# INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

**The quality of this reproduction is dependent upon the quality of the copy submitted.** Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

i.. ..

Topics in nonlinear filtering

Fan, Kaisheng, Ph.D.

Iowa State University, 1992

# U·M·I

# Topics in nonlinear filtering

by

Kaisheng Fan

A Dissertation Submitted to the

Graduate Faculty in Partial Fulfillment of the

Requirements for the Degree of

DOCTOR OF PHILOSOPHY

Department: Mathematics
Major: Applied Mathematics

Approved:

In Charge of Major Work

For the Major Department

For the Graduate College

Iowa State University
Ames, Iowa
1992

# TABLE OF CONTENTS

# GENERAL INTRODUCTION

Filtering is the general theory of extracting information about a prescribed quantity of interest from noisy observations. It concerns "estimating" something about an unobserved stochastic process $\{X_t\}$ given observations of a related process $\{Y_t\}$; the classic problem is to calculate, for each t, the conditional distribution of $X_t$ given $\{Y_s, 0 \leq s \leq t\}$. This was solved in the context of linear systems theory by Kalman and Bucy [5], [6] in 1960, 1961 and the resulting well-known Kalman and Kalman-Bucy filter for respectively discrete and continuous time processes have found wide spread applications. A generalization of the linear stochastic filtering problem was soon made to systems with nonlinear dynamics. This is an essentially more difficult problem, being in general infinite-dimensional, but equations describing the evolution of the conditional distribution were obtained by several authors in mid-sixties, for example Bucy [1], Kushner [7], Shiryaev [8], Stratonovich [9], Woham [10]. In 1969 Zakai [11] obtained these equations in a substantially simpler form using the so-called "reference probability" method. In 1968 Kailath [4] introduced the "innovations approach" to linear filtering, and the significance for nonlinear filtering was immediately appreciated [2]. The definitive treatment in the context of martingale theory was given in 1972 by Fujisaki, Kallianpur and Kunita [3]. They obtained the "basic" stochastic differential equation of filtering theory, which, in principle, gave a

complete solution of the nonlinear filtering problem. In the late seventies and during the eighties, most work on nonlinear filtering had concentrated on the areas: rigorous formulation of the theory of stochastic partial differential equations, Lie algebraic aspects and discovery of finite dimensional nonlinear filters, development of "robust" or "pathwise" solutions of the filtering equations and the implementation of nonlinear filtering algorithms.

Recently, the implementation of nonlinear filtering algorithms has received considerable attention both from a theoretical and a practical viewpoint. Due to technological advances, heavy computational burdens can be carried out in real time and people from applied fields, such as communications, radar image analysis, sonar tracing, flight control , now demand to have available efficient filtering algoritms. During the last three decades linear filtering algorithms, based on the Kalman-Bucy equations, have enjoyed immense success in a wide variety of applications and have been implemented in many systems. However, for nonlinear filtering there exists a wide gap between the well developed theory and the design of efficient algorithms. The problem is that the design of useful numerical algorithms is complicated by the mathematical complexity of the filter equations, the generic nonexistence of finite dimensional filters, and the lack of convergence results and error estimates for the implemented algorithmic schemes with respect to the true (optimal) filter.

In this dissertation, we study the implementation of nonlinear filtering algorithms that can be used in real time applications. This requires the use and development of a broad range of techniques in theoretical and applied probability and in systems theory.

In order to implement a filtering algorithm, one has to discretize the state space,

the observation space and the time interval. If one discretizes the observation space first, the corresponding equations for the optimal filter are considerably less complicated than in the diffusion case. This is the starting point of our method. A brief introduction to our procedure to solve the nonlinear fitering problem under consideration is given next.

First, we focus on the development of a general procedure to solve the filtering problem for Markov semimartingale state processes and jump observation processes. Consider a partially observable system $(X_t, Y_t)$, $t \geq 0$, where $X_t$ is a d-dimensional state process and $Y_t$ is a p-dimensional observable component, respectively. We assume that the joint process $\xi_t = (X_t, Y_t), t \geq 0$, has the structure of a general Itó process in $\Re^m$ (m=n+p), i.e., $\xi_t$ has the representation:

$$\xi_t = \xi_0 + \int_0^t b(\xi_s)ds + \int_0^t c(\xi_s)dW_s + \int_0^t \int_Z K(\xi_{s-}, z)(N(ds, dz) - ds\nu(dz))$$

where $W_t$ is a standard m-dimensional Wiener process and N is a standard Poisson random measure on $\Re_+ \times \Re$ with mean measure $ds\nu(dz)$. Besides the usual conditions on b and c, we assume that K is bounded.

(i) If $X_t$ is a Markov semimartingale state process and $Y_t$ is a jump process, then we split the resulting equation for the optimal filter into two equations, one governs the evolution of the filter between the jump times of the observation process; the other one updates the filter at the jump times. Then we ignore the nonlinear terms in both equations, and show that the resulting linear equations have at least one weak solution, which is a finite positive measure. It turns out that normalization of this solution yields the optimal nonlinear filter for the problem. which is the unique solution to the filter problem. We use a Girsanov type change of measure argument and the filtered martingale set-up of Kurtz and Ocone to show existence

and uniqueness of the solution the filter equation respectively.

(ii) If the state process $X_t$ is a continuous time Markov chain and the observation process $Y_t$ is a jump process, then, using the above procedure, we can obtain the solution of the nonlinear filtering equation by solving the following ordinary differential equation and updating linear system (unnormalized density),

$$\dot{Q}(t) = A^{(m_j)} Q(t) \quad \text{for } t \in [T_{j-1}, T_j), Q \in \Re^d \tag{0.1}$$

$$Q(T_j) = B^{(m_j, m_k)} Q(T_j^-) \text{ for } t = T_j , \tag{0.2}$$

where $A^{(m_j)}$ and $B^{(m_j, m_k)}$ are $d \times d$ matrices, which can be obtained from the joint generator for the joint process $\xi_t$, $T_j, j = 1, 2, \cdots$, are the random jump times of the process $Y_t$.

We have shown, by numerical simulation, that solving the filtering problem by this procedure can save considerable time as compared to solving the nonlinear filtering equation directly.

(iii) For the implementation of equations (0.1,0.2), in particular for real time applications, it is important to know, whether (0.1) and (0.2) have a lower dimensional realization, because any dimension reduction to $\hat{d} < d$ may save considerable computation time. We provide some necessary and sufficient conditions for (0.1,0.2) to have lower dimensional realizations by using three different approaches: invariant linear subspaces, invariant integral submanifolds and exact criteria.

(iv) Unfortunately, the exact dimension reduction for (0.1,0.2) is a rare situation, i.e., generically not possible. Hence, it is essential to study approximate dimension reduction for (0.1,0.2). I have developed an efficient and applicable procedure to do approximate dimension reduction . We classify the states of the joint process $\xi_t$

into fast states and slow states according to their jump rates. Then the procedure is based on the rt that the sojourn times of the fast states are very short. or even negligible with respect to the time scale of the problem and hence can be deleted by properly modifying the transition between slow states. We also obtain conditions. under which the approximate nonlinear optimal filter converges to the optimal filter of the problem under consideration and we find corresponding error estimates.

## Explanation of Dissertation Format

The dissertation contains three papers which have been submitted to scholarly journals for publication. Paper II is the joint work with Dr. G. Delgrsso, Dr. W. Kliemann, and Dr. F. Marchetti. They suggested this topic to me and gave a lot of useful suggestions and comments. The papers are followed by a General Summary and literature cited in the General Introduction and General Summary are listed following the General Summary.

6

PAPER I.

ON THE UNNORMALIZED SOLUTION OF THE FILTERING

PROBLEM WITH JUMP PROCESS OBSERVATIONS

# ABSTRACT

This paper presents a general procedure to solve the filtering problem for Markov semimartingale state processes and jump observation processes. We split the resulting equation for the optimal filter into two equations, one governs the evolution of the filter between the jump times of the observation process; the other updates the filter at the jump times. Then we ignore the nonlinear terms in both equations, and show that the resulting linear equations have at least one weak solution, which is a finite positive measure. It turns out that normalization of this solution yields the optimal nonlinear filter for the problem, which is the unique solution to the filter problem.

# 1. INTRODUCTION

Kliemann, Koch and Marchetti [11] have established a general procedure to solve the filtering problem for Markov semimartingale state processes and counting observation processes. They rewrite the resulting nonlinear equation for the optimal filter into two equations, one describes the evolution of the filter between the observation jump times, the other one updates the filter at the jump times. Instead of solving the nonlinear filterting equations, they first solve a linear integral equation and a linear algebraic equation recursively, then normalize the solution, which turns out to be the solution of the original nonlinear equation. They use the Girsanov type change of measure argument and the filtered martingale set-up of Kurtz and Ocone to show the existence and uniqueness of the solution of the filter equation respectively.

Their procedure of solving the nonlinear filtering equation via unnormalized linear equations is based on the specific structure of the counting process observations, namely, its piecewise constant sample paths and unit jump size. Like counting processes, the general jump process also has piecewise constant sample paths. This similarity seems to allow us to extend the above procedure to general jump process observations. Our aim in this paper is to generalize their results to the case that the state processes are the same ones as in [11], but the observation processes are general jump processes ,which have finitely many jumps in any finite interval. Unlike the

counting process, the jump process does not always have monotone increasing sample paths and unit jump size. These differences force the use of different mathematical techniques for the problem, but, on the other hand, we can get more information for the estimation of state process from the jump sizes.

The paper is organized in the following manner. Section 2 is the basic setup, which will be used in the succeeding development. Section 3 discusses the structures of the model and the operators, when the observation process is a jump process. Section 4 considers the dependence between the state and the observation processes. We will show that the quadratic variation between the state and the observation martingale terms can be described by a predictable quadratic covariance operator. Section 5 presents the filtered martingale problem together with its relation to the problem considered here. The existence and uniqueness of the solution of the filter equation will be proved in this section. Section 6 will show that the nonlinear filtering equation has a solution, which is a normalized version of the solution to the linearized equation. Section 7 deals with two examples. The numerical simulations are also given for them.

# 2. PRELIMINARIES AND FORMULATIONS

This section describes most of the results from the literature, which are necessary in the sequel.

## 2.1 Definitions and Notations

Throughout, $(\Omega, \mathcal{F})$ is a fixed measurable space, the sample space, on which a probability measure P is placed. The time interval is $\Re_+ = [0, \infty)$. We equip our sample space $(\Omega, \mathcal{F})$ with a filtration, that is, an increasing family $\{\mathcal{F}_t\}$, $t \geq 0$, of sub-$\sigma$-fields of $\mathcal{F}$. It will always be assumed that the $\mathcal{F}_t$ is right continus. Thus, one has a probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, P)$. We also always assume that the probability spaces are complete. We introduce the following notations.

$$\mu^1(\mathcal{F}_t) = \{M_t : t \in \Re_+, M_t \text{ is an integrable } \mathcal{F}_t - \text{martingale}\}$$

$$\mu^2(\mathcal{F}_t) = \{M_t : t \in \Re_+, M_t \text{ is a square integrable } \mathcal{F}_t - \text{martingale}\}$$

$$\mu^1_{loc}(\mathcal{F}_t) = \{M_t : t \in \Re_+, M_t \text{ is a local integrable } \mathcal{F}_t - \text{martingale}\}$$

$$\mu^2_{loc}(\mathcal{F}_t) = \{M_t : t \in \Re_+, M_t \text{ is a local square integrable } \mathcal{F}_t - \text{martingale}\}$$

$$\mathcal{A}^+(\mathcal{F}_t) = \{a_t : t \in \Re_+, a_t \text{ is an integrable, increasing } \mathcal{F}_t - \text{process}\}$$

$$\mathcal{A}(\mathcal{F}_t) = \{a_t - a'_t : t \in \Re_+, a_t, a'_t \in \mathcal{A}^+(\mathcal{F}_t)\}.$$

$\mathcal{A}_{loc}^{+}$ and $\mathcal{A}_{loc}$ are defined in a manner analogous to the previous definitions.

$$L^2(\langle\, M\, \rangle) = \{\ H:\ H \text{ is predictable, with } \|H\|_{\langle M,\, M\rangle}^2 = E(\int_0^{\infty} H_s^2\, d\langle M,\, M\rangle_s) < \infty$$

where $M$ is a martingale and $\langle\cdot,\cdot\rangle$ denotes the quadratic variation $\}$.

If $A$ is matrix or vector, $A^T$ is the transpose of the $A$. Throughout, the capital $C$ always denotes a positive constant and it may have different meaning in different context.

## 2.2 The Differential Equation for the Model

Since the model which will be used in this paper is the same as the one in the [11], so we quote their results in here.

Let $(X_t, Y_t)$, $t \in \Re_+$, be a partially observable stochastic system with $X_t \in \Re^n$, the state process, and $Y_t \in \Re^p$, the observable component respecticely. We assume: The joint process $\xi_t = (X_t, Y_t)$, $t \in \Re_+$, has the structure of a general Itô process in $\Re^m$ (m = n+p) [5], i.e., $\xi_t$ has the reprsentation:

$$\xi_t = \xi_0 + \int_0^t b(\xi_s)\, ds + \int_0^t c(\xi_s)\, dW_s + \int_0^t \int_Z K(\xi_{s-}, z)(N(ds, dz) - ds\nu(dz)) \quad (2.1)$$

where,

b: $\Re^m \longrightarrow \Re^m$ is a measurable function,

c: $\Re^m \longrightarrow gl(m, \Re)$ is a measurable $(m \times m)$ matrix valued map,

K: $\Re^m \times Z \longrightarrow \Re^m$ is a measurable function, and

$W_t$ is a standard m-dimensional Wiener process defined on $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, P)$, and $(Z, \mathcal{Z})$ is a measurable space. N is a standard Poisson random measure on $\Re_+ \times Z$, defined on $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, P)$, with mean measure $ds\nu(dz)$. We denote:

$$\bar{N}(ds, dz) = N(ds, dz) - ds\nu(dz).$$

$\xi_0$ is a m-dimensional random variable defined on $(\Omega, \mathcal{F})$ independent of $W_t$ and N, with $E(\xi_0^2) < \infty$. We denote the distribution of $\xi_0$ by $\eta_0$, and $\pi_0$ denotes the distribution of $X_0$. If $Y_t$ is to represent a jump process with $Y_0 = 0$, then $\xi_0$ has distribution $\eta_0 = \pi_0 \times \delta_0$ with $\delta_0$ the Dirac measure at 0. We will need the following notations:

$$\bar{b}(\xi_s) = b(\xi_s) - \int_Z K(\xi_{s-}, z)\, \nu(dz), \qquad (2.2)$$

$$V(\xi) = \{z \in Z : |K(\xi, z)| \neq 0\}. \qquad (2.3)$$

We will always assume the following conditions on the coefficients of (1):

(i) $\quad \nu(V(\xi)) \leq C(1 + |\xi|)$

(ii) $\quad \sup_{\xi \in \Re^m, z \in Z} |K(\xi, z)| \leq C$

(iii) $\quad \left|\bar{b}(\xi)\right|^2 + \left|c(\xi)c^T(\xi)\right| \leq C(1 + |\xi|^2), \forall \xi \in \Re^m$

(iv) $\quad \left|\bar{b}(\xi) - \bar{b}(\xi')\right|^2 + \left|c(\xi) - c(\xi')\right|^2 \leq C\left|\xi - \xi'\right|^2, \forall \xi, \xi' \in \Re^m.$

Throughout, because of (i), we will take $Z = \Re$, $\mathcal{Z}$ to be the Lebesgue $\sigma$-field, and $\nu$ the Lebesgue measure [4].

Several sets of sufficient conditions for the existence and uniqueness of the solution of the differential equation (2.1) are proved in [1]. Here, we quote a convenient and useful result from it.

**Theorem 2.2.1** Let $\bar{\xi}_t$ be the diffusion part of (2.1), that is, the process solution of the stochastic differential equation

$$\bar{\xi}_t = \bar{\xi}_0 + \int_0^t b(\bar{\xi}_s)\, ds + \int_0^t c(\bar{\xi}_s)\, dW_s. \qquad (2.4)$$

Assume the above (i) and (ii) are true and assume (2.4) has a unique weak Feller

solution $\bar{\xi}_t$ such that

$$(v) \qquad E(\sup_{0 \leq t \leq 1} |\bar{\xi}_t - \xi|^4 |\bar{\xi}_0 = \xi) \leq C.$$

Then (2.1) has a unique strong solution in the class of Ito processes, which is a time homogeneous Feller process with $t \in [0, T]$ for any $T < \infty$.

Throughout, we always assume (i)-(iv) and Theorem 2.2.1. Next, we introduce some facts concerning the strong generator L of the solution of the equation (2.1) that will be used later.

(1) The martingale problem is well posed for both generators, generator L and its adjoint generator $L^*$, with solution in $\mathcal{D}_{\Re^m}[0, \infty)$ [8]. ($\mathcal{D}_{\Re^m}[0, \infty)$ is the Skorohod space ).

(2) Denote by $\mathcal{D}(A)$ and $\mathcal{R}(A)$ the domain and the range of an operator A respectively. $C_b^2(\Re^m)$ is the space of twice continuously differentiable functions on $\Re^m$, bounded together with their derivatives. We have

$$C_b^2(\Re^m) \subset \mathcal{D}(L).$$

If $f \in C_b^2(\mathcal{R}^m)$, then [5]

$$Lf(\xi) = (\nabla f(\xi), b(\xi)) + \frac{1}{2} trace(\nabla^2 f(\xi), c(\xi)c^T(\xi)) \qquad (2.5)$$

$$+ \int_Z [f(\xi + K(\xi, z)) - f(\xi) - (\nabla f(\xi), K(\xi, z))] \nu(dz),$$

where $(\cdot, \cdot)$ denotes the inner product in $\Re^m$.

(3) The domain of the resolvent $(\alpha I - L)^{-1}$ is separating for each $\alpha > 0$ and bounded pointwise dense in the space of bounded measurable functions[5].

# 3.   THE OBSERVATION PROCESS

In this section, we consider a special case of the model (2.1), that is, the state process $X_t$ lives in $\Re^n$ and the observation process is a 1-dimensional jump process. Before specializing the model and the observation process, we briefly describe the jump process.

## 3.1   The Jump Process and its Parameter

We begin by recalling the definition of the jump process in [6].

**Definition 3.1.1** A jump process $\{U_t\}$, $0 \leq t$, defined on $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, P)$ with values in $(Z, \mathcal{Z})$ is defined by random sequences, $\{T_n\}$, and $\{Z_n\}$, n=0,1, $\cdots$, where $T_0 = 0, T_{n+1} > T_n > 0$, n = 1, $\cdots$ are the jump times and $Z_n \in Z$, n = 0, 1, $\cdots$ are $\mathcal{Z}$-measurable, such that

$$U_t = \begin{cases} Z_0 & \text{if } 0 \leq t < T_1 \\ Z_n & \text{if } T_n \leq t < T_{n+1} \\ Z_\infty & \text{if } T_\infty \leq t \end{cases}$$

where $T_n \uparrow \infty$ is allowed , as well as $Z_n \uparrow \infty$, but they are finite on $[0, T]$.

In [6], the jump process is determined by the joint distribution of $(T_n, Z_n$, n =0, 1, $\cdots$ ) and described by its local characteristics. We will characterize our jump process by its integral representaion and describe it by the jump parameter, which is

given in the following definitions.

**Definition 3.1.2** If a jump process $\{U_t\}$, $t \in \Re_+$ has the representation

$$U_t = \int_0^t \lambda_s \, ds + M_t$$

$t \in \Re_+$, where $\lambda_t \in \mathcal{A}^+_{loc}(\mathcal{F}_t)$, and $M_t \in \mu^2_{loc}(\mathcal{F}_t)$, then $\lambda_t$ is called the jump intensity of $U_t$.

**Definition 3.1.3** If a jump process $U_t$ can be decomposed as

$$U_t = U_t^I - U_t^O$$

where

$$U_t^I = \Sigma_{0 \leq T_{i-1} \leq t} U_{T_{i-1}} I\{0 \leq U_{T_{i-1}}\} \text{ and } U_t^O = \Sigma_{0 \leq T_{i-1} \leq t} \left| U_{T_{i-1}} \right| I\{U_{T_{i-1}} \leq 0\},$$

we call $U_t^I$ the input process and $U_t^O$ the output process of $U_t$.

**Definition 3.1.4** If a jump process $U_t$ has the reprsentation

$$U_t = \int_0^t \lambda_s \, ds + M_t$$

where $\lambda_t = \lambda_t^I - \lambda_t^O$, $\lambda_t^I$ is the jump (up) intensity of the input process $U_t^I$, and $\lambda_t^O$ is the jump (down) intensity of the output process $U_t^O$, and $M_t \in \mu^2_{loc}(\mathcal{F}_t)$, then $\lambda_t$ is called the jump parameter of $U_t$.

We turn next to the discusion of the observation process.

## 3.2 The Structure of the Observation Process

In order to get a structure of the model (2.1) for $X_t \in \Re^n$ and $Y_t \in \Re$, we introduce the following notations:

(i)      $K(\xi, z)_{m \times 1} = ((K_X(\xi, z))_{1 \times n}^T, (K_Y(\xi, z))_{1 \times 1})^T$,

(ii)      $V_X(\xi) = \{z \in Z : |K_X(\xi, z)| \neq 0\}$,

        $V_Y(\xi) = \{z \in Z : K_Y(\xi, z) \neq 0\}$,

(iii)     $b(\xi)_{m \times 1} = ((b_X(\xi))_{1 \times n}^T, (b_Y(\xi))_{1 \times 1})^T$,

        $\bar{b}(\xi) = (\bar{b}_X(\xi), \bar{b}_Y(\xi))^T$,

        $c(\xi) = (c_X(\xi), c_Y(\xi))$.

Using these notations, we get the equations for the state process $X_t$ and the observation process $Y_t$ as follows:

$$X_t = X_0 + \int_0^t \bar{b}_X(\xi_s)\, ds + \int_0^t c_X(\xi_s)\, dW_s + \int_0^t \int_Z K_X(\xi_{s-}, z)\, N(ds, dz),$$

$$Y_t = Y_0 + \int_0^t \bar{b}(\xi_s)\, ds + \int_0^t c_Y(\xi_s)\, dW_s + \int_0^t \int_Z K_Y(\xi_{s-}, z)\, N(ds, dz).$$

Furthermore, we have the following assumptions about our observation process $Y_t$. Assumptions:

(i)      $Y_0 = \bar{b}(\xi_s) = 0, c_Y = 0$.

(ii)     $\{T_n\}, n = 0, 1, \cdots, with, T_0 = 0, 0 < T_n < T_{n+1}, a.s., n = 1, \cdots,$

        are the jump times of $Y_t$.

(iii)    $K_Y(\xi_t, z) I_{V_Y} = k_n \quad n = 0, 1, \cdots,$

        where $t \in [T_n, T_{n+1}), k_n \in \Re$ are nonzero $\mathcal{F}$-measurable random variables.

**Remark 3.2.1**   By the assumptions, it is clear that $Y_t$ is a jump process and all of its sample paths are piecewise constant, moreover,

$$Y_t = \int_0^t \int_{V_Y(\xi_{s-})} K_Y(\xi_{s-}, z)\, N(ds, dz)$$

**Lemma 3.2.1** Under the above assumptions, the observation process $Y_t$ has the following representation

$$Y_t = \int_0^t \lambda(\xi_s)\,ds + M_t,$$

where $M_t = \int_0^t \int_{V_Y(\xi_{s-})} K_Y(\xi_{s-}, z)(\tilde{N})(ds, dz)$ and

$\lambda(\xi_t) = \int_{V_Y(\xi_{t-})} K_Y(\xi_{t-}, z)\nu(dz)$ with $M_t \in \mu_{loc}^2(\mathcal{F}_t)$ and that $\lambda(\xi_t)$, $t \in [0, T]$, satisfies the following conditions:

(i) $\qquad\qquad \lambda(\xi_t) \in \mathcal{A}_{loc}(\mathcal{F}_t)$,

(ii) $\qquad\qquad \lambda(\xi_t)$ has a $\mathcal{F}_t$ − predictable version,

(iii) $\qquad\qquad |\lambda(\xi_t)| \leq C(1 + |\xi_t|)$.

The proof of this Lemma can be found in [3] or [6]. From now on, the $\mathcal{F}_t$-parameter $\lambda(\xi_t)$ is the predictable version.

**Remark 3.2.2** (i) The $\lambda(\xi_t)$ defined in the lemma is the jump parameter of $Y_t$, and it satisfies

$$\lambda(\xi_t) = \lambda^I(\xi_t) - \lambda^O(\xi_t)$$

where $\lambda^I(\xi^t)$ and $\lambda^O(\xi_t)$ are the jump intensities of $Y_t^I$ and $Y_t^O$ respectively.

(ii) Fix $0 < T$ , define

$$\bar{\lambda}_t = \lambda_{t \wedge T}, \bar{M}_t = M_{t \wedge T}$$

then $\bar{\lambda}_t \in \mathcal{A}(\mathcal{F}_t)$, $\bar{M}_t \in \mu^2(\mathcal{F}_t)$. Since we only consider $t \in [0, T]$, we always write $M_t$ for $\bar{M}_t$ and $\lambda_t$ for $\bar{\lambda}_t$.

## 3.3 The Specific Structure of the Operator L

Using the specific equations for $X_t$ and $Y_t$, we can get the specific structure for the opertors defined in Section 2.2.

First of all, for $f \in C_b^2(\Re^{n+1})$, Equation (2.5) has the form

$$
\begin{aligned}
Lf(\xi) = & \sum_{i=1}^{n} (\bar{b}_X)_i(\xi)\frac{\partial f(\xi)}{\partial x_i} + \frac{1}{2}\sum_{i=1}^{n}\sum_{k=1}^{n}(c_X(\xi)c_X^T(\xi))_{ik}\frac{\partial^2 f(\xi)}{\partial x_i \partial x_k} \\
& + \int_Z [f(x + K_X(\xi,z),y) - f(x,y)]\,\nu(dz) \\
& + \int_Z [f(x + K_X(\xi,z),y + K_Y(\xi,z)) - f(x + K_X(\xi,z),y)]\,\nu(dz).
\end{aligned} \tag{3.1}
$$

If we define

$$
\begin{aligned}
L_1 f(\xi) = & \sum_{i=1}^{n} (\bar{b}_X)_i(\xi)\frac{\partial f(\xi)}{\partial x_i} + \frac{1}{2}\sum_{i=1}^{n}\sum_{k=1}^{n}(c_X(\xi)c_X^T(\xi))_{ik}\frac{\partial^2 f(\xi)}{\partial x_i \partial x_k} \\
& + \int_Z [f(x + K_X(\xi,z),y) - f(x,y)]\,\nu(dz), \quad \text{for y fixed,} \\
L_2 f(\xi) = & \; L_2 f(x + K_X,y) \\
= & \int_Z [f(x + K_X,y + K_Y) - f(x + K_X,y)]\,\nu(dz), \quad \text{for x fied,}
\end{aligned}
$$

then (3.1) becomes

$$
Lf(\xi) = L_1 f(\xi) + L_2 f(x + K_X(\xi,z),y)
$$

We should notice that L, $L_1$, $L_2$ are conservative generators [8]. Now we turn to consider the restriction of the operator L on $D_X(L)$ and $D_Y(L)$, which are defined as follows.

**Definition 3.3.1** We define:

$$
D_X(L) = \{f \in D(L) : f(\xi) = f(x,y) = \varphi(x), \varphi \in C_b^2(\Re^n), x \in \Re^n, y \in \Re\}
$$

$$
D_Y(L) = \{f \in D(L) : f(\xi) = f(x,y) = \psi(x), \psi \in C_b^2(\Re), x \in \Re^n, y \in \Re\}
$$

$L_X = L|_{D_X}(L)$, the restriction of L on $D_X(L)$. $L_Y = L|_{D_Y}(L)$, the restriction of L on $D_Y(L)$.

Using the notations in the above definitions, we have , for $f(\xi) \in D_X(L)$,

$$
\begin{aligned}
Lf(\xi) &= L_1 f(\xi) = L_X \varphi(x) \\
&= \sum_{i=1}^{n} (\bar{b}_X)_i \frac{\partial \varphi(x)}{\partial x_i} + \frac{1}{2} \sum_{i=1}^{n} \sum_{k=1}^{n} (c_X(\xi) c_X^T(\xi))_{ik} \frac{\partial^2 \varphi(x)}{\partial x_i \partial x_k} \\
&\quad + \int_Z [\varphi(x + K_X(\xi, z)) - \varphi(x)] \nu(dz),
\end{aligned}
$$

and for $f(\xi) \in D_Y(L)$,

$$
Lf(\xi) = L_2 f(\xi) = L_Y \psi(y) = \int_Z [\psi(y + K_Y(\xi, z)) - \psi(y)] \nu(dz).
$$

Moreover, recalling that $X_t$ and $Y_t$ are semimartingales, then for $f(\xi) \in D_X(L)$, we have the representation

$$
\varphi(X_t) = \varphi(X_0) + \int_0^t L_X \varphi(X_s) \, ds + M_t^\varphi, \tag{3.2}
$$

where

$$
\begin{aligned}
M_t^\varphi &= \int_0^t \sum_{i=1}^{n} (c_X(\xi_s))_i \frac{\partial \varphi(X_s)}{\partial x_i} \, dW_s \\
&\quad + \int_0^t \int_Z [\varphi(X_{s-} + K_X(\xi_{s-}, z)) - \varphi(X_{s-})] \, \bar{N}(ds, dz).
\end{aligned} \tag{3.3}
$$

It is not difficut to see that $\varphi(X_t)$ is a semimartingale and $M_t^\varphi \in \mu^2(\mathcal{F}_t)$ (recalling Remark 3.2.2). For $f(\xi) \in D_Y(L)$, we have the representation

$$
\psi(Y_t) = \psi(Y_0) + \int_0^t L_Y \psi(Y_s) \, ds + M_t^\psi, \tag{3.4}
$$

where $M_t^\psi = \int_0^t \int_Z [\psi(Y_{s-} + K_Y(\xi_{s-}, z)) - \psi(Y_{s-})] \, \bar{N}(ds, dz)$ and $M_t^\psi \in \mu^2(\mathcal{F}_t)$.

Define $\Delta Y = (K_Y(\xi, z)) I_{V_Y(\xi)}$, $\Delta \psi = \psi(Y + \Delta Y, z) - \psi(Y)$, and consider

$t \in [T_{n-1}, T_n)$, then we have $\Delta Y = k_n$ and $\Delta \psi = \psi(Y + k_n, z) - \psi(Y)$. Hence, by

the theorem of Fubini we can interchange the order of integration [14], and obtain

$$\int_{T_{n-1}}^{T_n} \int_Z [\psi(Y_{s-} + \Delta Y_{s-}, z) - \psi(Y_{s-})] \, \nu(dz) ds$$

$$= \int_{T_{n-1}}^{T_n} \int_{V_Y(\xi_{s-})} [\psi(Y_{s-} + \Delta Y_{s-}, z) - \psi(Y_{s-})] \, \nu(dz) ds$$

$$= \int_{V_Y(\xi_{s-})} \int_{T_{n-1}}^{T_n} [\psi(Y_{s-} + k_n, z) - \psi(Y_{s-})] \, ds\, \nu(dz)$$

$$= \int_{T_{n-1}}^{T_n} \int_{V_Y(\xi_{s-})} \frac{\psi(Y_{s-} + k_n) - \psi(Y_{s-})}{k_n} k_n \, \nu(dz) ds$$

$$= \int_{T_{n-1}}^{T_n} \frac{\psi(Y_{s-} + k_n) - \psi(Y_{s-})}{k_n} \int_{V_Y(\xi_{s-})} k_n \, \nu(dz) ds$$

$$= \int_{T_{n-1}}^{T_n} \frac{\Delta \psi(Y_{s-})}{\Delta Y_{s-}} \lambda(\xi_s) \, ds.$$

Thus

$$\int_0^t L_Y \psi(Y_s) \, ds = \int_0^t \frac{\Delta \psi(Y_s)}{\Delta Y_s} \lambda(\xi_s) \, ds \tag{3.5}$$

and therefore

$$\psi(Y_t) = \psi(Y_0) + \int_0^t \frac{\Delta \psi(Y_s)}{\Delta Y_s} \lambda(\xi_s) \, ds + \int_0^t \frac{\Delta \psi(Y_{s-})}{\Delta Y_{s-}} [dY_s - \lambda(\xi_s) ds]. \tag{3.6}$$

# 4. THE PREDICTABLE QUADRATIC COVARIANCE OPERATOR

The main purpose of this section is to describle the dependence between the state process $X_t$ and the observation process $Y_t$. It will be established that this dependency is given by the predictable quadratic covariance operator, which is defined as follows.

**Definition 4.1** A quadratic covariance operator R is defined to be a linear bounded operator defined on $D(L)$ , where L is defined in Section 2.2, such that, for $t \geq 0$

$$\langle M^f, N \rangle_t = \int_0^t Rf(\xi_s)\, ds$$

where $M^f, N \in \mu^2(\mathcal{F}_t)$, $f \in D(L)$ and $\xi_t$ is the Itô process which is defined in Section 2.2 . The quadratic covariance operator R is predictable, if the process $\{Rf(\xi_t)\}_t$ , $0 \leq t$, is a predictable process.

Recall that the jump parts of the semimartingale $\varphi(X_t)$ and the observation process $Y_t$ are described by

$$M_t^\varphi = \int_0^t \sum_{i=1}^n (c_X(\xi_s))_i \frac{\partial \varphi(X_s)}{\partial x_i}\, dW_s$$
$$+ \int_0^t \int_Z [\varphi(X_{s-} + K_X(\xi_{s-}, z)) - \varphi(X_{s-})]\, \bar{N}(ds, dz)$$
$$and \quad M_t = \int_0^t \int_Z K_Y(\xi_{s-}, dz)\, \bar{N}(ds, dz)$$

respectively. Now, the dependence between $X_t$ and $Y_t$ can be characterized by the following Theorem.

**Theorem 4.1** There exists a quadratic covariance operator R such that

$$\langle M^{\varphi}, M \rangle_t = \int_0^t R_X \varphi(X_s) \, ds$$

where $R_X = R|_{D_X(L)}$, is the restriction of R to $D_X(L)$, and $R_X \varphi(X_s)$ has a predictable version.

The Theorem 4.1 is an immediate consequence of the following lemmas.

**Lemma 4.1** Under the above assumptions, we have

$$\langle M^{\varphi}, M \rangle_t = \int_0^t \int_Z [\varphi(X_{s-} + K_X(\xi_{s-}, z)) - \varphi(X_{s-})] K_Y(\xi_{s-}, z) \, \nu(dz) ds.$$

Proof. We write

$$M_t^{\varphi} = M_t^{\varphi, c} + M_t^{\varphi, d} .$$

where $M_t^{\varphi, c} = \int_0^t \sum_{i=1}^n (c_X(\xi_{s-}))_i \frac{\partial \varphi(X_s)}{\partial x_i} dW_s$ is the continuous part of $M_t^{\varphi}$,

$$M_t^{\varphi, d} = \int_0^t \int_Z [\varphi(X_{s-} + K_X(\xi_{s-}, z)) - \varphi(X_{s-}))] \bar{N}(ds, dz)$$

is the jump part of $M_t^{\varphi}$. Then,

$$\langle M^{\varphi}, M \rangle_t = \langle M^{\varphi, c}, M \rangle_t + \langle M^{\varphi, d}, M \rangle_t.$$

Since M is a purely discontinuous martingale, it follows from [7] and [9], that $\langle M^{\varphi, c}, M \rangle_t = 0$, and

$$\langle M^{\varphi, d}, M \rangle_t = \int_0^t \int_Z [\varphi(X_{s-} + K_X(\xi_{s-}, z)) - \varphi(X_{s-})] K_Y(\xi_{s-}, z) \, \nu(dz) ds,$$

therefore,

$$\langle M^{\varphi}, M \rangle_t = \int_0^t \int_Z [\varphi(X_{s-} + K_X(\xi_{s-}, z)) - \varphi(X_s)] K_Y(\xi_{s-}, z) \, \nu(dz) ds.$$

Q.E.D.

Define

$$Rf(\xi_t) = \int_Z [f(\xi_{t-} + K(\xi_{t-}, z)) - f(\xi_{t-})] K_Y(\xi_{t-}, z) \, \nu(dz) \qquad (4.1)$$

for $f \in D(L)$, then the next lemma shows that R is a quadratic covariance operator.

**Lemma 4.2** (i) $\xi_t$, the solution of (2.1), has finite second moments in $[0,T]$.

(ii) If $f \in C_b^2(\Re^{n+1})$, then the operator defined in (4.1) is a bounded linear operator.

Proof. (i) is from [11]. (ii) follows by the assumptions in Section 2.2 .

$Q.E.D.$

Following the above lemmas, we know that R defined in (4.1) is a quadratic covariance operator, with

$$\langle M^\varphi, M \rangle_t = \int_0^t R_X \varphi(X_s) \, ds.$$

Now we need to show that the process $R_X \varphi(X_s)$ has a predictable version.

**Lemma 4.3** There exists a predictable process H such that

$$\langle M^\varphi, M \rangle_t = \int_0^t H_s \bar\lambda(\xi_s) \, ds,$$

where $\bar\lambda(\xi_t) = \int_Z K_Y^2(\xi_t, z) \, \nu(dz)$.

Proof. By using the notations and results in the proof of Lemma 4.1 , we have $\langle M^\varphi, M \rangle_t = \langle M^{\varphi,d}, M \rangle_t$. Recall that both $M_t^{\varphi,d}$ and $M_t$ are square integrable martingales, then the stable space generated by M is the set [7],

$$S = \{H \cdot M : H \in L^2(\langle M \rangle)\}.$$

where $(H \cdot M)_t = \int_0^t H_s \, dM_s$

Write $M^{\varphi,d} = M_1^{\varphi,d} \perp M_2^{\varphi,d}$, where $M_1^{\varphi,d}$ is the projection of $M^{\varphi,d}$ onto S, and $M_2^{\varphi,d} \in S^{\perp}$. Then

$$
\begin{aligned}
\langle M^{\varphi,d}, M \rangle_t &= \langle M_1^{\varphi,d}, M \rangle_t + \langle M_2^{\varphi,d}, M \rangle_t \\
&= \langle M_1^{\varphi,d}, M \rangle_t.
\end{aligned}
$$

Since $M_1^{\varphi,d} \in S$, it can be written as $M_1^{\varphi,d} = H \cdot M$, therefore

$$
\langle M^{\varphi}, M \rangle_t = \langle H \cdot M, M \rangle_t = \int_0^t H_s \, d\langle M, M \rangle_s.
$$

Since $\langle M, M \rangle_s = \int_0^s \int_Z K_Y^2(\xi_{\tau-}, z) \, \nu(dz) d\tau$, we have

$$
\langle M^{\varphi}, M \rangle_t = \int_0^t H_s \bar{\lambda}(\xi_s) \, ds,
$$

where $\bar{\lambda}(\xi_s) = \int_Z K_Y^2(\xi_{s-}, z) \, \nu(dz)$.

Q.E.D.

**Lemma 4.4** For $f = \varphi(X)$, $\varphi(X) \in D_X(L)$, we have

$$
R_X \varphi(X_s) = H_s \bar{\lambda}(\xi_s), \text{ a.s on } [0,T].
$$

Proof. By Lemmas 4.1 and 4.3,

$$
\int_0^t (R_X \varphi(X_s) - H_s \bar{\lambda}(\xi_s)) \, ds = 0, \text{ for all } t \in [0,T]
$$

then we have (compare [14]) $R_X \varphi(X_s) = H_s \bar{\lambda}(\xi_s)$, a.s on [0,T] .

Q.E.D.

# 5.  THE FILTERED MARTINGALE PROBLEM

In this section, we will introduce a martingale formulation for the solution of the filtering problem. We also will establish the filter equation for our specialized model in Section 3.2 and then show that the filter equation has a unique solution.

## 5.1    Basic Setup

Although we will follow subsection III.A in [11] closely, we find it useful to recall here the filtered martingale set-up of T.G. Kurtz and D. Ocone. First, in addition to those already introduced, we need to use the following notations for the following sections.

E, F will denote euclidean spaces, say $\Re^n$, for some $n \in \mathcal{N}$, B(E) the bounded Borel functions on E, $\mathcal{P}(E)$ the set of probability measures on the Borel sets of E and $\mathcal{D}_E[0, \infty)$ the space of cadlag (right continuous with left limits) E-valued paths. As usual, $\mathcal{P}(E)$ carries the topology of weak convergence and $\mathcal{D}_E[0, \infty)$ the Skorohod topology.

For a (finite) measure on E, and $f \in B(E)$, we write $\mu f = \int_E f(s)\,\mu(ds)$. Notice that the notation may be used for unbounded functions f, whenever $\int_E f(s)\,\mu(ds)$ makes sence.

Here we quote some results from [12] for our purpose.

**Definition 5.1.1** A process $(\mu_t, U_t)$ with sample paths in $D_{\mathcal{P}(E) \times F}[0,T]$ is a solution of the filtered martingale problem for the operator L of the equation (2.1), if $\mu_t$ is $\{\mathcal{F}_t^{U}\}$ -adapted and

$$\mu_t f(\cdot, U_t) - \mu_0 f(\cdot, U_0) - \int_0^t \mu_s Lf(\cdot, U_s)\, ds$$

is an $\{\mathcal{F}_t^{U}\}$-martingale for every $f \in D(L)$.

**Theorem 5.1.1** (Section 3 of [12]) If $(\mu_t, U_t)$ is $\mathcal{F}_t^{U}$-adapted for each t, then there exists a measurable function $H_t : D_E[0,T] \to \mathcal{P}(E)$ such that

$$\mu_t = H(U_s I_{[s<t]}, s \in [0,T]), \quad \text{for all } t \in [0,T] \tag{5.1}$$

**Theorem 5.1.2** (Theorem 3.2 in [12]) Suppose that the pair $(X_t, Y_t)$ in $E \times F$ is the solution of a well -posed martingale problem for the opertor L and that, moreover, $\mathcal{R}(\alpha I - L)$, the range of operator $\alpha I - L$, is bounded pointwise dense in $B(E \times F)$ for all $\alpha > 0$. Then the FMP (the filtered martingale problem) has at most one solution (in law) for each distribution of the initial value $(\mu_0, U_0)$.

**Theorem 5.1.3** (Corollary 3.4 in [12]) If $(X_t, Y_t)$ is a solution to the martingale problem for L and

$$E[\mu_0 f(\cdot, U_0)] = E[f(X_0, Y_0)] \quad \text{for all } f \in D(L) \tag{5.2}$$

holds, then $U_0 = Y_0$ in law. Furthermore, if the solution to the FMP is unique, then Theorem 5.1.2 implies that

$$(\mu_t, U_t) = (\pi_t, Y_t) \text{ in law}$$

where $\pi_t$ is the conditional law of $X_t$ given $\mathcal{F}_t^{Y}$, and there is a version of $\pi_t$ such that $\pi = H(Y)$ in the sense of (5.1).

## 5.2   The Filter Equation

We begin this subsection by considering the following filter equation for our specific model defined in Section 3, as they are given in [2,13]:

$$
\begin{aligned}
\pi_t f(\cdot, Y_t) &= \pi_0 f(\cdot, Y_0) + \int_0^t \pi_s L f(\cdot, Y_s)\, ds \\
&+ \int_0^t [\pi_{s-}\lambda(\cdot, Y_{s-})]^+ [\pi_{s-}(\lambda(\cdot, Y_{s-})f(\cdot, Y_{s-})) - \\
&- \pi_{s-}\lambda(\cdot, Y_{s-})\pi_{s-}f(\cdot, Y_{s-}) + \pi_{s-}Rf(\cdot, Y_{s-})]\,[dY_s - \pi_s\lambda(\cdot, Y_s)ds]
\end{aligned}
\tag{5.3}
$$

for $f \in D(L)$, where $(\cdot)^+$ denotes the generalized inverse:

$$
(a)^+ = \frac{1}{a} \quad \text{if } a \neq 0; \qquad (a)^+ = 0 \quad \text{if } a = 0.
$$

**Remark 5.2.1:**(i) By noticing the asuumptions that we made on L and $\lambda$ and the integrability of $\xi_t$, it is not difficult to see that all integrals on the right side are well defined for every $t \in [0, T]$, and hence (5.3) is a representation of the conditional distribution $\pi_t$.

(ii) The third term is martingale, due to the minimal reprsentation [13]

$$
Y_t = \int_0^t \pi_s \lambda(\xi_s)\, ds + M_t'
$$

where $M_t'$ is a $\{\mathcal{F}_t^Y\}$-martingale.

If $f \in D_X(L)$, then we have the following equation:

$$
\begin{aligned}
\mu_t \varphi &= \pi_0 \varphi + \int_0^t \mu_s L_X \varphi\, ds \\
&+ \int_0^t [\mu_{s-}\lambda(\cdot, U_{s-})]^+ [\mu_{s-}(\lambda(\cdot, U_{s-})\varphi) - \\
&- \mu_{s-}\lambda(\cdot, U_{s-})\mu_{s-}\varphi + \mu_{s-}R_X\varphi]\,[dU_s - \mu_s\lambda(\cdot, U_s)ds].
\end{aligned}
\tag{5.4}
$$

Now, we turn to show that Equation (5.4) has a unique solution. Recalling the facts about the operator L in subsection 2.2 and using Theorem 5.1.2, we immediately have the following theorem.

**Theorem 5.2.1** If the FMP for the model (2.1) has a solution, then this solution is unique (in law) and provides the required conditional distribution law $\pi_t$.

The existence of the solution of the filter equation (5.3) is from the next theorem.

**Theorem 5.2.2** Let $U_t$ be a jump process with parameter $\bar{\lambda}_t = \bar{\lambda}_t^I - \bar{\lambda}_t^O$, where $\bar{\lambda}_t^I$ and $\bar{\lambda}_t^O$ are the jump intensities of the input and output process $U_t^I$ and $U_t^O$ respectively, and having values in $\mathcal{D}_{\Re}[0,T]$. Let $\mu_t$ be a measure valued process with values in $\mathcal{D}_{\mathcal{P}(\Re^n)}[0,T]$, adapted to $\{\mathcal{F}_t^U\}$. Assume

(i) $\mu_t g \leq C'(1 + U_t^2)$ for all $t \in [0,T]$, $\bar{P}$-a.e, where $g(x) = |x|^2$, $x \in \Re^n$, and $\bar{P} = P|_{\mathcal{F}_t^U}$, the restriction of P to $\mathcal{F}_t^U$, and

(ii) $(\mu_t, U_t)$ satisfies Equation (5.4) $\bar{P}$-a.e, for each $f \in D_X(L)$.

Then (i) There exists a probability measure $\bar{P}_1$ on $\mathcal{F}_T^U$ such that, under $\mathcal{F}_T^U$, $U_t$ is a jump process with parameter $\mu_t \lambda(\cdot, U_t) = \mu_t \lambda^I(\cdot, U_t) - \mu_t \lambda^O(\cdot, U_t)$, where $\lambda(\cdot, U_t)$ is defined in section 3.

(ii) The pair $(\mu_t, U_t)$ is a solution to the FMP for the model defined in (2.1), which also satisfies equality (5.2).

Before we show this theorem, let us make the following useful remarks.

**Remark 5.2.2** (i) We will know that, by the construction, the new measure $\bar{P}_1$ is absolutly continuous with respect to $\bar{P}$, hence all properties , which hold $\bar{P}$-a.e., also hold $\bar{P}_1$-a.e.. In particular, if $\mu_t = H(U_s I_{(s<t)}, s \in [0,T])$ for all $t \in [0,T]$ holds $\bar{P}$-a.e, then it holds $\bar{P}_1$-a.e. By recalling (5.2), and Theorem 5.2.1, it follows that $\pi_t = H(Y_s I_{[0,t]}(s), s \in [0,T])$, $t \in [0,T]$, $\bar{P}$-a.e and therefore P-a.e.

(ii) By assumption (i) and Lemma 3.2.1, we get

$$\mu_t|\lambda(\cdot, U_t)| \le C'[\mu_t(1 + |\cdot|^2 + U_t^2)^{\frac{1}{2}}]$$

and then

$$\mu_t|\lambda(\cdot, U_t)| \le C'(1 + U_t)$$

for all $t \in [0, T]$, $\bar{P}$-a.e.

Proof of Theorem 5.2.2. First, we show the part(i). Recall that $\lambda^I$ and $\lambda^O$ are $\mathcal{F}_t$-predictable nonnegative process, and then $\mu_t\lambda^I(\cdot, U_t)$ and $\mu_t\lambda^O(\cdot, U_t)$ are $\mathcal{F}_t$-predictable nonnegative processes. Under $\bar{P}$, by Assumption (i) and Remark 5.2.2 (ii), we have

$$\bar{E}(\int_0^T \mu_t\lambda^I(\cdot, U_t)\,dt) < \infty \tag{5.5}$$

and

$$\bar{E}(\int_0^T \mu_t\lambda^O(\cdot, U_t)\,dt) < \infty. \tag{5.6}$$

If $\mu_t|\lambda(\cdot, U_t)|$ is $\bar{P}$-a.e uniformly bounded in $[0,T]$, we define

$$L_t = exp\{\int_0^t ln(\frac{\mu_s\lambda^I(\cdot, U_s)}{\bar{\lambda}_s^I})\,dU_s^I - \int_0^t (\mu_s\lambda_s^I(\cdot, U_s) - \bar{\lambda}_s^I)\,ds\}\cdot$$

$$exp\{\int_0^t ln(\frac{\mu_s\lambda^O(\cdot, U_s)}{\bar{\lambda}_s^O})\,dU_s^O - \int_0^t (\mu_s\lambda_s^O(\cdot, U_s) - \bar{\lambda}_s^O)\,ds\}.$$

Then $\{L_t\}$ is a $\mathcal{F}_t$-martingale [11] and $\bar{E}(L_T) = 1$. Therefore , if we define $\bar{P}_1$ by

$$\frac{d\bar{P}_1}{d\bar{P}} = L_t, t \in [0, T],$$

then $U_t$, under $\bar{P}_1$, is a jump process with parameter [11] $\mu_t\lambda(\cdot, U_t) = \mu_t\lambda^I(\cdot, U_t) - \mu_t\lambda^O(\cdot, U_t)$.

We turn next to the general case. Define

$$\tau_n = inf\{T, inf\{0 \le t : max((\mu_t\lambda^I(\cdot, U_t), (\mu_t\lambda^0(\cdot, U_t)) \ge n\}\}$$

n = 1, 2, $\cdots$. Then $\tau_n$ is a $\mathcal{F}_t^U$- stopping time.

Define, for each $n \geq 1$, the probability $\bar{P}^{(n)}$ by,

$$\frac{d\bar{P}^{(n)}}{d\bar{P}} = L_{t \wedge \tau_n}, t \in [0, T].$$

Since, for $0 \leq t \leq \tau_n$, $\mu_t |\lambda(\cdot, U_t)|$ is uniformly bounded, then by the same arguments as above, under $\bar{P}^{(n)}$, $U_t$ is a jump process with jump parameter $\mu_{t \wedge \tau_n} \lambda(\cdot, U_{t \wedge \tau_n})$. Thus, we have finished the proof of part(i).

We begin the proof of part (ii) by showing that $(\mu_t, U_t)$ solves, under $\bar{P}^{(n)}$, a stopped FMP, that is, for every $f \in D(L)$,

$$\mu_{t \wedge \tau_n} f(\cdot, U_{t \wedge \tau_n}) - \pi_0 f(\cdot, U_0) - \int_0^{t \wedge \tau_n} \mu_s L f(\cdot, U_s) \, ds \tag{5.7}$$

is a $\mathcal{F}_{t \wedge \tau_n}^U$-martingale.

To prove this, it is enough to show that (19) holds for a set dense in D(L). such as the set of $f \in D(L)$ with the form $f(x, y) = \varphi(x)\psi(y)$ (because of the linearity of the equation (5.7) ). Recalling (3.2), (3.5) and (5.4), applying the product rule for the semimartingles $\mu_t \varphi$ and $\psi(U_t)$, and noticing that

$$\mu_t [\varphi(\cdot)\psi(U_t)] = (\mu_t \varphi)\psi(U_t),$$

we can get, for such a function f and for $t \in [0, \tau_n]$,

$$
\begin{aligned}
\mu_t f(\cdot, U_t) &= \mu_t \varphi \psi(U_t) \\
&= \pi_0 \varphi \psi(U_0) + \int_0^t \mu_{s-} \varphi \, d\psi(U_s) \\
&\quad + \int_0^t \psi(U_{s-}) \, d\mu_s \varphi + \sum_{s \leq t} \Delta \mu_s \varphi \Delta \psi(U_s).
\end{aligned}
$$

By (3.6)

$$\int_0^t \mu_{s-}\varphi\, d\psi(U_s) = \int_0^t \mu_{s-}\varphi\frac{\Delta\psi(U_s)}{\Delta U_s}\lambda(\xi_s)\, ds$$
$$+ \int_0^t \mu_{s-}\varphi\frac{\Delta\psi(U_{s-})}{\Delta U_{s-}}\,[dU_s - \mu_s\lambda(\cdot,U_s)ds].$$

By (5.4)

$$\int_0^t \psi(U_{s-})\, d\mu_s\varphi = \int_0^t \psi(U_{s-})\mu_s L_X\varphi\, ds +$$
$$+ \int_0^t \psi(U_{s-})[\mu_{s-}\lambda(\cdot,U_{s-})]^+[\mu_{s-}(\lambda(\cdot,U_{s-})\varphi) -$$
$$- \mu_{s-}\lambda(\cdot,U_{s-})\mu_{s-}\varphi + \mu_{s-}R_X\varphi]\,[dU_s - \mu_s\lambda(\cdot,U_s)ds]$$

and

$$\sum_{s\le t}\Delta\mu_s\Delta\psi(U_s) = \int_0^t \Delta\varphi(U_{s-})[\psi_{s-}\lambda(\cdot,U_{s-})]^+[\mu_{s-}(\lambda(\cdot,U_{s-})\varphi) -$$
$$- \mu_{s-}\lambda(\cdot,U_{s-})\mu_{s-}\varphi + \mu_{s-}R_X\varphi]\,[dU_s - \mu_s\lambda(\cdot,U_s)ds].$$

Combining those terms, we have

$$\mu_t f(\cdot,U_t) = \pi_0\varphi(\cdot)\psi(\cdot) + \int_0^t \mu_s L(\varphi\psi)\, ds + M'_t, \tag{5.8}$$

where

$$M'_t = \int_0^t (\mu_{s-}\varphi\frac{\Delta\psi(U_{s-})}{\Delta U_{s-}} + \psi(U_{s-} + \Delta U_{s-}))[\mu_{s-}\lambda(\cdot,U_{s-})]^+\cdot$$
$$[\mu_{s-}\lambda(\cdot,U_{s-})\varphi - \mu_{s-}\lambda(\cdot,U_{s-})\mu_{s-}\varphi + \mu_{s-}R_X\varphi]\,[dU_s - \mu_s\lambda(\cdot,U_s)ds].$$

Clearly, $M'_t$ is a $\mathcal{F}_t^U$-martingale under $\bar{P}^{(n)}$. Thus, under $\bar{P}^{(n)}$, $(\mu_{t\wedge\tau_n}, U_{t\wedge\tau_n})$ has the same distribution as the $(\pi_{t\wedge\tau_n}, Y_{t\wedge\tau_n})$ under $\bar{P}$ [3]. Hence

$$\mu_t I_{(t<\tau_n)} = H(U_s I_{(s<t)}, s \in [0,T])I_{(t<\tau_n)}, \text{ for all } t \in [0,T], \bar{P}^{(n)} - a.e. \tag{5.9}$$

By using the sublinear growth condition on $|\lambda_t|$, it follows [1] that the sequence $\{\bar{P}^{(n)}\}$ admits appropriately coverging subsequences, that any limit $\bar{P}_1$ of such subsequence is absolutly continuous with respect to $\bar{P}^{(n)}$ and agrees with $\bar{P}^{(n)}$ on $\{\mathcal{F}_{\tau_n}^U\}$ and that $\tau_n \uparrow T$, $\bar{P}_1$-a.e. Hence, by (5.9)

$$\mu_t I_{(t<\tau_n)} = H(U_s I_{(s<t)}, s \in [0,T]) I_{(t<\tau_n)}, \text{ for all } t \in [0,T], \bar{P}_1 - a.e$$

for all $n \in \mathcal{N}$, and therefore, $\mu = H(U)$ in the sense of (5.1) with respect to $\bar{P}_1$, that is

$$\mu_t = H(U_s I_{(s<t)}, s \in [0,T]), \forall \ t \in [0,T] \ \bar{P} - a.e..$$

So $\mu_t = H(U_s I_{(s<t)}, s \in [0,T]), \forall t \in [0,T], \bar{P}_1 - a.e..$

Finally, to check that all terms in (5.4) are well-defined, we only need to use Assumption (i) and the growth conditions (i) and (ii) of Section 2.2. Thus, we finish the proof of the theorem.

*Q.E.D.*

# 6. SOLUTION OF THE FILTER EQUATION

In this section, we shall show that (5.4), under above assumptions and with $U_t = Y_t$, our observation process, has a solution.

For a probability law $\mu_t$, the equation (5.4) can be written

$$\mu_t\varphi = \pi_0\varphi + \int_0^t \mu_s L_X\varphi \, ds + \int_0^t [\mu_{s-}\lambda(\cdot, Y_{s-})]^+[\mu_{s-}\lambda(\cdot, Y_{s-})\varphi \quad (6.1)$$
$$- \mu_{s-}\lambda(\cdot, Y_{s-})\mu_{s-}\varphi + \mu_{s-}R_X\varphi] [dY_s - \mu_s\lambda(\cdot, Y_s)ds]$$

for all $\varphi(x) = f(\xi) \in D_X(L)$.

Recalling Remark 3.2.1, and noticing that $T_i$, $i = 1, 2, \cdots$, are the jump times of $Y_t$ ($T_0 = 0$), we can rewrite (6.1) as the following weak equation

$$\mu_t\varphi = \mu_{T_{i-1}}\varphi + \int_{T_{i-1}}^t \mu_s L_X\varphi \, ds \qquad (6.2)$$
$$- \int_{T_{i-1}}^t [\mu_{s-}\lambda(\cdot, Y_{s-})\varphi - \mu_{s-}\lambda(\cdot, Y_{s-})\mu_{s-}\varphi + \mu_{s-}R_X\varphi] \, ds$$

for $t \in [T_{i-1}, T_i)$, and

$$\mu_{T_i}\varphi = [\mu_{T_{i-}}\lambda(\cdot, Y_{T_{i-}})]^+[\mu_{T_{i-}}\lambda(\cdot, Y_{T_{i-}})\varphi + \mu_{T_{i-}}R_X\varphi] \qquad (6.3)$$

with the initial condition $\mu_0 = \pi_0$.

Ignoring the nonlinear terms in (6.2) and (6.3), we have the following equations

$$\rho_t\varphi = \rho_{T_{i-1}}\varphi + \int_{T_{i-1}}^t \rho_s[(L_X - R_X)\varphi] \, ds - \int_{T_{i-1}}^t \rho_{s-}\lambda(\cdot, Y_{s-})\varphi \, ds \qquad (6.4)$$

for $T_{i-1} \leq t < T_i$, and

$$\rho_{T_i}\varphi = \rho_{T_{i-}}\lambda(\cdot, Y_{T_{i-}})\varphi + \rho_{T_{i-}}R_X\varphi \qquad (6.5)$$

$i = 1, 2, \cdots$, with initial condition $\rho_0 = \pi_0$.

**Remark 6.1** Recalling the definition of $\lambda(\cdot, Y_t)$ in section 3.2, we know that

$$\lambda(\cdot, Y_t) = k_i \nu(V_Y(\xi_{t-}))$$

for $t \in [T_{i-1}, T_i)$ and $i = 1, 2, \cdots$.

**Theorem 6.1** (i) Equations (6.4) and (6.5) have at least on one weak solution, which is also a finite measure.

(ii) If this measure is nomalized to unit mass, it also provides a solution to (6.2) and (6.3) with values in $\mathcal{D}_{\mathcal{P}(\Re n)}[0, T]$, adapted to $\{\mathcal{F}_t^Y\}$.

Proof. By using Remark 6.1, the proof of this theorem can be seen in [11].

$Q.E.D.$

If (6.4) and (6.5) happen to have a classical solution, then it follows that this one, after normalization, will provide a version of the conditional density $p_t$. In such cases, we can consider the solution of the adjoint equation for the unnormalized density $q_t$,

$$\frac{\partial q_t(x)}{\partial t} = (L_X^* - R_X^*)q_t(x) - \lambda(x, Y_t)q_t(x), \ t \in [T_{i-1}, T_i) \qquad (6.6)$$

$$q_{T_i}(x) = \lambda(x, Y_{T_i-})q_{T_i-}(x) + R_X^* q_{T_i-}(x) \qquad (6.7)$$

with initial density $p_0$. Then, we obtain $\hat{p}_t$ as $\hat{p}_t = \frac{q_t}{\int_{\Re n} q_t(x) \, dx}$.

**Remark 6.2** Intuitively, if we observe $Y_t$ jumping at time $T_n$, then we expect that the process $X_t$ also changes its states at $T_i$, and the probability of such change should depend on the jump sizes of the $Y_t$. The unnormalized equations (6.5) and (6.7) verify, in some sense, this intuition. In section 7, we will see more about the

behavior of the normalized and unnormalized equations at jump times and between jump times.

If the state process $X_t$ is a continuous time Markov chain, then we can identify the equations (6.4) and (6.5) with a linear ordinary differential equation and linear system by the following methods.

Let $X_t, t \geq 0$, is a Markov chain with state space $\{1, \cdots, d\}$, where d is a positive integer, and $Y_t, t \geq 0$, is a jump process with state space $\{a_1, \cdots, a_n\}$, where n is a positive integer, then we can identify the terms in the above filtering equations using the generators, which are defined as follows.

Define the joint generator $Q = (q_{(i,m_j)(k,m_l)})$, which generates the Markov process $\xi_t = (X_t, Y_t)$, by

$$q_{(i,m_j)(k,m_l)} = \lim_{h \downarrow 0} \frac{1}{h} P\{X_{t+h} = k, Y_{t+h} = a_l | X_t = i, Y_t = a_{m_j}\}.$$

where $m_j, m_l \in J = \{1, 2, \cdots, n\}$

Furthermore, we define

$A^{(m_j)} = (q_{(i,m_j)(k,j)})$,      $d \times d$ matrix of $X_t$-transitions for $Y_t$ constant on the time interval $[T_{j-1}, T_j)$, and

$B^{(m_j)} = (q_{(i,m_j)(k,m_l)})$,      where $j \neq l$, $d \times d$ matrix of $X_t$-transitions for $Y_{T_j}$ jumping, and

$\bar{S}^{(m_j)} = diag(\sum_k q_{(i,m_j)(k,m_l)})$,   $j \neq l$.

Identifying the terms in the above filtering equation, we have

$$L_X^{(m_j)} : = A^{(m_j)} + B^{(m_j)} \qquad \text{is a generator of } X_t \text{ given } Y_t = y_j$$
$$R_X^{(m_j)} : = B^{(m_j)} - \bar{S}^{(m_j)}.$$

Furthermore

$$A^{(m_j)} = L_X^{(m_j)} - R_X^{(m_j)} - \bar{S}^{(m_j)},$$

$$B^{(m_j)} = R_X^{(m_j)} + \bar{S}^{(m_j)},$$

and therefore we have the filter equation

$$\dot{V} = A^{(m_j)}V, \qquad \text{for} \quad Y_t = a_{m_j} \text{ in the interval} [T_{j-1}, T_j)$$

$$V_{T_j} = B^{(m_j)}V_{T_j-} \qquad \text{for} \quad Y_t - \text{jumps at time } T_j,$$

where $j = 1, 2, \cdots$.

## 7. EXAMPLES

In this section, we use our results in two examples. The first example is a simple combination of diffusion state process and jump observations. A numerical simulation will show how the unnormalized and normalized equations help us to determine some properties of the unobservable state process $X_t$. The second example has been discussed by Yashin [15]. Yashin provided a nonlinear filter equation for the problem by using the theory of stochastic differential equations and derivatives of measures on function spaces. A numerical simulation will show that we can save considerable CPU time by use our procedure to solve the filtering equations.

**Example 1** Let $X_t$ be a stationary Ornstein-Uhlenbeck process, i.e., $X_t$ is the unique strong solution of the equation

$$dX_t = -\beta X_t dt + \sigma dW_t$$

where $\beta$ and $\sigma$ are positive constants and $W_t$ is a standard Brownian motion, and the initial random variable $X_0$ has a normal distribution with mean 0 and variance 1(here we assume $\frac{\sigma^2}{2\beta} = 1$)

Let $Y_t$ be a jump process with

$$Y_t = \begin{cases} 1 & \text{if } X_{t-} \geq 0 \\ -1 & \text{otherwise} \end{cases}$$

and with same jump rate $|\lambda(X_{t-})|$.

We can easily write the following equations for the unnormalized density $q_t$.

$$\frac{\partial q_t(x)}{\partial t} = \frac{1}{2}\frac{\partial^2 q_t(x)}{\partial^2 x} - \lambda(x, Y_t)q_t(x), \text{ for } t \in [T_{i-1}, T_i)$$

$$q_{T_i}(x) = \lambda(x, Y_{T_i^-})q_{T_i^-}(x), \text{ for } t = T_i.$$

We use $|\lambda(X_{t-})| = 0.5$ to simulate the jump times $T_i$. We observe that the $Y_t$ starts with value -1 and then has first jump in $T_1$ and so on. The jump times are given in Figure 7.1. We use a numerical routine to solve the above partial differential equation and algebraic equation. The simulation results are presented in Figures 7.1 and 7.2.

Figure 7.1 is a trajectory of the unnormalized density $q_t(1.5)$. If we observe that $Y_t$ has a +1 jump at $T_i$, then we can set the unnormalized conditional density $q_{T_i}$ to be 0 for $x < 0$ (since $1 = P(T_{i-1} \neq T_i) = P(X_{T_i} \geq 0)$ ).Figure 1 shows that after the jump time $T_1$ the unnormalized density $q_t(+1.5)$ decreases until the jump time $T_2$ , and then it increases until next jump time $T_3$ and so on.

The graphs in the Figure 7.2 show us that

(i) If $Y_t = +1$, then $P(X_t \geq 0) > P(X_t < 0)$, and if $Y_t = -1$, then $P(X_t \geq 0) < P(X_t < 0)$.

(ii) Comparing the graph A and graph B, we can see that between jumps the conditional density builds up again in this region and tends eventually to the stationary distribution of the Ornstein-Uhlenbeck process.
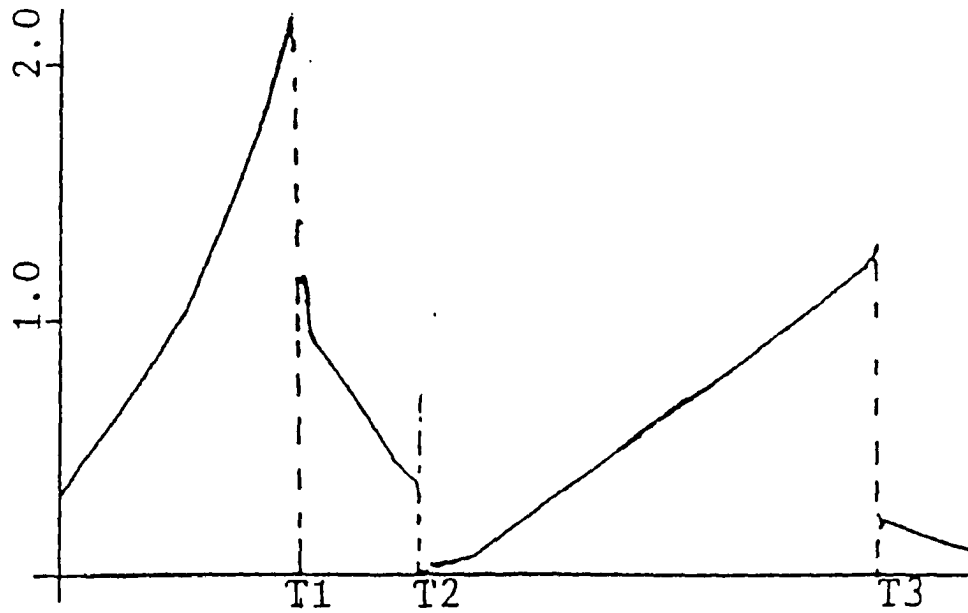
Fig. 7.1: Simulation record for the jump times, where T1= 3.76. T2=5.53, T3=12.61, and a typical trajectory of the unnormalized density $q_t(x)$ with $x \geq 0$.
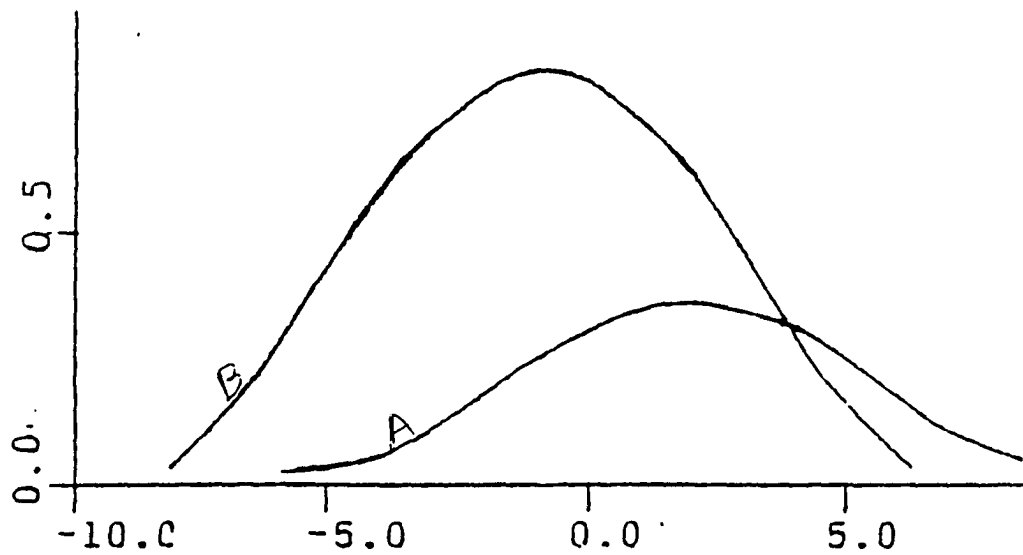


Fig. 7.2: Graph A is the graph of the conditional density $q_{3.75}(x)$. Graph B is the graph of the conditional density $q_{12.6}(x)$.

**Example 2** Let the state process $X_t$ be a Markov jump process and the state space consists of two points, 0 and 1. The corresponding matrix of transitions is of the form

$$\begin{pmatrix} q_{00} & q_{01} \\ q_{10} & q_{11} \end{pmatrix}$$

where $q_{00} = -q_{01}, q_{11} = -q_{10}$.

Let the observation process $Y_t$ be a jump process with values 0 and 1. Let the matrix of transition rates for the process $Y_t$ depend on the values of the process $X_t$ at the time t. The transition probability of the process $Y_t$ in this case are determined by $\gamma^{X_t}_{Y_t,1-Y_t}$ in the following form

$$P^{X_t}_{Y_t,1-Y_t}(t,t+\Delta t) = \gamma^{X_t}_{Y_t,1-Y_t}\Delta t + o(\Delta t)$$

$$P^{X_t}_{Y_t,Y_t}(t,t+\Delta t) = 1 - \gamma^{X_t}_{Y_t,1-Y_t}\Delta t + o(\Delta t).$$

We define the $\{T_n\}$ $n = 0,1,\cdots$, with $T_0 = 0, T_n < T_{n+1}$, to be the jump times of the Y process.

In [15] Yashin showed that $\pi_j(t)$, the conditional probability that $X_t = j$, given $Y_s$ for $s \le t$, satisfies the equation

$$d\pi_j(t) = \sum_{k=0}^{1} q_{kj}\pi_k(t)dt +$$
$$+ \pi_j(t)(\frac{\gamma^j_{Y_t,1-Y_t}}{\bar{\gamma}_{Y_t,1-Y_t}} - 1)(|dY_t| - \bar{\gamma}_{Y_t,1-Y_t})dt$$

for $j = 0,1$, where $\bar{\gamma}_{Y_t,1-Y_t} = \sum_{k=0}^{1} \pi_j(t)\gamma^k_{Y_t,1-Y_t}$.

Using the our results, we can have the following unnomalized linear equations

$$\rho_t\varphi = \rho_{T_i^-}\varphi - \int_{T_{i-1}}^t \rho_s\lambda(\cdot,Y_{s-})\varphi\,ds + \int_{T_{i-1}}^t \rho_s\cdot$$

$$[\varphi(X_{s-} + (-1)^{X_{s-}}) - \varphi(X_{s-})](q_{X_{s-}})\,ds \qquad (7.1)$$

$t \in [T_{i-1}, T_i)$ , and

$$\rho_{T_i}\varphi = \rho_{T_i^-}\lambda(\cdot,Y_{T_i^-})\varphi \qquad (7.2)$$

$i = 1, 2, \cdots$, with initial condition $\rho_0 = \pi_0$. where $q_o = q_{o1}, q_1 = q_{10}$.

We have simulated the solution of our filtering equation for the following parameters.

$$q_{01} = 0.8, \qquad q_{10} = 0.3, \qquad \gamma_{01}^0 = 0.1$$

$$\gamma_{10}^0 = 0.2, \qquad \gamma_{01}^1 = 0.325, \qquad \gamma_{10}^1 = 0.64,$$

with initial value $p_0 = P(X_0 = 0) = 0.4$ and T=22.0(seconds). The simulated $Y_t$ process starts in state 0, then transitions between the states 1 and 0 occurred at times $T_i, i = 1, 2, \cdots, 11$, which are shown in tsble 7.1. Using the above data and solving the linear ordinary differential equation and the linear system recursively, we find that the conditional probability of $X_{T_i-0.01} = 0$, given $Y_s, T_{i-1} \le s < T_i, i = 1, \cdots, 11$ (table 7.1).

We use the Runge-Kutta fourth order methords to solve the Yasin's nonlinear filtering equation with step size 0.01, we get approximate results, which have error order $O(10^{-8})$ with the above results.

We record the CPU time to compute $P(X_{20} = 0|Y_s, 0 \le s \le 20)$ in both methods as follows: 0.34159E-01 seconds (Yashin's procedure), and 0.21130E-02 seconds (our procedure). (All computations are carried on the NAS AS/9180 mainframe com-

puter at the Computation Center of Iowa State University.)

Table 7.1: The solutions of the filtering equations.

| Jumping Time $T_i$ | $P(X_{T_i - 0.01} = 0 | Y_s, T_{i-1} \leq s < T_i)$ |
|---|---|
| 1.077 | 0.34461076 |
| 2.843 | 0.32163193 |
| 8.870 | 0.31658635 |
| 9.280 | 0.19855931 |
| 12.673 | 0.30879970 |
| 14.166 | 0.30361160 |
| 14.949 | 0.22571545 |
| 15.080 | 0.11239086 |
| 16.195 | 0.22294083 |
| 17.644 | 0.28892280 |
| 21.111 | 0.31069577 |

# REFERENCES

[1] K.B. Athreya, W.H. Kliemann and G. Koch. *On sequential construction of solutions of stochastic differential equations with jump terms*. System and Control Letters, 10 (1988), 141-145.

[2] R. Boel, P. Varaiya and E.Wong. *Martingales on jump process, I and II*. SIAM, J. Control, 13 (1975), 999-1062.

[3] P. Bremaud. *Point Process and Queues: Martingale Dynamics*. Springer Verlag, Berlin, 1981.

[4] E. Cinlar and J. Jacod. *Semimartingales Defined on Markov Processes*. Lecture Notes in Control and Information Science, N.36, Springer Verlag, Berlin,1981.

[5] E. Cinlar, J. Jacod, P. Protter and M.J. Sharpe. *Semimartingales and Markov Process*. Zeitschrift für Warsch. theorie verw. Gebiete, 54 (1980), 161-219.

[6] M.H.A. Davis. *Lectures on Stochastic Control and Nonlinear Filtering*. Springer-Verlag, Berlin, 1984.

[7] R.J. Elliott. *Stochastic Calculus and Applications*. Springer-Verlag, Berlin, 1982.

[8] S.N. Ethier and T.G. Kurtz. *Markov Process: Characterization and Convergence*. Wiley, New York, 1986.

[9] J. Jacod and A.N. Shiryaev. *Limit Theorems for Stochastic Processes*. Springer-Verlag, Berlin, 1987.

[10] S. Karlin and H. Taylor. *A First Course in Stochastic Processes*. Academic Press, New York, 1975.

[11] W.H. Kliemann, G. Koch and F. Marchetti. *On the unnormalized solution of the filtering problem with counting process observations.* IEEE Trans. IT 36 (1990). 1415-1425.

[12] T.G. Kurtz and D. Ocone. *Unique characterzation of conditional distribution in nonlinear filtering.* Ann. Probab., 16 (1988), 80-107.

[13] R.S. Lipster and A.N. Shiryaev. *The Stochastics of Random Processes.* Springer Verlag, Berlin, 1982.

[14] H.L. Royden. *Real Analysis.* Second edition, Macmillan, New York, 1968.

[15] A.I. Yashin. *Filtering of jump process.* Automatikai Telemekhanika, 5 (1970), 725-730.

# PAPER II.

# DIMENSION REDUCTION FOR FILTERED MARKOV CHAINS

## ABSTRACT

It has been shown that the solution of a nonlinear filtering equation for continuous time Markov chain state processes $X_t$ and jump observation processes $Y_t$ can be obtained by solving the following filter equations (unnormalized density),

$$\dot{u}(t) = A^{(m_j)} u(t) \text{ for } t \in [T_{j-1}, T_j), u \in \Re^d$$

$$u(T_j) = B^{(m_j)} u(T_j^-) \text{ for } t = T_j,$$

where $T_j$ $j = 1, \cdots$, are the jump times of the process $Y_t$, $m_j \in J = \{1, 2, \cdots, n\}$ and $\lim_{t \uparrow T_j} u(t) = u(T_j^-)$, with certain restrictions on the matrices $A^{(m_j)}$ and $B^{(m_j)}$. This paper provides some necessary and sufficient conditions for the above filter equations to have lower dimensional realizations, i.e., the filter equations can be solved precisely in a manifold of dimension m, with $m < d$.

# 1. INTRODUCTION

Recently, Kliemann, Koch and Marchetti in [6] established a general procedure to solve the filtering problem for Markov semimartingale state processes and counting observation processes. In [5] Fan generalized their procedure to the case where observation processes are general jump processes. Fan also showed, by numerical simulation, that solving the filtering problem by this procedure can save considerable computation time as compared to solving the nonlinear filtering equations directly. In the implementation of filtering techniques, the computation time is a key factor, since larger computation time does not only lead to higher cost, but may also prevent nonlinear filters from being used in certain applications. One way to reduce the computation time for a given filtering problem is to reduce the the order of the filtering equations. It is the aim of this paper to investigate the existence of the reduced filter for the filtering equations presented in [5] and [6]. First, let us briefly introduce the procedure to solve the filtering problem for continuous time Markov chains state processes and jump observation processes in [5] and [6].

Consider a partially observable system, $(X_t, Y_t)$, $t \in \Re_+$, where $X_t \in \Re^d$ is the state process, and $Y_t \in \Re^p$ is the observable component respectively. We assume that the joint process $\xi_t = (X_t, Y_t)$, $t \in \Re_+$, has the structure of a general Itô process in

$\mathfrak{R}^m$ (m = d+p), i.e., $\xi_t$ has the representation:

$$\xi_t = \xi_0 + \int_0^t b(\xi_s)\,ds + \int_0^t c(\xi_s)\,dW_s + \int_0^t \int_Z K(\xi_{s-}, z)(N(ds, dz) - ds\nu(dz)) \quad (1.1)$$

where $W_t$ is a standard m-dimensional Wiener process and N is a standard Poisson random measure on $\mathfrak{R}_+ \times \mathfrak{R}$ with mean measure $ds\nu(dz)$. Besides the usual conditions on b and c, we assume that K is bounded. Existence and uniqueness of a strong Feller solution of (1.1) are guaranteed if the solution of the diffusion part of (1.1) has fourth moments [1].

Now if $Y_t$ is a jump process (in $\mathfrak{R}^1$), then in [5] and [6] it has been shown that

(i) $Y_t$ can be represented as

$$Y_t = \int_0^t \lambda(\xi_s)\,ds + M_t, \quad (1.2)$$

where $M_t$ and $\lambda(\xi_t)$ are the local intergrable martingale and the jump parameter of $Y_t$ respectively, for $t \in [0, T]$.

(ii) The filtered martingale problem for (1.1) and (1.2) has a unique solution.

(iii) The resulting equation for the optimal filter splits into two equations: one governs the evolution of the filter between jump times of $Y_t$; the other updates the filter at the jump times $T_i, i = 1, 2, \ldots$,

$$\mu_t \varphi = \mu_{T_{i-1}} \varphi + \int_{T_{i-1}}^t \mu_s L_X \varphi\,ds \quad (1.3)$$

$$- \int_{T_{i-1}}^t [\mu_{s-} \lambda(\cdot, Y_{s-})\varphi - \mu_{s-} \lambda(\cdot, Y_{s-})\mu_{s-}\varphi + \mu_{s-} R_X \varphi]\,ds$$

for $t \in [T_{i-1}, T_i)$, and

$$\mu_{T_i} \varphi = [\mu_{T_{i-}} \lambda(\cdot, Y_{T_{i-}})]^+ [\mu_{T_{i-}} \lambda(\cdot, Y_{T_{i-}})\varphi + \mu_{T_{i-}} R_X \varphi] \quad (1.4)$$

where $\mu_t$ is a probability measure , $R_X$ is a quadratic covariance operator and $L_X$ is the generator of the process $X_t$ (all those are defined in [5]), and $[\;]^+$ is defined as follows,

$$[a]^+ = \frac{1}{a} \text{ if } a \neq 0, \text{ and } [a]^+ = 0 \text{ if a } = 0.$$

(iv) A general procedure to solve the above nonlinear filtering equations is as follows. First, ignoring the nonlinear terms in (1.3) and (1.4), we have the following equations

$$\rho_t\varphi = \rho_{T_{i-1}}\varphi + \int_{T_{i-1}}^t \rho_s[(L_X - R_X)\varphi]\,ds - \int_{T_{i-1}}^t \rho_{s-}\lambda(\cdot,Y_{s-})\varphi\,ds \qquad (1.5)$$

for $T_{i-1} \leq t < T_i$, and

$$\rho_{T_i}\varphi = \rho_{T_{i-}}\lambda(\cdot,Y_{T_{i-}})\varphi + \rho_{T_{i-}}R_X\varphi \qquad (1.6)$$

where $i = 1,2,\cdots$. Then, it is shown in [5] and [6] that the normalization of the solutions of (1.5) and (1.6) yields the optimal nonlinear filter for the problem. which is the unique solution to the problem.

(v) If $X_t$, $t \geq 0$, is a Markov chain with state space $\{1,\cdots,d\}$, where d is a positive integer, and $Y_t, t \geq 0$, is a jump process with state space $\{a_1,\cdots,a_n\}$. where n is a positive integer, then we can identify the terms in the above filtering equations using the generators, which are defined as follows.

Define the joint generator $Q = (q_{(i,m_j)(k,m_l)})$, which generates the Markov process $\xi_t = (X_t,Y_t)$, by

$$q_{(i,m_j)(k,m_l)} = \lim_{h \downarrow 0} \frac{1}{h}P\{X_{t+h} = k, Y_{t+h} = a_{m_l}|X_t = i, Y_t = a_{m_j}\}.$$

where $m_j, m_l \in J = \{1,2,\cdots,n\}$.

Furthermore we define

$$A^{(m_j)} = (q_{(i,m_j)(k,m_j)}),$$ $d \times d$ matrix of $X_t$-transitions for $Y_t$ constant on

the time interval $[T_{j-1}, T_j)$, and

$$B^{(m_j)} = (q_{(i,m_j)(k,m_l)}),$$ where $j \neq l$, $d \times d$ matrix of $X_t$-transitions for $Y_{T_j}$

at the jump times $T_j$, and

$$\bar{S}^{(m_j)} = diag(\sum_k q_{(i,m_j)(k,m_l)}), \quad j \neq l,$$

$$S^{(m_j)} = diag(\sum_k q_{(i,m_j)(k,m_j)}) = -\bar{S}^{(m_j)}.$$

Identifying the terms in the above filtering equation, we have

$$L_X^{(m_j)} := A^{(m_j)} + B^{(m_j)},$$ which is a generator of $X_t$ given $Y_t = a_{m_j}$,

$$R_X^{(m_j)} := B^{(m_j)} - \bar{S}^{(m_j)}.$$

Furthermore

$$A^{(m_j)} = L_X^{(m_j)} - R_X^{(m_j)} - S^{(m_j)},$$

$$B^{(m_j)} = R_X^{(m_j)} + \bar{S}^{(m_j)},$$

and therefore we have the filter equation

$$\dot{u} = A^{(m_j)}u, \quad \text{for } Y_t = a_{m_j} \text{ in the interval } [T_{j-1}.T_j) \qquad (1.7)$$

$$u(T_j) = B^{(m_j)}u(T_j^-) \quad \text{for } Y_t - \text{jumps at time } T_j, \qquad (1.8)$$

where $j = 1, \cdots,$, and $T_j$ is a random variable.

For the implementation of equations (1.7,1.8) in particular for real time applications, we have to solve them numerically. Since the order of computation time depends on the dimension d for the ordinary differential equations (1.7) and for the updating operations (1.8), the computation time can be excessive, if the dimension

d is very large. Thus it is important to know, whether (1.7) and (1.8) has a lower dimensional realization, because any dimension reduction from d to $\hat{d} < d$ may save considerable computation time. Since the matrices $A^{(m_j)}$ and $B^{(m_j)}$ depend only on the system and observation dynamics, a lower dimensional realization can be computed beforehand, and at application time only the reduced order system has to be evaluted.

As we have seen, the idea of dimension reduction for (1.7,1.8) is different from that of minimal linear stochastic realizations. Losely speaking, minimal linear stochastic realization is about the syndissertation of stochastic processes such that the process is described equivalently by a state space with smallest possible dimension. However our purpose here is to analyze, whether the filter equations can be solved precisely in a manifold of dimension m, with $m < d$, without setting up a new system of dimension m.

In this paper it is investigated, whether there exists a lower dimensional realization for (1.7,1.8) under some conditions on $A^{(m_j)}$ and $B^{(m_j)}$. We will present some necessary and sufficient conditions for (1.7,1.8) to have lower dimensional realization by using three different approches : invariant linear subspaces, invariant integral submanifolds and exact criteria. The paper is organized as follows. In Section 2, assumptions, definitions and notations are given. In Section 3, it will be shown that if $B^{(m_j)}$ and $A^{(m_j)}$ are reducible, there are lower dimensional invariant linear subspaces for the problem, and if $B^{(m_j)} = exp(C^{(m_j)})$ , then there exist lower dimensional invariant integral manifolds for the problem. Finally, in Section 4, it will be investigated whether there exists any lower dimensional invariant submanifold for the problem without assuming $B^{(m_j)} = exp(C^{(m_j)})$ and whether there exists lower

dimensional invariant cone for the problem, when $B^{(m_j)}$ is irreducible.

## 2. PRELIMINARIES AND DEFINITIONS

We start this section by looking at the structures of $A^{(m_j)}$ and $B^{(m_j)}$ in

(1.7,1.8). From the constructions, we know that,

(i) $A^{(m_j)}$ has negative diagonal $q_{(i,m_j)(i,m_j)}$ and nonnegative off-diagonal $q_{(i,m_j)(k,m_j)}$,

$i \neq k, i, k = 1, 2, \ldots, d$.

(ii) All elements of the $B^{(m_j)}$ are nonnegative, and if the processes $X_t$ and $Y_t$ have

no common jumps, $B^{(m_j)}$ is a diagonal matrix.

We assume that:

(i) $B^{(m_j)} \in Gl(d, \Re) = \{G: G \text{ is } d \times d \text{ matrix}; \det(G) \neq 0\}, j = 1, \ldots$.

(ii) The jump times $\{T_j\}$ satisfy the condition $T_j - T_{j-1} > 0$, j=1,2,..., and $T_0 = 0$,

a.s.

Let us briefly comment on the above assumptions. Assumption (i) allows us

to restrict ourselves to the consideration of all nonsingular matrices $B^{(m_j)}$. This

assumption is without loss of generality since if there exists some matrix $B^{(m_j)}$

with rank$(B^{(m_j)}) = d_1 < d$ then we can get a $d_1$ dimensional invariant subspace

immediately. Assumption (ii) guarantees that $Y_t$ does jump at its jump time $T_j$.

We also should notice that the jump times $T_j$, $j = 1, \cdots$ are random, and

exponentially distributed. In order to analyze lower dimensional filters, one has to

consider (1.7) and (1.8) for arbitrary times $T_j > 0$, which leads to the following

solution manifolds.

First, it is obvious that the solution of

$(\sum 11)$ $\qquad \dot{u} = A^{(m_1)}u, \quad t \in [T_0, T_1)$

is of the form $M_1 = \{\phi_t^1 u_0 : t \in \Re\}$, where $\phi_t^1 = \exp(tA^{(m_1)})$ and $u_0 \in \Re^d$, which

defines a 1-dimensional manifold. Then at time $t = T_1$, the system has the first

jump: $B^{(m_1)}\phi_t^1 u_0$, which can be interpreted as the solution of

$(\sum 21)$ $\qquad \dot{u} = B^{(m_1)}A^{(m_1)}(B^{(m_1)})^{-1}u$, with $u(0) = B^{(m_1)}u_0$, and $t=T_1$.

The system continues via

$(\sum 22)$ $\qquad \dot{u} = A^{(m_2)}u$ with $u(0) \in B^{(m_1)}M_1 = \{B^{(m_1)}u : u \in M_1\}$.

The system $(\sum 21, \sum 22 )$ defines a manifold

$M_2 := \{\phi_{t2}^2 B^{(m_1)}\phi_{t1}^1 u_0 : t_1, t_2 \in \Re\}.$

Continuing the above procedures to n-1 jumps, we can get the system

$(\sum n1)$ $\qquad \dot{u} = B^{(m_1)}A^{(m_1)}(B^{(m_1)})^{-1}u,$ $\qquad\qquad$ $u(0)=B^{(m_1)}u_0$

$\qquad . \qquad\qquad . \qquad\qquad\qquad\qquad\qquad\qquad\qquad .$

$\qquad . \qquad\qquad . \qquad\qquad\qquad\qquad\qquad\qquad\qquad .$

$\qquad . \qquad\qquad . \qquad\qquad\qquad\qquad\qquad\qquad\qquad .$

$(\sum n(n-1))$ $\quad \dot{u} = B^{(m_{n-1})}A^{(m_{n-1})}(B^{(m_{n-1})})^{-1}u,$ $\quad u(0) \in B^{(m_{n-1})}M_{n-2}$

$(\sum nn)$ $\qquad \dot{u} = A^{(m_n)}u,$ $\qquad\qquad\qquad\qquad\qquad$ $u(0) \in B^{(m_n)}M_{n-1}$

The system $(\sum n1, \sum n2, \ldots, \sum nn )$ defines a manifold,

$M_n := \{\phi_{tn}^n B^{(m_{n-1})}...B^{(m_1)}\phi_{t1}^1 u_0 : t_j \in \Re, j = 1, 2, \ldots, n\}.$

By the construction of the $M_n$, we can see that $\dim(M_n) \leq d$, for every $n \in \mathcal{N}$.

where $\mathcal{N}$ is the set of positive integers. Our purpose is to analyze whether there exists

a lower dimensional submanifold of $\Re^d$, which contains $M_n$ for every for $n \in \mathcal{N}$ un-

der some conditions on $A^{(m_j)}$ and $B^{(m_j)}$. First, we want to introduce the following definitions.

**Definition 2.1** A filter of the form (1.7,1.8) has dimension m from $u_0 \in \Re^d$ if there is a submanifold $V \subset \Re^d$ with $\dim(V) = m$ such that $M_n(u_0) \subset V$, for every $n \in \mathcal{N}$.

**Remark 2.2** By Definition 2.1, if a filter has dimension m from $u_0 \in V$, then $u_0 \in V$ and $\phi_t^j u_0 \in V, B^{(m_j)} u_0 \in V$.

**Definition 2.2** A filter has dimension m if there is some submanifold V with $\dim(V)$ = m such that $M_n(u) \subset V$, for every $u \in V$ and $n \in \mathcal{N}$.

**Remark 2.2** It is easy to see that $M_n(u) \subset V$ for every $u \in V, n \in \mathcal{N}$ iff $\exp(t A^{(m_j)})V \subset V$ for every $t \in \Re$ and $B^{(m_j)}V \subset V, j \in \mathcal{N}$.

**Definition 2.3** We say that the filter (1.7,1.8) admits a dimension reduction, if the filter has dimension m with $m < d$.

# 3. DIMENSION REDUCTION BY LINEAR INVARIANT SUBSPACES AND INVARIANT INTEGRAL SUBMANIFOLDS

The equation (1.7,1.8) is a combination of a linear ordinary differential equation (1.7) and a linear updating operation (jump) (1.8). We can make use of just one of these two structures to obtain sufficient criteria for dimension reduction.

## · 3.1 Dimension Reduction by Linear Invariant Subspaces

In this subsection, we want to find a linear subspace $V \subset \Re^d$ such that $\dim(V) < d$ and $M_n(u) \subset V$ for every $n \in \mathcal{N}$ and $u \in V$. The following result gives a necessary and sufficient condition for such a lower dimensional linear subspace $V$ to exist.

**Theorem 3.1.1** (1.7,1.8) has a lower dimensional linear subspace iff there exists a transformation matrix $T \in Gl(d, \Re)$ such that $T \mathcal{U} T^{-1}$ is simulataneously reducible, where $\mathcal{U}$ is the algebra generated by $\{ A^{(m_j)}, B^{(m_j)}, j \in \mathcal{N} \}$, i.e., for all $C \in \mathcal{U}$, $T \mathcal{U} T^{-1}$ is of the form

$$\begin{pmatrix} * & * \\ 0 & * \end{pmatrix}$$

The theorem is an immediate consequence of the following lemma.

**Lemma 3.1.1** The following statements are equivalent.

(i) There exists an invariant subspace $V$ for (1.7,1.8).

(ii) There exists an invariant subspace V for $\mho$.

(iii) There exists a transformation matrix $T \in Gl(d, \Re)$ such that $T\mho T^{-1}$ is simultaneously reducible.

Proof: Let V be invariant for (1.7,1.8). By Remark 2.2, we have, $B^{(m_j)}V \subset V$ and $\exp(tA^{(m_j)})V \subset V$ for $t \in \Re, j \in \mathcal{N}$. Then $B^{(m_j)}V \subset V$, $\left(A^{(m_j)}\right)^n V \subset V$ for every $n, j \in \mathcal{N}$. This implies that V is invariant for $\mho$. Thus, we proved that "(i) —— (ii) ".

Now we show that "(ii)——(iii)". Choose a basis $e_1, \ldots, e_d$ for $\Re^d$ by extending a basis $e_1, \ldots, e_k$ of the $\mho$-invariant subspace V. Then for any $C \in \mho$, we have

$$T\mho T^{-1} = \begin{pmatrix} * & * \\ 0 & * \end{pmatrix},$$

where $T^{-1} = [e_1, \ldots, e_d]$. So, (ii) is true.

Finally, by linear algebra, it is easy to see that (iii) implies (i). Thus we proved the Lemma 3.1.

$$Q.E.D.$$

**Remark 3.1.1** (i) If $\mho$ is commutative, in particular if $A^{(m_j)}$ and $B^{(m_j)}$ commute for all $j \in \mathcal{N}$, then they have the same invariant subspaces and hence all $\mho$-invariant subspaces are direct sums of the invariant subspaces of $A^{(m_j)}$.

(ii) If $T\mho T^{-1}$ is simultaneously upper triangular, then the basis given by $T^{-1}$ defines all $\mho$-invariant subspaces.

(iii) If there exists a matrix $D \in gl(d, \Re)$, where $gl(d, \Re) = \{G : G$ is a $d \times d$ matrix $\}$, such that D commutes with $\mho$, i.e., $DC = CD$ for all $C \in \mho$, then range D is a $\mho$-invariant subspace.

Our next goal is to see the geometric characteristics of the system dynamics (1.7,1.8) on the above invariant lower dimensional subspace V. Consider the foliation of $\Re^d$ given by the affine subspaces

$$F_u = u + V \quad u \in \Re^d.$$

The intimate connection between the $\mho$-invariant subspace V and the foliation $F_u$ is given by the following theorem.

**Theorem 3.1.2** V is $\mho$-invariant iff the foliation $F_u = u + V$ is invariant under the system dynamics (1.7,1.8).

**Proof.** If V is $\mho$-invariant, then, by the Theorem 3.1, we know that $A^{(m_j)}$ and $B^{(m_j)}$ take the following forms

$$A^{(m_j)} = \begin{pmatrix} A_{11}^{(m_j)} & A_{12}^{(m_j)} \\ 0 & A_{22}^{(m_j)} \end{pmatrix}$$

and

$$B^{(m_j)} = \begin{pmatrix} B_{11}^{(m_j)} & B_{12}^{(m_j)} \\ 0 & B_{22}^{(m_j)} \end{pmatrix}$$

where $A_{11}^{(m_j)}$ and $B_{11}^{(m_j)}$ are k×k matrices and k=dim(V). Let $(u_1, \ldots, u_d)$ be the corresponding linear coordinate functions on $\Re^d$, i.e., the basis such that the matrices $A^{(m_j)}$ and $B^{(m_j)}$ have the above structures. We write $u^1 = (u_1, \ldots, u_k)^T$ and $u^2 = (u_{k+1}, \ldots, u_d)^T$. Then the system dynamics (1.7,1.8) can be written

$$
\begin{aligned}
\dot{u}^1 &= A_{11}^{(m_j)} u^1 + A_{12}^{(m_j)} u^2 \\
\dot{u}^2 &= A_{22}^{(m_j)} u^2
\end{aligned}
\tag{3.1}
$$

and

$$u^1(T_j) = B_{11}^{(m_j)} u^1(T_j^-) + B_{12}^{(m_j)} u^2(T_j^-) \qquad (3.2)$$

$$u^2(T_j) = B_{22}^{(m_j)} u^2(T_j^-).$$

It is easy to see that the foliation $F_u = u + V$ is the same as the $F_u = \{u = (u^1, u^2)|u^2 = \text{constant vector}\}$. Consider $q_1$ and $q_2$ on the same leaf, i.e., $q_1 - q_2 \in V$, or equivalently $u^2(q_1) = u^2(q_2)$. Then, by (9.1) and (10.1), we can see that the solutions $u(t,0,q_1)$ and $u(t,0,q_2)$ at every time instant $t \geq 0$ will be again on the same leaf. Thus the foliation $F_u$ is invariant under the system dynamics (1.7,1.8). Reversing each of the above statements, we can show the "only if" part.

$$Q.E.D.$$

The significance of this theorem is that,

(i) it shows that the foliation is a set of stacked V subspaces, and then the quotient of $\Re^d$ with respect to the foliation can be identified with $V'$ ($V'$ is a subspace of $\Re^d$ such that $\Re^d = V \oplus V'$ ), and

(ii) it provides a method to obtain the reduced dynamical syatem, namely, to get the reduced dynamical system we only need to project the linear vector field with respect to this foliation.

## 3.2 Dimension Reduction by Invariant Integral Submanifolds

In this subsection, we consider the case that the matrices $B^{(m_j)}$ have real logarithms, $B^{(m_j)} = \exp(C^{(m_j)})$. In this case, the solutions of (1.7,1.8) live on the maximal integral manifolds of the Lie algebra $\mathcal{L}$ generated by $\{A^{(m_j)}, C^{(m_j)}, j \in \mathcal{N}\}$,

and then the rank $\mathcal{L}(u)$ for $u \in \Re^d$, determines the dimension of the reduced system starting at u. First, we introduce some notations.

Let $\mathcal{M} \subset gl(d, \Re)$ be an arbitrary subset, $\{\mathcal{M}\}_{\mathcal{A}}$ denotes the Lie algebra generated by $\mathcal{M}$. If $\mathcal{U}$ is a set of matrices in $Gl(d, \Re)$, we let $\{\mathcal{U}\}_{\mathcal{G}}$ denote the multiplicative matrix group generated by $\mathcal{U}$, i.e., the smallest group in $Gl(d, \Re)$ which contains $\mathcal{U}$ and is closed under multiplication and inversion. If $\mathcal{M} \subset gl(d, \Re)$, then one can show [3] that

$$\{\exp(\mathcal{M})\}_{\mathcal{G}} = \{g: g = \exp(M_p) \ldots \exp(M_1), M_j \in \mathcal{M}, p = 1, \ldots\}.$$

The rank of any subalgebra $g$ of $gl(d, \Re)$ at $u \in \Re^d$ is the dimension of the subspace $\{Mu : M \in g\} \subseteq \Re^d$. The dimension of a Lie algebra is the dimension of its vector space.

Now we turn to our problem. First, we should notice that the matrix B has a real logarithm iff all eigenvalues of B are real and positive [8,p312]. If $B^{(m_j)} = \exp(C^{(m_j)})$, then $B^{(m_j)}$ is the fundamental matrix of the ordinary differential equation $\dot{u}(t) = C^{(m_j)}u(t)$ for $t = 1$. Hence one can consider dimension reduction for

$$\dot{u}(t) = A^{(m_j)}u(t) \tag{3.3}$$

$$\dot{u}(t) = C^{(m_j)}u(t). \tag{3.4}$$

Here, the solution is given by (3.3) at between jump times, and are described by (3.4) at jump times as the solution of (12) between $t = 0$ and $t = 1$.

If (11,12) has a dimension reduction, i.e., there exists a lower dimensional invariant submanifold V for (11,12), then we have

$\exp(t A^{(m_j)})u \in V$ and $\exp(t C^{(m_j)})u \in V$ for all $t \in \Re$ and $j \in \mathcal{N}$,

and hence $\sigma_t^j u \in V$ for all $t \in \Re$ and $exp(C^{(m_j)})u \in V$ for all $j \in \mathcal{N}$.

This implies

$$\sigma_t^j V \subset V \text{ and } B^{(m_j)} V \subset V \text{ for all } t \in \Re \text{ and } j \in \mathcal{N}.$$

Hence, recalling Remark 2.1, we know that V is also invariant for (1.7,1.8). Thus (1.7,1.8) admits a dimension reduction if (11,12) does.

**Theorem 3.2.1** Let $ln(B^{(m_j)}) = C^{(m_j)} \in gl(d, \Re), j \in \mathcal{N}$ exist. Then (1.7),(1.8) reduce the dimension from $u$ $in \Re^d$ iff $\dim(\{A^{(m_j)}, C^{(m_j)}, j \in \mathcal{N}\}_A)(u) < d$.

Proof: By the above notations, we have [3],

$$\{\exp\{A^{(m_j)}, C^{(m_j)}, j \in \mathcal{N}\}\}_G = \{\exp\{A^{(m_j)}, C^{(m_j)}, j \in \mathcal{N}\}_A\}_G$$

Clearly, $\text{Rank}(\{exp\{A^{(m_j)}, C^{(m_j)}, j \in \mathcal{N}\}_A\}_G)(u) = \dim(\{A^{(m_j)}, C^{(m_j)}, j \in \mathcal{N}\}_A)(u)$. Thus, if $\dim(\{A^{(m_j)}, C^{(m_j)}, j \in \mathcal{N}\}_A)(u) < d$ for all $u \in \Re^d$, then there is submanifold V with $\dim(V) < d$ such that $\mathcal{U}(u) \subset V$. That is, (1.7,1.8) admits a dimension reduction.

$$Q.E.D.$$

**Corollary 3.2.1** Under the assumption of the above theorem, (1.7,1.8) admits a dimension reduction iff $\dim(\{A^{(m_j)}, C^{(m_j)}, j \in \mathcal{N}\}_A)(u) < d$ for all $u \neq 0, u \in \Re^d$.

# 4. EXACT CRITERIA FOR LOWER DIMENSIONAL FILTERS

In this section, we will give some explicit criteria involving the matrices $\{A^{(m_j)}, B^{(m_j)}\}$, $j \in \mathcal{N}$ on the existence of a lower dimensional filter for (1.7,1.8).

## 4.1 Conditions for $M_n$ to be an Invariant Manifold

Observing the construction of $M_n$ in Section 2, we can see that $M_n$ consists of all the trajectories of (1.7,1.8). It is of interest to know if $M_n$ itself reduces the dimension for the problem. In this subsection, we look for conditions on the matrices $\{A^{(m_j)}, B^{(m_j)}, j \in \mathcal{N}\}$ such that $M_n$ is an invariant manifold for (1.7,1.8). First, we consider the special case $A^{(m_j)} = A, B^{(m_j)} = B$ for j=1,2, .... In this case definitions 2.1-2.2 are equivalent to the following conditions,

$$\phi_t M_n \quad \subset \quad M_n \text{ for all } t \in \Re \qquad (4.1)$$

$$B(M_n) \quad \subset \quad M_n$$

where $\phi_t = \exp(tA)$.

By the above definition, we know that a filter has dimension n from $u_0 \in \Re^d$ iff given $t_1, \cdots, t_n \in \Re$ there exist $s_1, \cdots, s_n \in \Re$ such that

(i) $\phi_t \phi_{t_n} B \ldots B \phi_{t_1} u_0 \subset \phi_s B \phi_{s_n} \ldots B \phi_{s_1} u_0$, this is always satisfied via construction,

(ii) $\qquad B \phi_{t_n} B \ldots B \phi_{t_1} u_0 = \phi_{s_n} B \ldots B \phi_{s_1} u_0.$

The task is to find explicitly conditions in terms of A and B for this. Before presenting the main results for this section, we remark a trivial case. If tr(A)= 0, then, by the structures of the generator Q and the matrices A and B, we know that all states of the joint processes $\xi_t = (X_t, Y_t)$ are absorbing and so are all states of the state processes $X_t$. Therefore, it is of no interest to do the estimation for the problem. Thus, we assume tr(A)$\neq$0 in the rest of this subsection.

**Theorem 4.1.1** (1) $M_n$ is B-invariant iff AB=BA and if $u_0 \in M_n$, then we have $Bu_0 \in M_n$).

(1) If $M_n$ is B-invariant, then dim($M_n$)=1.

To show this theorem, we need the following lemma.

**Lemma 4.1.1** If $M_n$ is B-invariant, i.e., given $t_1, \ldots, t_n \in \Re$, there exist $s_1, \ldots, s_n \in \Re$ such that $B\phi_{t_n} B \ldots B\phi_{t_1} = \phi_{s_n} B \ldots B\phi_{s_1}$, then

(i) det(B)>0,

(ii) if tr(A)$\neq$ 0, then $\sum_{k=1}^n s_k - \sum_{k=1}^n t_k = \frac{ln(det(B))}{tr(A)}$.

Proof: Suppose, given $t_1, \ldots, t_n \in \Re$, there exist $s_1, \ldots, s_n \in \Re$ such that

$$B\phi_{t_n} \ldots B\phi_{t_1} = \phi_{s_n} B \ldots B\phi_{t_1}.$$

Then det($B\phi_{t_n} \ldots B\phi_{t_1}$)= det($\phi_{s_n} \ldots B\phi_{s_1}$), thus,

$$(det(B))^{n+1} det(exp(\sum_{k=1}^n t_k)A) = (det(B))^n det(exp(\sum_{k=1}^n s_k)A).$$

Since det(B)$\neq$0, then det($B$) = exp$((\sum_{k=1}^n s_k - \sum_{k=1}^n t_j)tr(A))$, and we have det(B)>0 and

$$(\sum_{k=1}^n s_k - \sum_{k=1}^n t_k)tr(A) = ln(det(B))$$

If tr$(A)\neq 0$. $\sum_{k=1}^{n} s_k - \sum_{k=1}^{n} t_k = \frac{ln(det(B))}{tr(A)}$.

$$Q.E.D.$$

We should notice that if all the eigenvalues of B are positive, then

$$\sum_{k=1}^{n} s_k - \sum_{k=1}^{n} t_k = \frac{(\sum_{j=1}^{d} ln(\lambda_j(B)))}{\sum_{j=1}^{d} \lambda_j(A)}$$

where $\lambda_j(B)$ and $\lambda_j(A)$, j=1, ..., d, are the eigenvalues of A and B respectively.

Now we prove Theorem 4.1.1.

Proof: Suppose AB=BA and $u_0$ is B-invariant.

Let u=exp$(t_n A)B...B$exp$(t_1 A)u_0 \in M_n$, then ,

$$Bu = B\exp(t_n A)B...B\exp(t_1 A)u_0 = \exp(t_n A)B...B\exp(t_1 A)(Bu_0) \in M_n,$$

since $u_0$ is B-invariant. Thus $M_n$ is B-invariant.

If $M_n$ is B-invariant, it is clear that $u_0$ is B-invariant and hence we only need to show BA=AB.

Given $t_1,...,t_n \in \Re$, there exists $s_1,...,s_n \in \Re$ such that

$$B\exp(t_n A)B...B\exp(t_1 A) = \exp(s_n A)B...B\exp(s_1 A)$$

then

$$\frac{\partial}{\partial t_n}(B\exp(t_n A)...B\exp(t_1 A) = \frac{\partial}{\partial t_n}(\exp(s_n A)B...\exp(s_1 A))$$

Since $t_1,...,t_n$ and $s_1,...,s_n$ are arbitrary real numbers and, by Lemma 4.1.1.
$s_n = t_n + \sum_{k=1}^{n-1}(t_k - s_k) + \frac{ln(det(B))}{tr(A)}$, we have
$BA\phi_{t_n}B...B\phi_{t_1} = A\phi_{s_n}B...B\phi_{s_1}$ and hence $BA\phi_{t_n}B...B\phi_{t_1} = AB\phi_{t_n}...B\phi_{t_1}$.

Because det(B)$\neq$0, we have BA=AB. Thus we proved (1.1).

Now, it is clear, dim($M_n$)=1 for all $n \geq 1$.

$$Q.E.D.$$

For general case, we have the following corollary.

**Corollary 4.1.1** $M_n$ is an invariant manifold iff $A^{(m_j)}A^{(m_k)} = A^{(m_k)}A^{(m_j)}$, $B^{(m_j)}B^{(m_k)} = B^{(m_k)}B^{(m_j)}$, and $A^{(m_j)}B^{(m_k)} = B^{(m_k)}A^{(m_j)}$ for all j=1,..., k=1,2, ..., and $u_0$ is $B^{(m_j)}$-invariant, j=1,2,... and then dim($M_n$)=1.

## 4.2 Dimension Reduction Using the Structures of the Matrices A and B

From Section 3.1, we know that if $B^{(m_j)}$, $A^{(m_j)}$, $j = 1,\ldots,$ are reducible, we can find a lower dimensional invariant subspace for the problem. It is also of interest to know, whether there is any lower dimensional submanifold other than a linear subspace for the problem. It is the aim of this section to investigate, whether there is such lower dimensional manifold for (1.7,1.8). For simplicity, we only consider $A^{(m_j)} = A, B^{(m_j)} = B$ for j=1,.... First let us assume that B is irreducible.

Recalling the dimension reduction for our problem, we are interested in finding some subspace or submanifold $V \subset \Re^d_+$ with dim(V)<d such that

$$BV \subset V \text{ and } \phi_t V \subset V \text{ for all } t \geq 0 ,$$

and the matrix B satisfies the conditions in Section 2. Clearly, if B is irreducible, then there is no lower dimensional invariant linear subspace for the problem. Since

$\mathfrak{R}^d_-$ is a cone (a cone in a real vector space is a subset closed under addition and multiplication by positive real numbers ) and exp(tA) is nonnegative for all $t \geq 0$, then it is natural to look for the existence of a lower dimensional cone V in $\mathfrak{R}^d_-$ such that V or $\partial(V)$ (the boundary of V ) is invariant for the problem. The following results will give a negative answer to the above question, if B is irreducible.

Before we state our results for the problem, we introduce some notations.

Let M be a d×d matrix, we define

$$\sigma(M) = \text{the spectrum of M,}$$

$$\kappa(M) = \max\{|\lambda| : \lambda \in \sigma(M)\}.$$

**Theorem 4.2.1** Let B be d×d irreducible matrix and all its entries be nonnegative (we use the notation $B \geq 0$) with tr(B)>0. If there is a subcone $V \subset \mathfrak{R}^d_+$ such that $BV \subset V$, then eithere dim(V)=d or dim(V)=0.

Proof: Since $B \geq 0$, it is clear that $B(\mathfrak{R}^d_+) \subset \mathfrak{R}^d_+$. Since B is irreducible, B leaves no subcone $V \subset (\mathfrak{R}^d_+)^o$ with dim(V)<d and $BV \subset V$ invariant (see [2]), where $(\mathfrak{R}^d_+)^o$ is the interior of $\mathfrak{R}^d_+$.

Since B≥0, is irreducible and tr(B)>0, $\kappa(B) \in \sigma(B)$ is a simple eigenvalue and $\kappa(B) > |\lambda|$ for all $\lambda \in \sigma(B), \lambda \neq \kappa(B)$ [7,p11] and hence there exists $n \in \mathcal{N}$ such that $B^n > 0$ [8,p546]. This implies that, for any $u \in \partial(\mathfrak{R}^d_+ \backslash \{0\})$, $B^n u \in (\mathfrak{R}^d_+)^o$ and hence if $S \subset \partial(\mathfrak{R}^d_+)$ such that $BS \subset S$, then $S = \{0\}$. Thus we proved the theorem.

$$Q.E.D$$

We now turn our attention to the case when B is reducible. If B is reducible,

there exists a permutation matrix P such that

$$P^T BP = \begin{pmatrix} B11 & O \\ B21 & B22 \end{pmatrix}$$

where $B11 = diag(B_{11}, \ldots, B_{kk})$, the blocks $B_{jj}$, j=1,...,k are irreducible and $n_j \times n_j$ matrices. Let

$$\bar{P}_j^{-1} B_{jj} \bar{P} = J(B_{jj}) \quad j = 1, \ldots, k$$

where $J(B_{jj})$ is the Jordan form of $B_{jj}$. Since $B_{jj}$ is irreducible, $J(B_{jj})$ has only one block. Let

$$\bar{P} = diag(\bar{P}_1, \ldots, \bar{P}_k, I, \ldots, I) \text{ and } T = P\bar{P}$$

then

$$T^{-1}BT = \begin{pmatrix} J(B11) & O \\ * & * \end{pmatrix}.$$

It is easy to see that there exists lower a dimensional invariant submanifold for (1.7,1.8) iff there is one for       .

$$\dot{u} = T^{-1}ATu \quad t \in [T_{j-1}, T_j) \tag{4.2}$$

$$u(T_j) = T^{-1}BTu(T_j^-) \quad t = T_j. \tag{4.3}$$

If it happens that $T^{-1}AT = diag(J_1(A), \ldots, J_p(A))$, where $J_k(A), k = 1, \ldots, p$ are the Jordan blocks of A, and some $J_k(A)$ has the same structure as the corresponding $J_k(B)$, then there exist lower dimensional invariant submanifolds other than a linear subspace for (1.7,1.8) in the following cases.

(a) There is some $J_j(A)$, $1 \leq j \leq p$, which has the form

$$\begin{pmatrix} \beta & \alpha \\ -\alpha & \beta \end{pmatrix}$$

(b) $J_j(A) = \lambda I_{n_j \times n_j}$, for some $1 \leq j \leq p$.

Summary the above results, we have the following theorem.

**Theorem 4.2.2** Suppose there is a nonsingular matrix T such that $T^{-1}BT$ has the above form and $T^{-1}AT$ has Jordan form as the above, then there exists lower dimensional invariant submanifold for (1.7,1.8) if the following conditions are satisfied

(i) the Jordan form of $T^{-1}AT$ has at least one of the stuctures of (a)-(b), and

(ii) the corresponding $J(B_{jj})$ has the same structure as the $T^{-1}AT$ in (i).

Before ending this paper, we make two remarks.

(i) If all $B^{(m_j)}$ are diagonal matrices, then, by the construction of $B^{(m_j)}$, we know that the assumptions in Theorem 3.2.1 and Corollary 3.2.1 are satisfied and hence we can use invariant integral submanifolds to reduce the dimension for the problem. If $B^{(m_j)} = c_{m_j} I$, where $c_{m_j}$ is a positive constant and I is the $d \times d$ identity matrix, then we can use Remark 3.1.1 (i) to find invariant linear subspaces . Moreover the assumptions in Theorem 4.1.1 and Corollary 4.1.1 are automatically satisfied, and therefore the lower dimensional manifolds for (1.7,1.8) always exsit.

(ii) As we have seen, exact dimension reduction is rare situation, i.e., generically not possible. We will develop an approximate approach to the dimension reduction for the problem in a forthcoming paper.

# REFERENCES

[1] K. B. Athreya, W. H. Kliemann, and G. Koch *On sequential construction of solutions of stochastic differential equations with jump terms* Syst. Contr. Lett.. 10, pp. 141-146, 1988.

[2] G. P. Barker *On matrices having an invariant cone* Czchoslovak Mathematical Journal, 22, pp.49-68, 1972

[3] R. W. Brockett *System theory on group manifolds and coset spaces* SIAM J. Control, 10, pp. 226-284, 1972.

[4] R. W. Brockett *Lie algebras and lie groups in control theory* Proceedings of NATO Advanced Study Institute held at London, England, Aug. 27-Sep. 7.1973.

[5] K. Fan *On the unnormalized solution of the filtering problem with jump process observations* IEEE Trans. Infor. Theory (submitted).

[6] W. H. Kliemann, G. Koch, and F. Marchetti *On the unnormalized solution of the filtering problem with counting process observations* IEEE Trans. Infor. Theory, 36, pp. 1415-1425, 1990.

[7] H. Minc *Nonnegative Matrices* John Wiley and Sons, Inc., New York, 1988.

[8] P. Lancaster and M. Tismenetsky *The Theory of Matrices* Academic Press, New York, 1985.

PAPER III.

# APPROXIMATE DIMENSION REDUCTION FOR FILTERED MARKOV CHAINS

# ABSTRACT

It has been shown that the solution of a nonlinear filtering equation for continuous time Markov chain state processes $X_t$ and jump observation processes $Y_t$ can be obtained by solving the following filter equations (unnormalized density),

$$P'(t) = A^{(m_j)} P(t) \quad \text{for } t \in [T_{j-1}, T_j), P(t) \in \Re^d$$

$$P(T_j) = B^{(m_j)} P(T_j^-) \quad \text{for } t=T_j,$$

where d is a positive integer, $T_j$ $j = 1, \cdots$, are the jump times of the process $Y_t$, $m_j \in \{1, 2, \cdots, n\}$ and $\lim_{t \uparrow T_j} P(t) = P(T_j^-)$, with certain restrictions on the matrices $A^{(m_j)}$ and $B^{(m_j)}$.

This paper presents an efficient and applicable procedure for approximate dimension reduction of the above filter equations. The conditions under which the approximate optimal filter converges to the optimal filter of the problem under consideration and coresponding error estimates also are given.

# 1. INTRODUCTION

Filtering is the general theory of extracting information about a prescribed quantity of interest from noisy observations. Applications include such diverse areas as communications, radar image analysis, sonar tracking, flight control and general parameter for random systems. During the last three decades linear filtering algorithms, based on the Kalman-Bucy equations, have been implemented in many systems. For the nonlinear filtering problem genuine nonlinear filtering theories have been available for over twenty years, but there exists a wide gap between the well developed theory and its applications. The crucial task in bridging this gap is the development of efficient numerical methods, which obviously should be implementable on modern digital computers. However, the design of useful numerical algorithms is complicated by the mathematical complexity of the filter equations.

Recently, an efficient and applicable approach to study the implementation of nonlinear filtering algorithms has been presented in [3] and [7]. The authors have established a general procedure to solve the nonlinear filtering problem for Markov semimartingale state processes and general jump observation processes. They rewrite the resulting nonlinear equation for the optimal filter into two equations, one describes the evolution of the filter between the observation jump times, the other one updates the filter at the jump times. Instead of solving the nonlinear filtering equations, they

first solve a linear integral equation and a linear algebraic equation recursively, then normalize the solution which turns out to be the solution of the original nonlinear filtering equation. For practical and numerical purposes, one is more interested in the case, where the state process is a continuous time Markov chain. In this case, by using the procedure in [3,7], one can identify the nonlinear filtering equation as one linear ordinary differential equation and one linear algebraic system. First , let us briefly introduce the procedure to get these equations.

Let $X_t$ , $t \geq 0$ be a continuous time Markov chain with state space $\{1, 2, \cdots, d\}$, where d is a positive integer, and $Y_t$, $t \geq 0$, a jump process with state space $\{a_1, a_2, \cdots, a_n\}$, where n is a positive integer, and let the joint process $\xi_t = (X_t, Y_t)$ be a continuous time Markov chain with the generator $Q = (q_{(i,m_j)(k,m_l)})$, where

$$q_{(i,m_j)(k,m_l)} = \lim_{h \downarrow 0} \frac{1}{h} P(X_{t+h} = k, Y_{t+h} = a_{m_l} \mid X_t = i, Y_t = a_{m_j})$$

and $m_j, m_l \in \{1, 2, \cdots, n\}$.

Furthermore, we define $A^{(m_j)} = (q_{(i,m_j)(k,m_j)})$, $d \times d$ matrix of $X_t$-transitions for $Y_t$ constant on the interval $[T_{j-1}, T_j)$, and $B^{(m_j, m_k)} = (q_{(i,m_j)(k,m_l)})$, where $m_j \neq m_l$, $d \times d$ matrix of $X_t$ transitions for $Y_{T_j}$ jump.

Then, by identifying the terms in the nonlinear filtering equations, we can get the following filter equations (unnormalized density),

$$P'(t) = A^{(m_j)} P(t), \qquad \text{for } t \in [T_{j-1}, T_j) \qquad (1.1)$$

$$P(T_j) = B^{(m_j, m_k)} P(T_j^-) \quad \text{for } t = T_j \qquad (1.2)$$

In [3], it has been shown, by numerical simulations, that solving the filtering problem by the above procedure can save considerable computation time as com-

pared to solving the nonlinear filtering equations directly in the case where the state process has a small number of states. However, in many applications the state process $X_t$ has a larger number of states. This will lead to significant computation problems; e.g. to intolerable computation time in real time applications, or excessive computation time for other problems. Both problems prevent the above procedure from being used in applications. Therefore, it is essential to find a reduced order filter that estimates only the required minimum number of states. In [6] it has been shown that exact dimension reduction for (1) and (2) is a rare situation, i.e., generically not possible. In the present paper, we will propose an efficient and applicable procedure to reduce approximately the dimension for (1) and (2). There are several model reduction methods for dynamical systems available, for example, the method of principal component analysis [8], the method of balanced energy [1] and the method of balanced stochastic realization [2], etc. , but these existing methods of model reduction deal with either the continuous time systems or discrete time systems. Our systems (1) and (2) are combination of a continuous and a discrete times system. Moreover, the variety of $A^{(m_j)}, B^{(m_j, m_k)}$ and the random jump times make the problem quite complicated.

# 2.   PRELIMINARIES AND NOTATIONS

In this section we shall introduce some assumptions and look at the structures of the state and observation processes.

## 2.1   Assumptions and Comments

First, we assume:

(i) The state process is an irreducible continuous time Markov chain.

(ii) $B^{(m_j, m_k)} \in Gl(d, \Re) = \{G : G$ is a $d \times d$ matrix; $\det(G) \neq 0\}$, where $m_j, m_k \in \{1, 2, \cdots, n\}$.

(iii) The random jump times $\{T_j\}$ satisfy the condition $T_j - T_{j-1} > 0$, where $j = 1, 2, \cdots$, and $T_0 = 0$, a.s.

Let us briefly comment on the above assumptions. Assumption (i) is without loss of generality, since, otherewise, we can use the arguments in [6] to get lower dimensional filter equations for the problem. Assumption (ii) allows us to restrict ourselves to considering all nonsingular matrices $B^{(m_j, m_k)}$. This assumption holds also without loss of generality since, if there exists some matrix $B^{(m_j, m_k)}$ with rank($B^{(m_j, m_k)}) = d_1 < d$, then we can get a $d_1$ dimensional filter equation immediately. Assumption (iii) guarantees that $Y_t$ does jump at its jump time $T_j$.

From the construction of $A^{(m_j)}$ and $B^{(m_j, m_k)}$, we know that.

(i) $A^{(m_j)}$, $m_j \in \{1, 2, \cdots, n\}$, are dominant diagonal matrices with negative diagonal elements and nonnegative off-diagonal elements, and

(ii) all entries of the matrices $B^{(m_j, m_k)}$ are nonnegative and if the processes $X_t$ and $Y_t$ have no common jumps, then $B^{(m_j, m_k)}$ are diagonal matrices.

In the following two sections, we want to look at the structures of the state and the observation process.

## 2.2 Fast and Slow Transition States of a Continuous Time Markov Chain

In this subsection we will classify the states of the continuous time Markov chain state·process into fast and slow states in terms of their transition rates. First, we introduce the folllowing definitions.

**Definition 2.1** Let $X = \{X_t, t \in \Re^+\}$ be a continuous time Markov chain with transition rates matrix $\Lambda = (\lambda_{i,j})$, $i, j = 1, 2, \cdots, d$, $\sum_{i \neq j} \lambda_{i,j} = -\lambda_{j,j}$.

(i) We define $\rho = \frac{|trace(\Lambda)|}{d}$ to be the average transition rate of the process X.

(ii) Denote by $X(i), i = 1, 2, \cdots, d$ the state i of X. Then the state $X(i)$ is called a fast state if

$$|\lambda_{i,i}| \geq \alpha \rho,$$

where $\alpha$ is a positive constant, to be chosen later. The state $X(i)$ is called a slow state, if it is not a fast state.

**Definition 2.2** Let

$$\Omega_F(X) = \text{the set of the fast states of the process X},$$

$$\Omega_S(X) = \text{the set of the slow states of the process } X$$

(i) We say that state $X(i)$ is connected to state $X(j)$ via fast transitions if $\lambda_{ij} \geq \alpha\rho/(d-1)$, otherwise it is called a slow transition.

(ii) The subset $\Omega_F(X)$ is called a fast recurrent subset if any $X(i) \in \Omega_F(X)$ is connected via fast transitions to any other $X(j) \in \Omega_F(X)$, but is connected to any $X(j) \in \Omega_S(X)$ only by slow transitions.

(iii) The subset $\Omega_F(X)$ is called a fast transient subset if any $X(i) \in \Omega_F(X)$ is connected to the states $X(j) \in \Omega_S(X)$ by at least one fast transition.

**Definition 2.3**  We say that two positive real numbers $r_1$ and $r_2$ have comparable magnitude, if $\frac{r_1}{r_2} = c \times 10^{-1}$, where $1 \leq c \leq 10$. Two nonzero matrices $R1_{m \times n}$ and $R2_{m \times n}$ are of norms of comparable magnitude if $\|R1\|_\infty$ and $\|R2\|_\infty$ are of comparable magnitude.

By reordering the states of the original Markov chain following the partition into slow states and fast states, the original trnasition matrix $\Lambda$ assumes the following block form

$$\begin{pmatrix} \Lambda_{SS} & \Lambda_{FS} \\ \Lambda_{SF} & \Lambda_{FF} \end{pmatrix}$$

**Remark 2.1**  If $\alpha > 1$, let $\Lambda_{FF} = \alpha\bar{\Lambda}_{FF}$ and $\Lambda_{FS} = \alpha\bar{\Lambda}_{FS}$. If the $\Omega_F(X)$ is a fast recurrent subset, then $\Lambda_{SS}$, $\bar{\Lambda}_{FF}$, $\Lambda_{FS}$ and $\Lambda_{SF}$ have norms of comparable magnitude. If the $\Omega_F(X)$ is a fast transient subset, then $\Lambda_{SS}$, $\bar{\Lambda}_{FS}$, $\Lambda_{SF}$ and $\bar{\Lambda}_{FF}$ have norms of comparable magnitude.

For a continuous time Markov chain, one of the important statistics is its sojourn time. The following result will show the intimate connection between the sojourn times and the transition rates.

Let

$$S_T(i) = \int_0^T I_{(X(s)=i)}(s)ds$$

where $T > 0$. Then, $S_T(i)$ is the sojourn time of the process X in state i during $[0, T]$.

**Proposition 2.1**    If $|\lambda_{jj}| > |\lambda_{ii}| > 0, 1 \le i,j \le d$, where $\lambda_{ii}, i = 1, \cdots, d$, are defined in Definition 2.1, then

(i) $E(S_T(i)) > E(S_T(j))$ ,

(ii) $P(X(s) = i) > P(X(s) = j)$.

Proof    The result follows from the following facts

(i) $E(S_T(i)) = \frac{1}{|\lambda_{ii}|}(1 - exp(-|\lambda_{ii}|T))$ and

(ii) $E(S_T(i)) = \int_0^T P(X(s) = i)ds$.

$$Q.E.D.$$

By this proposition, we can see that the sojourn time in a slow state is longer than the sojourn time in a fast state. This will be one of the criteria to reduce the dimension for the filter equations (1) and (2).

## 2.3    Consistency of the Joint Process $\xi_t = (X_t, Y_t)$

Now, we turn our attention to the structures of the jump observation process and the joint process $\xi = (X_t, Y_t)$. By the structure of the partially observable system, we know that the jump rates of the observation process Y depend on the state process X. Thus it is reasonable to introduce the following consistency concept for the joint process $\xi = (X, Y)$.

First, let us introduce some notations.

Let Y be the jump observation process and $\Gamma^{(k)} = (\gamma_{ij}^{(k)})$ be the conditional jump rates matrix, i.e.,

$$P(X_{T_l} = k, Y_{T_l} = j | X_{T_l - h} = k, Y_{T_l - h} = i) = \gamma_{ij}^{(k)} h + o(h)$$

for $i \neq j$, and $\gamma_{jj}^{(k)} = 0$, where $i, j = 1, 2, \cdots, n$, $k = 1, \cdots, d$, and $h > 0$.

**Definition 2.4** (i) We say that $\xi_t = (X_t, Y_t)$ is consistent if $\gamma_{ij}^{(k)} > \gamma_{ij}^{(l)}$ for $k, l = 1, \cdots, d$, such that $X(k) \in \Omega_F(X)$ and $X(l) \in \Omega_S(X)$.

**Remarks 2.2** If the joint process $\xi_t = (X_t, Y_t)$ is consistent and $X(k) \in \Omega_F(X)$, then

(i) $|q_{(k,m_j)(k,m_j)}| > |q_{(i,m_l)(i,m_l)}|$, where $m_j, m_l \in \{1, \cdots, n\}$, $i \in \{1, 2, \cdots, m\}$ and $X(i) \in \Omega_S(X)$.

(ii) $\sum_{h=1}^{d} q_{(k,m_j)(h,m_l)} > \sum_{h=1}^{d} q_{(i,m_j)(h,m_l)}$, where $i = 1, \cdots, d$, such that $X_i \in \Omega_S(X)$, and $m_j \neq m_l$.

By reordering the states of the original joint process $\xi$ following the partition into slow states and fast states, and using the above remark, the matrices $A^{(m_j)}, B^{(m_j, m_k)}$ have the following corresponding block forms:

$$A^{(m_j)} = \begin{pmatrix} A_{SS}^{(m_j)} & A_{FS}^{(m_j)} \\ A_{SF}^{(m_j)} & A_{FF}^{(m_j)} \end{pmatrix},$$

and

$$B^{(m_j, m_l)} = \begin{pmatrix} B_{SS}^{(m_j, m_l)} & B_{FS}^{(m_j, m_l)} \\ B_{SF}^{(m_j, m_l)} & B_{FF}^{(m_j m_l)} \end{pmatrix}.$$

The unnormalized density vector $P(t)$ is partitioned as $P(t) = \begin{pmatrix} P_S(t) \\ P_F(t) \end{pmatrix}$. We assume that $P_S(t)$ is an $n_1$-dimensional vector and $P_F(t)$ is an $n_2$-dimensional vector

with $n_1 + n_2 = d$ (actually, the dimension $n_1$ of the vector $P_S(t)$ will be determined by the chosen constant $\alpha$).

# 3. THE APPROXIMATE DIMENSION REDUCTION ALGORITHM AND ERROR ESTIMATES

In this section we shall propose a procedure to reduce the the dimension of the filter equation (1.2) approximately.

## 3.1 The Approximate Dimension Reduction Algorithm

If the joint process $\xi_t = (X_t, Y_t)$ is uniformly consistent, and the matrices $A^{(m_j)}$ and $B^{(m_j, m_l)}$ are partitioned as above, then we have

$$P'_S(t) = A^{(m_j)}_{SS} P_S(t) + A^{(m_j)}_{FS} P_F(t) \tag{3.1}$$

$$P'_F(t) = A^{(m_j)}_{SF} P_S(t) + A^{(m_j)}_{FF} P_F(t) \tag{3.2}$$

for $t \in [T_{j-1}, T_j)$, and

$$P_S(T_j) = B^{(m_j, m_l)}_{SS} P_S(T_j^-) + B^{(m_j, m_l)}_{FS} P_F(T_j^-) \tag{3.3}$$

$$P_F(T_j) = B^{(m_j, m_l)}_{SF} P_S(T_j^-) + B^{(m_j, m_l)}_{FF} P_F(T_j^-) \tag{3.4}$$

for $t = T_j$.

Now, we want to find an approximate dimension reduction algorithm for the problem. The idea is based on the fact that the sojourn time of the fast state is very short, or even negligible with respect to the time scale of the problem and hence can

be deleted by properly modifying the transition between slow states. The following proposition will suggest how to approximate the $P_S(t)$.

**Proposition 3.1**   Let $P_S^*(s)$ be the Laplace transform of $P_S(t)$:

$$P_S^*(s) = [sI_S - A_{SS}^{(m_j)} - A_{FS}^{(m_j)}(sI_F - A_{FF}^{(m_j)})^{-1}A_{SF}^{(m_j)}]^{-1} \times$$
$$[P_S(T_j) + A_{FS}^{(m_j)}(sI_F - A_{FF}^{(m_j)})^{-1}P_F(T_j)],$$

where $I_S$ and $I_F$ are $n_1$ and $n_2$ dimensional identity matrices respectively.

If $\Omega_F(X)$ is a fast recurrent subset, then

$$\bar{P}_S^*(s) = \lim_{\alpha \to \infty} P_S^*(s) = (sI_S - A_{SS}^{(m_j)})^{-1}P_S(T_j)$$

and if $\Omega_F(X)$ is a fast transient subset, then

$$\bar{P}_S^*(s) = \lim_{\alpha \to \infty} P_S^*(s)$$
$$= (sI_S - A_{SS}^{(m_j)} + \overline{A_{FS}^{(m_j)}}(\overline{A_{FF}^{(m_j)}})^{-1}\overline{A_{SF}^{(m_j)}})^{-1} \cdot$$
$$(P_S(T_j) - \overline{A_{FS}^{(m_j)}}(\overline{A_{FF}^{(m_j)}})^{-1}P_F(T_j)),$$

where $\overline{()}$ is defined in the Remark 2.1.

Proof:   Proposition 3.1 follows from Remark 2.2.

$$Q.E.D$$

Proposition 3.1 suggests taking as an approximation $\bar{P}_S(t)$ of $P_S(t)$ the solution

$$\bar{P}'(t) = A_{SS}^{(m_j)}\bar{P}_S(T_j)$$

with $\bar{P}_S(T_j) = P_S(T_j)$ if $\Omega_F(X)$ is a fast recurrent subset, and

$$\bar{P}_S'(t) = (A_{SS}^{(m_j)} - A_{FS}^{(m_j)}(A_{FF}^{(m_j)})^{-1}A_{SF}^{(m_j)})\bar{P}_S(t) \qquad (3.5)$$
$$\bar{P}_F(t) = -(A_{FF}^{(m_j)})^{-1}A_{SF}^{(m_j)}\bar{P}_S(t) \qquad (3.6)$$

with

$$\bar{P}_S(T_j) = (B_{SS}^{(m_j,m_l)} - B_{FS}^{(m_j,m_l)}(A_{FF}^{(m_j)})^{-1}A_{SF}^{(m_j)})\bar{P}_S(T_j^-) \qquad (3.7)$$

$$\bar{P}_F(T_j) = (B_{SF}^{(m_j,m_l)} - B_{FF}^{(m_j,m_l)}(A_{FF}^{(m_j)})^{-1}A_{SF}^{(m_j)})\bar{P}_S(T_j^-) \qquad (3.8)$$

if $\Omega_F(X)$ is a fast transient subset.

It should be noticed that if $\Omega_F(X)$ is a fast recurrent subset then the process is nearly decomposable, i.e., the states can be divided into two nearly invariant sets ( sets each of which consists of communicating states but with the property that passage to other set is of very small probability). In this case, we can find the solutions for the (1, 2) from two separated low dimensional filter equations. Thus, we will limit our attension in the case where $\Omega_F(X)$ is a fast transient subset.

## 3.2    Model Reduction Error

First, let us look at the asymptotic distribution of normalized solution of (1) and (2). If $t \in [T_j, T_{j+1})$, then the solution of the filter equation (1) is

$$P(t) = \begin{pmatrix} P_S(t) \\ P_F(t) \end{pmatrix} = exp(A^{(m_j)}) \begin{pmatrix} P_S(T_j) \\ P_F(T_j) \end{pmatrix}$$

and, the normalized solution is

$$P^*(t) = \begin{pmatrix} P_S^*(t) \\ P_F^*(t) \end{pmatrix} = \frac{1}{U^T P(t)} P(t)$$

where $U^T = [1, \cdots, 1]_{1 \times d}$

Since $A^{(m_j)}$ is a dominant diagonal matrix·with negative diagonal elements, the real parts of all its eigenvalues are negative [4, p26]. Let $\lambda_j, j = 1, \cdots, d$ be its

eigenvalues with

$$|\lambda_1| \leq \cdots \leq |\lambda_d|.$$

The following proposition gives some useful properties of the eigenvalue $\lambda_1$.

**Proposition 3.2** (i) $\lambda_1$ is a real eigenvalue of $A^{(m_j)}$,

(ii) $\lambda_1$ is not an eigenvalue of $A_{FF}^{(m_j)}$,

(iii) $\det(\lambda_1 I_S - A_{SS}^{(m_j)} - A_{FS}^{(m_j)}(\lambda_1 I_F - A_{FF}^{(m_j)})^{-1}A_{SF}^{(m_j)})=0$.

Proof Since $-A^{(m_j)}$ is an M-matrix, then, by Theorem 11.4.7 in [5, p385], $-\lambda_1$ is

a positive real eigenvalue of $-A^{(m_j)}$ and hence (i) is true.

Since the matrix $A^{(m_j)}$ is reordered by the partition into slow states and fast

states, we have

$$a_{11} \leq a_{22} \leq \cdots \leq a_{n_1 n_1} < a_{n_1+1 n_1+1} \leq \cdots \leq a_{nn}.$$

By Theorem 11.4.8 in [4, p386],

$$-\lambda_1 \leq a_{11} < a_{n_1+1 n_1+1}.$$

Thus, $-A_{FF}^{(m_j)} + \lambda_1 I_F$ is an M-matrix and then $-A_{FF}^{(m_j)} + \lambda_1 I_F$ is nonsigular [5,

p383]. So $\lambda_1$ is not an eigenvalue of $A_{FF}^{(m_j)}$.

(iii) follows from (ii) and the following fact,

$$det(A^{(m_j)}-\lambda_1 I) = det(\lambda_1 I_S - A_{SS}^{(m_j)} - A_{FS}^{(m_j)}(\lambda_1 I_F - A_{FF}^{(m_j)})^{-1}A_{SF}^{(m_j)})det(A_{FF}^{(m_j)}-\lambda_1 I_F).$$

$$Q.E.D.$$

We now turn to investigate the asymptotic distribution of

$$exp((-\lambda_1 I + A^{(m_j)})t)P(T_j)$$

by using Laplace transform techniques.

Let $\hat{P}^*(s)$ be the Laplace transform of $exp((-\lambda_1 I + A^{(m_j)})t)P(T_j)$. Then

$$\hat{P}^*(s) = \begin{pmatrix} \hat{P}_S^*(s) \\ \hat{P}_F^*(s) \end{pmatrix} = [sI - (-\lambda_1 I + A^{(m_j)})]^{-1} P(T_j)$$

Thus,

$$\hat{P}_S^* = [sI_1 - (-\lambda_1 I_S + A_{SS}^{(m_j)}) - A_{FS}^{(m_j)}(sI_F - (-\lambda_1 I_F + A_{FF}^{(m_j)}))^{-1} A_{SF}^{(m_j)}]^{-1} \cdot$$

$$[P_S(T_j) + A_{FS}^{(m_j)}(sI_F - (-\lambda_1 I_F + A_{FF}^{(m_j)})P_F(T_j)]$$

Let $A_{11}^{*(m_j)} = sI_S - (-\lambda_1 I_S + A_{SS}^{(m_j)}) - A_{FS}^{(m_j)}(sI_S - (-\lambda_1 I_F + A_{FF}^{(m_j)})^{-1} A_{SF}^{(m_j)}$.

Since

$$det(A_{11}^{*(m_j)}) = C_0' + C_1' s + \cdots + C_{n_1}' s^{n_1}$$

where $C_r', 0 \le r \le n_1$ is equal to the sum of all principal minors of order $n_1 - r$ of $A_{11}^{*(m_j)}$ multiplied by $(-1)^{n_1-r}$. By Lemma 3.1 , $C_0' = 0$. Thus,

$$\lim_{s \to 0} \frac{1}{s} det(A_{11}^{*(m_j)}) = C_1'$$

where $C_1'$ is the sum of all principal minors of order $n_1 - 1$ of

$$A_0^* = \lambda_1 I_S - A_{SS}^{(m_j)} - A_{FS}^{(m_j)}(\lambda_1 I_F - A_{FF}^{(m_j)})^{-1} A_{SF}^{(m_j)}.$$

Therefore, it is easy to see that

$$\lim_{s \to 0} s\hat{P}_S^*(s) = (\frac{1}{C_1'} adj(A_0^*))[P_S(T_j) - A_{FS}^{(m_j)}(\lambda_1 I_F - A_{FF}^{(m_j)})P_F(T_j))].$$

By the finite value theorem of the Laplace transform [9,p21], we have

$$\lim_{t \to \infty} exp(-\lambda_1 t)P_S(t) = \lim_{s \to 0} s\hat{P}_S^*(s)$$

$$= (\frac{1}{C_1'} adj(A_0^*))[P_S(T_j) - A_{FS}^{(m_j)}(\lambda_1 I_F - A_{FF}^{(m_j)})P_F(T_j))].$$

If we use $exp(A^{(m_j)} - \lambda_1 I)$ to modify our reduced dimension filter equation (7), we have

$$\bar{P}'_S(t) = \bar{\Lambda}_C \bar{P}_S(t) \tag{3.9}$$

where $\bar{\Lambda}_C = A_{SS}^{(m_j)} - \lambda_1 I_S - A_{FS}^{(m_j)}(sI_F + \lambda_1 I_F - A_{FF}^{(m_j)})^{-1} A_{SF}^{(m_j)}$.

Using the Laplace transform to (11), we have

$$\bar{P}^*_S(s) = [sI_S - \bar{\Lambda}_C]^{-1} \bar{P}_s(T_j)$$

By the same arguments as above we can obtain,

$$\lim_{t \to \infty} \bar{P}_S(t) = \lim_{s \to \infty} \bar{P}^*_S = (\frac{1}{C_1} adj(A^*_0)) \bar{P}_S(T_j)$$

Summarizing the above dicussions, we have the following theorem for the convergence of the errors.

**Theorem 3.1** If $E(W_j) \to \infty$, where $W_j = T_{j+1} - T_j$ is the wating time of the jump observation process $Y_t$, then

$$E|e_S(t)| \to 0, \quad \text{as } t \to \infty$$

where $t \in [T_j, T_{j+1})$ and $e_S(t) = P^*_S(t) - \bar{P}^*(t)$, provided that both have the same initial values.

Next, we want to find the corresponding error estimate for the above procedure. First, let us introduce some notations.

Let $\bar{P}^*(t) = \frac{1}{P(t)}[(\bar{P}_S(t))^T, (\bar{P}_F(t))^T]^T$ and

$$e(t) = \begin{pmatrix} e_S(t) \\ e_F(t) \end{pmatrix} = P^*(t) - \bar{P}^*(t).$$

Then, it is easy to find the following equation for $e(t)$:

$$e'(t) = \begin{pmatrix} A_{SS}^{(m_j)} & A_{FS}^{(m_j)} \\ A_{SF}^{(m_j)} & A_{FF}^{(m_j)} \end{pmatrix} e(t) + G(t) + H_1(t) + H_2(t), \quad (12)$$

where $G(t) = [0^T, (-A_{FF}^{-1}A_{SF}\Lambda_C \bar{P}_S(t))^T]^T$ and $H_1(t) = \bar{P}^* U^T \bar{A}(m_j)\bar{P}^*$ with

$$\bar{A}^{(m_j)} = \begin{bmatrix} \Lambda_C & 0 \\ (-A_{FF}^{(m_j)})^{-1}A_{SF}^{(m_j)}\Lambda_C & 0 \end{bmatrix},$$

and $H_2(t) = -P^* U^T A^{(m_j)} P^*$ for $t \in [T_j, T_{j+1})$ and

$$e(T_{j+1}) = P^*(T_{j+1}) - \bar{P}^*(T_{j+1}) = \begin{bmatrix} B_{SS}^{(m_j,m_i)} & B_{FS}^{(m_j,m_i)} \\ B_{SF}^{(m_j,m_i)} & B_{FF}^{(m_j,m_i)} \end{bmatrix} e(T_{j+1}^-)$$

at jump times $T_{j+1}$.

It is easy to see that the first three terms in above inhomegeneous ordinary differential equation can be computed. $H_2(t)$ also can be computed because $P^*(t) = (exp(A^{(m_j)}(t - T_j)))P^*(T_j)$ and $P^*(T_j) = e(T_j) + \bar{P}^*(T_j)$. Thus we can compute the error at between jump times and at jump times by the above formulas. Unfortunately, the formula described above is of limited utility, because the numerical computation is very complicated. So it is essential to seek a simple and resonable approximate error estimates formula.

Before looking for such an approximate error estimates formula. we prove the following proposition.

**Proposition 3.3** The system (12 ) is asymptotically stable.

Proof    Since the real parts of all eigenvalues of $A^{(m_j)}$ are negative, the system $e'(t) = A^{(m_j)}e(t)$ is asymptotically stable. Since $\Lambda_C$ is diagonally dominant with

negative diagonal elements [4, p128], real parts of all eigenvalues of $\Lambda_{C'}$ are negative and hence

$$\lim_{t\to\infty}\int_{T_j}^{t} G(s)ds < \infty, \quad \lim_{t\to\infty}\int_{T_j}^{t} H_1(s)ds < \infty, \quad \lim_{t\to\infty}\int_{T_j}^{t} H_2(s)ds < \infty.$$

Hence the system (12) is asympoticlly stable.

$$Q.E.D$$

With Proposition 3.3 in mind, we can take the following approximate error estimates,

$$\bar{e}(t) = A^{(m_j)}\bar{e}(t),$$

for $T_j \le t < T_{j+1}$.

## 4. EXAMPLE

In order to illustrate the efficiency of our approximate dimension reduction procedure let us consider the following simulation example.

Let the continuous time Markov chain state process $X_t$, $t \geq 0$, have states $\{1, 2, \cdots, 6\}$, and the observation process $Y_t$, $t \geq 0$, have states $\{y_1, y_2, y_3\}$. We assume that $X_t$ and $Y_t$ do not have common jumps. Let the joint process $\xi_t = (X_t, Y_t)$ be the continuous time Markov chain with the following generator.

The $B^{(y_j, y_l)} = (b_{i,k}^{(y_j, y_l)})$ are given as follows,

$$B^{(y_1, y_2)} = diag(0.220, 0.240, 0.260, 0.280, 0.800, 0.900),$$

$$B^{(y_1, y_3)} = diag(0.240, 0.280, 0.320, 0.360, 0.840, 0.940),$$

$$B^{(y_2, y_1)} = diag(0.220, 0.290, 0.360, 0.430, 1.000, 1.100),$$

$$B^{(y_2, y_3)} = diag(0.295, 0.340, 0.385, 0.430, 0.880, 0.980),$$

$$B^{(y_3, y_1)} = diag(0.340, 0.410, 0.480, 0.550, 1.100, 1.200),$$

$$B^{(y_3, y_2)} = diag(0.270, 0.290, 0.310, 0.330, 0.880, 0.940).$$

The $A^{(m_j)} = (a_{ik}^{(y_j)})$ are given as follows.

$$A^{(y_1)} = \begin{pmatrix} 1.3 & 0.1 & 0.15 & 0.2 & 0.5 & 0.6 \\ 0.15 & 1.35 & 0.2 & 0.25 & 0.6 & 0.7 \\ 0.2 & 0.2 & 1.55 & 0.3 & 0.7 & 0.8 \\ 0.25 & 0.25 & 0.3 & 1.75 & 0.8 & 0.9 \\ 0.11 & 0.13 & 0.15 & 0.17 & 5.24 & 0.9 \\ 0.13 & 0.15 & 0.17 & 0.19 & 1.0 & 5.74 \end{pmatrix}.$$

and $A^{(y_2)}$ and $A^{(y_3)}$ are given as follows,

$$a_{ik}^{(y_2)} = a_{ik}^{(y_1)} + (-0.05)^i,$$

$$a_{ik}^{(y_3)} = a_{ik}^{(y_1)} + (-0.03)^i,$$

for $i \neq k$ and

$$a_{kk}^{(y_j)} = \sum_{k \neq i} a_{ik}^{(y_j)} + \sum_{i=1}^{d} \sum_{y_l \neq y_j} b_{ik}^{(y_j, y_l)}.$$

We have simulated the $Y_t$ process with the above parameters and initial conditions $P(X_t = i) = \frac{1}{6}, i = 1, 2 \cdots, 6$. The simulated $Y_t$ process starts at state $i = 1$, then jumps to state $i = 2$ at time $T_1 = 5.46$, jumps to state $i = 1$ at time $T_2 = 6.08$, jumps to state $i = 2$ at time $T_3 = 9.35$, jumps to state $i = 3$ at time $T_4 = 9.66$, jumps to state $i = 2$ at time $T_5 = 10.97$ and so on.

The average transition rates of the $X_t$, given $Y_t = y_j$ is $\rho = 1.875$ and $\beta = 1.03$. If we take $\alpha = 0.56$ then we can reduce the filter equation to a 3-dimensional filter equation. If we take $\alpha = 2.79$ the filter equation can be reduced to a 4-dimensional filter equation. The numerical solutions of the original filter equations and the two reduced dimension filter equations at time t=10.0 are given as follows.

The solution of the original filter equation is

$$P(X_{10.0} = 1 | Y_{10.0} = y_3) = 0.1516 \quad P(X_{10.0} = 2 | Y_{10.0} = y_3) = 0.2465$$

$$P(X_{10.0} = 3 | Y_{10.0} = y_3) = 0.2233 \quad P(X_{10.0} = 4 | Y_{10.0} = y_3) = 0.2663$$

$$P(X_{10.0} = 5 | Y_{10.0} = y_3) = 0.0490 \quad P(X_{10.0} = 6 | Y_{10.0} = y_3) = 0.0631.$$

The solution of the reduced dimension filter equation in the case $\alpha = 0.56$:

The solution for slow states is

$$P(X_{10.0} = 1 | Y_{10} = y_3) = 0.2077 \quad P(X_{10.0} = 2 | Y_{10.0} = y_3) = 0.3406$$

$$P(X_{10.0} = 3 | Y_{10.0} = y_3) = 0.2904.$$

The solution for fast states is

$$P(X_{10.0} = 4 | Y_{10.0} = y_3) = 0.1568 \quad P(X_{10.0} = 5 | Y_{10.0} = y_3) = 0.0043$$

$$P(X_{10.0} = 6 | Y_{10.0} = y_3) = 0.0016.$$

The solution of the reduced dimension filter equation in the case $\alpha = 2.79$:

The solution for slow states is

$$P(X_{10.0} = 1 | Y_{10.0} = y_3) = 0.1718 \quad P(X_{10.0} = 2 | Y_{10.0} = y_3) = 0.2781$$

$$P(X_{10.0} = 3 | Y_{10.0} = y_3) = 0.2480 \quad P(X_{10.0} = 4 | Y_{10.0} = y_3) = 0.2896.$$

The solution for fast states is

$$P(X_{10.0} = 5 | Y_{10.0} = y_3) = 0.045 \quad P(X_{10.0} = 6 | Y_{10.0} = y_3) = 0.077.$$

The approximate errors for the first and second cases are

$$e(10) = 0.3211 \quad \text{and} \quad e(10) = 0.0555.$$

respectively.

**Remark 4.1** As we have seen that the number of dimension reduction and errors depend on the constant $\alpha$. Small $\alpha$ results in large error and high dimension reduction and large $\alpha$ results in small error and low dimension reduction.

# REFERENCES

[1] J. Casti, *Linear Dynamical Systems.* Academic Press, Inc., 1986.

[2] U. B. Desai and D. Pal, *A transformation approach to stochastic model reduction.* IEEE Trans. Automat. Contr. vol. AC-29, pp1097-1100, 1984.

[3] K. Fan, *On the unnormalized solution of the filtering problem with jump process observations.* Submitted to IEEE Trans. IT

[4] M. Fiedler , *Special matrices and their applications in numerical mathematics* Martinus Nijhoff Publishers, New York, 1986.

[5] F. A. Grayill, *Matrices with Applications in Statistics.* Wadsworth, Inc., New York, 1983.

[6] G. Del. Grosso, K. Fan, W. Kliemann and F. Marchetti, *Dimension Reduction for Filtered Markov Chains.* Submitted to Opt. and Appl. Mathematics.

[7] W. Kliemann, G. Koch and F. Marchetti *On the unnormlized solution of the filtering problem with counting process observations.* IEEE Trans. IT 36. pp1415-1425, 1990.

[8] B. Moore, *Principal Component Analysis in Linear Syatems; Controllability, Observability, and Model Reduction.* IEEE Trans. Automat. Contr., AC-26, pp17-31. 1981.

[9] W. T. Thomson, *Laplace Transformation.* Prentice-Hall, Inc., New York, 1960.

# GENERAL SUMMARY

In this dissertation, we have studied the implementation of nonlinear filtering algorithems that can be used in real time applications.

In the first part, we developed a general procedure to solve the filtering problem for Markov semimartingale state processes and jump observation processes. By this procedure, we can obtain the solution of the nonlinear filtering equation from solving the corresponding linear filtering equations.

In the second part, we consider the discretization of the state space which leads to filtering equations that are a combination of ordinary differential equations and linear updating operations. For this we investigated the problem of dimension reduction. We provided some necessary and sufficient conditions for the problem to allow for exact dimension reduction.

In the third part, we considered the approximate dimension reduction. We established a procedure to reduce the dimension approximately. The computer simulation showed that this procedure is efficient and applicable.

# LITERATURE CITED

[1] R. S. Bucy, *Nonlinear Filtering* . IEEE Trans. Automatic Control, AC-10, pp95-108, 1965.

[2] P. A. Frost and T. Kailath, *An Innovations Approach to Least-Squares Estimation III*. IEEE Trans. Automatic Control AC-16, pp217-226, 1971.

[3] M. Fujisaki, G. Kallianpur, and H. Kunita, *Stochastic Differential Equations for the Nonlinear Filtering Problem*. Osaka J. Math., 9, pp19-40, 1972.

[4] T. Kailath, *An Innovations Approach to Least-Squares Estimation*. IEEE Trans. Automatic Control, AC-13, pp646-655, 1968.

[5] R. E. Kalman, *A New Approach to Linear Filtering and Prediction Problems*. J. Basic Eng. ASME, 82, pp33-45, 1960.

[6] R. E. Kalman and R. S. Bucy, *New Results in Linear Filtering and Prediction Theory*. J. Basic Engr. ASME D, 83, pp95-108, 1961.

[7] H. J. Kushner, *On the Differential Equations Satisfied by Conditional Probability Densities of Markov Processes*. SIAM J. Control, 2, pp106-119, 1964.

[8] A. N. Shiryaev, *Some New Results in the Theory of Controlled Stochastic Processes*. Trans. 4th Prague Conference on Information Theory. Czech. Academy of Sciences, Prague, 1967.

[9] R. L. Stratonovich, *Conditional Markov Processes and Their Application to the Theory of Optimal Control*. New York: Elsevier, 1968.

[10] W. M. Wonham, , *Some Applications of Stochastic Differential Equations to Optimal Nonlinear Filtering*. SIAM J. Control, 2, pp 347-369, 1965.

[11] M. Zakai, *On the Optimal Filtering of Diffusion Processes*. Z. Wahr. Verw. Geb., 11, pp230-243, 1969.

# ACKNOWLEDGMENTS

Looking back, I realize and appreciate the support and guidance provided to me by several people. These people have all played a part, directly or indirectly, in the production of this dissertation, and I would like to acknowledge some of them here.

Firstly, I would to thank my mother and my older sister for their constant love and support. My thanks also to my high school teachers Mr. Zheng Zhangzhi and Mr. Ke Xianxi for their encouragement and support during my high school and undergraduate period. I also want to thank Dr. Athreya, Dr. Lieberman, Dr. Murdock, Dr. Steiner and Dr. David for serving on my POS committee. I also thank all the people in the Stochastics Group at ISU, especially, Dr. Athreya, Dr. Goggin and Dr. Weersainghe for their help. I also thank Dr. F. Marchetti for his helpful suggestions and useful discussions. I would like to also thank Dr. Colwell and the Mathematics office secretaries, Mandonna, Jan, Ruth and Ellen who have always cheerfully helped me. I should like to express my heartfelt thanks to Dr. Kliemann, my major professor, for his profound concern, patient guidance, and warm friendship. I rejoice to have such a good teacher and helpful friend. I am also especially grateful to my wife, Xuanyuan Wang for her deep love and complete understanding.