A-posteriori estimates for backward SDEs

C. Bender, J. Steiner

Preprint 120
A-posteriori estimates for backward SDEs

C. Bender, J. Steiner

Preprint 120
The consecutive numbering of the publications is determined by their chronological order.

The aim of this preprint series is to make new research rapidly available for scientific discussion. Therefore, the responsibility for the contents is solely due to the authors. The publications will be distributed by the authors.
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
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 $\pi = \{t_0, \ldots, t_N\}$ which satisfies $0 = t_0 < t_1 < \cdots < t_N = T$. We assume that $(\Omega, \mathcal{F}, (G_t)_{t \in [0,T]}, P)$ is a filtered probability space which carries a $D$-dimensional Brownian mo-
tion $W_t = (W_t^1, \ldots, W_t^D)^\star$, (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 $\pi$ 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

$$
Y_{t_{i+1}} = Y_{t_i} - f^\pi(t_i, Y_{t_i}, \Delta_i^{-1}E_i[\Delta W_i^j M_{t_{i+1}}^j])\Delta_i - (M_{t_{i+1}}^\pi - M_{t_i}^\pi), \quad i = 0, \ldots, N - 1,
$$

$$
Y_{t_0}^\pi = \xi^\pi.
$$

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

$(H_0^\pi)$ $\xi^\pi$ is a square-integrable $\mathcal{G}_T$-measurable random variable. $f^\pi : \pi \times \Omega \times \mathbb{R} \times \mathbb{R}^D \rightarrow \mathbb{R}$ is measurable and, for every $t_i \in \pi$ and $(y, z) \in \mathbb{R}^{D+1}$, $f^\pi(t_i, y, z)$ is $\mathcal{G}_t$-measurable. Moreover $f^\pi$ is Lipschitz in $(y, z)$ with constant $K$ (independent of $\pi$) uniformly in $(t_i, \omega)$ 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 $(Y^\pi, M^\pi)$ are connected via the property

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

Under $(H_0^\pi)$ it is straightforward to check by a contraction mapping argument that this discrete BSDE admits a unique solution if the mesh $|\pi| = \max\{|t_{i+1} - t_i| : i = 0, \ldots, N - 1\}$ of $\pi$ is sufficiently small. In fact, denoting

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

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

$$
Y_{t_i} = \xi^\pi, \quad Z_{t_N}^\pi = 0, \\
Z_{t_i}^\pi = E_i[\Delta W_i^j Y_{t_{i+1}}^\pi], \quad i = N - 1, \ldots, 0, \\
Y_{t_i}^\pi = E_i[Y_{t_{i+1}}^\pi] - f^\pi(t_i, Y_{t_i}, Z_{t_i})\Delta_i, \quad i = N - 1, \ldots, 0.
$$

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 $(Y^\pi, M^\pi)$ is at hand. We merely assume that $(\hat{Y}^\pi, \hat{M}^\pi)$ is square-integrable, adapted to $\mathcal{G}_t$, and that $M^\pi$ 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 $(Y^\pi, M^\pi)$, which only requires knowledge of the approximation and the data $(f^\pi, \xi^\pi)$. 

2 A-posteriori estimates for discrete BSDEs
Theorem 2.1. Assume \( H_u^\pi \). Then there are constants \( c, C > 0 \) such that for sufficiently small \( |\pi| \) and for every pair of square-integrable \( G_t \)-adapted processes \((\hat{Y}_t^\pi, \hat{M}_t^\pi)\), where \( \hat{M}_t^\pi \) is a martingale starting in \( 0 \), it holds that

\[
\frac{1}{c} E_\pi(\hat{Y}_T^\pi, \hat{M}_T^\pi) \leq \max_{t \in \Phi} E_0 \left[ |Y_t^\pi - \hat{Y}_t^\pi|^2 \right] + E_0 \left[ |M_t^\pi - \hat{M}_t^\pi|^2 \right] \leq C E_\pi(\hat{Y}_T^\pi, \hat{M}_T^\pi)
\]

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 \( |\pi| \leq \Gamma^{-1} \).

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 \((X_t^\pi, F_t^\pi)\) taking values in \( R^M \) and deterministic functions \( y^\pi, z^\pi \) such that the solution \((Y^\pi, Z^\pi)\) of \([4]\) is of the form

\[
Y_t^\pi = y^\pi(t, X_t^\pi), \quad Z_t^\pi = z^\pi(t, X_t^\pi).
\]

We recall here that \((F_t^\pi)_{t \in [0, T]}\) is the augmented Brownian filtration, but it could also be replaced by a larger filtration \((H_t^\pi)_{t \in [0, T]}\) 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

\[
\hat{y}^\pi(t, x; \Xi), \quad \hat{z}^\pi(t, x; \Xi),
\]

where \( \Xi \) is a random vector independent of \( F_T \). 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 $X^\pi$ which we can collect in $\Xi$. We now define the filtration $\mathcal{G}_t$ by enlarging the Brownian filtration with the random vector $\Xi$, i.e. $\mathcal{G}_t = \mathcal{F}_t \vee \sigma(\Xi)$. Given $(\hat{y}^\pi, \hat{z}^\pi)$, we consider

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

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

\[
\hat{M}_0^\pi = 0, \quad \hat{M}_{t_{i+1}} - \hat{M}_t^\pi = \hat{z}^\pi(t_i, X^\pi_{t_i}; \Xi) \Delta W_{t_i},
\]

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

\[
\hat{Z}_{t_i}^\pi = \Delta^{-1} E_t[\Delta W_{t_i} \hat{M}_{t_{i+1}}^\pi],
\]

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

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

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

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

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

\[
\hat{\mathcal{E}}_\pi(\hat{Y}^\pi, \hat{Z}^\pi) = \frac{1}{\Lambda} \sum_{\lambda=1}^{\Lambda} \left[ \lambda \xi^\pi - \lambda \hat{Y}_{t_N}^\pi \right]^2 + \max_{i=1,\ldots,N} \frac{1}{\Lambda} \sum_{\lambda=1}^{\Lambda} \left[ \lambda \hat{Y}_{t_i}^\pi - \lambda \hat{Y}_{t_0}^\pi - \sum_{j=0}^{i-1} f^\pi(t_j, \lambda \hat{Y}_{t_j}^\pi, \lambda \hat{Z}_{t_j}^\pi) \Delta j - \sum_{j=0}^{i-1} \lambda \hat{Z}_{t_j}^\pi \Delta \lambda W_j \right]^2.
\]

In such situation we can, thanks to Theorem \ref{thm:2}, 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 $(Y^\pi, M^\pi)$ of the discrete BSDE $[\hat{7}]$.

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

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

and

\[
E \left[ \hat{y}^\pi(t_{i+1}, X^\pi_{t_{i+1}}; \Xi) \bigg| \Xi, X^\pi_{t_i} = x \right]
\]
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 \((X^\pi_t, \mathcal{F}_t)\) we can then define a \(\mathcal{G}_t\)-martingale \(\hat{M}^\pi_t\) via
\[
\hat{M}^\pi_0 = 0, \quad \hat{M}^\pi_{t+1} - \hat{M}^\pi_t = \hat{g}^\pi(t+1, X^\pi_{t+1}; \Xi) - \mathbb{E}\left[\hat{g}^\pi(t+1, X^\pi_{t+1}; \Xi) \mid X^\pi_t\right].
\]
Then, again,
\[
\hat{Z}^\pi_i = \hat{z}^\pi(t, X^\pi_t; \Xi) = \Delta^{-1}_i E_i[\Delta W^\pi_i \hat{M}^\pi_{t+1}],
\]
and the error criterion becomes
\[
\mathcal{E}_\pi(\hat{Y}^\pi, \hat{M}^\pi) = E_0[|\xi^\pi - \hat{Y}^\pi_N|] + \max_{1 \leq i \leq N} \mathbb{E}_0\left[\hat{Y}^\pi_i - \hat{Y}^\pi_0 - \sum_{j=0}^{i-1} f^\pi(t_j, \hat{Y}^\pi_j, \hat{Z}^\pi_j) \Delta_j - \hat{M}^\pi_i \right]^2.
\]

Given a sampling procedure for \((X^\pi_t)_{i=0,\ldots,N}\) we can now, conditional on \(\Xi\), sample independent copies of \((\hat{Y}^\pi_i, \hat{Z}^\pi_i, \hat{M}^\pi_i)_{i=0,\ldots,N}\). Assuming again that the data is good enough and so \(\Lambda\) independent copies
\[(\lambda \hat{Y}^\pi_i, \lambda \hat{Z}^\pi_i, f^\pi(t, \lambda \hat{Y}^\pi_i, \lambda \hat{Z}^\pi_i), \lambda \hat{M}^\pi_i, \lambda \xi^\pi; \ i = 0, \ldots, N; \lambda = 1, \ldots, \Lambda)\]
of \((\hat{Y}^\pi_i, \hat{Z}^\pi_i, f^\pi(t, \hat{Y}^\pi_i, \hat{Z}^\pi_i), \hat{M}^\pi_i, \xi; \ i = 0, \ldots, N)\) 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 \((Y^\pi, M^\pi)\) 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
\[
\mathcal{E}_\pi(\hat{Y}^\pi, \hat{M}^\pi) \leq c \left( \max_{t \in \pi} \mathbb{E}_0 \left[ |Y^\pi_t - \hat{Y}^\pi_t|^2 \right] + \mathbb{E}_0 \left[ |M^\pi_t - \hat{M}^\pi_t|^2 \right] \right).
\]
In order to simplify the notation we set
\[
\hat{Z}^\pi_i = \Delta^{-1}_i E_i[\Delta W^\pi_i \hat{M}^\pi_{t+1}].
\]
Then,
\[
\mathcal{E}_\pi(\hat{Y}^\pi, \hat{M}^\pi) = E_0[|\xi^\pi - \hat{Y}^\pi_N|] + \max_{1 \leq i \leq N} \mathbb{E}_0\left[\hat{Y}^\pi_i - \hat{Y}^\pi_0 - \sum_{j=0}^{i-1} f^\pi(t_j, \hat{Y}^\pi_j, \hat{Z}^\pi_j) \Delta_j - \hat{M}^\pi_i \right]^2 =: \langle A \rangle + \max_{i=1,\ldots,N} \langle B_i \rangle.
\]
By (1) and (3) we get
\[
Y^\pi_t - Y^\pi_0 - \sum_{j=0}^{i-1} f^\pi(t_j, Y^\pi_j, Z^\pi_j) \Delta_j - M^\pi_t = 0.
\]
Hence, by Young’s inequality, the Lipschitz condition on $f^\pi$ and the martingale property of $M^\pi - \hat{M}^\pi$, we have for every $\gamma > 0$,

\[
(B_i) = E_0 \left[ \hat{Y}_{t_i}^\pi - Y_{t_0}^\pi + Y_{t_0}^\pi - \sum_{j=0}^{i-1} \left[ f^\pi(t_j, \hat{Y}_{t_j}^\pi, \hat{Z}_{t_j}^\pi) - f^\pi(t_j, Y_{t_j}^\pi, Z_{t_j}^\pi) \right] \Delta_j - \hat{M}_{t_i}^\pi + M_{t_i}^\pi \right]^2 
\leq \frac{24}{5} (1 + \gamma) \max_{t_j \in \pi} E_0 \left[ |Y_{t_j}^\pi - \hat{Y}_{t_j}^\pi|^2 \right] + 6(1 + \gamma) E_0 \left[ |M_{t_i}^\pi - \hat{M}_{t_i}^\pi|^2 \right] 
+ (1 + \gamma^{-1}) T \sum_{j=0}^{N-1} \Delta_j E_0 \left[ |f^\pi(t_j, \hat{Y}_{t_j}^\pi, \hat{Z}_{t_j}^\pi) - f^\pi(t_j, Y_{t_j}^\pi, Z_{t_j}^\pi)|^2 \right] 
\leq 5(1 + \gamma) \max_{t_j \in \pi} E_0 \left[ |Y_{t_j}^\pi - \hat{Y}_{t_j}^\pi|^2 \right] + 6(1 + \gamma) E_0 \left[ |M_{t_N}^\pi - \hat{M}_{t_N}^\pi|^2 \right] 
+ (1 + \gamma^{-1}) T (1 + T) K^2 \left( \max_{t_j \in \pi} E_0 \left[ |Y_{t_j}^\pi - \hat{Y}_{t_j}^\pi|^2 \right] + \sum_{j=1}^{N-1} E_0 \left[ |Z_{t_j}^\pi - \hat{Z}_{t_j}^\pi|^2 \Delta_j \right] \right).
\]

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

\[
\sum_{j=1}^{N-1} E_0 \left[ |Z_{t_j}^\pi - \hat{Z}_{t_j}^\pi|^2 \right] \Delta_j = \sum_{j=1}^{N-1} \frac{1}{\Delta_j} E_0 \left[ E_j \left[ \Delta W^\pi_j (M_{t_{j+1}}^\pi - \hat{M}_{t_{j+1}}^\pi - M_{t_j}^\pi + \hat{M}_{t_j}^\pi) \right] \right] 
\leq D \sum_{j=1}^{N-1} \left( E_0 \left[ |M_{t_{j+1}}^\pi - \hat{M}_{t_{j+1}}^\pi|^2 \right] - E_0 \left[ |M_{t_j}^\pi - \hat{M}_{t_j}^\pi|^2 \right] \right) = DE_0 \left[ |M_{t_N}^\pi - \hat{M}_{t_N}^\pi|^2 \right],
\]

we immediately get

\[
E_\pi(\hat{Y}^\pi, \hat{M}^\pi) \leq \left( 6(1 + \gamma) + DT(1 + T) K^2 (1 + \gamma^{-1}) \right) \left( \max_{t_j \in \pi} E_0 \left[ |Y_{t_j}^\pi - \hat{Y}_{t_j}^\pi|^2 \right] + E_0 \left[ |M_{t_N}^\pi - \hat{M}_{t_N}^\pi|^2 \right] \right).
\]

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

\[
\bar{Y}_{t_0}^\pi = \hat{Y}_{t_0}^\pi, \quad \bar{Y}_{t_i}^\pi = \hat{Y}_{t_i}^\pi + f^\pi(t_i, \hat{Y}_{t_i}^\pi, \hat{Z}_{t_i}^\pi) \Delta_i + \hat{M}_{t_i}^\pi - \hat{M}_{t_i}^\pi, \quad i = 0, \ldots, N - 1,
\]

where again $\hat{Z}_{t_i}^\pi = \Delta_i^{-1} E_i[\Delta W_{t_i}^\pi \hat{M}_{t_{i+1}}^\pi]$. Then the pair $(\bar{Y}^\pi, \hat{M}^\pi)$ solves the discrete BSDE with terminal condition $\hat{\xi}^\pi = Y_{t_N}^\pi$ and driver $\hat{f}^\pi(t_i, y, z) = f^\pi(t_i, \hat{Y}_{t_i}^\pi, \hat{Z}_{t_i}^\pi)$. We will first estimate the error between $(\bar{Y}^\pi, \hat{M}^\pi)$ and $(Y^\pi, M^\pi)$ 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

\[
M_{t_{i+1}}^\pi - \hat{M}_{t_{i+1}}^\pi = Y_{t_{i+1}}^\pi - E_i[Y_{t_{i+1}}^\pi], \quad \hat{M}_{t_{i+1}}^\pi - \hat{M}_{t_{i}}^\pi = \bar{Y}_{t_{i+1}}^\pi - E_i[\bar{Y}_{t_{i+1}}^\pi].
\]

Hence,

\[
\sum_{i=0}^{N-1} q_i E_0 \left[ \left( |M_{t_{i+1}}^\pi - \hat{M}_{t_{i+1}}^\pi| - (\hat{M}_{t_{i+1}}^\pi - \hat{M}_{t_{i}}^\pi) \right)^2 \right] = \sum_{i=0}^{N-1} q_i E_0 \left[ |Y_{t_{i+1}}^\pi - \bar{Y}_{t_{i+1}}^\pi - E_i[Y_{t_{i+1}}^\pi - \bar{Y}_{t_{i+1}}^\pi]|^2 \right].
\]
Now, following the argument in step 1 of the proof of Lemma 7 in Bender and Denk (2007), we obtain,

\[
\sum_{i=1}^{N-1} q_i E_0 \left[ \left( (M_{i+1}^{\pi} - M_i^{\pi}) - (\hat{M}_{i+1}^{\pi} - \hat{M}_i^{\pi}) \right)^2 \right] \\
\leq q_N E_0 \left[ |Y_{t_N}^{\pi} - \hat{Y}_{t_N}^{\pi}|^2 \right] + \gamma \sum_{i=0}^{N-1} q_i \Delta_i E_0 \left[ |Y_{t_i}^{\pi} - \hat{Y}_{t_i}^{\pi}|^2 \right] \\
+ \frac{(1 + T)K^2}{\gamma T} \sum_{i=0}^{N-1} q_i \Delta_i E_0 \left[ Y_{t_i}^{\pi} - \hat{Y}_{t_i}^{\pi} \right]^2 + \frac{(1 + T)K^2}{\gamma T} \sum_{i=0}^{N-1} q_i \Delta_i E_0 \left[ Z_{t_i}^{\pi} - \hat{Z}_{t_i}^{\pi} \right]^2.
\]

The argument of step 2 of the same proof yields

\[
\max_{0 \leq i \leq N} E_0 \left[ |Y_{t_i}^{\pi} - \hat{Y}_{t_i}^{\pi}|^2 \right] \leq q_N E_0 \left[ |Y_{t_N}^{\pi} - \hat{Y}_{t_N}^{\pi}|^2 \right] \\
+ \frac{K^2(T + 1)(|\pi| + \Gamma^{-1})}{N} \left( \frac{1}{T} \sum_{i=0}^{N-1} q_i \Delta_i E_0 \left[ |Y_{t_i}^{\pi} - \hat{Y}_{t_i}^{\pi}|^2 \right] + \frac{K^2}{\gamma T} \sum_{i=0}^{N-1} q_i \Delta_i E_0 \left[ |Z_{t_i}^{\pi} - \hat{Z}_{t_i}^{\pi}|^2 \right] \right).
\]

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

\[
\max_{0 \leq i \leq N} E_0 \left[ |Y_{t_i}^{\pi} - \hat{Y}_{t_i}^{\pi}|^2 \right] + \sum_{i=1}^{N-1} q_i E_0 \left[ \left( (M_{i+1}^{\pi} - M_i^{\pi}) - (\hat{M}_{i+1}^{\pi} - \hat{M}_i^{\pi}) \right)^2 \right] \\
\leq (2 + \gamma T)q_N E_0 \left[ |Y_{t_N}^{\pi} - \hat{Y}_{t_N}^{\pi}|^2 \right] + \left[ (1 + \gamma T)K^2(T + 1)(|\pi| + \Gamma^{-1}) + \frac{(1 + T)K^2}{\gamma} \right] \\
\times \left( \max_{0 \leq i \leq N} q_i E_0 \left[ |Y_{t_i}^{\pi} - \hat{Y}_{t_i}^{\pi}|^2 \right] + \frac{D}{\gamma T} \sum_{i=1}^{N-1} q_i E_0 \left[ \left( (M_{i+1}^{\pi} - M_i^{\pi}) - (\hat{M}_{i+1}^{\pi} - \hat{M}_i^{\pi}) \right)^2 \right] \right) \\
\leq (2 + \gamma T)q_N E_0 \left[ |Y_{t_N}^{\pi} - \hat{Y}_{t_N}^{\pi}|^2 \right] + \left[ (1 + \gamma T)K^2(T + 1)(|\pi| + \Gamma^{-1}) + \frac{(1 + T)K^2}{\gamma} \right] \left( 1 + \frac{1}{D \vee 2} \right) \\
\times \left( \max_{0 \leq i \leq N} q_i E_0 \left[ |Y_{t_i}^{\pi} - \hat{Y}_{t_i}^{\pi}|^2 \right] + \sum_{i=1}^{N-1} q_i E_0 \left[ \left( (M_{i+1}^{\pi} - M_i^{\pi}) - (\hat{M}_{i+1}^{\pi} - \hat{M}_i^{\pi}) \right)^2 \right] \right) \\
+ \left[ (1 + \gamma T)K^2(T + 1)(|\pi| + \Gamma^{-1}) + \frac{(1 + T)K^2}{\gamma} \right] \left( 1 + \frac{1}{D \vee 2} \right) \left( 1 + \frac{1}{D \vee 2} \right) \\
\times \left( \max_{0 \leq i \leq N} q_i E_0 \left[ |\hat{Y}_{t_i}^{\pi} - \hat{Y}_{t_i}^{\pi}|^2 \right] \right).
\]

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 \( |\pi| \leq \Gamma^{-1} \)

\[
\max_{0 \leq i \leq N} E_0 \left[ |Y_{t_i}^{\pi} - \hat{Y}_{t_i}^{\pi}|^2 \right] + \sum_{i=1}^{N-1} q_i E_0 \left[ \left( (M_{i+1}^{\pi} - M_i^{\pi}) - (\hat{M}_{i+1}^{\pi} - \hat{M}_i^{\pi}) \right)^2 \right] \\
\leq 4(2 + \gamma T)q_N E_0 \left[ |Y_{t_N}^{\pi} - \hat{Y}_{t_N}^{\pi}|^2 \right] + \frac{3}{D \vee 2} \left( 1 + \frac{1}{D \vee 2} \right) \max_{0 \leq i \leq N} q_i E_0 \left[ |Y_{t_i}^{\pi} - \hat{Y}_{t_i}^{\pi}|^2 \right].
\]
3 A-posteriori estimates for continuous BSDEs

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

$$\max_{0 \leq t \leq N} E_0 \left[ \left| Y_{t_i}^\pi - \hat{Y}_{t_i}^\pi \right|^2 \right] + E_0 \left[ \left| (M_{t_i}^\pi - \hat{M}_{t_i}^\pi)^2 \right| \right]$$

$$\leq 8(2 + \gamma T) e^{GT} E_0 \left[ \left| Y_{t_N}^\pi - \hat{Y}_{t_N}^\pi \right|^2 \right] + \left( \frac{6}{D \vee 2} \left( 1 + \frac{1}{D \vee 2 - 1} \right) \right) e^{GT} + 2 \max_{0 \leq t \leq N} E_0 \left[ \left| \hat{Y}_{t_i}^\pi - \bar{Y}_{t_i}^\pi \right|^2 \right]$$

$$\leq 16(2 + \gamma T) e^{GT} E_0 \left[ \left| \hat{X} - \bar{X} \right|^2 \right]$$

$$+ \left( \frac{6}{D \vee 2} \left( 1 + \frac{1}{D \vee 2 - 1} \right) + 2 + 16(2 + \gamma T) \right) e^{GT} \max_{0 \leq t \leq N} E_0 \left[ \left| \hat{Y}_{t_i}^\pi - \bar{Y}_{t_i}^\pi \right|^2 \right]$$

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

$$\hat{Y}_{t_i}^\pi - \bar{Y}_{t_i}^\pi = \hat{Y}_{t_0}^\pi - \bar{Y}_{t_0}^\pi - \sum_{j=0}^{i-1} f^\pi(t_j, \hat{Y}_{t_j}^\pi, \Delta_j^{-1} E_j [\Delta W_j^\pi \hat{M}_j^\pi]) \Delta_j - \bar{M}_i^\pi.$$  \hfill (9)

\[\square\]

3 A-posteriori estimates for continuous BSDEs

We now turn to BSDEs in continuous time

$$Y_t = \xi - \int_t^T f(s, Y_s, Z_s) ds - Z_s dW_s$$  \hfill (10)

and assume

(H) $\xi$ 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 $(t, \omega)$ and $E[\int_0^T |f(t, 0, 0)|^2 dt] < \infty$.

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 $(\bar{Y}, \bar{Z})$ such that (10) is satisfied.

We now suppose that a square-integrable approximative solution $(\hat{Y}_{t_i}^\pi, \hat{Z}_{t_i}^\pi)_{t_i \in \pi}$ of $(Y, Z)$ has been computed on some time grid $\pi$ by some numerical algorithm. Again we allow that $(\hat{Y}_{t_i}^\pi, \hat{Z}_{t_i}^\pi)$ is adapted to a larger filtration $\mathcal{G}_t$ and wish to quantify the error between true solution and approximative solution. As, in general, it may not be possible to draw sample copies of $\xi$ and $f(t_i, \hat{Y}_{t_i}^\pi, \hat{Z}_{t_i}^\pi)$, we apply the error criterion (5) with approximative data $(\hat{\xi}, \hat{f})$. 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,

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

9
(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 \( \Xi \) denotes a family of random variables independent of \( \mathcal{F}_T \) (cf. the discussion in Example 2.2. Concerning the approximative data, we suppose \((H^T_H)\), i.e. assumption \((H^\Xi_H)\) as introduced in the previous section, but with the larger filtration \( \mathcal{G}_t \) replaced by the augmented Brownian filtration \( (\mathcal{F}_t) \) 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 \((H^T_H)\) and that \( \mathcal{G}_t = \mathcal{F}_t \vee \sigma(\Xi) \), where \( \Xi \) is independent of \( \mathcal{F}_T \). Then, there are constants \( c, C > 0 \) such that for every pair of \( \mathcal{G}_t \)-adapted square integrable processes \( (\hat{Y}^\pi_{t_i}, \hat{Z}^\pi_{t_i})_{t_i \in \pi} \)

\[
\max_{t_i \in \pi} E_0 \left[ |Y_{t_i} - \hat{Y}^\pi_{t_i}|^2 \right] + \sum_{i=0}^{N-1} E_0 \left[ \int_{t_i}^{t_{i+1}} \left( |Y_t - \hat{Y}^\pi_t|^2 + |Z_t - \hat{Z}^\pi_t|^2 \right) dt \right] \\
\leq C \left( \mathcal{E}_\pi(\hat{Y}^\pi, \hat{Z}^\pi) + |\pi| + E \left[ |\xi - \xi^\pi|^2 \right] + \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} E \left[ |f(t, Y_t, Z_t) - f^\pi(t_i, Y_t, Z_t)|^2 \right] dt \right).
\]

and

\[
\mathcal{E}_\pi(\hat{Y}^\pi, \hat{Z}^\pi) \leq c \left( \max_{t_i \in \pi} E_0 \left[ |Y_{t_i} - \hat{Y}^\pi_{t_i}|^2 \right] + \sum_{i=0}^{N-1} E_0 \left[ \int_{t_i}^{t_{i+1}} \left( |Y_t - \hat{Y}^\pi_t|^2 + |Z_t - \hat{Z}^\pi_t|^2 \right) dt \right] \\
+ E \left[ |\xi - \xi^\pi|^2 \right] + \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} E \left[ |f(t, Y_t, Z_t) - f^\pi(t_i, Y_t, Z_t)|^2 \right] dt \right)
\]

If, additionally, \( f \) and \( f^\pi \) do not depend on \( y \), then

\[
\max_{t_i \in \pi} E_0 \left[ |Y_{t_i} - \hat{Y}^\pi_{t_i}|^2 \right] + \sum_{i=0}^{N-1} E_0 \left[ \int_{t_i}^{t_{i+1}} |Z_t - \hat{Z}^\pi_t|^2 dt \right] \\
\leq C \left( \mathcal{E}_\pi(\hat{Y}^\pi, \hat{Z}^\pi) + E \left[ |\xi - \xi^\pi|^2 \right] + \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} E \left[ |f(t, Z_t) - f^\pi(t_i, Z_t)|^2 \right] dt \right)
\]

and

\[
\mathcal{E}_\pi(\hat{Y}^\pi, \hat{Z}^\pi) \leq c \left( \max_{t_i \in \pi} E_0 \left[ |Y_{t_i} - \hat{Y}^\pi_{t_i}|^2 \right] + \sum_{i=0}^{N-1} E_0 \left[ \int_{t_i}^{t_{i+1}} |Z_t - \hat{Z}^\pi_t|^2 dt \right] \\
+ E \left[ |\xi - \xi^\pi|^2 \right] + \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} E \left[ |f(t, Z_t) - f^\pi(t_i, Z_t)|^2 \right] dt \right)
\]

**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 \( E \left[ \int_0^T |f(t, 0, 0)|^2 dt \right] \) and \( E \left[ |\xi|^2 \right] \).
Lemma 3.3. Suppose (H) and (Hδ) and that \( G_t = \mathcal{F}_t \vee \sigma(\Xi) \), where \( \Xi \) is independent of \( \mathcal{F}_T \). Then, there is a constant \( C > 0 \) such that for every pair of \( G_{t_i} \)-adapted square integrable processes \((\hat{Y}_{t_i}^\pi, \hat{Z}_{t_i}^\pi)_{t_i \in \pi}\)

\[
\max_{0 \leq i \leq N} \mathbb{E} \left[ |Y_{t_i} - Y_{t_i}^\pi|^2 \right] + \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \mathbb{E} \left[ |Z_t - Z_{t_i}^\pi|^2 \right] dt 
\leq C \left( \mathbb{E} \left[ Y_{t,N}^\pi \right] + K_y |\pi| + \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \mathbb{E} \left[ \left| f(t, Y_t, Z_t) - f^\pi(t, Y_{t_i}^\pi, Z_{t_i}^\pi) \right|^2 \right] dt \right).
\]

Here, \( K_y \leq K \) denotes a Lipschitz constant of \( f^\pi \) with respect to the \( y \)-variable.

Proof. First notice that by assumption (Hδ) and the assumption on the filtration \( G_t \), it holds

\[
\begin{align*}
Y_{t_N}^\pi &= \xi^\pi, \quad Z_{t_N}^\pi = 0, \\
Z_{t_i}^\pi &= E[\Delta W_{t_i} Y_{t_{i+1}}^\pi | \mathcal{F}_{t_i}], \quad i = N - 1, \ldots, 0, \\
Y_{t_i}^\pi &= E[Y_{t_{i+1}}^\pi | \mathcal{F}_{t_i}] - f^\pi(t_i, Y_{t_i}^\pi, Z_{t_i}^\pi) \Delta_i, \quad i = N - 1, \ldots, 0,
\end{align*}
\]

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

\[
\max_{0 \leq i \leq N} \mathbb{E} \left[ |Y_{t_i} - Y_{t_i}^\pi|^2 \right] + \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \mathbb{E} \left[ |Z_t - Z_{t_i}^\pi|^2 \right] dt 
\leq C \left( \mathbb{E} \left[ |\xi - \xi^\pi|^2 \right] + K_y |\pi| + \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \mathbb{E} \left[ \left| Z_s^\pi dW_s - Z_{t_i}^\pi \Delta W_i \right|^2 \right] \right) 
+ \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \mathbb{E} \left[ \left| f(t, Y_t, Z_t) - f^\pi(t_i, Y_{t_i}^\pi, Z_{t_i}^\pi) \right|^2 \right] dt, \quad (13)
\]

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

\[
\int_{t_i}^{t_{i+1}} \hat{Z}^\pi dW_i = Y_{t_{i+1}}^\pi - E \left[ Y_{t_{i+1}}^\pi | \mathcal{F}_{t_i} \right]. \quad (14)
\]

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
\[
\sum_{i=0}^{N-1} \mathbb{E} \left[ \left( \int_{t_i}^{t_{i+1}} \tilde{Z}_s^\pi dW_s - Z_{t_i}^\pi \Delta W_i \right)^2 \right] \leq C \mathbb{E}_\pi (\tilde{Y}^\pi, \tilde{Z}^\pi),
\]
which, in view of (13), completes the proof of this lemma.

By (12) and the independence of \(G_0\) and \(\mathcal{F}_T\), Young’s inequality and Itô’s isometry, we obtain
\[
\sum_{i=0}^{N-1} \mathbb{E} \left[ \left( \int_{t_i}^{t_{i+1}} \tilde{Z}_s^\pi dW_s - Z_{t_i}^\pi \Delta W_i \right)^2 \right] = \sum_{i=0}^{N-1} \mathbb{E}_0 \left[ \left( \int_{t_i}^{t_{i+1}} \tilde{Z}_s^\pi dW_s - Z_{t_i}^\pi \Delta W_i \right)^2 \right] \leq 2 \sum_{i=0}^{N-1} \mathbb{E}_0 \left[ \left( \int_{t_i}^{t_{i+1}} \tilde{Z}_s^\pi dW_s - \tilde{Z}_{t_i}^\pi \Delta W_i \right)^2 \right] + 2 \sum_{i=0}^{N-1} \mathbb{E}_0 \left[ |\tilde{Z}_{t_i}^\pi - Z_{t_i}^\pi|^2 \right] \Delta_i.
\]

As in the proof of Theorem 2.1, we introduce the process \(\tilde{Y}_{t_i}^\pi\) by
\[
\tilde{Y}_{t_0} = \tilde{Y}_{t_0}^\pi, \quad \tilde{Y}_{t_{i+1}}^\pi = \tilde{Y}_{t_i}^\pi + f^\pi(t_i, \tilde{Y}_{t_i}^\pi, \tilde{Z}_{t_i}^\pi) \Delta_i + \tilde{Z}_{t_i}^\pi \Delta W_i, \quad i = 0, \ldots, N - 1,
\]
with the specific choice \(\tilde{M}_{t_{i+1}}^\pi - \tilde{M}_{t_i}^\pi = \tilde{Z}_{t_i}^\pi \Delta W_i\) for the martingale increment. Notice that with this choice of the martingale \(\tilde{M}^\pi\), \(\mathcal{E}_\pi(\tilde{Y}^\pi, \tilde{Z}^\pi)\) equals \(\mathcal{E}_\pi(\tilde{Y}^\pi, \tilde{M}^\pi)\) in the notation of Theorem 2.1. Then, by (14) and (4), and applying Young’s inequality and the Lipschitz property of \(f^\pi\), we get
\[
\mathbb{E}_0 \left[ \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \tilde{Z}_s^\pi dW_s - \tilde{Z}_{t_i}^\pi \Delta W_i \right]^2 = \mathbb{E}_0 \left[ \sum_{i=0}^{N-1} \tilde{Y}_{t_{i+1}}^\pi - \tilde{Y}_{t_i}^\pi - (\tilde{Y}_{t_i}^\pi - \tilde{Y}_{t_i}^\pi) \left( f^\pi(t_i, \tilde{Y}_{t_i}^\pi, Z_{t_i}^\pi) - f^\pi(t_i, \tilde{Y}_{t_i}^\pi, \tilde{Z}_{t_i}^\pi) \right) \Delta_i \right]^2 \leq C \left( \mathbb{E}_0 \left[ |\tilde{Y}_{t_N}^\pi - \tilde{Y}_{t_N}^\pi|^2 \right] + \max_{t_i \in \pi} \mathbb{E}_0 \left[ |Y_{t_i}^\pi - \tilde{Y}_{t_i}^\pi|^2 \right] + \sum_{i=0}^{N-1} \mathbb{E}_0 \left[ |Z_{t_i}^\pi - \tilde{Z}_{t_i}^\pi|^2 \right] \Delta_i \right).
\]

Gathering the above inequalities and applying (8) we have
\[
\sum_{i=0}^{N-1} \mathbb{E} \left[ \left( \int_{t_i}^{t_{i+1}} \tilde{Z}_s^\pi dW_s - Z_{t_i}^\pi \Delta W_i \right)^2 \right] \leq C \left( \mathbb{E}_0 \left[ |\tilde{Y}_{t_N}^\pi - \tilde{Y}_{t_N}^\pi|^2 \right] + \max_{t_i \in \pi} \mathbb{E}_0 \left[ |Y_{t_i}^\pi - \tilde{Y}_{t_i}^\pi|^2 \right] + \mathbb{E}_0 \left[ |M_{t_N}^\pi - \tilde{M}_{t_N}^\pi|^2 \right] \right).
\]

Inequality (15) now follows from Theorem 2.1 and (9) with the above choice of the martingale \(\tilde{M}^\pi\). \(\square\)
3 A-posteriori estimates for continuous BSDEs

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.

\[
\hat{M}^\pi_0 = 0, \quad \hat{M}^\pi_{t_{i+1}} - \hat{M}^\pi_{t_i} = \hat{Z}^\pi_{t_i} \Delta W_i,
\]

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

\[
\max_{t_i \in \pi} E_0 \left[ |Y_{t_i} - \hat{Y}_{t_i}^\pi|^2 \right] + \sum_{i=0}^{N-1} E_0 \left[ \int_{t_i}^{t_{i+1}} |Z_t - \hat{Z}_{t_i}^\pi|^2 dt \right] \\
\leq 2 \left( \max_{t_i \in \pi} E \left[ |Y_{t_i} - \hat{Y}_{t_i}^\pi|^2 \right] + \sum_{i=0}^{N-1} E \left[ \int_{t_i}^{t_{i+1}} |Z_t - Z_{t_i}^\pi|^2 dt \right] \right) \\
+ 2 \left( \max_{t_i \in \pi} E_0 \left[ |Y_{t_i}^\pi - \hat{Y}^\pi_{t_i}|^2 \right] + DE_0 \left[ |M^\pi_{t_N} - \hat{M}^\pi_{t_N}|^2 \right] \right)
\]

Applying Theorem 2.1 and Lemma 3.3 yields, for some constant \( C > 0 \),

\[
\max_{t_i \in \pi} E_0 \left[ |Y_{t_i} - \hat{Y}_{t_i}^\pi|^2 \right] + \sum_{i=0}^{N-1} E_0 \left[ \int_{t_i}^{t_{i+1}} |Z_t - \hat{Z}_{t_i}^\pi|^2 dt \right] \\
\leq C \left( \mathcal{E}_\pi(\hat{Y}^\pi, \hat{Z}^\pi) + K_y |\pi| + E \left[ |\xi - \xi^\pi|^2 \right] + \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} E \left[ |f(t, Y_t, Z_t) - f^\pi(t_i, Y_{t_i}, Z_{t_i})|^2 \right] dt \right).
\]

(16)

If \( f \) and \( f^\pi \) 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

\[
\sum_{i=0}^{N-1} E_0 \left[ \int_{t_i}^{t_{i+1}} |Y_t - \hat{Y}_{t_i}^\pi|^2 dt \right] \leq 2T \max_{t_i \in \pi} E_0 \left[ |Y_{t_i} - \hat{Y}_{t_i}^\pi|^2 \right] + 2 \sum_{i=0}^{N-1} E \left[ \int_{t_i}^{t_{i+1}} |Y_t - Y_{t_i}|^2 dt \right]
\]

As

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

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

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

(17)
After inserting (17) we obtain by the Itô isometry, Young’s inequality and Jensen’s inequality
\[
\mathcal{E}_\pi(\hat{Y}^\pi, \hat{Z}^\pi) = \mathbb{E}_0 \left[ |\hat{Y}^\pi_{t_N} - \xi^\pi|^2 \right] + \max_{1 \leq i \leq N} \mathbb{E}_0 \left[ \left( \hat{Y}^\pi_{t_i} - Y_{t_i} \right) + \left( Y_0 - \hat{Y}^\pi_0 \right) \right]
\]
\[
+ \sum_{j=0}^{t_{j+1}} \left( f(t, Y_t, Z_t) - f^\pi(t_j, \hat{Y}^\pi_{t_j}, \hat{Z}^\pi_{t_j}) \right) dt + \sum_{j=0}^{t_{j+1}} \int_{t_j}^{t_{j+1}} \left( Z_t - \hat{Z}^\pi_t \right) dW_t \right]^2
\]
\[
\leq c \left( \max_{0 \leq i \leq N} \mathbb{E}_0 \left[ |\hat{Y}^\pi_{t_i} - Y_{t_i}|^2 \right] + \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \mathbb{E}_0 \left[ |Z_t - \hat{Z}^\pi_t|^2 \right] dt \right)
\]
\[
+ \sum_{j=0}^{N-1} \int_{t_j}^{t_{j+1}} \mathbb{E}_0 \left[ |f^\pi(t_j, Y_t, Z_t) - f^\pi(t_j, \hat{Y}^\pi_{t_j}, \hat{Z}^\pi_{t_j})|^2 \right] dt \right).
\]

(18)

The Lipschitz property of \( f^\pi \) and Young’s inequality yield
\[
\sum_{j=0}^{N-1} \int_{t_j}^{t_{j+1}} \mathbb{E}_0 \left[ |f^\pi(t_j, Y_t, Z_t) - f^\pi(t_j, \hat{Y}^\pi_{t_j}, \hat{Z}^\pi_{t_j})|^2 \right] dt \leq 2K^2_y \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} |Y_t - \hat{Y}^\pi_t|^2 dt + 2K^2 \sum_{i=0}^{N-1} \mathbb{E}_0 \left[ \int_{t_i}^{t_{i+1}} |Z_t - \hat{Z}^\pi_t|^2 dt \right]
\]

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

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 \((\xi, f)\) 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
\[
dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t, \quad X_0 = x_0
\]
with constant initial condition \( x_0 \) and Lipschitz continuous coefficients \( b: [0, T] \times \mathbb{R}^D \to \mathbb{R}^D \) and \( \sigma: [0, T] \times \mathbb{R}^D \to \mathbb{R}^{D \times D} \). We also assume that \( \varphi \) is Lipschitz continuous with constant \( K_\varphi \) and \( F \) is \( \alpha \)-Hölder continuous in time and Lipschitz continuous in space, i.e.
\[
|F(t_1, x_1, y_1, z_1) - F(t_2, x_2, y_2, z_2)| \leq K_1 |t_1 - t_2|^{\alpha} + K_x |x_1 - x_2| + K_1 |y_1 - y_2| + K_1 |z_1 - z_2|,
\]
for some \( \alpha \geq 1/2 \). Given a strong order \( \alpha \) approximation \( X^\pi \) of \( X \) on a grid \( \pi \), i.e.
\[
\max_{t_i \in \pi} \mathbb{E} [ |X_{t_i} - X^\pi_{t_i}|^2 ] \leq C |\pi|^{2\alpha},
\]

14
we define the approximate data \((\xi^n, f^n)\) by

\[
\xi^n = \varphi(X^n_T), \quad f^n(t_i, y, z) = F(t_i, X^n_i, y, z).
\]

Then,

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

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

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

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

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

and

\[
\mathcal{E}_\pi(\hat{Y}^\pi, \hat{Z}^\pi) \leq c \left( \max_{i=0, \ldots, N-1} \sup_{t \in [t_i, t_{i+1}]} \left( E \left[ |Y_{t_i} - \hat{Y}_{t_i}|^2 \right] + \sum_{i=0}^{N-1} E \left[ \int_{t_i}^{t_{i+1}} |Z_t - \hat{Z}_{t_i}|^2 dt \right] \right) + (K_\varphi + K_t) |\pi|^{2\alpha} + K_x |\pi| \right).
\]

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 \(\alpha\) in the lower bound \((19)\), if \(F\) is sufficiently regular in time and a higher order approximation \(X^\pi\) of \(X\) is applied. It vanishes completely, when additionally \(F\) does not depend on \(t\) and one can sample perfectly from \(X^\pi\) on the grid, i.e. one can choose \(X^\pi_{t_i} = X_{t_i}\).

**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 \(\varphi\). For simplicity, we assume that we can sample perfectly from \(X\) on the grid and, hence, choose the true data as approximate data

\[
\xi^n = \varphi(X_T) = \xi, \quad f^n(t_i, y, z) = F(t_i, X_{t_i}, y, z) = f(t_i, y, z).
\]

In this situation, the first inequality in Theorem 3.1 becomes

\[
\max_{t \in \Pi} E_0 \left[ |Y_{t_i} - \hat{Y}_{t_i}|^2 \right] + \sum_{i=0}^{N-1} E_0 \left[ \int_{t_i}^{t_{i+1}} \left( |Y_t - \hat{Y}_{t_i}|^2 + |Z_t - \hat{Z}_{t_i}|^2 \right) dt \right] \leq C \left( \mathcal{E}_\pi(\hat{Y}^\pi, \hat{Z}^\pi) + |\pi| \right).
\]
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 $\pi$ vanish, namely

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

for some Lipschitz continuous function $F$ and a function $\varphi$ (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 $f^\pi$ do not depend on $y$ we can rewrite the third and fourth inequality in Theorem 3.1 as

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

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 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 $\varphi, F, b, \sigma$ are sufficiently smooth and bounded with bounded derivatives. We also assume for the moment that we can sample perfectly from $X$ on the grid $\pi$, 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

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

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

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

Here, we assume, of course, that the assumptions of Theorem 3.1 are in force and that $(\hat{Y}^\pi, \hat{Z}^\pi)$ is $\mathcal{G}_t$-adapted. Compared to (19), the extra term is now of order 1 in the mesh size instead of
4 Numerical examples

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

$$\xi^\pi = \varphi(X^\pi_T), \quad f^\pi(t_i, y, z) = F(t_i, X^\pi_{t_i}, y, z),$$

where $X^\pi$ 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

$$X_{d,t} = x_{d,0} + \int_0^t \sigma\left(\sum_{d'=1}^D \sin(X_{d',u})\right)dW_{d,u}, \quad d = 1, \ldots, D,$$

$$Y_t = \sum_{d=1}^D \sin(X_{d,T}) + \int_t^T \frac{1}{2}\sigma^2[(Y_u)^3]_{D^3} du - \sum_{d=1}^D \int_t^T Z_{d,u} dW_{d,u},$$

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

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

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

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

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 $\pi$. We denote the equidistant partition of $[0, T]$ into $N$ subintervals by $\pi_N$. 17
4 Numerical examples

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

\[
\Xi = (\left(\lambda X_{t_{i+1}}, \Delta \lambda W_i\right) | \lambda = 1, \ldots, \Lambda, \ i = 0, \ldots, \ N - 1)
\]  

(26)

of \(\Lambda\) independent copies of \((X_{t_{i+1}}, \Delta W_i)_{i=0,\ldots,N-1}\). Furthermore, we choose function bases

\[
\eta(i, x) = \{\eta_1(i, x), \ldots, \eta_K(i, x)\}, \ i = 0, \ldots, \ N - 1,
\]

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

\[
\hat{y}_N(t_N, x) = \phi(X_N)
\]

one defines iteratively, for \(i = N - 1\) to 0,

\[
\hat{\alpha}_{d,i} = \arg \min_{\alpha \in \mathbb{R}^K} \sum_{\lambda=1}^{\Lambda} \left| \frac{\Delta \lambda}{\Delta_i} \hat{y}_{i+1}^N - \eta(i, \lambda X_{t_{i+1}}) \alpha \right|^2, \ d = 1, \ldots, D,
\]

\[
\hat{\delta}_d(t_i, x) = \eta(i, x) \hat{\alpha}_{d,i}, \ d = 1, \ldots, D,
\]

\[
\hat{\alpha}_{0,i} = \arg \min_{\alpha \in \mathbb{R}^K} \sum_{\lambda=1}^{\Lambda} \left| \hat{y}_{i+1}^N (\lambda X_{t_{i+1}}) - \Delta_i f^\pi(t_i, \hat{y}_{i+1}^N (\lambda X_{t_{i+1}}), \hat{z}_d^N (\lambda X_{t_{i+1}})) - \eta(i, \lambda X_{t_{i+1}}) \alpha \right|^2,
\]

\[
\hat{y}^N(t_i, x) = \eta(i, x) \hat{\alpha}_{0,i}.
\]

We finally obtain

\[
\hat{Y}_t = \hat{y}^N(t_i, X_{t_{i+1}}), \ \hat{Z}_t = \hat{z}^N(t_i, X_{t_{i+1}})
\]

as approximations for \((Y, Z)\) on the grid \(\pi_N\). 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.

\[
\hat{X}_{d,t} = x_{d,0} + \sigma(\sum_{d'=1}^D \sin(x_{d',0}))W_{d,t}, \ d = 1, \ldots, D,
\]

\[
\hat{Y}_t = \sum_{d=1}^D \sin(\hat{X}_{d,T}) - \sum_{d=1}^D \int_t^T \hat{Z}_{d,u} dW_{d,u}.
\]

18
4 Numerical examples

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

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

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

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

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

$$Y_t^V = \int_t^T \frac{1}{2} \sigma^2 \left\{ [u(s, X_s) + Y_s^V]^3 \right\} ds + u(s, X_s) \left( \sum_{d=1}^D \sin(x_{d,0}) \right)^2 - \left( \sum_{d=1}^D \sin(x_{d,s}) \right)^2 \right\} ds$$
$$- \sum_{d=1}^D \int_t^T Z_{d,t}^V dW_{d,s}.$$ (27)

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

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

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

$$\eta_1(i, x) = \mathds{1}_{(x < 0)}(x), \quad \eta_{d,K} = \mathds{1}_{(x \geq 3)}(x),$$
$$\eta_k(i, x) = \mathds{1}_{(x \in [3(k-1)/(K-2), 3k/(K-2))]}(x), \quad k = 1, \ldots, K - 2,$$

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

$$N = \left[ 2\sqrt{2}^{j-1} \right], \quad K = \max \left\{ \left\lfloor \sqrt{2}^{l-1} \right\rfloor, 3 \right\}, \quad \Lambda = \left[ 2\sqrt{2}^{(j-1)} \right].$$
4 Numerical examples

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 \( \ell \), in which we simultaneously increase the parameters \( N, K \) and \( \Lambda \) 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 \( \Lambda \) is proportional to \( N^{2+\beta}K^2 \). With our choice of the parameters, we hence expect convergence of order 1/2 for \( \ell = 5 \). The choice \( \ell = 4 \) is just on the threshold of the theoretical convergence results, and we hence cannot expect convergence for \( \ell = 3 \). In order to illustrate the computational effort, the table below displays the number of simulated paths for the different choices of \( \ell \) in dependence of the number of time steps \( N \).

![Table 1: Sample size \( \Lambda \) in dependence of \( N \) and \( \ell \)]

<table>
<thead>
<tr>
<th>( N )</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>6</th>
<th>8</th>
<th>11</th>
<th>16</th>
<th>23</th>
<th>32</th>
<th>45</th>
<th>64</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \ell )</td>
<td>3</td>
<td>2</td>
<td>6</td>
<td>17</td>
<td>46</td>
<td>129</td>
<td>363</td>
<td>1025</td>
<td>2897</td>
<td>8193</td>
<td>23171</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>9</td>
<td>33</td>
<td>129</td>
<td>513</td>
<td>2049</td>
<td>8193</td>
<td>32769</td>
<td>131073</td>
<td>5241289</td>
<td>2097153</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>12</td>
<td>65</td>
<td>363</td>
<td>2049</td>
<td>11586</td>
<td>65537</td>
<td>370728</td>
<td>2097153</td>
<td>11863284</td>
<td>6710865</td>
</tr>
</tbody>
</table>

Note that the choice \( \ell = 5 \), which corresponds to convergence of order 1/2, yields tremendous simulation costs.

Given these parameters, we compute the coefficients \( \hat{\alpha}_{\pi N}^{0,i} \) and \( \hat{\alpha}_{\pi N}^{1,i} \) for the linear combination of the basis function by least-squares Monte Carlo 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

\[
\max_{0 \leq i \leq N} E_0|Y_{t_i} - \hat{Y}_{t_i}^{\pi N}|^2 + \sum_{i=1}^{N-1} \frac{T}{N} E_0|Z_{t_i} - \hat{Z}_{t_i}^{\pi N}|^2
= \max_{0 \leq i \leq N} E_0|\sin(X_{t_i}^{\pi N,MS}) - \hat{Y}_{t_i}^{\pi N}|^2 + \sum_{i=1}^{N-1} \frac{T}{N} E_0|\sigma \cos(X_{t_i}^{\pi N,MS}) \sin(X_{t_i}^{\pi N,MS}) - \hat{Z}_{t_i}^{\pi N}|^2 + O(N^{-2}),
\]

where \( X_{t_i}^{\pi N,MS} \) indicates the approximation of \( X \) by the Milstein scheme. Here again \( E_0 \) denotes the conditional expectation given \( \mathcal{G}_0 \), where \( \mathcal{G}_t = \mathcal{F}_t \lor \sigma(\Xi) \), and \( \Xi \) 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 1000\( N \) 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 \( \ell = 4 \) and \( \ell = 5 \), 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

$$
E_{\pi N}(\hat{Y}_{\pi N}, \hat{Z}_{\pi N})
= E_0[|\sin(X_{t_N}^{\pi N, MS}) - \hat{Y}_{t_N}^{\pi N}|^2] + \max_{1 \leq i \leq N} E_0\left[ \hat{Y}_{t_i}^{\pi N} - \hat{Y}_{t_0}^{\pi N} - \sum_{j=0}^{i-1} \frac{1}{2} \sigma^2[(\hat{Y}_{t_j}^{\pi N})^3] \Delta t_j - \sum_{j=0}^{i-1} \hat{Z}_{t_j}^{\pi N} \Delta W_j \right]^2.
$$

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

$$
\max_{i=0, \ldots, N-1} \sup_{t \in [t_i, t_{i+1}]} E_0\left[ |Y_t - \hat{Y}_{t_i}^{\pi N}|^2 \right] + \sum_{i=0}^{N-1} E_0\left[ \int_{t_i}^{t_{i+1}} |Z_t - \hat{Z}_{t_i}^{\pi N}|^2 dt \right] \leq C \left( E_{\pi N}(\hat{Y}_{\pi N}, \hat{Z}_{\pi N}) + \frac{1}{N} \right)
$$

and

$$
\max_{i=0, \ldots, N-1} \sup_{t \in [t_i, t_{i+1}]} E_0\left[ |Y_t - \hat{Y}_{t_i}^{\pi N}|^2 \right] + \sum_{i=0}^{N-1} E_0\left[ \int_{t_i}^{t_{i+1}} |Z_t - \hat{Z}_{t_i}^{\pi N}|^2 dt \right] \geq c E_{\pi N}(\hat{Y}_{\pi N}, \hat{Z}_{\pi N}) - \frac{1}{N^2}
$$

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

$$
E_{\pi N}(\hat{Y}_{\pi N}, \hat{Z}_{\pi N}) \geq \text{const.} \frac{1}{N}.
$$
4 Numerical examples

Figure 2 displays the a-posteriori error criterion for the three cases $l = 3, 4, 5$. As before, the expectation is replaced by a sample mean over $1000 N$ 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](image)

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
A Proof of inequality (13)

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

\[
\tilde{Y}^\pi_t = \mathbb{E}[Y^\pi_{t_{i+1}} | \mathcal{F}_t] - f^\pi(t_i, Y^\pi_{t_i}, Z^\pi_{t_i})(t_{i+1} - t)
\]

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

\[
Y_t = Y_{t_{i+1}} - \int_{t_i}^{t_{i+1}} f(s, Y_s, Z_s)ds - \int_{t_i}^{t_{i+1}} Z_s dW_s,
\]

\[
\tilde{Y}^\pi_t = Y^\pi_{t_{i+1}} - \int_{t_i}^{t_{i+1}} f^\pi(t_i, Y^\pi_{t_i}, Z^\pi_{t_i})dt - \int_{t_i}^{t_{i+1}} \tilde{Z}^\pi_s dW_s,
\]

By Itô’s formula we then obtain

\[
\mathbb{E}\left[|Y_t - \tilde{Y}^\pi_t|^2\right] + \int_{t_i}^{t_{i+1}} \mathbb{E}\left[|Z_s - \tilde{Z}^\pi_s|^2\right] ds
\]

\[
\leq \mathbb{E}\left[|Y_{t_{i+1}} - \tilde{Y}^\pi_{t_{i+1}}|^2\right] + 2 \int_{t_i}^{t_{i+1}} \mathbb{E}\left[(Y_s - \tilde{Y}^\pi_s) (f(s, Y_s, Z_s) - f^\pi(t_i, Y^\pi_{t_i}, Z^\pi_{t_i}))\right] ds
\]

\[
= (I) + (II).
\]

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

\[
(II) \leq \gamma \int_{t_i}^{t_{i+1}} \mathbb{E}\left[|Y_s - \tilde{Y}^\pi_s|^2\right] ds + \frac{2}{\gamma} \int_{t_i}^{t_{i+1}} \mathbb{E}\left[|f^\pi(t_i, Y_s, Z_s) - f^\pi(t_i, Y^\pi_{t_i}, Z^\pi_{t_i})|^2\right] ds
\]

\[
+ \frac{2}{\gamma} \int_{t_i}^{t_{i+1}} \mathbb{E}\left[|f(s, Y_s, Z_s) - f^\pi(t_i, Y_s, Z_s)|^2\right] ds
\]

\[
\leq \gamma \int_{t_i}^{t_{i+1}} \mathbb{E}\left[|Y_s - \tilde{Y}^\pi_s|^2\right] ds + \frac{4}{\gamma} \int_{t_i}^{t_{i+1}} K^2 \mathbb{E}\left[|Y_s - Y^\pi_{t_i}|^2\right] + K^2 \mathbb{E}\left[|Z_s - Z^\pi_{t_i}|^2\right] ds
\]

\[
+ C \int_{t_i}^{t_{i+1}} \mathbb{E}\left[|f(s, Y_s, Z_s) - f^\pi(t_i, Y_s, Z_s)|^2\right] ds.
\]
Due to [Zhang (2004)], Lemma 2.4 (i) we have
\[
E \left[ |Y_s - Y_{t_i}^\pi|^2 \right] \leq 2E \left[ |Y_s - Y_{t_i}|^2 \right] + 2E \left[ |Y_t - Y_{t_i}^\pi|^2 \right] \\
\leq C|\pi| + C \int_{t_i}^{t_{i+1}} E \left[ |Z_t|^2 \right] dt + 2E \left[ |Y_t - Y_{t_i}^\pi|^2 \right].
\]

Applying Itô’s isometry and Young’s inequality once more, we get
\[
(II) \leq \gamma \int_{t}^{t_{i+1}} E \left[ |Y_s - \hat{Y}_s^\pi|^2 \right] ds + \frac{8K^2}{\gamma} \left( \Delta_i E \left[ |Y_{t_i} - Y_{t_i}^\pi|^2 \right] + E \left[ |Z_s - \hat{Z}_s|^2 \right] ds \right) \\
+ K_y \left( C|\pi| \Delta_i + C\Delta_i \int_{t_i}^{t_{i+1}} E \left[ |Z_t|^2 \right] dt \right) + CE \left( \int_{t_i}^{t_{i+1}} \hat{Z}_s dW_s - \hat{Z}_{t_i} \Delta W_i \right) \\
+ C \int_{t_i}^{t_{i+1}} E \left[ |f(s, Y_s, Z_s) - f^\pi(t_i, Y_s, Z_s)|^2 \right] ds \\
= \gamma \int_{t}^{t_{i+1}} E \left[ |Y_s - \hat{Y}_s^\pi|^2 \right] ds + \frac{8K^2}{\gamma} A_i + B_i.
\]

Summarizing, we have
\[
E \left[ |Y_{t_i} - \hat{Y}_{t_i}^\pi|^2 \right] \leq E \left[ |Y_t - \hat{Y}_t^\pi|^2 \right] + \int_{t}^{t_{i+1}} E \left[ |Z_s - \hat{Z}_s|^2 \right] ds \\
\leq E \left[ |Y_{t_{i+1}} - \hat{Y}_{t_i}^\pi|^2 \right] + \gamma \int_{t}^{t_{i+1}} E \left[ |Y_s - \hat{Y}_s^\pi|^2 \right] ds + \frac{8K^2}{\gamma} A_i + B_i. \quad (29)
\]

By Gronwall’s lemma it follows that \( E \left[ |Y_{t_i} - \hat{Y}_{t_{i+1}}|^2 \right] \leq e^{\gamma \Delta_i} (E \left[ |Y_{t_{i+1}} - \hat{Y}_{t_{i+1}}^\pi|^2 \right] + 8K^2A_i/\gamma + B_i) \). Inserting this result into the second inequality of (29) yields
\[
E \left[ |Y_{t_i} - Y_{t_i}^\pi|^2 \right] + \int_{t_i}^{t_{i+1}} \left[ E \left[ |Z_t - \hat{Z}_{t_i}||^2 \right] dt \right] \\
\leq (1 + \gamma \Delta_i e^{\gamma \Delta_i}) (E \left[ |Y_{t_{i+1}} - \hat{Y}_{t_{i+1}}^\pi|^2 \right] + \frac{8K^2}{\gamma} A_i + B_i) \\
\leq (1 + C\gamma \Delta_i) \left( E \left[ |Y_{t_{i+1}} - \hat{Y}_{t_{i+1}}^\pi|^2 \right] + \frac{8K^2}{\gamma} A_i + B_i \right)
\]
for \(|\pi|\) small enough. Choosing \( \gamma = 32K^2 \) and \(|\pi| \leq 1/(C\gamma) \) leads to
\[
E \left[ |Y_{t_i} - Y_{t_i}^\pi|^2 \right] + \int_{t_i}^{t_{i+1}} E \left[ |Z_t - \hat{Z}_{t_i}||^2 \right] dt \leq (1 + C\gamma \Delta_i) \left( E \left[ |Y_{t_{i+1}} - \hat{Y}_{t_{i+1}}^\pi|^2 \right] + B_i \right) \\
+ \frac{1}{2} \Delta_i E \left[ |Y_{t_i} - Y_{t_i}^\pi|^2 \right] + \frac{1}{2} \int_{t_i}^{t_{i+1}} E \left[ |Z_t - \hat{Z}_{t_i}||^2 \right] dt.
\]

25
A Proof of inequality (13)

Hence, for \(|\pi|\) small enough

\[
\mathbb{E} \left[ |Y_t - Y_t^\pi|^2 \right] + \frac{1}{2} \int_{t_i}^{t_{i+1}} \mathbb{E} \left[ |Z_t - \hat{Z}_t^\pi|^2 \right] dt \\
\leq (1 + C\Delta_i) \left\{ \mathbb{E} \left[ |Y_{t_{i+1}} - Y_{t_{i+1}}^\pi|^2 \right] + K_y \left( C |\pi| \Delta_i + C \Delta_i \int_{t_i}^{t_{i+1}} \mathbb{E} \left[ |Z_t|^2 \right] dt \right) + CE \left( \int_{t_i}^{t_{i+1}} \hat{Z}_s^\pi dW_s - Z_t^\pi \Delta W_i \right)^2 \right\} + CE \left( \int_{t_i}^{t_{i+1}} \mathbb{E} \left[ |f(t, Y_t, Z_t) - f^\pi(t, Y_t, Z_t)|^2 \right] dt \right) \right\},
\]

(30)

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

\[
\mathbb{E} \left[ |Y_t - Y_t^\pi|^2 \right] \leq C \left\{ \mathbb{E} \left[ |\xi - \xi^\pi|^2 \right] + K_y |\pi| + \sum_{j=1}^{N-1} \mathbb{E} \left[ \left( \int_{t_i}^{t_{j+1}} \hat{Z}_s^\pi dW_s - Z_t^\pi \Delta W_i \right)^2 \right] \right\} + \sum_{j=1}^{N-1} \int_{t_j}^{t_{j+1}} \mathbb{E} \left[ |f(t, Y_t, Z_t) - f^\pi(t, Y_t, Z_t)|^2 \right] dt \right\},
\]

(31)

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

\[
\sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \mathbb{E} \left[ |Z_t - \hat{Z}_t^\pi|^2 \right] dt \\
\leq C \left\{ \mathbb{E} \left[ |\xi - \xi^\pi|^2 \right] + K_y |\pi| + \sum_{j=1}^{N-1} \mathbb{E} \left[ \left( \int_{t_i}^{t_{j+1}} \hat{Z}_s^\pi dW_s - Z_t^\pi \Delta W_i \right)^2 \right] \right\} + \sum_{j=1}^{N-1} \int_{t_j}^{t_{j+1}} \mathbb{E} \left[ |f(t, Y_t, Z_t) - f^\pi(t, Y_t, Z_t)|^2 \right] dt + \max_{t_i \in \pi} \mathbb{E} \left[ |Y_{t_i} - Y_{t_i}^\pi|^2 \right] \}
\]

(32)

Noting that

\[
\sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \mathbb{E} \left[ |Z_t - Z_t^\pi|^2 \right] dt \leq 2 \sum_{i=0}^{N-1} \left( \int_{t_i}^{t_{i+1}} \mathbb{E} \left[ |Z_t - \hat{Z}_t^\pi|^2 \right] dt + \mathbb{E} \left[ \left( \int_{t_i}^{t_{i+1}} \hat{Z}_s^\pi dW_s - Z_t^\pi \Delta W_i \right)^2 \right] \right),
\]

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


References


Preprint Series DFG-SPP 1324

http://www.dfg-spp1324.de

Reports


