

PROBABILISTIC METHODS FOR SEMILINEAR PARTIAL DIFFERENTIAL EQUATIONS. APPLICATIONS TO FINANCE*

DAN CRISAN¹ AND KONSTANTINOS MANOLARAKIS¹

Abstract. With the pioneering work of [Pardoux and Peng, *Syst. Contr. Lett.* **14** (1990) 55–61; Pardoux and Peng, *Lecture Notes in Control and Information Sciences* **176** (1992) 200–217]. We have at our disposal stochastic processes which solve the so-called *backward stochastic differential equations*. These processes provide us with a Feynman-Kac representation for the solutions of a class of nonlinear partial differential equations (PDEs) which appear in many applications in the field of Mathematical Finance. Therefore there is a great interest among both practitioners and theoreticians to develop reliable numerical methods for their numerical resolution. In this survey, we present a number of probabilistic methods for approximating solutions of semilinear PDEs all based on the corresponding Feynman-Kac representation. We also include a general introduction to backward stochastic differential equations and their connection with PDEs and provide a generic framework that accommodates existing probabilistic algorithms and facilitates the construction of new ones.

Mathematics Subject Classification. 65C30, 65C05, 60H07, 62G08.

Received 19 May, 2009. Revised 1st March, 2010.
Published online August 26, 2010.

1. INTRODUCTION

One of the most celebrated results in stochastic analysis is the so-called Feynman-Kac representation formula, first established in the late forties (see [21,28]) for the heat equation and extended since to more general parabolic equations. This formula provides us with a fundamental link between solutions of Partial Differential Equations and solutions of Stochastic Differential Equations (SDEs), thus allowing the dual formulation of numerous problems.

Keywords and phrases. Probabilistic methods, semilinear PDEs, BSDEs, Monte Carlo methods, Malliavin calculus, cubature methods.

* Research supported by EPSRC Grant No EP/H000550/1 and EP/H000100/1.

¹ Department of Mathematics, Imperial College London, 180 Queen's Gate, London, SW7 2AZ, UK.
d.crisan@imperial.ac.uk; km3@imperial.ac.uk

In brief, let $v : [0, T] \times \mathbb{R}^q \rightarrow \mathbb{R}$ be a function of polynomial growth, which is the solution of the Cauchy problem

$$\begin{aligned} (\partial_t + L)v &= 0, & (t, x) \in [0, T] \times \mathbb{R}^q \\ v(T, x) &= \Phi(x) & x \in \mathbb{R}^q \end{aligned}$$

where L is the second order differential operator

$$Lv = \sum_{i=1}^q V_0^i \partial_i v + \frac{1}{2} \sum_{i,j=1}^q a^{ij} \partial_i \partial_j v. \tag{1.1}$$

In (1.1), $V_0 = (V_0^i)_{i=1}^q : \mathbb{R}^q \rightarrow \mathbb{R}^q$ is a Lipschitz function and $A = (a^{ij})_{i,j=1}^q : \mathbb{R}^q \rightarrow \mathbb{R}^{q \times q}$ is a matrix valued function such that $A = V^*V$, where $V = (V_1, \dots, V_q)$ is a matrix valued function with columns given by the Lipschitz functions $V_j = (V_j^i)_{i=1}^q : \mathbb{R}^q \rightarrow \mathbb{R}^q, j = 1, \dots, d^2$. Then v has the following Feynman-Kac representation³

$$v(s, x) = \mathbb{E} [\Phi(X_T^{s,x})], \quad (s, x) \in [0, T] \times \mathbb{R}^q,$$

where $X^{s,x}$ is the solution to the SDE

$$X_t^{s,x} = x + \int_s^t V_0(X_u^{s,x}) du + \sum_{i=1}^q \int_s^t V_i(X_u^{s,x}) dW_u^i, \quad s \leq t \leq T. \tag{1.2}$$

Similar representations exists for more general classes of Cauchy, Dirichlet and free boundary problems. However, it was not until the early nineties that Feynman-Kac formulas were developed for nonlinear equations in the seminal work of Pardoux and Peng [40,41]. They considered the following Cauchy problem

$$\begin{cases} (\partial_t + L)u = -f(t, x, u, (\nabla u V)(x)), & t \in [0, T], \quad x \in \mathbb{R}^q \\ u(T, x) = \Phi(x), & x \in \mathbb{R}^q, \end{cases} \tag{1.3}$$

where ∇u is the row vector $\nabla u := (\partial_{x_1} u, \dots, \partial_{x_q} u)$. A stochastic representation to the solution of (1.3) is obtained by means of a solution to a *backward* stochastic differential equation (BSDE henceforth). The first occurrence of BSDEs in the literature may be traced back to 1973, when [4] studied a BSDE of a Ricatti type. The general problem was first addressed by [40] and a vast literature has been developed since. Notably, one of the driving forces behind the rapid expansion of the theory of BSDEs has been their various applications in Mathematical Finance. These equations are ideal tools for solving problems in a market framework beyond the Black-Scholes-Merton classical paradigm. For instance, BSDEs are used in the study of pricing contingent claims with trading constraints as in [12,27] or in incomplete markets as in [16]. They have also been applied to problems of recursive utility [13,14] as well as to the pricing of American options [19]. We review some of their applications in finance in Section 3.

As solutions of BSDEs provide a probabilistic representation (the Feynman-Kac formula) to the solution of the Cauchy problem (1.3), they form the basis of a number of algorithms of probabilistic nature suited for the numerical approximation of (1.3), thus offering an alternative to deterministic, grid-based algorithms. It is a well known fact that this type of algorithms, though very efficient in low dimensions, become difficult to handle or even intractable, as the dimension increases. The reason is that the size of the grid (and hence the computational effort) increases exponentially with the dimension of the state space. A probabilistic numerical method, one based on the numerical approximation of the corresponding BSDE does not require a grid, hence it can be applied to problems where the dimension of the state space goes beyond the current capabilities

²Note that V is not required to be a square matrix valued.
³For a proof of this result, see, for example, Theorem 5.7.6 in [29].

of its deterministic counter-parts. In this survey we present four such probabilistic algorithms: the quantization method (Sect. 5), the Malliavin calculus regression method (Sect. 6), the regression on function bases method (Sect. 7) and the cubature method (Sect. 8).

In the following section, we describe in detail backward SDEs and forward-backward SDEs and explain their connection with semilinear PDEs. In Section 3 we describe the applications of BSDEs to Mathematical Finance and in Section 4 we describe the discretization process for these equations which is the first step towards constructing probabilistic algorithms for solving the Cauchy problem (1.3).

Throughout we will be working with the following notation and assumptions:

Notation. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space on which we consider a q -dimensional standard Brownian motion $W = \{W_t, 0 \leq t \leq T\}$. We shall be working with the following objects:

- $\{\mathcal{F}_t\}_{0 \leq t \leq T}$ denotes the augmented Brownian filtration. We denote by \mathcal{P} the previsible σ -algebra on $[0, T] \times \Omega$. Also $\mathcal{B}(\mathbb{R}^q)$ is the Borel σ -algebra on the Euclidean space \mathbb{R}^q .
- $\mathbb{L}^p, p \geq 1$ stands for the \mathcal{F}_T measurable, p -integrable random variables, *i.e.* for any $X : \Omega \rightarrow \mathbb{R}$ such that

$$\|X\|_p := \mathbb{E}[|X|^p]^{1/p} < \infty.$$

- $\mathbb{H}_T^{p,d}$ denotes all \mathcal{P} measurable processes $\psi : [0, T] \times \Omega \rightarrow \mathbb{R}^d$ such that

$$\|\psi\|_p^p := \mathbb{E} \int_0^T |\psi_t|^p dt < \infty.$$

- For any vector $x \in \mathbb{R}^q, |x|$ stands for its Euclidean norm, while $x \cdot y$ denotes the inner product of two such vectors. Moreover for any matrix A with real valued entries, its Euclidean norm is given by $|A| := \sqrt{\text{tr}(AA^*)}$, where $*$ denotes transposition of the matrix.

Assumptions.

(H1a): The coefficients of the forward SDE $V_i : \mathbb{R}^q \rightarrow \mathbb{R}^q, i = 0, 1, \dots, q$ are uniformly Lipschitz, *i.e.* there exists a constant K such that

$$|V_i(x) - V_i(y)| \leq K|x - y|, \quad \forall i = 0, \dots, q, \quad x, y \in \mathbb{R}^q.$$

Moreover the matrix $V = (V_1, \dots, V_q)$, *i.e.* the matrix whose columns are V_1, \dots, V_q , is elliptic.

(H1b): The vector fields $V_i, i = 0, \dots, q$, belong to $C_b^\infty(\mathbb{R}^q)$ the space of bounded infinitely differentiable functions with all partial derivatives bounded.

Under assumption (H1a) the forward equation (1.2) is known to possess a unique solution such that $\mathbb{E}[\sup_{0 \leq t \leq T} |X_t|^2] < \infty$ (see for example Chap. 5 of [29]). We shall be working under (H1a) throughout. Some of the algorithms that we present in this article, require extra smoothness in the vector fields appearing in the forward equation (1.2) and in that case we will assume (H1b) as well. We will be explicit as to when (H1b) is in force.

2. BACKWARD SDEs AND FORWARD-BACKWARD SDEs

Let ξ be a random variable, $\xi \in \mathbb{L}^2$ and $f : [0, T] \times \Omega \times \mathbb{R} \times \mathbb{R}^q \rightarrow \mathbb{R}$ be a function measurable with respect to the σ -field $\mathcal{P} \times \mathcal{B}(\mathbb{R}) \times \mathcal{B}(\mathbb{R}^d)$ such that $f(\cdot, \cdot, 0, 0) \in \mathbb{H}_T^{2,1}$. Pardoux and Peng proved in [40] that there exists a unique pair of \mathcal{F}_t -adapted, progressively measurable processes (Y, Z) , taking values in \mathbb{R}, \mathbb{R}^d respectively, such that

$$\mathbb{E} \left[\sup_{t \leq s \leq T} |Y_s|^2 \right] < \infty, \quad Z_t \in \mathbb{H}_T^{2,q},$$

which satisfy the equation

$$Y_t = \xi + \int_t^T f(s, \omega, Y_s, Z_s) ds - \int_t^T Z_s \cdot dW_s, \quad t \in [0, T]. \tag{2.1}$$

Heuristically, one starts at the (final) time T with $Y_T = \xi$ and then Y is constructed backwards in time for the whole of the interval $[0, T]$. That is why equation (2.1) is suitably called a *backward* SDE. The remarkable fact about the result is that the process Y is constructed in such a way that it is adapted to the (forward) filtration $\{\mathcal{F}_t\}_{0 \leq t \leq T}$ throughout $[0, T]$. It is the freedom to choose the process Z that makes it possible to construct a non-anticipative solution to equation (2.1).

Pardoux and Peng proved the existence and uniqueness of the solution of (2.1) under a uniform Lipschitz assumptions on the *driver* of the BSDE f in its spatial variables. This assumption has since been relaxed. For instance the case of a BSDE with a driver of quadratic growth in its z variable, but a bounded terminal condition has been studied in [30,31]. See also [7] for a study of this case without the assumption of boundedness on ξ .

Though we shall only be looking at systems where the filtration is generated by a Brownian motion, let us remark that this condition may also be relaxed. For example El Karoui and Huang [15] study a BSDE driven by a general càdlàg martingale, and in [3,47] the question of existence and uniqueness to a equation of the form (2.1) in the presence of random jumps, is addressed.

A case of particular interest, is that where the final value ξ in (2.1) comes from a (forward) stochastic differential equation (SDE). On the one hand, this provides us with an ideal framework for financial applications as we shall see in Section 3. On the other hand, it allows us to build the Feynman-Kac representation for the solution of the Cauchy problem (1.3). To be precise, let us assume that there exists a real valued functional Φ such that $\xi = \Phi(X^{s,x})$ and that the driver takes the special form $f(t, X_t, Y_t, Z_t)$ for a deterministic function $f : [0, T] \times \mathbb{R}^q \times \mathbb{R} \times \mathbb{R}^q \rightarrow \mathbb{R}$, where $X^{t,x}$ is the solution to (1.2). Observe that any randomness on the backward equation now comes only through the dependence of the driver and the terminal condition on the *state* variable x . On these coefficients we make the following:

Assumptions II.

(H2a): The driver of the BSDE $f : [0, 1] \times \mathbb{R}^q \times \mathbb{R} \times \mathbb{R}^q \rightarrow \mathbb{R}$ is uniformly Lipschitz in its spatial variables and Hölder continuous with respect to time, *i.e.* there exists a constant K such that

$$|f(t, x, y, z) - f(s, x', y', z')| \leq K(|t - s|^{1/2} + |x - x'| + |y - y'| + |z - z'|).$$

The terminal condition $\Phi(\cdot)$ is also uniformly Lipschitz.

(H2b): The driver of the BSDE f belongs to $C_b^{1,m}([0, T] \times \mathbb{R}^q \times \mathbb{R} \times \mathbb{R}^q)$, *i.e.* it is one time continuously differentiable with respect to time and m times with respect to its spatial variables. All derivatives are assumed bounded.

Again, we are primarily working under (H2a). The additional assumption (H2b) shall be needed only for the algorithm presented in Section 8. By pairing up equations (1.2) and (2.1) we obtain a so-called forward-backward stochastic differential equation (FBSDE):

$$\begin{cases} dX_s^{t,x} &= V_0(X_s^{t,x})ds + \sum_{j=1}^q V_j(X_s^{t,x})dW_s^j, & s \in [t, T] \\ -dY_s^{t,x} &= f(s, X_s^{t,x}, Y_s^{t,x}, Z_s^{t,x})ds - Z_s^{t,x} \cdot dW_s, & s \in [t, T] \\ X_t^{t,x} &= x, \\ Y_T^{t,x} &= \Phi(X_T^{t,x}). \end{cases} \tag{2.2}$$

Under assumptions (H1a) and (H2a), there exists a triplet of progressively measurable processes $(X_s^{t,x}, Y_s^{t,x}, Z_s^{t,x})$ that solves (2.2) such that

$$\mathbb{E} \left[\sup_{t \leq s \leq T} |X_s^{t,x}|^2 + \sup_{t \leq s \leq T} |Y_s^{t,x}|^2 + \int_t^T |Z_s^{t,x}|^2 ds \right] < \infty.$$

As we shall see, the solution of the system (2.2) provides us with an extension to the Feynman-Kac representation for semilinear PDEs⁴.

A closely related class of SDEs are backward stochastic differential equations with reflection with respect to a continuous time barrier (RBSDEs henceforth). RBSDEs are the stochastic counterpart of obstacle problems. The algorithms presented in Sections 5 and 6 are applicable to BSDEs as well as RBSDEs (in fact the first of these was designed with such problems in mind). Hence, it makes sense to briefly present here the setup for reflected BSDEs.

Consider, as above the BSDE (2.1), where the vector fields satisfy assumption (H1a). We also consider a terminal condition ξ and a driver f as in (2.2) and also a real valued continuous time process S_t such that $\mathbb{E} [(S_t^+)^2] < \infty$, called the obstacle. We call a triplet (Y, Z, K) of \mathcal{F}_t -adapted, progressively measurable $\mathbb{R} \times \mathbb{R}^d \times \mathbb{R}_+$ -valued processes, a solution to a reflected BSDE if:

- (1) P -almost surely,

$$\begin{aligned} Y_s &= \xi + \int_s^T f(u, X_u, Y_u, Z_u) du + K_T - K_s - \int_s^T Z_u \cdot dW_u \\ Y_s &\geq S_s \end{aligned} \quad s \in [0, T]. \tag{2.3}$$

- (2) The process K_t is continuous, non decreasing, $K_0 = 0$ and

$$\int_0^T (Y_s - S_s) dK_s = 0.$$

- (3) The processes (Y, Z, K) satisfy the integrability conditions $Z \in \mathbb{H}_T^{2,q}$, $K_T \in \mathbb{L}^2$. As a consequence we also have that $\mathbb{E} [\sup_{t \leq s \leq T} |Y_s|^2] < \infty$.

The existence and uniqueness for solutions of equations of the form established by El Karaoui *et al.* [18]. The idea behind a RBSDES is as follows: As long as the solution stays above a certain barrier (modelled by the process S) it satisfies (2.1). When it hits the barrier, it is reflected back by means of the process K . As a result, the solution always stays above S . The action of the process K is minimal, in the sense that it occurs only in the critical situation, when $Y = S$.

A solution to a RBSDE connects naturally to optimal stopping problems (a fact which beyond its theoretical value, becomes also important when designing algorithms for solving RBSDEs as we shall see in Sect. 5). It is shown in Proposition 2.3 of [18] that the process Y is related to the *Snell envelope* by

$$Y_t = \operatorname{ess\,sup}_{\tau \in \mathcal{T}_t} \mathbb{E} \left[\int_t^\tau f(s, X_s, Y_s, Z_s) ds + S_\tau 1_{\tau < T} + \xi 1_{\tau = T} \mid \mathcal{F}_t \right], \tag{2.4}$$

where \mathcal{T}_t denotes the collection of all stopping times taking values in $[t, T]$.

⁴The FBSE obtained in this manner is called decoupled because the backward components (Y, Z) do not influence the dynamics of the forward part X . The coupled version of (2.2) can be found for example in [38,42] and it provides a Feynman-Kac representation to a class of quasilinear PDEs. Recently, there has also been progress to the direction of obtaining stochastic representations for fully nonlinear parabolic PDEs in the work of Cheridito *et al.* [9]. In this survey we shall restrict ourselves to systems of the form (2.2) or, to put it differently, to the stochastic counterpart of the Cauchy problem (1.3).

2.1. Connections with PDEs

Let $u \in C^{1,2}([0, T] \times \mathbb{R}^q)$ be a real valued function which is solution of the Cauchy problem (1.3), where the second order differential operator L is the infinitesimal generator of the forward part of (2.2). A straightforward application of Itô’s formula to the process $u(t, X_t^{s,x})$ shows that solution $(Y_t^{s,x}, Z_t^{s,x})$ to the backward part of (2.2) may be identified respectively with $(u(t, X_t^{s,x}), \nabla u(t, X_t^{s,x})V(X_t^{s,x}))$. In effect, we may identify the solution of the Cauchy problem (1.3) as $u(s, x) = Y_s^{s,x}$.

Under (H1a), (H2a) one can not expect that equation (1.3) will have a solution in the classical sense. However, it can be shown that there exists a solution in the viscosity sense. Pardoux and Peng [41] establish the link between the flow of the system (2.2) and the solution of (1.3) in the viscosity sense. Their work has been extended further by Ma *et al.* [36,37].

Theorem 2.1.

The smooth case [41]:

Define the function $u(s, x) = Y_s^{s,x}$, $(s, x) \in [0, T] \times \mathbb{R}^q$. Assume that assumptions (H1b) and (H2b) are in force with $m = 3$ for the latter, and also that $\Phi \in C_b^3(\mathbb{R}^n)$. Then $u \in C^{1,2}([0, T] \times \mathbb{R}^q)$ and solves (1.3) in the classical sense.

The viscosity case [36,41]:

If only (H1a) and (H2a) are in force, u is the unique viscosity solution to (1.3). If in addition the driver f and the terminal condition Φ are continuously differentiable with bounded first order partial derivatives then u , the viscosity solution of (1.3), is continuously differentiable with bounded first order partial derivatives in the spatial variable. Moreover $Z_s^{s,x} = \nabla u(s, x)V(x)$.

In the presence of a reflection, the forward backward system associates to an obstacle problem. Consider the system (2.3) where the reflection is determined through a function of the underlying diffusion, *i.e.* $S_t = h(t, X_t^{s,x})$ where h is a continuous function in (t, x) satisfying the polynomial growth condition

$$|h(t, x)| \leq K(1 + |x|^p), \quad t \in [0, T], x \in \mathbb{R}^q,$$

where p is a given positive integer. We also assume that $h(t, x) \leq \Phi(x)$, $x \in \mathbb{R}^q$. A solution to an obstacle problem in the weak (viscosity) sense is a continuous function u such that for any $(t, x) \in [0, T] \times \mathbb{R}^q$:

$$\max((\partial_t + L)u(t, x) + f(t, x, u(t, x), (\nabla u V)(t, x)), h(t, x) - u(t, x)) = 0 \tag{2.5}$$

and $u(T, x) = \Phi(x)$. It is shown in [18] that $u(s, x) = Y_s^{s,x}$ is the unique viscosity solution to (2.5) within the class of solutions with at most polynomial growth at infinity, where $Y_t^{s,x}$ is the corresponding flow associated to equation to (2.3), constructed in a similar manner to (2.2). With their work [37], Ma and Zhang recover a representation for the gradient of a solution to an obstacle problem similarly to the second part of Theorem 2.1. Naturally, the latter is valid under an additional regularity assumption for the obstacle function h .

3. APPLICATIONS IN FINANCE

The theory of BSDE has found many applications in the field of Mathematical Finance, primarily because it is suited for the study of financial problems where assumptions of the classical Black-Scholes theory do not hold. In this section, we briefly describe a few such situations, primerey to motivate the interest in BSDEs theory. It is by no means a complete list of references nor do we intend to present a sound introduction in the field of mathematical finance. For a more complete list of the various financial applications of BSDEs, one should consult the excellent survey article of [20] (which we follow closely).

3.1. The market framework

The typical model for a continuous time market is that of assuming that traded risky assets evolve as a geometric Brownian motion. We take up this assumption to keep things as simple as possible. Let us consider a market with $q + 1$ traded securities. One of these securities should be understood as being risk free (*e.g.* a bond), where as the rest may be interpreted as the stocks. The equations governing the evolution of these assets are

$$\begin{aligned} dX_t^i &= V_0^i(t)X_t^i dt + \sum_{j=1}^q V_j^i(t)X_t^i dW_t^j, \quad i = 1, \dots, q \\ dX_t^0 &= r_t X_t^0 dt \end{aligned} \tag{3.1}$$

where $V_i, i = 0, \dots, q$ are \mathcal{F}_t -adapted bounded processes. The process r models the risk free interest rate. We assume that r is a non-negative \mathcal{F}_t -adapted, progressively measurable bounded process. We also denote by $\theta_t \in \mathbb{R}^q$ the risk premium process, *i.e.* θ_t^i measures the gain in excess of the risk free interest rate per volatility unit, that an investor expects from stock i :

$$\theta_t = V^{-1}(t)(V_0(t) - r_t \underline{1}), \quad (dP \times dt) - a.s.,$$

where $\underline{1}$ is the vector with all entries equal to 1. We now consider a small investor (*i.e.* one whose actions do not affect market prices), who wishes to invest in the market modelled by the system (3.1) over a finite time horizon $[0, T]$. At time 0 her wealth is assumed to be Y_0 and we denote the portfolio process by $\pi_t = (\pi_t^1, \dots, \pi_t^q)$, *i.e.* π_t^i is the amount invested at time t in the i -th stock. Clearly, decisions made at time t , can only be based on the available information at time t , hence π is a predictable process. If there are no cash inflows or outflows from the portfolio up until time T , then the wealth process $Y_t = \sum_{i=0}^q \pi_t^i$, where $\pi_t^0 = Y_t - \sum_{i=1}^q \pi_t^i$ denotes the amount of money invested in the bond, must satisfy

$$\begin{aligned} dY_t &= \pi_t^i \frac{dP_t^i}{P_t^i} \\ &= r_t Y_t dt + \pi_t \cdot (V_0(t) - r_t \underline{1}) dt + \pi_t \cdot V_t dW_t \end{aligned} \tag{3.2}$$

i.e. any infinitesimal gain or loss realized between times t and $t + dt$ is determined only by her decision at time t and the infinitesimal market moves. A strategy (V, π) whose dynamics satisfy equation (3.2) is called *self-financing*.

3.2. Pricing a contingent claim

A contingent claim of European type is a contract that gives its owner a random payoff at time T modelled by an \mathcal{F}_T -measurable random variable ξ . In an arbitrage-free economy, pricing a contingent claim ξ is equivalent to constructing a self-financing trading strategy for which the terminal wealth, is equal to ξ almost surely. Assume that we are given such a hedging strategy and we form the dynamics of the corresponding wealth process according to (3.2) with initial wealth Y_0 . Since $Y_T = \xi$ by no arbitrage arguments we can deduce that the fair price for ξ at time 0 is Y_0 . Of course, it is not always possible to construct such a hedging strategy, nor is it *a priori* certain that such a strategy will be unique. However, one can show using BSDEs that given a contingent claim $\xi \in \mathbb{L}^2$, there exists a unique hedging strategy π in the class of strategies that satisfy certain integrability conditions. Viewing equation (3.2) as a BSDE with terminal condition ξ , we see that $\pi_t V_t$ can be identified with the Z -part of its solution. In this case Y_0 is the fair value for ξ at time 0. This is an alternative to the classical arguments of Black-Scholes theory for pricing a contingent claims. Observe that here, using the theory of BSDEs, we do not have to appeal to any change of measure arguments. Moreover, this approach can be extended to situations where the Black-Scholes assumptions do not hold and hence classical techniques *via* measure changes may fail.

3.3. Differential interest rates

We now consider a situation where non linear pricing arises, where we follow [12]. Assume that we wish to hedge a contingent claim ξ in an economy where there are different interest rates for borrowing and lending. That is, we can invest in the bond to gain a risk free interest rate r_t , but if we wish to borrow money, we would have to pay an interest rate $R_t > r_t$, which is as well a bounded and adapted process. Clearly, given any trading strategy π the dynamics of the wealth process Y are no longer described by equation (3.2). Here we invest in the bond when $Y_t > \sum_{i=1}^q \pi_t^i$, but we also have to pay an interest rate R_t on the complementary set.

Starting from the self financing assumption we have that

$$\begin{aligned} dY_t &= \sum_{i=1}^q \pi_t^i \frac{dX_t^i}{X_t^i} + \left(Y_t - \sum_{i=1}^q \pi_t^i \right)_+ r_t dt - R_t \left(Y_t - \sum_{i=1}^q \pi_t^i \right)_- dt \\ &= r_t Y_t dt + \pi_t \cdot V_t \theta_t dt - (R_t - r_t) \left(Y_t - \sum_{i=1}^q \pi_t^i \right)_- dt + \pi_t \cdot V_t dW_t \end{aligned}$$

where for any real number x , $(x)_+$, $(x)_-$ are its positive and negative parts respectively. Hence the dynamics of the wealth and portfolio process are linked in a nonlinear way. The nonlinearity occurs both with respect to the wealth and the portfolio process. The results of Pardoux and Peng allow us to deduce the existence of a unique square integrable strategy that hedges the contingent claim ξ . A similar situation of nonlinear dynamics occurs in a market where there are short selling constraints (see [27]).

3.4. Pricing American options

Assume the model (3.1) for the market. An American option on the stocks $\{X^i\}_{i=1}^q$ with expiry T , is a contract which gives its owner the right to exercise the option at any time t up to the maturity time T in order to obtain $h(X_t)$, where h is a given function describing the contract. Clearly the aim is to exercise the contract at the right time, *i.e.* to choose the (stopping) time τ such that $h(X_\tau) = \max_{0 \leq t \leq T} h(X_t)$. It is shown by El Karoui *et al.* [19] that the value of such an option is given by the solution to a reflected BSDE. Presenting the complete argument is quite lengthy, so we just briefly give some key points. The interested reader is referred to [19].

We have already seen that the value Y_t^{Eur} for a simple (European style) contingent claim ξ (with or without constraints), may be viewed as the solution of a BSDE

$$-dY_t^{\text{Eur}} = f(t, Y_t^{\text{Eur}}, \pi_t) dt - \pi_t \cdot V_t dW_t, \quad Y_T^{\text{Eur}} = \xi \tag{3.3}$$

where π_t is the hedging portfolio. In the absence of constraints, we have $f(t, y, z) = -r_t y - z \cdot V_t \theta_t$.

In the American option case it is shown that the value of the contingent claim is the solution to the reflected BSDE with obstacle $h(X_t)$ corresponding to (3.3). That is, the value Y_t^{Am} of the option is the first component of the progressively measurable triplet (Y^{Am}, Z, K) that solves the reflected BSDE

$$\begin{aligned} -dY_t^{\text{Am}} &= f(t, Y_t^{\text{Am}}, Z_t) dt - dK_t - Z_t \cdot V_t dW_t, \quad Y_T^{\text{Am}} = h(X_T) \\ Y_t^{\text{Am}} &\geq h(X_t), \quad t \leq T, \text{ a.s.} \\ \int_0^T (Y_t^{\text{Am}} - h(X_t)) dK_t &= 0. \end{aligned} \tag{3.4}$$

This agrees with basic no-arbitrage financial arguments according to which the value of an American option should never be below its intrinsic value, *i.e.*, below $h(X_t)$. The process K in the solution of (3.4) is interpreted as a consumption process. Finally let us remark that the above result verifies that the price of the American option

is the smallest initial endowment by which a self-financing super-hedging strategy may be constructed⁵. Indeed, using a comparison theorem for RBSDEs (see [19]) it holds that $Y_t^{\text{Am}} \leq Y'_t$ for any $t \in [0, T]$, where Y' is the first component of any other solution (Y', Z', K') for (3.4) that does not satisfy the condition $\int_0^T (Y'_t - h(X_t^i)) dK'_t = 0$ (Y' is the wealth process of a super-hedging strategy).

To conclude, we have given a brief account on how the value of European or American options with or without constraints on the market, may be represented as solutions to BSDEs, respectively RBSDEs. Moreover, by employing Theorem 2.1, we can give a PDE interpretation of these financial instruments. To be precise, the value of a contingent claim is the solution in the viscosity sense of the Cauchy problems (1.3), respectively (2.5) which involve (possibly) nonlinear PDEs, thus extending the results of Black and Scholes.

3.5. Non linear evaluations and g-expectations

The connection between risk measures and non linear expectations using tools from the theory of BSDEs has been an area of very active research in recent years. There are numerous occasions where the linearity properties of pricing mechanisms give rise to paradoxes, and hence the interest in modern economic theory for non linear evaluation techniques remains high. To the best of our knowledge, non-linear pricing mechanisms and their connections with BSDEs and g-expectations were first considered in [17], whereas an early axiomatic approach can be found in [43]. The interested reader should consult [44] for a more detailed exposition of recent advances in this field. The most prominent applications of these dynamic (non linear) expectations is to risk measures, see for example [22] and also the study of Rosazza Giannin [46] on the connections among risk measures and g-expectations. As we have done so far, we will take a brief look on the basic notions and definitions, merely to motivate interest.

We continue to work under the same market framework that we have been using throughout this section. We then consider the following family of operators.

Definition 3.1. The family

$$\mathcal{E}_{s,t} : \mathbb{L}^2(\mathcal{F}_t) \rightarrow \mathbb{L}^2(\mathcal{F}_s), \quad s \leq t \leq T$$

is called an \mathcal{F}_t -consistent family of evaluations if it satisfies the following set of axioms:

- (A1): $\mathcal{E}_{t,t}[Y] = Y$, a.s. for $Y \in \mathcal{F}_t$, in particular $\mathcal{E}_{t,t}[c] = c$.
- (A2): $\mathcal{E}_{s,t}[\mathcal{E}_{t,v}[Y]] = \mathcal{E}_{s,t}[Y]$, $s \leq t \leq v$ a.s.
- (A3): $\mathcal{E}_{s,t}[Y] \geq \mathcal{E}_{s,t}[\bar{Y}]$ a.s. if $Y \geq \bar{Y}$ a.s.
- (A4): For all $s \leq t$ we have $1_A \mathcal{E}_{s,t}[Y] = \mathcal{E}_{s,t}[1_A Y]$, $A \in \mathcal{F}_t$.

To interpret this set of conditions, one should think that the value at time t of any financial instrument should depend only on the trajectories of the underlying prices X_s , $0 \leq s \leq t$ up to time t . Then (A1) is obvious. (A2) states that the value of an F_v measurable contingent claim Y at time s is the same as the value of a portfolio at an intermediate time t , which is equal almost surely to Y at time v . The next axiom expresses the simple economic intuition that a claim which dominates another claim a.s. at time T , should also be worth more at any previous time. Finally (A4), also called the “zero-one law”, tells that at time t , we know whether the event A has occurred. Hence the value of Y is either 0 or equal to the value of Y restricted on the set A .

If instead of (A2), one considers

$$(A2'): \mathcal{E}_{s,t}[Y] = Y, \quad \text{a.s. } \forall Y \in \mathbb{L}^2(\mathcal{F}_s)$$

then, a family of operators that satisfies (A1)-(A2')-(A3)-(A4) is called an \mathcal{F}_t -consistent family of non linear expectations.

As it turns out, and is discussed in great detail in [44], one can generate non linear expectations and evaluation by means of BSDEs.

⁵A super-hedging strategy is a strategy that super-replicates the contingent claim, hence it is a strategy for which the corresponding wealth process is larger than or equal to the value of the contingent claim at exercise time.

Definition 3.2. Consider a market described by the set of equations (3.1) and assume that we are given a real valued function $g : \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d \rightarrow \mathbb{R}$ uniformly Lipschitz in all its spatial variables and such that

$$(i) \quad g(x, 0, 0) \equiv 0, \quad \forall x \in \mathbb{R}^d$$

or

$$(ii) \quad g(x, y, 0) = 0, \quad \forall x \in \mathbb{R}^d, y \in \mathbb{R}.$$

For any $t \leq s \leq T$ consider the family of operators $\mathcal{E}_{s,t}^g : \mathbb{L}^2(\mathcal{F}_t) \rightarrow \mathbb{L}^2(\mathcal{F}_s)$ with $\mathcal{E}_{s,t}^g[\xi] = Y_s$ where Y_s is the solution of the backward SDE

$$Y_s = \xi + \int_s^t g(X_u, Y_u, Z_u)du - \int_s^t Z_u dW_u.$$

$\mathcal{E}_{s,t}^g[\cdot]$ is called a g-expectation.

It is shown in [44] that if the generator g of a BSDE satisfies (i) of Definition 3.2 above, then $\mathcal{E}_{s,t}^g$ is a non linear evaluation, whereas if it satisfies (ii) it is an \mathcal{F}_t consistent non linear expectation. Roughly speaking properties (A1) and (A4) follow from the uniqueness results to solutions of BSDEs. Property (A2) is known in the BSDE literature as the flow property whereas (A3) is just the comparison theorem for BSDEs. All the above can be found in [20]. Examples of g-expectations of course include all examples that we have considered thus far in this section.

Of course the more interesting questions are whether all dynamic non linear expectations can be described as g expectations as well as, how can one identify a market pricing mechanism as a g-expectation. Some attempts to answer these questions are made in [45].

4. DISCRETIZING A FBSDE

Having seen a few of the applications of FBSDEs in finance and having established their connections with nonlinear PDEs, we now describe a time discretization procedure of the solution of a FBSDE. This procedure is the first step towards constructing probabilistic algorithms for nonlinear PDEs. Implicitly or explicitly, all numerical schemes for forward backward systems that we present in this article, with the exemption of the algorithm by Bally and Pages presented in Section 5, rely on the discretization of a BSDE in an Euler scheme fashion, obtained independently in [5,49].

The methods presented below follow the programme: We first fix a partition $\pi = \{0 = t_0 < t_1 < \dots < t_n = T\}$ of the time interval $[0, T]$ and we consider an exact sample of the forward diffusion at these points $(X_{t_1}, \dots, X_{t_n})$ if this is possible (for example if X is a geometric Brownian motion), or an approximate sample $(X_{t_1}^\pi, \dots, X_{t_n}^\pi)$ obtained by considering the corresponding Euler approximation X^π of the forward diffusion along the partition π .

Both the forward diffusion and its Euler approximation have a number of common characteristics, that are essential in the discretization of BSDEs, so we enumerate them in the following definition:

Definition 4.1. Consider a partition π of $[0, T]$ as above and let $\mathcal{R}(\pi)$ denote the class of step processes \bar{X}_t^π , constant between t_i, t_{i+1} , that satisfy the following:

- (1) There exists a $p > 1$ such that $\max_{1 \leq i \leq n} \|\bar{X}_{t_i}^\pi\|_p < \infty$.
- (2) $\{\bar{X}_{t_i}^\pi\}_{0 \leq i \leq n}$ is a Markov chain.
- (3) For any function $h : \mathbb{R}^q \rightarrow \mathbb{R}$ such that $h(\bar{X}_{t_{i+1}}^\pi)(W_{t_{i+1}} - W_{t_i}) \in L^1$ we have

$$\mathbb{E}[h(\bar{X}_{t_{i+1}}^\pi)(W_{t_{i+1}} - W_{t_i}) | \mathcal{F}_{t_i}^{\bar{X}^\pi}] = \mathbb{E}[h(\bar{X}_{t_{i+1}}^\pi)(W_{t_{i+1}} - W_{t_i}) | \bar{X}_{t_i}^\pi],$$

where $\mathcal{F}^{\bar{X}^\pi}$ denotes the natural filtration of the process \bar{X}^π .

Given a process $\bar{X}^\pi \in \mathcal{R}(\pi)$, the next step is to construct the corresponding discrete versions of the processes (Y, Z) . To see how this must be done, let us assume for simplicity, that the choice of process within the class $\mathcal{R}(\pi)$ is the diffusion itself. We also make the assumption that f and Φ are smooth functions with respect to their spatial variables and have bounded partial derivatives of all orders. At time T , we assign $Y_{t_n}^\pi = \Phi(\bar{X}_{t_n}^\pi) = \Phi(X_T) = Y_T$ and based on this we wish to construct an approximation $Y_{t_{n-1}}^\pi$ to $Y_{t_{n-1}}$. Let $\epsilon = t_n - t_{n-1}$ and observe that, by taking conditional expectations on (2.2) we have

$$\begin{aligned} Y_{t_{n-1}} &= Y_{T-\epsilon} \\ &= \mathbb{E} \left[\Phi(X_T) + \int_{T-\epsilon}^T f(s, X_s, Y_s, Z_s) ds \mid \mathcal{F}_{T-\epsilon}^X \right] \\ &= \mathbb{E} \left[\Phi(X_T) + \int_{T-\epsilon}^T f(s, X_s, u(s, X_s), (\nabla u V)(s, X_s)) ds \mid \mathcal{F}_{T-\epsilon}^X \right] \end{aligned}$$

and, by discretizing the Riemann integral using the left hand point of the interval, we obtain

$$Y_{T-\epsilon} \simeq \mathbb{E} [\Phi(X_T) \mid \mathcal{F}_{T-\epsilon}^X] + \epsilon f(T - \epsilon, x, u(T - \epsilon, X_{T-\epsilon}), \nabla u(T - \epsilon, X_{T-\epsilon}) V(X_{T-\epsilon})). \tag{4.1}$$

Since we also have that $Y_{T-\epsilon} \simeq u(T - \epsilon, X_{T-\epsilon})$ the above would be an implicit definition for an approximation of $Y_{T-\epsilon}$ but for the presence of the gradient. However, assuming that Φ is smooth enough to perform a second order Itô-Taylor expansion on $u(\cdot, X)$, we get

$$\begin{aligned} \Phi(X_T) &= u(T, X_T) \\ &= u(T - \epsilon, X_{T-\epsilon}) + (\partial_t + L)u(T - \epsilon, X_{T-\epsilon})\epsilon \\ &\quad + (\nabla u(T - \epsilon, X_{T-\epsilon}))^* V(X_{T-\epsilon})(W_T - W_{T-\epsilon}) \\ &\quad + \sum_{i,j=0}^q L_i L_j u(T - \epsilon, X_{T-\epsilon}) \int_{T-\epsilon}^T \int_{T-\epsilon}^s dW_u^j dW_s^i + O(\epsilon^{3/2}), \end{aligned}$$

where we understand that $O(\epsilon^{3/2})$ stands for terms with L^2 norms smaller than a constant multiple of $\epsilon^{3/2}$ and the first order differential operators L^j are defined as $L^j = \sum_{i=1}^q V_j^i \partial_{x_i}$ for $j = 1, \dots, q$. Hence, if we multiply by $W_T - W_{T-\epsilon}$ and condition with respect to $\mathcal{F}_{T-\epsilon}$ we obtain

$$(\nabla u(T - \epsilon, X_{T-\epsilon}))^* V(X_{T-\epsilon}) = \frac{1}{\epsilon} \mathbb{E} [\Phi(X_T) (W_T - W_{T-\epsilon}) \mid \mathcal{F}_{T-\epsilon}] + O(\epsilon^{3/2}). \tag{4.2}$$

Putting (4.1) and (4.2) together and using the Markov property of the diffusion, we may define an approximation of $Y_{t_{n-1}}$ as

$$Y_{t_{n-1}}^\pi = \mathbb{E} [\Phi(X_T) \mid X_{T-\epsilon}] + f \left(T - \epsilon, X_{T-\epsilon}, Y_{T-\epsilon}^\pi, \frac{1}{\epsilon} \mathbb{E} [\Phi(X_T) (W_T - W_{T-\epsilon}) \mid X_{T-\epsilon}] \right). \tag{4.3}$$

The above procedure is then repeated in the same manner n times until time 0 is reached. Observe that, (4.3) is also an implicit definition for the regression function that approximates $u(t_{n-1}, \cdot)$. It is this regression function that replaces Φ in (4.3) computing $Y_{t_{n-2}}^\pi$ from $Y_{t_{n-1}}^\pi$. To describe these regression functions we introduce the following family of operators.

Definition 4.2. Let $\pi = \{0 = t_0 < \dots < t_n = T\}$ be a partition of $[0, T]$ where we denote by $h_{i+1} = t_{i+1} - t_i$ and $\Delta W_{i+1} = W_{t_{i+1}} - W_{t_i}$. Given a process $\bar{X}_t^\pi \in \mathcal{R}(\pi)$ and function $g : \mathbb{R}^q \rightarrow \mathbb{R}$, we define inductively,

in a backward manner (that is for $i = n - 1, \dots, 0$), the one step operators

$$R_i g(x) = \mathbb{E}_{t_i, x}[g(\bar{X}_{t_{i+1}}^\pi)] + h_{i+1} f\left(t_i, x, R_i g(x), \frac{1}{h_{i+1}} \mathbb{E}_{t_i, x}[g(\bar{X}_{t_{i+1}}^\pi) \Delta W_{i+1}]\right), \tag{4.4}$$

where $\mathbb{E}_{t_i, x}[F] := \mathbb{E}[F | \bar{X}_{t_i}^\pi = x]$.

For this and other families of operators we use the notation

$$R_{i,j} \Phi(x) = \begin{cases} R_i \circ \dots \circ R_j \Phi(x) & i < j \\ \mathbf{1} & i \geq j \end{cases} \tag{4.5}$$

where $\mathbf{1}$ denotes the identity operator. Observe that, given an $x \in \mathbb{R}^q$ the definition (4.4) is an implicit one, i.e. $R_i g(x)$ is the unique fixed point of the mapping

$$x \rightarrow \mathbb{E}_{t_i, x}\left[g\left(\bar{X}_{t_{i+1}}^\pi\right)\right] + h_{i+1} f\left(t_i, x, R_i g(x), \frac{1}{h_{i+1}} \mathbb{E}_{t_i, x}\left[g\left(\bar{X}_{t_{i+1}}^\pi\right) \Delta W_{i+1}\right]\right).$$

This mapping is a contraction for $h_{i+1} < 1/K$ and hence $R_i g(x)$ is well defined. The definition of the corresponding family of operators $\{R_i\}_i$ used to approximate a reflected BSDE (equivalently an obstacle problem) is very similar to (4.4). We postpone the presentation of this until Section 5 where we discuss an algorithm that is particularly designed to solve obstacle problems.

As it turns out, the assumption of smoothness for f and Φ may be reduced to Lipschitz regularity, and one may use equation (4.4) as a starting point to produce a discrete time process that approximates Y_t .

Theorem 4.3 ([5,49]). *Given a partition π of $[0, T]$, consider the family of operators $\{R_i\}_{i=0}^{n-1}$ defined as in (4.4). Let $\{Y_t^\pi\}_{i=0}^{n-1}$, $\{Z_t^\pi\}_{i=0}^{n-1}$ be the random variables defined as*

$$Y_{t_i}^\pi = R_{i,n-1} \Phi(\bar{X}_{t_i}^\pi) \text{ and } Z_{t_i}^\pi = \frac{1}{h_{i+1}} \mathbb{E}_{t_i, \bar{X}_{t_i}^\pi} \left[Y_{t_{i+1}}^\pi \Delta W_{i+1} \right], \quad i = n - 1, \dots, 0$$

and (Y^π, Z^π) be the step processes

$$Y_t^\pi = \sum_{i=0}^{n-1} Y_{t_i}^\pi 1_{[t_i, t_{i+1})}(t) + \Phi(\bar{X}_{t_n}^\pi) 1_{t=T}, \quad Z_t^\pi = \sum_{i=0}^{n-1} Z_{t_i}^\pi 1_{[t_i, t_{i+1})}(t)$$

for $t \in [0, T]$. Then there exists a constant C independent of the partition π such that

$$\sup_{0 \leq t \leq T} \mathbb{E} [|Y_t - Y_t^\pi|^2] + \mathbb{E} \int_0^T |Z_t - Z_t^\pi|^2 dt \leq C|\pi|.$$

In particular $|Y_0 - Y_0^\pi| = O(|\pi|^{1/2})$.

Remark 4.4. In Theorem 4.3, the process Y is approximated pointwise. In particular Y_0^π is indeed an approximation of Y_0 . By contrast, the process Z is only approximated in the $L^2(d\mathbb{P} \times dt)$ -norm. Hence it is not necessarily the case that, for example, Z_0^π is an approximation of Z_0 . When working only under the Lipschitz assumptions of Theorem 4.3, one can not hope for a robust approximation of Z . However, by assuming more regularity on the coefficients of (2.2) Gobet and Labart [24] show that it is possible to obtain estimates on the error $|Z_{t_i}^\pi - Z_{t_i}|$ similar to the ones of $|Y_{t_i}^\pi - Y_{t_i}|$ in Theorem 4.3. As most of the designers of the forthcoming algorithms choose to work under minimal smoothness assumptions, we do so as well and focus on the Y part of the solution from now on.

Remark 4.5. It is worth noting that both the Bouchard-Touzi and the Zhang proof of Theorem 4.3 rely in an essential way, on a result obtained by Zhang [48] that is known as *the L^2 regularity of Z* . The regularity of Z states that under assumptions (H1a) and (H2a), for any partition π of $[0, T]$ there exists a constant $C = C(T, K)$ independent of π such that

$$\sum_{i=1}^n \mathbb{E} \left[\int_{t_{i-1}}^{t_i} |Z_t - Z_{t_{i-1}}|^2 + |Z_t - Z_{t_i}|^2 dt \right] \leq C(1 + |x|^2)|\pi| \tag{4.6}$$

given that $X_0 = x$.

The estimates of Theorem 4.3 complete the first task of discretizing a BSDE. We note that these estimates are sharp. One can show (similar to [24] where the authors prove the result for a slightly different discretization) the following error expansion

$$Y_{t_i} - Y_{t_i}^\pi = \nabla_x u(t_i, X_{t_i})(\bar{X}_{t_i}^\pi - X_{t_i}) + G|\pi| + O(|\bar{X}_{t_i}^\pi - X_{t_i}|^2),$$

where G is a random variable with $G \in \mathbb{L}^2$. Consequently, the error is of order $|\pi|$ even when $\bar{X}^\pi = X$, that is, exact sampling is used for the forward diffusion.

We note that every backward step requires the evaluation of certain (conditional) expectations which can not be solved explicitly in most cases. Hence, to obtain a fully implementable scheme a method of computation for these expectations, must be employed. The algorithms that we present below differ in the choice for this method. The methods employed are either of a Monte Carlo (MC) type or, as is the case of the cubature method described in Section 8, rely on a weak approximation of the law of the forward diffusion.

The arguments are further complicated as the regression function used at every step on our (nonlinear) backward projection comes from the empirical estimate achieved in one step above. Hence, one should be very careful when deriving the error estimates of any particular method to avoid compounding the corresponding local errors. The exact asymptotics of the error of any algorithm will clearly depend on the choice of method for approximating the conditional expectations involved in the definition of Y_t^π .

To see how the local error propagates, we remain at a generic level and introduce an operator $\bar{\mathbb{E}}_{t_i, x}[\cdot]$ which stands for an approximation of $\mathbb{E}_{t_i, x}[\cdot] = \mathbb{E}[\cdot | \bar{X}_{t_i}^\pi = x]$. Particular examples for the operator $\bar{\mathbb{E}}_{t_i, x}[\cdot]$ are given in the subsequent sections. With such an operator at hand, we define inductively, in a backward manner (that is for $i = n - 1, \dots, 0$), the family of operators $\{\bar{R}_i\}_{i=0}^{n-1}$ as

$$\bar{R}_i g(x) = \bar{\mathbb{E}}_{t_i, x}[g(\bar{X}_{t_{i+1}}^\pi)] + h_{i+1} f \left(t_i, x, \bar{R}_i g(x), \frac{1}{h_{i+1}} \bar{\mathbb{E}}_{t_i, x}[g(\bar{X}_{t_{i+1}}^\pi) \Delta W_{i+1}] \right). \tag{4.7}$$

Then, an approximation of Y_0 is given by $\bar{Y}_{t_0}^\pi := \bar{R}_{0, n-1} \Phi(x)$ given that $X_0 = x$. To obtain an evaluation for the error for any algorithm, one needs to look at the difference $R_{0, n-1} \Phi(x) - \bar{R}_{0, n-1} \Phi(x)$. To estimate this error, one can use a Trotter product expansion, that is

$$\begin{aligned} |R_{0, n-1} \Phi(x) - \bar{R}_{0, n-1} \Phi(x)| &= \left| \sum_{i=1}^{n-1} R_{0, i-1} \bar{R}_{i, n-1} \Phi(x) - R_{0, i} \bar{R}_{i+1, n-1} \Phi(x) \right|, \text{ or} \\ &= \left| \sum_{i=1}^{n-1} \bar{R}_{0, i-1} R_{i, n-1} \Phi(x) - \bar{R}_{0, i} R_{i+1, n-1} \Phi(x) \right|. \end{aligned} \tag{4.8}$$

Assume that we choose to work with the first expansion. It is an easy consequence of the defining properties of the family $\mathcal{R}(\pi)$, Hölder's inequality and basic estimates on the moments of Brownian motion that given two functions $g_1, g_2 : \mathbb{R}^q \rightarrow \mathbb{R}$ we have

$$|R_{0, i} g_1(x) - R_{0, i} g_2(x)| \leq C_{i, p} \mathbb{E} [|(g_1 - g_2)(\bar{X}_{t_i}^\pi)|^p | \bar{X}_0^\pi = x]^{1/p} \tag{4.9}$$

with $p > 1$ and $C_{i,p}$ are positive constants such that $\sup_i C_{i,p} := C < \infty$. Plugging (4.9) in (4.8) yields

$$\begin{aligned}
 |Y_0 - \bar{Y}_0| &= |R_{0,n-1}\Phi(x) - \bar{R}_{0,n-1}\Phi(x)| \\
 &\leq C \sum_{i=1}^{n-1} \mathbb{E} [|(R_i - \bar{R}_i)\bar{R}_{i+1,n-1}\Phi(\bar{X}_{t_i}^\pi)|^p | \bar{X}_0^\pi = x]^{1/p}.
 \end{aligned}
 \tag{4.10}$$

If the operators $\bar{\mathbb{E}}_{t_i,x}[\cdot]$, $i = 0, \dots, n - 1$, $x \in \mathbb{R}^q$ are defined by means of the cubature method (see Sect. 8), a similar property to (4.9) is enjoyed by the family $\{\bar{R}_i\}_{0 \leq i \leq n-1}$. Hence we can use the second form of the Trotter expansion appearing in (4.8) to obtain

$$|R_{0,n-1}\Phi(x) - \bar{R}_{0,n-1}\Phi(x)| \leq C \sum_{i=1}^{n-1} \bar{\mathbb{E}} [|(R_i - \bar{R}_i)R_{i+1,n-1}\Phi(\bar{X}_{t_i}^\pi)|^p | \bar{X}_0^\pi = x]^{1/p}.
 \tag{4.11}$$

The choice between (4.10) and (4.11) depends on the qualitative properties of the two families and the complexities that the choice of method for $\bar{\mathbb{E}}_{t_i,x}[\cdot]$ induces in providing asymptotic estimates for the error. Either way we proceed in the following manner:

Going back to (4.4) and (4.7) and using the Lipschitz property of the driver f , we have for any Lipschitz function $g : \mathbb{R}^q \rightarrow \mathbb{R}$

$$\begin{aligned}
 R_i g(x) - \bar{R}_i g(x) &= (\mathbb{E} - \bar{\mathbb{E}})_{t_i,x} \left[g(\bar{X}_{t_{i+1}}^\pi) \right] + \nu(x)(R_i g(x) - \bar{R}_i g(x)) \\
 &\quad + \zeta(x) \cdot \frac{1}{h_{i+1}} (\mathbb{E}_{t_i,x} - \bar{\mathbb{E}}_{t_i,x}) \left[g(\bar{X}_{t_{i+1}}^\pi) \Delta W_{i+1} \right]
 \end{aligned}
 \tag{4.12}$$

with $\nu(x)$, $\zeta(x)$ taking values in \mathbb{R} , \mathbb{R}^q respectively and being bounded by K . Hence,

$$|R_i g(x) - \bar{R}_i g(x)| \leq \frac{C}{1 - h_{i+1}K} \max_{0 \leq l \leq q} \left| (\mathbb{E} - \bar{\mathbb{E}})_{t_i,x} \left[g(\bar{X}_{t_{i+1}}^\pi) \Delta W_{i+1}^l \right] \right|
 \tag{4.13}$$

provided that the partition is fine enough to guarantee $h_{i+1} < 1/K$. Given that $\bar{X}_{t_0}^\pi = x$ we can put (4.10) together with (4.13) to obtain

$$\begin{aligned}
 |Y_0^{0,x} - \bar{Y}_0^{\pi,0,x}| &= |R_{0,n-1}\Phi(x) - \bar{R}_{0,n-1}\Phi(x)| \\
 &\leq C \sum_{i=0}^{n-1} \left\| \max_{0 \leq l \leq q} \left(\mathbb{E}_{t_i, \bar{X}_{t_i}^\pi} - \bar{\mathbb{E}}_{t_i, \bar{X}_{t_i}^\pi} \right) \left[\bar{R}_{i+1,n-1}\Phi(\bar{X}_{t_{i+1}}^\pi) \Delta W_{i+1}^l \right] \right\|_p
 \end{aligned}
 \tag{4.14}$$

with $p > 1$ and C a generic constant independent of the partition. Of course, one can use (4.11) instead of (4.10) and obtain the corresponding result.

Remark 4.6. Observe that similar to (4.4), the definition (4.7) is an implicit one and given any method $\bar{\mathbb{E}}_{t_i,x}$, the computed approximating value $\bar{R}_i g(x)$ comes as the unique fixed point of the corresponding mapping. Hence to compute $\bar{R}_i g(x)$ one has to perform successive Picard iterations at every discretization point t_i . A standard argument shows that with I iterations at every time step t_i , one obtains a value which is at most Ch^I far from the unique fixed point.

5. THE QUANTIZATION METHOD

The quantization method has been introduced by Bally and Pagès in [1,2]. This method was first developed to treat the case of reflected BSDEs when the driver is independent of the z variable, presumably because at the time the L^2 regularity of Z (see Rem. 4.5) was not known.

As this algorithm was particularly designed, for the numerical solution of obstacle problems, we will provide here the equivalent formulation of the families $\{R_i\}_{0 \leq i \leq n}$, $\{\bar{R}_i\}_{0 \leq i \leq n}$ in the presence of an obstacle. This will be nothing more than a straightforward extension of what we presented in the previous section. However, note that Bally and Pagès constructed their discretization based on the representation of a solution to a RBSDE as a Snell envelope (see (5.2) below) and showed its convergence independently of [5,49].

The quantization method is an interesting mixture of analytical and MC methods. One first builds a time-space grid and computes the transition probabilities for the underlying forward diffusion X_t on this grid, by MC simulation. Once these probabilities are settled, evaluating the BSDE on the grid (*i.e.* solving the corresponding obstacle problem) reduces to solving analytically a backward dynamic programming formula. As the authors remark, the error estimates for this method depend on the dimension as analytical methods do.

The aim here is to approximate a reflected BSDE of the form:

$$\begin{aligned}
 Y_t &= \Phi(X_T) + \int_t^T f(u, X_u, Y_u)du + K_T - K_s - \int_t^T Z_u \cdot dW_u. \\
 Y_t &\geq \Phi(X_t).
 \end{aligned}
 \tag{5.1}$$

In PDE language, we wish to approximate $u(0, x)$, $x \in \mathbb{R}^q$, where u solves the obstacle problem

$$\max(\partial_t u(t, x) + Lu(t, x) + f(t, x, u(t, x)), \Phi(x) - u(t, x)) = 0, \quad u(T, x) = \Phi(x),$$

where L is the second order operator associated with the forward diffusion. Evidently, the formulation of the family of operators (4.4) needs to be adjusted so as to take into account the presence of the reflection with respect to the barrier $\Phi(X_t)$. The discretization scheme that was considered by Bally and Pagès uses as a starting point the representation of the solution of (5.1) as the Snell envelope

$$Y_t = \text{ess sup}_{\tau \in \mathcal{T}_t} \mathbb{E} \left[\int_t^\tau f(s, X_s, Y_s)ds + \Phi(X_\tau) | \mathcal{F}_t \right].
 \tag{5.2}$$

Given a partition π of $[0, T]$ we consider a process $\bar{X}^\pi \in \mathcal{R}(\pi)$ from whose distribution we can sample directly and a Lipschitz continuous function $g : \mathbb{R}^q \rightarrow \mathbb{R}$. We then set

$$R_i g(x) = \max \left(\Phi(\bar{X}_{t_i}^\pi), \mathbb{E}_{t_i, x} \left[g(\bar{X}_{t_{i+1}}^\pi) \right] + h_{i+1} f(t_i, x, R_i g(x)) \right), \quad 0 \leq i < n.
 \tag{5.3}$$

That is, $R_i g(\bar{X}_{t_i}^\pi)$ is the discrete (Φ, f) -Snell envelope of $\{\bar{X}_{t_i}^\pi\}_{0 \leq i \leq n}$. It is then easy to show that

$$R_{i, n-1} \Phi(\bar{X}_{t_i}^\pi) = \text{ess sup}_{\tau \in \bar{\mathcal{T}}_i} \mathbb{E}_{t_i, \bar{X}_{t_i}^\pi} \left[\Phi(\bar{X}_\tau) + \sum_{k=i+1}^{\bar{\tau}} f(t_k, \bar{X}_{t_k}^\pi, R_{k, n-1} \Phi(\bar{X}_{t_k}^\pi)) \right]$$

where $\bar{\mathcal{T}}_i$ denotes the collection of all $\{t_i, \dots, t_n\}$ valued stopping times and $\bar{\tau} = (n\tau)/T$. As before, we consider $Y_{t_i}^\pi = R_{i, n-1} \Phi(\bar{X}_{t_i}^\pi)$ the discrete time approximation of the process Y . For the operators (5.3), one can obtain the same error estimates as the ones of Theorem 4.3 under the general case where the coefficients are only Lipschitz continuous and the choice within the class $\mathcal{R}(\pi)$ is, for example, the Euler approximation to the forward diffusion. It is also worth noting that Bally and Pagès [1] explain how these error estimates may be improved if one assumes greater smoothness of f, h in (5.1). This fact is analysed in great detail by Gobet and Labart [24].

Having explained how the discretization of the BSDE works out in the presence of a reflection, let us now give the description of the operator $\bar{\mathbb{E}}[\cdot]$. This operator is based on the space discretization by quantization of the process chosen from the family $\mathcal{R}(\pi)$. The quantization is constructed as follows:

We consider a grid $\Gamma_k = \{x_1^k, \dots, x_{N_k}^k\}$ of points in \mathbb{R}^q given for every $k = 1, \dots, n - 1$. We denote by $\hat{X}_k := \text{Proj}_{\Gamma_k}(\bar{X}_{t_k}^\pi)$, the projection of the \mathbb{R}^q vector $\bar{X}_{t_k}^\pi$ on the grid Γ_k following the closest neighbor rule. We define p_{ij}^k to be the transition probabilities for the discrete Markov chain \hat{X} on these grids, *i.e.*

$$\begin{aligned} p_{ij}^k &:= \mathbb{P} \left[\hat{X}_{k+1} = x_j^{k+1} \mid \hat{X}_k = x_i^k \right], \quad i = 1, \dots, N_k, j = 1, \dots, N_{k+1} \\ &= \mathbb{P} \left[\hat{X}_{k+1} \in C(x_j^{k+1}) \mid \hat{X}_k \in C(x_i^k) \right] \end{aligned}$$

where $C(x_i^k)$ is the cell of x_i^k in the Voronoi tessellation,

$$C(x_i^k) := \text{Proj}_{\Gamma_k}^{-1}(x_i^k) \subset \{x : |x - x_i^k| = \min_{1 \leq l \leq N_k} |x - x_l^k|\}.$$

Let us assume for the moment that these transition probabilities can be computed explicitly. Then the operator $\bar{\mathbb{E}}[\cdot]$ is defined as

$$\begin{aligned} \bar{\mathbb{E}}_{t_k, x} \left[h(\bar{X}_{t_{k+1}}^\pi) \right] &:= \mathbb{E} \left[h(\hat{X}_{k+1}) \mid \hat{X}_k = x_i^k \right], \quad x_i^k = \text{Proj}_{\Gamma_k}(x) \\ &= \sum_{l=1}^{N_{k+1}} h(x_l^{k+1}) p_{il}^k \end{aligned}$$

for any Borel function $h : \mathbb{R}^q \rightarrow \mathbb{R}$. Similarly to (4.7), we define

$$\bar{R}_i g(x) = \max \left(\Phi(x), \bar{\mathbb{E}}_{t_i, x} \left[g(\bar{X}_{t_{i+1}}) \right] + h_{i+1} f(t_i, x, \bar{R}_i g(x)) \right), \quad i = n - 1, \dots, 0,$$

and, as before, we set $\bar{Y}_{t_i}^\pi := \bar{R}_{i, n-1} \Phi(\bar{X}_{t_i})$. It is shown in [1] that if one chooses \bar{X}^π to be the Euler approximation to the forward diffusion and assumes Lipschitz coefficients in (5.1), then

$$\max_{0 \leq k \leq n-1} \|Y_{t_k} - \bar{Y}_{t_k}^\pi\|_p = O(n^{-1/2}), \quad \forall p > 1.$$

However, two points remain to be clarified. The first one is the choice of the grids Γ_k used in the quantization of the underlying diffusion. The objective is to construct optimal grids, in the sense that they minimize the corresponding L^p - error. To be specific, let X be an \mathbb{R}^q -valued random variable. Then, for a set of points Γ in \mathbb{R}^q (a grid), the corresponding L^p quantization error is

$$\|X - \text{Proj}_\Gamma(X)\|_p := \left(\int_{\mathbb{R}^q} |X - \text{Proj}_\Gamma(X)|^p \mathbb{P}_X(dx) \right)^{1/p}. \tag{5.4}$$

The grid Γ is called optimal if it minimizes the quantity (5.4). It can be shown that there exists at least one optimal grid and there exists algorithms for the computation by simulation of an optimal grid. See [2] for a stochastic recursive algorithm for the construction of an optimal grid called *the extended Competitive Learning Vector Quantization*. Bally and Pagès [1] explain how the parameters $n, N_k, k = 1, \dots, n - 1$ may be chosen to achieve optimal error estimates depending on the smoothness that the coefficients of (5.1) enjoy. In particular, if one decides to deploy N points for the quantization of the underlying diffusion $N = N_1 + \dots + N_n$ and if the coefficients of (5.1) are only Lipschitz continuous, then $n, N_k, k = 1, \dots, n - 1$ should be chosen according to the relationships

$$\begin{aligned} n &\sim \left(\frac{2d}{d+1} C_p(x) \right)^{\frac{2}{3d+2}} N^{2/(3d+2)} \\ N_k &\sim \frac{3d+1}{2(d+1)} (k/n)^{d/(2(d+1))} \frac{N}{n}. \end{aligned}$$

The second point to be clarified is the fact that the transition probabilities are not known explicitly in most interesting cases. Thus, they need to be estimated through MC simulation introducing a further *statistical error*. Given M independent copies of \bar{X} , $\{\bar{X}^l\}_{1 \leq l \leq M}$, we define

$$\hat{p}_{ij}^k := \frac{\sum_{l=1}^M 1_{\{\hat{X}_{k+1}^l = x_j^{k+1}\}} 1_{\{\hat{X}_k^l = x_i^k\}}}{\sum_{l=1}^M 1_{\{\hat{X}_k^l = x_i^k\}}},$$

where $\hat{X}_k^l = \text{Proj}_{\Gamma_k}(\bar{X}_{t_k}^l)$. We adjust the definition of the operator $\bar{\mathbb{E}}$ according to these *empirical* probabilities in a natural manner

$$\bar{\mathbb{E}}_{t_k, x} [h(\bar{X}_{t_{k+1}})] = \sum_{l=1}^{N_{k+1}} h(x_l^{k+1}) \hat{p}_{il}^k \tag{5.5}$$

and change the definition of the family $\{\bar{R}_i\}_{0 \leq i \leq n-1}$ accordingly. The end result is that we obtain the error estimate of the form

$$|Y_{t_0} - \bar{Y}_{t_0}^\pi| \leq C \left(\frac{\sqrt{nN} \sum_{k=1}^n \|X_{t_k} - \hat{X}_k\|_2 + \sqrt{n}}{\sqrt{M}} + \frac{N^2}{M} \right),$$

where $\bar{Y}_{t_0}^\pi = \bar{R}_{0, n-1} \Phi(x)$ and $\{\bar{R}_i\}_{0 \leq i \leq n-1}$ is defined by (5.4) with $\bar{\mathbb{E}}$ given by (5.5).

6. THE MALLIAVIN CALCULUS REGRESSION METHOD

The Malliavin calculus regression method was introduced by Bouchard and Touzi in [5]. The algorithm makes use of the Malliavin integration by parts formula, to obtain an alternative representation for the required conditional expectations which, in turn allows their Monte Carlo estimation. This representation was first established in [6].

In the following we use some basic facts of Malliavin calculus which may be found in [5,6]. For a detailed exposition of the subject, the interested reader should consult the excellent monograph of [39]. *Throughout this section, the stronger assumption (H1b) on the coefficients of the forward part of (2.2), is in force.*

To simplify matters, assume X is a one dimensional diffusion and $g : \mathbb{R} \rightarrow \mathbb{R}$ is a smooth function. We also assume that $X_t \in \mathbb{D}^{1,2}$ for any $t \in [0, T]$, the space of Malliavin differentiable random variables and we denote the Malliavin derivative process of X_t by $D_s X_t$. We also fix two times $s, t \in [0, T]$, $s < t$ and we consider the regression function $r(x) = \mathbb{E}[g(X_t) | X_s = x]$. Using Bayes' rule, we may formally write $r(x)$ as

$$r(x) = \frac{\mathbb{E}[g(X_t) \epsilon_x(X_s)]}{\mathbb{E}[\epsilon_x(X_s)]}$$

where ϵ_x stands for the Dirac delta function at x . Suppose now that we are given a process h_t such that

$$\int_0^T D_u X_s h_u du = 1, \quad \int_0^T D_u X_t h_u du = 0, \quad P\text{-a.s.}$$

Then, (still only formally) considering the Dirac function as the derivative of the Heaviside function $H_x(y) := 1_{[x,\infty)}(y)$, we have that

$$\begin{aligned} \mathbb{E}[g(X_t)\epsilon_x(X_s)] &= \mathbb{E}\left[g(X_t)\frac{d}{dx}H_x(X_s)\right] = \mathbb{E}\left[g(X_t)\frac{d}{dx}H_x(X_s)\int_0^T D_u X_s h_u du\right] \\ &= \mathbb{E}\left[\int_0^T g(X_t)D_u H_x(X_s)h_u du\right] \\ &= \mathbb{E}\left[\int_0^T g'(X_t)D_u X_t h_u + g(X_t)D_u H_x(X_s)h_u du\right] \\ &= \mathbb{E}[g(X_t)H_x(X_s)\delta(h.)]. \end{aligned}$$

The last equality following from the Malliavin integration by parts formula and $\delta(h.)$ is the *Skorohod integral* of the process h_t . A similar representation holds for the denominator of $r(x)$. In fact the actual representation is a bit more involved, but the idea is as above and it constitutes of integrating up the Dirac to the Heaviside function.

For this method the choice within the class $\mathcal{R}(\pi)$ is the Euler approximation to the forward diffusion. This is necessary as we will need to simulate the underlying process at the points of the partition. If the actual diffusion can be simulated (as is the case for example for geometric Brownian motion or the Ornstein-Uhlenbeck process), then we can work with the actual diffusion.

Given a multi index $I = (i_1, \dots, i_k)$ with $1 \leq i_1 < \dots < i_k \leq d$ we denote by $|I|$ its length. For any index i we write $\partial_i = \frac{\partial}{\partial x_i}$ and $\partial_I = \partial_{i_1} \dots \partial_{i_k}$. Let \mathcal{L} denote the collection of smooth functions $\phi \in C_b^\infty(\mathbb{R}^n)$, such that

$$\sum_{|I| \leq d} \int_{\mathbb{R}^d} |x|^m \partial_I \phi(x)^2 dx < \infty, \quad \text{for some fixed integer } m. \tag{6.1}$$

Given a matrix valued process h with columns denoted by h^i and a random variable F we denote by

$$S_i^h[F] := \int_0^\infty F(h_t^i)^* dW_t$$

and for a multi index $I = (i_1, \dots, i_k)$

$$S_I^h[F] := S_{i_1}^h \circ \dots \circ S_{i_k}^h[F],$$

where the integrals are understood in the Skorohod sense. We extend the definition to $I = \emptyset$ by setting $S_\emptyset^h[F] = F$. Now, under assumption (H1b), the Euler scheme (or the diffusion itself) $\bar{X}^\pi \in \mathbb{D}^\infty$, *i.e.* it is infinitely many times differentiable in the Malliavin sense. Moreover, there exists for every $i = 0, \dots, n$ a matrix valued process $h_i(t)$ (see [5,6]) such that

$$\int_0^1 D_t \bar{X}_{t_i}^\pi h_i(t) dt = I_q, \quad \int_0^1 D_t \bar{X}_{t_{i+1}}^\pi h_i(t) dt = 0, \tag{6.2}$$

where I_q is the identity matrix in $\mathbb{R}^{q \times q}$. The representation is as follows:

Let $\rho, a : \mathbb{R}^q \rightarrow \mathbb{R}$ be real valued functions, with a being affine and ρ being random function of the form $\rho(\cdot, \chi)$ with χ a random variable independent of X^π . Assume that the random variable $F = \rho(\bar{X}_{t_{i+1}}^\pi) a(\Delta W_{i+1}) \in L^2$.

Then the operator $\bar{\mathbb{E}}[\cdot]$ considered by Bouchard and Touzi [5] is defined as

$$\bar{\mathbb{E}} [F | \bar{X}_{t_i}^\pi = x] := \frac{\mathbb{E} [Q^F[h_i, \phi_F](x)]}{\mathbb{E} [Q^1[h_i, \phi_1](x)]}, \tag{6.3}$$

for all localizing functions $\phi_1, \phi_F \in \mathcal{L}$, where for any $\phi \in \mathcal{L}$

$$Q^F[h_i, \phi](x) := H_x(\bar{X}_{t_i}^\pi) \rho(\bar{X}_{t_{i+1}}^\pi) S_{(1, \dots, d)}^{h_i} [a(\Delta W_{i+1}) \phi(\bar{X}_{t_i}^\pi - x)],$$

$h_i(t)$ is a matrix valued process that satisfies (6.2) and $H_x(y) := \prod_{i=1}^d 1_{x_i \leq y_i}$ is the Heaviside function in d -dimension. Observe the introduction of the localizing functions ϕ_1, ϕ_F in (6.3) capture the idea that, all relevant information are located in a neighborhood of x . Moreover it allows the variance reduction of the corresponding estimator. For a study on the optimal choice of localizing functions see [6].

Equation (6.3) suggests a Monte Carlo estimation for the involved conditional expectations. However, there are certain numerical complexities regarding the implementation of the algorithm. Firstly, the presence of the Heaviside function in the numerator and denominator of (6.3) is an undesirable feature. This is addressed by appealing to advanced sorting algorithms in many dimensions. Another difficulty is the fact that the Monte Carlo estimation of the denominator of (6.3) introduces integrability problems. Indeed, if \hat{Q}^F, \hat{Q}^1 are Monte Carlo estimations of Q^F, Q^1 respectively then, by the central limit theorem \hat{Q}^1 has an asymptotic Gaussian distribution, thus creating integrability problems in the ratio \hat{Q}^F / \hat{Q}^1 . The way this problem is tackled in [5] is by taking advantage of preliminary bounds available for Y and Z . It can be shown that there exist two functions $\underline{\nu}$ and $\bar{\nu}$ with quadratic growth, *i.e.*

$$\limsup_{|\pi| \rightarrow \infty} \{|\underline{\nu}(x)|, |\bar{\nu}(x)|\} < C_1 x^2 + C_0$$

with C_1, C_0 independent of π , such that

$$\underline{\nu}(x) \leq Y_{t_i}^\pi, \mathbb{E}_{t_i, x} [Y_{t_{i+1}}^\pi], \frac{1}{\sqrt{h_i}} \mathbb{E}_{t_i, x} [Y_{t_{i+1}}^\pi \Delta W_{i+1}] \leq \bar{\nu}(x). \tag{6.4}$$

We use the above bounds we construct the operator $T^\nu[F] = \underline{\nu}(x) \vee F \wedge \bar{\nu}(x)$. The Monte Carlo simulation based on (6.3) is build as follows: Consider nN copies of the approximation X_t^π of the forward diffusion, $\{X^{\pi^j}, j \in \mathcal{N}_i\}, \mathcal{N}_i = \{Ni + 1, \dots, N(i + 1)\}$. We make use of N copies at every point t_k on the grid. The realization of the operator $\bar{\mathbb{E}}_{t_i, x}$ in this algorithm is

$$\bar{\mathbb{E}}_{t_i, x} [g(\bar{X}_{t_{i+1}}^\pi) \Delta W_{i+1}^l] = T^\nu [\hat{\mathbb{E}}_{t_i, x} [g(\bar{X}_{t_{i+1}}^\pi) \Delta W_{i+1}^l]],$$

with

$$\hat{\mathbb{E}}_{t_i, x} [g(\bar{X}_{t_{i+1}}^\pi) \Delta W_{i+1}^l] = \frac{\hat{Q}^F[h_i, \phi^F](x)}{\hat{Q}^1[h_i, \phi^1](x)} \tag{6.5}$$

where ϕ^1, ϕ^F are possibly different localizing functions in \mathcal{L}_0 , and

$$\begin{aligned} \hat{Q}_i^F[h_i, \phi^F](x) &= \frac{1}{N} \sum_{k \in \mathcal{N}_{i+1}} H_x(X_{t_i}^{\pi^{(k)}}) g(X_{t_{i+1}}^{\pi^{(k)}}) S^{h_i^{(k)}} [\phi^F(X_{t_{i+1}}^{\pi^{(k)}} - x) \Delta W_{i+1}^{l^{(k)}}] \\ \hat{Q}_i^1[h_i, \phi^1](x) &= \frac{1}{N} \sum_{k \in \mathcal{N}_{i+1}} H_x(X_{t_i}^{\pi^{(k)}}) S^{h_i^{(k)}} [\phi^1(X_{t_{i+1}}^{\pi^{(k)}} - x)] \end{aligned}$$

where $\Delta W_{i+1}^{l(k)}$ is the l -th entry of the k -th copy of the forward diffusion. Similarly $X_{t_i}^{\pi(k)}$, $h_i^{(k)}$ stand for the evaluation of the Euler scheme and the random variable h_i using this copy. At time t_0 , \mathbb{E} is just the truncation of the empirical mean

$$\bar{\mathbb{E}}_{0,x} [g(X_{t_1}^{\pi} \delta W_1^l)] = T^\nu \left(\frac{1}{N} \sum_{k \in \mathcal{N}_1} g(X_{t_{i+1}}^{\pi(k)}) \Delta W_1^{l(k)} \right).$$

Combining (4.4) with the above MC simulation and the preliminary bounds (6.4), provides us with the approximation $\bar{Y}_{t_i}^\pi = \bar{R}_{i,n-1} \Phi(X_{t_i}^\pi)$ for Y_{t_i} and in particular $\bar{R}_{0,n-1} \Phi(x)$ for Y_0 .

The Malliavin calculus algorithm presents two numerical complexities. The first one is due to the presence of the Heaviside function in (6.3). This is handled by appealing to advanced sorting algorithms in many dimensions. The second one is the fact that the Malliavin weights $S^{h_i} \left[\phi \left(X_{t_{i+1}}^{\pi(k)} - x \right) \Delta W_{i+1}^l \right]$ constitute of an exploding number of terms, once transformed to Itô integrals, as the dimension increases, thus making its computation intractable. This point is handled in [11] by truncating the process $h_i(t)$ and considering

$$\bar{h}_i = |\pi|^{-1} \left(\sigma^{-1}(\bar{X}_{t_{i-1}}^\pi) \mathbf{1}_{\{[t_{i-1}, t_i)\}} - \sigma^{-1}(\bar{X}_{t_i}^\pi) \mathbf{1}_{\{[t_i, t_{i+1})\}} \right), \tag{6.6}$$

in its place, $S^{\bar{h}_i}$ in place of S^{h_i} , and form a backward induction scheme based on these new weights. Such a substitution will require only $1/2^d$ of the original terms to be computed. Moreover, it is shown that by scaling the localizing function accordingly no precision is lost. The end result is that with this algorithm we obtain an error estimate of the form

$$\max_{0 \leq i \leq n} \|Y_{t_i} - \bar{Y}_{t_i}^\pi\|_p \leq \begin{cases} \frac{C_1}{\sqrt{|\pi|}} + C_2 \frac{|\pi|^{-d/4}}{N^{1/2}} & \text{if } h(t) \text{ is used} \\ \frac{C_1}{\sqrt{|\pi|}} + C_2 |\pi|^{\frac{1}{p}((\frac{1}{2} + \alpha) - 1)} + C_3 \frac{|\pi|^{-1 - \alpha d/2p}}{N^{1/2p}} & \text{if } \bar{h}(t) \text{ is used} \end{cases} \tag{6.7}$$

where C_1, C_2, C_3 are independent of the partition and with $p > 1, \alpha > 1/2$.

Remark 6.1. Let us comment that the algorithm of Bouchard and Touzi may be adapted to the case of reflected BSDEs. One merely has to change the definition of the family $\{R_i\}_{0 \leq i \leq n-1}$ to account for the reflection. The details are very similar to the previous section and are given in the last part of [5]

7. THE REGRESSION ON FUNCTION BASES METHOD

The regression on function bases method was introduced by Gobet *et al.* [25]. The algorithm is based on estimating the involved (conditional) expectations by considering the projections of the random variables on a (finite) function basis and then estimating these projections by MC simulation. This idea was first introduced by Longstaff and Schwartz [32] to address the problem of pricing American type options and further applied successfully in the nonlinear framework of BSDEs. It has also been extended to treat the case of generalized BSDEs where the underlying forward process may be a jump diffusion but we shall not present the details for this here. The interested reader should consult [26]. Let us also mention that this algorithm is applicable to BSDEs where the terminal condition can be path dependent, *i.e.* situations where $\xi = \Phi(X)$ for appropriate functionals Φ . Thus it links naturally to the problem of pricing path dependent contingent claims, such as look-back and barrier options, in the presence of market imperfections. In accordance to our practice so far, we will only consider simple terminal payoffs thus keeping the link with non linear PDEs⁶.

⁶The extension to path dependent payoffs for this algorithm is not difficult and it amounts to adding extra state variables so as to make the implicit dynamics of the system Markovian. For example, if the terminal condition was of the form $\xi = \Phi(X_T, \max_{0 \leq t \leq T} X_t)$ which is the case of European barrier option, then given a discretization $\bar{X}_{t_i}^\pi$ of the forward process, one should consider as state variables the augmented pair $(\bar{X}_{t_i}^\pi, \max_{0 \leq i \leq j} \bar{X}_{t_j}^\pi)_{0 \leq i \leq n}$ (see [25]).

Let $\mathcal{U}_i = \{U_j | U_j \in m\mathcal{F}_{t_i}, j = 1, \dots, k\}$ be a family of random variables measurable with respect to \mathcal{F}_{t_i} . Given another integrable random variable F we denote by $\text{Proj}_{\mathcal{U}_i}[F]$ the projection of F on the space spanned by \mathcal{U}_i and we also write $\text{Res}_{\mathcal{U}_i}[F]$ for the residue of this projection, *i.e.* $\text{Res}_{\mathcal{U}_i}[F] = F - \text{Proj}_{\mathcal{U}_i}[F]$. The idea behind the algorithm of Gobet *et al.* [25] is simple and yet powerful. We consider $d + 1$ families of measurable functions $\{(p_{l,t_i}(x))_{0 \leq l \leq d}\}_{0 \leq i \leq n-1}$ available at times $\{t_i\}_{0 \leq i \leq n-1}$. If $\{\bar{X}_{t_i}^\pi\}_{0 \leq i \leq n}$ is a directly simulable approximation of the forward diffusion then, $(p_{l,t_i}(\bar{X}_{t_i}^\pi))$ is going to be used to estimate $\mathbb{E}_{t_i, \bar{X}_{t_i}^\pi} \left[R_{i+1, n-1} \Phi(\bar{X}_{t_{i+1}}^\pi) \Delta W_{i+1}^l \right]$ with the understanding that $\Delta W_{i+1}^0 = 1$. For this algorithm, the operator $\bar{\mathbb{E}}$ is given by ($l = 0, \dots, d$):

$$\bar{\mathbb{E}}_{t_i, \bar{X}_{t_i}^\pi} \left[R_{i+1, n-1} \Phi(\bar{X}_{t_{i+1}}^\pi) \Delta W_{i+1}^l \right] = \text{Proj}_{\langle p_{l,t_i}(\bar{X}_{t_i}^\pi) \rangle} \left[R_{i+1, n-1} \Phi(\bar{X}_{t_{i+1}}^\pi) \Delta W_{i+1}^l \right], \tag{7.1}$$

where $\langle p_{l,t_i}(\bar{X}_{t_i}^\pi) \rangle$ denotes the subspace of $\mathbb{L}^2(\mathcal{F}_{t_i})$ spanned by $\{p_{l,t_i}(\bar{X}_{t_i}^\pi)\}$. At this level, the operator $\bar{\mathbb{E}}$ is still only a theoretical one. The goal is to estimate the coefficients of these projections *via* MC simulation, *i.e.* the $d + 1$ families of real numbers $\{(\alpha_{l,t_i})\}_{0 \leq l \leq d}$, such that

$$\bar{\mathbb{E}}_{t_i, \bar{X}_{t_i}^\pi} \left[R_{i+1, n-1} \Phi(\bar{X}_{t_{i+1}}^\pi) \Delta W_{i+1}^l \right] = \alpha_{l,t_i} \cdot p_{l,t_i}(\bar{X}_{t_i}^\pi), \quad \forall l = 0, \dots, d.$$

This problem may be formulated in different ways. As the authors in [25] remark, it is more convenient to view it as a *least squares problem*. Indeed, given that at time t_i we “know” $\bar{X}_{t_i}^\pi, R_{i+1, n-1} \Phi(\bar{X}_{t_{i+1}}^\pi)$ then $R_{i, n-1} \Phi(\bar{X}_{t_i}^\pi)$ and $\mathbb{E}_{t_i, \bar{X}_{t_i}^\pi} [R_{i+1, n-1} \Phi(\bar{X}_{t_{i+1}}^\pi) \Delta W_{i+1}^l]$ minimize

$$\mathbb{E} \left[\left(Y_{t_{i+1}}^\pi - \psi + h_i f(t_i, \bar{X}_{t_i}^\pi, \psi, \zeta) - \zeta \cdot \Delta W_{i+1} \right)^2 \right] \tag{7.2}$$

among all $\mathbb{L}^2(\mathcal{F}_{t_i})$ random variables ψ and ζ taking values in \mathbb{R}, \mathbb{R}^d respectively. The coefficients $\{(\alpha_{l,t_i})\}_{0 \leq l \leq d}$ of the projections on the finite function bases are obtained iteratively through (7.2) (recall that definition (4.4) is implicit and hence, Picard iterations are necessary to obtain the value of $R_{i, n-1} \Phi(x)$). In particular, let us denote by $\alpha_{l,i}^{j,J}, 0 \leq l \leq d$ the coefficients obtained by performing j iterations on the least squares problem (7.2) at time t_i where J iterations have been performed at any time after t_i . We set $\alpha_{l,i}^{0,J} = 0$ for all l and we also denote by $f(i, \alpha_i^{j-1,J}) = f\left(t_i, \bar{X}_{t_i}^\pi, \alpha_{0,i}^{j-1,J} \cdot p_{0,i}, \dots, \alpha_{d,i}^{j-1,J} \cdot p_{d,i}\right)$ where we suppress the dependence of the function bases families on the random variables $\{\bar{X}_{t_i}^\pi\}_{1 \leq i \leq n}$, *i.e.* $p_{l,i} \equiv p_{l,i}(\bar{X}_{t_i}^\pi)$. Then $\alpha_{l,i}^{j,J}, 0 \leq l \leq d$ are obtained as the argmin of the minimization problem

$$\min_{(\alpha_0, \dots, \alpha_d)} \mathbb{E} \left[\alpha_{0,i+1}^{J,J} p_{0,i+1} - \alpha_0 p_{0,i} + h_{i+1} f(i, \alpha_i^{j-1,J}) - \sum_{l=1}^d \alpha_l p_{l,i} \Delta W_{i+1}^l \right]. \tag{7.3}$$

If there are multiple solutions to the above then, we may choose the one of minimal norm. It is shown in [25] that the error induced by truncating the expectation projection operator from a projection on the space $\mathbb{L}^2(\mathcal{F}_{t_i})$ to the smaller space spanned by the function bases is

$$\begin{aligned} \max_{0 \leq i \leq n} \|R_{i, n-1} \Phi(\bar{X}_{t_i}^\pi) - \bar{R}_{i, n-1} \Phi(\bar{X}_{t_i}^\pi)\|_2^2 &\leq Ch^{2J-2} + C \sum_{i=0}^{n-1} \mathbb{E} [|\mathcal{R}_{p_{0,i}}(R_{i, n-1} \Phi(\bar{X}_{t_i}^\pi))|^2] \\ &+ Ch \sum_{i=0}^{n-1} \sum_{l=1}^d \mathbb{E} \left[\left| \mathcal{R}_{p_{l,i}} \left(\frac{1}{h_{i+1}} \mathbb{E} \left[R_{i+1, n-1} \Phi(\bar{X}_{t_{i+1}}^\pi) \Delta W_{i+1}^l \right] \right) \right|^2 \right]. \end{aligned} \tag{7.4}$$

In (7.4), J is the number of iterations performed at every discretization step. Observe that at this stage the value $J = 2$ is relevant, in order to keep the same error estimate as the one induced by the discretization

of the BSDE. Moreover, observe that the error is quantified in terms of the residue of the projection of the random variables $R_{i,n-1}\Phi(\bar{X}_{t_i}^\pi)$ rather than in terms of $\bar{R}_{i,n-1}\Phi(\bar{X}_{t_i}^\pi)$. In other words, at this step one is able to use a Trotter product expansion of the error, as in (4.11). This greatly facilitates the analysis of the error with respect to the influence of the function bases as in the alternative Trotter expansion, the influence of the function basis would be also hidden inside the residue of the projection.

Finally, one computes the coefficients $(\alpha_{l,i}^{J,J})_{0 \leq l \leq d}$, $0 \leq i \leq n$, by means of MC simulation. To that end consider M copies of \bar{X}^π denoted by $(\bar{X}^{m,\pi})_{1 \leq m \leq M}$. To estimate the coefficients we shall draw upon (7.3). Let $(\alpha_{l,i}^{j,J,M})_{0 \leq l \leq d}$ denote the MC estimation of the family $(\alpha_{l,i}^{J,J})_{0 \leq l \leq d}$ at time t_i where j Picard iterations have been performed at time t_i and J at every time after t_i . We also set $\alpha_{l,i}^{0,J,M} = 0$ for every l and i . Given that we have already computed $(\alpha_{l,i}^{j-1,J,M})_{0 \leq l \leq d}$ the coefficients $(\alpha_{l,i}^{j,J,M})_{0 \leq l \leq d}$ are obtained as the solution to the empirical least squares problem

$$\min_{(\alpha_0, \dots, \alpha_d)} \frac{1}{M} \sum_{k=1}^M \left(\alpha_{0,i+1}^{J,J,M} p_{0,i+1}^m - \alpha_0 p_{0,i}^m + h_{i+1} f(i, \alpha_i^{j-1,J,M}) - \sum_{l=1}^d \alpha_l p_{l,i}^m \Delta W_{i+1}^{l,m} \right)^2 \tag{7.5}$$

where $p_{l,i}^m = p_{l,t_i}(\bar{X}_{t_i}^{\pi,m})$ for $l = 0, \dots, d$ and $\Delta W_{i+1}^{l,m}$ is the l -th component of the m -th copy of the Brownian motion. Having solved the least square problem above, it is natural to define

$$\hat{\mathbb{E}}_{t_i,x} \left[R_{i+1,n-1} \Phi(\bar{X}_{t_{i+1}}^\pi) \Delta W_{i+1}^l \right] := T^\nu \left[\alpha_{l,t_i}^{J,J,M} \cdot p_{l,t_i}(x) \right], \quad \forall l = 0, \dots, d$$

as the MC approximation of $\bar{\mathbb{E}}[\cdot]$, where $T^\nu[\cdot]$ is the truncation operator introduced in (6.4) of Section 6 (in fact the actual truncation operator used in [25] is a bit more elaborate). Denote by $\{\hat{R}_i\}_{0 \leq i \leq n}$ the family of empirical operators defined exactly similar to (4.7) with $\hat{\mathbb{E}}[\cdot]$ in place of $\bar{\mathbb{E}}[\cdot]$. The error estimate is of the form⁷

$$\max_{0 \leq i \leq n} \mathbb{E} \left[\left| R_{i,n-1} \Phi(\bar{X}_{t_i}^\pi) - \hat{R}_{i,n-1} \Phi(\bar{X}_{t_i}^\pi) \right|^2 \right] \leq C_1 h^{J-1} + \frac{C_2}{hM} + C_3 h, \tag{7.6}$$

where the constants C_1, C_2, C_3 depend on the up to fourth moments of the random variables $p_{l,i}(\bar{X}_{t_i}^\pi)$, $0 \leq l \leq d$, $1 \leq i \leq n$. The study of how one should optimally choose between n, M and the number of functions in the function basis to achieve a certain error accuracy, is developed in great detail in [25,26] which the interested reader should consult.

8. THE CUBATURE METHOD⁸

The cubature method was recently introduced in the context of BSDEs in [10] and it is based on the work of Lyons and Victoir [34]. The method’s basic idea is that the approximation of the solution of a BSDE can be viewed as a problem of weak approximation of the forward diffusion. More precisely, the backward component of a BSDE is an integral of an implicitly defined functional with respect to the Wiener measure. The cubature method of Lyons and Victoir produces an approximation of the Wiener measure (and implicitly of the law of the forward component of the BSDE) based on the classical iterated integrals expansion of a diffusion processes developed by Chen [8]. The method replaces the Wiener measure with the law of a Markov process which is directly computable or at least a lot easier to sample from. The roots of the Lyons-Victoir cubature method lie in theory of rough paths developed by Lyons and his collaborators (see, for example, [23,33,35]). One of the outcomes of the rough paths theory is the possibility of describing the evolution of a complex dynamical system by using its so-called “signature” – essentially an algebraic equivalent of the iterated integrals expansion

⁷The exact error estimate can be found in Theorem 3 of [25] and is further analyzed in [26].

⁸Throughout this section, assumptions (H1b) and (H2b) are in force.

introduced by Chen in the late fifties. The approximating process is chosen so that its signature coincides in expectation with the signature of the original process up to some order. The expected value of the signature of the new process will offer an accurate approximation of that of the original process and, implicitly, also its distribution will match closely that of the original process. Therefore, as the cubature method can be used to approximate the law of the forward component of the BSDE, it also leads to a corresponding approximation of the backward one.

In the following, we denote by $\mathcal{M} := \bigcup_{k \in \mathbb{N}} \{0, \dots, q\}^k$ the set of all multi indices with entries in the set $\{0, \dots, q\}$ and consider the norm

$$\|\alpha\| := |\alpha| + \text{number of } 0 \text{ entries in } \alpha,$$

where $|\alpha|$ is the length of α , e.g. if $\alpha = (1, 2, 0)$, then $|\alpha| = 3$ and $\|\alpha\| = 4$. We also fix m to be an arbitrary positive integer and $C_{\text{Lip}}([0, t]; \mathbb{R}^q)$ denotes the space of Lipschitz \mathbb{R}^q valued continuous functions.

Definition 8.1. We say that the positive weights $\lambda_1, \dots, \lambda_N$ and the paths of bounded variation $\omega_{b,1}, \dots, \omega_{b,N} \in C_{\text{Lip}}([0, b]; \mathbb{R}^q)$ define a cubature formula of degree m at time b , if and only if, for any multi index $\alpha = (\alpha_1, \dots, \alpha_k)$ with $\|\alpha\| \leq m$

$$\mathbb{E} \left[\int_{0 < t_1 < \dots < t_k < b} \circ dW_{t_1}^{\alpha_1} \dots \circ dW_{t_k}^{\alpha_k} \right] = \sum_{j=1}^N \lambda_j \int_{0 < t_1 < \dots < t_k < b} d\omega_{b,j}^{\alpha_1}(t_1) \dots d\omega_{b,j}^{\alpha_k}(t_k). \tag{8.1}$$

In (8.1) we use the convention that $W^0(t) = t$ and similarly for any $\omega \in C_{\text{Lip}}([0, b]; \mathbb{R}^q)$, $j = 1, \dots, N$, $\omega^0(t) = t$.

The identity (8.1) can be re-written as

$$\mathbb{E} \left[\int_{0 < t_1 < \dots < t_k < b} \circ dW_{t_1}^{\alpha_1} \dots \circ dW_{t_k}^{\alpha_k} \right] = \mathbb{E}_{\mathbb{Q}_b^m} \left[\int_{0 < t_1 < \dots < t_k < b} \circ dW_{t_1}^{\alpha_1} \dots \circ dW_{t_k}^{\alpha_k} \right], \tag{8.2}$$

where \mathbb{Q}_b^m is the cubature measure defined on $(C_0([0, b]; [0, b] \times \mathbb{R}^q), \mathcal{B}([0, b]) \otimes \mathcal{F})$

$$\mathbb{Q}_b^m = \sum_{j=1}^N \lambda_j \delta_{\omega_{b,j}} \tag{8.3}$$

and δ_ω denotes the Dirac measure at the point $\omega \in C([0, b])$. Specific examples of cubature methods for $m = 3, 5$ are described in [34]. Moreover, given a function $g : \mathbb{R}^q \rightarrow \mathbb{R}$, one has

$$\mathbb{E}_{\mathbb{Q}_b^m} \left[g(X_t^{0,x}) \right] = \sum_{j=1}^N \lambda_j \delta_{\omega_{b,j}} \left(g(X_b^{0,x}) \right),$$

where $\delta_{\omega_{b,j}} \left(g(X_b^{0,x}) \right)$ is the solution of the *ordinary differential equation* obtained by formally replacing the Brownian motion W with the path $\omega_{b,j}$ in Stratonovitch version of the forward part of the BSDE. More precisely, $\delta_{\omega_{b,j}} \left(g(X_b^{0,x}) \right)$ is the solution at time b of the equation

$$\begin{cases} dY_{t,x} &= \bar{V}_0(y_{t,x}) d\omega_{b,j}^0(t) + \sum_{i=1}^q V_i(y_{t,x}) d\omega_{b,j}^i(t) \\ y_{0,x} &= x \end{cases}$$

driven by $\omega_{b,j}$ where, again, we use the convention $\omega_{b,j}^0(t) = t$ and $\bar{V}_0 = V_0 - \frac{1}{2} \sum_{i=1}^q L^i V_i$ and L^j are the differential operators $L^j = \sum_{i=1}^q V_j^i \partial_{x_i}$ for $j = 1, \dots, q$. We also define L_0 to be the differential operator $L^0 = \sum_{i=1}^q \bar{V}_0^i \partial_{x_i}$. The following result is reported from [34]:

Proposition 8.2. *Let the path $\omega_{t,1}, \dots, \omega_{t,N} \in C_{bv}([0, t]; \mathbb{R}^q)$ and the positive weights $\lambda_1, \dots, \lambda_N$ define a cubature formula of order m at time t . Let $g \in C_b^{m+2}(\mathbb{R}^q; \mathbb{R})$. Then*

$$\sup_x \left| \mathbb{E} \left[g(X_t^{0,x}) \right] - \mathbb{E}_{\mathbb{Q}_t^m} \left[g(X_t^{0,x}) \right] \right| \leq C \sum_{j=m+1}^{m+2} t^{j/2} \sup_{\alpha \in \mathcal{M}, \|\alpha\|=j} \|L^\alpha g\|_\infty, \tag{8.4}$$

where C is a constant independent of t and g . In (8.4), the differential operator L^α corresponding to the multi-index $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_k) \in \mathcal{M}$ is defined to be $L^\alpha := L^{\alpha_1} L^{\alpha_2} \dots L^{\alpha_k}$.

The above result is justified by combining the properties of the cubature paths with the Stratonovich-Taylor expansion of $g(X_t^{0,x})$. In a similar manner, we can also obtain:

Corollary 8.3. *Within the setting of Proposition 8.2, we have for any $l = 1, \dots, q$*

$$\sup_x \left| \mathbb{E} \left[g(X_t^{0,x}) W_t^l \right] - \mathbb{E}_{\mathbb{Q}_t^m} \left[g(X_t^{0,x}) W_t^l \right] \right| \leq C \sum_{j=m}^{m+2} t^{(j+1)/2} \sup_{\alpha \in \mathcal{M}, \|\alpha\|=j} \|L^\alpha g\|_\infty \tag{8.5}$$

again with C independent of t .

Proposition 8.2 and Corollary 8.3 tell us that the cubature measure $\mathbb{E}_{\mathbb{Q}_{h_{i+1}}^m}[\cdot]$ is, in effect, an excellent candidate for the operator $\bar{\mathbb{E}}_{t_i,x}[\cdot]$. More precisely we define the family of operators $\{\bar{R}_i\}_i$ as follows:

$$\begin{aligned} \bar{R}_i g(x) &= \mathbb{E}_{\mathbb{Q}_{h_{i+1}}^m} [g(X_{t_{i+1}}^{t_i,x})] \\ &+ h_{i+1} f \left(t_i, x, \bar{R}_i g(x), \frac{1}{h_{i+1}} \mathbb{E}_{\mathbb{Q}_{h_{i+1}}^m} [g(X_{t_{i+1}}^{t_i,x}) (W_{t_{i+1}} - W_{t_i})] \right). \end{aligned} \tag{8.6}$$

To explain how one computes the family $\{\bar{R}_i\}_{0 \leq i \leq n-1}$ we first take a closer look at the operator $\mathbb{E}_{\mathbb{Q}_{h_{i+1}}^m}[\cdot]$, $i = 0, \dots, n-1$. From the definition of the cubature measure, it is straightforward that, given a function $g : \mathbb{R}^q \rightarrow \mathbb{R}$, $\mathbb{E}_{\mathbb{Q}_{h_{i+1}}^m} [g(X_{t_{i+1}}^{t_i,x})]$ is equal to a weighted average of solutions of ODEs. Indeed, according to (8.3)

$$\mathbb{E}_{\mathbb{Q}_{h_{i+1}}^m} [g(X_{t_{i+1}}^{t_i,x})] = \sum_{j=1}^N \lambda_j \delta_{\omega_{h_{i+1},j}} \left(g(X_{t_{i+1}}^{t_i,x}) \right)$$

where $\delta_{\omega_{h_{i+1},j}} \left(g(X_{t_{i+1}}^{t_i,x}) \right)$ is just a solution of an ODE from t_i to t_{i+1} formulated by substituting the path $\omega_{h_{i+1},j}$ in place of the Brownian motion in the forward part of the BSDE.

More precisely, let $\Xi_{t,x}(\omega)$, $\omega \in C_{0,bv}^0([0, t]; \mathbb{R}^q)$ denote the solution at time t of the ODE

$$\begin{cases} dY_{t,x} &= \bar{V}_0(y_{t,x}) d\omega^0(t) + \sum_{i=0}^q V_i(y_{t,x}) d\omega^i(t) \\ y_{0,x} &= x \end{cases}$$

driven by ω , where we understand that $\omega^0(t) = t$. Given a partition $0 = t_0 < t_1 < \dots < t_n = T$ of $[0, T]$ and a set of N paths and N weights that define a cubature formula of order m , we proceed in building a tree

with depth n . The root of the tree is the point $x = X_0$. We then define the singleton $\mathcal{N}_0 = \{x\}$. At depth k , we will have N^k points, the totality of which we denote by \mathcal{N}_k . We define the sets \mathcal{N}_k recursively as follows:

Given an integer $k = 1, \dots, n$ and a point $\chi \in \mathcal{N}_{k-1}$ we set

$$\begin{aligned} \mathcal{N}_k^\chi &:= \{y \in \mathbb{R}^q \mid y = \Xi_{h_k, \chi}(\omega_{h_k, j}) \text{ for some } j = 1, \dots, N\} \\ \mathcal{N}_k &:= \bigcup_{\chi \in \mathcal{N}_{k-1}} \mathcal{N}_k^\chi. \end{aligned}$$

The family $\{\bar{R}_i\}_{0 \leq i \leq n-1}$ is then evaluated on the tree $\bigcup_{k=0}^n \mathcal{N}_k$. For every $i = 0, \dots, n-1$, $\bar{R}_{i, n-1} \Phi$ is evaluated at all the different knots in the set \mathcal{N}_i .

Algorithm.

$i = n$ Compute the function Φ on all points of \mathcal{N}_n .

for $i = n - 1$ to $i = 0$ **do**

(recall the convention (4.5)) Compute the N^i values of the function $\bar{R}_{i, n-1} \Phi(\cdot)$ at all points of \mathcal{N}_i , as the unique fixed point of equation (see (8.6))

$$\begin{aligned} \bar{R}_{i, n-1} \Phi(\chi) &= \mathbb{E}_{\mathbb{Q}_{h_{i+1}}^m} [\bar{R}_{i+1, n-1} \Phi(X_{t_{i+1}}^{t_i, \chi})] \\ &\quad + h_{i+1} f \left(t_i, \chi, \bar{R}_{i, n-1} \Phi(\chi), \frac{1}{h_{i+1}} \mathbb{E}_{\mathbb{Q}_{h_{i+1}}^m} [\bar{R}_{i+1, n-1} \Phi(X_{t_{i+1}}^{t_i, \chi})(W_{t_{i+1}} - W_{t_i})] \right) \end{aligned}$$

where $\chi \in \mathcal{N}_i$ and

$$\mathbb{E}_{\mathbb{Q}_{h_{i+1}}^m} [\bar{R}_{i+1, n-1} \Phi(X_{t_{i+1}}^{t_i, \chi})] = \sum_{y \in \mathcal{N}_{i+1}^\chi} \lambda_y \bar{R}_{i+1, n-1} \Phi(y)$$

where λ_y denotes the weight that corresponds to the ODE that lead to the point y .

end for

Return $\bar{R}_{0, n-1} \Phi(x)$.

The global error for this algorithm is analysed in detail in [10]. To quantify it, one needs to work with (4.11) rather than (4.10). In fact the Trotter expansion that is used is of the form

$$\begin{aligned} |Y_0^{0,x} - \bar{R}_{0, n-1} \Phi(x)| &\leq \sum_{i=1}^{n-1} |\bar{R}_{0, i-1} Y_{t_i}^{0,x} - \bar{R}_{0, i} Y_{t_{i+1}}^{0,x}| \\ &\leq \sum_{i=1}^{n-1} |\bar{R}_{0, i-1} u(t_i, \bar{X}_{t_i}^\pi) - \bar{R}_{0, i} u(t_{i+1}, \bar{X}_{t_{i+1}}^\pi)|, \end{aligned} \tag{8.7}$$

where u solves (1.3).

The choice to work with an expansion as above is not just a matter of taste. Recall that from (4.14), we need to quantify the error $(\mathbb{E} - \mathbb{E}_{\mathbb{Q}_{h_{i+1}}^m}) [g(X_{t_{i+1}}) | X_{t_i} = x]$ for appropriate functions g . However, we see from (8.2) and (8.5) that these errors will essentially depend on the derivatives of the functions inside the expectations. With an expansion of the form (4.10) or (4.11) we would have to quantify the derivatives of the functions $R_{i, n-1} \Phi(x)$ or $\bar{R}_{i, n-1} \Phi(x)$, a difficult task. Working with (8.7), these derivatives are estimated by using standard PDE arguments through (1.3).

However, it is not always the case that a classical solution of the PDE (1.3) exists. Under (H1b), (H2b) and the additional assumption that $\Phi \in C_b^2(\mathbb{R}^d)$ this is true. For problems where Φ is merely Lipschitz continuous,

one can only argue the PDE (1.3) has solution only in the viscosity sense (see [41]). There is a detailed discussion in [10] on how one can circumvent this problem and produce satisfactory error estimates. The price that has to be paid for this reduction on the assumption on the smoothness of Φ is that one has to work with a rather exotic partition. We report the main theorem from [10]:

Theorem 8.4. *Let assumptions (H1b) and (H2b) hold true and $\Phi \in C_{\text{Lip}}(\mathbb{R}^d)$, the space of Lipschitz continuous functions. Let the family of operators $\{\bar{R}_i\}_{i=0}^n$ be constructed using a cubature formula of order m . Consider also a partition of $[0, T]$, $\pi = \{0 = t_0 < t_1 < \dots < t_n = T\}$ such that $h_i < h_n < 1$, $\forall i = 1, \dots, n-1$. Then, there exists a constant C independent of the partition, such that the global error satisfies*

$$|Y_0 - \bar{R}_{0,n-1}\Phi(x)| \leq \begin{cases} C \left(h_n^{1/2} + \sum_{i=1}^{n-1} h_i^2 h_n^{-3/2} \right) & m = 3 \\ C \left(h_n^{1/2} + \sum_{i=1}^{n-1} \left(h_i^2 h_n^{-1} + h_i^{(m+1)/2} h_n^{-m/2} \right) \right) & m \geq 5. \end{cases} \quad (8.8)$$

In particular, if one considers a partition of the form $t_i = T(1 - 1/n)i/n^\alpha$, $t_{n+1} = T$, i.e. a partition with $M = n^\alpha + 1$ points where $\alpha = 2 \mathbf{1}_{m=3} + \frac{3}{2} \mathbf{1}_{m=5}$, then one obtains the error estimate

$$|Y_0 - \bar{R}_{0,n-1}\Phi(x)| = O\left(1/M^{1/4}\right) \mathbf{1}_{m=3} + O\left(1/M^{1/3}\right) \mathbf{1}_{m=5}.$$

REFERENCES

- [1] V. Bally and G. Pagès, Error analysis of the quantization algorithm for obstacle problems. *Stochastic Processes their Appl.* **106** (2003) 1–40.
- [2] V. Bally and G. Pagès, A quantization algorithm for solving multi dimensional discrete-time optional stopping problems. *Bernoulli* **6** (2003) 1003–1049.
- [3] D. Becherer, Bounded solutions to backward SDE's with jumps for utility optimization and indifference pricing. *Ann. Appl. Prob.* **16** (2006) 2027–2054.
- [4] J.M. Bismut, *Théorie probabiliste du contrôle des diffusions*, Mem. Amer. Math. Soc. **176**. Providence, Rhode Island (1973).
- [5] B. Bouchard and N. Touzi, Discrete time approximation and Monte Carlo simulation for Backward Stochastic Differential Equations. *Stochastic Processes their Appl.* **111** (2004) 175–206.
- [6] B. Bouchard, I. Ekeland and N. Touzi, On the Malliavin approach to Monte Carlo methods of conditional expectations. *Financ. Stoch.* **8** (2004) 45–71.
- [7] P. Briand and Y. Hu, BSDE with quadratic growth and unbounded terminal value. *Probab. Theor. Relat. Fields* **136** (2006) 604–618.
- [8] K.-T. Chen, Integration of paths, geometric invariants and a generalized Baker-Hausdorff formula. *Ann. Math.* **65** (1957) 163–178.
- [9] P. Cheridito, M. Soner, N. Touzi and N. Victoir, Second-order backward stochastic differential equations and fully non linear parabolic pdes. *Commun. Pure Appl. Math.* **60** (2007) 1081–1110.
- [10] D. Crisan and K. Manolarakis, *Numerical solution for a BSDE using the Cubature method*. Preprint available at <http://www2.imperial.ac.uk/~dcrisan/> (2007).
- [11] D. Crisan, K. Manolarakis and N. Touzi, On the Monte Carlo simulation of BSDEs: An improvement on the Malliavin weights. *Stochastic Processes their Appl.* **120** (2010) 1133–1158.
- [12] J. Cvitanic and I. Karatzas, Hedging contingent claims with constrained portfolios. *Ann. Appl. Prob.* **3** (1993) 652–681.
- [13] D. Duffy and L. Epstein, Asset pricing with stochastic differential utility. *Rev. Financ. Stud.* **5** (1992) 411–436.
- [14] D. Duffy and L. Epstein, Stochastic differential utility. *Econometrica* **60** (1992) 353–394.
- [15] N. El Karoui and S.J. Huang, A general result of existence and uniqueness of backward stochastic differential equations, in *Backward Stochastic Differential Equations*, N. El Karoui and L. Mazliak Eds., Longman (1996).
- [16] N. El Karoui and M. Quenez, Dynamic programming and pricing of contingent claims in incomplete markets. *SIAM J. Contr. Opt.* **33** (1995) 29–66.
- [17] N. El Karoui and M. Quenez, Non linear pricing theory and Backward Stochastic Differential Equations, in *Financial Mathematics* **1656**, Springer (1995) 191–246.
- [18] N. El Karoui, C. Kapoudjan, E. Pardoux, S. Peng and M.C. Quenez, Reflected solutions of backward SDEs and related obstacle problems. *Annals Probab.* **25** (1997) 702–737.
- [19] N. El Karoui, E. Pardoux and M. Quenez, Reflected backward SDEs and American Options, in *Numerical Methods in Finance*, Chris Rogers and Denis Talay Eds., Cambridge University Press, Cambridge (1997).
- [20] N. El Karoui, S. Peng and M. Quenez, Backward Stochastic Differential Equations in finance. *Mathematical Finance* **7** (1997) 1–71.

- [21] R. Feynman, Space-time approach to non-relativistic quantum mechanics. *Rev. Mod. Phys.* **20** (1948) 367–387.
- [22] H. Föllmer and A. Schied, Convex measures of risk and trading constraints. *Financ. Stoch.* **6** (2002) 429–447.
- [23] P. Friz and N. Victoir, *Multidimensional Stochastic Processes as Rough Paths: Theory and applications*. Cambridge studies in advanced mathematics, Cambridge University Press, Cambridge (2010).
- [24] E. Gobet and C. Labart, Error expansion for the discretization of Backward Stochastic Differential Equations. *Stochastic Processes their Appl.* **117** (2007) 803–829.
- [25] E. Gobet, J.P. Lemor and X. Warin, A regression based Monte Carlo method to solve Backward Stochastic Differential Equations. *Ann. Appl. Prob.* **15** (2005) 2172–2202.
- [26] E. Gobet, J.P. Lemor and X. Warin, Rate of convergence of an empirical regression method for solving generalized backward stochastic differential equations. *Bernoulli* **12** (2006) 889–916.
- [27] E. Jouini and H. Kallal, Arbitrage in securities markets with short sales constraints. *Mathematical Finance* **5** (1995) 178–197.
- [28] M. Kac, On distributions of certain Wiener functionals. *Trans. Amer. Math. Soc.* **65** (1949) 1–13.
- [29] I. Karatzas and S. Schreve, *Brownian Motion and Stochastic Calculus*. Springer Verlag, New York (1991).
- [30] M. Kobylanski, Backward Stochastic Differential Equations and Partial Differential Equations. *Ann. Appl. Prob.* **28** (2000) 558–602.
- [31] J.-P. Lepeltier and J. San Martin, Backward Stochastic Differential Equations with continuous coefficients. *Stat. Probab. Lett.* **32** (1997) 425–430.
- [32] F. Longstaff and E.S. Schwartz, Valuing American options by simulation: a simple least squares approach. *Rev. Financ. Stud.* **14** (2001) 113–147.
- [33] T. Lyons and Z. Qian, *System Control and Rough Paths*. Oxford Science publication, Oxford University Press, Oxford (2002).
- [34] T. Lyons and N. Victoir, Cubature on Wiener space. *Proc. Royal Soc. London* **468** (2004) 169–198.
- [35] T. Lyons, M. Caruana and T. Levy, *Differential Equations Driven by Rough Paths, Lecture Notes in Mathematics* **1908**. Springer (2004).
- [36] J. Ma and J. Zhang, Representation theorems for Backward Stochastic Differential Equations. *Ann. Appl. Prob.* **12** (2002) 1390–1418.
- [37] J. Ma and J. Zhang, Representation and regularities for solutions to BSDEs with reflections. *Stochastic Processes their Appl.* **115** (2005) 539–569.
- [38] J. Ma, P. Protter and J. Yong, Solving Forward-Backward SDEs explicitly – A four step scheme. *Probab. Theor. Relat. Fields* **122** (1994) 163–190.
- [39] D. Nualart, *The Malliavin calculus and related topics*. Springer-Verlag (1996).
- [40] E. Pardoux and S. Peng, Adapted solution to Backward Stochastic Differential Equations. *Syst. Contr. Lett.* **14** (1990) 55–61.
- [41] E. Pardoux and S. Peng, Backward Stochastic Differential Equations and quasi linear parabolic partial differential equations, in *Lecture Notes in Control and Information Sciences* **176**, Springer, Berlin/Heidelberg (1992) 200–217.
- [42] E. Pardoux and S. Tang, Forward-backward stochastic differential equations and quasilinear parabolic PDEs. *Probab. Theor. Relat. Fields* **114** (1999) 123–150.
- [43] S. Peng, Backward SDEs and related g-expectations, in *Pitman Research Notes in Mathematics Series* **364**, Longman, Harlow (1997) 141–159.
- [44] S. Peng, *Non linear expectations non linear evaluations and risk measures* **1856**. Springer-Verlag (2004).
- [45] S. Peng, *Modelling derivatives pricing mechanisms with their generating functions*. Preprint, [arxiv:math/0605599v1](https://arxiv.org/abs/math/0605599v1) (2006).
- [46] E. Rosazza Giannin, Risk measures via g expectations. *Insur. Math. Econ.* **39** (2006) 19–34.
- [47] S. Tang and X. Li, Necessary conditions for optimal control of stochastic systems with random jumps. *SIAM J. Contr. Opt.* **32** (1994) 1447–1475.
- [48] J. Zhang, *Some fine properties of backward stochastic differential equations*. Ph.D. Thesis, Purdue University, USA (2001).
- [49] J. Zhang, A numerical scheme for BSDEs. *Ann. Appl. Prob.* **14** (2004) 459–488.