ON EXACT FINITE DIMENSIONAL FILTERS
IN [illegible] TIME

[illegible] III

JOSEPH NELSON RUSHTON III
On Exact Finite Dimensional Filters in Mixed Time
(Under the direction of D. KANNAN)

This dissertation considers filtering for mixed-time models, i.e. models with continuous time signal and discrete time observations. We consider linear Gaussian models without switching, models in which the switching parameter is a Markov chain, and models with deterministic (but unknown) switching parameters. In all cases, exact filters are derived for the mean and covariance of the state, and for quantities associated with EM parameter estimation. All filters are derived using change of measure. We discuss the issue of convergence to continuous-time filters as the duration between successive observations becomes negligible.

INDEX WORDS: Stochastic filtering, Mixed time filtering, EM algorithm, Exact filter, Finite dimensional filter

TABLE OF CONTENTS

page

CHAPTER 1. Introduction 1

1.1 The Filtering Problem 1

1.2 Preliminaries 11

1.3 Summary of Contents 14

CHAPTER 2. Filters for the State in the Linear Model 17

2.0 Model 17

2.1 Exact Filter 19

2.2 Finite Dimensional Exact Filter 22

2.3 Limiting Case 28

CHAPTER 3. Parameter Estimation for the Linear Case 32

3.0 Model 32

3.1 Recursive Exact Filters 34

3.2 Limiting Case 42

3.3 Change of Measure and EM Algorithm 47

CHAPTER 4. Markov Switching Parameter 52

CHAPTER 5. Deterministic Switching 57

5.1 Model and Change of Measure 57

5.2 EM Algorithm 60

5.3 Filters 63

REFERENCES 69

Chapter I. Introduction

1.1. The Filtering Problem

We begin by describing the filtering problem and certain related topics relevant to this dissertation. We will be cavalier at first about the domains and ranges of various functions involved, because they vary widely. Suppose we are given two random processes x_t and y_t. We think of x_t as a hidden 'signal process,' such as an evasive enemy aircraft, which is observable only through the related 'observation process' y_t, such as radar data. Loosely speaking, the filtering problem is to 'find out about' the signal at time t based on the history of the observations up to time t. To be more precise, let $\mathcal{Y}_t$ be the sigma algebra generated by $\{y_s, s \leq t\}$. We would like to find, as a function of t, the conditional density (or some characteristic thereof), given $\mathcal{Y}_t$, of some given functional F_t of $\{(x_s, y_s), s \leq t\}$. For example, we might wish to compute the conditional density or expectation, given $\mathcal{Y}_t$, of any of the following:

- $\sum_{s \leq t} x_s$ (in the case where x_t is a discrete-time process);
- $\int_0^t x_t dy_t$ (in the case where x_t and y_t are continuous-time processes);
- The number N_t of times x_t has changed state up to time t (in the case x is a jump process).

One of the most important filtering problems is perhaps the simplest to state, namely the determination of $E[x_t \mid \mathcal{Y}_t]$.

Any unpredictable system whose behavior we wish to estimate based on incomplete data may be interpreted as an application of stochastic filtering. For example, during World War II, Norbert Wiener lead an investigation in the use of filtering theory to estimate the position of enemy aircraft given noisy radar data. Target tracking in background noise (incidental, or as a result of hostile countermeasures) is still an important application of filtering. Other uses range from navigation (Brown, [8]) to seismic signal processing (Sims and D'mello, [26]).

The greatest breakthrough in filtering theory to date came around 1960, with the discovery of the following theorems by Rudolf Kalman:

Theorem 1.1. DISCRETE-TIME KALMAN FILTER. *Let $\{w_k, k \geq 0\}$ and $\{v_k, k \geq 1\}$ be independent sequences of iid rv's distributed respectively $N(0, I_n)$ and $N(0, I_m)$. Let $x_0 \in \mathbb{R}^m$ have Gaussian density with known mean and covariance matrix, and let sequences $\{x_k, k \geq 1\}$ and $\{y_k, k \geq 0\}$, in $\mathbb{R}^m$ and $\mathbb{R}^n$ respectively, be given by*

$$x_k = A_k x_{k-1} + B_k v_k, \tag{1.1.1}$$

$$y_k = C_k x_k + D_k w_k, \tag{1.1.2}$$

where A, B, C, and D are sequences of real matrices of appropriate dimension with B_k and D_k being symmetric and positive definite for all k. Let $\mathcal{Y}_k := \sigma\{y_j, j \leq k\}$ and define

$$\hat{x}_k := E[x_k|\mathcal{Y}_k],$$

$$\bar{x}_k := E[x_k|\mathcal{Y}_{k-1}],$$

$$\hat{P}_k := \; cov[x_k|\mathcal{Y}_k],$$

$$\bar{P}_k := \; cov[x_k|\mathcal{Y}_{k-1}].$$

Then the conditional density of x_k given $\mathcal{Y}_k$ is Gaussian for each k; and $\hat{x}_k, \bar{x}_k$, $\hat{P}_k$, and $\bar{P}_k$ satisfy the following recursive equations:

$$\bar{x}_{k+1} = A_{k+1}\hat{x}_k$$

$$\bar{P}_{k+1} = A_{k+1}\hat{P}_k A'_{k+1} + B^2_{k+1}$$

$$\hat{P}_{k+1} = \bar{P}_{k+1} - \bar{P}_{k+1}C_{k+1}(C_{k+1}\bar{P}_{k+1}C'_{k+1} + D^2_{k+1})^{-1}C_{k+1}\bar{P}_{k+1}.$$

$$\hat{x}_{k+1} = K_{k+1}y_{k+1} + (I - K_{k+1}C_{k+1})\bar{x}_{k+1},$$

where K_k, the so-called Kalman gain at time k, is defined by

$$K_k := \hat{P}_k C_k' D_k^{-2}.$$

Theorem 1.2. Continuous-Time Kalman Filter. *Let W and V be independent standard Brownian motions in $\mathbb{R}^m$ and $\mathbb{R}^n$ respectively. Let $x_0 \in \mathbb{R}^m$ have Gaussian density with known mean and covariance matrix, and let processes $\{x_t, t \geq 0\}$ and $\{y_t, t \geq 0\}$, in $\mathbb{R}^m$ and $\mathbb{R}^n$ respectively, satisfy*

$$dx_t = A_t x_t dt + B_t dW_t, \qquad t \in [0, T], \tag{1.1.3}$$

$$dy_t = C_t x_t dt + D_t dV_t, \qquad t \in [0, T], \tag{1.1.4}$$

where A, B, C, D are real matrix-valued functions on $[0, T]$ (satisfying the usual conditions guaranteeing a unique solution to the system) with B_t and D_t symmetric and positive definite for all t. Let $\mathcal{Y}_t := \sigma\{y_s, s \leq t\}$, and define

$$\hat{x}_t : = E[x_t|\mathcal{Y}_t],$$
$$\hat{P}_t : = \ cov[x_k|\mathcal{Y}_t].$$

Then the conditional density of x_t given $\mathcal{Y}_t$ is Gaussian for each t, and $\hat{P}_t$ and $\hat{x}_t$ satisfy

$$d\hat{P} = \left[A_t\hat{P}_t + \hat{P}_t A_t' - \hat{P}_t C_t' D_t^{-2} C_t \hat{P}_t + B_t^2\right] dt,$$
$$\hat{x}_t = \int_0^t A_s \hat{x}_s ds + \int_0^t K_s dy_s - \int_0^t K_s C_s \hat{x}_s ds,$$

where K_s is defined on $[0, T]$ by

$$K_s = \hat{P}_s C_s' D_s^{-2}.$$

The practical (economic) and theoretical impact of the Kalman filters has been tremendous ([16], p. 15). Much, if not most, of the research in stochastic filtering for the past thirty years has revolved around extending these results in various directions, but in a sense with only limited success. Albeit for certain special models, the Kalman filters are very good in that they are 'exact finite dimensional filters,' *i.e.*, they completely characterize the conditional densities in terms of a fixed, finite number of statistics (four in the discrete-time filter, two in the continuous time filter). This has turned out to be difficult, if not (provably) impossible for even slightly more general models.

To further clarify this point, we give an example of an infinite dimensional filter:

Theorem 1.3. ELLIOTT, [12]. *Let x_k and $y_k, k \geq 0$, be as in Theorem 1.1, except that v_k and w_k are iid with densities ψ and φ respectively, not necessarily Gaussian. Let $p_k(x)$ be the conditional density of x_k given $\mathcal{Y}_k$. Let α_0 be the density of x_0 and for $k = 1, 2, \ldots$, define $\alpha_k : \mathbb{R}^m \to \mathbb{R}$ recursively as follows:*

$$\alpha_{k+1}(x) = \frac{\varphi(y_{k+1} - C_{k+1}(x))}{\varphi(y_{k+1})} \int_{\mathbb{R}^m} \psi(x - A_{k+1}(u))\, \alpha_k(u) du.$$

Then, for $k \geq 1$,

$$p_k(x) = \frac{\alpha_k(x)}{\int_{\mathbb{R}^m} \alpha_k(u) du}.$$

Theoretically, the above gives us an expression for the conditional density $p_k(x)$ for each k and x; so Theorem 1.3 gives an exact filter. Numerically, however, it is a nightmare. Suppose we have computed α_k. In order to compute α_{k+1}, we must perform an integration over all of $\mathbb{R}^m$ for each value of x for which we seek $\alpha_{k+1}(x)$. Before moving on to compute α_{k+2}, we must compute α_{k+1} for a large enough set of values to obtain a global approximation of α_{k+1}. The errors in such a process are significant and accumulate with each iteration; hence, for most practical purposes, this filter cannot be implemented *as is*. (We shall see, however, in that it does have redeeming value as a tool for developing more practical filters).

The class of finite dimensional, non-exact filters includes the Extended Kalman Filter (McGee and Schmidt, [23]) and 'pruned' versions of the Interactive Multiple

Model (IMM) algorithm (*e.g.*, Blom and Bar-Shalom, [7]). Such filters provide 'fast' algorithms, but often yield estimates of questionable value (Daum, [10]).

Given any filtering problem, we clearly would hope for a finite dimensional exact filter; but the search for such filters has been difficult. In fact, there are only two such filters in wide use: the Kalman filter (above) and the Wonham filter (Wonham, [28]), (see Elliott and Krishnamurthy, [15]).

Mixed-Time Filtering

Most filtering theory has dealt with one of two cases: continuous-time signal with continuous-time observations (as in Theorem 1.2), or discrete-time signal with discrete-time observations (as in Theorem 1.1). Recently however, applications have motivated researchers (*e.g.* Kouritzin [20]; Daum [10]) to consider 'mixed time models', *i.e.*, models with continuous-time signal and discrete-time observations. This dissertation investigates several topics in mixed time filtering.

We adopt the general approach to mixed time models treated by Jazwinski ([17]), whose proscription is roughly is follows (details vary by case, and are given in Chapters II-V): Suppose we have a diffusion signal x_s on $[0, t]$ observable only through observations z_i made at times $\delta, 2\delta, \ldots, k\delta$, where k is the greatest integer less than or equal to t/δ, i.e. $k = [t/\delta]$. Each observation z_i is taken to be a (stochastically) noisy function of $x_{i\delta}$. Our goal is to obtain the conditional density of x_t given $Z_k = \sigma\{z_i, i \leq k\}$. Letting $x_i := x_{i\delta}, i \leq k$, we use the semigroup property of diffusions (Theorem 1.9) to write a recursive formula for the x_i. We can now borrow from the extensive arsenal of purely discrete-time techniques to obtain the conditional density of x_k $(= x_{k\delta})$ given Z_k. Finally, we observe that in the the interval $(k\delta, t)$, the conditional density of x_t given Z_k evolves according to the Kolmogorov forward equation, which allows us to compute the conditional density of x_t given Z_k.

From the above, we might think that mixed time filtering is merely discrete-time filtering with a little extra baggage. To an extent, this is true—but this doesn't mean there are not good reasons to study mixed time filtering. First, many applications fall most naturally into the mixed-time domain. This is the case, for instance, in the quite common situation where a continuously evolving phenomenon is monitored by a digital device; and engineers nowadays prefer mixed time models over approximate 'pure time' models (i.e. purely continuous or purely discrete time) for applications such as target tracking and navigation.

There are also theoretical reasons for studying mixed-time models. The discrete-time models which arise out of mixed time problems are limited to certain special cases. The study of mixed-time filtering hence motivates giving special attention to these cases, and can lead the discrete-time field to new research problems whose importance might not otherwise have been seen. These special case discrete-time models also fall themselves under the umbrella of mixed time research. In fact, we will see that a large portion of this work is devoted to what are essentially discrete-time problems motivated by mixed-time considerations. Finally, mixed-time models can give new insights into the purely continuous-time set-up, especially with regards to the stability of continuous-time algorithms.

Models with Switching

One direction in which the work of Kalman may be generalized is to take the coefficients A, B, C, and D themselves as stochastic processes. Certain applications have motivated the consideration of models with 'switching,' in which one or more of the coefficients A, B, C, D are functions of a finite-state stochastic 'switching parameter,' often denoted by θ_t. Such a system is thought of as having multiple modes of evolution, characterized by different values of θ_t. The trajectory of an aircraft tracked by military radar, for example, might be thought of as having two modes of evolution, 'maneuvering' and 'not maneuvering': A maneuvering target is 'more random,' corresponding to a larger norm for the B matrix in Theorem

1.2. Applications of switched models include

* Systems subject to failures (*e.g.* Caglayan, [9]; Basseville and Nikiforov, [4]);
* Tracking maneuvering targets (*e.g.* Bar-Shalom and Li, [3]; Mazor, *et al*, [22]).
* Piecewise linearization of nonlinear systems : The mode parameter θ_t serves to identify the correct 'piece' to be used for linearization (*e.g.* Moose *et al*, [24]; Skeppstedt *et al*, [27]).

Mixed time models with switching are considered in Chapters IV and V of this dissertation.

Change of Measure

One technique used in filtering is a change of probability measure (see, *e.g.*, Elliott, Dufour, and Sworder 1996; Elliott and Krishnamurthy 1996). This is illustrated most simply with a discrete-time example: Suppose we have an unobservable stochastic sequence $x_k, k \geq 0$, given by (1.1.1), under the assumptions of Theorem 1.1. Then it is easy to give a recursive integral formula for the density of x_k. Now, suppose $y_k, k \geq 1$, is an observation process given by (1.1.2). To give a recursive formula for the *conditional* density of x_k given $\mathcal{Y}_k$ amounts to proving a version of Theorem 1.1, and is not so easy. However, it can be shown that under an alternate probability measure $\overline{P}$, the sequences $\{x_k\}$ and $\{y_k\}$ are independent, while the unconditional density of x_k remains unchanged. Under $\overline{P}$, the conditional density of x_k given $\mathcal{Y}_k$ is equal to the unconditional density of x_k, which is easy to compute. Now, if we can lay our hands on the Radon-Nikodym derivative of $\overline{P}$ with respect to P, we can use the conditional Bayes' Theorem (1.4), given *infra*, to 'transform' our results back to the original probability measure, obtaining the conditional density of x_k given $\mathcal{Y}_k$ under P. In practice, $\overline{P}$ is obtained by judiciously constructing its Radon-Nikodym derivative *w.r.t.* P so as to yield the desired independence properties. This technique can be used to give a recursive integral formula (as in Theorem 1.3) for the conditional density in any filtering problem where, 1) We can compute the desired *unconditional* density, and 2) We can construct, via Radon-Nikodym derivative, a probability measure under which the desired independence properties hold. The filters obtained by change of measure are always exact, and in some cases can be used to derive finite-dimensional exact filters. Change of measure will be a primary tool used in this dissertation. A survey of measure change methods used in filtering is given in (Elliott, Aggoun, and Moore, [12]).

Maximum Likelihood Estimation and the EM Algorithm

In many applications, the parameters of the model (such as A, B, C, and D in the Kalman filter case) are initially not known, or known poorly. Such applications include economic forecasting and the estimation of speech signals (Elliott and Krishnamurthy, [15]). If the parameters are modeled as deterministic, then the problem of estimating them falls into the domain of parametric statistics. A frequently used criterion is to choose a 'Maximum Likelihood' (ML) estimate for the parameter set. This is chosen roughly as follows: Suppose we are given a system of signal and observations supported on a complete probability space $(\Omega, \mathcal{F}, P)$, and that the system dynamics depend on a parameter set Θ. To each hypothetical parameter set $\hat{\Theta}$, associate a probability measure $P(\hat{\Theta})$ under which the system dynamics hold with Θ replaced by $\hat{\Theta}$, (the details of this construction are given for certain cases in Chapters III and V). An ML estimate of Θ at time k is a parameter set $\hat{\Theta}$ which maximizes

$$\mathcal{L}_k(\hat{\Theta}) := E\left[\frac{dP(\hat{\Theta})}{dP} \,\Big|\, \mathcal{Y}_k\right].$$

The intuition behind the ML estimate is that it gives the model under which the observations which actually occurred are 'most likely': Let $Y_k := (y_1, y_2, \ldots, y_k)$, and let φ be the unconditional density of Y_k. In this set-up, we do not know φ because the 'real world' system dynamics are assumed unknown. The Radon-Nikodym derivative $dP(\hat{\Theta})/dP$ on Ω induces, abusing notation slightly, a Radon-Nikodym derivative $dP(\hat{\Theta})/dP$ on the state space of Y_k. If $\tilde{\varphi}$ is the unconditional density of Y_k under $P(\Theta)$, then

$$\tilde{\varphi}(u) = \frac{dP(\Theta)}{dP}\varphi(u).$$

Hence,

$$\begin{aligned} E[\tilde{\varphi}(Y_k) \mid Y_k] &= E\left[\frac{dP(\Theta)}{dP}\varphi(Y_k) \,\middle|\, Y_k\right] \\ &= \varphi(Y_k)E\left[\frac{dP(\Theta)}{dP} \,\middle|\, Y_k\right] \\ &= \varphi(Y_k)\mathcal{L}_k(\Theta). \end{aligned}$$

For a given value of Y_k, the expression $\varphi(Y_k)$ is unknown, but fixed. Hence the likelihood (conditional density, actually) of the observations can be maximized by choosing a value of $\hat{\Theta}$ which maximizes $\mathcal{L}_k(\hat{\Theta})$, without knowing φ and without knowing the real world dynamics .

This sounds too good to be true—and it is. The expression for $\mathcal{L}$ always involves the real world parameters, which are thought of as being unknown. An ingenious solution to this problem was put forward in (Baum, Petrie, Soules, and Weiss, [5]), known as the Expectation Maximization (EM) algorithm, which has since obtained wide use (Lange, [21]). Fix k and let $\hat{\Theta}, \hat{\hat{\Theta}}$ be two hypothetical parameter sets. Let $\hat{E}$ denote the expectation under $P(\hat{\hat{\Theta}})$, and define

$$Q(\hat{\Theta}, \hat{\hat{\Theta}}) := \hat{E}\left[\log \frac{dP(\hat{\Theta})}{dP(\hat{\hat{\Theta}})} \,\middle|\, Y_k\right].$$

Then Jensen's Inequality and Conditional Bayes' Theorem (1.4) give us

$$
\begin{aligned}
Q(\hat{\Theta}, \tilde{\Theta}) &= \tilde{E}\left[\log \frac{dP(\hat{\Theta})}{dP(\tilde{\Theta})} \,\middle|\, Y_k\right] && (1.1.5) \\
&\leq \log \tilde{E}\left[\frac{dP(\hat{\Theta})}{dP(\tilde{\Theta})} \,\middle|\, Y_k\right] \\
&= \log \left(\frac{E\left[\frac{dP(\tilde{\Theta})}{dP} \frac{dP(\hat{\Theta})}{dP(\tilde{\Theta})} \,\middle|\, Y_k\right]}{E[\left[\frac{dP(\tilde{\Theta})}{dP} \,\middle|\, Y_k\right]} \right) \\
&= \log \mathcal{L}_k(\hat{\Theta}) - \log \mathcal{L}_k(\tilde{\Theta}).
\end{aligned}
$$

Now, given a hypothetical parameter set $\tilde{\Theta}$, we can reasonably hope to compute $Q(\hat{\Theta}, \tilde{\Theta})$ for any alternate parameter set $\hat{\Theta}$—since $Q(\hat{\Theta}, \tilde{\Theta})$ does not depend on unknown parameters. If we choose a value of $\hat{\Theta}$ which *maximizes* $Q(\cdot, \tilde{\Theta})$, then (1.1.5) guarantees that

$$\log \mathcal{L}_k(\hat{\Theta}) - \log \mathcal{L}_k(\tilde{\Theta}) \geq Q(\hat{\Theta}, \tilde{\Theta}) \geq Q(\tilde{\Theta}, \tilde{\Theta}) = 0.$$

The upshot of this is that, given a value of $\tilde{\Theta}$, we can find another one which is *at least as good as, and usually better than* $\tilde{\Theta}$. Iterating this procedure gives a non-decreasing sequence of values for $\mathcal{L}(\hat{\Theta}^i)$, with equality (and, hence, convergence) iff $\hat{\Theta}^i = \hat{\Theta}^{i-1}$. Sufficient conditions for convergence of the sequence $\hat{\Theta}^i$ are given in (Dempster, *et al*, [11]), where it is pointed out that "In almost all applications, the limiting of Θ^* ... will occur at a local, if not global, maximum of $\mathcal{L}(\Theta)$." Often, as in (Elliott and Krishnamurthy [14], [15]), several initial values of $\hat{\Theta}$ are tested to improve the chances of convergence to a global maximum of $\mathcal{L}$.

Each iteration of the EM algorithm turns out to involve re-estimating the sum (or integral) over time of second moment of the signal process. Since this estimation must be carried out multiple times between successive observations, its speed is of upmost importance. Until recently, the estimation was done by 'smoothing,' that is, estimating the second moment of the signal at each time point of the past, and then adding (or integrating) over the time variable to achieve an estimate for the sum. In 1996, Elliott and Krishnamurthy ([14], [15]) devised finite dimensional exact filters for the sum (or integral) itself, which do not involve estimating past values of the signal. Since so few finite dimensional exact filters are known, this discovery was of great theoretical importance. It was also of practical importance since the filter-based EM algorithm requires significantly less memory and is at least twice as fast as the standard smoother-based methods. In this dissertation, we extend the results of Elliott and Krishnamurthy to the mixed-time case (Chapter III) and the mixed-time case with switching (Chapter V).

Let us keep in mind that there is a fundamental theoretical difference between filtering (*e.g.* the Kalman Filter) and parameter estimation (*e.g.* EM algorithms), even though both give estimates based on the same observations. *Filtering* algorithms give us the conditional densities of random variables, while *parameter estimation* schemes give estimates for quantities which are fixed, but unknown. In terms of the phenomenon being modeled, the line between 'random' and 'unknown' may blur at times; but in terms of our mathematical model, we make a clear distinction.

1.2. Preliminaries

Before summarizing the work done in this dissertation, (see Section 1.3), we shall recall some theorems and definitions which serve as background material.

From the above discussions, we already see that the conditional Bayes' Theorem plays a crucial role in our change of measure approach. A version of this result is as follows, (a proof can be found in [12]):

Theorem 1.4. CONDITIONAL BAYES' THEOREM. *Suppose $(\Omega, \mathcal{F}, P)$ is a probability space and $\mathcal{G} \subset \mathcal{F}$ is a sub-sigma field. Suppose $\overline{P}$ is another probability measure absolutely continuous with respect to P and with Radon-Nikodym deriva-*

tive

$$\frac{d\overline{P}}{dP} = \Lambda.$$

Then if Z is any $\mathcal{F}$-measurable random variable, we have

$$\overline{E}[Z \mid \mathcal{G}] = \frac{E[\Lambda Z \mid \mathcal{G}]}{E[\Lambda \mid \mathcal{G}]}.$$

The following theorem says, loosely speaking, that given a collection of 'consistent' finite-dimensional distributions, there exists a stochastic process corresponding to them. It is used in this work to construct probability measures by constructing their Radon-Nikodym derivatives at each point in time.

Theorem 1.5. Kolmogorov Extension Theorem, [12]. *For all $t_1, \dots, t_k, k \in \mathbb{N}$ and t_i in an index set $\mathcal{T}$, let $P_{t_1,\dots,t_k}$ be the probability measures on $\mathbb{R}^{nk}$ such that*

$$P_{t_{\sigma(1)},\dots,t_{\sigma(k)}}(F_1 \times \cdots \times F_k) = P_{t_1,\dots,t_k}\left(F_{\sigma^{-1}(1)} \times \cdots \times F_{\sigma^{-1}(k)}\right)$$

for all permutations σ on $\{1, 2, \dots, k\}$ and

$$\begin{aligned} &P_{t_1,\dots,t_k}(F_1 \times \cdots \times F_k) \\ &\qquad = P_{t_1,\dots,t_k,t_{k+1},\dots,t_{k+m}}(F_1 \times \cdots \times F_k \times \mathbb{R}^n \times \cdots \times \mathbb{R}^n) \end{aligned}$$

for all $m \in \mathbb{N}$, where the set on the right-hand side has a total of $k+m$ factors. Then there exists a probability space $(\Omega, \mathcal{F}, P)$ and a stochastic process X_t on Ω into $\mathbb{R}^n$ such that

$$P_{t_1,\dots,t_k}(F_1 \times \cdots \times F_k) = P[X_{t_1} \in F_1, \dots, X_{t_k} \in F_k],$$

for all t_i in $\mathcal{T}$ and all Borel sets F_i.

After deriving filters for a mixed-time model, it is natural to ask whether they tend to the filters for the corresponding continuous-time model as the duration between successive observations approaches 0. We will examine this asymptotic 'tendency' in terms the following type of convergence:

Definition 1.6. *Let $\{X_k(\cdot), k \geq 1\}$ be $\mathbb{R}^d$-valued stochastic processes on a time domain $S \subset \mathbb{R}$. We say that $X_k(\cdot)$ converges to a process $X(\cdot)$* **uniformly on compacts in probability** *(abbreviated as convergence in ucp) provided for every compact set $T \subset S$ and every $\epsilon > 0$, we have*

$$P\left[\left(\sup_{t\in T} \|X_k(t) - X(t)\|\right) > \epsilon\right] \to 0$$

as $k \to \infty$.

Switching parameters (as discussed in Section 1.1) are often modeled by finite state Markov chains, defined as follows:

Definition 1.7. *A stochastic sequence Z_k taking values in a countable set $\mathcal{S}$ is said to be a discrete time* **Markov chain** *if for each k, the sigma algebras $\sigma\{Z_j, j \leq k\}$ and $\sigma\{Z_j, j \geq k\}$ are conditionally independent given Z_k.*

Definition 1.8. *A discrete-time Markov chain Z_k taking values in a set $\{1, \ldots, N\}$, where N is possibly infinite, is said to be* **stationary** *provided there exists a matrix $P_{ij}, 1 \leq i, j \leq N$, such that for all $k \in \mathbb{N}$ we have*

$$P[Z_{k+1} = j \mid Z_k = i] = P_{ij}.$$

The matrix P_{ij} of Definition 1.8 is known as the **transition matrix** for the Markov chain Z. For details on Markov chains we refer to Kannan [18].

As pointed out earlier, our approach to mixed-time filtering involves first discretizing the signal process. The following theorem is useful for constructing 'discretized' versions of continuous-time signals that are given by diffusions:

Theorem 1.9. *Suppose W_t is a standard Brownian motion in $\mathbb{R}^m$, and B_t, Q_t are $m \times m$ matrix valued functions on $[0, T]$ such that the solution x_t of the following Itô equation exists uniquely on $[0, T]$:*

$$dx_t = B_t x_t dt + Q_t dW_t, \quad x_0 = 0_{1\times m}.$$

Then there exists a collection of matrices $\{\Phi(s,t), 0 \leq s < t \leq T\}$ *such that*

$$x_t = \Phi(t,s)x_s + \int_s^t \Phi(t,\tau)Q_\tau dW_\tau$$

for all $0 < s < t < T$.

The matrix Φ of Theorem 1.9 is known as the 'fundamental matrix' associated with B_t (for details, see Arnold [2]).

1.3 Summary of Contents

We shall now present a summary of the work done in this dissertaton. Chapter II is devoted to mixed-time filtering for models with linear Gaussian dynamics—a sort of hybrid of the models of Theorems 1.1 and 1.2. Only the presentation and methods of proof are new in this chapter. The main results are either known or are modificiations or simple extensions of known results (especially see [12]). The purpose of the chapter is to introduce the techniques of the dissertation in a relatively familiar context, and to demonstrate the relation of these techniques to well known and widely used mathematics. We first derive, using change of measure techniques, an exact (infinite dimensional) filter for the conditional density of the signal state. From this, we then obtain finite dimensional filters for the first and second moments of the signal state. Our finite dimensional filters turn out to be, at the discrete observation times, a special case of the discrete-time Kalman filter (Theorem 1.1). In the final section of the chapter, we examine the behavior of our mixed time filters as the time δ between observations goes to 0. In the limiting case, we recover the continuous time Kalman filter (Theorem 1.2). Though not surprising in the case of Chapter II, this convergence to the continuous-time filter is not automatic (see Chapter III, Section 2).

We continue in Chapter III the study of the mixed-time linear Gaussian model. We consider there the problem of estimating the model parameters in the case where they are unknown, or known poorly. This chapter extends the work of

Elliott and Krishnamurthy ([14], [15]) from the continuous-time and discrete-time models to the mixed-time case. The approach is Maximum Likelihood estimation via an EM algorithm. As pointed out in Section 1.1, a primary concern in the EM method is the calculation of estimates for the sum over time of certain functions of the signal. It turns out that the discrete-time filters of [15], when applied to models arising from mixed-time problems, 'blow up' as the time δ between observations approaches 0. We alter the techniques of [15] slightly to obtain filters which stabilize as $\delta \downarrow 0$. Being finite dimensional and exact, these filters are of theoretical interest unto themselves, apart from their use in EM parameter estimation. Also, the filters can be derived for models more general than those in which they are practical for EM estimation. For these two reasons, we give the filters in Section 3.1. In Section 3.2, we examine the behavior of our filters as the time δ between observations goes to 0. In some cases, these stabilize to limits *other* than those given in [14] for the continuous-time case. This suggests that Elliott and Krishnamurthy's continuous-time techniques can not be used with impunity in mixed-time applications, even when the times between observations are 'small.' Section 3.3 givess the EM algorithm itself for the mixed-time case.

In chapter IV, we consider a mixed time model with switching, where the switching parameter θ is taken as a discrete-time Markov chain. We derive a recursive exact filter for the joint conditional density of the parameter $\theta_{k\delta}$ and the signal x_t. The results of this chapter are similar to those discovered independently by Elliott, Dufours, and Sworder ([13], 1996) in the discrete-time case. The obtained filter, though exact, is infinite dimensional. This is not a severe shortfall in light of Bjork's result ([6]) that, for the model we consider, finite dimensional exact filters exist only in the special case where observation equation does not explicitly involve the signal.

Recently, Fleming and McEnneny have obtained new results in filtering by going somewhat against the grain of tradition, modeling the noise terms in the signal and observation processes as deterministic but unknown functions of time. Motivated by this idea, and by the difficulty in finding finite dimensional exact filters for switched models , we consider in Chapter V a mixed time switched model in which the switching parameter θ is a fixed (non-random) but unknown function of time. Probabilistically speaking, this places us in the domain of Chapter II, and so we obtain finite-dimensional exact filters—probabilistically speaking. In practice, our filter is only as good as our estimates of the model parameters. We apply the methods of chapter III to obtain these estimates, designing an EM parameter estimation scheme based on finite dimensional exact filters. Recall from the discussion of switching (in Section 1.1) that the switching parameter θ is a finite state process. Hence estimating θ requires techniques slightly different from those of the Chapter III, where the parameter state space is continuous. In particular, the Q function is be maximized by 'brute force,' *i.e.*, by substituting all possible values of θ, rather than by a calculus-type maximization scheme. Another upshot of the finite parameter-state space is that our EM algorithm always converges in a finite number of steps.

Chapter II

Filters for the State in the Linear Model

2.0. Model

In this chapter we consider a mixed-time filtering problem in which the system dynamics are linear. Our goal is to derive expressions for the conditional mean and covariance of the state at any time t given the history of observations, and examine the behavior of these filters as the time δ between observations approaches 0. Our approach is to first derive the filters only for the discrete observation times, and then extend them to continuous time via the Kolmogorov Forward Equation.

We begin with the model itself. Fix $\delta > 0$. Our $\mathbb{R}^m$-valued signal process x_t and $\mathbb{R}^n$-valued observation process $\{z^\delta_k, k\delta \leq T\}$ are defined on $[0, T] \subset \mathbb{R}$ by

$$dx_t = B_t x_t dt + Q_t dw_t, \ x_0 = 0; \tag{2.0.1}$$

$$z^\delta_k = \delta H_{k\delta} x^\delta_k + \Delta^\delta_k. \tag{2.0.2}$$

Equation (2.0.2) gives an observation process intended as a discrete-time analog of the continuous time observation process

$$dy_t = H_t x_t dt + R_t dV_t. \tag{2.0.3}$$

Indeed, as $\delta \downarrow 0$, the stochastic process $y^\delta_t := \sum_{k\delta \leq t} z^\delta_k$ converges to the process $\{y_t\}$ uniformly in probability on $[0, T]$.

We make the following assumptions with regards to equations (2.0.1),(2.0.2),

where $A^\delta_{k+1} = \Phi(k\delta + \delta, k\delta)$. It is this discretized model we will be concerned with initially. After obtaining filters for the discretized System (2.0.6)-(2.0.7), we will discuss the extension to the mixed-time filter and the limiting behavior thereof.

2.1. Exact Filter

In this section, we employ a change of probability measure to obtain a recursive exact filter for the conditional density of x_k given $\mathcal{Z}_k$. For notational convenience, we supress all superscripts δ until further notice. Also, we will write H_k for $H_{k\delta}$.

Write $\mathcal{G}_k$ for the complete σ-field generated by $\{x_0, \ldots, x_k, z_0, \ldots, z_{k-1}\}$ and $\mathcal{Z}_k$ for the complete σ-field generated by $\{z_0, \ldots, z_k\}$. Define

$$\lambda_\ell := \frac{\varphi_\ell(z_\ell)}{\varphi_\ell(\Delta_\ell)}, \quad \ell\delta \le T,$$

$$\Lambda_k := \prod_{\ell=1}^{k} \lambda_\ell, \quad k\delta \le T.$$

For each $k \ge 1$, define a probability measure $\overline{P_k}$ on $(\Omega, \mathcal{G}_k)$ by

$$\frac{d\overline{P}}{dP} = \Lambda_k,$$

i.e, for any event $G \in \mathcal{G}_k$, we have

$$\overline{P}_k(G) = \int_G \Lambda_k \, dP = \int_G \frac{d\overline{P}}{dP} \, dP = \int_G d\overline{P}.$$

By Kolmogorov's Extension Theorem (1.5), we can now define a probability measure $\overline{P}$ on $(\Omega, \mathcal{G}_\infty)$ such that $\overline{P}$ agrees with $\overline{P}_k$ on each respective $\mathcal{G}_k$.

The following theorem tells us *a fortiori* that the under $\overline{P}$, the sequences $\{x_k\}$ and $\{z_k\}$ are independent.

Theorem 2.2. *Under $\overline{P}$, z_k is independent of $\mathcal{G}_k$ and has density φ_k, $k \le T/\delta$.*

Proof. For $t \in \mathbb{R}^n$, by the event $\{z_k < t\}$ we mean $\{z_k^i < t^i, i = 1, \ldots, n\}$. Then

$$\begin{aligned}
\overline{P}(z_k < t \,|\mathcal{G}_k) &= \overline{E}[I\{z_k < t\}|\mathcal{G}_k] \\
&= \frac{E[\Lambda_k I\{z_k < t\}|\mathcal{G}_k]}{E[\Lambda_k|\mathcal{G}_k]}, \quad \text{by abstract Bayes' Theorem,} \\
&= \frac{\Lambda_{k-1} E[\lambda_k I\{z_k < t\}|\mathcal{G}_k]}{\Lambda_{k-1} E[\lambda_k|\mathcal{G}_k]}. \qquad (2.1.1)
\end{aligned}$$

Now the disctetized version of (2.0.1)-(2.0.2) takes the recursive form

$$x_{k+1}^\delta = A_{k+1}^\delta x_k^\delta + v_{k+1}^\delta, \quad x_0 = 0; \qquad (2.0.6)$$

$$z_k^\delta = \delta H_{k\delta} x_k^\delta + \Delta_k^\delta. \qquad (2.0.7)$$

Now,

$$E[\lambda_k|\mathcal{G}_k] = \int_{\mathbb{R}^{m+n}} \frac{\varphi_k(\delta H_k x_k + \Delta_k)}{\varphi_k(\Delta_k)} \rho(x_k, \Delta_k)\, dx_k\, d\Delta_k$$

where $\rho(\cdot,\cdot)$ is the conditional joint density of x_k, Δ_k given $\mathcal{G}_k$ (under $\overline{P}$). But x_k is $\mathcal{G}_k$ measurable and Δ_k is independent of $\mathcal{G}_k$. Hence

$$\begin{aligned}
E[\lambda_k|\mathcal{G}_k] &= \int_{\mathbb{R}^n} \frac{\varphi_k(\delta H_k x_k + \Delta_k)}{\varphi_k(\Delta_k)} \varphi_k(\Delta_k)\, d\Delta_k \\
&= \int_{\mathbb{R}^n} \varphi(\delta H_k x_k + \Delta_k) d\, \Delta_k \\
&= 1.
\end{aligned}$$

Hence, the denominator of (2.1.1) is equal to Λ_{k-1}. Hence,

$$\begin{aligned}
\overline{P}(z_k < t\,|\mathcal{G}_k) &= \int_{\mathbb{R}^n} \frac{\varphi_k(\delta H_k x_k + \Delta_k)}{\varphi_k(\Delta_k)} I\{\delta H_k x_k + \Delta_k < t\} \varphi(\Delta_k) d\, \Delta_k \\
&= \int_{\mathbb{R}_n} \varphi_k(u) I\{u < t\}\, du \\
&= \int_{u<t} \varphi_k(u)\, du.
\end{aligned}$$

Since this quantity is independent of $\mathcal{G}_k$, the proof is done. ∎

Let $\overline{\Lambda}_k := 1/\Lambda_k$ and $\alpha_k(x) := \overline{E}[\overline{\Lambda}_k I\{x_k \in dx\}|\mathcal{Z}_k]$. Then $\alpha_k(x)$ is the 'unnormalized' density (*unnormalized* insofar as $\int \alpha_k(x)dx \neq 1$) on $\mathbb{R}^m$ such that for every Borel measurable function $f : \mathbb{R}^m \longrightarrow \mathbb{R}$, we have

$$\int_{\mathbb{R}^m} f(x)\alpha_{k+1}(x)\, dx = \overline{E}[f(x_{k+1})\overline{\Lambda}_{k+1}|\mathcal{Z}_{k+1}], \quad k \geq 1. \tag{2.1.2}$$

Theorem 2.3. *For $k \geq 1$, the following recursion holds for the unnormalized density α_k.*

$$\alpha_{k+1}(x) = \frac{\varphi_{k+1}(z_{k+1} - H_{k+1}x)}{\varphi_{k+1}(z_{k+1})} \int_{\mathbb{R}^m} \psi_{k+1}(x - A_{k+1}u)\alpha_k(u)\, du. \tag{2.1.3}$$

Proof. Let $f : \mathbb{R}^m \longrightarrow \mathbb{R}$ be an arbitrary Borel function. Then

$$\begin{aligned}
&\int_{\mathbb{R}^m} f(x)\alpha_{k+1}(x)\, dx \qquad (2.1.4)\\
&= \overline{E}[f(x_{k+1})\overline{\Lambda}_{k+1}|\mathcal{Z}_{k+1}] \\
&= \overline{E}\left[\overline{\Lambda}_k f(A_{k+1}x_k + v_{k+1}) \frac{\varphi_{k+1}(z_{k+1} - \delta H_{k+1}(A_{k+1}x_k + v_{k+1}))}{\varphi_{k+1}(z_{k+1})} \Big| \mathcal{Z}_{k+1}\right].
\end{aligned}$$

Let $dz := dz_1 dz_2 \cdots dz_k$ and $d\Delta := d\Delta_1 d\Delta_2 \cdots d\Delta_k$. Then the above expectation can be written as an integral

$$\int_{\mathbb{R}^{2kn+2m}} \overline{\Lambda}_k f(A_{k+1}x_k + v_{k+1}) \frac{\varphi_{k+1}[z_{k+1} - \delta H_{k+1}(A_{k+1}x_k + v_{k+1})]}{\varphi_{k+1}(z_{k+1})} \cdot \rho_{k+1}(z_1, \ldots, z_k, \Delta_1, \ldots, \Delta_k, x_k, v_{k+1}) dz d\Delta dx_k dv_{k+1},$$

where ρ_{k+1} is the appropriate joint conditional density given $\mathcal{Z}_{k+1}$ under $\overline{P}$. But under $\overline{P}$, we have by Theorem 2.2 that z_{k+1} is independent of all the arguments of ρ_{k+1} jointly; and so ρ_{k+1} may be replaced by the joint conditional density ρ_k of the arguments given $\mathcal{Z}_k$. Moreover, v_{k+1} is independent of all the other arguments of ρ_k jointly, and so we may factor out its density ψ_{k+1} from ρ_k. Writing $\tilde{\rho}_k$ for the conditional joint density given $\mathcal{Z}_k$ of the remaining arguments, (2.1.4) may hence be written

$$\begin{aligned}
&\int_{\mathbb{R}^{2kn+2m}} \overline{\Lambda}_k f(A_{k+1}x_k + v_{k+1}) \frac{\varphi_{k+1}[z_{k+1} - \delta H_{k+1}(A_{k+1}x_k + v_{k+1})]}{\varphi_{k+1}(z_{k+1})} \\
&\qquad \cdot \tilde{\rho}_k(z_1, \ldots, z_k, \Delta_1, \ldots, \Delta_k, x_k,) \psi_{k+1}(v_{k+1}) \, dz \, d\Delta \, dx_k \, dv_{k+1} \\
= \; & \overline{E}\Big[\int_{\mathbb{R}^m} \overline{\Lambda}_k f(A_{k+1}x_k + v_{k+1}) \frac{\varphi_{k+1}[z_{k+1} - \delta H_{k+1}(A_{k+1}x_k + v_{k+1})]}{\varphi_{k+1}(z_{k+1})} \\
&\qquad \times \psi_{k+1}(v_{k+1}) dv_{k+1} | \mathcal{Z}_k\Big].
\end{aligned}$$

By (2.1.2), this equals

$$\int_{\mathbb{R}^{2m}} f(A_{k+1}x_k + v_{k+1}) \frac{\varphi_{k+1}[z_{k+1} - \delta H_{k+1}(A_{k+1}x_k + v_{k+1})]}{\varphi_{k+1}(z_{k+1})} \times \psi_{k+1}(v_{k+1}) \alpha_k(x_k) \, dv_{k+1} \, dx_k.$$

Now, let $x := A_{k+1}x_k + v_{k+1}$ and $u := x_k$. Then the above equals

$$\int_{\mathbb{R}^{2m}} f(x) \frac{\varphi_{k+1}(z_{k+1} - \delta H_{k+1}x)}{\varphi_{k+1}(z_{k+1})} \alpha_k(u) \psi_{k+1}(x - A_{k+1}u) \, du \, dx$$

By setting the final integral equal to (2.1.4), the theorem follows since f was arbitrary. ∎

computing the density

Let $p_k(x)$ be the conditional density of x_k given $\mathcal{Z}_k$. Then by abstract Bayes' Theorem (1.4), we have

$$\begin{aligned} p_k(x) &= E[I\{x_k \in dx\}|\mathcal{Z}_k] \\ &= \frac{\overline{E}_k[\overline{\Lambda}_k I\{x_k \in dx\}|\mathcal{Z}_k]}{\overline{E}_k[\overline{\Lambda}_k\}|\mathcal{Z}_k]} \\ &= \frac{\alpha_k(x)}{\int_{\mathbb{R}^m} \alpha_k(z)\, dz} \end{aligned} \tag{2.1.5}$$

Hence, the conditional density of $x_{k\delta}$ given $\mathcal{Z}_k$ may be computed using Theorem 2.3 and (2.1.5).

Let $p_t(x)$ denote the conditional density of x_t given the observations up to time t. Between observations, $p_t(x)$ evolves in accordance with Kolmogorov's forward equation. Hence, $p_t(x)$ satisfies

$$\partial p/\partial t = -p' \mathrm{Tr}(B_t) - p_x'(B_t)x + \frac{1}{2}\mathrm{Tr}(Q_t^2 p_{xx}) \tag{2.1.6}$$

on any interval $[k\delta, k\delta + \delta)$, where p_x denotes the gradient of p with respect to the vector x ([17], p.195). The initial condition for a given interval is obtained via Theorem 2.3.

2.2. Finite Dimensional Exact Filter

The exact filter of Theorem 2.3 is infinite dimensional. Our next goal is to use Theorem 2.3 to derive finite dimensional filter equations for mean and variance of the state in the discrete time System (2.0.6)-(2.0.7). This is possible because the integral in (2.1.3) can be viewed as the expectation of a function of a Gaussian random variable and hence evaluated explicitly. We first recall a computational lemma, the proof of which is straightforward and can be found in ([1], p 138).

Lemma 2.4 **(Matrix Inversion Lemma).** *In terms of a $p \times q$ matrix Σ, a $q \times q$ matrix R, and a $p \times q$ matrix H, the following equalities hold on the assumption* *that the various inverses exist:*

$$\begin{aligned}(I + \Sigma H R^{-1} H')\Sigma^{-1} &= (\Sigma^{-1} + H R^{-1} H')^{-1} \\ &= \Sigma - \Sigma H (H' \Sigma H + R)^{-1} H' \Sigma\end{aligned}$$

and

$$\begin{aligned}(I + \Sigma H R^{-1} H')^{-1} \Sigma H R^{-1} &= (\Sigma^{-1} + H R^{-1} H')^{-1} H R^{-1} \\ &= \Sigma H (H' \Sigma H + R)^{-1}.\end{aligned}$$

Theorem 2.5. *Let x_k and z_k, $k\delta \le T$, be as given in model (2.0.6)-(2.0.7) under Assumptions(2.1). Define*

$$\begin{aligned}\hat{x}_k &:= E[x_k | Z_k], \\ \bar{x}_k &:= E[x_k | Z_{k-1}], \\ \hat{P}_k &:= cov[x_k | Z_k], \\ \bar{P}_k &:= cov[x_k | Z_{k-1}].\end{aligned}$$

Then the a posteriori and a priori estimates $\hat{x}_k, \bar{x}_k$, $\hat{P}_k$, and $\bar{P}_k$ satisfy the following filter equations:

$$\bar{x}_{k+1} = A_{k+1} \hat{x}_k \tag{2.2.1,}$$

$$\bar{P}_{k+1} = A_{k+1} \hat{P}_k A'_{k+1} + Q_{k+1}, \tag{2.2.2}$$

$$\begin{aligned}\hat{P}_{k+1} = \bar{P}_{k+1} - \bar{P}_{k+1} \delta H_{k+1} (\delta H_{k\delta+\delta} \bar{P}^{\delta}_{k+1} \delta H'_{k+1} + R_{k+1})^{-1} \\ \times \delta H_{k+1} \bar{P}_{k+1},\end{aligned} \tag{2.2.3}$$

$$\hat{x}_{k+1} = K_{k+1} z_{k+1} + (I - K_{k+1} \delta H_{k+1}) \bar{x}_{k+1}, \tag{2.2.4}$$

where K_k is defined by

$$K_k := \hat{P}_k \delta H'_k R_k^{-1}. \tag{2.2.5}$$

Proof. Recall that $\varphi_k \sim N(0, Q_k)$ and $\psi_k \sim N(0, R_k)$. The linearity of (2.0.6), (2.0.7) imply that the conditional distribution of x_k given $\mathcal{Z}_k$ is also normal. Then using the formula for normal density, Theorem 2.3 yields the following expression:

$$\alpha_{k+1} = K_{k+1}(x) \int_{\mathbb{R}^m} \exp\Big\{-\frac{1}{2}\big[(x - A_{k+1}u)' Q_{k+1}^{-1} (x - A_{k+1}u) \\ + (u - \hat{x}_k)' \hat{P}_k^{-1} (u - \hat{x}_k)\big]\Big\} du,$$

where

$$K_{k+1}(x) = \frac{\varphi_{k+1}(z_{k+1} - \delta H_{k+1} x)}{\varphi_{k+1}(z_{k+1})} (2\pi)^{-m/2} |Q_{k+1}|^{-1/2} |\hat{P}_k|^{-1/2}.$$

Carrying out the multiplication in the argument of the exponential, we have

$$\alpha_{k+1}(x) = K_{k+1}(x) \int_{\mathbb{R}^m} \exp\Big\{ -\frac{1}{2}\big[x' Q_{k+1}^{-1} x - x' Q_{k+1}^{-1} A_{k+1} u \\ + u' A_{k+1}' Q_{k+1}^{-1} x + u' A_{k+1}' Q_{k+1}^{-1} A_{k+1} u + u' \hat{P}_k^{-1} u \\ - u' \hat{P}_k^{-1} \hat{x}_k - \hat{x}_k' \hat{P}_k^{-1} u + \hat{x}_k' \hat{P}_k^{-1} \hat{x}_k\big]\Big\} du.$$

Since $x' Q_{k+1}^{-1} A_{k+1} u$ and $\hat{x}_k' \hat{P}_k^{-1} u$ are scalars, they are equal to their own transposes. Keeping only terms involving u under the integral, we then have

$$\alpha_{k+1}(x) = K_{k+1}(x) \exp\Big\{-\frac{1}{2}\big[x' Q_{k+1}^{-1} x + \hat{x}_k' \hat{P}_k^{-1} \hat{x}_k\big]\Big\} \\ \times \int_{\mathbb{R}^m} \exp\Big\{-\frac{1}{2}\big[u'(A_{k+1}' Q_{k+1}^{-1} A_{k+1} + \hat{P}_k^{-1}) u \\ - (2x' Q_{k+1}^{-1} A_{k+1} + 2\hat{x}_k' \hat{P}_k^{-1}) u\big]\Big\} du.$$

Define

$$L_{k+1}(x) = L_{k+1} := \exp\left[-\frac{1}{2}\left(x' Q_{k+1}^{-1} x + \hat{x}_k' \hat{P}_k^{-1} \hat{x}_k\right)\right],$$

$$M_{k+1} := A_{k+1}' Q_{k+1}^{-1} A_{k+1} + \hat{P}_k^{-1},$$

$$N_{k+1}'(x) = N_{k+1}' := 2(x' Q_{k+1}^{-1} A_{k+1} + \hat{x}_k' \hat{P}_k^{-1}).$$

For notational convenience, we will temporarily suspend all subscripts $k+1$. We then have from above

$$\begin{aligned}
\alpha(x) &= K(x)\cdot L\cdot \int_{\mathbb{R}^m} \exp\left\{-\frac{1}{2}(u'Mu - N'u)\right\}\, du \\
&= K(x)L\exp\left\{-\frac{1}{2}\left(-\frac{1}{4}N'M^{-1}N\right)\right\} \\
&\quad\times \int_{\mathbb{R}^m} \exp\left\{-\frac{1}{2}\left[u'Mu + \frac{1}{4}N'M^{-1}N - \frac{1}{2}u'N \ -\frac{1}{2}N'u\right]\right\}\, du \\
&= K(x)L\exp\left\{-\frac{1}{2}\left(-\frac{1}{4}N'M^{-1}N\right)\right\} \\
&\quad\times \int_{\mathbb{R}^m} \exp\left\{-\frac{1}{2}\left[u'Mu + \frac{(M^{-1}N)'}{2}M\frac{(M^{-1}N)}{2}\right.\right. \\
&\qquad\left.\left. -\left(\frac{1}{2}\right)u'MM^{-1}N - \frac{(M^{-1}N)'}{2}Mu\right]\right\}\, du \\
&= K(x)L\exp\left\{-\frac{1}{2}\left(-\frac{1}{4}N'M^{-1}N\right)\right\} \\
&\quad\times \int_{\mathbb{R}^m} \exp\left\{-\frac{1}{2}\left[\left(u-\frac{1}{2}M^{-1}N\right)' M\left(u-\frac{1}{2}M^{-1}N\right)\right]\right\}\, du \\
&= K(x)L\exp\left\{-\frac{1}{2}\left[-\frac{1}{4}N'M^{-1}N\right]\right\}|M|^{1/2}(2\pi)^{-m/2}.
\end{aligned}$$

Recalling the expressions for L and K, this becomes

$$\begin{aligned}
\alpha(x) &= \frac{\varphi(z-\delta Hx)}{\varphi(z)}(2\pi)^{-m}|Q|^{-1/2}|\hat{P}_k|^{-1/2}|M|^{-1/2} \\
&\quad\times\exp\{-\frac{1}{2}[x'Q^{-1}x + \hat{x}_k'\hat{P}_k\hat{x}_k - \frac{1}{4}N'M^{-1}N]\}.
\end{aligned} \tag{2.2.6}$$

Now, using the fact that $\varphi(=\varphi_{k+1})$ is normally distributed with mean zero and variance R_{k+1} (for which we will write R), we can write

$$\begin{aligned}
&\varphi(z-\delta Hx) \\
&= (2\pi)^{-\frac{m}{2}}|R|^{-1/2}\left\{-\frac{1}{2}[z-\delta Hx]'R^{-1}[z-\delta Hx]\right\} \\
&= (2\pi)^{-\frac{m}{2}}|R|^{-1/2}\exp\left\{-\frac{1}{2}[z'R^{-1}z - 2(\delta Hx)'R^{-1}z + x'H'\delta R^{-1}\delta Hx]\right\}.
\end{aligned}$$

So substituting this and the expression for N in (2.2.6), we get

$$\alpha_{k+1} = \frac{1}{\varphi(z)}(2\pi)^{-\frac{3m}{2}}|Q|^{-\frac{1}{2}}|\hat{P}|^{-\frac{1}{2}}||M|^{-\frac{1}{2}}|R|^{-\frac{1}{2}} \tag{2.2.7}$$
$$\times \exp\Big\{-\frac{1}{2}\Big[z'R^{-1}z - 2x'H'\delta R^{-1}z + x'H'\delta R^{-1}\delta H x$$
$$+ x'Q^{-1}x + \hat{x}'_k\hat{P}_k\hat{x}_k + x'Q^{-1}AM^{-1}AQ^{-1}x$$
$$- x'Q^{-1}AM^{-1}\hat{P}_k^{-1}\hat{x}_k - \hat{x}'_k\hat{P}_k^{-1}M^{-1}A'Q^{-1}x$$
$$- \hat{x}'_k\hat{P}_k^{-1}M^{-1}\hat{P}_k^{-1}\hat{x}_k\Big]\Big\}.$$

But $\alpha_{k+1}(x)$ is a scalar multiple of a normal distribution with mean $\hat{x}_{k+1}$ and variance $\hat{P}_{k+1}$. Hence, there exists a constant C, not depending on x, such that

$$\alpha_{k+1}(x) = C\exp\left\{-\frac{1}{2}\left[(x-\hat{x}_{k+1})'\hat{P}_{k+1}^{-1}(x-\hat{x}_{k+1})\right]\right\} \tag{2.2.8}$$
$$= C\exp\left\{-\frac{1}{2}\left[x'\hat{P}_{k+1}^{-1}x - 2x'\hat{P}_{k+1}^{-1}\hat{x}_{k+1} + \hat{x}'_{k+1}\hat{P}_{k+1}^{-1}\hat{x}_{k+1}\right]\right\}.$$

Setting equal like terms in (2.2.7) and (2.2.8) (in terms of the variable x) yields

$$\hat{P}_{k+1}^{-1} = H'_{k+1}\delta R_{k+1}^{-1}\delta H_{k+1} + Q_{k+1}^{-1} - Q_{k+1}^{-1}A_{k+1}M_{k+1}^{-1}A'_{k+1}Q_{k+1}^{-1}, \tag{2.2.9}$$

$$\hat{P}_{k+1}^{-1}\hat{x}_{k+1} = H'_{k+1}\delta R_{k+1}^{-1}z_{k+1} + Q_{k+1}^{-1}A_{k+1}M_{k+1}^{-1}\hat{P}_k^{-1}\hat{x}_k. \tag{2.2.10}$$

Note that (2.2.9) and (2.2.10) are the information filter version of the Kalman filter for the System (2.0.6)-(2.0.7).

We next use (2.2.9) and (2.2.10) to derive equations similar to the measurement update equations. Recall our definitions

$$\hat{x}_{k+1} := E[x_{k+1}|Z_k]$$
$$\hat{P}_{k+1} := \mathrm{cov}[x_{k+1}|Z_k]$$
$$K_k := \hat{P}_k\delta H'_k R_k^{-1}.$$

Taking the conditional (given Z_k) mean and covariance of both sides of (2.0.6) yields the desired Equations (2.2.1) and (2.2.2):

$$\begin{aligned}\bar{x}_{k+1} &= A_{k+1}\hat{x}_k \\ \bar{P}_{k+1} &= A_{k+1}\hat{P}_k A'_{k+1} + Q_{k+1}.\end{aligned}$$

Then by the Matrix Inversion Lemma and (2.2.2) we have (once again dropping all subscripts $k+1$),

$$\bar{P}^{-1} = Q^{-1} - Q^{-1}AM^{-1}A'Q^{-1}.$$

Combining this with (2.2.9) yields

$$\hat{P}^{-1} = \bar{P}^{-1} + H'\delta R^{-1}\delta H.$$

Taking inverses of both sides of this, again using the Matrix Inversion Lemma, yields

$$\hat{P} = \bar{P} - \bar{P}\delta H'(\delta H\bar{P}\delta H' + R)^{-1}\delta H\bar{P}.$$

Finally, re-introducing the subscripts $k+1$ into the above yields the desired measurement Update Equation (2.2.3) for the covariance.

We now derive the corresponding equation for the mean, again supressing subscripts of $k+1$. Multiplying both sides of (2.2.10) by $\hat{P}$ yields

$$\begin{aligned}\hat{x} &= \hat{P}H'\delta R^{-1}z + \hat{P}Q^{-1}AM^{-1}\hat{P}_k^{-1}\hat{x}_k \\ &= Kz + \hat{P}Q^{-1}AM^{-1}\hat{P}_k\hat{x}_k \\ &= Kz + \hat{P}Q^{-1}A(\hat{P}_k - \hat{P}_kA'\bar{P}^{-1}A\hat{P}_k)\hat{P}_k^{-1}\hat{x}_k, \\ &\qquad\qquad \text{by the Matrix Inversion Lemma,} \\ &= Kz + \hat{P}Q^{-1}\bar{x} - \hat{P}Q^{-1}A\hat{P}_kA'\bar{P}^{-1}\bar{x}, \quad \text{by (2.2.1),} \\ &= Kz + \hat{P}Q^{-1}(I - A\hat{P}_kA'\bar{P}^{-1})\bar{x}\end{aligned}$$

$$= Kz + \hat{P}Q^{-1}(I - \{\hat{P} - Q\}\hat{P}^{-1})\bar{x}), \text{ by (2.2.2)},$$

$$= Kz + \hat{P}Q^{-1}(I - \{I - Q\hat{P}^{-1}\})\bar{x}$$

$$= Kz + \hat{P}Q^{-1}(Q\hat{P}^{-1})\bar{x}$$

$$= Kz + \hat{P}\hat{P}^{-1}\bar{x}$$

$$= Kz + \hat{P}[Q^{-1} - Q^{-1}A(A'Q^{-1}A + \hat{P}_k^{-1})A'Q^{-1}]\bar{x},$$

by the Matrix Inversion Lemma and (2.2.2),

$$= Kz + \hat{P}(Q^{-1} - Q^{-1}AMA'Q^{-1})\bar{x},$$

by the definition of M,

$$= Kz + \hat{P}(\hat{P}^{-1} - \delta H'R^{-1}\delta H)\bar{x}, \text{ by (2.2.9)},$$

$$= Kz + (I - \hat{P}\delta H'R^{-1}\delta H)\bar{x}$$

$$= Kz + (I - K\delta H)\bar{x}$$

Re-introducing the subscript $k+1$ in the above gives the desired Equation (2.2.4). ∎

computing the mean and variance

The conditional mean and covariance of $x_{k\delta}$ given Z_k can be computed recursively using Theorem 2.5. Between observations, (2.1.6) yields ([17], p.196):

$$\frac{d\hat{x}}{dt} = B_t\hat{x}_t,$$
$$\frac{d\hat{P}}{dt} = B_t\hat{P}_t + \hat{P}_tB_t' + Q_t^2.$$

2.3. Limiting Case

We now examing the behavior of the finite dimensional exact filters of Section 2.2 as the time δ between observations goes to 0. Our mixed-time filters are found to converge to the Kalman filter for purely continuous time. Though not surprising here, this convergence to the continuous-time case is not automatic, as we shall see in chapter III.

For each $\delta > 0$, define a step function $\hat{P}^\delta_s$ on $[0, T]$ taking values $\hat{P}^\delta_0, \hat{P}^\delta_1, \dots, \hat{P}_{[T/\delta]}$ on respective intervals $[0, \delta), [\delta, 2\delta), \cdots, [[T/\delta]\delta, T)$, where $[\cdot]$ denotes the greatest integer function. Similarly define, for each $\delta > 0$, *cadlag* step functions R^δ_s, Q^δ_s, $\hat{P}^\delta_s$, H^δ_s, z^δ_s, x^δ_s, K^δ_s associated with respective sequences R^δ_k, Q^δ_k, $\hat{P}^\delta_k$, $H_{k\delta}$, z^δ_k, x^δ_k, K^δ_k.

Theorem 2.6. *Under Assumptions 2.1, the step functions $\hat{P}$ and $\hat{x}$ converge uniformly in probability on $[0, T]$ to the solutions of the continuous time Kalman equations*

$$d\hat{P} = \left[B_t\hat{P}_t + \hat{P}_tB'_t - \hat{P}_tH'_tR_t^{-2}\hat{H}_tP_t + Q_t^2\right]dt, \quad \hat{P}_0 = 0, \tag{2.3.1}$$

$$\hat{x}_t = \int_0^t B_s\hat{x}_s\, ds + \int_0^t K_s\, dy_s - \int_0^t K_sH_s\hat{x}_s\, ds, \tag{2.3.2}$$

where K_s is defined on $[0, T]$ by

$$K_s = \hat{P}_sH'_sR_s^{-2}.$$

Proof. Since H is continuous, we have $H^\delta \to H$ as $\delta \downarrow 0$ uniformly on $[0, T]$. Also,

$$(R^\delta_t) = \int_{[t/\delta]\delta-\delta}^{[t/\delta]\delta} (R^2_s)\, ds.$$

And so $\frac{1}{\delta}R^\delta \to R^2$ as $\delta \downarrow 0$. Further, this convergence is uniform on $[0, T]$, (see Theorem 9.3.1 of [2]). Similarly, $\frac{1}{\delta}Q^\delta \to Q^2$ uniformly on $[0, T]$ as $\delta \downarrow 0$. It follows that $\delta(R^\delta)^{-1} \to R^{-2}$ uniformly on $[0, T]$. In particular, the values of $\delta(R^\delta_k)^{-1}$ are bounded uniformly in δ and k.

Substituting (2.2.3) into (2.2.2) gives

$$\begin{aligned}\hat{P}^\delta_{k+1} = A^\delta_{k+1}\Big[\hat{P}^\delta_k - \hat{P}^\delta_k\delta H'_{k\delta}\Big(&\delta H_{k\delta}\hat{P}^\delta_k\delta H'_{k\delta} \\ &+ R^\delta_k\Big)^{-1}\delta H_{k\delta}\hat{P}^\delta_k\Big](A^\delta_{k+1})' + Q^\delta_{k+1}\end{aligned} \tag{2.3.3}$$

Now, by the Matrix Inversion Lemma, we have

$$(\delta H_{k\delta}\hat{P}_k^\delta \delta H'_{k\delta} + R_k^\delta)^{-1}$$
$$= (R_s^\delta)^{-1} - [\delta(R_k^\delta)^{-1}]\left(\delta H'_{k\delta}[\delta(R_k^\delta)^{-1}]H_{k\delta} + (\hat{P}_k^\delta)^{-1}\right) H_{k\delta}[\delta(R_k^\delta)^{-1}].$$

Since $[\delta(R_k^\delta)^{-1}]$ is uniformly bounded, so is the quantity $[\delta(R_k^\delta)^{-1}]\Big(\delta H'_{k\delta}[\delta(R_k^\delta)^{-1}]$ $H_{k\delta} + (\hat{P}_k^\delta)^{-1}\Big) H_{k\delta}[\delta(R_k^\delta)^{-1}]$. Observing now that, by Taylor's formula,

$$A_{k+1}^\delta = \Phi(k\delta + \delta, k\delta) = I + \delta B_{k\delta} + O(\delta^2), \tag{2.3.4}$$

(2.3.3) can be rewritten as

$$\begin{aligned}\bar{P}_{k+1}^\delta = \bar{P}_k^\delta + \delta\big[B_{k\delta}\bar{P}_k^\delta + \bar{P}_k^\delta B'_{k\delta} \\ - \bar{P}_k^\delta H'_{k\delta}(R_k^{\delta-1}\delta)H_k^\delta \bar{P}_k^\delta + \frac{1}{\delta}Q_{k+1}^\delta + O(\delta)\big].\end{aligned} \tag{2.3.5}$$

Let f^δ be an increasing step function on $[0, T]$ with jumps of size δ at times $0, \delta, 2\delta, \ldots$. Then (2.3.5) may be rewritten as an integral equation

$$\bar{P}_t^\delta = \int_0^t (B_s\bar{P}_s^\delta + \bar{P}_s^\delta B'_s - \bar{P}_s H'_s \delta(R_s^\delta)^{-1} H_s \bar{P}_s^\delta + \frac{1}{\delta}Q_{s+\delta}^\delta + O(\delta))_- \; df_s^\delta.$$

Now, f^δ converges uniformly on $[0, T]$ to the identity function. In light of the convergences already established, it now follows by stability of integral equations (see, for example, Theorem 15 p.209 and comments at bottom of p.201 of [25]) that $\bar{P}^\delta$ converges uniformly on $[0, T]$ to the solution of the Ricatti equation (2.3.1). By 2.3.4, A_k^δ converges uniformly to the identity matrix as $\delta \downarrow 0$. In light of (2.2.2), it follows that $\hat{P}^\delta$ converges uniformly as $\delta \downarrow 0$ to the solution of the Ricatti equation (2.3.1).

As for the mean, substituting (2.2.4) into (2.2.1) yeilds

$$\bar{x}_{k+1}^\delta = A_{k+1}\left[K_k^\delta z_k^\delta + (I - K_k^\delta \delta H_{k\delta})\bar{x}_k^\delta\right].$$

Using 2.3.4, we have

$$\begin{aligned}\bar{x}^{\delta}_{k+1} &= \left(K^{\delta}_k + O(\delta)\right) z^{\delta}_k + (I + \delta B_{k\delta})(I - K^{\delta}_k \delta H_{k\delta})\bar{x}^{\delta}_k + O(\delta^2) \\ &= \bar{x}^{\delta}_k + \delta\left[-K^{\delta}_k H_{k\delta}\bar{x}_k + B_{k\delta}\bar{x}^{\delta}_k + O(\delta)\right] + \left(K^{\delta}_k + O(\delta)\right) z^{\delta}_k.\end{aligned} \tag{2.3.6}$$

It follows from stability results for S.D.E.'s (see, for example, Theorem 15 p.209 of [25]) that as $\delta \downarrow 0$, $\bar{x}_t$ converges in the *ucp* topology to the solution of the Kalman filter equation (2.3.2). ∎

Chapter III

Parameter Estimation for the Linear Case

In this chapter, we develop a parameter estimation scheme for a mixed-time linear model, employing an EM algorithm based on finite-dimensional exact filters. This chapter extends the work of Elliott and Krishnamurthy ([14],[15]) to the mixed-time case in such a way as to expose the limiting behavior of the filters as the time δ between observations approaches 0. Section 3.1 gives the filters themselves. In Section 3.2, we examine the limiting behavior of the filters, which is in some instances surprisingly different from the purely continuous-time case. Section 3.3 gives the EM algorithm.

3.0. Model

We begin with the model. Fix $\delta > 0$. Our $\mathbb{R}^m$-valued signal process x_t and $\mathbb{R}^n$-valued observation process $\{z_k^\delta, k\delta \leq T\}$ are defined by

$$dx_t = B_t x_t dt + Q_t dw_t, \; t \in [0, T], \qquad x_0 = 0, \tag{3.0.1}$$

$$z_k = \delta H_{k\delta} x_{k\delta} + \Delta_k. \tag{3.0.2}$$

Equation (3.0.2) gives an observation process intended as a discrete-time analog of the continuous time observation process

$$dy_t = H_t x_t dt + R_t dV_t. \tag{3.0.3}$$

Indeed, as $\delta \downarrow 0$, the stochastic process $y_t^\delta := \sum_{k\delta \leq t} z_k$ converges to the process $\{y_t\}$ uniformly in probability on $[0, T]$.

With regards to Equations (3.0.1),(3.0.2), and (3.0.3) we make the following

Assumptions 3.1:

1. B, Q are continuous $m \times m$ matrix valued functions on $[0, T]$;
2. w_t is an m-dimensional standard Brownian motion;
3. Q is a continuos symmetric matrix valued function on $[0, T]$ such that Q_t^2 positive definite and symmetric for all t;
4. H is a continuous $n \times m$ matrix valued function on $[0, T]$;
5. V is a standard n-dimensional Brownian motion which is independent of w_t;
6. R is a continuous symmetric $n \times n$ matrix valued function on $[0, T]$ such that R_t is positive definite for all $t \in [0, T]$;
7. $\Delta_k := R_{k\delta}V_{k\delta} - R_{k\delta-\delta}V_{k\delta-\delta}$.

Writing N_k^2 for the variance of Δ_k, note that

$$N_k^2 = \int_{k\delta-\delta}^{k\delta} R_s^2 ds.$$

Put $x_k := x_{k\delta}$ and $H_k := H_{k\delta}$, $k\delta \leq T$, and let $\Phi(\cdot,\cdot)$ be the fundamental matrix associated with B_t. Then we have for $k \geq 1$,

$$x_{k+1} = \Phi(k\delta + \delta, k\delta)x_k + \int_{k\delta}^{k\delta+\delta} \Phi(k\delta + \delta, \tau)Q_\tau dw_\tau. \tag{3.0.4}$$

Denote $\int_{k\delta}^{k\delta+\delta} \Phi(k\delta+\delta, \tau)Q_\tau dw_\tau$ by v_{k+1}. As the Ito integrals $w.r.t.$ the Brownian motion w_t of deterministic functions over disjoint intervals, the v_k are independent and normally distributed with mean 0. Let M_k^2 be the covariance matrix of v_k, and Note

$$M_k^2 = \int_{k\delta-\delta}^{k\delta} (\Phi(k\delta, \tau)Q_\tau)' (\Phi(k\delta, \tau)Q_\tau)\ d\tau. \tag{3.0.5}$$

Let $A_{k+1}^\delta = \Phi(k\delta + \delta, k\delta)$. Then we may write the discretized version of system (3.0.1)-(3.0.2) as:

$$x_{k+1} = A_{k+1}x_k + M_{k+1}\nu_{k+1}, \quad x_0 = 0, \tag{3.0.6}$$

$$z_k = \delta H_k x_k + N_k w_k. \tag{3.0.7}$$

for two independent sequences $\{v_k\}$ and $\{w_k\}$ of *iid rv*'s with $v_k \sim N(0, I_m)$ and $w_k \sim N(0, I_n)$.

3.1. Recursive Exact Filters

In this section, we use change of measure techniques to derive the finite dimensional exact filters needed for parameter estimation. The arguments are in the spirit of Sections 2.1 and 2.2, but are computationally more complex.

Let φ, ψ be the standard normal densities in m and n dimensions respectively and let

$$\lambda_0 = \frac{\varphi(N_0^{-1}(z_0 - \delta H_0 x_0))}{|N_0|\varphi(z_0)},$$

$$\lambda_l = \frac{\varphi(N_l^{-1}(z_l - \delta H_l x_l))}{|N_l|\varphi(z_l)} \cdot \frac{\psi(M_l^{-1}(x_l - A_l x_{l-1}))}{|M_l|\psi(x_l)}$$

$$\Lambda_k = \prod_{l=0}^{k} \lambda_l.$$

Define a probability measure $\overline{P}$ by

$$\left.\frac{d\overline{P}}{dP}\right|_{\mathcal{G}_k} = \Lambda_k^{-1}.$$

Once again, the existence of $\overline{P}$ follows from Kolmogorov's extension theorem.

Theorem 3.2. *Under $\overline{P}$, the sequences $\{x_k\}$ and $\{z_k\}$ defined by (3.0.6)-(3.0.7) are iid and independent of each other. Moreover, under $\overline{P}$ we have $x_k \sim N(0, I_m)$ and $z_k \sim N(0, I_n)$.*

Proof. Let $f : \mathbb{R}^m \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}$ be measurable functions. Then

$$\begin{aligned}
&\overline{E}[f(x_k)g(z_k)|\mathcal{G}_{k-1}] \\
&\qquad = \frac{E[\Lambda_k^{-1} f(x_k)g(z_k)|\mathcal{G}_{k-1}]}{E[\Lambda_k^{-1}|\mathcal{G}_{k-1}]} \\
&\qquad = \frac{E[\lambda_k^{-1} f(x_k)g(z_k)|\mathcal{G}_{k-1}]}{E[\lambda_k^{-1}|\mathcal{G}_{k-1}]}.
\end{aligned}$$

But

$$\begin{aligned}
E[\lambda_k^{-1}|\mathcal{G}_{k-1}] &= \int_{\mathbb{R}^{m+n}} \lambda_k^{-1}\psi(\nu_k)\varphi(w_k)\,d\nu_k\,dw_k \\
&= \int_{\mathbb{R}^{m+n}} |M_k||N_k|\psi\Big(x_k(\nu_k)\Big)\varphi\Big(z_k(w_k)\Big)\,d\nu_k\,dw_k \\
&= \int_{\mathbb{R}^{m+n}} \psi(x_k)\varphi(z_k)\,dx_k\,dz_k, \quad \text{by (3.0.6) and (3.0.7),} \\
&= 1.
\end{aligned}$$

Similarly,

$$\begin{aligned}
&E[\lambda_k^{-1}f(x_k)g(z_k)|\mathcal{G}_{k-1}] \\
&\qquad = \int_{\mathbb{R}^{m+n}} f(x_k)g(z_k)\psi(x_k)\varphi(z_k)\,dx_k\,dz_k.
\end{aligned}$$

Since this quantity is independent of $\mathcal{G}_{k-1}$, this shows that under $\overline{P}$, x_k and z_k are independent of $\mathcal{G}_{k-1}$ with joint density $\psi(x_k)\varphi(z_k)$. This is tantamount to what was to be shown. ∎

Define

$$H_k^{ij(0)} := \delta\sum_{l=0}^{k} x_l^i x_l^j \qquad H_k^{ij(1)} := \delta\sum_{l=1}^{k} x_l^i x_{l-1}^j$$

$$H_k^{ij(2)} := \delta\sum_{l=1}^{k} x_{l-1}^i x_{l-1}^j \qquad J_k^{in} := \sum_{l=0}^{k} x_l^i z_l^n.$$

It turns out (see Section 3.3) these are the quantities which must be estimated in order to carry out the EM parameter estimation for our model. Traditionally, matrices J_k and $H_k^{(M)}$, $M = 0, 1, 2$, were estimated by computing an estimate for each summand individually. We will, however, derive recursive filters for the sums themselves, extending the work of [14] and [15] to the mixed time environment. Our recursive update approach is much faster than the traditional approach (at least twice as fast), which is highly important since the estimation is to be carried out *within each iteration* of the EM algorithm.

We begin by deriving recursive integral formulae for the 'unnormalized' conditional densities, as in Theorem 2.3. Define

$$\begin{aligned}\alpha_k(x) &:= \overline{E}[\Lambda_k I\{x_k \in dx\}|\mathcal{Z}_k],\\ \beta_k^{ij(M)}(x) &:= \overline{E}[\Lambda_k H_k^{ij(M)} I\{x_k \in dx\}|\mathcal{Z}_k],\ M = 0,\ 1,\ 2,\\ \gamma_k^{in}(x) &:= \overline{E}[\Lambda_k J_k^{in} I\{x_k \in dx\}|\mathcal{Z}_k].\end{aligned}$$

Then for every measurable function $g : \mathbb{R}^m \to \mathbb{R}$, and $M = 0,\ 1,\ 2,$

$$\begin{aligned}\overline{E}[\Lambda_k g(x_k)|\mathcal{Z}_k] &= \int_{\mathbb{R}^m} \alpha_k(x) g(x) dx\\ \overline{E}[\Lambda_k H_k^{ij(M)} g(x_k)|\mathcal{Z}_k] &= \int_{\mathbb{R}^m} \beta_k^{ij(M)}(x) g(x) dx.\end{aligned} \tag{3.1.1}$$

Also define

$$\mu_k(z, x) := \varphi(N_k^{-1}(z - \delta H_k x)), \quad \nu_k(x, z) := \psi(M_k^{-1}(x - A_k z)).$$

The next Lemma gives recursive exact filters for α, β and γ.

Lemma 3.3. *For $k \geq 1$, we have the following recursive equations for the unnormalized conditional densities α, β, and γ.*

$$\beta_k^{ij(0)}(x) = \frac{\mu_k(z_k, x)}{|M_k||N_k|\varphi(z_k)} \left(\int_{\mathbb{R}^m} \beta_{k-1}^{ij(0)}(z) \nu_k(x, z) dz + \delta x^i x^j \int_{\mathbb{R}^m} \alpha_{k-1}(z) \nu_k(x, z) dz \right) \tag{3.1.2}$$

$$\beta_k^{ij(1)}(x) = \frac{\mu_k(z_k, x)}{|M_k||N_k|\varphi(z_k)} \left(\int_{\mathbb{R}^m} \beta_{k-1}^{ij(1)}(z) \nu_k(x, z) dz + \delta x^i \int_{\mathbb{R}^m} z^j \alpha_{k-1}(z) \nu_k(x, z) dz \right) \tag{3.1.3}$$

$$\beta_k^{ij(2)}(x) = \frac{\mu_k(z_k, x)}{|M_k||N_k|\varphi(z_k)} \left(\int_{\mathbb{R}^m} \beta_{k-1}^{ij(2)}(z) \nu_k(x, z) dz + \delta \int_{\mathbb{R}^m} z^i z^j \alpha_{k-1}(z) \nu_k(x, z) dz \right) \tag{3.1.4}$$

$$\alpha_k(x) = \frac{\mu(z_k, x)}{|M_k||N_k|\varphi(z_k)} \left(\int_{\mathbb{R}^m} \alpha_{k-1}(z) \nu_k(x, z) dz \right) \tag{3.1.5}$$

$$\gamma_k^{in}(x) = \frac{\mu_k(z_k, x)}{|M_k||N_k|\varphi(z_k)} \left(\int_{\mathbb{R}^m} \gamma_{k-1}^{in}(z) \nu_k(x, z) dz + x^i z_k^n \int_{\mathbb{R}^m} \alpha_{k-1}(z) \nu_k(x, z) dz \right) \tag{3.1.6}$$

proof. We prove (3.1.2). The proofs of (3.1.3), (3.1.4), (3.1.5), (3.1.6) are similar.

Let $g : \mathbb{R}^m \to \mathbb{R}$ be an arbitrary Borel measurable function. Then

$$\begin{aligned}
\overline{E}[\Lambda_k H^{ij(0)} g(x_k)] &= \overline{E}\Big[\Lambda_{k-1} \frac{\mu_k(z_k, x_k)\nu_k(x_k, x_{k-1})}{|N_k||M_k|\varphi(z_k)\psi(x_k)} H_{k-1}^{ij(0)} g(x_k)|\mathcal{Z}_k\Big] \\
&\qquad + \overline{E}[\Lambda_{k-1} \frac{\mu_k(z_k, x_k)\nu_k(x_k, x_{k-1})}{|N_k||M_k|\varphi(z_k)\psi(x_k)} \delta x_k^i x_k^j g(x_k)|\mathcal{Z}_k] \\
&= \frac{1}{|M_k||N_k|\varphi(z_k)} \Big(\overline{E}\Big[\Lambda_{k-1} H_{k-1}^{ij(0)} \int_{\mathbb{R}^m} \mu_k(z_k, x)\nu_k(x, x_{k-1}) g(x)\, dx|\mathcal{Z}_{k-1}\Big] \\
&\qquad + \overline{E}\Big[\Lambda_{k-1} \int_{\mathbb{R}^m} \mu_k(z_k, x)\nu_k(x, x_{k-1}) \delta x^i x^j g(x)|\mathcal{Z}_k\Big]\Big) \\
&= \frac{1}{|M_k||N_k|\varphi(z_k)} \Big(\int_{\mathbb{R}^m}\int_{\mathbb{R}^m} \mu_k(z_k, x)\nu_k(x, z) g(x)\, dx\, dz \\
&\qquad + \int_{\mathbb{R}^m}\int_{\mathbb{R}^m} \alpha_{k-1}(z)\mu_k(z_k, x)\nu_k(x, z)\delta x^i x^j g(x)\, dx\, dz\Big)
\end{aligned} \tag{3.1.7}$$

Since g was arbitrary, equating (3.1.7) with the RHS of (3.1.1) yields (3.1.2).

∎

Remark: Since we have taken $x_0 = 0$, the initial conditions for all recursions given in Theorem 4 are point masses at the origin.

The densities $\beta^{(M)}$, $M = 0, 1, 2$, and γ are not necessarily Gaussian, and so the recursions of Lemma 3.3 are truly infinite dimensional. Remarkably, however, at each instant in time, $\beta(x)$ and $\gamma(x)$ are each given by a quadratic in x times $\alpha(x)$. The linearity of our model implies that $\alpha_k(x)$ itself is an unnormalized Gaussian density, and so $\alpha_k(x)$ can be completely characterized by finitely many statistics. This allows us to give finite dimensional filters for $\beta^{ij(M)}, M = 0, 1, 2$, and γ^{in}, which appear in Theorems 3.4, 3.5, and 3.6. First we prove the existence of the aforementioned quadratics, and determine their coefficients. Toward this, let , $\hat{x}, \hat{P}$ be as in Chapter II and define

$$\sigma_k := A_k' M_k^{-2} A_k + \hat{P}_{k-1}^{-1},$$

$$F_k := M_k^{-2} A_k \sigma_k^{-1},$$

$$G_k := \sigma_k^{-1} \hat{P}_{k-1}^{-1} \hat{x}_{k-1}.$$

Theorem 3.4. *For $1 \le i, j \le m$, define sequences $\{a_k^{ij(0)}\}, \{b_k^{ij(0)}\}, \{d_k^{ij(0)}\}$ by*

$$d_{k+1}^{ij(0)} := F_{k+1} d_k^{ij(0)} F'_{k+1} + \frac{\delta}{2}(e_i e'_j + e_j e'_i), \quad d_0^{ij(0)} := \frac{\delta}{2}(e_i e'_j + e_j e'_i), \tag{3.1.8}$$

$$b_{k+1}^{ij(0)} := F_{k+1}(b_k^{ij(0)} + 2 d_k^{ij(0)} G_{k+1}), \quad b_0^{ij(0)} = 0_{m\times 1}, \tag{3.1.9}$$

$$a_{k+1}^{ij(0)} := a_k^{ij(0)} + (b_k^{ij(0)})' G_{k+1} + Tr(d_k^{ij(0)} \sigma^{-1}) + G'_{k+1} d_k^{ij(0)} G_{k+1}, \; a_0^{ij(0)} = 0. \tag{3.1.10}$$

Then,

$$\beta_k^{ij(0)}(x) = [a_k^{ij(0)} + (b_k^{ij(0)})' x + x' d_k^{ij(0)} x] \alpha_k(x). \tag{3.1.11}$$

proof. We proceed by induction. Since $\beta_0^{ij(0)}$ and α_0 are point masses at the origin, (3.1.11) clearly holds for $k = 0$. Assume (3.1.11) holds at time k. Then at time $k+1$, we have, by Lemma 3.3 and induction hypothesis,

$$\begin{aligned} \beta_{k+1}^{ij(0)}(x) = & \frac{\mu_{k+1}(z_k, x)}{|M_{k+1}||N_{k+1}|\varphi(z_k)} \int_{\mathbb{R}^m} \nu_{k+1}(x,z)(a_k + b'_k z + z' d_k z)\alpha_k \, dz \\ & + \delta x^i x^j \frac{\mu_{k+1}(z_k, x)}{|M_{k+1}||N_{k+1}|\varphi(z_k)} \int_{\mathbb{R}^m} \alpha_k(z) \nu(x,z) \, dz. \end{aligned} \tag{3.1.12}$$

Let

$$\begin{aligned} \overline{\alpha}_k :&= \int_{\mathbb{R}^m} \alpha_k(z) \, dz, \\ K(x) :&= \frac{\mu_{k+1}(z_{k+1}, x)}{|M_{k+1}||N_{k+1}|\varphi(z_{k+1})} 2\pi^{-m} |M_{k+1}|^{-1} |\hat{P}_k|^{-1/2} \overline{\alpha}_k, \\ \delta_{k+1} :&= 2(x' M_{k+1}^{-2} A_{k+1} + \hat{x}_k \hat{P}_k^{-1})', \\ K_1(x) :&= K(x) \left[-\frac{1}{2}(x' M_{k+1}^{-2} x + \hat{x}'_{k+1} \hat{P}_{k+1} \hat{x}_{k+1}) \right]. \end{aligned}$$

And let L denote the first term on the RHS of (3.1.12). Then we have

$$\begin{aligned} L = & K(x) \int_{\mathbb{R}^m} \exp\Big(-\frac{1}{2} \Big[(x - A_{k+1} z)' M_{k+1}^{-2} (x - A_{k+1} z) \\ & + (z - \hat{x})' \hat{P}_K^{-1} (z - \hat{x})' \Big] \Big) (a_k + b'_k z + z' d_z) \, dz \\ = & K_1(x) \int_{\mathbb{R}^m} \exp(-\frac{1}{2}[z' \sigma z - \delta'_{k+1} z])(a_k + b'_k z + z' d_z) \, dz. \end{aligned} \tag{3.1.13}$$

Completing the square in the exponential in (3.1.13) yields

$$
\begin{aligned}
L = K_1(x)\exp&\left[-\frac{1}{2}\left(-\frac{\delta_{k+1}\sigma_{k+1}^{-1}\delta_{k+1}}{4}\right)\right] \\
&\times \int_{\mathbb{R}^m} \exp\left[-\frac{1}{2}\left(z-\frac{\sigma_{k+1}^{-1}\delta_{k+1}}{2}\right)'\sigma_{k+1}\left(z-\frac{\sigma_{k+1}^{-1}\delta_{k+1}}{2}\right)\right] \\
&\qquad\qquad \times (a_k + b_k' z + z' d_k z)\, dz
\end{aligned}
\tag{3.1.14}
$$

Thinking of z as a normal random variable with mean $\frac{1}{2}\sigma_{k+1}\delta_{k+1}$ and variance σ_{k+1}, the integral in (3.1.14) becomes an expectation, equalling

$$
(2\pi)^{m/2}|\sigma_{k+1}|^{-1/2}(a_k + b_k' E\{z\} + E\{z' d_k z\}).
$$

Now,

$$
\begin{aligned}
E\{z' d_k z\} &= E\{(z-E(z))' d_k (z-E(z))\} + E(z')\, d_k\, E(z) \\
&= \sum_{p,q=1}^{m} d_k(p,q) - \sigma_{k+1}^{-1}(p,q) \ + \frac{1}{4}(\sigma_{k+1}^{-1}\delta_{k+1})' d_k (\sigma_{k+1}^{-1}\delta_{k+1}) \\
&= \mathrm{Tr}(d_k \sigma_{k+1}^{-1}) + \frac{1}{4}(\sigma_{k+1}^{-1}\delta_{k+1})'\ d_k\ (\sigma_{k+1}^{-1}\delta_{k+1})
\end{aligned}
\tag{3.1.15}
$$

Note that the second term of (3.1.12) equals $\delta x^i x^j \cdot \alpha_{k+1}$ by Lemma 3.3. Hence combining (3.1.12),(3.1.13),(3.1.14),(3.1.15), we get

$$
\begin{aligned}
\beta_{k+1}^{ij(0)}(x) = \alpha_{k+1}(x)&\left[a_k + \frac{1}{2} b_k' \sigma_{k+1}^{-1}\delta_k + \mathrm{Tr}(d_k \sigma_{k+1}^{-1})\right. \\
&\qquad \left. + \frac{1}{4}\delta_{k+1}'\sigma_{k+1} d_k \delta_{k+1}\sigma_{k+1} + \delta x^i x^j\right].
\end{aligned}
$$

Substituting in this for δ_{k+1} yields

$$
\beta_{k+1}^{ij(0)}(x) = [a_{k+1} + b_{k+1}' x + x' d_{k+1} x]\alpha_{k+1}(x)
$$

where $a_{k+1}, b_{k+1}, d_{k+1}$ are as given by (3.1.8), (3.1.9), (3.1.10). ∎

The proofs of the following theorems are similar:

Theorem 3.5. *At time k, the unnormalized densities $\beta_k^{ij(M)}$, $M = 1, 2$, are given by*

$$\beta_k^{ij(M)}(x) = [a_k^{ij(M)} + (b_k^{ij(M)})'x + x'd_k^{ij(M)}x]\alpha_k(x)$$

where the coefficients a, b, and d are given by the following recursions:

$$d_{k+1}^{ij(1)} = F_{k+1}d_k^{ij(1)}F_{k+1}' + \frac{\delta}{2}[e_ie_j'F_{k+1}' + F_{k+1}e_je_i'], \qquad d_0^{ij(1)} = 0_{m\times m}; \tag{3.1.16}$$

$$d_{k+1}^{ij(2)} = F_{k+1}(d_k^{ij(2)} + \frac{\delta}{2}[e_ie_j' + e_je_i'])F_{k+1}', \qquad d_0^{ij(2)} = 0_{m\times m}; \tag{3.1.17}$$

$$b_{k+1}^{ij(1)} = F_{k+1}(b_k^{ij(1)} + 2d_k^{ij(1)}G_{k+1}) +_i e_j'G_{k+1}, \qquad b_0^{ij(1)} = 0_{m\times 1}; \tag{3.1.18}$$

$$b_{k+1}^{ij(2)} = F_{k+1}\left(b_k^{ij(2)} + (2d_k^{ij(2)} + e_ie_j' + e_je_i')G_{k+1}\right), \qquad b_0^{ij(2)} = 0_{m\times 1}; \tag{3.1.19}$$

$$a_{k+1}^{ij(M)} = a_k^{ij(M)} + (b_k^{ij(M)})'G_k + Tr(d_k^{ij(M)}\sigma_{k+1}^{-1}) + G_{k+1}'d_k^{ij(M)}G_{k+1}, \quad a_0^{ij(M)} = 0. \tag{3.1.20}$$

Theorem 3.6. *At time k, the unnormalized density $\gamma_k^{in}(x)$ is given by*

$$\gamma_k^{in}(x) = [\hat{a}_k^{in} + (\hat{b}_k^{in})'x]\alpha_k(x)$$

where $\hat{a}_k, \hat{b}_k$ are given by

$$\hat{b}_{k+1}^{in} = F_k\hat{b}_k^{in} + e_iz_{k+1}^n, \qquad \hat{b}_0^{in} = 0_{m\times 1};$$

$$\hat{a}_{k+1}^{in} = a_k^{in} + (b_k^{in})'G_k, \qquad a_0^{in} = 0.$$

With these unnormalized conditional densities in hand, it is straightforward to compute the conditional expectations of $H_k^{(M)}$ and J_k given Z_k; and we now give methods for doing so. These are the estimates we will use in EM parameter estimation. Define

$$\hat{H}_k^{ij(0)} := E[H_k^{ij(0)}|Z_k], \qquad \hat{H}_k^{ij(1)} := E[H_k^{ij(1)}|Z_k],$$

$$\hat{H}_k^{ij(2)} := E[H_k^{ij(2)}|Z_k], \qquad \hat{J}_k^{in} := E[J_k^{in}|Z_k].$$

Lemma 3.7. *Let Y be a random variable on Ω such that for some a, b, c in $\mathbb{R}, \mathbb{R}^m, \mathbb{R}^{(m\times m)}$ respectively, we have, for all $x \in \mathbb{R}^m$,*

$$\overline{E}[\Lambda_k Y I\{x_k \in dx\}|\mathcal{Z}_k] = (a + b'x + x'cx)\alpha_k(x).$$

Then the conditional expectation $E[Y|\mathcal{Z}_k]$ is given by

$$E[Y|\mathcal{Z}_k] = a + b'\hat{x}_k + \sum_{p,q} c(p,q)\hat{P}_k(p,q) + \hat{x}'_T c\hat{x}_k$$

Proof. Using abstract Bayes' rule we have

$$E[Y|\mathcal{Z}_k] = \frac{\overline{E}[\Lambda_k Y|\mathcal{Z}_k]}{\overline{E}[\Lambda_k|\mathcal{Z}_k]} = \frac{\int_{\mathbb{R}^m}(a + b'x + x'cx)\alpha_k(x)\,dx}{\int_{\mathbb{R}^m}\alpha_k(x)\,dx}. \tag{3.1.21}$$

But since $\alpha_k(x)$ is an unnormalized density of x_k given $\mathcal{Z}_k$, the numerator of the far RHS is equal to

$$\begin{aligned}&\int_{\mathbb{R}^m}\alpha_k(x)\,dx \cdot E[a + b'x_k + x'cx_k]\\ &= \left(\int_{\mathbb{R}^m}\alpha_k(x)\,dx\right)\left[a + b'\hat{x}_k + \sum_{p,q} c(p,q)\hat{P}_k(p,q) + \hat{x}'_k c\hat{x}_k\right]\end{aligned}$$

Substituting this in (3.1.21) proves the result. QED

The following corollary of Lemma 3.7, together with Thoerems 3.4, 3.5, 3.6, gives an algorithm for computing the conditional expectations $\hat{H}^{ij(M)}$, $M = 0, 1, 2$, and $\hat{J}^{in}$.

Corollary 3.8. *For $k \geq 0$, we have*

$$\begin{aligned}\hat{H}_k^{ij(M)} &= a_k^{ij(M)} + (b_k^{ij(M)})'\hat{x}_k + Tr(d_k^{ij(M)}\hat{P}_k) + \hat{x}'_k d_k^{ij(M)}\hat{x}_k,\ M = 0,1,2,\\ \hat{J}_k^{in} &= \hat{a}_k^{in} + (\hat{b}_k^{in})'\hat{x}_k.\end{aligned}$$

3.2. Limitimg Case

We are now interested in examining the behavior of the results from Section 3.2 as $\delta \downarrow 0$. We begin by writing the expressions F_k and G_k in a convenient way. Recall that for square matrices X, Y we have

$$(X+Y)^{-1} = X^{-1}(I + YX^{-1})^{-1} = X^{-1}[I - YX^{-1} + YX^{-1}YX^{-1} - \ldots]$$

provided the sequence converges. Hence, for sufficiently small δ,

$$\begin{aligned}
F_k &= M_k^{-2} A_k \sigma_k^{-1} \\
&= M_k^{-2} A_k [(A_k' M_k^{-2} A_k) + \hat{P}_{k-1}^{-1}]^{-1} \\
&= M_k^{-2} A_k (A_k' M_k^{-2} A_k)^{-1} [I - \hat{P}_{k-1}^{-1} (A_k' M_k^{-2} A_k)^{-1} + O(\delta^2)] \\
&= (A_k')^{-1} [I - \hat{P}_{k-1}^{-1} (A_k' M_k^{-2} A_k)^{-1} + O(\delta^2)] \\
&= (A_k')^{-1} - (A_k')^{-1} \hat{P}_{k-1}^{-1}]^{-1} A_k^{-1} M_k^2 (A_k')^{-1} + O(\delta^2)] \\
&= (I - \delta B_{k\delta}') - (I - \delta B_{k\delta}') \hat{P}_{k-1}^{-1} (I - \delta B_{k\delta}) M_k^2 (I - \delta B_{k\delta}') + O(\delta^2)].
\end{aligned}$$

Since M^2 is $O(\delta)$, this equals

$$\begin{aligned}
&I - \delta B_{k\delta}' - \hat{P}_{k-1}^{-1} M_k^2 + O(\delta^2) \\
&= I + \delta[-B_{k\delta}' - \hat{P}_{k-1}^{-1}(M_k^2/\delta) + O(\delta)] \qquad (3.2.1) \\
&= F_k.
\end{aligned}$$

Also, note

$$\begin{aligned}
\sigma_{k+1}^{-1} &= [(A_{k+1}' M_{k+1}^{-2} A_{k+1}) + \hat{P}_k^{-1}]^{-1} \\
&= (A_{k+1}^{-1} M_{k+1}^2 (A_{k+1}')^{-1})(I + O(\delta)) \\
&= [I + O(\delta)] M_{k+1}^2 [I + O(\delta)][I + O(\delta)] \\
&= M_{k+1}^2 + O(\delta^2) \\
&= \delta[M_{k+1}^2/\delta + O(\delta)]
\end{aligned}$$

Hence,

$$
\begin{aligned}
G_{k+1} &= \sigma_{k+1}^{-1}\hat{P}_k^{-1}\hat{x}_k \\
&= \delta\big[(M_{k+1}^2/\delta)\hat{P}_k^{-1}\hat{x}_k + O(\delta)\big].
\end{aligned} \tag{3.2.2}
$$

We are now armed to investigate the behavior of the statistics given in Theorems 3.4, 3.5, and 3.6 as $\delta \downarrow 0$. The theorems of this section hold for all i and j; but, for notational convenience, we will supress the superscripts ij.

Notation: If $\{r_k, 0 \le k \le T/\delta\}$ is any sequence, let ${}^\delta r$ denote the *cadlag* step function on [0,T] taking values r_k on respective intervals $[k\delta, \max\{k\delta+\delta, T\})$.

Theorem 3.9. *Define functions* $a : [0,T] \to \mathbb{R}$, $b : [0,T] \to \mathbb{R}^m, c : [0,T] \to \mathbb{R}^{m\times m}, \tilde{a} : [0,t] \to \mathbb{R}$, *and* $\tilde{b} : [0,T] \to \mathbb{R}^m$ *by*

$$
\begin{aligned}
dc/dt = -\Big(B_t' + \hat{P}_t^{-1}Q_t^2\Big)c_t - c_t\Big(B_t + (\hat{P}_tQ_t^2)\Big) + \frac{1}{2}\Big(e_ie_j' + e_je_i'\Big),& \\
c_0 = O_{m\times m},&
\end{aligned} \tag{3.2.3}
$$

$$
db/dt = -\Big(B_t' + \hat{P}_t^{-1}Q_t^2\Big)b_t + 2c_tQ_t^2\hat{P}_t^{-1}\hat{x}_t, \qquad b_0 = O_{m\times 1}, \tag{3.2.4}
$$

$$
da/dt = Tr(c_tQ_t^2) + b_t'Q_t^2\hat{P}_t^{-1}\hat{x}_t, \qquad a_0 = 0, \tag{3.2.5}
$$

$$
d\tilde{b}_t = -\Big(B_t' + \hat{P}_t^{-1}Q_t^2\Big)\tilde{b}_t d_t + e_i(dy_k, e_n), \qquad \tilde{b}_0 = 0_{m\times 1}, \tag{3.2.6}
$$

$$
d\tilde{a}/dt = \tilde{b}_t'Q_t^2\hat{P}_t^{-1}\hat{x}_t, \qquad \tilde{a}_0 = 0. \tag{3.2.7}
$$

Then as $\delta \downarrow 0$, *the we have the following convergences, uniformly on* $[0,T]$: ${}^\delta d^{(M)} \to c$, ${}^\delta b^{(M)} \to b$, ${}^\delta a^{(M)} \to a$, $M = 0,1,2$, ${}^\delta \hat{a} \to \tilde{a}$, ${}^\delta \hat{b} \to \tilde{b}$.

Proof. We prove the convergence only for ${}^\delta d^{(0)}, {}^\delta \hat{a}$, and ${}^\delta \hat{b}$. Proofs for the others are similar.

First, note that the Φ^2 term on the RHS of (3.0.5) converges uniformly to the identity matrix I as $\delta \downarrow 0$. It now follows from Theorem 9.3.1 of [2] that ${}^\delta M^2/\delta$ converges to Q^2 uniformly on $[0,T]$ as $\delta \downarrow 0$. Now, using (3.2.1) and (3.2.2), (3.1.8) may be written, (dropping the superscetipts (0) and δ),

$$
\begin{aligned}
d_{k+1} &= F_{k+1} d_k F'_{k+1} + \frac{\delta}{2}\Big(e_i e'_j + e_j e'_i\Big) \\
&= \Big\{I + \delta\Big(-B'_{k+1} - \hat{P}_k^{-1}(M^2_{k+1}/\delta)\Big)\Big\} d_k \Big\{I + \\
&\quad \delta\Big(-B'_{k+1} - \hat{P}_k^{-1}(M^2_{k+1}/\delta)\Big)\Big\}' + \frac{\delta}{2}\Big(e_i e'_j + e_j e'_i\Big) + O(\delta^2) \\
&= d_k + \delta\Big\{\Big(-B'_{k+1} + \hat{P}_k^{-1}(M^2_{k+1}/\delta)\Big) d_k \\
&\qquad + d_k\Big(-B_{k+1} - \hat{P}_k^{-1}(M^2_{k+1}/\delta)\Big) + \frac{1}{2}\Big(e_i e'_j + e_j e'_i\Big) \\
&\qquad + O(\delta)\Big\}, d_0 = \frac{\delta}{2}\Big(e_i e'_j + e_j e'_i\Big).
\end{aligned}
$$

In light of the convergence of ${}^\delta M^2/\delta$ to Q^2, it now follows from stability results for ODE's, (see, for example, Theorem 15, p209 of [25]), that ${}^\delta d^{(0)}$ converges uniformly on $[0,T]$ to the solution of (3.2.3). As for $\hat{a}^{in}, \hat{b}^{in}$, recall they are given by

$$
\begin{aligned}
\hat{b}_{k+1} &= F_k b_k + e_i\langle z_{k+1} e_n\rangle, \quad \hat{b}_0 = O_{m\times 1}, \\
\hat{a}_{k+1} &= \hat{a}_k + \Big(\hat{b}_k\Big)' G_k, \quad a_0 = 0.
\end{aligned}
$$

Substituting in this for F and G, and using the step functions gives

$$
\begin{aligned}
{}^\delta\hat{b}_{k+1} &= ((I - \delta B'_{k\delta+\delta} - \hat{P}_k^{-1}({}^\delta M^2_{k+1}) + O(\delta^{-2}))({}^\delta b_k) + e_i\langle z_{k+1} e_n\rangle, \quad {}^\delta\hat{b}_0 = O_{m\times 1}, \\
{}^\delta\hat{a}_{k+1} &= {}^\delta\hat{a}_k + \Big({}^\delta\hat{b}_k\Big)' \delta[({}^\delta M_k +^2/\delta) P_{k-1}^{-1}\hat{x}_{k-1} + O(\delta)], \quad {}^\delta a_0 = 0.
\end{aligned}
$$

It now follows by stability results for ODE's, (see, for example, Theorem 15, p209 of [25]), that ${}^\delta\hat{b}$ and ${}^\delta\hat{a}$ converge to the respective solutions of (3.2.6) and (3.2.7), uniformly on $[0,T]$ as $\delta \downarrow 0$. ∎

Let $\hat{x}_k := E[x_k \mid \mathcal{Z}_k]$, $\hat{P}_k := \mathrm{cov}[x_k \mid \mathcal{Z}_k]$, and let ${}^\delta\hat{x}$, ${}^\delta\hat{P}$ be the associated *cadlag* step functions on $[0,T]$. Also, let $\hat{x}$, $\hat{P}$ be the respective solutions of (2.3.2) and (2.3.1). Then it has already been shown (Theorem 2.6) that as $\delta \downarrow 0$, ${}^\delta x \to \hat{x}$ in *ucp* and ${}^\delta\hat{P} \to \hat{P}$ uniformly on $[0,T]$. In light of this, the following is a consequence of Theorems 3.4 and 3.5, Corollary 3.8, and Theorem 3.9:

Corollary 3.10. *Identify* $\{\hat{H}_k^{(M)}, 1 \leq k \leq T/\delta\}$, $M = 0,1,2$, *with cadlag step functions* ${}^\delta\hat{H}^{(M)}$ *on* $[0,T]$. *Then as* $\delta \downarrow 0$, ${}^\delta\hat{H}^{(M)}$, $M = 0,1,2$ *converge in the ucp topology to the common limit*

$$a_t + b_t'\hat{x}_t + Tr(c_t\hat{P}_t) + \hat{x}_t'c_t\hat{x}_t, \tag{3.2.8}$$

where a, b , *and* c *are as given in Theorem* 3.9.

Similarly, the following is a consequence of Theorem 3.6, Corollary 3.8, and Theorem 3.9:

Corollary 3.11. *Identify* $\hat{J}$ *with a cadlag step function* ${}^\delta\hat{J}$ *on* $[0,T]$. *Then as* $\delta \downarrow 0$, ${}^\delta\hat{J}$ *converges in ucp to*

$$\tilde{a}_t + \tilde{b}_t'\hat{x}_t, \tag{3.2.9}$$

where $\tilde{a}$ *and* $\tilde{b}$ *are as given in Theorem* 3.9.

Comparison with Purely Continuous Case

Thinking of x_t as the continuous-time solution of (3.0.1), note that

$$\begin{aligned}
{}^\delta H_t^{ij(0)} &:= \sum_{l\delta\leq t} \delta x^i_{l\delta} x^j_{l\delta} \,, \\
{}^\delta H_t^{ij(1)} &:= \sum_{l\delta\leq t} \delta x^i_{l\delta} x^j_{(l-1)\delta} \,, \\
{}^\delta H_t^{ij(2)} &:= \sum_{l\delta\leq t-\delta} \delta x^i_{l\delta} x^j_{(l\delta)} \,.
\end{aligned}$$

Hence ${}^\delta\hat{H}^{ij(M)}$, $M = 0,1,2$, converge *a.e.* to the common limit

$$H_t^{ij} := \int_0^t x_s^i x_s^j ds.$$

Also, as noted earlier, the process ${}^\delta y_t := \sum^\delta_{k\delta\leq t} z_k$ converges in *ucp* to the solution y_t of

$$dy_t = H_t x_t dt + R_t dV_t.$$

Define ${}^{\delta}\mathcal{Z}_t := \sigma\{{}^{\delta}z_k, k\delta \le t\}$ and $\mathcal{Y}_t := \sigma\{y_s, s \le t\}$. From the above, we might expect that for $M = 0, 1, 2$, we would have

$$E[{}^{\delta}H_t^{(M)} \mid {}^{\delta}\mathcal{Z}_t] \to E[H_t \mid \mathcal{Y}_t], \tag{3.2.10}$$

in *ucp* as $\delta \downarrow 0$. There are several heuristic reasons for making this conjecture: First of all, 'corresponding things' are converging to 'corresponding things.' More than that, for a given t, as $\delta(n) \downarrow 0$ through some sequence, the sequence of sigma algebras ${}^{\delta(n)}\mathcal{Z}_t$, $n \ge 1$, gives us, generally speaking, more and better information about H_t, reminiscent of a filtration. Thus we are not too far from the domain of martingale convergence. Finally, convergence to the continuous-time filter was obtained for another filter associated with this model in Theorem 2.6.

Now, limits in *ucp* are unique up to events of probability 0, and Corollary 3.10 tells us exactly what $E[\delta H_t^{(M)} \mid {}^{\delta}\mathcal{Z}_t]$ converges to as $\delta \downarrow 0$. Hence if all the above handwaving is worth anything, we expect the RHS of (3.2.10) to equal (3.2.8). In fact, this is true, and was proved by Elliott and Krishnamurthy (see [14], Theorem 3.2). The same type of intuition fails, however, in the case of $\hat{J}_t$, as the following theorem of Elliott and Krishnamurthy ([14]) shows:

Theorem 3.12. *For $t \in [0, T]$, we have*

$$E\left[\int_0^t x_s^i dy_s^j \;\middle|\; \mathcal{Y}_t\right] = \overline{a}_t^{ij} + \hat{x}_t' \overline{b}_t^{ij} + \sum_{p,q} \overline{c}_t^{ij}(p,q)\hat{P}_t(p,q) + \hat{x}_t'\overline{c}_t^{ij}, \tag{3.2.11}$$

where, dropping the superscripts ij, $\overline{a}, \overline{b}_,$, and $\overline{c}$ are given by

$$\begin{aligned}
\frac{d\overline{a}}{dt} &= Tr(\overline{c}_t Q_t^2) + \overline{b}_t Q_t^2 \hat{P}_t^{-1}\hat{x}_t, \qquad \overline{a}_0 = 0 \in \mathbb{R}. \\
d\overline{b}_t &= \left[-(B_t' + \hat{P}_t^{-1}Q_t^2)\overline{b}_t + 2\overline{c}_t Q_t^2 \hat{P}_t^{-1}\hat{x}_t\right] dt + e_i(dy_t, e_j), \qquad \overline{b}_0 \in \mathbb{R}^m. \\
\frac{d\overline{c}}{dt} &= -(B_t' + \hat{P}_t^{-1}Q_t^2)\overline{c}_t - \overline{c}_t'(B_t + Q_t^2\hat{P}_t^{-1}) \\
&\qquad + \frac{1}{2}(e_i e_j' H_t + H_t' e_j e_i'), \qquad \overline{c}_0 = 0 \in \mathbb{R}^{m \times m}.
\end{aligned}$$

This is qualitatively different from the limit (3.2.9) as $\delta \downarrow 0$ of its discrete-time counterpart . The RHS of (3.2.11) is quadratic in $\hat{x}_t$, while (3.2.9) is linear in $\hat{x}_t$. Also, (3.2.11) is given in terms of five statistics, while (3.2.9) is given in terms of only four.

3.3. Change of Measure and EM Algorithm

This section deals with parameter estimation in the time-invariant case of the model of Section 3.1, based on the filters of Section 3.2. The EM algorithm, like the filters themselves, is rooted in the technique of changing probability measure.

Fix $\delta > 0$ and consider the system

$$dx_t = B_0 x_t dt + Q_0 dw_t, \ t \in [0, T], \ \ x_0 = 0, \tag{3.3.1}$$

$$z_k = \delta H_0 x_{k\delta} + \Delta_k, \ \ k\delta \leq T, \tag{3.3.2}$$

subject to **Assumptions 3.13**:

1. $x_t \in \mathbb{R}^m$ and $z_k \in \mathbb{R}^n$.
2. w and V are independent standard Brownian motions in $\mathbb{R}^m$ and $\mathbb{R}^n$ respectively.
3. R_0 is a symmetric, positive definite matrix and $\Delta_k = R_0 V_{k\delta} - R_0 V_{k\delta-\delta}$.
4. H_0, C_0, and Q_0 are real matrices of appropriate dimension. Q_0 is symmetric and positive definite.

Let $x_k = x_{k\delta}$. For $1 \leq k \leq T/\delta$, let $\mathcal{Z}_k$ be the complete sigma field generated by $\{z_1, \ldots, z_k\}$ and let

$$\mathcal{G}_k := \sigma\{x_1, \ldots, x_k, z_1, \ldots, z_k\}.$$

Define

$$A_0 := \exp(\delta B_0),$$

$$M_0^2 := \delta Q_0^2,$$

$$N_0^2 := \delta R_0^2.$$

Then

$$x_{k+1} = A_0 x_k + M_0 v_{k+1}, \tag{3.3.3}$$

$$z_k = \delta H_0 x_k + N_0 w_k \tag{3.3.4}$$

for two sequences $\{v_k\}$ and $\{w_k\}$ of independent, identically distributed random variables with $v_k \sim N(0, I_m)$ and $w_k \sim N(0, I_n)$.

Let φ, ψ be the standard normal densities in m and n dimensions respectively. For a given hypothetical parameter set $\theta = (A, M, H, N)$, let

$$\lambda_0(\theta) = \frac{\varphi(N^{-1}(z_0 - \delta H x_0))}{|N|\varphi(z_0)},$$

$$\lambda_l(\theta) = \frac{\varphi(N^{-1}(z_l - \delta H x_l))}{|N|\varphi(z_l)} \cdot \frac{\psi(M^{-1}(x_l - A x_{l-1}))}{|M|\psi(x_l)},$$

$$\Lambda_k(\theta) = \prod_{l=0}^{k} \lambda_l(\theta).$$

Let $\overline{P}$ be a probability measure under which x_k, z_k are *iid* $N(0,I)$ r.v.'s . Given a parameter set $\theta = (A, M, H, N)$, define the probability measure $P(\theta)$ by

$$\frac{dP(\theta)}{d\overline{P}}\Big|_{\mathcal{G}_k} = \Lambda_k(\theta).$$

It is shown in [14] that $P(\theta)$ is a probability measure under which (3.3.3)-(3.3.4) hold with A_0,M_0,H_0, and N_0 replaced by A,M,H, and N, respectively, and with $\{v_k\}$ and $\{w_k\}$ *iid* $N(0, I)$. For a given k and parameter set θ, define

$$\mathcal{L}(\theta) := E\left[\frac{dP(\theta)}{dP_0}\,\Big|\mathcal{Z}_k\right]$$

where $\mathcal{Z}_k := \sigma(z_1, \ldots, z_k)$ and P_0 is the 'real life' probability measure under which (3.3.3)-(3.3.4) hold for fixed but unknown matrices A_0,H_0,M_0,N_0. We will define a maximum likelihood estimate of θ to be a choice of A, H, M, N which makes $\mathcal{L}(\theta)$ as large as possible. Our approach uses the EM algorithm: For parameter sets $\theta, \tilde{\theta}$, define

$$Q(\theta, \tilde{\theta}) := E_{\tilde{\theta}}\left[\log\left(\frac{dP(\theta)}{dP(\tilde{\theta})}\right)\Big|\mathcal{Z}_k\right].$$

For each k with $1 \leq k \leq T/\delta$, our EM parameter estimation scheme will go as follows:

Step 1) Input z_k. Set $j = 0$.

Step 2) Find an initial estimate $\hat{\theta}^0$ for θ_0, (see *Remarks* below).

Step 3) Using Theorem 2.5, obtain estimates for the conditional mean $\hat{x}_k$ and covariance $\hat{P}_k$ of x_k given $\mathcal{Z}_k$ under $P(\hat{\theta}^j)$.

Step 4) Use Theorems 3.4, 3.5, and 3.6 to update $\hat{H}_k^{(M)}, M = 0, 1, 2$, and $\hat{J}_k$ under $P(\hat{\theta}_j)$.

Step 5) Set $\hat{\theta}^{j+1}$ equal to a value which maximazes $Q(\cdot, \hat{\theta}^j)$.

Step 6) If a stopping criterion has been reached, output θ^{j+1} as the EM estimate of θ_0 and terminate the EM algorithm. Otherwise, set j to $j+1$ and go to Step 3.

Remarks : For $k > 1$, we will have previousely estimated the model parameters based upon $z_1, z_2, \ldots, z_{k-1}$, and can take this previous estimate for $\hat{\theta}^0$. The initial estimate for $k = 1$ will depend on the application.

The remainder of this section deals with Step 4 of the above algorithm, and hence k will remain fixed. For notational convenience, the role of $\hat{\theta}^j$ will be played by a fixed parameter set $\bar{\theta}$. (A, H, M, N) will denote the parameter values associated with $\hat{\theta}^j$.

For every admissible parameter set $\theta = (A, M, H, N)$, we have

$$\begin{aligned}
Q(\theta, \bar{\theta}) &= E_{\bar{\theta}}\left[\log \frac{dP(\theta)}{dP(\bar{\theta})}\middle| \mathcal{Z}_k\right] \\
&= E_{\bar{\theta}}\left[\log \frac{dP(\theta)/d\overline{P}}{dP(\bar{\theta})/d\overline{P}}\middle| \mathcal{Z}_k\right] \\
&= E_{\bar{\theta}}\left[\log \frac{\Lambda_k(\theta)}{\Lambda_k(\bar{\theta})}\middle| \mathcal{Z}_k\right] \\
&= E_{\bar{\theta}}\left[\sum_{l=0}^{k} \log \lambda_l(\theta) - \sum_{l=0}^{k} \log \lambda_l(\bar{\theta}) \,\middle|\, \mathcal{Z}_k\right].
\end{aligned}$$

Substituting for the λ's and recalling that φ and ψ are standard normal densities, the above becomes

$$
\begin{aligned}
Q(\theta,\hat{\theta}) = E_{\hat{\theta}}\Big[&- k\log|M| - (k+1)\log|N| \\
&- \frac{1}{2}\sum_{l=1}^{k}(x_l - Ax_{l-1})'M^{-2}(x_l - Ax_{l-1}) \\
&- \frac{1}{2}\sum_{l=1}^{k}(z_l - \delta H x_l)'N^{-2}(z_l - \delta H x_l) + R(\hat{\theta}) \;\Big|\; Z_k\Big],
\end{aligned}
\tag{3.3.5}
$$

where $R(\hat{\theta})$ does not depend on θ.

To maximize Q, we set the partial derivatives of (3.3.5) *w.r.t.* each of the parameters A, M, H, N equal to 0 and solve, yielding

$$
\begin{aligned}
A &= E\left[\sum_{l=1}^{k} x_l x_{l-1}'\Big| Z_k\right]\left(E[\sum_{l=1}^{k} x_{l-1}x_{l-1}'\Big| Z]\right)^{-1} \\
&= E\left[\delta\sum_{l=1}^{k} x_l x_{l-1}'\Big| Z_k\right]\left(E[\delta\sum_{l=1}^{k} x_{l-1}x_{l-1}'\Big| Z]\right)^{-1} \\
&= \hat{H}_k^{(1)}(\hat{H}_k^{(2)})^{-1}; \\
\delta H &= E\left[\sum_{l=0}^{k} z_l x_l'\Big| Z_k\right]\left(E[\sum_{l=0}^{k} x_l x_l'\Big| Z_k]\right)^{-1} \\
H &= E\left[\sum_{l=0}^{k} z_l x_l'\Big| Z_k\right]\left(E\left[\delta\sum_{l=0}^{k} x_l x_l'\Big| Z_k\right]\right)^{-1} \\
&= \hat{J}_k'(\hat{H}_k^{(0)})^{-1}; \\
M^2 &= E\left[\sum_{l=1}^{k}(x_l - Ax_{l-1})(x_l - Ax_{l-1})'\Big| Z_k\right] \\
&= \frac{1}{\delta}[\hat{H}_k^{(0)} - A\hat{H}_k^{(1)} - \hat{H}_k^{(1)}A' + A\hat{H}_k^{(0)}A']; \\
N^2 &= \left[\sum_{l=1}^{k}(Z_l - \delta H x_l)(z_l - \delta H x_l)'\Big| Z_k\right] \\
&= \left(\sum_{l=0}^{k} z_l z_l'\right) - \hat{J}_k'(\delta H)' - \delta H\hat{J}_k + \delta H\hat{H}_k^{(0)}\delta H'.
\end{aligned}
$$

where $H_k^{(M)}$ denotes the matrix with entries $H_k^{ij(M)}$, and J_k denotes the matrix with entries J_k^{ij}.

Chapter IV

Markov Switching Parameter

We now consider the case where the model parameters are functions of a time-varying parameter θ_t. We will again consider a continuous time signal with discrete time observations, but we will make the simplifying assumption that the parameter θ changes value only at discrete times $\delta, 2\delta, 3\delta, \ldots$, for some $\delta > 0$. Our switching parameter $\Theta := \{\theta_0, \theta_1, \theta_2, \ldots\}$ is a discrete time stationary Markov chain taking values in $\{1, 2, \ldots, N\}$, for some $N \in \mathbb{N}$, with initial and transition probabilities

$$p_i = P[\theta_0 = i],$$
$$p_{ij} = P[\theta_{k+1} = j | \theta_k = i].$$

We will use a change of probability measure to obtain a recursive exact filter for the joint contitional density of the parameter θ_t and the signal x_t given the history of observations up to time t. Our filter is infinite dimensional, which is to be expected in light of Bjorg's result, mentioned in Chapter I, that finite dimensional exact filters for switched models exist only in the case where the observation equation does not explicitly involve the signal. The model will be given by

$$dx_t = C(\theta_t)x_t dt + \sigma(\theta_t)dW_t, \; x \in [0, \tau], \; x_0 = 0; \tag{4.1}$$
$$z_k = \delta H(\theta_k)x_{k\delta} + \Delta_k, k \geq 0. \tag{4.2}$$

subject to the following assumpitions:

Assumptions 4.1:

1. In (4.1), we identify the sequence θ with a step function taking values θ_i on respective intervals $[i\delta - \delta, \delta)$.

2. W_t is a standard Brownian motion in $\mathbb{R}^m$;
3. $x_t \in \mathbb{R}^m$ and $z_k \in \mathbb{R}^n$;
4. C, H, and σ are matrix functions of θ of appropriate dimensions;
5. $\Delta_k = R_{k\delta}V_{k\delta} - R_{k\delta-\delta}V_{(k\delta-\delta)}$, where V is a standard Brownian motion in $\mathbb{R}^n$ and R_t is a positive definite $n \times n$ continuous real matrix valued function on $[0, \tau]$;
6. Θ is indepentent of W and V.

Letting $x_k := x_{k\delta}$, (4.1) yields

$$x_{k+1} = \Phi(k\delta + \delta, k\delta, \theta_{k+1})x_k + \int_0^\delta \Phi(k\delta + \delta, \tau, \theta_{k+1})\sigma(\theta_{k+1})\, dW_s. \tag{4.3}$$

Let $A(i) := \Phi(k\delta + \delta, k\delta, \theta_{k+1})$ and let

$$v_k^i := A(i) \int_0^\delta \Phi(k\delta + \delta, \tau, i)\sigma(i)\, dW_s.$$

Then for fixed i, $\{v_k^i, k \geq 1\}$ are *iid* [2]. Let the common distribution of $\{v_k^i, k \geq 1\}$ be called ψ^i. The noise term in (4.3) can now be written

$$\sum_{i=1}^{n} v_{k+1}^i I\{\theta_{k+1} = i\},$$

and hence

$$x_{k+1} = A(\theta_{k+1})x_k + \sum_{i=1}^{n} v_{k+1}^i I\{\theta_{k+1} = i\}.$$

Write $\mathcal{G}_k$ for the complete σ-field generated by $\{\theta_0, \ldots, \theta_k, x_0, \ldots, x_k, z_0, \ldots, z_{k-1}\}$. Let φ_k be the density of Δ_k and let

$$\Lambda_n = \prod_{k=1}^{n} \frac{\varphi(\Delta_k)}{\varphi(z_k)}.$$

Now define a probability measure $\overline{P}$ on $(\Omega, \mathcal{G}_\infty)$ by

$$\left.\frac{d\overline{P}}{dP}\right|_{\mathcal{G}_k} = \Lambda_k.$$

The existence of $\overline{P}$ follows, as before, from Kolmogorov's extension theorem.

The following theorem tells us *a fortiori* that under $\overline{P}$, the sequences $\{x_k\}$ and $\{z_k\}$ are independent.

Theorem 4.2. *Under $\overline{P}$, z_k is independent of $\mathcal{G}_k$ and has density φ_k, $k \leq T/\delta$.*

Proof. For $t \in \mathbb{R}^n$, by the event $\{z_k < t\}$ we mean $\{z_i^i < t^i, i = 1 \cdots n\}$. Then

$$\begin{aligned}
\overline{P}(z_k < t \,|\mathcal{G}_k) &= \overline{E}[I\{z_k < t\}|\mathcal{G}_k] \\
&= \frac{E[\Lambda_k I\{z_k < t\}|\mathcal{G}_k]}{E[\Lambda_k|\mathcal{G}_k]} \\
&= \frac{\Lambda_{k-1} E[\lambda_k I\{z_k < t\}|\mathcal{G}_k]}{\Lambda_{k-1} E[\lambda_k|\mathcal{G}_k]}.
\end{aligned} \tag{4.4}$$

Now,

$$E[\lambda_k|\mathcal{G}_k] = \sum_{\theta_k=1}^{N} \int_{\mathbb{R}^{m+n}} \frac{\varphi_k(\delta H(\theta_k)x_k + \Delta_k)}{\varphi_k(\Delta_k)} \rho(x_k, \Delta_k, \theta_k)\, dx_k\, d\Delta_k$$

where $\rho(\cdot,\cdot,\cdot)$ is the conditional joint density of x_k, Δ_k, and θ_k given $\mathcal{G}_k$. But x_k and θ_k are $\mathcal{G}_k$ measurable and Δ_k is independent of $\mathcal{G}_k$. Hence

$$\begin{aligned}
E[\lambda_k|\mathcal{G}_k] &= \int_{\mathbb{R}^n} \frac{\varphi_k(\delta H(\theta_k)x_k + \Delta_k)}{\varphi_k(\Delta_k)} \varphi_k(\Delta_k) d\Delta_k \\
&= \int_{\mathbb{R}^n} \varphi(\delta H(\theta_k)x_k + \Delta_k) d\Delta_k \\
&= 1.
\end{aligned}$$

Hence, the denominator of (4.4) is equal to Λ_{k-1}. Hence,

$$\begin{aligned}
\overline{P}(z_k < t \,|\mathcal{G}_k) &= \int_{\mathbb{R}^n} \frac{\varphi_k(\delta H(\theta_k)x_k + \Delta_k)}{\varphi_k(\Delta_k)} I(\delta H(\theta_k)x_k + \Delta_k < t)\varphi(\Delta_k)\, d\Delta_k \\
&= \int_{\mathbb{R}_n} \varphi_k(u) I(u < t)\, du \\
&= \int_{u<t} \varphi_k(u)\, du.
\end{aligned}$$

Since this quantity is independent of $\mathcal{G}_k$, the proof is done. ∎

Let $\mathcal{Z}_k = \sigma(z_0, ..., z_k)$ and let $q_k(z, i) = \overline{E}[\Lambda_k^{-1} I\{\theta_k = i\} I\{x_k \in dz\}]$. For any Borel function $f : \mathbb{R}^m \to \mathbb{R}$, we have

$$\overline{E}[\Lambda_k^{-1} f(x_k) I\{\theta_k = i\}|\mathcal{Z}_k] = \int_{R^m} f(z) q_k(z, i) dz.$$

Theorem 4.3. *For $k \geq 1$, q_k satisfies the recursion*

$$q_{k+1}(z,i) = \frac{\varphi_{k+1}(z_{k+1} - \delta H(i)z)}{\varphi_{k+1}(z_{k+1})} \sum_{j=1}^{N} p_{ij} \int_{\mathbb{R}^m} \psi^i(z - A(i)x) q_k(x,j) dx.$$

Proof. Let f be an arbitrary Borel function from $\mathbb{R}^m$ to $\mathbb{R}$. Then

$$\begin{aligned}
&\int_{\mathbb{R}^m} f(z) q_{k+1}(z,i) dz \qquad (4.5)\\
&= \overline{E}[\Lambda_{k+1}^{-1} f(x_{k+1}) I\{\theta_{k+1} = i\} | \mathcal{Z}_{k+1}]\\
&= \frac{1}{\varphi_{k+1}(z_{k+1})} \overline{E}\Big[\Lambda_k^{-1} \varphi_{k+1}(z_{k+1} - \delta H(\theta_{k+1}) x_{k+1})\\
&\qquad \cdot f\left(A(\theta_{k+1}) x_k + \Sigma_{i=1}^N v_{k+1}^i I\{\theta_{k+1} = i\}\right) I\{\theta_{k+1} = i\} | \mathcal{Z}_{k+1}\Big]\\
&= \frac{1}{\varphi_{k+1}(z_{k+1})} \sum_{j=1}^{N} \overline{E}\Big[\Lambda_k^{-1} \varphi_{k+1}\Big(z_{k+1} - \delta H(i)\left(A(i) x_k + v_{k+1}^i\right)\Big)\\
&\qquad \cdot f\left(A(i) x_k + v_{k+1}^i\right) \ I\{\theta_k = j \ \theta_{k+1} = i\} \ \Big| \ \mathcal{Z}_k\Big].
\end{aligned}$$

Letting $z = A(i)x + v$, the above is equal to

$$\frac{1}{\varphi_{k+1}(z_{k+1})} \sum_{j=1}^{N} \int_{\mathbb{R}^m} \int_{\mathbb{R}^m} \varphi_{k+1}(z_{k+1} - \delta H(i) z) f(z) \psi^i(A(i)x - z) q_k(x,j) p_{ji} dx dz.$$

Since f was arbitrary, equating this to (4.0.5) yields the result. ∎

We can use the above recursion to calculate the conditional density p_k of x_k given $\mathcal{Z}_k$ as follows: first, note that

$$\sum_{i=q}^{N} \int_{\mathbb{R}^m} q_k(z,i) dz = \sum_{i=q}^{N} \overline{E}[\Lambda_k^{-1} I\{\theta_k = i\} | \mathcal{Z}_k] = \overline{E}[\Lambda^{-1} | \mathcal{Z}_k].$$

So, by conditional Bayes' theorem,

$$\begin{aligned}
p_k(z) &= E[I\{x_k \in dz\} | \mathcal{Z}_k]\\
&= \frac{\overline{E}[\Lambda_k^{-1} I\{x_k \in dz\} | \mathcal{Z}_k]}{\overline{E}[\lambda_k^{-1} | \mathcal{Z}_k]}\\
&= \frac{\sum_{i=1}^{N} q_k(z,i)}{\sum_{i=1}^{N} \int_{\mathbb{R}^m} q_k(z,i) dz}
\end{aligned}$$

Similarly, the conditional density of θ_k given Z_k is given by

$$P[\theta_k = i|Z_k] = \frac{\int_{\mathbb{R}^m} q_k(z,i)dz}{\sum_{i=1}^{N} \int_{\mathbb{R}^m} q_k(z,i)dz}.$$

Between observations, we have $x_t = \sum_{j=1}^{N} x_t I\{\theta_{k+1} = j\}$. The density thus evolves according to the Kolmogorov forward equation

$$\begin{aligned}
\frac{\partial p}{\partial t} &= \sum_{j=1}^{N} \left\{ -p' Tr(B(j)) - p'_x (B(j)) x + \frac{1}{2} Tr(\sigma^2(j) p_{xx}) \right\} P[\theta_{k+1} = j|Z_k] \\
&= \sum_{j=1}^{N} \left\{ -p' Tr(B(j)) - p'_x (B(j)) x + \frac{1}{2} Tr(\sigma^2(j) p_{xx}) \right\} \sum_{i=1}^{N} P[\theta_k = i|Z_k] p_{ij}.
\end{aligned}$$

Chapter V

Deterministic Switching

The exact filter of the previous chapter is not finite dimensional. In this chapter, we consider an alternate model in which the switching parameter sequence is deterministic (but unknown). We employ a Kalman filter approach using an EM algorithm to estimate the model parameters. Consequently, we obtain finite dimensional filters.

5.1. Model and Change of Measure

Fix $\delta > 0$. We will now consider a model consisting of an $\mathbb{R}^m$ valued signal and $\mathbb{R}^n$ valued observation process

$$dx_t = C(\theta_t)x_t dt + \sigma(\theta_t)dW_t, \ x \in [0,\tau], \ x_0 = 0; \tag{5.1.1}$$

$$z_k = \delta H(\theta_k)x_{k\delta} + \Delta_k, \ \ k \geq 0 \tag{5.1.2}$$

subject to

Assumptions 5.1

1. W, V are independent standard Brownian motions;
2. R is a continuous positive definite matrix valued function on $[0,\tau]$, and
$$\Delta_k := R_{k\delta}V_{k\delta} - R_{k\delta-\delta}V_{k\delta-\delta};$$
3. H, C, σ are matrix valued functions on the finite parameter set $\{1, \dots, N\}$. $\sigma(i)$ is positive definite for each i;
4. $\Theta := \{\theta_1, \theta_2, \dots\}$ is a fixed (non-random) but unknown sequence on $\{1, 2, \dots, N\}$, i.e., $\Theta \in \{1, 2, \dots, N\}^{\mathbb{N}}$;
5. In (5.1.1), we identify the sequence θ with a step function taking values θ_i on respective intervals $[i\delta - \delta, i\delta)$.

Let $x_k = x_{k\delta}$. For $1 \le k \le \tau/\delta$, let $\mathcal{Z}_k$ be the complete sigma field generated by $\{z_1, \dots, z_k\}$ and let $\mathcal{G}_k$ be the complete sigma field generated by $\{x_1, \dots, x_k, z_1, \dots, z_k\}$. Let Φ denote the fundamental matrix associated with the matrix function $C(\theta_t)$. Then for $0 \le k\delta \le \tau$, $\Phi(k\delta - \delta, k\delta)$ depends only on θ_k, and we may write $\Phi(i)$ for the values of the fundamental matrix so generated. For $i = 1, 2, \dots, N$, define

$$
\begin{aligned}
A(i) &:= \Phi(i) \\
B^2(i) &:= A^2(i) \int_0^{\delta} e^{-2C(i)s} \sigma(i)^2 ds, \\
R_k^2 &:= \int_{k\delta-\delta}^{k\delta} R_s^2 ds.
\end{aligned}
$$

Then,

$$
x_{k+1} = A(\theta_{k+1}) x_k + \int_{k\delta}^{k\delta+\delta} \Phi(k\delta + \delta, s)\sigma(\theta_{k+1})\, dW_s.
$$

Hence,

$$
x_{k+1} = A(\theta_{k+1}) x_k + B(\theta_{k+1}) v_{k+1} \tag{5.1.3}
$$

$$
z_k = H(\theta_k) x_k + R_k w_k. \tag{5.1.4}
$$

for two independent sequences $\{v_k\}, \{w_k\}$ of *iid* $N(0, I)$ random variables.

Notation : Given a sequence $\hat{\Theta} = \{\hat{\theta}_1, \hat{\theta}_2, \dots\}$ taking values in $\{1, \dots, N\}$, we will write

$$
\begin{aligned}
\hat{A}_k &:= \Phi(\theta_k) \\
\hat{B}_k^2 &:= \int_{k\delta-\delta}^{k\delta} \left(Phi(k\delta, s)\sigma(\hat{\theta}_k)\right)^2 ds \\
\hat{H}_k &:= H(\hat{\theta}_k) \\
\hat{R}_k^2 &:= R_k
\end{aligned}
$$

Let ψ, ϕ be the standard normal densities in m, n dimensions respectively and let

$$
\lambda_0(\hat{\Theta}) := \frac{\phi\left(\hat{R}_0^{-1}(z_0 - \hat{H}_0 x_0)\right)}{|\hat{R}_0|\phi(z_0)},
$$

$$\lambda_k(\hat{\Theta}) := \frac{\phi\left((\hat{R}_k)^{-1}(z_k - \hat{H}_k x_k)\right)}{|\hat{R}_k|\phi(z_k)} \frac{\psi\left(\hat{B}_k^{-1}(x_k - \hat{A}_k x_{k-1})\right)}{|\hat{B}_k|\psi(x_k)},$$

$$\Lambda_n(\hat{\Theta}) := \prod_{k=1}^{n} \lambda_k(\hat{\Theta}).$$

Let $\overline{P}$ be a probability measure under which $\{x_k\}$ and $\{y_k\}$ are sequences of *iid* $N(0, I)$ random variables. For each possible choice of $\hat{\Theta}$, we can define a probability measure $P(\hat{\Theta})$ by

$$\left.\frac{dP(\hat{\Theta})}{d\overline{P}}\right|_{\mathcal{G}_k} = \Lambda_k(\hat{\Theta}).$$

We recall a theorem from [15]:

Theorem 5.2. *Let $\hat{\Theta} = \{\hat{\Theta}_1, \hat{\Theta}_2, \dots\}$ be a sequence on $\{1, 2, \dots, N\}$. Define a probability measure $\hat{P}$ by $d\hat{P}/d\overline{P}|_{\mathcal{G}_k} = \Lambda_k(\hat{\Theta})$. Then under $\hat{P}$, the sequences $\{x_k\}$ and $\{y_k\}$ satisfy*

$$x_{k+1} = \hat{A}_{k+1} x_k + \hat{B}_{k+1} \hat{v}_{k+1}, \tag{5.1.5}$$

$$z_k = \hat{H}_k x_k + \hat{R}_k \hat{w}_k, \tag{5.1.6}$$

for two sequences $\{\hat{v}_k\}, \{\hat{w}_k\}$ which, under $\hat{P}$, are iid $N(0, I)$.

Proof. Let $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^m \to \mathbb{R}$ be arbitrary Borel measurable functions, and let $\hat{E}$ (respectively $\overline{E}$) denote the expectation under $\hat{P}$ (respectively $\overline{P}$). We then have

$$\hat{E}[g(w_k)f(v_k)|\mathcal{G}_{k-1}] = \frac{\overline{E}[\Lambda_k g(w_k)f(v_k)|\mathcal{G}_{k-1}]}{\overline{E}[\Lambda_k|\mathcal{G}_{k-1}]}$$

by abstract Bayes' Theorem.

Now Λ_{k-1} is Z_{k-1} measurable. Therefore,

$$\hat{E}[g(w_k)f(v_k)|\mathcal{G}_{k-1}] = \frac{\overline{E}[\lambda_k g(w_k)f(v_k)|\mathcal{G}_{k-1}]}{\overline{E}[\lambda_k|\mathcal{G}_{k-1}]}.$$

However,

$$\overline{E}[\lambda_k|\mathcal{G}_{k-1}] = \frac{1}{|\hat{B}_k|} \int_{\mathbb{R}^m} \psi(\hat{B}_k^{-1}(x_k - \hat{A}_k x_{k-1}))dx_k = 1.$$

Consequently,

$$\begin{aligned}\hat{E}[g(w_k)f(v_k) \mid \mathcal{G}_{k-1}] &= \overline{E}[\lambda_k\, g(w_k)f(v_k) \mid \mathcal{G}_{k-1}] \\ &= \overline{E}\big[\lambda_k\, g(\hat{B}_k^{-1}(x_k - \hat{A}_k x_{k-1}))f(\hat{R}_k^{-1}(z_k - \hat{H}_k x_k)) \,\big|\mathcal{G}_{k-1}\big] \\ &= \int_{\mathbb{R}^n} \varphi(v)f(v)\, dv \;\times\; \int_{\mathbb{R}^m} \psi(w)g(w)\, dw,\end{aligned}$$

and the theorem is proved. ∎

5.2. EM Algorithm

The conditional mean and variance of x_k given $\mathcal{Z}_k$ (under the 'real world' measure P) can be computed using Theorem 2.5 if the parameters of the model are known. We will estimate the parameters using the EM algorithm that follows. For the remainder of this chapter, we will leave T fixed and assume we have already computed estimates $\theta_1^*, \theta_2^*, \ldots, \theta_{T-1}^*$ for respective values of $\theta_1, \ldots, \theta_{T-1}$.

Suppose we have given a hypothetical parameter sequence $\hat{\Theta}$ of length T on $\{1, \ldots, N\}$, whose first $T-1$ terms are $\{\theta_1^*, \theta_2^*, \ldots, \theta_{T-1}^*\}$, and define

$$\mathcal{L}(\hat{\Theta}) := E\left[\frac{dP(\hat{\Theta})}{dP} \,\Big|\mathcal{Z}_T\right].$$

We will define a maximum likelihood (ML) estimate of (the first T terms of) Θ to be one which makes $\mathcal{L}(\hat{\Theta})$ as large as possible. Since the expression for $\mathcal{L}$ involves the unknown parameter sequence Θ, we cannot directly compute the ML estimate. This is why an iterative technique (EM algorithm, in our approach) is necessary.

For two parameter sets $\hat{\Theta}, \hat{\Theta}$, define

$$Q(\hat{\Theta}, \hat{\Theta}) := E_{\hat{\Theta}}\left[\log\left(\frac{dP(\hat{\Theta})}{dP(\hat{\Theta})}\right) \;\Big|\; \mathcal{Z}_k\right]. \tag{5.2.1}$$

Our EM algorithm goes as follows:

Step 1) Input z_T. Set $k = 0$.

Step 2) Find an initial estimate $\hat{\theta}^0$ for θ_T, (see Remarks below).

Step 3) Set $\hat{\Theta}^k = \{\theta_1^*, \theta_2^*, \ldots, \theta_{T-1}^*, \hat{\theta}^k\}$. (Assuming that for $k > 0$, $\hat{\theta}^k$ has been calculated in Step 6 of the previous iteration.)

Step 4) Using Theorem 2.5, obtain estimates for the conditional mean $\hat{x}_T$ and covariance $\hat{P}_T$ of x_T given Z_T under $P(\hat{\Theta}^k)$.

Step 5) For each possible value i of θ_T, set $\hat{\Theta}^i = \{\theta_1^*, \theta_2^*, \ldots, \theta_{T-1}^*, i\}$ and compute $Q(\hat{\Theta}^i, \hat{\Theta}^k)$.

Step 6) Set $\hat{\theta}^{k+1}$ equal to a value of i which yields a maximal Q in Step 5.

Step 7) If a stopping criterion has been reached (see Remarks below), output $\theta_T^* = \hat{\theta}^{k+1}$ and terminate the EM algorithm. Otherwise, set k to $k+1$ and go to Step 3.

Remarks:

1. The initial estimate of θ_T in Step 2 depends on the application. In many tracking applications, the mode parameter θ changes infrequently compared to the observation rate. Hence it is reasonable in at least these cases to take θ_{T-1}^* as the initial estimate for θ_T.
2. By Jensen's inequality, we have

$$\log \mathcal{L}(\hat{\Theta}^{k+1}) - \log \mathcal{L}(\hat{\Theta}^k) \geq Q(\hat{\Theta}^{k+1}, \hat{\Theta}^k) \geq 0,$$

   with equality if and only if $\hat{\Theta}^{k+1} = \hat{\Theta}^k$. Hence, the generated sequence $\{\hat{\Theta}^k, k \geq 0\}$ gives a nondecreasing sequence of values of $\mathcal{L}$. Also, once leaving a value of $\hat{\Theta}$, the algorithm will never return to that value. Thus, we will terminate the algorithm when $\hat{\Theta}^{k+1} = \hat{\Theta}^k$. Since there are only N different values $\hat{\Theta}^{k+1}$ may take, the algorithm will terminate in at most $N+1$ iterations.

E-step

Step 5 in the above algorithm has only a little to say, but a lot to do. We now discuss the computation of $Q(\cdot, \cdot)$.

Let $\tilde{\Theta}, \hat{\Theta}$ be two sequences on $\{1, \ldots, N\}$ of length T which differ only in the T^{th} position. Let $\tilde{P}$, $\hat{P}$ denote the probability measures $P(\tilde{\Theta}), P(\hat{\Theta})$ respectively, and let $\tilde{A}_k, \tilde{B}_k, \tilde{H}_k, \tilde{R}_k$ denote $A(\tilde{\theta}_k), B(\tilde{\theta}_k), H(\tilde{\theta}_k), R(\tilde{\theta}_k)$ respectively, where $\tilde{\theta}_k$ is the k^{th} entry in $\tilde{\Theta}$. Then we have

$$\frac{d\hat{P}}{d\tilde{P}}\Big|_{\mathcal{G}_T} = \frac{d\hat{P}/d\overline{P}}{d\tilde{P}/d\overline{P}}\Big|_{\mathcal{G}_T} = \frac{\hat{\Lambda}_T}{\tilde{\Lambda}_T}.$$

We will see that the expression for $Q(\tilde{\Theta}, \hat{\Theta})$ is made more manageable by introducing the following notation:

$$\begin{aligned}
a_T^{ij} &:= x_T^i x_T^j, \\
b_T^{ij} &:= x_T^i x_{T-1}^j, \\
c_T^{ij} &:= x_{T-1}^i x_{T-1}^j, \\
\hat{a}_T^{ij} &:= \hat{E}[a_T^{ij} | \mathcal{Z}_T], \\
\hat{b}_T^{ij} &:= \hat{E}[b_T^{ij} | \mathcal{Z}_T], \\
\hat{c}_T^{ij} &:= \hat{E}[c_T^{ij} | \mathcal{Z}_T].
\end{aligned}$$

Now,

$$\begin{aligned}
Q(\tilde{\Theta}, \hat{\Theta}) &= E_{\hat{\Theta}}\left[\log \frac{dP(\tilde{\Theta})}{dP(\hat{\Theta})} \mid \mathcal{Z}_T\right] \\
&= \hat{E}\left[\log \frac{\Lambda_T(\tilde{\Theta})}{\Lambda_T(\hat{\Theta})} \mid \mathcal{Z}_T\right] \\
&= \hat{E}[-\log|\tilde{B}_T| - \log|\tilde{R}_T| - \frac{1}{2}(x_T - \tilde{A}_T(x_{T-1})'\tilde{B}_T^{-2}(x_T - A_T x_{T-1}) \\
&\qquad - \frac{1}{2}(z_T - \tilde{H}_T x_T)'\tilde{R}^{-2}(z_T - \tilde{H}_T x_T) + f(\hat{\Theta}_k) \mid \mathcal{Z}_T].
\end{aligned}$$

Where $f(\hat{\Theta}_k)$ does not depend on $\tilde{\Theta}$. Also,

$$\begin{aligned}
&(x_T - \tilde{A}_T x_{T-1})'\tilde{B}_T^{-2}(x_T - \tilde{A}_T x_{T-1}) \\
&= \sum_{i,j}\left\{x_T^i x_T^j [\tilde{B}_T^{-2}]_{ij} - 2x_{T-1}^i x_T^j [\tilde{A}_T' \tilde{B}_T^{-2}]_{ij} + x_{T-1}^i x_{T-1}^j [\tilde{A}_T' \tilde{B}_T^{-2} \tilde{A}_T]_{ij}\right\},
\end{aligned}$$

and

$$(z_T - \hat{H}_T x_T)' \hat{R}_T^{-2} (z_T - \hat{H}_T x_T)$$
$$= \sum_{i,j} \left\{ z_T^i z_T^j [\hat{R}_T^{-2}]_{ij} - 2 z_T^j x_T^i [\hat{H}_T' \hat{R}_T^{-2}]_{ij} + x_T^i x_T^j [\hat{H}_T' \hat{R}_T^{-2} \hat{H}]_{ij} \right\}.$$

Hence we have

$$Q(\hat{\Theta}, \hat{\Theta}) \tag{5.2.2}$$
$$= -\log |\hat{B}_T| - \log |\hat{R}_T|$$
$$- \frac{1}{2} \sum_{i,j} \left\{ \hat{a}_T^{ij} [\hat{B}_T^{-2}]_{ij} - 2 \hat{b}_T^{ji} [\hat{A}_T' \hat{B}_T^{-2}]_{ij} + \hat{c}_T^{ij} [\hat{A}_T' \hat{B}_T^{-2} \hat{A}_T]_{ij} \right.$$
$$\left. + z_T^i z_T^j [\hat{R}_T^{-2}]_{ij} - 2 z_T^j \hat{x}_T^i [\hat{H}_T' \hat{R}_T^{-2}]_{ij} + \hat{a}_T^{ij} [\hat{H}_T' \hat{R}_T^{-2} \hat{H}]_{ij} \right\}.$$

5.3. Filters

We see from (5.2.2) that to execute the E-step of our algorithm, we need to compute $\hat{a}, \hat{b}$, and $\hat{c}$. This section gives finite dimensional exact filters for doing so.

Define

$$\beta_T^{ij(\cdot)}(x) := \overline{E} \left[\hat{\Lambda}_T [(\cdot)_T^{ij}] I\{x_T \in dx\} \mid \mathcal{Z}_T \right]$$

where $(\cdot)$ is a, b, or c and $\overline{E}$ is the expectation under $\overline{P}$.

Let

$$\alpha_T(x) := \overline{E}[\hat{\Lambda}_T I\{x_T \in dx\} | \mathcal{Z}_T].$$

Then for every Borel measurable function f,

$$\overline{E}[\hat{\Lambda}_T f(x_T) | \mathcal{Z}_T] = \int_{\mathbb{R}^m} \alpha_T(x) f(x) \, dx, \tag{5.3.1}$$
$$\overline{E} \left[\hat{\Lambda}_T [(\cdot)_T^{ij}] f(x_T) | \mathcal{Z}_T \right] = \int_{\mathbb{R}^m} \beta_T^{ij(\cdot)}(x) \, dx. \tag{5.3.2}$$

Let

$$\nu_T(z_T, x) := \frac{\varphi\left(\hat{R}_T^{-1} (z_T - \hat{H}_T x) \right)}{|\hat{B}_T| |\hat{R}_T| \varphi(z_T)},$$
$$\rho_T(x, z) := \psi\left(\hat{B}_T^{-1} (x - \hat{A}_T z) \right).$$

Theorem 5.3. *The unnormalized conditional densities α and $\beta^{(\cdot)}$ are given inductively as follows:*

$$\alpha_T(x) = \nu_T(z_T, x) \int_{\mathbb{R}^m} \alpha_{T-1}(z)\rho_T(x, z)\, dz \tag{5.3.3}$$

$$\beta_T^{ij(a)}(x) = \nu_T(z_T, x) x^i x^j \int_{\mathbb{R}^m} \alpha_{T-1}(z)\rho_T(x, z)\, dz \tag{5.3.4}$$

$$\beta_T^{ij(b)}(x) = \nu_T(z_T, x) x^i \int_{\mathbb{R}^m} z^j \alpha_{T-1}(z)\rho_T(x, z)\, dz \tag{5.3.5}$$

$$\beta_T^{ij(c)}(x) = \nu_T(z_T, x) \int_{\mathbb{R}^m} z^i z^j \alpha_{T-1}(z)\rho_T(x, z)\, dz \tag{5.3.6}$$

Proof. We prove (5.3.4). The proofs of (5.3.3), (5.3.5), and (5.3.6))) are similar. Let f be an arbitrary Borel measurable function from $\mathbb{R}^m$ to $\mathbb{R}$. Then

$$\begin{aligned}
&\overline{E}[\hat{\Lambda}_T a_T^{ij} f(x_T) \mid \mathcal{Z}_T] \\
&= \overline{E}[\hat{\Lambda}_{T-1}\hat{\lambda}_T x_T^i x_T^j f(x_T) \mid \mathcal{Z}_T] \\
&= \overline{E}[\hat{\Lambda}_{T-1}\nu_T(z_T, x_T)\rho_T(x_T, x_{T-1})\frac{1}{\psi(x_T)} x_T^i x_T^j f(x_T) \mid \mathcal{Z}_T] \\
&= \overline{E}[\hat{\Lambda}_{T-1} \int_{\mathbb{R}^m} \nu_T(z_T, x)\rho_T(x, x_{T-1}) x_T^i x_T^j f(x) dx \mid \mathcal{Z}_{T-1}] \\
&= \int_{\mathbb{R}^m}\int_{\mathbb{R}^m} \alpha_{T-1}(z)\nu_T(z_T, x)\rho_T(x, z) f(x) dx dz
\end{aligned} \tag{5.3.7}$$

Since f is an arbitrary Borel measurable function, equating (5.3.7) with the RHS of (5.3.2) yields (5.3.4). ∎

Notation: Let

$$\begin{aligned}
\hat{x}_l &:= \hat{E}[x_l \mid \mathcal{Z}_l], \\
\hat{P}_l &:= \hat{E}[(x_l - \hat{x}_l)(x_l - \hat{x}_l)' \mid \mathcal{Z}_l], \\
\sigma_l &:= \hat{A}_l \hat{B}_l^{-2} \hat{A}_l + \hat{P}_{l-1}^{-1}, \\
\delta_l &:= \delta_l(x) = 2(x' \hat{B}_l^{-2} \hat{A}_l + \hat{x}_{l-1} \hat{P}_{l-1}^{-1})'.
\end{aligned}$$

We now recall a computation from [15].

Lemma 5.4. (Elliott and Krishnamurthy) *Let $a \in \mathbb{R}, b \in \mathbb{R}^m, c \in \mathbb{R}^{m\times m}$. Then for all $l \le T$ and all $x \in \mathbb{R}^m$ we have*

$$\begin{aligned}\nu_l(z_l, x) \int_{\mathbb{R}^m} \alpha_{l-1}(z)\rho_l(x,z)[a + b'z + z'cz]dz \\ = \alpha_l(x)\left[a + \frac{1}{2}b'\sigma_l^{-1}\delta_l + \Sigma_{p,q}c(p,q)\sigma_l^{-1}(p,q) + \frac{1}{4}\delta_l'\sigma_l^{-1}c_l\sigma_l^{-1}\delta_l\right].\end{aligned}$$

Lemma 5.5. *Let Y be a random variable on Ω such that for some a, b, c in $\mathbb{R}, \mathbb{R}^m, \mathbb{R}^{m\times m}$, respectively, we have, for all $x \in \mathbb{R}^m$,*

$$\overline{E}[\hat{\Lambda}_T Y I\{x_T \in dx\} \mid \mathcal{Z}_T] = (a + b'x + x'cx)\alpha_T(x).$$

Then the conditional expectation of Y given $\mathcal{Z}_T$ is given by

$$\hat{E}[Y|\mathcal{Z}_T] = a + b'\hat{x}_T + \sum_{p,q} c(p,q)\hat{P}_T(p,q) + \hat{x}_T'c\hat{x}_T.$$

Proof. Using abstract Bayes' rule we have

$$\hat{E}[Y|\mathcal{Z}_T] = \frac{\overline{E}[\hat{\Lambda}_T Y \mid \mathcal{Z}_T]}{\overline{E}[\hat{\Lambda}_T \mid \mathcal{Z}_T]} = \frac{\int_{\mathbb{R}^m}(a + b'x + x'cx)\alpha_T(x)dx}{\int_{\mathbb{R}^m}\alpha_T(x)}. \tag{5.3.8}$$

But since $\alpha_T(x)$ is an unnormalized conditional density of x_T given $\mathcal{Z}_T$, the numerator of the far RHS in the above is equal to

$$\begin{aligned}&\int_{\mathbb{R}^m} \alpha_T(x)dx \times \hat{E}[a + b'x_T + x'cx_T] \\ &= \int_{\mathbb{R}^m} \alpha_T(x)dx \times \left[a + b'\hat{x}_T + \sum_{p,q} c(p,q)\hat{P}_T(p,q) + \hat{x}_T'c\hat{x}_T\right]\end{aligned}$$

Substituting this in (5.3.8) proves the result. ∎

The following theorems give us the means to calculate $\hat{a}$, $\hat{b}$, and $\hat{c}$, and hence to execute Step 5 of our EM algorithm.

Theorem 5.6.

$$\hat{E}[a_T^{ij}|Z_T] = \hat{P}_T(i,j) + \hat{x}_T^i \hat{x}_T^j.$$

Proof. From Theorem 5.3, we have

$$\hat{E}[\hat{\Lambda}_T a_T^{ij} I\{x_T \in dx\}|Z_T] =: \beta_T^{ij(a)}(x) = x^i x^j \alpha_T(x)$$

Now, by Lemma 5.5 with $a = 0$, $b = 0_{m\times 1}$ and $c = e_i e_j'$,

$$\begin{aligned}\hat{a}^{ij} &= \hat{P}_T(i,j) + \hat{x}^i \hat{x}^j \\ &= \hat{P}_T(i,j) + \hat{x}_T' e_i e_j' \hat{x}_T.\end{aligned}$$

∎

Toward finding the filters for b and c, it will be useful to introduce the notation

$$\begin{aligned}F_k &:= B_k^{-2} A_k \sigma_{k+1}^{-1}, \\ G_{k+1} &:= \sigma_{k+1}^{-1} \hat{P}_k^{-1} \hat{x}_k.\end{aligned}$$

Here now is the filter for $\hat{b}$:

Theorem 5.7.

$$\begin{aligned}\hat{E}[b_T^{ij}|Z_T] = G_T' e_i e_j' \hat{x}_T + \sum_{p,q} [e_i e_j' F_T'](p,q) \hat{P}_T(p,q) \\ + \hat{x}_T e_i e_j F_T' \hat{x}_T]\end{aligned}$$

Proof. From Theorem 5.3, we have

$$\beta_T^{ij(b)}(x) = \nu_T(z_T, x) x^i \int_{\mathbb{R}^m} z^j \alpha_{T-1}(z) \rho_T(x,z) dz$$

Now, by Theorem 5.4 with $a = 0, b = e_j, c = 0_{m\times m}$, we have

$$\begin{aligned}\beta_T^{ij(b)}(x) &= \alpha_T(x)[\frac{1}{2} e_j' \sigma_T^{-1} \delta_T] \\ &= \alpha_T(x)[x' e_i e_j' \sigma_T^{-1} (x' \hat{B}_T^{-2} \hat{A}_T + \hat{x}_{T-1}' \hat{P}_{T-1}^{-1})'] \\ &= \alpha_T(x)[x' e_i e_j' \sigma_T^{-1} \hat{A}_T' \hat{B}_T^{-2} x + x' e_i e_j' \sigma_T^{-1} \hat{P}_{T-1}^{-1} \hat{x}_{T-1}] \\ &= \alpha_T(x)[x' e_i e_j' \sigma_T^{-1} \hat{A}_T' \hat{B}_T^{-2} x + \hat{x}_{T-1}' \hat{P}_{T-1}^{-1} \sigma_T^{-1} e_j e_i' x].\end{aligned}$$

Now, invoking Theorem 5.5 with

$$a = 0$$

$$b' = \hat{z}'_{T-1}\hat{P}^{-1}_{T-1}\sigma^{-1}_T e_i e'_j = G'_T e_i e'_j$$

$$c = e_i e'_j \sigma^{-1}_T \hat{A}'_T \hat{B}^{-2}_T = e_i e'_j F'_T$$

yields the result. ∎

Finally, the filter for $\hat{c}$ is given in the following

Theorem 5.8. *Let*

$$S_{ij} := \frac{1}{2}(e_i ej' + e_j e'_i).$$

Then

$$\hat{E}[c^{ij}_T | \mathcal{Z}_T] = L + M'\hat{x}_T + \hat{x}_T N \hat{x}_T + \sum_{p,q} N(p,q)\hat{P}_T(p,q),$$

where

$$L = \sigma^{-1}_T(i,j) + G'_T S_{ij} G_T \ \in \mathbb{R},$$

$$M = 2G'_T S_{ij} F_T \in \mathbb{R}^m,$$

$$N = F_T S_{ij} F'_T \in \mathbb{R}^{m \times m}.$$

Proof. From Theorem 5.3,

$$\beta^{ij(c)}_T(x) = \nu_T(z_T, x)\int_{\mathbb{R}^m} z^i z^j \alpha_{T-1}(z)\rho_T(x,z)dz.$$

Now, by Lemma 5.4 with $a = 0, b = 0_{m\times 1}$, and $c = S_{ij}$, we have

$$\begin{aligned}
\beta^{ij(c)}_T(x) &= \alpha_T(x)\left[\frac{1}{4}\delta'_T\sigma^{-1}_T S_{ij}\sigma^{-1}_T\delta_T + \sigma^{-1}_T(i,j)\right] \\
&= \alpha_T(x)[(x'\hat{B}^{-2}_T\hat{A}_T + \hat{z}'_{T-1}\hat{P}^{-1}_{T-1})\sigma^{-1}_T S_{ij}\sigma^{-1}_T(\hat{A}'_T\hat{B}^{-2}_T x + \hat{P}^{-1}_{T-1}\hat{z}_{T-1}) \\
&\qquad + \sigma^{-1}_T(i,j)] \\
&= \alpha_T(x)[\sigma^{-1}_T(i,j) + x'\hat{B}^{-2}_T\hat{A}_T\sigma^{-1}_T S_{ij}\sigma^{-1}_T\hat{A}_T\hat{B}^{-2}_T x \\
&\qquad + 2x'(\hat{B}^{-2}_T\hat{A}_T\sigma^{-1}_T S_{ij}\sigma^{-1}_T\hat{P}^{-1}_{T-1}\hat{z}_{T-1}) \\
&\qquad + \hat{z}'_{T-1}\hat{P}^{-1}_{T-1}\sigma^{-1}_T S_{ij}\sigma^{-1}_T\hat{P}^{-1}_{T-1}\hat{z}_{T-1}].
\end{aligned}$$

Now, Lemma 5.5 with $a = L, b' = M, c = N$ proves the claim. ∎

For the sake of completeness, we now examine the behavior of the above filters as the time δ between observations approaches 0.

Notation: For any sequence $\{r_k, 0 \le k \le T/\delta\}$, let ${}^\delta r$ denote the *cadlag* step function on [0,T] taking values r_k on respective intervals $[k\delta, \max\{k\delta + \delta, T\})$.

Proposition 5.9. *As $\delta \downarrow$, the functions ${}^\delta\tilde{b}_t^{ij}$, ${}^\delta\hat{b}_t^{ij}$, and ${}^\delta\tilde{c}_t^{ij}$ converge in the ucp topology to the common limit*

$$\hat{P}_t(i,j) + \hat{x}_t^i \hat{x}_t^j$$

for all i and j.

Proof. As in Chapter III, we have

$$
\begin{aligned}
{}^\delta F_t &\to I_{m \times m}, \\
{}^\delta G_t &\to 0_{m \times 1}, \\
{}^\delta \sigma_t &\to 0_{m \times m},
\end{aligned}
$$

as $\delta \downarrow 0$, all uniformly in t. Also, since Assumptions 2.1 hold, Theorem 2.6 guarantees that

$$
\begin{aligned}
{}^\delta \hat{x}_t &\to \hat{x}_t, \\
{}^\delta \hat{P}_t &\to \hat{P}_t,
\end{aligned}
$$

respectively in *ucp* and uniformly in t, as $\delta \downarrow 0$. The claim now follows from the above, together with Theorems 5.6, 5.7, and 5.8. ∎

References

[1] B. Anderson and J. B. Moore. *Optimal Filtering*. Prentice-Hall Inc., New Jersey, (1979).

[2] L. Arnold. *Stochastic Differential Equations*. Krieger Publishing, Malabar, Florida (1974).

[3] Y. Bar-Shalom and X.R. Li. *Estimation and Tracking: Principles, Techniques, and Software*. Artech House, Boston, (1993).

[4] M. Basseville and I. Nikiforov. *Detection of Abrupt Changes: Theory and Application*. Prentice-Hall, Englewood Cliffs, NJ., (1993).

[5] L.E. Baum, T. Petrie, G. Soules, and N. Weiss. "A Maximization Technique occuring in the Statistical Analysis of Probabilistic Functions of Markov Chains," *Annals of the Institute of Statistical Mathematics*, **41** (1970), 164-171.

[6] T. Bjork. "Finite Dimensional Filters for a Class of Nonlinear Diffusions with Jumping Parameters." *Stochastics*, **6** (1982), 121-138.

[7] H.A.P. Blom and Y. Bar-Shalom. "The Interacting Multiple Model Algorithm for Systems with Markovian Switching Coefficients," *IEEE Transactions on Auromatic Control*, **33** (1988), 780-783.

[8] R.G. Brown. "Integrated Navigation Systems and Kalman Filtering: A Perspective," *Navigation: Journal of the Innstitute of Navigation*, **19** (1973), 355-362.

[9] A.K. Caglayan."Simultaneous Failure Detection and Estimation in Linear System" in *Proc. 19th IEEE Conf. Decision and Control*, Albuquerque, NM, 1980.

[10] F. E. Daum. "New Exact Nonlinear Filters," in *Bayesian Analysis of Time Series and Dynamic Models*. Editor: J.C Spall, Marcell Decker Inc., New York, (1988).

[11] A.P. Dempster, N.M Laird, and D.B. Rubin. "Maximum Likelihood from Incomplete Data via the EM Algorithm," *J. Royal Stat. Soc. of London B*, **39** (1971), pp. 1-38.

[12] R. J. Elliott, L. Aggoun, and J.B. Moore. *Hidden Markov Models*. Springer-Verlag, New York, (1995).

[13] R.J Elliott, F. Dufour, and F. Sworder. "Exact Hybrid Filters in Discrete Time," *IEEE Transactions on Automatic Control*, **41** (1996), 1807-1810.

[14] R.J. Elliott and V. Krishnamurthy. "New finite Dimensional Filters for Parameter Estimation of Continuous-time Linear Gaussian Models," preprint, (1996)

[15] R.J. Elliott and V. Krishnamurthy. "New finite Dimensional Filters for Parameter Estimation of Discrete-time Linear Gaussian Models," preprint, (1996)

[16] M. S. Grewal and A.P. Andrews. *Kalman Filtering*, Prentice-Hall, Inc. New Jersey, (1993).

[17] A.H. Jazwinski. *Stochastic Processes and Filtering Theory*, Academic Press, New York, (1970).

[18] D Kannan: *Stochastic Processes*, North-Holland Publ., New York, 1980.

[19] D. Kannan, J.N. Rushton, and Q. Zhang. "Exact finite dimensional filters for mixed-time models." submitted.

[20] M. A. Kouritzin "On Exact Filters for Continuous Signals with Discrete Observations," preprint, (1996)

[21] K. Lange. "A Gradient Algorithm Locally equivalent to the EM Algorithm," *J.R Statist. Soc. B* **57**, (1995), pp 425-437 .

[22] E. Mazor, A. Averbuch, Y. Bar-Shalom, and J. Dayan, "Interacting Multiple Model Methods in Multitarget-Multisensor Tracking: A Survey," *IEEE Trans. Aerospace and Electronic Systems*, submitted (1995).

[23] L.A. Mcgee and S.F Schmidt. "Discovery of the Kalman Filter as a Practical Tool for Aerospace and Industry," NASA Technical Memorandum 86847, November 1985.

[24] R.L Moose, H.F. VanLandingham, and V.Z. Zwick. "Digital Set Point Control of Nonlinear Stochastic Systems," *IEEE Trans Indust. Electron. and Control Instrum*, **25** (1978), 39-45.

[25] Philip Protter. *Stochastic Integration and Differential Equations*. Springer-Verlag, New York, (1990).

[26] C.S Simms and M. R. D'mello. "Adaptive Deconvolution of Seismic Signals," *IEEE Trans. Geoscience Electronics*, **16** (1978), pp 99-103.

[27] A. Skeppstedt, L. Ljung, and M Millnert. "Construction of Composite Models from Observed Data," *International Journal of Control*, **55** (1992), pp 141-152.

[28] W.M. Wonham. "Some Applications for Stochastic Differential Equations to Optimal Nonlinear Filtering," *SIAM Journal on Control*, **2** (1965), pp 347-369.