The Annals of Applied Probability
2014, Vol. 24, No. 2, 652–678
DOI: 10.1214/13-AAP932

SECOND ORDER DISCRETIZATION OF BACKWARD SDES AND SIMULATION WITH THE CUBATURE METHOD[1]

BY DAN CRISAN AND KONSTANTINOS MANOLARAKIS[2]

Imperial College London

We propose a second order discretization for backward stochastic differential equations (BSDEs) with possibly nonsmooth boundary data. When implemented, the discretization method requires essentially the same computational effort with the Euler scheme for BSDEs of Bouchard and Touzi [*Stochastic Process. Appl.* **111** (2004) 175–206] and Zhang [*Ann. Appl. Probab.* **14** (2004) 459–488]. However, it enjoys a second order asymptotic rate of convergence, provided that the coefficients of the equation are sufficiently smooth. In the second part of the paper, we combine this discretization with higher order cubature formulas on Wiener space to produce a fully implementable second order scheme.

1. Introduction. The present paper is concerned with the problem of numerical approximation to forward backward stochastic differential equations (FBSDEs henceforth). Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space on which we have defined a triple of processes (X, Y, Z) which solve the decoupled forward-backward system

$$
\begin{aligned}
X_t &= x_0 + \int_0^t V_0(X_u)\,du + \sum_{j=1}^d \int_0^t V_j(X_u) \circ dB_u^j, \qquad t \in [0, T], \\
Y_t &= \Phi(X_T) + \int_t^T f(X_u, Y_u, Z_u)\,du - \sum_{j=1}^d \int_t^T Z_u^j\,dB_u^j, \qquad t \in [0, T],
\end{aligned}
\tag{1}
$$

where B is a d-dimensional Brownian motion and

$$
V_k : \mathbb{R}^q \to \mathbb{R}^q, \qquad k = 0, \dots, d,\ f : \mathbb{R}^q \times \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}
$$

are some appropriate functions. The stochastic integrals appearing in the equation for the process X are understood in the Stratonovich sense, whilst the stochastic integrals appearing in the equation for the process Y are understood in Itô sense. The system is called decoupled as the (backward) processes (Y, Z) do not appear

Received April 2012; revised March 2013.

[1]The results in this paper were announced by the second author at the workshop "New advances in backward SDEs for financial engineering applications," Tamerza (Tunisia), October 25–28, 2010.

[2]Supported in part by EPSRC Grant EP/H0005500/1.

MSC2010 subject classifications. Primary 60H10; secondary 60H35.

Key words and phrases. Backward SDEs, second order discretization, cubature methods, numerical analysis.

in the dynamics of the forward component X. Systems of form (1) have received a lot of attention over the past twenty years, primarily due to their applications in the field of mathematical finance.

Of equal importance is the fact that the stochastic flow associated with (1), that is, the triple of processes $(X^{(t,x)}, Y^{(t,x)}, Z^{(t,x)})$, $(t,x) \in [0,T] \times \mathbb{R}$ satisfying

$$
\begin{aligned}
X_s^{t,x} &= x + \int_t^s V_0(X_u^{t,x})\,du + \sum_{j=1}^d \int_t^s V_j(X_u^{t,x}) \circ dB_u^j, \qquad s \in [t,T],\\
Y_s^{t,x} &= \Phi(X_T^{t,x}) + \int_s^T f(X_u^{t,x}, Y_u^{t,x}, Z_u^{t,x})\,du - \int_t^T Z_u^{t,x} \cdot dB_u, \\
&\qquad s \in [t,T],
\end{aligned} \tag{2}
$$

provides a Feynman–Kac representation for the (viscosity) solution of a class of semi-linear partial differential equations. In particular, let $u(t,x) : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$, be the viscosity solution of

$$
\begin{aligned}
&\frac{du}{dt} + \tilde{V}(x) \cdot \nabla u + \frac{1}{2} \operatorname{Tr}[V(x)V^*(x)D^2 u] + f(t,x,u,\nabla u V(x)) = 0,\\
&u(T,x) = \Phi(x),
\end{aligned} \tag{3}
$$

where $\tilde{V}(x) = V_0(x) - \frac{1}{2}\sum_{j=1}^d \nabla V_j(x) V_j(x)$. In their seminal work, Pardoux and Peng [25] showed that

$$
u(t,x) = Y_t^{t,x} \qquad (t,x) \in [0,T] \times \mathbb{R}^d. \tag{4}
$$

In addition, Ma and Zhang [24] proved that when this viscosity solution is continuously differentiable in its spatial variables, we have the following representation for the solution of the partial differential equation (3) and its gradient:

$$
u(t,x) = Y_t^{t,x}, \qquad Z_t^{t,x} = \nabla u(t,x) V(x) \qquad \text{a.s.} \tag{5}
$$

From the perspective of numerical analysis, the above means that any numerical method for the resolution of (1) provides an algorithm for the resolution of semi-linear PDEs. As a result, there is a high interest in robust numerical algorithms for FBSDEs.

To the best of our knowledge, the first attempt for the numerical resolution of a BSDE is due to Chevance [6], where the approach is to approximate the driving noise by some discrete time process. Similarly, in Ma et al. [23] the authors propose a discrete scheme for the resolution of a BSDE based on an Euler discretization and approximation of the driving Brownian noise by a random walk. They show convergence of their scheme in a weak sense, and no convergence rates are obtained. On the other hand, Daglas, Ma and Protter [12] suggest a four-step scheme for the numerical resolution of FBSDEs which can be more strongly coupled that the ones we consider in this paper. Their method is based on a finite difference approximation of the associated PDE. We should also mention the work of Bally and Pagès [1, 2] who propose an algorithm that relies on quantization for the numerical resolution of reflected BSDEs when the driver is independent of Z.

However, it was the work of Bouchard and Touzi [5] and, in parallel, Zhang [26] [essentially scheme (6) below] that first suggested a method for the discretization of a fairly general class of BSDEs and paved the way for algorithms of probabilistic flavor. Following this work, algorithms designed to solve the problem of numerical approximation of the backward part of (1) consist of two parts: First, the backward equation is discretized. This step involves the use of one or more conditional expectations. Second, a numerical method is grafted onto the chosen discretization to compute the conditional expectations involved. To date, the method of choice for the discretization of a BSDE is the Euler scheme proposed in its explicit format by Zhang [26] and implicit by Bouchard and Touzi [5].[3] In its implicit form it reads as

$$
\begin{aligned}
&Y_{1,n}^{\pi} := \Phi(X_{t_n}), \qquad Z_{1,n}^{\pi} := 0,\\
&Z_{1,i}^{\pi} := \frac{1}{\delta_{i+1}} \mathbb{E}\big[Y_{1,i+1}^{\pi} \Delta B_{i+1} | \mathcal{F}_{t_i}\big], \qquad i = 0, \ldots, n-1, \\
&Y_{1,i}^{\pi} := \mathbb{E}\big[Y_{1,i+1}^{\pi} | \mathcal{F}_{t_i}\big] + f\big(t_i, X_{t_i}, Y_{1,i}^{\pi}, Z_{1,i}^{\pi}\big), \qquad i = 0, \ldots, n-1,
\end{aligned}
\tag{6}
$$

where π is a given partition $\pi := \{0 = t_0 < t_1 < \cdots < t_n = T\}$ of $[0, T]$, δ_{i+1} is the time step $\delta_{i+1} := t_{i+1} - t_i$ and $\Delta B_{i+1} := B_{t_{i+1}} - B_{t_i}$, $i = 0, \ldots, n-1$ is the Brownian increment.

Assuming that all coefficients of (1) are at least Lipschitz continuous in their spatial variables, we have, following [5] and [26], that

$$
\sup_{0 \le t \le T} \mathbb{E}\big[|Y_t^{\pi} - Y_t|^2\big] + \mathbb{E}\left[\int_0^T |Z_t^{\pi} - Z_t|^2 \, dt\right] \le C|\pi|,
\tag{7}
$$

where $\{(Y_t^{\pi}, Z_t^{\pi}), t \ge 0\}$ are the step processes

$$
\begin{aligned}
Y_t^{\pi} &:= \sum_{i=0}^{n-1} Y_{1,i}^{\pi} 1_{[t_i, t_{i+1})}(t) + Y_{1,n}^{\pi} 1_{t=t_n}, \\
Z_t^{\pi} &:= \sum_{i=0}^{n-1} Z_{1,i}^{\pi} 1_{[t_i, t_{i+1})}(t) + Z_{1,n}^{\pi} 1_{t=t_n}.
\end{aligned}
\tag{8}
$$

In other words, the above discretization of the backward part achieves a convergence of order $n^{-1/2}$, when n points are used, that is, the same order as the strong convergence order of the Euler scheme for a (classical) SDE. With additional smoothness assumptions imposed on the coefficients, Gobet and Labart [13]

[3] See Bender and Zhang [4] for an alternative approach based on Picard iterations.

that the rate of convergence of the processes (8) is of order n^{-1}. In fact an error expansion is obtained in [13] and the leading order coefficients in the error expansion are identified.

Our first and most important goal in this paper is to introduce a discretization method for FBSDEs, in the spirit of (6), that achieves second order accuracy; that is, it enjoys an asymptotic rate of convergence of order n^{-2} when n points are used in the discretization of the time interval. In designing the second order scheme, we follow a similar template to (6). We will use the trapezoidal method to discretize the Riemann integral. That provides us with a second order discretization of the finite variation part of Y. In order to recover a second order approximation for Z_t, we compute a Brownian weight, that when multiplied to $Y_{t_{i+1}}$ and conditioned with respect to $\mathcal{F}_{t_i}$ cancels out all lower order terms. Let us note that the heuristic as well as detailed arguments that lead to the second order scheme rely on the Stratonovich–Taylor expansion of $u(t, X_t)$; see Section 2. As a result they depend on the smoothness of the solution $u(t,x)$ of PDE (3). The latter will generally be smooth, provided that the driver and coefficients of the forward SDE are smooth enough, even when the boundary condition Φ is not smooth. To treat such anomalies, if any, at the boundary T, our scheme takes the first backward step using the Euler scheme as in (6) and after leaving the boundary, continues with higher order discrete time steps. More precisely, assuming that all coefficients involved in system (1) are sufficiently smooth, we propose the following discretization: $(Y^\pi_{2,i}, Z^\pi_{2,i})_{0\le i\le n}$:

• *Initialization.*

If Φ is Lipschitz continuous,

$$Y^\pi_{2,n} := \Phi(X_n), \qquad Z^\pi_{2,n} := 0 \quad \text{and} \quad Z^\pi_{2,n-1} := Z^\pi_{1,n-1}, \qquad Y^\pi_{2,n-1} := Y^\pi_{1,n-1}.$$

If Φ is smooth,

$$Y^\pi_{2,n} := \Phi(X_n), \qquad Z^\pi_{2,n} := \nabla\Phi(X_n)V(X_n).$$

• *Backward induction.*

$$\mathcal{Z}^l_i = 4\frac{W_{t_{i+1}} - W_{t_i}}{\delta_{i+1}} - 6\frac{\int_{t_i}^{t_{i+1}}(s-t_i)\,dB^l_s}{\delta^2_{i+1}}, \qquad l = 1,\dots,d,$$

$$\begin{aligned} Z^\pi_{2,i} &= \mathbb{E}_i\big[\big(Y^\pi_{2,i+1} + \delta_{i+1} f\big(X_{i+1}, Y^\pi_{2,i+1}, Z^\pi_{2,i+1}\big)\big)\mathcal{Z}_i\big], \\ &\qquad\qquad\qquad\qquad \mathcal{Z}_i := \big(\mathcal{Z}^1_i,\dots,\mathcal{Z}^d_i\big)^T, \end{aligned} \tag{9}$$

$$Y^\pi_{2,i} = \mathbb{E}_i\big[Y^\pi_{2,i+1}\big] + \frac{\delta_{i+1}}{2}\big(f\big(X_i, Y^\pi_{2,i}, Z^\pi_{2,i}\big) + \mathbb{E}_i\big[f\big(X_{i+1}, Y^\pi_{2,i+1}, Z^\pi_{2,i+1}\big)\big]\big).$$

Under appropriate condition on the coefficients of system (1), we show in Theorem 3.4 and Corollary 3.7 that

$$\sup_{0\le i\le n} \mathbb{E}\bigg[\big|Y_{t_i} - Y^\pi_{2,i}\big|^2 + \frac{1}{n}\big|Z_{t_i} - Z^\pi_{2,i}\big|^2\bigg]^{1/2} \le \frac{C}{n^2}.$$

In particular, there is a detailed discussion of how one should choose the points in the time partition, when the boundary condition is not smooth (see Corollary 3.7).

There are two advantages of scheme (9) over the usual Euler scheme that need to be underlined. First, observe that given a partition π of the time interval $[0, T]$, the second order scheme approximates Y_t, Z_t at exactly the same points in the partition as the Euler scheme; namely, it does not require any intermediate points. This means that the two schemes are of the same complexity up to a constant. Second, the scheme is derivative free. This not only saves us the overhead of computing the derivatives of the coefficients, which can be costly, particularly when these coefficients come from calibration procedures (compare with the Milstein scheme for forward SDEs), but also allows us to use the scheme even in cases where the coefficients are not smooth. One can easily adapt the arguments of the proof of Theorem 3.4 to show that when the coefficients of system (1) are assumed only Lipschitz continuous, scheme (9) still converges but with the same order as the Euler type scheme (6), that is,

$$\sup_{0\le i\le n} \mathbb{E}\big[\big|Y_{t_i} - Y^{\pi}_{2,i}\big|^2\big]^{1/2} \le \frac{C}{\sqrt{n}}.$$

Hence (9) *can be used in all circumstances with no significant overhead. If the corresponding PDE enjoys a smooth solution* (*at least away from the boundary*), *the approximation converges quadratically. Otherwise, it will still converge at least as fast as the Euler scheme*.

Having discretized the backward equation, one needs to employ a method for the approximation of the involved conditional expectations.[4] Various such methods have been introduced, based on Malliavin calculus [5], on projection on function basis [14, 19] and on quantization [1]. In [10], the authors suggested the application of the cubature method of [22], which is based on the ideas of Kusuoka [17]. The overall rate of convergence of this second approximation step is still of order $n^{-1/2}$ or of order n^{-1} when coefficients are smooth.

In the second part of our paper, we bring together the second order discretization (9) with the cubature and TBBA algorithm presented in [10]. The latter is a combination of two algorithms put together to provide an efficient method for the approximation of nested conditional expectations. In particular, as a first ingredient we have the cubature on Wiener space method of Lyons and Victoir [22], that constructs explicit and discretely supported measures that approximate the law of forward diffusions. In effect integrals against this law, such as conditional expectations, can be approximated by integrating against this discrete measure. The resulting algorithm enjoys a second order of convergence. However, in practice the measures constructed with cubature have an exponentially growing support,

[4]When the process X in (6) is replaced by its Euler approximation, then $\mathbb{E}[Y^{\pi}_{1,i+1}|\mathcal{F}^{\pi}_{t_i}]$ and $\mathbb{E}[Y^{\pi}_{1,i+1}\Delta B_{i+1}|\mathcal{F}^{\pi}_{t_i}]$ need to be computed.

a growth that needs to be controlled. To the best of our knowledge, there are two different algorithms that can achieve this. The recombination algorithm of Litterer and Lyons [21] and the tree based branching algorithm (TBBA) of Crisan and Lyons [9]. We will use the latter as a second ingredient for our algorithm. The application of cubature and TBBA in BSDEs, in conjunction with the Euler scheme (6), has already been presented by the authors in [10]. As our second order scheme follows a similar template with the Euler method, that is, the approximating processes are defined in a backward manner by means of conditional expectations, the application of cubature and TBBA in the current framework follow exactly the steps presented in [10].

This paper has benefited from the very careful reading of an anonymous referee. Among many constructive comments and suggestions, the referee brought to our attention the work of Li, Zhang and Zhao [27] were the authors suggest a second order discretization for backward SDEs that shares many features with the algorithm we present in the first part of the paper. However, the work in Li, Zhang and Zhao [27] requires that the underlying diffusion is a Brownian motion rather than a general diffusion. This is a significant advantage as the stochastic Taylor expansions of the involved functionals are greatly simplified. The reason is that the lower order cross derivative terms cancel out as the corresponding differential operators commute [one should consider the computations in (21), (22) when the underlying noise is simply a Brownian motion to appreciate the differences].

The paper is organized as follows: In Section 2 we present our main assumptions and give details on Stratonovich–Taylor expansions, which will be our main tool in the construction and analysis of our algorithm. We then proceed in Section 3 to present the new second order discretization. In Section 4 we give the details of cubature and TBBA method and explain how one should couple this method with the second order discretization. Note that here we also present cubature formulas supported on Lie polynomials and not just paths as in [10]. Though the two approaches are equivalent, we discuss the advantages of Lie polynomials, and we also discuss the details of the implementation. We then analyze the asymptotic convergence of the algorithm, and finally, we present a one-dimensional example where we compare the new discretization with the Euler scheme and validate its second order convergence properties.

2. Preliminaries. Throughout the paper, given a positive integer m, we will use the following assumptions:

(A) The coefficients of the forward SDE $V_i:\mathbb{R}^d \to \mathbb{R}^d$, $i = 0, 1, \ldots, d$ have all entries belonging to $C_b^\infty(\mathbb{R}^d)$, the space of bounded infinitely differentiable functions with all partial derivatives bounded.

We make no assumptions at this point on the ellipticity (or lack of), for the diffusion matrix. We will come back to this in the discussion on the PDE gradient estimates necessary in the derivation of the general convergence results; see Corollary 3.7 and the discussion preceding it.

(B(m)) The driver of the BSDE $f:[0,1]\times\mathbb{R}^d\times\mathbb{R}\times\mathbb{R}^d\to\mathbb{R}$ belongs to $C_b^{[m/2],m}$. The exact value for the parameter m shall be made precise as we proceed.

(C1) The terminal condition $\Phi:\mathbb{R}^d\to\mathbb{R}$ is Lipschitz continuous.

(C2(m)) The terminal condition Φ belongs to $C_b^m(\mathbb{R}^d;\mathbb{R})$, where value of m shall be determined further on.

We shall denote by K the constant that bounds all derivatives that appear in our assumptions. When (C1) is in force we shall assume that Φ is K-Lipschitz. Of course under (A), (B(m)) with $m\geq 1$ and (C1) system (1) has a unique solution such that

$$\mathbb{E}\left[\sup_{0\leq t\leq T}\left(Y_t^2+|X_t|^2\right)+\int_0^T |Z_s|^2\,ds\right]<\infty.$$

To abbreviate notation we shall denote by $M_i\equiv M_{t_i}, i=0,\ldots,n$, where M can be any of the processes that appear in this paper. Moreover we write $\mathbb{E}_s[\cdot]$ for $\mathbb{E}[\cdot|\mathcal{F}_s]$. Note that in the current (Markovian) set up $\mathbb{E}[\cdot|\mathcal{F}_s]=\mathbb{E}[\cdot|X_s]$. We shall also consider as given a partition $\pi:=\{0=t_0<t_1<\cdots<t_n=T\}$ of $[0,T]$ and consider the time step intervals $\delta_i:=t_i-t_{i-1}, l=1,\ldots,d, i=1,\ldots,n$.

Finally, the driver of the BSDE shall be abbreviated as

$$\bar{f}(s,x)=f\bigl(x,u(s,x),\nabla uV(s,x)\bigr),$$

where u is the solution of the semilinear PDE (3).

Working toward a higher order discretization of the backward part of (1) we shall rely heavily on the Stratonovich–Taylor expansions. Hence we need to fix notation and present some elementary facts regarding these expansions:

We denote by $\mathcal{A}$ the set of multi-indices $\mathcal{A}:=\{\varnothing\}\cup\bigcup_{m=1}^{\infty}\{0,1,\ldots,d\}^m$ endowed with the norms $|\cdot|$ and $\|\cdot\|$ given by

$$|\beta|=\text{length of }\beta,\qquad \|\beta\|:=|\beta|+\operatorname{Card}\bigl\{j:\beta_j=0, 1\leq j\leq|\beta|\bigr\}.$$

Clearly $|\varnothing|=\|\varnothing\|=0$. For a $\beta=(j_1,\ldots,j_l)\in\mathcal{A}$ we also write $\beta-=(j_1,\ldots,j_{l-1})$ and $-\beta=(j_2,\ldots,j_l)$. We define the subsets of $\mathcal{A}$,

$$\mathcal{A}(m)=\bigl\{\beta\in\mathcal{A}:\|\beta\|\leq m\bigr\}\quad\text{and}$$
$$\mathcal{A}^1(m)=\bigl\{\beta\in\mathcal{A}\setminus\{\varnothing,(0)\}:\|\beta\|\leq m\bigr\}.$$

Given two multi-indices $\alpha=(\alpha_1,\ldots,\alpha_k)$ and $\beta=(\beta_1,\ldots,\beta_l)$, we define their concatenation as $\alpha*\beta=(\alpha_1,\ldots,\alpha_k,\beta_1,\ldots,\beta_l)$. For a suitably chosen function f and a multi-index $\beta=(\beta_1,\ldots,\beta_l)$, we define the iterated Stratonovich/Riemann integrals as follows:

$$J^\beta[f]_{t,s}:=\begin{cases} f(s), & |\beta|=0,\\ \displaystyle\int_t^s J^{\beta-}[f]_{t,u}\,du, & l\geq 1, j_l=0,\\ \displaystyle\int_t^s J^{\beta-}[f]_{t,u}\circ dB^{j_l}(u), & l\geq 1, j_l\neq 0.\end{cases}$$

We also use the abbreviation $J^{\alpha}_{s,t} \equiv J^{\alpha}[\mathbf{1}]_{s,t}$ where $\mathbf{1}$ is the constant function $\mathbf{1}: r \to 1$ for all $r \in [s,t]$.

For any vector field $V \in C_b^{\infty}(\mathbb{R}^d)$ we make the usual identification with the first order differential operator

$$Vf(x) = \sum_{j=1}^{d} V_j(x)\partial_{x_j} f(x). \tag{10}$$

In particular we consider the differential operators $V_1, \dots, V_d$ and V_0.

The iteration of this family of operators is understood as $V_\alpha f := V_{\alpha_1} \cdots V_{\alpha_n} f$, $\alpha = (\alpha_1, \dots, \alpha_n)$, and we also use the convention $V_\varnothing f = f$.

Given a multi index $\alpha = (\alpha_1, \dots, \alpha_n)$ the following identity is proven in Proposition 5.2.10 of [16]:

$$B^j(t) J^{\alpha}_{0,t} = \sum_{k=0}^{n+1} J_{0,t}^{(\alpha_1,\dots,\alpha_k,j,\alpha_{k+1},\alpha_n)}, \qquad j = 1, \dots, d. \tag{11}$$

A direct consequence of (11) (Corollary 5.2.11 of [16]) is that if $\alpha = (k, k, \dots, k)$, $k = 0, \dots, d$, with $|\alpha| = m$, then

$$J^{\alpha}_{0,t} = \frac{1}{m!}\big(J^{(k)}_{0,t}\big)^k. \tag{12}$$

The (conditional) expectations of iterated integrals can also be computed. In particular we have a characterization for those iterated integrals that have nonzero expectation (see, e.g., [8]):

LEMMA 2.1. *Let $\alpha = (i_1, \dots, i_r)$ be an arbitrary multi-index with $\|\alpha\| = m$ and $t \in [0,T]$. If m is odd, then $\mathbb{E}[J^{\alpha}_{0,t}] = 0$, and if m is even, then*

$$\mathbb{E}[J^{\alpha}_{0,t}] = \begin{cases} \dfrac{t^{m/2}}{2^{r-m/2}(m/2)!}, & \text{if } \alpha \in \mathcal{A}(m,r), \\ 0, & \text{otherwise,} \end{cases}$$

where $A(m,r) \subset \mathcal{A}(m)$ is the set of multi-indices with $\alpha = \alpha_1 * \cdots * \alpha_{m/2} \in \mathcal{A}(m)$, such that $\alpha_i = (0)$ or $\alpha_i = (j,j)$, $j \in \{1, \dots, d\}$.

A nonempty subset $\mathcal{G} \subseteq \mathcal{A}$ is called a hierarchical set if $\sup_{\alpha \in \mathcal{G}} |\alpha| < \infty$ and $-\alpha \in \mathcal{G}$ for any $\alpha \in \mathcal{G} \setminus \{\varnothing\}$. Given a hierarchical set $\mathcal{G}$ we define the remainder set

$$\mathcal{B}(\mathcal{G}) = \{\beta \in \mathcal{A} \setminus \mathcal{G} : -\beta \in \mathcal{G}\}.$$

Note that $\mathcal{A}(m)$ and $\mathcal{A}^1(m)$ are hierarchical sets. If $g : \mathbb{R}^d \to \mathbb{R}$ is a smooth function with all partial derivatives bounded, then by repeatedly applying the Itô–Stratonovich formula, we obtain the Stratonovich–Taylor expansion

$$g\big(t, X_t^{s,x}\big) = \sum_{\alpha \in \mathcal{K}} V_\alpha g(0,x) J^{\alpha}_{s,t} + \sum_{\alpha \in \mathcal{B}(\mathcal{K})} J^{\alpha}\big[V_\alpha g\big(\cdot, X_\cdot^{s,x}\big)\big]_{s,t}, \qquad s \le t$$

for any hierarchical set $\mathcal{K}$. The second term on the right-hand side of the previous equation is called the remainder process. In particular, when the stochastic Taylor expansion is applied with respect to the hierarchical set $\mathcal{A}(m), m \in \mathbb{N}_+$, we shall denote the remainder by $R_m(0, t, g)$, that is,

$$g(t, X_t^{0,x}) = \sum_{\alpha \in \mathcal{A}(m)} V_\alpha g(0, x) J_{0,t}^\alpha + R_m(0, t, g). \tag{13}$$

Obviously as m increases, and for small t, the remainder gets smaller and smaller. In this paper we shall work only with stochastic expansions with respect to the hierarchical sets $\mathcal{A}(m), m \in \mathbb{N}_+$. For this set we have that

$$\mathcal{B}(\mathcal{A}(m)) = \{(j) \star \beta | j = 0, \dots, d, \|\beta\| = m\}.$$

To estimate the remainder that corresponds to $\mathcal{B}(\mathcal{A}(m))$, we need to estimate iterated Stratonovich integrals of the form $J^\alpha[g(\cdot, X_\cdot)]_{s,t}$ for appropriate function g such that the previous integral makes sense. For $0 \le s \le t$ with $t - s < 1$, we have

$$\begin{aligned} &\Big|\sup_{s \le r \le t} \mathbb{E}_s\big[J^\alpha[g(\cdot, X_\cdot)]_{s,r}\big]\Big| \\ &\quad \le C(t-s)^{\|\alpha\|/2}\Big(\sup_{s \le r \le t}\{\mathbb{E}_s[|V_{\alpha_n} g(r, X_r)| 1_{\alpha_n \neq 0}] \\ &\qquad + \mathbb{E}_s[|g(r, X_r)| 1_{\alpha_n = 0}]\}\Big). \end{aligned} \tag{14}$$

Proving estimate (14) amounts to a tedious induction, based on Itô's formula, similar to the proof of Lemma 5.7.2 of Chapter 5 of Kloeden and Platen [16]. Applying estimate (14) to every index in the set $\mathcal{B}(\mathcal{A}(m))$ provides us with the following estimate for the remainder of the Taylor formula for any $s \le t$:

$$\begin{aligned} &\mathbb{E}[|R_m(s, t, f)|^p]^{1/p} \\ &\quad \le C(t-s)^{(m+1)/2} \max_{m \le \|\gamma\| \le m+2} \sup_{s \le r \le t} \mathbb{E}[|V_\gamma f(r, X_r)|^p]^{1/p}. \end{aligned} \tag{15}$$

If the function g depends on time as well, $g \equiv g(t, x)$, then one of course may still apply the Stratonovich–Taylor expansion to obtain a similar result to (13). Differentiation with respect to time can be absorbed by the V_0 operator, namely $V_0 \equiv \partial_t + \sum_{i=1}^d V_0^i \partial_{x_i}$, since in applying the Itô formula, derivatives with respect to t or V_0, scale similarly; namely, they produce terms of order t. In effect estimate (15) remains valid. In the next section, we use Taylor–Stratonovich expansions extensively in order to design a second order discretization for (1).

3. One-step discretization. To understand the terms that a second order scheme should incorporate, we present first the intuitive arguments that lead to the discretization scheme (9).

REMARK 3.1. We will start by assuming perfect knowledge of the forward diffusion along a given partition, that is, of $(X_{t_1}, \dots, X_{t_n})$.

In the second part of the paper, we approximate the law of this multi dimensional random vector by means of cubature formulas without having to discretize its dynamics, and hence the second step will be consistent with the first.

For the benefit of methods that would combine the present discretization with a Monte Carlo simulation, that would most likely require some sort of discretization for the forward process X, we note that all results regarding the second order scheme (particularly Theorem 3.4 and Corollary 3.7) remain valid, once one replaces X with a second order approximation. A wealth of such approximations are presented in Chapters 12–15 of [16].[5]

We consider the backward part of (1) between two successive times of the partition π

$$Y_{t_i} = Y_{t_{i+1}} + \int_{t_i}^{t_{i+1}} f(X_s, Y_s, Z_s)\, ds - \int_{t_i}^{t_{i+1}} Z_s \cdot dB_s$$

and discretize the Riemann integral using the left-hand side point, as in [5], thus leading to an implicit equation for Y_{t_i} and the stochastic part in the usual way, to obtain

$$Y_{t_i} \simeq Y_{t_{i+1}} + \delta_{i+1} f(X_{t_i}, Y_{t_i}, Z_{t_i}) - Z_{t_i} \cdot \Delta B_{i+1}. \tag{16}$$

By conditioning (16) with respect to $\mathcal{F}_{t_i}$ in (16) we obtain a first order approximation for Y_{t_i}

$$Y_{t_i} \simeq \mathbb{E}[Y_{t_{i+1}} | \mathcal{F}_{t_i}] + \delta_{i+1} f(X_{t_i}, Y_{t_i}, Z_{t_i}), \tag{17}$$

but for the presence of Z_{t_i}. To treat the Z_{t_i}, we can multiply both sides of (16) by ΔB^l_{i+1}, $l = 1, \dots, d$ and condition with respect to $\mathcal{F}_{t_i}$, to obtain

$$Z^l_{t_i} \simeq \mathbb{E}\left[Y_{t_{i+1}} \frac{\Delta B^l_{i+1}}{\delta_{i+1}} \Big| \mathcal{F}_{t_i} \right], \qquad l = 1, \dots, d. \tag{18}$$

Observe that scheme (6) is just the backward iteration of equations (17) and (18). Another way of interpreting the approximation for Z (18), is via the Stratonovich–Taylor expansion. Recall that for every $l = 1, \dots, d$, Z^l_t is the directional derivative of the value function to the direction of V_l, that is, $Z^l_t = \sum_{i=1}^d V^i_l(x) \partial_{x_i} u(t, X_t)$. So if we apply expansion (13) using the hierarchical set $\mathcal{A}(2)$ and multiply by the Brownian increment and condition, we obtain

$$\begin{aligned}\mathbb{E}\left[Y_{t_{i+1}} \frac{\Delta B^l_{i+1}}{\delta_{i+1}} \Big| \mathcal{F}_{t_i} \right] &\equiv \mathbb{E}\left[u(t_{i+1}, X_{t_i+1}) \frac{\Delta B^l_{i+1}}{\delta_{i+1}} \Big| \mathcal{F}_{t_i} \right] \\ &= V_l u(t_i, X_{t_i}) + O(\delta_{i+1}).\end{aligned}$$

[5] Such approximations can be chosen to be completely derivative free.

Later on, we shall use this point of view to understand the type of terms that a second order approximation for Z should incorporate.

As far as the driver is concerned, the next elementary result tells us that the trapezoidal method indeed achieves a third order local error, thus leading to a second order global error:

LEMMA 3.2. *Let assumptions* (A) *and* (B(3)) *hold true*, *and assume that* $\Phi \in C_{\mathrm{Lip}}(\mathbb{R}^d)$. *Let us also assume that there exists a classical solution to PDE* (3) *on* $[0,T)\times\mathbb{R}^d$ *which we denote by* $u(t,x)$ *and also denote by* $\bar f(t,x) \equiv f(x,u(t,x),\nabla u(t,x)V(x))$. *Then there exists a constant* C *independent of the partition and of* u, *such that*

$$\left|\mathbb{E}_i\left[\int_{t_i}^{t_{i+1}} \bar f(s,X_s)\,ds - \frac{\delta_{i+1}}{2}\big(\bar f(t_i,X_{t_i}) + \bar f(t_{i+1},X_{t_{i+1}})\big)\right]\right| \le C\delta_{i+1}^3 \max_{\|\alpha\|=4,5}\|V_\alpha u(t_{i+1},\cdot)\|_\infty.$$

PROOF. Since we have assumed existence of a classical solution of the associated PDE, the nonlinear Feynman–Kac formula tells us that $Y_t = u(t,X_t)$ and $Z_t^l = V_l u(t,X_t)$, $l=1,\dots,d$. We then proceed by expanding the integrand of the Riemann integral using expansion (13), with the hierarchical set $\mathcal{A}(3)$,

$$\begin{aligned}\mathbb{E}_i\left[\int_{t_i}^{t_{i+1}} \bar f(s,X_s)\,ds\right] &= \delta_{i+1}\bar f(t_i,X_i)\\ &\quad + \mathbb{E}_i\left[\int_{t_i}^{t_{i+1}}\left(\sum_{\alpha\in\mathcal{A}_0(3)} V_\alpha \bar f(t_i,X_i) J^\alpha_{t_i,u}\right.\right.\\ &\quad \left.\left. + \sum_{\alpha\in\mathcal{A}(5)\setminus\mathcal{A}_0(3)} J^\alpha[V_\alpha \bar f(\cdot,X_\cdot)]_{t_i,u}\right)du\right]\end{aligned} \tag{19}$$

and to $\bar f(t_{i+1},X_{i+1})$,

$$\begin{aligned}&\frac{\delta_{i+1}}{2}\big(f(X_i,Y_i,Z_i) + \mathbb{E}_i[f(X_{i+1},Y_{i+1},Z_{i+1})]\big)\\ &\quad = \delta_{i+1}\bar f(t_i,X_i) + \frac{\delta_{i+1}}{2}\mathbb{E}_i\left(\sum_{\alpha\in\mathcal{A}_0(3)} V_\alpha\bar f(t_i,X_i)J^\alpha_{t_i,t_{i+1}}\right.\\ &\qquad \left. + \sum_{\alpha\in\mathcal{A}_0(5)\setminus\mathcal{A}_0(3)} J^\alpha[V_\alpha\bar f(\cdot,X_\cdot)]_{t_i,t_{i+1}}\right).\end{aligned} \tag{20}$$

If $\alpha\in\mathcal{A}_0(3)$, then one of the following holds:

$$\alpha = \begin{cases} (i), & i=0,\dots,d,\\ (0,i),(i,0), & i=1,\dots,d,\\ (i,j),(i,j,k) & i,j,k=1,\dots,d.\end{cases}$$

Elementary facts about stochastic integrals and Lemma 2.1 tell us that $\mathbb{E}_s[J^{\alpha}_{s,u}] = 0$, $u \geq s$ for all α's as above except when $\alpha = (0)$ or (j, j), $j = 1, \dots, d$. For these two cases, we have

$$\mathbb{E}_s[J^{(0)}_{s,u}] = u - s \quad \text{and} \quad \mathbb{E}_s[J^{(j,j)}_{s,u}] = \frac{1}{2}\mathbb{E}[(B^j_u - B^j_s)^2] = \frac{u-s}{2}.$$

Substituting the above into (19), (20) and taking their difference, we have

$$\begin{aligned}
&\left|\mathbb{E}_i\left[\int_{t_i}^{t_{i+1}} \bar{f}(s, X_s)\, ds - \frac{\delta_{i+1}}{2}(\bar{f}(t_i, X_{t_i}) + \bar{f}(t_{i+1}, X_{t_{i+1}}))\right]\right| \\
&\quad \leq \left|\mathbb{E}_i\left[\int_{t_i}^{t_{i+1}} \left(\sum_{\alpha \in \mathcal{A}_0(5)\setminus\mathcal{A}_0(3)} J^{\alpha}[V_\alpha \bar{f}(\cdot, X_\cdot)]_{t_i,u}\right) du \right.\right. \\
&\qquad\qquad \left.\left. - \frac{\delta_{i+1}}{2} \sum_{\alpha \in \mathcal{A}_0(5)\setminus\mathcal{A}_0(3)} J^{\alpha}[V_\alpha \bar{f}(\cdot, X_\cdot)]_{t_i,t_{i+1}}\right]\right|.
\end{aligned}$$

The estimates on the remainder process complete the proof. □

The above result dictates how one should treat the Riemann integral when working toward a second order scheme. Working toward a second order approximation for Z, we apply a Taylor expansion on the stochastic integral to obtain

$$\begin{aligned}
Y_{t_i} &= Y_{t_{i+1}} + \int_{t_i}^{t_{i+1}} \bar{f}(s, X_s)\, ds \\
&\quad - \sum_{l=1}^{d} \left\{ V_l u(t_i, X_{t_i}) \Delta B^l_{i+1} + \sum_{k=1}^{d} V_{(k,l)} u(t_i, X_{t_i}) \int_{t_i}^{t_{i+1}} J^{(k)}[1]_{t_i,s}\, dB^l_s \right. \\
&\qquad + \sum_{k,j=1}^{d} V_{(k,j,l)} u(t_i, X_{t_i}) \int_{t_i}^{t_{i+1}} J^{(k,j)}[1]_{t_i,s}\, dB^l_s \\
&\qquad + V_{(0,l)} u(t_i, X_{t_i}) \int_{t_i}^{t_{i+1}} (s - t_i)\, dB^l_s \\
&\qquad + \sum_{\|\alpha\|=3} V_{\alpha\star(l)} u(t_i, X_{t_i}) \int_{t_i}^{t_{i+1}} J^{\alpha}[1]_{t_i,s}\, dB^l_s \\
&\qquad \left. + \int_{t_i}^{t_{i+1}} R_3(t_i, s, V_l u)\, dB^l_s \right\}.
\end{aligned} \tag{21}$$

As in (18), we need to innovate a way to recover Z^l_i, $l = 1, \dots, d$, $i = 0, \dots, n-1$ from (21), but this time up to a second order error.

If we multiply both sides of (21) by $\frac{\Delta B_{i+1}}{\delta_{i+1}}$ and condition with respect to $\mathcal{F}_i$, we shall obtain $Z^l_{t_i}$, but some surviving terms of order δ_{i+1} will render the approximation first order. For example, consider

$$\frac{1}{\delta_{i+1}}\mathbb{E}\left[V_{(0,l)}u(t_i, X_{t_i})\int_{t_i}^{t_{i+1}} s\, dB^l_s \Delta B^l_{i+1}|\mathcal{F}_{t_i}\right] = \frac{\delta_{i+1}}{2}V_{(0,l)}u(t_i, X_{t_i}).$$

Hence, we need to find an appropriate weight which, when multiplying (21) and after conditioning with respect to $\mathcal{F}_{t_i}$, provides a second order approximation for Z_{t_i} by *canceling out all first order terms*. We make the following judicious choice:

$$\mathcal{Z}^l_i := \lambda_1 \frac{\Delta B^l_{i+1}}{\delta_{i+1}} + \lambda_2 \frac{J^{(0,l)}[1]_{t_i,t_{i+1}}}{\delta^2_{i+1}}, \qquad l = 1,\dots,d, i = 0,\dots,n-1. \tag{22}$$

With a few straightforward computations, we obtain for any $q = 1,\dots,d$,

$$\mathbb{E}\left[\sum_{l=1}^d V_l u(t_i, X_{t_i})\Delta B^l_{i+1}\mathcal{Z}^q_i \Big| \mathcal{F}_{t_i}\right] = \left(\lambda_1 + \frac{\lambda_2}{2}\right)V_q u(t_i, X_{t_i}),$$

$$\mathbb{E}\left[\sum_{l=1}^d V_{(0,l)}u(t_i, X_{t_i})\int_{t_i}^{t_{i+1}} s\, dB^l_s \mathcal{Z}^q_i \Big| \mathcal{F}_{t_i}\right] = \left(\frac{\lambda_1\delta_{i+1}}{2} + \frac{\lambda_2\delta_{i+1}}{3}\right)V_{(0,q)}u(t_i, X_{t_i}),$$

$$\mathbb{E}\left[\sum_{l=1}^d \sum_{k,j=1}^d V_{(k,j,l)}u(t_i, X_{t_i})\int_{t_i}^{t_{i+1}} J^{(k,j)}[1]_{t_i,s}\, dB^l_s \mathcal{Z}^q_i \Big| \mathcal{F}_{t_i}\right]$$

$$= \left(\frac{\lambda_1\delta_{i+1}}{4} + \frac{\lambda_2\delta_{i+1}}{6}\right)\sum_{k=1}^d V_{(k,k,q)}u(t_i, X_{t_i}),$$

$$\mathbb{E}\left[\sum_{l=1}^d \left(\sum_{k=1}^d \int_{t_i}^{t_{i+1}} J^{(k)}[1]_{t_i,s}\, dB^l_s + \sum_{\|\alpha\|=3}\int_{t_i}^{t_{i+1}} J^{\alpha}[1]_{t_i,s}\, dB^l_s\right)\mathcal{Z}^q_i \Big| \mathcal{F}_{t_i}\right] = 0.$$

By choosing $\lambda_1 = 4$, $\lambda_2 = -6$, we have the following:

LEMMA 3.3. *Let assumptions* (A), (B(m)) *hold true, and let* $u(t,x)$ *denote the classical solution of PDE* (3). *Set*

$$\mathcal{Z}^l_i := 4\frac{\Delta B^l_{i+1}}{\delta_{i+1}} - 6\frac{J^{(0,l)}[1]_{t_i,t_{i+1}}}{\delta^2_{i+1}}, \qquad l = 1,\dots,d, i = 0,\dots,n-2.$$

Then

$$\left|Z^l_{t_i} - \mathbb{E}_i\left[\left(Y_{t_{i+1}} + \delta_{i+1} f(X_{t_{i+1}}, Y_{t_{i+1}}, Z_{t_{i+1}})\right)\mathcal{Z}^l_i\right]\right| \le \delta^2_{i+1}\max_{\|\alpha\|=4,5}\|V_\alpha u(t_{i+1},\cdot)\|_\infty.$$

PROOF. We first apply a simple triangle inequality to

$$
\begin{aligned}
&\big|Z^l_{t_i} - \mathbb{E}_i\big[\big(Y_{t_{i+1}} + \delta_{i+1} f(X_{t_{i+1}}, Y_{t_{i+1}}, Z_{t_{i+1}})\big)\mathcal{Z}^l_i\big]\big| \\
&\quad = \bigg|Z^l_{t_i} \mp \mathbb{E}_i\bigg[\bigg(Y_{t_{i+1}} + \int_{t_i}^{t_{i+1}} f(X_s, Y_s, Z_s)\, ds\bigg)\mathcal{Z}^l_i\bigg] \\
&\qquad\qquad - \mathbb{E}_i\big[Y_{t_{i+1}} + \delta_{i+1} f(X_{t_{i+1}}, Y_{t_{i+1}}, Z_{t_{i+1}})\mathcal{Z}^l_i\big]\bigg|.
\end{aligned}
$$

Using the results of Lemma 2.1 and following the intuitive computations preceding Lemma 3.3, it is clear that

$$
\begin{aligned}
&\bigg|Z^l_{t_i} - \mathbb{E}_i\bigg[\bigg(Y_{t_{i+1}} + \int_{t_i}^{t_{i+1}} f(X_s, Y_s, Z_s)\, ds\bigg)\mathcal{Z}^l_i\bigg]\bigg| \\
&\qquad \le \bigg|\mathbb{E}_i\bigg[\mathcal{Z}^l_i \int_{t_i}^{t_{i+1}} R_3(t_i, s, V_l u)\, dB^l_s\bigg]\bigg|,
\end{aligned}
$$

where as applying a Stratonovich Taylor expansion to $f(X_s, Y_s, Z_s)$ and $f(X_{t_{i+1}}, Y_{t_{i+1}}, Z_{t_{i+1}})$ with the hierarchical set $\mathcal{A}(2)$ leads to

$$
\begin{aligned}
&\bigg|\mathbb{E}_i\bigg[\int_{t_i}^{t_{i+1}} f(X_s, Y_s, Z_s)\, ds\, \mathcal{Z}^l_i\bigg] - \mathbb{E}_i\big[\delta_{i+1} f(X_{t_{i+1}}, Y_{t_{i+1}}, Z_{t_{i+1}})\mathcal{Z}^l_i\big]\bigg| \\
&\qquad \le \bigg|\mathbb{E}_i\bigg[\mathcal{Z}^l_i \int_{t_i}^{t_{i+1}} R_3(t_i, s, \bar f)\, ds\bigg]\bigg|.
\end{aligned}
$$

The result then follows from estimate (15) on the remainder process and the Cauchy–Schwarz inequality. $\square$

Given the estimates of Lemmas 3.2 and 3.3, the rationale behind scheme (9) should now be clear. In the following Theorem, we provide the main estimate for the error of our scheme.

THEOREM 3.4. *Let assumptions* (A), (B(5)) *and either* (C1) *or* (C2) *hold true. Then there exists a constant* $C > 0$ *such that*

$$
\begin{aligned}
&\max_{0 \le i \le n-2} \bigg[|Y_{t_i} - Y^\pi_{2,i}|^2 + \frac{\delta_{i+1}}{4d} |Z_{t_i} - Z^\pi_{2,i}|^2\bigg] \\
&\qquad \le C\bigg(|Y_{t_{n-1}} - Y^\pi_{2,n-1}|^2 + \frac{\delta_n}{4d} |Z_{t_{n-1}} - Z^\pi_{2,n-1}|^2\bigg) \\
&\qquad\qquad + \sum_{i=1}^{n-1} \delta_i^5 \max_{\|\alpha\|=4,5} \|V_\alpha u(t_i, \cdot)\|^2_\infty.
\end{aligned}
$$

The above result shows that the approximation scheme is in general of order 2 modulus the error that we are making on the first backward step. The latter can be easily shown to be also of order 2 when the boundary function is smooth. In the more interesting case where Φ is merely Lipschitz continuous, our first backward step is done via the Euler scheme. In Corollary 3.7 we will see how a nonequidistant partition can compensate for this and provide a scheme with order of convergence 2.

PROOF OF THEOREM 3.4. In the following proof, C will denote a constant whose value might change from line to line. It will, however, be independent of the partition and of the bounds of the derivatives of the solution of (3). For ease of notation, we set

$$
\begin{aligned}
\Delta_i^\pi Y &:= Y_{t_i} - Y^\pi_{2,i}, \qquad \Delta_i^\pi Z := Z_{t_i} - Z^\pi_{2,i},\\
\Delta_i^\pi f &:= f(X_{t_i}, Y_{t_i}, Z_{t_i}) - f\big(X_{t_i}, Y^\pi_{2,i}, Z^\pi_{2,i}\big),\\
\Psi_{i+1} &:= Y_{t_{i+1}} + \frac{\delta_{i+1}}{2} f(X_{t_{i+1}}, Y_{t_{i+1}}, Z_{t_{i+1}}),\\
\Psi^\pi_{i+1} &:= Y^\pi_{2,i+1} + \frac{\delta_{i+1}}{2} f\big(X_{t_{i+1}}, Y^\pi_{2,i+1}, Z^\pi_{2,i+1}\big),\\
\Delta\Psi_{i+1} &= \Psi_{i+1} - \Psi^\pi_{i+1}.
\end{aligned}
$$

Let us fix a value for $i = 0, \dots, n-2$. We consider the difference of the solution of the BSDE at time t_i and of scheme (9),

$$
\begin{aligned}
\Delta_i^\pi Y = {}& \mathbb{E}_i[\Delta_{i+1}^\pi Y] + \frac{\delta_{i+1}}{2}\mathbb{E}_i\big[f\big(X_i, Y^\pi_{2,i}, Z^\pi_{2,i}\big) + f\big(X_{i+1}, Y^\pi_{2,i+1}, Z^\pi_{2,i+1}\big)\big]\\
&\mp \frac{\delta_{i+1}}{2}\mathbb{E}_i\big[\bar f(t_i, X_i) + \bar f(t_{i+1}, X_{i+1})\big] - \int_{t_i}^{t_{i+1}} \mathbb{E}_i\big[\bar f(s, X_s)\big]\,ds.
\end{aligned} \tag{23}
$$

According to the estimates of Lemma 3.2 we have that

$$
\begin{aligned}
&\big|\mathbb{E}_i\big[f\big(X_i, Y^\pi_{2,i}, Z^\pi_{2,i}\big) + f\big(X_{i+1}, Y^\pi_{2,i+1}, Z^\pi_{2,i+1}\big)\big]\\
&\qquad - \frac{\delta_{i+1}}{2}\mathbb{E}_i\big[\bar f(t_i, X_i) + \bar f(t_{i+1}, X_{i+1})\big]\big|\\
&\qquad\le C\delta_{i+1}^3 \max_{\|\alpha\|=4,5}\big\|V_\alpha u(t_{i+1}, \cdot)\big\|_\infty.
\end{aligned} \tag{24}
$$

Moreover, by the mean value theorem, there exists a real number and vector $\mu_1 \in \mathbb{R}$, $\nu_1 \in \mathbb{R}^d$ bounded by K, such that

$$
\frac{\delta_{i+1}}{2}\Delta_i^\pi f = \frac{\delta_{i+1}}{2}\big(\mu_1 \Delta_i^\pi Y + \nu_1 \cdot \Delta_i^\pi Z\big). \tag{25}
$$

Combining (23)–(25) with Young's inequality with $\gamma_1 > 0$, we have

$$
\begin{aligned}
|\Delta_i^\pi Y|^2 &\le (1+\gamma_1\delta_{i+1})|\mathbb{E}_i[\Delta_{i+1}^\pi \Psi]|^2 \\
&\quad + \left(1+\frac{1}{\gamma_1\delta_{i+1}}\right)C\delta_{i+1}^2(|\Delta_i^\pi Y|^2 + |\Delta_i^\pi Z|^2) \\
&\quad + \left(1+\frac{1}{\gamma_1\delta_{i+1}}\right)C\delta_{i+1}^6 \max_{\|\alpha\|=4,5}\|V_\alpha u(t_{i+1},\cdot)\|_\infty^2.
\end{aligned}
\tag{26}
$$

Next, observe that for any random variable F which is measurable with respect to $\mathcal{F}_{t_{i+1}}$, we have

$$
|\mathbb{E}_i[F\mathcal{Z}_i]|^2 = |\mathbb{E}_i[(F-\mathbb{E}_i[F])\mathcal{Z}_i]|^2 \le \frac{1}{\delta_{i+1}}(\mathbb{E}_i[F^2] - \mathbb{E}_i[F]^2).
$$

Combining this with definition of $Z_{2,i}^\pi, i = 0,\dots,n-2$ and the conclusion of Lemma 3.3, we have

$$
\begin{aligned}
\delta_{i+1}\mathbb{E}[|\Delta_i^\pi Z|^2] &\le 2d(\mathbb{E}[|\Delta\Psi_{i+1}|^2] - \mathbb{E}[|\mathbb{E}_i[\Delta\Psi_{i+1}]|^2]) \\
&\quad + C\delta_{i+1}^6 \max_{\|\alpha\|=4,5}\|V_\alpha u(t_{i+1},\cdot)\|_\infty^2.
\end{aligned}
\tag{27}
$$

Putting together (26) and (27) we get

$$
\begin{aligned}
&\mathbb{E}\left[|\Delta_i^\pi Y|^2 + \frac{\delta_{i+1}}{4d}|\Delta_i^\pi Z|^2\right] \\
&\quad\le (1+\gamma_1\delta_{i+1})\mathbb{E}|\mathbb{E}_i[\Delta_{i+1}^\pi\Psi]|^2 \\
&\qquad + C\delta_{i+1}\mathbb{E}[|\Delta_i^\pi Y|^2] + \left(\frac{C}{\gamma_1} + \frac{1}{4d} + C\delta_{i+1}\right)\delta_{i+1}\mathbb{E}[|\Delta_i^\pi Z|^2] \\
&\qquad + C\delta_{i+1}^5 \max_{\|\alpha\|=4,5}\|V_\alpha u(t_{i+1},\cdot)\|_\infty^2 \\
&\quad\le (1+\gamma_1\delta_{i+1})\mathbb{E}[|\Delta_{i+1}^\pi\Psi|^2] \\
&\qquad \times C\delta_{i+1}\mathbb{E}[|\Delta_i^\pi Y|^2] + C\delta_{i+1}^5 \max_{\|\alpha\|=4,5}\|V_\alpha u(t_{i+1},\cdot)\|_\infty^2,
\end{aligned}
\tag{28}
$$

where we have chosen $\gamma_1 = C4d$.

We can argue once more with the mean value theorem and Young's inequality to deduce that

$$
\begin{aligned}
&(1-C\delta_{i+1})\mathbb{E}\left[|\Delta_i^\pi Y|^2 + \frac{\delta_{i+1}}{4d}|\Delta_i^\pi Z|^2\right] \\
&\quad\le (1+C'\delta_{i+1})\left(\mathbb{E}\left[|\Delta_{i+1}^\pi Y|^2 + \frac{\delta_{i+1}}{4d}|\Delta_{i+1}^\pi Z|^2\right]\right) \\
&\qquad + C\delta_{i+1}^5 \max_{\|\alpha\|=4,5}\|V_\alpha u(t_{i+1},\cdot)\|_\infty^2
\end{aligned}
$$

for some different constant C'. By appealing to the discrete version of Gronwall's lemma we complete the proof. $\square$

Theorem 3.4 tells us that (9) produces a second order approximation of the BSDE, provided that the value function $u(t,x) = Y_x^{t,x}$ is smooth with bounded derivatives. In other words, for a "well behaved" PDE, one achieves an error of order $1/n^2$ when a uniform partition with n points is used. This is certainly the case when (A), (B(5)) hold true, the diffusion matrix satisfies the Hörmander condition and Φ is smooth.

A more interesting case occurs when Φ is only Lipschitz continuous (or perhaps belongs only in $\mathbb{L}^2$) and/or the diffusion matrix is degenerate. A priori, it is not clear whether the value function is differentiable, and if yes, how these derivatives behave. However, in a recent paper Crisan and Delarue [7] show that the value function is smooth, at least in the direction of interest, even when the diffusion matrix is degenerate, that is, it only satisfies the so called (*UFG*) condition which we now introduce.

Recalling the operators V_j, $j = 1, \dots, d$ (10), we define the vector field concatenation $V_{[\alpha]}$, $\alpha \in \mathcal{A}$ inductively, as follows:

$$\begin{aligned} V_{[\varnothing]} &:= 0, \\ V_{[i]} &:= V_i, \qquad i = 1, \dots, d, \\ V_{[\alpha * i]} &:= [V_{[\alpha]}, V_i], \qquad i = 0, 1, \dots, d, \end{aligned}$$

where for two first order differential operators $L, W, [L, W]$ denote the usual Lie bracket.

The UFG condition. We say that a system of smooth vector fields $\{V_i : i = 0, \dots, d\}$ satisfy the UFG condition if, for any $\alpha \in \mathcal{A}$, there exists $m \in \mathbb{N}$, and there exist $\varphi_{\alpha,\beta} \in \mathcal{C}_b^\infty(\mathbb{R}^N)$, with $\beta \in \mathcal{A}(m)$ such that

$$V_{[\alpha]}(x) = \sum_{\beta \in \mathcal{A}(m)} \varphi_{\alpha,\beta}(x) V_{[\beta]}(x) \qquad \forall x \in \mathbb{R}^N.$$

The following example is taken from Kusuoka [18]:

EXAMPLE 3.5. Assume $d = 1$ and $N = 2$. Let $V_0, V_1 \in \mathcal{C}_b^\infty(\mathbb{R}^2; \mathbb{R}^2)$ be given by

$$V_0(x_1, x_2) = \sin x_1 \frac{\partial}{\partial x_1}, \qquad V_1(x_1, x_2) = \sin x_1 \frac{\partial}{\partial x_2}.$$

One should note that the Hörmander condition is not satisfied. But the UFG condition is; for $m = 4$, see definition.

THEOREM 3.6 (Crisan and Delarue [7]). *Let* (A) *and* (B(m)) *hold true, and assume further that the vector fields* $\{V_i : i = 0, \dots, d\}$ *satisfy the UFG condi-*

tion. Then, if $\Phi \in C_b^m(\mathbb{R}^d)$ *there exists a unique* $u \in C_b^{\lceil m/2 \rceil, m}([0,T) \times \mathbb{R}^d)$ *that solves* (3). *Moreover, for any multi-index* $\alpha \in \mathcal{A}_m^1$, *there exist increasing function* $c_\alpha, \bar{c}_\alpha : [0,\infty) \to [0,\infty)$ *such that for any* $\Phi \in C_b^m(\mathbb{R}^d)$, *we have*

$$\|V_\alpha u(t,\cdot)\|_\infty \le c_\alpha\biggl(\sum_{\alpha \in \mathcal{A}_m} \|V_\alpha \Phi\|_\infty\biggr), \tag{29}$$

$$\|V_\alpha u(t,\cdot)\|_\infty \le \frac{\bar{c}_\alpha(\|\Phi\|_{\mathrm{Lip}})}{(T-t)^{(\|\alpha\|-1)/2}}, \qquad t \in [0,T). \tag{30}$$

Observe that the directional derivatives of the value function explode as $t \uparrow T$. To compensate for this behavior we choose to work with a nonequidistant partition, one that becomes more dense as we approach the boundary. In this manner, we manage to achieve an order of convergence of $1/n^2$ with n points on the grid.

COROLLARY 3.7. *Let assumptions* (A), (B(5)) *hold true, and assume that* Φ *is Lipschitz continuous. Consider the discretization* (9) *along the partition* π:

$$t_i = T\biggl(1 - \biggl(1 - \frac{i}{n}\biggr)^\beta\biggr), \qquad i = 0, \dots, n, \beta \ge 5.$$

Then there exists a constant C *independent of the partition and of the value function* u, *such that*

$$\max_{0 \le i \le n-1} \mathbb{E}\biggl[|Y_{t_i} - Y_{2,i}^\pi|^2 + \frac{1}{n^2}|Z_{t_i} - Z_{2,i}^\pi|^2\biggr]^{1/2} \le \frac{C}{n^2}.$$

PROOF. Theorem 3.4 tells us that the error is controlled by

$$C\biggl(|Y_{t_{n-1}} - Y_{2,n-1}^\pi|^2 + \frac{\delta_n}{4d}|Z_{t_{n-1}} - Z_{2,n-1}^\pi|^2\biggr) + \sum_{i=1}^{n-1} \delta_i^5 \max_{\|\alpha\|=4,5} \|V_\alpha u(t_i,\cdot)\|_\infty^2.$$

We examine every term separately.

According to Theorem 3.6, we have

$$\begin{aligned}
&\sum_{i=1}^{n-1} \delta_i^5 \max_{\|\alpha\|=4,5} \|V_\alpha u(t_i,\cdot)\|_\infty^2 \\
&\qquad \le \sum_{i=1}^{n-1} \delta_i^5 \frac{C\|\nabla\Phi\|_\infty}{(T-t_i)^4} \\
&\qquad = \sum_{i=1}^{n-1} T^5 \biggl(\int_{1-(i/n)}^{1-(i-1)/n} \beta s^{\beta-1}\, ds\biggr)^5 \frac{C\|\nabla\Phi\|_\infty}{T^4(1-(i/n))^{4\beta}} \\
&\qquad \le \sum_{i=1}^{n-1} \frac{T\beta^5}{n^5} \frac{(1-(i-1)/n)^{5(\beta-1)}}{(1-(i/n))^{4\beta}} \le C/n^4
\end{aligned} \tag{31}$$

since $\beta \geq 5$.

To complete our proof we estimate the error on the first backward step. Using the standard estimate on the Euler discretization error of Lebesgue integrals, we have

$$\begin{aligned}\left|Y_{n-1} - Y^{\pi}_{1,n-1}\right|^2 &= \biggl|\mathbb{E}_i\biggl[\int_{t_{n-1}}^{t_n} \bar{f}(s, X_s)\, ds\biggr] \mp f(X_{t_{n-1}}, Y_{t_{n-1}}, Z_{t_{n-1}})\delta_n \\ &\qquad - f(X_{t_{n-1}}, Y^{\pi}_{1,n-1}, Z^{\pi}_{1,n-1})\delta_n\biggr|^2 \\ &\leq C\delta_n^2\bigl(1 + \left|Y_{t_{n-1}} - Y^{\pi}_{1,n-1}\right|^2 + \left|Z_{t_{n-1}} - Z^{\pi}_{1,n-1}\right|^2\bigr),\end{aligned}$$

where once again, we have used the mean value theorem. Rearranging the terms, we may argue on the existence of a constant C such that

$$\begin{aligned}&(1 - C\delta_n)\biggl(\left|Y_{t_{n-1}} - Y^{\pi}_{1,n-1}\right|^2 + \frac{\delta_n}{4d}\left|Z_{t_{n-1}} - Z^{\pi}_{1,n-1}\right|^2\biggr) \\ &\qquad \leq C\delta_n^2 + C\delta_n\left|Z_{t_{n-1}} - Z^{\pi}_{1,n-1}\right|^2.\end{aligned} \tag{32}$$

From standard estimates on BSDEs we know that under (A), (B(m)) and (C1)

$$\sup_{0 \leq t \leq T} \mathbb{E}\bigl[|Z_t|^2\bigr] < +\infty.$$

Finally, under (C1) we have that

$$\begin{aligned}\mathbb{E}\bigl[\left|Z^{\pi}_{1,n-1}\right|^2\bigr] &= \frac{1}{\delta_n^2}\mathbb{E}\bigl[\left|\mathbb{E}_{n-1}[\Phi(X_{t_n})\Delta B_n]\right|^2\bigr] \\ &= \frac{1}{\delta_n^2}\mathbb{E}\bigl[\left|\mathbb{E}_{n-1}\bigl[\bigl(\Phi(X_{t_n}) - \Phi(X_{t_{n-1}})\bigr)\Delta B_n\bigr]\right|^2\bigr] \leq C.\end{aligned} \tag{33}$$

Substituting (33) into (32) and then taking square roots, completes the proof. □

4. An implementable second order scheme. As is the case with all other FBSDE discretization methods (see, e.g., the first order method of Bouchard and Touzi [5], Zhang [26] or the forward backward algorithm of Bender and Denk [3]), the algorithm suggested in the previous section does not constitute on its own, an implementable numerical scheme, as one needs to employ a method that approximates the involved conditional expectations. To the best of our knowledge, to date, four different methods have been suggested for computing these conditional expectations: The Malliavin calculus algorithm of [5] (with a further refinement presented in [11]), the regression on function basis approach of Gobet, Lemor and Warin [14, 19], the quantization method of Bally and Pagès [1] and finally, the algorithm presented by the authors in [10], based on the cubature on Wiener space method of Lyons and Victoir [22]. In this section, we combine the latter with the discretization of the previous section, to construct an implementable scheme for FBSDEs of second order.

In [10], the authors include a detailed presentation of the cubature algorithm and how it may be applied to BSDEs in conjunction with the first order discretization (6). We also suggest the use of the tree-based branching algorithm to control the computational effort required by the cubature method. TBBA is a minimal variance reduction method that keeps the computational effort per partition step constant.[6] As a result, the proposed algorithm not only enjoys nice asymptotic convergence properties (see Corollary 4.2 and Theorem 5.5 of [10]), but it is also competitive with other existing methods for BSDEs, even in high dimensions; see Section 6 of [10].

All details regarding the asymptotic analysis of the error as well as the implementation, found in [10], lend themselves perfectly when we combine cubature and TBBA with the second order scheme (9). This is of course of no surprise as the two discretization methods, (6) and (9), have a very similar functional formulation. We will not present here the details of this analysis. We restrict ourselves to the basic definition and main results. The interested reader can fill in the details by following the material in [10].

DEFINITION 4.1. Consider a fixed $t > 0$ and $m \in \mathbb{N}_+$. We say that the discrete measure $\mathbb{Q}_t^m$ supported on paths of finite variation $\omega_1, \dots, \omega_{c_d^m} \in C_{bv}([0,T];\mathbb{R}^d)$ with corresponding positive weights $\lambda_1, \dots, \lambda_{c_d^m}$, $\mathbb{Q}_t^m := \sum_{j=1}^{c_d^m} \lambda_j \delta_{\omega_{j,t}}$ is a cubature measure of degree m if

$$\begin{aligned}\mathbb{E}_{\mathbb{Q}_t^m}[J_{0,t}^\alpha] &:= \sum_{i=1}^{c_d^m} \lambda_i \int_{0<s_1<\cdots<s_k<t} d\omega_{j,t}^{\alpha_1}(s_1)\cdots d\omega_{j,t}^{\alpha_k}(s_k),\\ &= \mathbb{E}[J_{0,t}^\alpha] \qquad \forall \alpha = (\alpha_1,\dots,\alpha_k) \in \mathcal{A}(m),\end{aligned}$$

where $\omega_{j,t}(s) := (s, \omega_{j,t}^1(s), \dots, \omega_{j,t}^d(s))$.

The scaling properties of Brownian motion imply that if $\mathbb{Q}_1^m$ is a cubature measure at time $t = 1$ supported on the paths $\omega_1, \dots, \omega_n$, then we obtain the cubature measure $\mathbb{Q}_T^m$ by considering the paths

$$\omega_{1,T}(t) = \sqrt{T}\omega_1(t/T), \dots, \omega_{c_d^m,T}(t) = \sqrt{T}\omega_{c_d^m}(t/T)$$

and the same weights as in $\mathbb{Q}_1^m$. The existence of cubature measures is proved in [22]. However, their explicit construction is by no means trivial. This task is carried out in [22] for cubature formulas of order 3 and 5 and in Gyurkó and

[6] A different but very efficient approach for controlling the growth in the number of particles in the cubature method can be found in [21], the so-called recombination method.

Lyons [15], Litterer [20] for cubature of order 7 in dimensions 1, 2, 3 and in [15] for cubature of order 9 and dimension 1.

Given a smooth function $g:\mathbb{R}^d \to \mathbb{R}$ and a cubature measure $\mathbb{Q}_t^m$, the stochastic Taylor expansion (13) tells us that

$$\begin{aligned}&\sup_x \bigl|\mathbb{E}[g(X_{t_{i+1}}^{t_i,x})] - \mathbb{E}_{\mathbb{Q}_t^m}[g(X_{t_{i+1}}^{t_i,x})]\bigr| \\ &\qquad \le C \sum_{j=m+1}^{m+2} \delta_{i+1}^{j/2} \sup_{\alpha\in\mathcal{A}(j)\setminus\mathcal{A}(j-1)} \|V_\alpha g\|_\infty.\end{aligned} \tag{34}$$

Hence, for a small time step $\mathbb{E}_{\mathbb{Q}_t^m}[g(X_{t_{i+1}}^{t_i,x}))]$ is an efficient approximation of $\mathbb{E}[g(X_{t_{i+1}}^{t_i,x}))]$. We can easily show that we also have

$$\begin{aligned}&\sup_x \bigl|\mathbb{E}[g(X_{t_{i+1}}^{t_i,x})\mathcal{Z}_i] - \mathbb{E}_{\mathbb{Q}_t^m}[g(X_{t_{i+1}}^{t_i,x}))\mathcal{Z}_i]\bigr| \\ &\qquad \le C \sum_{j=m+1}^{2m} \delta_{i+1}^{(j-1)/2} \sup_{\alpha\in\mathcal{A}(j)\setminus\mathcal{A}(j-1)} \|V_\alpha g\|_\infty + \delta_{i+1}^{(m-1)/2}\|\nabla g\|_\infty.\end{aligned} \tag{35}$$

Using the cubature method, we can produce an evolving cloud of points at every point on our time partition π that approximate the integrals against the law of the forward process. We denote by $\Xi_{t,x}(\omega)$ the solution of the ODE

$$y_{t,x} = x + \sum_{j=0}^{d} \int_0^t V(y_{s,x})\, d\omega^j(s), \tag{36}$$

where we make the identification $\omega(t) \equiv (t, \omega(t)) \in \mathbb{R} \oplus \mathbb{R}^d$ for any $\omega \in C_{bv}([0,T];\mathbb{R}^d)$. We denote by $\mathcal{S}_k$, $k = 0,\dots,n$ the set of weights and points in the cloud at depth k (equivalently time t_k),

$$\begin{aligned}&\mathcal{S}_0 := \{(1, x_0)\}, \\ &\mathcal{S}_k := \{(\lambda_j\lambda_z, \Xi_{\delta_k,z}(\omega_{j,\delta_k}));\ j = 1,\dots,c_d^m, (\lambda_z, z) \in \mathcal{S}_{k-1}\}, \\ &\hspace{22em} k = 1,\dots,n.\end{aligned} \tag{37}$$

Given a point $x \in \mathcal{S}_k$ we can approximate the expectations

$$\mathbb{E}[g(X_{t_{i+1}}^{t_i,x})\mathcal{Z}_i], \qquad \mathbb{E}[g(X_{t_{i+1}}^{t_i,x})]$$

by averaging out the involved functions on all the offsprings of x in the set $\mathcal{S}_{k+1}$.

However, the cardinality of this sets of points grows exponentially, and this is issue needs to be addressed if one wants to use the method efficiently on any partition. The authors suggested in [10] the use of the tree-based branching algorithm to control this growth. Formally, given a cubature measure at depth t_k, $\mathbb{Q}_{t_k}^m$, TBBA constructs a random measure

$$\hat{\mathbb{Q}}_t^m := \sum_{(\lambda,z)\in\mathcal{S}_k} \hat{\lambda}_z \delta_z,$$

where the $\hat{\lambda}_z$'s are random variables, some of which may well be zero. Let $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$ denote a probability space that supports these random variables, and let also $\tilde{\mathbb{E}}$ denote integration with respect to $\tilde{\mathbb{P}}$. The defining property of the measure $\hat{\mathbb{Q}}_{t_k}^m$ is

$$\tilde{\mathbb{E}}[\hat{\mathbb{Q}}_k^m] = \mathbb{Q}_k^m \quad \Leftrightarrow \quad \tilde{\mathbb{E}}[\hat{\lambda}_z] = \lambda_z \tag{38}$$
$$\forall z \in \mathcal{S}_k, \text{ and } \#\{z \in \mathcal{S}_k : \hat{\lambda}_z > 0\} \leq N.$$

We denote by $\hat{\mathcal{S}}_k$ the set of points selected by the TBBA,

$$\hat{\mathcal{S}}_k := \{x \in \mathcal{S}_k, \hat{\lambda}_x > 0\}.$$

The offspring of every point $x \in \hat{\mathcal{S}}_k$ that survive the selection process are denoted by

$$\hat{\mathcal{S}}_k^x := \hat{\mathcal{S}}_k \cap \{\text{offspring of } x\}, \qquad x \in \hat{\mathcal{S}}_{k-1}, k = 1, \dots, n.$$

There is an extensive discussion in Section 5 of [10] on the theoretical and practical details of TBBA, along with the corresponding pseudo code to facilitate its implementation.

Using (9) as a template, we substitute integration with respect to the Wiener measure with integration against the cubature and TBBA measures $\{\hat{\mathbb{Q}}_{t_k}^m\}_{1 \leq k \leq n}$ where we consider a fixed value for m (the cubature degree) and fixed N (the number of particles that are sampled at every depth). We denote by $\hat{\mathbb{E}}_i$ the conditional expectation with respect to the family $\{\hat{\mathbb{Q}}_{t_k}^m\}_{1. \leq k \leq n}$. Given a function $g : \mathbb{R}^d \to \mathbb{R}$ we have

$$\hat{\mathbb{E}}_i[g(X_{t_{i+1}}) | X_{t_i} = x] = \sum_{(\lambda_z, z) \in \hat{\mathcal{S}}_{i+1}^x} \frac{\hat{\lambda}_z}{\hat{\lambda}_x} g(z), \qquad x \in \hat{\mathcal{S}}_i. \tag{39}$$

Also given a point $x \in \hat{\mathcal{S}}^i$, $z \in \hat{\mathcal{S}}_{i+1}^x$ we denote by ω_z the cubature path (i.e., one of the paths $\omega_1, \dots, \omega_{c_d^m}$) that was used to arrive at the point z starting from x. We then have the following approximations for $Y_{2,i}^\pi, Z_{2,i}^\pi, i = 0, \dots, n$:

- *Initialization.*

$$\hat{Y}_{2,n}^\pi(x) := \Phi(x), \qquad x \in \hat{\mathcal{S}}_n,$$
$$\hat{Z}_{2,n}^\pi(x) := \begin{cases} 0, & \text{if (C1) is in force,} \\ \nabla\Phi(x)V(x), & \text{if (C2) is in force,} \end{cases} \qquad x \in \hat{\mathcal{S}}_n$$

If (C1) is in force:

$$\hat{Z}_{2,n-1}^\pi(x) = \frac{1}{\delta_n} \mathbb{E}_{\hat{\mathbb{Q}}_n^m}[\hat{Y}_{2,n}^\pi \Delta\omega_n | X_{t_{n-1}} = x],$$
$$x \in \hat{\mathcal{S}}_{n-1}, \Delta\omega_n = \omega_{\delta_n}(t_n) - \omega_{\delta_n}(t_{n-1}),$$
$$\hat{Y}_{2,n-1}^\pi(x) = \mathbb{E}_{\hat{\mathbb{Q}}_n^m}[Y_{2,n}^\pi | X_{t_{n-1}} = x] + \delta_n f(x, \hat{Y}_{2,n-1}^\pi(x), \hat{Z}_{2,n-1}^\pi(x)),$$
$$x \in \hat{\mathcal{S}}_{n-1}.$$

• *Backward induction*.

For $x \in \hat{\mathcal{S}}_i$, $i = n-2, \dots, 0$,

$$\hat{\mathcal{Z}}_i(z) = \frac{4}{\delta_{i+1}}\big(\omega_{z,\delta_{i+1}}(t_{i+1}) - \omega_{z,\delta_{i+1}}(t_i)\big) - \frac{6}{\delta_{i+1}^2}\int_{t_i}^{t_{i+1}} (s-t_i)\, d\omega_{z,\delta_{i+1}}(s), \qquad z \in \hat{\mathcal{S}}_{i+1}^x,$$

$$\hat{Z}^\pi_{2,i}(x) = \mathbb{E}_{\hat{\mathbb{Q}}_i^m}\big[\big(\hat{Y}^\pi_{2,i+1} + \delta_{i+1} f\big(X_{t_{i+1}}, \hat{Y}^\pi_{2,i+1}, \hat{Z}^\pi_{2,i+1}\big)\big)\hat{\mathcal{Z}}_{i+1} | X_{t_i} = x\big],$$

$$\hat{\mathcal{Z}}_i := \big(\hat{\mathcal{Z}}_i^1, \dots, \hat{\mathcal{Z}}_i^d\big)^T, \tag{40}$$

$$\begin{aligned}\hat{Y}^\pi_{2,i}(x) = {} & \mathbb{E}_{\hat{\mathbb{Q}}_i^m}\big[Y^\pi_{2,i+1} | X_{t_i} = x\big] \\ & + \frac{\delta_{i+1}}{2}\big(f\big(x, \hat{Y}^\pi_{2,i}(x), \hat{Z}^\pi_{2,i}(x)\big) \\ & + \mathbb{E}_{\hat{\mathbb{Q}}_i^m}\big[f\big(X_{t_{i+1}}, \hat{Y}^\pi_{2,i+1}, \hat{Z}^\pi_{2,i+1}\big) | X_{t_i} = x\big]\big).\end{aligned}$$

The computation of the involved (conditional) expectations follows a straightforward logic. For example,

$$\begin{aligned}&\mathbb{E}_{\hat{\mathbb{Q}}_i^m}\big[\big(\hat{Y}^\pi_{2,i+1} + \delta_{i+1} f\big(X_{t_{i+1}}, \hat{Y}^\pi_{2,i+1}, \hat{Z}^\pi_{2,i+1}\big)\big)\hat{\mathcal{Z}}_i | X_{t_i} = x\big] \\ &\quad = \sum_{(\hat{\lambda}_z, z) \in \hat{\mathcal{S}}_i^x} \frac{\hat{\lambda}_z}{\hat{\lambda}_x}\big(\hat{Y}^\pi_{2,i+1}(z) + \delta_{i+1} f\big(z, \hat{Y}^\pi_{2,i+1}(z), \hat{Z}^\pi_{2,i+1}(z)\big)\big)\hat{\mathcal{Z}}_i(z).\end{aligned} \tag{41}$$

The analysis of the overall error of the algorithm $|Y_0 - \hat{Y}^\pi_{2,0}|$ presents no significant differences with similar analysis appearing elsewhere in the numerics for BSDEs literature. In the usual fashion, one breaks the error in two components. The discretization error $|Y_0 - Y^\pi_{2,0}|$ and the simulation error $|Y^\pi_{2,0} - \hat{Y}^\pi_{2,0}|$. The discretization error has already been addressed in Theorem 3.7. Regarding the simulation error, a derivation of its asymptotic behavior can just follow the steps in the proof of Theorem 5.5 of [10]. In fact, the only thing that appears to be different is the quantification of the error

$$\mathbb{E}\big[g\big(X^{t_i,x}_{t_{i+1}}\big)\mathcal{Z}_i\big] - \mathbb{E}_{\mathbb{Q}^m}\big[g\big(X^{t_i,x}_{t_{i+1}}\big)\mathcal{Z}_i\big].$$

However, this can be treated similar to equation (3.5) of [10]. Putting everything together, we have the following error estimate:

THEOREM 4.2. *Let the coefficients of system* (1) *satisfy assumptions* (A), (B(m)), (C1), (UFG) *and* (V0) *and consider a partition* π *with* n *points of* $[0,T]$ *defined as*

$$t_i = T\bigg(1 - \bigg(1 - \frac{i}{n}\bigg)^\beta\bigg), \qquad i = 0, \dots, n, \beta \geq 5.$$

Along the partition π we consider the random variables $\hat{Y}^{\pi}_{2,i}, \hat{Z}^{\pi}_{2,i}$ defined as in (40) *using a cubature measure of order $m \geq 7$ and at most N points in the support of the intermediate measures. Then there exist a constant $C > 0$ such that*

$$\mathbb{E}|Y_0 - \hat{Y}^{\pi}_{2,0}| \leq C\left(\frac{1}{n^2} + \frac{n}{N^{1/2}}\right).$$

4.1. *A numerical example.* A detailed analysis through numerous numerical examples of the performance of the cubature and TBBA algorithm for BSDEs in conjunction with a first order discretization is presented in [10]. We do not intend to engage here in a similar discussion. We merely want to present a simple case where we compare the performance of the second order algorithm with the first order method so as to convince the reader of the comparative advantage of the second order discretization.

To that end we focus on a smooth example also considered in [10], namely the FBSDE

$$X_t^{0,x_0} = x_0 + \int_0^t \mu X_s\, ds + \int_0^t \sqrt{1 + X_t^2}\, dB_t, \qquad 0 \leq t \leq T,$$

$$Y_t^{0,x_0} = \arctan\left(X_T^{0,x_0}\right) - \int_t^T rY_s + e^{r(T-s)}(\mu - 1)\frac{X_s^{0,x_0} Z_s^{0,x_0}}{\sqrt{1 + (X_t^{0,x_0})^2}}\, ds - \int_t^T Z_s^{0,x_0}\, dB_s. \tag{42}$$

It is easy to check, by means of Itô's lemma, that the solution to the above system is given by

$$Y_t^{0,x_0} = e^{-r(T-t)} \arctan\left(X_t^{0,x_0}\right), \qquad Z_t^{0,x_0} = \frac{e^{-r(T-t)}}{\sqrt{1 + (X_t^{0,x_0})^2}}.$$

We test our example with parameters

T	μ	r	x_0
1	0.1	0.2	1

We repeat the scheme 10 times and collect values. Let $y^m = \hat{Y}^{\pi}_{2,0}$ for $m = 1, \dots, 10$. We plot in Figure 1 the average relative error, that is,

$$\text{Error} = \frac{1}{10}\sum_{m=1}^{10}\left|\frac{y^m - Y_0}{Y_0}\right|.$$

It should be clear from these graphs that the cubature and TBBA algorithm achieves errors an order of magnitude better when combined with the second order discretization than when combined with first order one.

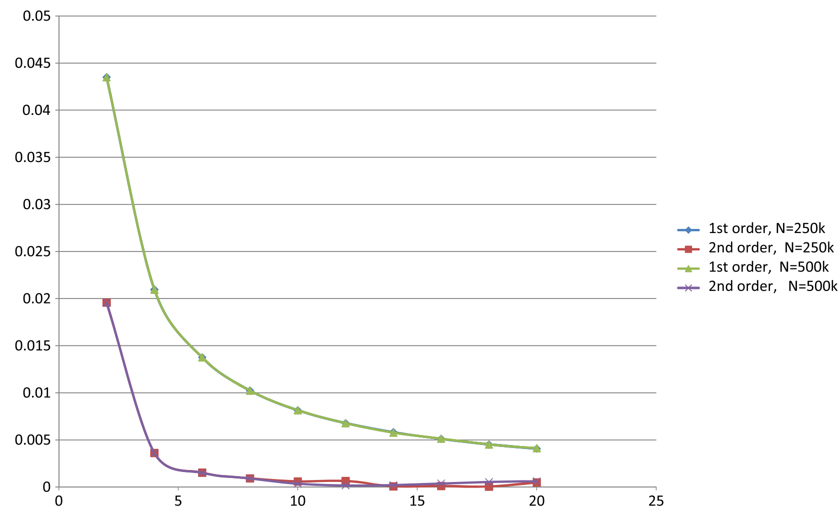

FIG. 1. *Numerical results on system* (42).

APPENDIX

PROOF OF THEOREM 4.2. We will only give some comments on the proof as it follows very closely the proofs of Corollary 4.2 and Theorem 5.5 of [10]. As mentioned in the beginning of this subsection, the discretization error is already analyzed in Corollary (3.7). The simulation error $|Y^{\pi}_{2,0} - \hat{Y}^{\pi}_{2,0}|$ is broken in its own turn in two parts:

$$\left|Y^{\pi}_{2,0} - \hat{Y}^{\pi}_{2,0}\right| \leq \left|Y^{\pi}_{2,0} - \bar{Y}^{\pi}_{t_0}\right| + \left|\bar{Y}^{\pi}_{t_0} - \hat{Y}^{\pi}_{2,0}\right|,$$

where $\bar{Y}^{\pi}_{t_i}, \bar{Z}^{\pi}_{t_i}, i = 0, \dots, n$ are random variables defined exactly as in (40) but with the pure cubature measures $\mathbb{Q}^m_i$ in place of $\hat{\mathbb{Q}}^m_i$. We can then, using estimates (34), (35), follow the steps in the proof of Theorem 3.4 and Corollary 4.2 of [10] to show that

$$\left|Y^{\pi}_{2,0} - \bar{Y}^{\pi}_{t_0}\right| \leq C/n^2.$$

Finally, the influence of the sampling process to the error may be assessed exactly as in Theorem 5.5 of [10]. Indeed, the only thing that differs between (40) and the first order algorithm presented in [10] is the weight in the conditional expectations approximating Z, that is, $\mathcal{Z}^l_i$ versus $\frac{\Delta B^l_{i+1}}{\delta_{i+1}}, l = 1, \dots, d$. However, note that $\mathcal{Z}_i$ is a random variable of the same order as $\frac{\Delta B^l_{i+1}}{\delta_{i+1}}$ which, and this is the crucial part, has conditional expectation 0 when conditioned against $\mathcal{F}_{t_i}$. Hence the Lipschitz like property as it appears in equations (5.7), (5.9) of the proof of Theorem 5.5 of [10] is enjoyed by scheme (40). All else follows in an identical manner. $\square$

REFERENCES

[1] BALLY, V. and PAGÈS, G. (2003). Error analysis of the optimal quantization algorithm for obstacle problems. *Stochastic Process. Appl.* **106** 1–40. MR1983041

[2] BALLY, V. and PAGÈS, G. (2003). A quantization algorithm for solving multi-dimensional discrete-time optimal stopping problems. *Bernoulli* **9** 1003–1049. MR2046816

[3] BENDER, C. and DENK, R. (2007). A forward scheme for backward SDEs. *Stochastic Process. Appl.* **117** 1793–1812. MR2437729

[4] BENDER, C. and ZHANG, J. (2008). Time discretization and Markovian iteration for coupled FBSDEs. *Ann. Appl. Probab.* **18** 143–177. MR2380895

[5] BOUCHARD, B. and TOUZI, N. (2004). Discrete-time approximation and Monte–Carlo simulation of backward stochastic differential equations. *Stochastic Process. Appl.* **111** 175–206. MR2056536

[6] CHEVANCE, D. (1997). Numerical methods for backward stochastic differential equations. In *Numerical Methods in Finance* (L. C. G. Rogers and D. Talay, eds.) 232–244. Cambridge Univ. Press, Cambridge. MR1470517

[7] CRISAN, D. and DELARUE, F. (2010). Sharp gradient bounds for solutions of degenerate semi-linear partial differential equations. Preprint. Available at https://math.unice.fr/~delarue.

[8] CRISAN, D. and GHAZALI, S. (2007). On the convergence rates of a general class of weak approximations of SDEs. In *Stochastic Differential Equations: Theory and Applications. Interdiscip. Math. Sci.* **2** 221–248. World Sci. Publ., Hackensack, NJ. MR2393578

[9] CRISAN, D. and LYONS, T. (2002). Minimal entropy approximations and optimal algorithms. *Monte Carlo Methods Appl.* **8** 343–355. MR1943203

[10] CRISAN, D. and MANOLARAKIS, K. (2012). Solving backward stochastic differential equations using the cubature method: Application to nonlinear pricing. *SIAM J. Financial Math.* **3** 534–571. MR2968045

[11] CRISAN, D., MANOLARAKIS, K. and TOUZI, N. (2010). On the Monte Carlo simulation of BSDEs: An improvement on the Malliavin weights. *Stochastic Process. Appl.* **120** 1133–1158. MR2639741

[12] DOUGLAS, J. JR., MA, J. and PROTTER, P. (1996). Numerical methods for forward-backward stochastic differential equations. *Ann. Appl. Probab.* **6** 940–968. MR1410123

[13] GOBET, E. and LABART, C. (2007). Error expansion for the discretization of backward stochastic differential equations. *Stochastic Process. Appl.* **117** 803–829. MR2330720

[14] GOBET, E., LEMOR, J.-P. and WARIN, X. (2005). A regression-based Monte Carlo method to solve backward stochastic differential equations. *Ann. Appl. Probab.* **15** 2172–2202. MR2152657

[15] GYURKÓ, L. G. and LYONS, T. J. (2011). Efficient and practical implementations of cubature on Wiener space. In *Stochastic Analysis* 2010 (D. Crisan, eds.) 73–111. Springer, Heidelberg. MR2789080

[16] KLOEDEN, P. E. and PLATEN, E. (1992). *Numerical Solution of Stochastic Differential Equations. Applications of Mathematics* (*New York*) **23**. Springer, Berlin. MR1214374

[17] KUSUOKA, S. (2001). Approximation of expectation of diffusion process and mathematical finance. In *Taniguchi Conference on Mathematics Nara '98. Adv. Stud. Pure Math.* **31** 147–165. Math. Soc. Japan, Tokyo. MR1865091

[18] KUSUOKA, S. (2003). Malliavin calculus revisited. *J. Math. Sci. Univ. Tokyo* **10** 261–277. MR1987133

[19] LEMOR, J.-P., GOBET, E. and WARIN, X. (2006). Rate of convergence of an empirical regression method for solving generalized backward stochastic differential equations. *Bernoulli* **12** 889–916. MR2265667

[20] LITTERER, C. (2008). The signature in numerical algorithms. Ph.D. thesis, Univ. Oxford.

[21] LITTERER, C. and LYONS, T. (2012). High order recombination and an application to cubature on Wiener space. *Ann. Appl. Probab.* **22** 1301–1327. MR2985162

[22] LYONS, T. and VICTOIR, N. (2004). Cubature on Wiener space. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **460** 169–198. MR2052260

[23] MA, J., PROTTER, P., SAN MARTÍN, J. and TORRES, S. (2002). Numerical method for backward stochastic differential equations. *Ann. Appl. Probab.* **12** 302–316. MR1890066

[24] MA, J. and ZHANG, J. (2002). Representation theorems for backward stochastic differential equations. *Ann. Appl. Probab.* **12** 1390–1418. MR1936598

[25] PARDOUX, É. and PENG, S. (1992). Backward stochastic differential equations and quasilinear parabolic partial differential equations. In *Stochastic Partial Differential Equations and Their Applications* (*Charlotte*, *NC*, 1991). *Lecture Notes in Control and Inform. Sci.* **176** 200–217. Springer, Berlin. MR1176785

[26] ZHANG, J. (2004). A numerical scheme for BSDEs. *Ann. Appl. Probab.* **14** 459–488. MR2023027

[27] ZHAO, W., LI, Y. and ZHANG, G. (2012). A generalized θ-scheme for solving backward stochastic differential equations. *Discrete Contin. Dyn. Syst. Ser. B* **17** 1585–1603. MR2903768

DEPARTMENT OF MATHEMATICS
IMPERIAL COLLEGE LONDON
180 QUEEN'S GATE
LONDON, SW7 2AZ
UNITED KINGDOM
E-MAIL: d.crisan@ic.ac.uk
km3@ic.ac.uk