

# On the convergence rates of a general class of weak approximations of SDEs

D. Crisan,<sup>\*</sup> S.Ghazali<sup>†</sup>

## Abstract

In this paper, the convergence analysis of a class of weak approximations of solutions of stochastic differential equations is presented. This class includes recent approximations such as Kusuoka's moment similar families method and the Lyons-Victoir cubature of Wiener Space approach. We show that the rate of convergence depends intrinsically on the smoothness of the chosen test function. For smooth functions (the required degree of smoothness depends on the order of the approximation), an equidistant partition of the time interval on which the approximation is sought is optimal. For functions that are less smooth (for example Lipschitz functions), the rate of convergence decays and the optimal partition is no longer equidistant. Our analysis rests upon Kusuoka-Stroock's results on the smoothness of the distribution of the solution of a stochastic differential equation. Finally the results are applied to the numerical solution of the filtering problem.

## 1 Introduction

Stochastic differential equations (SDEs) constitute an ideal mathematical model for a multitude of phenomena and processes encountered in areas such as filtering, optimal stopping, stochastic control, signal processes and mathematical finance, most notably in option pricing (see for example Oksendal[34] and Kloeden & Platen[16]). Unlike their deterministic counterparts, SDEs do not have explicit solutions, apart from in a few exceptional cases, hence the necessity for a sound theory of their numerical approximation.

In this paper we will be concerned with SDEs written in Stratonovich form, in other words, we will look at equations of the form,

$$X_t = X_0 + \int_0^t V_0(X_s)ds + \sum_{j=1}^k \int_0^t V_j(X_s) \circ dW_s^j, \quad (1)$$

---

<sup>\*</sup>Department of Mathematics, Imperial College London, 180 Queen's Gate, London SW7 2BZ. E-mail: d.crisan@imperial.ac.uk.

<sup>†</sup>Department of Mathematics, Imperial College London, 180 Queen's Gate, London SW7 2BZ. E-mail: saadia.ghazali@imperial.ac.uk.

where the last term is a stochastic integral of Stratonovich type. There are two classes of numerical methods for approximating SDEs. The objective of the first is to produce a pathwise approximation of the solution (strong approximation). The second method involves approximating the distribution of the solution at a particular instance in time (weak approximation). For example when one is only interested in the expectation  $\mathbb{E}[\varphi(X_t)]$  for some function  $\varphi$ , it is sufficient to have a good approximation of the distribution of the random variable  $X_t$  rather than of its sample paths. This observation was first made by Milstein [27] who showed that pathwise schemes and  $L^2$  estimates of the corresponding errors are irrelevant in this context since the objective is to approximate the law of  $X_t$ . This paper contains approximations that belong to this second class of algorithms.

Classical results in this area concentrate on solving numerically SDEs for which the so-called ‘ellipticity condition’, or more generally the ‘Uniform Hörmander condition’ (UH), is satisfied. For a survey of such schemes see, for example, Kloeden & Platen[16] or Burrage, Burrage & Tian [5]. Under this condition, for any bounded measurable function  $\varphi$ , the semigroup of operators  $\{P_t\}_{t \in [0, \infty)}$  defined,

$$(P_t \varphi)(x) = \mathbb{E}[\varphi(X_t(x))], \quad (2)$$

where  $X(x) = \{X_t(x)\}_{t \in [0, \infty)}$  solves (1) with initial condition  $X_0 = x$ , is smooth for any  $t > 0$ . It is this property upon which the majority of these schemes rely.

For example, the classical Euler-Maruyama scheme requires  $P_t \varphi$  to be four times differentiable in order to obtain the optimal rate of convergence. Talay([42],[43]) and, independently, Milstein[28] introduced the appropriate methodology to analyse this scheme. They express the error as a difference including a sum of terms involving the solution of a parabolic PDE. Their analysis also shows the relationship between the smoothness of  $\varphi$  and the corresponding error. Talay & Tubaro[44] prove an even more precise result showing that, under the same conditions, the errors corresponding to the Euler-Maruyama and many other schemes can be expanded in terms of powers of the discretisation step. Furthermore, Bally & Talay[1] show the existence of such an expansion under a much weaker hypothesis on  $\varphi$ : that  $\varphi$  need only be measurable and bounded (even the boundedness condition can be relaxed). Higher order schemes require additional smoothness properties of  $P_t \varphi$  (see for example, Platen & Wagner[37]).

In the eighties, Kusuoka & Stroock([21], [22], [23]) studied the properties of  $P_t \varphi$  under a weaker condition, the so-called UFG condition (see (5) in Section 2). Essentially, this condition states that the Lie algebra generated by the vector fields  $\{V_i\}_{i=1}^k \in C_b^\infty(\mathbb{R}^d; \mathbb{R}^d)$  is finite dimensional as a  $C_b^\infty(\mathbb{R}^d)$ -module. Kusuoka & Stroock conclude Malliavin’s undertaking to recover Hörmander’s hypo-ellipticity theory for degenerate second order elliptic operators. They show that under the UFG condition,  $P_t \varphi$  retains certain regularity properties, in particular, that  $V_{i_1} V_{i_2} \dots V_{i_n} P_t \varphi$  is well defined for any vector fields  $V_{i_r}$ , where  $i_r \in \{1, \dots, k\}$ . They also give upper bounds for the supremum norm of  $V_{i_1} V_{i_2} \dots V_{i_n} P_t \varphi$ . Besides the UFG condition, our analysis rests upon a second assumption which we call the **V0** condition. It states that  $V_0$  can be expressed in terms of

$\{V_1, \dots, V_k\} \cup \{[V_i, V_j], 1 \leq i < j \leq k\}$ . This premise is weaker than the ellipticity assumption and has been used, for example, by Jerison and Sánchez-Calle ([14],[38]) to obtain estimates for the heat kernel. This second condition enables one to control the supremum norm of  $V_{i_1} V_{i_2} \dots V_{i_n} P_t \varphi$  for  $i_r \in \{0, 1, \dots, k\}$  (see Corollary 18 in the appendix).

A number of schemes have recently been developed to work under these weaker conditions rather than the ellipticity condition, their convergence depending intrinsically on the above estimates of  $V_{i_1} V_{i_2} \dots V_{i_n} P_t \varphi$ . A further advantage of this new generation of schemes is a consequence of the classical result stating that the support of  $X(x)$  is the closure of the set  $S = \{x^\varphi : [0, \infty) \rightarrow \mathbb{R}^d\}$  where  $x^\varphi$  solves the ODE,

$$x_t^\varphi = x + \int_0^t V_0(x_s^\varphi) ds + \sum_{j=1}^k \int_0^t V_j(x_s^\varphi) \varphi(s) \circ ds,$$

and  $\varphi : [0, \infty) \rightarrow \mathbb{R}^d$  is an arbitrary smooth function (see Stroock & Varadhan [39], [40], [41], Millet & Sanz-Sole[26]). These schemes attempt to keep the support of the approximating process on the set  $S$ . In this way, stability problems that are known to affect classical schemes can be avoided. For example, Ninomiya & Victoir[33] give an explicit example where the Euler-Maruyama approximation fails whilst their algorithm succeeds (see Example 5 below for the algorithm). Their example involves an SDE related to the Heston stochastic volatility model in finance.

In this paper we give a general criterion for the convergence of a class of weak approximations incorporating this new category of schemes. This criterion is based upon the stochastic Stratonovich-Taylor expansion of  $P_t \varphi$  and demonstrates how the rate of convergence depends on the smoothness of the test function selected.

Our plan is as follows. In section 2 we set out some essential terminology, adopted partly from Kusuoka[17], which is required to index the stochastic Stratonovich-Taylor expansions that follow and to state the UFG and V0 conditions imposed on the given vector fields (see (5) & (6) below). Section 3 contains our main results on the convergence analysis of a class of weak approximations of solutions of SDEs, characterised by introducing the concept of an *m-perfect family* (Definition 2). The Lyons-Victoir and Ninomiya-Victoir approximations are both members of this class. Although the Kusuoka approximation is not within this family, it can effectively be categorised in the same way (see Example 7 and the subsequent comment). In our main Theorem 8, we show that the rate of convergence depends intrinsically on the smoothness of the chosen test function; the higher the order of the approximation, the smoother the test function required.

For a smooth function, an equidistant partition of the time interval on which the approximation is sought is optimal. For less smooth functions, this is no longer true. We emphasise that the UFG+V0 conditions are not required for a smooth test function. Furthermore, the Kusuoka approximation does not require the V0 condition. Finally, in Section 4 we present an application of this theory to the numerical solution of the filtering problem.

## 2 Preliminaries

Let  $(\Omega, \mathcal{F}, \mathbb{P})$  be a probability space satisfying the usual conditions and  $W = \{W_t\}_{t \in [0, \infty)}$  be a  $k$ -dimensional Brownian motion defined on it. We also set  $W_t^0 = t$  for  $t \in [0, \infty)$ . Let  $C_b^\infty(\mathbb{R}^d, \mathbb{R}^d)$  denote the space of smooth functions  $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}^d$  with bounded derivatives, that is, bounded partial derivatives (of all orders) of the  $d$  component functions  $\{\varphi^i : \mathbb{R}^d \rightarrow \mathbb{R}\}_{i=1}^d$  exist. We have a dual interpretation of this space, in that we also regard its elements as vector fields. In other words, a smooth function  $V : C_b^\infty(\mathbb{R}^d) \rightarrow C_b^\infty(\mathbb{R}^d)$  is equivalent to an operator on  $C_b^\infty(\mathbb{R}^d) \equiv C_b^\infty(\mathbb{R}^d, \mathbb{R})$  defined by,

$$V\varphi = \sum_{i=1}^d V^i \frac{\partial \varphi}{\partial x_i}$$

where  $V^i \in C_b^\infty(\mathbb{R}^d)$  is the  $i$ -th component of  $L$  for  $i = 1, \dots, d$ .

We next consider the Stratonovich SDE with drift vector  $\{V_0^i(x)\}_{i=1}^d$  and dispersion matrix  $\{V_j^i(x)\}_{i,j=1}^{d,k}$  for  $x \in \mathbb{R}^d$ , for some  $V_0, \dots, V_k \in C_b^\infty(\mathbb{R}^d, \mathbb{R}^d)$ . This is written componentwise as,

$$X_t^i = X_0^i + \int_0^t V_0^i(X_s) ds + \sum_{j=1}^k \int_0^t V_j^i(X_s) \circ dW_s^j \quad (3)$$

for  $i = 1, \dots, d$ .

The following essential terminology has been adopted from Kusuoka[17]. Let  $\mathcal{A}$  be the set of multi-indices,

$$\mathcal{A} = \{\emptyset\} \cup \bigcup_{m=1}^{\infty} \{0, 1, \dots, k\}^m$$

where  $k$  is the dimension of the Brownian Motion introduced above. This set will be used to index the Stratonovich-Taylor expansions that follow. Let  $|\cdot|$  and  $\|\cdot\|$  be the following two norms defined on  $\mathcal{A}$  by

$$|\emptyset| = 0, \quad |\alpha| = r \text{ if } \alpha = (i_1, \dots, i_r) \in \{0, 1, \dots, k\}^r \text{ for } r \in \mathbb{N}$$

and  $\|\alpha\| = |\alpha| + \text{card}\{1 \leq j \leq |\alpha| : i_j = 0\}$ . Furthermore, let  $\mathcal{A}_0 = \mathcal{A} \setminus \{\emptyset\}$ ,  $\mathcal{A}_1 = \mathcal{A} \setminus \{\emptyset, (0)\}$  and correspondingly,  $\mathcal{A}(m) = \{\alpha \in \mathcal{A} : \|\alpha\| \leq m\}$ ,  $\mathcal{A}_0(m) = \{\alpha \in \mathcal{A}_0 : \|\alpha\| \leq m\}$  and  $\mathcal{A}_1(m) = \{\alpha \in \mathcal{A}_1 : \|\alpha\| \leq m\}$  where the integer  $m \in \mathbb{N}$  above corresponds to the level of the truncation in the Stratonovich-Taylor expansions that follow (see (12)).

We also require the following concatenation on  $\mathcal{A}$ ,

$$(i_1, \dots, i_r) * (j_1, \dots, j_s) = (i_1, \dots, i_r, j_1, \dots, j_s)$$

and need to define a further operation on the vector fields  $\{V_j \in C_b^\infty(\mathbb{R}^d, \mathbb{R}^d)\}_{j=0}^k$  introduced above.

**Definition 1** For  $\alpha \in \mathcal{A}$ , the vector field  $V_{[\alpha]}$  is defined inductively by,

$$V_{[\phi]} = 0; \quad V_{[(j)]} = V_j; \quad V_{[\alpha*(j)]} = [V_\alpha, V_j]$$

for  $j = 0, 1, \dots, k$  where  $[V_i, V_j] := V_i V_j - V_j V_i$  for  $V_i, V_j \in C_b^\infty(\mathbb{R}^d, \mathbb{R}^d)$ .

In the following we will make use of the semi-norm,

$$\|\varphi\|_{V,i} = \sum_{u=1}^i \sum_{\substack{\alpha_1, \dots, \alpha_u \in \mathcal{A}_0 \\ \|\alpha_1 * \dots * \alpha_u\| = i}} \|V_{[\alpha_1]} \cdots V_{[\alpha_u]} \varphi\|_\infty.$$

for  $i \in \mathbb{N}$ . Furthermore, we introduce the semi-norm,

$$\|\varphi\|_p = \sum_{i=1}^p \|\nabla^i \varphi\|_\infty \tag{4}$$

for  $p \in \mathbb{N}$ ,  $\varphi \in C_b^p(\mathbb{R}^d)$  where,

$$\|\nabla^i \varphi\|_\infty = \max_{j_1, \dots, j_i \in \{1, \dots, d\}} \left\| \frac{\partial^i \varphi}{\partial x_{j_1} \cdots \partial x_{j_i}} \right\|_\infty$$

and note that it can easily be deduced from the chain rule, for  $\alpha \in \mathcal{A}_0$ ,  $\varphi \in C_b^{|\alpha|}(\mathbb{R}^d)$ , that  $\|V_\alpha \varphi\|_\infty \leq C \|\varphi\|_{|\alpha|}$  and hence  $\|\varphi\|_{V,i} \leq C \|\varphi\|_i$  for  $i \in \mathbb{N}$ . Also let  $\|\cdot\|_{p,\infty}$  be the norm  $\|\varphi\|_{p,\infty} := \|\varphi\|_\infty + \|\varphi\|_p$  for  $\varphi \in C_b^p(\mathbb{R}^d)$ . We define the space,

$$C_b^{V,i}(\mathbb{R}^d) = \left\{ \varphi : \|\varphi\|_{V,i} < \infty \right\}$$

which appears in the definition of an *m-perfect* family below (see Definition 2).

We now introduce the two conditions on the vector fields  $\{V_j \in C_b^\infty(\mathbb{R}^d, \mathbb{R}^d)\}_{j=0}^k$  required to treat the case when the test function  $\varphi$  is not smooth. We emphasise that the smooth case requires neither the UFG nor the V0 condition. Furthermore, the Kusuoka approximation (Example 7) does not require the V0 condition as the corresponding error is controlled in terms of  $\{V_{[\beta]} : \beta \in \mathcal{A}_1\}$  alone.

**The UFG condition** [Kusuoka & Stroock [17], [23]]. There exists some  $l \in \mathbb{N}$  such that for any  $\alpha \in \mathcal{A}_1$  we have,

$$V_{[\alpha]} = \sum_{\beta \in \mathcal{A}_1(l)} \varphi_{\alpha, \beta} V_{[\beta]}. \quad (5)$$

where  $\varphi_{\alpha, \beta} \in C_b^\infty(\mathbb{R}^d)$  for all  $\beta \in \mathcal{A}_1(l)$ .

**The V0 condition.** There exist  $\varphi_\beta \in C_b^\infty(\mathbb{R}^d)$ ,  $\beta \in \mathcal{A}_1(2)$  such that,

$$V_0 = \sum_{\beta \in \mathcal{A}_1(2)} \varphi_\beta V_{[\beta]}. \quad (6)$$

As mentioned in the introduction, the UFG condition states that the Lie algebra generated by the vector fields  $\{V_i\}_{i=1}^k \in C_b^\infty(\mathbb{R}^d; \mathbb{R}^d)$  is finite dimensional as a  $C_b^\infty(\mathbb{R}^d)$ -module. It is implied by the Uniform Hörmander Condition which states that,

$$\exists l \in \mathbb{N} \text{ and } c > 0 \text{ s.t. } \sum_{\alpha \in \mathcal{A}_1(l)} \langle V_{[\alpha]}(x), \xi \rangle^2 \geq c |\xi|^2 \quad (7)$$

for all  $x, \xi \in \mathbb{R}^d$ , where  $\langle V, \xi \rangle = \sum_{i=1}^d V^i \xi^i$  for  $V \in C_b^\infty(\mathbb{R}^d, \mathbb{R}^d)$ .

Under conditions (5)+(6), one can show that for any  $r \in \mathbb{Z}^+$ ,  $\{\alpha_i \in \mathcal{A}_0\}_{i=1}^r$  and  $p = 1, \dots, \|\alpha_1 * \dots * \alpha_r\|$ , there exists a constant  $C_p^T > 0$  such that,

$$\|V_{[\alpha_1]} \dots V_{[\alpha_r]} P_t \varphi\|_\infty \leq C_p^T t^{(p - \|\alpha_1 * \dots * \alpha_r\|)/2} \|\varphi\|_p, \quad t \in [0, T]. \quad (8)$$

where  $\{P_t\}_{t \in [0, \infty)}$  is the semigroup associated with  $X$  defined in (2). We remark that condition (5) will only give us (8) for  $\alpha_i \in \mathcal{A}_1$ ,  $i = 1, \dots, r$ . Under both (5)+(6) we have,

$$\|V_0 P_t \varphi\|_\infty \leq C t^{-1} \|\varphi\|_\infty.$$

However, under condition (5) alone,  $\|V_0 P_t \varphi\|_\infty$  may be of higher order. Kusuoka has given an explicit example in which,

$$c t^{-\frac{1}{2}} \|\varphi\|_\infty \leq \|V_0 P_t \varphi\|_\infty \leq C t^{-\frac{1}{2}} \|\varphi\|_\infty$$

for some constants  $c, C > 0$  and where  $l$  is the constant appearing in (5) (see Proposition 14 and Proposition 16 in [19]). This coarser bound on  $\|V_0 P_t \varphi\|_\infty$  results in lower rates of convergence. The authors believe that (6) is the most general condition required to preserve the same rates of convergence as those obtained when  $\{V_i\}_{i=1}^k$  satisfy the ellipticity condition (for non-smooth test functions  $\varphi$ ).

Result (8) is a corollary of a certain representation theorem proved in Kusuoka-Stroock[23]. For completeness, we state the representation theorem (Theorem 16) in the appendix where we also sketch a proof of inequality (8) in Corollary 18. We note that the case  $p = 1$  has been proved in [23] for  $\{\alpha_i \in \mathcal{A}_1\}_{i=1}^r$ . Inequality (8) proves to be crucial in obtaining upper bounds on the error of the class algorithms that we study below (see Theorem 8).

### 3 The Main Theorem

In this section we introduce the concept of an  $m$ -perfect family. Such families correspond to various weak approximations of SDEs, including the Lyons-Victoir and Ninomiya-Victoir schemes. The main result appears in Theorem 8 and Corollary 9.

For  $\alpha = (i_1, \dots, i_r) \in \mathcal{A}_0$  and  $\varphi \in C_b^r(\mathbb{R}^d)$  let  $f_{\alpha, \varphi}$  be defined as  $f_{(i_1, \dots, i_r), \varphi} := V_{i_1} \dots V_{i_r} \varphi$ . We also need to define the iterated Stratonovich integral

$$I_{f_{\alpha, \varphi}}(t) := \int_0^t \int_0^{s_0} \dots \left( \int_0^{s_{r-2}} f_{\alpha, \varphi}(X_{s_{r-1}}) \circ dW_{s_{r-1}}^{i_1} \right) \circ \dots \circ dW_{s_1}^{i_{r-1}} \circ dW_{s_0}^{i_r},$$

for  $t \geq 0$ . If  $i_1 = 0$  then  $I_{f_{\alpha, \varphi}}(t)$  is well defined for  $\varphi \in C_b^r(\mathbb{R}^d)$ . However, if  $i_1 \neq 0$  then  $I_{f_{\alpha, \varphi}}(t)$  is well defined provided  $\varphi \in C_b^{r+2}(\mathbb{R}^d)$ , since the semimartingale property of  $f_{\alpha, \varphi}(X)$  is required in the definition of the first Stratonovich integral  $\int_0^{s_{r-2}} f_{\alpha, \varphi}(X_{s_{r-1}}) \circ dW_{s_{r-1}}^{i_1}$ . Note that the Stratonovich integrals are evaluated innermost first. Finally let

$$I_{\alpha}(t) := \int_0^t \int_0^{s_0} \dots \left( \int_0^{s_{r-2}} 1 \circ dW_{s_{r-1}}^{i_1} \right) \circ \dots \circ dW_{s_1}^{i_{r-1}} \circ dW_{s_0}^{i_r}.$$

Let  $\alpha = (i_1, \dots, i_r) \in \mathcal{A}_0$  be an arbitrary multi-index such that  $\|\alpha\| = m \in \mathbb{N}$  (and  $|\alpha| = r \in \mathbb{N}$ ). If  $m$  is odd, then  $\mathbb{E}[I_{\alpha}(t)] = 0$  and if  $m$  is even then

$$\mathbb{E}[I_{\alpha}(t)] = \begin{cases} \frac{t^{\frac{m}{2}}}{2^{r-\frac{m}{2}} (\frac{m}{2})!} & \text{if } \alpha \in \mathcal{A}_0^{m,r} \\ 0 & \text{otherwise} \end{cases}, \quad (9)$$

where  $\mathcal{A}_0^{m,r}$  is the set of multi-indices  $\alpha = \alpha_1 * \dots * \alpha_{\frac{m}{2}} \in \mathcal{A}_0(m)$  such that each  $\alpha_i = (0)$  or  $(j, j)$  for some  $j \in \{1, \dots, k\}$ . Note that  $r - \frac{m}{2}$  is equal to the number of pairs of indices  $(j, j)$  occurring in  $\alpha$ . A proof of this result can be found in [12].

The set of iterated Stratonovich integrals plays a central role in the theory of approximation of solutions of SDEs and there are numerous papers that study its structure. Here, we adopt the hierarchical set approach introduced by Kloeden & Platen[15]. An alternative method can be found in Gaines[11] where it is shown how Lyndon words provide a basis for iterated Stratonovich integrals and also how shuffle products may be used to obtain moments of stochastic integrals. Pettersson[35] gives a notationally and computationally convenient Stratonovich-Taylor expansion. Furthermore, Burrage & Burrage[3] use rooted-tree theory to describe the aforementioned set and Burrage & Burrage [4] presents an approach based on B-series.

We state three further results in (10), (11) and (13). The proofs are all elementary and can be found in [12]. The first two give an upper bound on the  $L^2$  norm of  $I_{f_{\alpha, \varphi}}(t)$  for smooth  $\varphi$ . The third provides an explicit form for the remainder of  $\varphi(X_t)$  when expanded in terms of iterated integrals.

For  $\varphi \in C_b^{|\alpha|+2}(\mathbb{R}^d)$  and any multi-index  $\alpha = (i_1, \dots, i_r) \in \mathcal{A}_0$  such that  $i_1 \neq 0$ , we have

$$\|I_{f_{\alpha,\varphi}}(t)\|_2 \leq c_1 \|f_{\alpha,\varphi}\|_\infty t^{\frac{\|\alpha\|}{2}} + c_2 \sum_{i=1}^k \|V_i f_{\alpha,\varphi}\|_\infty t^{\frac{\|\alpha\|+1}{2}} \quad (10)$$

for some constants  $c_1 \equiv c_1(\alpha), c_2 \equiv c_2(\alpha) \geq 0$ . For  $\varphi \in C_b^{|\alpha|}(\mathbb{R}^d)$  and any multi-index  $\alpha = (i_1, \dots, i_r) \in \mathcal{A}_0$  such that  $i_1 = 0$  we have

$$\|I_{f_{\alpha,\varphi}}(t)\|_2 \leq c_1 \|f_{\alpha,\varphi}\|_\infty t^{\frac{\|\alpha\|}{2}}. \quad (11)$$

For  $m \in \mathbb{N}$ ,  $\varphi \in C_b^{m+3}(\mathbb{R}^d)$  and  $x \in \mathbb{R}^d$ , we define the truncation,

$$\varphi_t^m(x) := \varphi(x) + \sum_{\alpha \in \mathcal{A}_0(m)} f_{\alpha,\varphi}(x) I_\alpha(t). \quad (12)$$

Then for  $t \geq 0$  the remainder is

$$R_{m,t,\varphi}(x) := \varphi(X_t) - \varphi_t^m(x) = \left( \sum_{\|\alpha\|=m+1} + \sum_{\|\alpha\|=m+2, \alpha=0*\beta, \|\beta\|=m} \right) I_{f_{\alpha,\varphi}}(t). \quad (13)$$

In the following, we define a class of approximations of  $X$  expressed in terms of certain families of stochastic processes,  $\bar{X}(x) = \{\bar{X}_t(x)\}_{t \in [0, \infty)}$  for  $x \in \mathbb{R}^d$ , which are explicitly solvable. In particular, we can explicitly compute the operator,

$$(Q_t \varphi)(x) = \mathbb{E}[\varphi(\bar{X}_t(x))]. \quad (14)$$

The semigroup  $P_T$  will then be approximated by  $Q_{h_n}^m Q_{h_{n-1}}^m \dots Q_{h_1}^m$  where  $\{h_j := t_j - t_{j-1}\}_{j=1}^n$  and  $\pi_n = \{t_j := (\frac{j}{n})^\gamma T\}_{j=0}^n$  for  $n \in \mathbb{N}$ , is a sufficiently fine partition of the interval  $[0, T]$ . In particular  $h_j \in [0, 1)$  for  $j = 1, \dots, n$ . The underlying idea is that  $Q_t \varphi$  will have the same truncation as  $P_t \varphi$ .

So let  $\bar{X}(x) = \{\bar{X}_t(x)\}_{t \in [0, \infty)}$ , where  $x \in \mathbb{R}^d$ , be a family of progressively measurable stochastic processes such that,  $\lim_{y \rightarrow x_0} \bar{X}_{t_0}(y) = \bar{X}_{t_0}(x_0)$   $\mathbb{P}$ -almost surely, for any  $t_0 \geq 0$  and  $x_0 \in \mathbb{R}^d$ . As a result, the operator  $Q_t$  defined in (14) has the property that  $Q_t \varphi \in C_b(\mathbb{R}^d)$  for any  $\varphi \in C_b(\mathbb{R}^d)$ . In particular,  $Q_t : C_b(\mathbb{R}^d) \rightarrow C_b(\mathbb{R}^d)$  is a Markov operator.

**Definition 2** For  $m \in \mathbb{N}$ , the family  $\bar{X}(x) = \{\bar{X}_t(x)\}_{t \in [0, \infty)}$  where  $x \in \mathbb{R}^d$ , is said to be  **$m$ -perfect** for the process  $X$  if there exist constants  $C > 0$  and  $M \geq m + 1$  such that for  $\varphi \in C_b^{V,M}(\mathbb{R}^d)$ ,

$$\sup_{x \in \mathbb{R}^d} |Q_t \varphi(x) - \mathbb{E}[\varphi_t^m(x)]| \leq C \sum_{i=m+1}^M t^{i/2} \|\varphi\|_{V,i}. \quad (15)$$



As we can see from (15), the quantity  $\mathbb{E}[\varphi_t^m(x)]$  plays the same role as the classical truncation in the standard Taylor expansion of a function. Using (9) we deduce that,

$$\begin{aligned}
\mathbb{E}[\varphi_t^0(x)] &= \varphi(x) \\
\mathbb{E}[\varphi_t^2(x)] &= \varphi(x) + V_0\varphi(x)t + \sum_{i=1}^k V_i^2\varphi(x)\frac{t}{2} = \varphi(x) + L\varphi(x)t \\
\mathbb{E}[\varphi_t^4(x)] &= \mathbb{E}[\varphi_t^2(x)] + \sum_{i=1}^k V_0^2\varphi(x)\frac{t^2}{2} + \sum_{i=1}^k V_0V_i^2\varphi(x)\frac{t^2}{4} \\
&\quad + \sum_{i=1}^k V_i^2V_0\varphi(x)\frac{t^2}{4} + \sum_{i,j=1}^k V_j^2V_i^2\varphi(x)\frac{t^2}{8} \\
&= \varphi(x) + L\varphi(x)t + L^2\varphi(x)\frac{t^2}{2},
\end{aligned}$$

where  $L = V_0 + \frac{1}{2} \sum_{i=1}^k V_i^2$ . Furthermore, since  $\mathbb{E}[I_\alpha(t)] = 0$  for odd  $\|\alpha\|$ , it follows that  $\mathbb{E}[\varphi_t^1(x)] = \mathbb{E}[\varphi_t^0(x)]$ ,  $\mathbb{E}[\varphi_t^3(x)] = \mathbb{E}[\varphi_t^2(x)]$  and  $\mathbb{E}[\varphi_t^5(x)] = \mathbb{E}[\varphi_t^4(x)]$ .

There now follow some examples of  $m$ -perfect families corresponding to  $\{P_t\}_{t \in [0, \infty)}$  as described in (2), the Lyons-Victoir method and the Ninomiya-Victoir algorithm.

**Example 3** *The family of stochastic processes  $\{X_t(x)\}_{t \in [0, \infty)}$ , where  $x \in \mathbb{R}^d$ , is  $m$ -perfect.*

*More precisely there exists a constant  $c_3 > 0$  such that for  $\varphi \in C_b^{V, m+2}(\mathbb{R}^d)$ ,*

$$\sup_x |P_t\varphi(x) - \mathbb{E}[\varphi_t^m(x)]| \leq c_3 \sum_{i=m+1}^{m+2} t^{i/2} \|\varphi\|_{V,i}, \quad (16)$$

**Proof.** For  $\varphi \in C_b^{V, m+3}(\mathbb{R}^d)$ ,

$$|P_t\varphi(x) - \mathbb{E}[\varphi_t^m(x)]| = |\mathbb{E}[R_{m,t,\varphi}(x)]| = \left| \mathbb{E} \left[ \left( \sum_{\|\alpha\|=m+1} + \sum_{\|\alpha\|=m+2, \alpha=0*\beta, \|\beta\|=m} \right) I_{f_{\alpha,\varphi}}(t) \right] \right|$$

Applying inequality (10) to the first sum,

$$\begin{aligned}
\sum_{\|\alpha\|=m+1} \|I_{f_{\alpha,\varphi}}(t)\|_2 &\leq \sum_{\|\alpha\|=m+1} \{c_1(\alpha) \|f_{\alpha,\varphi}\|_\infty t^{\frac{m+1}{2}} + c_2(\alpha) \sum_{i=1}^k \|V_i f_{\alpha,\varphi}\|_\infty t^{\frac{m+2}{2}}\} \\
&\leq c_4 \sum_{i=m+1}^{m+2} t^{i/2} \|\varphi\|_{V,i}
\end{aligned} \quad (17)$$

for some constant  $c_4 > 0$ . Applying result (11) to the second sum,

$$\begin{aligned} \sum_{\|\alpha\|=m+2, \alpha=0*\beta, \|\beta\|=m} \|I_{f_{\alpha, \varphi}}(t)\|_2 &\leq \sum_{\|\alpha\|=m+2, \alpha=0*\beta, \|\beta\|=m} c_1(\alpha) \|f_{\alpha, \psi}\|_{\infty} t \\ &\leq c_5 \|\varphi\|_{V, m+2} t^{\frac{m+2}{2}} \end{aligned} \quad (18)$$

for some  $c_5 > 0$ . The result for  $\varphi \in C_b^{V, m+3}(\mathbb{R}^d)$  follows from combining (17) and (18). Since none of the terms in (16) depend on partial derivatives of order  $m+3$ , the inequality is also valid for any  $\varphi \in C_b^{V, m+2}(\mathbb{R}^d)$  (a standard approximation method can be used). ■

In the following example, the family of processes  $\bar{X}(x) = \{\bar{X}_t(x)\}_{t \in [0, 1]}$ , where  $x \in \mathbb{R}^d$ , corresponds to the Lyons-Victoir approximation (see [24]). The example involves a set of  $l$  finite variation paths,  $\omega_1, \dots, \omega_l \in C_0^0([0, 1], \mathbb{R}^k)$ , for some  $l \in \mathbb{N}$ , together with some weights  $\lambda_1, \dots, \lambda_l \in \mathbb{R}^+$  such that  $\sum_{j=1}^l \lambda_j = 1$ . These paths are said to define a **cubature formula on Wiener Space of degree  $m$**  if, for any  $\alpha \in \mathcal{A}_0(m)$ ,

$$\mathbb{E}[I_{\alpha}(1)] = \sum_{j=1}^l \lambda_j I_{\alpha}^{\omega_j}(1)$$

where,

$$I_{(i_1, \dots, i_r)}^{\omega_j}(1) := \int_0^1 \int_0^{s_0} \cdots \left( \int_0^{s_{r-2}} d\omega_j^{i_1}(s_{r-1}) \right) \cdots d\omega_j^{i_{r-1}}(s_1) d\omega_j^{i_r}(s_0).$$

From the scaling properties of the Brownian motion we can deduce, for  $t \geq 0$ ,

$$\mathbb{E}[I_{\alpha}(t)] = \sum_{j=1}^l \lambda_j I_{\alpha}^{\omega_{t,j}}(t)$$

where  $\omega_{t,1}, \dots, \omega_{t,l} \in C_0^0([0, t], \mathbb{R}^k)$  is defined by  $\omega_{t,j}(s) = \sqrt{t} \omega_j(\frac{s}{t})$ ,  $s \in [0, t]$ . In other words, the expectation of the iterated Stratonovich integrals  $I_{\alpha}(t)$  is the same under the Wiener measure as it is under the measure,

$$\mathbb{Q}_t := \sum_{j=1}^l \lambda_j \delta_{\omega_{t,j}}.$$

**Example 4** *If we choose  $\bar{X}$  to satisfy the evolution equation (3) but with the driving Brownian motion replaced by the paths  $\omega_{t,1}, \dots, \omega_{t,l}$  defined above then the family of*

processes,  $\{\bar{X}_t(x)\}_{t \in [0,1]}$ , with corresponding operator  $(\bar{Q}_t \varphi)(x) := \mathbb{E}_{\mathbb{Q}_t}[\varphi(\bar{X}_t(x))]$ , is  $m$ -perfect. More precisely, there exists a constant  $c_6 > 0$  such that for  $\varphi \in C_b^{V,m+2}(\mathbb{R}^d)$ ,

$$\sup_x |\bar{Q}_t \varphi(x) - \mathbb{E}[\varphi_t^m(x)]| \leq c_6 \sum_{i=m+1}^{m+2} t^{i/2} \|\varphi\|_{V,i}$$

For example, if  $(\lambda_j, \omega_{t,j})$  are chosen such that for  $l = 2^k$  the paths are  $\omega_{t,j} : t \mapsto t(1, z_j^1, \dots, z_j^k)$  for  $j = 1, \dots, 2^k$  with points  $z_j \in \{-1, 1\}^k$  and weights  $\lambda_j = 2^{-k}$ , we obtain a cubature formula of degree 3 and a corresponding 3-perfect family.

**Proof.** Let us first observe that  $I_\alpha^{\omega_{t,j}}(t) = t^{\frac{|\alpha|}{2}} I_\alpha^{\omega_j}(1)$ . Hence, for  $\varphi \in C_b^{V,m+2}(\mathbb{R}^d)$ ,

$$\begin{aligned} |\bar{Q}_t \varphi(x) - \mathbb{E}[\varphi_t^m(x)]| &= |\mathbb{E}_{\mathbb{Q}_t}[R_{m,t,\varphi}(x)]| \\ &\leq \left( \sum_{\|\alpha\|=m+1} + \sum_{\|\alpha\|=m+2, \alpha=0*\beta, \|\beta\|=m} \right) \|f_{\alpha,\varphi}\|_\infty \|\mathbb{E}_{\mathbb{Q}_t}[I_\alpha(t)]\|_2 \\ &\leq \left( \sum_{\|\alpha\|=m+1} + \sum_{\|\alpha\|=m+2, \alpha=0*\beta, \|\beta\|=m} \right) \|f_{\alpha,\varphi}\|_\infty \sum_{j=1}^l \lambda_j \|I_\alpha^{\omega_{t,j}}(t)\|_2 \\ &\leq \left( \sum_{\|\alpha\|=m+1} + \sum_{\|\alpha\|=m+2, \alpha=0*\beta, \|\beta\|=m} \right) k_\alpha \|f_{\alpha,\varphi}\|_\infty t^{\frac{|\alpha|}{2}}. \end{aligned}$$

where  $k_\alpha = \sum_{j=1}^l \lambda_j \|I_\alpha^{\omega_j}(1)\|_2$ . ■

### Remarks

(i) There has been no change to the underlying measure in the example above. Merely a representation in terms of the measure  $\mathbb{Q}_t$  has been introduced to ease the computation of  $\bar{Q}_t$ . More precisely, the family of processes  $\{\bar{X}_t(x)\}_{t \in [0,1]}$  where  $x \in \mathbb{R}^d$  is constructed as follows. We take,

$$\bar{X}_0(x) = x$$

and then randomly choose a path  $\omega_{t,r}$  from the set  $\{\omega_{t,1}, \dots, \omega_{t,l}\}$  with corresponding probabilities  $(\lambda_1, \dots, \lambda_l)$ . Each process then follows a deterministic trajectory driven by the solution of the ordinary differential equation,

$$d\bar{X}_t = V_0(\bar{X}_t)dt + \sum_{j=1}^k V_j(\bar{X}_t)d\omega_{t,k}^j$$

for some  $V_0, \dots, V_k \in C_b^\infty(\mathbb{R}^d, \mathbb{R}^d)$  as in (3). We can therefore compute the expected values of a functional of  $X_t(x)$  as integrals on the path space with respect to the Radon measure  $\mathbb{Q}_t$ . Hence the identities,

$$\overline{Q}_t \varphi(x) = \mathbb{E} [\varphi(\overline{X}_t(x))] = \mathbb{E}_{\mathbb{Q}_t} [\varphi(\overline{X}_t(x))]$$

(ii) The approach adopted by Lyons and Victoir to construct the above approximation resembles the ideas developed by Clark and Newton in a series of papers ([8],[9],[29],[30]). Heuristically, Clark and Newton constructed strong approximations of SDEs using flows driven by vector fields which were measurable with respect to the filtration generated by the driving Wiener process. In a similar vein, Castell & Gaines[7] provide a method of strongly approximating the solution of an SDE by means of exponential Lie series.

For the following example, we will denote by  $\exp(Vt)f$  the value at time  $t$  of the solution of the ODE  $y' = V(y)$ ,  $y(0) = f$  where  $V \in C_b^\infty(\mathbb{R}^d, \mathbb{R}^d)$ . In particular,  $\exp(Vt)(x)$  is  $\exp(Vt)f$  for  $f$  being the identity function. The family of processes  $Y(x) = \{Y_t(x)\}_{t \in [0,1]}$  below corresponds to the Ninomiya-Victoir approximation (see [33]).

**Example 5** Let  $\Lambda$  and  $Z$  be two independent random variables such that  $\Lambda$  is Bernoulli distributed  $\mathbb{P}(\Lambda = 1) = \mathbb{P}(\Lambda = -1) = \frac{1}{2}$  and  $Z = (Z^i)_{i=1}^k$  is a standard normal  $k$ -dimensional random variable. Consider the family of processes  $Y(x) = \{Y_t(x)\}_{t \in [0,1]}$  defined by

$$Y_t(x) = \begin{cases} \exp(\frac{V_0}{2}t) \prod_{i=1}^k \exp(Z^i V_i t^{1/2}) \exp(\frac{V_0}{2}t)(x) & \text{if } \Lambda = 1 \\ \exp(\frac{V_0}{2}t) \prod_{i=1}^k \exp(Z^{k+1-i} V_{k+1-i} t^{1/2}) \exp(\frac{V_0}{2}t)(x) & \text{if } \Lambda = -1 \end{cases}$$

with the corresponding operator  $(Q_t \varphi)(x) := \mathbb{E}[\varphi(Y_t(x))]$ . Then there exists a constant  $c_7 > 0$  such that for  $\varphi \in C_b^{V,8}(\mathbb{R}^d)$

$$\sup_x |\overline{Q}_t \varphi(x) - \mathbb{E}[\varphi_t^5(x)]| \leq c_7 t^3 \|\varphi\|_{V,6}$$

Hence  $\{Y_t(x)\}_{t \in [0,1]}$  is 5-perfect.

**Proof.** We first consider the case  $\Lambda = 1$ . Let  $\{Y^i\}_{i=0}^{k+1}$  be defined,

$$\begin{aligned} \frac{d\varphi}{ds}(Y_s^0) &= V_0 \varphi(Y_s^0) \text{ for } s \in [0, \frac{t}{2}], Y_0^0 = x \\ \frac{d\varphi}{ds}(Y_s^1) &= Z^1 V_1 \varphi(Y_s^1) \text{ for } s \in [0, \sqrt{t}], Y_0^1 = Y_{\frac{t}{2}}^0 \\ \frac{d\varphi}{ds}(Y_s^i) &= Z^i V_i \varphi(Y_s^i) \text{ for } s \in [0, \sqrt{t}], Y_0^i = Y_{\sqrt{t}}^{i-1}, i = 2, \dots, k \\ \frac{d\varphi}{ds}(Y_s^{k+1}) &= V_0 \varphi(Y_s^{k+1}) \text{ for } s \in [0, \frac{t}{2}], Y_0^{k+1} = Y_{\sqrt{t}}^k \end{aligned}$$

It follows from the definition of the algorithm and by Itô's Formula that,

$$\varphi(Y_{\frac{t}{2}}^{k+1}) = \varphi(Y_{\sqrt{t}}^k) + V_0 \varphi(Y_{\sqrt{t}}^k) \frac{t}{2} + V_0^2 \varphi(Y_{\sqrt{t}}^k) \frac{t^2}{8} + \int_0^{t/2} \int_0^{s_1} \int_0^{s_2} V_0^3 \varphi(Y_{s_3}^{k+1}) ds_3 ds_2 ds_1. \quad (19)$$

We need to expand the right hand side of (19) and divide the resulting expansion into two parts: the required truncation and a remainder whose expected value should be bounded by  $Ct^3 \|\varphi\|_{V,6}$ . The final term in (19) belongs to the remainder and indeed,

$$\mathbb{E} \left[ \left| \int_0^{t/2} \int_0^{s_1} \int_0^{s_2} V_0^3 \varphi(Y_{s_3}^{k+1}) ds_3 ds_2 ds_1 \right| \right] \leq \|V_0^3 \varphi\|_{\infty} \frac{(t/2)^3}{6} \leq \frac{t^3}{48} \|\varphi\|_{V,6} \quad (20)$$

Expanding the third term in (19),

$$\begin{aligned} V_0^2 \varphi(Y_{\sqrt{t}}^k) &= V_0^2 \varphi(Y_{\sqrt{t}}^{k-1}) + \int_0^{\sqrt{t}} Z^k V_k V_0^2 \varphi(Y_s^k) ds \\ &= V_0^2 \varphi(Y_{\sqrt{t}}^{k-1}) + Z^k V_k V_0^2 \varphi(Y_{\sqrt{t}}^{k-1}) \sqrt{t} + \int_0^{\sqrt{t}} \int_0^{s_1} (Z^k)^2 V_k^2 V_0^2 \varphi(Y_{s_2}^k) ds_2 ds_1 \\ &= \dots \\ &= V_0^2 \varphi(Y_{\frac{t}{2}}^0) + Z^1 V_1 V_0^2 \varphi(Y_{\frac{t}{2}}^0) \sqrt{t} + \sum_{i=2}^k Z^i V_i V_0^2 \varphi(Y_{\sqrt{t}}^{i-1}) \sqrt{t} \\ &\quad + \sum_{i=1}^k \int_0^{\sqrt{t}} \int_0^{s_1} (Z^i)^2 V_i^2 V_0^2 \varphi(Y_{s_2}^i) ds_2 ds_1 \end{aligned}$$

So

$$\begin{aligned} V_0^2 \varphi(Y_{\sqrt{t}}^k) &= V_0^2 \varphi(x) + \int_0^{\frac{t}{2}} V_0^3 \varphi(Y_s^0) ds + Z^1 V_1 V_0^2 \varphi(Y_{\frac{t}{2}}^0) \sqrt{t} + \sum_{i=2}^k Z^i V_i V_0^2 \varphi(Y_{\sqrt{t}}^i) \sqrt{t} \\ &\quad + \sum_{i=1}^k \int_0^{\sqrt{t}} \int_0^{s_1} (Z^i)^2 V_i^2 V_0^2 \varphi(Y_{s_2}^i) ds_2 ds_1 \end{aligned}$$

Now the last four terms are all  $O(t)$  since  $\mathbb{E} \left[ Z^1 V_1 V_0^2 \varphi(Y_{\frac{t}{2}}^0) \sqrt{t} \right] = \mathbb{E} \left[ Z^i V_i V_0^2 \varphi(Y_{\sqrt{t}}^{i-1}) \sqrt{t} \right] = 0$  because  $Z$  is normal,  $\mathbb{E} \left[ \left| \int_0^{\frac{t}{2}} V_0^3 \varphi(Y_s^0) ds \right| \right] \leq \|V_0^3 \varphi\|_{\infty} \frac{t}{2}$  and finally  $\mathbb{E} \left[ \left| \int_0^{\sqrt{t}} \int_0^{s_1} (Z^i)^2 V_i^2 V_0^2 \varphi(Y_{s_2}^i) ds_2 ds_1 \right| \right] \leq \|V_i^2 V_0^2 \varphi\|_{\infty} \frac{t}{2}$ .

So for the third term in (19) we have established,

$$\begin{aligned} \left| E \left[ V_0^2 \varphi(Y_{\sqrt{t}}^k) \frac{t^2}{8} \right] - V_0^2 \varphi(x) \frac{t^2}{8} \right| &\leq \frac{t^3}{16} \left( \|V_0^3 \varphi\|_{\infty} + \sum_{i=1}^k \|V_i^2 V_0^2 \varphi\|_{\infty} \right) \\ &\leq \frac{t^3}{16} \|\varphi\|_{V,6}. \end{aligned} \quad (21)$$

Similarly, for the second term on the RHS of (19),

$$\left| E \left[ V_0 \varphi(Y_{\sqrt{t}}^k) \frac{t}{2} \right] - \left( V_0 \varphi(x) + V_0^2 \varphi(x) \frac{t}{2} + \sum_{i=1}^k V_i^2 V_0 \varphi(x) \frac{t}{2} \right) \frac{t}{2} \right| \leq \frac{t^3}{8} \|\varphi\|_{V,6} \quad (22)$$

Finally, for the first term on the RHS of (19),

$$\left| E \left[ \varphi(Y_{\sqrt{t}}^k) \right] - \left( \varphi(x) + V_0 \varphi(x) \frac{t}{2} + V_0^2 \varphi(x) \frac{t^2}{8} + \sum_{i=1}^k V_i^4 \varphi(x) \frac{t^2}{4!} + \sum_{i=1}^k \left( V_i^2 \varphi(x) \frac{t}{2} + V_0 V_i^2 \varphi(x) \frac{t^2}{4} \right) + \sum_{i=2}^k \sum_{j=1}^{i-1} V_j^2 V_i^2 \varphi(x) \frac{t^2}{4} \right) \right| \leq \frac{t^3}{16} \|\varphi\|_{V,6} \quad (23)$$

Substituting (20), (21), (22) and (23) in (19) gives the bound for the case  $\Lambda = 1$ . An analogous bound is then established for the case  $\Lambda = -1$ . The final result is obtained by taking the average of the two cases. See [12] for details. ■

The following Lemma is required to prove the main theorem below.

**Lemma 6** For  $0 < s \leq t \leq 1$  and any  $m$ -perfect family  $\{\bar{X}_t(x)\}_{t \in (0,1]}$  with corresponding operator  $Q = \{Q_t\}_{t \in (0,1]}$  we have,

$$\|P_t(P_s \varphi) - Q_t(P_s \varphi)\|_\infty \leq c_8 \|\varphi\|_p \sum_{j=m+1}^M \frac{t^{j/2}}{s^{\frac{j-p}{2}}}, \quad (24)$$

where  $\varphi \in C_b^p(\mathbb{R}^d)$  for  $0 \leq p < \infty$  and some constant  $c_8 > 0$ . In particular, for  $\varphi \in C_b^M(\mathbb{R}^d)$ ,

$$\|P_t(P_s \varphi) - Q_t(P_s \varphi)\|_\infty \leq c_8 \|\varphi\|_p t^{\frac{m+1}{2}}. \quad (25)$$

**Proof.** Since  $C_b^\infty(\mathbb{R}^d)$  is dense in  $C_b^p(\mathbb{R}^d)$  in the topology generated by the norm  $\|\cdot\|_{p,\infty}$  it suffices to prove (24) and (25) only for a function  $\varphi \in C_b^\infty(\mathbb{R}^d)$ . By Corollary 18 in the appendix,

$$\begin{aligned} \|P_t \varphi\|_{V,j} &= \sum_{i=1}^j \sum_{\substack{\alpha_1, \dots, \alpha_i \in \mathcal{A}_0 \\ \|\alpha_1 * \dots * \alpha_i\| = j}} \|V_{[\alpha_1]} \cdots V_{[\alpha_i]} P_t \varphi\|_\infty \\ &\leq \sum_{i=1}^j \sum_{\substack{\alpha_1, \dots, \alpha_i \in \mathcal{A}_0 \\ \|\alpha_1 * \dots * \alpha_i\| = j}} \frac{C_p^T \|\varphi\|_p}{t^{(\|\alpha_1 * \dots * \alpha_i\| - p)/2}} \leq \frac{c_9 \|\varphi\|_p}{t^{\frac{i-p}{2}}} \end{aligned}$$

for some  $c_9 \equiv c_9(j, p) \geq 0$ . Then (24) and (25) follow from the definition of an  $m$ -perfect family. ■

The family of processes  $\bar{X}(x) = \{\bar{X}_t(x)\}_{t \in [0, \infty)}$  below corresponds to the Kusuoka approximation. We recall that Kusuoka's result requires only the UFG condition.

**Example 7** A family of random variables  $\{Z_\alpha : \alpha \in \mathcal{A}_0\}$  is said to be ***m-moment similar*** if  $\mathbb{E}[|Z_\alpha|^r] < \infty$  for any  $r \in \mathbb{N}$ ,  $\alpha \in \mathcal{A}_0$  and  $Z_{(0)} = 1$  with,

$$\mathbb{E}[Z_{\alpha_1} \dots Z_{\alpha_j}] = \mathbb{E}[I_{\alpha_1} \dots I_{\alpha_j}]$$

for any  $j = 1, \dots, m$  and  $\alpha_1, \dots, \alpha_j \in \mathcal{A}_0$  such that  $\|\alpha_1\| + \dots + \|\alpha_j\| \leq m$  where  $I_\alpha$  is defined as above.

Let  $\{Z_\alpha : \alpha \in \mathcal{A}_0\}$  be a family of *m-moment similar* random variables and let  $\bar{X}(x) = \{\bar{X}_t(x)\}_{t \in [0, \infty)}$  be the family of processes,

$$\bar{X}_t(x) = \sum_{j=0}^m \frac{1}{j!} \sum_{\substack{\alpha_1, \dots, \alpha_j \in \mathcal{A}_0, \\ \|\alpha_1\| + \dots + \|\alpha_j\| \leq m}} t^{\frac{\|\alpha_1\| + \dots + \|\alpha_j\|}{2}} (P_{\alpha_1}^0 \dots P_{\alpha_j}^0)(V_{[\alpha_1]} \dots V_{[\alpha_j]} H)(x) \quad (26)$$

where  $H : \mathbb{R}^d \rightarrow \mathbb{R}^d$  is defined  $H(x) = x$  and

$$P_\alpha^0 := |\alpha|^{-1} \sum_{j=0}^{|\alpha|} \frac{(-1)^{j+1}}{j} \sum_{\beta_1 * \dots * \beta_j = \alpha} Z_{\beta_1} \dots Z_{\beta_j}$$

with the corresponding operator  $Q = \{Q_t\}_{t \in (0, 1]}$  in  $C_b(\mathbb{R}^d)$  defined by,

$$Q_t \varphi(x) = \mathbb{E}[\varphi(\bar{X}_t(x))]$$

for  $\varphi \in C_b^\infty(\mathbb{R}^d)$  Then,

$$\|P_{t+s} \varphi - Q_t P_s \varphi\|_\infty \leq c_{10} \|\nabla \varphi\|_\infty \sum_{j=m+1}^{m^{m+1}} \frac{t^{j/2}}{s^{\frac{j-1}{2}}} \quad (27)$$

for some constant  $c_{10} > 0$ .

**Proof.** See Definition 1, Theorem 3 and Lemma 18 in Kusuoka[18] for (27). ■

The family  $\bar{X}(x)$ ,  $x \in \mathbb{R}^d$  as defined in (26) is not *m-perfect*. However, inequality (27) is a particular case of (24) where  $p = 1$  and  $M = m^{m+1}$ . Since (24) is the only result required to obtain (28), we deduce from the proof of Theorem 8 that (28), with  $p = 1$ , holds for Kusuoka's method as well. Similarly part (ii) of Corollary 9 holds for Kusuoka's method.

The set of vector fields appearing in (26) belong to the Lie algebra generated by the original vector fields  $\{V_0, V_1, \dots, V_k\}$ . Ben Arous[2] and Burrage & Burrage[3] employ the same set of vector fields to produce *strong* approximations of solutions of SDEs. Notably, the same ideas appear much earlier in Magnus[25], in the context of approximations of the solution of linear (deterministic) differential equations. Castell[6] also gives an explicit

formula for the solution of an SDE in terms of Lie brackets and iterated Stratonovich integrals.

We now prove our main result on  $m$ -perfect families, the gist of which can be conveyed by the concept of local and global order of an approximation. Local order measures how close an approximation is to the exact solution on a sub-interval of the integration, given an exact initial condition at the start of that subinterval. The global order of an approximation looks at the build up of errors over the entire integration range. The theorem below states that, in the best possible case, the global order of an approximation obtained using an  $m$ -perfect family is one less than the local order. More precisely, for a suitable partition, the global error is of order  $\frac{m-1}{2}$  whilst the local error is of order  $\frac{m+1}{2}$ .

Let us define the function,

$$\Upsilon^p(n) = \begin{cases} n^{-\frac{1}{2} \min(\gamma p, (m-1))} & \text{if } \gamma p \neq m-1 \\ n^{-(m-1)/2} \ln n & \text{for } \gamma p = m-1 \end{cases}$$

In the following,

$$\mathcal{E}^{\gamma, n}(\varphi) := \|P_T \varphi - Q_{h_n}^m Q_{h_{n-1}}^m \dots Q_{h_1}^m \varphi\|_\infty$$

for  $\gamma \in \mathbb{R}$ ,  $n \in \mathbb{N}$ .

**Theorem 8** *Let  $T, \gamma > 0$  and  $\pi_n = \{t_j = (\frac{j}{n})^\gamma T\}_{j=0}^n$  be a partition of the interval  $[0, T]$  where  $n \in \mathbb{N}$  is such that  $\{h_j = t_j - t_{j-1}\}_{j=1}^n \subseteq (0, 1]$ . Then for any  $m$ -perfect family  $\{\bar{X}_t(x)\}_{t \in [0, T]}$  with corresponding operator  $Q = \{Q_t\}_{t \in (0, 1]}$  we have, for  $\varphi \in C_b^p(\mathbb{R}^d)$  where  $p = 1, \dots, m$ ,*

$$\mathcal{E}^{\gamma, n}(\varphi) \leq c_{11} \Upsilon^p(n) \|\varphi\|_p + \|P_{h_1} \varphi - Q_{h_1}^m \varphi\|_\infty \quad (28)$$

for some constant  $c_{11} \equiv c_{11}(\gamma, M, T) > 0$  where  $M \geq m+1$ , as in Definition 2. In particular, if  $\gamma \geq \frac{m-1}{p}$  then,

$$\mathcal{E}^{\gamma, n}(\varphi) \leq \frac{c_{11}}{n^{\frac{m-1}{2}}} \|\varphi\|_p + \|P_{h_1} \varphi - Q_{h_1}^m \varphi\|_\infty$$

**Proof.** We have,

$$\begin{aligned} \mathcal{E}^{\gamma, n}(\varphi) &= P_{h_n}(P_{T-h_n} \varphi) - Q_{h_n}^m(P_{T-h_n} \varphi) \\ &\quad + \sum_{j=1}^{n-1} Q_{h_n}^m \dots Q_{h_{j+1}}^m (P_{T-h_{j+1}-\dots-h_n} \varphi - Q_{h_j}^m P_{T-h_j-\dots-h_n} \varphi) \\ &= P_{h_n}(P_{t_{n-1}} \varphi) - Q_{h_n}^m(P_{t_{n-1}} \varphi) \\ &\quad + \sum_{j=1}^{n-1} Q_{h_n}^m \dots Q_{h_{j+1}}^m (P_{h_j}(P_{t_{j-1}} \varphi) - Q_{h_j}^m(P_{t_{j-1}} \varphi)) \end{aligned}$$



By Lemma 6, there exists a constant  $c_8 > 0$  such that,

$$\|P_{h_n}(P_{t_{n-1}}\varphi) - Q_{h_n}^m(P_{t_{n-1}}\varphi)\|_\infty \leq c_8 \|\varphi\|_p \sum_{l=m+1}^M \frac{h_n^{l/2}}{t_{n-1}^{\frac{l-p}{2}}}$$

Since  $P$  is a semigroup and  $Q_{h_j}^m$  is a Markov operator for  $j = 2, \dots, n-1$ ,

$$\begin{aligned} \left\| Q_{h_n}^m \dots Q_{h_{j+1}}^m (P_{h_j}(P_{t_{j-1}}\varphi) - Q_{h_j}^m(P_{t_{j-1}}\varphi)) \right\|_\infty &\leq \left\| P_{h_j}(P_{t_{j-1}}\varphi) - Q_{h_j}^m(P_{t_{j-1}}\varphi) \right\|_\infty \\ &\leq c_{12} \|\varphi\|_p \sum_{l=m+1}^M \frac{h_j^{l/2}}{t_{j-1}^{\frac{l-p}{2}}} \end{aligned}$$

for some  $c_{12} > 0$ . Finally, since  $Q_{h_j}^m$  is a Markov operator, it follows from (32) that,

$$\left\| Q_{h_n}^m \dots Q_{h_2}^m (P_{h_1}\varphi - Q_{h_1}^m\varphi) \right\|_\infty \leq \|P_{h_1}\varphi - Q_{h_1}^m\varphi\|_\infty$$

Combining these last four results gives,

$$\mathcal{E}^{\gamma,n}(\varphi) = \|P_T\varphi - Q_{h_n}^m \dots Q_{h_1}^m\varphi\|_\infty \leq \|P_{h_1}\varphi - Q_{h_1}^m\varphi\|_\infty + c_{12} \|\varphi\|_p \sum_{j=2}^n \sum_{l=m+1}^M \frac{h_j^{l/2}}{t_{j-1}^{\frac{l-p}{2}}}$$

It follows, almost immediately from the definition of  $h_j$  that,

$$h_j = \frac{\gamma T(j-1)^{\gamma-1}}{n^\gamma} \int_{j-1}^j \left( \frac{u}{j-1} \right)^{\gamma-1} du$$

but for  $j \in \{2, \dots, n\}$ ,  $(\frac{u}{j-1})^{\gamma-1} \leq \max[(\frac{j}{j-1})^{\gamma-1}, 1] \leq \max[2^{\gamma-1}, 1]$ . Hence for  $l = m+1, \dots, M$ ,

$$\begin{aligned} \frac{h_j^{l/2}}{t_{j-1}^{(l-p)/2}} &\leq \frac{(\frac{\gamma T(j-1)^{\gamma-1}}{n^\gamma} \max[2^{\gamma-1}, 1])^{l/2}}{((\frac{j-1}{n})^\gamma T)^{(l-p)/2}} \\ &\leq c_{13} \left(\frac{T}{n^\gamma}\right)^{\frac{l}{2} - \frac{(l-p)}{2}} (j-1)^{\frac{(\gamma-1)l}{2} - \frac{\gamma(l-p)}{2}} = c_{13} \left(\frac{T}{n^\gamma}\right)^{\frac{l}{2}} (j-1)^{\frac{\gamma p-l}{2}} \end{aligned}$$

where  $c_{13} \equiv c_{13}(\gamma, M) = \max[1, (\gamma \max[2^{\gamma-1}, 1])^{M/2}]$ . It follows that,

$$\sum_{l=m+1}^M \frac{h_j^{l/2}}{t_{j-1}^{(l-p)/2}} \leq c_{14} \left(\frac{1}{n}\right)^{\frac{\gamma p}{2}} \sum_{l=m+1}^M (j-1)^{\frac{\gamma p-l}{2}}$$

where  $c_{14} \equiv c_{14}(\gamma, M, T) = T^{p/2} c_{13}$  and since  $\sum_{l=m+1}^M (j-1)^{\frac{\gamma p-l}{2}} = (j-1)^{\frac{\gamma p-(m+1)}{2}} \sum_{l=0}^{M-(m+1)} (j-1)^{-\frac{l}{2}} \leq (j-1)^{\frac{\gamma p-(m+1)}{2}} M$  we have,

$$\sum_{l=m+1}^M \frac{h_j^{l/2}}{t_{j-1}^{(l-p)/2}} \leq c_{14} M \left(\frac{1}{n}\right)^{\frac{\gamma p}{2}} (j-1)^{\frac{\gamma p-(m+1)}{2}} \quad (29)$$

We now consider (29) for three different ranges of  $\gamma$ .

For  $\gamma \in \left(0, \frac{m-1}{p}\right)$ ,  $\sum_{j=2}^n (j-1)^{\frac{\gamma p-(m+1)}{2}} \leq \sum_{j=2}^{\infty} (j-1)^{\frac{\gamma p-(m+1)}{2}}$  and since the series on the RHS is convergent, we have,

$$n^{-\frac{\gamma p}{2}} \sum_{j=2}^n (j-1)^{\frac{\gamma p-(m+1)}{2}} \leq c_{15} n^{-\frac{\gamma p}{2}}$$

for some constant  $c_{15} \equiv c_{15}(\gamma, M) > 0$ .

For  $\gamma = \frac{m-1}{p}$ ,  $\sum_{j=2}^n (j-1)^{-1} \leq c_{16} \ln n$  for some constant  $c_{16} \equiv c_{16}(\gamma, M) > 0$  so we have,

$$n^{-\frac{\gamma p}{2}} \sum_{j=2}^n (j-1)^{\frac{\gamma p-(m+1)}{2}} \leq c_{16} n^{-\frac{(m-1)}{2}} \ln n$$

For  $\gamma > \frac{m-1}{p}$ ,  $\sum_{j=2}^n \left(\frac{j-1}{n}\right)^{\frac{\gamma p-(m+1)}{2}} \frac{1}{n} \leq c_{17} \int_0^1 x^{\frac{\gamma p-(m+1)}{2}} dx = c_{17} \int_0^1 x^{-1+\frac{\gamma p-(m-1)}{2}} dx \leq c_{18}$  (like a Riemann integral) for some constants  $c_{17} \equiv c_{17}(\gamma, M)$ ,  $c_{18} \equiv c_{18}(\gamma, M)$  so,

$$n^{-\frac{\gamma p}{2}} \sum_{j=2}^n (j-1)^{\frac{\gamma p-(m+1)}{2}} = n^{-\frac{m-1}{2}} \sum_{j=2}^n \left(\frac{j-1}{n}\right)^{\frac{\gamma p-(m+1)}{2}} \frac{1}{n} \leq c_{18} n^{-\frac{m-1}{2}}$$

■

We observe that the rate of convergence is controlled by the maximum between  $\Upsilon(n)$  and the rate at which  $\|P_{h_1} \varphi - Q_{h_1}^m \varphi\|_{\infty}$  converges to 0. We define  $\tilde{\Upsilon}^{k_1, k_2}(n) = \Upsilon^{k_1}(n) + n^{-\frac{\gamma k_2}{2}}$ . Hence we have the following corollary:

### Corollary 9

(i) For any  $\varphi \in C_b^M(\mathbb{R}^d)$ ,

$$\mathcal{E}^{\gamma, n}(\varphi) \leq c_{19} \tilde{\Upsilon}^{m+1, m+1}(n) \|\varphi\|_M.$$

for some constant  $c_{19} > 0$ . In particular, if  $\gamma \geq 1$ , then  $\mathcal{E}^{\gamma, n}(\varphi) \leq \frac{c_{19}}{n^{\frac{m-1}{2}}} \|\varphi\|_M$ .

(ii) For any  $\varphi \in C_b^1(\mathbb{R}^d)$ ,

$$\mathcal{E}^{\gamma, n}(\varphi) \leq c_{21} \tilde{\Upsilon}^{1, 1}(n) \|\varphi\|_1$$

for some constant  $c_{21} > 0$ , if there exists a constant  $c_{20} > 0$  independent of  $t$  such that,

$$\sup_{x \in \mathbb{R}^d} |\bar{X}_t(x) - x| \leq c_{20} \sqrt{t}. \quad (30)$$

In particular, if  $\gamma \geq m - 1$ , then  $\mathcal{E}^{\gamma, n}(\varphi) \leq \frac{c_{21}}{n^{\frac{m-1}{2}}} \|\varphi\|_1$ .

(iii) For any  $\varphi \in C_b^l(\mathbb{R}^d)$  where  $1 < l < M$ ,

$$\mathcal{E}^{\gamma, n}(\varphi) \leq c_{24} \bar{\Upsilon}^{l, c_{23}}(n) \|\varphi\|_l$$

for some constant  $c_{24} > 0$ , if there exist two constants  $c_{22}, c_{23} > 0$  independent of  $t$  such that,

$$\|P_t \varphi - Q_t^m \varphi\|_\infty \leq c_{22} t^{\frac{c_{23}}{2}} \|\varphi\|_l. \quad (31)$$

In particular, if  $\gamma \geq m - 1$ , then  $\mathcal{E}^{\gamma, n}(\varphi) \leq \frac{c_{24}}{n^{\frac{m-1}{2}}} \|\varphi\|_l$ .

**Proof.**

(i) The result follows from the theorem and the definition of an  $m$ -perfect family.

(ii) If  $\varphi \in C_b(\mathbb{R}^d)$  is Lipschitz then,

$$|Q_t \varphi(x) - \varphi(x)| \leq c_{20} \|\nabla \varphi\|_\infty \sqrt{t} \quad (32)$$

hence,

$$\|P_{h_1} \varphi - Q_{h_1}^m \varphi\|_\infty \leq c_{20} \|\varphi\|_1 \sqrt{t}.$$

(iii) The result follows from the theorem and (31). ■

Finally we define  $\mu_t$  to be the law of  $X_t$ :

$$\mu_t(\varphi) = E[\varphi(X_t)] \text{ for } \varphi \in C_b(\mathbb{R}^d).$$

We also define  $\mu_t^N$  to be the probability measure defined by,

$$\mu_t^N(\varphi) = \mathbb{E}[Q_{h_n}^m Q_{h_{n-1}}^m \dots Q_{h_1}^m \varphi(X_0)] = \int_{\mathbb{R}^d} Q_{h_n}^m Q_{h_{n-1}}^m \dots Q_{h_1}^m \varphi(x) \mu_0(dx)$$

for  $\varphi \in C_b(\mathbb{R}^d)$  and need to introduce the family of norms on the set of signed measures:

$$|\mu|_l = \sup \left\{ |\mu(\varphi)|, \varphi \in C_b^l(\mathbb{R}^d), \|\varphi\|_{l, \infty} \leq 1 \right\}, \quad l \geq 1.$$

Obviously,  $|\mu|_l \leq |\mu|_{l'}$  if  $l \leq l'$ . In other words, the higher the value of  $l$ , the coarser the norm. We have the following.

**Corollary 10**

(i) For  $l \geq M$ , we have  $|\mu_t - \mu_t^N|_l \leq c_{19} \bar{\Upsilon}^{m+1, m+1}(n)$ . In particular, if  $\gamma \geq 1$ , then  $|\mu_t - \mu_t^N|_l \leq \frac{c_{19}}{n^{\frac{m-1}{2}}}$ .

(ii) If (30) is satisfied then  $|\mu_t - \mu_t^N|_l \leq c_{21} \bar{\Upsilon}^{1,1}(n)$ . In particular, if  $\gamma \geq m - 1$ , then  $|\mu_t - \mu_t^N|_l \leq \frac{c_{21}}{n^{\frac{m-1}{2}}}$ .

(iii) If (31) is satisfied then  $|\mu_t - \mu_t^N|_l \leq c_{24} \bar{\Upsilon}^{l, c_{23}}(n)$ . In particular, if  $\gamma \geq m - 1$ , then  $|\mu_t - \mu_t^N|_l \leq \frac{c_{24}}{n^{\frac{m-1}{2}}}$ .

Throughout, the constants  $c_{19}, c_{21}, c_{23}, c_{24} > 0$  correspond to those found in Corollary 9.

**Remark** We deduce that there is a payoff between the rate of convergence and the coarseness of the norm employed: the finer the norm the slower the rate of convergence. Hence intermediate results such as part (iii) of Corollaries 9 and 10 may prove useful in subsequent applications. The additional constraint (31) holds, for example, for the Lyons-Victoir method, as a cubature formula of degree  $m$  is also a cubature formula of degree  $m'$  for  $m' \leq m$ . Similarly, it holds for Kusuoka's approximation since an  $m$ -similar family is also  $m'$ -similar for any  $m' \leq m$ .

## 4 An Application to Filtering

We begin with a short description of the filtering problem. Let  $(X, Y)$  be a system of partially observed random processes. The process  $X$  satisfies the stochastic differential equation (1) and is the unobserved component. The process  $Y$  is the observable component and satisfies the evolution equation,

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

where  $\{B_t\}_{t \in [0, \infty)}$  is an  $l$ -dimensional Brownian motion independent of  $X$  and  $h = (h^i)_{i=1}^l \in C_b^\infty(\mathbb{R}^d, \mathbb{R}^l)$ . Let  $(\mathcal{Y}_t)_{t \geq 0}$  be the filtration generated by  $Y$ ,  $\mathcal{Y}_t = \sigma(Y_s, 0 \leq s \leq t)$ . The problem of stochastic filtering for the partially observed system  $(X, Y)$  involves the construction of  $\pi_t(\varphi)$ , where  $\pi = \{\pi_t, t \geq 0\}$  is the conditional distribution of  $X_t$  given  $\mathcal{Y}_t$  and  $\varphi$  belongs to a suitably large class of functions. If  $\varphi$  is square integrable with respect to the law of  $X_t$  then,

$$\pi_t(\varphi) = \mathbb{E}[\varphi(X_t) | \mathcal{Y}_t], \mathbb{P} - \text{almost surely.}$$

Using Girsanov's theorem, one can find a new probability measure  $\tilde{\mathbb{P}}$  absolutely continuous with respect to  $\mathbb{P}$  (and vice versa), so that  $Y$  is a Brownian motion under  $\tilde{\mathbb{P}}$ , independent of  $X$  and, almost surely,

$$\pi_t(\varphi) = \frac{\rho_t(\varphi)}{\rho_t(1)}, \tag{33}$$

where,

$$\rho_t(\varphi) = \tilde{\mathbb{E}} \left[ \varphi(X_t) \exp \left( \sum_{i=1}^l \int_0^t h^i(X_s) dY_s^i - \frac{1}{2} \sum_{i=1}^l \int_0^t h^i(X_s)^2 ds \right) \middle| \mathcal{Y}_t \right] \quad (34)$$

and  $\tilde{\mathbb{E}}$  is the expectation with respect to  $\tilde{\mathbb{P}}$ . The measure  $\rho_t$  is called the *unnormalised conditional distribution* of the signal. The identity (33) is called the Kallianpur-Striebel Formula. In the following, we will denote by  $\|\cdot\|_p$ , the  $L^p$ -norm with respect to the probability measure  $\tilde{\mathbb{P}}$ ,  $\|\xi\|_p = \tilde{\mathbb{E}} [|\xi|^p]^{\frac{1}{p}}$ , for any random variable  $\xi$ . The laws of the families  $X(x) = \{X_t(x)\}_{t \in [0, \infty)}$ ,  $x \in \mathbb{R}^d$  and  $\bar{X}(x) = \{\bar{X}_t(x)\}_{t \in [0, \infty)}$ ,  $x \in \mathbb{R}^d$  are not affected by the change of measure, hence, to avoid working with both  $\mathbb{P}$  and  $\tilde{\mathbb{P}}$  we can write,

$$(P_t \varphi)(x) = \tilde{\mathbb{E}}[\varphi(X_t(x))], \quad (Q_t \varphi)(x) = \tilde{\mathbb{E}}[\varphi(\bar{X}_t(x))].$$

In the following, we will only consider equidistant partitions and smooth functions. The method of approximation and the results closely follow the application of the classical Euler method as described in Picard [36] and Talay [42].

Let  $y_r$ ,  $r = 1, \dots, n$  be the observation process increments  $y_r = Y_{\frac{(r+1)t}{n}} - Y_{\frac{rt}{n}}$  and  $h_r \in C_b^\infty(\mathbb{R}^d)$ ,  $r = 0, \dots, n-1$ , be the (observation dependent) functions defined by  $h_r = \sum_{i=1}^l (h^i y_r^i - \frac{t}{2n} (h^i)^2)$ . Let  $R_s^r, \bar{R}_s^r : C_b^\infty(\mathbb{R}^d) \rightarrow C_b^\infty(\mathbb{R}^d)$ ,  $r = 0, 1, \dots, n$  be the following operators,

$$R_s^n \varphi(x) = P_s \varphi(x), \quad \bar{R}_s^n \varphi(x) = Q_s \varphi(x) \text{ for } \varphi \in C_b^\infty(\mathbb{R}^d), \quad x \in \mathbb{R}^d$$

and, for  $r = 0, 1, \dots, n-1$ , and for  $\varphi \in C_b^\infty(\mathbb{R}^d)$ ,  $x \in \mathbb{R}^d$ ,

$$\begin{aligned} R_s^r \varphi(x) &= \tilde{\mathbb{E}}[\varphi(X_s(x)) \exp(h_r(X_s(x))) | \mathcal{Y}_s] = P_s \varphi^r(x) \\ \bar{R}_s^r \varphi(x) &= \tilde{\mathbb{E}}[\varphi(\bar{X}_s(x)) \exp(h_r(\bar{X}_s(x))) | \mathcal{Y}_s] = Q_s \varphi^r(x), \end{aligned}$$

where  $\varphi^r = \varphi \exp(h_r)$  and  $s \in [0, 1]$ .

Firstly, one approximates  $\rho$  by replacing the (continuous) observation path with a discrete version. We choose the equidistant partition  $\{\frac{it}{n}, i = 0, 1, \dots, n\}$  of the interval  $[0, t]$  and consider only the observation data  $\{y_r, r = 0, 1, \dots, n\}$ . We define the measure,

$$\begin{aligned} \rho_t^n(\varphi) &= \tilde{\mathbb{E}} \left[ \varphi(X_t) \exp \left( \sum_{i=0}^{n-1} h_i \left( X_{\frac{it}{n}} \right) \right) \middle| \mathcal{Y}_t \right] \\ &= \tilde{\mathbb{E}} \left[ R_{\frac{t}{n}}^0 R_{\frac{t}{n}}^1 \dots R_{\frac{t}{n}}^{n-1} \varphi(X_0) \middle| \mathcal{Y}_t \right] \text{ for } \varphi \in C_b(\mathbb{R}^d). \end{aligned} \quad (35)$$

Following Theorem 1 from Picard [36], for any  $\varphi \in C_b^\infty(\mathbb{R}^d)$  there is a constant  $c \equiv c(t, \varphi)$  such that,

$$\|\rho_t(\varphi) - \rho_t^n(\varphi)\|_2 \leq \frac{c}{n}$$

The next step is to approximate  $\prod_{i=0}^n R_{\frac{t}{n}}^i$  with  $\prod_{i=0}^n \bar{R}_{\frac{t}{n}}^i$ . For this we need to adapt the definition of an  $m$ -perfect family so that we may use functions which are parametrized by the observation path  $Y$ . Let  $C_b^{Y,\infty}(\mathbb{R}^d)$  be the set of measurable functions,  $f : \mathbb{R}^d \times C([0, T], \mathbb{R}^l) \rightarrow \mathbb{R}$  with the following properties:

- i. for any  $y \in C([0, T], \mathbb{R}^l)$  the function  $x \rightarrow f(x, y)$  belongs to  $C_b^\infty(\mathbb{R}^d)$ .
- ii. for any multi-index  $\alpha \in \mathcal{A}$ , any  $x \in \mathbb{R}^d$  and  $p \geq 1$ ,  $|D_\alpha f(x, Y)|_p < \infty$ .
- iii. for any multi-index  $\alpha \in \mathcal{A}$  and  $p \geq 1$ ,  $\|D_\alpha f\|_{p,\infty} = \sup_{x \in \mathbb{R}^d} |D_\alpha f(x, Y)|_p < \infty$ .

For  $f \in C_b^{Y,\infty}(\mathbb{R}^d)$  we define the norm  $\|f\|_p^m = \sum_{\alpha \in \mathcal{A}(m)} \|D_\alpha f\|_{p,\infty}$ . We note that if

$f : \mathbb{R}^d \times C([0, T], \mathbb{R}^l) \rightarrow \mathbb{R}$  is constant in the  $y$  variable, then  $\|D_\alpha f(x, Y)\|_p = |D_\alpha f(x, Y)|$  and  $\|f\|_p^m = \|f\|_\infty + \|\nabla^1 f\|_\infty + \dots + \|\nabla^m f\|_\infty$ .

We now consider an  $m$ -perfect family  $\bar{X}(x)$  that satisfies the equivalent of (15) extended to functions in  $C_b^{Y,\infty}(\mathbb{R}^d)$ ; more precisely we will assume that for any  $f \in C_b^{Y,\infty}(\mathbb{R}^d)$ ,

$$\left\| \left\| Q_t f - \tilde{\mathbb{E}}[f_t^m | Y] \right\|_{p,\infty} \right\| \leq C \sum_{i=m+1}^M t^{i/2} \|f\|_p^i, \quad (36)$$

for some constants  $C > 0$  and  $M \geq m + 1$ , where  $f_t^m$  is the truncation defined in (12). Note that the original definition (15) implies (36) due to the inequality  $\|\varphi\|_{V,i} \leq C \|\varphi\|_i$  for  $i \in \mathbb{N}$  first established in Section 2. Indeed, if  $f \in C_b^\infty(\mathbb{R}^d)$ , in other words it is constant in the  $y$  variable, then (15) and (36) actually coincide. The original Markov family  $X(x)$ , the family generated by the Lyons-Victoir method and the one generated by the Ninomiya-Victoir algorithm satisfy (36).

The following two lemmas are required to prove the main theorem below.

**Lemma 11** *Let  $\bar{X}(x)$  be an  $m$ -perfect family  $\bar{X}(x)$  that satisfies (36). Then there is a constant  $c_{25} = c_{25}(t, m, p) > 0$  such that for any  $\varphi \in C_b^\infty(\mathbb{R}^d)$  and  $r = 1, \dots, n$ , we have,*

$$\left\| \left\| \bar{R}_{\frac{t}{n}}^{r-1} R_{\frac{t}{n}}^r R_{\frac{t}{n}}^{r+1} \dots R_{\frac{t}{n}}^n \varphi - R_{\frac{t}{n}}^{r-1} R_{\frac{t}{n}}^r R_{\frac{t}{n}}^{r+1} \dots R_{\frac{t}{n}}^n \varphi \right\|_{p,\infty} \right\| \leq c_{25} n^{-(m+1)/2} \|\varphi\|_M.$$

**Proof.** Again, using the variational argument in Friedman[10] p.122 (5.17) one can check that for any  $\varphi \in C_b^\infty(\mathbb{R}^d)$  and  $r = 1, \dots, n$  the functions,

$$\exp(h_{r-1}) R_{\frac{t}{n}}^r R_{\frac{t}{n}}^{r+1} \dots R_{\frac{t}{n}}^n \varphi \in C_b^{Y,\infty}(\mathbb{R}^d).$$

Moreover there is a constant  $c_{26} \equiv c_{26}(M, p)$  such that for any  $\varphi \in C_b^\infty(\mathbb{R}^d)$  and  $r = 1, \dots, n$ ,

$$\left\| \left\| \exp(h_{r-1}) R_{\frac{t}{n}}^r R_{\frac{t}{n}}^{r+1} \dots R_{\frac{t}{n}}^n \varphi \right\|_p \right\|^M \leq c_{26} \|\varphi\|_M \quad (37)$$

Let  $\bar{X}(x)$  be an  $m$ -perfect family that satisfies (36). Then since  $X(x)$  also satisfies (36), it follows by the triangle inequality that for  $s \in [0, 1]$ ,

$$\| \|Q_s f - P_s f\| \|_{p,\infty} \leq C s^{(m+1)/2} \| \|f\| \|_p^M$$

Hence,

$$\begin{aligned} & \left\| \left\| Q_{\frac{t}{n}} \exp(h_{r-1}) R_{\frac{t}{n}}^r R_{\frac{t}{n}}^{r+1} \dots R_{\frac{t}{n}}^n \varphi - P_{\frac{t}{n}} \exp(h_{r-1}) R_{\frac{t}{n}}^r R_{\frac{t}{n}}^{r+1} \dots R_{\frac{t}{n}}^n \varphi \right\| \right\|_{p,\infty} \\ & \leq C \left( \frac{t}{n} \right)^{(m+1)/2} \left\| \left\| \exp(h_{r-1}) R_{\frac{t}{n}}^r R_{\frac{t}{n}}^{r+1} \dots R_{\frac{t}{n}}^n \varphi \right\| \right\|_p^M \end{aligned} \quad (38)$$

The result now follows from (37) and (38) and the fact that  $\bar{R}_{\frac{t}{n}}^{r-1} R_{\frac{t}{n}}^r R_{\frac{t}{n}}^{r+1} \dots R_{\frac{t}{n}}^n \varphi = Q_{\frac{t}{n}} \exp(h_{r-1}) R_{\frac{t}{n}}^r R_{\frac{t}{n}}^{r+1} \dots R_{\frac{t}{n}}^n \varphi$  and  $R_{\frac{t}{n}}^{r-1} R_{\frac{t}{n}}^r R_{\frac{t}{n}}^{r+1} \dots R_{\frac{t}{n}}^n \varphi = P_{\frac{t}{n}} \exp(h_{r-1}) R_{\frac{t}{n}}^r R_{\frac{t}{n}}^{r+1} \dots R_{\frac{t}{n}}^n \varphi$ . ■

**Lemma 12** *Let  $\bar{X}(x)$  be an  $m$ -perfect family that satisfies (36). Then there is a constant  $c_{27} \equiv c_{27}(t, m, p) > 0$  such that for any  $\varphi \in C_b^\infty(\mathbb{R}^d)$  we have,*

$$\left\| \left\| R_{\frac{t}{n}}^0 R_{\frac{t}{n}}^1 \dots R_{\frac{t}{n}}^n \varphi - \bar{R}_{\frac{t}{n}}^0 \bar{R}_{\frac{t}{n}}^1 \dots \bar{R}_{\frac{t}{n}}^n \varphi \right\| \right\|_{p,\infty} \leq c_{27} n^{-(m-1)/2} \|\varphi\|_M.$$

**Proof.** Let us observe that,

$$\begin{aligned} R_{\frac{t}{n}}^0 R_{\frac{t}{n}}^1 \dots R_{\frac{t}{n}}^n \varphi - \bar{R}_{\frac{t}{n}}^0 \bar{R}_{\frac{t}{n}}^1 \dots \bar{R}_{\frac{t}{n}}^n \varphi &= R_{\frac{t}{n}}^0 R_{\frac{t}{n}}^1 \dots R_{\frac{t}{n}}^n \varphi - \bar{R}_{\frac{t}{n}}^0 R_{\frac{t}{n}}^1 \dots R_{\frac{t}{n}}^n \varphi \\ &+ \sum_{j=1}^{n-1} \bar{R}_{\frac{t}{n}}^0 \dots \bar{R}_{\frac{t}{n}}^{j-1} \left( \bar{R}_{\frac{t}{n}}^j R_{\frac{t}{n}}^{j+1} \dots R_{\frac{t}{n}}^n \varphi - R_{\frac{t}{n}}^j R_{\frac{t}{n}}^{j+1} \dots R_{\frac{t}{n}}^n \varphi \right) \\ &+ \bar{R}_{\frac{t}{n}}^0 \dots \bar{R}_{\frac{t}{n}}^{n-1} \left( \bar{R}_{\frac{t}{n}}^n \varphi - R_{\frac{t}{n}}^n \varphi \right) \end{aligned}$$

Also note that for  $p \geq 1$  and  $r = 1, \dots, n$ ,

$$\begin{aligned} & \left\| \left\| \bar{R}_{\frac{t}{n}}^0 \dots \bar{R}_{\frac{t}{n}}^{j-1} \left( \bar{R}_{\frac{t}{n}}^j R_{\frac{t}{n}}^{j+1} \dots R_{\frac{t}{n}}^n \varphi - R_{\frac{t}{n}}^j R_{\frac{t}{n}}^{j+1} \dots R_{\frac{t}{n}}^n \varphi \right) \right\| \right\|_{p,\infty}^p \\ & \leq c_{28} \tilde{\mathbb{E}} \left[ \left\| \left\| \bar{R}_{\frac{t}{n}}^{r-1} R_{\frac{t}{n}}^r R_{\frac{t}{n}}^{r+1} \dots R_{\frac{t}{n}}^n \varphi - R_{\frac{t}{n}}^{r-1} R_{\frac{t}{n}}^r R_{\frac{t}{n}}^{r+1} \dots R_{\frac{t}{n}}^n \varphi \right\| \right\|_{p,\infty}^p \right] \\ & \leq c_{28} n^{-(m+1)/2} \|\varphi\|_M. \end{aligned}$$

where  $c_{28} \equiv c_{28}(t, m, p) > 0$  is a constant independent of  $\varphi$  and  $r$ . The claim follows. ■  
Finally let us define now the measures,

$$\bar{\rho}_t^n(\varphi) = \tilde{\mathbb{E}} \left[ \bar{R}_{\frac{t}{n}}^0 \bar{R}_{\frac{t}{n}}^1 \dots \bar{R}_{\frac{t}{n}}^n \varphi(X_0) \middle| \mathcal{Y}_t \right] \quad \text{for } \varphi \in C_b(\mathbb{R}^d).$$

and let  $\bar{\pi}_t^n = \bar{\rho}_t^n / \bar{\rho}_t^n(1)$  be its normalized version.

**Theorem 13** *Let  $\bar{X}(x)$  be an  $m$ -perfect family that satisfies (36). Then there is a constant  $c_{29} \equiv c_{29}(t, m, p) > 0$  such that for any  $\varphi \in C_b^\infty(\mathbb{R}^d)$  we have,*

$$\|\rho_t^n(\varphi) - \bar{\rho}_t^n(\varphi)\|_p \leq c_{29} n^{-(m-1)/2} \|\varphi\|_M.$$

**Proof.** We have for  $p \geq 1$ ,

$$\begin{aligned} \|\rho_t^n(\varphi) - \bar{\rho}_t^n(\varphi)\|_p &\leq \tilde{\mathbb{E}} \left[ \tilde{\mathbb{E}} \left[ \left( R_{\frac{t}{n}}^0 R_{\frac{t}{n}}^1 \dots R_{\frac{t}{n}}^n \varphi - \bar{R}_{\frac{t}{n}}^0 \bar{R}_{\frac{t}{n}}^1 \dots \bar{R}_{\frac{t}{n}}^n \varphi \right) (X_0) \middle| \mathcal{Y}_t \right]^p \right]^{\frac{1}{p}} \\ &\leq \left\| \left\| R_{\frac{t}{n}}^0 R_{\frac{t}{n}}^1 \dots R_{\frac{t}{n}}^n \varphi - \bar{R}_{\frac{t}{n}}^0 \bar{R}_{\frac{t}{n}}^1 \dots \bar{R}_{\frac{t}{n}}^n \varphi \right\| \right\|_{p, \infty}, \end{aligned}$$

hence the result. ■

**Corollary 14** *Let  $\bar{X}(x)$  be an  $m$ -perfect family that satisfies (36) with  $m \geq 3$  and assume that  $X_0$  has all moments finite. Then for any  $\varphi \in C_b^\infty(\mathbb{R}^d)$  there is a constant  $c_{30} \equiv c_{30}(t, m, p, \varphi) > 0$  such that,*

$$\|\pi_t^n(\varphi) - \bar{\pi}_t^n(\varphi)\|_p \leq \frac{c_{30}(\varphi)}{n}.$$

**Proof.** Since,

$$|\pi_t(\varphi) - \bar{\pi}_t^n(\varphi)| \leq \frac{1}{\rho_t(1)} |\rho_t(\varphi) - \bar{\rho}_t^n(\varphi)| + \frac{\|\varphi\|}{\rho_t(1)} |\rho_t(1) - \bar{\rho}_t^n(1)|$$

the Corollary follows by applying Theorem 13 and the fact that  $\|\rho_t(1)^{-1}\|_p$  is finite for any  $p \geq 0$ . ■

## 5 Appendix

The main aim of the Appendix is to deduce inequality (41). In the following we will adopt the framework of Kusuoka-Stroock[23]. Let  $(\Theta, \mathcal{B}, \mathcal{W})$  be the standard Wiener space with continuous paths  $\theta : [0, \infty) \rightarrow \mathbb{R}^d$  satisfying  $\theta(0) = 0$ . Then  $\Theta$  with the topology of uniform convergence on compact intervals is a Polish space. Also let  $H \subset \Theta$  be the Hilbert space of absolutely continuous functions  $h \in \Theta$  such that  $\|h\|_H = (\int_0^\infty |h'(t)|^2 dt)^{1/2} < \infty$ .

Let  $W^1(\mathbb{R})$  denote the space of measurable  $\Phi : \Theta \rightarrow \mathbb{R}$  with the following two properties:

- (i) For all  $h \in H$ , there exists a measurable function  $\bar{\Phi}_h : \Theta \rightarrow \mathbb{R}$  such that  $\bar{\Phi}_h = \Phi$ ,  $\mathcal{W} - a.s.$  and  $t \in \mathbb{R} \rightarrow \bar{\Phi}_h(\theta + th) \in \mathbb{R}$  is strictly absolutely continuous for all  $\theta$ .
- (ii) There exist a measurable map,  $D\Phi : \Theta \rightarrow \mathbb{L}(H; \mathbb{R})$  such that, for all  $h \in H$  and  $\varepsilon > 0$ ,

$$\lim_{|t| \downarrow 0} \mathcal{W} \left( \theta : \left| \frac{\Phi(\theta + th) - \Phi(\theta)}{t} - D\Phi(\theta)(h) \right| \geq \varepsilon \right) = 0.$$



On  $W^1(\mathbb{R})$  the norm  $\|\cdot\|_{q;\mathbb{R}}^{(n)}$  is defined as follows,

$$\|\Phi\|_{q;\mathbb{R}}^{(n)} := \left\| \|\Phi\|_{\mathbb{R}}^{(n)} \right\|_{L^q(\mathcal{W})},$$

where  $\|\Phi\|_{\mathbb{R}}^{(n)} = \sum_{m=0}^n \|D^m \Phi\|_{\mathcal{H}^m(\mathbb{R})}$  for  $q \in [2, \infty)$ . Note that  $D^0 \Phi = \Phi$  with  $\|D^0 \Phi\|_{\mathcal{H}^0(\mathbb{R})} = |\Phi|$  so  $\mathbb{E}[|\Phi|] \leq \|\Phi\|_{q;\mathbb{R}}^{(n)}$ . Finally let  $\mathcal{G}(\mathcal{L})$  be the set of all  $\Phi \in W^1(\mathbb{R})$  to which  $D$  and the Ornstein-Uhlenbeck operator  $\mathcal{L}$  as defined in [21] can be applied infinitely often.

**Definition 15** We say that  $f \in \eta_r(\mathbb{R}^d; \mathbb{R})$  for  $r \in \mathbb{Z}$  if  $f$  is a measurable map from  $(0, \infty) \times \mathbb{R}^d \times \Omega$  into  $\mathbb{R}$  such that,

- (i)  $f(t, \cdot, \omega) : \mathbb{R}^d \rightarrow \mathbb{R}$  is smooth for each  $t \in (0, \infty)$  and  $\mathcal{W}$ -a.e  $\omega \in \Omega$ .
- (ii)  $f(\cdot, x, \cdot) : (0, \infty) \times \Omega \rightarrow \mathbb{R}$  is progressively measurable for each  $x \in \mathbb{R}^d$ .
- (iii)  $\frac{\partial^\alpha}{\partial x^\alpha} f(t, x, \cdot) \in \mathcal{G}(\mathcal{L})$  and is continuous in  $t \in (0, \infty)$  for any  $\alpha \in \mathcal{A}_1, x \in \mathbb{R}^d$
- (iv)  $\sup_{0 < t \leq T} \sup_{x \in \mathbb{R}^d} \frac{1}{t^{r/2}} \left\| \frac{\partial^\alpha}{\partial x^\alpha} f(t, x, \omega) \right\|_{q;\mathbb{R}}^{(k)} < \infty$  for any  $\alpha \in \mathcal{A}_1, k \in \mathbb{N}, T > 0$  and  $2 \leq q < \infty$ .

For a proof of the following Theorem see Kusuoka-Stroock [23], Theorem 2.15 (p. 405).

**Theorem 16** (Kusuoka-Stroock) Under the UFG condition, for any  $\Phi \in \eta_r(\mathbb{R}^d; \mathbb{R}), r \in \mathbb{Z}$  and  $\alpha \in \mathcal{A}_1$  there exists  $\bar{\Phi}_\alpha \in \eta_{r-\|\alpha\|}(\mathbb{R}^d; \mathbb{R})$  such that,

$$V_{[\alpha]} \mathbb{E}[\Phi(t, x) f(X_t(x))] = \mathbb{E}[\bar{\Phi}_\alpha(t, x) f(X_t(x))] \quad (39)$$

for any  $f \in C_b^\infty(\mathbb{R}^d), t > 0, x \in \mathbb{R}^d$ .

The following immediate extension of Theorem 16 holds under the additional condition V0.

**Lemma 17** Under the UFG+V0 condition, for any  $\Phi \in \eta_r(\mathbb{R}^d; \mathbb{R})$  there exists  $\bar{\Phi}_\alpha \in \eta_{r-2}(\mathbb{R}^d; \mathbb{R})$  such that,

$$V_0 \mathbb{E}[\Phi(t, x) f(X_t(x))] = \mathbb{E}[\bar{\Phi}_0(t, x) f(X_t(x))] \quad (40)$$

for any  $f \in C_b^\infty(\mathbb{R}^d), t > 0, x \in \mathbb{R}^d$ .

**Proof.** From Theorem 16 we get that for any  $\Phi \in \eta_r(\mathbb{R}^d; \mathbb{R}), r \in \mathbb{Z}$  and  $\beta \in \mathcal{A}_1$  (2) there exists  $\bar{\Phi}_\beta \in \eta_{r-\|\beta\|}(\mathbb{R}^d; \mathbb{R})$  such that,

$$V_{[\beta]} \mathbb{E}[\Phi(t, x) f(X_t(x))] = \mathbb{E}[\bar{\Phi}_\beta(t, x) f(X_t(x))]$$

for any  $f \in C_b^\infty(\mathbb{R}^d)$ ,  $t > 0$ ,  $x \in \mathbb{R}^d$ . Hence,

$$\begin{aligned} V_0 \mathbb{E}[\Phi(t, x) f(X_t(x))] &= \sum_{\beta \in \mathcal{A}_1(2)} \varphi_\beta V_{[\beta]} \mathbb{E}[\Phi(t, x) f(X_t(x))] \\ &= \mathbb{E}[\bar{\Phi}_0(t, x) f(X_t(x))] \end{aligned}$$

where  $\bar{\Phi}_0(t, x) = \sum_{\beta \in \mathcal{A}_1(2)} \varphi_\beta \bar{\Phi}_\beta(t, x) \in \eta_{r-2}(\mathbb{R}^d; \mathbb{R})$ . ■

The following Corollary is a generalization of Corollary 2.19 (p.407) in Kusuoka-Stroock[23].

**Corollary 18** *For any  $r \in \mathbb{Z}^+$ ,  $\{\alpha_i \in \mathcal{A}\}_{i=1}^r$  and  $\|\alpha_1 * \dots * \alpha_r\|$ , there exists a constant  $C_p^T > 0$  such that,*

$$\|V_{[\alpha_1]} \dots V_{[\alpha_r]} P_t \varphi\|_\infty \leq C_p^T t^{(p - \|\alpha_1 * \dots * \alpha_r\|)/2} \|\varphi\|_p, \quad t \in [0, T]. \quad (41)$$

for  $\varphi \in C_b^\infty(\mathbb{R}^d)$  where  $\|\cdot\|_p$  for  $p \in \mathbb{N}$  is defined in (4).

**Proof.**

For the case  $p = 1$ , let  $Y(x)$  be the matrix valued process  $Y_t^{i,j}(x) := \frac{\partial}{\partial x_j}(X^i(x))$  where  $i, j = 1, \dots, d$ . (see Ikeda & Watanabe[13], Chapter V, for details). Then for any  $i, j = 1, \dots, d$  we have  $Y^{i,j}(t, x) \in \eta_0(\mathbb{R}^d; \mathbb{R})$ , in particular,

$$\sup_{t \in [0, T]} \sup_{x \in \mathbb{R}^d} \mathbb{E} \left[ \sum_{i=1}^d Y^{i,j}(x) \right] \leq c(T). \quad (42)$$

Differentiating under the integral sign (see Friedman[10], p.122 (5.17)) gives,

$$\frac{\partial}{\partial x_j} \mathbb{E}[\varphi(X_t(x))] = \sum_{i=1}^d \mathbb{E} \left[ \frac{\partial \varphi}{\partial x_i}(X_t(x)) Y_t^{i,j}(x) \right]. \quad (43)$$

Hence,

$$V_{[\alpha_r]} P_t \varphi(x) = \sum_{i=1}^d \mathbb{E} \left[ \frac{\partial \varphi}{\partial x_i}(X_t(x)) \Phi^i(t, x) \right]$$

with  $\Phi^i(t, x) = \sum_{j=1}^d V_{[\alpha_r]}^j(x) Y_t^{i,j}(x) \in \eta_0(\mathbb{R}^d; \mathbb{R})$  for  $i = 1, \dots, d$ . Then by Theorem 16 & Lemma 17 there exists  $\bar{\Phi}_\alpha^i \in \eta_{-(\|\alpha_1 * \dots * \alpha_{r-1}\|)}(\mathbb{R}^d; \mathbb{R})$ ,  $i = 1, \dots, d$  such that, for any  $\varphi \in C_b^1(\mathbb{R}^d)$ ,

$$\begin{aligned} V_{[\alpha_1]} \dots V_{[\alpha_{r-1}]} V_{[\alpha_r]} P_t \varphi(x) &= \sum_{i=1}^d V_{[\alpha_1]} \dots V_{[\alpha_{r-1}]} \mathbb{E} \left[ \frac{\partial \varphi}{\partial x_i}(X_t(x)) \Phi^i(t, x) \right] \\ &= \sum_{i=1}^d \mathbb{E} \left[ \bar{\Phi}_\alpha^i(t, x) \frac{\partial \varphi}{\partial x_i}(X_t(x)) \right] \end{aligned}$$

for any  $\varphi \in C_b^\infty(\mathbb{R}^d)$ . Hence,

$$|V_{[\alpha_1]} \dots V_{[\alpha_r]} P_t \varphi(x)| \leq \|\varphi\|_1 \sum_{i=1}^d \mathbb{E}[|\bar{\Phi}_\alpha^i(t, x)|]$$

and (41) follows from the uniform bound (in  $(t, x)$ ) on the  $L^1$ -norm of  $\bar{\Phi}_\alpha^i(t, x)$ ,  $i = 1, \dots, d$ .

For  $p > 1$  one can obtain an analogue of (43) with higher derivatives of  $P_t \varphi$  in terms of derivatives of  $\varphi$  and a set of processes analogous to  $Y$ . The proof follows in a similar manner. ■

**Acknowledgments** The authors wish to thank the anonymous referee for his comments. We would also like to thank Alan Bain for carefully reading the paper and his many pertinent remarks.

## References

- [1] Bally, V. and Talay, D., *The Law of the Euler scheme for stochastic differential equations, I. Convergence Rate of the distribution function*, Probab. Th. Rel. Fields. **104**, pp. 43-60 (1996).
- [2] Ben Arous, G., *Flots et series de Taylor stochastiques*, Probab. Th. Rel. Fields. **81**, pp. 29-77 (1989).
- [3] Burrage, K. and Burrage, P. M., *High strong order methods for non-commutative stochastic ordinary differential equation systems and the Magnus formula*,. Phys. D **133** (1999), no. 1-4, pp. 34–48.
- [4] Burrage, K., and Burrage P.M., *Order conditions of stochastic Runge-Kutta methods by B-series*, *SIAM J. Numer. Analysis* **38**, pp. 1626-1646 (2000).
- [5] Burrage, K., Burrage P.M. and Tian, T., *Numerical Methods for strong solutions of stochastic differential equations: an overview*, Proc. R. Soc. Lond. A (2004), **460**, pp. 373-402.
- [6] Castell, F. *Asymptotic expansion of stochastic flows*, Probab. Theory Relat. Fields **96**, pp. 225-239 (1993).
- [7] Castell, F. and Gaines, *An efficient approximation method for stochastic differential equations by means of exponential Lie series*, Mathematics and computers in Simulation **38** (1995), pp. 13-19.

- [8] Clark, J. M. C., *An efficient approximation for a class of stochastic differential equations*, Advances in Filtering and Optimal Stochastic Control, pp. 69-78, Lecture Notes in Control and Inform. Sci. **42**, Springer, Berlin 1982.
- [9] Clark, J. M. C., *Asymptotically Optimal Quadrature Formulae for Stochastic Integrals*, Proceedings of the 23rd Conference on Decision and Control, Las Vegas, Nevada, IEEE, 1984, pp. 712-715.
- [10] Friedman, A., *Stochastic Differential Equations and Applications I*, Academic, NY (1975).
- [11] Gaines, G. J., *The Algebra of Iterated Stochastic Integrals*, Stochastics and Stochastics Reports, Vol. 49, pp. 169-179 (1994).
- [12] Ghazali, S., *The Global Error in Weak Approximations of Stochastic Differential Equations*, PhD Thesis, Imperial College London (2006).
- [13] Ikeda, N. and Watanabe, N., *Stochastic Differential Equations and Diffusion Processes*, Second Edition, Elsevier (1989).
- [14] Jerison, D. and Sánchez-Calle, A., *Estimates for the heat kernel for a sum of squares of vector fields*, Indiana Univ. Math. J. **35** (1986), no. 4, pp. 835–854.
- [15] Kloeden, P.E. and Platen, E., *Stratonovich and Itô Stochastic Taylor Expansions*, Math. Nachr. **151** (1991), pp. 33-50.
- [16] Kloeden, P.E. and Platen, E., *Numerical Solution of Stochastic Differential Equations*, Applications of Mathematics **23**, Springer-Verlag, NY (1992).
- [17] Kusuoka, S., *Approximation of Expectation of Diffusion Process and Mathematical Finance*, Advanced Studies in Pure Mathematics **31** (2001).
- [18] Kusuoka, S., *Approximation of Expectation of Diffusion Processes based on Lie Algebra and Malliavin Calculus*, UTMS 2003-34 (2003).
- [19] Kusuoka, S., *Malliavin Calculus Revisited*, J. Fac. Sci. Univ. Tokyo, **10** (2003), pp. 261-277.
- [20] Kusuoka, S. and Ninomiya, S., *A new simulation method of diffusion processes applied to finance*, Stochastic processes and applications to mathematical finance, pp. 233–253, World Sci. Publ., River Edge, NJ, 2004.
- [21] Kusuoka, S. and Stroock, D.W., *Applications of the Malliavin Calculus, Part I*, Stochastic Analysis, Proc. Taniguchi Internatl. Symp. Katata and Kyoto, 1982, (ed. by K. Itô), Kinokuniya, Tokyo, 1981, pp. 271-306.

- [22] Kusuoka, S. and Stroock, D.W., *Applications of the Malliavin Calculus, Part II*, J. Fac. Sci. Univ. Tokyo, Sect 1A, Math. **32** (1985), pp. 1-76.
- [23] Kusuoka, S. and Stroock, D.W., *Applications of the Malliavin Calculus, Part III*, J. Fac. Sci. Univ. Tokyo, Sect 1A, Math. **34** (1987), pp. 391-442.
- [24] Lyons, T. and Victoir, N., *Cubature on Wiener Space*, Proc. R. Soc. Lond. A (2004), pp. 169-198.
- [25] Magnus, W., *On the exponential solution of differential equations for a linear operator*, Commun. Pure Appl. Math. **7**, (1954) pp. 649-673.
- [26] Millet, A. and Sanz-Sole, M., *A simple proof of the support theorem for diffusion processes*, Séminaire de probabilités, XXVIII, pp. 36-48, Lecture Notes in Math., 1583, Springer, Berlin (1994).
- [27] Milstein, G. N., *A Method of second-order accuracy integration of stochastic differential equations*, Theory Probab. Appl. **23**, pp. 396-401 (1976).
- [28] Milstein, G. N., *Weak approximation of solutions of systems of stochastic differential equations*, Theory Probab. Appl. **30**, pp. 750-766 (1985).
- [29] Newton, N.J., *An asymptotically efficient difference formula for solving stochastic differential equations*, Stochastics **19** (1986), No. 3, pp. 175-206.
- [30] Newton, N.J., *An efficient approximation for stochastic differential equations on the partition of symmetrical first passage times*, Stochastics Stochastics Rep. **29** (1990), No. 2, pp. 227-258.
- [31] Ninomiya, S., *A New Simulation Scheme of Diffusion Processes: Application of the Kusuoka Approximation to Finance Problems*, 3rd IMACS Seminar on Monte Carlo Methods—MCM 2001 (Salzburg). Math. Comput. Simulation 62 (2003), no. 3-6, pp. 479-486.
- [32] Ninomiya, S., *A partial sampling method applied to the Kusuoka approximation*, Monte Carlo Methods Appl. 9 (2003), no. 1, pp. 27-38.
- [33] Ninomiya, S. and Victoir, N., *Weak Approximations of Stochastic Differential Equations and Application to Derivative Pricing*, [www.crest.fr/pageperso/elic/bachelier\\_fichiers/Programme\\_Fichiers/oxford2004.pdf](http://www.crest.fr/pageperso/elic/bachelier_fichiers/Programme_Fichiers/oxford2004.pdf).
- [34] Øksendal, B., *Stochastic differential equations : an introduction with applications*, 5th ed, Springer-Verlag, 2002.
- [35] Pettersson, R., *Stratonovich Taylor Expansion and Numerical Methods*, Stochastic Analysis and Applications, 10(5), pp. 603-612 (1992).

- [36] Picard, J., *Approximation of nonlinear filtering problems and order of convergence*, Filtering and control of random processes (Paris, 1983), pp. 219–236, Lecture Notes in Control and Inform. Sci., 61, Springer, Berlin, 1984.
- [37] Platen, E. and Wagner, W., *On a Taylor formula for a class of Itô Processes*, Probability and Mathematical Statistics, Vol 3, 1982, pp. 37-51.
- [38] Sánchez-Calle, A., *Fundamental solutions and geometry of the sum of squares of vector fields*, Invent. Math. 78 (1984), no. 1, pp. 143–160.
- [39] Stroock, D.W. and Varadhan, S.R.S., *Diffusion processes with continuous coefficients*, I, Comm. Pure Appl. Math., Vol. 22, 1969, pp. 345-400.
- [40] Stroock, D.W. and Varadhan, S.R.S., *On the support of diffusion processes with applications to the strong maximum principle*, Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Vol. III: Probability Theory, pp. 333-359 (1970).
- [41] Stroock, D.W. and Varadhan, S.R.S., *On degenerate elliptic-parabolic operators of second order and their associated diffusions*, Comm. Pure Appl. Math. XXII, (1972), pp. 651-714.
- [42] Talay, D., *Efficient numerical schemes for the approximation of expectations of functionals of the solution of an SDE and applications*, Filtering and Control of Random Processes, Vol. 61, Lecture Notes in Control and Inform. Sci., pp. 294-313, Springer (1984).
- [43] Talay, D., *Discrétisation d'une E.D.S. et calcul approché d'espérances de fontionelles de la solution*, Math. Modelling Numer. Anal. **20**, pp. 141-179 (1986).
- [44] Talay, D. and Tubaro, L., *Expansion of the Global Error for Numerical Schemes Solving Stochastic Differential Equations*, Stochastic Anal. Appl. (1990), Vol. 8, No. 4, pp. 483-509.