Journal of Stochastic Analysis

Volume 2 | Number 3 Article 15

September 2021

Particle Representation for the Solution of the Filtering Problem. Application to the Error Expansion of Filtering Discretizations

Dan Crisan
Imperial College London, Huxley's Building,180 Queen's Gate, London SW7 2AZ, United Kingdom, dcrisan@imperial.ac.uk

Thomas G. Kurtz
University of Wisconsin-Madison, 48 Lincoln Drive, Madison WI 53706-1388, USA, kurtz@math.wisc.edu

Salvador Ortiz-Latorre
University of Oslo, Niels Henrik Abels hus, Moltke Moes vei 35, 0851 Oslo, Norway, salvadoo@math.uio.no

Follow this and additional works at: https://digitalcommons.lsu.edu/josa

Part of the Analysis Commons, and the Other Mathematics Commons

Recommended Citation
Crisan, Dan; Kurtz, Thomas G.; and Ortiz-Latorre, Salvador (2021) "Particle Representation for the Solution of the Filtering Problem. Application to the Error Expansion of Filtering Discretizations," *Journal of Stochastic Analysis*: Vol. 2 : No. 3 , Article 15.
DOI: 10.31390/josa.2.3.15
Available at: https://digitalcommons.lsu.edu/josa/vol2/iss3/15

Journal of Stochastic Analysis
Vol. 2, No. 3 (2021) Article 15 (34 pages)
DOI: 10.31390/josa.2.3.15

LSU Libraries
digitalcommons.lsu.edu/josa

PARTICLE REPRESENTATION FOR THE SOLUTION OF THE FILTERING PROBLEM. APPLICATION TO THE ERROR EXPANSION OF FILTERING DISCRETIZATIONS

DAN CRISAN, THOMAS G. KURTZ, AND SALVADOR ORTIZ-LATORRE*

Dedicated to the memory of Professor Hiroshi Kunita

Abstract. We introduce a weighted particle representation for the solution of the filtering problem based on a suitably chosen variation of the classical de Finetti theorem. This representation has important theoretical and numerical applications. In this paper, we explore some of its theoretical consequences. The first is to deduce the equations satisfied by the solution of the filtering problem in three different frameworks: the signal independent Brownian measurement noise model, the spatial observations with additive white noise model and the cluster detection model in spatial point processes. Secondly we use the representation to show that a suitably chosen filtering discretisation converges to the filtering solution. Thirdly we study the leading error coefficient for the discretisation. We show that it satisfies a stochastic partial differential equation by exploiting the weighted particle representation for both the approximation and the limiting filtering solution.

1. Introduction

Many phenomena of interest are not completely observable, so it is natural to look for ways of estimating what is not observable about a phenomenon in terms of what is. A natural approach to this problem is to create a mathematical model for the phenomenon that relates what is observable to what is not. Then the model can be used to constrain or estimate the possibilities for the unobserved quantities in terms of the observed quantities. If the mathematical model is stochastic, then a natural way of formulating the solution of this problem is to compute the conditional distribution of what is not known given what is known. If the unknown and known quantities are evolving in time, the problem of computing these conditional distributions as functions of time is referred to as *stochastic filtering*.

Stochastic filtering has a illustrious history that can be traced back to the work of Kolmogorov, Krein and Wiener from the 1940's[1]. Over the last fifty years,

Received 2021-3-29; Accepted 2021-7-25; Communicated by S. Aida, D. Applebaum, Y. Ishikawa, A. Kohatsu-Higa, and N. Privault.

2010 *Mathematics Subject Classification.* 60G35, 60F05, 60F25, 60H35, 60H07, 93E11.

Key words and phrases. Non-linear filtering, Kallianpur-Striebel's formula, first order discretization, particle representation.

* Corresponding author.

[1]The interested reader can consult [1] and [7] for historical accounts of the subject.

Professor Kunita has made fundamental contributions to the study of the stochastic filtering problem. A representative selection of his work on the subject is incorporated in [10, 22, 16, 17, 18, 19, 20, 21], and covers : the stochastic partial differential equations connected with stochastic filtering for the classical model considered here in Section 2.1, the stochastic filtering problem for models in which the signal is a general semi-martingale, the ergodic properties of nonlinear filtering processes, the associated stability and approximation problems in nonlinear filtering theory, the nonlinear filtering Cauchy problem, the asymptotic behavior of the nonlinear filtering errors of Markov processes, the analysis of the innovation process, the long time behavior of the solution of the filtering equations, etc.

The work [10] deserves special consideration. It contains a self-contained, fully rigorous derivation of filtering equations. It is based on an approach that requires the innovation process first considered by Kailath only a year earlier. It also uses the existence of a reference probability measure obtained from the original one by means of a transformation due to Girsanov that was, at the time, barely a decade old. It used a (by now classical) representation of square integrable martingales (appearing in Kunita's earlier work with Watanabe) to remove the assumption of independence between the measurement noise and the signal. The treatment in [10] of the filtering problem was much cleaner than existing contemporary works and allowed, among other things, for the treatment of controlled system processes.

The current work offers an alternative to Kunita's treatment of the filtering equations. In common with [10], we still make use of the reference probability measure (as described below). However, the main tool for the derivation of filtering equations is a certain weighted particle representation for the solution of the filtering problem. Let us describe next the intuition behind this representation:

The simplest version of the filtering problem is one in which the model consists of two random variables, say X and Y, where Y is known to the observer and X is not. Although X is not known, we know the probability distribution of X and, in fact, the joint law of (X, Y). Assume the random variables are defined on a sample space $(\Omega, \mathcal{F}, P)$, with X taking values in a space $\mathcal{U}$ and Y taking values in a space $\mathcal{O}$, which we will always take to be complete, separable metric spaces. In all the examples we consider, these will be function spaces. Typically, we characterize the conditional distribution in terms of the conditional expectations $\mathbb{E}^P[f(X)|Y]$ for a sufficiently large class of functions f.

Central to our analysis is the notion of a *reference probability measure*. If $P << Q$ with $dP = LdQ$, Bayes formula says

$$\mathbb{E}^P[f(X)|Y] = \frac{\mathbb{E}^Q[f(X)L|Y]}{\mathbb{E}^Q[L|Y]}.$$

If $X = h(U, Y)$, $L = L(U, Y)$, and U and Y are independent under Q, then

$$\mathbb{E}^P[f(X)|Y] = \frac{\int f(h(u, Y))L(u, Y)\mu_U(du)}{\int L(u, Y)\mu_U(du)},$$

where μ_U is the distribution of U. If we can find such a Q, that will be our reference probability measure.

These comments suggest a method for simplifying the calculation of conditional distributions: Find a reference probability measure under which what we don't know is independent of what we do know. Then, as in [11], let $U_1, U_2, \ldots$ be iid with distribution μ_U, and we have

$$\mathbb{E}^P\left[f(X)|Y\right] = \lim_{N\to\infty} \frac{\sum_{k=1}^N f(h(U_k,Y))L(U_k,Y)}{\sum_{k=1}^N L(U_k,Y)}.$$

Note that $(U_1,Y),(U_2,Y),\ldots$ is an exchangeable sequence with tail σ-algebra $\mathcal{T} = \sigma(Y)$ (see Corollary 7.25 of [13]), and de Finetti's theorem gives

$$\begin{aligned}\lim_{N\to\infty}\frac{1}{N}\sum_{k=1}^N f(h(U_k,Y))L(U_k,Y) &= \mathbb{E}^Q\left[f(U_1,Y)L(U_1,Y)|\mathcal{T}\right]\\ &= \mathbb{E}^Q\left[f(U_1,Y)L(U_1,Y)|Y\right]\\ &= \int f(h(u,Y))L(u,Y)\mu_U(dx) \qquad a.s.\ Q.\end{aligned}$$

In the context of stochastic processes and the Kallianpur-Striebel formula ([14]), this limit suggests a natural approach to the derivation and representation of filtering equations. We introduce a weighted particle representation for the solution of the filtering problem based on a suitably chosen variation of the classical de Finetti theorem. This representation has important theoretical and numerical applications. In this paper, we explore some of its theoretical consequences.

The first is to deduce the equations satisfied by the solution of the filtering problem in three different frameworks: the signal independent Brownian measurement noise model, the spatial observations with additive white noise model, and the cluster detection model in spatial point processes. We cover this topic in Section 2.

Second, we use the representation to show that a suitably chosen filtering discretization converges to the filtering solution. We cover this topic in Section 3. This discretization is one of three procedures required to develop any numerical method approximating the solution of the filtering problem. See, for example, Chapters 8, 9 and 10 in [1] for concrete examples of numerical schemes for solving the filtering problem.

Third, we study the leading error coefficient for the discretization introduced in Section 3. In Section 4, we show that it satisfies a stochastic partial differential equation by exploiting the weighted particle representation for both the approximation and the limiting filtering solution. Based on these representations, an extension of the classical Richardson extrapolation result can also be obtained. This is the subject of a subsequent work.

Particle representations are flexible tools that can be used for many other stochastic dynamical systems. In this paper, particle representations are used to characterize the solution of the filtering problem by deducing the corresponding filtering equation, and it is also used to show the convergence of a certain discretization of the filtering solution. However, particle representations have many other applications. In [25], they are used to prove uniqueness for a class of stochastic partial differential equations that includes filtering equations. In [4], particle

representations are used to study the solution of a nonlinear stochastic partial differential equation. In particular, the authors show, under mild nondegeneracy conditions on the coefficients, that the solution charges every open set and, under slightly stronger conditions, that the solution is absolutely continuous with respect to Lebesgue measure with strictly positive density. Such results would be very hard to obtain (under the same general assumptions) by other methods such as PDE methods (Sobolev embedding theorems) or Malliavin calculus. Separately, in [5], a similar particle representation is used to study a class of semilinear stochastic partial differential equations with Dirichlet boundary conditions that includes the stochastic Allen-Cahn equation and the Φ^4_d equation of Euclidean quantum field theory. Particle representations arise naturally in the study of McKean-Vlasov type models, for example, [24, 15, 4] where the representations are used to prove limit theorems.

We should emphasize that what we are deriving here are *particle representations* of the filter rather than *particle approximations*.[2] There is a massive area of research regarding *particle approximations* of the distributions of evolving dynamical system, which we shall not discuss here.

2. Derivation of Filtering Equations

In the Introduction, we applied de Finetti's theorem to derive a representation of a conditional expectation in terms of what we called a reference probability measure. In this section, we use this argument to derive stochastic equations giving the solution of the filtering problem in three different settings. The first of these is the familiar observation of a diffusion in Gaussian white noise. The second is similar, but includes a noise process that is common to both the signal and the observation. In addition, the observation process is infinite dimensional. In the third example, the signal and observations are given by spatial point processes.

To avoid certain technicalities, we assume that all σ-algebras are complete and all filtrations are complete and right continuous.

2.1. Observation of a diffusion in Gaussian white noise.

2.1.1. *The model.* The signal is given by an Itô equation in $\mathbb{R}^{d_X}$,

$$X(t) = X(0) + \int_0^t \sigma(X(s))dB(s) + \int_0^t b(X(s))ds, \tag{2.1}$$

for d_B-dimensional standard Brownian motion B, continuous $d_X \times d_B$ matrix-valued σ, and continuous $\mathbb{R}^{d_X}$-valued b, and the observation by

$$Y(t) = \int_0^t h(X(s))ds + W(t), \tag{2.2}$$

where $h : \mathbb{R}^{d_X} \to \mathbb{R}^{d_Y}$ is measurable and W is a $\mathbb{R}^{d_Y}$-valued standard Brownian motion that is independent of B. What is known to the observer is Y and what

[2]The type of weighted particle representations considered here were mentioned briefly in [24]

is not known is X, or assuming uniqueness for (2.1), B. We assume that X does not explode, as would be the case if σ and b have at most linear growth, that is

$$|\sigma(x)| + |b(x)| \leq K_1 + K_2 |x| . \tag{2.3}$$

For simplicity, we assume $\mathbb{E}[|X(0)|^m] < \infty$ for all $m > 0$, and note that under the linear growth assumption, an exercise with Itô's formula shows that for each $m \geq 2$, implies

$$\mathbb{E}[\sup_{0 \leq s \leq t} |X(s)|^m] < D_1^m e^{D_2^m t}, \quad \forall t > 0, \tag{2.4}$$

for appropriate constants D_1, D_2, see Proposition 7.2 in [26].

2.1.2. *The reference probability space.* We take $(\Omega, \mathcal{F}, Q)$ to be a probability space on which are defined independent Brownian motions B and Y, both independent of $X(0)$, with the same dimensions as B and Y above. Let the signal X be defined on $(\Omega, \mathcal{F}, Q)$ as the solution of (2.1), that is, under Q, what is known is independent of what is not known. We note that many presentations of filtering problems begin with $(\Omega, \mathcal{F}, P)$ and obtain Q by change of measure from P. That approach requires P and Q to be equivalent in the sense that $P << Q$ and $Q << P$. In many settings, constructing the model starting with Q is more straightforward and does not require $Q << P$. Let

$$L(t) = \exp\left\{ \int_0^t h^T(X(s))dY(s) - \frac{1}{2}\int_0^t |h|^2(X(s))ds \right\},$$

that is,

$$L(t) = 1 + \int_0^t L(t)h^T(X(s))dY(s),$$

and assume that h satisfies conditions ensuring that L is a martingale. To ensure that L is a martingale one can proceed as follows: since L is a positive local martingale it is a supermartingale and it suffices to show that $\mathbb{E}^Q[L_t] = 1$. Assume that

$$\int_0^t |h|^2(X(s))ds < \infty, \quad Q\text{-a.s.} \tag{2.5}$$

Under Q, Y is a Brownian motion independent of X and the law of

$$\int_0^t h^T(X(s))dY(s),$$

given $\mathcal{F}_t^X := \sigma(X_s : s \leq t)$, is that of a Gaussian random variable with mean zero and variance $\int_0^t |h|^2(X(s))ds$. Therefore, we can write

$$\mathbb{E}^Q\left[\exp\left\{ \int_0^t h^T(X(s))dY(s) \right\} \middle| \mathcal{F}_t^X \right] = \exp\left\{ \frac{1}{2}\int_0^t |h|^2(X(s))ds \right\},$$

which yields $\mathbb{E}^Q\left[L_t | \mathcal{F}_t^X\right] \equiv 1$ and, therefore, $\mathbb{E}^Q[L_t] = 1$. For example, (2.5) is trivially satisfied if h is bounded. Then defining $dP_{|\mathcal{F}_t} = L(t)dQ_{|\mathcal{F}_t}$, under P, by Theorem A.8, B and W given by

$$W(t) = Y(t) - \int_0^t h(X(s))ds$$

are independent standard Brownian motions. Consequently, under P, X and Y have the joint distributions of (2.1) and (2.2).

2.1.3. *Filtering equations.* Let $\{\mathcal{F}_t^Y\}$ be the (completed) filtration generated by the observations Y. Then, assuming $\mathbb{E}^Q[|\varphi(X(t))L(t)|] < \infty$, we have the Kallianpur-Striebel formula [14]

$$\mathbb{E}^P\left[\varphi(X(t))|\mathcal{F}_t^Y\right] = \frac{\mathbb{E}^Q\left[\varphi(X(t))L(t)|\mathcal{F}_t^Y\right]}{\mathbb{E}^Q\left[L(t)|\mathcal{F}_t^Y\right]}.$$

Let $X_1, X_2, \ldots$ be iid copies of X that are independent of Y under Q, and let

$$L_k(t) = 1 + \int_0^t L_k(s)h(X_k(s))^T dY(s).$$

We define the *unnormalized conditional distribution* ρ by

$$\rho_s(\varphi) \equiv \mathbb{E}^Q\left[\varphi(X(s))L(s)|\mathcal{F}_s^Y\right] = \mathbb{E}^Q\left[\varphi(X_k(s))L_k(s)|\mathcal{F}_s^Y\right],$$

and the exchangeability of $\{(X_k, Y)\}$ ensures

$$\lim_{N\to\infty} \frac{1}{N}\sum_{k=1}^N \varphi(X_k(s))L_k(s) = \mathbb{E}^Q[\varphi(X(s))L(s)|\mathcal{F}_s^Y]. \tag{2.6}$$

With reference to Appendix A.1, since $\{(X_k, L_k)\}$ is exchangeable, by de Finetti's theorem, the sequence determines a random probability measure, which we will call the de Finetti measure,

$$\Xi = \lim_{N\to\infty} \frac{1}{N}\sum_{k=1}^N \delta_{(X_k, L_k)} \quad a.s.$$

on $C_{\mathbb{R}^{d_X}\times[0,\infty)}[0,\infty)$ and a probability measure-valued process

$$V(\cdot) = \lim_{N\to\infty} \frac{1}{N}\sum_{k=1}^N \delta_{(X_k(\cdot), L_k(\cdot))} \quad \text{in probability,}$$

in $C_{\mathcal{P}(\mathbb{R}^{d_X}\times[0,\infty))}[0,\infty)$. Of course, the unnormalized conditional distribution is

$$\rho_t(C) = \int_{\mathbb{R}^{d_X}\times[0,\infty)} a\mathbf{1}_C(x)V(dx\times da, t).$$

Lemma 2.1. *Assume $|h|$ is bounded. Then as with (2.4), for each $m \geq 2$, there exists D_3 such that*

$$\mathbb{E}^Q[\sup_{0\leq s\leq t} L(s)^m] \leq e^{D_3 t}. \tag{2.7}$$

We assume (2.4) and (2.7) for all $m > 0$. Then, for $\varphi \in C^2(\mathbb{R}^{d_X})$ satisfying $|\varphi(x)| \leq C_1 + C_2|x|^m$,

$$\varphi(X(t)) = \varphi(X(0)) + \int_0^t \nabla\varphi^T(X(s))\sigma(X(s))dB(s) + \int_0^t A\varphi(X(s))ds,$$

where for $a(x) = \sigma(x)\sigma^T(x)$

$$A\varphi(x) = \frac{1}{2}\sum_{i,j} a_{ij}(x)\partial_i\partial_j\varphi(x) + b(x)\cdot\nabla\varphi(x),$$

and

$$\begin{aligned}\varphi(X(t))L(t) &= \varphi(X(0)) + \int_0^t L(s)d\varphi(X(s))ds + \int_0^t \varphi(X(s))dL(s)\\ &= \varphi(X(0)) + \int_0^t L(s)\nabla\varphi(X(s))^T\sigma(X(s))dB(s)\\ &\quad + \int_0^t L(s)A\varphi(X(s))ds + \int_0^t \varphi(X(s))L(s)h(X(s))^T dY(s).\end{aligned}$$

Then for $\{(X_k, L_k)\}$ given above,

$$\begin{aligned}\varphi(X_k(t))L_k(t) &= \varphi(X_k(0)) + \int_0^t L_k(s)\nabla\varphi(X_k(s))^T\sigma(X_k(s))dB_k(s)\\ &\quad + \int_0^t L_k(s)A\varphi(X_k(s))ds\\ &\quad + \int_0^t \varphi(X_k(s))L_k(s)h(X_k(s))^T dY(s),\end{aligned} \tag{2.8}$$

where $\{B_k\}$ is the sequence of independent standard Brownian motions associated to the sequence of signals $\{X_k\}$. We claim that we can average both sides as in (2.6) and obtain the following:

Theorem 2.2. *For the model in Section 2.1, the unnormalized conditional distribution ρ satisfies*

$$\rho_t(\varphi) = \rho_0(\varphi) + \int_0^t \rho_s(A\varphi)ds + \int_0^t \rho_s(\varphi h)dY(s), \tag{2.9}$$

the Zakai equation, and by Itô's formula, we have the Kushner-Stratonovich equation.

$$\begin{aligned}\pi_t\varphi &= \mathbb{E}^P[\varphi(X(t))|\mathcal{F}_t^Y] = \frac{\rho_t(\varphi)}{\rho_t(1)}\\ &= \frac{\rho_0(\varphi)}{\rho_0(1)} + \int_0^t \frac{1}{\rho_s(1)}d\rho_s(\varphi) - \int_0^t \frac{\rho_s(\varphi)}{\rho_s(1)^2}d\rho_s(1)\\ &\quad + \int_0^t \frac{\rho_s(\varphi)}{\rho_s(1)^3}d[\rho_\cdot(1)]_s - \int_0^t \frac{1}{\rho_s(1)^2}d[\rho_\cdot(\varphi), \rho_\cdot(1)]_s\\ &= \pi_0\varphi + \int_0^t \pi_s A\varphi ds + \int_0^t (\pi_s\varphi h - \pi_s\varphi\pi_s h)dY(s)\\ &\quad + \int_0^t \sigma^2\pi_s\varphi\pi_s|h|^2 ds - \int_0^t \sigma^2\pi_s\varphi h\pi_s h ds\\ &= \pi_0\varphi + \int_0^t \pi_s A\varphi ds + \int_0^t (\pi_s\varphi h - \pi_s\varphi\pi_s h)(dY(s) - \pi_s h ds).\end{aligned}$$

Proof. The term on the left and the first term on the right of (2.8) average as in (2.6). The average over $1 \le k \le N$ of the second term on the right is a continuous,

mean zero martingale with quadratic variation

$$\frac{1}{N^2}\sum_{k=1}^{N}\int_0^t L_k(s)^2\nabla\varphi(X_k(s))^T\sigma(X_k(s))\sigma(X_k(s))^T\nabla\varphi(X_k(s))ds,$$

which, under the growth and moment conditions above, converges to zero implying the average converges to zero by Doob's inequality. The averages of the integrands in the last two terms converge to the integrands in the last two terms of (2.9) by Lemma A.6, so the next to the last terms converges by elementary calculus and the last term converges by the stochastic integral convergence result, Theorem 2.2 in [23]. □

In the terminology of [25], the infinite sequence $\{(X_k, L_k)\}$ gives a *particle representation* of the Zakai equation. The point here is not just that $\{(X_k, L_k)\}$ gives a derivation of the Zakai equation. The representation can, for example, be used to prove uniqueness (see [25]) and derive approximations (see Section 4).

2.2. Spatial observations with additive white noise.

2.2.1. *The model.* The basic outline of the argument above works in many different situations. We again take the signal to be a diffusion in $\mathbb{R}^{d_X}$, but now we assume that the stochastic inputs include both a d_B-dimensional standard Brownian motion B and a space-time Gaussian white noise W. In particular,

$$X(t) = X(0)+\int_0^t \sigma(X(s))dB(s)+\int_0^t b(X(s))ds+\int_{S_0\times[0,t]}\alpha(X(s),u)W(du\times ds), \tag{2.10}$$

where $\mathbb{E}[W(C,t)] = 0$ and

$$\mathbb{E}[W(C,t)W(D,s)] = \mu_0(C\cap D)t\wedge s,$$

$t, s \geq 0$ and $C, D \in \mathcal{B}(S_0)$, the Borel sets for some complete, separable metric space, S_0.

We assume that the observations are given by $Y(C,t)$, $t \geq 0$ and $C \in \mathcal{B}(S_0)$, where

$$Y(C,t) = \int_0^t\int_C h(X(s),u)\mu_0(du)ds + W(C,t). \tag{2.11}$$

Consequently, σ is a $d_X \times d_B$-dimensional matrix-valued function, b and α are $\mathbb{R}^{d_X}$-valued, and h is $\mathbb{R}$-valued. For simplicity, assume σ, b, α, and h are bounded and continuous and that μ_0 is a finite measure.

Then the generator for X is

$$A\varphi(x) = \frac{1}{2}\sum a_{ij}(x)\partial_i\partial_j\varphi(x) + \sum b_i(x)\partial_i\varphi(x), \quad \varphi\in C_c^2\left(\mathbb{R}^{d_X}\right),$$

where

$$a(x) = \sigma(x)\sigma(x)^T + \int_{S_0}\alpha(x,u)\alpha(x,u)^T\mu_0(du).$$

We can write

$$X(t) = X(0) + \int_0^t \sigma(X(s))dB(s) + \int_{S_0\times[0,t]}\alpha(X(s),u)Y(du\times ds) \tag{2.12}$$

$$+\int_0^t (b(X(s)) - \int_{S_0} \alpha(X(s),u)h(X(s),u)\mu_0(du))ds,$$

so Y is what we know and B is what we don't know.

2.2.2. *The reference probability space.* Consequently, we assume B and Y are defined on a measurable space $(\Omega, \mathcal{F})$, and there is a probability distribution Q on $\mathcal{F}$ such that under Q, Y is Gaussian white noise on $S_0 \times [0,\infty)$ with $\mathbb{E}[Y(C,t)] = 0$, $t \geq 0$, $C \in \mathcal{B}(S_0)$ and

$$\mathbb{E}[Y(C,t)Y(D,s)] = \mu_0(C \cap D)t \wedge s,$$

and B is a standard d_B dimensional Brownian motion independent of Y. Both are independent of $X(0)$.

Then take $dP_{|\mathcal{F}_t} = L(t)dQ_{|\mathcal{F}_t}$ where

$$L(t) = 1 + \int_{S_0 \times [0,t]} L(s)h(X(s),u)Y(du \times ds),$$

and under P, (X,Y) has the joint distribution of the original model.

2.2.3. *Filtering equations.* Under Q, X is a diffusion with generator

$$A^Q\varphi(x) = \frac{1}{2}\sum a_{ij}(x)\partial_i\partial_j\varphi(x) + \sum c_i(x)\partial_i\varphi(x),$$

where

$$c_i(x) = b_i(x) - \int_{S_0} \alpha(x,u)h(x,u)\mu_0(du),$$

and under P, X is a diffusion with the original generator

$$A\varphi(x) = \frac{1}{2}\sum a_{ij}(x)\partial_i\partial_j\varphi(x) + \sum b_i(x)\partial_i\varphi(x),$$

that is, X is the signal of the original model.

Then

$$\begin{aligned}
&\varphi(X(t))L(t)\\
&= \varphi(X(0)) + \int_0^t L(s)\nabla\varphi(X(s))^T\sigma(X(s))dB(s)\\
&\quad + \int_{S_0\times[0,t]} L(s)(\nabla\varphi(X(s))\cdot\alpha(X(s),u))Y(du\times ds)\\
&\quad + \int_0^t L(s)A^Q\varphi(X(s))ds + \int_{S_0\times[0,t]} L(s)\varphi(X(s))h(X(s),u))Y(du\times ds)\\
&\quad + \int_0^t L(s)\nabla\varphi(X(s))^T \int_{S_0} \alpha(X(s),u)h(X(s),u)\mu_0(du)ds.
\end{aligned}$$

To obtain the particle representation, we let B_k be independent, standard Brownian motions, independent of Y on $(\Omega, \mathcal{F}, Q)$. Let

$$X_k(t) = X_k(0) + \int_0^t \sigma(X_k(s))dB_k(s) + \int_{S_0\times[0,t]} \alpha(X_k(s),u)Y(du\times ds)$$

$$+\int_0^t \left(b(X_k(s)) - \int_{S_0} \alpha(X_k(s),u)h(X_k(s),u)\mu_0(du)\right) ds$$

$$L_k(t) = 1 + \int_{S_0\times[0,t]} L_k(s)h(X_k(s),u)Y(du\times ds).$$

Then, as before,

$$\rho_t(\varphi) = \mathbb{E}^P[\varphi(X(t))L(t)|\mathcal{F}_t^Y] = \lim_{N\to\infty}\frac{1}{N}\sum_{k=1}^N \varphi(X_k(t))L_k(t).$$

Since

$$\begin{aligned}
&\varphi(X_k(t))L_k(t)\\
&= \varphi(X_k(0)) + \int_0^t L_k(s)\nabla\varphi(X_k(s))^T\sigma(X_k(s))dB_k(s)\\
&\quad + \int_{S_0\times[0,t]} L_k(s)(\nabla\varphi(X_k(s))\cdot\alpha(X_k(s),u)Y(du\times ds)\\
&\quad + \int_0^t L_k(s)A\varphi(X_k(s))ds + \varphi(X_k(s))h(X_k(s),u))Y(du\times ds),
\end{aligned}$$

under appropriate moment conditions and applying convergence results of [24] to the Y integral instead of [23], averaging gives

Theorem 2.3.

$$\begin{aligned}
\rho_t(\varphi) &= \rho_0(\varphi) + \int_0^t \rho_s(A\varphi)ds\\
&\quad + \int_{S_0\times[0,t]} \rho_s(\nabla\varphi\cdot\alpha(\cdot,u) + \varphi h(\cdot,u))Y(du\times ds),
\end{aligned}$$

determines the unnormalized conditional distribution and the corresponding Kushner–Stratonovich equation is

$$\begin{aligned}
\pi_t\varphi &= \frac{\rho_t(\varphi)}{\rho_t(1)}\\
&= \pi_0\varphi + \int_0^t \pi_s A\varphi ds\\
&\quad + \int_{S_0\times[0,t]} \Big(\pi_s(\nabla\varphi\cdot\alpha(\cdot,u) + \varphi h(\cdot,u)) - \pi_s\varphi\pi_s h(\cdot,u)\Big)Y(du\times ds)\\
&\quad + \int_0^t\int_{S_0} \Big(\pi_s\varphi\pi_s h(\cdot,u) - \pi_s(\nabla\varphi\cdot\alpha(\cdot,u) + \varphi h(\cdot,u))\Big)\pi_s h(\cdot,u)\mu_0(du)ds\\
&= \pi_0\varphi + \int_0^t \pi_s A\varphi ds\\
&\quad \int_{S_0\times[0,t]} \Big(\pi_s(\nabla\varphi\cdot\alpha(\cdot,u) + \varphi h(\cdot,u)) - \pi_s\varphi\pi_s h(\cdot,u)\Big)\tilde{Y}(du\times ds)
\end{aligned}$$

where

$$\tilde{Y}(C,t) = Y(C,t) - \int_0^t\int_C \pi_s h(\cdot,u)\mu_0(du)ds.$$

2.3. Cluster detection in spatial point processes.

2.3.1. *The model.* The following example is a simplified version of the models considered in [29, 28]. Natural settings in which this problem might arise include internet packets that form a malicious attack on a computer system, financial transactions that form a collusive trading scheme, and, the example considered in [29], earthquakes that form a single seismic event.

Let E be a measurable space and $\mathcal{C}(E)$ be the collection of counting measures on E and $\mathcal{C}(E\times[0,\infty))$ the collection of counting measures on $E\times[0,\infty)$. The observations form a marked point process O with marks in E, that is for $t\geq 0$, $O(\cdot,t)\in\mathcal{C}(E)$, that include the cluster C (the signal) and the noise N

$$O(A,t)=N(A,t)+C(A,t),\quad A\in\mathcal{B}(E),t\geq 0.$$

For simplicity, we will assume the $O(E,t)$ is finite for all t.

For ξ_1 and ξ_2 independent Poisson random measures on $E\times[0,\infty)\times[0,\infty)$ with mean measure $\nu\times\ell\times\ell$, ℓ denoting Lebesgue measure, γ a nonnegative function on E, and $\lambda: E\times\mathcal{C}(E\times[0,\infty))\to[0,\infty)$, N and C can be written as solutions of

$$N(A,t)=\int_{A\times[0,\infty)\times[0,t]}\mathbf{1}_{[0,\gamma(u)]}(v)\xi_1(du\times dv\times ds) \tag{2.13}$$
$$C(A,t)=\int_{A\times[0,\infty)\times[0,t]}\mathbf{1}_{[0,\lambda(u,\eta_{s-})]}(v)\xi_2(du\times dv\times ds),$$

where η is given by

$$\eta_t(A\times[0,r])=\int_{A\times[0,t]}\mathbf{1}_A(u)\mathbf{1}_{[0,r]}(s)C(du\times ds),\quad A\in\mathcal{B}(E),r\in[0,t],$$

that is, η_t is the collection of points in the cluster up to time t. The noise, N, is a space-time Poisson process. We assume there exists λ_0 such that $\lambda(u,\eta)\leq\lambda_0(u)$ for all η and that $\int_E\gamma(u)\nu(du)<\infty$ and $\int_E\lambda_0(u)\nu(du)<\infty$. Assuming $N(E,0)=C(E,0)=0$, these assumptions assure that $N(E,t)$ is Poisson distributed with mean $t\int_E\gamma(u)\nu(du)$ and that $C(E,\cdot)$ is dominated by a Poisson process. Consequently, existence and uniqueness follow by construction from one jump to the next.

Of course, if E is a finite set, this model is essentially a filtering model for counting processes as studied by Bremaud [2].

2.3.2. *The reference probability space.* On $(\Omega,\mathcal{F},Q)$, let N and C be independent Poisson random measures with mean measures $\nu_0(du\times ds)=\gamma(u)\nu(du)ds$ and $\nu_1(du\times ds)=\lambda_0(u)\nu(du)ds$ respectively. At each point (u,t) in $O=N+C$, let $\theta(u,t)=1$, if $(u,t)\in C$ and $\theta(u,t)=0$ otherwise. Then $Q\{\theta(u,t)=1|O\}=\frac{\lambda_0(u)}{\lambda_0(u)+\gamma(u)}$, and hence, under Q, what is known, O, is independent of what is not known, $\{\theta(u,t):(u,t)\in O\}$. Note also that the $\theta(u,t)$ are independent of each other, and η and θ are related by

$$\eta_t(A\times[0,r])=\sum_{(u,s)\in O(A\times[0,r])}\theta(u,s),\quad A\in\mathcal{B}(E),r\in[0,t]. \tag{2.14}$$

However, under Q, O is not independent of the signal C. This is in contrast to the models in sections 2.1 and 2.2 Under Q,

$$\tilde{C}(A,t) = C(A,t) - \int_0^t \int_A \lambda_0(u)\nu(du)ds, \quad A \in \mathcal{B}(E), t \geq 0, \tag{2.15}$$

is a martingale random measure. In particular, for each $A \in \mathcal{B}(E)$, (2.15) is a $\{\mathcal{F}_t\}$-martingale for $\mathcal{F}_t = \sigma(C(A,s), N(A,s) : s \leq t, A \in \mathcal{B}(E))$.

Let L satisfy

$$\begin{aligned} L(t) &= 1 + \int_{E\times[0,t]} (\frac{\lambda(u,\eta_{s-})}{\lambda_0(u)} - 1)L(s-)(C(du\times ds) - \lambda_0(u)\nu(du)ds) \\ &= 1 + \int_{E\times[0,t]} (\frac{\lambda(u,\eta_{s-})}{\lambda_0(u)} - 1)L(s-)\theta(u,t)O(du\times ds) \\ &\quad - \int_{E\times[0,t]} (\lambda(u,\eta_{s-}) - \lambda_0(u))L(s)\nu(du)ds. \end{aligned}$$

At each point $(u,s) \in C$

$$L(s) = \frac{\lambda(u,\eta_{s-})}{\lambda_0(u)} L(s-),$$

so L is a nonnegative local martingale, and setting $\tau_n = \inf\{t : L(t) \geq n\}$, $L_n(t) = L(t \wedge \tau_n)$ is a martingale. Recall that for the original cluster detection model, O is dominated by a Poisson process with mean measure $dt \times (\gamma(u) + \lambda_0(u))\nu(du)$, and hence O has finitely many jumps per unit time. Consequently, under P, $\sup_{t\leq T} L(t) < \infty$ for all T, and $P\{\lim_{n\to\infty} \tau_n = \infty\} = 1$. Hence, by Lemma 3.3 of [12], pp.166, $P_{|\mathcal{F}_t} << Q_{|\mathcal{F}_t}$ with $\frac{dP}{dQ} = L(t)$ on $\mathcal{F}_t$. Under P, for all A, by the results in Section A.2 below,

$$C(A,t) - \int_{A\times[0,t]} \lambda(u,\eta_s)\nu(du)ds \tag{2.16}$$

is a local martingale, N is independent of C, and is a Poisson random measure with mean measure ν_0, that is, under P, (N,C) has the distribution of the solution of (2.13).

2.3.3. *Filtering equations.* Let $\theta^k(u,s)$ be independent realizations of $\theta(u,s)$ and η^k its associated counting process given by (2.14). Observing that

$$\begin{aligned} &\varphi(\eta_t^k)L_k(t) \\ &= \varphi(\eta_0^k) + \int_{E\times[0,t]} \left(\varphi(\eta_{s-}^k + \delta_{(u,s)})\frac{\lambda(u,\eta_{s-}^k)}{\lambda(u)} - \varphi(\eta_{s-}^k)\right) L_k(s-)C_k(du\times ds) \\ &\quad - \int_{E\times[0,t]} \varphi(\eta_s^k)\left(\lambda(u,\eta_s^k) - \lambda(u)\right) L_k(s)\nu(du)ds \\ &= \varphi(\eta_0^k) \\ &\quad + \int_{E\times[0,t]} \left(\varphi(\eta_{s-}^k + \delta_{(u,s)})\frac{\lambda(u,\eta_{s-}^k)}{\lambda(u)} - \varphi(\eta_{s-}^k)\right) L_k(s-)\theta^k(u,s)O(du\times ds) \\ &\quad - \int_{E\times[0,t]} \varphi(\eta_s^k)\left(\lambda(u,\eta_s^k) - \lambda(u)\right) L_k(s)\nu(du)ds, \end{aligned}$$

and after averaging over k, we deduce

Theorem 2.4. *The unnormalized conditional distributions satisfies*

$$\rho_t(\varphi) = \rho_0(\varphi) - \int_{E\times[0,t]} \rho_s\left(\varphi(\cdot)\left(\lambda(u,\cdot) - \lambda(u)\right)\right)\nu(du)ds$$
$$+ \int_{E\times[0,t]} \rho_{s-}(\varphi\left(\cdot + \delta_{(u,s)}\right)\frac{\lambda(u,\cdot)}{\lambda(u)} - \varphi(\cdot),)\frac{\lambda(u)}{\lambda(u)+\gamma(u)}O(du\times ds)$$

and

$$\pi_t\varphi = \pi_0\varphi$$
$$+ \int_{E\times[0,t]} \frac{\pi_{s-}\left(\varphi\left(\cdot + \delta_{(u,s)}\right)\lambda(u,\cdot)\right) - \pi_{s-}\lambda(u,\cdot)\pi_{s-}\varphi}{\pi_{s-}\lambda(u,\cdot)+\gamma(u)}O(du\times ds)$$
$$- \int_{E\times[0,t]} \left(\pi_s\left(\varphi(\cdot)\lambda(u,\cdot)\right) - \pi_s\varphi\pi_s\lambda(u,\cdot)\right)\nu(du)ds$$

Remark 2.5. In most settings, the difficulty of computing the distribution of $2^{O(E,t)}$ possible states would be prohibitive. The compromise in [29] is to structure the model in such a way that it is possible to compute $\pi_t\varphi = \mathbb{E}^P\left[\varphi(\eta_s)|\mathcal{F}_s\right]$ for a "small" collection of φ.

Suppose one observes u_i at time τ_i and $y_i = (u_i, \tau_i)$. Let

$$\theta(y_i)(\cdot) = \mathbf{1}_{\{y_i \text{ is a point in the cluster}\}}$$

and

$$\theta_0(y_i)(\cdot) = \mathbf{1}_{\{y_i \text{ is the latest point in the cluster}\}}.$$

One needs to be able to evaluate

$$\pi_t\lambda(u,\cdot)$$

which is accomplished under a Markov scenario.

Consider

$$\lambda(u,\eta_t) = \sum_{i=1}^{O(E,t)} \lambda(u,y_i)\theta_0(y_i) + \epsilon(u),$$

Then the goal is to obtain a closed system for

$$\pi_t\theta_0(y_i), \quad \pi_t\theta(y_i), \quad \pi_t\theta(y_i)\theta_0(y_j).$$

See Section 1.3 of [28].

3. Convergence of First Order Discretizations

The solution of the stochastic filtering problem depends on both the signal and on the observation process. However the manner in which it depends on both ingredients is different. Let us take, as an example, the framework described in Section 2.1. By Kallianpur-Striebel's formula, we have that

$$\pi_t(\varphi) = \mathbb{E}^P\left[\varphi(X(t))|\mathcal{F}_t^Y\right] = \frac{\mathbb{E}^Q\left[\varphi(X(t))L(t)|\mathcal{F}_t^Y\right]}{\mathbb{E}^Q\left[L(t)|\mathcal{F}_t^Y\right]},$$

where Q is the reference measure. Under Q, X and Y are independent and

$$L(t) = \exp\left\{\int_0^t h^T(X(s))dY(s) - \frac{1}{2}\int_0^t |h|^2(X(s))ds\right\}. \tag{3.1}$$

Because of the conditioning with respect to the observation σ-algebra, the observation path can be assumed to be fixed to a particular realisation (the one that is actually observed). Although this fact is not immediately clear from the expression appearing in (3.1), a simple integration by parts of the stochastic integral in (3.1) can justify the observation path dependence of L_t, see e.g. [3, 8] for further details. The signal process enters into the solution of the filtering problem through its law. The pathwise behavior of the signal plays no role; it is just its law that is needed to compute $\pi_t(\varphi)$. Any numerical resolution of the solution of the filtering problem involves a discretization of the observation path as well as the approximation of the law of the signal. The error in the numerical approximations of $\pi_t(\varphi)$ will then depend on the observation path discretization error as well as the error due to the approximation of the signal. In the following, we will analyze the error due to the observation path discretization by exploiting the particle representation of the various quantities involved. In this section we show, under very general conditions, that the discretization error tends to 0 as the time discretization mesh converges to 0, and in the next section we compute the order of convergence as well as the leading error coefficient.

Typically, the observation data is recorded at discrete times, and only these data are made available and used. For example, if the set of data $\{Y_{\frac{k}{n}}, k \geq 0\}$ is available, we can use the approximation

$$\pi_t^n(\varphi) = \frac{\mathbb{E}^Q\left[\varphi(X(t))L^n(t)|\mathcal{F}_t^Y\right]}{\mathbb{E}^Q\left[L^n(t)|\mathcal{F}_t^Y\right]}, \tag{3.2}$$

where for $t = k/n$,

$$\begin{aligned} &L^n(k/n) \\ &= \exp\left(\sum_{i=0}^{k-1}\left(h^T\left(X(i/n)\right)\left(Y\left((i+1)/n\right) - Y\left(i/n\right)\right) - \frac{1}{2n}|h|^2((X(i/n))\right)\right). \end{aligned} \tag{3.3}$$

More generally, we can embed the above approximation into the usual continuous time version. Let $\tau_n(s) = j\frac{1}{n}$ for $s \in \left[j\frac{1}{n}, (j+1)\frac{1}{n}\right)$, and let $L^n(t)$ be given by

$$L^n(t) = \exp\left(\int_0^t h^T(X(\tau_n(s)))dY(s) - \frac{1}{2}\int_0^t |h|^2(X(\tau_n(s)))ds\right) \tag{3.4}$$

and then re-write (3.2) as

$$\pi_t^n(\varphi) = \frac{\rho_t^n(\varphi)}{\rho_t^n(1)}, \tag{3.5}$$

where ρ_t^n is the Picard approximation with time step $1/n$ for the unnormalized conditional distribution ρ

$$\rho_t^n(\varphi) \triangleq E^Q\left[\varphi(X(t))L^n(t)|\mathcal{F}_t^Y\right].$$

Proposition 3.1. *For all $\varphi \in C_P\left(\mathbb{R}^{d_X}\right)$, we have that*

$$\lim_{n\to\infty} \rho_t^n(\varphi) = \rho_t(\varphi), \quad \text{and} \quad \lim_{n\to\infty} \pi_t^n(\varphi) = \pi_t(\varphi)$$

in probability. Moreover, $\lim_{n\to\infty} \rho^n = \rho$ and $\lim_{n\to\infty} \pi^n = \pi$ in probability, where the convergence is in the space of continuous measure-valued paths.

Proof. As announced, we use the particle representation for the respective measure valued processes ρ^n, ρ, π^n and π. Let $X_1, X_2, \ldots$ be i.i.d. copies of X that are independent of Y under Q, and let L_k^n be the exponentials corresponding to $X_1, X_2, \ldots$ defined as above, i.e.,

$$\begin{aligned} L_k^n(t) &= \exp\left(\int_0^t h^T(X_k(\tau_n(s)))dY(s) - \frac{1}{2}\int_0^t |h|^2(X_k(\tau_n(s)))ds\right) \\ L_k(t) &= \exp\left(\int_0^t h^T(X_k(s))dY(s) - \frac{1}{2}\int_0^t |h|^2(X_k(s))ds\right). \end{aligned}$$

Then, the particle representations give

$$\rho_t^n(\varphi) = \lim_{N\to\infty} \frac{1}{N}\sum_{k=1}^N \varphi(X_k(t))\, L_k^n(t)$$

and

$$\rho_t(\varphi) = \lim_{N\to\infty} \frac{1}{N}\sum_{k=1}^N \varphi(X_k(t))\, L_k(t),$$

where φ is a continuous function with at most polynomial growth, and convergence is assured by the exchangeability of $\{(L_k^n, X_k)\}$ and $\{(L_k^n, X_k)\}$.

We apply Lemma A.1 with $Z^n = ((X_1, L_1^n), (X_2, L_2^n), \ldots)$. Note that, in this example, $N_n = \infty$. Thanks to Remark A.2 it suffices to check the convergence of Z^n to Z in $C_{\left(\mathbb{R}^{d_X}\times[0,\infty)\right)^\infty}[0,\infty)$ in probability. For fixed k, one has that

$$\mathbb{E}^Q\left[\sup_{0\le s\le t} |L_k^n(s) - L_k(s)|^2\right] \to_{n\to\infty} 0.$$

for all $t \ge 0$. This follows from the inequality $|e^y - e^x| \le \frac{(e^x+e^y)}{2}|x-y|$ combined with similar estimates as those in Lemmas 3.6 and 3.9 in [6]. Therefore, (X_k, L_k^n) converges in probability to (X_k, L_k) when n tends to infinity. Lemma A.1 b) yields that

$$\lim_{n\to\infty} V^n = \lim_{n\to\infty}\lim_{N\to\infty} \frac{1}{N}\sum_{k=1}^N \delta_{(X_k, L_k^n)} = \lim_{N\to\infty} \frac{1}{N}\sum_{k=1}^N \delta_{(X_k, L_k)} = V,$$

in $C_{\mathcal{P}\left(\mathbb{R}^{d_X}\times[0,\infty)\right)}[0,\infty)$.

The moment estimates on X_k, L_k^n, and L_k ensure that for each $\psi \in C_P\left(\mathbb{R}^{d_X}\right)$ and $m > 0$ and each $T > 0$,

$$\begin{aligned} &\sup_n \mathbb{E}[\sup_{t\le T} \psi(X_k(t))(L_k^n(t)^m + 1)] \\ &= \sup_n \mathbb{E}[\sup_{t\le T} \int_{\mathbb{R}^{d_X}\times[0,\infty)} \psi(x)(a^m+1)V_n(dx\times da, t) < \infty. \end{aligned}$$

Then keeping in mind that

$$\rho_t^n(\varphi) = \int_{\mathbb{R}^{d_X}\times[0,\infty)} a\varphi(x)V^n(dx\times da,t)$$

and

$$\rho_t(\varphi) = \int_{\mathbb{R}^{d_X}\times[0,\infty)} a\varphi(x)V(dx\times da,t),$$

the result follows by applying Lemma A.5. In particular, for all $\varphi \in C_P\left(\mathbb{R}^{d_X}\right)$, we have that

$$\lim_{n\to\infty} \rho_t^n(\varphi) = \rho_t(\varphi)$$

in probability, with the convergence of the measure valued processes

$$\lim_{n\to\infty} \rho^n = \rho, \qquad \text{and} \qquad \lim_{n\to\infty} \pi^n = \pi,$$

being an immediate consequence of the above and of the Kallianpur-Striebel formula. □

A similar result can be obtained for the second framework (Spatial observations with additive white noise). The above convergence result does not give an estimate of the order of convergence. This is not possible under the general assumptions of stated in section 2.1.2 on the functions h, b and σ. However, we can do this under more restrictive assumptions. This is the goal of the next section.

4. Leading Error Coefficient for the Picard Discretization

In this section, we are using the same framework as in sections 2.1 and 3, as well as the same notation introduced therein. In addition, to prove the relative compactness result of Theorem 4.4 below, we will require that $\sigma, b, h \in C_b^6\left(\mathbb{R}^{d_X}\right)$. Heuristically, the main goal of this section is to show that

$$\rho_t^n = \rho_t - \frac{1}{n}U_t + o\left(\frac{1}{n}\right), \qquad t\in[0,\infty),$$

where U is a process characterized as a solution of a certain stochastic evolution equation.[3] The exact statement of the result is contained in Theorem 4.4 below. The main technical tool to do this is, again, the particle representations of the various processes involved. To be more precise, let C_P^k be the space of k-differentiable functions with at most polynomial growth. Then, for $\varphi \in C_P^k$ consider the quantities

$$E_t^{k,n}(\varphi) \triangleq n\varphi\left(X_k(t)\right)\left(L_k(t) - L_k^n(t)\right),$$

$$U_t^{N,n}(\varphi) \triangleq \frac{1}{N}\sum_{k=1}^{N} E_t^{k,n}(\varphi),$$

and note that, combining the results in the previous sections, we get

$$\lim_{N\to\infty} U_t^{N,n}(\varphi) = \mathbb{E}^Q\left[\varphi\left(X(t)\right)n\left(L(t) - L^n(t)\right)\middle|\,\mathcal{F}_t^Y\right]$$

[3] A similar expansion holds for π_t using a straightforward application of the Kallianpur-Striebel's formula.

$$= n\left(\rho_t\left(\varphi\right) - \rho_t^n\left(\varphi\right)\right) \triangleq U_t^n\left(\varphi\right).$$

The goal is to find an evolution equation for the limit of U^n when n tends to infinity. This is attained in Theorem 4.4. Let us introduce first two preliminary results. The first is related to a martingale process that will converge to a Brownian motion as n tends to infinity. The second is the evolution equation for the process U^n.

Consider the sequence of processes $R^n = \{R_t^n\}_{t\geq 0}, n \in \mathbb{N}$, defined by

$$R^n\left(t\right) \triangleq 2\sqrt{3}\int_0^t \left(n\left(s - \tau_n\left(s\right)\right) - \frac{1}{2}\right) dY\left(s\right). \tag{4.1}$$

Lemma 4.1. *Under Q, (Y, R^n) converges in distribution to $(Y, \mathcal{R})$, where $\mathcal{R}$ is a d_Y-dimensional standard Brownian motion independent of Y and the B_k.*

Proof. The process (Y, R^n) is a $2d_Y$-dimensional martingale as R^n is a stochastic integral with respect to the Brownian motion Y. Moreover, observe that

$$\begin{aligned}
\left[Y^i, Y^j\right]_t &= [R^{n,i}, R^{n,j}]_t = [Y^i, R^{n,j}]_t = 0, \qquad i \neq j,\\
\left[R^{n,i}\right]_t &= 12\int_0^t \left(n\left(s - \tau_n\left(s\right)\right) - \frac{1}{2}\right)^2 ds = t + \mathcal{O}\left(n^{-1}\right), \qquad (4.2)\\
\left[Y^i, (R^n)^i\right]_t &= 2\sqrt{3}\int_0^t \left(n\left(s - \tau_n\left(s\right)\right) - \frac{1}{2}\right) ds = \mathcal{O}\left(n^{-1}\right).
\end{aligned}$$

The result follows by the martingale central limit theorem, for example, Theorem 1.4, Chapter 7 in [9]. □

Let $\{S_n\}_{n\geq 1}$ be a sequence of real-valued random processes. In what follows we will use the notation $S_n = \mathcal{O}\left(n^{-p}\right)$ for some $p \geq 0$ to indicate that

$$\mathbb{E}^Q\left[\sup_{0\leq s\leq t}\left|S_n\left(s\right)\right|^2\right]^{1/2} \leq \frac{C\left(t\right)}{n^p}, \quad t \geq 0,$$

for some positive constant $C\left(t\right)$.

Proposition 4.2. *For each $\varphi \in C_P^4$ and $n \in \mathbb{N}$, the process U^n satisfies, the following approximate evolution equation:*

$$\begin{aligned}
U_t^n\left(\varphi\right) = &\int_0^t U_s^n\left(A\varphi\right) ds + \int_0^t U_s^n\left(\varphi h^T\right) dY\left(s\right) + \frac{1}{2}\int_0^t \rho_s^n\left(\varphi A h^T\right) dY\left(s\right) \qquad (4.3)\\
&+ \frac{1}{2\sqrt{3}}\int_0^t \rho_s^n\left(\varphi A h^T\right) dR^n\left(s\right)\\
&+ \frac{1}{2\sqrt{3}}\sum_{i=1}^{d_Y}\int_0^t \rho_{\tau_n(s)}^n\left(\operatorname{tr}\left(\widetilde{O}_{\sigma,\varphi,h_i}\right)\right) dR^{n,i}\left(s\right)\\
&+ \frac{1}{2}\sum_{i=1}^{d_Y}\int_0^t \rho_{\tau_n(s)}^n\left(\operatorname{tr}\left(\widetilde{O}_{\sigma,\varphi,h_i}\right)\right) dY^i\left(s\right) + \Gamma_t^n(\varphi),
\end{aligned}$$

where $\Gamma^n(\varphi)$ is a process satisfying $\Gamma^n(\varphi) = \mathcal{O}\left(n^{-1/2}\right)$ for all $\varphi \in C_P^3$, and

$$\widetilde{O}_{\sigma,\varphi,h_i}\left(x\right) = \frac{1}{2}\left(O_{\sigma,\varphi,h_i}\left(x\right) + O_{\sigma,\varphi,h_i}^T\left(x\right)\right),$$

with

$$O_{\sigma,\varphi,h_i}(x) = \sigma^T(x)\,\nabla\varphi(x)\,\nabla h_i^T(x)\,\sigma(x).$$

We give the proof of Proposition 4.2 in Section 4.1 below.

Lemma 4.3. *For each $\varphi \in C_P^4$ and $t > 0$, we have that*

$$\sup_{n\in\mathbb{N}} \mathbb{E}^Q\left[\sup_{0\le s\le t} |U_s^n(\varphi)|^2\right] \le C(t,\varphi),$$

for some positive constant $C(t,\varphi)$.

Proof. Theorem 2.3 in [6], with $m = 1$, states that, for any $\varphi \in C_P^2$, there exists a constant C (not depending on n but possibly on t, φ, b, σ, and h) such that

$$\sup_n \sup_{0\le s\le t} \mathbb{E}^Q\left[|U_s^n(\varphi)|^2\right] < \infty,$$

This estimate clearly yields that, for any $t \ge 0$ and $\varphi \in C_P^4$ (we need the additional smoothness to ensure that $A\varphi \in C_P^4$)

$$\bar{C}_t := \sup_n \left(\sup_{0\le s\le t} \mathbb{E}^Q\left[|U_s^n(A\varphi)|^2\right] + \sup_{0\le s\le t} \mathbb{E}^Q\left[|U_s^n(\varphi h)|^2\right]\right) < \infty,$$

By Proposition 4.2, we can write

$$U_t^n(\varphi) = \int_0^t U_s^n(A\varphi)\,ds + \int_0^t U_s^n\left(\varphi h^T\right) dY(s) + S_t(\varphi),$$

where in $S_t(\varphi)$ we put all the terms in equation (4.3) not containing U^n. We deduce that

$$\begin{aligned}
\mathbb{E}^Q\left[\sup_{0\le s\le t} |U_s^n(\varphi)|^2\right] &\le 3\mathbb{E}^Q\left[\sup_{0\le s\le t}\left|\int_0^s U_r^n(A\varphi)\,dr\right|^2\right] \\
&\quad + 3\mathbb{E}^Q\left[\sup_{0\le s\le t}\left|\int_0^s U_r^n\left(\varphi h^T\right) dY(r)\right|^2\right] \\
&\quad + 3\mathbb{E}^Q\left[\sup_{0\le s\le t} |S_s(\varphi)|^2\right] \\
&\triangleq 3(I_1^n + I_2^n + I_3^n).
\end{aligned}$$

and, therefore, it suffices to bound I_1^n, I_2^n and I_3^n to justify the claim. Using Cauchy-Schwarz' inequality and Fubini's theorem we obtain

$$I_1^n \le \mathbb{E}^Q\left[\sup_{0\le s\le t} s\int_0^s |U_r^n(A\varphi)|^2\,dr\right] \le t\int_0^t \mathbb{E}^Q\left[|U_r^n(A\varphi)|^2\right] dr \le t^2\bar{C}_t,$$

All the remaining terms are stochastic integrals with respect to continuous martingales and can be controlled by means of Doob's maximal inequality. □

We are now in a position to state and prove the main result of this section.

Theorem 4.4. *Let $\{U^n\}_{n\geq 1}$ be the sequence of measure-valued processes given by*

$$U_t^n = n\left(\rho_t - \rho_t^n\right), \qquad t \in [0,\infty).$$

Then for each $\varphi \in C_P^8$, the sequence $\left\{U^n(\varphi), U^n(A\varphi), U^n\left(\varphi h^i\right), i = 1,\ldots,d_Y\right\}$ is relatively compact in $C_{\mathbb{R}^{2+d_Y}}[0,\infty)$, and every limit point satisfies

$$\begin{aligned} U_t(\varphi) = &\int_0^t U_s(A\varphi)\,ds + \int_0^t U_s\left(\varphi h^T\right)dY(s) + \frac{1}{2}\int_0^t \rho_s\left(\varphi A h^T\right)dY(s) \\ &+ \frac{1}{2\sqrt{3}}\int_0^t \rho_s\left(\varphi A h^T\right)d\mathcal{R}(s) + \frac{1}{2\sqrt{3}}\sum_{i=1}^{d_Y}\int_0^t \rho_s\left(\operatorname{tr}\left(\widetilde{O}_{\sigma,\varphi,h_i}\right)\right)d\mathcal{R}^i(s) \\ &+ \frac{1}{2}\sum_{i=1}^{d_Y}\int_0^t \rho_s\left(\operatorname{tr}\left(\widetilde{O}_{\sigma,\varphi,h_i}\right)\right)dY^i(s), \end{aligned} \tag{4.4}$$

where $\mathcal{R}$ is a Brownian motion independent of Y and all B_k.

Proof. Fix $\varphi \in C_P^6$. This assumption along with the assumptions on σ, b, and h, assure that $A\varphi \in C_P^4$ and hence that the estimate in Lemma 4.3 applies. Using this estimate on the integrands in the first two integrals on the right of (4.3), we see that these integrals are relatively compact in $C_{\mathbb{R}}[0,\infty)$. The remaining terms on the right converge by the convergence of ρ^n. Consequently, $\{U^n(\varphi)\}$ is relatively compact in $C_{\mathbb{R}}[0,\infty)$. Moreover, if we take $\varphi \in C_P^8$, $A\varphi \in C_P^6$, so the integrands in the first two terms are relatively compact, and (4.4) is satisfied for any limit point. □

Remark 4.5. Note that we do not know if equation 4.4 has a unique solution and, therefore, we cannot claim that the sequence $\{U^n\}_{n\geq 1}$ is convergent.

4.1. Proof of Proposition 4.2. We can combine (2.8) and

$$\begin{aligned} \varphi(X_k(t))L_k^n(t) = &\varphi(X_k(0)) + \int_0^t L_k^n(s)\nabla\varphi^T\sigma(X_k(s))\,dB_k(s) \\ &+ \int_0^t L_k^n(s)A\varphi(X_k(s))\,ds + \int_0^t L_k^n(s)\varphi h^T(X_k(s))\,dY(s) \\ &- \int_0^t L_k^n(s)\varphi(X_k(s))\left\{h(X_k(s)) - h(X_k(\tau_n(s)))\right\}^T dY(s). \end{aligned} \tag{4.5}$$

to write a more convenient expression for $E_t^{k,n}(\varphi)$, that is,

$$\begin{aligned} E_t^{k,n}(\varphi) &= n\varphi(X_k(t))\left(L_k(t) - L_k^n(t)\right) \\ &= \int_0^t n\left\{L_k(s) - L_k^n(s)\right\}\nabla\varphi^T\sigma(X_k(s))\,dB_k(s) \\ &\quad + \int_0^t n\left\{L_k(s) - L_k^n(s)\right\}A\varphi(X_k(s))\,ds \\ &\quad + \int_0^t n\left\{L_k(s) - L_k^n(s)\right\}\varphi h^T(X_k(s))\,dY(s) \end{aligned}$$

$$+\sum_{i=1}^{d_Y}\int_0^t nL_k^n(s)\,\varphi(X_k(s))\,dM_k^{n,i}(s),$$

where the processes $M_k^{n,i}$ are defined by

$$M_k^{n,i}(t)\triangleq\int_0^t\left(h_i(X_k(s))-h_i(X_k(\tau_n(s)))\right)dY^i(s).$$

Moreover, using Itô's formula , we can write

$$\begin{aligned}
nM_k^{n,i}(t) &= \int_0^t n\left[\int_{\tau_n(s)}^s \nabla h_i^T\sigma(X_k(u))\,dB_k(u)+\int_{\tau_n(s)}^s Ah_i(X_k(u))\,du\right]dY^i(s)\\
&= \int_0^t\int_{\tau_n(s)}^s n\nabla h_i^T\sigma(X_k(u))\,dB_k(u)\,dY^i(s)\\
&\quad+\int_0^t\int_{\tau_n(s)}^s nAh_i(X_k(u))\,du\,dY^i(s)\\
&= \int_0^t\int_{\tau_n(s)}^s n\nabla h_i^T\sigma(X_k(u))\,dB_k(u)\,dY^i(s)\\
&\quad+\int_0^t Ah_i(X_k(\tau_n(s)))\,n(s-\tau_n(s))\,dY^i(s)\\
&\quad+\int_0^t\int_{\tau_n(s)}^s n\left\{Ah_i(X_k(u))-Ah_i(X_k(\tau_n(s)))\right\}du\,dY^i(s)\\
&= \int_0^t\int_{\tau_n(s)}^s n\nabla h_i^T\sigma(X_k(u))\,dB_k(u)\,dY^i(s)+K_k^{n,i}(t)+I_k^{n,i}(t),
\end{aligned}$$

where

$$\begin{aligned}
K_k^{n,i}(t) &\triangleq \int_0^t Ah_i(X_k(\tau_n(s)))\,n(s-\tau_n(s))\,dY^i(s),\\
I_k^{n,i}(t) &\triangleq \int_0^t\int_{\tau_n(s)}^s n\left\{Ah_i(X_k(u))-Ah_i(X_k(\tau_n(s)))\right\}du\,dY^i(s).
\end{aligned}$$

Since,

$$(n(s-\tau_n(s)))\,dY^i(s)=\frac{dR^{n,i}(s)}{2\sqrt{3}}+\frac{1}{2}dY^i(s),$$

we can write

$$K_k^{n,i}(t)=\frac{1}{2\sqrt{3}}\int_0^t Ah_i(X_k(s))\,dR^{n,i}(s)+\frac{1}{2}\int_0^t Ah_i(X_k(s))\,dY^i(s),$$

and

$$\begin{aligned}
nM_k^{n,i}(t) &= \int_0^t\int_{\tau_n(s)}^s n\nabla h_i^T\sigma(X_k(u))\,dB_k(u)\,dY^i(s)\\
&\quad+\frac{1}{2\sqrt{3}}\int_0^t Ah_i(X_k(s))\,dR^{n,i}(s)
\end{aligned}$$

$$+\frac{1}{2}\int_0^t Ah_i\left(X_k\left(s\right)\right)dY^i\left(s\right)+I_k^{n,i}\left(s\right).$$

Finally, we can also write

$$\begin{aligned}
E_t^{k,n}\left(\varphi\right)=&\int_0^t n\left\{L_k\left(s\right)-L_k^n\left(s\right)\right\}\nabla\varphi^T\sigma\left(X_k\left(s\right)\right)dB_k\left(s\right)\\
&+\int_0^t n\left\{L_k\left(s\right)-L_k^n\left(s\right)\right\}A\varphi\left(X_k\left(s\right)\right)ds\\
&+\int_0^t n\left\{L_k\left(s\right)-L_k^n\left(s\right)\right\}\varphi h^T\left(X_k\left(s\right)\right)dY\left(s\right)\\
&+\sum_{i=1}^{d_Y}\int_0^t L_k^n\left(s\right)\varphi\left(X_k\left(s\right)\right)\int_{\tau_n(s)}^s n\nabla h_i^T\sigma\left(X_k\left(u\right)\right)dB_k\left(u\right)dY^i\left(s\right)\\
&+\sum_{i=1}^{d_Y}\frac{1}{2}\int_0^t L_k^n\left(s\right)\varphi Ah_i\left(X_k\left(s\right)\right)dY^i\left(s\right)\\
&+\sum_{i=1}^{d_Y}\frac{1}{2\sqrt{3}}\int_0^t L_k^n\left(s\right)\varphi Ah_i\left(X_k\left(s\right)\right)dR^{n,i}\left(s\right)\\
&+\sum_{i=1}^{d_Y}\int_0^t L_k^n\left(s\right)\varphi\left(X_k\left(s\right)\right)dI_k^{n,i}\left(s\right)\triangleq\sum_{i=1}^{7}A_i.
\end{aligned}$$

The result follows by averaging over $1\leq k\leq N$ in the previous equation and taking limits when N tends to infinity, combined with Lemmas A.6, 4.6 and 4.7:

- For the term on the left hand side of the previous equation we have

$$\lim_{N\to\infty}\frac{1}{N}\sum_{k=1}^N E_t^{k,n}\left(\varphi\right)=\lim_{N\to\infty}U_t^{N,n}\left(\varphi\right)=U_t^n\left(\varphi\right).$$

- For the term A_1 we can write

$$\begin{aligned}
&\lim_{N\to\infty}\frac{1}{N}\sum_{k=1}^N\int_0^t n\left\{L_k\left(s\right)-L_k^n\left(s\right)\right\}\nabla\varphi^T\sigma\left(X_k\left(s\right)\right)dB_k\left(s\right)\\
&=\mathbb{E}^Q\left[\int_0^t n\left\{L\left(s\right)-L^n\left(s\right)\right\}\nabla\varphi^T\sigma\left(X\left(s\right)\right)dB\left(s\right)\middle|\mathcal{F}_t^Y\right]\\
&=\mathbb{E}^Q\left[\mathbb{E}^Q\left[\int_0^t n\left\{L\left(s\right)-L^n\left(s\right)\right\}\nabla\varphi^T\sigma\left(X\left(s\right)\right)dB\left(s\right)\middle|\mathcal{F}_t^Y\vee\mathcal{F}_0^V\right]\middle|\mathcal{F}_t^Y\right]=0.
\end{aligned}$$

- For the term A_2, using Proposition 3.15 in [1], we can write

$$\begin{aligned}
&\lim_{N\to\infty}\frac{1}{N}\sum_{k=1}^N\int_0^t n\left\{L_k\left(s\right)-L_k^n\left(s\right)\right\}A\varphi\left(X_k\left(s\right)\right)ds\\
&=\mathbb{E}^Q\left[\int_0^t n\left\{L\left(s\right)-L^n\left(s\right)\right\}A\varphi\left(X\left(s\right)\right)ds\middle|\mathcal{F}_t^Y\right]
\end{aligned}$$

$$= \int_0^t n\mathbb{E}^Q \left[\{L(s) - L^n(s)\} A\varphi(X(s)) | \mathcal{F}_s^Y \right] ds$$
$$= \int_0^t U_s^n(A\varphi)\, ds.$$

- The terms A_3, A_5 and A_6 are treated similarly as the term A_2. Note that the processes $R^{n,i}$ are $\mathcal{F}_t^Y$-adapted.
- For the term A_4 we apply Lemma 4.7.
- For the term A_7 we apply Lemma 4.6.

The process $\Gamma_t^n(\varphi) = \sum_{i=1}^{d_Y} \left(\Phi_t^{n,i}(\varphi) + \Psi_t^{n,i}(\varphi) \right)$, where $\Phi_t^{n,1}(\varphi)$ and $\Psi_t^{n,i}(\varphi)$ are the processes in the statement of Lemmas 4.6 and 4.7.

4.2. Auxiliary lemmas.

Lemma 4.6. *For all $i = 1, ..., d_Y$ and φ Borel measurable with at most polynomial growth, let*

$$I_k^{n,i}(t) = \int_0^t \int_{\tau_n(s)}^s n \{Ah_i(X_k(u)) - Ah_i(X_k(\tau_n(s)))\} du dY^i(s).$$

Then,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^N \int_0^t L_k^n(s) \varphi(X_k(s)) dI_k^{n,i}(s) = \Phi_t^{n,i}(\varphi),$$

where $\Phi_t^{n,i}(\varphi) = \mathcal{O}\left(n^{-1/2}\right)$.

Proof. Applying Lemma A.6 we have that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^N \int_0^t L_k^n(s) \varphi(X_k(s)) dI_k^{n,i}(s)$$
$$= \mathbb{E}^Q \left[\int_0^t L^n(s) \varphi(X(s)) dI^{n,i}(s) \middle| \mathcal{F}_t^Y \right],$$

where

$$dI^{n,i}(s) = \left(\int_{\tau_n(s)}^s n \{Ah_i(X(u)) - Ah_i(X(\tau_n(s)))\} du \right) dY^i(s).$$

In what follows we will use the more compact notation

$$\Delta_n^i(u) \triangleq Ah_i(X(u)) - Ah_i(X(\tau_n(s))), \quad \tau_n(s) \le u \le s.$$

Note that, using Proposition 3.15 in [1], we can write

$$\mathbb{E}^Q \left[\int_0^t L^n(s) \varphi(X(s)) dI^{n,i}(s) \middle| \mathcal{F}_t^Y \right]$$
$$= \mathbb{E}^Q \left[\int_0^t L^n(s) \varphi(X(s)) \left(\int_{\tau_n(s)}^s n\Delta_n^i(u) du \right) dY^i(s) \middle| \mathcal{F}_t^Y \right]$$
$$= \int_0^t \mathbb{E}^Q \left[L^n(s) \varphi(X(s)) \left(\int_{\tau_n(s)}^s n\Delta_n^i(u) du \right) \middle| \mathcal{F}_s^Y \right] dY^i(s)$$

$$\triangleq \Phi_t^{n,i}(\varphi)$$

Moreover, using Burkholder-Davis-Gundy inequality, Jensen's inequality for conditional expectation, the law of total expectation, Fubini's theorem and Cauchy-Schwarz inequality we obtain

$$\begin{aligned}
&\mathbb{E}^Q\left[\sup_{0\le t\le T}\left|\mathbb{E}^Q\left[\int_0^t L_1^n(s)\,\varphi(X_1(s))\,dI_1^{n,i}(s)\middle|\mathcal{F}_t^Y\right]\right|^2\right]\\
&\le \mathbb{E}^Q\left[\int_0^T\left|\mathbb{E}^Q\left[L^n(s)\,\varphi(X(s))\left(\int_{\tau_n(s)}^s n\Delta_n^i(u)\,du\right)\middle|\mathcal{F}_s^Y\right]\right|^2 ds\right]\\
&\le \mathbb{E}^Q\left[\int_0^T\left|L^n(s)\,\varphi(X(s))\left(\int_{\tau_n(s)}^s n\Delta_n^i(u)\,du\right)\right|^2 ds\right]\\
&= \int_0^T \mathbb{E}^Q\left[\left|L^n(s)\,\varphi(X(s))\left(\int_{\tau_n(s)}^s n\Delta_n^i(u)\,du\right)\right|^2\right] ds\\
&\le \int_0^T \mathbb{E}^Q\left[\left|L^n(s)\,\varphi(X(s))\right|^4\right]^{1/2}\mathbb{E}^Q\left[\left|\left(\int_{\tau_n(s)}^s n\Delta_n^i(u)\,du\right)\right|^4\right]^{1/2} ds.
\end{aligned}$$

Using Jensen's innequality and Itô's formula we get that

$$\begin{aligned}
&\mathbb{E}^Q\left[\left|\left(\int_{\tau_n(s)}^s n\Delta_n^i(u)\,du\right)\right|^4\right]\\
&\le (s-\tau_n(s))^3 n^4\int_{\tau_n(s)}^s \mathbb{E}^Q\left[\left|\Delta_n^i(u)\right|^4\right]du\\
&\le Cn\left\{\int_{\tau_n(s)}^s \mathbb{E}^Q\left[\left|\int_{\tau_n(s)}^u \nabla(Ah_i)^T\sigma(X(v))\,dB(v)\right|^4\right]du\right.\\
&\quad\left.+\int_{\tau_n(s)}^s \mathbb{E}^Q\left[\left|\int_{\tau_n(s)}^u A^2h_i(X(v))\,dv\right|^4\right]du\right\}\\
&= B_1+B_2
\end{aligned}$$

Due to the hypothesis on σ, b and h we have that $\nabla(Ah_i)^T\sigma$ has at most polynomial growth, which combined with the bound (2.4) yields that

$$\mathbb{E}^Q\left[\sup_{0\le s\le T}\left|\nabla(Ah_i)^T\sigma(X(s))\right|^4\right] = C(T,h,b,\sigma)<\infty.$$

Therefore, using Burkholder-Davis-Gundy innequality we obtain

$$B_1 \le Cn\int_{\tau_n(s)}^s \mathbb{E}^Q\left[\left|\int_{\tau_n(s)}^u\left|\nabla(Ah_i)^T\sigma(X(v))\right|^2 dv\right|^2\right]du$$

$$\leq Cn\int_{\tau_n(s)}^{s}(u-\tau_n(s))\int_{\tau_n(s)}^{u}\mathbb{E}^Q\left[\left|\nabla(Ah_i)^T\sigma(X(v))\right|^4\right]dvdu$$
$$\leq C(T,h,b,\sigma)\,n\int_{\tau_n(s)}^{s}(u-\tau_n(s))^2\,du$$
$$\leq C(T,h,b,\sigma)\,n^{-2}$$

For the term B_2 one can use similar reasonings as for B_1 to obtain that

$$B_2\leq C(T,h,b,\sigma)\,n^{-4},$$

and, hence,

$$\mathbb{E}^Q\left[\left|\left(\int_{\tau_n(s)}^{s}n\Delta_n^i(u)\,du\right)\right|^4\right]^{1/2}\leq Cn^{-1}$$

On the other hand, using Hölder's innequality, the bounds for L^n in Lemma 3.9 in [6], that φ has at most polynomial growth and (2.4) we get that

$$\sup_{s\in[0,T]}\sup_{n\in\mathbb{N}}\mathbb{E}^Q\left[\left|L^n(s)\,\varphi(X(s))\right|^4\right]<\infty.$$

Combining the previous estimates we can conclude that $\Phi_t^{n,i}(\varphi)=\mathcal{O}\left(n^{-1/2}\right)$. □

Lemma 4.7. *For all $i=1,...,d_Y$ and $\varphi\in C_P^3$, we have that*

$$\lim_{N\to\infty}\frac{1}{N}\sum_{k=1}^{N}n\int_0^t L_k^n(s)\,\varphi(X_k(s))\int_{\tau_n(s)}^{s}\nabla h_i^T\sigma(X_k(u))\,dB_k(u)\,dY^i(s)$$
$$=\frac{1}{2\sqrt{3}}\sum_{i=1}^{d_Y}\int_0^t\rho_{\tau_n(s)}^n\left(\operatorname{tr}\left(\widetilde{O}_{\sigma,\varphi,h_i}\right)\right)dR^{n,i}(s)$$
$$+\frac{1}{2}\sum_{i=1}^{d_Y}\int_0^t\rho_{\tau_n(s)}^n\left(\operatorname{tr}\left(\widetilde{O}_{\sigma,\varphi,h_i}\right)\right)dY^i(s)+\Psi_t^{n,i}(\varphi),$$

where $\Psi_t^{n,i}(\varphi)=\mathcal{O}\left(n^{-1/2}\right)$.

Proof. We can write

$$n\int_0^t L_k^n(s)\,\varphi(X_k(s))\int_{\tau_n(s)}^{s}\nabla h_i^T\sigma(X_k(u))\,dB_k(u)\,dY^i(s)$$
$$=n\int_0^t\left\{L_k^n(s)\,\varphi(X_k(s))-L_k^n(\tau_n(s))\,\varphi(X_k(\tau_n(s)))\right\}$$
$$\times\int_{\tau_n(s)}^{s}\nabla h_i^T\sigma(X_k(u))\,dB_k(u)\,dY^i(s)$$
$$+n\int_0^t L_k^n(\tau_n(s))\,\varphi(X_k(\tau_n(s)))\int_{\tau_n(s)}^{s}\nabla h_i^T\sigma(X_k(u))\,dB_k(u)\,dY^i(s)$$
$$\triangleq C_{k,1}+C_{k,2}.$$

Using similar reasonings as for the term A_1 in the proof of Proposition 4.2), one has that

$$\lim_{N\to\infty}\frac{1}{N}\sum_{k=1}^{N}C_{k,2}=0.$$

Using integration by parts we obtain

$$\begin{aligned}
&L_k^n(s)\,\varphi(X_k(s))-L_k^n(\tau_n(s))\,\varphi(X_k(\tau_n(s)))\\
&=\int_{\tau_n(s)}^{s}L_k^n(u)\,\varphi h^T(X_k(\tau_n(s)))\,dY(u)\\
&\quad+\int_{\tau_n(s)}^{s}L_k^n(u)\,A\varphi(X_k(u))\,du+\int_{\tau_n(s)}^{s}L_k^n(u)\,\nabla\varphi^T\sigma(X_k(u))\,dB_k(u)\\
&\triangleq D_{k,1}(s)+D_{k,2}(s)+D_{k,3}(s),
\end{aligned}$$

and using Itô's formula

$$\begin{aligned}
&\int_{\tau_n(s)}^{s}\nabla h_i^T\sigma(X_k(u))\,dB_k(u)\\
&=\nabla h_i^T\sigma(X_k(\tau_n(s)))\,(B_k(s)-B_k(\tau_n(s)))\\
&\quad+\int_{\tau_n(s)}^{s}\int_{\tau_n(s)}^{u}\nabla\left(\nabla h_i^T\sigma\right)^T\sigma(X_k(v))\,dB_k(v)\,dB_k(u)\\
&\quad+\int_{\tau_n(s)}^{s}\int_{\tau_n(s)}^{u}A\left(\nabla h_i^T\sigma\right)^T(X_k(v))\,dvdB_k(u)\\
&\triangleq E_{k,1}(s)+E_{k,2}(s)+E_{k,3}(s).
\end{aligned}$$

Hence, the term $C_{k,1}$ can be written as the sum of nine terms

$$C_{k,1}=\sum_{l,m=1}^{3}n\int_0^t D_{k,l}(s)E_{k,m}(s)dY^i(s).$$

The terms containing as a factor $D_{k,2}(s), E_{k,2}(s)$, and $E_{k,3}(s)$, after averaging over $1\le k\le N$ and taking limit when N tends to infinity, yield processes which are at least of order $\mathcal{O}\left(n^{-1/2}\right)$. With similar reasonings as in Lemma 4.6 we can identify these processes as:

$$\begin{aligned}
\Psi_t^{n,i,1}(\varphi)&\triangleq n\int_0^t\mathbb{E}^Q\left[\int_{\tau_n(s)}^{s}L^n(u)\,A\varphi(X(u))\,du\right.\\
&\qquad\left.\times\ \nabla h_i^T\sigma(X(\tau_n(s)))\,(B(s)-B(\tau_n(s)))\Big|\,\mathcal{F}_s^Y\right]dY^i(s),\\
\Psi_t^{n,i,2}(\varphi)&\triangleq n\int_0^t\mathbb{E}^Q\left[\int_{\tau_n(s)}^{s}L^n(u)\,A\varphi(X(u))\,du\right.\\
&\qquad\left.\times\int_{\tau_n(s)}^{s}\int_{\tau_n(s)}^{u}\nabla\left(\nabla h_i^T\sigma\right)^T\sigma(X(v))\,dB(v)\,dB(u)\Bigg|\,\mathcal{F}_s^Y\right]dY^i(s),
\end{aligned}$$

$$\Psi_t^{n,i,3}(\varphi) \triangleq n\int_0^t \mathbb{E}^Q\left[\int_{\tau_n(s)}^s L^n(u)\,A\varphi(X(u))\,du \right.$$
$$\left. \times \int_{\tau_n(s)}^s\int_{\tau_n(s)}^u A\left(\nabla h_i^T\sigma\right)^T(X(v))\,dvdB(u)\middle|\mathcal{F}_s^Y\right]dY^i(s),$$
$$\Psi_t^{n,i,4}(\varphi) \triangleq n\int_0^t \mathbb{E}^Q\left[\int_{\tau_n(s)}^s L^n(u)\,\varphi h^T(X(\tau_n(s)))\,dY(u) \right.$$
$$\left. \times \int_{\tau_n(s)}^s\int_{\tau_n(s)}^u \nabla\left(\nabla h_i^T\sigma\right)^T\sigma(X(v))\,dB(v)\,dB(u)\middle|\mathcal{F}_s^Y\right]dY^i(s),$$
$$\Psi_t^{n,i,5}(\varphi) \triangleq n\int_0^t \mathbb{E}^Q\left[\int_{\tau_n(s)}^s L^n(u)\,\nabla\varphi^T\sigma(X(u))\,dB(u) \right.$$
$$\left. \times \int_{\tau_n(s)}^s\int_{\tau_n(s)}^u \nabla\left(\nabla h_i^T\sigma\right)^T\sigma(X(v))\,dB(v)\,dB(u)\middle|\mathcal{F}_s^Y\right]dY^i(s),$$
$$\Psi_t^{n,i,6}(\varphi) \triangleq n\int_0^t \mathbb{E}^Q\left[\int_{\tau_n(s)}^s L^n(u)\,\varphi h^T(X(\tau_n(s)))\,dY(u) \right.$$
$$\left. \times \int_{\tau_n(s)}^s\int_{\tau_n(s)}^u A\left(\nabla h_i^T\sigma\right)^T(X(v))\,dvdB(u)\middle|\mathcal{F}_s^Y\right]dY^i(s),$$
$$\Psi_t^{n,i,7}(\varphi) \triangleq n\int_0^t \mathbb{E}^Q\left[\int_{\tau_n(s)}^s L^n(u)\,\nabla\varphi^T\sigma(X(u))\,dB(u) \right.$$
$$\left. \times \int_{\tau_n(s)}^s\int_{\tau_n(s)}^u A\left(\nabla h_i^T\sigma\right)^T(X(v))\,dvdB(u)\middle|\mathcal{F}_s^Y\right]dY^i(s),$$

There are two terms left:

$$F_{k,1}^{n,i}(t,\varphi) \triangleq n\int_0^t\left(\int_{\tau_n(s)}^s L_k^n(u)\,\varphi(X_k(u))\,h^T(X_k(\tau_n(s)))\,dY(u)\right)$$
$$\times \nabla h_i^T\sigma(X_k(\tau_n(s)))\,(B_k(s)-B_k(\tau_n(s)))\,dY^i(s),$$

and

$$F_{k,2}^{n,i}(t,\varphi) \triangleq n\int_0^t\left(\int_{\tau_n(s)}^s L_k^n(u)\,\nabla\varphi^T\sigma(X_k(u))\,dB_k(u)\right)$$
$$\times \nabla h_i^T\sigma(X_k(\tau_n(s)))\,(B_k(s)-B_k(\tau_n(s)))\,dY^i(s).$$

For the term $F_{k,1}^{n,i}(t,\varphi)$, using integration by parts with $L_k^n(u)\,\varphi(X_k(u))$, we can write

$$F_{k,1}^{n,i}(t,\varphi) = n\int_0^t L_k^n(\tau_n(s))\,\varphi h^T(X_k(\tau_n(s)))\,(Y(s)-Y(\tau_n(s)))$$
$$\times \nabla h_i^T\sigma(X_k(\tau_n(s)))\,(B_k(s)-B_k(\tau_n(s)))\,dY^i(s)$$
$$+\Psi_t^{k,n,i,8}(\varphi)+\Psi_t^{k,n,i,9}(\varphi)+\Psi_t^{k,n,i,10}(\varphi)$$

We get, using similar reasonings as for the term $C_{k,2}$, that

$$\lim_{N\to\infty}\frac{1}{N}\sum_{k=1}^{N}n\int_0^t L_k^n\left(\tau_n\left(s\right)\right)\varphi h^T\left(X_k\left(\tau_n\left(s\right)\right)\right)\left(Y\left(s\right)-Y\left(\tau_n\left(s\right)\right)\right)=0,$$

and the following terms

$$\begin{aligned}
\Psi_t^{n,i,8}\left(\varphi\right) &\triangleq \lim_{N\to\infty}\frac{1}{N}\sum_{k=1}^{N}\Psi_t^{k,n,i,8}\left(\varphi\right)\\
&= n\int_0^t \mathbb{E}^Q\left[\int_{\tau_n(s)}^{s}\int_{\tau_n(s)}^{u} L^n\left(r\right)\nabla\varphi^T\sigma\left(X\left(r\right)\right)dB\left(r\right)h^T\left(X\left(\tau_n\left(s\right)\right)\right)\right.\\
&\quad\times\left. dY\left(u\right)\nabla h_i^T\sigma\left(X\left(\tau_n\left(s\right)\right)\right)\left(B\left(s\right)-B\left(\tau_n\left(s\right)\right)\right)\middle|\,\mathcal{F}_s^Y\right]dY^i\left(s\right)\\
\Psi_t^{n,i,9}\left(\varphi\right) &\triangleq \lim_{N\to\infty}\frac{1}{N}\sum_{k=1}^{N}\Psi_t^{k,n,i,9}\left(\varphi\right)\\
&= n\int_0^t \mathbb{E}^Q\left[\int_{\tau_n(s)}^{s}\int_{\tau_n(s)}^{u} L^n\left(r\right)A\varphi\left(X\left(r\right)\right)drh^T\left(X\left(\tau_n\left(s\right)\right)\right)dY\left(u\right)\right.\\
&\quad\times\left.\nabla h_i^T\sigma\left(X\left(\tau_n\left(s\right)\right)\right)\left(B\left(s\right)-B\left(\tau_n\left(s\right)\right)\right)\middle|\,\mathcal{F}_s^Y\right]dY^i\left(s\right)\\
\Psi_t^{n,i,10}\left(\varphi\right) &\triangleq \lim_{N\to\infty}\frac{1}{N}\sum_{k=1}^{N}\Psi_t^{k,n,i,10}\left(\varphi\right)\\
&= n\int_0^t \mathbb{E}^Q\left[\int_{\tau_n(s)}^{s}\int_{\tau_n(s)}^{u} L^n\left(r\right)\varphi\left(X\left(r\right)\right)h^T\left(X\left(\tau_n\left(s\right)\right)\right)dY\left(r\right)\right.\\
&\quad\times h^T\left(X\left(\tau_n\left(s\right)\right)\right)dY\left(u\right)\\
&\quad\times\left.\nabla h_i^T\sigma\left(X\left(\tau_n\left(s\right)\right)\right)\left(B\left(s\right)-B\left(\tau_n\left(s\right)\right)\right)\middle|\,\mathcal{F}_s^Y\right]dY^i\left(s\right),
\end{aligned}$$

which are at least of order $\mathcal{O}\left(n^{-1/2}\right)$.

For the term $F_{k,2}^{n,i}\left(t,\varphi\right)$, using integration by parts with $L_k^n\left(u\right)\nabla\varphi^T\sigma\left(X_k\left(u\right)\right)$, we can write

$$\begin{aligned}
F_{k,2}^{n,i}\left(t,\varphi\right) = n\int_0^t & L_k^n\left(\tau_n\left(s\right)\right)\nabla\varphi^T\sigma\left(X_k\left(\tau_n\left(s\right)\right)\right)\left(B_k\left(s\right)-B_k\left(\tau_n\left(s\right)\right)\right)\\
&\times\nabla h_i^T\sigma\left(X_k\left(\tau_n\left(s\right)\right)\right)\left(B_k\left(s\right)-B_k\left(\tau_n\left(s\right)\right)\right)dY^i\left(s\right)\\
&+\Psi_t^{k,n,i,11}\left(\varphi\right)+\Psi_t^{k,n,i,12}\left(\varphi\right)+\Psi_t^{k,n,i,13}\left(\varphi\right).
\end{aligned}\tag{4.6}$$

Denote the first term on the right hand side of the previous equation by $\widehat{F}_{k,2}^{n,i}\left(t,\varphi\right)$, and consider the following matrices

$$O_{\sigma,\varphi,h_i}\left(x\right)\triangleq\sigma^T\left(x\right)\nabla\varphi\left(x\right)\nabla h_i^T\left(x\right)\sigma\left(x\right),$$

and

$$\widetilde{O}_{\sigma,\varphi,h_i}\left(x\right)\triangleq\frac{1}{2}\left(O_{\sigma,\varphi,h_i}\left(x\right)+O_{\sigma,\varphi,h_i}^T\left(x\right)\right).$$

Then, $\widehat{F}_{k,2}^{n,i}(t,\varphi)$ can be further expanded using again integration by parts

$$\begin{aligned}
\widehat{F}_{k,2}^{n,i}(t,\varphi) &= n\int_0^t L_k^n(\tau_n(s))\,\nabla\varphi^T\sigma(X_k(\tau_n(s)))(B_k(s)-B_k(\tau_n(s)))\\
&\quad\times\nabla h_i^T\sigma(X_k(\tau_n(s)))(B_k(s)-B_k(\tau_n(s)))\,dY^i(s).\\
&= n\int_0^t L_k^n(\tau_n(s))(B_k(s)-B_k(\tau_n(s)))^T\\
&\quad\times O_{\sigma,\varphi,h_i}(X_k(\tau_n(s)))(B_k(s)-B_k(\tau_n(s)))\,dY^i(s)\\
&= n\int_0^t L_k^n(\tau_n(s))\int_{\tau_n(s)}^s (B_k(u)-B_k(\tau_n(s)))^T\\
&\quad\times\left(O_{\sigma,\varphi,h_i}+O_{\sigma,\varphi,h_i}^T\right)(X_k(\tau_n(s)))\,dB_k(u)\,dY^i(s)\\
&\quad+\frac{n}{2}\int_0^t L_k^n(\tau_n(s))\\
&\quad\times\operatorname{tr}\left(\left(O_{\sigma,\varphi,h_i}+O_{\sigma,\varphi,h_i}^T\right)(X_k(\tau_n(s)))\right)(s-\tau_n(s))\,dY^i(s)\\
&= 2n\int_0^t L_k^n(\tau_n(s))\\
&\quad\times\int_{\tau_n(s)}^s (B_k(u)-B_k(\tau_n(s)))^T\widetilde{O}_{\sigma,\varphi,h_i}(X_k(\tau_n(s)))\,dB_k(u)\,dY^i(s)\\
&\quad+n\int_0^t L_k^n(\tau_n(s))\operatorname{tr}\left(\widetilde{O}_{\sigma,\varphi,h_i}(X_k(\tau_n(s)))\right)(s-\tau_n(s))\,dY^i(s)
\end{aligned}$$

On the one hand, using similar reasonings as for the term $C_{k,2}$, we have

$$\begin{aligned}
0 &= \lim_{N\to\infty}\frac{1}{N}\sum_{k=1}^N 2n\int_0^t L_k^n(\tau_n(s))\\
&\quad\times\int_{\tau_n(s)}^s (B_k(u)-B_k(\tau_n(s)))^T\widetilde{O}_{\sigma,\varphi,h_i}(X_k(\tau_n(s)))\,dB_k(u)\,dY^i(s).
\end{aligned}$$

On the other hand, recalling the expression for $n(s-\tau_n(s))\,dY^i(s)$, we get

$$\begin{aligned}
&n\int_0^t L_k^n(\tau_n(s))\operatorname{tr}\left(\widetilde{O}_{\sigma,\varphi,h_i}(X_k(\tau_n(s)))\right)(s-\tau_n(s))\,dY^i(s)\\
&=\frac{1}{2\sqrt{3}}\int_0^t L_k^n(\tau_n(s))\operatorname{tr}\left(\widetilde{O}_{\sigma,\varphi,h_i}(X_k(\tau_n(s)))\right)dR^{n,i}(s)\\
&\quad+\frac{1}{2}\int_0^t L_k^n(\tau_n(s))\operatorname{tr}\left(\widetilde{O}_{\sigma,\varphi,h_i}(X_k(\tau_n(s)))\right)dY^i(s),
\end{aligned}$$

and

$$\begin{aligned}
&\lim_{N\to\infty}\frac{1}{N}\sum_{k=1}^N\frac{1}{2\sqrt{3}}\int_0^t L_k^n(\tau_n(s))\operatorname{tr}\left(\widetilde{O}_{\sigma,\varphi,h_i}(X_k(\tau_n(s)))\right)dR^{n,i}(s)\\
&=\frac{1}{2\sqrt{3}}\int_0^t \rho_{\tau_n(s)}^n\left(\operatorname{tr}\left(\widetilde{O}_{\sigma,\varphi,h_i}\right)\right)dR^{n,i}(s),
\end{aligned}$$

$$\lim_{N\to\infty}\frac{1}{N}\sum_{k=1}^{N}\frac{1}{2}\int_0^t L_k^n\left(\tau_n\left(s\right)\right)\operatorname{tr}\left(\widetilde{O}_{\sigma,\varphi,h_i}\left(X_k\left(\tau_n\left(s\right)\right)\right)\right)dY^i\left(s\right)$$
$$=\frac{1}{2}\int_0^t \rho^n_{\tau_n(s)}\left(\operatorname{tr}\left(\widetilde{O}_{\sigma,\varphi,h_i}\right)\right)dY^i\left(s\right)$$

With similar reasonings as in Lemma 4.6, we can identify the processes in (4.6) after the averaging and limiting procedure. That is

$$\begin{aligned}\Psi_t^{n,i,11}\left(\varphi\right)&\triangleq\lim_{N\to\infty}\frac{1}{N}\sum_{k=1}^{N}\Psi_t^{k,n,i,11}\left(\varphi\right)\\&=n\int_0^t\mathbb{E}^Q\left[\int_{\tau_n(s)}^s\int_{\tau_n(s)}^u L^n\left(r\right)\nabla\left(\nabla\varphi^T\sigma\right)^T\sigma\left(X\left(r\right)\right)dB\left(r\right)dB\left(u\right)\right.\\&\quad\left.\times\ \nabla h_i^T\sigma\left(X\left(\tau_n\left(s\right)\right)\right)\left(B\left(s\right)-B\left(\tau_n\left(s\right)\right)\right)\middle|\mathcal{F}_s^Y\right]dY^i\left(s\right)\\\Psi_t^{n,i,12}\left(\varphi\right)&\triangleq\lim_{N\to\infty}\frac{1}{N}\sum_{k=1}^{N}\Psi_t^{k,n,i,12}\left(\varphi\right)\\&=n\int_0^t\mathbb{E}^Q\left[\int_{\tau_n(s)}^s\int_{\tau_n(s)}^u L^n\left(r\right)A\left(\nabla\varphi^T\sigma\right)^T\sigma\left(X\left(r\right)\right)dB\left(r\right)dB\left(u\right)\right.\\&\quad\left.\times\ \nabla h_i^T\sigma\left(X\left(\tau_n\left(s\right)\right)\right)\left(B\left(s\right)-B\left(\tau_n\left(s\right)\right)\right)\middle|\mathcal{F}_s^Y\right]dY^i\left(s\right)\\\Psi_t^{n,i,13}\left(\varphi\right)&\triangleq\lim_{N\to\infty}\frac{1}{N}\sum_{k=1}^{N}\Psi_t^{k,n,i,13}\left(\varphi\right)\\&=n\int_0^t\mathbb{E}^Q\left[\int_{\tau_n(s)}^s\int_{\tau_n(s)}^u L^n\left(r\right)\nabla\varphi^T\sigma\left(X\left(r\right)\right)h^T\left(X\left(\tau_n\left(s\right)\right)\right)dY\left(r\right)\right.\\&\quad\left.\times\ dB\left(u\right)\nabla h_i^T\sigma\left(X\left(\tau_n\left(s\right)\right)\right)\left(B\left(s\right)-B\left(\tau_n\left(s\right)\right)\right)\middle|\mathcal{F}_s^Y\right]dY^i\left(s\right),\end{aligned}$$

which are at least of order $\mathcal{O}\left(n^{-1/2}\right)$. Finally, $\Psi_t^{n,i}\left(\varphi\right)=\sum_{j=1}^{13}\Psi_t^{n,i,j}\left(\varphi\right)$.

□

Appendix A. Appendix

A.1. Limits for particle representations. As we will see, particle representations are useful in deriving approximations and computing limits. In the following lemma from [15], N_n may be finite or infinite. If $N_n=\infty$, then by $\frac{1}{N_n}\sum_{k=1}^{N_n}z_k$, we mean $\lim_{m\to\infty}\frac{1}{k}\sum_{k=1}^{m}z_k$.

Lemma A.1. *Let $Z^n=(Z_1^n,\ldots,Z_{N_n}^n)$ be exchangeable families of $D_E[0,\infty)$-valued random variables such that $N_n\Rightarrow\infty$ and $Z^n\Rightarrow Z$ in $D_E[0,\infty)^\infty$. Define*

$\Xi^n=\frac{1}{N_n}\sum_{k=1}^{N_n}\delta_{Z_k^n}\in\mathcal{P}(D_E[0,\infty))$

$\Xi=\lim_{m\to\infty}\frac{1}{m}\sum_{k=}^{m}\delta_{Z_k}$

$V^n(t)=\frac{1}{N_n}\sum_{k=1}^{N_n}\delta_{Z_k^n(t)}\in\mathcal{P}(E)$

$V(t)=\lim_{m\to\infty}\frac{1}{m}\sum_{k=1}^{m}\delta_{Z_k(t)}$

Then

a) *For* $t_1, \ldots, t_l \notin \{t : \mathbb{E}\left[\Xi\left\{x : x(t) \neq x(t-)\right\}\right] > 0\}$

$$(\Xi_n, V^n(t_1), \ldots, V^n(t_l)) \Rightarrow (\Xi, V(t_1), \ldots, V(t_l)).$$

b) *If* $Z^n \Rightarrow Z$ *in* $D_{E^\infty}[0,\infty)$*, then* $V^n \Rightarrow V$ *in* $D_{\mathcal{P}(E)}[0,\infty)$*. If* $Z^n \to Z$ *in probability in* $D_{E^\infty}[0,\infty)$*, then* $V^n \to V$ *in* $D_{\mathcal{P}(E)}[0,\infty)$ *in probability.*

Remark A.2. If the Z^n_k are in $C_E[0,\infty)$, then every D can be replaced by C. In particular, in this case, $D_E[0,\infty)^\infty = C_{E^\infty}[0,\infty)$.

Remark A.3. If z^n converges in $D_{\mathcal{P}(E)}[0,\infty)$ to $z \in D_{\mathcal{P}(E)}[0,\infty)$, then for any bounded continuous $f : E \to \mathbb{R}$, x^n given by $x^n(t) = \int_E f(u) z^n(t, du)$ converges to x given by $x(t) = \int_E f(u) z(t, du)$ in $D_{\mathbb{R}}[0,\infty)$. We need similar results for unbounded f.

Lemma A.4. *Let* E *be locally compact and* $C_0(E)$ *be the space of continuous functions vanishing at infinity. Let* $\{\mu_n\} \subset \mathcal{P}(E)$ *and* $\mu_n \Rightarrow \mu$*. Suppose* $\varphi \in C(E)$*,* $\varphi > 0$*, and* $\sup_n \int_E \varphi d\mu_n < \infty$*. Then if* $f \in C(E)$ *and* $\frac{f}{\varphi} \in C_0(E)$*,* $\lim_{n\to\infty} \int_E f d\mu_n = \int_E f d\mu$*.*

Proof. Let E^∞ denote the one point compactification and $d\nu_n = \varphi d\mu_n$. Then since $\sup_n \nu_n(E^\infty) < \infty$, $\{\nu_n\}$ is relatively compact in $\mathcal{M}_f(E^\infty)$, the finite measures on E^∞, and every limit point is of the form $\varphi\mu + p\delta_\infty$. Consequently, if $\frac{f}{\varphi} \in C_0(E)$,

$$\lim_{n\to\infty} \int_E f d\mu_n = \lim_{n\to\infty} \int_{E_\infty} \frac{f}{\varphi} d\nu_n = \int_E f d\mu.$$

□

Lemma A.5. *Let* E *be locally compact. Suppose* z^n *converges in* $D_{\mathcal{P}(E)}[0,\infty)$ *to* $z \in D_{\mathcal{P}(E)}[0,\infty)$ *and for each* $T = 1, 2, \cdots$*, there exists* $\varphi_T \in C(E)$*,* $\varphi_T > 0$*, such that*

$$\sup_n \sup_{t \le T} \int_E \varphi_T(u) z^n(t, du) < \infty.$$

If $f \in C(E)$ *and for each* T*,* $\frac{f}{\varphi_T} \in C_0(E)$*, then* $y_n(t) = \int_E f(u) z^n(t, du)$ *converges to* y *given by* $y(t) = \int_E f(u) z(t, du)$ *in* $D_{\mathbb{R}}[0,\infty)$*.*

Proof. Suppose $t_n \to t$. By Lemma A.4, if $z^n(t_n) \to z(t)$, then $y_n(t_n) \to y(t)$ and if $z^n(t_n) \to z(t-)$, $y_n(t_n) \to y(t-)$. Since $\{z^n\}$ converges in $D_{\mathcal{P}(E)}[0,\infty)$, by Lemma 3.6.5 of [9], $y_n \to y$ in $D_E[0,\infty)$. □

Lemma A.6. *Let* $\{\xi_k\} \subset C_{\mathbb{R}}[0,T]$ *be exchangeable with*

$$\mathbb{E}[\sup_{0\le t\le T} |\xi_k(t)|] < \infty, \tag{A.1}$$

and let $\mathcal{T}_t = \cap_N \sigma(\xi_N(s), \xi_{N+1}(s), \ldots : s \le t)$*. Then setting* $\zeta(t) = \mathbb{E}[\xi_1(t)|\mathcal{T}_t]$*,*

$$\lim_{N\to\infty} \sup_{0\le t\le T} \left| \frac{1}{N} \sum_{k=1}^N \xi_k(t) - \zeta(t) \right|. \quad a.s.$$

Proof. Let $U_k^0(t) = \xi_k([t])$ and

$$U_k^j(t) = \xi_k(\frac{[2^j t]}{2^j}) - \xi_k(\frac{[2^{j-1}t]}{2^{j-1}}).$$

Then

$$\xi_k(t) = \sum_{j=0}^{\infty} U_k^j(t)$$

and

$$\zeta(t) = \lim_{N\to\infty} \frac{1}{N}\sum_{k=1}^{N} \xi_k(t) = \sum_{j=0}^{\infty} \lim_{N\to\infty} \frac{1}{N}\sum_{k=1}^{N} U_k^j(t) \quad \text{for all } t \quad a.s. \tag{A.2}$$

Note that the convergence on the right involves countably many applications of de Finetti's theorem for real exchangeable sequences $\{\{\xi_k(\frac{i}{2^j})\} : i, j = 0, 1, \ldots\}$. The interchange of the limit and the sum follows from (A.1). Since we are assuming all σ-algebras are completed,

$$\mathcal{T}_t = \vee_{\frac{i}{2^j}\le t} \cap_N \sigma(\xi_N(\frac{i}{2^j}), \xi_{N+1}(\frac{i}{2^j}), \ldots).$$

Let $\delta > 0$ and let $\omega_k(\delta)$ be the modulus of continuity of ξ_k, that is,

$$\omega_k(\delta) = \sup_{0\le s,t\le T, |s-t|\le\delta} |\xi_k(t) - \xi_k(s)|.$$

Since the modulus of continuity of an average is less than or equal to the average of moduli of continuity, it follows that

$$\omega_\zeta(\delta) \le \mathbb{E}[\omega_1(\delta)|\mathcal{T}_C] \quad a.s. \tag{A.3}$$

Since it is sufficient to consider a sequence of δ converging to zero, there exists an event of probability one on which (A.2) and (A.3) hold and hence ζ is continuous. □

A.2. Martingales and change of measure. This section follows [27], Section III.8. Let $\{\mathcal{F}_t\}$ be a filtration and assume that $P|_{\mathcal{F}_t} << Q|_{\mathcal{F}_t}$, for all $t \ge 0$, and that $L(t)$ is the corresponding Radon-Nikodym derivative. Then L is an $\{\mathcal{F}_t\}$-martingale on $(\Omega, \mathcal{F}, Q)$.

Lemma A.7. *Z is a P-local martingale if and only if LZ is a Q-local martingale.*

Proof. Note that for a bounded stopping time τ, $Z(\tau)$ is P-integrable if and only if $L(\tau)Z(\tau)$ is Q-integrable. Let σ be any stopping time such that $Z(\sigma\wedge\cdot)$ is a P-martingale. We want to show that $L(\sigma\wedge\cdot)Z(\sigma\wedge\cdot)$ is a Q-martingale The integrability requirement follows from the previous observation taking $\tau = \sigma\wedge t$. For the martingale property, by Bayes formula, $\mathbb{E}^P[Z(\sigma\wedge(t+h)) - Z(\sigma\wedge t)|\mathcal{F}_t] = 0$ if and only if $\mathbb{E}^Q[L(\sigma\wedge(t+h))(Z(\sigma\wedge(t+h)) - Z(\sigma\wedge t))|\mathcal{F}_t] = 0$ which is equivalent to

$$\begin{aligned}\mathbb{E}^Q\left[L(\sigma\wedge(t+h))Z(\sigma\wedge(t+h))|\mathcal{F}_t\right] &= \mathbb{E}^Q\left[L(\sigma\wedge(t+h))Z(\sigma\wedge t)|\mathcal{F}_t\right] \\ &= L(\sigma\wedge t)Z(\sigma\wedge t).\end{aligned}$$

□

Theorem A.8. *If M is a Q-local martingale, then*

$$Z(t) = M(t) - \int_0^t \frac{1}{L(s)} d[L, M]_s \tag{A.4}$$

is a P-local martingale. (Note that the integrand is $\frac{1}{L(s)}$, not $\frac{1}{L(s-)}$.)

Proof. Note that $LM - [L, M]$ is a Q-local martingale. We need to show that LZ is a Q-local martingale. But letting V denote the second term on the right of (A.4), we have

$$L(t)Z(t) = L(t)M(t) - [L, M]_t - \int_0^t V(s-)dL(s),$$

and both terms on the right are Q-local martingales. □

Acknowledgment. Part of this research was funded within the project STORM: Stochastics for Time-Space Risk Models, from the Research Council of Norway (RCN). Project number: 274410.

References

1. Bain, A., and Crisan, D.: *Fundamentals of stochastic filtering*, volume 60 of *Stochastic Modelling and Applied Probability*. Springer, New York, 2009. 1, 1, 4.1, 4.2
2. Brémaud, P.: *Point processes and queues*. Springer-Verlag, New York-Berlin, 1981. Martingale dynamics, Springer Series in Statistics. 2.3.1
3. Clark, J. M. C., and Crisan, D.: On a robust version of the integral representation formula of nonlinear filtering. *Probab. Theory Related Fields*, 133(1): 43–56, 2005. 3
4. Crisan, D., Kurtz, T. G., and Lee, Y.: Conditional distributions, exchangeable particle systems, and stochastic partial differential equations. *Ann. Inst. Henri Poincaré Probab. Stat.*, 50(3):946–974, 2014. 1
5. Crisan, D., Janjigian, C., and Kurtz, T. G.: Particle representations for stochastic partial differential equations with boundary conditions. *Electron. J. Probab.*, 23:Paper No. 65, 29, 2018. 1
6. Crisan, D., and Ortiz-Latorre, S.: A high order time discretization of the solution of the non-linear filtering problem. *Stoch. Partial Differ. Equ. Anal. Comput.*, 8(4):693–760, 2020. 3, 4, 4.2
7. Crisan, D., and Rozovskiĭ, B.: *The Oxford handbook of nonlinear filtering*. Oxford Univ. Press, Oxford, 2011. 1
8. Crisan, D., Lobbe, A., and Ortiz-Latorre, S.: Pathwise approximations for the solution of the non-linear filtering problem, 2021. 3
9. Ethier, S. N., and Kurtz, T. G.: *Markov processes: Characterization and Convergence*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1986. 4, A.1
10. Fujisaki, M., Kallianpur, G., and Kunita, H.: Stochastic differential equations for the non linear filtering problem. *Osaka Math. J.*, 9:19–40, 1972. 1

11. Handschin, J. E., and Mayne, D. Q.: Monte Carlo techniques to estimate the conditional expectation in multi-stage non-linear filtering. *Internat. J. Control (1)*, 9:547–559, 1969. 1
12. Jacod, J., and Shiryaev, A. N.: *Limit theorems for stochastic processes*, volume 288 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, second edition, 2003. 2.3.2
13. Kallenberg, O.: *Foundations of modern probability*. Probability and its Applications (New York). Springer-Verlag, New York, second edition, 2002. 1
14. Kallianpur, G., and Striebel, C.: Estimation of stochastic systems: Arbitrary system process with additive white noise observation errors. *Ann. Math. Statist.*, 39:785–801, 1968. 1, 2.1.3
15. Kotelenez, P. M., and Kurtz, T. G.: Macroscopic limits for stochastic partial differential equations of McKean-Vlasov type. *Probab. Theory Related Fields*, 146(1-2):189–222, 2010. 1, A.1
16. Kunita, H.: Asymptotic behavior of the nonlinear filtering errors of Markov processes. *J. Multivariate Anal.*, 1:365–393, 1971. 1
17. Kunita, H.: Nonlinear filtering for the system with general noise. In *Stochastic control theory and stochastic differential systems (Proc. Workshop, Deutsch. Forschungsgemeinsch., Univ. Bonn, Bad Honnef, 1979)*, volume 16 of *Lecture Notes in Control and Information Sci.*, pages 496–509. Springer, Berlin-New York, 1979. 1
18. Kunita, H.: Cauchy problem for stochastic partial differential equations arising in nonlinear filtering theory. *Systems Control Lett.*, 1(1):37–41, 1981. 1
19. Kunita, H.: Stochastic partial differential equations connected with nonlinear filtering. In *Nonlinear filtering and stochastic control (Cortona, 1981)*, volume 972 of *Lecture Notes in Math.*, pages 100–169. Springer, Berlin, 1982. 1
20. Kunita, H.: Ergodic properties of nonlinear filtering processes. In *Spatial stochastic processes*, volume 19 of *Progr. Probab.*, pages 233–256. Birkhäuser Boston, Boston, MA, 1991. 1
21. Kunita, H.: The stability and approximation problems in nonlinear filtering theory. In *Stochastic analysis*, pages 311–330. Academic Press, Boston, MA, 1991. 1
22. Kunita, H.: Nonlinear filtering problems I: Bayes formulas and innovations. In *The Oxford handbook of nonlinear filtering*, pages 19–54. Oxford Univ. Press, Oxford, 2011. 1
23. Kurtz, T. G., and Protter, P. E.: Weak limit theorems for stochastic integrals and stochastic differential equations. *Ann. Probab.*, 19(3):1035–1070, 1991. 2.1.3, 2.2.3
24. Kurtz, T. G., and Protter, P. E.: Weak convergence of stochastic integrals and differential equations. II. Infinite-dimensional case. In *Probabilistic models for nonlinear partial differential equations (Montecatini Terme, 1995)*, volume 1627 of *Lecture Notes in Math.*, pages 197–285. Springer, Berlin, 1996. 1, 2, 2.2.3
25. Kurtz, T. G., and Xiong, J.: Particle representations for a class of nonlinear SPDEs. *Stochastic Process. Appl.*, 83(1):103–126, 1999. 1, 2.1.3

26. Pagès, G.: *Numerical probability*. Universitext. Springer, Cham, 2018. An introduction with applications to finance. 2.1.1
27. Protter, P. E.: *Stochastic integration and differential equations*, volume 21 of *Applications of Mathematics (New York)*. Stochastic Modelling and Applied Probability. Springer-Verlag, Berlin, second edition, 2004. A.2
28. Wu, Z.: *A filtering approach to abnormal cluster identification*. ProQuest LLC, Ann Arbor, MI, 2007. Thesis (Ph.D.)–The University of Wisconsin - Madison. 2.3.1, 2.5
29. Wu, Z.: A cluster identification framework illustrated by a filtering model for earthquake occurrences. *Bernoulli*, 15(2):357–379, 2009. 2.3.1, 2.5

Dan Crisan: Department of Mathematics, Imperial College London, Huxley's Building,180 Queen's Gate, London SW7 2AZ, United Kingdom
Email address: dcrisan@imperial.ac.uk
URL: https://www.ma.imperial.ac.uk/~dcrisan/

Thomas G. Kurtz: Departments of Mathematics and Statistics, University of Wisconsin-Madison, 48 Lincoln Drive, Madison WI 53706-1388, USA
Email address: kurtz@math.wisc.edu
URL: https://people.math.wisc.edu/~kurtz/

Salvador Ortiz-Latorre: Department of Mathematics, University of Oslo, Niels Henrik Abels hus, Moltke Moes vei 35, 0851 Oslo, Norway
Email address: salvadoo@math.uio.no
URL: https://www.mn.uio.no/math/english/people/aca/salvadoo/