Numerical solutions of BSDEs: A-posteriori estimates and enhanced least-squares Monte Carlo

Dissertation zur Erlangung des Grades der

Doktorin der Naturwissenschaften der Naturwissenschaftlich-Technischen Fakultäten der Universität des Saarlandes (Dr. rer. nat.)

eingereicht

am 31. Mai 2012

in Saarbrücken

von

Jessica Steiner

Tag des Kolloquiums:	11. September 2012
Mitglieder des Prüfungsausschusses:	
Vorsitzender:	Prof. Dr. Anselm Lambert
Protokollführer:	Dr. Moritz Weber
1. Berichterstatter:	Prof. Dr. Christian Bender
2. Berichterstatter:	Prof. Dr. Bruno Bouchard
Dekan:	Prof. Dr. Mark Groves

To my parents

Eidesstattliche Versicherung

Hiermit versichere ich an Eides statt, dass ich die vorliegende Arbeit selbstständig und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus anderen Quellen oder indirekt übernommenen Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet. Die Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form in einem Verfahren zur Erlangung eines akademischen Grades vorgelegt.

Saarbrücken, 31. Mai 2012

(Dipl.-Math. Jessica Steiner)

Abstract

Backward stochastic differential equations (BSDEs) are a powerful tool in financial mathematics. Important examples are option pricing or portfolio selection problems. In non-linear cases BSDEs are usually not solvable in closed form and approximation becomes then inevitable. Several proposals for solving BSDEs numerically have been published in recent years, including an analysis of the related approximation error.

The first part of this theses is devoted to the problem that a direct a-posteriori evaluation of the L^2-error between the true solution and some numerical solution is usually impossible. Therefore, we present an a-posteriori criterion on the approximation error, which is computable in terms of the numerical solution only and allows us to judge the numerical solution.

Secondly, we pick up the idea of Gobet, Lemor and Warin (Ann. Appl. Probab., 15, 2172 – 2202 (2005)) to generate numerical solutions by least-squares Monte Carlo. We suggest to use function bases that form a system of martingales. A complete analysis of the approximation error shows, that in contrast to original least-squares Monte Carlo, the convergence behaviour can be significantly enhanced by the martingale property of the bases.

Deutsche Zusammenfassung

Rückwärtsgerichtete stochastische Differentialgleichungen (BSDEs) sind ein vielseitiges Instrument in der Finanzmathematik. Optionsbepreisung oder Portfolio-Auswahlprobleme sind wichtige Beispiele dafür. In nichtlinearen Fällen sind BSDEs in der Regel jedoch nicht geschlossen lösbar, weshalb in den vergangenen Jahren zahlreiche numerische Ansätze zusammen mit einer theoretischen Analyse ihres Approximationsfehlers vorgestellt worden sind.

Der erste Teil dieser Arbeit beschäftigt sich mit dem Problem, dass eine direkte a-posteriori Berechnung des L^2-Fehlers zwischen der unbekannten echten und der numerischen Lösung oftmals unmöglich ist. Deshalb präsentieren wir ein a-posteriori Kriterium, das nur von der numerischen Lösung abhängt und eine Beurteilung dieser erlaubt.

Der zweite Teil baut auf der Idee von Gobet, Lemor und Warin (Ann. Appl. Probab., 15, 2172 – 2202 (2005)) auf, numerische Lösungen mit Hilfe eines Kleinste-Quadrate-Monte-Carlo-Verfahrens zu erzeugen. Wir schlagen Funktionenbasen vor, die ein System von Martingalen bilden. Eine vollständige Analyse des Approximationsfehlers zeigt, dass das Konvergenzverhalten durch die Martingaleigenschaft erheblich verbessert wird im Vergleich zum ursprünglichen Verfahren.

Contents

1 Introduction

1.1 Background on BSDEs and their numerical solution

The theory of backward stochastic differential equations (BSDEs) is a rather young research field and its subjects first popped up in the context of stochastic control. It was Bismut (1973) who carried Pontryagin's maximum principle over to stochastic control problems and showed that the pair of adjoint processes solves a linear BSDE.

The actual foundation for BSDE theory was laid later on by Pardoux and Peng (1990), who examined non-linear BSDEs and proved the well-posedness of such equations in case of a Lipschitz continuous driver. In the following, numerous publications were devoted to an extension of this result.

One branch was engaged with the relaxation of the Lipschitz condition on the driver. For instance, see Lepeltier and San Martín (1997), who examined BSDEs with continuous driver of linear growth, or Kobylanski (2000) on BSDEs with drivers of quadratic growth. An overview is given in El Karoui and Mazliak (1997). Another important aspect was the analysis of the connection between solutions of BSDEs and viscosity solutions for quasilinear parabolic partial differential equations by Pardoux and Peng (1992). Based on this, the notion of forward backward stochastic differential equations (FBSDEs) was developed and a generalization of the Feynman-Kac formula was obtained. A detailed introduction on this topic is also available in Ma and Yong (1999). Particularly, FBSDEs became a useful tool in the field of financial mathematics. Amongst these are pricing and hedging of European options in cases with constraints or utility optimization problems. An extension to American options by BSDEs with reflection was shown in El Karoui et al. (1997a). A comprehensive survey on the application of BSDEs in finance is given by El Karoui et al. (1997).

Whereas the research on BSDEs was indeed fruitful from its kick-off in the early nineties on, the pioneering work on the numerics of BSDEs initially advanced much slower. Bally (1997) was the first who proposed a time discretization scheme as a numerical approach towards the solution of BSDEs. Then it remained to solve a series of linear BSDEs within each time step. The main drawback of this approach is that the time steps have to be chosen randomly in order to avoid any stronger regularity assumptions on the coefficients of the BSDE that go beyond the Lipschitz continuity of the driver.

Chevance (1997) presented a fully implementable numerical attempt to solve a decoupled FBSDE with a deterministic time discretization. However, this was connected with quite strong regularity conditions on the coefficient of both the forward and the backward SDE. It was Zhang (2001) who offered a way out of this dilemma by formulating conditions such that the control part of the solution of a BSDE behaves somewhat 'nice'. These conditions include Lipschitz conditions on the coefficients of the forward SDE and the possibly path-dependent terminal condition of the BSDE and are merely an addition to the Lipschitz continuity of the driver. In particular, he introduced the notion of L^2-regularity for stochastic processes and showed that these mild extra conditions are sufficient for the L^2-regularity of the control part and thus also for the convergence of a deterministic time discretization with order 1/2 in the number of time steps.

Slightly different, but somewhat more natural ways of time discretization for decoupled FBSDEs were examined in Bouchard and Touzi (2004) and Lemor et al. (2006), however they both benefit from the L^2-regularity results obtained by Zhang (2001). In contrast to the algorithm suggested in Lemor et al. (2006), the approach by Bouchard and Touzi (2004) is characterized by its implicit formulation. There are several proposals to turn this idea into a tractable algorithm by using some sort of Picard iteration. This can be done within each time step, see Gobet et al. (2005), or globally by an iteration that restarts at terminal time after having completed the iteration step along the entire partition, see Bender and Denk (2007). Both methods have to deal with the problem of nested conditional expectations, on the one hand along the partition of the time interval and on the other one along the Picard iterations. Bender and Denk (2007) showed that the global Picard iteration is more favorable concerning the error propagation that arises when estimating conditional expectations. The work of Gobet and Labart (2010) is also in the spirit of global Picard iteration connected with a control variate technique. Another way of variance reduction within a global Picard scheme was presented in Bender and Moseler (2010), who applied the so-called importance sampling technique that makes use of measure change to receive more samples in 'interesting' regions.

Extensions to this research can be found in Gobet and Makhlouf (2010) and Geiss et al. (2011), who supposed the terminal condition to be irregular. Even then the error due to time discretization tends to zero, although the convergence rate is in this case slower for equidistant partitions of the time interval. However, a clever choice of partition can improve this rate significantly, in certain cases up to 1/2 in the number of time steps. Worth mentioning is the work of Imkeller et al. (2010) and also Richou (2011) on numerical approximation of BSDEs with drivers of quadratic growth in the control part. In the first case, the non-Lipschitz continuity was tackled by imposing a truncation on the driver and approximating the true BSDE by a series of truncated ones. In the latter one time-dependent bounds for the control part within the time discretization scheme were incorporated.

Regarding coupled FBSDEs, Bender and Zhang (2008) proposed a combination of time discretization and Markovian iteration to tackle the coupling. They formulated also sufficient conditions for a time discretization error that decreases with rate 1/2 in the number of time steps. The case of FBSDEs with jumps was covered by Bouchard and Elie (2008).

Whatever type of time discretization is chosen, at the end of the day one is confronted with the problem of estimating conditional expectations. This stems from the backward property of BSDEs and the necessity to adapt the approximation to the available information at each time step. In recent years several proposals have been established to cope with this problem. Bouchard and Touzi (2004) applied Malliavin

Monte Carlo for the estimation of conditional expectations. By means of Malliavin integration by parts, these can be expressed by a ratio of expectations, that can be estimated via Monte Carlo simulation. See also Bouchard et al. (2004).

An alternative was considered in Bally and Pagès (2003), who chose the quantization tree method for the estimation of conditional expectations. Roughly speaking, the idea is to project the time-discretized underlying diffusion process on discrete state spaces and to estimate the transition probabilities between the single time steps by simulation. The conditional expectations are then easily computable weighted sums. Delarue and Menozzi (2006) transferred this idea to the numerical solution of coupled FBSDEs.

Only recently, Crisan and Manolarakis (2010) exploited the cubature method for the estimation of conditional expectations for the generation of numerical solutions of BSDEs.

Last but not least, Gobet et al. (2005) tackled the estimation of conditional expectations by least-squares Monte Carlo. This approach can be understood as a two-step procedure that starts with a projection on a function basis and next solves the resulting minimization problem by Monte Carlo simulation. We will explain this idea later on in more detail.

1.2 Problem description

Let $(\Omega, \mathbb{F}, P)$ be a probability space, where $\mathbb{F} = (\mathcal{F}_t, t \geqslant 0)$ is the augmented filtration generated by a D-dimensional Brownian motion $W = (W_1, \dots, W_D)^*$. Here the star denotes matrix transposition. We fix further a terminal time $T > 0$. Then our first branch of studies starts with a backward stochastic differential equation (BSDE) of the form

$$Y_t = \xi - \int_t^T f(u, Y_u, Z_u)du - \int_t^T Z_u dW_u, \tag{1.1}$$

where the data is assumed to satisfy

Assumption 1. *(i) The terminal condition* ξ *is a real valued, square-integrable,* $\mathcal{F}_T$*-measurable random variable.*

(ii) The driver is a measurable function $f : \Omega \times [0,T] \times \mathbb{R} \times \mathbb{R}^D \to \mathbb{R}$*, such that* $(f(t,0,0), 0 \leqslant t \leqslant T)$ *is a continuous,* $\mathbb{F}$*-adapted process with* $\int_0^T E|f(t,0,0)|^2 dt < \infty$*. Moreover,* f *is Lipschitz in its spatial variables with constant* κ *uniformly in* (t, ω)*. Note, that the stochastic variable is suppressed in the above equation.*

The solution of (1.1) consists of a pair of adapted stochastic processes (Y, Z), where Y_t is real valued and $Z_t = (Z_{1,t}, \dots, Z_{D,t})$ is $\mathbb{R}^D$-valued. However, in most cases we cannot state a closed-form solution for (1.1) and a workaround by numerical approaches becomes highly interesting in order to obtain at least an approximation of (Y, Z).

Let us assume, we conducted some arbitrary numerical scheme, that is based on a discretization $\pi = \{t_0, \dots, t_N\}$ of the interval $[0,T]$, namely $0 = t_0 < t_1 < \dots <$

$t_N = T$, and its result is the pair $(\hat{Y}^\pi_{t_i}, \hat{Z}^\pi_{t_i})_{t_i\in\pi}$. Then, a quite natural wish is to get information about the approximation error. Precisely, we want to check

$$\sup_{0\leqslant t\leqslant T} E|Y_t - \hat{Y}^\pi_{t_i}|^2 + \int_0^T E|Z_t - \hat{Z}^\pi_{t_i}|^2 dt$$

and judge thereby, if the chosen numerical approach was successful. Here, the pair $(\hat{Y}^\pi_t, \hat{Z}^\pi_t)_{0\leqslant t\leqslant T}$ denotes the RCLL-extension of $(\hat{Y}^\pi_{t_i}, \hat{Z}^\pi_{t_i})_{t_i\in\pi}$ by constant interpolation. But, as the true solution is usually unknown to us, it is not possible to compute the approximation error directly or even estimate it, e. g. by Monte Carlo simulation.

Nevertheless, we want to shed some light on the question, whether the pair $(\hat{Y}^\pi_{t_i}, \hat{Z}^\pi_{t_i})_{t_i\in\pi}$ is a good approximation. For this purpose we introduce a so-called 'global' a-posteriori error criterion. Suppose that $(\hat{Y}^\pi_{t_i}, \hat{Z}^\pi_{t_i})_{t_i\in\pi}$ is adapted to a filtration $\mathbb{G} = (\mathcal{G}_t, 0 \leqslant t \leqslant T)$ such that $\mathcal{F}_t \subseteq \mathcal{G}_t$ for $t \in [0,T]$. That means, $\mathbb{G}$ is enlarged in comparison to $\mathbb{F}$ and W is still a Brownian motion with respect to $\mathbb{G}$. But $\mathcal{G}_{t_i}$ can also contain additional information, for instance induced by copies of W_{t_i} that were required for the approximation of $(\hat{Y}^\pi_{t_i}, \hat{Z}^\pi_{t_i})_{t_i\in\pi}$. Least-squares Monte Carlo simulation for BSDEs is an example for the incorporation of such copies. Then the global a-posteriori criterion checks by

$$\begin{aligned}\mathcal{E}_\pi(\hat{Y}^\pi, \hat{Z}^\pi) := E[|\xi^\pi - \hat{Y}^\pi_{t_N}|^2|\,\mathcal{G}_{t_0}] + \max_{1\leqslant j\leqslant N} E[|\hat{Y}^\pi_{t_j} - \hat{Y}^\pi_{t_0} \\ - \sum_{i=0}^{j-1} f^\pi(t_i, \hat{Y}^\pi_{t_i}, \hat{Z}^\pi_{t_i})(t_{i+1} - t_i) - \sum_{i=0}^{j-1} \hat{Z}^\pi_{t_i}(W_{t_{i+1}} - W_{t_i})|^2|\,\mathcal{G}_{t_0}], \quad (1.2)\end{aligned}$$

if the approximate solution is 'close to solving' (1.1). Here, (ξ^π, f^π) denotes an approximation of (ξ, f) living on the time grid π. Contrary to the approximation error, it is possible to simulate (1.2), as it involves only approximate, hence known solutions and approximate data. In a first step, we will develop upper and lower estimates on the approximation error in terms of this criterion. These estimates require only standard Lipschitz conditions on the driver f.

After that, we apply the global error criterion on a forward backward stochastic differential equation (FBSDE) denoted by

$$\begin{aligned} S_t &= s_0 + \int_0^t b(u, S_u)du + \int_0^t \sigma(u, S_u)dW_u \\ Y_t &= \phi(S) - \int_t^T F(u, S_u, Y_u, Z_u)du - \int_t^T Z_u dW_u. \end{aligned} \quad (1.3)$$

This system is supposed to fulfill

Assumption 2. *We call* $s_0 \in \mathbb{R}^{\tilde{D}}$ *the initial condition of* S. *The functions* $b : [0,T]\times\mathbb{R}^{\tilde{D}} \to \mathbb{R}^{\tilde{D}}$, $\sigma : [0,T] \times \mathbb{R}^{\tilde{D}} \to \mathbb{R}^{\tilde{D}\times D}$ *and* $F : [0,T] \times \mathbb{R}^{\tilde{D}} \times \mathbb{R} \times \mathbb{R}^D \to \mathbb{R}$ *are deterministic and there is a constant* κ *such that*

$$\begin{aligned}|b(t,s) - b(t',s')| + |\sigma(t,s) - \sigma(t',s')| + |F(t,s,y,z) - F(t',s',y',z')| \\ \leqslant \kappa(\sqrt{t-t'} + |s-s'| + |y-y'| + |z-z'|)\end{aligned}$$

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

$$|\phi(\boldsymbol{s}) - \phi(\boldsymbol{s}')| \leqslant \kappa \sup_{0 \leqslant t \leqslant T} |\boldsymbol{s}(t) - \boldsymbol{s}'(t)|$$

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

$$\sup_{0 \leqslant t \leqslant T} (|b(t, 0)| + |\sigma(t, 0)| + |F(t, 0, 0, 0)|) + |\phi(\boldsymbol{0})| \leqslant \kappa,$$

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

Clearly, we look at a BSDE with data $\xi = \phi(S)$ and $f(t, y, z) = F(t, S, y, z)$, where F is stochastic through S only. The above system is called decoupled as the forward SDE is independent of the pair (Y, Z). Given Assumption 2 it is easy to check, that the conditions of Assumptions 1 are satisfied as well.

Concerning this type of FBSDEs, we will take a closer look on a numerical method that combines a backwards time discretization scheme with the least-squares Monte Carlo approach for the estimation of conditional expectations to generate approximations of the processes Y and Z. This method was already employed by Gobet et al. (2005) and Lemor et al. (2006) in this setting and aims at replacing the conditional expectations by a projection on a subspace of $L^2(\mathcal{F}_{t_i})$ for each time step t_i.

We will review the approximation error of this scheme and explain its error sources, in particular the time discretization error, the projection error and the simulation error. Moreover, we will recall how the parameters of the latter one can be fixed such that the overall approximation error converges with the same rate as the time discretization error.

Additionally, we present for this setting a 'local' a-posteriori error criterion, that is denoted by

$$\mathcal{E}^{loc}_{\pi,j}(\hat{Y}^\pi, \hat{Z}^\pi) := \sum_{i=j}^{N-1} E|\hat{Y}^\pi_{t_{i+1}} - \hat{Y}^\pi_{t_i} - F(t_i, S^\pi_{t_i}, \hat{Y}^\pi_{t_i}, \hat{Z}^\pi_{t_i})(t_{i+1} - t_i) - \hat{Z}^\pi_{t_i}(W_{t_{i+1}} - W_{t_i})|^2,$$

for $j = 0, \ldots, N-1$. It is meant to give further information about the projection error, which is expressed in terms of the L^2-error between a time-discretized solution $(Y^\pi_{t_i}, Z^\pi_{t_i})_{t_i \in \pi}$ and its best projection on the selected function basis. Precisely, it turns out that a small local error criterion is a necessary condition for a small projection error. Furthermore, it allows us to detect those time steps for which the projection functions were picked inappropriately.

The second branch of our studies is devoted to a modification of the least-squares Monte Carlo approach. Induced by the time discretization we are confronted with the estimation of

$$\begin{aligned}&E[\hat{Y}^{\pi}_{t_{i+1}} - F(t_i, S^{\pi}_{t_i}, \hat{Y}^{\pi}_{t_i}, \hat{Z}^{\pi}_{t_i})(t_{i+1} - t_i)|\mathcal{F}_{t_i}]\\ &\frac{1}{(t_{i+1} - t_i)}E[(W_{t_{i+1}} - W_{t_i})\hat{Y}^{\pi}_{t_{i+1}}|\mathcal{F}_{t_i}],\end{aligned}$$

by a linear combination of basis functions. The estimation of the first conditional expectation leads to the definition of $\hat{Y}^{\pi}_{t_i}$, whereas the estimation of the latter one is required for $\hat{Z}^{\pi}_{t_i}$. Motivated by a kind of variance reduction for FBSDEs, we assume that the function bases form a system of martingales. Let $(X^{\pi}_{t_i})_{t_i \in \pi}$ be a $\mathbb{F}$-adapted Markov process and $\eta(i+1, X^{\pi}_{t_{i+1}})$ a basis function at time t_{i+1} such that

(i) its conditional expectation related to $\mathcal{F}_{t_i}$ is computable in closed-form,

(ii) the conditional expectation of this function multiplicated with the dth component of the Brownian increment $W_{t_{i+1}} - W_{t_i}$ can be evaluated related to $\mathcal{F}_{t_i}$ for all $d = 1, \ldots, D$.

This suggestion is inspired by Glasserman and Yu (2004) in the field of pricing American options. Assumption (i) is related to the approximation of Y and ensures that $(\eta(i, X^{\pi}_{t_i})_{t_i \in \pi}$ forms a martingale with respect to $\mathbb{F}$, that is available in closed form. In this setting, the estimation of conditional expectations becomes obsolete for all linear terms, as they can be computed in closed form under the assumption of martingale basis functions. E. g., let $\hat{Y}^{\pi}$ at time t_{i+1} be a linear combination of so-called martingale basis functions, then we can figure out its conditional expectation by the martingale property. This simplifies the approximation of $\hat{Y}^{\pi}_{t_i}$, as only $E[F(t_i, S^{\pi}_{t_i}, \hat{Y}^{\pi}_{t_i}, \hat{Z}^{\pi}_{t_i})|\mathcal{F}_{t_i}]$ remains to be estimated.

Moreover, by (ii) the evaluation of $E[(W_{d,t_{i+1}} - W_{d,t_i})\hat{Y}^{\pi}_{t_{i+1}}|\mathcal{F}_{t_i}], d = 1, \ldots, D$, which stems from the time discretization of Z_d, becomes possible in closed form. That means, we do not require any additional estimation of conditional expectations for the approximate solution of Z_d. This is particularly interesting in high-dimensional problems, when $D > 1$. Clearly, in the martingale basis approach the amount of conditional expectations to be estimated is the same, no matter if the Brownian motion W is one-dimensional or multi-dimensional.

We give several examples for 'martingale type' basis functions and conduct afterwards a detailed analysis of the approximation error and its error sources. It turns out that the projection error and the simulation error can be reduced significantly in contrast to the original least-squares Monte Carlo approach.

The rest of this thesis is organized as follows. In Chapter 2 we review some important results on BSDEs that are essential for this paper. Additionally, we explain the least-squares Monte Carlo approach and its approximation error in detail. Chapter 3 is devoted to the a-posteriori error criteria. Apart from the global criterion, we present the local one for approximate solutions that where obtained by replacing conditional expectations by projections. This chapter ends with the introduction of non-linear control variates for (F)BSDEs inspired by variance reduction methods. Therefore, we will diminish the original BSDE by some BSDE, that is solvable in closed-form and is likely to 'explain' the main part of the original one. Approximation has then to be applied to the remainder BSDE. The chapter also includes numerical examples. In Chapter 4 we introduce the enhanced least-squares Monte Carlo scheme and examine the approximation error in its very detail. Again, the chapter is finished by numerical examples.

2 Preliminaries

2.1 Some important results on BSDEs

Before turning to the numerical solution of BSDEs and their validation, it is essential to know if the problem in (1.1) is well-defined. To this end, we cite a result of Pardoux and Peng (1990).

Theorem 1. *We suppose that the data* (ξ, f) *satisfy Assumption 1. Then there is a unique pair of predictable processes* (Y, Z) *with*

$$E\int_0^T |Y_t|^2 dt < \infty, \quad E\int_0^T |Z_t|^2 dt < \infty,$$

that solves the differential equation (1.1).

During our thesis we require some standard regularity estimates on the processes S and Y several times. These results can also be found in Zhang (2004), who more generally considers the L^p-norm for $p \geqslant 2$ instead of the case $p = 2$ only.

Lemma 2. *Let Assumption 1 be fulfilled and* (Y, Z) *be an adapted solution of* (1.1). *Then there is a constant* C *depending on* T, κ *and the data* (ξ, f) *only such that*

$$E|Y_t - Y_s|^2 \leqslant C|t-s| + C\int_s^t E|Z_u|^2 du.$$

Lemma 3. *Let Assumption 2 be fulfilled and* S *be a solution of the forward SDE in* (1.3). *Then there is a constant* C *depending on* T, κ, s_0 *and the data* (b, σ) *such that*

$$E|S_t - S_s|^2 \leqslant C|t-s|.$$

It was Zhang (2001), who made an important contribution concerning the regularity of the process Z. Beyond inventing the notion of L^2-regularity by

$$\sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} E\left| Z_t - \frac{1}{\Delta_i} E\left[\int_{t_i}^{t_{i+1}} Z_s ds \,\middle|\, \mathcal{F}_{t_i} \right] \right|^2 dt,$$

he showed that rather mild conditions are sufficient to obtain a regularity rate of order 1/2 in the number of time steps of a deterministic partition of the time interval $[0, T]$. This result is essential for the convergence of a discrete-time approximation of (Y, Z), as will be reviewed in the next subsection.

Theorem 4. *We suppose that Assumption 2 is satisfied. Let* $\pi = \{t_0, \dots, t_N\}$ *be a partition of* $[0, T]$ *with* $0 = t_0 < \dots < t_i < \dots < t_N = T$. *Then there is a constant* $C > 0$ *depending on* T, κ *and* s_0 *only such that*

$$\sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} E\left| Z_t - \frac{1}{\Delta_i} E\left[\int_{t_i}^{t_{i+1}} Z_s ds \middle| \mathcal{F}_{t_i} \right] \right|^2 dt \leqslant C \max_{0 \leqslant i \leqslant N-1} |t_{i+1} - t_i|,$$

where C *is independent of* π.

2.2 The least-squares Monte Carlo algorithm for BSDEs

The least-squares Monte Carlo algorithm for BSDEs was initially proposed by Gobet et al. (2005) for the numerical solution of FBSDEs as formulated in (1.3) and is based on a discrete-time approximation of (Y, Z). Then least-squares Monte Carlo comes into play in order to tackle the estimation of conditional expectations, that arise during the time discretization. We explain both steps in detail in the following subsections.

2.2.1 Discrete-time approximators

There are several proposals for the time discretization of (Y, Z), for instance see Bouchard and Touzi (2004) or Zhang (2004). Here, we will explain step by step the scheme that was proposed by Lemor et al. (2006). Considering the time grid $\pi = \{t_0, \dots, t_N\}$ of $[0, T]$ with $0 = t_0 < \dots < t_i < \dots < t_N = T$, we define $|\pi| = \max_{0 \leqslant i \leqslant N-1} |t_{i+1} - t_i|$ and suppose that a discrete-time approximation $S^{\pi}_{t_i}$, $t_i \in \pi$ of the forward SDE S is at hand that fullfills

Assumption 3. *The process* $(S^{\pi}_{t_i})_{t_i \in \pi}$ *is an adapted Markov process. Moreover, there is a constant* $C > 0$ *such that*

$$\max_{0 \leqslant i \leqslant N} E|S_{t_i} - S^{\pi}_{t_i}|^2 \leqslant C|\pi|.$$

In numerous financial settings the forward SDE consists of a geometric Brownian motion, that can be sampled perfectly on the time grid π. For many other cases the Euler scheme, e. g., provides a suitable approximation satisfying Assumption 3.

For the time discretization of (Y, Z) we define $\Delta_i := t_{i+1} - t_i$, $\Delta W_{d,i} = W_{d,t_{i+1}} - W_{d,t_i}$ and $\Delta W_i = (\Delta W_{1,i}, \dots, \Delta W_{D,i})^*$. Due to the definition of the BSDE we have

$$Y_{t_i} = Y_{t_{i+1}} - \int_{t_i}^{t_{i+1}} F(u, S_u, Y_u, Z_u) du - \int_{t_i}^{t_{i+1}} Z_u dW_u.$$

Inspired by this equality, we replace the integrals by their discrete counterparts and receive the relation

$$Y_{t_i} \approx Y_{t_{i+1}} - \Delta_i F(t_i, S_{t_i}, Y_{t_i}, Z_{t_i}) - Z_{t_i} \Delta W_i. \tag{2.1}$$

Next, we multiply (2.1) with the Brownian increment $\Delta W_{d,i}$ and take after that the conditional expectation. Thus, we can derive from

$$0 \approx E[Y_{t_{i+1}} \Delta W_{d,i} | \mathcal{F}_{t_i}] - Z_{d,t_i} \Delta_i$$

an approximation $Z^\pi_{t_i}$, provided that $Y^\pi_{t_{i+1}}$ is given:

$$Z^\pi_{t_i} = \frac{1}{\Delta_i} E[(\Delta W_i)^* Y^\pi_{t_{i+1}} | \mathcal{F}_{t_i}].$$

For the time discretization of the Y-part, we take the conditional expectation in (2.1) and obtain

$$\begin{aligned} Y_{t_i} &\approx E[Y_{t_{i+1}} - \Delta_i F(t_i, S_{t_i}, Y_{t_i}, Z_{t_i}) | \mathcal{F}_{t_i}] \\ &\approx E[Y_{t_{i+1}} - \Delta_i F(t_i, S_{t_i}, Y_{t_{i+1}}, Z_{t_i}) | \mathcal{F}_{t_i}]. \end{aligned}$$

In the last step we switched from Y_{t_i} to $Y_{t_{i+1}}$, which turns the relation into an explicit one. Hence, we define for $S^\pi_{t_i}$, $Y^\pi_{t_{i+1}}$ and $Z^\pi_{t_i}$ known,

$$Y^\pi_{t_i} = E[Y^\pi_{t_{i+1}} - \Delta_i F(t_i, S^\pi_{t_i}, Y^\pi_{t_{i+1}}, Z^\pi_{t_i}) | \mathcal{F}_{t_i}].$$

Now, we want to combine these considerations to a full description of the time discretization scheme, that starts backwards in time with an approximation ξ^π of the terminal condition. We achieve a time-discretized approximation (Y^π, Z^π) of (Y, Z) by conducting for all $i = N-1, \ldots, 0$

$$\begin{aligned} Y^\pi_{t_N} &= \xi^\pi, \\ Z^\pi_{t_i} &= \frac{1}{\Delta_i} E[(\Delta W_i)^* Y^\pi_{t_{i+1}} | \mathcal{F}_{t_i}], \\ Y^\pi_{t_i} &= E[Y^\pi_{t_{i+1}} - \Delta_i F(t_i, S^\pi_{t_i}, Y^\pi_{t_{i+1}}, Z^\pi_{t_i}) | \mathcal{F}_{t_i}]. \end{aligned} \tag{2.2}$$

Using constant interpolation we get processes (Y^π_t, Z^π_t), $t \in [0, T]$. Zhang (2004) and Bouchard and Touzi (2004) introduced quite similar schemes. Roughly speaking, they differ from (2.2) due to the variables that are plugged in the driver. Particularly, the latter authors evaluate the driver F at $(t_i, S^\pi_{t_i}, Y^\pi_{t_i}, Z^\pi_{t_i})$, which leads to an implicit definition of $Y^\pi_{t_i}$. All approaches have in common that under Assumptions 2 and 3 the time discretization error in the L^2-sense is of order 1/2 in the number of time steps plus an error concerning the approximate terminal condition, i. e.

$$\sup_{0 \leqslant t \leqslant T} E|Y_t - Y^\pi_t|^2 + \int_0^T E|Z_t - Z^\pi_t|^2 dt \leqslant C|\pi| + CE|\xi - \xi^\pi|^2,$$

see Lemor et al. (2006) for a proof with respect to the above setting.

Although (2.2) is formulated explicitly in time, it incorporates the computation of (nested) conditional expectations, that in many cases cannot be figured out in closed form. Thus, estimation of conditional expectations is an important problem, when it comes to solving BSDEs numerically. In the next subsection we will review the least-squares Monte Carlo method as an estimation tool for this purpose.

Before going into the details, we endow the time-discretized solution with a kind of Markovian structure. To this end, we establish a multivariate Markov process $(X^\pi_{t_i})_{t_i\in\pi}$ such that its first component matches the discretized SDE $(S^\pi_{t_i})_{t_i\in\pi}$. In such a framework we can formulate the approximate terminal condition by $\xi^\pi = \phi^\pi(X^\pi_{t_N})$, even if the true terminal condition is path dependent, e.g. $\phi(S) = \max_{0\leqslant t\leqslant T} S_t$ or $\phi(S) = 1/T\int_0^T S_t dt$. Several examples for an appropriate construction of $(X^\pi_{t_i})_{t_i\in\pi}$ can be found in Gobet et al. (2005). In view of the Markovianity of $(X^\pi_{t_i}, \mathcal{F}_{t_i})_{t_i\in\pi}$ we can then rephrase algorithm (2.2). For $i = N-1,\dots,0$ we have

$$\begin{aligned} Y^\pi_{t_N} &= \phi^\pi(X^\pi_{t_N}),\\ Z^\pi_{t_i} &= \frac{1}{\Delta_i} E[(\Delta W_i)^* Y^\pi_{t_{i+1}} | X^\pi_{t_i}],\\ Y^\pi_{t_i} &= E[Y^\pi_{t_{i+1}} - \Delta_i F(t_i, S^\pi_{t_i}, Y^\pi_{t_{i+1}}, Z^\pi_{t_i}) | X^\pi_{t_i}]. \end{aligned} \tag{2.3}$$

Hence, there are functions $y^\pi_i(x)$ and $z^\pi_i(x)$ such that

$$Y^\pi_{t_i} = y^\pi_i(X^\pi_{t_i}), \quad Z^\pi_{t_i} = z^\pi_i(X^\pi_{t_i}), \quad i = 0,\dots,N.$$

That means, the estimation of conditional expectation aims at finding deterministic functions as approximations for y^π_i and z^π_i. In the following we describe how this can be done by least-squares Monte Carlo.

2.2.2 Least-squares Monte Carlo estimation of conditional expectations

The least-squares Monte Carlo approach to the estimation of conditional expectations was suggested in the context of pricing American options, see Longstaff and Schwartz (2001). Let U and $\tilde{X}$ be some random variables. Then the computation of $E[U|\tilde{X}]$ is equivalent to finding a function $\tilde{v}(x)$ such that

$$\tilde{v}(\tilde{X}) = \arg\min_{v} E|v(\tilde{X}) - U|^2, \tag{2.4}$$

where v is taken from the set of measurable functions with the property $E|v(\tilde{X})|^2 < \infty$. We simplify the infinite-dimensional minimization problem to a finite-dimensional one by defining a function basis $\eta(x)$ with

$$\eta(x) = \{\eta_1(x),\dots,\eta_K(x)\}, \quad K \in \mathbb{N}.$$

Thus, substituting (2.4) by the K-dimensional minimization problem

$$\tilde{\alpha} = \arg\min_{\alpha\in\mathbb{R}^K} E|\eta(\tilde{X})\alpha - U|^2 \tag{2.5}$$

reduces the original problem of finding a minimizing function to the problem of finding minimizing coefficients $\tilde{\alpha}$. This yields an orthogonal projection of U on the subspace of $L^2(\sigma(\tilde{X}))$ spanned by $\eta(\tilde{X})$. Still, we have a problem that is in general not solvable in closed form. Therefore, we replace the expectation operator in (2.5) by the sample mean and compute

$$\tilde{\alpha}^L = \arg\min_{\alpha \in \mathbb{R}^K} \frac{1}{L} \sum_{\lambda=1}^{L} |\eta({}_\lambda\tilde{X})\alpha - {}_\lambda U|^2, \tag{2.6}$$

where $({}_\lambda\tilde{X}, {}_\lambda U)$, $\lambda = 1, \dots, L$ are independent copies of $(\tilde{X}, U)$. After setting

$$\mathcal{A}^L := \frac{1}{\sqrt{L}} \begin{pmatrix} \eta_1({}_\lambda\tilde{X}) & \cdots & \eta_K({}_\lambda\tilde{X}) \end{pmatrix}_{\lambda=1,\dots,L},$$

we get a solution for (2.6) by

$$\tilde{\alpha}^L = \frac{1}{\sqrt{L}} \left((\mathcal{A}^L)^* \mathcal{A}^L\right)^{-1} (\mathcal{A}^L)^* \begin{pmatrix} {}_1U \\ \vdots \\ {}_LU \end{pmatrix}. \tag{2.7}$$

In case $(\mathcal{A}^L)^* \mathcal{A}^L$ is not invertible, we employ the pseudo inverse $\mathcal{A}^{L,+}$ of $\mathcal{A}^L$ and compute instead of (2.7) the following coefficients,

$$\tilde{\alpha}^L = \frac{1}{\sqrt{L}} \mathcal{A}^{L,+} \begin{pmatrix} {}_1U \\ \vdots \\ {}_LU \end{pmatrix}.$$

In sum, we receive by $\eta(\tilde{X})\tilde{\alpha}^L$ the least-squares Monte Carlo estimator for $E[U|\,\tilde{X}]$. The related approximation error is determined by two components, namely the projection error, that reflects the adequacy of the chosen basis functions, and the simulation error caused by the step from (2.5) to (2.6).

2.2.3 Projection error within least-squares Monte Carlo estimation

This subsection is devoted to the analysis of the projection error that occurs when applying the first step, see (2.5), of least-squares Monte Carlo estimation on (2.3). Since we are located in the setting of Lemor et al. (2006), the below stated result is of course part of their error analysis. However, Lemor et al. (2006) examine only the overall approximation error between a truncated version of the time-discretized solution and the simulated solution and the impact of the projection error is only mentioned in passing. In order to distinguish different error sources, we provide Lemma 5. To this end, we define for all $i = 0, \dots, N-1$ function bases

$$\eta_0(i, x) := \{\eta_{0,1}(i, x), \dots, \eta_{0,K_{0,i}}(i, x)\}$$

for the estimation of $y_i^\pi(x)$ and

$$\eta_d(i, x) := \{\eta_{d,1}(i, x), \dots, \eta_{d,K_{d,i}}(i, x)\}, \quad d = 1, \dots, D$$

for the estimation of the dth component of $z_i^\pi(x)$. Here, $K_{d,i}$ stands for the dimension of the function basis for $d = 0, \dots, D$ at time t_i. In particular, we can select in each time step and for each of the $D+1$ estimation tasks a different basis. However, many numerical examples for least-squares Monte Carlo are based on an identical basis for the estimation of all conditional expectations within the same time step. Later on we will show, how the estimation can benefit from different bases.

For the sake of clarity we denote by $\mathcal{P}_{d,i}$, $d = 0, \dots, D$, $i = 0, \dots N-1$ the operator such that for some F_T-measurable random variable U

$$\mathcal{P}_{d,i}(U) := \eta_d(i, X_{t_i}^\pi)\alpha_{d,i}$$

with

$$\alpha_{d,i} = \arg \min_{\alpha \in \mathbb{R}^{K_{d,i}}} E|\eta_d(i, X_{t_i}^\pi)\alpha - U|^2.$$

In other words $\mathcal{P}_{d,i}$ carries out an orthogonal projection on the subspace spanned by $\eta_d(i,x)$, $d = 0, \dots, D$. Replacing the conditional expectations in (2.3) by the projection operator yields then the following algorithm:

$$\begin{aligned} Y_{t_N}^{\pi,K_{0,N}} &= \phi^\pi(X_{t_N}^\pi), \\ Z_{d,t_i}^{\pi,K_{d,i}} &= \frac{1}{\Delta_i}\mathcal{P}_{d,i}\big(\Delta W_{d,i} Y_{t_{i+1}}^{\pi,K_{0,i+1}}\big), \quad d = 1, \dots, D, \\ Y_{t_i}^{\pi,K_{0,i}} &= \mathcal{P}_{0,i}\big(Y_{t_{i+1}}^{\pi,K_{0,i+1}} - \Delta_i F(t_i, S_{t_i}^\pi, Y_{t_{i+1}}^{\pi,K_{0,i+1}}, Z_{t_i}^{\pi,K_i})\big), \end{aligned} \tag{2.8}$$

where $Z_{t_i}^{\pi,K_i} = (Z_{d,t_i}^{\pi,K_{d,i}})_{d=1,\dots,D}$. Again, for all $i = 0, \dots, N-1$ there are deterministic functions $y_i^{\pi,K_{0,i}}(x)$ and $z_{d,i}^{\pi,K_{d,i}}(x)$ such that

$$Y_{t_{i+1}}^{\pi,K_{0,i+1}} = y_i^{\pi,K_{0,i}}(X_{t_i}^\pi), \quad Z_{d,t_i}^{\pi,K_{d,i}} = z_{d,i}^{\pi,K_{d,i}}(X_{t_i}^\pi), \quad d = 1, \dots, D. \tag{2.9}$$

In view of the definition of $\mathcal{P}_{d,i}$, $d = 0, \dots, D$, these functions can be written as linear combinations of $\eta_d(i,x)$, respectively.

Lemma 5. *Let* F *be Lipschitz continuous in its spatial variables* (y,z) *with constant* κ*. Then*

$$\begin{aligned} &\max_{j \leqslant i \leqslant N} E|Y_{t_i}^\pi - Y_{t_i}^{\pi,K_{0,i}}|^2 + \sum_{i=j}^{N-1} \Delta_i E|Z_{t_i}^\pi - Z_{t_i}^{\pi,K_i}|^2 \\ &\quad < C \sum_{i=j}^{N-1} E|\mathcal{P}_{0,i}(Y_{t_i}^\pi) - Y_{t_i}^\pi|^2 + C \sum_{i=j}^{N-1} \sum_{d=1}^{D} \Delta_i E|\mathcal{P}_{d,i}(Z_{d,t_i}^\pi) - Z_{d,t_i}^\pi|^2, \end{aligned}$$

for $j = 0, \dots, N-1$ *with* $C > 0$ *being a constant depending on* κ, T *and* D.

Gobet et al. (2005) provide an analysis of the projection error in a setting that combines least-squares Monte Carlo with Picard iterations in each time step.

Proof. We set $\Delta F_i = F(t_i, S^\pi_{t_i}, Y^{\pi,K_{0,i+1}}_{t_{i+1}}, Z^{\pi,K_i}_{t_i}) - F(t_i, S^\pi_{t_i}, Y^\pi_{t_{i+1}}, Z^\pi_{t_i})$ and exploit the Lipschitz condition on F and Young's inequality for some Γ to be defined later on. Hence,

$$E|\Delta F_i|^2 \leqslant \kappa^2(1+\Gamma D)E\Big[|Y^\pi_{t_{i+1}} - Y^{\pi,K_{0,i+1}}_{t_{i+1}}|^2 + \frac{1}{\Gamma D}|Z^\pi_{t_i} - Z^{\pi,K_i}_{t_i}|^2\Big]. \tag{2.10}$$

Then we define

$$\overline{Y}^\pi_{t_i} = E[Y^{\pi,K_{0,i+1}}_{t_{i+1}} - \Delta_i F(t_i, S^\pi_{t_i}, Y^{\pi,K_{0,i+1}}_{t_{i+1}}, Z^{\pi,K_i}_{t_i})| X^\pi_{t_i}].$$

and apply again Young's inequality. Due to (2.10), we obtain for $\Gamma = 1$

$$\begin{aligned} E|Y^\pi_{t_i} - \overline{Y}^\pi_{t_i}|^2 \leqslant & (1+(1+D)\kappa^2\Delta_i)E|E[Y^\pi_{t_{i+1}} - Y^{\pi,K_{0,i+1}}_{t_{i+1}}| X^\pi_{t_i}]|^2 \\ & + \Big(\Delta_i + \frac{1}{(1+D)\kappa^2}\Big)\Delta_i\kappa^2(1+D)E|Y^\pi_{t_{i+1}} - Y^{\pi,K_{0,i+1}}_{t_{i+1}}|^2 \\ & + \Big(\Delta_i + \frac{1}{(1+D)\kappa^2}\Big)\Delta_i\kappa^2\frac{1+D}{D}E|Z^\pi_{t_i} - Z^{\pi,K_i}_{t_i}|^2. \end{aligned} \tag{2.11}$$

Using the orthogonality of the projection $\mathcal{P}_{d,i}$ we receive

$$\begin{aligned} E|Z^\pi_{d,t_i} - Z^{\pi,K_{d,i}}_{d,t_i}|^2 &= E|Z^\pi_{d,t_i} - \mathcal{P}_{d,i}\left(Z^\pi_{d,t_i}\right)|^2 + E|\mathcal{P}_{d,i}\left(Z^\pi_{d,t_i}\right) - Z^{\pi,K_{d,i}}_{d,t_i}|^2 \\ &= (I) + (II). \end{aligned} \tag{2.12}$$

As for (II), the definition of Z^π_{d,t_i} and $Z^{\pi,K_{d,i}}_{d,t_i}$ in (2.3) and (2.8) yields

$$\begin{aligned} (II) &= E|\mathcal{P}_{d,i}\big(\Delta_i^{-1}E[\Delta W_{d,i}(Y^\pi_{t_{i+1}} - Y^{\pi,K_{0,i+1}}_{t_{i+1}})| X^\pi_{t_i}]\big)|^2 \\ &\leqslant E|\Delta_i^{-1}E[\Delta W_{d,i}\{Y^\pi_{t_{i+1}} - Y^{\pi,K_{0,i+1}}_{t_{i+1}} - E[Y^\pi_{t_{i+1}} - Y^{\pi,K_{0,i+1}}_{t_{i+1}}| X^\pi_{t_i}]\}| X^\pi_{t_i}]|^2 \\ &\leqslant \Delta_i^{-1}E|Y^\pi_{t_{i+1}} - Y^{\pi,K_{0,i+1}}_{t_{i+1}} - E[Y^\pi_{t_{i+1}} - Y^{\pi,K_{0,i+1}}_{t_{i+1}}| X^\pi_{t_i}]|^2 \\ &\leqslant \Delta_i^{-1}\big(E|Y^\pi_{t_{i+1}} - Y^{\pi,K_{0,i+1}}_{t_{i+1}}|^2 - E|E[Y^\pi_{t_{i+1}} - Y^{\pi,K_{0,i+1}}_{t_{i+1}}| X^\pi_{t_i}]|^2\big), \end{aligned} \tag{2.13}$$

where the second step followed by the contraction property of the projection $\mathcal{P}_{d,i}$ and the third step by Hölder's inequality. Now we define a sequence $(q_i)_{i\in\mathbb{N}}$ with $q_0 = 1$ and $q_{i+1} = q_i(1+(1+D)\kappa^2\Delta_i)(1+\Delta_i)$. Turning back to (2.11), we first exploit the estimates on the Z-part and multiply then with q_i. Thus, for $i < N-1$,

$$\begin{aligned} q_i E|Y^\pi_{t_i} - \overline{Y}^\pi_{t_i}|^2 \leqslant & \, q_i(1+(1+D)\kappa^2\Delta_i)(1+\Delta_i)E|Y^\pi_{t_{i+1}} - Y^{\pi,K_{0,i+1}}_{t_{i+1}}|^2 \\ & + q_i(1+(1+D)\kappa^2\Delta_i)\frac{\Delta_i}{D}\sum_{d=1}^{D} E|\mathcal{P}_{d,i}(Z^\pi_{d,t_i}) - Z^{\pi,K_{d,i}}_{d,t_i}|^2 \\ \leqslant & \, q_{i+1}E|Y^\pi_{t_{i+1}} - \overline{Y}^\pi_{t_{i+1}}|^2 + q_{i+1}E|P_{0,i+1}(Y^\pi_{t_{i+1}}) - Y^\pi_{t_{i+1}}|^2 \\ & + q_{i+1}\frac{\Delta_i}{D}\sum_{d=1}^{D} E|\mathcal{P}_{d,i}(Z^\pi_{d,t_i}) - Z^{\pi,K_{d,i}}_{d,t_i}|^2, \end{aligned}$$

where we incorporated the relation

$$\mathcal{P}_{0,i}\left(\bar{Y}^{\pi}_{t_i} - Y^{\pi}_{t_i}\right) = Y^{\pi,K_{0,i}}_{t_i} - \mathcal{P}_{0,i}(Y^{\pi}_{t_i}) \tag{2.14}$$

as well as the orthogonality and the contraction property of the projection $\mathcal{P}_{0,i}$. In case $i = N-1$ we have

$$q_{N-1}E|Y^{\pi}_{t_{N-1}} - \overline{Y}^{\pi}_{t_{N-1}}|^2 \leqslant q_N \frac{\Delta_{N-1}}{D} \sum_{d=1}^{D} E|\mathcal{P}_{d,N-1}(Z^{\pi}_{d,t_{N-1}}) - Z^{\pi,K_{d,N-1}}_{d,t_{N-1}}|^2,$$

since $Y^{\pi}_{t_N} = Y^{\pi,K_{0,N}}_{t_N}$. Taking the sum from i to $N-1$ leads to

$$\begin{aligned} &q_i E|Y^{\pi}_{t_i} - \overline{Y}^{\pi}_{t_i}|^2 \\ &\quad\leqslant \sum_{j=i+1}^{N-1} q_j E|P_{0,j}(Y^{\pi}_{t_j}) - Y^{\pi}_{t_j}|^2 + \sum_{j=i}^{N-1} q_{j+1}\frac{\Delta_j}{D}\sum_{d=1}^{D} E|\mathcal{P}_{d,j}(Z^{\pi}_{d,t_j}) - Z^{\pi,K_{d,j}}_{d,t_j}|^2. \end{aligned}$$

As $\Delta_i < |\pi| < CT/N$ for some $C > 0$, we can conclude

$$q_N < \left(1 + \frac{(1+D)\kappa^2 CT}{N}\right)^N \left(1 + \frac{CT}{N}\right)^N \xrightarrow[N\to\infty]{} e^{CT(1+(1+D)\kappa^2)}.$$

Hence,

$$\begin{aligned} \max_{j\leqslant i\leqslant N-1} E|Y^{\pi}_{t_i} - \overline{Y}^{\pi}_{t_i}|^2 &\leqslant e^{CT(1+(1+D)\kappa^2)} \sum_{i=j+1}^{N-1} E|\mathcal{P}_{0,i}(Y^{\pi}_{t_i}) - Y^{\pi}_{t_i}|^2 \\ &\quad + e^{CT(1+(1+D)\kappa^2)} \sum_{i=j}^{N-1} \Delta_i \sum_{d=1}^{D} E|\mathcal{P}_{d,i}(Z^{\pi}_{d,t_i}) - Z^{\pi,K_{d,i}}_{d,t_i}|^2. \end{aligned} \tag{2.15}$$

In view of (2.14) and by exploiting the orthogonality of the projections, we receive immediately

$$\begin{aligned} &\max_{j\leqslant i\leqslant N-1} E|Y^{\pi}_{t_i} - Y^{\pi,K_{0,i}}_{t_i}|^2 \\ &\quad\leqslant 2\max_{j\leqslant i\leqslant N-1} E|Y^{\pi}_{t_i} - \mathcal{P}_{0,i}(Y^{\pi}_{t_i})|^2 + 2\max_{j\leqslant i\leqslant N-1} E|\mathcal{P}_{0,i}(Y^{\pi}_{t_i} - \overline{Y}^{\pi}_{t_i})|^2 \\ &\quad\leqslant C\Big(\sum_{i=j}^{N-1} E|\mathcal{P}_{0,i}(Y^{\pi}_{t_i}) - Y^{\pi}_{t_i}|^2 + \sum_{i=j}^{N-1}\sum_{d=1}^{D} \Delta_i E|\mathcal{P}_{d,i}(Z^{\pi}_{d,t_i}) - Z^{\pi,K_{d,i}}_{d,t_i}|^2\Big). \end{aligned}$$

Coming back to the estimates in (2.12) and (2.13), we apply the definition of $\overline{Y}^{\pi}_{t_i}$ and the orthogonality of the projections. Clearly, we have for $i = 0, \dots, N-2$

$$\begin{aligned}
&\Delta_i E|Z^\pi_{d,t_i} - Z^{\pi,K_{d,i}}_{d,t_i}|^2 \\
&\quad \leqslant E|Y^\pi_{t_{i+1}} - \overline{Y}^\pi_{t_{i+1}}|^2 + E|\mathcal{P}_{0,i+1}(Y^\pi_{t_{i+1}}) - Y^\pi_{t_{i+1}}|^2 - E|Y^\pi_{t_i} - \overline{Y}^\pi_{t_i} + \Delta_i \Delta F_i|^2 \\
&\qquad + E|\mathcal{P}_{d,i}(Z^\pi_{d,t_i}) - Z^{\pi,K_{d,i}}_{d,t_i}|^2 \\
&\quad \leqslant E|Y^\pi_{t_{i+1}} - \overline{Y}^\pi_{t_{i+1}}|^2 - E|Y^\pi_{t_i} - \overline{Y}^\pi_{t_i}|^2 + 2\Delta_i E|(Y^\pi_{t_i} - \overline{Y}^\pi_{t_i})\Delta F_i|^2 \\
&\qquad + E|\mathcal{P}_{0,i+1}(Y^\pi_{t_{i+1}}) - Y^\pi_{t_{i+1}}|^2 + \Delta_i E|\mathcal{P}_{d,i}(Z^\pi_{d,t_i}) - Z^{\pi,K_{d,i}}_{d,t_i}|^2 \\
&\quad \leqslant E|Y^\pi_{t_{i+1}} - \overline{Y}^\pi_{t_{i+1}}|^2 - E|Y^\pi_{t_i} - \overline{Y}^\pi_{t_i}|^2 + \gamma\Delta_i E|Y^\pi_{t_i} - \overline{Y}^\pi_{t_i}|^2 + \frac{\Delta_i}{\gamma} E|\Delta F_i|^2 \\
&\qquad + E|\mathcal{P}_{0,i+1}(Y^\pi_{t_{i+1}}) - Y^\pi_{t_{i+1}}|^2 + \Delta_i E|\mathcal{P}_{d,i}(Z^\pi_{d,t_i}) - Z^{\pi,K_{d,i}}_{d,t_i}|^2,
\end{aligned}$$

for some $\gamma > 0$. Now we apply (2.10) with $\Gamma = 2$ and consider also relation (2.14). Thus,

$$\begin{aligned}
\Delta_i E|Z^\pi_{d,t_i} - Z^{\pi,K_{d,i}}_{d,t_i}|^2 &\leqslant E|Y^\pi_{t_{i+1}} - \overline{Y}^\pi_{t_{i+1}}|^2 - E|Y^\pi_{t_i} - \overline{Y}^\pi_{t_i}|^2 + \gamma\Delta_i E|Y^\pi_{t_i} - \overline{Y}^\pi_{t_i}|^2 \\
&\quad + \frac{\kappa^2(1+2D)\Delta_i}{\gamma} E\Big[|\mathcal{P}_{0,i+1}(Y^\pi_{t_{i+1}} - \overline{Y}^\pi_{t_{i+1}})|^2 + \frac{1}{2D}|Z^\pi_{t_i} - Z^{\pi,K_i}_{t_i}|^2\Big] \\
&\quad + \Big(1 + \frac{\kappa^2(1+2D)\Delta_i}{2D\gamma}\Big) E|\mathcal{P}_{0,i+1}(Y^\pi_{t_{i+1}}) - Y^\pi_{t_{i+1}}|^2 \\
&\quad + \Delta_i E|\mathcal{P}_{d,i}(Z^\pi_{d,t_i}) - Z^{\pi,K_{d,i}}_{d,t_i}|^2.
\end{aligned} \tag{2.16}$$

Concerning $E|\mathcal{P}_{0,i+1}(Y^\pi_{t_{i+1}} - \overline{Y}^\pi_{t_{i+1}})|^2$, we will make use of the contraction property of the projections. Then, we set $\gamma = \kappa^2(1+2D)$ and define a second sequence $(\tilde{q}_i)_{i\in\mathbb{N}}$ with $\tilde{q}_0 = 1$ and $\tilde{q}_{i+1} = \tilde{q}_i(1+\Delta_i)$. Multiplying (2.16) with $\tilde{q}_i$ and summing up from $d = 1, \dots, D$ and $i = 0, \dots, N-1$ yields

$$\begin{aligned}
&\sum_{i=j}^{N-1} \tilde{q}_i \Delta_i E|Z^\pi_{t_i} - Z^{\pi,K_i}_{t_i}|^2 \\
&\quad \leqslant D\kappa^2(1+2D)e^{CT} \max_{j\leqslant i\leqslant N} \tilde{q}_i E|Y^\pi_{t_i} - \overline{Y}^\pi_{t_i}|^2 + \frac{1}{2}\sum_{i=0}^{N-1} \tilde{q}_i \Delta_i E|Z^\pi_{t_i} - Z^{\pi,K_i}_{t_i}|^2 \\
&\qquad + C\sum_{i=j}^{N-1}\Big(E|\mathcal{P}_{0,i}(Y^\pi_{t_i}) - Y^\pi_{t_i}|^2 + \sum_{d=1}^{D} \Delta_i E|\mathcal{P}_{d,i}(Z^\pi_{d,t_i}) - Z^{\pi,K_{d,i}}_{d,t_i}|^2\Big).
\end{aligned}$$

In view of (2.15) and the definition of $\tilde{q}_i$, it holds true that

$$\begin{aligned}
&\sum_{i=j}^{N-1} \Delta_i E|Z^\pi_{t_i} - Z^{\pi,K_i}_{t_i}|^2 \\
&\quad \leqslant C\Big(\sum_{i=j}^{N-1} E|\mathcal{P}_{0,i}(Y^\pi_{t_i}) - Y^\pi_{t_i}|^2 + \sum_{i=j}^{N-1}\sum_{d=1}^{D} \Delta_i E|\mathcal{P}_{d,i}(Z^\pi_{d,t_i}) - Z^{\pi,K_{d,i}}_{d,t_i}|^2\Big). \qquad \square
\end{aligned}$$

2.2.4 Simulation error within least-squares Monte Carlo estimation

In this subsection we will review the proposal of Lemor et al. (2006) how to use the simulation step of least-squares Monte Carlo, see 2.6, to get a fully implementable algorithm for the approximation of BSDEs. The result of Lemor et al. (2006) considering the simulation error will be dicussed as well. Looking back in the last subsection, we received approximate solutions for $Y^\pi_{t_i}$ and $Z^\pi_{t_i}$ by replacing conditional expectations by projections on subspaces of $L^2(\mathcal{F}_{t_i})$. Clearly, we obtained functions

$$y_i^{\pi,K_{0,i}}(x) = \eta_0(i,x)\alpha_{0,i}^{\pi,K_{0,i}}, \quad z_{d,i}^{\pi,K_{d,i}}(x) = \eta_d(i,x)\alpha_{d,i}^{\pi,K_{d,i}}, \quad d = 1,\ldots,D,$$

where $\alpha_{0,i}^{\pi,K_{0,i}}$ and $\alpha_{d,i}^{\pi,K_{d,i}}$, $d = 1,\ldots,D$ are solutions of minimization problems of the form (2.5). The application of least-squares Monte Carlo implies to substitute $\alpha_{d,i}^{\pi,K_{d,i}}$, $d = 0,\ldots,D$ by coefficients that solve minimization problems of type (2.6). To this end, we introduce L independent copies of $(\Delta W_i, X^\pi_{t_{i+1}})_{i=0,\ldots,N-1}$. We denote these samples by $(\Delta_\lambda W_i, {}_\lambda X^\pi_{t_{i+1}})_{i=0,\ldots,N-1}$, $\lambda = 1,\ldots,L$ and by $\mathcal{X}^L$ the set that contains these samples. The least-squares Monte Carlo approximations $y_i^{\pi,K_{0,i},L}(x)$ and $z_{d,i}^{\pi,K_{d,i},L}(x)$, $d = 1,\ldots,D$ are evaluated by carrying out for $i = N-1,\ldots,0$:

$$\begin{aligned}
&y_N^{\pi,K_{0,N},L}(x) = \phi^\pi(x),\\
&\quad \alpha_{d,i}^{\pi,K_{d,i},L} = \arg\min_{\alpha\in\mathbb{R}^{K_{d,i}}} \frac{1}{L}\sum_{\lambda=1}^{L}\left|\eta_d(i,{}_\lambda X^\pi_{t_i})\alpha - \frac{\Delta_\lambda W_{d,i}}{\Delta_i} y_{i+1}^{\pi,K_{0,i+1},L}({}_\lambda X^\pi_{t_{i+1}})\right|^2,\\
&\hspace{20em} d = 1,\ldots,D,\\
&z_{d,i}^{\pi,K_{d,i},L}(x) = \eta_d(i,x)\alpha_{d,i}^{\pi,K_{d,i},L}, \quad d = 1,\ldots,D,\\
&\quad \alpha_{0,i}^{\pi,K_{0,i},L} = \arg\min_{\alpha\in\mathbb{R}^{K_{0,i}}} \frac{1}{L}\sum_{\lambda=1}^{L}\Big|\eta_0(i,{}_\lambda X^\pi_{t_i})\alpha - y_{i+1}^{\pi,K_{0,i+1},L}({}_\lambda X^\pi_{t_{i+1}})\\
&\hspace{10em} + \Delta_i F\big(t_i,{}_\lambda S^\pi_{t_i}, y_{i+1}^{\pi,K_{0,i+1},L}({}_\lambda X^\pi_{t_{i+1}}), z_i^{\pi,K_i,L}({}_\lambda X^\pi_{t_i})\big)\Big|^2,\\
&y_i^{\pi,K_{0,i},L}(x) = \eta_0(i,x)\alpha_{0,i}^{\pi,K_{0,i},L},
\end{aligned} \tag{2.17}$$

where $z_i^{\pi,K_i,L}(x) = (z_{d,i}^{\pi,K_{d,i},L}(x))_{d=1,\ldots,D}$. Setting

$$Y_{t_i}^{\pi,K_{0,i},L} = y_i^{\pi,K_{0,i},L}(X^\pi_{t_i}), \quad Z_{d,t_i}^{\pi,K_{d,i},L} = z_{d,i}^{\pi,K_{d,i},L}(X^\pi_{t_i}), \quad d = 1,\ldots,D$$

gives then the least-squares Monte Carlo estimators for $(Y^\pi_{t_i}, Z^\pi_{t_i})_{t_i\in\pi}$. The analysis of the L^2-error induced by the simulation step of least-squares Monte Carlo can be found in Lemor et al. (2006), Theorem 2 and Remark 2. It is rather involved, since the approximation error has to be traced back to the error related to the law of

$(\Delta_\lambda W_i, {}_\lambda X^\pi_{t_{i+1}})_{i=1,\dots,N-1}$, $\lambda = 1, \dots, L$, namely

$$\max_{0\leqslant i\leqslant N} E\left[\frac{1}{L}\sum_{\lambda=1}^{L}|y_i^\pi({}_\lambda X^\pi_{t_i}) - y_i^{\pi,K_{0,i},L}({}_\lambda X^\pi_{t_i})|^2\right] + \sum_{d=1}^{D}\sum_{i=0}^{N-1}\Delta_i E\left[\frac{1}{L}\sum_{\lambda=1}^{L}|z^\pi_{d,i}({}_\lambda X^\pi_{t_i}) - z_{d,i}^{\pi,K_{d,i},L}({}_\lambda X^\pi_{t_i})|^2\right].$$

Recall, that $y_i^{\pi,K_{0,i},L}(x)$ and $z_{d,i}^{\pi,K_{d,i},L}(x)$ are estimated via the samples of future time steps. Hence, one has to deal with a quite complicated dependency structure between the approximators in the different time steps.

What is more, the examination of this error requires the implementation of a truncation structure in the pure backward scheme (2.3) (which is based on the assumption of computable conditional expectations) and in the least-squares Monte Carlo algorithm (2.17). The aim is to receive a Lipschitz continuous, bounded estimation of $y_i^\pi(x)$ and $z^\pi_{d,i}(x)$ on the one hand and a bounded estimation of $y_i^{\pi,K_{0,i},L}(x)$ and $z_{d,i}^{\pi,K_{d,i},L}(x)$ on the other one. The Lipschitz continuity requires certain additional assumptions on the approximate terminal condition $\phi^\pi(x)$ and the Markov process $(X^\pi_{t_i})_{t_i\in\pi}$. As the truncation is generally omitted in practice, we refrain from stating detailed information on the truncation error and refer the reader to Lemor et al. (2006).

Neglecting the truncation error, the squared approximation error is bounded as follows, see Lemor et al. (2006). Given an equidistant partition of $[0, T]$ with $\Delta_i = h := T/N, i = 0, \dots, N-1$ and $\beta \in (0, 1]$ we have

$$\begin{aligned}
&\max_{0\leqslant i\leqslant N} E|Y^\pi_{t_i} - Y_{t_i}^{\pi,K_{0,i},L}|^2 + \sum_{d=1}^{D}\sum_{i=0}^{N-1}\Delta_i E|Z^\pi_{d,t_i} - Z_{d,t_i}^{\pi,K_{d,i},L}|^2 \\
&\leqslant \tilde{C}h^\beta + \tilde{C}\Bigg(\frac{\log(L)}{L}\sum_{i=0}^{N-1}\sum_{d=0}^{D}K_{i,d} \\
&\quad + \sum_{i=0}^{N-1}\frac{K_{0,i}}{h}\exp\left\{\tilde{C}K_{0,i+1}\log\frac{\tilde{C}\sqrt{K_{0,i}}}{h^{\frac{\beta+2}{2}}} - \frac{Lh^{\beta+2}}{72\tilde{C}K_{0,i}}\right\} \\
&\quad + \sum_{i=0}^{N-1}\sum_{d=1}^{D}K_{d,i}\exp\left\{\tilde{C}K_{0,i+1}\log\frac{\tilde{C}\sqrt{K_{d,i}}}{h^{\frac{\beta+1}{2}}} - \frac{Lh^{\beta+1}}{72\tilde{C}K_{d,i}}\right\} \\
&\quad + \sum_{i=0}^{N-1}\frac{1}{h}\exp\left\{\tilde{C}K_{0,i}\log\frac{\tilde{C}}{h^{\frac{\beta+2}{2}}} - \frac{Lh^{\beta+2}}{72\tilde{C}}\right\}\Bigg) \\
&\quad + \tilde{C}\left(\sum_{i=0}^{N-1}E|\mathcal{P}_{0,i}(Y^\pi_{t_i}) - Y^\pi_{t_i}|^2 + \sum_{i=0}^{N-1}\sum_{d=1}^{D}\Delta_i E|\mathcal{P}_{d,i}(Z^\pi_{d,t_i}) - Z^\pi_{d,t_i}|^2\right) \\
&\quad + \text{truncation error},
\end{aligned} \tag{2.18}$$

where $\tilde{C}$ is a constant depending on the Lipschitz constant κ, T, s_0, the dimensions $\widetilde{D}$ and D as well as the truncation parameters. Particularly, the first and the second summand mark the additional error terms that arise from the simulation step in least-squares Monte Carlo.

2.2.5 Qualitative analysis of the error sources and their configuration

When neglecting the implementation of truncations, the approximation error of least-squares Monte Carlo is driven by three main error sources, the time discretization error, the projection error and the simulation error. In the following we give a short qualitative recapitulation of the previous subsections. Moreover, we describe what it takes to bound all error sources by $C|\pi|^{\beta/2}$ in L^2-sense for $\beta \in (0,1]$.

- The squared **time discretization error** is bounded by

  $$C(|\pi| + E|\xi - \xi^{\pi}|^2).$$

  Hence, it is enough to suppose that the L^2-error regarding the terminal condition decreases with order $\beta/2$ in the number of time steps. For instance this case is fulfilled if the terminal condition can be expressed via some Lipschitz-continuous function ϕ such that $\xi = \phi(S_T)$ and $\xi^{\pi} = \phi(S^{\pi}_{t_N})$ and the L^2-error between S_{t_i} and its approximation $S^{\pi}_{t_i}$ decreases with rate $|\pi|^{\beta/2}$.

- The squared **projection error** is determined by the chosen function bases and is bounded by terms of the squared L^2-distance between the time-discretized solution $(Y^{\pi}_{t_i}, \sqrt{h}Z^{\pi}_{t_i})$ and its best projections on the function bases. Precisely, the squared error is bounded by a constant times

  $$\sum_{i=0}^{N-1} E|\mathcal{P}_{0,i}(Y^{\pi}_{t_i}) - Y^{\pi}_{t_i}|^2 + \sum_{i=0}^{N-1}\sum_{d=1}^{D} \Delta_i E|\mathcal{P}_{d,i}(Z^{\pi}_{d,t_i}) - Z^{\pi}_{d,t_i}|^2. \tag{2.19}$$

  Note that $(Y^{\pi}_{t_i}, Z^{\pi}_{t_i})_{t_i \in \pi}$ is based on an evaluation of nested conditional expectations. Thus, the errors due to the estimation of conditional expectations propagate and the approximation error of $Y^{\pi,K_{0,i},L}_{t_i}$ and $Z^{\pi,K_{d,i},L}_{d,t_i}$, $d = 1,\dots,D$ is influenced by all previous projection errors. Consequently, (2.19) contains the sum over all L^2-distances between $(Y^{\pi}_{t_i}, \sqrt{\Delta_i}Z^{\pi}_{t_i})$ and its best projection for $i = 0,\dots,N-1$.
  Both the time-discretized solution and its projection, are unknown. Hence, these error terms cannot be quantified in general. An exception to this are indicator functions related to hypercubes, that form a partition of the state space of $X^{\pi}_{t_i}$. For this case Gobet et al. (2005) have shown that each of the summands of (2.19) is bounded by $C\delta^2$ for all $i = 0,\dots N-1$, $d = 0,\dots,D$, where δ denotes the edge length of the hypercubes. Setting $\delta = (T/N)^{(\beta+1)/2}$ yields the desired convergence rate. Then, the dimension of the function bases $K_{d,i}$ grows proportional to $N^{D(\beta+1)/2}$ for all $d = 0,\dots,D$ and $i = 0,\dots,N-1$.

- The squared **simulation error** causes the additional terms

$$
\begin{aligned}
&\tilde{C}|\pi|^{\beta} + \tilde{C}\Bigg(\frac{\log(L)}{L}\sum_{i=0}^{N-1}\sum_{d=0}^{D} K_{i,d} \\
&\quad + \sum_{i=0}^{N-1}\frac{K_{0,i}}{h}\exp\left\{\tilde{C}K_{0,i+1}\log\frac{\tilde{C}\sqrt{K_{0,i}}}{h^{\frac{\beta+2}{2}}} - \frac{Lh^{\beta+2}}{72\tilde{C}K_{0,i}}\right\} \\
&\quad + \sum_{i=0}^{N-1}\sum_{d=1}^{D} K_{d,i}\exp\left\{\tilde{C}K_{0,i+1}\log\frac{\tilde{C}\sqrt{K_{d,i}}}{h^{\frac{\beta+1}{2}}} - \frac{Lh^{\beta+1}}{72\tilde{C}K_{d,i}}\right\} \\
&\quad + \sum_{i=0}^{N-1}\frac{1}{h}\exp\left\{\tilde{C}K_{0,i}\log\frac{\tilde{C}}{h^{\frac{\beta+2}{2}}} - \frac{Lh^{\beta+2}}{72\tilde{C}}\right\}\Bigg)
\end{aligned}
\tag{2.20}
$$

  in the upper bound on the squared approximation error, see (2.18). Given an appropriate choice of $K_{d,i}$, $d = 0, \dots, D$, $i = 0, \dots, N-1$ and L it can be designed to grow with order β in the number of time steps N. To this end we fix the dimension of the function bases $K_{d,i}$ by $\tilde{C}N^{\rho}$ for some $\rho \geqslant 0$ and the sample size L by $\tilde{C}N^{\beta+2+2\rho}$ for some constant $\tilde{C} \geqslant 0$. Here, the logarithmic terms were neglected.

3 Error criteria for BSDEs

3.1 Global a-posteriori error criterion

As the true approximation error cannot be evaluated, the success of a numeric solution of a BSDE is often judged by the approximation of Y_0, see for instance Bender and Denk (2007). Precisely, an approximation $(\hat{Y}^\pi_{t_i}, \hat{Z}^\pi_{t_i})_{t_i \in \pi}$ is supposed to be successful, if for a finer getting time grid π the approximate initial value $\hat{Y}^\pi_{t_0}$ stabilizes, i. e. converges to some value for $|\pi| \to 0$. There are two major problems connected with this procedure. First, in most cases the true Y_0 is not available in closed form. Hence, as $\hat{Y}^\pi_{t_0}$ is a point estimator, it might converge to a biased initial value.

Second, this method provides no statement on the quality of the approximation of the entire paths Y and Z. However, this information is highly interesting, e. g. in financial settings, where the hedging portfolio can be expressed in terms of Z. Inspired by the identity

$$Y_{t_{i+1}} - Y_{t_i} - \int_{t_i}^{t_{i+1}} f(t_i, Y_t, Z_t)dt - \int_{t_i}^{t_{i+1}} Z_t dW_t = 0$$

we argue that a successful approximation $(\hat{Y}^\pi_{t_i}, \hat{Z}^\pi_{t_i})_{t_i \in \pi}$ should satisfy

$$\hat{Y}^\pi_{t_{i+1}} - \hat{Y}^\pi_{t_i} - \Delta_i f^\pi(t_i, \hat{Y}^\pi_{t_i}, \hat{Z}^\pi_{t_i}) - \hat{Z}^\pi_{t_i} \Delta W_i \approx 0. \tag{3.1}$$

From these considerations we derive the global a-posteriori error criterion by summing up the left-hand side of (3.1) from $i = 0$ up to $i = j - 1$. Applying the L^2-norm and then taking the maximum over $j = 1, \dots, N$ yields the definition of the global error criterion, see (1.2):

$$\begin{aligned}\mathcal{E}_\pi(\hat{Y}^\pi, \hat{Z}^\pi) := {} & E[|\xi^\pi - \hat{Y}^\pi_{t_N}|^2 | \mathcal{G}_{t_0}] \\ & + \max_{1 \leqslant j \leqslant N} E[|\hat{Y}^\pi_{t_j} - \hat{Y}^\pi_{t_0} - \sum_{i=0}^{j-1} \Delta_i f^\pi(t_i, \hat{Y}^\pi_{t_i}, \hat{Z}^\pi_{t_i}) - \sum_{i=0}^{j-1} \hat{Z}^\pi_{t_i} \Delta W_i|^2 | \mathcal{G}_{t_0}],\end{aligned}$$

where $\mathbb{G} = (\mathcal{G}_t, 0 \leqslant t \leqslant T)$ is an enlarged filtration such that $\mathcal{F}_t \subseteq \mathcal{G}_t$ for $t \in [0, T]$ and $(\hat{Y}^\pi_{t_i}, \hat{Z}^\pi_{t_i})_{t_i \in \pi}$ is $\mathbb{G}$-adapted. This criterion can be interpreted as a necessary condition for the convergence of the approximation error, because it gives information, if the numeric solution is 'close to solving' the BSDE, when considering it as a forward SDE. Therefore, it is interesting in its own right.

However, we require information, if the approximation is close to the true solution, precisely if the approximation error is tending to zero. The main result of this section contains estimates on the L^2-error between the true solution and $(\hat{Y}^\pi_{t_i}, \hat{Z}^\pi_{t_i})_{t_i\in\pi}$ in terms of this global criterion and the L^2-error between true and approximate data. Given certain assumptions on the approximate driver and the approximate terminal condition, these estimates can be extended to an equivalence result between the global a-posteriori criterion on the one hand and the squared approximation error on the other one, up to terms of order 1 in the number of time steps (the usual time discretization error). Hence, the criterion can also be seen as a sufficient condition for the convergence behavior of the approximation error. Moreover, as the a-posteriori criterion only depends on the available approximate solution, we can estimate it consistently by Monte Carlo simulation.

Previous to this we will formulate in a first step the global a-posteriori error criterion for a discrete-time BSDE that is equipped with data (ξ^π, f^π). Mainly based on the Lipschitz continuity of f^π, we can derive an equivalence relation between the error criterion and the approximation error for the discrete-time setting. This result comes along with examples for its application. Next, we consider the solution of the time-discretized BSDE as a time discretization of the original continuous BSDE. The estimates on the approximation error of the continuous BSDE are then easily shown by means of the time discretization error and the equivalence result regarding the global a-posteriori error criterion for time-discretized BSDEs.

Finally, we review typical examples of BSDEs and explain how the estimates on the approximation error look like in these special cases.

3.1.1 Global a-posteriori estimates for discrete-time BSDEs

Before deriving a-posteriori estimates for BSDEs as introduced in (1.1), we first focus on discrete-time BSDEs, that live on the time grid π. In our setting we admit an enlarged filtration $\mathbb{G} = (\mathcal{G}_t, t \geqslant 0)$ such that for some random vector Ξ, that is independent of $\mathbb{F}$, $\mathcal{G}_{t_i} = \mathcal{F}_{t_i} \vee \sigma(\Xi)$ for all $t_i \in \pi$. Recall, that $\mathcal{F}_{t_i}$ is the σ-algebra generated by $(W_t)_{0\leqslant t\leqslant t_i}$. Thus, W is also a Brownian motion with respect to $\mathbb{G}$. The subject of consideration is then

$$\begin{aligned} Y^{\pi,\star}_{t_N} &= \xi^\pi, \\ Y^{\pi,\star}_{t_i} &= Y^{\pi,\star}_{t_{i+1}} - \Delta_i f^\pi(t_i, Y^{\pi,\star}_{t_i}, \Delta_i^{-1} E[(\Delta W_i)^* M^{\pi,\star}_{t_{i+1}} | \mathcal{G}_{t_i}]) - (M^{\pi,\star}_{t_{i+1}} - M^{\pi,\star}_{t_i}), \end{aligned} \tag{3.2}$$

for $i = N-1, \dots, 0$. The solution of (3.2) is formed by a pair of square-integrable, $\mathbb{G}$-adapted processes $(Y^{\pi,\star}_{t_i}, M^{\pi,\star}_{t_i})_{t_i\in\pi}$ such that the process $(M^{\pi,\star}_{t_i})_{t_i\in\pi}$ is a $(\mathcal{G}_{t_i})_{t_i\in\pi}$-martingale starting in 0. Analogously to our continuous-time setting, determined by Assumption 1, we suppose that the data (ξ^π, f^π) fulfill

Assumption 4. *(i) The terminal condition ξ^π is a real valued, square-integrable, $\mathcal{G}_{t_N}$-measurable random variable.*

(ii) The driver is a function $f^\pi : \Omega \times \pi \times \mathbb{R} \times \mathbb{R}^D \to \mathbb{R}$ such that $f^\pi(t_i, y, z)$ is $\mathcal{G}_{t_i}$-measurable for every $(t_i, y, z) \in \pi \times \mathbb{R} \times \mathbb{R}^D$ and $f^\pi(t_i, 0, 0)$ is square-integrable for every $t_i \in \pi$. Furthermore, f^π is Lipschitz continuous in (y, z) with constant κ uniformly in (t_i, ω) and independent of π.

It follows for $i = 0, \ldots, N-1$ that

$$M^{\pi,\star}_{t_{i+1}} - M^{\pi,\star}_{t_i} = Y^{\pi,\star}_{t_{i+1}} - E[Y^{\pi,\star}_{t_{i+1}} | \mathcal{G}_{t_i}]. \tag{3.3}$$

Given $|\pi|$ small enough, the existence of a solution follows by a contraction mapping argument. Considering the relation

$$Z^{\pi,\star}_{t_i} = \frac{1}{\Delta_i} E[(\Delta W_i)^* M^{\pi,\star}_{t_{i+1}} | \mathcal{G}_{t_i}] \tag{3.4}$$

we receive a reformulation of the discrete BSDE studied in Bouchard and Touzi (2004), i.e. for $i = N-1, \ldots, 0$ we have

$$\begin{aligned}
Y^{\pi,\star}_{t_N} &= \xi^\pi, \\
Z^{\pi,\star}_{t_i} &= \frac{1}{\Delta_i} E[(\Delta W_i)^* Y^{\pi,\star}_{t_{i+1}} | \mathcal{G}_{t_i}], \\
Y^{\pi,\star}_{t_i} &= E[Y^{\pi,\star}_{t_{i+1}} | \mathcal{G}_{t_i}] - \Delta_i f^\pi(t_i, Y^{\pi,\star}_{t_i}, Z^{\pi,\star}_{t_i}).
\end{aligned}$$

Now, let $(\hat{Y}^\pi_{t_i}, \hat{M}^\pi_{t_i})_{t_i \in \pi}$ be an arbitrary, square-integrable approximation of the pair $(Y^{\pi,\star}_{t_i}, M^{\pi,\star}_{t_i})_{t_i \in \pi}$, that is $(\mathcal{G}_{t_i})_{t_i \in \pi}$-adapted. At this point the way of approximation does not have to be specified any further. Our aim is to judge the L^2-error between $(Y^{\pi,\star}_{t_i}, M^{\pi,\star}_{t_i})_{t_i \in \pi}$ and $(\hat{Y}^\pi_{t_i}, \hat{M}^\pi_{t_i})_{t_i \in \pi}$ by means of the approximate solution and the data (ξ^π, f^π) only.

As already mentioned above, we want to use for this purpose a criterion that analyzes, if the approximate solution is close to solving (3.2). Hence, we examine

$$\begin{aligned}
\mathcal{E}_\pi(\hat{Y}^\pi, \hat{M}^\pi) := E[|\xi^\pi - \hat{Y}^\pi_{t_N}|^2 | \mathcal{G}_{t_0}] + \max_{1 \leqslant j \leqslant N} E[|\hat{Y}^\pi_{t_j} - \hat{Y}^\pi_{t_0} \\
- \sum_{i=0}^{j-1} \Delta_i f^\pi(t_i, \hat{Y}^\pi_{t_i}, \Delta_i^{-1} E[(\Delta W_i)^* \hat{M}^\pi_{t_{i+1}} | \mathcal{G}_{t_i}]) - \hat{M}^\pi_{t_{i+1}}|^2 | \mathcal{G}_{t_0}].
\end{aligned} \tag{3.5}$$

The next theorem will show that this criterion is equivalent to the squared L^2-error between true solution and approximation.

Theorem 6. *Let Assumption 4 be fulfilled and* $(\hat{Y}^\pi_{t_i}, \hat{M}^\pi_{t_i})_{t_i \in \pi}$ *be a pair of square-integrable,* $(\mathcal{G}_{t_i})_{t_i \in \pi}$*-adapted processes such that* $\hat{M}^\pi$ *is a* $\mathbb{G}$*-martingale starting in 0. Then there are constants* $C, c > 0$ *such that for* $|\pi|$ *small enough*

$$\begin{aligned}
\frac{1}{c} \mathcal{E}_\pi(\hat{Y}^\pi, \hat{M}^\pi) &\leqslant \max_{0 \leqslant i \leqslant N} E[|Y^{\pi,\star}_{t_i} - \hat{Y}^\pi_{t_i}|^2 | \mathcal{G}_{t_0}] + E[|M^{\pi,\star}_{t_N} - \hat{M}^\pi_{t_N}|^2 | \mathcal{G}_{t_0}] \\
&\leqslant C \mathcal{E}_\pi(\hat{Y}^\pi, \hat{M}^\pi).
\end{aligned}$$

More precisely, the inequalities hold with the choice

$$c = 6(1 + \kappa^2 T(T+D)) + 1, \quad C = \left(3 + 8(3 + 4(2T+D)\kappa^2 T)\right) e^{\Gamma T} + 2,$$

where $\Gamma = 4\kappa^2(2T+D)(2 + 4(2T+D)\kappa^2 T)$ *and* $|\pi| < \Gamma^{-1}$.

Proof. The condition on the mesh size $|\pi|$ ensures that a unique solution $(Y^{\pi,\star}, M^{\pi,\star})$ to the discrete BSDE (3.2) exists, see e.g. Theorem 5 and Remark 6 in Bender and Denk (2007). First we show the lower bound

$$\mathcal{E}_\pi(\hat{Y}^\pi, \hat{M}^\pi) \leqslant c \left(\max_{0\leqslant i\leqslant N} E[|Y^{\pi,\star}_{t_i} - \hat{Y}^\pi_{t_i}|^2|\mathcal{G}_{t_0}] + E[|M^{\pi,\star}_{t_N} - \hat{M}^\pi_{t_N}|^2|\mathcal{G}_{t_0}] \right). \tag{3.6}$$

In order to simplify the notation we set

$$\hat{Z}^\pi_{t_i} = \Delta_i^{-1} E[(\Delta W_i)^* \hat{M}^\pi_{t_{i+1}}|\mathcal{G}_{t_i}].$$

Hence,

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

Thanks to the definition in (3.2) and (3.4),

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

Next, we insert this relation in B_i. By applying Young's inequality and the martingale property of $M^{\pi,\star} - \hat{M}^\pi$, we have for every $\gamma > 0$,

$$\begin{aligned}B_i &= E[|\hat{Y}^\pi_{t_i} - Y^{\pi,\star}_{t_i} - \hat{Y}^\pi_{t_0} + Y^{\pi,\star}_{t_0} \\ &\quad - \sum_{j=0}^{i-1} \Delta_j \big(f^\pi(t_j, \hat{Y}^\pi_{t_j}, \hat{Z}^\pi_{t_j}) - f^\pi(t_j, Y^{\pi,\star}_{t_j}, Z^{\pi,\star}_{t_j})\big) - \hat{M}^\pi_{t_i} + M^{\pi,\star}_{t_i}|^2|\mathcal{G}_{t_0}] \\ &\leqslant (1+\gamma)\Big[\frac{5}{4}4 \max_{0\leqslant i\leqslant N} E[|Y^{\pi,\star}_{t_i} - \hat{Y}^\pi_{t_i}|^2|\mathcal{G}_{t_0}] + 5E[|M^{\pi,\star}_{t_N} - \hat{M}^\pi_{t_N}|^2|\mathcal{G}_{t_0}]\Big] \\ &\quad + (1+\gamma^{-1})T \sum_{j=0}^{N-1} \Delta_j E[|f^\pi(t_j, \hat{Y}^\pi_{t_j}, \hat{Z}^\pi_{t_j}) - f^\pi(t_j, Y^{\pi,\star}_{t_j}, Z^{\pi,\star}_{t_j})|^2|\mathcal{G}_{t_0}].\end{aligned}$$

Then we make use of the Lipschitz condition on f^π. Thus,

$$\begin{aligned}B_i &\leqslant 5(1+\gamma)\Big[\max_{0\leqslant i\leqslant N} E[|Y^{\pi,\star}_{t_i} - \hat{Y}^\pi_{t_i}|^2|\mathcal{G}_{t_0}] + E[|M^{\pi,\star}_{t_N} - \hat{M}^\pi_{t_N}|^2|\mathcal{G}_{t_0}]\Big] \\ &\quad + (1+\gamma^{-1})T(T+D)\kappa^2 \\ &\qquad \times \Big[\max_{0\leqslant i\leqslant N-1} E[|Y^{\pi,\star}_{t_i} - \hat{Y}^\pi_{t_i}|^2|\mathcal{G}_{t_0}] + \sum_{j=0}^{N-1} \frac{\Delta_j}{D} E[|Z^{\pi,\star}_{t_j} - \hat{Z}^\pi_{t_j}|^2|\mathcal{G}_{t_0}]\Big].\end{aligned}$$

Due to the definition of $Z^{\pi,\star}$ and $\hat{Z}^\pi$ and the martingale property of $M^{\pi,\star} - \hat{M}^\pi$,

$$\begin{aligned}
&\sum_{j=0}^{N-1} \Delta_j E[|Z^{\pi,\star}_{t_j} - \hat{Z}^\pi_{t_j}|^2 | \mathcal{G}_{t_0}] \\
&\quad = \sum_{j=0}^{N-1} \frac{1}{\Delta_j} E[|E[(\Delta W_j)^* (M^{\pi,\star}_{t_{j+1}} - \hat{M}^\pi_{t_{j+1}} - M^{\pi,\star}_{t_j} + \hat{M}^\pi_{t_j}) | \mathcal{G}_{t_j}]|^2 | \mathcal{G}_{t_0}] \\
&\quad \leqslant D \sum_{j=0}^{N-1} \Big(E[|M^{\pi,\star}_{t_{j+1}} - \hat{M}^\pi_{t_{j+1}}|^2 | \mathcal{G}_{t_0}] - E[|M^{\pi,\star}_{t_j} - \hat{M}^\pi_{t_j}|^2 | \mathcal{G}_{t_0}] \Big) \\
&\quad = D E[|M^{\pi,\star}_{t_N} - \hat{M}^\pi_{t_N}|^2 | \mathcal{G}_{t_0}].
\end{aligned} \tag{3.7}$$

By plugging (3.7) in B_i, we obtain

$$\begin{aligned}
\mathcal{E}_\pi(\hat{Y}^\pi, \hat{M}^\pi) \leqslant \Big(&5(1+\gamma) + T(T+D)\kappa^2(1+\gamma^{-1}) + 1\Big) \\
&\times \left(\max_{0 \leqslant i \leqslant N} E[|Y^{\pi,\star}_{t_i} - \hat{Y}^\pi_{t_i}|^2 | \mathcal{G}_{t_0}] + E[|M^{\pi,\star}_{t_N} - \hat{M}^\pi_{t_N}|^2 | \mathcal{G}_{t_0}] \right).
\end{aligned}$$

Setting $\gamma = T(T+D)\kappa^2$, we receive the lower bound (3.6) with $c = 6(1+\kappa^2 T(T+D))+1$. For the proof of the upper bound we first introduce the process $\overline{Y}^\pi$ by defining for $i = 0, \dots, N-1$

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

where again $\hat{Z}^\pi_{t_i} = \Delta_i^{-1} E[(\Delta W_i)^* \hat{M}^\pi_{t_{i+1}} | \mathcal{G}_{t_i}]$. The pair $(\overline{Y}^\pi, \hat{M}^\pi)$ can also be considered as solution of the discrete BSDE with terminal condition $\overline{\xi}^\pi = \overline{Y}^\pi_{t_N}$ and driver $\overline{f}^\pi(t_i, y, z) = f^\pi(t_i, \hat{Y}^\pi_{t_i}, z)$. We will derive the upper bound by examining the error between $(\overline{Y}^\pi, \hat{M}^\pi)$ and $(Y^{\pi,\star}, M^{\pi,\star})$. To this end we use a slight modification of the weighted a-priori estimates of Lemma 7 in Bender and Denk (2007). Let $\Gamma, \gamma > 0$ be constants to be defined later on and $q_i = \prod_{j=0}^{i-1}(1 + \Gamma \Delta_j)$ the mentioned weights. Due to (3.3) we have

$$M^{\pi,\star}_{t_{i+1}} - M^{\pi,\star}_{t_i} = Y^{\pi,\star}_{t_{i+1}} - E[Y^{\pi,\star}_{t_{i+1}} | \mathcal{G}_{t_i}], \quad \hat{M}^\pi_{t_{i+1}} - \hat{M}^\pi_{t_i} = \overline{Y}^\pi_{t_{i+1}} - E[\overline{Y}^\pi_{t_{i+1}} | \mathcal{G}_{t_i}].$$

Hence,

$$\begin{aligned}
&\sum_{i=0}^{N-1} q_i E[|(M^{\pi,\star}_{t_{i+1}} - M^{\pi,\star}_{t_i}) - (\hat{M}^\pi_{t_{i+1}} - \hat{M}^\pi_{t_i})|^2 | \mathcal{G}_{t_0}] \\
&\quad = \sum_{i=0}^{N-1} q_i E[|Y^{\pi,\star}_{t_{i+1}} - \overline{Y}^\pi_{t_{i+1}} - E[Y^{\pi,\star}_{t_{i+1}} - \overline{Y}^\pi_{t_{i+1}} | \mathcal{G}_{t_i}]|^2 | \mathcal{G}_{t_0}].
\end{aligned}$$

By adapting the argumentation in Step 1 of the proof of Lemma 7 in Bender and Denk (2007) to our setting, we get,

$$\begin{aligned}
&\sum_{i=0}^{N-1} q_i E[|(M^{\pi,\star}_{t_{i+1}} - M^{\pi,\star}_{t_i}) - (\hat{M}^{\pi}_{t_{i+1}} - \hat{M}^{\pi}_{t_i})|^2 | \mathcal{G}_{t_0}] \\
&\quad \leqslant q_N E[|Y^{\pi,\star}_{t_N} - \overline{Y}^{\pi}_{t_N}|^2 | \mathcal{G}_{t_0}] + \gamma \sum_{i=0}^{N-1} q_i \Delta_i E[|Y^{\pi,\star}_{t_i} - \overline{Y}^{\pi}_{t_i}|^2 | \mathcal{G}_{t_0}] \\
&\qquad + \frac{(2T+D)\kappa^2}{\gamma} \sum_{i=0}^{N-1} q_i \Delta_i E\Big[\frac{1}{2T}|Y^{\pi,\star}_{t_i} - \hat{Y}^{\pi}_{t_i}|^2 + \frac{1}{D}|Z^{\pi,\star}_{t_i} - \hat{Z}^{\pi}_{t_i}|^2 \Big| \mathcal{G}_{t_0}\Big].
\end{aligned}$$

The line of argument of Step 2 of the same proof leads to

$$\begin{aligned}
&\max_{0\leqslant i\leqslant N} q_i E[|Y^{\pi,\star}_{t_i} - \overline{Y}^{\pi}_{t_i}|^2 | \mathcal{G}_{t_0}] \leqslant q_N E[|Y^{\pi,\star}_{t_N} - \overline{Y}^{\pi}_{t_N}|^2 | \mathcal{G}_{t_0}] \\
&\quad + \kappa^2(2T+D)(|\pi| + \Gamma^{-1}) \sum_{i=0}^{N-1} q_i \Delta_i E\Big[\frac{1}{2T}|Y^{\pi,\star}_{t_i} - \hat{Y}^{\pi}_{t_i}|^2 + \frac{1}{D}|Z^{\pi,\star}_{t_i} - \hat{Z}^{\pi}_{t_i}|^2 \Big| \mathcal{G}_{t_0}\Big].
\end{aligned}$$

Next, we combine the last two inequalities. For convenience, we abbreviate

$$\begin{aligned}
\tilde{\mathcal{E}}(Y^{\pi,\star} - \overline{Y}^{\pi}, M^{\pi,\star} - \hat{M}^{\pi}) &:= 2 \max_{0\leqslant i\leqslant N} q_i E[|Y^{\pi,\star}_{t_i} - \overline{Y}^{\pi}_{t_i}|^2 | \mathcal{G}_{t_0}] \\
&\quad + \sum_{i=1}^{N-1} q_i E[|(M^{\pi,\star}_{t_{i+1}} - M^{\pi,\star}_{t_i}) - (\hat{M}^{\pi}_{t_{i+1}} - \hat{M}^{\pi}_{t_i})|^2 | \mathcal{G}_{t_0}].
\end{aligned}$$

Thus,

$$\begin{aligned}
&\tilde{\mathcal{E}}(Y^{\pi,\star} - \overline{Y}^{\pi}, M^{\pi,\star} - \hat{M}^{\pi}) \leqslant (3+\gamma T) q_N E[|Y^{\pi,\star}_{t_N} - \overline{Y}^{\pi}_{t_N}|^2 | \mathcal{G}_{t_0}] \\
&\quad + \tilde{C}\left[\max_{0\leqslant i\leqslant N} q_i E[|Y^{\pi,\star}_{t_i} - \overline{Y}^{\pi}_{t_i}|^2 | \mathcal{G}_{t_0}] + \frac{1}{D}\sum_{i=0}^{N-1} q_i \Delta_i E[|Z^{\pi,\star}_{t_i} - \hat{Z}^{\pi}_{t_i}|^2 | \mathcal{G}_{t_0}]\right] \\
&\quad + \tilde{C} \max_{0\leqslant i\leqslant N} q_i E[|\overline{Y}^{\pi}_{t_i} - \hat{Y}^{\pi}_{t_i}|^2 | \mathcal{G}_{t_0}]
\end{aligned}$$

with

$$\tilde{C} = \left[(2+\gamma T)\kappa^2(2T+D)(|\pi| + \Gamma^{-1}) + \frac{(D+2T)\kappa^2}{\gamma}\right].$$

Considering a weighted formulation of the estimate in (3.7), we have for $\gamma = 4(2T + D)\kappa^2$ and $\Gamma = 4\kappa^2(2T+D)(2+\gamma T)$

$$\begin{aligned}
&\tilde{\mathcal{E}}(Y^{\pi,\star} - \overline{Y}^{\pi}, M^{\pi,\star} - \hat{M}^{\pi}) \leqslant (3+\gamma T) q_N E[|Y^{\pi,\star}_{t_N} - \overline{Y}^{\pi}_{t_N}|^2 | \mathcal{G}_{t_0}] \\
&\quad + \left(\frac{\Gamma|\pi|+1}{4} + \frac{1}{4}\right)\left[\tilde{\mathcal{E}}(Y^{\pi,\star} - \overline{Y}^{\pi}, M^{\pi,\star} - \hat{M}^{\pi}) + \max_{0\leqslant i\leqslant N} q_i E[|\hat{Y}^{\pi}_{t_i} - \overline{Y}^{\pi}_{t_i}|^2 | \mathcal{G}_{t_0}]\right].
\end{aligned}$$

Then, we receive for $|\pi| \leqslant \Gamma^{-1}$

$$\begin{aligned}
&\tilde{\mathcal{E}}(Y^{\pi,\star} - \overline{Y}^{\pi}, M^{\pi,\star} - \hat{M}^{\pi}) \\
&\quad \leqslant 4(3+\gamma T) q_N E[|Y^{\pi,\star}_{t_N} - \overline{Y}^{\pi}_{t_N}|^2 | \mathcal{G}_{t_0}] + 3 \max_{0\leqslant i\leqslant N} q_i E[|\hat{Y}^{\pi}_{t_i} - \overline{Y}^{\pi}_{t_i}|^2 | \mathcal{G}_{t_0}].
\end{aligned}$$

Now, it remains to make use of Young's inequality twice. Bearing in mind the definition of q_i, we have

$$\begin{aligned}
&\max_{0\leqslant i\leqslant N} E[|Y^{\pi,\star}_{t_i} - \hat{Y}^{\pi}_{t_i}|^2 | \mathcal{G}_{t_0}] + E[|M^{\pi,\star}_{t_N} - \hat{M}^{\pi}_{t_N}|^2 | \mathcal{G}_{t_0}] \\
&\quad \leqslant \tilde{\mathcal{E}}(Y^{\pi,\star} - \overline{Y}^{\pi}, M^{\pi,\star} - \hat{M}^{\pi}) + 2 \max_{0\leqslant i\leqslant N} q_i E[|\hat{Y}^{\pi}_{t_i} - \overline{Y}^{\pi}_{t_i}|^2 | \mathcal{G}_{t_0}] \\
&\quad = 4(3+\gamma T) e^{\Gamma T} E[|Y^{\pi,\star}_{t_N} - \overline{Y}^{\pi}_{t_N}|^2 | \mathcal{G}_{t_0}] + (3e^{\Gamma T} + 2) \max_{0\leqslant i\leqslant N} E[|\hat{Y}^{\pi}_{t_i} - \overline{Y}^{\pi}_{t_i}|^2 | \mathcal{G}_{t_0}] \\
&\quad \leqslant 8(3+\gamma T) e^{\Gamma T} E[|\xi^{\pi} - \hat{Y}^{\pi}_{t_i}|^2 | \mathcal{G}_{t_0}] \\
&\qquad + \left((3 + 8(3+\gamma T))\, e^{\Gamma T} + 2 \right) \max_{0\leqslant i\leqslant N} E[|\hat{Y}^{\pi}_{t_i} - \overline{Y}^{\pi}_{t_i}|^2 | \mathcal{G}_{t_0}] \\
&\quad \leqslant \left((3 + 8(3 + 4(2T+D)\kappa^2 T))\, e^{\Gamma T} + 2 \right) \mathcal{E}_{\pi}(\hat{Y}^{\pi}, \hat{M}^{\pi}),
\end{aligned}$$

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

$$\hat{Y}^{\pi}_{t_i} - \overline{Y}^{\pi}_{t_i} = \hat{Y}^{\pi}_{t_i} - \hat{Y}^{\pi}_{t_0} - \sum_{j=0}^{i-1} \Delta_j f^{\pi}(t_j, \hat{Y}^{\pi}_{t_j}, \Delta_j^{-1} E[(\Delta W_j)^* \hat{M}^{\pi}_{t_{j+1}} | \mathcal{G}_{t_j}]) - \hat{M}^{\pi}_{t_i}. \qquad \square$$

3.1.2 Examples for the application on numerical approaches

In order to illustrate the global a-posteriori criterion in more detail, we will quite roughly describe the generic background of some numerical approaches and how the error criterion works in these settings. Here, we focus on time-discretized Markovian BSDEs. That means, we suppose that there is a $(\mathcal{F}_{t_i})_{t_i\in\pi}$-adapted Markov process $(X^{\pi}_{t_i})_{t_i\in\pi}$ such that $Y^{\pi,\star}_{t_i}$ and $Z^{\pi,\star}_{t_i}$, $i = 0, \dots, N-1$ can be expressed by discrete functions $(y^{\pi,\star}_i(x), z^{\pi,\star}_i(x))$, $i = 0, \dots, N-1$ that will be applied on $X^{\pi}_{t_i}$, i. e.

$$Y^{\pi,\star}_{t_i} = y^{\pi,\star}_i(X^{\pi}_{t_i}), \quad Z^{\pi,\star}_{t_i} = z^{\pi,\star}_i(X^{\pi}_{t_i})$$

for $i = 0, \dots, N-1$. For the sake of simplicity we also assume here, that ξ^{π} can be written as deterministic function $\phi^{\pi}(X^{\pi}_{t_N})$. Then we are in a comparable situation as in Subsection 2.2.1. Now, one aims at estimating the deterministic functions $(y^{\pi,\star}_i(x), z^{\pi,\star}_i(x))$, $i = 0, \dots, N-1$. Let these estimators be of the form

$$\hat{y}^{\pi}_i(x, \Xi), \quad \hat{z}^{\pi}_i(x, \Xi), \quad i = 0, \dots, N-1,$$

where Ξ is some random vector independent of $\mathbb{F}$, which is the natural filtration generated by the Brownian motion W. Then we define the enlarged σ-algebra $\mathbb{G}$ by setting $\mathcal{G}_t = \mathcal{F}_t \vee \sigma(\Xi)$. Note that W remains a Brownian motion with respect to $\mathbb{G}$.

Example 7. This quite generally formulated setting contains also least-squares Monte Carlo estimation for BSDEs as explained in Section 2.2, where $\hat{y}_i^\pi(x,\Xi)$ and $\hat{z}_{d,i}^\pi(x,\Xi)$, $d = 1,\dots,D$ are constructed by linear combinations of functions $\eta_d(i,x)$, $d = 0,\dots,D$. Looking back in Subsection 2.2.4 shows, that the computation of the corresponding coefficients involves independent copies of $(X_{t_i}^\pi)_{t_i\in\pi}$. These can be gathered within the random vector Ξ. Now, we define the $(\mathcal{G}_{t_i})_{t_i\in\pi}$-adapted approximate solution of (3.2) by

$$\hat{Y}_{t_i}^\pi = \hat{y}_i^\pi(X_{t_i}^\pi,\Xi), \quad \hat{M}_{t_{i+1}}^\pi - \hat{M}_{t_i}^\pi = \hat{z}_i^\pi(X_{t_i}^\pi,\Xi)\Delta W_i,$$

where the last definition is obviously a martingale with respect to $(\mathcal{G}_{t_i})_{t_i\in\pi}$ but not to $(\mathcal{F}_{t_i})_{t_i\in\pi}$. As $\hat{Z}_{t_i}^\pi = \Delta_i^{-1}E[(\Delta W_i)^*\hat{M}_{t_{i+1}}^\pi|\,\mathcal{G}_{t_i}]$, the global a-posteriori criterion can as well be formulated as follows:

$$\begin{aligned}\mathcal{E}_\pi(\hat{Y}^\pi,\hat{Z}^\pi) &:= E[|\xi^\pi - \hat{Y}_{t_N}^\pi|^2|\,\mathcal{G}_{t_0}] \\ &\quad + \max_{1\leqslant i\leqslant N} E[|\hat{Y}_{t_i}^\pi - \hat{Y}_{t_0}^\pi - \sum_{j=0}^{i-1}\Delta_j f^\pi(t_j,\hat{Y}_{t_j}^\pi,\hat{Z}_{t_j}^\pi) - \sum_{j=0}^{i-1}\hat{Z}_{t_j}^\pi\Delta W_j|^2|\,\mathcal{G}_{t_0}].\end{aligned}$$

In order to derive information about the approximation error from this a-posteriori criterion, we estimate it by Monte Carlo simulation. To this end, we suppose that a realization of Ξ is given and that it is possible to draw independent copies of $(X_{t_i}^\pi)_{t_i\in\pi}$ and of the Brownian increments $(\Delta W_i)_{i=0,\dots,N-1}$. Precisely, let $\mathcal{X}^L$ be such a set of samples, i. e.

$$\mathcal{X}^L = \{({}_\lambda X_{t_{i+1}}^\pi, \Delta_\lambda W_i)_{i=0,\dots,N-1}|\,\lambda = 1,\dots,L\}.$$

Thanks to the definition of $\hat{y}_i^\pi(x,\Xi)$, $\hat{z}_i^\pi(x,\Xi)$ and $\phi^\pi(x)$ we can produce samples

$$({}_\lambda\hat{Y}_{t_i}^\pi, {}_\lambda\hat{Z}_{t_i}^\pi, f^\pi(t_i,{}_\lambda\hat{Y}_{t_i}^\pi,{}_\lambda\hat{Z}_{t_i}^\pi), \Delta_\lambda W_i, {}_\lambda\xi^\pi)_{i=0,\dots,N}, \quad \lambda = 1,\dots,L,$$

that are independent conditional to Ξ. Hence, we can estimate $\mathcal{E}_\pi(\hat{Y}^\pi,\hat{Z}^\pi)$ by

$$\begin{aligned}\hat{\mathcal{E}}_\pi(\hat{Y}^\pi,\hat{Z}^\pi) &:= \frac{1}{L}\sum_{\lambda=1}^{L}|{}_\lambda\xi^\pi - {}_\lambda\hat{Y}_{t_N}^\pi|^2 \\ &\quad + \max_{1\leqslant i\leqslant N}\frac{1}{L}\sum_{\lambda=1}^{L}|{}_\lambda\hat{Y}_{t_i}^\pi - {}_\lambda\hat{Y}_{t_0}^\pi - \sum_{j=0}^{i-1}\Delta_i f^\pi(t_j,{}_\lambda\hat{Y}_{t_j}^\pi,{}_\lambda\hat{Z}_{t_j}^\pi) - \sum_{j=0}^{i-1}{}_\lambda\hat{Z}_{t_i}^\pi\Delta_\lambda W_i|^2.\end{aligned}$$

Considering the result of Theorem 6 we get thereby estimations on the lower and upper bound of the approximation error between $(Y_{t_i}^{\pi,\star}, M_{t_i}^{\pi,\star})_{t_i\in\pi}$ and $(\hat{Y}_{t_i}^\pi,\hat{M}_{t_i}^\pi)_{t_i\in\pi}$.

Example 8. In Chapter 4 we will examine a simplification of least-squares Monte Carlo. There, we assume that

$$\hat{z}_i^\pi(x,\Xi) := \frac{1}{\Delta_i}E[(\Delta W_i)^*\hat{y}_{i+1}^\pi(X_{t_{i+1}}^\pi,\Xi)|\,\Xi, X_{t_i}^\pi = x]$$

and

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

are computable in closed form. This allows us to define

$$\hat{M}^{\pi}_{t_0} = 0, \quad \hat{M}^{\pi}_{t_{i+1}} - \hat{M}^{\pi}_{t_i} = \hat{y}^{\pi}_{i+1}(X^{\pi}_{t_{i+1}}, \Xi) - E[\hat{y}^{\pi}_{i+1}(X^{\pi}_{t_{i+1}}, \Xi) | \Xi, X^{\pi}_{t_i} = x]$$

for $i = 0, \ldots, N-1$. Note, that in Example 7 it was impossible to define the martingale differences $\hat{M}^{\pi}_{t_{i+1}} - \hat{M}^{\pi}_{t_i}$ in such a way, since we require these martingale differences in closed form. Like before,

$$\hat{Z}^{\pi}_{t_i} = \hat{z}^{\pi}_i(X^{\pi}_{t_i}, \Xi) = \Delta_i^{-1} E[(\Delta W_i)^* \hat{M}^{\pi}_{t_{i+1}} | \mathcal{G}_{t_i}].$$

Here, the global a-posteriori criterion equals

$$\begin{aligned}\mathcal{E}_{\pi}(\hat{Y}^{\pi}, \hat{M}^{\pi}) := E[|\xi^{\pi} - \hat{Y}^{\pi}_{t_N}|^2 | \mathcal{G}_{t_0}] \\ + \max_{1 \leqslant i \leqslant N} E[|\hat{Y}^{\pi}_{t_i} - \hat{Y}^{\pi}_{t_0} - \sum_{j=0}^{i-1} \Delta_j f^{\pi}(t_j, \hat{Y}^{\pi}_{t_j}, \hat{Z}^{\pi}_{t_j}) - \hat{M}^{\pi}_{t_i}|^2 | \mathcal{G}_{t_0}].\end{aligned}$$

Similarly to Example 7, we use independent copies of $(X^{\pi}_{t_i})_{t_i \in \pi}$ and the definition of $\hat{y}^{\pi}_{t_i}(x, \Xi)$ and $\hat{z}^{\pi}_{t_i}(x, \Xi)$ to get samples

$$({}_{\lambda}\hat{Y}^{\pi}_{t_i}, {}_{\lambda}\hat{Z}^{\pi}_{t_i}, {}_{\lambda}\hat{M}^{\pi}_{t_i}, f^{\pi}(t_i, {}_{\lambda}\hat{Y}^{\pi}_{t_i}, {}_{\lambda}\hat{Z}^{\pi}_{t_i}), {}_{\lambda}\xi^{\pi})_{i=0,\ldots,N}, \quad \lambda = 1, \ldots, L.$$

Then, the estimator $\hat{\mathcal{E}}_{\pi}(\hat{Y}^{\pi}, \hat{M}^{\pi})$ is analogously defined as in Example 7.

3.1.3 Global a-posteriori criterion for continuous BSDEs

Now we return to the original setting, where we dealt with continuous BSDEs, as formulated in (1.1):

$$Y_t = \xi - \int_t^T f(u, Y_u, Z_u) du - \int_t^T Z_u dW_u.$$

Again we received by some arbitrary numerical algorithm an approximate solution $(\hat{Y}^{\pi}_{t_i}, \hat{Z}^{\pi}_{t_i})_{t_i \in \pi}$, that is defined on the discretized time interval π. We assume that it is square-integrable and adapted to $(\mathcal{G}_{t_i})_{t_i \in \pi}$. Like before, $\mathbb{G}$ is the σ-algebra defined by $\mathcal{G}_{t_i} = \mathcal{F}_{t_i} \vee \sigma(\Xi)$, where Ξ is some random vector independent of $\mathcal{F}_T$. This time we want to judge the approximation error between (Y, Z) and $(\hat{Y}^{\pi}_{t_i}, \hat{Z}^{\pi}_{t_i})_{t_i \in \pi}$ by

$$\begin{aligned}\mathcal{E}_{\pi}(\hat{Y}^{\pi}, \hat{Z}^{\pi}) := E[|\xi^{\pi} - \hat{Y}^{\pi}_{t_N}|^2 | \mathcal{G}_{t_0}] \\ + \max_{1 \leqslant i \leqslant N} E[|\hat{Y}^{\pi}_{t_i} - \hat{Y}^{\pi}_{t_0} - \sum_{j=0}^{i-1} \Delta_j f^{\pi}(t_j, \hat{Y}^{\pi}_{t_j}, \hat{Z}^{\pi}_{t_j}) - \sum_{j=0}^{i-1} \hat{Z}^{\pi}_{t_j} \Delta W_j|^2 | \mathcal{G}_{t_0}].\end{aligned}$$

In contrast to (3.5), we replace $\hat{M}^{\pi}_{t_i}$ by the sum over $\hat{Z}^{\pi}_{t_i}\Delta W_i$, which are martingale differences with respect to $\mathcal{G}_{t_i}$ as well. However, $\mathcal{E}_{\pi}(\hat{Y}^{\pi}, \hat{Z}^{\pi})$ still measures, whether $(\hat{Y}^{\pi}_{t_i}, \hat{Z}^{\pi}_{t_i})_{t_i \in \pi}$ is close to solving the time-discretized BSDE, even though we are situated in a continuous case. The reason is that it might be impossible to draw samples of ξ and $f(t, \hat{Y}^{\pi}_{t_i}, \hat{Z}^{\pi}_{t_i})$. As we want to ensure that $\mathcal{E}_{\pi}(\hat{Y}^{\pi}, \hat{Z}^{\pi})$ can be estimated via Monte Carlo simulation, we have replaced (ξ, f) by their approximations (ξ^{π}, f^{π}).

Assumption 5. *(i) The approximate terminal condition ξ^{π} is a real valued, square-integrable, and $\mathcal{F}_{t_N}$-measurable random variable.*

(ii) The approximate driver is a function $f^{\pi} : \Omega \times \pi \times \mathbb{R} \times \mathbb{R}^D \to \mathbb{R}$ such that $f^{\pi}(t_i, y, z)$ is $\mathcal{F}_{t_i}$-measurable for every $(t_i, y, z) \in \pi \times \mathbb{R} \times \mathbb{R}^D$ and $f^{\pi}(t_i, 0, 0)$ is square-integrable for every $t_i \in \pi$. Furthermore, f^{π} is Lipschitz continuous in (y, z) with constant κ uniformly in (t_i, ω) and independent of π.

The next theorem provides estimates on the L^2-error between the true solution of the BSDE and its approximation. These estimates consist of terms of the approximate solution $(\hat{Y}^{\pi}, \hat{Z}^{\pi})$, the approximate data (ξ^{π}, f^{π}) and the L^2-error between approximate and original data.

Theorem 9. *We assume that Assumption 1 and 5 are satisfied. Let $\mathcal{G}_{t_0}$ be independent of $\mathbb{F}$. We also define the abbreviation $\Delta f^{\pi}_i(t) = f(t, Y_t, Z_t) - f^{\pi}(t_i, Y_t, Z_t)$. Then there are constants C, $c > 0$ depending on κ, T, D and the data (ξ, f) such that for every pair of $(\mathcal{G}_{t_i})_{t_i \in \pi}$-adapted, square-integrable processes $(\hat{Y}^{\pi}_{t_i}, \hat{Z}^{\pi}_{t_i})_{t_i \in \pi}$ and $|\pi|$ small enough*

$$\max_{0 \leqslant i \leqslant N} E[|Y_{t_i} - \hat{Y}^{\pi}_{t_i}|^2 | \mathcal{G}_{t_0}] + \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} E\left[|Y_t - \hat{Y}^{\pi}_{t_i}|^2 + |Z_t - \hat{Z}^{\pi}_{t_i}|^2 | \mathcal{G}_{t_0}\right] dt$$
$$\leqslant C\Big(\mathcal{E}_{\pi}(\hat{Y}^{\pi}, \hat{Z}^{\pi}) + |\pi| + E|\xi - \xi^{\pi}|^2 + \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} E|\Delta f^{\pi}_i(t)|^2 dt\Big).$$

and

$$\mathcal{E}_{\pi}(\hat{Y}^{\pi}, \hat{Z}^{\pi}) \leqslant c\Big(\max_{0 \leqslant i \leqslant N} E[|Y_{t_i} - \hat{Y}^{\pi}_{t_i}|^2 | \mathcal{G}_{t_0}] + \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} E[|Y_t - \hat{Y}^{\pi}_{t_i}|^2 | \mathcal{G}_{t_0}] dt$$
$$+ \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} E[|Z_t - \hat{Z}^{\pi}_{t_i}|^2 | \mathcal{G}_{t_0}] dt + E|\xi - \xi^{\pi}|^2 + \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} E|\Delta f^{\pi}_i(t)|^2 dt\Big).$$

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

$$\max_{0 \leqslant i \leqslant N} E[|Y_{t_i} - \hat{Y}^{\pi}_{t_i}|^2 | \mathcal{G}_{t_0}] + \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} E[|Z_t - \hat{Z}^{\pi}_{t_i}|^2 | \mathcal{G}_{t_0}] dt$$
$$\leqslant C\Big(\mathcal{E}_{\pi}(\hat{Y}^{\pi}, \hat{Z}^{\pi}) + E|\xi - \xi^{\pi}|^2 + \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} E|\Delta f^{\pi}_i(t)|^2 dt\Big)$$

and

$$\mathcal{E}_\pi(\hat{Y}^\pi, \hat{Z}^\pi) \leqslant c\Big(\max_{0\leqslant i\leqslant N} E[|Y_{t_i} - \hat{Y}^\pi_{t_i}|^2|\,\mathcal{G}_{t_0}] + \sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}} E[|Z_t - \hat{Z}^\pi_{t_i}|^2|\,\mathcal{G}_{t_0}]dt$$
$$+ E|\xi - \xi^\pi|^2 + \sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}} E|\Delta f_i^\pi(t)|^2 dt\Big).$$

The above inequalities can quickly be shown by means of Theorem 6 and the L^2-distance between the true solution (Y, Z) of the continuous BSDE and the pair $(Y^{\pi,\star}_{t_i}, Z^{\pi,\star}_{t_i})_{t_i\in\pi}$, that we derived from the solution of the discrete-time BSDE, see (3.2). The following Lemma provides an upper bound for this L^2-distance. Recalling the definition of $Z^{\pi,\star}_{t_i}$ in (3.4), we obtain

Lemma 10. *Let Assumption 1 and 5 be satisfied. Furthermore, we suppose that* f^π *is Lipschitz continuous in the way that*

$$|f^\pi(t_i, y, z) - f^\pi(t_i, y', z')| \leqslant \kappa_y|y - y'| + \kappa|z - z'|, \quad \kappa_y \leqslant \kappa$$

for all $(y, z), (y', z') \in \mathbb{R}\times\mathbb{R}^D$ *uniformly in* (t_i, ω) *and independent of* π*. Then there is a constant* $C > 0$ *depending on* κ*,* T *and the data* (ξ, f) *such that for* $|\pi|$ *small enough*

$$\max_{0\leqslant i\leqslant N} E|Y_{t_i} - Y^{\pi,\star}_{t_i}|^2 + \sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}} E|Z_t - Z^{\pi,\star}_{t_i}|^2 dt$$
$$\leqslant C\Big\{E|Y_{t_N} - \xi^\pi|^2 + \kappa_y^2|\pi| + \sum_{i=0}^{N-1} E|Z^{\pi,\star}_{t_i}\Delta W_i - (Y^{\pi,\star}_{t_{i+1}} - E[Y^{\pi,\star}_{t_{i+1}}|\,\mathcal{F}_{t_i}])|^2 \tag{3.8}$$
$$+ \sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}} E|f(t, Y_t, Z_t) - f^\pi(t_i, Y_t, Z_t)|^2 dt\Big\}.$$

Note, that the proof of the next lemma follows the argumentation in Bouchard and Touzi (2004), Theorem 3.1.

Proof. The pairs (Y, Z) and $(\check{Y}^{\pi,\star}_t, \check{Z}^{\pi,\star}_t)$ are solving for $t \in [t_i, t_{i+1})$ the following differential equations

$$Y_t = Y_{t_{i+1}} - \int_t^{t_{i+1}} f(s, Y_s, Z_s)ds - \int_t^{t_{i+1}} Z_s dW_s,$$
$$\check{Y}^{\pi,\star}_t = Y^{\pi,\star}_{t_{i+1}} - f^\pi(t_i, Y^{\pi,\star}_{t_i}, Z^{\pi,\star}_{t_i})(t - t_i) - \int_t^{t_{i+1}} \check{Z}^{\pi,\star}_s dW_s,$$

where $\check{Z}^{\pi,\star}_t$ can be obtained by the martingale representation theorem, i. e.

$$\int_{t_i}^{t_{i+1}} \check{Z}^{\pi,\star}_t dW_t = Y^{\pi,\star}_{t_{i+1}} - E[Y^{\pi,\star}_{t_{i+1}}|\,\mathcal{F}_{t_i}]. \tag{3.9}$$

At time t_i we have $\check{Y}^{\pi,\star}_{t_i} = Y^{\pi,\star}_{t_i}$ by definition. By Itô's Lemma follows then

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

Concerning summand (II), we receive due to Young's inequality for some $\gamma > 0$

$$\begin{aligned}
(II) \leqslant{}& \gamma \int_t^{t_{i+1}} E|Y_s - \check{Y}^{\pi,\star}_s|^2 ds + \frac{2}{\gamma}\int_{t_i}^{t_{i+1}} E|f^\pi(t_i, Y_s, Z_s) - f^\pi(t_i, Y^{\pi,\star}_{t_i}, Z^{\pi,\star}_{t_i})|^2 ds \\
&+ \frac{2}{\gamma}\int_{t_i}^{t_{i+1}} E|f(s, Y_s, Z_s) - f^\pi(t_i, Y_s, Z_s)|^2 ds.
\end{aligned}$$

Next the Lipschitz condition on f^π yields together with Young's inequality

$$\begin{aligned}
(II) \leqslant{}& \gamma \int_t^{t_{i+1}} E|Y_s - \check{Y}^{\pi,\star}_s|^2 ds + \frac{4}{\gamma}\int_{t_i}^{t_{i+1}} \left(\kappa_y^2 E|Y_s - Y^{\pi,\star}_{t_i}|^2 + \kappa^2 E|Z_s - Z^{\pi,\star}_{t_i}|^2\right) ds \\
&+ C\int_{t_i}^{t_{i+1}} E|f(s, Y_s, Z_s) - f^\pi(t_i, Y_s, Z_s)|^2 ds.
\end{aligned}$$

In view of the setting explained in (1.1) and the Lipschitz condition on f, we can make use of Lemma 2. Hence,

$$\begin{aligned}
E|Y_s - Y^{\pi,\star}_{t_i}|^2 &\leqslant 2E|Y_s - Y_{t_i}|^2 + 2E|Y_{t_i} - Y^{\pi,\star}_{t_i}|^2 \\
&\leqslant C|\pi| + C\int_{t_i}^{t_{i+1}} E|Z_t|^2 dt + 2E|Y_{t_i} - Y^{\pi,\star}_{t_i}|^2.
\end{aligned}$$

Coming back to summand (II), we have as $\kappa_y < \kappa$,

$$\begin{aligned}
(II) \leqslant{}& \gamma \int_t^{t_{i+1}} E|Y_s - \check{Y}^{\pi,\star}_s|^2 ds + \frac{8\kappa^2}{\gamma}\left(\Delta_i E|Y_{t_i} - Y^{\pi,\star}_{t_i}|^2 + \frac{1}{4}\int_{t_i}^{t_{i+1}} E|Z_s - Z^{\pi,\star}_{t_i}|^2 ds\right) \\
&+ C\kappa_y^2|\pi|\left(\Delta_i + \int_{t_i}^{t_{i+1}} E|Z_t|^2 dt\right) \\
&+ C\int_{t_i}^{t_{i+1}} E|f(s, Y_s, Z_s) - f^\pi(t_i, Y_s, Z_s)|^2 ds \\
=:{}& \gamma \int_t^{t_{i+1}} E|Y_s - \check{Y}^{\pi,\star}_s|^2 ds + \frac{8\kappa^2}{\gamma} A_i + B_i.
\end{aligned}$$

Summarizing (I) and (II), we get

$$\begin{aligned}
E|Y_t - \check{Y}^{\pi,\star}_t|^2 &\leqslant E|Y_t - \check{Y}^{\pi,\star}_t|^2 + \int_t^{t_{i+1}} E|Z_s - \check{Z}^{\pi,\star}_s|^2 ds \\
&\leqslant E|Y_{t_{i+1}} - Y^{\pi,\star}_{t_{i+1}}|^2 + \gamma\int_t^{t_{i+1}} E|Y_s - \check{Y}^{\pi,\star}_s|^2 ds + \frac{8\kappa^2}{\gamma} A_i + B_i
\end{aligned} \tag{3.10}$$

and by Gronwall's lemma follows

$$E|Y_t - \check{Y}_t^{\pi,\star}|^2 \leqslant e^{\gamma\Delta_i}(E|Y_{t_{i+1}} - Y_{t_{i+1}}^{\pi,\star}|^2 + 8\kappa^2 A_i/\gamma + B_i).$$

Inserting this result into the second inequality of (3.10) yields

$$\begin{aligned} &E|Y_{t_i} - Y_{t_i}^{\pi,\star}|^2 + \int_{t_i}^{t_{i+1}} E|Z_t - \check{Z}_t^{\pi,\star}|^2 dt \\ &\quad \leqslant (1+\gamma\Delta_i e^{\gamma\Delta_i})(E|Y_{t_{i+1}} - Y_{t_{i+1}}^{\pi,\star}|^2 + \frac{8\kappa^2}{\gamma}A_i + B_i) \\ &\quad \leqslant (1+C\gamma\Delta_i)(E|Y_{t_{i+1}} - Y_{t_{i+1}}^{\pi,\star}|^2 + \frac{8\kappa^2}{\gamma}A_i + B_i) \end{aligned}$$

for $|\pi|$ small enough. Then, choosing $\gamma = 64\kappa^2$ and $|\pi| \leqslant 1/(C\gamma)$ leads to

$$\begin{aligned} E|Y_{t_i} - Y_{t_i}^{\pi,\star}|^2 + \int_{t_i}^{t_{i+1}} E|Z_t - \check{Z}_t^{\pi,\star}|^2 dt &\leqslant (1+C\gamma\Delta_i)\left(E|Y_{t_{i+1}} - Y_{t_{i+1}}^{\pi,\star}|^2 + B_i\right) \\ &+ \frac{1}{4}\Delta_i E|Y_{t_i} - Y_{t_i}^{\pi,\star}|^2 + \frac{1}{16}\int_{t_i}^{t_{i+1}} E|Z_t - Z_{t_i}^{\pi,\star}|^2 dt. \end{aligned}$$

Hence, we have for $|\pi|$ small enough

$$\begin{aligned} &E|Y_{t_i} - Y_{t_i}^{\pi,\star}|^2 + \int_{t_i}^{t_{i+1}} E|Z_t - \check{Z}_t^{\pi,\star}|^2 dt \\ &\quad \leqslant (1+C\Delta_i)\left\{E|Y_{t_{i+1}} - Y_{t_{i+1}}^{\pi,\star}|^2 + B_i\right\} + \frac{1}{4}\int_{t_i}^{t_{i+1}} E|Z_t - Z_{t_i}^{\pi,\star}|^2 dt. \end{aligned} \tag{3.11}$$

Next, we make use of

$$\int_{t_i}^{t_{i+1}} E|Z_t - Z_{t_i}^{\pi,\star}|^2 dt \leqslant 2\int_{t_i}^{t_{i+1}} \left(E|Z_t - \check{Z}_t^{\pi,\star}|^2 + E|\check{Z}_t^{\pi,\star} - Z_{t_i}^{\pi,\star}|^2\right) dt \tag{3.12}$$

and it turns out that

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

Thanks to the discrete Gronwall lemma we get an upper bound for the Y-part, i. e.

$$\begin{aligned} &E|Y_{t_i} - Y_{t_i}^{\pi,\star}|^2 \\ &\quad \leqslant e^{CT}\left\{E|Y_{t_N} - \xi^\pi|^2 + C\sum_{j=i}^{N-1} B_i + C\sum_{j=i}^{N-1}\int_{t_i}^{t_{i+1}} E|\check{Z}_t^{\pi,\star} - Z_{t_i}^{\pi,\star}|^2 dt\right\}. \end{aligned} \tag{3.13}$$

By summing (3.11) up from $i = 0$ to $N - 1$ we obtain

$$\sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} E|Z_t - \check{Z}_t^{\pi,\star}|^2 dt \leqslant C \max_{0 \leqslant i \leqslant N} E|Y_{t_i} - Y_{t_i}^{\pi,\star}|^2 + C \sum_{i=0}^{N-1} B_i + \frac{1}{4} \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} E|Z_t - Z_{t_i}^{\pi,\star}|^2 dt$$

and applying this result on (3.12) yields

$$\sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} E|Z_t - Z_{t_i}^{\pi,\star}|^2 dt \leqslant C \max_{0 \leqslant i \leqslant N} E|Y_{t_i} - Y_{t_i}^{\pi,\star}|^2 + C \sum_{i=0}^{N-1} B_i + C \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} E|\check{Z}_t^{\pi,\star} - Z_{t_i}^{\pi,\star}|^2 dt. \tag{3.14}$$

Merging the results in (3.13) and (3.14) gives

$$\max_{0 \leqslant i \leqslant N} E|Y_{t_i} - Y_{t_i}^{\pi,\star}|^2 + \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} E|Z_t - Z_{t_i}^{\pi,\star}|^2 dt \leqslant CE|Y_{t_N} - \xi^\pi|^2 + C \sum_{i=0}^{N-1} B_i + C \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} E|\check{Z}_t^{\pi,\star} - Z_{t_i}^{\pi,\star}|^2 dt. \tag{3.15}$$

Regarding the second summand, we have by definition

$$\begin{aligned} \sum_{i=0}^{N-1} B_i \leqslant\ & C\kappa_y^2 T|\pi| + C\kappa_y^2 |\pi| \int_0^T E|Z_t|^2 dt \\ &+ C \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} E|f(t, Y_t, Z_t) - f^\pi(t_i, Y_t, Z_t)|^2 dt \\ &\leqslant C|\pi| + C \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} E|f(t, Y_t, Z_t) - f^\pi(t_i, Y_t, Z_t)|^2 dt, \end{aligned}$$

as $\int_0^T E|Z_t|^2 dt < \infty$. As far as the third summand of the right-hand side of (3.15)is concerned, we use Itô's isometry and the definition of $\int_{t_i}^{t_{i+1}} \check{Z}_t^{\pi,\star} dW_t$ in (3.9) to complete the proof. □

Remark 11. The third term of the right-hand side of (3.8) has a meaningful interpretation concerning the L^2-regularity of the true control process $(Z_t)_{0 \leqslant t \leqslant T}$. The notion of L^2-regularity was introduced in Zhang (2001) and is defined by

$$\sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} E\left|Z_t - \frac{1}{\Delta_i} E\left[\int_{t_i}^{t_{i+1}} Z_s ds \,\middle|\, \mathcal{F}_{t_i}\right]\right|^2 dt, \tag{3.16}$$

see also Subsection 2.1. In order to show the relation between (3.16) and

$$\sum_{i=0}^{N-1} E|Z_{t_i}^{\pi,\star}\Delta W_i - (Y_{t_{i+1}}^{\pi,\star} - E[Y_{t_{i+1}}^{\pi,\star}|\mathcal{F}_{t_i}])|^2 \tag{3.17}$$

we make some insertions and apply Young's inequality.

$$\begin{aligned}
&\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}} E\Big|Z_t - \frac{1}{\Delta_i}E\Big[\int_{t_i}^{t_{i+1}} Z_s\,ds\Big|\mathcal{F}_{t_i}\Big]\Big|^2 dt \\
&\quad\leqslant 2\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}}\left(E|Z_t - Z_{t_i}^{\pi,\star}|^2 dt + E\Big|Z_{t_i}^{\pi,\star} - \frac{1}{\Delta_i}E\Big[\int_{t_i}^{t_{i+1}} Z_s\,ds\Big|\mathcal{F}_{t_i}\Big]\Big|^2\right) dt \\
&\quad\leqslant 4\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}} E|Z_t - Z_{t_i}^{\pi,\star}|^2 dt,
\end{aligned}$$

where the last step followed by Jensen's inequality. Assuming $E|\xi - \xi^{\pi}|^2 \leqslant C|\pi|$ and $\sup_{t_i\leqslant t\leqslant t_{i+1}} E|f(t,y,z) - f^{\pi}(t_i,y,z)|^2 \leqslant C|\pi|$ for all $t_i \in \pi$, we obtain by Lemma 10

$$\begin{aligned}
&\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}} E\Big|Z_t - \frac{1}{\Delta_i}E\Big[\int_{t_i}^{t_{i+1}} Z_s\,ds\Big|\mathcal{F}_{t_i}\Big]\Big|^2 dt \\
&\quad\leqslant C|\pi| + C\sum_{i=0}^{N-1} E|Z_{t_i}^{\pi,\star}\Delta W_i - (Y_{t_{i+1}}^{\pi,\star} - E[Y_{t_{i+1}}^{\pi,\star}|\mathcal{F}_{t_i}])|^2.
\end{aligned}$$

On the other hand we have by the definition of $\check{Z}_t^{\pi,\star}$ in the previous proof, Itô' s isometry and Young' s inequality

$$\begin{aligned}
&\sum_{i=0}^{N-1} E|Z_{t_i}^{\pi,\star}\Delta W_i - (Y_{t_{i+1}}^{\pi,\star} - E[Y_{t_{i+1}}^{\pi,\star}|\mathcal{F}_{t_i}])|^2 \\
&\quad\leqslant 2\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}} E|Z_{t_i}^{\pi,\star} - Z_t|^2 dt + 2\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}} E|Z_t - \check{Z}_t^{\pi,\star}|^2 dt.
\end{aligned}$$

Now we apply Young's inequality on the first summand of the above right-hand side and use then the relation $\Delta_i Z_{t_i}^{\pi,\star} = E[\int_{t_i}^{t_{i+1}} \check{Z}_t^{\pi,\star} dt|\mathcal{F}_{t_i}]$, see Lemma 3.1 in Bouchard and Touzi (2004). Due to Jensen's inequality we receive

$$\begin{aligned}
&E|Z_{t_i}^{\pi,\star}\Delta W_i - (Y_{t_{i+1}}^{\pi,\star} - E[Y_{t_{i+1}}^{\pi,\star}|\mathcal{F}_{t_i}])|^2 \\
&\quad\leqslant 4\int_{t_i}^{t_{i+1}} E\Big|Z_{t_i}^{\pi,\star} - \frac{1}{\Delta_i}E\Big[\int_{t_i}^{t_{i+1}} Z_s\,ds\Big|\mathcal{F}_{t_i}\Big]\Big|^2 dt \\
&\qquad + 4\int_{t_i}^{t_{i+1}} E\Big|\frac{1}{\Delta_i}E\Big[\int_{t_i}^{t_{i+1}} Z_s\,ds\Big|\mathcal{F}_{t_i}\Big] - Z_t\Big|^2 dt + 2\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}} E|Z_t - \check{Z}_t^{\pi,\star}|^2 dt \\
&\quad\leqslant 4\int_{t_i}^{t_{i+1}} E\Big|Z_t - \frac{1}{\Delta_i}E\Big[\int_{t_i}^{t_{i+1}} Z_s\,ds\Big|\mathcal{F}_{t_i}\Big]\Big|^2 dt + 6\int_{t_i}^{t_{i+1}} E|Z_s - \check{Z}_s^{\pi,\star}|^2 ds.
\end{aligned}$$

After replacing (3.12) through

$$
\begin{aligned}
&\int_{t_i}^{t_{i+1}} E|Z_t - Z_{t_i}^{\pi,\star}|^2 dt \\
&\leqslant 2\int_{t_i}^{t_{i+1}} E\Big|Z_t - \frac{1}{\Delta_i}E\Big[\int_{t_i}^{t_{i+1}} Z_s\, ds\Big|\,\mathcal{F}_{t_i}\Big]\Big|^2 dt \\
&\quad + 2\int_{t_i}^{t_{i+1}} E\Big|\frac{1}{\Delta_i}E\Big[\int_{t_i}^{t_{i+1}} Z_s\, ds\Big|\,\mathcal{F}_{t_i}\Big] - Z_{t_i}^{\pi,\star}\Big|^2 dt \\
&\leqslant 2\int_{t_i}^{t_{i+1}} E\Big|Z_t - \frac{1}{\Delta_i}E\Big[\int_{t_i}^{t_{i+1}} Z_s\, ds\Big|\,\mathcal{F}_{t_i}\Big]\Big|^2 dt + 2\int_{t_i}^{t_{i+1}} E|Z_s - \check{Z}_s^{\pi,\star}|^2 ds,
\end{aligned}
$$

we repeat the remaining steps of Lemma 10. Together with the assumptions $E|\xi - \xi^\pi|^2 \leqslant C|\pi|$ and $|f(t,y,z) - f^\pi(t_i,y,z)|^2 \leqslant C|\pi|$ we obtain

$$
\begin{aligned}
&\sum_{i=0}^{N-1} E|Z_{t_i}^{\pi,\star}\Delta W_i - (Y_{t_{i+1}}^{\pi,\star} - E[Y_{t_{i+1}}^{\pi,\star}|\,\mathcal{F}_{t_i}])|^2 \\
&\leqslant C|\pi| + C\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}} E\Big|Z_t - \frac{1}{\Delta_i}E\Big[\int_{t_i}^{t_{i+1}} Z_s\, ds\Big|\,\mathcal{F}_{t_i}\Big]\Big|^2 dt.
\end{aligned}
$$

for some constant $C > 0$. Summing up, we can say that (3.16) and (3.17) are equivalent up to a term of order $|\pi|$. That means, (3.17) reflects a property of the original BSDE, precisely the L^2-regularity of Z.

In case we are located in the setting of (1.3) and Assumption 2 is fulfilled, the squared L^2-regularity of Z is of order $|\pi|$ and (3.16) converges with the same rate. However, for the results of Theorem 9 the much weaker Assumptions 1 and 5 are sufficient. Indeed, we estimate (3.17) by the global a-posteriori criterion basically by using the Lipschitz condition on f and f^π.

Proof of Theorem 9. Recall the notation

$$\Delta f_i^\pi(t) = f(t, Y_t, Z_t) - f^\pi(t_i, Y_t, Z_t).$$

We start with the first and third inequality. Therefore, we define the $(\mathcal{G}_{t_i})_{t_i\in\pi}$-martingale $(\hat{M}_{t_i}^\pi)_{t_i\in\pi}$ by setting $\hat{M}_{t_0}^\pi = 0$ and $\hat{M}_{t_{i+1}}^\pi - \hat{M}_{t_i}^\pi := \hat{Z}_{t_i}^\pi \Delta W_i$ for $i = 0,\dots,N-1$. Due to Young's inequality and the independence between $\mathcal{G}_{t_0}$ and $\mathbb{F}$, we have

$$
\begin{aligned}
&\max_{0\leqslant i\leqslant N} E[|Y_{t_i} - \hat{Y}_{t_i}^\pi|^2|\,\mathcal{G}_{t_0}] + \sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}} E[|Z_t - \hat{Z}_{t_i}^\pi|^2|\,\mathcal{G}_{t_0}]dt \\
&\leqslant 2\Big(\max_{0\leqslant i\leqslant N} E[|Y_{t_i} - Y_{t_i}^{\pi,\star}|^2] + \sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}} E[|Z_t - Z_{t_i}^{\pi,\star}|^2]dt\Big) \\
&\quad + 2\Big(\max_{0\leqslant i\leqslant N} E[|Y_{t_i}^{\pi,\star} - \hat{Y}_{t_i}^\pi|^2|\,\mathcal{G}_{t_0}] + \sum_{i=0}^{N-1}\Delta_i E[|Z_{t_i}^{\pi,\star} - \hat{Z}_{t_i}^\pi|^2|\,\mathcal{G}_{t_0}]\Big) \\
&= (I) + (II).
\end{aligned}
$$

Regarding the first summand, we employ the result of Lemma 10. Hence, (I) is bounded by

$$E|Y_{t_N} - \xi^\pi|^2 + \kappa_y^2|\pi| + \sum_{i=0}^{N-1} E|Z_{t_i}^{\pi,\star}\Delta W_i - (Y_{t_{i+1}}^{\pi,\star} - E[Y_{t_{i+1}}^{\pi,\star}|\mathcal{F}_{t_i}])|^2 \\ + \sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}} E|\Delta f_i^\pi(t)|^2 dt.$$

Bear in mind that there is a process $\check{Z}_t^{\pi,\star}$ such that (3.9) holds. Again we incorporate the independence between $\mathcal{G}_{t_0}$ and $\mathbb{F}$ and receive

$$\begin{aligned}
&\sum_{i=0}^{N-1} E|Z_{t_i}^{\pi,\star}\Delta W_i - (Y_{t_{i+1}}^{\pi,\star} - E[Y_{t_{i+1}}^{\pi,\star}|\mathcal{F}_{t_i}])|^2 \\
&\quad = \sum_{i=0}^{N-1} E[|Z_{t_i}^{\pi,\star}\Delta W_i - \int_{t_i}^{t_{i+1}} \check{Z}_s^{\pi,\star} dW_s|^2|\mathcal{G}_{t_0}] \\
&\quad \leqslant 2\sum_{i=0}^{N-1} E[|Z_{t_i}^{\pi,\star}\Delta W_i - \hat{Z}_{t_i}^\pi \Delta W_i|^2 + |\hat{Z}_{t_i}^\pi\Delta W_i - \int_{t_i}^{t_{i+1}} \check{Z}_s^{\pi,\star} dW_s|^2|\mathcal{G}_{t_0}] \\
&\quad = 2\sum_{i=0}^{N-1}\Delta_i E[|Z_{t_i}^{\pi,\star} - \hat{Z}_{t_i}^\pi|^2|\mathcal{G}_{t_0}] + 2E\Big[\Big|\sum_{i=0}^{N-1}\Big(\hat{Z}_{t_i}^\pi\Delta W_i - \int_{t_i}^{t_{i+1}}\check{Z}_s^{\pi,\star}dW_s\Big)\Big|^2\Big|\mathcal{G}_{t_0}\Big].
\end{aligned}$$

Similarly to Theorem 6 we define

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

and recall the identity arising from (2.2):

$$\int_{t_i}^{t_{i+1}} \check{Z}_s^{\pi,\star} dW_s = Y_{t_{i+1}}^{\pi,\star} - Y_{t_i}^{\pi,\star} + \Delta_i f^\pi(t_i, Y_{t_i}^{\pi,\star}, Z_{t_i}^{\pi,\star}).$$

Then, we obtain

$$\begin{aligned}
&E\Big[\Big|\sum_{i=0}^{N-1}\Big(\hat{Z}_{t_i}^\pi\Delta W_i - \int_{t_i}^{t_{i+1}}\check{Z}_s^{\pi,\star}dW_s\Big)\Big|^2\Big|\mathcal{G}_{t_0}\Big] \\
&\quad = E[|\bar{Y}_{t_N}^\pi - \hat{Y}_{t_0}^\pi - \sum_{i=0}^{N-1}\Delta_i f(t_i, \hat{Y}_{t_i}^\pi, \hat{Z}_{t_i}^\pi) \\
&\qquad - (\xi^\pi - Y_{t_0}^{\pi,\star} - \sum_{i=0}^{N-1}\Delta_i f(t_i, Y_{t_i}^{\pi,\star}, Z_{t_i}^{\pi,\star}))|^2|\mathcal{G}_{t_0}] \\
&\quad \leqslant CE[|\bar{Y}_{t_N}^\pi - \hat{Y}_{t_N}^\pi|^2|\mathcal{G}_{t_0}] \\
&\qquad + C\Big(\max_{0\leqslant i\leqslant N} E|Y_{t_i}^{\pi,\star} - \hat{Y}_{t_i}^\pi|^2|\mathcal{G}_{t_0}] + \sum_{i=0}^{N-1}\Delta_i E[|Z_{t_i}^{\pi,\star} - \hat{Z}_{t_i}^\pi|^2|\mathcal{G}_{t_0}]\Big).
\end{aligned} \tag{3.18}$$

The first summand of the right-hand side of (3.18) is bounded by the error criterion by definition of $\bar{Y}^\pi_{t_N}$. The remaining two summands are bounded by a constant times (II). Turning to this summand, we apply the estimate (3.7) and get

$$(II) \leqslant C\Big(\max_{0\leqslant i\leqslant N} E[|Y^{\pi,\star}_{t_i} - \hat{Y}^\pi_{t_i}|^2|\,\mathcal{G}_{t_0}] + D E[|M^{\pi,\star}_{t_N} - \hat{M}^\pi_{t_N}|^2|\,\mathcal{G}_{t_0}]\Big).$$

Then we find ourselves in the setting of Theorem 6 and thus can deduce that $(II) \leqslant C\mathcal{E}_\pi(\hat{Y}^\pi, \hat{M}^\pi)$, i. e. Summand (II) is bounded by terms of the global a-posteriori criterion for discrete-time BSDEs. Due to the definition of $\hat{M}^\pi_{t_i}$, we immediately obtain $(II) \leqslant C\mathcal{E}_\pi(\hat{Y}^\pi, \hat{Z}^\pi)$. In sum,

$$\begin{aligned} &\max_{0\leqslant i\leqslant N} E[|Y_{t_i} - \hat{Y}^\pi_{t_i}|^2|\,\mathcal{G}_{t_0}] + \sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}} E[|Z_t - \hat{Z}^\pi_{t_i}|^2|\,\mathcal{G}_{t_0}]dt \\ &\quad \leqslant C\Big(\mathcal{E}_\pi(\hat{Y}^\pi, \hat{Z}^\pi) + E|\xi - \xi^\pi|^2 + \kappa_y^2|\pi| + \sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}} E|\Delta f^\pi_i(t)|^2 dt\Big). \end{aligned} \tag{3.19}$$

As far as the third inequality is concerned, the proof is complete, since $\kappa_y = 0$ in case f^π does not depend on y. For the first inequality, it remains to give an estimate for

$$\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}} E[|Y_t - \hat{Y}^\pi_{t_i}|^2|\,\mathcal{G}_{t_0}],$$

which is bounded by

$$2\max_{0\leqslant i\leqslant N} E[|Y_{t_i} - \hat{Y}^\pi_{t_i}|^2|\,\mathcal{G}_{t_0}] + 2\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}} E|Y_t - Y_{t_i}|^2.$$

Concerning the first summand, there is an estimate given by (3.19). On the second summand we can apply Lemma 2. Hence,

$$\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}} E|Y_t - Y_{t_i}|^2 \leqslant \sum_{i=0}^{N-1}\Delta_i(|\pi| + \int_{t_i}^{t} E|Z_s|^2 ds) \leqslant C|\pi|,$$

as $\int_0^T E|Z_t|^2 dt < \infty$. This completes the proof on the first inequality. The second part of the proof considers the second and forth inequality. Therefore, we make use of the identity

$$Y_{t_i} - Y_0 = \int_0^{t_i} f(t, Y_t, Z_t)dt + \int_0^{t_i} Z_t dW_t. \tag{3.20}$$

Inserting (3.20) gives

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

Then we obtain by the Itô isometry, Young's inequality and Jensen's inequality

$$\begin{aligned}\mathcal{E}_\pi(\hat{Y}^\pi, \hat{Z}^\pi) \leqslant c\Bigg(& E|\xi - \xi^\pi|^2 + \max_{0 \leqslant i \leqslant N} E[|\hat{Y}^\pi_{t_i} - Y_{t_i}|^2 | \mathcal{G}_{t_0}] \\ & + \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} E[|Z_t - \hat{Z}^\pi_{t_i}|^2 | \mathcal{G}_{t_0}] dt + \sum_{j=0}^{N-1} \int_{t_j}^{t_{j+1}} E|\Delta f^\pi_i(t)|^2 dt \\ & + \sum_{j=0}^{N-1} \int_{t_j}^{t_{j+1}} E[|f^\pi(t_j, Y_t, Z_t) - f^\pi(t_j, \hat{Y}^\pi_{t_j}, \hat{Z}^\pi_{t_j})|^2 | \mathcal{G}_{t_0}] dt \Bigg).\end{aligned} \tag{3.21}$$

Due to the Lipschitz condition of f^π and Young's inequality, we obtain

$$\begin{aligned}&\sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} E[|f^\pi(t_i, Y_t, Z_t) - f^\pi(t_i, \hat{Y}^\pi_{t_i}, \hat{Z}^\pi_{t_i})|^2 | \mathcal{G}_{t_0}] dt \\ &\quad \leqslant 2 \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \left(\kappa_y^2 E[|Y_t - \hat{Y}^\pi_{t_i}|^2 | \mathcal{G}_{t_0}] + \kappa^2 E[|Z_t - \hat{Z}^\pi_{t_i}|^2 | \mathcal{G}_{t_0}] \right) dt.\end{aligned}$$

Combining this inequality with (3.21) yields the second inequality. In case f^π does not depend on y, we have $\kappa_y = 0$. Thus, the fourth inequality is shown as well. □

3.1.4 The a-posteriori error criterion for typical examples of BSDEs

Let S be the solution of the forward SDE

$$S_t = s_0 + \int_0^t b(u, S_u) du + \int_0^t \sigma(u, S_u) dW_u,$$

where the deterministic functions $b : [0, T] \times \mathbb{R}^{\tilde{D}} \to \mathbb{R}^{\tilde{D}}$ and $\sigma : [0, T] \times \mathbb{R}^{\tilde{D}} \to \mathbb{R}^{\tilde{D} \times D}$ are 1/2-Hölder-continuous in time and Lipschitz in its spatial variables.

Irregular terminal condition and Lipschitz continuous driver

We define ξ by $\phi(S_T)$, where ϕ is a deterministic function that is considered to be irregular, as no Lipschitz condition is imposed on ϕ. Many cases in the literature on BSDEs involve a driver, that consists of a deterministic function $F : [0, T] \times \mathbb{R}^{\tilde{D}} \times \mathbb{R} \times \mathbb{R}^D \to \mathbb{R}$, namely

$$f(t, y, z) = F(t, S, y, z),$$

where F is β-Hölder-continuous in t for some $\beta \geqslant 1/2$. Here, we assume that S can be sampled perfectly on the mesh π. Thus, we can set $\xi^\pi = \xi$ and $f^\pi(t_i, y, z) = f(t_i, y, z)$. Then, the first inequality in Theorem 9 simplifies to

$$\begin{aligned}&\max_{0 \leqslant i \leqslant N} E[|Y_{t_i} - \hat{Y}^\pi_{t_i}|^2 | \mathcal{G}_{t_0}] + \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} E\left[|Y_t - \hat{Y}^\pi_{t_i}|^2 + |Z_t - \hat{Z}^\pi_{t_i}|^2 | \mathcal{G}_{t_0}\right] dt \\ &\quad \leqslant C\Big(\mathcal{E}_\pi(\hat{Y}^\pi, \hat{Z}^\pi) + |\pi|\Big).\end{aligned}$$

For ϕ irregular and F Lipschitz in its spatial variables and 1/2-Hölder in t and an equidistant time grid, the time discretization error converges with rate $|\pi|^p$, where p can be smaller than 1/2, see e.g. Gobet and Makhlouf (2010). Then, the global error criterion provides information about the time discretization error.

Lipschitz driver depending on z only

As before, we suggest a terminal condition $\xi = \phi(S_T)$ without any further conditions. But this time we look at the special case $f(t, y, z) = F(z)$ with F being a deterministic Lipschitz function. For the sake of simplicity, we suppose again that S can be sampled perfectly on the grid such that (ξ^π, f^π) can be defined by

$$\xi^\pi = \phi(S_T) = \xi, \quad f^\pi(t_i, y, z) = F(z) = f(t_i, y, z).$$

Since f is independent of y, we have $\kappa_y = 0$ and by the third and fourth inequality of Theorem 9 consequently

$$\begin{aligned}\frac{1}{c}\mathcal{E}_\pi(\hat{Y}^\pi, \hat{Z}^\pi) &\leqslant \max_{0\leqslant i\leqslant N} E[|Y_{t_i} - \hat{Y}^\pi_{t_i}|^2 | \mathcal{G}_{t_0}] + \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} E[|Z_t - \hat{Z}^\pi_{t_i}|^2 | \mathcal{G}_{t_0}] dt \\ &\leqslant C\mathcal{E}_\pi(\hat{Y}^\pi, \hat{Z}^\pi).\end{aligned}$$

It is worth noting that in this case the squared approximation error between the true solution of the continuous BSDE and the approximate solution is equivalent to the global a-posteriori criterion. This is insofar striking, as it is evaluated only by means of the approximate solution. However this equivalence result considers the error between Y and $\hat{Y}^\pi$ merely on the time grid.

Lipschitz continuous terminal condition and Lipschitz continuous driver

Again we look at the case $\xi = \phi(S_T)$ and $f(t, y, z) = F(t, S, y, z)$, where F is deterministic. In contrast to the previous examples, let Assumption 2 be satisfied with the difference that F shall be β-Hölder continuous in time and its Lipschitz constant corresponding to S is denoted by κ_s. Precisely,

$$\begin{aligned}&|\phi(s_1) - \phi(s_2)|^2 \leqslant \kappa|s_1 - s_2|, \\ &|F(t_1, s_1, y_1, z_1) - F(t_2, s_2, y_2, z_2)| \\ &\quad \leqslant \kappa|t_1 - t_2|^\beta + \kappa_s|s_1 - s_2| + \kappa|y_1 - y_2| + \kappa|z_1 - z_2|,\end{aligned}$$

for some $\beta \geqslant 1/2$. Initially, we suppose that for S the approximation S^π_t is at hand, e. g. produced by the Euler scheme. Then we define the approximate data (ξ^π, f^π) by

$$\xi^\pi = \phi^\pi(S^\pi_{t_N}), \quad f^\pi(t_i, y, z) = F(t_i, S^\pi_{t_i}, y, z),$$

where ϕ^π is Lipschitz with constant κ and it holds that $\max_{t_i \in \pi} E|S_{t_i} - S^\pi_{t_i}|^2 \leqslant C|\pi|^{2\beta}$. Under these assumptions,

$$E|\xi - \xi^\pi|^2 + \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} E|F(t, S_t, Y_t, Z_t) - F(t_i, S^\pi_{t_i}, Y_t, Z_t)|^2 dt \\ \leqslant C(\kappa^2|\pi|^{2\beta} + \kappa_s^2|\pi|).$$

Here we also made use of the estimate $E|S_t - S_{t_i}|^2 < C|t - t_i|$, that is valid according to Zhang (2004), see Lemma 3. Considering Assumption 2 we have by Lemma 2 and Lemma 3.2 in Zhang (2004) that

$$\max_{0 \leqslant i \leqslant N-1} \sup_{t_i \leqslant t < t_{i+1}} E|Y_t - Y_{t_i}|^2 < C|\pi|.$$

In view of these estimates the first and second inequality of Theorem 9 reduce to

$$\max_{0 \leqslant i \leqslant N-1} \sup_{t_i \leqslant t < t_{i+1}} E[|Y_t - \hat{Y}^\pi_{t_i}|^2 | \mathcal{G}_{t_0}] + \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} E[|Z_t - \hat{Z}^\pi_{t_i}|^2 | \mathcal{G}_{t_0}] dt \\ \leqslant C(\mathcal{E}_\pi(\hat{Y}^\pi, \hat{Z}^\pi) + |\pi|) \tag{3.22}$$

and

$$\mathcal{E}_\pi(\hat{Y}^\pi, \hat{Z}^\pi) \leqslant c\Big(\max_{0 \leqslant i \leqslant N-1} \sup_{t_i \leqslant t < t_{i+1}} E[|Y_t - \hat{Y}^\pi_{t_i}|^2 | \mathcal{G}_{t_0}] \\ + \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} E[|Z_t - \hat{Z}^\pi_{t_i}|^2 | \mathcal{G}_{t_0}] dt + \kappa^2|\pi|^{2\beta} + \kappa_s^2|\pi| \Big).$$

Due to $|\pi|^{2\beta} < |\pi|$ for $|\pi| < 1$, the error criterion is equivalent to the squared approximation error between the true solution an $(\hat{Y}^\pi, \hat{Z}^\pi)$ up to terms of order $|\pi|$ (which matches the above mentioned squared time discretization error). Contrary to the previous example, this equivalence works with respect to the complete time interval $[0, T]$ and is not restricted to the time grid π.

If the function F does not depend on S, the additional error term in the lower bound reduces to $c\kappa^2|\pi|^{2\beta}$. In case F does not depend on t and the process S can be sampled perfectly on the grid, i. e. $S^\pi_t = S$, we obtain $\xi^\pi = \xi$ and $f = f^\pi$. Then the additional error term $c(\kappa^2|\pi|^{2\beta} + \kappa_s^2|\pi|)$ disappears completely.

Coefficient functions with certain smoothness and boundedness conditions

In the last example we deal with the same data as in the previous example, but this time we assume that the coefficient functions b, σ, ϕ and f satisfy beside Assumption 2 certain smoothness and boundedness conditions. Based on the assumption that S can be sampled perfectly on an equidistant grid π, we set

$$\xi^\pi = \phi(S_T) = \xi, \quad f^\pi(t_i, y, z) = F(t_i, S_{t_i}, y, z) = f(t_i, y, z).$$

For this setting, Gobet and Labart (2007) have shown that

$$\max_{0\leqslant i\leqslant N} E|Y^{\pi,\star}_{t_i} - Y_{t_i}|^2 + \sum_{i=0}^{N-1} \Delta_i E|Z^{\pi,\star}_{t_i} - Z_{t_i}|^2 \leqslant C|\pi|^2. \tag{3.23}$$

In view of (3.4), the combination of Theorem 6 and (3.23) yields

$$\max_{0\leqslant i\leqslant N} E[|\hat{Y}^{\pi}_{t_i} - Y_{t_i}|^2|\,\mathcal{G}_{t_0}] + \sum_{i=0}^{N-1} \Delta_i E[|\hat{Z}^{\pi}_{t_i} - Z_{t_i}|^2|\,\mathcal{G}_{t_0}] \leqslant C(\mathcal{E}_\pi(\hat{Y}^\pi, \hat{Z}^\pi) + |\pi|^2),$$

where $(\hat{Y}^\pi_{t_i}, \hat{Z}^\pi_{t_i})_{t_i\in\pi}$ is $(\mathcal{G}_{t_i})_{t_i\in\pi}$-adapted and $\mathcal{G}_{t_0}$ is independent of $\mathcal{F}_T$. In contrast to (3.22), the additional error term decreases here with rate 1 instead of 1/2 in the L^2-sense. In other words, due to the stronger assumptions we are rewarded with a faster convergence of the additional error term. However, the upper bound is related to the approximation error on the time grid π only.

The estimate on the approximation error is still valid in case

$$\xi^\pi = \phi(S^\pi_{t_N}), \quad f^\pi(t_i, y, z) = F(t_i, S^\pi_{t_i}, y, z)$$

and S^π is a strong order 1 approximation of S, for example generated by the Milstein scheme. This result can be obtained by a comparison of the error criteria with respect to the data (ξ, f) and (ξ^π, f^π), respectively. Clearly, we have

$$\begin{aligned}
&E|\xi - \hat{Y}^\pi_{t_N}|^2 + \max_{1\leqslant i\leqslant N} E[|\hat{Y}^\pi_{t_i} - \hat{Y}^\pi_{t_0} - \sum_{j=0}^{i-1} \Delta_j f(t_j, \hat{Y}^\pi_{t_j}, \hat{Z}^\pi_{t_j}) - \sum_{j=0}^{i-1} \hat{Z}^\pi_{t_j} \Delta W_j|^2|\,\mathcal{G}_{t_0}] \\
&\quad\leqslant 2E|\xi - \xi^\pi|^2 + C\sum_{j=0}^{i-1} \Delta_j E[|F(t_j, S_{t_j}, \hat{Y}^\pi_{t_j}, \hat{Z}^\pi_{t_j}) - F(t_j, S^\pi_{t_j}, \hat{Y}^\pi_{t_j}, \hat{Z}^\pi_{t_j})|^2|\,\mathcal{G}_{t_0}] \\
&\qquad + 2E|\xi - \hat{Y}^\pi_{t_N}|^2 + \max_{1\leqslant i\leqslant N} 2E[|\hat{Y}^\pi_{t_i} - \hat{Y}^\pi_{t_0} \\
&\qquad\quad - \sum_{j=0}^{i-1} \Delta_j f^\pi(t_j, \hat{Y}^\pi_{t_j}, \hat{Z}^\pi_{t_j}) - \sum_{j=0}^{i-1} \hat{Z}^\pi_{t_j} \Delta W_j|^2|\,\mathcal{G}_{t_0}] \\
&\quad\leqslant \max_{0\leqslant i\leqslant N} 2E|S^\pi_{t_i} - S_{t_i}|^2 + 2\Big(E|\xi^\pi - \hat{Y}^\pi_{t_N}|^2 \\
&\qquad + \max_{1\leqslant i\leqslant N} E[|\hat{Y}^\pi_{t_i} - \hat{Y}^\pi_{t_0} - \sum_{j=0}^{i-1} \Delta_j f^\pi(t_j, \hat{Y}^\pi_{t_j}, \hat{Z}^\pi_{t_j}) - \sum_{j=0}^{i-1} \hat{Z}^\pi_{t_j} \Delta W_j|^2|\,\mathcal{G}_{t_0}]\Big).
\end{aligned}$$

3.2 Local error criterion for approximate solutions obtained by projections

During the review of typical BSDEs in the previous subsection, we already indicated the suggestion of an additional 'local' error criterion. In contrast to the globally natured criterion it considers the violation of (3.1) along the partial interval $[t_i, t_{i+1}]$ for all $i = j, \dots, N-1$. Clearly, we define it by taking the L^2-norm and summing up from $i = j$ to $N-1$ the local criterion, i.e.

$$\mathcal{E}^{loc}_{\pi,j}(\hat{Y}^\pi, \hat{Z}^\pi) = \sum_{i=j}^{N-1} E|\hat{Y}^\pi_{t_{i+1}} - \hat{Y}^\pi_{t_i} - \Delta_i f^\pi(t_i, \hat{Y}^\pi_{t_i}, \hat{Z}^\pi_{t_i}) - \hat{Z}^\pi_{t_i}\Delta W_i|^2.$$

Situated in the setting of (1.3), see also Subsection 3.1.4, we will examine this criterion merely for two cases. First, we set $(\hat{Y}^\pi_{t_i}, \hat{Z}^\pi_{t_i})_{t_i\in\pi} = (Y^\pi_{t_i}, Z^\pi_{t_i})_{t_i\in\pi}$, which is the solution of the explicit time discretization scheme in Subsection 2.2.1. The results of this step will primarily have a supporting function for the second step. There we look at $(\hat{Y}^\pi_{t_i}, \hat{Z}^\pi_{t_i})_{t_i\in\pi} = (Y^{\pi,K_{0,i}}_{t_i}, Z^{\pi,K_i}_{t_i})_{t_i\in\pi}$, that means we refer to the 'projection' step of least-squares Monte Carlo, where conditional expectations were replaced by projections on subspaces of $L^2(\mathcal{F}_{t_i})$ spanned by $\eta_{d,i}(X^\pi_{t_i})$, $d = 0, \dots, D$, see (2.8).

A natural third step would be to regard $(Y^{\pi,K_{0,i},L}_{t_i}, Z^{\pi,K_i,L}_{t_i})_{t_i\in\pi}$, which is the numerical solution obtained by (2.17). However, this analysis is similar to that of the approximation error of least-squares Monte Carlo rather intricate. As the emphasis of this work is on the global a-posteriori-criterion and the enhanced least-squares Monte Carlo approach, we neglect this topic here.

Lemma 12. *In the setting of* (1.3) *let Assumptions 2 and 3 be satisfied. Suppose further, there exists a constant such that*

$$E|\xi - \xi^\pi|^2 \leqslant const.|\pi|.$$

Then there is a constant $C > 0$ *depending on* s_0, κ, T, $\widetilde{D}$ *and* D *such that* $\mathcal{E}^{loc}_{\pi,0}(Y^\pi, Z^\pi) \leqslant C|\pi|$.

Proof. In view of (1.3), we have $f^\pi(t_i, y, z) = F(t_i, S^\pi_{t_i}, y, z)$. Then we define

$$\Delta f^\pi_i(u) = F(t_i, S^\pi_{t_i}, Y^\pi_{t_{i+1}}, Z^\pi_{t_i}) - F(u, S_u, Y_u, Z_u).$$

Step 1: We show

$$\int_{t_i}^{t_{i+1}} E|\Delta f^\pi_i(u)|^2 du \leqslant C\Delta_i^2 + C\Delta_i \int_{t_i}^{t_{i+1}} E|Z_u|^2 du + C\big(\Delta_i E|Y^\pi_{t_{i+1}} - Y_{t_{i+1}}|^2 + \int_{t_i}^{t_{i+1}} E|Z^\pi_{t_i} - Z_u|^2 du\big). \quad (3.24)$$

Due to the Lipschitz condition on F, there is a generic constant $C > 0$ depending on κ such that

$$\int_{t_i}^{t_{i+1}} E|\Delta f^\pi_i(u)|^2 du \leqslant C\Delta_i^2 + C\Delta_i \sup_{t_i\leqslant u\leqslant t_{i+1}} E[|S_{t_i} - S_u|^2 + |Y_{t_{i+1}} - Y_u|^2] + C\Delta_i E|S^\pi_{t_i} - S_{t_i}|^2 + C\Big(\Delta_i E|Y^\pi_{t_{i+1}} - Y_{t_{i+1}}|^2 + \int_{t_i}^{t_{i+1}} E|Z^\pi_{t_i} - Z_u|^2 du\Big).$$

Thanks to the assumptions in the present lemma we have for the third summand the estimation $C\Delta_i^2$. Assumption 2 allows us to employ the regularity results on S and Y in Lemmas 3 and 2. Combining these steps yields (3.24).

Step 2: We will insert the equality

$$Y_{t_{i+1}} - Y_{t_i} = \int_{t_i}^{t_{i+1}} F(u, S_u, Y_u, Z_u)du + \int_{t_i}^{t_{i+1}} Z_u dW_u$$

in the summands of $\mathcal{E}_{\pi,0}^{loc}(Y^\pi, Z^\pi)$. Recall that

$$\begin{aligned} Y_{t_i} &= E[Y_{t_{i+1}} - \int_{t_i}^{t_{i+1}} F(u, S_u, Y_u, Z_u)du | \mathcal{F}_{t_i}], \\ Y_{t_i}^\pi &= E[Y_{t_{i+1}}^\pi - \Delta_i F(t_i, S_{t_i}^\pi, Y_{t_{i+1}}^\pi, Z_{t_i}^\pi) | \mathcal{F}_{t_i}]. \end{aligned} \tag{3.25}$$

The first equation arises from the formulation of the BSDE, the second from the backward scheme (2.2). Together with Young's inequality and Itô's isometry we get

$$E|Y_{t_i}^\pi - \left(Y_{t_{i+1}}^\pi - \Delta_i f^\pi(t_i, Y_{t_i}^\pi, Z_{t_i}^\pi)\right) + Z_{t_i}^\pi \Delta W_i|^2 \leqslant (I) + (II) + (III), \tag{3.26}$$

with

$$\begin{aligned} (I) &= 3E|Y_{t_i}^\pi - Y_{t_i} - \left(Y_{t_{i+1}}^\pi - Y_{t_{i+1}} - \int_{t_i}^{t_{i+1}} \Delta f_i^\pi(u)du\right)|^2, \\ (II) &= 3\Delta_i^2 E|f^\pi(t_i, Y_{t_i}^\pi, Z_{t_i}^\pi) - f^\pi(t_i, Y_{t_{i+1}}^\pi, Z_{t_i}^\pi)|^2, \\ (III) &= 3\int_{t_i}^{t_{i+1}} E|Z_{t_i}^\pi - Z_s|^2 ds. \end{aligned}$$

In view of (3.25) we work out the quadratic term of summand (I) under consideration of the rules for conditional expectations. Thus,

$$\begin{aligned} (I) \leqslant\ & 3E|Y_{t_{i+1}}^\pi - Y_{t_{i+1}} - \int_{t_i}^{t_{i+1}} \Delta f_i^\pi(u)du|^2 \\ & - 3E|E[Y_{t_{i+1}}^\pi - Y_{t_{i+1}} - \int_{t_i}^{t_{i+1}} \Delta f_i^\pi(u)du | \mathcal{F}_{t_i}]|^2. \end{aligned}$$

The definition of $Y_{t_i}^\pi - Y_{t_i}$ yields

$$\begin{aligned} (I) &\leqslant 3E|Y_{t_{i+1}}^\pi - Y_{t_{i+1}} - \int_{t_i}^{t_{i+1}} \Delta f_i^\pi(u)du|^2 - 3E|Y_{t_i}^\pi - Y_{t_i}|^2 \\ &\leqslant 3(1+\Delta_i)\left(E|Y_{t_{i+1}}^\pi - Y_{t_{i+1}}|^2 + \int_{t_i}^{t_{i+1}} E|\Delta f_i^\pi(u)|^2 du\right) - 3E|Y_{t_i}^\pi - Y_{t_i}|^2, \end{aligned}$$

where the last step followed by Young's inequality and concerning the integral also by Jensen's inequality. Thanks to the Lipschitz condition on F we have

$$\begin{aligned} (II) &\leqslant 3\kappa^2\Delta_i^2 E|Y_{t_i}^\pi - Y_{t_{i+1}}^\pi|^2 \leqslant C\Delta_i^2\big(\max_{0\leqslant i\leqslant N} E|Y_{t_i}^\pi - Y_{t_i}|^2 + E|Y_{t_i} - Y_{t_{i+1}}|^2\big) \\ &\leqslant C\Delta_i^2 \max_{0\leqslant i\leqslant N} E|Y_{t_i}^\pi - Y_{t_i}|^2 + C\Delta_i^3 + C\Delta_i^2 \int_{t_i}^{t_{i+1}} E|Z_u|^2 du, \end{aligned}$$

where we again made use of Lemma 2. Summing (3.26) up from $i = 0$ to $N-1$ and considering (3.24) leads to

$$\mathcal{E}^{loc}_{\pi,0}(Y^\pi, Z^\pi) \leqslant C\big(\max_{0\leqslant i\leqslant N} E|Y^\pi_{t_i} - Y_{t_i}|^2 + \int_0^T E|Z^\pi_{t_i} - Z_s|^2 ds\big) + C|\pi| \int_0^T E|Z_u|^2 du + C|\pi|.$$

Applying the assumption on the terminal condition and $\int_0^T E|Z_u|^2 du < \infty$ yields

$$\mathcal{E}^{loc}_{\pi,0}(Y^\pi, Z^\pi) \leqslant C\big(\max_{0\leqslant i\leqslant N} E|Y^\pi_{t_i} - Y_{t_i}|^2 + \int_0^T E|Z^\pi_{t_i} - Z_s|^2 ds\big) + C|\pi|.$$

The result on the time discretization error by Lemor et al. (2006) completes the proof. □

Theorem 13. *Let Assumptions 2 and 3 be fulfilledfor the setting in* (1.3)*. Suppose further there exists a constant such that*

$$E|\xi - \xi^\pi|^2 \leqslant const.|\pi|.$$

Then there is a constant $C > 0$ *depending on* s_0*,* κ*,* T*,* $\widetilde{D}$ *and* D *such that for every* $j = 0, \dots, N-1$

$$\sum_{i=j}^{N-1} E|\mathcal{P}_{0,i}(Y^\pi_{t_i}) - Y^\pi_{t_i}|^2 + \sum_{d=1}^{D}\sum_{i=j}^{N-1} \Delta_i E|\mathcal{P}_{d,i}(Z^\pi_{d,t_i}) - Z^\pi_{d,t_i}|^2 \geqslant C\mathcal{E}^{loc}_{\pi,j}(Y^{\pi,K}, Z^{\pi,K}) - |\pi|,$$

where $(Y^{\pi,K}, Z^{\pi,K})$ *denotes the pair* $(Y^{\pi,K_{0,i}}_{t_i}, Z^{\pi,K_i}_{t_i})_{t_i\in\pi}$*.*

Theorem 13 provides a lower bound on the error between the time-discretized solution and the unknown best approximation of the discretized solution in terms of the function basis. A large summand in the local error criterion suggests that the choice of the basis functions at this time step may be unsuccessful. In particular, for $i = N-1$ we get

$$E|\mathcal{P}_{0,N-1}(Y^\pi_{t_{N-1}}) - Y^\pi_{t_{N-1}}|^2 + \Delta_{N-1}\sum_{d=1}^{D} E|\mathcal{P}_{d,N-1}(Z^\pi_{d,t_{N-1}}) - Z^\pi_{d,t_{N-1}}|^2 \geqslant C\mathcal{E}^{loc}_{\pi,N-1}(Y^{\pi,K}, Z^{\pi,K}) - |\pi|.$$

Proof. Recall that within the explicit time discretization scheme (2.2) the generator F is applied on the vector $(t_i, S^\pi_{t_i}, Y^\pi_{t_{i+1}}, Z^\pi_{t_i})$ in the case of computable conditional expectations and on $(t_i, S^\pi_{t_i}, Y^{\pi,K_{0,i+1}}_{t_{i+1}}, Z^{\pi,K_i}_{t_i})$, when conditional expectations are estimated. Hence, we have to adapt the local criterion concerning the time points, at which the Y-processes are evaluated. Therefore, we abbreviate

$$\Delta f^\pi_i := f^\pi(t_i, Y^\pi_{t_{i+1}}, Z^\pi_{t_i}) - f^\pi(t_i, Y^{\pi,K_{0,i+1}}_{t_{i+1}}, Z^{\pi,K_i}_{t_i})$$

and define

$$A_i := Y^{\pi,K_{0,i+1}}_{t_{i+1}} - Y^{\pi}_{t_{i+1}} + \Delta_i \Delta f^{\pi}_i.$$

The orthogonal projections $\mathcal{P}_{0,i}$ are mappings on a subspace of $L^2(\mathcal{F}_{t_i})$. We have,

$$\begin{aligned} \mathcal{P}_{0,i}\left(Y^{\pi}_{t_i}\right) &= \mathcal{P}_{0,i}\big(E[Y^{\pi}_{t_{i+1}} - \Delta_i f^{\pi}(t_i, Y^{\pi}_{t_{i+1}}, Z^{\pi}_{t_i}) | \mathcal{F}_{t_i}]\big) \\ &= \mathcal{P}_{0,i}\big(Y^{\pi}_{t_{i+1}} - \Delta_i f^{\pi}(t_i, Y^{\pi}_{t_{i+1}}, Z^{\pi}_{t_i})\big). \end{aligned} \tag{3.27}$$

After adding a zero we employ Young's inequality and receive

$$\begin{aligned} \mathcal{E}^{loc}_{\pi,j}(Y^{\pi,K}, Z^{\pi,K}) \leqslant 3 \sum_{i=j}^{N-1} & E|Y^{\pi,K_{0,i+1}}_{t_{i+1}} - Y^{\pi}_{t_{i+1}} - (Y^{\pi,K_{0,i}}_{t_i} - Y^{\pi}_{t_i}) \\ & - \Delta_i\big(f^{\pi}(t_i, Y^{\pi,K_{0,i}}_{t_i}, Z^{\pi,K_i}_{t_i}) - f^{\pi}(t_i, Y^{\pi}_{t_i}, Z^{\pi}_{t_i})\big)|^2 \\ & + 3 \sum_{i=0}^{N-1} E|Y^{\pi}_{t_i} - Y^{\pi}_{t_{i+1}} + \Delta_i f^{\pi}(t_i, Y^{\pi}_{t_i}, Z^{\pi}_{t_i}) + Z^{\pi}_{t_i}\Delta W_i|^2 \\ & + 3 \sum_{i=j}^{N-1} E|(Z^{\pi,K_i}_{t_i} - Z^{\pi}_{t_i})\Delta W_i|^2 \\ =: & B_j + (I) + (II). \end{aligned}$$

Due to Lemma 12, summand $(I) \leqslant C|\pi|$. Now, we use the relation in (3.27) to add again a zero. By Young's inequality follows

$$\begin{aligned} B_j \leqslant & C \sum_{i=j}^{N-1} E|\mathcal{P}_{0,i}\big(A_i\big) - A_i|^2 + C \sum_{i=j}^{N-1} E|\mathcal{P}_{0,i}\big(Y^{\pi}_{t_i}\big) - Y^{\pi}_{t_i}|^2 + C \sum_{i=j}^{N-1} \Delta_i^2 E|\Delta f^{\pi}_i|^2 \\ & + C \sum_{i=j}^{N-1} \Delta_i^2 E|f^{\pi}(t_i, Y^{\pi}_{t_i}, Z^{\pi}_{t_i}) - f^{\pi}(t_i, Y^{\pi,K_{0,i}}_{t_i}, Z^{\pi,K_i}_{t_i})|^2 \\ = & (III) + (IV) + (V) + (VI). \end{aligned}$$

The Lipschitz condition on F yields

$$(V) + (VI) \leqslant C|\pi| \left(\max_{j \leqslant i \leqslant N} E|Y^{\pi}_{t_i} - Y^{\pi,K_{0,i}}_{t_i}|^2 + \sum_{i=j}^{N-1} \Delta_i E|Z^{\pi}_{t_i} - Z^{\pi,K_i}_{t_i}|^2 \right).$$

Thanks to the definitions of $Y^{\pi,K_{0,i}}_{t_i}$ and $Y^{\pi}_{t_i}$ the following equality holds true for all $i = 0, \dots, N-2$:

$$A_i = P_{0,i+1}(A_{i+1}) + \mathcal{P}_{0,i+1}\big(Y^{\pi}_{t_{i+1}}\big) - Y^{\pi}_{t_{i+1}} + \Delta f^{\pi}_i \Delta_i. \tag{3.28}$$

Due to the orthogonality of $\mathcal{P}_{0,i}$ we have

$$E[\mathcal{P}_{0,i}\,(A_i)\,A_i] = E|\mathcal{P}_{0,i}\big(A_i\big)|^2$$

and consequently

$$\sum_{i=j}^{N-1} E|\mathcal{P}_{0,i}(A_i) - A_i|^2 = \sum_{i=j}^{N-1} E|A_i|^2 - E|\mathcal{P}_{0,i}(A_i)|^2.$$

The following calculation takes place in view of (3.28), the orthogonality of the projections and the equality $Y_{t_N}^{\pi,K_{0,N}} - Y_{t_N}^{\pi} = 0$.

$$\begin{aligned}(III) &\leqslant (1+\Delta_{N-1})E|Y_{t_N}^{\pi,K_{0,N}} - Y_{t_N}^{\pi}|^2 \\ &\quad + \sum_{i=j}^{N-2} (1+\Delta_i)E\left[|\mathcal{P}_{0,i+1}(A_{i+1})|^2 + |\mathcal{P}_{0,i+1}(Y_{t_{i+1}}^{\pi}) - Y_{t_{i+1}}^{\pi}|^2\right] \\ &\quad + \sum_{i=j}^{N-1} (1+\Delta_i)\Delta_i E|\Delta f_i^{\pi}|^2 - \sum_{i=j}^{N-1} E|\mathcal{P}_{0,i}(A_i)|^2 \\ &\leqslant C\sum_{i=j}^{N-2} \left(\Delta_i E|Y_{t_{i+2}}^{\pi,K_{0,i+2}} - Y_{t_{i+2}}^{\pi}|^2 + E|P_{0,i+1}(Y_{t_{i+1}}^{\pi}) - Y_{t_{i+1}}^{\pi}|^2\right) \\ &\quad + C\sum_{i=j}^{N-1} \Delta_i E|\Delta f_i^{\pi}|^2.\end{aligned}$$

Because of the Lipschitz condition on F we get

$$\begin{aligned}(III) &\leqslant C\big(\max_{j\leqslant i\leqslant N} E|Y_{t_i}^{\pi,K_{0,i}} - Y_{t_i}^{\pi}|^2 + \sum_{i=j}^{N-1} \Delta_i E|Z_{t_i}^{\pi,K_i} - Z_{t_i}^{\pi}|^2\big) \\ &\quad + C\sum_{i=j}^{N-2} E|P_{0,i+1}(Y_{t_{i+1}}^{\pi}) - Y_{t_{i+1}}^{\pi}|^2.\end{aligned}$$

In sum, we achieve

$$\begin{aligned}\mathcal{E}_{\pi,j}^{loc}(Y^{\pi,K}, Z^{\pi,K}) &\leqslant C\big(\max_{j\leqslant i\leqslant N} E|Y_{t_i}^{\pi,K_{0,i}} - Y_{t_i}^{\pi}|^2 + \sum_{i=j}^{N-1} \Delta_i E|Z_{t_i}^{\pi,K_i} - Z_{t_i}^{\pi}|^2\big) \\ &\quad + C\sum_{i=j}^{N-1} E|\mathcal{P}_{0,i}(Y_{t_i}^{\pi}) - Y_{t_i}^{\pi}|^2 + C|\pi|.\end{aligned}$$

Finally, we obtain by employing Lemma 5 the proof. □

3.3 Non-linear control variates for BSDEs

In this section we propose a method for reducing the approximation error within least-squares Monte Carlo under suitable assumptions. Precisely, we suggest to split the original BSDE into the sum of two BSDEs and assume that one of them can be solved in closed form and only the other one requires numerical approximation. We call this procedure non-linear control variate inspired by the variance reduction technique for simulating expectations. The original BSDE is given by

$$Y_t = \xi - \int_t^T f(s, Y_s, Z_s)ds - \int_t^T Z_s dW_s. \tag{3.29}$$

Instead of (3.29), we examine the following BSDEs:

$$\tilde{Y}_t = \xi - \int_t^T \tilde{f}(s, \tilde{Y}_s, \tilde{Z}_s)ds - \int_t^T \tilde{Z}_s dW_s,$$
$$Y_t^{\mathcal{V}} = -\int_t^T \big(f(s, Y_s^{\mathcal{V}} + \tilde{Y}_s, Z_s^{\mathcal{V}} + \tilde{Z}_s) - \tilde{f}(s, \tilde{Y}_s, \tilde{Z}_s)\big)ds - \int_t^T Z_s^{\mathcal{V}} dW_s,$$

where $\mathcal{V}$ denotes the application of a control variate. Then, we receive the solution (Y, Z) of (3.29) by adding $(\tilde{Y}, \tilde{Z})$ and $(Y^{\mathcal{V}}, Z^{\mathcal{V}})$. Note that Gobet and Makhlouf (2010) made use of this decomposition in their proof of the L^2-regularity of Z in cases of irregular terminal conditions. Concerning $(Y^{\mathcal{V}}, Z^{\mathcal{V}})$ we employ least-squares Monte Carlo, see Section 2.2.

Example 14. Think of an European option pricing problem with pay-off function ξ and non-linear driver f. Typically, the non-linearity of f is 'small' compared to the terminal condition. In many settings the BSDE

$$\tilde{Y}_t = \xi - \int_t^T \tilde{Z}_s dW_s$$

has closed-form solutions or very accurate approximations. So, heuristically, the 'main' part $(\tilde{Y}, \tilde{Z})$ of the solution is correctly or almost correctly computed and only a small part, here $(Y^{\mathcal{V}}, Z^{\mathcal{V}})$, is affected by approximation errors.

3.4 Numerical examples

3.4.1 A non-linear decoupled FBSDE with known closed-form solution

We begin with a modification of an example in Bender and Zhang (2008) that is solvable in closed form as far as (Y, Z) is concerned. That enables us in a way to compare the Monte Carlo estimates on the global a-posteriori criterion and the approximation error for some given approximation. Concretely, we consider

$$S_{d,t} = s_{d,0} + \int_0^t \sigma\Big(\sum_{d'=1}^D \sin(S_{d',u})\Big)dW_{d,u}, \quad d = 1, \dots, D$$
$$Y_t = \sum_{d=1}^D \sin(S_{d,T}) + \int_t^T \frac{1}{2}\sigma^2(Y_u)^3 du - \sum_{d=1}^D \int_t^T Z_{d,u} dW_{d,u},$$

where $W = (W_1, \dots, W_D)$ is a D-dimensional Brownian motion and $\sigma > 0$ and $s_{d,0}$, $d = 1, \dots, D$ are constants. The true solution for (Y, Z) is given by

$$Y_t = \sum_{d=1}^{D} \sin(S_{d,t}), \quad Z_{d,t} = \sigma \cos(S_{d,t})\Big(\sum_{d'=1}^{D} \sin(S_{d',t})\Big), \quad d = 1, \dots, D,$$

which can be verified by Itô' s formula. But there is no closed-form solution for S. Therefore, we will incorporate the Euler or the Miltstein scheme to obtain an approximation S^π. Since the terminal condition is not path-dependent we can refrain here from constructing an extra Markov chain, as described in Subsection 2.2.1 and simply set $X^\pi = S^\pi$. For the approximate solution of (Y, Z) we intend to use then least-squares Monte Carlo as explained in Section 2.2. This requires, however the Lipschitz continuity of the driver. Let $[\cdot]_R$ be a truncation function such that

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

for some constant $R > 0$ that will be replaced by suitable values as the case may be. Instead of approximating (Y, Z), we will generate numerical solutions for

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

where $\mathcal{T}$ indicates the BSDE with truncated driver.

Case 1: One-dimensional Brownian motion and indicator function bases

In the first case we fix the parameters by

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

Drawing samples of $X_{t_N}^\pi = S_{t_N}^\pi$ shows that they are primarily located in the interval $[0, 3]$. Hence, let $K \geqslant 3$ be the dimension of the function bases $\eta(i, x)$, that is composed of indicator functions of equidistant partial intervals of $[0, 3]$ for all $i = 0, \dots, N-1$. Clearly, we set

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

for $i = 0, \dots, N-1$. The simulation parameter consist of the number of time steps N, the dimension of the function bases K and the sample size L. For $m = 1, \dots, 11$ and $l = 3, \dots, 5$ they are fixed by

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

where $[a]$ is the closest integer to a and $\lceil a \rceil$ is the closest upper integer to a. To be precise, we will observe three different choices of l, in which we simultaneously increase the parameters N, K and L through their dependence on m. For a better distinction of the simulation results we will denote the partitions by π_N.

The main advantage of indicator function bases is the possibility to control the projection error through the choice of the dimension K. According to the explanations in Subsection 2.2.5, the above definition yields a convergence rate for the corresponding L^2-error of order 1/2 in the number of times steps. However, this basis choice is also connected with a severe drawback. Recalling the remarks in Subsection 2.2.5 on the simulation error, the theoretical convergence threshold is located at $l = 4$. The L^2-error due to simulation theoretically decreases with rate $N^{-1/2}$ when the sample size L grows proportional to N^3K^2, which is satisfied for $l = 5$. Hence, the growing dimension K blows the required sample size much more up than a constant choice for K would. Keep in mind, that enlarging the sample size L leads to increasing computational cost. For a better illustration, see the absolute values of L in dependence of m and l in the below table.

Table 3.1: Sample size L in dependence of m and l

	m	1	2	3	4	5	6	7	8	9	10	11
	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

Given these parameters, we initialize the approximation by $\hat{Y}^{\mathcal{T},\pi_N}_{t_i} = \sin(S^{\pi_N}_{t_i})$ and compute the coefficients $\hat{\alpha}^{\mathcal{T},\pi_N}_{0,i}$ and $\hat{\alpha}^{\mathcal{T},\pi_N}_{1,i}$ for the linear combination of the basis functions by least-squares Monte Carlo and receive the approximate solution by setting

$$\hat{Y}^{\mathcal{T},\pi_N}_{t_i} = \eta(i, X^{\pi_N}_{t_i})\hat{\alpha}^{\mathcal{T},\pi_N}_{0,i}, \quad \hat{Z}^{\mathcal{T},\pi_N}_{t_i} = \eta(i, X^{\pi_N}_{t_i})\hat{\alpha}^{\mathcal{T},\pi_N}_{1,i}.$$

As S cannot be sampled perfectly, we measure the squared approximation error by

$$\max_{0\leqslant i\leqslant N} E|\sin(S^{\pi_N,MS}_{t_i}) - \hat{Y}^{\mathcal{T},\pi_N}_{t_i}|^2 + \sum_{i=1}^{N-1} \frac{T}{N} E|\sigma\cos(S^{\pi_N,MS}_{t_i})\sin(S^{\pi_N,MS}_{t_i}) - \hat{Z}^{\mathcal{T},\pi_N}_{t_i}|^2, \quad (3.30)$$

where $S^{\pi_N,MS}$ denotes the approximation of S by the Milstein scheme. This error term is equivalent to

$$\max_{0\leqslant i\leqslant N} E|Y_{t_i} - \hat{Y}^{\mathcal{T},\pi_N}_{t_i}|^2 + \sum_{i=1}^{N-1} \frac{T}{N} E|Z_{t_i} - \hat{Z}^{\mathcal{T},\pi_N}_{t_i}|^2$$

up to terms of order $|\pi_N|^2$, as the L^2-error between S and $S^{\pi_N,MS}$ decreases with rate $|\pi_N|$ rather than $|\pi_N|^{1/2}$ as in the Euler scheme. Note that $\xi^\pi = \sin(S^{\pi_N,MS}_{t_N})$

and $f^{\pi}(t_i, y, z) = -\frac{1}{2}\sigma^2[y^3]_1$. According to Subsection 3.1.4 the global a-posteriori criterion $\mathcal{E}_{\pi_N}(\hat{Y}^{\mathcal{J},\pi_N}, \hat{Z}^{\mathcal{J},\pi_N})$ satisfies the inequalities

$$\max_{0 \leqslant i \leqslant N-1} \sup_{t_i \leqslant t < t_{i+1}} E[|Y_t - \hat{Y}_{t_i}^{\mathcal{J},\pi_N}|^2 | \mathcal{G}_{t_0}] + \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} E[|Z_t - \hat{Z}_{t_i}^{\mathcal{J},\pi_N}|^2 | \mathcal{G}_{t_0}] dt$$
$$\leqslant C(\mathcal{E}_{\pi_N}(\hat{Y}^{\mathcal{J},\pi_N}, \hat{Z}^{\mathcal{J},\pi_N}) + |\pi_N|)$$

and

$$\max_{0 \leqslant i \leqslant N-1} \sup_{t_i \leqslant t < t_{i+1}} E[|Y_t - \hat{Y}_{t_i}^{\mathcal{J},\pi_N}|^2 | \mathcal{G}_{t_0}] + \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} E[|Z_t - \hat{Z}_{t_i}^{\mathcal{J},\pi_N}|^2 | \mathcal{G}_{t_0}] dt$$
$$\geqslant \frac{1}{c}\mathcal{E}_{\pi_N}(\hat{Y}^{\mathcal{J},\pi_N}, \hat{Z}^{\mathcal{J},\pi_N}) - |\pi_N|^2.$$

Thus, in case $\mathcal{E}_{\pi_N}(\hat{Y}^{\mathcal{J},\pi_N}, \hat{Z}^{\mathcal{J},\pi_N}) \geqslant$ const. $(1/N)$ the global error criterion is equivalent to the squared approximation error. For the estimation of both the criterion and the error term (3.30) we draw 1000N copies of the increments of the Brownian motion, denoted by $(\Delta W_i)_{i=0,\ldots,N-1}$, and generate thereby samples of $X^{\pi_N} = S^{\pi_N}$ and $S^{\pi_N, MS}$.

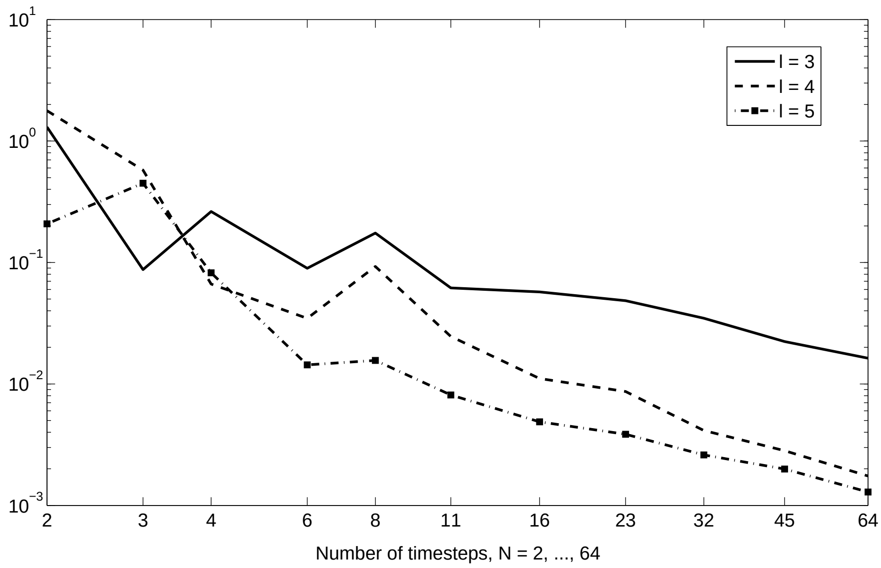

Figure 3.1: Development of the global a-posteriori criterion in Case 1

Figure 3.1 shows the estimated global a-posteriori criterion and in Figure 3.2 we can see the estimated squared approximation error. In both figures the different paths correspond to the cases $l = 3, \ldots, 5$ with simultaneously growing number of time steps, dimension of function bases and sample size as described above. The horizontal as well as the vertical axes are chosen logarithmically for a better illustration of the results. A comparison of these figures reveals that the a-posteriori criterion neatly reflects the convergence behaviour of the approximation error. In this example, also the absolute values of the criterion and the squared approximation error almost coincide.

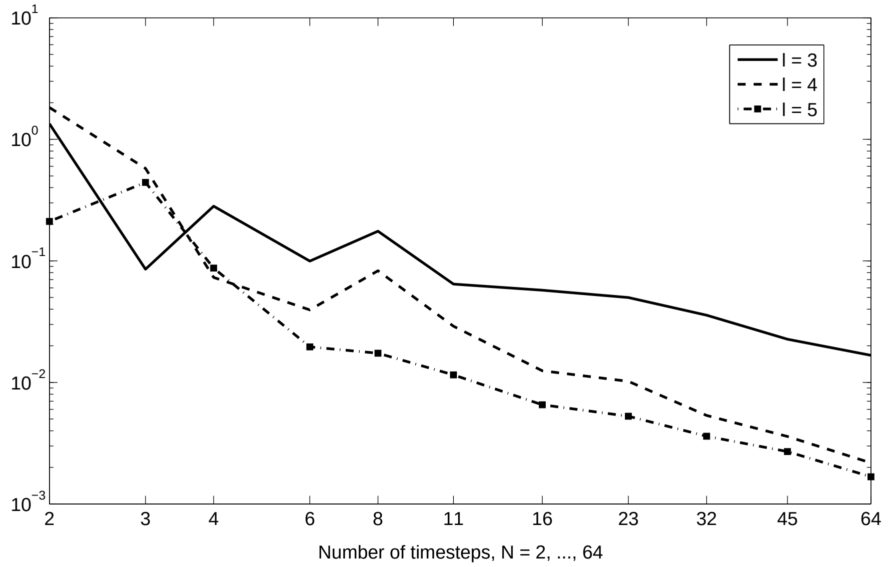

Figure 3.2: Development of the squared approximation error in Case 1

Contrary to the theoretical results the global a-posteriori criterion tends to zero in all three cases for l. Considering the results for $N = 32, 45, 64$ we receive an empirical convergence rate of -1.09 for $l = 3$, -1.25 for $l = 4$ and -1.02 for $l = 5$. Hence, only the expensive example ($l = 5$) matches the theoretical results as described above. Nevertheless, the levels of the three paths demonstrate the connection between sample size L and approximation error. Neglecting the simulations with only few time steps, we can see that larger values for m and thereby higher computational cost lead to smaller approximation errors. However, the distance between the error criteria of $l = 4$ and $l = 5$ seems to vanish for a growing number of time steps. That means for N large enough the error level of the high-expensive case might as well be achieved by a simulation with smaller computational cost than determined by $l = 5$.

Case 2: Three-dimensional Brownian motion and polynomial function bases

In this example we also apply the method of non-linear control variates. To this end, we freeze the diffusion coefficient of S at time 0 and consider a rather simple case of decoupled FBSDEs, namely

$$\check{S}_{d,t} = S_{d,0} + \sum_{d=1}^{D} \sin(s_{d,0})\sigma W_{d,t}, \quad d = 1, \dots, D,$$
$$\check{Y}_t = \sum_{d=1}^{D} \sin(\check{S}_{d,T}) - \int_t^T \check{Z}_u dW_u.$$

The process $\check{Y}_t$ can easily be obtained in closed form. Precisely,

$$\begin{aligned}\check{Y}_t &= \sum_{d=1}^{D} E[\sin\left(\check{S}_{d,t} + \sigma(W_{d,T} - W_{d,t})\right)] \\ &= \exp\Big\{ -\frac{1}{2}\sigma^2 \Big(\sum_{d=1}^{D} \sin(s_{d,0}) \Big)^2 (T-t) \Big\} \sum_{d=1}^{D} \sin\left(\check{S}_{d,t}\right) =: u(t, \check{S}_t).\end{aligned}$$

This result inspires to figure out $\tilde{Y}_t := u(t, S_t)$ and define thereby the non-linear control variate. For the sake of convenience we abbreviate

$$g(t) = \exp\Big\{ -\frac{1}{2}\sigma^2 \Big(\sum_{d=1}^{D} \sin(s_{d,0}) \Big)^2 (T-t) \Big\}.$$

The application of Itô' s formula yields

$$\begin{aligned}\tilde{Y}_t &= \sum_{d=1}^{D} \sin(S_{d,T}) \\ &\quad - \frac{1}{2}\sigma^2 \int_t^T g(u) \big(\sum_{d=1}^{D} \sin(S_{d,u}) \big) \Big(\big(\sum_{d=1}^{D} \sin(s_{d,0}) \big)^2 - \big(\sum_{d=1}^{D} \sin(S_{d,u}) \big)^2 \Big) du \\ &\quad - \sum_{d=1}^{D} \int_t^T g(u) \sigma \cos(S_{d,u}) \big(\sum_{d'=1}^{D} \sin(S_{d',u}) \big) dW_{d,u} \\ &= \sum_{d=1}^{D} \sin(S_{d,T}) - \frac{1}{2}\sigma^2 \int_t^T \tilde{Y}_u \Big(\big(\sum_{d=1}^{D} \sin(s_{d,0}) \big)^2 - \big(\sum_{d=1}^{D} \sin(S_{d,u}) \big)^2 \Big) du \\ &\quad - \sum_{d=1}^{D} \int_t^T \tilde{Z}_{d,u} dW_{d,u},\end{aligned}$$

with

$$\tilde{Z}_{d,t} = g(t)\sigma \cos(S_{d,t}) \big(\sum_{d'=1}^{D} \sin(S_{d',t}) \big), \quad d = 1 \dots, D.$$

Hence, there is a BSDE, that has the same terminal condition as the original one and is solvable in closed form. As described in Section 3.3 it remains now to approximate the residual BSDE

$$Y_t^{\mathcal{V}} = \int_t^T \frac{1}{2}\sigma^2 \Big\{ [(\tilde{Y}_u + Y_u^{\mathcal{V}})^3]_{D^3} + \tilde{Y}_u \Big(\big(\sum_{d=1}^{D} \sin(s_{d,0}) \big)^2 - \big(\sum_{d=1}^{D} \sin(S_{d,u}) \big)^2 \Big) \Big\} du - \sum_{d=1}^{D} \int_t^T Z_{d,u}^{\mathcal{V}} dW_{d,u}.$$

The upper index $\mathcal{V}$ refers to the application of non-linear control variates. As concrete parameters of the BSDE we choose

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

For the construction of function bases we use this time polynomials. Clearly,

$$\eta_1(i,x) = 1, \quad \eta_k(i,x) = x_{k-1}, \quad k = 2, \dots, 4$$
$$\eta_k(i,x) = x_{k-4}x_j, \quad (k,j) \in \{(5,2),(6,3),(7,1)\},$$

for $i = 0, \dots, N-1$ and $d = 0, 1$. Thus, the bases are again identical for all $d = 0, \dots, 3$. Following the analysis in 2.2.5 we fix the simulation parameters for $m = 1, \dots, 15$ by

$$N = \left[2\sqrt{2}^{m-1} \right], \quad K = 7, \quad L = \left[2\sqrt{2}^{3(m-1)} \right],$$

which corresponds to a simulation error that decreases with rate $N^{-1/2}$. Exploiting least-squares Monte Carlo both for the approximation of $(Y^{\mathcal{T}}, Z^{\mathcal{T}})$ and $(Y^{\mathcal{V}}, Z^{\mathcal{V}})$ gives the numerical solutions

$$\hat{Y}_{t_i}^{\mathcal{T},\pi_N} = \eta(i, X_{t_i}^{\pi_N})\hat{\alpha}_{0,i}^{\mathcal{T},\pi_N}, \quad \hat{Z}_{d,t_i}^{\mathcal{T},\pi_N} = \eta(i, X_{t_i}^{\pi_N})\hat{\alpha}_{d,i}^{\mathcal{T},\pi_N}, \quad d = 1, \dots, 3,$$
$$\hat{Y}_{t_i}^{\mathcal{V},\pi_N} = \eta(i, X_{t_i}^{\pi_N})\hat{\alpha}_{0,i}^{\mathcal{V},\pi_N}, \quad \hat{Z}_{d,t_i}^{\mathcal{V},\pi_N} = \eta(i, X_{t_i}^{\pi_N})\hat{\alpha}_{d,i}^{\mathcal{V},\pi_N}, \quad d = 1, \dots, 3.$$

Based on these results we estimate the global a-posteriori criteria $\mathcal{E}_{\pi_N}(\hat{Y}^{\mathcal{T},\pi_N}, \hat{Z}^{\mathcal{T},\pi_N})$ and $\mathcal{E}_{\pi_N}(\hat{Y}^{\mathcal{V},\pi_N} + \tilde{Y}, \hat{Z}^{\mathcal{V},\pi_N} + \tilde{Z})$ by Monte Carlo simulation, for that we use 1000N samples of $X^{\pi_N} = S^{\pi_N}$. In contrast to the previous example the approximate terminal condition is this time based on the Euler scheme, namely $\xi^\pi = \sum_{d=1}^{3} \sin(S_{d,T}^{\pi})$. Figure 3.3 allows a comparison of the estimated criteria. Again both axes are logarithmic.

In the original least-squares Monte Carlo approach we can observe for small values of N that the criterion decreases faster than N^{-1}, whereas from $N = 64$ the reduction rate gets significantly smaller than N^{-1}. At $N = 256$ the error criterion settles down at about 0.03. Following the theoretical results, the contribution of the squared time discretization error and the squared simulation error should tend to zero with rate 1 in the number of time steps. Hence, the over all approximation error must be mainly determined by the non-converging projection error.

For $N = 256$ we have now a closer look on the projection error. Therefore, we evaluate the local criterion $\mathcal{E}_{\pi_N,j}^{loc}(\hat{Y}^{\mathcal{T},\pi_N}, \hat{Z}^{\mathcal{T},\pi_N})$ for $j = 0, \dots, 255$. Recall, that this

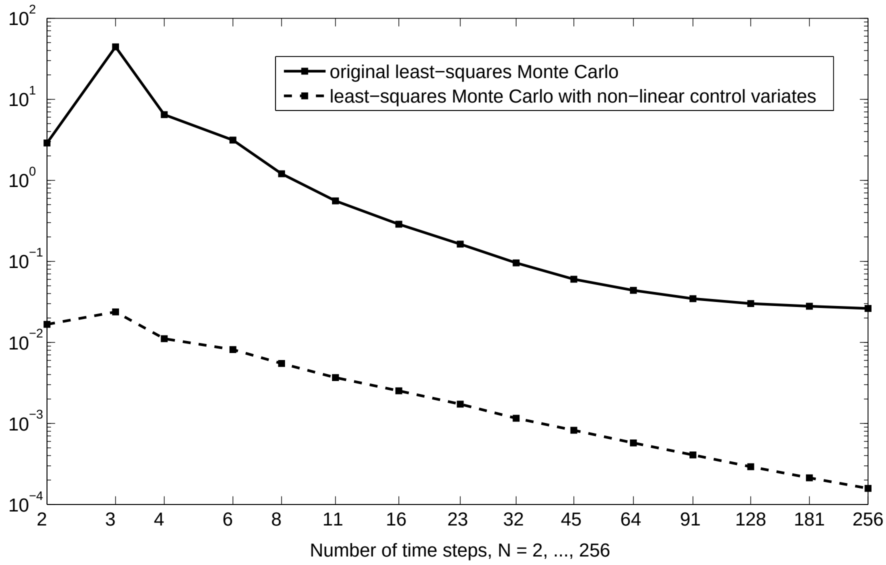

Figure 3.3: Development of the a-posteriori criterion in Case 2 - Original least-squares Monte Carlo vs. least-squares Monte Carlo with non-linear control variates

criterion is a sum over $i = j$ to $i = N - 1$. According to Section 3.2 the sum of the projection errors from $i = j$ to $N - 1$ is bounded from below by a constant times the local criterion less the negligible term $|\pi_N|$. The below Figure 3.4 shows that the local criterion amounts already at $j = 255$ to 0.025 and then increases nearly linearly for decreasing j. Finally, we end up at a criterion value of 0.026 at $j = 0$.

Hence, the results for the local criterion at $j \leqslant 255$ are primarily influenced by summand $i = 255$. This indicates that the projection error at time step $i = N-1 = 255$ has chief impact on the local criterion, whereas the projections at the remaining time steps of least-squares Monte Carlo make only minor contribution to this criterion. Thus, it takes a more suitable function basis a time step $i = 255$ for a reduction of the projection error. A first natural step would be the addition of $\sum_{d=1}^{3} \sin(x_d)$, as the absolute value of $\hat{Y}^{\mathcal{T},\pi_N}_{t_{N-1}}$ is mainly determined by the terminal condition.

Turning to the application of non-linear control variates, we can observe a global a-posteriori criterion that empirically decreases with rate 1.03 in the number of time steps. This matches rougly the theoretical convergence rate of both the squared time discretization and the squared simulation error. Theses error sources seem to dominate the over all approximation error, whereas the projection error has negligible influence up to $N = 256$. At $N = 256$ the global error criterion amounts only to about a 170th part of the value achieved with the original scheme.

Concerning the local criterion for $N = 256$, we observe that the estimation of $\mathcal{E}^{loc}_{\pi_N,j}(\hat{Y}^{\mathcal{V},\pi_N} + \tilde{Y}, \hat{Z}^{\mathcal{V},\pi_N} + \tilde{Z})$ totals $4.667 * 10^{-7}$ for $j = 255$ and increases up to 0.0001 for $j = 0$. In contrast to least-squares Monte Carlo without control variates, we cannot identify one particular time step whose projection error has major impact on

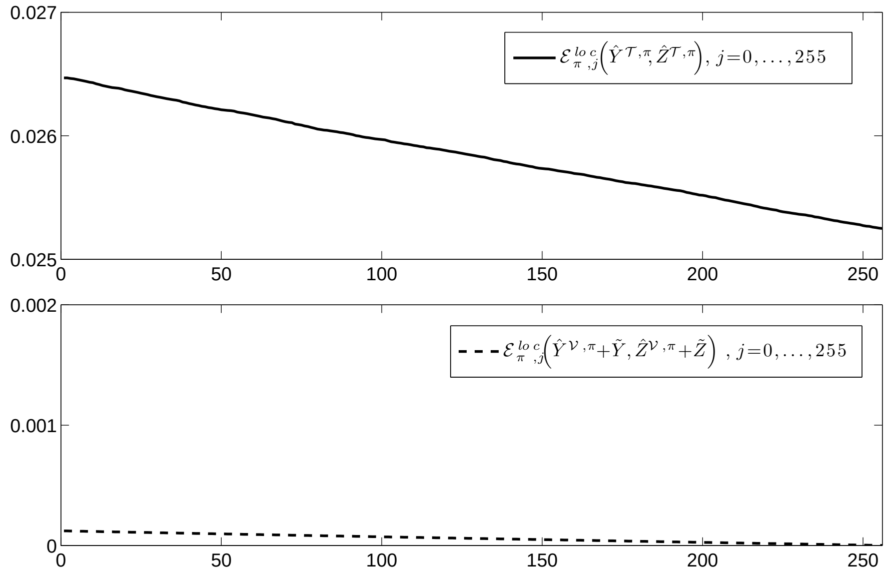

Figure 3.4: Development of the local criterion in Case 2 - Original least-squares Monte Carlo vs. least-squares Monte Carlo with non-linear control variates

the local criterion. This corresponds to the fact that here the terminal condition is not subject of estimation due to the application of non-linear control variates. For the approximation of $(Y^{\mathcal{V}}, Z^{\mathcal{V}})$ the chosen function bases seem to be suitable enough to achieve a small overall approximation error.

3.4.2 A non-linear option pricing problem

The last numerical example of this chapter deals with a non-linear option pricing problem that was already presented in Lemor et al. (2006). Precisely, we assume that the underlying stock price is modeled by a geometric Brownian motion according to Black-Scholes, i. e.

$$S_t = s_0 \exp\left\{\left(\mu - \sigma^2/2\right) t + \sigma W_t\right\},$$

with $\mu, \sigma > 0$ and W being a one-dimensional Brownian motion. We aim at finding the price process of an European call-spread option with pay-off

$$\phi(S_T) = (S_T - \kappa_1)_+ - 2\,(S_T - \kappa_2)_+,$$

where $\kappa_1, \kappa_2 > 0$ are strike prices. Thus, we can again set $X = S$ and $\xi^\pi = \xi = \phi(S_T)$. We also assume to act in a market with different interest rates for borrowing and lending. That means, we can invest money in riskless assets at rate $r > 0$, whereas bonds can be emitted at rate $R > r$. According to Bergman (1995), the dynamic of the price process is then described by

$$Y_t = \phi(S_T) - \int_t^T \left(rY_u + \frac{\mu - r}{\sigma} Z_u - (R-r) \left(Y_u - \frac{Z_u}{\sigma} \right)_+ \right) du - \int_t^T Z_u dW_u.$$

As concrete market parameters we choose

$$T = 0.25, \quad s_0 = 100, \quad r = 0.01, \quad R = 0.06, \quad \mu = 0.05, \quad \sigma = 0.2.$$

The strike prices are fixed with $\kappa_1 = 95$ and $\kappa_2 = 105$. The numerical solution will be obtained by least-squares Monte Carlo. For this purpose we define the function bases for $i = 0, \dots, N-1$

$$\eta_1(x) = (x - 95)_+ - 2(x - 105)_+,$$
$$\eta_2(x) = \mathbb{1}_{\{x<40\}}(x), \quad \eta_3(x) = \mathbb{1}_{\{x \geqslant 180\}}(x),$$
$$\eta_k(x) = \mathbb{1}_{\{x \in [40+140(k-1)/(K-3),\, 40+140k/(K-3))\}}(x), \quad k = 1, \dots, K-3,$$

where K is the dimension of the function bases. Again the bases are identical for $d = 0, 1$ within each time step. The simulation parameter grow depending on $m = 1, \dots, 10$ and $l = 3, \dots, 5$, clearly

$$N = \left\lceil 2\sqrt{2}^{m-1} \right\rceil, \quad K = \left\lceil 3\sqrt{2}^{m-1} \right\rceil + 1, \quad L = \left\lceil 2\sqrt{2}^{l(m-1)} \right\rceil.$$

See also the explanations concerning the basis choice in Case 1 of 3.4.1. Note, that this time the approximators are functions of $X = S$ and not X^π, since the geometric Brownian motion can be sampled perfectly. Given these specification, we receive by least-squares Monte Carlo the approximators for (Y, Z), that is

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

The global a-posteriori criterion $\mathcal{E}_\pi(\hat{Y}^\pi, \hat{Z}^\pi)$ is now estimated by drawing 1000N samples of $X = S$ and applying then Monte Carlo simulation. The results are shown in Figure 3.5.

Like before the three paths correspond to the different choices of l. Each path represents the estimated criterion for a simultaneously growing number of time steps N, dimension K and sample size L. Whereas the a-posteriori criterion does not seem to converge in the low-cost case $l = 3$, we have a growth rate of -1.09 in the expensive case $l = 5$. This is consistent with the theoretical results. Apart from that we observe that the criterion decreases with rate -1 for $l = 4$. Here, the numerical results turn out to behave better than the theory suggests. Nevertheless, the absolute values of the a-posteriori criterion proceed on a higher level for $l = 4$ than for $l = 5$. In case of 45 time steps we end up with a criterion value of 1.39 for the middle-cost simulation ($l = 4$) compared to 0.86 in the expensive case.

In the present example it might be prohibitive to turn the sample size of the expensive case any higher due to the computational complexity required by the evaluation of the pseudo-inverse of

$$\frac{1}{\sqrt{L}} \begin{pmatrix} \eta_1({}_\lambda X_{t_i}) & \eta_2({}_\lambda X_{t_i}) & \cdots & \eta_K({}_\lambda X_{t_i}) \end{pmatrix}_{\lambda=1,\dots,L},$$

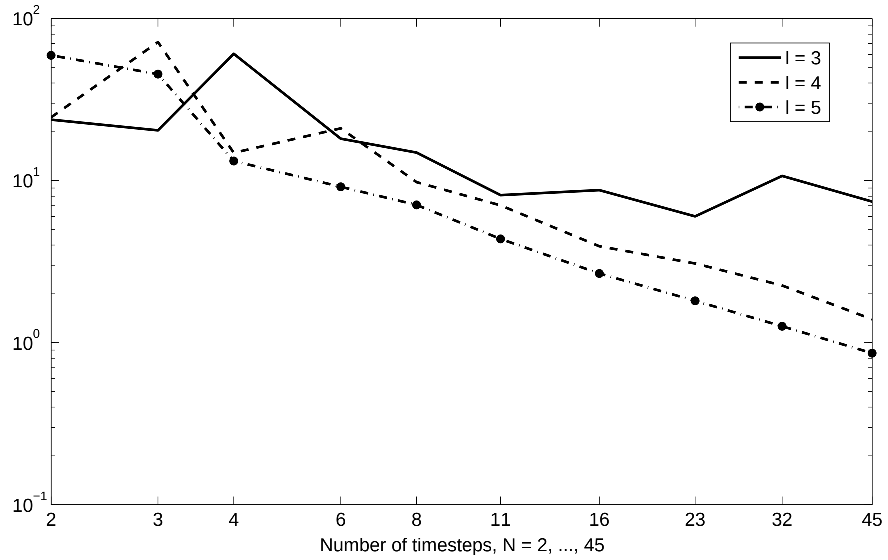

Figure 3.5: Development of the global a-posteriori criterion for a call-spread option

see also Subsection 2.2.2. Here, we have to deal with function bases that consist of the pay-off function and indicator functions. Thus the above matrix is generally not orthogonal. In contrast to that, the bases of Case 1 in Subsection 3.4.1 are composed by indicator functions only and thus the corresponding matrix used for least-squares Monte Carlo is orthogonal. Then the calculation of the pseudo-inverse in order to receive a solution of the minimization problem of type (2.6) can be avoided. Indeed, computing projections on orthogonal bases are connected with smaller computational complexity. For an overview of the absolute values of the sample size L we refer to Subsection 3.4.1.

4 Enhancing the least-squares MC approach by exploiting martingale basis functions

4.1 Construction of the simplified algorithm and examples for martingale bases

In subsection 2.2.2 we reviewed the least-squares Monte Carlo approach on estimating conditional expectations. The objective was to tackle the conditional expectations that appear in the time discretization scheme (2.3). Clearly, there are $(D+2)$ conditional expectations to be calculated in every time step, i. e.

$$E[\Delta W_{d,i} Y^{\pi}_{t_{i+1}} | X^{\pi}_{t_i}], \quad d = 1, \dots, D \tag{4.1}$$
$$E[Y^{\pi}_{t_{i+1}} | X^{\pi}_{t_i}], \tag{4.2}$$
$$E[F(t_i, S^{\pi}_{t_i}, Y^{\pi}_{t_{i+1}}, Z^{\pi}_{t_i}) | X^{\pi}_{t_i}]. \tag{4.3}$$

Our contribution is now to provide a certain structure such that (4.1) and (4.2) are computable in closed form and only (4.3) remains to be estimated via least-squares Monte Carlo.

Roughly speaking, we suppose that at time t_{i+1} an approximation $y^{\pi,K,L}_{i+1}(X^{\pi}_{t_{i+1}}) = Y^{\pi,K,L}_{t_{i+1}}$ of $Y^{\pi}_{t_{i+1}}$ is at hand such that $y^{\pi,K,L}_{i+1}(x)$ can be expressed as linear combination of basis functions $\eta_{0,k}(i+1,x)$, i. e.

$$y^{\pi,K,L}_{i+1}(x) = \sum_{k=1}^{K} \tilde{\alpha}_k \eta_{0,k}(i+1,x),$$

where K is the dimension of the function basis

$$\eta_0(i+1,x) = \{\eta_{0,1}(i+1,x), \dots, \eta_{0,K}(i+1,x)\}.$$

Note, that the dimension of the function bases stays constant over all time steps. Then, we assume that the basis functions form a system of martingales in the sense that for all $k = 1, \dots, K$

$$\begin{aligned} E[\eta_{0,k}(i+1, X^{\pi}_{t_{i+1}}) | X^{\pi}_{t_i}] &=: \eta_{0,k}(i, X^{\pi}_{t_i}), \\ E[\Delta W_{d,i} \eta_{0,k}(i+1, X^{\pi}_{t_{i+1}}) | X^{\pi}_{t_i}] &=: \eta_{d,k}(i, X^{\pi}_{t_i}), \quad d = 1, \dots, D. \end{aligned}$$

By this construction we receive for each $k = 1, \dots, K$ martingales $\big(\eta_{0,k}(i, X^{\pi}_{t_i})\big)_{0 \leqslant i \leqslant N}$. Because of this definition we have

$$E[\Delta W_{d,i} y^{\pi,K,L}_{i+1}(X^{\pi}_{t_{i+1}}) | X^{\pi}_{t_i}] = \sum_{k=1}^{K} \tilde{\alpha}_k \eta_{d,k}(i, X^{\pi}_{t_i}), \quad d = 1, \dots, D,$$
$$E[y^{\pi,K,L}_{i+1}(X^{\pi}_{t_{i+1}}) | X^{\pi}_{t_i}] = \sum_{k=1}^{K} \tilde{\alpha}_k \eta_{0,k}(i, X^{\pi}_{t_i}).$$

However, the non-linearity of F calls for the application of some estimator for the conditional expectation in (4.3). Like before, we choose for this purpose least-squares Monte Carlo. Before giving a complete description of the algorithm, we fix the necessary conditions for the martingale bases setting.

Assumption 6. *Let* $\eta_0(N, x) = \{\eta_{0,1}(N, x), \dots, \eta_{0,K}(N, x)\}$ *be a* K*-dimensional basis such that*

(a) $E[\eta_{0,k}(N, X^{\pi}_{t_{i+1}}) | X^{\pi}_{t_i} = x] =: \eta_{0,k}(i, x)$,

(b) $E[\Delta W_{d,i} \eta_{0,k}(N, X^{\pi}_{t_{i+1}}) | X^{\pi}_{t_i} = x] =: \eta_{d,k}(i, x)$

are computable in closed form for all $k = 1, \dots, K$ *and* $i = 0, \dots, N-1$. *Then we define the bases* $\eta_d(i, x)$ *by* $\{\eta_{d,1}(i, x), \dots, \eta_{d,K}(i, x)\}$, $d = 0, \dots, D$.

Now, we give a description of the algorithm. Similarly to Subsection 2.2.4, we make use of a set $\mathcal{X}^L$ of independent copies of $(X^{\pi}_{t_i})_{t_i \in \pi}$, precisely we define

$$\mathcal{X}^L = \{(\Delta_\lambda W_i, {}_\lambda X^{\pi}_{t_{i+1}}), i = 0, \dots, N-1, \lambda = 1, \dots, L\}.$$

First, we check if

$$E[\phi^{\pi}(X^{\pi}_{t_N}) | X^{\pi}_{t_i} = x], \quad E[\Delta W_{d,i} \phi^{\pi}(X^{\pi}_{t_N}) | X^{\pi}_{t_i} = x]$$

are available in closed form. If so, we add $\phi^{\pi}(x)$ to the function basis at time t_N. Otherwise we approximate $\phi^{\pi}(x)$ by a linear combination whose coefficients solve the minimization problem

$$\hat{\alpha}^{\pi,K,L}_N = \arg\min_{\alpha \in \mathbb{R}^K} \frac{1}{L} \sum_{\lambda=1}^{L} |\eta_0(N, {}_\lambda X^{\pi}_{t_N})\alpha - \phi^{\pi}({}_\lambda X^{\pi}_{t_N})|^2.$$

Whatever the case, we can proceed from the assumption that a coefficient vector $\hat{\alpha}^{\pi,K,L}_N$ has been chosen, either by perfect evaluation or by least-squares Monte Carlo estimation. Similarly as before we start with $\hat{y}^{\pi,K,L}_N(x) = \eta_0(N, x)\hat{\alpha}^{\pi,K,L}_N$ and repeat then for $i = N-1, \dots, 0$

$$\begin{aligned}
\hat{z}_{d,i}^{\pi,K,L}(x) &= \frac{1}{\Delta_i}\eta_d(i,x)\hat{\alpha}_{i+1}^{\pi,K,L}, \quad d = 1,\dots,D,\\
\bar{\alpha}_i^{\pi,K,L} &= \arg\min_{\alpha\in\mathbb{R}^K} \frac{1}{L}\sum_{\lambda=1}^{L} |\eta_0'(i,{}_\lambda X_{t_i}^\pi)\alpha \\
&\qquad\qquad - F(t_i, {}_\lambda S_{t_i}^\pi, \hat{y}_{i+1}^{\pi,K,L}({}_\lambda X_{t_{i+1}}^\pi), \hat{z}_i^{\pi,K,L}({}_\lambda X_{t_i}^\pi))|^2,\\
\hat{\alpha}_i^{\pi,K,L} &= \hat{\alpha}_{i+1}^{\pi,K,L} - \Delta_i\bar{\alpha}_i^{\pi,K,L},\\
\hat{y}_i^{\pi,K,L}(x) &= \eta_0(i,x)\hat{\alpha}_{i,k}^{\pi,K,L}.
\end{aligned} \tag{4.4}$$

The comparison of (4.4) with the original scheme in (2.17) shows that in the setting of Assumption 6 only the conditional expectations of type (4.3) have to be estimated via least-squares Monte Carlo. This point right away reveals a main advantage of the simplification. Particularly, in high-dimensional problems the computational effort is thereby reduced significantly (from $D+2$ estimations to one estimation only per time step).

Nevertheless, the remaining application of least-squares Monte Carlo related to (4.3) causes a projection error due to the basis choice and a simulation error. Similar to the original scheme in Lemor et al. (2006), the simplified least-squares Monte Carlo scheme as well requires the implementation of truncations in order to attain a converging simulation error. Hence, we also have to consider a truncation error. Before analyzing how the different error sources contribute to the approximation error in the enhanced approach, we will illustrate by several examples the construction of function bases, that form a system of martingales according to Assumption 6.

Example 15. This example is based on the assumption that the terminal condition fulfills $\xi = \phi(S_T)$ and the forward SDE in (1.3) is solved by a (possibly multi-variate) geometric Brownian motion. We model S by D identically and independently distributed Markov processes $(S_{d,t})_{t\in[0,T]}$ with

$$S_{d,t} = s_{d,0}\exp\{(\mu - \frac{1}{2}\sigma^2)t + \sigma W_{d,t}\}, \quad d = 1,\dots,D,$$

where $s_{d,0}$, $\sigma > 0$ and $\mu \in \mathbb{R}$. In this setting the approximation of S by S^π becomes obsolete as S can be sampled perfectly. We will explain the creation of martingale basis functions for three different cases. As the terminal condition is not path-dependent in the present case, we simply set $X = S$.

Precisely, we suppose that $\eta_0(N,x)$ is (i) a set of indicator functions of hypercubes of the state space of X, (ii) a set of monomials depending on X or (iii) includes the pay-off function of a European max-call option.

(i) Indicator functions of hypercubes: Let $\eta_0(N,x)$ be a set of functions

$$\eta_{[\mathbf{a},\mathbf{b}]} := \mathbb{1}_{[\mathbf{a},\mathbf{b}]} = \mathbb{1}_{[a_1,b_1]\times\cdots\times[a_D,b_D]}.$$

Due to the independence of $(X_{d,t})_{t\in[0,T]}$ for all $d = 1,\dots,D$, we receive

$$E[\eta_{[\mathbf{a},\mathbf{b}]}(X_T)|X_{t_i} = x] = \prod_{d=1}^{D} E[\mathbb{1}_{[a_d,b_d]}(X_{d,T})|X_{d,t_i} = x_d]$$
$$= \prod_{d=1}^{D} \mathcal{N}(\tilde{b}_d) - \mathcal{N}(\tilde{a}_d)\,.$$

Here $\mathcal{N}$ is the cumulative distribution function of a standard normal applied on

$$\tilde{a}_d = \frac{\log(a_d/x_d) - (\mu - 0.5\sigma^2)(T - t_i)}{\sigma\sqrt{T - t_i}}$$

and an analogously defined $\tilde{b}_d$.

(ii) Monomials: For monomials $\eta_{\mathbf{p}}(x) := x_1^{p_1}\cdots x_D^{p_D}$ one has

$$E[\eta_{\mathbf{p}}(X_T)|X_{t_i} = x] = \prod_{d=1}^{D} x_d^{p_d}\exp\{(p_d\mu + 0.5p_d(p_d - 1)\sigma^2)(T - t_i)\}\,.$$

(iii) For the payoff function of a max-call option $\eta_\kappa(x) = (\max_{d=1,\dots,D} x_d - \kappa)_+$, it can be derived from the results by Johnson (1987) that

$$E[\eta_\kappa(X_T)|X_{t_i} = x] = \sum_{d=1}^{D} e^{\mu(T-t_i)}x_d\mathcal{N}_{0,\Sigma}(a_{d,+})$$
$$- \kappa\Big(1 - \prod_{d=1}^{D}\mathcal{N}\Big(\frac{\log(\kappa/x_d) - (\mu - 0.5\sigma^2)(T - t_i)}{\sigma\sqrt{T - t_i}}\Big)\Big)\,,$$

where $\mathcal{N}_{0,\Sigma}$ is the distribution function of a D-variate normal with mean vector 0 and covariance matrix Σ. Precisely,

$$a_{d,+} = \frac{1}{\sigma\sqrt{T - t_i}}\begin{pmatrix} \log(x_d/\kappa) + (\mu + 0.5\sigma^2)(T - t_i) \\ \frac{1}{\sqrt{2}}(\log(x_d/x_{\bar{d}}) + \sigma^2(T - t_i)) \\ \vdots \\ \frac{1}{\sqrt{2}}(\log(x_d/x_D) + \sigma^2(T - t_i)) \end{pmatrix},$$

with $\bar{d} = 1,\dots,D$, $\bar{d} \neq d$,and

$$\Sigma = \begin{pmatrix} 1 & 1/\sqrt{2} & 1/\sqrt{2} & \cdots & 1/\sqrt{2} \\ 1/\sqrt{2} & 1 & 1/2 & \cdots & 1/2 \\ 1/\sqrt{2} & 1/2 & 1 & & 1/2 \\ \vdots & \vdots & & \ddots & \vdots \\ 1/\sqrt{2} & 1/2 & \cdots & 1/2 & 1 \end{pmatrix}.$$

Now we assume that $\eta_0(i,x)$ is computable in closed form according to Assumption 6 (a) and is continuously differentiable with respect to x_d, $d = 1,\dots,D$. When it comes to calculating conditional expectations of the form $E[\Delta W_{d,i}\eta_{0,k}(N, X^{\pi}_{t_{i+1}})| X^{\pi}_{t_i} = x]$ in the present setting for $X = S$ we can apply for $i < N$ the following rule:

$$\eta_d(i,x) = \sigma x_d \frac{\partial}{\partial x_d}\eta_0(i,x). \tag{4.5}$$

Indeed, for the one-dimensional case $(D = 1)$ one easily computes

$$\begin{aligned}
\sigma x \frac{d}{dx}\eta_0(i,x) &= \sigma x \frac{d}{dx} E\left[\eta_0(i+1, X_{t_{i+1}})\,|X_{t_i} = x\right] \\
&= \sigma x \frac{1}{\sqrt{2\pi\Delta_i}}\int_{-\infty}^{\infty} e^{-\frac{u^2}{2\Delta_i}} \frac{d}{dx}\eta_0(i+1, xe^{\sigma u + (\mu - 0.5\sigma^2)\Delta_i})du \\
&= \frac{1}{\sqrt{2\pi\Delta_i}}\int_{-\infty}^{\infty} e^{-\frac{u^2}{2\Delta_i}} \frac{d}{du}\eta_0(i+1, xe^{\sigma u + (\mu - 0.5\sigma^2)\Delta_i})du \\
&= \frac{1}{\sqrt{2\pi\Delta_i}}\int_{-\infty}^{\infty} \eta_0(i+1, xe^{\sigma u + (\mu - 0.5\sigma^2)\Delta_i})\frac{d}{du}\left(-e^{-\frac{u^2}{2\Delta_i}}\right) du \\
&= \frac{1}{\sqrt{2\pi\Delta_i}}\int_{-\infty}^{\infty} \eta_0(i+1, xe^{\sigma u + (\mu - 0.5\sigma^2)\Delta_i})\frac{u}{\Delta_i}e^{-\frac{u^2}{2\Delta_i}} du \\
&= \frac{1}{\Delta_i}E[\Delta W_i \eta_0(i+1, X_{t_{i+1}})| X_{t_i} = x] \\
&= \frac{1}{\Delta_i}E[\Delta W_i \eta_0(N, X^{\pi}_{t_N})| X_{t_i} = x].
\end{aligned}$$

Analogously we receive the multi-dimensional case. Using formula (4.5) we can then calculate the conditional expectations of type $E[\Delta W_{d,i}\eta_{0,k}(N, X^{\pi}_{t_{i+1}})| X^{\pi}_{t_i} = x]$ for the above examples of $\eta_0(N,x)$, e.g. indicator functions, monomials, and pay-off function of a European call.

Remark 16. It might be objected, that Assumption 6 oversimplifies the problem of estimating conditional expectations that appear in the time discretization scheme (2.3). Indeed, the crucial point consists of finding appropriate basis functions, that fulfill the martingale property. A way out might be to find basis functions that match the conditions of the martingale setting at least approximately. When it comes to pricing and hedging European options, there are often approximative solutions for the price and its delta available, which can be used in this sense.

Generally, one can exploit the approximative terminal condition and estimate

$$\begin{aligned}
&\eta_0(i,x) := E[\phi^{\pi}(X^{\pi}_{t_N})| X^{\pi}_{t_i} = x], \\
&\eta_d(i,x) := E[\Delta W_{d,i}\phi^{\pi}(X^{\pi}_{t_N})| X^{\pi}_{t_i} = x], \quad d = 1,\dots,D
\end{aligned}$$

by Monte Carlo simulation. To this end, we use samples of $X^{\pi,t_i,x}_{t_N}$, where the upper index denotes that the Markov process starts in x at time t_i. Both approaches to finding basis functions should be complemented by further functions for the least-squares Monte Carlo estimation of $E[F(t_i, S^{\pi}_{t_i}, Y^{\pi}_{t_{i+1}}, Z^{\pi}_{t_i})| X^{\pi}_{t_i}]$. For this purpose, see the above proposals. A related numerical example can be found at the end of this chapter.

Similarly to Section 2.2, we will proceed with the analysis of the approximation error step by step. Again, we will start with the projection error.

4.2 Error sources of the simplified scheme and their contribution to the approximation error

4.2.1 Projection error

We first examine the projection error of the simplified least-squares Monte Carlo scheme. To this end, we assume that (4.1) and (4.2) are computable in closed form and (4.3) is replaced by

$$\mathcal{P}_{0,i}\big(F(t_i, S^\pi_{t_i}, Y^\pi_{t_{i+1}}, Z^\pi_{t_i})\big) = \eta_0(i, X^\pi_{t_i})\tilde{\alpha}^{\pi,K}_i,$$

with

$$\tilde{\alpha}^{\pi,K}_i = \arg\min_{\alpha\in\mathbb{R}^K} E|F(t_i, S^\pi_{t_i}, Y^\pi_{t_{i+1}}, Z^\pi_{t_i}) - \eta_0(i, X^\pi_{t_i})\alpha|^2.$$

Thus, the adjusted scheme reads then for all $i = N-1, \dots, 0$ as follows:

$$\begin{aligned} \hat{Y}^{\pi,K}_{t_N} &= \phi^\pi(X^\pi_{t_N}), \\ \hat{Z}^{\pi,K}_{t_i} &= \frac{1}{\Delta_i} E[(\Delta W_i)^* \hat{Y}^{\pi,K}_{t_{i+1}} | X^\pi_{t_i}], \\ \hat{Y}^{\pi,K}_{t_i} &= E[\hat{Y}^{\pi,K}_{t_{i+1}} | X^\pi_{t_i}] - \Delta_i \mathcal{P}_{0,i}\big(F(t_i, S^\pi_{t_i}, \hat{Y}^{\pi,K}_{t_{i+1}}, \hat{Z}^{\pi,K}_{t_i})\big). \end{aligned} \tag{4.6}$$

Lemma 17. *Let Assumption 2 be satisfied. Then there is a constant* C *depending on* κ, T *and* D *such that*

$$\begin{aligned} &\max_{0\leqslant i\leqslant N} E|Y^\pi_{t_i} - \hat{Y}^{\pi,K}_{t_i}|^2 + \sum_{i=0}^{N-1} \Delta_i E|Z^\pi_{t_i} - \hat{Z}^{\pi,K}_{t_i}|^2 \\ &\quad \leqslant C \sum_{i=0}^{N-1} \Delta_i E|\mathcal{P}_{0,i}\big(F(t_i, S^\pi_{t_{t_i}}, Y^\pi_{t_{i+1}}, Z^\pi_{t_i})\big) - E[F(t_i, S^\pi_{t_{t_i}}, Y^\pi_{t_{i+1}}, Z^\pi_{t_i}) | X^\pi_{t_i}]|^2. \end{aligned}$$

As the proof of Lemma 17 involves procedures that will be repeated for the analysis of the truncation error, we first show general estimates on the L^2-distance of two processes $\tilde{Y}^k_{t_i}$, $k = 1, 2$ and $\tilde{Z}^k_{t_i}$, $k = 1, 2$, respectively. For an $(\mathcal{F}_{t_i})_{t_i\in\pi}$-adapted triple $(s^k_{t_i}, y^k_{t_i}, z^k_{t_i})_{t_i\in\pi}$, these processes are defined for $i = N-1, \dots, 0$ by

$$\begin{aligned} \tilde{Y}^k_{t_N} &= y^k_{t_N}, \\ \tilde{Z}^k_{t_i} &= \frac{1}{\Delta_i} E[(\Delta W_i)^* y^k_{t_{i+1}} | \mathcal{F}_{t_i}], \\ \tilde{Y}^k_{t_i} &= E[y^k_{t_{j+1}} | \mathcal{F}_{t_i}] - \Delta_i \Psi^{(k)}\big(i, F(t_i, s^k_{t_i}, y^k_{t_{i+1}}, z^k_{t_i})\big), \end{aligned} \tag{4.7}$$

where $\Psi^{(k)}(i, \cdot)$, $k = 1, 2$ are operators that map $\tilde{U}$ on a $\mathcal{F}_{t_i}$-measurable random variable $\Psi^{(k)}(i, \tilde{U})$, $k = 1, 2$, respectively. Precisely, $\Psi^{(k)}(i, \cdot)$ can e.g. be the conditional expectation or some other orthogonal projection on a subspace of $L^2(\mathcal{F}_{t_i})$.

Lemma 18. *Let* $\Psi^{(1)}(i,\cdot) = \mathcal{P}_{0,i}(\cdot)$ *and*

$$\Psi^{(2)}(i,\cdot) = \mathcal{P}_{0,i}(\cdot) \quad \textit{or} \quad \Psi^{(2)}(i,\cdot) = E[\cdot|\mathcal{F}_{t_i}].$$

Supposing that γ_i, $i = 0,\dots,N-1$ *is a series of positive real numbers and* F *is Lipschitz in* (s,y,z) *with constant* κ*, we receive for* $q_i = (1 + \mathbb{1}_{\{s^1_{t_i} \neq s^2_{t_i}\}})\kappa^2(1+D)$*,* $i = 0,\dots,N-1$ *that*

$$\begin{aligned} E|\tilde{Y}^1_{t_i} - \tilde{Y}^2_{t_i}|^2 &\leqslant (1+q_i\Delta_i)E|E[y^1_{t_{i+1}} - y^2_{t_{i+1}}|\mathcal{F}_{t_i}]|^2 + \frac{1+q_i\Delta_i}{1+D}E|s^1_{t_i} - s^2_{t_i}|^2 \\ &+ (1+q_i\Delta_i)\Delta_i E\left[|y^1_{t_{i+1}} - y^2_{t_{i+1}}|^2 + \frac{1}{D}|z^1_{t_i} - z^2_{t_i}|^2\right] \\ &+ \frac{1+q_i\Delta_i}{q_i}\Delta_i E|\mathcal{P}_{0,i}\big(F(t_i, s^2_{t_i}, y^2_{t_{i+1}}, z^2_{t_i})\big) - \Psi^{(2)}\big(i, F(t_i, s^2_{t_i}, y^2_{t_{i+1}}, z^2_{t_i})\big)|^2, \end{aligned} \tag{4.8}$$

$$\Delta_i E|\tilde{Z}^1_{d,t_i} - \tilde{Z}^2_{d,t_i}|^2 \leqslant E\left[|y^1_{t_{i+1}} - y^2_{t_{i+1}}|^2 - E[y^1_{t_i} - y^2_{t_i}|\mathcal{F}_{t_i}]^2\right], \tag{4.9}$$

and

$$\begin{aligned} \Delta_i E|\tilde{Z}^1_{d,t_i} - \tilde{Z}^2_{d,t_i}|^2 &\leqslant \left(1 + \frac{q_i}{\gamma_i}\Delta_i\right) E|y^1_{t_{i+1}} - y^2_{t_{i+1}}|^2 \\ &+ (\gamma_i\Delta_i - 1)E|\tilde{Y}^1_{t_i} - \tilde{Y}^2_{t_i}|^2 + \frac{q_i}{\gamma_i(1+D)}\Delta_i E|s^1_{t_i} - s^2_{t_i}|^2 + \frac{q_i}{D\gamma_i}\Delta_i E|z^1_{t_i} - z^2_{t_i}|^2 \\ &+ \frac{1}{\gamma_i}\Delta_i E|\mathcal{P}_{0,i}\big(F(t_i, s^2_{t_i}, y^2_{t_{i+1}}, z^2_{t_i})\big) - \Psi^{(2)}\big(i, F(t_i, s^2_{t_i}, y^2_{t_{i+1}}, z^2_{t_i})\big)|^2. \end{aligned} \tag{4.10}$$

Proof. From now on we abbreviate as follows:

$$\Delta\Psi_i := \mathcal{P}_{0,i}\big(F(t_i, s^1_{t_i}, y^1_{t_{i+1}}, z^1_{t_i})\big) - \Psi^{(2)}\big(i, F(t_i, s^2_{t_i}, y^2_{t_{i+1}}, z^2_{t_i})\big).$$

In view of (4.7) we can write for $d = 1,\dots,D$

$$\tilde{Z}^k_{d,t_i} = \frac{1}{\Delta_i}E[\Delta W_{d,i} y^k_{t_{i+1}}|\mathcal{F}_{t_i}].$$

Thanks Hölder's inequality we have

$$\sqrt{\Delta_i}|\tilde{Z}^1_{d,t_i} - \tilde{Z}^2_{d,t_i}| \leqslant E\left[|\big(y^1_{t_{i+1}} - y^2_{t_{i+1}}| - E[y^1_{t_i} - y^2_{t_i}|\mathcal{F}_{t_i}]\big)^2|\mathcal{F}_{t_i}\right]^{1/2}$$

and (4.9) follows immediately by computing the quadratic term and by considering the rules concerning conditional expectations. Due to the definition of $\tilde{Y}^k_{t_i}$ we obtain

$$\Delta_i E|\tilde{Z}^1_{d,t_i} - \tilde{Z}^2_{d,t_i}|^2 \leqslant E|y^1_{t_{i+1}} - y^2_{t_{i+1}}|^2 - E|\tilde{Y}^1_{t_i} - \tilde{Y}^2_{t_i}|^2 - 2\Delta_i E[(\tilde{Y}^1_{t_i} - \tilde{Y}^2_{t_i})(\Delta\Psi_i)].$$

Young's inequality yields for some $\gamma_i > 0$

$$\begin{aligned}&\Delta_i E|\tilde{Z}^1_{d,t_i} - \tilde{Z}^2_{d,t_i}|^2 \\ &\quad \leqslant E|y^1_{t_{i+1}} - y^2_{t_{i+1}}|^2 + (\gamma_i\Delta_i - 1)E|\tilde{Y}^1_{t_i} - \tilde{Y}^2_{t_i}|^2 + \frac{1}{\gamma_i}\Delta_i E|\Delta\Psi_i|^2.\end{aligned} \tag{4.11}$$

Taking the possible definitions of $\Psi^{(2)}(i,\cdot)$ into account, we can either make use of the orthogonality of $\mathcal{P}_{0,i}$ or of the identity $\Psi^{(2)}(i,\cdot) = \mathcal{P}_{0,i}(\cdot)$. Thus, it holds true that

$$\begin{aligned}\Delta_i E|\Delta\Psi_i|^2 \leqslant{}& \Delta_i E|\mathcal{P}_{0,i}\big(F(t_i, s^1_{t_i}, y^1_{t_{i+1}}, z^1_{t_i}) - F(t_i, s^2_{t_i}, y^2_{t_{i+1}}, z^2_{t_i})\big)|^2 \\ &+ \Delta_i E|\mathcal{P}_{0,i}\big(F(t_i, s^2_{t_i}, y^2_{t_{i+1}}, z^2_{t_i})\big) - \Psi^{(2)}\big(i, F(t_i, s^2_{t_i}, y^2_{t_{i+1}}, z^2_{t_i})\big)|^2.\end{aligned}$$

The contraction property of the projections and the Lipschitz condition on F lead to

$$\begin{aligned}\Delta_i E|\Delta\Psi_i|^2 &\leqslant \kappa^2\Delta_i E\big[|s^1_{t_i} - s^2_{t_i}| + |y^1_{t_{i+1}} - y^2_{t_{i+1}}| + |z^1_{t_i} - z^2_{t_i}|\big]^2 \\ &\quad + \Delta_i E|\mathcal{P}_{0,i}\big(F(t_i, s^2_{t_i}, y^2_{t_{i+1}}, z^2_{t_i})\big) - \Psi^{(2)}\big(i, F(t_i, s^2_{t_i}, y^2_{t_{i+1}}, z^2_{t_i})\big)|^2 \\ &\leqslant (1 + \mathbb{1}_{\{s^1_{t_i} \neq s^2_{t_i}\}})\kappa^2 E|s^1_{t_i} - s^2_{t_i}|^2 \\ &\quad + (1 + \mathbb{1}_{\{s^1_{t_i} \neq s^2_{t_i}\}})\kappa^2(1 + D)\Delta_i E\big[|y^1_{t_{i+1}} - y^2_{t_{i+1}}|^2 + \frac{1}{D}|z^1_{t_i} - z^2_{t_i}|^2\big] \\ &\quad + \Delta_i E|\mathcal{P}_{0,i}\big(F(t_i, s^2_{t_i}, y^2_{t_{i+1}}, z^2_{t_i})\big) - \Psi^{(2)}\big(i, F(t_i, s^2_{t_i}, y^2_{t_{i+1}}, z^2_{t_i})\big)|^2,\end{aligned} \tag{4.12}$$

where the last step followed by Young' s inequality. After setting $q_i = (1 + \mathbb{1}_{\{s^1_{t_i} \neq s^2_{t_i}\}})\kappa^2(1+D)$, we apply (4.12) on (4.11) and receive immediately (4.10). Turning to the Y-part we obtain by Young's inequality

$$E|\tilde{Y}^1_{t_i} - \tilde{Y}^2_{t_i}|^2 \leqslant (1 + q_i\Delta_i)E|E[y^1_{t_{i+1}} - y^2_{t_{i+1}}|\mathcal{F}_{t_i}]|^2 + \frac{1 + q_i\Delta_i}{q_i}\Delta_i E|\Delta\Psi_i|^2.$$

The estimate in (4.12) completes the proof of (4.8). □

After these preparations we turn to the

Proof of Lemma 17. We want to apply Lemma 18. To this end we set

$$\begin{aligned}(s^1_{t_i}, y^1_{t_i}, z^1_{t_i})_{t_i\in\pi} &= (S^\pi_{t_i}, \hat{Y}^{\pi,K}_{t_i}, \hat{Z}^{\pi,K}_{t_i})_{t_i\in\pi}, \\ (s^2_{t_i}, y^2_{t_i}, z^2_{t_i})_{t_i\in\pi} &= (S^\pi_{t_i}, Y^\pi_{t_i}, Z^\pi_{t_i})_{t_i\in\pi}\end{aligned}$$

and $\Psi^{(2)}(i,\cdot) = E[\cdot|X^\pi_{t_i}]$. Then $q_i = \kappa^2(1 + D)$ for all $i = 0, \ldots, N-1$. That means, we are now in the setting of (4.6) and (2.3). Hence, we receive by (4.8)

$$\begin{aligned}E|Y^\pi_{t_i} - \hat{Y}^{\pi,K}_{t_i}|^2 &\leqslant (1 + q_i\Delta_i)E|E[Y^\pi_{t_{i+1}} - \hat{Y}^{\pi,K}_{t_{i+1}}|X^\pi_{t_i}]|^2 \\ &+ (1 + q_i\Delta_i)\Delta_i E\big[|Y^\pi_{t_{i+1}} - \hat{Y}^{\pi,K}_{t_{i+1}}|^2 + \frac{1}{D}|Z^\pi_{t_i} - \hat{Z}^{\pi,K}_{t_i}|^2\big] \\ &+ C\Delta_i E|\mathcal{P}_{0,i}\big(F(t_i, S^\pi_{t_i}, Y^\pi_{t_{i+1}}, Z^\pi_{t_i})\big) - E[F(t_i, S^\pi_{t_i}, Y^\pi_{t_{i+1}}, Z^\pi_{t_i})|X^\pi_{t_i}]|^2.\end{aligned}$$

By exploiting (4.9) we obtain

$$
\begin{aligned}
E|Y^\pi_{t_i} - \hat{Y}^{\pi,K}_{t_i}|^2 &\leqslant (1+q_i\Delta_i)(1+\Delta_i)E|Y^\pi_{t_{i+1}} - \hat{Y}^{\pi,K}_{t_{i+1}}|^2 \\
&\quad + C\Delta_i E|\mathcal{P}_{0,i}\big(F(t_i, S^\pi_{t_i}, Y^\pi_{t_{i+1}}, Z^\pi_{t_i})\big) - E[F(t_i, S^\pi_{t_i}, Y^\pi_{t_{i+1}}, Z^\pi_{t_i})|\, X^\pi_{t_i}]|^2.
\end{aligned}
$$

Gronwall's inequality leads to

$$
\begin{aligned}
E|Y^\pi_{t_i} - \hat{Y}^{\pi,K}_{t_i}|^2 &\leqslant e^{T(1+q_i(1+|\pi|))}\Big\{E|Y^\pi_{t_N} - \hat{Y}^{\pi,K}_{t_N}|^2 \\
&\quad + C\sum_{j=i}^{N-1}\Delta_j E|\mathcal{P}_{0,j}\big(F(t_j, S^\pi_{t_j}, Y^\pi_{t_{j+1}}, Z^\pi_{t_j})\big) - E[F(t_j, S^\pi_{t_j}, Y^\pi_{t_{j+1}}, Z^\pi_{t_j})|\, X^\pi_{t_j}]|^2\Big\}.
\end{aligned}
$$

Since $Y^\pi_{t_i} = \hat{Y}^{\pi,K}_{t_i}$, the upper bound for the Y-part is proven. Thanks to (4.10) we get

$$
\begin{aligned}
\Delta_i E|Z^\pi_{d,t_i} - \hat{Z}^{\pi,K}_{d,t_i}|^2 &\leqslant (1+\frac{q_i}{\gamma_i}\Delta_i)E|Y^\pi_{t_{i+1}} - \hat{Y}^{\pi,K}_{t_{i+1}}|^2 \\
&\quad + (\gamma_i\Delta_i - 1)E|Y^\pi_{t_i} - \hat{Y}^{\pi,K}_{t_i}|^2 + \frac{q_i}{D\gamma_i}\Delta_i E|Z^\pi_{t_i} - \hat{Z}^{\pi,K}_{t_i}|^2 \\
&\quad + \frac{1}{\gamma_i}\Delta_i E|\mathcal{P}_{0,i}\big(F(t_i, S^\pi_{t_i}, Y^\pi_{t_{i+1}}, Z^\pi_{t_i})\big) - E[F(t_i, S^\pi_{t_i}, Y^\pi_{t_{i+1}}, Z^\pi_{t_i})|\, X^\pi_{t_i}]|^2.
\end{aligned}
$$

Summing up from $i = 0$ to $N-1$ and setting $\gamma_i = 2q_i$ yields

$$
\begin{aligned}
\sum_{i=0}^{N-1}\Delta_i E|Z^\pi_{t_i} - \hat{Z}^{\pi,K}_{t_i}|^2 &\leqslant D(1+4q_i)T\max_{0\leqslant i\leqslant N-1} E|Y^\pi_{t_i} - \hat{Y}^{\pi,K}_{t_i}|^2 \\
&\quad + \frac{D}{q_i}\sum_{i=0}^{N-1}\Delta_i E|\mathcal{P}_{0,i}\big(F(t_i, S^\pi_{t_i}, Y^\pi_{t_{i+1}}, Z^\pi_{t_i})\big) - E[F(t_i, S^\pi_{t_i}, Y^\pi_{t_{i+1}}, Z^\pi_{t_i})|\, X^\pi_{t_i}]|^2.
\end{aligned}
$$

We finish the proof by applying the upper bound on $E|Y^\pi_{t_i} - \hat{Y}^{\pi,K}_{t_i}|^2$. □

4.2.2 Truncation error

For technical reasons we require an approximation of (Y, Z) that is bounded. Precisely, we modify the scheme in (4.6) by applying a truncation function on $\hat{Y}^{\pi,K}_{t_{i+1}}$ and $\hat{Z}^{\pi,K}_{t_i}$ for all $i = 0, \dots, N-1$. For this purpose we define for some $\mathbb{R}$-valued random variable U and $R > 0$

$$
[U]_R := -R \wedge U \vee R, \quad [U]_{R/\sqrt{\Delta_i}} := -\frac{R}{\sqrt{\Delta_i}} \wedge U \vee \frac{R}{\sqrt{\Delta_i}}.
$$

By implementing the truncations in (4.6) we obtain for $i = N-1, \dots, 0$

$$
\begin{aligned}
\hat{Y}^{\pi,K,R}_{t_N} &= \big[\phi^\pi(X^\pi_{t_N})\big]_R, \\
\hat{Z}^{\pi,K,R}_{d,t_i} &= \Big[\Delta_i^{-1}E[\Delta W_{d,i}\hat{Y}^{\pi,K,R}_{t_{i+1}}|\, X^\pi_{t_i}]\Big]_{R/\sqrt{|\Delta_i|}}, \quad d = 1, \dots, D \\
\hat{Y}^{\pi,K,R}_{t_i} &= \Big[E[\hat{Y}^{\pi,K,R}_{t_{i+1}}|\, X^\pi_{t_i}] - \Delta_i\mathcal{P}_{0,i}\big(F(t_i, S^\pi_{t_i}, \hat{Y}^{\pi,K,R}_{t_{i+1}}, \hat{Z}^{\pi,K,R}_{t_i})\big)\Big]_R.
\end{aligned}
$$

However, introducing truncation functions cancels out the advantage of Assumption 6. This is insofar no critical factor as truncations in practice are generally neglected. The next lemma gives information about the truncation error, which determines the difference between $(\hat{Y}^{\pi,K}_{t_i}, \hat{Z}^{\pi,K}_{t_i})_{t_i \in \pi}$ and $(\hat{Y}^{\pi,K,R}_{t_i}, \hat{Z}^{\pi,K,R}_{t_i})_{t_i \in \pi}$.

Lemma 19. *Let Assumption 2 be satisfied. Then there is a constant* C *depending on* κ, T *and* D *such that*

$$\begin{aligned}
&\max_{0 \leqslant i \leqslant N} E|\hat{Y}^{\pi,K}_{t_i} - \hat{Y}^{\pi,K,R}_{t_i}|^2 + \sum_{i=0}^{N-1} \Delta_i E|\hat{Z}^{\pi,K}_{t_i} - \hat{Z}^{\pi,K,R}_{t_i}|^2 \\
&\leqslant CN \frac{K^{2\epsilon}}{R^{2\epsilon-2}} \max_{0 \leqslant i \leqslant N} E|Y^{\pi}_{t_i}|^{2\epsilon} \\
&\quad + C \sum_{i=0}^{N-1} \Delta_i E|\mathcal{P}_{0,i}\big(F(t_i, S^{\pi}_{t_i}, Y^{\pi}_{t_{i+1}}, Z^{\pi}_{t_i})\big) - E[F(t_i, S^{\pi}_{t_i}, Y^{\pi}_{t_{i+1}}, Z^{\pi}_{t_i})|X^{\pi}_{t_i}]|^2.
\end{aligned}$$

Proof. By Young's inequality we receive

$$\begin{aligned}
&\max_{0 \leqslant i \leqslant N} E|\hat{Y}^{\pi,K}_{t_i} - \hat{Y}^{\pi,K,R}_{t_i}|^2 + \sum_{i=0}^{N-1} \Delta_i E|\hat{Z}^{\pi,K}_{t_i} - \hat{Z}^{\pi,K,R}_{t_i}|^2 \\
&\leqslant 2\Big(\max_{0 \leqslant i \leqslant N} E|\hat{Y}^{\pi,K}_{t_i} - Y^{\pi}_{t_i}|^2 + \sum_{i=0}^{N-1} \Delta_i E|\hat{Z}^{\pi,K}_{t_i} - Z^{\pi}_{t_i}|^2\Big) \\
&\quad + 2\Big(\max_{0 \leqslant i \leqslant N} E|Y^{\pi}_{t_i} - \hat{Y}^{\pi,K,R}_{t_i}|^2 + \sum_{i=0}^{N-1} \Delta_i E|Z^{\pi}_{t_i} - \hat{Z}^{\pi,K,R}_{t_i}|^2\Big).
\end{aligned}$$

An upper bound for the first summand is given by Lemma 17 and it remains to analyse the second summand. This will be done in two steps.

Step 1: We start with calculating estimates for

$$E\big[|Y^{\pi}_{t_i} - \hat{Y}^{\pi,K,R}_{t_i}|^2 \mathbb{1}_{\{|Y^{\pi}_{t_i}|>R\}}\big], \quad \Delta_i E\big[|Z^{\pi}_{d,t_i} - \hat{Z}^{\pi,K,R}_{d,t_i}|^2 \mathbb{1}_{\{|Z^{\pi}_{d,t_i}|>R/\sqrt{\Delta_i}\}}\big].$$

The application of Young's inequality and then Hölder's inequality yields

$$\begin{aligned}
E\big[|Y^{\pi}_{t_i} - \hat{Y}^{\pi,K,R}_{t_i}|^2 \mathbb{1}_{\{|Y^{\pi}_{t_i}|>R\}}\big] &\leqslant 2E\big[(|Y^{\pi}_{t_i}|^2 + R^2)\mathbb{1}_{\{|Y^{\pi}_{t_i}|>R\}}\big] \\
&\leqslant 2E[|Y^{\pi}_{t_i}|^{2\epsilon}]^{1/\epsilon}(\mathbb{P}\{|Y^{\pi}_{t_i}| > R\})^{1/\zeta} + 2R^2\mathbb{P}\{|Y^{\pi}_{t_i}| > R\},
\end{aligned}$$

where $\zeta > 1$ is determined by $\epsilon^{-1} + \zeta^{-1} = 1$. Due to Markov's inequality we have

$$E\big[|Y^{\pi}_{t_i} - \hat{Y}^{\pi,K,R}_{t_i}|^2 \mathbb{1}_{\{|Y^{\pi}_{t_i}|>R\}}\big] \leqslant 2E|Y^{\pi}_{t_i}|^{2\epsilon}(R^{-\frac{2\epsilon}{\zeta}} + R^{2-2\epsilon}) \leqslant 4R^{2-2\epsilon} \max_{0 \leqslant i \leqslant N} E|Y^{\pi}_{t_i}|^{2\epsilon}.$$

Analogously, we obtain

$$\Delta_i E\big[|Z^{\pi}_{d,t_i} - \hat{Z}^{\pi,K,R}_{d,t_i}|^2 \mathbb{1}_{\{|Z^{\pi}_{d,t_i}|>R/\sqrt{\Delta_i}\}}\big] \leqslant 4E|\sqrt{\Delta_i} Z^{\pi}_{d,t_i}|^{2\epsilon} R^{2-2\epsilon}.$$

By the definition of Z^{π}_{d,t_i} and Hölder's inequality, we achieve

$$E|\sqrt{\Delta_i}Z^{\pi}_{d,t_i}|^{2\epsilon} \leqslant E\Big|E\Big[\frac{\Delta W_{d,i}}{\sqrt{\Delta_i}}Y^{\pi}_{t_{i+1}}\Big| X^{\pi}_{t_i}\Big]\Big|^{2\epsilon} \leqslant E|E[|Y^{\pi}_{t_{i+1}}|^2| X^{\pi}_{t_i}]|^{\epsilon} \leqslant E|Y^{\pi}_{t_{i+1}}|^{2\epsilon}.$$

Thus, we receive for $\Delta_i E[|Z^{\pi}_{d,t_i} - \hat{Z}^{\pi,K,R}_{d,t_i}|^2 \mathbb{1}_{\{|Z^{\pi}_{d,t_i}|>R/\sqrt{\Delta_i}\}}]$ the same upper bound as for $E[|Y^{\pi}_{t_i} - \hat{Y}^{\pi,K,R}_{t_i}|^2 \mathbb{1}_{\{|Y^{\pi}_{t_i}|>R\}}]$.

Step 2: For the application of Lemma 18 we define

$$(s^1_{t_i}, y^1_{t_i}, z^1_{t_i})_{t_i\in\pi} = (S^{\pi}_{t_i}, \hat{Y}^{\pi,K,R}_{t_i}, \hat{Z}^{\pi,K,R}_{t_i})_{t_i\in\pi},$$
$$(s^1_{t_i}, y^2_{t_i}, z^2_{t_i})_{t_i\in\pi} = (S^{\pi}_{t_i}, Y^{\pi}_{t_i}, Z^{\pi}_{t_i})_{t_i\in\pi}$$

and set $\Psi^{(2)}(i,\cdot) = E[\cdot| X^{\pi}_{t_i}]$. Then we have $q_i = \kappa^2(1+D)$ for all $i = 0,\dots,N-1$. Note, that in view of this definition the Lipschitz continuity of $[\cdot]_R$ yields

$$\begin{aligned} E|\hat{Y}^{\pi,K,R}_{t_i} - Y^{\pi}_{t_i}|^2 &= E\big[|\hat{Y}^{\pi,K,R}_{t_i} - Y^{\pi}_{t_i}|^2 \mathbb{1}_{\{|Y^{\pi}_{t_i}|\leqslant R\}}\big] + E\big[|\hat{Y}^{\pi,K,R}_{t_i} - Y^{\pi}_{t_i}|^2 \mathbb{1}_{\{|Y^{\pi}_{t_i}|>R\}}\big] \\ &\leqslant E|\tilde{Y}^1_{t_i} - \tilde{Y}^2_{t_i}|^2 + E\big[|\hat{Y}^{\pi,K,R}_{t_i} - Y^{\pi}_{t_i}|^2 \mathbb{1}_{\{|Y^{\pi}_{t_i}|>R\}}\big] \end{aligned} \tag{4.13}$$

and analogously

$$E|\hat{Z}^{\pi,K,R}_{d,t_i} - Z^{\pi}_{d,t_i}|^2 \leqslant E|\tilde{Z}^1_{t_i} - \tilde{Z}^2_{t_i}|^2 + E\big[|\hat{Z}^{\pi,K,R}_{d,t_i} - Z^{\pi}_{d,t_i}|^2 \mathbb{1}_{\{|Z^{\pi}_{d,t_i}|>R/\sqrt{\Delta_i}\}}\big]. \tag{4.14}$$

We obtain by (4.8),

$$\begin{aligned} E|\tilde{Y}^1_{t_i} - \tilde{Y}^2_{t_i}|^2 &\leqslant (1+q_i\Delta_i)E|E[\hat{Y}^{\pi,K,R}_{t_{i+1}} - Y^{\pi}_{t_{i+1}}| X^{\pi}_{t_i}]|^2 \\ &+ (1+q_i\Delta_i)\Delta_i E\big[|\hat{Y}^{\pi,K,R}_{t_{i+1}} - Y^{\pi}_{t_{i+1}}|^2 + \frac{1}{D}|\hat{Z}^{\pi,K,R}_{t_i} - Z^{\pi}_{t_i}|^2\big] \\ &+ \frac{1+q_i\Delta_i}{q_i}\Delta_i E|\mathcal{P}_{0,i}\big(F(t_i, S^{\pi}_{t_i}, Y^{\pi}_{t_{i+1}}, Z^{\pi}_{t_i})\big) - E[F(t_i, S^{\pi}_{t_i}, Y^{\pi}_{t_{i+1}}, Z^{\pi}_{t_i})| X^{\pi}_{t_i}]|^2. \end{aligned}$$

Due to (4.14) and (4.9) it holds true that

$$\begin{aligned} E|\tilde{Y}^1_{t_i} - \tilde{Y}^2_{t_i}|^2 &\leqslant (1+q_i\Delta_i)(1+\Delta_i)E|\hat{Y}^{\pi,K,R}_{t_{i+1}} - Y^{\pi}_{t_{i+1}}|^2 \\ &+ (1+q_i\Delta_i)\Delta_i E\big[|\hat{Z}^{\pi,K,R}_{d,t_i} - Z^{\pi}_{d,t_i}|^2 \mathbb{1}_{\{|Z^{\pi}_{d,t_i}|>R/\sqrt{\Delta_i}\}}\big] \\ &+ \frac{1+q_i\Delta_i}{q_i}\Delta_i E|\mathcal{P}_{0,i}\big(F(t_i, S^{\pi}_{t_i}, Y^{\pi}_{t_{i+1}}, Z^{\pi}_{t_i})\big) - E[F(t_i, S^{\pi}_{t_i}, Y^{\pi}_{t_{i+1}}, Z^{\pi}_{t_i})| X^{\pi}_{t_i}]|^2. \end{aligned}$$

Considering (4.13) and the upper bounds derived in Step 1, we can employ Gronwall's inequality. Hence,

$$
\begin{aligned}
E|\tilde{Y}^1_{t_i} - \tilde{Y}^2_{t_i}|^2 &\leqslant e^{T(1+q_i(1+|\pi|))}\Big\{E|\tilde{Y}^1_{t_N} - \tilde{Y}^2_{t_N}|^2 + C\frac{NK^{2\epsilon}}{R^{2\epsilon-2}}\max_{0\leqslant i\leqslant N} E|Y^\pi_{t_i}|^{2\epsilon} \\
&\quad + C\sum_{j=i}^{N-1}\Delta_j E|\mathcal{P}_{0,j}\big(F(t_j, S^\pi_{t_j}, Y^\pi_{t_{j+1}}, Z^\pi_{t_j})\big) - E[F(t_j, S^\pi_{t_j}, Y^\pi_{t_{j+1}}, Z^\pi_{t_j})|X^\pi_{t_j}]|^2\Big\} \\
&\leqslant CN\frac{K^{2\epsilon}}{R^{2\epsilon-2}}\max_{0\leqslant i\leqslant N} E|Y^\pi_{t_i}|^{2\epsilon} \\
&\quad + C\sum_{j=i}^{N-1}\Delta_j E|\mathcal{P}_{0,j}\big(F(t_j, S^\pi_{t_j}, Y^\pi_{t_{j+1}}, Z^\pi_{t_j})\big) - E[F(t_j, S^\pi_{t_j}, Y^\pi_{t_{j+1}}, Z^\pi_{t_j})|X^\pi_{t_j}]|^2,
\end{aligned} \tag{4.15}
$$

as $\tilde{Y}^1_{t_N} - \tilde{Y}^2_{t_N} = 0$. Inserting this result in (4.13) and using again the upper bounds of Step 1, has the consequence

$$
\begin{aligned}
E|\hat{Y}^{\pi,K,R}_{t_i} - Y^\pi_{t_i}|^2 &\leqslant CNR^{2-2\epsilon}K^{2\epsilon}\max_{0\leqslant i\leqslant N} E|Y^\pi_{t_i}|^{2\epsilon} \\
&\quad + C\sum_{j=i}^{N-1}\Delta_j E|\mathcal{P}_{0,j}\big(F(t_j, S^\pi_{t_j}, Y^\pi_{t_{j+1}}, Z^\pi_{t_j})\big) - E[F(t_j, S^\pi_{t_j}, Y^\pi_{t_{j+1}}, Z^\pi_{t_j})|X^\pi_{t_j}]|^2.
\end{aligned}
$$

Exploiting (4.10) and (4.13) gives

$$
\begin{aligned}
\Delta_i E|\tilde{Z}^1_{d,t_i} - \tilde{Z}^2_{d,t_i}|^2 &\leqslant (1 + \frac{q_i}{\gamma_i}\Delta_i)\Big[E|\tilde{Y}^1_{t_{i+1}} - \tilde{Y}^2_{t_{i+1}}|^2 + C\frac{K^{2\epsilon}}{R^{2\epsilon-2}}\max_{0\leqslant i\leqslant N} E|Y^\pi_{t_i}|^{2\epsilon}\Big] \\
&\quad + (\gamma_i\Delta_i - 1)E|\tilde{Y}^1_{t_i} - \tilde{Y}^2_{t_i}|^2 + \frac{q_i}{D\gamma_i}\Delta_i E|\hat{Z}^{\pi,K,R}_{t_i} - Z^\pi_{t_i}|^2 \\
&\quad + \frac{1}{\gamma_i}\Delta_i E|\mathcal{P}_{0,i}\big(F(t_i, S^\pi_{t_i}, Y^\pi_{t_{i+1}}, Z^\pi_{t_i})\big) - E[F(t_i, S^\pi_{t_i}, Y^\pi_{t_{i+1}}, Z^\pi_{t_i})|X^\pi_{t_i}]|^2.
\end{aligned}
$$

Taking (4.14) into account and summing up from $i = 0$ to $N-1$, it turns out that

$$
\begin{aligned}
\sum_{i=0}^{N-1}\Delta_i E|\hat{Z}^{\pi,K,R}_{t_i} - Z^\pi_{t_i}|^2 &\leqslant D(1 + \frac{q_{N-1}}{\gamma_{N-1}}\Delta_{N-1})E|\tilde{Y}^1_{t_N} - \tilde{Y}^2_{t_N}|^2 \\
&\quad + \sum_{i=0}^{N-1} D(\frac{q_{i-1}}{\gamma_{i-1}} + \gamma_i)\Delta_i E|\tilde{Y}^1_{t_i} - \tilde{Y}^2_{t_i}|^2 + \sum_{i=0}^{N-1}\frac{q_i}{\gamma_i}\Delta_i E|\hat{Z}^{\pi,K,R}_{t_i} - Z^\pi_{t_i}|^2 \\
&\quad + C\sum_{i=0}^{N-1} D(2 + \frac{q_i}{\gamma_i}\Delta_i)\frac{K^{2\epsilon}}{R^{2\epsilon-2}}\max_{0\leqslant i\leqslant N} E|Y^\pi_{t_i}|^{2\epsilon} \\
&\quad + D\sum_{i=0}^{N-1}\frac{\Delta_i}{\gamma_i}E|\mathcal{P}_{0,i}\big(F(t_i, S^\pi_{t_i}, Y^\pi_{t_{i+1}}, Z^\pi_{t_i})\big) - E[F(t_i, S^\pi_{t_i}, Y^\pi_{t_{i+1}}, Z^\pi_{t_i})|X^\pi_{t_i}]|^2,
\end{aligned}
$$

where $q_{-1}/\gamma_{-1} := 0$. By definition, $\tilde{Y}^1_{t_N} - \tilde{Y}^2_{t_N} = 0$. Choosing $\gamma_i = 2q_i$ yields

$$\begin{aligned}\sum_{i=0}^{N-1} \Delta_i E|\hat{Z}^{\pi,K,R}_{t_i} - Z^\pi_{t_i}|^2 &\leqslant TD(1+4q_i) \max_{0\leqslant i\leqslant N-1} E|\tilde{Y}^1_{t_i} - \tilde{Y}^2_{t_i}|^2 \\ &+ CND(2+\frac{1}{2}\Delta_i)R^{2-2\epsilon}K^{2\epsilon} \max_{0\leqslant i\leqslant N} E|Y^\pi_{t_i}|^{2\epsilon} \\ &+ C\sum_{i=0}^{N-1} \Delta_i E|\mathcal{P}_{0,i}\big(F(t_i, S^\pi_{t_i}, Y^\pi_{t_{i+1}}, Z^\pi_{t_i})\big) - E[F(t_i, S^\pi_{t_i}, Y^\pi_{t_{i+1}}, Z^\pi_{t_i})|X^\pi_{t_i}]|^2.\end{aligned}$$

Due to (4.15), we have

$$\begin{aligned}\sum_{i=0}^{N-1} \Delta_i E|\hat{Z}^{\pi,K,R}_{t_i} - Z^\pi_{t_i}|^2 &\leqslant CN\frac{K^{2\epsilon}}{R^{2\epsilon-2}} \max_{0\leqslant i\leqslant N} E|Y^\pi_{t_i}|^{2\epsilon} \\ &+ C\sum_{i=0}^{N-1} \Delta_i E|\mathcal{P}_{0,i}\big(F(t_i, S^\pi_{t_i}, Y^\pi_{t_{i+1}}, Z^\pi_{t_i})\big) - E[F(t_i, S^\pi_{t_i}, Y^\pi_{t_{i+1}}, Z^\pi_{t_i})|X^\pi_{t_i}]|^2. \qquad \square\end{aligned}$$

Recall, that the above approximators can be expressed by deterministic functions of $(X^\pi_{t_i})_{t_i\in\pi}$. Thus, there are functions $\hat{y}^{\pi,K,R}_i(x)$ and $\hat{z}^{\pi,K,R}_i(x)$ such that

$$\hat{Y}^{\pi,K,R}_{t_i} = \hat{y}^{\pi,K,R}_i(X^\pi_{t_i}), \quad \hat{Z}^{\pi,K,R}_{t_i} = \hat{z}^{\pi,K,R}_i(X^\pi_{t_i}).$$

For the analysis of the simulation error of the martingale based least-squares Monte Carlo approach we require $\hat{y}^{\pi,K,R}_i(x)$ to be Lipschitz continuous in x. Therefore, we have to endow the approximative terminal condition ϕ^π, the approximation of S^π and the Markov process $(X^\pi_{t_i})_{t_i\in\pi}$ with additional properties, that imply the desired Lipschitz continuity.

Assumption 7. *(i) The approximative terminal condition* $\phi^\pi(x)$ *is Lipschitz continuous (uniformly in* π*) and* $\sup_N |\phi^\pi(\mathbf{0})| < \infty$.

(ii) We denote by $S^{\pi,i_0,s}_{t_i}$, $i_0 \leqslant i \leqslant N$ *the approximation of* $(S_t)_{t\in[t_{i_0},T]}$ *that starts with* $S^\pi_{t_{i_0}} = s$. *Moreover, we call* $X^{\pi,i_0,x}_{t_i}$, $i_0 \leqslant i \leqslant N$ *the related multivariate Markov process that we require for the Markovian formulation of the time discretization, see Subsection 2.2.1. That means,* $X^{\pi,i_0,x}_{t_{i_0}} = x$, *where* x *is determined by* s *only and its first component is equal to* s. *There is a* $C_X > 0$ *such that for all* $i = i_0, \ldots, N-1$

$$E|X^{\pi,i_0,x}_{t_N} - X^{\pi,i_0,x'}_{t_N}|^2 + E|S^{\pi,i_0,s}_{t_i} - S^{\pi,i_0,s'}_{t_i}|^2 \leqslant C_X|x-x'|^2,$$

uniformly in i_0 *and* π.

(iii) There is a $C > 0$ *such that for any* x

$$E|X^{\pi,i_0,x}_{t_{i_0+1}} - x|^2 \leqslant C\Delta_{i_0}(1+|x|^2).$$

uniformly in i_0 *and* π.

Remark 20. The above assumption on S^π is naturally fulfilled in case S satisfies Assumption 2 and is approximated via Euler scheme.

Lemma 21. *Let Assumptions 2 and 7 be fulfilled. Then there is a Lipschitz constant $\kappa_R > 0$ depending on κ, T, D and C_X such that*

$$|\hat{y}^{\pi,K,R}_{i_0}(x) - \hat{y}^{\pi,K,R}_{i_0}(x')| < \kappa_R |x - x'|$$

for $i_0 \in \{0, \ldots, N-1\}$ and x, x' real-valued samples of $X^\pi_{t_i}$.

Proof. Let s and s' be the first component of the vectors x and x', respectively. First, we define analogously to (1.3) forward SDEs that start at time t_{i_0}. Precisely, we set

$$S^{i_0,x}_t = s + \int_{t_{i_0}}^t b(u, S^{i_0,x}_u)du + \int_{t_{i_0}}^t \sigma(u, S^{i_0,x}_u)dW_u$$

for $t \in [t_{i_0}, T]$. The forward SDE $S^{i_0,x'}_t$ is constructed analogously. We call $S^{\pi,i_0,x}_{t_i}$ and $S^{\pi,i_0,x'}_{t_i}$ for $i = i_0, \ldots, N-1$ the time-discrete approximations of $(S^{i_0,x}_t)_{t\in[t_{i_0},T]}$ and $(S^{i_0,x'}_t)_{t\in[t_{i_0},T]}$. The related multivariate Markov processes, that we need for the Markovian formulation of the time-discrete BSDE (see Subsection 2.2.1), are denoted by $X^{\pi,i_0,x}_{t_i}$ and $X^{\pi,i_0,x'}_{t_i}$ for $i = i_0, \ldots, N$ and shall fulfill Assumption 7. Then, $\hat{y}^{\pi,K,R}_{i_0}(x)$ is the solution of the following scheme. For $i = N-1, \ldots, i_0$ we conduct

$$\begin{aligned}
\hat{Y}^{\pi,K,R,i_0,x}_{t_N} &= \Big[\phi^\pi(X^{\pi,i_0,x}_{t_N})\Big]_R, \\
\hat{Z}^{\pi,K,R,i_0,x}_{d,t_i} &= \Big[\Delta_i^{-1} E[\Delta W_{d,i} \hat{Y}^{\pi,K,R,i_0,x}_{t_{i+1}} | \mathcal{F}_{t_i}]\Big]_{R/\sqrt{\Delta_i}}, \quad d = 1, \ldots, D \\
\hat{Y}^{\pi,K,R,i_0,x}_{t_i} &= \Big[E[\hat{Y}^{\pi,K,R,i_0,x}_{t_{i+1}} | \mathcal{F}_{t_i}] - \Delta_i \mathcal{P}_{0,i}\big(F(t_i, S^{\pi,i_0,x}_{t_i}, \hat{Y}^{\pi,K,R,i_0,x}_{t_{i+1}}, \hat{Z}^{\pi,K,R,i_0,x}_{t_i})\big)\Big]_R.
\end{aligned}$$

Hence, $\hat{y}^{\pi,K,R}_{i_0}(x) = \hat{Y}^{\pi,K,R,i_0,x}_{t_{i_0}}$. Analogously, we can evaluate $\hat{y}^{\pi,K,R}_{i_0}(x')$. Again we exploit Lemma 18. Therefore, we set

$$\begin{aligned}
(s^1_{t_i}, y^1_{t_i}, z^1_{t_i})_{i=i_0,\ldots,N} &= (S^{\pi,i_0,x}_{t_i}, \hat{Y}^{\pi,K,R,i_0,x}_{t_i}, \hat{Z}^{\pi,K,R,i_0,x}_{t_i})_{i=i_0,\ldots,N}, \\
(s^2_{t_i}, y^2_{t_i}, z^2_{t_i})_{i=i_0,\ldots,N} &= (S^{\pi,i_0,x'}_{t_i}, \hat{Y}^{\pi,K,R,i_0,x'}_{t_i}, \hat{Z}^{\pi,K,R,i_0,x'}_{t_i})_{i=i_0,\ldots,N},
\end{aligned}$$

and $\Psi^{(2)}(i, \cdot) = \mathcal{P}_{0,i}$. Here we have $q_i = (1+\mathbb{1}_{\{s^1_{t_i} \neq s^2_{t_i}\}})\kappa^2(1+D)$ for all $i = i_0, \ldots, N-1$. Note that $[\cdot]_R$ is 1-Lipschitz. Thus, due to (4.8) follows

$$\begin{aligned}
E|\hat{Y}^{\pi,K,R,i_0,x}_{t_i} - \hat{Y}^{\pi,K,R,i_0,x'}_{t_i}|^2 &\leqslant (1 + q_i\Delta_i) E|E[\hat{Y}^{\pi,K,R,i_0,x}_{t_{i+1}} - \hat{Y}^{\pi,K,R,i_0,x'}_{t_{i+1}} | \mathcal{F}_{t_i}]|^2 \\
&+ \frac{(1 + q_i\Delta_i)}{1 + D} \Delta_i E|S^{\pi,i_0,x}_{t_i} - S^{\pi,i_0,x'}_{t_i}|^2 \\
&+ (1 + q_i\Delta_i)\Delta_i E\Big[|\hat{Y}^{\pi,K,R,i_0,x}_{t_{i+1}} - \hat{Y}^{\pi,K,R,i_0,x'}_{t_{i+1}}|^2 + \frac{1}{D}|\hat{Z}^{\pi,K,R,i_0,x}_{t_i} - \hat{Z}^{\pi,K,R,i_0,x'}_{t_i}|^2\Big].
\end{aligned} \tag{4.16}$$

Note, that

$$\sqrt{\Delta_i}\hat{Z}^{\pi,K,R,i_0,x}_{d,t_i} = \Big[\sqrt{\Delta_i}^{-1}E[\Delta W_{d,i}\hat{Y}^{\pi,K,R,i_0,x}_{t_{i+1}}|\mathcal{F}_{t_i}]\Big]_R, \quad d = 1,\dots,D.$$

In view of the Lipschitz continuity of $[\cdot]_R$ and (4.9) we achieve then

$$\begin{aligned}&\Delta_i\hat{E}|Z^{\pi,K,R,i_0,x}_{d,t_i} - \hat{Z}^{\pi,K,R,i_0,x'}_{d,t_i}|^2\\&\quad\leqslant E|\hat{Y}^{\pi,K,R,i_0,x}_{t_{i+1}} - \hat{Y}^{\pi,K,R,i_0,x'}_{t_{i+1}}|^2 - E|E[\hat{Y}^{\pi,K,R,i_0,x}_{t_{i+1}} - \hat{Y}^{\pi,K,R,i_0,x'}_{t_{i+1}}|\mathcal{F}_{t_i}]|^2,\end{aligned}$$

Applying this result on (4.16) together with Assumption 7 (ii) on S^π yields

$$\begin{aligned}&E|Y^{\pi,K,R,i_0,x}_{t_i} - Y^{\pi,K,R,i_0,x'}_{t_i}|^2\\&\quad\leqslant \big(1+\Delta_i[q_i(1+\Delta_i)+1]\big)E|Y^{\pi,K,R,i_0,x}_{t_{i+1}} - Y^{\pi,K,R,i_0,x'}_{t_{i+1}}|^2 + C\Delta_i|x-x'|^2.\end{aligned}$$

Making use of Gronwall's inequality and after that of the Lipschitz continuity of $[\cdot]_R$ and the Lipschitz condition on ϕ^π leads to

$$\begin{aligned}&E|\hat{Y}^{\pi,K,R,i_0,x}_{t_i} - \hat{Y}^{\pi,K,R,i_0,x'}_{t_i}|^2\\&\quad\leqslant e^{T(q_i(1+|\pi|)+1)}\Big(E|\hat{Y}^{\pi,K,R,i_0,x}_{t_N} - \hat{Y}^{\pi,K,R,i_0,x'}_{t_N}|^2 + CT|x-x'|^2\Big)\\&\quad\leqslant CE|\phi^\pi(X^{\pi,i_0,x}_{t_N}) - \phi^\pi(X^{\pi,i_0,x'}_{t_N})|^2 + C|x-x'|^2\\&\quad\leqslant C\Big(E|X^{\pi,i_0,x}_{t_N} - X^{\pi,i_0,x'}_{t_N}|^2 + |x-x'|^2\Big).\end{aligned}$$

Recalling Assumption 7 we can finish the proof. □

4.2.3 Simulation error

First, we translate the 'function'-based scheme (4.4) in a 'random'-variable based approach. To this end, we denote by $\sigma(\mathcal{X}^L \cup X^\pi_{t_i})$ the σ-algebra generated by $\mathcal{X}^L$ and $X^\pi_{t_i}$. Moreover, let $\mathcal{P}^L_i$ be an operator defined by

$$\begin{aligned}&\mathcal{P}^L_i\Big(\big(F(t_i,{}_\lambda S^\pi_{t_i},\hat{y}^{\pi,K,L}_{i+1}({}_\lambda X^\pi_{t_{i+1}}),\hat{z}^{\pi,K,L}_i({}_\lambda X^\pi_{t_i}))\big)_{\lambda=1,\dots,L}\Big)\\&\qquad = \Big(\eta_0'(i,{}_\lambda X^\pi_{t_i})\bar{\alpha}^{\pi,K,L}_i\Big)_{\lambda=1,\dots,L},\end{aligned}$$

where, by (4.4),

$$\bar{\alpha}^{\pi,K,L}_i = \arg\min_{\alpha\in\mathbb{R}^K}\frac{1}{L}\sum_{\lambda=1}^{L}|\eta_0(i,{}_\lambda X^\pi_{t_i})\alpha - F(t_i,{}_\lambda S^\pi_{t_i},\hat{y}^{\pi,K,L}_{i+1}({}_\lambda X^\pi_{t_{i+1}}),\hat{z}^{\pi,K,L}_i({}_\lambda X^\pi_{t_i}))|^2.$$

In other words, given some function $g(x)$ the operator $\mathcal{P}^L_i$ is an orthogonal projection with respect to the norm $(\frac{1}{L}\sum_{\lambda=1}^{L}|g({}_\lambda X^\pi_{t_i})|^2)^{1/2}$. Based on the definition of $\mathcal{P}^L_i$, we define also

$$\tilde{\mathcal{P}}^L_i\Big(F(t_i,S^\pi_{t_i},\hat{y}^{\pi,K,L}_{i+1}(X^\pi_{t_{i+1}}),\hat{z}^{\pi,K,L}_i(X^\pi_{t_i}))\Big) = \eta_0(i,X^\pi_{t_i})\bar{\alpha}^{\pi,K,L}_i.$$

With these definitions, we can reformulate (4.4). By definition, we have

$$\hat{Y}^{\pi,K,L}_{t_{i+1}} = \hat{y}^{\pi,K,L}_{i+1}(X^\pi_{t_{i+1}}) = \eta_0(i+1, X^\pi_{t_{i+1}})\hat{\alpha}^{\pi,K,L}_{i+1}.$$

Considering Assumption 6, we can also write

$$\begin{aligned}\hat{Z}^{\pi,K,L}_{d,t_i} &= \hat{z}^{\pi,K,L}_{d,i}(X^\pi_{t_i}) = \frac{1}{\Delta_i}\eta_d(i, X^\pi_{t_i})\hat{\alpha}^{\pi,K,L}_{i+1}\\ &= \frac{1}{\Delta_i}E[\Delta W_{d,i}\eta_0(i+1, X^\pi_{t_{i+1}})\hat{\alpha}^{\pi,K,L}_{i+1}|\,\sigma(\mathcal{X}^L \cup X^\pi_{t_i})]\\ &= \frac{1}{\Delta_i}E[\Delta W_{d,i}\hat{Y}^{\pi,K,L}_{t_{i+1}}|\,\sigma(\mathcal{X}^L \cup X^\pi_{t_i})].\end{aligned}$$

Similarly, we obtain

$$\begin{aligned}\hat{Y}^{\pi,K,L}_{t_i} &= \eta_0(i, X^\pi_{t_i})\hat{\alpha}^{\pi,K,L}_i\\ &= E[\eta_0(i+1, X^\pi_{t_{i+1}})\hat{\alpha}^{\pi,K,L}_{i+1}|\,\sigma(\mathcal{X}^L \cup X^\pi_{t_i})] - \Delta_i\eta_0(i, X^\pi_{t_i})\bar{\alpha}^{\pi,K,L}_i\\ &= E[\hat{Y}^{\pi,K,L}_{t_{i+1}}|\,\sigma(\mathcal{X}^L \cup X^\pi_{t_i})] - \Delta_i\tilde{\mathcal{P}}^L_i\big(F(t_i, S^\pi_{t_i}, \hat{Y}^{\pi,K,L}_{t_{i+1}}, \hat{Z}^{\pi,K,L}_{t_i})\big).\end{aligned}$$

For technical reasons, we additionally have to impose a truncation structure on (4.4) such that $(\hat{Y}^{\pi,K,L}_{t_i}, \hat{Z}^{\pi,K,L}_{t_i})_{t_i\in\pi}$ are bounded processes. However, we emphasize, that the truncations in essence have a technical character and are usually neglected in practical implementation. Hence, we set for $i = N-1, \dots, 0$

$$\begin{aligned}\hat{Y}^{\pi,K,R,L}_{t_N} &= \Big[\eta_0(N, X^\pi_{t_N})\hat{\alpha}^{\pi,K,L}_N\Big]_R,\\ \hat{Z}^{\pi,K,R,L}_{d,t_i} &= \Big[\frac{1}{\Delta_i}E[\Delta W_{d,i}\hat{Y}^{\pi,K,R,L}_{t_{i+1}}|\,\sigma(\mathcal{X}^L \cup X^\pi_{t_i})]\Big]_{R/\sqrt{\Delta_i}}, \quad d = 1, \dots, D\\ \hat{Y}^{\pi,K,R,L}_{t_i} &= \Big[E[\hat{Y}^{\pi,K,R,L}_{t_{i+1}}|\,\sigma(\mathcal{X}^L \cup X^\pi_{t_i})] - \Delta_i\tilde{\mathcal{P}}^L_i\big(F(t_i, S^\pi_{t_i}, \hat{Y}^{\pi,K,R,L}_{t_{i+1}}, \hat{Z}^{\pi,K,R,L}_{t_i})\big)\Big]_R.\end{aligned}$$

Our aim is now to examine the error

$$\max_{0\leqslant i\leqslant N}\frac{1}{L}\sum_{\lambda=1}^{L}E|\hat{Y}^{\pi,K,R}_{t_i} - \hat{Y}^{\pi,K,R,L}_{t_i}|^2 + \sum_{i=0}^{N-1}\Delta_i\frac{1}{L}\sum_{\lambda=1}^{L}E|\hat{Z}^{\pi,K,R}_{t_i} - \hat{Z}^{\pi,K,R,L}_{t_i}|^2.$$

Like in the original least-squares Monte Carlo scheme, we have to trace this error back to

$$\begin{aligned}&\max_{0\leqslant i\leqslant N}\frac{1}{L}\sum_{\lambda=1}^{L}E|\hat{y}^{\pi,K,R}_i({}_\lambda X^\pi_{t_i}) - \hat{y}^{\pi,K,R,L}_i({}_\lambda X^\pi_{t_i})|^2\\ &\qquad\qquad + \sum_{i=0}^{N-1}\Delta_i\frac{1}{L}\sum_{\lambda=1}^{L}E|\hat{z}^{\pi,K,R}_i({}_\lambda X^\pi_{t_i}) - \hat{z}^{\pi,K,R,L}_i({}_\lambda X^\pi_{t_i})|^2.\end{aligned}$$

For this purpose, we introduce for $i = 0, \dots, N-1$ the norms

$$\|g\|_{\mathcal{X}^L_{t_{i+1}}} = \sqrt{\frac{1}{L}\sum_{\lambda=1}^{L}|g({}_\lambda X^\pi_{t_{i+1}})|^2}, \quad \|g\|_{\bar{\mathcal{X}}^{\pi,t_i}_{t_{i+1}}} = \sqrt{\frac{1}{L}\sum_{\lambda=1}^{L}|g({}_\lambda \bar{X}^{\pi,t_i}_{t_{i+1}})|^2},$$

where $g : \mathbb{R}^{\tilde{D}} \to \mathbb{R}$ is some measurable function and $\bar{\mathcal{X}}^{\pi,t_i}_{t_{i+1}}$ is a set of so-called ghost samples. Clearly, we denote by $\bar{\mathcal{X}}^{\pi,t_i}_{t_{i+1}} = \{(\Delta_\lambda \bar{W}^{t_i}_i, {}_\lambda\bar{X}^{\pi,t_i}_{t_{i+1}}) | \lambda = 1,\dots,L\}$ an independent copy of $\mathcal{X}^L_{t_{i+1}} = \{(\Delta_\lambda W_i, {}_\lambda X^\pi_{t_{i+1}}) | \lambda = 1,\dots,L\}$ conditional to $\{{}_\lambda X^\pi_{t_i} | \lambda = 1,\dots,L\}$.

Lemma 22. *For all* $i = 0,\dots,N-1$ *we define by*

$$G_i = \left\{ [\eta_0(i,x)\alpha]_R - \hat{y}^{\pi,K,R}_i(x) | \, \alpha \in \mathbb{R}^K \right\}$$

sets of bounded functions. Furthermore, we denote for all $i = 0,\dots,N-1$

$$\mathcal{A}_{i+1} = \left\{ \forall g \in G_{i+1} \; : \; \|g\|_{\bar{\mathcal{X}}^{\pi,t_i}_{t_{i+1}}} - \|g\|_{\mathcal{X}^L_{t_{i+1}}} \leqslant \Delta_i^{\frac{\beta+2}{2}} \right\}.$$

Under the Assumptions 2 and 7 we have for $|\pi|$ *small enough and* $\beta \in (0,1]$

$$\begin{aligned}
&\max_{0\leqslant i\leqslant N} E\|\hat{y}^{\pi,K,R}_i - \hat{y}^{\pi,K,R,L}_i\|^2_{\mathcal{X}^L_{t_i}} + \sum_{i=0}^{N-1} \Delta_i E\|\hat{z}^{\pi,K,R}_i - \hat{z}^{\pi,K,R,L}_i\|^2_{\mathcal{X}^L_{t_i}} \\
&\quad \leqslant C \inf_{\alpha\in\mathbb{R}^K} E|\phi^\pi(X^\pi_{t_N}) - \eta_0(N, X^\pi_{t_N})\alpha|^2 \\
&\qquad + C\Big(\max_{0\leqslant i\leqslant N} E|Y^\pi_{t_i} - \hat{Y}^{\pi,K}_{t_i}|^2 + \sum_{i=0}^{N-1} \Delta_i E|Z^\pi_{t_i} - \hat{Z}^{\pi,K}_{t_i}|^2 \Big) \\
&\qquad + C\Big(\max_{0\leqslant i\leqslant N} E|\hat{Y}^{\pi,K}_{t_i} - \hat{Y}^{\pi,K,R}_{t_i}|^2 + \sum_{i=0}^{N-1} \Delta_i E|\hat{Z}^{\pi,K}_{t_i} - \hat{Z}^{\pi,K,R}_{t_i}|^2 \Big) \\
&\qquad + C \sum_{i=0}^{N-1} \Delta_i E|\mathcal{P}_{0,j}\big(F(t_i, S^\pi_{t_i}, Y^\pi_{t_{i+1}}, Z^\pi_{t_i})\big) - E[F(t_i, S^\pi_{t_i}, Y^\pi_{t_{i+1}}, Z^\pi_{t_i}) | X^\pi_{t_i}]|^2 \\
&\qquad + C|\pi|^\beta + CR^2 \sum_{i=0}^{N-1} \frac{1}{\Delta_i} \mathbb{P}\{[\mathcal{A}_{i+1}]^c\},
\end{aligned} \tag{4.17}$$

where C *is a constant depending on* κ, T, D, C_X *and* κ_R.

The following proof adapts the argumentation in Lemor et al. (2006) on our setting.

Proof. Preliminary definitions and abbreviations: First, we will introduce the coefficient $\beta^{\pi,K,R,L}_i$, which solves

$$\beta^{\pi,K,R,L}_i = \arg\min_{\alpha\in\mathbb{R}^K} \frac{1}{L} \sum_{\lambda=1}^{L} |\eta_0(i, {}_\lambda X^\pi_{t_i})\alpha - F(t_i, {}_\lambda S^\pi_{t_i}, \hat{y}^{\pi,K,R}_{i+1}({}_\lambda\bar{X}^{\pi,t_i}_{t_{i+1}}), \hat{z}^{\pi,K,R}_i({}_\lambda X^\pi_{t_i}))|^2.$$

In view of the definition of ${}_\lambda\bar{X}^{\pi,t_i}_{t_{i+1}}$ we have the following identities.

$$\begin{aligned}
&E[F(t_i, {}_\lambda S^\pi_{t_i}, \hat{y}^{\pi,K,R}_{i+1}({}_\lambda\bar{X}^{\pi,t_i}_{t_{i+1}}), \hat{z}^{\pi,K,R}_i({}_\lambda X^\pi_{t_i})) | \sigma(\mathcal{X}^L)] \\
&\quad = E[F(t_i, {}_\lambda S^\pi_{t_i}, \hat{y}^{\pi,K,R}_{i+1}({}_\lambda X^\pi_{t_{i+1}}), \hat{z}^{\pi,K,R}_i({}_\lambda X^\pi_{t_i})) | {}_\lambda X^\pi_{t_i}],
\end{aligned}$$

Thus, $E[\beta_i^{\pi,K,R,L} | \sigma(\mathcal{X}^L)]$ is the minimizer of

$$\frac{1}{L}\sum_{\lambda=1}^{L} |\eta_0(i, {}_\lambda X_{t_i}^\pi)\alpha - E[F(t_i, {}_\lambda S_{t_i}^\pi, \hat{y}_{i+1}^{\pi,K,R}({}_\lambda X_{t_{i+1}}^\pi), \hat{z}_i^{\pi,K,R}({}_\lambda X_{t_i}^\pi)) | {}_\lambda X_{t_i}^\pi]|^2.$$

For reasons of space, we will abbreviate the projection error of some $\mathcal{F}_T$-measurable random variable U. Clearly, we denote

$$\mathcal{R}_i(U) = E|\mathcal{P}_{0,i}(U) - E[U| X_{t_i}^\pi]|^2.$$

Error due to sample changes: For technical reasons the proof involves several so-called sample changes. To this end, we repeatedly carry out the following estimation:

$$\begin{aligned} &E\|\hat{y}_{i+1}^{\pi,K,R} - \hat{y}_{i+1}^{\pi,K,R,L}\|^2_{\bar{\mathcal{X}}^{\pi,t_i}_{t_{i+1}}} \\ &\quad\leqslant (1+\Delta_i)E\|\hat{y}_{i+1}^{\pi,K,R} - \hat{y}_{i+1}^{\pi,K,R,L}\|^2_{\mathcal{X}^L_{t_{i+1}}} \\ &\qquad + \frac{C}{\Delta_i}E\left[\left(\|\hat{y}_{i+1}^{\pi,K,R} - \hat{y}_{i+1}^{\pi,K,R,L}\|_{\bar{\mathcal{X}}^{\pi,t_i}_{t_{i+1}}} - \|\hat{y}_{i+1}^{\pi,K,R} - \hat{y}_{i+1}^{\pi,K,R,L}\|_{\mathcal{X}^L_{t_{i+1}}}\right)_+^2\right]. \end{aligned}$$

By the definition of $\mathcal{A}_{i+1}$ we receive

$$\begin{aligned} &E\|\hat{y}_{i+1}^{\pi,K,R} - \hat{y}_{i+1}^{\pi,K,R,L}\|^2_{\bar{\mathcal{X}}^{\pi,t_i}_{t_{i+1}}} \\ &\quad\leqslant (1+\Delta_i)E\|\hat{y}_{i+1}^{\pi,K,R} - \hat{y}_{i+1}^{\pi,K,R,L}\|^2_{\mathcal{X}^L_{t_{i+1}}} + C\Delta_i^{\beta+1} + \frac{C}{\Delta_i}R^2\mathbb{P}\{[\mathcal{A}_{i+1}]^c\}. \end{aligned} \tag{4.18}$$

Main proof: Our proof goes through the following steps. In Step 1 we give proof for the following estimate. Let $\bar{\alpha}_i^{\pi,K,R,L} \in \mathbb{R}^K$ be the minimizing coefficient vector of

$$\frac{1}{L}\sum_{\lambda=1}^{L} |\eta_0(i, {}_\lambda X_{t_i}^\pi)\alpha - F(t_i, {}_\lambda S_{t_i}^\pi, \hat{y}_{i+1}^{\pi,K,R,L}({}_\lambda X_{t_{i+1}}^\pi), \hat{z}_i^{\pi,K,R,L}({}_\lambda X_{t_i}^\pi))|^2.$$

Then, for every $\Gamma > 0$,

$$\begin{aligned} &\frac{1}{L}\sum_{\lambda=1}^{L} E|\eta_0(i, {}_\lambda X_{t_i}^\pi)\bar{\alpha}_i^{\pi,K,R,L} - \mathcal{P}_{0,i}\big(F(t_i, {}_\lambda S_{t_i}^\pi, \hat{y}_{i+1}^{\pi,K,R}({}_\lambda \bar{X}_{t_{i+1}}^{\pi,t_i}), \hat{z}_i^{\pi,K,R}({}_\lambda X_{t_i}^\pi))\big)|^2 \\ &\quad\leqslant \gamma\mathcal{R}_i\big(F(t_i, S_{t_i}^\pi, \hat{Y}_{t_{i+1}}^{\pi,K,R}, \hat{Z}_{t_i}^{\pi,K,R})\big) + \gamma E\|\hat{y}_{i+1}^{\pi,K,R} - \hat{y}_{i+1}^{\pi,K,R,L}\|^2_{\mathcal{X}^L_{t_{i+1}}} \\ &\qquad + \frac{\gamma}{\Gamma}E\|\hat{z}_i^{\pi,K,R} - \hat{z}_i^{\pi,K,R,L}\|^2_{\mathcal{X}^L_{t_i}} + C|\pi|. \end{aligned} \tag{4.19}$$

with $\gamma = 4 + (2+\Gamma)\kappa^2$. Applying Step 1, we will show in Step 2 that

$$\begin{aligned} E\|\hat{y}_i^{\pi,K,R} - \hat{y}_i^{\pi,K,R,L}\|^2_{\mathcal{X}^L_{t_i}} &\leqslant C\inf_{\alpha\in\mathbb{R}^K} E|\phi^\pi(X_{t_N}^\pi) - \eta_0(N, X_{t_N}^\pi)\alpha|^2 \\ &\quad + C\sum_{j=i}^{N-1}\Delta_j\mathcal{R}_j\big(f(t_j, S_{t_j}^\pi, \hat{Y}_{t_{j+1}}^{\pi,K,R}, \hat{Z}_{t_j}^{\pi,K,R})\big) + C|\pi|^\beta + CR^2\sum_{j=i}^{N-1}\frac{1}{\Delta_j}\mathbb{P}\{[A_{j+1}]^c\}. \end{aligned} \tag{4.20}$$

In Step 3 we will turn to Z-part and deduce that

$$\sum_{i=0}^{N-1} \Delta_i E\|\hat{z}_i^{\pi,K,R} - \hat{z}_i^{\pi,K,R,L}\|^2_{\mathcal{X}^L_{t_i}} \leqslant C \inf_{\alpha\in\mathbb{R}^K} E|\varphi^{\pi}(X^{\pi}_{t_N}) - \eta_0(N, X^{\pi}_{t_N})\alpha|^2$$
$$+ C \sum_{i=0}^{N-1} \Delta_i \mathcal{R}_i\big(F(t_i, S^{\pi}_{t_i}, \hat{Y}^{\pi,K,R}_{t_{i+1}}, \hat{Z}^{\pi,K,R}_{t_i})\big) + C|\pi|^{\beta} + CR^2 \sum_{i=0}^{N-1} \frac{1}{\Delta_i}\mathbb{P}\{[\mathcal{A}_{i+1}]^c\}. \tag{4.21}$$

Combining the results of Step 2 and 3 with the following calculation completes then the proof.

$$\begin{aligned}
&\mathcal{R}_i\big(F(t_i, S^{\pi}_{t_i}, \hat{Y}^{\pi,K,R}_{t_{i+1}}, \hat{Z}^{\pi,K,R}_{t_i})\big)\\
&\leqslant CE\big|\mathcal{P}_{0,i}\big(F(t_i, S^{\pi}_{t_i}, Y^{\pi}_{t_{i+1}}, Z^{\pi}_{t_i})\big) - E[F(t_i, S^{\pi}_{t_i}, Y^{\pi}_{t_{i+1}}, Z^{\pi}_{t_i})|\, X^{\pi}_{t_i}]\big|^2\\
&\quad + CE\big|F(t_i, S^{\pi}_{t_i}, Y^{\pi}_{t_{i+1}}, Z^{\pi}_{t_i}) - F(t_i, S^{\pi}_{t_i}, \hat{Y}^{\pi,K}_{t_{i+1}}, \hat{Z}^{\pi,K}_{t_i})\big|^2\\
&\quad + CE\big|F(t_i, S^{\pi}_{t_i}, \hat{Y}^{\pi,K}_{t_{i+1}}, \hat{Z}^{\pi,K}_{t_i}) - F(t_i, S^{\pi}_{t_i}, \hat{Y}^{\pi,K,R}_{t_{i+1}}, \hat{Z}^{\pi,K,R}_{t_i})\big|^2\\
&\leqslant C\mathcal{R}_i\big(F(t_i, S^{\pi}_{t_i}, Y^{\pi}_{t_{i+1}}, Z^{\pi}_{t_i})\big) + C\left(E|Y^{\pi}_{t_{i+1}} - \hat{Y}^{\pi,K}_{t_{i+1}}|^2 + E|Z^{\pi}_{t_i} - \hat{Z}^{\pi,K}_{t_i}|^2\right)\\
&\quad + C\left(E|\hat{Y}^{\pi,K}_{t_{i+1}} - \hat{Y}^{\pi,K,R}_{t_{i+1}}|^2 + E|\hat{Z}^{\pi,K}_{t_i} - \hat{Z}^{\pi,K,R}_{t_i}|^2\right).
\end{aligned}$$

Step 1: Considering the definition of $E[\beta_i^{\pi,K,R,L}|\, \sigma(\mathcal{X}^L)]$ and by Young's inequality we receive for some $\Gamma > 0$ and $\gamma = 4 + (2+\Gamma)\kappa^2$

$$\begin{aligned}
&\frac{1}{L}\sum_{\lambda=1}^{L} E|\eta_0(i, {}_{\lambda}X^{\pi}_{t_i})\bar{\alpha}_i^{\pi,K,R,L} - \mathcal{P}_{0,i}\big(F(t_i, {}_{\lambda}S^{\pi}_{t_i}, \hat{y}^{\pi,K,R}_{i+1}({}_{\lambda}\bar{X}^{\pi,t_i}_{t_{i+1}}), \hat{z}_i^{\pi,K,R}({}_{\lambda}X^{\pi}_{t_i}))\big)|^2\\
&\leqslant \frac{\gamma}{4} 2\frac{1}{L}\sum_{\lambda=1}^{L} E|\mathcal{P}_{0,i}\big(F(t_i, {}_{\lambda}S^{\pi}_{t_i}, \hat{y}^{\pi,K,R}_{i+1}({}_{\lambda}\bar{X}^{\pi,t_i}_{t_{i+1}}), \hat{z}_i^{\pi,K,R}({}_{\lambda}X^{\pi}_{t_i}))\big)\\
&\qquad\qquad - E[F(t_i, {}_{\lambda}S^{\pi}_{t_i}, \hat{y}^{\pi,K,R}_{i+1}({}_{\lambda}\bar{X}^{\pi,t_i}_{t_{i+1}}), \hat{z}_i^{\pi,K,R}({}_{\lambda}X^{\pi}_{t_i}))|\, \sigma(\mathcal{X}^L)]|^2\\
&\quad + \frac{\gamma}{4} 2\frac{1}{L}\sum_{\lambda=1}^{L} E|E[F(t_i, {}_{\lambda}S^{\pi}_{t_i}, \hat{y}^{\pi,K,R}_{i+1}({}_{\lambda}\bar{X}^{\pi,t_i}_{t_{i+1}}), \hat{z}_i^{\pi,K,R}({}_{\lambda}X^{\pi}_{t_i}))|\, \sigma(\mathcal{X}^L)]\\
&\qquad\qquad - \eta_0(i, {}_{\lambda}X^{\pi}_{t_i})E[\beta_i^{\pi,K,R,L}|\, \sigma(\mathcal{X}^L)]|^2\\
&\quad + \frac{\gamma}{4}\frac{4}{(2+\Gamma)\kappa^2}\frac{1}{L}\sum_{\lambda=1}^{L} E|\eta_0(i, {}_{\lambda}X^{\pi}_{t_i})E[\beta_i^{\pi,K,R,L}|\, \sigma(\mathcal{X}^L)] - \eta_0(i, {}_{\lambda}X^{\pi}_{t_i})\bar{\alpha}_i^{\pi,K,R,L}|^2\\
&= (\mathrm{I}) + (\mathrm{II}) + (\mathrm{III}).
\end{aligned}$$

The summands of (I) are identically distributed for all $\lambda = 1, \ldots, L$. Hence, we have

$$(\mathrm{I}) = \frac{\gamma}{2}\mathcal{R}_i\big(f(t_i, S^{\pi}_{t_i}, \hat{Y}^{\pi,K,R}_{t_{i+1}}, \hat{Z}^{\pi,K,R}_{t_i})\big).$$

In view of the definition of $E[\beta_i^{\pi,K,R,L}|\,\sigma(\mathcal{X}^L)]$ we obtain

$$\begin{aligned}
(II) &= \frac{\gamma}{2} E\Big[\inf_{\alpha\in\mathbb{R}^K}\frac{1}{L}\sum_{\lambda=1}^{L}|\eta_0(i,{}_\lambda X_{t_i}^\pi)\alpha \\
&\qquad - E[F(t_i,{}_\lambda S_{t_i}^\pi,\hat{y}_{i+1}^{\pi,K,R}({}_\lambda X_{t_{i+1}}^\pi),\hat{z}_i^{\pi,K,R}({}_\lambda X_{t_i}^\pi))|\,{}_\lambda X_{t_i}^\pi]|^2\Big] \\
&\leqslant \frac{\gamma}{2}\inf_{\alpha\in\mathbb{R}^K} E\Big[|\eta_0(i,X_{t_i}^\pi)\alpha - E[F(t_i,S_{t_i}^\pi,\hat{y}_{i+1}^{\pi,K,R}(X_{t_{i+1}}^\pi),\hat{z}_i^{\pi,K,R}(X_{t_i}^\pi))|\,X_{t_i}^\pi]|^2\Big] \\
&= \frac{\gamma}{2}\mathcal{R}_i\big(F(t_i,S_{t_i}^\pi,\hat{Y}_{t_{i+1}}^{\pi,K,R},\hat{Z}_{t_i}^{\pi,K,R})\big).
\end{aligned}$$

Turning to (III) we exploit first the fact that $\bar{\alpha}_i^{\pi,K,R,L}$ is $\sigma(\mathcal{X}^L)$-measurable, then the contraction property of the operator $\mathcal{P}_i^L$ and the Lipschitz continuity of F and finally Young's inequality.

$$\begin{aligned}
(III) &\leqslant \gamma\frac{1}{(2+\Gamma)\kappa^2}\frac{1}{L}\sum_{\lambda=1}^{L} E|\eta_0(i,{}_\lambda X_{t_i}^\pi)\beta_i^{\pi,K,R,L} - \eta_0(i,{}_\lambda X_{t_i}^\pi)\bar{\alpha}_i^{\pi,K,R,L}|^2 \\
&\leqslant \gamma\frac{1}{(2+\Gamma)\kappa^2}\frac{1}{L}\sum_{\lambda=1}^{L} E|F(t_i,{}_\lambda S_{t_i}^\pi,\hat{y}_{i+1}^{\pi,K,R}({}_\lambda\bar{X}_{t_{i+1}}^{\pi,t_i}),\hat{z}_i^{\pi,K,R}({}_\lambda X_{t_i}^\pi)) \\
&\qquad\qquad - F(t_i,{}_\lambda S_{t_i}^\pi,\hat{y}_{i+1}^{\pi,K,R,L}({}_\lambda X_{t_{i+1}}^\pi),\hat{z}_i^{\pi,K,R,L}({}_\lambda X_{t_i}^\pi))|^2 \\
&\leqslant \gamma\frac{1}{(2+\Gamma)}(1+\Gamma/2)\frac{1}{L}\sum_{\lambda=1}^{L} E|\hat{y}_{i+1}^{\pi,K,R}({}_\lambda\bar{X}_{t_{i+1}}^{\pi,t_i}) - \hat{y}_{i+1}^{\pi,K,R,L}({}_\lambda X_{t_{i+1}}^\pi)|^2 \\
&\quad + \gamma\frac{1}{(2+\Gamma)}(1+\frac{2}{\Gamma})E\|\hat{z}_i^{\pi,K,R} - \hat{z}_i^{\pi,K,R,L}\|_{\mathcal{X}_{t_i}^L}^2 \\
&= (IIIa) + (IIIb).
\end{aligned}$$

The Lipschitz continuity of $\hat{y}_{i+1}^{\pi,K,R}(x)$ and Assumption 7 (iii) lead to

$$\begin{aligned}
(IIIa) &\leqslant \gamma E\|\hat{y}_{i+1}^{\pi,K,R} - \hat{y}_{i+1}^{\pi,K,R,L}\|_{\mathcal{X}_{t_{i+1}}^L}^2 + \gamma\kappa_R\frac{1}{L}\sum_{\lambda=1}^{L} E|{}_\lambda\bar{X}_{t_{i+1}}^{\pi,t_i} - {}_\lambda X_{t_{i+1}}^\pi|^2 \\
&\leqslant \gamma E\|\hat{y}_{i+1}^{\pi,K,R} - \hat{y}_{i+1}^{\pi,K,R,L}\|_{\mathcal{X}_{t_{i+1}}^L}^2 + C|\pi|.
\end{aligned}$$

Summarizing the estimates of (I), (II) and (III) we get the result in (4.19).
Step 2: Note, that

$$\begin{aligned}
\hat{y}_i^{\pi,K,R}({}_\lambda X_{t_i}^\pi) &= \Big[E[\hat{y}_{i+1}^{\pi,K,R}({}_\lambda\bar{X}_{t_{i+1}}^{\pi,t_i})|\,\sigma(\mathcal{X}^L)] \\
&\qquad - \Delta_i\mathcal{P}_{0,i}\big(F(t_i,{}_\lambda S_{t_i}^\pi,\hat{y}_{i+1}^{\pi,K,R}({}_\lambda\bar{X}_{t_{i+1}}^{\pi,t_i}),\hat{z}_i^{\pi,K,R}({}_\lambda X_{t_i}^\pi))\big)\Big]_R, \\
\hat{y}_i^{\pi,K,R,L}({}_\lambda X_{t_i}^\pi) &= \Big[E[\hat{y}_{i+1}^{\pi,K,R,L}({}_\lambda\bar{X}_{t_{i+1}}^{\pi,t_i})|\,\sigma(\mathcal{X}^L)] - \Delta_i\eta_0(i,{}_\lambda X_{t_i}^\pi)\bar{\alpha}_i^{\pi,K,R,L}\Big]_R.
\end{aligned}$$

Bearing these identities in mind, we first employ the Lipschitz-continuity of $[\cdot]_R$ and then Young's inequality.

$$
\begin{aligned}
&E\|\hat{y}_i^{\pi,K,R} - \hat{y}_i^{\pi,K,R,L}\|^2_{\mathcal{X}^L_{t_i}} \\
&\leqslant (1+\tilde{\gamma}\Delta_i)\frac{1}{L}\sum_{\lambda=1}^{L} E|E[\hat{y}_{i+1}^{\pi,K,R}({}_\lambda\bar{X}^{\pi,t_i}_{t_{i+1}}) - \hat{y}_{i+1}^{\pi,K,R,L}({}_\lambda\bar{X}^{\pi,t_i}_{t_{i+1}})|\,\sigma(\mathcal{X}^L)]|^2 \\
&\quad + (1+\tilde{\gamma}\Delta_i)\frac{\Delta_i}{\tilde{\gamma}}\frac{1}{L}\sum_{\lambda=1}^{L} E|\eta_0(i,{}_\lambda X^\pi_{t_i})\bar{\alpha}_i^{\pi,K,R,L} \\
&\qquad - \mathcal{P}_{0,i}\big(F(t_i,{}_\lambda S^\pi_{t_i},\hat{y}_{i+1}^{\pi,K,R}({}_\lambda\bar{X}^{\pi,t_i}_{t_{i+1}}),\hat{z}_i^{\pi,K,R}({}_\lambda X^\pi_{t_i}))\big)|^2,
\end{aligned}
\tag{4.22}
$$

where $\tilde{\gamma}$ is a positive constant. The application of (4.19) with $\Gamma = D$ and $\tilde{\gamma} = \gamma = 4 + (2+D)\kappa^2$ yields

$$
\begin{aligned}
&E\|\hat{y}_i^{\pi,K,R} - \hat{y}_i^{\pi,K,R,L}\|^2_{\mathcal{X}^L_{t_i}} \\
&\leqslant (1+\tilde{\gamma}\Delta_i)\frac{1}{L}\sum_{\lambda=1}^{L} E|E[\hat{y}_{i+1}^{\pi,K,R}({}_\lambda\bar{X}^{\pi,t_i}_{t_{i+1}}) - \hat{y}_{i+1}^{\pi,K,R,L}({}_\lambda\bar{X}^{\pi,t_i}_{t_{i+1}})|\,\sigma(\mathcal{X}^L)]|^2 \\
&\quad + (1+\tilde{\gamma}\Delta_i)\Delta_i E\left[\|\hat{y}_{i+1}^{\pi,K,R} - \hat{y}_{i+1}^{\pi,K,R,L}\|^2_{\mathcal{X}^L_{t_{i+1}}} + \frac{1}{D}\|\hat{z}_i^{\pi,K,R} - \hat{z}_i^{\pi,K,R,L}\|^2_{\mathcal{X}^L_{t_i}}\right] \\
&\quad + (1+\tilde{\gamma}\Delta_i)\Delta_i\mathcal{R}_i\big(F(t_i,S^\pi_{t_i},\hat{Y}^{\pi,K,R}_{t_{i+1}},\hat{Z}^{\pi,K,R}_{t_i})\big) + C\Delta_i|\pi|.
\end{aligned}
\tag{4.23}
$$

Regarding the third summand of the right-hand side of the above summand, we employ the sample set $\bar{\mathcal{X}}^{t_i,L}$ in order to consider the dependency structure of $\hat{z}_i^{\pi,K,R,L}(x)$ correctly. In view of the definitions of $\hat{z}_i^{\pi,K,R}(x)$ and $\hat{z}_i^{\pi,K,R,L}(x)$, respectively, and the Lipschitz continuity of $[\cdot]_R$ we achieve

$$
\begin{aligned}
&\sqrt{\Delta_i}|\hat{z}_{d,i}^{\pi,K,R}({}_\lambda X^\pi_{t_i}) - \hat{z}_{d,i}^{\pi,K,R,L}({}_\lambda X^\pi_{t_i})| \\
&\leqslant |(\sqrt{\Delta_i})^{-1}E[\Delta_\lambda \bar{W}^{t_i}_{d,i}\left\{\hat{y}_{i+1}^{\pi,K,R}({}_\lambda\bar{X}^{\pi,t_i}_{t_{i+1}}) - \hat{y}_{i+1}^{\pi,K,R,L}({}_\lambda\bar{X}^{\pi,t_i}_{t_{i+1}})\right\}|\,\sigma(\mathcal{X}^L)]|
\end{aligned}
$$

For an analogous application of Lemma 18, (4.9) we set $y^1_{t_{i+1}} = \hat{y}_{i+1}^{\pi,K,R}({}_\lambda\bar{X}^{\pi,t_i}_{t_{i+1}})$ and $y^2_{t_{i+1}} = \hat{y}_{i+1}^{\pi,K,R,L}({}_\lambda\bar{X}^{\pi,t_i}_{t_{i+1}})$. Considering $\sigma(\mathcal{X}^L)$ instead of $\mathcal{F}_{t_i}$, we get

$$
\begin{aligned}
&\Delta_i E|\hat{z}_{d,i}^{\pi,K,R}({}_\lambda X^\pi_{t_i}) - \hat{z}_{d,i}^{\pi,K,R,L}({}_\lambda X^\pi_{t_i})|^2 \\
&\leqslant E|\hat{y}_{i+1}^{\pi,K,R}({}_\lambda\bar{X}^{\pi,t_i}_{t_{i+1}}) - \hat{y}_{i+1}^{\pi,K,R,L}({}_\lambda\bar{X}^{\pi,t_i}_{t_{i+1}})|^2 \\
&\quad - E|E[\hat{y}_{i+1}^{\pi,K,R}({}_\lambda\bar{X}^{\pi,t_i}_{t_{i+1}}) - \hat{y}_{i+1}^{\pi,K,R,L}({}_\lambda\bar{X}^{\pi,t_i}_{t_{i+1}})|\,\sigma(\mathcal{X}^L)]|^2.
\end{aligned}
\tag{4.24}
$$

Inserting this inequality in (4.23) gives

$$\begin{aligned} E\|\hat{y}_i^{\pi,K,R} - \hat{y}_i^{\pi,K,R,L}\|^2_{\mathcal{X}^L_{t_i}} &\leqslant (1+\tilde{\gamma}\Delta_i) E\|\hat{y}_{i+1}^{\pi,K,R} - \hat{y}_{i+1}^{\pi,K,R,L}\|^2_{\bar{\mathcal{X}}^{\pi,t_i}_{t_{i+1}}} \\ &+ (1+\tilde{\gamma}\Delta_i)\Delta_i E\|\hat{y}_{i+1}^{\pi,K,R} - \hat{y}_{i+1}^{\pi,K,R,L}\|^2_{\mathcal{X}^L_{t_{i+1}}} \\ &+ (1+\tilde{\gamma}\Delta_i)\Delta_i \mathcal{R}_i\big(F(t_i, S^\pi_{t_i}, \hat{Y}^{\pi,K,R}_{t_{i+1}}, \hat{Z}^{\pi,K,R}_{t_i})\big) + C\Delta_i|\pi|. \end{aligned}$$

A sample change in $\hat{y}_{i+1}^{\pi,K,R}({}_\lambda\bar{X}^{\pi,t_i}_{t_{i+1}}) - \hat{y}_{i+1}^{\pi,K,R,L}({}_\lambda\bar{X}^{\pi,t_i}_{t_{i+1}})$ leads to

$$\begin{aligned} E\|\hat{y}_i^{\pi,K,R} - \hat{y}_i^{\pi,K,R,L}\|^2_{\mathcal{X}^L_{t_i}} &\leqslant (1+\tilde{\gamma}\Delta_i)(1+2\Delta_i) E\|\hat{y}_{i+1}^{\pi,K,R} - \hat{y}_{i+1}^{\pi,K,R,L}\|^2_{\mathcal{X}^L_{t_{i+1}}} \\ &+ C\Delta_i \mathcal{R}_i\big(F(t_i, S^\pi_{t_i}, \hat{Y}^{\pi,K,R}_{t_{i+1}}, \hat{Z}^{\pi,K,R}_{t_i})\big) + C\Delta_i|\pi| + C\Delta_i^{\beta+1} + \frac{C}{\Delta_i} R^2 \mathbb{P}\{[\mathcal{A}_{i+1}]^c\}. \end{aligned}$$

Thanks to Gronwall's inequality we receive

$$\begin{aligned} E\|\hat{y}_i^{\pi,K,R} - \hat{y}_i^{\pi,K,R,L}\|^2_{\mathcal{X}^L_{t_i}} &\leqslant e^{(\gamma(1+2|\pi|)+2)T} E\|\hat{y}_N^{\pi,K,R} - \hat{y}_N^{\pi,K,R,L}\|^2_{\mathcal{X}^L_{t_N}} \\ &+ C\sum_{j=i}^{N-1} \Delta_j \mathcal{R}_j\big(f(t_j, S^\pi_{t_j}, \hat{Y}^{\pi,K,R}_{t_{j+1}}, \hat{Z}^{\pi,K,R}_{t_j})\big) + C|\pi|^\beta + CR^2 \sum_{j=i}^{N-1} \frac{1}{\Delta_j} \mathbb{P}\{[\mathcal{A}_{j+1}]^c\}. \end{aligned}$$

The definition of $\hat{y}_N^{\pi,K,R}(x)$ and $\hat{y}_N^{\pi,K,R,L}(x)$ and the Lipschitz continuity of $[\cdot]_R$ yield

$$\begin{aligned} &E\|\hat{y}_i^{\pi,K,R} - \hat{y}_i^{\pi,K,R,L}\|^2_{\mathcal{X}^L_{t_i}} \\ &\leqslant e^{(\gamma(1+2|\pi|)+2)T} E\Big[\inf_{\alpha\in\mathbb{R}^K} \frac{1}{L}\sum_{\lambda=1}^L |\phi({}_\lambda X^\pi_{t_N}) - \eta_0(N, {}_\lambda X^\pi_{t_N})\alpha|^2\Big] \\ &\quad + C\sum_{j=i}^{N-1} \Delta_j \mathcal{R}_j\big(f(t_j, S^\pi_{t_j}, \hat{Y}^{\pi,K,R}_{t_{j+1}}, \hat{Z}^{\pi,K,R}_{t_j})\big) + C|\pi|^\beta + CR^2 \sum_{j=i}^{N-1} \frac{1}{\Delta_j} \mathbb{P}\{[\mathcal{A}_{j+1}]^c\} \\ &\leqslant C \inf_{\alpha\in\mathbb{R}^K} E\Big[|\phi(X^\pi_{t_N}) - \eta_0(N, X^\pi_{t_N})\alpha|^2\Big] \\ &\quad + C\sum_{j=i}^{N-1} \Delta_j \mathcal{R}_j\big(f(t_j, S^\pi_{t_j}, \hat{Y}^{\pi,K,R}_{t_{j+1}}, \hat{Z}^{\pi,K,R}_{t_j})\big) + C|\pi|^\beta + CR^2 \sum_{j=i}^{N-1} \frac{1}{\Delta_j} \mathbb{P}\{[\mathcal{A}_{j+1}]^c\}. \end{aligned}$$

This completes Step 2.

Step 3: Recalling the estimate in (4.24), we get by a change of samples

$$\begin{aligned} &\Delta_i E\|\hat{z}_{d,i}^{\pi,K,R} - \hat{z}_{d,i}^{\pi,K,R,L}\|^2_{\mathcal{X}^L_{t_i}} \\ &\leqslant (1+\Delta_i)\frac{1}{L} E\|\hat{y}_{i+1}^{\pi,K,R} - \hat{y}_{i+1}^{\pi,K,R,L}\|^2_{\mathcal{X}^L_{t_{i+1}}} + C\Delta_i^{\beta+1} + \frac{C}{\Delta_i}\mathbb{P}\{[\mathcal{A}_{i+1}]^c\} \\ &\quad - \frac{1}{L}\sum_{\lambda=1}^L E|E[\hat{y}_{i+1}^{\pi,K,R}({}_\lambda\bar{X}^{\pi,t_i}_{t_{i+1}}) - \hat{y}_{i+1}^{\pi,K,R,L}({}_\lambda\bar{X}^{\pi,t_i}_{t_{i+1}})|\,\sigma(\mathcal{X}^L)]|^2, \end{aligned}$$

for $i = 0, \dots, N-2$. Making use of the inequality in (4.22) gives

$$
\begin{aligned}
&\Delta_i E\|\hat{z}^{\pi,K,R}_{d,i} - \hat{z}^{\pi,K,R,L}_{d,i}\|^2_{\mathcal{X}^L_{t_i}} \\
&\quad\leqslant (1+\Delta_i)(1+\tilde{\gamma}\Delta_{i+1}) \\
&\qquad\quad \times \frac{1}{L}\sum_{\lambda=1}^{L} E|E[\hat{y}^{\pi,K,R}_{i+2}({}_\lambda\bar{X}^{\pi,t_{i+1}}_{t_{i+2}}) - \hat{y}^{\pi,K,R,L}_{i+2}({}_\lambda\bar{X}^{\pi,t_{i+1}}_{t_{i+2}}) |\, \sigma(\mathcal{X}^L)]|^2 \\
&\qquad + (1+\Delta_i)(1+\tilde{\gamma}\Delta_{i+1})\frac{\Delta_{i+1}}{\tilde{\gamma}}\frac{1}{L}\sum_{\lambda=1}^{L} E|\eta_0(i+1, {}_\lambda X^\pi_{t_{i+1}})\bar{\alpha}^{\pi,K,R,L}_{i+1} \\
&\qquad\qquad - \mathcal{P}_{0,i+1}\big(f(t_{i+1}, {}_\lambda S^\pi_{t_{i+1}}, \hat{y}^{\pi,K,R}_{i+2}({}_\lambda\bar{X}^{\pi,t_{i+1}}_{t_{i+2}}), \hat{z}^{\pi,K,R}_{i+1}({}_\lambda X^\pi_{t_{i+1}}))\big)|^2 \\
&\qquad + C\Delta_i^{\beta+1} + \frac{C}{\Delta_i}\mathbb{P}\{[\mathcal{A}_{i+1}]^c\} \\
&\qquad - \frac{1}{L}\sum_{\lambda=1}^{L} E|E[\hat{y}^{\pi,K,R}_{i+1}({}_\lambda\bar{X}^{\pi,t_i}_{t_{i+1}}) - \hat{y}^{\pi,K,R,L}_{i+1}({}_\lambda\bar{X}^{\pi,t_i}_{t_{i+1}}) |\, \sigma(\mathcal{X}^L)]|^2.
\end{aligned}
$$

By summing up from $i = 0$ to $N-1$, we get

$$
\begin{aligned}
&\sum_{i=0}^{N-1}\Delta_i E\|\hat{z}^{\pi,K,R}_{d,i} - \hat{z}^{\pi,K,R,L}_{d,i}\|^2_{\mathcal{X}^L_{t_i}} \leqslant (1+\Delta_{N-1})E\|\hat{y}^{\pi,K,R}_N - \hat{y}^{\pi,K,R,L}_N\|^2_{\mathcal{X}^L_{t_N}} \\
&\quad + C\sum_{i=0}^{N-2}(\Delta_i + \Delta_{i+1}) \\
&\qquad\quad \times \frac{1}{L}\sum_{\lambda=1}^{L} E|E[\hat{y}^{\pi,K,R}_{i+2}({}_\lambda\bar{X}^{\pi,t_{i+1}}_{t_{i+2}}) - \hat{y}^{\pi,K,R,L}_{i+2}({}_\lambda\bar{X}^{\pi,t_{i+1}}_{t_{i+2}}) |\, \sigma(\mathcal{X}^L)]|^2 \\
&\quad + \sum_{i=0}^{N-2}(1+\Delta_i)(1+\tilde{\gamma}\Delta_{i+1})\frac{\Delta_{i+1}}{\tilde{\gamma}}\frac{1}{L}\sum_{\lambda=1}^{L} E|\eta_0(i+1, {}_\lambda X^\pi_{t_{i+1}})\bar{\alpha}^{\pi,K,R,L}_{i+1} \\
&\qquad\qquad - \mathcal{P}_{0,i+1}\big(f(t_{i+1}, {}_\lambda S^\pi_{t_{i+1}}, \hat{y}^{\pi,K,R}_{i+2}({}_\lambda\bar{X}^{\pi,t_{i+1}}_{t_{i+2}}), \hat{z}^{\pi,K,R}_{i+1}({}_\lambda X^\pi_{t_{i+1}}))\big)|^2 \\
&\quad + C|\pi|^\beta + \sum_{i=0}^{N-1}\frac{CR^2}{\Delta_i}\mathbb{P}\{[\mathcal{A}_{i+1}]^c\}.
\end{aligned}
$$

Now, we conduct a sample change in the second summand of the above inequality and exploit (4.19) with $\Gamma = 1$. Hence,

$$\begin{aligned}
\sum_{i=0}^{N-1} \Delta_i E\|\hat{z}_{d,i}^{\pi,K,R} - \hat{z}_{d,i}^{\pi,K,R,L}\|^2_{\mathcal{X}^L_{t_i}} &\leqslant C \max_{0\leqslant i\leqslant N} E\|\hat{y}_i^{\pi,K,R} - \hat{y}_i^{\pi,K,R,L}\|^2_{\mathcal{X}^L_{t_i}} \\
&+ \sum_{i=0}^{N-2} (1+\Delta_i)(1+\tilde{\gamma}\Delta_{i+1})\frac{\gamma}{\tilde{\gamma}}\Delta_{i+1} E\|\hat{y}_{i+2}^{\pi,K,R} - \hat{y}_{i+2}^{\pi,K,R,L}\|^2_{\mathcal{X}^L_{t_{i+2}}} \\
&+ \sum_{i=0}^{N-2} (1+\Delta_i)(1+\tilde{\gamma}\Delta_{i+1})\frac{\gamma}{\tilde{\gamma}}\Delta_{i+1} E\|\hat{z}_{i+1}^{\pi,K,R} - \hat{z}_{i+1}^{\pi,K,R,L}\|^2_{\mathcal{X}^L_{t_{i+1}}} \\
&+ C\sum_{i=0}^{N-2} \Delta_{i+1}\mathcal{R}_{i+1}\big(f(t_{+1}i, S^\pi_{t_{i+1}}, \hat{Y}^{\pi,K,R}_{t_{i+2}}, \hat{Z}^{\pi,K,R}_{t_{i+1}})\big) \\
&+ C|\pi|^\beta(1+|\pi|) + \sum_{i=0}^{N-1} CR^2(\frac{1}{\Delta_i}+1)\mathbb{P}\{[\mathcal{A}_{i+1}]^c\}.
\end{aligned}$$

For $\tilde{\gamma} = 8D\gamma$ and $|\pi| < \min\{1, 1/\tilde{\gamma}\}$ we obtain then

$$\begin{aligned}
\sum_{i=0}^{N-1} \Delta_i E\|\hat{z}_{i}^{\pi,K,R} - \hat{z}_{i}^{\pi,K,R,L}\|^2_{\mathcal{X}^L_{t_i}} &\leqslant C \max_{0\leqslant i\leqslant N} E\|\hat{y}_i^{\pi,K,R} - \hat{y}_i^{\pi,K,R,L}\|^2_{\mathcal{X}^L_{t_i}} \\
&+ C\sum_{i=0}^{N-2} \Delta_{i+1}\mathcal{R}_{i+1}\big(f(t_{+1}i, S^\pi_{t_{i+1}}, \hat{Y}^{\pi,K,R}_{t_{i+2}}, \hat{Z}^{\pi,K,R}_{t_{i+1}})\big) \\
&+ C|\pi|^\beta + CR^2\sum_{i=0}^{N-1} \frac{1}{\Delta_i}\mathbb{P}\{[\mathcal{A}_{i+1}]^c\}.
\end{aligned}$$

By employing the result of Step 2, see (4.20), we can finish the proof of Step 3. Hence, the proof is complete. □

Next, we aim at giving an upper bound for $\mathbb{P}\{[\mathcal{A}_{i+1}]^c\}$, $i = 0, \dots, N-1$ with

$$\mathcal{A}_{i+1} = \left\{\forall g \in G_{i+1} \;:\; \|g\|_{\bar{\mathcal{X}}^{\pi,t_i}_{t_{i+1}}} - \|g\|_{\mathcal{X}^L_{t_{i+1}}} \leqslant \Delta_i^{\frac{\beta+2}{2}}\right\}.$$

Concerning the original least-squares Monte Carlo approach, Lemor et al. (2006) used in their analysis of the approximation error rather similar sets $\mathcal{A}^M_{i+1}$. The only difference is that our sets $\mathcal{A}_{i+1}$ are based on a general partial interval Δ_i, whereas the sets $[\mathcal{A}^M_{i+1}]^c$ consider $h := \Delta_i = T/N$ for all $i = 0, \dots, N-1$.

Lemma 23. *Under the Assumption of Lemma 22 it holds true that for some* $C > 0$

$$\mathbb{P}\{[\mathcal{A}_{i+1}]^c\} \leqslant C\exp\left\{CK\log\frac{CR}{\Delta_i^{(\beta+2)/2}} - \frac{L\Delta_i^{\beta+2}}{72R^2}\right\}$$

for $i = 0, \dots, N-1$.

We omit the proof, because it works in exactly the same manner as the proof of Proposition 4 in Lemor et al. (2006), except that h is replaced by Δ_i. Now it remains to derive the L^2-error between $(\hat{y}_i^{\pi,K,R}(\cdot), \hat{z}_i^{\pi,K,R}(\cdot))_{t_i\in\pi}$ and $(\hat{y}_i^{\pi,K,R,L}(\cdot), \hat{z}_i^{\pi,K,R,L}(\cdot))_{t_i\in\pi}$ with respect to $X^\pi_{t_i}$ instead of $\{{}_\lambda X^\pi_{t_i}, \lambda = 1,\dots,L\}$ as done in Lemma 22. Recall,

$$\hat{Y}^{\pi,K,R}_{t_i} = \hat{y}_i^{\pi,K,R}(X^\pi_{t_i}), \qquad \hat{Z}^{\pi,K,R}_{t_i} = \hat{z}_i^{\pi,K,R}(X^\pi_{t_i}),$$
$$\hat{Y}^{\pi,K,R,L}_{t_i} = \hat{y}_i^{\pi,K,R,L}(X^\pi_{t_i}), \qquad \hat{Z}^{\pi,K,R,L}_{t_i} = \hat{z}_i^{\pi,K,R,L}(X^\pi_{t_i}).$$

Lemma 24. *Under the assumptions of Lemma 22 there is a constant $C > 0$ depending on κ, T, D, C_X and κ_R such that for $|\pi|$ small enough and $\beta \in (0,1]$*

$$\begin{aligned}
&\max_{0\leqslant i\leqslant N} E|\hat{Y}^{\pi,K,R}_{t_i} - \hat{Y}^{\pi,K,R,L}_{t_i}|^2 + \sum_{i=0}^{N-1}\Delta_i E|\hat{Z}^{\pi,K,R}_{t_i} - \hat{Z}^{\pi,K,R,L}_{t_i}|^2 \\
&\quad\leqslant CR^2NK\frac{\log L}{L} + C\inf_{\alpha\in\mathbb{R}^K} E|\phi^\pi(X^\pi_{t_N}) - \eta_0(N, X^\pi_{t_N})\alpha|^2 \\
&\quad\quad + C\Big(\max_{0\leqslant i\leqslant N} E|Y^\pi_{t_i} - \hat{Y}^{\pi,K}_{t_i}|^2 + \sum_{i=0}^{N-1}\Delta_i E|Z^\pi_{t_i} - \hat{Z}^{\pi,K}_{t_i}|^2\Big) \\
&\quad\quad + C\Big(\max_{0\leqslant i\leqslant N} E|\hat{Y}^{\pi,K}_{t_i} - \hat{Y}^{\pi,K,R}_{t_i}|^2 + \sum_{i=0}^{N-1}\Delta_i E|\hat{Z}^{\pi,K}_{t_i} - \hat{Z}^{\pi,K,R}_{t_i}|^2\Big) \\
&\quad\quad + C\sum_{i=0}^{N-1}\Delta_i E|\mathcal{P}_{0,j}\big(f(t_i, S^\pi_{t_i}, Y^\pi_{t_{i+1}}, Z^\pi_{t_i})\big) - E[f(t_i, S^\pi_{t_i}, Y^\pi_{t_{i+1}}, Z^\pi_{t_i})|\,X^\pi_{t_i}]|^2 \\
&\quad\quad + C|\pi|^\beta + CR^2\sum_{i=0}^{N-1}\frac{1}{\Delta_i}\exp\left\{CK\log\frac{CR}{\Delta_i^{(\beta+2)/2}} - \frac{L\Delta_i^{\beta+2}}{72R^2}\right\}.
\end{aligned}$$

By and large, the following proof matches that of Theorem II.3 in Lemor (2005), who adopted the line of argumentation of Theorem 11.3 in Györfi et al. (2002).

Proof. We denote by P_i^X the distribution of $X^\pi_{t_i}$. Additionally, we have for some measurable function g the norms

$$\|g\|_i = \sqrt{\int |g(x)|^2 dP_i^X(x)}, \quad \|g\|_{\mathcal{X}^L_{t_i}} = \sqrt{\frac{1}{L}\sum_{\lambda=1}^L |g({}_\lambda X^\pi_{t_i})|^2}.$$

Then,

$$\begin{aligned}
&E\|\hat{y}_i^{\pi,K,R} - \hat{y}_i^{\pi,K,R,L}\|_i^2 \\
&\quad= E\Big(\|\hat{y}_i^{\pi,K,R} - \hat{y}_i^{\pi,K,R,L}\|_i - 2\|\hat{y}_i^{\pi,K,R} - \hat{y}_i^{\pi,K,R,L}\|_{\mathcal{X}^L_{t_i}} \\
&\quad\quad + 2\|\hat{y}_i^{\pi,K,R} - \hat{y}_i^{\pi,K,R,L}\|_{\mathcal{X}^L_{t_i}}\Big)^2 \\
&\quad\leqslant E\Big(\max\big\{\|\hat{y}_i^{\pi,K,R} - \hat{y}_i^{\pi,K,R,L}\|_i - 2\|\hat{y}_i^{\pi,K,R} - \hat{y}_i^{\pi,K,R,L}\|_{\mathcal{X}^L_{t_i}}, 0\big\} \\
&\quad\quad + 2\|\hat{y}_i^{\pi,K,R} - \hat{y}_i^{\pi,K,R,L}\|_{\mathcal{X}^L_{t_i}}\Big)^2.
\end{aligned}$$

Making use of Young' s inequality gives

$$\begin{aligned}E\|\hat{y}_i^{\pi,K,R} - \hat{y}_i^{\pi,K,R,L}\|_i^2 &\leqslant 2E\Big(\max\big\{\|\hat{y}_i^{\pi,K,R} - \hat{y}_i^{\pi,K,R,L}\|_i - 2\|\hat{y}_i^{\pi,K,R} - \hat{y}_i^{\pi,K,R,L}\|_{\mathcal{X}^L_{t_i}}, 0\big\}\Big)^2 \\ &\quad + 8E\|\hat{y}_i^{\pi,K,R} - \hat{y}_i^{\pi,K,R,L}\|^2_{\mathcal{X}^L_{t_i}}.\end{aligned}$$

Similarly, we have

$$\begin{aligned}\Delta_i E\|\hat{z}_{d,i}^{\pi,K,R} - \hat{z}_{d,i}^{\pi,K,R,L}\|_i^2 &\leqslant 2E\Big(\max\big\{\|\sqrt{\Delta_i}(\hat{z}_{d,i}^{\pi,K,R} - \hat{z}_{d,i}^{\pi,K,R,L})\|_i - 2\|\sqrt{\Delta_i}(\hat{z}_{d,i}^{\pi,K,R} - \hat{z}_{d,i}^{\pi,K,R,L})\|_{\mathcal{X}^L_{t_i}}, 0\big\}\Big)^2 \\ &\quad + 8\Delta_i E\|\hat{z}_{d,i}^{\pi,K,R} - \hat{z}_{d,i}^{\pi,K,R,L}\|^2_{\mathcal{X}^L_{t_i}}.\end{aligned}$$

Due to Lemma 22 and Lemma 23, the upper bound for

$$\max_{0\leqslant i\leqslant N} E\|\hat{y}_i^{\pi,K,R} - \hat{y}_i^{\pi,K,R,L}\|^2_{\mathcal{X}^L_{t_i}} + \sum_{d=1}^{D}\sum_{i=0}^{N-1}\Delta_i E\|\hat{z}_{d,i}^{\pi,K,R} - \hat{z}_{d,i}^{\pi,K,R,L}\|^2_{\mathcal{X}^L_{t_i}}$$

is given by the right-hand side of (4.17) and it suffices to provide an estimate for

$$\begin{aligned}&E\Big(\max\big\{\|\hat{y}_i^{\pi,K,R} - \hat{y}_i^{\pi,K,R,L}\|_i - 2\|\hat{y}_i^{\pi,K,R} - \hat{y}_i^{\pi,K,R,L}\|_{\mathcal{X}^L_{t_i}}, 0\big\}\Big)^2, \\ &E\Big(\max\big\{\|\sqrt{\Delta_i}(\hat{z}_{d,i}^{\pi,K,R} - \hat{z}_{d,i}^{\pi,K,R,L})\|_i - 2\|\sqrt{\Delta_i}(\hat{z}_{d,i}^{\pi,K,R} - \hat{z}_{d,i}^{\pi,K,R,L})\|_{\mathcal{X}^L_{t_i}}, 0\big\}\Big)^2.\end{aligned}$$

for $d = 1, \dots, D$. We first take care for the Y-part and explain then, how the results can be transfered to the Z-part. Let a be some positive variable. It holds true that

$$\begin{aligned}P\Big\{\big(\max\big\{\|\hat{y}_i^{\pi,K,R} - \hat{y}_i^{\pi,K,R,L}\|_i - 2\|\hat{y}_i^{\pi,K,R} - \hat{y}_i^{\pi,K,R,L}\|_{\mathcal{X}^L_{t_i}}, 0\big\}\big)^2 > a\Big\} \\ < P\Big\{\exists g \in G_i \,\Big|\, \|g\|_i - 2\|g\|_{\mathcal{X}^L_{t_i}} > \sqrt{a}\Big\}.\end{aligned}$$

The application of Lemma 28, Appendix A, yields

$$\begin{aligned}&P\Big\{\big(\max\big\{\|\hat{y}_i^{\pi,K,R} - \hat{y}_i^{\pi,K,R,L}\|_i - 2\|\hat{y}_i^{\pi,K,R} - \hat{y}_i^{\pi,K,R,L}\|_{\mathcal{X}^L_{t_i}}, 0\big\}\big)^2 > a\Big\} \\ &\quad\leqslant 3\exp\Big\{-\frac{La}{288(2R)^2}\Big\} E\left[\mathcal{N}_2\Big(\frac{\sqrt{2a}}{24}, G_i, \mathcal{X}_i^{2L}\Big)\right],\end{aligned}$$

where $\mathcal{X}_i^{2L} = \{{}_\lambda X^\pi_{t_i} | \lambda = 1, \dots, 2L\}$ is a set of i.i.d. copies of $X^\pi_{t_i}$. For an explanation of $\mathcal{N}_2$, see Definition 27. Recalling

$$G_i = \Big\{[\eta_0(i,x)\alpha]_R - \hat{y}_i^{\pi,K,R}(x) | \alpha \in \mathbb{R}^K\Big\},$$

we can write by definition of $\mathcal{N}_2$ that

$$\mathcal{N}_2(\frac{\sqrt{2a}}{24}, G_i, \mathcal{X}_i^{2L}) = \mathcal{N}_2(\frac{\sqrt{2a}}{24}, [\eta_0(i,x)\alpha]_R, \mathcal{X}_i^{2L}) = \mathcal{N}_2(\frac{\sqrt{2a}}{24}, \tilde{G}_i, \mathcal{X}_i^{2L}),$$

where

$$\tilde{G}_i = \{[\eta_0(i,x)\alpha]_R + R | \alpha \in \mathbb{R}^K\}$$

is a set of positive functions bounded by 2R. Let $\tilde{G}_i^+ = \{\{(x,t) \in \mathbb{R}^{\widetilde{D}} \times \mathbb{R} | t \leqslant \tilde{g}(x)\}, \tilde{g} \in \tilde{G}_i\}$. By Lemma 29 we obtain for $0 < a < 72R^2$

$$\begin{aligned}\mathcal{N}_2(\frac{\sqrt{2a}}{24}, \tilde{G}_i, \mathcal{X}_i^{2L}) &\leqslant 3\left(\frac{2e(2R)^2 24^2}{2a}\log\frac{3e(2R)^2 24^2}{2a}\right)^{\mathcal{V}_{\tilde{G}_i^+}} \\ &\leqslant 3\left(\frac{\sqrt{6}e(2R)^2 24^2}{2a}\right)^{2\mathcal{V}_{\tilde{G}_i^+}} \leqslant 3\left(\frac{1152\sqrt{6}eR^2}{a}\right)^{2\mathcal{V}_{\tilde{G}_i^+}},\end{aligned}$$

where $\mathcal{V}_{\tilde{G}_i^+}$ is the Vapnik-Chervonenkis (VC) dimension of $\tilde{G}_i^+$. See Definition 26 for an explanation on this dimension and the related topic of shattering coefficients. It remains to show

$$\mathcal{V}_{\tilde{G}_i^+} \overset{(I)}{=} \mathcal{V}_{\{[\eta_0(i,x)\alpha]_R | \alpha\in\mathbb{R}^K\}^+} \overset{(II)}{\leqslant} \mathcal{V}_{\{\eta_0(i,x)\alpha | \alpha\in\mathbb{R}^K\}^+} \overset{(III)}{\leqslant} K+1. \tag{4.25}$$

Concerning (I) we assume that $\mathcal{V}_{\tilde{G}_i^+} = n$. Hence, there is a set

$$\tilde{\mathcal{A}} := \{(x_1,t_1),\dots,(x_n,t_n)\} \subset \mathbb{R}^{\widetilde{D}} \times \mathbb{R}$$

that is shattered by $\tilde{G}_i^+$. Namely, for an arbitrary subset $J \subseteq \{1,\dots,n\}$ there is a $\tilde{g} \in \tilde{G}_i$ such that

$$\begin{aligned}\tilde{g}(x_j) &= [\eta_0(i,x_j)\tilde{\alpha}]_R + R \geqslant t_j, \quad j \in J,\\ \tilde{g}(x_j) &= [\eta_0(i,x_j)\tilde{\alpha}]_R + R < t_j, \quad j \notin J.\end{aligned}$$

Considering the set

$$\mathcal{A} = \{(x_1,t_1-R),\dots,(x_n,t_n-R)\} \subset \mathbb{R}^{\widetilde{D}} \times \mathbb{R},$$

we can then pick out the points determined by the index set J by means of the function $[\eta_0(i,x)\tilde{\alpha}]_R$. As J was chosen arbitrary, we can deduce that $\{[\eta_0(i,x)\alpha]_R | \alpha \in \mathbb{R}^K\}^+$ shatters $\mathcal{A}$. Thus, $\mathcal{V}_{\tilde{G}_i^+} \leqslant \mathcal{V}_{\{[\eta_0(i,x)\alpha]_R | \alpha\in\mathbb{R}^K\}^+}$. The reverse direction can be proven in the same manner.

Turning to (II), we suppose again $\mathcal{V}_{\{[\eta_0(i,x)\alpha]_R | \alpha\in\mathbb{R}^K\}^+} = n$. Let $\tilde{\mathcal{A}}$ again be a subset of n points of $\mathbb{R}^{\widetilde{D}} \times \mathbb{R}$ that is shattered by $\{[\eta_0(i,x)\alpha]_R | \alpha \in \mathbb{R}^K\}^+$. Clearly, there is a $g(x)$ such that

$$\begin{aligned}g(x_j) &= [\eta_0(i,x_j)\tilde{\alpha}]_R \geqslant t_j, \quad j \in J,\\ g(x_j) &= [\eta_0(i,x_j)\tilde{\alpha}]_R < t_j, \quad j \notin J.\end{aligned}$$

We claim now, $\eta_0(i, x_j)\tilde{\alpha} \geqslant [\eta_0(i, x_j)\tilde{\alpha}]_R$ for $j \in J$ and $\eta_0(i, x_j)\tilde{\alpha} < [\eta_0(i, x_j)\tilde{\alpha}]_R$ for $j \notin J$. Suppose there is a $j^\star \in J$ with $\eta_0(i, x_{j^\star})\tilde{\alpha} < g(x_{j^\star})$. Consequently, by definition of $g(x_{j^\star})$ we have $\eta_0(i, x_{j^\star})\tilde{\alpha} < -R$ and $g(x_{j^\star}) = -R$. Then, $t_{j^\star} \leqslant -R$. Regarding the complement of $j^\star$ in $\{1, \dots, n\}$ there must be a $g^\star(x) \in \{[\eta_0(i, x)\alpha]_R | \, \alpha \in \mathbb{R}^K\}$ such that

$$\begin{aligned} g^\star(x_j) &= [\eta_0(i, x_j)\alpha^\star]_R \geqslant t_j, \quad j \neq j^\star, \\ g^\star(x_{j^\star}) &= [\eta_0(i, x_{j^\star})\alpha^\star]_R < t_{j^\star}. \end{aligned}$$

But $-R < t_{j^\star} \leqslant -R$ is a contradiction and we get the desired result $\eta_0(i, x_j)\tilde{\alpha} \geqslant [\eta_0(i, x_j)\tilde{\alpha}]_R$ for $j \in J$. The inequality $\eta_0(i, x_j)\tilde{\alpha} < [\eta_0(i, x_j)\tilde{\alpha}]_R$ for $j \notin J$ can be shown analogously. In sum, $\tilde{\mathcal{A}}$ is also shattered by $\{\eta_0(i, x)\alpha | \alpha \in \mathbb{R}^K\}^+$.

As far as (III) is concerned, we adopt the argument from page 152, Györfi et al. (2002). We have

$$\begin{aligned} \{\eta_0(i, x)\alpha | \, \alpha \in \mathbb{R}^K\}^+ &= \Big\{\{(x, t) | \, \eta_0(i, x)\alpha \geqslant t\}, \alpha \in \mathbb{R}^K\Big\} \\ &\subset \Big\{\{(x, t) | \, \eta_0(i, x)\alpha + b \cdot t \geqslant 0\}, \alpha \in \mathbb{R}^K, b \in \mathbb{R}\Big\} \end{aligned}$$

The vector space $\{\eta_0(i, x)\alpha + b \cdot t | \, \alpha \in \mathbb{R}^K, b \in \mathbb{R}\}$ is $K+1$-dimensional and by Lemma 30, the proof of (4.25) is complete. Now, we have the estimate

$$\begin{aligned} &P\Big\{\Big(\max\big\{\|\hat{y}_i^{\pi,K,R} - \hat{y}_i^{\pi,K,R,L}\|_i - 2\|\hat{y}_i^{\pi,K,R} - \hat{y}_i^{\pi,K,R,L}\|_{\mathcal{X}^L_{t_i}}, 0\big\}\Big)^2 > a\Big\} \\ &\quad < 9\left(\frac{1152\sqrt{6}eR^2}{a}\right)^{2(K+1)} \exp\left\{-\frac{La}{1152R^2}\right\} \\ &\quad < 9\left(\sqrt{6}eL\right)^{2(K+1)} \exp\left\{-\frac{La}{1152R^2}\right\}, \end{aligned}$$

for $a > 1152R^2/L$. This enables us to give an upper bound for the expectation of (I). Clearly,

$$\begin{aligned} &E\Big(\max\big\{\|\hat{y}_i^{\pi,K,R} - \hat{y}_i^{\pi,K,R,L}\|_i - 2\|\hat{y}_i^{\pi,K,R} - \hat{y}_i^{\pi,K,R,L}\|_{\mathcal{X}^L_{t_i}}, 0\big\}\Big)^2 \\ &\quad = \int_0^\infty P\Big\{\Big(\max\big\{\|\hat{y}_i^{\pi,K,R} - \hat{y}_i^{\pi,K,R,L}\|_i - 2\|\hat{y}_i^{\pi,K,R} - \hat{y}_i^{\pi,K,R,L}\|_{\mathcal{X}^L_{t_i}}, 0\big\}\Big)^2 > t\Big\} dt \\ &\quad \leqslant a + 9\left(\sqrt{6}eL\right)^{2(K+1)} \int_a^\infty \exp\left\{-\frac{Lt}{1152R^2}\right\} dt \\ &\quad \leqslant a + 9\left(\sqrt{6}eL\right)^{2(K+1)} \frac{1152R^2}{L} \exp\left\{-\frac{La}{1152R^2}\right\}. \end{aligned}$$

The last term can be minimized by choosing

$$a = \frac{1152R^2}{L} \log(9(\sqrt{6}eL)^{2(K+1)}).$$

Hence,

$$\begin{aligned} &E\Big(\max\big\{\|\hat{y}_i^{\pi,K,R} - \hat{y}_i^{\pi,K,R,L}\|_i - 2\|\hat{y}_i^{\pi,K,R} - \hat{y}_i^{\pi,K,R,L}\|_{\mathcal{X}^L_{t_i}}, 0\big\}\Big)^2 \\ &\quad \leqslant \frac{1152R^2}{L}\Big(\log(9) + 2(K+1)\log(\sqrt{6}eL) + 1\Big) \leqslant CR^2K\frac{\log L}{L}. \end{aligned}$$

Concerning the Z-part, we get the same upper bound for

$$E\Big(\max\{\|\sqrt{\Delta_i}(\hat{z}^{\pi,K,R}_{d,i}-\hat{z}^{\pi,K,R,L}_{d,i})\|_i-2\|\sqrt{\Delta_i}(\hat{z}^{\pi,K,R}_{d,i}-\hat{z}^{\pi,K,R,L}_{d,i})\|_{\mathcal{X}^L_{t_i}},0\}\Big)^2$$

by replacing G_i by $\{[\sqrt{\Delta_i}\eta_d(i,x)\alpha]_R-\sqrt{\Delta_i}\hat{z}^{\pi,K,R}_{d,i}|\,\alpha\in\mathbb{R}^K\}$. The functions of this set are also bounded by 2R. Therefore, the result follows by a straightforward repetition of the single steps of the proof for the Y-part. Then the proof is complete. □

4.3 The overall approximation error and its comparison with the original LSMC approach

Just like the original least-squares Monte Carlo approach, the approximation error of the simplified algorithm is determined by the errors that are caused by time discretization, projection, truncation and last but not least simulation.

However, the simplification has no impact on the squared **time discretization error**, that is

$$\sup_{0\leqslant t\leqslant T}E|Y_t-Y^\pi_t|^2+\int_0^T E|Z_t-Z^\pi_t|^2dt\leqslant C|\pi|+CE|\xi-\xi^\pi|^2,$$

see Subsection 2.2.1. The error term $E|\xi-\xi^\pi|^2$ decreases with rate $|\pi|^\beta$, for $\beta\in(0,1]$ for instance, if there is a Lipschitz-continuous function ϕ such that $\xi=\phi(S_T)$ and $\xi^\pi=\phi(S^\pi_{t_N})$ with $\max_{0\leqslant i\leqslant N}E|S_{t_i}-S^\pi_{t_i}|^2\leqslant|\pi|^\beta$. As for the remaining error sources, the combination of Lemmas 17, 19 and 24 yields the overall L^2-error between the time-discrete solution and the approximation generated by simplified least-squares Monte Carlo.

Theorem 25. *Let Assumption 2 and 7 be satisfied. Then there is a constant* $C>0$ *depending on* κ, T, D, C_X *and* κ_R *such that for* $|\pi|$ *small enough,* $\epsilon>1$ *and* $\beta\in(0,1]$

$$\begin{aligned}&\max_{0\leqslant i\leqslant N}E|Y^\pi_{t_i}-\hat{Y}^{\pi,K,R,L}_{t_i}|^2+\sum_{i=0}^{N-1}\Delta_i E|Z^\pi_{t_i}-\hat{Z}^{\pi,K,R,L}_{t_i}|^2\\&\leqslant CR^2NK\frac{\log L}{L}+C\inf_{\alpha\in\mathbb{R}^K}E|\phi^\pi(X^\pi_{t_N})-\eta_0(N,X^\pi_{t_N})\alpha|^2\\&\quad+C\frac{NK^{2\epsilon}}{R^{2(\epsilon-1)}}\max_{0\leqslant i\leqslant N}E|Y^\pi_{t_i}|^{2\epsilon}\\&\quad+C\sum_{i=0}^{N-1}\Delta_iE|\mathcal{P}_{0,j}\big(F(t_i,S^\pi_{t_i},Y^\pi_{t_{i+1}},Z^\pi_{t_i})\big)-E[F(t_i,S^\pi_{t_i},Y^\pi_{t_{i+1}},Z^\pi_{t_i})|X^\pi_{t_i}]|^2\\&\quad+C|\pi|^\beta+CR^2\sum_{i=0}^{N-1}\frac{1}{\Delta_i}\exp\left\{CK\log\frac{CR}{\Delta_i^{(\beta+2)/2}}-\frac{L\Delta_i^{\beta+2}}{72R^2}\right\}.\end{aligned}$$

Referring to Lemma 17 and the defintion of $\hat{Y}^{\pi,K,R,L}_{t_N}$ as a projection on the space spanned by $\eta_0(N, X^\pi_{t_N})$, the squared **projection error** is bounded by

$$C \inf_{\alpha \in \mathbb{R}^K} E|\phi^\pi(X^\pi_{t_N}) - \eta_0(N, X^\pi_{t_N})\alpha|^2 \\ + C \sum_{i=0}^{N-1} \Delta_i E|\mathcal{P}_{0,j}\big(F(t_i, S^\pi_{t_i}, Y^\pi_{t_{i+1}}, Z^\pi_{t_i})\big) - E[F(t_i, S^\pi_{t_i}, Y^\pi_{t_{i+1}}, Z^\pi_{t_i})| X^\pi_{t_i}]|^2.$$

The first error term stems from the projection error of the approximate terminal condition. It vanishes, if the conditional expectations of the approximate terminal condition are available in closed form, which means that it can be included in the system of martingale basis functions. Contrary to that, the squared projection error of the original least-squares Monte Carlo scheme was bounded by a constant times the sum of the L^2-errors regarding $(Y^\pi_{t_i}, Z^\pi_{t_i})_{t_i \in \pi}$ and their best projection. In other words, the original least-squares Monte Carlo scheme suffers from a propagation of the projection errors, that can be avoided in our proposal.

The additional error term

$$CR^{-2(\epsilon-1)} N K^{2\epsilon} \max_{0 \leqslant i \leqslant N} E|Y^\pi_{t_i}|^{2\epsilon}$$

arises from the squared **truncation error**. Due to $|Y^\pi_{t_i}| < C(1 + |X^\pi_{t_i}|)$, see Gobet et al. (2005), the term $E|Y^\pi_{t_i}|^{2\epsilon}$ is bounded under appropriate integrability conditions. Thus, the squared truncation error can be designed to converge with rate $|\pi|^\beta$ for R proportional to $N^{(1+\beta)/(2\epsilon-2)} K^{\epsilon/(\epsilon-1)}$. But, usually, this error term is simply neglected when it comes to conducting simulations.

The second important difference between original and simplified least-squares Monte Carlo lies in the additional terms caused by the squared **simulation error**. They sum up to

$$CR^2 NK \frac{\log L}{L} + CR^2 \sum_{i=0}^{N-1} \frac{1}{\Delta_i} \exp\left\{ CK \log \frac{CR}{\Delta_i^{(\beta+2)/2}} - \frac{L\Delta_i^{\beta+2}}{72R^2} \right\} + C|\pi|^\beta$$

These error terms are also contained in the squared simulation error of the original scheme, see Subsection 2.2.5. It is worth noting, that these terms require a much slower increase of the sample size L than the remaining terms in (2.20). Precisely, if the dimension K grows proportional to N^δ for some $\delta \geqslant 0$, then choosing L proportional to $N^{\beta+2+\delta} \log(N) R^2$ is sufficient for a convergence rate of $|\pi|^\beta$. In general, the log-term and the truncation constant are neglected, when determining the sample size. Hence, we have for L a growth rate of $\beta + 2 + \delta$ in the simplified scheme versus $\beta + 2 + 2\delta$ in the original least-squares Monte Carlo algorithm.

4.4 Numerical examples for non-linear European option pricing problems

Again we look at option pricing problems, where the price of the underlying stocks S is modeled by a geometric Brownian motion according to Black-Scholes, i. e.

$$S_{t,d} = s_{0,d} \exp\left\{\left(\mu - \sigma^2/2\right) t + \sigma W_{t,d}\right\}, \quad d = 1, \dots, D,$$

with μ, $\sigma > 0$ and $W = (W_1, \dots, W_D)$ being a D-dimensional Brownian motion. That means, for $D > 1$ we have options that are based on a basket of several stocks. As S can be sampled perfectly, we can simply set $S^\pi = S$. The pay-off function will be of type $\xi = \phi(S_T)$, that means we concentrate on non-path-dependent termincal conditions. Hence, the construction of a larger Markov process X^π, that includes S^π, becomes obsolete and we define $X^\pi = X = S$.

The assumption of a market with different interest rates for borrowing R and lending r with $R > r$ makes our problem a non-linear one. Following Bergman (1995), the option price for a possibly multidimensional underlying is described by the BSDE

$$Y_t = \phi(S_T) - \int_t^T \left(rY_u + \frac{\mu - r}{\sigma} \sum_{d=1}^{D} Z_{d,u}\right) du + (R - r) \int_t^T \left(Y_u - \frac{1}{\sigma} \sum_{d=1}^{D} Z_{d,u}\right)_+ du - \sum_{d=1}^{D} \int_t^T Z_{d,u} dW_{d,u}.$$

The following examples contain a call-spread option (either one-dimensional and multi-dimensional) and a straddle. In the latter case, we will try the Monte Carlo estimation of martingale basis functions. For a better distinction of the simulation results we write again π_N instead of π to indicate how many time steps the partition π has.

4.4.1 Call-spread option

The payoff-function is a composition of max-call options, clearly

$$\phi(S_T) = \left(\max_{d=1,\dots,D} S_{T,d} - \kappa_1\right)_+ - 2\left(\max_{d=1,\dots,D} S_{T,d} - \kappa_2\right)_+,$$

where κ_1 and κ_2 are the corresponding strike values. The market parameters are the same as in Subsection 3.4.2, thus

$$T = 0.25, \quad s_{d,0} = 100, \quad r = 0.01, \quad R = 0.06, \quad \mu = 0.05, \quad \sigma = 0.2.$$

for $d = 1, \dots, D$. The strike prices are again $\kappa_1 = 95$ and $\kappa_2 = 105$. Note, that the case $D = 1$ matches the example in Subsection 3.4.2.

Case 1: One-dimensional Brownian motion and indicator functions at terminal time

The first example considers $D = 1$. For the numerical solution we fix the basis functions at terminal time by

$$\eta_{0,1}(N, x) = (x - 95)_+ - (x - 105)_+,$$
$$\eta_{0,k}(N, x) = 3(K-1)\mathbb{1}_{\{x \in [a_{k-2}, a_{k-1})\}}, \quad k = 2, \dots, K$$

where K is the dimension of the function bases and $\{a_0, \dots, a_{K-1}\}$ a partition of the real line such that the probability of S_T ending up in $[a_{k-2}, a_{k-1})$ is the same for all $k = 2, \dots, K$. This kind of interval construction was also applied by Bouchard and Warin (2012) in the field of pricing American options with Monte Carlo methods. The function bases $\eta_0(i, x)$ and $\eta_1(i, x)$ are then generated by the martingale property for $i = 0, \dots, N-1$. The factor $3(K-1)$ prevents too small function values that might produce problems when computing the pseudo-inverse of $(\eta_0(i, {}_\lambda X_{t_i})_{\lambda=1,\dots,L}$ for $i < N$. In contrast to a pure indicator function basis, we are not able to quantify the projection error that arises in the present case. Like before, we fix the simulation parameters in dependence on $l = 3, \dots, 5$ and $m = 1, \dots, m(l)$. To be precise, $m(3) = 14, m(4) = 12, m(5) = 10$. Then, the number of time steps N, the dimension of the function bases K and the sample size L are given by

$$N = \left[2\sqrt{2}^{m-1}\right], \quad K = \left\lceil 3\sqrt{2}^{m-1} \right\rceil + 1, \quad L = \left[2\sqrt{2}^{l(m-1)}\right].$$

Concerning the simulation error, the cases $l = 3$ and $l = 4$ are the convergence thresholds in the simplified and the original least-squares Monte Carlo scheme, respectively. According to the theoretical results the L^2-error due to simulation decreases with rate 1/2 in the number of time steps for $l = 4$ in the simplified and $l = 5$ in the original approach. We denote by

$$\check{Y}^{\pi_N}_{t_i} = \eta_0(i, X_{t_i})\check{\alpha}^{\pi_N}_{0,i}, \quad \check{Z}^{\pi_N}_{t_i} = \eta_1(i, X_{t_i})\check{\alpha}^{\pi_N}_{1,i}.$$

the approximators of (Y, Z) generated by original least-squares Monte Carlo and by

$$\hat{Y}^{\pi_N}_{t_i} = \eta_0(i, X_{t_i})\hat{\alpha}^{\pi_N}_{0,i}, \quad \hat{Z}^{\pi_N}_{t_i} = \eta_1(i, X_{t_i})\hat{\alpha}^{\pi_N}_{1,i}$$

those, that result from the simplified approach. Again the global a-posteriori criteria $\mathcal{E}_{\pi_N}(\check{Y}^{\pi_N}, \check{Z}^{\pi_N})$ and $\mathcal{E}_{\pi_N}(\hat{Y}^{\pi_N}, \hat{Z}^{\pi_N})$ are in each case for l estimated by Monte Carlo simulation for which we incorporate 1000N samples of $X = S$. For a better view on the results we have separated them in two figures. The first one, Figure 4.1, shows the criterion for the original least-squares Monte Carlo scheme $\mathcal{E}_{\pi_N}(\check{Y}^{\pi_N}, \check{Z}^{\pi_N})$ for $l = 3, 4, 5$ and that for our enhanced proposal $\mathcal{E}_{\pi_N}(\hat{Y}^{\pi_N}, \hat{Z}^{\pi_N})$ for $l = 3$. As in the previous chapter, all figures will have logarithmic axes for a better view on convergence rates and details in the smaller range of values.

Concerning original least-squares Monte Carlo first, a comparison with the results in Subsection 3.4.2 gives information how the switch to a system of martingale basis

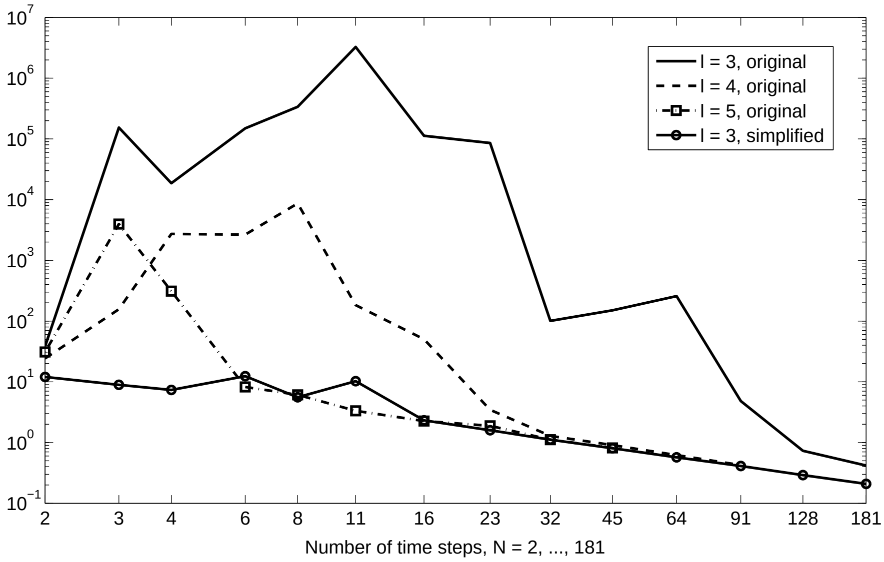

Figure 4.1: Case 1: Development of the global a-posteriori criterion for the original least-squares Monte Carlo approach in case of a one-dimensional call-spread

functions affects the projection error and thereby the over-all approximation error. Recall, that Subsection 3.4.2 differs from the present example only in the choice of bases, which there consisted of the pay-off function and indicator functions in all time steps $i = 0, \dots, N-1$.

Starting with the low-cost case $l = 3$, the global error criterion seems to be worsened by the chosen martingale basis functions. Not until the number of time steps takes values larger than 91, we can observe a trend tending to zero. Even for $l = 4$ the new basis functions deteriorate the results on the error criterion when looking at $N = 2, \dots, 23$. However, the numerics for larger numbers of time steps nearly coincide with the results in Subsection 3.4.2 as far as available. For $N > 23$ the case $l = 4$ decreases with rate -1.06. Looking at the case $l = 5$, the difference between the absolute values of the error criterion in Subsection 3.4.2 and that for martingale basis functions is negligible. Here, the empirical rate of convergence is -1.

It remains to mention the path in Figure 4.1 that corresponds to the global a-posteriori criterion when applying the simplified least-squares Monte Carlo approach for $l = 3$. We can see that for simulations with 16 or even more time steps the error criterion amounts roughly to the same absolute values as in case $l = 5$ when using original least-squares Monte Carlo. Taking a closer look at the numerics for 45 time steps, we observe an absolute value of 0.82 in the original scheme and 0.80 in the simplified algorithm. Particularly remarkable is here, that the first value was obtained by using $11,863,284$ samples, whereas the latter result gets along with $23,171$ samples only.

In Figure 4.2 we show the numerics for $\mathcal{E}_{\pi_N}(\hat{Y}^{\pi_N}, \hat{Z}^{\pi_N})$ for $l = 3, 4, 5$. Apparently, the results for larger numbers of time steps, precisely for $N > 16$, nearly coincide as far as calculated. For larger N all paths decrease with mean rate roughly about 0.96. This is insofar surprising as the theoretical results on the simulation error suggest that such a rate of convergence is attained not until $l = 4$. But we can also see, that the a-posteriori criterion does not benefit from larger sample sizes as used in the expensive case $l = 5$. This is also supported by the theoretical analysis.

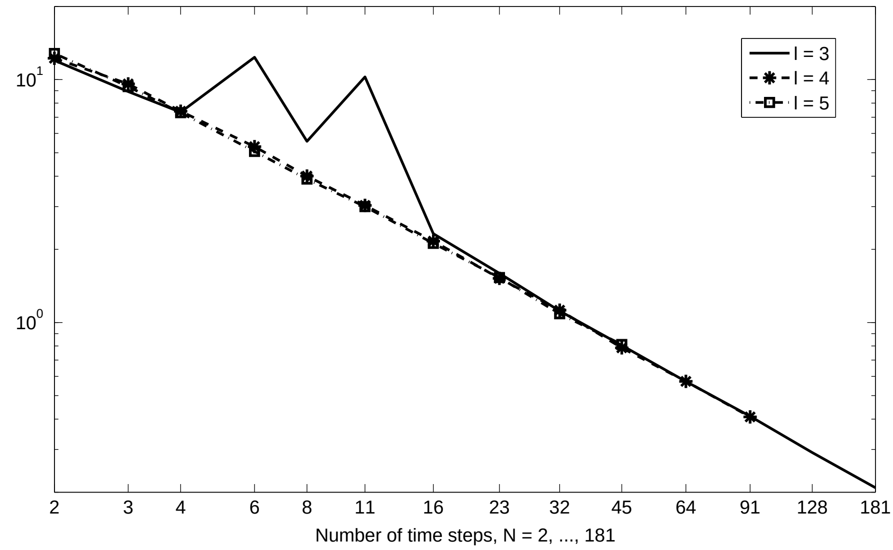

Figure 4.2: Case 1: Development of the a-posteriori criterion for the simplified least-squares Monte Carlo approach in case of a one-dimensional call-spread

The present example shows nicely how the computational cost can be reduced by enhanced least-squares Monte Carlo, when the dimension of the function bases grows with the number of time steps. The smaller effort can be exploited to simulate even finer partitions than possible in original least-squares Monte Carlo due to computational limitations. This has the effect that the approximation can be calculated for larger N than in the original proposal such that the corresponding error can be further reduced. Here, we finished the simulations at $N = 181$, where we achieved a global error criterion of 0.21 for $l = 3$. Recall, that the simulations for the call-spread in Subsection 3.4.2 stopped at $N = 45$ with a global error criterion of 0.86 in the expensive case $l = 5$.

Case 2: Three-dimensional Brownian motion and monomials at terminal time

This time we set $D = 3$ such that our basket includes 3 stocks. As basis functions at terminal time we pick

$$\eta_{0,1}(N, x) = (x - 95)_+ - (x - 105)_+, \quad \eta_{0,2}(N, x) = 1,$$
$$\eta_{0,3}(N, x) = x_1, \quad \eta_{0,4}(N, x) = x_2, \quad \eta_{0,5}(N, x) = x_3$$

and determine $\eta_{d,k}(i,x)$ by the martingale property explained in Assumption 6. The simulation parameter are defined by

$$N = \left[2\sqrt{2}^{m-1}\right], \quad K = 5, \quad L = \left[2\sqrt{2}^{3(m-1)}\right],$$

for $m = 1, \dots, 11$. We try three types of numerical solution. The first one exploits original least-squares Monte Carlo with basis functions

$$\tilde{\eta}_{d,k}(i,x) = \tilde{\eta}_{0,k}(i,x) = \eta_{0,k}(N,x), \quad d = 1, \dots, 3, \quad k = 1, \dots, 5$$

for all $i = 0, \dots, N-1$. This generates approximators

$$\tilde{Y}^{\pi_N}_{t_i} = \tilde{\eta}_0(i, X_{t_i})\tilde{\alpha}^{\pi_N}_{0,i}, \quad \tilde{Z}^{\pi_N}_{d,t_i} = \tilde{\eta}_d(i, X_{t_i})\tilde{\alpha}^{\pi_N}_{d,i}, \quad d = 1, \dots, 3.$$

The second simulation combines original least-squares Monte Carlo with the system of martingale function bases $\eta_d(i,x)$, $d = 0, \dots, 3$, $i = 0, \dots, N$ and yields

$$\check{Y}^{\pi_N}_{t_i} = \eta_0(i, X_{t_i})\check{\alpha}^{\pi_N}_{0,i}, \quad \check{Z}^{\pi_N}_{d,t_i} = \eta_d(i, X_{t_i})\check{\alpha}^{\pi_N}_{d,i}, \quad d = 1, \dots, 3.$$

The third attempt uses simplified least-squares Monte Carlo with martingale function bases and we receive

$$\hat{Y}^{\pi_N}_{t_i} = \eta_0(i, X_{t_i})\hat{\alpha}^{\pi_N}_{0,i}, \quad \hat{Z}^{\pi_N}_{d,t_i} = \eta_d(i, X_{t_i})\hat{\alpha}^{\pi_N}_{d,i}, \quad d = 1, \dots, 3.$$

Concerning both algorithms, original as well as simplified least-squares Monte Carlo, this parameter choice leads to a simulation error that decreases with rate $|\pi_N|^{1/2}$. The following figure compares the global a-posteriori criterion of all three approaches. Note, that the approximation of $(\tilde{Y}^{\pi_N}_{t_i}, \tilde{Z}^{\pi_N}_{t_i})$ and $(\check{Y}^{\pi_N}_{t_i}, \check{Z}^{\pi_N}_{t_i})$ varies only in the choice of basis functions. Apparently, the projection error connected with $(\check{Y}^{\pi_N}_{t_i}, \check{Z}^{\pi_N}_{t_i})$ is far smaller than that caused by $(\tilde{Y}^{\pi_N}_{t_i}, \tilde{Z}^{\pi_N}_{t_i})$ due to the choice of basis functions. As expected, the functions $\eta_d(i,x)$, $d = 0, \dots, 3$, $i = 0, \dots, N$ are much more suitable as projection bases thanks to their martingale property. Moreover, the error criterion $\mathcal{E}_{\pi_N}(\tilde{Y}^{\pi_N}, \tilde{Z}^{\pi_N})$ seems to tend to a constant value of about 13.70. Hence, the projection error superposes the effects from the time discretization error and the simulation error, which both decrease with rate $|\pi_N|^{1/2}$ in this setting.

Contrary to that, the absolute value of $\mathcal{E}_{\pi_N}(\check{Y}^{\pi_N}, \check{Z}^{\pi_N})$ amounts to 0.90 at $N = 64$. Looking at the entire path gives the impression that the convergence rate of $\mathcal{E}_{\pi_N}(\check{Y}^{\pi_N}, \check{Z}^{\pi_N})$ gets closer to that of $\mathcal{E}_{\pi_N}(\hat{Y}^{\pi_N}, \hat{Z}^{\pi_N})$, where we tried simplified least-squares Monte Carlo. Indeed, the path that represents the empirical error criterion $\mathcal{E}_{\pi_N}(\hat{Y}^{\pi_N}, \hat{Z}^{\pi_N})$ for $m = 1, \dots, 11$ tends to zero with rate -0.88 and ends up at $N = 64$ with an absolute value of 0.77.

A possible reasons for the difference between the error criteria for the approximations $(\check{Y}^{\pi_N}_{t_i}, \check{Z}^{\pi_N}_{t_i})$ and $(\hat{Y}^{\pi_N}_{t_i}, \hat{Z}^{\pi_N}_{t_i})$ is the following: The squared projection error in the latter case does not sum up, see Lemma 5, but is an average over time of the L^2-error between $E[F(t_i, S^{\pi}_{t_i}, Y^{\pi}_{t_{i+1}}, Z^{\pi}_{t_i}) | X^{\pi}_{t_i}]$ and its best projection on the function bases, see also Lemma 17.

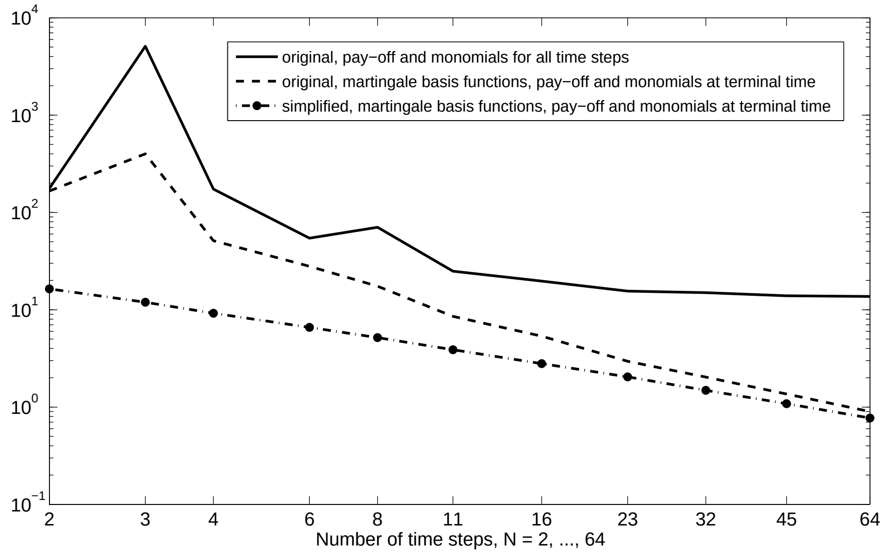

Figure 4.3: Case 2: Development of the global a-posteriori criterion in case of a three-dimensional call-spread

4.4.2 Pricing of a straddle - Simulation with estimated martingales

In the previous case we exploited the possibility to compute the conditional expectation of the basis functions in closed form. Several examples for such functions were already introduced in Example 15. The last numerical setting will pick up the question what to do if this possibility is not available. Let $\eta_{0,k}(N,x)$, $k = 1,\dots,K$ be a function basis at terminal time. When carrying out enhanced least-squares Monte Carlo estimation, we have to compute

$$\begin{aligned}\eta_{0,k}(i,{}_\lambda X_{t_i}) &= E\left[\eta_{0,k}(N,X_{t_N})\,\middle|\,X_{t_i} = {}_\lambda X_{t_i}\right],\\ \eta_{d,k}(i,{}_\lambda X_{t_i}) &= E\left[\Delta_\lambda W_{d,i}\eta_{0,k}(N,X_{t_N})\,\middle|\,X_{t_i} = {}_\lambda X_{t_i}\right], \quad d = 1,\dots,D\end{aligned}$$

for $\lambda = 1,\dots,L$, $k = 1,\dots,K$ and $i = 0,\dots,N-1$. In case this is not computable in closed form, we estimate these conditional expectations by Monte Carlo simulation. To this end, we generate for $\lambda = 1,\dots,L$ a set of $M_{i,\lambda}$ copies of $\{(\Delta_\lambda W_i, {}_\lambda X_{t_{i+1}})|j = i,\dots,N-1\}$, called

$$\mathcal{X}^{t_i,\lambda} := \{(\Delta_\mu W_j^{t_i,\lambda}, {}_\mu X_{t_{j+1}}^{t_i,\lambda})|\, j = i,\dots,N-1, \mu = 1,\dots,M_{i,\lambda}\}.$$

Here, the upper index (t_i,λ) signals, that the Markov process $(X_{t_j}^{t_i,\lambda})_{i\leqslant j\leqslant N}$ starts at time t_i in ${}_\lambda X_{t_i}$. Then we define

$$\check{\eta}_{0,k}(i,{}_\lambda X_{t_i}) = \frac{1}{M_{i,\lambda}}\sum_{\mu=1}^{M_{i,\lambda}} \eta_{0,k}(N,{}_\mu X_{t_N}^{t_i,\lambda}).$$

For the estimation of $\eta_{d,k}(i, {}_\lambda\tilde{x})$, $d = 1, \ldots, D$ we use the identity

$$
\begin{aligned}
&E\left[\Delta W_{d,i}\eta_{0,k}(N, X_{t_N}) \,|\, X_{t_i} = x\right] \\
&\qquad = E\left[\Delta W_{d,i}\left(\eta_{0,k}(N, X_{t_N}) - E\left[\eta_{0,k}(N, X_{t_N}) \,|\, X_{t_i} = x\right]\right) | \, X_{t_i} = x\right]
\end{aligned}
$$

in order to improve Monte Carlo simulation by variance reduction. With an independent copy

$$
\tilde{\mathcal{X}}^{t_i,\lambda} := \{(\Delta_\mu \tilde{W}_j^{t_i,\lambda}, {}_\mu X_{t_{j+1}}^{t_i,\lambda}) | \, j = i, \ldots, N-1, \mu = 1, \ldots, M_{i,\lambda}\}
$$

of $\mathcal{X}^{t_i,\lambda}$ we set

$$
\check{\eta}_{d,k}(i, {}_\lambda X_{t_i}) = \frac{1}{M_{i,\lambda}} \sum_{\mu=1}^{M_{i,\lambda}} \Delta_\mu \tilde{W}_{d,i}^{t_i,\lambda} \left(\eta_{0,k}(N, {}_\mu \tilde{X}_{t_N}^{t_i,\lambda}) - \check{\eta}_{0,k}(i, {}_\lambda X_{t_i})\right),
$$

for $d = 1, \ldots, D$. Now, we have for a fixed sample ${}_\lambda X_{t_i}$ of X_{t_i} at least estimations for the function values $\eta_{0,k}(i, {}_\lambda X_{t_i})$ and $\eta_{d,k}(i, {}_\lambda X_{t_i})$, even if the martingales $(\eta_{0,k}(i, X_{t_i}))_{0 \leqslant i \leqslant N}$ and the processes $(\eta_{d,k}(i, X_{t_i}))_{0 \leqslant i \leqslant N}$, $d = 1, \ldots, D$, for $k = 1, \ldots, K$ are not available in closed form. With this workaround simplified least-squares Monte Carlo becomes possible. Even though a theoretical analysis of the impact of this idea on the approximation has yet to be worked out, the following numerical example will show that this approach is quite promising.

Once again we are concerned with the pricing and hedging of a European option with dimension $D = 1$, see the introductory explanations of the current section. The pay-off function is this time defined by

$$
\phi(S_T) = |S_T - \kappa_1|.
$$

The parameters of the stock are determined by

$$
T = 0.5, \quad s_{0,1} = 100, \quad r = 0.01, \quad R = 0.01, \quad \mu = 0.05, \quad \sigma = 0.2.
$$

The strike price is fixed by $\kappa_1 = 110$ and the function basis $\eta_0(N, x)$ at terminal time is formed by

$$
\eta_{0,1}(N, x) = |x - \kappa_1|, \quad \eta_{0,2}(N, x) = \mathbb{1}, \quad \eta_{0,3}(N, x) = x, \quad \eta_{0,4} = x^2.
$$

By the martingale property we receive $\eta_0(i, x)$ and $\eta_1(i, x)$ for $i = 0, \ldots, N-1$, see Assumption 6. It remains to define the simulation parameter. Clearly,

$$
N = \left[2\sqrt{2}^{m-1}\right], \quad K = 4, \quad L = \left[2\sqrt{2}^{3(m-1)}\right],
$$

for $m = 1, \ldots, 15$. With these preliminaries we carry out three different numerical approaches. We apply original least-squares Monte Carlo with

$$
\tilde{\eta}_1(i, x) = \tilde{\eta}_0(i, x) = \eta_0(N, x)
$$

for all $i = 0, \dots, N-1$. Then we receive

$$\tilde{Y}^{\pi_N}_{t_i} = \tilde{\eta}_0(i, X_{t_i})\tilde{\alpha}^{\pi_N}_{0,i}, \quad \tilde{Z}^{\pi_N}_{t_i} = \tilde{\eta}_0(i, X_{t_i})\tilde{\alpha}^{\pi_N}_{1,i}.$$

The second approximation of (Y, Z) uses simplified least-squares Monte Carlo with the above defined function bases $\eta_d(i, x)$, $d = 0, 1$ and $i = 0, \dots, N$. This gives the approximators

$$\hat{Y}^{\pi_N}_{t_i} = \eta_0(i, X_{t_i})\hat{\alpha}^{\pi_N}_{0,i}, \quad \hat{Z}^{\pi_N}_{t_i} = \eta_1(i, X_{t_i})\hat{\alpha}^{\pi_N}_{1,i}.$$

The last numerical solution arises from the combination of simplified least-squares Monte Carlo with estimated function values $\check{\eta}_{0,k}(i, {}_\lambda X_{t_i})$ and $\check{\eta}_{d,k}(i, {}_\lambda X_{t_i})$, $\lambda = 1, \dots, L$, $d = 1, \dots, D$, $k = 1, \dots, 4$, that were computed by an 'inner' Monte Carlo simulation as explained above. The amount $M_{i,\lambda}$ of inner samples, that are used for this Monte Carlo simulation, is set to $200(N - i)$ independent of λ. Then we define

$$\check{Y}^{\pi_N}_{t_i} = \check{\eta}_0(i, X_{t_i})\check{\alpha}^{\pi_N}_{0,i}, \quad \check{Z}^{\pi_N}_{t_i} = \check{\eta}_1(i, X_{t_i})\check{\alpha}^{\pi_N}_{1,i}.$$

The empirical global a-posteriori criteria for all three attempts are shown in the following figure. Each of the three paths refers to one of the different numerical approaches. Not surprisingly, the empirical error criterion for $(\tilde{Y}^{\pi_N}_{t_i}, \tilde{Z}^{\pi_N}_{t_i})_{t_i \in \pi_N}$ does not tend to zero but levels out at 9.15 for 256 time steps.

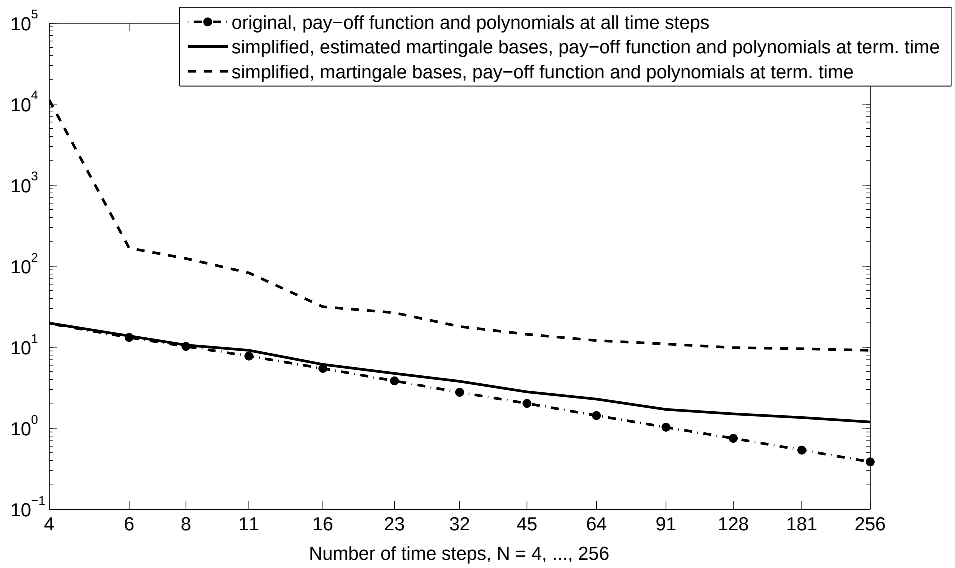

Figure 4.4: Development of the global a-posteriori criterion in case of a straddle

In contrast to that the a-posteriori criterion $\mathcal{E}_{\pi_N}(\hat{Y}^{\pi_N}, \hat{Z}^{\pi_N})$ has a empirical convergence rate of -0.94 and we obtain at $N = 256$ the absolute value of 0.39. These results are our benchmark for judging the approximation of $(\check{Y}^{\pi_N}_{t_i}, \check{Z}^{\pi_N}_{t_i})_{t_i \in \pi_N}$. For this approach with approximate martingale basis functions we observe that the error criterion $\mathcal{E}_{\pi_N}(\check{Y}^{\pi_N}, \check{Z}^{\pi_N})$ runs on a higher level compared to the results for $\mathcal{E}_{\pi_N}(\hat{Y}^{\pi_N}, \hat{Z}^{\pi_N})$ and gets down to an absolute value of 1.19 at $N = 256$. The distance between both criteria stays nearly constant and amounts to 0.78 on average. Although the empirical criterion $\mathcal{E}_{\pi_N}(\check{Y}^{\pi_N}, \check{Z}^{\pi_N})$ decreases with significantly smaller rate than $\mathcal{E}_{\pi_N}(\hat{Y}^{\pi_N}, \hat{Z}^{\pi_N})$, we can observe a significant improvement in contrast to the results for $\mathcal{E}_{\pi_N}(\tilde{Y}^{\pi_N}, \tilde{Z}^{\pi_N})$. The empirical results for $\mathcal{E}_{\pi_N}(\check{Y}^{\pi_N}, \check{Z}^{\pi_N})$ can be further improved by using a larger size of inner samples $M_{i,\lambda}$ for the computation of $\check{\eta}_{0,k}(i, {}_\lambda X_{t_i})$ and $\check{\eta}_{d,k}(i, {}_\lambda X_{t_i})$.

By and large, the combination of simplified least-squares Monte Carlo with approximate martingales seems to be a good alternative to original least-squares Monte Carlo if no appropriate system of martingale basis functions is available in closed form, even if it is more expensive to implement due to the simulation of inner samples.

A Some results on nonparametric regression and VC dimension

For the sake of convenience, we list here some results on nonparametric regression, that are required for the proof of Lemma 24. Precisely, we start by citing the contents of Definition 9.5 and 9.6 in Györfi et al. (2002).

Definition 26. Let $\mathcal{A}$ be a class of subsets of $\mathbb{R}^{\tilde{D}}$ and let $n \in \mathbb{N}$.
(i) For $x_1, \dots, x_n \in \mathbb{R}^{\tilde{D}}$ define

$$s(\mathcal{A}, \{x_1, \dots, x_n\}) = \#\big\{\{A \cap \{x_1, \dots, x_n\} | A \in \mathcal{A}\}\big\},$$

that is, $s(\mathcal{A}, \{x_1, \dots, x_n\})$ is the number of different subsets of $\{x_1, \dots, x_n\}$ of the form $A \cap \{x_1, \dots, x_n\}$ for $A \in \mathcal{A}$.
(ii) Let B be a subset of $\mathbb{R}^{\tilde{D}}$ of size n. One says that $\mathcal{A}$ shatters B if $s(\mathcal{A}, B) = 2^n$, i. e. if each subset of B can be represented in the form $A \cap B$ for some $A \in \mathcal{A}$.
(iii) The nth shatter coefficient of $\mathcal{A}$ is

$$\mathcal{S}(\mathcal{A}, n) = \max_{\{x_1,\dots,x_n\} \subseteq \mathbb{R}^{\tilde{D}}} s(\mathcal{A}, \{x_1, \dots, x_n\}).$$

That is the shatter coefficient is the maximal number of different subsets of n points that can be picked out by sets from $\mathcal{A}$.
(iv) Let $\mathcal{A} \neq \emptyset$. The VC dimension (or Vapnik-Chervonenkis dimension) $\mathcal{V}_{\mathcal{A}}$ of $\mathcal{A}$ is defined by

$$\mathcal{V}_{\mathcal{A}} = \sup\{n \in \mathbb{N} | \mathcal{S}(\mathcal{A}, n) = 2^n\},$$

i. e. the VC dimension $\mathcal{V}_{\mathcal{A}}$ is the largest integer n such that there exists a set of n points in $\mathbb{R}^{\tilde{D}}$ which can be shattered by $\mathcal{A}$.

Now, we introduce for a set $\mathcal{U}$ of functions $u : \mathbb{R}^{\tilde{D}} \to \mathbb{R}$ the norms

$$\|u\| = \sqrt{\int |u(x)|^2 dP^X(x)}, \quad \|u\|_L = \sqrt{\frac{1}{L} \sum_{\lambda=1}^{L} |u(X_\lambda)|^2},$$

where P^X is the law of a random variable X and $X^L := \{X_\lambda | \lambda = 1, \dots, L\}$ a set of independent copies of X.

The following definitions of covers and covering numbers are taken from Definition 9.3 in Györfi et al. (2002).

Definition 27. Let $\epsilon > 0$.
(i) An $L_2 - \epsilon$-cover of $\mathcal{U}$ on X^L is a finite set of functions $u_1, \dots, u_n : \mathbb{R}^{\tilde{D}} \to \mathbb{R}$ such that for every $u \in \mathcal{U}$ there is a $j \in \{1, \dots, n\}$ with

$$\left\|u - u_j\right\|_L < \epsilon.$$

(ii) The ϵ-covering number $\mathcal{N}_2(\epsilon, \mathcal{U}, X^L)$ of $\mathcal{U}$ with respect to $\|u\|_L$ is the smallest number n such that an $L_2 - \epsilon$-cover of $\mathcal{U}$ on X^L exists. Note that, as X^L is a random set, the covering number $\mathcal{N}_2(\epsilon, \mathcal{U}, X^L)$ is also a random variable.

By Theorem 11.2 of Györfi et al. (2002) we have

Lemma 28. *Let $\mathcal{U}$ be a class of functions $u : \mathbb{R}^{\tilde{D}} \to \mathbb{R}$ that is bounded in absolute value by R. Given $\epsilon > 0$ we have*

$$P\{\exists u \in \mathcal{U} : \|u\| - 2\|u\|_L > \epsilon\} \leqslant 3 \exp\{-\frac{L\epsilon^2}{288R^2}\} E\left[\mathcal{N}_2(\frac{\sqrt{2}}{24}\epsilon, \mathcal{U}, X^{2L})\right],$$

where $X^{2L} = \{X_1, \dots, X_L, X_{L+1}, \dots, X_{2L}\}$ is as set of i.i.d. copies of X.

Combining Lemma 9.2 and Theorem 9.4 of Györfi et al. (2002), we receive

Lemma 29. *Let $\mathcal{U}$ be a class of functions $u : \mathbb{R}^{\tilde{D}} \to [0, R]$ and*

$$\mathcal{U}^+ := \left\{\{(x, t) \in \mathbb{R}^{\tilde{D}} \times \mathbb{R} | t \leqslant u(x)\}, u \in \mathcal{U}\right\}$$

with $\mathcal{V}_{\mathcal{U}^+} \geqslant 2$ and let $0 < \epsilon < R/4$. Then

$$\mathcal{N}_2(\epsilon, \mathcal{U}, X^L) \leqslant 3\left(\frac{2eR^2}{\epsilon^2} \log \frac{3eR^2}{\epsilon^2}\right)^{\mathcal{V}_{\mathcal{U}^+}}.$$

Furthermore, we quote a result on the VC dimension of linear vector spaces, that can be found in Theorem 9.5 of Györfi et al. (2002).

Lemma 30. *Let $\mathcal{U}$ be a K-dimensional vector space of real-valued functions on $\mathbb{R}^{\tilde{D}}$, and set*

$$\mathcal{A} = \left\{\{x | u(x) \geqslant 0\}, u \in \mathcal{U}\right\}.$$

Then $\mathcal{V}_{\mathcal{A}} \leqslant K$.

Acknowledgements

I would like to express my gratitude to my advisor Prof. Dr. Christian Bender, who not only introduced me to this interesting research field, but also gave me the opportunity to prepare this thesis and offered me a position in his group. I very much appreciated that he shared his experience and his broad knowledge in stochastic analysis with me and always had time for discussions.

I am also very much obliged to Prof. Bruno Bouchard who agreed to act as co-referee for this thesis.

This research was financially supported by Deutsche Forschungsgemeinschaft under grant BE3933/3-1.

Bibliography

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

Bally, V.: Approximation scheme for solutions of BSDE. In: El Karoui, N., Mazliak, L. (eds.) Backward Stochastic Differential Equations, pp. 177–191. Longman (1997).

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., Moseler, T.: Importance sampling for backward SDEs. Stoch. Analysis Appl. **28**, 226–253 (2010).

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

Bergman, Y. Z.: Option pricing with differential interest rates. Rev. Financ. Stud. **8**, 475–500 (1995).

Bouchard, B., Elie, R.: Discrete-time approximation of decoupled forward-backward SDE with jumps. Stochastic Process. Appl. **118**, 53–75 (2008).

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

Bouchard, B., Ekeland, I., Touzi, N.: On the Malliavin approach to Monte Carlo approximation of conditional expectations. Finance Stoch. **8**, 45–71 (2004).

Bouchard, B., Warin, X.: Monte-Carlo Valuation of American options: Facts and new algorithms to improve existing methods. In: Carmona, R. A. et al. (eds.) Numerical methods in finance, pp. 215–256. Springer, Berlin (2012).

Chevance, D.: Numerical methods for backward stochastic differential equations. In: Numerical methods in finance, pp. 232–244. Publ. Newton Inst., Cambridge Univ. Press, Cambridge (1997).

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

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

El Karoui, N., Mazliak, L.: Backward stochastic differential equations. Pitman research notes in mathematical series. Longman Harlow (1997).

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

El Karoui, N., Pardoux, E., Quenez, M. C.: Reflected backward SDEs and American options, Numerical Methods in Finance, pp. 215-231. Publ. Newton Inst., Cambridge University Press, Cambridge (1997a).

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).

Glasserman, P., Yu, B.: Simulation for American options: Regression now or regression later? In: H. Niederreiter (ed.), Monte Carlo and Quasi-Monte Carlo Methods 2002, pp. 213–226. Springer, Berlin (2004).

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

Gobet, E., Labart, C.: Solving BSDE with adaptive control variate. SIAM J. Numer. Anal. **48**, 257–277 (2010).

Gobet, E., Lemor, J.-P., Warin, X.: A regression-based Monte Carlo method to solve backward stochastic differential equations. Ann. Appl. Probab. **15**, 2172–2202 (2005).

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

Györfi, L., Kohler, M., Krzyżak, A., Walk, H.: A distribution-free theory of nonparametric regression. Springer, Berlin (2002).

Imkeller, P., Dos Reis, G., Zhang, J.: Results on numerics for FBSDE with drivers of quadratic growth. In: Contemporary Quantitative Finance (Essays in Honour of Eckhard Platen), pp. 159–182. Springer, Berlin (2010).

Johnson, H.: Options on the maximum or the minimum of several assets. J. Fin. Quant. Analysis **22**, 277–283 (1987).

Kobylanski, M.: Backward stochastic differential equations and partial differential equations with quadratic growth. Ann. Probab. **28** , 558–602 (2000).

Lepeltier, J. P., San Martín, J.: Backward stochastic differential equations with continuous coefficient. Statist. Probab. Lett. **32**, 425–430 (1997).

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).

Lemor, J.-P.: Approximation par projections et simulations de Monte-Carlo des équations différentielles stochatiques rétrogades. Ph.D. dissertation, Ecole Polytechnique (2005).

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

Ma, J., Yong, J.: Forward-backward differential equations and their applications. Springer (1999).

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 quasilinear 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).

Richou, A.: Numerical simulation of BSDEs with drivers of quadratic growth. Ann. Appl. Probab. **21**, 1933–1964 (2011).

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

Zhang, J.: Some fine properties of backward stochastic differential equations, Ph.D. dissertation, Purdue Univ. (2001).

List of Figures

List of Tables