

# A Forward Scheme for Backward SDEs<sup>1</sup>

Christian Bender and Robert Denk  
WIAS Berlin and University of Konstanz

July 21, 2005

## Abstract

We introduce a forward scheme to simulate backward SDEs. Compared to existing schemes, we avoid high order nestings of conditional expectations backwards in time. In this way the error, when approximating the conditional expectation, in dependence of the time partition is significantly reduced. Besides this generic result, we present an implementable algorithm and provide an error analysis for it. Finally, we demonstrate the strength of the new algorithm by solving some financial problems numerically.

*AMS classifications:* Primary 65C05; Secondary 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 (1990). Mainly motivated by financial problems (see e.g. the survey article by El Karoui et al. (1997)) 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 4-step-scheme from Ma et al. (1994), numerical algorithms for BSDEs have been developed by Douglas et al. (1996) and more recently by Milstein and Tretyakov (2004). The main focus of these algorithms is the numerical solution of a parabolic PDE which is related to the BSDE.

A second type of algorithms works backwards through time and tries to tackle the stochastic problem directly. Bally (1997) and Chevance (1997) were the first to study this type of algorithm with a (hardly implementable) random time partition respectively under strong regularity assumptions. The work of Ma et al. (2002) is in the same spirit, replacing, however, the Brownian motion by a binary random walk in the approximative equation. See also Briand et al. (2001) 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 Zhang (2004),

---

<sup>1</sup>C. Bender is supported by the DFG Research Center MATHEON ‘Mathematics for key technologies’ in Berlin. R. Denk is partially supported by the AFF grant 28/04 of the University of Konstanz. C. Bender thanks Shanjian Tang for the kind invitation to present this paper at the ‘4th Colloquium on BSDEs and Their Applications’.

which allowed to prove convergence of this backward approach with deterministic partitions under rather weak regularity assumptions, see Zhang (2004), Bouchard and Touzi (2004), and Gobet et al. (2004) for different algorithms.

A main drawback of the backward schemes is, that nestings of conditional expectations backwards through the time steps have to be evaluated. For a practical implementation the conditional expectations must be replaced by some estimator. A generic result of Bouchard and Touzi (2004) 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 geometric rate (Theorem 2.6). At this cost the error, when approximating the conditional expectations by a generic estimator, in dependence of the time partition is reduced by order  $1/2$  compared to existing backward schemes (Theorem 3.1). In fact, in our scheme this error does neither explode when the number of time steps nor when the number of iterations tends to infinity. We believe that this is a striking advantage compared to the backward scheme.

Besides this 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. (2004) in the context of the backward scheme. We analyze the error, when replacing the conditional expectation by the orthogonal projections on subspaces (Theorem 4.1), and also provide rates of convergence when the projection coefficients are substituted by their simulation-based analogues (Theorem 4.9). Again we have an error reduction of order  $1/2$  in the mesh size of the time partition compared to the results in Gobet et al. (2004). Depending on the number  $L$  of simulated paths, the best expected rate of  $L^{-1/2}$  can be achieved for appropriate projection spaces (Theorem 4.9).

Finally, we present some simulations related to financial problems (Section 5). We consider the hedging problem under different interest rates for investing and borrowing and the superhedging problem under borrowing constraints, which lead 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 2.1.** *There is a constant  $K$  such that*

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

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

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

where  $\Phi$  is a functional on the space of  $\mathbb{R}^M$ -valued RCLL-functions on  $[0, T]$  satisfying the  $L^\infty$ -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 neither assume that the matrix  $\sigma$  is quadratic nor that  $\sigma\sigma^*$  is invertible.

**Remark 2.2.** 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. Yong and Zhou (2000), theorem 7.3.4. Here  $(Y^{(0)}, Z^{(0)}) \equiv (0, 0)$ , and  $(Y^{(n)}, Z^{(n)})$  is the solution of the simple BSDE

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

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 Yong and Zhou (2000), ch. 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 but for the occurrence of conditional expectations.

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  $\xi$ . 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  $\Delta W_{d,i} = W_{d,t_{i+1}} - W_{d,t_i}$ ,

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

$d = 1, \dots, D$ . (Here we used the convention  $\Delta W_{d,N} = 0$ ). 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 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 convergence of the discretized Picard-type iteration:

**Theorem 2.3.** *Suppose Assumption 2.1 holds, and for some constant  $C$  depending on the data*

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

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 2.4.** (i) Note, 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,

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

and, thanks to the  $L^\infty$ -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 2.3 is split into two parts. Given the partition  $\pi$  and a corresponding discretization  $X^{(\pi)}$  of  $X$  we define  $(Y^{(\infty,\pi)}, Z^{(\infty,\pi)})$  as the solution of

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

It exists, when the mesh  $|\pi|$  of the partition  $\pi$  is sufficiently fine. Again, the processes  $Y^{(\infty,\pi)}$  and  $Z^{(\infty,\pi)}$  are extended to RCLL processes by constant interpolation. Note,  $(Y^{(\infty,\pi)}, Z^{(\infty,\pi)})$  is (up to the interpolation of the  $Z$ -part) the backward scheme considered in Bouchard and Touzi (2004). 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 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 (2004). The following variant of theorem 3.1 in Bouchard and Touzi (2004) is a slight generalization concerning the assumptions on the coefficients. Moreover, it allows for path-dependent terminal data and the approximating processes are piecewise constant.

**Theorem 2.5.** *Suppose Assumption 2.1 holds, and 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

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

provided  $|\pi|$  is sufficiently small.

The proof combines ideas of Bouchard and Touzi (2004) and Zhang (2004), who suggests a different time discretization. For the reader's convenience we sketch the proof of Theorem 2.5 in the Appendix.

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

**Theorem 2.6.** *Under the assumptions of theorem 2.3 there are constants  $C_1$  and  $C_2$  depending on the data such that*

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

provided  $|\pi|$  is sufficiently small.

Clearly, Theorem 2.3 follows from a straightforward combination of Theorems 2.5 and 2.6.

**Remark 2.7.** Let  $K$  denote the Lipschitz constant of  $f$ . Then it follows from the proof below that Theorem 2.6 holds, for instance, for  $|\pi| \leq \Gamma$  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 2.6 with some a priori estimates.

**Lemma 2.8.** *Suppose  $\Gamma$  and  $\gamma$  are positive real numbers,  $\tilde{y}^{(\iota)}, \tilde{z}^{(\iota)}$ ,  $\iota = 1, 2$  are adapted processes and*

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

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

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

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

*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

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

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

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

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

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

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),

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

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

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

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

**Step 2:** We show

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

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

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

Multiplying with  $\lambda_j$  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 of theorem 2.6.* Denote,

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

By Lemma 2.8,

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

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  $\lambda_i$  from Lemma 2.8 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)–(3). Moreover, for  $n \in \mathbb{N}$ ,

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

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 2.9.** *Under the assumptions of theorem 2.3, 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 inequality 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

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

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  $|\pi| \leq 1$ ,

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

Analogously to step 1 in Lemma 2.8 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  $\Delta_i$  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 2.10.** *Under the assumptions of Theorem 2.3, 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 2.6 we get for sufficiently small  $|\pi|$ ,

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

with a constant  $C$  depending on the data only. Lemma 2.9 concludes.  $\square$

### 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, 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 (2004), 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 (2004), Theorem 4.1, prove, under slightly stronger assumptions than Assumption 2.1, that

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

for some constant  $C$  depending on the data.

This means, given the same accuracy of the conditional expectation estimator 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 the more paths the finer the partition. This increases the computational costs. This effect is particularly unfavorable when the constant in Theorem 2.5 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 described effect has also been observed in the numerical examples by Gobet et al. (2004).

We shall now show that the error due to the approximation of the conditional expectation by its generic estimator does not explode for the discretized Picard iteration. We define

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

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

**Theorem 3.1.** *Under Assumption 2.1 there is a constant  $C$  depending on the*

data such that for any sufficiently fine partition  $\pi$ ,

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

holds for all  $n \in \mathbb{N}$ .

*Proof.* Define,

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

Then, by Young's inequality, and with the notation from Lemma 2.9,

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

Lemma 2.9 can be applied to the second term. Hence, with a suitable choice of  $\Gamma$  and  $\gamma$ ,

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

Now for  $|\pi|$  sufficiently small (e.g. less or equal  $(4\Gamma)^{-1}$ ) the above estimate can be iterated to obtain the theorem. Note,  $1 \leq \lambda_i \leq e^{\Gamma T}$ . 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 Longstaff and Schwartz (2001) in the context of American options and is also applied to the backward scheme in Gobet et al. (2004). 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. Subsection 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  $L^2(\mathcal{F}_{t_i})$ . Precisely, 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 3.1 – is that the error does not propagate backwards in time. Neither does it explode, when the number of iteration tends to infinity. This is an important advantage compared to the scheme proposed in Gobet et al. (2004) where the projection errors sum up over the time steps. Roughly speaking, in the Gobet et al. (2004)-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 4.1.** *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

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

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

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

*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 2.8 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:

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

for any  $\gamma, \Gamma > 0$  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 3.1 taking into account that, due to the orthogonality of the orthogonal projection,

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

□

We also get uniform  $L^2$ -bounds for  $\widehat{Y}^{(n, \pi)}$  and  $\widehat{Z}^{(n, \pi)}$ .

**Corollary 4.2.** *Under the assumptions of Theorem 2.3, 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.

*Proof.* This assertion directly follows from Corollary 2.10 and Theorem 4.1, because the orthogonal projection has norm 1. □

## 4.2 A Monte-Carlo Least-Squares Method to Approximate 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

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

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

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

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

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

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

$$\begin{aligned} \alpha_{i,\cdot}^{(n, \pi)} &= \mathcal{B}_i^{-1} E \left[ \eta^i \left( \xi^{(\pi)} - \sum_{j=i}^{N-1} f(t_j, X_{t_j}^{(\pi)}, \widehat{Y}_{t_j}^{(n-1, \pi)}, \widehat{Z}_{t_j}^{(n-1, \pi)}) \Delta_j \right) \right], \\ \tilde{\alpha}_{i,\cdot}^{(n, \pi)} &= \tilde{\mathcal{B}}_i^{-1} E \left[ \tilde{\eta}^i \frac{\Delta W_i}{\Delta_i} \left( \xi^{(\pi)} - \sum_{j=i+1}^{N-1} f(t_j, X_{t_j}^{(\pi)}, \widehat{Y}_{t_j}^{(n-1, \pi)}, \widehat{Z}_{t_j}^{(n-1, \pi)}) \Delta_j \right) \right]. \end{aligned} \tag{9}$$

The expectations in (9) will be replaced by their simulation based estimators. We shall therefore assume that we have  $L \geq \max_i \{K(i) \vee \tilde{K}(i)\}$  independent samples  $(\Delta W_i^{(\lambda)}, \xi^{(\pi, \lambda)}, X_{t_i}^{(\pi, \lambda)}, \eta_k^{(i, \lambda)}, \tilde{\eta}_k^{(i, \lambda)})$ ,  $\lambda = 1, \dots, L$ , of  $(\Delta W_i, \xi^{(\pi)}, X_{t_i}^{(\pi)}, \eta_k^i, \tilde{\eta}_k^i)$ . We define

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

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

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

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

$$\begin{aligned} \alpha_{i,k}^{(0, \pi, L)} &= \tilde{\alpha}_{i,k}^{(0, \pi, L)} = 0 \\ Y_{t_i}^{(n-1, \pi, \lambda)} &= \sum_{k=1}^{K(i)} \alpha_{i,k}^{(n-1, \pi, L)} \eta_k^{(i, \lambda)} \\ Z_{t_i}^{(n-1, \pi, \lambda)} &= \sum_{k=1}^{\tilde{K}(i)} \tilde{\alpha}_{i,k}^{(n-1, \pi, L)} \tilde{\eta}_k^{(i, \lambda)} \\ \alpha_{i,\cdot}^{(n, \pi, L)} &= \frac{1}{\sqrt{L}} (\mathcal{A}_i^L)^+ \left( \xi^{(\pi, \cdot)} - \sum_{j=i}^{N-1} f(t_j, X_{t_j}^{(\pi, \cdot)}, Y_{t_j}^{(n-1, \pi, \cdot)}, Z_{t_j}^{(n-1, \pi, \cdot)}) \Delta_j \right) \\ \tilde{\alpha}_{i,\cdot}^{(n, \pi, L)} &= \frac{1}{\sqrt{L}} (\tilde{\mathcal{A}}_i^L)^+ \\ &\quad \times \left( \frac{\Delta W_i^{(\cdot)}}{\Delta_i} \left( \xi^{(\pi, \cdot)} - \sum_{j=i+1}^{N-1} f(t_j, X_{t_j}^{(\pi, \cdot)}, Y_{t_j}^{(n-1, \pi, \cdot)}, Z_{t_j}^{(n-1, \pi, \cdot)}) \Delta_j \right) \right). \end{aligned}$$

The simulation based estimators are now defined by,

$$\begin{aligned} Y_{t_i}^{(n, \pi, L, *)} &= \sum_{k=1}^{K(i)} \alpha_{i,k}^{(n, \pi, L)} \eta_k^i, \\ Z_{t_i}^{(n, \pi, L, *)} &= \sum_{k=1}^{\tilde{K}(i)} \tilde{\alpha}_{i,k}^{(n, \pi, L)} \tilde{\eta}_k^i. \end{aligned}$$

**Remark 4.3.** For  $t_i = t_0 = 0$  the only choice of the projection space is  $\Lambda_{0,0} = \mathbb{R}$ . Taking  $\{1\}$  as basis we observe that  $Y_{t_0}^{(n, \pi, L, *)}$  reduces to the plain Monte-Carlo estimator

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

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



We will next prove almost sure convergence of the simulation-based estimators. To this end we first derive a lemma.

**Lemma 4.4.** *Under the Lipschitz condition of Theorem 4.1  $(\alpha_{i,k}^{(n,\pi,L)}, \tilde{\alpha}_{i,k}^{(n,\pi,L)})$  converges  $P$ -almost surely to  $(\alpha_{i,k}^{(n,\pi)}, \tilde{\alpha}_{i,k}^{(n,\pi)})$ , when  $L$  tends to infinity.*

*Proof.* We prove the claim by induction over  $n$ . The case  $n = 0$  is trivial. Suppose now the convergence is already proved for some  $n - 1 \in \mathbb{N}$ . We show the convergence of  $\tilde{\alpha}_{i,k}^{(n,\pi,L)}$ , the argument for  $\alpha_{i,k}^{(n,\pi,L)}$  is similar. First observe that by the law of large numbers

$$\lim_{L \rightarrow 0} \tilde{\mathcal{B}}_i^L = \tilde{\mathcal{B}}_i; \quad P\text{-a.s.} \quad (10)$$

Since  $\tilde{\mathcal{B}}_i$  is invertible, the same holds for  $\tilde{\mathcal{B}}_i^L$  provided  $L$  is sufficiently large (which we assume for the rest of the proof). In particular,  $\tilde{\mathcal{A}}_i^L$  then has full rank, and consequently the pseudo-inverse may be rewritten as

$$\left(\tilde{\mathcal{A}}_i^L\right)^+ = \left(\tilde{\mathcal{B}}_i^L\right)^{-1} \left(\tilde{\mathcal{A}}_i^L\right)^*.$$

Hence,

$$\begin{aligned} \tilde{\alpha}_{i,\cdot}^{(n,\pi,L)} &= \left(\tilde{\mathcal{B}}_i^L\right)^{-1} \left( \frac{1}{L} \sum_{\lambda=1}^L \tilde{\eta}^{(i,\lambda)} \frac{\Delta W_i^{(\lambda)}}{\Delta_i} \left( \xi^{(\pi,\lambda)} \right. \right. \\ &\quad \left. \left. - \sum_{j=i+1}^{N-1} f(t_j, X_{t_j}^{(\pi,\lambda)}, Y_{t_j}^{(n-1,\pi,\lambda)}, Z_{t_j}^{(n-1,\pi,\lambda)}) \Delta_j \right) \right). \end{aligned}$$

By (10) it suffices to prove that for all  $1 \leq l \leq \tilde{K}(i)$ ,

$$\begin{aligned} &\frac{1}{L} \sum_{\lambda=1}^L \tilde{\eta}_l^{(i,\lambda)} \frac{\Delta W_i^{(\lambda)}}{\Delta_i} \left( \xi^{(\pi,\lambda)} - \sum_{j=i+1}^{N-1} f(t_j, X_{t_j}^{(\pi,\lambda)}, Y_{t_j}^{(n-1,\pi,\lambda)}, Z_{t_j}^{(n-1,\pi,\lambda)}) \Delta_j \right) \\ \rightarrow &E \left[ \tilde{\eta}_l^i \frac{\Delta W_i}{\Delta_i} \left( \xi^{(\pi)} - \sum_{j=i+1}^{N-1} f(t_j, X_{t_j}^{(\pi)}, \hat{Y}_{t_j}^{(n-1,\pi)}, \hat{Z}_{t_j}^{(n-1,\pi)}) \Delta_j \right) \right]; \quad P\text{-a.s.} \quad (11) \end{aligned}$$

Define

$$\begin{aligned} \hat{Y}_{t_i}^{(n-1,\pi,\lambda)} &= \sum_{k=1}^{K(i)} \alpha_{i,k}^{(n-1,\pi)} \eta_k^{(i,\lambda)}, \\ \hat{Z}_{t_i}^{(n-1,\pi,\lambda)} &= \sum_{k=1}^{\tilde{K}(i)} \tilde{\alpha}_{i,k}^{(n-1,\pi)} \tilde{\eta}_k^{(i,\lambda)}. \end{aligned} \quad (12)$$

By the law of large numbers,

$$\begin{aligned} &\frac{1}{L} \sum_{\lambda=1}^L \tilde{\eta}_l^{(i,\lambda)} \frac{\Delta W_i^{(\lambda)}}{\Delta_i} \left( \xi^{(\pi,\lambda)} - \sum_{j=i+1}^{N-1} f(t_j, X_{t_j}^{(\pi,\lambda)}, \hat{Y}_{t_j}^{(n-1,\pi,\lambda)}, \hat{Z}_{t_j}^{(n-1,\pi,\lambda)}) \Delta_j \right) \\ \rightarrow &E \left[ \tilde{\eta}_l^i \frac{\Delta W_i}{\Delta_i} \left( \xi^{(\pi)} - \sum_{j=i+1}^{N-1} f(t_j, X_{t_j}^{(\pi)}, \hat{Y}_{t_j}^{(n-1,\pi)}, \hat{Z}_{t_j}^{(n-1,\pi)}) \Delta_j \right) \right]; \quad P\text{-a.s.} \quad (13) \end{aligned}$$

Moreover,

$$\begin{aligned}
& \left| \frac{1}{L} \sum_{\lambda=1}^L \tilde{\eta}_l^{(i,\lambda)} \frac{\Delta W_i^{(\lambda)}}{\Delta_i} \left( \xi^{(\pi,\lambda)} - \sum_{j=i+1}^{N-1} f(t_j, X_{t_j}^{(\pi,\lambda)}, Y_{t_j}^{(n-1,\pi,\lambda)}, Z_{t_j}^{(n-1,\pi,\lambda)}) \Delta_j \right) \right. \\
& - \left. \frac{1}{L} \sum_{\lambda=1}^L \tilde{\eta}_l^{(i,\lambda)} \frac{\Delta W_i^{(\lambda)}}{\Delta_i} \left( \xi^{(\pi,\lambda)} - \sum_{j=i+1}^{N-1} f(t_j, X_{t_j}^{(\pi,\lambda)}, \hat{Y}_{t_j}^{(n-1,\pi,\lambda)}, \hat{Z}_{t_j}^{(n-1,\pi,\lambda)}) \Delta_j \right) \right| \\
& \leq K \frac{1}{L} \sum_{\lambda=1}^L \left| \tilde{\eta}_l^{(i,\lambda)} \frac{\Delta W_i^{(\lambda)}}{\Delta_i} \right| \\
& \quad \times \sum_{j=i+1}^{N-1} |Y_{t_j}^{(n-1,\pi,\lambda)} - \hat{Y}_{t_j}^{(n-1,\pi,\lambda)}| + |Z_{t_j}^{(n-1,\pi,\lambda)} - \hat{Z}_{t_j}^{(n-1,\pi,\lambda)}| \\
& \leq K \frac{1}{L} \sum_{\lambda=1}^L \left| \tilde{\eta}_l^{(i,\lambda)} \frac{\Delta W_i^{(\lambda)}}{\Delta_i} \right| \sum_{j=i+1}^{N-1} \left( \sum_{k=1}^{K(j)} |\eta_k^{(j,\lambda)}| |\alpha_{j,k}^{(n-1,\pi,L)} - \alpha_{j,k}^{(n-1,\pi)}| \right. \\
& \quad \left. + \sum_{k=1}^{\tilde{K}(j)} |\tilde{\eta}_k^{(j,\lambda)}| |\tilde{\alpha}_{j,k}^{(n-1,\pi,L)} - \tilde{\alpha}_{j,k}^{(n-1,\pi)}| \right) \\
& \leq \max_{0 \leq j \leq N-1} \left( \max_{1 \leq k \leq K(j)} |\alpha_{j,k}^{(n-1,\pi,L)} - \alpha_{j,k}^{(n-1,\pi)}| + \max_{1 \leq k' \leq \tilde{K}(j)} |\tilde{\alpha}_{j,k'}^{(n-1,\pi,L)} - \tilde{\alpha}_{j,k'}^{(n-1,\pi)}| \right) \\
& \quad \times K \frac{1}{L} \sum_{\lambda=1}^L \left| \tilde{\eta}_l^{(i,\lambda)} \frac{\Delta W_i^{(\lambda)}}{\Delta_i} \right| \sum_{j=i+1}^{N-1} \left( \sum_{k=1}^{K(j)} |\eta_k^{(j,\lambda)}| + \sum_{k=1}^{\tilde{K}(j)} |\tilde{\eta}_k^{(j,\lambda)}| \right).
\end{aligned}$$

The right hand side tends to zero, since the first factor tends to zero by induction hypothesis and the second converges to a finite number by the law of large numbers. In view of (11)–(13) the proof is complete.  $\square$

An immediate consequence is the convergence of the simulation-based estimators:

**Theorem 4.5.** *Under the Lipschitz condition of Theorem 4.1  $(Y_{t_i}^{(n,\pi,L,*)}, Z_{t_i}^{(n,\pi,L,*)})$  converges  $P$ -almost surely to  $(\hat{Y}_{t_i}^{(n,\pi)}, \hat{Z}_{t_i}^{(n,\pi)})$ , when  $L$  tends to infinity.*

To obtain  $L^2$ -convergence we will introduce truncations of the estimators  $(Y_{t_i}^{(n,\pi,L,*)}, Z_{t_i}^{(n,\pi,L,*)})$ . The following lemma prepares the construction. Here,  $\lambda_{\min}(\mathcal{M})$  denotes the minimal eigenvalue of a symmetric matrix  $\mathcal{M}$ .

**Lemma 4.6.** *Under the conditions of theorem 2.3 there are positive constants  $c_i^{(n,\pi)}$ ,  $\tilde{c}_i^{(n,\pi)}$ , and a constant  $c$  depending on the data such that for sufficiently small  $|\pi|$*

$$\begin{aligned}
|\hat{Y}_{t_i}^{(n,\pi)}| & \leq c_i^{(n,\pi)} \lambda_{\min}(\mathcal{B}_i)^{-1/2} |\eta^i|, \\
|\hat{Z}_{t_i}^{(n,\pi)}| & \leq \tilde{c}_i^{(n,\pi)} \lambda_{\min}(\tilde{\mathcal{B}}_i)^{-1/2} |\tilde{\eta}^i|,
\end{aligned}$$

and

$$\max_{0 \leq i \leq N} |c_i^{(n,\pi)}|^2 + \sum_{i=0}^{N-1} |\tilde{c}_i^{(n,\pi)}|^2 \Delta_i \leq c.$$

*Proof.* As in Gobet et al. (2004), by the Cauchy-Schwarz inequality, and since the symmetric matrix  $\mathcal{B}_i$  satisfies  $\mathcal{B}_i \geq \lambda_{\min}(\mathcal{B}_i)$ ,

$$\begin{aligned} |\widehat{Y}_{t_i}^{(n,\pi)}|^2 &\leq |\alpha_i^{(n,\pi)}|^2 |\eta^i|^2 \leq \lambda_{\min}(\mathcal{B}_i)^{-1} \langle \alpha_i^{(n,\pi)}, \mathcal{B}_i \alpha_i^{(n,\pi)} \rangle |\eta^i|^2 \\ &= \lambda_{\min}(\mathcal{B}_i)^{-1} E \left[ |\widehat{Y}_{t_i}^{(n,\pi)}|^2 \right] |\eta^i|^2. \end{aligned}$$

A similar estimate holds for  $|\widehat{Z}_{t_i}^{(n,\pi)}|^2$ . Hence, in view of corollary 4.2, we can choose

$$\begin{aligned} c_i^{(n,\pi)} &= \left( E \left[ |\widehat{Y}_{t_i}^{(n,\pi)}|^2 \right] \right)^{1/2} \vee 1, \\ \tilde{c}_i^{(n,\pi)} &= \left( E \left[ |\widehat{Z}_{t_i}^{(n,\pi)}|^2 \right] \right)^{1/2} \vee 1, \end{aligned}$$

to obtain the assertion for some constant  $c$  depending on the data only.  $\square$

**Definition 4.7.** We call a  $((c_i^{(n,\pi)}), (\tilde{c}_i^{(n,\pi)}), c, \rho)$  a *truncation tuple*, if  $((c_i^{(n,\pi)}), (\tilde{c}_i^{(n,\pi)}), c)$  satisfy the conditions and estimates of lemma 4.6, and  $\rho: \mathbb{R} \rightarrow \mathbb{R}$  is a Lipschitz continuous function with Lipschitz constant 1, bounded by 2, which coincides with the identity on  $[-1, 1]$ .

By lemma 4.6 we have for every truncation tuple,

$$\begin{aligned} \widehat{Y}_{t_i}^{(n,\pi)} &= c_i^{(n,\pi)} \lambda_{\min}(\mathcal{B}_i)^{-1/2} |\eta^i| \rho \left( \frac{\widehat{Y}_{t_i}^{(n,\pi)}}{c_i^{(n,\pi)} \lambda_{\min}(\mathcal{B}_i)^{-1/2} |\eta^i|} \right), \\ \widehat{Z}_{t_i}^{(n,\pi)} &= \tilde{c}_i^{(n,\pi)} \lambda_{\min}(\tilde{\mathcal{B}}_i)^{-1/2} |\tilde{\eta}^i| \rho \left( \frac{\widehat{Z}_{t_i}^{(n,\pi)}}{\tilde{c}_i^{(n,\pi)} \lambda_{\min}(\tilde{\mathcal{B}}_i)^{-1/2} |\tilde{\eta}^i|} \right). \end{aligned} \quad (14)$$

Suppose, a truncation tuple is given. Then a truncated version of the algorithm is defined by

$$\begin{aligned} \alpha_{i,k}^{(0,\pi,L,\rho)} &= \tilde{\alpha}_{i,k}^{(0,\pi,L,\rho)} = 0 \\ \overline{Y}_{t_i}^{(n-1,\pi,\lambda,\rho)} &= \sum_{k=1}^{K(i)} \alpha_{i,k}^{(n-1,\pi,L,\rho)} \eta_k^{(i,\lambda)} \\ \overline{Z}_{t_i}^{(n-1,\pi,\lambda,\rho)} &= \sum_{k=1}^{\tilde{K}(i)} \tilde{\alpha}_{i,k}^{(n-1,\pi,L,\rho)} \tilde{\eta}_k^{(i,\lambda)} \\ Y_{t_i}^{(n-1,\pi,\lambda,\rho)} &= c_i^{(n-1,\pi)} \lambda_{\min}(\mathcal{B}_i)^{-1/2} |\eta^i| \rho \left( \frac{\overline{Y}_{t_i}^{(n-1,\pi,\lambda,\rho)}}{c_i^{(n-1,\pi)} \lambda_{\min}(\mathcal{B}_i)^{-1/2} |\eta^i|} \right) \\ Z_{t_i}^{(n-1,\pi,\lambda,\rho)} &= \tilde{c}_i^{(n-1,\pi)} \lambda_{\min}(\tilde{\mathcal{B}}_i)^{-1/2} |\tilde{\eta}^i| \rho \left( \frac{\overline{Z}_{t_i}^{(n-1,\pi,\lambda,\rho)}}{\tilde{c}_i^{(n-1,\pi)} \lambda_{\min}(\tilde{\mathcal{B}}_i)^{-1/2} |\tilde{\eta}^i|} \right) \\ \alpha_{i,\cdot}^{(n,\pi,L,\rho)} &= \frac{1}{\sqrt{L}} (\mathcal{A}_i^L)^+ \left( \xi^{(\pi,\cdot)} - \sum_{j=i}^{N-1} f(t_j, X_{t_j}^{(\pi,\cdot)}, Y_{t_j}^{(n-1,\pi,\cdot,\rho)}, Z_{t_j}^{(n-1,\pi,\cdot,\rho)}) \Delta_j \right) \\ \tilde{\alpha}_{i,\cdot}^{(n,\pi,L,\rho)} &= \frac{1}{\sqrt{L}} (\tilde{\mathcal{A}}_i^L)^+ \\ &\times \left( \frac{\Delta W_i^{(\cdot)}}{\Delta_i} \left( \xi^{(\pi,\cdot)} - \sum_{j=i+1}^{N-1} f(t_j, X_{t_j}^{(\pi,\cdot)}, Y_{t_j}^{(n-1,\pi,\cdot,\rho)}, Z_{t_j}^{(n-1,\pi,\cdot,\rho)}) \Delta_j \right) \right). \end{aligned}$$

Moreover, we define truncated versions of  $(Y_{t_i}^{(n,\pi,L,*)}, Z_{t_i}^{(n,\pi,L,*)})$  by

$$\begin{aligned}\widehat{Y}_{t_i}^{(n,\pi,L,\rho)} &= c_i^{(n,\pi)} \lambda_{\min}(\mathcal{B}_i)^{-1/2} |\eta^i| \rho \left( \frac{\sum_k \alpha_{i,k}^{(n,\pi,L,\rho)} \eta_k^i}{c_i^{(n,\pi)} \lambda_{\min}(\mathcal{B}_i)^{-1/2} |\eta^i|} \right), \\ \widehat{Z}_{t_i}^{(n,\pi,L,\rho)} &= \tilde{c}_i^{(n,\pi)} \lambda_{\min}(\tilde{\mathcal{B}}_i)^{-1/2} |\tilde{\eta}^i| \rho \left( \frac{\sum_k \tilde{\alpha}_{i,k}^{(n,\pi,L,\rho)} \tilde{\eta}_k^i}{\tilde{c}_i^{(n,\pi)} \lambda_{\min}(\tilde{\mathcal{B}}_i)^{-1/2} |\tilde{\eta}^i|} \right).\end{aligned}\quad (15)$$

We have,

**Theorem 4.8.** *Under the assumptions of theorem 2.3  $(\widehat{Y}_{t_i}^{(n,\pi,L,\rho)}, \widehat{Z}_{t_i}^{(n,\pi,L,\rho)})$  converges  $P$ -almost surely to  $(\widehat{Y}_{t_i}^{(n,\pi)}, \widehat{Z}_{t_i}^{(n,\pi)})$ , when  $L$  tends to infinity. Moreover,*

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

*Proof.* The Lipschitz condition on  $\rho$  yields in view of (14) and (15)

$$\begin{aligned}& |\widehat{Y}_{t_i}^{(n,\pi,L,\rho)} - \widehat{Y}_{t_i}^{(n,\pi)}| + |\widehat{Z}_{t_i}^{(n,\pi,L,\rho)} - \widehat{Z}_{t_i}^{(n,\pi)}| \sqrt{\Delta_i} \\ & \leq \sum_k |\alpha_{i,k}^{(n,\pi,L,\rho)} - \alpha_{i,k}^{(n,\pi)}| |\eta_k^i| + \sqrt{\Delta_i} \sum_k |\tilde{\alpha}_{i,k}^{(n,\pi,L,\rho)} - \tilde{\alpha}_{i,k}^{(n,\pi)}| |\tilde{\eta}_k^i|.\end{aligned}$$

$P$ -almost sure convergence of  $|\alpha_{i,k}^{(n,\pi,L,\rho)} - \alpha_{i,k}^{(n,\pi)}|$  and  $|\tilde{\alpha}_{i,k}^{(n,\pi,L,\rho)} - \tilde{\alpha}_{i,k}^{(n,\pi)}|$  to zero follows basically with the same argument as in Lemma 4.4. This shows the claimed almost sure convergence. Since  $\rho$  is bounded, the dominated convergence theorem yields the  $L^2$ -convergence.  $\square$

Under some extra conditions on the basis, rates of the  $L^2$ -convergence can be derived.

**Theorem 4.9.** *Suppose the bases  $(\eta^i)$ ,  $(\tilde{\eta}^i)$  are orthonormal for all  $i$  and have finite moments of order  $p \geq 4$ . Moreover assume that*

$$E[|\xi^{(\pi)}|^4] + \max_i E[|X_{t_i}^{(\pi)}|^4] < \infty.$$

*Then there is a constant  $C$  depending on the data such that for  $|\pi|$  sufficiently small*

$$\begin{aligned}& \max_{0 \leq i \leq N} E \left[ |\widehat{Y}_{t_i}^{(n,\pi,L,\rho)} - \widehat{Y}_{t_i}^{(n,\pi)}|^2 \right] + \sum_{i=0}^{N-1} E \left[ |\widehat{Z}_{t_i}^{(n,\pi,L,\rho)} - \widehat{Z}_{t_i}^{(n,\pi)}|^2 \right] \Delta_i \\ & \leq \frac{C}{L^{(p-2)/p}} \left( \max_i E[|\eta^i|^p]^{2/p} + \max_i E[|\tilde{\eta}^i|^p]^{2/p} \right) \\ & \quad \times \left( 8\kappa + 4 \max_{0 \leq i \leq N} E[|\eta^i|_4^4] + 4 \max_{0 \leq i \leq N} E[|\tilde{\eta}^i|_4^4] \right)^{(p-2)/p} \\ & \quad + \left( \frac{C}{L} \max_i E[|\eta^i|_4^4]^{1/2} + \frac{CN}{L} \max_i E[|\tilde{\eta}^i|_4^4]^{1/2} \right) \left( 1 + E[|\xi^{(\pi)}|^4] \right) \\ & \quad + \max_{0 \leq j \leq N} E[|X_{t_j}^{(\pi)}|^4] + \max_{0 \leq j \leq N} E[|\eta^j|_4^4] + \max_{0 \leq j \leq N} E[|\tilde{\eta}^j|_4^4] \Big)^{1/2} \\ & \quad + \frac{C}{L} (1 + \kappa) \left( 2\kappa + \max_{0 \leq i \leq N} E[|\eta^i|_4^4] + \max_{0 \leq i \leq N} E[|\tilde{\eta}^i|_4^4] \right),\end{aligned}$$

where  $\kappa$  is the maximal length of the bases, i.e.  $\kappa = \max_i \{K(i) \vee \tilde{K}(i)\}$  and  $|\cdot|_4$  denotes the 4-norm on a Euclidean space.

The lengthy proof is given in the Appendix.

**Remark 4.10.** If  $X$  is discretized by the Euler scheme, then  $\max_i E[|X_{t_i}^{(\pi)}|^4]$  will be bounded by a constant depending on the data only, see, for instance, Zhang (2004), Lemma 4.1.

**Remark 4.11.** (i) The convergence is of the best expected order 1/2 in  $L$ , provided the elements of the basis are bounded.

(ii) Compared to the error estimates in Gobet et al. (2004), Theorem 3, the error estimates for our scheme are of order 1/2 better in  $N$  than for their scheme. This is in accordance with the discussion in Section 3 above.

(iii) After a first draft of this paper was finished we became aware of the recent paper Lemor et al. (2005). In Lemor et al. (2005) the estimates of Theorem 3 in Gobet et al. (2004) are improved for a localized version of their algorithm. We conjecture that the same techniques can be applied to a localized version of our algorithm.

### 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

$$\begin{aligned} X_0^{(\pi)} &= x \\ X_{t_i}^{(\pi)} &= X_{t_{i-1}}^{(\pi)} + b(t_{i-1}, X_{t_{i-1}}^{(\pi)})\Delta_{i-1} + \sigma(t_{i-1}, X_{t_{i-1}}^{(\pi)})\Delta W_{i-1}, \end{aligned}$$

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 Gobet et al. (2004). 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 Zhang (2004).

**Example 4.12.** (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:** As for 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 4.1,  $(\widehat{Y}^{(n,\pi)}, \widehat{Z}^{(n,\pi)})$  converges (in the  $L^2$ -sense of Theorem 4.1) to  $(Y^{(n,\pi)}, Z^{(n,\pi)})$  as  $\kappa$  tends to infinity. Hence, Theorems 2.3 and 4.8 provide  $L^2$ -convergence of the truncated algorithm (15) in this situation.

## 5 Simulations

In this section we present some simulations of financial problems.

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  $\pi_N$ .

## 5.1 Different Interest Rate for Borrowing

In the first 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 Bergman (1995). 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. Precisely,

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

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 (resp. plus) two times their empirical standard deviation.

Figure 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  $(e_k^{(1)})_k$ , respectively  $\kappa = 21$  for  $(e_k^{(2)})_k$ . In both cases we simulate  $L = 100000$  paths. One can see from Figure 1 that there is a minimal number  $N_{min}$  of time partition points after which the computed price is independent from  $N \geq N_{min}$ . For the linear case this  $N_{min}$  is smaller as for the nonlinear case where  $N_{min}$  is in the range of 15 – 20. We remark that the number of iterations  $n_{stop}$  are 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 by the Black-Scholes formula. We also note that the relative standard error in the calculation of  $Y_0^{(n_{stop}, \pi_N, 100000, *)}$  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 3.1.

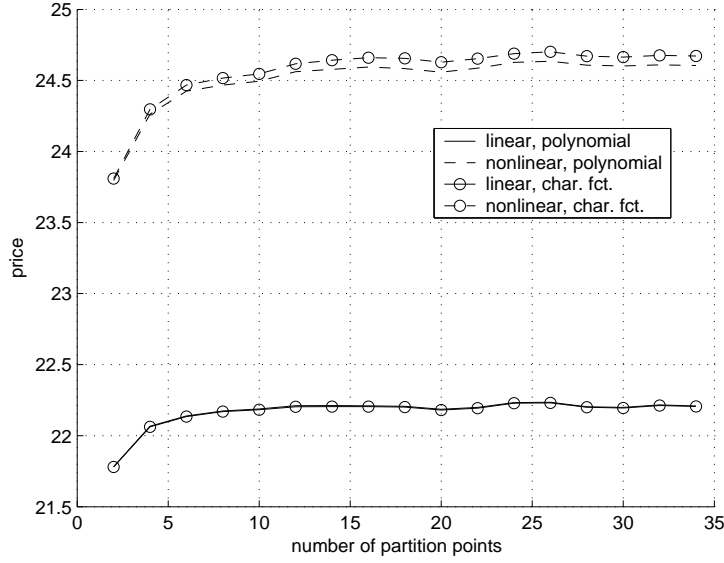


Figure 1:  $Y_0^{(n_{stop}, \pi_N, 100000, *)}$  as a function of  $N$ .

Figure 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  $L$  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. This bias will be explained in the description of Figure 3 below, where it also appears. 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. Precisely, we replace

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

by

$$\xi^{(\pi, \lambda)} - \sum_{j=i}^{N-1} \left( f(t_j, X_{t_j}^{(\pi, \lambda)}, Y_{t_j}^{(n-1, \pi, \lambda)}, Z_{t_j}^{(n-1, \pi, \lambda)}) \Delta_j + Z_{t_j}^{(n-1, \pi, \lambda)} \Delta W_j^{(\lambda)} \right).$$

From Figure 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, 300000 paths, where the relative difference is less than 0.2%.

## 5.2 Constraints on Borrowing

The second example concerns borrowing constraints. Suppose an investor must not borrow an arbitrary amount of money from the money market account



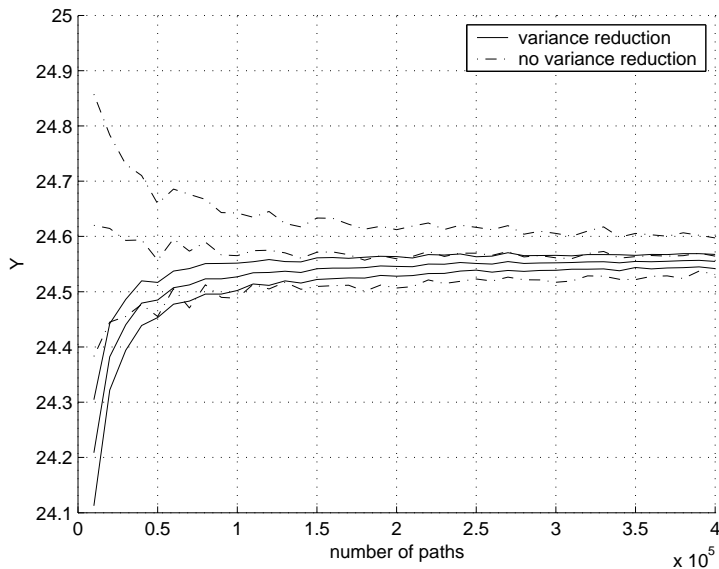


Figure 2: Empirical mean and standard deviation of 100 launches as function of  $L$ .

but a given fraction of his total wealth only. His goal is to super-replicate a given contingent claim (in our case a call option) with minimal initial wealth. This problem is known as superhedging problem. It is shown in Bender and Kohlmann (2004), extending results of El Karoui et al. (1997), that for quite general constraints the solution of the superhedging problem can be obtained as a limit of the solution of a sequence of nonlinear BSDEs. This sequence has an intuitive meaning: The investor is bound to yield an increasing penalization payment when he fails to meet the constraint. In the simple borrowing constraint under consideration the optimal superhedging price can be obtained as the limit of  $Y_0^\epsilon$  (as  $\epsilon$  tends to zero), where

$$\begin{aligned}
 dY_t^\epsilon &= \left[ rY_t^\epsilon + \frac{\mu - r}{\sigma} Z_t^\epsilon - \frac{1}{\epsilon} \left( \frac{Z_t^\epsilon}{\sigma} - \rho Y_t^\epsilon \right)_+ \right] dt + Z_t^\epsilon dW_t \\
 Y_T^\epsilon &= (X_T - K)_+.
 \end{aligned}$$

Here  $\rho - 1$  is the fraction of his total wealth, which the investor is allowed to borrow. We consider the case  $\rho = 10$  with the parameters  $\sigma = 0.2$ ,  $\mu = r = 0.05$ , and  $X_0 = K = 100$ . The maturity is  $T = 0.5$  years. Note, in this example the superhedging price can be determined analytically by calculating an equivalent dominating, but unconstrained, claim, see Broadie et al. (1998). It is 8.06.

We compute numerical approximations for different values of  $\epsilon$ . The stopping criterion for the Picard iteration is 0.001 and we choose  $N = 40$  and the monomial basis with  $\kappa = 5$ , but the straddle payoff replaced by the call payoff. Figure 3 shows the corresponding approximation of  $Y_0^\epsilon$  as function of  $\epsilon$  for different numbers of simulated paths.

We first explain, why the nonlinearity of the generator in the penalization approach is generally expected to yield a positive bias which is increasing in

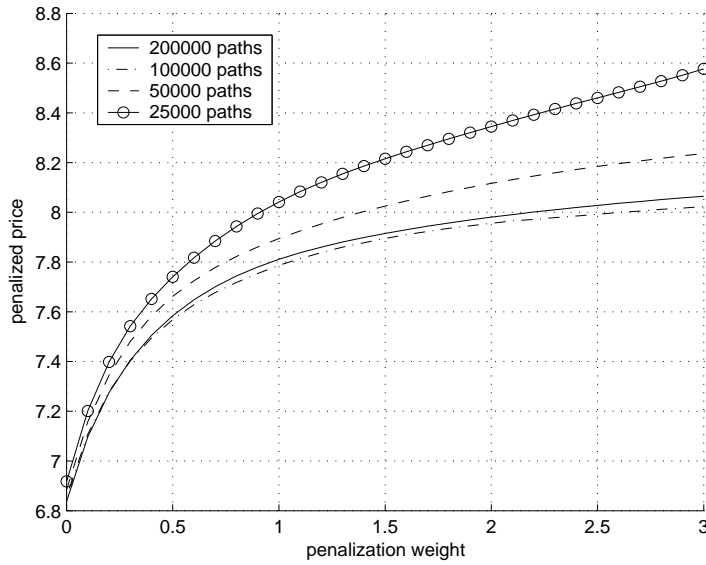


Figure 3:  $\epsilon$ -approximation of the superhedging price as function of  $\epsilon^{-1}$ .

the weight: Suppose violation of the constraint is penalized with a large weight. Due to the penalization there will be more values of  $(Y^\epsilon, Z^\epsilon)$  close to the area, which is forbidden by the constraint, but outside this area than close to the allowed area inside the forbidden one. This effect is more prominent the higher the weight. Hence, by simulation errors supposedly more simulated paths are wrongly penalized (and pushed upwards) than wrongly not penalized. This causes a positive bias increasing in the weight. We remark, this also explains the positive bias in Figure 2, since the higher interest rate for borrowing can be interpreted as a slight penalization for borrowing.

This positive bias can clearly be seen in Figure 3. Indeed, the simulated  $\epsilon$ -approximation tend to merge into a straight line (as function of  $\epsilon^{-1}$ ), when  $\epsilon$  (depending on the number of paths) is sufficient small. Note that the gradient of this straight line is decreasing in the number of simulated path and is already rather small for 100000 paths. Comparing with the simulation for 200000 paths we may deduce that the curve does not become significantly flatter by solely increasing the number of simulated paths. Indeed, in this simulation, the curve with 100000 paths is even a little flatter than the one for 200000 paths. Admittedly, given a number of simulated paths, it seems to be hard to identify the exact weight, from which on the bias dominates. Nonetheless, as the example shows, the penalization method can be applied to approximate the range of the superhedging price. We emphasize that one is typically not interested in the exact superhedging price but whether the difference between the superhedging price and the non-penalized Black-Scholes price is big or not.

## A Appendix: Proof of Theorem 2.5

To ease the notation we only consider the case  $D = 1$ . The extension to the general case is straightforward.

*Proof of Theorem 2.5 .* We recall that  $C$  denotes a constant depending on the data, which may vary from line to line.

**Step 1:** Preliminary estimates:

We first introduce a process  $\tilde{Z}^{(\pi)}$  by

$$Y_{t_{i+1}}^{(\infty, \pi)} = E \left[ Y_{t_{i+1}}^{(\infty, \pi)} \middle| \mathcal{F}_{t_i} \right] + \int_{t_i}^{t_{i+1}} \tilde{Z}_s^{(\pi)} dW_s \quad (16)$$

via the martingale representation theorem. Then,

$$\begin{aligned} & Y_{t_i}^{(\infty, \pi)} - Y_{t_i} + \int_{t_i}^{t_{i+1}} (\tilde{Z}_s^{(\pi)} - Z_s) dW_s \\ = & Y_{t_{i+1}}^{(\infty, \pi)} - Y_{t_{i+1}} - \int_{t_i}^{t_{i+1}} \left( f(t_i, X_{t_i}^{(\pi)}, Y_{t_i}^{(\infty, \pi)}, Z_{t_i}^{(\infty, \pi)}) - f(s, X_s, Y_s, Z_s) \right) ds. \end{aligned}$$

Squaring and taking expectation yields,

$$\begin{aligned} & E \left[ |Y_{t_i}^{(\infty, \pi)} - Y_{t_i}|^2 \right] + E \int_{t_i}^{t_{i+1}} |\tilde{Z}_s^{(\pi)} - Z_s|^2 ds \\ = & E \left[ \left( - \int_{t_i}^{t_{i+1}} \left( f(t_i, X_{t_i}^{(\pi)}, Y_{t_i}^{(\infty, \pi)}, Z_{t_i}^{(\infty, \pi)}) - f(s, X_s, Y_s, Z_s) \right) ds \right. \right. \\ & \left. \left. + Y_{t_{i+1}}^{(\infty, \pi)} - Y_{t_{i+1}} \right)^2 \right] \\ \leq & E \left[ \left( |Y_{t_{i+1}}^{(\infty, \pi)} - Y_{t_{i+1}}| + \Delta_i^{3/2} + K \Delta_i \left( |X_{t_i} - X_{t_i}^{(\pi)}| + \sup_{t_i \leq t \leq t_{i+1}} |X_t - X_{t_i}| \right) \right. \right. \\ & \left. \left. + K \Delta_i \left( |Y_{t_i} - Y_{t_i}^{(\infty, \pi)}| + \sup_{t_i \leq t \leq t_{i+1}} |Y_t - Y_{t_i}| \right) \right. \right. \\ & \left. \left. + K \int_{t_i}^{t_{i+1}} |Z_s - Z_{t_i}^{(\infty, \pi)}| ds \right)^2 \right]. \end{aligned}$$

We can now apply Young's inequality, (1), and Theorem 3.4.3 of Zhang (2001), (see also Lemma 3.2 in Zhang (2004) and observe that no additional path regularity of  $Z$  is required for the proof), to get,

$$\begin{aligned} & E \left[ |Y_{t_i}^{(\infty, \pi)} - Y_{t_i}|^2 \right] + E \int_{t_i}^{t_{i+1}} |\tilde{Z}_s^{(\pi)} - Z_s|^2 ds \\ \leq & \left( 1 + \frac{\Delta_i}{\epsilon} \right) E \left[ |Y_{t_{i+1}}^{(\infty, \pi)} - Y_{t_{i+1}}|^2 \right] + C \left( 1 + \frac{\epsilon}{\Delta_i} \right) \Delta_i^3 \\ & + C \left( 1 + \frac{\epsilon}{\Delta_i} \right) \left( \Delta_i^2 E \left[ |Y_{t_i}^{(\infty, \pi)} - Y_{t_i}|^2 \right] \right. \\ & \left. + \Delta_i E \int_{t_i}^{t_{i+1}} |Z_s - Z_{t_i}^{(\infty, \pi)}|^2 ds \right). \quad (17) \end{aligned}$$

We will next estimate the last term on the right hand side. To this end let us introduce the random variables

$$\widehat{Z}_{t_i}^{(\pi)} = \frac{1}{\Delta_i} E \left[ \int_{t_i}^{t_{i+1}} Z_s ds \middle| \mathcal{F}_{t_i} \right]. \quad (18)$$

It is shown in Zhang (2001), Theorem 3.4.3, that

$$\sum_{i=0}^{N-1} E \int_{t_i}^{t_{i+1}} |Z_s - \widehat{Z}_{t_i}^{(\pi)}|^2 ds \leq C|\pi|. \quad (19)$$

Note also that by (16) and Itô's isometry,

$$Z_{t_i}^{(\infty, \pi)} = \frac{1}{\Delta_i} E \left[ \int_{t_i}^{t_{i+1}} \widetilde{Z}_s^{(\pi)} ds \middle| \mathcal{F}_{t_i} \right]. \quad (20)$$

The identities (18) and (20) can be easily combined to get,

$$\begin{aligned} & E \int_{t_i}^{t_{i+1}} |Z_s - Z_{t_i}^{(\infty, \pi)}|^2 ds \\ & \leq 2 \left( E \int_{t_i}^{t_{i+1}} |Z_s - \widehat{Z}_{t_i}^{(\pi)}|^2 ds + E \int_{t_i}^{t_{i+1}} |Z_s - \widetilde{Z}_s^{(\pi)}|^2 ds \right). \end{aligned} \quad (21)$$

We can now fix  $\epsilon$  sufficiently small such that for small  $|\pi|$  (combining (17) and (21)),

$$\begin{aligned} & \left(1 - \frac{\Delta_i}{4}\right) E \left[ |Y_{t_i}^{(\infty, \pi)} - Y_{t_i}|^2 \right] + \frac{1}{2} E \int_{t_i}^{t_{i+1}} |\widetilde{Z}_s^{(\pi)} - Z_s|^2 ds \\ & \leq \left(1 + \frac{\Delta_i}{\epsilon}\right) E \left[ |Y_{t_{i+1}}^{(\infty, \pi)} - Y_{t_{i+1}}|^2 \right] + C\Delta_i^2 + \frac{1}{2} E \int_{t_i}^{t_{i+1}} |Z_s - \widehat{Z}_{t_i}^{(\pi)}|^2 ds. \end{aligned}$$

Note that for sufficiently small  $|\pi|$ ,

$$\left(1 + \frac{\Delta_i}{\epsilon}\right) \left(1 - \frac{\Delta_i}{4}\right)^{-1} \leq \left(1 + \frac{\Delta_i}{\epsilon} + \frac{\Delta_i}{2}\right).$$

Thus,

$$\begin{aligned} & E \left[ |Y_{t_i}^{(\infty, \pi)} - Y_{t_i}|^2 \right] + \frac{1}{2} E \int_{t_i}^{t_{i+1}} |\widetilde{Z}_s^{(\pi)} - Z_s|^2 ds \\ & \leq \left(1 + C\Delta_i\right) E \left[ |Y_{t_{i+1}}^{(\infty, \pi)} - Y_{t_{i+1}}|^2 \right] + C\Delta_i^2 + CE \int_{t_i}^{t_{i+1}} |Z_s - \widehat{Z}_{t_i}^{(\pi)}|^2 ds \end{aligned} \quad (22)$$

**Step 2:** Convergence of  $Y^{(\infty, \pi)}$ :

We may now conclude from (22), the discrete Gronwall Lemma and (19) that

$$\begin{aligned} & \max_{0 \leq i \leq N} E \left[ |Y_{t_i}^{(\infty, \pi)} - Y_{t_i}|^2 \right] \\ & \leq C \left( E[|\xi - \xi^{(\pi)}|^2] + \sum_{i=0}^{N-1} \Delta_i^2 + E \int_{t_i}^{t_{i+1}} |Z_s - \widehat{Z}_{t_i}^{(\pi)}|^2 ds \right) \\ & \leq C \left( E[|\xi - \xi^{(\pi)}|^2] + |\pi| \right). \end{aligned} \quad (23)$$

This shows the estimate for  $Y^{(\infty, \pi)}$  at the points of the partition. The extension to the piecewise constant interpolation is rather straightforward and identical to the argument in Theorem 5.6 of Zhang (2004).

**Step 3:** Convergence of  $Z^{(\infty, \pi)}$ :

We sum (22) from 0 to  $N - 1$  and obtain,

$$\begin{aligned}
& \sum_{i=0}^{N-1} E \int_{t_i}^{t_{i+1}} |\tilde{Z}_s^{(\pi)} - Z_s|^2 ds \\
& \leq C \sum_{i=1}^{N-1} E \left[ |Y_{t_i}^{(\infty, \pi)} - Y_{t_i}|^2 \Delta_i \right] + CE[|\xi - \xi^{(\pi)}|^2] \\
& \quad + C|\pi| + C \sum_{i=0}^{N-1} E \int_{t_i}^{t_{i+1}} |\hat{Z}_{t_i}^{(\pi)} - Z_s|^2 ds \\
& \leq C \left( E[|\xi - \xi^{(\pi)}|^2] + |\pi| \right)
\end{aligned} \tag{24}$$

due to (19) and (23). By (20) and the mean-square minimizing property of the conditional expectation,

$$E \left[ \left( \int_{t_i}^{t_{i+1}} (\tilde{Z}_s^{(\pi)} - Z_{t_i}^{(\infty, \pi)}) ds \right)^2 \right] \leq E \left[ \left( \int_{t_i}^{t_{i+1}} (\tilde{Z}_s^{(\pi)} - \hat{Z}_{t_i}^{(\pi)}) ds \right)^2 \right].$$

Elementary manipulations show that this is equivalent to

$$E \left[ \int_{t_i}^{t_{i+1}} (\tilde{Z}_s^{(\pi)} - Z_{t_i}^{(\infty, \pi)})^2 ds \right] \leq E \left[ \int_{t_i}^{t_{i+1}} (\tilde{Z}_s^{(\pi)} - \hat{Z}_{t_i}^{(\pi)})^2 ds \right].$$

The estimate for  $Z^{(\infty, \pi)}$  may now be easily derived from (19) and (24).  $\square$

## B Appendix: Proof of Theorem 4.9

We first introduce the sets

$$M_i^L = \left\{ \|\mathcal{B}_i^L - I\| \leq \frac{1}{2} \wedge \|\tilde{\mathcal{B}}_i^L - I\| \leq \frac{1}{2} \right\},$$

where  $\|\cdot\|$  denotes the operator norm with respect to the Euclidean norm, i.e. the maximal eigenvalue of the matrix, and  $I$  is the unit matrix.

**Lemma B.1.** *We have*

$$\begin{aligned}
& \max_{0 \leq i \leq N} E \left[ \|\mathcal{B}_i^L - I\|^2 \right] + \max_{0 \leq i \leq N} E \left[ \|\tilde{\mathcal{B}}_i^L - I\|^2 \right] \\
& \leq \frac{1}{L} \left( 2\kappa + \max_{0 \leq i \leq N} E[|\eta^i|_4^4] + \max_{0 \leq i \leq N} E[|\tilde{\eta}^i|_4^4] \right).
\end{aligned}$$

*In particular,*

$$P \left( (M_i^L)^c \right) \leq \frac{4}{L} \left( 2\kappa + \max_{0 \leq i \leq N} E[|\eta^i|_4^4] + \max_{0 \leq i \leq N} E[|\tilde{\eta}^i|_4^4] \right).$$

*Proof.* We have, due to orthonormality and independence,

$$\begin{aligned}
E [\|\mathcal{B}_i^L - I\|^2] &\leq \sum_{k,l=1}^{K(i)} E \left[ \left| \frac{1}{L} \sum_{\lambda=1}^L \eta_k^{(i,\lambda)} \eta_l^{(i,\lambda)} - \delta_{k,l} \right|^2 \right] \\
&= \frac{1}{L} \sum_{k,l=1}^{K(i)} E \left[ |\eta_k^i \eta_l^i - \delta_{k,l}|^2 \right] \\
&\leq \frac{1}{L} \sum_{k,l=1; k \neq l}^{K(i)} \frac{1}{2} E [|\eta_k^i|^4 + |\eta_l^i|^4] + \sum_{k=1}^{K(i)} E [|\eta_k^i|^4] + \kappa \\
&= \frac{\kappa}{L} + \frac{1}{L} \sum_{k=1}^{K(i)} E [|\eta_k^i|^4] = \frac{1}{L} (\kappa + E[|\eta^i|^4]).
\end{aligned}$$

The estimate for  $E [\|\widetilde{\mathcal{B}}_i^L - I\|^2]$  follows the same lines. The probability of  $(M_i^L)^c$  can now be estimated by Chebyshev's inequality.  $\square$

We next give estimates concerning

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

**Lemma B.2.** *There is a constant  $C$  depending on the data only such that*

$$\begin{aligned}
E [|b_i|^4] &\leq C \left( 1 + E[|\xi^{(\pi)}|^4] + \max_{0 \leq j \leq N} E[|X_{t_j}^{(\pi)}|^4] \right. \\
&\quad \left. + \max_{0 \leq j \leq N} E[|\eta^j|^4] + \max_{0 \leq j \leq N} E[|\widetilde{\eta}^j|^4] \right), \tag{25}
\end{aligned}$$

$$E [|b_i|^2] \leq C(1 + \kappa), \tag{26}$$

and

$$\sum_{i=0}^{N-1} E \left[ \left| E \left[ \frac{\Delta W_i}{\Delta_i} b_{i+1} \middle| \mathcal{F}_{t_i} \right] \right|^2 \right] \Delta_i \leq C(1 + \kappa). \tag{27}$$

*Proof.* By the Lipschitz continuity of  $f$  we obtain,

$$|b_i| \leq |\xi^{(\pi)}| + \sum_{j=0}^{N-1} |f(t_j, 0, 0, 0)| + K \left( |X_{t_j}^{(\pi)}| + |\widehat{Y}_{t_j}^{(n-1,\pi,L,\rho)}| + |\widehat{Z}_{t_j}^{(n-1,\pi,L,\rho)}| \right) \Delta_j. \tag{28}$$

Note that by (15) and Hölder's inequality

$$\begin{aligned}
\left( \sum_{j=0}^{N-1} |\widehat{Z}_{t_j}^{(n-1,\pi,L,\rho)}| \Delta_j \right)^4 &\leq \left( \sum_{j=0}^{N-1} |\widehat{Z}_{t_j}^{(n-1,\pi,L,\rho)}|^2 |\widetilde{\eta}^j|^{-2} \Delta_j \right)^2 \left( \sum_{j=0}^{N-1} |\widetilde{\eta}^j|^2 \Delta_j \right)^2 \\
&\leq 16c^2 T \sum_{j=0}^{N-1} |\widetilde{\eta}^j|^4 \Delta_j.
\end{aligned}$$

Estimating the other terms in (28) in a similar but easier way we obtain, due to Young's inequality,

$$|b_i|^4 \leq C \left( 1 + |\xi^{(\pi)}|^4 + \max_{0 \leq j \leq N} |X_{t_j}^{(\pi)}|^4 + \max_{0 \leq j \leq N} |\eta^j|^4 + \max_{0 \leq j \leq N} |\tilde{\eta}^j|^4 \right),$$

whence (25). Along the same lines we obtain

$$|b_i|^2 \leq C \left( 1 + |\xi^{(\pi)}|^2 + \max_{0 \leq j \leq N} |X_{t_j}^{(\pi)}|^2 + \sum_{j=0}^{N-1} |\eta^j|^2 \Delta_j + \sum_{j=0}^{N-1} |\tilde{\eta}^j|^2 \Delta_j \right),$$

which in view of the orthonormality of the bases yields (26). As in the proof of lemma 2.8, step 1, we get,

$$\begin{aligned} & \sum_{i=0}^{N-1} E \left[ \left| E \left[ \frac{\Delta W_i}{\Delta_i} b_{i+1} \middle| \mathcal{F}_{t_i} \right] \right|^2 \right] \Delta_i \\ & \leq E[|\xi^{(\pi)}|^2] - 2 \sum_{i=0}^{N-1} E \left[ E[b_i | \mathcal{F}_i] f(t_i, X_{t_i}^{(\pi)}, \widehat{Y}_{t_i}^{(n-1, \pi, L, \rho)}, \widehat{Z}_{t_i}^{(n-1, \pi, L, \rho)}) \right] \Delta_i \\ & \leq E[|\xi^{(\pi)}|^2] + \sum_{i=0}^{N-1} E[|b_i|^2] \Delta_i \\ & \quad + CE \left[ \sum_{i=0}^{N-1} |f(t_i, 0, 0, 0)|^2 + |X_{t_i}^{(\pi)}|^2 + |\widehat{Y}_{t_i}^{(n-1, \pi, L, \rho)}|^2 + |\widehat{Z}_{t_i}^{(n-1, \pi, L, \rho)}|^2 \Delta_j \right]. \end{aligned}$$

Estimate (27) now easily follows.  $\square$

With these estimates at hand we can now give the

*Proof of Theorem 4.9.* Define  $\lambda_0 = 1$ ,  $\lambda_{i+1} = \lambda_i(\Gamma \Delta_i + 1)$  with a constant  $\Gamma$  depending on the data only which is to be fixed later. We further introduce simulation based analogues of  $b_i$ , namely,

$$\begin{aligned} b_i^{(\lambda)} &= \xi^{(\pi)} - \sum_{j=i}^{N-1} f(t_j, X_{t_j}^{(\pi, \lambda)}, Y_{t_j}^{(n-1, \pi, \lambda, \rho)}, Z_{t_j}^{(n-1, \pi, \lambda, \rho)}) \Delta_j, \\ \widehat{b}_i^{(\lambda)} &= \xi^{(\pi)} - \sum_{j=i}^{N-1} f(t_j, X_{t_j}^{(\pi, \lambda)}, \widehat{Y}_{t_j}^{(n-1, \pi, \lambda)}, \widehat{Z}_{t_j}^{(n-1, \pi, \lambda)}) \Delta_j, \end{aligned}$$

where  $(\widehat{Y}_{t_j}^{(n-1,\pi,\lambda)}, \widehat{Z}_{t_j}^{(n-1,\pi,\lambda)})$  are defined in (12). We decompose

$$\begin{aligned}
& \max_{0 \leq i \leq N} \lambda_i E \left[ |\widehat{Y}_{t_i}^{(n,\pi,L,\rho)} - \widehat{Y}_{t_i}^{(n,\pi)}|^2 \right] + \sum_{i=0}^{N-1} \lambda_i E \left[ |\widehat{Z}_{t_i}^{(n,\pi,L,\rho)} - \widehat{Z}_{t_i}^{(n,\pi)}|^2 \right] \Delta_i \\
= & \max_{0 \leq i \leq N} \lambda_i E \left[ \mathbf{1}_{M_i^L} |\widehat{Y}_{t_i}^{(n,\pi,L,\rho)} - \widehat{Y}_{t_i}^{(n,\pi)}|^2 \right] \\
& + \sum_{i=0}^{N-1} \lambda_i E \left[ \mathbf{1}_{M_i^L} |\widehat{Z}_{t_i}^{(n,\pi,L,\rho)} - \widehat{Z}_{t_i}^{(n,\pi)}|^2 \right] \Delta_i \\
& + \max_{0 \leq i \leq N} \lambda_i E \left[ \mathbf{1}_{(M_i^L)^c} |\widehat{Y}_{t_i}^{(n,\pi,L,\rho)} - \widehat{Y}_{t_i}^{(n,\pi)}|^2 \right] \\
& + \sum_{i=0}^{N-1} \lambda_i E \left[ \mathbf{1}_{(M_i^L)^c} |\widehat{Z}_{t_i}^{(n,\pi,L,\rho)} - \widehat{Z}_{t_i}^{(n,\pi)}|^2 \right] \Delta_i =: (I) + (II). \tag{29}
\end{aligned}$$

We obtain from Lemma B.1,

$$\begin{aligned}
(II) & \leq 16e^{\Gamma T} \left( \max_i E[(c_i^{(n,\pi)})^2 |\eta^i|^2 \mathbf{1}_{(M_i^L)^c}] + \sum_i E[(\widetilde{c}_i^{(n,\pi)})^2 |\widetilde{\eta}^i|^2 \mathbf{1}_{(M_i^L)^c}] \Delta_i \right) \\
& \leq 16e^{\Gamma T} c \left( \max_i E[|\eta^i|^2 \mathbf{1}_{(M_i^L)^c}] + \max_i E[|\widetilde{\eta}^i|^2 \mathbf{1}_{(M_i^L)^c}] \right) \\
& \leq 16e^{\Gamma T} c \max_i P((M_i^L)^c)^{(p-2)/p} \left( \max_i E[|\eta^i|^p]^{2/p} + \max_i E[|\widetilde{\eta}^i|^p]^{2/p} \right) \\
& \leq \frac{C}{L^{(p-2)/p}} \left( \max_i E[|\eta^i|^p]^{2/p} + \max_i E[|\widetilde{\eta}^i|^p]^{2/p} \right) \\
& \quad \times \left( 8\kappa + 4 \max_{0 \leq i \leq N} E[|\eta^i|_4^4] + 4 \max_{0 \leq i \leq N} E[|\widetilde{\eta}^i|_4^4] \right)^{(p-2)/p}. \tag{30}
\end{aligned}$$

We now turn to (I). By the Lipschitz property of  $\rho$  and since the bases  $(\eta^i)$ ,  $(\widetilde{\eta}^i)$  are orthonormal,

$$\begin{aligned}
(I) & \leq \max_{0 \leq i \leq N} \lambda_i E \left[ \mathbf{1}_{M_i^L} \sum_{k=1}^{K(i)} |\alpha_{i,k}^{(n,\pi,L,\rho)} - \alpha_{i,k}^{(n,\pi)}|^2 \right] \\
& \quad + \sum_{i=0}^{N-1} \lambda_i E \left[ \mathbf{1}_{M_i^L} \sum_{k=1}^{\widetilde{K}(i)} |\widetilde{\alpha}_{i,k}^{(n,\pi,L,\rho)} - \widetilde{\alpha}_{i,k}^{(n,\pi)}|^2 \right] \Delta_i.
\end{aligned}$$

Note now that

$$\begin{aligned}
\widetilde{\alpha}_{i,\cdot}^{(n,\pi,L,\rho)} & = \frac{1}{\sqrt{L}} (\widetilde{\mathcal{A}}_i^L)^+ \left( \frac{\Delta W_i^{(\cdot)}}{\Delta_i} \widehat{b}_{i+1}^{(\cdot)} \right) \\
\widetilde{\alpha}_{i,\cdot}^{(n,\pi)} & = \frac{1}{\sqrt{L}} E \left[ (\widetilde{\mathcal{A}}_i^L)^* \left( \frac{\Delta W_i^{(\cdot)}}{\Delta_i} \widehat{b}_{i+1}^{(\cdot)} \right) \right]
\end{aligned}$$



and analogous expressions hold without the tilde. Hence,

$$\begin{aligned}
(I) &\leq \max_{0 \leq i \leq N} \lambda_i \frac{2}{L} E \left[ \mathbf{1}_{M_i^L} |((\mathcal{A}_i^L)^+ - (\mathcal{A}_i^L)^*) b_i^{(\cdot)}|^2 \right] \\
&\quad + \sum_{i=0}^{N-1} \lambda_i \frac{2}{L} E \left[ \mathbf{1}_{M_i^L} \left| ((\tilde{\mathcal{A}}_i^L)^+ - (\tilde{\mathcal{A}}_i^L)^*) \left( \frac{\Delta W_i^{(\cdot)}}{\Delta_i} b_{i+1}^{(\cdot)} \right) \right|^2 \right] \\
&\quad + \max_{0 \leq i \leq N} \lambda_i \frac{2}{L} E \left[ \mathbf{1}_{M_i^L} \left| (\mathcal{A}_i^L)^* b_i^{(\cdot)} - E \left[ (\mathcal{A}_i^L)^* \hat{b}_i^{(\cdot)} \right] \right|^2 \right] \\
&\quad + \sum_{i=0}^{N-1} \lambda_i \frac{2}{L} E \left[ \mathbf{1}_{M_i^L} \left| (\tilde{\mathcal{A}}_i^L)^* \left( \frac{\Delta W_i^{(\cdot)}}{\Delta_i} b_{i+1}^{(\cdot)} \right) - E \left[ (\tilde{\mathcal{A}}_i^L)^* \left( \frac{\Delta W_i^{(\cdot)}}{\Delta_i} \hat{b}_{i+1}^{(\cdot)} \right) \right] \right|^2 \right] \\
&=: (Ia) + (Ib).
\end{aligned}$$

We first estimate (Ia). On the set  $M_i^L$  the pseudo-inverse may be rewritten as

$$(\tilde{\mathcal{A}}_i^L)^+ = (\tilde{\mathcal{B}}_i^L)^{-1} (\tilde{\mathcal{A}}_i^L)^*.$$

As in Gobet et al. (2004) we expand  $(\tilde{\mathcal{B}}_i^L)^{-1} - I$  in a von-Neumann series and observe on  $M_i^L$

$$\|(\tilde{\mathcal{B}}_i^L)^{-1} - I\|^2 \leq \frac{\|\tilde{\mathcal{B}}_i^L - I\|^2}{(1 - \|\tilde{\mathcal{B}}_i^L - I\|)^2} \leq 1.$$

Hence,

$$\begin{aligned}
(Ia) &\leq \max_{0 \leq i \leq N} \lambda_i \frac{4}{L} E \left[ |(\mathcal{A}_i^L)^* b_i^{(\cdot)} - E[(\mathcal{A}_i^L)^* b_i^{(\cdot)}]|^2 \right] \\
&\quad + \sum_{i=0}^{N-1} \lambda_i \frac{4}{L} E \left[ \left| (\tilde{\mathcal{A}}_i^L)^* \left( \frac{\Delta W_i^{(\cdot)}}{\Delta_i} b_{i+1}^{(\cdot)} \right) - E \left[ (\tilde{\mathcal{A}}_i^L)^* \left( \frac{\Delta W_i^{(\cdot)}}{\Delta_i} b_{i+1}^{(\cdot)} \right) \right] \right|^2 \right] \Delta_i \\
&\quad + \max_{0 \leq i \leq N} \lambda_i \frac{16}{L} E \left[ \|\mathcal{B}_i^L - I\|^2 |E[(\mathcal{A}_i^L)^* b_i^{(\cdot)}]|^2 \right] \\
&\quad + \sum_{i=0}^{N-1} \lambda_i \frac{16}{L} E \left[ \|\tilde{\mathcal{B}}_i^L - I\|^2 \left| E \left[ (\tilde{\mathcal{A}}_i^L)^* \left( \frac{\Delta W_i^{(\cdot)}}{\Delta_i} b_{i+1}^{(\cdot)} \right) \right] \right|^2 \right] \\
&= (III) + (IV).
\end{aligned}$$

To estimate (III) we first note that

$$\begin{aligned}
&E \left[ \left| (\tilde{\mathcal{A}}_i^L)^* \left( \frac{\Delta W_i^{(\cdot)}}{\Delta_i} b_{i+1}^{(\cdot)} \right) - E \left[ \tilde{\mathcal{A}}_i^L \left( \frac{\Delta W_i^{(\cdot)}}{\Delta_i} b_{i+1}^{(\cdot)} \right) \right] \right|^2 \right] \\
&= E \left[ \frac{1}{L} \sum_{\lambda=1}^L \left( \frac{\Delta W_i^{(\lambda)}}{\Delta_i} b_{i+1}^{(\lambda)} \tilde{\eta}^{(i,\lambda)} - E \left[ \frac{\Delta W_i^{(\lambda)}}{\Delta_i} b_{i+1}^{(\lambda)} \tilde{\eta}^{(i,\lambda)} \right] \right)^2 \right] \\
&= E \left[ \left( \frac{\Delta W_i}{\Delta_i} b_{i+1} \tilde{\eta}^i - E \left[ \frac{\Delta W_i}{\Delta_i} b_{i+1} \tilde{\eta}^i \right] \right)^2 \right] \\
&\leq E[|b_{i+1}|^4]^{1/2} E[|\tilde{\eta}^i|^4]^{1/2} \frac{1}{\Delta_i} E[|U|^4]^{1/2}
\end{aligned}$$

where  $U$  is standard Gaussian. A similar estimate holds for the  $Y$ -part. Thus, by lemma B.2

$$\begin{aligned}
(III) &\leq \left( \frac{C}{L} \max_i E[|\eta^i|^4]^{1/2} + \frac{CN}{L} \max_i E[|\tilde{\eta}^i|^4]^{1/2} \right) \max_i E[|b_i|^4]^{1/2} \\
&\leq \left( \frac{C}{L} \max_i E[|\eta^i|^4]^{1/2} + \frac{CN}{L} \max_i E[|\tilde{\eta}^i|^4]^{1/2} \right) \left( 1 + E[|\xi^{(\pi)}|^4] \right. \\
&\quad \left. + \max_{0 \leq j \leq N} E[|X_{t_j}^{(\pi)}|^4] + \max_{0 \leq j \leq N} E[|\eta^j|^4] + \max_{0 \leq j \leq N} E[|\tilde{\eta}^j|^4] \right)^{1/2} \quad (31)
\end{aligned}$$

Concerning (IV) note,

$$\begin{aligned}
&\frac{1}{L} \left| E \left[ (\tilde{\mathcal{A}}_i^L)^* \left( \frac{\Delta W_i^{(\cdot)}}{\Delta_i} b_{i+1}^{(\cdot)} \right) \right] \right|^2 = \left| E \left[ \tilde{\eta}^i \left( \frac{\Delta W_i}{\Delta_i} b_{i+1} \right) \right] \right|^2 \\
&= E \left[ \left| P_{1,i} \left( \frac{\Delta W_i}{\Delta_i} b_{i+1} \right) \right|^2 \right] \leq E \left[ \left| E \left[ \frac{\Delta W_i}{\Delta_i} b_{i+1} \mid \mathcal{F}_{t_i} \right] \right|^2 \right]
\end{aligned}$$

and similarly for the  $Y$ -part. Thus, by Lemmas B.1 and B.2,

$$(IV) \leq \frac{C}{L} (1 + \kappa) \left( 2\kappa + \max_{0 \leq i \leq N} E[|\eta^i|^4] + \max_{0 \leq i \leq N} E[|\tilde{\eta}^i|^4] \right). \quad (32)$$

We now decompose (Ib) as

$$\begin{aligned}
(Ib) &\leq \max_i \lambda_i \frac{4}{L} E \left[ \left| (\mathcal{A}_i^L)^* b_i^{(\cdot)} - E \left[ (\mathcal{A}_i^L)^* b_i^{(\cdot)} \right] \right|^2 \right] \\
&\quad + \sum_i \lambda_i \frac{4}{L} E \left[ \left| (\tilde{\mathcal{A}}_i^L)^* \left( \frac{\Delta W_i^{(\cdot)}}{\Delta_i} b_{i+1}^{(\cdot)} \right) - E \left[ (\tilde{\mathcal{A}}_i^L)^* \left( \frac{\Delta W_i^{(\cdot)}}{\Delta_i} b_{i+1}^{(\cdot)} \right) \right] \right|^2 \right] \\
&\quad + \max_i \lambda_i \frac{4}{L} \left| E \left[ (\mathcal{A}_i^L)^* (\hat{b}_i^{(\cdot)} - b_i^{(\cdot)}) \right] \right|^2 \\
&\quad + \sum_i \lambda_i \frac{4}{L} \left| E \left[ (\mathcal{A}_i^L)^* \left( \left( \frac{\Delta W_i^{(\cdot)}}{\Delta_i} \hat{b}_{i+1}^{(\cdot)} \right) - \left( \frac{\Delta W_i^{(\cdot)}}{\Delta_i} b_{i+1}^{(\cdot)} \right) \right) \right] \right|^2 \\
&= (III) + (V). \quad (33)
\end{aligned}$$

Concerning (V) we obtain

$$\begin{aligned}
&\frac{1}{L} E \left[ \left| (\mathcal{A}_i^L)^* \left( \left( \frac{\Delta W_i^{(\cdot)}}{\Delta_i} \hat{b}_{i+1}^{(\cdot)} \right) - \left( \frac{\Delta W_i^{(\cdot)}}{\Delta_i} b_{i+1}^{(\cdot)} \right) \right) \right|^2 \right] \\
&= E \left[ \left| \tilde{\eta}^i \frac{\Delta W_i}{\Delta_i} (b_{i+1} - \hat{b}_{i+1}) \right|^2 \right] = E \left[ \left| P_{1,i} \left( \frac{\Delta W_i}{\Delta_i} (b_{i+1} - \hat{b}_{i+1}) \right) \right|^2 \right] \\
&\leq E \left[ \left| E \left[ \frac{\Delta W_i}{\Delta_i} (b_{i+1} - \hat{b}_{i+1}) \mid \mathcal{F}_{t_i} \right] \right|^2 \right]
\end{aligned}$$

and similarly for the  $Y$ -part. Hence, we can apply Lemma 2.8 and get, with an

appropriate choice of  $\gamma$  and  $\Gamma$ , and provided  $|\pi|$  is sufficiently small,

$$(V) \leq \left(\frac{1}{4} + \Gamma|\pi|\right) \max_{0 \leq i \leq N} \lambda_i E \left[ |\widehat{Y}_{t_i}^{(n-1, \pi, L, \rho)} - \widehat{Y}_{t_i}^{(n-1, \pi)}|^2 \right] \\ + \left(\frac{1}{4} + \Gamma|\pi|\right) \sum_{i=0}^{N-1} \lambda_i E \left[ |\widehat{Z}_{t_i}^{(n-1, \pi, L, \rho)} - \widehat{Z}_{t_i}^{(n-1, \pi)}|^2 \right] \Delta_i. \quad (34)$$

Gathering (29)–(34) yields, for sufficiently small  $|\pi|$ ,

$$\max_{0 \leq i \leq N} \lambda_i E \left[ |\widehat{Y}_{t_i}^{(n, \pi, L, \rho)} - \widehat{Y}_{t_i}^{(n, \pi)}|^2 \right] + \sum_{i=0}^{N-1} \lambda_i E \left[ |\widehat{Z}_{t_i}^{(n, \pi, L, \rho)} - \widehat{Z}_{t_i}^{(n, \pi)}|^2 \right] \Delta_i \\ \leq \frac{1}{2} \max_{0 \leq i \leq N} \lambda_i E \left[ |\widehat{Y}_{t_i}^{(n-1, \pi, L, \rho)} - \widehat{Y}_{t_i}^{(n-1, \pi)}|^2 \right] \\ + \frac{1}{2} \sum_{i=0}^{N-1} \lambda_i E \left[ |\widehat{Z}_{t_i}^{(n-1, \pi, L, \rho)} - \widehat{Z}_{t_i}^{(n-1, \pi)}|^2 \right] \Delta_i \\ + \frac{C}{L^{(p-2)/p}} \left( \max_i E[|\eta^i|^p]^{2/p} + \max_i E[|\tilde{\eta}^i|^p]^{2/p} \right) \\ \times \left( 8\kappa + 4 \max_{0 \leq i \leq N} E[|\eta^i|_4^4] + 4 \max_{0 \leq i \leq N} E[|\tilde{\eta}^i|_4^4] \right)^{(p-2)/p} \\ + \left( \frac{C}{L} \max_i E[|\eta^i|^4]^{1/2} + \frac{CN}{L} \max_i E[|\tilde{\eta}^i|^4]^{1/2} \right) \left( 1 + E[|\xi^{(\pi)}|^4] \right) \\ + \max_{0 \leq j \leq N} E[|X_{t_j}^{(\pi)}|^4] + \max_{0 \leq j \leq N} E[|\eta^j|^4] + \max_{0 \leq j \leq N} E[|\tilde{\eta}^j|^4] \Big)^{1/2} \\ + \frac{C}{L} (1 + \kappa) \left( 2\kappa + \max_{0 \leq i \leq N} E[|\eta^i|_4^4] + \max_{0 \leq i \leq N} E[|\tilde{\eta}^i|_4^4] \right).$$

The assertion now follows by iterating this inequality.  $\square$

## References

- Bally, V. (1997) Approximation Scheme for Solutions of BSDE. In: El Karoui, N., Mazliak, L. (eds.) *Backward Stochastic Differential Equations*, Addison Wesley Longman, 177-191.
- Bender, C., Kohlmann, M. (2004) Optimal Superhedging under Nonconvex Constraints – a BSDE Approach. WIAS-Preprint 928.
- Bergman, Y. Z. (1995) Option Pricing with Differential Interest Rates. *Rev. of Financial Studies*, **8**, 475-500.
- Bouchard, B., Touzi, N. (2004) Discrete-Time Approximation and Monte-Carlo Simulation of Backward Stochastic Differential Equations. *Stochastic Process. Appl.*, **111**, 175-206.
- Briand, P., Delyon, B., Mémmin, J. (2001) Donsker-Type Theorem for BSDES. *Elect. Comm. in Probab.*, **6**, 1-14.

- Broadie, M., Cvitanić, J., Mete Soner, H. (1998) Optimal Replication of Contingent Claims under Portfolio Constraints. *Rev. of Financial Studies*, **11**, 59-79.
- Chevance, D. (1997) Numerical Methods for Backward Stochastic Differential Equations. In: Rogers, L. C. G., Talay, D. (eds.), *Numerical Methods in Finance*, Cambridge: University Press, 232-244.
- Douglas, J., Ma, J., Protter, P. (1996) Numerical Methods for Forward Backward Stochastic Differential Equations. *Ann. Appl. Probab.*, **6**, 940-968.
- El Karoui, N., Peng, S., Quenez, M. C. (1997) Backward Stochastic Differential Equations in Finance. *Math. Finance*, **7**, 1-71.
- Gobet, E., Lemor, J.-P., Warin, X. (2004) A Regression-Based Monte-Carlo Method to Solve Backward Stochastic Differential Equations. *Ann. Appl. Probab.*, forthcoming.
- Lemor, J.-P., Gobet, E., Warin, X. (2005) Rate of Convergence of an Empirical Regression Method for Solving Generalized Backward Stochastic Differential Equations. Preprint.
- Longstaff, F. A., Schwartz, R. S. (2001) Valuing American Options by Simulation: A Simple Least-Square Approach. *Review of Financial Studies*, **14**, 113-147.
- Ma, J., Protter, P., San Martín, J., Soledad, S. (2002) Numerical Method for Backward Stochastic Differential Equations. *Ann. Appl. Probab.*, **12**, 302-316.
- Ma, J., Protter, P., Yong, J. (1994) Solving Forward-Backward Stochastic Differential Equations Explicitly – a Four Step Scheme. *Prob. Th. Rel. Fields*, **98**, 339-359.
- Milstein, G. N., Tretyakov, M. V. (2004) Numerical Algorithms for Forward-Backward Stochastic Differential Equations Connected with Semilinear Parabolic Equations. Preprint.
- Pardoux, E., Peng, S. (1990) Adapted Solutions of a Backward Stochastic Differential Equation. *Syst. Contr. Lett.*, **14**, 55-61.
- Yong, J., Zhou, X. Y. (2000) *Stochastic Controls: Hamiltonian Systems and HJB Equations*. Berlin: Springer.
- Zhang, J. (2001) *Some Fine Properties of Backward Stochastic Differential Equations*. PhD Thesis, Purdue University.
- Zhang, J. (2004) A Numerical Scheme for BSDEs. *Ann. Appl. Probab.*, **14**, 459-488.

Christian Bender  
 Weierstrass Institute for Applied  
 Analysis and Stochastics  
 Mohrenstr. 39  
 D-10117 Berlin  
 Germany  
 bender@wias-berlin.de

Robert Denk  
 Department of Mathematics  
 and Statistics  
 University of Konstanz  
 D-78457 Konstanz  
 Germany  
 robert.denk@uni-konstanz.de