8-2018

Asymptotics for dynamical systems driven by jump noise

Xiaoyang Pan
University of Tennessee, xpan10@vols.utk.edu

To the Graduate Council:

I am submitting herewith a dissertation written by Xiaoyang Pan entitled "Asymptotics for dynamical systems driven by jump noise." 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.

Vasileios Maroulas, Major Professor

We have read this dissertation and recommend its acceptance:

Xia Chen, Haileab T. Hilafu, Kody J. Law, Jan Rosinski, Abner J. Salgado-Gonzalez

Accepted for the Council:

Dixie L. Thompson

Vice Provost and Dean of the Graduate School

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

Asymptotics for dynamical systems driven by jump noise

A Dissertation Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Xiaoyang Pan

August 2018

This work is dedicated to my dear parents Mingjie and Cuxiang, my brother Rubin, my beloved wife Qizhu, my lovely son Matthew Zirui.

Acknowledgments

First and foremost, I would like to express my sincerest and deepest gratitude to my advisor, Dr. Vasileios Maroulas, for his constant support, encouragement and inspiring guidance during the whole study period of PhD program in the Department of Mathematics at The University of Tennessee, Knoxville. His motivating knowledge and discussion, and the ability to conduct extensive research has played an important role in my degree. I greatly appreciate all his advice both on mathematics and personal life, which has guided me how to grow up to be a researcher and a critical thinker. All of these were and will definitely be invaluable in my work career and my life.

My acknowledgments also go to my committee members, Drs. Xia Chen, Jan Rosinski, Kody Law, Abner Salgado and Haileab Hilafu. I am very grateful for the willingness of each of them to assist me, as well as their comments and suggestions. In addition, a special thank must go to Drs. Jie Xiong and Seddik Djouadi. I would deeply appreciate Dr. Xiong's recommendation for me to US. And, the research discussion each time with Dr. Xiong was always helpful to solve problems. Also Dr. Djouadi's financial support from NSF is gratefully acknowledged.

I am thankful to all my colleagues and friends, Liguo Wang, Bo Gao, Le Yin, Kai Kang, Jin Dong, Yichen Zhang, James Ren, Ibrahim Aslan, Mustafa Elmas and Vy Nguyen for the camaraderie and the good time that we had. The interaction and daily conversations with them made the research life relaxing and will become a nice memory of my overseas study.

I would never forget the appreciation for my family in China. Without the endless support and love from my parents Mingjie Pan and Cuxiang Pan, my brother Rubin Pan, my uncles, my parents-in-law and all beloved family members, it would not have been possible to complete the PhD program and write this thesis.

Specially saving the best for the last, I know no words can express my love and gratitude to them, my dear wife, Qizhu Liang, and son, Zirui Pan. It was them who stood by me every hard time and gave me confidence, encourage and power to face the difficulties. I should not have missed every moment that I ought to stay with them. I owe them everything and wish I could show them just how much I love and appreciate them.

— Thank you all!!!

Abstract

This dissertation studies asymptotic estimates for dynamical systems with jumps. We first focus on the parameter estimation problem for a linear partially observed system. A least-squares estimator for the intensity of a Poisson process is proposed, where the signal process is driven by the mixture of a Brownian motion and a Poisson precess and the observation is a diffusion process. Precisely, we verify the unbiasedness, consistency for the estimator of the intensity. Furthermore, the asymptotic distribution and convergence rate of the consistent estimator are studied as well as a statistics for statistical inference is constructed employing the central limit theorem, large and moderate deviation principles. The last part of this dissertation is concerned with large deviation principles for the optimal filtering of a general nonlinear model. First, the uniqueness of the solution of the Zakai and Kushner-Stratonovich equations are proved, by applying a pertinent transformation of the associated equations into SDEs in an appropriate Hilbert space. Taking into account the controlled analogue of Zakai and Kushner-Stratonovich equations, respectively, the large deviation principle follows by employing some qualitative properties of their solutions using weak convergence arguments.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

A stochastic dynamical system typically describes the evolution of a phenomenon over time driven by a noise, such as Brownian motion and Poisson noise. A Brownian motion is a stochastic process depicting the path in a continuous fashion while a Poisson noise may model discontinuities or jumps of the trajectories. In particular, the partially observed system is a dynamical system that studies the hidden state or signal of interest with a related process giving the information. The estimation problems under partial information have been studied within the framework of stochastic filtering theory, for example, see this partial list for filtering models and their applications [6, 44, 56, 39, 70]. Given two stochastic processes, the signal process X_t and observation process Y_t, the filtering problem aims to estimate a functional of the unknown signal, $F(X_t)$, for some pertinent function F, based on the observed information up to time t. The best estimate in the sense of least squares error for $F(X_t)$ is given by the conditional expectation, $\mathbb{E}(F(X_t)|\mathcal{F}_t^Y)$, where $\mathcal{F}_t^Y$ is the σ-field generated by $\{Y_s : 0 \leq s \leq t\}$ [77].

Many works have been devoted to this subject. For example, for systems with a continuous driving noise, the pioneering works include, e.g., Kushner [49], Zakai [82], Kallianpur and Striebel [42], and the extended works [5, 41, 68, 47, 77, 80]. Over the last decade, systems with jump noise have gotten more attention. In [83] the filtering model where the observation process is a counting process was proposed to describe micromovement of asset prices, and a branching particle approximation for such a filtering problem was proposed in [78]. The paper [62] derived filtering equations for jump-diffusion models where the hidden signal was driven by a mixture of Brownian motion and Poisson random measure, while the multidimensional system was considered in [63]. The model in which signal is a diffusion and observation contains Gaussian and jump noise was studied in [54]. The optimal filter for a jump-diffusion model was treated by Bayesian algorithms in [19, 39]. Those work [13, 14] studied the model where signal and observation have common jump times. The recent work [66] verified the existence of the filtering for a more general nonlinear system. At the same time, applications in finance [19, 39, 58, 78, 83], engineering [70, 60, 32, 74], biology [33], environment pollution [25], etc., have also been devoted to using stochastic dynamics driven by a pertinent jump noise.

However, estimation of the parameters associated with the hidden signal are oftentimes the primary goal of a study. For instance, as an application in engineering, identification of stochastic dynamical systems deals with the problem of building a mathematical model from recorded observation data generated by the system. Given the recorded data and a class of models, system identification amounts to selecting the model parameters which best describe the data, see [15]. Several studies have been devoted to the estimation of relevant parameters related to a jump-diffusion process. The early work [1] considered the generalized Itô process type containing both jump and diffusion components for a fully-observed system. A simple counterexample was employed in order to verify that, the common approach of estimating parameters of an Itô process by applying maximum likelihood to a discretization of the SDE does not yield a consistent estimator. The work in [18] demonstrated estimates of stochastic volatility jump-diffusion process using a simulation-based estimator. The study in [71] proposed a maximum likelihood estimator from discrete observations for diffusion processes with jumps under restrictive assumptions and established consistency and asymptotic normality. The consistency, rate of convergence of least-squares estimator as well as asymptotic normality were established in [51] for Ornstein-Uhlenbeck processes with small Lévy noise. The work in [59] presented asymptotic normality and convergence of moments of diffusion process with jumps using a quasi-maximum likelihood estimator.

Chapter 2 examines the consistency and asymptotic properties of a least-squares estimator for the intensity of the driving Poisson noise in a partially observed process. Unlike the problems under full information, inference for partially observed jump-diffusion processes is more complex and challenging due to the fact that such distributions typically do not enjoy a closed form. To that end, the typical maximum likelihood estimation (MLE) method for diffusion cases (e.g. [20, 15] and the references therein), which involves a conditional expectation of the density, is rarely tractable and thus need to be estimated. Indeed, very few studies have investigated the estimation of Poisson intensity for partially observed jump-diffusion systems. The paper [39] studied a discretized model by particle filter and it is well known the simulation method is sensitive to the choice of proposal density, and also there was no study for convergence properties. The work in [72] investigated the MLE for the intensity of a counting process using expectation maximization (EM) algorithm, where the intensity is restricted to a pertinent dynamic process and relies on the observations, which turned to be a filtering problem.

To construct a statistics for inference, we usually make use of the asymptotic distribution that deals with the tail probabilities. The central limit theorem with a normal distribution oftentimes is the first priority to be used. However, in some practical problems, the central limit theorem yields that the asymptotic variance depends on the true parameter, see [30]. In this case, other types of asymptotic estimates may be considered, for example, large and moderate deviation principles. Large deviation principle (LDP) is a theory that investigates the probability of rare events that decay exponentially fast to zero. The decay scale is described by a rate function. One may refer to [21, 23], for detailed discussion in large deviation theory. In contrast, the moderate deviation principle (MDP) considers an intermediate scale of convergence, that is, a scale that bridges large deviation principle and central limit theorem.

The large deviation and moderate deviation results of estimators provide us with the rates of convergence and a useful method for constructing a statistics. The approach to statistical problems that bases on the study of probabilities of large deviations has been used in statistical inference since the papers by Chernoff [17] and Bahadur [4]. For some developments of large deviations and moderate deviations in statistics, see [27, 46, 36, 65, 3, 31] for large deviations of estimators; [40, 28, 24, 30] for moderate deviations of estimators;

[52, 75, 29, 50] for large deviations and moderate deviations of kernel density estimators, and references therein.

In Chapter 3, we study the convergence rate of the consistent LSE by examining large and moderate deviation principles. As discussed above, large deviation principle provides approaches to investigate the exponential rate of decay for a sequence of random variables, so is the moderate deviation principle. With the deviation principles at hand, constructing a statistics based on the asymptotic distribution enable us to study the hypothesis testing and conference intervals, taking advantage of the convergence speed and rate function.

Chapter 4 focuses on a rather general dynamical system for the asymptotic behavior of the nonlinear filter with jumps in both signal and observation processes, and investigates a large deviation principle of the optimal filter, which is associated with the rare events of a small signal-to-noise ratio. As mentioned above, the best estimate for the hidden process, optimal filter, is defined as the conditional expectation given the observation. The optimal filter is a solution of a stochastic partial differential equation, the Kushner-Stratonovich equation. Relying on a Bayesian transformation, the unnormalized filter, given by the so-called Zakai equation, plays a key role on the derivation of the optimal filter.

Several studies have been concerned with the convergence rate of the optimal filter perturbed by a small parameter. The early work [35] derived large deviations for the conditional density for diffusion systems in which both the signal and observation are diffusions and their noises were small. Discussion in [38] later extended it by allowing random initial value and connected it to a control problem. The study [61] established a quenched large deviation principle with small-noise observation. In a similar setup as herein, [76] took aim at a model where the signal and observation processes were driven by Brownian motions, whereas [57] considered a memory observation process driven by a fractional Brownian motion.

The strategy we apply is to prove a tantamount argument to the large deviation principle, the Laplace principle, using a weak convergence fashion as proposed in [11, 12, 55]. Weak convergence method is an approach that has been increasingly used, e.g., the large deviations for a variety of SPDEs [7, 11, 22, 67, 79, 81], based on variational representations of the functionals of driving Brownian motions and counting random measures. The novelty of such a method is that, it does not require the exponential continuity or exponential tightness, and in contrast only basic qualitative properties of existence, uniqueness of controlled analogues of the stochastic dynamical systems of interest need to be shown.

To that end, we first prove the uniqueness of the controlled unnormalized filtering equation, i.e. the controlled Zakai equation, and subsequently this of the controlled filtering equation, the controlled Kushner-Stratonovich equation. Some studies have been devoted to verifying the uniqueness of the filtering equation. In [14], the uniqueness was shown by the Filtered Martingale Problem approach which was proposed in [47], however, that model has a limitation that the signal and observation are driven by the same Poisson random measure having common jump times. A more general setting was suggested in [66] and the uniqueness was also proved by the approach of Filtered Martingale Problem, however, it was shown under the assumption that the correlated Poisson random measure is independent of the signal. Moreover, in these studies the uniqueness of the Filtered Martingale Problem was assumed when the Filtered Martingale Problem method was used, which requires the regularity conditions of the coefficients of the equations. In our work, based on a method using the Brownian motion semigroup, we bypass these restrictions and establish the uniqueness for the general model with a mild assumption on the coefficients of the Poissonian noise.

The dissertation yielded three manuscripts is organized in the same way. Chapter 2 focuses on a linear partially observed jump diffusion. A least-squares estimator for the intensity of the Poisson process is proposed. Consistency is verified as well as asymptotic normality when the drift coefficient in the signal process is negative. For other case, we show that moments up to fourth order of the least-squares estimator are bounded but inconsistent. Simulation demonstrating the results is implemented. In Chapter 3 we study inferential problems for the intensity of the Poisson process. Large and moderate deviation principles are studied to describe the convergence behavior of the estimator. The inferential results, the power of hypothesis testing and confidence intervals, are presented in Section 3.3 by proving a moderate deviation principle Chapter 4 studies the general dynamical system with presence of Lévy noises in both signal and observation processes. The uniform large deviation principle is established using the weak convergent argument. To apply such a strategy, we prove the uniqueness of the solution to Zakai and Kushner-Stratonovich equations. Then the large deviation for the optimal filter is verified using the qualitative properties–existence, uniqueness and tightness of an controlled analogue of the equation.

Main notations

- $(\Omega, \mathcal{F}, \mathbb{P})$: probability space
- $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t\in[0,T]})$: filtered probability space
- $\mathcal{F}_t^Y$: σ-field generated by $\{Y_s : 0 \leq s \leq t\}$
- $\mathbb{R}^m$: m-dimensional vector space over real number $\mathbb{R}$
- $\mathcal{G}$: sub-σ-field of $\mathcal{F}$
- $\mathbb{U}$: locally compact Polish space
- $\mathcal{E}$: Polish space
- $\mathcal{B}(\mathbb{U})$: Borel σ-field on $\mathbb{U}$
- $\mathcal{M}(\mathbb{U})$: space of all measures ν on $(\mathbb{U}, \mathcal{B}(\mathbb{U}))$
- $\mathbb{U}_T$: $[0, T] \times \mathbb{U}$
- $\mathbb{M}$: $\mathcal{M}(\mathbb{U}_T)$
- $\mathcal{P}_2$: L^2 Hilbert space containing the control function for a Brownian motion
- $\bar{\mathcal{A}}$ be the class of all measurable maps $\phi : \mathbb{U}_T \times \bar{\mathbb{M}} \to [0, \infty)$.
- $\tilde{S}^M$: $\{\psi \in L^2([0, T], \mathbb{R}^m) : \int_0^T |\psi(s)|^2 ds \leq M\}$
- S^M: $\{\phi : \mathbb{U}_T \to [0, \infty) : L_T(\phi) = \int_{\mathbb{U}_T} l(\phi(t, u))\ \nu(du)ds \leq M\}$,
- $\mathcal{U}^M$: $\{(\psi, \phi) : \tilde{L}_T(\psi) \leq M, L_T(\phi) \leq M\}$
- $C([0, T], \mathbb{R}^m)$: space of all continuous functions from $[0, T]$ to $\mathbb{R}^m$
- $D([0, T], \mathcal{E})$: space of right continuous functions with left limits from $[0, T]$ to $\mathcal{E}$.
- $\mathbb{W}_m$: $C([0, T], \mathbb{R}^m)$
- $\mathbb{V}$: $\mathbb{W}_m \times \mathbb{W}_n \times \mathbb{M}$
- $\bar{\mathbb{M}}$: $\mathcal{M}(\mathbb{Y}_T)$
- λ_T: Lebesgue measure on $[0, T]$
- $\mathbb{Y}_T$: $[0, T] \times \mathbb{Y}$

- $\bar{\mathbb{V}}$: $\mathbb{W}_m \times \mathbb{W}_n \times \bar{\mathbb{M}}$
- $N^{\phi}_{\lambda^\epsilon}(dt, du)$: $\tilde{\mathbb{P}}^\epsilon$-Poisson random measure with compensator $\phi dt \nu_2(du)$
- H_0: Hilbert space consisting of square-integrable functions on $\mathbb{R}^d$
- $\|\phi\|_0^2$: L^2-norm for $\phi \in H_0$
- $\langle \phi, \psi \rangle_0$: inner product for $\phi, \psi \in H_0$
- $\mathcal{M}_G(\mathbb{R}^d)$: space of finite signed measures on $\mathbb{R}^d$
- $\mathcal{M}_F(\mathbb{R}^d)$: collection of all finite Borel measures on $\mathbb{R}^d$
- $\mathcal{P}(\mathbb{R}^d)$: space of probability measures on $\mathbb{R}^d$
- $\langle \nu, f \rangle$: $\int f(x)\nu(dx)$, for each measure ν
- N_p: $\mathbb{P}$-Poisson random measure with compensator $t\nu_1(du)$
- N_λ: $\mathbb{P}$-integer-valued random measure with predictable compensator $\lambda(t, X_{t-}, u)t\nu_2(du)$
- $\|\cdot\|_{\mathbb{U}}$: norm on the measurable space $(\mathbb{U}, \mathcal{B}(\mathbb{U}))$
- ν_i: measures on $\mathbb{U}$ with $\nu_i(A) < \infty, A \in \mathcal{B}(\mathbb{U})$ and $\int_{\mathbb{U}_i} \|u\|^2_{\mathbb{U}} \nu_i(du) < \infty, \mathbb{U}_i \subset \mathbb{U}, i = 1, 2.$
- $\mathcal{L}$: infinitesimal generator of a function of the signal process (4.3.1a)
- $\mathcal{D}(\mathcal{L})$: domain of infinitesimal generator $\mathcal{L}$
- $C_b^2(\mathbb{R}^d)$: collection of bounded functions with continuous 2nd-order derivatives.
- $\mathcal{G}^\epsilon(\mu_0, \sqrt{\epsilon}\tilde{W}^\epsilon, \epsilon N^{\epsilon^{-1}}_{\lambda^\epsilon})$: solution of small noise Zakai Eq. (4.4.3)
- $\mathcal{G}^\epsilon(\mu_0, \sqrt{\epsilon}\tilde{W}^\epsilon + \int_0^\cdot \psi(s)ds, \epsilon N^{\epsilon^{-1}\phi}_{\lambda^\epsilon})$: solution of controlled small-noise Zakai Eq. (4.4.5)
- $\mathcal{G}^0$: solution of noise-free Zakai Eq. (4.4.6)

Chapter 2

LSE for Poisson intensity

2.1 Introduction

In this chapter we study the jump diffusion with a linear setting and focus on the estimation of the intensity of a Poisson process under partial information. In Section 2.2 we propose and verify the least-squares estimator of the Poisson intensity. Section 2.3.1 establishes the consistency of the least-squares estimator and the asymptotic normality when data are streamed online. Asymptotic properties in the case of data collection within a fixed time horizon are presented in Section 2.3.2. Last, Section 2.4 presents some simulation results which visualize the theoretical landscape of this chapter.

2.2 Models and least-squares estimator for the intensity

Let $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t\in[0,T]})$ be a filtered probability space. A latent jump-diffusion signal process, X_t, is considered partially observed by data Y_t expressed by a diffusion process related to X_t. The stochastic model is given below

$$\begin{aligned} dX_t &= aX_t dt + bdW_t + \delta dN_t, \quad X_0 = x_0 \in \mathbb{R} \\ dY_t &= hX_t dt + dB_t, \quad Y_0 = 0, \quad 0 \le t \le T, \end{aligned} \tag{2.2.1}$$

where W_t, B_t are independent standard Brownian motions, and N_t is a Poisson process with intensity λ on the stochastic basis $(\Omega, \mathcal{F}, P, (\mathcal{F}_t)_{t\in[0,T]})$ which is independent of W_t and B_t. We assume that Y_t is observable at discrete times $\{t_0^n, t_1^n, \cdots t_n^n\}$.

Let Λ denote the parameter space of the intensity of the driving Poisson process. Our goal is to estimate the intensity, $\lambda \in \Lambda$, using a least-squares estimator, denoted herein by $\hat{\lambda}_n$, as a function of the other associated parameters assuming they are known. This assumption is well justified in several applications, e.g. in biology [56] or defense [53]. Employing Itô's formula [34], we obtain the solution of signal process

$$X_t = e^{at}\left(x_0 + b\int_0^t e^{-as}dW_s + \delta\int_0^t e^{-as}dN_s\right). \tag{2.2.2}$$

Consider the compensated Poisson process, $\hat{N}_t = N_t - \lambda t$, which is a martingale. Therefore the expectation satisfies $\mathbb{E}(\int_0^t e^{-as}dN_s) = \lambda\int_0^t e^{-as}ds$. Define the innovation process

$$Z_{t_i^n} := Y_{t_i^n} - Y_{t_{i-1}^n} = h\int_{t_{i-1}^n}^{t_i^n} X_s ds + \varepsilon_i,$$

where ε_i denote the normally distributed and independent Brownian increments $B_{t_i^n} - B_{t_{i-1}^n} \sim$ i.i.d. $N(0, \Delta t_{i-1}^n)$, $i = 1, ..., n$. Thus, denoting the time difference between two observations by $\Delta t_{i-1}^n = t_i^n - t_{i-1}^n$, the expectation of $Z_{t_i^n}$ is given by

$$\mathbb{E}(Z_{t_i^n}) = (\frac{hx_0}{a} + \frac{\delta\lambda h}{a^2})(e^{at_i^n} - e^{at_{i-1}^n}) - \frac{\lambda\delta h}{a}\Delta t_{i-1}^n. \tag{2.2.3}$$

Denote the residual sum of squares (RSS) by

$$Q(\lambda) = \sum_{i=1}^n (Z_{t_i^n} - \mathbb{E}(Z_{t_i^n}))^2. \tag{2.2.4}$$

Then, the least-square estimator $\hat{\lambda}_n$ of λ is to find the intensity which minimizes the RSS. That is,

$$\hat{\lambda}_n = \arg\min_{\lambda\in\Lambda} Q(\lambda).$$

Taking into account Eq. (2.2.3) and the RSS given in Eq. (2.2.4), we have that

$$Q(\lambda)=\sum_{i=1}^{n}\left(Z_{t_i^n}-(\frac{hx_0}{a}+\frac{\delta\lambda h}{a^2})(e^{at_i^n}-e^{at_{i-1}^n})+\frac{\lambda\delta h}{a}\Delta t_{i-1}^n\right)^2.$$

The associated derivative of $Q(\lambda)$ with respect to λ can be easily verified that

$$\begin{aligned} Q'(\lambda)= \quad & -2\sum_{i=1}^{n}\left(Z_{t_i^n}-(\frac{hx_0}{a}+\frac{\delta\lambda h}{a^2})(e^{at_i^n}-e^{at_{i-1}^n})+\frac{\lambda\delta h}{a}\Delta t_{i-1}^n\right) \\ & \times\left(\frac{\delta h}{a^2}(e^{at_i^n}-e^{at_{i-1}^n})-\frac{\delta h}{a}\Delta t_{i-1}^n\right). \end{aligned}$$

Setting $Q'(\lambda)=0$, the least-squares estimator is given by

$$\hat{\lambda}_n=\frac{\sum_{i=1}^{n}\left(Z_{t_i^n}-\frac{hx_0}{a}(e^{at_i^n}-e^{at_{i-1}^n})\right)\left(\frac{1}{a}(e^{at_i^n}-e^{at_{i-1}^n})-\Delta t_{i-1}^n\right)}{\frac{\delta h}{a}\sum_{i=1}^{n}\left(\frac{1}{a}(e^{at_i^n}-e^{at_{i-1}^n})-\Delta t_{i-1}^n\right)^2}. \tag{2.2.5}$$

The second order derivative of $Q(\lambda)$ with respect to λ is

$$Q''(\lambda)=\left(\frac{\delta h}{a^2}(e^{at_i^n}-e^{at_{i-1}^n})-\frac{\delta h}{a}\Delta t_{i-1}^n\right)^2>0,$$

and thus indeed $\hat{\lambda}_n$ is the maximizing argument of the RSS as expressed in Eq. (2.2.4).

2.3 Main results

2.3.1 Consistency and asymptotic normality

In this section we establish the consistency and verify the asymptotic normality of the least-squares estimator $\hat{\lambda}_n$ defined in Eq. (2.2.5) when data are collected *ad infinitum* and the drift coefficient $a<0$. The case where $a>0$ does not make sense because the trajectory of X_t stays bounded a.s. only if the drift coefficient is negative. Precisely, we take into account that $t_i^n=i,\ i=1,...,n$ (and $n\to\infty,$ or $T\to\infty$). Let $d_i=e^{ai}-e^{ai-a}$, Eq. (2.2.5) yields that the least-squares estimator is

$$\hat{\lambda}_n = \frac{1}{\frac{h\delta}{a}\sum_{i=1}^{n}(\frac{1}{a}d_i - 1)^2}\sum_{i=1}^{n}\left(Z_i - \frac{hx_0}{a}d_i\right)\left(\frac{1}{a}d_i - 1\right). \tag{2.3.1}$$

Lemma 2.1. *Suppose that we receive data ad infinitum ($T \to \infty$). Then $\hat{\lambda}_n$ is unbiased. Furthermore, Define*

$$\tilde{\lambda}_n := \frac{|a|}{n\delta}\int_0^n X_s ds + \frac{|a|}{hn\delta}\sum_{i=1}^{n}\varepsilon_i. \tag{2.3.2}$$

Then,

$$\hat{\lambda}_n = \frac{n}{n+k_n}\left(\tilde{\lambda}_n + \frac{1}{n}\xi_n\right) + \frac{\ell_n}{n}, \tag{2.3.3}$$

where the sequence of (k_n, ℓ_n) converges to (k, ℓ), and the random sequence ξ_n converges to a finite random variable ξ almost surely.

Proof. Substituting the expectation of Z_i into Eq. (3.2.1) gives the unbiasedness. Note that

$$\sum_{i=1}^{n}(\frac{1}{a}d_i - 1)^2 = n + k_n,$$

where $k_n \to \frac{2}{a} - \frac{e^a-1}{a^2(e^a+1)}$. As $Z_i = h\int_{i-1}^{i} X_s ds + \epsilon_i$, the numerator in Eq. (3.2.1) can be rewritten as

$$\sum_{i=1}^{n} Z_i\left(\frac{1}{a}d_i - 1\right) - \frac{hx_0}{a}\sum_{i=1}^{n} d_i\left(\frac{1}{a}d_i - 1\right) = A - B,$$

where

$$A = \frac{1}{a}\sum_{i=1}^{n} e^{ai}(1-e^{-a})Z_i - (h\int_0^n X_s ds + \sum_{i=1}^{n}\epsilon_i),$$

in which $\frac{1}{a}\sum_{i=1}^{n} e^{ai}(1-e^{-a})Z_i < \infty$ a.s. and

$$B \to \frac{hx_0}{a^2(e^a+1)}\left(a(e^a+1) - (e^a - 1)\right) < \infty.$$

Next, by denoting

$$\xi_n = \frac{(1-e^{-a})}{h\delta}\sum_{i=1}^{n} e^{ai}Z_i, \quad \ell_n = -\frac{an}{h\delta(n+k_n)}B,$$

the conclusion then follows. ■

Theorem 2.2. *The least-squares estimator $\hat{\lambda}_n$ is consistent.*

Proof. Plugging that

$$\mathbb{E}(X_t) = e^{at}x_0 + \lambda\delta \int_0^t e^{a(t-s)}ds$$

into Eq. (2.3.2), we verify that

$$\mathbb{E}\left(\frac{|a|}{n\delta}\int_0^n X_s ds\right) = \lambda - \frac{1}{an\delta}(ax_0e^{an} + \lambda\delta e^{an} - ax_0 - \lambda\delta) \to \lambda, \text{ as } n \to \infty. \tag{2.3.4}$$

Therefore, $\hat{\lambda}_n$ is an asymptotically unbiased estimator for the intensity λ. Since ε_i are standard normal independent random variables, one deduces that

$$\mathrm{Var}\left(\frac{1}{hn\delta}\sum_{i=1}^n \varepsilon_i\right) = \frac{1}{nh^2\delta^2} \to 0, \text{ as } n \to \infty.$$

Note that ε_i is independent of X_t, thus it is sufficient to show that $\frac{|a|}{n\delta}\int_0^n X_s ds \to \lambda$ in probability as $n \to \infty$. For $s < t$,

$$\mathbb{E}\left(\frac{|a|}{\delta n}\int_0^n X_s ds\right)^2 = 2\mathbb{E}\left(\frac{a^2}{\delta^2 n^2}\int_0^n\int_0^t X_t X_s ds dt\right).$$

Applying Itô's formula, the "covariance" between two different time instants for the signal, $\mathbb{E}(X_tX_s)$, is given by

$$\begin{aligned} &e^{at+as}\mathbb{E}\left[\left(x_0 + b\int_0^t e^{-ar}dW_r + \delta\int_0^t e^{-ar}dN_r\right)\right.\\ &\qquad\left.\times\left(x_0 + b\int_0^s e^{-ar}dW_r + \delta\int_0^s e^{-ar}dN_r\right)\right]\\ &= e^{at+as}\left[x_0^2 + b^2\int_0^s e^{-2ar}dr + x_0\delta\mathbb{E}\left(\int_0^s e^{-ar}dN_r + \int_0^t e^{-ar}dN_r\right)\right.\\ &\qquad\left.+\delta^2\mathbb{E}\left(\int_0^t e^{-ar}dN_r \cdot \int_0^s e^{-ar}dN_r\right)\right]. \end{aligned} \tag{2.3.5}$$

Note that

$$\mathbb{E}\Big(\int_0^s e^{-ar}dN_r + \int_0^t e^{-ar}dN_r\Big) = \lambda\int_0^s e^{-ar}dr + \lambda\int_0^t e^{-ar}dr, \tag{2.3.6}$$

and

$$
\begin{aligned}
&\mathbb{E}\big(\int_0^t e^{-ar}dN_r \cdot \int_0^s e^{-ar}dN_r\big)\\
=&\mathbb{E}\big[\big(\int_0^t e^{-ar}d\hat{N}_r + \lambda\int_0^t e^{-ar}dr\big)\big(\int_0^s e^{-ar}d\hat{N}_r + \lambda\int_0^s e^{-ar}dr\big)\big]\\
=&\lambda\int_0^s e^{-2ar}dr + \lambda^2\int_0^t e^{-ar}dr \cdot \int_0^s e^{-ar}dr. \qquad (2.3.7)
\end{aligned}
$$

Combining Eqs. (2.3.5), (2.3.6) and (2.3.7) gives $\mathbb{E}\left(\frac{a^2}{\delta^2 n^2}\int_0^n\int_0^t X_t X_s ds dt\right) \to \frac{\lambda^2}{2}$. With this limit and Eq. (2.3.4) in hand, we obtain that $\mathrm{Var}(\frac{|a|}{n\delta}\int_0^n X_s ds)$ converges to 0. Employing Chebychev's inequality, for all $\varepsilon > 0$

$$
\mathbb{P}\left(|\frac{|a|}{n\delta}\int_0^n X_s ds - \lambda| \geq \varepsilon\right) \leq \frac{\mathrm{Var}(\frac{|a|}{n\delta}\int_0^n X_s ds)}{\varepsilon^2} \to 0, \quad \text{as } n \to \infty.
$$

The conclusion then follows from Eq. (2.3.3). ■

Next, we establish the asymptotic normality of the least-squares estimator, $\hat{\lambda}_n$ by verifying the Central Limit Theorem (CLT). Here, since the estimator can be decomposed into a random sequence which is not identical but independent, the Lyapunov Central Limit Theorem (Theorem 2.3 below) will be invoked in the proof of our result. The reader may refer to [8] for its proof.

Theorem 2.3. (Lyapunov CLT) *Suppose $X_1, X_2 \ldots$ is a sequence of independent random variables, each with finite expected value μ_i and variance σ_i^2. Define $S_n^2 = \sum_{i=1}^n \sigma_i^2$. If for some $\delta > 0$, the Lyapunov's condition*

$$
\lim_{n\to\infty} \frac{1}{S_n^{2+\delta}} \sum_{i=1}^n \mathbb{E}\left(|X_i - \mu_i|^{2+\delta}\right) = 0
$$

is satisfied, then

$$
\frac{1}{S_n}\sum_{i=1}^n (X_i - \mu_i) \xrightarrow{L} N(0,1), \quad \text{as } n \to \infty
$$

where $\xrightarrow{L}$ denotes convergence in law.

Before proceeding with the demonstration of the CLT for the least-squares estimator, we state a lemma which will be also employed in the proof.

Lemma 2.4. *Assume that X, Y are integrable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. If $\mathbb{E}(|X|^r) < \infty$ and $\mathbb{E}(|Y|^r) < \infty$ for some $r > 0$, then*

$$\mathbb{E}\left(|X+Y|^r\right) \leq \mathbb{E}\left(|X|^r\right) + \mathbb{E}\left(|Y|^r\right), \qquad \text{if } 0 \leq r < 1;$$
$$\mathbb{E}\left(|X+Y|^r\right) \leq 2^{r-1}\mathbb{E}\left(|X|^r\right) + 2^{r-1}\mathbb{E}\left(|Y|^r\right), \qquad \text{if } r \geq 1.$$

Theorem 2.5. *Consider the partially observed signal process X_t and the associated ad infinitum collection of data Y_t as described in Eq. (2.2.1). Then the least-squares estimator $\hat{\lambda}_n$ in Eq. (3.2.1) satisfies the following CLT*

$$\sqrt{n}\left(\hat{\lambda}_n - \lambda\right) \xrightarrow{L} N\left(0, \frac{b^2}{\delta^2} + \lambda + \frac{a^2}{h^2\delta^2}\right), \qquad \text{as } n \to \infty. \tag{2.3.8}$$

Proof. Based on the signal process of X_t, we have that

$$|a| \int_0^t X_s ds = bW_t + \delta N_t - X_t + x_0. \tag{2.3.9}$$

In turn, Eq. (2.3.9) yields that

$$\begin{aligned}\sqrt{n}(\tilde{\lambda}_n - \lambda) = &\frac{b}{\delta\sqrt{n}} \int_0^n (1 - e^{an-as}) dW_s + \frac{1}{\sqrt{n}} \int_0^n (1 - e^{an-as}) dN_s \\ &+ \frac{|a|}{h\delta\sqrt{n}} \sum_{i=1}^n \varepsilon_i - \frac{1}{\sqrt{n}\delta}(e^{an}x_0 - x_0) - \sqrt{n}\lambda.\end{aligned}$$

Let us first consider the Poissonian term and denote $A_i = \int_{i-1}^i (1 - e^{an-as}) dN_s$. Then $\int_0^n (1 - e^{an-as}) dN_s = \sum_{i=1}^n A_i$. Since N_t is a Poisson process with independent increments, $\{A_i\}$ is a sequence of independent random variables for $1 \leq i \leq n$. Let $\lambda S_n^2 := \sum_{i=1}^n \mathrm{Var}(A_i)$.

$$S_n^2 = \frac{1}{2a}(2an - 4e^{an} + e^{2an} + 3) = O(n). \tag{2.3.10}$$

Let $\hat{N}_t = N_t - \lambda t$ be the compensated process of N_t. Using Lemma 2.4, we have

$$\begin{aligned}\sum_{i=1}^{n} \mathbb{E}\left(|A_i|^4\right) &= \sum_{i=1}^{n} \mathbb{E}\left|\int_{i-1}^{i}(1-e^{an-as})d\hat{N}_s + \lambda\int_{i-1}^{i}(1-e^{an-as})ds\right|^4 \\ &\le \frac{1}{8}\sum_{i=1}^{n}\mathbb{E}\left(|\int_{i-1}^{i}(1-e^{an-as})d\hat{N}_s|^4\right) + \frac{1}{8}\sum_{i=1}^{n}\left(|\lambda\int_{i-1}^{i}(1-e^{an-as})ds|^4\right).\end{aligned}$$

By Lemma B.5, there exists a proper constant C_p such that

$$\begin{aligned}\sum_{i=1}^{n} \mathbb{E}\left(|A_i|^4\right) &\le \frac{C_p}{8}\sum_{i=1}^{n}\mathbb{E}\left|\int_{i-1}^{i}(1-e^{an-as})^2 dN_s\right|^2 + \frac{\lambda^4 n}{8} \\ &\le \frac{C_p}{8}\sum_{i=1}^{n}\mathbb{E}\left(|N_i - N_{i-1}|^2\right) + \frac{\lambda^4 n}{8} \\ &= \frac{C_p(\lambda+\lambda^2)n}{8} + \frac{\lambda^4 n}{8}.\end{aligned}$$

Thus, $\sum_{i=1}^{n}\mathbb{E}(|A_i|^4) = O(n)$. By

$$\sum_{i=1}^{n}\mathbb{E}\left(|A_i - \mathbb{E}(A_i)|^4\right) \le \frac{1}{8}\sum_{i=1}^{n}\mathbb{E}(|A_i|^4) + \frac{1}{8}\sum_{i=1}^{n}|\mathbb{E}(A_i)|^4.$$

Therefore $\sum_{i=1}^{n}\mathbb{E}\left(|A_i - \mathbb{E}(A_i)|^4\right) = O(n)$ by Jensen's inequality, and we obtain the Lyapunov's condition in Theorem 2.3 with $\delta = 2$. Hence by Theorem 2.3, we have

$$\frac{1}{\sqrt{\lambda}S_n}\left(\int_0^n (1-e^{an-as})dN_s - \lambda\int_0^n(1-e^{an-as})ds\right) \xrightarrow{L} N(0,1). \tag{2.3.11}$$

On the other hand, one can get the normality according to the definition of Itô integral that

$$\frac{1}{S_n}\int_0^n (1-e^{an-as})dW_s \xrightarrow{L} N(0,1). \tag{2.3.12}$$

Now the results in (2.3.11) and (2.3.12) together with the independence of W_t and N_t provide that

$$\begin{aligned}
&\frac{\sqrt{n}}{S_n}\left(\frac{b}{\delta\sqrt{n}}\int_0^n \left(1-e^{an-as}\right) dW_s \right.\\
&\qquad\qquad \left. +\frac{1}{\sqrt{n}}\int_0^n \left(1-e^{an-as}\right) dN_s\right) - \frac{\lambda}{S_n}\int_0^n \left(1-e^{an-as}\right) ds \\
=&\frac{n}{S_n}(\tilde{\lambda}_n-\lambda) - \frac{|a|}{h\delta S_n}\sum_{i=1}^n \varepsilon_i + \frac{1}{\delta S_n}\left(e^{an}x_0 - x_0\right) - \frac{1-e^{an}}{aS_n}\lambda \xrightarrow{L} N(0, \frac{b^2}{\delta^2}+\lambda).
\end{aligned} \tag{2.3.13}$$

Moreover, since $\epsilon \sim N(0,1)$ are i.i.d.,

$$\frac{|a|}{h\delta S_n}\sum_{i=1}^n \varepsilon_i \xrightarrow{L} N(0, \frac{a^2}{h^2\delta^2}).$$

Therefore, Eq. (2.3.8) follows from Eqs. (2.3.13) and (2.3.3). ■

2.3.2 Properties of the least-squares estimator for a fixed time horizon

This section focuses on properties of the estimator $\hat{\lambda}_n$ when data are collected within a fixed time horizon at discrete time instances $t_i^n = i/n$. The drift coefficient is considered any non-zero number ($a \neq 0$) since horizon is finite and the solution of X_t is always bounded almost surely. To what it follows, without loss of generality we adopt that $a = 1$. Note that $\varepsilon_i = B_{t_i^n} - B_{t_{i-1}^n} \sim$ i.i.d. $N(0, \frac{1}{n})$ and thus

$$\mathbb{E}\left(Z_{t_i^n}\right) = hx_0(\frac{1}{n}+D_i) + h\delta\lambda D_i,$$

where $D_i = e^{\frac{i}{n}} - e^{\frac{i-1}{n}} - \frac{1}{n}$. The least-squares estimator, $\hat{\lambda}_n$ is given by

$$\hat{\lambda}_n = \frac{\sum_{i=1}^n D_i Z_{t_i^n} - hx_0\sum_{i=1}^n (\frac{1}{n}+D_i)D_i}{h\delta\sum_{i=1}^n D_i^2}. \tag{2.3.14}$$

Lemma 2.6. *Consider the partially observed jump diffusion X_t of Eq. (2.2.1), where data Y_t are collected at discrete times $t_i^n = i/n, i = 1, ..., n$. Then the least-squares estimator, $\hat{\lambda}_n$, is an unbiased estimator of the intensity λ. Furthermore,*

$$\hat{\lambda}_n = p_n \tilde{\lambda}_n + q_n + o(1), \tag{2.3.15}$$

where

$$\tilde{\lambda}_n := \frac{1}{\delta} \sum_{i=1}^{n} (e^{\frac{i}{n}} - 1) \int_{t_{i-1}^n}^{t_i^n} X_s ds + \frac{1}{h\delta} \sum_{i=1}^{n} (e^{\frac{i}{n}} - 1)\varepsilon_i, \tag{2.3.16}$$

and real sequences of (p_n, q_n) converges to (p, q). Here, $o(1)$ means a random sequence converging to 0 in probability.

Proof. The unbiasedness follows from substituting $\mathbb{E}(Z_{t_i^n})$ into Eq. (2.3.14). By Taylor expansions for $e^{\frac{1}{n}}$ and $e^{-\frac{1}{n}}$, we get

$$\begin{aligned} \sum_{i=1}^{n} D_i^2 =& \frac{1}{n(e^{\frac{1}{n}} + 1)} \left(-e^{\frac{1}{n}}(n-3) - 2e^{\frac{1}{n}+1} + ne^{\frac{1}{n}+2} - e^2 n + n - 2e + 3 \right) \\ =& \frac{e^2 - 4e + 5}{2n} + o(n^{-2}) \equiv \frac{\alpha}{n} + k_n, \end{aligned} \tag{2.3.17}$$

where $\alpha = (e^2 - 4e + 5)/2$, and

$$D_i = e^{\frac{i}{n}} \left(1 - e^{-\frac{1}{n}} \right) - \frac{1}{n} = \frac{1}{n} \left(e^{\frac{i}{n}} - 1 \right) + o\left(\frac{1}{n} \right). \tag{2.3.18}$$

Note that

$$n \sum_{i=1}^{n} (\frac{1}{n} + D_i) D_i = \frac{e-1}{(e^{\frac{1}{n}} + 1)} \left(e^{\frac{1}{n}}(n-1) + ne^{\frac{1}{n}+1} - en - n - 1 \right).$$

Then taking the limit, this turns to be

$$(e-1) \lim_{n\to\infty} \frac{e^{\frac{1}{n}}(1 - \frac{1}{n}) + e^{\frac{1}{n}+1} - e - 1 - \frac{1}{n}}{\frac{e^{\frac{1}{n}}}{n} + \frac{1}{n}} = \frac{1}{2}(e-1)^2. \tag{2.3.19}$$

Combining Eqs. (2.3.17), (2.3.18) and (2.3.19), we can rewrite $\hat{\lambda}_n$ as

$$\begin{aligned}
&\frac{n}{h\delta(\alpha+nk_n)}\left(\sum_{i=1}^{n}(\frac{1}{n}(e^{\frac{i}{n}}-1)+o(\frac{1}{n}))(\int_{t_{i-1}^n}^{t_i^n} hX_s ds+\varepsilon_i)-hx_0\sum_{i=1}^{n}(\frac{1}{n}+D_i)D_i\right)\\
&=\frac{1}{\alpha+nk_n}\sum_{i=1}^{n}(e^{\frac{i}{n}}-1)\left(\frac{1}{\delta}\int_{t_{i-1}^n}^{t_i^n} X_s ds+\frac{1}{h\delta}\varepsilon_i\right)\\
&\quad-\frac{nx_0}{\delta(\alpha+nk_n)}\sum_{i=1}^{n}(\frac{1}{n}+D_i)D_i+o(1)\cdot\left(\int_0^1 X_s ds+B_1\right).
\end{aligned}$$

As $\int_0^1 X_s ds$ and B_1 are finite random variables, Eq. (2.3.15) follows with

$$p_n=\frac{1}{\alpha+nk_n},\qquad q_n=-\frac{nx_0}{\delta(\alpha+nk_n)}\sum_{i=1}^{n}(\frac{1}{n}+D_i)D_i.$$

This then finished the proof. ■

Next, for the sake of simplicity and without loss of generality, we assume that initial signal $x_0=0$ and $b=\delta=h=1$.

Theorem 2.7. *The asymptotic variance of the least-squares estimator $\hat{\lambda}_n$ is strictly positive, that is,*

$$\liminf_{n\to\infty}\mathrm{Var}(\hat{\lambda}_n)>0.$$

Proof. Note that ε_i's are the increments of standard Brownian motion B_t, so

$$\mathrm{Var}\left(\sum_{i=1}^{n}(e^{\frac{i}{n}}-1)\varepsilon_i\right)=\frac{1}{n}\sum_{i=1}^{n}(e^{\frac{i}{n}}-1)^2\to\frac{1}{2}(e^2-4e+5),\ \text{as } n\to\infty.$$

Then the conclusion follows from Lemma 2.6 and the independency of X and ϵ's. ■

Theorem 2.8. *The fourth moment of the least-squares estimator $\hat{\lambda}_n$ is bounded.*

Proof. First using Lemma 2.4 follows that

$$
\begin{aligned}
\mathbb{E}\left(\sum_{i=1}^{n}(e^{\frac{i}{n}}-1)\int_{t_{i-1}^n}^{t_i^n} X_s ds\right)^4 &= \mathbb{E}\left(\sum_{i=1}^{n}(e^{\frac{i}{n}}-1)\int_{t_{i-1}^n}^{t_i^n}\Big(\int_0^s e^{s-r}dW_r+\int_0^s e^{s-r}dN_r\Big)ds\right)^4 \\
&\leq K_0\mathbb{E}\left(\sum_{i=1}^{n}(e^{\frac{i}{n}}-1)\int_{t_{i-1}^n}^{t_i^n}\int_0^s e^{s-r}dW_r ds\right)^4 \\
&\quad + K_0\mathbb{E}\left(\sum_{i=1}^{n}(e^{\frac{i}{n}}-1)\int_{t_{i-1}^n}^{t_i^n}\int_0^s e^{s-r}dN_r ds\right)^4,
\end{aligned}
\tag{2.3.20}
$$

then changing the order of integration variables, we have

$$
\begin{aligned}
&= K_0\mathbb{E}\left(\sum_{i=1}^{n}(e^{\frac{i}{n}}-1)\Big(\int_0^{t_{i-1}^n}\int_{t_{i-1}^n}^{t_i^n} e^{s-r}dsdW_r+\int_{t_{i-1}^n}^{t_i^n}\int_r^{t_i^n} e^{s-r}dsdW_r\Big)\right)^4 \\
&\quad + K_0\mathbb{E}\left(\sum_{i=1}^{n}(e^{\frac{i}{n}}-1)\Big(\int_0^{t_{i-1}^n}\int_{t_{i-1}^n}^{t_i^n} e^{s-r}dsdN_r+\int_{t_{i-1}^n}^{t_i^n}\int_r^{t_i^n} e^{s-r}dsdN_r\Big)\right)^4 \\
&\leq K_1\mathbb{E}\left(\sum_{i=1}^{n}(e^{\frac{i}{n}}-1)\Big(\int_0^{t_{i-1}^n}\int_{t_{i-1}^n}^{t_i^n} e^{s-r}dsdW_r\Big)^4\right) \\
&\quad + K_1\mathbb{E}\left(\sum_{i=1}^{n}(e^{\frac{i}{n}}-1)\Big(\int_{t_{i-1}^n}^{t_i^n}\int_r^{t_i^n} e^{s-r}dsdW_r\Big)^4\right) \\
&\quad + K_1\mathbb{E}\left(\sum_{i=1}^{n}(e^{\frac{i}{n}}-1)\Big(\int_0^{t_{i-1}^n}\int_{t_{i-1}^n}^{t_i^n} e^{s-r}dsdN_r\Big)^4\right) \\
&\quad + K_1\mathbb{E}\left(\sum_{i=1}^{n}(e^{\frac{i}{n}}-1)\Big(\int_{t_{i-1}^n}^{t_i^n}\int_r^{t_i^n} e^{s-r}dsdN_r\Big)^4\right) \\
&\equiv E_1+E_2+E_3+E_4,
\end{aligned}
\tag{2.3.21}
$$

where, K_0, $K_1 > 0$ are appropriate constants and so are the following $\{K_i\}_{i=2}^{7}$. Next taking into account Lemma 2.4 and Lemma B.5, the convergence of E_1 in Eq. (2.3.21) is shown. Indeed, along with the inequality $0 < e^{\frac{i}{n}} - 1 < e - 1$,

$$
\begin{aligned}
E_1 &\leq K_2\sum_{i=1}^{n}\left(e^{\frac{i}{n}}-1\right)^4\mathbb{E}\left(\int_0^{t_{i-1}^n}\int_{t_{i-1}^n}^{t_i^n} e^{s-r}dsdW_r\right)^4 \\
&\leq K_3\sum_{i=1}^{n}\left(\int_0^{t_{i-1}^n}\Big(\int_{t_{i-1}^n}^{t_i^n} e^{s-r}ds\Big)^2dr\right)^2.
\end{aligned}
$$

Notice that for $0 \le r \le t_{i-1}^n$, $0 < \int_{t_{i-1}^n}^{t_i^n} e^{s-r} ds \le \frac{e}{n}$. Thus, $E_1 \le \frac{K_3 e^4}{n^6} \sum_{i=1}^n (i-1)^2$ which converges to zero. For the term E_3 which encapsulates the Poissonian noise in Eq. (2.3.21), employing the compensated Poisson process again, we have that E_3

$$\begin{aligned}
&\le K_4 \sum_{i=1}^n \mathbb{E}\left(\int_0^{t_{i-1}^n} \int_{t_{i-1}^n}^{t_i^n} e^{s-r} ds dN_r\right)^4 \\
&\le K_4 \mathbb{E}\left(\int_0^{t_{i-1}^n} \int_{t_{i-1}^n}^{t_i^n} e^{s-r} ds d\hat{N}_r + \lambda \int_0^{t_{i-1}^n} \int_{t_{i-1}^n}^{t_i^n} e^{s-r} ds dr\right)^4 \\
&\le K_5 \sum_{i=1}^n \mathbb{E}\left(\int_0^{t_{i-1}^n} \int_{t_{i-1}^n}^{t_i^n} e^{s-r} ds d\hat{N}_r\right)^4 + K_5 \lambda^4 \sum_{i=1}^n \left(\int_0^{t_{i-1}^n} \int_{t_{i-1}^n}^{t_i^n} e^{s-r} ds dr\right)^4 \\
&\le K_6 \sum_{i=1}^n \mathbb{E}\left(\int_0^{t_{i-1}^n} (\int_{t_{i-1}^n}^{t_i^n} e^{s-r} ds)^2 dN_r\right)^2 + \frac{K_5 \lambda^4 e^4}{n^3} \\
&\le \frac{K_6 e^4}{n^4} \sum_{i=1}^n \mathbb{E}\left(N_{t_{i-1}^n}^2\right) + \frac{K_5 \lambda^4 e^4}{n^3}.
\end{aligned}$$

Note that $\mathbb{E}(N_{t_{i-1}^n}^2) = \lambda \frac{i-1}{n} + \lambda^2 \frac{(i-1)^2}{n^2}$. Hence we have the desired result that E_3 converges to zero.

Similarly, one may show that the terms E_2 and E_4 of Eq. (2.3.21) converge to 0 as $n \to \infty$. Besides,

$$\mathbb{E}\left(\sum_{i=1}^n (e^{\frac{i}{n}} - 1)\varepsilon_i\right)^4 \le K_7 \sum_{i=1}^n \mathbb{E}\left(\varepsilon_i\right)^4 = \frac{3K_7}{n} \to 0.$$

This implies that $\mathbb{E}(|\tilde{\lambda}_n|)^4$ is convergent and thus bounded. Therefore, the boundedness of the fourth moment follows from Eq. (2.3.15). ■

Theorem 2.9. *The least-squares estimator $\hat{\lambda}_n$ is inconsistent.*

Proof. We suppose the estimator $\hat{\lambda}_n$ is consistent, i.e. $\hat{\lambda}_n \to \lambda$ in probability as $n \to \infty$. By the last theorem, $\mathbb{E}(|\hat{\lambda}_n - \lambda|^4)$ is bounded, then $|\hat{\lambda}_n - \lambda|^2$ is uniformly integrable. Thus, we have

$$\lim_{n\to\infty} \mathbb{E}\left(|\hat{\lambda}_n - \lambda|^2\right) = \mathbb{E}\left(\lim_{n\to\infty} |\hat{\lambda}_n - \lambda|^2\right) = 0.$$

This contradicts the result of Theorem 2.7 that the liminf of the variance of the unbiased estimator is $c > 0$, so conclusion of the theorem is proved. ■

2.4 Simulation

This section is a demonstration of the theoretical landscape of Sections 2.3.1 and 2.3.2 respectively. The initial parameters are set $X_0 = Y_0 = 0$ and $b = \delta = h = 1$. The simulation setup with respect to Section 2.3.1 considers $a = -1$, $t_i^n = i$, $i = 1, ..., n$, $T = n$ and the corresponding scenario of Section 2.3.2 adopts $a = \pm 1$, $t_i^n = \frac{i}{n}, T = 1$. The estimate of the intensity for different n and λ are summarized in Table 2.1, Table 2.2 and Table 2.3. The associated squared error, $(\hat{\lambda}_n - \lambda)^2$, is listed as well. As it was anticipated by the associated theory developed in Section 2.3.1, we observe from Table 2.1 that the estimation is accurate and the error decreases as more data are collected. This result could be fruitful for big data analysis associated with a partially observed jump-diffusion process, for example stochastic volatility of bitcoins. In contrast, the least-squares estimation for a fixed time horizon is inaccurate as verified in Table 2.2 and Table 2.3, respectively. Furthermore, we repeated the simulation procedure $1,000$ times with $n = 100,000$ and $\lambda = 1$. We observe in Figure 2.1 that the estimator converges to the true value ($\lambda = 1$) and the associated error is close to 0. Figure 2.4 shows the asymptotic normality of $\sqrt{n}(\hat{\lambda}_n - \lambda)$. Figure 2.5, similarly, verifies the empirical CDF of Monte Carlo samples and contrasts it against the theoretical distribution. A QQ-plot is also displayed to visually notice the alignment with the theoretical normal distribution. By the same token, Figure 2.2 and Figure 2.3 display the inconsistency of the least-squares estimator and the associated error which is very large in comparison to the one depicted in Figure 2.1.

Table 2.1: Simulation result when $a = -1$, $t_i^n = i$

	$\lambda = 1$		$\lambda = 3$	
n	$\hat{\lambda}_n$	Error	$\hat{\lambda}_n$	Error
1000	1.04952434	2.4526606e-03	3.04025959	1.6208348e-03
10000	0.97242533	7.6036266e-04	2.97121824	8.2838998e-04
100000	1.00299904	8.9942631e-06	2.99799043	4.0383843e-06
1000000	1.00010593	3.6325661e-08	2.99974629	1.5717925e-08

Table 2.2: Simulation result when $a = 1$, $t_i^n = i/n$

	$\lambda = 1$		$\lambda = 3$	
n	$\hat{\lambda}_n$	Error	$\hat{\lambda}_n$	Error
1000	2.98851393	3.9541876475	4.557670318	2.4263368205
10000	1.38856624	0.1509837267	2.417325080	0.3395100619
100000	3.44658968	5.9858010721	3.973874657	0.9579448659
1000000	-0.80372271	3.2534156242	-0.18654475	10.154067441

Table 2.3: Simulation result when $a = -1$, $t_i^n = i/n$

	$\lambda = 1$		$\lambda = 3$	
n	$\hat{\lambda}_n$	Error	$\hat{\lambda}_n$	Error
1000	1.65440831	0.4282502402	0.81466433	4.7756919576
10000	-2.88579644	15.099413996	7.06558363	16.528970303
100000	-0.53224464	2.3477736521	2.058130036	0.8871190285
1000000	5.481551578	20.084304549	2.852386349	0.0217897896

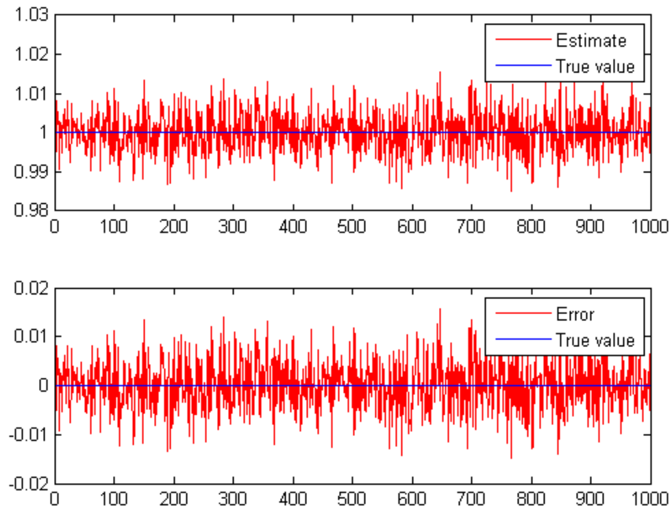

Figure 2.1: Simulations and errors for $\lambda = 1, a = -1, t_i^n = i$, when repeating $1,000$ times

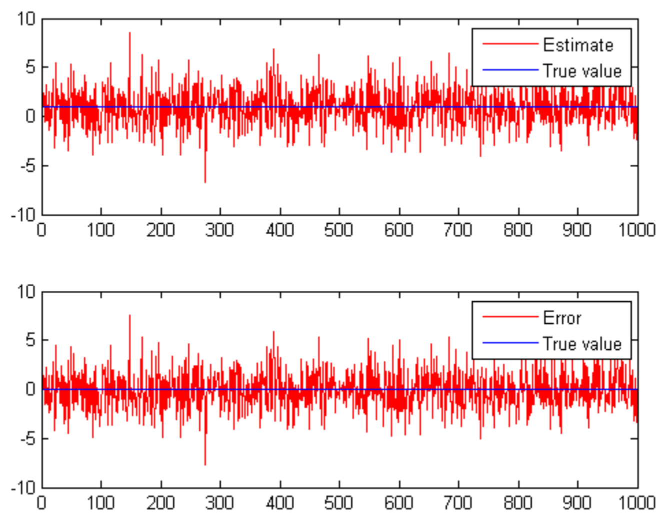

Figure 2.2: Simulations and errors, $\lambda = 1, a = 1, t_i^n = i/n$, when repeating 1000 times.

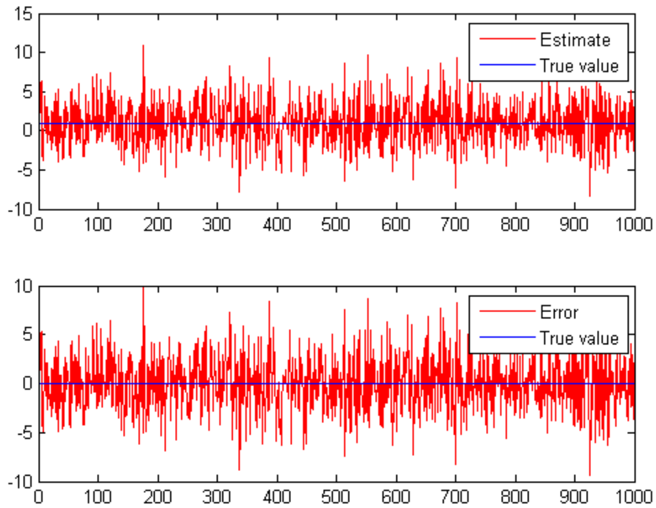

Figure 2.3: Simulations and errors $\lambda = 1, a = -1, t_i^n = i/n$, when repeating 1000 times

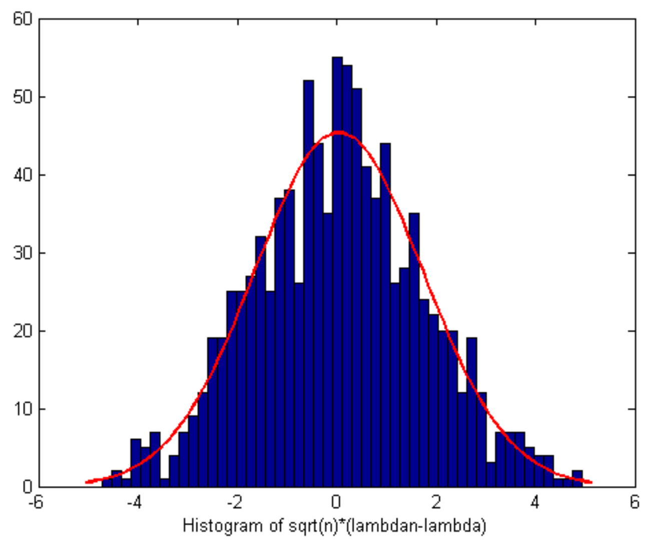

Figure 2.4: Monte Carlo samples verifying asymptotic normality when $\lambda = 1, a = -1, t_i^n = i$

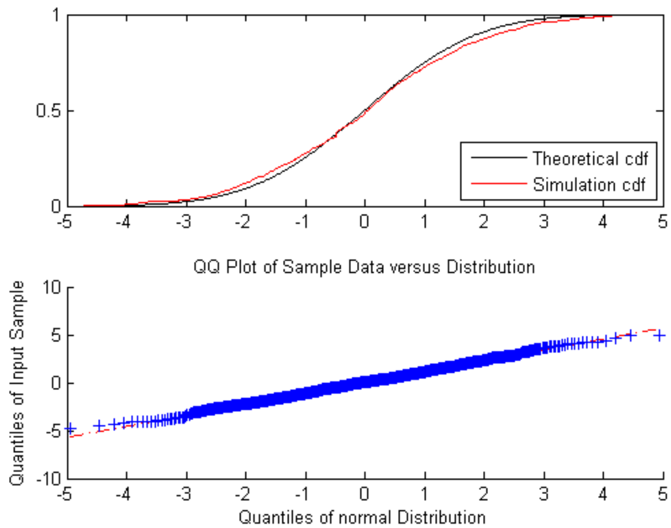

Figure 2.5: Empirical cumulative function estimate and QQ-plot of samples $\sqrt{n}(\hat{\lambda}_n - \lambda)$ when $\lambda = 1, a = -1,\ t_i^n = i$

Chapter 3

Statistical inference

3.1 Introduction

To study the statistical inference for the parameter of interest, one needs an associated distribution of the parameter. We have proposed an unbiased, consistent least-squares estimator for the intensity in the jump-diffusion under partial information and derived a central limit theorem in the last chapter, however, the limit variance in the central limit theorem depends on the true value, and thus the approximation cannot be employed for resolving the inference problem of the intensity of the signals' jumps. To that end we study the continuation of the central limit theorem, large and moderate deviation principles. Consider a sequence of random variables $\sqrt{n}b_n^{-1}\gamma_n$ with $b_n \to \infty$. If $\sqrt{n}b_n^{-1} \to c > 0$, the result is called large deviation principle (LDP) of $\{\gamma_n\}$; if $\sqrt{n}b_n^{-1} \to \infty$, the result is called moderate deviation principle (MDP). We note that if $b_n \to c$, the result belongs to the regime of central limit theorem. Deviation principles provide us with the rates of convergence of LSE and prescribe a strategy for approximating the power of the associated hypothesis testing as well as computing $100\%(1-\alpha_0)$ confidence intervals for the intensity. In Section 3.2 we give some elements of large deviation theory and state our method, Gartner-Ellis theorem. The inference results, the power of hypothesis testing and confidence intervals, are presented in Section 3.3 based on the moderate deviation principle.

3.2 Preliminary

In this chapter, the asymptotic behavior of the least-squares estimator $\hat{\lambda}_n$ of the intensity λ in the signal-observation system (2.2.1) is considered when the drift coefficient $a < 0$. According to Chapter 2, $a \geq 0$ does not yield an interesting behavior, in fact one should notice that the trajectory of X_t blows up when $a > 0$. Assume that Y_t is observable at discrete time $\{t_0^n, t_1^n, ..., t_n^n\}$ and precisely $t_i^n = i,\ i = 1, ..., n$ (and $T = n \to \infty$).

As discussed in Chapter 2, we define the innovation process $Z_{t_i^n} := Y_{t_i^n} - Y_{t_{i-1}^n} = h\int_{t_{i-1}^n}^{t_i^n} X_s ds + \varepsilon_i$, where ε_i denote the normally distributed and independent Brownian increments $B_{t_i^n} - B_{t_{i-1}^n} \sim$ i.i.d. $N(0, \Delta t_{i-1}^n)$, $\Delta t_{i-1}^n = t_i^n - t_{i-1}^n$, $i = 1, ..., n$. Let Θ denote the parameter space of the intensity of the driving Poisson process. The least-square estimator $\hat{\lambda}_n$ of λ is to find the intensity which minimizes the residual sum of squares (RSS). That is, $\hat{\lambda}_n = \arg\min_{\lambda\in\Theta} Q(\lambda)$, where $Q(\lambda) = \sum_{i=1}^{n}(Z_{t_i^n} - \mathbb{E}(Z_{t_i^n}))^2$. The solution of the optimization problem derived in Chapter 2 yields an unbiased and consistent least-squares estimator of the intensity λ:

$$\hat{\lambda}_n = \frac{a}{h\delta\sum_{i=1}^{n}(\frac{1}{a}d_i - 1)^2}\sum_{i=1}^{n}\left(Z_i - \frac{hx_0}{a}d_i\right)\left(\frac{1}{a}d_i - 1\right). \tag{3.2.1}$$

where $d_i = e^{ai} - e^{ai-a}$.

We recall that Theorem 2.5 provides the limit distribution of the least-squares estimator, $\hat{\lambda}_n$, however its asymptotic variance depends on the true intensity. To that end, different convergence arguments, precisely a moderate deviation principle for the $\hat{\lambda}_n$ need to be established. The moderate deviations principle is related to the large deviations. Therefore, we present below large and moderate deviations formalism. The reader may refer to [21] for more details.

Definition 3.1. *A function $I : \mathbb{R} \to [0, \infty]$ is called a rate function if it is lower semi-continuous: For any $l > 0$, the level set $I_l = \{x \in \mathbb{R}; I(x) \leq l\}$ is a closed set. Further, a rate function is said to be good if every level is compact in $\mathbb{R}$.*

Definition 3.2. *The sequence $\{\gamma_n\}$ is said to satisfy the large deviation principle on Γ with rate function I, if there exist $c_n \to \infty$ as $n \to \infty$ and the following hold.*

1. *For each closed subset F of Γ,*

$$\limsup_{n\to\infty} \frac{1}{c_n} \log \mathbb{P}(\gamma_n \in F) \le - \inf_{x\in F} I(x). \tag{3.2.2}$$

2. *For each open subset G of Γ,*

$$\liminf_{n\to 0} \frac{1}{c_n} \log \mathbb{P}(\gamma_n \in G) \ge - \inf_{x\in G} I(x). \tag{3.2.3}$$

Assumption 3.2.1. *Let $\{\gamma_n\}$ be a sequence of random variables with topological state space Γ. For each $\mu \in \mathbb{R}$ and a sequence $c_n \to \infty$ as $n \to \infty$, the logarithmic moment generating function $\Lambda(\mu)$, defined as the limit*

$$\Lambda(\mu) = \lim_{n\to\infty} \frac{1}{c_n} \log \mathbb{E} \exp\{\mu c_n \gamma_n\} \tag{3.2.4}$$

exists in $(-\infty, +\infty]$. Further, the origin belongs to the interior $\mathcal{D}_\Lambda^\circ$ of the domain $\mathcal{D}_\Lambda = \{\mu \in \mathbb{R} : \Lambda(\mu) < \infty\}$ of the function $\Lambda(\mu)$.

By Hölder inequality, $\Lambda(\mu)$ is a convex function. Define the *Fenchel-Legendre transform*

$$I(x) = \sup_{\mu\in\mathbb{R}} \{\mu x - \Lambda(\mu)\}, \; x \in \mathbb{R}. \tag{3.2.5}$$

Then the function $I(x)$ is a good rate function, see [21, Section 4.5].

Definition 3.3. *A convex function $\Lambda : \mathbb{R} \to (-\infty, \infty]$ is essentially smooth if*

(1) $\mathcal{D}_\Lambda^\circ = (a, b)$ is non-empty for some $-\infty \le a < b \le \infty$.

(2) $\Lambda(\mu)$ is differentiable in $\mathcal{D}_\Lambda^\circ$.

(3) $\Lambda(\cdot)$ is steep, i.e. $\lim_{\mu\to a^+} \Lambda'(\mu) = \lim_{\mu\to b^-} \Lambda'(\mu) = \infty$.

The following theorem is known as the *Gartner-Ellis theorem* (see e.g. [16, 21]) on large deviations, which shows the correspondence between the logarithmic moment function and large deviations.

Theorem 3.4. *Suppose Assumption 3.2.1 holds. Then for any closed set $F \subset \mathbb{R}$, inequality (3.2.2) is satisfied. If we further assume that the logarithmic moment function $\Lambda(\mu)$ is essentially smooth, then for any open set $G \subset \mathbb{R}$, inequality (3.2.3) is satisfied.*

Next, a motivating large deviation principle for the LSE, $\hat{\lambda}_n$, is presented whose proof is delegated to Section 3.4. Take $c_n = n$ in Theorem 3.5.

Theorem 3.5. *Consider system (2.2.1). The sequence of least-squares estimates $\{\hat{\lambda}_n, n \geq 1\}$ for the intensity, given by (3.2.1), satisfies the large deviation principle in $\mathbb{R}$, with rate function $I_\ell(x) = \sup_{\mu\in\mathbb{R}} \{\mu x - \Lambda_\ell(\mu)\}$ where the logarithmic moment generating function $\Lambda_\ell(\mu)$ defined with $c_n = n$ in Eq. (3.2.4) is given by*

$$\Lambda_\ell(\mu) = \frac{(h^2b^2 + a^2)\mu^2}{2h^2\delta^2} + \lambda e^\mu - \lambda, \quad \mu \in \mathbb{R}. \tag{3.2.6}$$

Theorem 3.5 provides the rate function of convergence, I_ℓ, however it depends on the true value of the unknown parameter (as the asymptotic variance in Theorem 2.3.8). Therefore, Theorem 3.5 is insufficient to conduct statistical inference relying on this asymptotic tail probability. To that end, moderate deviations need to be examined instead.

3.3 Statistical inference

3.3.1 Moderate deviations

We are now concerned with the weak convergence of the quantity $n^\alpha(\hat{\lambda}_n - \lambda)$, for $0 < \alpha < \frac{1}{2}$, i.e. the moderate deviation principles, using the logarithmic moment generating function $\frac{1}{c_n} \log \mathbb{E} \exp\left(c_n \mu n^\alpha (\hat{\lambda}_n - \lambda)\right)$, where $c_n = n^\beta$ for some fixed $\beta > 0$. The result is presented in the following theorem.

Theorem 3.6. *Consider system (2.2.1). The least-squares estimate, $\{\hat{\lambda}_n\}$, given by (3.2.1), satisfies a moderate deviation principle, precisely, for any $\alpha \in (0, \frac{1}{2})$, the sequence $\{n^\alpha(\hat{\lambda}_n - \lambda), n \geq 1\}$ satisfies the large deviation principle in $\mathbb{R}$ with speed $n^{1-2\alpha}$ and the rate function*

$$I_m(x) = \frac{x^2}{2\kappa^2 + 2}, \quad \textit{where} \quad \kappa^2 = \frac{h^2b^2 + a^2}{h^2\delta^2}. \tag{3.3.1}$$

3.3.2 Hypothesis testing and confidence intervals

According to Theorem 2.5 (CLT), the variance of the asymptotic distribution is a function of the true intensity, which confines the use of the statistics in the applications of hypothesis testing and confidence intervals.

Consider the null and alternative hypotheses

$$H_0 : \lambda = \lambda_0 \text{ v.s. } H_1 : \lambda \neq \lambda_0,$$

and the test statistic, $T_n := \sqrt{n}(\hat{\lambda}_n - \lambda_0)$. For $0 < \alpha_0 < 1$, denote the rejection region for testing the null hypothesis H_0 against H_1 to be $\mathcal{R} = \{|T_n| \geq c(\alpha_0)\}$, where $c(\alpha_0)$ is a positive constant in $\mathbb{R}$ such that $\alpha_0/2 = 1 - \Phi(\frac{c(\alpha_0)}{\sigma_0})$ with $\sigma_0^2 = \frac{b^2}{\delta^2} + \lambda_0 + \frac{a^2}{h^2\delta^2}$, where $\Phi(\cdot)$ is the cdf of a standard normal distribution. We consider the *power* of the hypothesis testing, i.e. the probability of correctly rejecting a false null hypothesis, $1 - \beta_n$, where β_n is the probability of Type II Error given by

$$\beta_n = \mathbb{P}\left(|T_n| < c(\alpha_0) \;\Big|\; \lambda = \lambda_1 \neq \lambda_0 \text{ under } H_1\right).$$

Proposition 3.7. *Consider system (2.2.1). Then the power of the hypothesis testing tends to* 1 *with exponential speed* $\exp\{-rn^{1-2\alpha}\}$ *for any* $\alpha \in (0, 1/2)$ *and any* $r > 0$. *In other words,*

$$\lim_{n\to\infty} \frac{1}{n^{1-2\alpha}} \log \beta_n = -\infty.$$

Proof. Note that by triangle inequality, $|\hat{\lambda}_n - \lambda_0| \geq |\lambda_1 - \lambda_0| - |\hat{\lambda}_n - \lambda_1|$. Then we have

$$\beta_n = \mathbb{P}\left(|T_n| < c(\alpha_0) \;\Big|\; \lambda = \lambda_1\right) \leq \mathbb{P}\left(n^\alpha|\hat{\lambda}_n - \lambda_1| \geq n^\alpha|\lambda_1 - \lambda_0| - \frac{c(\alpha_0)}{n^{\frac{1}{2}-\alpha}} \;\Big|\; \lambda = \lambda_1\right).$$

Then Theorem 3.6 implies that $\lim_{n\to\infty} \frac{1}{n^{1-2\alpha}} \log \beta_n = -\infty$. ■

Next, a direct application of the asymptotic normality derived in Theorem 2.5 yields an approximate $(1 - \alpha_0) \cdot 100\%$ confidence intervals for λ as

$$\left(\hat{\lambda}_n - \frac{\sigma}{\sqrt{n}} Z_{\alpha_0/2}, \hat{\lambda}_n + \frac{\sigma}{\sqrt{n}} Z_{\alpha_0/2}\right),$$

where $Z_{\alpha_0/2}$ is a critical value of the standard random variable Z such that $\mathbb{P}(Z > Z_{\alpha_0/2}) = \alpha_0/2$, and σ^2 is defined in Theorem 2.5. Consider the inequality $-Z_{\alpha_0/2} \leq \frac{\sqrt{n}(\hat{\lambda}_n - \lambda)}{\sigma} \leq Z_{\alpha_0/2}$, which is equivalent to

$$\lambda^2 - \left(2\hat{\lambda}_n + \frac{Z^2_{\alpha_0/2}}{n}\right)\lambda + \hat{\lambda}_n^2 - \left(\frac{b^2}{n\delta^2} + \frac{a^2}{nh^2\delta^2}\right)Z^2_{\alpha_0/2} \leq 0. \tag{3.3.2}$$

Thus, solving inequality (3.3.2) gives the $(1-\alpha_0)\cdot 100\%$ confidence intervals by the weak convergence

$$\left(\hat{\lambda}_n + \frac{Z^2_{\alpha_0/2}}{2n} \pm \frac{1}{2}\left(\left(2\hat{\lambda}_n + \frac{Z^2_{\alpha_0/2}}{n}\right)^2 - 4\left(\frac{b^2}{n\delta^2} + \frac{a^2}{nh^2\delta^2}\right)Z^2_{\alpha_0/2}\right)^{1/2}\right). \tag{3.3.3}$$

We can see from (3.3.3) that the confidence intervals by the central limit theorem have large errors when the true value and accordingly the unbiased point estimate $\hat{\lambda}_n$ is large. We now apply Theorem 3.6 to constructing confidence intervals for the intensity λ.

Proposition 3.8. *Consider system (2.2.1). The MDP based* $(1-\alpha_0)\cdot 100\%$ *confidence intervals for the intensity* λ*,* $\alpha_0 \in (0,1)$*, are given by*

$$\left(\hat{\lambda}_n + \sqrt{\frac{2\kappa^2+2}{n}}\,(\log\alpha_0)^{1/2}, \hat{\lambda}_n - \sqrt{\frac{2\kappa^2+2}{n}}\,(\log\alpha_0)^{1/2}\right), \tag{3.3.4}$$

where κ^2 *is given in (3.3.1).*

Proof. By Theorem 3.6, for a fixed $u > 0$ we have

$$\mathbb{P}\left(\left|n^{\alpha}(\hat{\lambda}_n - \lambda)\right| > u\right) \approx \exp\left(-n^{1-2\alpha}\inf_{|x|>u} I_m(x)\right) = \exp\left(-\frac{n^{1-2\alpha}u^2}{2\kappa^2+2}\right),$$

where $\kappa^2 = \frac{h^2b^2+a^2}{h^2\delta^2}$. For a given confidence level $1-\alpha_0$, set

$$\tau_{\alpha_0} = \left(\frac{2\kappa^2+2}{n^{1-2\alpha}}\log\frac{1}{\alpha_0}\right)^{1/2} = n^{\alpha}\left(\frac{2\kappa^2+2}{n}\log\frac{1}{\alpha_0}\right)^{1/2}.$$

Then the $(1-\alpha_0)\cdot 100\%$ confidence intervals by the moderate deviation principle for the intensity λ are approximately $\left(\hat{\lambda}_n - n^{-\alpha}\tau_{\alpha_0}, \hat{\lambda}_n + n^{-\alpha}\tau_{\alpha_0}\right)$, which follows expression (3.3.4).

■

3.4 Proof of Theorems 3.5 and 3.6

In this section, we give the proofs of the results of large and moderate deviations which were presented in Section 3.3.1. The main tools we employ are Fenchel-Legendre transform and Gartner-Ellis theorem that were recalled in Section 3.2.

Proof of Theorem 3.5. Employing Itô's formula [34], we obtain the solution of signal process

$$X_t = e^{at}(x_0 + b\int_0^t e^{-as}dW_s + \delta\int_0^t e^{-as}dN_s), \tag{3.4.1}$$

and consequently,

$$\int_0^n X_r dr = \frac{b}{|a|}\int_0^n (1-e^{an-as})dW_s + \frac{\delta}{|a|}\int_0^n (1-e^{an-as})dN_s + \frac{1}{|a|}x_0(1-e^{an}). \tag{3.4.2}$$

This implies that

$$\begin{aligned}\mathbb{E}\exp\left(\frac{\mu|a|}{\delta}\int_0^n X_s ds\right) =& \mathbb{E}\exp\left(\frac{\mu b}{\delta}\int_0^n (1-e^{a(n-s)})dW_s\right)\\ &\times \mathbb{E}\exp\left(\mu\int_0^n (1-e^{a(n-s)})dN_s\right)\\ &\times \mathbb{E}\exp\left(\frac{\mu x_0}{\delta}(1-e^{an})\right).\end{aligned} \tag{3.4.3}$$

Note that $\frac{\mu b}{\delta}\int_0^n (1-e^{a(n-s)})dW_s$ follows a normal distribution $N(0,\tau_n^2)$, for a pertinent variance τ_n^2, $n\in\mathbb{N}$. Hence

$$\mathbb{E}\exp\left(\frac{\mu b}{\delta}\int_0^n (1-e^{a(n-s)})dW_s\right) = \exp\left(\frac{1}{2}\tau_n^2\right), \tag{3.4.4}$$

where $\tau_n^2 = \mathbb{E}\left(\frac{\mu b}{\delta}\int_0^n (1-e^{a(n-s)})dW_s\right)^2 = \frac{\mu^2 b^2}{\delta^2}\int_0^n (1-e^{a(n-s)})^2 ds$. Next, by Itô's formula [34, Rule 4.23], for a suitable function g we have the SDE

$$d\left\{\exp\left(\int_0^t g(s)dN_s\right)\right\} = \exp\left\{\int_0^t g(s)dN_s\right\}\left(e^{g(t)}-1\right)dN_t. \tag{3.4.5}$$

Taking expectation on the integral form of Eq. (3.4.5) gives

$$\mathbb{E}\exp\left\{\int_0^t g(s)dN_s\right\} = 1+\lambda\int_0^t \left(e^{g(r)}-1\right)\mathbb{E}\exp\left\{\int_0^r g(s)dN_s\right\}dr.$$

Let $G(t) := \mathbb{E}\exp\left\{\int_0^t g(s)dN_s\right\}$. Then solving the ODE $G'(t) = \lambda G(t)\left(e^{g(t)}-1\right)$ yields that

$$G(t) = G(0)\exp\left\{\lambda\int_0^t \tilde{g}(s)ds\right\}, \tag{3.4.6}$$

where $\tilde{g}(s) := e^{g(s)}-1$. Consequently, substitution with $g(s) := \mu(1-e^{a(n-s)})$ into Eq. (3.4.6) follows that

$$\mathbb{E}\exp\left(\mu\int_0^n (1-e^{a(n-s)})dN_s\right) = \exp\left\{\lambda\int_0^n \left(e^{\mu-\mu e^{a(n-s)}}-1\right)ds\right\}. \tag{3.4.7}$$

Thus, in combination with Eqs. (3.4.3), (3.4.4), and (3.4.7), we have

$$\begin{aligned}
&\frac{1}{n}\log\mathbb{E}\exp\left(\frac{\mu|a|}{\delta}\int_0^n X_s ds\right)\\
=&\frac{1}{n}\left(\frac{1}{2}\sigma_n^2+\lambda\int_0^n \left(e^{\mu-\mu e^{a(n-s)}}-1\right)ds+\frac{\mu x_0}{\delta}(1-e^{an})\right)\\
\to&\frac{\mu^2 b^2}{2\delta^2}+\lambda(e^{\mu}-1).
\end{aligned} \tag{3.4.8}$$

Next, notice that $\frac{\mu|a|}{h\delta}\sum_{i=1}^n \varepsilon_i \sim N(0, \frac{\mu^2 a^2 T}{h^2\delta^2})$. Therefore, $\mathbb{E}\exp\left(\frac{\mu|a|}{h\delta}\sum_{i=1}^n \varepsilon_i\right) = \exp\left(\frac{\mu^2 a^2 T}{2h^2\delta^2}\right)$, and furthermore $\lim_{n\to\infty}\frac{1}{n}\log\mathbb{E}\exp\left(\frac{\mu|a|}{h\delta}\sum_{i=1}^n \varepsilon_i\right) = \frac{\mu^2 a^2}{2h^2\delta^2}$. Now we claim that, $\frac{1}{n}\xi_n$ converges to 0 exponentially fast, i.e.,

$$\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}\left(\left|\frac{1}{n}\xi_n\right| > \epsilon\right) = -\infty. \tag{3.4.9}$$

Indeed, by Markov's inequality

$$\mathbb{P}\left(\frac{1}{n}|\xi_n|>\epsilon\right)=\mathbb{P}\left(\left|\frac{(1-e^{-a})}{h\delta}\sum_{i=1}^{n}e^{ai}Z_i\right|>n\epsilon\right)\le e^{-n^2\epsilon^2}\mathbb{E}\exp\left\{\left(\frac{(1-e^{-a})}{h\delta}\sum_{i=1}^{n}e^{ai}Z_i\right)^2\right\},$$

and then the claim follows from Lemma 2.1. Therefore, the logarithmic moment generating function (3.2.6) is concluded by Eqs. (3.4.9), (3.4.8) and (2.3.3), and the rate function follows from Theorem 3.4. ■

Proof of Theorem 3.6. By Eq. (3.4.9) and Lemma 2.1, we only need to consider

$$\frac{1}{n^{\beta}}\log\mathbb{E}\exp\left(\mu n^{\beta+\alpha}(\tilde{\lambda}_n-\lambda)\right),\ \text{for some } \beta>0,\ \alpha\in(0,\frac{1}{2}). \tag{3.4.10}$$

It follows from Eq. (2.3.2) that, (3.4.10) is given by

$$\begin{aligned}&\frac{1}{n^{\beta}}\log\mathbb{E}\exp\left(\mu n^{\beta+\alpha}\left(\frac{|a|}{n\delta}\int_0^n X_s ds+\frac{|a|}{hn\delta}B_T-\lambda\right)\right) \qquad (3.4.11)\\ =&\frac{1}{n^{\beta}}\log\mathbb{E}\exp\left(\delta^{-1}|a|\mu n^{\beta+\alpha-1}\int_0^n X_s ds\right)+\frac{1}{n^{\beta}}\log\mathbb{E}\exp\left(\mu n^{\beta+\alpha-1}\frac{|a|}{h\delta}B_T\right)-\mu\lambda n^{\alpha}.\end{aligned}$$

Letting $a_n:=\frac{|a|\mu}{\delta}n^{\beta+\alpha-1}$, Eq. (3.4.2) yields

$$\begin{aligned}\mathbb{E}\exp\left(a_n\int_0^n X_s ds\right)&=\mathbb{E}\exp\left(\frac{a_n b}{|a|}\int_0^n(1-e^{a(n-s)})dW_s\right) \qquad (3.4.12)\\ &\times\mathbb{E}\exp\left(\frac{a_n\delta}{|a|}\int_0^n(1-e^{a(n-s)})dN_s\right)\times\mathbb{E}\exp\left(\frac{a_n}{|a|}x_0(1-e^{an})\right).\end{aligned}$$

First, employing Eq. (3.4.4), we have

$$\mathbb{E}\exp\left(\frac{a_n b}{|a|}\int_0^n(1-e^{a(n-s)})dW_s\right)=\exp\left\{\frac{\mu^2b^2n^{2\beta+2\alpha-2}}{4a\delta^2}(2an+3-4e^{an}+e^{2an})\right\}. \tag{3.4.13}$$

Making use of Eq. (3.4.6) with $g(s):=\frac{a_n\delta}{|a|}(1-e^{a(n-s)})$ yields

$$\mathbb{E}\exp\left(\frac{a_n\delta}{|a|}\int_0^n(1-e^{a(n-s)})dN_s\right)\exp\left\{\lambda\int_0^n\left\{\exp\left(\mu n^{\beta+\alpha-1}(1-e^{a(n-s)})\right)-1\right\}ds\right\}. \tag{3.4.14}$$

Recall that $a < 0$, thus the term $\int_0^n \exp\{\mu n^{\beta+\alpha-1} e^{an-as}\} ds$ decays exponentially. Choosing $\beta + \alpha - 1 < 0$, by Taylor's formula, Eq. (3.4.14) is further given by

$$\exp\left\{\lambda \mu n^{\beta+\alpha} + \frac{1}{2}\mu^2 n^{2\beta+2\alpha-1} + o(n^{2\beta+2\alpha-1})\right\}. \tag{3.4.15}$$

Note that B_T is a standard Brownian motion at time T, and hence it is easy to get that $\mathbb{E}\exp\left(\mu n^{\beta+\alpha-1}\frac{|a|}{h\delta}B_T\right) = \exp\left\{\frac{a^2\mu^2}{2h^2\delta^2}n^{2\beta+2\alpha-1}\right\}$. Consequently, substituting this and Eqs. (3.4.13), (3.4.15) into (3.4.12) and taking into account Eq. (3.4.11), the logarithm moment generating function, eq. (3.4.10) is expressed as

$$\begin{aligned} &\frac{\mu^2 b^2 n^{\beta+2\alpha-2}}{4a\delta^2}(2an + 3 - 4e^{an} + e^{2an}) + \frac{1}{2}\mu^2 n^{\beta+2\alpha-1} + o(n^{\beta+2\alpha-1}) \\ &+ \frac{\mu x_0}{\delta} n^{\alpha-1}(1 - e^{an}) + \frac{a^2\mu^2}{2h^2\delta^2} n^{\beta+2\alpha-1}. \end{aligned} \tag{3.4.16}$$

Now we see that letting $\beta = 1 - 2\alpha$, eq. (3.4.16) converges to a constant as $n \to \infty$. Similarly to the proof of eq. (3.4.9), one can show that $\frac{1}{n}\xi_n$ converges to 0 exponentially fast with a scale n^β, i.e.,

$$\limsup_{n\to\infty} \frac{1}{n^\beta} \log \mathbb{P}\left(\left|\frac{1}{n}\xi_n\right| > \epsilon\right) = -\infty. \tag{3.4.17}$$

Thus, together with Lemma 2.1, Eqs. (3.4.16) and (3.4.17) imply that the logarithmic moment generating function

$$\Lambda_m(\mu) := \lim_{n\to\infty} \frac{1}{n^{1-2\alpha}} \log \mathbb{E}\exp\left(\mu n^{1-\alpha}(\hat{\lambda}_n - \lambda)\right) = \frac{(h^2b^2 + a^2)\mu^2}{2h^2\delta^2} + \frac{1}{2}\mu^2,$$

for any $\alpha \in (0, \frac{1}{2})$. The rate function then is given by the Fenchel-Legendre transform (3.2.5): $I_m(x) = \sup_{\mu\in\mathbb{R}}\{\mu x - \Lambda_m(\mu)\} = \frac{x^2}{2\kappa^2+2}$, with $\kappa^2 = \frac{h^2b^2+a^2}{h^2\delta^2}$. ■

Chapter 4

Nonlinear dynamical systems with Lévy noise

4.1 Introduction

In this chapter we study the general nonlinear dynamical systems where the signal and observation processes are both driven by Lévy noises. A large deviation principle for the optimal filter is investigated by the weak convergence method. To use this argument, uniqueness of the solutions to the controlled Zakai and Kushner-Stratonovich equations is needed. Next section lists definitions and notations used later on, as well as a general criterion of large deviations. Most of these notations can be also found in [12] but they are presented here for the sake of completeness. We prove the uniqueness of the solutions to the Zakai and Kushner-Stratonovich equations in Section 4.3 which are given in Theorem 4.13 and Theorem 4.14 respectively. Section 4.4 focuses on the establishment of the large deviation principle for the optimal filter. We start with deriving the controlled version and zero-noise version of Zakai equation, then the existence and uniqueness of these two versions of Zakai equation are verified and finally the results are concluded by demonstrating the sufficient conditions presented in Propositions 4.23 and 4.24.

4.2 Preliminaries

4.2.1 Definitions and conventional notations

Let $\mathbb{U}$ be a locally compact Polish space and denote by $\mathcal{M}(\mathbb{U})$ the space of all measures ν on $(\mathbb{U}, \mathcal{B}(\mathbb{U}))$, satisfying $\nu(A) < \infty$ for every compact subset A of $\mathbb{U}$, and $\mathcal{B}(\mathbb{U})$ is the Borel σ-field on $\mathbb{U}$. Endow $\mathcal{M}(\mathbb{U})$ with the weakest topology such that for every continuous function f on $\mathbb{U}$ with compact support, the function $\nu \to \int_{\mathbb{U}} f(u)\nu(du)$, $\nu \in \mathcal{M}(\mathbb{U})$ is continuous. This topology can be metrized such that $\mathcal{M}(\mathbb{U})$ is a Polish space, e.g. see [12]. For a fixed $T \in (0, \infty)$, denote by $\mathbb{M} = \mathcal{M}(\mathbb{U}_T)$ the space of measures on $\mathbb{U}_T = [0, T] \times \mathbb{U}$ and let $\nu_T = \lambda_T \otimes \nu$, λ_T is the Lebesgue measure on $[0, T]$. We recall that a general Poisson random measure (PRM) $\mathbf{n}$ on $\mathbb{U}_T$ with intensity measure ν_T is an $\mathbb{M}$-valued random variable such that for each $A \in \mathcal{B}(\mathbb{U}_T)$ with $\nu_T(A) < \infty$, $\mathbf{n}(A)$ is Poisson distributed with mean $\nu_T(A)$ and for disjoint $A_1, ..., A_k \in \mathcal{B}(\mathbb{U}_T)$, $\mathbf{n}(A_1), ..., \mathbf{n}(A_k)$ are mutually independent random variables.

For any $m \geq 1$, let $\mathbb{W}_m = C([0, T], \mathbb{R}^m)$ be the space of all continuous functions from $[0, T]$ to $\mathbb{R}^m$. Correspondingly for the use later, $D([0, T], \mathcal{E})$ denotes the space of right continuous functions with left limits from $[0, T]$ to a Polish space $\mathcal{E}$. Take $\mathbb{V} = \mathbb{W}_m \times \mathbb{W}_n \times \mathbb{M}$ and let $\mathbb{P}$ be the probability measure on $(\mathbb{V}, \mathcal{B}(\mathbb{V}))$ such that (i) $N : \mathbb{V} \to \mathbb{M}$ is a Poisson random measure with intensity measure $\theta\nu_T$, and $\nu_T(A) < \infty$ for all $A \in \mathcal{B}(\mathbb{U}_T)$; (ii) $W : \mathbb{V} \to \mathbb{W}_m$ is a $\mathbb{R}^m$-valued Brownian motion and $B : \mathbb{V} \to \mathbb{W}_n$ is a $\mathbb{R}^n$-valued Brownian motion; and (iii) $\{W_t\}_{t\in[0,T]}, \{B_t\}_{t\in[0,T]}$ and $\{N([0, t] \times A), t \in [0, T]\}$ are $\mathcal{G}_t$-martingales for every $A \in \mathcal{B}(\mathbb{U})$, where the filtration $\mathcal{G}_t := \sigma\{N([0, t] \times A) - \theta t\nu(A), W_s, B_s : 0 \leq s \leq t, A \in \mathcal{B}(\mathbb{U})\}$.

To adopt the strategy of weak convergence arguments in order to prove the large deviations, we introduce a properly controlled Poisson random measure. Define $\mathbb{Y}_T = [0, T] \times \mathbb{Y}$, where $\mathbb{Y} = \mathbb{U} \times [0, \infty)$ and then denote $\bar{\mathbb{M}} = \mathcal{M}(\mathbb{Y}_T)$. Suppose $\bar{N}$ is a Poisson random measure on $\bar{\mathbb{V}} = \mathbb{W}_m \times \mathbb{W}_n \times \bar{\mathbb{M}}$ with intensity measure $\bar{\nu}_T = \lambda_T \otimes \nu \otimes \lambda_\infty$ where λ_∞ is the Lebesgue measure on $[0, \infty)$. Similarly abusing notations, B and W, are a Brownian motions on $\bar{\mathbb{V}}$. Next define $(\bar{\mathbb{P}}, \{\bar{\mathcal{G}}_t\})$ on $(\bar{\mathbb{V}}, \mathcal{B}(\bar{\mathbb{V}}))$ analogous to $(\mathbb{P}, \{\mathcal{G}_t\})$ by replacing $(N, \theta\nu_T)$ with $(\bar{N}, \bar{\nu}_T)$. Consider the $\bar{\mathbb{P}}$-completion of the filtration $\{\mathcal{G}_t\}$ and denote it by $\{\bar{\mathcal{F}}_t\}$. We denote by $\bar{\mathcal{P}}$ the predictable σ-field on $[0, T] \times \bar{\mathbb{V}}$ with the filtration $\{\bar{\mathcal{F}}_t : 0 \leq t \leq T\}$ on $(\bar{\mathbb{V}}, \mathcal{B}(\bar{\mathbb{V}}))$. Let $\bar{\mathcal{A}}$ be the class of all $(\bar{\mathcal{P}} \otimes \mathcal{B}(\mathbb{U})) \setminus \mathcal{B}[0, \infty)$ measurable maps $\phi : \mathbb{U}_T \times \bar{\mathbb{M}} \to [0, \infty)$. For the variable $\phi \in \bar{\mathcal{A}}$ which basically controls the intensity at time s on location u, define the counting process N^ϕ on $\mathbb{U}_T$ by

$$N^\phi((0,t] \times A) = \int_{(0,t]\times A\times(0,\infty)} 1_{[0,\phi(s,u)]}(r)\bar{N}(ds, du, dr) \tag{4.2.1}$$

where $t \in [0, T], A \in \mathcal{B}(\mathbb{U})$. If $\phi(s, u) = \theta$ for all $(s, u) \in \mathbb{U}_T$, then we write $N^\phi = N^\theta$, where N^θ has the same distribution on $\bar{\mathbb{M}}$ with respect to $\bar{\mathbb{P}}$ as N has on $\mathbb{M}$ with respect to $\mathbb{P}$. For any $\phi \in \bar{\mathcal{A}}$, the quantity

$$L_T(\phi) = \int_{\mathbb{U}_T} l(\phi(t, u))\ \nu(du)ds \tag{4.2.2}$$

is well-defined as a $[0, \infty]-$valued random variable where $l(r) := r \log r - r + 1, r \in [0, \infty)$. We denote by $L^2([0, T], \mathbb{R}^m)$ the Hilbert space from $[0, T]$ to $\mathbb{R}^m$ satisfying $|\psi(s)|^2 = \sum_{i=1}^m \psi_i(s)^2 < \infty$. Define

$$\mathcal{P}_2 = \left\{ \psi = (\psi_i)_{i=1}^m : \psi_i \text{ is } \bar{\mathcal{P}} \setminus \mathcal{B}(\mathbb{R}) \text{ measurable and } \int_0^T |\psi(s)|^2 ds < \infty, \bar{\mathbb{P}} - a.s. \right\}$$

and set $\mathcal{U} = \mathcal{P}_2 \times \bar{\mathcal{A}}$. For $\psi \in \mathcal{P}_2$ define

$$\tilde{L}_T(\psi) := \frac{1}{2} \int_0^T \psi(s)^2 ds \tag{4.2.3}$$

and for $u = (\psi, \phi) \in \mathcal{U}$, set

$$\bar{L}_T(u) = L_T(\phi) + \tilde{L}_T(\psi). \tag{4.2.4}$$

4.2.2 A general criterion of large deviations

Let $\epsilon > 0$, $x_0 \in \mathcal{E}_0$, where $\mathcal{E}_0$ is a Polish space and $\{X^{\epsilon,x_0}\}$ be $\mathcal{E}_0-$valued random variables defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The theory of small noise large deviations concerns with the asymptotic behavior of solutions of SPDEs which converge exponentially fast. The decay rate is expressed via a rate function. An equivalent argument of the large deviations principle is the Laplace principle. A reader may refer to [23, Theorem 1.2.1 and Theorem

1.2.3]. In this work, we establish a uniform large deviation principle with respect to the initial data, x_0, of the SPDE.

Definition 4.1. *A family of rate functions I_{x_0} on $\mathcal{E}$, parameterized by $x_0 \in \mathcal{E}_0$, is said to have compact level sets on compacts if for all compact subsets K of $\mathcal{E}_0$ and each $M < \infty$, $\Lambda_{M,K} \doteq \cup_{x_0 \in K}\{x \in \mathcal{E} : I_{x_0}(x) \leq M\}$ is a compact subset of $\mathcal{E}$.*

Definition 4.2. *Let I_{x_0} be a family of rate functions on $\mathcal{E}$ parameterized by $x_0 \in \mathcal{E}_0$ and assume that this family has compact level sets on compacts. The family $\{X^{\epsilon,x_0}\}$ is said to satisfy the Laplace principle on $\mathcal{E}$ with rate function I_{x_0}, uniformly on compacts, if for all compact subsets K of $\mathcal{E}_0$ and all bounded continuous functions h mapping $\mathcal{E}$ into $\mathbb{R}$,*

$$\lim_{\epsilon \to 0} \sup_{x_0 \in K} \left| \epsilon \log \mathbb{E}_{x_0} \left\{ \exp \left[-\frac{1}{\epsilon} h(X^{\epsilon,x_0}) \right] \right\} + \inf_{x \in \mathcal{E}} \{h(x) + I_{x_0}(x)\} \right| = 0.$$

Next, a set of sufficient conditions for a uniform large deviation principle for functionals of a Brownian motion and Poisson random measure is presented. Consider the family of measurable maps, for any $\epsilon > 0$, $\mathcal{G}^\epsilon : \mathcal{E}_0 \times \mathbb{W}_m \times \mathbb{M} \to \mathcal{E}$ defined as follows:

$$\{X^{\epsilon,x_0} \doteq \mathcal{G}^\epsilon(x_0, \sqrt{\epsilon}W, \epsilon N^{\epsilon^{-1}})\}.$$

Define the space, for some $M \in \mathbb{N}$, $\tilde{S}^M := \{\psi \in L^2([0,T], \mathbb{R}^m) : \tilde{L}_T(\psi) \leq M\}$ and $S^M := \{\phi : \mathbb{U}_T \to [0,\infty) : L_T(\phi) \leq M\}$, where L_T and $\tilde{L}_T$ are defined in Eqs. (4.2.2) and (4.2.3), respectively. For a function ϕ, define a measure $\nu_T^\phi \in \mathbb{M}$, such that

$$\nu_T^\phi(A) = \int_A \phi(s,u)\nu(du)ds, \quad A \in \mathcal{B}(\mathbb{U}_T).$$

Throughout we encapsulate the topology on S^M obtained through this identification which makes S^M a compact space. Let $\bar{S}^M = \tilde{S}^M \times S^M$ with the usual product topology and $\mathbb{S} = \cup_{M=1}^\infty \bar{S}^M$. Define the space of controls

$$\mathcal{U}^M = \{u = (\psi,\phi) \in \mathcal{U} : u(\omega) \in \bar{S}^M, \bar{\mathbb{P}} \text{ a.e. } \omega\}, \text{ where } \mathcal{U} = \mathcal{P}_2 \times \bar{\mathcal{A}}.$$

The following condition in [12, 55] plays a key role in proving uniform large deviation estimates for the filtering equation driven by a Brownian motion and an independent Poisson random measure. The corresponding rate function is given in Eq. (4.2.5) below.

Condition 4.2.1. *There exists a measurable map $\mathcal{G}^0 : \mathcal{E}_0 \times \mathbb{W}_m \times \mathbb{M} \to \mathcal{E}$ such that the following holds.*

1. *For $M \in \mathbb{N}$, let $(f_n, g_n), (f, g) \in \bar{S}^M$ and $x_n, x_0 \in \mathcal{E}_0$ be such that $(x_n, f_n, g_n) \to (x_0, f, g)$ as $n \to \infty$, then*

$$\mathcal{G}^0\left(x_n, \int_0^{\cdot} f_n(s)ds, \nu_T^{g_n}\right) \to \mathcal{G}^0\left(x_0, \int_0^{\cdot} f(s)ds, \nu_T^{g}\right).$$

2. *For $M \in \mathbb{N}$, let $\xi^\epsilon = (\psi^\epsilon, \phi^\epsilon), \xi = (\psi, \phi) \in \mathcal{U}^M$ and $x^\epsilon, x_0 \in \mathcal{E}_0$ be such that ξ^ϵ converges in distribution to ξ and $x^\epsilon \to x_0$, as $\epsilon \to 0$. Then*

$$\mathcal{G}^\epsilon(x^\epsilon, \sqrt{\epsilon}W + \int_0^{\cdot} \psi^\epsilon(s)ds, \epsilon N^{\epsilon^{-1}\phi^\epsilon}) \to \mathcal{G}^0(x_0, \int_0^{\cdot} \psi(s)ds, \nu_T^{\phi}), \text{ in distribution }.$$

For $\zeta \in \mathcal{E}$, define $\mathbb{S}_\zeta = \{u \in \mathbb{S} : \zeta = \mathcal{G}^0(x_0, \int_0^{\cdot} \psi(s)ds, \nu_T^{\phi})\}$. Let $I_{x_0} : \mathcal{E} \to [0, \infty]$ be defined by

$$I_{x_0}(\zeta) = \inf_{u \in \mathbb{S}_\zeta} \left\{\bar{L}_T(u)\right\}. \tag{4.2.5}$$

By convention, $I_{x_0}(\zeta) = \infty$ if $\mathbb{S}_\zeta = \emptyset$. Under Condition 4.2.1, we have the following theorem shown in [12].

Theorem 4.3. *For $\epsilon > 0$, let $X^{\epsilon,x_0} \doteq \mathcal{G}^\epsilon(x_0, \sqrt{\epsilon}W, \epsilon N^{\epsilon^{-1}})$, and suppose that Condition 4.2.1 holds. Then for every $x_0 \in \mathcal{E}_0, I_{x_0} : \mathcal{E} \to [0, \infty]$, defined by Eq. (4.2.5), is a rate function on $\mathcal{E}$ and the family $\{X^{\epsilon,x_0}\}$ satisfies the Laplace principle on $\mathcal{E}$ (equivalently the large deviation principle) with rate function I_{x_0} uniformly on compact subsets of $\mathcal{E}_0$.*

4.3 Existence and uniqueness of filtering equations

4.3.1 Existence

This section first presents the filtering model and filtering equations for the system driven by Lévy noise. Consider the following signal-observation system (X_t, Y_t) on $\mathbb{R}^d \times \mathbb{R}^m$:

$$dX_t = b_1(X_t)dt + \sigma_1(X_t)dB_t + \int_{\mathbb{U}_1} f_1(X_{t-}, u)\tilde{N}_p(dt, du) + \int_{\mathbb{U}\setminus\mathbb{U}_1} g_1(X_{t-}, u)N_p(dt, du), \tag{4.3.1a}$$

$$dY_t = b_2(t, X_t)dt + \sigma_2(t)dW_t + \int_{\mathbb{U}_2} f_2(t, u)\tilde{N}_\lambda(dt, du) + \int_{\mathbb{U}\setminus\mathbb{U}_2} g_2(t, u)N_\lambda(dt, du), \tag{4.3.1b}$$

where B_t, W_t are n-dimensional and m-dimensional Brownian motions respectively defined on the filtered probability space $(\mathbb{V}, \mathcal{B}(\mathbb{V}), \mathbb{P}, \{\mathcal{F}_t\}_{t\in[0,T]})$. N_p is a Poisson random measure such that $\mathbb{E}N_p([0,t]\times A) = t\nu_1(A)$, where ν_1 is a σ-finite measure on $\mathbb{U}$. For any $A \in \mathcal{B}(\mathbb{U})$, $\nu_1(A) < \infty$ and $\int_{\mathbb{U}_1} \|u\|_{\mathbb{U}}^2 \nu_1(du) < \infty$ where $\|\cdot\|_{\mathbb{U}}$ denotes the norm on the measurable space $(\mathbb{U}, \mathcal{B}(\mathbb{U}))$ and $\mathbb{U}_1 \subset \mathbb{U}$. Denote the compensated measure $\tilde{N}_p([0,t] \times du) = N_p([0,t], du) - t\nu_1(du)$. Let $N_\lambda([0,t]\times du)$ be an integer-valued random measure and its predictable compensator is given by $\lambda(t, X_{t-}, u)t\nu_2(du)$, where the function $\lambda(t, x, u) \in [l, 1), 0 < l < 1$, and $\tilde{N}_\lambda([0,t] \times du) = N_\lambda([0,t] \times du) - \lambda(t, X_{t-}, u)t\nu_2(du)$ such that ν_2 is a σ-finite measure on $\mathbb{U}$ and for each $A \in \mathcal{B}(\mathbb{U})$, $\nu_2(A) < \infty$ and $\int_{\mathbb{U}_2} \|u\|_{\mathbb{U}}^2 \nu_2(du) < \infty$ with $\mathbb{U}_2 \subset \mathbb{U}$. Moreover, B_t, W_t, N_p, N_λ are mutually independent.

We assume that the mappings $b_1 : \mathbb{R}^d \to \mathbb{R}^d, b_2 : [0,T] \times \mathbb{R}^d \to \mathbb{R}^m, \sigma_1 : \mathbb{R}^d \to \mathbb{R}^{d\times n}, \sigma_2 : [0,T] \to \mathbb{R}^{m\times m}, f_1 : \mathbb{R}^d \times \mathbb{U}_1 \to \mathbb{R}^d, f_2 : [0,T] \times \mathbb{U}_2 \to \mathbb{R}^m, g_1 : \mathbb{R}^d \times (\mathbb{U} \setminus \mathbb{U}_1) \to \mathbb{R}^d$, and $g_2 : [0,T] \times (\mathbb{U} \setminus \mathbb{U}_2) \to \mathbb{R}^m$ are all Borel measurable, and satisfy the following conditions:

Assumption 4.3.1. *For each $x_1, x_2 \in \mathbb{R}^d$, there exists a constant K such that*

$$|b_1(x_1) - b_1(x_2)|^2 + |\sigma_1(x_1) - \sigma_1(x_2)|^2 + \int_{\mathbb{U}_1} |f_1(x_1, u) - f_1(x_2, u)|^2 \nu_1(du) \le K|x_1 - x_2|^2.$$

where $|\cdot|$ denotes the Hilbert-Schmidt norm for a matrix and the length for a vector.

Assumption 4.3.2. *$\sigma_2(t)$ is invertible for $t \in [0,T]$, and for each $x \in \mathbb{R}^d$,*

$$|\sigma_1(x)|^2 + \int_{\mathbb{U}_1} |f_1(x,u)|^2 \nu_1(du) + |b_2(t,x)|^2 + |\sigma_2^{-1}(t)|^2 + \int_{\mathbb{U}_2} |f_2(t,u)|^2 \nu_2(du) \le K.$$

Set $\pi_t(F) := \mathbb{E}(F(X_t)|\mathcal{F}_t^Y)$, for any $F \in C_b^2(\mathbb{R}^d)$, where $C_b^2(\mathbb{R}^d)$ denotes the set of all bounded functions $F : \mathbb{R}^d \to \mathbb{R}^d$ that have continuous second order derivative. Taking into account the Radon-Nikodym derivative (see [66] and [37, Theorem 3.17], also [9, Theorem 3.12], [69] or [10, Thm. VIII, T10]), there is an equivalent probability measure $\tilde{\mathbb{P}}$ such that, $d\tilde{\mathbb{P}} = \Lambda_T^{-1} d\mathbb{P}$, where

$$\begin{aligned}\Lambda_T^{-1} = \exp\Bigg(& -\int_0^T \sigma_2^{-1}(s) b_2(s,X_s)^* dW_s - \frac{1}{2}\int_0^T |\sigma_2^{-1}(s) b_2(s,X_s)|^2 ds \\ & - \int_0^T \int_{\mathbb{U}_2} \log \lambda(s,X_{s-},u) N_\lambda(ds,du) - \int_0^T \int_{\mathbb{U}_2} (1-\lambda(s,X_s,u))\nu_2(du)ds \Bigg). \qquad (4.3.2)\end{aligned}$$

where $*$ denotes the transpose operator. In turn, the Kallianpur-Striebel formula [77] gives $\pi_t(F) = \frac{\mu_t(F)}{\mu_t(1)}$, where $\mu_t(F) := \tilde{\mathbb{E}}(F(X_t)\Lambda_t|\mathcal{F}_t^Y)$ and $\tilde{\mathbb{E}}$ denotes expectation under the measure $\tilde{\mathbb{P}}$. The existence of the optimal filter, derived in [66, Theorem 3.1 and 3.2], is given by the following two SPDEs respectively.

Proposition 4.4 (Zakai equation)**.** *Assume Assumptions 4.3.1 and 4.3.2. For $F \in D(\mathcal{L})$, the Zakai equation of Eq. (4.3.1) is given by*

$$\begin{aligned}\mu_t(F) = \mu_0(F) + \int_0^t \mu_s(\mathcal{L}F)ds + \sum_{i=1}^m \int_0^t \mu_s(F(\sigma_2^{-1}(s)b_2(s,\cdot))^i) d\tilde{W}_s^i \qquad (4.3.3) \\ + \int_0^t \int_{\mathbb{U}_2} \mu_{s-}(F(\lambda(s,\cdot,u)-1))\tilde{N}(ds,du),\end{aligned}$$

where $\tilde{W}_t = W_t + \int_0^t \sigma_2^{-1}(s) b_2(s, X_s) ds$ *and* $\tilde{N}(dt, du) = N_\lambda(dt, du) - dt\nu_2 d(u)$*, and for any* $f \in C_b^2(\mathbb{R}^d)$*, the infinitesimal generator,* $\mathcal{L}$*, is given by*

$$\begin{aligned}\mathcal{L}f(x) = &\sum_{i=1}^{d} \frac{\partial f(x)}{\partial x_i} b_1^i(x) + \frac{1}{2} \sum_{i,j=1}^{d} \sum_{k=1}^{n} \frac{\partial^2 f(x)}{\partial x_i \partial x_j} \sigma_1^{ik}(x) \sigma_1^{jk}(x) \\ &+ \int_{\mathbb{U} \setminus \mathbb{U}_1} [f(x + g_1(x, u)) - f(x)] \nu_1(du) \\ &+ \int_{\mathbb{U}_1} [f(x + f_1(x, u)) - f(x) - \sum_{i=1}^{d} \frac{\partial f(x)}{\partial x_i} f_1^i(x, u)] \nu_1(du),\end{aligned} \tag{4.3.4}$$

where σ_1^{ik} *is the* (i,k)*th entry of the diffusion coefficient* σ_1*.*

Proposition 4.5 (Kushner-Stratonovich equation)**.** *Assume Assumptions 4.3.1 and 4.3.2. For* $F \in D(\mathcal{L})$*, the solution of the following equation exists*

$$\begin{aligned}\pi_t(F) = &\pi_0(F) + \int_0^t \pi_s(\mathcal{L}F) ds \\ &+ \sum_{i=1}^{m} \int_0^t (\pi_s(F(\sigma_2^{-1}(s) b_2(s, \cdot))^i) - \pi_s(F) \pi_s(\sigma_2^{-1}(s) b_2(s, \cdot))^i) d\hat{W}_s^i \\ &+ \int_0^t \int_{\mathbb{U}_2} \frac{\pi_{s-}(F\lambda(s, \cdot, u)) - \pi_{s-}(F) \pi_{s-}(\lambda(s, \cdot, u))}{\pi_{s-}(\lambda(s, \cdot, u))} \hat{N}(ds, du),\end{aligned} \tag{4.3.5}$$

where $\hat{W}_t = \tilde{W}_t - \int_0^t \pi_s(\sigma_2^{-1}(s) b_2(s, \cdot)) ds$ *is the innovation process and* $\hat{N}(dt, du) = N_\lambda(dt, du) - \pi_{t-}(\lambda(t, \cdot, u)) \nu_2(du) dt$*.*

4.3.2 Uniqueness

In this section, we prove the uniqueness for the solutions to the Zakai and Kushner-Stratonovich equations for the signal-observation model (4.3.1). Although the uniqueness was investigated in [66], it was assumed that the Poisson noise in the observation is independent of the signal, i.e., $\lambda(t, x, u) = \lambda(t, u)$. This reduces the complexity of the Zakai equation; that is, the Zakai equation is independent of the Poisson noise. This makes the problem more tractable since the Poissonian part in Eq. (4.3.5) vanishes, see [66, Section 4]. Furthermore, therein, the uniqueness of the Filtered Martingale Problem is assumed, and regularity conditions on the coefficients of the signal and observation processes are required (see [66, Remark 4.1] and [47]). Next, we show the uniqueness of Zakai and Kushner-Stratonovich equations by bypassing the above restrictive assumptions and instead imposing the following mild assumption.

Assumption 4.3.3. $|\det(J_{f_1}+I)| > \frac{1}{C}$ *and* $|\det(J_{g_1}+I)| > \frac{1}{C}$ *for a constant* $C > 0$, *where* J_{f_1} *and* J_{g_1} *are the Jacobian matrices of* f_1 *and* g_1 *with respect to* x*, respectively.*

The uniqueness for the solution to Zakai equation is proved by transforming it to an SDE in a pertinent Hilbert space and by making use of estimates based on Hilbert-space techniques, which was studied in [48, 77]. Recall that the optimal filter $\mathbb{E}(F(X_t)|\mathcal{F}_t^Y)$ is the solution to the filtering Eq. (4.3.5) characterized by the conditional probability π_t. Denote $\mathcal{P}(\mathbb{R}^d)$ the collection of all Borel probability measures on $\mathbb{R}^d$ such that $\pi_t \in \mathcal{P}(\mathbb{R}^d)$. Denote by $\langle \nu, F\rangle$ the integral of a function F with respect to a measure ν, e.g., for any $F \in C_b(\mathbb{R}^d)$, $\mathbb{E}(F(X_t)|\mathcal{F}_t^Y) = \langle \pi_t, F\rangle$. Let $\mathcal{M}_F(\mathbb{R}^d)$ be the collection of all finite Borel measures on $\mathbb{R}^d$ such that the unnormalized filter μ_t is an $\mathcal{M}_F(\mathbb{R}^d)$-valued process. Let $H_0 = L^2(\mathbb{R}^d)$ be the Hilbert space consisting of square-integrable functions on $\mathbb{R}^d$ with the usual L^2-norm and the inner product given by

$$\|\phi\|_0^2 = \int_{\mathbb{R}^d} |\phi(x)|^2 dx \quad \text{and} \quad \langle \phi, \psi\rangle_0 = \int_{\mathbb{R}^d} \phi(x)\psi(x)dx.$$

We introduce an operator to transform a measure-valued process to an H_0-valued process. Denote by $\mathcal{M}_G(\mathbb{R}^d)$ the space of finite signed measures on $\mathbb{R}^d$. For any $\nu \in \mathcal{M}_G(\mathbb{R}^d)$ and $\delta > 0$, let

$$(T_\delta \nu)(x) = \int_{\mathbb{R}^d} G_\delta(x-y)\nu(dy), \tag{4.3.6}$$

where G_δ is the heat kernel given by

$$G_\delta(x) = (2\pi\delta)^{-\frac{d}{2}} \exp\left(-\frac{|x|^2}{2\delta}\right).$$

We use the same notation as in Eq. (4.3.6) for the Brownian motion semigroup on H_0, i.e., for $t \geq 0$, define operator $T_t : H_0 \to H_0$ by

$$T_t\phi(x) = \int_{\mathbb{R}^d} G_t(x-y)\phi(y)dy,$$

for any $\phi \in H_0$. Then Lemma 4.6 below obtains bounds for the partial derivative of T_δ, and Lemma 4.7 is directly applied to Theorem 4.8 below. The reader should refer to [77] for the proofs of these two lemmas.

Lemma 4.6. *(i) The family of operators $\{T_t : t \geq 0\}$ forms a contraction semigroup on H_0, i.e. for any $t, s \geq 0$ and $\phi \in H_0$, we have $T_{t+s} = T_t T_s$ and $\|T_t\phi\|_0 \leq \|\phi\|_0$.*

(ii) If $\nu \in \mathcal{M}_G(\mathbb{R}^d)$ and $\delta > 0$, then $T_\delta \nu \in H_0$.

(iii) If $\nu \in \mathcal{M}_G(\mathbb{R}^d)$ and $\delta > 0$, then $\|T_{2\delta}|\nu|\|_0 \leq \|T_\delta|\nu|\|_0$, where $|\nu|$ is the total variation measure of ν.

Lemma 4.7. *For any $\delta > 0$, $\nu \in \mathcal{M}_G(\mathbb{R}^d)$ and $\phi \in H_0$, we have*

(i)
$$\langle T_\delta \nu, \phi \rangle_0 = \langle \nu, T_\delta \phi \rangle. \tag{4.3.7}$$

(ii) If, in addition, $\partial_i \phi \in H_0$, where $\partial_i \phi = \frac{\partial \phi}{\partial x_i}$, then
$$\partial_i T_\delta \phi = T_\delta \partial_i \phi. \tag{4.3.8}$$

The next theorem presents an expression for the expectation of the transformation applying to the solution to Zakai equation.

Theorem 4.8. *Let $\mu_t \in \mathcal{M}_F(\mathbb{R}^d)$ be a solution to Zakai Eq. (4.3.3) and let $Z_t^\delta = T_\delta \mu_t$. Considering the probability measure $\tilde{\mathbb{P}}$ defined by (4.3.2), the following holds.*

$$\tilde{\mathbb{E}}\|Z_t^\delta\|_0^2 = A_1 - 2A_2 + A_3 + 2A_4 + A_5 + A_6 + A_7, \tag{4.3.9}$$

where

$$
\begin{aligned}
&A_1 = \|Z_0^\delta\|_0^2, \quad A_2 = \sum_{i=1}^d \int_0^t \tilde{\mathbb{E}}\langle Z_s^\delta, \partial_i T_\delta(b_1^i \mu_s)\rangle_0 ds,\\
&A_3 = \sum_{i,j=1}^d \sum_{k=1}^n \int_0^t \tilde{\mathbb{E}}\langle Z_s^\delta, \partial_{ij}^2 T_\delta(\sigma_1^{ik}\sigma_1^{kj}\mu_s)\rangle_0 ds,\\
&A_4 = \int_0^t \int_{\mathbb{U}\setminus\mathbb{U}_1} [\tilde{\mathbb{E}}\langle Z_s^\delta, \tilde{Z}_s^{\delta,g_1}\rangle_0 - \tilde{\mathbb{E}}\|Z_s^\delta\|_0^2]\nu_1(du)ds, \qquad (4.3.10)\\
&A_5 = \int_0^t \int_{\mathbb{U}_1} [\tilde{\mathbb{E}}\langle Z_s^\delta, \tilde{Z}_s^{\delta,f_1}\rangle_0 - \tilde{\mathbb{E}}\|Z_s^\delta\|_0^2 - \sum_{i=1}^d \tilde{\mathbb{E}}\langle Z_s^\delta, \partial_i T_\delta(f_1^i(\cdot,u)\mu_s)\rangle_0]\nu_1(du)ds,\\
&A_6 = \int_0^t \tilde{\mathbb{E}}\|\langle T_\delta(\sigma_2^{-1}(s)b_2(s,\cdot)\mu_s)\|_0^2 ds, \ A_7 = \int_0^t \int_{\mathbb{U}_2} \tilde{\mathbb{E}}\|T_\delta\left((\lambda(s,\cdot,u)-1)\mu_s\right)\|_0^2 \nu_2(du)ds;
\end{aligned}
$$

and

$$
\langle Z_t^\delta, \tilde{Z}_t^{\delta,f_1}\rangle_0 = \sum_\eta \langle Z_t^\delta, \eta\rangle_0 \langle Z_t^\delta, \eta(\cdot + f_1(\cdot,u))\rangle_0, \qquad (4.3.11)
$$

$$
\langle Z_t^\delta, \tilde{Z}_t^{\delta,g_1}\rangle_0 = \sum_\eta \langle Z_t^\delta, \eta\rangle_0 \langle Z_t^\delta, \eta(\cdot + g_1(\cdot,u))\rangle_0, \qquad (4.3.12)
$$

here the set of functions $\{\eta\}$ is a complete orthonormal system (CONS) of H_0.

Proof. Considering Eq. (4.3.7) and replacing F by $T_\delta F$ in the Zakai Eq. (4.3.3), we have that

$$
\begin{aligned}
\langle Z_s^\delta, F\rangle_0 =& \langle \mu_0, T_\delta F\rangle + \int_0^t \langle \mu_s, \mathcal{L} T_\delta F\rangle ds + \sum_{i=1}^m \int_0^t \langle \mu_s, (T_\delta F)(\sigma_2^{-1}(s)b_2(s,\cdot))^i\rangle d\tilde{W}_s^i\\
&+ \int_0^t \int_{\mathbb{U}_2} \langle \mu_{s-}, (T_\delta F)(\lambda(s,\cdot,u)-1)\rangle \tilde{N}(ds,du). \qquad (4.3.13)
\end{aligned}
$$

Note that for any $\nu \in \mathcal{M}_F(\mathbb{R}^d)$,

$$\begin{aligned}\langle \nu, \mathcal{L}T_\delta F\rangle &= \langle \nu, \sum_{i=1}^d \partial_i(T_\delta F)b_1^i + \frac{1}{2}\sum_{i,j=1}^d\sum_{k=1}^n \partial_{ij}^2(T_\delta F)\sigma_1^{ik}\sigma_1^{jk}\rangle \\ &+ \langle \nu, \int_{\mathbb{U}\setminus\mathbb{U}_1}[T_\delta F(x+g_1(x,u)) - T_\delta F(x)]\nu_1(du)\rangle \\ &+ \langle \nu, \int_{\mathbb{U}_1}[T_\delta F(x+f_1(x,u)) - T_\delta F(x) - \sum_{i=1}^d \partial_i(T_\delta F)f_1^i(x,u)]\nu_1(du)\rangle. \end{aligned} \tag{4.3.14}$$

Furthermore, by Lemma 4.7, we have

$$\begin{aligned}&\langle \nu, \sum_{i=1}^d \partial_i(T_\delta F)b_1^i + \frac{1}{2}\sum_{i,j=1}^d\sum_{k=1}^n \partial_{ij}^2(T_\delta F)\sigma_1^{ik}\sigma_1^{jk}\rangle \\ &\qquad = \sum_{i=1}^d \langle b_1^i\nu, T_\delta\partial_i F\rangle + \frac{1}{2}\sum_{i,j=1}^d\sum_{k=1}^n\langle \sigma_1^{ik}\sigma_1^{jk}\nu, T_\delta\partial_{ij}^2 F\rangle \\ &\qquad = \sum_{i=1}^d \langle T_\delta(b_1^i\nu), \partial_i F\rangle_0 + \frac{1}{2}\sum_{i,j=1}^d\sum_{k=1}^n\langle T_\delta(\sigma_1^{ik}\sigma_1^{jk}\nu), \partial_{ij}^2 F\rangle_0 \\ &\qquad = -\sum_{i=1}^d \langle \partial_i T_\delta(b_1^i\nu), F\rangle_0 + \frac{1}{2}\sum_{i,j=1}^d\sum_{k=1}^n\langle \partial_{ij}^2 T_\delta(\sigma_1^{ik}\sigma_1^{jk}\nu), F\rangle_0. \end{aligned} \tag{4.3.15}$$

In addition,

$$\langle \nu, (T_\delta F)(\sigma_2^{-1}(s)b_2(s,\cdot))^i\rangle = \langle T_\delta(\sigma_2^{-1}(s)b_2(s,\cdot)^i\nu), F\rangle_0,$$

where $(T_\delta F)(\sigma_2^{-1}(s)b_2(s,\cdot))^i$ is the ith entry of vector $(T_\delta F)(\sigma_2^{-1}(s)b_2(s,\cdot))$, as well as

$$\langle \nu, (T_\delta F)(\lambda(s,\cdot,u)-1)\rangle = \langle T_\delta((\lambda(s,\cdot,u)-1)v), F\rangle_0.$$

Suppose that $F \geq 0$, then the case for general F follows from the linearity. By Fubini's theorem,

$$\begin{aligned}&\langle \nu, \int_{\mathbb{U}\setminus\mathbb{U}_1} T_\delta F(x+g_1(x,u))\nu_1(du)\rangle = \int_{\mathbb{R}^d}\int_{\mathbb{U}\setminus\mathbb{U}_1} T_\delta F(x+g_1(x,u))\nu_1(du)\nu(dx) \\ &= \int_{\mathbb{U}\setminus\mathbb{U}_1}\int_{\mathbb{R}^d} T_\delta F(x+g_1(x,u))\nu(dx)\nu_1(du) = \int_{\mathbb{U}\setminus\mathbb{U}_1}\langle \nu, T_\delta F(x+g_1(x,u))\rangle\nu_1(du). \end{aligned} \tag{4.3.16}$$

Consequently, taking into consideration Eqs. (4.3.14), (4.3.15) and (4.3.16) into Eq. (4.3.13), one can get

$$\begin{aligned}
\langle Z_s^\delta, F\rangle_0 &= \langle Z_0^\delta, F\rangle_0 - \sum_{i=1}^{d}\int_0^t \langle \partial_i T_\delta(b_1^i\mu_s), F\rangle_0 ds \\
&+ \frac{1}{2}\sum_{i,j=1}^{d}\sum_{k=1}^{n}\int_0^t \langle \partial_{ij}^2 T_\delta(\sigma_1^{ik}\sigma_1^{jk}\mu_s), F\rangle_0 ds \\
&+ \int_0^t\int_{\mathbb{U}\setminus\mathbb{U}_1} \left[\langle T_\delta\mu_s, F(\cdot + g_1(\cdot,u))\rangle_0 - \langle T_\delta\mu_s, F\rangle_0\right]\nu_1(du)ds \\
&+ \sum_{i=1}^{m}\int_0^t \langle T_\delta(\sigma_2^{-1}(s)b_2(s,\cdot))^i\mu_s), F\rangle_0 d\tilde{W}_s^i \\
&+ \int_0^t\int_{\mathbb{U}_2} \langle T_\delta\left((\lambda(s,\cdot,u)-1)\mu_{s-}\right), F\rangle_0 \tilde{N}(ds,du \\
&+ \int_0^t\int_{\mathbb{U}_1}\left[\langle T_\delta\mu_s, F(\cdot + f_1(\cdot,u))\rangle_0 - \langle T_\delta\mu_s, F\rangle_0 - \sum_{i=1}^{d}\langle \partial_i T_\delta(f_1^i(\cdot,u)\mu_s), F\rangle_0\right]\nu_1(du)ds.
\end{aligned}$$

Then by Itô's formula, we have $\langle Z_t^\delta, F\rangle_0^2$ equals to

$$\begin{aligned}
&\langle Z_0^\delta, F\rangle_0^2 + 2\int_0^t \langle Z_s^\delta, F\rangle_0 \Bigg\{ -\sum_{i=1}^{d}\langle \partial_i T_\delta(b_1^i\mu_s), F\rangle_0 + \frac{1}{2}\sum_{i,j=1}^{d}\sum_{k=1}^{n}\langle \partial_{ij}^2 T_\delta(\sigma_1^{ik}\sigma_1^{jk}\mu_s), F\rangle_0 \\
&+ \int_{\mathbb{U}\setminus\mathbb{U}_1}\left[\langle Z_s^\delta, F(\cdot + g_1(\cdot,u))\rangle_0 - \langle Z_s^\delta, F\rangle_0\right]\nu_1(du) \\
&+ \int_{\mathbb{U}_1}\left[\langle Z_s^\delta, F(\cdot + f_1(\cdot,u))\rangle_0 - \langle Z_s^\delta, F\rangle_0 - \langle \partial_i T_\delta(f_1^i(\cdot,u)\mu_s), F\rangle_0\right]\nu_1(du)\Bigg\} ds \\
&+ 2\sum_{i=1}^{m}\int_0^t \langle Z_s^\delta, F\rangle_0 \langle T_\delta((\sigma_2^{-1}(s)b_2(s,\cdot))^i\mu_s), F\rangle_0 d\tilde{W}_s^i + \int_0^t \left|\langle T_\delta(\sigma_2^{-1}(s)b_2(s,\cdot)\mu_s), F\rangle_0\right|^2 ds \\
&+ \int_0^t\int_{\mathbb{U}_2}\left[\left(\langle Z_{s-}^\delta, F\rangle_0 + \langle T_\delta\left((\lambda(s,\cdot,u)-1)\mu_{s-}\right), F\rangle_0\right)^2 - \langle Z_{s-}^\delta, F\rangle_0^2\right]\tilde{N}(ds,du) \\
&+ \int_0^t\int_{\mathbb{U}_2}\Big[\left(\langle Z_s^\delta, F\rangle_0 + \langle T_\delta\left((\lambda(s,\cdot,u)-1)\mu_s\right), F\rangle_0\right)^2 - \langle Z_s^\delta, F\rangle_0^2 \\
&-2\langle T_\delta\left((\lambda(s,\cdot,u)-1)\mu_s\right), F\rangle_0\langle Z_s^\delta, F\rangle_0\Big]\nu_2(du)ds.
\end{aligned}$$

Summing over η in a CONS of H_0 we have

$$\sum_{\eta}\langle Z_0^\delta,\eta\rangle_0^2=\|Z_0^\delta\|_0^2,\qquad \sum_{\eta}\langle Z_t^\delta,\eta\rangle_0\langle\partial_iT_\delta(b_1^i\mu_t),\eta\rangle_0=\langle Z_t,\partial_iT_\delta(b_1^i\mu_t)\rangle_0.$$

Similarly, we eventually get

$$\begin{aligned}\|Z_t^\delta\|_0^2&=\|Z_0^\delta\|_0^2-2\sum_{i=1}^d\int_0^t\langle Z_s^\delta,\partial_iT_\delta(b_1^i\mu_s)\rangle_0ds+\sum_{i,j=1}^d\sum_{k=1}^n\int_0^t\langle Z_s^\delta,\partial_{ij}^2T_\delta(\sigma_1^{ik}\sigma_1^{jk}\mu_s)\rangle_0ds\\&+2\int_0^t\int_{\mathbb{U}\setminus\mathbb{U}_1}[\langle Z_s^\delta,\tilde{Z}_s^{\delta,g_1}\rangle_0-\|Z_s^\delta\|_0^2]\nu_1(du)ds\\&+\int_0^t\int_{\mathbb{U}_1}[\langle Z_s^\delta,\tilde{Z}_s^{\delta,f_1}\rangle_0-\|Z_s^\delta\|_0^2-\sum_{i=1}^d\langle Z_s^\delta,\partial_iT_\delta(f_1^i(\cdot,u)\mu_s)\rangle_0]\nu_1(du)ds\\&+2\sum_{i=1}^m\int_0^t\langle Z_s^\delta,T_\delta((\sigma_2^{-1}(s)b_2(s,\cdot))^i\mu_s)\rangle_0d\tilde{W}_s^i+\int_0^t\|T_\delta(\sigma_2^{-1}(s)b_2(s,\cdot)\mu_s)\|_0^2ds\\&+\int_0^t\int_{\mathbb{U}_2}[2\langle Z_{s-}^\delta,T_\delta\left((\lambda(s,\cdot,u)-1)\mu_s\right)\rangle_0+\|T_\delta\left((\lambda(s,\cdot,u)-1)\mu_{s-}\right)\|_0^2]\tilde{N}(ds,du)\\&+\int_0^t\int_{\mathbb{U}_2}\|T_\delta\left((\lambda(s,\cdot,u)-1)\mu_s\right)\|_0^2\nu_2(du)ds,\end{aligned}\tag{4.3.17}$$

provided $Z_t^{\delta,g_1}, Z_t^{\delta,f_1}$ are defined as Eqs. (4.3.11) and (4.3.12) respectively. Eventually, taking expectation on Eq. (4.3.17) gives Eq. (4.3.9). ■

In order to get estimates of the terms as defined in Eq. (4.3.10), we proceed with the following lemmas.

Lemma 4.9. *Suppose Assumption 4.3.3 holds. Then there exists a constant $C_0>0$ such that*

$$\left|\langle Z_t^\delta,\tilde{Z}_t^{\delta,f_1}\rangle_0\right|\le C_0\|Z_t^\delta\|_0^2,\tag{4.3.18}$$

and

$$\left|\langle Z_t^\delta,\tilde{Z}_t^{\delta,g_1}\rangle_0\right|\le C_0\|Z_t^\delta\|_0^2.\tag{4.3.19}$$

Proof. We only need to show inequality (4.3.18). Let $y = x + f_1(x,u)$ and assume $x = h(y,u)$. Denote the Jacobian of h with respect to y by $|\partial_y h(y,u)|$. According to Assumption 4.3.3, $|\det(I + J_{f_1})| > \frac{1}{C}$, then

$$|\partial_y h(y,u)| = |\det(I + J_{f_1})|^{-1} \le C. \tag{4.3.20}$$

Note that

$$\langle Z_t^\delta, \eta(\cdot + f_1(\cdot,u))\rangle_0 = \int_{\mathbb{R}^d} Z_t^\delta(h(y,u))\eta(y)|\partial_y h(y,u)|dy = \langle Z_t^\delta(h(\cdot,u))|\partial_y h(\cdot,u)|, \eta\rangle_0.$$

Furthermore, summing over $\{\eta\}$ in a CONS of H_0, we have

$$\sum_\eta \langle Z_t^\delta(h(\cdot,u))|\partial_y h(\cdot,u)|, \eta\rangle_0 \langle Z_t^\delta, \eta\rangle_0 = \langle Z_t^\delta(h(\cdot,u))|\partial_y h(\cdot,u)|, Z_t^\delta\rangle_0.$$

Hence, by Eq. (4.3.20) we get

$$\begin{aligned}\|Z_t^\delta(h(y,u))|\partial_y h(y,u)|\|_0^2 &= \int_{\mathbb{R}^d} |T_\delta \mu_t(h(y,u))|^2 \, |\partial_y h(y,u)|^2 dy \\ &= \int_{\mathbb{R}^d} |T_\delta \mu_t(x)|^2 \, |\partial_y h(y,u)| dx \le C\|Z_t^\delta\|_0^2,\end{aligned} \tag{4.3.21}$$

and the bound of Eq. (4.3.18) then follows from the Cauchy-Schwarz inequality. ■

Lemma 4.10 below, verified in [48, Lemma 3.2], is useful to estimate the transformation T_δ and the derivatives of T_δ in Lemma 4.11 and Theorem 4.12.

Lemma 4.10. *Let $(H, \mathcal{H}, \eta)$ be a measure space and $\mathbb{H} = Ł_2(\eta)$. Let $\phi_i : \mathbb{R}^d \to \mathbb{H}, i = 1,2$ such that there exists a constant $K > 0$, for any $x, y \in \mathbb{R}^d$, $\|\phi_i(x)\|_{\mathbb{H}} \le K$. Let $\zeta \in \mathcal{M}_G(\mathbb{R}^d)$. Then there exists a constant $K_1 \equiv K_1(\phi_i)$ such that*

$$\|\|T_\delta(\phi_1\zeta)\|_0\|_{\mathbb{H}} \le K_1\|T_\delta(|\zeta|)\|_0. \tag{4.3.22}$$

If, additionally, (We are interested in $H = \mathbb{U}$ with $\eta = \nu_1$) $\|\phi_i(x) - \phi_i(y)\|_{\mathbb{H}} \le K|x-y|$, then

$$|\langle T_\delta(\phi_2\zeta), \partial_i T_\delta(\phi_1\zeta)\rangle_{H_0\otimes\mathbb{H}}| \le K_1 \| |T_\delta(|\zeta|)\|_0^2, \tag{4.3.23}$$

where $\otimes$ denotes convolution.

The next lemma gives a bound of the derivatives of the transformation on σ_1.

Lemma 4.11. *There exists constant K_1 such that for any $\zeta \in \mathcal{M}_G(\mathbb{R}^d)$, we have*

$$\sum_{i,j=1}^d \langle T_\delta\zeta, \partial_{ij}^2 T_\delta(\sigma_1\sigma_1^*)_{ij}\zeta\rangle_0 + \sum_{i=1}^n \left\| \sum_{i=1}^d \partial_i T_\delta(\sigma_1^{ik}\zeta)\right\|_0^2 \le K_1 \|T_\delta(|\zeta|)\|_0^2. \tag{4.3.24}$$

Proof. Note that $\sum_{i,j=1}^d \langle T_\delta\zeta, \partial_{ij}^2 T_\delta(\sigma_1\sigma_1^*)_{ij}\zeta\rangle_0$ equals to

$$\sum_{i,j=1}^d \int_{\mathbb{R}^d} dx \int_{\mathbb{R}^d} \zeta(dy) G_\delta(x-y) \int_{\mathbb{R}^d} \zeta(dz) \partial_{x_i x_j}^2 G_\delta(x-z) \sum_{k=1}^n \sigma_1^{ik}(z)\sigma_1^{jk}(z). \tag{4.3.25}$$

Employing the semigroup property of G_δ in Lemma 4.6, we have

$$\int_{\mathbb{R}^d} G_\delta(x-y) G_\delta(x-z) dx = G_{2\delta}(y-z).$$

Noticing that $\partial_i G_\delta(x) = -\frac{x_i}{\delta} G_\delta(x)$, we have

$$\partial_{ij}^2 G_\delta(x) = \left(\frac{x_i x_j}{\delta^2} - \frac{1_{i=j}}{\delta} \right) G_\delta(x).$$

Due to the fact that

$$\partial_{x_i x_j}^2 G_\delta(x-z) = \partial_{z_i z_j}^2 G_\delta(x-z),$$

Eq. (4.3.25) is written as

$$\sum_{i,j=1}^{d}\int_{\mathbb{R}^d}\zeta(dy)\int_{\mathbb{R}^d}\zeta(dz)\sum_{k=1}^{n}\sigma_1^{ik}(z)\sigma_1^{kj}(z)\partial^2_{z_iz_j}\int_{\mathbb{R}^d}G_\delta(x-y)G_\delta(x-z)dx$$
$$=\sum_{i,j=1}^{d}\int_{\mathbb{R}^d}\zeta(dy)\int_{\mathbb{R}^d}\zeta(dz)\left(\frac{(z_i-y_i)(z_j-y_j)}{4\delta^2}-\frac{1_{i=j}}{2\delta}\right)G_{2\delta}(z-y)\sum_{k=1}^{n}\sigma_1^{ik}(z)\sigma_1^{jk}(z). \tag{4.3.26}$$

By symmetry of y and z in Eq. (4.3.26), Eq. (4.3.25) is further given by

$$\sum_{i,j=1}^{d}\int_{\mathbb{R}^d}\zeta(dy)\int_{\mathbb{R}^d}\zeta(dz)\left(\frac{(z_i-y_i)(z_j-y_j)}{4\delta^2}-\frac{1_{i=j}}{2\delta}\right)$$
$$\times G_{2\delta}(z-y)\frac{1}{2}\sum_{k=1}^{n}\left(\sigma_1^{ik}(z)\sigma_1^{jk}(z)+\sigma_1^{ik}(y)\sigma_1^{jk}(y)\right). \tag{4.3.27}$$

Similarly, the second term on the left hand side of Eq. (4.3.24) can be expressed as

$$-\sum_{i=1}^{n}\sum_{i,j=1}^{d}\langle T_\delta(\sigma_1^{ik}\zeta),\partial_i\partial_jT_\delta(\sigma_1^{jk}\zeta)\rangle_0$$
$$=-\sum_{i,j=1}^{d}\int_{\mathbb{R}^d}\zeta(dy)\int_{\mathbb{R}^d}\zeta(dz)\left(\frac{(z_i-y_i)(z_j-y_j)}{4\delta^2}-\frac{1_{i=j}}{2\delta}\right)$$
$$\times G_{2\delta}(z-y)\frac{1}{2}\sum_{k=1}^{n}\left(\sigma_1^{ik}(y)\sigma_1^{jk}(z)+\sigma_1^{ik}(z)\sigma_1^{jk}(y)\right). \tag{4.3.28}$$

Now adding Eqs. (4.3.27) and (4.3.28), the left-hand side of (4.3.24) is given by

$$\sum_{i,j=1}^{d}\int_{\mathbb{R}^d}\zeta(dy)\int_{\mathbb{R}^d}\zeta(dz)\left(\frac{(z_i-y_i)(z_j-y_j)}{4\delta^2}-\frac{1_{i=j}}{2\delta}\right)$$
$$\times G_{2\delta}(z-y)\frac{1}{2}\sum_{k=1}^{n}\left(\sigma_1^{ik}(y)-\sigma_1^{ik}(z)\right)\left(\sigma_1^{jk}(y)-\sigma_1^{jk}(z)\right).$$

Using the identity that $G_\delta(x) = \exp\left(-\frac{|x|^2}{4\delta}\right) 2^{d/2} G_{2\delta}(x)$, the Lipschitz continuity of σ_1 and Lemma 4.6, we have the quantity above estimated by

$$\begin{aligned}
&\sum_{i,j=1}^{d} \int_{\mathbb{R}^d} \zeta(dy) \int_{\mathbb{R}^d} \zeta(dz) \left(\frac{|z-y|^2}{4\delta^2} + \frac{1}{2\delta}\right) \exp\left(-\frac{|z-y|^2}{4\delta}\right) 2^{d/2} G_{4\delta}(z-y) \frac{1}{2} K^2 |z-y|^2 \\
&\leq 4K^2 \sum_{i,j=1}^{d} \int_{\mathbb{R}^d} |\zeta|(dy) \int_{\mathbb{R}^d} |\zeta|(dz) 2^{d/2} G_{4\delta}(z-y) \\
&= d^2 2^{2+d/2} K^2 \|T_{2\delta}(|\zeta|)\|_0^2 \\
&\leq d^2 2^{2+d/2} K^2 \|T_\delta(|\zeta|)\|_0^2.
\end{aligned}$$

The lemma then follows with $K_1 = d^2 2^{2+d/2} K^2$. ■

Now applying Lemmas 4.9, 4.10 and 4.11 yields the following theorem.

Theorem 4.12. *If μ is a measure-valued solution of the Zakai Eq. (4.3.3) and $Z^\delta = T_\delta \mu$, then*

$$\tilde{\mathbb{E}} \|Z_t^\delta\|_0^2 \leq \|Z_0^\delta\|_0^2 + K_1 \int_0^t \tilde{\mathbb{E}} \|T_\delta(|\mu_s|)\|_0^2 ds, \tag{4.3.29}$$

where K_1 is a suitable constant.

Proof. Consider Eq. (4.3.9) such that $\tilde{\mathbb{E}}\|Z_t^\delta\|_0^2 = A_1 - 2A_2 + A_3 + 2A_4 + A_5 + A_6 + A_7$ and $A_1, \ldots, A_7$ are defined in Eq. (4.3.10). Then by inequality (4.3.23), A_2 is bounded by a constant times $\|T_\delta(|\mu_s|)\|_0^2$. The bound for A_3 follows from inequality (4.3.24). A_4 follows from Lemma (4.9) and inequality (4.3.22), and A_5 is bounded by Lemma (4.9), inequalities (4.3.22) and (4.3.23). The bounds for A_6 and A_7 follow from inequality (4.3.22). ■

Theorem 4.13. *The solution of Zakai Eq. (4.3.3) is unique.*

Proof. Let μ_t^1 and μ_t^2 be two measure-valued solution with the same initial value μ_0. Then by Theorem 4.12

$$\tilde{\mathbb{E}} \|T_\delta D_t\|_0^2 \leq K_1 \int_0^t \tilde{\mathbb{E}} \|T_\delta(|D_s|)\|_0^2 ds.$$

Note that by Lemma 4.6,

$$\int_0^t \tilde{\mathbb{E}}\|T_\delta(|D_s|)\|_0^2 ds \le \int_0^t \tilde{\mathbb{E}}\|D_s\|_0^2 ds < \infty. \tag{4.3.30}$$

Then letting $\delta \to 0$, by dominated convergence based on Eq. (4.3.30) we have

$$\tilde{\mathbb{E}}\|D_t\|_0^2 \le K_1 \int_0^t \tilde{\mathbb{E}}\|D_s\|_0^2 ds,$$

and Gronwall's inequality yields $D_t = 0$. ■

Consequently, the uniqueness of Kushner-Stratonovich equation then follows from Theorem 4.13.

Theorem 4.14. *The solution of Kushner-Stratonovich Eq. (4.3.5) is unique.*

Proof. Let π_t^1 and π_t^2 be two solutions to the Kushner-Stratonovich Eq. (4.3.5). Note that for $i = 1, 2$ and $F \in C_b(\mathbb{R})$ we have $\pi_t^i(F)\mu_t^i(1) = \mu_t^i(F)$, where μ_t^i are the corresponding solutions to Zakai equation. From Theorem 4.13, we have $\mu_t^1 = \mu_t^2$, *a.s.* for all $t \ge 0$. Hence for all $t \ge 0$,

$$\pi_t^1(F) = \frac{\mu_t^1(F)}{\mu_t^1(1)} = \frac{\mu_t^2(F)}{\mu_t^2(1)} = \pi_t^2(F), \ a.s.$$

■

4.4 Large deviation principle

4.4.1 LDP result for the optimal filter

We study the limiting behavior of the optimal filter with a small signal-to-noise ratio, i.e., consider the signal given in Eq. (4.3.1a) and the observation process below, for $\epsilon \downarrow 0$,

$$\begin{aligned} Y_t^\epsilon =& \sqrt{\epsilon}\int_0^t b_2(s, X_s) ds + \int_0^t \sigma_2(s) dW_s + \int_0^t \int_{\mathbb{U}_2} f_2(s,u)\tilde{N}_{\lambda^\epsilon}(ds, du) \\ &+ \int_0^t \int_{\mathbb{U}\setminus\mathbb{U}_2} g_2(s,u) N_{\lambda^\epsilon}(ds, du), \end{aligned} \tag{4.4.1}$$

where $N_{\lambda^\epsilon}(dt,du)$ is an integer-valued random measure with a predictable compensator given by $\lambda^\epsilon(t,x,u)\nu_2(du)dt$ and $\lambda^\epsilon(t,x,u) = \epsilon\lambda(t,x,u)+1-\epsilon$. For any test function $F \in C_b^2(\mathbb{R}^d)$, set $\pi_t^\epsilon(F) = \mathbb{E}(F(X_t)|\mathcal{F}_t^{Y^\epsilon})$ and define similarly to Eq. (4.3.2) an equivalent probability measure which makes the signal and observation processes independent, i.e.,

$$\Lambda_t^\epsilon = \exp\left\{\sqrt{\epsilon}\int_0^t (\sigma_2^{-1}(s)b_2(s,X_s))^* dW_s + \frac{\epsilon}{2}\int_0^t |\sigma_2^{-1}(s)b_2(s,X_s)|^2 ds \right.$$
$$\left. + \int_0^t\int_{\mathbb{U}_2}(1-\lambda(s,X_s,u))\nu_2(du)ds + \int_0^t\int_{\mathbb{U}_2}\log(\epsilon\lambda(s,X_{s-},u)+1-\epsilon)N_{\lambda^\epsilon}^{\epsilon^{-1}}(ds,du)\right\},$$

where $N_{\lambda^\epsilon}^{\epsilon^{-1}}(dt,du)$ is an integer-valued random measure with its predictable compensator $\epsilon^{-1}\lambda^\epsilon(t,X_t,u)\nu_2(du)dt$. Consider $\tilde{\mathbb{P}}^\epsilon$ such that $d\tilde{\mathbb{P}}^\epsilon/d\mathbb{P} = (\Lambda_T^\epsilon)^{-1}$ and $\mu_t^\epsilon(F) = \tilde{\mathbb{E}}^\epsilon(\Lambda_t^\epsilon F(X_t)|\mathcal{F}_t^{Y^\epsilon})$, where $\tilde{\mathbb{E}}^\epsilon$ denotes the expectation under the measure $\tilde{\mathbb{P}}^\epsilon$. Now we establish the existence of the small-noise optimal filter, given in Proposition 4.15 below.

Proposition 4.15. *Let the signal be defined as in Eq. (4.3.1a) and the observation process, Y_t^ϵ, be as in Eq. (4.4.1). Then we have the following small-noise Zakai equation, for any $F \in \mathcal{D}(\mathcal{L})$*

$$\mu_t^\epsilon(F) = \mu_0(F) + \int_0^t \mu_s^\epsilon(\mathcal{L}F)ds + \sqrt{\epsilon}\sum_{i=1}^m\int_0^t \mu_s^\epsilon(F(\sigma_2^{-1}(s)b_2(s,\cdot))^i)d\tilde{W}_s^{\epsilon,i}$$
$$+ \int_0^t\int_{\mathbb{U}_2}\mu_{s-}^\epsilon(F(\lambda(s,\cdot,\mu)-1))(\epsilon N_{\lambda^\epsilon}^{\epsilon^{-1}}(ds,du)-\nu_2(du)ds). \qquad (4.4.2)$$

The corresponding small-noise Kushner-Stratonovich equation is given by

$$\pi_t^\epsilon(F) = \pi_0(F) + \int_0^t \pi_s^\epsilon(\mathcal{L}F)ds$$
$$+ \sqrt{\epsilon}\sum_{i=1}^m\int_0^t(\pi_s^\epsilon(F(\sigma_2^{-1}(s)b_2(s,\cdot))^i) - \pi_s^\epsilon(F)\pi_s^\epsilon(\sigma_2^{-1}(s)b_2(s,\cdot))^i)d\hat{W}_s^{\epsilon,i}$$
$$+ \int_0^t\int_{\mathbb{U}_2}\frac{\pi_{s-}^\epsilon(F\lambda(s,\cdot,u)) - \pi_{s-}^\epsilon(F)\pi_{s-}^\epsilon(\lambda(s,\cdot,u))}{\pi_{s-}^\epsilon(\epsilon\lambda(s,\cdot,u)+1-\epsilon)}\epsilon\hat{N}^{\epsilon^{-1}}(ds,du), \qquad (4.4.3)$$

where $\hat{W}_t^\epsilon = \tilde{W}_t^\epsilon - \sqrt{\epsilon}\int_0^t \pi_s^\epsilon(\sigma_2^{-1}(s)b_2(s,\cdot))ds$ is the innovation process and

$$\epsilon\hat{N}^{\epsilon^{-1}}(dt,du) = \epsilon N_{\lambda^\epsilon}^{\epsilon^{-1}}(dt,du) - \pi_{t-}^\epsilon(\epsilon\lambda(s,\cdot,u)+1-\epsilon))\nu_2(du)dt.$$

Proof. By Itô's formula,

$$
\begin{aligned}
d\Lambda_t^\epsilon =&\Lambda_t^\epsilon\Big\{\sqrt{\epsilon}(\sigma_2^{-1}(t)b_2(t,X_t))^*dW_t \\
&+\frac{\epsilon}{2}|\sigma_2^{-1}(t)b_2(t,X_t)|^2dt+\int_{\mathbb{U}_2}(1-\lambda(t,X_t,u))\nu_2(du)dt\Big\} \\
&+\frac{\epsilon}{2}\Lambda_t^\epsilon|\sigma_2^{-1}(t)b_2(t,X_t)|^2dt+\int_{\mathbb{U}_2}\Lambda_t^\epsilon(\lambda(t,X_{t-},u)-1)\epsilon N_{\lambda^\epsilon}^{\epsilon^{-1}}(dt,du).
\end{aligned}
$$

Then

$$
\begin{aligned}
\Lambda_t^\epsilon = 1+\sqrt{\epsilon}\int_0^t \Lambda_s^\epsilon(\sigma_2^{-1}(s)b_2(s,X_s))^*d\tilde{W}_s^\epsilon& \\
+\int_0^t\int_{\mathbb{U}_2}\Lambda_{s-}(\lambda(s,X_{s-},u)-1)(\epsilon N_{\lambda^\epsilon}^{\epsilon^{-1}}(ds,du)-\nu_2(du)ds).&
\end{aligned}
$$

By Girsanov's Theorem (see [66] and [37, Theorem 3.17], also [9, Theorem 3.12], [69] or [10, Thm. VIII, T10]), $\tilde{W}_t^\epsilon = W_t+\sqrt{\epsilon}\int_0^t \sigma_2^{-1}(s)b_2(s,X_s)ds$, is a $\tilde{\mathbb{P}}^\epsilon$-Brownian motion, and $\tilde{N}^{\epsilon^{-1}} = N_{\lambda^\epsilon}^{\epsilon^{-1}}(dt,du)-\epsilon^{-1}\nu_2(du)dt$, is a $\tilde{\mathbb{P}}^\epsilon$-Poisson martingale measure. The Zakai Eq. (4.4.2) is obtained by the same argument in [66, Theorem 3.1]. We now show the derivation of the Kushner-Stratonovich equation (4.4.3). Note that

$$
\begin{aligned}
\mu_t^\epsilon(1)=\mu_0(1)+\sqrt{\epsilon}\sum_{i=1}^m\int_0^t\mu_s^\epsilon((\sigma_2^{-1}(s)b_2(s,\cdot))^i)d\tilde{W}_s^{\epsilon,i}& \\
+\int_0^t\int_{\mathbb{U}_2}\mu_{s-}^\epsilon(\lambda(s,\cdot,\mu)-1)(\epsilon N_{\lambda^\epsilon}^{\epsilon^{-1}}(ds,du)-\nu_2(du)ds).&
\end{aligned}
$$

By Itô's formula, we have

$$
\begin{aligned}
d\frac{1}{\mu_t^\epsilon(1)} =& -\frac{\sqrt{\epsilon}}{\mu_t^\epsilon(1)^2}\sum_{i=1}^m\mu_t^\epsilon((\sigma_2^{-1}(t)b_2(t,\cdot))^i)d\tilde{W}_t^{\epsilon,i}+\frac{\epsilon}{\mu_t^\epsilon(1)^3}\mu_t^\epsilon(\sigma_2^{-1}(t)b_2(t,\cdot))dt \\
&+\int_{\mathbb{U}_2}\left(\frac{1}{\mu_t^\epsilon(1)+\epsilon\mu_t^\epsilon(\lambda(t,\cdot,u)-1)}-\frac{1}{\mu_t^\epsilon(1)}\right)\left(N_{\lambda^\epsilon}^{\epsilon^{-1}}(dt,du)-\epsilon^{-1}\nu_2(du)dt\right) \\
&+\int_{\mathbb{U}_2}\left(\frac{1}{\mu_t^\epsilon(1)+\epsilon\mu_t^\epsilon(\lambda(t,\cdot,u)-1)}-\frac{1}{\mu_t^\epsilon(1)}+\epsilon\mu_t^\epsilon(\lambda(t,\cdot,u)-1)\frac{1}{\mu_t^\epsilon(1)^2}\right)\epsilon^{-1}\nu_2(du)dt.
\end{aligned}
$$

Hence, the quadratic variation is given by

$$d[\mu^\epsilon(F),\mu^\epsilon(1)^{-1}]_t = -\frac{\epsilon}{\mu_t^\epsilon(1)^2}\sum_{i=1}^m \mu_t^\epsilon((\sigma_2^{-1}(t)b_2(t,\cdot))^i)\mu_t^\epsilon(F(\sigma_2^{-1}(t)b_2(t,\cdot))^i)dt$$
$$+\int_{\mathbb{U}_2}\left(\frac{1}{\mu_t^\epsilon(1)+\epsilon\mu_t^\epsilon(\lambda(t,\cdot,u)-1)}-\frac{1}{\mu_t^\epsilon(1)}\right)\cdot\epsilon\cdot\mu_{s-}^\epsilon(F(\lambda(s,\cdot,\mu)-1))N_{\lambda^\epsilon}^{\epsilon^{-1}}(ds,du).$$

Eventually, one can easily get Eq. (4.4.3) by the following product of Itô's formula

$$d\frac{\mu_t^\epsilon(F)}{\mu_t^\epsilon(1)} = \frac{1}{\mu_t^\epsilon(1)}d\mu_t^\epsilon(F)+\mu_t^\epsilon(F)d\frac{1}{\mu_t^\epsilon(1)}+d[\mu^\epsilon(F),\mu^\epsilon(1)^{-1}]_t.$$

This completes the proof. ■

What it follows verifies Condition 4.2.1 such that we show the LDP for the signal described in Eq. (4.3.1a) and observation process as in Eq. (4.4.1). To proceed with the demonstration of the LDP, the assumption on the boundedness of the infinitesimal generator $\mathcal{L}$ is necessary. However, this condition does not contradict with the well-posedness of the optimal filtering framework, as seen in [66] and the proof of Proposition 4.15.

Assumption 4.4.1. *The test function F has continuous and bounded derivatives up to order 2.*

The images of $\mathcal{G}^\epsilon, \mathcal{G}^0$ considered in Condition 4.2.1 are solutions of versions of Zakai equation with or without noise respectively. Recall that $\mathcal{M}_F(\mathbb{R}^d)$ denotes the collection of finite Borel measure on $\mathbb{R}^d$, and μ_t^ϵ is an $\mathcal{M}_F(\mathbb{R}^d)$-valued process. For each $\epsilon > 0$, let $\mathcal{G}^\epsilon : \mathcal{M}_F(\mathbb{R}^d)\times C([0,T],\mathbb{R}^m)\times\mathbb{M}(\mathbb{U}_T)\to D([0,T],\mathcal{M}_F(\mathbb{R}^d))$ be a measurable map, such that

$$\mu^{\epsilon,\mu_0} := \mathcal{G}^\epsilon(\mu_0,\sqrt{\epsilon}\tilde{W}^\epsilon,\epsilon N_{\lambda^\epsilon}^{\epsilon^{-1}}). \tag{4.4.4}$$

Adopting the arguments of Theorem 4.13, the following holds.

Theorem 4.16. *Under Assumptions 4.3.1, 4.3.2 and 4.3.3, the unnormalized filtered μ^{ϵ,μ_0} defined in Eq. (4.4.4) is the unique solution of the Zakai Eq. (4.4.2) with initial condition μ_0.*

Let $\xi = (\psi, \phi) \in \mathcal{U}^M$. The controlled version of Eq. (4.4.2) for all $F \in \mathcal{D}(\mathcal{L})$, is given by

$$\begin{aligned}\mu_t^{\epsilon,\xi}(F) &= \mu_0(F) + \int_0^t \mu_s^{\epsilon,\xi}(\mathcal{L}F)ds + \sqrt{\epsilon}\sum_{i=1}^m \int_0^t \mu_s^{\epsilon,\xi}(F(\sigma_2^{-1}(s)b_2(s,\cdot))^i)d\tilde{W}_s^{\epsilon,i} \\ &+ \int_0^t \mu_s^{\epsilon,\xi}(F(\sigma_2^{-1}(s)b_2(s,\cdot)))\psi(s)ds \\ &+ \int_0^t \int_{\mathbb{U}_2} \mu_{s-}^{\epsilon,\xi}(F(\lambda(s,\cdot,\mu)-1))(\epsilon N_{\lambda^\epsilon}^{\epsilon^{-1}\phi}(ds,du) - \nu_2(du)ds). \qquad (4.4.5)\end{aligned}$$

Let $\mu_t^{0,\xi}$ be the solution of the noise-free controlled version of Eq. (4.4.5), i.e.

$$\begin{aligned}\mu_t^{0,\xi}(F) &= \mu_0(F) + \int_0^t \mu_s^{0,\xi}(\mathcal{L}F)ds + \int_0^t \mu_s^{0,\xi}(F(\sigma_2^{-1}(s)b_2(s,\cdot)))\psi(s)ds \\ &+ \int_0^t \int_{\mathbb{U}_2} \mu_s^{0,\xi}(F(\lambda(s,\cdot,\mu)-1))(\phi(s,u)-1)\nu_2(du)ds. \qquad (4.4.6)\end{aligned}$$

For $g : \mathcal{B}(\mathbb{U}_T) \to [0,\infty)$, define $\nu_2^g(A) = \int_A g(s,u)\nu_2(du)ds$ for any $A \in \mathcal{B}(\mathbb{U}_T)$ where ν_2 is the intensity measure of N_{λ^ϵ} in the observation process. Let $\mathcal{G}^0 : \mathcal{M}_F(\mathbb{R}^d) \times C([0,T],\mathbb{R}^m) \times \mathbb{M}(\mathbb{U}_T) \to C([0,T],\mathcal{M}_F(\mathbb{R}^d))$ be a measurable map such that $\mathcal{G}^0(\mu_0, w, m) = \mu_t^{0,\xi}$ if $(w,m) = (\int_0^\cdot \psi(s)ds, \nu_2^\phi) \in C([0,T],\mathbb{R}^m) \times \mathbb{M}(\mathbb{U}_T)$ for $\xi = (\psi,\phi)$, otherwise $\mathcal{G}^0 = 0$. For $\mu \in C([0,T],\mathcal{M}_F(\mathbb{R}^d))$ and $\mu_0 \in \mathcal{M}_F(\mathbb{R}^d)$ define

$$I_1 \equiv I_{1,\mu_0}(\mu) := \inf_{\{\xi=(\psi,\phi)\in\mathbb{S}_\mu : \mu = \mathcal{G}^0(\mu_0, \int_0^\cdot \psi(s)ds, \nu_2^\phi)\}} \bar{L}_T(\xi), \qquad (4.4.7)$$

where $\bar{L}_T$ is defined in Eq. (4.2.4) by replacing ν with ν_2. The following theorem is the main one that establishes the uniform large deviations for the unnormalized filter. Its proof is delegated to the next section.

Theorem 4.17. *Let μ^{ϵ,μ_0} be as in Eq. (4.4.4). Then I_{1,μ_0}, defined in (4.4.7), is a rate function on $D([0,T],\mathcal{M}_F(\mathbb{R}^d))$ and the family of rate functions $\{I_{1,\mu_0}, \mu_0 \in \mathcal{M}_F(\mathbb{R}^d)\}$ has compact level sets. Furthermore, $\{\mu^{\epsilon,\mu_0}\}$ satisfies the large deviation principle on $D([0,T],\mathcal{M}_F(\mathbb{R}^d))$ with the rate function I_{1,μ_0} uniformly for μ_0 in compact sets of $\mathcal{M}_F(\mathbb{R}^d)$.*

Recall that $\mathcal{P}(\mathbb{R}^d)$ denotes the collection of the optimal filter π_t. We define $\tilde{\mathcal{G}}^0 : \mathcal{P}(\mathbb{R}^d) \times L^2([0,T],\mathbb{R}^m) \times [0,\infty) \to C([0,T],\mathcal{P}(\mathbb{R}^d))$ a measurable function and suppose that $\pi^{0,\xi} =$

$\tilde{\mathcal{G}}^0(\pi_0, \psi, \phi)$, where $\pi^{0,\xi}$ is the solution of the following controlled equation.

$$\pi_t^{0,\xi}(F) = \pi_0(F) + \int_0^t \pi_s^{0,\xi} \left\{ \mathcal{L}F + \left(\sigma_2^{-1}(s)b_2(s,\cdot) - \pi_s^{0,\xi}(\sigma_2^{-1}(s)b_2(s,\cdot))\right)\psi(s)F \right\} ds \quad (4.4.8)$$
$$+ \int_0^t \int_{\mathbb{U}_2} \left[\pi_s^{0,\xi}(F\lambda(s,\cdot,u)) - \pi_s^{0,\xi}(F)\pi_s^{0,\xi}(\lambda(s,\cdot,u))\right](\phi(s,u)-1)\nu_2(du)dt.$$

For $\pi \in C([0,T], \mathcal{P}(\mathbb{R}^d))$ and $\pi_0 \in \mathcal{P}(\mathbb{R}^d)$ define

$$I_2 \equiv I_{2,\pi_0}(\pi) := \inf_{\{\xi=(\psi,\phi)\in\mathbb{S}_\pi : \pi = \mathcal{G}^0(\pi_0, \int_0^\cdot \psi(s)ds, \nu_2^\phi)\}} \bar{L}_T(\xi). \quad (4.4.9)$$

Recall that a topological space, $\mathcal{X}$ is Hausdorff if, for every pair of distinct points x and y in $\mathcal{X}$, there exist disjoint neighborhoods of x and y. We present the theorem of contraction principle of LDP, whose proof is referred to [21].

Theorem 4.18 (**Contraction Principle**). *Let $\mathcal{X}$ and $\mathcal{Y}$ be Hausdorff topological spaces and $f : \mathcal{X} \to \mathcal{Y}$ a continuous function. Consider a good rate function $I : \mathcal{X} \to [0, \infty]$. (a) For each $y \in \mathcal{Y}$, define*

$$I'(y) := \inf\{I(x) : x \in \mathcal{X}, y = f(x)\}.$$

Then I is a good rate function on $\mathcal{Y}$, where as usual the infimum over the empty set is taken as ∞. (b) If I controls the LDP associated with a family of probability measures $\{\mu^\epsilon\}$ on $\mathcal{X}$, then I controls the LDP associated with the family of probability measures $\{\mu^\epsilon \circ f^{-1}\}$ on $\mathcal{Y}$.

Lemma 4.19. *For $\xi = (\psi, \phi) \in \mathcal{U}^M$, let*

$$M_t^\xi = \exp\left(\int_0^t \pi_s(\sigma_2^{-1}(s)b_2(s,\cdot)\psi(s))ds + \int_0^t \int_{\mathbb{U}_2} \pi_s(\lambda(s,\cdot,\mu)-1)(\phi(s,u)-1)\nu_2(du)ds\right).$$

Then $\mu_t^{0,\xi}(F) = \pi_t^{0,\xi}(F)M_t^\xi$ satisfies the noise-free Zakai Eq. (4.4.6).

Proof. We first note that

$$dM_t^\xi = M_t^\xi \left(\pi_t(\sigma_2^{-1}(t)b_2(t,\cdot)\psi(t))dt + \int_{\mathbb{U}_2} \pi_t(\lambda(t,\cdot,\mu)-1)(\phi(t,u)-1)\nu_2(du)dt\right).$$

Differentiating the product, we have

$$
\begin{aligned}
&dtd\pi_t^{0,\xi}(F)M_t^{\xi} = \pi_t^{0,\xi}(F)M_t^{\xi}\pi_t(\sigma_2^{-1}(t)b_2(t,\cdot)\psi(t))dt \\
&\quad + \pi_t^{0,\xi}(F)M_t^{\xi}\int_{\mathbb{U}_2}\pi_t(\lambda(t,\cdot,\mu)-1)(\phi(t,u)-1)\nu_2(du)dt \\
&\quad + \pi_t^{0,\xi}(\mathcal{L}F)M_t^{\xi}dt + \pi_t(F\sigma_2^{-1}(t)b_2(t,\cdot)\psi(t))M_t^{\xi}dt - \pi_t^{0,\xi}(F)M_t^{\xi}\pi_t(\sigma_2^{-1}(t)b_2(t,\cdot)\psi(t))dt \\
&\quad + \int_{\mathbb{U}_2} M_t^{\xi}\left[\pi_s^{0,\xi}(F\lambda(s,\cdot,u)) - \pi_s^{0,\xi}(F)\pi_s^{0,\xi}(\lambda(s,\cdot,u))\right](\phi(s,u)-1)\nu_2(du)dt. \qquad (4.4.10)
\end{aligned}
$$

Regroup the right-hand side of Eq. (4.4.10) and then the integral form coincides with Eq. (4.4.6). ■

The following theorem establishes a uniform large deviation principle for the optimal filtering defined in Eq. (4.4.3).

Theorem 4.20. *Suppose π_t^{ϵ,π_0} is the optimal filter described by the Kushner-Stratonovich Eq. (4.4.3). Then $\{\pi^{\epsilon,\pi_0}\}$ satisfies the large deviation principle on $D([0,T],\mathcal{P}(\mathbb{R}^d))$ with the rate function I_{2,π_0}, defined in Eq. (4.4.9), uniformly for π_0 in compact sets of $\mathcal{P}(\mathbb{R}^d)$.*

Proof. Define a continuous map $G : D([0,T],\mathcal{M}_F(\mathbb{R}^d)\setminus\{0\}) \to D([0,T],\mathcal{P}(\mathbb{R}^d))$ such that

$$
(G\mu)_t = \frac{\mu_t(F)}{\mu_t(1)}.
$$

Then, by the Theorem 4.18, $\{\pi^{\epsilon,\pi_0} = G(\mu^{\epsilon,\pi_0})\}$ satisfies the large deviation principle with rate function $I_2'(\pi) = \inf\{I_1(\mu) : G(\mu) = \pi\}$. Suppose $I_2'(\pi) < \infty$, then for all $\delta > 0$ there exists μ such that $G(\mu) = \pi$ and $I_1(\mu) < I_2'(\pi) + \delta$. Choose a control $\xi \in \mathbb{S}_\pi$ such that $\gamma(\xi) = \mu$, where γ is the solution of Eq. (4.4.6) and $\bar{L}_T(\xi) < I_1(\mu) + \delta$. Taking $\tilde{\gamma} \doteq G\circ\gamma$ we have $\tilde{\gamma}(\xi) = \xi$, where $\tilde{\gamma}$ is the solution of Eq. (4.4.8), and $\bar{L}_T(\xi) < I_2'(\pi) + 2\delta$. Thus, by definition of I_2 we have

$$
I_2(\pi) \le I_2'(\pi). \qquad (4.4.11)
$$

Now if $I_2(\pi) < \infty$. Then for all $\delta > 0$, there exists $\xi \in \mathbb{S}_\pi$ such that $\tilde{\gamma}(\xi) = \pi$ and $\bar{L}_T(\xi) < I_2(\pi) + \delta$. By Lemma 4.19, we have $\mu_t = M_t^{\xi}\pi_t$. Then $\mu = \gamma(\xi)$ is the solution of Eq. (4.4.6) and $G(\mu) = \pi$. Hence

$$I_2'(\pi) \leq I_1(\mu) \leq \bar{L}_T(\xi) < I_2(\pi) + \delta. \tag{4.4.12}$$

Eqs. (4.4.11) and (4.4.12) give $I_2 = I_2'$ and then the result follows. ■

4.4.2 Proof of Theorem 4.17

In the subsection we will demonstrate the Condition 4.2.1 is verified in order to show the uniform large deviation estimates for the unnormalized filter. The following theorem establishes the existence and uniqueness of the controlled version of Zakai equation given in Eq. (4.4.5).

Theorem 4.21. *Suppose $\mathcal{G}^\epsilon$ is given by $\mu^{\epsilon,\mu_0} = \mathcal{G}^\epsilon(\mu_0, \sqrt{\epsilon}\tilde{W}^\epsilon, \epsilon N_{\lambda^\epsilon}^{\epsilon^{-1}})$, and $\xi = (\psi, \phi) \in \mathcal{U}^M$ for some $M > 0$. For $\epsilon > 0$ and $\mu_0 \in \mathcal{M}_F(\mathbb{R}^d)$, define*

$$\mu^{\epsilon,\xi,\mu_0} = \mathcal{G}^\epsilon(\mu_0, \sqrt{\epsilon}\tilde{W}^\epsilon + \int_0^\cdot \psi(s)ds, \epsilon N_{\lambda^\epsilon}^{\epsilon^{-1}\phi}).$$

Then μ^{ϵ,ξ,μ_0} is the unique solution of Eq. (4.4.5).

Proof. Take a control $\xi = (\psi, \phi) \in \mathcal{U}^M$, and consider

$$\begin{aligned}
M_t^{\epsilon,\xi} := \exp\Bigg\{ &-\frac{1}{\sqrt{\epsilon}} \int_0^t \psi(s) d\tilde{W}_s^\epsilon - \frac{1}{2\epsilon} \int_0^t |\psi(s)|^2 ds \\
&+ \int_0^t \int_{\mathbb{U}_2 \times [0,\epsilon^{-1}]} \log \phi(s,u) \bar{N}_{\lambda^\epsilon}(ds,du,dr) + \int_0^t \int_{\mathbb{U}_2 \times [0,\epsilon^{-1}]} (1 - \phi(s,u)) \bar{\nu}_2(ds,du,dr) \Bigg\},
\end{aligned}$$

where $\bar{N}_{\lambda^\epsilon}$ and the corresponding intensity $\bar{\nu}_2$ satisfy the definitions of $\bar{N}$ and $\bar{\nu}_T$ in Section 4.2.1 respectively. By Itô's formula,

$$\begin{aligned}
dM_t^{\epsilon,\xi} =& M_t^{\epsilon,\xi} \left(-\frac{1}{\sqrt{\epsilon}} \psi(t) d\tilde{W}_t^\epsilon - \frac{1}{2\epsilon} |\psi(t)|^2 dt + \int_{\mathbb{U}_2 \times [0,\epsilon^{-1}]} (1 - \phi(s,u)) \bar{\nu}_2(ds,du,dr) \right) \\
&+ \frac{1}{2\epsilon} M_t^{\epsilon,\xi} |\psi(t)|^2 dt + \int_{\mathbb{U}_2 \times [0,\epsilon^{-1}]} M_{t-}^{\epsilon,\xi} (\phi(t,u) - 1) \bar{N}_{\lambda^\epsilon}(ds,du,dr).
\end{aligned}$$

Thus,

$$M_t^{\epsilon,\xi} = 1 - \frac{1}{\sqrt{\epsilon}} \int_0^t M_s^{\epsilon,\xi} \psi(s) d\tilde{W}_s^\epsilon + \int_0^t \int_{\mathbb{U}_2} M_{s-}^{\epsilon,\xi} (\phi(s,u) - 1) \left(\bar{N}_{\lambda^\epsilon}(ds, du, dr) - \bar{\nu}_2(ds, du, dr) \right).$$

Define a new probability measure $\mathbb{P}^{\epsilon,\xi}$ by $\frac{d\mathbb{P}^{\epsilon,\xi}}{d\tilde{\mathbb{P}}^\epsilon} = M_T^{\epsilon,\xi}$, then $\mathbb{P}^{\epsilon,\xi}$ is an equivalent probability measure with respect to $\tilde{\mathbb{P}}^\epsilon$. By Girsanov's Theorem, $W_t^{\epsilon,\xi} := \tilde{W}_t^\epsilon + \frac{1}{\sqrt{\epsilon}} \int_0^t \psi(s) ds$ is a $\mathbb{P}^{\epsilon,\xi}-$Brownian motion and $N^{\epsilon,\xi}(dt, du) := N_{\lambda^\epsilon}(dt, du) - \phi(t,u)\nu_2(du)dt$ is a $\mathbb{P}^{\epsilon,\xi}-$Poisson martingale measure. Hence, the desired result follows from replacing $(\tilde{W}, N_{\lambda^\epsilon})$ in Eq. (4.4.2) with $(W^{\epsilon,\xi}, N^{\epsilon,\xi})$ and Theorem 4.16. ■

We now verify the existence and uniqueness of the solution of the zero-noise controlled Zakai Eq. (4.4.6).

Theorem 4.22. *For any fixed $\mu_0 \in \mathcal{M}_F(\mathbb{R}^d)$ and a pair of well-defined control $\xi = (\psi, \phi) \in \mathcal{U}^M$ for some $M > 0$, there is a unique solution $\mu_t^{0,\xi} \in C([0,T], \mathcal{M}_F(\mathbb{R}^d))$ of Eq. (4.4.6).*

Proof. We start with showing the existence of the solution. Let

$$\Lambda_t^{0,\xi} = \exp\left(\int_0^t \sigma_2^{-1}(s) b_2(s, X_s) \psi(s) ds + \int_0^t \int_{\mathbb{U}_2} (\lambda(s, X_s, \mu) - 1)(\phi(s,u) - 1)\nu_2(du) ds \right).$$

Next, we define $\mu_t^{0,\xi} = \mathbb{E}(\Lambda_t^{0,\xi} F(X_t))$ and we show that it is a solution of Eq. (4.4.6). First of all,

$$d\Lambda_t^{0,\xi} = \Lambda_t^{0,\xi} \left(\sigma_2^{-1}(t) b_2(t, X_t) \psi(t) dt + \int_{\mathbb{U}_2} (\lambda(t, X_t, \mu) - 1)(\phi(t,u) - 1)\nu_2(du) dt \right).$$

Note that for any $F \in \mathcal{D}(\mathcal{L})$ we have $M_t^F \equiv F(X_t) - \int_0^t \mathcal{L}F(X_s) ds$ is a martingale. Then by Itô's formula, we have

$$\begin{aligned} d\Lambda_t^{0,\xi} F(X_t) = \Lambda_t^{0,\xi} \left(dM_t^F + \mathcal{L}F(X_t) dt \right) + F(X_t) \Lambda_t^{0,\xi} \sigma_2^{-1}(t) b_2(t, X_t) \psi(t) dt \\ + \int_{\mathbb{U}_2} F(X_t) \Lambda_t^{0,\xi} (\lambda(t, X_t, \mu) - 1)(\phi(t,u) - 1)\nu_2(du) dt. \end{aligned}$$

Hence,

$$\Lambda_t^{0,\xi}F(X_t) = F(X_0) + \int_0^t \Lambda_s^{0,\xi}(\mathcal{L}F(X_s) + \sigma_2^{-1}(s)b_2(s,X_s)\psi(s))ds. \tag{4.4.13}$$
$$+ \int_0^t \int_{\mathbb{U}_2} \Lambda_s^{0,\xi}F(X_s)(\lambda(s,X_s,\mu)-1)(\phi(s,u)-1)\nu_2(du)ds + \int_0^t \Lambda_s^{0,\xi}dM_s^F.$$

Taking expectation on both sides of Eq. (4.4.13) yields Eq. (4.4.6). Thus $\mu_t^{0,\xi}(F) = \mathbb{E}(\Lambda_t^{0,\xi}F(X_t))$ is a solution of Eq. (4.4.6).

Next, we verify the uniqueness by a similar strategy to Theorem 4.13. Recall that for any $\mu \in \mathcal{M}_F(\mathbb{R}^d)$, $\langle \mu, F\rangle = \int_{\mathbb{R}} F(x)\mu(dx)$. Then Eq. (4.4.6) is written as

$$\langle \mu_t^{0,\epsilon}, F\rangle = \langle \mu_0, F\rangle + \int_0^t \langle \mu_s^{0,\epsilon}, \mathcal{L}F\rangle ds + \int_0^t \langle \mu_s^{0,\epsilon}, F(\sigma_2^{-1}(s)b_2(s,\cdot)\psi(s))\rangle ds$$
$$+ \int_0^t \int_{\mathbb{U}_2} \langle \mu_s^{0,\epsilon}, F(\lambda(s,\cdot,u)-1)(\phi(s,u)-1)\rangle \nu_2(du)ds.$$

Let $Z_t^\delta = T_\delta \mu_t^{0,\epsilon}$ be defined as in Eq. (4.3.6). Replacing F by $T_\delta F$ and noting that $\langle T_\delta \nu, F\rangle_0 = \langle \nu, T_\delta F\rangle$, we have that

$$\langle Z_t^\delta, F\rangle_0 = \langle \mu_0, T_\delta F\rangle + \int_0^t \langle \mu_s^{0,\epsilon}, \mathcal{L}T_\delta F\rangle ds + \int_0^t \langle \mu_s^{0,\epsilon}, T_\delta F(\sigma_2^{-1}(s)b_2(s,\cdot)\psi(s))\rangle ds$$
$$+ \int_0^t \int_{\mathbb{U}_2} \langle \mu_s^{0,\epsilon}, T_\delta F(\lambda(s,\cdot,u)-1)(\phi(s,u)-1)\rangle \nu_2(du)ds. \tag{4.4.14}$$

Employing Lemma 4.7, Eq. (4.4.14) is given by

$$\langle Z_0^\delta, F\rangle + \int_0^t \langle \partial_i T_\delta(b_1^i \mu_s^{0,\epsilon}), F\rangle_0 ds + \frac{1}{2}\int_0^t \langle \partial_{ij}^2 T_\delta(\sigma_1^{ik}\sigma_1^{kj}\mu_s^{0,\epsilon}), F\rangle_0 ds$$
$$+ \int_0^t \langle T_\delta(\sigma_2^{-1}(s)b_2(s,\cdot)\psi(s)\mu_s^{0,\epsilon}), F\rangle_0 ds$$
$$+ \int_0^t \int_{\mathbb{U}\setminus\mathbb{U}_1} \left[\langle T_\delta \mu_s^{0,\epsilon}, F(\cdot + g_1(\cdot,u))\rangle_0 - \langle T_\delta \mu_s^{0,\epsilon}, F\rangle_0\right] \nu_1(du)ds$$
$$+ \int_0^t \int_{\mathbb{U}_1} \left[\langle T_\delta \mu_s^{0,\epsilon}, F(\cdot + f_1(\cdot,u))\rangle_0 - \langle T_\delta \mu_s^{0,\epsilon}, F\rangle_0 - \langle \partial_i T_\delta(f_1^i(\cdot,u)\mu_s^{0,\epsilon}), F\rangle_0\right] \nu_1(du)ds$$
$$+ \int_0^t \int_{\mathbb{U}_2} \langle T_\delta(\lambda(s,\cdot,u)-1)(\phi(s,u)-1)\mu_s^{0,\epsilon}), F\rangle_0 \nu_2(du)ds.$$

Furthermore, summing $\langle Z_t^\delta, \eta\rangle_0^2 = \langle Z_0^\delta, \eta\rangle_0^2 + \int_0^t 2\langle Z_s^\delta, \eta\rangle_0 d\langle Z_s^\delta, \eta\rangle_0$ over η in a CONS of L^2-space, there exists a constant $K_1 > 0$ such that

$$\|Z_t^\delta\|_0^2 \le \|Z_0^\delta\|_0^2 + K_1 \int_0^t \|T_\delta \mu_s^{0,\epsilon}\|_0^2 ds. \tag{4.4.15}$$

Now let $\mu_{1,t}^{0,\epsilon}$ and $\mu_{2,t}^{0,\epsilon}$ be two solutions with the same initial value μ_0. Then by Eq. (4.4.15), we get

$$\|T_\delta(\mu_{1,t}^{0,\epsilon} - \mu_{2,t}^{0,\epsilon})\|_0^2 \le K_1 \int_0^t \|T_\delta(\mu_{1,s}^{0,\epsilon} - \mu_{2,s}^{0,\epsilon})\|_0^2 ds.$$

Taking $\delta \to 0$ gives

$$\|\mu_{1,t}^{0,\epsilon} - \mu_{2,t}^{0,\epsilon}\|_0^2 \le K_1 \int_0^t \|\mu_{1,s}^{0,\epsilon} - \mu_{2,s}^{0,\epsilon}\|_0^2 ds,$$

and by Gronwall's inequality, we have $\mu_{1,t}^{0,\epsilon} - \mu_{2,t}^{0,\epsilon} = 0$. ■

The following proposition demonstrates the first statement in Condition 4.2.1.

Proposition 4.23. *For $M < \infty$, suppose that $\mu_0^n \to \mu_0$, and $\xi_n = (\psi_n, \phi_n)$, $\xi = (\psi, \phi) \in \mathcal{U}^M$ such that $\xi_n \to \xi$, as $n \to \infty$. Then $\mathcal{G}^0(\mu_0^n, \int_0^\cdot \psi_n(s)ds, \nu_2^{\phi_n}) \to \mathcal{G}^0(\mu_0, \int_0^\cdot \psi(s)ds, \nu_2^{\phi})$ as $n \to \infty$.*

Proof. Consider

$$\Lambda_t^{0,\xi} = \exp\left(\int_0^t \sigma_2^{-1}(s)b_2(s, X_s)\psi(s)ds + \int_0^t \int_{\mathbb{U}_2} (\lambda(s, X_s, \mu) - 1)(\phi(s,u) - 1)\nu_2(du)ds\right).$$

The proof of the existence in Theorem 4.22 shows that $\rho_t F = \int_{\mathbb{R}^d} \mu_0(dx)\mathbb{E}_x \Lambda_t^{0,\xi} F(X_t)$ is a solution of zero-noise version of Zakai Eq. (4.4.6). Furthermore, the solution is unique according to Theorem 4.22 and hence we get $\mu_t^{0,\xi} = \rho_t$. Notice that, by the boundedness of σ_2^{-1}, b_2 and $(\psi_n, \phi_n) \in \mathcal{U}^M$, there exists a constant $K > 0$ such that

$$\begin{aligned}\int_0^T |\sigma_2^{-1}(s)b_2(s, X_s)\psi_n(s)|ds &\le \left(\int_0^T |\sigma_2^{-1}(s)b_2(s, X_s)|^2 ds\right)^{1/2} \left(\int_0^T |\psi_n(s)|^2 ds\right)^{1/2} \\ &\le (KT)^{1/2} M^{1/2}.\end{aligned}$$

In addition, by the inequality, for any $\phi \in [0, \infty)$,

$$|\phi - 1| \leq 2 + \phi \log \phi - \phi + 1, \tag{4.4.16}$$

we have

$$\begin{aligned}
&\int_0^T \int_{\mathbb{U}_2} |(\lambda(s, X_s, \mu) - 1)(\phi_n(s,u) - 1)| \nu_2(du) ds \\
\leq &2 \int_0^T \int_{\mathbb{U}_2} (2 + \phi_n(s,u) \log \phi_n(s,u) - \phi_n(s,u) + 1) \, \nu_2(du) ds \\
\leq &4T\nu_2(\mathbb{U}_2) + 2M.
\end{aligned}$$

The dominated convergence theorem yields the convergence below, as $n \to \infty$,

$$\begin{aligned}
\mu_t^{0,\xi_n}(F) = &\int_{\mathbb{R}} \mu_0^n(dx) \mathbb{E}_x F(X_t) \exp \bigg(\int_0^t \sigma_2^{-1}(s) b_2(s, X_s) \psi_n(s) ds \\
&+ \int_0^t \int_{\mathbb{U}_2} (\lambda(s, X_s, \mu) - 1)(\phi_n(s,u) - 1) \nu_2(du) ds \bigg) \\
\to &\int_{\mathbb{R}} \mu_0(dx) \mathbb{E}_x F(X_t) \exp \bigg(\int_0^t \sigma_2^{-1}(s) b_2(s, X_s) \psi(s) ds \\
&+ \int_0^t \int_{\mathbb{U}_2} (\lambda(s, X_s, \mu) - 1)(\phi(s,u) - 1) \nu_2(du) ds \bigg) \\
= &\mu_t^{0,\xi}(F).
\end{aligned}$$

Then the proof is completed. ■

The next proposition verifies the second part of Condition 4.2.1.

Proposition 4.24. *For $M < \infty$, let $\xi^\epsilon = (\psi^\epsilon, \phi^\epsilon)$, $\xi = (\psi, \phi) \in \mathcal{U}^M$ be such that ξ^ϵ converges to ξ in distribution and $\mu_0^\epsilon \to \mu_0$ as $\epsilon \to 0$. Then $\mathcal{G}^\epsilon(\mu_0^\epsilon, \sqrt{\epsilon}\tilde{W}^\epsilon + \int_0^\cdot \psi^\epsilon(s) ds, \epsilon N_{\lambda^\epsilon}^{\epsilon^{-1}\phi^\epsilon})$ converges to $\mathcal{G}^0(\mu_0, \int_0^\cdot \psi(s) ds, \nu_2^\phi)$ in distribution as $\epsilon \to 0$.*

Proof. First, we prove that the family $\{\mu^{\epsilon,\xi^\epsilon,\mu_0} = \mathcal{G}^\epsilon(\mu_0, \sqrt{\epsilon}\tilde{W}^\epsilon + \int_0^\cdot \psi^\epsilon(s)ds, \epsilon N_{\lambda^\epsilon}^{\epsilon^{-1}\phi^\epsilon}), \epsilon \in (0, \epsilon_0)\}$ is tight in $D([0,T], \mathcal{M}_F(\mathbb{R}^d))$ for some $\epsilon_0 > 0$. Note that

$$
\begin{aligned}
\mu_t^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}(1) = \mu_0(1) &+ \sqrt{\epsilon}\sum_{i=1}^m \int_0^t \mu_s^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}((\sigma_2^{-1}(s)b_2(s,\cdot))^i)d\tilde{W}_s^{\epsilon,i} \\
&+ \int_0^t \mu_s^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}(\sigma_2^{-1}(s)b_2(s,\cdot))\psi^\epsilon(s)ds \\
&+ \int_0^t \int_{\mathbb{U}_2} \mu_s^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}(\lambda(s,\cdot,\mu)-1)\left(\phi^\epsilon(s,u)-1\right)\nu_2(du)ds \\
&+ \epsilon\int_0^t \int_{\mathbb{U}_2} \mu_{s-}^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}(\lambda(s,\cdot,\mu)-1)(N_{\lambda^\epsilon}^{\epsilon^{-1}\phi^\epsilon}(ds,du) - \epsilon^{-1}\phi^\epsilon(s,u)\nu_2(du)ds).
\end{aligned}
$$

By Itô's formula, we get $\mu_t^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}(1)^2 = A_t^1 + A_t^2 + A_t^3 + A_t^4 + A_t^5 + A_t^6 + A_t^7$ where

$$
\begin{aligned}
A_t^1 &= \mu_0^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}(1)^2, \quad A_t^2 = 2\int_0^t \mu_s^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}(1)\mu_s^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}(\sigma_2^{-1}(s)b_2(s,\cdot))\psi^\epsilon(s)ds, \\
A_t^3 &= 2\int_0^t \int_{\mathbb{U}_2} \mu_s^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}(1)\mu_s^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}(\lambda(s,\cdot,\mu)-1)\left(\phi^\epsilon(s,u)-1\right)\nu_2(du)ds, \\
A_t^4 &= \epsilon\int_0^t \left|\mu_s^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}(\sigma_2^{-1}(s)b_2(s,\cdot))\right|^2 ds, \\
A_t^5 &= \int_0^t \int_{\mathbb{U}_2} \epsilon^2 \mu_s^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}(\lambda(s,\cdot,\mu)-1)^2\phi^\epsilon(s,u)\nu_2(du)ds, \\
A_t^6 &= 2\sqrt{\epsilon}\sum_{i=1}^m \int_0^t \mu_s^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}(1)\mu_s^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}((\sigma_2^{-1}(s)b_2(s,\cdot))^i)d\tilde{W}_s^{\epsilon,i}, \\
A_t^7 &= \int_0^t \int_{\mathbb{U}_2} \Big[2\epsilon\mu_{s-}^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}(1)\mu_{s-}^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}(\lambda(s,\cdot,\mu)-1)\,, \\
&\quad + \epsilon^2\mu_{s-}^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}((\lambda(s,\cdot,\mu)-1))^2\Big]\left(N_{\lambda^\epsilon}^{\epsilon^{-1}\phi^\epsilon}(ds,du) - \epsilon^{-1}\phi^\epsilon(s,u)\nu_2(du)ds\right).
\end{aligned}
$$

Considering Assumption 4.3.2 and Cauchy–Schwarz inequality, A_t^3 is bounded by

$$
\begin{aligned}
&2K\int_0^t \mu_s^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}(1)^2|\psi^\epsilon(s)|ds \\
&\le 2K\left(\int_0^t \mu_s^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}(1)^2|\psi^\epsilon(s)|^2ds\right)^{1/2}\left(\int_0^t \mu_s^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}(1)^2ds\right)^{1/2} \\
&\le K\int_0^t \mu_s^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}(1)^2|\psi^\epsilon(s)|^2ds + K\int_0^t \mu_s^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}(1)^2ds, \qquad (4.4.17)
\end{aligned}
$$

and $A_t^3 + A_t^4 + A_t^5$ is bounded by

$$4\int_0^t \int_{\mathbb{U}_2} \mu_s^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}(1)^2 \, |\phi^\epsilon(s,u) - 1| \, \nu_2(du)ds \\ + K\epsilon \int_0^t \mu_s^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}(1)^2 ds + 4\epsilon^2 \int_0^t \int_{\mathbb{U}_2} \mu_s^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}(1)^2 |\phi^\epsilon(s,u)| \nu_2(du)ds. \quad (4.4.18)$$

Set

$$A_t^{7,1} = \int_0^t \int_{\mathbb{U}_2} 2\epsilon \mu_{s-}^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}(1) \mu_{s-}^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}(\lambda(s,\cdot,\mu) - 1) \left(N_{\lambda^\epsilon}^{\epsilon^{-1}\phi^\epsilon}(ds,du) - \epsilon^{-1}\phi^\epsilon(s,u)\nu_2(du)ds \right) \quad (4.4.19)$$

and

$$A_t^{7,2} = \int_0^t \int_{\mathbb{U}_2} \epsilon^2 \mu_{s-}^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}((\lambda(s,\cdot,\mu) - 1))^2 \left(N_{\lambda^\epsilon}^{\epsilon^{-1}\phi^\epsilon}(ds,du) - \epsilon^{-1}\phi^\epsilon(s,u)\nu_2(du)ds \right). \quad (4.4.20)$$

Combined Eqs. (4.4.17) and (4.4.18) implies that

$$\mu_t^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}(1)^2 \leq \mu_0^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}(1)^2 + \sup_{0\leq t\leq T} |A_t^6| + \sup_{0\leq t\leq T} |A_t^{7,1}| + \sup_{0\leq t\leq T} |A_t^{7,2}| \\ + \int_0^t \mu_s^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}(1)^2 \left(K|\psi^\epsilon(s)|^2 + K + \int_{\mathbb{U}_2} \left(|\phi^\epsilon(s,u) - 1| + K\epsilon + 4\epsilon^2 \phi^\epsilon(s,u) \right) \nu_2(du) \right) ds.$$

Furthermore, inequality (4.4.16) gives

$$\begin{aligned} \int_0^T \int_{\mathbb{U}_2} |\phi^\epsilon(s,u)| \, \nu_2(du)ds &\leq \int_0^T \int_{\mathbb{U}_2} (|\phi^\epsilon(s,u) - 1| + 1) \, \nu_2(du)ds) \\ &\leq \int_0^T \int_{\mathbb{U}_2} (3 + \phi^\epsilon(s,u) \log \phi^\epsilon(s,u) - \phi^\epsilon(s,u) + 1) \, \nu_2(du)ds) \\ &\leq 3T\nu_2(\mathbb{U}_2) + M. \end{aligned} \quad (4.4.21)$$

Then by Gronwall's inequality and Eq. (4.4.21), we have

$$\mu_t^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}(1)^2 \leq C_0 \left(\mu_0^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}(1)^2 + \sup_{0\leq t\leq T} |A_t^6| + \sup_{0\leq t\leq T} |A_t^{7,1}| + \sup_{0\leq t\leq T} |A_t^{7,2}| \right), \quad (4.4.22)$$

where $C_0 = \exp\left(KM + KT + K\epsilon T + 2T\nu_2(\mathbb{U}_2) + M + 4\epsilon^2(3T\nu_2(\mathbb{U}_2) + M)\right)$. Next, we have

$$\begin{aligned}
\tilde{\mathbb{E}}^\epsilon\left(\sup_{0\le t\le T}|A_t^6|\right) \le& 8K\sqrt{\epsilon}\tilde{\mathbb{E}}^\epsilon\sqrt{\int_0^T \mu_s^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}(1)^4 ds} \\
\le& 8K\sqrt{T}\sqrt{\epsilon}\tilde{\mathbb{E}}^\epsilon\left(\sup_{0\le t\le T}\mu_s^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}(1)^2\right). \qquad (4.4.23)
\end{aligned}$$

For Eq. (4.4.19), by BDG inequality, we have

$$\begin{aligned}
&\tilde{\mathbb{E}}^\epsilon\left(\sup_{0\le t\le T}|A_t^{7,1}|\right) \\
\le& 8\epsilon\tilde{\mathbb{E}}^\epsilon\left(\int_0^T\int_{\mathbb{U}_2}\mu_{s-}^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}(1)^2\mu_{s-}^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}(\lambda(s,\cdot,\mu)-1)^2 N_{\lambda^\epsilon}^{\epsilon^{-1}\phi^\epsilon}(ds,du)\right)^{\frac{1}{2}} \\
\le& 16\epsilon\tilde{\mathbb{E}}^\epsilon\left(\sup_{0\le t\le T}\mu_t^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}(1)\cdot\left(\int_0^T\int_{\mathbb{U}_2}\mu_{s-}^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}(1)^2 N_{\lambda^\epsilon}^{\epsilon^{-1}\phi^\epsilon}(ds,du)\right)^{\frac{1}{2}}\right) \\
\le& \frac{1}{C}\tilde{\mathbb{E}}^\epsilon\left(\sup_{0\le t\le T}\mu_t^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}(1)^2\right) + 64\epsilon^2 C\tilde{\mathbb{E}}^\epsilon\left(\int_0^T\int_{\mathbb{U}_2}\mu_{s-}^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}(1)^2 N_{\lambda^\epsilon}^{\epsilon^{-1}\phi^\epsilon}(ds,du)\right) \\
=& \frac{1}{C}\tilde{\mathbb{E}}^\epsilon\left(\sup_{0\le t\le T}\mu_t^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}(1)^2\right) + 64\epsilon C\left(\int_0^T\int_{\mathbb{U}_2}\mu_s^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}(1)^2\phi^\epsilon(s,u)\nu_2(du)ds\right) \\
\le& \left(\frac{1}{C} + 64\epsilon C(3T\nu_2(\mathbb{U}_2)+M)\right)\tilde{\mathbb{E}}^\epsilon\left(\sup_{0\le t\le T}\mu_t^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}(1)^2\right), \qquad (4.4.24)
\end{aligned}$$

where C is any positive number. Eq. (4.4.19) is bounded by

$$\begin{aligned}
\tilde{\mathbb{E}}^\epsilon\left(\sup_{0\le t\le T}|A_t^{7,2}|\right) \le& \tilde{\mathbb{E}}^\epsilon\left(\int_0^T\int_{\mathbb{U}_2}\epsilon^2\mu_{s-}^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}((\lambda(s,\cdot,\mu)-1))^2 N_{\lambda^\epsilon}^{\epsilon^{-1}\phi^\epsilon}(ds,du)\right) \\
&+ \epsilon\tilde{\mathbb{E}}^\epsilon\left(\int_0^T\int_{\mathbb{U}_2}\mu_s^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}((\lambda(s,\cdot,\mu)-1))^2\phi^\epsilon(s,u)\nu_2(du)ds)\right) \\
\le& 8\epsilon\tilde{\mathbb{E}}^\epsilon\left(\int_0^T\int_{\mathbb{U}_2}\mu_s^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}(1)^2\phi^\epsilon(s,u)\nu_2(du)ds)\right) \\
\le& 8\epsilon\tilde{\mathbb{E}}^\epsilon\left(\sup_{0\le t\le T}\mu_t^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}(1)^2\right)\left(\int_0^T\int_{\mathbb{U}_2}\phi^\epsilon(s,u)\nu_2(du)ds)\right) \\
\le& \epsilon\left(24T\nu_2(\mathbb{U}_2)+8M\right)\tilde{\mathbb{E}}^\epsilon\left(\sup_{0\le t\le T}\mu_t^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}(1)^2\right), \qquad (4.4.25)
\end{aligned}$$

where inequality (4.4.21) was used. By Eqs. (4.4.23), (4.4.24) and (4.4.25), Eq. (4.4.22) turns out to be

$$\begin{aligned}\tilde{\mathbb{E}}^\epsilon\left(\sup_{0\le t\le T}\mu_t^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}(1)^2\right) &\le C_0\mu_0^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}(1)^2\\ &\quad + C_0\left(8K\sqrt{T}\sqrt{\epsilon}+\epsilon\left(24T\nu_2(\mathbb{U}_2)+8M\right)+\frac{1}{C}+64\epsilon C(3T\nu_2(\mathbb{U}_2)+M)\right).\end{aligned}$$

Recall that the constant C can be arbitrarily large, we hence select C and ϵ_0 small enough such that

$$C_0\left(8K\sqrt{T}\sqrt{\epsilon}+\epsilon\left(24T\nu_2(\mathbb{U}_2)+8M\right)+\frac{1}{C}+64\epsilon C(3T\nu_2(\mathbb{U}_2)+M)\right)<\frac{1}{2}.$$

Therefore, there is a constant K_1 such that

$$\sup_{0<\epsilon\le\epsilon_0}\tilde{\mathbb{E}}^\epsilon\left(\sup_{0\le t\le T}\mu_t^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}(1)^2\right)\le K_1. \tag{4.4.26}$$

We now establish the tightness of $\{\mu^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}\}$. It is well-known, e.g. see [43] that we only need to prove the tightness of $\{\mu_t^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}(F)\}$ in $D([0,T],\mathbb{R})$ for every test function F in $\mathcal{D}(\mathcal{L})$. By Eq. (4.4.26) and the boundness of F, we have

$$\sup_{0<\epsilon\le\epsilon_0}\tilde{\mathbb{E}}^\epsilon\left(\sup_{0\le t\le T}\mu_t^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}(F)^2\right)\le K_2. \tag{4.4.27}$$

Denote

$$\begin{aligned}A_t^\epsilon &= \int_0^t\mu_s^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}(\mathcal{L}F)ds+\int_0^t\mu_s^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}(F(\sigma_2^{-1}(s)b_2(s,\cdot)))\psi^\epsilon(s)ds\\ &\quad+\int_0^t\int_{\mathbb{U}_2}\mu_s^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}(F(\lambda(s,\cdot,\mu)-1))\left(\phi^\epsilon(s,u)-1\right)\nu_2(du)ds.\end{aligned}$$

Let $M_t^\epsilon=M_t^{\epsilon,1}+M_t^{\epsilon,2}$, where

$$M_t^{\epsilon,1}=\sqrt{\epsilon}\sum_{i=1}^m\int_0^t\mu_s^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}(F(\sigma_2^{-1}(s)b_2(s,\cdot))^i)d\tilde{W}_s^{\epsilon,i}$$

and

$$M_t^{\epsilon,2} = \int_0^t \int_{\mathbb{U}_2} \mu_{s-}^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}(F(\lambda(s,\cdot,\mu)-1)) \left(\epsilon N_{\lambda^\epsilon}^{\epsilon^{-1},\phi^\epsilon}(ds,du) - \phi^\epsilon(s,u)\nu_2(du)ds\right).$$

To verify the tightness of A_t^ϵ in $D([0,T],\mathbb{R})$, it suffices to show (see Lemma 6.1.2 of [43]) that for all $\delta > 0$, there exists $\tau = \tau_\delta > 0$ such that

$$\sup_{0\le\epsilon\le\epsilon_0} \tilde{\mathbb{P}}^\epsilon \left(\sup_{0<t_1<t_2<\tau} |A_{t_1}^\epsilon - A_{t_2}^\epsilon| > \delta \right) < \delta.$$

Then for arbitrary $\tau > 0$ and a fixed $\delta > 0$

$$\begin{aligned}
&\sup_{0\le\epsilon\le\epsilon_0} \tilde{\mathbb{P}}^\epsilon \left(\sup_{0<t_1<t_2<\tau} |A_{t_1}^\epsilon - A_{t_2}^\epsilon| > \delta \right) \\
&= \sup_{0\le\epsilon\le\epsilon_0} \tilde{\mathbb{P}}^\epsilon \left(\sup_{0<t_1<t_2<\tau} \left| \int_{t_1}^{t_2} \mu_s^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}(\mathcal{L}F)ds + \int_{t_1}^{t_2} \mu_s^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}(F(\sigma_2^{-1}(s)b_2(s,\cdot)))\psi^\epsilon(s)ds \right.\right. \\
&\qquad\qquad \left.\left. + \int_{t_1}^{t_2}\int_{\mathbb{U}_2} \mu_s^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}(F(\lambda(s,\cdot,\mu)-1))\left(\phi^\epsilon(s,u)-1\right)\nu_2(du)ds \right| > \delta \right) \\
&\le \sup_{0\le\epsilon\le\epsilon_0} \tilde{\mathbb{P}}^\epsilon \left(\sup_{0<t_1<t_2<\tau} \left| \int_{t_1}^{t_2} \mu_s^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}(\mathcal{L}F)ds \right| > \frac{\delta}{3} \right) \\
&\quad + \sup_{0\le\epsilon\le\epsilon_0} \tilde{\mathbb{P}}^\epsilon \left(\sup_{0<t_1<t_2<\tau} \left| \int_{t_1}^{t_2} \mu_s^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}(F(\sigma_2^{-1}(s)b_2(s,\cdot)))\psi^\epsilon(s)ds \right| > \frac{\delta}{3} \right) \qquad (4.4.28) \\
&\quad + \sup_{0\le\epsilon\le\epsilon_0} \tilde{\mathbb{P}}^\epsilon \left(\sup_{0<t_1<t_2<\tau} \left| \int_{t_1}^{t_2}\int_{\mathbb{U}_2} \mu_s^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}(F(\lambda(s,\cdot,\mu)-1))\left(\phi^\epsilon(s,u)-1\right)\nu_2(du)ds \right| > \frac{\delta}{3} \right) \\
&:= D_1 + D_2 + D_3.
\end{aligned}$$

The term D_1 in Eq. (4.4.28) is bounded by

$$\sup_{0\le\epsilon\le\epsilon_0} \frac{9}{\delta^2}\tilde{\mathbb{E}}^\epsilon \left(\sup_{0<t_1<t_2<\tau} \int_{t_1}^{t_2} \mu_s^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}(\mathcal{L}F)ds \right)^2 \le \sup_{0\le\epsilon\le\epsilon_0} \frac{9}{\delta^2}\tilde{\mathbb{E}}^\epsilon \left(\delta^2 \sup_{0\le t\le T} \mu_t^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}(\mathcal{L}F)^2 \right). \qquad (4.4.29)$$

By Eq. (4.4.27) and Assumption 4.4.1, we can find $\tau_1 > 0$ such that for all $\tau \le \tau_1$, Eq. (4.4.29) is bounded by $\delta/3$.

For D_2 in Eq. (4.4.28), we note that

$$\begin{aligned}
&\tilde{\mathbb{E}}^\epsilon \left| \int_{t_1}^{t_2} \mu_s^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}(F(\sigma_2^{-1}(s)b_2(s,\cdot)))\psi^\epsilon(s)ds \right|^2 \\
\leq &\tilde{\mathbb{E}}^\epsilon \left(\int_{t_1}^{t_2} \left| \mu_s^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}(F(\sigma_2^{-1}(s)b_2(s,\cdot))) \right|^2 ds \int_{t_1}^{t_2} |\psi^\epsilon(s)|^2 ds \right) \leq MK_2|t_2 - t_1|.
\end{aligned}$$

Thus, we can find $\tau_2 > 0$ such that for all $\tau \leq \tau_2$ the second term D_2 in Eq. (4.4.28) is bounded by $\delta/3$.

Now let us consider D_3 in Eq. (4.4.28). Using the fundamental inequality, for $a, b \in (0, \infty)$ and any $c \in [1, \infty)$, $ab \leq e^{ac} + \frac{1}{c}(b \log b - \log b + 1) = e^{ac} + \frac{1}{c}l(b)$, twice (once with $b = \phi^\epsilon$ and once with $b = 1$), for any $C_1 \in (1, \infty)$ we have

$$\begin{aligned}
&\left| \int_{t_1}^{t_2} \int_{\mathbb{U}_2} \mu_s^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}(F(\lambda(s,\cdot,\mu) - 1)) \left(\phi^\epsilon(s,u) - 1\right) \nu_2(du)ds \right| \\
\leq &2 \sup_{0\leq t\leq T} \mu_t^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}(F) \int_{t_1}^{t_2} \int_{\mathbb{U}_2} (\phi^\epsilon(s,u) + 1)\nu_2(du)ds \\
\leq &2 \sup_{0\leq t\leq T} \mu_t^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}(F) \left(2\tau\nu_2(\mathbb{U}_2)e^{C_1} + \frac{M}{C_1} \right). \qquad (4.4.30)
\end{aligned}$$

Given any $\epsilon > 0$, we can choose C_1 such that $\frac{M}{C_1} \leq \epsilon$, and choose $\tau > 0$ such that $2\tau\nu_2(\mathbb{U}_2)e^{C_1} \leq \epsilon$. Hence combining Eqs. (4.4.27) and (4.4.30) we proved that

$$\lim_{\tau\to 0} \sup_{0<t_1<t_2<\tau} \left| \int_{t_1}^{t_2} \int_{\mathbb{U}_2} \mu_s^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}(F(\lambda(s,\cdot,\mu) - 1)) \left(\phi^\epsilon(s,u) - 1\right) \nu_2(du)ds \right| = 0.$$

Thus we find $\tau_3 > 0$ such that for $\tau \leq \tau_3$, D_3 in Eq. (4.4.28) is bounded by $\delta/3$. Consequently, the tightness of $\{A^\epsilon\}_{\epsilon\leq\epsilon_0}$ follows from Eq. (4.4.28) by taking $\tau = \min\{\tau_1, \tau_2, \tau_3\}$.

Next, considering M^ϵ we have $\tilde{\mathbb{E}}^\epsilon\langle M^{\epsilon,1}\rangle_T \leq \epsilon K \tilde{\mathbb{E}}^\epsilon \int_0^T \mu_t^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}(F)^2 ds$ and

$$\begin{aligned}
\tilde{\mathbb{E}}^\epsilon\langle M^{\epsilon,2}\rangle_T &= \epsilon\tilde{\mathbb{E}}^\epsilon \left(\int_0^T \int_{\mathbb{U}_2} \mu_t^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}(F(\lambda(s,\cdot,\mu) - 1))^2 \phi^\epsilon(s,u)\nu_2(du)ds \right) \\
&\leq \epsilon 4\tilde{\mathbb{E}}^\epsilon \left(\sup_{0\leq t\leq T} \mu_t^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}(F)^2 \int_0^T \int_{\mathbb{U}_2} \phi^\epsilon(s,u)\nu_2(du)ds \right).
\end{aligned}$$

By Eqs. (4.4.21) and (4.4.27), we have $\tilde{\mathbb{E}}^\epsilon \sup_{0\le t\le T}\langle M^\epsilon\rangle_t$ converges to 0 as $\epsilon \to 0$. Then according to Theorem 6.1.1 in [43], for any $F \in \mathcal{D}(\mathcal{L})$, the sequence of semimartingales $\mu_t^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}(F) = \mu_0^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}(F) + A_t^\epsilon + M_t^\epsilon$ is tight in $D([0,T],\mathbb{R})$.

Next we show that $\mathcal{G}^0(\mu_0, \int_0^\cdot \psi(s)ds, \nu_2^\phi)$ is the weak limit of $\mu^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}$. Note that

$$\tilde{\mathbb{E}}^\epsilon\left(\sup_{0\le t\le T}\left|M_t^{\epsilon,1}\right|^2\right) \le \epsilon K \tilde{\mathbb{E}}^\epsilon \int_0^T \mu_t^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}(F)^2 ds$$

and

$$\tilde{\mathbb{E}}^\epsilon\left(\sup_{0\le t\le T}\left|M_t^{\epsilon,2}\right|^2\right) \le \epsilon K \tilde{\mathbb{E}}^\epsilon\left(\sup_{0\le t\le T}\mu_t^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}(F)^2 \int_0^T \int_{\mathbb{U}_2} \phi^\epsilon(s,u)\nu_2(du)ds\right).$$

Therefore, $M_t^{\epsilon,1}$ and $M_t^{\epsilon,2}$ converge to 0 in distribution as $\epsilon \to 0$. Let $(\mu_t^{0,\xi,\mu_0}, \psi, \phi, 0, 0)$ be any limit point of the tight sequence $\{(\mu_t^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}, \psi^\epsilon, \phi^\epsilon, M_t^{\epsilon,1}, M_t^{\epsilon,2}), \epsilon \in (0,\epsilon_0)\}$. Without loss of generality, we assume that the convergence is almost sure by using the Skorokhod representation theorem. Note that for any test function F in $\mathcal{D}(\mathcal{L})$, we have

$$\begin{aligned}\mu_t^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}(F) &= \mu_0^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}(F) + \int_0^t \mu_s^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}(\mathcal{L}F)ds + M_t^{\epsilon,1} + \int_0^t \mu_s^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}(F(\sigma_2^{-1}(s)b_2(s,\cdot)))\psi^\epsilon(s)ds \\ &\quad + M_t^{\epsilon,2} + \int_0^t \int_{\mathbb{U}_2} \mu_s^{\epsilon,\xi^\epsilon,\mu_0^\epsilon}(F(\lambda(s,\cdot,\mu)-1))\left(\phi^\epsilon(s,u)-1\right)\nu_2(du)ds.\end{aligned}$$

Taking $\epsilon \to 0$, along the lines of Theorem 4.22 and Proposition 4.23 we see that μ^{0,ξ,μ_0} must solve Eq. (4.4.6). Then the uniqueness of the solution to Eq. (4.4.6) completes the proof. ∎

Bibliography

[1] Andrew, W. L. (1988). Maximum likelihood estimation of generalized Itô processes with discretely sampled data. *Econometric Theory*, 4(02):231–247. 2

[2] Applebaum, D. (2009). *Lévy processes and stochastic calculus*. Cambridge university press. 83, 84, 85

[3] Arcones, M. A. (2006). Large deviations for M-estimators. *Annals of the Institute of Statistical Mathematics*, 58(1):21–52. 3

[4] Bahadur, R. R. (1960). On the asymptotic efficiency of tests and estimates. *Sankhyā: The Indian Journal of Statistics*, pages 229–252. 3

[5] Bain, A. and Crisan, D. (2009). *Fundamentals of stochastic filtering*, volume 3. Springer. 1

[6] Bao, F. and Maroulas, V. (2017). Adaptive meshfree backward SDE filter. *SIAM Journal on Scientific Computing*, 39(6):A2664–A2683. 1

[7] Bao, J. and Yuan, C. (2015). Large deviations for neutral functional SDEs with jumps. *Stochastics*, 87(1):48–70. 4

[8] Billingsley, P. (1995). *Probability and measure*. Wiley Interscience, Series in Probability and Mathematical Statistics, New York. 13

[9] Björk, T., Kabanov, Y., and Runggaldier, W. (1997). Bond market structure in the presence of marked point processes. *Mathematical Finance*, 7(2):211–239. 42, 56

[10] Brémaud, P. (1981). *Point processes and queues: martingale dynamics*, volume 50. Springer-Verlag. 42, 56

[11] Budhiraja, A., Chen, J., and Dupuis, P. (2013). Large deviations for stochastic partial differential equations driven by a Poisson random measure. *Stochastic Processes and their Applications*, 123(2):523–560. 4

[12] Budhiraja, A., Dupuis, P., and Maroulas, V. (2011). Variational representations for continuous time processes. *Ann. Inst. Henri Poincaré Probab. Stat*, 47(3):725–747. 4, 36, 37, 40

[13] Ceci, C. and Colaneri, K. (2012). Nonlinear filtering for jump diffusion observations. *Advances in Applied Probability*, 44(03):678–701. 2

[14] Ceci, C. and Colaneri, K. (2014). The Zakai equation of nonlinear filtering for jump-diffusion observations: existence and uniqueness. *Applied Mathematics & Optimization*, 69(1):47–82. 2, 5

[15] Charalambous, C. D. and Logothetis, A. (2000). Maximum likelihood parameter estimation from incomplete data via the sensitivity equations: The continuous-time case. *IEEE Transactions on Automatic Control*, 45(5):928–934. 2, 3

[16] Chen, X. (2010). *Random walk intersections: Large deviations and related topics*. Number 157. American Mathematical Soc. 28

[17] Chernoff, H. (1952). A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *The Annals of Mathematical Statistics*, pages 493–507. 3

[18] Craine, R., Lochstoer, L. A., and Syrtveit, K. (2000). Estimation of a stochastic-volatility jump-diffusion model. *Revista de Analisis Economico*, 15:61–87. 2

[19] Creal, D. D. (2008). Analysis of filtering and smoothing algorithms for Lévy-driven stochastic volatility models. *Computational Statistics & Data Analysis*, 52(6):2863–2876. 2

[20] Dembo, A. and Zeitouni, O. (1986). Parameter estimation of partially observed continuous time stochastic processes via the EM algorithm. *Stochastic Processes and their Applications*, 23(1):91–113. 3

[21] Dembo, A. and Zeitouni, O. (1998). *Large deviations techniques and applications, 2nd edition*. Springer-Verlag, New York. 3, 27, 28, 59

[22] Duan, J. and Millet, A. (2009). Large deviations for the Boussinesq equations under random influences. *Stochastic Processes and their Applications*, 119(6):2052–2081. 4

[23] Dupuis, P. and Ellis, R. S. (2011). *A weak convergence approach to the theory of large deviations*, volume 902. John Wiley & Sons. 3, 38

[24] Ermakov, M. (2008). On semiparametric statistical inferences in the moderate deviation zone. *Journal of Mathematical Sciences*, 152(6):869–874. 3

[25] Evangelou, E. and Maroulas, V. (2017). Sequential empirical Bayes method for filtering dynamic spatiotemporal processes. *Spatial Statistics*, 21:114–129. 2

[26] Fima, C. K. (2005). *An introduction to stochastic calculus with applications (2nd Edition)*. Imperial College Press. 88

[27] Fu, J. C. (1982). Large sample point estimation: A large deviation theory approach. *The Annals of Statistics*, pages 762–771. 3

[28] Gao, F. (2001). Moderate deviations for the maximum likelihood estimator. *Statistics & Probability Letters*, 55(4):345–352. 3

[29] Gao, F. (2003). Moderate deviations and large deviations for kernel density estimators. *Journal of Theoretical Probability*, 16(2):401–418. 4

[30] Gao, F., Xiong, J., and Zhao, X. (2017). Moderate deviations and nonparametric inference for monotone functions. *The Annals of Statistics, to appear*. 3

[31] Gao, F. and Zhao, X. (2011). Delta method in large deviations and moderate deviations for estimators. *The Annals of Statistics*, pages 1211–1240. 3

[32] Gerencsér, L. and Mánfay, M. (2014). Empirical characteristic function identification of linear stochastic systems with possibly unstable zeros. In *Control Conference (ECC), 2014 European*, pages 412–417. IEEE. 2

[33] Grenander, U. and Miller, M. (1994). Representations of knowledge in complex systems. *Journal of the Royal Statistical Society. Series B*, 56(4):549–603. 2

[34] Hanson, F. B. (2007). *Applied stochastic processes and control for jump-diffusions: modeling, analysis, and computation*, volume 13. Society for Industrial and Applied Mathematics. 9, 32, 33

[35] Hijab, O. (1984). Asymptotic Bayesian estimation of a first order equation with small diffusion. *The Annals of Probability*, pages 890–902. 4

[36] Inglot, T. and Ledwina, T. (1990). On probabilities of excessive deviations for kolmogorov-smirnov, cramer-von mises and chi-square statistics. *The Annals of Statistics*, pages 1491–1495. 3

[37] Jacod, J. and Shiryaev, A. (2003). *Limit theorems for stochastic processes*. Berlin : Springer. 42, 56

[38] James, M. R. and Baras, J. S. (1988). Nonlinear filtering and large deviations: A pde-control theoretic approach. *Stochastics*, 23(3):391–412. 4

[39] Johannes, M. S., Polson, N. G., and Stroud, J. R. (2009). Optimal filtering of jump diffusions: Extracting latent states from asset prices. *Review of Financial Studies*, 22(7):2759–2799. 1, 2, 3

[40] Kallenberg, W. C. (1983). On moderate deviation theory in estimation. *The Annals of Statistics*, pages 498–504. 3

[41] Kallianpur, G. (2013). *Stochastic filtering theory*, volume 13. Springer Science & Business Media. 1

[42] Kallianpur, G. and Striebel, C. (1968). Estimation of stochastic systems: Arbitrary system process with additive white noise observation errors. *The Annals of Mathematical Statistics*, 39(3):785–801. 1

[43] Kallianpur, G. and Xiong, J. (1995). Stochastic differential equations in infinite dimensional spaces. *Lecture Notes-Monograph Series*, 26:iii–342. 69, 70, 72

[44] Kang, K., Maroulas, V., Schizas, I., and Bao, F. (2018). Improved distributed particle filters for tracking in a wireless sensor network. *Computational Statistics & Data Analysis*, 117:90–108. 1

[45] Karatzas, I. and Shreve, S. (1998). *Brownian motion and stochastic calculus*. Springer New York. 83, 84, 87

[46] Kester, A. D. M. and Kallenberg, W. C. M. (1986). Large deviations of estimators. *The Annals of Statistics*, pages 648–664. 3

[47] Kurtz, T. G. and Ocone, D. L. (1988). Unique characterization of conditional distributions in nonlinear filtering. *The Annals of Probability*, 16(1):80–107. 1, 5, 44

[48] Kurtz, T. G. and Xiong, J. (1999). Particle representations for a class of nonlinear SPDEs. *Stochastic Processes and their Applications*, 83(1):103–126. 44, 50

[49] Kushner, H. J. (1964). On the dynamical equations of conditional probability density functions, with applications to optimal stochastic control theory. *Journal of Mathematical Analysis and Applications*, 8(2):332–344. 1

[50] Lei, L. and Wu, L. (2005). Large deviations of kernel density estimator in L1 (Rd) for uniformly ergodic Markov processes. *Stochastic Processes and their Applications*, 115(2):275–298. 4

[51] Long, H. (2009). Least squares estimator for discretely observed Ornstein-Uhlenbeck processes with small levy noises. *Statistics & Probability Letters*, 79(19):2076–2085. 2

[52] Louani, D. (1998). Large deviations limit theorems for the kernel density estimator. *Scandinavian Journal of Statistics*, 25(1):243–253. 4

[53] Mahler, R. P. S. (2007). *Statistical multisource-multitarget information fusion.* Artech House. 9

[54] Mandrekar, V., Meyer-Brandis, T., and Proske, F. (2011). A Bayes formula for nonlinear filtering with Gaussian and Cox noise. *Journal of Probability and Statistics*, 2011. 2

[55] Maroulas, V. (2011). Large deviations for infinite-dimensional stochastic systems with jumps. *Mathematika*, 57(01):175–192. 4, 40

[56] Maroulas, V. and Nebenfuhr, A. (2015). Tracking rapid intracellular movements: a Bayesian random set approach. *Annals of Applied Statistics*, 9(2):926–949. 1, 9

[57] Maroulas, V. and Xiong, J. (2013). Large deviations for optimal filtering with fractional Brownian motion. *Stochastic Processes and their Applications*, 123(6):2340–2352. 4

[58] Meyer-Brandis, T. and Proske, F. (2004). Explicit solution of a non-linear filtering problem for Lévy processes with application to finance. *Applied Mathematics & Optimization*, 50(2):119–134. 2

[59] Ogihara, T. and Yoshida, N. (2011). Quasi-likelihood analysis for the stochastic differential equation with jumps. *Statistical Inference for Stochastic Processes*, 14(3):189–229. 2

[60] Pan, X. and Djouadi, S. M. (2017). Estimation and identification for wireless sensor network undergoing uncertain jumps. In *2017 Proceedings of the Conference on Control and its Applications*, pages 46–53. SIAM. 2

[61] Pardoux, É. and Zeitouni, O. (2004). Quenched large deviations for one dimensional nonlinear filtering. *SIAM Journal on Control and Optimization*, 43(4):1272–1297. 4

[62] Poklukar, D. R. (2006). Nonlinear filtering for jump-diffusions. *Journal of computational and applied mathematics*, 197(2):558–567. 1

[63] Popa, S. and Sritharan, S. (2009). Nonlinear filtering of Itô–Lévy stochastic differential equations with continuous observations. *Commun. Stoch. Anal*, 3(313):30. 2

[64] Protter, P. E. (2005). Stochastic differential equations. In *Stochastic integration and differential equations*, pages 249–361. Springer. 88

[65] Puhalskii, A., Spokoiny, V., et al. (1998). On large-deviation efficiency in statistical inference. *Bernoulli*, 4(2):203–272. 3

[66] Qiao, H. and Duan, J. (2015). Nonlinear filtering of stochastic dynamical systems with Lévy noises. *Advances in Applied Probability*, 47(03):902–918. 2, 5, 42, 43, 44, 56, 57

[67] Röckner, M., Zhang, T., and Zhang, X. (2010). Large deviations for stochastic tamed 3D Navier-Stokes equations. *Applied Mathematics & Optimization*, 61(2):267–285. 4

[68] Rozovskii, B. (1990). *Stochastic evolution systems. linear theory and applications to nonlinear filtering*. Kluwer, Dordrcht. 1

[69] Runggaldier, W. J. (2003). Jump-diffusion models. *Handbook of heavy tailed distributions in finance*, 1:169–209. 42, 56

[70] Shi, P., Mahmoud, M., Nguang, S. K., and Ismail, A. (2006). Robust filtering for jumping systems with mode-dependent delays. *Signal Processing*, 86(1):140–152. 1, 2

[71] Shimizu, Y. and Yoshida, N. (2006). Estimation of parameters for diffusion processes with jumps from discrete observations. *Statistical Inference for Stochastic Processes*, 9(3):227–277. 2

[72] Vacarescu, A. P. (2011). *Filtering and parameter estimation for partially observed generalized Hawkes processes*. PhD thesis, Stanford University. 3

[73] van der Vaart, A. W. and Wellner, J. A. (1997). Weak convergence and empirical processes with applications to statistics. *Journal of the Royal Statistical Society-Series A Statistics in Society*, 160(3):596–608. 84

[74] Wei, Y., Qiu, J., Karimi, H. R., and Wang, M. (2013). A new design of H_∞ filtering for continuous-time Markovian jump systems with time-varying delay and partially accessible mode information. *Signal processing*, 93(9):2392–2407. 2

[75] Worms, J. (2001). Moderate deviations of some dependent variables, part II: Some kernel estimators. *Mathematical Methods of Statistics*, 10(2):161–193. 4

[76] Xiong, J. (2003). Large deviation principle for optimal filtering. *Applied Mathematics and Optimization*, 47(2):151–165. 4

[77] Xiong, J. (2008). *An introduction to stochastic filtering theory*, volume 18. Oxford University Press. 1, 42, 44, 45

[78] Xiong, J. and Zeng, Y. (2011). A branching particle approximation to a filtering micromovement model of asset price. *Statistical Inference for Stochastic Processes*, 14(12):111–140. 1, 2

[79] Xiong, J. and Zhai, J. (2018). Large deviations for locally monotone stochastic partial differential equations driven by Lévy noise. *Bernoulli*, 24(4A):2842–2874. 4

[80] Xiong, J. and Zhao, X. (2005). Nonlinear filtering with fractional Brownian motion noise. *Stochastic Analysis and Applications*, 23(1):55–67. 1

[81] Yang, X., Zhai, J., and Zhang, T. (2015). Large deviations for SPDEs of jump type. *Stochastics and Dynamics*, 15(04):1550026. 4

[82] Zakai, M. (1969). On the optimal filtering of diffusion processes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 11(3):230–243. 1

[83] Zeng, Y. (2003). A partially observed model for micromovement of asset prices with Bayes estimation via filtering. *Mathematical Finance*, 13(3):411–444. 1, 2

Appendices

A Some basic definitions

In this appendix, we review some basic knowledge in stochastic analysis. These definitions and principle theorems are used in the main chapters.

Definition A.1 ([2]). *Let $L = (L_t)_{t\geq 0}$ be a stochastic process defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We say L is a Lévy process if*

- $L_0 = 0$ *almost surely;*
- L *has independent and stationary increments, i.e.,* $L_{t_n} - L_{t_{n-1}}, \cdots, L_{t_1} - L_{t_0}$ *are independent for each* $n \geq 1, 0 \leq t_0 < t_1 < t_2 < \cdots < t_n < \infty$*;*
- L *has stationary increments, i.e.,* $L_t - L_s \overset{d}{=} L_{t+h} - L_{s+h}$ *for all* $0 \leq s < t, h \geq s$*;*
- L *is continuous in probability, i.e. for all* $a > 0$ *and for all* $s \geq 0$

$$\lim_{t\to s} \mathbb{P}(|L_t - L_s| > a) = 0.$$

The following are two most important examples of Lévy processes: Brownian motion, Poisson process.

Definition A.2 ([45]). *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A $d-$dimensional standard Brownian motion $W = (W_t)_{t\geq 0}$ is a stochastic process indexed by $[0, \infty)$ taking values in $\mathbb{R}^d$ such that*

- $W_0 = 0$ *almost surely;*
- W_t *has independent increments, i.e.,* $W_{t_n} - W_{t_{n-1}}, \cdots, W_{t_1} - W_{t_0}$ *are independent for each* $n \geq 1, 0 \leq t_0 < t_1 < t_2 < \cdots < t_n < \infty$*;*
- $W_t - B_s \overset{d}{=} W_{t+h} - W_{s+h}$ *are d-dimensional normal distribution with mean equals to zero and variance matrix equals to* $(t-s)I_d$*, for all* $0 \leq s < t, h \geq s$*;*
- $t \to W_t(\omega), t \geq 0$ *is continuous for all* $\omega \in \Omega$*, i.e.* $\lim_{t\to s} W_t = W_s$ *almost surely.*

Definition A.3 ([45]). *The Poisson process, $N = (N_t)_{t\geq 0}$, with intensity $\lambda > 0$ is a stochastic process with the following properties:*

- $N_0 = 0$ *almost surely;*
- N_t *has independent increments, i.e.,* $N_{t_n} - N_{t_{n-1}}, \cdots, N_{t_1} - N_{t_0}$ *are independent for each* $n \geq 1, 0 \leq t_0 < t_1 < t_2 < \cdots < t_n < \infty$*;*
- N_t *has Poisson increments, i.e.,* $N_t - N_s \overset{d}{=} N_{t+h} - N_{s+h}$ *has a Poisson distribution with parameter* $\lambda(t-s)$*for all* $0 \leq s < t, h \geq s$*;*

Definition A.4 ([2]). *Let $(A, \mathcal{A}, \nu)$ be some measure space with σ-finite measure ν.The Poisson random measure with intensity measure ν is a family of random variables $\{N(t, B), t \geq 0, B \in \mathcal{A}\}$ defined on some filtered probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that*

1. *For any $\omega \in \Omega$, $t > 0$, $N(t, \cdot)(\omega)$ is a counting measure on $(A, \mathcal{A})$.*
2. *For any $t \geq 0, B \in \mathcal{A}$, $N(t, B)$ is a Poisson process with intensity $\nu(B) = \mathbb{E}(N(1, B))$.*

Definition A.5 ([73]). *a. A sequence of random variables $\{X_n\}$ is (uniformly) tight if for every $\epsilon > 0$ there exists a compact set K such that $\mathbb{P}(X_n \in K) \geq 1 - \epsilon$ for every K.*

b. A sequence of probability measures $\{\mu_n\}$ is tight if there exists a compact subset K_ϵ such that $\mu_n(K_\epsilon) \geq 1 - \epsilon$, for each $n \geq 1$.

Consider a stochastic differential equation driven by a Brownian motion and Poisson random measure:

$$dY_t = G(t, Y_t)dt + F(t, Y_t)dW_t + \int_{\mathbb{U}_1} H(t, u)\tilde{N}(dt, du) + \int_{\mathbb{U}\backslash\mathbb{U}_1} K(t, u)N(dt, du), \tag{A.1}$$

where W are d-dimensional Brownian motion, N is a Poisson random measure on $[0, T] \times \mathbb{U}$ and $\tilde{N}(dt, du) = N(dt, du) - dt\nu(du)$ which is called *compensated Poisson random measure.*

Theorem A.1 (Itô formula)**.** *If Y is a Lévy-type stochastic integral of the form (A.1), then for each $f \in C^2(\mathbb{R}^d)$, $t \geq 0$, with probability 1 we have*

$$\begin{aligned} f(Y_t) - f(Y_0) = & \sum_{i=1}^{d} \int_0^t \partial_i f(Y_{s-}) d\left(G(s, Y_s)dt + F(t, Y_s)dW_s\right)^i \\ & + \frac{1}{2} \sum_{i,j=1}^{d} \int_0^t \partial_i \partial_j f(Y_{s-}) F^2(s, Y_{s-}) ds \\ & + \int_0^t \int_{\mathbb{U} \setminus \mathbb{U}_1} \left(f(Y_{s-} + K(s, x)) - f(Y_{s-})\right) N(ds, du) \\ & + \int_0^t \int_{\mathbb{U}_1} \left(f(Y_{s-} + H(s, x)) - f(Y_{s-})\right) \tilde{N}(ds, du) \\ & + \int_0^t \int_{\mathbb{U}_1} \left(f(Y_{s-} + H(s, x)) - f(Y_{s-}) - \sum_{i=1}^{d} \partial_i f(Y_{s-})\right) \nu(du) ds, \end{aligned}$$

where $()^i$ is the ith entry of the vector, $\partial_i f(y) = \frac{\partial f}{\partial y^i}$.

Proof. See Theorem 4.4.7 in [2]. ■

B Inequalities

Lemma B.1 (Hölder's inequality and Cauchy-Schwarz inequality with $p = 2$)**.** *Let $(X, \mathcal{A}, \mu)$ be a σ-finite measure space. If $1 < p, q < \infty$ and $p^{-1} + q^{-1} = 1$, then*

$$\int |fg| d\mu \leq \left(\int |f(x)|^p d\mu \right)^{1/p} \left(\int |g(x)|^p d\mu \right)^{1/p}.$$

This also holds if $p = \infty$ and $q = 1$.

Proof. If $M = \|f\|_\infty = inf\{M : \mu(\{x : |f(x)| \geq M\}) = 0\}$, then $|f| < M$ a.e. and $\int |fg| d\mu \leq M \int |g| d\mu$. The case $p = \infty$ and $q = 1$ follows.

Now assume $1 < p, q < \infty$. If $\int |f(x)|^p d\mu = 0$ then $f = 0$ a.e. and $\int |fg| d\mu = 0$, so the result is clear and similarly if $\int |g(x)|^p d\mu = 0$. Let $K(x) = |f(x)| / \left(\int |f(x)|^p d\mu\right)^{1/p}$ and $G(x) = |g(x)| / \left(\int |g(x)|^p d\mu\right)^{1/p}$. Note that $\int |K(x)|^p d\mu = 1$ and $\int |G(x)|^p d\mu = 1$ and it suffices to show that $\int KGd\mu = 1$. For $0 \leq \lambda \leq 1$, we have

$$e^{\lambda a+(1-\lambda)b} \leq \lambda e^a + (1-\lambda)e^b \tag{B.1}$$

for every pair of real number $a \leq b$. If $K(x), G(x) = 0$, let $a = p\log K(x)$, $b = q\log G(x)$, $\lambda = 1/p$, and $1-\lambda = 1/q$. We then obtain from (B.1) that $K(x)G(x) \leq \frac{K(x)^p}{p} + \frac{G(x)^q}{q}$. Clearly this inequality also holds if $K(x) = 0$ or $G(x) = 0$. Integration of it gives

$$\int KGd\mu \leq \frac{1}{p}\int |K|^p d\mu + \frac{1}{q}\int |K|^q d\mu = \frac{1}{p} + \frac{1}{q} = 1.$$

This complete the proof. ■

Lemma B.2 (Minkowski's inequality)**.** *Let* $(X, \mathcal{A}, \mu)$ *be a* σ*-finite measure space. For* $1 \leq p \leq \infty$,

$$\left(\int |f(x)+g(x)|^p d\mu\right)^{1/p} \leq \left(\int |f(x)|^p d\mu\right)^{1/p} + \left(\int |g(x)|^p d\mu\right)^{1/p}.$$

Proof. Since $|(f+g)(x)| \leq |f(x)| + |g(x)|$, integrating gives the case when $p = 1$. The case $p = \infty$ is also easy. Now suppose $1 < p < \infty$. If $\int |f(x)|^p d\mu$ and $\int |g(x)|^p d\mu$ is infinite, the result is obvious, so we may assume both are finite. The inequality $(a+b)^p \leq 2^{p-1}a^p + 2^{p-1}b^p$, for $a, b \geq 0$ yields, after an integration, $\int |f(x)+g(x)|^p d\mu \leq 2^{p-1}\int |f(x)|^p d\mu + 2^{p-1}\int |g(x)|^p d\mu$. We then have $\int |f(x)+g(x)|^p d\mu < \infty$. Now write

$$|f+g|^p \leq |f||f+g|^{p-1} + |g||f+g|^{p-1}$$

and apply Hölder's inequality with $q = (1-\frac{1}{p})^{-1}$. We obtain

$$\int |f(x)+g(x)|^p d\mu \leq \left(\int |f+g|^{(p-1)q}\right)^{1/q} \left(\left(\int |f(x)|^p d\mu\right)^{1/p} + \left(\int |g(x)|^p d\mu\right)^{1/p}\right).$$

The result follows by $(p-1)q = p$. ■

Lemma B.3 (Gronwall's inequality)**.** *Let $f(t)$ be integrable function on $[0,T]$ which satisfies for each $0 \leq t \leq T$ the integral inequality*

$$f(t) \leq C_1 + C_2 \int_0^t f(s)ds$$

for constants $C_1, C_2 \geq 0$. Then

$$f(t) \leq C_1 e^{C_2 t}$$

for each $0 \leq t \leq T$.

Proof. Note that

$$\begin{aligned} f(t) \leq & C_1 + C_2 \int_0^t \left(C_1 + C_2 \int_0^s f(r)dr \right) ds \\ &= C_1 + C_1C_2t + C_2^2 \int_0^t (t-r)f(r)dr \\ &\leq C_1 + C_1C_2t + C_2^2 \int_0^t (t-r)\left(C_1 + C_2 \int_0^r f(s)ds \right) dr \\ &= C_1 \left(1 + C_2t + \frac{(C_2t)^2}{2} \right) + C_2^3 \int_0^t \frac{(t-s)^2}{2} f(s)ds. \end{aligned}$$

Using induction, we can show that

$$f(t) \leq C_1 \left(1 + C_2t + \frac{(C_2t)^2}{2} + \cdots \frac{(C_2t)^n}{n!} \right) + C_2^{n+1} \int_0^t \frac{(t-s)^n}{n!} f(s)ds.$$

Taking $n \to \infty$, we finish the proof. ■

All the inequalities in the following are defined in a general probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

Lemma B.4 (Doob's martingale inequality)**.** *Suppose $M(t)$ is a martingale (or a positive submartingale) on the interval $[0,\infty)$, then for any $p > 1$ and $t \in [0,\infty)$,*

$$\mathbb{E}\left(\sup_{0\leq s\leq t} |M(s)|^p \right) \leq \left(\frac{p}{p-1} \right)^p \mathbb{E}\left(|M(t)|^p\right) \tag{B.2}$$

Proof. See e.g. Page 14 in [45]. ■

Lemma B.5 (Burkholder-Davis-Gundy(BDG) inequality)**.** *Suppose $M(t)$ is a local martingale on the interval $[0,T]$. There are constants c_p and C_p depending only on p, such that*

$$c_p\mathbb{E}\left([M,M]_T\right)^{p/2} \leq \mathbb{E}\left(\sup_{0\leq t\leq T}|M(t)|\right)^p \leq C_p\mathbb{E}\left([M,M]_T\right)^{p/2} \tag{B.3}$$

for $1 \leq p < \infty$. Here $[M,M]_t$ is the quadratic variation process of M. If moreover, $M(t)$ is continuous, then the result holds also for $0 < p < 1$.

Proof. See e.g. Page 201 of [26], or Page 195 of [64]. ■

Vita

Xiaoyang Pan was born in Guangdong, China in 1990. He attended the Department of Mathematics at South China Normal University in Guangzhou in 2008 and received a Bachelor's degree in statistics in 2012. After that he was recommended to the postgraduate program of Faculty of Science and Technology at the University of Macau, which is located in a nice and international city of China. In August 2013, he came to the Department of Mathematics at the University of Tennessee, Knoxville as a visiting student and earned his Master of Science in mathematics in July 2014. In August 2014 Xiaoyang continued his study in mathematics at the University of Tennessee, Knoxville as a PhD student with concentration in probability and statistics. Xiaoyang will graduate in the summer of 2018 with his PhD in mathematics, along with a Master of Science in statistics. After graduation, he will start his career as a quantitative analyst at BB&T Corporation in North Carolina.