle in linear filtering **53**
3.1 Introduction 53
3.2 Preliminaries 54
3.3 Difference Riccati and linear equations 55
3.4 Decomposition formulas 58
3.5 Discussion 62

4 Sampled linear switching systems **63**
4.1 Introduction 63
4.2 General model 64
4.3 State estimation problem 66

4.4 Theoretical solution 67
4.5 Conditioned filtering and prediction 68
4.6 Recursive approach 69
4.7 Single switch model 71
4.8 State estimation and fault detection 74
4.9 Discussion 77

5 Fault detection algorithms **79**
5.1 Introduction 79
5.2 Parametric discretization 81
5.3 Conditioned Kalman filtering revisited 86
5.4 Implementation 89
5.5 Low-complexity approximations 92
5.6 Numerical results 95
5.7 Discussion 102

6 Mode estimation by hypothesis testing **105**
6.1 Introduction 105
6.2 Trajectory selection 107
6.3 Bhattacharyya product 110
6.4 Stopping time of sequential testing 112
6.5 Accumulation of errors 115
6.6 Discussion 118

Perspectives **121**

Acknowledgements **123**

Bibliography **125**

List of symbols

$\mathbb{R}$	Real numbers
$\mathbb{R}_+$	Nonnegative real numbers
$\mathbb{N}$	Positive integers
$\mathbb{N}_0$	Nonnegative integers
$\mathcal{L}_1$	Space of integrable functions
$\mathcal{L}_2$	Space of square-integrable functions
$\langle \cdot, \cdot \rangle$	$\mathcal{L}_2$-Hilbert product
$\|\| \cdot \|\|$	$\mathcal{L}_2$-Hilbert norm
$(\cdot)_{i,j}$	Matrix element in row i and column j
$\|\mathcal{A}\|$	Cardinality of set $\mathcal{A}$
$\mathcal{A}^c$	Complement of set $\mathcal{A}$
$\mathbb{1}_{\mathcal{A}}(\cdot)$	Indicator function of set $\mathcal{A}$
$\mathbb{P}[\cdot]$	Probability (measure)
$p(\cdot)$	Probability mass function
$\mathcal{F}(\cdot)$	Probability distribution function
$f(\cdot)$	Probability density function
$\mathbb{E}[\cdot]$	Expectation
$\mathcal{N}(\mu, \Sigma)$	Gaussian density function with mean μ and covariance matrix Σ
$\mathcal{E}(\lambda)$	Unilateral exponential density function of parameter λ
$\chi^2(p)$	Chi-squared density function with p degrees of freedom

Introduction and preview

Diverse applications of automated estimation and control deal with systems subject to sudden changes. Air traffic management requires monitoring the actions of aircrafts to guarantee safe flight conditions. In this context, the motion of an aircraft may be split into take off, cruising, landing and other manoeuvres [30, 43]. Industrial plants undergo ruptures that modify the dynamics of the process. Interest is in monitoring and detecting faults for fast repair or reconfiguration [6]. Automotive power train control requires the online identification of the transmission parameters of a car. They change in time according to the gear engaged and the clutch connection state [4]. Cardiac activity is subject to different types of arrhythmias, which need to be recognized based on electrocardiogram data [54]. The flow of information in a networked system is regulated by communication protocols, and congestion or dead-lock events may be avoided by suitable control [31]. Human motion and gaits may by tracked and classified by a computer vision system provided the learning of a set of typical dynamics [9, 50]. In these settings, continuous time variables characterizing the evolution of the system are combined with quantities and events that are discrete in nature. This results in a so-called *hybrid* system, a generic term used to indicate a vast and heterogeneous class of systems with discrete and continuous components.

In a number of settings, jumps are not determined by the evolution of the system, or the relationship between jumps and continuous dynamics cannot be explained by a model of reasonable complexity. In other words, the sequence of modes the system goes through follows its own *stochastic* dynamics. Since the late 60's, the engineering community has been interested in *jump Markov linear systems* (JMLS). A jump Markov linear system consists of a linear Gaussian stochastic difference equation for the evolution of a continuous state x_k, of a static linear equation describing measurements y_k of x_k corrupted by Gaussian noise, and of an N-valued discrete state q_k that determines the parameters of the two equations at each time step k. For fixed values of q_k, one gets a stochastic linear state-space model. However, the

value of q_k changes in time according to the laws of a *discrete-time Markov chain*, which makes the system *nonlinear*. Given a jump Markov model, one is concerned with the recursive estimation of the continuous state x and of the mode of the system, i.e. the discrete state q, from the collection of measurements $y_0, \dots, y_k$. Since N^{k+1} different sequences $q_0, \dots, q_k$ exist, optimal Bayesian filtering and prediction of x may be obtained by a weighted average of N^{k+1} *conditioned Kalman filters*, each matched to a different discrete state sequence. However, this solution is impracticable due the exponentially increasing complexity of the algorithm. A related problem is that of detecting the whole sequence $q_0, \dots, q_k$. As will be shown, the a posteriori statistics of the whole sequence may be determined by the same bank of Kalman filters used in the estimation of x.

Since the pioneering work of Ackerson and Fu [1], several suboptimal estimation strategies have been proposed trying to achieve good performance with finite-complexity algorithms. Successful applications were found in medicine and fault detection, see [54] and references therein, tracking of manoeuvering targets [46, 21, 23], signal detection and noise suppression [24], stochastic deconvolution [21], among others. However, it was already pointed out in [54] that performance analysis of the available algorithms is intractable. Moreover, numerical simulations show that performance depends heavily on the model considered, and no algorithm can be preferred in general. To date, there is general agreement on the fact that the theoretical understanding of the JMLS estimation problem is still loose. In fact, most algorithms in literature are somewhat arbitrary approximations of the optimal solution and have been evaluated on the basis of few numerical simulations.

Jump Markov linear systems may be used to approximate the dynamics of a continuous-time switching system. In this case, an implicit assumption is that the mode of the system does not change in-between measurement instants. This may be a serious limitation if the system modes or the switch dynamics are fast compared to the rate of measurements. In medical applications, for instance, measurements such as blood samples or radiographies may only be taken at reasonably sparse time instants. On the contrary, the evolution of a disease or the effects of a therapy should be monitored with as much detail as possible. In a remote control problem with communication constraints, few measurements are available for state estimation, therefore discrete-time approximations of dynamics and switching might be unacceptable [14]. Continuous-time stochastic hybrid systems have been investigated e.g. in [32, 35, 47, 58]. In this setting, a continuum of measurements is available for estimation purposes.

Motivations of the work

We are concerned with the study of continuous-time switching dynamical systems observed at sparse sample times. As we mentioned above, a discrete-time jump Markov model is a poor description of certain systems of actual interest. In particular, the problem of detecting the exact time of a switch cannot be posed unless the approximation to the closest sample time is accepted. On the other hand, no significant effort has been dedicated to model continuous-time systems undergoing Markovian switches. A large part of this thesis is dedicated to the generalization of the JMLS framework to systems with continuous-time dynamics and switching. We consider a state-space model where the continuous state ξ obeys a stochastic *differential* equation, and noisy measurements of ξ are taken at fixed sample times. The parameters of the two equations are determined by the $N-valued$ discrete state q, which is now assumed to jump in time according to a *continuous-time Markov chain*. Therefore, switches in-between measurements are explicitly accounted for. State estimation problems are considered together with the problem of determining the time of a jump. The continuous nature of switching makes estimation hard to treat in a parametric way. However, in the basic setting of a single-switch system, an effective solution is found by convenient application of conditioned Kalman filtering. This gives rise to optimal algorithms for state estimation and fault detection, approximations being limited to an adaptive numerical evaluation of a small number of finite-support integrals.

A second issue of our concern is the detection of the discrete-state sequence $q_0, \ldots, q_k$ of a JMLS. The problem raises questions of theoretical interest such as the distinguishability of the sequences on the basis of the output data y^k. The ultimate aim would be to determine the achievable performance on the basis of structural properties of the system such as rate of switching, observability of the modes, et cetera. At the same time, we wish to devise detection algorithms supported by a sound performance analysis. Literature lacks theoretical results on the JMLS detection problem. Recent works [44, 45] report encouraging results based on the use of the tools of information theory. Our effort is based on the application of hypothesis testing. Using maximum-a-posteriori criteria, we select a subset of most probable sequences and reduce estimation to an exhaustive search *within* this set. Based on the equivalence to a family of sequential likelihood ratio tests, we relate the performance in pruning wrong sequences to a measure of similarity between conditioned predictors. The results we obtain in this sense are preliminary. On the other hand, general results on the probability of discarding the true sequence shed light on the complexity of the selection task.

Thesis outline

Chapter 1 provides an overview of Markov chains and statistical estimation methods. The main properties of discrete-time and continuous-time Markov chains are exposed for later use in jump Markov modeling. Bayesian estimation is illustrated in its essential lines. Basic results of linear and nonlinear estimation are presented along with recursive linear filtering for state-space models. One section is dedicated to hypothesis testing. Fundamental concepts such as the Neyman-Pearson theorem and Wald's sequential testing are adapted to the context of jump Markov systems.

Chapter 2 introduces jump Markov models in discrete-time and the relevant state estimation problems. Optimal Bayesian solutions are derived in the form of an average of conditioned linear estimates and in terms of a recursion on the a posteriori statistics of the state. This shows the exponentially increasing complexity of the problem and reveals that the use of an optimal estimator is impracticable. The main suboptimal algorithms proposed in the literature are reviewed. A final discussion points out the open issues that motivate our research.

Chapter 3 describes an alternative formulation of the Kalman filtering algorithm. Estimates are split in a term associated to zero initial conditions and in a second term which is a parametric function of the initial conditions. Both terms are updated by simple matrix recursions. The result resembles the superposition principle holding for linear difference equations, and allows direct evaluation of the estimates for changing initial conditions at any step of the recursion. This will be fundamental in the development of fault detection algorithms for continuous-dynamics switching systems.

Chapter 4 introduces a natural generalization of the discrete-time jump Markov models to continuous-time dynamics and switching. A continuous-time Markov chain description of the discrete state is used. Statement and analysis of Bayesian estimation problems are developed in analogy with Chapter 2. Both the averaged and the recursive solutions are shown to be formally equivalent to the discrete-time counterparts, however, they are complicated by the continuous variety of discrete state trajectories. The basic instance of a single-switch system, well suited to fault detection applications, is examined in depth. State estimation is discussed and is shown to encompass the estimation of the whole discrete-state trajectory.

Chapter 5 concentrates on the implementation of the fault detection strategies developed in Chapter 4. We initially discuss the prerequisites of a numerical solution. Estimation formulas are modified accordingly based on

the application of the results of Chapter 3 to a conditioned discretization of the continuous-dynamics model. Optimal and suboptimal estimation algorithms are proposed. An implementation of the optimal estimator is tested by computer simulation. A qualitative analysis of the results is reported.

Chapter 6 treats the estimation of the discrete-state sequence of discrete-time jump Markov systems. We consider the use of sequential testing for isolating a set of sequences with high a posteriori probability. Maximum-a-posteriori estimation is reduced to a search within this set. We show that the selection rule is equivalent to a family of sequential likelihood ratio tests between two alternative sequences. We study the connection between the properties of the system modes and the performance in the isolation of the true sequence. We derive very general results on the probability of missing the true sequence. To conclude, we discuss the limitations of the method and possible remedies.

Main contributions

The reformulation of discrete-time Kalman filtering proposed in Chapter 3 is original. Besides its use in this thesis, a number of utilizations may be foreseen, including the application to extended Kalman filtering and smoothing. The work was submitted in the form of a journal paper [13] and is currently under review.

The sampled-measurement Markov switching model of Chapter 4 is not investigated in the literature. Along with the fault detection algorithms of Chapter 5, it seems to be the first attempt to extend JMLS to systems where estimation from sparse measurements cannot be tackled by trivial system discretization. The first results on state estimation and fault detection were published in the proceedings of two international conferences [17], [16]. A journal paper was recently submitted [15] and is currently under review.

Chapter 6 represents an attempt to answer open theoretical questions concerning JMLS detection problems. Perhaps the most interesting result is the study of the update map for the probability of missing the true discrete-state sequence. Although rather simple, it leads to debating the complexity of the detection problem and suggests possible directions of research. The study of JMLS detection problems constitutes our current research activity.

Chapter 1

Statistical background

This chapter provides an overview of Markov chains and stochastic estimation. The main statistical properties of continuous-time and discrete-time Markov chains are reviewed. In the continuous-time case, attention is paid to the analytical properties of the trajectories. The equivalence between continuous-time Markov chains and uniform chains, in which discrete Markovian jumps are subordinated to the occurrence of Poisson events, is illustrated. Bayesian estimation methods for continuous and discrete variables are briefly reviewed. Finally, the essential concepts of hypothesis testing are presented. Most of the material is taken from [11],[52] and [57].

1.1 Discrete-time Markov chains

Let $\{q_k\}$, $k \in \mathbb{N}_0$, be a sequence of random variables taking values in a finite set $\mathcal{Q}$.

Definition 1.1 *Process* $\{q_k\}$ *is said to be a* Markov chain *if, for any* $k \in \mathbb{N}_0$ *and any* $i_0, \ldots, i_{k-1}, i, j \in \mathcal{Q}$

$$\mathbb{P}[q_{k+1} = j | q_k = i, q_{k-1} = i_{k-1}, \ldots, q_0 = i_0] = \mathbb{P}[q_{k+1} = j | q_k = i].$$

If, in addition, the above probability is independent of k*, the Markov chain is said to be* homogeneous.

In the sequel, we will always consider homogeneous chains.

Definition 1.2 *The* transition probability matrix π *of a homogeneous Markov chain is defined as*

$$\pi_{i,j} \triangleq \mathbb{P}[q_{k+1} = j | q_k = i].$$

It is easily verified that π is a *stochastic matrix*, i.e.

$$\sum_{j \in \mathcal{Q}} \pi_{i,j} = 1$$

for all $i \in \mathcal{Q}$. Let p_k be the probability distribution of q_k,

$$p_k(i) \triangleq \mathbb{P}[q_k = i].$$

In general, p_k will be thought of as a column vector having $p_k(i)$ on its i-th row. We will call p_0 the *initial probability distribution* of the chain.

Proposition 1.1 *The probability distribution of a discrete-time Markov chain is determined by the transition probability matrix π and the initial distribution p_0. For any $k \in \mathbb{N}_0$ and any $i_0, \ldots, i_k \in \mathcal{Q}$ it holds that*

$$\mathbb{P}[q_0 = i_0, \ldots, q_k = i_k] = p_0(i_0)\pi_{i_0,i_1} \cdot \ldots \cdot \pi_{i_{k-1},i_k}.$$

Therefore, two discrete-time homogeneous Markov chains are probabilistically equivalent if they have the same initial distribution and the same transition probabilities.

Proposition 1.2 *For any $k, h \in \mathbb{N}_0$ and any $i, j \in \mathcal{Q}$ it holds that*

$$\mathbb{P}[q_{k+h} = j | q_k = i] = (\pi^h)_{i,j}.$$

Corollary 1.1 *For any $k, h \in \mathbb{N}_0$, it holds that*

$$p_{k+h}^T = p_k^T \pi^h.$$

In particular, $p_{k+1}^T = p_k^T \pi$.

Quantity q_k will be called the *state* of the chain at time k. With an abuse of terminology, the elements of $\mathcal{Q}$ will also be called the *states* of the chain.

Definition 1.3 *A state $i \in \mathcal{Q}$ is said to be* absorbing *if, for every $h \in \mathbb{N}_0$,*

$$\mathbb{P}[q_{k+h} = i | q_k = i] = 1.$$

Proposition 1.3 *A state $i \in \mathcal{Q}$ is absorbing if and only if*

$$\pi_{i,i} = 1.$$

Observe that, if i is absorbing, $\pi_{i,j} = 0$ for all $j \neq i$.

1.2 Continuous-time Markov chains

Let $q(t)$, $t \in \mathbb{R}_+$, be a continuous-time stochastic process. Let $q(t) in \mathcal{Q}$, where $\mathcal{Q}$ is finite set.

Definition 1.4 *Process $q(\cdot)$ is said to be a* Markov chain *if, for any $t, \delta \in \mathbb{R}_+$, any $k \in \mathbb{N}_0$, any $s_0, \ldots, s_k$ such that $0 \leq s_0 < \ldots s_\ell < s_{\ell+1} \ldots < s_k < t$ and any $i, j, i_0, \ldots, i_k \in \mathcal{Q}$ it holds that*

$$\mathbb{P}[q(t+\delta) = j | q(t) = i, q(s_k) = i_k, \ldots, q(s_0) = i_0] = \mathbb{P}[q(t+\delta) = j | q(t) = i].$$

If, in addition, the above probability is independent of t, the Markov chain is said to be homogeneous.

It this work, we will consider homogeneous Markov chains only.

Definition 1.5 *The* transition probability function $\mathbf{T}(\delta)$, $\delta \in \mathbb{R}_+$, *of a continuous-time Markov chain is defined as*

$$\mathbf{T}_{i,j}(\delta) \triangleq \mathbb{P}[q(t+\delta) = j | q(t) = i].$$

One may verify that $\mathbf{T}(\delta)$ is a stochastic matrix for every value of δ.

Proposition 1.4 *It holds that*

$$\begin{aligned} \mathbf{T}(0) &= I \\ \mathbf{T}(t+s) &= \mathbf{T}(t)\mathbf{T}(s) \end{aligned}$$

for every $t, s \in \mathbb{R}_+$.

In light of these properties, matrix function $\mathbf{T}(\delta)$ is also called a *transition semigroup* on $\mathcal{Q}$. Let p_t denote the probability distribution of $q(t)$,

$$p_t(i) \triangleq \mathbb{P}[q(t) = i].$$

As usual, we will call p_0 the initial probability distribution of the chain, and we will interpret p_t as a column vector of functions of t.

Proposition 1.5 *For any $t, \delta \in \mathbb{R}_+$,it holds that*

$$p_{t+\delta}^T = p_t^T \mathbf{T}(\delta).$$

In particular, $p_t^T = p_0^T \mathbf{T}(t)$.

Proposition 1.6 *The probability distribution of a continuous-time Markov chain is determined by the transition probability function* $\mathbf{T}(\cdot)$ *and the initial distribution* p_0*. For any* $k \in \mathbb{N}_0$*, any* $i_0, \ldots, i_k \in \mathcal{Q}$ *and any* $s_0, \ldots, s_k$ *such that* $0 \leq s_0 < \ldots s_\ell < s_{\ell+1} \ldots < s_k$ *it holds that*

$$\mathbb{P}[q(s_0) = i_0, \ldots, q(s_k) = i_k] = p_0(i_0)\mathbf{T}_{i_0,i_1}(s_1 - s_0) \cdot \ldots \cdot \mathbf{T}_{i_{k-1},i_k}(s_k - s_{k-1}).$$

Therefore, two continuous-time homogeneous Markov chains are probabilistically equivalent if they have the same initial distribution and the same transition semigroup.

Definition 1.6 *A transition semigroup* $\mathbf{T}(\delta)$ *is said to be* continuous *if*

$$\lim_{\delta \to 0^+} \mathbf{T}(t + \delta) = \mathbf{T}(t)$$

for every $t \in \mathbb{R}_+$*, where the limit is evaluated elementwise.*

Proposition 1.7 *A transition semigroup* $\mathbf{T}(\delta)$ *is continuous if and only if it is continuous at the origin.*

For a continuous semigroup, one may consider local properties which do not have a discrete-time counterpart.

Definition 1.7 *The* infinitesimal generator *of a continuous transition semigroup is defined as*

$$\mathbf{G} \triangleq \lim_{\delta \to 0^+} \frac{\mathbf{T}(\delta) - I}{\delta},$$

i.e. it is the right-derivative of $\mathbf{T}(\delta)$ *at* 0*.*

It may be shown that $\mathbf{G}_{i,j} \in [0, +\infty)$ for every $i, j \in \mathcal{Q}$, $i \neq j$, and that

$$\mathbf{G}_{i,i} = -\sum_{j \neq i} \mathbf{G}_{i,j}.$$

Proposition 1.8 *If* $\mathbf{T}$ *is continuous, it holds that*

$$\frac{d}{dt}\mathbf{T}(t) \triangleq \lim_{\delta \to 0^+} \frac{\mathbf{T}(t + \delta) - T(t)}{\delta} = \mathbf{G}\mathbf{T}(t).$$

Corollary 1.2 *If* $\mathbf{T}$ *is continuous, it holds that*

$$\mathbf{T}(t) = e^{\mathbf{G}t}. \tag{1.1}$$

We shall now complete the analogy with the discrete-time chains.

Definition 1.8 *A state $i \in \mathcal{Q}$ is said to be* absorbing *if*

$$\mathbb{P}[q(t+\delta) = i | q(t) = i] = 1$$

for all $t, \delta \in \mathbb{R}_+$.

Hence, if i is an absorbing state, $\mathbf{T}_{i,i}(\delta) = 1$. As a consequence, $\mathbf{T}_{i,j}(\delta) = 0$ for all $j \neq i$.

Proposition 1.9 *If i is an absorbing state, then*

$$\mathbf{G}_{i,j} = 0$$

for all $j \in \mathcal{Q}$.

So far, we were concerned with the statistical properties of continuous-time Markov chains. Let us now consider the analytical properties of the trajectories of the process. We will write $q(t, \omega)$ to make the dependency on the event ω explicit.

Definition 1.9 *A (generic) continuous-time process $q(t, \omega)$ is a* jump process *if, for almost every ω and all $t \geq 0$,*

$$q(t+\delta, \omega) = q(t, \omega), \quad \delta \in [0, \varepsilon)$$

for some $\varepsilon = \varepsilon(t, \omega) > 0$.

Let us denote with $\{\tau_k(\omega)\}$, with $k \in \mathbb{N}_0$, $\tau_k \geq 0$ and $\tau_k < \tau_{k+1}$, the set of discontinuities of $q(t, \omega)$. Furthermore, let $\mathcal{D}_T(\omega) \triangleq \{\tau_k(\omega)\} \cap [0, T]$, $T \in \mathbb{R}_+$, denote the discontinuities falling within the interval $[0, T]$.

Definition 1.10 *A jump process is* regular *if*

$$|\mathcal{D}_T(\omega)| < +\infty$$

for almost every ω and all $T \in \mathbb{R}_+$.

In the above definitions, no restriction is imposed on the codomain $\mathcal{Q}$ of $q(t, \omega)$.

Proposition 1.10 *Let $q(t, \omega)$ be a continuous-time Markov chain with continuous transition semigroup $\mathbf{T}(\delta)$. Then $q(t, \omega)$ is a regular jump process. For almost every ω, $q(t, \omega)$ is a right-continuous function.*

Therefore, the outcome of a continuous-time Markov chain is, with probability 1, a piecewise constant, right-continuous trajectory with a finite number of discontinuities in any finite time.

1.3 Poisson processes

Let $\{\tau_k\}$, $k \in \mathbb{N}_0$, be a stochastic sequence such that, with probability one, $\tau_0 = 0$ and $\tau_k \leq \tau_{k+1}$ for all k.

Definition 1.11 *For $a, b \in \mathbb{R}_+$, $a \leq b$, the* counter *process of $\{\tau_k\}$ is*

$$c_{(a,b]} \triangleq \sum_{k \in \mathbb{N}_0} \mathbb{1}_{(a,b]}(\tau_k),$$

where $\mathbb{1}_{(a,b]}(\cdot)$ is the indicator function of the set $(a, b]$.

Definition 1.12 *The stochastic sequence $\{\tau_k\}$ is a* Poisson process *if:*

i. for every $k \in \mathbb{N}_0$ and every $\{a_\ell\} \subset \mathbb{R}_+$ such that $a_\ell \leq a_{\ell+1}$, the random variables $\{c_{(a_\ell, a_{\ell+1}]}\}$, $\ell = 0, \ldots, k-1$, are independent;

ii. for a given $\nu > 0$, the random variable $c_{(a,b]}$ has distribution of Poisson of parameter $\nu(b-a)$, i.e., for every $k \in \mathbb{N}_0$,

$$\mathbb{P}[c_{(a,b]} = k] = e^{-\nu(b-a)} \frac{\nu^k (b-a)^k}{k!}.$$

Note that $\nu(b-a)$ is the mean of the Poisson distribution. Therefore, ν represents the average density of the events τ_k, and will be called the rate of the Poisson process. Let now $c(t) \triangleq c_{(0,t]}$. Because the interval is closed on the right, the trajectories of $c(t)$ are right-continuous.

Proposition 1.11 *$c(t)$ is a regular jump homogeneous Markov process.*

1.4 Uniform chains

Consider a Poisson process $\{\tau_k\}$, $k \in \mathbb{N}_0$, of rate ν and a discrete-time Markov chain q_ℓ, $\ell \in \mathbb{N}_0$, with $q_\ell \in \mathcal{Q}$ and $\mathcal{Q}$ finite, having transition probability matrix π and initial distribution p_0. Assume that $\{\tau_k\}$ and $\{q_\ell\}$ are independent.

Definition 1.13 *A* uniform chain *$q(t)$, $t \in \mathbb{R}_+$, is defined as $q(t) \triangleq q_{c(t)}$, where $c(t)$ is the counter process of τ_k.*

According to this definition, process $q(t)$ takes values in $\mathcal{Q}$ and is allowed to switch in correspondence of the arrivals of a Poisson process. Switching itself follows the laws of a discrete-time Markov chain, which is called the *subordinated chain*.

Proposition 1.12 *Every uniform chain is a continuous-time homogeneous Markov chain with transition semigroup given by*

$$\mathbf{T}(\delta) = e^{-\nu\delta} \sum_{k=0}^{+\infty} \frac{(\nu\delta)^k}{k!} \pi^k. \tag{1.2}$$

Proof: ([38], pp.173) Recall that $\mathbb{P}[c_{(t,t+\delta]} = k] = e^{-\nu\delta}(\nu\delta)^k/k!$. Then, for any $\{s_0, \ldots, s_\ell\} \subset [0, t)$ with $s_0 < s_1 < \ldots < s_\ell$,

$$\begin{aligned}
&\mathbb{P}[q(t+\delta) = j | q(t) = i, q(s_0), \ldots, q(s_\ell)] = \\
&\quad = \sum_{k=0}^{+\infty} \frac{(\nu\delta)^k}{k!} e^{-\nu\delta} \mathbb{P}[q(t+\delta) = j | q(t) = i, q(s_0), \ldots, q(s_\ell), c_{(t,t+\delta]} = k] \\
&\quad = \sum_{k=0}^{+\infty} \frac{(\nu\delta)^k}{k!} e^{-\nu\delta} \mathbb{P}[q_{c(t+\delta)} = j | q_{c(t)} = i, c_{(t,t+\delta]} = k]
\end{aligned}$$

because q_k is Markov. Therefore one gets

$$\begin{aligned}
&\mathbb{P}[q(t+\delta) = j | q(t) = i, q(s_0), \ldots, q(s_\ell)] = \mathbb{P}[q(t+\delta) = j | q(t) = i] = \\
&\qquad = \sum_{k=0}^{+\infty} \frac{(\nu\delta)^k}{k!} e^{-\nu\delta} \mathbb{P}[q_{c(t)+k} = j | q_{c(t)} = i] = \sum_{k=0}^{+\infty} \frac{(\nu\delta)^k}{k!} e^{-\nu\delta} (\pi^k)_{i,j}.
\end{aligned}$$

□

Equation (1.2) may be written concisely as $\mathbf{T}(\delta) = e^{-\nu\delta} e^{\nu\delta\pi}$. In general, this expression may be computed by means of the Jordan canonical form of π [27]. Note that the initial distribution of $q(t)$ is the same as the initial distribution of q_k, namely, p_0.

Proposition 1.13 *The infinitesimal generator of a uniform chain is*

$$\mathbf{G} = \nu(\pi - I). \tag{1.3}$$

Proof: It holds that

$$\mathbf{G} = \left[\frac{d}{dt}\mathbf{T}(t)\right]_{t=0} = \left[\frac{d}{dt} e^{-\nu t} e^{\nu t \pi}\right]_{t=0} = -\nu\mathbf{T}(0) + \nu\pi\mathbf{T}(0),$$

hence the result. □

Proposition 1.14 *Every continuous-time Markov chain having continuous transition semigroup is probabilistically equivalent to a uniform Markov chain.*

Proof: Consider a Markov chain with transition semigroup $\mathbf{T}(\delta)$ and initial probability p_0. Since $\mathbf{T}(\delta)$ is continuous, one may equivalently consider the infinitesimal generator $\mathbf{G}$. Let q_k be a Markov chain with initial probability p_0. Let τ_k be a Poisson process of rate ν. Choose ν and the transition probability matrix π of q_k so that, for every $i, j \in \mathcal{Q}$, $j \neq i$,

$$\begin{aligned} \mathbf{G}_{i,j} &= \nu \pi_{i,j}, \\ \mathbf{G}_{i,i} &= \nu(\pi_{i,i} - 1). \end{aligned}$$

Then $q(t) = q_{c(t)}$ has initial probability p_0 and infinitesimal generator $\mathbf{G}$. In fact, one may choose any value of ν such that $\nu \geq -\min_{i \in \mathcal{Q}} \mathbf{G}_{i,i}$, and define $\pi_{i,i} \triangleq \mathbf{G}_{i,i}/\nu + 1$ and $\pi_{i,j} \triangleq \mathbf{G}_{i,j}/\nu$. □

Continuous-time Markov chains and uniform chains are therefore equivalent. Last proof shows that the interaction between the Poisson process and the subordinated Markov chain introduces redundancy. The ambiguity is removed by fixing the value of ν. The *minimal* realization of a continuous-time Markov chain is obtained by choosing $\nu = -\min_{i \in \mathcal{Q}} \mathbf{G}_{i,i}$, in which case $q(t)$ is forced to jump away from the state $\arg\min_{i \in \mathcal{Q}} \mathbf{G}_{i,i}$ at the first Poisson event.

1.5 Case study

In this section we consider a very simple case of continuous-time Markov chain which will be of use later on. Let $\mathcal{Q} = \{0, \ldots, N-1\}$, and consider a discrete-time Markov chain q_k taking values in $\mathcal{Q}$. Let the states $\mathcal{Q} \setminus \{0\}$ be absorbing. Then, the transition probability matrix of q_k has the form

$$\pi = \left[\begin{array}{c|c} \pi_{0,0} & \pi_{0,1} \ \cdots \ \pi_{0,N-1} \\ \hline 0 & \\ \vdots & I \\ 0 & \end{array}\right]. \tag{1.4}$$

Proposition 1.15 *It holds that*

$$\pi^k = \left[\begin{array}{c|c} (\pi_{0,0})^k & \cdots\ \pi_{0,j}\sum_{\ell=0}^{k-1}(\pi_{0,0})^\ell\ \cdots \\ \hline 0 & \\ \vdots & I \\ 0 & \end{array}\right]. \tag{1.5}$$

Proof: By induction on k. □

Next, consider the uniform chain $q(t) \triangleq q_{c(t)}$, where $c(t)$ is the counter of a Poisson process of rate ν.

Corollary 1.3 *Let π be as in (1.4). The transition semigroup of $q(t)$ is*

$$\mathbf{T}(\delta) = \left[\begin{array}{c|c} \mathbf{T}_{0,0}(\delta) & \mathbf{T}_{0,1}(\delta)\ \cdots\ \mathbf{T}_{0,N-1}(\delta) \\ \hline 0 & \\ \vdots & I \\ 0 & \end{array}\right],$$

where

$$\mathbf{T}_{0,j}(\delta) = \begin{cases} e^{-\nu\delta(1-\pi_{0,0})}, & j = 0; \\ \dfrac{\pi_{0,j}}{1-\pi_{0,0}}(1 - e^{-\nu\delta(1-\pi_{0,0})}), & j = 1, \ldots, N-1. \end{cases}$$

In particular, $\mathcal{Q} \setminus \{0\}$ is a set of absorbing states of $q(t)$.

Proof: For every $i, j \in \mathcal{Q}$ let us compute the expression

$$\mathbf{T}_{i,j}(\delta) = e^{-\nu\delta}\sum_{k=0}^{+\infty}\frac{(\nu\delta)^k}{k!}(\pi^k)_{i,j},$$

with π^k given by (1.5). For $i = j = 0$,

$$\mathbf{T}_{0,0}(\delta) = e^{-\nu\delta}\sum_{k=0}^{+\infty}\frac{(\nu\delta\pi_{0,0})^k}{k!} = e^{-\nu\delta}e^{\nu\delta\pi_{0,0}} = e^{-\nu\delta(1-\pi_{0,0})}.$$

For $i = j > 0$,

$$\mathbf{T}_{i,i}(\delta) = e^{-\nu\delta}\sum_{k=0}^{+\infty}\frac{(\nu\delta)^k}{k!} = e^{-\nu\delta}e^{\nu\delta} = 1.$$

For $i \neq j$, $i > 0$,

$$\mathbf{T}_{i,j}(\delta) = e^{-\nu\delta} \sum_{k=0}^{+\infty} \frac{(\nu\delta)^k}{k!} \cdot 0 = 0.$$

For $i = 0$, $j > 0$,

$$\begin{aligned}
\mathbf{T}_{0,j}(\delta) &= e^{-\nu\delta} \sum_{k=0}^{+\infty} \frac{(\nu\delta)^k}{k!} \pi_{0,j} \frac{(\pi_{0,0})^k - 1}{\pi_{0,0} - 1} \\
&= e^{-\nu\delta} \frac{\pi_{0,j}}{\pi_{0,0} - 1} \left(\sum_{k=0}^{+\infty} \frac{(\nu\delta\pi_{0,0})^k}{k!} - \sum_{k=0}^{+\infty} \frac{(\nu\delta)^k}{k!} \right) \\
&= e^{-\nu\delta} \frac{\pi_{0,j}}{\pi_{0,0} - 1} (e^{\nu\delta\pi_{0,0}} - e^{\nu\delta}) \\
&= \frac{\pi_{0,j}}{\pi_{0,0} - 1} (e^{-\nu\delta(1-\pi_{0,0})} - 1),
\end{aligned}$$

where the identity

$$\sum_{\ell=0}^{k-1} (\pi_{0,0})^\ell = \frac{(\pi_{0,0})^k - 1}{\pi_{0,0} - 1}$$

has been used. Alternative, one may compute (1.1) with $\mathbf{G}$ given by (1.3). □

1.6 Bayesian estimation

Let x and y be *continuous* random vectors. Consider the problem of estimating the value of x from the outcome of y. In this context, y is referred to as *measurements* (of x) or *data*. Define $X \triangleq \mathbb{R}^n$.

Definition 1.14 *Let $x \in X$. An* estimator z *of x is any measurable function of the data of the form*

$$z : y \mapsto z(y) \in X.$$

The quality of an estimate may be evaluated in terms of the cost of the error $x - z(y)$.

Definition 1.15 *A* cost function $c : X \to \mathbb{R}$ *is any measurable nonnegative convex function of the form*

$$c(x - z)$$

with $c(0) = 0$.

A natural way to define an estimator is to require that it minimizes the expected error given the available measurements. Let $f_{x|y}$ be a posteriori probability density of x given y.

Definition 1.16 *Given a cost function c, the conditioned expected risk of an estimator z given y is defined as*

$$\mathcal{R}(z, y) \triangleq \mathbb{E}[c(x - z)|y],$$

where the expectation is taken w.r.t. $f_{x|y}$ *and is assumed to exist.*

Consider the following optimization problem:

$$\mathcal{P}_c : \quad \min_z \mathcal{R}(z, y).$$

Definition 1.17 *A function z solving problem* $\mathcal{P}_c$ *will be called an* optimal Bayesian estimator *of x w.r.t. the cost function c.*

In fact, the optimal Bayesian estimator minimizes the unconditioned expectation $\mathbb{E}[c(x - z(y))]$ as well, see [36]. In this work we will always consider the *quadratic* cost function

$$c(x - z) \triangleq ||x - z||^2. \tag{1.6}$$

Therefore, the optimal Bayesian estimator of x will be called the *minimum mean-squared error* estimator.

Proposition 1.16 *Let c be defined as in (1.6). For an* arbitrary *density* $f_{x|y}$*, the optimal Bayesian estimator of x is given by*

$$\hat{x} = \mathbb{E}[x|y],$$

provided the expectation exists.

A proof of this result may be found e.g. in [52]. Note that in this case

$$\mathcal{R}(\hat{x}|y) = \mathbb{E}[||x - \mathbb{E}[x|y]||^2|y],$$

that is, the risk function coincides with the variance of x given y.

Assume that the random vectors x and y are jointly Gaussian, that is, $f(x, y) = \mathcal{N}(\mu, \Sigma)$. Consider the partitioning

$$\mu = \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix} \qquad \Sigma = \begin{bmatrix} \Sigma_x & \Sigma_{xy} \\ \Sigma_{xy}^T & \Sigma_y \end{bmatrix},$$

with obvious meaning of the symbols. For estimation purposes, we may assume that $\Sigma_y > 0$ without loss of generality [52].

Proposition 1.17 *The minimum-mean-squared-error estimator of x given y and the associated conditioned error covariance matrix are as follows:*

$$\hat{x} = \mu_x + \Sigma_{xy}\Sigma_y^{-1}(y - \mu_y),$$
$$Var(\tilde{x}|y) = \Sigma_x - \Sigma_{xy}\Sigma_y^{-1}\Sigma_{xy}^T,$$

where $\tilde{x}$ denotes the estimation error $x - \hat{x}$.

In this case, the optimal Bayesian estimator of x is a *linear* function of y. Moreover, the error covariance matrix is *independent of the data*, as it depends on the joint statistics of x and y only.

Let now q denote a *discrete* random variable, i.e. q takes values in a countable set $\mathcal{Q}$. In particular, we will be concerned with situations where $\mathcal{Q}$ is finite. Assume that we want to estimate the value of q from the measured data y.

Definition 1.18 *An estimator of q is any measurable function*

$$z : y \mapsto z(y) \in \mathcal{Q}.$$

Since q is a discrete variable, it is natural to construct an estimator by ensuring that the probability of an error $z(y) \neq q$ is minimized for the given data y. That is, the following optimization problem is considered:

$$\mathcal{P}_d : \quad \min_z \mathbb{P}[q \neq z(y)|y].$$

Definition 1.19 *A function z solving problem $\mathcal{P}_d$ will be called a* minimum-probability-of-error *estimator of q.*

Let $p_{q|y}$ be the a posteriori probability of q given y. Then we have the following result [56].

Proposition 1.18 *The minimum-probability-of-error estimator of q is*

$$\hat{q} = \arg\max_z p_{q|y}(z|y).$$

That is, the optimal Bayesian estimator of q is the *maximum-a-posteriori estimator*. Again, it is easily verified that $z(y) = \hat{q}$ also minimizes the unconditioned probability $\mathbb{P}[q \neq z(y)]$.

Minimum-probability-of-error estimation fits perfectly into the optimal Bayesian estimation framework that we illustrated for continuous random variables. Replace X with Q, x with q and define the cost function $c(q-z) \triangleq \mathbb{1}_{\{0\}^c}(q-z)$, where "$\{0\}^c$" denotes values different from zeros. Then one obtains

$$\mathcal{R}(z,y) = \mathbb{E}[\mathbb{1}_{\{0\}^c}(q-z)|y] = \mathbb{P}[q \neq z|y],$$

i.e. problem $\mathcal{P}_c$ specializes to $\mathcal{P}_d$. However, care must be taken in the definition of the notion of convexity (of the cost function) over discrete sets.

1.7 Kalman filtering

For $k \in \mathbb{N}_0$ and assigned matrix parameters $\{A_k, B_k, C_k, D_k\}$, $\hat{x}_0$, P_0, consider the discrete-time linear state-space model

$$\begin{cases} x_{k+1} = A_k x_k + B_k u_k \\ y_k = C_k x_k + D_k v_k \end{cases}$$

where $x_k \in \mathbb{R}^n$, $y_k \in \mathbb{R}^p$, $u_k \in \mathbb{R}^m$, $v_k \in \mathbb{R}^r$, $x_0 \sim \mathcal{N}(\hat{x}_0, P_0)$, $\{u_k\}$ and $\{v_k\}$ are zero-mean normalized white Gaussian sequences, and $x_0, \{u_k\}, \{v_k\}$ are mutually uncorrelated. Since the system is linear, the joint distribution of x_ℓ and $y^k \triangleq \{y_0, \dots, y_k\}$ will be a multivariate Gaussian for every $k, \ell \in \mathbb{N}_0$. Consider the minimum-mean-squared error estimation of x_ℓ from the measurements y^k. Let

$$\begin{aligned} \hat{x}_{\ell|k} &\triangleq \mathbb{E}[x_\ell|y^k], \\ P_{\ell|k} &\triangleq \mathrm{Var}(\tilde{x}_{\ell|k}|y^k), \end{aligned}$$

where $\tilde{x}_{\ell|k}$ denotes the estimation error $x_\ell - \hat{x}_{\ell|k}$. Since x_ℓ and y^k are jointly Gaussian, $\hat{x}_{\ell|k}$ will be a linear function of y^k, and $P_{\ell|k}$ will be independent of y^k. For $\ell = k$ and $k+1$, their computation may be carried out iteratively. Let for simplicity $D_k D_k^T > 0$.

Theorem 1.1 (Kalman-Bucy filter) *The following two-step iteration on index k holds:*

i. Measurement update:

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + L_k[y_k - C_k \hat{x}_{k|k-1}], \tag{1.7a}$$

$$P_{k|k} = P_{k|k-1} - L_k C_k P_{k|k-1}, \tag{1.7b}$$

with gain *matrix* L_k *given by*

$$P_{k|k-1}C_k^T \left[D_k D_k^T + C_k P_{k|k-1} C_k^T\right]^{-1};$$

ii. Time update:

$$\hat{x}_{k+1|k} = A_k \hat{x}_{k|k}, \tag{1.8a}$$
$$P_{k+1|k} = A_k P_{k|k} A_k^T + B_k B_k^T. \tag{1.8b}$$

Iterations are initialized by $\hat{x}_{0|-1} = \hat{x}_0$ *and* $P_{0|-1} = P_0$.

For any $\ell > k$, $\hat{x}_{\ell|k}$ and $P_{\ell|k}$ are found by iterating the time update equations. Assuming $D_k D_k^T > 0$ guarantees that the matrix inverse in the expression of L_k is well-defined for all k. However, the assumption is not essential and may be circumvented, see [26].

Next, for $t \in \mathbb{R}_+$, assigned regular matrix functions $F(t), G(t)$, assigned matrix parameters $\{H_k, G_k\}$ and assigned sequence of measurement instants $\{t_k\} \subset \mathbb{R}_+$ such that

$$0 = t_0 < t_1 < \ldots < t_k < t_{k+1} < \ldots,$$

consider the continuous/discrete-time linear state-space model

$$\begin{cases} \dot{\xi} = F(t)\xi(t) + G(t)u(t) \\ y_k = H_k \xi(t_k) + K_k v_k \end{cases}.$$

Assume that $\xi(t) \in \mathbb{R}^n$, $u(t) \in \mathbb{R}^m$, $\xi(0) \sim \mathcal{N}(\hat{\xi}_0, \Sigma_0)$, $u(t)$ is a zero-mean normalized white Gaussian noise, and $\xi(0)$, $u(t)$, $\{v_k\}$ are mutually uncorrelated for all $t \in \mathbb{R}_+$. Let

$$\hat{\xi}_{t|k} \triangleq \mathbb{E}[\xi(t)|y^k],$$
$$\Sigma_{t|k} \triangleq \mathrm{Var}(\tilde{\xi}_{t|k}|y^k),$$

with $\tilde{\xi}_{t|k}$ being the estimation error $\xi(t) - \hat{\xi}_{t|k}$. The system is linear, therefore $\xi(t)$ and y^k are jointly Gaussian random variables. As a consequence, $\hat{\xi}_{t|k}$ will be a linear function of y^k, and $\Sigma_{t|k}$ will be independent of y^k. For $t \geq t_k$, their computation may be carried out as follows. It is assumed that $H_k H_k^T > 0$.

Proposition 1.19 *The following iteration on index* k *holds:*

i. Measurement update:

$$\begin{aligned}\hat{\xi}_{t_k|k} &= \hat{\xi}_{t_k|k-1} + L_k[y_k - H_k\hat{\xi}_{t_k|k-1}],\\ \Sigma_{t_k|k} &= \Sigma_{t_k|k-1} - L_k H_k \Sigma_{t_k|k-1},\end{aligned}$$

with

$$L_k = \Sigma_{t_k|k-1}H_k^T[H_k\Sigma_{t_k|k-1}H_k^T + K_kK_k^T]^{-1};$$

ii. Time update: solve, for $t \geq t_k$, the differential equations

$$\begin{aligned}d\hat{\xi}_{t|k}/dt &= F(t)\hat{\xi}_{t|k},\\ d\Sigma_{t|k}/dt &= F(t)\Sigma_{t|k} + \Sigma_{t|k}F(t)^T + G(t)G(t)^T.\end{aligned}$$

Iterations are initialized by $\hat{\xi}_{t_0|-1} = \hat{\xi}_0$ and $\Sigma_{t_0|-1} = \Sigma_0$.

1.8 Hypothesis testing

Consider a sequence of random variables $\{y_\ell\}$, $y_\ell \in \mathbb{R}^p$, $\ell \in \mathbb{N}_0$, distributed according to some joint density function f. Let f_i, $i = 0, \ldots, N-1$, be N assigned distributions of sequences in $\mathbb{R}^p$. That is, for $i = 0, \ldots N-1$, the density functions

$$f_i(y_0, \ldots, y_k)$$

are known for all k and satisfy the conditions of marginalization. Assume that $f \in \{f_0, \ldots, f_{N-1}\}$, i.e., there exists one index i such that

$$f_i(y_0, \ldots, y_k) = f(y_0, \ldots, y_k)$$

for all k and all values of $y_0, \ldots, y_k$. One wants to build a procedure to decide which distribution f_i is the true one given the outcome of the variables y_ℓ, which we will call data or measurements. The eventuality that $f = f_i$ will be called *hypothesis i* and will be indicated with $\mathcal{H}_i$.

Assume that a decision has to be taken based on the k measurements $y_0, \ldots, y_k$.

Definition 1.20 *A* test function *is a measurable function*

$$T: \quad (y_1, \ldots, y_k) \mapsto T(y_1, \ldots, y_k) \in \{0, \ldots, N-1\}.$$

For a given test function T, the procedure consists in applying T to the measured data and to *accept hypothesis* $\mathcal{H}_i$ if and only if $T(y_1, \ldots, y_k) = i$. Of course, different kinds of errors may occur.

Definition 1.21 *An* error of type i *occurs if* $T(y_1, \ldots, y_k) \neq i$ *when* $\mathcal{H}_i$ *holds. The probability of an error of type* i *is by definition*

$$\alpha_i \triangleq \mathbb{P}[T(y_1, \ldots, y_k) \neq i | \mathcal{H}_i].$$

Of course, the error probabilities α_i depend on the definition of the test function T, i.e. $\alpha_i = \alpha_i(T)$. The test function shall be chosen as to minimize certain error probabilities, or to keep them below a certain bound, in relation with the context of use.

Let now $N = 2$, i.e., only two different hypotheses are admissible. It is evident that one can make either α_0 or α_1 as small as desired by a suitable choice of T. However, for a fixed probability of the error of type 0, $\alpha_0 = \bar{\alpha}$, the probability of error α_1 shall not be smaller than a certain lower bound depending on the nature of the distributions f_0 and f_1 and on the number of measurements k. In general, the larger the k, the smaller the lowerbound. Denote with $\mathcal{C}_{\bar{\alpha}}$ the class of test functions yielding probability of error of type 0 equal to $\bar{p}$. Consider the *likelihood ratio*

$$r^k(y_0, \ldots, y_k) \triangleq \frac{f_1(y_0, \ldots, y_k)}{f_0(y_0, \ldots, y_k)}.$$

Definition 1.22 *The* likelihood ratio test *is defined by*

$$T_\kappa(y_0, \ldots, y_k) = \begin{cases} 0, & \text{if } r^k(y_0, \ldots, y_k) < \kappa \\ 1, & \text{if } r^k(y_0, \ldots, y_k) \geq \kappa \end{cases}$$

with threshold *parameter* $\kappa \in \mathbb{R}_+$.

It may be shown by continuity arguments that there must exist a value of $\kappa = \kappa(\bar{\alpha})$ for which the likelihood ratio test has probability of error $\alpha_0(T_{\kappa(\bar{\alpha})}) = \bar{\alpha}$.

Theorem 1.2 (Neyman-Pearson) *For any* $T \in \mathcal{C}_{\bar{\alpha}}$ *it holds that*

$$\alpha_1(T) \geq \alpha_1(T_{\kappa(\bar{\alpha})}).$$

By the symmetry of the problem, an identical result is obtained if α_1 is fixed instead of α_0. In this sense, the likelihood ratio test is optimal among the set of tests with one probability of error fixed.

Remark. In a Bayesian setting, one assumed that each hypothesis is assigned a prior probability, say, p_i. In this case, the optimal estimator of $\mathcal{H}_i$ is given by the maximum-a-posteriori estimator

$$\hat{q} = \arg\max_q \mathbb{P}[\mathcal{H}_q | y_0, \ldots, y_k],$$

where $q = 0, \ldots, N-1$. If $N = 2$, one may consider the ratio

$$\frac{\mathbb{P}[\mathcal{H}_1 | y_0, \ldots, y_k]}{\mathbb{P}[\mathcal{H}_0 | y_0, \ldots, y_k]} = r^k(y_0, \ldots, y_k) \frac{p_1}{p_0} \tag{1.9}$$

and reformulate $\hat{q}$ as

$$\hat{q} = \begin{cases} 0, & \text{if } r^k(y_0, \ldots, y_k) < p_0/p_1, \\ 1, & \text{if } r^k(y_0, \ldots, y_k) \geq p_0/p_1. \end{cases}$$

Therefore, the optimal Bayesian estimator is essentially a likelihood ratio test with threshold $\kappa = p_0/p_1$. The overall estimation error probability is given by

$$\alpha_0 p_0 + \ldots + \alpha_{N-1} p_{N-1},$$

for an arbitrary number of hypotheses N. □

1.9 Sequential hypothesis testing

Assume that the outcomes of y_ℓ, $\ell = 0, 1, \ldots$ are measured sequentially.

Definition 1.23 *A* sequential test function *at time ℓ is a measurable function*

$$T^\ell : (y_0, \ldots, y_\ell) \mapsto T^\ell(y_0, \ldots, y_\ell) \in \{0, \ldots, N-1, \epsilon\},$$

where ϵ is an arbitrary value different from $0, \ldots, N-1$.

Given a family of sequential test functions $\{T_\ell\}$, $\ell \in \mathbb{N}_0$, a *sequential test* is the following recursive algorithm. At time ℓ, apply T_ℓ to $y_0, \ldots, y_\ell$. Let $T_\ell(y_0, \ldots, y_\ell) = i$. If $i \in \{0, \ldots, N-1\}$, *accept hypothesis $\mathcal{H}_i$*. If instead $i = \epsilon$, then draw a new measurement $y_{\ell+1}$ and repeat the cycle at time $\ell + 1$.

Definition 1.24 *The* stopping time *of a test with sequential test functions $\{T^\ell\}$ is the random variable*

$$\bar{k} = \inf\{\ell : T^\ell(y_0, \ldots, y_\ell) \neq \epsilon\}.$$

The stopping time may be finite or infinite. In $\bar{k}$ is finite, we say that the *test terminates* at time $\bar{k}$, otherwise we say that the test *never terminates*. One wants to choose $\{T_\ell\}$ so to ensure that $\mathbb{P}[\bar{k} < +\infty] = 1$.

Definition 1.25 *An error of type i occurs if $T^{\bar{k}}(y_0, \ldots, y_{\bar{k}}) \neq i$ when $\mathcal{H}_i$ holds. Associated to it is the probability of error*

$$\alpha_i = \mathbb{P}[T^{\bar{k}}(y_0, \ldots, y_{\bar{k}}) \neq i | \mathcal{H}_i].$$

Observe that a non-sequential test with test function $T(y_0, \ldots, y_k)$ may always be put in the form of a sequential test: just set $T^\ell = \epsilon$, $\ell = 0, \ldots, k-1$ and $T^k \triangleq T$. Therefore, one may always construct a sequential test yielding probabilities of error equal to $\alpha_i(T)$, $i = 0, \ldots, N-1$, for which $\bar{k} \leq k$ with probability 1. However, one is interested in the expected stopping time $\mathbb{E}[\bar{k}]$. In most cases, it is possible to construct $\{T_\ell\}$ so that $\mathbb{E}[\bar{k}] < k$.

Now consider the case $N = 2$.

Definition 1.26 *The* sequential likelihood ratio *test is defined by*

$$T^\ell_{\kappa_0,\kappa_1}(y_0, \ldots, y_\ell) = \begin{cases} 0, & \text{if } r^\ell(y_0, \ldots, y_\ell) < \kappa_1 \\ 1, & \text{if } r^\ell(y_0, \ldots, y_\ell) \geq \kappa_0 \\ \epsilon, & \text{if } \kappa_1 \leq r^\ell(y_0, \ldots, y_\ell) < \kappa_0 \end{cases}$$

for all $\ell \in \mathbb{N}_0$ and parameters $\kappa_1, \kappa_0 \in \mathbb{R}_+$, $\kappa_1 \leq \kappa_0$.

In more generality, thresholds κ_0 and κ_1 may depend on ℓ. In particular, one may define a *truncated* test by setting $\kappa_1 = \kappa_0$ at a certain $\ell = k$, in which case $\bar{k} \leq k$ w.p.1. For fixed values of κ_0 and κ_1, there may be a nonzero probability that the test never terminates. A special case is the following.

Proposition 1.20 *Assume that $\{y_\ell\}$ are i.i.d. variables, i.e. $f(y_0, \ldots, y_k) = f(y_0) \cdot \ldots \cdot f(y_k)$ for every $k \in \mathbb{N}_0$. Let f_i, $i = 0, 1$ be such that $f_i(y_0, \ldots, y_k) = f_i(y_0) \cdot \ldots \cdot f_i(y_k)$, with $f_0 \neq f_1$ on a set of nonzero probability. Then, the probability that the sequential likelihood ratio test never terminates is zero.*

For arbitrary functions f and f_i let us define

$$\gamma_i(\kappa_0, \kappa_1) = \mathbb{P}[\bar{k} = +\infty | \mathcal{H}_i],$$

where $\bar{k} = +\infty$ stands for $T^\ell_{\kappa_0,\kappa_1}(y_0, \ldots, y_\ell) = \epsilon$, $\forall \ell \in \mathbb{N}_0$.

Proposition 1.21 *The following inequalities hold:*

$$\begin{aligned} \alpha_0\kappa_0 &\leq 1-\alpha_1-\gamma_1, \\ \alpha_1 &\leq \kappa_1(1-\alpha_0-\gamma_0). \end{aligned}$$

Proof: Following [57], it holds that

$$\begin{aligned} \mathbb{P}[\text{accept } \mathcal{H}_1|\mathcal{H}_1] &\geq \kappa_0\mathbb{P}[\text{accept } \mathcal{H}_1|\mathcal{H}_0], \\ \mathbb{P}[\text{accept } \mathcal{H}_0|\mathcal{H}_1] &\leq \kappa_1\mathbb{P}[\text{accept } \mathcal{H}_0|\mathcal{H}_0]. \end{aligned}$$

On the other hand, $\mathbb{P}[\text{accept } \mathcal{H}_1|\mathcal{H}_1]$ may be rewritten as

$$1-\mathbb{P}[(\text{accept } \mathcal{H}_1)^c|\mathcal{H}_1] = 1-\mathbb{P}[\text{accept } \mathcal{H}_0|\mathcal{H}_1]-\mathbb{P}[\bar{k}=+\infty|\mathcal{H}_1],$$

since the event that the test never terminates is disjoint from the event of accepting $\mathcal{H}_0$. Similarly, $\mathbb{P}[\text{accept } \mathcal{H}_0|\mathcal{H}_0]$ may be rewritten as

$$1-\mathbb{P}[(\text{accept } \mathcal{H}_0)^c|\mathcal{H}_0] = 1-\mathbb{P}[\text{accept } \mathcal{H}_1|\mathcal{H}_0]-\mathbb{P}[\bar{k}=+\infty|\mathcal{H}_0].$$

Recognizing the quantities α_i and γ_i yields the result. □

As a consequence, the probabilities of error α_0 and α_1 obey the bounds

$$\begin{aligned} \alpha_0 &\leq 1/\kappa_0, \\ \alpha_1 &\leq \kappa_1, \end{aligned}$$

regardless of the values of $\gamma_0(\kappa_0,\kappa_1)$ and $\gamma_1(\kappa_0,\kappa_1)$. Thus, arbitrarily small probabilities of error may be attained by a suitable choice of the parameters κ_0 and κ_1. If, moreover, $\gamma_0=\gamma_1=0$ for all choices of κ_0 and κ_1 – which is the case, for instance, if $\{y_\ell\}$ is an i.i.d. sequence – the test terminates w.p.1. It is conjectured in [57] that the sequential likelihood ratio test is optimal in the sense of minimizing $\mathbb{E}[\bar{k}]$ over all sequential tests with given error probabilities.

Remark. In the Bayesian setting discussed in Section 1.8, a sequential test procedure is set up as follows. At time ℓ, accept $\mathcal{H}_0$ if ratio (1.9) is below κ_0; accept $\mathcal{H}_1$ if (1.9) it exceeds κ_0; otherwise, repeat the procedure at time $\ell+1$. That is, the Bayesian test becomes the likelihood test

$$\kappa_0\frac{p_0}{p_1} \lesseqgtr r^\ell(y_0,\ldots,y_\ell) \lesseqgtr \kappa_1\frac{p_0}{p_1},$$

where the symbol "$\lessgtr$" indicates the comparison of the likelihood ratio with the thresholds. It follows that the overall probability of error is

$$\alpha_0 p_0 + \alpha_1 p_1 \leq p_1 \frac{1}{\kappa_0} + p_0 \kappa_1,$$

i.e. it can be made arbitrarily small by a proper choice of κ_0 and κ_1. A natural generalization of the sequential Bayesian test to an arbitrary *fixed* number of hypotheses N is presented in [7]. □

Chapter 2

Jump Markov linear systems

In this chapter, the problem of state estimation for discrete-time switching linear systems is considered. We introduce the class of jump Markov linear systems and the relevant state estimation problem. The optimal Bayesian solution is derived and is shown to depend on a mixing of an exponential number of Kalman filters. A recursive formulation of the solution is then presented. Based on this, currently available suboptimal algorithms are illustrated. Limits of the model and theoretical issues of our concern are finally discussed.

2.1 Model statement

For a certain index $N \in \mathbb{N}_0$, let $\mathcal{Q} = \{0, \dots, N-1\}$. Assume that we are given the matrix-valued functions

$$\begin{aligned} A &: \mathcal{Q} \to \mathbb{R}^{n \times n} &&: q \mapsto A_q \\ B &: \mathcal{Q} \to \mathbb{R}^{n \times m} &&: q \mapsto B_q \\ C &: \mathcal{Q} \to \mathbb{R}^{p \times n} &&: q \mapsto C_q \\ D &: \mathcal{Q} \to \mathbb{R}^{p \times r} &&: q \mapsto D_q \end{aligned}$$

with $n, m, p, r \in \mathbb{N}_0$. For $k \in \mathbb{N}_0$, let $u_k, v_k \in \mathbb{R}^m$ be zero-mean, normalized white Gaussian sequences, and let q_k be a sequence taking values in $\mathcal{Q}$. Consider the following discrete-time, time-varying model

$$\begin{cases} x_{k+1} = A_{q_k} x_k + B_{q_k} u_k \\ y_k = C_{q_k} x_k + D_{q_k} v_k \end{cases} \tag{2.1}$$

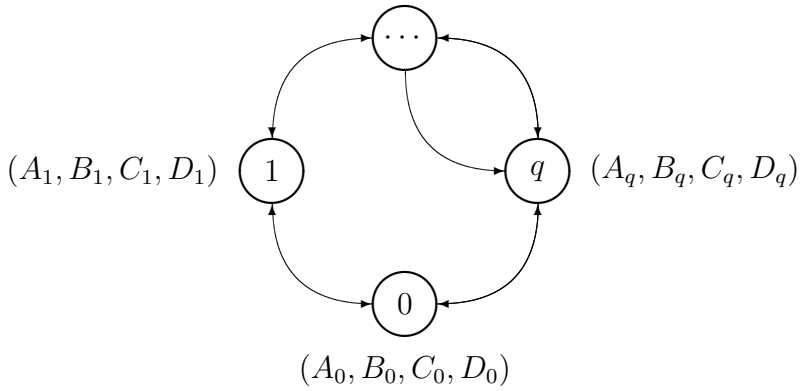


Figure 2.1: Switching system

with $x_0 \sim \mathcal{N}(\hat{x}_0, P_0)$, and $\hat{x}_0$, P_0 known. For a given sequence $\{q_k\}$, the first equation is a linear stochastic difference equation describing the evolution of x_k, and the second equation defines y_k as a linear noisy measurement of x_k. We will assume that $\{u_k\}$, $\{v_k\}$ and x_0 are mutually uncorrelated, and that $\mathrm{Var}(D_q v_k) = D_q D_q^T > 0$ for every $q \in \mathcal{Q}$.

At any time index k, the value of q_k defines the current mode of the system by fixing the parameters of the model. Since q_k is allowed to take different values at different times, equation (2.1) describes a switching system, as depicted in Figure 2.1

Let $\{q_k\}$ follow the laws of a discrete-time homogeneous Markov chain with known transition probability matrix π and initial probability distribution p_0. That is, for every $i, j \in \mathcal{Q}$, and any $k \in \mathbb{N}_0$,

$$\begin{aligned} \pi_{i,j} &= \mathbb{P}[q_{k+1} = j | q_k = i], \\ p_0(j) &= \mathbb{P}[q_0 = j]. \end{aligned}$$

We shall assume q_{k+1} to be conditionally independent of x_k given q_k. More precisely, for every $k \in \mathbb{N}_0$, we will assume that

$$\mathbb{P}[q_{k+1} | q_k, x_0, w, v] = \mathbb{P}[q_{k+1} | q_k].$$

The joint process (x, q) is Markovian: for every $k \in \mathbb{N}_0$ it holds that

$$\mathcal{F}(x_{k+\ell}, q_{k+\ell} | x_k, q_k, x_{k-1}, q_{k-1}, \ldots, x_0, q_0) = \mathcal{F}(x_{k+\ell}, q_{k+\ell} | x_k, q_k),$$

where $\mathcal{F}$ denotes probability distribution. Therefore, the couple (x_k, q_k) will be called the state of the system at time k. In particular, x_k will be called the continuous state, as opposed to q_k, which will be called the discrete state.

For a fixed sequence $\{q_k\}$, equation (2.1) represents a linear Gaussian system. Thus, letting q^k denote the subsequence $q_0, \ldots, q_k$, the joint density function

$$f(x_\ell, y_{\ell'}|q^k)$$

is a multivariate Gaussian for all $\ell, \ell' \leq k$. However, the unconditioned system is not linear, since it depends on the outcomes of $\{q_k\}$. Note that

$$f(x_\ell, y_{\ell'}) = \sum_{q^k} f(x_\ell, y_{\ell'}|q^k)\mathbb{P}[q^k].$$

Therefore, the unconditioned joint distribution of x_ℓ and $y_{\ell'}$ is a mixture of N^{k+1} Gaussian distributions.

2.2 State estimation problem

Let y^k denote the measurements available up to time k, i.e.,

$$y^k \triangleq \{y_0, \ldots, y_k\}.$$

We are concerned in the estimation of the state (x_ℓ, q_ℓ) given y^k. In particular, we will consider the case of filtering, if $\ell = k$, and prediction, if $\ell > k$. Let

$$\hat{x}_{\ell|k} \triangleq \arg\min_z \mathbb{E}[||x_\ell - z(y^k)||^2], \tag{2.2a}$$

$$\hat{q}_{\ell|k} \triangleq \arg\min_z \mathbb{P}[q_\ell \neq z(y^k)], \tag{2.2b}$$

where z is in both cases a measurable function of the data. By definition, $\hat{x}_{\ell|k}$ is the minimum-mean-squared-error estimator of x_ℓ, and $\hat{q}_{\ell|k}$ is the minimum-probability-of-error estimator of q_ℓ. In light of the results reviewed in Chapter 1, it holds that

$$\begin{aligned}\hat{x}_{\ell|k} &= \mathbb{E}[x_\ell|y^k],\\ \hat{q}_{\ell|k} &= \arg\max_{j\in\mathcal{Q}} \mathbb{P}[q_\ell = j|y^k].\end{aligned}$$

For $\ell = k, k+1$, consider the a posteriori density $f(x_\ell|y^k)$, from which $\hat{x}_{\ell|k}$ follows. One may write

$$f(x_\ell|y^k) = \sum_{q^k} f(x_\ell|y^k, q^k)p(q^k|y^k),$$

where $p(q^k|y^k)$ is the a posteriori probability of q^k. Since $f(x_\ell|y^k, q^k) \propto f(x_\ell, y^k|q^k)$, it must be the case that

$$f(x_\ell|y^k, q^k) = \mathcal{N}(\hat{x}^{q^k}_{\ell|k}, P^{q^k}_{\ell|k}),$$

for some mean and covariance parameters depending on q^k. Applying Proposition 1.16 w.r.t. the a posteriori density $f(x_\ell|y^k, q^k)$, it follows that

$$\hat{x}^{q^k}_{\ell|k} = \arg\min_z \mathbb{E}[||x_\ell - z(y^k)||^2|q^k] = \mathbb{E}[x_\ell|y^k, q^k],$$

i.e. $\hat{x}^{q^k}_{\ell|k}$ is itself the solution of the state estimation problem conditioned on q^k. Correspondingly, $P^{q^k}_{\ell|k}$ is the error covariance matrix $\mathrm{Var}(x_\ell - \hat{x}^{q^k}_{\ell|k}|y^k, q^k)$, which may be shown to be data-independent [36]. That is,

$$P^{q^k}_{\ell|k} = \mathrm{Var}(x_\ell - \hat{x}^{q^k}_{\ell|k}|q^k).$$

Proposition 2.1 *It holds that*

$$\hat{x}_{\ell|k} = \sum_{q^k} \hat{x}^{q^k}_{\ell|k} p(q^k|y^k). \tag{2.3}$$

Proof: By the total probability law,

$$\begin{aligned}\mathbb{E}[x_\ell|y^k] &= \int x_\ell f(x_\ell|y^k)dx_\ell = \int x_\ell \sum_{q^k} f(x_\ell|y^k, q^k)p(q^k|y^k)dx_\ell \\ &= \sum_{q^k}\{\int x_\ell f(x_\ell|y^k, q^k)dx_\ell\}p(q^k|y^k) = \sum_{q^k}\mathbb{E}[x_\ell|y^k, q^k]p(q^k|y^k).\end{aligned}$$

□

Therefore, $\hat{x}_{\ell|k}$ is a weighted average of the solutions of N^{k+1} estimation problems. This fact was first noticed by Ackerson and Fu [1] in a slightly less general context. Consider now the a posteriori probability $p(q^k|y^k)$.

Proposition 2.2 *It holds that*

$$p(q^k|y^k) \propto f(y^k|q^k)p(q^k) = f(y_k|y^{k-1}, q^k) \cdot \ldots f(y_0|y^{-1}, q^0)p(q^k)$$

where $f(y_0|y^{-1}, q^0) \triangleq f(y_0|q_0)$ *and* $p(q^k)$ *is the a priori probability of* q^k. *In turn, for* $l = 0, \ldots, k$,

$$f(y_l|y^{l-1}, q^l) = \mathcal{N}(C_{q_l}\hat{x}^{q^l}_{l|l-1}, C_{q_l}P^{q^l}_{l|l-1}C^T_{q_l} + D_{q_l}D^T_{q_l}), \tag{2.4}$$

where q_l *indicates the last element of the sequence* q^l.

Proof: First expression is obtained by Bayes'rule,

$$p(q^k|y^k) = \frac{f(y^k|q^k)p(q^k)}{f(y^k)},$$

and iterated application of the relation

$$f(y^k|q^k) = f(y_k|y^{k-1}, q^k)f(y^{k-1}|q^k).$$

Because of the conditional independence of q_k on y^{k-1} given q_{k-1},

$$f(y^{k-1}|q^k) = f(y^{k-1}|q_k, q^{k-1}) = f(y^{k-1}|q^{k-1}).$$

Second expression follows from the measurement equation of (2.1). Consider w.l.o.g. $l = k$. As q^k is fixed, y_k is Gaussian and the sum is between independent vectors. Thus,

$$\mathbb{E}[y_k|y^{k-1}, q^k] = \mathbb{E}[C_{q_k}x_k|y^{k-1}, q^k] + \mathbb{E}[D_{q_k}v_k|y^{k-1}, q^k] = C_{q_k}\mathbb{E}[x_k|y^{k-1}, q^{k-1}],$$

and

$$\begin{aligned}\mathrm{Var}(y_k|y^{k-1}, q^k) &= \mathrm{Var}(C_{q_k}x_k|y^{k-1}, q^k) + \mathrm{Var}(D_{q_k}v_k|y^{k-1}, q^k)\\ &= C_{q_k}\mathrm{Var}(x_k|y^{k-1}, q^{k-1})C_{q_k}^T + D_{q_k}D_{q_k}^T,\end{aligned}$$

where the conditional independence of q_k on x_k, y^{k-1} given q_{k-1} was used. □

Note that estimates $\hat{q}_{\ell|k}$ may be formulated in terms of $p(q^k|y^k)$. In fact, they amount to the maximization of $p(q_\ell|y^k)$. If $\ell \leq k$, $p(q_\ell|y^k)$ follows from $p(q^k|y^k)$ by marginalization. Otherwise, one may write

$$p(q_\ell|y^k) = \sum_{q^k} p(q^k|y^k)(\pi^{\ell-k})_{q_k,q_\ell},$$

where q_k denotes the last element of q^k and the conditional independence of q_ℓ from y^k given q^k has been used.

Therefore, both the continuous and the discrete state estimation problem may be reduced to the problem of computing $\hat{x}_{\ell|k}^{q^k}$ and $P_{\ell|k}^{q^k}$ for increasing values of k. If $\ell = k, k+1$, for every possible sequence $\{q_k\}$, this may be done iteratively by way of a *conditioned* Kalman filter.

Proposition 2.3 *The following two-step iteration holds:*

$$\begin{aligned}\hat{x}_{k|k}^{q^k} &= \hat{x}_{k|k-1}^{q^k} + L_k^{q^k}[y_k - C_{q_k}\hat{x}_{k|k-1}^{q^k}],\\ P_{k|k}^{q^k} &= P_{k|k-1}^{q^k} - L_k^{q^k}C_{q_k}P_{k|k-1}^{q^k},\end{aligned}$$

where $L_k^{q^k} = P_{k|k-1}^{q^k} C_{q_k}^T [D_{q_k} D_{q_k}^T + C_{q_k} P_{k|k-1}^{q^k} C_{q_k}^T]^{-1}$, *and*

$$\hat{x}_{k+1|k}^{q^k} = A_{q_k} \hat{x}_{k|k}^{q^k},$$
$$P_{k+1|k}^{q^k} = A_{q_k} P_{k|k}^{q^k} A_{q_k}^T + B_{q_k} B_{q_k}^T,$$

with $\hat{x}_{0|-1}^{q^k} = \hat{x}_0$ *and* $P_{0|-1}^{q^k} = P_0$.

Proof: Observe that system (2.1) conditioned on q^k is linear with parameters $(A_{q_k}, B_{q_k}, C_{q_k}, D_{q_k})$ fixed by the last element q_k of the sequence q^k. Hence, apply Theorem 1.1. □

In practice, running N^{k+1} Kalman filters in parallel is prohibitive, hence the need of suboptimal estimation strategies. These will be commented in Section 2.4.

2.3 Recursive solution

From a theoretical point of view, one may alternatively solve filtering and prediction by a recursion on the functions $p(q_\ell|y^k)$ and $f(x_\ell|q_\ell, y^k)$. Observe that

$$f(x_\ell|y^k) = \sum_{q_\ell} f(x_\ell|q_\ell, y^k) p(q_\ell|y^k),$$

from which the estimate $\hat{x}_{\ell|k}$ follows.

Proposition 2.4 *For $k \geq 0$, the following two-step recursion holds:*

1. *Measurement update:*

$$p(q_k|y^k) = f(y_k|y^{k-1}, q_k) p(q_k|y^{k-1}) / f(y_k|y^{k-1}) \quad (2.5a)$$
$$f(x_k|q_k, y^k) = f(y_k|x_k, q_k) f(x_k|q_k, y^{k-1}) / f(y_k|y^{k-1}, q_k) \quad (2.5b)$$

   with

$$f(y_k|y^{k-1}) = \sum_{q_k \in \mathcal{Q}} f(y_k|y^{k-1}, q_k) f(q_k|y^{k-1})$$

2. *Time update:*

$$p(q_{k+1}|y^k) = \sum_{q_k} \pi_{q_k, q_{k+1}} p(q_k|y^k) \quad (2.6a)$$

$$f(x_{k+1}|q_{k+1}, y^k) = \sum_{q_k} f(x_{k+1}|q_k, q_{k+1}, y^k) \frac{p(q_k|y^k)}{p(q_{k+1}|y^k)} \pi_{q_k, q_{k+1}} \quad (2.6b)$$

Initial conditions are given by $p(q_0|y^{-1}) = p_0(q_0)$ *and* $f(x_0|q_0, y^{-1}) = f(x_0)$.

Proof: Equations (2.5a) and (2.5b) are simple applications of the Bayes'rule. Equation (2.6a) follows by application of the total probability law and of the conditional independence of q_{k+1} on y^k given q_k. Similarly,

$$f(x_{k+1}|q_{k+1}, y^k) = \sum_{q_k} f(x_{k+1}|q_k, q_{k+1}, y^k) p(q_k|q_{k+1}, y^k)$$

where

$$p(q_k|q_{k+1}, y^k) = \frac{p(q_{k+1}, q_k|y^k)}{p(q_{k+1}|y^k)} = \frac{p(q_{k+1}|q_k)p(q_k|y^k)}{p(q_{k+1}|y^k)},$$

and the statement is proven. □

The remaining terms of the measurement update step follow from the measurement equation of (2.1). As v_k is conditionally independent of x_k and of y^{k-1} given q_k, $f(y_k|y^{k-1}, q_k)$ is equal to the convolution of functions of $\mathbb{R}^p$

$$f(C_{q_k} x_k|q_k, y^{k-1}) \circledast f(D_{q_k} v_k|q_k). \tag{2.7}$$

The left factor follows from $f(x_k|q_k, y^{k-1})$ in in accordance with the linear transformation $C_{q_k} x_k$, while $f(D_{q_k} v_k|q_k) = \mathcal{N}(0, D_{q_k} D_{q_k}^T)$. Furthermore,

$$f(y_k|q_k, x_k) = \mathcal{N}(C_{q_k} x_k, D_{q_k} D_{q_k}^T). \tag{2.8}$$

On the other hand, the expression of $f(x_{k+1}|q_k, q_{k+1}, y^k)$ follows from the state update equation and is

$$f(A_{q_k} x_k|q_k, y^k) \circledast f(B_{q_k} u_k|q_k) \tag{2.9}$$

because $A_{q_k} x_k$ is conditionally independent of $B_{q_k} u_k$. The rightmost term is simply $\mathcal{N}(0, B_{q_k} B_{q_k}^T)$, whereas the leftmost term follows from $f(x_k|q_k, y^k)$ under the linear transformation $A_{q_k} x_k$.

Remark. Both $f(C_{q_k} x_k|q_k, y^{k-1})$ and $f(A_{q_k} x_k|q_k, y^k)$ are finite Gaussian mixtures,

$$f(C_{q_k} x_k|q_k, y^{k-1}) = \sum_{q^{k-1}} f(C_{q_k} x_k|q_k, q^{k-1}, y^{k-1}) p(q^{k-1}|q_k, y^{k-1}),$$
$$f(A_{q_k} x_k|q_k, y^k) = \sum_{q^{k-1}} f(A_{q_k} x_k|q_k, q^{k-1}, y^k) p(q^{k-1}|q_k, y^k),$$

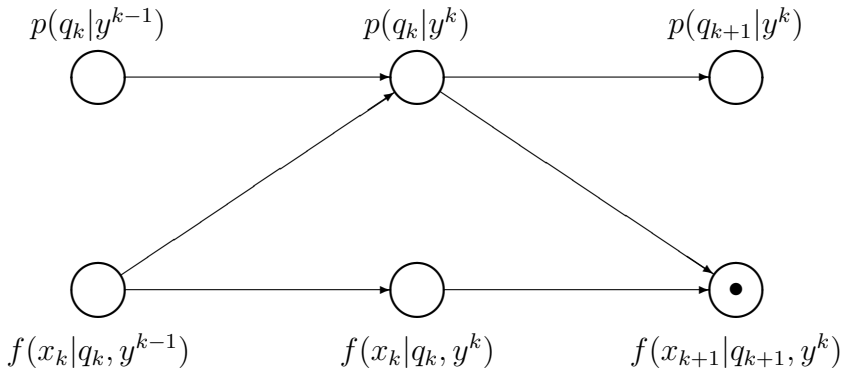


Figure 2.2: Recursive computation of $f(x_\ell|q_\ell, y^k)$ and $p(q_\ell|y^k)$

where

$$f(C_{q_k}x_k|q_k, q^{k-1}y^{k-1}) = \mathcal{N}(C_{q_k}\hat{x}^{q^{k-1}}_{k|k-1}, C_{q_k}P^{q^{k-1}}_{k|k-1}C^T_{q_k}),$$
$$f(A_{q_k}x_k|q_k, q^{k-1}, y^k) = \mathcal{N}(A_{q_k}\hat{x}^{q^k}_{k|k}, A_{q_k}P^{q^k}_{k|k}A^T_{q_k}),$$

while $p(q^{k-1}|q_k, y^{k-1})$ and $p(q^{k-1}|q_k, y^k)$ are easily written in terms of $p(q^\ell|y^\ell)$, $\ell = k-1, k$, and of π. Correspondingly, (2.7) and (2.9) may be rewritten as a weighted sum of convolutions between Gaussian distributions, each resulting in

$$\mathcal{N}(C_{q_k}\hat{x}^{q^{k-1}}_{k|k-1}, C_{q_k}P^{q^{k-1}}_{k|k-1}C^T_{q_k} + D_{q_k}D^T_{q_k}),$$
$$\mathcal{N}(A_{q_k}\hat{x}^{q^k}_{k|k}, A_{q_k}P^{q^k}_{k|k-1}A^T_{q_k} + B_{q_k}B^T_{q_k}),$$

in the order. The first expression yields the conditioned predictor of y_k given q^k and the associated error covariance matrix, whereas the second expression is the update step of Proposition 2.3. Therefore, computing (2.7) and (2.9) at each step k is equivalent to carrying over $\hat{x}^{q^k}_{k|k-1}$, $P^{q^k}_{k|k-1}$ and the weights $p(q^\ell|y^\ell)$. □

One cycle – measurement update and time update – of the recursion is depicted in Figure 2.2. Following on the above remark, Gaussian mixing is introduced by the update formula (2.6b), whose location in the diagram is marked with a symbol "•".

2.4 Approximate estimation methods

The recursive method outlined by Proposition 2.4 does not overcome the exponential complexity of the estimation problem. However, diverse approximate algorithms may be derived from this scheme. Will shall now present the classical suboptimal estimation algorithms and a brief overview of other solutions.

Generalized pseudo-Bayes (GPS) algorithm. The basic form of the method was proposed in [1] with a slightly different definition of model 2.1. It is based on the approximation of the Gaussian mixtures with a single Gaussian. In our setting, the algorithm would be as follows. At each step of the recursion (2.5)÷(2.6), the mixture of N^k Gaussian distributions $f(x_k|q_k, y^{k-1})$ is approximated by a single Gaussian of mean $\hat{x}_{k|k-1}$ and covariance matrix $P_{k|k-1}$. Given a new measurement y_k, the measurement update is carried out individually for each value of $q_k \in \mathcal{Q}$. Equation (2.5b) is replaced by N conditioned Kalman updates

$$\begin{aligned}\hat{x}^{q_k}_{k|k} &= \hat{x}_{k|k-1} + L^{q_k}_k[y_k - C_{q_k}\hat{x}_{k|k-1}],\\ P^{q_k}_{k|k} &= P_{k|k-1} - L^{q_k}_k C_{q_k} P_{k|k-1},\end{aligned}$$

where $L^{q_k}_k \triangleq P_{k|k-1}C^T_{q_k}[D_{q_k}D^T_{q_k} + C_{q_k}P_{k|k-1}C^T_{q_k}]^{-1}$. Equation (2.5a) is unchanged, but now one takes

$$f(y_k|y_{k-1}, q_k) = \mathcal{N}(C_{q_k}\hat{x}_{k|k-1}, C_{q_k}P_{k|k-1}C^T_{q_k} + D_{q_k}D^T_{q_k}).$$

The time update step amounts to computing (2.6a) along with the N conditioned predictions

$$\begin{aligned}\hat{x}^{q_k}_{k+1|k} &= A_{q_k}\hat{x}^{q_k}_{k|k},\\ P^{q_k}_{k+1|k} &= A_{q_k}P^{q_k}_{k|k}A^T_{q_k} + B_{q_k}B^T_{q_k}\end{aligned}$$

in place of (2.6b). The Gaussian approximation

$$f(x_{k+1}|q_{k+1}, y^k) \simeq \mathcal{N}(\hat{x}_{k+1|k}, P_{k+1|k})$$

is introduced by defining

$$\begin{aligned}\hat{x}_{k+1|k} &\triangleq \sum_{q_k} \hat{x}^{q_k}_{k+1|k} p(q_k|y^k),\\ P_{k+1|k} &\triangleq \sum_{q_k} P^{q_k}_{k+1|k} p(q_k|y^k) + \{\sum_{q_k}[x^{q_k}_{k+1|k}(x^{q_k}_{k+1|k})^T - \hat{x}_{k+1|k}\hat{x}^T_{k+1|k}]p(q_k|y^k)\},\end{aligned}$$

where the term in braces compensates for the variance nonlinearity. That is, the mixing step (2.6b) is reduced to the computation of ensemble conditioned mean and covariance matrix. At any time k, the prediction of state x_k and the corresponding error covariance matrix are approximated by $\hat{x}_{k|k-1}$ and $P_{k|k-1}$. On the other hand, the estimation of x_k given y^k is formulated as

$$\begin{aligned}\hat{x}_{k|k} &= \sum_{q_k} \hat{x}_{k|k}^{q_k} p(q_k|y^k),\\ P_{k|k} &= \sum_{q_k} P_{k|k}^{q_k} p(q_k|y^k) + \{\sum_{q_k} [x_{k|k}^{q_k}(x_{k|k}^{q_k})^T - \hat{x}_{k|k}\hat{x}_{k|k}^T] p(q_k|y^k)\}.\end{aligned}$$

Therefore, the estimation burden is reduced from N^{k+1} down to N conditioned Kalman filters. Subsequent versions of the algorithm – see [54] and references therein – use the less drastic assumption that

$$f(x_k|q_k, \dots, q_{k-\ell}, y^{k-1}) = \mathcal{N}(\hat{x}_{k|k-1}, P_{k|k-1}).$$

This results in approximating $f(x_k|q_k, y^{k-1})$ with a mixture of N^ℓ Gaussian distributions, for an overall complexity of $\mathcal{O}(N^{\ell+1})$. For $\ell = k$, the optimal estimator is recovered.

Detection-estimation (DE) algorithm. The detection estimation algorithm is based on a direct approximation of (2.3). It was first presented in [55]. Let Θ_k denote a set of up to M trajectories q^k, where M is a design parameter describing the allowable complexity. An approximate estimate of $\hat{x}_{\ell|k}$ is given by

$$\hat{x}_{\ell|k} \simeq \sum_{q^k \in \Theta_k} \hat{x}_{\ell|k}^{q^k} \tilde{p}(q^k|y^k), \tag{2.10}$$

where the a posteriori probabilities $\tilde{p}(q^k|y^k) \propto p(q^k|y^k)$ are defined as

$$\tilde{p}(q^k|y^k) \triangleq \frac{p(y^k|q^k)p(q^k)}{\sum_{q_k \in \Theta_k} p(y^k|q^k)p(q^k)}.$$

In light of Proposition 2.2, these may be updated recursively in terms of $\hat{x}_{k|k-1}^{q^k}$ and $P_{k|k-1}^{q^k}$. Because of renormalization, up to M such predictors are needed. The set Θ_k is chosen so to contain highly probable trajectories as follows. Let Θ_{k-1} contain M_{k-1} trajectories. Given some fixed threshold $\kappa > 0$, consider the set $\tilde{\Theta}_k$ of all trajectories $q^k = (q^{k-1}, q)$, $q_{k-1} \in \Theta_{k-1}$, $q \in \mathcal{Q}$ satisfying

$$(e_k^{q^k})^T (\Lambda_k^{q^k})^{-1} (e_k^{q^k}) \leq \kappa, \tag{2.11}$$

where

$$\begin{aligned} e_k^{q^k} &= y_k - C_{q_k}\hat{x}_{k|k-1}^{q^k}, \\ \Lambda_k^{q^k} &= C_{q_k}P_{k|k-1}C_{q_k}^T + D_{q_k}D_{q_k}^T. \end{aligned}$$

If $|\tilde{\Theta}_k| \leq M$, set $\Theta_k = \tilde{\Theta}_k$, otherwise, form Θ_k by choosing the M elements of $\tilde{\Theta}_k$ with the largest values of $p(q^k|y^k)$. Observe that, if q^k is the trajectory underlying the generation of y^k, then $e_k^{q^k}$ is the *minimum* mean-squared error for the prediction of y_k given y^{k-1}. In this case, the left-hand side of (2.11) has known distribution $\chi^2(p)$ (compare [55]). Therefore κ may be chosen as to limit the probability of rejecting the true trajectory q^k. Moreover, trajectories q^k "far away" from the true outcome will lead to larger values of the left-hand side of (2.11) and are likely to be discarded. This guarantees that only good estimates of q^k are employed in the approximation (2.10). In this sense, the state estimation step is *subordinated to* a *detection* step where the most probable trajectories q^k are elicited. Finally, the reduction of $\tilde{\Theta}_k$ to Θ_k ensures that $\Theta_k \leq M$. It becomes essential when system (2.1) is such that different trajectories q^k lead to similar statistics of $e_k^{q^k}$. The overall complexity of the algorithm is of $\mathcal{O}(M)$. For $\kappa \to +\infty$ and $M = M_k = N^{k+1}$, the optimal estimation algorithm is recovered.

Remark. A further refinement of the algorithm consists in "merging" those trajectories for which $e_k^{q^k}$ and $\Lambda_k^{q^k}$ are almost the same. This amounts to carrying on a "representative" q^k and the associated conditioned estimates, and to collecting into $\tilde{p}(q^k|y^k)$ the a posteriori probabilities of the trajectories that are no longer considered. In [55], two trajectories are merged if the Bhattacharyya coefficient [37] between the corresponding densities $f(y_k|y^{k-1}, q^k)$ exceeds a certain bound. A similar idea will be developed in Chapter 6 to study the ability of a sequential test to discriminate between different trajectories. □

Interactive multiple-model (IMM) algorithm. This suboptimal estimation paradigm relies on running exactly N Kalman filters in parallel, one for each of the possible modes (A_q, B_q, C_q, D_q), and to find suboptimal estimates of x_k by suitable averaging. For $q^k \equiv q$, let $\hat{x}_{\ell|k}^q = \hat{x}_{\ell|k}^{q^k}$ and $P_{\ell|k}^q = P_{\ell|k}^{q^k}$ denote the N matched filters. Then [46], the basic form of the IMM estimator

of x_k given y^k is

$$\hat{x}_{k|k} \triangleq \sum_{q\in\mathcal{Q}} \hat{x}^q_{\ell|k}\eta^q_k,$$
$$P_{k|k} \triangleq \sum_{q\in\mathcal{Q}} (P^q_{k|k} + (\hat{x}_{k|k} - \hat{x}^q_{k|k})(\hat{x}_{k|k} - \hat{x}^q_{k|k})^T)\eta^q_k,$$

where the weights η_q are approximations of the a posteriori probability $p(q_k|y^k)$ evaluated for $q_k = q$. The standard update of $\hat{x}^q_{k|k}$ and $P^q_{k|k}$ given in Theorem 1.1 is modified so to take into account the possible *interaction* of the modes, i.e. the possible occurrence of a switch between different modes. For $i, j \in \mathcal{Q}$, one defines the mixing probabilities $\eta^{i,j}_k \propto \pi_{i,j}\eta^i_k$, and the quantities

$$\hat{x}^{0,j}_{k|k} \triangleq \sum_{i\in\mathcal{Q}} \hat{x}^i_{k|k}\eta^{i,j}_k,$$
$$P^{0,j}_{k|k} \triangleq \sum_{i\in\mathcal{Q}} (P^i_{k|k} + (\hat{x}^i_{k|k} - \hat{x}^{0,j}_{k|k})(\hat{x}^i_{k|k} - \hat{x}^{0,j}_{k|k})^T)\eta^{i,j}_k.$$

These replace $\hat{x}^j_{k|k}$ and $P^j_{k|k}$ as initial conditions of the time update (1.8),

$$\hat{x}^j_{k+1|k} = A_j\hat{x}^{0,j}_{k|k},$$
$$P^j_{k+1|k} = A_jP^{0,j}_{k|k}A^T_j + B_jB^T_j,$$

while the computation of $\hat{x}^j_{k+1|k+1}$ and $P^j_{k+1|k+1}$ from $\hat{x}^j_{k+1|k}$ and $P^j_{k+1|k}$ follows unaltered from (1.7). Finally, the update of the mixing coefficients is done by setting

$$\eta^j_{k+1} \propto f^j(y_{k+1}|y^k)\sum_{i\in\mathcal{Q}} \eta^{i,j}_k,$$

where $f^j(y_{k+1}|y^k)$ is the evaluation at y_{k+1} of the likelihood function

$$\mathcal{N}(C_j\hat{x}^j_{k+1|k}, C_jP^j_{k+1|k}C^T_j + D_jD^T_j).$$

A derivation of this algorithm from a recursion similar to that of Proposition 2.4 is shown in [10]. The update of η^j_k by way of the mixing coefficients $\eta^{i,j}_k$ is just a restatement of equations (2.5a) and (2.6a), with

$$\sum_{i\in\mathcal{Q}} \eta^{i,j}_k \simeq \mathbb{P}[q_{k+1} = j|y^k].$$

The Gaussian mixing operated by (2.6b) is pushed back to the mixed initial conditions $\hat{x}^{0,j}_{k|k}$ and $P^{0,j}_{k|k}$. To conclude, the measurement update (2.5b) is reduced to a separate update of the type (1.7) for each of the modes $j \in \mathcal{Q}$. The complexity of this algorithm is given by the computation of N Kalman filters in parallel. A whole family of algorithms based on mixing of mode-matched filters is illustrated in [46].

Random sampling (RS) algorithm. An alternative approach is based on the Monte Carlo generation of a large number of discrete state trajectories and the use of an approximation like (2.10). The idea was first presented in [2], and may be explained in simple terms as follows. Let Θ_k denote a set of M trajectories q^k. An *importance* weight $\eta_k^{q^k}$ is assigned to each trajectory $q^k \in \Theta_k$. Assume that

$$\eta^{q^k} \simeq p(q^k|y^k). \tag{2.12}$$

Then, a reasonable approximation of $\hat{x}_{k|k}$ is

$$\hat{x}_{k|k} = \sum_{q^k \in \Theta_k} \hat{x}_{k|k}^{q^k} \eta^{q^k},$$

where $\hat{x}_{k|k}^{q^k}$ is the Kalman filter conditioned on q^k. Given a new measurement y_{k+1}, one wants to (1) extend every trajectory $q^k \in \Theta_k$ to $q^{k+1} = (q^k, q_{k+1})$ by *generating at random* its next state $q_{k+1} \in \mathcal{Q}$, thus forming Θ_{k+1}, and (2) update the weights $\eta_k^{q^k}$. The idea is to sample the new state q_{k+1} according to its a posteriori probability

$$p(q_{k+1}|y^{k+1}, q^k) \propto f(y_{k+1}|q_{k+1}, q^k, y^k) p(q_{k+1}|q^k)$$

and to update the weights by privileging those trajectories that best explain the data, i.e. by setting

$$\eta_{k+1}^{q^{k+1}} \propto f(y_{k+1}, q_{k+1}|q^k, y^k) \eta_k^{q^k} = f(y_{k+1}|q_{k+1}, q^k, y^k) p(q_{k+1}|q^k) \eta_k^{q^k} \tag{2.13}$$

with normalization factors chosen so that $\sum_{q^k \in \Theta_k} \eta_k^{q^k} = 1$ at all k. Computing factors $f(y_{k+1}|q_{k+1}, q^k, y^k)$ amounts to evaluate (2.4) for each of the M Kalman predictors $\hat{x}_{k+1|k}^{q^k}$, $P_{k+1|k}^{q^k}$. Factor $p(q_{k+1}|q^k)$ is computed directly in terms of the transition probability matrix π. Note that (2.13) is, besides normalization, the same recursion satisfied by $p(q^k|y^k)$, provided the initialization by p_0. The stochastic sampling mechanism guarantees that trajectories with larger probability $p(q^k|y^k)$ will be generated. In principle, if $q_k \notin \Theta_k$, then $p(q^k|y^k) \simeq 0$. This motivates approximation (2.12). Using the same approximation, the distribution of the current discrete state may be evaluated by

$$p_{k|k}(q) = \sum_{q^k} \mathbb{1}_{\{q_k = q\}}(q^k) p(q^k|y^k) \simeq \sum_{q^k \in \Theta_k} \mathbb{1}_{\{q_k = q\}}(q^k) \eta_k^{q^k}.$$

The computational burden of the algorithm is given by the M Kalman filters and predictors that need be run in parallel. In order for the approximations to have statistical significance, a rather large number of trajectories M needs be considered. For $M \to +\infty$, random sampling yields optimal estimates with probability one [22].

Other methods In the last few decades, a number of suboptimal estimation strategies have been considered. In [29], polynomial functions rather than linear functions of the data y^k have been considered for estimation. Joint state-mode estimation has been address in [18, 19], where the random variable $x_k \cdot \mathbb{1}_{q_k}$ is estimated in a linear fashion. Sequential Monte Carlo methods have been further investigated in [21, 24, 39, 40, 42]. Sequential hypothesis testing, whose application to jump Markov systems is explored in Chapter 6, is used in [6, 49] for the detection of abrupt changes in linear systems. An information-theoretic approach has recently been proposed in [44, 45] for the fundamental case of switching finite-impulse-response systems with known input. Further tecniques are illustrated in [12, 25, 33, 41, 48, 53].

2.5 Discussion

State estimation for systems undergoing Markovian jumps has been an active field of research in the last thirty years. Following a common approach, we have considered a switching model where the system evolves according to one of a known finite set of linear modes, and jumps are regulated by a Markov chain q_k with transition probabilities known and independent of the continuous state x_k. We have shown that both the continuous state x_k and the discrete state q_k may be estimated by solving an exponentially growing number of conditioned linear estimation problems. An alternative method was illustrated based on the joint two-step recursive update of the a posteriori probability $p(q_\ell|y^k)$ and density $f(x_\ell|q_\ell, y^k)$. Both solutions are not practicable in that they require exponentially increasing computational capabilities. However, they are at the basis of the currently available suboptimal estimation algorithms, which were commented in Section 2.4.

Diverse suboptimal solutions, among which the DE and RS algorithms, are essentially composed of two steps. In a first step, a certain number of discrete-state trajectories are selected. These are used in a second step for both estimation of q_k and approximation of the minimum-mean-squared-error estimator of x_k. In a sense, state estimation is built on top of the solution of a *discrete-state trajectory detection* problem. In fact, once a finite subset of most probable trajectories $\{q^k\}$ has been determined, one may estimate q^k by searching for the sequence in the set that best matches the data y^k. However, this subset is often built with regard to the estimate of x_k and q_k only. The estimation of the whole trajectory is considered explicitly in [44, 42, 24], among others. In Chapter 6, we will consider the application of sequential hypothesis testing to trajectory estimation.

The discrete-time jump Markov system (2.1) is often used to model stochastic systems with switching continuous-time dynamics and sampled measurements. In this case, the discrete-time state x_k stands for the value of a continuous-time state $\xi(t)$ at measurement instants t_k, and the update matrix A_{q_k} represents the evolution of ξ between two sample times t_k and t_{k+1}. Implicit in the scheme is the assumption that the evolution of the state between two measurements depends on a constant value of the discrete state q. In other words, it is assumed that switches take place in correspondence of measurements. In fact, the is no general reason for the switches of a continuous-time dynamical system to be confined to the fixed sample times $\{t_k\}$. In this context, model (2.1) should be regarded as an approximate discretization of the original system. When the system's dynamics are slow and the average dwell time of q is large compared to the time intervals (t_k, t_{k+1}), one may assign the occurrence of a switch to the closest measurement time t_k, and the error committed is negligible. If, however, measurements are sparse, model (2.1) is no longer a good approximation of the original system, regardless of the choice of the parameters (A_q, B_q, C_q, D_q). This issue will be addressed in Chapter 4, where a generalized jump Markov model based on a continuous-time Markov chain description of q will be considered.

Chapter 3

Superposition principle in linear filtering

In this chapter, decomposition formulas for the discrete-time Kalman filter are presented. Both the state estimate and the error covariance matrix are expressed as the sum of two terms, the first being the estimate corresponding to zero initial conditions, and the second an explicit function of the initial estimates $\hat{x}_0$ and P_0. The representation is updated in time by well-behaved finite-complexity matrix recursions, and allows for a direct evaluation of the estimates for variable initial conditions.

3.1 Introduction

In Kalman filtering, the initial state estimate, $\hat{x}_0$, and the initial error covariance matrix, P_0, are usually fixed. However, it is sometimes necessary to regard the estimates as functions of $\hat{x}_0$ and P_0. In the applications of our concern, the estimates need to be evaluated in correspondence of a continuous set of initial values. Therefore, the standard implementations of the Kalman filter are not appropriate.
This chapter presents an alternative formulation of Kalman filtering where the initial values $\hat{x}_0$ and P_0 appear in explicit form. Based on the connection between Riccati equation and linear difference equations, the one-step prediction and the error covariance matrix are split in two parts: the estimate and the error covariance matrix corresponding to the initial values $\hat{x}_0 = 0$, $P_0 = 0$ and two parametric functions of $\hat{x}_0$ and P_0. Finite-complexity matrix recursions are provided for the update of the predictor and are shown to be

numerically well-behaved in a time invariant setting. Prediction formulas are then extended to filtering.

This chapter is organized as follows. In the next section, we elaborate on the problems of filtering and prediction for discrete-time models and discuss the Kalman filtering algorithm presented in Chapter 1. In Section 3.3, the relationship between difference Riccati equations and linear dynamical equations is illustrated. The main results are presented in Section 3.4, where decomposition formulas for the Kalman filter are established. A discussion of the results is reported in Section 3.5.

3.2 Preliminaries

Consider the discrete-time linear Gaussian state-space model of Chapter 1, namely,

$$\begin{cases} x_{k+1} = A_k x_k + B_k u_k \\ y_k \;\; = C_k x_k + D_k v_k \end{cases}, \tag{3.1}$$

where $u_k \in \mathbb{R}^{m\times 1}$ and $v_k \in \mathbb{R}^{r\times 1}$ are zero-mean white noise sequences, $x_k \in \mathbb{R}^{n\times 1}$, $y_k \in \mathbb{R}^{p\times 1}$, $\{A_k, B_k, C_k, D_k\}$ are given matrices of proper dimensions with $D_k D_k > 0$, and u, v, x_0 are mutually independent. According to Theorem 1.1, the problem of computing, for $\ell = k$ and $\ell = k+1$, the minimum-mean-squared-error estimate

$$\hat{x}_{\ell|k} = \mathbb{E}[x_\ell | y^k]$$

and the estimation error covariance matrix

$$P_{\ell|k} \triangleq \mathbb{E}[(\hat{x}_{\ell|k} - x_\ell)(\hat{x}_{\ell|k} - x_\ell)^T]$$

given measurements y^k is solved by the Kalman recursion

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + L_k(y_k - C_k \hat{x}_{k|k-1}), \tag{3.2a}$$
$$P_{k|k} = P_{k|k-1} - L_k C_k P_{k|k-1}, \tag{3.2b}$$

with gain $L_k = P_{k|k-1}C_k^T(C_k P_{k|k-1} C_k^T + D_k D_k^T)^{-1}$, and

$$\hat{x}_{k+1|k} = A_k \hat{x}_{k|k}, \tag{3.3a}$$
$$P_{k+1|k} = A_k P_{k|k} A_k^T + Q_k, \tag{3.3b}$$

where $Q_k \triangleq B_k B_k^T$. If $x(0)$ has mean $\hat{x}_0$ and covariance matrix P_0, the recursion is started off by setting $\hat{x}_{0|-1} = \hat{x}_0$ and $P_{0|-1} = P_0$.

It is evident that the initial conditions of the filter influence both $\hat{x}$ and P at every step k. That is, $\hat{x} = \hat{x}(\hat{x}_0, P_0)$, $P = P(P_0)$. On the other hand, the dependency on $\hat{x}_0$ and P_0 is implicit and the algorithm gives no chance to evaluate the estimates for changing values of $\hat{x}_0$ and P_0, unless the whole computation is repeated from the initial step.

We shall show that equations (3.2)–(3.3) can be rearranged so to obtain explicit functions of $\hat{x}_0$ and P_0 at every step of the recursion. We will initially concentrate on the prediction problem. A simple substitution yields

$$\hat{x}_{k+1|k} = A_k(I - L_kC_k)\hat{x}_{k|k-1} + A_kL_ky_k, \tag{3.4a}$$
$$P_{k+1|k} = A_k(I - L_kC_k)P_{k|k-1}A_k^T + Q_k. \tag{3.4b}$$

These equations represent a time-varying filter whose dynamics is described by the state evolution matrix $\Phi_k \triangleq A_k(I - L_kC_k)$. By the algebraic identity

$$(I - P_{k|k-1}C_k^T(C_kP_{k|k-1}C_k^T + D_kD_k^T)^{-1}C_k) \cdot (I + P_{k|k-1}\Delta_k) = I, \tag{3.5}$$

where $\Delta_k \triangleq C_k^T(D_kD_k^T)^{-1}C_k$, a useful alternative expression of Φ_k is

$$\Phi_k = A_k(I + \Delta_kP_{k|k-1})^{-T}.$$

Equation (3.5) also shows the invertibility of Φ_k whenever A_k in nonsingular.

3.3 Difference Riccati and linear equations

The evolution of the error covariance matrix $P_{k|k-1}$ is governed by the difference Riccati equation (3.4b). It is well-known that the Riccati equation is related to linear difference equations.

Assumption 3.1 *Matrices A_k and $D_kD_k^T$ are invertible for all $k \geq 0$.*

Define the $2n \times 2n$ matrix

$$Z_k \triangleq \begin{bmatrix} A_k^{-T} & A_k^{-T}\Delta_k \\ Q_kA_k^{-T} & A_k + Q_kA_k^{-T}\Delta_k \end{bmatrix}.$$

The relationship is established by the following classical result (see e.g. [3]).

Lemma 3.1 *Consider the matrix difference equation*

$$\begin{bmatrix} X_{k+1} \\ Y_{k+1} \end{bmatrix} = Z_k \begin{bmatrix} X_k \\ Y_k \end{bmatrix} \tag{3.6}$$

with initial conditions $Y_0 = P_0X_0$, and X_0 any invertible matrix. Then, for all $k \geq 0$, X_k is invertible, and it holds that $P_{k|k-1} = Y_kX_k^{-1}$.

Remark. Assumption 3.1 will be maintained throughout the rest of the chapter. Were A_k not invertible, the Riccati equation should be studied in terms of generalized linear difference equations and symplectic pencils [51]. This would complicate the presentation and will not be done here. However, the assumption that A_k be nonsingular is appropriate for a quite general set of problems. □

Equation (3.6) may be written as

$$\begin{bmatrix} X_{k+1} \\ Y_{k+1} \end{bmatrix} = \Pi_k \begin{bmatrix} X_0 \\ Y_0 \end{bmatrix} \tag{3.7}$$

where Π_k satisfies the recursion $\Pi_k = Z_k \Pi_{k-1}$, provided $\Pi_{-1} \triangleq I$. Matrix Π_k inherits from Z_k the *symplectic* property: $\Pi_k^T J \Pi_k = J$, where J is the $2n \times 2n$ matrix

$$J = \begin{bmatrix} 0 & -I \\ I & 0 \end{bmatrix}$$

satisfying $J^T = J^{-1} = -J$. The property may be verified by induction, and has relevant consequences: in particular, it implies that the eigenvalues of Π_k come in reciprocal pairs. Thus, if we actually were to compute Π_k for, say, a time-invariant system, i.e. $\Pi_k = Z^{k+1}$ for some fixed $Z_k \equiv Z$, we would end up with an extremely ill-conditioned matrix even for small values of k.

Lemma 3.2 *The state evolution matrix and the gain of the one-step predictor obey the following equalities:*

$$\Phi_k = X_{k+1}^{-T} X_k^T,$$

$$A_k L_k = X_{k+1}^{-T} Y_k^T C_k^T (D_k D_k^T)^{-1}.$$

Proof: The first equality follows from identity (3.5) and Lemma 3.1:

$$A_k(I - L_k C_k) = A_k(I + \Delta_k P_{k|k-1})^{-T} = (A_k^{-T} X_k + A_k^{-T} \Delta_k Y_k)^{-T} X_k^T = X_{k+1}^{-T} X_k^T.$$

To verify the second equality, left-multiply both sides by $X_k^{-T} X_{k+1}^T$ so to write

$$X_k^{-T} X_{k+1}^T A_k L_k = X_k^{-T} Y_k^T C_k^T (D_k D_k^T)^{-1}. \tag{3.8}$$

Use first equality and identity (3.5) to write $X_k^{-T} X_{k+1}^T A_k$ as $(I + P_{k|k-1}\Delta_k)$. Also use Lemma 3.1 to replace $X_k^{-T} Y_k^T$ by $P_{k|k-1}$. Then (3.8) takes the form

$$(I + P_{k|k-1}\Delta_k) L_k = P_{k|k-1} C_k^T (D_k D_k^T)^{-1}.$$

Expand L_k, Δ_k and right-multiply by $C_k P_{k|k-1} C_k^T + D_k D_k^T$ to get

$$(I + P_{k|k-1} C_k^T (D_k D_k^T)^{-1} C_k) P_k C_k^T = P_{k|k-1} C_k^T (D_k D_k^T)^{-1} (C_k P_{k|k-1} C_k^T + D_k D_k^T).$$

Last equality is easily checked by inspection. □

Let the superscript "(i,j)" indicate the (i,j)-th $n \times n$ matrix block. Denote with $P^{\diamond}_{k|k-1} \triangleq P_{k|k-1}(0)$ the prediction error covariance matrix associated with zero initial conditions, and with $\Phi^{\diamond}_k$ and $L^{\diamond}_k$ the corresponding predictor state evolution matrix and gain:

$$\begin{aligned} \Phi^{\diamond}_k &\triangleq A_k (I + \Delta_k P^{\diamond}_{k|k-1})^{-T}, \\ L^{\diamond}_k &\triangleq P^{\diamond}_{k|k-1} C_k^T (C_k P^{\diamond}_{k|k-1} C_k^T + D_k D_k^T)^{-1}. \end{aligned} \tag{3.9}$$

Corollary 3.1 *1. The following recursion holds:*

$$\Pi^{1,1}_k = (\Phi^{\diamond}_k)^{-T} \Pi^{1,1}_{k-1},$$

initialized by $\Pi^{1,1}_{-1} = I$. In particular, $\Pi^{1,1}_k$ is invertible for all k;

2. It holds that $\Pi^{2,1}_{k-1} (\Pi^{1,1}_{k-1})^{-1} = P^{\diamond}_{k|k-1}$;

3. It holds that $A_k L^{\diamond}_k = (\Pi^{1,1}_k)^{-T} (\Pi^{2,1}_{k-1})^T C_k^T (D_k D_k^T)^{-1}$.

Proof: First observe that $X_k = (\Pi^{1,1}_{k-1} X_0 + \Pi^{1,2}_{k-1} P_0 X_0)$ and $Y_k = (\Pi^{2,1}_{k-1} X_0 + \Pi^{2,2}_{k-1} P_0 X_0)$. Thus, by Lemmas 3.1–3.2,

$$\begin{aligned} \Phi_k &= (\Pi^{1,1}_k + \Pi^{1,2}_k P_0)^{-T} (\Pi^{1,1}_{k-1} + \Pi^{1,2}_{k-1} P_0)^T, \\ P_{k|k-1} &= (\Pi^{2,1}_{k-1} + \Pi^{2,2}_{k-1} P_0)(\Pi^{1,1}_{k-1} + \Pi^{1,2}_{k-1} P_0)^{-1}, \\ A_k L_k &= (\Pi^{1,1}_k + \Pi^{1,2}_k P_0)^{-T} (\Pi^{2,1}_{k-1} + \Pi^{2,2}_{k-1} P_0)^T C_k^T (D_k D_k^T)^{-1}. \end{aligned}$$

The result follows by evaluation at $P_0 = 0$. □

Define $U_k \triangleq (\Pi^{1,1}_k)^{-T}$. Then, the result at point 1. may be restated as

$$U_k = \Phi^{\diamond}_k U_{k-1},$$

with $U_{-1} = I$. Consider in addition $S_k \triangleq U_k^T \Pi^{1,2}_k$.

Lemma 3.3 *Matrix S_k obeys the recursion*

$$S_k = U_{k-1}^T (I + \Delta_k P^{\diamond}_{k|k-1})^{-1} \Delta_k U_{k-1} + S_{k-1}$$

initialized by $S_{-1} = 0$. Moreover, $S_k \geq 0$ for all k.

Proof: By the definition of Π_k, $\Pi_k^{1,2} = A_k^{-T}\Pi_{k-1}^{1,2} + A_k^{-T}\Delta_k\Pi_{k-1}^{2,2}$. Notice the identity

$$\Pi^{2,2} = (\Pi^{1,1})^{-T} + \Pi^{2,1}(\Pi^{1,1})^{-1}\Pi^{1,2}, \tag{3.10}$$

holding for any symplectic matrix Π with nonsingular $\Pi^{1,1}$. Then

$$\Pi_k^{1,2} = A_k^{-T}\Delta_k(\Pi_{k-1}^{1,1})^{-T} + A_k^{-T}(I + \Delta_k\Pi_{k-1}^{2,1}(\Pi_{k-1}^{1,1})^{-1})\Pi_{k-1}^{1,2}.$$

Hence, using equation (3.9) and the results of Corollary 3.1,

$$\begin{aligned} S_k &= (\Pi_k^{1,1})^{-1}\{A_k^{-T}\Delta_k(\Pi_{k-1}^{1,1})^{-T} + A_k^{-T}(I + \Delta_k\Pi_{k-1}^{2,1}(\Pi_{k-1}^{1,1})^{-1})\Pi_{k-1}^{1,2}\} \\ &= (\Pi_{k-1}^{1,1})^{-1}(\Phi_k^{\diamond})^T\{A_k^{-T}\Delta_k(\Pi_{k-1}^{1,1})^{-T} + (\Phi_k^{\diamond})^{-T}\Pi_{k-1}^{1,2}\} \\ &= (\Pi_{k-1}^{1,1})^{-1}(I + \Delta_k P_{k|k-1}^{\diamond})^{-1}\Delta_k(\Pi_{k-1}^{1,1})^{-T} + S_{k-1}. \end{aligned} \tag{3.11}$$

To prove that $S_k \geq 0$, let us proceed by induction. That $S_{-1} \geq 0$ is obvious. Now assume $S_{k-1} \geq 0$. For any $\Delta \geq 0$ and $P \geq 0$, it holds that

$$(I + \Delta P)^{-1}\Delta = (I + \Delta P)^{-1}(\Delta + \Delta P\Delta)(I + \Delta P)^{-T} \geq 0. \tag{3.12}$$

Thus, both addends of (3.11) are nonnegative definite. The result follows. □

Quantities U_k, $P_k^{\diamond}$ and S_k will be the building blocks of an explicit representation of $\hat{x}(\hat{x}_0, P_0)$ and $P(P_0)$. Their properties may be promptly studied in a time-invariant setting. Assume $(A_k, B_k, C_k, D_k) \equiv (A, B, C, D)$ for all k. If (A, C) is detectable, then the limiting matrix $P_{k+1|k}^{\diamond} \xrightarrow{k} \overline{P}$ is positive semidefinite and bounded. If, in addition, (A, B) is controllable, then the limiting matrix $\Phi_k^{\diamond} \xrightarrow{k} \overline{\Phi}$ has all the eigenvalues strictly within the unit circle. As

$$U_k = \Phi_k^{\diamond}U_{k-1} = \ldots = \Phi_k^{\diamond} \cdot \ldots \cdot \Phi_0^{\diamond},$$

it follows that $U_k \xrightarrow{k} 0$ with ultimate exponential decay. This guarantees that $S_k \xrightarrow{k} \overline{S} < +\infty$. Therefore, all recursions prove numerically bounded and well-conditioned. If the system is not detectable and controllable, or it is not time-invariant, the analysis is more involved. In any case, the stability of the recursions is intimately related to that of the corresponding Kalman filter.

3.4 Decomposition formulas

We are now ready to establish the main results of this chapter. Based on the linear representation of the Riccati equation, an explicit expression for $P_{k+1|k}(P_0)$ is derived first.

Proposition 3.1 *The prediction error covariance matrix admits the following decomposition:*

$$P_{k+1|k}(P_0) = P^{\diamond}_{k+1|k} + U_k(I + S_kP_0)^{-T}P_0U_k^T. \tag{3.13}$$

Proof: By Lemma 3.1 and equation (3.7),

$$P_{k+1|k}(P_0) = (\Pi_k^{2,1} + \Pi_k^{2,2}P_0)(\Pi_k^{1,1} + \Pi_k^{1,2}P_0)^{-1}.$$

In light of identity (3.10), the first factor may be rewritten as

$$(\Pi_k^{1,1})^{-T}P_0 + \Pi_k^{2,1}(\Pi_k^{1,1})^{-1}(\Pi_k^{1,1} + \Pi_k^{1,2}P_0).$$

Therefore, using Corollary 3.1 and the definitions of U_k and S_k,

$$\begin{aligned} P_{k+1|k}(P_0) &= (\Pi_k^{1,1})^{-T}P_0(\Pi_k^{1,1} + \Pi_k^{1,2}P_0)^{-1} + \Pi_k^{2,1}(\Pi_k^{1,1})^{-1} \\ &= U_kP_0(I + S_kP_0)^{-1}U_k^T + P^{\diamond}_{k+1|k}. \end{aligned}$$

Expression (3.13) follows from the symmetry of $P_{k+1|k}$. □

The expression of $P_{k+1|k}(P_0)$ is a superposition of two terms. The first term, $P^{\diamond}_{k+1|k}$, corresponds to the initial value $P_0 = 0$. The second term accounts explicitly for P_0. It is characterized by a *finite set of parameters*, U_k and S_k, *independent of* P_0, which are *updated* by well-behaved recursions. Let now $\hat{x}^{\diamond}_{k+1|k} \triangleq \hat{x}_{k+1|k}(0,0)$ indicate the state prediction associated to $\hat{x}_0 = 0$ and $P_0 = 0$. Along the lines of Proposition 3.1, an explicit formula is found for $\hat{x}_{k+1|k}(\hat{x}_0, P_0)$.

Proposition 3.2 *The one-step predictor may be decomposed as*

$$\hat{x}_{k+1|k}(\hat{x}_0, P_0) = \hat{x}^{\diamond}_{k+1|k} + U_k(I + S_kP_0)^{-T}(\hat{x}_0 + P_0M_k), \tag{3.14}$$

where the quantity M_k satisfies the recursions

$$M_k = U_{k-1}^T(C_k^T(D_kD_k^T)^{-1}y_k - (I + \Delta_kP^{\diamond}_{k|k-1})^{-1}\Delta_k\widetilde{M}_k) + M_{k-1}$$

initialized by $M_{-1} = 0$. In turn, $\widetilde{M}_k$ satisfies

$$\widetilde{M}_k = \Phi^{\diamond}_{k-1}\widetilde{M}_{k-1} + P^{\diamond}_{k|k-1}C_k^T(D_kD_k^T)^{-1}y_k$$

initialized by $\widetilde{M}_{-1} = 0$.

Proof: Let $z_k \triangleq C_k^T(D_kD_k^T)^{-1}y_k$. By Lemma 3.2,

$$\hat{x}_{k+1|k} = X_{k+1}^{-T}X_k^T\hat{x}_{k|k-1} + X_{k+1}^{-T}Y_k^T z_k.$$

Developing this recursion backwards, one gets

$$\hat{x}_{k+1|k} = X_{k+1}^{-T}X_0^T\hat{x}_{0|-1} + X_{k+1}^{-T}\sum_{i=0}^{k} Y_i^T z_i. \tag{3.15}$$

Set w.l.o.g. $X_0 = I$. Then $X_{k+1} = \Pi_k^{1,1} + \Pi_k^{1,2}P_0 = \Pi_k^{1,1}(I + S_kP_0)$. Moreover,

$$\begin{aligned} Y_i &= \Pi_{i-1}^{2,1} + \Pi_{i-1}^{2,2}P_0 \\ &= \Pi_{i-1}^{2,1}(\Pi_{i-1}^{1,1})^{-1}(\Pi_{i-1}^{1,1} + \Pi_{i-1}^{1,2}P_0) + (\Pi_{i-1}^{1,1})^{-T}P_0 \\ &= P_{i|i-1}^{\diamond}\Pi_{i-1}^{1,1}(I + S_{i-1}P_0) + (\Pi_{i-1}^{1,1})^{-T}P_0, \end{aligned}$$

where (3.10) was used. By Lemma 3.3, term S_{i-1} may be written as $S_k - \sum_{l=i}^{k}\widetilde{S}_l$, with $\widetilde{S}_l = (\Pi_{l-1}^{1,1})^{-1}(I + \Delta_l P_{l|l-1}^{\diamond})^{-1}\Delta_l(\Pi_{l-1}^{1,1})^{-T}$, whence

$$Y_i = P_{i|i-1}^{\diamond}\Pi_{i-1}^{1,1}(I + S_kP_0) + \{(\Pi_{i-1}^{1,1})^{-T} - P_{i|i-1}^{\diamond}\Pi_{i-1}^{1,1}\sum_{l=i}^{k}\widetilde{S}_l\}P_0.$$

Substituting the expressions of X_{k+1} and Y_i into (3.15) and rearranging,

$$\begin{aligned} \hat{x}_{k+1|k} = \underbrace{(\Pi_k^{1,1})^{-T}\sum_{i=0}^{k}(\Pi_{i-1}^{1,1})^T P_{i|i-1}^{\diamond} z_i}_{\triangleq N_k} &+ (\Pi_k^{1,1})^{-T}(I + S_kP_0)^{-T}\cdot \\ \cdot\Big\{\hat{x}_0 + P_0\underbrace{\sum_{i=0}^{k}(\Pi_{i-1}^{1,1})^{-1}z_i}_{\triangleq M_k'} &- P_0\underbrace{\sum_{i=0}^{k}(\sum_{l=i}^{k}\widetilde{S}_l)(\Pi_{i-1}^{1,1})^T P_{i|i-1}^{\diamond} z_i}_{\triangleq M_k''}\Big\}. \end{aligned}$$

Let us first show that $N_k \equiv \hat{x}_{k+1|k}^{\diamond}$. In light of Corollary 3.1, one gets[1]

$$\begin{aligned} N_k &= \sum_{i=0}^{k}(\Pi_k^{1,1})^{-T}(\Pi_{i-1}^{1,1})^T(\Pi_{i-1}^{1,1})^{-T}(\Pi_{i-1}^{2,1})^T z_i \\ &= \sum_{i=0}^{k}(\Pi_k^{1,1})^{-T}\cdot(\Pi_{k-1}^{1,1})^T(\Pi_{k-1}^{1,1})^{-T}\cdot\ldots\cdot(\Pi_i^{1,1})^T(\Pi_i^{1,1})^{-T}\cdot(\Pi_{i-1}^{2,1})^T z_i \\ &= \sum_{i=0}^{k}\prod_{l=i+1}^{k}\Phi_l^{\diamond}A_iL_i^{\diamond}y_i. \end{aligned}$$

[1]Given a sequence of square matrices Φ_l of the same size, we let $\prod_{l=h}^{k}\Phi_l$ be equal to $\Phi_k\cdot\Phi_{k-1}\cdot\ldots\cdot\Phi_h$ if $h \le k$, and to the identity matrix otherwise.

It follows that $N_k = \Phi_k^\diamond N_{k-1} + A_k L_k^\diamond y_k$, with $N_{-1} \triangleq 0$. For $\hat{x}_0 = 0$ and $P_0 = 0$, this recursion is identical to recursion (3.4a). Hence, $N_k \equiv \hat{x}_{k+1|k}^\diamond$. Next, setting $M'_{-1} \triangleq 0$,

$$M'_k = \sum_{i=0}^{k-1} (\Pi_{i-1}^{1,1})^{-1} z_i + (\Pi_{k-1}^{1,1})^{-1} z_k = M'_{k-1} + (\Pi_{k-1}^{1,1})^{-1} z_k.$$

Consider now M''_k. Make use of Corollary 3.1 to get

$$\widetilde{S}_l (\Pi_{i-1}^{1,1})^T = (\Pi_{l-1}^{1,1})^{-1} (I + \Delta_l P_{l|l-1}^\diamond)^{-1} \Delta_l \cdot \prod_{j=i}^{l-1} \Phi_j^\diamond.$$

Substituting the above into the expression of M''_k and rearranging yields

$$\begin{aligned} M''_k &= \sum_{i=0}^{k} \sum_{l=i}^{k} (\Pi_{l-1}^{1,1})^{-1} (I + \Delta_l P_{l|l-1}^\diamond)^{-1} \Delta_l \cdot \prod_{j=i}^{l-1} \Phi_j^\diamond \cdot P_{i|i-1}^\diamond z_i \\ &= \sum_{i=0}^{k-1} \sum_{l=i}^{k-1} (\Pi_{l-1}^{1,1})^{-1} (I + \Delta_l P_{l|l-1}^\diamond)^{-1} \Delta_l \cdot \prod_{j=i}^{l-1} \Phi_j^\diamond \cdot P_{i|i-1}^\diamond z_i + \\ &\quad + (\Pi_{k-1}^{1,1})^{-1} (I + \Delta_k P_{k|k-1}^\diamond)^{-1} \Delta_k \cdot \sum_{i=0}^{k} \prod_{j=i}^{k-1} \Phi_j^\diamond \cdot P_{i|i-1}^\diamond z_i \\ &= M''_{k-1} + (\Pi_{k-1}^{1,1})^{-1} (I + \Delta_k P_{k|k-1}^\diamond)^{-1} \Delta_k \widetilde{M}_k, \end{aligned}$$

with $M''_{-1} \triangleq 0$ and $M_k \triangleq \sum_{i=0}^{k} \prod_{j=i}^{k-1} \Phi_j^\diamond \cdot P_{i|i-1}^\diamond z_i$. In turn,

$$\widetilde{M}_k = P_{k|k-1}^\diamond z_k + \Phi_{k-1}^\diamond \sum_{i=0}^{k-1} \prod_{j=i}^{k-2} \Phi_j^\diamond \cdot P_{i|i-1}^\diamond z_i = P_{k|k-1}^\diamond z_k + \Phi_{k-1}^\diamond \widetilde{M}_{k-1},$$

where $\widetilde{M}_{-1} \triangleq 0$. To conclude, just set $M_k \triangleq M'_k - M''_k$. □

Sequences $\widetilde{M}_k$ and M_k are independent of $\hat{x}_0$ and P_0. Together with $\hat{x}_{k+1|k}^\diamond$, they account for measurements y^k. Under suitable assumptions – in particular, for time-invariant detectable and controllable systems – their expected norms are bounded.

Decomposition formulas for $\hat{x}_{k|k}(\hat{x}_0, P_0)$ and $P_{k|k}(P_0)$ follow easily. Let as usual $\hat{x}_{k|k}^\diamond \triangleq \hat{x}_{k|k}(0,0)$ and $P_{k|k}^\diamond \triangleq P_{k|k}(0)$.

Corollary 3.2 *The Kalman filter and the associated error covariance matrix admit the following decomposition:*

$$\begin{aligned}\hat{x}_{k|k}(\hat{x}_0, P_0) &= \hat{x}^{\diamond}_{k|k} + A_k^{-1}U_k(I + S_kP_0)^{-T}(\hat{x}_0 + P_0M_k),\\ P_{k|k}(P_0) &= P^{\diamond}_{k|k} + A_k^{-1}U_k(I + S_kP_0)^{-T}P_0U_k^TA_k^{-T}.\end{aligned}$$

Proof: By equations (3.3a)–(3.3b), it is simply

$$\begin{aligned}\hat{x}_{k|k} &= A_k^{-1}\hat{x}_{k+1|k}\\ &= A_k^{-1}\hat{x}^{\diamond}_{k+1|k} + A_k^{-1}U_k(I + S_kP_0)^{-T}(\hat{x}_0 + P_0M_k),\end{aligned}$$

where indeed $A_k^{-1}\hat{x}^{\diamond}_{k+1|k} = \hat{x}^{\diamond}_{k|k}$, and

$$\begin{aligned}P_{k|k} &= A_k^{-1}(P_{k+1|k} - Q_k)A_k^{-T}\\ &= A_k^{-1}(P^{\diamond}_{k+1|k} + U_k(\ldots)^{-T}P_0U_k^TQ_k)A_k^{-T}\\ &= A_k^{-1}(P^{\diamond}_{k+1|k} - Q_k)A_k^{-T} + A_k^{-1}U_k(\ldots)^{-T}P_0U_k^TA_k^{-T},\end{aligned}$$

where $A_k^{-1}(P^{\diamond}_{k+1|k} - Q_k)A_k^{-T} = P^{\diamond}_{k|k}$. □

Note that all terms accounting for $\hat{x}_0$ and P_0 involve U_k as a factor. In general, this makes the contribution of the initialization negligible for large k: in the standard time-invariant setting, for instance, $U_k \xrightarrow{k} 0$, whereas all the remaining factors are bounded. Finally observe that the quantities $\hat{x}^{\diamond}_{k|k}$, $P^{\diamond}_{k|k}$, $\hat{x}^{\diamond}_{k+1|k}$ and $P^{\diamond}_{k+1|k}$ may be computed at once by way of the standard algorithm (3.2)–(3.3) initialized by $\hat{x}_0 = 0$, $P_0 = 0$.

3.5 Discussion

In this chapter we have presented an implementation of the Kalman filter which keeps the dependence on the a priori state estimate $\hat{x}_0$ and error covariance matrix P_0 explicit. The decomposition formulas we obtained may be interpreted as a nonlinear version of the superposition principle holding for linear difference equations. The recursive nature of the parameters describing $\hat{x}(\hat{x}_0, P_0)$ and $P(P_0)$ makes this formulation well-suited for applications. Compared to the former arrangement described in [17], the algorithm has been relieved of numerical ill-conditioning and of certain unessential assumptions. Further investigation of its numerical properties and application to several nonlinear settings are possible directions of research.

Chapter 4

Sampled linear switching systems

In this chapter we introduce a jump Markov model with continuous-time dynamics and sampled measurements. A general state estimation problem is stated. The optimal solution is discussed and is shown to relate to the solutions of certain linear estimation problems. A recursive solution to the problems of filtering and prediction at measurement times is derived. The fundamental case of a single switch model is examined in more depth. In this setting, state estimation is further investigated in connection with the more specific problem of fault detection.

4.1 Introduction

In Chapter 2 we observed that the discrete-time model (2.1) may be unsuited for applications where the rate of measurement is slow compared to the rate of switching or to the system's dynamics. This limitation may be overcome if the evolution of the discrete-state is described by a continuous-time Markov chain $q(t)$. By the results of Chapter 1, this allows to model switches as the events τ_k of a Poisson process. Moreover, the dynamics of a continuous state $\xi(t)$ may be expressed in terms of the piecewise constant trajectories of $q(t)$. The nature of measurements is unchanged, but now the instants of switching τ_k are completely unrelated with the measurement times t_k. As for discrete-time jump Markov linear systems, one is interested in the estimation of the states $\xi(t)$ and $q(t)$. In addition, the problem of isolating the switches in between measurements may be posed. In general, the vastity of the discrete-

state trajectories that may be generated complicates the estimation tasks.

In the next section, we will formalize the continuous-type dynamics, sampled measurement model and discuss its fundamental properties. In the following section, we will state the estimation problems we are concerned with. In analogy with discrete-time jump Markov systems, we will illustrate the connection of the problem with a set of conditioned linear estimation problems. Conditional Kalman filtering and a recursive scheme for the estimation of the joint state $(\xi(t), q(t))$ will be introduced in Sections 4.5÷ 4.6 as a natural extension of the tools of Chapter 2. In Sections 4.7 and 4.8, we will focus on the case study of a system subject to a single Markovian switch. A simple characterization of the trajectories of $q(t)$ will lead to a straightforward solution of state estimation and of the related problem of fault detection.

4.2 General model

For a certain index $N \in \mathbb{N}_0$, let $\mathcal{Q} = \{0, \dots, N-1\}$. Assume that we are given the matrix-valued functions

$$\begin{aligned} F &: \mathcal{Q} \to \mathbb{R}^{n\times n} \ : q \mapsto F_q \\ G &: \mathcal{Q} \to \mathbb{R}^{n\times m} : q \mapsto G_q \end{aligned}$$

with $n, m \in \mathbb{N}$. For $t \geq 0$, let $w(t) \in \mathbb{R}^m$ be a zero-mean, normalized white Gaussian process, and let $\xi(t) \in \mathbb{R}^n$ obey the stochastic differential equation

$$\dot{\xi} = F_{q(t)}\xi(t) + G_{q(t)}w(t) \tag{4.1a}$$

with $\xi(0) \sim \mathcal{N}(\hat{\xi}_0, \Sigma_0)$, $\hat{\xi}_0$ and Σ_0 known, uncorrelated with w. Next, assume we are given functions

$$\begin{aligned} H &: \mathcal{Q} \to \mathbb{R}^{p\times n} : q \mapsto H_q \\ K &: \mathcal{Q} \to \mathbb{R}^{p\times r} \ : q \mapsto K_q \end{aligned}$$

with $p, r \in \mathbb{N}$. For $k \in \mathbb{N}_0$, let $v_k \in \mathbb{R}^r$ be a zero-mean, normalized white Gaussian sequence, and let $y_k \in \mathbb{R}^p$ be defined as

$$y_k = H_{q(t_k)}\xi(t_k) + K_{q(t_k)}v_k \tag{4.1b}$$

where

$$\mathcal{T} \triangleq \{t_k\}_{k\in\mathbb{N}_0} \subseteq \mathbb{R}_+$$

is an assigned sequence satisfying $t_k < t_{k+1}$, $t_0 = 0$, and v is uncorrelated with $\xi(0)$ and w. We will always assume that, for any $q \in \mathcal{Q}$,

$$\mathrm{Var}(K_q v_k) = K_q K_q^T > 0.$$

Finally, let $q(t) \in \mathcal{Q}$ be the outcome of a homogeneous Markov chain of assigned (continuous) transition probability matrix $\mathbf{T}(\delta)$ and initial probability distribution p_0. That is, for every $i, j \in \mathcal{Q}$,

$$\mathbf{T}_{i,j}(\delta) = \mathbb{P}[q(t+\delta) = j | q(t) = i] \tag{4.1c}$$

and $p_0(j) \triangleq \mathbb{P}[q(0) = j]$. We will make the further hypothesis that

$$\mathbb{P}[q(t+\delta)|q(t), \xi(0), w, v] = \mathbb{P}[q(t+\delta)|q(t)].$$

As a consequence,

$$\mathbb{P}[q(t+\delta) = j | q(t) = i, \xi(t)] = \mathbf{T}_{i,j}(\delta),$$

i.e. $q(t+\delta)$ is conditionally independent of $\xi(t)$ given $q(t)$.

For all $t \in \mathbb{R}_+$, all $s \geq t$, any $k \in \mathbb{N}_0$ and any $\{s_0, \ldots, s_k\} \subset [0, t)$, the following *separation principle* holds:

$$\mathcal{F}(\xi(s), q(s)|\xi(t), q(t), \xi(s_0), q(s_0), \ldots, \xi(s_k), q(s_k)) = \mathcal{F}(\xi(s), q(s)|\xi(t), q(t)).$$

Therefore, the couple $(\xi(t), q(t))$ will be called the state of the system. In particular, ξ and q will be referred to as the continuous state and the discrete state, respectively.

The set of equations (4.1) is a non-Gaussian switching model which cannot be studied with the standard tools of stochastic linear systems. However, the outcomes of $q(t)$ are, with probability 1, piecewise-constant, right-continuous functions. This was illustrated in Chapter 1. Therefore, for any fixed trajectory $q(t)$, say $\tilde{q}$, equations (4.1a) and (4.1b) form the well-defined time-varying linear state-space model

$$\begin{cases} \dot{\xi}(t) &= F_{\tilde{q}(t)}\xi(t) + G_{\tilde{q}(t)}w(t) \\ y_k &= H_{\tilde{q}(t_k)}\xi(t_k) + K_{\tilde{q}(t_k)}v_k \end{cases}. \tag{4.2}$$

The parameters of this system, which we will refer to as the conditioned system, are fixed in time by $\tilde{q}(t)$. As a consequence, for any $t \geq 0$ and $\ell \in \mathbb{N}_0$, the joint distribution

$$f(\xi(t), y_\ell | \tilde{q})$$

is a multivariate Gaussian.

4.3 State estimation problem

The problem we are interested in is the optimal estimation of the system state from the available measurements. Precisely, let

$$y^k = \{y_0, \dots, y_k\}$$

denote the set of measurements up to index k. We want to compute the min-square-error estimate estimate of the continuous state,

$$\arg\min_z \mathbb{E}[||\xi(t) - z(y^k)||^2], \tag{4.3}$$

and the minimum-probability-of-error estimate of the discrete state,

$$\arg\min_z \mathbb{P}[q(t) \neq z(y^k)], \tag{4.4}$$

where z indicates in both cases a measurable function of data y^k. For any density function $f(\xi(t)|y^k)$, problem (4.3) is equivalent to the minimization of

$$\mathbb{E}[||\xi(t) - z(y_k)||^2|y^k],$$

which results in the conditional expectation of $\xi(t)$ given y^k. Similarly, for any distribution $p(q(t)|y^k)$, problem (4.4) is equivalent to minimizing

$$\mathbb{P}[q(t) \neq z(y^k)|y^k],$$

which is solved by the maximum a posteriori estimate of $q(t)$ given y^k. Thus, we wish to compute

$$\hat{\xi}_{t|k} \triangleq \mathbb{E}[\xi(t)|y^k] \tag{4.5}$$

as an estimate of $\xi(t)$, and

$$\hat{q}_{t|k} \triangleq \arg\max_{j \in \mathcal{Q}} \mathbb{P}[q(t) = j|y^k] \tag{4.6}$$

as an estimate of $q(t)$. As a measure of quality of estimate (4.5), we would also like to compute the conditional variance

$$\Sigma_{t|k} \triangleq \mathbb{E}[\tilde{\xi}_{t|k}\tilde{\xi}_{t|k}^T|y^k] = \mathrm{Var}(\xi(t)|y^k), \tag{4.7}$$

where $\tilde{\xi}_{t|k} \triangleq \xi(t) - \hat{\xi}_{t|k}$ is the estimation error. On the other hand, it would be useful to determine the whole conditional distribution

$$p_{t|k}(j) \triangleq \mathbb{P}[q(t) = j|y^k], \tag{4.8}$$

with $j = 0, \dots, N-1$, from which estimate (4.6) follows. In particular, we are concerned with the cases of filtering, i.e. $t = t_k$, and one-step prediction, i.e. $t = t_{k+1}$.

4.4 Theoretical solution

In more generality, let us consider the computation of

$$f(\xi(t)|y^k),$$

from which expressions (4.5) and (4.7) follow. Where unessential, we will simply write ξ instead of $\xi(t)$. By the total probability law,

$$f(\xi|y^k) = \int d\mathbb{P}[\tilde{q}|y^k] f(\xi|y^k, \tilde{q}), \tag{4.9}$$

where $\tilde{q}$ indicates one trajectory of q, and integration is taken over a set of trajectories of $q(t)$ of probability 1. Since

$$f(\xi|y^k, \tilde{q}) \propto f(\xi, y^k|\tilde{q}),$$

it must hold that, for certain parameters $\hat{\xi}^{\tilde{q}}$ and $\Sigma^{\tilde{q}}$,

$$f(\xi|y^k, \tilde{q}) \sim \mathcal{N}(\hat{\xi}^{\tilde{q}}, \Sigma^{\tilde{q}}). \tag{4.10}$$

Thus, equation (4.9) represents an average of Gaussian densities weighted by the a posteriori distribution of process $q(t)$ given y^k. Note that

$$d\mathbb{P}[\tilde{q}|y^k] \propto f(y^k|\tilde{q}) d\mathbb{P}[\tilde{q}], \tag{4.11}$$

i.e. weighting only depends on the statistics of the measurements of the conditioned system (4.2) and on the a priori description of $q(t)$. In general, density $f(\xi|y^k)$ is non-Gaussian, nor it can written in terms of a finite number of parameters.

For every possible trajectory $\tilde{q}$, the mean of (4.10) is given by

$$\hat{\xi}^{\tilde{q}} = \arg\min_z \mathbb{E}[||\xi - z(y^k)||^2|\tilde{q}],$$

i.e. it may be found by solving the continuous state estimation problem associated to system (4.2). To verify this, it suffices to apply Proposition 1.16 to problem (4.3) using the conditioned distribution $f(\xi|y^k, \tilde{q})$. Correspondingly, the covariance matrix of (4.10) is given by

$$\Sigma^{\tilde{q}} = \mathbb{E}[\tilde{\xi}^{\tilde{q}} \tilde{\xi}^{\tilde{q}T}|\tilde{q}],$$

where $\tilde{\xi}^{\tilde{q}}$ is the estimation error $\xi - \hat{\xi}^{\tilde{q}}$.

Therefore, the solution of the continuous state estimation problem is essentially an average of the solutions of the linear estimation problems associated to every possible trajectory $\tilde{q}$. It follows from (4.9) that, at a generic time t, estimate (4.5) is given by

$$\hat{\xi} = \int d\mathbb{P}[\tilde{q}|y^k]\hat{\xi}^{\tilde{q}}, \tag{4.12}$$

i.e. it is itself a weighted average of the optimal estimates conditioned on $\tilde{q}$. Similarly, covariance matrix (4.7) is given by

$$\Sigma = \int d\mathbb{P}[\tilde{q}|y^k]\Sigma^{\tilde{q}} + \{\int d\mathbb{P}[\tilde{q}|y^k]\hat{\xi}^{\tilde{q}}\hat{\xi}^{\tilde{q}T} - \hat{\xi}\hat{\xi}^T\} \tag{4.13}$$

where the extra term in braces is due to the nonlinearity of the variance operator.

Turning to the discrete state estimation problem, consider the computation of (4.8). For any $j \in \mathcal{Q}$, it holds that

$$p_{t|k}(j) = \int_{\{\tilde{q}:\tilde{q}(t)=j\}} d\mathbb{P}[\tilde{q}|y^k] \propto \int_{\{\tilde{q}:\tilde{q}(t)=j\}} d\mathbb{P}[\tilde{q}]f(y^k|\tilde{q}), \tag{4.14}$$

where (4.11) has been used. Hence, in principle, the estimation of q at any time t depends only on the a priori statistics of process $q(t)$ itself and on the statistics of the data y^k given $\tilde{q}$.

In general, the actual computation of integrals (4.12)÷(4.14) is impracticable, mainly due to the large amount of trajectories of $q(t)$ to be considered. However, for certain definitions of the transition function (4.1c), explicit expressions may be found for (4.11) as well as for $\hat{\xi}^{\tilde{q}}$ and $\Sigma^{\tilde{q}}$. In this case, integrations may be either carried out exactly or approximated numerically with arbitrary accuracy.

4.5 Conditioned filtering and prediction

For any fixed $\tilde{q}$, the linearity of (4.2) allows to compute $\hat{\xi}^{\tilde{q}}$ and $\Sigma^{\tilde{q}}$ very efficiently. For t equal to t_k and t_{k+1}, in particular, they may be computed at once by way of conditional Kalman filtering and prediction.

Proposition 4.1 *The following recursion holds: (i) Measurement update: compute*

$$\begin{aligned}\hat{\xi}^{\tilde{q}}_{t_k|k} &= \hat{\xi}^{\tilde{q}}_{t_k|k-1} + L^{\tilde{q}}_k[y_k - H_{\tilde{q}(t_k)}\hat{\xi}^{\tilde{q}}_{t_k|k-1}],\\ \Sigma^{\tilde{q}}_{t_k|k} &= \Sigma^{\tilde{q}}_{t_k|k-1} - L^{\tilde{q}}_k H_{\tilde{q}(t_k)}\Sigma^{\tilde{q}}_{t_k|k-1},\end{aligned}$$

with $L_k^{\tilde{q}} = \Sigma_{t_k|k-1}^{\tilde{q}} H_{\tilde{q}(t_k)}^T [H_{\tilde{q}(t_k)} \Sigma_{t_k|k-1}^{\tilde{q}} H_{\tilde{q}(t_k)}^T + K_{\tilde{q}(t_k)} K_{\tilde{q}(t_k)}^T]^{-1}$*; (ii) Time update: solve, from* $t = t_k$ *to* $t = t_{k+1}$*, the differential equations*

$$\begin{aligned} d\hat{\xi}_{t|k}^{\tilde{q}}/dt &= F_{\tilde{q}(t)} \hat{\xi}_{t|k}^{\tilde{q}}, \\ d\Sigma_{t|k}^{\tilde{q}}/dt &= F_{\tilde{q}(t)} \Sigma_{t|k}^{\tilde{q}} + \Sigma_{t|k}^{\tilde{q}} F_{\tilde{q}(t)}^T + G_{\tilde{q}(t)} G_{\tilde{q}(t)}^T. \end{aligned}$$

The initialization of the recursion is given by $\hat{\xi}_{t_0|-1}^{\tilde{q}} = \hat{\xi}_0$ *and* $\Sigma_{t_0|-1}^{\tilde{q}} = \Sigma_0$.

The differential equations of the time update step may be used to compute $\hat{\xi}_{t|k-1}^{\tilde{q}}$ and $\Sigma_{t|k-1}^{\tilde{q}}$ at any time $t \geq t_{k-1}$. Due to the nature of process $q(t)$, carrying out this step amounts to solving differential equations with piecewise constant coefficients.

The a priori statistics of measurements y^k given $\tilde{q}$ may also be computed based on the equations of the linear system (4.2). However, it proves convenient to exploit the results of Proposition 4.1.

Proposition 4.2 *The a priori density function of measurements* y^k *given* $\tilde{q}$ *may be computed iteratively as follows:*

$$f(y^k|\tilde{q}) = f(y_k|y^{k-1}, \tilde{q}) f(y^{k-1}|\tilde{q}), \tag{4.15}$$

with

$$f(y_k|y^{k-1}, \tilde{q}) = \mathcal{N}(H_{\tilde{q}(t_k)} \hat{\xi}_{t_k|k-1}^{\tilde{q}}, H_{\tilde{q}(t_k)} \Sigma_{t_k|k-1}^{\tilde{q}} H_{\tilde{q}(t_k)}^T + K_{\tilde{q}(t_k)} K_{\tilde{q}(t_k)}^T). \tag{4.16}$$

Iterations are initialized by

$$f(y_0|y^{-1}) = \mathcal{N}(H_{\tilde{q}(t_0)} \hat{\xi}_0, H_{\tilde{q}(t_0)} P_0 H_{\tilde{q}(t_0)}^T + K_{\tilde{q}(t_0)} K_{\tilde{q}(t_0)}^T).$$

Proof: Equality (4.15) is a simple application of Bayes'rule. Expression (4.16) follows from the measurement equation of (4.2) as a consequence of the conditional independence of v_k and y^{k-1} given $\tilde{q}$. □

Together, Propositions 4.1 and 4.2 yield an iterative method for the computation of $f(y^k|\tilde{q})$.

4.6 Recursive approach

In analogy to the discrete-time case, one may solve estimation at measurement times by computing (4.5) and (4.6) recursively. Following the previous

section, consider the computation of the a posteriori density of $\xi(t_\ell)$. By the total probability law,

$$f(\xi(t_\ell)|y^k) = \sum_{j\in\mathcal{Q}} f(\xi(t_\ell)|q(t_\ell)=j, y^k) p_{t_\ell|k}(j).$$

Thus, computing $f(\xi(t_\ell)|y^k)$ and $p_{t_\ell|k}(\cdot)$, $\ell = k, k+1$, is equivalent to computing the likelihood function

$$f_{t_\ell|k}(j) \triangleq f(\xi(t_\ell)|q(t_\ell)=j, y^k),$$

and $p_{t_\ell|k}(j)$, for $j = 0, \ldots, N-1$.

Proposition 4.3 *For $\ell = k$ and $\ell = k+1$, $p_{t_\ell|k}$ and $f_{t_\ell|k}$ obey the following two-step recursion on index k:*

1. *Measurement update: for $j = 0, \ldots, N-1$,*

$$\begin{aligned} p_{t_k|k}(j) &\propto f(y_k|y^{k-1}, q(t_k)=j) p_{t_k|k-1}(j) \\ f_{t_k|k}(j) &\propto f(y_k|\xi(t_k), q(t_k)=j) f_{t_k|k-1}(j) \end{aligned}$$

2. *Time update: for $j = 0, \ldots, N-1$ and $\delta_k \triangleq t_{k+1} - t_k$,*

$$\begin{aligned} p_{t_{k+1}|k}(j) &= \sum_{i\in\mathcal{Q}} \mathbf{T}_{i,j}(\delta_k) p_{t_k|k}(i) \\ f_{t_{k+1}|k}(j) &= \sum_{i\in\mathcal{Q}} f(\xi(t_{k+1})|q(t_k)=i, q(t_{k+1})=j, y^k) \frac{p_{t_k|k}(i)}{p_{t_{k+1}|k}(j)} \mathbf{T}_{i,j}(\delta_k) \end{aligned}$$

The recursion is initialized by $p_{t_0|-1} \triangleq p_0$ and $f_{t_0|-1} \triangleq \mathcal{N}(\hat{\xi}_0, \Sigma_0)$.

Proof: Identical to that of Proposition 2.4, provided the substitution of x_ℓ, q_ℓ, $\pi_{q_k,q_{k+1}}$ with ξ_{t_ℓ}, $q(t_\ell)$ and $\mathbf{T}_{i,j}(\delta_k)$. □

This result parallels Proposition 2.4, and the structure of the recursion is again that of Figure 2.2. Quantities $f(y_k|y^{k-1}, q(t_k))$ and $f(y_k|\xi(t_k), q(t_k))$ appearing in step 1 are promptly computed by equations similar to (2.7) and (2.8). On the contrary, the computation of $f(\xi(t_{k+1})|q(t_k), q(t_{k+1}), y^k)$ is not at all trivial. Take $q(t_k) = i$ and $q(t_{k+1}) = j$. Let $\tilde{q}_k(t)$, $t_k \le t \le t_{k+1}$, denote a portion of trajectory $q(t)$ such that $\tilde{q}_k(t_k) = i$ and $\tilde{q}_k(t_{k+1}) = j$. Then the above term is given by the integral mixing

$$\int d\mathbb{P}[\tilde{q}_k|q(t_k)=i, q(t_{k+1})=j, y^k] f(\xi(t_{k+1})|\tilde{q}_k, y^k), \tag{4.17}$$

extended over all possible such $\tilde{q}_k$. For every fixed $\tilde{q}$, density $f(\xi(t_{k+1})|\tilde{q}_k, y^k)$ may be deduced from $f_{t_k|k}(i)$ by solving the diffusion associated to the linear, time-varying equation (4.1a), see [36]. On the other hand, the weighting term may be computed for all $\tilde{q}_k$ in terms of $p_{t_k|k}$ and $\mathbf{T}$.

From a practical standpoint, applying this recursion to the calculation of (4.5) is critical. The computation of (4.17) suffers from limitations analogous to that of (4.12), and needs be carried out at each step k. Then, one further integration, namely,

$$\hat{\xi}_{t_\ell|k} = \int d\xi(t_\ell)\xi(t_\ell)f(\xi(t_\ell)|y^k)$$

is required to get $\hat{\xi}_{t_\ell|k}$ from $f(\xi(t_\ell)|y^k)$. On the other hand, implicit in the scheme is the joint solution of both the discrete state and the continuous state estimation problem. This gives a deeper understanding of the problem at hand and might be the basis for the development of approximate recursive estimation algorithms.

4.7 Single switch model

For certain choices of the transition matrix $\mathbf{T}$, the complexity of the estimation problem gets drastically reduced. This is the case, in particular, when all but one of the states of process $q(t)$ are absorbing. Let λ_j, $j = 1, \ldots, N-1$ be assigned real parameters such that $\lambda_j > 0$. Let process $q(t)$ have generator matrix

$$\mathbf{G} = \left[\begin{array}{c|c} -\Lambda & \lambda_1 \cdots \lambda_j \cdots \lambda_{N-1} \\ \hline 0 & \\ \vdots & 0 \\ 0 & \end{array}\right], \tag{4.18}$$

where $\Lambda \triangleq \lambda_1 + \ldots + \lambda_{N-1}$. Then, by the equation $\mathbf{T}(\delta) = e^{\mathbf{G}\delta}$, the transition matrix of $q(t)$ takes the form

$$\mathbf{T}(\delta) = \left[\begin{array}{c|c} e^{-\Lambda\delta} & \cdots \frac{\lambda_j}{\Lambda}\left(1 - e^{-\Lambda\delta}\right) \cdots \\ \hline 0 & \\ \vdots & I \\ 0 & \end{array}\right]. \tag{4.19}$$

Note that, for any $i \in \mathcal{Q} \setminus \{0\}$, any $j \in \mathcal{Q} \setminus \{i\}$ and any $\delta > 0$, $\mathbf{T}_{i,j}(\delta) = 0$.

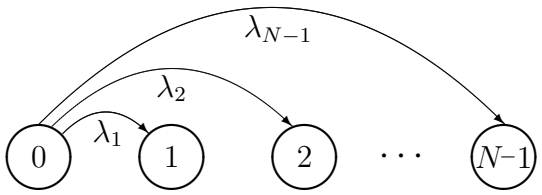


Figure 4.1: Graphical representation of Markov process (4.19).

It follows that

$$\mathbb{P}[q(t+\delta) \neq i | q(t) = i] = 0.$$

In simple terms, process $q(t)$ is not allowed to jump from a state $i \in \mathcal{Q} \setminus \{0\}$ to a different one, that is, states $1, \ldots, N-1$ are absorbing. On the other hand, the probability of eventually jumping from state 0 to any of the states $j \in \mathcal{Q} \setminus \{0\}$ is given by

$$\lim_{\delta \to +\infty} 1 - \mathbf{T}_{0,0}(\delta) = 1. \tag{4.20}$$

A pictorial view of this setting is given in Figure 4.1.

Proposition 4.4 *A continuous-time Markov chain $q(t)$, $t \geq 0$, with transition probability matrix (4.19) is of the form*

$$q(t) = \begin{cases} 0, & t < t^\star \\ q^\star, & t \geq t^\star \end{cases}. \tag{4.21}$$

For $s \geq 0$ and $j \in \mathcal{Q} \setminus \{0\}$, the joint probability distribution of the random variables $t^\star$ and $q^\star$ is given by

$$\mathcal{F}_{t^\star, q^\star}(s, j) \triangleq \mathbb{P}[t^\star < s, q^\star = j] = \frac{\lambda_j}{\Lambda}(1 - e^{-\Lambda s}) p_0(0) + p_0(j). \tag{4.22}$$

Proof: Due to the absorbing property of states $Q \setminus \{0\}$, at most one jump from state 0 to a state $q^\star \in Q \setminus \{0\}$ is allowed. Denote with $t^\star$ the time at which this jump occurs. The following chain of equalities hold:

$$\begin{aligned} \mathbb{P}[t^\star < s, q^\star = j] &= \mathbb{P}[q(s) = j] = \mathbb{P}[q(s) = j | q(0) = 0]\mathbb{P}[q(0) = 0] + \\ &+ \mathbb{P}[q(s) = j | q(0) = j]\mathbb{P}[q(0) = j] = \mathbf{T}_{0,j}(s) p_0(0) + \mathbf{T}_{j,j}(s) p_0(j). \end{aligned}$$

□

Therefore, the outcomes of $q(t)$ are completely characterized by the *switching time* $t^\star$ and the *final state* $q^\star$.

Corollary 4.1 *(i) For $s \geq 0$ and $j \in \mathcal{Q}\setminus\{0\}$, the following expressions hold:*

$$\mathcal{F}_{t^\star}(s) \triangleq \mathbb{P}[t^\star < s] = 1 - p_0(0)e^{-\Lambda t},$$
$$p_{q^\star}(j) \triangleq \mathbb{P}[q^\star = j] = \frac{\lambda_j}{\Lambda} p_0(0) + p_0(j).$$

(ii) In the case $p_0(0) = 1$, the "density" functions

$$f_{t^\star,q^\star}(s,j) \triangleq \frac{\partial}{\partial s}\mathcal{F}_{t^\star,q^\star}(s,j) \tag{4.23}$$
$$f_{t^\star}(s) \triangleq \frac{\partial}{\partial s}\mathcal{F}_{t^\star}(s) \tag{4.24}$$

are well defined over $s \geq 0$, and it holds that

$$f_{t^\star,q^\star}(s,j) = \lambda_j e^{-\Lambda s},$$
$$f_{t^\star}(s) = \Lambda e^{-\Lambda s},$$
$$p_{q^\star}(j) = \frac{\lambda_j}{\Lambda}.$$

In particular, $t^\star$ and $q^\star$ are independent.

Proof: (i) The results follow by marginalization of $\mathcal{F}_{t^\star,q^\star}(s,j)$ with respect to $t^\star$ and $q^\star$, in the order. (ii) For $p_0(0) = 1$, $\mathcal{F}_{t^\star,q^\star}(0,j) = 0$ for all $j \in \mathcal{Q}\setminus\{0\}$, i.e. the probability mass of the joint variable $(t^\star, q^\star)$ is concentrated in $s \geq 0$. Thus, density functions are correctly defined in this domain. The expressions of $f_{t^\star}$ and $f_{t^\star,q^\star}$ are found by differentiation of $\mathcal{F}_{t^\star}$ and $\mathcal{F}_{t^\star,q^\star}$ w.r.t. s, whereas $p_{q^\star}$ just needs be evaluated for $p_0(0) = 1$. Independence of $t^\star$ and $q^\star$ may be verified by inspection. □

Thus, if $p_0(0) = 1$, $t^\star \sim \mathcal{E}(\Lambda)$. Note that the exponential distribution enjoys the property of being memoryless. This will become apparent in the applications to state estimation. Moreover, parameters λ_j are naturally interpreted as the relative probability of states $j \in \mathcal{Q}\setminus\{0\}$. Otherwise, if $p_0(j) > 0$ for some $j \in \mathcal{Q}\setminus\{0\}$, the distribution of $(t^\star, q^\star)$ gets slightly modified in accordance with the probability that process $q(t)$ already starts from an absorbing state. For the purpose of applications, this complication does not add much insight and is not worth considering. Therefore, the assumption that $p_0(0) = 1$ will always be maintained in the sequel.

Let us now move to system (4.1). In light of Proposition 4.4, the model may be rewritten as

$$\begin{cases} \dot{\xi}(t) &= \widehat{F}(t)\xi(t) + \widehat{G}(t)w(t) \\ y_k &= \widehat{H}(t_k)\xi(t_k) + \widehat{K}(t_k)v_k \end{cases} \tag{4.25}$$

where, for $s \geq 0$,

$$(\widehat{F}, \widehat{G}, \widehat{H}, \widehat{K})(s) = \begin{cases} (F_0, G_0, H_0, K_0), & s < t^\star \\ (F_{q^\star}, G_{q^\star}, H_{q^\star}, K_{q^\star}), & s \geq t^\star \end{cases}.$$

The system starts in the linear mode associated with $q = 0$ and jumps at time $t^\star$ to the linear mode associated with $q = q^\star$. This model is well suited for (but not restricted to) the description of a real-world system subject to abrupt faults, in which case the initial mode 0 corresponds to nominal operating conditions, whereas modes 1 through $N - 1$ represent a set of possible faulty behaviors. In the present context, the interest is focused on the detection of the fault, i.e. on the estimation of $t^\star$ and $q^\star$ from the available measurements. For this reason, in a loose sense, the expression "fault detection" is used throughout this work to indicate either the model resulting from (4.19) or the relevant estimation problem.

Remark. In certain situations, it is reasonable to think of a system fault as depending on two factors: impulsive events stressing the system at unknown times, and the intrinsic robustness of the system with respect to one such event. An immediate way of modeling the occurrence of shocks is by way of random, memoryless arrival times τ_k. On the other hand, based on suitable test data, the response of the system to a shock may be described in terms of the probability $\pi_{0,j}$ that the event will result in a fault of type j, for a known set of typical faults indexed by $j = 1, \ldots, N - 1$, with $\pi_{0,0}$ indicating the probability of sticking to the normal operating mode. Therefore, uniform chains naturally fit the description of a system subject to faults. Moreover, their equivalence to continuous-time Markov chains supports the use of this tool for modeling systems undergoing abrupt changes. □

4.8 State estimation and fault detection

In the setting of fault detection, it is convenient to exploit the characterization of $q(t)$ in terms of $t^\star$ and $q^\star$. First consider the continuous state estimation problem. Equation (4.9) specializes to

$$f(\xi|y^k) = \int d\mathbb{P}[t^\star, q^\star|y^k] f(\xi|y^k, t^\star, q^\star).$$

Next, consider the problem of estimating the discrete state. Since for every $j \in \mathcal{Q} \setminus \{0\}$ the event $\{q(t) = j\}$ is equivalent to $\tilde{Q} \triangleq \{t^\star \leq t, q^\star = j\}$,

equation (4.14) becomes

$$p_{t|k}(j) = \int_{\tilde{Q}} d\mathbb{P}[t^\star, q^\star | y^k].$$

On the other hand, it holds that

$$p_{t|k}(0) = 1 - \sum_{j=1}^{N-1} p_{t|k}(j).$$

Let us define, for $j \in \mathcal{Q} \setminus \{0\}$ and $s \geq 0$,

$$\begin{aligned} f^k_{t^\star,q^\star}(s,j) &\triangleq \frac{\partial}{\partial s}\mathbb{P}[t^\star \leq s, q^\star = j | y^k], \\ &\propto f(y^k | t^\star = s, q^\star = j) f_{t^\star,q^\star}(s,j). \end{aligned} \tag{4.26}$$

with normalization factor

$$f(y^k) = \sum_{j=1}^{N-1} \int_0^{+\infty} f(y^k | t^\star = s, q^\star = j) f_{t^\star,q^\star}(s,j) ds. \tag{4.27}$$

For $k = -1$, (4.26) reduces to (4.23).

Proposition 4.5 *The following expressions hold*

$$\begin{aligned} f(\xi | y^k) &\propto \sum_{j=1}^{N-1} \int f(\xi | y^k, t^\star = s, q^\star = j) f(y^k | t^\star = s, q^\star = j) f_{t^\star,q^\star}(s,j) ds, \\ p_{t|k}(j) &\propto \int_0^t f(y^k | t^\star = s, q^\star = j) f_{t^\star,q^\star}(s,j) ds, \end{aligned}$$

for every $j \in \mathcal{Q} \setminus \{0\}$, with normalization factor given by (4.27).

Proof: The expression of $f(\xi | y^k)$ may be rewritten as

$$\sum_{j=1}^{N-1} \int f(\xi | y^k, t^\star = s, q^\star = j) f^k_{t^\star,q^\star}(s,j) ds,$$

to which (4.26) applies. Similarly, for any $j \in \mathcal{Q} \setminus \{0\}$ and $\tilde{Q}$ as above, the expression of $p_{t|k}(j)$ may be recast as

$$\int_{\tilde{Q}} f^k_{t^\star,q^\star}(s,j) ds.$$

Again, equation (4.26) applies. □

Therefore, the joint state estimation problem may be restated as that of computing, for any $t^\star$ and any $q^\star$, the quantities

$$f(\xi|y^k, t^\star, q^\star) = \mathcal{N}(\hat{\xi}^{t^\star,q^\star}, \Sigma^{t^\star,q^\star}), \tag{4.28a}$$

i.e. the best conditioned estimate of ξ, and

$$f(y^k|t^\star, q^\star), \tag{4.28b}$$

i.e. the a priori conditioned distribution of the data. Observe that Propositions (4.1) and (4.2) may be rewritten in terms of $t^\star$ and $q^\star$ as well. This yields an iterative method for the computation of (4.28a), with $\xi = \xi(t_k)$ and $\xi = \xi(t_{k+1})$, and of (4.28b).

It is also of interest to consider a more specific estimation problem, which we refer to as fault detection. Precisely, we wish to compute the minimum mean-squared error estimate of $t^\star$, i.e.

$$\arg\min_z \mathbb{E}[(t^\star - z(y^k))^2],$$

and the minimum-probability-of-error estimate of $q^\star$, i.e.

$$\arg\min_z \mathbb{P}[q^\star \neq z(y^k)],$$

where z indicates a measurable function of data y^k. By the arguments of Section 1.6, the desired estimator of $t^\star$ turns out to be

$$\hat{t}_k^\star \triangleq \mathbb{E}[t^\star|y^k], \tag{4.29}$$

whereas that of $q^\star$ is given by

$$\hat{q}_k^\star \triangleq \arg\max_{j\in\mathcal{Q}\setminus\{0\}} \mathbb{P}[q^\star = j|y^k].$$

In order to evaluate the quality of these estimates, we would also like to compute the estimate error variance

$$\mathbb{E}[(t^\star - \hat{t}_k^\star)^2|y^k] = \mathrm{var}(t^\star|y^k) \tag{4.30}$$

In more generality, we shall consider the density function

$$f_k^\star(s) \triangleq \frac{\partial}{\partial s}\mathbb{P}[t^\star \leq s|y^k],$$

from which (4.29) and (4.30) follow. Similarly, in place of $\hat{q}_k^\star$, we shall consider the distribution

$$p_k^\star(j) = \mathbb{P}[q^\star = j|y^k].$$

Proposition 4.6 *For every $s \geq 0$ and $j \in \mathcal{Q} \setminus \{0\}$, it holds that*

$$p_k^\star(j) \propto \int_0^{+\infty} f(y^k | t^\star = s, q^\star = j) f_{t^\star, q^\star}(s, j) ds,$$
$$f_k^\star(s) \propto \sum_{j=1}^{N-1} f(y^k | t^\star = s, q^\star = j) f_{t^\star, q^\star}(s, j),$$

where the normalization factor is given by (4.27).

Proof: The result is obtained by marginalization of (4.26). □

Of course, this problem is strictly related to the discrete state estimation problem. Indeed, the above quantities may also be computed in terms of the distribution $p_{t|k}$. In practice, though, both problems depend on the computation of the (4.28). In this sense, they may be considered equivalent.

4.9 Discussion

In this chapter we have introduced the state estimation problem for jump Markov linear systems with continuous dynamics and sampled measurements. The problem has been studied in full generality and the theoretical properties of the optimal solution have been discussed. Also, a recursive solution of filtering and prediction at measurement time instants has been illustrated. The simple scenario of a single Markovian switch has been investigated in more depth. The general results have been particularized to this case. Moreover, a specific estimation problem, referred to as fault detection, has been solved. The substantial equivalence of fault detection and discrete state estimation has been pointed out.

From an analytic point of view, it is clear how the optimal state estimates should be computed. This is especially evident in the case of a single switch model, thanks to a very convenient characterization of the discrete trajectories $q(t)$. On a practical ground, however, several issues need to be addressed. First of all, most of the random variables into play are distributed according a continuous mixture of Gaussian densities. In general, this kind of distribution is intractable, unless some approximations are introduced. The second question to address is how to carry out integrations. Explicit solutions are not conceivable, let apart certain trivial instances. Thus one has to resort to numerical quadrature. Moreover, when the solution has to

be computed online, elementary methods such as a priori gridding are prone to fail, since the functions to be integrated are data-dependent. Therefore, using adaptive numerical strategies is almost mandatory. To do this, one fundamental requirement is that the integrand terms be in explicit form w.r.t. the integration variable. Finally, in order to make the solution practicable, the computational complexity of the estimators should be of low polynomial order. Needless to say, further approximations are required to achieve this.

Valuable answers to these issues may be given for the fault detection problem. This will be done in Chapter 5, where an online algorithm for the solution of the state estimation problem is derived. The theoretical tools of Chapter 3 will be used.

Chapter 5

Fault detection algorithms

State estimation procedures for sampled switching systems are studied in depth. We first discuss numerical issues arising in the implementation of the estimators, and establish the basic requirements that the implementation must fulfill. Hence, the problem of conditioned Kalman filtering on the sampled model is reduced to parametric discrete-time filtering over an equivalent discrete-time model. The results of Chapter 3 are then applied to make the dependency of the filter on the parameters explicit. This is used to formulate numerical estimation procedures of arbitrary accuracy and reasonable complexity. Linear complexity approximations are proposed based on a limited-memory approximation. Numerical results are illustrated.

5.1 Introduction

By the results of Chapter 4, it is quite clear that the problem of fault detection given data y^k may be reduced to the computation of integrals such as

$$\int g_k(s, j, y^k) f_{t^\star, q^\star}(s, j) ds, \tag{5.1}$$

where the values taken by $t^\star$ and $q^\star$, namely s and j, play the role of the fixed trajectory $\tilde{q}$ and enter g_k only through

$$\begin{array}{l} \hat{x}^{s,j}_{\ell+1|\ell} = \hat{\xi}^{\tilde{q}}_{t_{\ell+1}|\ell} \\ P^{s,j}_{\ell+1|\ell} = \Sigma^{\tilde{q}}_{t_{\ell+1}|\ell} \end{array} \tag{5.2}$$

with $\ell = 0, \ldots, k-1$. For the time being, equation (5.2) is a mere change of notation. For any fixed values of s and j, quantities $\hat{x}^{s,j}_{\ell+1|\ell}$ and $P^{s,j}_{\ell+1|\ell}$ may

be computed iteratively in the index ℓ. This amounts to apply the Kalman filter of Proposition 4.1 to the system (4.25).

We are now concerned with the online calculation of (5.1) for increasing values of the index k. Since the closed-form solution is in general not known, it is necessary to resort to some numerical quadrature method, i.e. an approximation such as

$$\int g_k(s, j, y^k) f_{t^\star,q^\star}(s, j) ds \simeq \sum_{l=0}^{L} a_l^k g_k(s_l^k, j, y^k) f_{t^\star,q^\star}(s_l^k, j), \tag{5.3}$$

where the sampling points s_l^k and the weights a_l^k depend on the method used.

If the sample points are independent of k, i.e. $\{s_l^k\} \equiv \{\bar{s}_l\}$, the computation of $g_k(j, \bar{s}_l, y^k)$ for increasing values of k is immediate: for every l, one only needs to update the conditioned estimates $\hat{x}_{k|k-1}^{\bar{s}_l,j}$ and $P_{k|k-1}^{\bar{s}_l,j}$. This leads to a very fast recursive integration strategy. Unfortunately, fixing the sampling grid $\{\bar{s}_l\}$ yields very poor approximations, unless a very large number of samples L is used.

More effective approximations may be obtained by the use of adaptive techniques. That is, for any index k, the initial choice of $\{s_l^k\}$ should be refined driven by the actual form of the integrand of (5.1). In principle, this allows to achieve better approximations with fewer samples. However, sampling becomes data-dependent, i.e. $\{s_l^k\} = \{s_l^k(y^k)\}$. As a consequence, in general,

$$\{s_l^{k+1}(y^{k+1})\} \neq \{s_l^k(y^k)\}.$$

In this case, for $s = s_l^{k+1}$, quantities $\hat{x}_{k|k-1}^{s,j}$ and $P_{k|k-1}^{s,j}$ are not available from the previous step. Hence, computing $\hat{x}_{k+1|k}^{s,j}$ and $P_{k+1|k}^{s,j}$ at $s = s_l^{k+1}$, $l = 1, \ldots, L$ requires starting L "new" Kalman filters and running them up to step $k+1$. This is to say, the evaluation of g_k at sample points is no longer recursive.

The gap between adaptive quadrature and recursive Kalman filtering may be overcome by rearranging the computation of (5.2). Assume we are given two parametric functions ψ, ϕ such that

$$\begin{aligned} \hat{x}_{\ell+1|\ell}^{s,j} &= \psi\{\Xi_\ell(j, y^\ell), s\}, \\ P_{\ell+1|\ell}^{s,j} &= \phi\{\Theta_\ell(j), s\}, \end{aligned} \tag{5.4a}$$

where Ξ_ℓ and Θ_ℓ are vectors of parameters independent of s. Assume that the update of these parameters is independent of s. That is, for suitable

transition functions Ψ_ℓ and Φ_ℓ,

$$\begin{aligned} \Xi_{\ell+1} &= \Psi_\ell\{\Xi_\ell, j, y^\ell\}, \\ \Theta_{\ell+1} &= \Phi_\ell\{\Theta_\ell, j\}. \end{aligned} \tag{5.4b}$$

In this case, at any step k, the evaluation of g_k, i.e. of $\hat{x}^{s,j}_{\ell+1|\ell}$ and $P^{s,j}_{\ell+1|\ell}$, $\ell = 0, \dots, k-1$, at an arbitrary sampling point s_l amounts to substituting $s = s_l$ and the parameters Ξ_ℓ and Θ_ℓ in the expressions of ψ and ϕ. On the other hand, the computation of $\hat{x}^{s,j}_{k+1|k}$ and $P^{s,j}_{k+1|k}$ from $\hat{x}^{s,j}_{k|k-1}$ and $P^{s,j}_{k|k-1}$ is replaced by the computation of the parameters Ξ_{k+1} and Θ_{k+1} from Ξ_k and Θ_k. In simple words, the recursive part is "split away" from the evaluation part: instead of updating the expressions (5.2) at fixed values of $t^\star$, "the whole functions" are updated. As a consequence, *any* approximation of the form (5.3) shall just require the recursive update of Ξ_k and Θ_k.

In the next sections, we will show that expressions similar to (5.4a) may indeed be derived, along with the update formulas (5.4b). To do this, we will apply the tools of Chapter 3 to a suitable discretization of system (4.25). Hence, we will build algorithms for the adaptive numerical solution of fault detection as well as of continuous state filtering and prediction at measurement instants. With the online implementation in mind, one section will be devoted to suboptimal algorithms of lower computational complexity. Finally, we will illustrate numerical results showing the effectiveness of our method.

5.2 Parametric discretization

Let the values of $t^\star$ and $q^\star$ be fixed. For any $k \in \mathbb{N}_0$, define the sampled state

$$x_k \triangleq \xi(t_k). \tag{5.5}$$

For $k \in \mathbb{N}_0$, consider the model

$$\begin{cases} x_{k+1} = A_k(t^\star, q^\star)\, x_k + u_k \\ y_k \;\; = C_k(t^\star, q^\star)\, x_k + D_k(t^\star, q^\star)\, v_k \end{cases}. \tag{5.6}$$

We wish to choose the parameters $A_k(t^\star, q^\star)$, $C_k(t^\star, q^\star)$, $D_k(t^\star, q^\star)$ and the joint statistics of x_0, $\{u_k\}$, $\{v_k\}$ that make the discrete-time system (5.6) equivalent to the sampled measurement system (4.25) at the sample times $\mathcal{T} = \{t_k\}$. Precisely, we want to guarantee that the joint statistics of $\{y_k\}$ and $\{x_k\}$ be the same as the joint statistics of $\{y_k\}$ and $\{\xi(t_k)\}$.

In the rest of the section, we will make the following assumption.

Assumption 5.1 *The spectra of F_j and of $-F_j$ are disjoint, $\forall j \in \mathcal{Q}$.*

Let us introduce a couple of preliminary results. The first is a classical result in the theory of the Lijapunov equations.

Lemma 5.1 ([8], pp. 203–204) *For any $j \in \mathcal{Q}$, the equation*

$$F_j J_j + J_j F_j^T = -G_j G_j^T \tag{5.7}$$

admits a unique (symmetric) solution in $J_j \in \mathbb{R}^{n\times n}$.

Next, for arbitrary values $a, b \in \mathbb{R}$ and $j \in \mathcal{Q}$, consider

$$Q \triangleq \int_a^b e^{F_j(b-s)} G_j G_j^T e^{F_j^T(b-s)} ds. \tag{5.8}$$

Integrals like (5.8) arise in the computation of the statistics of u_k.

Lemma 5.2 *It holds that*

$$Q = J_j - e^{F_j(b-a)} J_j e^{F_j^T(b-a)}$$

where $J_j \in \mathbb{R}^{n\times n}$ is the solution of $F_j J_j + J_j F_j^T = -G_j G_j^T$.

Proof: For conciseness, let us drop the subscript indexes. First observe that Assumption 5.1 implies the invertibility of F. Also notice that F and e^{Ft} commute for any $t \in \mathbb{R}$, and so do their transposed. Integration by parts of Q – reading $e^{F(b-s)}G$ as derivative of $-F^{-1}e^{F(b-s)}G$ and $G^T e^{F^T(b-s)}$ as primitive of $-G^T e^{F^T(b-s)} F^T$ – followed by left multiplication by F yields

$$FQ = -(GG^T - e^{F(b-a)} GG^T e^{F^T(b-a)}) - QF^T.$$

By Lemma 5.1, this equation admits a unique solution in Q. By linearity, it must be $Q = J + \bar{J}$, where J and $\bar{J}$ are the unique solutions of

$$\begin{aligned} FJ + JF^T &= -GG^T, \\ F\bar{J} + \bar{J}F^T &= e^{F(b-a)} GG^T e^{F^T(b-a)}. \end{aligned}$$

The latter equation may be rewritten as

$$F(-e^{-F(b-a)} \bar{J} e^{-F^T(b-a)}) + (-e^{-F(b-a)} \bar{J} e^{-F^T(b-a)}) F^T = -GG^T,$$

which is identical to the equation for J. Thus, it must hold that

$$-e^{-F(b-a)} \bar{J} e^{-F^T(b-a)} = J.$$

Solving for $\bar{J}$ and substituting the result into the expression of Q yields

$$Q = J + \bar{J} = J - e^{F(b-a)} J e^{F^T(b-a)},$$

and the assertion is proven. □

For any $k \in \mathbb{N}_0$ and any $j \in \mathcal{Q}$, define

$$\begin{aligned} \widehat{A}_{k,j} &\triangleq e^{F_j(t_{k+1}-t_k)}, \\ \widetilde{A}_k(t^\star, j) &\triangleq \begin{cases} e^{F_0(t^\star - t_k)}, & j = 0, \\ e^{F_j(t_{k+1}-t^\star)}, & j = 1, \dots, N-1, \end{cases} \\ \widetilde{Q}_k(t^\star, j) &\triangleq J_j - \widetilde{A}_k(t^\star, j) J_j \widetilde{A}_k^T(t^\star, j), \end{aligned}$$

where J_j is the unique solution of equation (5.7).

Proposition 5.1 *Assume that $t^\star \in (t_h, t_{h+1})$, $t_h, t_{h+1} \in \mathcal{T}$, for some $h \in \mathbb{N}_0$. The following definition of the parameters $A_k(t^\star, q^\star)$, $C_k(t^\star, q^\star)$, $D_k(t^\star, q^\star)$ and of x_0, $\{u_k\}$, $\{v_k\}$ guarantees the equivalence of systems (4.25) and (5.6) at the sample times $\mathcal{T}$:*

- *$\{u_k\}$ and $\{v_k\}$ are white processes, mutually independent and independent of x_0, with $x_0 \sim \mathcal{N}(\hat{x}_0, P_0)$, $v_k \sim \mathcal{N}(0, I)$ and $u_k \sim \mathcal{N}(0, Q_k(t^\star, q^\star))$, where $\hat{x}_0 \triangleq \hat{\xi}_0$ and $P_0 \triangleq \Sigma_0$;*
- *For $k < h$,*

$$\begin{aligned} A_k(t^\star, q^\star) &= \widehat{A}_{k,0} & \qquad Q_k(t^\star, q^\star) &= J_0 - \widehat{A}_{k,0}\, J_0\, \widehat{A}_{k,0}^T \\ C_k(t^\star, q^\star) &= H_0 & \qquad D_k(t^\star, q^\star) &= K_0\ ; \end{aligned}$$

- *For $k > h$,*

$$\begin{aligned} A_k(t^\star, q^\star) &= \widehat{A}_{k,q^\star} & \qquad Q_k(t^\star, q^\star) &= J_{q^\star} - \widehat{A}_{k,q^\star}\, J_{q^\star}\, \widehat{A}_{k,q^\star}^T \\ C_k(t^\star, q^\star) &= H_{q^\star} & \qquad D_k(t^\star, q^\star) &= K_{q^\star}; \end{aligned}$$

- *For $k = h$,*

$$\begin{aligned} A_k(t^\star, q^\star) &= \widetilde{A}_k(t^\star, q^\star) \widetilde{A}_k(t^\star, 0) \\ Q_k(t^\star, q^\star) &= \widetilde{A}_k(t^\star, q^\star) \widetilde{Q}_k(t^\star, 0) \widetilde{A}_k(t^\star, q^\star) + \widetilde{Q}_k(t^\star, q^\star) \\ C_k(t^\star, q^\star) &= H_0 \\ D_k(t^\star, q^\star) &= K_0. \end{aligned}$$

Proof: The choice of the joint description of x_0 and $\{v_k\}$ is obvious. The definition of parameters C_k and D_k is also immediate: indeed,

$$(C_k, D_k)(t^\star, q^\star) = \begin{cases} (H_0, K_0), & t^\star > t_k \\ (H_{q^\star}, K_{q^\star}), & t^\star < t_k \end{cases},$$

where $t^\star > t_k$ if $h \geq k$ and $t^\star < t_k$ if $h < k$. Finding the parameters $\{A_k\}$ and the description of $\{u_k\}$ amounts to solving the dynamics of the original system over the interval $T_k \triangleq (t_k, t_{k+1})$, for every index $k \in \mathbb{N}_0$. If $k \neq h$, i.e. $t^\star \notin T_k$, $F_{q(t)}$ and $G_{q(t)}$ are constant over T_k, hence

$$\xi(t_{k+1}) = e^{F_j(t_{k+1}-t_k)}\xi(t_k) + u_k,$$

$$u_k \triangleq \int_{t_k}^{t_{k+1}} e^{F_j(t_{k+1}-s)} G_j w(s) ds,$$

where j is equal to 0 or $q^\star$ according to whether $k < h$ or $k > h$, i.e. $t^\star > t_{k+1}$ or $t^\star < t_k$. As u_k is a linear function of $w(s)$, it is zero-mean Gaussian with variance $Q_k(t^\star, q^\star)$ given by

$$\begin{aligned} \mathbb{E}[u_k u_k^T] &= \int\int_{T_k \times T_k} \mathbb{E}[w(s)w(s')] e^{F_j(t_{k+1}-s)} G_j G_j^T e^{F_j^T(t_{k+1}-s')} ds ds' \\ &= \int\int_{T_k \times T_k} \delta(s-s') e^{F_j(t_{k+1}-s)} G_j G_j^T e^{F_j^T(t_{k+1}-s')} ds ds' \\ &= \int_{T_k} e^{F_j(t_{k+1}-s)} G_j G_j^T e^{F_j^T(t_{k+1}-s)} ds \\ &= J_j - e^{F_j(t_{k+1}-t_k)} J_j e^{F_j^T(t_{k+1}-t_k)}, \end{aligned}$$

where Lemma 5.2 has been applied and j is defined as above. If instead $k = h$, i.e. $t^\star \in T_k$, then $F_{q(t)}$ and $G_{q(t)}$ are piecewise constant:

$$(F_{q(t)}, G_{q(t)}) = \begin{cases} (F_0, G_0), & t \in (t_k, t^\star) \\ (F_{q^\star}, G_{q^\star}), & t \in (t^\star, t_{k+1}) \end{cases}.$$

Discretization over $(t_k, t^\star)$ yields

$$\xi(t^\star) = e^{F_0(t^\star - t_k)}\xi(t_k) + u_{k,0},$$

$$u_{k,0} \triangleq \int_{t_k}^{t^\star} e^{F_0(t^\star - s)} G_0 w(s) ds.$$

Discretization over $(t^\star, t_{k+1})$ yields

$$\xi(t_{k+1}) = e^{F_{q^\star}(t_{k+1}-t^\star)}\xi(t^\star) + u_{k,q^\star} = e^{F_{q^\star}(t_{k+1}-t^\star)} e^{F_0(t^\star - t_k)}\xi(t_k) + u_k$$

with $u_k \triangleq e^{F_{q^\star}(t_{k+1}-t^\star)}u_{k,0} + u_{k,q^\star}$ and

$$u_{k,q^\star} \triangleq \int_{t^\star}^{t_{k+1}} e^{F_{q^\star}(t_{k+1}-s)}G_{q^\star}w(s)ds.$$

Again, $u_{k,0}$ and $u_{k,q^\star}$ are both zero-mean Gaussian. Their variances are

$$\begin{aligned}
\widetilde{Q}_k(t^\star,0) &= \int_{t_k}^{t^\star} e^{F_0(t^\star-s)}G_0G_0^T e^{F_0^T(t^\star-s)}ds = J_0 - e^{F_0(t^\star-t_k)}J_0 e^{F_0^T(t^\star-t_k)},\\
\widetilde{Q}_k(t^\star,q^\star) &= \int_{t^\star}^{t_{k+1}} e^{F_{q^\star}(t_{k+1}-s)}G_{q^\star}G_{q^\star}^T e^{F_{q^\star}^T(t_{k+1}-s)}ds\\
&= J_{q^\star} - e^{F_{q^\star}(t_{k+1}-t^\star)}J_{q^\star}e^{F_{q^\star}^T(t_{k+1}-t^\star)},
\end{aligned}$$

in the order, where again Lemma 5.2 has been applied. Moreover, as they depend on disjoint portions of $w(t)$ and $w(t)$ is white, variables $u_{k,0}$ and $u_{k,q^\star}$ are mutually independent. Then u_k is zero mean Gaussian with variance

$$Q_k(t^\star,q^\star) = e^{F_{q^\star}(t_{k+1}-t^\star)}\widetilde{Q}_k(t^\star,0)e^{F_{q^\star}^T(t_{k+1}-t^\star)} + \widetilde{Q}_k(t^\star,q^\star).$$

Finally, since $w(\cdot)$ is independent x_0 and $\{v_k\}$, and $\{u_k\}$ is a linear transformation of $w(\cdot)$, $\{u_k\}$ is also independent of x_0 and $\{v_k\}$. □

Therefore, for every value of $q^\star$, all the parameters of system (5.6), namely

$$(A_k, Q_k, C_k, D_k)(t^\star, q^\star) \tag{5.9}$$

may be written in the form of explicit functions of $t^\star$, with

$$t^\star \in \bigcup_{h\in\mathbb{N}_0}(t_h, t_{h+1}).$$

In particular, they depend on $t^\star$ only through simple matrix exponentials. For simplicity, values of $t^\star$ in $\mathcal{T}$ are not considered in the proposition. This constitutes no loss of generality: because $t^\star$ is a continuous random variable,

$$\mathbb{P}[t^\star \in \mathcal{T}] = 0.$$

Remark. If Assumption 5.1 fails to hold, equation (5.7) may not have a solution. It may be shown (see [34], pp. 278–281) that the equation is satisfied by some J_j if and only if

$$\begin{bmatrix} F_j & -G_jG_j^T \\ 0 & -F_j^T \end{bmatrix} \sim \begin{bmatrix} F_j & 0 \\ 0 & -F_j^T \end{bmatrix}$$

where symbol "$\sim$" denotes matrix similarity. If a solution exists, Lemma 5.2 and Proposition 5.1 hold without modifications. If, however, no such J_j exists, Lemma 5.2 no longer holds. As a consequence, the integration yielding $Q_k(t^\star, q^\star)$ cannot be solved explicitly. □

5.3 Conditioned Kalman filtering revisited

We are now interested in conditioned Kalman filtering and prediction at the sample times $\mathcal{T}$. Based on the discrete-time system (5.6), we wish to compute, for $\ell = k, k+1$, explicit expressions for the quantities

$$\hat{x}_{\ell|k}^{t^\star,q^\star} \triangleq \mathbb{E}[x_\ell | y^k, t^\star, q^\star], \tag{5.10a}$$

$$P_{\ell|k}^{t^\star,q^\star} \triangleq \mathrm{Var}(\tilde{x}_{\ell|k} | y^k, t^\star, q^\star), \tag{5.10b}$$

where $\tilde{x}_{\ell|k}$ is the estimation error $x_\ell - \hat{x}_{\ell|k}$. Observe that, by the equivalence of systems (5.6) and (4.25) at sample times,

$$\begin{aligned} \hat{x}_{\ell|k}^{t^\star,q^\star} &= \hat{\xi}_{t_\ell|k}^{\tilde{q}}, \\ P_{\ell|k}^{t^\star,q^\star} &= \Sigma_{t_\ell|k}^{\tilde{q}}, \end{aligned}$$

where $\tilde{q}$ is fixed by the values of $t^\star$ and $q^\star$.

Proposition 5.2 *For $\ell = k$ and $\ell = k+1$, the conditioned estimate (5.10a) and the error covariance matrix (5.10b) may be computed iteratively as follows:*

i. Measurement update: compute

$$\hat{x}_{k|k}^{t^\star,q^\star} = \hat{x}_{k|k-1}^{t^\star,q^\star} + L_k(t^\star, q^\star)[y_k - C_k(t^\star, q^\star)\hat{x}_{k|k-1}^{t^\star,q^\star}], \tag{5.11a}$$

$$P_{k|k}^{t^\star,q^\star} = P_{k|k-1}^{t^\star,q^\star} - L_k(t^\star, q^\star)C_k(t^\star, q^\star)P_{k|k-1}^{t^\star,q^\star}, \tag{5.11b}$$

with gain matrix $L_k(t^\star, q^\star)$ given by

$$P_{k|k-1}^{t^\star,q^\star} C_k^T(t^\star, q^\star)\left[D_k(t^\star, q^\star)D_k^T(t^\star, q^\star) + C_k(t^\star, q^\star)P_{k|k-1}^{t^\star,q^\star}C_k^T(t^\star, q^\star)\right]^{-1};$$

ii. Time update: compute

$$\hat{x}_{k+1|k}^{t^\star,q^\star} = A_k(t^\star, q^\star)\hat{x}_{k|k}^{t^\star,q^\star}, \tag{5.12a}$$

$$P_{k+1|k}^{t^\star,q^\star} = A_k(t^\star, q^\star)P_{k|k}^{t^\star,q^\star}A_k^T(t^\star, q^\star) + Q_k(t^\star, q^\star). \tag{5.12b}$$

Iterations are initialized by $\hat{x}_{0|-1}^{t^\star,q^\star} = \hat{x}_0$ *and* $P_{0|-1}^{t^\star,q^\star} = P_0$.

Proof: Apply Proposition 1.1 to the system conditioned on $t^\star$ and $q^\star$. □

Consider the values of $t^\star$ in the interval (t_h, t_{h+1}), $t_h, t_{h+1} \in \mathcal{T}$, for some fixed index $h \in \mathbb{N}_0$. It is evident by Proposition 5.1 that, for $k \neq h$, parameters (5.9) are constant w.r.t. to $t^\star$. It follows that most of the recursion steps (5.11) and (5.12) do not depend on $t^\star$ themselves. However, in general, they update quantities that do depend on $t^\star$. Therefore, it may be expected that a suitable reformulation of the update rules allows to isolate the dependency on $t^\star$ out of the recursion. For any $k \in \mathbb{N}_0$ and $\ell = k, k+1$, define

$$\hat{x}_{\ell|k}^{\infty} \triangleq \hat{x}_{\ell|k}^{+\infty,q^\star}, \tag{5.13a}$$

$$P_{\ell|k}^{\infty} \triangleq P_{\ell|k}^{+\infty,q^\star}, \tag{5.13b}$$

where the informal notation "$t^\star = +\infty$" stands for $q(t) \equiv 0$.

Corollary 5.1 *(i) Expressions (5.13) are independent of* $q^\star$. *They may be computed by running iteration (5.11)÷(5.12) with parameters*

$$(A_k, Q_k, C_k, D_k)(t^\star, q^\star) = (\widehat{A}_{k,0}, J_0 - \widehat{A}_{k,0} J_0 \widehat{A}_{k,0}^T, H_0, K_0).$$

(ii) If $t^\star > t_h$, $h \in \mathbb{N}_0$, $t_h \in \mathcal{T}$, *then for any* $k \leq h$ *it holds that*

$$\begin{aligned} \hat{x}_{k|k-1}^{t^\star,q^\star} &= \hat{x}_{k|k-1}^{\infty}, \qquad & \hat{x}_{k|k}^{t^\star,q^\star} &= \hat{x}_{k|k}^{\infty}, \\ P_{k|k-1}^{t^\star,q^\star} &= P_{k|k-1}^{\infty}, \qquad & P_{k|k}^{t^\star,q^\star} &= P_{k|k}^{\infty}. \end{aligned}$$

In particular, the above quantities are independent of $t^\star$.

Proof: (i) Just notice that the results of Proposition 5.1 for the case $k < h$ equally hold for any trajectory $q(t)$ such that $q(t) = 0$, $t \in (t_k, t_{k+1})$. (ii) By (i), right-hand side and left-hand side quantities obey the same recursion in k up to index $k = h$. □

Thus, over the time intervals before $t^\star$, the recursion is completely independent of $t^\star$, the system being conditioned on the trajectory $q(t) \equiv 0$.

Corollary 5.2 *If* $t^\star \in (t_h, t_{h+1})$, $h \in \mathbb{N}_0$, $t_h, t_{h+1} \in \mathcal{T}$, *then it holds that*

$$\begin{aligned} \hat{x}_{h+1|h}^{t^\star,q^\star} &= A_h(t^\star, q^\star) \hat{x}_{h|h}^{\infty}, \\ P_{h+1|h}^{t^\star,q^\star} &= A_h(t^\star, q^\star) P_{h|h}^{\infty} A_h^T(t^\star, q^\star) + Q_h(t^\star, q^\star), \end{aligned}$$

with

$$A_h(t^\star, q^\star) = \widetilde{A}_h(t^\star, q^\star)\widetilde{A}_h(t^\star, 0),$$
$$Q_h(t^\star, q^\star) = \widetilde{A}_h(t^\star, q^\star)\widetilde{Q}_h(t^\star, 0)\widetilde{A}_h(t^\star, q^\star) + \widetilde{Q}_h(t^\star, q^\star).$$

Proof: For $k = h$, substitute the results of Corollary 5.1 and Proposition 5.1 into equations (5.12). □

The time update step of index $k = h$ is where the value of $t^\star$ comes into play in the recursion (5.11)÷(5.12). Let us now apply the results of Chapter 3 to the recursion starting at index $k = h + 1$ with initial state given by $\hat{x}^{t^\star,q^\star}_{h+1|h}$ and $P^{t^\star,q^\star}_{h+1|h}$.

Corollary 5.3 *If $t^\star \in (t_h, t_{h+1})$, $h \in \mathbb{N}_0$, $t_h, t_{h+1} \in \mathcal{T}$, then, for $k \geq h + 1$,*

$$\hat{x}^{t^\star,q^\star}_{k|k} = \hat{x}^{\diamond}_{k|k} + A_k^{-1}U_k(I + S_kP^{t^\star,q^\star}_{h+1|h})^{-T}(\hat{x}^{t^\star,q^\star}_{h+1|h} + P^{t^\star,q^\star}_{h+1|h}M_k), \quad (5.14a)$$
$$P^{t^\star,q^\star}_{k|k} = P^{\diamond}_{k|k} + A_k^{-1}U_k(I + S_kP^{t^\star,q^\star}_{h+1|h})^{-T}P^{t^\star,q^\star}_{h+1|h}U_k^TA_k^{-T}, \quad (5.14b)$$

and

$$\hat{x}^{t^\star,q^\star}_{k+1|k} = \hat{x}^{\diamond}_{k+1|k} + U_k(I + S_kP^{t^\star,q^\star}_{h+1|h})^{-T}(\hat{x}^{t^\star,q^\star}_{h+1|h} + P^{t^\star,q^\star}_{h+1|h}M_k), \quad (5.15a)$$
$$P^{t^\star,q^\star}_{k+1|k} = P^{\diamond}_{k+1|k} + U_k(I + S_kP^{t^\star,q^\star}_{h+1|h})^{-T}P^{t^\star,q^\star}_{h+1|h}U_k^T. \quad (5.15b)$$

Quantities U_k, S_k, M_k, $\hat{x}^\diamond$ and $P^\diamond$ obey the usual iterations, provided the initializations $U_h = I$, $S_h = 0$, $M_h = 0$, $\hat{x}^\diamond_{h+1,h} = 0$, $P^\diamond_{h+1,h} = 0$ and $\widetilde{M}_h = 0$.

Proof: Note that $f(x_{h+1}|y^h, t^\star, q^\star) = \mathcal{N}(\hat{x}^{t^\star,q^\star}_{h+1|h}, P^{t^\star,q^\star}_{h+1|h})$. Then, for $k \geq h+1$, the results of Propositions 3.1÷ 3.2 and Corollary 3.2 may be applied to system (5.6) with time origin $k = h + 1$ conditioned on $t^\star$, $q^\star$ and on y^h. □

For any $k \geq h + 1$, these equations make the dependency on $t^\star$ of estimates (5.10) explicit. Observe that parameters U_k, S_k, M_k, $\hat{x}^\diamond$ and $P^\diamond$ do not depend on $t^\star$. However, they depend on h through the initialization, and on $q^\star$. Where necessary, we will write $U_k(h, q^\star)$, $S_k(h, q^\star)$, $M_k(h, q^\star)$, $\hat{x}^\diamond(h, q^\star)$ and $P^\diamond(h, q^\star)$.

So far, we have given formulas for the computation of the functions (5.10) by fixing the value of $t^\star$ within intervals. Let us now fix index k. Assume that the quantities $U_k(h, q^\star)$, $S_k(h, q^\star)$, $M_k(h, q^\star)$, $\hat{x}^\diamond_{\ell|k}(h, q^\star)$, $P^\diamond_{\ell|k}(h, q^\star)$, with $\ell = k, k + 1$ and $\hat{x}^\infty_{\ell|h}$, $P^\infty_{\ell|h}$, with $\ell = h, h + 1$, are available for $h = 0, \ldots, k$.

For changing values of $t^\star$, estimates (5.10), $\ell = k, k+1$ may be evaluated piecewise according to the partitioning

$$(t_0, t_1), \dots, (t_h, t_{h+1}), \dots, (t_k, t_{k+1}), (t_{k+1}, +\infty). \tag{5.16}$$

Specifically, Corollary 5.3 and Corollary 5.2 provide evaluation formulas for $t^\star \in (t_h, t_{h+1})$, with $h \leq k$, whereas Corollary 5.1 accounts for $t^\star \geq t_{k+1}$.

5.4 Implementation

All the quantities required in the evaluation of $\hat{x}_{\ell|k}^{t^\star,q^\star}$ and $P_{\ell|k}^{t^\star,q^\star}$ may be computed iteratively in the index k. Algorithm 5.1 outlines an implementation of the method. The procedure is identical for every $q^\star \in \mathcal{Q} \setminus \{0\}$. For shortness, we dropped $q^\star$ from the notation. At step k, the measurement update yields all the parameters needed to evaluate $\hat{x}_{k|k}^{t^\star,q^\star}$ and $P_{k|k}^{t^\star,q^\star}$ for almost every value of $t^\star$. Similarly, the time update provides all parameters for the evaluation of $\hat{x}_{k+1|k}^{t^\star,q^\star}$ and $P_{k+1|k}^{t^\star,q^\star}$. The update formulas are those given in the previous sections. They require the computation of the discrete-time system parameters. In principle, these may be computed offline except for $\widetilde{A}_k$ and $\widetilde{Q}_k$ – see Proposition 5.1 – whose values depend on $t^\star$. Matrix J_j needs to be computed for every $j \in \mathcal{Q}$ before execution. In practice, this is done by standard numerical techniques which solve equation (5.7) via Schur transformation of matrices F_j and F_j^T [5].

Remark. Parameters S_k, U_k and $P_{\ell|k}^\diamond$, $P_{\ell|k}^\infty$, $\ell = k, k+1$, do not depend on data y^k. Therefore, they may also be computed offline. If, for a certain $T > 0$ and every $k \in \mathbb{N}_0$, $t_{k+1} - t_k = T$, with $t_k, t_{k+1} \in \mathcal{T}$, then

$$\begin{aligned} S_k(h, q^\star) &= S_{k-h}(0, q^\star), \\ U_k(h, q^\star) &= U_{k-h}(0, q^\star), \\ P_{\ell|k}^\diamond(h, q^\star) &= P_{\ell-h|k-h}^\diamond(0, q^\star). \end{aligned}$$

In this case, it suffices to compute S_k, U_k and $P_{\ell|k}^\diamond$ for $h = 0$. □

Consider Proposition 4.2. For $\ell = 0, \dots, k$, the following function may also be evaluated piecewise:

$$f(y_\ell | y^{\ell-1}, t^\star, q^\star) = \mathcal{N}(H_j \hat{x}_{\ell|\ell-1}^{t^\star,q^\star}, H_j \Sigma_{\ell|\ell-1}^{t^\star,q^\star} H_j^T + K_j K_j^T), \tag{5.17}$$

Algorithm 5.1 Conditioned estimation algorithm

{Initialization}
set $\hat{x}^{\infty}_{0|-1} = \mu_0$
set $P^{\infty}_{0|-1} = P_0$

$k \Leftarrow 0$

loop {Iterations}

{Measurement update}
for $h = 0, \ldots, k-1$ **do**

$$\text{compute} \begin{cases} \hat{x}^{\diamond}_{k|k}(h) \\ P^{\diamond}_{k|k}(h) \\ U_k(h) \\ S_k(h) \\ \widetilde{M}_k(h) \\ M_k(h) \end{cases} \text{from} \begin{cases} \hat{x}^{\diamond}_{k|k-1}(h) & P^{\diamond}_{k|k-1}(h) \\ P^{\diamond}_{k|k-1}(h) & \\ U_{k-1}(h) & \\ S_{k-1}(h) & U_{k-1}(h) \quad P^{\diamond}_{k|k-1}(h) \\ \widetilde{M}_{k-1}(h) & P^{\diamond}_{k|k-1}(h) \\ M_{k-1}(h) & \widetilde{M}_{k-1}(h) \quad U_{k-1}(h) \quad P^{\diamond}_{k|k-1}(h) \end{cases}$$

end for
compute $\hat{x}^{\infty}_{k|k}$ from $\hat{x}^{\infty}_{k|k-1}$, $P^{\infty}_{k|k-1}$
compute $P^{\infty}_{k|k}$ from $P^{\infty}_{k|k-1}$

{Time update}
for $h = 0, \ldots, k-1$ **do**
compute $\hat{x}^{\diamond}_{k+1|k}(h)$ from $\hat{x}^{\diamond}_{k|k}(h)$
compute $P^{\diamond}_{k+1|k}(h)$ from $P^{\diamond}_{k|k}(h)$
end for
compute $\hat{x}^{\infty}_{k+1|k}$ from $\hat{x}^{\infty}_{k|k}$
compute $P^{\infty}_{k+1|k}$ from $P^{\infty}_{k|k}$

{Initialization of the next step}

$$\text{set} \begin{cases} \hat{x}^{\diamond}_{k+1|k}(k) \\ P^{\diamond}_{k+1|k}(k) \\ U_k(k) \\ S_k(k) \\ \widetilde{M}_k(k) \\ M_k(k) \end{cases} = \begin{cases} 0 \\ 0 \\ I \\ 0 \\ 0 \\ 0 \end{cases}$$

$k \Leftarrow k+1$

end loop

where $j = 0$ or $j = q^\star$ according to whether $t^\star > t_{k+1}$ or $t^\star < t_{k+1}$. Hence,

$$f(y^k|t^\star, q^\star) = \prod_{\ell=0}^{k} f(y_\ell|y^{\ell-1}, t^\star, q^\star).$$

Therefore, all the quantities involved in the state estimation problem are expressed in the form of piecewise explicit functions of $t^\star$.

For $\xi = \xi(t_\ell) = x_\ell$ and convenient functions g and z, the solution to the state estimation problem given in Propositions 4.5÷4.6 may be reduced to integrals such as

$$\int_0^{+\infty} g(\hat{x}_{\ell|k}^{t^\star,q^\star}, P_{\ell|k}^{t^\star,q^\star}) f(y^k|t^\star, q^\star) z(t^\star) dt^\star. \tag{5.18}$$

In particular, integral

$$\int_0^{+\infty} \hat{x}_{\ell|k}^{t^\star,q^\star} f(y^k|t^\star, q^\star) f(t^\star, q^\star) dt^\star,$$

involved in the computation of $\hat{\xi}_{t_\ell|k}$, is found by setting $g(\hat{x}_{\ell|k}^{t^\star,q^\star}, P_{\ell|k}^{t^\star,q^\star}) = \hat{x}_{\ell|k}^{t^\star,q^\star}$ and $z(t^\star) = f(t^\star, q^\star)$. Similarly, in the computation of

$$\hat{t}_k^\star \propto \sum_{q^\star} \int t^\star f(y^k|t^\star, q^\star) f(t^\star, q^\star) dt^\star,$$

one just needs to set $g \equiv 1$ and $z(t^\star) = t^\star f(t^\star, q^\star)$. Along the same lines, integrals for the computation of $f(y^k)$, $p_{t|k}$ and $\Sigma_{t_\ell|k}$ are easily recovered. Integration over a finite interval is obtained by using an indicator function in the definition of $z(t^\star)$. Define

$$f^\infty(y_\ell|y^{\ell-1}) \triangleq \mathcal{N}(H_0\hat{x}_{\ell|k}^\infty, H_0 P_{\ell|k}^\infty H_0^T + K_0 K_0^T).$$

Proposition 5.3 *Let g be a function of $\hat{x}_{\ell|k}^{t^\star,q^\star}$ and $P_{\ell|k}^{t^\star,q^\star}$, with $\ell = k$ or $k+1$, and let z be a function of $t^\star$ having primitive $\mathcal{Z}$. Then, integral (5.18) may be computed as follows:*

$$\sum_{h=0}^{k} \mathcal{V}_h \mathcal{I}_h + g(\hat{x}_{\ell|k}^\infty, P_{\ell|k}^\infty) \mathcal{V}_k (\mathcal{Z}(+\infty) - \mathcal{Z}(t_{k+1})) \tag{5.19}$$

where

$$\mathcal{V}_h \triangleq \prod_{l=0}^{h} f^\infty(y_l|y^{l-1}),$$

$$\mathcal{I}_h \triangleq \int_{t_h}^{t_{h+1}} g(\hat{x}_{\ell|k}^{t^\star,q^\star}, P_{\ell|k}^{t^\star,q^\star}) \prod_{l=h+1}^{k} f(y_l|y^{l-1}, t^\star, q^\star) z(t^\star) dt^\star.$$

Proof: The result is obtained by partitioning the integral according to (5.16) and recalling the independence of $\hat{x}_{\ell|k}^{t^\star,q^\star}$ and $P_{\ell|k}^{t^\star,q^\star}$ from $t^\star$ for $t^\star \geq t_{k+1}$. □

Using Proposition 5.3, we can exploit the properties of Algorithm 5.1 for the numerical solution of the estimation problems. Every integration appearing in (5.19) is over an interval of type (t_h, t_{h+1}), with $t_h, t_{h+1} \in \mathcal{T}$. It may be evaluated by any quadrature method using the parametric expressions of $\hat{x}_{\ell|k}^{t^\star,q^\star}$ and $P_{\ell|k}^{t^\star,q^\star}$ with parameters fixed by index h. Infinite support integrations are avoided by the use of estimates $\hat{x}_{\ell|k}^{\infty}$, $P_{\ell|k}^{\infty}$ and direct evaluation of $\mathcal{Z}(t_\ell)$. Incidentally, the primitive of the functions of our interest, namely $z(t^\star) = (t^\star)^d f(t^\star, q^\star)$ with $d \in \mathbb{N}_0$, is

$$\mathcal{Z}(t^\star) = -d! \frac{\lambda_{q^\star}}{\Lambda^{d+1}} \sum_{l=0}^{d} \frac{(\Lambda t^\star)^{d-l}}{(d-l)!} e^{-\Lambda t^\star},$$

for which $\mathcal{Z}(+\infty) = 0$.

5.5 Low-complexity approximations

According to Proposition 5.3, the computation of integral (5.18) requires approximating $\mathcal{O}(k)$ finite-support integrals, each involving the evaluation of $\mathcal{O}(k)$ factors depending on $t^\star$. Therefore, the computational demand grows rather fast with index k. In certain situations, however, estimates (5.10) become practically independent of the remote values of $t^\star$. In this case, the computation of (5.18) may be largely simplified.

Proposition 5.4 *Assume that $t_{k+1} - t_k = T$, $\exists T > 0$, for any $t_k, t_{k+1} \in \mathcal{T}$. (i) The system*

$$(\widehat{A}_{k,j}, J_j - \widehat{A}_{k,j} J_j \widehat{A}_{k,j}^T, H_j, K_j) \tag{5.20}$$

is time-invariant. (ii) If system (5.20) is controllable and detectable, then, for any fixed $t^\star$,

$$\hat{x}_{\ell|k}^{t^\star,j} \xrightarrow{k} \hat{x}_{\ell|k}^{0,j}, \tag{5.21a}$$

$$P_{\ell|k}^{t^\star,j} \xrightarrow{k} P_{\ell|k}^{0,j}. \tag{5.21b}$$

Proof: (i) The result follows by the definition of the parameters. (ii) (*Sketch*) Assume w.l.o.g. that $t^\star < t_{h+1}$ for some index h. For $k > h$, system (5.6)

is time-invariant with parameters (A, Q, C, D) given by (5.20). Consider $\ell = k + 1$. Controllability and detectability of (A, Q, C, D) guarantee that

$$P^{t^\star,j}_{k+1|k} \xrightarrow{k} \overline{P},$$

regardless of the initial value $P^{t^\star,j}_{h+1|h}$ [52], that is, regardless of the value of $t^\star$. This proves (5.21b). Moreover, $\overline{P}$ is stabilizing, i.e. the predictor state evolution matrix $A(I - \overline{L}C)$, with

$$\overline{L} = \overline{P}C^T(C\overline{P}C^T + DD^T)^{-1},$$

is strictly stable. Denote with $\hat{x}^{ss}_{k+1|k}$ the steady-state predictor associated to the constant gain $\overline{L}$. Then [52],

$$\hat{x}^{t^\star,j}_{k+1|k} \xrightarrow{k} \hat{x}^{ss}_{k+1|k}$$

regardless of the initial condition $\hat{x}^{t^\star,j}_{h+1|h}$, that is, regardless of the value of $t^\star$. Convergence (5.21a) follows. □

In the light of (5.21a) and (5.21b), for $t^\star \in (t_h, t_{h+1})$ and $k \geq h + d$, $d \in \mathbb{N}$, it is reasonable to consider the approximation

$$(\hat{x}^{t^\star,j}_{\ell|k}, P^{t^\star,j}_{\ell|k}) \simeq (\hat{x}^{0,j}_{\ell|k}, P^{0,j}_{\ell|k}). \tag{5.22}$$

Index d plays the role of a "forgetting time" and is chosen based on the properties of system (5.20). As a consequence, for $l > h + d$, we get the approximation

$$f(y_l|y^{l-1}, t^\star, q^\star) \simeq f^{0,q^\star}(y_l|y^{l-1}), \tag{5.23}$$

where

$$f^{0,q^\star}(y_l|y^{l-1}) \triangleq \mathcal{N}(H_{q^\star}\hat{x}^{0,q^\star}_{l|l-1}, H_{q^\star}P^{0,q^\star}_{l|l-1}H^T_{q^\star} + K_{q^\star}K^T_{q^\star}).$$

Corollary 5.4 *Assume that (5.22) holds with equality. Then, for $k \geq d$, integral (5.18) is given by*

$$g(\hat{x}^{0,q^\star}_{\ell|k}, P^{0,q^\star}_{\ell|k})\mathcal{S}_k + \sum_{h=k-d+1}^{k} \mathcal{V}_h\mathcal{I}_h + g(\hat{x}^{\infty}_{\ell|k}, P^{\infty}_{\ell|k})\mathcal{V}_k(\mathcal{Z}(+\infty) - \mathcal{Z}(t_{k+1})) \tag{5.24}$$

where $\mathcal{S}_k$ is computed iteratively as follows:

$$\mathcal{S}_k = f^{0,q^\star}(y_k|y^{k-1})\mathcal{S}_{k-1} + \mathcal{V}_{k-d}\widetilde{\mathcal{I}}_k,$$

with

$$\widetilde{\mathcal{I}}_k = \int_{t_{k-d}}^{t_{k-d+1}} \prod_{l=k-d+1}^{k} f(y_l|y^{l-1}, t^\star, q^\star) z(t^\star) dt^\star.$$

Iterations are initialized by $S_{d-1} \triangleq 0$.

Proof: For $k \geq d$, it holds that $g(\hat{x}_{\ell|k}^{t^\star,q^\star}, P_{\ell|k}^{t^\star,q^\star}) = g(\hat{x}_{\ell|k}^{0,q^\star}, P_{\ell|k}^{0,q^\star})$. Hence, the summation in (5.19) may be written as

$$g(\hat{x}_{\ell|k}^{0,q^\star}, P_{\ell|k}^{0,q^\star})\mathcal{S}_k + \sum_{h=k-d+1}^{k} \mathcal{V}_h \mathcal{I}_h$$

with

$$\mathcal{S}_k \triangleq \sum_{h=0}^{k-d} \mathcal{V}_h \int_{t_h}^{t_{h+1}} \prod_{l=h+1}^{k} f(y_l|y^{l-1}, t^\star, q^\star) z(t^\star) dt^\star.$$

Observe that, for $t^\star \in (t_h, t_{h+1})$ and $h = 0, \ldots, k-d-1$, (5.23) holds for $l = k$. Then, for any $k > d$,

$$\begin{aligned}\mathcal{S}_k = &\sum_{h=0}^{k-d-1} \mathcal{V}_h \int_{t_h}^{t_{h+1}} f^{0,q^\star}(y_k|y^{k-1}) \prod_{l=h+1}^{k-1} f(y_l|y^{l-1}, t^\star, q^\star) z(t^\star) dt^\star + \\ &\mathcal{V}_{k-d} \int_{t_{k-d}}^{t_{k-d+1}} \prod_{l=k-d+1}^{k} f(y_l|y^{l-1}, t^\star, q^\star) z(t^\star) dt^\star \\ = &\, f^{0,q^\star}(y_k|y^{k-1})\mathcal{S}_{k-1} + \mathcal{V}_{k-d}\widetilde{\mathcal{I}}_k.\end{aligned}$$

Finally, for $k = d$,

$$\mathcal{S}_d = \mathcal{V}_0 \widetilde{I}_d = \mathcal{V}_0 \widetilde{I}_d + f^{0,q^\star}(y_d|y^{d-1})\mathcal{S}_{d-1},$$

provided $\mathcal{S}_{d-1} \triangleq 0$. □

The result is identical for every $q^\star \in \mathcal{Q} \setminus \{0\}$. In this case, at step k, integration (5.18) only requires the computation of $\mathcal{O}(d)$ finite-support integrals, each involving the evaluation of $\mathcal{O}(d)$ factors depending on $t^\star$. In particular, $\hat{x}_{l,l-1}^{t^\star q^\star}$ and $P_{l,l-1}^{t^\star q^\star}$ only need to be evaluated for $h = k-d+1, \ldots, k$ and $l = h+1, \ldots, k$, with $t^\star \in (t_h, t_{h+1})$. Similarly, $\hat{x}_{\ell,k}^{t^\star q^\star}$ and $P_{\ell,k}^{t^\star q^\star}$ only need to be evaluated for $h = k-d+1, \ldots, k$. Therefore, parameters U_l, S_l, M_l, $\hat{x}_{l,l-1}^{\diamond}$, and $P_{l,l-1}^{\diamond}$ only need to be available for $h = k-d+1, \ldots, k$ and $l = h+1, \ldots, k$, whereas $\hat{x}_{\ell,k}^{\diamond}$ and $P_{\ell,k}^{\diamond}$ only need to be available for

$h = k - d + 1, \dots, k$. In practice, a modification of Algorithm 5.1 of $\mathcal{O}(d)$ complexity may be implemented for their computation and storage. Finally, $\mathcal{V}_h$ may be computed by the recursion

$$\mathcal{V}_h = f^{\infty}(y_h|y^{h-1})\mathcal{V}_{h-1},$$

for $h \geq 0$, with $\mathcal{V}_{-1} \triangleq 1$. At step k, it needs to be available for $h = k-d, \dots, k$ only. The new procedure for the computation of integral (5.18) is summarized in Algorithm 5.2. The overall complexity is of $\mathcal{O}(d)$ in both memory and computations.

Remark. The assumption that (5.22) hold with equality is never satisfied. However, the approximation of (5.18) provided by Corollary 5.4 can be made arbitrarily good by choosing a sufficiently large value for d, at the expense of an increase in computational effort. As a last note, Corollary 5.4 is of practical use whenever (5.22) is a reasonable approximation. In this sense, the assumptions of Proposition 5.4 are not strictly necessary. □

5.6 Numerical results

In this section we will show numerical results for a specific example of single switch system. We will pursue a qualitative analysis of the estimation of both the switching time and the state of the system at the sample times t_k. In particular, we are interested in the probability distribution of the current discrete state, i.e. $p_{t_k|k}$, and in the minimum-mean-squared-error estimates of $t^\star$, that is $\hat{t}_k^\star$, and of the continuous state ξ_{t_k}, that is $\hat{\xi}_{t_k|k}$. This amounts to running Algorithm 5.1 and to carry out a numerical evaluation of the estimates, in accordance with the results of Section 5.4.

Let $\mathcal{Q} = \{0, 1, 2\}$. Consider the system (4.1) with $t_k \triangleq k \cdot T$, $T = 0.5$, and parameters $\hat{\xi}_0 = 0$, $\Sigma_0 = 0.1 \cdot I$. We chose all 4-tuples (F_q, G_q, H_q, K_q) to be

$$\left(\begin{bmatrix} -0.4 & 0.6 \\ c_q & -0.5 \end{bmatrix}, \begin{bmatrix} 0.2 & 0.1 \\ 0.1 & 0.2 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0.1 & 0 \\ 0 & 0.1 \end{bmatrix} \right)$$

where $c_0 = 0$, $c_1 = 1$, $c_2 = -2$. That is, only the state evolution matrix changes with q. This modifies the character of the continuous-time system from stable ($q = 0$, stable node) to unstable ($q = 1$, saddle) or oscillatory ($q = 2$, stable focus), according to the different spectra $\sigma(F_0) = \{-0.4, -0.5\}$, $\sigma(F_1) = \{-1.22, 0.32\}$, $\sigma(F_2) = \{-0.45 \pm i\, 1.09\}$. In this setting y is simply

Algorithm 5.2 Approximate estimation

{Initialization}
set $\mathcal{V}_{-1} = 1$
set $\hat{x}^{0,q^\star}_{0|-1} = \hat{x}_0$
set $P^{0,q^\star}_{0|-1} = P_0$
for $k = 0, \dots d-1$ **do**

$$\text{compute} \begin{cases} f^\infty(y_k|y^{k-1}) \\ \mathcal{V}_k \\ \hat{x}^{0,q^\star}_{k|k} \\ P^{0,q^\star}_{k|k} \\ \hat{x}^{0,q^\star}_{k+1|k} \\ P^{0,q^\star}_{k+1|k} \end{cases} \text{from} \begin{cases} \hat{x}^\infty_{k|k-1} & P^\infty_{k|k-1} \\ \mathcal{V}_{k-1} & f^\infty(y_k|y^{k-1}) \\ \hat{x}^{0,q^\star}_{k|k-1} & P^{0,q^\star}_{k|k-1} \\ P^{0,q^\star}_{k|k-1} & \\ \hat{x}^{0,q^\star}_{k|k} & \\ P^{0,q^\star}_{k|k} & \end{cases}$$

  for $h = 0, \dots, k$ **do**
    approximate $\mathcal{I}_h$
  end for
  compute (5.19) {Integral}
end for

set $\mathcal{S}_{d-1} = 0$
$k \Leftarrow d$

{Iterations}
loop

$$\text{compute} \begin{cases} f^\infty(y_k|y^{k-1}) \\ \mathcal{V}_k \\ \hat{x}^{0,q^\star}_{k|k} \\ P^{0,q^\star}_{k|k} \\ \hat{x}^{0,q^\star}_{k+1|k} \\ P^{0,q^\star}_{k+1|k} \end{cases} \text{from} \begin{cases} \hat{x}^\infty_{k|k-1} & P^\infty_{k|k-1} \\ \mathcal{V}_{k-1} & f^\infty(y_k|y^{k-1}) \\ \hat{x}^{0,q^\star}_{k|k-1} & P^{0,q^\star}_{k|k-1} \\ P^{0,q^\star}_{k|k-1} & \\ \hat{x}^{0,q^\star}_{k|k} & \\ P^{0,q^\star}_{k|k} & \end{cases}$$

  for $h = k-d+1, \dots, k$ **do**
    approximate $\mathcal{I}_h$
  end for
  compute $f^{0,q^\star}(y_k|y^{k-1})$ from $\hat{x}^{0,q^\star}_{k|k-1}$ and $P^{0,q^\star}_{k|k-1}$
  approximate $\widetilde{\mathcal{I}}_k$
  compute S_k from $f^{0,q^\star}(y_k|y^{k-1})$, $\mathcal{S}_{k-1}$, $\mathcal{V}_{k-d}$ and $\widetilde{\mathcal{I}}_k$
  compute 5.24 {Approximate integral}

  $k \Leftarrow k+1$

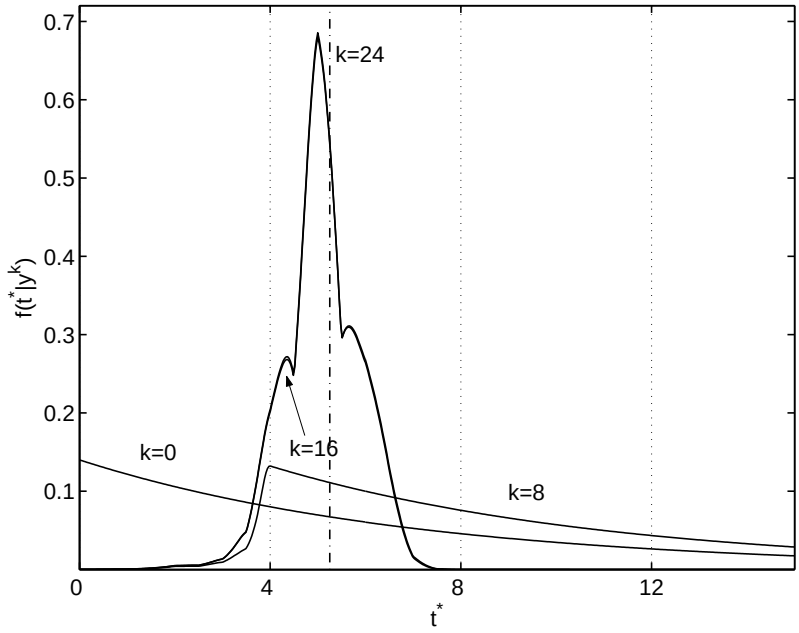
end loop

a noisy version of the state x. The Markov chain underlying the evolution of $q(t)$ is assumed to have generator of the form (4.18). The chain is set to start from $q(0) = 0$ with probability one, i.e. $p_0 = 1$. Switching intensities are fixed to $\lambda_1 = 0.06$, $\lambda_2 = 0.08$. With this choice, $\mathbb{P}[q^\star = 1] \simeq 0.43$, $\mathbb{P}[q^\star = 2] \simeq 0.57$, i.e. jumps towards $q = 2$ are privileged. Moreover, the a priori expected switching time is $\mathbb{E}[t^\star] \simeq 7.14$.

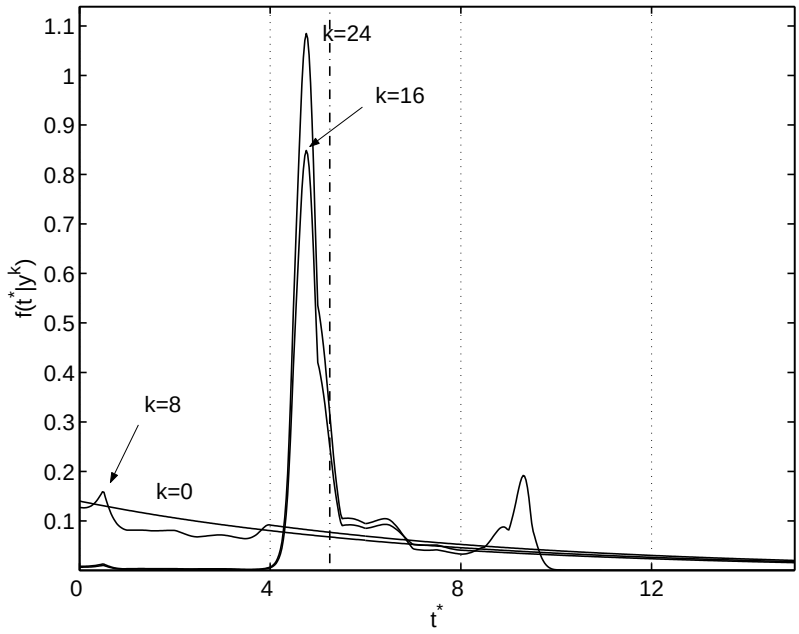
In the simulations, we started off the system from $\xi(0) = 0$. We then randomly generated ξ and y up to time $k_{\max} \cdot T$, with $k_{\max} = 30$, for a jump of $q(t)$ occurring at time $\bar{t}^\star = 5.25$, i.e. significantly before the expected value of $t^\star$. We considered both $\bar{q}^\star = 1$ and $\bar{q}^\star = 2$ as final discrete state. Note that the values assumed by $t^\star$ (i.e. $\bar{t}^\star$) and $q^\star$ (i.e. $\bar{q}^\star = 1$ and $\bar{q}^\star = 2$, in turn) have been chosen manually by the programmer, i.e. they have not been simulated as random variables. This simplifies the analysis and of course does not affect the validity of our estimation methods. Also notice the exiguity of measurements, which is precisely the kind of situation for which model (4.1) is conceived. The procedures of Section 5.4 are then applied to the data y_k, $k = 0, \dots, k_{\max}$. Approximations of Section 5.5 are not used here. Numerical integrations are carried out by a standard iterative Simpson's adaptive quadrature algorithm [28]. Terminating conditions are chosen so to guarantee a relative error less than 10^{-6}.

Figures 5.1(a) and 5.1(b) show, for different values of k, the a posteriori density of $t^\star$ given y^k, for the cases $\bar{q}^\star = 1$ and $\bar{q}^\star = 2$. The evolution from the exponential prior to a density roughly concentrated around the true switching instant may be observed. Notice the exponential tails of the curves for $t^\star > t_k$. For $\bar{q}^\star = 1$, in particular, the almost indistinguishability of the curves associated to $k = 18$ and $k = 24$ reflects the fast convergence of $f(t^\star|y^k)$ to an almost invariant density.

The evolution of the conditioned expectation of $t^\star$ and of the conditioned probability distribution of $q(t_k)$ given y^k are reported in Figure 5.2(a) and Figure 5.2(b), for $\bar{q}^\star = 1$ and $\bar{q}^\star = 2$. One may note that, even before the switch happens, $p_{t_k|k}(0)$ adjusts to values that are significantly smaller than the prior probability $p_0(0) = 1$. This fact reveals that certain fluctuations of the state due to the input noise u may also be explained in terms of a mode switch, and is accompanied by an increase in $p_{t_k|k}(1)$ and $p_{t_k|k}(2)$. Typically, the latter grows faster than the former, since $\lambda_2 > \lambda_1$. For $\bar{q}^\star = 1$, where the initial fluctuations of $p_{t_k|k}(1)$ are rather limited, $\hat{t}_k^\star$ grows in a quasi-linear fashion. This is primarily due to the memoryless nature of (unconditioned) random variable $t^\star$, and may be explained as follows. Under the condition

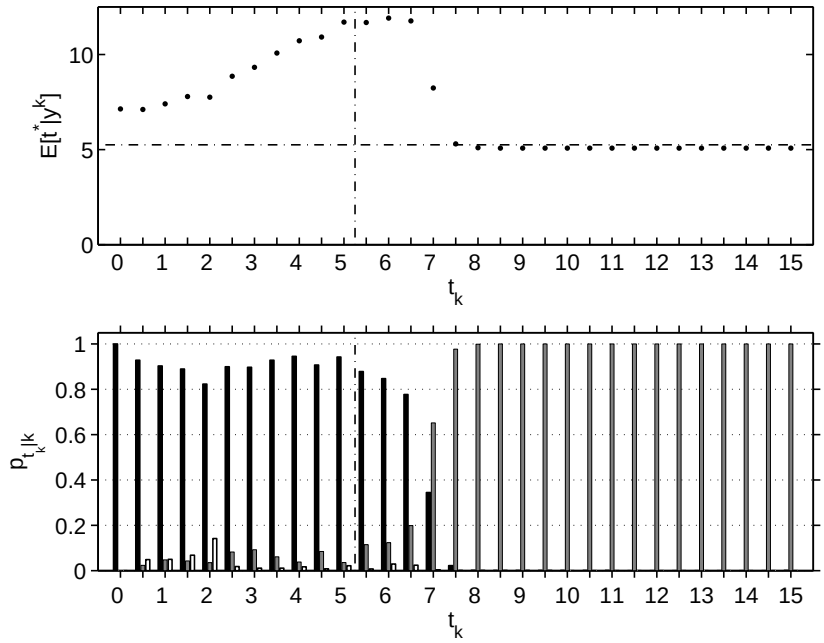


(a) $q^\star = 1$

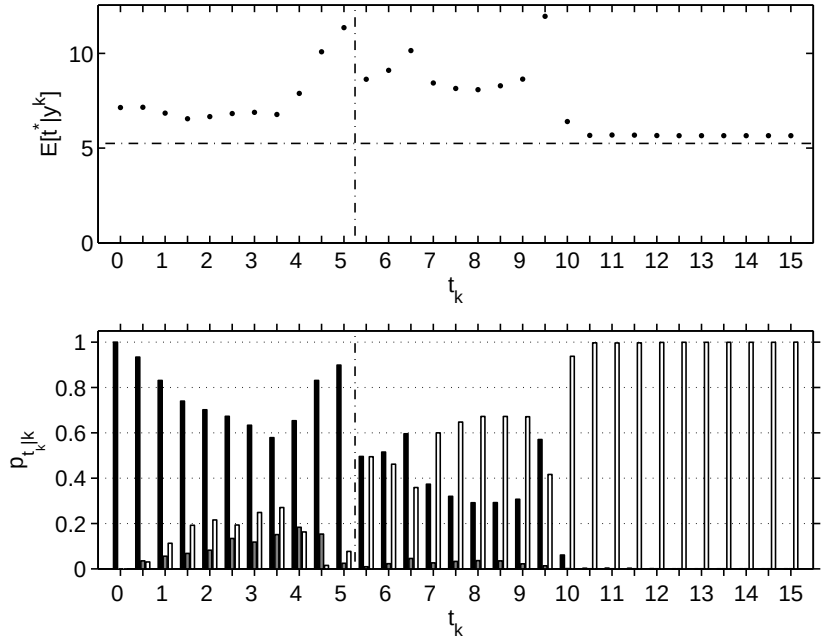


(b) $q^\star = 2$

Figure 5.1: Density functions $f(t^\star|y^k)$ plotted for $k = 0, 8, 16$ and 24. Actual switch time: $\bar{t}^\star = 5.25$ (dash-dotted line).



(a) $q^\star = 1$



(b) $q^\star = 2$

Figure 5.2: Evolution of the estimate $\hat{t}_k^\star$ (above) and of the distribution $p_{t_k|k}$ (below; left bar: $p_{t_k|k}(0)$, center bar: $p_{t_k|k}(1)$; right bar: $p_{t_k|k}(2)$). Actual switch time: $\bar{t}^\star = 5.25$ (dash-dotted line).

that $t^\star > t_k$,

$$\begin{aligned}\mathbb{E}[t^\star|y^k, t^\star > t_k] &= \int_0^{+\infty} t^\star f(t^\star|y^k, t^\star > t_k)\mathrm{d}t^\star = \int_{t_k}^{+\infty} t^\star \Lambda e^{-\Lambda(t^\star - t_k)}\mathrm{d}t^\star \\ &= \Lambda^{-1} + t_k,\end{aligned}$$

where Λ^{-1} is the a priori expectation of $t^\star$. Thus, if condition $t^\star > t_k$ could be determined with certainty from the data y^k, the estimate of $t^\star$ would just increase linearly in time, because $t_k = kT$. In general, however, probability $\mathbb{P}[t^\star > t_k|y^k] = p_{t_k|k}(0)$ is less than one even before the actual switch. Indeed, the optimal estimate of $t^\star$ may be written as

$$\mathbb{E}[t^\star|y^k] = (\Lambda^{-1} + t_k)p_{k|k}(0) + \mathbb{E}[t^\star|y^k, t^\star \leq t_k](1 - p_{k|k}(0)).$$

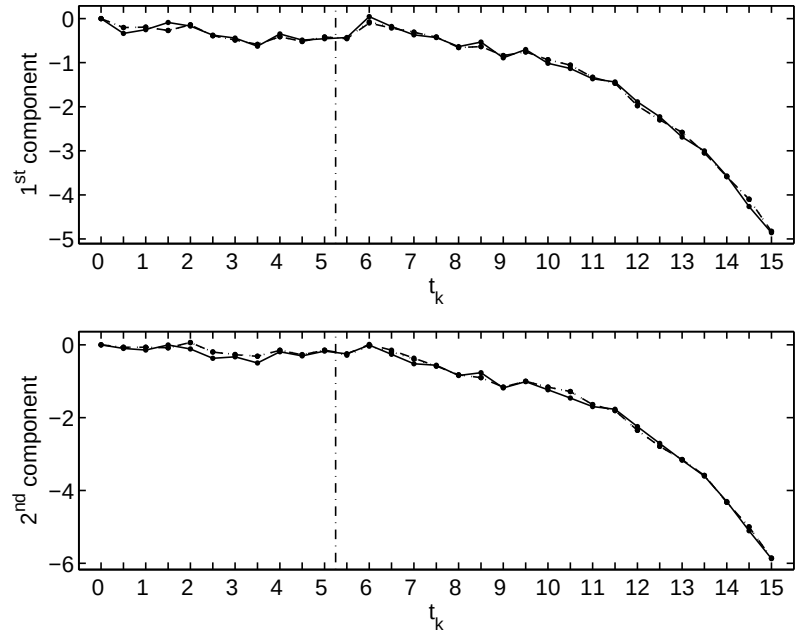
Moreover, as $0 \leq \mathbb{E}[t^\star|y^k, t^\star \leq t_k] \leq t_k$ for all y^k, one gets

$$\Lambda^{-1}p_{k|k}(0) + t_k p_{k|k}(0) \leq \mathbb{E}[t^\star|y^k] \leq \Lambda^{-1}p_{k|k}(0) + t_k.$$

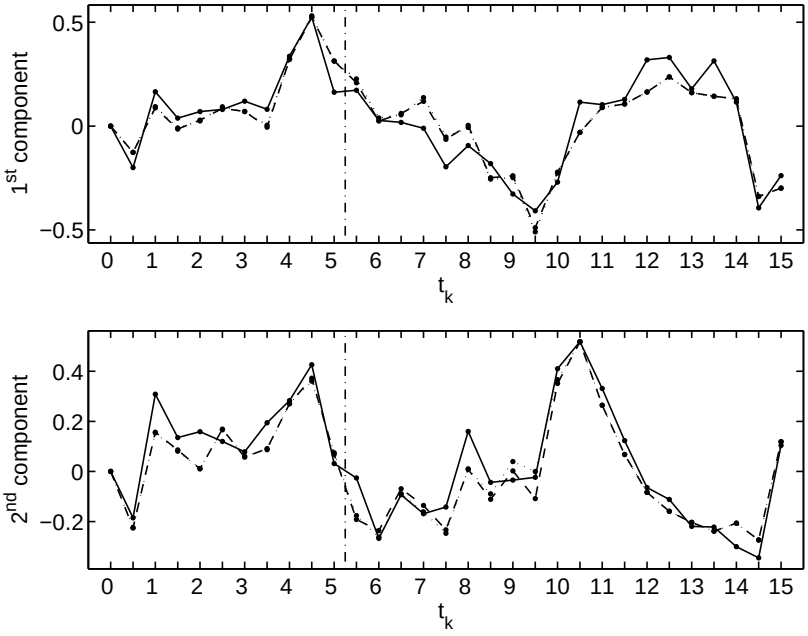
This provides bounds on the estimates $\hat{t}_k^\star$ in terms of the a posteriori probability of a switching event – the larger the value of $p_{t_k|k}(0)$, the tighter the bounds.

After the switch occurs, a small number of measurements suffice to detect the new discrete state of the system. The estimates of the switch time $t^\star$ also converge to the true value $\bar{t}^\star$ quite closely. However, comparison of the plots for $\bar{q}^\star = 1$ and $\bar{q}^\star = 2$ suggests that detecting a switch toward the unstable mode is "easier" than detecting a switch toward the damped oscillatory one. This is especially evident in the transient period following the switch. In fact, contrary to the discrete state value $q = 1$, both $q = 0$ and $q = 2$ give rise to stable modes, which keep the state of the system close to zero. Thus, due to the stochastic nature of the system, the first few measurements taken after $\bar{t}^\star$ are not sufficient to distinguish mode 0 from mode 2 when $\bar{q}^\star = 2$. On the other hand, they are quite indicative of the new system's dynamics when $\bar{q}^\star = 1$. In general, the more "different" the modes are, the quicker the algorithm is to detect the switch.

Plots of the estimates of the continuous state ξ at sample times t_k are finally drawn in Figure 5.3. The optimal estimates $\hat{\xi}_{t_k|k}$ are compared with the true values ξ_k and with the best estimates $\hat{x}_{k|k}^{\bar{t}^\star,\bar{q}^\star}$ one could produce in case the switching event were known in advance. In both the cases $\bar{q}^\star = 1$ and $\bar{q}^\star = 2$, estimates $\hat{\xi}_{t_k|k}$ follow the benchmark $\hat{x}_{k|k}^{\bar{t}^\star,\bar{q}^\star}$ quite accurately, even in the "transient" between the switch instant and the time when $p_{k|k}(\bar{q}^\star) \simeq 1$, that is, when the final value of $q(t)$ is determined with almost certainty. This



(a) $q^\star = 1$



(b) $q^\star = 2$

Figure 5.3: Evolution of the sampled state $\xi(t_k)$ (solid line) and of its estimates $\hat{x}_{k|k}^{\bar{t}^\star,\bar{q}^\star}$ (dotted line) and $\hat{\xi}_{t_k|k}$ (dashed line). Actual switch time: $\bar{t}^\star = 5.25$ (dash-dotted line).

is only partially surprising. Indeed, $\hat{\xi}_{t_k|k}$ is obtained by computing a weighted average of $\hat{x}_{k|k}^{t^\star,q^\star}$. The weighting term, $f(t^\star, q^\star|y^k)$, is proportional to factors such as (5.17). In our case, for a fixed matrix H, $y_k = x_k + Hv_k$, regardless of the value of $q(t_k)$. This suggests that the quality of the state estimate $\hat{x}_{k|k}^{t^\star,q^\star}$ directly relates with the value of the corresponding weight $f(t^\star, q^\star|y^k)$: the better the estimate is, the more relevant it is considered.

5.7 Discussion

We have presented an effective numerical implementation of state estimation for single-switch sampled stochastic systems. The method we illustrated relies on the characterization of the discrete-state trajectory in terms of the switch time $t^\star$ and of the final state $q^\star$. The linear system obtained by conditioning on $t^\star$ and $q^\star$ has been replaced by an equivalent discrete-time system with parameters depending on $t^\star$ and $q^\star$. This allowed to operate conditioned Kalman filtering in discrete-time. Using the superposition principle presented in Chapter 3, we made the dependence on the continuous values of $t^\star$ explicit and suited for adaptive numerical integration. Simulations show numerically accurate results and prove the effectiveness of Bayesian state estimation for switching systems with sampled measurements.

In terms of numerical complexity, it should be noted that Algorithm 5.1 yields an explicit representation of $\hat{x}_{\ell|k}^{t^\star,q^\star}$ and $P_{\ell|k}^{t^\star,q^\star}$ with $\mathcal{O}(k)$ parameters and $\mathcal{O}(k)$ operations for each update cycle. If one considers a single-switch discrete-time system, $k+1$ different trajectories q^k exist at time k for every value of the final state $q^\star$, one for each possible switch instant $k^\star$, $0 \leq k^\star \leq k$. Associated to each value of $k^\star$ are different conditioned Kalman estimates $\hat{x}_{\ell|k}^{k^\star,q^\star}$, $P_{\ell|k}^{k^\star,q^\star}$. Again, at time k, $\mathcal{O}(k)$ parameters are required to represent these estimates, and $\mathcal{O}(k)$ operations need be carried out for their update. Therefore, Algorithm 5.1 computes the parameters for the explicit representation of the continuous functions $\hat{x}_{\ell|k}^{t^\star,q^\star}$ and $P_{\ell|k}^{t^\star,q^\star}$ *at the same computational cost* of its discrete-time counterpart.

In Section 5.4, the optimal Bayesian estimators derived in Chapter 4 were implemented. However, simulations suggest that certain suboptimal approximations may perform nearly as well at a reduced computational cost. This is the rationale of the suboptimal algorithm 5.2, which was proposed in Section 5.5. In the discrete-time setting, the same idea motivates trajectory merging, see e.g. [10], [55] and the remark in Section 2.4. In more generality, the information provided by estimates conditioned on different discrete-state

trajectories may be redundant. Therefore, devising suboptimal approximations based on a restricted set of trajectories may still yield good estimators of the continuous-state. On the other hand, discriminating discrete-state trajectories on the basis of similar data statistics is a hard task. This raises the question of the distinguishability of discrete-state trajectories, which is an unsolved problem for both discrete-time and sampled systems with Markovian jumps.

The strategy we developed to solve state estimation for single-switch systems may appear a little peculiar. However, it should be noted that any switching system may be viewed as a concatenation in time of single-switch systems, provided the average dwell time in each discrete state is comparable to the time the system modes take to reach steady-state. In more generality, one may easily adapt the results of Sections 5.2÷5.3 to the case of an arbitrary – though bounded – number of switches over a finite time interval, the only requirement being that at most one switch occurs between two subsequent measurements. In most cases, this approximation is expected to be negligible and may lead to effective suboptimal estimators. In a discrete-time setting, the idea of considering a maximum number of switches in a finite number of steps was examined e.g. in [55]. It was shown that a drastic reduction in complexity may be obtained at a low probability of missing the true trajectory. Derivation of analogous results and their application to generic sampled switching systems is part of the aims of our future research.

Chapter 6

Mode estimation by hypothesis testing

This chapter explores the application of sequential hypothesis testing to the estimation of the discrete state trajectory of discrete-time switching linear systems. Although the interest is focused on jump Markov systems, the topic is discussed in rather general terms. We introduce a sequential rule for selecting a subset of most probable trajectories, and show its equivalence to a family of sequential likelihood ratio tests. We then focus on two issues. First, we study the ability of the test to discriminate between different discrete trajectories in connection with the structural properties of the system. Second, we present very general results on the probability of discarding the true trajectory and on how to choose the selection parameters so to keep this probability below a prescribed bound.

6.1 Introduction

In the context of jump Markov linear systems, one important issue is to estimate the whole discrete-state trajectory from a posteriori measurements. A first glimpse to the issue was given in Chapter 2, where it was shown that the a posteriori probability $p(q^k|y^k)$ intervenes in the averaging of the conditioned continuous state estimates. However, the problem is of interest *per se*, for at least two reasons: 1) In certain applications such as communication by randomly generated codes, see for instance [44], the continuous state measurements play the role of a "signature" of the discrete state sequence. The emphasis here is in recovering the sequence of symbols, less in estimating the

continuous-state trajectory; 2) It is experimental evidence, see [54] and references therein, that the performance in continuous-state and discrete-state estimation are not "commensurable". This is to say, situations exist where producing a good estimate of the continuous state is far more difficult than detecting the discrete state trajectory, or conversely, to the point that one of the two solutions may not have practical meaning. In this sense, the two problems should be regarded as different.

Given the switching model (2.1), recall that

$$p(q^k|y^k) \propto f(y^k|q^k)p(q^k) \tag{6.1}$$

where in turn

$$f(y^k|q^k) = \prod_{\ell=0}^{k} f(y_\ell|q^\ell, y^{\ell-1}).$$

Each of the factors is determined by the conditioned Kalman estimates $\hat{x}^{q^\ell}_{\ell|\ell-1}$ and $P^{q^\ell}_{\ell|\ell-1}$ as in (2.4). Therefore, besides normalization, quantity (6.1) is computable by k recursive operations for each possible trajectory q^k. On the other hand, since $|\mathcal{Q}| = N$,

$$|\{q^k\}| = |\{q_0\}| \times |\{q_1\}| \times \ldots \times |\{q_k\}| = N^{k+1},$$

i.e. N^{k+1} distinct trajectories exist at time k. As a consequence, that maximum-a-posteriori estimation of q^k is impracticable, because of the exponentially increasing number of probabilities $p(q^k|y^k)$ that should be considered. On the other hand, the a posteriori probability of a large number of trajectories shall be nearly zero.[1] Since

$$p(q^{k+1}|y^{k+1}) \propto f(y_{k+1}|y^k, q^{k+1})p(q^k|y^k),$$

a trajectory q^k for which $p(q^k|y^k) \simeq 0$ is candidate to extend to trajectories $q^{k+1} = (q^k, q_{k+1})$ for which $p(q^{k+1}|y^{k+1}) \simeq 0$. Therefore, a reasonable simplification is to discard at time k the trajectories with smaller $p(q^k|y^k)$, and to recast maximum-a-posteriori estimation as a search in the set of most probable trajectories. The problem becomes that of determining a suitable discard rule and to study how its application reflects into the estimation performance.

[1] Information theoretic arguments show that the true trajectory lyes, with high probability, in a *typical set* [20], that is, a small subset of trajectories with high *a priori* probability. Extensive application of these tools to discrete-state trajectory estimation is a current direction of research. First results of this type may be found in [44, 45].

In the next section, we will reformulate the discrete-state trajectory estimation problem in abstract terms. Based on a hypothesis testing approach, we will introduce a rule for trajectory selection and show its equivalence to a family of sequential likelihood ratio tests between two alternatives. Next, we will describe a measure of similarity between probability distributions. This will be used to determine the ability of the test to discriminate between different discrete trajectories. One section will be dedicated to discuss the overall performance of the trajectory selection approach. Final comments, hints and perspectives for our future research are reported in the concluding section.

6.2 Trajectory selection

Consider the jump Markov linear system (2.1). At any time $k \geq 0$, we want to determine a set $\Theta_k \subseteq \mathcal{Q}^{k+1}$ containing the trajectories with larger a posteriori probability $p(q^k|y^k)$.

Definition 6.1 *Let $q^k \in \mathcal{Q}^{k+1}$ and $\bar{q}^{k+\ell} \in \mathcal{Q}^{k+\ell+1}$. We say that q^k is a* prefix *of $\bar{q}^{k+\ell}$, and $\bar{q}^{k+\ell}$ is an* extension *of q^k, if*

$$\bar{q}^{k+\ell} = (q^k, q_{k+1}, \ldots, q_{k+\ell})$$

for some elements $q_{k+1}, \ldots, q_{k+\ell} \in \mathcal{Q}$.

Let $\Theta_{k-1} \times \mathcal{Q}$ denote the set of all possible extensions in $\mathcal{Q}^{k+1}$ of trajectories in Θ_{k-1}. Define

$$\tilde{p}_k \triangleq \max_{q^k \in \Theta_{k-1} \times \mathcal{Q}} p(q^k|y^k).$$

If $\Theta_{k-1} = \mathcal{Q}^k$, then $\tilde{p}_k$ is in correspondence with the maximum-a-posteriori estimator of q^k given y^k.

Definition 6.2 *At time k, we call* selection rule *of parameter κ, $0 \leq \kappa \leq 1$, a function $\mathcal{S}_k : \Theta_{k-1} \times \mathcal{Q} \to \{0, 1\}$ defined as follows:*

$$\mathcal{S}_k(q^k) = \begin{cases} 0, & \text{if } p(q^k|y^k) < \kappa \tilde{p}_k, \\ 1, & \text{if } p(q^k|y^k) \geq \kappa \tilde{p}_k. \end{cases} \tag{6.2}$$

In general, threshold κ may depend on k.

Definition 6.3 *The* most probable *set of trajectories Θ_k is defined recursively as follows:*

i. $\Theta_0 \triangleq \mathcal{Q}$;

ii. for $k > 0$, $\Theta_k = \{q^k \in \Theta_{k-1} \times \mathcal{Q} : S_k(q^k) = 1\}$.

Therefore, Θ_k is built from Θ_{k-1} by including the extensions having a posteriori probability sufficiently close to the maximum. Threshold κ regulates the size of set Θ_k: the larger the value of κ, the smaller the set Θ_k.

Proposition 6.1 *Let* $\bar{q}^{k-1} \in \Theta_{k-1}$. *An extension* $\bar{q}^k$ *of* $\bar{q}^{k-1}$ *belongs to* Θ_k *if and only if*

$$p(\bar{q}^k|y^k) \geq \kappa p(q^k|y^k)$$

for all $q^k \in \Theta_{k-1} \times \mathcal{Q}$.

Proof: Sufficiency: Let $\tilde{q}^k$ be an element of $\Theta_{k-1} \times \mathcal{Q}$ such that $p(\tilde{q}^k|y^k) = \tilde{p}_k$. If $p(\bar{q}^k|y^k) \geq \kappa p(q^k|y^k)$ for all $q^k \in \Theta_{k-1} \times \mathcal{Q}$, then, in particular,

$$p(\bar{q}^k|y^k) \geq \kappa p(\tilde{q}^k|y^k) = \kappa \tilde{p}_k,$$

whence $\mathcal{S}_k(\bar{q}^k) = 1$. Necessity: Let q^k be any element in $\Theta_{k-1} \times \mathcal{Q}$. Because $\mathcal{S}_k(\bar{q}^k) = 1$,

$$p(\bar{q}^k|y^k) \geq \kappa \tilde{p}_k \geq \kappa p(q^k|y^k),$$

where the definition of $\tilde{p}_k$ has been used. □

Thus, the selection of the most probable extensions may be done by comparing every possible pair of extensions. This result has an important theoretical consequence.

Corollary 6.1 *Let* $\bar{q}^{k-1} \in \Theta_{k-1}$. *An extension* $\bar{q}^k$ *of* $\bar{q}^{k-1}$ *belongs to* Θ_k *if and only if*

$$\frac{f(y^k|\bar{q}^k)}{f(y^k|q^k)} \geq \kappa \frac{p(q^k)}{p(\bar{q}^k)} \tag{6.3}$$

for all $q^k \in \Theta_{k-1} \times \mathcal{Q}$.

Proof: Simple application of Bayes'rule. □

Therefore, the selection rule $\mathcal{S}_k$ is equivalent to a set of likelihood ratio tests on the alternatives $f(y^k|q^k)$, $q^k \in \Theta_{k-1} \times \mathcal{Q}$. The test may be carried out

by computing $f(y^k|q^k)$ in place of $p(q^k|y^k)$ for all trajectories q^k whose prefix q^{k-1} belongs to Θ_{k-1}, according to the update rule

$$f(y^k|q^k) = f(y_k|q^k, y^{k-1}) f(y^{k-1}|q^{k-1}).$$

Note that, in general, neither of these alternatives is the true hypothesis, since $\Theta_{k-1} \times \mathcal{Q} \neq \mathcal{Q}^{k+1}$.

Remark. As a matter of fact, given two infinite sequences q^∞ and $\bar{q}^\infty$, it may happen that $q^\infty$is *never* discarded on the basis of $\bar{q}^\infty$ even when $\bar{q}^\infty$ is the true sequence. For instance, if $f(y^k|q^k) = f(y^k|\bar{q}^k)$ for every $k \in \mathbb{N}_0$ – which is the case if e.g. $q^\infty \equiv i$, $\bar{q}^\infty \equiv j$, $i, j \in \mathcal{Q}$, $i \neq j$ and $(A_i, B_i, C_i, D_i) = (A_j, B_j, C_j, D_j)$ – then the test will be decided on the basis of the a priori probabilities $p(q^k)$, $p(\bar{q}^k)$ only. This raises the question of the "distinguishability" of two sequences, and of course influences the expected size of the most probable set Θ_k. As we shall see, since $f(y_k|q^k, y^{k-1})$ is determined by the Kalman estimate $\hat{x}^{q^k}_{k|k-1}$, $P^{q^k}_{k|k-1}$, it is possible to establish a connection between the performance of the test (6.3) and the structure of the switching system (2.1). □

Remark. Given a set of most probable trajectories Θ_k, one may estimate q^k by solving

$$\max_{q^k \in \Theta_k} p(q^k|y^k) \propto \max_{q^k \in \Theta_k} f(y^k|q^k) p(q^k).$$

In general, the solution differs from the maximum-a-posteriori estimate

$$\hat{q}^k = \arg \max_{q^k \in \mathcal{Q}^{k+1}} p(q^k|y^k)$$

because the prefix q^ℓ of $\hat{q}^k$ may have been discarded at time $\ell < k$. Therefore, two kinds of error need be considered. The first error is a wrong estimate of the true trajectory when this belongs to Θ_k. This is essentially the same error possibly affecting $\hat{q}^k$, and may be studied by the tools of Chapter 1. The second error is the absence of the true trajectory in the set Θ_k. This eventuality is typical of a trajectory selection scheme, and will be investigated in a later section for a quite general choice of the selection rule. □

In the following, we will write (6.3) as the comparison

$$r^k(y^k) = \frac{f_1(y^k)}{f_0(y^k)} \lessgtr \kappa_k. \tag{6.4}$$

Here, $f_0(y^k)$ and $f_1(y^k)$ indicate the likelihood functions associated to two sequences in $\mathcal{Q}^{k+1}$ whose a priori probabilities are taken on by the time-varying threshold κ_k. The nature of the two sequences will be specified when necessary.

6.3 Bhattacharyya product

Let f_0 and f_1 be two generic density functions in $\mathbb{R}^p$.

Definition 6.4 *The* Bhattacharyya product *of f_0 and f_1 is given by*

$$\mathcal{B}(f_0, f_1) \triangleq \int \sqrt{f_0(y) f_1(y)} dy. \tag{6.5}$$

Note that $f_0, f_1 \in \mathcal{L}_1$. Therefore, $f_0^{1/2}$ and $f_1^{1/2}$ are functions of $\mathcal{L}_2$, and one may write

$$\mathcal{B}(f_0, f_1) = (\langle f_0^{1/2}, f_1^{1/2} \rangle)^2,$$

where $\langle \cdot, \cdot \rangle$ stands for Hilbert product. This shows that the Bhattacharyya product is well defined. An alternative expression is

$$\mathcal{B}(f_0, f_1) = \mathbb{E}_0[\sqrt{f_0(y)/f_1(y)}],$$

where $\mathbb{E}_0$ indicates expectation w.r.t density f_0.

Proposition 6.2 *The following properties hold:*

1. $0 \leq \mathcal{B}(f_0, f_1) \leq 1$, *with* $\mathcal{B}(f_0, f_1) = 1$ *if and only if* $f_0 = f_1$ *almost everywhere;*
2. $\mathcal{B}(f_0, f_1) = \mathcal{B}(f_1, f_0)$*;*
3. $\mathcal{B}(f_0, f_1) = \mathcal{B}(g_0, g_1)$ *for every* g_0, g_1 *such that* $g_i(y) = f_i(y - z)$, $z \in \mathbb{R}^n$*;*
4. $\mathbb{P}_0[\frac{f_1(y)}{f_0(y)} > \kappa] \leq \kappa^{-1/2} \mathcal{B}(f_0, f_1)$ *for every* $\kappa > 0$,

where $\mathbb{P}_0[\cdot]$ *is the probability measure induced by* f_0.

Proof: Property 1 follows from $f_i^{1/2} \in \mathcal{L}_2$ and $||f_i^{1/2}|| = 1$. Property 2 is obvious. Property 3 follows from the translation invariance of integration. To prove Property 4, let

$$\mathcal{A} \triangleq \{y : \frac{f_1(y)}{f_0(y)} > \kappa\} = \{y : \kappa^{-1/2} \frac{f_1^{1/2}(y)}{f_0^{1/2}(y)} > 1\}.$$

Then

$$\mathbb{P}_0[\frac{f_1(y)}{f_0(y)} > \kappa] = \int_{\mathcal{A}} f_0(y)dy \leq \int_{\mathbb{R}^p} \kappa^{-1/2} \frac{f_1^{1/2}(y)}{f_0^{1/2}(y)} f_0(y)dy = \kappa^{-1/2}\mathcal{B}(f_1, f_0).$$

□

In particular, Property 4 will be used to evaluate the probability of error α_0 in a likelihood ratio test.

Proposition 6.3 *Let $f_i = \mathcal{N}(\mu_i, \Sigma_i)$, $i = 0, 1$, with Σ_i nonsingular. Then $\mathcal{B}(f_0, f_1)$ is given by*

$$\rho \cdot \exp\left\{-\frac{1}{4}(\mu_0 - \mu_1)^T(\Sigma_0^{-1} - \Sigma_0^{-1}(\Sigma_0^{-1} + \Sigma_1^{-1})^{-1}\Sigma_0^{-1})(\mu_0 - \mu_1)\right\} \quad (6.6)$$

where

$$\rho = 2^{p/2} \det(\Sigma_0^{-1} + \Sigma_1^{-1})^{-1/2} \det(\Sigma_0)^{-1/4} \det(\Sigma_1)^{-1/4}.$$

Proof: Writing integral (6.5) explicitly and completing the square at the exponent, one gets

$$\mathcal{B}(f_0, f_1) = \rho \cdot \exp\{-\frac{1}{4}\mu_0^T\Sigma_0^{-1}\mu_0 - \frac{1}{4}\mu_1^T\Sigma_1^{-1}\mu_1 + \frac{1}{4}\bar{\mu}^T\bar{\Sigma}^{-1}\bar{\mu}\}$$

with $\bar{\mu} \triangleq \Sigma_0^{-1}\mu_0 + \Sigma_1^{-1}\mu_1$ and $\bar{\Sigma} \triangleq \Sigma_0^{-1} + \Sigma_1^{-1}$, compare [37]. Let now $\eta \triangleq \mu_0 - \mu_1$. Substitute $\bar{\mu} = \Sigma_0^{-1}\eta + \bar{\Sigma}\mu_1$ in $\bar{\mu}^T\bar{\Sigma}^{-1}\bar{\mu}$ and $\mu_0 = \eta + \mu_1$ in $\mu_0^T\Sigma_0^{-1}\mu_0$. Expand the products, collect all terms having leftmost factor η^T and rightmost factor η, and verify that the remaining terms cancel out. □

Corollary 6.2 *Let $f_i = \mathcal{N}(\mu_i, \Sigma_i)$, $i = 0, 1$, with Σ_i nonsingular. Then*

$$\max_{\mu_0, \mu_1} \mathcal{B}(f_0, f_1) = \rho,$$

attained for $\mu_0 = \mu_1$, and

$$\lim_{||\mu_0 - \mu_1|| \to +\infty} \mathcal{B}(f_0, f_1) = 0.$$

Proof: It is shown in [3] that

$$\Sigma_0^{-1} - \Sigma_0^{-1}(\Sigma_0^{-1} + \Sigma_1^{-1})^{-1}\Sigma_0^{-1} > 0.$$

Hence, the exponent of (6.6) is a negative definite quadratic form in $\mu_0 - \mu_1$. Its maximum value is equal to zero and is found for $\mu_0 - \mu_1 = 0$, whereas its limit for $||\mu_0 - \mu_1|| \to +\infty$ is $-\infty$. □

6.4 Stopping time of sequential testing

Consider a sequential test of the form (6.4). Assuming that f_0 is the true distribution of the measurements y^k, we wish to evaluate the probability that hypothesis f_1 is *rejected* in finite time, that is,

$$\frac{f_1(y^k)}{f_0(y^k)} < \kappa_k$$

at some $k \in \mathbb{N}_0$. Also, an evaluation of the expected stopping time $\mathbb{E}_0[\bar{k}]$ would be desirable, where

$$\bar{k} = \inf\{k : \frac{f_1(y^k)}{f_0(y^k)} < \kappa_k\}$$

and $\mathbb{E}_0$ indicates expectation taken w.r.t. f_0. Define

$$\mathcal{B}_k \triangleq \mathcal{B}(f_0(y^k), f_1(y^k)).$$

Proposition 6.4 *It holds that*

$$\mathbb{P}_0[\bar{k} > k] \leq \kappa_k^{-1/2} \mathcal{B}_k.$$

Proof: Apply Property 4 to the inequality

$$\mathbb{P}_0[\bar{k} > k] = \mathbb{P}_0[\bigcap_{\ell=0}^{k} \{\frac{f_1(y^\ell)}{f_0(y^\ell)} \geq \kappa_\ell\}] \leq \mathbb{P}_0[\frac{f_1(y^k)}{f_0(y^k)} \geq \kappa_k].$$

□

This provides a bound on the probability that the stopping time exceeds k. To check whether

$$\lim_{k \to +\infty} \mathbb{P}_0[\bar{k} > k] = 0,$$

and to establish a bound on $\mathbb{E}_0[\bar{k}]$, one needs to study the evolution in time of $\mathcal{B}_k$. This will be done by considering the recursive equation

$$f_i(y^k) = f_i(y_k|y^{k-1}) f_i(y^{k-1}). \tag{6.7}$$

Let

$$\hat{\mathcal{B}}_{k|k-1} \triangleq \sup_{y^{k-1}} \mathcal{B}(f_0(y_k|y^{k-1}), f_1(y_k|y^{k-1})).$$

Observe that, in light of expression (2.4),

$$f_i(y_k|y^{k-1}) = \mathcal{N}(\hat{y}^i_{k|k-1}, \Lambda^i_k) \tag{6.8}$$

where $\hat{y}^i_{k|k-1}$ is a linear function of y^{k-1} and Λ^i_k is independent of the data.

Proposition 6.5 *It holds that* $\hat{\mathcal{B}}_{k|k-1} \leq \rho_k$*, where*

$$\rho_k \triangleq 2^{p/2} \det\{(\Lambda_k^0)^{-1} + (\Lambda_k^1)^{-1}\}^{-1/2} \det(\Lambda_k^0)^{-1/4} \det(\Lambda_k^1)^{-1/4}.$$

Moreover, $\rho_k < 1$ *or* $\rho_k = 1$ *according to whether* $\Lambda_k^0 \neq \Lambda_k^1$ *or* $\Lambda_k^0 = \Lambda_k^1$*.*

Proof: Application of Corollary 6.2 to (6.8) yields

$$\mathcal{B}(f_0(y_k|y^{k-1}), f_1(y_k|y^{k-1})) \leq \rho_k.$$

Thus,

$$\hat{\mathcal{B}}_{k|k-1} = \sup_{y^{k-1}} \mathcal{B}(f_0(y_k|y^{k-1}), f_1(y_k|y^{k-1})) \leq \sup_{y^{k-1}} \rho_k = \rho_k,$$

because ρ_k does not depend on y^{k-1}. Observe that ρ_k is equal to the Bhattacharyya product of two Gaussian distributions with equal mean and variances Λ_k^0 and Λ_k^1. Hence, $\rho_k \leq 1$, with $\rho_k = 1$ if and only if $\Lambda_k^0 = \Lambda_k^1$. □

Proposition 6.6 *The following inequality holds:*

$$\mathcal{B}_k \leq \hat{\mathcal{B}}_{k|k-1} \mathcal{B}_{k-1}.$$

Proof: Applying equation (6.7) to the definition of $\mathcal{B}_k$,

$$\begin{aligned}
\mathcal{B}_k &= \int \sqrt{f_0(y^k) f_1(y^k)} dy^k \\
&= \int \left\{ \sqrt{f_0(y^{k-1}) f_1(y^{k-1})} \cdot \int \sqrt{f_0(y_k|y^{k-1}) f_1(y_k|y^{k-1})} dy_k \right\} dy^{k-1} \\
&\leq \int \left\{ \sqrt{f_0(y^{k-1}) f_1(y^{k-1})} \cdot \sup_{y^{k-1}} \int \sqrt{f_0(y_k|y^{k-1}) f_1(y_k|y^{k-1})} dy_k \right\} dy^{k-1} \\
&= \int \sqrt{f_0(y^{k-1}) f_1(y^{k-1})} dy^{k-1} \cdot \sup_{y^{k-1}} \int \sqrt{f_0(y_k|y^{k-1}) f_1(y_k|y^{k-1})} dy_k \\
&= \mathcal{B}_{k-1} \cdot \hat{B}_{k|k-1},
\end{aligned}$$

where the inequality holds because all integrands are nonnegative. □

Therefore, the sequence $\{\mathcal{B}_k\}$ is non-increasing. A sufficient condition for $\{\mathcal{B}_k\}$ to be strictly decreasing after a certain time $h \in \mathbb{N}_0$ is that $\Lambda_k^0 \neq \Lambda_k^1$ for all $k \geq h$.

Corollary 6.3 *It holds that*

$$\mathbb{P}_0[\bar{k} > k] \leq \kappa_k^{-1/2} \rho_k \cdot \rho_{k-1} \cdot \ldots \cdot \rho_0.$$

Proof: Straightforward from Propositions 6.4÷6.6. □

These tools allow to evaluate the performance of test (6.4) on the basis of the mutual properties of the (linear) systems indexed by $i = 0$ and $i = 1$.

Example. Consider a switching system (2.1) with $\mathcal{Q} = \{0, 1\}$. For every $q \in \mathcal{Q}$ let (A_q, B_q, C_q, D_q) be detectable and controllable, so that the algebraic Riccati equation

$$\bar{P}_q = A_q[\bar{P}_q - \bar{P}_q C_q^T (C_q \bar{P}_q C_q^T + D_q D_q^T)^{-1} C_q \bar{P}_q] A_q^T + B_q B_q^T$$

has a unique solution $\bar{P}_q \geq 0$ [52]. Assume that $\bar{\Lambda}_0 \neq \bar{\Lambda}_1$, where

$$\bar{\Lambda}_i \triangleq C_i \bar{P}_i C_i^T + D_i D_i^T.$$

Let $i = 0$ and $i = 1$ indicate the sequences

$$q_k = \begin{cases} 1, & k < h, \\ 0, & k \geq h, \end{cases},$$

for a certain $h \in \mathbb{N}_0$, and $\bar{q}_k \equiv 1$, in the order. Choose $\kappa_k \equiv \kappa < 1$.

For $k < h$, $q^k = \bar{q}^k$. Thus, $f_0(y^k) = f_1(y^k)$, and the test will not reject hypothesis 1. Note that $\rho_k = 1$. For $k = h + \ell$, $\ell \in \mathbb{N}_0$, q^k and $\bar{q}^k$ differ in the last ℓ values, and one has

$$\mathbb{P}_0[\bar{k} > h + \ell] \leq \kappa^{-1/2} \rho_{h+1} \cdot \ldots \cdot \rho_{h+\ell},$$

with $\rho_{h+1}, \ldots, \rho_{h+\ell}$ strictly less than 1. Since $\Lambda_{h+\ell}^i \to \bar{\Lambda}^i$ as $\ell \to +\infty$, then, for ℓ sufficiently large, $\rho_{h+\ell}$ is approximately equal to

$$\bar{\rho} \triangleq 2^{p/2} \det\{(\bar{\Lambda}^0)^{-1} + (\bar{\Lambda}^1)^{-1}\}^{-1/2} \det(\bar{\Lambda}^0)^{-1/4} \det(\bar{\Lambda}^1)^{-1/4}.$$

Therefore one gets the exponential decay

$$\mathbb{P}_0[\bar{k} > h + \ell] \approx \kappa^{-1/2} \bar{\rho}^\ell$$

whence [7] a finite expected stopping time $\mathbb{E}_0[\bar{k}]$, proportional to $\kappa^{-1/2}$, and $\mathbb{P}_0[\bar{k} = +\infty] = 0$. □

6.5 Accumulation of errors

The selection rule (6.2) may exclude the discrete-state trajectory underlying the generation of the data from the most probable set Θ_k. For a given trajectory q^k, let

$$\beta_k^{q^k} \triangleq \mathbb{P}[q^k \notin \Theta_k | q^k]$$

be the probability that q^k *is not in* Θ_k *at time* k. Similarly, let

$$\gamma_k^{q^k} \triangleq \mathbb{P}[q^k \notin \Theta_k | q^{k-1} \in \Theta_{k-1}, q^k],$$

with q^{k-1} prefix of q^k, be the probability that q^k *is discarded at time* k. For every fixed k, note that $\gamma_k^{q^k}$ decreases with κ_k, and that $\kappa_k = 0$ implies $\gamma_k^{q^k} = 0$. Hence, in principle, one may always choose a value of $\kappa_k > 0$ for which $\gamma_k^{q^k}$ is strictly less than a prescribed positive bound.

Proposition 6.7 *The following recursion holds:*

$$\beta_k^{q^k} = \beta_{k-1}^{q^{k-1}} + (1 - \beta_{k-1}^{q^{k-1}})\gamma_k^{q^k}.$$

In particular, $\{\beta_k^{q^k}\}$ *is non-decreasing, and* $\beta_{k-1}^{q^k} = 1$ *implies* $\beta_k^{q^k} = 1$.

Proof: Observe that $\mathbb{P}[q^{k-1} \in \Theta_{k-1} | q^{k-1}] = 1 - \mathbb{P}[q^{k-1} \notin \Theta_{k-1} | q^{k-1}]$. Then

$$\begin{aligned}\beta_k^{q^k} &= \mathbb{P}[q^k \notin \Theta_k | q^{k-1} \in \Theta_{k-1}, q^k] \cdot \mathbb{P}[q^{k-1} \in \Theta_{k-1} | q^{k-1}] + \\ &\quad \mathbb{P}[q^k \notin \Theta_k | q^{k-1} \notin \Theta_{k-1}, q^k] \cdot \mathbb{P}[q^{k-1} \notin \Theta_{k-1} | q^{k-1}] \\ &= \gamma_k^{q^k} \cdot (1 - \beta_{k-1}^{q^{k-1}}) + 1 \cdot \beta_{k-1}^{q^{k-1}}\end{aligned}$$

because $q^{k-1} \notin \Theta_{k-1}$ implies that $q^k \notin \Theta_k$. □

Next, denote with β_k the unconditioned probability that the outcome of q^k will not be contained in Θ_k. That is,

$$\beta_k \triangleq \mathbb{P}[q^k \notin \Theta_k],$$

where the probability is computed w.r.t. to the joint distribution $\mathcal{F}(y^k, q^k)$. Observe that

$$\beta_k = \sum_{q^k \in \mathcal{Q}^{k+1}} \beta_k^{q^k} p(q^k). \tag{6.9}$$

Proposition 6.8 *It holds that*

$$\beta_k = \beta_{k-1} + \sum_{q^{k-1}\in\mathcal{Q}^k} \Big\{ (1-\beta_{k-1}^{q^{k-1}}) p(q^{k-1}) \sum_{q_k\in\mathcal{Q}} \gamma_k^{q^k} p(q_k|q^{k-1}) \Big\}.$$

In particular, $\{\beta_k\}$ *is non-decreasing, and* $\beta_{k-1}=1$ *implies* $\beta_k = 1$.

Proof: Applying Proposition 6.7 to the equation (6.9) one gets

$$\begin{aligned}\beta_k &= \sum_{q^{k-1}} \sum_{q_k} (\beta_{k-1}^{q^{k-1}} + (1-\beta_{k-1}^{q^{k-1}})\gamma_k^{q^k}) p(q^{k-1}) p(q_k|q^{k-1}) \\ &= \sum_{q^{k-1}} \beta_{k-1}^{q^{k-1}} p(q^{k-1}) + \sum_{q^{k-1}} (1-\beta_{k-1}^{q^{k-1}}) p(q^{k-1}) \sum_{q_k} \gamma_k^{q^k} p(q_k|q^{k-1}),\end{aligned}$$

where the identity $\sum_{q_k} p(q_k|q^{k-1}) = 1$ was used. □

Corollary 6.4 *Assume that* $\gamma_k^{q^k} \equiv \gamma_k$, $\gamma_k \in [0,1]$, *for all* $q^k \in \mathcal{Q}^{k+1}$. *Then,*

$$\beta_k = \beta_{k-1} + (1-\beta_{k-1})\gamma_k. \tag{6.10}$$

Proof: By Proposition 6.8, if $\gamma_k^{q^k} \equiv \gamma_k$, then

$$\begin{aligned}\beta_k &= \beta_{k-1} + \gamma_k \sum_{q^{k-1}} (1-\beta_{k-1}^{q^{k-1}}) p(q^{k-1}) \sum_{q_k} p(q_k|q^{k-1}) \\ &= \beta_{k-1} + \gamma_k \{1 - \sum_{q^{k-1}} \beta_{k-1}^{q^{k-1}} p(q^{k-1})\},\end{aligned}$$

whence the result. □

When $\{\gamma_k^{q^k}\} \equiv \{\gamma_k\}$ – e.g. if thresholds κ_k are chosen so that the probabilities $\gamma_k^{q^k}$ are independent of q^k – both $\{\beta_k\}$ and $\{\beta_k^{q^k}\}$ take the form (6.10). Let us now study how the probability of error β_k evolves in correspondence to a specific sequence $\{\gamma_k\}$.

Proposition 6.9 *(i) For* $l=1,2$, *let* $\beta_k^{(l)}$ *and* $\gamma_k^{(l)}$ *satisfy (6.10). If, at a given* $k>0$, $\beta_{k-1}^{(1)} \geq \beta_{k-1}^{(2)}$ *and* $\gamma_k^{(1)} \geq \gamma_k^{(2)}$, *then* $\beta_k^{(1)} \geq \beta_k^{(2)}$. *(ii) Let* $\beta_0^{(l)} = 0$, $l=0,1$. *If* $\gamma_k^{(1)} \geq \gamma_k^{(2)}$ *for all* $k>0$, *then* $\beta_k^{(1)} \geq \beta_k^{(2)}$ *for all* $k \in \mathbb{N}_0$.

Proof: (i) It holds that

$$\begin{aligned}\beta_k^{(1)} &= \beta_{k-1}^{(1)} + (1-\beta_{k-1}^{(1)})\gamma_k^{(1)} \geq \beta_{k-1}^{(1)} + (1-\beta_{k-1}^{(1)})\gamma_k^{(2)} = \gamma_k^{(2)} + (1-\gamma_k^{(2)})\beta_{k-1}^{(1)} \\ &\geq \gamma_k^{(2)} + (1-\gamma_k^{(2)})\beta_{k-1}^{(2)}.\end{aligned}$$

(ii) The result follows by applying (i) iteratively on $k > 0$. □

This result is quite intuitive. In essence, it tells that smaller values of κ_k – i.e. larger values of γ_k – result in a larger probability that $q^k \notin \Theta_k$. Since κ_k influences the size of Θ_k – the larger the κ_k, the larger the Θ_k – we are interested in choosing the smallest values of κ_k that keep β_k within some prescribed bound. For $d \geq 1$ and an arbitrary $c \in (0,1)$, consider the sequence

$$\gamma_k = c \cdot k^{-d}, \quad k > 0.$$

Proposition 6.10 *For $\beta_0 < 1$, the following holds:*

i. If $d = 1$, $\lim_{k\to\infty} \beta_k = 1$;

ii. If $d > 1$, $\lim_{k\to\infty} \beta_k < 1$.

Proof: (i) It is equivalent to show that $\eta_k \to +\infty$, where $\eta_k = \beta_k/(1-\beta_k)$. Simple manipulations yield

$$\eta_k = \frac{\eta_{k-1} + \gamma_k}{1-\gamma_k}.$$

Since $\gamma_k \leq c < 1$, it holds that $\eta_k \geq \tilde{\eta}_k$, where, for $k > 0$, $\tilde{\eta}_k$ obeys

$$\tilde{\eta}_k = \tilde{\eta}_{k-1} + c \cdot k^{-1},$$

and $\tilde{\eta}_0 \triangleq \eta_0$. Hence,

$$\lim_{k\to\infty} \eta_k \geq \lim_{k\to\infty} \tilde{\eta}_k = \eta_0 + c \cdot \lim_{k\to\infty} \sum_{\ell=1}^{k} \ell^{-1} = +\infty.$$

(ii) We may equivalently prove that $\eta_k \triangleq -\log(1-\beta_k) \to \bar{\eta}$ with $\bar{\eta} < +\infty$. Using (6.10) one gets

$$\eta_k = -\log\{(1-\beta_{k-1})(1-\gamma_k)\} = \eta_{k-1} - \log(1-\gamma_k).$$

For $t \in (0, +\infty)$, consider $\gamma(t) \triangleq c \cdot t^{-d}$. Note that $\gamma(k) = \gamma_k$. Using the de l'Hospital rule, it is easily proven that, for any $\varepsilon \in (0, d-1)$,

$$\lim_{t \to +\infty} \frac{-\log(1 - c \cdot t^{-d})}{t^{-(1+\varepsilon)}} \stackrel{\text{H}}{=} \frac{c \cdot d}{1+\varepsilon} \cdot \lim_{t \to +\infty} \frac{t^{-(d-1)+\varepsilon}}{1 - c \cdot t^{-d}} = 0.$$

Therefore, there must exist some index ℓ such that $-\log(1-\gamma_k) \leq k^{-(1+\varepsilon)}$ for all $k > \ell$. It follows that, for $k > \ell$, $\eta_k \leq \tilde{\eta}_k$, where $\tilde{\eta}_k$ obeys

$$\tilde{\eta}_k = \tilde{\eta}_{k-1} + k^{-(1+\varepsilon)}.$$

with $\tilde{\eta}_\ell = \eta_\ell$. Then,

$$\lim_{k \to +\infty} \eta_k \leq \lim_{k \to \infty} \tilde{\eta}_k = \eta_\ell + c \cdot \sum_{k=\ell+1}^{+\infty} k^{-(1+\epsilon)} < +\infty.$$

□

In our case, $\beta_0 = 0$ because $\Theta_0 = \mathcal{Q}$ by definition. For choices of κ_k such that $\gamma_k \leq c \cdot k^{-d}$, $d > 1$, the probability β_k that $q^k \notin \Theta_k$ may be bounded away from 1. Note that, for $c = 0$, $\beta_k \equiv 0$, whereas, for $c = 1$, $\beta_k = 1$ for $k > 0$. With the aid of continuity arguments, Proposition 6.9 suggests that the bound may be made as close to zero as desired, provided the choice of a value of c small enough. In fact, one may alternatively consider the brute-force approximation

$$\beta_k = \beta_{k-1} + (1 - \beta_{k-1})\gamma_k \leq \beta_{k-1} + \gamma_k = \beta_0 + \sum_{\ell=1}^{k} \gamma_k = c \cdot \sum_{\ell=1}^{k} \ell^{-d},$$

which yields

$$\lim_{k \to +\infty} \beta_k \leq c \cdot \sum_{\ell=1}^{+\infty} \ell^{-d} = c \cdot \bar{\gamma},$$

where $\bar{\gamma} = \sum \ell^{-d}$ is finite for $d > 1$. Thus, if $\bar{\beta}$ is a desired bound for β_k, one shall choose the sequence $\{\kappa_k\}$ so that, for some $d > 1$, $\gamma_k \leq c \cdot k^{-d}$, with $c = \bar{\beta}/\bar{\gamma}$.

6.6 Discussion

We have studied the application of hypothesis testing to the problem of estimating the trajectory of the discrete-state jump Markov system. The

estimation problem was restated in terms of a maximum-a-posteriori search in a restricted set of most probable trajectories. This set is updated in time by a sequential trajectory selection rule based on the a posteriori probability of the trajectories given data y^k. We have shown the equivalence of this rule to a family of sequential tests between two alternative trajectories. The ability of one such test to distinguish trajectories has been related to the mutual properties of the system modes. This was done by way of a convenient measure of similarity between likelihood functions. Next, we considered the eventuality that, at a certain time k, the true trajectory lyes out of the most probable set of trajectories. The probability β_k of this event was shown to increase monotonically according to a simple update depending on the probability of discarding the true trajectory at each step k. Conditions were found for keeping the asymptotic probability of error below a prescribed bound.

The selection of a subset of most probable trajectories has been considered in several works, compare [55, 10, 2, 24] among the others. This is motivated by the need of handling the exponentially growing number of possible trajectories by finite complexity estimation algorithms. Moreover, it is a fact that most of the trajectories prove indistinguishable on the basis of the measured data y^k [54]. However, a large part of the currently available algorithms for trajectory estimation do not seem to have solid theoretical foundations. Our work attempts to determine the performance that an estimation algorithm may achieve on the basis of the structural properties of the switching system. A second aim is to provide an algorithm with known estimation performance. In the same spirit, an information-theoretic approach was recently pursued in [44, 45].

A number of questions deserve more investigations. In Section 6.5 we showed that bounding the error probability β_k away from one requires the probability γ_k of discarding the true trajectory at time k to decrease as k^{-d}, with $d > 1$. In principle, this decay may be achieved by choosing increasingly large thresholds κ_k. However, raising the threshold κ_k inevitably leads to larger trajectory sets Θ_k. We argue that, in general, no choice of the sequence $\{\kappa_k\}$ can simultaneously guarantee an upperbound for β_k strictly less than 1 and a finite bound for the size of Θ_k. If this were the case, the very statement of the problem should be put into question. In fact, the results of Section 6.4 suggest that trajectories that differ only in the remote past may not be distinguishable on the basis of future data. Therefore, one may consider reformulating trajectory estimation as the detection of states $q_{k-\ell}, \dots q_k$ given y^k, where index ℓ relates to the actual "memory" of the system. The same idea motivates the truncated maximum likelihood method presented in [53].

Along these lines, it seems reasonable to determine the set of most probable trajectories Θ_k out of a *finite* set of trajectories that differ in the last ℓ states. It is our intention to make these observations more precise by introducing notions of distinguishabiliy of two sequences. Hopefully, this will lead to consistent problem statements and will help both the development and the comparison of finite-complexity estimation algorithms.

Perspectives

We introduced a continuous-time dynamics jump Markov model and considered a general state estimation problem. In the basic case of a single-switch model we derived algorithms for both state estimation and fault detection. An algorithm of linearly increasing complexity has been tested in a simulated example. We plan to extend this work in several directions. Qualitative evaluation of the fault detection algorithm shall be supported by a quantitative analysis based on multiple Monte Carlo runs, for a sufficiently large set of model choices. We expect to obtain indications on the variance of the estimates of the continuous state and of the switch time as well as on the probability of error in the detection of the final discrete state. Performance comparison with the suboptimal estimation algorithm 5.2 would also be of interest. Implementation of state estimation algorithms for models with more complex switching is our second concern. This will require the study of convenient approximations of the optimal Bayesian estimators. In the spirit of the DE algorithm for JMLS, selection of a subclass of most probable trajectories shall be studied and exploited in approximate conditioned Kalman filter averaging. We believe that an adaptation of the results of Chapters 3 and 5 will come to help. Testing and application of the estimation algorithms on real-world problems is our ultimate goal.

The application of hypothesis testing to the detection of the discrete-state sequence of JMLS was also considered. This strategy is aimed at restricting the detection problem to an exhaustive maximum-a-posteriori search over a subset of most probable sequences. The ability of the test to discriminate sequences was examined on the basis of a measure of similarity between conditioned predictors. Results on the probability of pruning the true sequence were derived. The work we illustrated in Chapter 6 is by no means definitive. It represents our initial effort to produce finite-complexity detection algorithms with known statistical performance. We are currently working on an adaptive rule for the iterative pruning of unlikely sequences. In our intentions, this should yield a small, possibly bounded-size subset of

sequences where the true trajectory lies with high prescribed probability. In a system-theoretic perspective, intuition leads to consider an issue of sequence distinguishability. However, no accepted statement of this property exists. The problem is blurred by the stochastic nature of switching and typically involves finite-memory estimators, which puts the statement of the sequence detection problem into question. In the future, we plan to exploit tools of information theory to establish a precise connection between achievable performance and the properties of the model at hand.

Acknowledgements

My first thanks go to Prof. Giorgio Picci, for all the things I learned from him throughout these years. His encouragement and support were really appreciated all along. Personal thanks to Prof. Augusto Ferrante, for his devotion in helping the students of the group.

Thanks to all the people of the third floor, for the uncountable discussions and their precious advice. A special thought goes to Marione, for giving me The Idea in hopeless times (he knows what indecent proposal I am talking about). This makes me think of all my friends, Saccoman, Kekko, Ticuz, Nanni, Masie', Peràbo, Jean Batèin, Riki, Richard, Lzbth, Speps, and all the people that contributed to make the Ph.D. a fun experience of life.

Thanks to my family, Mutti, Co', Silvietta, for making me feel strong and loved. Thank you Lucia, here with me, for making me feel Home.

A note of disappointment goes to Er Finestra, for killing any reasonable hope to do something, and to February, for having three days less than a normal month.

This work was supported in part by the European Community through the project RECSYS of the V framework program.

Bibliography

[1] G. A. Ackerson and K. S. Fu. On state estimation in switching environments. *IEEE Transactions on Automatic Control*, 15(1):10–17, 1970.

[2] H. Akashi and H. Kumamoto. Random sampling approach to state estimation in switching environments. *Automatica*, 13, 1977.

[3] B.D.O. Anderson and J.B. Moore. *Optimal Filtering*. Prentice-Hall Inc., Englewood Cliffs, New Jersey, USA, 1979.

[4] A. Balluchi, L. Benvenuti, M. D. Di Benedetto, C. Pinello, and A. L. Sangiovanni-Vincentelli. Automotive engine control and hybrid systems: Challenges and opportunities. *Proceedings of the IEEE, special issue on hybrid systems*, 88(7):888–912, July 2000.

[5] R.H. Bartels and G.W. Stewart. Solution of the matrix equation $AX + XB = C$. *Communications of the ACM*, 15(9), 1972.

[6] M. Basseville and I. Nikiforov. *Detection of abrupt changes: theory and application*. Prentice-Hall, Inc., New Jersey, 1993.

[7] C.W. Baum and V.V. Veeravalli. A sequential procedure for multihypothesis testing. *IEEE Transactions on Information Theory*, 40(6):1994–2007, November 1994.

[8] R. Bhatia. *Matrix Analysis*. Springer-Verlag New York Inc., New York, USA, 1997.

[9] A. Blake, B. Bascle, M. Isard, and J. MacCormick. Statistical models of visual shape and motion. *Philosophical Transactions of the Royal Society*, 356:1283–1302, 1998.

[10] H. A. P. Blom and Y. Bar-Shalom. The interacting multiple model algorithm for systems with Markovian switching coefficients. *IEEE Transactions on Automatic Control*, 33(8):780–783, August 1988.

[11] P. Brémaud. *Markov chains: Gibbs fields, Monte Carlo simulation, and queues*. Springer-Verlag, New York, 1999.

[12] R. Chen and J. S. Liu. Mixture Kalman filters. *Journal of the Royal Statistical Society – Series B*, 62:493–508, 2000.

[13] E. Cinquemani. A superposition principle for the Kalman filter. *Systems and Control Letters*, 2004. Submitted.

[14] E. Cinquemani, A. Chiuso, A. Ferrante, R. Frezza, G.B. Gennari, A. Masiero, F. Nori, and G. Picci. Real-time hybrid estimation and identification algorithms for distributed sensor data processing. Technical report, University of Padova, Italy, May 2004. Project IST-2201-37170 RECSYS.

[15] E. Cinquemani and M. Micheli. State estimation in stochastic hybrid systems with sparse measurements. *IEEE transactions on automatic control*, January 2005. Submitted.

[16] E. Cinquemani, M. Micheli, and G. Picci. Fault detection in a class of stochastic hybrid systems. In *Proceedings of the 43th IEEE conference on decision and control*, Paradise Island, Bahamas, December 2004.

[17] E. Cinquemani, M. Micheli, and G. Picci. State estimation and prediction in a class of stochastic hybrid systems. In *Proceedings of the 16th International Symposium on Mathematical Theory of Networks and Systems*, Leuven, BE, July 2004.

[18] O. L. V. Costa. Linear minimum mean square error estimation for discrete-time Markovian jump linear systems. *IEEE Transactions on Automatic Control*, 39(8):1685–1689, August 1994.

[19] O. L. V. Costa and S. Guerra. Stationary filter for linear minimum mean square error estimator of discrete-time Markovian jump systems. *IEEE Transactions on Automatic Control*, 47(8):1351–1356, August 2003.

[20] T.M. Cover and J.A Thomas. *Elements of information theory*. John Wiley & sons, New York, 1991.

[21] A. Doucet and C. Andrieu. Iterative algorithms for state estimation of Jump Markov linear systems. *IEEE Transactions on Signal Processing*, 49(6):1216–1227, June 2001.

[22] A. Doucet, N. de Freitas, and N. Gordon. *Sequential Monte Carlo methods in practice*. Springer-Verlag, New York, 2001.

[23] A. Doucet, N.J. Gordon, and V. Krishnamurthy. Particle filters for state estimation of jump Markov linear systems. *IEEE Transactions on Signal Processing*, 49(3):613–624, March 2001.

[24] A. Doucet, A. Logothetis, and V. Krishnamurthy. Stochastic sampling algorithms for state estimation of Jump Markov linear systems. *IEEE Transactions on Automatic Control*, 45(2):188–201, 2000.

[25] R. J. Elliott, F. Dufour, and D. D. Sworder. Exact hybrid filters in discrete time. *IEEE Transactions on Automatic Control*, 41(12):1807–1810, December 1996.

[26] A. Ferrante, G. Picci, and S. Pinzoni. Silverman algorithm and the structure of discrete-time stochastic systems. *Linear Algebra and its Applications (Special Issue on Linear Systems and Control)*, 351–352:219–242, 2002.

[27] E. Fornasini and G. Marchesini. *Appunti di teoria dei sistemi*. Edizioni libreria Progetto, Padova, 1994.

[28] W. Gander and W. Gautschi. Adaptive quadrature – revisited. *BIT Numerical Mathematics*, 40:84–101, 2000.

[29] A. Germani, C. Manes, and P. Palumbo. Polynomial filtering for stochastic systems with Markovian switching coefficients. In *Proceedings of the 42nd IEEE Conference on Decision and Control (CDC 2003)*, pages 1392–1397, Maui, Hawaii, December 2003.

[30] W. Glover and J. Lygeros. A stochastic hybrid model for air traffic control simulation. In *Hybrid Systems: Computation and Control (Proceedings of the 7th International Workshop)*, Berlin, March 2004. Springer-Verlag.

[31] J. Hespanha. Stochastic hybrid systems: application to communication networks. In *Hybrid Systems: Computation and Control (proceedings)*, pages 387–401, Berlin, 2004. Springer-Verlag.

[32] J. L. Hibey and C. D. Charalambous. Conditional densities for continuous-time nonlinear hybrid systems with application to fault detection. *IEEE Transactions on Automatic Control*, 44(11):2164–2169, November 1999.

[33] M. W. Hofbaur and B. C. Williams. Mode estimation of probabilistic hybrid systems. In *Hybrid Systems: Computation and Control (HSCC 2002)*, Lecture Notes on Computer Sciences. Springer Verlag, 2002.

[34] R.A. Horn and C.R. Johnson. *Topics in matrix analysis*. Cambridge University Press, New York, 1991.

[35] J. Hu, J. Lygeros, and S. S. Sastry. Towards a theory of stochastic hybrid systems. In *Third International Workshop on Hybrid Systems: Computation and Control*, Pittsburgh, PA, 2000. Springer Verlag Lecture Notes on Computer Science, vol. 1790.

[36] A. H. Jazwinski. *Stochastic Processes and Filtering Theory*. Academic Press, London, 1970.

[37] T. Jebara and R. Kondor. Bhattacharyya and expected likelihood kernels. In *Proceedings of the Sixteenth Annual Conference on Computational Learning Theory*, Washington D.C., August 2003.

[38] M. Kijima. *Markov Processes for Stochastic Modeling*. CRC Press, London, UK, 1997.

[39] X. Koutsoukos, J. Kurien, and F. Zhao. Monitoring and diagnosis of hybrid systems using particle filtering methods. In *Proceedings of the Fifteenth International Symposium on the Mathematical Theory of Networks and Systems (MTNS '02)*, University of Notre Dame, South Bend, Indiana, August 2002.

[40] X. Koutsoukos, J. Kurien, and F. Zhao. Estimation of distributed hybrid systems using particle filtering methods. In *Hybrid Systems: Computation and Control (HSCC 2003)*. Springer Verlag Lecture Notes on Computer Science, vol. 2623, Pittsburgh, PA, 2003.

[41] U. Lerner, R. Parr, D. Koller, and G. Biswas. Bayesian fault detection and diagnosis in dynamic systems. In *Proceedings of the 17th National Conference on Artificial Intelligence (AAAI)*, pages 531–537, Austin, Texas, July 2000.

[42] A. Logothetis and V. Krishnamurthy. Expectation maximization algorithms for map estimation of jump Markov linear systems. *IEEE Transactions on Signal Processing*, 47(8):2139–2156, August 1999.

[43] J. Lygeros and M. Prandini. Aircraft and weather models for probabilistic collision avoidance in air traffic control. In *Proceedings of the 41st IEEE conference on decision and control*, volume 3, pages 2427–2432, Berlin, December 2002. Springer-Verlag.

[44] N.C. Martins and M.A. Dahleh. An information theoretic approach to the mode estimation of randomly switching FIR systems. In *Proceedings of the 11th Mediterranenan Control Conference*, Rhodes, Greece, 2003.

[45] N.C. Martins and M.A. Dahleh. Rate distortion in the modal estimation of switching FIR linear systems. In *Proceedings of the 43rd Conference on Decision and Control*, pages 3575–3580, Paradise Island, Bahamas, 2004.

[46] E. Mazor, A. Averbuch, Y. Bar-Shalom, and J. Dayan. Interactive multiple model methods in target tracking: a survey. *IEEE transactions on aerospace and electronic systems*, 34(1):103–123, January 1998.

[47] B. M. Miller and W. J. Runggaldier. Kalman filtering for linear systems with coefficients driven by a hidden Markov jump process. *Systems and Control Letters*, 31:93–102, 1997.

[48] K. P. Murphy. Switching Kalman filters. Report 98-10, Compaq Cambridge Research Laboratory, 1998.

[49] I. V. Nikiforov. Optimal sequential detection and isolation of changes in stochastic systems. Technical Report 2063, Institut National de Recherche en Informatique et en Automatique (INRIA), September 1993.

[50] B. North, A. Blake, M. Isard, and J. Rittscher. Learning and classification of complex dynamics. *IEEE transactions on pattern analysis and machine intelligence*, pages 1016–1034, 2000.

[51] T. Pappas, A.J. Laub, and N.R. Sandell. On the numerical solution of the discrete-time algebraic Riccati equation. *IEEE Transactions on Automatic Control*, 25(4):631–641, August 1980.

[52] G. Picci. *Filtraggio statistico (Wiener, Levinson, Kalman) e applicazioni*. Edizioni libreria Progetto, Padova, 1998.

[53] J. K. Tugnait. Adaptive estimation and identification for discrete systems with Markov jump parameters. *IEEE Transactions on Automatic Control*, 27(5):1054–1065, October 1982.

[54] J. K. Tugnait. Detection and estimation for abruptly changing systems. *Automatica*, 18(5):607–615, 1982.

[55] J.K. Tugnait and A.H. Haddad. A detection-estimation scheme for state estimation in switching environments. *Automatica*, 15:477–481, 1979.

[56] H.L. Van Trees. *Detection, estimation and modulation theory (Part 1)*. John Wiley & sons, New York, 1968.

[57] A. Wald. *Sequential analysis*. John Wiley & sons, New York, 1947.

[58] Q. Zhang. Hybrid filtering for linear systems with non-Gaussian disturbances. *IEEE Transactions on Automatic Control*, 45(1):50–61, January 2000.