ARTICLE IN PRESS

Available online at www.sciencedirect.com
ScienceDirect
Stochastic Processes and their Applications ▮ (▮▮▮▮) ▮▮▮–▮▮▮
stochastic processes and their applications
www.elsevier.com/locate/spa

On the Monte Carlo simulation of BSDEs: An improvement on the Malliavin weights

D. Crisan[a,*], K. Manolarakis[a], N. Touzi[b]

[a] *Department of Mathematics, Imperial College London, 180 Queen's Gate, London, SW7 2AZ, UK*
[b] *CMAP, Ecole Polytechnique Paris, France*

Received 15 April 2008; received in revised form 11 February 2010; accepted 5 March 2010

Abstract

We propose a generic framework for the analysis of Monte Carlo simulation schemes of backward SDEs. The general results are used to re-visit the convergence of the algorithm suggested by Bouchard and Touzi (2004) [6]. By keeping the higher order terms in the expansion of the Skorohod integrals resulting from the Malliavin integration by parts in [6], we introduce a variant of the latter algorithm which allows for a significant reduction of the numerical complexity. We prove the convergence of this improved Malliavin-based algorithm, and derive a bound on the induced error. In particular, we show that the price to pay for our simplification is to use a more accurate localizing function.

MSC: 65C30; 65C05; 60H07; 62G08

Keywords: BSDEs; Weak approximations; Monte Carlo methods; Malliavin calculus

1. Introduction

Let $(\Omega, \mathcal{F}, P)$ be a probability space on which we have defined the decoupled forward–backward stochastic differential equation (BSDE henceforth)

$$X_t = X_0 + \int_0^t \sigma^0(X_s)\mathrm{d}s + \int_0^t \sigma(X_s)\mathrm{d}W_s \quad t \in [0, 1] \tag{1.1}$$

* Corresponding author.
E-mail addresses: d.crisan@imperial.ac.uk (D. Crisan), nizar.touzi@polytechnique.edu (N. Touzi).

0304-4149/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.spa.2010.03.015

$$Y_t = \Phi(X_1) + \int_t^1 f(s, X_s, Y_s, Z_s)\mathrm{d}s - \int_t^1 Z_s \mathrm{d}W_s \quad t \in [0, 1], \tag{1.2}$$

where $W = \{W_t, t \geq 0\}$ is a d-dimensional Brownian motion. The coefficients of the forward equation (1.1) are the functions $\sigma^0 : \mathbb{R}^d \longrightarrow \mathbb{R}^d$, $\sigma : \mathbb{R}^d \longrightarrow \mathbb{R}^{d\times d}$ which satisfy the classical assumptions to ensure existence and uniqueness of its solution and convergence of the corresponding Euler discrete-time approximation. The backward equation (1.2) is defined by the generator $f : [0, 1]\times\mathbb{R}^d\times\mathbb{R}\times\mathbb{R}^d \longrightarrow \mathbb{R}$ which is also assumed to satisfy the classical condition for existence and uniqueness of its solution.

Discrete-time schemes based on the approximation of the Brownian motion have been analyzed by Chevance [9], Coquet, Macquevicius and Mémin [10], Briand, Delyon and Mémin [8], Antonelli and Kohatsu-Higa [1], Ma, Protter, San Martin and Torres [20], Bally and Pagès [3]. In contrast with the above literature, we concentrate on approximations of solutions of backward stochastic differential equations based on the Monte Carlo simulation of the underlying Brownian motion, thus continuing the studies of Bouchard and Touzi [6], Zhang [23], Gobet, Lemor and Warin [16] and Gobet and Labart [14]. See [5] for the case of jump–diffusions and [2] for a related Malliavin Calculus approach for pricing and hedging American options.

As a first contribution, we propose a generic framework for the analysis of such approximations. By isolating the one-step approximation operator of the algorithm, we provide a transparent set of sufficient conditions for the convergence of the approximation. The main tool is an expansion of the error as a Trotter product. This generic framework applies to various numerical methods. For instance, the above methodology is applied by Crisan and Manolarakis [11] to prove convergence of a probabilistic numerical algorithm where the regressions are estimated by using the cubature method of Lyons and Victoir [19] to approximate the distribution of (X, W).

This methodology is applied in this paper to the Malliavin integration-by-parts-based algorithm introduced in [6]. The algorithm exploits the representation of the regression function $\psi(x) := \mathbb{E}[Y|X = x] = \mathbb{E}[Y\varepsilon_{\{x\}}(X)]/\mathbb{E}[\varepsilon_{\{x\}}(X)]$, where $\varepsilon_{\{x\}}$ denotes the Dirac mass at x, as

$$\psi(x) = \frac{\mathbb{E}\left[Y\mathbf{1}_{\mathbb{R}^d_+}(X - x)S^h\left(\phi(X - x)\right)\right]}{\mathbb{E}\left[\mathbf{1}_{\mathbb{R}^d_+}(X - x)S^h\left(\phi(X - x)\right)\right]}$$

for *smooth*, in the Malliavin sense, random variables X and Y valued respectively in $\mathbb{R}^d$ and $\mathbb{R}^k$, for some integers $d, k \geq 1$. Here S^h is an iterated *Skorohod integral* and ϕ is a certain smooth function with $\phi(0) = 1$ called *localizing function*.

The Malliavin integration-by-parts algorithm of [6] turns out to have high numerical complexity. As we shall see in Section 3 the involved weights are actually iterated Skorohod integrals (the length being equal to the dimension of the space where the forward diffusion lives in), the definition of which involves the derivatives of the functions $\sigma^0(x), \sigma(x)$. Hence, on the one hand the computation of these weights will involve an exponentially increasing number of terms with respect to the dimension of the problem. On the other hand, when the functions $\sigma^0(x), \sigma(x)$ are not known explicitly (e.g. when they are computed via a calibration procedure), numerical methods need to be employed for the computations of the involved derivatives. Our second goal is then to show that one can avoid these complexities by using alternative weights. These new weights are easier to handle algebraically, involve significantly less number of terms and yet preserve the asymptotic properties of the original algorithm (see Theorem 4.5 and the last

Please cite this article in press as: D. Crisan, et al., On the Monte Carlo simulation of BSDEs: An improvement on the Malliavin weights, Stochastic Processes and their Applications (2010), doi:10.1016/j.spa.2010.03.015

section on numerical experiments). For example, for $d = 5$, our simplified algorithm requires the calculation of 32 terms instead of 1024, hence we reduce the effort by more than 96%.

The proposed variation of the Malliavin integration-by-parts algorithm eliminates lower order terms contained in the Malliavin weights by a judicious expansion. This leads to a drastic reduction of the number of terms required in the case of nonconstant coefficients σ^0 and σ, for every weight. Moreover, these lower order terms that get ignored, are exactly the ones that involve the derivatives of the coefficient functions. Hence we manage to avoid the numerical complexities that may arise from the computation of these derivatives. The reduction in the numerical complexity requires a change of scale in the localizing function $\phi^h(x) := \phi(xh^{-\alpha})$ of [4,6], where h is the time step, in order to keep an equivalent rate of convergence. While $\alpha = 1/2$ in [6] is the optimal scale, we find in the present setting that α must be chosen strictly larger than $1/2$.

We discuss in the following the effect of the change of scale in the localizing function. In the original algorithm the error was of order $|\pi|^{-1-d/4p}N^{-1/2p} + |\pi|^{\frac{1}{2}}$ for any $p > 1$. This implies that one requires the number of paths N to be of order $\frac{1}{|\pi|^{p+\frac{d}{2p}}}$ to match the discretization error. In the new version, the error is of order (see Theorem 4.5)

$$|\pi|^{-1-\alpha d/2p}N^{-1/2p} + |\pi|^{\frac{1}{p}\left(\frac{1}{2}+\alpha\right)-1} + |\pi|^{\frac{1}{2}}.$$

Hence, for $\alpha > 1/2$ the algorithm converges, whereas for $\alpha = (3p-1)/2$ and $N = \frac{1}{|\pi|^{p+\frac{d}{2p}+\frac{3d(p-1)}{4p}}}$ one achieves the same error. However, the numerical results in the last section show identical convergence rates for both algorithms where the same number of paths are used.

The paper is organised as follows. After collecting some preliminaries in Section 2, we describe our one-step operator approach for the discretization of BSDEs in Section 2.1. The original Malliavin integration-by-parts algorithm [6] is reviewed in Section 3. Section 4 introduces the modified algorithm which includes truncated weights and provides the corresponding asymptotic results for the error analysis. In Section 5 we present the numerical results. In particular, we give details of the implementation of the algorithm and compare the method with the original method of [6]. The comparison is done on a nontrivial two dimensional example and it shows that no precision is lost by the transition to the new weights.

2. Preliminaries

Let $(\Omega, \mathcal{F}, \mathbb{P}, \{\mathcal{F}_t\})$ be a complete filtered probability space on which we have defined a d-dimensional Brownian motion $\{W_t, 0 \le t \le 1\}$. In the following we will assume that the filtration $\{\mathcal{F}_t\}$ is the augmented Brownian filtration and that W and all the other processes appearing below are defined on the finite time interval $[0, 1]$. Let $\mathbb{M}^d$ be the space of $d \times d$ matrices equipped with the norm $|A| := \left(\sum_{i,j=1}^d |a_{ij}|^2\right)^{1/2}$ for $A = (a_{ij}) \in \mathbb{M}^d$ and let A^* denote the transpose of $A \in \mathbb{M}^d$. In what follows, all Euclidean spaces are equipped with the usual Euclidean norm. The inner product of two elements in $\mathbb{R}^d$ is denoted by $x \cdot y$.

We will be using the following assumptions on the coefficients of system ((1.1)–(1.2)):

(H1a) The functions $\sigma^0 : \mathbb{R}^d \to \mathbb{R}^d, \sigma : \mathbb{R}^d \to \mathbb{M}^d$ are Lipschitz continuous

$$|\sigma^0(x) - \sigma^0(y)| + |\sigma(x) - \sigma(y)| \le K|x - y|,$$

and σ is uniformly positive definite.

Please cite this article in press as: D. Crisan, et al., On the Monte Carlo simulation of BSDEs: An improvement on the Malliavin weights, Stochastic Processes and their Applications (2010), doi:10.1016/j.spa.2010.03.015

(H1b) The functions $\sigma^0, \sigma, \sigma^{-1} \in C_b^\infty$, the space of infinite differentiable functions with bounded derivatives of all orders.

(H2) The driver of the BSDE $f : [0,1] \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}$ and its terminal condition $\Phi : \mathbb{R}^d \longrightarrow \mathbb{R}$ are Lipschitz in the spatial variables and 1/2-Hölder continuous with respect to t.

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

A solution to the backward equation (1.2) is an adapted pair of processes (Y, Z) taking values in $\mathbb{R} \times \mathbb{R}^d$ with

$$\mathbb{E}\left[\sup_{0\leq t\leq 1} |Y_t|^2 + \int_0^1 |Z_s|^2 \mathrm{d}s\right] < \infty.$$

Following for example, Theorem 2.1 of [13], such a solution exists when conditions (H1a) and (H2) are satisfied. In what follows we will be primarily working with these two hypotheses. The enhanced condition (H1b) will be required to justify the additional smoothness, in the Malliavin sense, of the Euler approximation of the forward component (1.1) below.

2.1. Discretizing the BSDE

Let $\pi = \{0 = t_0 < t_1 < \cdots < t_n = 1\}$ be a partition of the time interval along which we discretize the system ((1.1)–(1.2)). Denote by

$$\Delta_i := t_i - t_{i-1}, \qquad \Delta W_i := W_{t_i} - W_{t_{i-1}}, \quad \text{and} \quad |\pi| := \max_{1\leq i\leq n} \Delta_i$$

the time step, the increment of the Brownian motion, and the partition mesh, respectively. We also write $\Delta W_{i+1}^l, l = 1, \ldots, d$ for the entries of the random vector ΔW_{i+1} for any $i = 0, \ldots, n-1$. Let X^π be the Euler approximation of X, the forward component (1.1). That is, X^π is defined recursively as:

$$X_0^\pi = X_0 \quad \text{and} \quad X_{t_i}^\pi := X_{t_{i-1}}^\pi + \sigma^0\left(X_{t_{i-1}}^\pi\right)\Delta_i + \sigma\left(X_{t_{i-1}}^\pi\right)\Delta W_i, \quad 0 \leq i \leq n-1.$$

Under conditions (H1a) and (H2), the following standard estimate holds true:

$$\max_{0\leq i\leq n-1} \mathbb{E}\left[|X_{t_i} - X_{t_i}^\pi|^2\right] \leq C(1 + |X_0|^2)|\pi|, \tag{2.1}$$

for some constant C, see e.g. [17], Theorem 10.2.2. In order to define a discrete-approximation of the process Y, we introduce a family of operators $\{R_i, i = 0, \ldots, n-1\}$ which we subsequently call the regression operators. In the following we will use the notation $\mathbb{E}_{t_i,x}[\cdot]$ for $\mathbb{E}[\cdot|X_{t_i}^\pi = x]$ and $C_{\mathrm{Lip}}(\mathbb{R}^d)$ to denote the set of Lipschitz continuous functions $g : \mathbb{R}^d \to \mathbb{R}$. For $0 \leq i < n$ and $g \in C_{\mathrm{Lip}}(\mathbb{R}^d)$, $R_i g$ is defined implicitly by:

$$R_i g(x) := \mathbb{E}_{t_i,x}\left[g\left(X_{t_{i+1}}^\pi\right)\right] + \Delta_{i+1} f\left(t_i, x, R_i g(x), (\Delta_{i+1})^{-1} \times \mathbb{E}_{t_i,x}\left[g\left(X_{t_{i+1}}^\pi\right)\Delta W_{i+1}\right]\right). \tag{2.2}$$

Please cite this article in press as: D. Crisan, et al., On the Monte Carlo simulation of BSDEs: An improvement on the Malliavin weights, Stochastic Processes and their Applications (2010), doi:10.1016/j.spa.2010.03.015

Moreover, let $\{R_{i,j}, 0 \le i \le j < n\}$ be the composition operators

$$R_{i,j} := \begin{cases} R_i \dots R_j & \text{for } 0 \le i \le j < n \\ \mathbf{1} & \text{for } i > j, \end{cases} \tag{2.3}$$

where $\mathbf{1}$ denotes the identity operator. Let us observe that $R_i g$ can be defined in the same manner for a larger class of functions, for example for every function $g : \mathbb{R}^d \to \mathbb{R}$ with polynomial growth. However, we are only concerned with the definition of the operators R_i on the set $C_{\mathrm{Lip}}(\mathbb{R}^d)$. Indeed, by Proposition 2 of [16] or the proof of Theorem 6.1 of [23] for any $g \in C_{\mathrm{Lip}}(\mathbb{R}^d)$ we have that

$$\max_{0 \le i \le n-1} \sup_{\substack{x,y \in \mathbb{R}^d \\ x \ne y}} \frac{\left| R_{i,n-1} g(x) - R_{i,n-1} g(y) \right|}{|x - y|} < \infty, \tag{2.4}$$

i.e., the composition of the family of operators $\{R_i, i = 0, \dots, n-1\}$ applied to a Lipschitz function, produces a sequence of uniformly (in i) Lipschitz functions. Moreover, for any $p > 1$ and for sufficiently small $|\pi|$ (such that $|\pi| < 1/K$), the following Lipschitz-type property of the operators R_i is crucially used in [6], although not outlined in a clear statement:

$$|R_i g_1 - R_i g_2|(x) \le \frac{1 + C\Delta_{i+1}}{1 - K\Delta_{i+1}} \|g_1 - g_2\|_{\mathbb{L}^p(\mathbb{P}^x_{t_i, t_{i+1}})}, \tag{2.5}$$

where C is a constant depending on d, p, and K, the Lipschitz constant of f, and

$$\|f\|_{\mathbb{L}^p(\mathbb{P}^x_{t_i,t_j})} := \left(\mathbb{E}[|f(X^\pi_{t_j})|^p | X^\pi_{t_i} = x] \right)^{1/p}.$$

For completeness, we report in Appendix the proof of (2.5). By direct iteration of (2.5), we see that

$$\left| R_{0,i} g - R_{0,i} h \right|(x) \le \prod_{j=1}^{i+1} \frac{1 + C\Delta_j}{1 - K\Delta_j} \|g - h\|_{\mathbb{L}^p(\mathbb{P}^x_{t_0,t_{i+1}})}, \quad \text{for all } 0 \le i < n. \tag{2.6}$$

Let (Y^π, Z^π) be the pair of processes defined recursively as: $(Y^\pi_{t_n}, Z^\pi_{t_n}) = (\Phi(X^\pi_{t_n}), 0)$ and for $0 \le i < n$ we have

$$Y^\pi_{t_i} := R_{i,n-1} \Phi\left(X^\pi_{t_i}\right), \qquad Z^\pi_{t_i} = (\Delta_{i+1})^{-1} \mathbb{E}_{t_i, X^\pi_{t_i}}\left[Y^\pi_{t_{i+1}} \Delta W_{i+1} \right], \quad 0 \le i < n.$$

Then (Y^π, Z^π) are a discrete-time approximation of the solution (Y, Z) of the BSDE. The following estimate of the corresponding approximation error may be found in [6] or, in a slightly different formulation, in [23] for the path dependent case:

$$\limsup_{|\pi| \to 0} \frac{1}{|\pi|} \mathbb{E}\left[\sup_{0 \le i \le n} |Y_{t_i} - Y^\pi_{t_i}|^2 + \sum_{0 \le i < n} |Z_{t_i} - Z^\pi_{t_i}|^2 \Delta_i \right] < \infty.$$

2.2. Approximating the regression operator

The conditional expectations involved in (2.2) are not computable in most cases. Hence, the next step is to replace the expectation operator $\mathbb{E}_{t_i,x}[\cdot]$ by a simulation-based approximating

Please cite this article in press as: D. Crisan, et al., On the Monte Carlo simulation of BSDEs: An improvement on the Malliavin weights, Stochastic Processes and their Applications (2010), doi:10.1016/j.spa.2010.03.015

operator $\hat{\mathbb{E}}_{t_i,x}[\cdot]$ that is explicitly computable. The simulation could rely on a Monte Carlo method as is the case in [6,15], or on an evaluation on a tree as in [11]. Given such an operator $\hat{\mathbb{E}}_{t_i,x}[\cdot]$ we define the corresponding family of operators $\{\hat{R}_i, i = 0, \ldots, n-1\}$,

$$\hat{R}_i g(x) = \hat{\mathbb{E}}_{t_i,x}\left[g(X^\pi_{t_{i+1}})\right] + \Delta_{i+1} f\left(t_i, x, \hat{R}_i g(x), \Delta_{i+1}^{-1}\hat{\mathbb{E}}_{t_i,x}\left[g(X^\pi_{t_{i+1}})\Delta W_{i+1}\right]\right), \quad (2.7)$$

and we also define the family of iterations of these operators $\hat{R}_{i,j}$, $i, j = 0, \ldots, n$ exactly as in (2.3). Next let $\hat{Y}^\pi$ be the process defined recursively as: $Y^\pi_{t_n} = \Phi(X^\pi_{t_n})$ and

$$\hat{Y}^\pi_{t_i} = \hat{R}_{i,n-1}\Phi\left(X^\pi_{t_i}\right), \quad \text{for } 0 \le i < n.$$

It is shown in [6], that the global error between the processes Y^π, $\hat{Y}^\pi$ is controlled by the sum of the local regression errors. Observe that, from the Lipschitz property of f:

$$\left|R_i g(x) - \hat{R}_i g(x)\right| \le \frac{K}{1 - K\Delta_{i+1}} \sum_{l=0}^{d} \left|\left(\mathbb{E}_{t_i,x} - \hat{\mathbb{E}}_{t_i,x}\right)\left[g\left(X^\pi_{t_{i+1}}\right)\Delta W^l_{i+1}\right]\right|, \quad (2.8)$$

using the convention that $\Delta W^0_{i+1} = 1$. We then fix the time to t_0 and analyze the error $|Y^\pi_{t_0} - \hat{Y}^\pi_{t_0}|$ by means of a Trotter product expansion. Namely we add and subtract n terms of the form $R_{0,i}\hat{R}_{i+1,n-1}\Phi(X^\pi_{t_0})$ each of which being a combination of $n-i$ empirical backward projections and i exact ones:

$$\begin{aligned}|Y^\pi_{t_0} - \hat{Y}^\pi_{t_0}| &\le \sum_{i=0}^{n-1}\left|R_{0,i-1}\hat{R}_{i,n-1}\Phi(X^\pi_{t_0}) - R_{0,i}\hat{R}_{i+1,n-1}\Phi(X^\pi_{t_0})\right| \\ &\le \sum_{i=0}^{n-1}\prod_{j=0}^{i-1}\frac{1 + C\Delta_{j+1}}{1 - K\Delta_{j+1}}\left\|(R_i - \hat{R}_i)\hat{R}_{i+1,n-1}\Phi(X^\pi_{t_i})\right\|_p,\end{aligned}$$

by (2.6). Using (2.8), this provides:

$$|Y^\pi_{t_0} - \hat{Y}^\pi_{t_0}| \le A(\pi)\frac{K(d+1)}{1 - K|\pi|}\max_{0\le l\le d}\left\|\left(\mathbb{E}_{t_i,X^\pi_{t_i}} - \hat{\mathbb{E}}_{t_i,X^\pi_{t_i}}\right)\left[\hat{R}_{i,n-1}\Phi(X^\pi_{t_{i+1}})\Delta W^l_{i+1}\right]\right\|_p,$$

where

$$A(\pi) = \sum_{i=0}^{n-1}\prod_{j=0}^{i-1}\frac{1 + C\Delta_{j+1}}{1 - K\Delta_{j+1}} \le |\pi|^{-1}e^{C+K}.$$

The above arguments are of course valid for any t_i, $i = 0, \ldots, n-1$. Hence we have:

Proposition 2.1. *Under the assumptions* (H1a) *and* (H2) *for every* $p > 1$ *and* $i < n$*:*

$$|Y^\pi_{t_i} - \hat{Y}^\pi_{t_i}| \le C|\pi|^{-1}\max_{\substack{0\le l\le d\\ 0\le i\le n-1}}\left\|\left(\mathbb{E}_{t_i,X^\pi_{t_i}} - \hat{\mathbb{E}}_{t_i,X^\pi_{t_i}}\right)\left[\hat{R}_{i,n-1}\Phi\left(X^\pi_{t_{i+1}}\right)\Delta W^l_{i+1}\right]\right\|_p,$$

where π *is a partition of* $[0,1]$ *with mesh size* $|\pi| < \frac{1}{K}$ *and* C *is a constant independent of the choice of partition.*

Please cite this article in press as: D. Crisan, et al., On the Monte Carlo simulation of BSDEs: An improvement on the Malliavin weights, Stochastic Processes and their Applications (2010), doi:10.1016/j.spa.2010.03.015

3. Simulating the BSDE by the Malliavin calculus approach

The purpose of this section is to review the Malliavin calculus-based simulation method suggested in [6] so as to introduce the necessary notation and some results which will be needed in the subsequent section.

Throughout this section the stronger condition (H1b) is assumed to hold. Also, the partition is assumed to be equidistant so that $\Delta_i = |\pi|$ for all $1 \le i \le n$. The definition of the simulation-based approximation $\hat{\mathbb{E}}$ of $\mathbb{E}$ requires the following notation:

We denote by $\mathcal{J}_k$ the subset of $\mathbb{N}^k$ with elements (multi-indices) $I = (i_1, \dots, i_k)$ that satisfy $1 \le i_1 < \cdots < i_k \le d$, and we set $\mathcal{J}_0 := \emptyset$. Given two elements $I \in \mathcal{J}_k$, $J \in \mathcal{J}_q$ we define their concatenation $I * J := (r_1, \dots, r_l)$ with $k \vee q \le l \le d \wedge (k+q)$, $r_i \in I \cup J$ for every $i \le l$, and $1 \le r_1 < \cdots < r_l \le d$. For $J \in \mathcal{J}_k$, $k \le d$ we write J^c for its complementary set, i.e. the unique disjoint set from J with $J * J^c = \{1, 2, \dots, d\}$. Finally, given a multi-index I with length k we shall write $-I$ for the multi index of length $k-1$ obtained from I by deleting the left most entry.

Given a matrix valued process h with columns denoted by h^i, a random variable F, and a multi-index $I = (i_1, \dots, i_k)$, we denote by

$$S_i^h[F] := \int_0^\infty F(h_t^i)^* \mathrm{d}W_t, \qquad S_I^h[F] := S_{i_1}^h \circ \cdots \circ S_{i_k}^h[F],$$

where the integrals are understood in the Skorohod sense. We extend the definition to $I = \emptyset$ by setting $S_\emptyset^h[F] = F$. Let ϕ be a bounded and continuous real valued function with $\phi(0) = 1$. We denote by $\partial_i \phi$ the partial derivative with respect to x_i and $\partial_I \phi := \partial_{i_k} \cdots \partial_{i_1} \phi$ when $I = (i_1, \dots, i_k)$. We will say that ϕ is a smooth localizing function if

$$\partial_I \phi(x) \in C_b^0(\mathbb{R}^d), \quad I \in \mathcal{J}_k,\ k \le d, \quad \text{and} \quad \sum_{I \in \cup_{k=1}^d \mathcal{J}_k} \int_{\mathbb{R}^d} |x|^m \partial_I \phi(x)^2 \mathrm{d}x < \infty,$$

where $\partial_\emptyset \phi := \phi$. By $\mathcal{L}_0$, we denote the collection of all smooth localizing functions. The latter integrability condition was needed in [6] for technical reasons.

Under condition (H1b), X^π is infinitely many times differentiable in the Malliavin sense, and the derivative is given by

$$D_t X_{t_i}^\pi = \sigma(X_{t_{i-1}}^\pi) 1_{(t_{i-1}, t_i]}(t) + \left(I_d + \sum_{j=0}^d \nabla \sigma^j(X_{t_{i-1}}^\pi) \Delta W_i^j \right) D_t X_{t_{i-1}}^\pi.$$

Then, it follows that the matrix-valued processes $\{h_i, i = 0, \dots, n-1\}$ defined as

$$h_i(t) := \frac{1}{|\pi|} \sigma^{-1}(X_{t_{i-1}}^\pi) 1_{[t_{i-1}, t_i)}(t) - \frac{1}{|\pi|} \sigma^{-1}(X_{t_i}^\pi) \times \left(I_d + \sum_{j=0}^d \nabla \sigma^j(X_{t_i}^\pi) \Delta W_{i+1}^j \right) 1_{[t_i, t_{i+1})}(t), \quad i \le n,$$

satisfy the identities

$$\int_0^1 D_t X_{t_i}^\pi h_i(t) \mathrm{d}t = I_d \quad \text{and} \quad \int_0^1 D_t X_{t_{i+1}}^\pi h_i(t) \mathrm{d}t = 0. \tag{3.1}$$

Please cite this article in press as: D. Crisan, et al., On the Monte Carlo simulation of BSDEs: An improvement on the Malliavin weights, Stochastic Processes and their Applications (2010), doi:10.1016/j.spa.2010.03.015

For $I \in \cup_{k=0}^{d} \mathcal{J}_k$, $\phi \in \mathcal{L}_0$, and $l \leq d$, the iterated Skorohod integrals

$$S_I^{h_i}\left[\Delta W_{i+1}^l \phi(X_{t_i}^\pi - x)\right] = \sum_{J \subseteq I} (-1)^{|J|} \partial_J \phi(X_{t_i}^\pi - x) S_{J^c}^{h_i}[\Delta W_{i+1}^l], \tag{3.2}$$

are well defined, and belong to $\mathbb{D}^{1,2}$, the space of *Malliavin differentiable random variables*. The latter equality follows from the first part of (3.1).

The following representation is reported from [4,6].

Theorem 3.1. *For $0 \leq l \leq d$ and $\rho : \mathbb{R}^d \to \mathbb{R}$ with $F := \rho(X_{t_{i+1}}^\pi)\Delta W_{i+1}^l \in \mathbb{L}^2(\mathbb{P})$:*

$$\mathbb{E}\left[F | X_{t_i}^\pi = x\right] = \frac{\mathbb{E}\left[Q^F[h_i, \phi^F](x)\right]}{\mathbb{E}\left[Q^1[h_i, \phi^1](x)\right]} \quad \text{for all } \phi^1, \phi^F \in \mathcal{L}_0,$$

where $\Delta W_{i+1}^0 \equiv 1$, and for $\phi \in \mathcal{L}_0$,

$$Q^F[h_i, \phi](x) := H_x(X_{t_i}^\pi)\rho(X_{t_{i+1}}^\pi) S_{(1,\ldots,d)}^{h_i}[\Delta W_{i+1}^l \phi(X_{t_i}^\pi - x)], \quad \text{and}$$

$$H_x(y) := \prod_{i=1}^{d} 1_{x_i \leq y_i},$$

$$Q^1[h_i, \phi](x) := H_x(X_{t_i}^\pi) S_{(1,\ldots,d)}^{h_i}[\phi(X_{t_i}^\pi - x)].$$

We note that $q_i^1(x) = \mathbb{E}[Q^1[h_i, \phi^1](x)]$ is the density of $X_{t_i}^\pi$. The above representation paves the way for a Monte Carlo approach to approximate the regression function $\mathbb{E}[F | X_{t_i}^\pi = x]$ by using a ratio of empirical means. Since empirical means are asymptotically distributed as Gaussian random variables, the potential integrability problems are avoided by making use of the a priori bounds derived in Lemma 3.3, [6] and the corresponding truncation operators:

$$\begin{aligned}
&\underline{\Psi}_i(X_{t_i}^\pi) \leq Y_{t_i}^\pi \leq \overline{\Psi}_i(X_{t_i}^\pi), \qquad T_i^\Psi(y) := \underline{\Psi}_i(X_{t_i}^\pi) \wedge y \vee \overline{\Psi}_i(X_{t_i}^\pi) \\
&\underline{\zeta}_{i-1}(X_{t_{i-1}}^\pi) \leq \mathbb{E}_{i-1}[Y_{t_i}^\pi \Delta W_i^l] \leq \overline{\zeta}_{i-1}(X_{t_{i-1}}^\pi), \\
&T_i^\zeta(y) := \underline{\zeta}_i(X_{t_i}^\pi) \wedge y \vee \overline{\zeta}_i(X_{t_i}^\pi), \quad i \leq n,\ l \leq d.
\end{aligned} \tag{3.3}$$

We now have all ingredients for the backward algorithm suggested in [6]. Let us denote by $(\Omega^0, \mathcal{F}^0, \mathbb{P}^0, \{\mathcal{F}_t^0\})$ the original filtered probability space and $\{(\Omega^i, \mathcal{F}^i, \mathbb{P}^i, \{\mathcal{F}_t^i\})\}_{i=1}^n$ be n independent copies of it. In each of these spaces, we let $\{X^{\pi^j}, j \in \mathcal{N}_i\}, \mathcal{N}_i = \{Ni+1, \ldots, N(i+1)\}$ be N copies of the Euler approximation of the forward diffusion. Therefore we consider nN copies of the Euler approximation of the forward diffusion, N copies at every point t_k on the grid. By abuse of notation, we denote by

$$(\Omega, \mathcal{F}, \mathbb{P}) = \left(\otimes_{i=0}^n \Omega^i, \otimes_{i=0}^n \mathcal{F}^i, \otimes_{i=0}^n \mathbb{P}^i\right) \tag{3.4}$$

the product probability space on which all processes are lifted. On this space, integration with respect to marginal measures $\mathbb{P}^i$, $\mathbb{P}^i \otimes \mathbb{P}^k$, is denoted by $\mathbb{E}^i$, $\mathbb{E}^{i,k}$, etc.

For a Lipschitz function $g : \mathbb{R}^d \to \mathbb{R}$ and an integer $0 \leq l \leq d$, following Theorem 3.1 we introduce the approximation $\hat{\mathbb{E}}_{t_i,x}[g(X_{t_{i+1}}^\pi)\Delta W_{i+1}^l]$ defined by:

$$\hat{\mathbb{E}}_{t_i,x}\left[g(X_{t_{i+1}}^\pi)\Delta W_{i+1}^l\right] = T_i^\zeta\left(\frac{\hat{Q}^F[h_i, \phi^F](x)}{\hat{Q}^1[h_i, \phi^1](x)}\right),$$

Please cite this article in press as: D. Crisan, et al., On the Monte Carlo simulation of BSDEs: An improvement on the Malliavin weights, Stochastic Processes and their Applications (2010), doi:10.1016/j.spa.2010.03.015

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

$$\hat{Q}^F[h_i,\phi^F](x)=\frac{1}{N}\sum_{k\in\mathcal{N}_{i+1}} H_x(X_{t_i}^{\pi^{(k)}})g(X_{t_{i+1}}^{\pi^{(k)}})S^{h_i^{(k)}}\left[\phi^F\left(X_{t_i}^{\pi^{(k)}}-x\right)\Delta W_{i+1}^{l^{(k)}}\right] \tag{3.5}$$

$$\hat{Q}^1[h_i,\phi^1](x)=\frac{1}{N}\sum_{k\in\mathcal{N}_{i+1}} H_x(X_{t_i}^{\pi^{(k)}})S^{h_i^{(k)}}\left[\phi^1\left(X_{t_i}^{\pi^{(k)}}-x\right)\right]. \tag{3.6}$$

Here, $\Delta W_{i+1}^{0^{(k)}}\equiv 1$ and, for $1\le l\le d$, $\Delta W_{i+1}^{l^{(k)}}$ is the k-th copy in bundle $\mathcal{N}_{i+1}$, of the l-th component of the Brownian motion. 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, $\hat{\mathbb{E}}$ is just the truncation of the empirical mean

$$\hat{\mathbb{E}}_{0,x}\left[g\left(X_{t_1}^{\pi}\Delta W_1{}^{l}\right)\right]=T_0^{\zeta}\left(\frac{1}{N}\sum_{k\in\mathcal{N}_1} g\left(X_{t_1}^{\pi^{(k)}}\right)\Delta W_1^{l^{(k)}}\right).$$

The bounds in the values of Y^{π} are used in the definition of the operators $\hat{R}_i$, $i<n$:

$$\begin{aligned}&\hat{R}_i g(x)=T_i^{\Psi}\left(\hat{R}_i^0 g(x)\right),\quad i=0,1,\dots,n-1,\\ &\hat{R}_i^0 g(x)=\hat{\mathbb{E}}_{t_i,x}\left[g(X_{t_{i+1}}^{\pi})\right]+|\pi| f\left(t_i,x,\hat{R}_i^0 g(x),|\pi|^{-1}\hat{\mathbb{E}}_{t_i,x}\left[g(X_{t_{i+1}}^{\pi})\Delta W_{i+1}\right]\right).\end{aligned} \tag{3.7}$$

Remark 3.2. By the structure of the above simulations, a careful inspection of the algorithm shows that Theorem 3.1 needs to be applied to random functions of the form $\varrho(x,\xi)$ where ξ is independent of $\mathcal{F}^{\pi}$. This is due to the fact that at each step t_i of the grid, the algorithm makes use of the N copies of the bundle $\mathcal{N}_i$ which are independent of those used in the other steps. Notice that the conclusions of Theorem 3.1 remain valid in this context.

In the remaining part of this section, we review the main steps of the proof of the error estimate $|Y_{t_0}^{\pi}-\hat{Y}_{t_0}^{\pi}|$ which will be needed for the proof of our main result in the subsequent section. The keypoint of the proof is the following estimate on the regression error, proved in [6]:

Proposition 3.3. *Let $i<n$, and $l\le d$ be fixed, and consider the random variable $F:=\varrho(X_{t_{i+1}}^{\pi},\omega)\Delta W_{i+1}^{l}$, where $\varrho(\cdot,\omega)$ is a random function with $\omega\in\Omega^{i+2}\times\cdots\times\Omega^n$.*

For $\phi^F,\phi^1\in\mathcal{L}_0$, let $q_i^F(x):=\mathbb{E}[Q^F[h_i,\phi^F](x)]$, where $Q^F[h_i,\phi^F](x)$ is defined as in Theorem 3.1, and set

$$r_i^{\pi}(x):=q_i^F(x)/q_i^1(x)\quad\text{and}\quad \hat{r}_i^{\pi}(x):=T_i^{\zeta}\left(\frac{\hat{Q}_i^F[h_i,\phi](x)}{\hat{Q}_i^1[h_i,\psi](x)}\right).$$

Then, for any $p>1$, we have the error estimate:

$$\left\|r_i^{\pi}(X_{t_i}^{\pi})-\hat{r}_i^{\pi}(X_{t_i}^{\pi})\right\|_p\le N^{-1/2p}\left(\Gamma(F,\phi^F,\phi^1)\right)^{1/p},$$

where

$$\begin{aligned}\Gamma(F,\phi^F,\phi^1):=2\int_{\mathbb{R}^d}\gamma(x)\Big(&\|Q^F[h_i,\phi](x)\|_{L^2(\bar{\mathbb{P}})}\\&+(|\bar{r}(x)|+\gamma(x))\|Q^1[h_i,\psi](x)\|_{L^2(\bar{\mathbb{P}})}\Big)\,\mathrm{d}x,\end{aligned}$$

Please cite this article in press as: D. Crisan, et al., On the Monte Carlo simulation of BSDEs: An improvement on the Malliavin weights, Stochastic Processes and their Applications (2010), doi:10.1016/j.spa.2010.03.015

$$\gamma(x) := \sup_{|\pi|} \max_{0\le i\le n} |\overline{\Psi}_i(x) - \underline{\Psi}_i(x)| \vee |\overline{\zeta}_i(x) - \underline{\zeta}_i(x)|, \qquad \bar{r}(x) := \sup_{|\pi|} \max_{0\le i\le n} r_i^\pi(x).$$

For the sake of completeness, we include the proof of Proposition 3.3 in the Appendix. In particular, the factorisation (3.4) of the probability space $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{\mathbb{P}})$ clarifies some points in the corresponding proof of [6].

The above result provides a control on the error of the simulation of the conditional expectation. The error is controlled by means of the parameter N, which is the number of simulated paths used at time t_i for the Monte Carlo method. However, there is a further negative influence of the partition mesh $|\pi|$ hidden in the functional Γ because the norms of the involved random variables explode as the partition mesh shrinks to zero. In order to control this effect, a further dependence on the partition mesh is included in the localizing functions:

Lemma 3.4 (*[6]*). *For $\phi \in \mathcal{L}_0$, set $\phi_{1/2}(x) := \phi(|\pi|^{-1/2}x)$ and let F be defined as in Proposition 3.3. Assume further that $F \in L^{2+\epsilon}$ for some $\epsilon > 0$. Then, for any function $\mu : \mathbb{R}^d \to \mathbb{R}$ with polynomial growth, we have:*

$$\limsup_{|\pi|\to 0} |\pi|^{d/4} \max_{1\le i\le n} \int_{\mathbb{R}^d} \mu(x) \left\| Q^F[h_i, \phi_{1/2}](x) \right\|_{L^2(\bar{\mathbb{P}})} \mathrm{d}x < \infty.$$

With this additional estimate, we obtain the following:

Theorem 3.5 (*[6]*). *Assume that* (H1b) *and* (H2) *hold true. Then, the following error estimate holds*

$$|Y_{t_0}^\pi - \hat{Y}_{t_0}^\pi| \le C|\pi|^{-1-d/4p} N^{-1/2p},$$

for a constant C independent of the partition and N.

Proof of Theorem 3.5. An easy induction argument shows that the random variable $V_i(x) := \hat{R}_{i,n-1} g(x)$ is of the form $\varrho(x, \omega)$ for $\omega \in \Omega^{i+2} \times \cdots \times \Omega^n$, for some appropriate random function ϱ. By Proposition 3.3, this provides

$$\left\| \left(\mathbb{E} - \hat{\mathbb{E}}\right) \left[V_i\left(X_{t_i}^\pi\right) \Delta W_{i+1}^l \right] \right\|_p \le N^{-1/2p} \Gamma\left(V_i\left(X_{t_i}^\pi\right) \Delta W_{i+1}^l, \phi_{1/2}^F, \phi_{1/2}^1 \right)^{1/p}$$
$$\text{for } 0 \le l \le d.$$

Observing that the functions $r(x)$ and $\gamma(x)$ have linear growth, see [6], it follows from Lemma 3.4 that

$$\max_{0\le i\le n} \left\| \left(\mathbb{E} - \hat{\mathbb{E}}\right) \left[V_i\left(X_{t_i}^\pi\right) \Delta W_{i+1}^l \right] \right\|_p \le C N^{-1/2p} |\pi|^{-d/4p},$$

and the required result follows from Proposition 2.1. □

4. An improvement on the Malliavin weights

4.1. The reduced scheme

The Malliavin calculus algorithm provides an efficient method for the numerical solution of a BSDE. However, a serious drawback of this method is that it can become quite heavy from the implementation viewpoint when the dimension is high. Our aim in this section is to present a

Please cite this article in press as: D. Crisan, et al., On the Monte Carlo simulation of BSDEs: An improvement on the Malliavin weights, Stochastic Processes and their Applications (2010), doi:10.1016/j.spa.2010.03.015

variation of the algorithm of [6] that reduces the computational effort. We will show that one can consider the matrix process

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

in place of h_i and $S^{\bar{h}_i}$ in place of S^{h_i}, and form a backward induction scheme based on these new weights.

Of course when σ^0 and σ are constant, there is no difference between h and $\bar{h}$. Let us try to appreciate the gain from the proposed truncation when σ^0 and/or σ are not constant. The truncation involves the following two terms:

$$\sigma^{-1}(X^\pi_{t_i})\nabla\sigma^0(X^\pi_{t_i})1_{[t_i,t_{i+1})} \quad \text{and} \quad \frac{1}{|\pi|}\sigma^{-1}(X^\pi_{t_i})\sum_{j=1}^{d}\nabla\sigma^j(X^\pi_{t_i})\Delta W^j_{i+1}1_{[t_i,t_{i+1})}$$

(to simplify the discussion we treat the sum as being a single term). Under this assumption, h_i has four terms whilst $\bar{h}_i$ has only two terms. It follows that the iterated Skorohod integral $S^{h_i}[\Delta W^l_{i+1}\phi(X^\pi_{t_i} - x)]$ has 4^d terms whilst $S^{\bar{h}_i}[\Delta W^l_{i+1}\phi(X^\pi_{t_i} - x)]$ has 2^d terms. Therefore the new algorithm requires $1/2^d$ of the original terms to be implemented per Malliavin weight (recall that there are $d+1$ expectations to be computed per discretization step). Moreover by truncating the h_i's, we avoid the computation of the derivatives of σ, σ^0, hence we expect the new algorithm to be more stable than the original.

However, this reduction in the implementation complexity does not get equivalently reflected on the time performance of the algorithm. This is due to the fact that every backward empirical projection is of complexity $O(N\log^{d-1}N)$, N being the number of copies of the forward Euler scheme generated at each step and d the dimension of the state space. On the other hand, one can implement the algorithm in a way such that the complexity of the computation of the weights is merely $O(N)$. Hence as N increases, the time spent in the computation of the weights will be dominated by other tasks. We will explain how this works in Section 5 where we include a numerical example.

Remark 4.1. The truncation of the h process does not affect the denominators in the representation of Theorem 3.1. Indeed, a repetitive application of Lemma 3.2.1 of [22] shows that

$$q^1_i(x) = \mathbb{E}\left[H_x(X^\pi_{t_i})\mathbb{E}_i\left[S^{h_i}[\phi(X^\pi_{t_i} - x)]\right]\right] = \mathbb{E}\left[H_x(X^\pi_{t_i})S^{\bar{h}_i}[\phi(X^\pi_{t_i} - x)]\right]. \tag{4.1}$$

In other words, the truncation induces no loss of precision in the denominator. Given this, our effort will focus on the truncation of the numerator.

We need to re-define the regression function according to the new "truncated" weights. For a localizing function $\phi \in \mathcal{L}_0$ we denote by

$$\phi_\alpha(x) := \phi\left(x|\pi|^{-\alpha}\right), \quad \alpha \geq 1/2,\ x \in \mathbb{R}^d. \tag{4.2}$$

Obviously $\phi_\alpha \in \mathcal{L}_0$. The parameter α is chosen according to the asymptotic result of Proposition 4.3 below.

Please cite this article in press as: D. Crisan, et al., On the Monte Carlo simulation of BSDEs: An improvement on the Malliavin weights, Stochastic Processes and their Applications (2010), doi:10.1016/j.spa.2010.03.015

Next, for $F = \varrho(X^\pi_{t_{i+1}})\Delta W^l_{i+1}$, $0 \le i < n$, $0 \le l \le d$, and $\phi, \psi \in \mathcal{L}_0$, we introduce the alternative regression approximation operator:

$$\bar{\mathbb{E}}_{t_i,x}\left[\varrho(X^\pi_{t_{i+1}})\Delta W^l_{i+1}\right] = T_i^\zeta\left(\bar{q}^F(x)/\bar{q}^1(x)\right)$$

where

$$\bar{q}^F(x) := \mathbb{E}\left[Q^F[\bar{h}_i, \phi_\alpha](x)\right], \qquad \bar{q}^1(x) := \mathbb{E}\left[Q^1[\bar{h}_i, \psi_\alpha](x)\right]$$

and the corresponding one-step ahead approximation of the BSDE:

$$\begin{aligned}
&\bar{R}_i g(x) = T_i^\Psi\left(\bar{R}^0_i g(x)\right),\\
&\bar{R}^0_i g(x) = \bar{\mathbb{E}}_{t_i,x}\left[g\left(X^\pi_{t_{i+1}}\right)\right] + |\pi| f\left(t_i, x, \bar{R}^0_i g(x), |\pi|^{-1}\bar{\mathbb{E}}_{t_i,x}\left[g\left(X^\pi_{t_{i+1}}\right)\Delta W_{i+1}\right]\right),
\end{aligned} \tag{4.3}$$

for $i = n-1, \dots, 0$ and any function g in a suitable class. The iteration of the above family will give us the values

$$\bar{Y}^\pi_{t_i} := \bar{R}_i \cdots \bar{R}_{n-1}\Phi(X^\pi_{t_i}), \quad 0 \le i < n.$$

Finally we introduce the family of operators $\{\hat{\bar{R}}_i\}_{i=1}^n$ corresponding to the Monte Carlo simulation-based estimation of the family $\{\bar{R}_i\}_{i=1}^n$, i.e. the analogue of (3.7), with $\bar{h}_i(t)$ substituted to $h_i(t)$ for every step i:

$$\hat{\bar{\mathbb{E}}}_{t_i,x}\left[g(X^\pi_{t_{i+1}})\Delta W^l_{i+1}\right] = T_i^\zeta\left(\frac{\hat{Q}^F[\bar{h}_i, \phi^F_\alpha](x)}{\hat{Q}^1[\bar{h}_i, \phi^1_\alpha](x)}\right),$$

where ϕ^1, ϕ^F are possibly different localizing functions in $\mathcal{L}_0$, $\hat{Q}^F_i$ are defined in (3.5), and

$$\begin{aligned}
&\hat{\bar{R}}_i g(x) = T_i^\Psi\left(\hat{\bar{R}}^0_i g(x)\right), \quad i = 0, 1, \dots, n-1,\\
&\hat{\bar{R}}^0_i g(x) = \hat{\bar{\mathbb{E}}}_{t_i,x}\left[g(X^\pi_{t_{i+1}})\right] + |\pi| f\left(t_i, x, \hat{\bar{R}}^0_i g(x), |\pi|^{-1}\hat{\bar{\mathbb{E}}}_{t_i,x}\left[g(X^\pi_{t_{i+1}})\Delta W_{i+1}\right]\right).
\end{aligned} \tag{4.4}$$

The main result of this section provides a rate of convergence of the simplified scheme:

$$\hat{\bar{Y}}^\pi_{t_n} := \Phi(X_{t_n}) \quad \text{and, for } 0 \le i < n, \quad \hat{\bar{Y}}^\pi_{t_i} := \hat{\bar{R}}_{i,n-1}\Phi(X_{t_i}),$$

where $\{\hat{\bar{R}}_{i,n-1}\Phi\}_{0\le i\le n-1}$ is defined as in (2.3).

Remark 4.2. As in the previous section, it is an easy induction argument to show that for any $i = 0, 1, \dots, n-1$, the iteration of the family $\{\hat{\bar{R}}_i\}_{i=1}^n$ produces random functions that may be written as $\hat{\bar{R}}_i \dots \hat{\bar{R}}_n \Phi(x, \chi)$ where χ is a functional of the copies of the Euler scheme $\{X^{\pi^{(k)}}, k = i+1, \dots, n\}$.

Please cite this article in press as: D. Crisan, et al., On the Monte Carlo simulation of BSDEs: An improvement on the Malliavin weights, Stochastic Processes and their Applications (2010), doi:10.1016/j.spa.2010.03.015

4.2. The weight truncation error

The first step in order to analyze the rate of convergence of the algorithm defined by the backward sequence $\{\hat{\bar{R}}_i, i \leq n\}$ is to estimate the error $\mathbb{E}_i - \bar{\mathbb{E}}_i$ for $i \leq n-1$, when applied to random functions of the form described in Remark 4.2 above.

Proposition 4.3. *Let $F := \varrho(X^\pi_{t_{i+1}}, \omega)$ for some function $\varrho(\cdot, \omega)$ with $\omega \in \Omega^{i+2} \times \cdots \times \Omega^n$.*

(i) *Assume that $\varrho(\cdot, \omega)$ is uniformly Lipschitz, then for $i < n$ and $p > 1$:*

$$\limsup_{|\pi|\to 0} |\pi|^{-\frac{1}{p}\left(\alpha+\frac{1}{2}\right)} \max_{0\leq l\leq d} \left\| \left(\mathbb{E} - \bar{\mathbb{E}}\right)_{t_i, X^\pi_{t_i}} [F \Delta W^l_{i+1}] \right\|_p < \infty.$$

(ii) *Assume that $F \in \mathbb{L}^q$, $q > 1$, then for $i < n$ and $p > 1$:*

$$\limsup_{|\pi|\to 0} |\pi|^{-\frac{\alpha}{p}} \max_{0\leq l\leq d} \left\| \left(\mathbb{E} - \bar{\mathbb{E}}\right)_{t_i, X^\pi_{t_i}} [F \Delta W^l_{i+1}] \right\|_p < \infty.$$

The rest of this subsection is dedicated to the proof of Proposition 4.3. We first observe that $\int D_t X^\pi_{t_i} \bar{h}_i(t) \mathrm{d}t = \mathbb{I}_d$, and therefore the expansion (3.2) holds for $\bar{h}_i$ as well:

$$S^{\bar{h}_i}_I [\Delta W^l_{i+1} \varphi(X^\pi_{t_i} - x)] = \sum_{J\subseteq I} (-1)^{|J|} \partial_J \varphi(X^\pi_{t_i} - x) S^{\bar{h}_i}_{J^c} [\Delta W^l_{i+1}], \tag{4.5}$$

for every $0 \leq l \leq d$. We then concentrate on the errors $S^{h^i}_I(\Delta W^l_{i+1}) - S^{\bar{h}^i}_I(\Delta W^l_{i+1})$.

Lemma 4.4. *For $I \in \mathcal{J}_m$, $1 \leq l \leq d$, and $p > 1$, we have:*

$$\left\| S^{h_i}_I[1] - S^{\bar{h}_i}_I[1] \right\|_p \leq C|\pi|^{(1-m)/2} \quad \text{and}$$

$$\left\| S^{h_i}_I[\Delta W^l_{i+1}] - S^{\bar{h}_i}_I[\Delta W^l_{i+1}] \right\|_p \leq C|\pi|^{1-m/2},$$

where C is a constant independent of the partition.

Proof. The linearity of the Skorohod integral implies that $S^{h_i}_j = S^{\bar{h}_i}_j + S^{h_i - \bar{h}_i}_j$ and hence, $S^{h_i}_{(i_i, i_2, \ldots, i_m)}[\cdot]$ can be written as a sum of 2^m terms of the form

$$S^{v_1}_{i_1} \circ \cdots \circ S^{v_m}_{i_m}[\cdot], \tag{4.6}$$

where vectors v_i are either columns of $\bar{h}_i$ or of $h_i - \bar{h}_i$. Hence we may focus on iterated integrals of the form (4.6) to prove our claim.

Given a real vector $v \in \mathbb{R}^m$ and a multi-index $\beta := (\beta_1, \ldots, \beta_m)$ with nonnegative integer entries, we denote by $v^\beta := v_1^{\beta_1} \ldots v_m^{\beta_m}$. We also write $|\beta| := \sum_{i=1}^m \beta_i$. Consider the collection of matrix valued processes

$$\mathcal{V}_i = \left\{ v(t) | v^j(t) = h^j_i(t) - \bar{h}^j_i(t) \text{ or } \bar{h}^j_i(t),\ j = 1, \ldots, d \right\},$$

i.e., it is the collection of all matrix valued processes with columns either equal to the corresponding column of $h_i(t) - \bar{h}_i(t)$ or $\bar{h}_i(t)$.

Elementary properties of the Skorohod integral imply that the iterated Skorohod integrals that we are considering here may be written as sums of products of Brownian increments. In fact

Please cite this article in press as: D. Crisan, et al., On the Monte Carlo simulation of BSDEs: An improvement on the Malliavin weights, Stochastic Processes and their Applications (2010), doi:10.1016/j.spa.2010.03.015

we claim that if k is a positive integer and β, γ are multi-indices then, for any $m \in \mathbb{Z}_+$ and $I = (i_1, \ldots, i_m) \in \mathcal{J}_m$ we have that,

$$S_I^{\nu}[1] = \frac{1}{|\pi|^m} \sum_{\substack{(k,\beta,\gamma):\\ m-q(I)/2 \le k + \frac{|\beta|+|\gamma|}{2} \le \frac{3m-q(I)}{2}}} \theta^I_{k,\beta,\gamma}\left(X^{\pi}_{t_{i-1}}, X^{\pi}_{t_i}\right) |\pi|^k \, (\Delta W_i)^{\beta} \, (\Delta W_{i+1})^{\gamma} \tag{4.7}$$

and, for $l > 0$,

$$\begin{aligned} S_I^{\nu}[\Delta W_{i+1}^l] = \frac{1}{|\pi|^m} & \sum_{\substack{(k,\beta,\gamma):\\ m-\frac{q(I)}{2}+\frac{1}{2} \le k + \frac{|\beta|+|\gamma|}{2} \le \frac{3m-q(I)+1}{2}}} \theta^I_{k,\beta,\gamma}\left(X^{\pi}_{t_{i-1}}, X^{\pi}_{t_i}\right) \\ & \times |\pi|^k \, (\Delta W_i)^{\beta} \, (\Delta W_{i+1})^{\gamma} \end{aligned} \tag{4.8}$$

where $q(I) := \#\{j \in I | v^j \text{ is a column from } \bar{h}_i\}$ and $(x, y) \to \theta^{(i_1,\ldots,i_m)}_{k,\alpha,\beta}(x, y) \in C_b^{\infty}(\mathbf{R}^d \times \mathbf{R}^d)$ (some of which may be identically equal to zero).

Assume for the moment that (4.7) and (4.8) hold true. Then, for any $\nu \in \mathcal{V}_i \setminus \{\bar{h}_i\}$ we have that,

$$\left\| S_I^{h_i}[1] - S_I^{\bar{h}_i}[1] \right\|_p \le \sum_{\nu \in \mathcal{V}_i \setminus \{\bar{h}_i\}} \|S_I^{\nu}[1]\|_p.$$

According to (4.7),

$$\left\| S_I^{\nu}[1] \right\|_p \le C|\pi|^{-q(I)/2} \le C|\pi|^{(1-m)/2}$$

since $\nu \ne \bar{h}_i$ means that $q(I) \le m - 1$. Similarly, (4.8) is used to prove the second claimed estimate.

It remains to show (4.7) and (4.8). The proof is with induction on m. First observe that

$$\begin{aligned} S_j^{h_i - \bar{h}_i}[F] = \frac{1}{|\pi|} \Bigg\{ & F\left(\sum_{s=0}^{d} \sigma^{-1}(t_i) \nabla \sigma^s(t_i) \Delta W_{i+1}^s\right)^j \cdot \Delta W_{i+1} \\ & - \int_{t_i}^{t_{i+1}} \left(\sum_{s=0}^{d} \sigma^{-1}(t_i) \nabla \sigma^s(t_i) \Delta W_{i+1}^s\right)^j \cdot D_s F \mathrm{d}s \\ & + \int_{t_i}^{t_{i+1}} F D_s \left(\sum_{s=0}^{d} \sigma^{-1}(t_i) \nabla \sigma^s(t_i) \Delta W_{i+1}^s\right)^j \mathrm{d}s \Bigg\} \end{aligned} \tag{4.9}$$

and

$$\begin{aligned} S_j^{\bar{h}_i}[F] = \frac{1}{|\pi|} \Bigg\{ & F\left(\sigma^{-1}(t_{i-1})\right)^j \cdot \Delta W_i - F\left(\sigma^{-1}(t_i)\right)^j \cdot \Delta W_{i+1} \\ & - \int_{t_{i-1}}^{t_i} (\sigma^{-1}(t_{i-1}))^j \cdot D_s F \mathrm{d}s - \int_{t_i}^{t_{i+1}} (\sigma^{-1}(t_{i_1}))^j \cdot D_s F \mathrm{d}s \Bigg\}. \end{aligned} \tag{4.10}$$

By substituting $F = 1$ or ΔW_{i+1}^l, $l = 1, \ldots, d$ it is easy to verify (4.7) and (4.8) for $m = 1$. Let us assume that our claim is true for m and focus on proving (4.7) for $m + 1$. Given the induction

Please cite this article in press as: D. Crisan, et al., On the Monte Carlo simulation of BSDEs: An improvement on the Malliavin weights, Stochastic Processes and their Applications (2010), doi:10.1016/j.spa.2010.03.015

hypothesis, a multi-index $I \in \mathcal{J}_{m+1}$ and a matrix valued process $\nu \in \mathcal{V}_i$, we have

$$S_I^{\nu}[1] = \frac{1}{|\pi|^m} S^{\nu^{i_1}} \times \left[\sum_{\substack{(k,\beta,\gamma):\\ m-q(-I)/2 \le k + \frac{|\beta|+|\gamma|}{2} \le \frac{3m-q(-I)}{2}}} \theta_{k,\alpha,\beta}^{-I} \left(X_{t_{i-1}}^{\pi}, X_{t_i}^{\pi} \right) |\pi|^k \, (\Delta W_i)^{\beta} \, (\Delta W_{i+1})^{\gamma} \right].$$

Assume that $\nu^{i_1} = \bar{h}_i^{i_1}$. The application of the operator $S^{\nu_{i_1}}[\cdot]$ to a random variable of the form

$$F = \theta_{k_F,\beta_F,\gamma_F}^{-I} \left(X_{t_{i-1}}^{\pi}, X_{t_i}^{\pi} \right) |\pi|^k \, (\Delta W_i)^{\beta_F} \, (\Delta W_{i+1})^{\gamma_F}, \tag{4.11}$$

where (k_F, β_F, γ_F) is such that $m - q(-I)/2 \le k_F + \frac{|\beta_F|+|\gamma_F|}{2} \le \frac{3m-q(-I)}{2}$, generates four terms according to (4.7). If we write $\lambda_0 = k_F + \frac{|\beta_F|+|\gamma_F|}{2}$ then,

$$F_1 := F\left(\sigma^{-1}(t_{i-1})\right)^j \cdot \Delta W_i, \qquad F_2 := F\left(\sigma^{-1}(t_i)\right)^j \cdot \Delta W_{i+1}$$

are terms with

$$k_{F_j} + \frac{|\beta_{F_j}| + |\gamma_{F_j}|}{2} = \lambda_0 + \frac{1}{2}, \quad j = 1, 2. \tag{4.12}$$

As for the last two terms generated by the application of $S_{i_1}^{\bar{h}_i}[\cdot]$ on F observe that, according to the chain rule for Malliavin differentiation, we have for $j = 1, \ldots, d$

$$\begin{aligned} D_s^j F = {} & \nabla_y \theta_{k,\beta,\gamma}^{-I} \left(X_{t_{i-1}}^{\pi}, X_{t_i}^{\pi} \right) \cdot \sigma(t_{i-1})^j |\pi|^k (\Delta W_i)^{\beta} (\Delta W_{i+1})^{\gamma} 1_{[t_{i-1},t_i)}(s) \\ & + |\pi|^k \theta_{k,\beta,\gamma}^{-I} \left(X_{t_{i-1}}^{\pi}, X_{t_i}^{\pi} \right) (\Delta W_{i+1})^{\gamma} \beta_j (\Delta W_i)^{(\beta_1,\ldots,\beta_j-1,\ldots,\beta_d)} 1_{[t_{i-1},t_i)}(s) \\ & + |\pi|^k \theta_{k,\beta,\gamma}^{-I} \left(X_{t_{i-1}}^{\pi}, X_{t_i}^{\pi} \right) (\Delta W_i)^{\beta} \gamma_j (\Delta W_{i+1})^{(\gamma_1,\ldots,\gamma_j-1,\ldots,\gamma_d)} 1_{[t_i,t_{i+1})}(s). \end{aligned} \tag{4.13}$$

Substituting (4.13) in the last two terms of (4.7) we obtain two terms with

$$k_F + \frac{|\beta_F| + |\gamma_F|}{2} = \lambda_0 + 1. \tag{4.14}$$

The linearity of the Skorohod integral, (4.12) and (4.14) prove the induction hypothesis when $\nu^{i_1} = \bar{h}^{i_1}$. The case $\nu^{i_1} = h^{i_1} - \bar{h}^{i_1}$ is treated with identical arguments. The proof, by induction, for (4.8) is of course completely analogous. □

We now turn to the

Proof of Proposition 4.3. Throughout this proof, all expectations are in fact conditional on $\omega \in \Omega^{i+2}, \ldots, \Omega^n$, and we omit the dependence of ϱ on this independent source of randomness. All calculations are then similar to the case where ϱ is a Lipschitz deterministic function. Let

$$F := \varrho(X_{t_{i+1}}^{\pi}) \Delta W_{i+1}^l, \qquad \nu_i(x) := \bar{q}^F(x)/\bar{q}^1(x) \quad \text{and}$$
$$\mathcal{S}_J(G) := \left(S_J^{h_i} - S_J^{\bar{h}_i} \right) [\Delta W_{i+1}^l G].$$

Please cite this article in press as: D. Crisan, et al., On the Monte Carlo simulation of BSDEs: An improvement on the Malliavin weights, Stochastic Processes and their Applications (2010), doi:10.1016/j.spa.2010.03.015

Using the preliminary bounds (3.3), we have

$$\begin{aligned}\left\|\left(\mathbb{E}-\bar{\mathbb{E}}\right)_{t_i,X^\pi_{t_i}}[F]\right\|_p^p &\le \mathbb{E}\left[\left|v_i(X^\pi_{t_i})-r^\pi_i(X^\pi_{t_i})\right|^p \wedge \gamma(X^\pi_{t_i})^p\right]\\ &\le \mathbb{E}\left[\left|v_i(X^\pi_{t_i})-r^\pi_i(X^\pi_{t_i})\right|\gamma(X^\pi_{t_i})^{p-1}\right]\\ &= \mathbb{E}\left[\left(\left|\frac{q^F-\bar{q}^F}{q^1}\right|\gamma\right)(X^\pi_{t_i})^{p-1}\right]\\ &= \int_{\mathbb{R}^d}\left|\mathbb{E}\left[H_x(X^\pi_{t_i})\varrho(X^\pi_{t_{i+1}})\mathcal{S}\left(\phi_\alpha(X^\pi_{t_i}-x)\right)\right]\right|\gamma(x)^{p-1}\mathrm{d}x\end{aligned}$$

since q^1 is just the density of $X^\pi_{t_i}$. Denoting $\mathcal{J}:=\bigcup_{k<d}\mathcal{J}_k$, and using the representation (3.2), this provides

$$\begin{aligned}&\left\|\left(\mathbb{E}-\bar{\mathbb{E}}\right)_{t_i,X^\pi_{t_i}}[F]\right\|_p^p\\ &\le \sum_{J\in\mathcal{J}}\int_{\mathbb{R}^d}\left|\mathbb{E}\left[H_x(X^\pi_{t_i})\varrho(X^\pi_{t_{i+1}})\partial_J\phi_\alpha(X^\pi_{t_i}-x)\mathcal{S}_{J^c}(1)\right]\right|\gamma(x)^{p-1}\,\mathrm{d}x\\ &= \sum_{J\in\mathcal{J}}\int_{\mathbb{R}^d}\left|\mathbb{E}\left[H_x(X^\pi_{t_i})\left(\varrho(X^\pi_{t_{i+1}})-\varrho(X^\pi_{t_i})\right)\partial_J\phi_\alpha(X^\pi_{t_i}-x)\mathcal{S}_{J^c}(1)\right]\right|\gamma(x)^{p-1}\,\mathrm{d}x\end{aligned}$$

where the last equality follows from the fact that $\mathbb{E}[\mathcal{S}_{J^c}(1)|\mathcal{F}_{t_i}]=0,\forall J^c\in\mathcal{J}$, which can be seen by arguing as in (4.1) and using Lemma 3.2.1 of [22].

Our next step is to perform an ω-by-ω change of variables for the Riemann integral by setting $u=(X^\pi_{t_i}-x)/|\pi|^\alpha$. Observe that the definition of the localizing function ϕ_α yields $\partial_J\phi_\alpha(x)=|\pi|^{-\alpha|J|}\partial_J\phi(|\pi|^{-\alpha}x)$. Hence we have that

$$\begin{aligned}\left\|\left(\mathbb{E}-\bar{\mathbb{E}}\right)_{t_i,X^\pi_{t_i}}[F]\right\|_p^p \le \sum_{J\in\mathcal{J}}|\pi|^{\alpha(d-|J|)}\int_{\mathbb{R}^d_+}&\left|\mathbb{E}\left[\left(\varrho(X^\pi_{t_{i+1}})-\varrho(X^\pi_{t_i})\right)\right.\right.\\ &\left.\left.\times\mathcal{S}_{J^c}(1)\gamma(X^\pi_{t_i}-|\pi|^\alpha u)^{p-1}\right]\right||\partial_J\phi(u)|\,\mathrm{d}u.\end{aligned}$$

The Lipschitz assumption on ϱ, the obvious estimate $\|X^\pi_{t_{i+1}}-X^\pi_{t_i}\|_k\le C|\pi|^{1/2}$ for any $k\ge1$ and Hölder's inequality give us

$$\begin{aligned}\left\|\left(\mathbb{E}-\bar{\mathbb{E}}_{t_i,X^\pi_{t_i}}\right)[F]\right\|_p^p \le C\sum_{J\in\mathcal{J}}&|\pi|^{\alpha(d-|J|)}\|\varrho(X^\pi_{t_{i+1}})-\varrho(X^\pi_{t_i})\|_p\|\mathcal{S}_{J^c}(1)\|_{q_1}\\ &\times\int_{\mathbb{R}^d_+}\|\gamma(X^\pi_{t_i}-|\pi|^\alpha u)^{p-1}\|_{q_2}|\partial_J\phi(u)|\mathrm{d}u\end{aligned}$$

for some $q_1,q_2>1$. It is clear from Lemma 4.4, that the term

$$\|\mathcal{S}_{J^c}(1)\|_{q_1}\equiv\left\|\left(S^{h_i}_{J^c}-S^{\bar{h}_i}_{J^c}\right)\left[\Delta W^l_{i+1}\right]\right\|_{q_1}$$

has a negative impact on the error estimate above. Moreover, according to the results of this lemma, this impact is worse when $l=0$. So let us fix this value for l to obtain

$$\left\|\left(\mathbb{E}-\bar{\mathbb{E}}_{t_i,X^\pi_{t_i}}\right)[F]\right\|_p^p$$

Please cite this article in press as: D. Crisan, et al., On the Monte Carlo simulation of BSDEs: An improvement on the Malliavin weights, Stochastic Processes and their Applications (2010), doi:10.1016/j.spa.2010.03.015

$$\leq C \sum_{J\in\mathcal{J}} |\pi|^{1+\left(\alpha-\frac{1}{2}\right)(d-|J|)} \int_{\mathbb{R}^d_+} \|\gamma(X^\pi_{t_i} - |\pi|^\alpha u)^{p-1}\|_{q_2} |\partial_J \phi(u)| \mathrm{d}u. \tag{4.15}$$

Eq. (4.15) should explain why we need to choose $\alpha > 1/2$, as for $\alpha = 1/2$ we obtain a local error merely of order $|\pi|$ which will cancel under summation (one has $1/|\pi|$ local errors to sum up).

Finally, since the function $\gamma(\cdot)$ is of polynomial growth we have, with r an integer, that

$$\int_{\mathbb{R}^d_+} \mathbb{E}\left[\gamma(X^\pi_{t_i} - |\pi|^\alpha u)^{q_2(p-1)}\right]^{1/q_2} |\partial_J \phi(u)| \mathrm{d}u$$

$$\leq C \int_{\mathbb{R}^d_+} \mathbb{E}\left[\left(1 + \sum_{k=0}^{r} \binom{r}{k} |X^\pi_{t_i}|^k |\,|\pi|^\alpha u|^{r-k}\right)^{q_2(p-1)}\right]^{1/q_2} |\partial_J \phi(u)| \mathrm{d}u,$$

and since $|\partial_J \phi(u)|$ integrates against polynomials, by the assumptions on the localizing function, we have shown the first assertion of the theorem. The second part of the proposition follows from (4.15) without using the extra $|\pi|^{1/2}$ coming from the $\mathbb{L}^p$ estimate of $\|\varrho(X^\pi_{t_{i+1}}) - \varrho(X^\pi_{t_i})\|_p$. □

4.3. The main result

We have already seen at the end of the proof of Proposition 4.3 that we need to consider an $\alpha > 1/2$ in the normalization of the localizing function and this will have a negative impact, relative to the results of the previous section, on the integrated variance that controls the error of the simulation-based estimation of the family $\{\bar{R}_i\}_{i=0}^{n-1}$. In particular using the assumptions and the notation of Lemma 3.4, we may show with identical arguments that for a function $\mu(\cdot)$ of polynomial growth and a random variable F defined as in 4.3:

$$\limsup_{|\pi|\to 0} |\pi|^{\alpha d/2} \max_{1\leq i\leq n} \int_{\mathbb{R}^d} \mu(x) \|Q^F[\bar{h}_i, \phi_\alpha](x)\|_2 \mathrm{d}x < \infty. \tag{4.16}$$

The following main result provides an upper bound on the rate of convergence for the global error, which says roughly that in order to guarantee an accuracy of order $n^{-1/2}$, one needs to simulate $N = n^{3p+\frac{(3p-1)d}{2}}$ copies of the Euler scheme at every time step.

Theorem 4.5. *Let π be a partition of $[0,1]$ with $K|\pi| < 1$, $\phi \in \mathcal{L}_0$, and define $\phi_\alpha \in \mathcal{L}_0$ as in (4.2). Consider the three families of operators $\{R_i\}_{i=0}^{n-1}, \{\bar{R}_i\}_{i=0}^{n-1}, \{\hat{\bar{R}}_i\}_{i=0}^{n-1}$ defined with ϕ_α as the localizing function, on the partition π. Then, for any $p > 1$ and $\alpha \geq (3p-1)/2$, we have*

$$\max_{0\leq i\leq n-1} \|Y^\pi_{t_i} - \hat{\bar{Y}}^\pi_{t_i}\|_p \leq C\left(|\pi|^{\frac{1}{p}\left(\frac{1}{2}+\alpha\right)-1} + |\pi|^{-1-\alpha d/2p} N^{-1/2p}\right)$$

for a constant C independent of the partition.

Proof. We will estimate $|Y^\pi_{t_0} - \hat{\bar{Y}}^\pi_{t_0}|$ and remark that the estimate for $i > 0$ follows identical arguments. Once again, we use a Trotter product expansion for the error:

$$|Y^\pi_{t_0} - \hat{\bar{Y}}^\pi_{t_0}| = |R_0 \ldots R_{n-1}\Phi(x) - \hat{\bar{R}}_0 \ldots \hat{\bar{R}}_{n-1}\Phi(x)|$$

$$\leq \sum_{i=0}^{n-1} |R_{0,i}\hat{\bar{R}}_{i+1,n-1}\Phi(x) - R_{0,i-1}\hat{\bar{R}}_{i,n-1}\Phi(x)|$$

Please cite this article in press as: D. Crisan, et al., On the Monte Carlo simulation of BSDEs: An improvement on the Malliavin weights, Stochastic Processes and their Applications (2010), doi:10.1016/j.spa.2010.03.015

$$\leq \sum_{i=0}^{n-1} \left(\frac{1 + C|\pi|}{1 - K|\pi|} \right)^{i-1} \left\| (R_i - \hat{\bar{R}}_i) \hat{\bar{R}}_{i+1,n-1} \Phi(X^\pi_{t_i}) \right\|_p ,$$

where we have used (2.6) with $p > 1$. For every i we clearly have

$$\begin{aligned} &\left\| (R_i - \hat{\bar{R}}_i) \hat{\bar{R}}_{i+1,n-1} \Phi(X^\pi_{t_i}) \right\|_p \\ &\leq \left\| (R_i - \bar{R}_i) \hat{\bar{R}}_{i+1,n-1} \Phi(X^\pi_{t_i}) \right\|_p + \left\| (\bar{R}_i - \hat{\bar{R}}_i) \hat{\bar{R}}_{i+1,n-1} \Phi(X^\pi_{t_i}) \right\|_p . \end{aligned} \tag{4.17}$$

To estimate the second term above, we may appeal to Proposition 3.3. The proof of the latter comes through in this case to give us,

$$\begin{aligned} &\left\| (\bar{R}_i - \hat{\bar{R}}_i) \hat{\bar{R}}_{i+1,n-1} \Phi(X^\pi_{t_i}) \right\|_p \\ &\leq \left(\frac{2}{N^{1/2}} \int_{\mathbb{R}^d} \gamma(x) \left[\| Q^F[\bar{h}_i, \phi](x) \|_p + (|r(x)| + \gamma(x)) \| Q^1[\bar{h}_i, \psi](x) \|_p \right] \mathrm{d}x \right)^{1/p} . \end{aligned}$$

Plugging the estimate (4.16) in the above we get

$$\left\| (\bar{R}_i - \hat{\bar{R}}_i) \hat{\bar{R}}_{i+1,n-1} \Phi(X^\pi_{t_i}) \right\|_p \leq \frac{C}{|\pi|^{\alpha d/2p} N^{1/2p}} . \tag{4.18}$$

As for the first term in (4.17) we have

$$\begin{aligned} &\left\| (R_i - \bar{R}_i) \hat{\bar{R}}_{i+1,n-1} \Phi(X^\pi_{t_i}) \right\|_p \\ &\leq \frac{C}{1 - K|\pi|} \max_{0 \leq l \leq d} \left\| \left(\mathbb{E} - \bar{\mathbb{E}} \right)_{t_i, X^\pi_{t_i}} [\hat{\bar{R}}_{i+1,n-1} \Phi(X^\pi_{t_{i+1}}) \Delta W^l_{i+1}] \right\|_p \\ &\leq \frac{C}{1 - K|\pi|} \left(\max_{0 \leq l \leq d} \left\| \left(\mathbb{E} - \bar{\mathbb{E}} \right)_{t_i, X^\pi_{t_i}} [R_{i+1,n-1} \Phi(X^\pi_{t_{i+1}}) \Delta W^l_{i+1}] \right\|_p \right. \\ &\quad \left. + \max_{0 \leq l \leq d} \left\| \left(\mathbb{E} - \bar{\mathbb{E}} \right)_{t_i, X^\pi_{t_i}} [(R_{i+1,n-1} - \hat{\bar{R}}_{i+1,n-1}) \Phi(X^\pi_{t_{i+1}}) \Delta W^l_{i+1}] \right\|_p \right). \end{aligned}$$

We know from (2.4) that the functions $R_{i+1,n-1} \Phi$ are Lipschitz uniformly in i. On the other hand, at this stage nothing can be said about the Lipschitz regularity of $(R_{i+1,n-1} - \hat{\bar{R}}_{i+1,n-1}) \Phi(\cdot)$. Hence, we apply the first assertion of Proposition 4.3 to the first term above, while the second assertion is applied to the other one, to obtain

$$\begin{aligned} &\left\| (R_i - \bar{R}_i) \hat{\bar{R}}_{i+1,n-1} \Phi(X^\pi_{t_i}) \right\|_p \\ &\leq C \left(|\pi|^{\frac{\alpha+1/2}{p}} + |\pi|^{\frac{\alpha}{p}} \left\| (\hat{\bar{R}}_{i+1,n-1} - R_{i+1,n-1}) \Phi(X^\pi_{t_{i+1}}) \right\|_p^{1/p} \right). \end{aligned} \tag{4.19}$$

We would now like to eliminate the $1/p$-root above. To this end, observe that for any $x > 0$ and $p > 1$, it holds that $x^{1/p} \leq x + 1$. Hence

$$\begin{aligned} &\left\| (\hat{\bar{R}}_{i+1,n-1} - R_{i+1,n-1}) \Phi(X^\pi_{t_{i+1}}) \right\|_p^{1/p} \\ &\leq |\pi|^{\frac{1}{2p} - \frac{1}{2}} \left\| (\hat{\bar{R}}_{i+1,n-1} - R_{i+1,n-1}) \Phi(X^\pi_{t_{i+1}}) \right\|_p + |\pi|^{1/2p} . \end{aligned} \tag{4.20}$$

Please cite this article in press as: D. Crisan, et al., On the Monte Carlo simulation of BSDEs: An improvement on the Malliavin weights, Stochastic Processes and their Applications (2010), doi:10.1016/j.spa.2010.03.015

Putting together (4.18)–(4.20) we have that

$$\begin{aligned}\varepsilon_i &:= \left\| (R_i - \hat{\hat{R}}_i)\hat{\hat{R}}_{i+1,n-1}\,\Phi(X^\pi_{t_i}) \right\|_p \\ &\le C\left(|\pi|^{\frac{\alpha+1/2}{p}} + \frac{1}{|\pi|^{\alpha d/2p} N^{1/2p}} + |\pi|^{\frac{\alpha}{p}+\frac{1}{2p}-\frac{1}{2}} \left\| (\hat{\hat{R}}_{i+1,n-1} - R_{i+1,n-1})\,\Phi(X^\pi_{t_{i+1}}) \right\|_p \right),\end{aligned}$$

which provides, by performing a Trotter product expansion to the last term,

$$\varepsilon_i \le C\left(|\pi|^{\frac{\alpha+1/2}{p}} + \frac{1}{|\pi|^{\alpha d/2p} N^{1/2p}} + |\pi|^{\frac{\alpha}{p}+\frac{1}{2p}-\frac{1}{2}} \sum_{j=i+1}^{n-1} \varepsilon_j \right),$$

and the required estimate follows from the discrete version of Gronwall's inequality. □

Remark 4.6. Recently, Malliavin and Thalmaier [21] introduced an integration by parts formula to replace the one stated in Theorem 3.1 which is more advantageous when simulating high dimensional probability density functions. In particular, the Malliavin and Thalmaier integration by parts formula induces the following representation for $F^\pi_i := \rho(X^\pi_{t_{i+1}})\Delta W^l_{i+1}$

$$\mathbb{E}\left[F^\pi | X^\pi_{t_i} = x\right] = \frac{q_i^{F^\pi}(x)}{q_i^1(x)}, \tag{4.21}$$

where

$$\begin{aligned} q_i^{F^\pi}(x) &= \mathbb{E}\left[\sum_{j=1}^d \Lambda_d^i \left(X^\pi_{t_i} - x\right) H_i(X^\pi_{t_i}, F^\pi) \right], \\ q_i^{F^\pi}(x) &= \mathbb{E}\left[\sum_{j=1}^d \partial_j \Lambda_d \left(X^\pi_{t_i} - x\right) H_i(X^\pi_{t_i}, 1) \right] \end{aligned}$$

and $\Lambda_d = \{\Lambda_d^i\}_{i=1}^d : \mathbb{R}^d \to \mathbb{R}^d$ is the Riesz kernel

$$\Lambda_d(y) = \begin{cases} \dfrac{1}{a_2}\dfrac{y}{|y|^2} & \text{if } d = 2 \\ \dfrac{1}{a_d\,(d-2)}\dfrac{y}{|y|^d} & \text{if } d \ge 3 \end{cases}, \quad |y|^2 = \sum_{i=1}^d y_i^2.$$

Here a_d is the area of the unit sphere in $\mathbb{R}^d$ and $H_i(X^\pi_{t_i}, F^\pi)$ and $H_i(X^\pi_{t_i}, 1)$ are random variables that can be expressed as *single* Skorohod integrals instead of the previous multiple Skorohod integrals (see [21,18] for details and [2] for general conditions under which the representation holds true). Unfortunately, due to the singularity at zero of the Riesz kernel Λ_d, the variance of the estimator is infinite. In turn, this implies that no rates of convergence are available should we use a Monte Carlo approximation based on the representation (4.21). Finding a suitable way to deal with the instability of the approximations of Riesz integrals remains a hard problem (see [12] for another application of Riesz integral in a different context). A subsequent modification is required to make the formula applicable. Recently Kohatsu-Higa and Yasuda [18] propose the following approximation to the Malliavin–Thalmaier formula, where the (unbounded) Riesz

Please cite this article in press as: D. Crisan, et al., On the Monte Carlo simulation of BSDEs: An improvement on the Malliavin weights, Stochastic Processes and their Applications (2010), doi:10.1016/j.spa.2010.03.015

kernel is replaced by the bounded modification

$$\Lambda_d^h(y) = \begin{cases} \dfrac{1}{a_2}\dfrac{y}{|y|_h^2} & \text{if } d = 2 \\ \dfrac{1}{a_d\,(d-2)}\dfrac{y}{|y|_h^d} & \text{if } d \geq 3 \end{cases}, \quad |y|_h = \sqrt{|y|^2 + h}.$$

The variance of the truncated Riesz integral is bounded, hence one can approximate it via Monte Carlo with explicit rates. However, even though the L_1-norm of the bias introduced in this manner is controllable in terms of the truncation parameter p, the L_p-norm of the error is infinite for $p > 1$ rendering the method unusable in the BSDE context.

5. Numerical results

In this section we compare numerically the original version of Malliavin integration-by-parts algorithm [6] with the new version on a nonlinear model for which the computation of the Malliavin weights is nontrivial. The implementation of the algorithm has benefited greatly from many fruitful comments by Bruno Bouchard and Xavier Warin. In particular we have benefited from access to the work of [7], where Bouchard and Warin perform a comparison analysis of existing methods for the numerical approximation to BSDEs (quantization/Malliavin calculus method/regression on function bases method).

For our purposes we choose a model so that it has an explicit solution to enable us to quantify the convergence error. In particular, we assume that the forward component satisfies the following two-dimensional equation:

$$\mathrm{d}X_t^i = \mu_i X_t^i \mathrm{d}t + \sqrt{1 + (X_t^i)^2}\,\mathrm{d}W_t^i, \quad i = 1, 2, \tag{5.1}$$

and that the backward component satisfies the equation

$$Y_t = \sum_{i=1}^{2} \arctan\left(X_1^i\right) + \int_t^1 f(s, X_s, Y_s, Z_s)\mathrm{d}s - \int_t^1 Z_s \mathrm{d}W_s \quad t \in [0, 1], \tag{5.2}$$

where the driver $f : [0, 1] \times \mathbb{R}^2 \times \mathbb{R} \times \mathbb{R}^2 \to \mathbb{R}$ is given by

$$f(s, x_1, x_2, y, z_1, z_2) = \mathrm{e}^{r(1-s)} \sum_{i=1}^{2} (1 - \mu_i) x_i z_i^2 - ry. \tag{5.3}$$

An immediate application of Itô's formula gives us the explicit solution of (5.2): The process Y is given by

$$Y_t = \mathrm{e}^{-r(1-t)} \left(\arctan\left(X_t^1\right) + \arctan\left(X_t^2\right)\right), \quad t \in [0, 1],$$

and the process Z is given by

$$Z_t^i = \mathrm{e}^{-r(1-t)} \frac{1}{\sqrt{1 + (X_t^i)^2}}, \quad i = 1, 2,\ t \in [0, 1]. \tag{5.4}$$

We note that the driver (5.2) is only locally Lipschitz continuous. However, in view of (5.4), we know that the process Z is in fact bounded. Hence, we can consider the driver

Please cite this article in press as: D. Crisan, et al., On the Monte Carlo simulation of BSDEs: An improvement on the Malliavin weights, Stochastic Processes and their Applications (2010), doi:10.1016/j.spa.2010.03.015

$$\bar{f}(s, x_1, x_2, y, z_1, z_2) = \mathrm{e}^{r(1-s)} \sum_{i=1}^{2} (1-\mu_i) x_i \left(z_i \wedge \frac{2}{\sqrt{1+x_i^2}} \right)^2 - ry$$

in place of f, as the BSDE with driver $\bar{f}$ and the one with driver f have the same solution a.s.

The reduction in the required terms for the Malliavin weights reduces the computational time of the step where the Skorohod integrals are computed. However, the operation that dominates the time performance of the algorithm is the handling of the Heaviside function, which is an operation of complexity $O(N \log^{d-1} N)$ per step, in dimension *d and not of* $O(dN^2)$. The latter reduction in complexity is achieved by appealing to advanced sorting algorithms in d dimensions. A very detailed discussion on how one should handle the presence of the Heaviside function (with an explicit algorithm) can be found in [7], which we follow exactly.

On the other hand, all other involved operations can be reduced to complexity $O(N)$ *in the case where one is using a separable localizing function*, for example $\phi(x) = \mathrm{e}^{-x}$. To explain this, observe that, given that we have dealt with the Heaviside function by means of a sorting algorithm, the weights are implemented by means of the formulae (3.2) (*resp.* (4.5)). To estimate Y_{t_i}, we need to MC estimate the integrals $S^{h_i}[\phi(X_{t_i}^{\pi} - x)\Delta W_{i+1}^{l}]$ for $l = 0, 1, 2$ and some localizing function $\phi(\cdot)$ (and similarly for the new algorithm with $\bar{h}_i$ in place of h_i). Observe now, that according to (3.2) (*resp.* (4.5)), at time t_i the weight (e.g. $S^{h_i}[\cdot]$) in dimension 2 will read as (for $k \in \mathcal{N}_{i+1}$)

$$\begin{aligned} & S^{h_i^{(k)}} \left[\phi^F \left(X_{t_i}^{\pi^{(k)}} - x \right) \Delta W_{i+1}^{l^{(k)}} \right] \\ & = \phi^F \left(X_{t_i}^{\pi^{(k)}} - x \right) S_{1,2}^{h_i^{(k)}} \left[\Delta W_{i+1}^{l^{(k)}} \right] - \partial_{x_1} \phi^F \left(X_{t_i}^{\pi^{(k)}} - x \right) S_2^{h_i^{(k)}} \left[\Delta W_{i+1}^{l^{(k)}} \right] \\ & \quad - \partial_{x_2} \phi^F \left(X_{t_i}^{\pi^{(k)}} - x \right) S_1^{h_i^{(k)}} \left[\Delta W_{i+1}^{l^{(k)}} \right] + \partial_{x_1\, x_2}^2 \phi^F \left(X_{t_i}^{\pi^{(k)}} - x \right) \Delta W_{i+1}^{l^{(k)}}. \end{aligned} \tag{5.5}$$

Eq. (5.5) will be called with argument x ranging among the $X_{t_i}^{\pi^{(l)}}$, $l \in \mathcal{N}_i$. To avoid calling the exponential function $O(N \log^{d-1} N)$ times, observe that[1] for $k \in \mathcal{N}_{i+1}$, $l \in \mathcal{N}_i$

$$\exp\left(-(X_{t_i}^{\pi^{(k)}} - X_{t_i}^{\pi^{(l)}}) \right) = \exp\left(-(X_{t_i}^{\pi^{(k)}} - X_{t_i}^{\pi^{(l=1)}}) \right) \exp\left(-(X_{t_i}^{\pi^{(l=1)}} - X_{t_i}^{\pi^{(l)}}) \right).$$

Hence, to estimate Y_{t_i}, we take the following steps:

1. *Compute and store the weights*:
$$S_{1,2}^{h_i^{(k)}} \left[\Delta W_{i+1}^{j^{(k)}} \right], \qquad S_1^{h_i^{(k)}} \left[\Delta W_{i+1}^{j^{(k)}} \right] S_2^{h_i^{(k)}} \left[\Delta W_{i+1}^{j^{(k)}} \right], \quad k \in \mathcal{N}_{i+1},\ j = 0, 1, 2.$$
2. *Compute and store the two localizing function vectors*:
$$\exp\left(-(X_{t_i}^{\pi^{(k)}} - X_{t_i}^{\pi^{(l=1)}}) \right), \qquad \exp\left(-(X_{t_i}^{\pi^{(l=1)}} - X_{t_i}^{\pi^{(l)}}) \right), \quad k \in \mathcal{N}_{i+1}, l \in \mathcal{N}_i.$$
3. *Compute the sums involved in the definition* (3.5) *and* (3.6), *by means of the sorting algorithm of* [7]. *When required, call* (5.5) *using the stored vectors of steps* 1, 2.

The above procedure will then be dominated by step 3, and hence the overall method is of complexity $O(N \log^{d-1} N)$.

[1] This was mentioned to us by Bruno Bouchard.

Please cite this article in press as: D. Crisan, et al., On the Monte Carlo simulation of BSDEs: An improvement on the Malliavin weights, Stochastic Processes and their Applications (2010), doi:10.1016/j.spa.2010.03.015

Table 1
$n = 20$ steps.

N	8000	10 000	12 000	14 000
MIP	0.081 (0.008)	0.068 (0.006)	0.0521 (0.006)	0.041 (0.008)
MIPnew	0.073 (0.01)	0.071 (0.006)	0.046 (0.006)	0.038 (0.008)

Table 2
$n = 30$ steps.

N	12 000	16 000	20 000	24 000
MIP	0.12 (0.011)	0.091 (0.008)	0.06 (0.005)	0.035 (0.005)
MIPnew	0.124 (0.014)	0.07 (0.006)	0.068 (0.003)	0.04 (0.004)

We test our algorithm with parameters

T	μ_1	μ_2	r	X_0^1	X_0^2
0.5	0.06	0.08	0.03	3	4

For the new weights, we use for α the value 1. Also, we choose to work with $\phi(x) = \mathrm{e}^{-x}$ for localizing function. To summarize the results we call the implementation of the original algorithm of [6] MIP (Malliavin Integration-by-Parts), whereas the implementation with the new weights is called MIPnew. We run every algorithm 10 times and report on the absolute value of the error and (in parenthesis) the standard deviation in Table 1, when the number of steps $n = 20$ and in Table 2 when $n = 30$.

We see from the results that no accuracy has been lost by the transition from the old to the new weights. We also see the two opposite forces at work here. Namely, on the one hand, with more points in the partition, one achieves a finer discretization of the backward SDE, but that has a negative effect on the MC simulation of the involved expectations, as predicted by Lemma 3.4 and Theorem 3.5. For example, one needs 12 000 paths, when 30 discretization points are used (see the second column of Table 2) to achieve similar accuracy with 20 points on the partition and 8000 paths (see the first column of Table 1). Moreover, though we expected from Theorem 4.5, that the new weights would require a greater number of paths to achieve a similar accuracy with the original algorithm, in practice, at least for this example, we see that the two algorithms perform similarly with the same N. This is most likely due to the presence of well behaved coefficients in the forward backward system.

Acknowledgements

The authors would like to thank the referees for their extensive comments which certainly contributed to the many improvements of the earlier versions of the paper.

The first author's research was partially supported by EPSRC grant no. EP/H000550/1. The second author's research was partially supported by EPSRC and the Greek state scholarship's foundation. The third author's research was supported by the Chair *Financial Risks* of the *Risk Foundation* sponsored by Société Générale, the Chair *Derivatives of the Future* sponsored by the Fédération Bancaire Française, and the Chair *Finance and Sustainable Development* sponsored by EDF and Calyon.

Please cite this article in press as: D. Crisan, et al., On the Monte Carlo simulation of BSDEs: An improvement on the Malliavin weights, Stochastic Processes and their Applications (2010), doi:10.1016/j.spa.2010.03.015

Appendix

Proof of (2.5) and (2.6). Let us fix a value for i. Starting from the definition of R_i we have that

$$\begin{aligned} R_i g_1(x) - R_i g_2(x) &= \mathbb{E}_{t_i,x}\left[(g_1-g_2)(X_{t_{i+1}})\right] \\ &\quad + \Delta_{i+1}\left\{ f\left(t_i, x, R_i g_1(x), \frac{1}{\Delta_{i+1}}\mathbb{E}_{t_i,x}\left[g_1(X_{t_{i+1}})\Delta W_{i+1}\right]\right)\right. \\ &\quad \left. - f\left(t_i, x, R_i g_2(x), \frac{1}{\Delta_{i+1}}\mathbb{E}_{t_i,x}\left[g_2(X_{t_{i+1}})\Delta W_{i+1}\right]\right)\right\}. \end{aligned}$$

By the K-Lipschitz property of f, we can find deterministic functions $\nu(x) : \mathbb{R}^d \to \mathbb{R}$, $\zeta(x) : \mathbb{R}^d \to \mathbb{R}^d$, uniformly bounded by K, such that

$$\begin{aligned} &\Delta_{i+1}\left(f\left(t_i, x, R_i g_1(x), \frac{1}{\Delta_{i+1}}\mathbb{E}_{t_i,x}\left[g_1(X_{t_{i+1}})\Delta W_{i+1}\right]\right)\right. \\ &\quad \left. - f\left(t_i, x, R_i g_2(x), \frac{1}{\Delta_{i+1}}\mathbb{E}_{t_i,x}\left[g_2(X_{t_{i+1}})\Delta W_{i+1}\right]\right)\right) \\ &\quad = \Delta_{i+1}\nu(x)\left(R_i g_1(x) - R_i g_2(x)\right) + \zeta(x)\cdot\mathbb{E}_{t_i,x}\left[(g_1-g_2)(X_{t_{i+1}})\Delta W_{i+1}\right], \end{aligned}$$

and we have that

$$\begin{aligned} &(1-\Delta_{i+1}\nu(x))(R_i g_1(x) - R_i g_2(x)) \\ &\quad = \mathbb{E}_{t_i,x}\left[(g_1-g_2)(X_{t_{i+1}})\right] + \zeta(x)\cdot\mathbb{E}_{t_i,x}\left[\left(g_1(X_{t_{i+1}}) - g_2(X_{t_{i+1}})\right)\Delta W_{i+1}\right]. \end{aligned} \tag{A.1}$$

From (A.1) we deduce that

$$\begin{aligned} (1-\Delta_{i+1}K)|(R_i g_1(x) - R_i g_2(x))| &\le \left|\mathbb{E}_{t_i,x}\left[(g_1-g_2)(X_{t_{i+1}})\left(1+\zeta(x)\cdot\Delta W_{i+1}\right)\right]\right| \\ &\le \|g_1-g_2\|_{\mathbb{L}^p(\mathbb{P}^x_{t_i,t_{i+1}})}\mathbb{E}\left[|1+\zeta(x)\cdot\Delta W_{i+1}|^q\right]^{\frac{1}{q}} \\ &\le \|g_1-g_2\|_{\mathbb{L}^p(\mathbb{P}^x_{t_i,t_{i+1}})}\mathbb{E}\left[\left(1+\zeta(x)\cdot\Delta W_{i+1}\right)^{2k}\right]^{\frac{1}{2k}}, \end{aligned}$$

where p, q are conjugate and k an integer with $k > q/2$. Since $\zeta(x)$ is deterministic and bounded by the Lipschitz constant of f, K, we have that

$$\begin{aligned} \mathbb{E}\left[\left(1+\zeta(x)\cdot\Delta W_{i+1}\right)^{2k}\right] &= \sum_{j=0}^{2k}\binom{2k}{j}\mathbb{E}\left[\left(\zeta(x)\cdot\Delta W_{i+1}\right)^j\right] \\ &= \sum_{j=0}^{2k}\binom{2k}{j}\mathbb{E}\left[\left(\zeta(x)\cdot\Delta W_{i+1}\right)^{2j}\right] \\ &\le 1 + C\Delta_{i+1}, \end{aligned}$$

which completes the proof of (2.5).

As for the second claim let us fix an integer value $k = 0, \ldots, n-2$. We apply (2.5) to $|R_k R_{k+1} g_1 - R_k R_{k+1} g_2|(x)$ and to $|R_{k+1} g_1(x) - R_{k+1} g_2(x)|$ to get

$$|R_k R_{k+1} g_1(x) - R_k R_{k+1} g_2(x)| \le \frac{1+C\Delta_{k+1}}{1-K\Delta_{k+1}}\|R_{k+1} g_1 - R_{k+1} g_2\|_{\mathbb{L}^p\left(\mathbb{P}^x_{t_k,t_{k+1}}\right)},$$

Please cite this article in press as: D. Crisan, et al., On the Monte Carlo simulation of BSDEs: An improvement on the Malliavin weights, Stochastic Processes and their Applications (2010), doi:10.1016/j.spa.2010.03.015

and

$$|R_{k+1}g_1(x) - R_{k+1}g_2(x)| \leq \frac{1 + C\Delta_{k+2}}{1 - K\Delta_{k+2}} \|g_1 - g_2\|_{\mathbb{L}^p\left(\mathbb{P}^x_{t_{k+1},t_{k+2}}\right)}.$$

Then

$$\begin{aligned}&|R_k R_{k+1}g_1(x) - R_k R_{k+1}g_2(x)|\\ &\quad\leq \left(\frac{1 + C\Delta_{k+1}}{1 - K\Delta_{k+1}}\right)\left(\frac{1 + C\Delta_{k+2}}{1 - K\Delta_{k+2}}\right)\|g_1 - g_2\|_{\mathbb{L}^p\left(\mathbb{P}^x_{t_k,t_{k+2}}\right)},\end{aligned}$$

by the semigroup property. The generalization to an arbitrary number of iterations is now straightforward. □

Proof of Proposition 3.3. Set

$$\tilde{r}_i(x) := \frac{\hat{Q}^F_i[h_i, \phi](x)}{\hat{Q}^1_i[h_i, \psi](x)} \quad \text{so that } \hat{r}_i(x) = T^\zeta_i\left(\tilde{r}_i(x)\right).$$

Using the bounds on r_i (see the definition of $\gamma(x)$ in Proposition 3.3), we have

$$\begin{aligned}&\mathbb{E}^{0,i+1}\left[(r_i(X^\pi_{t_i}) - \hat{r}_i(X^\pi_{t_i}))^p\right] \leq \mathbb{E}^{0,i+1}\left[|\tilde{r}_i(X^\pi_{t_i}) - r_i(X^\pi_{t_i})|^p \wedge \gamma(X^\pi_{t_i})^p\right]\\ &\quad= \mathbb{E}^{0,i+1}\left[\left|\frac{\varepsilon^F(X^\pi_{t_i}) - r_i(X^\pi_{t_i})\varepsilon^1(X^\pi_{t_i})}{\hat{Q}^1(X^\pi_{t_i})}\right|^p \wedge \gamma(X^\pi_{t_i})^p\right],\end{aligned} \tag{A.2}$$

where

$$\varepsilon^F(x) := \hat{Q}^F[h_i, \phi](x) - q^F_i(x), \quad \text{and} \quad \varepsilon^1(x) := \hat{Q}^1[h_i, \psi](x) - q^1_i(x).$$

We will be using later the fact that

$$\mathbb{E}^{i+1}|\varepsilon^F(x)| \leq \|\varepsilon^F(x)\|_{L^2(\mathbb{P}^{i+1})} \leq \frac{1}{N^{1/2}} V_F(x)^{1/2},$$

where $V_F(x) = \mathrm{Var}[Q^F[h_i, \phi](x)]$. Notice that the last inequality holds in the marginal measure $\mathbb{P}^{i+1}$. A similar result is true for ε^1.

Next, for $x \in \mathbb{R}^d$ let us consider the event

$$\mathcal{M}(x) = \left\{\omega : |\hat{Q}^1(x, \omega) - q^1_i(x)| \leq \frac{1}{2} q^1_i(x)\right\}.$$

Using this set we may split the expectation above

$$\begin{aligned}&\mathbb{E}^{0,i+1}\left[\left|\frac{\varepsilon^F(X^\pi_{t_i}) - r_i(X^\pi_{t_i})\varepsilon^1(X^\pi_{t_i})}{\hat{Q}^1(X^\pi_{t_i})}\right|^p \wedge \gamma(X^\pi_{t_i})^p\right]\\ &\quad\leq \mathbb{E}^{0,i+1}\left[2\left|\frac{\varepsilon^F(X^\pi_{t_i}) - r_i(X^\pi_{t_i})\varepsilon^1(X^\pi_{t_i})}{q^1_i(X^\pi_{t_i})}\right|\gamma(X^\pi_{t_i})^{p-1}1_{\mathcal{M}(X^\pi_{t_i})}\right]\\ &\qquad+\mathbb{E}^{0,i+1}\left[\gamma(X^\pi_{t_i})^p \quad 1_{\mathcal{M}^c(X^\pi_{t_i})}\right],\end{aligned}$$

where we have used the inequality $a^p \wedge b^p \leq ab^{p-1}$. For the first term on the right hand side we compute

$$2\int_{\mathbb{R}^d} \mathbb{E}^{i+1}\left[|\varepsilon^F(x) - r_i(x)\varepsilon^1(x)|\gamma(x)\right]\mathrm{d}x$$

Please cite this article in press as: D. Crisan, et al., On the Monte Carlo simulation of BSDEs: An improvement on the Malliavin weights, Stochastic Processes and their Applications (2010), doi:10.1016/j.spa.2010.03.015

$$\leq 2\int_{\mathbb{R}^d}(\|\varepsilon^F(x)\|_{L^2(\mathbb{P}^{i+1})}+|r_i(x)|\|\varepsilon^1(x)\|_{L^2(\mathbb{P}^{i+1})})\gamma(x)^{p-1}\mathrm{d}x$$
$$=\frac{2}{N^{1/2}}\int_{\mathbb{R}^d}(V_F(x)^{1/2}+|r_i(x)|V_1(x)^{1/2})\gamma(x)^{p-1}\mathrm{d}x$$
$$\leq \frac{2}{N^{1/2}}\int_{\mathbb{R}^d}(\|Q^F\|_{L^2(\mathbb{P}^0)}+|r_i(x)|\|Q^1\|_{L^2(\mathbb{P}^0)})\gamma(x)^{p-1}\mathrm{d}x.$$

As for the second term we estimate it by means of the Chebychev inequality

$$\begin{aligned}\mathbb{E}^{0,i+1}\left[\gamma(X^\pi_{t_i})^p \quad 1_{\mathcal{M}^c(X^\pi_{t_i})}\right] &= \mathbb{E}^0\mathbb{E}^{i+1}\left[\gamma(X^\pi_{t_i})^p \quad 1_{\mathcal{M}^c(X^\pi_{t_i})}\right]\\ &= \mathbb{E}^0\left[\gamma(X^\pi_{t_i})^p\mathbb{P}^{i+1}[\mathcal{M}^c(X^\pi_{t_i})]\right]\\ &\leq \mathbb{E}^0\left[\gamma(X^\pi_{t_i})^p\frac{1}{q_i^1(X^\pi_{t_i})}\mathbb{E}^{i+1}[2|\hat{Q}^1(X^\pi_{t_i})-q_i^1(X^\pi_{t_i})|]\right]\\ &= \mathbb{E}^{i+1}\left[\int_{\mathbb{R}^d}2|\hat{Q}^1(x)-q_i^1(x)|\gamma(x)^p\,\mathrm{d}x\right]\\ &\leq \frac{1}{N^{1/2}}\int_{\mathbb{R}^d}\gamma(x)^pV_{Q^1}(x)^{1/2}\mathrm{d}x.\end{aligned}$$

We now have an estimate of the error with respect to the measure $\mathrm{d}\mathbb{P}^0\times\mathrm{d}\mathbb{P}^{i+1}$. The result now follows from an application of Fubini's theorem. □

References

[1] F. Antonelli, A. Kohatsu-Higa, Filtration stability of backward SDEs, Stochastic Analysis and its Applications 18 (2000) 11–37.
[2] V. Bally, L. Caramellino, A. Zanette, Pricing and hedging American options by Monte Carlo methods using a Malliavin calculus approach, Monte Carlo Methods and Applications 11 (2) (2005) 97–133.
[3] Vlad Bally, Gilles Pagès, Error analysis of the quantization algorithm for obstacle problems, Stochastic Processes and their Applications (2003) 1–40.
[4] B. Bouchard, I. Ekeland, N. Touzi, On the Malliavin approach to Monte Carlo methods of conditional expectations, Finance and Stochastics 8 (2004) 45–71.
[5] B. Bouchard, R. Elie, Discrete time approximation of decoupled forward–backward SDE with jumps, Stochastic Processes and their Applications 118 (1) (2008) 53–75.
[6] B. Bouchard, N. Touzi, Discrete time approximation and Monte Carlo simulation for backward stochastic differential equations, Stochastic Processes and their Applications 111 (2004) 175–206.
[7] B. Bouchard, X. Warin, Valuation of american options—new algorithm to improve on existing methods, Working Paper, 2010.
[8] P. Briand, B. Delyon, J. Mémin, Donsker-type theorem for BSDEs, Electronic Communications in Probability 6 (2001) 1–14.
[9] D. Chevance, Numerical methods for backward stochastic differential equations, in: L.C.G. Rogers, D. Talay (Eds.), Numerical Methods in Finance, Cambridge University Press, 1997, pp. 232–244.
[10] F. Coquet, V. Mackevicius, J. Mémin, Stability in d of martingales and backward equations under discretization of filtration, Stochastic Processes and their Applications 75 (1998) 235–248.
[11] D. Crisan, K. Manolarakis, Numerical solution for a BSDE using the cubature method, Preprint, 2007. Available at: http://www2.imperial.ac.uk/dcrisan/.
[12] D. Crisan, J. Xiong, Approximate Mckean–Vlasov representations for a class of SPDEs, Stochastics 82 (2010) 1744–2508.
[13] N. El Karoui, S. Peng, M. Quenez, Backward stochastic differential equations in finance, Mathematical Finance 7 (1) (1997) 1–71.

Please cite this article in press as: D. Crisan, et al., On the Monte Carlo simulation of BSDEs: An improvement on the Malliavin weights, Stochastic Processes and their Applications (2010), doi:10.1016/j.spa.2010.03.015

[14] E. Gobet, C. Labart, Error expansion for the discretization of backward stochastic differential equations, Stochastic Processes and their Applications 117 (7) (2007) 803–829.
[15] E. Gobet, J.P Lemor, X. Warin, A regression based Monte Carlo method to solve backward stochastic differential equations, Annals of Applied Probability 15 (3) (2005) 2172–2202.
[16] E. Gobet, J.-P. Lemor, X. Warin, Rate of convergence of an empirical regression method for solving generalized backward stochastic differential equations, Bernoulli 12 (5) (2006) 889–916.
[17] P. Kloeden, E. Platen, Numerical Solutions of Stochastic Differential Equations, Springer, 1999.
[18] A. Kohatsu-Higa, K. Yasuda, Estimating multidimensional density functions for random variables in Wiener space, Comptes Rendus Mathématique. Académie des Sciences 346 (5–6) (2008) 335–338.
[19] T. Lyons, N. Victoir, Cubature on Wiener space, Proceedings of the Royal Society of London 468 (2004) 169–198.
[20] J. Ma, P. Protter, J. San Martin, S. Torres, Numerical methods for backward stochastic differential equations, Annals of Applied Probability 12 (1) (2002) 302–316.
[21] Paul Malliavin, Anton Thalmaier, Stochastic Calculus of Variations in Mathematical Finance, in: Springer Finance, Springer-Verlag, Berlin, 2006.
[22] D. Nualart, The Malliavin Calculus and Related Topics, Springer-Verlag, 1996.
[23] Jianfeng Zhang, A numerical scheme for BSDEs, Annals of Applied Probability 14 (1) (2004) 459–488.

Please cite this article in press as: D. Crisan, et al., On the Monte Carlo simulation of BSDEs: An improvement on the Malliavin weights, Stochastic Processes and their Applications (2010), doi:10.1016/j.spa.2010.03.015