

# CUBATURE ON WIENER SPACE FOR MCKEAN–VLASOV SDES WITH SMOOTH SCALAR INTERACTION<sup>1</sup>

BY DAN CRISAN AND EAMON MCMURRAY

*Imperial College London*

We present two cubature on Wiener space algorithms for the numerical solution of McKean–Vlasov SDEs with smooth scalar interaction. First, we consider a method introduced in [Stochastic Process. Appl. **125** (2015) 2206–2255] under a uniformly elliptic assumption and extend the analysis to a uniform strong Hörmander assumption. Then we introduce a new method based on Lagrange polynomial interpolation. The analysis hinges on sharp gradient to time-inhomogeneous parabolic PDEs bounds. These bounds may be of independent interest. They extend the classical results of Kusuoka and Stroock [J. Fac. Sci., Univ. Tokyo, Sect. 1A, Math. **32** (1985) 1–76] and Kusuoka [J. Math. Sci. Univ. Tokyo **10** (2003) 261–277] further developed in [J. Funct. Anal. **263** (2012) 3024–3101; J. Funct. Anal. **268** (2015) 1928–1971; Cubature Methods and Applications (2013), Springer, Cham] and, more recently, in [Probab. Theory Related Fields **171** (2016) 97–148]. Both algorithms are tested through two numerical examples.

**1. Introduction.** In this paper, we analyse the error in two different algorithms that use cubature on Wiener space to weakly approximate the solution of a McKean–Vlasov SDE with smooth scalar interaction. By scalar interaction, we mean that the dependence on the measure is through the integral against a scalar function, so the McKean–Vlasov SDE takes the form

$$(1) \quad X_t^x = x + \int_0^t V_0(X_s^x, \mathbb{E}[\varphi_0(X_s^x)]) ds + \sum_{i=1}^d \int_0^t V_i(X_s^x, \mathbb{E}[\varphi_i(X_s^x)]) \circ dB_s^i,$$

where  $\varphi_i \in C_b^\infty(\mathbb{R}^N; \mathbb{R})$ ,  $V_i \in C_b^\infty(\mathbb{R}^{N+1}; \mathbb{R}^N)$  and  $B = (B^1, \dots, B^d)$  is a Brownian motion. We wish to approximate  $\mathbb{E}[f(X_T^x)]$  for  $f$  Lipschitz continuous and  $T > 0$  a fixed time.

One common way of approaching this problem is to consider a discretisation of the equation, such as the Euler–Maruyama scheme, along with a Monte Carlo approximation. At each time step, an approximation of the law of  $X_T^x$  is then given

---

Received August 2016; revised January 2018.

<sup>1</sup>Supported in part by the UK EPSRC Grant no. EP/M506345/1.

*MSC2010 subject classifications.* Primary 60H30, 60H10, 60H07; secondary 34F05, 35K55, 35K65.

*Key words and phrases.* Cubature on Wiener space, Lagrange polynomial interpolation, McKean–Vlasov SDEs, Kusuoka–Stroock functions.

by the empirical distribution of the entire Monte Carlo population. However, estimating the error due to approximating the expectation inside the coefficients by the Monte Carlo estimator is exactly the problem one wishes to solve in the first place. This leads to a more difficult analysis than for classical SDEs. Nonetheless, this analysis has been carried out under a number of different assumptions when the coefficients have the form

$$V_i \left( X_t^x, \int h_i(X_t^x - y) P_{X_t^x}(dy) \right)$$

[different from (1)]. This type of scheme was studied in papers by Bossy, alone [4] and along with Talay [5]; Kohatsu-Higa and Ogawa [23], and Antonelli and Kohatsu-Higa [2]. In all of these papers, the total error is composed of a discretisation error of order  $\sqrt{h}$  or  $h$ , where  $h$  is size of the largest time step, and statistical error of order  $N_{MC}^{-1/2}$  where  $N_{MC}$  is the number of Monte Carlo samples. A Milstein discretisation is also analysed in Ogawa [39]. In [41], Tachet des Combes proposes a deterministic numerical scheme based on discretising the PDE satisfied by the density function of the solution to (1). More recently, in [40], a multi-level Monte Carlo scheme has been analysed for equations of the type (1).

Systems of the time (1) have been studied in [26]. The particle representation of the solution of equation (1) appears explicitly in [28]. A generalization of the model can be found in [20]. Section 1.8 in [25] presents the origins of this model which can be traced back to the Vlasov equation for plasma which was then augmented by McKean with a stochastic term. In the same section, the Landau–Fokker–Planck equation is discussed. However, the nonlinearity is different here as, in the case of Landau–Fokker–Planck, the coefficients depend on the convolutions (shifted expectations), rather than just the expectations. The introduction of diffusive noise can be considered as arising from natural scaling of various finite state space models; see Section 5 in [24]. Equation (1) appears in many applications: for example, see [37] for an applications to Mathematics Biology, [6, 34] and the more recent [8] for applications to mean field games and [11] for an application to asset pricing through competing traders valuations.<sup>2</sup>

Cubature on Wiener space is a high-order alternative to Monte Carlo methods. It is part of a class of methods called Kusuoka–Lyons–Victoir methods that have been shown to be highly effective in practice; see, for example, [19], [38]. Applications include the nonlinear filtering problem [10, 17, 33, 35], backward stochastic differential equations [13, 14] and calculating Greeks [42] in finance. Convergence of the cubature approximation for some path dependent functionals has also been shown in [3]. The starting point of the cubature on Wiener space method is to view  $X_T^x$  as a functional of the Brownian path  $(B_s)_{s \in [0, T]}$ , say

$$X_T^x = \alpha_{x, T}((B_s)_{s \in [0, T]}), \quad \alpha_{x, T} : \mathcal{C}_0([0, T]; \mathbb{R}^d) \rightarrow \mathbb{R}^N,$$

---

<sup>2</sup>The authors would like to thank Vassili Kolokoltsov and Tom Kurtz for pointing out relevant references to the model considered here.

and to view the expectation  $\mathbb{E}[f(X_T^x)]$  as an integral over the Wiener space

$$(2) \quad \mathbb{E}[f(X_T^x)] = \int_{\Omega} (f \circ \alpha_{x,T})(\omega) \mathbb{P}(d\omega),$$

where  $\Omega = \mathcal{C}_0([0, T]; \mathbb{R}^d)$  and  $\mathbb{P}$  is the Wiener measure. The key idea of the cubature on Wiener space method is that one can approximate such integrals by replacing the Wiener measure,  $\mathbb{P}$ , by a discrete measure supported on finitely many bounded variation paths, called a cubature measure. If the cubature measure is chosen so that iterated Stratonovich integrals of Brownian motion up to some order have the same expectation under the cubature and Wiener measures and the time interval  $[0, T]$  is small, then by considering the Stratonovich–Taylor expansion of the solution of the SDE, one can show that the target expectations under the cubature and Wiener measures agree up to some high order error. When the time interval is not small,  $[0, T]$  can be partitioned into sub-intervals and the approximation performed over each sub-interval. In the original paper [36], dealing with ordinary SDEs, evaluating the functional  $\alpha_{x,T}$  at a bounded variation path  $\omega$  amounts to solving an ordinary differential equation (ODE) and evaluating integrals such as (2) under a cubature measure amounts to computing weighted sums of solutions of ODEs. The complication for McKean–Vlasov equations like (1) is that the functional  $\alpha_{x,T}$  depends on the paths  $\{(\mathbb{E}\varphi_i(X_s^x))_{s \in [0, T]} : i = 0, \dots, d\}$  which are unknown. Instead, one must include an approximation of the functional  $\alpha_{x,T}$  in the design of the algorithm.

To our knowledge, the first algorithm involving cubature on Wiener Space in relation to McKean–Vlasov SDE was introduced by Chaudru de Raynal and Garcia-Trillos [18]. Their idea is to partition  $[0, T]$  into  $\{0 = t_0 < t_1 < \dots < t_n = T\}$  and over the interval  $[t_j, t_{j+1}]$  to replace  $\mathbb{E}\varphi_i(X_t^x)$  appearing in the coefficients with the cubature approximation of the Taylor expansion of the path  $t \mapsto \mathbb{E}\varphi_i(X_t^x)$  around  $t_j$  up to some order,  $q$ . The global error can, as in the original case, be decomposed as a sum of local errors, and these local errors naturally split into an error due to the approximation of  $\mathbb{E}\varphi_i(X_t^x)$  in the coefficients, and an error due to replacing the Wiener measure by a cubature measure. In [18], the authors consider the case of smooth and bounded uniformly elliptic coefficients and prove that the error is of order  $n^{-[(q+1) \wedge (l-1)/2]}$  where  $n$  is the number of time steps and  $l$  is the degree of the cubature formula. In this paper, we show how to extend the error analysis to the case when the coefficients satisfy a uniform strong Hörmander condition. One of the reasons the authors of [18] choose to impose a uniformly elliptic condition on the coefficients of equation (1) is the lack of available sharp derivative estimates for time-inhomogeneous parabolic PDEs (which are necessary for the error analysis) under any more general conditions. For this reason, a secondary goal of this work is to develop derivative estimates for time-inhomogeneous parabolic PDEs under more general conditions and to analyse the error for the cubature on Wiener space algorithm in this case. These estimates may be of independent interest. They

extend the classical results of Kusuoka and Stroock [31] and Kusuoka [30], further developed in [9, 12, 15] and, more recently, in [16].

In the second algorithm, which we call the Lagrange interpolation method, over the interval  $[t_j, t_{j+1}]$ , one simply replaces  $\mathbb{E}\varphi_i(X_t^x)$  with the Lagrange polynomial which interpolates the cubature approximation of the quantity  $\mathbb{E}\varphi_i(X_t^x)$  at the previous  $r$  points in the time partition.

For both algorithms, we deduce upper bounds for their corresponding rates of convergence when approximating  $\mathbb{E}[f(X_T^x)]$ . The bounds are presented as explicit functions of the type of partition used, the number of points in the partition, the smoothness of the the function  $f$ , the cubature order, the order of the Taylor expansion (for the first algorithm) and, respectively, the order of the Lagrange approximation (for the Lagrange interpolation algorithm). See Theorem 1.4 below for details.

The paper is organized as follows: In the following, we present a brief introduction of cubature on the Wiener space followed by a detailed description of the paper framework, the assumptions, the algorithms and the main results. In Section 2, we collect a number of preliminary results required to prove the main result. The proof of the main result, Theorem 1.4 is done in Section 3. In Section 4, we test both algorithms on two numerical benchmarks. We complete the paper with an Appendix containing derivative bounds for time-inhomogeneous parabolic PDEs under the so-called UFG condition which may be of independent interest. These bounds are the cornerstone of the convergence analysis of the numerical algorithms. Several other generic results used in the earlier proofs are also incorporated in the Appendix.

1.1. *Cubature on Wiener space.* First, we detail what we mean by a cubature method on Wiener space. We need to introduce notation for iterated integrals with respect to components of the  $(d + 1)$ -dimensional process  $(B^0, B^1, \dots, B^d)$  consisting of time and the  $d$ -dimensional Brownian motion. We use the following notation for multi-indices on  $\{0, \dots, d\}$ :

$$\mathcal{A} := \{\emptyset\} \cup \bigcup_{k \geq 1} \{0, 1, \dots, d\}^k \quad \text{and} \quad \mathcal{A}_1 := \mathcal{A} \setminus \{\emptyset, (0)\}.$$

We endow  $\mathcal{A}$  with the concatenation operation

$$\alpha * \beta := (\alpha_1, \dots, \alpha_k, \beta_1, \dots, \beta_l), \quad \alpha = (\alpha_1, \dots, \alpha_k), \beta = (\beta_1, \dots, \beta_l) \in \mathcal{A}$$

and we define  $\alpha' = (\alpha_1)$  and  $-\alpha := (\alpha_2, \dots, \alpha_k)$ , so that  $\alpha = \alpha' * -\alpha$ . We define the following  $n$ -tuples lengths:

$$|\alpha| := \begin{cases} k, & \text{if } \alpha = (\alpha_1, \dots, \alpha_k), \\ 0, & \text{if } \alpha = \emptyset, \end{cases}$$

$$\|\alpha\| := |\alpha| + \text{card}\{i : \alpha_i = 0, i = 1, \dots, d\}$$

and define the set  $\mathcal{A}(l) := \{\alpha \in \mathcal{A} : \|\alpha\| \leq l\}$  and define  $\mathcal{A}_1(l)$  similarly. For  $\alpha \in \mathcal{A}$ , we denote by  $I_{t,s}^\alpha(Y)$  the iterated Stratonovich integral of the process  $Y$  over the interval  $[t, s]$ :

$$I_{t,s}^\alpha(Y) := \int_t^s \int_t^{s_n} \cdots \int_t^{s_2} Y_{s_1} \circ dB_{s_1}^{\alpha_1} \circ \cdots \circ dB_{s_{n-1}}^{\alpha_{n-1}} \circ dB_{s_n}^{\alpha_n}.$$

Similarly, for a bounded variation path  $\omega = (\omega^1, \dots, \omega^d) \in \mathcal{C}_{bv}([t, s]; \mathbb{R}^d)$ , we set  $\omega^0(s) = s$  and denote the iterated integral of the stochastic process  $Y$  by  $I_{t,s}^\alpha[\omega](Y)$ :

$$I_{t,s}^\alpha[\omega](Y) := \int_t^s \int_t^{s_n} \cdots \int_t^{s_2} Y_{s_1} d\omega_{s_1}^{\alpha_1} \cdots d\omega_{s_{n-1}}^{\alpha_{n-1}} d\omega_{s_n}^{\alpha_n}.$$

With this notation in hand, we can define a cubature formula (we follow here [36]).

**DEFINITION 1.1 (Cubature formula).** A set of  $N_{\text{Cub}}$  bounded variation paths,  $\omega_1, \dots, \omega_{N_{\text{Cub}}} \in \mathcal{C}_{bv}([0, 1]; \mathbb{R}^d)$ , for some  $N_{\text{Cub}} \in \mathbb{N}$ , together with some weights  $\lambda_1, \dots, \lambda_{N_{\text{Cub}}} \in \mathbb{R}^+$  such that  $\sum_{j=1}^{N_{\text{Cub}}} \lambda_j = 1$  define a cubature formula on Wiener Space of degree  $l$  if, for any  $\alpha \in \mathcal{A}(l)$ ,

$$\mathbb{E}[I_{0,1}^\alpha(1)] = \sum_{j=1}^{N_{\text{Cub}}} \lambda_j I_{0,1}^\alpha[\omega_j](1).$$

We note that for a given  $l \in \mathbb{N}$ , Lyons and Victoir [36] proved that there exists a cubature formula on Wiener Space of degree  $l$ , with concrete examples given, for certain pairs  $(l, d)$ , in [36] and [19]. From the scaling properties of the Brownian motion, we can deduce, for  $0 \leq t < s$ ,

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

where  $\omega_j(t, s)$  is the rescaled path defined by  $\omega_j(t, s)(u) = \sqrt{s-t} \omega_j(\frac{u}{s-t})$ ,  $u \in [t, s]$ . In other words, the expectation of the iterated Stratonovich integrals  $I_{t,s}^\alpha(1)$  with  $\alpha \in \mathcal{A}(l)$  is the same under the Wiener measure as it is under the cubature measure,

$$\mathbb{Q}_{t,s} := \sum_{j=1}^{N_{\text{Cub}}} \lambda_j \delta_{\omega_j(t,s)}.$$

Once we have a cubature measure  $\mathbb{Q}$  and a partition  $\Pi_n$ , we can extend this to a measure  $\mathbb{Q}^{\Pi_n}$  on  $[0, T]$ , supported on  $(N_{\text{Cub}})^n$  paths along a tree. We use the notation  $\mathcal{M}_k$  to denote multi-indices over  $\{1, \dots, N_{\text{Cub}}\}$  of length exactly  $k$ . We use this set to index the nodes in the cubature tree after  $k$  time-steps or, equivalently, the unique path leading to that node. To create the tree, one first creates the paths by concatenating the rescaled paths: for  $p = (p_1, \dots, p_n) \in \mathcal{M}_n$ , define the path

$$\omega_p(t) = \omega_p(t_{i-1}) + \omega_{p_i}(t_{i-1}, t_i)(t) \quad \text{when } t \in [t_{i-1}, t_i).$$

Then one can attach a new weight to each path by

$$\Lambda_p := \prod_{p_i \in p} \lambda_{p_i}.$$

Finally, we can define a measure on all paths along the tree by

$$\mathbb{Q}^{\Pi_n} := \sum_{p \in \mathcal{M}_n} \Lambda_p \delta_{\omega_p}.$$

1.2. *Outline and main results.* In this section, let us make more precise our contribution. We introduce the following processes:

$$(3) \quad X_t^{s,x} = x + \sum_{i=0}^d \int_s^t V_i(X_u^{s,x}, \mathbb{E}\varphi_i(X_u^{s,x})) \circ dB_u^i$$

and

$$(4) \quad X_t^{s,x,y} = x + \sum_{i=0}^d \int_s^t V_i(X_u^{s,x,y}, \mathbb{E}\varphi_i(X_u^{0,y})) \circ dB_u^i.$$

The first is just the McKean–Vlasov SDE started from  $x$  at time  $s$ . The second process starts from  $x$  at time  $s$  as well, but with the path  $u \mapsto \mathbb{E}[\varphi_i(X_u^{0,y})]$  appearing in the coefficients instead of the McKean–Vlasov term. This process is therefore not a true McKean–Vlasov process but an SDE with coefficients depending on time and a parameter,  $y$ . We introduce the operators

$$P_{s,t} f(x) := \mathbb{E}[f(X_t^{s,x})] \quad \text{and} \quad P_{s,t}^y f(x) := \mathbb{E}[f(X_t^{s,x,y})].$$

We note that  $X_T^{0,x} = X_T^{0,x,y}|_{y=x}$ , so the quantity we wish to compute is

$$P_{0,T} f(x) = P_{0,T}^y f(x)|_{y=x}.$$

Now, let us denote by  $E_t^y(\varphi_i)$  a generic approximation of  $\mathbb{E}[\varphi_i(X_t^{0,y})]$ . Later, we will introduce specific approximations  $E_t^{T,y}(\varphi_i)$  and  $E_t^{L,y}(\varphi_i)$ , corresponding to the Taylor and Lagrange interpolation methods, respectively. We then introduce the approximating process

$${}^E X_t^{s,x,y} = x + \sum_{i=0}^d \int_s^t V_i({}^E X_u^{s,x,y}, E_u^y(\varphi_i)) \circ dB_u^i,$$

and the operators

$$P_{s,t}^{E,y} g(x) := \mathbb{E}[g({}^E X_t^{s,x,y})] \quad \text{and} \quad Q_{s,t}^{E,y} g(x) := \mathbb{E}_{\mathbb{Q}_{s,t}}[g({}^E X_t^{s,x,y})].$$

In a similar way, we will denote the local approximation operator by  $Q_{s,t}^{E,x}$  and, once a partition  $\Pi_n$  of  $[0, T]$  is fixed, we define

$$Q_{0,t}^{E,x,\Pi_n} := Q_{0,t_1}^{E,x} \cdots Q_{t_j,t}^{E,x} \quad \text{for } t \in [t_j, t_{j+1}).$$

Then  $Q_{0,T}^{E,x,\Pi_n} f(x)$  will be the final approximation of  $P_{0,T} f(x)$ , with the global error

$$(5) \quad \mathcal{E}(T, x, l, \Pi_n) := (P_{0,T} - Q_{0,T}^{E,x,\Pi_n}) f(x).$$

We note that

$$\sup_{x \in \mathbb{R}^N} |\mathcal{E}(T, x, l, \Pi_n)| \leq \sup_{x, y \in \mathbb{R}^N} |(P_{0,T}^y - Q_{0,T}^{E,y,\Pi_n}) f(x)|.$$

Now, for fixed  $y$ ,  $\{P_{s,t}^y : 0 \leq s \leq t \leq T\}$  forms a two-parameter semigroup of operators. This allows us to decompose the global error of the scheme as follows:

$$\begin{aligned} & [P_{0,T}^y - Q_{0,t_1}^{E,y} Q_{t_1,t_2}^{E,y} \cdots Q_{t_{n-1},t_n}^{E,y}] f(x) \\ &= \sum_{j=0}^{N-1} Q_{0,t_j}^{E,y} [P_{t_j,t_{j+1}}^y - Q_{t_j,t_{j+1}}^{E,y}] P_{t_{j+1},T}^y f(x). \end{aligned}$$

Then, since  $\|Q_{0,t_j}^{E,y} \phi\|_\infty \leq \|\phi\|_\infty$ , we are left to estimate the *local error*

$$[P_{t_j,t_{j+1}}^y - Q_{t_j,t_{j+1}}^{E,y}] u^y(t_{j+1}, x)$$

uniformly in  $x$  and  $y$ , where

$$(6) \quad u^y(t, x) := P_{t,T}^y f(x)$$

solves a parabolic PDE with coefficients depending on the parameter  $y \in \mathbb{R}^N$ . The resulting error analysis relies on regularity estimates for the solution of this PDE.

Now, let us specify what the approximation  $E_t^x(\varphi_i)$  is for each scheme. First, the Taylor method: we wish to perform a Taylor expansion of the path  $t \mapsto \mathbb{E}\varphi_i(X_t^{0,x})$ , but since the coefficients in the SDE satisfied by  $X^{0,x}$  are of the form  $V_i(X_t^{0,x}, \mathbb{E}\varphi_i(X_t^{0,x}))$ , we instead consider the Taylor expansion of this more general form. For a pair of functions  $g \in C_b^\infty(\mathbb{R}^N \times \mathbb{R}; \mathbb{R})$  and  $\varphi \in C_b^\infty(\mathbb{R}^N; \mathbb{R})$ , Itô's formula yields

$$\begin{aligned} & \mathbb{E}[g(X_t^{0,x}, \mathbb{E}\varphi(X_t^{0,x}))] \\ (7) \quad &= g(x, \varphi(x)) + \int_0^t \mathbb{E}[(\mathcal{L}_u^x g)(X_u^{0,x}, \mathbb{E}\varphi(X_u^{0,x}))] \\ & \quad + \mathbb{E}[(\partial_y g)(X_u^{0,x}, \mathbb{E}\varphi(X_u^{0,x}))] \mathbb{E}[(\mathcal{L}_u^x \varphi)(X_u^{0,x})] du, \end{aligned}$$

where  $\partial_y$  is the derivative in the second argument of  $g$  and  $\mathcal{L}_s^x$  is the differential operator

$$\mathcal{L}_s^x := V_0(\cdot, \mathbb{E}\varphi_0(X_s^{0,x})) + \frac{1}{2} \sum_{i=1}^d V_i(\cdot, \mathbb{E}\varphi_i(X_s^{0,x}))^2.$$

Note that each term under the integral in the right-hand side of (7) is again a product of terms of the form  $\mathbb{E}[g(X_u^{0,x}, \mathbb{E}\varphi(X_u^{0,x}))]$  for different functions  $g$  and  $\varphi$ . Let us denote by  $\mathcal{S}$  the set of all paths  $(S_t)_{t \in [0, T]}$  of the form

$$S_t = \mathbb{E}[g(X_t^{0,x}, \mathbb{E}\varphi(X_t^{0,x}))]$$

for some  $g \in C_b^\infty(\mathbb{R}^N \times \mathbb{R}; \mathbb{R})$ , and  $\varphi \in C_b^\infty(\mathbb{R}^N; \mathbb{R})$  and denote by  $\bar{\mathcal{S}}$  the set of all paths  $(\bar{S}_t)_{t \in [0, T]}$  of the form

$$\bar{S}_t = \sum_{i=1}^{n_1} \prod_{j=1}^{n_2} S_t^{i,j},$$

where  $n_1$  and  $n_2$  are positive integers and  $S^{i,j} \in \mathcal{S}$  for  $i = 1, \dots, n_1$  and  $j = 1, \dots, n_2$ . For  $t \in [0, T]$ , we introduce the notation  $\mathcal{T} : \mathcal{S} \rightarrow \bar{\mathcal{S}}$  for the map

$$\begin{aligned} \mathbb{E}g(X^{0,x}, \mathbb{E}\varphi(X^{0,x})) &\mapsto \mathbb{E}[(\mathcal{L}^x g)(X^{0,x}, \mathbb{E}\varphi(X^{0,x}))] \\ &\quad + \mathbb{E}[(\partial_y g)(X^{0,x}, \mathbb{E}\varphi(X^{0,x}))]\mathbb{E}[(\mathcal{L}^x \varphi)(X^{0,x})], \end{aligned}$$

so that the expansion of  $G = (G_t)_{t \geq 0}$  defined as

$$G_t := \mathbb{E}[g(X_t^{0,x}, \mathbb{E}\varphi(X_t^{0,x}))]$$

in (7) can alternatively be written as

$$(8) \quad G_t = G_0 + \int_0^t \mathcal{T}_s(G) ds.$$

For the Taylor expansion, we would like to apply  $\mathcal{T}$  to the term  $\mathcal{T}(G)$ . To do so, we extend  $\mathcal{T}$  to an operator from  $\bar{\mathcal{S}}$  to itself by linearity

$$\mathcal{T}\left(\sum_{i=1}^{n_1} S^i\right) = \sum_{i=1}^{n_1} \mathcal{T}(S^i),$$

and a product rule

$$\mathcal{T}\left(\prod_{j=1}^{n_2} S^j\right) = \sum_{k=1}^{n_2} \prod_{j=1, j \neq k}^{n_2} S^j \mathcal{T}(S^k).$$

Since now we have  $\mathcal{T}$  defined as an operator on  $\bar{\mathcal{S}}$ , we can iterate the expansion in (8) to get the Taylor expansion of order  $q \geq 1$ ,

$$(9) \quad G_t = \sum_{k=0}^q \frac{t^k}{k!} (\mathcal{T}_0)^k(G) + \frac{1}{(q+1)!} \int_0^t (\mathcal{T}_s)^{q+1}(G) ds.$$

Now, for all  $s \in [0, T]$ ,  $k \geq 1$  and  $G \in \bar{\mathcal{S}}$  we define  $(\mathcal{T}_s^Q)^k(G)$  to be the same expression as  $(\mathcal{T}_s)^k(G)$  with all expectations under  $\mathbb{P}$  replaced by expectations



under  $\mathbb{Q}^{\Pi_n}$ . Then the approximation of  $\mathbb{E}\varphi_i(X_s^{0,y})$  for the Taylor method, which we henceforth denote by  $E_s^{T,y}(\varphi_i)$ , is

$$E_s^{T,y}(\varphi_i) := \sum_{k=0}^q \frac{1}{k!} (\mathcal{T}_{t_j}^Q)^k (\mathbb{E}\varphi_i(X^{0,y})) (s - t_j)^k \quad \text{for } s \in [t_j, t_{j+1}).$$

To define the approximation for the Lagrange interpolation method, we denote by  $L[\{(t_i, x_i)\}_{i=1}^k]$  the Lagrange interpolating polynomial of degree at most  $k - 1$  with  $L[\{(t_i, x_i)\}_{i=1}^k](t_j) = x_j$  for all  $j = 1, \dots, k$  (see p. 149 for further details). We define the approximation of order  $r$  by

$$E_t^y(\varphi_i) := L[\{(t_{j-k}, \mathbb{E}^{\mathbb{Q}^{\Pi_n}} \varphi_i(X_{t_{j-k}}^{0,y}))\}_{k=0}^{j \wedge (r-1)}](t) \quad t \in [t_j, t_{j+1}].$$

In other words,  $E_t^y(\varphi_i)$  for  $t \in [t_j, t_{j+1}]$  is the unique polynomial of degree minimal degree which passes through the points

$$(t_{j+1-(j+1) \wedge r}, \mathbb{E}^{\mathbb{Q}^{\Pi_n}} \varphi_i(X_{t_{j+1-(j+1) \wedge r}}^{0,y})), \dots, (t_j, \mathbb{E}^{\mathbb{Q}^{\Pi_n}} \varphi_i(X_{t_j}^{0,y})).$$

That is, if the time index  $j$  is greater than  $r - 1$ , we interpolate through the last  $r$  cubature approximations of  $\mathbb{E}\varphi_i(X_t^{0,y})$  along the partition. If  $j < r - 1$ , we interpolate through all of the available previous points. We now detail Algorithms 1 and 2.

**REMARK 1.2.** The Taylor method requires finding an expression for the quantity  $(\mathcal{T}_t)^k(\mathbb{E}\varphi_i(X^{0,y}))$  for  $k = 1, \dots, q$  and  $i = 0, \dots, d$  either by hand or using some symbolic computation. The Lagrange interpolation method does not require this; the interpolating polynomial is defined at each time step as part of the algorithm.

We state next the main assumptions of the paper. To do this, we introduce the notation  $V_{[\alpha]}$  for iterated Lie brackets of the vector fields. In this setting, each  $V_0, \dots, V_d : \mathbb{R}^N \times \mathbb{R} \rightarrow \mathbb{R}^N$  and we think of these as vector fields  $V_i(\cdot, x')$  on  $\mathbb{R}^N$  parametrised by the second variable,  $x' \in \mathbb{R}$ , with the Lie bracket between any two given by

$$[V_i, V_j](x, x') = \partial_x V_j(x, x') V_i(x, x') - \partial_x V_i(x, x') V_j(x, x'),$$

where  $\partial V_i(x, x') := (\partial_{x_l} V_i^k(x, x'))_{1 \leq k, l \leq N}$  is the Jacobian matrix of  $V_i$  and similarly for  $\partial_x V_j$ . Then, for  $\alpha \in \bigcup_{k \geq 1} \{1, \dots, N\}^k$  and  $i \in \{1, \dots, N\}$ , we define inductively

$$V_{[i]} := V_i, \quad V_{[\alpha * i]} := [V_i, V_{[\alpha]}].$$

**Algorithm 1** Taylor method

- 
- 1: Set  $(X^\emptyset, \Lambda_0) = (x, 1)$
  - 2: **for**  $0 \leq j \leq n - 1$  **do**
  - 3:     Let  $\mathcal{T}^Q$  be as in the main text.
  - 4:     **for**  $p \in \mathcal{M}_j$  **do**
  - 5:         **for**  $0 \leq i \leq d$  **do**
  - 6:             Set  $E_i^T(t) = \sum_{k=0}^q \frac{1}{k!} (t - t_j)^k (\mathcal{T}_{t_j}^Q)^k (\mathbb{E} \varphi_i(X^{0,x}))$
  - 7:             **end for**
  - 8:         **for**  $1 \leq l \leq N_{\text{cub}}$  **do**
  - 9:             Define  $X_{t_{j+1}}^{p*l}$  as the solution of the ODE:
$$dX_t^{p*l} = \sum_{i=0}^d V_i(X_t^{p*l}, E_i^T(t)) d\omega_t^i(t_j, t_{j+1})(t),$$

$$X_{t_j}^{p*l} = X_{t_j}^p.$$
  - 10:             Set the associated weight:  $\Lambda_{p*l} = \Lambda_p \lambda_l$
  - 11:         **end for**
  - 12:         **for**  $0 \leq i \leq d$  **do**
  - 13:             Store  $\sum_{p \in \mathcal{M}_{j+1}} \Lambda_p \varphi_i(X_{t_{j+1}}^p)$ .
  - 14:         **end for**
  - 15:     **end for**
  - 16: **end for**
  - 17: Final approximation of  $\mathbb{E}[f(X_T^x)]$  is

$$\sum_{p \in \mathcal{M}_n} \Lambda_p f(X_{t_n}^p)$$


---

ASSUMPTION 1.3. (A1) Uniform strong Hörmander condition: There exist  $\delta > 0$  and  $m \in \mathbb{N}$  such that, for all  $\xi \in \mathbb{R}^N$ ,

$$\inf_{(x, x') \in \mathbb{R}^N \times \mathbb{R}} \sum_{\alpha \in \bigcup_{k=1}^m \{1, \dots, N\}^k} \langle V_{[\alpha]}(x, x'), \xi \rangle^2 \geq \delta |\xi|^2.$$

(A2) Smoothness of coefficients:

$$\varphi_i \in C_b^\infty(\mathbb{R}^N; \mathbb{R}), \quad V_i \in C_b^\infty(\mathbb{R}^N \times \mathbb{R}; \mathbb{R}^N) \quad i = 0, \dots, d.$$

(A3) We assume the paths in any cubature formula we use are absolutely continuous.

As is common with cubature on Wiener space methods, when the terminal function  $f$  is not smooth, we will use an uneven partition of the time interval  $[0, T]$ .

**Algorithm 2** Lagrange interpolation method

- 
- 1: Set  $(X^\emptyset, \Lambda_0) = (x, 1)$
  - 2: **for**  $0 \leq j \leq n - 1$  **do**
  - 3:     **for**  $p \in \mathcal{M}_j$  **do**
  - 4:         **for**  $0 \leq i \leq d$  **do**
  - 5:             Set  $E_i^L(t) := L[\{(t_{j-k}, \sum_{p \in \mathcal{M}_{j-k}} \Lambda_p \varphi_i(X_{t_{j-k}}^p)\}_{k=0}^{j \wedge (r-1)}](t)$
  - 6:         **end for**
  - 7:         **for**  $1 \leq l \leq N_{\text{cub}}$  **do**
  - 8:             Define  $X_{t_{j+1}}^{p*l}$  as the solution of the ODE:

$$dX_t^{p*l} = \sum_{i=0}^d V_i(X_t^{p*l}, E_i^L(t)) d\omega_i^l(t_j, t_{j+1})(t),$$

$$X_{t_j}^{p*l} = X_{t_j}^p.$$

- 9:             Set the associated weight:  $\Lambda_{p*l} = \Lambda_p \lambda_l$
- 10:         **end for**
- 11:         **for**  $0 \leq i \leq d$  **do**
- 12:             **for**  $0 \leq k \leq ((j+1) \wedge r)$  **do**
- 13:                 Store  $\sum_{p \in \mathcal{M}_{j+1-k}} \Lambda_p \varphi_i(X_{t_{j+1-k}}^p)$ .
- 14:             **end for**
- 15:         **end for**
- 16:     **end for**
- 17: **end for**
- 18: Final approximation of  $\mathbb{E}[f(X_T^x)]$  is

$$\sum_{p \in \mathcal{M}_n} \Lambda_p f(X_{t_n}^p)$$


---

Here, we introduce the Kusuoka partition and a modified version. We denote by  $\Pi_n^\gamma$  the Kusuoka [29] partition of the interval  $[0, T]$  with  $(n+1)$  points and parameter  $\gamma \geq 1$ , defined by

$$t_j = T \left( 1 - \left( 1 - \frac{j}{n} \right)^\gamma \right) \quad \text{for } j = 0, \dots, n-1,$$

$$t_n = T.$$

We denote by  $\Pi_n^{\gamma,r}$  the modified Kusuoka partition, with  $r$  smaller steps at the start whose size is determined by the overall order of the method we require. It is defined as follows: for a fixed integer  $r$  and real parameter  $\gamma$ , we fix the first  $(r+1)$  points as  $t_0 = 0$ ,  $t_{k+1} - t_k = T n^{-r/(k+1)}$  for  $k = 0, \dots, r-1$ . Thereafter,

and we split the rest of the interval  $[t_r, T]$  using the Kusuoka partition, that is,

$$t_j = (T - t_r) \left( 1 - \left( 1 - \frac{j-r}{n-r} \right)^\gamma \right) + t_r, \quad j \in \{r+1, \dots, n-1\},$$

$$t_n = T.$$

Recall that  $\mathcal{E}(T, x, l, \Pi_n)$ , as defined in (5), is the global error of the respective algorithm approximation. Then we have the following result, which is the main result of this work.

**THEOREM 1.4.** *Let  $f \in C_b^\infty(\mathbb{R}^N; \mathbb{R})$ . Then, assuming (A2), the error for the Taylor method satisfies the following:*

$$\sup_{x \in \mathbb{R}^N} |\mathcal{E}(T, x, l, \Pi_n)| \leq C \sum_{j=0}^{n-1} (t_{j+1} - t_j)^{A(q,l)},$$

where  $A(q, l) := (q + 2) \wedge (l + 1)/2$ . Under the same assumptions, the error in the Lagrange interpolation method is

$$\begin{aligned} & \sup_{x \in \mathbb{R}^N} |\mathcal{E}(T, x, l, \Pi_n)| \\ & \leq C \sum_{j=0}^{n-1} \left\{ \frac{(t_{j+1} - t_j)^{j \wedge (r-1)}}{((j+1) \wedge r)!} \prod_{k=0}^{j \wedge (r-1)} (t_{j+1} - t_{j-k}) + (t_{j+1} - t_j)^{(l+1)/2} \right\}. \end{aligned}$$

Now, suppose  $f$  is only Lipschitz continuous. Assuming (A1)–(A3) and that we use the Kusuoka partition  $\Pi_n^\gamma$  with  $\gamma > l - 1$ , we can bound the error in the Taylor method according to the size of  $m$

$$(10) \quad m = 1 : \quad \sup_{x \in \mathbb{R}^N} |\mathcal{E}(T, x, l, \Pi_n^\gamma)| \leq C n^{-B(q,l)-1/2},$$

$$(11) \quad m \geq 2 : \quad \sup_{x \in \mathbb{R}^N} |\mathcal{E}(T, x, l, \Pi_n^\gamma)| \leq C n^{-B(q,l)},$$

where  $B(q, l) = (q + \frac{1}{2}) \wedge \frac{l-2}{2}$ . Assuming (A1)–(A3) and that we use the modified Kusuoka partition  $\Pi_n^{\gamma,r}$  with  $\gamma \in (l - 1, l)$ , we can bound the error in the Lagrange interpolation method according to the size of  $m$ :

$$(12) \quad m = 1 : \quad \sup_{x \in \mathbb{R}^N} |\mathcal{E}(T, x, l, \Pi_n^{\gamma,r})| \leq C n^{-D(r,l)-1/2} (1 - r/n)^{-l/2},$$

$$(13) \quad m \geq 2 : \quad \sup_{x \in \mathbb{R}^N} |\mathcal{E}(T, x, l, \Pi_n^{\gamma,r})| \leq C n^{-D(r,l)} (1 - r/n)^{-l/2},$$

where  $D(r, l) = (r - \frac{3}{2}) \wedge \frac{l-2}{2}$ .

REMARK 1.5. 1. Let us comment on the term  $(1 - r/n)^{-1/2}$  appearing in the error for the Lagrange interpolation method. This term comes from the need to take small steps at the start (for an accurate polynomial approximation) and end of the partition [due to blow up of the derivatives of  $u^y(t, \cdot)$ ]. We need to take  $r$  small steps at the start, leaving  $n - r$  steps split in the style of the Kusuoka partition. This term reflects the need to balance the size of  $r$  and  $n$ . Since it goes to zero as  $n \rightarrow \infty$ , it can be bounded by a constant for sufficiently large  $n$ . For example, in the case  $m = 1$  for a second-order method, one must use a cubature formula of degree  $l = 5$  and choose  $r = 4$ . Then, for  $n \geq 10$ , we have that  $(1 - r/n)^{-1/2} < 3.6$ .

2. The uniformly elliptic case covered by [18] is the case  $m = 1$ . Choosing the parameter  $r$  appropriately, we also recover the same rate for the Lagrange interpolation method up to the multiplication of the term  $(1 - r/n)^{-1/2}$ .

3. In the case where  $m \geq 2$ , we lose  $1/2$  an order of convergence. This is due to the difference in the way we split the error, which we explain in Section 3.

REMARK 1.6. The cubature method will still converge even if the test function is not chosen to be Lipschitz, but only bounded and measurable. However, in this case, a control on the rate of convergence of the cubature method will not be possible, regardless of the choice of partition (Kusuoka, uniform) or the choice of the method (Taylor, Lagrange). A similar result was obtained by the first author for the cubature method applied to classical SDEs (not McKean–Vlasov); see Theorem 3.7 in [10].

**2. Preliminary results.** First, we have a lemma on the existence, uniqueness and moment bounds for the solution of equation (4).

LEMMA 2.1. *Under assumption (A2), there exist unique strong solutions to equations (3) and (4). Moreover, for all  $s \in [t, T]$ , the mapping  $x \mapsto X_s^{t,x,y}$  is  $\mathbb{P}$ -a.s. smooth, and for all multi-indices  $\eta$  on  $\{1, \dots, N\}$ ,*

$$(14) \quad \sup_{x,y \in \mathbb{R}^N} \|\partial_x^\eta X_s^{t,x,y}\|_p < \infty \quad \forall p \geq 1.$$

PROOF. Under assumption (A2), existence and uniqueness of strong solutions to equations (3) and (4) is easy to prove and can be found in, for example, [21]. Now, we note that we can view  $(X_s^{t,x,y})_{s \in [t, T]}$  as the solution of an SDE with coefficients

$$W_i^y(s, z) := V_i(z, \mathbb{E}\varphi_i(X_s^{0,y})),$$

depending on time and a parameter. Due to Assumption 1.3 (A2), the map  $s \mapsto h(s) := \mathbb{E}\varphi_i(X_s^{0;x})$  is continuous in view of the path continuity of the process  $X_s^{0;x}$  and Lebesgue's dominated convergence theorem. Therefore, Itô's formula applied to  $\varphi_i(X_s^{0;x})$  shows that the map  $h(s)$  is continuously differentiable with

a bounded first derivative. This allows one to apply Itô’s formula to  $\mathcal{L}_s^y \varphi_i(X_s^{0;x})$  and get that the map  $s \mapsto \mathbb{E} \mathcal{L}_s^y \varphi_i(X_s^{0;x})$  enjoys the same properties as  $h(s)$ . Iterating the procedure, one obtains that the map  $(s, z) \mapsto W_i^y(s, z)$  is smooth, with bounded derivatives of all orders, with all bounds uniform in  $y$ . The differentiability in  $z$  is assumed, and the differentiability in  $s$  comes from the smoothness of each  $\varphi_i$ , which allows us to apply Itô’s formula to  $\varphi_i(X_s^{0,y})$ , giving

$$\partial_s [\mathbb{E} \varphi_i(X_s^{0,y})] = \mathbb{E}[(\mathcal{L}_s^y \varphi_i)(X_s^{0,y})].$$

Then Kunita [27], Theorem 4.6.5, guarantees that the moment bound (14) holds. □

In the next lemma, we collect some results on the regularity of  $u^y(t, x)$  [defined in (6)] and the pure cubature part of the error. We use the notation, for  $\psi \in \mathcal{C}_b^2(\mathbb{R}^N; \mathbb{R})$ ,

$$\|\psi\|_{2,\infty} := \sup_{x \in \mathbb{R}^N} \{|\psi(x)| + |\nabla \psi(x)| + |\nabla^2 \psi(x)|\}.$$

LEMMA 2.2. *Let  $j \in \{0, \dots, n - 2\}$  and  $t \in [0, T)$ .*

1. *If  $f \in \mathcal{C}_b^\infty(\mathbb{R}^N; \mathbb{R})$ , then for both schemes corresponding to  $E = E^T$  and  $E = E^L$*

$$(15) \quad \sup_{x, y \in \mathbb{R}^N} |[P_{t_j, t_{j+1}}^{E, y} - Q_{t_j, t_{j+1}}^{E, y}] u^y(t_{j+1}, x)| \leq C(t_{j+1} - t_j)^{(l+1)/2}.$$

*Moreover, the first two derivatives of  $u^y$  are bounded:*

$$(16) \quad \sup_{(t, y) \in [0, T] \times \mathbb{R}^N} \|u^y(t, \cdot)\|_{2,\infty} \leq C.$$

2. *If  $f$  is Lipschitz, then*

$$(17) \quad \begin{aligned} & \sup_{x, y \in \mathbb{R}^N} |[P_{t_j, t_{j+1}}^{y} - Q_{t_j, t_{j+1}}^{y}] u^y(t_{j+1}, x)| \\ & \leq C \sum_{k=l}^{l+1} (T - t_{j+1})^{-k/2} (t_{j+1} - t_j)^{(k+1)/2}. \end{aligned}$$

*In addition, the first derivative of  $u^y$  is bounded*

$$(18) \quad \sup_{(t, x, y) \in [0, T] \times \mathbb{R}^N \times \mathbb{R}^N} |\nabla u^y(t, x)| \leq C$$

*and we have the estimate on the first two derivatives:*

$$(19) \quad \sup_{y \in \mathbb{R}^N} \|u^y(t, \cdot)\|_{2,\infty} \leq C(T - t)^{-m/2}.$$

*Finally, for both schemes corresponding to  $E = E^L$  and  $E = E^T$ ,*

$$(20) \quad \sup_{x, y \in \mathbb{R}^N} |[P_{s, t}^{y} - Q_{s, t}^{E, y}] f(x)| \leq C \|f\|_{\text{Lip}} |t - s|^{1/2}.$$

PROOF. We think of  $(X_s^{t,x,y})_{s \in [t,T]}$  and  $({}^E X_s^{t,x,y})_{s \in [t,T]}$  as the solutions of the SDEs with coefficients

$$\begin{aligned} W_i^y(s, z) &:= V_i(z, \mathbb{E}\varphi_i(X_s^{0,y})), \\ {}^E W_i^y(s, z) &:= V_i(z, E_s^y(\varphi_i)), \quad i = 0, \dots, d, \end{aligned}$$

respectively. We think of  $\{W_i^y(t, \cdot), {}^E W_i^y(t, \cdot) : i = 0, \dots, d\}$  as vector fields on  $\mathbb{R}^N$  depending on time  $t \in [0, T]$  and the parameter  $y \in \mathbb{R}^N$ . In the proof of Lemma 2.1, we explained that  $W_0^y, \dots, W_d^y \in \mathcal{C}_b^\infty([0, T] \times \mathbb{R}^N; \mathbb{R}^N)$  with all bounds uniform in  $y$ . The same is true of the functions  ${}^E W_0^y, \dots, {}^E W_d^y$ . To see this, we note that for the both schemes, the map  $s \mapsto E_s^y(\varphi_i)$  is a polynomial, therefore, smooth with bounded derivatives on  $[0, T]$ .

We use the notation  $W_{[\alpha]}^y$  and  ${}^E W_{[\alpha]}^y$  for iterated Lie brackets of the vector fields introduced just before Assumption 1.3. Then we note that for all  $(s, z, y) \in [0, T] \times \mathbb{R}^N \times \mathbb{R}^N$  and  $\alpha \in \bigcup_{k \geq 1} \{1, \dots, N\}^k$ ,

$$\langle W_{[\alpha]}^y(s, z), \xi \rangle^2 \geq \inf_{x' \in \mathbb{R}^N} \langle V_{[\alpha]}(z, x'), \xi \rangle^2,$$

so that

$$\inf_{(s,z,y) \in [0,T] \times \mathbb{R}^N \times \mathbb{R}^N} \langle W_{[\alpha]}^y(s, z), \xi \rangle^2 \geq \inf_{(z,x') \in \mathbb{R}^N \times \mathbb{R}^N} \langle V_{[\alpha]}(z, x'), \xi \rangle^2.$$

Hence, under the uniform strong Hörmander condition (A1),

$$\inf_{(s,z,y) \in [0,T] \times \mathbb{R}^N \times \mathbb{R}^N} \sum_{\alpha \in \bigcup_{k=1}^m \{1, \dots, N\}^k} \langle W_{[\alpha]}^y(s, z, y), \xi \rangle^2 \geq \delta |\xi|^2,$$

so the vector fields  $\{W_i^y : i = 1, \dots, d\}$  satisfy a uniform strong Hörmander condition. Exactly the same holds true for the vector fields  $\{{}^E W_i^y : i = 1, \dots, d\}$ . This uniform strong Hörmander condition is stronger than the UFG condition, hence, we have the results of Section A.3 available to us, subject to slight modification since the coefficients in the current setting also depend on a parameter.

Now, for  $f \in \mathcal{C}_b^\infty(\mathbb{R}^N; \mathbb{R})$ , by differentiating under the expectation and using the moment bounds on  $\partial_x^\eta X_s^{t,x,y}$  contained in (14) we see that for all multi-indices  $\eta$  on  $\{1, \dots, N\}$  with length at least one,

$$\sup_{x,y \in \mathbb{R}^N} |\partial_x^\eta u^y(t, x)| < \infty,$$

so (16) holds [recall that  $u^y(t, x)$  was defined in (6)]. The one step cubature error contained in (15) follows from a stochastic Taylor expansion, noting that for all  $\beta \in \mathcal{A}_1$ ,

$$\sup_{x,y \in \mathbb{R}^N} |\widetilde{W}_\beta^y u^y(t, x)| < \infty.$$

This again follows from the boundedness of derivatives of  $u^y(t, \cdot)$  and  $W_i^y(t, \cdot)$  uniformly in  $y$ .

For  $f$  Lipschitz, the bound in (17) is the same as (77), adapted to the case where coefficients also depend on a parameter. Both estimates (18) and (19) come from Corollary A.10 adapted to the case where coefficients also depend on a parameter.

Finally, when  $f$  is Lipschitz,

$$\begin{aligned} |[P_{s,t}^y - Q_{s,t}^{E,y}]f(x)| &= |\mathbb{E}[f(X_t^{s,x,y})] - \mathbb{E}_{\mathbb{Q}}[f({}^E X_t^{s,x,y})]| \\ &\leq |\mathbb{E}[f(X_t^{s,x,y})] - f(x)| + |\mathbb{E}_{\mathbb{Q}}[f({}^E X_t^{s,x,y})] - f(x)| \\ &\leq \|f\|_{\text{Lip}}(\mathbb{E}|X_t^{s,x,y} - x| + \mathbb{E}_{\mathbb{Q}}|{}^E X_t^{s,x,y} - x|). \end{aligned}$$

That  $\mathbb{E}|X_t^{s,x,y} - x| \leq C|t - s|^{1/2}$  is a standard result for SDEs with bounded coefficients. For the other term,

$$\mathbb{E}_{\mathbb{Q}}|{}^E X_t^{s,x,y} - x| = \sum_{i=1}^{N_{\text{Cub}}} \lambda_i |{}^E X_t^{s,x,y,i} - x|,$$

where  ${}^E X_t^{s,x,y,i}$  is the solution of the ODE along the  $i$ th cubature path. Then we have

$$\mathbb{E}_{\mathbb{Q}}|{}^E X_t^{s,x,y} - x| \leq C|t - s|,$$

due to a standard estimate on the solution of an ODE with bounded coefficients. □

Before we discuss how accurate the polynomial approximations are, we need a lemma concerning the time partitions we use and a type of sum involving its increments which will appear in the error analysis.

LEMMA 2.3. 1. Let  $a > b \geq 0$ , let  $\gamma > \frac{a-1}{a-b}$  and let  $t_j$  be times in points in the Kusuoka partition, then there is a constant  $\tilde{C} = C(\gamma) > 0$  such that

$$(21) \quad \sum_{j=0}^{n-2} (t_{j+1} - t_j)^a (T - t_{j+1})^{-b} \leq Cn^{-(a-1)}.$$

2. For the partition  $\Pi_n^{\gamma,r}$ ,

$$(22) \quad \frac{1}{((j+1) \wedge r)!} \prod_{k=0}^{j \wedge (r-1)} (t_{j+1} - t_{j-k}) \leq Cn^{-r}$$

and for  $a > b \geq 0$  and  $\gamma > \frac{a-1}{a-b}$

$$(23) \quad \sum_{j=0}^{n-2} (t_{j+1} - t_j)^a (T - t_{j+1})^{-b} \leq Cn^{-(a-1)}(1 - r/n)^{-(a+b-1)}.$$



PROOF. 1. This is proved in a slightly different format in Crisan and Ghazali [10]. First, note that

$$\begin{aligned} t_{j+1} - t_j &= T \left[ \left(1 - \frac{j}{n}\right)^\gamma - \left(1 - \frac{j+1}{n}\right)^\gamma \right] \\ &= T\gamma \int_{(1-\frac{j+1}{n})}^{(1-\frac{j}{n})} u^{\gamma-1} du \\ &\leq T\gamma \frac{1}{n} \left(1 - \frac{j}{n}\right)^{\gamma-1}. \end{aligned}$$

Then we use that  $(1 - \frac{j}{n}) \leq 2(1 - \frac{j+1}{n})$  for  $j \in \{0, \dots, n-2\}$  to get

$$(24) \quad t_{j+1} - t_j \leq C \frac{1}{n} \left(1 - \frac{j+1}{n}\right)^{\gamma-1}.$$

By definition,

$$T - t_{j+1} = T \left(1 - \frac{j+1}{n}\right)^\gamma,$$

so that

$$\begin{aligned} &\sum_{j=0}^{n-2} (t_{j+1} - t_j)^a (T - t_{j+1})^{-b} \\ &\leq C \sum_{j=0}^{n-2} n^{-a} \left(1 - \frac{j+1}{n}\right)^{a(\gamma-1)} \left(1 - \frac{j+1}{n}\right)^{-b\gamma}. \end{aligned}$$

Reordering the terms, we get

$$(25) \quad \sum_{j=0}^{n-2} (t_{j+1} - t_j)^a (T - t_{j+1})^{-b} \leq n^{-(a-1)} \sum_{j=1}^{n-1} n^{-1} \left(\frac{j}{n}\right)^{a(\gamma-1)-b\gamma}.$$

We note that

$$\sum_{j=1}^{n-1} n^{-1} \left(\frac{j}{n}\right)^{a(\gamma-1)-b\gamma} \leq \int_0^1 x^{a(\gamma-1)-b\gamma} dx$$

and the condition  $\gamma > \frac{a-1}{a-b}$  guarantees that the exponent  $a(\gamma - 1) - b\gamma > -1$ , so that the integral is finite.

2. First, note  $t_{j+1} - t_{j-k} = (t_{j+1} - t_j) + (t_j - t_{j-1}) + \dots + (t_{j-k+1} - t_{j-k})$ . There are  $(k + 1)$  terms in this sum, and, in the case  $j < r - 1$ , from the definition of the first  $r$  steps of the partition, the biggest of these is  $T n^{-r/(j+1)}$ . Hence,  $t_{j+1} -$

$t_{j-k} \leq (k + 1)Tn^{-r/(j+1)}$ . So, in this case,

$$\begin{aligned} \frac{1}{((j + 1) \wedge r)!} \prod_{k=0}^{j \wedge (r-1)} (t_{j+1} - t_{j-k}) &= \frac{1}{(j + 1)!} \prod_{k=0}^j (t_{j+1} - t_{j-k}) \\ &\leq \frac{1}{(j + 1)!} \prod_{k=0}^j (k + 1)Tn^{-r/(j+1)} \\ &= T^j n^{-r}. \end{aligned}$$

In the case  $j \geq r - 1$ , there is a constant  $K$  such that for each interval  $t_{j+1} - t_j \leq K/n$ , so that  $t_{j+1} - t_{j-k} \leq K(k + 1)/n$  and

$$\begin{aligned} \frac{1}{(r \wedge j)!} \prod_{k=0}^{(r \wedge j)-1} (t_{j+1} - t_{j-k}) &= \frac{1}{r!} \prod_{k=0}^{r-1} (t_{j+1} - t_{j-k}) \\ &\leq \frac{1}{r!} \prod_{k=0}^{r-1} K(k + 1)/n \\ &= K^r n^{-r}. \end{aligned}$$

Now considering the sum  $\sum_{j=0}^{n-2} (t_{j+1} - t_j)^a (T - t_{j+1})^{-b}$ , we split it into two parts: when  $0 \leq j \leq r - 1$ ,  $t_{j+1} - t_j = Tn^{-r/(j+1)} \leq Cn^{-1}$  and  $T - t_{j+1} \geq T(1 - rn^{-1})$ , using that  $r \leq n/2$ . So,

$$\begin{aligned} (26) \quad \sum_{j=0}^{r-1} (t_{j+1} - t_j)^a (T - t_{j+1})^{-b} &\leq Crn^{-a} (1 - rn^{-1})^{-b} \\ &\leq Cn^{-a} (1 - rn^{-1})^{-b}. \end{aligned}$$

For  $r \leq j \leq n - 2$ , the same analysis as in Lemma 2.3 gives

$$t_{j+1} - t_j \leq C \frac{1}{n} \left(1 - \frac{j + 1 - r}{n - r}\right)^{\gamma-1}.$$

By definition,

$$T - t_{j+1} = (T - t_r) \left(1 - \frac{j + 1 - r}{n - r}\right)^\gamma \geq T(1 - r/n) \left(1 - \frac{j + 1 - r}{n - r}\right)^\gamma.$$

The proof then follows as in the first part of this lemma to give for  $\gamma > \frac{a-1}{a-b}$

$$\sum_{j=r}^{n-2} (t_{j+1} - t_j)^a (T - t_{j+1})^{-b} \leq C(1 - r/n)^{-b} (n - r)^{-(a-1)}.$$

We then note that  $\frac{1}{n-r} = \frac{1}{n} \cdot \frac{1}{1-r/n}$ . Combining this with (26) gives the result.  $\square$

LEMMA 2.4 (Polynomial approximations). *For the Taylor approximation, there exists a finite collection of functions  $\mathcal{H} \subset C_b^\infty(\mathbb{R}^N; \mathbb{R})$  such that*

$$(27) \quad \begin{aligned} & \sup_{s \in [t_j, t_{j+1}]} |\mathbb{E}\varphi_i(X_s^{0,y}) - E_s^{T,y}(\varphi_i)| \\ & \leq C \left\{ (t_{j+1} - t_j)^{q+1} + \sum_{\psi \in \mathcal{H}} |(\mathcal{Q}_{t_j}^{E^{T,y}, \Pi_n} - P_{0,t_j}^y)\psi(y)| \right\}. \end{aligned}$$

*For the Lagrange interpolation method,*

$$(28) \quad \begin{aligned} & \sup_{s \in [t_j, t_{j+1}]} |\mathbb{E}\varphi_i(X_s^{0,y}) - E_s^y(\varphi_i)| \\ & \leq C \left\{ \frac{1}{((j+1) \wedge r)!} \prod_{k=0}^{j \wedge (r-1)} (t_{j+1} - t_{j-k}) \right. \\ & \quad \left. + \sum_{k=0}^{j \wedge (r-1)} |(\mathcal{Q}_{t_{j-k}}^{E,y, \Pi_n} - P_{0,t_{j-k}}^y)\varphi_i| \right\}. \end{aligned}$$

PROOF. *Taylor method*

Now we recall, for the Taylor method with  $s \in [t_j, t_{j+1}]$ ,

$$E_s^{T,y}(\varphi_i) := \sum_{k=0}^q \frac{1}{k!} (\mathcal{T}_{t_j}^Q)^k (\mathbb{E}\varphi_i(X^{0,y}))(s - t_j)^k.$$

We will estimate the error  $|\mathbb{E}\varphi_i(X_s^{0,y}) - E_s^{T,y}(\varphi_i)|$  by splitting it into

$$(29) \quad \begin{aligned} & |\mathbb{E}\varphi_i(X_s^{0,y}) - E_s^{T,y}(\varphi_i)| \\ & \leq |\mathbb{E}\varphi_i(X_s^{0,y}) - \widehat{E}_s^y(\varphi_i)| + |\widehat{E}_s^y(\varphi_i) - E_s^{T,y}(\varphi_i)|, \end{aligned}$$

where

$$\widehat{E}_s^y(\varphi_i) := \sum_{k=0}^q \frac{1}{k!} (\mathcal{T}_{t_j})^k (\mathbb{E}\varphi_i(X^{0,y}))(s - t_j)^k, \quad s \in [t_j, t_{j+1})$$

is the truncated Taylor expansion of  $s \mapsto \mathbb{E}\varphi_i(X_s^{0,y})$  of order  $q$  around  $t_j$ . It is straightforward that

$$|\mathbb{E}\varphi_i(X_s^{0,y}) - \widehat{E}_s^y(\varphi_i)| \leq C(s - t_j)^{q+1}$$

and

$$(30) \quad \begin{aligned} & |\widehat{E}_s^y(\varphi_i) - E_s^{T,y}(\varphi_i)| \\ & \leq \sum_{k=0}^q \frac{1}{k!} (s - t_j)^k |[(\mathcal{T}_{t_j})^k - (\mathcal{T}_{t_j}^Q)^k](\mathbb{E}\varphi_i(X^{0,y}))|. \end{aligned}$$

Now, recall that  $\mathcal{T}_{t_j}^k(\mathbb{E}\varphi_i(X^{0,x})) \in \bar{\mathcal{S}}$ , so it can be written as a sum of products of terms of the form

$$\mathbb{E}g(X_{t_j}^{0,x}, \mathbb{E}\varphi(X_{t_j}^{0,x})),$$

for some  $g \in C_b^\infty(\mathbb{R}^N \times \mathbb{R}; \mathbb{R})$ , and  $\varphi \in C_b^\infty(\mathbb{R}^N; \mathbb{R})$ . For this type of term, the error in replacing expectations under  $\mathbb{P}$  with expectations under  $\mathbb{Q}^{\Pi_n}$  can be bounded by

$$\begin{aligned} & |\mathbb{E}g(X_{t_j}^{0,x}, \mathbb{E}\varphi(X_{t_j}^{0,x})) - \mathbb{E}_{\mathbb{Q}^{\Pi_n}}g(X_{t_j}^{0,x}, \mathbb{E}_{\mathbb{Q}^{\Pi_n}}\varphi(X_{t_j}^{0,x}))| \\ & \leq |\mathbb{E}\varphi(X_{t_j}^{0,x}) - \mathbb{E}_{\mathbb{Q}^{\Pi_n}}\varphi(X_{t_j}^{0,x})| \\ & \quad + |\mathbb{E}g(X_{t_j}^{0,x}, \mathbb{E}_{\mathbb{Q}^{\Pi_n}}\varphi(X_{t_j}^{0,x})) - \mathbb{E}_{\mathbb{Q}^{\Pi_n}}g(X_{t_j}^{0,x}, \mathbb{E}_{\mathbb{Q}^{\Pi_n}}\varphi(X_{t_j}^{0,x}))|. \end{aligned}$$

Due to the form of  $\mathcal{T}^k(\mathbb{E}\varphi_i(X^{0,x})) \in \bar{\mathcal{S}}$ , the error

$$[(\mathcal{T}_{t_j})^k - (\mathcal{T}_{t_j}^Q)^k](\mathbb{E}\varphi_i(X^{0,y}))$$

can be decomposed as a sum of products of such errors for different functions  $g$  and  $\varphi$ . We define  $\mathcal{H}_{i,k}$  to be the collection of these functions appearing in the expression for  $\mathcal{T}^k(\mathbb{E}\varphi_i(X^{0,y}))$ . Then the error

$$|[(\mathcal{T}_{t_j})^k - (\mathcal{T}_{t_j}^Q)^k](\mathbb{E}\varphi_i(X^{0,y}))|$$

can be bounded by a constant multiple of

$$\sum_{\psi \in \mathcal{H}_{i,k}} |(\mathbb{E}_{\mathbb{Q}^{\Pi_n}} - \mathbb{E})\psi(X_{t_j}^{0,y})| = \sum_{\psi \in \mathcal{H}_{i,k}} |(Q_{t_j}^{E^T, y, \Pi_n} - P_{0,t_j}^y)\psi(y)|.$$

So,

$$\begin{aligned} & \sup_{s \in [t_j, t_{j+1}]} |\widehat{E}_s^y(\varphi_i) - E_s^{T,y}(\varphi_i)| \\ & \leq C \sum_{k=0}^q \frac{1}{k!} (t_{j+1} - t_j)^k \sum_{i=0}^d \sum_{\psi \in \mathcal{H}_{i,k}} |(Q_{t_j}^{E^T, y, \Pi_n} - P_{0,t_j}^y)\psi(y)|, \end{aligned}$$

and the largest term in the outer sum on the right-hand side occurs when  $k = 0$ , so using this and defining  $\mathcal{H} := \bigcup_{i=1}^d \mathcal{H}_{i,0}$ , the estimate (29) becomes

$$\begin{aligned} & \sup_{s \in [t_j, t_{j+1}]} |\mathbb{E}\varphi_i(X_s^{0,y}) - E_s^{T,y}(\varphi_i)| \\ & \leq C \left\{ (t_{j+1} - t_j)^{q+1} + q \sum_{\psi \in \mathcal{H}} |(Q_{t_j}^{E^T, y, \Pi_n} - P_{0,t_j}^y)\psi(y)| \right\}. \end{aligned}$$

*Lagrange method*

Let  $z \in C^{k+1}([0, T]; \mathbb{R})$ . Recall that we denote by

$$L[(t_1, x_1), \dots, (t_k, x_k)]$$

the Lagrange interpolating polynomial passing through the points

$$(t_1, x_1), \dots, (t_k, x_k).$$

It is a standard result (see, e.g., Chapter 25 in Abramovitz and Stegun [1]) that the error in approximating  $z(t)$  with the polynomial  $L[(t_1, z(t_1)), \dots, (t_k, z(t_k))]$  for any  $t \in [0, T]$  is

$$(31) \quad z(t) - L[(t_1, z(t_1)), \dots, (t_k, z(t_k))](t) = \frac{1}{k!} z^{(k)}(\xi) \prod_{j=1}^k (t - t_j),$$

where  $\xi$  is some point in  $[0, T]$ . Note also that we can write  $L[(t_1, x_1), \dots, (t_k, x_k)]$  terms of the Lagrange basis polynomials  $L_j : [0, T] \rightarrow \mathbb{R}$  as

$$L[(t_1, x_1), \dots, (t_k, x_k)](t) = \sum_{j=1}^k x_j L_j(t),$$

where

$$L_j(t) = \prod_{i=1, i \neq j}^k \frac{t - t_i}{t_j - t_i}.$$

So, the difference between polynomials interpolating different points on the same time grid is given by

$$L[(t_1, x_1), \dots, (t_k, x_k)](t) - L[(t_1, y_1), \dots, (t_k, y_k)](t) = \sum_{j=0}^k (x_j - y_j) L_j(t)$$

and, in particular,

$$(32) \quad \sup_{t \in [0, T]} |L^k[x_1, \dots, x_k](t) - L^k[y_1, \dots, y_k](t)| \leq C(T) \sum_{j=1}^k |x_j - y_j|.$$

Now, recall the definition of the Lagrange interpolation approximation is

$$E_t^{L,y}(\varphi_i) := L[\{(t_{j-k}, \mathbb{E}^{\mathbb{Q}^{\Pi_n}} \varphi_i(X_{t_{j-k}}^{0,y}))\}_{k=0}^{j \wedge (r-1)}](t), \quad t \in [t_j, t_{j+1}],$$

and consider the same object but with all expectations under the Wiener measure,  $\mathbb{P}$ :

$$E_t^{L,y,\mathbb{P}}(\varphi_i) := L[\{(t_{j-k}, \mathbb{E} \varphi_i(X_{t_{j-k}}^{0,y}))\}_{k=0}^{j \wedge (r-1)}](t), \quad t \in [t_j, t_{j+1}].$$

Then we can split the error error  $\mathbb{E} \varphi_i(X_t^{0,y}) - E_t^y(\varphi_i)$  into

$$[\mathbb{E} \varphi_i(X_t^{0,y}) - E_t^{L,y,\mathbb{P}}(\varphi_i)] + [E_t^{L,y,\mathbb{P}}(\varphi_i) - E_t^{L,y}(\varphi_i)].$$

We are able to control using the first term using (31) and the second term using (32). The result follows immediately.  $\square$

**3. Proof of Theorem 1.4.** For the Taylor method, our proof is essentially the same as that in [18], when  $f \in C_b^\infty(\mathbb{R}^N; \mathbb{R})$ . When  $f$  is Lipschitz, however, we split the local errors differently. In [18], the error is split into<sup>3</sup>

$$(33) \quad \begin{aligned} & [P_{t_j, t_{j+1}}^y - Q_{t_j, t_{j+1}}^{E, y}] u^y \\ &= [P_{t_j, t_{j+1}}^y - P_{t_j, t_{j+1}}^{E, y}] u^y + [P_{t_j, t_{j+1}}^{E, y} - Q_{t_j, t_{j+1}}^{E, y}] u^y. \end{aligned}$$

The term  $[P_{t_j, t_{j+1}}^y - P_{t_j, t_{j+1}}^{E, y}] u^y$  is then estimated in terms of the difference of the generators of the processes  $X$  and  ${}^E X$  applied to  $u^y$ , which in turn depends on an estimate on  $\nabla^2 u^y$ . In the uniformly elliptic case with  $f$  Lipschitz, in [18],  $|\nabla^2 u^y(t, x)| \leq C(T-t)^{-1/2}$ . However, in the Hörmander case,  $|\nabla^2 u^y(t, x)| \leq C(T-t)^{-m/2}$ , where  $m$  is the order of the Hörmander condition, which could be very large. Instead, here, we split the error into

$$(34) \quad \begin{aligned} & [P_{t_j, t_{j+1}}^y - Q_{t_j, t_{j+1}}^{E, y}] u^y \\ &= [Q_{t_j, t_{j+1}}^y - Q_{t_j, t_{j+1}}^{E, y}] u^y + [P_{t_j, t_{j+1}}^y - Q_{t_j, t_{j+1}}^y] u^y. \end{aligned}$$

We control the term  $[Q_{t_j, t_{j+1}}^y - Q_{t_j, t_{j+1}}^{E, y}] u^y$  using only the Lipschitz constant of  $u^y$  which is uniformly bounded in time.

**3.1. Smooth bounded terminal condition.** We introduce the generator associated to the process  $X^{0, y}$

$$\mathcal{L}_s^y := V_0(\cdot, \mathbb{E}\varphi_0(X_s^{0, y})) + \frac{1}{2} \sum_{i=1}^d V_i(\cdot, \mathbb{E}\varphi_i(X_s^{0, y}))^2,$$

and note that, as defined in (6),  $u^y(t, x) = P_{t, T}^y f(x)$  solves the PDE

$$(35) \quad \begin{aligned} (\partial_t + \mathcal{L}_t^y) u^y(t, x) &= 0, \\ u^y(T, x) &= f(x). \end{aligned}$$

In the analysis of each scheme, we split the local error into

$$(36) \quad [P_{t_j, t_{j+1}}^y - P_{t_j, t_{j+1}}^{E, y}] u^y(t_{j+1}, x) = [P_{t_j, t_{j+1}}^y - P_{t_j, t_{j+1}}^{E, y}] u^y(t_{j+1}, x)$$

$$(37) \quad + [P_{t_j, t_{j+1}}^{E, y} - Q_{t_j, t_{j+1}}^{E, y}] u^y(t_{j+1}, x).$$

Equation (36) is the error due to approximating the  $\mathbb{E}\varphi_i(X_t^{0, y})$  by  $E_t^y(\varphi_i)$ , and (37) is a one-step cubature error. Now,

$$\begin{aligned} & [P_{t_j, t_{j+1}}^y - P_{t_j, t_{j+1}}^{E, y}] u^y(t_{j+1}, x) = \mathbb{E}[u^y(t_{j+1}, X_{t_{j+1}}^{t_j, x, y}) - u^y(t_{j+1}, {}^E X_{t_{j+1}}^{t_j, x, y})] \\ &= \mathbb{E} \int_{t_j}^{t_{j+1}} (\mathcal{L}_s^y - \mathcal{L}_s^{E, y}) u^y(s, {}^E X_s^{t_j, x, y}) ds. \end{aligned}$$

<sup>3</sup>Recall that  $u^y$  was defined in (6).

Using the Lipschitz property of the coefficients, we get

$$|(\mathcal{L}_s^y - \mathcal{L}_s^{E,y})u^y(s, {}^E X_s^{t_j, x, y})| \leq C \|u^y(s, \cdot)\|_{2, \infty} \sum_{i=0}^d |\mathbb{E} \varphi_i(X_s^{0,y}) - E_s^y(\varphi_i)|.$$

Now we recall from Lemma 2.2 that  $\|u^y(s, \cdot)\|_{2, \infty} \leq C$  when  $f \in \mathcal{C}_b^\infty(\mathbb{R}^N; \mathbb{R})$ , so

$$(38) \quad [P_{t_j, t_{j+1}}^y - P_{t_j, t_{j+1}}^{E,y}]u^y(t_{j+1}, x) \leq C \int_{t_j}^{t_{j+1}} \sum_{i=0}^d |\mathbb{E} \varphi_i(X_s^{0,y}) - E_s^y(\varphi_i)| ds$$

Now, to control the right-hand side above, we use Lemma 2.4 and we split the proof depending on the individual scheme.

3.1.1. *Taylor method.* Using Lemma 2.4, for the Taylor method, (38) becomes

$$(39) \quad [P_{t_j, t_{j+1}}^y - P_{t_j, t_{j+1}}^{E^T, y}]u^y(t_{j+1}, x) \leq C(t_{j+1} - t_j) \left\{ (t_{j+1} - t_j)^{q+1} + \sum_{\psi \in \mathcal{H}} |(Q_{t_j}^{E^T, y, \Pi_n} - P_{0, t_j}^y)\psi(y)| \right\}.$$

Summing up the local errors, the global error is then given by

$$(40) \quad [P_{0, t_n}^y - Q_{0, t_n}^{E^T, y, \Pi_n}]f(x) \leq C \sum_{j=0}^{n-1} \left\{ (t_{j+1} - t_j)^{q+2} + q(t_{j+1} - t_j) \sum_{\psi \in \mathcal{H}} |(Q_{t_j}^{E^T, y, \Pi_n} - P_{0, t_j}^y)\psi(y)| + (t_{j+1} - t_j)^{(l+1)/2} \right\}.$$

The above holds for any  $f \in \mathcal{C}_b^\infty(\mathbb{R}^N; \mathbb{R})$ . In particular, we can take  $f = \psi$  for any  $\psi \in \mathcal{H}$ . Doing this and repeatedly applying the discrete version of the Gronwall inequality,

$$(41) \quad [P_{0, t_n}^y - Q_{0, t_n}^{E,y}]f(x) \leq C \exp\left(q \sum_{j=0}^{n-1} (t_{j+1} - t_j)\right) \times \sum_{j=0}^{n-1} \left\{ (t_{j+1} - t_j)^{q+2} + (t_{j+1} - t_j)^{(l+1)/2} \right\}.$$

3.1.2. *Lagrange interpolation method.* Using Lemma 2.4, for the Lagrange interpolation method, (38) becomes

$$(42) \quad \begin{aligned} & [P_{t_j, t_{j+1}}^y - P_{t_j, t_{j+1}}^{E^L, y}] u^y(t_{j+1}, x) \\ & \leq C(t_{j+1} - t_j) \left\{ \frac{1}{((j+1) \wedge r)!} \prod_{k=0}^{j \wedge (r-1)} (t_{j+1} - t_{j-k}) \right. \\ & \quad \left. + \sum_{k=0}^{j \wedge (r-1)} |(\mathcal{Q}_{t_{j-k}}^{E^T, y, \Pi_n} - P_{0, t_{j-k}}^y) \varphi_i| \right\}. \end{aligned}$$

The global error is then given by

$$(43) \quad \begin{aligned} & [P_{0, t_n}^y - \mathcal{Q}_{0, t_n}^{E^L, y}] f(x) \\ & \leq C \sum_{j=0}^{n-1} \left[ (t_{j+1} - t_j) \left\{ \frac{1}{((j+1) \wedge r)!} \prod_{k=0}^{j \wedge (r-1)} (t_{j+1} - t_{j-k}) \right. \right. \\ & \quad \left. \left. + \sum_{k=0}^{j \wedge (r-1)} |(\mathcal{Q}_{t_{j-k}}^{E^T, y, \Pi_n} - P_{0, t_{j-k}}^y) \varphi_i| \right\} + (t_{j+1} - t_j)^{(l+1)/2} \right]. \end{aligned}$$

Taking  $f = \varphi_i$  for any  $i = 0, \dots, d$  and using discrete Gronwall inequality, we obtain

$$(44) \quad \begin{aligned} & [P_{0, t_n}^y - \mathcal{Q}_{0, t_n}^{E^L, y}] f(x) \\ & \leq C \exp\left(r \sum_{j=0}^{n-1} (t_{j+1} - t_j)\right) \\ & \quad \times \sum_{j=0}^{n-1} \left\{ \frac{(t_{j+1} - t_j)}{((j+1) \wedge r)!} \prod_{k=0}^{j \wedge (r-1)} (t_{j+1} - t_{j-k}) + (t_{j+1} - t_j)^{(l+1)/2} \right\}. \end{aligned}$$

3.2. *Lipschitz terminal condition,  $m = 1$ .* In this case, the estimate we have on the first two derivatives of  $u^y$  is  $\|u^y(t, \cdot)\|_{2, \infty} \leq C(T - t)^{-1/2}$ . Using this estimate we get, similar to (38),

$$(45) \quad \begin{aligned} & [P_{t_j, t_{j+1}}^y - P_{t_j, t_{j+1}}^{E, y}] u^y(t_{j+1}, x) \\ & \leq C(T - t_{j+1})^{-1/2} \int_{t_j}^{t_{j+1}} \sum_{i=0}^d |\mathbb{E} \varphi_i(X_s^{0, y}) - E_s^y(\varphi_i)| ds. \end{aligned}$$



3.2.1. *Taylor method.* The same arguments as the previous section give the global error is

$$\begin{aligned}
& [P_{0,t_n}^y - Q_{t_n}^{E^T, y, \Pi_n}]f(x) \\
& \leq C \sum_{j=0}^{n-2} (T - t_{j+1})^{-1/2} \left\{ (t_{j+1} - t_j)^{q+2} + (t_{j+1} - t_j) \right. \\
(46) \quad & \left. \times \sum_{\psi \in \mathcal{H}} |(Q_{t_j}^{E^T, y, \Pi_n} - P_{0,t_j}^y)\psi(y)| + (t_{j+1} - t_j)^{(l+1)/2} \right\} \\
& + [P_{t_{n-1}, t_n}^y - Q_{t_{n-1}, t_n}^{E^T, y}]f(x).
\end{aligned}$$

Since  $\psi \in C_b^\infty(\mathbb{R}^N; \mathbb{R})$ , in particular it is Lipschitz. The above estimate holds for all Lipschitz  $f$ , so taking  $f = \psi \in \mathcal{H}$  and using the discrete Gronwall inequality, we get

$$\begin{aligned}
& [P_{0,t_n}^y - Q_{t_n}^{E^T, y, \Pi_n}]f(x) \\
& \leq C \exp\left(\sum_{j=0}^{n-2} (T - t_{j+1})^{-1/2} (t_{j+1} - t_j)\right) \\
(47) \quad & \times \sum_{j=0}^{n-2} (T - t_{j+1})^{-1/2} [(t_{j+1} - t_j)^{q+2} + (t_{j+1} - t_j)^{(l+3)/2}] \\
& + [P_{t_{n-1}, t_n}^y - Q_{t_{n-1}, t_n}^{E^T, y}]f(x).
\end{aligned}$$

Now, we recall that in this setting we use the Kusuoka partition  $\Pi_n^\gamma$  with  $\gamma > l - 1$ . Using Lemma 2.3, we can see that  $\sum_{j=0}^{n-1} (T - t_{j+1})^{-1/2} (t_{j+1} - t_j)$  is bounded independently of  $n$ . For the other two sums, we also use Lemma 2.3 and for the final term we use (20) to get

$$[P_{0,t_n}^y - Q_{t_n}^{E^T, y, \Pi_n}]f(x) \leq C(n^{-(l-1)/2} + n^{-(q+1)} + n^{-\gamma/2}).$$

Noting  $\gamma > l - 1$  gives the result (10).

3.2.2. *Lagrange interpolation method.* The same arguments as the previous section give the global error is

$$\begin{aligned}
 & [P_{0,t_n}^y - Q_{0,t_n}^{E^L,y}]f(x) \\
 & \leq C \sum_{j=0}^{n-2} (T - t_{j+1})^{-1/2} \\
 (48) \quad & \times \left[ (t_{j+1} - t_j) \left\{ \frac{1}{((j+1) \wedge r)!} \prod_{k=0}^{j \wedge (r-1)} (t_{j+1} - t_{j-k}) \right. \right. \\
 & \left. \left. + \sum_{k=0}^{(r \wedge j)-1} |(Q_{t_{j-k}}^{E^L,y,\Pi_n} - P_{0,t_{j-k}}^y)\varphi_i| \right\} + (t_{j+1} - t_j)^{(l+1)/2} \right] \\
 & + [P_{t_{n-1},t_n}^y - Q_{t_{n-1},t_n}^{E^L,y}]f(x).
 \end{aligned}$$

Since  $\varphi_i \in \mathcal{C}_b^\infty(\mathbb{R}^N; \mathbb{R})$ , in particular it is Lipschitz. The above estimate holds for all Lipschitz  $f$ , so taking  $f = \varphi_i$ ,  $i = 0, \dots, d$  and using the discrete Gronwall inequality, we get

$$\begin{aligned}
 & [P_{0,t_n}^y - Q_{t_n}^{E^L,y,\Pi_n}]f(x) \\
 & \leq C \exp\left(r \sum_{j=0}^{n-2} (T - t_{j+1})^{-1/2} (t_{j+1} - t_j)\right) \\
 (49) \quad & \times \sum_{j=0}^{n-2} \left[ \frac{(T - t_{j+1})^{-1/2}}{((j+1) \wedge r)!} \prod_{k=0}^{j \wedge (r-1)} (t_{j+1} - t_{j-k}) + (t_{j+1} - t_j)^{(l+1)/2} \right] \\
 & + [P_{t_{n-1},t_n}^y - Q_{t_{n-1},t_n}^{E^L,y}]f(x).
 \end{aligned}$$

By part 1 Lemma 2.3,

$$\sum_{j=0}^{n-2} (T - t_{j+1})^{-1/2} (t_{j+1} - t_j)^{(l+1)/2} \leq C n^{-(l-1)/2} (1 - r/n)^{-l/2}.$$

Now, we recall that in this setting we use the modified Kusuoka partition  $\Pi_n^{\gamma,r}$  with  $\gamma > l - 1$ . We note

$$\begin{aligned}
 & \sum_{j=0}^{n-2} (T - t_{j+1})^{-1/2} \frac{1}{((j+1) \wedge r)!} \prod_{k=0}^{j \wedge (r-1)} (t_{j+1} - t_{j-k}) \\
 & = \sum_{j=0}^{n-2} (T - t_{j+1})^{-1/2} (t_{j+1} - t_j) \times \left[ \frac{1}{((j+1) \wedge r)!} \prod_{k=1}^{j \wedge (r-1)} (t_{j+1} - t_{j-k}) \right]
 \end{aligned}$$

with the left-hand term in the product above being bounded uniformly in  $n$  by part 1, Lemma 2.3 and the second term being less than  $n^{-(r-1)}$  by part 2 of the same lemma. For the final term in (49), we use (20) to get

$$[P_{t_{n-1}, t_n}^y - Q_{t_{n-1}, t_n}^{E^L, y}]f(x) \leq C(n-r)^{-\gamma/2},$$

then noting that

$$(n-r)^{-\gamma/2} = n^{-\gamma/2}(1-r/n)^{-\gamma/2} \leq n^{-(l-1)/2}(1-r/n)^{-l/2}$$

when  $\gamma \in (l-1, l)$ , we finally obtain

$$[P_{0, t_n}^y - Q_{t_n}^{E^L, y, \Pi_n^{\gamma, r}}]f(x) \leq C(n^{-(l-1)/2}(1-r/n)^{-l/2} + n^{-(r-1)}).$$

This proves the result (12).

3.3. *Lipschitz terminal condition,  $m \geq 2$ .* When  $f$  is Lipschitz and  $m \geq 2$ , we split the local error into

$$(50) \quad [P_{t_j, t_{j+1}}^y - Q_{t_j, t_{j+1}}^{E, y, \Pi_n}]u(t_{j+1}, x) = [Q_{t_j, t_{j+1}}^{y, \Pi_n} - Q_{t_j, t_{j+1}}^{E, y, \Pi_n}]u(t_{j+1}, x)$$

$$(51) \quad + [P_{t_j, t_{j+1}}^y - Q_{t_j, t_{j+1}}^{y, \Pi_n}]u(t_{j+1}, x).$$

Equation (50) is the error due to approximating the  $\mathbb{E}\varphi_i(X_t^{0, y})$  by  $E_t^y(\varphi_i)$ , and (51) is a one-step cubature error. For the term in (50), we note that

$$(52) \quad \begin{aligned} & |[Q_{t_j, t_{j+1}}^{y, \Pi_n} - Q_{t_j, t_{j+1}}^{E, y, \Pi_n}]u^y(t_{j+1}, x)| \\ &= \mathbb{E}_{\mathbb{Q}^{\Pi_n}} |u^y(t_{j+1}, X_{t_{j+1}}^{t_j, x, y}) - u^y(t_{j+1}, E X_{t_{j+1}}^{t_j, x, y})| \\ &\leq \|\nabla u^y(t_{j+1}, \cdot)\|_{\infty} \mathbb{E}_{\mathbb{Q}^{\Pi_n}} |X_{t_{j+1}}^{t_j, x, y} - E X_{t_{j+1}}^{t_j, x, y}|. \end{aligned}$$

Now, using the Lipschitz property of the coefficients, we note that

$$\begin{aligned} & \mathbb{E}_{\mathbb{Q}^{\Pi_n}} |X_{t_{j+1}}^{t_j, x, y} - E X_{t_{j+1}}^{t_j, x, y}| \\ &\leq \sum_{k=1}^{N_{\text{Cub}}} \lambda_k \sum_{i=0}^d \int_{t_j}^{t_{j+1}} |V_i(X_s^{t_j, x, y}(\omega_k), \mathbb{E}\varphi_i(X_s^{0, y})) \\ &\quad - V_i(E X_s^{t_j, x, y}(\omega_k), E_s(\varphi_i))| d\omega_k^i(t_j, t_{j+1})(s). \end{aligned}$$

We recall the rescaled path  $\omega_k^i(t_j, t_{j+1})(s) = \sqrt{t_{j+1} - t_j} \omega_k^i(\frac{s-t_j}{t_{j+1}-t_j})$ , so that, under the assumption that  $\omega_k$  is absolutely continuous,

$$\sup_{s \in [t_j, t_{j+1}]} \left| \frac{d}{ds} \omega_k(t_j, t_{j+1})(s) \right| \leq \frac{1}{\sqrt{t_{j+1} - t_j}} \sup_{s \in [0, 1]} \left| \frac{d}{ds} \omega_k(s) \right|.$$

So, there exists a constant  $C$  which depends on

$$\max_{k=1,\dots,N_{\text{Cub}}} \sup_{s \in [0,1]} \left| \frac{d}{ds} \omega_k(s) \right|$$

such that

$$\begin{aligned} & \mathbb{E}_{\mathbb{Q}^{\Pi_n}} |X_{t_{j+1}}^{t_j, x, y} - E X_{t_{j+1}}^{t_j, x, y}| \\ & \leq C \frac{1}{\sqrt{t_{j+1} - t_j}} \\ & \quad \times \sum_{i=0}^d \int_{t_j}^{t_{j+1}} \mathbb{E}_{\mathbb{Q}^{\Pi_n}} |X_s^{t_j, x, y} - E X_s^{t_j, x, y}| + |\mathbb{E} \varphi_i(X_s^{0, y}) - E_s(\varphi_i)| ds. \end{aligned}$$

Then, using Gronwall's inequality, we have

$$\mathbb{E}_{\mathbb{Q}^{\Pi_n}} |X_{t_{j+1}}^{t_j, x, y} - E X_{t_{j+1}}^{t_j, x, y}| \leq C \sqrt{t_{j+1} - t_j} \sup_{s \in [t_j, t_{j+1}]} |\mathbb{E} \varphi_i(X_s^{0, y}) - E_s(\varphi_i)|.$$

Now, going back to (52), we have

$$(53) \quad \begin{aligned} & |[\mathcal{Q}_{t_j, t_{j+1}}^{y, \Pi_n} - \mathcal{Q}_{t_j, t_{j+1}}^{E, y, \Pi_n}] u^y(t_{j+1}, x)| \\ & \leq C \sqrt{t_{j+1} - t_j} \sup_{s \in [t_j, t_{j+1}]} |\mathbb{E} \varphi_i(X_s^{0, y}) - E_s(\varphi_i)|. \end{aligned}$$

From this point on, the arguments depend on the individual scheme.

3.3.1. *Taylor method.* Using Lemma 2.4, (53) becomes

$$(54) \quad \begin{aligned} & |[\mathcal{Q}_{t_j, t_{j+1}}^{y, \Pi_n} - \mathcal{Q}_{t_j, t_{j+1}}^{E^T, y, \Pi_n}] u^y(t_{j+1}, x)| \\ & \leq C(t_{j+1} - t_j)^{1/2} \left\{ (t_{j+1} - t_j)^{q+1} \right. \\ & \quad \left. + \sum_{\psi \in \mathcal{H}} |(\mathcal{Q}_{t_j}^{E^T, y, \Pi_n} - P_{0, t_j}^y) \psi(y)| \right\}. \end{aligned}$$

Since  $\psi \in C_b^\infty(\mathbb{R}^N; \mathbb{R})$ , we can use the global error from the last section for smooth terminal conditions contained in (41) to obtain

$$(55) \quad \begin{aligned} & |[\mathcal{Q}_{t_j, t_{j+1}}^{y, \Pi_n} - \mathcal{Q}_{t_j, t_{j+1}}^{E^T, y, \Pi_n}] u^y(t_{j+1}, x)| \\ & \leq C(t_{j+1} - t_j)^{1/2} \\ & \quad \times \left\{ (t_{j+1} - t_j)^{q+1} + \sum_{i=0}^j [(t_{i+1} - t_i)^{q+2} + (t_{i+1} - t_i)^{(l+1)/2}] \right\}. \end{aligned}$$

Now, we simply use that  $t_{j+1} - t_j \leq C/n$  to get

$$(56) \quad \begin{aligned} & |[\mathcal{Q}_{t_j, t_{j+1}}^{y, \Pi_n} - \mathcal{Q}_{t_j, t_{j+1}}^{E^T, y, \Pi_n}]u^y(t_{j+1}, x)| \\ & \leq Cn^{-1/2}\{n^{-(q+1)} + n\{n^{-(q+2)} + n^{-(l+1)/2}\}\} \\ & \leq C(n^{-(q+3/2)} + n^{-l/2}). \end{aligned}$$

Combining with the local cubature errors and summing up, we get the global error

$$[P_{0, t_n}^y - \mathcal{Q}_{t_n}^{E^T, y, \Pi_n}]f(x) \leq C(n^{-(q+1/2)} + n^{-(l-2)/2}).$$

3.3.2. *Lagrange interpolation method.* Here, we only consider the modified Kusuoka partition  $\Pi_n^{y, r}$ . Using Lemma 2.3, part 2 and Lemma 2.4, inequality (53) becomes

$$(57) \quad \begin{aligned} & |[\mathcal{Q}_{t_j, t_{j+1}}^y - \mathcal{Q}_{t_j, t_{j+1}}^{E^L, y}]u^y(t_{j+1}, x)| \\ & \leq C(t_{j+1} - t_j)^{1/2} \left\{ n^{-r} + \sum_{k=0}^{j \wedge (r-1)} |(\mathcal{Q}_{t_{j-k}}^{E^L, y, \Pi_n} - P_{0, t_{j-k}}^y)\varphi_i| \right\}. \end{aligned}$$

Using the local error for functions  $\varphi_i \in \mathcal{C}_b^\infty(\mathbb{R}^N; \mathbb{R})$  contained in (44), we get

$$(58) \quad \begin{aligned} & |[\mathcal{Q}_{t_j, t_{j+1}}^y - \mathcal{Q}_{t_j, t_{j+1}}^{E^L, y}]u^y(t_{j+1}, x)| \\ & \leq C(t_{j+1} - t_j)^{1/2} \left\{ n^{-r} + r \sum_{i=0}^j \{n^{-r} + (t_{i+1} - t_i)^{(l+1)/2}\} \right\}. \end{aligned}$$

Now, we simply use that  $t_{j+1} - t_j \leq C/n$  to get

$$(59) \quad \begin{aligned} & |[\mathcal{Q}_{t_j, t_{j+1}}^y - \mathcal{Q}_{t_j, t_{j+1}}^{E^L, y}]u^y(t_{j+1}, x)| \\ & \leq Cn^{-1/2}\{n^{-r} + rn\{n^{-r} + n^{-(l+1)/2}(1 - r/n)^{-l/2}\}\} \\ & \leq C(n^{1/2-r} + n^{-l/2}(1 - r/n)^{-l/2}). \end{aligned}$$

Combining with the local cubature errors and summing up, we get the global error

$$[P_{0, t_n}^y - \mathcal{Q}_{t_n}^{E^L, y, \Pi_n^{y, r}}]f(x) \leq C(n^{3/2-r} + n^{-(l-2)/2}(1 - r/n)^{-l/2}).$$

## 4. Numerical examples.

4.1. *Example 1.* In this section, we implement and compare both algorithms. We consider the following example with dimensions  $N = d = 1$ :

$$X_t^{0, x} = x + \int_0^t \mathbb{E}[X_s^{0, x}] ds + B_t,$$

which has the explicit solution

$$X_t^x = xe^t + B_t.$$

In this case,  $\mathbb{E}X_t^{0,x} = xe^t$  so the Taylor approximation of order  $q$  is easy to compute:

$$\mathcal{T}_t^q(\mathbb{E}X_t^{0,x}) = \sum_{k=0}^q \frac{x}{k!} t^k.$$

We choose the Lipschitz terminal function  $f(x) = x^+ := \max\{x, 0\}$  and, by integrating the Gaussian density, we can compute

$$\mathbb{E}(X_t^{0,x})^+ = \sqrt{t}\phi\left(\frac{xe^t}{\sqrt{t}}\right) + xe^t\left(1 - \Phi\left(-\frac{xe^t}{\sqrt{t}}\right)\right),$$

where  $\phi$  and  $\Phi$  are the density and cumulative distribution function, respectively, of a standard Gaussian random variable. We use the cubature formula of degree 5 contained in Lyons and Victoir [36]. We use a fourth-order adaptive Runge–Kutta scheme to solve the ODEs. We choose our parameters in order to achieve the optimal rate of convergence as given by Theorem 1.4. Since the coefficients are uniformly elliptic, expect to be able to achieve order 2 convergence with a cubature formula of degree 5. So, we only need to choose  $q \geq 1$  and  $\gamma \in (4, 5)$  to achieve quadratic convergence in the Taylor method, and  $r \geq 3$  in the Lagrange interpolation method. We choose the parameters  $(x, T, \gamma, q, r) = (0.5, 10, 4.5, 2, 3)$  and the results are presented in Figure 1. We fit a line to the last four points on the log-log error plot and calculate its gradient as an estimate of the rate of convergence.

We see that both methods achieve the expected quadratic convergence rate. In this simple example, the convergence is quite smooth and the Taylor method performs better than the Lagrange interpolation method.

4.2. *Example 2.* We implement an example where the coefficients are not uniformly elliptic and  $N = d = 2$ . We write  $X_t^{0,x} = (X_t^1, X_t^2)$  to lighten notation slightly.

The example we consider is

$$\begin{pmatrix} X_t^1 \\ X_t^2 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \int_0^t \begin{pmatrix} [2 + \sin(\mathbb{E}X_s^2)] \\ X_s^1 \end{pmatrix} \circ dB_s^1 + \int_0^t \begin{pmatrix} X_s^2 \\ X_s^1 \end{pmatrix} \circ dB_s^2,$$

where the coefficients are

$$\begin{aligned} V_0 &\equiv 0, & V_1(x_1, x_2, x') &= \begin{pmatrix} 2 + \sin(x') \\ x_1 \end{pmatrix}, \\ V_2(x_1, x_2, x') &= \begin{pmatrix} x_2 \\ x_1 \end{pmatrix}, & \varphi_1(x_1, x_2) &= x_2, \end{aligned}$$

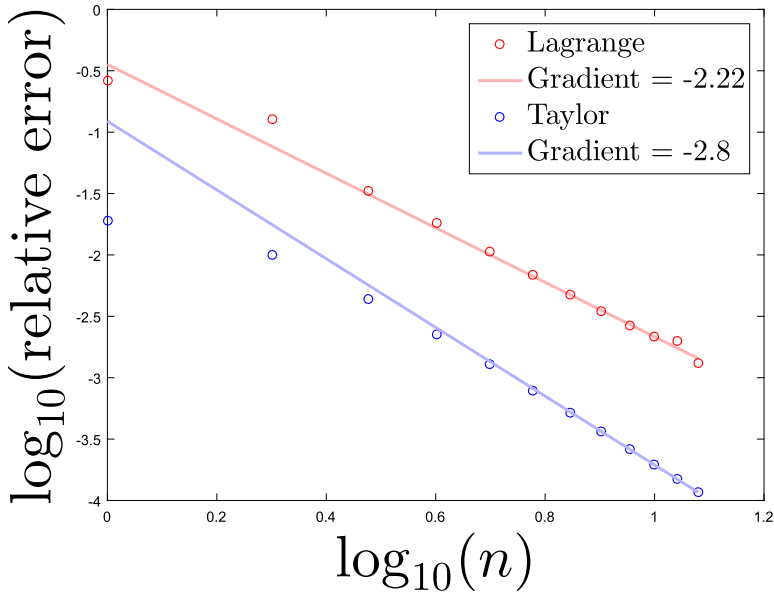


FIG. 1.  $\log$ - $\log$  error plot comparison between the Lagrange interpolation and Taylor methods for Example 1. The gradient of each solid line is given by a linear regression on the last 5 points.

for all  $(x_1, x_2, x') \in \mathbb{R}^3$ . We note that at  $x_1 = 0$  the coefficients degenerate. Second, we note that

$$V_{[(1,2)]}(x_1, x_2, x') = \begin{pmatrix} x_1 \\ 2 + \sin(x') - x_1 \end{pmatrix}.$$

Since  $x_1$  and  $2 + \sin(x') - x_1$  cannot both be zero at the same time, we see that  $V_1$ ,  $V_2$  and  $V_{[(1,2)]}$  span  $\mathbb{R}^2$ . The coefficients therefore satisfy Assumption 1.3 (A1), the uniform strong Hörmander condition, for  $m = 2$ . For  $m = 2$ , with a cubature formula of degree 5, we expect to achieve a convergence rate of  $3/2$  according to Theorem 1.4. To do so, we have to choose  $\gamma \in (4, 5)$  and  $r > 7/2$ . We choose the parameters  $(x_1, x_2, T, \gamma, r) = (1, 0.5, 1, 4.5, 4)$ , with the terminal function  $f(x) = x^+$ . We implement the cubature formula of degree 5 in dimension  $d = 2$  from Lyons and Victoir [36]. In this case, the cubature measure is supported on  $N_{\text{Cub}} = 13$  paths. We could not find an explicit solution, so we compare the cubature approximation to a Monte Carlo approximation with Euler–Maruyama discretisation. The results are presented in Figure 2. In this example, the convergence is not as smooth as Example 1, but the log-log error plot looks approximately linear after 7 steps. After this, the performance of each algorithm is remarkably similar. Empirically we observe second-order convergence, whereas Theorem 1.4 predicts a rate of  $3/2$ .

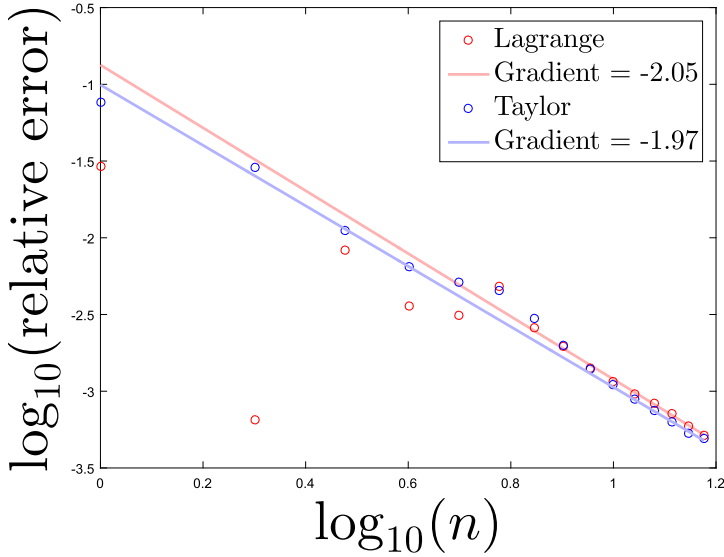


FIG. 2.  $\log$ - $\log$  error plot comparison between the Lagrange interpolation and Taylor methods for Example 2. The gradient of each solid line is given by a linear regression on the last 5 points.

### APPENDIX: DERIVATIVE ESTIMATES FOR TIME-INHOMOGENEOUS PARABOLIC PDES

In this section we obtain estimates on the derivatives of the solution of the linear parabolic partial differential equation (PDE)

$$(60) \quad \begin{aligned} (\partial_t + \mathcal{L}_t)u(t, x) &= 0, & (t, x) \in [0, T) \times \mathbb{R}^N, \\ u(T, x) &= f(x), & x \in \mathbb{R}^N, \end{aligned}$$

where  $f$  is either Lipschitz or continuous and bounded, and  $\mathcal{L}_t$  is the *time-inhomogeneous* differential operator, written in Hörmander form,

$$\mathcal{L}_t = W_0(t) + \frac{1}{2} \sum_{i=1}^d W_i(t)^2.$$

The connection between parabolic PDEs and stochastic differential equations has been well studied. Under various types of conditions on the vector fields  $W_0, \dots, W_d$  and terminal condition  $f$ , the solution to (60) is given by  $u(t, x) = \mathbb{E}[f(X_T^{t,x})]$ , where  $(X_s^{t,x})_{s \in [t, T]}$  solves the following Stratonovich SDE driven by a Brownian motion  $B = (B^1, \dots, B^d)$ , with the convention  $B_t^0 = t$ ,

$$(61) \quad X_s^{t,x} = x + \sum_{i=0}^d \int_t^s W_i(u, X_u^{t,x}) \circ dB_u^i.$$



In the homogeneous case, Kusuoka and Stroock [31] under a uniform Hörmander condition and subsequently Kusuoka [30] under the weaker UFG condition, establish sharp estimates on the derivatives of the solution of (60). Crisan and Delarue [9] extend this analysis to semi-linear equations. To our knowledge, these results have not been obtained without using Malliavin calculus.

When the vector fields defining the time-inhomogeneous stochastic differential equation are smooth in time and space, we can consider the space-time process  $(t, X_t^{0,x})_{t \in [0, T]}$  on  $\mathbb{R}^{N+1}$  and adapt existing results in the literature. There are three main works we draw on: we first show how to adapt the results of Kusuoka [30] to derive gradient bounds in the directions of the vector fields except  $W_0$ ; we then adapt an argument from Crisan and Delarue [9] to prove that the time-inhomogeneous semigroup is a generalised classical solution to a parabolic PDE, and finally using this PDE as a tool, we adapt a result from Crisan, Manolarakis and Nee [15] to derive gradient bounds in the direction of all the vector fields, including  $W_0$ .

We introduce the first assumption which we make throughout this section.

ASSUMPTION A.1.  $W_0, \dots, W_d \in \mathcal{C}_b^\infty([0, T] \times \mathbb{R}^N; \mathbb{R}^N)$ .

Under this assumption, (61) has a unique strong solution. Now, let us define the space-time process  $(\tilde{X}_s^{(t,x)})_{s \in [t, T]} := (s, X_s^{t,x})_{s \in [t, T]}$ , taking values in  $\mathbb{R}^{N+1}$ , which solves the equation

$$(62) \quad \tilde{X}_s^{(t,x)} = \begin{pmatrix} t \\ x \end{pmatrix} + \int_t^s \begin{pmatrix} 1 \\ W_0(\tilde{X}_u^{(t,x)}) \end{pmatrix} du + \sum_{i=1}^d \int_t^s \begin{pmatrix} 0 \\ W_i(\tilde{X}_u^{(t,x)}) \end{pmatrix} \circ dB_u^i.$$

Defining  $\tilde{x} := (t, x) \in \mathbb{R}^{N+1}$  and  $\tilde{W}_0, \dots, \tilde{W}_d : \mathbb{R}^{N+1} \rightarrow \mathbb{R}^{N+1}$  as follows:

$$\forall \tilde{x} \in \mathbb{R}^{N+1} : \quad \tilde{W}_0(\tilde{x}) := \begin{pmatrix} 1 \\ W_0(\tilde{x}) \end{pmatrix}, \quad \tilde{W}_i(\tilde{x}) := \begin{pmatrix} 0 \\ W_i(\tilde{x}) \end{pmatrix}$$

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

we can rewrite (62) more compactly as

$$(63) \quad \tilde{X}_s^{\tilde{x}} = \tilde{x} + \sum_{i=0}^d \int_t^s \tilde{W}_i(\tilde{X}_u^{\tilde{x}}) \circ dB_u^i.$$

As we shall see, by working under the relaxed  $\widetilde{\text{UFG}}$  condition (see Assumption A.2), the solution of the PDE (60) is not necessarily differentiable in each co-ordinate direction in  $\mathbb{R}^N$ . The solution of (60) remains differentiable in certain directions, determined by the vector fields  $\tilde{W}_0, \dots, \tilde{W}_d$ . We now explain what we mean by such a directional derivative.

We can identify vector fields,  $U : \mathbb{R}^m \rightarrow \mathbb{R}^m$  with differential operators acting on sufficiently smooth functions  $\varphi : \mathbb{R}^m \rightarrow \mathbb{R}$  by

$$(64) \quad \forall y \in \mathbb{R}^m : \quad U\varphi(y) := \nabla\varphi(y) \cdot U(y) \equiv \sum_{i=1}^m U^i(y) \partial_{y_i} \varphi(y).$$

We can define a directional derivative of  $\varphi$  in the direction  $U$ , even when  $\partial_{y_j} \varphi$  does not exist classically for all  $j = 1, \dots, m$ . Let  $w_s(y)$  be the solution to the ODE

$$(65) \quad \begin{aligned} \frac{dw_s(y)}{ds} &= U(w_s(y)), & s \geq 0, \\ w_0(y) &= y. \end{aligned}$$

We say that  $\varphi$  is differentiable in the direction  $U$  if the function  $s \mapsto \varphi(w_s(y))$  is differentiable at 0. Then we denote

$$U\varphi(y) = \left. \frac{d}{ds} \varphi(w_s(y)) \right|_{s=0},$$

which coincides with (64) when  $\varphi \in C^1(\mathbb{R}^m; \mathbb{R})$ . In fact, we will see that the semi-group associated to equation (61) is differentiable in directions determined by commutators of the vector fields. The Lie bracket, or commutator, between two vector fields  $U$  and  $W$  is then defined the differential operator

$$[U, W]\varphi := U(W\varphi) - W(U\varphi),$$

which can be identified with the vector field

$$[U, W](y) = \partial W(y)U(y) - \partial U(y)W(y),$$

where  $\partial W(y) := (\partial_{y_j} W^i(y))_{1 \leq i, j \leq m}$  is the Jacobian matrix of  $W$  and similarly for  $\partial U$ .

Since we have assumed the vector fields  $\widetilde{W}_0, \dots, \widetilde{W}_d$  to be smooth, we can repeatedly take commutators of them. Recall the notation  $\mathcal{A}$  for multi-indices on  $\{0, \dots, d\}$  from Section 1.1. We define  $\widetilde{W}_{[\alpha]}$ , for  $\alpha \in \mathcal{A}$  inductively by forming Lie brackets on  $\mathbb{R}^{N+1}$ :

$$\widetilde{W}_{[i]} := \widetilde{W}_i, \quad \widetilde{W}_{[\alpha^*i]} := [\widetilde{W}_{[\alpha]}, W_i] \quad \text{for } i = 0 \dots, d, \alpha \in \mathcal{A}.$$

We note that for all  $\alpha \in \mathcal{A}_1(m)$  [i.e., for  $\alpha \neq (0)$ ] the first component of the vector field  $\widetilde{W}_{[\alpha]}$  is zero. So for  $\alpha \in \mathcal{A}_1(m)$ , a derivative in the direction  $\widetilde{W}_{[\alpha]}$  of a function  $\mathbb{R}^{N+1} \ni (t, x) \mapsto \phi(t, x) \in \mathbb{R}$  only acts in the  $x$  variable. We can therefore write  $\{\widetilde{W}_{[\alpha]}(t) : \alpha \in \mathcal{A}_1(m)\}$  and think of these as differential operators parametrised by  $t$  and acting in the  $x$  variable. Only the vector field  $\widetilde{W}_0$  acts in the  $t$ -direction.

With these concepts in mind, we can now introduce the second assumption we make on the vector fields.

ASSUMPTION A.2.  $\widetilde{\text{UFG}}(m)$  condition: there exists a positive integer  $m$  such that, for all  $\alpha \in \mathcal{A}$  with  $\|\alpha\| > m$ , there exist  $\varphi_{\alpha,\beta} \in C_b^\infty([0, T] \times \mathbb{R}^N; \mathbb{R})$  with

$$\widetilde{W}_{[\alpha]}(t, x) = \sum_{\beta \in \mathcal{A}_1(m)} \varphi_{\alpha,\beta}(t, x) \widetilde{W}_{[\beta]}(t, x).$$

REMARK A.3. Assumption A.2 is not sufficient to obtain the regularity of the law of the associated nonhomogeneous (time dependent) diffusion process  $X$ . Indeed, as observed in [7], the process  $X$  may not have a (smooth) density with respect to the Lebesgue measure even when the vector fields  $W_0, \dots, W_d$  satisfy the Hörmander condition. However, as we show below, Assumption A.2 suffices to allow us to derive gradient bounds of the corresponding (time-inhomogeneous) semigroup. These results hinge on the integration by parts formulae included in Theorem A.7 below which give representations of directional derivatives

$$\widetilde{W}_{[\alpha_1]}(t) \cdots \widetilde{W}_{[\alpha_n]}(t) (P_{t,s} f)(x)$$

of the semigroup  $P_{t,s} f$ . In order to duplicate the results in [7], we would need to deduce integration by parts formulae (and subsequent bounds) for quantities of the form

$$P_{t,s} (\widetilde{W}_{[\alpha_1]}(t) \cdots \widetilde{W}_{[\alpha_n]}(t) f)(x).$$

For classical SDEs, such results were obtained in the paper [15]. As the differentiability of the transition kernel of the solution time-inhomogeneous stochastic differential equation (61) is not required for the analysis of the cubature method (or for it to work), we did not need to develop the corresponding integration by parts formulae here.

**A.1. Kusuoka–Stroock processes.** This class of process was introduced by Kusuoka and Stroock [32]. They will appear as Malliavin weights in our integration by parts formulas. The definition and properties we give here record the regularity and growth of these processes with respect to different parameters. The results allow one to develop integration by parts formulas in a systematic and transparent way, which automatically leads to nice derivative estimates.

DEFINITION A.4 (Kusuoka–Stroock processes). Let  $E$  be a separable Hilbert space and let  $r \in \mathbb{R}$ ,  $M \in \mathbb{N}$ . We denote by  $\mathcal{K}_r(t, E, M)$  the set of functions:  $\Phi_t : (t, T] \times \mathbb{R}^N \rightarrow \mathbb{D}^{M,\infty}(E)$  satisfying the following:

1. For all  $s \in (t, T]$ , the map  $\mathbb{R}^N \ni x \mapsto \Phi_t(s, x) \in L^p(\Omega)$  is  $M$ -times continuously differentiable for all  $p \geq 1$ .
2. For any  $p \geq 1$ , any multi-index  $\alpha$  on  $\{1, \dots, N\}$  and  $m \in \mathbb{N}$  with  $|\alpha| + m \leq M$ , we have

$$(66) \quad \sup_{x \in \mathbb{R}^N} \sup_{s \in (t, T]} (s - t)^{-r/2} \|\partial_x^\alpha \Phi_t(s, x)\|_{\mathbb{D}^{m,p}(E)} < \infty.$$

REMARK A.5. 1. The number  $M$  denotes how many times the Kusuoka–Stroock function can be differentiated and  $r$  measures the growth in  $(s - t)$ .

2. This definition is slightly different to that in [32]: here our processes are defined on  $(t, T]$  instead of  $(0, T]$  and we require continuity in  $L^p(\Omega)$  rather than almost surely.

We record here some properties which help when building Malliavin weights later.

LEMMA A.6. (a) If  $\Phi_t \in \mathcal{K}_r^q(t, \mathbb{R}, M)$  is  $\mathbb{F}$ -adapted, and we define

$$\Psi_t^i(s, x) := \int_t^s \Phi_t(u, x) dB_u^i$$

for  $i \in \{1, \dots, d\}$  and  $\Gamma(s, x) := \int_t^s \Phi_t(u, x) du$ , then  $\Psi_t^i \in \mathcal{K}_{r+1}^q(t, \mathbb{R}, M)$  and  $\Gamma_t \in \mathcal{K}_{r+2}^q(t, \mathbb{R}, M)$ .

(b) If  $\Phi_{t,i} \in \mathcal{K}_{r_i}^{q_i}(t, \mathbb{R}, M_i)$ ,  $i = 1, \dots, n$ , then

$$\sum_{i=1}^n \Phi_{t,i} \in \mathcal{K}_{\min_i r_i}^{\max_i q_i}(t, \mathbb{R}, \min_i M_i)$$

and  $\prod_{i=1}^n \Phi_{t,i} \in \mathcal{K}_{r_1+\dots+r_n}^{q_1+\dots+q_n}(t, \mathbb{R}, \min_i M_i)$ .

PROOF. The proof is essentially the same as the proof of Lemma 75 in [15]. □

**A.2. Integration by parts and derivative bounds.** We are now in a position to prove the main integration by parts result.

THEOREM A.7. Assume that  $\widetilde{\text{UFG}}(m)$  holds and fix  $s \in (t, T]$ . Then, for any  $\alpha_1, \dots, \alpha_n \in \mathcal{A}_1(m)$ , there exist  $\Phi_{t,\alpha_1,\dots,\alpha_n}^1 \in \mathcal{K}_0(t, \mathbb{R})$  and  $\Phi_{t,\alpha_1,\dots,\alpha_n}^2 \in \mathcal{K}_0(t, \mathbb{R}^N)$  such that for  $f \in \mathcal{C}_b^\infty(\mathbb{R}^N; \mathbb{R})$ ,

$$(67) \quad \begin{aligned} & \widetilde{W}_{[\alpha_1]}(t) \cdots \widetilde{W}_{[\alpha_n]}(t)(P_{t,s}f)(x) \\ &= (s - t)^{-\frac{\|\alpha_1\| + \dots + \|\alpha_n\|}{2}} \mathbb{E}[f(X_s^{t,x}) \Phi_{t,\alpha_1,\dots,\alpha_n}^1(s, x)] \end{aligned}$$

and

$$(68) \quad \begin{aligned} & \widetilde{W}_{[\alpha_1]}(t) \cdots \widetilde{W}_{[\alpha_n]}(t)(P_{t,s}f)(x) \\ &= (s - t)^{-\frac{\|\alpha_1\| + \dots + \|\alpha_{n-1}\|}{2}} \mathbb{E}[\nabla f(X_s^{t,x}) \Phi_{t,\alpha_1,\dots,\alpha_n}^2(s, x)]. \end{aligned}$$

Moreover, for  $f$  continuous and bounded or Lipschitz,  $P_{t,s}f(x)$  is differentiable in the directions  $\{\widetilde{W}_{[\alpha]}(t) : \alpha \in \mathcal{A}_1(m)\}$  with

$$(69) \quad \sup_{x \in \mathbb{R}^N} |\widetilde{W}_{[\alpha_1]}(t) \cdots \widetilde{W}_{[\alpha_n]}(t)(P_{t,s}f)(x)| \leq C \|f\|_\infty (s - t)^{-\frac{(\|\alpha_1\| + \dots + \|\alpha_n\|)}{2}}$$

and

$$(70) \quad \sup_{x \in \mathbb{R}^N} |\widetilde{W}_{[\alpha_1]}(t) \cdots \widetilde{W}_{[\alpha_n]}(t)(P_{t,s}f)(x)| \leq C \|f\|_{\text{Lip}}(s-t)^{\frac{1-(|\alpha_1|+\dots+|\alpha_n|)}{2}},$$

in each case, respectively.

REMARK A.8. We emphasise here that each  $\widetilde{W}_{[\alpha_1]}(t)$  is a differential operator acting in  $x$ , parametrised by  $t$ , applied to the function  $x \mapsto P_{t,s}f(x)$ . These results are not valid, for example,  $\widetilde{W}_{[\alpha_1]}(u)P_{t,s}f(x)$  when  $u \neq t$ . We develop estimates on the derivative of  $(t, x) \mapsto P_{t,s}f(x)$  as a function on  $[0, T] \times \mathbb{R}^N$  in Proposition A.15.

PROOF OF THEOREM A.7. Since the  $\widetilde{\text{UFG}}$  condition is precisely the UFG condition introduced in Kusuoka [30] on  $\mathbb{R}^{N+1}$ , we can use the results there. We use the notation  $\widetilde{P}_s g(\tilde{x}) := \mathbb{E}[g(\widetilde{X}_s^{\tilde{x}})]$  for a suitably integrable function  $g : \mathbb{R}^{N+1} \rightarrow \mathbb{R}$ . By Kusuoka [30], Lemma 8 (see also [15], Corollary 32), we know for any  $\alpha_1, \dots, \alpha_n \in \mathcal{A}_1(m)$ , there exist  $\Phi_{\alpha_1, \dots, \alpha_n}^1 \in \mathcal{K}_0(\mathbb{R})$  and  $\Phi_{\alpha_1, \dots, \alpha_n}^2 \in \mathcal{K}_0(\mathbb{R}^{N+1})$  such that for  $g \in C_b^\infty(\mathbb{R}^{N+1}; \mathbb{R})$ ,

$$\widetilde{W}_{[\alpha_1]}(t) \cdots \widetilde{W}_{[\alpha_n]}(t)(\widetilde{P}_s g)(\tilde{x}) = (s-t)^{-\frac{\|\alpha_1\|+\dots+\|\alpha_n\|}{2}} \mathbb{E}[g(\widetilde{X}_s^{\tilde{x}})\Phi_{\alpha_1, \dots, \alpha_n}^1(s, \tilde{x})]$$

and

$$\begin{aligned} & \widetilde{W}_{[\alpha_1]}(t) \cdots \widetilde{W}_{[\alpha_n]}(t)(\widetilde{P}_s g)(\tilde{x}) \\ &= (s-t)^{-\frac{\|\alpha_1\|+\dots+\|\alpha_{n-1}\|}{2}} \mathbb{E}[\nabla g(\widetilde{X}_s^{\tilde{x}})\Phi_{\alpha_1, \dots, \alpha_n}^2(s, \tilde{x})]. \end{aligned}$$

Now for any function  $f \in C_b^\infty(\mathbb{R}^N; \mathbb{R})$ , we can extend it to  $g \in C_b^\infty(\mathbb{R}^{N+1}; \mathbb{R})$  by  $g(t, x) := f(x)$ . We then immediately have the integration by parts formulas (67) and (68). We get the bound stated in (69) with the constant

$$C = \sup_{s \in [t, T]} \sup_{x \in \mathbb{R}^N} \mathbb{E}|\Phi_{t, \alpha_1, \dots, \alpha_n}^1(s, x)|,$$

and a standard approximation argument gives the same estimate for bounded and continuous  $f$ . Similarly, we obtain the bound in (70) with constant

$$C = \sup_{s \in [t, T]} \sup_{x \in \mathbb{R}^N} \mathbb{E}|\Phi_{t, \alpha_1, \dots, \alpha_n}^2(s, x)|,$$

and a standard approximation argument allows one to obtain the same bound for  $f$  Lipschitz with  $\|\nabla f\|_\infty$  replaced by  $\|f\|_{\text{Lip}}$ .  $\square$

**A.3. Uniform Hörmander setting.** In this section, let us consider a stronger assumption than the  $\widetilde{\text{UFG}}(m)$  condition. Suppose that a uniform strong Hörmander condition of order  $m$  holds, that is, we have the following.

ASSUMPTION A.9.  $\text{USH}(m)$ : There exists  $\delta > 0$  and  $m \in \mathbb{N}$  such that for all  $\xi \in \mathbb{R}^N$ ,

$$\inf_{(t,z) \in [0,T] \times \mathbb{R}^N} \sum_{\alpha \in \mathcal{A}_{\geq 1}(m)} \langle W_{[\alpha]}(t, z), \xi \rangle^2 \geq \delta |\xi|^2.$$

In this case, we recover differentiability of  $x \mapsto P_{t,T} f(x)$  in all directions.

COROLLARY A.10. Assume  $\text{USH}(m)$  holds. Let  $\eta$  be a multi-index on  $\{1, \dots, N\}$  and let  $f$  be Lipschitz. Then

$$\sup_{x \in \mathbb{R}^N} |\partial^\eta (P_{t,T} f)(x)| \leq C \|f\|_{\text{Lip}} (T-t)^{-(|\eta|-1)m/2}.$$

PROOF. First, let  $f \in C_b^\infty(\mathbb{R}^N; \mathbb{R})$ . For the first-order derivatives,

$$\partial_x^i \mathbb{E}[f(X_T^{t,x})] = \sum_{k=1}^N \mathbb{E}[\partial^k f(X_T^{t,x}) \partial_x^i (X_T^{t,x})^k].$$

For the higher order derivatives, we note that there exist  $F_\alpha^i \in C_b^\infty([0, T] \times \mathbb{R}^N; \mathbb{R})$  such that

$$e_j = \sum_{\alpha \in \mathcal{A}_{\geq 1}(m)} F_\alpha^j(t, x) W_{[\alpha]}(t, x),$$

where  $e_j$  is the  $j$ th standard basis vector in  $\mathbb{R}^N$ . To see this, define  $\mathcal{W}(t, x)$  to be the  $N \times \text{card}(\mathcal{A}_{\geq 1}(m))$  matrix whose columns are the vector fields  $(W_{[\alpha]})_{\alpha \in \mathcal{A}_{\geq 1}(m)}$  evaluated at  $(t, x)$ .  $\text{USH}(m)$  guarantees that  $\mathcal{W}\mathcal{W}^\top(t, x)$  is invertible. Then

$$F_\alpha^j(t, x) := (\mathcal{W}^\top [\mathcal{W}\mathcal{W}^\top]^{-1}(t, x) e_j)_\alpha$$

satisfies the above relation. Then, for the second-order derivatives,

$$\begin{aligned} \partial_x^{(j,i)} \mathbb{E}[f(X_T^{t,x})] &= \sum_{k=1}^N \partial_x^j \mathbb{E}[\partial^k f(X_T^{t,x}) \partial_x^i (X_T^{t,x})^k] \\ &= \sum_{k=1}^N \sum_{\alpha \in \mathcal{A}_{\geq 1}(m)} F_\alpha^j(t, x) W_{[\alpha]}(t) (\mathbb{E}[\partial^k f(X_T^{t,x}) \partial_x^i (X_T^{t,x})^k]). \end{aligned}$$

We note that  $\partial_x^i (X_T^{t,x})^k \in \mathcal{K}_0(t, \mathbb{R})$ , so we can apply the IBPF in Kusuoka [30] Lemma 8 to obtain the existence of  $H_{t,\alpha} \in \mathcal{K}_0(t, \mathbb{R})$  such that

$$\begin{aligned} & \partial_x^{(j,i)} \mathbb{E}[f(X_T^{t,x})] \\ &= \sum_{k=1}^N \sum_{\alpha \in \mathcal{A}_{\geq 1}(m)} F_\alpha^j(t, x) (T-t)^{-|\alpha|/2} \mathbb{E}[\partial^k f(X_T^{t,x}) H_{t,\alpha}(T, x)] \\ &= (T-t)^{-m/2} \sum_{k=1}^N \mathbb{E}[\partial^k f(X_T^{t,x}) \bar{H}_{t,i,j,k}(T, x)], \end{aligned}$$

where

$$\bar{H}_{t,i,j,k}(T, x) := \sum_{\alpha \in \mathcal{A}_{\geq 1}(m)} (T-t)^{(m-|\alpha|)/2} F_\alpha^j(t, x) H_{t,\alpha}(T, x) \in \mathcal{K}_0(t, \mathbb{R}),$$

and we have used that for all  $\alpha \in \mathcal{A}_{\geq 1}(m)$ ,  $\|\alpha\| = |\alpha| \leq m$ . We get the bound

$$\sup_{x \in \mathbb{R}^N} |\partial_x^{(j,i)} \mathbb{E}[f(X_T^{t,x})]| \leq (T-t)^{-m/2} \|\nabla f\|_\infty \sum_{k=1}^N \sup_{x \in \mathbb{R}^N} \mathbb{E}|\bar{H}_{t,i,j,k}(T, x)|.$$

We can iterate this argument to arbitrary order. Then we can get the bound for a Lipschitz  $f$  using the same approximation as before.  $\square$

**A.4. Connection with PDE.** In this section we make use of the integration by parts formulae of Theorem A.7 to extend the notion of classical solution to the PDE (60) to the case when the solution is not classically differentiable in all directions. The notation and arguments in this section closely follows Crisan and Delarue [9], who provide a similar notion of solution to semilinear PDEs with coefficients which do not depend on time. The idea is very simple: it is a standard result that for a terminal condition  $f \in \mathcal{C}_p^\infty(\mathbb{R}^N; \mathbb{R})$  the PDE (60) has a classical solution. For  $f \in \mathcal{C}_p(\mathbb{R}^N; \mathbb{R})$ , we consider a sequence of smooth approximations  $(f_l)_{l \geq 1}$  to which we can associate solutions  $(v_l)_{l \geq 1}$  to (60). For each  $v_l$ , we can use the integration by parts formula of Theorem A.7 to write the derivatives  $W_t^2(t)v_l(t, x)$  in a form which does not depend on any derivatives of  $f_l$ . We then show that the PDE still holds in the limit  $l \rightarrow \infty$ .

We introduce some function spaces we will need to define what we mean by a classical solution. We denote by  $\mathbb{B}(0, R)$  the open ball in  $\mathbb{R}^N$  of radius  $R > 0$  centred at zero. Let  $\phi \in \mathcal{C}_b^\infty([0, T-1/r] \times \mathbb{B}(0, R); \mathbb{R})$  and define

$$\begin{aligned} & \|\phi\|_{[0, T-1/r] \times \mathbb{B}(0, R); \infty}^{\widetilde{W}_0, 1} \\ & := \|\phi\|_{[0, T-1/r] \times \mathbb{B}(0, R); \infty} + \|\widetilde{W}_0(t)\phi\|_{[0, T-1/r] \times \mathbb{B}(0, R); \infty}. \end{aligned}$$

Define  $\mathcal{D}_{\widetilde{W}_0}^{1,\infty}([0, T - 1/R] \times \mathbb{B}(0, R))$  as the closure of  $\mathcal{C}_b^\infty([0, T - 1/R] \times \mathbb{B}(0, R); \mathbb{R})$  in  $\mathcal{C}_b([0, T - 1/R] \times \overline{\mathbb{B}}(0, R); \mathbb{R})$  with respect to the norm  $\|\cdot\|_{[0, T - 1/R] \times \mathbb{B}(0, R); \infty}^{\widetilde{W}_0, 1}$ . And define

$$\mathcal{D}_{\widetilde{W}_0}^{1,\infty}([0, T] \times \mathbb{R}^N) := \bigcap_{R \geq 1} \mathcal{D}_{\widetilde{W}_0}^{1,\infty}([0, T - 1/R] \times \mathbb{B}(0, R)).$$

Now, take  $\psi \in \mathcal{C}_b^\infty(\mathbb{R}^N; \mathbb{R})$  and, for any ball  $\mathbb{B}$ , define

$$\|\psi\|_{\mathbb{B}, \infty}^{W(t), 1} := \|\psi\|_{\mathbb{B}, \infty} + \sum_{i=1}^d \|W_i(t)\psi\|_{\mathbb{B}, \infty}$$

and

$$\|\psi\|_{\mathbb{B}, \infty}^{W(t), 2} := \|\psi\|_{\mathbb{B}, \infty}^{W(t), 1} + \sum_{i=1}^d \|W_i^2(t)\psi\|_{\mathbb{B}, \infty}.$$

We define  $\mathcal{D}_{W(t)}^{2,\infty}(\mathbb{B})$  to be the closure of  $\mathcal{C}_b^\infty(\mathbb{B}; \mathbb{R})$  in  $\mathcal{C}_b(\overline{\mathbb{B}}; \mathbb{R})$  with respect to  $\|\cdot\|_{\mathbb{B}, \infty}^{W(t), 2}$  and

$$\mathcal{D}_{W(t)}^{2,\infty}(\mathbb{R}^N) := \bigcap_{R \geq 1} \mathcal{D}_{W(t)}^{2,\infty}(\mathbb{B}(0, R)).$$

**DEFINITION A.11 (Classical solution).** We define a function  $v : [0, T] \times \mathbb{R}^N \rightarrow \mathbb{R}$  to be a classical solution to (60) if the following three conditions are satisfied:

1.  $v \in \mathcal{D}_{\widetilde{W}_0}^{1,\infty}([0, T] \times \mathbb{R}^N)$  and for each  $t \in [0, T]$ ,  $v(t, \cdot) \in \mathcal{D}_{W(t)}^{2,\infty}(\mathbb{R}^N)$ , such that for  $i = 1, \dots, d$ ,

$$[0, T] \times \mathbb{R}^N \ni (t, x) \mapsto (W_i(t)v(t, x), W_i^2(t)v(t, x))$$

is a continuous function.

2. For all  $(t, x) \in [0, T] \times \mathbb{R}^N$ ,

$$\widetilde{W}_0 v(t, x) + \frac{1}{2} \sum_{i=1}^d W_i^2 v(t, x) = 0.$$

3.  $\lim_{(t,y) \rightarrow (T,x)} v(t, y) = f(x)$  for all  $x \in \mathbb{R}^N$ .

**REMARK A.12.** 1. Note that since, in general, the space  $\mathcal{D}_{W(t)}^{2,\infty}(\mathbb{R}^N)$  is different for each  $t \in [0, T]$ , our definition requires that  $v(t, \cdot)$  belongs to a different space at each time  $t \in [0, T]$ .



2. If  $\phi \in \mathcal{C}^{1,2}([0, T] \times \mathbb{R}^N; \mathbb{R})$ , then  $\phi \in \mathcal{D}_{\widetilde{W}_0}^{1,\infty}([0, T] \times \mathbb{R}^N)$  and  $\phi(t, \cdot) \in \mathcal{D}_{W(t)}^{2,\infty}(\mathbb{R}^N)$  for all  $t \in [0, T]$ . Moreover,

$$\begin{aligned} \|\phi\|_{[0, T-1/R] \times \mathbb{B}(0, r); \infty}^{\widetilde{W}_0, 1} &\leq C_R \{ \|\phi\|_{[0, T-1/R] \times \mathbb{B}(0, r); \infty} + \|\partial_t \phi\|_{[0, T-1/R] \times \mathbb{B}(0, R); \infty} \\ &\quad + \|\nabla \phi\|_{[0, T-1/R] \times \mathbb{B}(0, R); \infty} \}, \\ \|\phi\|_{\mathbb{B}(0, R); \infty}^{W(t), 2} &\leq C_R \{ \|\phi(t, \cdot)\|_{\mathbb{B}(0, R); \infty} + \|\nabla \phi(t, \cdot)\|_{\mathbb{B}(0, R); \infty} \\ &\quad + \|\nabla^2 \phi(t, \cdot)\|_{\mathbb{B}(0, R); \infty} \}, \end{aligned}$$

where

$$\begin{aligned} C_R &= 1 + \|W_0\|_{[0, T-1/R] \times \mathbb{B}(0, R); \infty} \\ &\quad + \sum_{i=1}^d \|W_i(t, \cdot)\|_{\mathbb{B}(0, R); \infty} + \|\partial W_i(t, \cdot)\|_{\mathbb{B}(0, R); \infty}. \end{aligned}$$

It is then clear that our definition truly is an *extension* of the usual definition of classical solution.

With this definition in hand, we have the following theorem.

**THEOREM A.13.** *Assume that  $\widetilde{\text{UFG}}(m)$  holds and let  $f : \mathbb{R}^N \rightarrow \mathbb{R}$  be continuous with polynomial growth. Then  $v(t, x) := P_{t, T} f(x)$  is a classical solution to (60). It is also the unique solution amongst those which satisfy the following polynomial growth condition: there exists  $q > 0$  such that*

$$|v(t, x)| \leq C(1 + |x|)^q \quad \forall t \in [0, T], x \in \mathbb{R}^N.$$

The proof of uniqueness relies on an Itô formula valid for functions differentiable in the directions of the vector fields. We will also need a stochastic Taylor expansion based on this formula in Section A.6 for the analysis of the error in the cubature on Wiener space algorithm.

**LEMMA A.14.** *Let  $v : [0, T] \times \mathbb{R}^N \rightarrow \mathbb{R}$  satisfy part (1) of Definition A.11 and be of at most polynomial growth. Then, for all  $u \in [t, T]$ ,*

$$\begin{aligned} (71) \quad v(u, X_u^{t,x}) &= v(t, x) + \int_t^u \left[ \widetilde{W}_0(s) v(s, X_s^{t,x}) + \frac{1}{2} \sum_{i=1}^d W_i^2(s) v(s, X_s^{t,x}) \right] ds \\ &\quad + \sum_{i=1}^d \int_t^u W_i(s) v(s, X_s^{t,x}) dB_s^i. \end{aligned}$$

PROOF. This can be proved by a mollification argument as in Proposition 7.1 in [9].  $\square$

We can now prove Theorem A.13.

PROOF OF THEOREM A.13. *Existence:* Denote by  $(f_l)_{l \geq 1}$  a sequence of mollifications of  $f$ . Since  $f$  is continuous,  $f_l$  converges to  $f$  uniformly on compact subsets of  $\mathbb{R}^N$ . Since  $v_l(t, x) - v(t, x) = \mathbb{E}[f_l(X_T^{t,x}) - f(X_T^{t,x})]$ , it is clear that  $v_l$  converges to  $v$  uniformly on compact subsets of  $[0, T] \times \mathbb{R}^N$ . Therefore,  $v$  is continuous up to the boundary at  $t = T$ . Now, consider the integration by parts formula for  $W_i v_l$  and  $W_i^2 v_l$  provided by (67) as part of Theorem A.7. We get

$$\begin{aligned} W_i v_l(t, x) &= (T - t)^{-1/2} \mathbb{E}[f_l(X_T^{t,x}) \Phi_{t,(i)}^1(T, x)], \\ W_i^2 v_l(t, x) &= (T - t)^{-1} \mathbb{E}[f_l(X_T^{t,x}) \Phi_{t,(i,i)}^1(T, x)], \end{aligned}$$

where, crucially,  $\Phi_{t,(i)}^1, \Phi_{t,(i,i)}^1$  are independent of  $f_l$ . Then, considering the differences  $W_i v_l(t, x) - W_i v_m(t, x)$  and  $W_i^2 v_l(t, x) - W_i^2 v_m(t, x)$  over compact subsets of  $[0, T] \times \mathbb{R}^N$ , we see that  $(W_i v_l, W_i^2 v_l)_{l \geq 1}$  converges uniformly on compact subsets of  $[0, T] \times \mathbb{R}^N$ . This proves that  $W_i v, W_i^2 v$  exist and are continuous. Now, each  $f_l \in \mathcal{C}_p^\infty(\mathbb{R}^N; \mathbb{R})$ , so associated to each, there is a classical solution  $v_l$  of the PDE (60). Since  $\widetilde{W}_0 v_l = -\frac{1}{2} \sum_{i=1}^d W_i^2 v_l$ , and  $W_i^2 v_l \rightarrow W_i^2 v$  uniformly on compacts in  $[0, T] \times \mathbb{R}^N$ , we get that  $v \in \mathcal{D}_{\widetilde{W}_0}^{1,\infty}([0, T] \times \mathbb{R}^N)$ . Moreover, taking the limit in the PDE satisfied by  $v_l$  shows that it is also satisfied by  $v$ .

*Uniqueness:* Using the Itô formula in Lemma A.14, we have for  $u < T$ ,

$$\begin{aligned} v(u, X_u^{t,x}) &= v(t, x) + \int_t^u \left[ \widetilde{W}_0(s) v(s, X_s^{t,x}) + \frac{1}{2} \sum_{i=1}^d W_i^2(s) v(s, X_s^{t,x}) \right] ds \\ &\quad + \sum_{i=1}^d \int_t^u W_i(s) v(s, X_s^{t,x}) dB_s^i. \end{aligned}$$

Using part (2) of the definition, the drift term is zero and

$$(72) \quad v(u, X_u^{t,x}) = v(t, x) + \sum_{i=1}^d \int_t^u W_i(s) v(s, X_s^{t,x}) dB_s^i.$$

Now, using that  $v$  has polynomial growth and  $X_u^{t,x}$  has moments of all orders, we can easily show that the left-hand side of (72) is square integrable, and so the right-hand side is also. Hence the right-hand side is a true martingale and we can take expectation in (72) to get

$$\mathbb{E}v(u, X_u^{t,x}) = v(t, x)$$

and using part (3) of the definition (continuity of  $v$  at the boundary  $t = T$ ) we can take  $u \nearrow T$  to get

$$\mathbb{E}f(X_T^{t,x}) = v(t, x),$$

which proves uniqueness.  $\square$

**A.5. Derivatives in the direction  $\widetilde{W}_0$ .** In Theorem A.7, we established integration by parts formulae for derivatives of  $x \mapsto P_{t,T}f(x)$  in the directions  $\{\widetilde{W}_{[\alpha]}(t), \alpha \in \mathcal{A}_1(m)\}$ . However,  $0 \notin \mathcal{A}_1(m)$ , so we have no control over derivatives in the direction  $\widetilde{W}_0$ . Using that  $P_{t,T}f(x)$  solves PDE (60) we are now able to estimate derivatives in the  $\widetilde{W}_0$  direction.

PROPOSITION A.15. *Assume  $\widetilde{\text{UFG}}(m)$  holds. Let  $\alpha = (\alpha_1, \dots, \alpha_n) \in \mathcal{A}$  and use the notation  $\widetilde{W}_\alpha(t) = \widetilde{W}_{\alpha_1}(t) \cdots \widetilde{W}_{\alpha_n}(t)$ . Then the function  $v(t, x) := P_{t,T}f(x)$  is differentiable in the directions  $\widetilde{W}_0(t), W_1(t), \dots, W_d(t)$  and the following bounds hold for all  $t \in [0, T)$ : for  $f$  continuous and bounded,*

$$(73) \quad \sup_{x \in \mathbb{R}^N} |\widetilde{W}_\alpha(t)v(t, x)| \leq C \|f\|_\infty (T - t)^{\frac{-(\|\alpha_1\| + \dots + \|\alpha_n\|)}{2}}.$$

For  $f$  Lipschitz,

$$(74) \quad \sup_{x \in \mathbb{R}^N} |\widetilde{W}_\alpha(t)v(t, x)| \leq C \|f\|_{\text{Lip}} (T - t)^{\frac{1 - (\|\alpha_1\| + \dots + \|\alpha_n\|)}{2}}.$$

PROOF. Thinking of the  $\widetilde{W}_0, W_1, \dots, W_d$  as differential operators acting on functions in

$C^\infty([0, T] \times \mathbb{R}^N; \mathbb{R})$ , Corollary 78 in [15] shows that  $\widetilde{W}_\alpha, \alpha \in \mathcal{A}$  satisfies the following convenient identity:

$$(75) \quad \widetilde{W}_\alpha v = \sum_{i=1}^{\|\alpha\|} \sum_{\substack{\beta_1, \dots, \beta_i \in \mathcal{A}_1, \\ \|\beta_1\| + \dots + \|\beta_i\| = \|\alpha\|}} c_{\alpha, \beta_1, \dots, \beta_i} \widetilde{W}_{[\beta_1]} \cdots \widetilde{W}_{[\beta_i]} v,$$

where  $c_{\alpha, \beta_1, \dots, \beta_i} \in \mathbb{R}$ . The importance of this identity is that the left-hand side contains derivatives possibly in the direction  $\widetilde{W}_0$  whereas on the right-hand side, there are only derivatives in directions  $\widetilde{W}_{[\alpha]}, \alpha \in \mathcal{A}_1$  which does not include  $\widetilde{W}_0$ .

Hence,

$$|\widetilde{W}_\alpha v| \leq C \sup_{\substack{\beta_1, \dots, \beta_i \in \mathcal{A}_1, \\ \|\beta_1\| + \dots + \|\beta_i\| = \|\alpha\|}} |\widetilde{W}_{[\beta_1]} \cdots \widetilde{W}_{[\beta_i]} v|,$$

this being exactly the type of term we can control by Theorem A.7.  $\square$

Now, define, for  $\varphi \in C_b^\infty([0, T - 1/R] \times \mathbb{B}(0, R); \mathbb{R})$ , the norm

$$\|\varphi\|_{[0, T-1/R] \times \mathbb{B}(0, R); \infty}^{W, n} := \sum_{\alpha \in \mathcal{A}(n)} \|\widetilde{W}_\alpha \varphi\|_{[0, T-1/R] \times \mathbb{B}(0, R); \infty},$$

and define  $\widehat{\mathcal{D}}^n([0, T - 1/R] \times \mathbb{B}(0, R))$  as the closure of  $\mathcal{C}_b^\infty([0, T - 1/R] \times \mathbb{B}(0, R); \mathbb{R})$  in  $\mathcal{C}_b([0, T - 1/R] \times \mathbb{B}(0, R); \mathbb{R})$  with respect to this norm. Then set

$$\widehat{\mathcal{D}}^\infty([0, T] \times \mathbb{R}^N) := \bigcap_{R \geq 1, n \geq 1} \widehat{\mathcal{D}}^n([0, T - 1/R] \times \mathbb{B}(0, R)).$$

LEMMA A.16. *The function  $v(t, x) := P_{t,T} f(x)$  is a member of  $\widehat{\mathcal{D}}^\infty([0, T] \times \mathbb{R}^N)$  for all  $f \in \mathcal{C}_p(\mathbb{R}^N; \mathbb{R})$ .*

PROOF. We take a sequence  $(f_l)_{l \geq 1}$  of smooth approximations of  $f$  and associate a  $v_l$  to each. For any  $n \in \mathbb{N}$  and any  $\alpha \in \mathcal{A}(n)$ , we can use the identity (75) to write  $\widetilde{W}_\alpha(t)v_l(t, x)$  as a linear combination of terms of the form  $\widetilde{W}_{[\beta]}(t)v_l(t, x)$  where  $\beta \in \mathcal{A}_1(n)$ . This allows us to apply the integration by parts formulae in Theorem A.7 to write

$$\widetilde{W}_\alpha(t)v_l(t, x) = t^{-\|\alpha\|/2} \mathbb{E}[f_l(X_T^{t,x})\Phi_{t,\alpha}(T, x)]$$

for some  $\Phi_{t,\alpha} \in \mathcal{K}_0(t, \mathbb{R})$ . This converges over compact subsets of  $[0, T] \times \mathbb{R}^N$ . □

The above lemma is used in the next section where we need to perform a stochastic Taylor expansion of  $v(t, x) := P_{t,T} f(x)$  for Lipschitz  $f$ .

### A.6. Stochastic Taylor expansion.

PROPOSITION A.17. *Let  $f$  be Lipschitz continuous and assume that  $\widetilde{\text{UFG}}(m)$  holds for some  $m \in \mathbb{N}$ . Then,  $u$ , the solution of equation (60) admits a stochastic Taylor expansion for  $s < T$ ,*

$$u(s, X_s^{t,x}) = \sum_{\alpha \in \mathcal{A}(l)} \widetilde{W}_\alpha u(t, x) I_{t,s}^\alpha(1) + R(l, t, s, x),$$

with the following estimate on the remainder:

$$(76) \quad \sup_{x \in \mathbb{R}^N} \|R(l, s, t, x)\|_2 \leq C \sum_{k=l+1}^{l+2} (T-s)^{-(k-1)/2} (s-t)^{-k/2}.$$

This leads to a one-step cubature error estimate of

$$(77) \quad \begin{aligned} & \sup_{x \in \mathbb{R}^N} |\mathbb{E}[u(s, X_s^{t,x})] - \mathbb{E}_{\mathbb{Q}_{t,s}}[u(s, X_s^{t,x})]| \\ & \leq C \sum_{k=l+1}^{l+2} (T-s)^{-(k-1)/2} (s-t)^{-k/2}. \end{aligned}$$

PROOF. For any  $g \in C_b^\infty([t, s] \times \mathbb{R}^N; \mathbb{R})$ , the following Stratonovich–Taylor expansion is contained in, for example, Kloeden and Platen ([22], Theorem 5.6.1),

$$g(s, X_s^{t,x}) = \sum_{\alpha \in \mathcal{A}(l)} \widetilde{W}_\alpha g(t, x) I_{t,s}^\alpha(1) + R(l, t, s, x)$$

where

$$R(l, t, s, x, g) = \sum_{-\beta \in \mathcal{A}(l), \beta \notin \mathcal{A}(l)} I_{t,s}^\beta(\widetilde{W}_\beta g(\cdot, X^{t,x})).$$

It is not immediate that this expansion is valid for  $g = u$ , the solution of equation (60) since it is not differentiable in all directions. However, the Stratonovich–Taylor expansion follows from repeated application of the Itô formula contained in Lemma A.14. We recall Lemma A.16, which says that  $(t, x) \mapsto P_{t,T} f(x) \in \widehat{\mathcal{D}}^\infty([0, T] \times \mathbb{R}^N)$ . This guarantees we can apply Itô’s formula as many times as we wish and so the Stratonovich–Taylor expansion is still valid. We then have the following estimate for  $g = u$ :

$$\begin{aligned} & \sup_{x \in \mathbb{R}^N} \|R(l, s, t, x)\|_2 \\ & \leq \sum_{-\beta \in \mathcal{A}(l), \beta \notin \mathcal{A}(l)} \|I_{t,s}^\beta(\widetilde{W}_\beta u(\cdot, X^{t,x}))\|_2 \\ (78) \quad & \leq \sum_{j=l+1}^{l+2} \sup_{\beta \in \mathcal{A}(j)} \sup_{(p,x) \in [t,s] \times \mathbb{R}^N} |\widetilde{W}_\beta u(p, x)| \|I_{t,s}^\beta(1)\|_2 \\ & \leq C \sum_{j=l+1}^{l+2} \sup_{\beta \in \mathcal{A}(j)} \sup_{(p,x) \in [t,s] \times \mathbb{R}^N} |\widetilde{W}_\beta u(p, x)| (s-t)^{j/2}, \end{aligned}$$

where we have used the standard moment estimate on iterated Stratonovich integrals  $\|I_{t,s}^\beta(1)\|_2 \leq C(s-t)^{\|\beta\|/2}$ . A similar estimate holds under the one step cubature measure,  $\mathbb{Q}_{t,s}$ :

$$\begin{aligned} & \sup_{x \in \mathbb{R}^N} |\mathbb{E}_{\mathbb{Q}_{t,s}} R(l, t, s, x)| \\ (79) \quad & \leq C \sum_{j=l+1}^{l+2} \sup_{\beta \in \mathcal{A}(j)} \sup_{(p,x) \in [t,s] \times \mathbb{R}^N} |\widetilde{W}_\beta u(p, x)| (s-t)^{j/2}. \end{aligned}$$

This is a standard estimate on iterated integrals of bounded variation paths. The constant  $C$  depends on  $d, l$  and the length of the cubature paths. Inequalities (78) and (79) give us control over the error in approximating  $P_{t,s}g(t, \cdot)$  by

$Q_{t,s}g(t, \cdot),$

$$\begin{aligned} & \sup_{x \in \mathbb{R}^N} |\mathbb{E}[u(s, X_s^{t,x})] - \mathbb{E}_{Q_{t,s}}[u(s, X_s^{t,x})]| \\ &= \sup_{x \in \mathbb{R}^N} |(\mathbb{E} - \mathbb{E}_{Q_{t,s}})R(l, t, s, x)| \\ &\leq C \sum_{j=l+1}^{l+2} \sup_{\beta \in \mathcal{A}(j)} \sup_{(p,x) \in [t,s] \times \mathbb{R}^N} |\widetilde{W}_\beta u(p, x)|(s-t)^{j/2}. \end{aligned}$$

To bound

$$\sup_{(p,x) \in [t_j, t_{j+1}] \times \mathbb{R}^N} |\widetilde{W}_\beta u(p, x)|,$$

we use the estimate provided in (74) and taking the supremum over  $p \in [t_j, t_{j+1}]$ , we get

$$\sup_{(p,x) \in [t_j, t_{j+1}] \times \mathbb{R}^N} \sup_{\beta \in \mathcal{A}(j)} |\widetilde{W}_\beta u(p, x)| \leq C(T - t_{j+1})^{(1-j)/2}. \quad \square$$

## REFERENCES

- [1] ABRAMOWITZ, M. and STEGUN, I. A. (1972). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*.
- [2] ANTONELLI, F. and KOHATSU-HIGA, A. (2002). Rate of convergence of a particle method to the solution of the McKean–Vlasov equation. *Ann. Appl. Probab.* **12** 423–476. [MR1910635](#)
- [3] BAYER, C. and FRIZ, P. K. (2013). Cubature on Wiener space: Pathwise convergence. *Appl. Math. Optim.* **67** 261–278.
- [4] BOSSY, M. (2005). Some stochastic particle methods for nonlinear parabolic PDEs. In *ESAIM: Proceedings* **15** 18–57. EDP Sciences.
- [5] BOSSY, M. and TALAY, D. (1997). A stochastic particle method for the McKean–Vlasov and the Burgers equation. *Math. Comp.* **66** 157–192.
- [6] CARDALIAGUET, P. (2012). Notes on mean field games (from P.-L. Lions’ lectures at Collège de France). <https://www.ceremade.dauphine.fr/~cardaliaguet/MFG20130420.pdf>.
- [7] CATTIAUX, P. and MESNAGER, L. (2002). Hypocoelliptic non-homogeneous diffusions. *Probab. Theory Related Fields* **123** 453–483.
- [8] CHASSAGNEUX, J.-F., CRISAN, D. and DELARUE, F. (2014). A Probabilistic approach to classical solutions of the master equation for large population equilibria. ArXiv e-prints. Available at [arXiv:1411.3009](https://arxiv.org/abs/1411.3009).
- [9] CRISAN, D. and DELARUE, F. (2012). Sharp derivative bounds for solutions of degenerate semi-linear partial differential equations. *J. Funct. Anal.* **263** 3024–3101.
- [10] CRISAN, D. and GHAZALI, S. (2007). On the convergence rates of a general class of weak approximations of SDEs. In *Stochastic Differential Equations: Theory and Applications. Interdiscip. Math. Sci.* **2** 221–248. World Sci. Publ., Hackensack, NJ.
- [11] CRISAN, D., KURTZ, T. G. and LEE, Y. (2014). Conditional distributions, exchangeable particle systems, and stochastic partial differential equations. *Ann. Inst. Henri Poincaré Probab. Stat.* **50** 946–974. [MR3224295](#)

- [12] CRISAN, D., LITTERER, C. and LYONS, T. (2015). Kusuoka–Stroock gradient bounds for the solution of the filtering equation. *J. Funct. Anal.* **268** 1928–1971.
- [13] CRISAN, D. and MANOLARAKIS, K. (2012). Solving backward stochastic differential equations using the cubature method: Application to nonlinear pricing. *SIAM J. Financial Math.* **3** 534–571.
- [14] CRISAN, D. and MANOLARAKIS, K. (2014). Second order discretization of backward SDEs and simulation with the cubature method. *Ann. Appl. Probab.* **24** 652–678. [MR3178494](#)
- [15] CRISAN, D., MANOLARAKIS, K. and NEE, C. (2013). Cubature methods and applications. In *Paris–Princeton Lectures on Mathematical Finance 2013. Lecture Notes in Math.* **2081** 203–316. Springer, Cham.
- [16] CRISAN, D. and MCMURRAY, E. (2016). Smoothing properties of McKean–Vlasov SDEs. *Probab. Theory Related Fields* **171** 97–148. [MR3800831](#)
- [17] CRISAN, D. and ORTIZ-LATORRE, S. (2013). A Kusuoka–Lyons–Victoir particle filter. In *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **469** 20130076.
- [18] DE RAYNAL, P. C. and TRILLOS, C. G. (2015). A cubature based algorithm to solve decoupled McKean–Vlasov forward–backward stochastic differential equations. *Stochastic Process. Appl.* **125** 2206–2255.
- [19] GYURKÓ, L. G. and LYONS, T. J. (2010). Efficient and practical implementations of cubature on Wiener space. In *Stochastic Analysis 2010*. 73–111. Springer, Berlin.
- [20] JOURDAIN, B., MÉLÉARD, S. and WOYCZYNSKI, W. A. (2008). Nonlinear SDEs driven by Lévy processes and related PDEs. *ALEA Lat. Am. J. Probab. Math. Stat.* **4** 1–29.
- [21] JOURDAIN, B., MÉLÉARD, S. and WOYCZYNSKI, W. A. (2008). Nonlinear SDEs driven by Lévy processes and related PDEs. *ALEA Lat. Am. J. Probab. Math. Stat.* **4** 1–29.
- [22] KLOEDEN, P. E. and PLATEN, E. (1992). *Numerical Solution of Stochastic Differential Equations. Applications of Mathematics (New York)* **23**. Springer, Berlin.
- [23] KOHATSU-HIGA, A. and OGAWA, S. (1997). Weak rate of convergence for an Euler scheme of nonlinear SDE’s. *Monte Carlo Methods Appl.* **3** 327–345.
- [24] KOLOKOLTSOV, V. N. (2004). Hydrodynamic limit of coagulation-fragmentation type models of  $k$ -nary interacting particles. *J. Stat. Phys.* **115** 1621–1653.
- [25] KOLOKOLTSOV, V. N. (2010). *Nonlinear Markov Processes and Kinetic Equations. Cambridge Tracts in Mathematics* **182**. Cambridge Univ. Press, Cambridge.
- [26] KOTELENEZ, P. M. and KURTZ, T. G. (2010). Macroscopic limits for stochastic partial differential equations of McKean–Vlasov type. *Probab. Theory Related Fields* **146** 189–222. [MR2550362](#)
- [27] KUNITA, H. (1990). *Stochastic Flows and Stochastic Differential Equations. Cambridge Studies in Advanced Mathematics* **24**. Cambridge Univ. Press, Cambridge.
- [28] KURTZ, T. G. and PROTTER, P. E. (1996). Weak convergence of stochastic integrals and differential equations. In *Probabilistic Models for Nonlinear Partial Differential Equations. Lecture Notes in Math.* **1627** 1–41. Springer, Berlin.
- [29] KUSUOKA, S. (2001). Approximation of expectation of diffusion process and mathematical finance. In *Taniguchi Conference on Mathematics Nara ’98. Adv. Stud. Pure Math.* **31** 147–165. Math. Soc. Japan, Tokyo.
- [30] KUSUOKA, S. (2003). Malliavin calculus revisited. *J. Math. Sci. Univ. Tokyo* **10** 261–277.
- [31] KUSUOKA, S. and STROOCK, D. (1985). Applications of the Malliavin calculus. II. *J. Fac. Sci., Univ. Tokyo, Sect. 1A, Math.* **32** 1–76.
- [32] KUSUOKA, S. and STROOCK, D. (1987). Applications of the Malliavin calculus. III. *J. Fac. Sci., Univ. Tokyo, Sect. 1A, Math.* **34** 391–442.
- [33] LEE, W. and LYONS, T. (2015). The adaptive patched cubature filter and its implementation. arXiv preprint. Available at [arXiv:1509.04239](https://arxiv.org/abs/1509.04239).
- [34] LIONS, P.-L. (2014) Cours au collège de france. Available at <http://www.college-de-france.fr/site/pierre-louis-lions/seminar-2014-11-14-11h15.htm>.

- [35] LITTERER, C. and LYONS, T. (2011). Introducing cubature to filtering. In *The Oxford Handbook of Nonlinear Filtering* 768–796. Oxford Univ. Press, Oxford.
- [36] LYONS, T. and VICTOIR, N. (2004). Cubature on Wiener space. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **460** 169–198.
- [37] MORALE, D., CAPASSO, V. and OELSCHLÄGER, K. (2005). An interacting particle system modelling aggregation behavior: From individuals to populations. *J. Math. Biol.* **50** 49–66.
- [38] NINOMYIA, S. and VICTOIR, N. (2008). Weak approximation scheme of stochastic differential equations and applications to derivatives pricing. *Appl. Math. Finance.* **15** 107–121.
- [39] OGAWA, S. (1995). Some problems in the simulation of nonlinear diffusion processes. *Math. Comput. Simulation* **38** 217–223.
- [40] RICKETSON, L. F. (2015). A multilevel Monte Carlo method for a class of McKean–Vlasov processes. Available at <https://arxiv.org/abs/1508.02299>.
- [41] TACHET, R. (2011). Des Combes. Calibration non paramétriques de modèles en finance. Theses, Ecole Centrale Paris.
- [42] TEICHMANN, J. (2006). Calculating the Greeks by cubature formulae. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **462** 647–670. [MR2269681](#)

DEPARTMENT OF MATHEMATICS  
IMPERIAL COLLEGE LONDON  
180 QUEEN’S GATE  
LONDON SW7 2AZ  
UNITED KINGDOM  
E-MAIL: [d.crisan@ic.ac.uk](mailto:d.crisan@ic.ac.uk)