

A forward scheme for backward SDEs

Christian Bender^{a,*}, Robert Denk^b

^a *Institute for Mathematical Stochastics, Braunschweig University of Technology, Pockelsstr. 14,
D-38107 Braunschweig, Germany*

^b *Department for Mathematics and Statistics, University of Konstanz, D-78457 Konstanz, Germany*

Received 23 January 2006; received in revised form 6 October 2006; accepted 19 March 2007

Available online 30 March 2007

Abstract

We introduce a forward scheme for simulating backward SDEs. Compared to existing schemes, ours avoids high order nestings of conditional expectations backwards in time. In this way the error, when approximating the conditional expectation, depending on the time partition, is significantly reduced. Besides this generic result, we present an implementable algorithm and prove its convergence. Finally, we demonstrate the strength of the new algorithm by solving a financial problem numerically.

© 2007 Elsevier B.V. All rights reserved.

MSC: 65C05; 65C30; 91B28

Keywords: BSDE; Numerics; Monte Carlo simulation; Finance

1. Introduction

The study of nonlinear backward stochastic differential equations (BSDEs) was initiated by Pardoux and Peng [14]. Mainly motivated by financial problems (see e.g. the survey article by El Karoui et al. [8]), the theory of BSDEs was developed at high speed during the 1990s. Comparably slow progress has been made on the numerics of BSDEs.

Up to now basically two types of schemes have been considered. Based on the theoretical four-step scheme from [12], numerical algorithms for BSDEs have been developed by Douglas et al. [7] and more recently by Milstein and Tretyakov [13]. The main focus of these algorithms is the numerical solution of a parabolic PDE which is related to the BSDE.

* Corresponding author.

E-mail address: c.bender@tu-bs.de (C. Bender).

A second type of algorithms works backwards through time and tries to tackle the stochastic problem directly. Bally [1] and Chevance [6] were the first to study this type of algorithm with a (hardly implementable) random time partition and under strong regularity assumptions, respectively. The work of Ma et al. [11] is in the same spirit, replacing, however, the Brownian motion by a binary random walk in the approximative equation. See also [5] for the binary random walk approach. Only recently, a new notion of L^2 -regularity on the control part of the solution was introduced in [17], which allowed proof of convergence of this backward approach with deterministic partitions under rather weak regularity assumptions; see [17,4,9] for different algorithms.

A main drawback of the backward schemes is that nestings of conditional expectations backwards through the time steps must be evaluated. For a practical implementation the conditional expectations must be replaced by some estimator. A generic result of Bouchard and Touzi [4] shows that the error due to the approximation of the conditional expectation grows with order $1/2$, as the number of time steps goes to infinity. This leads to high computational costs when a fine mesh of the time discretization is required.

In this paper we propose a new forward scheme which avoids nestings of conditional expectations backwards through the time steps. Instead it mimics the Picard type iteration for BSDEs and, consequently, has nestings of conditional expectation along the iterations.

We show that the additional error due to the iteration converges to zero at a geometric rate (Theorem 5). At this cost the error, when approximating the conditional expectations by a generic estimator, depending on the time partition, is reduced by order $1/2$ compared to existing backward schemes (Theorem 10). In fact, in our scheme this error does not explode when the number of time steps tends to infinity or when the number of iterations tends to infinity. We believe that this is a striking advantage compared to the backward scheme.

Besides these generic results, we develop a practically implementable numerical scheme. In particular, we use the regression based least squares Monte Carlo method to approximate the conditional expectation as was suggested by Gobet et al. [9] in the context of the backward scheme. We analyze the error when replacing the conditional expectation by the orthogonal projections on subspaces (Theorem 11), and also prove convergence when the projection coefficients are substituted by their simulation based analogues (Theorem 15).

Finally, we present some simulations related to financial problems (Section 5). To be precise, we consider the hedging problem under different interest rates for investing and borrowing, which leads to nonlinear BSDEs.

2. A discretization of the Picard type iteration

In this section we introduce a discretized Picard iteration and prove its convergence for the following type of BSDE:

$$\begin{aligned} dX_t &= b(t, X_t)dt + \sigma(t, X_t)dW_t \\ dY_t &= f(t, X_t, Y_t, Z_t)dt + Z_t dW_t \\ X_0 &= x \\ Y_T &= \xi. \end{aligned}$$

Here $W_t = (W_{1,t}, \dots, W_{D,t})^*$ (the star denoting matrix transposition) is a D -dimensional Brownian motion on $[0, T]$ and $Z_t = (Z_{1,t}, \dots, Z_{D,t})$. The process X is \mathbb{R}^M -valued and the process Y is \mathbb{R} -valued. Throughout the paper we assume:

Assumption 1. There is a constant K such that

$$|b(t, x) - b(t', x')| + |\sigma(t, x) - \sigma(t', x')| + |f(t, x, y, z) - f(t', x', y', z')| \\ \leq K(\sqrt{|t - t'|} + |x - x'| + |y - y'| + |z - z'|)$$

for all $(t, x, y, z), (t', x', y', z') \in [0, T] \times \mathbb{R}^M \times \mathbb{R} \times \mathbb{R}^D$,

$$\xi = \Phi(X)$$

where Φ is a functional on the space of \mathbb{R}^M -valued RCLL functions on $[0, T]$ satisfying the L^∞ -Lipschitz condition

$$|\Phi(\mathbf{x}) - \Phi(\mathbf{x}')| \leq K \sup_{0 \leq t \leq T} |\mathbf{x}(t) - \mathbf{x}'(t)|$$

for all RCLL functions \mathbf{x}, \mathbf{x}' . Moreover,

$$\sup_{0 \leq t \leq T} (|b(t, 0)| + |\sigma(t, 0)| + |f(t, 0, 0, 0)|) + |\Phi(\mathbf{0})| \leq K$$

where $\mathbf{0}$ denotes the constant function taking value 0 on $[0, T]$.

Note that we do not assume that the matrix σ is quadratic or that $\sigma\sigma^*$ is invertible.

Remark 1. We shall say that a constant depends on the data if it depends on K, T, x_0 and the dimensions M and D only. Throughout the paper, C denotes a generic constant depending on the data which may vary from line to line.

Theoretically, the backward part (Y, Z) can be obtained as the limit of a Picard type iteration $(Y^{(n)}, Z^{(n)})$; see e.g. [15], Theorem 7.3.4. Here $(Y^{(0)}, Z^{(0)}) \equiv (0, 0)$, and $(Y^{(n)}, Z^{(n)})$ is the solution of the simple BSDE

$$dY_t^{(n)} = f(t, X_t, Y_t^{(n-1)}, Z_t^{(n-1)})dt + Z_t^{(n)}dW_t \\ Y_T^{(n)} = \xi$$

with X as above.

The solution is given by

$$Y_t^{(n)} = E \left[\xi - \int_t^T f(s, X_s, Y_s^{(n-1)}, Z_s^{(n-1)})ds \middle| \mathcal{F}_t \right]$$

and $Z^{(n)}$ is obtained via the martingale representation theorem. As is emphasized in [15], Chapter 7, the above Picard iteration is still implicit due to the use of the martingale representation theorem.

We will now introduce a time discretization of the above iteration which is explicit in time, but still requires the evaluation of conditional expectations. The error due to the approximation of the conditional expectations is investigated in Sections 3 and 4 below.

Suppose a partition $\pi = \{t_0, t_1, \dots, t_N\}$ of $[0, T]$ with mesh size $|\pi| := \max_i |t_{i+1} - t_i|$ is given and a corresponding discretization $X^{(\pi)}$ of X as well as some approximation $\xi^{(\pi)}$ of ξ . Let $(Y^{(0,\pi)}, Z^{(0,\pi)}) \equiv (0, 0)$. Then define iteratively for $i = 0, 1, \dots, N$, with $\Delta_i = t_{i+1} - t_i$ and $\Delta W_{d,i} = W_{d,t_{i+1}} - W_{d,t_i}$,

$$Y_{t_i}^{(n,\pi)} = E \left[\xi^{(\pi)} - \sum_{j=i}^{N-1} f(t_j, X_{t_j}^{(\pi)}, Y_{t_j}^{(n-1,\pi)}, Z_{t_j}^{(n-1,\pi)})\Delta_j \middle| \mathcal{F}_{t_i} \right],$$

$$Z_{d,t_i}^{(n,\pi)} = E \left[\frac{\Delta W_{d,i}}{\Delta_i} \left(\xi^{(\pi)} - \sum_{j=i+1}^{N-1} f(t_j, X_{t_j}^{(\pi)}, Y_{t_j}^{(n-1,\pi)}, Z_{t_j}^{(n-1,\pi)}) \Delta_j \right) \middle| \mathcal{F}_{t_i} \right],$$

$d = 1, \dots, D$. (Here we used the convention that $\Delta W_{d,N} = 0$.) The processes $Y^{(n,\pi)}$ and $Z^{(n,\pi)}$ are extended to RCLL processes by constant interpolation. Note that the discretized Picard type iteration has no nestings of conditional expectations backward in time, but does have ones forward in the number of Picard iterations. This turns out to be an advantage from the numerical point of view (see Section 3 below).

We can now state the convergence of the discretized Picard type iteration:

Theorem 2. Let Assumption 1 be satisfied and suppose that for some constant C depending on the data we have

$$\sup_{0 \leq t \leq T} E \left[|X_t - X_t^{(\pi)}|^2 \right] \leq C|\pi|, \quad \sup_{|\pi| \leq 1} E \left[|\xi^{(\pi)}|^2 \right] \leq C.$$

Then there is a constant C depending on the data such that

$$\begin{aligned} \sup_{0 \leq t \leq T} E \left[|Y_t - Y_t^{(n,\pi)}|^2 \right] + E \int_0^T |Z_t - Z_t^{(n,\pi)}|^2 dt \\ \leq C \left(|\pi| + E[|\xi - \xi^{(\pi)}|^2] + \left(\frac{1}{2} + C|\pi| \right)^n \right) \end{aligned}$$

provided $|\pi|$ is sufficiently small.

Remark 3. (i) Note that the condition on the discretization $X^{(\pi)}$ of X is, for instance, satisfied by the Euler scheme.

(ii) The condition on $\xi^{(\pi)}$ is satisfied whenever for $|\pi| \leq 1$

$$E[|\xi - \xi^{(\pi)}|^2] \leq C|\pi|^\alpha$$

with some constant C depending on the data and some $\alpha > 0$. Indeed,

$$E[|\xi^{(\pi)}|^2] \leq 2E[|\xi|^2] + 2E[|\xi - \xi^{(\pi)}|^2],$$

and, thanks to the L^∞ -Lipschitz condition and a classical estimate for SDEs,

$$\begin{aligned} E[|\xi|^2] &\leq 2K^2 E \left[\sup_{0 \leq t \leq T} |X_t|^2 \right] + 2|\Phi(\mathbf{0})|^2 \\ &\leq C \left(|x|^2 + \int_0^T |b(t, 0)|^2 + |\sigma(t, 0)|^2 dt \right) + 2K^2 \leq C. \end{aligned}$$

The proof of Theorem 2 is split into two parts. Given the partition π and a corresponding discretization $X^{(\pi)}$ of X we define $(Y^{(\infty,\pi)}, Z^{(\infty,\pi)})$ as the solution of

$$\begin{aligned} Y_{t_N}^{(\infty,\pi)} &= \xi^{(\pi)}, \\ Z_{d,t_i}^{(\infty,\pi)} &= E \left[\frac{\Delta W_{d,i}}{\Delta_i} Y_{t_{i+1}}^{(\infty,\pi)} \middle| \mathcal{F}_{t_i} \right], \\ Y_{t_i}^{(\infty,\pi)} &= E[Y_{t_{i+1}}^{(\infty,\pi)} | \mathcal{F}_{t_i}] - f(t_i, X_{t_i}^{(\pi)}, Y_{t_i}^{(\infty,\pi)}, Z_{t_i}^{(\infty,\pi)}) \Delta_i. \end{aligned}$$

This solution exists when the mesh $|\pi|$ of the partition π is sufficiently fine. Again, the processes $Y^{(\infty,\pi)}$ and $Z^{(\infty,\pi)}$ are extended to RCLL processes by constant interpolation. Note

that $(Y^{(\infty,\pi)}, Z^{(\infty,\pi)})$ is (up to the interpolation of the Z -part) the backward scheme considered in [4]. We remark that this backward scheme is still implicit, and inner iterations are required for numerical implementation.

We shall separately consider the convergence of $(Y^{(n,\pi)}, Z^{(n,\pi)})$ to $(Y^{(\infty,\pi)}, Z^{(\infty,\pi)})$ and that of $(Y^{(\infty,\pi)}, Z^{(\infty,\pi)})$ to (Y, Z) .

Concerning the backward scheme we need an extension of the results by Bouchard and Touzi [4]. The following variant of Theorem 3.1 in [4] is a slight generalization concerning the assumptions on the coefficients. Moreover, it allows path-dependent terminal data and the approximating processes are piecewise constant.

Theorem 4. *Let Assumption 1 be satisfied and suppose that the discretization $X^{(\pi)}$ of X satisfies*

$$\sup_{0 \leq t \leq T} E \left[|X_t - X_t^{(\pi)}|^2 \right] \leq C |\pi| \quad (1)$$

for some constant C depending on the data. Then there is a constant C depending on the data such that

$$\sup_{0 \leq t \leq T} E \left[|Y_t - Y_t^{(\infty,\pi)}|^2 \right] + E \int_0^T |Z_t - Z_t^{(\infty,\pi)}|^2 dt \leq C \left(|\pi| + E[|\xi - \xi^{(\pi)}|^2] \right)$$

provided $|\pi|$ is sufficiently small.

The proof combines ideas of Bouchard and Touzi [4] and Zhang [16,17], who suggests a slightly different time discretization. As the argumentation is fairly standard by now, the proof is omitted.

We now investigate the iteration for a fixed partition. Our aim is to derive rates of convergence uniform in π .

Theorem 5. *Under the assumptions of Theorem 2 there are constants C_1 and C_2 depending on the data such that*

$$\max_{0 \leq i \leq N} E \left[|Y_{t_i}^{(\infty,\pi)} - Y_{t_i}^{(n,\pi)}|^2 \right] + \sum_{i=0}^{N-1} E \left[|Z_{t_i}^{(\infty,\pi)} - Z_{t_i}^{(n,\pi)}|^2 \right] \Delta_i \leq C_1 \left(\frac{1}{2} + C_2 |\pi| \right)^n$$

provided $|\pi|$ is sufficiently small.

Clearly, Theorem 2 follows from a straightforward combination of Theorems 4 and 5.

Remark 6. Let K denote the Lipschitz constant of f . Then it follows from the proof below that Theorem 5 holds, for instance, for $|\pi| \leq \Gamma^{-1}$ with

$$C_2 = \frac{\Gamma}{4},$$

where

$$\Gamma = 16T(T+1)^2 D^2 K^4 + 4K(T+1)K^2.$$

We prepare the proof of Theorem 5 with some a priori estimates.

Lemma 7. Suppose Γ and γ are positive real numbers, $\tilde{y}^{(i)}, \tilde{z}^{(i)}, i = 1, 2$, are adapted processes and

$$\begin{aligned}\tilde{Y}_{t_i}^{(i)} &= E \left[\xi^{(\pi)} - \sum_{j=i}^{N-1} f(t_j, X_{t_j}^{(\pi)}, \tilde{y}_{t_j}^{(i)}, \tilde{z}_{t_j}^{(i)}) \Delta_j \middle| \mathcal{F}_{t_i} \right], \\ \tilde{Z}_{d,t_i}^{(i)} &= E \left[\frac{\Delta W_{d,i}}{\Delta_i} \left(\xi^{(\pi)} - \sum_{j=i+1}^{N-1} f(t_j, X_{t_j}^{(\pi)}, \tilde{y}_{t_j}^{(i)}, \tilde{z}_{t_j}^{(i)}) \Delta_j \right) \middle| \mathcal{F}_{t_i} \right].\end{aligned}$$

Moreover, assume that f is Lipschitz in (y, z) uniformly in (t, x) with constant K . Then we have

$$\begin{aligned}& \max_{0 \leq i \leq N} \lambda_i E \left[|\tilde{Y}_{t_i}^{(1)} - \tilde{Y}_{t_i}^{(2)}|^2 \right] + \sum_{i=0}^{N-1} \lambda_i E \left[|\tilde{Z}_{t_i}^{(1)} - \tilde{Z}_{t_i}^{(2)}|^2 \right] \Delta_i \\ & \leq K^2(T+1) \left((|\pi| + \Gamma^{-1}) (\gamma DT + 1) + \frac{D}{\gamma} \right) \\ & \quad \times \left(\frac{1}{T} \sum_{i=0}^{N-1} \lambda_i E \left[|\tilde{y}_{t_i}^{(1)} - \tilde{y}_{t_i}^{(2)}|^2 \right] \Delta_i + \sum_{i=0}^{N-1} \lambda_i E \left[|\tilde{z}_{t_i}^{(1)} - \tilde{z}_{t_i}^{(2)}|^2 \right] \Delta_i \right),\end{aligned}$$

where $\lambda_0 = 1$ and $\lambda_i = (1 + \Gamma \Delta_{i-1}) \lambda_{i-1}$.

This lemma will be applied several times throughout this paper. The choice of the constants Γ and γ may vary for applications in different theorems.

Proof. The proof goes through several steps. For notational convenience, let us introduce

$$\begin{aligned}y_{t_i} &= \tilde{y}_{t_i}^{(1)} - \tilde{y}_{t_i}^{(2)}, \\ z_{t_i} &= \tilde{z}_{t_i}^{(1)} - \tilde{z}_{t_i}^{(2)}, \\ \Delta f_i &= f(t_i, X_{t_i}^{(\pi)}, \tilde{y}_{t_i}^{(1)}, \tilde{z}_{t_i}^{(1)}) - f(t_i, X_{t_i}^{(\pi)}, \tilde{y}_{t_i}^{(2)}, \tilde{z}_{t_i}^{(2)}).\end{aligned}$$

First note that

$$\tilde{Y}_{t_i}^{(i)} = E[\tilde{Y}_{t_{i+1}}^{(i)} | \mathcal{F}_{t_i}] - f(t_i, X_{t_i}^{(\pi)}, \tilde{y}_{t_i}^{(i)}, \tilde{z}_{t_i}^{(i)}) \Delta_i \quad (2)$$

and, for the d th component of $\tilde{Z}^{(i)}$,

$$\tilde{Z}_{d,t_i}^{(i)} = E \left[\frac{\Delta W_{d,i}}{\Delta_i} \tilde{Y}_{t_{i+1}}^{(i)} \middle| \mathcal{F}_{t_i} \right]. \quad (3)$$

Step 1: We prove that for any $1 \leq d \leq D$

$$\begin{aligned}& \sum_{i=0}^{N-1} \lambda_i E \left[|\tilde{Z}_{d,t_i}^{(1)} - \tilde{Z}_{d,t_i}^{(2)}|^2 \right] \Delta_i \\ & \leq \gamma \sum_{i=0}^{N-1} \lambda_i E \left[|\tilde{Y}_{t_i}^{(1)} - \tilde{Y}_{t_i}^{(2)}|^2 \right] \Delta_i + \frac{(1+T)K^2}{\gamma} \sum_{i=0}^{N-1} \lambda_i E \left[|z_{t_i}|^2 \right] \Delta_i \\ & \quad + \frac{(1+T)K^2}{T\gamma} \sum_{i=0}^{N-1} \lambda_i E \left[|y_{t_i}|^2 \right] \Delta_i.\end{aligned} \quad (4)$$

First note that by (3) and Hölder's inequality,

$$\begin{aligned}\tilde{Z}_{d,t_i}^{(1)} - \tilde{Z}_{d,t_i}^{(2)} &= E \left[\frac{\Delta W_{d,i}}{\Delta_i} \left(\tilde{Y}_{t_{i+1}}^{(1)} - \tilde{Y}_{t_{i+1}}^{(2)} \right) \middle| \mathcal{F}_{t_i} \right] \\ &= E \left[\frac{\Delta W_{d,i}}{\Delta_i} \left(\tilde{Y}_{t_{i+1}}^{(1)} - \tilde{Y}_{t_{i+1}}^{(2)} - E[\tilde{Y}_{t_{i+1}}^{(1)} - \tilde{Y}_{t_{i+1}}^{(2)} | \mathcal{F}_{t_i}] \right) \middle| \mathcal{F}_{t_i} \right] \\ &\leq \sqrt{\frac{1}{\Delta_i}} E \left[\left(\tilde{Y}_{t_{i+1}}^{(1)} - \tilde{Y}_{t_{i+1}}^{(2)} - E[\tilde{Y}_{t_{i+1}}^{(1)} - \tilde{Y}_{t_{i+1}}^{(2)} | \mathcal{F}_{t_i}] \right)^2 \middle| \mathcal{F}_{t_i} \right]^{1/2}.\end{aligned}$$

Thus, by (2),

$$\begin{aligned}E \left[|\tilde{Z}_{d,t_i}^{(1)} - \tilde{Z}_{d,t_i}^{(2)}|^2 \right] &\leq \frac{1}{\Delta_i} E \left[|\tilde{Y}_{t_{i+1}}^{(1)} - \tilde{Y}_{t_{i+1}}^{(2)}|^2 - E[\tilde{Y}_{t_{i+1}}^{(1)} - \tilde{Y}_{t_{i+1}}^{(2)} | \mathcal{F}_{t_i}]^2 \right] \\ &= \frac{1}{\Delta_i} E \left[|\tilde{Y}_{t_{i+1}}^{(1)} - \tilde{Y}_{t_{i+1}}^{(2)}|^2 - |\tilde{Y}_{t_i}^{(1)} - \tilde{Y}_{t_i}^{(2)} + \Delta f_i \Delta_i|^2 \right] \\ &\leq \frac{1}{\Delta_i} E \left[|\tilde{Y}_{t_{i+1}}^{(1)} - \tilde{Y}_{t_{i+1}}^{(2)}|^2 - |\tilde{Y}_{t_i}^{(1)} - \tilde{Y}_{t_i}^{(2)}|^2 - 2(\tilde{Y}_{t_i}^{(1)} - \tilde{Y}_{t_i}^{(2)}) \Delta f_i \Delta_i \right].\end{aligned}$$

Multiplying both sides with the weights $\lambda_i \Delta_i$ and summing from 0 to $N - 1$ yields for $\gamma > 0$,

$$\begin{aligned}&\sum_{i=0}^{N-1} \lambda_i E \left[|\tilde{Z}_{t_i}^{(1)} - \tilde{Z}_{t_i}^{(2)}|^2 \right] \Delta_i + \lambda_0 E \left[|\tilde{Y}_{t_0}^{(1)} - \tilde{Y}_{t_0}^{(2)}|^2 \right] \\ &\leq \lambda_N E \left[|\tilde{Y}_{t_N}^{(1)} - \tilde{Y}_{t_N}^{(2)}|^2 \right] - 2 \sum_{i=0}^{N-1} \lambda_i E \left[(\tilde{Y}_{t_i}^{(1)} - \tilde{Y}_{t_i}^{(2)}) \Delta f_i \right] \Delta_i \\ &\leq \gamma \sum_{i=0}^{N-1} \lambda_i E \left[|\tilde{Y}_{t_i}^{(1)} - \tilde{Y}_{t_i}^{(2)}|^2 \right] \Delta_i + \frac{K^2}{\gamma} \sum_{i=0}^{N-1} \lambda_i E \left[(|y_{t_i}| + |z_{t_i}|)^2 \right] \Delta_i.\end{aligned}$$

Here we used $\tilde{Y}_{t_N}^{(1)} - \tilde{Y}_{t_N}^{(2)} = 0$ and Young's inequality. Inequality (4) can now be obtained by another application of Young's inequality.

Step 2: We show

$$\begin{aligned}&\max_{0 \leq i \leq N} \lambda_i E \left[|\tilde{Y}_{t_i}^{(1)} - \tilde{Y}_{t_i}^{(2)}|^2 \right] \\ &\leq K^2(T+1) \left(|\pi| + \frac{1}{\Gamma} \right) \left(\sum_{i=0}^{N-1} \lambda_i E \left[|z_{t_i}|^2 \Delta_i \right] + \frac{1}{T} \sum_{i=0}^{N-1} \lambda_i E \left[|y_{t_i}|^2 \Delta_i \right] \right).\end{aligned}\quad (5)$$

By (2), Jensen's inequality, and Young's inequality we get

$$\begin{aligned}E \left[|\tilde{Y}_{t_j}^{(1)} - \tilde{Y}_{t_j}^{(2)}|^2 \right] &\leq (1 + \Gamma \Delta_j) E \left[|\tilde{Y}_{t_{j+1}}^{(1)} - \tilde{Y}_{t_{j+1}}^{(2)}|^2 \right] + (\Delta_j + \Gamma^{-1}) E[(\Delta f_j)^2] \Delta_j \\ &\leq (1 + \Gamma \Delta_j) E \left[|\tilde{Y}_{t_{j+1}}^{(1)} - \tilde{Y}_{t_{j+1}}^{(2)}|^2 \right] \\ &\quad + \left(|\pi| + \Gamma^{-1} \right) K^2(T+1) E[|z_{t_j}|^2] \Delta_j \\ &\quad + \left(|\pi| + \Gamma^{-1} \right) K^2 \frac{T+1}{T} E[|y_{t_j}|^2] \Delta_j.\end{aligned}$$

Multiplying with λ_j and summing from $j = i$ to $N - 1$ easily yields (5), since $\tilde{Y}_{t_N}^{(1)} - \tilde{Y}_{t_N}^{(2)} = 0$.

Final step: The assertion follows from a straightforward combination of (4) and (5). \square

Proof (Theorem 5). Define

$$\begin{aligned} y_{t_i}^{(n+1,\pi)} &= Y_{t_i}^{(n+1,\pi)} - Y_{t_i}^{(n,\pi)}, \\ z_{t_i}^{(n+1,\pi)} &= Z_{t_i}^{(n+1,\pi)} - Z_{t_i}^{(n,\pi)}. \end{aligned}$$

By Lemma 7,

$$\begin{aligned} & \max_{0 \leq i \leq N} \lambda_i E \left[|y_{t_i}^{(n+1,\pi)}|^2 \right] + \sum_{i=0}^{N-1} \lambda_i E \left[|z_{t_i}^{(n+1,\pi)}|^2 \right] \Delta_i \\ & \leq K^2(T+1) \left((|\pi| + \Gamma^{-1}) (\gamma DT + 1) + \frac{D}{\gamma} \right) \\ & \quad \times \left(\max_{0 \leq i \leq N} \lambda_i E \left[|y_{t_i}^{(n,\pi)}|^2 \right] + \sum_{i=0}^{N-1} \lambda_i E \left[|z_{t_i}^{(n,\pi)}|^2 \right] \Delta_i \right). \end{aligned}$$

We now choose $\gamma = 4DK^2(T+1)$ and $\Gamma = 4K^2(T+1)(\gamma DT + 1)$ and iterate the above inequality to obtain

$$\begin{aligned} & \max_{0 \leq i \leq N} \lambda_i E \left[|y_{t_i}^{(n+1,\pi)}|^2 \right] + \sum_{i=0}^{N-1} \lambda_i E \left[|z_{t_i}^{(n+1,\pi)}|^2 \right] \Delta_i \\ & \leq \left(\frac{\Gamma|\pi|}{4} + \frac{1}{2} \right)^n \left(\max_{0 \leq i \leq N} \lambda_i E \left[|Y_{t_i}^{(1,\pi)}|^2 \right] + \sum_{i=0}^{N-1} \lambda_i E \left[|Z_{t_i}^{(1,\pi)}|^2 \right] \Delta_i \right). \end{aligned}$$

Recalling the definition of λ_i from Lemma 7 we have

$$\begin{aligned} & \max_{0 \leq i \leq N} E \left[|y_{t_i}^{(n+1,\pi)}|^2 \right] + \sum_{i=0}^{N-1} E \left[|z_{t_i}^{(n+1,\pi)}|^2 \right] \Delta_i \\ & \leq e^{\Gamma T} \left(\frac{\Gamma|\pi|}{4} + \frac{1}{2} \right)^n \left(\max_{0 \leq i \leq N} E \left[|Y_{t_i}^{(1,\pi)}|^2 \right] + \sum_{i=0}^{N-1} E \left[|Z_{t_i}^{(1,\pi)}|^2 \right] \Delta_i \right). \end{aligned}$$

Denote the square root of the right-hand side by $A(\pi, n)$. Clearly the series $\sum_n A(\pi, n)$ converges when $|\pi|$ is sufficiently small. This shows that $(Y^{(n,\pi)}, Z^{(n,\pi)})$ is a Cauchy sequence and thus converges to $(Y^{(\infty,\pi)}, Z^{(\infty,\pi)})$ (when $|\pi|$ is sufficiently small) by means of (2) and (3). Moreover, for $n \in \mathbb{N}$,

$$\begin{aligned} & \max_{0 \leq i \leq N} E \left[|Y_{t_i}^{(\infty,\pi)} - Y_{t_i}^{(n,\pi)}|^2 \right] + \sum_{i=0}^{N-1} E \left[|Z_{t_i}^{(\infty,\pi)} - Z_{t_i}^{(n,\pi)}|^2 \right] \Delta_i \\ & \leq \left(\sum_{v=n}^{\infty} A(\pi, v) \right)^2 \\ & \leq e^{\Gamma T} \left(\max_{0 \leq i \leq N} E \left[|Y_{t_i}^{(1,\pi)}|^2 \right] + \sum_{i=0}^{N-1} E \left[|Z_{t_i}^{(1,\pi)}|^2 \right] \Delta_i \right) \left(1 - \sqrt{\frac{\Gamma|\pi|}{4} + \frac{1}{2}} \right)^{-2} \\ & \quad \times \left(\frac{\Gamma|\pi|}{4} + \frac{1}{2} \right)^n. \end{aligned}$$

It remains to prove a uniform bound for

$$\left(\max_{0 \leq i \leq N} E \left[|Y_{t_i}^{(1,\pi)}|^2 \right] + \sum_{i=0}^{N-1} E \left[|Z_{t_i}^{(1,\pi)}|^2 \right] \Delta_i \right),$$

which is given in the following lemma. \square

Lemma 8. *Under the assumptions of Theorem 2, there is a constant C depending on the data only such that*

$$\max_{0 \leq i \leq N} E \left[|Y_{t_i}^{(1,\pi)}|^2 \right] + \sum_{i=0}^{N-1} E \left[|Z_{t_i}^{(1,\pi)}|^2 \right] \Delta_i \leq C$$

provided $|\pi| \leq 1$.

Proof. By Young's and Hölder's inequalities we have

$$\max_{0 \leq i \leq N} E \left[|Y_{t_i}^{(1,\pi)}|^2 \right] \leq 2E[|\xi^{(\pi)}|^2] + 2T \sum_{j=0}^{N-1} E \left[|f(t_j, X_{t_j}^{(\pi)}, 0, 0)|^2 \right] \Delta_j.$$

The first term on the right-hand side is bounded by a constant depending on the data for $|\pi| \leq 1$ by assumption. For the second term we observe that

$$\begin{aligned} E \left[|f(t_j, X_{t_j}^{(\pi)}, 0, 0)|^2 \right] &\leq 2E \left[|f(t_j, X_{t_j}^{(\pi)}, 0, 0) - f(t_j, 0, 0, 0)|^2 \right] + 2|f(t_j, 0, 0, 0)|^2 \\ &\leq 2K^2 \left(\sup_{0 \leq t \leq T} E[|X_t^{(\pi)}|^2] + 1 \right). \end{aligned}$$

Now, by assumption and a classical result on SDEs,

$$\begin{aligned} \sup_{0 \leq t \leq T} E[|X_t^{(\pi)}|^2] &\leq 2 \sup_{0 \leq t \leq T} E[|X_t^{(\pi)} - X_t|^2] + 2 \sup_{0 \leq t \leq T} E[|X_t|^2] \\ &\leq C|\pi| + C \left(|x|^2 + \int_0^T |b(t, 0)|^2 + |\sigma(t, 0)|^2 dt \right) \leq C(1 + |\pi|). \end{aligned}$$

We have thus shown that for $|\pi| \leq 1$,

$$\max_{0 \leq i \leq N} E \left[|Y_{t_i}^{(1,\pi)}|^2 \right] + \max_{0 \leq i \leq N} E \left[|f(t_j, X_{t_j}^{(\pi)}, 0, 0)|^2 \right] \Delta_i \leq C. \quad (6)$$

Analogously to step 1 in Lemma 7 we obtain

$$E \left[|Z_{d,t_i}^{(1,\pi)}|^2 \right] \leq \frac{1}{\Delta_i} E \left[|Y_{t_{i+1}}^{(1,\pi)}|^2 - |Y_{t_i}^{(1,\pi)}|^2 - 2Y_{t_i}^{(1,\pi)} f(t_i, X_{t_i}^{(\pi)}, 0, 0) \Delta_i \right].$$

Multiplying with Δ_i and summing i from 0 to $N - 1$ easily gives the L^2 -bound for $Z^{(1,\pi)}$ in view of (6). \square

As a corollary we obtain a uniform bound for the L^2 -norms:

Corollary 9. *Under the assumptions of Theorem 2 there is a constant C depending on the data only such that*

$$\max_{0 \leq i \leq N} E \left[|Y_{t_i}^{(n,\pi)}|^2 \right] + \sum_{i=0}^{N-1} E \left[|Z_{t_i}^{(n,\pi)}|^2 \right] \Delta_i \leq C$$

provided $|\pi|$ is sufficiently small.

Proof. With the notation from the proof of [Theorem 5](#) we get for sufficiently small $|\pi|$,

$$\begin{aligned} & \max_{0 \leq i \leq N} E \left[|Y_{t_i}^{(n,\pi)}|^2 \right] + \sum_{i=0}^{N-1} E \left[|Z_{t_i}^{(n,\pi)}|^2 \right] \Delta_i \\ & \leq \max_{0 \leq i \leq N} \sum_{v=1}^n \left(E \left[|y_{t_i}^{(n,\pi)}|^2 \right] + \sum_{i=0}^{N-1} E \left[|z_{t_i}^{(n,\pi)}|^2 \right] \Delta_i \right) \leq \left(\sum_{v=1}^{\infty} A(\pi, v) \right)^2 \\ & \leq C \left(\max_{0 \leq i \leq N} E \left[|Y_{t_i}^{(1,\pi)}|^2 \right] + \sum_{i=0}^{N-1} E \left[|Z_{t_i}^{(1,\pi)}|^2 \right] \Delta_i \right) \end{aligned}$$

with a constant C depending on the data only. Application of [Lemma 8](#) concludes the proof. \square

3. Generic analysis of the error propagation

For numerical implementation of the iteration proposed in the previous section, one has to approximate the conditional expectations. This section is devoted to an analysis of the error due to the replacement of the conditional expectation by a generic estimator. It turns out that the error grows moderately when the mesh of the partition goes to zero and the number of Picard iterations tends to infinity. We believe that this is an important advantage over the backward scheme, where the error explodes when the mesh tends to zero.

Suppose a generic estimator $\widehat{E}^\pi[\cdot|\mathcal{F}_t]$ of the conditional expectation is given. We consider first the corresponding approximation of the backward scheme of Bouchard and Touzi [\[4\]](#), namely

$$\begin{aligned} \widehat{Y}_{t_N}^{(\infty,\pi)} &= \xi^{(\pi)}, \\ \widehat{Z}_{d,t_i}^{(\infty,\pi)} &= \widehat{E}^\pi \left[\frac{\Delta W_{d,i}}{\Delta_i} \widehat{Y}_{t_{i+1}}^{(\infty,\pi)} \middle| \mathcal{F}_{t_i} \right], \\ \widehat{Y}_{t_i}^{(\infty,\pi)} &= \widehat{E}^\pi [\widehat{Y}_{t_{i+1}}^{(\infty,\pi)} | \mathcal{F}_{t_i}] - f(t_i, X_{t_i}^{(\pi)}, \widehat{Y}_{t_i}^{(\infty,\pi)}, \widehat{Z}_{t_i}^{(\infty,\pi)}) \Delta_i. \end{aligned} \quad (7)$$

Bouchard and Touzi [\[4\]](#), Theorem 4.1, prove, under slightly stronger assumptions than [Assumption 1](#), that

$$\begin{aligned} \max_{0 \leq i \leq N} E[|\widehat{Y}_{t_i}^{(\infty,\pi)} - Y_{t_i}^{(\infty,\pi)}|^2] &\leq \frac{C}{|\pi|} \max_{0 \leq j \leq N} E \left(\left| \widehat{E}^\pi [\widehat{Y}_{t_{i+1}}^{(\infty,\pi)} | \mathcal{F}_{t_i}] - E[\widehat{Y}_{t_{i+1}}^{(\infty,\pi)} | \mathcal{F}_{t_i}] \right|^2 \right. \\ &\quad \left. + \left| \widehat{E}^\pi \left[\frac{W_{t_{i+1}} - W_{t_i}}{t_{i+1} - t_i} \widehat{Y}_{t_{i+1}}^{(\infty,\pi)} \middle| \mathcal{F}_{t_i} \right] - E \left[\frac{W_{t_{i+1}} - W_{t_i}}{t_{i+1} - t_i} \widehat{Y}_{t_{i+1}}^{(\infty,\pi)} \middle| \mathcal{F}_{t_i} \right] \right|^2 \right) \end{aligned}$$

for some constant C depending on the data.

This means, given the same accuracy of the conditional expectation estimator, that the error due to the approximation of the conditional expectation explodes when the mesh of the partition tends to zero. Put differently, due to the numerical approximation of the conditional expectation by a Monte Carlo based estimator one has to simulate more paths as the partition becomes finer. This increases the computational costs. This effect is particularly unfavorable when the constant in [Theorem 4](#) is large (e.g. due to a large Lipschitz constant or time horizon) and, thus, a fine mesh is needed for $Y_t^{(\infty,\pi)}$ to be a good approximation of Y_t . We note that the effect described has also been observed in numerical examples by Gobet et al. [\[9\]](#).

It is intuitively clear that there should be no significant error propagation backwards in time in our forward scheme, because we do not have to estimate nested conditional expectations

backwards in time. We shall now show that the error due to the approximation of the conditional expectation by its generic estimator does not explode in the number of iterations. To this end we define

$$\begin{aligned}\widehat{Y}_{t_i}^{(n,\pi)} &= \widehat{E}^\pi \left[\xi^{(\pi)} - \sum_{j=i}^{N-1} f(t_j, X_{t_j}^{(\pi)}, \widehat{Y}_{t_j}^{(n-1,\pi)}, \widehat{Z}_{t_j}^{(n-1,\pi)}) \Delta_j \middle| \mathcal{F}_{t_i} \right], \\ \widehat{Z}_{d,t_i}^{(n,\pi)} &= \widehat{E}^\pi \left[\frac{\Delta W_{d,i}}{\Delta_i} \left(\xi^{(\pi)} - \sum_{j=i+1}^{N-1} f(t_j, X_{t_j}^{(\pi)}, \widehat{Y}_{t_j}^{(n-1,\pi)}, \widehat{Z}_{t_j}^{(n-1,\pi)}) \Delta_j \right) \middle| \mathcal{F}_{t_i} \right],\end{aligned}$$

initialized at $(\widehat{Y}^{(0,\pi)}, \widehat{Z}^{(0,\pi)}) = (0, 0)$.

Theorem 10. Define

$$\widehat{R}(i, n, \pi) = \xi^{(\pi)} - \sum_{j=i}^{N-1} f(t_j, X_{t_j}^{(\pi)}, \widehat{Y}_{t_j}^{(n-1,\pi)}, \widehat{Z}_{t_j}^{(n-1,\pi)}) \Delta_j.$$

Under [Assumption 1](#) there is a constant C depending on the data such that for any sufficiently fine partition π , and for all $n \in \mathbb{N}$,

$$\begin{aligned}& \max_{0 \leq i \leq N} E[|\widehat{Y}_{t_i}^{(n,\pi)} - Y_{t_i}^{(n,\pi)}|^2] + \sum_{i=0}^{N-1} E[|\widehat{Z}_{t_i}^{(n,\pi)} - Z_{t_i}^{(n,\pi)}|^2] \Delta_i \\ & \leq C \max_{1 \leq v \leq n} \left(\max_{0 \leq i \leq N} E \left[|\widehat{E}^\pi[\widehat{R}(i, v, \pi) | \mathcal{F}_{t_i}] - E[\widehat{R}(i, v, \pi) | \mathcal{F}_{t_i}]|^2 \right] \right. \\ & \quad \left. + E \sum_{i=0}^{N-1} \left| \widehat{E}^\pi \left[\frac{\Delta W_i}{\Delta_i} \widehat{R}(i+1, v, \pi) \middle| \mathcal{F}_{t_i} \right] - E \left[\frac{\Delta W_i}{\Delta_i} \widehat{R}(i+1, v, \pi) \middle| \mathcal{F}_{t_i} \right] \right|^2 \Delta_i \right).\end{aligned}$$

Proof. We define

$$R(i, n, \pi) = \xi^{(\pi)} - \sum_{j=i}^{N-1} f(t_j, X_{t_j}^{(\pi)}, Y_{t_j}^{(n-1,\pi)}, Z_{t_j}^{(n-1,\pi)}) \Delta_j.$$

Then, by Young's inequality, and with the notation from [Lemma 7](#), i.e. with $\lambda_0 = 1$, $\lambda_i = (1 + \Gamma \Delta_{i-1}) \lambda_{i-1}$, and $\Gamma > 0$ to be chosen later,

$$\begin{aligned}& \max_{0 \leq i \leq N} \lambda_i E[|\widehat{Y}_{t_i}^{(n,\pi)} - Y_{t_i}^{(n,\pi)}|^2] + \sum_{i=0}^{N-1} \lambda_i E[|\widehat{Z}_{t_i}^{(n,\pi)} - Z_{t_i}^{(n,\pi)}|^2] \Delta_i \\ & \leq 2 \left(\max_{0 \leq i \leq N} \lambda_i E \left[|\widehat{E}^\pi[\widehat{R}(i, n, \pi) | \mathcal{F}_{t_i}] - E[\widehat{R}(i, n, \pi) | \mathcal{F}_{t_i}]|^2 \right] \right. \\ & \quad \left. + E \sum_{i=0}^{N-1} \lambda_i \left| \widehat{E} \left[\frac{\Delta W_i}{\Delta_i} \widehat{R}(i+1, n, \pi) \middle| \mathcal{F}_{t_i} \right] - E \left[\frac{\Delta W_i}{\Delta_i} \widehat{R}(i+1, n, \pi) \middle| \mathcal{F}_{t_i} \right] \right|^2 \Delta_i \right) \\ & \quad + 2 \left(\max_{0 \leq i \leq N} \lambda_i E \left[|E[\widehat{R}(i, n, \pi) - R(i, n, \pi) | \mathcal{F}_{t_i}]|^2 \right] \right. \\ & \quad \left. + \sum_{i=0}^{N-1} \lambda_i E \left[\left| E \left[\frac{\Delta W_i}{\Delta_i} \widehat{R}(i+1, n, \pi) - \frac{\Delta W_i}{\Delta_i} R(i+1, n, \pi) \middle| \mathcal{F}_{t_i} \right] \right|^2 \Delta_i \right] \right).\end{aligned}$$

Lemma 7 can be applied to the second term. We choose $\gamma = 16DK^2(T+1)$ and $\Gamma = 16K^2(T+1)(\gamma DT+1)$ and obtain

$$\begin{aligned} & \max_{0 \leq i \leq N} \lambda_i E[|\widehat{Y}_{t_i}^{(n,\pi)} - Y_{t_i}^{(n,\pi)}|^2] + \sum_{i=0}^{N-1} \lambda_i E[|\widehat{Z}_{t_i}^{(n,\pi)} - Z_{t_i}^{(n,\pi)}|^2] \Delta_i \\ & \leq 2 \left(\max_{0 \leq i \leq N} \lambda_i E \left[|\widehat{E}^\pi[\widehat{R}(i, n, \pi) | \mathcal{F}_{t_i}] - E[\widehat{R}(i, n, \pi) | \mathcal{F}_{t_i}]|^2 \right] \right. \\ & \quad \left. + E \sum_{i=0}^{N-1} \lambda_i \left| \widehat{E}^\pi \left[\frac{\Delta W_i}{\Delta_i} \widehat{R}(i+1, n, \pi) \middle| \mathcal{F}_{t_i} \right] - E \left[\frac{\Delta W_i}{\Delta_i} \widehat{R}(i+1, n, \pi) \middle| \mathcal{F}_{t_i} \right] \right|^2 \Delta_i \right) \\ & \quad + \left(\frac{1}{4} + \frac{\Gamma}{4} |\pi| \right) \left(\max_{0 \leq i \leq N} \lambda_i E[|\widehat{Y}_{t_i}^{(n-1,\pi)} - Y_{t_i}^{(n-1,\pi)}|^2] \right. \\ & \quad \left. + \sum_{i=0}^{N-1} \lambda_i E[|\widehat{Z}_{t_i}^{(n-1,\pi)} - Z_{t_i}^{(n-1,\pi)}|^2] \Delta_i \right). \end{aligned}$$

Now for $|\pi|$ sufficiently small (e.g. less than or equal to Γ^{-1}) the above estimate can be iterated to obtain the theorem. Note that $1 \leq \lambda_i \leq e^{\Gamma T}$. Thus, we can choose $C = 2e^{\Gamma T} \vee \Gamma$. \square

4. A numerical forward scheme

In this section we specify an estimator for the conditional expectation. We shall utilize the so-called least squares Monte Carlo regression method, which was introduced in [10] in the context of American options and is also applied to the backward scheme in [9]. The approximation takes place in two steps. First, the conditional expectation is replaced by an orthogonal projection on finite dimensional subspaces. Then, the coefficients of the orthogonal projections are estimated from a sample of independent simulations by the least squares method. Convergence of these two steps will be analyzed in the following subsections. Section 4.3 summarizes the results in a Markovian setting relevant for the practical implementation of the numerical scheme.

4.1. Orthogonal projection on subspaces of $L^2(\mathcal{F}_{t_i})$

We will first replace the conditional expectations $E[\cdot | \mathcal{F}_{t_i}]$ by orthogonal projections on subspaces of $L^2(\mathcal{F}_{t_i})$. To be precise, we fix $D+1$ subspaces $\Lambda_{d,i}$, $0 \leq d \leq D$, of $L^2(\mathcal{F}_{t_i})$ for each $0 \leq i \leq k$. The orthogonal projection on $\Lambda_{d,i}$ is denoted by $P_{d,i}$.

We now consider the algorithm

$$\begin{aligned} \widehat{Y}_{t_i}^{(n,\pi)} &= P_{0,i} \left[\xi^{(\pi)} - \sum_{j=i}^{N-1} f(t_j, X_{t_j}^{(\pi)}, \widehat{Y}_{t_j}^{(n-1,\pi)}, \widehat{Z}_{t_j}^{(n-1,\pi)}) \Delta_j \right], \\ \widehat{Z}_{d,t_i}^{(n,\pi)} &= P_{d,i} \left[\frac{\Delta W_{d,i}}{\Delta_i} \left(\xi^{(\pi)} - \sum_{j=i+1}^{N-1} f(t_j, X_{t_j}^{(\pi)}, \widehat{Y}_{t_j}^{(n-1,\pi)}, \widehat{Z}_{t_j}^{(n-1,\pi)}) \Delta_j \right) \right], \end{aligned}$$

initialized at $(\widehat{Y}^{(0,\pi)}, \widehat{Z}^{(0,\pi)}) = 0$.

Our aim is to analyze the error of $(\widehat{Y}^{(n,\pi)}, \widehat{Z}^{(n,\pi)})$ as compared to $(Y^{(n,\pi)}, Z^{(n,\pi)})$ in terms of the projection errors $|Y_{t_i}^{(n,\pi)} - P_{0,i}[Y_{t_i}^{(n,\pi)}]|$ and $|Z_{d,t_i}^{(n,\pi)} - P_{d,i}[Z_{d,t_i}^{(n,\pi)}]|$. The main feature of

the algorithm – as can be expected in view of [Theorem 10](#) – is that the error does not propagate backwards in time. Moreover, the error does not explode when the number of iterations tends to infinity. This is an important advantage compared to the scheme proposed in [9] where the projection errors sum over the time steps. Roughly speaking, in the Gobet et al. [9] scheme the L^2 -error is bounded by \sqrt{N} times a constant times the worst L^2 -projection error (see their [Theorem 2](#)). The following theorem states that in our scheme the L^2 -error is bounded by a constant times the worst L^2 -projection error.

Theorem 11. *Suppose f is Lipschitz in (y, z) uniformly in (t, x) with constant K . Then there is a constant C depending on the data such that*

$$\begin{aligned} & \max_{0 \leq i \leq N} E \left[|\widehat{Y}_{t_i}^{(n, \pi)} - Y_{t_i}^{(n, \pi)}|^2 \right] + \sum_{i=0}^{N-1} E \left[|\widehat{Z}_{t_i}^{(n, \pi)} - Z_{t_i}^{(n, \pi)}|^2 \right] \Delta_i \\ & \leq C \sum_{v=0}^n \left(\frac{1}{2} + C|\pi| \right)^{n-v} \left(\sum_{i=0}^{N-1} E \left[|Y_{t_i}^{(v, \pi)} - P_{0,i}[Y_{t_i}^{(v, \pi)}]|^2 \right] \Delta_i \right. \\ & \quad \left. + \sum_{d=1}^D \sum_{i=0}^{N-1} E \left[|Z_{d,t_i}^{(v, \pi)} - P_{d,i}[Z_{d,t_i}^{(v, \pi)}]|^2 \right] \Delta_i \right) \end{aligned}$$

for sufficiently small $|\pi|$. In particular, with a possibly different constant C ,

$$\begin{aligned} & \max_{0 \leq i \leq N} E \left[|\widehat{Y}_{t_i}^{(n, \pi)} - Y_{t_i}^{(n, \pi)}|^2 \right] + \sum_{i=0}^{N-1} E \left[|\widehat{Z}_{t_i}^{(n, \pi)} - Z_{t_i}^{(n, \pi)}|^2 \right] \Delta_i \\ & \leq C \max_{0 \leq v \leq n} \max_{0 \leq i \leq N} \left(E \left[|Y_{t_i}^{(v, \pi)} - P_{0,i}[Y_{t_i}^{(v, \pi)}]|^2 \right] + \sum_{d=1}^D E \left[|Z_{d,t_i}^{(v, \pi)} - P_{d,i}[Z_{d,t_i}^{(v, \pi)}]|^2 \right] \right). \end{aligned}$$

Proof. We define

$$\begin{aligned} \overline{Y}_{t_i}^{(n, \pi)} &= E \left[\xi^{(\pi)} - \sum_{j=i}^{N-1} f(t_j, X_{t_j}^{(\pi)}, \widehat{Y}_{t_j}^{(n-1, \pi)}, \widehat{Z}_{t_j}^{(n-1, \pi)}) \Delta_j \middle| \mathcal{F}_{t_i} \right], \\ \overline{Z}_{d,t_i}^{(n, \pi)} &= E \left[\frac{\Delta W_{d,i}}{\Delta_i} \left(\xi^{(\pi)} - \sum_{j=i}^{N-1} f(t_j, X_{t_j}^{(\pi)}, \widehat{Y}_{t_j}^{(n-1, \pi)}, \widehat{Z}_{t_j}^{(n-1, \pi)}) \Delta_j \right) \middle| \mathcal{F}_{t_i} \right]. \end{aligned}$$

Notice that

$$\begin{aligned} P_{0,i} \left(\overline{Y}_{t_i}^{(n, \pi)} - Y_{t_i}^{(n, \pi)} \right) &= \widehat{Y}_{t_i}^{(n, \pi)} - P_{0,i} \left(Y_{t_i}^{(n, \pi)} \right), \\ P_{d,i} \left(\overline{Z}_{d,t_i}^{(n, \pi)} - Z_{d,t_i}^{(n, \pi)} \right) &= \widehat{Z}_{d,t_i}^{(n, \pi)} - P_{d,i} \left(Z_{d,t_i}^{(n, \pi)} \right). \end{aligned}$$

Since the orthogonal projection has norm 1 and applying [Lemma 7](#) with $\tilde{Y}^{(1)} = \overline{Y}^{(n, \pi)}$, $\tilde{Z}^{(1)} = \overline{Z}^{(n, \pi)}$, $\tilde{Y}^{(2)} = Y^{(n, \pi)}$, and $\tilde{Z}^{(2)} = Z^{(n, \pi)}$, we obtain

$$\begin{aligned} & \max_{0 \leq i \leq N} \lambda_i E \left[|\widehat{Y}_{t_i}^{(n, \pi)} - P_{0,i}(Y_{t_i}^{(n, \pi)})|^2 \right] + \sum_{d=1}^D \sum_{i=0}^{N-1} \lambda_i E \left[|\widehat{Z}_{d,t_i}^{(n, \pi)} - P_{d,i}(Z_{d,t_i}^{(n, \pi)})|^2 \right] \Delta_i \\ & \leq \max_{0 \leq i \leq N} \lambda_i E \left[|\overline{Y}_{t_i}^{(n, \pi)} - Y_{t_i}^{(n, \pi)}|^2 \right] + \sum_{i=0}^{N-1} \lambda_i E \left[|\overline{Z}_{t_i}^{(n, \pi)} - Z_{t_i}^{(n, \pi)}|^2 \right] \Delta_i \end{aligned}$$

$$\leq K^2(T+1) \left((|\pi| + \Gamma^{-1}) (\gamma DT + 1) + \frac{D}{\gamma} \right) \\ \times \left(\frac{1}{T} \sum_{i=0}^{N-1} \lambda_i E \left[|\widehat{Y}_{t_i}^{(n-1,\pi)} - Y_{t_i}^{(n-1,\pi)}|^2 \right] + \sum_{i=0}^{N-1} \lambda_i E \left[|\widehat{Z}_{t_i}^{(n-1,\pi)} - Z_{t_i}^{(n-1,\pi)}|^2 \right] \Delta_i \right)$$

for any γ , $\Gamma > 0$ with $\lambda_0 = 1$ and $\lambda_i = (1 + \Gamma \Delta_{i-1}) \lambda_{i-1}$. The rest of the proof now follows the same lines as the proof of [Theorem 10](#) taking into account that, due to the orthogonality of the orthogonal projection,

$$E \left[|\widehat{Y}_{t_i}^{(v,\pi)} - Y_{t_i}^{(v,\pi)}|^2 \right] = E \left[|\widehat{Y}_{t_i}^{(v,\pi)} - P_{0,i}[Y_{t_i}^{(v,\pi)}]|^2 \right] + E \left[|Y_{t_i}^{(v,\pi)} - P_{0,i}[Y_{t_i}^{(v,\pi)}]|^2 \right].$$

□

In view of [Corollary 9](#) and [Theorem 11](#) we also get uniform L^2 -bounds for $\widehat{Y}^{(n,\pi)}$ and $\widehat{Z}^{(n,\pi)}$.

Corollary 12. *Under the assumptions of [Theorem 2](#), there is a constant C depending on the data only such that*

$$\max_{0 \leq i \leq N} E \left[|\widehat{Y}_{t_i}^{(n,\pi)}|^2 \right] + \sum_{i=0}^{N-1} E \left[|\widehat{Z}_{t_i}^{(n,\pi)}|^2 \right] \Delta_i \leq C$$

provided $|\pi|$ is sufficiently small.

4.2. A Monte Carlo least squares method for approximating conditional expectations

In a next step we replace the projection on subspaces by a simulation based least squares estimator.

To avoid an overload in notation and since the generalization is straightforward, we shall consider the case $D = 1$ only.

We now assume that the projection spaces from the previous section are all finite dimensional and denote by

$$\{\eta_1^i, \dots, \eta_{K(i)}^i\} \quad (\text{resp.} \quad \{\tilde{\eta}_1^i, \dots, \tilde{\eta}_{\tilde{K}(i)}^i\})$$

a basis of $A_{0,i}$ and $A_{1,i}$, respectively. The inner-product matrices associated with these bases are denoted by

$$\mathcal{B}_i = \left(E[\eta_k^i \eta_l^i] \right)_{k,l=0,\dots,K(i)} \quad \left(\text{resp.} \quad \tilde{\mathcal{B}}_i = \left(E[\tilde{\eta}_k^i \tilde{\eta}_l^i] \right)_{k,l=0,\dots,\tilde{K}(i)} \right).$$

In this situation the processes $\widehat{Y}^{(n,\pi)}$ and $\widehat{Z}^{(n,\pi)}$ may be rewritten as

$$\widehat{Y}_{t_i}^{(n,\pi)} = \sum_{k=1}^{K(i)} \alpha_{i,k}^{(n,\pi)} \eta_k^i, \tag{8} \\ \widehat{Z}_{t_i}^{(n,\pi)} = \sum_{k=1}^{\tilde{K}(i)} \tilde{\alpha}_{i,k}^{(n,\pi)} \tilde{\eta}_k^i,$$

where, e.g. with $\eta^i = (\eta_1^i, \dots, \eta_{K(i)}^i)^*$, $\alpha_i^{(n,\pi)} = (\alpha_{i,1}^{(n,\pi)}, \dots, \alpha_{i,K(i)}^{(n,\pi)})^*$,

$$\alpha_i^{(n,\pi)} = \mathcal{B}_i^{-1} E \left[\eta^i \left(\xi^{(\pi)} - \sum_{j=i}^{N-1} f(t_j, X_{t_j}^{(\pi)}, \widehat{Y}_{t_j}^{(n-1,\pi)}, \widehat{Z}_{t_j}^{(n-1,\pi)}) \Delta_j \right) \right], \quad (9)$$

$$\widetilde{\alpha}_i^{(n,\pi)} = \widetilde{\mathcal{B}}_i^{-1} E \left[\widetilde{\eta}^i \frac{\Delta W_i}{\Delta_i} \left(\xi^{(\pi)} - \sum_{j=i+1}^{N-1} f(t_j, X_{t_j}^{(\pi)}, \widehat{Y}_{t_j}^{(n-1,\pi)}, \widehat{Z}_{t_j}^{(n-1,\pi)}) \Delta_j \right) \right].$$

The expectations in (9) will be replaced by their simulation based estimators. We shall therefore assume that we have $L \geq \max_i \{K(i) \vee \widetilde{K}(i)\}$ independent copies $(\Delta_\lambda W_i, {}_\lambda \xi^{(\pi)}, {}_\lambda X_{t_i}^{(\pi)}, {}_\lambda \eta_k^i, {}_\lambda \widetilde{\eta}_k^i)$, $\lambda = 1, \dots, L$, of $(\Delta W_i, \xi^{(\pi)}, X_{t_i}^{(\pi)}, \eta_k^i, \widetilde{\eta}_k^i)$. The column vectors of these copies are denoted by $(\Delta \mathbf{W}_i, \mathbf{x}^{(\pi)}, \mathbf{X}_{t_i}^{(\pi)}, \mathbf{e}_k^i, \widetilde{\mathbf{e}}_k^i)$, e.g.

$$\mathbf{x}^{(\pi)} = ({}_1 \xi^{(\pi)}, \dots, {}_L \xi^{(\pi)})^*.$$

We define

$$\mathcal{A}_i^L = \frac{1}{\sqrt{L}} \left({}_\lambda \eta_k^i \right)_{\lambda=1, \dots, L, k=1, \dots, K(i)}$$

and $\widetilde{\mathcal{A}}_i^L$ similarly. Note that

$$\mathcal{B}_i^L = (\mathcal{A}_i^L)^* \mathcal{A}_i^L = \frac{1}{L} \left(\sum_{\lambda=1}^L {}_\lambda \eta_k^i {}_\lambda \eta_l^i \right)_{k,l=1, \dots, K(i)}$$

is the simulation based analogue of \mathcal{B}_i . Since the inverse of \mathcal{B}_i^L , in general, does not exist, we shall make use of the pseudo-inverses $(\mathcal{A}_i^L)^+$, $(\widetilde{\mathcal{A}}_i^L)^+$ to define simulation based analogues of (9) recursively by the least squares method, i.e.

$$\alpha_{i,k}^{(0,\pi,L)} = \widetilde{\alpha}_{i,k}^{(0,\pi,L)} = 0$$

$$\widehat{\mathbf{Y}}_{t_i}^{(n-1,\pi)} = ({}_ \lambda \widehat{Y}_{t_i}^{(n-1,\pi)})_{\lambda=1, \dots, L} = \sum_{k=1}^{K(i)} \alpha_{i,k}^{(n-1,\pi,L)} \mathbf{e}_k^i$$

$$\widehat{\mathbf{Z}}_{t_i}^{(n-1,\pi)} = ({}_ \lambda \widehat{Z}_{t_i}^{(n-1,\pi)})_{\lambda=1, \dots, L} = \sum_{k=1}^{\widetilde{K}(i)} \widetilde{\alpha}_{i,k}^{(n-1,\pi,L)} \widetilde{\mathbf{e}}_k^i$$

$$\mathbf{f}(t_j) = \left(f(t_j, {}_\lambda X_{t_j}^{(\pi)}, {}_\lambda \widehat{Y}_{t_j}^{(n-1,\pi)}, {}_\lambda \widehat{Z}_{t_j}^{(n-1,\pi)}) \right)_{\lambda=1, \dots, L}, \quad j \geq i$$

$$\alpha_i^{(n,\pi,L)} = \frac{1}{\sqrt{L}} (\mathcal{A}_i^L)^+ \left(\mathbf{x}^{(\pi)} - \sum_{j=i}^{N-1} \mathbf{f}(t_j) \Delta_j \right)$$

$$\widetilde{\alpha}_i^{(n,\pi,L)} = \frac{1}{\sqrt{L}} (\widetilde{\mathcal{A}}_i^L)^+ \left(\frac{\Delta \mathbf{W}_i}{\Delta_i} \bullet \left(\mathbf{x}^{(\pi)} - \sum_{j=i+1}^{N-1} \mathbf{f}(t_j) \Delta_j \right) \right),$$

where \bullet denotes the componentwise multiplication of two vectors.

The simulation based estimators are now defined by

$$\widehat{Y}_{t_i}^{(n,\pi,L)} = \sum_{k=1}^{K(i)} \alpha_{i,k}^{(n,\pi,L)} \eta_k^i,$$

$$\widehat{Z}_{t_i}^{(n,\pi,L)} = \sum_{k=1}^{\widetilde{K}(i)} \widetilde{\alpha}_{i,k}^{(n,\pi,L)} \widetilde{\eta}_k^i.$$

Remark 13. For $t_i = t_0 = 0$ the only choice of the projection space is $\Lambda_{0,0} = \mathbb{R}$. Taking $\{1\}$ as the basis we observe that $\widehat{Y}_{t_0}^{(n,\pi,L)}$ reduces to the plain Monte Carlo estimator

$$\widehat{Y}_{t_0}^{(n,\pi,L)} = \frac{1}{L} \sum_{\lambda=1}^L \left(\lambda \xi^{(\pi)} - \sum_{j=0}^{N-1} f(t_j, {}_{\lambda}X_{t_j}^{(\pi)}, {}_{\lambda}\widehat{Y}_{t_j}^{(n-1,\pi)}, {}_{\lambda}\widehat{Z}_{t_j}^{(n-1,\pi)}) \Delta_j \right).$$

Of course, the same remark applies to $\widehat{Z}_{t_0}^{(n,\pi,L)}$.

Almost sure convergence of the simulation based estimators is a direct consequence of the following lemma.

Lemma 14. Under the Lipschitz condition of Theorem 11 $(\alpha_{i,k}^{(n,\pi,L)}, \widetilde{\alpha}_{i,k}^{(n,\pi,L)})$ converges P -almost surely to $(\alpha_{i,k}^{(n,\pi)}, \widetilde{\alpha}_{i,k}^{(n,\pi)})$, when L tends to infinity.

The lemma can be proved by induction on n , making use of the law of large numbers in a fairly straightforward way. We therefore omit the proof.

We can now easily obtain the following theorem.

Theorem 15. Under the Lipschitz condition of Theorem 11 $(\widehat{Y}_{t_i}^{(n,\pi,L)}, \widehat{Z}_{t_i}^{(n,\pi,L)})$ converges P -almost surely to $(\widehat{Y}_{t_i}^{(n,\pi)}, \widehat{Z}_{t_i}^{(n,\pi)})$, when L tends to infinity.

Remark 16. If the estimator is appropriately truncated, L^2 -convergence can be proved and rates of convergence can be derived; see [2]. It can be shown that the convergence is of the best expected order $1/2$ in L , provided the elements of the basis are bounded. Moreover, compared to the error estimates in [9], Theorem 3, the error estimates for our scheme are of order $1/2$ better in N than for their backward scheme. This is in accordance with the discussion in Section 3 above.

4.3. A Markovian setting

Now the results from the previous sections can be put together and made more explicit in a Markovian setting.

(1) *Discretization of X* : We discretize X by the Euler scheme

$$X_0^{(\pi)} = x$$

$$X_{t_i}^{(\pi)} = X_{t_{i-1}}^{(\pi)} + b(t_{i-1}, X_{t_{i-1}}^{(\pi)}) \Delta_{i-1} + \sigma(t_{i-1}, X_{t_{i-1}}^{(\pi)}) \Delta W_{i-1},$$

and extend $X^{(\pi)}$ to an RCLL process by piecewise constant interpolation. When X is known to be strictly positive, it can be more convenient to apply the Euler scheme to $\ln(X)$ instead of X ; see [9]. Note that $(X_{t_i}^{(\pi)}, \mathcal{F}_{t_i})$ forms a Markov chain.

(2) *Terminal Condition* $\xi^{(\pi)}$: The terminal condition $\xi^{(\pi)}$ is supposed to be of the form

$$\xi^{(\pi)} = \Phi^{(\pi)}(\Xi_{t_N}^{(\pi)})$$

where $(\Xi_{t_i}^{(\pi)}, \mathcal{F}_{t_i})$ is an M' -dimensional Markov chain with $X_{t_i}^{(\pi)}$ as its first M components and $\Phi^{(\pi)}$ is a deterministic function

Typical extensions for the last components of $\Xi_{t_i}^{(\pi)}$ are $\max_{0 \leq j \leq i} X_{t_j}^{(\pi)}$, $\min_{0 \leq j \leq i} X_{t_j}^{(\pi)}$, or $\sum_{j=0}^{i-1} X_{t_j}^{(\pi)}$. These extensions are of crucial importance for financial problems related to exotic options such as Asian options and lookback options. We now give some convergence results for terminal conditions $\xi^{(\pi)}$ of the above type, which are simple consequences of Corollary 4.4 in [17].

Example 17. (i) Suppose $\phi : \mathbb{R}^{2M} \rightarrow \mathbb{R}$ is Lipschitz-continuous. Then

$$E \left[\left| \phi \left(X_T, \int_0^T X_s ds \right) - \phi \left(X_T^{(\pi)}, \sum_{i=0}^{N-1} X_{t_i}^{(\pi)} \Delta_i \right) \right|^2 \right] \leq C |\pi|.$$

(ii) Suppose $\phi : \mathbb{R}^{4M} \rightarrow \mathbb{R}$ is Lipschitz-continuous. Then

$$E \left[\left| \phi \left(X_T, \int_0^T X_s ds, \max_{0 \leq t \leq T} X_t, \min_{0 \leq t \leq T} X_t \right) - \phi \left(X_T^{(\pi)}, \sum_{i=0}^{N-1} X_{t_i}^{(\pi)} \Delta_i, \max_{0 \leq j \leq i} X_{t_j}^{(\pi)}, \min_{0 \leq j \leq i} X_{t_j}^{(\pi)} \right) \right|^2 \right] \leq C |\pi| \ln \left(\frac{1}{|\pi|} \right).$$

(3) *Choice of the basis*: One may choose a set of functions $\{e_1(x), \dots, e_\kappa(x)\}$ and define the basis via

$$\eta_k^i = e_k(\Xi_{t_i}^{(\pi)}).$$

Typical choices are indicator functions or (exponentially damped) polynomials. In principle the basis functions e_k may depend on d , but for simulations it might be more convenient to work with one set of functions only.

In the situation described above it is easily checked that

$$Y_{t_i}^{(n,\pi)} = E \left[\xi^{(\pi)} - \sum_{j=i}^{N-1} f(t_j, X_{t_j}^{(\pi)}, Y_{t_j}^{(n-1,\pi)}, Z_{t_j}^{(n-1,\pi)}) \Delta_j \middle| \Xi_{t_i}^{(\pi)} \right],$$

$$Z_{d,t_i}^{(n,\pi)} = E \left[\frac{\Delta W_{d,i}}{\Delta_i} \left(\xi^{(\pi)} - \sum_{j=i+1}^{N-1} f(t_j, X_{t_j}^{(\pi)}, Y_{t_j}^{(n-1,\pi)}, Z_{t_j}^{(n-1,\pi)}) \Delta_j \right) \middle| \Xi_{t_i}^{(\pi)} \right].$$

Hence, if $\{e_1(x), \dots, e_\kappa(x)\}$ are the initial elements of a sequence $(e_k)_{k \in \mathbb{N}}$ such that

$$(e_k(\Xi_{t_i}^{(\pi)}))_{k \in \mathbb{N}}$$

is total in $L^2(\sigma(\Xi_{t_i}^{(\pi)}))$ and are linearly independent for all $0 \leq i \leq N-1$, then, by virtue of Theorem 11, $(\hat{Y}^{(n,\pi)}, \hat{Z}^{(n,\pi)})$ converges (in the L^2 -sense of Theorem 11) to $(Y^{(n,\pi)}, Z^{(n,\pi)})$ as κ tends to infinity. Hence, Theorem 2 and Remark 16 provide L^2 -convergence for (an appropriately truncated version of) the constructed Monte Carlo estimator in this situation.

5. Simulations

In this section we present some simulations for a hedging problem with different interest rates for borrowing and investing. Throughout the section the process X is one-dimensional representing a stock in the standard Black–Scholes model, i.e.

$$X_t = X_0 \exp\{\sigma W_t + \mu t - 1/2\sigma^2 t\}.$$

It is discretized by the log-Euler scheme. In all cases we will apply an equidistant partition of the interval $[0, T]$ with $N + 1$ points denoted by π_N .

In the example we numerically evaluate a straddle, i.e. the sum of a call and a put option, under different rates for borrowing and investing in the money market account. The rate for borrowing is denoted by R , the one for investing by r . The fair price of a straddle in this model is given by Y_0 , where (Y, Z) is the solution of the nonlinear BSDE

$$\begin{aligned} dY_t &= \left[rY_t + \frac{\mu - r}{\sigma} Z_t - (R - r) \left(Y_t - \frac{Z_t}{\sigma} \right)_- \right] dt + Z_t dW_t \\ Y_T &= |X_T - K|, \end{aligned}$$

see [3]. In the following we fix the parameters $X_0 = 100$, $\sigma = 0.2$, $\mu = 0.05$, $r = 0.01$, $R = 0.06$, and the straddle is supposed to be at the money, i.e. $K = 100$, with maturity $T = 2$ years. In the figures below this situation is the ‘nonlinear case’, which will be compared with the standard ‘linear case’ where $R = 0.01$, i.e. the same interest rate is applied for borrowing and investing. We stop the Picard iteration, when the distance of two subsequent time-zero values is less than 0.001. The total number of calculated iterations is denoted by n_{stop} . We compare two different bases. The first basis consists of monomials and the straddle payoff, the second of characteristic functions. To be precise,

$$\begin{aligned} e_1^{(1)}(x) &= |x - K|, & e_k^{(1)}(x) &= (x - X_0)^{k-2}, \quad 2 \leq k \leq \kappa, \\ e_1^{(2)}(x) &= \mathbf{1}_{[0,l)}(x), & e_2^{(2)}(x) &= \mathbf{1}_{[u,\infty)}(x), \\ e_k^{(2)}(x) &= \mathbf{1}_{[l+(k-3)(u-l)/(\kappa-2), l+(k-2)(u-l)/(\kappa-2))}(x), & 3 \leq k \leq \kappa. \end{aligned}$$

Here, the lower bound l and the upper bound u depend on i and the simulations. They are calculated as the empirical mean of $X_{t_i}^{(\pi_N, \lambda)}$ minus and plus, respectively, two times their empirical standard deviation.

Fig. 1 shows the simulated price of the straddle as a function of the number of partition points for both bases. We choose $\kappa = 7$ for the basis $(e_k^{(1)})_k$, and $\kappa = 21$ for $(e_k^{(2)})_k$. In both cases we simulate $L = 100\,000$ paths. One can see from Fig. 1 that there is a minimal number N_{\min} of time partition points after which the computed price is independent of $N \geq N_{\min}$. For the linear case this N_{\min} is smaller than for the nonlinear case where N_{\min} is in the range of 15–20. We remark that the number of iterations n_{stop} is about 5–6, so the computational costs are still relatively low. In the linear case the computed value is quite close to the exact price of 22.32 computed using the Black–Scholes formula. We also note that the relative standard error in the calculation of $\widehat{Y}_0^{(n_{\text{stop}}, \pi_N, 100\,000)}$ is about 0.28% for the nonlinear case and 0.29% for the linear case for both bases. The relative standard error does not change significantly in the number of partition points N . Thus, the simulation complements the assertion of Theorem 10.

Fig. 2 shows the empirical mean and the empirical standard deviation of the simulated price calculated from 100 launches of the algorithm as a function of the number of simulated paths L

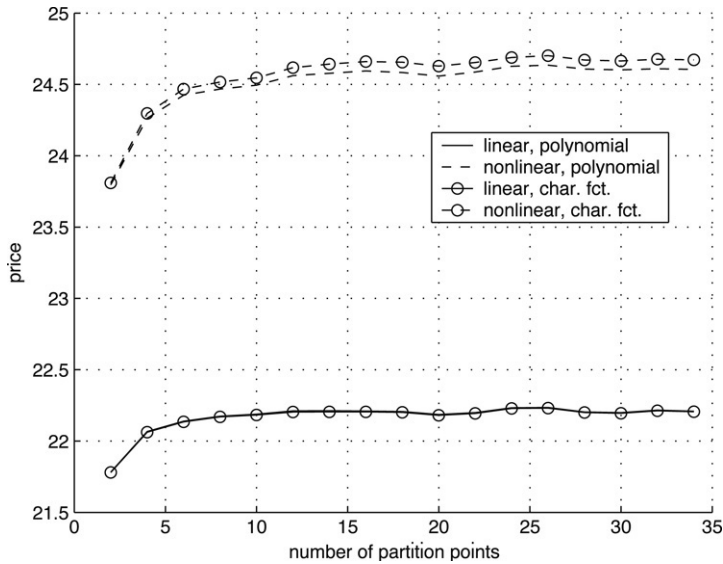


Fig. 1. $Y_0^{(n_{\text{stop}}, \pi_N, 100\,000, *)}$ as a function of N .

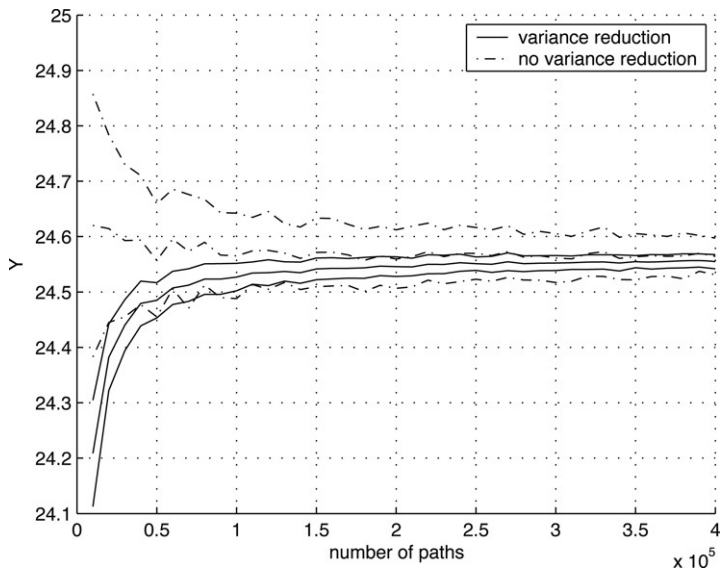


Fig. 2. Empirical mean and standard deviation of 100 launches as function of L .

per launch. Here we choose $N = 20$. The simulations have been performed with the monomial basis and $\kappa = 5$ for the nonlinear case. One can see a small positive bias of the empirical mean value which is decreasing with increasing number of paths. The standard deviation as a function of L decreases like $L^{-1/2}$ which is the expected rate. Additionally, we launched a variance reduced variant of the algorithm. To be precise, we replace

$$_{\lambda}\xi^{(\pi)} - \sum_{j=i}^{N-1} f(t_j, {}_{\lambda}X_{t_j}^{(\pi)}, {}_{\lambda}\widehat{Y}_{t_j}^{(n-1,\pi)}, {}_{\lambda}\widehat{Z}_{t_j}^{(n-1,\pi)})\Delta_j$$

by

$$_{\lambda}\xi^{(\pi)} - \sum_{j=i}^{N-1} \left(f(t_j, {}_{\lambda}X_{t_j}^{(\pi)}, {}_{\lambda}\widehat{Y}_{t_j}^{(n-1,\pi)}, {}_{\lambda}\widehat{Z}_{t_j}^{(n-1,\pi)})\Delta_j + {}_{\lambda}\widehat{Z}_{t_j}^{(n-1,\pi)} \Delta_{\lambda} W_j \right).$$

From Fig. 2 we clearly see the effect of the variance reduction. However, now there is a negative bias which is of order L^{-1} . Note that both computed values (with a positive and with a negative bias) are very close for, say, 300 000 paths, where the relative difference is less than 0.2%.

Acknowledgements

Christian Bender was partially supported by the DFG Research Center MATHEON ‘Mathematics for key technologies’ in Berlin.

Robert Denk was partially supported by the AFF grant 28/04 of the University of Konstanz.

References

- [1] V. Bally, Approximation scheme for solutions of BSDE, in: N. El Karoui, L. Mazliak (Eds.), *Backward Stochastic Differential Equations*, Addison-Wesley-Longman, 1997, pp. 177–191.
- [2] C. Bender, R. Denk, A forward scheme for BSDEs: Rates of convergence for a Monte Carlo estimator, 2006 (in preparation).
- [3] Y.Z. Bergman, Option pricing with differential interest rates, *Rev. Financ. Stud.* 8 (1995) 475–500.
- [4] B. Bouchard, N. Touzi, Discrete-time approximation and Monte Carlo Simulation of backward stochastic differential equations, *Stochastic Process. Appl.* 111 (2004) 175–206.
- [5] P. Briand, B. Delyon, J. Mémin, Donsker-type theorem for BSDEs, *Electron. Comm. Probab.* 6 (2001) 1–14.
- [6] D. Chevance, Numerical methods for backward stochastic differential equations, in: L.C.G. Rogers, D. Talay (Eds.), *Numerical Methods in Finance*, University Press, Cambridge, 1997, pp. 232–244.
- [7] J. Douglas, J. Ma, P. Protter, Numerical methods for forward backward stochastic differential equations, *Ann. Appl. Probab.* 6 (1996) 940–968.
- [8] N. El Karoui, S. Peng, M.C. Quenez, Backward stochastic differential equations in finance, *Math. Finance* 7 (1997) 1–71.
- [9] E. Gobet, J.-P. Lemor, X. Warin, A regression-based Monte Carlo method to solve backward stochastic differential equations, *Ann. Appl. Probab.* 15 (2005) 2172–2202.
- [10] F.A. Longstaff, R.S. Schwartz, Valuing American options by simulation: A simple least-square approach, *Rev. Financ. Stud.* 14 (2001) 113–147.
- [11] J. Ma, P. Protter, J. San Martín, S. Soledad, Numerical method for backward stochastic differential equations, *Ann. Appl. Probab.* 12 (2002) 302–316.
- [12] J. Ma, P. Protter, J. Yong, Solving forward–backward stochastic differential equations explicitly — a four step scheme, *Probab. Theory Related Fields* 98 (1994) 339–359.
- [13] G.N. Milstein, M.V. Tretyakov, Numerical algorithms for forward–backward stochastic differential equations, *SIAM J. Sci. Comput.* 28 (2006) 561–582.
- [14] E. Pardoux, S. Peng, Adapted solutions of a backward stochastic differential equation, *Systems Control Lett.* 14 (1990) 55–61.
- [15] J. Yong, X.Y. Zhou, *Stochastic Controls: Hamiltonian Systems and HJB Equations*, Springer, Berlin, 2000.
- [16] J. Zhang, Some fine properties of backward stochastic differential equations, Ph.D. Thesis, Purdue University, 2001.
- [17] J. Zhang, A numerical scheme for BSDEs, *Ann. Appl. Probab.* 14 (2004) 459–488.