Weierstraß-Institut
für Angewandte Analysis und Stochastik

im Forschungsverbund Berlin e.V.

Preprint ISSN 0946 – 8633

Forward Simulation of Backward SDEs

Christian Bender[1], Robert Denk[2]

submitted: May 4, 2005

[1] Weierstrass Institute
for Applied Analysis and Stochastics
Mohrenstr. 39
D-10117 Berlin
Germany
E-mail: bender@wias-berlin.de

[2] Department of Mathematics
University of Konstanz
D-78457 Konstanz
Germany
E-mail: robert.denk@uni-konstanz.de

No. 1026

Berlin 2005

2000 *Mathematics Subject Classification.* 65C05; 65C30; 91B28.

Key words and phrases. BSDE, Numerics, Monte-Carlo simulation, Picard iteration, Finance.

Supported by the DFG Research Center MATHEON 'Mathematics for key technologies' in Berlin and the AFF grant 28/04 of the University of Konstanz.

Edited by
Weierstraß-Institut für Angewandte Analysis und Stochastik (WIAS)
Mohrenstraße 39
10117 Berlin
Germany

Fax: + 49 30 2044975
E-Mail: preprint@wias-berlin.de
World Wide Web: http://www.wias-berlin.de/

Abstract

We introduce a forward scheme to simulate backward SDEs and analyze the error of the scheme. Finally, we demonstrate the strength of the new algorithm by solving some financial problems numerically.

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. Only recently, a new notion of L^2-regularity on the control part of the solution was introduced in Zhang (2004), 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 slightly different algorithms.

A main drawback of the backward schemes is, that nestings of conditional expectations backwards through the time steps have to been 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 explodes linearly, when 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 Picard iterations.

In Section 2 we prove convergence of the discretized Picard iteration under quite general assumptions. In particular, we show that the additional error (compared to the backward scheme) due to the Picard iteration converges to zero at a geometric rate.

The error due to a generic approximation of the conditional expectation is analyzed in Section 3. We show that 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.

Section 4 is devoted to the development of 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, and prove convergence when the projection coefficients are substituted by their simulation-based analogues.

Finally, in Section 5, we present some simulations related to financial problems.

2 A Discretization of the Picard 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 $W_t = (W_{1,t}, \dots, W_{D,t})^*$ 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 Φ is a functional on the space of RCLL-functions on $[0,T]$ satisfying the L^∞-Lipschitz condition,

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

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

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

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

Note, that we do neither assume that the matrix σ 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 Picard iteration, which is explicit but for the occurrence of conditional expectations.

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

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

The processes $Y^{(n,\pi)}$ and $Z^{(n,\pi)}$ are extended to RCLL processes by constant interpolation. Note that the discretized Picard 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 iteration:

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

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

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

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

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 $|\pi| \le 1$

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

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

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

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

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

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

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

It exists, when the mesh $|\pi|$ of the partition π is sufficiently fine. Again, the processes $Y^{(\infty,\pi)}$ and $Z^{(\infty,\pi)}$ are extended to RCLL processes by constant interpolation. Note, $(Y^{(\infty,\pi)}, Z^{(\infty,\pi)})$ is (up to the interpolation of the Z-part) the backward scheme considered in Bouchard and Touzi (2004).

We 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-depending 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*

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

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

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

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 Picard iteration for a fixed partition. Our aim is to derive rates of convergence uniform in π.

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

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

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 Theorem 2.6 holds, for instance, for $|\pi| \le \Gamma$ with*

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

where

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

We prepare the proof with a technical lemma.

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

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

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

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

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

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

First note that

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

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

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

Step 1: For any $1 \le d \le D$

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

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

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

Thus, by (2),

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

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

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

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

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

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

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

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

Final Step: The assertion follows from a straightforward combination of (4) and (5). ∎

Proof of theorem 2.6. Denote,

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

By Lemma 2.8,

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

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

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

Recalling the definition of λ_i from Lemma 2.8 we have,

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

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 Cauchy and thus converges to $(Y^{(\infty,\pi)}, Z^{(\infty,\pi)})$ (when $|\pi|$ is sufficiently small) by means of (2)–(3). Moreover, for $n \in \mathbb{N}$,

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

It remains to prove a uniform bound for

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

which is given in the following lemma. ∎

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 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 $|\pi| \leq 1$,

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

Analogously to step 1 in Lemma 2.8 we obtain,

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

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

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 $|\pi|$,

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

with a constant C depending on the data only. Lemma 2.9 concludes. ■

3 Generic Analysis of the Error Propagation

To numerically implement the discretized Picard 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).

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

Bouchard and Touzi (2004), Theorem 4.1, prove, under slightly stronger assumptions than Assumption 2.1, that

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

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 $Y^{(\infty,\pi)}_t$ 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

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

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 π,*

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

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,

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

Lemma 2.9 can be applied to the second term. Hence, with a suitable choice of Γ and γ,

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

Now for $|\pi|$ sufficiently small (e.g. less or equal $(4\Gamma)^{-1}$) the above estimate can be iterated to obtain the theorem. Note, $1 \le \lambda_i \le e^{\Gamma T}$. Thus, we can choose $C = 2e^{\Gamma T} \vee \Gamma$. ∎

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 $\Lambda_{d,i}$, $0 \le d \le D$, of $L^2(\mathcal{F}_{t_i})$ for each $0 \le i \le k$. The orthogonal projection on $\Lambda_{d,i}$ is denoted by $P_{d,i}$.

We now consider the algorithm

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

initiated 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 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*

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

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

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

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 is norm contracting 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:

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

for any γ, $\Gamma > 0$ with $\lambda_0 = 1$ and $\lambda_i = (1 + \Gamma\Delta_{i-1})\lambda_{i-1}$. We choose $\gamma = 4DK^2(T+1)(1+1/T)$ and $\Gamma = 4K^2(T+1)(\gamma DT+1)(1+1/T)$. Since, due to the orthogonality of the orthogonal projection,

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

we get

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

Iterating this inequality and applying the orthogonality of the orthogonal projection once more (with $\nu = n$) gives the claim. (Note, $1 \le \lambda_i \le e^{\Gamma T}$. Thus, we can choose $C = e^{\Gamma T} \vee \Gamma$.) ■

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*

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

provided $|\pi|$ is sufficiently small.

Proof. This assertion directly follows from Corollary 2.10 and Theorem 4.1, because the orthogonal projection is norm-contracting. ∎

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 plain, 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

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

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

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

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

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

where (with componentwise evaluation of the expectation and an obvious notation)

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

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 $(\Delta W_i^{(\lambda)}, \xi^{(\pi,\lambda)}, X_{t_i}^{(\pi,\lambda)}, \eta_k^{(i,\lambda)}, \tilde{\eta}_k^{(i,\lambda)})$, $\lambda = 1, \ldots, L$, of $(\Delta W_i, \xi^{(\pi)}, X_{t_i}^{(\pi)}, \eta_k^i, \tilde{\eta}_k^i)$. We define

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

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

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

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 $(\mathcal{A}_i^L)^+$, $(\widetilde{\mathcal{A}}_i^L)^+$ to define simulation-based analogues of (9) recursively by:

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

The simulation based estimators are now defined by,

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

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* $(\alpha_{i,k}^{(n,\pi,L)}, \widetilde{\alpha}_{i,k}^{(n,\pi,L)})$ *converges P-almost surely to* $(\alpha_{i,k}^{(n,\pi)}, \widetilde{\alpha}_{i,k}^{(n,\pi)})$*, when L tends to infinity.*

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 $\widetilde{\alpha}_{i,k}^{(n,\pi,L)}$, the argument for $\alpha_{i,k}^{(n,\pi,L)}$ is similar. First observe that by the law of large numbers

$$\lim_{L\to 0} \widetilde{\mathcal{B}}_i^L = \widetilde{\mathcal{B}}_i; \quad P\text{-a.s.} \tag{10}$$

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

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

Hence,

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

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

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

Define

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

By the law of large numbers,

$$\begin{aligned}&\frac{1}{L}\sum_{\lambda=1}^{L}\widetilde{\eta}_l^{(i,\lambda)}\frac{\Delta W_i^{(\lambda)}}{\Delta_i}\left(\xi^{(\pi,\lambda)} - \sum_{j=i+1}^{N-1} f(t_j, X_{t_j}^{(\pi,\lambda)}, \widehat{Y}_{t_j}^{(n-1,\pi,\lambda)}, \widehat{Z}_{t_j}^{(n-1,\pi,\lambda)})\Delta_j\right) \\ \to\ & E\left[\widetilde{\eta}_l^i\frac{\Delta W_i}{\Delta_i}\left(\xi^{(\pi)} - \sum_{j=i+1}^{N-1} f(t_j, X_{t_j}^{(\pi)}, \widehat{Y}_{t_j}^{(n-1,\pi)}, \widehat{Z}_{t_j}^{(n-1,\pi)})\Delta_j\right)\right]; P\text{-a.s.}\end{aligned} \tag{12}$$

Moreover,

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

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)–(12) the proof is complete. ■

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* $(\widehat{Y}_{t_i}^{(n,\pi)}, \widehat{Z}_{t_i}^{(n,\pi)})$*, 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 is a positive constant* c *depending on the data such that for sufficiently small* $|\pi|$

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

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 $|\widehat{Z}_{t_i}^{(n,\pi)}|^2\Delta_i$. Hence, in view of Corollary 4.2, the proof is complete. ■

Definition 4.7 *We call a pair (c,ρ) a* truncation pair, *if c satisfies the estimates of Lemma 4.6, and $\rho : \mathbb{R} \to \mathbb{R}$ is a Lipschitz continuous function with constant 1, bounded by 2, which coincides with the identity on $[-1,1]$.*

By Lemma 4.6 we have for every truncation pair,

$$\begin{aligned} \widehat{Y}_{t_i}^{(n,\pi)} &= c\lambda_{\min}(\mathcal{B}_i)^{-1/2}|\eta^i|\rho\left(\frac{\widehat{Y}_{t_i}^{(n,\pi)}}{c\lambda_{\min}(\mathcal{B}_i)^{-1/2}|\eta^i|}\right) \\ \widehat{Z}_{t_i}^{(n,\pi)} &= \frac{1}{\sqrt{\Delta_i}}c\lambda_{\min}(\widetilde{\mathcal{B}}_i)^{-1/2}|\widetilde{\eta}^i|\rho\left(\frac{\sqrt{\Delta_i}\widehat{Z}_{t_i}^{(n,\pi)}}{c\lambda_{\min}(\widetilde{\mathcal{B}}_i)^{-1/2}|\widetilde{\eta}^i|}\right) \end{aligned} \tag{13}$$

This motivates to define (c,ρ)-truncations of $(Y_{t_i}^{(n,\pi,L,*)}, Z_{t_i}^{(n,\pi,L,*)})$ by

$$\begin{aligned} Y_{t_i}^{(n,\pi,L,\rho)} &= c\lambda_{\min}(\mathcal{B}_i)^{-1/2}|\eta^i|\rho\left(\frac{Y_{t_i}^{(n,\pi,L,*)}}{c\lambda_{\min}(\mathcal{B}_i)^{-1/2}|\eta^i|}\right) \\ Z_{t_i}^{(n,\pi,L,\rho)} &= \frac{1}{\sqrt{\Delta_i}}c\lambda_{\min}(\widetilde{\mathcal{B}}_i)^{-1/2}|\widetilde{\eta}^i|\rho\left(\frac{\Delta_i Z_{t_i}^{(n,\pi,L,*)})}{c\lambda_{\min}(\widetilde{\mathcal{B}}_i)^{-1/2}|\widetilde{\eta}^i|}\right) \end{aligned} \tag{14}$$

An immediate consequence of the dominated convergence theorem is the L^2-convergence of the truncated estimators.

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

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

Remark 4.9 *We conjecture that, possibly with a more sophisticated truncation, $\frac{1}{\sqrt{L}}$ can be derived as rate of convergence in the above theorem. This issue will be addressed in our future research.*

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 $\max_{0\le j\le i} X_{t_j}^{(\pi)}$, $\min_{0\le j\le i} X_{t_j}^{(\pi)}$, 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 look-back 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 4.10 *(i) Suppose $\phi : \mathbb{R}^{2M} \to \mathbb{R}$ is Lipschitz-continuous. Then*

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

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

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

3. **Choice of the basis:** As for the basis on may choose a set of functions $\{e_1(x), \ldots, e_\kappa(x)\}$ and define the basis via

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

Typical choices are indicator functions or (exponentially damped) polynomials such as Hermite functions. 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

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

Hence, if $\{e_1(x), \ldots, e_\kappa(x)\}$ 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 $0 \le i \le N-1$, 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 κ tends to infinity. Hence, Theorems 2.3 and 4.8 provide L^2-convergence of the truncated algorithm (14) 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 π_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. $K = 100$. 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 0.0001. 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,

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

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 for a maturity of $T = 2$ years 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. 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 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 $N = 20$ and $T = 0.5$. The simulations have been performed with the monomial basis and $\kappa = 5$ for the nonlinear case.

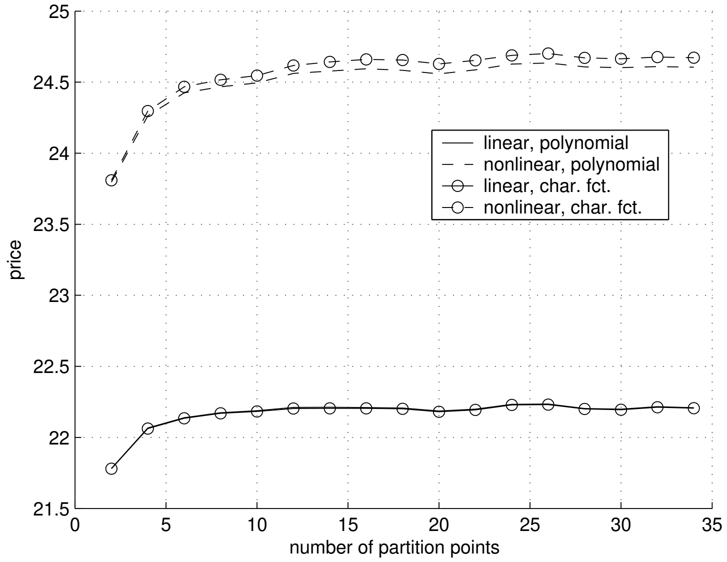

Figure 1: $Y_0^{(n_{stop},\pi_N,100000,*)}$ as a function of N for $T=2$.

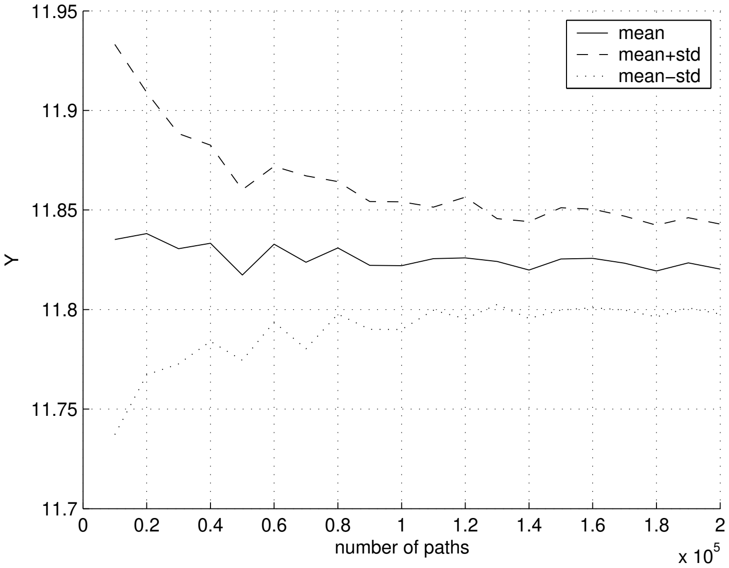

Figure 2: Empirical mean and standard deviation of 100 launches as function of L for $T=0.5$.

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 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 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^ϵ (as ϵ 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 8.058.

We compute numerical approximations for different values of ϵ. 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^ϵ as function of ϵ for different numbers of simulated paths.

Figure 3 indicates that, due to the nonlinearity, the estimator for the conditional expectation has a positive bias. Indeed, the simulated ϵ-approximation tend to merge into a straight line (as function of ϵ^{-1}), when ϵ (depending on the number of paths) is sufficient small. Since the curves for 100000 paths and 200000 paths are almost parallel this effect can not be mended by solely enlarging the number of simulations. Preliminary simulations suggest that a larger number of partition points, an enlarged basis, and the simulation of more paths are needed to obtain accurate approximations of the ϵ-price, the higher the penalization. To achieve this with reasonable computational cost, variance reduction techniques are called for. This issue is left to future research.

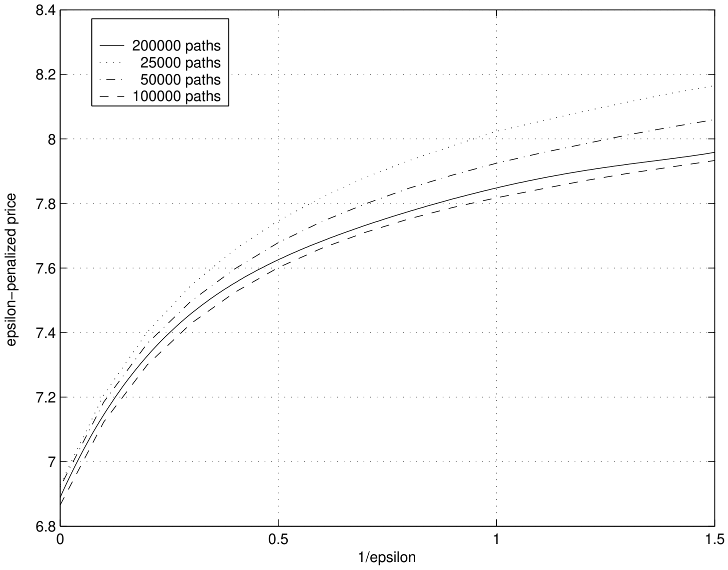

Figure 3: ϵ-approximation of the superhedging price as function of ϵ^{-1}.

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

$$Y^{(\infty,\pi)}_{t_{i+1}} = E\left[Y^{(\infty,\pi)}_{t_{i+1}} \middle| \mathcal{F}_{t_i} \right] + \int_{t_i}^{t_{i+1}} \tilde{Z}^{(\pi)}_s dW_s \tag{15}$$

via the martingale representation theorem. Then,

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

Squaring and taking expectation yields,

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

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,

$$
\begin{aligned}
& E\left[|Y_{t_i}^{(\infty,\pi)} - Y_{t_i}|^2\right] + E\int_{t_i}^{t_{i+1}} |\tilde{Z}_s^{(\pi)} - Z_s|^2 ds \\
\leq \quad & (1+\frac{\Delta_i}{\epsilon})E\left[|Y_{t_{i+1}}^{(\infty,\pi)} - Y_{t_{i+1}}|^2\right] + C(1+\frac{\epsilon}{\Delta_i})\Delta_i^3 \\
& +C(1+\frac{\epsilon}{\Delta_i})\left(\Delta_i^2 E\left[|Y_{t_i}^{(\infty,\pi)} - Y_{t_i}|^2\right] + \Delta_i E\int_{t_i}^{t_{i+1}} |Z_s - Z_{t_i}^{(\infty,\pi)}|^2 ds\right) \quad (16)
\end{aligned}
$$

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

$$
\widehat{Z}_{t_i}^{(\pi)} = \frac{1}{\Delta_i} E\left[\int_{t_i}^{t_{i+1}} Z_s ds \middle| \mathcal{F}_{t_i}\right]. \tag{17}
$$

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

$$
\sum_{i=0}^{N-1} E\int_{t_i}^{t_{i+1}} |Z_s - \widehat{Z}_{t_i}^{(\pi)}|^2 ds \leq C|\pi| \tag{18}
$$

Note also that by (15) and Itô's isometry,

$$
Z_{t_i}^{(\infty,\pi)} = \frac{1}{\Delta_i} E\left[\int_{t_i}^{t_{i+1}} \tilde{Z}_s^{(\pi)} ds \middle| \mathcal{F}_{t_i}\right]. \tag{19}
$$

The identities (17) and (19) can be easily combined to get,

$$\begin{aligned}
& E\int_{t_i}^{t_{i+1}} |Z_s - Z_{t_i}^{(\infty,\pi)}|^2 ds \\
\leq \; & 2\left(E\int_{t_i}^{t_{i+1}} |Z_s - \widehat{Z}_{t_i}^{(\pi)}|^2 ds + E\int_{t_i}^{t_{i+1}} |Z_s - \widetilde{Z}_s^{(\pi)}|^2 ds\right) \qquad (20)
\end{aligned}$$

We can now fix ϵ sufficiently small such that for small $|\pi|$ (combining (16) and (20)),

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

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

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

Thus,

$$\begin{aligned}
& E\left[|Y_{t_i}^{(\infty,\pi)} - Y_{t_i}|^2\right] + \frac{1}{2}E\int_{t_i}^{t_{i+1}} |\tilde{Z}_s^{(\pi)} - Z_s|^2 ds \\
\leq \; & (1+C\Delta_i)E\left[|Y_{t_{i+1}}^{(\infty,\pi)} - Y_{t_{i+1}}|^2\right] + C\Delta_i^2 + CE\int_{t_i}^{t_{i+1}} |Z_s - \widehat{Z}_{t_i}^{(\pi)}|^2 ds \qquad (21)
\end{aligned}$$

Step 2: Convergence of $Y^{(\infty,\pi)}$:
We may now conclude from (21), the discrete Gronwall lemma and (18) that

$$\begin{aligned}
& \max_{0\leq i\leq N} E\left[|Y_{t_i}^{(\infty,\pi)} - Y_{t_i}|^2\right] \\
\leq \; & C\left(E[|\xi - \xi^{(\pi)}|^2] + \sum_{i=0}^{N-1}\Delta_i^2 + E\int_{t_i}^{t_{i+1}} |Z_s - \widehat{Z}_{t_i}^{(\pi)}|^2\right) \\
\leq \; & C\left(E[|\xi - \xi^{(\pi)}|^2] + |\pi|\right) \qquad (22)
\end{aligned}$$

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 (21) from 0 to $N-1$ and obtain,

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

due to (18) and (22). By (19) and the mean-square minimizing property of the conditional expectation,

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

Elementary manipulations show that this is equivalent to

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

The estimate for $Z^{(\infty,\pi)}$ may now be easily derived from (18) and (23). ■

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.

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 Forwad 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. 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. Prob.*, **14**, 459-488.