5-2012

Stability of Nonlinear Filters and Branching Particle Approximations to The Filtering Problems

Zhiqiang Li
zli21@utk.edu

To the Graduate Council:

I am submitting herewith a dissertation written by Zhiqiang Li entitled "Stability of Nonlinear Filters and Branching Particle Approximations to The Filtering Problems." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Mathematics.

Jie Xiong, Major Professor

We have read this dissertation and recommend its acceptance:

Xia Chen, Jan Rosinski, Mary Sue Younger

Accepted for the Council:

Carolyn R. Hodges

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

Stability of Nonlinear Filters and Branching Particle Approximations to The Filtering Problems

A Thesis Presented for
The Doctor of Philosophy
Degree
The University of Tennessee, Knoxville

Zhiqiang Li

May 2012

I dedicate this dissertation to my wife Wenfang, son Dylan, parents Qunying and Yuzhi. Their support and encouragement have been my motivation.

Acknowledgements

I would like to express my gratitude to the people who have made this dissertation possible.

First, I am most grateful to my advisor: Jie Xiong, who has been an unflagging guide. His support and his excellent mathematical judgment helped me to express myself mathematically. It has been my pleasure to work under his guidance.

I also would like to thank the members of my committee, Dr. Jan Rosinski, Dr. Xia Chen and Dr. Mary Sue Younger. I really appreciate what Dr. Rosinski and Dr. Chen taught me during these years. Dr. Younger was always happy to give me suggestions on statistics.

Submitted in fulfillment of the requirements for the degree of Doctor of Philosophy in Mathematics in the Graduate School of the University of Tennessee at Knoxville, 2012

Abstract

Various particle filters have been proposed and their convergence to the optimal filter are obtained for finite time intervals. However, uniform convergence results have been established only for discrete time filters. We prove the uniform convergence of a branching particle filter for continuous time setup when the optimal filter itself is exponentially stable.

The short interest rate process is modeled by an asymptotically stationary diffusion process. With the counting process observations, a filtering problem is formulated and its exponential stability is derived. Base on the stability result, the uniform convergence of a branching particle filter is proved.

Contents

Chapter 1

Introduction

1.1 Basic knowledge

The goal of the stochastic filtering theory is to estimate a function of an unknown Markov process based on the partial information obtained by observation process. The filtering problem consists of two processes: the signal process X_t, which is what we want to estimate, and the observation process Y_t that provides the information we can use. Let $\mathcal{G}_t$ be the information up to time t which is a σ-field generated by $\{Y_s, 0 \le s \le t\}$. To estimate $f(X_t)$ for bounded function f, we use the conditional expectation $\mathbb{E}(f(X_t)|\mathcal{G}_t)$. The following lemma shows that it has the minimum square error among all the $\mathcal{G}_t$-measurable square-integrable random variables.

Lemma 1.1.1. *Let η be any $\mathcal{G}_t$-measurable square-integrable random variable, then we have*

$$\mathbb{E}\left((f(X_t) - \mathbb{E}(f(X_t)|\mathcal{G}_t))^2\right) \le \mathbb{E}\left((f(X_t) - \eta)^2\right). \tag{1.1}$$

Proof.

$$\begin{aligned}
& \mathbb{E}\left((f(X_t)-\eta)^2\right) - \mathbb{E}\left((f(X_t)-\mathbb{E}(f(X_t)|\mathcal{G}_t))^2\right) \\
= & \ \mathbb{E}\left((\mathbb{E}(f(X_t)|\mathcal{G}_t)-\eta)(2f(X_t)-\eta-\mathbb{E}(f(X_t)|\mathcal{G}_t))\right) \\
= & \ \mathbb{E}\left(\mathbb{E}((\mathbb{E}(f(X_t)|\mathcal{G}_t)-\eta)(2f(X_t)-\eta-\mathbb{E}(f(X_t)|\mathcal{G}_t))|\mathcal{G}_t)\right) \\
= & \ \mathbb{E}\left((\mathbb{E}(f(X_t)|\mathcal{G}_t)-\eta)\mathbb{E}(2f(X_t)-\eta-\mathbb{E}(f(X_t)|\mathcal{G}_t)|\mathcal{G}_t)\right) \\
= & \ \mathbb{E}\left((\mathbb{E}(f(X_t)|\mathcal{G}_t)^2\right) \geq 0.
\end{aligned}$$

□

Let $\pi_t(\cdot) \equiv \mathbb{P}(X_t \in \cdot|\mathcal{G}_t)$ be the regular conditional probability distribution of X_t given $\mathcal{G}_t$; i.e. π_t is a map from $\mathcal{B}(\mathbb{R}^d) \times \Omega$ to $[0,1]$ such that

i) For any $\omega \in \Omega$, $\pi_t(\cdot,\omega)$ is a probability measure on $\mathbb{R}^d$.

ii) For any $A \in \mathcal{B}(\mathbb{R}^d)$, $\pi_t(A,\cdot)$ is a $\mathcal{G}_t$-measurable random variable.

iii) For any $A \in \mathcal{B}(\mathbb{R}^d)$, we have

$$\pi_t(A,\omega) = \mathbb{P}(X_t \in A|\mathcal{G}_t)(\omega), \quad a.s.\ \omega.$$

We use $\langle \mu, f\rangle$ to denote the integral of a function with respect to the measure μ. Then it can be shown that the conditional expectation $\mathbb{E}(f(X_t)|\mathcal{G}_t)$ is given by the integral of f with respect to the regular conditional probability distribution π_t.

Lemma 1.1.2. *For any $f \in C_b(\mathbb{R}^d)$ and $t \geq 0$, we have*

$$\mathbb{E}(f(X_t)|\mathcal{G}_t) = \langle \pi_t, f\rangle \quad a.s.$$

Based on the above two lemmas, we call π_t the optimal filter. Let $\hat{\mathbb{P}}$ be the measure on Ω that is absolutely continuous with respect to $\mathbb{P}$ and the Radon-Nickodym derivative on $(\Omega, \mathcal{F}_t)$ is M_t^{-1}, that is

$$\frac{d\hat{\mathbb{P}}}{d\mathbb{P}} = M_t^{-1}.$$

The following theorem plays a very important role in the filtering theory.

Theorem 1.1.1. (Kallianpur-Striebel formula) *The optimal filter π_t can be represented as*

$$\langle \pi_t, f \rangle = \frac{\langle V_t, f \rangle}{\langle V_t, 1 \rangle}, \quad \forall f \in C_b(\mathbb{R}^d), \tag{1.2}$$

where

$$\langle V_t, f \rangle = \hat{\mathbb{E}}(M_t f(X_t) | \mathcal{G}_t), \tag{1.3}$$

and $\hat{\mathbb{E}}$ refers to the expectation with respect to the measure $\hat{\mathbb{P}}$.

1.2 Nonlinear filtering model with Brownian motion

Let the signal process X_t be a $\mathbb{R}^d$-valued process governed by the following stochastic differential equation (SDE):

$$dX_t = b(X_t)dt + c(X_t)dW_t + \sigma(X_t)dB_t,$$

where B and W are independent Brownian motions taking values in $\mathbb{R}^d$ and $\mathbb{R}^m$, respectively, and $b : \mathbb{R}^d \to \mathbb{R}^d, c : \mathbb{R}^d \to \mathbb{R}^{d\times m}$ and $\sigma : \mathbb{R}^d \to \mathbb{R}^{d\times d}$ are continuous mappings. The observation process is given by:

$$Y_t = \int_0^t h(X_s)ds + W_t,$$

where $h \in C_b(\mathbb{R}^d)$.

The stability of this nonlinear filter model is an important concept in the filtering theory. We investigate the following question: Under what conditions does the distance between π_t and $\bar{\pi}_t$ tends to 0 as $t \to \infty$? Here π_t and $\bar{\pi}_t$ are two optimal filters with initial distribution π_0 and $\bar{\pi}_0$, respectively.

Definition 1.2.1. *The filtering model is asymptotically stable if for any* $\pi_0, \bar{\pi}_0 \in \mathcal{P}(\mathbb{R}^d)$*, we have*

$$\lim_{t\to\infty} d(\pi_t, \bar{\pi}_t) = 0,$$

where $d(\cdot,\cdot)$ *is a suitable metric in the space of probability measure on* $\mathbb{R}^d$.

The investigation of this problem has a long history, starting with the pioneering work of Kunita [20] (in continuous time setting) on the stationary behavior of the mean square estimation error of the nonlinear filter. But there is a serious gap in the proof of the main result in [20] (see [29]). After that, the stability of the optimal filter has received considerable attention in many years, e.g. . Atar and Zeitouni [2] showed the exponential stability of the nonlinear filter in the compact space. Budhiraja and Ocone [3] proved the exponential stability of discrete time filters for bounded observation noise. Atar [1] considered one dimensional nonlinear filtering with linear observations in a noncompact domain. Recently, Van Handel [14] partially solved Kunita's problem by checking Von Weizsácker's conditions for exchange of intersection and supremum of $\sigma-$fields.

Another problem in the filtering theory is numerical methods for solving optimal filter. Even though, the filtering problem has been studied in the literature extensively, there are only a few cases which have explicit solutions. Therefore, to solve the filtering problems, we have to resort to numerical approximations.

An efficient way is to use random particle systems to approximate the filtering problem numerically. Such approximation of the optimal filter was studied in heuristic schemes in the beginning of the 1990s by Gordon *et al* [28], Kitagawa [18], Carvalho *et al* [13]. The rigorous proof of the convergence results for the particle filter were published in 1996 by Del Moral [23], and indepently, by Crisan and Lyons [8] in 1997. Since then, many improvements have been made, e.g. Crisan *et al* [5] [11] [6] [10], Del

Moral and Miclo [26]. There is one thing is common in these methods: the number of particles doesn't change along the time line. So it always require a lot on computation ability.

A different type of particle system is introduced by Crisan and Xiong in [9] and a central-limit-type theorem was proved. In [9], Crisan and Xiong studied a branching particle system to solve the continuous time filtering problem. In their construction, particles move according to the law of the signal, conditionally (given the observation) independent of each other, in a small time interval whose length tends to 0 while the initial number of particles tends to ∞. At the end of each time interval, the particles die and give birth to random numbers of offsprings. The offsprings move from the positions of their mothers with weight 1. The expected numbers of offsprings are the weight of the corresponding particles decided according to the paths during the period prior to that time step. In this setting, the number of offsprings decreases to a small number with a large probability. Therefore, the approximation is easy to calculate after the long time period.

Uniform convergence of particle filter with fixed number of particles was first studied by Del Moral and Guionnet in [25]. It was also studied by some other authors (e.g. Crisan and Heine [7], Del Moral [24]) for discrete time filters. In [9], since the numbers of offsprings are random numbers at each time step, the proof of uniform convergence is not trivial. In fact, as mentioned above, the number of particles is more likely small number after a long time. From the reference above, we see that at the discrete-time setting, there is a close connection between the stability of the filter and the uniform convergence of its (particle) approximation. We will study the uniform convergence of the particle filter defined by the particle system under the some stability condition. In [22], a scheme without integration in weights of the particles is defined. The uniform convergence is also proved for this case.

1.3 The filtering model with Poisson observations

Zeng [35] proposes a general Filtering Micromovement model for asset price (FM model, as we simply call it), where the sample characteristics of micro- and macro-movements are tied in a consistent manner. Economically, the proposed model has the structure similar to a class of time series structural models developed in many early market microstructure papers (see [15], a survey paper on this topic, and [16]). Namely, price can be decomposed as a permanent component and a transient component. The permanent component has a long-term impact on price while the transient component has only a short-term impact. In FM model, there is an unobservable intrinsic value process for an asset, which corresponds to the usual price process in the option pricing literature and in the empirical econometric literature of macro-movement. The intrinsic value process is the permanent component and has a long-term impact on price. Prices are observed only at random trading times which are modeled by a conditional Poisson process. Moreover, prices are distorted observations of the intrinsic value process at the trading times and trading (or market microstructure) noise is explicitly modeled. It is the transient component and only has short-term impact (when a trade happens) on price.

The most prominent feature of FM model is that trade-by-trade prices are viewed as a collection of counting processes of price level and the model is framed as a filtering problem with counting process observations. Then, the unnormalized and normalized filtering equations, which correspond to Duncan-Mortensen-Zakai, and Kushner-Stratonovich or Fujisaki-Kallianpur-Kunita equations in classical nonlinear filtering, are derived by [35]. These equations characterize the evaluation of the integrated likelihoods and the conditional distribution of the intrinsic value process (the signal). The Markov chain approximation method is applied to numerically solve the filtering equations. Then Bayes estimation via filtering for the intrinsic value process and the related parameters in the model is developed in [35]. Bayesian hypothesis testing or model selection via filtering for this class of models is developed in [19]. Furthermore, a risk minimization hedging strategy for a FM model is considered in [21], and a mean-variance portfolio selection for a FM model is studied in [34].

The short interest rate process could be modeled by an asymptotically stationary process, e.g. the Vasicek model and the Cox-Ingersoll-Ross model. For short interest rate process, the stationary assumption is natural and there are ultra-high frequency data for short interest rate, too.

The underlining intrinsic process $X(t)$ is modeled by an asymptotically stationary diffusion process. The observation process $Y(t)$ is a counting process and it describes the numbers of trades at each price level. The filtering problem is established the same as in [35]. We will study the stability of the filtering with counting process observations and numerical method for approximation. The following is the mathematical model:

$X(t)$, the intrinsic interest rate process follows a diffusion process:

$$dX(t) = \mu(X(t)) + \sigma(X(t))dB(t),$$

where $B(t)$ is a standard Brownian motion. The generator associated with X is

$$Lf(x) = \frac{1}{2}\sigma^2(x)\frac{\partial^2 f}{\partial x^2}(x) + \mu(x)\frac{\partial f}{\partial x}(x).$$

The intrinsic rate process can only be partially observed through the price process, Y. Due to price discreteness, Y is in a discrete state space given by the multiples of tick, the minimum price variation set by trading regulation. Y is a distorted observation of X at some random times. We view the transaction prices in the levels of price due to price discreteness. That is, we view the prices as a collection of counting processes in the following form:

$$\vec{Y}(t) = \begin{pmatrix} N_1(\int_0^t \lambda_1(X(s))ds) \\ N_2(\int_0^t \lambda_2(X(s))ds) \\ . \\ . \\ . \\ N_w(\int_0^t \lambda_w(X(s))ds) \end{pmatrix}, \tag{1.4}$$

where $Y_k(t) = N_k(\int_0^t \lambda_k(X(s), s)ds)$, $k = 1, 2, ..., w$, is the counting process recording the cumulative number of trades that have corrupted at the kth price level (denoted by y_k) up to time t. The stability of the filter is studied and the exponential stability is derived. A branching particle filter is introduced by J. Xiong and Y. Zeng in [34] and the convergence on the finite time interval is proved. We prove the uniform convergence on the whole time line for a diffusion intrinsic process with linear growth.

1.4 Notational conventions

Many of our notational conventions are outlined in the following table.

Notation	Meaning
$(\Omega, \mathcal{F}, \mathbb{P}, \mathcal{F}_t)$	stochastic basis
$\mathbb{R}^d$	d-dimensional vector space over $\mathbb{R}$
$\tanh(x)$	hyperbolic tangent function of x
$\mathcal{S}_+^d$	the space of all nonnegative-definite symmetric $d \times d$matrices
$\mathbb{R}_+$	all positive real numbers
$C(\mathbb{R}_+, \mathbb{R}^d)$	the space of all continuous mappings from $\mathbb{R}_+$ to $\mathbb{R}^d$
$\mathcal{B}(\mathbb{R}^d)$	the collection of all the Borel sets in $\mathbb{R}^d$
$\mathcal{M}_F(\mathbb{R}^d)$	the space of finite measures on $\mathbb{R}^d$
$\mathcal{P}(\mathbb{R}^d)$	the space of probability measures on $\mathbb{R}^d$
$d_{TV}(\cdot, \cdot)$	total variation metric on $\mathcal{P}(\mathbb{R}^d)$
$C_b^k(\mathbb{R}^d)$	the collection of all bounded continuous mappings on $\mathbb{R}^d$ with bounded partial derivatives up to order k
$W_p^w(\mathbb{R}^d)$	the collection of all functions on $\mathbb{R}^d$ with generalized partial derivatives up to order k with both the functions and all its partial derivatives being p-integrable
$\log x$	natural logarithm of x

Chapter 2

Nonlinear filtering model with Brownian motion

2.1 Stability of the filtering

As mentioned in chapter 1, the exponential stability has been studied a lot. In this section, some assumptions are stated and they are used in the proof the uniform convergence of the numerical method in the next section. Furthermore, two examples are presented for the assumptions check.

2.1.1 Assumptions

First, let's recall our model. The signal process X_t follows

$$dX_t = b(X_t)dt + c(X_t)dW_t + \sigma(X_t)dB_t,$$

and observation process Y_t is given by

$$Y_t = \int_0^t h(X_s)ds + W_t.$$

Therefore,

$$dW_t = dY_t - h(X_t)dt.$$

Rewrite the SDE for signal process X_t, we have

$$dX_t = (b(X_t) - c(X_t)h)X_t))\, dt + c(X_t)dY_t + \sigma(X_t)dB_t.$$

The following assumptions are made on the parameters in the above model.

Assumption 2.1.1. *The mappings σ, b, c, h are in $C_b^k(\mathbb{R}, \mathcal{X})$ with $k = [\frac{d}{2}] + 2$ and $\mathcal{X}$ being $\mathbb{R}^{d\times d}, \mathbb{R}^d, \mathbb{R}^{d\times m}$ and $\mathbb{R}^m$.*

let K be the uniform bound of functions mentioned above. Since h is bounded,

$$\begin{aligned} M_t^{-1} &\equiv \exp\left(-\int_0^t \langle h(X_t), dW_s\rangle - \frac{1}{2}\int_0^t |h(X_s)|^2 ds\right) \\ &= \exp\left(\int_0^t h^*(X_s)dY_s - \frac{1}{2}\int_0^t |h(X_s)|^2 ds\right) \end{aligned}$$

is a martingale. Let $\hat{\mathbb{P}}$ be the measure on Ω that is absolutely continuous with respect to $\mathbb{P}$ and the Radon-Nickodym derivative on $(\Omega, \mathcal{F}_t)$ is

$$\frac{d\hat{\mathbb{P}}}{d\mathbb{P}}\Big|_{\mathcal{F}_t} = M_t^{-1}.$$

Then,

$$\langle V_t, f\rangle = \hat{\mathbb{E}}(M_t f(X_t)|\mathcal{G}_t),$$

and

$$\langle \pi_t, f\rangle = \frac{\langle V_t, f\rangle}{\langle V_t, 1\rangle}$$

where $f \in C_b(\mathbb{R}^d)$.

The followings two theorems are two main equations in filtering theory. The first one is a linear equation of unnormalized filter V_t.

Theorem 2.1.1. (Zakai's equation) *The unnormalized filter* V_t *satisfies the following stochastic differential equation:*

$$\langle V_t, f\rangle = \langle V_0, f\rangle + \int_0^t \langle V_s, Lf\rangle\, ds + \int_0^t \langle V_s, \nabla^* f + fh^*\rangle\, dY_s, \tag{2.1}$$

where

$$Lf = \frac{1}{2}\sum_{i,j=1}^{d} a_{ij}\partial_{ij}^2 f + \sum_{i}^{d} b_i \partial_i f$$

is the generator of the signal process, and the $d \times d$ *matrix* $a = (a_{ij})$ *is given by* $a = cc^* + \sigma\sigma^*$.

We define the innovation process ν_t by

$$d\nu_t = dY_t - \langle \pi_t, h\rangle\, dt. \tag{2.2}$$

Then, for $t > s$, we have

$$\begin{aligned} \mathbb{E}(\nu_t|\mathcal{G}_s) &= \mathbb{E}\left(Y_t - \int_0^t \langle \pi_r, h\rangle\, dr|\mathcal{G}_s\right) \\ &= \mathbb{E}\left(Y_t - Y_s - \int_s^t \langle \pi_r, h\rangle\, dr|\mathcal{G}_s\right) + \nu_s \\ &= \mathbb{E}\left(W_t - W_S - \int_0^t (h(X_r) - \langle \pi_r, h\rangle)\, dr|\mathcal{G}_s\right) + \nu_s \\ &= \int_s^t \mathbb{E}\left(h(X_r) - \mathbb{E}(h(X_r)|\mathcal{G}_s)|\mathcal{G}_s\right) + \nu_s \\ &= \nu_s. \end{aligned} \tag{2.3}$$

As Y_t is Brownian motion under $\hat{\mathbb{P}}$, its quadratic variation is given by

$$\left\langle Y^i, Y^j\right\rangle_t = \delta_{ij} t, \quad i, j = 1, ..., d,$$

where

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases}$$

Since the quadratic variation of the second term in (2.2) is 0, we get

$$\left\langle \nu^i, \nu^j \right\rangle_t = \left\langle Y^i, Y^j \right\rangle_t = \delta_{ij} t, \quad i, j = 1, ..., d.$$

Therefore ν_t is a $\mathcal{G}_t$-Brownian motion under the probability measure $\mathbb{P}$.

Theorem 2.1.2. (Kushner-FKK equation) *The optimal filter π_t satisfies the following stochastic differential equation: for all $f \in C_b^2(\mathbb{R}^d)$,*

$$\langle \pi_t, f \rangle = \langle \pi_0, f \rangle + \int_0^t \langle \pi_s, Lf \rangle \, ds + \int_0^t (\langle \pi_s, \nabla^* f + f h^* \rangle) - \langle \pi_s, f \rangle \langle \pi_s, h^* \rangle \, d\nu_s. \quad (2.4)$$

Let $\mathcal{M}_F(\mathbb{R}^d)$ be the collection of all finite Borel measures on $\mathbb{R}^d$. Denote by $d_{TV}(\cdot, \cdot)$ the total variation distance on $\mathcal{M}_F(\mathbb{R}^d)$. For any $\mu, \nu \in \mathcal{M}_F(\mathbb{R}^d)$, $d_{TV}(\mu, \nu)$ is defined as following:

$$d_{TV}(\mu, \nu) = \sup_{\|f\| \leq 1} |\langle \mu, f \rangle - \langle \nu, f \rangle| \quad (2.5)$$

The exponential stability assumption for the optimal filter is made as following:

Assumption 2.1.2. *The filter is stable in the following sense: There exist constants $C > 0$ and $\beta \in (0, 1)$ such that $\forall \epsilon > 0$, whenever $\mathbb{E} d_{TV}(\pi_0, \bar{\pi}_0) < \epsilon$, there exists $T > 0$ such that when $t > T$, we have*

$$\mathbb{E} d_{TV}(\pi_t, \bar{\pi}_t) \leq C g(t) \mathbb{E} d_{TV}(\pi_0, \bar{\pi}_0),$$

where π_t and $\bar{\pi}_t$ are the optimal filters with initial π_0 and $\bar{\pi}_0$, respectively and $g(\cdot)$ is a function on $\mathbb{R}$ which satisfies $\sum_{n=0}^{\infty} g(n) < \infty$.

2.1.2 Examples

In this section, we state two examples that satisfy the assumption 2.1.2. First example is the Kalman-Bucy filtering. The filtering model whose signal is given by

$$dX_t = bX_t dt + c dW_t + \sigma dB_t \tag{2.6}$$

and the observation process is

$$dY_t = hX_t dt + dW_t, \tag{2.7}$$

where X_0 is a $d-$dimensional normal random vector with mean $\hat{X}_0 \in \mathbb{R}^d$ and covariance matrix $\gamma_0 \in \mathbb{R}^d_+$, the space of the all non-negative-definite symmetric $d \times d-$matrices, (W_t, B_t) is an $m + d$-dimensional Brownian motion, the coefficients b, c, σ, h are matrices of dimensions $d \times d, d \times m, d \times d$ and $m \times d$, respectively.

Theorem 2.1.3. *For any $t \geq 0$ and $\omega \in \Omega$ being fixed, $\pi_t(\omega)$ is a multivariate normal probability measure on $\mathbb{R}^d$.*

Proof. Let $D_N = \{0 = s_1^N < \cdots < s_k^N = t\}$ be an increasing sequence of sets whose union is dense in $[0, t]$. Since (X_t, Y_t) is a Gaussian process, the conditional distribution $\pi_t^N \equiv \mathbb{P}(X_t \in \cdot | Y_s, s \in D_N)$ is normal with conditional mean $\hat{X}_t^N$ and conditional covariance matrix γ_t^N. We now consider the characteristic function corresponding to π_t^N. For $\lambda \in \mathbb{R}^d$, we define

$$\begin{aligned} \phi_N(\lambda) &= \int_{\mathbb{R}^d} e^{i\lambda^* x} \pi_t^N(dx) \\ &= \mathbb{E}\left(e^{i\lambda^* X_t} | Y_s, s \in D_N\right). \end{aligned}$$

Note that for $\lambda \in \mathbb{R}^d$ fixed, $\{\phi_N(\lambda) : N \geq 1\}$ is a martingale. By martingale convergence theorem (see Theorem 27.3 in [17]), we have

$$\lim_{N \to \infty} \phi_N(\lambda) = \phi_\infty(\lambda) \quad a.s.$$

Since the characteristic function of a multivariate normal distribution π_t^N is given by

$$\phi_N(\lambda) = \exp\left(\lambda^* \hat{X}_t^N - \frac{1}{2}\lambda^* \gamma_t^N \lambda\right),$$

we have the convergence of $\hat{X}_t^N$ and γ_t^N as $N \to \infty$. Denote the limits by $\hat{X}_t$ and γ_t, respectively. Then

$$\phi_\infty(\lambda) = \exp\left(\lambda^* \hat{X}_t - \frac{1}{2}\lambda^* \gamma_t \lambda\right).$$

Thus, $\pi_t(\omega)$ is a multivariate normal distribution on $\mathbb{R}^d$. □

Let $\hat{X}_t = \mathbb{E}(X_t|\mathcal{G}_t)$ and $\gamma_t = \mathbb{E}((X_t - \hat{X}_t)(X_t - \hat{X}_t)^*)$. Then $\hat{X}_t$ and γ_t satisfy the following equations:

$$d\hat{X}_t = (b - ch - \gamma_t h^* h)\hat{X}_t dt + (c + \gamma_t h^*)dY_t, \tag{2.8}$$

and

$$\dot{\gamma}_t = \gamma_t(b - ch) + (b - ch)\gamma_t + \sigma^*\sigma - \gamma_t h^* h \gamma_t. \tag{2.9}$$

For any $z \in \mathbb{R}^d$ and $R \in \mathbb{R}_+^d$, we define the $d-$dimensional stochastic process Z_t and $\mathbb{R}_+^d$-valued function P_t by

$$\begin{cases} dZ_t = (b - ch - P_t h^* h)Z_t dt + (c + P_t h^*)dY_t \\ Z_0 = z, \end{cases} \tag{2.10}$$

and

$$\begin{cases} \dot{P}_t = P_t(b - ch) + (b - ch)P_t + \sigma^*\sigma - P_t h^* h P_t \\ P_0 = R. \end{cases} \tag{2.11}$$

Note that for $z = \hat{X}_0$ and $R = \gamma_0$, we have $Z_t = \hat{X}_t$ and $P_t = \gamma_t$. Thus (Z_t, P_t) can be regarded as the linear filter with "incorrect" initial.

First, we define the asymptotically stable matrix.

Definition 2.1.1. *Let A be a $d \times d-$ matrix, A is asymptotically stable if all its eigenvalues have negative real parts.*

We make the following assumption on the coefficients of the system.

Assumption 2.1.3. *There exists a matrix $\gamma_\infty \in \mathbb{R}^d_+$ such that*

$$\gamma_\infty(b - ch)^* + (b - ch)\gamma_\infty + \sigma^*\sigma - \gamma_\infty h^* h \gamma_\infty = 0,$$

and $b - ch - \gamma_\infty h^* h$ is asymptotically stable.

Let

$$0 < \lambda_0 < \inf\{-Re\lambda : \lambda \text{ is an eigenvalue of } b - ch - \gamma_\infty h^* h\}.$$

Note that

$$\begin{aligned} \frac{d}{dt}(P_t - \gamma_\infty) &= \left(b - ch - \frac{1}{2}(P_t - \gamma_\infty)hh^*\right)(P_t - \gamma_\infty) \\ &\quad + \left(\left(b - ch - \frac{1}{2}(P_t - \gamma_\infty)hh^*\right)\right)^*. \end{aligned}$$

Thus, there exists a constant K_1 such that

$$|P_t - \gamma_\infty| \leq K_1 |R - \gamma_0| e^{-\lambda_0 t}. \tag{2.12}$$

Similarly, we have

$$|P_t - \gamma_t| \leq K_2 |R - \gamma_0| e^{-\lambda_0 t}, \tag{2.13}$$

and

$$|\gamma_\infty - \gamma_t| \leq K_3|R - \gamma_0|e^{-\lambda_0 t}. \tag{2.14}$$

By (2.8) and (2.10), we get

$$\begin{aligned} d(\hat{X}_t - Z_t) &= \Big\{(b - ch - \gamma_\infty h^*h)\left(\hat{X} - Z_t\right) + (\gamma_\infty - \gamma_t)h^*h\hat{X}_t \\ &\quad +(\gamma_\infty - P_t)h^*hZ_t\Big\}dt + (\gamma_t - P_t)h^*dY_t. \end{aligned}$$

Applying It$\hat{o}$'s formula, we have

$$\begin{aligned} & d\left((\hat{X}_t - Z_t)e^{-(b-ch-\gamma_\infty h^*h)t}\right) \\ = \;& e^{-(b-ch-\gamma_\infty h^*h)t}\Big\{(\gamma_\infty - \gamma_t)h^*h\hat{X}_t + (\gamma_\infty - P_t)h^*hZ_t\Big\} \\ = \;& e^{-(b-ch-\gamma_\infty h^*h)t}(\gamma_t - P_t)h^*dY_t. \end{aligned}$$

Then,

$$\begin{aligned} \mathbb{E}|\hat{X}_t - Z_t|^2 &\leq 3|\hat{X}_0 - z|^2|e^{2(b-ch-\gamma_\infty h^*h)t}| \\ &\quad +6t\mathbb{E}\int_0^t |e^{2(b-ch-\gamma_\infty h^*h)(t-s)}||\gamma_\infty - \gamma_t||h^*h|^2|\hat{X}_s|^2ds \\ &\quad +6t\mathbb{E}\int_0^t |e^{2(b-ch-\gamma_\infty h^*h)(t-s)}||\gamma_\infty - P_s||h^*h|^2|Z_s|^2ds \\ &\quad +3t\mathbb{E}\left|\int_0^t e^{(b-ch-\gamma_\infty h^*h)(t-s)}(\gamma_t - P_t)h^*dY_s\right|^2 \end{aligned} \tag{2.15}$$

By (2.7)and (2.13), we have

$$\begin{aligned}
& \mathbb{E}\left|\int_0^t e^{(b-ch-\gamma_\infty h^*h)(t-s)}(\gamma_t - P_t)h^* dY_s\right|^2 \\
\leq\ & 2\mathbb{E}\int_0^t e^{2(b-ch-\gamma_\infty h^*h)(t-s)}|(\gamma_t - P_t)|^2|h^*|^2 ds \\
\leq\ & 2t\mathbb{E}\int_0^t e^{2(b-ch-\gamma_\infty h^*h)(t-s)}|(\gamma_t - P_t)|^2|h^*h|^2|\hat{X}_s|^2 ds \\
\leq\ & K_4|R-\gamma_0|^2\int_0^t e^{-2\lambda_0(t-s)}e^{-2\lambda_0 s}ds \\
=\ & K_4|R-\gamma_0|^2 te^{-2\lambda_0 t}
\end{aligned} \tag{2.16}$$

Combining (2.14), (2.12) (2.16) and (2.15), we get

$$\mathbb{E}|\hat{X}_t - Z_t|^2 \leq K_5\left(|\hat{X}_0 - z|^2 + |R-\gamma_0|^2\right)e^{-2\lambda_0 t} \tag{2.17}$$

Proposition 2.1. *Under assumption 2.1.3 , the optimal filter for the model (2.6-2.7) is exponential stable in the following sense:*

$$\mathbb{E}d_{TV}(\pi_t, \bar{\pi}_t) \leq K d_{TV}(\pi_0, \bar{\pi}_0)e^{-\frac{1}{2}\lambda_0 t}$$

Proof. Let $\phi(u)$ be the probability density function of the d-dimensional standard normal random vector. Then, for $f \in C_b(\mathbb{R}^d)$ with $\|f\|_\infty \leq 1$, we have

$$\begin{aligned}
& \left|\int_{\mathbb{R}^d} f(x)\pi_t(dx) - \int_{\mathbb{R}^d} f(x)\bar{\pi}_t(dx)\right| \\
=\ & \left|\int_{\mathbb{R}^d} f(\hat{X}_t + \sqrt{\gamma_t}u)\phi(u)du - \int_{\mathbb{R}^d} f(Z_t + \sqrt{P_t}u)\phi(u)du\right| \\
\leq\ & |\hat{X}_t - Z_t| + |\sqrt{\gamma_t} - \sqrt{P_t}|\int_{\mathbb{R}^d}|u|\phi(u)du \\
\leq\ & |\hat{X}_t - Z_t| + \sqrt{d}|\sqrt{\gamma_t} - \sqrt{P_t}|
\end{aligned}$$

By the definition of $d(\cdot,\cdot)$ discussion above, we have

$$\begin{aligned} & d_{TV}(\pi_t, \bar{\pi}_t) \\ \leq \;& |\hat{X}_t - Z_t| + \sqrt{d}|\sqrt{\gamma_t} - \sqrt{P_t}| \\ \leq \;& K_1 e^{-1/2\lambda_0 t}(|\hat{X}_0 - Z_0| + \sqrt{d}|\sqrt{\gamma_0} - \sqrt{P_0}|). \end{aligned}$$

Note that the convergence in distribution is equivalent to the convergences of mean and variance for normal random variable. Therefore, for $\forall\epsilon > 0$, when $d_{TV}(\pi_0, \bar{\pi}_0) < \epsilon$, we have

$$|\hat{X}_0 - Z_0| + \sqrt{d}|\sqrt{\gamma_0} - \sqrt{P_0}| \leq K_2 d_{TV}(\pi_0, \bar{\pi}_0).$$

Hence

$$\mathbb{E} d_{TV}(\pi_t, \bar{\pi}_t) \leq K_3 d_{TV}(\pi_0, \bar{\pi}_0) e^{-1/2\lambda_0 t}.$$

□

Next we consider a nonlinear filtering in one dimension. The model for the state and observation processes is as follows:

$$dX_t = f(X_t)dt + dB_t, \quad X_t \in \mathbb{R}, \tag{2.18}$$

$$dY_t = X_t dt + N_0^{1/2} dW_t, \quad Y_t \in \mathbb{R}, \quad Y_0 = 0. \tag{2.19}$$

Here $(B_t)_{t\geq 0}$ and $(W_t)_{t\geq 0}$ are independent standard Browian motions.

Assumption 2.1.4. *The function $V(x) = f'(x) + f^2(x)$ is twice continuously differentiable with a bounded second derivative.*

Assumption 2.1.5. *There exists an initial density $\tilde{p}$, such that under $P^{\tilde{p}}$, X_t is stationary and ergodic.*

Assumption 2.1.6. *For some $t \geq 0$, the marginal law of X_t under P is absolutely continuous with respect to that under $P^{\tilde{p}}$.*

Assumption 2.1.7. *One has that $\mathbb{E}^{P^{\tilde{p}}} X_t^2 < \infty$.*

Examine the proof of the Theorem 1 in [1], it's easy to get the following:

Proposition 2.2. *Under assumptions 2.1.4, 2.1.5 and 2.1.6, there exist nonrandom constants C, C_1 and C_2, independent of N_0 and of the initial distributions π_0 and $\bar{\pi}_0$, such that $P-a.s.$,*

$$\mathbb{E}d_{TV}(\pi_t, \bar{\pi}_t) \leq C\mathbb{E}(d_{TV}(\pi_0, \bar{\pi}_0)) \exp\left\{(C_1 \log N_0 + C_2)\, t\right\}. \tag{2.20}$$

As a consequence, the optimal filter is exponential stable when N_0 is small enough.

Proof. Let $\rho_h(\cdot,\cdot)$ be Hilbert metric on $\mathcal{M}_F(\mathbb{R}^d)$ (see definition on page 6 of [12]). By Corollary 1 in [1] and Lemma 3.4 in [12], for $n-1 \leq t \leq n$, we have

$$d_{TV}(\pi_t, \bar{\pi}_t) \leq K d_{TV}(\pi_0, \bar{\pi}_0) \prod_{k=1}^{n} \frac{\rho_h(\pi_k, \bar{\pi}_k)}{\rho_h(\pi_{k-1}, \bar{\pi}_{k-1})}$$

Let $\mathcal{T}$ be an linear operator on $\mathcal{M}_F(\mathbb{R}^d)$, that possesses a kernel $T(\cdot,\cdot)$. By Lemma 10.31 in [33], we have

$$\sup\left\{\frac{\rho_h(\mathcal{T}\mu, \mathcal{T}\nu)}{\rho_h(\mu,\nu)} : 0 < \rho_h(\mu,\nu) < \infty\right\} = \tanh\frac{H(\mathcal{T})}{4},$$

where $H(\mathcal{T}) = \log esssup \frac{T(x,y)T(x',y')}{T(x,y')T(x',y)}$ with convention $0/0 = 1$ and $1/0 = \infty$. The supremum above is strict over $x, x' \in \mathbb{R}^d$, and is essential over $y, y' \in \mathbb{R}^d$ with respect to the distribution at the beginning of each time interval.

To obtain the exponential decay, it's sufficient to show the boundness of the kernel in the time interval $[0, 1]$ which is proved in [1]. $\square$

2.2 Numerical method

2.2.1 Branching particle system

In this section, we introduce the branching particle system and the particle filter studied by Crisan and Xiong [9]. The main idea of using branching mechanism is to reduce the variance of the weight of the particles in the system. We divide the time interval into small subintervals, and the weight for each particle is restarted at every partition time.

Now we proceed to defining the branching particle system. Initially, there are n particles of weight 1 each at locations $x_i^n, i = 1, 2, ..., n$, satisfying the following condition:

Assumption 2.2.1. *The initial positions $\{x_i^n : i = 1, 2, ..., n\}$ of the particles are i.i.d. random vectors in $\mathbb{R}^d$ with the common distribution $\pi_0 \in \mathcal{P}(\mathbb{R}^d)$.*

Let $\delta = \delta_n = n^{-2\alpha}, 0 < \alpha < 1$. For $j = 0, 1, 2, ...$, there are m_j^n number of particles alive at time $t = j\delta$. Note that $m_0^n = n$.

During the time interval $[j\delta, (j+1)\delta)$, the particles move according to the following diffusions: For $i = 1, 2, ..., m_j^n$,

$$X_t^i = X_{j\delta}^i + \int_{j\delta}^t \sigma(X_s^i)dB_s^i + \int_{j\delta}^t \tilde{b}(X_s^i)ds + \int_{j\delta}^t c(X_s^i)dY_s \tag{2.21}$$

where $\tilde{b} = b - ch$.

At the end of the interval, the ith particle ($i = 1, 2, ..., m_j^n$) branches (conditionally independent of others with $\mathcal{F}_{(j+1)\delta}$ given) into a random number ξ_{j+1}^i of offsprings such that the conditional expectation and the conditional variance given the information prior to the branching satisfy

$$\hat{\mathbb{E}}(\xi_{j+1}^i | \mathcal{F}_{(j+1)\delta-}) = \tilde{M}_{j+1}^n(X^i),$$

and

$$Var^{\hat{P}}(\xi^i_{j+1}|\mathcal{F}_{(j+1)\delta-}) = \gamma^n_{j+1}(X^i),$$

where $\gamma^n_{j+1}(X^i)$ is arbitrary,

$$\tilde{M}^n_{j+1}(X^i) = \frac{M^n_{j+1}(X^i)}{\frac{1}{m^n_j}\sum_{l=1}^{m^n_j} M^n_{j+1}(X^l)}$$

and

$$M^n_{j+1}(X^i) = \exp\left(\int_{j\delta}^{(j+1)\delta} h^*(X^i_t)dY_t - \frac{1}{2}\int_{j\delta}^{(j+1)\delta} |h(X^i_t)|^2 dt\right) \tag{2.22}$$

To minimize $\gamma^n_{j+1}(X^i)$, we take

$$\xi^i_{j+1} = \begin{cases} [\tilde{M}^n_{j+1}(X^i)] & \text{with probability } 1 - \{\tilde{M}^n_{j+1}(X^i)\}, \\ [\tilde{M}^n_{j+1}(X^i)] + 1 & \text{with probability } \{\tilde{M}^n_{j+1}(X^i)\} \end{cases}$$

where $\{x\} = x - [x]$ is the fraction of x, and $[x]$ is the largest integer that is not greater that x. In this case, we have

$$\gamma^n_{j+1}(X^i) = \{\tilde{M}^n_{j+1}(X^i)\}(1 - \{\tilde{M}^n_{j+1}(X^i)\}).$$

Now we define the approximate filter as follows:

$$\pi^n_t = \frac{1}{m^n_j}\sum_{i=1}^{m^n_j} \tilde{M}^n_{j+1}(X^i, t)\delta_{X^i_t}, \qquad j\delta \le t < (j+1)\delta,$$

where

$$M^n_j(X^i, t) = \exp\left(\int_{j\delta}^{t} h^*(X^i_s)dY_s - \frac{1}{2}\int_{j\delta}^{t} |h(X^i_s)|^2 ds\right) \tag{2.23}$$

and

$$\tilde{M}_j^n(X^i,t) = \frac{M_j^n(X^i,t)}{\frac{1}{m_j^n}\sum_{l=1}^{m_j^n} M_j^n(X^l,t)}$$

Namely, the ith particle has a time-dependent weight $\tilde{M}_j^n(X^i,t)$. At the end of the interval, i.e. $t = (j+1)\delta$, this particle dies and gives birth to a random number of offsprings, whose conditional expectation is equal to the pre-death weight of the particle. The new particles start from their mother's position with weight 1 each.

The process π_t^n is called the hybrid filter since it involves a branching particle system and the empirical measure of these weighted particles.

To show the uniform convergence, we also define the approximation for the unnormalized filter V_t^n as following: For $k\delta \le t < (k+1)\delta$,

$$V_t^n = \frac{1}{n}\eta_k^n \sum_{i=1}^{m_k^n} M_{k+1}^n(X^i,t)\delta_{X_t^i},$$

where

$$\eta_k^n = \Pi_{j=1}^k \frac{1}{m_{j-1}^n}\sum_{l=1}^{m_{j-1}^n} M_j^n(X^l).$$

We state the following lemmas whose proof can be found in [33].

Lemma 2.2.1. *Let $\phi \in C_b^k(\mathbb{R}^d) \cap W_2^k(\mathbb{R}^d)$ and ψ is a solution of the following backward SPDE:*

$$\begin{cases} d\psi_s = -L\psi_s ds - (\nabla^*\psi_s c + h^*\psi_s)\,\hat{d}Y_s \\ \psi_t = \phi, \end{cases} \tag{2.24}$$

where $\hat{d}$ denotes the backward Itô integral. Than, for every $t \ge 0$, we have

$$\psi_t(X_t)M_t - \psi_0(X_0) = \int_0^t M_s \nabla^*\psi_s \sigma(X_s) dB_s, \quad a.s.$$

Lemma 2.2.2. *There exist constants K and K' such that for any $i = 1, 2, ..., m_j^n$, we have*

$$\hat{\mathbb{E}}((M_j^n(X^i, s))^2 | \mathcal{F}_{j\delta} \vee \mathcal{F}^i_{j\delta,(j+1)\delta}) \leq e^{K^2\delta}, \qquad j\delta \leq s \leq (j+1)\delta$$

and

$$\hat{\mathbb{E}}(|M_{j+1}^n(X^i) - 1|^2 | \mathcal{F}_{j\delta}) \leq K'\delta,$$

where $\mathcal{F}^i_{j\delta,(j+1)\delta} = \sigma\{B_s^i - B_{j\delta}^i : j\delta \leq s \leq (j+1)\delta\}$ is the σ-filed generated by the increments of B_t^i in $t \in [j\delta, (j+1)\delta)]$.

Lemma 2.2.3. *There exists a constant K'' such that for any $j \geq 0$ and $i = 1, 2, ..., m_j^n$, we have*

$$\hat{\mathbb{E}}\left(\gamma_{j+1}^n(X^i)(\frac{\eta_{j+1}^n}{\eta_j^n})^2 \Big| \mathcal{F}_{j\delta}\right) \leq K''\sqrt{\delta}.$$

Remark 2.3. *In [22], another branching particle filter is defined to avoid the stochastic integral in (4.1) and (4.2). We describe it for the convenience of the reader.*

Let m_j^n be the number of particles at time $j\delta$. During the time interval $[j\delta, (j+1)\delta)$, the particles move according to the following equation:

$$X_t^i = X_{j\delta}^i + \tilde{b}(X_{j\delta}^i)(t - \delta) + c(X_{j\delta}^i)(Y_t - Y_{j\delta}) + \sigma(X_{j\delta}^i)(B_t^i - B_{j\delta}^i), \quad i = 1, ..., m_j^n$$

At the end of the interval, the ith particle ($i = 1, ..., m_j^n$) branches into a random number ξ_{j+1}^i of off-springs such that the conditional expectation and the conditional variance are given by:

$$\hat{\mathbb{E}}(\xi_{j+1}^i | \mathcal{F}_{(j+1)\delta-}) = \tilde{M}_{j+1}^n(X^i),$$

and

$$Var^{\hat{\mathbb{P}}}(\xi_{j+1}^i | \mathcal{F}_{(j+1)\delta-}) = \gamma_{j+1}^n(X^i) = \{\tilde{M}_{j+1}^n(X^i)\}\left(1 - \{\tilde{M}_{j+1}^n(X^i)\}\right),$$

where

$$\tilde{M}^n_{j+1}(X^i) = \frac{M^n_{j+1}(X^i)}{\frac{1}{m^n_j}\sum_{l=1}^{m^n_j} M^n_{j+1}(X^l)},$$

and

$$M^n_{j+1}(X^i) = \exp\left(h^*(X^i_{j\delta})(Y_{(j+1)\delta} - Y_{j\delta}) - \frac{1}{2}|h(X^i_{j\delta})|^2\delta\right)$$

The branching particle filter is defined by

$$\pi^n_t = \frac{1}{m^n_j}\sum_{i=1}^{m^n_j} \delta_{X^i_{j\delta}}, \qquad j\delta \le t < (j+1)\delta. \tag{2.25}$$

Similar convergence result can be proved for this branching filter by using the idea which will be given in the next section.

2.2.2 Uniform convergence

Define the distance on $\mathcal{M}_F(\mathbb{R}^d)$ by

$$d(\mu,\nu) = \sum_{i=0}^{\infty} 2^{-i}\left(|\langle \mu-\nu, f_i\rangle| \wedge 1\right), \quad \forall \mu,\nu \in \mathcal{M}_F(\mathbb{R}^d), \tag{2.26}$$

where $f_0 = 1$ and for $i \ge 1$, $f_i \in C_b^{k+2}(\mathbb{R}^d) \cup W_2^{k+2}(\mathbb{R}^d)$ with $\|f_i\|_{k+2,\infty} \le 1$ and also $\|f_i\|_{k+2,2} \le 1$, where $k = [\frac{d}{2}] + 2$ is given in assumption 2.1.1.

Theorem 2.2.1. *Under assumptions 2.1.1, 2.1.2 and 2.2.1, the branching particle filter uniformly converges to the optimal filter in the following sense:*

$$\lim_{n\to\infty} \sup_{t>0} \mathbb{E} d(\pi_t, \pi^n_t) = 0. \tag{2.27}$$

Remark 2.4. *Note that many exponential stability results are proved in compact state space case. Similar convergence results hold true and our proof can also be applied to them.*

Let $p(t,x,A)$ be the transition probability of the Markov process X_t. There exists a probability measure $P_{s,x}$ on $C(\mathbb{R}_+,\mathbb{R}^d)$ such that for $t>s$ and $A\in\mathcal{B}(\mathbb{R}^d)$,

$$P_{s,x}(\xi_t\in A|\mathcal{F}_s^{\xi}) = p(t-s,x,A), \qquad P_{s,x}-a.s.,$$

and

$$P_{s,x}(\xi_u = x, 0\le u\le s) = 1$$

where ξ_t is the co-ordinate process on $C(\mathbb{R}_+,\mathbb{R}^d)$, i.e. $\xi_t(\theta)=\theta_t$ for all $\theta\in C(\mathbb{R}_+,\mathbb{R}^d)$.

Then λ is the initial distribution of X_t and $\eta\in C(\mathbb{R}_+,\mathbb{R}^m)$. We define an $\mathcal{M}_F(\mathbb{R}^d)$-valued process $\Gamma_{s,t}(\lambda)$ and a $\mathcal{P}(\mathbb{R}^d)$-valued process $\Lambda_{s,t}(\lambda)$ on $C(\mathbb{R}_+,\mathbb{R}^m)$ as

$$\langle\Gamma_{s,t}(\lambda)(\eta),f\rangle = \int_{\mathbb{R}^d}\int_{C(\mathbb{R}_+,\mathbb{R}^d)} f(\xi_t(\theta))q_{st}(\theta,\eta)P_{s,x}(d\theta)\lambda(dx),$$

and

$$\Lambda_{s,t}(\lambda)(\eta) = \frac{\langle\Gamma_{s,t}(\lambda)(\eta),f\rangle}{\langle\Gamma_{s,t}(\lambda)(\eta),1\rangle},$$

where $q_{st}(\theta,\eta)=\exp\left(\int_s^t h(\xi_u(\theta))^* d\beta_u(\eta) - \frac{1}{2}\int_s^t |h(\xi_u(\theta))|^2du\right)$ and $\beta_t(\eta)=\eta_t$ is the co-ordinate process on $C(\mathbb{R}_+,\mathbb{R}^m)$.

Let $\Lambda_{k\delta,(k+1)\delta}(\lambda)(Y)$ be the optimal filter at time $(k+1)\delta$ using the observation $\sigma(Y_t, k\delta\le t\le(k+1)\delta)$ starting with λ at time $k\delta$. We define the following $\mathcal{P}(\mathbb{R}^d)$-valued processes, for $j\le k$

$$\pi^n_{j\delta,k\delta} := \Lambda_{j\delta,k\delta}(\pi^n_{j\delta})(Y) \;\; = \;\; \Lambda_{(k-1)\delta,k\delta}\circ\cdots\circ\Lambda_{j\delta,(j+1)\delta}(\pi^n_{j\delta})(Y),$$

The following is our strategy of the proof. For $k\delta \le t < (k+1)\delta$, we estimate the distance between π_t and π_t^n by the sum of three distances: $d(\pi_{k\delta}, \pi_t), d(\pi_{k\delta}, \pi_{k\delta}^n)$ and $d(\pi_{k\delta}^n, \pi_t^n)$ through the triangle inequality. To estimate the distance of $\pi_{k\delta}$ and $\pi_{k\delta}^n$, we bound it by a sum of k distances. Each term in the sum is the distance of two filters at the same time with different initials which are not far away from each other.

Namely, for $k\delta \le t < (k+1)\delta$, we have

$$\begin{aligned} d(\pi_t, \pi_t^n) &\le d(\pi_t, \pi_{k\delta}) + d(\pi_{k\delta}, \pi_{k\delta}^n) + d(\pi_{k\delta}^n, \pi_t^n) \\ &\le d(\pi_t, \pi_{k\delta}) + \sum_{i=0}^{k-1} d(\pi_{i\delta,k\delta}^n, \pi_{(i+1)\delta,k\delta}^n) + d(\pi_{k\delta}^n, \pi_t^n) \end{aligned} \tag{2.28}$$

Remark 2.5. *Note that, in the definition (2.25), $\pi_t^n = \pi_{k\delta}^n$ for $k\delta \le t < (k+1)\delta$. Then the third of (2.28) vanishes for the filter studied by [22].*

We start with the estimate of the first term on the right side of (2.28).

Lemma 2.2.4. *There exists a constant K_1 such that*

$$\mathbb{E} d(\pi_t, \pi_{k\delta}) \le K_1 \sqrt{\delta}$$

Proof. Since π_t satisfies Kushner-FKK equation, we have, for $f_i \in C_b^{k+2}(\mathbb{R}^d) \cap W_2^{k+2}(\mathbb{R}^d)$,

$$I_i := \langle \pi_t, f_i \rangle - \langle \pi_{k\delta}, f_i \rangle = \int_{k\delta}^t \langle \pi_s, L f_i \rangle \, ds + \int_{k\delta}^t (\langle \pi_s, \nabla^* f_i c + f_i h^* \rangle - \langle \pi_s, f_i \rangle \langle \pi_s, h^* \rangle) d\nu_s,$$

Then

$$\begin{aligned} \mathbb{E}(I_i^2) &\le 3\delta \mathbb{E} \int_{k\delta}^t \langle \pi_s, L f_i \rangle^2 \, ds + 3\mathbb{E} \int_{k\delta}^t (\langle \pi_s, \nabla^* f_i c + f_i h^* \rangle - \langle \pi_s, f_i \rangle \langle \pi_s, h^* \rangle)^2 ds \\ &\le K_2 \delta. \end{aligned}$$

Thus

$$\mathbb{E}d(\pi_t, \pi_{k\delta}) \leq \sum_{i=0}^{\infty} 2^{-i}\sqrt{K_2\delta} = 2\sqrt{K_2\delta}. \tag{2.29}$$

□

We now estimate the third term on the right side of (2.28).

Lemma 2.2.5. *There exists a constant K_3 such that*

$$\mathbb{E}d(\pi_t^n, \pi_{k\delta}^n) \leq K_3\sqrt{\delta}$$

Proof. Let $f \in C_b^{k+2}(\mathbb{R}^d) \cap W_2^{k+2}(\mathbb{R}^d)$ with $\|f\|_{k+2,\infty} \leq 1$. By the definition of π^n, we have

$$\begin{aligned}
|\langle \pi_t^n - \pi_{k\delta}^n, f\rangle| &= \Big|\frac{1}{m_k^n}\sum_{i=1}^{m_k^n} \tilde{M}(X^i,t)f(X_t^i) - \frac{1}{m_k^n}\sum_{i=1}^{m_k^n} f(X_{k\delta}^i)\Big| \\
&= \Big|\frac{1}{m_k^n}\sum_{i=1}^{m_k^n}(\tilde{M}(X^i,t)f(X_t^i) - f(X_t^i) + f(X_t^i) - f(X_{k\delta}^i))\Big| \\
&\leq \frac{1}{m_k^n}\sum_{i=1}^{m_k^n}|\tilde{M}(X^i,t)f(X_t^i) - f(X_t^i)| + \frac{1}{m_k^n}\sum_{i=1}^{m_k^n}|f(X_t^i) - f(X_{k\delta}^i)| \\
&\leq K\frac{1}{m_k^n}\sum_{i=1}^{m_k^n}|\tilde{M}(X^i,t) - 1| + K\frac{1}{m_k^n}\sum_{i=1}^{m_k^n}|X_t^i - X_{k\delta}^i|
\end{aligned}$$

By the proof of Lemma 3.1 in [9], we have

$$\tilde{M}^n(X^i,t) = 1 + \int_{k\delta}^t \tilde{M}^n(X^i,s)(h^*(X_s^i) - \bar{h}_s^*)(dY_s - \bar{h}_s)ds,$$

where $\bar{h}_s = \frac{1}{m_k^n}\sum_{i=1}^{m_k^n} \tilde{M}_j^n(X^i,s)h(X_s^i)$.

Since $h(X) \leq K$ and $\sum_{i=1}^{m_k^n} \tilde{M}(X^i, s) = m_k^n$, we have $|\bar{h}_s| \leq K$. Thus,

$$\begin{aligned}
\hat{\mathbb{E}}|\tilde{M}^n(X^i,t)|^2 &\leq 3 + 3\hat{\mathbb{E}}\Big|\int_{j\delta}^t \tilde{M}^n(X^i,s)(h^*(X_s^i) - \bar{h}_s^*)dY_s\Big|^2 \\
&\quad +3\delta\hat{\mathbb{E}}\int_{k\delta}^t |\tilde{M}^n(X^i,s)(h^*(X_s^i) - \bar{h}_s^*)\bar{h}_s|^2 ds \\
&\leq 3 + 3\hat{\mathbb{E}}\int_{j\delta}^t \tilde{M}^n(X^i,s)^2(h^*(X_s^i) - \bar{h}_s^*)^2 ds \\
&\quad +3\delta\hat{\mathbb{E}}\int_{k\delta}^t |\tilde{M}^n(X^i,s)(h^*(X_s^i) - \bar{h}_s^*)\bar{h}_s|^2 ds \\
&\leq 3 + K\int_{j\delta}^t \hat{\mathbb{E}}\tilde{M}^n(X^i,s)^2 dr.
\end{aligned}$$

By Gronwall's inequality,

$$\hat{\mathbb{E}}|\tilde{M}^n(X^i,t)|^2 \leq K_4.$$

Then

$$\begin{aligned}
\hat{\mathbb{E}}|\tilde{M}^n(X^i,t) - 1|^2 &= \hat{\mathbb{E}}\Big|\int_{k\delta}^t \tilde{M}^n(X^i,s)(h^*(X_s^i) - \bar{h}_s^*)(dY_s - \bar{h}_s ds)\Big|^2 \\
&\leq 2\hat{\mathbb{E}}\Big|\int_{k\delta}^t \tilde{M}^n(X^i,s)(h^*(X_s^i) - \bar{h}_s^*)dY_s\Big|^2 \\
&\quad +2\hat{\mathbb{E}}\Big|\int_{k\delta}^t \tilde{M}^n(X^i,s)(h^*(X_s^i) - \bar{h}_s^*)\bar{h}_s ds\Big|^2 \\
&\leq 2\hat{\mathbb{E}}\int_{k\delta}^t \tilde{M}^n(X^i,s)^2(h^*(X_s^i) - \bar{h}_s^*)^2 ds \\
&\quad +2\delta\hat{\mathbb{E}}\int_{k\delta}^t \tilde{M}^n(X^i,s)^2(h^*(X_s^i) - \bar{h}_s^*)^2\bar{h}_s^2 ds \\
&\leq K_5\delta. \qquad (2.30)
\end{aligned}$$

On the other hand,

$$\begin{aligned}
\hat{\mathbb{E}}|X_t^i - X_{k\delta}^i|^2 &= \hat{\mathbb{E}}|\int_{j\delta}^t \tilde{b}(X_s^i)ds + \int_{j\delta}^t c(X_s^i)dY_s + \int_{j\delta}^t \sigma(X_s^i)dB_s^i|^2 \\
&\leq 3\delta\hat{\mathbb{E}}\int_{j\delta}^t |\tilde{b}(X_s^i)|^2 ds + 3\hat{\mathbb{E}}\int_{j\delta}^t |c(X_s^i)|^2 ds + \hat{\mathbb{E}}\int_{j\delta}^t |\sigma(X_s^i)|^2 ds \\
&\leq K_6\delta. \qquad (2.31)
\end{aligned}$$

Therefore,

$$\begin{aligned}\hat{\mathbb{E}}\left\langle \pi^n_t-\pi^n_{k\delta},f\right\rangle^2 &\le K\hat{\mathbb{E}}\frac{1}{m^n_k}\sum_{i=1}^{m^n_k}\hat{\mathbb{E}}(|\tilde{M}^n(X^i,t)-1|^2+|X^i_t-X^i_{k\delta}|^2|\mathcal{F}_{k\delta})\\ &\le K_7\delta.\end{aligned}$$

Let

$$M_{s,t}\equiv\exp\Big(\int_s^t h^*(X_s)dY_s-\frac{1}{2}\int_s^t|h(X_s)|^2ds\Big).$$

Then

$$\begin{aligned}\hat{\mathbb{E}}M^2_{s,t} &= \hat{\mathbb{E}}\exp\Big(2\int_s^t h^*(X_s)dY_s-\int_s^t|h(X_s)|^2ds\Big)\\ &= \hat{\mathbb{E}}\exp\Big(\int_s^t 2h^*(X_s)dY_s-\frac{1}{2}\int_s^t|2h(X_s)|^2ds\Big)\exp\Big(\int_s^t|2h(X_s)|^2\Big)\\ &\le e^{K(t-s)}.\end{aligned}\tag{2.32}$$

By Cauchy-Schwarz inequality, we have

$$\begin{aligned}\mathbb{E}d(\pi^n_t,\pi^n_{k\delta}) &= \hat{\mathbb{E}}(d(\pi^n_t,\pi^n_{k\delta})M_{k\delta,t})\\ &\le (\hat{\mathbb{E}}d(\pi^n_t,\pi^n_{k\delta})^2)^{1/2}(\hat{\mathbb{E}}M^2_{k\delta,t})^{1/2}\\ &\le (\sum_{i=0}^{\infty}2^{-i}\hat{\mathbb{E}}\left\langle\pi^n_t-\pi^n_{k\delta},f_i\right\rangle^2)^{1/2}(\hat{\mathbb{E}}M^2_{k\delta,t})^{1/2}\\ &\le K_3\sqrt{\delta}.\end{aligned}$$

□

To estimate $d(\pi^n_{i\delta,k\delta},\pi^n_{(i+1)\delta,k\delta})$ in (2.28), we will use the stability assumption. Note that, $\pi^n_{j\delta,k\delta}$ is the optimal filter at time $k\delta$ starting at time $(j+1)\delta$ with measure $\pi^n_{j\delta,(j+1)\delta}$. Similarly, $\pi^n_{(j+1)\delta,k\delta}$ is the optimal filter at the same time but with initial $\pi^n_{(j+1)\delta}$ at the initial time $(j+1)\delta$. For any $j\in\mathbb{N}$, we have

Lemma 2.2.6. *There exists a constant K_8 such that*

$$\mathbb{E}d(\pi^n_{j\delta,(j+1)\delta}, \pi^n_{(j+1)\delta}) \leq K_8 \delta^{1/4}$$

Proof. Note that $\pi^n_{j\delta,(j+1)\delta}$ and $\pi^n_{(j+1)\delta}$ have the same initial distribution $\pi^n_{j\delta}$ at time $j\delta$. Let $V^n_{j\delta,(j+1)\delta}$ and $V^n_{(j+1)\delta}$ be the unnormalized optimal filter and unnormalized particle filter, respectively, with the same initial distribution $\pi^n_{j\delta}$. Note that for ϕ bounded by 1, we have

$$\left| \left\langle \pi^n_{j\delta,(j+1)\delta} - \pi^n_{(j+1)\delta}, \phi \right\rangle \right| \leq \frac{\left| \left\langle V^n_{j\delta,(j+1)\delta} - V^n_{(j+1)\delta}, 1 \right\rangle \right|}{\left\langle V^n_{j\delta,(j+1)\delta}, 1 \right\rangle} + \frac{\left| \left\langle V^n_{j\delta,(j+1)\delta} - V^n_{(j+1)\delta}, \phi \right\rangle \right|}{\left\langle V^n_{j\delta,(j+1)\delta}, 1 \right\rangle}$$

As

$$d \log \left\langle V^n_{j\delta,t}, 1 \right\rangle = \left\langle \pi^n_{j\delta,t}, h^* \right\rangle dY_t - \frac{1}{2} \left| \left\langle \pi^n_{j\delta,t}, h^* \right\rangle \right|^2 dt, \qquad j\delta \leq t \leq (j+1)\delta$$

we have

$$\left\langle V^n_{j\delta,(j+1)\delta}, 1 \right\rangle^{-1} = \left\langle V^n_{j\delta}, 1 \right\rangle^{-1} \exp \left(- \int_{j\delta}^{(j+1)\delta} \left\langle \pi^n_{j\delta,t}, h^* \right\rangle dY_t + \frac{1}{2} \int_{j\delta}^{(j+1)\delta} \left| \left\langle \pi^n_{j\delta,t}, h^* \right\rangle \right|^2 dt \right)$$

Then

$$\begin{aligned} & \left(\hat{\mathbb{E}} \left| \frac{\left\langle V^n_{(j+1)\delta} - V^n_{j\delta,(j+1)\delta}, \phi \right\rangle}{\left\langle V^n_{j\delta,(j+1)\delta}, 1 \right\rangle} \right| \right)^2 \\ \leq \; & \hat{\mathbb{E}} \left| \frac{\left\langle V^n_{(j+1)\delta} - V^n_{j\delta,(j+1)\delta}, \phi \right\rangle}{\left\langle V^n_{j\delta}, 1 \right\rangle} \right|^2 \hat{\mathbb{E}} \exp \left(-2 \int_{j\delta}^{(j+1)\delta} \left\langle \pi_t, h^* \right\rangle dY_t + \int_{j\delta}^{(j+1)\delta} \left| \left\langle \pi_t, h^* \right\rangle \right|^2 dt \right) \\ \leq \; & K \hat{\mathbb{E}} \left| \frac{\left\langle V^n_{(j+1)\delta} - V^n_{j\delta,(j+1)\delta}, \phi \right\rangle}{\left\langle V^n_{j\delta}, 1 \right\rangle} \right|^2 \qquad (2.33) \end{aligned}$$

where $f \in C_b^{k+2}(\mathbb{R}^d) \cap W_2^{k+2}(\mathbb{R}^d)$ is any test function. Now we show

$$\left(\hat{\mathbb{E}}\left|\frac{\left\langle V^n_{(j+1)\delta} - V^n_{j\delta,(j+1)\delta}, \phi\right\rangle}{\left\langle V^n_{j\delta}, 1\right\rangle}\right|\right)^2 \le K\delta^{1/2}. \tag{2.34}$$

Then, by corollary 6.22 in [33], we get

$$\begin{aligned}
& \frac{\left\langle V^n_{(j+1)\delta}, \phi\right\rangle - \left\langle V^n_{j\delta,(j+1)\delta}, \phi\right\rangle}{\left\langle V^n_{j\delta}, 1\right\rangle} \\
= & \frac{\left\langle V^n_{(j+1)\delta}, \psi_{(j+1)\delta}\right\rangle - \left\langle V^n_{j\delta}, \psi_{j\delta}\right\rangle}{\left\langle V^n_{j\delta}, 1\right\rangle} \\
= & \frac{\left\langle V^n_{(j+1)\delta}, \psi_{(j+1)\delta}\right\rangle - \hat{\mathbb{E}}\Big(\left\langle V^n_{(j+1)\delta}, \psi_{(j+1)\delta}\right\rangle \Big| \mathcal{F}_{(j+1)\delta-}\Big)}{\left\langle V^n_{j\delta}, 1\right\rangle} \\
& + \frac{\hat{\mathbb{E}}\Big(\left\langle V^n_{(j+1)\delta}, \psi_{(j+1)\delta}\right\rangle \Big| \mathcal{F}_{(j+1)\delta-}\Big) - \left\langle V^n_{j\delta}, \psi_{j\delta}\right\rangle}{\left\langle V^n_{j\delta}, 1\right\rangle}
\end{aligned} \tag{2.35}$$

By the definition of V^n, we have

$$\begin{aligned}
& \frac{\left\langle V^n_{(j+1)\delta}, \psi_{(j+1)\delta}\right\rangle - \hat{\mathbb{E}}\left(\left\langle V^n_{(j+1)\delta}, \psi_{(j+1)\delta}\right\rangle \Big| \mathcal{F}_{(j+1)\delta-}\right)}{\left\langle V^n_{j\delta}, 1\right\rangle} \\
= & \frac{\frac{1}{n}\eta_{j+1}\sum_{i=1}^{m_j^n}\psi(X^i_{(j+1)\delta})(\xi^i_{j+1} - \tilde{M}^n_{j+1}(X^i))}{\frac{1}{n}\eta_j m_j^n} \\
= & \frac{\eta_{j+1}}{\eta_j}\frac{1}{m_j^n}\sum_{i=1}^{m_j^n}\psi(X^i_{(j+1)\delta})(\xi^i_{j+1} - \tilde{M}^n_{j+1}(X^i)) \\
\equiv & J_1.
\end{aligned}$$

Note that

$$\gamma^n_{j+1}(X^i) \le |\tilde{M}^n_{j+1}(X^i) - 1|.$$

By (4.15), we have

$$\begin{aligned}\hat{\mathbb{E}}(\gamma_{j+1}^n(X^i)|\mathcal{F}_{j\delta}) &\leq \mathbb{E}(|\tilde{M}_{j+1}^n(X^i)-1||\mathcal{F}_{j\delta}) \\ &\leq K_9\sqrt{\delta}.\end{aligned}$$

It follows from the independent increment property of Y that

$$\hat{\mathbb{E}}(\xi_j^i-\tilde{M}_j^n(X^i)|\mathcal{F}_{j\delta-}\vee\mathcal{G}_t)=\hat{\mathbb{E}}(\xi_j^i-\tilde{M}_j^n(X^i)|\mathcal{F}_{j\delta-})=0.$$

Then

$$\begin{aligned}\hat{\mathbb{E}}(J_1^2) &= \hat{\mathbb{E}}\Big|\frac{\eta_{j+1}}{\eta_j}\frac{1}{m_j^n}\sum_{i=1}^{m_j^n}\psi(X_{(j+1)\delta}^i)(\xi_{j+1}^i-\tilde{M}_{j+1}^n(X^i))\Big|^2 \\ &= \hat{\mathbb{E}}\left(\hat{\mathbb{E}}\left(\left(\frac{1}{m_j^n}\sum_{i=1}^{m_j^n}\psi(X_{(j+1)\delta}^i)(\xi_{j+1}^i-\tilde{M}_{j+1}^n(X^i))\right)^2|\mathcal{F}_{(j+1)\delta-}\vee\mathcal{G}_t\right)(\frac{\eta_{j+1}}{\eta_j})^2\right) \\ &= \hat{\mathbb{E}}\frac{1}{(m_j^n)^2}\sum_{i=1}^{m_j^n}\psi^2(X_{(j+1)\delta}^i)\gamma_{j+1}^n(X^i)(\frac{\eta_{j+1}}{\eta_j})^2 \\ &= \hat{\mathbb{E}}\frac{1}{(m_j^n)^2}\sum_{i=1}^{m_j^n}\hat{\mathbb{E}}\left(\psi^2(X_{(j+1)\delta}^i)\gamma_{j+1}^n(X^i)(\frac{\eta_{j+1}}{\eta_j})^2|\mathcal{F}_{j\delta}\right).\end{aligned}$$

By Lemma (2.2.3) and note that $m_j^n\geq 1$, we can continue the above estimate with

$$\begin{aligned}\hat{\mathbb{E}}(J_1^2) &\leq K''\|\psi_{(j+1)\delta}\|_{0,\infty}^2\hat{\mathbb{E}}\frac{1}{m_j^n}\sqrt{\delta} \\ &\leq K_{10}\sqrt{\delta}. \qquad (2.36)\end{aligned}$$

On the other hand,

$$
\begin{aligned}
& \frac{\hat{\mathbb{E}}(\left\langle V^n_{(j+1)\delta}, \psi_{(j+1)\delta}\right\rangle |\mathcal{F}_{(j+1)\delta-}) - \left\langle V^n_{j\delta}, \psi_{j\delta}\right\rangle}{\left\langle V^n_{j\delta}, 1\right\rangle} \\
= & \frac{\frac{1}{n}\eta_{j+1}\sum_{i=1}^{m_j^n}\psi(X^i_{(j+1)\delta})\tilde{M}^n_{j+1}(X^i) - \frac{1}{n}\eta_j\sum_{i=1}^{m_j^n}\psi(X^i_{j\delta})}{\frac{1}{n}\eta_j m_j^n} \\
= & \frac{1}{m_j^n}\sum_{i=1}^{m_j^n}\left(\psi(X^i_{(j+1)\delta})M^n_{j+1}(X^i) - \psi(X^i_{j\delta})\right) \\
= & \frac{1}{m_j^n}\sum_{i=1}^{m_j^n}\int_{j\delta}^{(j+1)\delta} M^n_j(X^i,s)\nabla^*\psi_s\sigma(X^i_s)dB^i_s \\
\equiv & J_2.
\end{aligned}
$$

Note that

$$
\hat{\mathbb{E}}\left(\int_{j\delta}^{(j+1)\delta} M^n_j(X^i,s)\nabla^*\psi_s\sigma(X^i_s)dB^i_s\middle|\mathcal{F}_{j\delta}\right) = 0
$$

Thus, we have

$$
\begin{aligned}
\hat{\mathbb{E}}|J_2|^2 = & \hat{\mathbb{E}}\left(\frac{1}{(m_j^n)^2}\sum_{i=1}^{m_j^n}\int_{j\delta}^{(j+1)\delta} M^n_j(X^i,s)^2|\nabla^*\psi_s\sigma(X^i_s)|^2 ds\right) \\
= & \hat{\mathbb{E}}\left(\frac{1}{(m_j^n)^2}\sum_{i=1}^{m_j^n}\int_{j\delta}^{(j+1)\delta} \hat{\mathbb{E}}\left(M^n_j(X^i,s)^2|\nabla^*\psi_s\sigma(X^i_s)|^2 ds|\mathcal{F}_{j\delta}\vee\mathcal{F}^i_{j\delta,(j+1)\delta}\right)\right) \\
= & \hat{\mathbb{E}}\Big(\frac{1}{(m_j^n)^2}\sum_{i=1}^{m_j^n}\int_{j\delta}^{(j+1)\delta} \hat{\mathbb{E}}\left(M^n_j(X^i,s)^2|\mathcal{F}_{j\delta}\vee\mathcal{F}^i_{j\delta,(j+1)\delta}\right) \\
& \hat{\mathbb{E}}\left(|\nabla^*\psi_s\sigma(X^i_s)|^2|\mathcal{F}_{j\delta}\vee\mathcal{F}^i_{j\delta,(j+1)\delta}\right) ds\Big) \\
\leq & K_{11}\delta, \qquad (2.37)
\end{aligned}
$$

where the last inequality follows from $\frac{1}{m_j^n} \leq 1$ and the boundness of $\hat{\mathbb{E}}(|\nabla^*\psi_s|^2|\mathcal{F}_{j\delta})$ which is showed on page 146 in [33].

Combine(2.35), (2.36) and (2.37), we prove (4.37). By (4.37) and (4.37), we have

$$\hat{\mathbb{E}}d(\pi^n_{j\delta,(j+1)\delta}, \pi^n_{(j+1)\delta})^2 \le K_{12}\sqrt{\delta}.$$

Then by (4.17), we have

$$\begin{aligned}
\mathbb{E}d(\pi^n_{j\delta,(j+1)\delta}, \pi^n_{(j+1)\delta}) &\le \hat{\mathbb{E}}(d(\pi^n_{j\delta,(j+1)\delta}, \pi^n_{(j+1)\delta})M_{j\delta,(j+1)\delta}) \\
&\le (\hat{\mathbb{E}}d(\pi^n_{j\delta,(j+1)\delta}, \pi^n_{(j+1)\delta})^2)^{1/2}(\hat{\mathbb{E}}M_{j\delta,(j+1)\delta})^2)^{1/2} \\
&\le (2K_{12}\sqrt{\delta})^{1/2}e^{K\delta} \\
&\le K_8\delta^{1/4}.
\end{aligned}$$

□

In fact, we can prove a stronger result for total variation distance as following:

Lemma 2.2.7.

$$\lim_{\delta\to 0}\mathbb{E}d_{TV}(\pi^n_{j\delta,(j+1)\delta}, \pi^n_{(j+1)\delta}) = 0 \tag{2.38}$$

Proof. For continuous f bounded by 1, we have

$$\left|\left\langle \pi^n_{j\delta,(j+1)\delta} - \pi^n_{(j+1)\delta}, f\right\rangle\right| \le \left|\left\langle \pi^n_{j\delta,(j+1)\delta}, f\right\rangle - \left\langle \pi^n_{j\delta}, f\right\rangle\right| + \left|\left\langle \pi^n_{j\delta}, f\right\rangle - \left\langle \pi^n_{(j+1)\delta}, f\right\rangle\right|$$

Similarly to lemma 2.2.7, by using Kushner-FKK equation, we have

$$\hat{\mathbb{E}}\left|\left\langle \pi^n_{j\delta,(j+1)\delta}, f\right\rangle - \left\langle \pi^n_{j\delta}, f\right\rangle\right| \le K\delta^{1/2}$$

On the other hand,

$$\hat{\mathbb{E}}\left|\left\langle \pi^n_{j\delta}, f\right\rangle - \left\langle \pi^n_{(j+1)\delta}, f\right\rangle\right| \le \hat{\mathbb{E}}\frac{\left|\left\langle V^n_{(j+1)\delta} - V^n_{j\delta}, f\right\rangle\right|}{\left\langle V^n_{j\delta}, 1\right\rangle} + \hat{\mathbb{E}}\frac{\left|\left\langle V^n_{(j+1)\delta} - V^n_{j\delta}, 1\right\rangle\right|}{\left\langle V^n_{j\delta}, 1\right\rangle}$$

Now we show

$$\hat{\mathbb{E}}\frac{\left|\left\langle V^n_{(j+1)\delta}-V^n_{j\delta,(j+1)\delta},f\right\rangle\right|}{\left\langle V^n_{j\delta},1\right\rangle}\to\infty,\quad asn\to\infty. \tag{2.39}$$

By the definition of V^n, we have

$$\begin{aligned}
&\hat{\mathbb{E}}\frac{\left|\left\langle V^n_{(j+1)\delta},f\right\rangle-\left\langle V^n_{j\delta},f\right\rangle\right|}{\left\langle V^n_{j\delta},1\right\rangle}\\
=\ &\hat{\mathbb{E}}\frac{\left|\frac{1}{n}\eta_j\sum_{i=1}^{m^n_j}\left(f(X^i_{(j+1)\delta})M^n_{j+1}(X^i)-f(X^i_{j\delta})\right)\right|}{\frac{1}{n}\eta_j m^n_j}\\
\leq\ &\hat{\mathbb{E}}\left(\frac{1}{m^n_j}\sum_{i=1}^{m^n_j}\left(\left|f(X^i_{(j+1)\delta})M^n_{j+1}(X^i)-f(X^i_{(j+1)\delta})\right|+\left|f(X^i_{(j+1)\delta})-f(X^i_{j\delta})\right|\right)\right)\\
=\ &\hat{\mathbb{E}}\left(\frac{1}{m^n_j}\sum_{i=1}^{m^n_j}\hat{\mathbb{E}}\left(\left|f(X^i_{(j+1)\delta})(M^n_{j+1}(X^i)-1)\right|+\left|f(X^i_{(j+1)\delta})-f(X^i_{j\delta})\right|\Big|\mathcal{F}_{j\delta}\right)\right)\\
\leq\ &\hat{\mathbb{E}}\left(\frac{1}{m^n_j}\sum_{i=1}^{m^n_j}\hat{\mathbb{E}}\left(\left|M^n_{j+1}(X^i)-1\right|\Big|\mathcal{F}_{j\delta}\right)+\frac{1}{m^n_j}\sum_{i=1}^{m^n_j}\hat{\mathbb{E}}\left(\left|f(X^i_{(j+1)\delta})-f(X^i_{j\delta})\right|\Big|\mathcal{F}_{j\delta}\right)\right)\\
\leq\ &e(\delta),
\end{aligned}$$

where the last inequality is from Lemma 8.6 in [33], (2.21) and the continuity of f, and $e(\delta)$ approaches 0 as δ tends to 0.

Then by (4.17), we have

$$\begin{aligned}
\mathbb{E}d_{TV}(\pi^n_{j\delta,(j+1)\delta},\pi^n_{(j+1)\delta}) &\leq \hat{\mathbb{E}}(d_{TV}(\pi^n_{j\delta,(j+1)\delta},\pi^n_{(j+1)\delta})M_{j\delta,(j+1)\delta})\\
&\leq (\hat{\mathbb{E}}d_{TV}(\pi^n_{j\delta,(j+1)\delta},\pi^n_{(j+1)\delta})^2)^{1/2}(\hat{\mathbb{E}}M_{j\delta,(j+1)\delta})^2)^{1/2}\\
&\leq (K\sqrt{\delta}+e(\delta))^{1/2}e^{K\delta}
\end{aligned}$$

Let δ tend to 0, we have (2.2.7). □

By assumption 2.1.2 and lemma 2.2.7, choose large n such that $K_8\delta^{1/4} < \epsilon$ and when $(k-(j+1))\delta > T$, we have

$$\begin{aligned}
\mathbb{E}d(\pi^n_{j\delta,k\delta}, \pi^n_{(j+1)\delta,k\delta}) &\le \mathbb{E}d_{TV}(\pi^n_{j\delta,k\delta}, \pi^n_{(j+1)\delta,k\delta}) \\
&\le C\beta^{(k-j-1)\delta}\mathbb{E}d_{TV}(\pi^n_{j\delta,(j+1)\delta}, \pi^n_{(j+1)\delta}) \\
&\le K'\beta^{(k-j-1)\delta}(K\sqrt{\delta}+e(\delta))^{1/2}
\end{aligned}$$

Now we are ready to prove the main theorem.

Proof of Theorem 3.1. By lemma 2.2.6, We have

$$\lim_{n\to\infty} \mathbb{E}d(\pi^n_{j\delta,(j+1)\delta}, \pi^n_{(j+1)\delta}) = \lim_{n\to\infty} K_8 n^{\alpha/2} = 0$$

For any $t > 0$, let k be such that $k\delta \le t < (k+1)\delta$. Then

$$\begin{aligned}
&\mathbb{E}d(\pi_t, \pi^n_t) \\
\le\;& \mathbb{E}d(\pi_t, \pi_{k\delta}) + \mathbb{E}\sum_{i=0}^{k-1} d(\pi^n_{i\delta,k\delta}, \pi^n_{(i+1)\delta,k\delta}) + \mathbb{E}d(\pi^n_{k\delta}, \pi^n_t) \\
\le\;& \mathbb{E}d(\pi_t, \pi_{k\delta}) + \mathbb{E}\sum_{i=k-j_0+1}^{k-1} d(\pi^n_{i\delta,k\delta}, \pi^n_{(i+1)\delta,k\delta}) + \mathbb{E}\sum_{i=0}^{k-j_0} d(\pi^n_{i\delta,k\delta}, \pi^n_{(i+1)\delta,k\delta}) \\
&+\mathbb{E}d(\pi^n_{k\delta}, \pi^n_t) \\
\le\;& K_1\sqrt{\delta} + \mathbb{E}\sum_{i=k-j_0+1}^{k-1} d(\pi^n_{i\delta,k\delta}, \pi^n_{(i+1)\delta,k\delta}) + \mathbb{E}\sum_{i=0}^{\infty} d_{TV}(\pi^n_{i\delta,k\delta}, \pi^n_{(i+1)\delta,k\delta}) + K_3\sqrt{\delta} \\
\le\;& K_1\sqrt{\delta} + \mathbb{E}\sum_{i=k-j_0+1}^{k-1} d(\pi^n_{i\delta,k\delta}, \pi^n_{(i+1)\delta,k\delta}) + K_2(K\sqrt{\delta}+e(\delta))^{1/2} + K_3\sqrt{\delta} \qquad (2.40)
\end{aligned}$$

where j_0 is the smallest integer such that $(k-(j+1))\delta > T$, and the last inequality follows $\sum_{i=0}^{\infty}\beta^i < \infty$.

By Zakai's equation, for any $t > 0$ and $f \in C_b^k(\mathbb{R}^d) \cup W_2^k(\mathbb{R}^d)$, we have

$$
\begin{aligned}
|\langle \pi_t - \bar{\pi}_t, f\rangle| &\leq \frac{|\langle \pi_t - \bar{\pi}_t, f\rangle|}{\langle V_t, 1\rangle} + \frac{|\langle \pi_t - \bar{\pi}_t, 1\rangle|}{\langle V, 1\rangle} \\
&\leq \frac{\left|\left\langle V_0 - \bar{V}_0, f\right\rangle + \int_0^t \left\langle V_s - \bar{V}_s, Lf\right\rangle ds + \int_0^t \left\langle V_s - \bar{V}_s, \nabla^* f + f h^*\right\rangle dY_s\right|}{\langle V_t, 1\rangle} \\
&\quad + \frac{\left|\left\langle V_0 - \bar{V}_0, 1\right\rangle + \int_0^t \left\langle V_s - \bar{V}_s, h^*\right\rangle dY_s\right|}{\langle V_t, 1\rangle}
\end{aligned}
$$

Then by Gronwall's inequality, for any t bounded, we have

$$
\mathbb{E} d(\pi_t, \bar{\pi}_t) \leq K d(\pi_0, \bar{\pi}_0) \tag{2.41}
$$

Therefore, by lemma (2.2.6), we get

$$
\begin{aligned}
\mathbb{E} \sum_{i=k-j_0+1}^{k-1} d(\pi^n_{i\delta, k\delta}, \pi^n_{(i+1)\delta, k\delta}) &\leq \sum_{i=k-j_0+1}^{k-1} K \mathbb{E} d(\pi^n_{i\delta,(i+1)\delta}, \pi^n_{(i+1)\delta,(i+1)\delta}) \\
&\leq K_4 \delta^{1/4}
\end{aligned} \tag{2.42}
$$

Therefore,

$$
\begin{aligned}
\lim_{n\to\infty} \sup_{t>0} \mathbb{E} d(\pi_t, \pi_t^n) &\leq \lim_{n\to\infty} ((K_1 + K_3)\delta^{1/2} + K_2(K\sqrt{\delta} + e(\delta))^{1/2} + K_4 \delta^{1/4}) \\
&= \lim_{n\to\infty} ((K_1 + K_3) n^{-\alpha} + K_2(K n^{-\alpha} + e(\delta)) + K_4 n^{-\frac{\alpha}{2}}) \\
&= 0.
\end{aligned}
$$

We state the uniform convergence theorem for the branching particle filter in [22] and give a sketch of the proof.

Theorem 2.2.2. *Under assumption 2.1.1, 2.1.2 and 2.2.1, the branching particle filter which is defined in [22] uniformly converges to the optimal filter in the following* *sense:*

$$\lim_{n\to\infty}\sup_{t>0}\mathbb{E}d(\pi_t,\pi_t^n)=0. \tag{2.43}$$

The following lemma (see[22]) is needed in the proof. Let

$$\Theta_s(x)=\hat{\mathbb{E}}(\psi_s(x)\tilde{\theta}_f^Y(s)|\mathcal{F}_s),\quad \forall x\in\mathbb{R}^d,$$

where

$$\mathcal{F}_s=\mathcal{G}_s\vee\mathcal{F}_s^B,$$

and

$$\tilde{\theta}_f^Y(s)=\exp\left(\sqrt{(-1)}\int_s^t f_s^*dY_s+\frac{1}{2}\int_s^t|f_s|^2ds\right)$$

Lemma 2.2.8. *Let ψ_t be the solution of (2.24). Then for every $t\geq 0$, we have*

$$\begin{aligned}\psi_t(X_t)M_t-\psi_0(X_s) &= \int_s^t M_u\nabla^*\psi_u\sigma(X_u)dB_u-\int_s^t M_u\Theta(X_u)(h(X_u)-h(X_s))^*dY_u\\ &\quad -\int_s^t M_s\nabla^*\psi_u c(X_u)(h(X_u)-h(X_s))^*du,\end{aligned}$$

where

$$M_t=\exp\left\{h^*(X_t)(Y_t-Y_s)-\frac{1}{2}|h(X_s)|^2(t-s)\right\}.$$

To prove Theorem 2.2.2, we use the similar idea to the proof of Theorem 4.3.1. By Remark 2.3, we only need to estimate $d(\pi^n_{j\delta,(j+1)\delta},\pi^n_{(j+1)\delta})$. The proof follows the same procedure. But to estimate J_2, we have to do more work. So we only give the proof for this part.

Proof. By Lemma 3.1 in [22], J_2 can be divided into three parts.

$$\begin{aligned} J_2 &= \frac{1}{m_j^n}\sum_{i=1}^{m_j^n}\int_{j\delta}^{(j+1)\delta} M_j^n(X^i,s)\nabla^*\psi_s\sigma(X_s^i)dB_s^i \\ &+\frac{1}{m_j^n}\sum_{i=1}^{m_j^n}\int_{j\delta}^{(j+1)\delta} M_j^n(X^i,s)c(X_s^i)\left(h(X_s^i)-h(X_{j\delta}^i)\right)^* ds \\ &+\frac{1}{m_j^n}\sum_{i=1}^{m_j^n}\int_{j\delta}^{(j+1)\delta} M_j^n(X^i,s)\Theta_s^*(X^i,s)\left(h(X_s^i)-h(X_{j\delta}^i)\right) dY_s \\ &\equiv J_{21}+J_{22}+J_{23} \end{aligned} \tag{2.44}$$

Note that

$$\begin{aligned} & \hat{\mathbb{E}}\left(\frac{1}{m_j^n}\sum_{i=1}^{m_j^n}\int_{j\delta}^{(j+1)\delta} M_j^n(X^i,s)\nabla^*\psi_s\sigma(X_s^i)dB_s^i\right) \\ = & \hat{\mathbb{E}}\left(\frac{1}{m_j^n}\sum_{i=1}^{m_j^n}\hat{\mathbb{E}}(\int_{j\delta}^{(j+1)\delta} M_j^n(X^i,s)\nabla^*\psi_s\sigma(X_s^i)dB_s^i|\mathcal{F}_j\delta)\right) \\ = & 0 \end{aligned}$$

Thus

$$\begin{aligned} \hat{\mathbb{E}}(J_{21})^2 &= \hat{\mathbb{E}}\left(\hat{\mathbb{E}}((\frac{1}{m_j^n}\sum_{i=1}^{m_j^n}\int_{j\delta}^{(j+1)\delta} M_j^n(X^i,s)\nabla^*\psi_s\sigma(X_s^i)dB_s^i)^2|\mathcal{F}_{j\delta}\vee\mathcal{G}_{(j+1)\delta}))\right) \\ &= \hat{\mathbb{E}}\frac{1}{(m_j^n)^2}\sum_{i=1}^{m_j^n}\int_{j\delta}^{(j+1)\delta} M_j^n(X^i,s)^2|\nabla^*\psi_s\sigma(X_s^i)|^2 ds \\ &\leq e^{K^2\delta}\hat{\mathbb{E}}\left(\frac{\delta}{m_j^n}\|\nabla^*\psi_s\|^2\|\sigma\|_{0,\infty}^2\right) \\ &\leq K_6\delta \end{aligned} \tag{2.45}$$

Note that h is bounded, we get

$$
\begin{aligned}
& \hat{\mathbb{E}}(J_{22})^2 \\
= \; & \hat{\mathbb{E}}\left(\frac{1}{m_j^n}\sum_{i=1}^{m_j^n}\int_{j\delta}^{(j+1)\delta} M_j^n(X^i,s)c(X_s^i)\left(h(X_s^i)-h(X_{j\delta}^i)\right)^* ds\right)^2 \\
= \; & \hat{\mathbb{E}}\left(\left(\frac{1}{(m_j^n)^2}\hat{\mathbb{E}}\left(\sum_{i=1}^{m_j^n}\int_{j\delta}^{(j+1)\delta} M_j^n(X^i,s)c(X_s^i)\left(h(X_s^i)-h(X_{j\delta}^i)\right)^* ds\right)^2 \Big| \mathcal{F}_{j\delta}\right)\right) \\
\leq \; & \hat{\mathbb{E}}\left(\frac{1}{(m_j^n)^2} m_j^n \sum_{i=1}^{m_j^n}\int_{j\delta}^{(j+1)\delta} M_j^n(X^i,s)^2 c(X_s^i)^2\left(h^*(X_s^i)-h^*(X_{j\delta}^i)\right)^2 ds\right) \\
\leq \; & K_7\delta^2
\end{aligned}
\tag{2.46}
$$

Since $\Theta_s(X^i, s)$ is bounded, by the similar argument to J_{22}, we have

$$
\hat{\mathbb{E}}(J_{23})^2 \leq K_8\delta^2 \tag{2.47}
$$

Combine (2.44), (2.45), (2.46) and (2.47), we get

$$
\hat{\mathbb{E}}|J_2|^2 \leq K\delta \tag{2.48}
$$

□

Chapter 3

Stability of filtering with Poisson observation

In this section, we study the stability of the filtering which is mentioned in section 1.3 and the exponential stability is derived.

3.1 Filtering model

Let's recall the intrinsic interest rate process $X(t)$ and observation process $\vec{Y}(t)$ in the model. $X(t)$ follows a diffusion process:

$$dX(t) = \mu(X(t))dt + \sigma(X(t))dB(t),$$

and $\vec{Y}(t)$ is defined as following:

$$\vec{Y}(t) = \begin{pmatrix} N_1(\int_0^t \lambda_1(X(s), s)ds) \\ N_2(\int_0^t \lambda_2(X(s), s)ds) \\ . \\ . \\ . \\ N_w(\int_0^t \lambda_w(X(s), s)ds) \end{pmatrix}. \tag{3.1}$$

The following four mild assumptions are invoked.

Assumption 3.1.1. *N_k's are unit Poisson processes under the physical measure $\mathbb{P}$.*

Assumption 3.1.2. *$X, N_1, N_2, ..., N_w$ are independent under P.*

Assumption 3.1.3. *The intensity at price level k, $\lambda_k(x,t) = a(x,t)p(y_k|x)$, where $a(x,t)$ is the total trading intensity at time t with $x = X(t)$ and $p(y_k|x)$ is the transition probability from x to y_k, the kth price level. Moreover, λ_k's are increasing in x and there exist two constants C_1 and C_2 such that $C_1 \le \lambda_k \le C_2, k = 1, 2, ..., w$.*

Remark 3.1. *Under this setup, $X(t)$ becomes the signal process, which cannot be observed directly, and $\vec{Y}(t)$ becomes the observation process. Hence $(X, \vec{Y})$ is framed as filtering problem with counting process observations.*

We assume that $(X, \vec{Y})$ is in a filtered complete probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ where $\mathbb{F} := (\mathcal{F}_t)_{0\le t\le\infty}$ is a given filtration. There is a reference measure $\hat{\mathbb{P}}$ under which, X and $\vec{Y}$ become independent, the probability distribution of X remains the same and $Y_1, Y_2, ..., Y_n$ become unit Poisson processes. The Radon-Nikodym derivative is:

$$M(t) = \frac{d\mathbb{P}}{d\hat{\mathbb{P}}}\Big|_{\mathcal{F}_t} = \prod_{k=1}^{w} \exp\left\{\int_0^t \log \lambda_k(X(s))dY_k(s) - \int_0^t [\lambda_k(X(s)) - 1]ds\right\} \tag{3.2}$$

Note that $M(t)$ satisfies the following SDE:

$$dM(t) = \sum_{k=1}^{w} (\lambda_k - 1) M(t-) d(Y_k(t) - t). \tag{3.3}$$

Let $\mathcal{G}_t = \sigma\{\vec{Y}(s) | 0 \leq s \leq t\}$ be all the available information up to time t and let π_t be the conditional distribution of $X(t)$ given $\mathcal{G}_t$. Define

$$\langle V_t, f \rangle = \hat{\mathbb{E}}[f(X(t))M(t)|\mathcal{G}_t] \quad \text{and} \quad \langle \pi_t, f \rangle = \mathbb{E}[f(X(t))|\mathcal{G}].$$

By Kallianpur-Striebel formula, the optimal filter in the sense of least mean square error can be written as $\langle \pi_t, f \rangle = \langle V_t, f \rangle / \langle V_t, 1 \rangle$.

The following proposition is a theorem from [35] summarizing both filtering equations.

Proposition 3.2. *Suppose that $(\theta, X, \vec{Y})$ satisfies assumptions 3.1.1, 3.1.2 and 3.1.3. Then, V_t is the unique measure-valued solution of the following SDE, the unnormalized filtering equation,*

$$\langle V_t, f \rangle = \langle V_0, f \rangle + \int_0^s \langle V_t, Lf \rangle + \sum_{k=1}^{w} \int_0^t \langle V_{s-}, (ap_k - 1) f \rangle \, d(Y_k(s) - s), \tag{3.4}$$

for $t > 0$ and $f \in D(L)$, the domain of generator L, where $a = a(X(t), t)$, is the trading intensity, and $p_k = p(y_k|x)$ is the transition probability from x to y_k, the kth price level.

π_t is the unique measure valued solution of the SDE, the normalized filtering equation,

$$\begin{aligned} \langle \pi_t, f \rangle \quad &= \quad \langle \pi_0, f \rangle + \int_0^t [\langle \pi_s, Lf \rangle - \langle \pi_s, fa \rangle + \langle \pi_s, f \rangle \langle \pi_s, a \rangle] ds \\ &\quad + \sum_{k=1}^{w} \int_0^t \left[\frac{\langle \pi_{s-}, fap_k \rangle}{\langle \pi_{s-}, ap_k \rangle} - \langle \pi_{s-}, f \rangle \right] dY_k(s) \end{aligned} \tag{3.5}$$

When $a(X(t),t) = a(t)$, the above equation is simplified as:

$$\langle \pi_t, f\rangle = \langle \pi_0, f\rangle + \int_0^t \langle \pi_s, Lf\rangle\, ds + \sum_{k=1}^{w} \int_0^t \left[\frac{\langle \pi_{s-}, f a p_k\rangle}{\langle \pi_{s-}, a p_k\rangle} - \langle \pi_{s-}, f\rangle \right] dY_k(s) \quad (3.6)$$

3.2 Examples

In this section, we make an assumption for the signal process $X(t)$. We give four examples which satisfy this assumption.

Assumption 3.2.1. *For any a, b and $c > 0$ and a large positive number x, $X(t)$ is an asymptotically stationary process with the following property:*

$$\mathbb{P}\left(\inf_{a \le s \le b} X(s) \le c \middle| X(b) = x \right) \le K(c)(b-a)^{\beta} g(x), \quad (3.7)$$

where $K(c)$ is a positive constant depends on c and $\beta > 0$. $g(\cdot)$ is a real function with $\lim_{x\to\infty} \frac{g(x)}{x} = 0$.

The following examples are the diffusion processes which satisfy the above assumption.

Example 1.

In finance, the Vasicek model (see [32]) is a mathematical model describing the evolution of interest rates. Vasicek's model was the first one to capture mean reversion, an essential characteristic of the interest rate that sets it apart from other financial prices. Thus, as opposed to stock prices for instance, interest rates cannot rise indefinitely. This is because at very high levels they would hamper economic activity, prompting a decrease in interest rates. Similarly, interest rates can not decrease below 0. As a result, interest rates move in a limited range, showing a tendency to revert to a long run value. The model specifies that the instantaneous interest rate follows the stochastic differential equation:

$$dX(t) = \theta(\mu - X(t))dt + \sigma dB(t), \quad (3.8)$$

where $B(t)$ is a Brownian motion and θ, μ and σ are parameters. The typical parameters θ, μ, and σ can be quickly characterized as follows:

μ: long term mean level. All future trajectories of X will evolve around a mean level μ in the long run;

θ: speed of reversion. θ characterizes the velocity at which such trajectories will regroup around μ in time;

σ: instantaneous volatility, measures instant by instant the amplitude of randomness entering the system. Higher σ implies more randomness.

The interest rate process is an Ornstein-Uhlenbeck process. Applying It$\hat{o}$ formula to function $f(X_t, t) = X_t e^{\theta t}$, we get

$$df(X(t), t) = e^{\theta t}\theta\mu dt + \sigma e^{\theta t} dB(t).$$

Let x_0 be the initial value of the process. Integrating the above equation from s to t, we have

$$X(t)e^{\theta t} = X(s)e^{\theta s} + \mu(e^{\theta t} - e^{\theta s}) + \int_s^t \sigma e^{\theta u} dB(u).$$

Then we have the following estimate:

$$\mathbb{P}\left(\inf_{a\leq s\leq b} X(s) \leq c \middle| X(b) = x\right) \leq K(c)(b-a)^{1/2} e^{-K'x^2}, \tag{3.9}$$

where K' is a constant.

The proof of above property is given in Appendix. Therefore, the process satisfies the assumption 3.2.1.

Example 2.

In mathematical finance, the CoxIngersollRoss model or CIR model (see [4]) describes the evolution of interest rates. It is a type of "one factor model" (short rate model) as it describes interest rate movements as driven by only one source of market risk. The model can be used in the valuation of interest rate derivatives. It was introduced in 1985 by John C. Cox, Jonathan E. Ingersoll and Stephen A. Ross as an extension of the Vasicek model. The CIR model specifies that the instantaneous interest rate follows the stochastic differential equation, also named the CIR process:

$$dX(t) = \theta(\mu - X(t))dt + \sigma\sqrt{X(t)}dB(t), \tag{3.10}$$

where $B(t)$ is a standard Brownian motion and θ, μ and σ are the parameters. The parameter θ corresponds to the speed of adjustment, μ to the mean and σ to volatility.

The drift factor, $\theta(\mu - X(t))$,is exactly the same as in the Vasicek model. It ensures mean reversion of the interest rate towards the long run value μ, with speed of adjustment governed by the strictly positive parameter θ. The standard deviation factor, $\sigma\sqrt{X(t)}$, avoids the possibility of negative interest rates for all positive values of θ and μ. An interest rate of zero is also precluded if the condition $2\theta\mu \geq \sigma^2$ is met. More generally, when the rate is at a low level (close to zero), the standard deviation also becomes close to zero, which dampens the effect of the random shock on the rate. Consequently, when the rate gets close to zero, its evolution becomes dominated by the drift factor, which pushes the rate upwards (towards equilibrium).

$X(t)$ is an ergodic process, possesses a stationary distribution, which is a gamma. The process satisfies the assumption 3.2.1 and the proof is given in Appendix. This process is widely used in finance to model short term interest rate.

Example 3.

The Rendleman-Bartter model (see [31]) in finance is a short rate model describing the evolution of interest rates. It is a “one factor model” as it describes interest rate movements as driven by only one source of market risk. It can be used in the valuation of interest rate derivatives. The model specifies that the instantaneous interest rate follows a geometric Brownian motion:

$$dX(t) = \theta X(t)dt + \sigma X(t)dB(t), \tag{3.11}$$

where $B(t)$ is a Wiener process modeling the random market risk factor. The drift parameter, θ, represents a constant expected instantaneous rate of change in the interest rate, while the standard deviation parameter, σ, determines the volatility of the interest rate.

For any initial value x_0, we apply itô formula to function $f(X(t)) = \log X(t)$

$$d \log X(t) = \left(\theta - \frac{\sigma^2}{2}\right) dt - \sigma dB(t).$$

Integrating the above equation from s to t, we get:

$$X(t) = X(s) \exp\left\{\left(\theta - \frac{\sigma^2}{2}\right) t - \sigma B(t)\right\}. \tag{3.12}$$

This is one of the early models of the short term interest rates, using the same stochastic process as the one already used to describe the dynamics of the underlying price in stock options. Its main disadvantage is that it does not capture the mean reversion of interest rates (their tendency to revert toward some value or range of values rather than wander without bounds in either direction). When $\theta < \frac{\sigma^2}{2}$, $X(t)$ is asymptotically stationary and approaches 0 as t tends to ∞. This process satisfies the assumption 3.2.1. The proof is given in Appendix. On the other hand, when $\theta > \frac{\sigma^2}{2}$, $X(t)$ blows up as t tends to ∞. But the inequality in the assumption 3.2.1 still holds.

Example 4.

In this example, we consider more general diffusion processes which are defined as following:

$$dX(t) = \mu(X(t))dt + \sigma(X(t))dB(t), \tag{3.13}$$

where $\mu(\cdot)$ and $\sigma(\cdot)$ are bounded.

It can be shown that the process defined as above satisfies the assumption 3.2.1. We give the proof in the Appendix.

3.3 Exponential stability

In this section, we state the stability theorem and give its proof.

Theorem 3.3.1. *The optimal filter is stable in the following sense: for any* $0 < \alpha < 1$*, we have*

$$\limsup_{t\to\infty} \frac{1}{t^\alpha} \log d_{TV}(\pi_t, \bar{\pi}_t) < 0, \tag{3.14}$$

where π_t *and* $\bar{\pi}_t$ *are optimal filters with initials* π_0 *and* $\bar{\pi}_0$ *respectively and* d_{TV} *is the total variation distance.*

The idea of the proof is to use the truncated filter to approximate the optimal filter. The interest rate process is truncated by a finite interval with length 2Δ. The stability of the truncated filter is derived by using Hilbert metric on the compact space. Then we replace Δ by Δ_t which depends on time t. Choosing a proper Δ_t, the truncated filter could be a good approximation for optimal filter. The truncated filter is defined as following:

Let $\Delta > 0$ be given. Define $B(\Delta) = \{x \in \mathbb{R} : |x| \leq \Delta\}$. For $\delta > 0$, let $n = [\frac{t}{\delta}]$. We divide the time line into small subintervals with the length δ. We define $I_{(i-1)\delta,i\delta}(x, x'; \vec{Y})$ on the subinterval $((i-1)\delta, i\delta)$ as following:

$$I_{(i-1)\delta,i\delta}(x, x'; \vec{Y}) := \hat{\mathbb{E}} \left(M_{(i-1)\delta,i\delta} \middle| \mathcal{G}_{(i-1)\delta,i\delta}, X((i-1)\delta) = x, X(i\delta) = x' \right),$$

where

$$M_{(i-1)\delta,i\delta} = \prod_{k=1}^{w} \exp\left\{ \int_{(i-1)\delta}^{i\delta} \log \lambda_k(X(s))dY_k(s) - \int_{(i-1)\delta}^{i\delta} [\lambda_k(X(s)) - 1]ds \right\},$$

and

$$\mathcal{G}_{(i-1)\delta,t} = \sigma\left(\vec{Y}_t - \vec{Y}_{(i-1)\delta} : (i-1)\delta < t \leq i\delta\right).$$

For $\mu \in \mathcal{P}(\mathbb{R})$, we define Q_i and $Q_i^\Delta,\ i = 1, ..., n$ on $\mathcal{P}(\mathbb{R})$:

$$\langle Q_i(\mu), f\rangle = \frac{\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} f(x')p_\delta(x, dx')I_{(i-1)\delta,i\delta}(x, x'; \vec{Y})\mu(dx)}{\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} f(x')p_\delta(x, dx')I_{(i-1)\delta,i\delta}(x, x'; \vec{Y})\mu(dx)}$$

and

$$\left\langle Q_i^\Delta(\mu), f\right\rangle = \frac{\int_{\mathbb{R}}\int_{\mathbb{R}} f(x')p_\delta(x, dx')I^\Delta_{(i-1)\delta,i\delta}(x, x'; \vec{Y})\mu(dx)}{\int_{\mathbb{R}}\int_{\mathbb{R}} p_\delta(x, dx')I^\Delta_{(i-1)\delta,i\delta}(x, x'; \vec{Y})\mu(dx)}, \tag{3.15}$$

where $I^\Delta_{(i-1)\delta,i\delta}(x, x'; \vec{Y}) = 1_{\{x' \in B(\Delta)\}} I_{(i-1)\delta,i\delta}(x, x'; \vec{Y})$. It's easy to prove that $\pi_{i\delta} = Q_i(\pi_{(i-1)\delta})$ and by induction, we define $\pi^\Delta_{i\delta} = Q_i(\pi^\Delta_{(i-1)\delta})$. Let $\left\langle V_t^\Delta, f\right\rangle$ be the unnormalized truncated filter.

We start with two truncated filters with different initial distributions.

Theorem 3.3.2. *For $\vec{Y} \in C(\mathbb{R}_+, \mathbb{R}^w)$ fixed and $\alpha \in (0, 1)$, we have*

$$\limsup_{t\to\infty} \frac{1}{t^\alpha} \log d_{TV}(\pi_t^{\Delta_t}, \bar{\pi}_t^{\Delta_t}) < 0, \tag{3.16}$$

where $\Delta_t = K\sqrt{\log t}$ and K is a constant.

To prove the theorem, we adapt the approach in [2] for the compact space case to the current step. First, we introduce the Hilbert metric ρ_h on $\mathcal{M}_F(\mathbb{R})$, where $\mathcal{M}_F(\mathbb{R})$ is the collection of all finite Borel measures on $\mathbb{R}$. We need the following definition:

Definition 3.3.1. *i) For λ, $\mu \in \mathcal{M}_F(\mathbb{R})$, $\lambda \leq \mu$ means that $\lambda(A) \leq \mu(A)$ for all $A \in \mathcal{B}(\mathbb{R})$.*

ii) Two measures λ, $\mu \in \mathcal{M}_F(\mathbb{R})$ are comparable if there are two positive constants K_1 and K_2 such that

$$K_1\lambda \leq \mu \leq K_2\lambda.$$

Then,for λ, $\mu \in \mathcal{M}_F(\mathbb{R})$, ρ_h is defined as

$$\rho_h(\lambda,\mu) = \begin{cases} \log \frac{\sup\{\lambda(A)/\mu(A): A\in\mathcal{B}(S), \mu(A)>0\}}{\inf\{\lambda(A)/\mu(A): A\in\mathcal{B}(S), \mu(A)>0\}} & \text{if } \lambda,\ \mu \text{ are comparable,} \\ 0 & \text{if } \lambda = \mu = 0, \\ \infty & \text{otherwise.} \end{cases}$$

Note that for any positive constants K_1 and K_2, we have

$$\rho_h(K_1\lambda, K_2\mu) = \rho_h(\lambda,\mu).$$

Thus,

$$\rho_h(\pi_t, \bar{\pi}_t) = \rho_h(V_t, \bar{V}_t), \quad \forall t > 0.$$

The following lemma will be useful in the proof of the stability of truncated filter (see lemma 10.31 on page 214 of [33]).

Lemma 3.3.1. *Let $\mathcal{T}$ be a linear transformation on $\mathcal{M}_F(\mathbb{R})$ which has the kernal representation*

$$\langle \mathcal{T}\mu, f\rangle = \int_{\mathbb{R}}\int_{\mathbb{R}} G(x,x')f(x')\mu(dx)\lambda(dx'),$$

where G(x,x') is non-negative. Then $\mathcal{T}$ is a contraction under the Hilbert metric and

$$\sup\left\{\frac{\rho_h(\mathcal{T}\lambda, \mathcal{T}\mu)}{\rho_h(\lambda,\mu)} : 0 < \rho_h(\lambda,\mu) < \infty\right\} = \tanh\frac{H(\mathcal{T})}{4},$$

where

$$H(\mathcal{T}) = \log esssup \frac{G(x,y)G(x',y')}{G(x,y')G(x'y)},$$

with the convention $0/0 = 1$ *and* $1/0 = \infty$. *The supremum above is strict over* $x, x' \in \mathbb{R}$, *and is essential over* $y, y' \in \mathbb{R}$ *with respect to* λ.

The next lemma shows the relationship between total variation distance and Hilbert metric.

Lemma 3.3.2. *For any* λ, $\mu \in \mathcal{P}$, *we have*

$$\begin{aligned} d_{TV}(\lambda,\mu) &\equiv 2 \sup_{A \in \mathcal{B}(\mathbb{R})} |\lambda(A) - \mu(A)| \\ &\leq \frac{2}{\log 3} \rho_h(\lambda,\mu). \end{aligned} \tag{3.17}$$

Proof. If λ and ν are not comparable, then $\rho_h(\lambda,\nu) = \infty$, and hence, equation (3.17) clearly holds. Now, we suppose λ and ν are comparable.

let $\mathcal{A} \equiv \{A \in \mathcal{B}(\mathbb{R} : \lambda(A) > \mu(A)\}$. Then, for $A \in \mathcal{A}$ with $\mu(A) > 0$, we have

$$\begin{aligned} 0 \leq \frac{\lambda(A)}{\mu(A)} - 1 &= \frac{\lambda(A)/\mu(A)}{\lambda(\mathbb{R})/\mu(\mathbb{R})} - 1 \\ &\leq \frac{\sup\{\lambda(A)/\mu(A) : A \in \mathcal{B}(S), \mu(A) > 0\}}{\inf\{\lambda(A)/\mu(A) : A \in \mathcal{B}(S), \mu(A) > 0\}} - 1 \\ &= e^{\rho_h(\lambda,\mu)} - 1. \end{aligned}$$

Hence

$$0 \leq \lambda(A) - \mu(A) \leq \mu(e^{\rho_h(\lambda,\mu)} - 1). \tag{3.18}$$

It's clear that (3.18) holds even if $\mu(A) = 0$ since $\lambda(A) = 0$ by the comparability.

By symmetry, for $A \in \mathcal{A}$, we have

$$0 \leq \lambda(A^c) - \mu(A^c) \leq \lambda(A^c)(e^{\rho_h(\lambda,\mu)} - 1).$$

Therefore,

$$\begin{aligned} d_{TV}(\lambda,\mu) &= \sup_{A\in\mathcal{A}} (\lambda(A)-\mu(A)+\mu(A^c)-\lambda(A^c)) \\ &\leq \sup_{A\in\mathcal{A}} (\mu(A)+\lambda(A^c)) \left(e^{\rho_h(\lambda,\mu)}-1\right) \\ &\leq \sup_{A\in\mathcal{A}} (\lambda(A)+\lambda(A^c)) \left(e^{\rho_h(\lambda,\mu)}-1\right) \\ &= e^{\rho_h(\lambda,\mu)}-1. \end{aligned}$$

Note that $d_{TV}(\lambda,\mu)$ is bounded by 2. By the following inequality

$$2 \wedge (e^x - 1) \leq \frac{2x}{\log 3}, \quad \forall x \geq 0$$

we have

$$d_{TV}(\lambda,\mu) \leq 2 \wedge (e^{\rho_h(\lambda,\mu)} - 1) \leq \frac{2\rho_h(\lambda,\mu)}{\log 3}.$$

□

Now we are ready for the proof of the Theorem 3.3.2.

Proof. It is well known that the transition probability for the diffusion process $X(t)$ exists and satisfies

$$K_1 e^{-K_2\Delta_t^2 t^{-1}} t^{-\frac{d}{2}} \leq p_t(x,x') \leq K_3 t^{-\frac{d}{2}}, \tag{3.19}$$

where $x, x' \in B(\Delta_t)$.

Let $p^{\Delta_t}_{(i-1)\delta,i\delta}(x,x') = p_{(i-1)\delta,i\delta}(x,x')\mathbf{1}\{x' \in B(\Delta_t)\}$. By lemma 3.3.1 and lemma 3.3.2, we have

$$\begin{aligned}
& \frac{1}{t^\alpha}\log d_{TV}(\pi_t,\bar{\pi}_t) - \frac{1}{t^\alpha}\log\frac{2}{\log 3} \\
\leq\ & \frac{1}{t^\alpha}\log\rho_h(V^{\Delta_t}_{n\delta},\bar{V}^{\Delta_t}_{n\delta}) \\
=\ & \frac{1}{t^\alpha}\sum_{i=1}^{n}\log\frac{\rho_h(V^{\Delta_t}_{i\delta},\bar{V}^{\Delta_t}_{i\delta})}{\rho_h(V^{\Delta_t}_{(i-1)\delta},\bar{V}^{\Delta_t}_{(i-1)\delta})} + \frac{1}{t^\alpha}\log\rho_h(\pi_0,\bar{\pi}_0) \\
\leq\ & \delta^{-1}t^{1-\alpha}\frac{1}{n}\sum_{i=1}^{n}\log\tanh(4^{-1}H^{\Delta_t}_{(i-1)\delta,i\delta}) + \frac{1}{t^\alpha}\log\rho_h(\pi_0,\bar{\pi}_0), \qquad (3.20)
\end{aligned}$$

where

$$\begin{aligned}
H^{\Delta_t}_{(i-1)\delta,i\delta} &= \log esssup\frac{p^{\Delta_t}_{(i-1)\delta,i\delta}(x,z)p^{\Delta_t}_{(i-1)\delta,i\delta}(x',z')}{p^{\Delta_t}_{(i-1)\delta,i\delta}(x,z')p^{\Delta_t}_{(i-1)\delta,i\delta}(x',z)} \\
&\quad + \log esssup\frac{I^{\Delta_t}_{(i-1)\delta,i\delta}(x,z;\vec{Y})I^{\Delta_t}_{(i-1)\delta,i\delta}(x',z';\vec{Y})}{I^{\Delta_t}_{(i-1)\delta,i\delta}(x,z';\vec{Y})I^{\Delta_t}_{(i-1)\delta,i\delta}(x',z;\vec{Y})} \\
&= \log esssup\frac{p^{\Delta_t}_{(i-1)\delta,i\delta}(x,z)p^{\Delta_t}_{(i-1)\delta,i\delta}(x',z')}{p^{\Delta_t}_{(i-1)\delta,i\delta}(x,z')p^{\Delta_t}_{(i-1)\delta,i\delta}(x',z)} \\
&\quad + \log esssup\frac{I_{(i-1)\delta,i\delta}(x,z;\vec{Y})I_{(i-1)\delta,i\delta}(x',z';\vec{Y})}{I_{(i-1)\delta,i\delta}(x,z';\vec{Y})I_{(i-1)\delta,i\delta}(x',z;\vec{Y})}.
\end{aligned}$$

Let $\mathcal{F}^{X,\vec{Y}}_{i-1,i} = \mathcal{G}_\delta \vee \{\omega\in\Omega : X(i\delta)=x', X((i-1)\delta)=x\}$.

$$\begin{aligned}
& I_{(i-1)\delta,i\delta}(x,x';\vec{Y}) \\
=\ & \hat{\mathbb{E}}\left(M_{(i-1)\delta,i\delta}|\mathcal{F}^{X,\vec{Y}}_{i-1,i}\right) \\
\leq\ & \hat{\mathbb{E}}\left(\prod_{k=1}^{w}\exp(\log C_2\left(Y_k(i\delta)-Y_k((i-1)\delta)\right)-(C_1-1)\delta)|\mathcal{F}^{X,\vec{Y}}_{i-1,i}\right) \\
=\ & \exp\left(\log C_2\sum_{k=1}^{w}\left(Y_k(i\delta)-Y_k((i-1)\delta)\right)-w(C_1-1)\delta\right) \\
\leq\ & \exp\left(2w\log C_2\|\vec{Y}^{i\delta}_{(i-1)\delta}\|_\infty - w(C_1-1)\delta\right), \qquad (3.21)
\end{aligned}$$

where $\|\vec{Y}^{i\delta}_{(i-1)\delta}\|_\infty = \max\{\|[Y_k]^{i\delta}_{(i-1)\delta}\|_\infty : 1 \leq k \leq w\}$.

Let $K_4(\vec{Y}, \delta) = \max\{w \log C_1 \|\vec{Y}^{i\delta}_{(i-1)\delta}\|_\infty - w(C_2 - 1)\delta : 1 \leq i \leq n\}$, then

$$I_{(i-1)\delta, i\delta}(x, x'; \vec{Y}) \leq \exp(K_4(\vec{Y}, \delta)), \tag{3.22}$$

Similarly, there exists $K_5(\vec{Y}, \delta)$ such that

$$\exp(K_5(\vec{Y}, \delta)) \leq I_{(i-1)\delta, i\delta}(x, x'; \vec{Y}). \tag{3.23}$$

Therefore,

$$\log esssup \frac{I^{\Delta_t}_{(i-1)\delta, i\delta}(x, z; \vec{Y}) I^{\Delta_t}_{(i-1)\delta, i\delta}(x', z'; \vec{Y})}{I^{\Delta_t}_{(i-1)\delta, i\delta}(x, z'; \vec{Y}) I^{\Delta_t}_{(i-1)\delta, i\delta}(x', z; \vec{Y})} \leq K_4(\vec{Y}, \delta)^2 K_5(\vec{Y}, \delta)^2. \tag{3.24}$$

On the other hand, by (3.19), when $x, x', z, z' \in B(\Delta_t)$ we have

$$\begin{aligned} \log esssup \frac{p^{\Delta_t}_\delta(x, z) p^{\Delta_t}_\delta(x', z')}{p^{\Delta_t}_\delta(x, z') p^{\Delta_t}_\delta(x', z)} &\leq \log \frac{K_3 t^{-\frac{d}{2}} K_3 t^{-\frac{d}{2}}}{K_1 e^{-K_2 \Delta_t^2 t^{-1}} t^{-\frac{d}{2}} K_1 e^{-K_2 \Delta_t^2 t^{-1}} t^{-\frac{d}{2}}} \\ &\leq \log(K_1^{-2} K_3^2 \exp(2K_2 \Delta_t^2 \delta^{-1})). \end{aligned} \tag{3.25}$$

Let $K_6(\vec{Y}, \delta, \Delta_t) = \log(K_1^{-2} K_3^2 \exp(2K_2 \Delta_t^2 \delta^{-1})) + K_4(\vec{Y}, \delta)^2 K_5(\vec{Y}, \delta)^2$, by the definition of $H^{\Delta_t}_{(i-1)\delta, i\delta}$, we have

$$H^{\Delta_t}_{(i-1)\delta, i\delta} \leq K_6(\vec{Y}, \delta, \Delta_t).$$

Note that

$$\tanh x = 1 - \frac{2}{e^{2x} + 1},$$

then

$$
\begin{aligned}
& \tanh(4^{-1}H^{\Delta_t}_{(i-1)\delta,i\delta}) \\
\leq\ & 1-\frac{2}{\exp\{2\cdot 4^{-1}K_6(\vec{Y},\delta,\Delta_t)+1\}} \\
=\ & 1-\frac{2}{\exp\{2^{-1}\log(K_1^{-2}K_3^2\exp(2K_2\Delta_t^2\delta^{-1}))+2K_4(\vec{Y},\delta)^2K_5(\vec{Y},\delta)^2\}+1} \\
=\ & 1-\frac{2}{K_7\exp\{K_8\Delta_t^2\}+1}, \qquad (3.26)
\end{aligned}
$$

where $K_7=(K_1^{-2}K_3^2)^{2^{-1}}\exp\left\{2\left(K_4(\vec{Y},\delta)^2K_5(\vec{Y},\delta)^2\right)\right\}$ and $K_8=K_2\delta^{-1}$.

Combining (3.20) and (3.26), we have

$$
\begin{aligned}
& \frac{1}{t^\alpha}\log d_{TV}(\pi_t^{\Delta_t},\bar{\pi}_t^{\Delta_t})-\frac{1}{t^\alpha}\log\frac{2}{\log 3} \\
\leq\ & \delta^{-1}t^{1-\alpha}\frac{1}{n}\sum_{i=1}^n\log\tanh(4^{-1}H^{\Delta_t}_{(i-1)\delta,i\delta})+\frac{1}{t^\alpha}\log\rho_h(\pi_0,\bar{\pi}_0) \\
\leq\ & \delta^{-1}\log\left(1-\frac{2}{\exp(2^{-1}K_6(\vec{Y},\delta,\Delta_t))+1}\right)^{t^{1-\alpha}}+\frac{1}{t^\alpha}\log\rho_h(\pi_0,\bar{\pi}_0) \\
\leq\ & \delta^{-1}\log\left(1-\frac{2}{K_7\exp\{K_8\Delta_t^2\}+1}\right)^{t^{1-\alpha}}+\frac{1}{t^\alpha}\log\rho_h(\pi_0,\bar{\pi}_0) \qquad (3.27)
\end{aligned}
$$

Let $\Delta_t^2=K_9\log t$, where $K_9=\frac{1-\alpha}{K_8}$. Then

$$
\lim_{t\to\infty}\frac{t^{1-\alpha}}{K_7\exp\{K_8\Delta_t^2\}+1}=\frac{1}{K_7}.
$$

Take $t\to\infty$ on (3.27), we have

$$
\lim_{t\to\infty}\left(\frac{1}{t^\alpha}\log d_{TV}(\pi_t^{\Delta_t},\bar{\pi}_t^{\Delta_t})-\frac{1}{t^\alpha}\log\frac{2}{\log 3}\right)\leq -\delta^{-1}K_7^{-1}.
$$

This implies that

$$
\lim_{t\to\infty}\frac{1}{t^\alpha}\log d_{TV}(\pi_t^{\Delta_t},\bar{\pi}_t^{\Delta_t})<0
$$

□

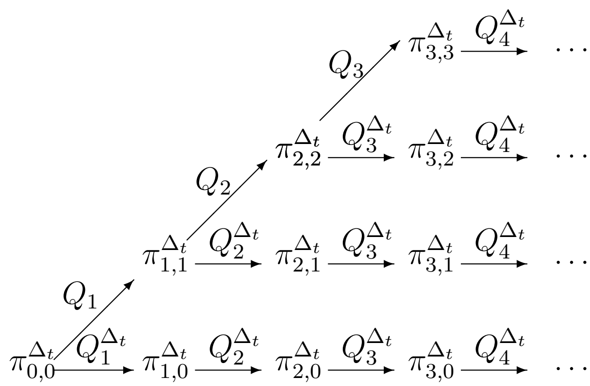

Figure 3.3 The exact filter process and approximations base on truncation.

Then we estimate the approximation of truncated filter for the optimal filter with same initial distribution. Denote by $\pi_{0,0}^{\Delta_t}$ the initial distribution π_0. For any $\mu \in \mathcal{M}_F(\mathbb{R})$, let $Q_{i,j} = Q_j\left(Q_{j-1}\left(\cdots\left(Q_{i+1}(\mu)\right)\right)\right)$ and $Q_{j,i}^{\Delta_t}(\mu) = Q_j^{\Delta_t}\left(Q_{j-1}^{\Delta_t}\left(\cdots\left(Q_{i+1}^{\Delta_t}(\mu)\right)\right)\right)$. Then, we define

$$\begin{aligned} \pi_{j,j}^{\Delta_t} &\equiv Q_{i,j}(\pi_{i,i}^{\Delta_t}) \\ &= Q_j\left(Q_{j-1}\left(\cdots\left(Q_{i+1}(\pi_{i,i}^{\Delta_t})\right)\right)\right), \end{aligned}$$

and

$$\begin{aligned} \pi_{j,i}^{\Delta_t} &\equiv Q_{j,i}^{\Delta_t}(\pi_{i,i}^{\Delta_t}) \\ &= Q_j^{\Delta_t}\left(Q_{j-1}^{\Delta_t}\left(\cdots\left(Q_{i+1}^{\Delta_t}(\pi_{i,i}^{\Delta_t})\right)\right)\right), \end{aligned}$$

where $i < j \leq n$.

For any $t \in [n\delta, (n+1)\delta)$, we have

$$d_{TV}(\pi_t, \pi_t^{\Delta_t}) \leq \sum_{i=0}^{n} d_{TV}(\pi_{t,i-1}^{\Delta_t}, \pi_{t,i}^{\Delta_t}) \tag{3.28}$$

Note that

$$\pi_{t,i-1}^{\Delta_t} = Q_{i+1,t}^{\Delta_t}(\pi_{i,i-1}^{\Delta_t}),$$

and

$$\pi_{t,i-1}^{\Delta_t} = Q_{i+1,t}^{\Delta_t}(\pi_{i,i}^{\Delta_t}). \tag{3.29}$$

So $\pi_{t,i-1}^{\Delta_t}$ and $\pi_{t,i-1}^{\Delta_t}$ are two truncated filters with different initial distributions $\pi_{i,i-1}^{\Delta_t}$ and $\pi_{i,i}^{\Delta_t}$ respectively. Then by the proof of stability of truncated filter, we have

$$d_{TV}(\pi_{t,i-1}^{\Delta_t}, \pi_{t,i}^{\Delta_t}) \leq K_1 e^{-K_2(t-i\delta)^\alpha} d_{TV}(\pi_{i,i-1}^{\Delta_t}, \pi_{i,i}^{\Delta_t}).$$

Therefore, we need to estimate $d_{TV}(\pi_{i,i-1}^{\Delta_t}, \pi_{i,i}^{\Delta_t})$.

Theorem 3.3.3.

$$\lim_{t\to\infty} \frac{1}{t} \log \sup_i d_{TV}(\pi_{i,i-1}^{\Delta_t}, \pi_{i,i}^{\Delta_t}) < 0. \tag{3.30}$$

Proof. For any function f bounded by 1, we have

$$\begin{aligned}
&\left|\left\langle \pi_{i,i-1}^{\Delta_t}, f\right\rangle - \left\langle \pi_{i,i}^{\Delta_t}, f\right\rangle\right| \\
\leq\ & \frac{\left|\left\langle V_{i,i-1}^{\Delta_t}, f\right\rangle - \left\langle V_{i,i}^{\Delta_t}, f\right\rangle\right|}{\left\langle V_{i,i}^{\Delta_t}, 1\right\rangle} + \frac{\left|\left\langle V_{i,i-1}^{\Delta_t}, 1\right\rangle - \left\langle V_{i,i}^{\Delta_t}, 1\right\rangle\right|}{\left\langle V_{i,i}^{\Delta_t}, 1\right\rangle} \\
=\ & \frac{\left|\int_{\mathbb{R}}\int_{\mathbb{R}} f(x_i) I_{(i-1)\delta,i\delta}(x_{i-1}, x_i; \vec{Y}) 1_{x_i \notin B(\Delta_t)} p_\delta(x_{i-1}, dx_i)\pi_{i-1}(dx_{i-1})\right|}{\int_{\mathbb{R}}\int_{\mathbb{R}} I_{(i-1)\delta,i\delta}(x_{i-1}, x_i; \vec{Y}) p_\delta(x_{i-1}, dx_i)\pi_{i-1}(dx)} \\
& + \frac{\left|\int_{\mathbb{R}}\int_{\mathbb{R}} I_{(i-1)\delta,i\delta}(x_{i-1}, x_i; \vec{Y}) 1_{x_i \notin B(\Delta_t)} p_\delta(x_{i-1}, dx_i)\pi_{i-1}(dx_{i-1})\right|}{\int_{\mathbb{R}}\int_{\mathbb{R}} I_{(i-1)\delta,i\delta}(x_{i-1}, x_i; \vec{Y}) p_\delta(x_{i-1}, dx_i)\pi_{i-1}(dx)} \\
\leq\ & \frac{2\int_{\mathbb{R}}\int_{\mathbb{R}} I_{(i-1)\delta,i\delta}(x_{i-1}, x_i; \vec{Y}) 1_{x_i \notin B(\Delta_t)} p_\delta(x_{i-1}, dx_i)\pi_{i-1}(dx_{i-1})}{\int_{\mathbb{R}}\int_{\mathbb{R}} I_{(i-1)\delta,i\delta}(x_{i-1}, x_i; \vec{Y}) p_\delta(x_{i-1}, dx_i)\pi_{i-1}(dx)}
\end{aligned} \tag{3.31}$$

let $M^i(\delta,\vec{Y}) = \frac{1}{w}\sum_{k=1}^{w}|Y_k(i\delta) - Y_k((i-1)\delta) - \delta|$ and $g_k(x) := \lambda_k(x) - 1 - \log\lambda_k(x),\ k = 1, 2, ..., w$, we have

$$
\begin{aligned}
& I_{(i-1)\delta,i\delta}(x_{i-1}, x_i; \vec{Y}) \\
= \ & \hat{\mathbb{E}}\left(M_{(i-1)\delta,i\delta}\middle|\mathcal{F}^{X,\vec{Y}}_{i-1,i}\right) \\
= \ & \hat{\mathbb{E}}\Big(\prod_{k=1}^{w}\exp\left\{\int_{(i-1)\delta}^{i\delta}\log\lambda_k(X(s))d(Y_k(s) - s)\right\}\cdot \\
& \exp\left\{-\int_{(i-1)\delta}^{i\delta}(\lambda_k(X(s)) - 1 - \log\lambda_k(X(s)))ds\right\}\Big|\mathcal{F}^{X,\vec{Y}}_{i-1,i}\Big) \\
\leq \ & \exp\left\{\log C_2\sum_{k=1}^{w}|Y_k(i\delta) - Y_k((i-1)\delta) - \delta|\right\}\cdot \\
& \hat{\mathbb{E}}\left(\prod_{k=1}^{w}\exp\left\{-\int_{(i-1)\delta}^{i\delta}g_k(x)ds\right\}\Big|\mathcal{F}^{X,\vec{Y}}_{i-1,i}\right) \\
\equiv \ & \exp\{w\log C_2 M^i(\delta,\vec{Y})\}\cdot \\
& \hat{\mathbb{E}}\left(\prod_{k=1}^{w}\exp\left\{-\int_{(i-1)\delta}^{i\delta}g_k(x)ds\right\}\Big|\mathcal{F}^{X,\vec{Y}}_{i-1,i}\right).
\end{aligned}
\tag{3.32}
$$

Let $A_k = \{\omega \in \Omega | \inf_{(i-1)\delta\leq s\leq i\delta} X(s) \leq \lambda_k^{-1}(1 + e^{\Delta_i^2})\},\ k = 1, 2, ..., w$. It's easy to check that $g_k(x) \geq 0$ for all x and k and the equality can be reached only when $\lambda_k(x) = 1$. By the monotonicity assumption on λ_k's, we have

$$
\begin{aligned}
& \hat{\mathbb{E}}\left(\prod_{k=1}^{w}\exp\left\{-\int_{(i-1)\delta}^{i\delta}g_k(x)ds\right\}\Big|\mathcal{F}^{X,\vec{Y}}_{i-1,i}\right) \\
\leq \ & \hat{\mathbb{E}}\left(\exp\left\{-\sum_{k=1}^{w}\int_{(i-1)\delta}^{i\delta}g_k(X(s))ds\right\}\prod_{k=1}^{w}\mathbf{1}_{A_k^c}\Big|\mathcal{F}^{X,\vec{Y}}_{i-1,i}\right) \\
& +\hat{\mathbb{E}}\left(\exp\left\{-\sum_{k=1}^{w}\int_{(i-1)\delta}^{i\delta}g_k(X(s))ds\right\}\sum_{k=1}^{w}\mathbf{1}_{A_k}\Big|\mathcal{F}^{X,\vec{Y}}_{i-1,i}\right)
\end{aligned}
$$

$$
\begin{aligned}
&\leq \exp\{-\sum_{k=1}^{w}(1+e^{\Delta_t^2}-1-\log(1+e^{\Delta_t^2}))\delta\} \\
&\quad + \sum_{k=1}^{w}\hat{\mathbb{E}}\left(\mathbf{1}_{A_k}\middle| X(i\delta)=x_i, X((i-1)\delta)=x_{i-1}\right) \\
&= \exp\{-w\delta(e^{\Delta_t^2}-\log(1+e^{\Delta_t^2}))\} + \sum_{k=1}^{w}\hat{\mathbb{E}}\left(\mathbf{1}_{A_k}\middle| X(i\delta)=x_i, X((i-1)\delta)=x_{i-1}\right) \\
&\equiv U_1^i + U_2^i \qquad (3.33)
\end{aligned}
$$

Therefore, we have

$$
I_{(i-1)\delta,i\delta}(x_{i-1},x_i;\vec{Y}) \leq \exp\left\{w\log C_2 M^i(\delta,\vec{Y})\right\}\cdot(U_1^i+U_2^i)
$$

Let $\delta_n \downarrow (i-1)\delta$ and $F_t^k = \mathbf{1}\{\inf_{t\leq s\leq i\delta} X(s) \leq \lambda_k^{-1}(1+e^{\Delta_t^2})\}$, by Fatou's lemma, we get

$$
\begin{aligned}
U_2^i &= \hat{\mathbb{E}}\left(F_{(i-1)\delta}^k\middle| X(i\delta)=x_i, X((i-1)\delta)=x_{i-1}\right) \\
&\leq \lim_{n\to\infty}\hat{\mathbb{E}}\left(F_{\delta_n}^k\middle| X(i\delta)=x_i, X((i-1)\delta)=x_{i-1}\right) \qquad (3.34)
\end{aligned}
$$

For any $\mathcal{F}_{\delta_n^i,i\delta}^X \equiv \sigma\{X_s, \delta_n^i \leq s \leq i\delta\}$-measurable random variable Z and Borel measurable function ϕ, it follows from the Markov property of $X(t)$ that

$$
\begin{aligned}
&\int_{\mathbb{R}}\hat{\mathbb{E}}(Z|X((i-1)\delta)=x, X(i\delta)=x')p_\delta(x,x')\phi(x)dx \\
&= \hat{\mathbb{E}}(\hat{\mathbb{E}}(Z|X(i\delta),X(i\delta)=x')\phi(X_{i\delta})|X(i\delta)=x') \\
&= \hat{\mathbb{E}}(Z\phi(X_{i\delta})X(i\delta)=x') \\
&= \hat{\mathbb{E}}(Z\hat{\mathbb{E}}(\phi(X_{i\delta})|\mathcal{F}_{\delta_n}^X)|X(i\delta)=x') \\
&= \hat{\mathbb{E}}\left(Z\int_{\mathbb{R}}p_{\delta_n^i-(i-1)\delta}(x,X(\delta_n^i))\phi(x)dx|X(i\delta)=x'\right) \\
&= \int_{\mathbb{R}}\hat{\mathbb{E}}\left(Zp_{\delta_n^i-(i-1)\delta}(x,X(\delta_n^i))|X(i\delta)=x'\right)\phi(x)dx
\end{aligned}
$$

Thus,

$$\begin{aligned}
& \hat{\mathbb{E}}\left(F^k_{\delta_n}\Big|X(i\delta)=x_i, X((i-1)\delta)=x_{i-1}\right)p_\delta(x_{(i-1)},x_i)\\
= & \hat{\mathbb{E}}\left(F^k_{\delta_n}p_{\delta^i_n-(i-1)\delta}(x_{i-1},X(\delta^i_n))\Big|X(i\delta)=x_i\right).
\end{aligned}$$

Let $b_{n-1}(x_{i-1},x_i)=\hat{\mathbb{E}}\left(F^k_{\delta_{n-1}}p_{\delta^i_n-(i-1)\delta}(x_{i-1},X(\delta^i_n))\Big|X(i\delta)=x_i\right)$, we have

$$\begin{aligned}
& |b_n(x_{i-1},x_i)-b_{n-1}(x_{i-1},x_i)|\\
\le & \hat{\mathbb{E}}\left(|F^k_{\delta_n}-F^k_{\delta_{n-1}}|p_{\delta^i_n-(i-1)\delta}(x_{i-1},X(\delta^i_n))\Big|X(i\delta)=x_i\right)\\
\le & \left\{\hat{\mathbb{E}}\left(|F^k_{\delta_n}-F^k_{\delta_{n-1}}|^2\Big|X(i\delta)=x_i\right)\right\}^{1/2}\left\{\hat{\mathbb{E}}\left(p_{\delta^i_n-(i-1)\delta}(x_{i-1},X(\delta^i_n))^2\Big|X(i\delta)=x_i\right)\right\}^{1/2}\\
\equiv & R^i_1R^i_2. \qquad (3.35)
\end{aligned}$$

By (3.19), we have

$$\begin{aligned}
R^i_2 &= \left\{\int p_{\delta^i_n-(i-1)\delta}(x_{i-1},z)^2p_{\delta_n}(z,x_i)dz\right\}^{1/2}\\
&\le K_3(\delta^i_n-(i-1)\delta)^{-1/2}\left\{\int p_{\delta^i_n-(i-1)\delta}(x_{i-1},z)p_{\delta_n}(z,x_i)dz\right\}^{1/2}\\
&\le K_3(\delta^i_n-(i-1)\delta)^{-1/2}p_\delta(x_{i-1},x_i)^{1/2}. \qquad (3.36)
\end{aligned}$$

By the definition of $F^k_{\delta_n}$ and assumption 3.2.1, we have

$$\begin{aligned}
& R^i_1\mathbf{1}\{x_i>\Delta_t\}\\
= & \left\{\hat{\mathbb{E}}\left(\mathbf{1}\{\inf_{\delta^i_n\le s\le\delta^i_{n-1}}X(s)\le\lambda_k^{-1}(1+e^{\Delta^2_t})\}\Big|X(i\delta)=x_i\right)\right\}^{1/2}\mathbf{1}\{x_i>\Delta_t\}\\
\le & K_{10}(\delta^i_{n-1}-\delta^i_n)^\beta g(\Delta^2_t)\mathbf{1}\{x_i>\Delta_t\}, \qquad (3.37)
\end{aligned}$$

where the last inequality is from the assumption 3.2.1.

Choosing $\delta_n = (i-1)\delta + 2^{-n}\delta$ and combining estimates (3.38), (3.36) and (3.37), we get

$$\begin{aligned}
& |b_n(x_{i-1}, x_i) - b_{n-1}(x_{i-1}, x_i)|\mathbf{1}\{x_i > \Delta_t\} \\
\leq \; & K_{11}(\delta_n^i - (i-1)\delta)^{-1/2} p_\delta(x_{i-1}, x_i)^{1/2} (\delta_{n-1} - \delta_n)^\beta g(\Delta_t^2)\mathbf{1}\{x_i > \Delta_t\} \\
\leq \; & K_{12} p_\delta(x_{i-1}, x_i)^{1/2} 2^{-n\beta} g(\Delta_t^2)\mathbf{1}\{x_i > \Delta_t\}
\end{aligned}$$

Note that, when $x_i > \Delta$, we have $x_i > \lambda_k^{-1}(1+e^{\Delta_t^2})$ by the boundness of λ_k. Therefore,

$$\begin{aligned}
& b_0(x_{i-1}, x_i)\mathbf{1}\{x_i > \Delta_t\} \\
= \; & \hat{\mathbb{E}}\left(F_{i\delta}^k p_{i\delta-\delta_1}(X_{i\delta}, x_i)\Big| X(i\delta) = x_i\right)\mathbf{1}\{x_i > \Delta_t\} \\
= \; & \hat{\mathbb{E}}\left(1\{x_i \leq \lambda_k^{-1}(1+e^{\Delta_t^2})\} p_{i\delta-\delta_1}(X_{\delta_0}, x_i)\Big| X(i\delta) = x_i\right)\mathbf{1}\{x_i > \Delta_t\} \\
= \; & 0.
\end{aligned}$$

Therefore,

$$b_n(x_{i-1}, x_i)\mathbf{1}\{x_i > \Delta_t\} \leq K_{13} p_\delta(x_{i-1}, x_i)^{1/2} g(\Delta_t^2) \tag{3.38}$$

By (3.34) and (3.38), we have

$$U_2^i \mathbf{1}\{x_i > \Delta_t\} \;\; \leq \;\; K_{13} p_\delta(x_{i-1}, x_i)^{1/2} g(\Delta_t^2). \tag{3.39}$$

Combine (3.32), (3.33) and (3.39), we have

$$\begin{aligned}
& I_{(i-1)\delta, i\delta}(x_{i-1}, x_i; \vec{Y})\mathbf{1}\{x_i > \Delta_t\} p_\delta(x_{i-1}, x_i) \\
\leq \; & \left\{w \log C_2 M^i(\delta, \vec{Y})\right\} \cdot (U_1^i + U_2^i)\mathbf{1}\{x_i > \Delta\} p_\delta(x_{i-1}, x_i) \\
\leq \; & \left\{w \log C_2 M^i(\delta, \vec{Y})\right\} \Big(p_\delta(x_{i-1}, x_i) \exp\{-w\delta(e^{\Delta_t^2} - \log(1+e^{\Delta_t^2}))\} \\
& + K_{13} p_\delta(x_{i-1}, x_i)^{1/2} g(\Delta_t^2)\Big) \\
\leq \; & \left\{w \log C_2 M^i(\delta, \vec{Y})\right\} p_\delta(x_{i-1}, x_i)^{1/2} K_{14}(\Delta_t),
\end{aligned} \tag{3.40}$$

where $K_{14}(\Delta_t) = \exp\{-w\delta(e^{\Delta_t^2} - \log(1+e^{\Delta_t^2}))\} + K_{13}g(\Delta_t^2)$.

As $\Delta_t^2 = K_9 \log t$,

$$\lim_{t\to\infty} \frac{1}{t} \log \exp\{-w\delta(e^{\Delta_t^2} - \log(1+e^{\Delta_t^2})\} < 0,$$

and

$$\lim_{t\to\infty} \frac{1}{t} \log g(\Delta_t^2) < 0.$$

Therefore

$$\lim_{t\to\infty} \frac{1}{t} \log K_{14}(\Delta_t) < 0. \tag{3.41}$$

On the other hand,we have

$$\begin{aligned} & I_{(i-1)\delta,i\delta}(x_{i-1}, x_i; \vec{Y}) \\ = \;& \hat{\mathbb{E}}\left(\exp\left\{\sum_{k=1}^{w}\left(\int_{(i-1)\delta}^{i\delta} \log \lambda_k(X(s))dY_k(s) - \int_{(i-1)\delta}^{i\delta} (\lambda_k(X(s)) - 1)\, ds\right)\right\} \Big| \mathcal{F}_{i-1,i}^{X,\vec{Y}}\right) \\ \geq \;& \hat{\mathbb{E}}\left(\exp\left\{\sum_{k=1}^{w} (\log C_1(Y_k(i\delta) - Y_k((i-1)\delta) - (C_2 - 1)\delta)\right\} \Big| \mathcal{G}_{(i-1)\delta}^{i\delta}\right) \\ \geq \;& K(\vec{Y}_{(i-1)\delta}^{i\delta}), \end{aligned} \tag{3.42}$$

where $K(\vec{Y}_{(i-1)\delta}^{i\delta})$ depends on $Y_k(i\delta) - Y_k((i-1)\delta),\ 1 \leq k \leq w$.

Note that $\int_{\mathbb{R}}\int_{\mathbb{R}} p_\delta(x_{i-1}, x_i)^{1/2} dx_i \pi_{i-1}(dx_{i-1}) < \infty$. Combining (3.31), (3.40) and (4.31), we get

$$\begin{aligned} & \left|\left\langle \pi_{i,i-1}^{\Delta_t}, f\right\rangle - \left\langle \pi_{i,i}^{\Delta_t}, f\right\rangle\right| \\ \leq \;& \frac{2\int_{\mathbb{R}}\int_{\mathbb{R}} I_{(i-1)\delta,i\delta}(x_{i-1}, x_i; \vec{Y}) 1_{x_i \notin B(\Delta_t)} p_\delta(x_{i-1}, dx_i)\pi_{i-1}(dx_{i-1})}{\int_{\mathbb{R}}\int_{\mathbb{R}} I_{(i-1)\delta,i\delta}(x_{i-1}, x_i; \vec{Y}) p_\delta(x_{i-1}, dx_i)\pi_{i-1}(dx)} \\ \leq \;& \frac{2\left\{w \log C_2 M^i(\delta, \vec{Y})\right\} K_{14}(\Delta_t) \int_{\mathbb{R}}\int_{\mathbb{R}} p_\delta(x_{i-1}, x_i)^{1/2} dx_i \pi_{i-1}(dx_{i-1})}{K(\vec{Y}_{(i-1)\delta}^{i\delta})}. \end{aligned} \tag{3.43}$$

Thus, by (3.41), we can prove (3.30). □

Similarly, we can prove

$$\lim_{t\to\infty} \frac{1}{t} \log d_{TV}(\pi^{\Delta_t}_{t,t}, \pi^{\Delta_t} n\delta, t) < 0. \tag{3.44}$$

Combining (3.28), (3.29), (3.30) and (4.36), we have

$$\lim_{t\to\infty} \frac{1}{t} \log d_{TV}(\pi_t, \pi^{\Delta_t}_t) < 0. \tag{3.45}$$

Now, we are ready to prove the main theorem.

Proof. By triangle inequality, we have

$$d_{TV}(\pi_t, \bar{\pi}_t) \leq d_{TV}(\pi_t, \pi^{\Delta_t}_t) + d_{TV}(\pi^{\Delta_t}_t, \bar{\pi}^{\Delta_t}_t) + d_{TV}(\bar{\pi}_t, \bar{\pi}^{\Delta_t}_t) \tag{3.46}$$

Combining (3.46), (3.3.2) and (3.45), we have

$$\lim_{t\to\infty} \frac{1}{t} \log d_{TV}(\pi_t, \bar{\pi}_t) < 0. \tag{3.47}$$

□

Chapter 4

Numerical method for filtering model with Poisson observation

4.1 Branching particle system

In [34], a branching particle filter is studied as a approximation of the optimal filter mentioned in previous section. In this section, we introduce a branching particle system without using the integration in the weight of the particles.

We proceed to defining the branching particle system. Initially, there are n particles of weight 1 each at locations $x_i^n, i = 1, 2, ..., n$, satisfying the following condition:

Assumption 4.1.1. *The initial positions $\{x_i^n : i = 1, 2, ..., n\}$ of the particles are i.i.d. random vectors in $\mathbb{R}^d$ with the common distribution $\pi_0 \in \mathcal{P}(\mathbb{R}^d)$.*

Let $\delta = \delta_n = n^{-2\alpha}, 0 < \alpha < 1$. For $j = 0, 1, 2, ...$, there are m_j^n number of particles alive at time $t = j\delta$. Note that $m_0^n = n$.

During the time interval $(j\delta, (j+1)\delta)$, the particles move according to the following equations: For $i = 1, 2, ..., m_j^n$,

$$X^i(t) = X^i(j\delta) + \mu(X^i(j\delta))(t - j\delta) + \sigma(X^i(j\delta))(B^i(t) - B^i(j\delta))$$

where $\{B^i, i = 1, 2, ..., n\}$ are independent standard Brownian motions.

At the end of the interval, the ith particle ($i = 1, 2, ..., m_j^n$) branches (independent of others) into a random number ξ_{j+1}^i of offsprings such that the conditional expectation and the conditional variance given the information prior to the branching satisfy

$$\hat{\mathbb{E}}(\xi_{j+1}^i|\mathcal{F}_{(j+1)\delta-}) = \tilde{M}_{j+1}^n(X^i),$$

and

$$Var^{\hat{P}}(\xi_{j+1}^i|\mathcal{F}_{(j+1)\delta-}) = \gamma_{j+1}^n(X^i),$$

where $\gamma_{j+1}^n(X^i)$ is arbitrary,

$$\tilde{M}_{j+1}^n(X^i) = \frac{M_{j+1}^n(X^i)}{\frac{1}{m_j^n}\sum_{l=1}^{m_j^n} M_{j+1}^n(X^l)}$$

and

$$M_{j+1}^n(X^i) = \prod_{k=1}^{w} \exp\{\log ap_k(X^i(j\delta))(Y_k(j+1)\delta) - Y_k(j\delta)) - \left(ap_k(X^i(j\delta)) - 1\right)\delta\} \quad (4.1)$$

To minimize $\gamma_{j+1}^n(X^i)$, we take

$$\xi_{j+1}^i = \begin{cases} [\tilde{M}_{j+1}^n(X^i)] & \text{with probability } 1 - \{\tilde{M}_{j+1}^n(X^i)\}, \\ [\tilde{M}_{j+1}^n(X^i)] + 1 & \text{with probability } \{\tilde{M}_{j+1}^n(X^i)\} \end{cases}$$

where $\{x\} = x - [x]$ is the fraction of x, and $[x]$ is the largest integer that is not greater that x. In this case, we have

$$\gamma_{j+1}^n(X^i) = \{\tilde{M}_{j+1}^n(X^i)\}(1 - \{\tilde{M}_{j+1}^n(X^i)\}).$$

Now we define the approximate filter as follows:

$$\pi_t^n = \frac{1}{m_j^n} \sum_{i=1}^{m_j^n} \tilde{M}_{j+1}^n(X^i, t)\delta_{X_t^i}, \qquad j\delta \leq t < (j+1)\delta,$$

where

$$M_j^n(X^i, t) = \prod_{k=1}^{w} \exp\{\log ap_k(X^i(j\delta))(Y_k(t) - Y_k(j\delta)) - [ap_k(X^i(j\delta)) - 1](t - j\delta)\} \quad (4.2)$$

and

$$\tilde{M}_j^n(X^i, s) = \frac{M_j^n(X^i, s)}{\frac{1}{m_j^n}\sum_{l=1}^{m_j^n}(X^l, s)}$$

Namely, the ith particle has a time-dependent weight $\tilde{M}_j^n(X^i, t)$. At the end of the interval, i.e. $t = (j+1)\delta$, this particle dies and gives birth to a random number of offsprings, whose conditional expectation is equal to the pre-death weight of the particle. The new particles start from their mother's position with weight 1 each.

The process π_t^n is called the hybrid filter since it involves a branching particle system and the empirical measure of these weighted particles.

To show the uniform convergence, we also define the approximation for the unnormalized filter V_t as following: For $k\delta \leq t < (k+1)\delta$,

$$V_t^n = \frac{1}{n}\eta_k^n \sum_{i=1}^{m_k^n} M_{k+1}^n(X^i, t)\delta_{X_t^i},$$

where

$$\eta_k^n = \Pi_{j=1}^k \frac{1}{m_{j-1}^n} \sum_{l=1}^{m_{j-1}^n} M_j^n(X^l).$$

We derive some estimates for the branching particle system introduced above in the following lemmas.

Lemma 4.1.1. *There exists constant K such that for any $i = 1, ..., m_j^n$ and j is bounded, we have*

$$\hat{\mathbb{E}}\left(M_j^n(X^i,t)^2\Big|\mathcal{F}_{j\delta}\right) \le e^{K^2\delta}, \tag{4.3}$$

and

$$\hat{\mathbb{E}}\left(|M_j^n(X^i,t)-1|^2\Big|\mathcal{F}_{j\delta}\right) \le K\delta \tag{4.4}$$

Proof. It's easy to show (4.3) by (4.1). Note that $M_j^n(X^i,t)$ satisfies the following SDE:

$$dM_j^n(X^i,t) = \sum_{k=1}^{w}[ap_k(X^i(t))-1]M_j^n(X^i,t)d\tilde{Y}_k(t),$$

where $\tilde{Y}_k(t) = Y_k(t) - t$ for $k = 1, ..., w$. Thus,

$$\begin{aligned}\hat{\mathbb{E}}\left(|M_j^n(X^i,t)-1|^2\Big|\mathcal{F}_{j\delta}\right) &= \hat{\mathbb{E}}\left(\int_{j\delta}^{t}\left|\sum_{k=1}^{w}[ap_k(X^i(s))-1]M_j^n(X^i,s)\right|^2 ds\right)\\ &\le (C_2-1)^2e^{K\delta}\delta\end{aligned}$$

□

The following lemma is proved in [34]

Lemma 4.1.2. *For any finite j, we have*

$$\hat{\mathbb{E}}\left(m_j^n(\eta_j^n)^2\right) \le Kn \tag{4.5}$$

4.2 Uniform convergence in finite time interval

In this section, we consider the uniform convergence of the branching particle filter over finite time interval $[0, S]$. We consider the backward SPDE:

$$\begin{cases} d\psi_s = -L\psi_s ds - \sum_{k=1}^{w}(ap_k - 1)\psi_{s+}\hat{d}(Y^k(s) - s), \quad 0 \le s \le t \\ \psi_t = \phi \end{cases} \tag{4.6}$$

where $\hat{d}$ denotes the backward Itô integral and ϕ is a bounded function.

Let $f_k, k = 1, 2, ..., w$ and g be bounded functions on $[0, t]$, for $r \in [0, t]$, we define

$$\theta_f^{\vec{Y}}(r) = \prod_{k=1}^{w} \exp\left\{\sqrt{-1}\int_0^r \log f_k(s-)dY_k(s) - \int_0^r (f_k(s) - 1)\, ds\right\}$$

and

$$\theta_g^B(r) = \exp\left\{\sqrt{-1}\int_0^r g_s dB(s) + \frac{1}{2}\int_0^r g_s^2 ds\right\}$$

The following lemma plays an important role in the proof of main theory.

Lemma 4.2.1. *Let* $M_t = \exp\left\{\sum_{k=1}^{w}\left(\log ap_k(X(0))(Y_k(t) - Y_k(0)) - (ap_k(X(0)) - 1)t\right)\right\}$ *and* $\tilde{Y}_k(t) = Y_k(t) - t, k = 1, 2, ..., w$*. Then almost surely, we have*

$$\begin{aligned} \psi_t(X(t))M_t - \psi_0(X(0)) &= \int_0^t M_s\psi_s'\sigma(X(s))dB(s) \\ &\quad + \int_0^t M_s\Theta_s(X(s))\sum_{k=1}^{w}(ap_k(X(0)) - ap_k(X(s)))d\tilde{Y}_k(s), \end{aligned} \tag{4.7}$$

where

$$\Theta_r = \mathbb{E}\left(\psi_r\tilde{\theta}_f(r)|\mathcal{F}_r^{\vec{Y}} \wedge \mathcal{F}_r^B\right)$$

with

$$\tilde{\theta}_f(r) = \theta_f^{\vec{Y}}(t)/\theta_t^{\vec{Y}}(r) = \prod_{k=1}^{w} \exp\left\{\sqrt{-1}\int_r^t \log f_k(s-)dY_k(s) - \int_r^t (f_k(s)-1)\, ds\right\}.$$

Proof. By the proof of lemma 4.2 in [34], we have the followings:

$$d\Theta_r(X(r)) = -\sqrt{1}\left(\sum_{k=1}^{w}(f_k(r)-1)(ap_k(X(r))-1)\Theta_r(X(r))\right)dr + \Theta_r'\sigma(X(r))dB(r),$$

$$dM_r = \sum_{k=1}^{w}[ap_k(X(r))-1]M_r d\tilde{Y}_k(r),$$

$$d\theta_f^{\vec{Y}}(r) = \sqrt{-1}\theta_f^{\vec{Y}}(r-)\sum_{k=1}^{w}(f_k(r)-1)\, d\tilde{Y}_k(r),$$

and

$$d\theta_g^B(r) = \sqrt{-1}\theta_g^B(r)g_r dB(r).$$

Apply Itô's formula to the four equations above, we have

$$\begin{aligned} & d(\Theta_r(X(r))M_r\theta_f^{\vec{Y}}(r)\theta_g^B(r)) \\ = \; & -\sqrt{1}\left(\sum_{k=1}^{w}(f_k(r)-1)(ap_k(X(r))-1)\Theta_r(X(r))\right)M_r\theta_f^{\vec{Y}}(r)\theta_g^B(r)dr \\ & +\Theta_r'\sigma(X(r))M_r\theta_f^{\vec{Y}}(r)\theta_g^B(r)dB(r) + \sum_{k=1}^{w}[ap_k(X(r))-1]\Theta_r M_r\theta_f^{\vec{Y}}(r)\theta_g^B(r)d\tilde{Y}_k(r) \\ & +\sqrt{-1}\theta_f^{\vec{Y}}(r-)\sum_{k=1}^{w}(f_k(r)-1)\,\Theta_r M_r\theta_g^B(r)d\tilde{Y}_k(r) \end{aligned}$$

$$
\begin{aligned}
& +\sqrt{-1}\Theta_r(X(r))M_r\theta_f^{\vec{Y}}(r)\theta_g^B(r)g_r dB(r) + \Theta_r'\sigma(X(r))\sqrt{-1}\theta_g^B(r)g_r M_r\theta_f^{\vec{Y}} dr \\
& +\sum_{k=1}^{w}[ap_k(X(r)) - 1]M_r\sqrt{-1}\theta_f^{\vec{Y}}(r-)\left(f_k(r)-1\right)\Theta_r(X(r))\theta_g^B(r)dr \\
= \;& \Theta_r'\sigma(X(r))\sqrt{-1}\theta_g^B(r)g_r M_r\theta_f^{\vec{Y}} dr \\
& +\sqrt{-1}\sum_{k=1}^{w}\left(ap_k(X(0)) - ap_k(X(r))\right)\left(f_k(r)-1\right)\Theta_r(X(r))M_r\theta_f^{\vec{Y}}(r)\theta_g^B(r)dr \\
& +d(\text{mart.})
\end{aligned}
$$

Note that for $r > 0$,

$$
\mathbb{E}\left(\psi_r(X(r))M_r\theta_f^{\vec{Y}}(t)\theta_g^B(t)|\mathcal{F}_r^{\vec{Y}} \wedge \mathcal{F}_r^B\right) = \Theta_r(X(r))M_r\theta_f^{\vec{Y}}(r)\theta_g^B(r).
$$

Let $Q_k^{\Theta,M}(t) = \left(ap_k(X(0)) - ap_k(X(t))\right)\Theta_t(X(t))M_t$. Combining above two equations, we get

$$
\begin{aligned}
& \mathbb{E}\left((\psi_t(X(t))M_t - \psi_0(X(0)))\theta_f^{\vec{Y}}\theta_g^B(t)\right) \\
= \;& \mathbb{E}\left(\Theta_r(X(r))M_r\theta_f^{\vec{Y}}(r)\theta_g^B(r) - \Theta_0(X(0))\theta_f^{\vec{Y}}(0)\theta_g^B(0)\right) \\
= \;& \int_0^t \mathbb{E}\left(\Theta_r'\sigma(X(r))\theta_g^B(r)g_r M_r\theta_f^{\vec{Y}}\right)dr \\
& +\sqrt{-1}\int_0^t \mathbb{E}\left(\sum_{k=1}^{w} Q_k^{\Theta,M}(r)\left(f_k(r)-1\right)\theta_f^{\vec{Y}}(r)\theta_g^B(r)\right)dr
\end{aligned}
\tag{4.8}
$$

It is showed on page 23 in [34] that

$$
\int_0^t \mathbb{E}\left(\Theta_r'\sigma(X(r))\theta_g^B(r)g_r M_r\theta_f^{\vec{Y}}\right)dr = \mathbb{E}\left(\int_0^t M_s\psi_s'\sigma(X(s))dB(s)\theta_f^{\vec{Y}}(t)\theta_g^B(t)\right). \tag{4.9}
$$

On the other hand, applying integration by parts, we have

$$\begin{aligned}
& \int_0^t \sum_{k=1}^w Q_k^{\Theta,M}(r)d\tilde{Y}_k(r)\theta_f^{\vec{Y}}(t)\theta_g^B(t) \\
= & \int_0^t \sum_{k=1}^w Q_k^{\Theta,M}(r)\theta_f^{\vec{Y}}(r)\theta_g^B(r)d\tilde{Y}_k(r) \\
& + \int_0^t \sqrt{-1}\sum_{k=1}^w Q_k^{\Theta,M}(r)\left(f_k(r)-1\right)\theta_f^{\vec{Y}}(r)\theta_g^B(r)dr \\
& + \int_0^t \cdots dB(r)
\end{aligned}$$

Therefore,

$$\begin{aligned}
& \mathbb{E}\left(\int_0^t \sqrt{-1}\sum_{k=1}^w Q_k^{\Theta,M}(r)\left(f_k(r)-1\right)\theta_f^{\vec{Y}}(r)\theta_g^B(r)dr\right) \\
= & \mathbb{E}\left(\int_0^t \sum_{k=1}^w Q_k^{\Theta,M}(r)d\tilde{Y}_k(r)\theta_f^{\vec{Y}}(t)\theta_g^B(t)\right) \qquad (4.10)
\end{aligned}$$

Combining (4.8), (4.9) and (4.10), we prove (4.7). □

By the definition of M_j^n, almost surely, we have

$$\begin{aligned}
& \psi_{(j+1)\delta}(X^i((j+1)\delta))M_j^n(X^i,(j+1)\delta) - \psi_{j\delta}(X^i(j\delta)) \\
= & \int_{j\delta}^{(j+1)\delta} M_s^n(X^i,s)\psi_s'(X^i(s))\sigma(X^i(s))dB^i(s) \\
& + \int_{j\delta}^{(j+1)\delta} M_s^n(X^i,s)\Theta_s(X^i(s))\sum_{k=1}^w (ap_k(X^i(j\delta)) - ap_k(X^i(s)))d\tilde{Y}_k(s) \qquad (4.11)
\end{aligned}$$

By triangle inequality, for $R\delta \le t < (R+1)\delta$, we have

$$\mathbb{E}d(\pi_t,\pi_t^n) \le \mathbb{E}d(\pi_t,\pi_{R\delta}) + \mathbb{E}d(\pi_{R\delta},\pi_{R\delta}^n) + \mathbb{E}d(\pi_{R\delta}^n,\pi_t^n).$$

Then we begin with the first term on the right side of (4.12).

Lemma 4.2.2. *There exists a constant K such that*

$$\mathbb{E}d(\pi_t, \pi_{R\delta}) \leq K\delta \tag{4.12}$$

Proof. Since π_t satisfies Kushner-FKK equation, we have, for $f_i \in C_b^{k+2}(\mathbb{R}^d) \cap W_2^{k+2}(\mathbb{R}^d)$,

$$\begin{aligned}
\mathbb{E}| \langle \pi_t, f_i \rangle - \langle \pi_{R\delta}, f_i \rangle | &= \mathbb{E}\Big| \int_{R\delta}^t [\langle \pi_s, Lf_i \rangle - \langle \pi_s, fa \rangle + \langle \pi_s, f \rangle \langle \pi_s, a \rangle]\, ds \\
&\quad + \sum_{k=1}^{w} \int_{R\delta}^t \left[\frac{\langle \pi_{s-}, fap_k \rangle}{\langle \pi_{s-}, ap_k \rangle} - \langle \pi_{s-}, f \rangle \right] dY_k(s) \Big| \\
&\leq K_1(t - R\delta) + K_2 \sum_{k=1}^{w} \mathbb{E}(Y_k(t) - Y_k(k\delta)) \\
&\leq K_3(t - R\delta)
\end{aligned}$$

Thus

$$\mathbb{E}d(\pi_t, \pi_{k\delta}) \leq \sum_{i=0}^{\infty} 2^{-i} K_3 \delta = 2K_3\delta. \tag{4.13}$$

□

We now estimate the third term on the right side of (4.12).

Lemma 4.2.3. *There exists a constant K such that*

$$\mathbb{E}d(\pi_t^n, \pi_{R\delta}^n) \leq K\sqrt{\delta} \tag{4.14}$$

Proof. Let $f \in C_b^4(\mathbb{R}^d) \cap W_2^4(\mathbb{R}^d)$ with $\|f\|_{4,\infty} \leq 1$. By the definition of π^n, we have

$$\begin{aligned}
|\langle \pi_t^n - \pi_{R\delta}^n, f\rangle| &= \Big|\frac{1}{m_R^n}\sum_{i=1}^{m_R^n} \tilde{M}(X^i,t)f(X_t^i) - \frac{1}{m_R^n}\sum_{i=1}^{m_R^n} f(X_{R\delta}^i)\Big| \\
&= \Big|\frac{1}{m_R^n}\sum_{i=1}^{m_R^n}(\tilde{M}(X^i,t)f(X_t^i) - f(X_t^i) + f(X_t^i) - f(X_{R\delta}^i))\Big| \\
&\leq \frac{1}{m_R^n}\sum_{i=1}^{m_R^n}|\tilde{M}(X^i,t)f(X_t^i) - f(X_t^i)| + \frac{1}{m_R^n}\sum_{i=1}^{m_R^n}|f(X_t^i) - f(X_{R\delta}^i)| \\
&\leq \frac{1}{m_R^n}\sum_{i=1}^{m_R^n}|\tilde{M}(X^i,t) - 1| + K_1\frac{1}{m_R^n}\sum_{i=1}^{m_R^n}|X_t^i - X_{R\delta}^i|
\end{aligned}$$

By the proof of Lemma 4.4 in [34], we have

$$\tilde{M}^n(X^i,t) = 1 + \int_{R\delta}^t \tilde{M}^n(X^i,s)\sum_{k=1}^w\left[\frac{\lambda_k(X^i(R\delta))}{\bar{h}_k(s-)+1} - 1\right]dY_k(s),$$

where $\bar{h}_s = \frac{1}{m_R^n}\sum_{i=1}^{m_R^n}\tilde{M}_j^n(X^i,s)(\lambda_k(X^i(R\delta)) - 1)$.

Since $C_1 \leq \lambda_k \leq C_2$ and $\sum_{i=1}^{m_k^n}\tilde{M}(X^i,s) = m_k^n$, we have $C_1 \leq \bar{h}_s + 1 \leq C_2$. Thus,

$$\begin{aligned}
\hat{\mathbb{E}}|\tilde{M}^n(X^i,t)|^2 &\leq 2 + 2\hat{\mathbb{E}}\Big|\int_{k\delta}^t \tilde{M}^n(X^i(R\delta))\sum_{k=1}^w\left[\frac{\lambda_k(X^i(R\delta))}{\bar{h}_k(s-)+1} - 1\right]dY_k(s)\Big|^2 \\
&\leq 2 + 4\hat{\mathbb{E}}\Big|\int_{R\delta}^t \tilde{M}^n(X^i,s)\sum_{k=1}^w\left[\frac{\lambda_k(X^i(k\delta))}{\bar{h}_k(s-)+1} - 1\right](dY_k(s) - ds)\Big|^2 \\
&\quad +4\hat{\mathbb{E}}\Big|\int_{R\delta}^t \tilde{M}^n(X^i,s)\sum_{k=1}^w\left[\frac{\lambda_k(X^i(R\delta))}{\bar{h}_k(s-)+1} - 1\right]ds\Big|^2 \\
&\leq 2 + 4\hat{\mathbb{E}}\int_{R\delta}^t\left(\tilde{M}^n(X^i,s)\right)^2\left(\sum_{k=1}^w\left[\frac{\lambda_k(X^i(R\delta))}{\bar{h}_k(s-)+1} - 1\right]\right)^2 ds \\
&\quad +4\delta\hat{\mathbb{E}}\int_{k\delta}^t\left(\tilde{M}^n(X^i,s)\right)^2\left(\sum_{k=1}^w\left[\frac{\lambda_k(X^i(R\delta))}{\bar{h}_k(s-)+1} - 1\right]\right)^2 ds \\
&\leq 2 + 4w(1+\delta)\frac{C_2 - C_1}{C_1}\int_{R\delta}^t\hat{\mathbb{E}}\left(\tilde{M}^n(X^i,s)\right)^2 ds
\end{aligned}$$

By Gronwall's inequality,

$$\hat{\mathbb{E}}|\tilde{M}^n(X^i,t)|^2 \le K_2.$$

Then

$$\begin{aligned}
\hat{\mathbb{E}}|\tilde{M}^n(X^i,t)-1|^2 &= \hat{\mathbb{E}}\Big|\int_{R\delta}^t \tilde{M}^n(X^i,s)\sum_{k=1}^w \left[\frac{\lambda_k(X^i(R\delta))}{\bar{h}_k(s-)+1}-1\right] dY_k(s)\Big|^2 \\
&\le 2\hat{\mathbb{E}}\Big|\int_{R\delta}^t \tilde{M}^n(X^i,s)\sum_{k=1}^w \left[\frac{\lambda_k(X^i(R\delta))}{\bar{h}_k(s-)+1}-1\right] (dY_k(s)-ds)\Big|^2 \\
&\quad +2\hat{\mathbb{E}}\Big|\int_{R\delta}^t \tilde{M}^n(X^i,s)\sum_{k=1}^w \left[\frac{\lambda_k(X^i(R\delta))}{\bar{h}_k(s-)+1}-1\right] ds\Big|^2 \\
&\le 2\hat{\mathbb{E}}\int_{R\delta}^t \left(\tilde{M}^n(X^i,s)\right)^2 \left(\sum_{k=1}^w \left[\frac{\lambda_k(X^i(R\delta))}{\bar{h}_k(s-)+1}-1\right]\right)^2 ds \\
&\quad +2\delta\hat{\mathbb{E}}\int_{R\delta}^t \left(\tilde{M}^n(X^i,s)\right)^2 \left(\sum_{k=1}^w \left[\frac{\lambda_k(X^i(R\delta))}{\bar{h}_k(s-)+1}-1\right]\right)^2 ds \\
&\le K_3\delta. \qquad (4.15)
\end{aligned}$$

On the other hand,

$$\begin{aligned}
\hat{\mathbb{E}}|X_t^i - X_{R\delta}^i|^2 &= \hat{\mathbb{E}}\left|\mu(X^i(R\delta))(t-k\delta)+\sigma(X^i(R\delta))(B^i(s)-B^i(R\delta)\right|^2 \\
&\le 2\delta\hat{\mathbb{E}}\int_{R\delta}^t |\mu(X_s^i)|^2 ds + 2\hat{\mathbb{E}}\int_{R\delta}^t |\sigma(X_s^i)|^2 ds \\
&\le K_4\delta, \qquad (4.16)
\end{aligned}$$

where the last inequality follows the linear growth condition of coefficients and stationary assumption.

Therefore,

$$\begin{aligned}
\hat{\mathbb{E}}\left\langle \pi_t^n - \pi_{R\delta}^n, f\right\rangle^2 &\le K\hat{\mathbb{E}}\frac{1}{m_R^n}\sum_{i=1}^{m_R^n}\hat{\mathbb{E}}(|\tilde{M}^n(X^i,t)-1|^2 + |X_t^i - X_{R\delta}^i|^2|\mathcal{F}_{R\delta}) \\
&\le K_5\delta.
\end{aligned}$$

Let

$$M_{s,t} \equiv \exp\left(\sum_{k=1}^{w}\int_s^t \lambda_k(X_s)dY_k(s) - \frac{1}{2}\int_s^t [\lambda_k(X_s)-1]ds\right).$$

Then by the boundness of λ_k, we have

$$\hat{\mathbb{E}}M_{s,t}^2 \le e^{K(t-s)}. \tag{4.17}$$

By Cauchy-Schwarz inequality, we have

$$\begin{aligned}
\mathbb{E}d(\pi_t^n, \pi_{R\delta}^n) &= \hat{\mathbb{E}}(d(\pi_t^n, \pi_{R\delta}^n)M_{R\delta,t}) \\
&\le (\hat{\mathbb{E}}d(\pi_t^n, \pi_{R\delta}^n)^2)^{1/2}(\hat{\mathbb{E}}M_{R\delta,t}^2)^{1/2} \\
&\le (\sum_{i=0}^{\infty} 2^{-i}\hat{\mathbb{E}}\left\langle \pi_t^n - \pi_{R\delta}^n, f_i\right\rangle^2)^{1/2}(\hat{\mathbb{E}}M_{R\delta,t}^2)^{1/2} \\
&\le K\sqrt{\delta}.
\end{aligned}$$

□

Finally, we estimate the middle term of (4.12).

Lemma 4.2.4. *There exists a constant K such that*

$$\sup_{0\le l\le R} \mathbb{E}d(\pi_{l\delta}^n, \pi_{l\delta}) \le Kn^{-1}, \tag{4.18}$$

where R is a finite number.

Proof. Let ψ be the solution of (4.6) with final condition $\psi_{k\delta} = \phi$, where $\phi \in C_b^4(\mathbb{R}) \cap W_2^4(\mathbb{R})$ with $\|\phi\|_{4,\infty} \le 1$ and also $\|\phi\|_{4,2} \le 1$. Then

$$|\langle \pi_{l\delta}, \phi\rangle - \langle \pi_{l\delta}^n, \phi\rangle| \le \frac{|\left\langle V_{l\delta}, \phi\right\rangle - \left\langle V_{l\delta}^n, \phi\right\rangle|}{\langle V_{l\delta}, 1\rangle} + \frac{|\left\langle V_{l\delta}, 1\right\rangle - \left\langle V_{l\delta}^n, 1\right\rangle|}{\langle V_{l\delta}, 1\rangle} \tag{4.19}$$

First we show that

$$\sup_{0\le l\le R} \mathbb{E}d(V_{l\delta}^n, V_{l\delta}) \le Kn^{-1} \tag{4.20}$$

As $\psi_{l\delta} = \phi$, we get

$$
\begin{aligned}
\langle V_{l\delta}, \phi \rangle - \langle V^n_{l\delta}, \phi \rangle &= (\langle V_{l\delta}, \phi \rangle - \langle V^n_0, \psi_0 \rangle) - \sum_{j=1}^{l} \left(\left\langle V^n_{j\delta}, \psi_{j\delta} \right\rangle - \left\langle V^n_{(j-1)\delta}, \psi_{(j-1)\delta} \right\rangle \right) \\
&= \langle V_{l\delta}, \phi \rangle - \langle V^n_0, \psi_0 \rangle \\
&\quad - \sum_{j=1}^{l} \left(\left\langle V^n_{j\delta}, \psi_{j\delta} \right\rangle - \hat{\mathbb{E}} \left(\left\langle V^n_{j\delta}, \psi_{j\delta} \right\rangle \middle| \mathcal{F}_{j\delta-} \vee \mathcal{G}_{j\delta,k\delta} \right) \right) \\
&\quad - \sum_{j=1}^{l} \left(\hat{\mathbb{E}} \left(\left\langle V^n_{j\delta}, \psi_{j\delta} \right\rangle \middle| \mathcal{F}_{j\delta-} \vee \mathcal{G}_{j\delta,k\delta} \right) - \left\langle V^n_{(j-1)\delta}, \psi_{(j-1)\delta} \right\rangle \right) \\
&\equiv I^n_1 - I^n_2 - I^n_3,
\end{aligned}
$$

where $\mathcal{G}_{s,t} = \sigma(\vec{Y}_u - \vec{Y}_s : s \leq u \leq t)$.

By corollary 6.22 in [33], we have

$$
\begin{aligned}
\hat{\mathbb{E}} \left| \langle V_{k\delta}, \phi \rangle - \langle V^n_0, \psi_0 \rangle \right|^2 &= \hat{\mathbb{E}} \left| \langle V_0, \psi_0 \rangle - \langle V^n_0, \psi_0 \rangle \right| \\
&\leq Kn^{-1} \qquad (4.21)
\end{aligned}
$$

By the proof of Theorem 4.1 in [34], we have

$$
\begin{aligned}
\hat{\mathbb{E}}(I^n_2)^2 &= \hat{\mathbb{E}} \left| \sum_{j=1}^{l} \eta^n_j \frac{1}{n} \sum_{i=1}^{m^n_{j-1}} \psi_{j\delta}(X^i_{j\delta})(\xi^i_j - \tilde{M}^n_j(X^i)) \right|^2 \\
&\leq \hat{\mathbb{E}} \sum_{j=1}^{l} \frac{1}{n^2} \sum_{i=1}^{m^n_{j-1}} \hat{\mathbb{E}} \left(\|\psi_{j\delta}\|^2_{0,\infty} \gamma^n_j(X^i)(\eta^n_j)^2 \middle| \mathcal{F}_{(j-1)\delta} \right) \\
&\leq Kn^{-1} \qquad (4.22)
\end{aligned}
$$

By the definition of V^n and (8.26) in [33], we have

$$\begin{aligned} I_3^n &= \sum_{j=1}^{l} \eta_{j-1}^n \frac{1}{n} \sum_{i=1}^{m_{j-1}^n} \left(\psi_{j\delta}(X_{j\delta}^i) M_j^n(X^i) - \psi_{(j-1)\delta}(X_{(j-1)\delta}^i)\right) \\ &= \sum_{j=1}^{l} \eta_{j-1}^n \frac{1}{n} \sum_{i=1}^{m_{j-1}^n} \Big(\int_{(j-1)\delta}^{j\delta} M_s \psi_s' \sigma(X_s^i) dB_s^i \\ &\quad + \int_{(j-1)\delta}^{j\delta} M_s \Theta_s(X_s^i) \sum_{k=1}^{w} (ap_k(X_0^i) - ap_k(X_s^i)) d\tilde{Y}_k(s) \Big) \end{aligned}$$

Note that

$$\hat{\mathbb{E}}\left(\int_{(j-1)\delta}^{j\delta} M_s \psi_s' \sigma(X_s^i) dB^i(s) \Big| \mathcal{F}_{(j-1)\delta} \vee \mathcal{G}_{j\delta} \right) = 0$$

and

$$\hat{\mathbb{E}}\left(\int_{(j-1)\delta}^{j\delta} M_s \Theta_s(X(s)) \sum_{k=1}^{w} (ap_k(X_0^i) - ap_k(X_s^i)) d\tilde{Y}_k(s) \Big| \mathcal{F}_{(j-1)\delta} \right) = 0.$$

Hence,

$$\begin{aligned} \hat{\mathbb{E}}(I_3^n)^2 &= \sum_{j=1}^{l} \hat{\mathbb{E}}\Big(\eta_{j-1}^n \frac{1}{n} \sum_{i=1}^{m_{j-1}^n} \Big(\int_{(j-1)\delta}^{j\delta} M_s \psi_s' \sigma(X_s^i) dB^i(s) \\ &\quad + \int_{(j-1)\delta}^{j\delta} M_s \Theta_s(X_s^i) \sum_{k=1}^{w} (ap_k(X_0^i) - ap_k(X_s^i)) d\tilde{Y}_k(s) \Big) \Big)^2 \\ &= \sum_{j=1}^{l} \hat{\mathbb{E}}\Big((\eta_{j-1}^n)^2 \frac{1}{n^2} \hat{\mathbb{E}}\Big(\sum_{i=1}^{m_{j-1}^n} \Big(\int_{(j-1)\delta}^{j\delta} M_s \psi_s' \sigma(X_s^i) dB^i(s) \\ &\quad + \int_{(j-1)\delta}^{j\delta} M_s \Theta_s(X_s^i) \sum_{k=1}^{w} (ap_k(X_0^i) - ap_k(X_s^i)) d\tilde{Y}_k(s) \Big) \Big| \mathcal{F}_{(j-1)\delta} \vee \mathcal{G}_{j\delta} \Big)^2 \Big) \end{aligned} \tag{4.23}$$

$$
\begin{aligned}
&= \sum_{j=1}^{l} \hat{\mathbb{E}}\Big((\eta_{j-1}^n)^2 \frac{1}{n^2} \sum_{i=1}^{m_{j-1}^n} \Big(\int_{(j-1)\delta}^{j\delta} M_s \psi_s' \sigma(X_s^i) dB^i(s) \\
&\quad + \int_{(j-1)\delta}^{j\delta} M_s \Theta_s(X_s^i) \sum_{k=1}^{w} (ap_k(X_0^i) - ap_k(X_s^i)) d\tilde{Y}_k(s) \Big)^2 \Big) \\
&\leq 2 \sum_{j=1}^{l} \hat{\mathbb{E}}\Big((\eta_{j-1}^n)^2 \frac{1}{n^2} \sum_{i=1}^{m_{j-1}^n} \Big(\int_{(j-1)\delta}^{j\delta} M_s^2 |\psi_s' \sigma(X_s^i)|^2 ds \\
&\quad + \int_{(j-1)\delta}^{j\delta} M_s^2 |\Theta_s(X_s^i)|^2 \sum_{k=1}^{w} (ap_k(X_0^i) - ap_k(X_s^i))^2 ds \Big) \Big) \\
&\leq K l \delta n^{-2} \hat{\mathbb{E}} \left(m_{j-1}^n (\eta_{j-1}^n)^2 \right) \\
&\leq K n^{-1}, \qquad (4.24)
\end{aligned}
$$

where the last inequality follows from (4.5).

Combining (4.21), (4.22) and (4.24), we have

$$
\hat{\mathbb{E}} \left(| \langle V_{l\delta}, \phi \rangle - \langle V_{l\delta}^n, \phi \rangle |^2 \right) \leq K n^{-1}. \qquad (4.25)
$$

Therefore,

$$
\sup_{0 \leq l \leq R} \hat{\mathbb{E}} d(V_{l\delta}^n, V_{l\delta})^2 \leq K n^{-1}.
$$

As

$$
d \langle V_t, \phi \rangle = \langle V_t, L\phi \rangle \, dt + \sum_{k=1}^{w} \langle V_{t-}, (\lambda_k - 1)\phi \rangle \, d\tilde{Y}_k(t),
$$

we have

$$
d \langle V_t, 1 \rangle = \sum_{k=1}^{w} \langle V_{t-}, (\lambda_k - 1) \rangle \, d\tilde{Y}_k(t). \qquad (4.26)
$$

By Itô's formula (see page 78 of [30]), we have

$$\begin{aligned}
\log \langle V_t, 1\rangle &= \int_0^t \langle V_{s-}, 1\rangle^{-1} d\langle V_s, 1\rangle - \frac{1}{2}\int_0^t \langle V_{s-}, 1\rangle^{-2} d\left[\langle V_s, 1\rangle, \langle V_s, 1\rangle\right]^c \\
&\quad + \sum_{0<s\le t} \left\{\log \langle V_s, 1\rangle - \log \langle V_{s-}, 1\rangle - \langle V_{s-}, 1\rangle^{-1} \Delta \langle V_s, 1\rangle\right\} \\
&= \int_0^t \sum_{k=1}^{w} \frac{\langle V_{s-}, (\lambda_k - 1)\rangle}{\langle V_{s-}, 1\rangle} d\tilde{Y}_k(s) + \sum_{0<s\le t} \log \frac{\langle V_s, 1\rangle}{\langle V_{s-}, 1\rangle} \\
&\quad - \sum_{0<s\le t} \frac{\Delta \langle V_s, 1\rangle}{\langle V_{s-}, 1\rangle} \\
&= \int_0^t \sum_{k=1}^{w} \langle \pi_{s-}, (\lambda_k - 1)\rangle \, d\tilde{Y}_k(s) + \sum_{0<s\le t} \log \frac{\langle V_s, 1\rangle}{\langle V_{s-}, 1\rangle} \\
&\quad - \sum_{0<s\le t} \frac{\Delta \langle V_s, 1\rangle}{\langle V_{s-}, 1\rangle} \\
&\equiv I_1 + I_2 - I_3 \qquad (4.27)
\end{aligned}$$

By (4.26), we have

$$\begin{aligned}
\langle V_t, 1\rangle &= \langle V_{t-}, 1\rangle + \int_{t-}^t \sum_{k=1}^{w} \langle V_{s-}, (\lambda_k - 1)\rangle \, d\tilde{Y}_k(s) \\
&= \langle V_{t-}, 1\rangle + \sum_{k=1}^{w} \langle V_{t-}, (\lambda_k - 1)\rangle \, \Delta\tilde{Y}_k(t) \qquad (4.28)
\end{aligned}$$

Plugging (4.28) in I_2 and I_3, we have

$$\begin{aligned}
I_2 &= \sum_{0<s\le t} \log \left(\frac{\langle V_{s-}, 1\rangle + \sum_{k=1}^{w} \langle V_{s-}, (\lambda_k - 1)\rangle \, \Delta\tilde{Y}_k(t)}{\langle V_{s-}, 1\rangle} \right) \\
&= \sum_{0<s\le t} \log \left(1 + \sum_{k=1}^{w} \langle \pi_{s-}, (\lambda_k - 1)\rangle \, \Delta\tilde{Y}_k(t) \right) \qquad (4.29)
\end{aligned}$$

and

$$
\begin{aligned}
I_3 &= \sum_{0<s\le t} \frac{\sum_{k=1}^{w} \langle V_{s-}, (\lambda_k - 1)\rangle \, \Delta\tilde{Y}_k(t)}{\langle V_{s-}, 1\rangle} \\
&= \hat{\mathbb{E}} \left(\sum_{0<s\le t} \sum_{k=1}^{w} \left\langle \pi^n_{s-}, (\lambda_k - 1)\right\rangle \Delta\tilde{Y}_k(t) \right)
\end{aligned}
\tag{4.30}
$$

Combining (4.27), (4.29) and (4.30), we have

$$
\langle V_t, 1\rangle = \exp\left\{ \int_0^t \log\left(1 + \sum_{k=1}^{w} \langle \pi_{s-}, \lambda_k - 1\rangle \right) dY_k(s) + \int_0^t \sum_{k=1}^{w} \langle \pi_{s-}, (\lambda_k - 1)\rangle \, ds \right\}
$$

Therefore,

$$
\sup_{0\le t\le S} \hat{\mathbb{E}} \, \langle V_t, 1\rangle^{-4} \le \infty \tag{4.31}
$$

By the boundness of λ_k, we have $\sup_{0\le t\le S} \hat{\mathbb{E}} M_S^4 < \infty$. Therefore,

$$
\begin{aligned}
\sup_{0\le l\le R} \mathbb{E} d(\pi_{l\delta}, \pi^n_{l\delta}) &\le \sup_{0\le l\le R} \hat{\mathbb{E}} \left\{ \frac{|\,\langle V_{l\delta} - V^n_{l\delta}, 1\rangle\,|}{\langle V_{l\delta}, 1\rangle} + \frac{d(V_{l\delta}, V^n_{l\delta})}{\langle V_{l\delta}, 1\rangle} \right\} M_S \\
&\le \left(\sup_{0\le l\le R} \hat{\mathbb{E}} |\,\langle V_{l\delta} - V^n_{l\delta}, 1\rangle\,|^2 \right)^{1/2} \left(\sup_{0\le l\le R} \hat{\mathbb{E}} \frac{M_S^2}{\langle V_{l\delta}, 1\rangle^2} \right)^{1/2} \\
&\quad + \left(\sup_{0\le l\le R} \hat{\mathbb{E}} d(V_{l\delta}, V^n_{l\delta})^2 \right)^{1/2} \left(\sup_{0\le l\le R} \hat{\mathbb{E}} \frac{M_S^2}{\langle V_{l\delta}, 1\rangle^2} \right)^{1/2} \\
&\le K n^{-1}
\end{aligned}
$$

□

Theorem 4.2.1. *For any finite positive number S, we have*

$$
\lim_{n\to\infty} \sup_{0\le t\le S} \mathbb{E} d(\pi_t, \pi^n_t) = 0. \tag{4.32}
$$

Proof. Combining (4.12), (4.12), (4.18) and (4.14), we have

$$\sup_{0\le t\le S} \mathbb{E}d(\pi_t,\pi_t^n)\le n^{-\alpha}.$$

This implies (4.32). □

4.3 Uniform convergence over the real line

In this section, we consider the uniform convergence of the branching particle filter over the real line.

Theorem 4.3.1. *Under assumptions 3.1.1, 3.1.2, 3.1.3 and 4.1.1, the branching particle filter uniformly converges to the optimal filter in the following sense:*

$$\lim_{n\to\infty}\sup_{t>0}\mathbb{E}d(\pi_t,\pi_t^n)=0. \tag{4.33}$$

Let $p(t,x,A)$ be the transition probability of the Markov process X_t. There exists a probability measure $P_{s,x}$ on $C(\mathbb{R}_+,\mathbb{R}^d)$ such that for $t>s$ and $A\in\mathcal{B}(\mathbb{R}^d)$,

$$P_{s,x}(\xi_t\in A|\mathcal{F}_s^\xi)=p(t-s,x,A),\qquad P_{s,x}-a.s.,$$

and

$$P_{s,x}(\xi_u=x,0\le u\le s)=1$$

where ξ_t is the co-ordinate process on $C(\mathbb{R}_+,\mathbb{R}^d)$, i.e. $\xi_t(\theta)=\theta_t$ for all $\theta\in C(\mathbb{R}_+,\mathbb{R}^d)$.

Let λ be the initial distribution of X_t and $\eta\in C(\mathbb{R}_+,\mathbb{R}^m)$. We define an $\mathcal{M}_F(\mathbb{R}^d)$-valued process $\Gamma_{s,t}(\lambda)$ and a $\mathcal{P}(\mathbb{R}^d)$-valued process $\Lambda_{s,t}(\lambda)$ on $C(\mathbb{R}_+,\mathbb{R}^m)$ as

$$\langle\Gamma_{s,t}(\lambda)(\eta),f\rangle=\int_{\mathbb{R}^d}\int_{C(\mathbb{R}_+,\mathbb{R}^d)}f(\xi_t(\theta))q_{st}(\theta,\eta)P_{s,x}(d\theta)\lambda(dx),$$

and

$$\Lambda_{s,t}(\lambda)(\eta) = \frac{\langle \Gamma_{s,t}(\lambda)(\eta), f\rangle}{\langle \Gamma_{s,t}(\lambda)(\eta), 1\rangle},$$

where $q_{st}(\theta,\eta) = \exp\left(\int_s^t h(\xi_u(\theta))^* d\beta_u(\eta) - \frac{1}{2}\int_s^t |h(\xi_u(\theta))|^2 du\right)$ and $\beta_t(\eta) = \eta_t$ is the co-ordinate process on $C(\mathbb{R}_+, \mathbb{R}^m)$.

Let $\Lambda_{R\delta,(R+1)\delta}(\lambda)(\vec{Y})$ be the optimal filter at time $(R+1)\delta$ using the observation $\sigma(\vec{Y}_t, R\delta \le t \le (R+1)\delta)$ starting with λ at time $R\delta$. We define the following $\mathcal{P}(\mathbb{R}^d)$-valued processes

$$\pi^n_{R\delta,(R+1)\delta} := \Lambda_{R\delta,(R+1)\delta}(\pi^n_{R\delta})(\vec{Y}),$$

and for $j < R$

$$\begin{aligned}
\pi^n_{j\delta,R\delta} := \Lambda_{j\delta,R\delta}(\pi^n_{j\delta})(\vec{Y}) &= \Lambda_{(R-1)\delta,R\delta} \circ \cdots \circ \Lambda_{j\delta,(j+1)\delta}(\pi^n_{j\delta})(\vec{Y}), \\
\pi^n_{0,j\delta} := \pi_{j\delta} &= \Lambda_{(j-1)\delta,j\delta} \circ \cdots \circ \Lambda_{0,\delta}(\pi_0)(\vec{Y}), \qquad (4.34)\\
\pi^n_{j\delta,j\delta} &:= \pi^n_{j\delta} \qquad (4.35)
\end{aligned}$$

The following is our strategy of the proof. For $R\delta \le t < (R+1)\delta$, we write the distance between π_t and π^n_t as a sum of three distances: $d(\pi_{R\delta}, \pi_t), d(\pi_{R\delta}, \pi^n_{R\delta})$ and $d(\pi^n_{R\delta}, \pi^n_t)$ by the triangle inequality. The estimates of $d(\pi_{R\delta}, \pi_t)$ and $d(\pi^n_{R\delta}, \pi^n_t)$ are showed in previous section. The following lemma and exponential stability which is proved in chapter 3 are used to estimate $d(\pi_{R\delta}, \pi^n_{R\delta})$.

Lemma 4.3.1. *There exists a constant K_8 such that*

$$\lim_{\delta\to 0} \mathbb{E} d_{TV}(\pi^n_{j\delta,(j+1)\delta}, \pi^n_{(j+1)\delta}) = 0 \qquad (4.36)$$

Proof. Note that $\pi^n_{j\delta,(j+1)\delta}$ and $\pi^n_{(j+1)\delta}$ have the same initial distribution $\pi^n_{j\delta}$ at time $j\delta$. Let $V^n_{j\delta,(j+1)\delta}$ and $V^n_{(j+1)\delta}$ be the unnormalized optimal filter and unnormalized particle filter, respectively, with the same initial distribution $\pi^n_{j\delta}$. Note that for continuous f bounded by 1, we have

$$\left|\left\langle \pi^n_{j\delta,(j+1)\delta} - \pi^n_{(j+1)\delta}, f\right\rangle\right| \leq \left|\left\langle \pi^n_{j\delta,(j+1)\delta}, f\right\rangle - \left\langle \pi^n_{j\delta}, f\right\rangle\right| + \left|\left\langle \pi^n_{j\delta}, f\right\rangle - \left\langle \pi^n_{(j+1)\delta}, f\right\rangle\right|$$

By using Kushner-FKK equation, we have

$$\hat{\mathbb{E}}\left|\left\langle \pi^n_{j\delta,(j+1)\delta}, f\right\rangle - \left\langle \pi^n_{j\delta}, f\right\rangle\right| \leq K\delta^{1/2}$$

On the other hand,

$$\hat{\mathbb{E}}\left|\left\langle \pi^n_{j\delta}, f\right\rangle - \left\langle \pi^n_{(j+1)\delta}, f\right\rangle\right| \leq \hat{\mathbb{E}}\frac{\left|\left\langle V^n_{(j+1)\delta} - V^n_{j\delta}, f\right\rangle\right|}{\left\langle V^n_{j\delta}, 1\right\rangle} + \hat{\mathbb{E}}\frac{\left|\left\langle V^n_{(j+1)\delta} - V^n_{j\delta}, 1\right\rangle\right|}{\left\langle V^n_{j\delta}, 1\right\rangle}$$

Now we show

$$\hat{\mathbb{E}}\frac{\left|\left\langle V^n_{(j+1)\delta} - V^n_{j\delta,(j+1)\delta}, f\right\rangle\right|}{\left\langle V^n_{j\delta}, 1\right\rangle} \leq K\delta^{1/2}. \tag{4.37}$$

By the definition of V^n, we have

$$\begin{aligned}
&\hat{\mathbb{E}}\frac{\left|\left\langle V^n_{(j+1)\delta}, f\right\rangle - \left\langle V^n_{j\delta}, f\right\rangle\right|}{\left\langle V^n_{j\delta}, 1\right\rangle} \\
= \;&\hat{\mathbb{E}}\frac{\left|\frac{1}{n}\eta_j \sum_{i=1}^{m^n_j}\left(f(X^i_{(j+1)\delta})M^n_{j+1}(X^i) - f(X^i_{j\delta})\right)\right|}{\frac{1}{n}\eta_j m^n_j}
\end{aligned}$$

$$
\begin{aligned}
&\leq \hat{\mathbb{E}}\left(\frac{1}{m_j^n}\sum_{i=1}^{m_j^n}\left(\left|f(X^i_{(j+1)\delta})M^n_{j+1}(X^i)-f(X^i_{(j+1)\delta})\right|+\left|f(X^i_{(j+1)\delta})-f(X^i_{j\delta})\right|\right)\right)\\
&= \hat{\mathbb{E}}\left(\frac{1}{m_j^n}\sum_{i=1}^{m_j^n}\hat{\mathbb{E}}\left(\left|f(X^i_{(j+1)\delta})(M^n_{j+1}(X^i)-1)\right|+\left|f(X^i_{(j+1)\delta})-f(X^i_{j\delta})\right|\Big|\mathcal{F}_{j\delta}\right)\right)\\
&\leq \hat{\mathbb{E}}\left(\frac{1}{m_j^n}\sum_{i=1}^{m_j^n}\hat{\mathbb{E}}\left(\left|M^n_{j+1}(X^i)-1\right|\Big|\mathcal{F}_{j\delta}\right)+\frac{1}{m_j^n}\sum_{i=1}^{m_j^n}\hat{\mathbb{E}}\left(\left|f(X^i_{(j+1)\delta})-f(X^i_{j\delta})\right|\Big|\mathcal{F}_{j\delta}\right)\right)\\
&\leq e(\delta),
\end{aligned}
$$

where the last inequality is from (4.15) and $e(\delta)$ approaches 0 as δ tends to 0.

Then by (4.17), we have

$$
\begin{aligned}
\mathbb{E}d_{TV}(\pi^n_{j\delta,(j+1)\delta},\pi^n_{(j+1)\delta}) &\leq \hat{\mathbb{E}}(d_{TV}(\pi^n_{j\delta,(j+1)\delta},\pi^n_{(j+1)\delta})M_{j\delta,(j+1)\delta})\\
&\leq (\hat{\mathbb{E}}d_{TV}(\pi^n_{j\delta,(j+1)\delta},\pi^n_{(j+1)\delta})^2)^{1/2}(\hat{\mathbb{E}}M_{j\delta,(j+1)\delta})^2)^{1/2}\\
&\leq (K\sqrt{\delta}+e(\delta))^{1/2}e^{K\delta}
\end{aligned}
$$

Let δ tend to 0, we have (4.36). □

To estimate $d(\pi_{R\delta},\pi^n_{R\delta})$, we rewrite as

$$
d(\pi_{R\delta},\pi^n_{R\delta})=\sum_{j=1}^{R}d(\pi^n_{j\delta,R\delta},\pi^n_{(j+1)\delta,R\delta}). \tag{4.38}
$$

By the exponential stability which is proved in chapter 3, $\forall\epsilon>0$, there exist positive constants K_1, K_2 and $T(\epsilon)$, such that when $t\geq T$, we have

$$
\mathbb{E}d_{TV}(\pi_t,\bar{\pi}_t)\leq K_1\mathbb{E}d_{TV}(\pi_0,\bar{\pi}_0)e^{-K_2t^{\alpha}}. \tag{4.39}
$$

Note that, $\pi^n_{j\delta,R\delta}$ is the optimal filter at time $R\delta$ starting at time $(j+1)\delta$ with measure $\pi^n_{j\delta,(j+1)\delta}$. Similarly, $\pi^n_{(j+1)\delta,R\delta}$ is the optimal filter at the same time but with initial

$\pi^n_{(j+1)\delta}$ at the initial time $(j+1)\delta$.Therefore, when $(R-j-1)\delta > T(\epsilon)$, we have

$$\begin{aligned}\mathbb{E}d_{TV}(\pi^n_{j\delta,R\delta},\pi^n_{(j+1)\delta,R\delta}) &\leq K_1 e^{-K_2((R-j-1)\delta)^\alpha}\mathbb{E}d_{TV}(\pi^n_{j\delta,(j+1)\delta},\pi^n_{(j+1)\delta}) \\ &\leq K_1(K\sqrt{\delta}+e(\delta))^{1/2}e^{-K_2((R-j-1)\delta)^\alpha} \qquad (4.40)\end{aligned}$$

Let j_0 be the largest j such that $(R-j-1)\delta > T(\epsilon)$, we have

$$\begin{aligned}\mathbb{E}d(\pi_t,\pi^n_t) &\leq \mathbb{E}d(\pi_t,\pi_{R\delta})+\mathbb{E}d(\pi_{R\delta},\pi^n_{R\delta})+\mathbb{E}d(\pi^n_{R\delta},\pi^n_t) \\ &\leq \mathbb{E}d(\pi_t,\pi_{R\delta})+\sum_{j=0}^{R-1}\mathbb{E}d(\pi^n_{j\delta,R\delta},\pi^n_{(j+1)\delta,R\delta})+\mathbb{E}d(\pi^n_{R\delta},\pi^n_t) \\ &\leq d\mathbb{E}(\pi_t,\pi_{R\delta})+\sum_{j=R-j_0}^{R}\mathbb{E}d(\pi^n_{j\delta,R\delta},\pi^n_{(j+1)\delta,R\delta}) \\ &\quad +\sum_{j=0}^{R-j_0-1}\mathbb{E}d(\pi^n_{j\delta,R\delta},\pi^n_{(j+1)\delta,R\delta})+\mathbb{E}d(\pi^n_{R\delta},\pi^n_t) \\ &\leq \mathbb{E}d(\pi_t,\pi_{R\delta})+\sum_{j=R-j_0}^{R}\mathbb{E}d(\pi^n_{j\delta,R\delta},\pi^n_{(j+1)\delta,R\delta}) \\ &\quad +\sum_{j=0}^{\infty}\mathbb{E}d_{TV}(\pi^n_{j\delta,R\delta},\pi^n_{(j+1)\delta,R\delta})+\mathbb{E}d(\pi^n_{R\delta},\pi^n_t) \\ &\leq +K_3\sqrt{\delta}+\sum_{j=R-j_0}^{R}\mathbb{E}d(\pi^n_{j\delta,R\delta},\pi^n_{(j+1)\delta,R\delta}) \\ &\quad +K_1(K\sqrt{\delta}+e(\delta))^{1/2}\sum_{j=0}^{\infty}e^{-K_2((R-j-1)\delta)^\alpha}+K_4\sqrt{\delta}, \qquad (4.41)\end{aligned}$$

where the last inequality follows from (4.12) and (4.14).

Similarly to (2.41), for finite time t, there exists a constant K such that

$$\mathbb{E}d(\pi_t,\bar{\pi}_t) \leq K\mathbb{E}d(\pi_0,\bar{\pi}_0).$$

Therefore, for $R - j_0 \le j \le R$, we have

$$\begin{aligned} \mathbb{E}d(\pi^n_{j\delta,R\delta}, \pi^n_{(j+1)\delta,R\delta}) &\le K\mathbb{E}d(\pi^n_{j\delta,(j+1)\delta}, \pi^n_{(j+1)\delta,(j+1)\delta}) \\ &\le Kn^{-1}, \end{aligned} \tag{4.42}$$

where the last inequality follows from (4.18).

Combining (4.41) and (4.42), we have

$$\lim_{n\to\infty} \sup_{t>0} \mathbb{E}d(\pi_t, \pi^n_t) = 0.$$

Bibliography

Bibliography

[1] Atar, R. (1998). Exponential stability for nonlinear filtering of diffusion processes in a noncompact domain. *Ann. Probab*, 26(4):1552–1574. 4, 20

[2] Atar, R. and Zeitouni, O. (1997). Exponential stability for nonlinear filtering. *Ann. Inst. H. Poincar Probab. Statist*, 33:697–725. 4, 50

[3] Budhiraja, A. and Ocone, D. (1997). Exponential stability of discrete time filters for bounded observation noise. *System and Control Letters,*, 30:185–193. 4

[4] Cox, J.C., J. I. and Ross, S. (1985). A theory of the term structure of interest rates. *Econometrica*, 53:385–407. 46

[5] Crisan, D. (2006). Partial approximation for a class of stochastic partial differential equations. *Applied mathematics and Optimization Journal*, 54(3):293–314. 4

[6] Crisan, D. and Doucet, A. (2002). A survey of convergence results on partial filtering methods for practitioners. *IEEE Trans. signal process*, 50(3):736–746. 4

[7] Crisan, D. and Heine, K. (2008). Uniform approximation of discrete time filters. *Adv. in Appl. Probab.*, 40(4):979–1001. 5

[8] Crisan, D. and Lyons, T. (1997). Nonlinear filtering and measure-valued processes. *Probab. Theory Related Fields*, 109:219–244. 4

[9] Crisan, D. and Xiong, J. (2007). A central limit type theorem for particle filter. *Comm. Stoch. Analysis*, 1(1):103–122. 5, 21, 28

[10] D. Crisan, J. G. and Lyons, T. (1998). Convergence of a branching particle mothod to the solution of the zakai equation. *SIAM J. Appl. Math.*, 58(5):1568–1590. 4

[11] D. Crisan, P. D. M. and Lyons, T. (1999). Interacting particle systems approximations of the kushner–stratonovitch equation. *Adv. Appl. Probab.*, 31(3):819–838. 4

[12] Gland, F. and Oudjane, N. (2004). Stability and uniform approximation of nonlinear filters using the hilbert metric and application to particle filters. *Ann Appl Probab*, Vol. 14(1):144–187. 20

[13] H. Carvalho, P. Del Moral, A. M. and Salut (1997). Optimal nonlinear filtering in gps/ins integration. *IEEE trans. Aerosp. Electron. Syst*, 33(3):865–850. 4

[14] Handel, R. V. (2009). The stability of conditional markov processes and markov chains in random environments. *Annals of Probability*, 37:1876–1925. 4

[15] Hasbrouck, J. (1996). Modeling market microstructure time series. *Handbook of statistics editted by G.S. Maddala and C.R. Rao, North-Holland*, pages 647–692. 6

[16] Hasbrouck, J. (2002). Stalking the "efficient price" in market microstructure specifications: an overview. *Journal of Financial Markets*, 5:329–339. 6

[17] Jacod, J. and Preotter, P. (2003). *Probability essentials*. Springer. 14

[18] Kitagawa, G. (1996). Monte-carlo filter and smoother for non–gaussian non-linear state space models. *J. Comput. Graphical. Stat.*, 5(1):1–25. 4

[19] Kouritzin, M. and Zeng, Y. (2005). Bayesian model selection via filtering for a class of micromovement models of asset price. *International Journal of Theoretical and Applied Finance*, 8:97–121. 6

[20] Kunita, H. (1971). Asymptotic behavior of the nonlinear filtering errors of markov processes. *J. Multivariate Anal.*, 1:365–393. 4

[21] Lee, K. and Zeng, Y. (2006). Risk minimization for a filtering micromovement model of asset price. *Working paper.* 7

[22] Liu, H. and Xiong, J. (2010). Branching particle system approximation of the nonlinear filter. *Submitted.* 5, 24, 27, 38, 39, 40

[23] Moral, P. D. (1996). Nonlinear filtering: interacting particle resolution. *Markov Procese and Related Fields.*, 2(4):555–581. 4

[24] Moral, P. D. (1998). A uniform convergence theorem for the numerical solving of the nonlinear filtering problem. *J. Appl. Probab.*, 4:873–884. 5

[25] Moral, P. D. and Guionnet, A. (1999). On the stability of interacting processes with applications to filtering and genetic algorithms. *Ann. Inst. H. Poincare Probab. Statist.*, 37:155–194. 5

[26] Moral, P. D. and Miclo, L. (2000). Branching and interacting particle systems approximations of feynman - kac formulae with applications to non-linear filtering. *Séminaire de Probabilités, XXXIV, Lecture Notes in Math.*, 1729:1–145. 5

[27] Morters, P. and Peres, Y. (2010). *Brownian motion.* Cambridge University Press. 94

[28] N. J. Gordon, D. J. S. and Smith, A. F. M. (1993). Novel approach to nonlinear/non-gaussian bayesian state estimation. *IEEE Proc. F.*, 140:107–113. 4

[29] P. Baxendale, P. C. and Liptser, R. (2004). Asymptotic stability of the wonham filter: Ergodic and nonergodic signals. *SIAM J. Control Optim.*, 43:643–669. 4

[30] Protter, P. E. (2004). *Stochastic integration and differential equations.* Springer. 80

[31] Rendleman, R. and Bartter, B. (1980). The pricing of options on debt securities. *Journal of Financial and Quantitative Analysis*, 15:11–24. 47

[32] Vasicek, O. (1977). An equilibrium characterisation of the term structure. *Journal of Financial Economics*, 5:177–188. 45

[33] Xiong, J. (2008). *An introduction to stochastic filtering theory*. Oxford University Press. 20, 23, 32, 34, 36, 51, 77, 78

[34] Xiong, J. and Zeng, Y. (2006). Mean-variance porfolio selection for a filtering point process model of asset price. *Working paper*. 7, 8, 65, 68, 70, 71, 74, 77

[35] Zeng, Y. (2003). A partialy observed model for micromovement of asset prices with bayes estimation via filtering. *Mathematical Finance*, 13:414–444. 6, 7, 44

Appendix

Appendix A

Important estimates for interest rate processes

Without loss of generality, we assume $a = 0$ in this section. The following tail estimate (see remark 2.20 on page 54 of [27]) plays an important role in the proofs of this section:

$$\mathbb{P}(\sup_{0\leq s\leq t} B(s) > a) \leq \frac{\sqrt{2t}}{a\sqrt{\pi}} \exp\left\{-\frac{a^2}{2t}\right\} \tag{A.1}$$

Theorem A.0.2. *Let $X(t)$ be an Ornstein-Uhlenbeck process defined by (3.8), we have the estimate for large x:*

$$\mathbb{P}\left(\inf_{0\leq s\leq b} X(s) \leq c \middle| X(b) = x\right) \leq K(c)b^{1/2}\exp\{-K'x^2\},$$

where $K(c)$ is a constant depends on c and K' is a constant.

Proof. It is well known that the solution of the SDE (3.8) can be represent conditionally (given initial value x) as:

$$X(t)e^{\theta t} = X(s)e^{\theta s} + \mu(e^{\theta t} - e^{\theta s}) + \frac{\sigma}{2\sqrt{\theta}}\tilde{B}(e^{2\theta t} - e^{2\theta s}),$$

where $\tilde{B}$ is a standard Brownian motion.

Let $K_1(c) = \frac{-2c\sqrt{\theta}}{\sigma}$ we have

$$\begin{aligned}
& \mathbb{P}\left(\inf_{0\le s\le b} X(s) \le c \middle| X(b) = x\right) \\
= \;& \mathbb{P}\left(\inf_{0\le s\le b} xe^{\theta(b-s)} + \mu(e^{\theta(b-s)} - 1) + \frac{\sigma}{2\sqrt{\theta}}\tilde{B}(e^{2\theta(b-s)} - 1) \le c\right) \\
\le \;& \mathbb{P}\left(\inf_{0\le s\le b} \tilde{B}(e^{2\theta(b-s)} - 1) \le \frac{2\sqrt{\theta}}{\sigma}(c - x)\right) \\
= \;& \mathbb{P}\left(\sup_{a\le s\le b} \tilde{B}(e^{2\theta(b-s)} - 1) \ge K_1(c) + \frac{2\sqrt{\theta}}{\sigma}x\right) \\
\le \;& \mathbb{P}\left(\sup_{0\le s\le e^{2\theta b}-1} \tilde{B}(s) \ge K_1(c) + \frac{2\sqrt{\theta}}{\sigma}x\right) \\
\le \;& \frac{\sqrt{2(e^{2\theta b} - 1)}}{(K_1(c) + \frac{2\theta}{\sigma}x)\sqrt{\pi}} \exp\left\{-\frac{(K_1(c) + \frac{2\sqrt{\theta}}{\sigma}x)^2}{2(e^{2\theta b} - 1)}\right\} \\
\le \;& K(c)b^{1/2}\exp\{-K'x^2\}.
\end{aligned}$$

□

Theorem A.0.3. *Let $X(t)$ be a CIR process which follows (3.10) with the initial state x_0, then $X(t)$ satisfies assumption 3.2.1.*

Proof.

$$\begin{aligned}
& \mathbb{P}(\inf_{0\le s\le b} X_s \le c | X(b) = x) \\
\le \;& \mathbb{P}(\sup_{0\le s\le b} |X(s) - x| \ge x - c) \\
\le \;& \mathbb{P}\left(\sup_{0\le s\le b}\left|\int_s^b \theta(\mu - X(u))du + \int_s^b \sigma\sqrt{X(u)}dB(u)\right| \ge x - c\right) \\
\le \;& \mathbb{P}\left(\sup_{0\le s\le b}\left|\int_s^b \theta(\mu - X(u))du + \int_s^b \sigma dB(u)\right| \ge \frac{x - c}{2}\right) \\
& + \mathbb{P}\left(\sup_{0\le s\le b}\left|\int_s^b \sigma(\sqrt{X(u)} - 1)dB(u)\right| \ge \frac{x - c}{2}\right) \\
\equiv \;& I_1 + I_2
\end{aligned}$$

Note that the process in I_1 is an Ornstein-Uhlenbeck process with initial value $X(s) = 0$. Similarly to the discuss in the previous theorem, we have the estimate of it as following:

$$\mathbb{P}\left(\sup_{0\leq s\leq t}\left|\int_s^b \theta(\mu - X(s))ds + \int_s^b \sigma dB(s)\right| \geq \frac{x-c}{2}\right) \leq K_1 b^{1/2}\exp\{-K_2 x^2\} \quad \text{(A.2)}$$

By Doob's inequality, we have

$$\begin{aligned}
& \mathbb{P}\left(\sup_{0\leq s\leq b}\left|\int_0^s \sigma(\sqrt{X(s)}-1)dB(s)\right| \geq \frac{x-c}{2}\right) \\
\leq & \frac{4}{(x-c)^2}\mathbb{E}\left(\left|\int_0^b \sigma(\sqrt{X(s)}-1)dB(s)\right|^2\right) \\
= & \frac{4}{(x-c)^2}\mathbb{E}\left(\int_0^b \left|\sigma(\sqrt{X(s)}-1)\right|^2 ds\right) \\
= & \frac{4\sigma^2}{(x-c)^2}\int_0^b \mathbb{E}(\sqrt{X(s)}-1)^2 ds \qquad \text{(A.3)}
\end{aligned}$$

It's known that $2a(t)X(t)$ follows a non-central chi-squared distribution with degree $\frac{4\theta\mu}{\sigma^2}$ and non-centrality parameter $2a(t)xe^{-\theta t}$, where $a(t) = \frac{2\theta}{\sigma^2(1-e^{-\theta t})}$. Therefore, $\mathbb{E}(\sqrt{X(s)}-1)^2$ is bounded. Then, we have

$$I_2 \leq K_3 b(x-c)^2 \qquad \text{(A.4)}$$

Combining (A.2) and (A.4), we can prove the assumption 5. □

Theorem A.0.4. *Let $X(t)$ be a geometric Brownian motion which is defined by (3.11), then $X(t)$ satisfies assumption 3.2.1.*

Proof. By (3.12) and $\theta < \frac{\sigma^2}{2}$, we have

$$\begin{aligned}
& \mathbb{P}\left(\inf_{0\le s\le b} X(s) \le c \middle| X(b) = x\right) \\
= \ & \mathbb{P}\left(\inf_{0\le s\le b} x\exp\left\{-(\theta - \frac{\sigma^2}{2})(b-s) + \sigma(B(b) - B(s))\right\} \le c\right) \\
\le \ & \mathbb{P}\left(\inf_{0\le s\le b} \exp\{\sigma(B(b) - B(s)\} \le cx^{-1}\right) \\
= \ & \mathbb{P}\left(\sup_{0\le s\le b} (B(s) - B(b)) \ge -\frac{1}{\sigma}\log\left\{cx^{-1}\right\}\right) \\
\le \ & K(c)b^{1/2}\exp\left\{-\frac{K'}{\sigma^2}\left\{\log\left\{cx^{-1}\right\}\right\}^2\right\}
\end{aligned}$$

Notice that

$$\lim_{x\to\infty}\exp\left\{-\frac{K'}{\sigma^2}\left\{\log\left\{cx^{-1}\right\}\right\}^2\right\} = 0,$$

this means that $X(t)$ satisfies assumption 3.2.1. □

Theorem A.0.5. *The process defined as (3.13) satisfies assumption 3.2.1.*

Proof. Let K_1 be the bound of two coefficients μ and σ.

$$\begin{aligned}
& \mathbb{P}\left(\inf_{0\le s\le b} X(s) \le c \middle| X(b) = x\right) \\
\le \ & \mathbb{P}(\sup_{0\le s\le b} |X(s) - x| \ge x - c) \\
\le \ & \mathbb{P}\left(\sup_{0\le s\le b}\left|\int_s^b \mu(X(u))du\right| + \sup_{0\le s\le b}\left|\int_s^b \sigma(X(u))dB(u)\right| \ge x - c\right) \\
\le \ & \mathbb{P}\left(K_1 b + \sup_{0\le s\le b}\left|\int_s^b \sigma(X(u))dB(u)\right| \ge x - c\right) \\
\le \ & 2\mathbb{P}\left(\sup_{0\le s\le b}\left|\int_0^s \sigma(X(u))dB(u)\right| \ge \frac{1}{2}(x - c - K_1 b)\right) \\
\le \ & \frac{8}{(x - c - K_1 b)^2}\mathbb{E}\left(\left|\int_0^b \sigma(X(u))dB(u)\right|^2\right) \\
\le \ & \frac{K_2 b}{(x - c - K_1 b)^2}
\end{aligned}$$

Therefore, assumption 3.2.1 is satisfied. □

Vita

Zhiqiang Li was born May 8, 1981 and grew up in Shijiazhuang, Hebei China. He received his Bachelor degree in Mathematics from Hebei Normal University in 2003 and Master degree in Mathematics from Hebei Normal University in 2006. While completing the Doctorate of Philosophy in Mathematics at University of Tennessee, he completed a Master of Science in Statistics through the Intercollegiate Graduate Statistics Program. He married his wife Wenfang Wang in July of 2006. In August of 2009, their son, Dylan Li, was born. In his spare time, he enjoys soccer and Chinese chess.