SIAM J. NUMER. ANAL.
Vol. 57, No. 2, pp. 563–591
© 2019 Society for Industrial and Applied Mathematics

Downloaded 03/28/19 to 130.70.8.131. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

AN ARBITRARY HIGH ORDER WEAK APPROXIMATION OF SDE AND MALLIAVIN MONTE CARLO: ANALYSIS OF PROBABILITY DISTRIBUTION FUNCTIONS*

TOSHIHIRO YAMADA†

Abstract. This paper provides an arbitrary high order weak approximation scheme for multidimensional Stratonovich stochastic differential equations using Malliavin calculus. The scheme efficiently works whether the test function is smooth or not. The Malliavin Monte Carlo method, a simple numerical algorithm, is introduced to implement the scheme. Numerical examples illustrate the validity of the method.

Key words. stochastic differential equations, weak approximation, Malliavin calculus, Monte Carlo simulation

AMS subject classifications. 60H07, 65C05, 65C30, 90C31, 91G60

DOI. 10.1137/17M114412X

1. Introduction.

The theory of weak approximation of stochastic differential equations (SDEs) is so important in many fields since it gives useful computational schemes for various quantities in stochastic modeling. The most fundamental weak approximation is the Euler–Maruyama scheme introduced by Maruyama [9], which has been widely used in practice as the benchmark method. The scheme is known as the first order weak approximation; in other words, the order is $O(1/n)$ for the number of time steps n. Remarkably, the result still holds even if the condition of the test function is measurable only (Bally and Talay [2]).

Higher order weak approximations have been studied to obtain more efficient numerical schemes; see the book [13] for more details. Basically, higher order schemes are constructed through the stochastic Taylor expansion, where the test functions are usually assumed to be smooth. Then, the smoothness conditions are relaxed by Malliavin calculus (Kusuoka [14]). Let us consider the higher order weak approximation problem of discretizing expectation $(P_T f)(x) = \mathbb{E}[f(X(T,x))]$ with an irregular test function f and a solution $X(T,x)$ at $T > 0$ of an N-dimensional Stratonovich SDE starting from $x \in \mathbb{R}^N$ driven by d-dimensional Brownian motion $\{B_t\}_{t\geq 0}$. For the Lipschitz continuous test function case, i.e., not a sufficiently smooth case, Kusuoka [14], [16] and Lyons and Victoir [22] showed a general weak scheme which is today called the cubature method on Wiener space or the KLV (Kusuoka–Lyons–Victoir) scheme. Key features of the cubature method are to choose appropriate finite variation paths instead of Brownian motions and use a nonuniform partition in the discretization to deal with Lipschitz continuous test function f (with the Lipschitz constant $C_{\mathrm{Lip}}[f]$). The m-th order weak approximation is obtained from the estimate

$$\left|P_T f(x) - Q^{\mathrm{KLV},m}_{(s_1)} \cdots Q^{\mathrm{KLV},m}_{(s_n)} f(x)\right| \leq C_{\mathrm{Lip}}[f]\Big(C\sum_{k=1}^{n-1}\Big(\sum_{i=2m+1}^{2(m+1)} \frac{s_k^{m+1}}{(T-t_k)^{(i-1)/2}}\Big) + c s_n^{\frac{1}{2}}\Big) \tag{1}$$

for some $C, c > 0$, where the operator $Q^{\mathrm{KLV},m}_{(t)}$ is constructed based on the cubature

*Received by the editors August 18, 2017; accepted for publication (in revised form) December 21, 2018; published electronically March 19, 2019.
http://www.siam.org/journals/sinum/57-2/M114412.html
Funding: This work is supported by JSPS KAKENHI (grant 16K13773) from MEXT, Japan, and a research fund from Tokio Marine Kagami Memorial Foundation.
†Hitotsubashi University, Tokyo, Japan (toshihiro.yamada@r.hit-u.ac.jp).

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

formula on Wiener space and satisfies $\mathbb{E}[\varphi(X(t,x))] = Q_{(t)}^{\mathrm{KLV},m}\varphi(x) + O(t^{m+1})$, $t \downarrow 0$ for a smooth function φ. Here, $t_k = T(1-(1-k/n)^{\gamma})$, $k = 0, 1, \cdots, n$, $\gamma > 0$, is a nonuniform partition (with its intervals $s_k = t_k - t_{k-1}$) which was initially introduced by Kusuoka [14] using Malliavin calculus (Kusuoka's trick), and we call this nonuniform partition the Kusuoka partition [10]. The error of (1) will be $O(n^{-m})$ by choosing the parameter $\gamma > 1$ appropriately. The efficient algorithms for the cubature method were proposed after the works of [14], [22]; see [4], [10], [11], [20], [26], for example. The algorithms of the cubature method are obtained for some specific weak orders, but the explicit arbitrary order method is unknown in general.

This paper introduces a general arbitrary high order weak approximation scheme using a Malliavin weight, a stochastic weight of a linear combination of explicit polynomials of Brownian motions, which will work under *uniform partition* whether the test function is smooth or not:

$$\left| P_T f(x) - (Q_{(T/n)}^{\mathrm{Mall},m})^n f(x) \right| \leq C\|f\|_\infty \frac{1}{n^m}. \tag{2}$$

Here, the operator $Q_{(t)}^{\mathrm{Mall},m}$ is constructed by the expectation of form $Q_{(t)}^{\mathrm{Mall},m}\varphi(x) = \mathbb{E}[\varphi(\overline{X}(t,x))\widetilde{\mathcal{M}}^{(m)}(t,x,B_t)]$ with a Gaussian process $\overline{X}(t,x)$ and a Malliavin weight $\widetilde{\mathcal{M}}^{(m)}(t,x,B_t)$, $t > 0$, $x \in \mathbb{R}^N$, satisfying $\mathbb{E}[\varphi(X(t,x))] = Q_{(t)}^{\mathrm{Mall},m}\varphi(x) + O(t^{m+1})$, $t \downarrow 0$, for a smooth function φ. The scheme (2) will be an extension of [30], [33], [34] in the sense that (2) is the arbitrary order method and does not rely on the Kusuoka partition; nevertheless we treat an irregular test function. We call $\widetilde{\mathcal{M}}^{(m)}(t,x,B_t)$ *universal weight* since it gives a universal high order scheme even if the test function is smooth or not under uniform partition. The use of uniform partition would be numerically optimal but developing a theoretical high order scheme with the uniform grids is a difficult task when f is measurable only. We briefly explain why constructing the higher order scheme is difficult under uniform partition if the test function is irregular and how the computational cost will be reduced through the universal weight.

In general, weak approximation is constructed based on a small time approximation. For the construction of the weak approximation with Malliavin weights for an irregular test function f, a trade-off between the small time approximation and the weak approximation exists. There are two cases as follows:

(i) [Small time approximation of $P_t\varphi$] Smoothness of $\varphi \downarrow$
[Weak approximation of $P_T f$ for nonsmooth f]
Complexity on computing Malliavin weight $\uparrow$,
Nonuniformity of Kusuoka partition $\downarrow$.

(ii) [Small time approximation of $P_t\varphi$] Smoothness of $\varphi \uparrow$
[Weak approximation of $P_T f$ for nonsmooth f]
Complexity on computing Malliavin weight $\downarrow$,
Nonuniformity of Kusuoka partition $\uparrow$.

Takahashi and Yamada [30] took the approach (i) and developed a weak scheme for irregular test functions using Watanabe's small noise expansion [32] with respect to a model parameter ε of SDEs. In the numerical examples in [30], only a lower order discretization was obtained, but the explicit higher order scheme was not obtained because of its large complexity.

Yamada in [33] and [34] avoided the approach (i) and obtained an explicit high order scheme using approach (ii) by controlling the Kusuoka partition. Using the

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 03/28/19 to 130.70.8.131. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Downloaded 03/28/19 to 130.70.8.131. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

method of [33], one can show

$$\left|P_1 f(x) - Q^{\mathrm{Mall},m}_{(s_1)} \cdots Q^{\mathrm{Mall},m}_{(s_n)} f(x)\right| \le \|f\|_\infty \Big(C \sum_{k=1}^{n-1} \Big(\sum_{i=1}^{e(m)} \frac{s_k^{m+1}}{(1-t_k)^{|\alpha_i|/2}} \Big) + c s_n^{\frac{1}{2}+q} \Big) \tag{3}$$

for some $C, c, e(m), q \ge 0$, where $t_k = (1-(1-k/n)^\gamma)$, $k = 0, 1, \cdots, n$, is the Kusuoka partition with a parameter $\gamma > 0$ and $s_k = t_k - t_{k-1}$, $k = 1, \cdots, n$, are its intervals. The error in (3) will be $O(n^{-m})$ when we appropriately choose the parameter $\gamma > 1$. Although a higher order method can be obtained through the approach (ii), there is room for improvement since the error estimate would be not optimal. This is the reason we introduce a new scheme in this paper.

In the current paper, we propose the universal weight to overcome these problems. In the main theorem (Theorem 3.13), we obtain an arbitrary high order method with uniform partition and an improved error estimate. The weak approximation (2) is particularly given by

$$\left|P_1 f(x) - (Q^{\mathrm{Mall},m}_{(1/n)})^n f(x)\right| \le C\|f\|_\infty \Big(\sum_{k=1}^{n} \sum_{i=1}^{e(m)} \left(\frac{1}{n}\right)^{m+1} \Big) = C' \|f\|_\infty \frac{1}{n^m} \tag{4}$$

for some $C, C' > 0$. Note that the factors $\frac{1}{(1-t_k)^{|\alpha_i|/2}}$, $1 \le k \le n-1$, $i = 1, \cdots, e(m)$, in (3) are all removed, and then the approximation (4) will be sharper than (3).

The scheme of the paper enables us to approximate quantities such as probability distribution functions $P(X(T,x) \le z) = \mathbb{E}[f(X(T,x))]$ where $f(\cdot) = \mathbf{1}_{\{\cdot \le z\}}$, $z \in \mathbb{R}^N$, more efficiently compared to the Euler–Maruyama scheme in [2] and the schemes in [30], [33]. The universal weight is obtained through Itô–Stratonovich calculus and the Malliavin integration by parts with the theory of Kusuoka–Stroock and is implemented by *Malliavin Monte Carlo*, an efficient weighted simulation scheme. Numerical examples are shown to illustrate the validity of the scheme. In particular, we check that the accuracy of (4) outperforms that of (3) in numerical experiments.

The organization of the paper is as follows. Section 2 summarizes basic notation and results on stochastic calculus. In section 3, we propose an arbitrary high order weak approximation using Malliavin calculus. The main result is shown in Theorem 3.13. Section 4 provides various numerical examples, and section 5 concludes on the method. The proofs of theorems and technical lemmas are in the appendices.

2. Stochastic calculus.

2.1. Notation. For a multi-index $\alpha = (\alpha_1, \cdots, \alpha_k) \in \{0, 1, \cdots, d\}^k$, $k = 0, 1, \cdots$, we define $|\alpha| = k$ and $\|\alpha\| = \#\{i; \alpha_i \ne 0\} + 2\#\{i; \alpha_i = 0\}$. For two multi-indices $\alpha = (\alpha_1, \cdots, \alpha_k)$ and $\beta = (\beta_1, \cdots, \beta_l)$, define a multi-index $\alpha * \beta = (\alpha_1, \cdots, \alpha_k, \beta_1, \cdots, \beta_l)$. For a multi-index $\alpha = (\alpha_1, \cdots, \alpha_k) \in \{0, 1, \cdots, d\}^k$, define $\alpha^{-i} = (\alpha_1, \cdots, \alpha_{k-i}) \in \{0, 1, \cdots, d\}^{k-i}$, $i \le k$, and $\alpha^\star = (\alpha_1^\star, \cdots, \alpha_{K(\alpha)}^\star) = (\alpha_{j_1}, \cdots, \alpha_{j_{K(\alpha)}})$, where $K(\alpha) = k - \#\{i; \alpha_i = 0\}$ and $\alpha_{j_i} \ne 0$, $i = 1, \cdots, K(\alpha)$, $j_1, \cdots, j_{K(\alpha)} \in \{1, \cdots, k\}$. For $k \in \mathbb{N}$, let S_k be symmetric group of permutations of the index set $\{1, \cdots, k\}$. For a permutation $\sigma \in S_k$ and a word $I = (i_1, \cdots, i_k)$, we write $\sigma \cdot I = (i_{\sigma(1)}, \cdots, i_{\sigma(k)})$.

For $N, n \in \mathbb{N}$, let $C_b^\infty(\mathbb{R}^N, \mathbb{R}^n)$ (resp., $C_p^\infty(\mathbb{R}^N, \mathbb{R}^n)$) be the set of all infinitely continuously differentiable functions $f : \mathbb{R}^N \to \mathbb{R}^n$ such that f and all of its partial derivatives at any order are bounded (resp., at most polynomial growth). We write $C_b^\infty(\mathbb{R}^N)$ (resp., $C_p^\infty(\mathbb{R}^N)$) for $C_b^\infty(\mathbb{R}^N, \mathbb{R})$ (resp., $C_p^\infty(\mathbb{R}^N, \mathbb{R})$). Further, let $\mathcal{S}(\mathbb{R}^N)$ and $\mathcal{S}'(\mathbb{R}^N)$ be the space of Schwartz rapidly decreasing functions $\mathbb{R}^N \to \mathbb{R}$ and the space of Schwartz distributions, respectively; that is, $\mathcal{S}'(\mathbb{R}^N)$ is the dual of $\mathcal{S}(\mathbb{R}^N)$.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 03/28/19 to 130.70.8.131. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

For a function $f : \mathbb{R}^N \to \mathbb{R}$, we write $\|f\|_\infty = \sup_{x\in\mathbb{R}^N} |f(x)|$. For $f \in C_b^\infty(\mathbb{R}^N)$ and $\alpha = (\alpha_1, \cdots, \alpha_k) \in \{1, \cdots, N\}^k$, $k \in \mathbb{N}$, we use abbreviated notation $\partial^\alpha f(x) = \partial_{\alpha_1} \cdots \partial_{\alpha_k} f(x) = \frac{\partial^k f(x)}{\partial x_{\alpha_1} \cdots \partial x_{\alpha_k}}$ and $\|\nabla^k f\|_\infty = \max_{\alpha_1,\cdots,\alpha_k\in\{1,\cdots,N\}} \|\frac{\partial^k}{\partial x_{\alpha_1} \cdots \partial x_{\alpha_k}} f\|_\infty$, $k \in \mathbb{N}$.

2.2. Analysis on Wiener space. We summarize notation and basic facts on Malliavin calculus; see [12], [27] for more details.

Let $(\mathscr{W}, H, P)$ be the Wiener space, that is, $\mathscr{W}$ is the space of continuous functions $w : [0, \infty) \to \mathbb{R}^d$ such that $w(0) = 0$, H is the Cameron–Martin space (i.e., a Hilbert space $H = \{h \in \mathscr{W}; h$ absolutely continuous, $\dot{h} \in L^2([0, \infty), \mathbb{R}^d)\}$ equipped with the inner product $\langle h_1, h_2\rangle_H = \sum_{j=1}^d \int_0^\infty \dot{h}_1^j(s)\dot{h}_2^j(s)ds$ for $h_1, h_2 \in H$), and P is the Wiener measure. Throughout this paper, we work on the probability space $(\mathscr{W}, \mathscr{B}(\mathscr{W}), P)$, where $\mathscr{B}(\mathscr{W})$ is the Borel field over $\mathscr{W}$. For a Hilbert space V with the norm $\|\cdot\|_V$ and $p \in [1, \infty)$, the L^p-space of V-valued Wiener functionals is denoted by $L^p(\mathscr{W}; V)$; that is, $L^p(\mathscr{W}; V)$ is a real Banach space of all P-measurable functionals $F : \mathscr{W} \to V$ such that $\|F\|_p = \mathbb{E}[\|F\|_V^p]^{1/p} < \infty$ with the identification $F = G$ if and only if $F(w) = G(w)$, a.s. When $V = \mathbb{R}$, we write $L^p(\mathscr{W})$. Let $B(h)$ be the Wiener integral $B(h) = \sum_{j=1}^d \int_0^\infty \dot{h}^j(s)dB_s^j$ for $h \in H$. Let $\mathscr{S}(\mathscr{W})$ denote the class of smooth random variables of the form $F = f(B(h_1), \cdots, B(h_n))$ where $f \in C_p^\infty(\mathbb{R}^n)$, $h_1, \cdots, h_n \in H$, $n \geq 1$. For $F \in \mathscr{S}(\mathscr{W})$, we define the derivative DF as the H-valued random variable $DF = \sum_{i=1}^n \partial_i f(B(h_1), \cdots, B(h_n))h_i$ and $D_{j,s}F = \sum_{i=1}^n \partial_i f(B(h_1), \cdots, B(h_n))\dot{h}_i^j(s)$, $s \geq 0$, $j = 1, \cdots, d$. For $F \in \mathscr{S}(\mathscr{W})$, we set D^jF, $j \in \mathbb{N}$, as the $H^{\otimes j}$-valued random variable obtained iterating j-times the operator D. For a real separable Hilbert space V, consider $\mathscr{S}_V$ of V-valued smooth Wiener functionals of the form $F = \sum_{i=1}^l F_i v_i$, $v_i \in V$, $F_i \in \mathscr{S}(\mathscr{W})$, $i \leq l$, $l \in \mathbb{N}$. Define $D^jF = \sum_{i=1}^l D^jF_i \otimes v_i$, $j \in \mathbb{N}$. Then D^j is a closable operator from $\mathscr{S}_V$ into $L^p(\mathscr{W}; H^{\otimes j} \otimes V)$ for any $p \in [1, \infty)$. For $k \in \mathbb{N}$, $p \in [1, \infty)$, we define $\|F\|_{k,p,V}^p = \mathbb{E}[\|F\|_V^p] + \sum_{j=1}^k \mathbb{E}[\|D^jF\|_{H^{\otimes j}\otimes V}^p]$, $F \in \mathscr{S}_V$. Then the space $\mathbb{D}^{k,p}(V)$ is defined as the completion of $\mathscr{S}_V$ with respect to the norm $\|F\|_{k,p,V}$. Moreover, let $\mathbb{D}^\infty(V)$ be the space of smooth Wiener functionals in the sense of Malliavin $\mathbb{D}^\infty(V) = \cap_{p\geq 1} \cap_{k\in\mathbb{N}} \mathbb{D}^{k,p}(V)$. We write $\mathbb{D}^{k,p}$, $k \in \mathbb{N}$, $p \in [1, \infty)$, and $\mathbb{D}^\infty$ when $V = \mathbb{R}$. Let δ be an unbounded operator from $L^2(\mathscr{W}; H)$ into $L^2(\mathscr{W})$ such that the domain of δ, denoted by $\mathrm{Dom}(\delta)$, is the set of H-valued square integrable random variables u such that $|\mathbb{E}[\langle DF, u\rangle_H]| \leq C\|F\|_2$ for all $F \in \mathbb{D}^{1,2}$ where C is some constant depending on u, and if $u \in \mathrm{Dom}(\delta)$, $\delta(u)$ is characterized by $\mathbb{E}[F\delta(u)] = \mathbb{E}[\langle DF, u\rangle_H]$ for all $F \in \mathbb{D}^{1,2}$. $\delta(u)$ is called the Skorohod integral of the process u. The operator δ is continuous from $\mathbb{D}^\infty(H) \subset \mathrm{Dom}(\delta)$ into $\mathbb{D}^\infty$.

For $F = (F^1, \cdots, F^N) \in (\mathbb{D}^\infty)^N$, define the Malliavin covariance matrix of F, $\sigma^F = (\sigma_{ij}^F)_{1\leq i,j\leq N}$, by $\sigma_{ij}^F = \langle DF^i, DF^j\rangle_H = \sum_{k=1}^d \int_0^\infty D_{k,s}F^i D_{k,s}F^j ds$, $1 \leq i, j \leq N$. We say that $F \in (\mathbb{D}^\infty)^N$ is nondegenerate if the matrix σ^F is invertible a.s. and satisfies $\|(\det \sigma^F)^{-1}\|_p < \infty$, $1 \leq p < \infty$. Under the nondegeneracy, we have the Malliaivin integration by parts (see Proposition 2.14 of [27], for example).

PROPOSITION 2.1. *Let $F \in (\mathbb{D}^\infty)^N$ be nondegenerate and $G \in \mathbb{D}^\infty$. Then for any $f \in C_b^\infty(\mathbb{R}^N)$ and multi-index $\alpha = (\alpha_1, \cdots, \alpha_k) \in \{1, \cdots, N\}^k$, $k \in \mathbb{N}$, there exists $H_\alpha(F, G)$ such that*

$$\mathbb{E}[\partial^\alpha f(F)G] = \mathbb{E}[f(F)H_\alpha(F, G)]. \tag{5}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 03/28/19 to 130.70.8.131. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Moreover, $H_\alpha(F,G)$ is recursively given by

$$H_{(i)}(F,G) = \delta\Big(\sum_{j=1}^{N} G\gamma_{ij}^F DF^j\Big), \quad i = 1,\cdots,N,$$
$$H_\alpha(F,G) = H_{(\alpha_k)}(F, H_{(\alpha_1,\cdots,\alpha_{k-1})}(F,G)),$$

where γ^F is the inverse matrix of the Malliavin covariance of F.

The Wiener functional of type $H_\alpha(F,G)$ above is called Malliavin weight.

We introduce distributions on the Wiener space. Let $\mathbb{D}^{-\infty}$ be the space of Watanabe distributions, i.e., the dual of $\mathbb{D}^\infty$. Then, the natural coupling $\langle F,G\rangle$ is well-defined for $F \in \mathbb{D}^{-\infty}$, $G \in \mathbb{D}^\infty$ as an extension of the usual expectation of Wiener functionals. For a Schwartz distribution $T \in \mathcal{S}'(\mathbb{R}^N)$ and a nondegenerate Wiener functional $F \in (\mathbb{D}^\infty)^N$, the composition $T(F) = T \circ F$ is characterized as $T(F) \in \mathbb{D}^{-\infty}$, and moreover for $G \in \mathbb{D}^\infty$, we can extend the integration by parts (5) as

$$\langle \partial_i T(F), G\rangle = \langle T(F), H_{(i)}(F,G)\rangle, \qquad i = 1,\cdots,N. \tag{6}$$

For a bounded measurable function $\phi : \mathbb{R}^N \to \mathbb{R}$, a nondegenerate $F \in (\mathbb{D}^\infty)^N$, and $G \in \mathbb{D}^\infty$ the usual expectation can be represented as follows:

$$\mathbb{E}[\phi(F)G] = \int_{\mathbb{R}^N} \phi(y)\langle \delta_y(F), G\rangle dy, \tag{7}$$

where $\delta_y(\cdot) \in \mathcal{S}'(\mathbb{R}^N)$ is the Dirac's delta function mass at $y \in \mathbb{R}^N$. For example, see [23], [24], [29] for the computation with Watanabe distributions.

3. Weak approximation for probability distribution functions.

3.1. Stratonovich SDE. On the Wiener space, we consider a d-dimensional Brownian motion $\{B_t\}_{t\geq 0} = \{(B_t^1,\cdots,B_t^d)\}_{t\geq 0}$ and the solution to the following N-dimensional Stratonovich-type SDE $\{X(t,x)\}_{t\geq 0}$:

$$dX(t,x) = V_0(X(t,x))dt + \sum_{i=1}^{d} V_i(X(t,x)) \circ dB_t^i, \qquad X(0,x) = x \in \mathbb{R}^N,$$

where V_i, $i = 0,1,\cdots,d$, satisfy the following conditions:

[A1] $V_i \in C_p^\infty(\mathbb{R}^N,\mathbb{R}^N)$, $i = 0,1,\cdots,d$.

[A2] There exists $\epsilon > 0$ such that $\sum_{i=1}^d V_i(x)\otimes V_i(x) > \epsilon I_N$ for all $x \in \mathbb{R}^N$.

Here, given $v \in \mathbb{R}^N$, $v \otimes v$ is the tensor product of v with itself, and I_N is the $N \times N$ identity matrix.

For $j = 1,\cdots,N$, each component $X^j(t,x)$, $t \geq 0$, $x \in \mathbb{R}^N$, can be written in the following integral form:

$$X^j(t,x) = x_j + \int_0^t V_0^j(X(s,x))ds + \sum_{i=1}^{d}\int_0^t V_i^j(X(s,x)) \circ dB_s^i.$$

Remark 1 (nondegeneracy of $X(t,x)$). Under **[A1]** and **[A2]**, $X(t,x) \in (\mathbb{D}^\infty)^N$, $t \in (0,T]$, is nondegenerate, and then we can apply integration by parts for $F = X(t,x)$ in (5); see [17], for instance. Further, $X(t,x)$ has a smooth density.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 03/28/19 to 130.70.8.131. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Let $\{P_t\}_t$ be a semigroup of linear operators given by $(P_t\varphi)(x) = E[\varphi(X(t,x))]$, $(t,x) \in [0,T] \times \mathbb{R}^N$, where $\varphi : \mathbb{R}^N \to \mathbb{R}$ is a test function appropriately chosen.

The purpose of this paper is to construct an arbitrary order discretization scheme with a time step $n \in \mathbb{N}$ for the target expectation

$$(P_T f)(x) = \mathbb{E}[f(X(T,x))], \quad T \geq 1,\ x \in \mathbb{R}^N,$$

where $f : \mathbb{R}^N \to \mathbb{R}$ is assumed to be bounded and measurable only. In particular, the Heaviside-type function $f(\cdot) = \mathbf{1}_{\{\cdot \leq z\}}$ (or $\mathbf{1}_{\{\cdot \geq z\}}$), $z \in \mathbb{R}^N$, is of interest; in other words, we aim to compute the probability distribution function $(P_T f)(x) = P(X(T,x) \leq z)$. We show a discretization scheme using Stratonovich–Taylor expansion and Malliavin calculus.

We introduce iterated integrals of Stratonovich and Itô type.

Definition 3.1. *Given a multi-index $J = (j_1, \cdots, j_k) \in \{0, 1, \cdots, d\}^k$, we define the iterated Stratonovich integral*

$$\mathbb{B}^{\mathrm{Strat}}_{J,t} = \int_{0<t_1<\cdots<t_k<t} \circ dB^{j_1}_{t_1} \circ \cdots \circ dB^{j_k}_{t_k}.$$

Also, the iterated Itô integral is denoted by

$$\mathbb{B}^{\mathrm{Itô}}_{J,t} = \int_{0<t_1<\cdots<t_k<t} dB^{j_1}_{t_1} \cdots dB^{j_k}_{t_k}.$$

Here, $\circ dB^0_t$ and dB^0_t represent $\circ dB^0_t = dB^0_t = dt$.

The next two basic lemmas (Lemmas 3.2 and 3.4) will play a role in the derivation of arbitrary order small time approximation. Lemma 3.2 suggests that the products of Stratonovich iterated integrals can be written by the sum of single Stratonovich iterated integrals (see [10] or [6], [21] for the derivation).

Lemma 3.2. *For multi-indices $I \in \{0, 1, \cdots, d\}^m$ and $J \in \{0, 1, \cdots, d\}^n$, we have*

$$\mathbb{B}^{\mathrm{Strat}}_{I,t}\mathbb{B}^{\mathrm{Strat}}_{J,t} = \sum_{\substack{\sigma \in S_{m+n}, s.t. \\ \sigma(1)<\cdots<\sigma(m),\ \sigma(m+1)<\cdots<\sigma(m+n)}} \mathbb{B}^{\mathrm{Strat}}_{\sigma^{-1}\cdot(I*J),t}.$$

Also, the Stratonovich iterated integral can be transformed by a linear combinations of Itô integrals. To show this, we introduce the following relation between multi-indices (Definition 2.4 of [10]).

Definition 3.3. *Let us consider a partition $(J_1, \cdots, J_k)$ of a multi-index J, i.e., $J = J_1 * \cdots * J_k$, for some k such that $\|J_i\| \leq 2$ for $i = 1, \cdots, k$. Suppose there exists a multi-index I which can be partitioned into subindices $I_1, \cdots, I_k$ such that $I = I_1 * \cdots * I_k$ and for each $i = 1, \cdots, k$ either $J_i = I_i$ with both length 1 or $J_i = (l, l)$ and $I_i = (0)$ for some $l \in \{1, \cdots, d\}$. Then we will say that I is related to J through the partitions $(I_1, \cdots, I_k)$ and $(J_1, \cdots, J_k)$ of length k and denote this relationship by $J \sim_k I$, where k denotes the number of subsets in the related partitions. When $J \sim_k I$, we define $\eta(J, I) := \#\{i; J_i \neq I_i\}$.*

Using the notation, we get the following Stratonovich–Itô transformation; see Lemma 2.1 of [10].

Lemma 3.4. *For any multi-index J,*

$$\mathbb{B}^{\mathrm{Strat}}_{J,t} = \sum_{J \sim_k I,\ k \in \mathbb{N}} \frac{1}{2^{\eta(J,I)}} \mathbb{B}^{\mathrm{Itô}}_{I,t}.$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 03/28/19 to 130.70.8.131. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Let $\hat{V}_i$, $i = 0, 1, \cdots, d$, be differential operators which are identified as vector fields on $\mathbb{R}^N$, i.e., $\hat{V}_i\varphi(x) = \sum_{j=1}^N V_i^j(x)\partial\varphi(x)/\partial x_j$ for $\varphi \in C_b^\infty(\mathbb{R}^N)$. By iteratively using the Itô formula, the Stratonovich–Taylor expansion holds:

$$X(t,x) = x + \sum_{i=1}^d V_i(x)B_t^i + \sum_{k=2}^e \sum_{\alpha\in\{0,1,\cdots,d\}^k} \hat{V}_{\alpha_1}\cdots\hat{V}_{\alpha_{k-1}}V_{\alpha_k}(x)\mathbb{B}_{\alpha,t}^{\mathrm{Strat}} + r_{e+1}(t,x), \tag{8}$$

$t \in [0,T]$, where $r_{e+1}(t,x)$ is the residual; see [13], [22]. We define an N-dimensional Gaussian process $\{\overline{X}(t,x)\}_{t\geq 0}$, $x \in \mathbb{R}^N$, as

$$\overline{X}(t,x) = x + \sum_{i=1}^d V_i(x)B_t^i, \quad t \in [0,T]. \tag{9}$$

Roughly speaking, our approximation for $\mathbb{E}[\varphi(X(t,x))]$, $\varphi \in C_b^\infty(\mathbb{R}^N)$, is based on a type of Taylor expansion around $\mathbb{E}[\varphi(\overline{X}(t,x))]$:

$$\begin{aligned}\mathbb{E}[\varphi(X(t,x))] &= \mathbb{E}[\varphi(\overline{X}(t,x))] \\ &+ \sum_{k=1}^m \frac{1}{k!}\sum_{I\in\{1,\cdots,N\}^k} \mathbb{E}\Big[\partial_{I_1}\cdots\partial_{I_k}\varphi(\overline{X}(t,x))\prod_{l=1}^k (X^{I_l}(t,x) - \overline{X}^{I_l}(t,x))\Big] + \cdots.\end{aligned} \tag{10}$$

Then, the terms $\mathbb{E}\big[\partial_{I_1}\cdots\partial_{I_k}\varphi(\overline{X}(t,x))\prod_{\nu=1}^k \mathbb{B}_{\alpha^\nu,t}^{\mathrm{Strat}}\big]$, $\|\alpha^\nu\| \geq 2$, $\nu = 1, \cdots, k$, appear in the approximation (10). Applying Lemmas 3.2 and 3.4 and the Malliavin integration by parts iteratively, we have the following representation without Stratonovich (or Itô) iterated integrals.

LEMMA 3.5. *Assume* **[A1]** *and* **[A2]**. *Let* $k \in \mathbb{N}$ *and* $\alpha^\nu \in \{0,1,\cdots,d\}^{r(\nu)}$, $r(\nu) \in \mathbb{N}$, $\nu = 1,\cdots,k$, *be multi-indices such that* $\|\alpha^\nu\| \geq 2$, $\nu = 1,\cdots,k$. *Then, for* $g \in C_b^\infty(\mathbb{R}^N)$, *we have*

$$\begin{aligned}\mathbb{E}\left[g(\overline{X}(t,x))\prod_{\nu=1}^k \mathbb{B}_{\alpha^\nu,t}^{\mathrm{Strat}}\right] = &\sum_{\substack{\sigma_l\in S_{m_l+n_l},\ l=1,\cdots,k-1,\\ s.t.\ m_l=\sum_{i=1}^l r(i),\ n_l=r(l+1),\\ \sigma_l(1)<\cdots<\sigma_l(m_l),\ \sigma_l(m_l+1)<\cdots<\sigma_l(m_l+n_l)}} \sum_{\substack{\gamma,\beta\ s.t.\ \beta\sim_p\gamma,\ p\in\mathbb{N},\\ \beta=\sigma_{k-1}^{-1}\cdot(\cdots(\sigma_2^{-1}\cdot(\sigma_1^{-1}\cdot(\alpha^1*\alpha^2)*\alpha^3)*\cdots)*\alpha^k)}} \\ &\sum_{J\in\{1,\cdots,N\}^{K(\gamma)}} \prod_{h=1}^{K(\gamma)} V_{\gamma_h^*}^{J_h}(x)\frac{1}{|\gamma|!}t^{|\gamma|}\mathbb{E}\left[g(\overline{X}(t,x))H_{(J_1,\cdots,J_{K(\gamma)})}(\overline{X}(t,x),1)\right].\end{aligned}$$

Proof. See Appendix A. ∎

We remark that the weights $H_{(J_1,\cdots,J_k)}(\overline{X}(t,x),1)$, $(J_1,\cdots,J_k) \in \{1,\cdots,N\}^k$, $k \in \mathbb{N}$, appearing in Lemma 3.5 are represented by polynomials of Brownian motions. By computation of Skorohod integral (see Proposition 1.3.3 of [27], for example) we can see

$$\begin{aligned}&H_{(i)}(\overline{X}(t,x),G) \\ &\quad = \sum_{k=1}^d\sum_{j=1}^N t^{-1}[\Lambda^{-1}]_{ij}(x)V_k^j(x)\Big[GB_t^k - \int_0^t D_{k,s}G ds\Big], \quad G \in \mathbb{D}^\infty,\ 1 \leq i \leq N,\end{aligned} \tag{11}$$

where $\Lambda^{-1}(x)$ is the inverse matrix of the deterministic matrix $\Lambda(x) = (\Lambda_{ij}(x))_{1\leq i,j\leq N}$, $x \in \mathbb{R}^N$, defined by $\Lambda_{ij}(x) = \sum_{k=1}^d V_k^i(x)V_k^j(x)$, $1 \leq i,j \leq N$. In particular, we have $H_{(i)}(\overline{X}(t,x),1) = \sum_{k=1}^d\sum_{j=1}^N t^{-1}[\Lambda^{-1}]_{ij}(x)V_k^j(x)B_t^k$, $1 \leq i \leq N$. Then, for a multi-index $J \in \{1,\cdots,N\}^k$, the Malliavin weight $H_{(J_1,\cdots,J_k)}(\overline{X}(t,x),1)$ is given by kth order Brownian polynomials.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 03/28/19 to 130.70.8.131. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Hereafter, we denote by $C(T,x)>0$ a generic function of nondecreasing in T and of at most polynomial growth order in x, which does not depend on the number of discretizations n.

3.2. "Naive" weak approximation with Malliavin weight under uniform partition. For preparation, we show a weak approximation of SDEs for bounded measurable test functions under uniform partition. First we give a small time approximation of $(P_t\varphi)(x)=\mathbb{E}[\varphi(X(t,x))]$, $t\in(0,1]$, with an error bound with $\|\varphi\|_\infty$.

THEOREM 3.6 (small time approximation I). *Assume* [**A1**] *and* [**A2**]. *For* $m\in\mathbb{N}$, *we have*

$$\mathbb{E}[\varphi(X(t,x))]=\mathbb{E}[\varphi(\overline{X}(t,x))\mathcal{M}_m(t,x,B_t)]+\mathcal{E}_1^{m+1,\varphi}(t,x) \tag{12}$$

for $\varphi\in C_b^\infty(\mathbb{R}^N)$, $t\in(0,1]$, *and* $x\in\mathbb{R}^N$, *where* $\mathcal{M}_m(t,x,B_t)$ *is a Malliavin weight given by*

$$\begin{aligned}\mathcal{M}_m(t,x,B_t) &:= \sum_{\substack{(j,k),\\ 1\le j\le 2m+1,\\ 1\le k\le j}}\sum_b\sum_I\sum_\alpha\sum_{\substack{\gamma,\\ \beta\sim_p\gamma}}\sum_J \mathbf{W}(t,x,b,I,\alpha,\gamma,\beta,J,B_t)\\ &:= 1+\sum_{j=1}^{2m+1}\sum_{k=1}^{j}\sum_{\substack{b=\{b_i\}_{i=1}^k \ s.t.\ \sum_{i=1}^k b_i=j+k,\\ b_i\ge 2,\ i=1,\cdots,k}}\sum_{I=(I_1,\cdots,I_k)\in\{1,\cdots,N\}^k}\frac{1}{k!}\sum_{\substack{\alpha^\nu\in\{0,1,\cdots,d\}^{r(\nu)}\ s.t.\ \|\alpha^\nu\|=b_\nu,\\ r(\nu)\in\mathbb{N},\ \nu=1,\cdots,k}}\\ &\quad\prod_{\nu=1}^k \hat{V}_{\alpha_1^\nu}\cdots\hat{V}_{\alpha_{r(\nu-1)}^\nu}V^{I_\nu}_{\alpha^\nu_{r(\nu)}}(x)\sum_{\substack{\sigma_l\in S_{m_l+n_l},\ l=1,\cdots,k-1,\ s.t.\\ m_l=\sum_{i=1}^l r(i),\ n_l=r(l+1),\\ \sigma_l(1)<\cdots<\sigma_l(m_l),\ \sigma_l(m_l+1)<\cdots<\sigma_l(m_l+n_l)}}\\ &\quad\sum_{\substack{\gamma,\beta\ s.t.\ \beta\sim_p\gamma,\ p\in\mathbb{N},\\ \beta=\sigma_{k-1}^{-1}\cdot(\cdots(\sigma_2^{-1}\cdot(\sigma_1^{-1}\cdot(\alpha^1*\alpha^2)*\alpha^3)*\cdots)*\alpha^k)}}\frac{1}{2^{\eta(\gamma,\beta)}}\sum_{J\in\{1,\cdots,N\}^{K(\gamma)}}\prod_{h=1}^{K(\gamma)}\\ &\qquad V^{J_h}_{\gamma_h^*}(x)\frac{1}{|\gamma|!}t^{|\gamma|}H_{(I_1,\cdots,I_k,J_1,\cdots,J_{K(\gamma)})}\Big(\overline{X}(t,x),1\Big),\end{aligned}$$

and $\mathcal{E}_1^{m+1,\varphi}(t,x)$ *satisfies* $|\mathcal{E}_1^{m+1,\varphi}(t,x)|\le Ct^{m+1}\|\varphi\|_\infty(1+|x|^Q)$ *for some constant* $C>0$ *and* $Q>0$ *independent to* φ, t *and* x.

Proof. See Appendix B. ∎

The weight $\mathcal{M}_m(t,x,B_t)$ is given by a linear combination of some polynomials of Brownian motions through the explicit computation of $H_{I*J}(\overline{X}(t,x),1)$. Note that the bound of $\mathcal{E}_1^\varphi(t,x)$ does not involve any derivatives of φ. Then, essentially, we do not need the regularity of φ. From the result, a theoretically tractable weak approximation holds.

For bounded measurable function φ, define operators $\{Q^m_{(t)}\}_{t>0}$, $m\in\mathbb{N}$ as $Q^m_{(t)}\varphi(x)=\mathbb{E}[\varphi(\overline{X}(t,x))\mathcal{M}_m(t,x,B_t)]$, $t>0$, $x\in\mathbb{R}^N$. We note that one has $|Q^m_{(t)}\varphi|\le\|\varphi\|_\infty\{1+O(\sqrt{t})\}$ since $\mathcal{M}_m(t,x,B_t)$ is of the form "1+polynomials of B_t."

Using Theorem 3.6 with operators $\{Q^m_{(t)}\}_{t>0}$, we have the following approximation with an easy proof.

THEOREM 3.7 (weak approximation I). *Under* [**A1**] *and* [**A2**], *it holds that*

$$\left|P_Tf(x)-(Q^m_{(T/n)})^nf(x)\right|\le C(T,x)\|f\|_\infty\frac{1}{n^m}$$

for all bounded measurable functions $f:\mathbb{R}^N\to\mathbb{R}$.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 03/28/19 to 130.70.8.131. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Proof. We easily see

$$P_T f(x) - (Q^m_{(T/n)})^n f(x) = \sum_{k=0}^{n-1} (Q^m_{(T/n)})^k (P_{T/n} - Q^m_{(T/n)}) P_{T-(k+1)T/n} f(x). \tag{13}$$

Since

$$\begin{aligned} &|(Q^m_{(T/n)})^k (P_{T/n} - Q^m_{(T/n)}) P_{T-(k+1)T/n} f(x)| \\ &\le C(T,x) \frac{1}{n^{m+1}} \|P_{T-(k+1)T/n} f\|_\infty \\ &\le C(T,x) \frac{1}{n^{m+1}} \|f\|_\infty \end{aligned}$$

by Theorem 3.6, we have

$$|P_T f(x) - (Q^m_{(T/n)})^n f(x)| \le C(T,x)\|f\|_\infty \sum_{k=0}^{n-1} \frac{1}{n^{m+1}} \le C(T,x)\|f\|_\infty \frac{1}{n^m}.$$

∎

The weak approximation (Theorem 3.7) is mathematically accessible once we obtain a small time approximation of $(P_t\varphi)(x)$ with an error bound of type $C\|\varphi\|_\infty t^{m+1}$ without derivatives of φ. However, from the point of view of numerical computation, it has a serious problem. Although we have an explicit representation of $\mathcal{M}_m(t,x,B_t)$ in Theorem 3.7, it is quite difficult to implement the scheme when we want to get the weak approximation of order $m \ge 2$ for multidimensional SDEs since much higher order polynomials of Brownian motions are required. Indeed, when $m = 2$, the maximum order of polynomial of Brownian motions in $\mathcal{M}_2(t,x,B_t)$ is 15, i.e., the weight $\mathcal{M}_2(t,x,B_t)$ involves $\prod_{k=1}^{15} B_t^{i_k}$, $(i_1,\cdots,i_{15}) \in \{1,\cdots,d\}^{15}$, which comes from the term $\mathbb{E}\big[\partial_{I_1}\cdots\partial_{I_5}\varphi(\overline{X}(t,x)) \prod_{\nu=1}^{5} \hat{V}_{\alpha_1^\nu} V^{I_\nu}_{\alpha_2^\nu}(x) \mathbb{B}^{\mathrm{Strat}}_{\alpha^\nu,t}\big]$. Also, function evaluations of $\hat{V}_{i_1}\cdots\hat{V}_{i_4}V_{i_5}(x)$, $(i_1,\cdots,i_5) \in \{1,\cdots,d\}^5$, are required since we need to compute the term $\mathbb{E}\big[\partial_{I_1}\varphi(\overline{X}(t,x))\hat{V}_{i_1}\cdots\hat{V}_{i_4}V^{I_1}_{i_5}(x)\mathbb{B}^{\mathrm{Strat}}_{(i_1,\cdots,i_5),t}\big]$. Thus it is hard to implement the current method when the dimension and the order of the weak approximation are high, i.e., $d \ge 2$, $N \ge 2$, and $m \ge 2$.

However, we can attain a higher order discretization even if we only use some lower order polynomials of Brownian motions. We justify this phenomenon by using a technical tool of Malliavin calculus and then see the validity through the numerical examples.

3.3. Main result: Weak approximation with universal weight. In this section, we show a new scheme as an improvement of Theorem 3.7. We still aim at obtaining higher order weak approximation for nonsmooth test functions under uniform partition, but we will use a simplified scheme. Theorem 3.13 is our main result of this paper. First, we give a modified small time approximation of $P_t\varphi(x)$ by reducing some terms in the Malliavin weight of Theorems 3.6 and 3.7. The following estimate will be a key result for the construction.

LEMMA 3.8. *Assume* **[A1]** *and* **[A2]**. *Let $m \in \mathbb{N}$ and $\gamma \in \{0,1,\cdots,d\}^{m+1}$ be a multi-index. Then, there are $C > 0$ and $Q > 0$ such that*

$$\left|\mathbb{E}\Big[\partial_{I_1}\cdots\partial_{I_k}\varphi(\overline{X}(t,x))\mathbb{B}^{\mathrm{It\hat{o}}}_{\gamma,t}\Big]\right| \le Ct^{m+1}\|\nabla^{k+K(\gamma)}\varphi\|_\infty(1+|x|^Q) \tag{14}$$

for all $\varphi \in C_b^\infty(\mathbb{R}^N)$, $t \in (0,1]$, $x \in \mathbb{R}^N$, and $I_l \in \{1,\cdots,N\}$, $l \le k \in \mathbb{N}$.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Proof. By Lemma 3.5, we have "reverse" Malliavin integration by parts:

$$\mathbb{E}\Big[\partial_{I_1}\cdots\partial_{I_k}\varphi(\overline{X}(t,x))\mathbb{B}^{\mathrm{It\hat{o}}}_{\gamma,t}\Big] = \sum_{J\in\{1,\cdots,N\}^{K(\gamma)}}\prod_{h=1}^{K(\gamma)} V^{J_h}_{\gamma^*_h}(x)\frac{1}{|\gamma|!}t^{|\gamma|}\mathbb{E}\Big[\partial_{I_1}\cdots\partial_{I_k}\partial_{J_1}\cdots\partial_{J_{K(\gamma)}}\varphi(\overline{X}(t,x))\Big]. \tag{15}$$

Then we obtain the following estimate:

$$\Big|\mathbb{E}\Big[\partial_{I_1}\cdots\partial_{I_k}\varphi(\overline{X}(t,x))\mathbb{B}^{\mathrm{It\hat{o}}}_{\gamma,t}\Big]\Big| \le Ct^{m+1}\|\nabla^{k+K(\gamma)}\varphi\|_\infty(1+|x|^Q).$$

∎

In principle, one has $\mathbb{E}[\partial_{I_1}\cdots\partial_{I_k}\varphi(\overline{X}(t,x))\mathbb{B}^{\mathrm{It\hat{o}}}_{\gamma,t}] = O(t^{\|\gamma\|/2})$ since the standard estimate gives that for $p\ge 1$, $\|\mathbb{B}^{\mathrm{It\hat{o}}}_{\gamma,t}\|_p = O(t^{\|\gamma\|/2})$. For example, if $\|\gamma\| = |\gamma| = m+1$, i.e., $K(\gamma) = m+1$, we only have $\mathbb{E}[\partial_{I_1}\cdots\partial_{I_k}\varphi(\overline{X}(t,x))\mathbb{B}^{\mathrm{It\hat{o}}}_{\gamma,t}] = O(t^{(m+1)/2})$. However, Lemma 3.8 suggests $\mathbb{E}[\partial_{I_1}\cdots\partial_{I_k}\varphi(\overline{X}(t,x))\mathbb{B}^{\mathrm{It\hat{o}}}_{\gamma,t}] = O(t^{m+1})$ through the "reverse" Malliavin integration by parts (15). From Lemma 3.8, the following rule holds under smoothness of φ: *The expectation*

$$\mathbb{E}[\partial_{I_1}\cdots\partial_{I_k}\varphi(\overline{X}(t,x))(\textit{iterated Itô integral of length } |\gamma|\ge m+1)]$$

will be $O(t^{m+1})$ (*not only* $O(t^{(m+1)/2})$). Therefore, if iterated Itô integrals of length $|\gamma|\ge m+1$ appear in the construction of the small time approximation of Theorem 3.6, those can be treated as the residuals.

Hence we obtain the following tractable small time approximation by truncating some redundant terms of order $O(t^{m+1})$.

THEOREM 3.9 (small time approximation II). *Assume* **[A1]** *and* **[A2]**. *For* $m\in\mathbb{N}$, *we have*

$$\mathbb{E}[\varphi(X(t,x))] = \mathbb{E}[\varphi(\overline{X}(t,x))\widetilde{\mathcal{M}_m}(t,x,B_t)] + \mathcal{E}_1^{m+1,\varphi}(t,x) + \mathcal{E}_2^{m+1,\varphi}(t,x) \tag{16}$$

for $\varphi\in C_b^\infty(\mathbb{R}^N)$, $t\in(0,1]$, *and* $x\in\mathbb{R}^N$, *where* $\widetilde{\mathcal{M}_m}(t,x,B_t)$ *is a Malliavin weight given by*

$$\begin{aligned}
\widetilde{\mathcal{M}_m}(t,x,B_t) &:= \sum_{\substack{(j,k),\\ 1\le j\le 2m,\\ 1\le k\le j}}\sum_b\sum_I\sum_\alpha\sum_{\substack{\gamma,\beta,\\ \beta\sim_p\gamma,\\ |\gamma|\le m}}\sum_J \mathbf{W}(t,x,b,I,\alpha,\gamma,\beta,J,B_t)\\
&:= 1+\sum_{j=1}^{2m}\sum_{k=1}^{j}\sum_{\substack{b=\{b_i\}_{i=1}^k,\sum_{i=1}^k b_i=j+k,\\ b_i\ge 2,\ i=1,\cdots,k}}\sum_{I=(I_1,\cdots,I_k)\in\{1,\cdots,N\}^k}\frac{1}{k!}\sum_{\substack{\alpha^\nu\in\{0,1,\cdots,d\}^{r(\nu)}\ s.t.\ \|\alpha^\nu\|=b_\nu,\\ r(\nu)\in\mathbb{N},\ \nu=1,\cdots,k}}\\
&\quad\prod_{\nu=1}^k \hat{V}_{\alpha_1^\nu}\cdots\hat{V}_{\alpha^\nu_{r(\nu-1)}}V^{I_\nu}_{\alpha^\nu_{r(\nu)}}(x)\sum_{\substack{\sigma_l\in S_{m_l+n_l},\ l=1,\cdots,k-1,\ s.t.\\ m_l=\sum_{i=1}^l r(i),\ n_l=r(l+1),\\ \sigma_l(1)<\cdots<\sigma_l(m_l),\ \sigma_l(m_l+1)<\cdots<\sigma_l(m_l+n_l)}}\\
&\quad\sum_{\substack{\gamma,\beta\ s.t.\ \beta\sim_p\gamma,\ p\in\mathbb{N},\\ \beta=\sigma_{k-1}^{-1}\cdot(\cdots(\sigma_2^{-1}\cdot(\sigma_1^{-1}\cdot(\alpha^1*\alpha^2)*\alpha^3)*\cdots)*\alpha^k),\\ |\gamma|\le m}}\frac{1}{2^{\eta(\gamma,\beta)}}\sum_{J\in\{1,\cdots,N\}^{K(\gamma)}}\prod_{h=1}^{K(\gamma)}\\
&\qquad V^{J_h}_{\gamma^*_h}(x)\frac{1}{|\gamma|!}t^{|\gamma|}H_{(I_1,\cdots,I_k,J_1,\cdots,J_{K(\gamma)})}\Big(\overline{X}(t,x),1\Big),
\end{aligned}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 03/28/19 to 130.70.8.131. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Downloaded 03/28/19 to 130.70.8.131. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

and $\mathcal{E}_1^{m+1,\varphi}(t,x)$ *is same as in Theorem* 3.6 *(i.e.,* $|\mathcal{E}_1^{m+1,\varphi}(t,x)| \le C_1 t^{m+1}\|\varphi\|_\infty(1+|x|^{Q_1})$ *for some constants* $C_1>0$ *and* $Q_1>0$*), and further* $\mathcal{E}_2^{m+1,\varphi}(t,x)$ *is given by*

$$\mathcal{E}_2^{m+1,\varphi}(t,x) = \mathbb{E}[\varphi(\overline{X}(t,x))(\mathcal{M}_m(t,x,B_t) - \widetilde{\mathcal{M}_m}(t,x,B_t))] \tag{17}$$

$$= t^{m+1}\sum_{i=1}^{e(m)} \mathbb{E}[\partial^{\alpha^i}\varphi(\overline{X}(t,x))]K_{\alpha^i}(t,x) \tag{18}$$

for an integer $e(m)\in\mathbb{N}$*, multi-indices* α^i *with* $|\alpha^i|\le 2m+1$*,* $i=1,\cdots,e(m)$*, and functions* $K_{\alpha^i}:[0,T]\times\mathbb{R}^N\to\mathbb{R}$*, with the following estimate:*

$$|\mathcal{E}_2^{m+1,\varphi}(t,x)| \le C_2 t^{m+1}\sum_{i=1}^{e(m)}\|\nabla^{|\alpha_i|}\varphi\|_\infty(1+|x|^{Q_2}) \tag{19}$$

for some $C_2>0$ *and* $Q_2>0$ *independent of* φ*,* t*, and* x*.*

Proof. See Appendix C. ∎

Remark 2 (comparison of the weights in Theorem 3.6 and 3.9). In the weight $\widetilde{\mathcal{M}_m}(t,x,B_t)$ above, the terms in $\mathcal{M}_m(t,x,B_t)$ of Theorem 3.6 (and Theorem 3.7) are substantially reduced. The factors $\frac{1}{|\gamma|!}t^{|\gamma|}H_{(I_1,\cdots,I_k,J_1,\cdots,J_{K(\gamma)})}(\overline{X}(t,x),1)$ satisfying $|\gamma|\ge m+1$ are all removed, that is, a lot of higher order Brownian polynomials are not necessary in the computation of the new weight $\widetilde{\mathcal{M}_m}(t,x,B_t)$. According to this, many costly terms involving higher order derivatives in $\hat{V}_{\alpha_1}\cdots\hat{V}_{\alpha_{r-1}}V_{\alpha_r}(x)$, $\alpha\in\{0,1,\cdots,d\}^r$, are also not required.

In the following, we write $\widehat{\mathcal{M}_m}(t,x,B_t)$ for $\mathcal{M}_m(t,x,B_t)-\widetilde{\mathcal{M}_m}(t,x,B_t)$ in (17).

A new weak approximation in Theorem 3.13 will be constructed using the reduced weight in Theorem 3.9. Although the small time approximation above highly depends on the smoothness of test function φ, we can still show an arbitrary order weak approximation for $P_Tf(x)$ under uniform partition without using the derivatives of the test function $f:\mathbb{R}^N\to\mathbb{R}$.

Let $\{\widetilde{Q^m_{(t)}}\}_{t>0}$, $m\in\mathbb{N}$, be operators given by

$$\widetilde{Q^m_{(t)}}\varphi(x) = \mathbb{E}[\varphi(\overline{X}(t,x))\widetilde{\mathcal{M}_m}(t,x,B_t)], \quad t>0,\ x\in\mathbb{R}^N,$$

for a function $\varphi:\mathbb{R}^N\to\mathbb{R}$ appropriately chosen. Then (16) in Theorem 3.9 is represented by

$$P_t\varphi(x) - \widetilde{Q^m_{(t)}}\varphi(x) = \mathcal{E}_1^{m+1,\varphi}(t,x) + \mathcal{E}_2^{m+1,\varphi}(t,x), \quad \varphi\in C_b^\infty(\mathbb{R}^N). \tag{20}$$

Let $f:\mathbb{R}^N\to\mathbb{R}$ be a bounded measurable function. Our purpose here is to show the order of $P_Tf-(\widetilde{Q^m_{(T/n)}})^n f$ using the bound $\|f\|_\infty$. By an argument similar to that in (13), we have

$$P_Tf(x) - (\widetilde{Q^m_{(T/n)}})^n f(x) = \sum_{k=0}^{n-1}(\widetilde{Q^m_{(T/n)}})^k(P_{T/n}-\widetilde{Q^m_{(T/n)}})P_{T-(k+1)T/n}f(x). \tag{21}$$

We analyze each term of the right-hand side of (21). Let $k=0,1,\cdots,n-1$; by (20) we have

$$\begin{aligned}&(\widetilde{Q^m_{(T/n)}})^k(P_{T/n}-\widetilde{Q^m_{(T/n)}})P_{T-(k+1)T/n}f(x)\\ &= (\widetilde{Q^m_{(T/n)}})^k\mathcal{E}_1^{m+1,P_{T-(k+1)T/n}f}(T/n,x) + (\widetilde{Q^m_{(T/n)}})^k\mathcal{E}_2^{m+1,P_{T-(k+1)T/n}f}(T/n,x).\end{aligned} \tag{22}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

We will prove $|(\widetilde{Q^m_{(T/n)}})^k(P_{T/n}-\widetilde{Q^m_{(T/n)}})P_{T-(k+1)T/n}f(x)|\le C(T,x)\|f\|_\infty n^{-(m+1)}$ to obtain the desired weak approximation

$$\left|P_Tf(x)-(\widetilde{Q^m_{(T/n)}})^nf(x)\right|\le C(T,x)\|f\|_\infty n^{-m} \tag{23}$$

with new operators $\{\widetilde{Q^m_{(t)}}\}_t$. For the first term of the right-hand side of (22), we immediately see the following.

LEMMA 3.10. *Under* [**A1**] *and* [**A2**], *we have, for* $k=0,1,\cdots,n-1$,

$$|(\widetilde{Q^m_{(T/n)}})^k\mathcal{E}_1^{m+1,P_{T-(k+1)T/n}f}(T/n,x)|\le C(T,x)\|f\|_\infty\frac{1}{n^{m+1}}.$$

Proof. As in the proof in Theorem 3.7 we have

$$|(\widetilde{Q^m_{(T/n)}})^k\mathcal{E}_1^{m+1,P_{T-(k+1)T/n}f}(T/n,x)|\le C(T,x)\|P_{T-(k+1)T/n}f\|_\infty n^{-(m+1)}\le C(T,x)\|f\|_\infty n^{-(m+1)}$$

by Theorem 3.9 (essentially by Theorem 3.7). Then we get the result. ∎

Therefore, in order to obtain the weak approximation (23), it suffices to show

$$\left|(\widetilde{Q^m_{(T/n)}})^k\mathcal{E}_2^{m+1,P_{T-(k+1)T/n}f}(T/n,x)\right|\le C(T,x)\|f\|_\infty\frac{1}{n^{m+1}}.$$

To see this, we will take several steps. Let $\{\overline{X}_{(T/n)}(t,x)\}_{t\ge0}$, $x\in\mathbb{R}^N$ be the process given by

$$\overline{X}_{(T/n)}(0,x)=x,$$
$$\overline{X}_{(T/n)}(t,x)=\overline{X}_{(T/n)}((k-1)T/n,x)+\sum_{i=1}^d V_i(\overline{X}_{(T/n)}((k-1)T/n,x))(B^i_t-B^i_{(k-1)T/n})$$

for $t\in((k-1)T/n,kT/n]$, $k=1,\cdots,n$. It can be written in the Itô process form as

$$\overline{X}_{(T/n)}(t,x)=x+\sum_{i=1}^d\int_0^t V_i(\overline{X}_{(T/n)}(\phi(s),x))dB^i_s,\quad 0\le t\le T, \tag{24}$$

where $\phi(s)=\sup\{kT/n;kT/n\le s\}$. In particular, (24) is an elliptic Itô process under [**A2**] (and [**A1**]). Using (18), for $k<n-1$, we have

$$\begin{aligned}(25)\quad &(\widetilde{Q^m_{(T/n)}})^k\mathcal{E}_2^{m+1,P_{T-(k+1)T/n}f}(T/n,x)\\ &=\mathbb{E}\Big[\mathcal{E}_2^{m+1,P_{T-(k+1)T/n}f}(T/n,\overline{X}_{(T/n)}(kT/n,x))\\ &\qquad\qquad\prod_{j=1}^k\widetilde{\mathcal{M}_m}(T/n,\overline{X}_{(T/n)}((j-1)T/n,x),B_{jT/n}-B_{(j-1)T/n})\Big]\\ &=\left(\frac{T}{n}\right)^{m+1}\sum_{i=1}^{e(m)}\mathbb{E}\Big[\partial^{\alpha^i}P_{T-(k+1)T/n}f\left(\overline{X}_{(T/n)}((k+1)T/n,x)\right)\Psi^{k,m}_{\alpha^i}\Big],\end{aligned}$$

where $\Psi^{k,m}_{\alpha^i}$ is a Wiener functional given by

$$\Psi^{k,m}_{\alpha^i}=K_{\alpha^i}(T/n,\overline{X}_{(T/n)}(kT/n,x))\prod_{j=1}^k\widetilde{\mathcal{M}_m}(T/n,\overline{X}_{(T/n)}((j-1)T/n,x),B_{jT/n}-B_{(j-1)T/n}).$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 03/28/19 to 130.70.8.131. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Downloaded 03/28/19 to 130.70.8.131. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

We note that it holds that for all $l \in \mathbb{N}$, $p > 1$, $\|\Psi_{\alpha^i}^{k,m}\|_{l,p} \leq C(T,x)\prod_{i=1}^{k}\{1 + O(\sqrt{T/n})\} = O(1)$ since the weight $\widetilde{\mathcal{M}_m}(t,x,B_t)$ (Theorem 3.9) is constructed by "1+polynomials of Brownian motions."

We will show the upper bound of the error term (25) without using the derivatives of f. For $t \in (0,T)$, if we directly estimate $(\widetilde{Q^m_{(T/n)}})^k \mathcal{E}_2^{m+1,P_{T-t}f}(T/n,x)$ using Hölder's inequality, the bounds of $|\partial^\alpha P_{T-t}f(\cdot)|$ and $\|\Psi_\alpha^{k,m}\|_p$ for multi-index α are required. Here, for multi-index α, the bound of $|\partial^\alpha P_{T-t}f(\cdot)|$ is given as

$$|\partial^\alpha P_{T-t}f(x)| \leq C(T,x)\|f\|_\infty \frac{1}{(T-t)^{|\alpha|/2}}, \quad t \in (0,T), \tag{26}$$

under ellipticity; see [17], for example. Then, when $T-t$ is large, i.e., the case $t \in (0,T/2)$ is small, we immediately have $\left|(\widetilde{Q^m_{(T/n)}})^k \mathcal{E}_2^{m+1,P_{T-t}f}(T/n,x)\right| \leq C(T,x)\, n^{-(m+1)}$. A problem occurs when $T-t$ is small, i.e., $t \in [T/2,T)$ is large. In this case, the bound (26) can be huge. Thus, the standard estimate of $(\widetilde{Q^m_{(T/n)}})^k \mathcal{E}_2^{m+1,P_{T-t}f}(T/n,x)$ cannot give what we are looking for.

However, if we apply the Malliavin integration by parts to remove the derivatives of $P_{T-t}f$, we can have a sharp estimate. To give the upper bound of (25), we use the integration by parts of Kusuoka and Stroock [17] for the elliptic Itô process $\{\overline{X}_{(T/n)}(t,x)\}_t$.

LEMMA 3.11. *Let α be a multi-index and $G \in \mathbb{D}^\infty$ such that for all $l \in \mathbb{N}$, $p > 1$, $\|G\|_{l,p} \leq C(T,x)$. Under* **[A1]** *and* **[A2]**, *we have*

$$\begin{aligned}&\left|\mathbb{E}[\partial^\alpha P_{T-t}f(\overline{X}_{(T/n)}(t,x))G]\right| \\ &= \left|\mathbb{E}[P_{T-t}f(\overline{X}_{(T/n)}(t,x))H_\alpha(\overline{X}_{(T/n)}(t,x),G)]\right| \leq C(T,x)\|f\|_\infty\end{aligned}$$

for all $t \in [T/2,T)$.

Proof. See Appendix D. ∎

The integration by parts in Lemma 3.11 holds for $S \in \mathcal{S}'(\mathbb{R}^N)$ and $t \leq T$ as

$$\langle \partial^\alpha S(\overline{X}_{(T/n)}(t,x)), G\rangle = \langle S(\overline{X}_{(T/n)}(t,x)), H_\alpha(\overline{X}_{(T/n)}(t,x),G)\rangle.$$

For $k \geq 0$ such that $(k+1)T/n \in [T/2,T]$, we have the representation

$$\begin{aligned}&(\widetilde{Q^m_{(T/n)}})^k \mathcal{E}_2^{m+1,P_{T-(k+1)T/n}f}(T/n,x) \\ &= \left(\frac{T}{n}\right)^{m+1}\sum_{i=1}^{e(m)} \mathbb{E}\Big[P_{T-(k+1)T/n}f\left(\overline{X}_{(T/n)}((k+1)T/n,x)\right) H_{\alpha^i}(\overline{X}_{(T/n)}((k+1)T/n,x),\Psi_{\alpha^i}^{k,m})\Big].\end{aligned}$$

We then attain the following.

LEMMA 3.12. *Under* **[A1]** *and* **[A2]**, *we have, for $k = 0,1,\cdots,n-1$,*

$$\left|(\widetilde{Q^m_{(T/n)}})^k \mathcal{E}_2^{m+1,P_{T-(k+1)T/n}f}(T/n,x)\right| \leq C(T,x)\|f\|_\infty \frac{1}{n^{m+1}}.$$

Proof. For $k \geq 0$ such that $(k+1)T/n \in [T/2,T]$, it holds that

$$\left|(\widetilde{Q^m_{(T/n)}})^k \mathcal{E}_2^{m+1,P_{T-(k+1)T/n}f}(T/n,x)\right| \leq C(T,x)\|f\|_\infty \frac{1}{n^{m+1}}.$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Further, we have for $k \geq 0$ such that $(k+1)T/n \in (0, T/2)$,

$$\left|(\widetilde{Q^m_{(T/n)}})^k \mathcal{E}_2^{m+1, P_{T-(k+1)T/n} f}(T/n, x)\right| \leq C(T,x)\|f\|_\infty \frac{1}{n^{m+1}} \sum_{i=1}^{e(m)} \frac{1}{T^{|\alpha^i|/2}},$$

by combining the gradient bound (26) and $T/2 \leq T - (k+1)T/n$. Since $T \geq 1$, we obtain the assertion. ∎

We now state the main result of this paper. Using operators $\{\widetilde{Q^m_{(s)}}\}_s$, we have the following arbitrary order weak approximation under uniform partition.

THEOREM 3.13 (weak approximation II). *Under* **[A1]** *and* **[A2]**, *it holds that*

$$\left|P_T f(x) - (\widetilde{Q^m_{(T/n)}})^n f(x)\right| \leq C(T,x)\|f\|_\infty \frac{1}{n^m}$$

for all bounded measurable functions $f : \mathbb{R}^N \to \mathbb{R}$. *In particular, we have*

$$\sup_{z \in \mathbb{R}^N} \Big| P\big(X(T,x) \leq z\big) - \mathbb{E}\Big[\mathbf{1}_{\{\overline{X}_{(T/n)}(T,x) \leq z\}} \prod_{j=1}^n \widetilde{\mathcal{M}_m}(T/n, \overline{X}_{(T/n)}((j-1)T/n, x), B_{jT/n} - B_{(j-1)T/n})\Big]\Big| \leq C(T,x)\frac{1}{n^m}.$$

Proof. Applying Lemmas 3.10 and 3.12 to

$$P_T f(x) - (\widetilde{Q^m_{(T/n)}})^n f(x) = \sum_{k=0}^{n-1} (\widetilde{Q^m_{(T/n)}})^k (P_{T/n} - \widetilde{Q^m_{(T/n)}}) P_{T-(k+1)T/n} f(x)$$

$$= \sum_{k=0}^{n-1} \left[(\widetilde{Q^m_{(T/n)}})^k \mathcal{E}_1^{m+1, P_{T-(k+1)T/n} f}(T/n, x) + (\widetilde{Q^m_{(T/n)}})^k \mathcal{E}_2^{m+1, P_{T-(k+1)T/n} f}(T/n, x)\right],$$

we obtain $|P_T f(x) - (\widetilde{Q^m_{(T/n)}})^n f(x)| \leq C(T,x)\|f\|_\infty \sum_{k=0}^{n-1} \frac{1}{n^{m+1}} = C(T,x)\|f\|_\infty \frac{1}{n^m}$. By the definition of $\{\widetilde{Q^m_{(t)}}\}_t$, we can see

$$(\widetilde{Q^m_{(T/n)}})^n f(x) = \mathbb{E}\Big[f(\overline{X}_{(T/n)}(T,x)) \prod_{j=1}^n \widetilde{\mathcal{M}_m}(T/n, \overline{X}_{(T/n)}((j-1)T/n, x), B_{jT/n} - B_{(j-1)T/n})\Big].$$

When we take $f_z(\cdot) = \mathbf{1}_{\{\cdot \leq z\}}$, we have the uniform bound with respect to z since $\|f_z\|_\infty \equiv 1$. ∎

Remark 3 (comparison of the schemes of Theorems 3.7 and 3.13). As mentioned in Remark 2, each one-step weight $\widetilde{\mathcal{M}_m}(t, x, B_t)$ in Theorem 3.13 is much simplified compared to that in Theorem 3.7. Then the product of the one-step weights in Theorem 3.13 can be easily implemented.

The weak approximation works under uniform partition whether the test function is smooth or not. Theorem 3.13 obviously holds if $f \in C_b^\infty(\mathbb{R}^N)$ and also holds if f is a Lipschitz continuous function when we replace $\|f\|_\infty$ with the Lipschitz constant $C_{Lip}[f]$ in Lemmas 3.10, 3.11, and 3.12 and Theorem 3.13. The weight is universal in the sense that the weak approximation does not depend on the smoothness of f. We call it *universal weight*.

As an example of the scheme, we show a tractable second order scheme ($m = 2$). Let $\tilde{V}_0$ be the differentiation operator $\tilde{V}_0\varphi(x) = \hat{V}_0\varphi(x) + 1/2\sum_{i=1}^d \hat{V}_i V_i \varphi(x)$, $\varphi \in C_p^\infty(\mathbb{R}^N)$, and $\bar{V}_0^j \in C_p^\infty(\mathbb{R}^N)$ be a function $\bar{V}_0^j(x) = V_0^j(x) + 1/2\sum_{i=1}^d \hat{V}_i V_i^j(x)$.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 03/28/19 to 130.70.8.131. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Downloaded 03/28/19 to 130.70.8.131. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

COROLLARY 3.14 (second order method with universal weight). *Under* **[A1]** *and* **[A2]**, *we have the following second order discretization with uniform partition:*

$$\Big|\mathbb{E}[f(X(T,x))] - \mathbb{E}\Big[f(\overline{X}_{(T/n)}(T,x))\prod_{j=1}^{n}\widetilde{\mathcal{M}_2}(T/n,\overline{X}_{(T/n)}((j-1)T/n,x),B_{jT/n}-B_{(j-1)T/n})\Big]\Big| = O\left(\frac{1}{n^2}\right)$$

for all bounded measurable functions $f:\mathbb{R}^N\to\mathbb{R}$, *where* $\widetilde{\mathcal{M}_2}(t,x,B_t)$ *is the universal weight given by*

$$\widetilde{\mathcal{M}_2}(t,x,B_t) = 1 + \sum_{r=1,2,3}\ \sum_{1\le k_1,\cdots,k_r\le N} c_{k_1,\cdots,k_r}(t,x)H_{(k_1,\cdots,k_r)}(\overline{X}(t,x),1),$$

with coefficients

$$\begin{aligned}
c_{k_1}(t,x) &= \bar{V}_0^{k_1}(x)t + \tilde{V}_0\bar{V}_0^{k_1}(x)\frac{1}{2}t^2,\\
c_{k_1,k_2}(t,x) &= \bar{V}_0^{k_1}(x)\bar{V}_0^{k_2}(x)\frac{1}{2}t^2 + \sum_{1\le i_1\le d}\{\tilde{V}_0V_{i_1}^{k_1}(x)+\hat{V}_{i_1}\bar{V}_0^{k_1}(x)\}V_{i_1}^{k_2}(x)\frac{1}{2}t^2\\
&\quad + \sum_{1\le i_1,i_2\le d}\hat{V}_{i_1}V_{i_2}^{k_1}(x)\hat{V}_{i_1}V_{i_2}^{k_2}(x)\frac{1}{4}t^2,\\
c_{k_1,k_2,k_3}(t,x) &= \sum_{1\le i_1,i_2\le d}\hat{V}_{i_1}V_{i_2}^{k_1}(x)V_{i_1}^{k_2}(x)V_{i_2}^{k_3}(x)\frac{1}{2}t^2.
\end{aligned}$$

Proof. See Appendix E. ∎

Remark 4. Similarly, the third order discretization ($m=3$) also holds as in Corollary 3.14. We will see the numerical accuracy in section 4.2.4.

We summarize the features of the proposed method (Theorem 3.13) and comment on the computational effort by comparing it with the classical methods: the Euler–Maruyama (first order) scheme of Bally and Talay [2] and Talay's second order scheme of [31].

To compare the schemes, let us consider the following Itô SDE:

$$dX(t,x) = A_0(X(t,x))dt + \sum_{i=1}^{d}A_i(X(t,x))dB_t^i,\quad X(0,x)=x. \tag{27}$$

Bally and Talay [2] proved the first order result $|E[f(X(T,x))]-E[f(\bar{X}^{\mathrm{EM}}_{(T/n)}(T,x))]| = O(n^{-1})$ for a bounded measurable function f, where $\bar{X}^{\mathrm{EM}}_{(T/n)}(T,x)$ is the n-step Euler–Maruyama scheme. Talay [31] showed $|E[f(X(T,x))]-E[f(\bar{X}^{\mathrm{Talay}}_{(T/n)}(T,x))]| = O(n^{-2})$ using a scheme $\bar{X}^{\mathrm{Talay}}_{(T/n)}(T,x)$ with some discrete random variables. In Talay's scheme, the test function f is assumed to be six order continuous differentiable and has polynomial growth order derivatives. The iterated Itô integrals up to second order ($A_i^j(x)B_t$ and $\hat{A}_{i_1}A_{i_2}^j(x)\mathbb{B}^{\mathrm{Itô}}_{(i_1,i_2),t}$) are used in the construction, where $\hat{A}_0 = \sum_{j=1}^N A_0^j(x)\frac{\partial}{\partial x_j} + 1/2\sum_{i,j=1}^N\sum_{k=1}^d A_k^i(x)A_k^j(x)\frac{\partial^2}{\partial x_i\partial x_j}$, $\hat{A}_i = \sum_{j=1}^N A_i^j(x)\frac{\partial}{\partial x_j}$, $i=1,\cdots,d$. To simulate the terms $\hat{A}_{i_1}A_{i_2}^j(x)\mathbb{B}^{\mathrm{Itô}}_{(i_1,i_2),t}$, a set of $d(d+1)/2$th random variables are used in the implementation, that is, the number of random variables increase as $O(d^2)$ in Talay's scheme.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 03/28/19 to 130.70.8.131. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

In this paper, we have proposed the arbitrary order weak scheme which is constructed by some polynomials of Brownian motion only. So, the number of required random variables is always d (see section 4.1). Further, the scheme can be applied to measurable test functions since the smoothness assumption is removed by Malliavin calculus. For the second order scheme (Corollary 3.14), the required evaluating functions are $A_0^j(\cdot)$, $A_i^j(\cdot)$, $\hat{A}_{i_1}A_{i_2}^j(\cdot)$, which are the same as in Talay's scheme. Then the numbers of evaluating functions must be N, Nd, N^2d, and N^3d since we have to compute derivatives $\partial_{j_1}A_{i_2}^{j_2}(\cdot)$ (if $i_1 \neq 0$), $\partial_{j_1}\partial_{j_2}A_{i_2}^{j_3}(\cdot)$ (if $i_1 = 0$) for $1 \leq j_1, j_2, j_3 \leq N$, and $1 \leq i_2 \leq d$. In addition, our method needs to specify an $N \times N$-inverse matrix $\Lambda^{-1}(\cdot)$ of $\Lambda_{ij}(\cdot) = \sum_{k=1}^d A_k^i(\cdot)A_k^j(\cdot)(= \sum_{k=1}^d V_k^i(\cdot)V_k^j(\cdot))$, $1 \leq i, j \leq N$, in the implementation, because the method relies on the Malliavin integration by parts under the elliptic condition. This may be a cost of not using additional random variables for the Lévy area term and a cost of treating nonsmooth test function f.

We note that a kind of elliptic condition and a smoothness of coefficients of SDEs are usually required to get a weak/strong approximation for irregular functionals. For the Euler–Maruyama scheme of Bally and Talay [2], the Hörmander hypoelliptic condition and the bounded smoothness condition for coefficients of SDEs are imposed (but $A_i(\cdot)$, $i = 0, 1, \cdots, d$, are not supposed to be bounded themselves). Also, Avikainen [1] (strong approximation for irregular functionals) and Yamada [33] (higher order weak scheme for irregular functionals) assumed the uniform elliptic condition and the bounded smoothness condition.

In this paper, we assume the conditions **[A1]** and **[A2]** for the arbitrary order weak approximation. We still need the uniform elliptic condition **[A2]**, which is stronger than that in [2], but the bounded smoothness condition is weakened in **[A1]**, which means that $A_i(\cdot)$, $i = 0, 1, \cdots, d$, themselves and their derivatives are assumed to be at most polynomial growth.

The methodologies are summarized in Table 1.

TABLE 1
Comparison of features of weak approximations and computational efforts in one-step simulation.

Method	*Bally and Talay* [2]	*Talay* [31]	*Theorem* 3.13 ($m = 2$)
Weak order	$O(n^{-1})$	$O(n^{-2})$	$O(n^{-2})$
Smoothness of coefficient of SDE	Smooth	Smooth	Smooth
Smoothness of test function	Nonsmooth (or smooth)	Smooth	Nonsmooth (or smooth)
Ellipticity	Hypoelliptic	–	Elliptic
Time partition	Uniform partition	Uniform partition	Uniform partition
Required number of random numbers	d	$d(d+1)/2$	d
Evaluations of required functions	$A_i(x)$	$A_i(x)$, $\hat{A}_{i_1}A_{i_2}(x)$	$A_i(x), \hat{A}_{i_1}A_{i_2}(x), \Lambda^{-1}(x)$
Order of computational cost	$O(Nd)$	$O(N^3d)$	$O(N^3d)$

Note that the number of required random numbers is still d in the arbitrary order scheme even if the target order m increases ($m \geq 3$), while evaluations of required functions such as $\hat{A}_{i_1}\cdots\hat{A}_{i_{k-1}}A_{i_k}^j(\cdot)$ increase at most $k \leq m$. Then the properties of the arbitrary order scheme are summarized in Table 2.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 03/28/19 to 130.70.8.131. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

TABLE 2
Features of the arbitrary order weak approximation of Theorem 3.13 *and computational efforts in one-step simulation.*

Method	*Theorem* 3.13 ($m \geq 2$)
Weak order	$O(n^{-m})$
Smoothness of coefficient of SDE	Smooth
Smoothness of test function	Nonsmooth (or smooth)
Ellipticity	Elliptic
Time partition	Uniform partition
Required number of random numbers	d
Evaluations of required functions	$A_i(x), \hat{A}_{i_1}A_{i_2}(x), \cdots, \hat{A}_{i_1}\cdots\hat{A}_{i_{m-1}}A_{i_m}(x), \Lambda^{-1}(x)$
Order of computational cost	$O(N^{2m-1}d)$

4. Malliavin Monte Carlo method.

4.1. Algorithm. We introduce *Malliavin Monte Carlo* for the proposed scheme in Theorem 3.13 for probability distribution functions as

$$
\begin{aligned}
(\widetilde{Q^m_{(T/n)}})^n f(x) &= \mathbb{E}\Big[\Big((\widetilde{Q^m_{(T/n)}})^{n-1} f\Big)(\overline{X}_{(T/n)}(T/n,x))\widetilde{\mathcal{M}_m}\Big(T/n, x, \Delta_1 B_{T/n}\Big)\Big] = \cdots \\
&= \mathbb{E}\Big[\Big(\widetilde{Q^m_{(T/n)}} f\Big)(\overline{X}_{(T/n)}((n-1)T/n,x)) \prod_{k=1}^{n-1} \widetilde{\mathcal{M}_m}\Big(T/n, \overline{X}_{(T/n)}((k-1)T,x), \Delta_k B_{T/n}\Big)\Big] \\
&\approx \frac{1}{M}\sum_{j=1}^{M} f(\overline{X}^{[j]}_{(T/n)}(T,x)) \prod_{k=1}^{n} \widetilde{\mathcal{M}_m}\Big(T/n, \overline{X}^{[j]}_{(T/n)}((k-1)T,x), \sqrt{T/n}\; z_k^{[j]}\Big),
\end{aligned} \tag{28}
$$

where $\Delta_k B_{T/n}$, $k = 1, \cdots, n$, are the Brownian increments defined by $\Delta_k B_{T/n} = B_{kT/n} - B_{(k-1)T/n}$. Here, $\overline{X}^{[j]}_{(T/n)}(kT/n, x)$, $1 \leq k \leq n$, $1 \leq j \leq M$, are given by

$$
\overline{X}^{[j]}_{(T/n)}(kT/n,x) = \overline{X}^{[j]}_{(T/n)}((k-1)T/n,x) + \sum_{i=1}^{d} V_i(\overline{X}^{[j]}_{(T/n)}((k-1)T/n,x))\sqrt{T/n}\; z_k^{i,[j]},
$$

starting from $\overline{X}^{[j]}_{(T/n)}(0,x) = x \in \mathbb{R}^N$, and $z_k^{[j]} = (z_k^{1,[j]}, \cdots, z_k^{d,[j]})$ is computed by $z_k^{l,[j]} = \Phi^{-1}(y_k^{l,[j]})$ from a *low-discrepancy* sequence $\{y_k^{l,[j]}\}_{k=1,\cdots,n,l=1,\cdots,d,j=1,\cdots,M}$ (in our case, the Sobol sequence) with the inverse of the normal distribution function Φ^{-1} when we use the quasi–Monte Carlo method (QMC), or is sampled as the realization $Z_k^{[j]}(\omega) = z_k^{[j]}$ ($\omega \in \Omega$) of independent d-dimensional standard Gaussian random variable $Z_k^{[j]}$ on a probability space $(\Omega, \mathcal{F}, \mathbb{Q})$ when the Monte Carlo method (MC) is applied. The algorithm of the Malliavin Monte Carlo is given in Algorithm 1.

Algorithm 1 Malliavin Monte Carlo method.

for $j = 1$ **to** M **do**
  $\overline{X}^{[j]}_{(T/n)}(0,x) = x$, $\mathscr{M}^{[j]}_{m,(T/n)}(0) = 1$.
  for $k = 1$ **to** n **do**
    Compute Malliavin weight $\mathscr{M}^{[j]}_{m,(T/n)}(kT/n) = \mathscr{M}^{[j]}_{m,(T/n)}((k-1)T/n) \times \widetilde{\mathcal{M}_m}(kT/n, \overline{X}^{[j]}_{(T/n)}((k-1)T/n,x), \sqrt{T/n}\; z_k^{[j]})$
    Compute $\overline{X}^{[j]}_{(T/n)}(kT/n,x) = \overline{X}^{[j]}_{(T/n)}((k-1)T/n,x) + V_0(\overline{X}^{[j]}_{(T/n)}((k-1)T/n,x))T/n + \sum_{i=1}^{d} V_i(\overline{X}^{[j]}_{(T/n)}((k-1)T/n,x))\sqrt{T/n}\; z_k^{i,[j]}$
  end for
end for
Return $\frac{1}{M}\sum_{j=1}^{M} f(\overline{X}^{[j]}_{(T/n)}(T,x))\mathscr{M}^{[j]}_{m,(T/n)}(T)$.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Remark 5 (complexity). If we apply QMC for the proposed scheme, the integration error would be close to $O(M^{-1})$ under a suitable setting, i.e., there exists $c(f)$ such that for all $M \in \mathbb{N}$,

$$\Big|(\widetilde{Q^m_{(T/n)}})^n f(x) - \frac{1}{M}\sum_{j=1}^{M} f(\overline{X}^{[j]}_{(T/n)}(T,x)) \prod_{k=1}^{n} \widetilde{\mathcal{M}_m}\Big(T/n, \overline{X}^{[j]}_{(T/n)}((k-1)T,x), \sqrt{T/n}\ z^j_k\Big)\Big| \le c(f)\frac{(\log M)^{nd}}{M}.$$

Then, in principle, the cost will be approximately $O(\varepsilon^{-\frac{1+m}{m}})$ in order to attain an error tolerance ε—for example, $O(\varepsilon^{-1.5})$ when $m = 2$ and $O(\varepsilon^{-1.333\cdots})$ when $m = 3$.

If we apply MC for the scheme, the statistical error behaves as approximately $O(M^{-1/2})$ since we have

$$\begin{aligned}
&\frac{1}{M}\sum_{j=1}^{M} f(\overline{X}^{[j]}_{(T/n)}(T,x)) \prod_{k=1}^{n} \widetilde{\mathcal{M}_m}\Big(T/n, \overline{X}^{[j]}_{(T/n)}((k-1)T,x), \sqrt{T/n}Z^j_k\Big)\\
\approx\ &\mathcal{N}\Big(\mathbb{E}\Big[f(\overline{X}_{(T/n)}(T,x)) \prod_{k=1}^{n} \widetilde{\mathcal{M}_m}\Big(T/n, \overline{X}_{(T/n)}((k-1)T,x), \Delta_k B_{T/n}\Big)\Big],\\
&\mathbb{V}\Big[f(\overline{X}_{(T/n)}(T,x)) \prod_{k=1}^{n} \widetilde{\mathcal{M}_m}\Big(T/n, \overline{X}_{(T/n)}((k-1)T,x), \Delta_k B_{T/n}\Big)\Big]/M\Big),
\end{aligned}$$

by the central limit theorem. Then, the computational cost will be $O(\varepsilon^{-\frac{1+2m}{m}})$ for a tolerance ε in the root mean square error sense in principle. For example, the cost of the scheme with $m = 2$ becomes $O(\varepsilon^{-2.5})$, while that of Bally and Talay [2] ($m = 1$) will be $O(\varepsilon^{-3})$. Note that the error of MC does not depend on the dimension.

Finally, for comparison, we comment on the multilevel Monte Carlo simulation (MLMC). Basically, MLMC works under the condition that the test function is at least Lipschitz continuous because MLMC is based on not only weak approximation but also strong approximation of SDEs; see [7] for the details. For a general scalar SDE, Avikainen [1] has proved that MLMC still works for a class of irregular functionals such as $\mathbf{1}_{\{X(T,x)\ge K\}}$ using a novel strong approximation. Then, the cost of MLMC for the Euler–Maruyama case is of $O(\varepsilon^{-2.5})$; that is, MLMC will be more efficient than the standard MC even for irregular functionals. We will compare the Malliavin Monte Carlo and MLMC in the section below.

4.2. Numerical experiments. In this subsection, we test the accuracy of the Malliavin Monte Carlo method. In the cost computations below, we take into account the number of arithmetic operations of scalar function evaluations as well as the number of realizations of random variables for each scheme for each step (based on Tables 1 and 2). Then the computational cost is computed by multiplying this value by the number of steps and the number of simulations. For each figure or table below, we use either QMC or MC for the schemes for fair comparison. In the error analysis, the absolute error and the root mean square error (RMSE) are used for QMC and MC, respectively. All computations are performed by a MacBook Pro with Intel Core i7 CPU with 3.3 GHz and 16 GB RAM.

4.2.1. Malliavin Monte Carlo in 2-dimensional SDE. Consider the following Stratonovich SDE ($N = d = 2$ in the previous section) known as the SABR model (local stochastic volatility model) which is typically used in practice in finance:

$$dX^j(t,x) = V^j_0(X(t,x))dt + \sum_{i=1}^{2} V^j_i(X(t,x)) \circ dB^i_t, \quad X^j(0,x) = x^j_0 > 0, \quad j = 1,2,$$

where $V^1_0(x) = -\frac{1}{2}[x_2^2 C(x_1)(\partial/\partial x_1)C(x_1) + \nu\rho x_2 C(x_1)]$, $V^1_1(x) = x_2 C(x_1)$, $V^1_2(x) = 0$, $V^2_0(x) = -\frac{1}{2}\nu^2 x_2$, $V^2_1(x) = \nu x_2 \rho$, $V^2_2(x) = \nu x_2\sqrt{1-\rho^2}$ with $\nu > 0$, $\rho \in (-1,1)$, and

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 03/28/19 to 130.70.8.131. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Downloaded 03/28/19 to 130.70.8.131. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

the local volatility function $C(\cdot)$. Here the processes $\{X^1(t,x)\}_{t\geq 0}$ and $\{X^2(t,x)\}_{t\geq 0}$ represent the underlying asset and its volatility. In the experiments, the local volatility function is specified as $C(x_1) = (x_0^1)^{1-\beta}x_1^{\beta}$ with the skewness parameter β and the Black–Scholes scaling $(x_0^1)^{1-\beta}$. Although the model above may not satisfy the smoothness assumption in the previous section rigorously, we can still use the proposed method by applying a smooth modification technique of [28]. We compute the binary option price or equivalently the probability given by the expectation $\mathbb{E}[f(X^1(T,x))]$ with $f(\cdot) = \mathbf{1}_{\{\cdot\geq K\}}$. The benchmark value (0.456578) is computed by the Euler–Maruyama scheme using Monte Carlo simulation with number of time steps $n = 2^{10} = 1,024$ and sample paths $M = 10^7$. The parameters are put as $T = 2.0$, $x_0^1 = 100$, $K = 100$, $\beta = 0.5$, $x_0^2 = 0.4$, $\nu = 0.1$, $\rho = -0.5$.

Figure 1 shows the absolute error versus the computational cost for the schemes computed by QMC. By the figure, we can check that the Malliavin Monte Carlo method attains a smaller error with a smaller cost.

Figure 2 gives the rate of convergence of the Euler–Maruyama (order one) scheme and the Malliavin Monte Carlo (order two) method using QMC. The plots ($n = 2, 4, 8, 16, 32$) are given under the computational costs 1.0×10^7, 2.1×10^8, 1.0×10^9, 2.6×10^9, 7.6×10^9. Here, the unit of the computational cost (1.0×10^7) is computed in 0.290 seconds. We can see that the weak convergence rate of the Malliavin Monte Carlo outperforms that of the Euler–Maruyama scheme, which is consistent with the theoretical result.

Table 3 reports the accuracy of the Malliavin Monte Carlo scheme and the Ninomiya–Victoir scheme [26], an algorithm of the cubature on Wiener space [14], [22]. Both methods are order two weak approximation schemes and are computed using QMC under the same condition with the number of time-steps $n = 2$ and the number of simulations $M = 128,000$ in order to compare the accuracy. The parameters are set to be $T = 1.0$, $x_0^1 = 1.0$, $K = 1.05$, $C(x) = x^{\beta}$, $x_0^2 = 0.3$, $\beta = 0.9$, $\nu = 0.4$, $\rho = -0.7$, and $f(\cdot) = \max\{\ \cdot - K, 0\}$. Under this setting, Malliavin Monte Carlo is computed in 0.098 seconds. The benchmark value and the numerical value of the Ninomiya–Victoir scheme are borrowing from [3]. Here, the reported Ninomiya–Victoir scheme [3] is an improved version using the semi–closed form cubature method which makes use of a special feature of the SABR model to speed up the computations, which does not apply in general. By Table 3, we can see the efficiency of the proposed scheme. Therefore, the computational cost is reduced by the Malliavin Monte Carlo method.

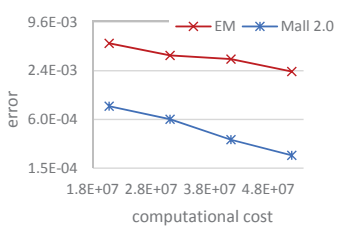

FIG. 1. *Numerical error versus computational cost.*

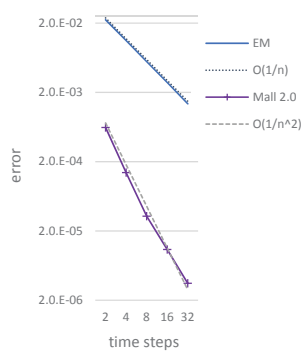

FIG. 2. *Convergence rate of weak approximation.*

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 03/28/19 to 130.70.8.131. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

TABLE 3
Comparison with Ninomiya–Victoir scheme.

Method	Numerical error
Ninomiya–Victoir [3] ($n = 2$)	0.0014
Malliavin Monte Carlo ($n = 2$)	0.000017

4.2.2. Comparison with the previous methods [29], [33]. We compare the accuracy of the schemes based on [29], [33], and the scheme of this paper under the SABR model which is implemented by QMC under the same parameters in the previous subsection. To see the efficiency, we check numerical error versus computational cost. Note that the scheme in [29] (approach (i)) can theoretically work near uniform partition but the scheme (3) in [33] (approach (ii)) does not work, as mentioned in section 1. On the other hand, the accuracy of the scheme in [33] may beat the lower order scheme in [29]. The scheme (4) of this paper would beat both schemes. Figure 3 shows these conjectures. Here, the labels Mall + K (Y) and Mall + K (T-Y) represent the schemes used in Yamada (2017) [33] and in Takahashi and Yamada (2016) [29], respectively. By the figure, we can check that the scheme with the universal weight provides faster convergence with lower computational cost. Consequently, computational cost will be reduced by the method of this paper.

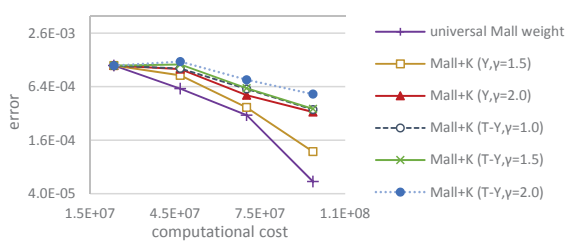

FIG. 3. *Comparison of accuracy under uniform and nonuniform partitions.*

4.2.3. Malliavin Monte Carlo in 10-dimensional SDE. Next, we apply MC to the proposed weak approximation scheme. Let us consider the following multidimensional Stratonovich SDE ($N = d = 10$):

$$(29)\quad dX^j(t,x) = V_0^j(X(t,x))dt + V_j^j(X(t,x)) \circ dB_t^j, \quad X^j(0,x) = x_0^j > 0, \quad j = 1, \cdots, 10,$$

where $V_0^j(x) = rx_j - \frac{1}{2}\hat{V}_j V_j^j(x)$, $V_j^j(x) = \sigma_j x_j$, $j = 1, \cdots, 10$. We compute the digital basket call option price with a bounded Borel payoff function $f(\xi_1, \cdots, \xi_{10}) = \mathbf{1}_{\{(1/10)\sum_{i=1}^{10}\xi_i \geq K\}}$, $K \geq 0$. The parameters are put as $T = 1.0$, $K = 100$, $r = 0.01$, $\sigma_j = 0.15$, $x_0^j = 100$, $j = 1, \cdots, 10$. Figure 4 plots the RMSE versus the computational cost for the Euler–Maruyama scheme and the Malliavin Monte Carlo. Here, the benchmark value (0.568043) is computed by MC with a number of simulations $M = 10^8$ through the explicit solution of the SDE by the Itô formula. Similar to the SABR case, the Malliavin Monte Carlo method gives smaller error with smaller cost. Also, Figure 5 shows the rate of weak convergence of the Euler-Maruyama scheme and the Malliavin Monte Carlo method in terms of RMSE, where the plots ($n = 2, 4, 8, 16, 32$) are given under the computational costs 8.0×10^9, 1.6×10^{10}, 3.2×10^{10}, 6.4×10^{10}, 1.3×10^{11}. By the results, the Malliavin Monte Carlo method can be an efficient weak scheme for multidimensional SDEs.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 03/28/19 to 130.70.8.131. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

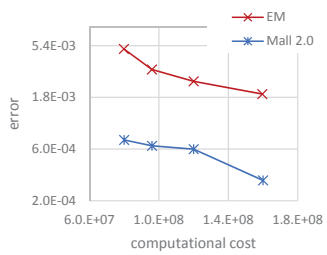

FIG. 4. *Numerical error versus computational cost.*

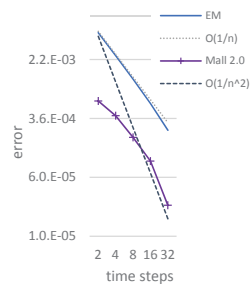

FIG. 5. *Convergence rate of weak approximation.*

4.2.4. Comparison of second and third order Malliavin Monte Carlo and MLMC. Finally, we compare the Malliavin Monte Carlo (of weak order $m = 2$ and $m = 3$) and MLMC for an irregular functional. Applicability of MLMC for irregular functionals was justified by Avikanen [1] for a general scalar case. We adopt the following scalar Stratonovich SDE as an example:

$$dX(t,x) = V_0(X(t,x))dt + V_1(X(t,x)) \circ dB_t^1, \quad X(0,x) = x_0 > 0,$$

where $V_0(x) = rx - \frac{1}{2}\hat{V}_1 V_1(x)$, $V_1(x) = \sigma x$. We compute the binary option price $e^{-rT}\mathbb{E}[\mathbf{1}_{X(T,x)\geq K}]$, $K \geq 0$. The parameters are put as $T = 2$, $x_0 = 100$, $K = 100$, $r = 0.01$, $\sigma = 0.2$. For the model, the benchmark value can be obtained from the closed form formula. Figure 6 plots the RMSE versus the computational cost for the schemes: the Euler–Maruyama scheme, MLMC, and the Malliavin Monte Carlo of weak order $m = 2$ and $m = 3$. The figure indicates that the Malliavin Monte Carlo gives faster convergence than MLMC, while both schemes beat the standard Euler–Maruyama scheme.

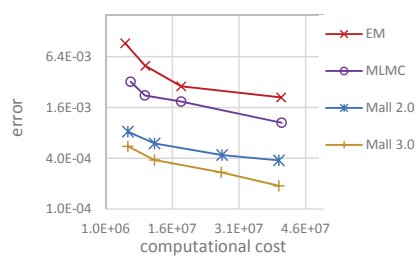

FIG. 6. *Numerical error versus computational cost.*

5. Concluding remarks. This paper showed an arbitrary high order weak approximation of stochastic differential equations. We introduced the universal weight in order to obtain a general weak approximation scheme under uniform partition which works whether the test function is smooth or not. The explicit formula of the universal weight was derived through Itô–Stratonovich calculus and Malliavin integration by parts. We performed various numerical experiments for the scheme to confirm the validity.

In future work, it should be considered that the scheme can work under a weaker condition (e.g., the hypoelliptic case). Also numerical tests for the higher order ($m \geq 3$) scheme for various models should be performed in more detail. Improving

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

the scheme or constructing a more efficient technique would be an interesting topic. For example, it is worth considering an extension of the present work to MLMC as in Debrabant and Rössler [5]. Debrabant and Rössler [5] provided an idea and a framework to combine MLMC with general higher order weak approximation in realistic situations. Then they obtained a more efficient scheme than the standard MLMC. In [5], the final level of MLMC is replaced by a higher order weak approximation, while the Euler–Maruyama scheme, a lower order weak approximation (of which strong order is known), is used until the final level. It could be possible to construct an efficient "Malliavin" MLMC for irregular test functions through the idea of [5] and the strong approximation of Avikainen [1].

Appendix A. Proof of Lemma 3.5. Iterating Lemma 3.2, we have

$$\mathbb{B}^{\mathrm{Strat}}_{\alpha^\nu,t} = \sum_{\substack{\sigma_l\in S_{m_l+n_l},\ l=1,\cdots,k-1,\\ s.t.\ m_l=\sum_{i=1}^{l} r(i),\ n_l=r(l+1),\\ \sigma_l(1)<\cdots<\sigma_l(m_l),\ \sigma_l(m_l+1)<\cdots<\sigma_l(m_l+n_l)}} \mathbb{B}^{\mathrm{Strat}}_{\sigma_{k-1}^{-1}\cdot(\cdots(\sigma_2^{-1}\cdot(\sigma_1^{-1}\cdot(\alpha^1*\alpha^2)*\alpha^3)*\cdots)*\alpha^k),t}\ .$$

Then, by Lemma 3.4, we obtain

$$\mathbb{B}^{\mathrm{Strat}}_{\alpha^\nu,t} = \sum_{\substack{\sigma_l\in S_{m_l+n_l},\ l=1,\cdots,k-1,\\ s.t.\ m_l=\sum_{i=1}^{l} r(i),\ n_l=r(l+1),\\ \sigma_l(1)<\cdots<\sigma_l(m_l),\ \sigma_l(m_l+1)<\cdots<\sigma_l(m_l+n_l)}} \sum_{\substack{\gamma,\beta\ s.t.\ \beta\sim_p\gamma,\ p\in\mathbb{N},\\ \beta=\sigma_{k-1}^{-1}\cdot(\cdots(\sigma_2^{-1}\cdot(\sigma_1^{-1}\cdot(\alpha^1*\alpha^2)*\alpha^3)*\cdots)*\alpha^k)}} \frac{1}{2^{\eta(\gamma,\beta)}}\mathbb{B}^{\mathrm{Itô}}_{\gamma,t}\ . \tag{30}$$

We will use the inductive method to obtain the following representation:

$$\begin{aligned}\mathbb{E}[g(\overline{X}(t,x))\mathbb{B}^{\mathrm{Itô}}_{\gamma,t}] &= \sum_{J\in\{1,\cdots,N\}^{K(\gamma)}}\prod_{h=1}^{K(\gamma)} V^{J_h}_{\gamma_h^\star}(x)\frac{1}{|\gamma|!}t^{|\gamma|}\mathbb{E}\left[g(\overline{X}(t,x))H_{(J_1,\cdots,J_{K(\gamma)})}(\overline{X}(t,x),1)\right].\end{aligned} \tag{31}$$

Assume that for any $h\in C_b^\infty(\mathbb{R}^N)$ and $\alpha\in\{0,1,\cdots,d\}^r$, $r<|\gamma|$,

$$\mathbb{E}[h(\overline{X}(t,x))\mathbb{B}^{\mathrm{Itô}}_{\alpha,t}] = \sum_{J\in\{1,\cdots,N\}^{K(\alpha)}}\prod_{h=1}^{K(\alpha)} V^{J_h}_{\alpha_h^\star}(x)\frac{1}{|\alpha|!}t^{|\alpha|}\mathbb{E}\left[h(\overline{X}(t,x))H_{(J_1,\cdots,J_{K(\alpha)})}(\overline{X}(t,x),1)\right].$$

Define $h^\gamma_{i+1}\in C_b^\infty(\mathbb{R}^N)$ recursively,

$$h^\gamma_{i+1}(\cdot) = \sum_{1\le\kappa_i\le N}\partial_{\kappa_i}h^\gamma_i(\cdot)V^{\kappa_i}_{\gamma_{|\gamma|-i}}(x)\mathbf{1}_{|\gamma|-i=0} + h^\gamma_i(\cdot)\mathbf{1}_{|\gamma|-i\neq 0},\quad 1\le i\le|\gamma|,$$

with $h^\gamma_1(\cdot)=\sum_{1\le\kappa_1\le N}\partial_{\kappa_1}g(\cdot)V^{\kappa_1}_{\gamma_{|\gamma|}}(x)\mathbf{1}_{|\gamma|=0}+g(\cdot)\mathbf{1}_{|\gamma|\neq 0}$. Using the Malliavin inte-

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 03/28/19 to 130.70.8.131. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Downloaded 03/28/19 to 130.70.8.131. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

gration by parts and the Itô formula, we can see the following:

$$
\begin{aligned}
&\mathbb{E}[g(\overline{X}(t,x))\mathbb{B}^{\mathrm{It\hat{o}}}_{\gamma,t}]\\
&= \sum_{1\le\kappa_1\le N} \mathbb{E}[\partial_{\kappa_1} g(\overline{X}(t,x))V^{\kappa_1}_{\gamma_{|\gamma|}}(x)\int_0^t \mathbb{B}^{\mathrm{It\hat{o}}}_{\gamma^{1-},s}dB^0_s]\mathbf{1}_{\gamma_{|\gamma|}\neq 0} + \mathbb{E}[g(\overline{X}(t,x))\int_0^t \mathbb{B}^{\mathrm{It\hat{o}}}_{\gamma^{1-},s}dB^0_s]\mathbf{1}_{\gamma_{|\gamma|}=0}\\
&= \mathbb{E}[h^\gamma_1(\overline{X}(t,x))\int_0^t \mathbb{B}^{\mathrm{It\hat{o}}}_{\gamma^{1-},s}dB^0_s]\\
&= \mathbb{E}[h^\gamma_1(\overline{X}(t,x))\mathbb{B}^{\mathrm{It\hat{o}}}_{(0),t}\mathbb{B}^{\mathrm{It\hat{o}}}_{\gamma^{1-},t}] - \mathbb{E}[h^\gamma_1(\overline{X}(t,x))\int_0^t \mathbb{B}^{\mathrm{It\hat{o}}}_{(0),s}\mathbb{B}^{\mathrm{It\hat{o}}}_{\gamma^{2-},s}dB^{\gamma_{|\gamma|-1}}_s]\\
&= \mathbb{E}[h^\gamma_1(\overline{X}(t,x))\mathbb{B}^{\mathrm{It\hat{o}}}_{(0),t}\mathbb{B}^{\mathrm{It\hat{o}}}_{\gamma^{1-},t}]\\
&- \sum_{1\le\kappa_2\le N} \mathbb{E}[\partial_{\kappa_2} h^\gamma_1(\overline{X}(t,x))V^{\kappa_2}_{\gamma_{|\gamma|-1}}(x)\int_0^t \mathbb{B}^{\mathrm{It\hat{o}}}_{(0),s}\mathbb{B}^{\mathrm{It\hat{o}}}_{\gamma^{2-},s}dB^0_s]\mathbf{1}_{\gamma_{|\gamma|-1}\neq 0} - \mathbb{E}[g(\overline{X}(t,x))\int_0^t \mathbb{B}^{\mathrm{It\hat{o}}}_{\gamma^{1-},s}dB^0_s]\mathbf{1}_{\gamma_{|\gamma|-1}=0}\\
&= \mathbb{E}[h^\gamma_1(\overline{X}(t,x))\mathbb{B}^{\mathrm{It\hat{o}}}_{(0),t}\mathbb{B}^{\mathrm{It\hat{o}}}_{\gamma^{1-},t}] - \mathbb{E}[h^\gamma_2(\overline{X}(t,x))\int_0^t \mathbb{B}^{\mathrm{It\hat{o}}}_{(0),s}\mathbb{B}^{\mathrm{It\hat{o}}}_{\gamma^{2-},s}dB^0_s]\\
&= \mathbb{E}[h^\gamma_1(\overline{X}(t,x))\mathbb{B}^{\mathrm{It\hat{o}}}_{(0),t}\mathbb{B}^{\mathrm{It\hat{o}}}_{\gamma^{1-},t}] - \mathbb{E}[h^\gamma_2(\overline{X}(t,x))\mathbb{B}^{\mathrm{It\hat{o}}}_{(0,0),t}\mathbb{B}^{\mathrm{It\hat{o}}}_{\gamma^{2-},t}] + \mathbb{E}[h^\gamma_2(\overline{X}(t,x))\int_0^t \mathbb{B}^{\mathrm{It\hat{o}}}_{(0,0),s}\mathbb{B}^{\mathrm{It\hat{o}}}_{\gamma^{3-},s}dB^{\gamma_{|\gamma|-2}}_s].
\end{aligned}
$$

Then we iterate this procedure and use the assumption to get

$$
\begin{aligned}
&\mathbb{E}[g(\overline{X}(t,x))\mathbb{B}^{\mathrm{It\hat{o}}}_{\gamma,t}] = \sum_{i=1}^{|\gamma|}(-1)^{i-1}\mathbb{E}[h^\gamma_i(\overline{X}(t,x))\mathbb{B}^{\mathrm{It\hat{o}}}_{\gamma^{i-},t}]\frac{1}{i!}t^i\\
&= \sum_{i=1}^{|\gamma|}(-1)^{i-1}\frac{1}{i!}t^i \sum_{J\in\{1,\cdots,N\}^{K(\gamma^{i-})}} \prod_{h=1}^{K(\gamma^{i-})} V^{J_h}_{\gamma^{i-\star}_h}(x)\frac{1}{|\gamma^{i-}|!}t^{|\gamma^{i-}|}\mathbb{E}\left[h^\gamma_i(\overline{X}(t,x))H_{(J_1,\cdots,J_{K(\gamma^{i-})})}(\overline{X}(t,x),1)\right].
\end{aligned}
$$

By definition of h^γ_i and (5), we have

$$
\begin{aligned}
&\mathbb{E}[g(\overline{X}(t,x))\mathbb{B}^{\mathrm{It\hat{o}}}_{\gamma,t}]\\
&= \sum_{i=1}^{|\gamma|}(-1)^{i-1}\frac{1}{i!}t^i \sum_{J\in\{1,\cdots,N\}^{K(\gamma)}} \prod_{h=1}^{K(\gamma)} V^{J_h}_{\gamma^\star_h}(x)\frac{1}{|\gamma^{i-}|!}t^{|\gamma^{i-}|}\mathbb{E}\left[g(\overline{X}(t,x))H_{(J_1,\cdots,J_{K(\gamma)})}(\overline{X}(t,x),1)\right]\\
&= \sum_{J\in\{1,\cdots,N\}^{K(\gamma)}} \prod_{h=1}^{K(\gamma)} V^{J_h}_{\gamma^\star_h}(x)\mathbb{E}\left[g(\overline{X}(t,x))H_{(J_1,\cdots,J_{K(\gamma)})}(\overline{X}(t,x),1)\right]\sum_{i=1}^{|\gamma|}(-1)^{i-1}\frac{1}{i!}t^i\frac{1}{|\gamma^{i-}|!}t^{|\gamma^{i-}|}.
\end{aligned}
$$

We easily see that

$$
\sum_{i=1}^{|\gamma|}(-1)^{i-1}\frac{1}{i!}t^i\frac{1}{|\gamma^{i-}|!}t^{|\gamma^{i-}|} = \sum_{i=1}^{|\gamma|}(-1)^{i-1}\frac{1}{i!}t^i\frac{1}{(|\gamma|-i)!}t^{|\gamma|-i} = \frac{1}{|\gamma|!}t^{|\gamma|}\sum_{i=1}^{|\gamma|}(-1)^{i-1}\ {}_{|\gamma|}C_i = \frac{1}{|\gamma|!}t^{|\gamma|}.
$$

Then the representation (31) is obtained. We have the assertion by combining (30) and (31).

Appendix B. Proof of Theorem 3.6. We introduce an SDE with a parameter $\lambda\in(0,1]$:

$$
dX^\lambda(t,x) = \lambda^2 V_0(X^\lambda(t,x))dt + \lambda\sum_{i=1}^d V_i(X^\lambda(t,x))\circ dB^i_t,\quad X^\lambda(0,x)=x, \tag{32}
$$

and a nondegenerate Wiener functional $\vartheta^\lambda(t,x) = \frac{X^\lambda(t,x)-x}{\lambda}$, $t\in(0,1]$, $x\in\mathbb{R}^N$. For $e\ge 2$, we have

$$
\vartheta^\lambda(t,x) = \sum_{i=1}^d V_i(x)B^i_t + \sum_{k=2}^e \lambda^{k-1}\sum_{\substack{\alpha=(\alpha_1,\cdots,\alpha_k)\in\{0,1,\cdots,d\}^{r(k)},\\ r(k)\in\mathbb{N},\|\alpha\|=k}} \hat{V}_{\alpha_1}\cdots\hat{V}_{\alpha_{k-1}}V_{\alpha_k}(x)\mathbb{B}^{\mathrm{Strat}}_{\alpha,t} + r^\lambda_{e+1}(t,x)
$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

for $t \in (0,1]$, $x \in \mathbb{R}^N$, by the Stratonovich–Taylor expansion, where $r^{\lambda}_{e+1}(t,x)$ is the residual. We set $\vartheta^0(t,x) = \sum_{i=1}^d V_i(x)B^i_t$. For $\varphi \in C^\infty_b(\mathbb{R}^N)$, one has

$$\mathbb{E}[\varphi(X^\lambda(t,x))] = \int_{\mathbb{R}^N} \varphi(x+\lambda y)\langle \delta_y(\vartheta^\lambda(t,x)), 1\rangle dy \tag{33}$$

by (7). We expand $\langle \delta_y(\vartheta^\lambda(t,x)), 1\rangle$ in (33) using (6) as

$$\langle \delta_y(\vartheta^\lambda(t,x)), 1\rangle = \left\langle \delta_y\left(\sum_{i=1}^d V_i(x)B^i_t\right), 1\right\rangle + \sum_{j=1}^{2m+1}\lambda^j \sum_{k=1}^{j} \sum_{\substack{\sum_{i=1}^k b_i=j+k,\ I=(I_1,\cdots,I_k)\in\{1,\cdots,N\}^k \\ b_i\geq 2,\ i=1,\cdots,k}} \frac{1}{k!}$$

$$\left\langle \delta_y\left(\sum_{i=1}^d V_i(x)B^i_t\right), H_{(I_1,\cdots,I_k)}\Big(\sum_{i=1}^d V_i(x)B^i_t, \prod_{\nu=1}^k \sum_{\substack{\alpha^\nu\in\{0,1,\cdots,d\}^{r(\nu)}, \\ \|\alpha^\nu\|=b_\nu, r(\nu)\in\mathbb{N}}} \hat{V}_{\alpha^\nu_1}\cdots\hat{V}_{\alpha^\nu_{b_\nu-1}} V^{I_\nu}_{\alpha^\nu_{b_\nu}}(x)\mathbb{B}^{\mathrm{Strat}}_{\alpha^\nu,t}\Big)\right\rangle$$

$$+\lambda^{2m+2}\int_0^1 \frac{(1-\xi)^{2m+1}}{(2m+1)!}\sum^{(2m+2)}\left\langle \delta_y(\vartheta^{\lambda\xi}(t,x)), H_{(I_1,\cdots,I_k)}\left(\vartheta^{\lambda\xi}(t,x), \prod_{\nu=1}^k \frac{1}{d_\nu!}\partial^{d_\nu}_\eta \vartheta^{\eta,I_\nu}(t,x)|_{\eta=\lambda\xi}\right)\right\rangle d\xi.$$

Here,

$$\sum^{(2m+2)} = (2m+2)\sum_{k=1}^{2m+2}\sum_{\substack{\sum_{i=1}^k d_i=2m+2,\ I=(I_1,\cdots,I_k)\in\{1,\cdots,N\}^k \\ d_i\geq 1,\ i=1,\cdots,k}} \frac{1}{k!}.$$

By (33) and the above expansion with letting $\lambda = 1$, we obtain

$$\mathbb{E}[\varphi(X(t,x))] = \mathbb{E}[\varphi(\overline{X}(t,x))] + \sum_{j=1}^{2m+1}\sum_{k=1}^{j}\sum_{\substack{\sum_{i=1}^k b_i=j+k,\ I=(I_1,\cdots,I_k)\in\{1,\cdots,N\}^k \\ b_i\geq 2,\ i=1,\cdots,k}} \frac{1}{k!} \tag{34}$$

$$\mathbb{E}\Big[\varphi(\overline{X}(t,x))H_{(I_1,\cdots,I_k)}\Big(\sum_{i=1}^d V_i(x)B^i_t, \prod_{\nu=1}^k \sum_{\substack{\alpha^\nu\in\{0,1,\cdots,d\}^{r(\nu)}, \\ \|\alpha^\nu\|=b_\nu, r(\nu)\in\mathbb{N}}} \hat{V}_{\alpha^\nu_1}\cdots\hat{V}_{\alpha^\nu_{b_\nu-1}} V^{I_\nu}_{\alpha^\nu_{b_\nu}}(x)\mathbb{B}^{\mathrm{Strat}}_{\alpha^\nu,t}\Big)\Big] + \mathcal{E}^{m,\varphi}_1(t,x),$$

where

$$\mathcal{E}^{m,\varphi}_1(t,x) = \int_0^1 \frac{(1-\xi)^{2m+1}}{(2m+1)!}\sum^{(2m+2)}\mathbb{E}\Big[\varphi\left(x+\vartheta^\xi(t,x)\right)H_{(I_1,\cdots,I_k)}\Big(\vartheta^\xi(t,x), \prod_{\nu=1}^k \frac{1}{d_\nu!}\partial^{d_\nu}_\eta\vartheta^{\eta,I_\nu}(t,x)|_{\eta=\xi}\Big)\Big]d\xi.$$

For the error estimate, we prepare the following.

Lemma B.1.

1. *For $r \in \mathbb{N}$, $k \in \mathbb{N}$, $p \geq 1$, there exist $C > 0$ and $Q > 0$ such that for all $t \in (0,1]$, $x \in \mathbb{R}^N$, and $1 \leq i \leq N$,*

$$\|\partial^r_\lambda \vartheta^{\lambda,i}(t,x)\|_{k,p} \leq Ct^{(r+1)/2}(1+|x|^Q). \tag{35}$$

2. *For $k \in \mathbb{N}$, $(I_1,\cdots,I_k) \in \{1,\cdots,N\}^k$, there exist $C > 0$ and $Q > 0$ such that for all $t \in (0,1]$, $x \in \mathbb{R}^N$, $\varphi \in C^\infty_b(\mathbb{R}^N)$, and $G \in \mathbb{D}^\infty$,*

$$\sup_{\lambda\in[0,1]}\left|\mathbb{E}[\varphi(x+\vartheta^\lambda(t,x))H_{(I_1,\cdots,I_k)}(\vartheta^\lambda(t,x),G)]\right| \leq C\|\varphi\|_\infty t^{-k/2}\|G\|_{k,q}(1+|x|^Q) \tag{36}$$

for some $q > 1$.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 03/28/19 to 130.70.8.131. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Downloaded 03/28/19 to 130.70.8.131. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Proof of Lemma B.1. We essentially use the approach of [33] for the Wiener functional $\vartheta^\lambda(t,x) = \frac{X^\lambda(t,x)-x}{\lambda}$ with the solution to the Stratonovich SDE $X^\lambda(t,x)$. As in the proof of Lemma 1 of [33] or Lemma 2 of [30], $\partial_\lambda^r \vartheta^{\lambda,i}(t,x)$ involves iterated stochastic integrals of length $|\alpha| \geq r+1$ and then is in the space of local Kusuoka–Stroock functions $\mathcal{K}_{r+1}^{loc}$ (Definition 22 of [4]) by Lemma 23 of [4], a generalization of 1 of Lemma 7 of [15] and Lemma 1.3 of [19]. Then we have (35) by taking into account the polynomial growth order of V_j, $j = 0, 1, \cdots, d$, and their derivatives. Applying Corollary 3.7 of [17] to our case as in the proof of Lemma 1 of [33], we have that for $p \geq 2$,

$$\sup_{\lambda\in[0,1]} \|H_{(I_1,\cdots,I_k)}(\vartheta^\lambda(t,x),G)\|_p \leq Ct^{-k/2}\|G\|_{k,q}(1+|x|^Q) \tag{37}$$

for some $q > 1$, and therefore we get (36). ∎

Applying (36) with (35) and Hölder's inequality for the Sobolev norm, we have

$$\begin{aligned}&\left|\mathbb{E}\Big[\varphi(x+\vartheta^\xi(t,x))H_{(I_1,\cdots,I_k)}\Big(\vartheta^\xi(t,x),\prod_{\nu=1}^k \partial_\eta^\nu \vartheta^{\eta,I_\nu}(t,x)|_{\eta=\xi}\Big)\Big]\right| \\ &\leq Ct^{-k/2}t^{(2m+2+k)/2}\|\varphi\|_\infty(1+|x|^Q) \;=\; Ct^{m+1}\|\varphi\|_\infty(1+|x|^Q)\end{aligned}$$

for $k \in \mathbb{N}$ and $(I_1,\cdots,I_k) \in \{1,\cdots,N\}^k$, uniformly in $\xi \in [0,1]$. Then, we get $|\mathcal{E}_1^{m,\varphi}(t,x)| \leq Ct^{m+1}\|\varphi\|_\infty(1+|x|^Q)$. Note that the terms in (34) can be computed as

$$\begin{aligned}&\sum_{j=1}^{2m+1}\sum_{k=1}^{j}\sum_{\substack{\sum_{i=1}^k b_i=j+k,\\ b_i\geq 2,\ i=1,\cdots,k}}\sum_{I=(I_1,\cdots,I_k)\in\{1,\cdots,N\}^k}\frac{1}{k!}\mathbb{E}\Big[\varphi(\overline{X}(t,x))H_{(I_1,\cdots,I_k)}\Big(\sum_{i=1}^d V_i(x)B_t^i, \\ &\qquad\prod_{\nu=1}^k\sum_{\substack{\alpha^\nu\in\{0,1,\cdots,d\}^{r(\nu)},\\ \|\alpha^\nu\|=b_\nu,r(\nu)\in\mathbb{N}}}\hat{V}_{\alpha_1^\nu}\cdots\hat{V}_{\alpha_{b_{\nu-1}}^\nu}V_{\alpha_{b_\nu}^\nu}^{I_\nu}(x)\mathbb{B}_{\alpha^\nu,t}^{\mathrm{Strat}}\Big)\Big] \\ &= \sum_{j=1}^{2m+1}\sum_{k=1}^{j}\sum_{\substack{\sum_{i=1}^k b_i=j+k,\\ b_i\geq 2,\ i=1,\cdots,k}}\sum_{I=(I_1,\cdots,I_k)\in\{1,\cdots,N\}^k}\frac{1}{k!}\sum_{\substack{\alpha^\nu\in\{0,1,\cdots,d\}^{r(\nu)},\\ \|\alpha^\nu\|=b_\nu,r(\nu)\in\mathbb{N},\\ \nu=1,\cdots,k}} \\ &\qquad\prod_{\nu=1}^k\hat{V}_{\alpha_1^\nu}\cdots\hat{V}_{\alpha_{b_{\nu-1}}^\nu}V_{\alpha_{b_\nu}^\nu}^{I_\nu}(x)\mathbb{E}\Big[\partial_{I_1}\cdots\partial_{I_k}\varphi(\overline{X}(t,x))\prod_{\nu=1}^k\mathbb{B}_{\alpha^\nu,t}^{\mathrm{Strat}}\Big].\end{aligned}$$

Apply Lemma 3.5 and the Malliavin integration by parts to get

$$\begin{aligned}&\mathbb{E}[\varphi(X(t,x))] = \mathbb{E}[\varphi(\overline{X}(t,x))] + \sum_{j=1}^{2m+1}\sum_{k=1}^{j}\sum_{\substack{\sum_{i=1}^k b_j=j+k,\\ b_i\geq 2,\ i=1,\cdots,k}}\sum_{I=(I_1,\cdots,I_k)\in\{1,\cdots,N\}^k}\frac{1}{k!} \\ &\sum_{\substack{\alpha^\nu\in\{0,1,\cdots,d\}^{r(\nu)},\\ \|\alpha^\nu\|=b_\nu,r(\nu)\in\mathbb{N},\\ \nu=1,\cdots,k}}\prod_{\nu=1}^k\hat{V}_{\alpha_1^\nu}\cdots\hat{V}_{\alpha_{b_{\nu-1}}^\nu}V_{\alpha_{b_\nu}^\nu}^{I_\nu}(x)\sum_{\substack{\sigma_l\in S_{m_l+n_l},\ l=1,\cdots,k-1,\\ s.t.\ m_l=\sum_{i=1}^l b_i,\ n_l=b_{l+1},\\ \sigma_l(1)<\cdots<\sigma_l(m_l),\ \sigma_l(m_l+1)<\cdots<\sigma_l(m_l+n_l)}} \\ &\sum_{\substack{\beta\sim_p\gamma,\ p\in\mathbb{N},\\ \beta=\sigma_{k-1}^{-1}\cdot(\cdots(\sigma_2^{-1}\cdot(\sigma_1^{-1}\cdot(\alpha^1*\alpha^2)*\alpha^3)*\cdots)*\alpha^k)}}\frac{1}{2^{\eta(\gamma,\beta)}}\sum_{J\in\{1,\cdots,N\}^{K(\gamma)}}\prod_{h=1}^{K(\gamma)}V_{\gamma_h^*}^{J_h}(x)\frac{1}{|\gamma|!}t^{|\gamma|} \\ &\qquad\mathbb{E}\Big[\varphi(\overline{X}(t,x))H_{(I_1,\cdots,I_k,J_1,\cdots,J_{K(\gamma)})}(\overline{X}(t,x),1)\Big] + \mathcal{E}_1^{m,\varphi}(t,x).\end{aligned}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 03/28/19 to 130.70.8.131. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Appendix C. Proof of Theorem 3.9. Some expectations in the approximation in Theorem 3.6 will be $O(t^{m+1})$-order using Lemma 3.8 if we use the smoothness of $\varphi \in C_b^\infty(\mathbb{R}^N)$, and then expectations of $O(t^{m+1})$-order can be treated as the residual of expansion. Observe that

$$\sum_{J\in\{1,\cdots,N\}^{K(\gamma)}} \prod_{h=1}^{K(\gamma)} V_{\gamma_h^*}^{J_h}(x)\frac{1}{|\gamma|!}t^{|\gamma|}\mathbb{E}\Big[\varphi(\overline{X}(t,x))H_{(I_1,\cdots,I_k,J_1,\cdots,J_{K(\gamma)})}(\overline{X}(t,x),1)\Big]$$

$$= \sum_{J\in\{1,\cdots,N\}^{K(\gamma)}} \prod_{h=1}^{K(\gamma)} V_{\gamma_h^*}^{J_h}(x)\frac{1}{|\gamma|!}t^{|\gamma|}\mathbb{E}\Big[\partial_{I_1}\cdots\partial_{I_k}\partial_{J_1}\cdots\partial_{J_{K(\gamma)}}\varphi(\overline{X}(t,x))\Big] \tag{38}$$

$$= \mathbb{E}\Big[\partial_{I_1}\cdots\partial_{I_k}\varphi(\overline{X}(t,x))\mathbb{B}_{\gamma,t}^{\mathrm{Itô}}\Big] \tag{39}$$

by (5) and the proof of Lemma 3.8. By the definition of γ, we have $\|\gamma\| = \|\beta\| = \sum_{j=1}^k b_j = j+k$. Then, if $\|\gamma\| = j+k \geq 2m+2$, we obviously see

$$\left|\frac{1}{|\gamma|!}t^{|\gamma|}\mathbb{E}\Big[\varphi(\overline{X}(t,x))H_{(I_1,\cdots,I_k,J_1,\cdots,J_{K(\gamma)})}(\overline{X}(t,x),1)\Big]\right| \leq Ct^{m+1}\|\nabla^k\varphi\|_\infty(1+|x|^Q)$$

by (39). Even if $\|\gamma\| = j+k \leq 2m+1$ with $|\gamma| \geq m+1$, we have

$$\left|\frac{1}{|\gamma|!}t^{|\gamma|}\mathbb{E}\Big[\varphi(\overline{X}(t,x))H_{(I_1,\cdots,I_k,J_1,\cdots,J_{K(\gamma)})}(\overline{X}(t,x),1)\Big]\right| \leq Ct^{m+1}\|\nabla^{k+K(\gamma)}\varphi\|_\infty(1+|x|^Q)$$

by Lemma 3.8. From these observations, we can decompose (12) into

$$\mathbb{E}[\varphi(X(t,x))] = \mathbb{E}[\varphi(\overline{X}(t,x))\{1 + \sum_{\substack{(j,k),\\ j\leq 2m+1,\\ k\leq j,\\ j+k\leq 2m+1}} \sum_b\sum_I\sum_\alpha \sum_{\substack{\gamma,\\ \beta\sim_p\gamma,\\ |\gamma|\leq m}} \sum_J \mathbf{W}(t,x,b,I,\alpha,\gamma,\beta,J,B_t)\}]$$
$$+\mathcal{E}_1^{m,\varphi}(t,x) + \mathcal{E}_2^{m,\varphi}(t,x),$$

where

$$\mathcal{E}_2^{m,\varphi}(t,x) = \sum_{\substack{(j,k),\\ j\leq 2m+1,\\ k\leq j,\\ j+k\geq 2m+2}} \sum_b\sum_I\sum_\alpha \sum_{\substack{\gamma,\\ \beta\sim_p\gamma}} \sum_J \mathbf{W}(t,x,b,I,\alpha,\gamma,\beta,J,B_t)$$
$$+ \sum_{\substack{(j,k),\\ j\leq 2m+1,\\ k\leq j,\\ j+k\leq 2m+1}} \sum_b\sum_I\sum_\alpha \sum_{\substack{\gamma,\\ \beta\sim_p\gamma,\\ |\gamma|\geq m+1}} \sum_J \mathbf{W}(t,x,b,I,\alpha,\gamma,\beta,J,B_t).$$

Therefore, it holds that $\mathbb{E}[\varphi(X(t,x))] = \mathbb{E}[\varphi(\overline{X}(t,x))\widetilde{\mathcal{M}_m}(t,x,B_t)] + \mathcal{E}_1^{m+1,\varphi}(t,x) + \mathcal{E}_2^{m+1,\varphi}(t,x)$. Note that we already see that $|\mathcal{E}_1^{m+1,\varphi}(t,x)| \leq C_1 t^{m+1}\|\varphi\|_\infty(1+|x|^{Q_1})$. Further we obtain $\mathcal{E}_2^{m+1,\varphi}(t,x) = t^{m+1}\sum_{i=1}^{e(m)}\mathbb{E}[\partial^{\alpha^i}\varphi(\overline{X}(t,x))]K_{\alpha^i}(t,x)$ by (38), with an integer $e(m)\in\mathbb{N}$, multi-indices α^i satisfying $|\alpha^i| < 2(m+1)$, $i = 1,\cdots,e(m)$, and functions $K_{\alpha^i}:[0,T]\times\mathbb{R}^N\to\mathbb{R}$ satisfying the following: there exist $C_2 > 0$ and $Q_2 > 0$ such that $\sup_{t\in[0,1]}|K_{\alpha^i}(t,x)| \leq C_2(1+|x|^{Q_2})$. Taking the bounds of $\partial^{\alpha^i}\varphi$ into account, we have $|\mathcal{E}_2^{m+1,\varphi}(t,x)| = O(t^{m+1})$.

Appendix D. Proof of Lemma 3.11. By the integration by parts of Kusuoka–Stroock [17], we have

$$E[\partial^\alpha P_{T-t}f(\overline{X}_{(T/n)}(t,x))G] = E[P_{T-s}f(\overline{X}_{(T/n)}(t,x))H_\alpha(\overline{X}_{(T/n)}(t,x),G)].$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 03/28/19 to 130.70.8.131. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

For all $p > 1$,

$$\begin{aligned}&\left\|H_\alpha\left(\overline{X}_{(T/n)}(t,x),G\right)\right\|_p\\ &\le C(T,x)\left\|\det(\sigma^{\overline{X}_{(T/n)}(t,x)})^{-1}\right\|_{(8N+4)|\alpha|(|\alpha|+1)p}^{(|\alpha|+1)(2|\alpha|-1)}\|D\overline{X}_{(T/n)}(t,x)\|_{|\alpha|,(16N+8)N|\alpha|^2,H}^{2N|\alpha|(2|\alpha|-1)}\end{aligned}$$

by Corollary 3.7 of [17]. It holds that, for all $p > 1$,

$$\sup_{T/2\le t\le T}\|\det(\sigma^{\overline{X}_{(T/n)}(t,x)})^{-1}\|_p \le C(T,x)\frac{1}{T^N}$$

by Theorem 3.5 of [17]. Also, note that it holds that, for all $k \in \mathbb{N}$, $p > 1$,

$$\|D\overline{X}_{(T/n)}(t,x)\|_{k,p,H} \le C(T,x) \tag{40}$$

by Theorem 2.19 of [17]. Then, we obtain the assertion.

Appendix E. Proof of Corollary 3.14. Since we need to compute the case $(j,k) \in \mathbb{N}^2$ such that $1 \le j \le 4$, $1 \le k \le j$, and $|\gamma| \le 2$ for the weight in Theorem 3.9, we have the formula

$$\begin{aligned}&\widetilde{\mathcal{M}_2}(t,x,B_t)\\ &= \sum_{1\le k_1\le N} V_0^{k_1}(x)tH_{(k_1)}(\overline{X}(t,x),1) + \sum_{\substack{1\le i_1,i_2\le d\\ 1\le k_1,k_2,k_3\le N}} \hat{V}_{i_1}V_{i_2}^{k_1}(x)V_{i_1}^{k_2}(x)V_{i_2}^{k_3}(x)\frac{1}{2}t^2H_{(k_1,k_2,k_3)}(\overline{X}(t,x),1)\\ &+ \sum_{\substack{1\le i_1\le d\\ 1\le k_1,k_2\le N}} \{\hat{V}_0V_{i_1}^{k_1}(x)+\hat{V}_{i_1}V_0^{k_1}(x)\}V_{i_1}^{k_2}(x)\frac{1}{2}t^2H_{(k_1,k_2)}(\overline{X}(t,x),1)\\ &+ \sum_{1\le k_1\le N}\hat{V}_0V_0^{k_1}(x)\frac{1}{2}t^2H_{(k_1)}(\overline{X}(t,x),1) + \sum_{\substack{1\le i_1\le d\\ 1\le k_1\le N}}\hat{V}_{i_1}V_{i_1}^{k_1}(x)\frac{1}{2}tH_{(k_1)}(\overline{X}(t,x),1)\\ &+ \sum_{\substack{1\le i_1,i_2\le d\\ 1\le k_1,k_2\le N}} \{\hat{V}_{i_1}\hat{V}_{i_1}V_{i_2}^{k_1}(x)V_{i_2}^{k_2}(x)+\hat{V}_{i_1}\hat{V}_{i_2}V_{i_2}^{k_1}(x)V_{i_1}^{k_2}(x)\}\frac{1}{4}t^2H_{(k_1,k_2)}(\overline{X}(t,x),1)\\ &+ \sum_{\substack{1\le i_1\le d\\ 1\le k_1\le N}} \{\hat{V}_{i_1}\hat{V}_{i_1}V_0^{k_1}(x)+\hat{V}_0\hat{V}_{i_1}V_{i_1}^{k_1}(x)\}\frac{1}{4}t^2H_{(k_1)}(\overline{X}(t,x),1)\\ &+ \sum_{\substack{1\le i_1,i_2\le d\\ 1\le k_1\le N}}\hat{V}_{i_1}\hat{V}_{i_1}\hat{V}_{i_2}V_{i_2}^{k_1}(x)\frac{1}{8}t^2H_{(k_1)}(\overline{X}(t,x),1)\\ &+ \frac{1}{2}\sum_{\substack{1\le i_1,i_2\le d\\ 1\le k_1,k_2\le N}}\hat{V}_{i_1}V_{i_2}^{k_1}(x)\hat{V}_{i_1}V_{i_2}^{k_2}(x)\frac{1}{2}t^2H_{(k_1,k_2)}(\overline{X}(t,x),1)\\ &+ \frac{1}{2}\sum_{1\le k_1,k_2\le N}V_0^{k_1}(x)V_0^{k_2}(x)t^2H_{(k_1,k_2)}(\overline{X}(t,x),1)\\ &+ \sum_{\substack{1\le i_1\le d\\ 1\le k_1,k_2\le N}}V_0^{k_1}(x)\hat{V}_{i_1}V_{i_1}^{k_2}(x)\frac{1}{2}t^2H_{(k_1,k_2)}(\overline{X}(t,x),1)\\ &+ \frac{1}{2}\sum_{\substack{1\le i_1,i_2\le d\\ 1\le k_1,k_2\le N}}\hat{V}_{i_1}V_{i_1}^{k_1}(x)\hat{V}_{i_2}V_{i_2}^{k_2}(x)\frac{1}{4}t^2H_{(k_1,k_2)}(\overline{X}(t,x),1).\end{aligned}$$

Using notation $\tilde{V}_0$, $\bar{V}_0^j(x)$, $j = 1,\cdots,N$, we have the assertion.

Acknowledgments. I am very grateful to the associate editor and two anonymous referees for their valuable comments and suggestions. I thank Shigeo Kusuoka

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

and Shoichi Ninomiya for advice on the method and its related topics. I also thank Yuga Iguchi, Daisuke Kataoka, and Riu Naito for fruitful discussions on theoretical and numerical aspects of the scheme.

Downloaded 03/28/19 to 130.70.8.131. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

REFERENCES

[1] R. Avikainen, *On irregular functionals of SDEs and the Euler scheme*, Finance Stoch., 13 (2009), pp. 381–401.

[2] V. Bally and D. Talay, *The law of the Euler scheme for stochastic differential equations* I. *Convergence rate of the distribution function*, Probab. Theory Related Fields, 104 (1996), pp. 43–60.

[3] C. Bayer, P. Friz, and R. Loeffen, *Semi-closed form cubature and applications to financial diffusion models*, Quant. Finance, 13 (2013), pp. 769–782.

[4] D. Crisan, K. Manolarakis, and C. Nee, *Cubature Methods and Applications*, Paris-Princeton Lectures on Mathematical Finance, Springer, New York, 2013, pp. 203–316.

[5] K. Debrabant and A. Rössler, *On the acceleration of the multi-level Monte Carlo method*, J. Appl. Probab., 52 (2015), pp. 307–322.

[6] P. Friz and M. Hairer, *A Course of Rough Paths*, Springer, New York, 2014.

[7] M. Giles, *Multilevel Monte Carlo path simulation*, Oper. Res., 56 (2008), pp. 607–617.

[8] M. Giles, D. Higham, and X. Mao, *Analysing multi-level Monte Carlo for options with non-globally Lipschitz payoff*, Finance Stoch., 13 (2009), pp. 403–413.

[9] P. Glasserman, *Monte Carlo Method in Financial Engineering*, Springer, New York, 2004.

[10] L. G. Gyurkó and T. Lyons, *Efficient and practical implementations of cubature on Wiener space*, in Stochastic Analysis 2010, Springer, Heidelberg, 2011, pp. 73–111.

[11] L. G. Gyurkó and T. Lyons, *Rough paths based numerical algorithms in computational finance*, in Mathematics in Finance, Contemp. Math. 515, AMS, Providence, RI, 2010, pp. 17–46.

[12] N. Ikeda and S. Watanabe, *Stochastic Differential Equations and Diffusion Processes*, 2nd ed., North–Holland Math. Library 24, North–Holland, Amsterdam, 1989.

[13] P. E. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations*, Springer, New York, 1999.

[14] S. Kusuoka, *Approximation of expectation of diffusion process and mathematical finance*, Adv. Stud. Pure Math., 31 (2001), pp. 147–165.

[15] S. Kusuoka, *Malliavin calculus revisited*, J. Math. Sci. Univ. Tokyo, 10 (2003), pp. 261–277.

[16] S. Kusuoka, *Approximation of expectation of diffusion processes based on Lie algebra and Malliavin calculus*, in Advances in Mathematical Economics, Adv. Math. Econ. 6, Springer, Tokyo, 2004, pp. 69–83.

[17] S. Kusuoka and D. Stroock, *Applications of the Malliavin calculus.* I, in Stochastic Analysis (Katata/Kyoto 1982), K. Itô, ed., North–Holland Math. Library 32, North–Holland, Amsterdam, 1984, pp. 271–306.

[18] S. Kusuoka and D. Stroock, *Applications of the Malliavin calculus.* II, J. Fac. Sci. Univ. Tokyo Sect IA Math., 32 (1985), pp. 1–76.

[19] S. Kusuoka and D. Stroock, *Applications of the Malliavin calculus.* III, J. Fac. Sci. Univ. Tokyo Sect IA Math., 34 (1987), pp. 391–442.

[20] C. Litterer and T. Lyons, *High order recombination and an application to cubature on Wiener space*, Ann. Appl. Probab., 22 (2012), pp. 1301–1327.

[21] T. Lyons, M. Caruana, and T. Lévy, *Differential Equations Driven by Rough Paths*, Springer, New York, 2006.

[22] T. Lyons and N. Victoir, *Cubature on Wiener space*, Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci., 460 (2004), pp. 169–198.

[23] P. Malliavin, *Stochastic Analysis*, Springer, New York, 1997.

[24] P. Malliavin and A. Thalmaier, *Stochastic Calculus of Variations in Mathematical Finance*, Springer, New York, 2006.

[25] G. Maruyama, *Continuous Markov processes and stochastic equations*, Rend. Circ. Mat. Palermo, 4 (1955), pp. 48–90.

[26] S. Ninomiya and N. Victoir, *Weak approximation of stochastic differential equations and application to derivative pricing*, Appl. Math. Finance, 15 (2008) pp. 107–121.

[27] D. Nualart, *The Malliavin Calculus and Related Topics*, Springer, New York, 2006.

[28] A. Takahashi, *Asymptotic expansion approach in finance*, in Large Deviations and Asymptotic Methods in Finance, P. Friz et al., eds., Springer Proc. Math. Stat. 110, Springer, Cham, 2015, pp. 345–411.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

[29] A. Takahashi and T. Yamada, *An asymptotic expansion with push-down of Malliavin weights*, SIAM J. Financial Math., 3 (2012), pp. 95–136, https://doi.org/10.1137/100807624.

[30] A. Takahashi and T. Yamada, *A weak approximation with asymptotic expansion and multidimensional Malliavin weights*, Ann. Appl. Probab., 26 (2016), pp. 818–856.

[31] D. Talay, *Efficient Numerical Schemes for the Approximation of Expectations of Functionals of the Solution of an SDE and Applications*, in Filtering and Control of Random Processes, Lecture Notes in Control and Inform. Sci. 61, Springer, Berlin, 1984, pp. 294–313.

[32] S. Watanabe, *Analysis of Wiener functionals (Malliavin calculus) and its applications to heat kernels*, Ann. Probab., 15 (1987), pp. 1–39.

[33] T. Yamada, *A higher order weak approximation of multidimensional stochastic differential equations using Malliavin weights*, J. Comput. Appl. Math., 321 (2017), pp. 427–447.

[34] T. Yamada and K. Yamamoto, *A second order discretization with Malliavin weight and quasi-Monte Carlo method for option pricing*, Quantitative Finance, (published Online First) 2018.

Downloaded 03/28/19 to 130.70.8.131. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.