DFG-Schwerpunktprogramm 1324

„Extraktion quantifizierbarer Information aus komplexen Systemen"

A-posteriori estimates for backward SDEs

C. Bender, J. Steiner

Preprint 120

Edited by

AG Numerik/Optimierung
Fachbereich 12 - Mathematik und Informatik
Philipps-Universität Marburg
Hans-Meerwein-Str.
35032 Marburg

DFG-Schwerpunktprogramm 1324

„Extraktion quantifizierbarer Information aus komplexen Systemen"

A-posteriori estimates for backward SDEs

C. Bender, J. Steiner

Preprint 120

A-posteriori estimates for backward SDEs

Christian Bender[1], Jessica Steiner[1]

April 25, 2012

Suppose an approximation to the solution of a backward SDE is pre-computed by some numerical algorithm. In this paper we provide a-posteriori estimates on the L^2-approximation error between true solution and approximate solution. These a-posteriori estimates provide upper and lower bounds for the approximation error. They can be expressed solely in terms of the approximate solution and the data of the backward SDE, and can be estimated consistently by simulation in typical situations. We also illustrate by some numerical experiments in the context of least-squares Monte Carlo how the a-posteriori estimates can be applied in practice.

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

1 Introduction

Backward stochastic differential equations (BSDEs) were first introduced by Bismut (1973) as adjoint equations in the stochastic maximum principle. Apart from their applications in stochastic control (for which we refer e.g. to Yong and Zhou, 1999), BSDEs turned out to be an extremely useful tool in mathematical finance (see e.g. the survey paper by El Karoui et al., 1997). Moreover, Feynman-Kac representation formulas for partial differential equations can be generalized via BSDEs, as shown e.g. by Pardoux and Peng (1992). Motivated by these applications, many numerical algorithms for backward stochastic differential equations (BSDEs) were developed during the last years. They typically consist of two steps. In the first step a time discretization is performed. The most popular choice for the time discretization is an Euler-type discretization for BSDEs, which was introduced by Zhang (2004) and Bouchard and Touzi (2004). The corresponding discrete BSDE still requires to evaluate high order nestings of conditional expectations, which in general cannot be computed in closed form. Several techniques have been suggested to approximate these conditional expectations. Among those are Malliavin Monte Carlo (Bouchard and Touzi, 2004), least-squares Monte Carlo (Lemor et al., 2006), quantization (Bally and Pagès, 2003), and cubature on Wiener

1 Saarland University, Department of Mathematics, PO Box 151150, D-66041 Saarbrücken, Germany. bender@math.uni-sb.de, steiner@math.uni-sb.de.

space (Crisan and Manolarakis, 2010). For more information on the rich literature on numerical schemes for BSDEs we also refer to the survey part in Bender and Steiner (2012).

The quality of the different numerical approximations mentioned above is typically difficult to judge and compare apart from rare situations in some test examples where the true solution is known in closed form. As a way-out, we suggest an a-posteriori error criterion in this paper, which can be expressed in terms of the approximate solution (generated by whatever algorithm) on a given time grid and the data of the BSDE. The idea of the criterion is to check how close the approximate solution is to solving the corresponding SDE run forward in time on the grid and how well it approximates the terminal condition. In Section 2, we show that this error criterion is equivalent to the squared L^2-error between the approximate solution and the unknown implicit backward Euler time discretization of the true solution on the grid. From this point of view the criterion measures in the first place how successful the numerical procedure for approximating the conditional expectation performs, once the time discretization is already done.

We then also study to which extent the time discretization error can be measured by the a-posteriori error criterion. It is well known by the results of Zhang (2004) that the discretization error between true solution and implicit time discretization heavily depends on the L^2-regularity of the Z-part of the solution of the BSDE. We show that the term which is related to this regularity can be estimated by our error criterion under rather weak conditions. In this way we are able to derive upper and lower bounds of the L^2-error between an approximate solution (generated by whatever algorithm) and the true continuous time solution in Section 3. We merely need to assume Lipschitz continuity of the driver and L^2-integrability of the terminal condition. No further regularity condition needs to be imposed on the terminal condition such as the popular assumption that the terminal condition is a Lipschitz functional on the path of a diffusion process. Several examples explain how our generic a-posteriori estimates can be applied under typical assumptions in the literature on numerical approximations of BSDEs including the situation of irregular terminal conditions or the situation of smooth data.

Finally we apply the a-posteriori criterion to a test example in which we compute the conditional expectations numerically by the least-squares Monte Carlo algorithm which was made popular in the context of American option pricing by Longstaff and Schwartz (2001) and was studied for BSDEs in Lemor et al. (2006). We explain how the a-posteriori error criterion can help to tailor the algorithm concerning the basis choice, the number of time points and the number of simulation paths. In particular in situations when the theoretical worst-case error estimates by Lemor et al. (2006) lead to prohibitive simulation costs in practice, our criterion can be applied to justify the use of a more moderate sample size or a small function basis.

This paper is organized as follows: In Section 2 we study the a-posteriori error criterion for discrete time BSDEs. The continuous time case is treated in Section 3, while the numerical example is discussed in Section 4.

2 A-posteriori estimates for discrete BSDEs

In this section we provide a-posteriori estimates for some kind of discrete BSDEs living on a time grid $\pi = \{t_0, \dots, t_N\}$ which satisfies $0 = t_0 < t_1 < \cdots < t_N = T$. We assume that $(\Omega, \mathcal{F}, (\mathcal{G}_t)_{t\in[0,T]}, P)$ is a filtered probability space which carries a D-dimensional Brownian motion $W_t = (W_t^1, \dots W_t^D)^*$, (the star denoting transposition). The filtration $\mathcal{G}_t$ can in general be larger than the augmented filtration generated by the Brownian motion W which we denote by $\mathcal{F}_t$. The time increments and the increments of the Brownian motion on the time grid π will be abbreviated by $\Delta_i = t_{i+1} - t_i$ and $\Delta W_i = W_{t_{i+1}} - W_{t_i}$. $E_i[\cdot]$ stands for the conditional expectation $E[\cdot|\mathcal{G}_{t_i}]$ with respect to the larger filtration.

The type of discrete BSDE, which we consider, is of the form

$$\begin{aligned} Y_{t_i}^\pi &= Y_{t_{i+1}}^\pi - f^\pi(t_i, Y_{t_i}^\pi, \Delta_i^{-1} E_i[\Delta W_i^* M_{t_{i+1}}^\pi])\Delta_i - (M_{t_{i+1}}^\pi - M_{t_i}^\pi), \quad i = 0, \dots, N-1, \\ Y_{t_N}^\pi &= \xi^\pi. \end{aligned} \tag{1}$$

Here the data (f^π, ξ^π) is given and a solution consists of a pair $(Y_{t_i}^\pi, M_{t_i}^\pi)$ of square integrable, $\mathcal{G}_{t_i}$-adapted processes such that M^π is a $(\mathcal{G}_{t_i})$-martingale starting in 0 and (1) is satisfied. Concerning the data we assume:

($\mathbf{H}_\pi^\mathcal{G}$) ξ^π is a square-integrable $\mathcal{G}_T$-measurable random variable. $f^\pi : \pi \times \Omega \times \mathbb{R} \times \mathbb{R}^D \to \mathbb{R}$ is measurable and, for every $t_i \in \pi$ and $(y, z) \in \mathbb{R}^{D+1}$, $f^\pi(t_i, y, z)$ is $\mathcal{G}_{t_i}$-measurable. Moreover f^π is Lipschitz in (y, z) with constant K (independent of π) uniformly in (t_i, ω) and $f^\pi(t_i, 0, 0)$ is square-integrable for every $t_i \in \pi$.

It follows immediately from (1), that the two solution processes (Y^π, M^π) are connected via the property

$$M_{t_{i+1}}^\pi - M_{t_i}^\pi = Y_{t_{i+1}}^\pi - E_i[Y_{t_{i+1}}^\pi]. \tag{2}$$

Under ($\mathrm{H}_\pi^\mathcal{G}$) it is straightforward to check by a contraction mapping argument that this discrete BSDE admits a unique solution if the mesh $|\pi| = \max\{|t_{i+1} - t_i|;\ i = 0, \dots, N-1\}$ of π is sufficiently small. In fact, denoting

$$Z_{t_i}^\pi = \Delta_i^{-1} E_i[\Delta W_i^* M_{t_{i+1}}^\pi] \tag{3}$$

and in view of (2), the discrete BSDE (1) can be rewritten in terms of (Y^π, Z^π) as

$$\begin{aligned} Y_{t_N}^\pi &= \xi^\pi, \quad Z_{t_N}^\pi = 0, \\ Z_{t_i}^\pi &= E_i[\Delta W_i^* Y_{t_{i+1}}^\pi], \quad i = N-1, \dots, 0, \\ Y_{t_i}^\pi &= E_i[Y_{t_{i+1}}^\pi] - f^\pi(t_i, Y_{t_i}^\pi, Z_{t_i}^\pi)\Delta_i, \quad i = N-1, \dots, 0. \end{aligned} \tag{4}$$

So the discrete BSDE in (1) is just a reformulation of the implicit time discretization scheme studied e.g. by Bouchard and Touzi (2004). It turns out that this reformulation is more convenient for our purposes.

Let us suppose now that some approximation $(\hat{Y}^\pi, \hat{M}^\pi)$ of (Y^π, M^π) is at hand. We merely assume that $(\hat{Y}^\pi, \hat{M}^\pi)$ is square-integrable, adapted to $\mathcal{G}_{t_i}$ and that $\hat{M}^\pi$ is a martingale starting in 0. No further assumptions on the algorithm which was used to compute $(\hat{Y}^\pi, \hat{M}^\pi)$ are necessary at this stage. Our purpose is to obtain quantifiable information about the L^2-error between the approximation $(\hat{Y}^\pi, \hat{M}^\pi)$ and the solution (Y^π, M^π), which only requires knowledge of the approximation and the data (f^π, ξ^π).

The basic idea is to check how close $(\hat{Y}^\pi, \hat{M}^\pi)$ is to solving (1) instead of checking how close it is to the solution of (1). This leads to the following error criterion:

$$\mathcal{E}_\pi(\hat{Y}^\pi, \hat{M}^\pi) = E_0[|\xi^\pi - \hat{Y}^\pi_{t_N}|^2] + \max_{1\le i\le N} \mathrm{E}_0 \left[\left| \hat{Y}^\pi_{t_i} - \hat{Y}^\pi_{t_0} - \sum_{j=0}^{i-1} f^\pi(t_j, \hat{Y}^\pi_{t_j}, \Delta_j^{-1} E_j[\Delta W_j^* \hat{M}^\pi_{t_{j+1}}])\Delta_j - \hat{M}^\pi_{t_i} \right|^2 \right] \tag{5}$$

The first term is just the squared L^2-error of the approximation at terminal time. The second term measures whether the approximation is close to solving the difference equation corresponding to (1) run forward in time.

The main result of this section shows that this error criterion is equivalent to the squared L^2-error between $(\hat{Y}^\pi, \hat{M}^\pi)$ and (Y^π, M^π).

Theorem 2.1. *Assume ($H^{\mathcal{G}}_\pi$). Then there are constants $c, C > 0$ such that for sufficiently small $|\pi|$ and for every pair of square-integrable $\mathcal{G}_{t_i}$-adapted processes $(\hat{Y}^\pi_{t_i}, \hat{M}^\pi_{t_i})$, where $\hat{M}^\pi$ is a martingale starting in 0, it holds that*

$$\frac{1}{c}\mathcal{E}_\pi(\hat{Y}^\pi, \hat{M}^\pi) \le \max_{t_i\in\pi} \mathrm{E}_0 \left[|Y^\pi_{t_i} - \hat{Y}^\pi_{t_i}|^2 \right] + \mathrm{E}_0 \left[|M^\pi_{t_N} - \hat{M}^\pi_{t_N}|^2 \right] \le C\mathcal{E}_\pi(\hat{Y}^\pi, \hat{M}^\pi)$$

More precisely, with the choice $c = (6 + D)(1 + K^2T(1 + T))$ and

$$C = \left(\frac{6}{(D \vee 2) - 1} + 2 + 16(2 + 4(1 + T)TK^2(D \vee 2)) \right) e^{\Gamma T},$$

where

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

the inequalities hold for $|\pi| \le \Gamma^{-1}$.

Before we present the proof of this theorem we explain by some generic examples how this result can be applied in practice.

Example 2.2. *a) In this example we assume that there is a Markovian process $(X^\pi_{t_i}, \mathcal{F}_{t_i})$ taking values in $\mathbb{R}^M$ and deterministic functions y^π, z^π such that the solution (Y^π, Z^π) of (4) is of the form*

$$Y^\pi_{t_i} = y^\pi(t_i, X^\pi_{t_i}), \quad Z^\pi_{t_i} = z^\pi(t_i, X^\pi_{t_i}).$$

We recall here that $(\mathcal{F}_t)_{t\in[0,T]}$ is the augmented Brownian filtration, but it could also be replaced by a larger filtration $(\mathcal{H}_t)_{t\in[0,T]}$ with respect to which W is still a Brownian motion. Numerical algorithms for (discrete) BSDEs typically try to approximate the pair of functions $(y^\pi,\ z^\pi)$. We assume that the approximation of $(y^\pi,\ z^\pi)$ is of the form

$$\hat{y}^\pi(t_i, x; \Xi), \quad \hat{z}^\pi(t_i, x; \Xi),$$

where Ξ is a random vector independent of $\mathcal{F}_T$. This form covers Monte Carlo algorithms such as least-squares Monte Carlo where the approximative functions $(\hat{y}^\pi, \hat{z}^\pi)$ depend on independently *simulated sample paths of X^π which we can collect in Ξ. We now define the filtration $\mathcal{G}_t$ by enlarging the Brownian filtration with the random vector Ξ, i.e. $\mathcal{G}_t = \mathcal{F}_t \vee \sigma(\Xi)$. Given $(\hat{y}^\pi, \hat{z}^\pi)$, we consider*

$$\hat{Y}^\pi_{t_i} = \hat{y}^\pi(t_i, X^\pi_{t_i}; \Xi), \quad \hat{Z}^\pi_{t_i} = \hat{z}^\pi(t_i, X^\pi_{t_i}; \Xi)$$

as approximations of (Y^π, Z^π). An approximation for the martingale M^π can be defined in terms of $\hat{Z}^\pi$ by

$$\hat{M}^\pi_0 = 0, \quad \hat{M}^\pi_{t_{i+1}} - \hat{M}^\pi_{t_i} = \hat{z}^\pi(t_i, X^\pi_{t_i}; \Xi)\Delta W_i,$$

which clearly is a martingale with respect to the larger filtration $\mathcal{G}_{t_i}$, but, in general, not with respect to $\mathcal{F}_{t_i}$. Then,

$$\hat{Z}^\pi_{t_i} = \Delta_i^{-1} E_i[\Delta W_i^* \hat{M}^\pi_{t_{i+1}}],$$

and, consequently, the error criterion can be rewritten in terms of $(\hat{Y}^\pi, \hat{Z}^\pi)$ (replacing $\hat{M}^\pi$ by $\hat{Z}^\pi$ on the left hand side with a slight abuse of notation) as

$$\mathcal{E}_\pi(\hat{Y}^\pi, \hat{Z}^\pi) = E_0[|\xi^\pi - \hat{Y}^\pi_{t_N}|^2] + \max_{1 \le i \le N} \mathrm{E}_0 \left[\left| \hat{Y}^\pi_{t_i} - \hat{Y}^\pi_{t_0} - \sum_{j=0}^{i-1} f^\pi(t_j, \hat{Y}^\pi_{t_j}, \hat{Z}^\pi_{t_j})\Delta_j - \sum_{j=0}^{i-1} \hat{Z}^\pi_{t_j} \Delta W_j \right|^2 \right]$$

If an algorithm for simulating independent sample paths of $(X^\pi_{t_i}, \Delta W_i)_{i=0\ldots,N}$ is at hand, we can draw independent copies of $(\hat{Y}^\pi_{t_i}, \hat{Z}^\pi_{t_i}, \Delta W_i)_{i=0\ldots,N}$ given Ξ. Now suppose that the data (f^π, ξ^π) is sufficiently good such that, given a realization of Ξ, we are even able to draw Λ independent copies

$$({}_\lambda\hat{Y}^\pi_{t_i}, {}_\lambda\hat{Z}^\pi_{t_i}, f^\pi(t_i, {}_\lambda\hat{Y}^\pi_{t_i}, {}_\lambda\hat{Z}^\pi_{t_i}), \Delta\, {}_\lambda W_i, {}_\lambda\xi^\pi;\ i = 0, \ldots N)_{\lambda=1,\ldots,\Lambda}$$

of $(\hat{Y}^\pi_{t_i}, \hat{Z}^\pi_{t_i}, f^\pi(t_i, \hat{Y}^\pi_{t_i}, \hat{Z}^\pi_{t_i}), \Delta W_i, \xi^\pi;\ i = 0, \ldots N)$. Then, $\mathcal{E}_\pi(\hat{Y}^\pi, \hat{Z}^\pi)$ can be estimated consistently by

$$\begin{aligned} &\hat{\mathcal{E}}^\Lambda_\pi(\hat{Y}^\pi, \hat{Z}^\pi) \\ = \ &\frac{1}{\Lambda} \sum_{\lambda=1}^{\Lambda} |\, {}_\lambda\xi^\pi - {}_\lambda\hat{Y}^\pi_{t_N}|^2 \\ &+ \max_{i=1,\ldots,N} \frac{1}{\Lambda} \sum_{\lambda=1}^{\Lambda} \left| {}_\lambda\hat{Y}^\pi_{t_i} - {}_\lambda\hat{Y}^\pi_{t_0} - \sum_{j=0}^{i-1} f^\pi(t_j, {}_\lambda\hat{Y}^\pi_{t_j}, {}_\lambda\hat{Z}^\pi_{t_j})\Delta_j - \sum_{j=0}^{i-1} {}_\lambda\hat{Z}^\pi_{t_j} \Delta\, {}_\lambda W_j \right|^2. \end{aligned} \tag{6}$$

In such situation we can, thanks to Theorem 2.1, consistently estimate upper and lower bounds for the L^2-approximation error between the given numerical solution $(\hat{Y}^\pi, \hat{M}^\pi)$ and the true solution (Y^π, M^π) of the discrete BSDE (1).

b) Suppose that we are in the situation of part a) of this example, but that, additionally, $\hat{z}^\pi$ and $\hat{y}^\pi$ are linked via

$$\hat{z}^\pi(t_i, x; \Xi) = E\left[\frac{\Delta W_i}{\Delta_i} \hat{y}^\pi(t_{i+1}, X^\pi_{t_{i+1}}; \Xi) \middle| \Xi,\, X^\pi_{t_i} = x \right]$$

and

$$E\left[\hat{y}^\pi(t_{i+1}, X^\pi_{t_{i+1}}; \Xi) \middle| \Xi,\, X^\pi_{t_i} = x \right]$$

can be calculated in closed form. This is e.g. the case in the martingale basis variant of least-squares Monte Carlo proposed by Bender and Steiner (2012). Thanks to the Markov property of $(X^\pi_{t_i}, \mathcal{F}_{t_i})$ we can then define a $\mathcal{G}_{t_i}$-martingale $\hat{M}^\pi$ via

$$\hat{M}^\pi_0 = 0, \quad \hat{M}^\pi_{t_{i+1}} - \hat{M}^\pi_{t_i} = \hat{y}^\pi(t_{i+1}, X^\pi_{t_{i+1}}; \Xi) - E\left[\hat{y}^\pi(t_{i+1}, X^\pi_{t_{i+1}}; \Xi) \middle| \Xi,\, X^\pi_{t_i}\right].$$

Then, again,

$$\hat{Z}^\pi_{t_i} = \hat{z}^\pi(t_i, X^\pi_{t_i}; \Xi) = \Delta_i^{-1} E_i[\Delta W_i^* \hat{M}^\pi_{t_{i+1}}],$$

and the error criterion becomes

$$\mathcal{E}_\pi(\hat{Y}^\pi, \hat{M}^\pi) = E_0[|\xi^\pi - \hat{Y}^\pi_{t_N}|^2] + \max_{1\le i\le N} \mathrm{E}_0\left[\left|\hat{Y}^\pi_{t_i} - \hat{Y}^\pi_{t_0} - \sum_{j=0}^{i-1} f^\pi(t_j, \hat{Y}^\pi_{t_j}, \hat{Z}^\pi_{t_j})\Delta_j - \hat{M}^\pi_{t_i}\right|^2\right].$$

Given a sampling procedure for $(X^\pi_{t_i})_{i=0,\dots,N}$ we can now, conditional on Ξ, sample independent copies of $(\hat{Y}^\pi_{t_i}, \hat{Z}^\pi_{t_i}, \hat{M}^\pi_{t_i})_{i=0\dots,N}$. Assuming again that the data is good enough and so Λ independent copies

$$({}_\lambda\hat{Y}^\pi_{t_i},\, {}_\lambda\hat{Z}^\pi_{t_i}, f^\pi(t_i,\, {}_\lambda\hat{Y}^\pi_{t_i},\, {}_\lambda\hat{Z}^\pi_{t_i}),\, {}_\lambda\hat{M}^\pi_{t_i},\, {}_\lambda\xi^\pi;\ i = 0, \dots N)_{\lambda=1,\dots,\Lambda}$$

of $(\hat{Y}^\pi_{t_i}, \hat{Z}^\pi_{t_i}, f^\pi(t_i, \hat{Y}^\pi_{t_i}, \hat{Z}^\pi_{t_i}), \hat{M}^\pi_{t_i}, \xi;\ i = 0, \dots N)$ are at hand, we can estimate $\mathcal{E}_\pi(\hat{Y}^\pi, \hat{M}^\pi)$ analogously to part a) of this example.

We now give the proof of Theorem 2.1.

Proof of Theorem 2.1. Notice first that the condition on the mesh size $|\pi|$ ensures that a unique solution (Y^π, M^π) to the discrete BSDE (1) exists, see e.g. Theorem 5 and Remark 6 in Bender and Denk (2007).

We first prove the easier lower bound

$$\mathcal{E}_\pi(\hat{Y}^\pi, \hat{M}^\pi) \le c\left(\max_{t_i\in\pi} \mathrm{E}_0\left[|Y^\pi_{t_i} - \hat{Y}^\pi_{t_i}|^2\right] + \mathrm{E}_0\left[|M^\pi_{t_N} - \hat{M}^\pi_{t_N}|^2\right]\right). \tag{7}$$

In order to simplify the notation we set

$$\hat{Z}^\pi_{t_i} = \Delta_i^{-1} E_i[\Delta W_i^* \hat{M}^\pi_{t_{i+1}}].$$

Then,

$$\begin{aligned}\mathcal{E}_\pi(\hat{Y}^\pi, \hat{M}^\pi) &= E_0[|\xi^\pi - \hat{Y}^\pi_{t_N}|^2] + \max_{1\le i\le N} \mathrm{E}_0\left[\left|\hat{Y}^\pi_{t_i} - \hat{Y}^\pi_{t_0} - \sum_{j=0}^{i-1} f^\pi(t_j, \hat{Y}^\pi_{t_j}, \hat{Z}^\pi_{t_j})\Delta_j - \hat{M}^\pi_{t_i}\right|^2\right] \\ &=: (A) + \max_{i=1,\dots,N} (B_i).\end{aligned}$$

By (1) and (3) we get

$$Y^\pi_{t_i} - Y^\pi_{t_0} - \sum_{j=0}^{i-1} f^\pi(t_j, Y^\pi_{t_j}, Z^\pi_{t_j})\Delta_j - M^\pi_{t_i} = 0.$$

Hence, by Young's inequality, the Lipschitz condition on f^π and the martingale property of $M^\pi - \hat{M}^\pi$, we have for every $\gamma > 0$,

$$\begin{aligned}
(B_i) &= \mathrm{E}_0\left[\left|\hat{Y}^\pi_{t_i} - Y^\pi_{t_i} - \hat{Y}^\pi_{t_0} + Y^\pi_{t_0} - \sum_{j=0}^{i-1}[f^\pi(t_j,\hat{Y}^\pi_{t_j},\hat{Z}^\pi_{t_j}) - f^\pi(t_j,Y^\pi_{t_j},Z^\pi_{t_j})]\Delta_j - \hat{M}^\pi_{t_i} + M^\pi_{t_i}\right|^2\right] \\
&\le \frac{24}{5}(1+\gamma)\max_{t_j\in\pi}\mathrm{E}_0\left[|Y^\pi_{t_j} - \hat{Y}^\pi_{t_j}|^2\right] + 6(1+\gamma)\mathrm{E}_0\left[|M^\pi_{t_i} - \hat{M}^\pi_{t_i}|^2\right] \\
&\quad + (1+\gamma^{-1})T\sum_{j=0}^{N-1}\Delta_j\mathrm{E}_0\left[\left|f^\pi(t_j,\hat{Y}^\pi_{t_j},\hat{Z}^\pi_{t_j}) - f^\pi(t_j,Y^\pi_{t_j},Z^\pi_{t_j})\right|^2\right] \\
&\le 5(1+\gamma)\max_{t_j\in\pi}\mathrm{E}_0\left[|Y^\pi_{t_j} - \hat{Y}^\pi_{t_j}|^2\right] + 6(1+\gamma)\mathrm{E}_0\left[|M^\pi_{t_N} - \hat{M}^\pi_{t_N}|^2\right] \\
&\quad + (1+\gamma^{-1})T(1+T)K^2\left(\max_{t_j\in\pi}\mathrm{E}_0\left[|Y^\pi_{t_j} - \hat{Y}^\pi_{t_j}|^2\right] + \sum_{j=1}^{N-1}\mathrm{E}_0\left[|Z^\pi_{t_j} - \hat{Z}^\pi_{t_j}|^2\Delta_j\right]\right).
\end{aligned}$$

As, by the definition of Z^π and $\hat{Z}^\pi$ and the martingale property of M^π and $\hat{M}^\pi$

$$\begin{aligned}
&\sum_{j=1}^{N-1}\mathrm{E}_0\left[|Z^\pi_{t_j} - \hat{Z}^\pi_{t_j}|^2\right]\Delta_j = \sum_{j=1}^{N-1}\frac{1}{\Delta_j}\mathrm{E}_0\left[\left|\mathrm{E}_j\left[\Delta W_j^*(M^\pi_{t_{j+1}} - \hat{M}^\pi_{t_{j+1}} - M^\pi_{t_j} + \hat{M}^\pi_{t_j})\right]\right|^2\right] \\
\le\ & D\sum_{j=1}^{N-1}\left(\mathrm{E}_0\left[|M^\pi_{t_{j+1}} - \hat{M}^\pi_{t_{j+1}}|^2\right] - \mathrm{E}_0\left[|M^\pi_{t_j} - \hat{M}^\pi_{t_j}|^2\right]\right) = D\mathrm{E}_0\left[|M^\pi_{t_N} - \hat{M}^\pi_{t_N}|^2\right], \qquad (8)
\end{aligned}$$

we immediately get

$$\mathcal{E}_\pi(\hat{Y}^\pi,\hat{M}^\pi) \le \left(6(1+\gamma) + DT(1+T)K^2(1+\gamma^{-1})\right)\left(\max_{t_j\in\pi}\mathrm{E}_0\left[|Y^\pi_{t_j} - \hat{Y}^\pi_{t_j}|^2\right] + \mathrm{E}_0\left[|M^\pi_{t_N} - \hat{M}^\pi_{t_N}|^2\right]\right).$$

Choosing $\gamma = T(1+T)K^2$ we obtain the lower bound (7) with $c = (6+D)(1+K^2T(1+T))$.

In order to derive the upper bound, we first introduce the process $\bar{Y}^\pi$ via

$$\bar{Y}^\pi_{t_0} = \hat{Y}^\pi_{t_0}, \quad \bar{Y}^\pi_{t_{i+1}} = \bar{Y}^\pi_{t_i} + f^\pi(t_i,\hat{Y}^\pi_{t_i},\hat{Z}^\pi_{t_i})\Delta_i + \hat{M}^\pi_{t_{i+1}} - \hat{M}^\pi_{t_i}, \quad i = 0,\dots,N-1,$$

where again $\hat{Z}^\pi_{t_i} = \Delta_i^{-1}E_i[\Delta W_i^*\hat{M}^\pi_{t_{i+1}}]$. Then the pair $(\bar{Y}^\pi,\hat{M}^\pi)$ solves the discrete BSDE with terminal condition $\bar{\xi}^\pi = \bar{Y}^\pi_{t_N}$ and driver $\bar{f}^\pi(t_i,y,z) = f^\pi(t_i,\hat{Y}^\pi_{t_i},z)$. We will first estimate the error between $(\bar{Y}^\pi,\hat{M}^\pi)$ and (Y^π,M^π) by a slight modification of the weighted a-priori estimates in Lemma 7 in Bender and Denk (2007). For some constants $\Gamma,\gamma > 0$, which will be fixed later, we consider the weights $q_i = \prod_{j=0}^{i-1}(1+\Gamma\Delta_j)$. Now recall from (2) that

$$M^\pi_{t_{i+1}} - M^\pi_{t_i} = Y^\pi_{t_{i+1}} - E_i[Y^\pi_{t_{i+1}}], \quad \hat{M}^\pi_{t_{i+1}} - \hat{M}^\pi_{t_i} = \bar{Y}^\pi_{t_{i+1}} - E_i[\bar{Y}^\pi_{t_{i+1}}].$$

Hence,

$$\sum_{i=0}^{N-1}q_i\mathrm{E}_0\left[|(M^\pi_{t_{i+1}} - M^\pi_{t_i}) - (\hat{M}^\pi_{t_{i+1}} - \hat{M}^\pi_{t_i})|^2\right] = \sum_{i=0}^{N-1}q_i\mathrm{E}_0\left[|Y^\pi_{t_{i+1}} - \bar{Y}^\pi_{t_{i+1}} - E_i[Y^\pi_{t_{i+1}} - \bar{Y}^\pi_{t_{i+1}}]|^2\right].$$

Now, following the argument in step 1 of the proof of Lemma 7 in Bender and Denk (2007), we obtain,

$$
\begin{aligned}
&\sum_{i=1}^{N-1} q_i \mathrm{E}_0\left[|(M^\pi_{t_{i+1}}-M^\pi_{t_i})-(\hat{M}^\pi_{t_{i+1}}-\hat{M}^\pi_{t_i})|^2\right] \\
\leq\ & q_N \mathrm{E}_0\left[\left|Y^\pi_{t_N}-\bar{Y}^\pi_{t_N}\right|^2\right]+\gamma\sum_{i=0}^{N-1} q_i\Delta_i \mathrm{E}_0\left[\left|Y^\pi_{t_i}-\bar{Y}^\pi_{t_i}\right|^2\right] \\
&+\frac{(1+T)K^2}{\gamma T}\sum_{i=0}^{N-1} q_i\Delta_i \mathrm{E}_0\left[\left|Y^\pi_{t_i}-\hat{Y}^\pi_{t_i}\right|^2\right]+\frac{(1+T)K^2}{\gamma}\sum_{i=0}^{N-1} q_i\Delta_i \mathrm{E}_0\left[\left|Z^\pi_{t_i}-\hat{Z}^\pi_{t_i}\right|^2\right].
\end{aligned}
$$

The argument of step 2 of the same proof yields

$$
\begin{aligned}
&\max_{0\leq i\leq N} q_i \mathrm{E}_0\left[\left|Y^\pi_{t_i}-\bar{Y}^\pi_{t_i}\right|^2\right] \leq q_N \mathrm{E}_0\left[\left|Y^\pi_{t_N}-\bar{Y}^\pi_{t_N}\right|^2\right] \\
+\ & K^2(T+1)(|\pi|+\Gamma^{-1})\left(\frac{1}{T}\sum_{i=0}^{N-1} q_i\Delta_i \mathrm{E}_0\left[\left|Y^\pi_{t_i}-\hat{Y}^\pi_{t_i}\right|^2\right]+\sum_{i=0}^{N-1} q_i\Delta_i \mathrm{E}_0\left[\left|Z^\pi_{t_i}-\hat{Z}^\pi_{t_i}\right|^2\right]\right).
\end{aligned}
$$

Thus, combining these two inequalities with a straightforward weighted version of (8), we get

$$
\begin{aligned}
&\max_{0\leq i\leq N} q_i \mathrm{E}_0\left[\left|Y^\pi_{t_i}-\bar{Y}^\pi_{t_i}\right|^2\right]+\sum_{i=1}^{N-1} q_i \mathrm{E}_0\left[|(M^\pi_{t_{i+1}}-M^\pi_{t_i})-(\hat{M}^\pi_{t_{i+1}}-\hat{M}^\pi_{t_i})|^2\right] \\
&\leq (2+\gamma T)q_N \mathrm{E}_0\left[\left|Y^\pi_{t_N}-\bar{Y}^\pi_{t_N}\right|^2\right]+\left[(1+\gamma T)K^2(T+1)(|\pi|+\Gamma^{-1})+\frac{(1+T)K^2}{\gamma}\right] \\
&\quad\times\left(\max_{0\leq i\leq N} q_i \mathrm{E}_0\left[\left|Y^\pi_{t_i}-\hat{Y}^\pi_{t_i}\right|^2\right]+D\sum_{i=1}^{N-1} q_i \mathrm{E}_0\left[|(M^\pi_{t_{i+1}}-M^\pi_{t_i})-(\hat{M}^\pi_{t_{i+1}}-\hat{M}^\pi_{t_i})|^2\right]\right) \\
&\leq (2+\gamma T)q_N \mathrm{E}_0\left[\left|Y^\pi_{t_N}-\bar{Y}^\pi_{t_N}\right|^2\right]+\left[(1+\gamma T)K^2(T+1)(|\pi|+\Gamma^{-1})+\frac{(1+T)K^2}{\gamma}\right](D\vee 2) \\
&\quad\times\left(\max_{0\leq i\leq N} q_i \mathrm{E}_0\left[\left|Y^\pi_{t_i}-\bar{Y}^\pi_{t_i}\right|^2\right]+\sum_{i=1}^{N-1} q_i \mathrm{E}_0\left[|(M^\pi_{t_{i+1}}-M^\pi_{t_i})-(\hat{M}^\pi_{t_{i+1}}-\hat{M}^\pi_{t_i})|^2\right]\right) \\
&\quad+\left[(1+\gamma T)K^2(T+1)(|\pi|+\Gamma^{-1})+\frac{(1+T)K^2}{\gamma}\right]\left(1+\frac{1}{(D\vee 2)-1}\right) \\
&\quad\times\max_{0\leq i\leq N} q_i \mathrm{E}_0\left[\left|\hat{Y}^\pi_{t_i}-\bar{Y}^\pi_{t_i}\right|^2\right].
\end{aligned}
$$

Choosing $\gamma=4(1+T)K^2(D\vee 2)$, $\Gamma=4K^2(T+1)(1+\gamma T)(D\vee 2)$, it, thus, holds, for $|\pi|\leq\Gamma^{-1}$

$$
\begin{aligned}
&\max_{0\leq i\leq N} q_i \mathrm{E}_0\left[\left|Y^\pi_{t_i}-\bar{Y}^\pi_{t_i}\right|^2\right]+\sum_{i=1}^{N-1} q_i \mathrm{E}_0\left[|(M^\pi_{t_{i+1}}-M^\pi_{t_i})-(\hat{M}^\pi_{t_{i+1}}-\hat{M}^\pi_{t_i})|^2\right] \\
&\leq 4(2+\gamma T)q_N \mathrm{E}_0\left[\left|Y^\pi_{t_N}-\bar{Y}^\pi_{t_N}\right|^2\right]+\frac{3}{D\vee 2}\left(1+\frac{1}{(D\vee 2)-1}\right)\max_{0\leq i\leq N} q_i \mathrm{E}_0\left[\left|\hat{Y}^\pi_{t_i}-\bar{Y}^\pi_{t_i}\right|^2\right].
\end{aligned}
$$

Now, applying Young's inequality twice and taking the definition of the weights into account, we have

$$
\begin{aligned}
&\max_{0\le i\le N} \mathrm{E}_0\left[\left|Y^\pi_{t_i}-\hat{Y}^\pi_{t_i}\right|^2\right]+\mathrm{E}_0\left[|(M^\pi_{t_N}-\hat{M}^\pi_{t_N})|^2\right]\\
\le\;& 8(2+\gamma T)e^{\Gamma T}\mathrm{E}_0\left[|Y^\pi_{t_N}-\bar{Y}^\pi_{t_N}|^2\right]+\left(\frac{6}{D\vee 2}\left(1+\frac{1}{(D\vee 2)-1}\right)e^{\Gamma T}+2\right)\max_{0\le i\le N}\mathrm{E}_0\left[\left|\hat{Y}^\pi_{t_i}-\bar{Y}^\pi_{t_i}\right|^2\right]\\
\le\;& 16(2+\gamma T)e^{\Gamma T}\mathrm{E}_0\left[\left|\xi^\pi-\hat{Y}^\pi_{t_N}\right|^2\right]\\
&+\left(\frac{6}{D\vee 2}\left(1+\frac{1}{(D\vee 2)-1}\right)+2+16(2+\gamma T)\right)e^{\Gamma T}\max_{0\le i\le N}\mathrm{E}_0\left[\left|\hat{Y}^\pi_{t_i}-\bar{Y}^\pi_{t_i}\right|^2\right]\\
\le\;& \left(\frac{6}{(D\vee 2)-1}+2+16(2+4(1+T)TK^2(D\vee 2))\right)e^{\Gamma T}\mathcal{E}_\pi(\hat{Y}^\pi,\hat{M}^\pi),
\end{aligned}
$$

because, by the construction of $\bar{Y}^\pi$,

$$
\hat{Y}^\pi_{t_i}-\bar{Y}^\pi_{t_i}=\hat{Y}^\pi_{t_i}-\hat{Y}^\pi_{t_0}-\sum_{j=0}^{i-1}f^\pi(t_j,\hat{Y}^\pi_{t_j},\Delta_j^{-1}E_j[\Delta W_j^*\hat{M}^\pi_{t_{j+1}}])\Delta_j-\hat{M}^\pi_{t_i}. \tag{9}
$$

□

3 A-posteriori estimates for continuous BSDEs

We now turn to BSDEs in continuous time

$$
Y_t=\xi-\int_t^T f(s,Y_s,Z_s)ds-Z_sdW_s \tag{10}
$$

and assume

(H) ξ is a square-integrable $\mathcal{F}_T$-measurable random variable. $f:[0,T]\times\Omega\times\mathbb{R}\times\mathbb{R}^D\to\mathbb{R}$ is measurable and, for every $(y,z)\in\mathbb{R}^{D+1}$, $f(\cdot,y,z)$ is $\mathcal{F}_t$-adapted. Moreover f is Lipschitz in (y,z) with constant K uniformly in (t,ω) and $\mathrm{E}[\int_0^T|f(t,0,0)|^2]dt<\infty$.

Under this set of assumptions a classical result by Pardoux and Peng (1990) ensures that there is a unique pair of square integrable $\mathcal{F}_t$-adapted processes (Y,Z) such that (10) is satisfied.

We now suppose that a square-integrable approximative solution $(\hat{Y}^\pi_{t_i},\hat{Z}^\pi_{t_i})_{t_i\in\pi}$ of (Y,Z) has been computed on some time grid π by some numerical algorithm. Again we allow that $(\hat{Y}^\pi_{t_i},\hat{Z}^\pi_{t_i})$ is adapted to a larger filtration $\mathcal{G}_{t_i}$ and wish to quantify the error between true solution and approximate solution. As, in general, it may not be possible to draw sample copies of ξ and $f(t_i,\hat{Y}^\pi_{t_i},\hat{Z}^\pi_{t_i})$, we apply the error criterion (5) with approximative data (ξ^π,f^π). It now reads with a slight abuse of notation, writing again $\hat{Z}^\pi$ instead of the the martingale difference of $\hat{Z}^\pi$ on the left-hand side,

$$
\mathcal{E}_\pi(\hat{Y}^\pi,\hat{Z}^\pi)=E_0[|\xi^\pi-\hat{Y}^\pi_{t_N}|^2]+\max_{1\le i\le N}\mathrm{E}_0\left[\left|\hat{Y}^\pi_{t_i}-\hat{Y}^\pi_{t_0}-\sum_{j=0}^{i-1}f^\pi(t_j,\hat{Y}^\pi_{t_j},\hat{Z}^\pi_{t_i})\Delta_j-\sum_{j=0}^{i-1}\hat{Z}^\pi_{t_j}\Delta W_j\right|^2\right], \tag{11}
$$

(where we recall that E_0 denotes the conditional expectation with respect to $\mathcal{G}_0$). We assume throughout this section that the larger filtration $\mathcal{G}_t$ is obtained from the augmented Brownian filtration by an enlargement at time 0, i.e. $\mathcal{G}_t = \mathcal{F}_t \vee \sigma(\Xi)$, where Ξ denotes a family of random variables independent of $\mathcal{F}_T$ (cf. the discussion in Example 2.2). Concerning the approximative data, we suppose ($\mathrm{H}_\pi^{\mathcal{F}}$), i.e. assumption ($\mathrm{H}_\pi^{\mathcal{G}}$) as introduced in the previous section, but with the larger filtration $\mathcal{G}_{t_i}$ replaced by the augmented Brownian filtration $(\mathcal{F}_{t_i})$ on the grid.

In this situation we obtain the following estimates on the squared L^2-error between true and approximative solution. The estimates can be computed in terms of the approximate solution on the grid, the approximate data and the error between true and approximate data.

Theorem 3.1. *Suppose (H) and ($H_\pi^{\mathcal{F}}$) and that $\mathcal{G}_t = \mathcal{F}_t \vee \sigma(\Xi)$, where Ξ is independent of $\mathcal{F}_T$. Then, there are constants $c, C > 0$ such that for every pair of $\mathcal{G}_{t_i}$-adapted square integrable processes $(\hat{Y}^\pi_{t_i}, \hat{Z}^\pi_{t_i})_{t_i \in \pi}$*

$$\begin{aligned} &\max_{t_i \in \pi} \mathrm{E}_0 \left[|Y_{t_i} - \hat{Y}^\pi_{t_i}|^2 \right] + \sum_{i=0}^{N-1} \mathrm{E}_0 \left[\int_{t_i}^{t_{i+1}} \left(|Y_t - \hat{Y}^\pi_{t_i}|^2 + |Z_t - \hat{Z}^\pi_{t_i}|^2 \right) dt \right] \\ \leq\;& C \left(\mathcal{E}_\pi(\hat{Y}^\pi, \hat{Z}^\pi) + |\pi| + \mathrm{E}\left[|\xi - \xi^\pi|^2 \right] + \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \mathrm{E}\left[|f(t, Y_t, Z_t) - f^\pi(t_i, Y_t, Z_t)|^2 \right] dt \right). \end{aligned}$$

and

$$\begin{aligned} \mathcal{E}_\pi(\hat{Y}^\pi, \hat{Z}^\pi) \;\leq\;& c \Bigg(\max_{t_i \in \pi} \mathrm{E}_0 \left[|Y_{t_i} - \hat{Y}^\pi_{t_i}|^2 \right] + \sum_{i=0}^{N-1} \mathrm{E}_0 \left[\int_{t_i}^{t_{i+1}} \left(|Y_t - \hat{Y}^\pi_{t_i}|^2 + |Z_t - \hat{Z}^\pi_{t_i}|^2 \right) dt \right] \\ & + \mathrm{E}\left[|\xi - \xi^\pi|^2 \right] + \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \mathrm{E}\left[|f(t, Y_t, Z_t) - f^\pi(t_i, Y_t, Z_t)|^2 \right] dt \Bigg) \end{aligned}$$

If, additionally, f and f^π do not depend on y, then

$$\begin{aligned} &\max_{t_i \in \pi} \mathrm{E}_0 \left[|Y_{t_i} - \hat{Y}^\pi_{t_i}|^2 \right] + \sum_{i=0}^{N-1} \mathrm{E}_0 \left[\int_{t_i}^{t_{i+1}} |Z_t - \hat{Z}^\pi_{t_i}|^2 dt \right] \\ \leq\;& C \left(\mathcal{E}_\pi(\hat{Y}^\pi, \hat{Z}^\pi) + \mathrm{E}\left[|\xi - \xi^\pi|^2 \right] + \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \mathrm{E}\left[|f(t, Z_t) - f^\pi(t_i, Z_t)|^2 \right] dt \right) \end{aligned}$$

and

$$\begin{aligned} \mathcal{E}_\pi(\hat{Y}^\pi, \hat{Z}^\pi) \;\leq\;& c \Bigg(\max_{t_i \in \pi} \mathrm{E}_0 \left[|Y_{t_i} - \hat{Y}^\pi_{t_i}|^2 \right] + \sum_{i=0}^{N-1} \mathrm{E}_0 \left[\int_{t_i}^{t_{i+1}} |Z_t - \hat{Z}^\pi_{t_i}|^2 \; dt \right] \\ & + \mathrm{E}\left[|\xi - \xi^\pi|^2 \right] + \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \mathrm{E}\left[|f(t, Z_t) - f^\pi(t_i, Z_t)|^2 \right] dt \Bigg) \end{aligned}$$

Remark 3.2. If one inspects the proof below carefully, then the constants C and c can be made explicit. They only depend on the time horizon T, the dimension D, the Lipschitz constant K of f and on $\mathrm{E}\left[\int_0^T |f(t,0,0)|^2 dt \right]$ and $\mathrm{E}\left[|\xi|^2 \right]$.

In view of Theorem 2.1, the above results can be easily deduced from error estimates between (Y, Z) and the implicit time discretization (Y^π, Z^π) corresponding to the approximate data (ξ^π, f^π). It is well-known from the results by Zhang (2004) that such time discretization crucially depends on some L^2-regularity of Z. For this reason some extra conditions on the data are usually imposed. For example, ξ is often assumed to be a Lipschitz functional of the path of a forward SDE X and $f(t, y, z)$ is supposed to depend on ω only through X_t. In our setting, the crucial observation is that the term which corresponds to this L^2-regularity can be estimated by the error criterion (11) without any such assumptions on the data.

Lemma 3.3. *Suppose (H) and* $(\mathrm{H}_\pi^{\mathcal{F}})$ *and that* $\mathcal{G}_t = \mathcal{F}_t \vee \sigma(\Xi)$, *where* Ξ *is independent of* $\mathcal{F}_T$. *Then, there is a constant* $C > 0$ *such that for every pair of* $\mathcal{G}_{t_i}$*-adapted square integrable processes* $(\hat{Y}^\pi_{t_i}, \hat{Z}^\pi_{t_i})_{t_i \in \pi}$

$$\max_{0\le i\le N} \mathrm{E}\left[\left|Y_{t_i} - Y^\pi_{t_i}\right|^2\right] + \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \mathrm{E}\left[\left|Z_t - Z^\pi_{t_i}\right|^2\right] dt$$

$$\le \; C\left(\mathcal{E}_\pi(\hat{Y}^\pi, \hat{Z}^\pi) + K_y\,|\pi| + \mathrm{E}\left[|\xi - \xi^\pi|^2\right] + \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \mathrm{E}\left[|f(t, Y_t, Z_t) - f^\pi(t_i, Y_t, Z_t)|^2\right] dt\right).$$

Here, $K_y \le K$ *denotes a Lipschitz constant of* f^π *with respect to the* y*-variable.*

Proof. First notice that by assumption $(\mathrm{H}_\pi^{\mathcal{F}})$ and the assumption on the filtration $\mathcal{G}_t$, it holds

$$\begin{aligned} Y^\pi_{t_N} &= \xi^\pi, \quad Z^\pi_{t_N} = 0, \\ Z^\pi_{t_i} &= E[\Delta W_i^* Y^\pi_{t_{i+1}} | \mathcal{F}_{t_i}], \quad i = N-1, \ldots, 0, \\ Y^\pi_{t_i} &= E_i[Y^\pi_{t_{i+1}} | \mathcal{F}_{t_i}] - f^\pi(t_i, Y^\pi_{t_i}, Z^\pi_{t_i})\Delta_i, \quad i = N-1, \ldots, 0, \end{aligned} \tag{12}$$

i.e. the filtration $\mathcal{G}_t$ in (4) can be replaced by the augmented Brownian filtration $\mathcal{F}_t$. The proof consists of two steps. In the first step we observe that

$$\begin{aligned} &\max_{0\le i\le N} \mathrm{E}\left[\left|Y_{t_i} - Y^\pi_{t_i}\right|^2\right] + \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \mathrm{E}\left[\left|Z_t - Z^\pi_{t_i}\right|^2\right] dt \\ &\quad \le C\Bigg(\mathrm{E}\left[|\xi - \xi^\pi|^2\right] + K_y\,|\pi| + \sum_{i=0}^{N-1} \mathrm{E}\left[\left(\int_{t_i}^{t_{i+1}} \check{Z}^\pi_s dW_s - Z^\pi_{t_i}\Delta W_i\right)^2\right] \\ &\qquad + \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \mathrm{E}\left[|f(t, Y_t, Z_t) - f^\pi(t_i, Y_t, Z_t)|^2\right] dt\Bigg), \end{aligned} \tag{13}$$

where $\check{Z}^\pi$ is defined on $[t_i, t_{i+1})$, $i = 0, \ldots, N-1$, via the martingale representation theorem as

$$\int_{t_i}^{t_{i+1}} \check{Z}^\pi_t dW_t = Y^\pi_{t_{i+1}} - \mathrm{E}\left[Y^\pi_{t_{i+1}} | \mathcal{F}_{t_i}\right]. \tag{14}$$

Inequality (13) can be derived by slightly modifying the argument in Theorem 3.1 of Bouchard and Touzi (2004). For the reader's convenience we provide the details in the Appendix.

As a second step we will now show that

$$\sum_{i=0}^{N-1} \mathrm{E}\left[\left(\int_{t_i}^{t_{i+1}} \check{Z}_s^\pi dW_s - Z_{t_i}^\pi \Delta W_i\right)^2\right] \leq C\mathcal{E}_\pi(\hat{Y}^\pi, \hat{Z}^\pi), \tag{15}$$

which, in view of (13), completes the proof of this lemma.

By (12) and the independence of $\mathcal{G}_0$ and $\mathcal{F}_T$, Young's inequality and Itô's isometry, we obtain

$$\begin{aligned}
&\sum_{i=0}^{N-1} \mathrm{E}\left[\left(\int_{t_i}^{t_{i+1}} \check{Z}_s^\pi dW_s - Z_{t_i}^\pi \Delta W_i\right)^2\right] = \sum_{i=0}^{N-1} \mathrm{E}_0\left[\left(\int_{t_i}^{t_{i+1}} \check{Z}_s^\pi dW_s - Z_{t_i}^\pi \Delta W_i\right)^2\right] \\
\leq\ & 2\sum_{i=0}^{N-1} \mathrm{E}_0\left[\left(\int_{t_i}^{t_{i+1}} \check{Z}_s^\pi dW_s - \hat{Z}_{t_i}^\pi \Delta W_i\right)^2\right] + 2\sum_{i=0}^{N-1} \mathrm{E}_0\left[|(\hat{Z}_{t_i}^\pi - Z_{t_i}^\pi)\Delta W_i|^2\right] \\
=\ & 2\mathrm{E}_0\left[\left(\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}} \check{Z}_s^\pi dW_s - \hat{Z}_{t_i}^\pi \Delta W_i\right)^2\right] + 2\sum_{i=0}^{N-1} \mathrm{E}_0\left[|\hat{Z}_{t_i}^\pi - Z_{t_i}^\pi|^2\right]\Delta_i
\end{aligned}$$

As in the proof of Theorem 2.1, we introduce the process $\bar{Y}_{t_i}^\pi$ by

$$\bar{Y}_{t_0}^\pi = \hat{Y}_{t_0}^\pi, \quad \bar{Y}_{t_{i+1}}^\pi = \bar{Y}_{t_i}^\pi + f^\pi(t_i, \hat{Y}_{t_i}^\pi, \hat{Z}_{t_i}^\pi)\Delta_i + \hat{Z}_{t_i}^\pi \Delta W_i, \quad i = 0, \dots, N-1,$$

with the specific choice $\hat{M}_{t_{i+1}}^\pi - \hat{M}_{t_i}^\pi = \hat{Z}_{t_i}^\pi \Delta W_i$ for the martingale increment. Notice that with this choice of the martingale $\hat{M}^\pi$, $\mathcal{E}_\pi(\hat{Y}^\pi, \hat{Z}^\pi)$ equals $\mathcal{E}_\pi(\hat{Y}^\pi, \hat{M}^\pi)$ in the notation of Theorem 2.1. Then, by (14) and (4), and applying Young's inequality and the Lipschitz property of f^π, we get

$$\begin{aligned}
&\mathrm{E}_0\left[\left(\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}} \check{Z}_s^\pi dW_s - \hat{Z}_{t_i}^\pi \Delta W_i\right)^2\right] \\
=\ & \mathrm{E}_0\left[\left|\sum_{i=0}^{N-1} Y_{t_{i+1}}^\pi - \bar{Y}_{t_{i+1}}^\pi - (Y_{t_i}^\pi - \bar{Y}_{t_i}^\pi) - \left(f^\pi(t_i, Y_{t_i}^\pi, Z_{t_i}^\pi) - f^\pi(t_i, \hat{Y}_{t_i}^\pi, \hat{Z}_{t_i}^\pi)\right)\Delta_i\right|^2\right] \\
\leq\ & C\left(\mathrm{E}_0\left[|\hat{Y}_{t_N}^\pi - \bar{Y}_{t_N}^\pi|^2\right] + \max_{t_i\in\pi}\mathrm{E}_0\left[|Y_{t_i}^\pi - \hat{Y}_{t_i}^\pi|^2\right] + \sum_{i=0}^{N-1}\mathrm{E}_0\left[|Z_{t_i}^\pi - \hat{Z}_{t_i}^\pi|^2\right]\Delta_i\right).
\end{aligned}$$

Gathering the above inequalities and applying (8) we have

$$\begin{aligned}
&\sum_{i=0}^{N-1} \mathrm{E}\left[\left(\int_{t_i}^{t_{i+1}} \check{Z}_s^\pi dW_s - Z_{t_i}^\pi \Delta W_i\right)^2\right] \\
\leq\ & C\left(\mathrm{E}_0\left[|\hat{Y}_{t_N}^\pi - \bar{Y}_{t_N}^\pi|^2\right] + \max_{t_i\in\pi}\mathrm{E}_0\left[|Y_{t_i}^\pi - \hat{Y}_{t_i}^\pi|^2\right] + \mathrm{E}_0\left[|M_{t_N}^\pi - \hat{M}_{t_N}^\pi|^2\right]\right).
\end{aligned}$$

Inequality (15) now follows from Theorem 2.1 and (9) with the above choice of the martingale $\hat{M}^\pi$. □

Proof of Theorem 3.1. We first prove the first and third inequality. These two inequalities basically follow by combining Theorem 2.1 and Lemma 3.3. Denote the martingale difference of $\hat{Z}^\pi$ with respect to the Brownian increments by $\hat{M}^\pi$, i.e.

$$\hat{M}_0^\pi = 0, \quad \hat{M}_{t_{i+1}}^\pi - \hat{M}_{t_i}^\pi = \hat{Z}_{t_i}^\pi \Delta W_i,$$

Then, $\mathcal{E}_\pi(\hat{Y}^\pi, \hat{Z}^\pi)$ equals $\mathcal{E}_\pi(\hat{Y}^\pi, \hat{M}^\pi)$ in the notation of the previous section. By the independence of $\mathcal{G}_0$ and $\mathcal{F}_T$ and in view of (8), we get

$$\begin{aligned}
&\max_{t_i \in \pi} \mathrm{E}_0 \left[|Y_{t_i} - \hat{Y}_{t_i}^\pi|^2 \right] + \sum_{i=0}^{N-1} \mathrm{E}_0 \left[\int_{t_i}^{t_{i+1}} |Z_t - \hat{Z}_{t_i}^\pi|^2 dt \right] \\
\leq\ & 2 \left(\max_{t_i \in \pi} \mathrm{E}\left[|Y_{t_i} - Y_{t_i}^\pi|^2 \right] + \sum_{i=0}^{N-1} \mathrm{E} \left[\int_{t_i}^{t_{i+1}} |Z_t - Z_{t_i}^\pi|^2 dt \right] \right) \\
& +2 \left(\max_{t_i \in \pi} \mathrm{E}_0 \left[|Y_{t_i}^\pi - \hat{Y}_{t_i}^\pi|^2 \right] + D\mathrm{E}_0 \left[|M_{t_N}^\pi - \hat{M}_{t_N}^\pi|^2 \right] \right)
\end{aligned}$$

Applying Theorem 2.1 and Lemma 3.3 yields, for some constant $C > 0$,

$$\begin{aligned}
&\max_{t_i \in \pi} \mathrm{E}_0 \left[|Y_{t_i} - \hat{Y}_{t_i}^\pi|^2 \right] + \sum_{i=0}^{N-1} \mathrm{E}_0 \left[\int_{t_i}^{t_{i+1}} |Z_t - \hat{Z}_{t_i}^\pi|^2 dt \right] \\
\leq\ & C \left(\mathcal{E}_\pi(\hat{Y}^\pi, \hat{Z}^\pi) + K_y |\pi| + \mathrm{E}\left[|\xi - \xi^\pi|^2 \right] + \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \mathrm{E}\left[|f(t, Y_t, Z_t) - f^\pi(t_i, Y_t, Z_t)|^2 \right] dt \right).
\end{aligned} \tag{16}$$

If f and f^π are independent of y, then we can choose $K_y = 0$ and the third inequality follows. In order to complete the proof of the first inequality we estimate

$$\sum_{i=0}^{N-1} \mathrm{E}_0 \left[\int_{t_i}^{t_{i+1}} |Y_t - \hat{Y}_{t_i}^\pi|^2 dt \right] \leq 2T \max_{t_i \in \pi} \mathrm{E}_0 \left[|Y_{t_i} - \hat{Y}_{t_i}^\pi|^2 \right] + 2 \sum_{i=0}^{N-1} \mathrm{E} \left[\int_{t_i}^{t_{i+1}} |Y_t - Y_{t_i}|^2 dt \right]$$

As

$$\sum_{i=0}^{N-1} \mathrm{E} \left[\int_{t_i}^{t_{i+1}} |Y_t - Y_{t_i}|^2 dt \right] \leq C |\pi|$$

e.g. by Lemma 2.4 in Zhang (2004), we immediately obtain the first inequality thanks to (16).

For the second and fourth inequality we we make use of the identity

$$Y_{t_i} - Y_0 = \int_0^{t_i} f(t, Y_t, Z_t) dt + \int_0^{t_i} Z_t dW_t. \tag{17}$$

After inserting (17) we obtain by the Itô isometry, Young's inequality and Jensen's inequality

$$\begin{aligned}
\mathcal{E}_\pi(\hat{Y}^\pi, \hat{Z}^\pi) &= \mathrm{E}_0\left[|\hat{Y}^\pi_{t_N} - \xi^\pi|^2\right] + \max_{1\le i\le N} \mathrm{E}_0\left[\left|\left(\hat{Y}^\pi_{t_i} - Y_{t_i}\right) + \left(Y_0 - \hat{Y}^\pi_{t_0}\right)\right.\right.\\
&\quad \left.\left. + \sum_{j=0}^{i-1}\int_{t_j}^{t_{j+1}} \left(f(t,Y_t,Z_t) - f^\pi(t_j, \hat{Y}^\pi_{t_j}, \hat{Z}^\pi_{t_j})\right) dt + \sum_{j=0}^{i-1}\int_{t_j}^{t_{j+1}} \left(Z_t - \hat{Z}^\pi_{t_j}\right) dW_t\right|^2\right]\\
&\le c\left(\max_{0\le i\le N} \mathrm{E}_0\left[\left|\hat{Y}^\pi_{t_i} - Y_{t_i}\right|^2\right] + \sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}} \mathrm{E}_0\left[\left|Z_t - \hat{Z}^\pi_{t_i}\right|^2\right] dt\right.\\
&\quad + \mathrm{E}\left[|\xi - \xi^\pi|^2\right] + \sum_{j=0}^{N-1}\int_{t_j}^{t_{j+1}} \mathrm{E}\left[|f(t,Y_t,Z_t) - f^\pi(t_j, Y_t, Z_t)|^2\right] dt\\
&\quad \left. + \sum_{j=0}^{N-1}\int_{t_j}^{t_{j+1}} \mathrm{E}_0\left[\left|f^\pi(t_j,Y_t,Z_t) - f^\pi(t_j, \hat{Y}^\pi_{t_j}, \hat{Z}^\pi_{t_j})\right|^2\right] dt\right). \qquad (18)
\end{aligned}$$

The Lipschitz property of f^π and Young's inequality yield

$$\begin{aligned}
&\sum_{j=0}^{N-1}\int_{t_j}^{t_{j+1}} \mathrm{E}_0\left[\left|f^\pi(t_j,Y_t,Z_t) - f^\pi(t_j, \hat{Y}^\pi_{t_j}, \hat{Z}^\pi_{t_j})\right|^2\right] dt\\
\le\ & 2K_y^2 \sum_{i=0}^{N-1} \mathrm{E}_0\left[\int_{t_i}^{t_{i+1}} |Y_t - \hat{Y}^\pi_{t_i}|^2 dt\right] + 2K^2 \sum_{i=0}^{N-1} \mathrm{E}_0\left[\int_{t_i}^{t_{i+1}} |Z_t - \hat{Z}^\pi_{t_i}|^2 dt\right]
\end{aligned}$$

Combining this estimate with (18) we obtain the second inequality. If f^π is independent of y, we can again choose $K_y = 0$, and get the fourth inequality. □

We close this section with some examples which cover typical situations in the literature on numerical algorithms for BSDEs.

Example 3.4. *Suppose that the data (ξ, f) satisfies*

$$\xi = \varphi(X_T), \quad f(t,y,z) = F(t, X_t, y, z)$$

where X_t is the solution of an SDE

$$dX_t = b(t,X_t)dt + \sigma(t,X_t)dW_t, \quad X_t = x_0$$

with constant initial condition x_0 and Lipschitz continuous coefficients $b : [0,T] \times \mathbb{R}^{\mathcal{D}} \to \mathbb{R}^{\mathcal{D}}$ and $\sigma : [0,T] \times \mathbb{R}^{\mathcal{D}} \to \mathbb{R}^{\mathcal{D}\times D}$. We also assume that φ is Lipschitz continuous with constant K_φ and F is α-Hölder continuous in time and Lipschitz continuous in space, i.e.

$$|F(t_1,x_1,y_1,z_2) - F(t_2,x_2,y_2,z_2)| \le K_t|t_1 - t_2|^\alpha + K_x|x_1 - x_2| + K|y_1 - y_2| + K|z_1 - z_2|,$$

for some $\alpha \ge 1/2$. Given a strong order α approximation X^π of X on a grid π, i.e.

$$\max_{t_i\in\pi} \mathrm{E}\left[|X_{t_i} - X^\pi_{t_i}|^2\right] \le C|\pi|^{2\alpha},$$

we define the approximate data (ξ^π, f^π) *by*

$$\xi^\pi = \varphi(X_T^\pi), \quad f^\pi(t_i, y, z) = F(t_i, X_{t_i}^\pi, y, z).$$

Then,

$$\mathrm{E}\left[|\xi - \xi^\pi|^2\right] + \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \mathrm{E}\left[|f(t, Y_t, Z_t) - f^\pi(t_i, Y_t, Z_t)|^2\right] dt \leq C(K_\varphi |\pi|^{2\alpha} + K_t |\pi|^{2\alpha} + K_x |\pi|)$$

Recall also from Zhang (2004) that under the above conditions

$$\max_{i=0,\dots N} \sup_{t \in [t_i, t_{i+1}]} \mathrm{E}\left[|Y_{t_i} - Y_t|^2\right] \leq C |\pi| .$$

Hence, the first two inequalities in Theorem 3.1 can be rewritten as

$$\max_{i=0,\dots,N-1} \sup_{t \in [t_i, t_{i+1}]} \mathrm{E}_0\left[|Y_t - \hat{Y}_{t_i}^\pi|^2\right] + \sum_{i=0}^{N-1} \mathrm{E}_0\left[\int_{t_i}^{t_{i+1}} |Z_t - \hat{Z}_{t_i}^\pi|^2 dt\right] \leq C\left(\mathcal{E}_\pi(\hat{Y}^\pi, \hat{Z}^\pi) + |\pi|\right) \quad (19)$$

and

$$\begin{aligned} \mathcal{E}_\pi(\hat{Y}^\pi, \hat{Z}^\pi) \quad \leq \quad & c\Bigg(\max_{i=0,\dots,N-1} \sup_{t \in [t_i, t_{i+1}]} \mathrm{E}_0\left[|Y_t - \hat{Y}_{t_i}^\pi|^2\right] + \sum_{i=0}^{N-1} \mathrm{E}_0\left[\int_{t_i}^{t_{i+1}} |Z_t - \hat{Z}_{t_i}^\pi|^2 dt\right] \\ & + (K_\varphi + K_t) |\pi|^{2\alpha} + K_x |\pi|\Bigg). \end{aligned} \quad (20)$$

Thus, the square root of the error criterion is, up to a term of order $1/2$ *in the time step (which corresponds to the time discretization error under the above assumptions), equivalent to the* L^2*-error between true solution and approximate solution over the whole time interval (not only on the grid). If* F *does not depend on* x*, the additional error in the time step can be reduced to order* α *in the lower bound (20), if* F *is sufficiently regular in time and a higher order approximation* X^π *of* X *is applied. It vanishes completely, when additionally* F *does not depend on* t *and one can sample perfectly from* X *on the grid, i.e. one can choose* $X_{t_i}^\pi = X_{t_i}$*.*

Example 3.5. *Let us now turn to the case of an irregular terminal condition. We impose the same assumptions as in the previous example, but remove the Lipschitz condition on the terminal condition* φ*. For simplicity we assume that we can sample perfectly from* X *on the grid and, hence, choose the true data as approximate data*

$$\xi^\pi = \varphi(X_T) = \xi, \quad f^\pi(t_i, y, z) = F(t_i, X_{t_i}, y, z) = f(t_i, y, z).$$

In this situation, the first inequality in Theorem 3.1 becomes

$$\max_{t_i \in \pi} \mathrm{E}_0\left[|Y_{t_i} - \hat{Y}_{t_i}^\pi|^2\right] + \sum_{i=0}^{N-1} \mathrm{E}_0\left[\int_{t_i}^{t_{i+1}} \left(|Y_t - \hat{Y}_{t_i}^\pi|^2 + |Z_t - \hat{Z}_{t_i}^\pi|^2\right) dt\right] \leq C\left(\mathcal{E}_\pi(\hat{Y}^\pi, \hat{Z}^\pi) + |\pi|\right). \quad (21)$$

It is known that for irregular terminal conditions and equidistant time grids, the time discretization error for BSDEs can be of a smaller order than 1/2 in the mesh of the time grid (see e.g. the survey paper by Geiss et al., 2011), but order 1/2 convergence in the number of time steps may be retained by a suitable choice of a non-equidistant time grid under appropriate assumptions as shown in Gobet and Makhlouf (2010). Hence, under irregular terminal conditions, the error criterion $\mathcal{E}_\pi(\hat{Y}^\pi, \hat{Z}^\pi)$ *does contain significant information about the time discretization error.*

Example 3.6. *We now consider some very specific assumptions on the data under which all extra terms involving the mesh size* $|\pi|$ *of* π *vanish, namely*

$$\xi = \varphi(X_T), \quad f(t,x,y,z) = F(z)$$

for some Lipschitz continuous function F *and a function* φ *(which is not necessarily Lipschitz). This type of driver* f *may occur e.g. in the context of* g*-expectations as introduced in Peng (1997). We impose the same assumptions on* X *as in the previous example and additionally suppose that we can sample perfectly from* X *on the grid. We can then choose* $(\xi^\pi, f^\pi) = (\xi, f)$*. As* f *and* f^π *do not depend on* y *we can rewrite the third and fourth inequality in Theorem 3.1 as*

$$\frac{1}{c}\mathcal{E}_\pi(\hat{Y}^\pi, \hat{Z}^\pi) \le \max_{t_i \in \pi} \mathrm{E}_0\left[|Y_{t_i} - \hat{Y}^\pi_{t_i}|^2\right] + \sum_{i=0}^{N-1} \mathrm{E}_0\left[\int_{t_i}^{t_{i+1}} |Z_t - \hat{Z}^\pi_{t_i}|^2 dt\right] \le C\mathcal{E}_\pi(\hat{Y}^\pi, \hat{Z}^\pi) \tag{22}$$

Here the squared L^2*-error between true solution and approximate solution is equivalent to the error criterion even in continuous time. This is somehow surprising, because the error criterion can be computed solely in terms of the approximate solution on the grid. The* L^2*-error of the* Y*-part is, however, only considered on the grid* π *in (22), whereas it was estimated on the whole interval in the previous examples.*

Example 3.7. *Our last example treats the case of smooth data. Precisely, in the setting of Example 3.4 we suppose that all coefficient functions* φ, F, b, σ *are sufficently smooth and bounded with bounded derivatives. We also assume for the moment that we can sample perfectly from* X *on the grid* π*, which is here taken as equidistant. We can hence choose*

$$\xi^\pi = \varphi(X_T) = \xi, \quad f^\pi(t_i, y, z) = F(t_i, X_{t_i}, y, z) = f(t_i, y, z).$$

By the results of Gobet and Labart (2007) it follows that

$$\max_{t_i \in \pi} \mathrm{E}\left[|Y_{t_i} - Y^\pi_{t_i}|^2\right] + \sum_{i=0}^{N-1} \mathrm{E}\left[|Z_{t_i} - Z^\pi_{t_i}|^2\right] \Delta_i \le C|\pi|^2.$$

A straightforward combination with Theorem 2.1, taking (8) into account, leads to the upper bound

$$\max_{t_i \in \pi} \mathrm{E}_0\left[|Y_{t_i} - \hat{Y}^\pi_{t_i}|^2\right] + \sum_{i=0}^{N-1} \mathrm{E}_0\left[|Z_{t_i} - \hat{Z}^\pi_{t_i}|^2\right] \Delta_i \le C\left(\mathcal{E}_\pi(\hat{Y}^\pi, \hat{Z}^\pi) + |\pi|^2\right). \tag{23}$$

Here, we assume, of course, that the assumptions of Theorem 3.1 are in force and that $(\hat{Y}^\pi, \hat{Z}^\pi)$ *is* $\mathcal{G}_{t_i}$*-adapted. Compared to (19), the extra term is now of order 1 in the mesh size instead of* *order 1/2. The price to pay for this, is that the L^2-error between true solution and approximate solution is estimated on the grid only. It is straightforward to check that (23) still holds true when the approximate data is of the form*

$$\xi^\pi = \varphi(X_T^\pi), \quad f^\pi(t_i, y, z) = F(t_i, X_{t_i}^\pi, y, z),$$

where X^π is a strong order 1 approximation of X such as a Milstein scheme. Indeed, one must just compare the error criteria based on the true data and the approximate data.

4 Numerical examples

In this section we apply the a-posteriori estimates to a numerical example. The test BSDE in this example is a slight modification of the one suggested by Bender and Zhang (2008) in the context of coupled FBSDEs. We here consider the following Markovian BSDE

$$\begin{aligned} X_{d,t} &= x_{d,0} + \int_0^t \sigma\big(\sum_{d'=1}^D \sin(X_{d',u})\big) dW_{d,u}, \quad d = 1, \ldots, D, \\ Y_t &= \sum_{d=1}^D \sin(X_{d,T}) + \int_t^T \frac{1}{2}\sigma^2[(Y_u)^3]_{D^3} du - \sum_{d=1}^D \int_t^T Z_{d,u} dW_{d,u}, \end{aligned} \tag{24}$$

where

$$[x]_R = -R \wedge x \vee R$$

is the truncation function at level $\pm R$, $W = (W_1, \ldots, W_D)$ is a D-dimensional Brownian motion and $\sigma > 0$ and $x_{d,0}$, $d = 1, \ldots, D$, are constants. The truncation function was merely implemented in order to ensure the Lipschitz continuity of the driver.

In this example, the true solution for (Y, Z) is related to X by

$$Y_t = y(t, X(t)) = \sum_{d=1}^D \sin(X_{d,t}), \quad Z_{d,t} = z_d(t, X(t)) = \sigma \cos(X_{d,t})\big(\sum_{d'=1}^D \sin(X_{d',t})\big), \tag{25}$$

which can be verified by Itô's formula. This relation will be used later to compare the error criterion proposed in this paper and the squared L^2-error between true solution and approximation quantitatively.

For the evaluation of the error criterion we apply as approximate data

$$f^\pi(t, y, z) = f(t, y, z) = \frac{1}{2}\sigma^2[y^3]_{D^3}, \quad \xi^\pi = \sum_{d=1}^D \sin(X_{d,T}^\pi)$$

where $X^\pi = X^{\pi,E}$ or $X^\pi = X^{\pi,MS}$ is either the Euler scheme or the Milstein scheme of X with respect to the partition π. We denote the equidistant partition of $[0, T]$ into N subintervals by π_N.

For the numerical approximation of (y, z) we will use least-squares Monte Carlo as suggested by Lemor et al. (2006) based on the Euler scheme. To this end we first sample a family

$$\Xi = ((_\lambda X^{\pi_N}_{t_{i+1}}, \Delta_\lambda W_i)|\lambda = 1, \dots, \Lambda,\ i = 0, \dots, N-1) \tag{26}$$

of Λ independent copies of $(X^{\pi_N,E}_{t_{i+1}}, \Delta W_i)_{i=0,\dots,N-1}$. Furthermore, we choose function bases

$$\eta(i, x) = \{\eta_1(i, x), \dots, \eta_K(i, x)\}, \quad i = 0, \dots, N-1,$$

where K is the number of basis functions at each time step. In principle, different basis functions can be used for the Y-part and the Z-part of the solution. But below we apply, for simplicity, the same basis functions for both parts of the solution. Initializing the algorithm at terminal time $t_N = T$ by

$$\hat{y}^{\pi_N}(t_N, x) = \phi(X^{\pi_N}_{t_N})$$

one defines iteratively, for $i = N-1$ to 0,

$$\begin{aligned}
\hat{\alpha}_{d,i} &= \arg\min_{\alpha\in\mathbb{R}^K} \sum_{\lambda=1}^{\Lambda} |\frac{\Delta_\lambda W_{d,i}}{\Delta_i}\hat{y}^{\pi_N}_{i+1} - \eta(i,_\lambda X^{\pi_N}_{t_i})\alpha|^2, \quad d = 1, \dots, D,\\
\hat{z}^{\pi_N}_d(t_i, x) &= \eta(i, x)\hat{\alpha}_{d,i}, \quad d = 1, \dots, D,\\
\hat{\alpha}_{0,i} &= \arg\min_{\alpha\in\mathbb{R}^K} \sum_{\lambda=1}^{\Lambda} |\hat{y}^{\pi_N}_{i+1}(_\lambda X^{\pi_N}_{t_{i+1}}) - \Delta_i f^\pi\big(t_i, \hat{y}^{\pi_N}_{i+1}(_\lambda X^{\pi_N}_{t_{i+1}}), \hat{z}^{\pi_N}_i(_\lambda X^{\pi_N}_{t_i})\big) - \eta(i,_\lambda X^{\pi_N}_{t_i})\alpha|^2,\\
\hat{y}^{\pi_N}(t_i, x) &= \eta(i, x)\hat{\alpha}_{0,i}.
\end{aligned}$$

We finally obtain

$$\hat{Y}^{\pi_N}_{t_i} = \hat{y}^{\pi_N}(t_i, X^{\pi_N,E}_{t_i}), \quad \hat{Z}^{\pi_N}_{t_i} = \hat{z}^{\pi_N}(t_i, X^{\pi_N,E}_{t_i})$$

as approximations for (Y, Z) on the grid π_N. They are extended by piecewise constant interpolation on the whole time interval $[0, T]$.

In the multidimensional case we also apply a slight modification of this procedure, which has a flavour of the classical control variate technique in the present non-linear BSDE setting and can easily be applied to various other BSDEs in a similar way. Instead of approximating Y_t directly, we approximate $Y_t - u(t, X_t)$ for a function u, for which we hope that $u(t, X_t)$ explains a significant part of Y. In the present example we construct u as follows: We freeze the diffusion coefficient of X and consider the simple BSDE with the same terminal function as in the original BSDE, i.e.

$$\begin{aligned}
\breve{X}_{d,t} &= x_{d,0} + \sigma\big(\sum_{d'=1}^{D} \sin(x_{d',0})\big)W_{d,t}, \quad d = 1, \dots, D,\\
\breve{Y}_t &= \sum_{d=1}^{D} \sin(\breve{X}_{d,T}) - \sum_{d=1}^{D} \int_t^T \breve{Z}_{d,u} dW_{d,u}.
\end{aligned}$$

A direct calculation shows that $\breve{Y}_t = u(t, \breve{X}_t)$ for

$$u(t, x_1, \dots, x_D) = \exp\left\{-\frac{1}{2}\sigma^2\left(\sum_{d'=1}^{D}\sin(x_{d',0})\right)^2 (T-t)\right\}\sum_{d=1}^{D}\sin(x_d)$$

Then, applying Itô's formula to $u(t, X_t)$, we observe that

$$\begin{aligned} Y_t &= Y_t^{\mathcal{V}} + u(t, X_t) \\ Z_{d,t} &= Z_{d,t}^{\mathcal{V}} + \sigma\left(\sum_{d'=1}^{D}\sin(X_{d',t})\right)\exp\left\{-\frac{1}{2}\sigma^2\left(\sum_{d'=1}^{D}\sin(x_{d',0})\right)^2 (T-t)\right\}\cos(X_{d,t}), \end{aligned}$$

where $(Y_t^{\mathcal{V}}, Z_t^{\mathcal{V}})$ solves the following BSDE with zero terminal condition:

$$\begin{aligned} Y_t^{\mathcal{V}} &= \int_t^T \frac{1}{2}\sigma^2\left\{[(u(s,X_s)+Y_s^{\mathcal{V}})^3]_{D^3} + u(s,X_s)\Big(\big(\sum_{d=1}^{D}\sin(x_{d,0})\big)^2 - \big(\sum_{d=1}^{D}\sin(X_{d,s})\big)^2\Big)\right\}ds \\ &\quad -\sum_{d=1}^{D}\int_t^T Z_{d,s}^{\mathcal{V}} dW_{d,s}. \end{aligned} \tag{27}$$

When we say below, that the non-linear control variate technique is applied, this means that the pair $(Y_t^{\mathcal{V}}, Z_t^{\mathcal{V}})$ is approximated by least-squares Monte-Carlo instead of (Y, Z). Then the corresponding approximations to (Y, Z) are defined via

$$\begin{aligned} &\hat{Y}_{t_i}^{\mathcal{V},\pi_N} + u(t_i, X_{t_i}^{\pi_N,E}) \\ &\hat{Z}_{d,t_i}^{\mathcal{V},\pi_N} + \sigma\left(\sum_{d'=1}^{D}\sin(X_{d',t_i}^{\pi_N,E})\right)\exp\left\{-\frac{1}{2}\sigma^2\left(\sum_{d'=1}^{D}\sin(x_{d',0})\right)^2 (T-t_i)\right\}\cos(X_{d,t_i}^{\pi_N,E}). \end{aligned}$$

4.1 Case 1: One-dimensional Brownian motion and local basis functions

In the first case we fix the parameters as follows:

$$D = 1, \quad T = 1, \quad x_{0,1} = \pi/2, \quad \sigma = 0.4\ .$$

Let $K \geq 3$ be the number of basis functions. We consider a local basis of indicator functions which partition the interval $[0, 3]$ into equidistant subintervals. Precisely, we set

$$\begin{aligned} &\eta_1(i,x) = \mathbb{1}_{\{x<0\}}(x), \quad \eta_{d,K} = \mathbb{1}_{\{x\geq 3\}}(x), \\ &\eta_k(i,x) = \mathbb{1}_{\{x\in[3(k-1)/(K-2),\ 3k/(K-2))\}}(x), \quad k = 1, \dots, K-2, \end{aligned}$$

for $i = 0, \dots, N-1$. The numerical procedure now depends of the number of time steps N, the dimension of the function basis K and the sample size Λ. For $j = 1, \dots, 11$ and $l = 3, \dots, 5$ they are fixed as

$$N = \left[2\sqrt{2}^{j-1}\right], \quad K = \max\left\{\left\lceil\sqrt{2}^{j-1}\right\rceil, 3\right\}, \quad \Lambda = \left[2\sqrt{2}^{l(j-1)}\right],$$

where $[a]$ is the closest integer to a and $\lceil a \rceil$ is the smallest integer larger or equal to a. To be precise, we will discuss three different choices of l, in which we simultaneously increase the parameters N, K and Λ through their dependence on j. The results in Lemor et al. (2006) suggest that in this one-dimensional setting the L^2-approximation error converges to zero at a rate of $N^{-\beta/2}$ for $0 < \beta \leq 1$ (up to a logarithmic factor), if the number of sample paths Λ is proportional to $N^{2+\beta}K^2$. With our choice of the parameters, we hence expect convergence of order 1/2 for $l = 5$. The choice $l = 4$ is just on the threshold of the theoretical convergence results, and we hence cannot expect convergence for $l = 3$. In order to illustrate the computational effort, the table below displays the number of simulated paths for the different choices of l in dependence of the number of time steps N.

Table 1: Sample size Λ in dependence of N and l

N	2	3	4	6	8	11	16	23	32	45	64
l											
3	2	6	17	46	129	363	1 025	2 897	8 193	23 171	65 537
4	2	9	33	129	513	2 049	8 193	32 769	131 073	524 289	2 097 153
5	2	12	65	363	2049	11 586	65 537	370 728	2 097 153	11 863 284	67 108 865

Note that the choice $l = 5$, which corresponds to convergence of order 1/2, yields tremendous simulation costs.

Given these parameters, we compute the coefficients $\hat{\alpha}_{0,i}^{\pi_N}$ and $\hat{\alpha}_{1,i}^{\pi_N}$ for the linear combination of the basis function by least-squares Monte Carlo and and receive the approximate solution by setting

$$\hat{Y}_{t_i}^{\pi_N} = \eta(i, X_{t_i}^{\pi_N,E})\hat{\alpha}_{0,i}^{\pi}, \quad \hat{Z}_{t_i}^{\pi_N} = \eta(i, X_{t_i}^{\pi_N,E})\hat{\alpha}_{1,i}^{\pi}.$$

We here recall that $X_{t_i}^{\pi_N,E}$ denotes the Euler scheme. Thanks to (25) we observe that the squared approximation error on the grid is given by

$$\begin{aligned} &\max_{0\leq i\leq N} E_0|Y_{t_i} - \hat{Y}_{t_i}^{\pi_N}|^2 + \sum_{i=1}^{N-1} \frac{T}{N} E_0|Z_{t_i} - \hat{Z}_{t_i}^{\pi_N}|^2 \\ = &\max_{0\leq i\leq N} E_0|\sin(X_{t_i}^{\pi_N,MS}) - \hat{Y}_{t_i}^{\pi_N}|^2 + \sum_{i=1}^{N-1} \frac{T}{N} E_0|\sigma\cos(X_{t_i}^{\pi_N,MS})\sin(X_{t_i}^{\pi_N,MS}) - \hat{Z}_{t_i}^{\pi_N}|^2 + \mathcal{O}(N^{-2}), \end{aligned} \tag{28}$$

where $X^{\pi_N,MS}$ indicates the approximation of X by the Milstein scheme. Here again E_0 denotes the conditional expectation given $\mathcal{G}_0$, where $\mathcal{G}_t = \mathcal{F}_t \vee \sigma(\Xi)$, and Ξ is the collection of random variables generated to determine the coefficients for the least-squares Monte Carlo estimator.

Figure 1 displays a log-log-plot of the right-hand side of (28) with the expectations replaced by the empirical mean using $1000N$ independent paths. We observe that the convergence behaviour in this example is much better than the theoretical bounds in Lemor et al. (2006) suggest. For the cases $l = 4$ and $l = 5$, Figure 1 indicates that the L^2-approximation error converges at a rate

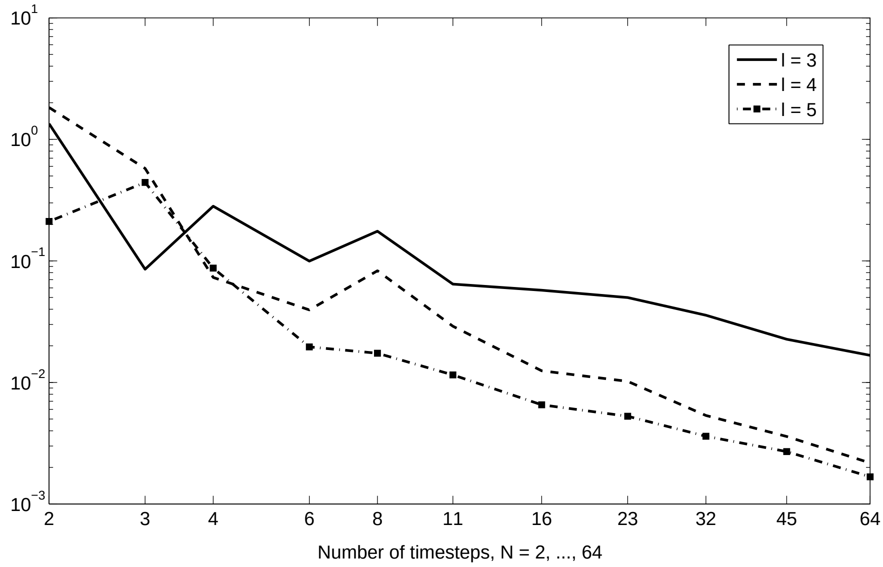

Figure 1: Squared approximation error for different choices of l.

of 1/2. For $l = 3$, the figure is less conclusive. The scheme seems to converge in this case as well (contrarily to what we expected in view of the theoretical results), but possibly at a lower rate.

We now show how to recover these results by applying the error criterion, which we introduced in this paper, without making use of the explicit form of the solution in (25). To this end we approximate the terminal condition based on the Milstein scheme and hence the error criterion becomes

$$\begin{aligned}&\mathcal{E}_{\pi_N}(\hat{Y}^{\pi_N}, \hat{Z}^{\pi_N})\\ =\;& E_0[|\sin(X_{t_N}^{\pi_N,MS}) - \hat{Y}_{t_N}^{\pi_N}|^2] + \max_{1\le i\le N} \mathrm{E}_0\left[\left|\hat{Y}_{t_i}^{\pi_N} - \hat{Y}_{t_0}^{\pi_N} - \sum_{j=0}^{i-1}\frac{1}{2}\sigma^2[(\hat{Y}_{t_j}^{\pi_N})^3]_{D^3}\Delta_j - \sum_{j=0}^{i-1}\hat{Z}_{t_j}^{\pi_N}\Delta W_j\right|^2\right].\end{aligned}$$

By Example 3.4 there are constants c and C such that squared approximation error on the whole interval satisfies

$$\max_{i=0,\dots,N-1}\sup_{t\in[t_i,t_{i+1}]}\mathrm{E}_0\left[|Y_t-\hat{Y}_{t_i}^{\pi_N}|^2\right] + \sum_{i=0}^{N-1}\mathrm{E}_0\left[\int_{t_i}^{t_{i+1}}|Z_t-\hat{Z}_{t_i}^{\pi_N}|^2dt\right] \le C\left(\mathcal{E}_{\pi_N}(\hat{Y}^{\pi_N},\hat{Z}^{\pi_N}) + \frac{1}{N}\right)$$

and

$$\max_{i=0,\dots,N-1}\sup_{t\in[t_i,t_{i+1}]}\mathrm{E}_0\left[|Y_t-\hat{Y}_{t_i}^{\pi_N}|^2\right] + \sum_{i=0}^{N-1}\mathrm{E}_0\left[\int_{t_i}^{t_{i+1}}|Z_t-\hat{Z}_{t_i}^{\pi_N}|^2dt\right] \ge c\,\mathcal{E}_{\pi_N}(\hat{Y}^{\pi_N},\hat{Z}^{\pi_N}) - \frac{1}{N^2}$$

for sufficiently large N. Consequently, the squared approximation error on the whole interval is equivalent to the error criterion for sufficiently large N, if

$$\mathcal{E}_{\pi_N}(\hat{Y}^{\pi_N},\hat{Z}^{\pi_N}) \ge const.\,\frac{1}{N}.$$

Figure 2 displays the a-posteriori error criterion for the three cases $l = 3, 4, 5$. As before, the

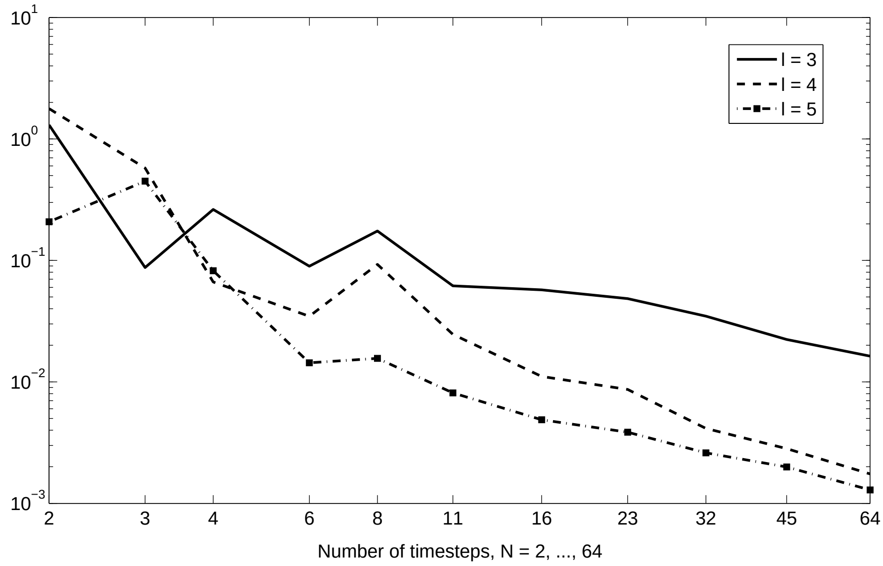

Figure 2: A-posteriori error criterion for different choices of l.

expectation is replaced by a sample mean over $1000N$ independent paths. Comparing Figures 1 and 2 we observe that the squared approximation error on the grid (Figure 1) and the error criterion (Figure 2) look almost identically, not only qualitatively, but also quantitatively. We can derive from Figure 2 that the a-posteriori error criterion converges to zero at a rate of N^{-1} for the cases $l = 4, 5$ and, consequently, the approximation error between approximate solution and true solution on the whole interval converges to zero at rate of $N^{-1/2}$ for these two cases. In particular, we can conclude that it is unnecessary to run the extensive scheme with $l = 5$ in this example, as an approximation of almost the same quality can be achieved with moderate simulation costs in the case $l = 4$. The cheap scheme with $l = 3$ leads, however, to a significantly larger error.

4.2 Case 2: Three-dimensional Brownian motion and global basis functions

We now consider the case of a three dimensional driving Brownian motion and apply a small global basis consisting of just a few monomials. The number of simulated paths will be adjusted in a way that the theoretical results in Lemor et al. (2006) support convergence of the simulation error of order 1/2 in the number of time steps. Apparently with the basis functions fixed, the scheme cannot converge to the true solution, but eventually, the projection error due to the choice of the basis will dominate the time discretization error and the simulation error. We now demonstrate how the error criterion can be applied to check whether a small global basis is sufficiently 'good' compared to the choice of the time grid.

As parameters of the BSDE we choose

$$D = 3, \quad T = 1, \quad s_{1,0} = s_{3,0} = \pi/2, \quad s_{2,0} = -\pi/2, \quad \sigma = 0.4 \ .$$

The basis consists of $K = 7$ functions, the constant function with value 1, the three monomials of order 1 and the three mixed monomial of order 2. The simulation parameters are given by

$$N = \left[2\sqrt{2}^{j-1}\right], \quad \Lambda = \left[2\sqrt{2}^{3(j-1)}\right],$$

which corresponds to a time discretization error and a simulation error that decrease with rate $N^{-1/2}$. We apply the a-posteriori error criterion to the original least-squares Monte Carlo scheme and to the modified one, which makes use of the non-linear control variate technique as described and designed at the beginning of this chapter. In contrast to the results in the previous section, the approximate terminal condition in the error criterion is based on the Euler scheme, i.e. $\xi^{\pi_N} = \sum_{d=1}^{3} \sin(X_{d,t_N}^{\pi_N,E})$.

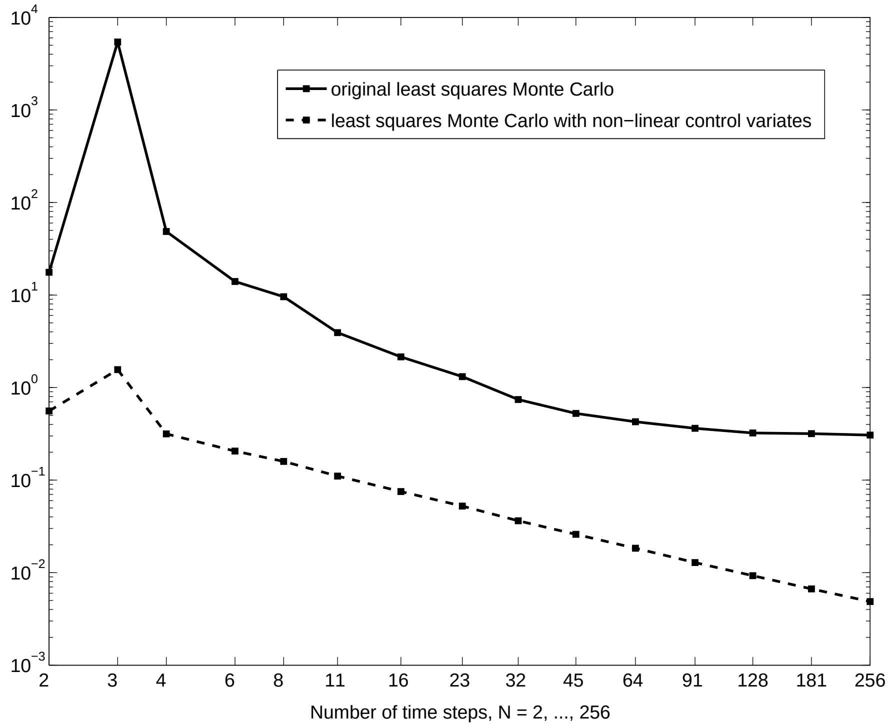

Figure 3: A-posteriori error criterion - original least squares Monte Carlo vs. least squares Monte Carlo with non-linear control variates

Figure 3 shows a log-log-plot for the error criterion of both implementations (with and without non-linear control variates) where, as before, the expectation is replaced by a sample mean with $1000N$ independent copies. In the case without control variates, the error criterion decreases roughly with a rate of N^{-1} for small values of N (roughly up to $N = 45$ time steps). This corresponds to an approximation error of order 1/2 which stems from the time discretization and the simulation. Starting from $N = 64$ the error criterion does not decrease significantly anymore, which suggests that the projection error dominates the two other error sources. A significant improvement of the approximative solution can, hence, not be achieved by increasing the number of time steps or the number of simulated paths, but a better choice of the basis would be called for. The situation turns out to be quite different for the implementation with non-linear control variates. Here, the error criterion decreases with a rate of N^{-1} for the whole range of time steps which we consider in this example (up to $N = 256$). This suggests that the small global basis consisting of seven monomials is sufficiently good in this situation and the dominating error sources are the time discretization and the simulation error even for a rather fine time grid consisting of 256 points. This example demonstrates on the one hand how the error criterion can be applied in order to judge the success of a small global basis. On the other hand it shows that the non-linear control variate technique can impressively improve the quality of the least-squares Monte Carlo scheme.

A Proof of inequality (13)

The proof follows the lines of Theorem 3.1 in Bouchard and Touzi (2004). We first define the process $\check{Y}^\pi$ on $[t_i, t_{i+1})$, $i = 0, \dots, N-1$, by

$$\check{Y}_t^\pi = \mathrm{E}\left[Y_{t_{i+1}}^\pi | \mathcal{F}_t\right] - f^\pi(t_i, Y_{t_i}^\pi, Z_{t_i}^\pi)(t_{i+1} - t)$$

with $\check{Y}_T^\pi = \xi^\pi$. Then, $\check{Y}_{t_i}^\pi = Y_{t_i}^\pi$ for $t_i \in \pi$ by (12). Moreover, thanks to (14), the pairs (Y, Z) and $(\check{Y}_t^\pi, \check{Z}_t^\pi)$ solve on $t \in [t_i, t_{i+1})$ the following differential equations

$$\begin{aligned}
Y_t &= Y_{t_{i+1}} - \int_t^{t_{i+1}} f(s, Y_s, Z_s) ds - \int_t^{t_{i+1}} Z_s dW_s, \\
\check{Y}_t^\pi &= Y_{t_{i+1}}^\pi - \int_t^{t_{i+1}} f^\pi(t_i, Y_{t_i}^\pi, Z_{t_i}^\pi) dt - \int_t^{t_{i+1}} \check{Z}_s^\pi dW_s,
\end{aligned}$$

By Itô's formula we then obtain

$$\begin{aligned}
&\mathrm{E}\left[\left|Y_t - \check{Y}_t^\pi\right|^2\right] + \int_t^{t_{i+1}} \mathrm{E}\left[\left|Z_s - \check{Z}_s^\pi\right|^2\right] ds \\
&\quad \leq \mathrm{E}\left[\left|Y_{t_{i+1}} - Y_{t_{i+1}}^\pi\right|^2\right] + 2\int_t^{t_{i+1}} \mathrm{E}\left[(Y_s - \check{Y}_s^\pi)\left(f(s, Y_s, Z_s) - f^\pi(t_i, Y_{t_i}^\pi, Z_{t_i}^\pi)\right)\right] ds \\
&\quad = (I) + (II).
\end{aligned}$$

Young's inequality and the Lipschitz condition on f^π yield for some $\gamma > 0$ (to be fixed later),

$$\begin{aligned}
(II) &\leq \gamma \int_t^{t_{i+1}} \mathrm{E}\left[\left|Y_s - \check{Y}_s^\pi\right|^2\right] ds + \frac{2}{\gamma}\int_{t_i}^{t_{i+1}} \mathrm{E}\left[\left|f^\pi(t_i, Y_s, Z_s) - f^\pi(t_i, Y_{t_i}^\pi, Z_{t_i}^\pi)\right|^2\right] ds \\
&\quad + \frac{2}{\gamma}\int_{t_i}^{t_{i+1}} \mathrm{E}\left[\left|f(s, Y_s, Z_s) - f^\pi(t_i, Y_s, Z_s)\right|^2\right] ds \\
&\leq \gamma \int_t^{t_{i+1}} \mathrm{E}\left[\left|Y_s - \check{Y}_s^\pi\right|^2\right] ds + \frac{4}{\gamma}\int_{t_i}^{t_{i+1}} K_y^2 \,\mathrm{E}\left[\left|Y_s - Y_{t_i}^\pi\right|^2\right] + K^2\,\mathrm{E}\left[\left|Z_s - Z_{t_i}^\pi\right|^2\right] ds \\
&\quad + C\int_{t_i}^{t_{i+1}} \mathrm{E}\left[\left|f(s, Y_s, Z_s) - f^\pi(t_i, Y_s, Z_s)\right|^2\right] ds.
\end{aligned}$$

Due to Zhang (2004), Lemma 2.4 (i) we have

$$\begin{aligned}
\mathrm{E}\left[\left|Y_s - Y_{t_i}^\pi\right|^2\right] &\leq 2\mathrm{E}\left[\left|Y_s - Y_{t_i}\right|^2\right] + 2\mathrm{E}\left[\left|Y_{t_i} - Y_{t_i}^\pi\right|^2\right] \\
&\leq C\left|\pi\right| + C\int_{t_i}^{t_{i+1}} \mathrm{E}\left[\left|Z_t\right|^2\right] dt + 2\mathrm{E}\left[\left|Y_{t_i} - Y_{t_i}^\pi\right|^2\right].
\end{aligned}$$

Applying Itô's isometry and Young's inequality once more, we get

$$\begin{aligned}
(II) \leq{}& \gamma \int_t^{t_{i+1}} \mathrm{E}\left[\left|Y_s - \check{Y}_s^\pi\right|^2\right] ds + \frac{8K^2}{\gamma}\left(\Delta_i \mathrm{E}\left[\left|Y_{t_i} - Y_{t_i}^\pi\right|^2\right] + \int_{t_i}^{t_{i+1}} \mathrm{E}\left[\left|Z_s - \check{Z}_s^\pi\right|^2\right] ds\right) \\
&+ K_y\left(C\left|\pi\right|\Delta_i + C\Delta_i \int_{t_i}^{t_{i+1}} \mathrm{E}\left[\left|Z_t\right|^2\right] dt\right) + C\mathrm{E}\left[\left(\int_{t_i}^{t_{i+1}} \check{Z}_s^\pi dW_s - Z_{t_i}^\pi \Delta W_i\right)^2\right] \\
&+ C\int_{t_i}^{t_{i+1}} \mathrm{E}\left[\left|f(s, Y_s, Z_s) - f^\pi(t_i, Y_s, Z_s)\right|^2\right] ds \\
=:{}& \gamma \int_t^{t_{i+1}} \mathrm{E}\left[\left|Y_s - \check{Y}_s^\pi\right|^2\right] ds + \frac{8K^2}{\gamma} A_i + B_i.
\end{aligned}$$

Summarizing, we have

$$\begin{aligned}
\mathrm{E}\left[\left|Y_t - \check{Y}_t^\pi\right|^2\right] &\leq \mathrm{E}\left[\left|Y_t - \check{Y}_t^\pi\right|^2\right] + \int_t^{t_{i+1}} \mathrm{E}\left[\left|Z_s - \check{Z}_s^\pi\right|^2\right] ds \\
&\leq \mathrm{E}\left[\left|Y_{t_{i+1}} - Y_{t_{i+1}}^\pi\right|^2\right] + \gamma \int_t^{t_{i+1}} \mathrm{E}\left[\left|Y_s - \check{Y}_s^\pi\right|^2\right] ds + \frac{8K^2}{\gamma} A_i + B_i.
\end{aligned} \tag{29}$$

By Gronwall's lemma it follows that $\mathrm{E}\left[|Y_t - \check{Y}_t^\pi|^2\right] \leq e^{\gamma\Delta_i}(\mathrm{E}\left[|Y_{t_{i+1}} - Y_{t_{i+1}}^\pi|^2\right] + 8K^2 A_i/\gamma + B_i)$. Inserting this result into the second inequality of (29) yields

$$\begin{aligned}
&\mathrm{E}\left[\left|Y_{t_i} - Y_{t_i}^\pi\right|^2\right] + \int_{t_i}^{t_{i+1}} \mathrm{E}\left[\left|Z_t - \check{Z}_t^\pi\right|^2\right] dt \\
&\quad\leq (1 + \gamma\Delta_i e^{\gamma\Delta_i})(\mathrm{E}\left[\left|Y_{t_{i+1}} - Y_{t_{i+1}}^\pi\right|^2\right] + \frac{8K^2}{\gamma} A_i + B_i) \\
&\quad\leq (1 + C\gamma\Delta_i)(\mathrm{E}\left[\left|Y_{t_{i+1}} - Y_{t_{i+1}}^\pi\right|^2\right] + \frac{8K^2}{\gamma} A_i + B_i)
\end{aligned}$$

for $|\pi|$ small enough. Choosing $\gamma = 32K^2$ and $|\pi| \leq 1/(C\gamma)$ leads to

$$\begin{aligned}
\mathrm{E}\left[\left|Y_{t_i} - Y_{t_i}^\pi\right|^2\right] + \int_{t_i}^{t_{i+1}} \mathrm{E}\left[\left|Z_t - \check{Z}_t^\pi\right|^2\right] dt &\leq (1 + C\gamma\Delta_i)\left(\mathrm{E}\left[\left|Y_{t_{i+1}} - Y_{t_{i+1}}^\pi\right|^2\right] + B_i\right) \\
&\quad + \frac{1}{2}\Delta_i \mathrm{E}\left[\left|Y_{t_i} - Y_{t_i}^\pi\right|^2\right] + \frac{1}{2}\int_{t_i}^{t_{i+1}} \mathrm{E}\left[\left|Z_t - \check{Z}_t^\pi\right|^2\right] dt.
\end{aligned}$$

Hence, for $|\pi|$ small enough

$$\begin{aligned}
&\mathrm{E}\left[\left|Y_{t_i}-Y^\pi_{t_i}\right|^2\right]+\frac{1}{2}\int_{t_i}^{t_{i+1}}\mathrm{E}\left[\left|Z_t-\check{Z}^\pi_t\right|^2\right]dt\\
&\quad\le(1+C\Delta_i)\Big\{\mathrm{E}\left[\left|Y_{t_{i+1}}-Y^\pi_{t_{i+1}}\right|^2\right]+K_y\left(C\,|\pi|\,\Delta_i+C\Delta_i\int_{t_i}^{t_{i+1}}\mathrm{E}\left[|Z_t|^2\right]dt\right)\\
&\qquad+C\mathrm{E}\left[\left(\int_{t_i}^{t_{i+1}}\check{Z}^\pi_s dW_s-Z^\pi_{t_i}\Delta W_i\right)^2\right]\\
&\qquad+C\int_{t_i}^{t_{i+1}}\mathrm{E}\left[|f(t,Y_t,Z_t)-f^\pi(t_i,Y_t,Z_t)|^2\right]dt\Big\}.
\end{aligned}\tag{30}$$

Thanks to the discrete Gronwall lemma we get for some larger constant C

$$\begin{aligned}
\mathrm{E}\left[\left|Y_{t_i}-Y^\pi_{t_i}\right|^2\right]\quad&\le\quad C\Big\{\mathrm{E}\left[|\xi-\xi^\pi|^2\right]+K_y\,|\pi|+\sum_{j=i}^{N-1}\mathrm{E}\left[\left(\int_{t_i}^{t_{i+1}}\check{Z}^\pi_s dW_s-Z^\pi_{t_i}\Delta W_i\right)^2\right]\\
&\qquad+\sum_{j=i}^{N-1}\int_{t_j}^{t_{j+1}}\mathrm{E}\left[|f(t,Y_t,Z_t)-f^\pi(t_j,Y_t,Z_t)|^2\right]dt\Big\},
\end{aligned}\tag{31}$$

because $\int_0^T\mathrm{E}\left[|Z_t|^2\right]dt<\infty$. Next we sum (30) up from $i=0$ to $N-1$ and obtain

$$\begin{aligned}
&\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}}\mathrm{E}\left[\left|Z_t-\check{Z}^\pi_t\right|^2\right]dt\\
\le\quad&C\Big\{\mathrm{E}\left[|\xi-\xi^\pi|^2\right]+K_y\,|\pi|+\sum_{j=i}^{N-1}\mathrm{E}\left[\left(\int_{t_i}^{t_{i+1}}\check{Z}^\pi_s dW_s-Z^\pi_{t_i}\Delta W_i\right)^2\right]\\
&+\sum_{j=i}^{N-1}\int_{t_j}^{t_{j+1}}\mathrm{E}\left[|f(t,Y_t,Z_t)-f^\pi(t_j,Y_t,Z_t)|^2\right]dt+\max_{t_i\in\pi}\mathrm{E}\left[\left|Y_{t_i}-Y^\pi_{t_i}\right|^2\right]\Big\}
\end{aligned}\tag{32}$$

Noting that

$$\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}}\mathrm{E}\left[\left|Z_t-Z^\pi_{t_i}\right|^2\right]dt\le2\sum_{i=0}^{N-1}\left(\int_{t_i}^{t_{i+1}}\mathrm{E}\left[\left|Z_t-\check{Z}^\pi_t\right|^2\right]dt+\mathrm{E}\left[\left(\int_{t_i}^{t_{i+1}}\check{Z}^\pi_s dW_s-Z^\pi_{t_i}\Delta W_i\right)^2\right]\right),$$

inequality (13) is a direct consequence of (31) and (32).

Acknowledgement

Financial support by the Deutsche Forschungsgemeinschaft under grant BE3933/3-1 is gratefully acknowledged.

References

Bally, V., Pagès, G.: A quantization algorithm for solving multi-dimensional discrete-time optimal stopping problems. Bernoulli **9**, 1003–1049 (2003).

Bender, C., Denk, R.: A forward scheme for backward SDEs. Stochastic Process. Appl. **117**, 1793–1812 (2007).

Bender, C., Steiner, J.: Least-squares Monte Carlo for BSDEs. In: Carmona, R. A. et al. (eds.) Numerical methods in finance, pp. 257–289. Springer (2012).

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

Bismut, J. M.: Conjugate convex functions in optimal stochastic control. J. Math. Anal. Apl. **44**, 384 – 404 (1973).

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

Crisan, D., Manolarakis, K.: Solving backward stochastic differential equations using the cubature method. Preprint (2010).

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

Geiss, C., Geiss, S., Gobet, E.: Fractional smoothness and applications to finance. In: Di Nunno, G., Øksendal, B. (eds.) Advanced Mathematical Methods for Finance, pp. 313–332, Springer (2011).

Gobet, E., Labart, C.: Error expansion for the discretization of backward stochastic differential equations. Stochastic Process. Appl. **117**, 803–829 (2007).

Gobet, E., Makhlouf, A.: L^2-time regularity of BSDEs with irregular terminal functions. Stochastic Process. Appl. **120**, 1105–1132 (2010).

Lemor, J.-P., Gobet, E., Warin, X.: Rate of convergence of an empirical regression method for solving generalized backward stochastic differential equations. Bernoulli **12**, 889–916 (2006).

Longstaff, F. A., Schwartz, R. S.: Valuing American options by simulation: A simple least-squares approach. Rev. Financ. Stud. **14**, 113–147 (2001).

Pardoux, E., Peng, S. G.: Adapted solution of a backward stochastic differential equation. System Control Lett. **14**, 55–61 (1990).

Pardoux, E., Peng, S. G.: Backward stochastic differential equations and quasiliner parabolic partial differential equations. In: Rozovskii, B. L., Sowers, R. S. (eds.) Stochastic partial differential equations and their applications (Lect. Notes Control. Inf. Sci. 176, pp. 200 – 217), Springer (1992).

Peng, S.: Backward SDE and related g-expectations. In: El Karoui, N., Mazliak, L. (eds.) Backward stochastic differential equations, pp. 141–159, Longman (1997).

Yong, J., Zhou, X. Y.: Stochastic controls: Hamiltonian systems and HJB equations. Springer (1999).

Zhang, J.: A numerical scheme for BSDEs, Ann. Appl. Probab. **14**, 459–488 (2004).

Preprint Series DFG-SPP 1324

http://www.dfg-spp1324.de

Reports

[1] R. Ramlau, G. Teschke, and M. Zhariy. A Compressive Landweber Iteration for Solving Ill-Posed Inverse Problems. Preprint 1, DFG-SPP 1324, September 2008.

[2] G. Plonka. The Easy Path Wavelet Transform: A New Adaptive Wavelet Transform for Sparse Representation of Two-dimensional Data. Preprint 2, DFG-SPP 1324, September 2008.

[3] E. Novak and H. Woźniakowski. Optimal Order of Convergence and (In-) Tractability of Multivariate Approximation of Smooth Functions. Preprint 3, DFG-SPP 1324, October 2008.

[4] M. Espig, L. Grasedyck, and W. Hackbusch. Black Box Low Tensor Rank Approximation Using Fibre-Crosses. Preprint 4, DFG-SPP 1324, October 2008.

[5] T. Bonesky, S. Dahlke, P. Maass, and T. Raasch. Adaptive Wavelet Methods and Sparsity Reconstruction for Inverse Heat Conduction Problems. Preprint 5, DFG-SPP 1324, January 2009.

[6] E. Novak and H. Woźniakowski. Approximation of Infinitely Differentiable Multivariate Functions Is Intractable. Preprint 6, DFG-SPP 1324, January 2009.

[7] J. Ma and G. Plonka. A Review of Curvelets and Recent Applications. Preprint 7, DFG-SPP 1324, February 2009.

[8] L. Denis, D. A. Lorenz, and D. Trede. Greedy Solution of Ill-Posed Problems: Error Bounds and Exact Inversion. Preprint 8, DFG-SPP 1324, April 2009.

[9] U. Friedrich. A Two Parameter Generalization of Lions' Nonoverlapping Domain Decomposition Method for Linear Elliptic PDEs. Preprint 9, DFG-SPP 1324, April 2009.

[10] K. Bredies and D. A. Lorenz. Minimization of Non-smooth, Non-convex Functionals by Iterative Thresholding. Preprint 10, DFG-SPP 1324, April 2009.

[11] K. Bredies and D. A. Lorenz. Regularization with Non-convex Separable Constraints. Preprint 11, DFG-SPP 1324, April 2009.

[12] M. Döhler, S. Kunis, and D. Potts. Nonequispaced Hyperbolic Cross Fast Fourier Transform. Preprint 12, DFG-SPP 1324, April 2009.

[13] C. Bender. Dual Pricing of Multi-Exercise Options under Volume Constraints. Preprint 13, DFG-SPP 1324, April 2009.

[14] T. Müller-Gronbach and K. Ritter. Variable Subspace Sampling and Multi-level Algorithms. Preprint 14, DFG-SPP 1324, May 2009.

[15] G. Plonka, S. Tenorth, and A. Iske. Optimally Sparse Image Representation by the Easy Path Wavelet Transform. Preprint 15, DFG-SPP 1324, May 2009.

[16] S. Dahlke, E. Novak, and W. Sickel. Optimal Approximation of Elliptic Problems by Linear and Nonlinear Mappings IV: Errors in L_2 and Other Norms. Preprint 16, DFG-SPP 1324, June 2009.

[17] B. Jin, T. Khan, P. Maass, and M. Pidcock. Function Spaces and Optimal Currents in Impedance Tomography. Preprint 17, DFG-SPP 1324, June 2009.

[18] G. Plonka and J. Ma. Curvelet-Wavelet Regularized Split Bregman Iteration for Compressed Sensing. Preprint 18, DFG-SPP 1324, June 2009.

[19] G. Teschke and C. Borries. Accelerated Projected Steepest Descent Method for Nonlinear Inverse Problems with Sparsity Constraints. Preprint 19, DFG-SPP 1324, July 2009.

[20] L. Grasedyck. Hierarchical Singular Value Decomposition of Tensors. Preprint 20, DFG-SPP 1324, July 2009.

[21] D. Rudolf. Error Bounds for Computing the Expectation by Markov Chain Monte Carlo. Preprint 21, DFG-SPP 1324, July 2009.

[22] M. Hansen and W. Sickel. Best m-term Approximation and Lizorkin-Triebel Spaces. Preprint 22, DFG-SPP 1324, August 2009.

[23] F.J. Hickernell, T. Müller-Gronbach, B. Niu, and K. Ritter. Multi-level Monte Carlo Algorithms for Infinite-dimensional Integration on $\mathbb{R}^{\mathbb{N}}$. Preprint 23, DFG-SPP 1324, August 2009.

[24] S. Dereich and F. Heidenreich. A Multilevel Monte Carlo Algorithm for Lévy Driven Stochastic Differential Equations. Preprint 24, DFG-SPP 1324, August 2009.

[25] S. Dahlke, M. Fornasier, and T. Raasch. Multilevel Preconditioning for Adaptive Sparse Optimization. Preprint 25, DFG-SPP 1324, August 2009.

[26] S. Dereich. Multilevel Monte Carlo Algorithms for Lévy-driven SDEs with Gaussian Correction. Preprint 26, DFG-SPP 1324, August 2009.

[27] G. Plonka, S. Tenorth, and D. Roşca. A New Hybrid Method for Image Approximation using the Easy Path Wavelet Transform. Preprint 27, DFG-SPP 1324, October 2009.

[28] O. Koch and C. Lubich. Dynamical Low-rank Approximation of Tensors. Preprint 28, DFG-SPP 1324, November 2009.

[29] E. Faou, V. Gradinaru, and C. Lubich. Computing Semi-classical Quantum Dynamics with Hagedorn Wavepackets. Preprint 29, DFG-SPP 1324, November 2009.

[30] D. Conte and C. Lubich. An Error Analysis of the Multi-configuration Time-dependent Hartree Method of Quantum Dynamics. Preprint 30, DFG-SPP 1324, November 2009.

[31] C. E. Powell and E. Ullmann. Preconditioning Stochastic Galerkin Saddle Point Problems. Preprint 31, DFG-SPP 1324, November 2009.

[32] O. G. Ernst and E. Ullmann. Stochastic Galerkin Matrices. Preprint 32, DFG-SPP 1324, November 2009.

[33] F. Lindner and R. L. Schilling. Weak Order for the Discretization of the Stochastic Heat Equation Driven by Impulsive Noise. Preprint 33, DFG-SPP 1324, November 2009.

[34] L. Kämmerer and S. Kunis. On the Stability of the Hyperbolic Cross Discrete Fourier Transform. Preprint 34, DFG-SPP 1324, December 2009.

[35] P. Cerejeiras, M. Ferreira, U. Kähler, and G. Teschke. Inversion of the noisy Radon transform on $SO(3)$ by Gabor frames and sparse recovery principles. Preprint 35, DFG-SPP 1324, January 2010.

[36] T. Jahnke and T. Udrescu. Solving Chemical Master Equations by Adaptive Wavelet Compression. Preprint 36, DFG-SPP 1324, January 2010.

[37] P. Kittipoom, G. Kutyniok, and W.-Q Lim. Irregular Shearlet Frames: Geometry and Approximation Properties. Preprint 37, DFG-SPP 1324, February 2010.

[38] G. Kutyniok and W.-Q Lim. Compactly Supported Shearlets are Optimally Sparse. Preprint 38, DFG-SPP 1324, February 2010.

[39] M. Hansen and W. Sickel. Best m-Term Approximation and Tensor Products of Sobolev and Besov Spaces – the Case of Non-compact Embeddings. Preprint 39, DFG-SPP 1324, March 2010.

[40] B. Niu, F.J. Hickernell, T. Müller-Gronbach, and K. Ritter. Deterministic Multi-level Algorithms for Infinite-dimensional Integration on $\mathbb{R}^{\mathbb{N}}$. Preprint 40, DFG-SPP 1324, March 2010.

[41] P. Kittipoom, G. Kutyniok, and W.-Q Lim. Construction of Compactly Supported Shearlet Frames. Preprint 41, DFG-SPP 1324, March 2010.

[42] C. Bender and J. Steiner. Error Criteria for Numerical Solutions of Backward SDEs. Preprint 42, DFG-SPP 1324, April 2010.

[43] L. Grasedyck. Polynomial Approximation in Hierarchical Tucker Format by Vector-Tensorization. Preprint 43, DFG-SPP 1324, April 2010.

[44] M. Hansen und W. Sickel. Best m-Term Approximation and Sobolev-Besov Spaces of Dominating Mixed Smoothness - the Case of Compact Embeddings. Preprint 44, DFG-SPP 1324, April 2010.

[45] P. Binev, W. Dahmen, and P. Lamby. Fast High-Dimensional Approximation with Sparse Occupancy Trees. Preprint 45, DFG-SPP 1324, May 2010.

[46] J. Ballani and L. Grasedyck. A Projection Method to Solve Linear Systems in Tensor Format. Preprint 46, DFG-SPP 1324, May 2010.

[47] P. Binev, A. Cohen, W. Dahmen, R. DeVore, G. Petrova, and P. Wojtaszczyk. Convergence Rates for Greedy Algorithms in Reduced Basis Methods. Preprint 47, DFG-SPP 1324, May 2010.

[48] S. Kestler and K. Urban. Adaptive Wavelet Methods on Unbounded Domains. Preprint 48, DFG-SPP 1324, June 2010.

[49] H. Yserentant. The Mixed Regularity of Electronic Wave Functions Multiplied by Explicit Correlation Factors. Preprint 49, DFG-SPP 1324, June 2010.

[50] H. Yserentant. On the Complexity of the Electronic Schrödinger Equation. Preprint 50, DFG-SPP 1324, June 2010.

[51] M. Guillemard and A. Iske. Curvature Analysis of Frequency Modulated Manifolds in Dimensionality Reduction. Preprint 51, DFG-SPP 1324, June 2010.

[52] E. Herrholz and G. Teschke. Compressive Sensing Principles and Iterative Sparse Recovery for Inverse and Ill-Posed Problems. Preprint 52, DFG-SPP 1324, July 2010.

[53] L. Kämmerer, S. Kunis, and D. Potts. Interpolation Lattices for Hyperbolic Cross Trigonometric Polynomials. Preprint 53, DFG-SPP 1324, July 2010.

[54] G. Kutyniok and W.-Q Lim. Shearlets on Bounded Domains. Preprint 54, DFG-SPP 1324, July 2010.

[55] A. Zeiser. Wavelet Approximation in Weighted Sobolev Spaces of Mixed Order with Applications to the Electronic Schrödinger Equation. Preprint 55, DFG-SPP 1324, July 2010.

[56] G. Kutyniok, J. Lemvig, and W.-Q Lim. Compactly Supported Shearlets. Preprint 56, DFG-SPP 1324, July 2010.

[57] A. Zeiser. On the Optimality of the Inexact Inverse Iteration Coupled with Adaptive Finite Element Methods. Preprint 57, DFG-SPP 1324, July 2010.

[58] S. Jokar. Sparse Recovery and Kronecker Products. Preprint 58, DFG-SPP 1324, August 2010.

[59] T. Aboiyar, E. H. Georgoulis, and A. Iske. Adaptive ADER Methods Using Kernel-Based Polyharmonic Spline WENO Reconstruction. Preprint 59, DFG-SPP 1324, August 2010.

[60] O. G. Ernst, A. Mugler, H.-J. Starkloff, and E. Ullmann. On the Convergence of Generalized Polynomial Chaos Expansions. Preprint 60, DFG-SPP 1324, August 2010.

[61] S. Holtz, T. Rohwedder, and R. Schneider. On Manifolds of Tensors of Fixed TT-Rank. Preprint 61, DFG-SPP 1324, September 2010.

[62] J. Ballani, L. Grasedyck, and M. Kluge. Black Box Approximation of Tensors in Hierarchical Tucker Format. Preprint 62, DFG-SPP 1324, October 2010.

[63] M. Hansen. On Tensor Products of Quasi-Banach Spaces. Preprint 63, DFG-SPP 1324, October 2010.

[64] S. Dahlke, G. Steidl, and G. Teschke. Shearlet Coorbit Spaces: Compactly Supported Analyzing Shearlets, Traces and Embeddings. Preprint 64, DFG-SPP 1324, October 2010.

[65] W. Hackbusch. Tensorisation of Vectors and their Efficient Convolution. Preprint 65, DFG-SPP 1324, November 2010.

[66] P. A. Cioica, S. Dahlke, S. Kinzel, F. Lindner, T. Raasch, K. Ritter, and R. L. Schilling. Spatial Besov Regularity for Stochastic Partial Differential Equations on Lipschitz Domains. Preprint 66, DFG-SPP 1324, November 2010.

[67] E. Novak and H. Woźniakowski. On the Power of Function Values for the Approximation Problem in Various Settings. Preprint 67, DFG-SPP 1324, November 2010.

[68] A. Hinrichs, E. Novak, and H. Woźniakowski. The Curse of Dimensionality for Monotone and Convex Functions of Many Variables. Preprint 68, DFG-SPP 1324, November 2010.

[69] G. Kutyniok and W.-Q Lim. Image Separation Using Shearlets. Preprint 69, DFG-SPP 1324, November 2010.

[70] B. Jin and P. Maass. An Analysis of Electrical Impedance Tomography with Applications to Tikhonov Regularization. Preprint 70, DFG-SPP 1324, December 2010.

[71] S. Holtz, T. Rohwedder, and R. Schneider. The Alternating Linear Scheme for Tensor Optimisation in the TT Format. Preprint 71, DFG-SPP 1324, December 2010.

[72] T. Müller-Gronbach and K. Ritter. A Local Refinement Strategy for Constructive Quantization of Scalar SDEs. Preprint 72, DFG-SPP 1324, December 2010.

[73] T. Rohwedder and R. Schneider. An Analysis for the DIIS Acceleration Method used in Quantum Chemistry Calculations. Preprint 73, DFG-SPP 1324, December 2010.

[74] C. Bender and J. Steiner. Least-Squares Monte Carlo for Backward SDEs. Preprint 74, DFG-SPP 1324, December 2010.

[75] C. Bender. Primal and Dual Pricing of Multiple Exercise Options in Continuous Time. Preprint 75, DFG-SPP 1324, December 2010.

[76] H. Harbrecht, M. Peters, and R. Schneider. On the Low-rank Approximation by the Pivoted Cholesky Decomposition. Preprint 76, DFG-SPP 1324, December 2010.

[77] P. A. Cioica, S. Dahlke, N. Döhring, S. Kinzel, F. Lindner, T. Raasch, K. Ritter, and R. L. Schilling. Adaptive Wavelet Methods for Elliptic Stochastic Partial Differential Equations. Preprint 77, DFG-SPP 1324, January 2011.

[78] G. Plonka, S. Tenorth, and A. Iske. Optimal Representation of Piecewise Hölder Smooth Bivariate Functions by the Easy Path Wavelet Transform. Preprint 78, DFG-SPP 1324, January 2011.

[79] A. Mugler and H.-J. Starkloff. On Elliptic Partial Differential Equations with Random Coefficients. Preprint 79, DFG-SPP 1324, January 2011.

[80] T. Müller-Gronbach, K. Ritter, and L. Yaroslavtseva. A Derandomization of the Euler Scheme for Scalar Stochastic Differential Equations. Preprint 80, DFG-SPP 1324, January 2011.

[81] W. Dahmen, C. Huang, C. Schwab, and G. Welper. Adaptive Petrov-Galerkin methods for first order transport equations. Preprint 81, DFG-SPP 1324, January 2011.

[82] K. Grella and C. Schwab. Sparse Tensor Spherical Harmonics Approximation in Radiative Transfer. Preprint 82, DFG-SPP 1324, January 2011.

[83] D.A. Lorenz, S. Schiffler, and D. Trede. Beyond Convergence Rates: Exact Inversion With Tikhonov Regularization With Sparsity Constraints. Preprint 83, DFG-SPP 1324, January 2011.

[84] S. Dereich, M. Scheutzow, and R. Schottstedt. Constructive quantization: Approximation by empirical measures. Preprint 84, DFG-SPP 1324, January 2011.

[85] S. Dahlke and W. Sickel. On Besov Regularity of Solutions to Nonlinear Elliptic Partial Differential Equations. Preprint 85, DFG-SPP 1324, January 2011.

[86] S. Dahlke, U. Friedrich, P. Maass, T. Raasch, and R.A. Ressel. An adaptive wavelet method for parameter identification problems in parabolic partial differential equations. Preprint 86, DFG-SPP 1324, January 2011.

[87] A. Cohen, W. Dahmen, and G. Welper. Adaptivity and Variational Stabilization for Convection-Diffusion Equations. Preprint 87, DFG-SPP 1324, January 2011.

[88] T. Jahnke. On Reduced Models for the Chemical Master Equation. Preprint 88, DFG-SPP 1324, January 2011.

[89] P. Binev, W. Dahmen, R. DeVore, P. Lamby, D. Savu, and R. Sharpley. Compressed Sensing and Electron Microscopy. Preprint 89, DFG-SPP 1324, March 2011.

[90] P. Binev, F. Blanco-Silva, D. Blom, W. Dahmen, P. Lamby, R. Sharpley, and T. Vogt. High Quality Image Formation by Nonlocal Means Applied to High-Angle Annular Dark Field Scanning Transmission Electron Microscopy (HAADF-STEM). Preprint 90, DFG-SPP 1324, March 2011.

[91] R. A. Ressel. A Parameter Identification Problem for a Nonlinear Parabolic Differential Equation. Preprint 91, DFG-SPP 1324, May 2011.

[92] G. Kutyniok. Data Separation by Sparse Representations. Preprint 92, DFG-SPP 1324, May 2011.

[93] M. A. Davenport, M. F. Duarte, Y. C. Eldar, and G. Kutyniok. Introduction to Compressed Sensing. Preprint 93, DFG-SPP 1324, May 2011.

[94] H.-C. Kreusler and H. Yserentant. The Mixed Regularity of Electronic Wave Functions in Fractional Order and Weighted Sobolev Spaces. Preprint 94, DFG-SPP 1324, June 2011.

[95] E. Ullmann, H. C. Elman, and O. G. Ernst. Efficient Iterative Solvers for Stochastic Galerkin Discretizations of Log-Transformed Random Diffusion Problems. Preprint 95, DFG-SPP 1324, June 2011.

[96] S. Kunis and I. Melzer. On the Butterfly Sparse Fourier Transform. Preprint 96, DFG-SPP 1324, June 2011.

[97] T. Rohwedder. The Continuous Coupled Cluster Formulation for the Electronic Schrödinger Equation. Preprint 97, DFG-SPP 1324, June 2011.

[98] T. Rohwedder and R. Schneider. Error Estimates for the Coupled Cluster Method. Preprint 98, DFG-SPP 1324, June 2011.

[99] P. A. Cioica and S. Dahlke. Spatial Besov Regularity for Semilinear Stochastic Partial Differential Equations on Bounded Lipschitz Domains. Preprint 99, DFG-SPP 1324, July 2011.

[100] L. Grasedyck and W. Hackbusch. An Introduction to Hierarchical (H-) Rank and TT-Rank of Tensors with Examples. Preprint 100, DFG-SPP 1324, August 2011.

[101] N. Chegini, S. Dahlke, U. Friedrich, and R. Stevenson. Piecewise Tensor Product Wavelet Bases by Extensions and Approximation Rates. Preprint 101, DFG-SPP 1324, September 2011.

[102] S. Dahlke, P. Oswald, and T. Raasch. A Note on Quarkonial Systems and Multilevel Partition of Unity Methods. Preprint 102, DFG-SPP 1324, September 2011.

[103] A. Uschmajew. Local Convergence of the Alternating Least Squares Algorithm For Canonical Tensor Approximation. Preprint 103, DFG-SPP 1324, September 2011.

[104] S. Kvaal. Multiconfigurational time-dependent Hartree method for describing particle loss due to absorbing boundary conditions. Preprint 104, DFG-SPP 1324, September 2011.

[105] M. Guillemard and A. Iske. On Groupoid C*-Algebras, Persistent Homology and Time-Frequency Analysis. Preprint 105, DFG-SPP 1324, September 2011.

[106] A. Hinrichs, E. Novak, and H. Woźniakowski. Discontinuous information in the worst case and randomized settings. Preprint 106, DFG-SPP 1324, September 2011.

[107] M. Espig, W. Hackbusch, A. Litvinenko, H. Matthies, and E. Zander. Efficient Analysis of High Dimensional Data in Tensor Formats. Preprint 107, DFG-SPP 1324, September 2011.

[108] M. Espig, W. Hackbusch, S. Handschuh, and R. Schneider. Optimization Problems in Contracted Tensor Networks. Preprint 108, DFG-SPP 1324, October 2011.

[109] S. Dereich, T. Müller-Gronbach, and K. Ritter. On the Complexity of Computing Quadrature Formulas for SDEs. Preprint 109, DFG-SPP 1324, October 2011.

[110] D. Belomestny. Solving optimal stopping problems by empirical dual optimization and penalization. Preprint 110, DFG-SPP 1324, November 2011.

[111] D. Belomestny and J. Schoenmakers. Multilevel dual approach for pricing American style derivatives. Preprint 111, DFG-SPP 1324, November 2011.

[112] T. Rohwedder and A. Uschmajew. Local convergence of alternating schemes for optimization of convex problems in the TT format. Preprint 112, DFG-SPP 1324, December 2011.

[113] T. Görner, R. Hielscher, and S. Kunis. Efficient and accurate computation of spherical mean values at scattered center points. Preprint 113, DFG-SPP 1324, December 2011.

[114] Y. Dong, T. Görner, and S. Kunis. An iterative reconstruction scheme for photoacoustic imaging. Preprint 114, DFG-SPP 1324, December 2011.

[115] L. Kämmerer. Reconstructing hyperbolic cross trigonometric polynomials by sampling along generated sets. Preprint 115, DFG-SPP 1324, February 2012.

[116] H. Chen and R. Schneider. Numerical analysis of augmented plane waves methods for full-potential electronic structure calculations. Preprint 116, DFG-SPP 1324, February 2012.

[117] J. Ma, G. Plonka, and M.Y. Hussaini. Compressive Video Sampling with Approximate Message Passing Decoding. Preprint 117, DFG-SPP 1324, February 2012.

[118] D. Heinen and G. Plonka. Wavelet shrinkage on paths for scattered data denoising. Preprint 118, DFG-SPP 1324, February 2012.

[119] T. Jahnke and M. Kreim. Error bound for piecewise deterministic processes modeling stochastic reaction systems. Preprint 119, DFG-SPP 1324, March 2012.

[120] C. Bender and J. Steiner. A-posteriori estimates for backward SDEs. Preprint 120, DFG-SPP 1324, April 2012.