Error Criteria for Numerical Solutions of Backward SDEs

Christian Bender[1], Jessica Steiner[1]

April 7, 2010

Many option pricing and portfolio selection problems in mathematical finance can be reformulated in terms of backward SDEs (BSDEs). As the corresponding BSDE can rarely be solved in closed form, simulation of BSDEs is of prime importance. However, the quality of the simulated solution depends on the interplay of different error sources, such as the time discretization error, the simulation error, and e.g. the choice of basis functions (if the conditional expectations are estimated by least squares Monte Carlo). In this paper we suggest error criteria which can be calculated explicitly in terms of the simulated solutions. Under suitable conditions the convergence behaviour of these observable error criteria can be linked to the approximation error between the simulated solution and the unknown true solution. Finally, we illustrate how the error criteria can be applied to judge the quality of simulated solutions for a non-linear option pricing problem. We also suggest some kind of non-linear control variates which reduces the numerical error considerably in our simulation results.

MSC (2010): Primary 65C30, secondary: 60H10, 65C05, 91G60
Keywords: BSDE, Numerics, Monte Carlo simulation, Option Pricing

1 Introduction

Backward stochastic differential equations (BSDEs) are a powerful tool to solve many option pricing problems and related problems in mathematical finance, see the review paper [16] or e.g. [3, 6, 21, 27, 30] for some more recent results. Moreover, they can be considered as general adjoint equations for stochastic optimal control problems, see [33] and the references therein, and yield representation formulae for quasi-linear PDEs [29] and even some fully non-linear parabolic PDEs [13]. Therefore the design of numerical schemes for BSDEs is an important task. Since BSDEs are terminal value problems for stochastic differential equations, a natural time discretization for

[1]Saarland University, Department of Mathematics, PO Box 151150, D-66041 Saarbrücken, Germany. bender@math.uni-sb.de, steiner@math.uni-sb.de.
Financial support by the Deutsche Forschungsgemeinschaft under grant BE3933/3-1 is gratefully acknowledged.

BSDEs works backwards in time. However, the solution must be adapted to the information which increases forwards in time, which makes the construction of numerical solutions to BSDEs a more challenging problem compared to initial value problems for stochastic differential equations.

In recent years many approximation schemes for BSDEs have been suggested. One branch of algorithms exploits the connection to parabolic PDEs via the Ma-Protter-Yong four step scheme [25]. Here the numerical problem roughly reduces to the approximation of a quasi-linear parabolic Cauchy problem, see [15, 26, 28]. This approach relies on some smoothness assumptions on the coefficients of the BSDE, and the numerical solvability of the PDE may become difficult, if not prohibitive, in high spatial dimensions.

A second branch of algorithms tackles the BSDE directly, see e.g. [5, 10, 14, 18, 22, 23]. These algorithms generally can be divided into two steps. The first one is a time discretization and it is based on a notion of L^2-regularity introduced by Zhang [34] for terminal conditions, which are Lipschitz continuous functionals of the paths of a forward SDE (under the assumption that the solution of the BSDE does not couple into the coefficients of the SDE). These convergence results for the time discretization were recently generalized to coupled systems of forward-backward SDEs by Bender and Zhang [8] and to non-Lipschitz terminal conditions by Gobet and Makhlouf [19]. We note that all these results assume a Lipschitz continuous non-linearity of the BSDE, but refer to Imkeller et al. [22] for a first approach to drivers with quadratic growth.

In order to keep the time discretized solution adapted to the filtration, the time discretization scheme requires nestings of conditional expectation backwards in time (or forwards in the iterations in the Picard type scheme of [5, 7, 17]). Therefore, in the second step, an estimator for the conditional expectation has to be applied which can be nested several times without running into exploding cost. Possible choices for these simulation based procedures are least-squares Monte Carlo [5, 18, 23], quantization [2, 14], Malliavin Monte Carlo [10], or non-parametric regression [10, 12].

These estimators for conditional expectations have been successfully applied to American option pricing problems on many underlying securities, before they have been transferred to BSDEs. Due to the structure of the American option pricing problem as a linear BSDE with reflection, its connection to optimal stopping problems allows to numerically calculate solutions which have a bias low by the Longstaff-Schwartz algorithm [24] and related algorithms [11, 12, 32]. Moreover, by duality, upper biased numerical solutions can be obtained (see [1, 4, 20, 31]). In this way confidence intervals on the price of the American option can be constructed, which allow to judge the success of the numerical procedure which was applied. For general non-linear BSDEs a similiar way to construct lower and upper approximations is not available. Nonetheless, the intricate interplay between the time discretization and the design of the estimator for the conditional expectation persists for non-linear BSDEs in a similar way as for American option problems (see e.g. the analysis by Lemor et al. [23] for a least-squares Monte Carlo algorithm for BSDEs).

The aim of the present paper is to introduce some error criteria for numerical solutions of BSDEs, which – to some extent – help to judge the quality of numerical solutions of BSDEs and indicate, at which stages of the algorithms an estimation of the conditional expectations might have been unsuccessful. Suppose (Y, Z) is the solution of the BSDE

$$Y_t = \xi - \int_t^T f(s, Y_s, Z_s)ds - \int_t^T Z_s dW_s, \quad 0 \le t \le T,$$

where W is a Brownian motion and the data (ξ, f) are given. Then by the very definition the solution satisfies, for every $0 \leq t \leq s \leq T$,

$$Y_t = Y_s - \int_t^s f(s, Y_s, Z_s)ds - \int_t^s Z_s dW_s. \tag{1}$$

We now assume that $(y^{(\pi)}, z^{(\pi)})$ is some adapted approximation of the solution (Y, Z), which is piecewise constant with respect to a time partition π of $[0, T]$. If $(y^{(\pi)}, z^{(\pi)})$ is a 'good' approximation of (Y, Z), Eq. (1) suggests that, for any $t_i < t_j \in \pi$,

$$y^{(\pi)}_{t_j} - y^{(\pi)}_{t_i} - \sum_{t_k \in \pi; t_i \leq t_k < t_j} f(t_k, y^{(\pi)}_{t_k}, z^{(\pi)}_{t_k})(t_{k+1} - t_k) - \sum_{t_k \in \pi; t_i \leq t_k < t_j} z^{(\pi)}_{t_k}(W_{t_{k+1}} - W_{t_k}) \approx 0. \tag{2}$$

For the first error criterion we set $t_i = 0$ and consider the maximum (over $t_j \in \pi$) of the L^2-norms of the left-hand side of (2). We call this criterion a 'global' one, because violation of (1) over the whole segment of the paths from 0 to t_j is penalized. Our main result shows that, under standard Lipschitz conditions, the global error criterion is equivalent to the L^2-approximation error between $(y^{(\pi)}, z^{(\pi)})$ and the unknown solution (Y, Z) up to a term of the order $|\pi|^{1/2}$, provided $y^{(\pi)}_T$ approximates the terminal condition ξ reasonably well. This result holds independently of the algorithm which is used to create $(y^{(\pi)}, z^{(\pi)})$.

The second criterion is based on the L^2-norm of the left-hand side of (2) for the case $t_j = t_{i+1}$ and, hence, locally penalizes violation of (1) over one time step of the time partition. This 'local' error criterion is studied in the context of the least-squares Monte Carlo scheme of Lemor et al. [23]. We link the local error criterion to the L^2-error between the true time discretization of the solution (Y, Z) and the best projection of this time discretization on the function basis, which is applied for estimating conditional expectations. In this way, the local error criterion can help to detect time steps at which a given function basis is inappropriately chosen.

The paper is organized as follows. In Section 2 we introduce the setting of our paper and study the global error criterion. Section 3 explains how the global error criterion can be applied in the context of the least-squares Monte Carlo framework, while Section 4 is devoted to the local error criterion. A simulation study of a non-linear option pricing problem illustrates how information about the numerical solutions can be extracted from the error criteria in practice. Finally, we introduce a technique for non-linear BSDEs, which has a flavour of control variates, in Section 6. We show that this technique can reduce the numerical error significantly for the non-linear option pricing problem. Section 7 concludes.

2 Global error criterion

We consider a forward backward stochastic differential equation (FBSDE) with the following properties:

$$\begin{aligned} X_t &= x + \int_0^t b(s, X_s)ds + \int_0^t \sigma(s, X_s)dW_s \\ Y_t &= \xi - \int_t^T f(s, X_s, Y_s, Z_s)ds - \int_t^T Z_s dW_s, \end{aligned} \tag{3}$$

where $b : [0,T] \times \mathbb{R}^M \to \mathbb{R}^M$, $\sigma : [0,T] \times \mathbb{R}^M \times \mathbb{R}^{M\times D}$ and $f : [0,T] \times \mathbb{R}^M \times \mathbb{R} \times \mathbb{R}^D \to \mathbb{R}$ are deterministic functions and $W_t = (W_{1,t}, \dots, W_{D,t})^*$ is a D-dimensional Brownian motion on $[0,T]$. The initial condition of the forward SDE $x \in \mathbb{R}^M$ and the $\mathcal{F}_T$-measurable and real valued terminal condition of the backward SDE (BSDE) ξ are given. With these data, the solution of the forward SDE X is an adapted $\mathbb{R}^M$-valued process. The solution of the BSDE consists of a pair of adapted, square-integrable processes (Y,Z), where Y_t is $\mathbb{R}$-valued and $Z_t = (Z_{1,t}, \dots, Z_{D,t})$ is $\mathbb{R}^D$-valued. The FBSDE is decoupled in the sense that the coefficients of the forward SDE do not depend on Y and Z. Our FBSDE fulfils the following assumptions:

Assumption 2.1. *There is a constant K such that*

$$\begin{aligned}&\left|b(t,x) - b(t',x')\right| + \left|\sigma(t,x) - \sigma(t',x')\right| + \left|f(t,x,y,z) - f(t',x',y',z')\right| \\ &\quad \le K\left(\sqrt{t-t'} + \left|x-x'\right| + \left|y-y'\right| + \left|z-z'\right|\right)\end{aligned}$$

for all (t,x,y,z), $(t',x',y',z') \in [0,T] \times \mathbb{R}^M \times \mathbb{R} \times \mathbb{R}^D$. The terminal value $\xi = \phi(X)$ is a functional on the space of $\mathbb{R}^M$-valued RCLL functions on $[0,T]$, that satisfies the L^∞-Lipschitz condition

$$\left|\phi(\boldsymbol{x}) - \phi(\boldsymbol{x}')\right| \le K \sup_{0\le t\le T} \left|\boldsymbol{x}(t) - \boldsymbol{x}'(t)\right|$$

for all RCLL functions $\boldsymbol{x}, \boldsymbol{x}'$. In addition to that

$$\sup_{0\le t\le T} \left(|b(t,0)| + |\sigma(t,0)| + |f(t,0,0,0)|\right) + |\phi(\boldsymbol{0})| \le K,$$

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

Remark 2.2. Throughout the paper we will use generic constants C and c, that will be called 'depending on the data' if they are influenced by T, K, $X_0 = x$, M and D only. The constants may vary from line to line.

In most cases it is not possible to find a closed-form solution (Y,Z) of (3). Hence, let $(y^{(\pi)}, z^{(\pi)})$ be an approximation of the solution (Y,Z) such that $(y^{(\pi)}, z^{(\pi)})$ is piecewise constant with respect to a partition $\pi = \{0 = t_0 < t_1 < \cdots t_N = T\}$ of $[0,T]$ and for all $i = 0, \dots, N-1$

$$\mathrm{E}\left[\left|y_{t_i}^{(\pi)}\right|^2\right] < \infty, \quad \mathrm{E}\left[\left|z_{t_i}^{(\pi)}\right|^2\right] < \infty.$$

Our aim is to judge the approximation error

$$\sup_{0\le t\le T} \mathrm{E}\left[\left|Y_t - y_t^{(\pi)}\right|^2\right] + \int_0^T \mathrm{E}\left[\left|Z_t - z_t^{(\pi)}\right|^2\right] dt.$$

Since the true solution (Y,Z) is typically not available in closed form, we cannot directly compute the approximation error. Therefore we will introduce criteria, which allow simulation and from which we can derive information about the approximation error. They are inspired by the identity

$$Y_{t_i} - Y_{t_0} - \int_0^{t_i} f(t, X_t, Y_t, Z_t)dt - \int_0^{t_i} Z_t dW_t = 0 \tag{4}$$

for all $i = 1, \dots, N$. After replacing (Y, Z) by its approximation and the integrals by sums, the equality does not hold anymore. We denote by

$$\max_{1\le i\le N} \mathrm{E}\left[\left|y_{t_i}^{(\pi)} - y_{t_0}^{(\pi)} - \sum_{j=0}^{i-1} f(t_j, X_{t_j}^{(\pi)}, y_{t_j}^{(\pi)}, z_{t_j}^{(\pi)})\Delta_j - \sum_{j=0}^{i-1} z_{t_j}^{(\pi)} \Delta W_j\right|^2\right]$$
$$= \max_{1\le i\le N} \mathrm{E}\left[\left|\sum_{j=0}^{i-1}\left(y_{t_{j+1}}^{(\pi)} - y_{t_j}^{(\pi)} - f(t_j, X_{t_j}^{(\pi)}, y_{t_j}^{(\pi)}, z_{t_j}^{(\pi)})\Delta_j - z_{t_j}^{(\pi)}\Delta W_j\right)\right|^2\right]$$

a 'global' error criterion, that penalizes violation of (4) along the time discretization from time 0 to t_i, $i = 1, \dots, N$. Note that $\Delta W_i = W_{t_{i+1}} - W_{t_i}$ and $\Delta_i = t_{i+1} - t_i$. The next theorem states that this global error is almost equivalent to the approximation error.

Theorem 2.3. *Let Assumption 2.1 be fulfilled. Suppose further, there is a constant depending on the data such that*

$$\max_{0\le i\le N} \mathrm{E}\left[\left|X_{t_i} - X_{t_i}^{(\pi)}\right|^2\right] \le const.\,|\pi|, \quad \mathrm{E}\left[\left|\xi - y_T^{(\pi)}\right|^2\right] \le const.\,|\pi|.$$

Then there are constants C and c depending on the data such that

$$\begin{aligned}
&\sup_{0\le t\le T} \mathrm{E}\left[\left|Y_t - y_t^{(\pi)}\right|^2\right] + \int_0^T \mathrm{E}\left[\left|Z_t - z_t^{(\pi)}\right|^2\right] dt \\
&\quad \le C\left(\max_{1\le i\le N} \mathrm{E}\left[\left|\sum_{j=0}^{i-1}\left(y_{t_{j+1}}^{(\pi)} - y_{t_j}^{(\pi)} - f(t_j, X_{t_j}^{(\pi)}, y_{t_j}^{(\pi)}, z_{t_j}^{(\pi)})\Delta_j - z_{t_j}^{(\pi)}\Delta W_j\right)\right|^2\right]\right) \\
&\qquad + C\,|\pi|\left(\max_{0\le i\le N} \Delta_i \mathrm{E}\left[\left|y_{t_i}^{(\pi)}\right|^2\right] + \sum_{i=0}^{N-1} \Delta_i \mathrm{E}\left[\left|z_{t_i}^{(\pi)}\right|^2\right]\right) + C\,|\pi|
\end{aligned} \tag{5}$$

and

$$\begin{aligned}
&\max_{1\le i\le N} \mathrm{E}\left[\left|\sum_{j=0}^{i-1}\left(y_{t_{j+1}}^{(\pi)} - y_{t_j}^{(\pi)} - f(t_j, X_{t_j}^{(\pi)}, y_{t_j}^{(\pi)}, z_{t_j}^{(\pi)})\Delta_j - z_{t_j}^{(\pi)}\Delta W_j\right)\right|^2\right] \\
&\quad \le c\left(\sup_{0\le t\le T} \mathrm{E}\left[\left|Y_t - y_t^{(\pi)}\right|^2\right] + \int_0^T \mathrm{E}\left[\left|Z_t - z_t^{(\pi)}\right|^2\right] dt\right) + c\,|\pi|.
\end{aligned} \tag{6}$$

Remark 2.4. Under the assumptions of Theorem 2.3, let $(y^{(\pi_n)}, z^{(\pi_n)})$ be a sequence of approximations of (Y, Z) with $|\pi_n| \to 0$ for $n \to \infty$. Then

$$\sup_{0\le t\le T} \mathrm{E}\left[\left|Y_t - y_t^{(\pi_n)}\right|^2\right] + \int_0^T \mathrm{E}\left[\left|Z_t - z_t^{(\pi_n)}\right|^2\right] dt \xrightarrow[n\to\infty]{} 0$$

if and only if

$$\max_{1\le i\le N} \mathrm{E}\left[\left|\sum_{j=0}^{i-1}\left(y^{(\pi_n)}_{t_{j+1}} - y^{(\pi_n)}_{t_j} - f(t_j, X^{(\pi_n)}_{t_j}, y^{(\pi_n)}_{t_j}, z^{(\pi_n)}_{t_j})\Delta_j - z^{(\pi_n)}_{t_j}\Delta W_j\right)\right|^2\right] \xrightarrow[n\to\infty]{} 0$$

and

$$\max_{0\le i\le N} \mathrm{E}\left[\left|y^{(\pi_n)}_{t_i}\right|^2\right] + \sum_{i=0}^{N-1}\Delta_i \mathrm{E}\left[\left|z^{(\pi_n)}_{t_i}\right|^2\right] \le const.$$

(Here, we suppressed that t_i and N depend on n). In this case the convergence behaviour of the approximation error is equivalent to the convergence behaviour of the global error criterion up to a term of order $|\pi_n|$, which matches the usual squared time discretization error.

The above remark in mind, the essence of Theorem 2.3 is that the approximation error can be judged via the global error criterion. Contrary to the approximation error, the error criterion only depends on the known approximative solutions. Since the criterion is a maximum of expectations, it can be estimated by Monte Carlo simulations in practice. For more details, see Sections 3 and 5.

Proof of Theorem 2.3. We begin with the first inequality and define

$$\begin{aligned}\bar{Y}_t &= y^{(\pi)}_{t_0} + \int_0^t f(s, X^{(\pi)}_s, y^{(\pi)}_s, z^{(\pi)}_s)ds + \int_0^t z^{(\pi)}_s dW_s \\ &= \bar{Y}_T - \int_t^T f(s, X^{(\pi)}_s, y^{(\pi)}_s, z^{(\pi)}_s)ds - \int_t^T z^{(\pi)}_s dW_s.\end{aligned}$$

Employing Young's inequality we achieve

$$\begin{aligned}&\sup_{0\le t\le T} \mathrm{E}\left[\left|Y_t - y^{(\pi)}_t\right|^2\right] + \int_0^T \mathrm{E}\left[\left|Z_t - z^{(\pi)}_t\right|^2\right] dt \\ &\quad\le C \max_{0\le i\le N}\sup_{t_i\le t<t_{i+1}} \mathrm{E}\left[|Y_t - Y_{t_i}|^2\right] + C\max_{0\le i\le N}\mathrm{E}\left[\left|\bar{Y}_{t_i} - y^{(\pi)}_{t_i}\right|^2\right] \\ &\qquad\quad + C\left(\max_{0\le i\le N}\mathrm{E}\left[\left|Y_{t_i} - \bar{Y}_{t_i}\right|^2\right] + \int_0^T \mathrm{E}\left[\left|Z_t - z^{(\pi)}_t\right|^2\right] dt\right) \\ &\quad= (I) + (II) + (III).\end{aligned}$$

The Assumptions 2.1 allow us to apply the following result of Zhang [34], that concerns the L^2-regularity of X and Y:

$$\max_{0\le i\le N-1}\sup_{t\in(t_i,t_{i+1}]} \mathrm{E}\left[|X_t - X_{t_i}|^2 + |Y_t - Y_{t_i}|^2\right] \le C\,|\pi|\,. \tag{7}$$

Thus, $(I) \le C\,|\pi|$. Note that the pair $(\bar{Y}, z^{(\pi)})$ is the solution of the BSDE with terminal condition $\bar{Y}_T$ and generator $g(t,y,z) = f(t, X^{(\pi)}_t, y^{(\pi)}_t, z)$. Hence, standard a-priori-estimates on the difference of solutions of BSDEs (see e.g. El Karoui et al. [16], Section 2) yield

$$(III) \le C\left(\mathrm{E}\left[\left|Y_T - \bar{Y}_T\right|^2\right] + \int_0^T \mathrm{E}\left[\left|f(t, X_t, \bar{Y}_t, z^{(\pi)}_t) - f(s, X^{(\pi)}_t, y^{(\pi)}_t, z^{(\pi)}_t)\right|^2\right] dt\right).$$

Employing the Lipschitz condition of f together with the assumptions on the approximation error regarding X and the terminal condition ξ we obtain

$$
\begin{aligned}
(III) \leq C \Bigg(&\mathrm{E}\left[\left|Y_T - y_{t_N}^{(\pi)}\right|^2\right] + \mathrm{E}\left[\left|y_{t_N}^{(\pi)} - \bar{Y}_T\right|^2\right] + \int_0^T \mathrm{E}\left[\left|\bar{Y}_t - y_t^{(\pi)}\right|^2\right] dt \\
&+ \int_0^T \mathrm{E}\left[\left|X_t - X_t^{(\pi)}\right|^2\right] dt \Bigg) \\
\leq C \Bigg(&\mathrm{E}\left[\left|\bar{Y}_T - y_{t_N}^{(\pi)}\right|^2\right] + \sum_{i=0}^{N-1} \int_{\Delta_i} \mathrm{E}\left[\left|\bar{Y}_t - \bar{Y}_{t_i}\right|^2\right] dt + \sum_{i=0}^{N-1} \Delta_i \mathrm{E}\left[\left|\bar{Y}_{t_i} - y_{t_i}^{(\pi)}\right|^2\right] + |\pi| \Bigg) \\
\leq C &\max_{0 \leq i \leq N} \mathrm{E}\left[\left|\bar{Y}_{t_i} - y_{t_i}^{(\pi)}\right|^2\right] + C \sum_{i=0}^{N-1} \int_{\Delta_i} \mathrm{E}\left[\left|\bar{Y}_t - \bar{Y}_{t_i}\right|^2\right] dt + C\,|\pi|,
\end{aligned}
$$

where $\int_{\Delta_i}$ is an abbreviation for $\int_{t_i}^{t_{i+1}}$. Now we make use of the definition of $\bar{Y}$ and express the difference in the second summand by integrals.

$$
\begin{aligned}
&\sum_{i=0}^{N-1} \int_{\Delta_i} \mathrm{E}\left[\left|\bar{Y}_t - \bar{Y}_{t_i}\right|^2\right] dt \\
&= \sum_{i=0}^{N-1} \int_{\Delta_i} \mathrm{E}\left[\left|\int_{t_i}^t f(s, X_s^{(\pi)}, y_s^{(\pi)}, z_s^{(\pi)}) ds + \int_{t_i}^t z_s^{(\pi)} dW_s\right|^2\right] dt \\
&\leq 2 \sum_{i=0}^{N-1} \int_{\Delta_i} \mathrm{E}\left[\left|\int_{t_i}^t f(s, X_s^{(\pi)}, y_s^{(\pi)}, z_s^{(\pi)}) ds\right|^2\right] dt + 2 \sum_{i=0}^{N-1} \int_{\Delta_i} \int_{t_i}^t \mathrm{E}\left[\left|z_s^{(\pi)}\right|^2\right] ds dt \\
&\leq 2 \sum_{i=0}^{N-1} \int_{\Delta_i} \mathrm{E}\left[\left|\int_{t_i}^t f(s, X_s^{(\pi)}, y_s^{(\pi)}, z_s^{(\pi)}) ds\right|^2\right] dt + 2 \sum_{i=0}^{N-1} \Delta_i^2 \mathrm{E}\left[\left|z_{t_i}^{(\pi)}\right|^2\right].
\end{aligned}
\tag{8}
$$

We approximate the first summand of the right-hand side of (8) separately. Because of the Lipschitz condition of f and the boundedness of $f(s,0,0,0)$ we get

$$
\begin{aligned}
&\mathrm{E}\left[\left|\int_{t_i}^t f(s, X_s^{(\pi)}, y_s^{(\pi)}, z_s^{(\pi)}) ds\right|^2\right] \\
&\leq \Delta_i \int_{t_i}^t \mathrm{E}\left[\left(f(s, X_s^{(\pi)}, y_s^{(\pi)}, z_s^{(\pi)})\right)^2\right] ds \\
&\leq 2\Delta_i \int_{t_i}^t \mathrm{E}\left[\left(f(s, X_s^{(\pi)}, y_s^{(\pi)}, z_s^{(\pi)}) - f(s,0,0,0)\right)^2\right] ds + 2\Delta_i \int_{t_i}^t f^2(s,0,0,0) ds \\
&\leq C\Delta_i^2 \mathrm{E}\left[\left|X_{t_i}^{(\pi)}\right|^2 + \left|y_{t_i}^{(\pi)}\right|^2 + \left|z_{t_i}^{(\pi)}\right|^2\right] + 2\Delta_i^2 \sup_{0 \leq t \leq T} |f(t,0,0,0)|^2 \\
&\leq C\Delta_i^2 \left(\max_{0 \leq i \leq N} \mathrm{E}\left[\left|X_{t_i}^{(\pi)} - X_{t_i}\right|^2\right] + \mathrm{E}\left[\sup_{0 \leq t \leq T} |X_t|^2\right] + \mathrm{E}\left[\left|y_{t_i}^{(\pi)}\right|^2\right] + \mathrm{E}\left[\left|z_{t_i}^{(\pi)}\right|^2\right] \right) + C\Delta_i^2.
\end{aligned}
$$

As $\mathrm{E}\left[\sup_{0\le t\le T}|X_t|^2\right]$ is finite and the approximation error of X converges with order $|\pi|^{1/2}$, we can simplify as follows:

$$\mathrm{E}\left[\left|\int_{t_i}^{t} f(s,X_s^{(\pi)},y_s^{(\pi)},z_s^{(\pi)})ds\right|^2\right] \le C\Delta_i^2\left(\mathrm{E}\left[\left|y_{t_i}^{(\pi)}\right|^2\right]+\mathrm{E}\left[\left|z_{t_i}^{(\pi)}\right|^2\right]+1\right).$$

Applying this approximation on (8) we receive

$$\sum_{i=0}^{N-1}\int_{\Delta_i}\mathrm{E}\left[\left|\bar{Y}_t-\bar{Y}_{t_i}\right|^2\right]dt \le C\,|\pi|\left(\max_{0\le i\le N}\Delta_i\mathrm{E}\left[\left|y_{t_i}^{(\pi)}\right|^2\right]+\sum_{i=0}^{N-1}\Delta_i\mathrm{E}\left[\left|z_{t_i}^{(\pi)}\right|^2\right]+|\pi|\right).$$

Hence,

$$(III) \le C\max_{0\le i\le N}\mathrm{E}\left[\left|\bar{Y}_{t_i}-y_{t_i}^{(\pi)}\right|^2\right]+C\,|\pi|\left(\max_{0\le i\le N}\Delta_i\mathrm{E}\left[\left|y_{t_i}^{(\pi)}\right|^2\right]+\sum_{i=0}^{N-1}\Delta_i\mathrm{E}\left[\left|z_{t_i}^{(\pi)}\right|^2\right]\right)+C\,|\pi|\,.$$

The first summand of the last inequality matches with the term (II), which remains to be estimated. To this end, we replace $\bar{Y}_{t_i}$ by its definition and achieve

$$\begin{aligned}
(II) &= \max_{0\le i\le N}\mathrm{E}\left[\left|\sum_{j=0}^{i-1} y_{t_{j+1}}^{(\pi)}-y_{t_j}^{(\pi)}-\int_{t_j}^{t_{j+1}} f(s,X_{t_j}^{(\pi)},y_{t_j}^{(\pi)},z_{t_j}^{(\pi)})ds-z_{t_j}^{(\pi)}\Delta W_j\right|^2\right]\\
&\le C\max_{0\le i\le N}\mathrm{E}\left[\left|\sum_{j=0}^{i-1} y_{t_{j+1}}^{(\pi)}-y_{t_j}^{(\pi)}-f(t_j,X_{t_j}^{(\pi)},y_{t_j}^{(\pi)},z_{t_j}^{(\pi)})\Delta_j-z_{t_j}^{(\pi)}\Delta W_j\right|^2\right]\\
&\quad+C\max_{0\le i\le N}\mathrm{E}\left[\left|\sum_{j=0}^{i-1}\int_{t_j}^{t_{j+1}}\left(f(t_j,X_{t_j}^{(\pi)},y_{t_j}^{(\pi)},z_{t_j}^{(\pi)})-f(t,X_{t_j}^{(\pi)},y_{t_j}^{(\pi)},z_{t_j}^{(\pi)})\right)dt\right|^2\right]\\
&\le C\max_{0\le i\le N}\mathrm{E}\left[\left|\sum_{j=0}^{i-1} y_{t_{j+1}}^{(\pi)}-y_{t_j}^{(\pi)}-f(t_j,X_{t_j}^{(\pi)},y_{t_j}^{(\pi)},z_{t_j}^{(\pi)})\Delta_j-z_{t_j}^{(\pi)}\Delta W_j\right|^2\right]\\
&\quad+C\max_{0\le i\le N}\mathrm{E}\left[\left(\sum_{j=0}^{i-1}\int_{t_j}^{t_{j+1}}\sqrt{t-t_j}dt\right)^2\right]\\
&\le C\max_{0\le i\le N}\mathrm{E}\left[\left|\sum_{j=0}^{i-1} y_{t_{j+1}}^{(\pi)}-y_{t_j}^{(\pi)}-f(t_j,X_{t_j}^{(\pi)},y_{t_j}^{(\pi)},z_{t_j}^{(\pi)})\Delta_j-z_{t_j}^{(\pi)}\Delta W_j\right|^2\right]+C\,|\pi|\,.
\end{aligned}$$

Thus,

$$
\begin{aligned}
&\sup_{0\le t\le T} \mathrm{E}\left[\left|Y_t - y_t^{(\pi)}\right|^2\right] + \int_0^T \mathrm{E}\left[\left|Z_t - z_t^{(\pi)}\right|^2\right] dt \\
&\quad \le C \max_{0\le i\le N} \mathrm{E}\left[\left|\sum_{j=0}^{i-1} y_{t_{j+1}}^{(\pi)} - y_{t_j}^{(\pi)} - f(t_j, X_{t_j}^{(\pi)}, y_{t_j}^{(\pi)}, z_{t_j}^{(\pi)})\Delta_j - z_{t_j}^{(\pi)}\Delta W_j\right|^2\right] \\
&\qquad + C\left|\pi\right| \left(\max_{0\le i\le N} \Delta_i \mathrm{E}\left[\left|y_{t_i}^{(\pi)}\right|^2\right] + \sum_{i=0}^{N-1} \Delta_i \mathrm{E}\left[\left|z_{t_i}^{(\pi)}\right|^2\right]\right) + C\left|\pi\right|.
\end{aligned}
$$

For the second inequality we make use of the identity

$$
Y_{t_i} - Y_0 = \int_0^{t_i} f(t, X_t, Y_t, Z_t)dt + \int_0^{t_i} Z_t dW_t.
$$

After inserting this identity we obtain by the Itô isometry, Young's inequality and Jensen's inequality

$$
\begin{aligned}
&\max_{0\le i\le N} \mathrm{E}\left[\left|\sum_{j=0}^{i-1} y_{t_{j+1}}^{(\pi)} - y_{t_j}^{(\pi)} - f(t_j, X_{t_j}^{(\pi)}, y_{t_j}^{(\pi)}, z_{t_j}^{(\pi)})\Delta_j - z_{t_j}^{(\pi)}\Delta W_j\right|^2\right] \\
&\quad = \max_{0\le i\le N} \mathrm{E}\Bigg[\Bigg| \left(y_{t_i}^{(\pi)} - Y_{t_i}\right) + \left(Y_0 - y_{t_0}^{(\pi)}\right) \\
&\qquad + \sum_{j=0}^{i-1} \int_{t_j}^{t_{j+1}} \left(f(t, X_t, Y_t, Z_t) - f(t_j, X_{t_j}^{(\pi)}, y_{t_j}^{(\pi)}, z_{t_j}^{(\pi)})\right) dt + \int_0^{t_i} \left(Z_t - z_t^{(\pi)}\right) dW_t \Bigg|^2\Bigg] \\
&\quad \le c \Bigg(\max_{0\le i\le N} \mathrm{E}\left[\left|y_{t_i}^{(\pi)} - Y_{t_i}\right|^2\right] + \int_0^T \mathrm{E}\left[\left|Z_t - z_t^{(\pi)}\right|^2\right] dt \\
&\qquad\quad + \max_{0\le i\le N} \sum_{j=0}^{i-1} \int_{t_j}^{t_{j+1}} \mathrm{E}\left[\left|\left(f(t, X_t, Y_t, Z_t) - f(t_j, X_{t_j}^{(\pi)}, y_{t_j}^{(\pi)}, z_{t_j}^{(\pi)})\right)\right|^2\right] dt \Bigg).
\end{aligned}
$$

Due to the Lipschitz condition of f, we obtain

$$\begin{aligned}
&\max_{0\le i\le N} \mathrm{E}\left[\left|\sum_{j=0}^{i-1} y_{t_{j+1}}^{(\pi)} - y_{t_j}^{(\pi)} - f(t_j, X_{t_j}^{(\pi)}, y_{t_j}^{(\pi)}, z_{t_j}^{(\pi)})\Delta_j - z_{t_j}^{(\pi)}\Delta W_j\right|^2\right] \\
&\quad\le c \max_{0\le i\le N} \mathrm{E}\left[\left|Y_{t_i} - y_{t_i}^{(\pi)}\right|^2\right] + c\int_0^T \mathrm{E}\left[\left|Z_t - z_t^{(\pi)}\right|^2\right] dt \\
&\qquad + c\sum_{i=0}^{N-1}\int_{\Delta_i}\left(|t-t_i| + \mathrm{E}\left[\left|X_t - X_{t_i}^{(\pi)}\right|^2\right] + \mathrm{E}\left[\left|Y_t - y_{t_i}^{(\pi)}\right|^2\right] dt\right) \\
&\quad\le c\left(\max_{0\le i\le N} \mathrm{E}\left[\left|Y_{t_i} - y_{t_i}^{(\pi)}\right|^2\right] + \int_0^T \mathrm{E}\left[\left|Z_t - z_t^{(\pi)}\right|^2\right] dt\right) + c\max_{0\le i\le N}\mathrm{E}\left[\left|X_{t_i}^{(\pi)} - X_{t_i}\right|^2\right] \\
&\qquad + c\sum_{i=0}^{N-1}\int_{\Delta_i}\left(\mathrm{E}\left[|X_{t_i} - X_t|^2\right] + \mathrm{E}\left[|Y_{t_i} - Y_t|^2\right]\right) dt + c\,|\pi|\,.
\end{aligned}$$

The second summand is bounded by $c\,|\pi|$ on account of the assumptions of the theorem. For the third summand we receive the same upper bound due to (7). Therefore, the proof is completed. □

3 Application of the global error criterion

Many numerical schemes for BSDEs are based on a two-step procedure. In a first step a time discretization is performed pretending that conditional expectations can be computed in closed form, e.g. in Bouchard and Touzi [10] or Gobet et al. [18]. In a second step conditional expectations are replaced by some estimator. Different techniques, e.g. Malliavin Monte Carlo (see Bouchard and Touzi [10]), nonparametric regression (see Carriere [12]) or quantization (see Bally and Pagès [2] or Delarue and Menozzi [14]), have been applied in the literature to approximate conditional expectations. Here we will concentrate on the least squares Monte Carlo approach, which is popular in the context of American options pricing (see Longstaff and Schwartz [24]) and was applied to BSDEs in Gobet et al. [18], Lemor et al. [23] and Bender and Denk [5]. We first turn to the time discretization. Assuming that conditional expectations can be computed we define for $i = N-1, \ldots, 0$

$$\begin{aligned}
Y_{t_N}^{(\pi)} &= \xi^{(\pi)}, \\
Z_{d,t_i}^{(\pi)} &= \mathrm{E}^i\left[Y_{t_{i+1}}^{(\pi)}\frac{\Delta W_{d,i}}{\Delta_i}\right], \quad d = 1,\ldots,D, \\
Y_{t_i}^{(\pi)} &= \mathrm{E}^i\left[Y_{t_{i+1}}^{(\pi)} - f(t_i, X_{t_i}^{(\pi)}, Y_{t_{i+1}}^{(\pi)}, Z_{t_i}^{(\pi)})\Delta_i\right],
\end{aligned} \tag{9}$$

where $\mathrm{E}^i[\cdot]$ is the conditional expectation $\mathrm{E}^i[\cdot] = \mathrm{E}[\cdot \mid \mathcal{F}_{t_i}]$ and $X^{(\pi)}$ a corresponding approximation of X with respect to the partition π. This procedure is called backward scheme since it works iteratively backward through time and therefore requires the computation of nested conditional expectations. Note that it is explicit apart from the calculation of conditional expectations.

From now on we assume that

$$\max_{0\le i\le N} \mathrm{E}\left[\left|X_{t_i}-X^{(\pi)}_{t_i}\right|^2\right] \le const.\,|\pi|\,,\quad \mathrm{E}\left[\left|\xi-\xi^{(\pi)}\right|^2\right] \le const.\,|\pi|\,.$$

Referring to Zhang [34] and Gobet et al. [18], the approximation error due to time discretization is

$$\sup_{0\le t\le T} \mathrm{E}\left[\left|Y_t-Y^{(\pi)}_t\right|^2\right] + \mathrm{E}\left[\int_0^T \left|Z_t-Z^{(\pi)}_t\right|^2 dt\right] \le C\,|\pi|\,. \tag{10}$$

Hence, the global error criterion applied on the discretized solution $(Y^{(\pi)},Z^{(\pi)})$ converges to zero as well, see (6) in Theorem 2.3. Clearly,

$$\max_{1\le i\le N} \mathrm{E}\left[\left|\sum_{j=0}^{i-1}\left(Y^{(\pi)}_{t_{j+1}}-Y^{(\pi)}_{t_j}-f(t_j,X^{(\pi)}_{t_j},Y^{(\pi)}_{t_j},Z^{(\pi)}_{t_j})\Delta_j-Z^{(\pi)}_{t_j}\Delta W_j\right)\right|^2\right] \le C\,|\pi|\,.$$

In a next step, the conditional expectations $\mathrm{E}^i\left[\cdot\right]$ are replaced by their least squares Monte Carlo simulation based estimators. That means, that an estimation error concerning the conditional expectations will add to the time discretization error. Lemor et al. [23] have chosen an orthogonal projection on finite-dimensional subspaces of $L^2(\mathcal{F}_{t_i})$ as estimator for $\mathrm{E}^i\left[\cdot\right]$. Precisely, we denote by $\Lambda_{d,i}$, $0\le d\le D$ finite dimensional subspaces of $L^2(\mathcal{F}_{t_i})$ and by $P_{d,i}$ the orthogonal projections on these subspaces. The backward scheme reads then for all $i=N-1,\dots,0$ as follows:

$$\begin{aligned}
\hat{Y}^{(\pi)}_{t_N} &= \xi^{(\pi)},\\
\hat{Z}^{(\pi)}_{d,t_i} &= P_{d,i}\left[\hat{Y}^{(\pi)}_{t_{i+1}}\frac{\Delta W_{d,i}}{\Delta_i}\right],\quad d=1,\dots,D,\\
\hat{Y}^{(\pi)}_{t_i} &= P_{0,i}\left[\hat{Y}^{(\pi)}_{t_{i+1}}-f(t_i,X^{(\pi)}_{t_i},\hat{Y}^{(\pi)}_{t_{i+1}},\hat{Z}^{(\pi)}_{t_i})\Delta_i\right].
\end{aligned} \tag{11}$$

The estimators $(\hat{Y}^{(\pi)},\hat{Z}^{(\pi)})$ are extended by constant interpolation to RCLL processes. The approximation error in scheme (11) consists of the time discretization error and the estimation error, that is bounded by terms of certain projection errors. The latter one makes the difference between $(Y^{(\pi)},Z^{(\pi)})$ and $(\hat{Y}^{(\pi)},\hat{Z}^{(\pi)})$. Indeed, for $j=0,\dots,N-1$,

$$\begin{aligned}
&\max_{j\le i\le N} \mathrm{E}\left[\left|Y^{(\pi)}_{t_i}-\hat{Y}^{(\pi)}_{t_i}\right|^2\right] + \sum_{i=j}^{N-1}\Delta_i\mathrm{E}\left[\left|Z^{(\pi)}_{t_i}-\hat{Z}^{(\pi)}_{t_i}\right|^2\right]\\
&\le C\left(\sum_{i=j}^{N-1}\mathrm{E}\left[\left|P_{0,i}\left(Y^{(\pi)}_{t_i}\right)-Y^{(\pi)}_{t_i}\right|^2\right]+\sum_{d=1}^{D}\sum_{i=j}^{N-1}\Delta_i\mathrm{E}\left[\left|P_{d,i}\left(Z^{(\pi)}_{d,t_i}\right)-Z^{(\pi)}_{d,t_i}\right|^2\right]\right).
\end{aligned} \tag{12}$$

Combining the results in (10) and (12) we can derive an upper bound for the approximation error between (Y,Z) and $(\hat{Y}^{(\pi)},\hat{Z}^{(\pi)})$. The projection errors in the above inequality are in terms of the unknown solution $(Y^{(\pi)},Z^{(\pi)})$ of the time discretization scheme (9). Clearly, the errors are expressed by the error between the discretized solution and their unknown best projection on the subspaces $\Lambda_{d,i}$ for all $i = j, \dots, N-1$. Hence, the estimation error, that arises from the estimation of conditional expectations, is mainly driven by the choice of basis functions of $\Lambda_{d,i}$. The better the basis functions can describe the discretized solution, the smaller are the projection errors and therefore the estimation error as well. For the proof of a slightly different statement from (12), see Lemor et al. [23].

We now assume that the discretized BSDE is 'Markovian' in the sense that there is a multivariate Markov process $(\tilde{X}^{(\pi)}_{t_i})_{0\le i\le N}$ such that the first components consist of the discretized SDE $X^{(\pi)}$ and the approximative terminal condition is of the form $\xi^{(\pi)} = \phi^{(\pi)}(\tilde{X}^{(\pi)}_{t_N})$. This setting covers terminal conditions depending e.g. on the maximum $\max_{0\le t\le T} X_t$ or the average $\frac{1}{T}\int_0^T X_t dt$. The discretized solutions $Y^{(\pi)}_{t_i}$ and $Z^{(\pi)}_{t_i}$ are then deterministic functions of time and $\tilde{X}^{(\pi)}_{t_i}$. Consequently, we determine the projection subspaces $\Lambda_{d,i}$, $0 \le d \le D$ by function bases

$$\eta_{d,i} = \left\{ \eta^{d,i}_1(\tilde{X}^{(\pi)}_{t_i}), \dots, \eta^{d,i}_{K(d,i)}(\tilde{X}^{(\pi)}_{t_i}) \right\},$$

respectively. $K(d,i)$ denotes the dimension of the projection subspace $\Lambda_{d,i}$. Then the estimators $(\hat{Y}, \hat{Z})$ can be written as linear combination of deterministic functions that depend on $\tilde{X}^{(\pi)}$. Given the coefficients of the linear combination, we can calculate samples of the global error criterion just by drawing samples of the Markov process $\tilde{X}^{(\pi)}$ and of the increments of the Brownian motion ΔW_i. Therefore, we can estimate the global error criterion by Monte Carlo simulation. Thanks to the law of large numbers,

$$\begin{aligned}
&\lim_{M\to\infty} \max_{1\le i\le N} \frac{1}{M} \sum_{\mu=1}^{M} \left| \sum_{j=0}^{i-1} \left({}_\mu\hat{Y}^{(\pi)}_{t_{j+1}} - {}_\mu\hat{Y}^{(\pi)}_{t_j} - f(t_j, {}_\mu X^{(\pi)}_{t_j}, {}_\mu\hat{Y}^{(\pi)}_{t_j}, {}_\mu\hat{Z}^{(\pi)}_{t_j})\Delta_j - {}_\mu\hat{Z}^{(\pi)}_{t_j} \Delta\, {}_\mu W_j \right) \right|^2 \\
&= \max_{1\le i\le N} \mathrm{E}\left[\left| \sum_{j=0}^{i-1} \left(\hat{Y}^{(\pi)}_{t_{j+1}} - \hat{Y}^{(\pi)}_{t_j} - f(t_j, X^{(\pi)}_{t_j}, \hat{Y}^{(\pi)}_{t_j}, \hat{Z}^{(\pi)}_{t_j})\Delta_j - \hat{Z}^{(\pi)}_{t_j} \Delta W_j \right) \right|^2 \right],
\end{aligned}$$

where the lower-left index μ denotes the samples and M the sample size. Note that the Monte Carlo estimator is consistent, yet biased high since we draw the maximum of all timesteps. Nevertheless, we can judge the approximation error via the global error criterion, as the approximation error is bounded by a constant times the error criterion plus the time discretization error.

In practice, the coefficients of the linear combination of $\eta_{d,i}$ cannot be computed, but have to be estimated. Let $\alpha^{(\pi)}_{0,i} = \left(\alpha^{(\pi)}_{0,i,1}, \dots, \alpha^{(\pi)}_{0,i,K(0,i)}\right)^*$ and $\alpha^{(\pi)}_{d,i} = \left(\alpha^{(\pi)}_{d,i,1}, \dots, \alpha^{(\pi)}_{d,i,K(d,i)}\right)^*$, $1 \le d \le D$ be the coefficients related to the approximation of $\hat{Y}^{(\pi)}_{t_i}$ and $\hat{Z}^{(\pi)}_{d,t_i}$. They solve the following minimisation problems:

$$\min_{\alpha^{(\pi)}_{0,i}} \mathrm{E}\left[\left| \hat{Y}^{(\pi)}_{t_{i+1}} - f(t_i, X^{(\pi)}_{t_i}, \hat{Y}^{(\pi)}_{t_{i+1}}, \hat{Z}^{(\pi)}_{t_i})\Delta_i - \eta_{0,i}\alpha^{(\pi)}_{0,i} \right|^2 \right]$$

and for $d = 1, \dots, D$

$$\min_{\alpha^{(\pi)}_{d,i}} \mathrm{E}\left[\left| \frac{\Delta W_{d,i}}{\Delta_i} \hat{Y}^{(\pi)}_{t_{i+1}} - \eta_{d,i}\alpha^{(\pi)}_{d,i} \right|^2 \right].$$

One way to tackle these tasks is to estimate the above expectations with Monte Carlo simulations and to compute solutions of the least squares problems

$$\min_{\alpha_{0,i}^{(\pi,L)}} \frac{1}{L}\sum_{\lambda=1}^{L}\left|\,{}_{\lambda}\hat{Y}_{t_{i+1}}^{(\pi)} - f(t_i,\, {}_{\lambda}X_{t_i}^{(\pi)},\, {}_{\lambda}\hat{Y}_{t_{i+1}}^{(\pi)},\, {}_{\lambda}Z_{t_i}^{(\pi)})\Delta_i - \,{}_{\lambda}\eta_{0,i}\alpha_{0,i}^{(\pi,L)}\right|^2$$

and for $d = 1, \dots, D$

$$\min_{\alpha_{d,i}^{(\pi,L)}} \frac{1}{L}\sum_{\lambda=1}^{L}\left|\frac{\Delta\,{}_{\lambda}W_{d,i}}{\Delta_i}\,{}_{\lambda}\hat{Y}_{i+1}^{(\pi)} - \,{}_{\lambda}\eta_{d,i}\alpha_{d,i}^{(\pi,L)}\right|^2,$$

where the lower-left indicator λ marks the Monte Carlo samples of the random variables and L is the number of samples drawn for the Monte Carlo simulation. Then we receive the simulation based estimators

$$\hat{Y}_{t_i}^{(\pi,L)} = \sum_{k=1}^{K(0,i)} \alpha_{0,i,k}^{L}\eta_k^{0,i}(\tilde{X}_{t_i}^{(\pi)}), \qquad \hat{Z}_{d,t_i}^{(\pi,L)} = \sum_{k=1}^{\tilde{K}(d,i)} \alpha_{d,i,k}^{L}\eta_k^{d,i}(\tilde{X}_{t_i}^{(\pi)}). \tag{13}$$

The approximation error between (Y, Z) and $(\hat{Y}^{(\pi,L)}, \hat{Z}^{(\pi,L)})$, and hence the global error criterion as well, are determined by the time discretization error, the estimation error concerning the conditional expectations and a simulation error, which results from the simulation of the coefficients for the linear combination of the basis functions of the projection spaces. The time discretization error converges to zero with order $1/2$ as the number of timesteps grows to infinity. The analysis of the simulation error is rather intricate, but has already been done by Lemor et al. [23]. Their results can be applied to choose the number of simulated paths in dependence of the timesteps and the dimension of the function basis in a way that the simulation error matches the discretization error. They explain in detail, how the simulation error converges to zero, if, given the dimension of the function basis, the number of Monte Carlo samples are chosen suitably. Thus, both the time discretization error and the simulation error can be controlled and the choice of the function basis is the crucial step in designing the algorithm.

So, by and large, within the least-squares Monte Carlo backward scheme the global error judges the quality of the function basis. If the global error is not sufficiently small, we require an additional criterion for a more detailed analysis of the projection errors and their allocation over the period $[0, T]$.

4 Local criterion

We define the local error criterion by

$$\sum_{i=j}^{N-1} \mathrm{E}\left[\left|y_{t_{i+1}}^{(\pi)} - y_{t_i}^{(\pi)} - f(t_i, X_{t_i}^{(\pi)}, y_{t_i}^{(\pi)}, z_{t_i}^{(\pi)})\Delta_i - z_{t_i}^{(\pi)}\Delta W_i\right|^2\right].$$

The next theorem gives information about the relation between the local error criterion applied on the backward scheme and the projection errors.

Theorem 4.1. *Let Assumption 2.1 be fulfilled. Suppose further there exists a constant depending on the data such that*

$$\max_{0\le i\le N} \mathrm{E}\left[\left|X_{t_i} - X_{t_i}^{(\pi)}\right|^2\right] \le const. \left|\pi\right|, \quad \mathrm{E}\left[\left|\xi - \xi^{(\pi)}\right|^2\right] \le const. \left|\pi\right|.$$

Then there is a constant C depending on the data such that for every $j = 0, \ldots, N-1$

$$\begin{aligned}\sum_{i=j}^{N-1} \mathrm{E}\left[\left|P_{0,i}\left(Y_{t_i}^{(\pi)}\right) - Y_{t_i}^{(\pi)}\right|^2\right] &+ \sum_{d=1}^{D}\sum_{i=j}^{N-1} \Delta_i \mathrm{E}\left[\left|P_{d,i}\left(Z_{d,t_i}^{(\pi)}\right) - Z_{d,t_i}^{(\pi)}\right|^2\right] \\ &\ge C \sum_{i=j}^{N-1} \mathrm{E}\left[\left|\hat{Y}_{t_i}^{(\pi)} - \left(\hat{Y}_{t_{i+1}}^{(\pi)} - f(t_i, X_{t_i}^{(\pi)}, \hat{Y}_{t_i}^{(\pi)}, \hat{Z}_{t_i}^{(\pi)})\Delta_i\right) + \hat{Z}_{t_i}^{(\pi)}\Delta W_i\right|^2\right] - \left|\pi\right|.\end{aligned}$$

Theorem 4.1 provides a lower bound on the error between the time discretized solution and the unknown best approximation of the discretized solution in terms of the function basis. A large summand in the local error criterion suggests that the choice of the basis functions at this time step may be unsuccessful. In particular, for $i = N-1$ we get

$$\begin{aligned}\mathrm{E}&\left[\left|P_{0,N-1}\left(Y_{t_{N-1}}^{(\pi)}\right) - Y_{t_{N-1}}^{(\pi)}\right|^2\right] + \Delta_{N-1}\mathrm{E}\left[\left|P_{1,N-1}\left(Z_{t_{N-1}}^{(\pi)}\right) - Z_{t_{N-1}}^{(\pi)}\right|^2\right] \\ &\ge C(\mathrm{E}\left[\left|\hat{Y}_{t_{N-1}}^{(\pi)} - \left(\hat{Y}_{t_N}^{(\pi)} - f(t_{N-1}, X_{t_{N-1}}^{(\pi)}, \hat{Y}_{t_{N-1}}^{(\pi)}, \hat{Z}_{t_{N-1}}^{(\pi)})\Delta_{N-1}\right) + \hat{Z}_{t_{N-1}}^{(\pi)}\Delta W_{N-1}\right|^2\right]) - \left|\pi\right|.\end{aligned} \tag{14}$$

The following lemma is required for the proof of the theorem.

Lemma 4.2. *Under the assumptions of Theorem 4.1 there is a constant C depending on the data such that*

$$\sum_{i=0}^{N-1} \mathrm{E}\left[\left(Y_{t_i}^{(\pi)} - \left(Y_{t_{i+1}}^{(\pi)} - f(t_i, X_{t_i}^{(\pi)}, Y_{t_{i+1}}^{(\pi)}, Z_{t_i}^{(\pi)})\Delta_i\right) + Z_{t_i}^{(\pi)}\Delta W_i\right)^2\right] \le C\left|\pi\right|. \tag{15}$$

Proof. We define

$$\Delta f_i^{(\pi)}(s) = f(t_i, X_{t_i}^{(\pi)}, Y_{t_{i+1}}^{(\pi)}, Z_{t_i}^{(\pi)}) - f(s, X_s, Y_s, Z_s).$$

Step 1: We show

$$\mathrm{E}\left[\left|\int_{\Delta_i} \Delta f_i^{(\pi)}(s)ds\right|^2\right] \le C\left(\Delta_i^2\left|\pi\right| + \Delta_i^2\mathrm{E}\left[\left|Y_{t_{i+1}}^{(\pi)} - Y_{t_{i+1}}\right|^2\right] + \Delta_i\int_{\Delta_i}\mathrm{E}\left[\left|Z_{t_i}^{(\pi)} - Z_s\right|^2\right]ds\right). \tag{16}$$

Due to Jensen's inequality and the Lipschitz condition of f there is a generic constant C such that

$$
\begin{aligned}
&\mathrm{E}\left[\left|\int_{\Delta_i}\left(\Delta f_i^{(\pi)}(s)\right)ds\right|^2\right] \\
&\quad\le C\Delta_i^3 + C\Delta_i^2\left(\sup_{t_i\le s\le t_{i+1}}\mathrm{E}\left[|X_{t_i}-X_s|^2\right] + \sup_{t_i\le s\le t_{i+1}}\mathrm{E}\left[\left|Y_{t_{i+1}}-Y_s\right|^2\right]\right) \\
&\qquad + C\Delta_i^2\mathrm{E}\left[\left|X_{t_i}^{(\pi)}-X_{t_i}\right|^2\right] + C\Delta_i^2\mathrm{E}\left[\left|Y_{t_{i+1}}^{(\pi)}-Y_{t_{i+1}}\right|^2\right] + C\Delta_i\int_{\Delta_i}\mathrm{E}\left[\left|Z_{t_i}^{(\pi)}-Z_s\right|^2\right]ds.
\end{aligned}
\tag{17}
$$

Again the Assumptions 2.1 allow us to employ the L^2-regularity of X and Y, see inequality (7). Hence, we can approximate the second summand of the right-hand side of (17) by $C\Delta_i^2\,|\pi|$. Thanks to the assumptions in the present lemma we have for the third summand the estimation $C\Delta_i^2\,|\pi|$. Putting these results together we obtain (16).

Step 2: We will insert the equality

$$
Y_{t_{i+1}} - Y_{t_i} = \int_{\Delta_i} f(s,X_s,Y_s,Z_s)ds + \int_{\Delta_i} Z_s dW_s
$$

in the left-hand side of (15). Recall that

$$
\begin{aligned}
Y_{t_i} &= \mathrm{E}^i\Big[Y_{t_{i+1}} - \int_{\Delta_i} f(s,X_s,Y_s,Z_s)ds\Big], \\
Y_{t_i}^{(\pi)} &= \mathrm{E}^i\Big[Y_{t_{i+1}}^{(\pi)} - f(t_i,X_{t_i}^{(\pi)},Y_{t_{i+1}}^{(\pi)},Z_{t_i}^{(\pi)})\Delta_i\Big].
\end{aligned}
$$

The first equation arises from the formulation of the BSDE, the second from the backward scheme (9). Together with Young's inequality and the Itô isometry we get

$$
\begin{aligned}
&\sum_{i=0}^{N-1}\mathrm{E}\left[\left|Y_{t_i}^{(\pi)} - \left(Y_{t_{i+1}}^{(\pi)} - f(t_i,X_{t_i}^{(\pi)},Y_{t_{i+1}}^{(\pi)},Z_{t_i}^{(\pi)})\Delta_i\right) + Z_{t_i}^{(\pi)}\Delta W_i\right|^2\right] \\
&\quad\le 2\sum_{i=0}^{N-1}\mathrm{E}\left[\left|\mathrm{E}^i\Big[Y_{t_{i+1}}^{(\pi)} - Y_{t_{i+1}} - \int_{\Delta_i}\Delta f_i^{(\pi)}(s)ds\Big] - \left(Y_{t_{i+1}}^{(\pi)} - Y_{t_{i+1}} - \int_{\Delta_i}\Delta f_i^{(\pi)}(s)ds\right)\right|^2\right] \\
&\qquad + 2\sum_{i=0}^{N-1}\int_{\Delta_i}\mathrm{E}\left[\left|Z_{t_i}^{(\pi)} - Z_s\right|^2\right]ds.
\end{aligned}
$$

By working out the quadratic term of the first sum and recalling the definiton of $Y_{t_i}^{(\pi)}$, we obtain

$$\begin{aligned}
&\sum_{i=0}^{N-1} \mathrm{E}\left[\left|Y_{t_i}^{(\pi)} - \left(Y_{t_{i+1}}^{(\pi)} - f(t_i, X_{t_i}^{(\pi)}, Y_{t_{i+1}}^{(\pi)}, Z_{t_i}^{(\pi)})\Delta_i\right) + Z_{t_i}^{(\pi)}\Delta W_i\right|^2\right] \\
&\quad \leq 2\sum_{i=0}^{N-1} \mathrm{E}\left[\left|Y_{t_{i+1}}^{(\pi)} - Y_{t_{i+1}} - \int_{\Delta_i} \Delta f_i^{(\pi)}(s)ds\right|^2\right] \\
&\qquad - 2\sum_{i=0}^{N-1} \mathrm{E}\left[\mathrm{E}^i\left[Y_{t_{i+1}}^{(\pi)} - Y_{t_{i+1}} - \int_{\Delta_i} \Delta f_i^{(\pi)}(s)ds\right]^2\right] + 2\sum_{i=0}^{N-1}\int_{\Delta_i} \mathrm{E}\left[\left|Z_{t_i}^{(\pi)} - Z_s\right|^2\right] ds \\
&\quad \leq 2\sum_{i=0}^{N-1} \mathrm{E}\left[\left|Y_{t_{i+1}}^{(\pi)} - Y_{t_{i+1}} - \int_{\Delta_i} \Delta f_i^{(\pi)}(s)ds\right|^2\right] - 2\sum_{i=0}^{N-1} \mathrm{E}\left[\left|Y_{t_i}^{(\pi)} - Y_{t_i}\right|^2\right] \\
&\qquad + 2\sum_{i=0}^{N-1}\int_{\Delta_i} \mathrm{E}\left[\left|Z_{t_i}^{(\pi)} - Z_s\right|^2\right] ds.
\end{aligned}$$

Then Young's inequality leads to

$$\begin{aligned}
&\sum_{i=0}^{N-1} \mathrm{E}\left[\left|Y_{t_i}^{(\pi)} - \left(Y_{t_{i+1}}^{(\pi)} - f(t_i, X_{t_i}^{(\pi)}, Y_{t_{i+1}}^{(\pi)}, Z_{t_i}^{(\pi)})\Delta_i\right) + Z_{t_i}^{(\pi)}\Delta W_i\right|^2\right] \\
&\quad \leq 2\sum_{i=0}^{N-1} (1+\Delta_i)\,\mathrm{E}\left[\left|Y_{t_{i+1}}^{(\pi)} - Y_{t_{i+1}}\right|^2\right] + 2\sum_{i=0}^{N-1}\left(1+\frac{1}{\Delta_i}\right)\mathrm{E}\left[\left|\int_{\Delta_i} \Delta f_i^{(\pi)}(s)ds\right|^2\right] \\
&\qquad - 2\sum_{i=0}^{N-1} \mathrm{E}\left[\left|Y_{t_i}^{(\pi)} - Y_{t_i}\right|^2\right] + 2\sum_{i=0}^{N-1}\int_{\Delta_i} \mathrm{E}\left[\left|Z_{t_i}^{(\pi)} - Z_s\right|^2\right] ds \\
&\quad \leq C\mathrm{E}\left[\left|Y_{t_N}^{(\pi)} - Y_{t_N}\right|^2\right] + 2\sum_{i=0}^{N-1} \Delta_i \mathrm{E}\left[\left|Y_{t_i}^{(\pi)} - Y_{t_i}\right|^2\right] \\
&\qquad + 2\sum_{i=0}^{N-1}\left(1+\frac{1}{\Delta_i}\right)\mathrm{E}\left[\left|\int_{\Delta_i} \Delta f_i^{(\pi)}(s)ds\right|^2\right] + 2\sum_{i=0}^{N-1}\int_{\Delta_i} \mathrm{E}\left[\left|Z_{t_i}^{(\pi)} - Z_s\right|^2\right] ds.
\end{aligned}$$

Applying the assumption on the terminal condition and (16) we receive

$$\begin{aligned}
&\sum_{i=0}^{N-1} \mathrm{E}\left[\left|Y_{t_i}^{(\pi)} - \left(Y_{t_{i+1}}^{(\pi)} - f(t_i, X_{t_i}^{(\pi)}, Y_{t_{i+1}}^{(\pi)}, Z_{t_i}^{(\pi)})\Delta_i\right) + Z_{t_i}^{(\pi)}\Delta W_i\right|^2\right] \\
&\quad \leq C\left(\max_{0\leq i\leq N} \mathrm{E}\left[\left|Y_{t_i}^{(\pi)} - Y_{t_i}\right|^2\right] + \int_0^T \mathrm{E}\left[\left|Z_s^{(\pi)} - Z_s\right|^2\right] ds\right) + C\left|\pi\right|.
\end{aligned}$$

Employing (10) completes the proof. □

Proof of Theorem 4.1. Recall that within the backward scheme the generator f is applied on the vector $(t_i, X_{t_i}^{(\pi)}, Y_{t_{i+1}}^{(\pi)}, Z_{t_i}^{(\pi)})$ in the case of computable conditional expectations and on the vector

$(t_i, X_{t_i}^{(\pi)}, \hat{Y}_{t_{i+1}}^{(\pi)}, \hat{Z}_{t_i}^{(\pi)})$, when conditional expectations are estimated. Hence, we have to adapt the local criterion concerning the time points, at which the Y-processes are evaluated.

$$\begin{aligned}
&\sum_{i=j}^{N-1} \mathrm{E}\left[\left|\hat{Y}_{t_i}^{(\pi)} - \left(\hat{Y}_{t_{i+1}}^{(\pi)} - f(t_i, X_{t_i}^{(\pi)}, \hat{Y}_{t_i}^{(\pi)}, \hat{Z}_{t_i}^{(\pi)})\Delta_i\right) + \hat{Z}_{t_i}^{(\pi)}\Delta W_i\right|^2\right] \\
&\quad \leq 2\sum_{i=j}^{N-1} \mathrm{E}\left[\left|\hat{Y}_{t_i}^{(\pi)} - \left(\hat{Y}_{t_{i+1}}^{(\pi)} - f(t_i, X_{t_i}^{(\pi)}, \hat{Y}_{t_{i+1}}^{(\pi)}, \hat{Z}_{t_i}^{(\pi)})\Delta_i\right) + \hat{Z}_{t_i}^{(\pi)}\Delta W_i\right|^2\right] \\
&\qquad + 2\sum_{i=j}^{N-1} \Delta_i^2 \mathrm{E}\left[\left|f(t_i, X_{t_i}^{(\pi)}, \hat{Y}_{t_i}^{(\pi)}, \hat{Z}_{t_i}^{(\pi)}) - f(t_i, X_{t_i}^{(\pi)}, \hat{Y}_{t_{i+1}}^{(\pi)}, \hat{Z}_{t_i}^{(\pi)})\right|^2\right] \\
&\quad = (I) + (II).
\end{aligned}$$

Step 1: Thanks to the Lipschitz condition of f we receive for summand (II)

$$\begin{aligned}
(II) &\leq C\sum_{i=j}^{N-1} \Delta_i^2 \mathrm{E}\left[\left|\hat{Y}_{t_i}^{(\pi)} - \hat{Y}_{t_{i+1}}^{(\pi)}\right|^2\right] \\
&\leq C\sum_{i=j}^{N-1} \Delta_i^2 \mathrm{E}\left[\left|\hat{Y}_{t_i}^{(\pi)} - Y_{t_i}\right|^2 + \left|Y_{t_i} - Y_{t_{i+1}}\right|^2 + \left|Y_{t_{i+1}} - \hat{Y}_{t_{i+1}}^{(\pi)}\right| 2\right] \\
&\leq C\,|\pi| \sup_{t_j \leq t \leq T} \mathrm{E}\left[\left|\hat{Y}_t^{(\pi)} - Y_t\right|^2\right] + C\,|\pi| \max_{0 \leq i \leq N-1} \sup_{t \in (t_i, t_{i+1}]} \mathrm{E}\left[\left|Y_t - Y_{t_i}\right|^2\right].
\end{aligned}$$

The first summand can be estimated due to the approximation error, see (10) and (12), whereas we can apply on the second summand the L^2-regularity of Y (see (7)). Thus,

$$(II) \leq C\,|\pi| \left(\sum_{i=j}^{N-1} \mathrm{E}\left[\left|P_{0,i}\left(Y_{t_i}^{(\pi)}\right) - Y_{t_i}^{(\pi)}\right|^2\right] + \sum_{d=1}^{D}\sum_{i=j}^{N-1} \Delta_i \mathrm{E}\left[\left|P_{d,i}\left(Z_{d,t_i}^{(\pi)}\right) - Z_{d,t_i}^{(\pi)}\right|^2\right]\right) + C\,|\pi|^2.$$

Step 2: For the estimation of summand (I) define

$$\Delta f_i := f(t_i, X_{t_i}^{(\pi)}, Y_{t_{i+1}}^{(\pi)}, Z_{t_i}^{(\pi)}) - f(t_i, X_{t_i}^{(\pi)}, \hat{Y}_{t_{i+1}}^{(\pi)}, \hat{Z}_{t_i}^{(\pi)})$$

and

$$\eta_i := \hat{Y}_{t_{i+1}}^{(\pi)} - Y_{t_{i+1}}^{(\pi)} + \Delta f_i \Delta_i.$$

The orthogonal projections $P_{0,i}$ are maps from $L^2(\mathcal{F}_T)$ into a subspace $\Lambda_{0,i}$ of $L^2(\mathcal{F}_{t_i})$. We have,

$$\begin{aligned}
P_{0,i}\left(Y_{t_i}^{(\pi)}\right) &= P_{0,i}\left(\mathrm{E}^i\left[Y_{t_{i+1}}^{(\pi)} - f(t_i, X_{t_i}^{(\pi)}, Y_{t_{i+1}}^{(\pi)}, Z_{t_i}^{(\pi)})\Delta_i\right]\right) \\
&= P_{0,i}\left(Y_{t_{i+1}}^{(\pi)} - f(t_i, X_{t_i}^{(\pi)}, Y_{t_{i+1}}^{(\pi)}, Z_{t_i}^{(\pi)})\Delta_i\right).
\end{aligned} \tag{18}$$

We use the relation in (18) to add a zero and then employ Young's inequality.

$$
\begin{aligned}
(I) = 2\sum_{i=j}^{N-1} \mathrm{E}\Bigg[\bigg|\Big[P_{0,i}\left(\hat{Y}^{(\pi)}_{t_{i+1}} - Y^{(\pi)}_{t_{i+1}} + \Delta f_i \Delta_i\right) - \left(\hat{Y}^{(\pi)}_{t_{i+1}} - Y^{(\pi)}_{t_{i+1}} + \Delta f_i \Delta_i\right)\Big] \\
+ \Big[P_{0,i}\left(Y^{(\pi)}_{t_i}\right) - Y^{(\pi)}_{t_i}\Big] + \Big[Y^{(\pi)}_{t_i} - Y^{(\pi)}_{t_{i+1}} + f(t_i, X^{(\pi)}_{t_i}, Y^{(\pi)}_{t_{i+1}}, Z^{(\pi)}_{t_i})\Delta_i + Z^{(\pi)}_{t_i}\Delta W_i\Big] \\
+ \Big[\hat{Z}^{(\pi)}_{t_i}\Delta W_i - Z^{(\pi)}_{t_i}\Delta W_i\Big]\bigg|^2\Bigg] \\
\le 8\sum_{i=j}^{N-1} \mathrm{E}\left[|P_{0,i}(\eta_i) - \eta_i|^2\right] \\
+ 8\sum_{i=j}^{N-1} \mathrm{E}\left[\left|Y^{(\pi)}_{t_i} - Y^{(\pi)}_{t_{i+1}} + f(t_i, X^{(\pi)}_{t_i}, Y^{(\pi)}_{t_{i+1}}, Z^{(\pi)}_{t_i})\Delta_i + Z^{(\pi)}_{t_i}\Delta W_i\right|^2\right] \\
+ 8\sum_{i=j}^{N-1} \mathrm{E}\left[\left|P_{0,i}\left(Y^{(\pi)}_{t_i}\right) - Y^{(\pi)}_{t_i}\right|^2\right] + 8\sum_{i=j}^{N-1} \mathrm{E}\left[\left|\left(\hat{Z}^{(\pi)}_{t_i} - Z^{(\pi)}_{t_i}\right)\Delta W_i\right|^2\right].
\end{aligned}
\tag{19}
$$

First we take a closer look at the first summand of the right-hand side of (19). Thanks to the definitions of $\hat{Y}^{(\pi)}_{t_i}$ and $Y^{(\pi)}_{t_i}$ the following equality holds true for all $i = 0, \ldots, N-2$:

$$
\eta_i = P_{0,i+1}(\eta_{i+1}) + P_{0,i+1}\left(Y^{(\pi)}_{t_{i+1}}\right) - Y^{(\pi)}_{t_{i+1}} + \Delta f_i \Delta_i. \tag{20}
$$

Due to the orthogonality of P_i we have

$$
\mathrm{E}\left[P_{0,i}(\eta_i)\,\eta_i\right] = \mathrm{E}\left[(P_{0,i}(\eta_i))^2\right]
$$

and consequently

$$
\sum_{i=j}^{N-1} \mathrm{E}\left[|P_{0,i}(\eta_i) - \eta_i|^2\right] = \sum_{i=j}^{N-1} \mathrm{E}\left[|\eta_i|^2\right] - \mathrm{E}\left[|P_{0,i}(\eta_i)|^2\right].
$$

The following calculation takes place in view of (20), the orthogonality of the projections and the equality $\hat{Y}^{(\pi)}_{t_N} - Y^{(\pi)}_{t_N} = 0$.

$$
\begin{aligned}
&\sum_{i=j}^{N-1} \mathrm{E}\left[\left|P_{0,i}\left(\eta_i\right) - \eta_i\right|^2\right] \\
&\quad \leq \sum_{i=j}^{N-2} \left(1+\Delta_i\right) \mathrm{E}\left[\left|P_{0,i+1}\left(\eta_{i+1}\right)\right|^2\right] + \sum_{i=j}^{N-2} \left(1+\Delta_i\right) \mathrm{E}\left[\left|P_{0,i+1}\left(Y^{(\pi)}_{t_{i+1}}\right) - Y^{(\pi)}_{t_{i+1}}\right|^2\right] \\
&\qquad + \left(1+\Delta_{N-1}\right) \mathrm{E}\left[\left|\hat{Y}^{(\pi)}_{t_N} - Y^{(\pi)}_{t_N}\right|^2\right] + \sum_{i=j}^{N-1} \left(1+\Delta_i\right) \Delta_i \mathrm{E}\left[\left|\Delta f_i\right|^2\right] - \sum_{i=j}^{N-1} \mathrm{E}\left[\left|P_{0,i}\left(\eta_i\right)\right|^2\right] \\
&\quad \leq C \sum_{i=j}^{N-2} \Delta_i \mathrm{E}\left[\left|\hat{Y}^{(\pi)}_{t_{i+2}} - Y^{(\pi)}_{t_{i+2}}\right|^2\right] + \sum_{i=j}^{N-2} \left(1+\Delta_i\right) \mathrm{E}\left[\left|P_{0,i+1}\left(Y^{(\pi)}_{t_{i+1}}\right) - Y^{(\pi)}_{t_{i+1}}\right|^2\right] \\
&\qquad + C \sum_{i=j}^{N-1} \Delta_i \mathrm{E}\left[\left|\Delta f_i\right|^2\right].
\end{aligned}
$$

Because of the Lipschitz condition of f we get

$$
\begin{aligned}
\sum_{i=j}^{N-1} \mathrm{E}\left[\left|P_{0,i}\left(\eta_i\right) - \eta_i\right|^2\right] \leq C &\left(\max_{j\leq i\leq N} \mathrm{E}\left[\left|\hat{Y}^{(\pi)}_{t_i} - Y^{(\pi)}_{t_i}\right|^2\right] + \sum_{i=j}^{N-1} \Delta_i \mathrm{E}\left[\left|\hat{Z}^{(\pi)}_{t_i} - Z^{(\pi)}_{t_i}\right|^2\right]\right) \\
&+ C \sum_{i=j}^{N-2} \mathrm{E}\left[\left|P_{0,i+1}\left(Y^{(\pi)}_{t_{i+1}}\right) - Y^{(\pi)}_{t_{i+1}}\right|^2\right].
\end{aligned}
\tag{21}
$$

Applying result (21) on (19) we achieve

$$
\begin{aligned}
(I) \leq C &\left(\max_{j\leq i\leq N} \mathrm{E}\left[\left|\hat{Y}^{(\pi)}_{t_i} - Y^{(\pi)}_{t_i}\right|^2\right] + \sum_{i=j}^{N-1} \Delta_i \mathrm{E}\left[\left|\hat{Z}^{(\pi)}_{t_i} - Z^{(\pi)}_{t_i}\right|^2\right]\right) \\
&+ C \sum_{i=j}^{N-1} \mathrm{E}\left[\left|P_{0,i}\left(Y^{(\pi)}_{t_i}\right) - Y^{(\pi)}_{t_i}\right|^2\right] \\
&+ C \sum_{i=j}^{N-1} \mathrm{E}\left[\left|Y^{(\pi)}_{t_i} - Y^{(\pi)}_{t_{i+1}} + f(t_i, X^{(\pi)}_{t_i}, Y^{(\pi)}_{t_{i+1}}, Z^{(\pi)}_{t_i})\Delta_i + Z^{(\pi)}_{t_i}\Delta W_i\right|^2\right].
\end{aligned}
$$

By employing Lemma 4.2 and inequality (12) we obtain

$$
(I) \leq C \left(\sum_{i=j}^{N-1} \mathrm{E}\left[\left|P_{0,i}\left(Y^{(\pi)}_{t_i}\right) - Y^{(\pi)}_{t_i}\right|^2\right] + \sum_{d=1}^{D} \sum_{i=j}^{N-1} \Delta_i \mathrm{E}\left[\left|P_{d,i}\left(Z^{(\pi)}_{d,t_i}\right) - Z^{(\pi)}_{d,t_i}\right|^2\right]\right) + C\left|\pi\right|.
$$

Summarizing the results on (I) and (II) completes the proof. □

5 Simulations

In this section we illustrate the global and local error criteria by a numerical example. It concerns a call-spread option within a money market with different interest rates for borrowing and lending and was suggested in Lemor et al. [23]. The rate for borrowing will be denoted by R, the rate for lending by r. The FBSDE reads then

$$\begin{aligned} X_t &= x + \int_0^t \mu X_s ds + \int_0^t \sigma X_s dW_s \\ Y_t &= \xi - \int_t^T \left(rY_s + \frac{(\mu - r)}{\sigma} Z_s - (R-r)\left(Y_t - \frac{Z_t}{\sigma}\right)_- \right) ds - \int_t^T Z_s dW_s, \\ \xi &= (X_T - K_1)_+ - 2\,(X_T - K_2)_+\,, \end{aligned} \tag{22}$$

see also Bergman [9]. As parameters we choose $x = 100$, $r = 0.01$, $R = 0.06$, $\mu = 0.05$, $\sigma = 0.2$ and $T = 0.25$. The strike prices are fixed with $K_1 = 95$ and $K_2 = 105$. As basis functions we take the payoff function together with indicator functions:

$$\begin{aligned} & e_1(X_{t_i}^{(\pi)}) = (X_{t_i}^{(\pi)} - K_1)_+ - 2(X_{t_i}^{(\pi)} - K_2)_+, \quad e_2(X_{t_i}^{(\pi)}) = \mathbb{1}_{[0,l]}(X_{t_i}^{(\pi)}), \\ & e_3(X_{t_i}^{(\pi)}) = \mathbb{1}_{[u,\infty)}(X_{t_i}^{(\pi)}), \\ & e_k(X_{t_i}^{(\pi)}) = \mathbb{1}_{[l+(k-4)(u-l)/(\kappa-3),\,l+(k-3)(u-l)/(\kappa-3)]}(X_{t_i}^{(\pi)}), \quad 4 \le k \le \kappa, \end{aligned}$$

where κ is the dimension of the function basis of $\Lambda_{0,i} = \Lambda_{1,i}$ and l and u are defined as the empirical mean of the samples of $X_{t_i}^{(\pi)}$ minus and plus the empirical standard deviation. The backward scheme explained in Section 3 is repeated for an increasing number of timesteps, samples and dimension of function basis, i.e. we set for $j = 1, \ldots, 9$ and $\beta = 3, 4, 5$

$$N = \left[2 * \sqrt{2}^{(j-1)} \right], \qquad \kappa = \left[\frac{14}{5} \sqrt{2}^{j-1} \right] + 2, \qquad L = M = \left[2\sqrt{2}^{\beta(j-1)} \right],$$

where $[a]$ is the closest integer to a. Recall that L is the number of samples required for the Monte Carlo simulation based least squares estimation of the projection coefficients and M is the number of samples necessary for the Monte Carlo simulation of the error criteria. The following tabular states the values for L considering the different values for β and j:

	j								
β	1	2	3	4	5	6	7	8	9
3	2	6	16	45	128	362	1,024	2,896	8,192
4	2	8	32	128	512	2,048	8,192	32,768	131,072
5	2	11	64	362	2,048	11,585	65,536	370,728	2,097,152

In Figure 1 the global error criterion is plotted on a logarithmic scale against the number of timesteps N for $\beta = 3, 4, 5$. We observe that the global error decreases for all β, when N and the sample sizes L and M are growing simultaneously. Moreover, increasing β from 3 to 5 significantly reduces the global error criterion. We observe, however, that even in the case $\beta = 5$ the global error is, despite the large simulation costs, still larger than one. A further error reduction requires a finer time grid and hence increasing of j, which in turn leads to tremendous cost.

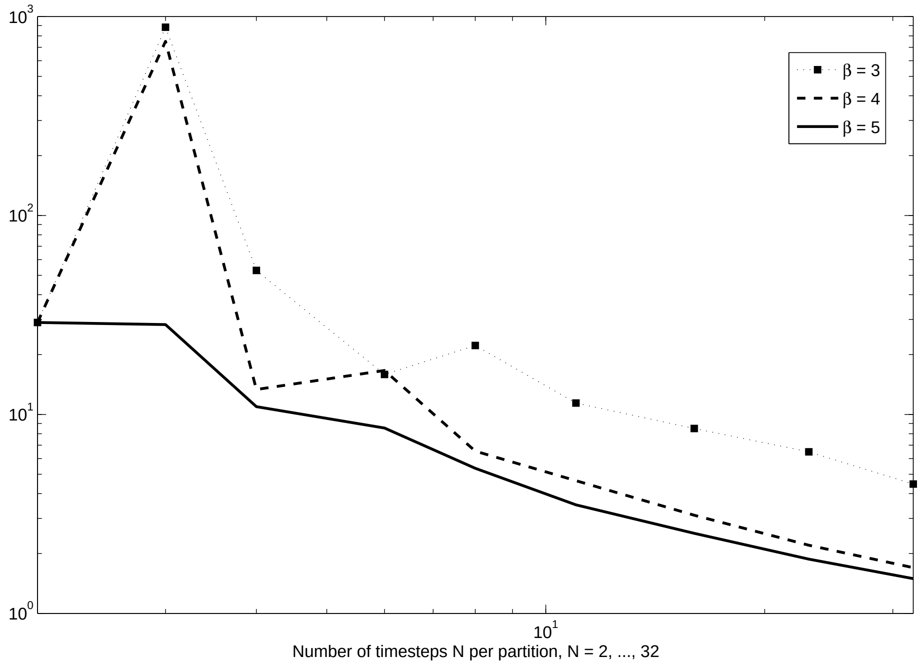

Figure 1: Development of the global error criterion for a call spread option

In Figure 2 we examine the case $j = 9$, where we have according to the above definition 32 timesteps. The figure shows how much every single timestep contributes to the local error criterion for different β. The case $\beta = 5$, which yields the lowest local errors, is presented by the lowest solid line.

For all β the local error related to each single timestep increases from timestep $t_0 = 0$ to $t_N = T$. Recalling that the approximation scheme works backward through time, it strikes that the first step of the algorithm is responsible for the highest local error per timestep. That means in view of (14) that high projection errors already occur in the first step. This observation suggests that the choice of the basis does not even fit well to the terminal condition.

In addition to this we present in Figure 3 the approximation of $\hat{Y}_{t_0}^{(\pi)}$. As one can see, good convergence results for $\hat{Y}_{t_0}^{(\pi)}$ are obtained already at relatively low cost ($\beta = 3, j = 9$). Higher values for β boost the number of samples and thereby the computational costs, but do not improve the estimation of Y_0 significantly. When looking at Figure 3 it is tempting to conclude that a good approximation is achieved for $\beta = 3$. However, an eximanation of the gobal error criterion in Figure 1 shows that it is not sufficient to set $\beta = 3$ for satisfactory results as far as one is interested in an approximation of the whole solution processes (Y, Z).

In conclusion, we observe that theoretically the backward scheme with the indicator basis converges thanks to the results in Gobet et al. [18], but the computational cost may become prohibitive, if one is interested in a good approximation of the solution processes (Y, Z) and not in the option price Y_0 only.

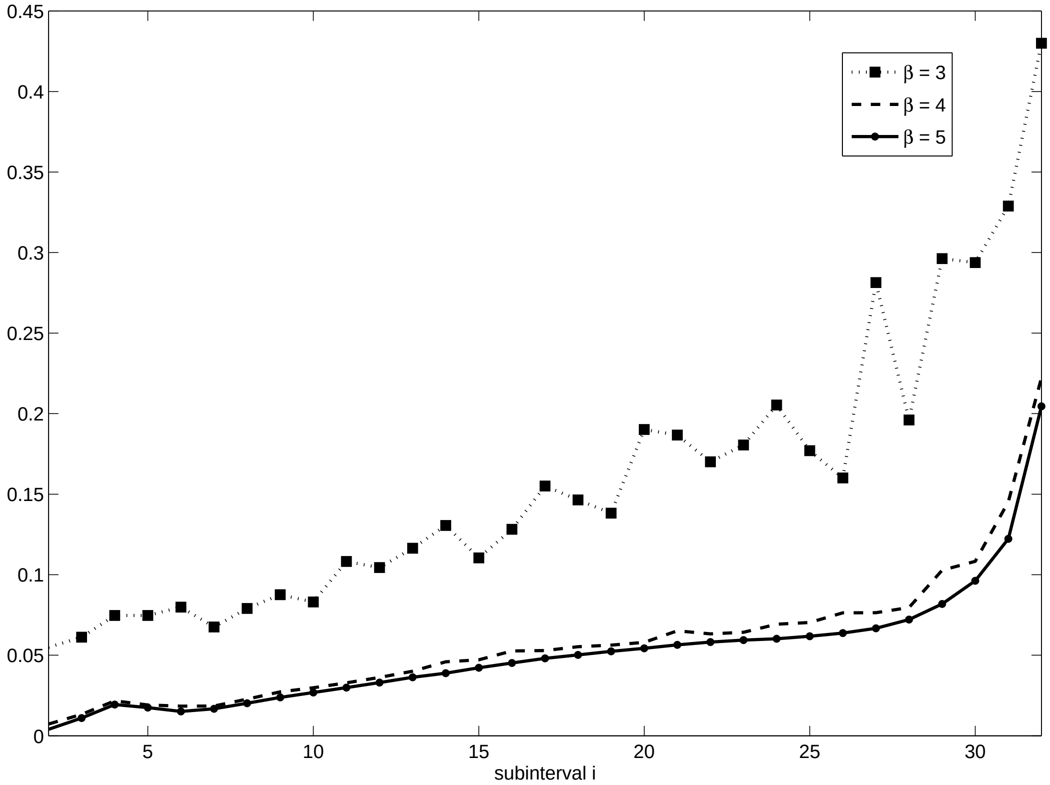

Figure 2: Contribution of the single timesteps to the local error criterion for a call spread option

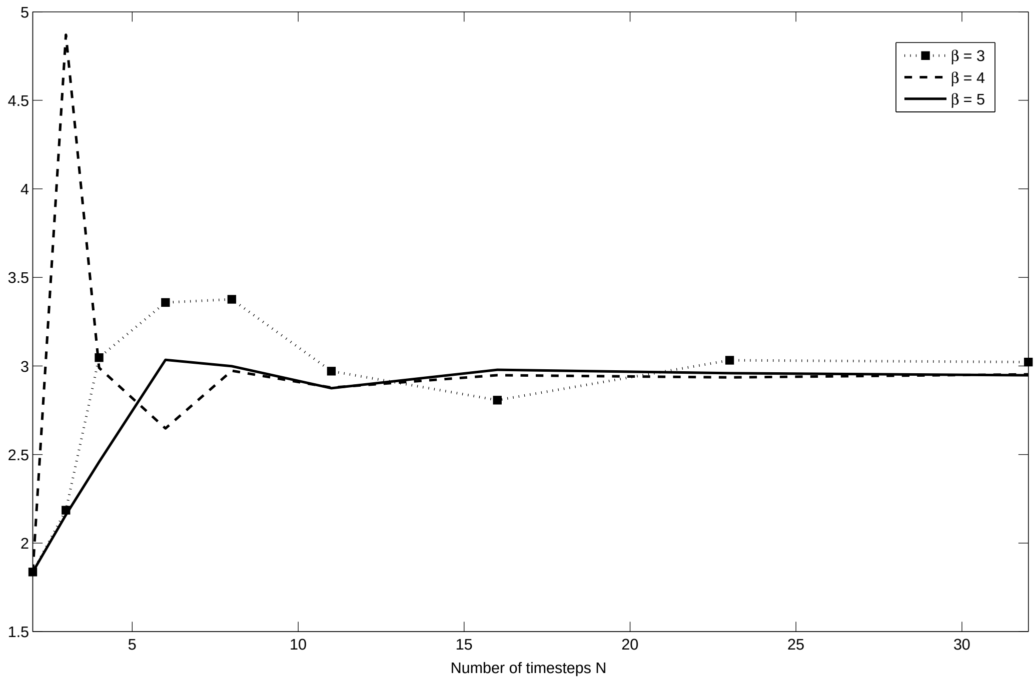

Figure 3: Development of $\hat{Y}_{t_0}^{(\pi)}$ for a call spread option

6 Non-linear control variates

In this section we propose a method for reducing the computational cost within the backward scheme under suitable assumptions. Precisely, we propose to split the original BSDE into the sum of two BSDEs and assume that one of them can be solved in closed form and only the other one needs to be approximated numerically. We call this procedure non-linear control variate as it has a flavor of the control variate technique for simulating expectations (i.e. the case of a linear BSDE). Suppose the original BSDE is given by

$$\begin{aligned} X_t &= x + \int_0^t b(s, X_s) ds + \int_0^t \sigma(s, X_s) dW_s \\ Y_t &= \phi(X_T) - \int_t^T f(s, X_s, Y_s, Z_s) ds - \int_t^T Z_s dW_s. \end{aligned} \tag{23}$$

Instead of (23), we examine the following FBSDEs:

$$\begin{aligned} X_t &= x + \int_0^t b(s, X_s) ds + \int_0^t \sigma(s, X_s) dW_s \\ \tilde{Y}_t &= \phi(X_T) - \int_t^T \tilde{Z}_s dW_s, \\ y_t &= -\int_t^T f(s, X_s, y_s + \tilde{Y}_s, z_s + \tilde{Z}_s) ds - \int_t^T z_s dW_s. \end{aligned} \tag{24}$$

Hence, we get the solution (Y, Z) of (23) by adding $(\tilde{Y}, \tilde{Z})$ and (y, z). Note that Gobet and Makhlouf [19] made use of this decomposition in their proof of the L^2-regularity of Z in cases of irregular terminal conditions. In many examples, closed form solutions or very accurate approximations of $(\tilde{Y}, \tilde{Z})$ are available (this BSDE corresponds to a linear option pricing problem). As far as (y, z) are regarded we employ the algorithm proposed in Section 3. The simulation based estimators are then defined by

$$\hat{Y}_{t_i}^{(\pi,L,\diamond)} = \hat{y}_{t_i}^{(\pi,L)} + \tilde{Y}_{t_i}, \qquad \hat{Z}_{t_i}^{(\pi,L,\diamond)} = \hat{z}_{t_i}^{(\pi,L,)} + \tilde{Z}_{t_i}.$$

In some non-linear option pricing problems (e.g. the problem with different interest rates) the non-linearity is typically 'small' compared to the terminal condition. So, heuristically, the 'main' part of the solution $(\tilde{Y}, \tilde{Z})$ is correctly computed and only a small part, here (y, z), is affected by approximation errors.

We illustrate now the proposed procedure at the case of the call-spread option introduced in Section 5. As function basis we fix the call-spread payoff and monomials, i. e.

$$e_1(X_{t_i}^{(\pi)}) = (X_{t_i}^{(\pi)} - K_1)_+ - 2(X_{t_i}^{(\pi)} - K_2)_+, \quad e_k(X_{t_i}^{(\pi)}) = (X_{t_i}^{(\pi)} - X_0)^{k-2}, \quad 2 \le k \le 6.$$

The sample sizes are $L = M = 200,000$ and the simulation is repeated for $N = 20, 25, \ldots, 200$, where N is the number of timesteps. The next figures show the simulation results of our call-spread option. First, we run the backward scheme to get $(\hat{Y}, \hat{Z})$. In addition to that, we use non-linear control variates to generate $(\hat{Y}^{(\pi,L,\diamond)}, \hat{Z}^{(\pi,L,\diamond)})$.

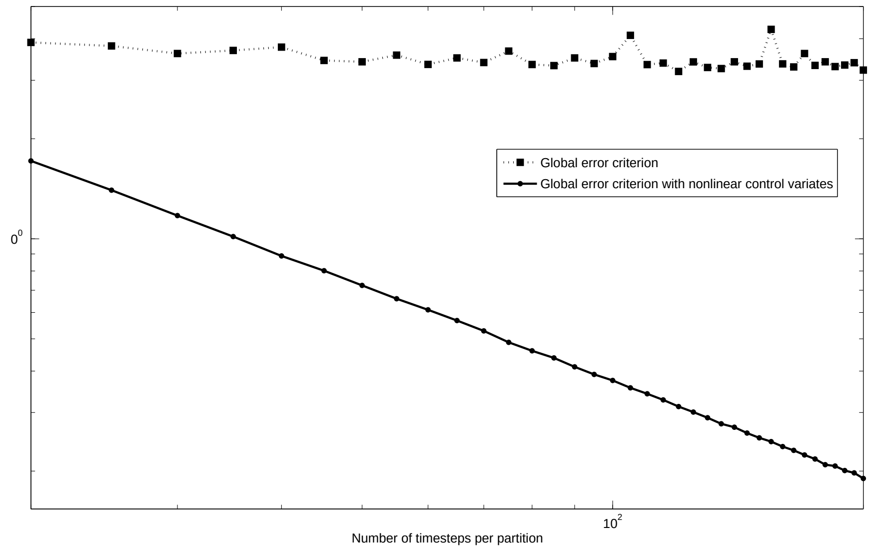

Figure 4: Development of the global error criterion for a call-spread option

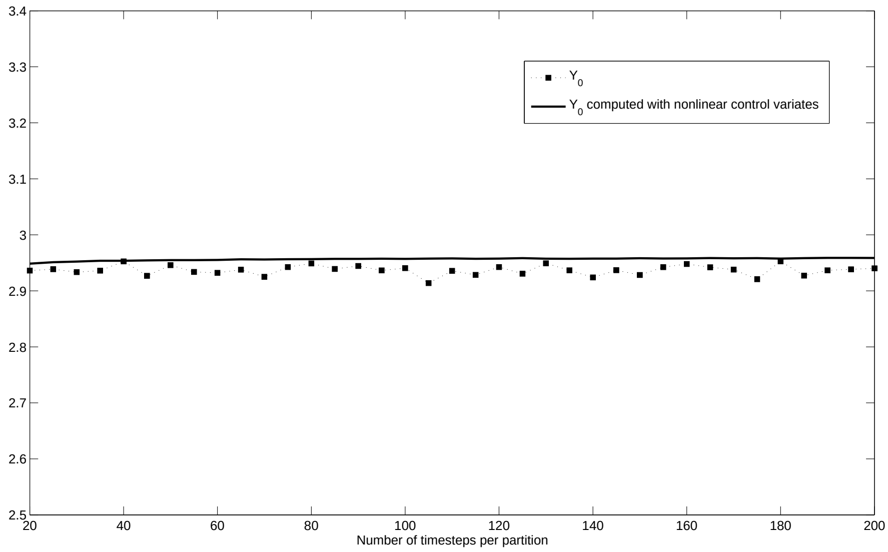

Figure 5: Development of $\hat{Y}^{(\pi)}_{t_0}$ for a call-spread option

Figure 4 compares the global error criterion for the original backward scheme and for the modified one using non-linear control variates. In the original algorithm the global error roughly stays constant as a function of the number of time steps. The reason is that the error in estimating the conditional expectation with the simple basis functions and a moderate sample size dominates the time discretization error. Contrarily, with non-linear control variates the global error decreases almost with the best possible rate of order N^{-1} in the number of time steps (the slope of the line in the logarithmic plot is -0.96). One also observes that the absolute size of the global error with non-linear control variates is smaller by a factor 10 compared to the results in Section 5, which were computed with much higher simulation cost.

In contrast to the global error, there is only a small difference between the estimation of $\hat{Y}_{t_0}^{(\pi)}$ by the original algorithm on the one hand and with non-linear control variates on the other hand. Since the global error in the latter case is smaller than in the case of the original algorithm, Figure 5 suggests that $\hat{Y}_{t_0}^{(\pi)}$ is slightly underestimated without control variates. In accordance with the simulation results in Section 5, we find that the option price Y_0 is approximated well in all settings which we implemented, even if the approximation of the whole solution (Y_t, Z_t) is not successful.

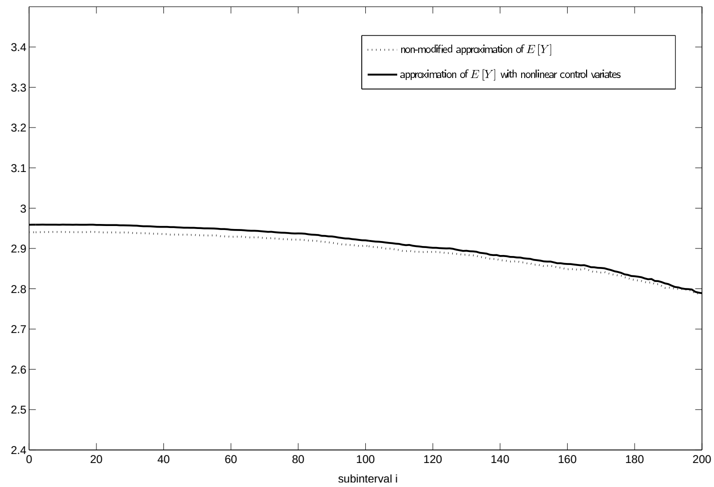

Figure 6: Approximation of $\mathrm{E}[Y_t]$ by unchanged backward scheme and with non-linear control variates

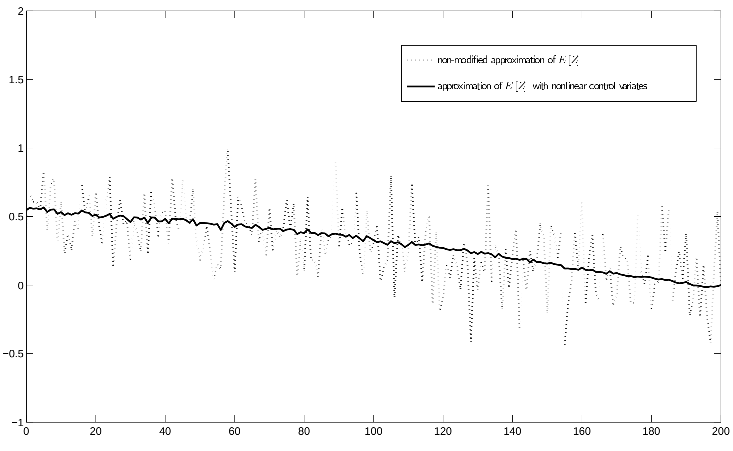

Figure 7: Approximation of $\mathrm{E}[Z_t]$ by unchanged backward scheme and with non-linear control variates

The Figures 6 and 7 clarify the effect of introducing control variates. These figures show the Monte Carlo approximations of $\mathrm{E}[Y_t]$ and $\mathrm{E}[Z_t]$ computed with and without non-linear control variates. All in all, the difference between both approximation results for $\mathrm{E}[Y_t]$ is very small. But when turning to the estimation of $\mathrm{E}[Z_t]$, we observe that the estimation without control variates is oszillating very strongly, whereas the improved algorithm reduces these oszillations significantly.

7 Conclusion

In this paper we have introduced two error criteria in order to judge the quality of numerical solutions of BSDEs. The global error criterion is basically equivalent to the approximation error between the numerical solution and the unknown true solution. The local error criterion provides bounds on the quality of the function basis in a least squares Monte Carlo framework for estimating conditional expectations. Numerical experiments illustrate that the initial value Y_0 may be estimated reasonably within the backward scheme, although the results of the local error criterion indicate high projection errors and therefore a bad choice of basis functions. Hence, for sufficient results on the approximation error of the entire processes Y and Z we have to improve the estimation of conditional expectations. Two observations might help us to tackle this problem. On the one hand it turned out that the projection errors are large especially at the terminal timestep of the partition. An interesting question is therefore, which factors influence the projection errors in this first step of the algorithm. Answering this question would improve the local error criterion related to the entire period, since the projection errors propagate in time. This indicates that the choice of the basis functions should be adjusted to the terminal condition, which we will discuss in detail in future research. On the other hand it became evident that the approximation of the process Z may strongly oszillate. These oszillations could be reduced by the introduction of non-linear control variates.

References

[1] L. Andersen and M. Broadie. A primal-dual simulation algorithm for pricing mutidimensional American options. *Manage. Sci.*, 50:1222–1234, 2004.

[2] V. Bally and G. Pagès. A quantization algorithm for solving multidimensional discrete-time optimal stopping problems. *Bernoulli*, 9:1003–1049, 2003.

[3] D. Becherer. From bounds on optimal growth towards a theory of good-deal hedging. In H. Albrecher, W. J. Runggaldier, and W. Schachermayer, editors, *Advanced Financial Modelling*, pages 27–52. de Gruyter, 2009.

[4] D. Belomestny, C. Bender, and J. Schoenmakers. True upper bounds for Bermudan products via non-nested Monte-Carlo. *Math. Finance*, 19:53–71, 2009.

[5] C. Bender and R. Denk. A forward scheme for backward SDEs. *Stochastic Process. Appl.*, 117:1793–1812, 2007.

[6] C. Bender and M. Kohlmann. Optimal superhedging under nonconvex constraints: a BSDE approach. *Int. J. Theor. Appl. Finance*, 11:363–380, 2008.

[7] C. Bender and T. Moseler. Importance sampling for backward SDEs. *Stoch. Analysis Appl.*, 28:226–253, 2010.

[8] C. Bender and J. Zhang. Time discretization and Markovian iteration for coupled FBSDEs. *Ann. Appl. Probab.*, 18:143–177, 2008.

[9] Y. Z. Bergman. Option pricing with differential interest rates. *The Review of Financial Studies*, 8:475–500, 1995.

[10] B. Bouchard and N. Touzi. Discrete-time approximation and Monte-Carlo simulation of backward stochastic differential equations. *Stochastic Process. Appl.*, 111:175–206, 2004.

[11] M. Broadie and P. Glasserman. A stochastic mesh method for pricing high-dimensional American options. *J. Comput. Finance*, 7:35–72, 2004.

[12] J. F. Carriere. Valuation of the early-exercise price for options using simulations and nonparametric regression. *Insurance Math. Econom.*, 19:19–30, 1996.

[13] P. Cheridito, H. M. Soner, T. Nizar, and N. Victoir. Second-order backward stochastic differential equations and fully nonlinear parabolic PDEs. *Comm. Pure Appl. Math.*, 60:1081–1110, 2007.

[14] F. Delarue and S. Menozzi. A forward-backward stochastic algorithm for quasi-linear PDEs. *Ann. Appl. Probab.*, 16:140–184, 2006.

[15] J. Douglas, J. Ma, and P. Protter. Numerical method for forward backward stochastic differential equations. *Ann. Appl. Probab.*, 6:940–968, 1996.

[16] N. El Karoui, S. Peng, and M. C. Quenez. Backward stochastic differential equations in finance. *Math. Finance*, 7:1–71, 1997.

[17] E. Gobet and C. Labart. Solving BSDE with adaptive control variate. *SIAM J. Numer. Anal.*, forthcoming.

[18] E. Gobet, J.-P. Lemor, and X. Warin. A regression-based Monte Carlo method to solve backward stochastic differential dquations. *Ann. Appl. Probab.*, 15:2172–2202, 2005.

[19] E. Gobet and A. Makhlouf. L_2-regularity of BSDEs with irregular terminal functions. *Stochastic Process. Appl.*, forthcoming.

[20] M. B. Haugh and L. Kogan. Pricing American options: A duality approach. *Oper. Res.*, 52:258–270, 2004.

[21] U. Horst, T. A. Pirvu, and G. Dos Reis. On securitization, market completion and equilibrium risk transfer. *Math. Financ. Econ.*, 2:211–252, 2010.

[22] P. Imkeller, G. Dos Reis, and J. Zhang. Results on numerics for FBSDE with drivers of quadratic growth. In C. Chiarella and A. Novikov, editors, *Contemporary Quantitative Finance*. Springer, 2010.

[23] J.-P. Lemor, E. Gobet, and X. Warin. Rate of convergence of an empirical regression method for solving generalized backward stochastic differential equations. *Bernoulli*, 5:889–916, 2006.

[24] F. A. Longstaff and E. S. Schwartz. Valuing American options by simulation: A simple least-squares approach. *The Review of Financial Studies*, 14:113–147, 2001.

[25] J. Ma, P. Protter, and J. Yong. Solving forward-backward stochastic differential equations explicitly - a four step scheme. *Probab. Theory Related Fields*, 98:339–359, 1994.

[26] J. Ma, J. Shen, and Y. Zhao. On numerical approximations of forward-backward stochastic differential equations. *SIAM J. Numer. Anal.*, 46:2636–2661, 2008.

[27] M. Mania and M. Schweizer. Dynamic exponential utility indifference valuation. *Ann. Appl. Probab.*, 15:2113–2143, 2005.

[28] G. N. Milstein and M. V. Tretyakov. Numerical algorithms for forward-backward stochastic differential equations. *SIAM J. Sci. Comp.*, 28:561–582, 2006.

[29] E. Pardoux and S. Peng. Backward stochastic differential equations and quasilinear parabolic partial differential equations. In *Stochastic Partial Differential Equations and Their Applications*, volume 176 of *Lecture Notes in Control and Inform. Sci.*, pages 200–217. Springer, 1992.

[30] S. Peng. Nonlinear expectations, nonlinear evaluations and risk measures. In M. Frittelli and W. Runggaldier, editors, *Stochastic Methods in Finance (LNM 1856)*, pages 165–253. Springer, 2004.

[31] L. C. G. Rogers. Monte Carlo valuation of American options. *Math. Finance*, 12:271–286, 2002.

[32] J. R. Tsitsiklis and B. Van Roy. Regression methods for pricing complex American style options. *IEEE Trans. Neural Networks*, 12:694–703, 1999.

[33] J. Yong and X. Y. Zhou. *Stochastic Controls – Hamiltonian Systems and HJB Equations.* Springer, 1999.

[34] J. Zhang. A numerical scheme for BSDEs. *Ann. Appl. Probab.*, 14:459–488, 2004.