# CARF Working Paper

CARF-F-398

# Solving Backward Stochastic Differential Equations by Connecting the Short-term Expansions

Masaaki Fujii, Akihiko Takahashi
Quantitative Finance Course, Graduate School of Economics,
The University of Tokyo

First version: 14 June, 2016
Second version: 29 June, 2016
This version: 5 November, 2016

CARF is presently supported by Bank of Tokyo-Mitsubishi UFJ, Ltd., Dai-ichi Mutual Life Insurance Company, Nomura Holdings, Inc. and Sumitomo Mitsui Banking Corporation (in alphabetical order). This financial support enables us to issue CARF Working Papers.

CARF Working Papers can be downloaded without charge from:
http://www.carf.e.u-tokyo.ac.jp/workingpaper/index.html

Working Papers are a series of manuscripts in their draft form. They are not intended for circulation or distribution except as indicated by the author. For that reason Working Papers may not be reproduced or distributed without the written consent of the author.

# Solving Backward Stochastic Differential Equations by Connecting the Short-term Expansions *

Masaaki Fujii† & Akihiko Takahashi‡

November 5, 2016

**Abstract**

This article proposes a new approximation scheme for quadratic-growth BSDEs in a Markovian setting by connecting a series of semi-analytic asymptotic expansions applied to short-time intervals. Although there remains a condition which needs to be checked a posteriori, one can avoid altogether time-consuming Monte Carlo simulation and other numerical integrations for estimating conditional expectations at each space-time node. Numerical examples of quadratic-growth as well as Lipschitz BSDEs suggest that the scheme works well even for large quadratic coefficients, and a fortiori for large Lipschitz constants.

**Keywords :** asymptotic expansion, discretization, quadratic-growth BSDEs

## 1 Introduction

The research on backward stochastic differential equations (BSDEs) was initiated by Bismut (1973) [8] for a linear case and followed by Pardoux & Peng (1990) [42] for general non-linear setups. Since then, BSDEs have attracted strong interests among researchers and now large amount of literature is available. See for example, El Karoui et al. (1997) [25], El Karoui & Mazliak (eds.) (1997) [24], Ma & Yong (2000) [38], Yong & Zhou (1999) [45], Cvitanić & Zhang (2013) [20] and Delong (2013) [21] for excellent reviews and various applications, and also Pardoux & Rascanu (2014) [43] for a recent thorough textbook for BSDEs in the diffusion setup. In particular, since the financial crisis in 2009, the importance of BSDEs in the financial industry has grown significantly. This is because that the BSDEs have found to be indispensable to understand various valuation adjustments collectively referred to XVAs as well as the optimal risk control under the new regulations. For market developments and related issues, see Brigo, Morini & Pallavicini (2013) [14], Bianchetti & Morini (eds.) (2013) [7] and Crépey & Bielecki (2014) [17].

In the past decade, there has been also significant progress of numerical computation methods for BSDEs. In particular, based on the so-called $\mathbb{L}^2$-regularity of the control variables established by Zhang (2001, 2004) [51, 50], now standard backward Monte Carlo schemes for Lipschitz BSDEs have been developed by Bouchard & Touzi (2004) [11], Gobet, Lemor

*All the contents expressed in this research are solely those of the author and do not represent any views or opinions of any institutions. The authors are not responsible or liable in any manner for any losses and/or damages caused by the use of any contents in this research.

†Quantitative Finance Course, Graduate School of Economics, The University of Tokyo.

‡Quantitative Finance Course, Graduate School of Economics, The University of Tokyo.

& Warin (2005) [31]. One can find many variants and extensions such as Bouchard & Elie (2008) [10] for BSDEs with jumps, Bouchard & Chassagneux (2008) [9] for reflected BSDEs, and Chassagneux & Richou (2016) [16] for a system of reflected BSDEs arising from optimal switching problems. Bender & Denk (2007) [4] proposed a forward scheme free from the linearly growing regression errors existing in the backward schemes; Bender & Steiner (2012) [5] suggested a possible improvement of the scheme [31] by using martingale basis functions for regressions; Crisan & Monolarakis (2014) [19] developed a second-order discretization using the cubature method. A different scheme based on the optimal quantization was proposed by Bally & Pagès (2003) [6]. See Pagès & Sagna (2015) [41] and references therein for its recent developments. Delarue & Menozzi (2006, 2008) [22, 23] studied a class of quasi-linear PDEs via a coupled forward-backward SDE with Lipschitz functions, which, in particular, becomes equivalent to a special type of quadratic-growth (qg) BSDE (so called deterministic KPZ equation) under a certain setting. Recently, Chassagneux & Richou (2016) [15] extended the standard backward scheme to qg-BSDEs with bounded terminal conditions in a Markovian setting.

As another approach, a semi-analytic approximation scheme was proposed by Fujii & Takahashi (2012) [26] and justified in the Lipschitz case by Takahashi & Yamada (2015) [47]. An efficient implementation algorithm based on an interacting particle method by Fujii & Takahashi (2015) [28] has been successfully applied to a large scale credit portfolio by Crépey & Song (2015) [18]. This is an asymptotic expansion around a linear driver motivated by the observation that, for many financial applications, the non-linear part of the driver is proportional to an interest rate spread and/or a default intensity which is, at most, of the order of a few percentage points. Although it cuts the cost of numerical computation drastically under many interesting situations, the non-linear effects may grow and cease to be perturbative when one deals with longer maturities, higher volatilities, or general risk-sensitive control problems for highly concave utility functions. For example, the quadratic-growth terms of the control variables appearing in the exponential utility optimization may give rise to significant non-linearity when the risk-averseness is high.

In this paper, we propose a new approximation scheme for Markovian qg-BSDEs with bounded terminal functions, which aims to achieve the advantages of both the standard Monte Carlo scheme, in terms of generality, and also the semi-analytic approximation scheme, in terms of the lesser numerical cost. The main idea is to decompose the original qg-BSDE into a sequence of qg-BSDEs each of which is defined in a short-time interval. We then employ an asymptotic expansion method to solve each of them approximately.[1] In order to obtain the total error estimate, we first investigate the propagation of error for a sequence of qg-BSDEs with terminal conditions perturbed by bounded functions, say $\{\delta^i\}_{1\leq i\leq n}$. The first main result is thus obtained as Theorem 3.1. We then substitute the error function associated to the asymptotic expansion in each period for the function $\delta^i$, which then leads to our second main result of Theorem 4.1 providing the total error estimate for the proposed scheme.

Although there remains assumptions which cannot be confirmed a priori, which is a drawback of the current scheme, they can still be checked numerically for each model (under one plausible assumption). Once it is confirmed, the convergence with the rate of $n^{-1/2+\delta}$ for all $\delta > 0$ in the strong sense is obtained for a finite range of discretization. In the case of the standard scheme with Monte Carlo simulation, although the convergence is guaranteed a priori for *sufficiently small discretization and many paths*, it is not completely free from a

[1] Similar ideas have been applied to stochastic filtering by Fujii (2014) [27] and to European option pricing by Takahashi & Yamada (2016) [48].

posteriori checks, either. One still needs to run heavy tests with varying discretization sizes and number of paths to confirm whether a given result provides a reasonable approximation or not. The advantage of the proposed scheme is that one can avoid time-consuming simulation for estimating conditional expectations at each space-time node by using simple semi-analytic results and finer discretization becomes easier to implement. We give numerical examples for both qg- and Lipschitz BSDEs to illustrate the empirical performance. They suggest that the scheme works well even for very large quadratic coefficients, and all the more so for large Lipschitz constants. Note also that, the short-term asymptotic expansion of qg-BSDEs in the strong sense is provided for the first time, which may be useful for other applications.

The organization of the paper is as follows: Section 2 explains the general setting and notations, Section 3 gives the time-discretization and investigates a sequence of qg-BSDEs perturbed in the terminal values; Section 4 applies the short-term expansion to the result of Section 3 which yields the total error estimate of the proposed scheme. Section 5 explains a concrete implementation using a discretized space-time grid and the corresponding error estimates. Finally Appendix A summarizes the properties of BMO-martingales, Appendices B and C derive the formula of the short-term asymptotic expansion and obtain the error estimates relevant for the analysis in the main text.

# 2 Preliminaries

## 2.1 General setting and notations

Throughout the paper, we fix the terminal time $T > 0$. We work on the filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ carrying a $d$-dimensional independent standard Brownian motion $W$. $\mathbb{F} = (\mathcal{F}_t)_{t\in[0,T]}$ is the Brownian filtration satisfying the usual conditions augmented by the $\mathbb{P}$-zero sets. We denote a generic positive constant by $C$, which may change line by line and it is sometimes associated with several subscripts (such as $C_{p,K}$) when there is a need to emphasize its dependency on those parameters. $\mathcal{T}_0^T$ denotes the set of all $\mathbb{F}$-stopping times $\tau : \Omega \to [0, T]$. We denote the sup-norm of $\mathbb{R}^k$-valued function $x : [0, T] \to \mathbb{R}^k$, $k \in \mathbb{N}$ by the symbol $||x||_{[a,b]} := \sup\big\{|x_t|, t \in [a, b]\big\}$ and write $||x||_t := ||x||_{[0,t]}$.

Let us introduce the following spaces for stochastic processes with $p \geq 2$ and $k \in \mathbb{N}$. For the convenience of the reader, we separately summarize the relevant properties of the BMO martingales and the associated function spaces in Appendix A.

• $\mathcal{S}^p_{[s,t]}(\mathbb{R}^k)$ is the set of $\mathbb{R}^k$-valued adapted processes $X$ satisfying

$$||X||_{\mathcal{S}^p_{[s,t]}} := \mathbb{E}\big[||X||^p_{[s,t]}\big]^{1/p} < \infty \ .$$

• $\mathcal{S}^\infty_{[s,t]}(\mathbb{R}^k)$ is the set of $\mathbb{R}^k$-valued essentially bounded adapted processes $X$ satisfying

$$||X||_{\mathcal{S}^\infty_{[s,t]}} := \big|\big|\sup_{r\in[s,t]} |X_r|\big|\big|_\infty < \infty \ .$$

• $\mathcal{H}^p_{[s,t]}(\mathbb{R}^k)$ is the set of $\mathbb{R}^k$-valued progressively measurable processes $Z$ satisfying

$$||Z||_{\mathcal{H}^p_{[s,t]}} := \mathbb{E}\Big[\Big(\int_s^t |Z_r|^2 dr\Big)^{\frac{p}{2}}\Big]^{\frac{1}{p}} < \infty.$$

• $\mathcal{K}^p[s, t]$ is the set of functions $(Y, Z)$ in the space $\mathcal{S}^p_{[s,t]}(\mathbb{R}) \times \mathcal{H}^p_{[s,t]}(\mathbb{R}^{1\times d})$ with the norm defined by

$$||(Y, Z)||_{\mathcal{K}^p[s,t]} := \big(||Y||^p_{\mathcal{S}^p_{[s,t]}} + ||Z||^p_{\mathcal{H}^p_{[s,t]}}\big)^{1/p} \ .$$

We frequently omit the argument $\mathbb{R}^k$ and subscript $[s,t]$ if they are obvious from the context. We use $\big(\Theta_s, s \in [0,T]\big)$ as a collective argument $\Theta_s := (Y_s, Z_s)$ to lighten the notation. We use the following notation for partial derivatives with respect to $x \in \mathbb{R}^d$ such that

$$\partial_x = (\partial_{x^1}, \cdots, \partial_{x^d}) = \big(\frac{\partial}{\partial x^1}, \cdots, \frac{\partial}{\partial x^d}\big)$$

and $\partial_\theta = (\partial_y, \partial_z)$ for the collective argument $\Theta$.

## 2.2 Setup

Firstly, we introduce the underlying forward process $X_t, t \in [0,T]$:

$$X_t = x_0 + \int_0^t b(r, X_r)dr + \int_0^t \sigma(r, X_r)dW_r \ , \tag{2.1}$$

where $x_0 \in \mathbb{R}^d$, $b : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$, $\sigma : [0,T] \times \mathbb{R}^d \to \mathbb{R}^{d\times d}$.[2]

**Assumption 2.1.** *(i) For all $t, t' \in [0,T]$ and $x, x' \in \mathbb{R}^d$, there exists a positive constant $K$ such that $|b(t,x) - b(t',x')| + |\sigma(t,x) - \sigma(t',x')| \le K\big(|t-t'|^{\frac{1}{2}} + |x - x'|\big)$.*
*(ii) $||b(\cdot,0)||_T + ||\sigma(\cdot,0)||_T \le K$.*
*(iii) $b$ and $\sigma$ are 3-time continuously differentiable with respect to $x$ and satisfy*

$$\begin{aligned} &|\partial_x^m b(t,x)| + |\partial_x^m \sigma(t,x)| \le K \ , \\ &|\partial_x^m b(t,x) - \partial_x^m b(t',x)| + |\partial_x^m \sigma(t,x) - \partial_x^m \sigma(t',x)| \le K|t-t'|^{1/2} \ , \end{aligned} \tag{2.2}$$

*for all $1 \le m \le 3$, $t, t' \in [0,T]$ and $x \in \mathbb{R}^d$.*

Let us now introduce a qg-BSDE which is a target of our investigation:

$$Y_t = \xi(X_T) + \int_t^T f(r, X_r, Y_r, Z_r)dr - \int_t^T Z_r dW_r, \quad t \in [0,T] \tag{2.3}$$

where $\xi : \mathbb{R}^d \to \mathbb{R}$, $f : [0,T] \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^{1\times d} \to \mathbb{R}$.

**Assumption 2.2.** *(i) $f$ satisfies the quadratic structure condition [2]:*

$$|f(t,x,y,z)| \le l_t + \beta|y| + \frac{\gamma}{2}|z|^2$$

*for all $(t,x,y,z) \in [0,T] \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^{1\times d}$, where $\beta \ge 0, \gamma > 0$ are constants, $l : [0,T] \to \mathbb{R}_+$ is a positive function bounded by a constant $K$, i.e. $||l||_T \le K$.*
*(ii) $f$ satisfies, for all $t, t' \in [0,T]$, $y, y' \in \mathbb{R}$, $x, x' \in \mathbb{R}^d$, $z, z' \in \mathbb{R}^{1\times d}$,*

$$\begin{aligned} |f(t,x,y,z) - f(t',x,y,z)| &\le K|t-t'|^{1/2} \ , \\ |f(t,x,y,z) - f(t,x,y',z)| &\le K|y-y'| \ , \\ |f(t,x,y,z) - f(t,x,y,z')| &\le K\big(1 + |z| + |z'|\big)|z-z'| \ , \\ |f(t,x,y,z) - f(t,x',y,z)| &\le K\big(1 + |y| + |z|^2\big)|x-x'|. \end{aligned}$$

[2] Useful standard estimates on the Lipschitz SDEs can be found, for example, in Appendix A of [30].

*(iii) the driver $f$ is 1-time continuously differentiable with respect to the spacial variables with continuous derivatives. In particular, we assume that*

$$|\partial_y f(t,x,y,z)| \le K, \quad |\partial_z f(t,x,y,z)| \le K(1+|z|), \quad |\partial_x f(t,x,y,z)| \le K(1+|y|+|z|^2)$$

*for all $(t,x,y,z) \in [0,T] \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^{1\times d}$.*
*(iv) $\xi$ is a 3-time continuously differentiable function satisfying that $||\xi(\cdot)||_{\mathbb{L}^\infty} \le K$ and $||\partial_x^m \xi(\cdot)||_{\mathbb{L}^\infty} \le K$ for every $1 \le m \le 3$.*

**Remark 2.1.** *The 2nd- and 3rd-order differentiability in Assumptions 2.1 and 2.2 is relevant only for the later part of the discussions (Section 4∼), where the error estimate of the short-term expansions are required.*

It has been well-known since the work of Kobylanski (2000) [35] that there exists a unique solution $(Y,Z)$ to (2.3) in the space $(Y,Z) \in \mathcal{S}^\infty \times \mathcal{H}^2_{BMO}$.

**Lemma 2.1.** *(universal bound) Under Assumptions 2.1 and 2.2, the solution $(Y,Z) \in \mathcal{S}^\infty \times \mathcal{H}^2_{BMO}$ of (2.3) satisfies*

$$||Y||_{\mathcal{S}^\infty} \le e^{\beta T}\Big(||\xi(\cdot)||_{\mathbb{L}^\infty} + T||l||_T\Big) ,$$
$$||Z||^2_{\mathcal{H}^2_{BMO}} \le \frac{e^{4\gamma||Y||_{\mathcal{S}^\infty}}}{\gamma^2}\Big(3 + 6\gamma T(\beta||Y||_{\mathcal{S}^\infty} + ||l||_T)\Big) .$$

*Proof.* This follows straightforwardly from the quadratic structure condition [2] which is given by Assumption 2.2 (i). See, for example, Lemma 3.1 and 3.2 in [29]. □

# 3 A sequence of qg-BSDEs perturbed in terminals

In this section, we investigate a sequence of qg-BSDEs. For each connecting point, we introduce a bounded perturbation function $\delta^i : \mathbb{R}^d \to \mathbb{R}$. We then investigate the propagation of its effects to the total error.

## 3.1 Setup

Let us introduce a time partition $\pi : 0 = t_0 < t_1 < \cdots < t_n = T$. We put $h_i := t_i - t_{i-1}$, $|\pi| := \max_{1\le i\le n} h_i$. We denote each interval by $I_i := [t_{i-1}, t_i]$, $i \in \{1,\cdots,n\}$ and assume that there exists some positive constant $C$ such that $|\pi|n \le C$ as well as $|\pi|/h_i \le C$ for every $i \in \{1,\cdots,n\}$. In order to approximate the BSDE (2.3), we introduce a sequence of qg-BSDEs perturbed in the terminal values for each interval $t \in I_i$, $i \in \{1,\cdots,n\}$ in the following way:

$$\overline{Y}^i_t = \widehat{u}^{i+1}(X_{t_i}) + \int_t^{t_i} f(r, X_r, \overline{Y}^i_r, \overline{Z}^i_r)dr - \int_t^{t_i} \overline{Z}^i_r dW_r , \tag{3.1}$$

where $\widehat{u}^{i+1} : \mathbb{R}^d \to \mathbb{R}$ with $\widehat{u}^{n+1}(x) := \xi(x)$.

Each terminal function $\widehat{u}^{i+1}(x)$, $x \in \mathbb{R}^d$ of the period $I_i$ is defined by $\big(\overline{Y}^{i+1,t_i,x}_t, t \in [t_i, t_{i+1}]\big)$, which is the solution of (3.1) for the period $I_{i+1}$ corresponding to the underlying

process $X$ with the initial data $(t_i, X_{t_i} = x)$ [3], and the additional perturbation term $\delta^{i+1}$

$$\widehat{u}^{i+1}(x) \quad := \quad \overline{Y}_{t_i}^{i+1,t_i,x} - \delta^{i+1}(x), \quad i \in \{1, \cdots, n-1\}. \tag{3.2}$$

In the next section, $\{\delta^i\}_{i \le n}$ will be related to the errors from the short-term expansion.

**Assumption 3.1.** *(i)The perturbation terms $\delta^{i+1} : \mathbb{R}^d \to \mathbb{R}$, $i \in \{1, \cdots, n\}$ are absolutely bounded functions that keep $(\widehat{u}^{i+1})_{1 \le i \le n}$ satisfying the conditions*

$$\begin{aligned}
&(a) \qquad \max_{1 \le i \le n} ||\widehat{u}^{i+1}(\cdot)||_{\mathbb{L}^\infty} \le K', \\
&(b) \qquad \max_{1 \le i \le n} ||\partial_x \widehat{u}^{i+1}(\cdot)||_{\mathbb{L}^\infty} \le K', \\
&(c) \qquad \max_{1 \le i \le n} ||\partial_x^m \widehat{u}^{i+1}(\cdot)||_{\mathbb{L}^\infty} \le K', \ (m \in \{2, 3\})
\end{aligned}$$

*with some $n$-independent positive constant $K'$.*[4]
*(ii) There exists an $n$-independent positive constant $C$ such that $\sum_{i=1}^{n-1} ||\delta^{i+1}(\cdot)||_{\mathbb{L}^\infty} \le C$.*

We use the convention $\delta^{n+1} \equiv 0$ and $\overline{Y}_{t_n}^{n+1} = \xi(X_{t_n})$ in the following.

**Remark 3.1.** *The condition (i)(c) becomes relevant only for the short-term expansion in the next section.*

**Remark 3.2.** *The classical (as well as variational) differentiability of qg-BSDEs is well-known by the works of Ankirchner et al. (2007) [1], Briand & Confortola (2008) [13] and Imkeller & Reis (2010) [32]. See Fujii & Takahashi (2015) [29] for the extension of these results to qg-BSDEs with Poisson random measures.*

## 3.2 Properties of the solution

Applying the known results of qg-BSDE for each period, one sees that there exists a unique solution $(\overline{Y}^i, \overline{Z}^i) \in \mathcal{S}^\infty_{[t_{i-1},t_i]} \times \mathcal{H}^2_{BMO[t_{i-1},t_i]}$. Applying Lemma 2.1 for each period $I_i$, one also sees

$$||\overline{Y}^i||_{\mathcal{S}^\infty[t_{i-1},t_i]} \le ||\overline{Y}||_{\mathcal{S}^\infty} := e^{\beta|\pi|}\Big(K' + |\pi|||l||_T\Big) \ , \tag{3.3}$$

which is bounded uniformly in $i \in \{1, \cdots, n\}$, and so is $||\overline{Z}^i||_{\mathcal{H}^2_{BMO}[t_{i-1},t_i]}$.

**Proposition 3.1.** *Under Assumptions 2.1, 2.2 and Assumption 3.1 [i(a,b)], there exists some positive $(i, n)$-independent constant $C$ such that the process $\overline{Z}^i_t, t \in I_i$ of the solution to the BSDE (3.1) satisfies*

$$|\overline{Z}^i_t| \le C(1 + |X_t|), \ t \in I_i$$

*uniformly in $i \in \{1, \cdots, n\}$.*

*Proof.* We use the representation theorem for the control variable (Theorem 8.5 in [1]) and follow the arguments of Theorem 3.1 in Ma & Zhang (2002) [39]. Let us introduce the

[3] In other words, the underlying forward process is given by $X_s^{t_i,x} = x + \int_{t_i}^s b(r, X_r^{t_i,x})dr + \int_{t_i}^s \sigma(r, X_r^{t_i,x})dW_r$, $s \in I_{i+1}$, and hence $\overline{Y}_{t_i}^{i+1,t_i,x}$ is a deterministic function of $x \in \mathbb{R}^d$.

[4] The exact size of $K'$ is somewhat arbitrary if it is big enough not to contradict the true solution of (2.3).

parameterized solution $(X^{t,x}, \overline{Y}^{i,t,x}, \overline{Z}^{i,t,x})$ with the initial data $(t,x) \in [t_{i-1}, t_i] \times \mathbb{R}^d$:

$$X_s^{t,x} = x + \int_t^s b(r, X_r^{t,x})dr + \int_t^s \sigma(r, X_r^{t,x})dW_r, \tag{3.4}$$

$$\overline{Y}_s^{i,t,x} = \widehat{u}^{i+1}(X_{t_i}^{t,x}) + \int_s^{t_i} f(r, X_r^{t,x}, \overline{Y}_r^{i,t,x}, \overline{Z}_r^{i,t,x})dr - \int_s^{t_i} \overline{Z}_r^{i,t,x} dW_r, \tag{3.5}$$

$s \in [t, t_i]$ where the classical differentiability of (3.4) and (3.5) with respect to the position $x$ is known [1]. The differential processes $(\partial_x X^{t,x}, \partial_x \overline{Y}^{i,t,x}, \partial_x \overline{Z}^{i,t,x})$ are given by the solutions to the following forward- and backward-SDE:

$$\begin{aligned}
\partial_x X_s^{t,x} &= \mathbb{I} + \int_t^s \partial_x b(r, X_r^{t,x})\partial_x X_r^{t,x} dr + \int_t^s \partial_x \sigma(r, X_r^{t,x})\partial_x X_r^{t,x} dW_r, \\
\partial_x \overline{Y}_s^{i,t,x} &= \partial_x \widehat{u}^{i+1}(X_{t_i}^{t,x})\partial_x X_{t_i}^{t,x} + \int_t^{t_i} \Big\{ \partial_x f(r, X_r^{t,x}, \overline{\Theta}_r^{i,t,x})\partial_x X_r^{t,x} \\
&\quad + \partial_\theta f(r, X_r^{t,x}, \overline{\Theta}_r^{i,t,x})\partial_x \overline{\Theta}_r^{i,t,x} \Big\} dr - \int_t^{t_i} \partial_x \overline{Z}_r^{i,t,x} dW_r,
\end{aligned} \tag{3.6}$$

where $\mathbb{I}$ is the $d \times d$ identity matrix and $\partial_x \overline{\Theta}^{i,t,x} = (\partial_x \overline{Y}^{i,t,x}, \partial_x \overline{Z}^{i,t,x})$. Note that $|\partial_y f|$ is bounded and $|\partial_z f(r, X_r^{t,x}, \overline{\Theta}_r^{i,t,x})| \leq K\big(1 + |\overline{Z}_r^{i,t,x}|\big)$ by Assumption 2.2 (iv). By the facts given in (3.3) and the remark that follows, one sees $||\partial_z f(\cdot, X_\cdot^{t,x}, \overline{\Theta}_\cdot^{i,t,x})||_{\mathcal{H}^2_{BMO[t,t_i]}} \leq C$ with some constant $C$. Thus Corollary 9 in [13] or Theorem A.1 in [29] implies that the BSDE (3.6) has a unique solution satisfying, for any $p \geq 2$,

$$\begin{aligned}
\big|\big|\partial_x \overline{\Theta}^{i,t,x}\big|\big|^p_{\mathcal{K}^p[t,t_i]} &\leq C_{p,\bar{q}} \mathbb{E}\Big[ |\partial_x \widehat{u}^{i+1}(X_{t_i}^{t,x})\partial_x X_{t_i}^{t,x}|^{p\bar{q}^2} + \Big( \int_t^{t_i} |\partial_x f(r, X_r^{i,t,x}, \overline{\Theta}_r^{i,t,x})\partial_x X_r^{i,t,x}| dr \Big)^{p\bar{q}^2} \Big]^{\frac{1}{\bar{q}^2}} \\
&\leq C_{p,\bar{q}} \mathbb{E}\Big[ ||\partial_x X^{t,x}||^{2p\bar{q}^2}_{[t,t_i]} \Big]^{\frac{1}{2\bar{q}^2}} \Big( 1 + h_i^p \mathbb{E}\Big[ ||\overline{Y}^i||^{2p\bar{q}^2}_{[t,t_i]} \Big]^{\frac{1}{2\bar{q}^2}} + \mathbb{E}\Big[ \Big( \int_t^{t_i} |\overline{Z}_r^{i,t,x}|^2 dr \Big)^{2p\bar{q}^2} \Big]^{\frac{1}{2\bar{q}^2}} \Big)
\end{aligned} \tag{3.7}$$

where $\bar{q}$ is a positive constant satisfying $q_* \leq \bar{q} < \infty$. Here, $q_* = \frac{r^*}{r^*-1} > 1$ is the conjugate exponent of $r^*$ the upper bound of power with which the Reverse Hölder inequality holds for $\mathcal{E}(\partial_z f * W)$. We have used Assumption 2.2 (iv) and Hölder inequality in the last line.

By the standard estimate of SDE [5], one can show that $||\partial_x X^{t,x}||_{\mathcal{S}^{2p\bar{q}^2}} \leq C$ with some positive constant $C$ that is independent of the initial data $(t,x)$. The boundedness of $\overline{Y}^i$ in (3.3) and the following remark on $\overline{Z}^i$ together with Lemma A.1 show that the right-hand side of (3.7) is bounded by some positive constant. In particular, one can choose a common constant $C$ for every $i \in \{1, \cdots, n\}$ such that $|\partial_x \overline{Y}_t^{i,t,x}| \leq ||\partial_x \overline{Y}^{i,t,x}||_{\mathcal{S}^p[t,t_i]} \leq C$ uniformly in $(t,x) \in [t_{i-1}, t_i] \times \mathbb{R}^d$. By the representation theorem [1, 39], we have $\overline{Z}_t^i = \partial_x \overline{u}^i(t, X_t)\sigma(t, X_t)$, $t \in [t_{i-1}, t_i]$ where the function $\partial_x \overline{u}^i : [t_{i-1}, t_i] \times \mathbb{R}^d \to \mathbb{R}^{1\times d}$ is defined by $\partial_x \overline{u}^i(t,x) := \partial_x \overline{Y}_t^{i,t,x}$. Now the Lipschitz property of $\sigma$ gives the desired result. □

Let us now define a progressively measurable process $\big(\overline{Z}_t, t \in [0,T]\big)$ by

$$\overline{Z}_t := \sum_{i=1}^n \overline{Z}_t^i \mathbf{1}_{\{t_{i-1} \leq t < t_i\}}, \ t \in [0,T] \tag{3.8}$$

[5] See, for example, Appendix A in [30].

so that $\int_0^T |\overline{Z}_t|^2 dt = \sum_{i=1}^n \int_{t_{i-1}}^{t_i} |\overline{Z}_t^i|^2 dt$.

**Proposition 3.2.** *Under Assumptions 2.1, 2.2 and 3.1 [i(a), ii], the process* $\big(\overline{Z}_t,\ t\in[0,T]\big)$ *defined by (3.8) belongs to* $\mathcal{H}^2_{BMO[0,T]}$ *satisfying* $||\overline{Z}||_{\mathcal{H}^2_{BMO}[0,T]} \le C$ *with some n-independent positive constant* $C$.

*Proof.* Applying Itô formula to $e^{2\gamma\overline{Y}^i}$, one obtains for $t\in I_i$ that

$$\int_t^{t_i} e^{2\gamma\overline{Y}^i_r} 2\gamma^2|\overline{Z}^i_r|^2 dr = e^{2\gamma\overline{Y}^i_{t_i}} - e^{2\gamma\overline{Y}^i_t} + \int_t^{t_i} e^{2\gamma\overline{Y}^i_r} 2\gamma f(r,X_r,\overline{Y}^i_r,\overline{Z}^i_r)dr - \int_t^{t_i} e^{2\gamma\overline{Y}^i_r} 2\gamma\overline{Z}^i_r dW_r\ .$$

The quadratic structure condition in Assumption 2.2 (i) gives

$$\int_t^{t_i} e^{2\gamma\overline{Y}^i_r}\gamma^2|\overline{Z}^i_r|^2 dr \le e^{2\gamma\overline{Y}^i_{t_i}} - e^{2\gamma\overline{Y}^i_t} + \int_t^{t_i} e^{2\gamma\overline{Y}^i_r} 2\gamma\big(l_r+\beta|\overline{Y}^i_r|\big)dr - \int_t^{t_i} e^{2\gamma\overline{Y}^i_r} 2\gamma\overline{Z}^i_r dW_r\ .$$

Since $\overline{Y}^i_{t_i} = \widehat{u}^{i+1}(X_{t_i}) = \overline{Y}^{i+1}_{t_i} - \delta^{i+1}(X_{t_i})$ and $||\delta^{i+1}(\cdot)||_{\mathbb{L}^\infty}\le C$ uniformly in $i\in\{1,\cdots,n\}$, one has $e^{2\gamma\overline{Y}^i_{t_i}} \le e^{2\gamma\overline{Y}^{i+1}_{t_i}} + Ce^{2\gamma\overline{Y}^{i+1}_{t_i}}|\delta^{i+1}(X_{t_i})|$ with some positive constant $C$. It follows that, with the choice $t=t_{i-1}$,

$$\begin{aligned}\int_{t_{i-1}}^{t_i} e^{2\gamma\overline{Y}^i_r}\gamma^2|\overline{Z}^i_r|^2 dr &\le \big(e^{2\gamma\overline{Y}^{i+1}_{t_i}} - e^{2\gamma\overline{Y}^i_{t_{i-1}}}\big) + Ce^{2\gamma\overline{Y}^{i+1}_{t_i}}|\delta^{i+1}(X_{t_i})| \\ &\quad + \int_{t_{i-1}}^{t_i} e^{2\gamma\overline{Y}^i_r} 2\gamma\big(l_r+\beta|\overline{Y}^i_r|\big)dr - \int_{t_{i-1}}^{t_i} e^{2\gamma\overline{Y}^i_r} 2\gamma\overline{Z}^i_r dW_r.\end{aligned}$$

Thus for any $\tau\in\mathcal{T}_0^T$ and $j:=\min\big(j\in\{1,\cdots,n\}:\tau\le t_j\big)$,

$$\begin{aligned}\int_\tau^{t_j} e^{2\gamma\overline{Y}^j_r}\gamma^2|\overline{Z}^j_r|^2 dr + \sum_{i=j+1}^n\int_{t_{i-1}}^{t_i} e^{2\gamma\overline{Y}^i_r}\gamma^2|\overline{Z}^i_r|^2 dr &\le e^{2\gamma\overline{Y}^{n+1}_{t_n}} - e^{2\gamma\overline{Y}^j_\tau} + C\sum_{i=j}^n e^{2\gamma\overline{Y}^{i+1}_{t_i}}|\delta^{i+1}(X_{t_i})| \\ &\quad + \int_\tau^{t_j} e^{2\gamma\overline{Y}^j_r} 2\gamma\big(l_r+\beta|\overline{Y}^j_r|\big)dr + \sum_{i=j+1}^n\int_{t_{i-1}}^{t_i} e^{2\gamma\overline{Y}^i_r} 2\gamma\big(l_r+\beta|\overline{Y}^i_r|\big)dr \\ &\quad - \int_\tau^{t_j} e^{2\gamma\overline{Y}^j_r} 2\gamma\overline{Z}^j_r dW_r - \sum_{i=j+1}^n\int_{t_{i-1}}^{t_i} e^{2\gamma\overline{Y}^i_r} 2\gamma\overline{Z}^i_r dW_r\ .\end{aligned}$$

Since $e^{2\gamma\overline{Y}^j_\tau}>0$ and $\delta^{n+1}\equiv 0$, one obtains

$$\begin{aligned}&\mathbb{E}\Big[\int_\tau^{t_j} e^{2\gamma\overline{Y}^j_r}\gamma^2|\overline{Z}^j_r|^2 dr + \sum_{i=j+1}^n\int_{t_{i-1}}^{t_i} e^{2\gamma\overline{Y}^i_r}\gamma^2|\overline{Z}^i_r|^2 dr\Big|\mathcal{F}_\tau\Big] \\ &\quad\le \mathbb{E}\Big[e^{2\gamma\xi(X_T)} + C\sum_{i=j}^{n-1} e^{2\gamma\overline{Y}^{i+1}_{t_i}}|\delta^{i+1}(X_{t_i})| + \sum_{i=j}^n\int_{t_{i-1}\vee\tau}^{t_i} e^{2\gamma\overline{Y}^i_r} 2\gamma\big(l_r+\beta|\overline{Y}^i_r|\big)dr\Big|\mathcal{F}_\tau\Big]\ .\end{aligned}$$

By Assumption 3.1 $[i(a), ii]$ and (3.3), the above inequality implies that there exists some

$n$-independent constant $C$ such that

$$\mathbb{E}\Big[\int_\tau^T |\overline{Z}_r|^2 dr\Big|\mathcal{F}_\tau\Big] \le \frac{e^{4\gamma||\overline{Y}||_{\mathcal{S}^\infty}}}{\gamma^2}\Big(1+2\gamma T\big(||l||_T+\beta||\overline{Y}||_{\mathcal{S}^\infty}\big)+C\sum_{i=1}^{n-1}||\delta^{i+1}(\cdot)||_{\mathbb{L}^\infty}\Big) \le C \ ,$$

and thus the claim is proved. □

### 3.3 Error estimates for the perturbed BSDEs in the terminals

Let $(Y,Z)$ be the solution to the BSDE (2.3) and $(\overline{Y}^i,\overline{Z}^i)$ to (3.1). Let us put

$$\delta Y_t^i := Y_t - \overline{Y}_t^i, \quad \delta Z_t^i := Z_t - \overline{Z}_t^i, \quad \delta f^i(t) := f(t,X_t,Y_t,Z_t) - f(t,X_t,\overline{Y}_t^i,\overline{Z}_t^i) \ ,$$

for $t \in I_i$, $i \in \{1,\cdots,n\}$. Then $(\delta Y^i, \delta Z^i)$ follows the BSDE

$$\delta Y_t^i = \delta Y_{t_i}^{i+1} + \delta^{i+1}(X_{t_i}) + \int_t^{t_i} \delta f^i(r)dr - \int_t^{t_i} \delta Z_r^i dW_r \ , \tag{3.9}$$

for $t \in I_i$. Our first main result is given as follows.

**Theorem 3.1.** *Under Assumptions 2.1, 2.2 and 3.1 [i(a,b),ii], there exist some $n$-independent positive constants $\bar{q} > 1$ and $C_{p,\bar{q}}$ such that, for any $p > 1$,*

$$\max_{1\le i\le n} \mathbb{E}\left[\sup_{r\in I_i}|\delta Y_r^i|^{2p}\right]^{\frac{1}{p}} + \sum_{i=1}^n \int_{t_{i-1}}^{t_i} \mathbb{E}|\delta Z_r^i|^2 dr \le \frac{C_{p,\bar{q}}}{|\pi|}\mathbb{E}\Big[\Big(\sum_{i=1}^{n-1}|\delta^{i+1}(X_{t_i})|^2\Big)^{p\bar{q}}\Big]^{\frac{1}{p\bar{q}}} \ .$$

*Proof.* It directly follows from the next Propositions 3.3 and 3.4. □

**Remark 3.3.** *Theorem 3.1 implies that we need $E|\delta|^2 \propto n^{-k}$ with $k > 2$ for the right-hand side to converge. We shall see in fact that $k = 3$ is achieved by the short-term expansion.*

**Proposition 3.3.** *Under Assumptions 2.1, 2.2 and Assumption 3.1 [i(a,b)], the inequality*

$$\sum_{i=1}^n \int_{t_{i-1}}^{t_i} \mathbb{E}|\delta Z_r^i|^2 dr \le C_p \max_{1\le i\le n} \mathbb{E}\Big[\sup_{r\in I_i}|\delta Y_r^i|^{2p}\Big]^{1/p} + \frac{C}{|\pi|}\sum_{i=1}^{n-1}\mathbb{E}|\delta^{i+1}(X_{t_i})|^2 \ ,$$

*holds for any $p > 1$ with some $n$-independent positive constants $C$ and $C_p$.*

*Proof.* For each interval $I_i$, let us define new progressively measurable processes $\big(\beta_r^i, r \in I_i\big)$ and $\big(\gamma_r^i, r \in I_i\big)$ as follows:

$$\beta_r^i := \frac{f(r,X_r,Y_r,Z_r) - f(r,X_r,\overline{Y}_r^i,Z_r)}{\delta Y_r^i}\mathbf{1}_{\delta Y_r^i \ne 0}, \quad \gamma_r^i := \frac{f(r,X_r,\overline{Y}_r^i,Z_r) - f(r,X_r,\overline{Y}_r^i,\overline{Z}_r^i)}{|\delta Z_r^i|^2}\mathbf{1}_{\delta Z_r^i \ne 0}(\delta Z_r^i)^\top \ .$$

Then, $|\beta^i| \le K$ is a bounded process by the Lipschitz property, and by Proposition 3.1, there exists some $(i,n)$-independent positive constant $C$ such that

$$|\gamma_r^i| \le K(1+|Z_r|+|\overline{Z}_r^i|) \le C(1+|X_r|) \tag{3.10}$$

for $r \in I_i$, $i \in \{1, \cdots, n\}$. The BSDE (3.9) can now be written as

$$\delta Y_t^i = \delta Y_{t_i}^{i+1} + \delta^{i+1}(X_{t_i}) + \int_t^{t_i} \Big(\beta_r^i \delta Y_r^i + \delta Z_r^i \gamma_r^i\Big) dr - \int_t^{t_i} \delta Z_r^i dW_r \ . \tag{3.11}$$

A simple application of Itô formula gives

$$\mathbb{E}|\delta Y_{t_{i-1}}^i|^2 + \int_{t_{i-1}}^{t_i} \mathbb{E}|\delta Z_r^i|^2 dr = \mathbb{E}\big|\delta Y_{t_i}^{i+1} + \delta^{i+1}(X_{t_i})\big|^2 + \int_{t_{i-1}}^{t_i} \mathbb{E}\Big[2\delta Y_r^i\big(\beta_r^i \delta Y_r^i + \delta Z_r^i \gamma_r^i\big)\Big] dr \ .$$

By Hölder inequality and (3.10), one obtains with some positive constants $C, C_p$ that

$$\begin{aligned}
\frac{1}{2}\int_{t_{i-1}}^{t_i} \mathbb{E}|\delta Z_r^i|^2 dr &\leq \Big(\mathbb{E}|\delta Y_{t_i}^{i+1}|^2 - \mathbb{E}|\delta Y_{t_{i-1}}^i|^2\Big) + C|\pi|\mathbb{E}|\delta Y_{t_i}^{i+1}|^2 + \frac{C}{|\pi|}\mathbb{E}|\delta^{i+1}(X_{t_i})|^2 \\
&\quad + C\int_{t_{i-1}}^{t_i} \mathbb{E}\Big[|\delta Y_r^i|^2(1+|\gamma_r^i|^2)\Big] dr \\
&\leq \Big(\mathbb{E}|\delta Y_{t_i}^{i+1}|^2 - \mathbb{E}|\delta Y_{t_{i-1}}^i|^2\Big) + C|\pi|\mathbb{E}|\delta Y_{t_i}^{i+1}|^2 + \frac{C}{|\pi|}\mathbb{E}|\delta^{i+1}(X_{t_i})|^2 + C_p|\pi|\mathbb{E}\Big[\sup_{r\in I_i}|\delta Y_r^i|^{2p}\Big]^{1/p} \ ,
\end{aligned}$$

where $p$ is an arbitrary constant satisfying $p > 1$. Summing up for $i \in \{1, \cdots, n\}$, one obtains

$$\begin{aligned}
\frac{1}{2}\sum_{i=1}^n \int_{t_{i-1}}^{t_i} \mathbb{E}|\delta Z_r^i|^2 dr &\leq \mathbb{E}|\delta Y_{t_n}^{n+1}|^2 - \mathbb{E}|\delta Y_{t_0}^1|^2 + C|\pi|\sum_{i=1}^n \mathbb{E}|\delta Y_{t_i}^{i+1}|^2 \\
&\quad + \frac{C}{|\pi|}\sum_{i=1}^n \mathbb{E}|\delta^{i+1}(X_{t_i})|^2 + C_p|\pi|\sum_{i=1}^n \mathbb{E}\Big[\sup_{r\in I_i}|\delta Y_r^i|^{2p}\Big]^{1/p} \ .
\end{aligned}$$

Since $\delta Y_{t_n}^{n+1} = \delta^{n+1} = 0$, one gets the desired result. □

**Proposition 3.4.** *Under Assumptions 2.1, 2.2 and 3.1 [i(a),ii], there exists some $n$-independent positive constants $\bar{q} > 1$ and $C_{p,\bar{q}}$ such that, for any $p > 1$,*

$$\mathbb{E}\Big[\max_{1\leq i\leq n}\sup_{r\in I_i}|\delta Y_r^i|^p\Big] \leq C_{p,\bar{q}}\mathbb{E}\Big[\Big(\sum_{i=1}^{n-1}|\delta^{i+1}(X_{t_i})|\Big)^{p\bar{q}}\Big]^{1/\bar{q}} \ .$$

*Proof.* Let us use the same notations $\beta_r^i, \gamma_r^i$ defined in Proposition 3.3. We also introduce the process $(\overline{\gamma}_r, r \in [0,T])$ by $\overline{\gamma}_r := \sum_{i=1}^n \gamma_r^i \mathbf{1}_{\{t_{i-1}\leq r<t_i\}}$. With $\overline{Z}$ defined by (3.8), it satisfies $|\overline{\gamma}_r| \leq K(1+|Z_r|+|\overline{Z}_r|)$. By Lemma 2.1 and Proposition 3.2, both $Z$ and $\overline{Z}$ are in $\mathcal{H}^2_{BMO}$, and so is $\overline{\gamma}$. In particular, $||\overline{\gamma}||_{\mathcal{H}^2_{BMO}} \leq C$ by some $n$-independent constant.

From the remark following Definition A.2, one can show that $\overline{\gamma} * W \in BMO(\mathbb{P})$. Thus the new probability measure $\mathbb{Q}$ can be defined by $d\mathbb{Q}/d\mathbb{P} = \mathcal{E}_T$, where $\mathcal{E}$ is a Doléan-Dade exponential $\mathcal{E}_t := \mathcal{E}\big(\int_0^T \overline{\gamma}_r^\top dW_r\big)$. The Brownian motion $W^{\mathbb{Q}}$ under the measure $\mathbb{Q}$ is given by $W_t^{\mathbb{Q}} = W_t - \int_0^t \overline{\gamma}_r dr$ for $t \in [0,T]$. Furthermore, it follows from Lemma A.2 that there exists a constant $r^*$ satisfying $1 < r^* < \infty$ such that, for every $1 < \bar{r} \leq r^*$, the reverse Hölder inequality of power $\bar{r}$ holds: $\big(1/\mathcal{E}_\tau\big)\mathbb{E}\big[\mathcal{E}_T^{\bar{r}}|\mathcal{F}_\tau\big]^{1/\bar{r}} \leq C_{\bar{r}}$. Here, $\tau \in \mathcal{T}_0^T$ is an arbitrary $\mathbb{F}$-stopping time, $C_{\bar{r}}$ is some positive constant depending only on $\bar{r}$ and $||\overline{\gamma}||_{\mathcal{H}^2_{BMO}}$. We put $\bar{q} > 1$ as the conjugate exponent of this $\bar{r}$ in the following. By the last observation, all of these constants can be chosen independently of $n$.

Under the new measure $\mathbb{Q}$, the BSDE (3.11) is given by

$$\delta Y_t^i = \delta Y_{t_i}^{i+1} + \delta^{i+1}(X_{t_i}) + \int_t^{t_i} \beta_r^i \delta Y_r^i dr - \int_t^{t_i} \delta Z_r^i dW_r^{\mathbb{Q}} \ ,$$

which can be solved as $\delta Y_t^i = \mathbb{E}^{\mathbb{Q}}\big[e^{\int_t^{t_i} \beta_r^i dr}\big(\delta Y_{t_i}^{i+1} + \delta^{i+1}(X_{t_i})\big)\big|\mathcal{F}_t\big]$. for all $t \in I_i$. Since $|\beta^i| \le K$, one obtains $|\delta Y_t^i| \le e^{Kh_i}\mathbb{E}^{\mathbb{Q}}\big[|\delta Y_{t_i}^{i+1}| + |\delta^{i+1}(X_{t_i})|\big|\mathcal{F}_t\big]$. It then follows by iteration

$$|\delta Y_t^i| \le \mathbb{E}^{\mathbb{Q}}\Big[e^{K\sum_{j=i}^n h_j}|\delta Y_{t_n}^{n+1}| + \sum_{j=i}^n e^{K\sum_{k=i}^j h_k}|\delta^{j+1}(X_{t_j})|\Big|\mathcal{F}_t\Big] \ .$$

Since $\delta Y_{t_n}^{n+1} = \delta^{n+1}(X_{t_n}) = 0$, one concludes $|\delta Y_t^i| \le \mathbb{E}^{\mathbb{Q}}\big[\sum_{j=i}^{n-1} e^{K\sum_{k=i}^j h_k}|\delta^{j+1}(X_{t_j})|\big|\mathcal{F}_t\big]$ for $t \in I_i$, $i \in \{1, \cdots, n\}$. The reverse Hölder inequality gives

$$\begin{aligned} |\delta Y_t^i| &\le e^{KT}\mathbb{E}^{\mathbb{Q}}\Big[\sum_{j=i}^{n-1}|\delta^{j+1}(X_{t_j})|\Big|\mathcal{F}_t\Big] = \frac{e^{KT}}{\mathcal{E}_t}\mathbb{E}\Big[\mathcal{E}_T\sum_{j=i}^{n-1}|\delta^{j+1}(X_{t_j})|\Big|\mathcal{F}_t\Big] \\ &\le C_{\bar{q}}e^{KT}\mathbb{E}\Big[\Big(\sum_{j=i}^{n-1}|\delta^{j+1}(X_{t_j})|\Big)^{\bar{q}}\Big|\mathcal{F}_t\Big]^{1/\bar{q}}, \end{aligned}$$

which then yields

$$\max_{1\le i\le n}\sup_{t\in I_i}|\delta Y_t^i| \le C_{\bar{q}}\sup_{t\in[0,T]}\mathbb{E}\Big[\Big(\sum_{i=1}^{n-1}|\delta^{i+1}(X_{t_i})|\Big)^{\bar{q}}\Big|\mathcal{F}_t\Big]^{1/\bar{q}}.$$

Using Jensen and Doob's maximal inequalities, one finally obtains

$$\begin{aligned} \mathbb{E}\Big[\max_{1\le i\le n}\sup_{t\in I_i}|\delta Y_t^i|^p\Big] &\le C_{p,\bar{q}}\mathbb{E}\Big[\sup_{t\in[0,T]}\mathbb{E}\Big[\Big(\sum_{i=1}^{n-1}|\delta^{i+1}(X_{t_i})|\Big)^{\bar{q}}\Big|\mathcal{F}_t\Big]^{p/\bar{q}}\Big] \\ &\le C_{p,\bar{q}}\mathbb{E}\Big[\sup_{t\in[0,T]}\mathbb{E}\Big[\Big(\sum_{i=1}^{n-1}|\delta^{i+1}(X_{t_i})|\Big)^{\bar{q}}\Big|\mathcal{F}_t\Big]^p\Big]^{1/\bar{q}} \le C_{p,\bar{q}}\mathbb{E}\Big[\Big(\sum_{i=1}^{n-1}|\delta^{i+1}(X_{t_i})|\Big)^{p\bar{q}}\Big]^{1/\bar{q}} \end{aligned}$$

which proves the claim. □

# 4 Connecting the sequence of qg-BSDEs

## 4.1 Short-term expansion of a qg-BSDE

We now give an analytic approximate solution $(\overline{Y}^i, \overline{Z}^i)$ of the BSDE (3.1) as a short-term expansion $(\widehat{Y}^i, \widehat{Z}^i)$. We need two steps involving the linearization method as well as the small-variance expansion method for BSDEs proposed in Fujii & Takahashi (2012) [26] and (2015) [30], respectively. We set aside technical details until Appendices B and C so that we can focus on the main story.

We obtain the approximated solution $(\widehat{Y}^i, \widehat{Z}^i)$ as

$$\widehat{Y}_t^i := \widehat{Y}_t^{i,[0]} + \widehat{Y}_t^{i,[1]}, \quad \widehat{Z}_t^i := \widehat{Z}_t^{i,[0]}, \tag{4.1}$$

for every interval $t \in I_i, i \in \{1, \cdots, n\}$, for which the exact expressions can be read from (C.15), (C.16), (C.17) and (C.18). For numerical implementation, the values at each connecting point $\{t_i\}$ are the most relevant; Under the condition $X_{t_{i-1}} = x, x \in \mathbb{R}^d$, the approximate solutions $\widehat{Y}^{i,t_{i-1},x}_{t_{i-1}} := \widehat{Y}^i_{t_{i-1}}\big|_{X_{t_{i-1}}=x}$, $\widehat{Z}^{i,t_{i-1},x}_{t_{i-1}} := \widehat{Z}^i_{t_{i-1}}\big|_{X_{t_{i-1}}=x}$ are given by the following simple explicit formulas;

$$\begin{aligned}
\widehat{Y}^{i,t_{i-1},x}_{t_{i-1}} &= \overline{y}(t_{i-1},x) + \frac{1}{2}\overline{y}^{[2]}_0(t_{i-1},x) \\
&\quad + h_i f\Big(t_{i-1}, x, \overline{y}(t_{i-1},x) + \frac{1}{2}\overline{y}^{[2]}_0(t_{i-1},x), \overline{\mathbf{y}}^{[1]\top}(t_{i-1},x)\sigma(t_{i-1},x)\Big) \ , \qquad (4.2)\\
\widehat{Z}^{i,t_{i-1},x}_{t_{i-1}} &= \overline{\mathbf{y}}^{[1]\top}(t_{i-1},x)\sigma(t_{i-1},x) \ , \qquad (4.3)
\end{aligned}$$

$$\text{where} \quad \begin{cases}
\overline{\chi}(t_i,x) = x + h_i b\big(t_{i-1},x\big) \ , \\
\overline{y}(t_{i-1},x) = \widehat{u}^{i+1}\big(\overline{\chi}(t_i,x)\big) \ , \\
\overline{\mathbf{y}}^{[1]}(t_{i-1},x) = \Big(\mathbb{I} + h_i\big[\partial_x b(t_i,\overline{\chi}(t_i,x))\big]\Big)\partial_x \widehat{u}^{i+1}(\overline{\chi}(t_i,x)) \ , \\
\overline{y}^{[2]}_0(t_{i-1},x) = h_i \mathrm{Tr}\Big(\partial^2_{x,x}\widehat{u}^{i+1}(\overline{\chi}(t_i,x))\big[\sigma\sigma^\top\big](t_i,\overline{\chi}(t_i,x))\Big) \ ,
\end{cases}$$

where $\mathbb{I}$ denotes $d \times d$-identity matrix. The main result regarding the error estimate for the short-term approximation is given by Theorem C.1. We emphasize that the theorem is interesting in its own sake. It provides the short-term asymptotic expansion of a $qg$-BSDE explicitly in the strong sense.

## 4.2 Connecting procedures

We now connect these approximate solutions by the following scheme.

**Definition 4.1.** *(Connecting Scheme)*
*(i) Setting $\widehat{u}^{n+1}(x) = \xi(x)$, $x \in \mathbb{R}^d$.*
*(ii) Repeating from $i = n$ to $i = 1$ that*
*(a) Calculate the short-term approximation of the BSDE (3.1) by using (4.2) and store the values $\big\{\widehat{Y}^{i,t_{i-1},x'}_{t_{i-1}}\big\}_{x' \in B_i}$ for a finite subset $B_i$ of $\mathbb{R}^d$.*
*(b) Define the terminal function $\widehat{u}^i(x), x \in \mathbb{R}^d$ for the next period $I_{i-1}$ by*

$$\widehat{u}^i(x) := \mathrm{Interpolation}\big(\big\{\widehat{Y}^{i,t_{i-1},x'}_{t_{i-1}}\big\}_{x' \in B_i}\big)(x)$$

*where "Interpolation" stands for some smooth interpolating function satisfying the bounds in Assumption 3.1 (i) .*

From the definition of $\delta^i$ in (3.2), we have

$$\begin{aligned}
&\delta^i(x) = \overline{Y}^{i,t_{i-1},x}_{t_{i-1}} - \widehat{u}^i(x) = \delta^i_{SE}(x) + \mathcal{R}^i(x), \\
\text{where} \qquad &\delta^i_{SE}(x) := \Big(\overline{Y}^{i,t_{i-1},x}_{t_{i-1}} - \widehat{Y}^{i,t_{i-1},x}_{t_{i-1}}\Big) \ , \qquad (4.4)\\
&\mathcal{R}^i(x) := \Big(\widehat{Y}^{i,t_{i-1},x}_{t_{i-1}} - \widehat{u}^i(x)\Big) \ . \qquad (4.5)
\end{aligned}$$

Here, $\delta^i_{SE}$ denotes the error of the short-term approximation (see Theorem C.1), and $\mathcal{R}^i$ the interpolation error as well as the regularization effects rendering the approximated function $\widehat{Y}^{i,t_{i-1},x}_{t_{i-1}}$ into the bounds satisfying Assumption 3.1 (i).

**Remark 4.1.** *Despite the similarity in its appearance, it differs from four-step-scheme (Ma et.al.(1994) [37]), its extensions such as [22, 23], and other PDE discretization approaches. They typically require differentiability in time $t$, the uniform ellipticity of $\sigma\sigma^\top$ and the global Lipschitz continuity of the driver. In particular, we do not know any literature treating Markovian BSDEs with drivers of quadratic growth in the control variables.*

### 4.3 Total error estimate

**Lemma 4.1.** *Under Assumptions 2.1 and 2.2, the solution $Y_t, t \in [0,T]$ of the BSDE (2.3) satisfies the continuity property $\mathbb{E}\big[\sup_{s\le u\le t}\big|Y_u - Y_s\big|^p\big] \le C_p\big|t-s\big|^{p/2}$ for any $0 \le s \le t \le T$ and $p \ge 2$ with some positive constant $C_p$.*

*Proof.* Using the Burkholder-Davis-Gundy inequality and Assumption 2.2 (i), one obtains

$$\begin{aligned}\mathbb{E}\Big[\sup_{s\le u\le t}|Y_u - Y_s|^p\Big] &\le C_p\mathbb{E}\Big[\Big(\int_s^t |f(r,X_r,Y_r,Z_r)|dr\Big)^p + \Big(\int_s^t |Z_r|^2 dr\Big)^{p/2}\Big]\\ &\le C_p\mathbb{E}\Big[\Big(\int_s^t \big[l_r + \beta|Y_r| + \frac{\gamma}{2}|Z_r|^2\big]dr\Big)^p + \Big(\int_s^t |Z_r|^2 dr\Big)^{p/2}\Big] .\end{aligned}$$

Since $l, |Y|$ are bounded and $|Z_t| \le C(1+|X_t|)$ with a constant $C$, the claim is proved. □

Let us now give the main result of the paper:

**Theorem 4.1.** *Define the piecewise constant process $(Y_t^\pi, Z_t^\pi), t \in [0,T]$ by*

$$Y_t^\pi := \widehat{u}^i(X_{t_{i-1}}), \quad Z_t^\pi := \overline{\mathbf{y}}^{[1]\top}(t_{i-1}, X_{t_{i-1}})\sigma(t_{i-1}, X_{t_{i-1}}) ,$$

*for $t_{i-1} \le t < t_i$, $i \in \{1,\cdots,n\}$ and $Y_{t_n}^\pi = \xi(X_{t_n})$, $Z_{t_n}^\pi = 0$, where the $\widehat{u}^i$ and $\overline{\mathbf{y}}^{[1]}$ are those determined by the connecting scheme in Definition 4.1. Then, under Assumptions 2.1, 2.2 and 3.1, there exist some $n$-independent positive constants $\bar{q} > 1$ and $C_{p,\bar{q}}$ such that the following inequality holds for any $p > 1$*

$$\begin{aligned}&\max_{1\le i\le n}\mathbb{E}\left[\big|\big|Y - Y^\pi\big|\big|^{2p}_{[t_{i-1},t_i]}\right]^{\frac{1}{2p}} + \Big(\sum_{i=1}^n \int_{t_{i-1}}^{t_i} \mathbb{E}\Big[\big|Z_t - Z_t^\pi\big|^2\Big]dt\Big)^{1/2}\\ &\le C_{p,\bar{q}}\sqrt{|\pi|} + C_{p,\bar{q}}\sqrt{n}\mathbb{E}\Big[\Big(\sum_{i=1}^n\big|\mathcal{R}^i(X_{t_{i-1}})\big|^2\Big)^{p\bar{q}}\Big]^{\frac{1}{2p\bar{q}}} .\end{aligned}$$

*Proof.* One obtains, by simple manipulation, that

$$\begin{aligned}&\max_{1\le i\le n}\mathbb{E}\left[\big|\big|Y - Y^\pi\big|\big|^{2p}_{[t_{i-1},t_i]}\right]^{1/p} + \sum_{i=1}^n\int_{t_{i-1}}^{t_i}\mathbb{E}\big|Z_t - Z_t^\pi\big|^2 dt\\ &\le C_p\Big(\max_{1\le i\le n}\mathbb{E}\Big[\big|Y_{t_{i-1}} - \overline{Y}^i_{t_{i-1}}\big|^{2p}\Big]^{1/p} + \sum_{i=1}^n\int_{t_{i-1}}^{t_i}\mathbb{E}\big|Z_t - \overline{Z}_t^i\big|^2 dt\Big)\\ &+C_p\Big(\max_{1\le i\le n}\mathbb{E}\Big[\big|\overline{Y}^i_{t_{i-1}} - \widehat{u}^i(X_{t_{i-1}})\big|^{2p}\Big]^{1/p} + \sum_{i=1}^n\int_{t_{i-1}}^{t_i}\mathbb{E}\big|\overline{Z}_t^i - \widehat{Z}_t^i\big|^2 dt\Big)\\ &+C_p\Big(\max_{1\le i\le n}\mathbb{E}\Big[\sup_{t\in I_i}\big|Y_t - Y_{t_{i-1}}\big|^{2p}\Big]^{1/p} + \sum_{i=1}^n\int_{t_{i-1}}^{t_i}\mathbb{E}\big|\widehat{Z}_t^i - \widehat{Z}^i_{t_{i-1}}\big|^2 dt\Big) .\end{aligned}$$

It follows that, by applying Theorems 3.1, C.1, Lemmas 4.1, C.7, and expressions (4.4), (4.5),

$$
\begin{aligned}
&\max_{1\le i\le n}\mathbb{E}\Big[\big|\big|Y-Y^{\pi}\big|\big|^{2p}_{[t_{i-1},t_i]}\Big]^{1/p}+\sum_{i=1}^{n}\int_{t_{i-1}}^{t_i}\mathbb{E}|Z_t-Z_t^{\pi}|^2dt\\
&\le \frac{C_{p,\bar q}}{|\pi|}\mathbb{E}\Big[\Big(\sum_{i=1}^{n-1}|\delta^{i+1}(X_{t_i})|^2\Big)^{p\bar q}\Big]^{\frac{1}{p\bar q}}+C_p\Big(\max_{1\le i\le n}\mathbb{E}\Big[|\delta^i(X_{t_{i-1}})|^{2p}\Big]^{1/p}+\sum_{i=1}^{n}h_i^3\Big)+C_p|\pi|\\
&\le C_p|\pi|+\frac{C_{p,\bar q}}{|\pi|}\Big\{n^{p\bar q-1}\sum_{i=1}^{n}\mathbb{E}\Big[\big|\delta^i_{SE}(X_{t_{i-1}})\big|^{2p\bar q}\Big]+\mathbb{E}\Big[\Big(\sum_{i=1}^{n}|\mathcal{R}^i(X_{t_{i-1}})|^2\Big)^{p\bar q}\Big]\Big\}^{\frac{1}{p\bar q}}\\
&\le C_{p,\bar q}|\pi|+C_{p,\bar q}n\mathbb{E}\Big[\Big(\sum_{i=1}^{n}|\mathcal{R}^i(X_{t_{i-1}})|^2\Big)^{p\bar q}\Big]^{\frac{1}{p\bar q}}\ ,
\end{aligned}
$$

which proves the desired result. □

**Remark 4.2. Comments on the Lipschitz BSDEs and the terminal function**
*The proposed scheme can be equally applicable to the standard Lipschitz BSDEs with a driver $f\in C_b^1$ and also with the terminal function $\xi\in C^3$ of linear growth. Except Proposition 3.2, which is not necessary anymore (and so is Assumption 3.1 (ii)), all the relevant results can be shown with slightly sharper estimates by following the arguments similar to those of Theorem 3.1 in Bouchard & Touzi (2004) [11].*

# 5 An example of implementation

## 5.1 Finite-difference scheme

The remaining problem for us is to find a concrete method of constructing a smooth bounded functions $(\widehat{u}^i)_{1\le i\le n+1}$ used in the step (ii) of Definition 4.1. It is important to notice that there is no need to specify $\widehat{u}^i(x)$ for whole space $x\in\mathbb{R}^d$ but only for those used in the interpolation. Based on this observation, we consider a finite-difference scheme as a method of *non-parametric* coarse graining. Let us suppose every $B_i$ a grid-cube in $\mathbb{R}^d$ centered at the origin, equally spaced by the size $\Delta$. For a function $v:\mathbb{R}^d\to\mathbb{R}$, we denote the 1st-order central difference as $\partial_\Delta v:\mathbb{R}^d\to\mathbb{R}^d$; $(\partial_\Delta v(x))_j:=(v(x+\Delta\ e_j)-v(x-\Delta\ e_j))/(2\Delta),\ j=\{1,\cdots,d\}$ where $e_j$ is the unit vector of direction $j$. Note that, when $v\in C^3$, $|\partial_x v-\partial_\Delta v|\propto\Delta^2$. For higher-order differences, we use a non-central scheme denoted similarly as $\partial_\Delta^m v$, $m=2,\cdots$. Let us arrange a sequence of grids $B_i\subseteq B_{i+1}$ satisfying that

$$
M/2\le|B_1|\le|B_{n+1}|\le M, \tag{5.1}
$$

where $M$ is some sufficiently large constant and $|B_i|$ the length of the cube's edge. Let us write $x\in\overline{B_i}$ when $x$ is inside the cube but not necessary on the grid point and denote the boundary of the cube by $\partial\overline{B_i}$.

**Lemma 5.1.** *Suppose $\widehat{u}^i\in C_b^\infty\cap\mathbb{L}^\infty$ and $dist(y,\partial\overline{B_i})\le C'\Delta$ with some positive constant $C'$. Then, denoting $x\in B_i$ as the nearest grid point to $y$, there exists a constant $C$ such that*

$$
\begin{aligned}
&\big|\widehat{u}^i(y)-[\widehat{u}^i(y)]\big|\le C\Delta^3,\ \big|\partial_x\widehat{u}^i(y)-[\partial_x\widehat{u}^i(y)]\big|\le C\Delta^2,\\
&\big|\partial_x^2\widehat{u}^i(y)-[\partial_x^2\widehat{u}^i(y)]\big|\le C\Delta\ ,
\end{aligned}
$$

$$\textit{where} \quad \begin{cases} [\widehat{u}^i(y)] := \widehat{u}^i(x) + \partial_x \widehat{u}^i(x) \cdot (y-x) + \frac{(y-x)^\top}{2} \partial_x^2 \widehat{u}^i(x) \cdot (y-x) \\ [\partial_x \widehat{u}^i(y)] := \partial_x \widehat{u}^i(x) + \partial_x^2 \widehat{u}^i(x) \cdot (y-x) \\ [\partial_x^2 \widehat{u}^i(y)] := \partial_x^2 \widehat{u}^i(x) \end{cases} .$$

*Proof.* It is obvious from the Taylor formula. □

Let us scale the size of the grid-cube according to the number of discretization as

$$M = Cn^{\delta/2} \tag{5.2}$$

with a given constant $C$ and an arbitrary $\delta \in (0, 1/2)$. In this case, the quadratic-growth term $h_i|x|^2$ existing in $\widehat{Y}^{i,t_{i-1},x}_{t_{i-1}}$ of (4.2) is bounded by $Cn^{-1+\delta}$ within the cube. We now modify the connecting scheme of Definition 4.1 as follows.

**Definition 5.1.** *(Connecting scheme with finite-difference)*
*(0) Set a scaling rule $\Delta = \zeta|\pi|^\nu$ with constants $\zeta, \nu > 0$, and then construct a sequence of grid-cubes $B_1 \subseteq B_2 \subseteq \cdots \subseteq B_{n+1} \subseteq \mathbb{R}^d$ satisfying (5.1) and (5.2).*
*(i) Suppose $\widehat{u}^{i+1}$ is in the class $C_b^\infty \cap \mathbb{L}^\infty$ and the values of $\{\widehat{u}^{i+1}(x), \partial_x \widehat{u}^{i+1}(x), \partial_x^2 \widehat{u}^{i+1}(x) \mid x \in B_{i+1}\}$ are known.* [6]
*(ii) Store the values of $\{[\widehat{Y}^{i,t_{i-1},x}_{t_{i-1}}], x \in B_i\}$ where $[\widehat{Y}^{i,t_{i-1},x}_{t_{i-1}}]$ is equal to $\widehat{Y}^{i,t_{i-1},x}_{t_{i-1}}$ calculated by (4.2) with $\widehat{u}^{i+1}(\overline{\chi}(t_i, x))$ and its derivatives replaced by their approximations as in Lemma 5.1* [7].
*(iii) Store the values of $\{\partial_\Delta[\widehat{Y}^{i,t_{i-1},x}_{t_{i-1}}], \partial_\Delta^2[\widehat{Y}^{i,t_{i-1},x}_{t_{i-1}}] \mid x \in B_i\}$.* [8]
*(iv) Consider $\widehat{u}^i$ as an appropriate $C_b^\infty \cap \mathbb{L}^\infty$-class function (see Remark 5.1) satisfying*

$$\Big(\widehat{u}^i(x), \partial_x \widehat{u}^i(x), \partial_x^2 \widehat{u}^i(x)\Big) \simeq \Big([\widehat{Y}^{i,t_{i-1},x}_{t_{i-1}}], \partial_\Delta[\widehat{Y}^{i,t_{i-1},x}_{t_{i-1}}], \partial_\Delta^2[\widehat{Y}^{i,t_{i-1},x}_{t_{i-1}}]\Big), \quad x \in B_i \ .$$

*where $\simeq$ is the approximate equality with the size of error bounded respectively by $(C\Delta^3, C\Delta^2, C\Delta)$ with some constant $C$, and smoothly tracks $\overline{Y}^{i,t_{i-1},x}_{t_{i-1}}$ outside the grid-cube while keeping the same order of accuracy as inside i.e., $\sup_{x \notin \overline{B_i}} |\overline{Y}^{i,t_{i-1},x}_{t_{i-1}} - \widehat{u}^i(x)| \le C \sup_{x \in \overline{B_i}} |\overline{Y}^{i,t_{i-1},x}_{t_{i-1}} - \widehat{u}^i(x)|$.*

**Remark 5.1.** *The existence of functions $\widehat{u}^i \in C_b^\infty \cap \mathbb{L}^\infty$ satisfying (iv) of the above scheme can be easily seen. Suppose $v : \mathbb{R}^d \to \mathbb{R}$ is an arbitrary smooth and bounded function satisfying $\max_{x \in B_i} |v(x) - [\widehat{Y}^{i,t_{i-1},x}_{t_{i-1}}]| \le C\Delta^3$. Then, by construction, $(\partial_\Delta v, \partial_\Delta^2 v)$ is equal to $(\partial_\Delta[\widehat{Y}^{i,t_{i-1},x}_{t_{i-1}}], \partial_\Delta^2[\widehat{Y}^{i,t_{i-1},x}_{t_{i-1}}])$ with the desired accuracy. Since $(\partial_\Delta v, \partial_\Delta^2 v)$ is equal to the true derivatives with the accuracy $(C\Delta^2, C\Delta)$ with $C = ||\partial_x^3 v||_\infty$, $v$ is in fact a valid candidate for $\widehat{u}^i$. Although there is no need to single out $\widehat{u}^i$, one can just suppose that a function with the smallest total variation is chosen among the candidates in order to avoid unnecessary oscillations between the neighboring grid points. Adjusting $v$ outside the cube so that it smoothly tracks $\overline{Y}^{i,t_{i-1},x}_{t_{i-1}}$ is always possible.*

## 5.2 Error estimate

Unfortunately, we cannot prove Assumption 3.1 for the above scheme in the limit $n \to \infty$ because of the non-linearity from the driver. Therefore, in the following analysis, we are

[6] In this scheme, $\widehat{u}^{n+1} \in C_b^\infty$ is a bounded function constructed as in step (iv) using $\widehat{Y}^{n+1,t_n,x}_{t_n} = \xi(x)$.

[7] If $b$ is $x$-independent, the adjustment of Lemma 5.1 is unnecessary by shifting the center of the grids. If $b$ is proportional to $x$, *log*-transformation may be used to make the drift $x$-independent once again.

[8] If $x$ is at the edge of the grid $B_i$, apply a non-central difference scheme.

forced to restrict out attention to the approximation within a finite range of $n$. In this case, Assumption 3.1 becomes trivial by simply taking the maximum for a given range and all the estimates including Theorem 4.1 can be used. However, this does not tell if the approximation is improved by using a finer discretization. We need to confirm that the bounds of Assumption 3.1 remain stable within the relevant range of $n$.

**Assumption 5.1.** *(i) (A posteriori check) The values*

$$\max_{1\le i\le n}\max_{x\in B_i}\Big\{|\widehat{u}^i(x)|, |\partial_x\widehat{u}^i(x)|, |\partial_x^2\widehat{u}^i(x)|, |\partial_x^3\widehat{u}^i(x)|\Big\}$$

*are confirmed, a posteriori, to be stable for a given range of discretization, say, $n\in[n_0,n_1]$.*[9] *(ii) Under the above condition, the same uniform boundedness can be maintained just slightly outside of each grid-cube $B_i$; for $x\in\mathbb{R}^d$ satisfying $dist(x,\partial\overline{B_i})\le|B_i|(\lambda\eta_i\sqrt{h_i})$, where $\lambda>1$ is some constant and $\eta_i$ the maximum size of the log-volatility of $X$ $\big($i.e., $||\sigma(\cdot,x)||_{[t_{i-1},t_i]}/|x|\big)$ for $x$ near the boundary $\partial\overline{B_i}$.*

**Remark 5.2.** *One easily sees that if $\Delta\downarrow 0$ faster than $|\pi|$, the higher-order derivatives goes to diverge due to the fact that $f\in C^1$. On the other hand, we need at least $\nu>1/3$ for the sum of $C\Delta^3$ to converge. Interestingly, for numerical example given in the next section, we find that the stability (i) is achieved with $\nu=1/2$, i.e., the scaling $\Delta=\zeta|\pi|^{1/2}$ with the coefficient $\zeta$ of the order of $X$'s volatility. The scaling rule also suggests a connection to the stability problem of parabolic PDEs with explicit finite-difference scheme, but we need further research to understand whether proving Assumption 3.1 is possible or not for the current scheme in the limit $n\to\infty$ with some $\nu>1/3$.*

**Remark 5.3.** *Note that the condition (ii) assures the uniform boundedness of $|\partial_x^m\widehat{u}^i(\cdot)|_{0\le m\le 3}$ for the $(\lambda)$-sigma range of the path $(X_t, t\in I_i)$ provided $X_{t_{i-1}}\in\overline{B_i}$. Now consider the conditional version of Theorem C.1 for the period $I_i=[t_{i-i},t_i]$ given $X_{t_{i-1}}\in\overline{B_i}$. Under Assumption 5.1, the contribution from the path $(X_t, t\in I_i)$ outside this $(\lambda)$-sigma range is suppressed by an arbitrary power of $\lambda$. Thus, one can take $n$-independent constant $C$ such that*

$$|\overline{Y}^{i,t_{i-1},x}_{t_{i-1}}-\widehat{Y}^{i,t_{i-1},x}_{t_{i-1}}|\le C|B_i|^2h_i^{3/2},\quad \forall x\in\overline{B_i}$$

*at least for a finite range of $n$. The factor of $|B_i|^2$ appears due to the quadratic growth property of $\widehat{Y}^{i,t_{i-1},x}_{t_{i-1}}$ in $x$.*[10] *In practice, $\lambda\simeq 5$ would be large enough and hence in most of the natural settings for financial problems, $\lambda\eta_i\simeq 1$ is expected to be good for the purpose.*

**Proposition 5.1.** *Suppose Assumptions 2.1, 2.2 and also Assumption 5.1 with the scaling rule $\Delta=\zeta|\pi|^{1/2}$, then there exists a constant $C$ such that $\sum_{i=1}^n||\delta^{i+1}||_{\mathbb{L}^\infty}\le C$ uniformly for $n\in[n_0,n_1]$. In particular $(||\widehat{u}^i||_{\mathbb{L}^\infty})_{1\le i\le n}$ are also uniformly bounded.*

*Proof.* By definition, we have

$$\sum_{i=1}^n||\delta^{i+1}||_{\mathbb{L}^\infty}\le\sum_{i=2}^{n+1}\Big\{\sup_{x\in\overline{B_i}}\big(|\overline{Y}^{i,t_{i-1},x}_{t_{i-1}}-\widehat{Y}^{i,t_{i-1},x}_{t_{i-1}}|+|\widehat{Y}^{i,t_{i-1},x}_{t_{i-1}}-\widehat{u}^i(x)|\big)+\sup_{x\notin\overline{B_i}}|\overline{Y}^{i,t_{i-1},x}_{t_{i-1}}-\widehat{u}^i(x)|\Big\}.$$

[9] Note that the first three quantities are already used for the connecting scheme itself. Unless $(\partial_x^2\widehat{u})$ rapidly oscillate, the direct evaluation of $(\partial_x^3\widehat{u})$ may be skipped.

[10] The condition (ii) is relevant only for $x$ near the boundary of the grid-cube.

By Remark 5.3, one can take a version satisfying $\sup_{x\in\overline{B}_i}|\overline{Y}^{i,t_{i-1},x}_{t_{i-1}}-\widehat{Y}^{i,t_{i-1},x}_{t_{i-1}}|\le C|M|^2|\pi|^{3/2}\le Cn^{\delta-3/2}$ with some constant $C$ uniformly. Next, by construction, $|\widehat{Y}^{i,t_{i-1},x}_{t_{i-1}}-\widehat{u}^i(x)|\le C\Delta^3$, $|\partial_\Delta\widehat{Y}^{i,t_{i-1},x}_{t_{i-1}}-\partial_x\widehat{u}^i(x)|\le C\Delta^2$ and $|\partial^2_\Delta\widehat{Y}^{i,t_{i-1},x}_{t_{i-1}}-\partial^2_x\widehat{u}^i(x)|\le C\Delta$ at each grid point $x\in B_i$. The function $\widehat{u}^i$ within the grid-cube $\overline{B_i}$ can be separated into two parts; one is the regularization of the first term of (4.2), and the other is the remaining two terms proportional to $h_i$. Since the first term of (4.2) is confirmed to have bounded derivatives up to the third order, the difference of the second (first) order derivatives is bounded by $C\Delta$ ($C\Delta^2$) between any interval of the grid points. For the latter, since they are in class-$C^1$, proportional to $h$ and with at most quadratic growth in $x$, the difference of the first order derivatives is bounded by $CM^2h$. Combining these two, one sees that $|\widehat{Y}^{i,t_{i-1},x}_{t_{i-1}}-\widehat{u}^i(x)|\le C\Delta^3+C(\Delta^2+M^2h)\Delta\le Cn^{\delta-3/2}$ for whole interval between the two neighboring grid points.[11]

By repeating the same arguments, one sees $\max_{x\in\overline{B_i}}|\widehat{Y}^{i,t_{i-1},x}_{t_{i-1}}-\widehat{u}^i(x)|\le Cn^{\delta-3/2}$. Since $\widehat{u}^i(x)$ outside the cube is constructed to follow $\overline{Y}^{i,t_{i-1},x}_{t_{i-1}}\in C^1_b\cap\mathbb{L}^\infty$ so that it keeps the same order of accuracy, one also has $\sup_{x\notin\overline{B_i}}|\overline{Y}^{i,t_{i-1},x}_{t_{i-1}}-\widehat{u}^i(x)|\le Cn^{\delta-3/2}$. Thus $||\delta^i||_{\mathbb{L}^\infty}\le Cn^{\delta-3/2}$ and the first claim is proved. Using the universal bound iteratively, one also has

$$||\widehat{u}^i||_{\mathbb{L}^\infty}\quad\le\quad e^{\beta|\pi|}\Big(||\widehat{u}^{i+1}||_{\mathbb{L}^\infty}+|\pi|||l||_T\Big)+Cn^{\delta-3/2}\le e^{\beta T}\Big(||\xi||_{\mathbb{L}^\infty}+T\big(||l||_T+Cn^{\delta-\frac12}\big)\Big)$$

which yields the second. □

**Proposition 5.2.** *Suppose Assumptions 2.1, 2.2 and also Assumption 5.1 with the scaling rule $\Delta=\zeta|\pi|^{1/2}$. Then there exist some constant $C$ satisfying that*

$$\mathbb{E}\Big[\Big(\sum_{i=1}^{n+1}|\mathcal{R}^i(X_{t_{i-1}})|^2\Big)^p\Big]^{\frac{1}{2p}}\le Cn^{-1+\delta}\tag{5.3}$$

*uniformly for $n\in[n_0,n_1]$, for every $p>1$.*

*Proof.* Using $\max_{x\in\overline{B_i}}|\widehat{Y}^{i,t_{i-1},x}_{t_{i-1}}-\widehat{u}^i(x)|\le Cn^{\delta-3/2}$, $||\widehat{u}^i||_{\mathbb{L}^\infty}\le C$ shown in Proposition 5.1 and the quadratic-growth property of $\widehat{Y}^{i,t_{i-1},x}_{t_{i-1}}$ in $x$, one obtains that

$$\begin{aligned}&\mathbb{E}\Big[\Big(\sum_{i=1}^{n+1}|\mathcal{R}^i(X_{t_{i-1}})|^2\Big)^p\Big]^{\frac{1}{2p}}\le|n+1|^{\frac12}\max_{1\le i\le n+1}\mathbb{E}\Big[\Big|\widehat{Y}^{i,t_{i-1},X_{t_{i-1}}}_{t_{i-1}}-\widehat{u}^i(X_{t_{i-1}})\Big|^{2p}\Big]^{\frac{1}{2p}}\\&\le Cn^{-1+\delta}+C\sqrt{n}\max_{1\le i\le n+1}\mathbb{E}\Big[(1+h_i||X||^2_{I_i})^{2p}\Big(\frac{||X||_{I_i}}{M/2}\Big)^{4pk}\Big]^{\frac{1}{2p}}\\&\le Cn^{-1+\delta}+C_{p,k}n^{-k\delta+1/2}\quad\text{ for any }\quad k>1.\end{aligned}$$

Since $k>1$ is arbitrary, one obtains the desired result. □

Let us define the approximate piecewise constant processes $(Y^{FD}_t,Z^{FD}_t),t\in[0,T]$ by using the bounded function $(\widehat{u}^i)_{1\le i\le n+1}$ constructed as in Definition 5.1 as

$$Y^{FD}_t:=\widehat{u}^i(x),\quad Z^{FD}_t:=\overline{\mathbf{y}}^{[1]\top}(t_{i-1},x)\sigma(t_{i-1},x)$$

[11] Although one may only have $|\partial_\Delta\widehat{Y}^{i,t_{i-1},x}_{t_{i-1}}-\partial_x\widehat{u}^i(x)|\le C\Delta$ when $x\in\partial\overline{B_i}$ and the direction of the derivative is orthogonal to the boundary (since the central difference cannot be taken for $\partial_\Delta[\widehat{Y}^{i,t_{i-1},x}_{t_{i-1}}]$ in the step (iii) of Definition 5.1), one obtains the same conclusion by estimating from the neighboring internal point.

for $t \in [t_{i-1}, t_i), i \in \{1, \cdots, n+1\}$. $x$ is a $B_i$-valued $\mathcal{F}_{t_{i-1}}$-measurable r.v. nearest to $X_{t_{i-1}}$.

**Corollary 5.1.** *Suppose Assumptions 2.1, 2.2 and also Assumption 5.1 with the scaling rule $\Delta = \zeta|\pi|^{1/2}$. Then there exist positive constants $\bar{q} > 1$ and $C_{p,\bar{q}}$ such that*

$$\max_{1\le i\le n} \mathbb{E}\Big[||Y - Y^{FD}||^{2p}_{[t_{i-1},t_i]}\Big]^{\frac{1}{2p}} + \Big(\sum_{i=1}^{n}\int_{t_{i-1}}^{t_i} \mathbb{E}\Big[|Z_t - Z_t^{FD}|^2\Big]dt\Big)^{1/2} \le C_{p,\bar{q}} n^{-1/2+\delta}$$

*for $\forall p > 1$ uniformly for $n \in [n_0, n_1]$.*

*Proof.* For $t \in I_i$, one has

$$\begin{aligned} |Y_t^\pi - Y_t^{FD}| &\le |\widehat{u}^i(X_{t_{i-1}}) - \widehat{u}^i(x)|\big(\mathbf{1}_{\{X_{t_{i-1}}\in\overline{B}_i\}} + \mathbf{1}_{\{X_{t_{i-1}}\notin\overline{B}_i\}}\big) \\ &\le C\Delta + 2||\widehat{u}^i||_{\mathbb{L}^\infty}\mathbf{1}_{\{X_{t_{i-1}}\notin\overline{B}_i\}} \end{aligned}$$

with $x \in B_i$ is the nearest grid point to $X_{t_{i-1}}$, and thus $\mathbb{E}\big[||Y^\pi - Y^{FD}||^{2p}_{I_i}\big]^{\frac{1}{2p}} \le C\sqrt{|\pi|} + C_{p,k}\Big(\frac{1+|x_0|}{M}\Big)^k$ for $\forall k > 1$. Similar estimate also holds for $Z^\pi - Z^{FD}$. The conclusion follows from Propositions 5.1, 5.2 and Theorem 4.1. Note that the contribution from $||X||_T > M$ in the (non-conditional) short-term expansion used in Theorem 4.1 is also suppressed by an arbitrary power of $M$. □

**Remark 5.4.** *In the case of the standard Lipschitz BSDEs, $\widehat{u}^i$ is not bounded and has the linear-growth property. We can still use the same scheme since the errors from $|x| > M$ is still suppressed by an arbitrary power of $M$.*

### 5.3 Application of sparse grids

In order to mitigate the so called *curse of dimensionality*, there exists a very interesting result on high dimensional polynomial interpolation using sparse grids. By Theorem 8 (as well as Remark 9) of Barthelmann et al. [3], it is known that there exists an interpolating function satisfying the following uniform estimates on the compact set for a function $f : \mathbb{R}^d \to \mathbb{R}$ in the class $C^k$;

$$\sup_{|x|\le M}\Big|f(x) - \mathcal{A}^{q,d}\big(f(x)\big)\Big| \le C_{q,d} N_{(q,d)}^{-k}(\log(N_{(q,d)}))^{(k+1)(d-1)} \ . \tag{5.4}$$

Here, $\mathcal{A}^{q,d}(f) : \mathbb{R}^d \to \mathbb{R}$ is an interpolating polynomial function of degree $q$ ($\ge d$) based on the Smolyak algorithm. The interpolating function is uniquely determined by the values of $f(x_i), x_i \in H(q,d)$ where $H(q,d)$ is the sparse grid whose number of nodes is give by $N_{(q,d)}$. $C_{q,d}$ is some positive constant depending only on $(q,d)$ and $\sup_{|x|\le M}|\partial_x^m f|$ of $m = \{0, \cdots, k\}$. The sparse grid $H(q,d)$ is the set of points on which the Chebyshev polynomials take the extrema. For details, see [3, 44] and references therein. The sparse grid method looks very attractive since (5.4) has only weak log dependency on the dimension $d$.

For our purpose, we want to interpolate $f(x) = \widehat{Y}_{t_{i-1}}^{i,t_{i-1},x}$ of (4.2) efficiently so that the right-hand side of (5.4) is of the order of $C\Delta^3$. Consider the interpolation for the first and the remaining two terms of (4.2), separately. The former has $k = 3$, and the latter has only $k = 1$ but it is proportional to $h_i \propto \Delta^2$. Therefore, for interpolation of $\widehat{Y}_{t_{i-1}}^{i,t_{i-1},x}$, the number of nodes $N_{(q,d)} \propto \Delta^\varepsilon$ ($\propto n^{\varepsilon/2}$), $\varepsilon > 1$ can maintain the same error estimate of Corollary 5.1.

# 6 Numerical examples

In the remainder of the paper, we demonstrate our computation scheme and its empirical convergence rate using illustrative models. For simplicity, we use a full grid (instead of a sparse grid) at each time step with the scaling rule $\Delta = \zeta|\pi|^{1/2}$. As existing literature, we focus on approximating the initial value of the BSDE ($Y_0 = Y_0^{0,x_0}$) and thus restrict the computations only to the relevant grid points. More extensive tests on higher dimensional setups with sparse grids will be left for the future works[12].

## 6.1 A solvable qg-BSDE

Let us first consider the following model with $d = 2$ similar to those studied in [15]:

$$X_t = x_0 + \int_0^t \begin{pmatrix} b_1 X_s^1 \\ b_2 X_s^2 \end{pmatrix} ds + \int_0^t \begin{pmatrix} \sigma_1 X_s^1 & 0 \\ 0 & \sigma_2 X_s^2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{pmatrix} dW_s \ , \tag{6.1}$$

$$Y_t = \xi(X_T) + \int_t^T \frac{a}{2}|Z_s|^2 ds - \int_t^T Z_s dW_s \ , \tag{6.2}$$

where $b_i, \sigma_i$, $i \in \{1,2\}$, $\rho \in [-1,1]$ and $a$ are all constants. For this example, by using a exponential transformation $\big(e^{aY_t}, t \in [0,T]\big)$, we obtain a closed form solution:

$$Y_t = \frac{1}{a} \log\big(\mathbb{E}\big[\exp\big(a\xi(X_T)\big)\big|\mathcal{F}_t\big]\big) \ , \tag{6.3}$$

whose expectation can be obtained semi-analytically by integrating over the density of $X$. We use

$$\xi(x) = 3\big(\sin^2(x^{\mathbf{1}}) + \sin^2(x^{\mathbf{2}})\big) \tag{6.4}$$

as the terminal value function, and set $x_0 = (1,1)^\top$, $T = 1$, $b_1 = b_2 = 0.05$, $\rho = 0.3$.

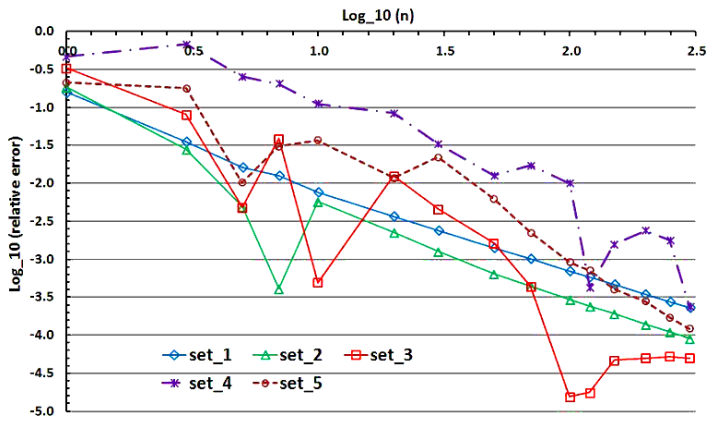


Figure 1: Empirical convergence of the proposed scheme for (6.2) with $\text{set}_i, i \in \{1, \cdots, 5\}$.

We have tested the following five sets of parameters $(\sigma_{i=1,2}, a)$:

$$\text{set}_1 = \big\{\sigma_i = 0.5, a = 1.0\big\},\ \text{set}_2 = \big\{\sigma_i = 0.5, a = 2.0\big\},\ \text{set}_3 = \big\{\sigma_i = 0.5, a = 3.0\big\},$$
$$\text{set}_4 = \big\{\sigma_i = 1.0, a = 3.0\big\},\ \text{set}_5 = \big\{\sigma_i = 0.5, a = 4.0\big\} \tag{6.5}$$

[12] For example, see [40, 49].

by changing the number of partitions from $n = 1$ to $n = 300$. In Figure 1, we have plotted $\log_{10}$(relative error) against the $\log_{10}(n)$ for $\text{set}_i, i \in \{1, \cdots, 5\}$, where the relative error is defined by

$$\frac{\text{estimated } Y_0 \text{ by the proposed scheme} - \text{the value obtained from (6.3)}}{\text{the value obtained from (6.3)}} .$$

As naturally expected, the bigger "$a$" we use, the bigger $\zeta$ is needed to keep the derivatives (calculated by finite-difference scheme) non-divergent as Assumption 5.1 requires.

### Driver truncation

As we have emphasized, it is crucial to have stable derivatives as Assumption 5.1 for the proposed scheme to converge. For the qg-BSDE (6.2), if we increase the coefficient "$a$" while keeping the factor $\zeta$ of $\Delta = \zeta|\pi|^{1/2}$ constant, we have observed that these derivatives (and hence the estimate of $Y$) are, in fact, divergent. In the remainder, instead of making $\zeta$ larger, let us study the truncation of the driver $f$ so that it has a global Lipschitz constant $N$ following the scaling rule ( see Section 2.1 of [15] )

$$N \propto n^{\alpha}, \quad 0 < \alpha < 1 . \tag{6.6}$$

The error estimates for the qg-BSDEs under this truncation have been studied by Imkeller & Reis (2010) [32] (Theorem 6.2) and applied to the backward numerical scheme by Chassagneux & Richou (2016) [15]. From Theorem 6.2 [32], one easily sees that this truncation does not affect the theoretical bound on the convergence rate of Theorem 4.1, which is also the case for the scheme studied in [15].

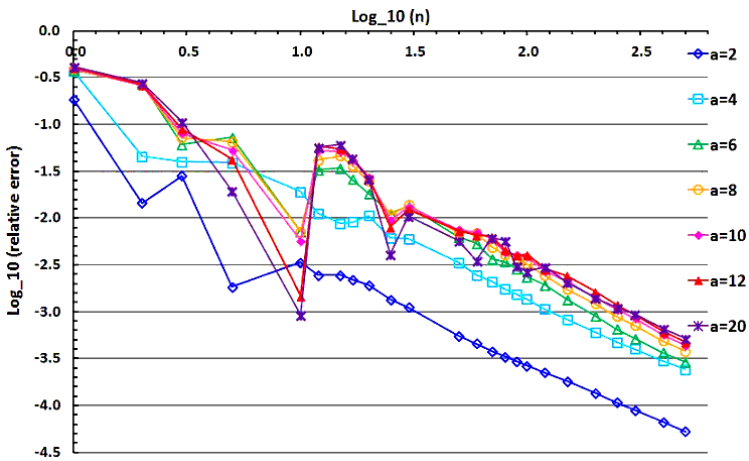


Figure 2: Empirical convergence of the proposed scheme for (6.2) with a truncated driver so that the Lipschitz constant scales as $N \propto n^{1/3}$.

We have chosen the constant $\zeta$ so that it marginally works for the $\text{set}_3$ in (6.5) without any truncation and adopted the scaling factor $\alpha = 1/3$. We tested the following seven cases of large quadratic coefficients;

$$\{a = 2,\ a = 4,\ a = 6,\ a = 8,\ a = 10,\ a = 12,\ a = 20\} \tag{6.7}$$

while keeping the other parameters the same, i.e.,

$$x_0 = (1,1)^\top,\ T = 1,\ b_1 = b_2 = 0.05,\ \rho = 0.3,\ \sigma_1 = \sigma_2 = 0.5\ . \tag{6.8}$$

In Figure 2, we have plotted the $\log_{10}$(relative error) against the $\log_{10}(n)$ changing the number of partitions from $n = 1$ to $n = 500$. Except for coarse partitions $n \lesssim 10$, the truncation of the driver yields quite stable convergence even for very large quadratic coefficients. We find no significant change in the empirical convergence rate, and it is close to one. The introduction of the truncation (6.6) looks quite attractive since there is no need to adjust $\zeta$ according to different size of the coefficient $a$. There seems a deep relation among the stability (Assumption 5.1), the scaling rule of finite difference scheme as well as the truncation of the driver ($N \propto n^\alpha$). This interesting problem requires further research.

### Non-differentiable terminal function

We now test a case of non-differentiable terminal function

$$\xi(x) = \min\big(\max(x^{\mathbf{1}}, 1), 3\big) + \max(2 - x^{\mathbf{2}}, 0) \tag{6.9}$$

in (6.2). If we apply the finite-difference scheme given in the last section directly, the 2nd and 3rd-order derivatives appearing in Assumption 5.1 grow with the rate of $1/\Delta$ and $1/\Delta^2$, at least, near the boundary $t = T$. Thus one naturally expects some instability appears if the space discretization $\Delta = \zeta|\pi|^{1/2}$ becomes sufficiently small. Note that, one can always apply the technique of Section 5 as long as an appropriate mollified function is chosen and kept fixed. In this case, however, the total error of Corollary 5.1 contains, of course, an additional term arising from the mollification.

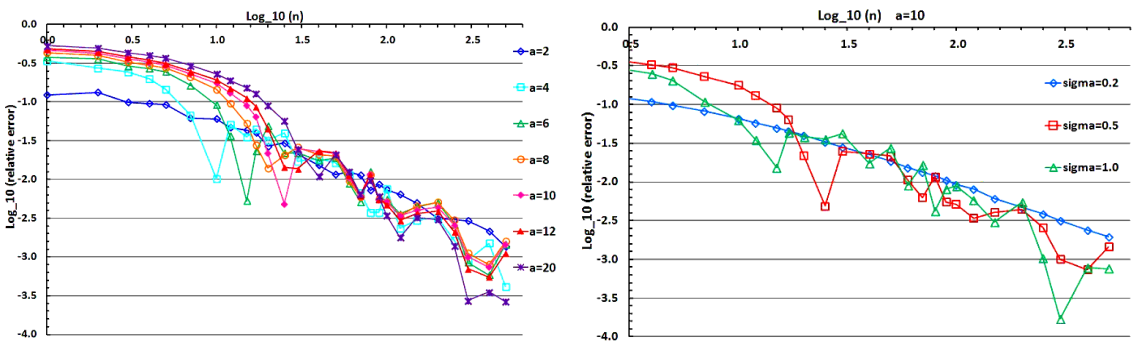


Figure 3: Empirical convergence of the proposed scheme for (6.2) with a non-differentiable terminal function (6.9). The left one uses (6.8) with $a$'s in (6.7), and the right one uses $a = 10$ and (6.8) but with difference choices of volatilities $(\sigma_i)_{1\le i\le 2} = 0.2,\ 0.5,\ 1.0$.

In Figure 3, we have tested the model (6.2) with the terminal function (6.9) and plotted the $\log_{10}$(relative error) for the number of partitions from $n = 1$ to $n = 500$ by directly applying the finite-difference scheme. The same truncation of the driver $N \propto n^{1/3}$ has been used as in the previous example. In the *left* figure, we have tested the same seven cases of "$a$" (6.7) with the same set of parameters (6.8). In the *right* one, we have kept $a = 10$ but tested (6.8) replaced by three different choices of volatilities $(\sigma_i)_{1\le i\le 2} = 0.2,\ 0.5,\ 1.0$. The results are very encouraging. From the left one, although one actually observes some instability, the overall rate of convergence is not much different from the previous example in Fig 2 of a differentiable terminal function. From the right one, one observes that the size of volatilities

does not meaningfully affect the empirical convergence rate.

In the computation, we observed that the maximum of the derivatives decays rather quickly when $t$ is away from the maturity. In fact, from the expression (6.3) and the *integration-by-parts* formula, one can show that the solution $Y_t^{t,\cdot} : \mathbb{R}^d \to \mathbb{R}$ is a smooth function of $x$ for $t < T$ as long as $X_T$ has a smooth density, which is the case for the current log-normal model (6.1) for $(X_t)_{t\in[0,T]}$. The above numerical result (and also the examples of the next subsection) suggests that the relaxing the conditions in Assumptions 2.2, 3.1 (and hence 5.1) may be possible under appropriate conditions. Further studies on the regularity of the true solution as well as its approximation are needed.

### 6.2 Lipschitz BSDEs: Option pricing with different interest rates

In the reminder, let us test the proposed scheme for a Lipschitz BSDE for completeness. We use the same scaling rule $\Delta = \zeta|\pi|^{1/2}$ but, of course, no truncation of the driver. We consider a very popular valuation problem of European options under two different interest rates, $r$ for investing and $R$ ($\neq r$) for borrowing. Since this problem has been often used for testing the numerical schemes for Lipschitz BSDEs, it would be informative to compare the current scheme to the existing numerical examples based on the Monte-Carlo simulation.

Let us assume the dynamics of the security price as

$$X_t = x_0 + \int_0^t \mu X_s ds + \int_0^t \sigma X_s dW_s \ ,$$

where $d = 1$ and $\mu, \sigma$ are positive constants. For the option payoff $\Phi(X_T)$ at the expiry $T$, the option price $Y_t$ implied by the self-financing replication is given by

$$Y_t = \Phi(X_T) - \int_t^T \Big\{ rY_s + \frac{\mu - r}{\sigma} Z_s - \Big( Y_s - \frac{Z_s}{\sigma} \Big)^- (R - r) \Big\} ds - \int_t^T Z_s dW_s \ . \tag{6.10}$$

Although both of the terminal and driver functions are not smooth, we can expect rather accurate results considering the result in Fig 3 for the qg-BSDE [13].

Firstly, we study the cases where the payoff function is equal to that of a call option: $\Phi(x) = (x - K)^+$, where $K > 0$ is the strike price. As suggested by [31], this example provides a very interesting test since the price must be exactly equal to that of Black-Scholes model with interest rate $R$. This is because the replicating portfolio consists of the long-only position and hence the investor must always borrow money to fund her position. We have chosen the common parameters as $\{T = 1, r = 0.01,\ R = 0.06,\ \mu = 0.06,\ X_0 = 100\}$ and tested the following five sets of $(K, \sigma)$ [14] with $n = 10$ to $n = 3000$ in Figure 4:

$$\text{set}_1 = \{K = 106,\ \sigma = 0.3\}\ , \text{set}_2 = \{K = 166,\ \sigma = 0.3\}\ , \text{set}_3 = \{K = 106,\ \sigma = 1.0\}\ ,$$
$$\text{set}_4 = \{K = 306,\ \sigma = 1.0\}\ , \text{set}_5 = \{K = 106,\ \sigma = 2.0\}.$$

The Black-Scholes price for each set is given by BS $= \{12.000, 1.117, 38.346, 11.662, 68.296\}$ respectively. Although the relative errors for OTM options are slightly higher, the convergence rate to the exact BS prices is close to 1 for every case. It is a bit striking that we do

[13] Although we tested the same model with mollified functions, we found no meaningful difference in the empirical convergence rate.

[14] $K = 106$ is close to *at the money forward* for $T = 1$ with 6% interest rate. The bigger strikes correspond to $2\sigma$ *out of the money*.

not see any deterioration in convergence rate in spite of the non-smooth functions and rather high volatilities. The observed irregularity of the error size is likely due to the change of the configuration of the grids close to the terminal time relative to the non-differentiable points of the terminal function.

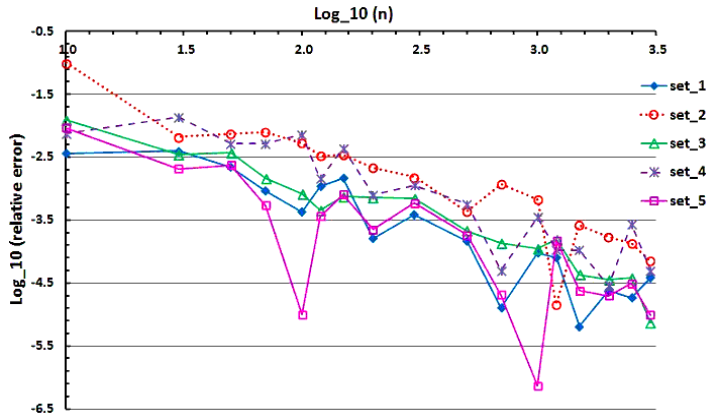


Figure 4: Empirical convergence of the proposed scheme for (6.10) for call options.

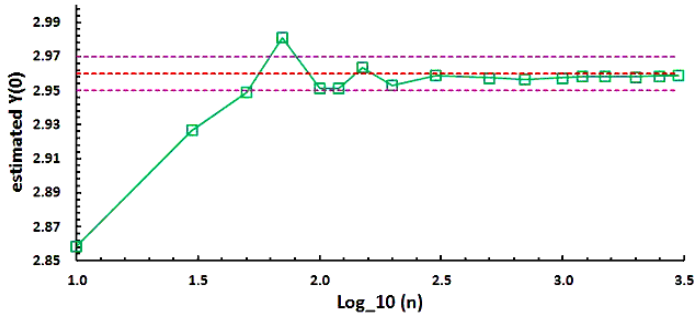


Figure 5: Empirical convergence of the proposed scheme for (6.10) for a call spread.

Next, let us consider a call-spread case: $\Phi(x) = (x - K_1)^+ - 2(x - K_2)^+$. This is exactly the same setup studied in [31] and hence we can test the performance of our scheme relative to the standard regression-based Monte Carlo simulation. Let us choose the same parameter sets as in [31]:

$$\{r = 0.01,\ R = 0.06,\ \mu = 0.05,\ X_0 = 100,\ T = 0.25,\ \sigma = 0.2,\ K_1 = 95,\ K_2 = 105\} \quad (6.11)$$

The result of [31] suggests that $Y_0 = 2.96 \pm 0.01$ or $Y_0 = 2.95 \pm 0.01$ with one standard deviation dependent on the choice of basis functions for the regressions. In Figure 5, we have compared the estimated $Y_0$ from our scheme and the one in [31]. The dotted lines represent $2.96 \pm 0.01$ for ease of comparison. In our scheme, $Y_0$ converges toward 2.96. In fact, the improvement of the regression method of [31] using martingale basis functions proposed by Bender & Steiner (2012) [5] suggests 2.96 which is perfectly consistent with our result.

We have also tested the convergence with a longer maturity and higher volatilities for the final payoff $\Phi(x) = (x - K_1)^+ - (x - K_2)^+$. We have used $\{r = 0.01,\ R = 0.06,\ \mu = 0.05,\ X_0 = 100,\ K_1 = 95,\ K_2 = 105\}$ as before, but with longer maturity $T = 1.0$ and $\text{set}_1 := \{\sigma = 0.3\}$, $\text{set}_2 := \{\sigma = 0.5\}$ and $\text{set}_3 := \{\sigma = 1.0\}$. From Figure 6, one observes smooth convergence for all the cases. The decrease in price for higher volatilities is natural from the fact that $K_2$ is closer to the *at-the-money-forward* point and hence the short position

has higher sensitivity on the volatility.

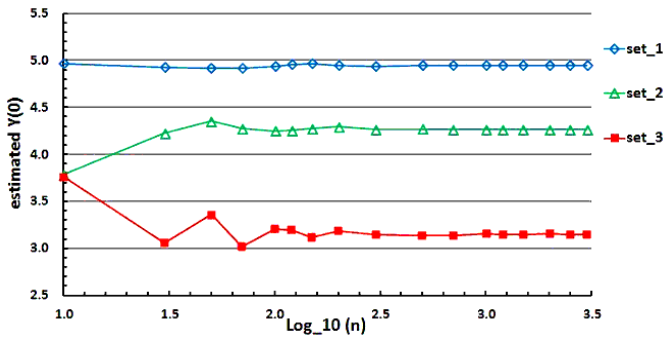


Figure 6: Empirical convergence of the proposed scheme for (6.10) for a call spread with $T = 1.0$ and higher volatilities.

### An example with a large Lipschitz constant

Bender & Steiner [5] have tested an extreme scenario with a parameter set (6.11) replaced by $R = 3.01$. In this case, the the non-linearity of the driver has a Lipschitz constant $(R-r)/\sigma = 15$. Their experiments suggest that the standard method of [31] fails to converge for this example under the simulation settings they tried. Their improved method with martingale basis functions (see Table 3 in [5]) gives $Y_0 \simeq 6.47$ with $n = 128$ and $Y_0 \simeq 6.44$ with the finest partition $n = 181$.

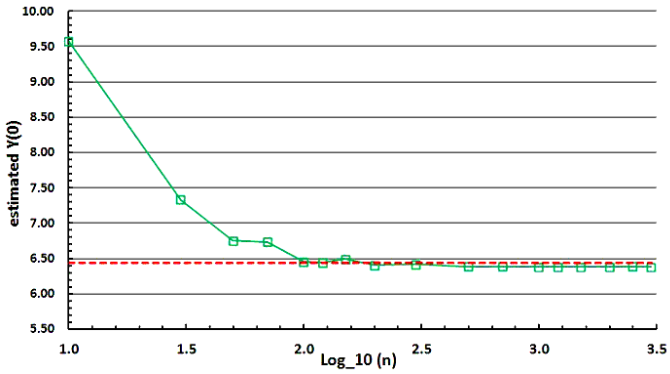


Figure 7: Empirical convergence of the proposed scheme for (6.10) with $R = 3.01$.

In Figure 7, we have plotted estimated $Y_0$ from our scheme with $n = 10$ to $n = 3000$. The dotted line corresponds to the value 6.44 given in [5]. In our scheme, $Y_0$ seems to converge 6.38. In particular, with the same discretization $n = 181$, our scheme yields $Y_0 \simeq 6.43$ showing a nice consistency. Note that the method [5] requires to change the basis functions based on the law of $X$.

# A BMO-martingale and its properties

In this section, let us summarize the properties BMO-martingales, the associated $\mathcal{H}^2_{BMO}$-space and their properties which play an important role in the discussions.

**Definition A.1.** *A BMO-martingale $M$ is a square integrable martingale satisfying $M_0 = 0$ and*

$$||M||^2_{BMO} := \sup_{\tau \in \mathcal{T}_0^T} \Big|\Big| \mathbb{E}\Big[\langle M\rangle_T - \langle M\rangle_\tau | \mathcal{F}_\tau\Big] \Big|\Big|_\infty < \infty \ ,$$

*where the supremum is taken over all stopping times $\tau \in \mathcal{T}_0^T$.*

**Definition A.2.** *$\mathcal{H}^2_{BMO}(\mathbb{R}^k)$ is the set of $\mathbb{R}^k$-valued progressively measurable processes $Z$ satisfying*

$$||Z||^2_{\mathcal{H}^2_{BMO}} := \sup_{\tau \in \mathcal{T}_0^T} \Big|\Big|\mathbb{E}\Big[\int_\tau^T |Z_s|^2 ds \Big| \mathcal{F}_\tau\Big]\Big|\Big|_\infty < \infty \ .$$

Note that if $Z \in \mathcal{H}^2_{BMO}(\mathbb{R}^{1\times d})$, we have

$$\Big|\Big|\int_0^\cdot Z_s dW_s\Big|\Big|^2_{BMO} = \sup_{\tau \in \mathcal{T}_0^T} \Big|\Big|\mathbb{E}\Big[\int_\tau^T |Z_s|^2 ds \Big| \mathcal{F}_\tau\Big]\Big|\Big|_\infty = ||Z||^2_{\mathcal{H}^2_{BMO}} < \infty \ ,$$

and hence $Z * W$ is a BMO-martingale. The next result is well-known as *energy inequality.*

**Lemma A.1.** *Let $Z$ be in $\mathcal{H}^2_{BMO}$. Then, for any $n \in \mathbb{N}$,*

$$\mathbb{E}\Big[\Big(\int_0^T |Z_s|^2 ds\Big)^n\Big] \leq n!\Big(||Z||^2_{\mathcal{H}^2_{BMO}}\Big)^n \ .$$

*Proof.* See proof of Lemma 9.6.5 in [20]. □

Let $\mathcal{E}(M)$ be a Doléan-Dade exponential of $M$.

**Lemma A.2.** *(Reverse Hölder inequality) Let $M$ be a BMO-martingale. Then, $\big(\mathcal{E}_t(M), t \in [0,T]\big)$ is a uniformly integrable martingale, and for every stopping time $\tau \in \mathcal{T}_0^T$, there exists some positive constant $r^* > 1$ such that the inequality*

$$\mathbb{E}\Big[\mathcal{E}_T(M)^r | \mathcal{F}_\tau\Big] \leq C_{r,M} \mathcal{E}_\tau(M)^r \ ,$$

*holds for every $1 < r \leq r^*$ with some positive constant $C_{r,M}$ depending only on $r$ and $||M||_{BMO}$.*

*Proof.* See Theorem 3.1 of Kazamaki (1994) [33]. □

**Lemma A.3.** *Let $M$ be a square integrable martingale and $\widehat{M} := \langle M \rangle - M$. Then, $M \in BMO(\mathbb{P})$ if and only if $\widehat{M} \in BMO(\mathbb{Q})$ with $d\mathbb{Q}/d\mathbb{P} = \mathcal{E}_T(M)$. Furthermore, $||\widehat{M}||_{BMO(\mathbb{Q})}$ is determined by some function of $||M||_{BMO(\mathbb{P})}$.*

*Proof.* See Theorem 2.4 and 3.3 in [33]. □

**Remark A.1.** *Theorem 3.1 [33] also tells that there exists some decreasing function $\Phi(r)$ with $\Phi(1+) = \infty$ and $\Phi(\infty) = 0$ such that if $||M||_{BMO(\mathbb{P})}$ satisfies $||M||_{BMO(\mathbb{P})} < \Phi(r)$ then $\mathcal{E}(M)$ allows the reverse Hölder inequality with power $r$. This implies together with Lemma A.3, one can take a common positive constant $\bar{r}$ satisfying $1 < \bar{r} \leq r^*$ such that both of the $\mathcal{E}(M)$ and $\mathcal{E}(\hat{M})$ satisfy the reverse Hölder inequality with power $\bar{r}$ under the respective probability measure $\mathbb{P}$ and $\mathbb{Q}$. Furthermore, the upper bound $r^*$ is determined only by $||M||_{BMO(\mathbb{P})}$ (or equivalently by $||M||_{BMO(\mathbb{Q})}$).*

# B Short-term expansion: Step 1

In the following two sections, we approximate the solution $(\overline{Y}^i, \overline{Z}^i)$ of the BSDE (3.1) semi-analytically and also obtain its error estimate. We need two steps for achieving this goal, which involve the linearization method and the small-variance expansion method for BSDEs proposed in Fujii & Takahashi (2012) [26] and (2015) [30], respectively [15].

[15] Note that the small-variance asymptotic expansion has been widely applied for the pricing of European contingent claims since the initial attempts by Takahashi (1999) [46] and Kunitomo & Takahashi (2003) [36].

When there is no confusion, we adopt the so-called Einstein convention assuming the obvious summation of duplicate indexes (such as $i \in \{1, \cdots, d\}$ of $x^i$) without explicitly using the summation symbol $\sum$. For example, $\partial_{x^i,x^j}\xi(X_T)\partial_x X_T^i \partial_x X_T^j$ assumes the summation about indexes $i$ and $j$ so that it denotes $\sum_{i,j=1}^d \partial_{x^i,x^j}\xi(X_T)\partial_x X_T^i \partial_x X_T^j$.

*(Standing Assumptions for Section B) We make Assumptions 2.1, 2.2 and Assumption 3.1 (i) the standing assumptions for this section.*

Let us introduce the next decomposition of the BSDE (3.1) for each interval $t \in I_i$, $i \in \{1, \cdots, n\}$:

$$\overline{Y}_t^{i,[0]} = \widehat{u}^{i+1}(X_{t_i}) - \int_t^{t_i} \overline{Z}_r^{i,[0]} dW_r \ , \tag{B.1}$$

$$\overline{Y}_t^{i,[1]} = \int_t^{t_i} f\big(r, X_r, \overline{Y}_r^{i,[0]}, \overline{Z}_r^{i,[0]}\big) dr - \int_t^{t_i} \overline{Z}_r^{i,[1]} dW_r \ . \tag{B.2}$$

They are the leading contributions in the linearization method [26, 47].

**Lemma B.1.** *For every interval $I_i, i \in \{1, \cdots, n\}$, there exists a unique solution $(\overline{Y}^{i,[0]}, \overline{Z}^{i,[0]})$ to the BSDE (B.1) satisfying, with some $(i,n)$-independent positive constants $C$ and $C_p$, that $||\overline{Y}^{i,[0]}||_{\mathcal{S}^\infty[t_{i-1},t_i]} + ||\overline{Z}^{i,[0]}||_{\mathcal{H}^2_{BMO}[t_{i-1},t_i]} \leq C$, and also $||\overline{Z}^{i,[0]}||_{\mathcal{S}^p[t_{i-1},t_i]} \leq C_p$ for any $p \geq 2$.*

*Proof.* The boundedness $||\overline{Y}^{i,[0]}||_{\mathcal{S}^\infty} \leq C$ follows easily from Assumption 3.1 (i), which then implies $||\overline{Z}^{i,[0]}||_{\mathcal{H}^2_{BMO}} \leq C$. The second claim follows from the similar arguments used in Proposition 3.1 . □

**Lemma B.2.** *For every interval $I_i$, $i \in \{1, \cdots, n\}$, there exists a unique solution $(\overline{Y}^{i,[1]}, \overline{Z}^{i,[1]})$ to the BSDE (B.2) satisfying, with some $(i,n)$-independent positive constant $C_p$, that*

$$||\overline{Y}^{i,[1]}||_{\mathcal{S}^p[t_{i-1},t_i]} + ||\overline{Z}^{i,[1]}||_{\mathcal{H}^p[t_{i-1},t_i]} \leq C_p$$

*for any $p \geq 2$.*

*Proof.* Since it is a Lipschitz BSDE (with zero Lipschitz constant), the existence of a unique solution easily follows. The standard estimate (see, for example, [12]) and Assumption 2.2 (i) implies

$$\begin{aligned}
\Big|\Big|(\overline{Y}^{i,[1]}, \overline{Z}^{i,[1]})\Big|\Big|^p_{\mathcal{K}^p[t_{i-1},t_i]} &\leq C_p \mathbb{E}\Big[\Big(\int_{t_{i-1}}^{t_i} |f(r, X_r, \overline{Y}_r^{i,[0]}, \overline{Z}_r^{i,[0]})| dr\Big)^p\Big] \\
&\leq C_p \mathbb{E}\Big[\Big(\int_{t_{i-1}}^{t_i} \big[l_r + \beta|\overline{Y}_r^{i,[0]}| + \frac{\gamma}{2}|\overline{Z}_r^{i,[0]}|^2\big] dr\Big)^p\Big] \\
&\leq C_p\Big(||l||_T^p + ||\overline{Y}^{i,[0]}||^p_{\mathcal{S}^p[t_{i-1},t_i]} + ||\overline{Z}^{i,[0]}||^{2p}_{\mathcal{S}^{2p}[t_{i-1},t_i]}\Big) \ .
\end{aligned}$$

Thus one obtains the desired result by Lemma B.1. □

We now define the process $(\widetilde{\overline{Y}}^i, \widetilde{\overline{Z}}^i)$ for each interval $t \in I_i$, $i \in \{1, \cdots, n\}$ by

$$\widetilde{\overline{Y}}_t^i := \overline{Y}_t^{i,[0]} + \overline{Y}_t^{i,[1]} \ , \qquad \widetilde{\overline{Z}}_t^i := \overline{Z}_t^{i,[0]} + \overline{Z}_t^{i,[1]} \ .$$

**Proposition B.1.** *There exists some $(i,n)$-independent positive constant $C_p$ such that the inequality*

$$\mathbb{E}\Big[\big|\big|\overline{Y}^i - \widetilde{\overline{Y}}^i\big|\big|^p_{[t_{i-1},t_i]} + \Big(\int_{t_{i-1}}^{t_i} \big|\overline{Z}_r^i - \widetilde{\overline{Z}}_r^i\big|^2 dr\Big)^{\frac{p}{2}}\Big] \leq C_p h_i^{3p/2}$$

*holds for every interval $I_i$, $i \in \{1, \cdots, n\}$ with any $p \geq 2$.*

*Proof.* For notational simplicity, let us put

$$\delta Y_t^{i,[0]} := \overline{Y}_t^i - \overline{Y}_t^{i,[0]}, \quad \delta Z_t^{i,[0]} := \overline{Z}_t^i - \overline{Z}_t^{i,[0]}$$
$$\delta Y_t^{i,[1]} := \overline{Y}_t^i - \widetilde{\overline{Y}}_t^i, \qquad \delta Z_t^{i,[1]} := \overline{Z}_t^i - \widetilde{\overline{Z}}_t^i$$

for each interval $t \in I_i, i \in \{1, \cdots, n\}$. Then, they are given by the solutions to the following BSDEs respectively:

$$\delta Y_t^{i,[0]} = \int_t^{t_i} f(r, X_r, \overline{Y}_r^i, \overline{Z}_r^i)dr - \int_t^{t_i} \delta Z_r^{i,[0]} dW_r \ ,$$
$$\delta Y_t^{i,[1]} = \int_t^{t_i} \Big( f(r, X_r, \overline{Y}_r^i, \overline{Z}_r^i) - f(r, X_r, \overline{Y}_r^{i,[0]}, \overline{Z}_r^{i,[0]}) \Big) dr - \int_t^{t_i} \delta Z_r^{i,[1]} dW_r \ .$$

By the stability result for the Lipschitz BSDEs (see, for example, [12]), Assumption 2.2 (i), (3.3) and Proposition 3.1, one obtains

$$\begin{aligned} \Big|\Big|(\delta Y^{i,[0]}, \delta Z^{i,[0]})\Big|\Big|^p_{\mathcal{K}^p[t_{i-1},t_i]} &\le C_p \mathbb{E}\Big[\Big(\int_{t_{i-1}}^{t_i} \big[l_r + \beta|\overline{Y}_r^i| + \frac{\gamma}{2}|\overline{Z}_r^i|^2\big]dr\Big)^p\Big] \\ &\le C_p h_i^p \left( ||l||_T^p + ||\overline{Y}^i||^p_{\mathcal{S}^\infty[t_{i-1},t_i]} + ||\overline{Z}^i||^{2p}_{\mathcal{S}^{2p}[t_{i-1},t_i]} \right) \le C_p h_i^p, \end{aligned} \tag{B.3}$$

with some $(i, n)$-independent positive constant $C_p$ for $\forall p \ge 2$. Similar analysis for $(\delta Y^{i,[1]}, \delta Z^{i,[1]})$ using Assumption 2.2 (ii) yields

$$\begin{aligned} \Big|\Big|(\delta Y^{i,[1]}, \delta Z^{i,[1]})\Big|\Big|^p_{\mathcal{K}^p[t_{i-1},t_i]} &\le C_p \mathbb{E}\Big[\Big(\int_{t_{i-1}}^{t_i} \Big[|\delta Y_r^{i,[0]}| + (1 + |\overline{Z}_r^i| + |\overline{Z}_r^{i,[0]}|)|\delta Z_r^{i,[0]}|\Big]dr\Big)^p\Big] \\ &\le C_p\Big(h_i^p ||\delta Y^{i,[0]}||^p_{\mathcal{S}^p[I_i]} + \mathbb{E}\Big[1 + ||\overline{Z}^i||^{2p}_{I_i} + ||\overline{Z}^{i,[0]}||^{2p}_{I_i}\Big]^{\frac{1}{2}} \mathbb{E}\Big[\Big(h_i \int_{t_{i-1}}^{t_i} |\delta Z_r^{i,[0]}|^2 dr\Big)^p\Big]^{\frac{1}{2}}\Big). \end{aligned}$$

By applying Proposition 3.1, Lemma B.1 and the previous estimate (B.3), one obtains the desired result. □

# C Short-term expansion: Step 2

In the second step, we obtain simple analytic approximation for the BSDEs (B.1) and (B.2) while keeping the same order of accuracy given in Proposition B.1. We use the small-variance expansion method for BSDEs proposed in [30] which renders all the problems into a set of simple ODEs. Furthermore, we shall see that these ODEs can be approximated by a *single-step* Euler method for each interval $I_i$.

*(Standing Assumptions for Section C) Similarly to the last section, we make Assumptions 2.1, 2.2 and Assumption 3.1 (i) the standing assumptions for this section.*

## C.1 Approximation for $(\overline{Y}^{i,[0]}, \overline{Z}^{i,[0]})$

For each interval, we introduce a new parameter $\epsilon$ satisfying $\epsilon \in (-c, c)$ with some constant $c > 1$ to perturb (2.1) and (B.1):

$$X_t^\epsilon = X_{t_{i-1}} + \int_{t_{i-1}}^t b(r, X_r^\epsilon)dr + \int_{t_{i-1}}^t \epsilon\sigma(r, X_r^\epsilon)dW_r \ . \tag{C.1}$$

$$\overline{Y}_t^{i,[0],\epsilon} = \widehat{u}^{i+1}(X_{t_i}^\epsilon) - \int_t^{t_i} \overline{Z}_r^{i,[0],\epsilon} dW_r \ . \tag{C.2}$$

for $t \in I_i = [t_{i-1}, t_i]$, $i \in \{1, \cdots, n\}$. Notice that the way $\epsilon$ is introduced to $X^\epsilon$, by which we have a different process for each interval $I_i$.[16] In the following, in order to avoid confusion between the index specifying the interval and the one for the component of $x \in \mathbb{R}^d$, we use the bold Gothic symbols such as $\{\mathbf{i}, \mathbf{j}, \cdots\}$ for the latter, each of which runs through 1 to $d$.

**Lemma C.1.** *The classical derivatives of* $(X^\epsilon, \overline{Y}^{i,[0],\epsilon}, \overline{Z}^{i,[0],\epsilon})$ *with respect to* $\epsilon$

$$\partial_\epsilon^k X_t^\epsilon := \frac{\partial^k}{\partial \epsilon^k} X_t^\epsilon, \quad \partial_\epsilon^k \overline{Y}^{i,[0],\epsilon} := \frac{\partial^k}{\partial \epsilon^k} \overline{Y}_t^{i,[0],\epsilon}, \quad \partial_\epsilon^k \overline{Z}^{i,[0],\epsilon} := \frac{\partial^k}{\partial \epsilon^k} \overline{Z}_t^{i,[0],\epsilon}$$

*for* $k = \{1, 2, 3\}$ *are given by the solutions to the following forward- and backward-SDEs:*

$$\begin{aligned}
\partial_\epsilon X_t^{\epsilon,\mathbf{i}} &= \int_{t_{i-1}}^t \partial_{x^\mathbf{j}} b^\mathbf{i}(r, X_r^\epsilon) \partial_\epsilon X_r^{\epsilon,\mathbf{j}} dr + \int_{t_{i-1}}^t \big[\sigma^\mathbf{i}(r, X_r^\epsilon) + \epsilon(\partial_\epsilon X_r^{\epsilon,\mathbf{j}}) \partial_{x^\mathbf{j}} \sigma^\mathbf{i}(r, X_r^\epsilon)\big] dW_r \ , \\
\partial_\epsilon^2 X_t^{\epsilon,\mathbf{i}} &= \int_{t_{i-1}}^t \big[\partial_{x^\mathbf{j}} b^\mathbf{i}(r, X_r^\epsilon) \partial_\epsilon^2 X_r^{\epsilon,\mathbf{j}} + \partial^2_{x^\mathbf{j}, x^\mathbf{k}} b^\mathbf{i}(r, X_r^\epsilon) \partial_\epsilon X_r^{\epsilon,\mathbf{j}} \partial_\epsilon X_r^{\epsilon,\mathbf{k}}\big] dr \\
&\quad + \int_{t_{i-1}}^t \big[2(\partial_\epsilon X_r^{\epsilon,\mathbf{j}}) \partial_{x^\mathbf{j}} \sigma^\mathbf{i}(r, X_r^\epsilon) + \epsilon(\partial_\epsilon^2 X_r^{\epsilon,\mathbf{j}}) \partial_{x^\mathbf{j}} \sigma^\mathbf{i}(r, X_r^\epsilon) + \epsilon(\partial_\epsilon X_r^{\epsilon,\mathbf{j}})(\partial_\epsilon X_r^{\epsilon,\mathbf{k}}) \partial^2_{x^\mathbf{j}, x^\mathbf{k}} \sigma^\mathbf{i}(r, X_r^\epsilon)\big] dW_r \ , \\
\partial_\epsilon^3 X_t^{\epsilon,\mathbf{i}} &= \int_{t_{i-1}}^t \big[\partial_{x^\mathbf{j}} b^\mathbf{i}(r, X_r^\epsilon) \partial_\epsilon^3 X_r^{\epsilon,\mathbf{j}} + 3\partial^2_{x^\mathbf{j}, x^\mathbf{k}} b^\mathbf{i}(r, X_r^\epsilon) \partial_\epsilon^2 X_r^{\epsilon,\mathbf{j}} \partial_\epsilon X_r^{\epsilon,\mathbf{k}} \\
&\quad + \partial^3_{x^\mathbf{j}, x^\mathbf{k}, x^\mathbf{m}} b^\mathbf{i}(r, X_r^\epsilon) \partial_\epsilon X_r^{\epsilon,\mathbf{j}} \partial_\epsilon X_r^{\epsilon,\mathbf{k}} \partial_\epsilon X_r^{\epsilon,\mathbf{m}}\big] dr + \int_{t_{i-1}}^t \big[3(\partial_\epsilon^2 X_r^{\epsilon,\mathbf{j}}) \partial_{x^\mathbf{j}} \sigma^\mathbf{i}(r, X_r^\epsilon) \\
&\quad + 3(\partial_\epsilon X_r^{\epsilon,\mathbf{j}})(\partial_\epsilon X_r^{\epsilon,\mathbf{k}}) \partial^2_{x^\mathbf{j}, x^\mathbf{k}} \sigma^\mathbf{i}(r, X_r^\epsilon) + \epsilon(\partial_\epsilon^3 X_r^{\epsilon,\mathbf{j}}) \partial_{x^\mathbf{j}} \sigma^\mathbf{i}(r, X_r^\epsilon) \\
&\quad + 3\epsilon(\partial_\epsilon^2 X_r^{\epsilon,\mathbf{j}})(\partial_\epsilon X_r^{\epsilon,\mathbf{k}}) \partial^2_{x^\mathbf{j}, x^\mathbf{k}} \sigma^\mathbf{i}(r, X_r^\epsilon) + \epsilon(\partial_\epsilon X_r^{\epsilon,\mathbf{j}})(\partial_\epsilon X_r^{\epsilon,\mathbf{k}})(\partial_\epsilon X_r^{\epsilon,\mathbf{m}}) \partial^3_{x^\mathbf{j}, x^\mathbf{k}, x^\mathbf{m}} \sigma^\mathbf{i}(r, X_r^\epsilon)\big] dW_r \ , \\
\partial_\epsilon \overline{Y}_t^{i,[0],\epsilon} &= \partial_{x^\mathbf{j}} \widehat{u}^{i+1}(X_{t_i}^\epsilon) \partial_\epsilon X_{t_i}^{\epsilon,\mathbf{j}} - \int_t^{t_i} \partial_\epsilon \overline{Z}_r^{i,[0],\epsilon} dW_r \ , \\
\partial_\epsilon^2 \overline{Y}_t^{i,[0],\epsilon} &= \partial_{x^\mathbf{j}} \widehat{u}^{i+1}(X_{t_i}^\epsilon) \partial_\epsilon^2 X_{t_i}^{\epsilon,\mathbf{j}} + \partial^2_{x^\mathbf{j}, x^\mathbf{k}} \widehat{u}^{i+1}(X_{t_i}^\epsilon) \partial_\epsilon X_{t_i}^{\epsilon,\mathbf{j}} \partial_\epsilon X_{t_i}^{\epsilon,\mathbf{k}} - \int_t^{t_i} \partial_\epsilon^2 \overline{Z}_r^{i,[0],\epsilon} dW_r \ , \\
\partial_\epsilon^3 \overline{Y}_t^{i,[0],\epsilon} &= \partial_{x^\mathbf{j}} \widehat{u}^{i+1}(X_{t_i}^\epsilon) \partial_\epsilon^3 X_{t_i}^{\epsilon,\mathbf{j}} + 3\partial^2_{x^\mathbf{j}, x^\mathbf{k}} \widehat{u}^{i+1}(X_{t_i}^\epsilon)(\partial_\epsilon^2 X_{t_i}^{\epsilon,\mathbf{j}})(\partial_\epsilon X_{t_i}^{\epsilon,\mathbf{k}}) \\
&\quad + \partial^3_{x^\mathbf{j}, x^\mathbf{k}, x^\mathbf{m}} \widehat{u}^{i+1}(X_{t_i}^\epsilon)(\partial_\epsilon X_{t_i}^{\epsilon,\mathbf{j}})(\partial_\epsilon X_{t_i}^{\epsilon,\mathbf{k}})(\partial_\epsilon X_{t_i}^{\epsilon,\mathbf{m}}) - \int_t^{t_i} \partial_\epsilon^3 \overline{Z}_r^{i,[0],\epsilon} dW_r \ ,
\end{aligned}$$

*for* $t \in I_i = [t_{i-1}, t_i]$. *Einstein convention is used with* $\{\mathbf{i}, \mathbf{j}, \cdots\}$ *running through* 1 *to* $d$.

*Proof.* The classical differentiability can be shown by following the arguments of Theorem 3.1 in [39]. See Section 6 of [30] for more details. □

**Lemma C.2.** *For* $k = \{1, 2, 3\}$, *there exists some* $(i, n)$*-independent positive constant* $C_{p,k}$ *such that the inequality*

$$\mathbb{E}\Big[\big|\big|\partial_\epsilon^k X^\epsilon\big|\big|^p_{[t_{i-1}, t_i]}\Big] \leq C_{p,k} h_i^{kp/2}$$

*holds for every interval* $I_i$, $i \in \{1, \cdots, n\}$ *with any* $p \geq 2$.

*Proof.* This can be shown by applying the standard estimates for the Lipschitz SDEs given, for

[16] It would be more appropriate to write $X_t^{i,\epsilon}$ to emphasize the dependence on the interval $t \in I_i$, but we have omitted "$i$" to lighten the notation.

example, in Appendix A of [30]. For $k=1$,

$$\mathbb{E}\Big[||\partial_\epsilon X^\epsilon||^p_{[t_{i-1},t_i]}\Big] \quad \leq \quad C_p\mathbb{E}\Big[\Big(\int_{t_{i-1}}^{t_i}|\sigma(r,X^\epsilon_r)|^2dr\Big)^{p/2}\Big] \leq C_p h_i^{p/2} \ .$$

For $k=2$, one obtains

$$\begin{aligned}\mathbb{E}\Big[||\partial^2_\epsilon X^\epsilon||^p_{[t_{i-1},t_i]}\Big] &\leq C_p\mathbb{E}\Big[\Big(\int_{t_{i-1}}^{t_i}|\partial_\epsilon X^\epsilon_r|^2dr\Big)^p + \Big(\int_{t_{i-1}}^{t_i}\big[|\partial_\epsilon X^\epsilon_r|^2+|\partial_\epsilon X^\epsilon_r|^4\big]dr\Big)^{\frac{p}{2}}\Big]\\ &\leq C_p\left(h_i^p\mathbb{E}\Big[||\partial_\epsilon X^\epsilon||^{2p}_{I_i}\Big] + h_i^{p/2}\mathbb{E}\Big[||\partial_\epsilon X^\epsilon||^p_{I_i}+||\partial_\epsilon X^\epsilon||^{2p}_{I_i}\Big]\right) \leq C_p h_i^p \ ,\end{aligned}$$

as desired. One can show the last case $k=3$ in a similar manner. □

Let introduce the following processes, with $k\in\{0,1,2\}$,

$$X^{[k]}_t := \frac{\partial^k}{\partial\epsilon^k}X^\epsilon_t\Big|_{\epsilon=0}, \quad \overline{Y}^{i,[0],[k]}_t := \frac{\partial^k}{\partial\epsilon^k}\overline{Y}^{i,[0],\epsilon}_t\Big|_{\epsilon=0}, \quad \overline{Z}^{i,[0],[k]}_t := \frac{\partial^k}{\partial\epsilon^k}\overline{Z}^{i,[0],\epsilon}_t\Big|_{\epsilon=0}$$

and also

$$\widetilde{\overline{Y}}^{i,[0]}_t := \sum_{k=0}^{2}\frac{1}{k!}\overline{Y}^{i,[0],[k]}_t, \quad \widetilde{\overline{Z}}^{i,[0]}_t := \sum_{k=0}^{2}\frac{1}{k!}\overline{Z}^{i,[0],[k]}_t \tag{C.3}$$

for each interval $t\in I_i,\ i\in\{1,\cdots,n\}$.

**Lemma C.3.** *There exists some $(i,n)$-independent positive constant $C_p$ such that the inequality*

$$\mathbb{E}\Big[\big|\big|\overline{Y}^{i,[0]}-\widetilde{\overline{Y}}^{i,[0]}\big|\big|^p_{[t_{i-1},t_i]} + \Big(\int_{t_{i-1}}^{t_i}|\overline{Z}^{i,[0]}_r-\widetilde{\overline{Z}}^{i,[0]}_r|^2dr\Big)^{p/2}\Big] \leq C_p h_i^{3p/2}$$

*holds for every interval $I_i, i\in\{1,\cdots,n\}$ with any $p\geq 2$.*

*Proof.* We can use the residual formula of Taylor expansion thanks to the classical differentiability of $\overline{\Theta}^{i,[0],\epsilon}$ with respect to $\epsilon$;

$$\begin{aligned}&\mathbb{E}\Big[\big|\big|\overline{Y}^{i,[0]}-\widetilde{\overline{Y}}^{i,[0]}\big|\big|^p_{[t_{i-1},t_i]} + \Big(\int_{t_{i-1}}^{t_i}|\overline{Z}^{i,[0]}_r-\widetilde{\overline{Z}}^{i,[0]}_r|^2dr\Big)^{p/2}\Big]\\ &\leq C_p\mathbb{E}\Big[\sup_{r\in I_i}\Big|\frac{1}{2}\int_0^1(1-\epsilon)^2\partial^3_\epsilon\overline{Y}^{i,[0],\epsilon}_r d\epsilon\Big|^p + \Big(\int_{t_{i-1}}^{t_i}\Big|\frac{1}{2}\int_0^1(1-\epsilon)^2\partial^3_\epsilon\overline{Z}^{i,[0],\epsilon}_r d\epsilon\Big|^2dr\Big)^{p/2}\Big]\\ &\leq C_p\int_0^1\Big(\mathbb{E}\Big[\big|\big|\partial^3_\epsilon\overline{Y}^{i,[0],\epsilon}\big|\big|^p_{[t_{i-1},t_i]} + \Big(\int_{t_{i-1}}^{t_i}\big|\partial^3_\epsilon\overline{Z}^{i,[0],\epsilon}_r\big|^2dr\Big)^{p/2}\Big]\Big)d\epsilon \ .\end{aligned}$$

Applying the standard estimates of the Lipschitz BSDEs (see, for example, [12]), the boundedness of $\partial^k_x\widehat{u}^{i+1}$ as well as Lemma C.2, one obtains

$$\begin{aligned}&\mathbb{E}\Big[\big|\big|\overline{Y}^{i,[0]}-\widetilde{\overline{Y}}^{i,[0]}\big|\big|^p_{[t_{i-1},t_i]} + \Big(\int_{t_{i-1}}^{t_i}\big|\overline{Z}^{i,[0]}_r-\widetilde{\overline{Z}}^{i,[0]}_r\big|^2dr\Big)^{p/2}\Big]\\ &\leq C_p\int_0^1\Big(\mathbb{E}\Big[||\partial^3_\epsilon X^\epsilon||^p_{I_i}+||\partial^2_\epsilon X^\epsilon||^p_{I_i}||\partial_\epsilon X^\epsilon||^p_{I_i}+||\partial_\epsilon X^\epsilon||^{3p}_{I_i}\Big]\Big)d\epsilon \leq C_p h_i^{3p/2}\end{aligned}$$

as desired. □

The last lemma implies that it suffices to obtain $(\overset{\sim}{\overline{Y}}{}^{i,[0]}, \overset{\sim}{\overline{Z}}{}^{i,[0]})$ for our purpose[17], which is the second order approximation of $(\overline{Y}^{i,[0]}, \overline{Z}^{i,[0]})$. Furthermore, as we shall see next, the solution of these BSDEs can be obtained explicitly by simple ODEs thanks to the grading structure introduced by the asymptotic expansion. The relevant system of FBSDEs is summarized below:

$$
\begin{aligned}
X_t^{[0]} &= X_{t_{i-1}} + \int_{t_{i-1}}^t b(r, X_r^{[0]})dr \ , \\
X_t^{[1],\mathbf{i}} &= \int_{t_{i-1}}^t \partial_{x^{\mathbf{j}}} b^{\mathbf{i}}(r, X_r^{[0]}) X_r^{[1],\mathbf{j}} dr + \int_{t_{i-1}}^t \sigma^{\mathbf{i}}(r, X_r^{[0]}) dW_r \ , \\
X_t^{[2],\mathbf{i}} &= \int_{t_{i-1}}^t \Big( \partial_{x^{\mathbf{j}}} b^{\mathbf{i}}(r, X_r^{[0]}) X_r^{[2],\mathbf{j}} + \partial^2_{x^{\mathbf{j}},x^{\mathbf{k}}} b^{\mathbf{i}}(r, X_r^{[0]}) X_r^{[1],\mathbf{j}} X_r^{[1],\mathbf{k}} \Big) dr + \int_{t_{i-1}}^t 2X_r^{[1],\mathbf{j}} \partial_{x^{\mathbf{j}}} \sigma^{\mathbf{i}}(r, X_r^{[0]}) dW_r \ , \\
\overline{Y}_t^{i,[0],[0]} &= \widehat{u}^{i+1}(X_{t_i}^{[0]}) - \int_t^{t_i} \overline{Z}_r^{i,[0],[0]} dW_r \ , \qquad \text{(C.4)} \\
\overline{Y}_t^{i,[0],[1]} &= \partial_{x^{\mathbf{j}}} \widehat{u}^{i+1}(X_{t_i}^{[0]}) X_{t_i}^{[1],\mathbf{j}} - \int_t^{t_i} \overline{Z}_r^{i,[0],[1]} dW_r \ , \qquad \text{(C.5)} \\
\overline{Y}_t^{i,[0],[2]} &= \partial_{x^{\mathbf{j}}} \widehat{u}^{i+1}(X_{t_i}^{[0]}) X_{t_i}^{[2],\mathbf{j}} + \partial^2_{x^{\mathbf{j}},x^{\mathbf{k}}} \widehat{u}^{i+1}(X_{t_i}^{[0]}) X_{t_i}^{[1],\mathbf{j}} X_{t_i}^{[1],\mathbf{k}} - \int_t^{t_i} \overline{Z}_r^{i,[0],[2]} dW_r \ , \qquad \text{(C.6)}
\end{aligned}
$$

for $t \in I_i, i \in \{1, \cdots, n\}$ with Einstein convention for $\{\mathbf{i}, \mathbf{j}, \cdots\}$.

**Definition C.1.** *(Coefficient functions)*
*We define the set of functions* $\chi : I_i \times \mathbb{R}^d \to \mathbb{R}^d$, $y : I_i \times \mathbb{R}^d \to \mathbb{R}$, $\mathbf{y}^{[1]} : I_i \times \mathbb{R}^d \to \mathbb{R}^d$, $\mathbf{y}^{[2]} : I_i \times \mathbb{R}^d \to \mathbb{R}^d$, $G^{[2]} : I_i \times \mathbb{R}^d \to \mathbb{R}^{d\times d}$, $y_0^{[2]} : I_i \times \mathbb{R}^d \to \mathbb{R}$ *by*

$$
\begin{aligned}
\chi(t,x) &:= x + \int_{t_{i-1}}^t b(r, \chi(r,x)) dr \ , \\
y(t,x) &:= \widehat{u}^{i+1}(\chi(t_i, x)) \ , \\
\mathbf{y}_{\mathbf{j}}^{[1]}(t,x) &:= \partial_{x^{\mathbf{j}}} \widehat{u}^{i+1}(\chi(t_i,x)) + \int_t^{t_i} \partial_{x^{\mathbf{j}}} b^{\mathbf{k}}(r, \chi(r,x)) \mathbf{y}_{\mathbf{k}}^{[1]}(r,x) dr \ , \\
G_{\mathbf{j},\mathbf{k}}^{[2]}(t,x) &:= \partial^2_{x^{\mathbf{j}},x^{\mathbf{k}}} \widehat{u}^{i+1}(\chi(t_i,x)) + \int_t^{t_i} \Big\{ \Big( \big[\partial_x b(r, \chi(r,x))\big] G^{[2]}(r,x) \Big)^{\leftrightarrow}_{\mathbf{j},\mathbf{k}} \\
&\qquad + \partial^2_{x^{\mathbf{j}},x^{\mathbf{k}}} b^{\mathbf{m}}(r, \chi(r,x)) \mathbf{y}_{\mathbf{m}}^{[2]}(r,x) \Big\} dr \ , \\
y_0^{[2]}(t,x) &:= \int_t^{t_i} \mathrm{Tr}\Big( G^{[2]}(r,x) [\sigma\sigma^\top](r, \chi(r,x)) \Big) dr \ ,
\end{aligned}
$$

*and* $\mathbf{y}^{[2]} = \mathbf{y}^{[1]}$ *for* $(t,x) \in I_i \times \mathbb{R}^d$, $i \in \{1, \cdots, n\}$. *We have used Einstein convention and the notation* $\big([\partial_x b(r,x)]_{\mathbf{i},\mathbf{j}} = \partial_{x^{\mathbf{i}}} b^{\mathbf{j}}(r,x), \mathbf{i}, \mathbf{j} \in \{1, \cdots, d\}\big)$. *We denote the symmetrization by* $A^{\leftrightarrow} := A + A^\top$ *for a* $d \times d$*-matrix* $A$ [18].

Note that the above coefficient functions are given by the ODEs for a given $x \in \mathbb{R}^d$ in each period. The solution of the BSDEs are expressed by these functions in the following way:

**Lemma C.4.** *For each period* $t \in I_i, i \in \{1, \cdots, n\}$, *the solutions of the BSDEs (C.4), (C.5) and (C.6) are given by, with Einstein convention,*

$$
\begin{aligned}
&\overline{Y}_t^{i,[0],[0]} = y(t, X_{t_{i-1}}), \quad \overline{Z}_t^{i,[0],[0]} \equiv 0 \qquad \textit{(0-th order)} \\
&\overline{Y}_t^{i,[0],[1]} = \mathbf{y}_{\mathbf{j}}^{[1]}(t, X_{t_{i-1}}) X_t^{[1],\mathbf{j}}, \quad \overline{Z}_t^{i,[0],[1]} = \mathbf{y}_{\mathbf{j}}^{[1]}(t, X_{t_{i-1}}) \sigma^{\mathbf{j}}(t, \chi(t, X_{t_{i-1}})), \quad \textit{(1st order)}
\end{aligned}
$$

[17] See Remark 3.3.

[18] Hence, $G^{[2]}$ is symmetric matrix valued.

*and lastly, for the 2nd order*

$$\begin{aligned}\overline{Y}_t^{i,[0],[2]} &= \mathbf{y}_{\mathbf{j}}^{[2]}(t,X_{t_{i-1}})X_t^{[2],\mathbf{j}} + G_{\mathbf{j},\mathbf{k}}^{[2]}(t,X_{t_{i-1}})X_t^{[1],\mathbf{j}}X_t^{[1],\mathbf{k}} + y_0^{[2]}(t,X_{t_{i-1}}) \ , \\ \overline{Z}_t^{i,[0],[2]} &= 2\Big(\mathbf{y}_{\mathbf{j}}^{[2]}(t,X_{t_{i-1}})X_t^{[1],\mathbf{k}}\partial_{x^{\mathbf{k}}}\sigma^{\mathbf{j}}(t,\chi(t,X_{t_{i-1}})) + G_{\mathbf{j},\mathbf{k}}^{[2]}(t,X_{t_{i-1}})X_t^{[1],\mathbf{j}}\sigma^{\mathbf{k}}(t,\chi(t,X_{t_{i-1}}))\Big) \ .\end{aligned}$$

*Proof.* This is a special case of the results of Section 8 of [30]. The existence of the unique solution to the BSDEs (C.4), (C.5) and (C.6) is obvious. The expression can be directly checked by applying Itô formula to the suggested forms using the ODEs given in Definition C.1, and compare the results with the BSDEs. □

Since each interval $I_i$ has a very short span $h_i$, we expect that we can approximate the above ODEs by just a single-step of Euler method without affecting the order of error given in Lemma C.3.

**Definition C.2.** *(Approximated coefficient functions)*
*We define the set of functions;* $\overline{\chi} : I_i \times \mathbb{R}^d \to \mathbb{R}^d$, $\overline{y} : I_i \times \mathbb{R}^d \to \mathbb{R}$, $\overline{\mathbf{y}}^{[1]} : I_i \times \mathbb{R}^d \to \mathbb{R}^d$, $\overline{\mathbf{y}}^{[2]} : I_i \times \mathbb{R}^d \to \mathbb{R}^d$, $\overline{G}^{[2]} : I_i \times \mathbb{R}^d \to \mathbb{R}^{d\times d}$, $\overline{y}_0^{[2]} : I_i \times \mathbb{R}^d \to \mathbb{R}$ *by*

$$\begin{aligned}\overline{\chi}(t,x) &:= x + \Delta(t)b(t_{i-1},x) \ , \\ \overline{y}(t,x) &:= \widehat{u}^{i+1}(\overline{\chi}(t_i,x)) \ , \\ \overline{\mathbf{y}}_{\mathbf{j}}^{[1]}(t,x) &:= \partial_{x^{\mathbf{j}}}\widehat{u}^{i+1}(\overline{\chi}(t_i,x)) + \delta(t)\partial_{x^{\mathbf{j}}}b^{\mathbf{k}}(t_i,\overline{\chi}(t_i,x))\partial_{x^{\mathbf{k}}}\widehat{u}^{i+1}(\overline{\chi}(t_i,x)) \ , \\ \overline{G}_{\mathbf{j},\mathbf{k}}^{[2]}(t,x) &:= \partial^2_{x^{\mathbf{j}},x^{\mathbf{k}}}\widehat{u}^{i+1}(\overline{\chi}(t_i,x)) + \delta(t)\Big\{\Big([\partial_x b(t_i,\overline{\chi}(t_i,x))]\partial^2_{x,x}\widehat{u}^{i+1}(\overline{\chi}(t_i,x))\Big)^{\leftrightarrow}_{\mathbf{j},\mathbf{k}} \\ &\qquad + \partial^2_{x^{\mathbf{j}},x^{\mathbf{k}}}b^{\mathbf{m}}(t_i,\overline{\chi}(t_i,x))\partial_{x^{\mathbf{m}}}\widehat{u}^{i+1}(\overline{\chi}(t_i,x))\Big\} \ , \\ \overline{y}_0^{[2]}(t,x) &:= \delta(t)\mathrm{Tr}\Big(\overline{G}^{[2]}(t_i,x)[\sigma\sigma^\top](t_i,\overline{\chi}(t_i,x))\Big) \ ,\end{aligned}$$

*and* $\overline{\mathbf{y}}^{[2]} = \overline{\mathbf{y}}^{[1]}$ *for* $(t,x) \in I_i \times \mathbb{R}^d$, $i \in \{1,\cdots,n\}$. *We have used Einstein convention and the notations* $\Delta(t) := t - t_{i-1}$, $\delta(t) := t_i - t$.

The functions in Definition C.2 provide good approximations for the coefficient functions in Definition C.1 in the following sense:

**Lemma C.5.** *There exists some* $(i,n)$*-independent positive constant* $C_p$ *satisfying*

$$\begin{aligned}&\mathbb{E}\Big\{\sup_{t\in I_i}\Big|\chi(t,X_{t_{i-1}}) - \overline{\chi}(t,X_{t_{i-1}})\Big|^p + \sup_{t\in I_i}\Big|y(t,X_{t_{i-1}}) - \overline{y}(t,X_{t_{i-1}})\Big|^p \\ &\quad + \sum_{k=1}^{2}\sup_{t\in I_i}\Big|\mathbf{y}^{[k]}(t,X_{t_{i-1}}) - \overline{\mathbf{y}}^{[k]}(t,X_{t_{i-1}})\Big|^p + \sup_{t\in I_i}\Big|G^{[2]}(t,X_{t_{i-1}}) - \overline{G}^{[2]}(t,X_{t_{i-1}})\Big|^p \\ &\quad + \sup_{t\in I_i}\Big|y_0^{[2]}(t,X_{t_{i-1}}) - \overline{y}_0^{[2]}(t,X_{t_{i-1}})\Big|^p\Big\} \le C_p h_i^{3p/2} \ ,\end{aligned}$$

*for every interval* $I_i, i \in \{1,\cdots,n\}$ *with any* $p \ge 2$.

*Proof.* Firstly, let us consider $(\chi,\overline{\chi})$. Using 1/2-Hölder continuity in $t$, the global Lipschitz and linear growth properties of $b$ in $x$, we have

$$\begin{aligned}|\chi(t,x) - \overline{\chi}(t,x)| &\le \int_{t_{i-1}}^t |b(r,\chi(r,x)) - b(t_{i-1},x)|dr \\ &\le K\int_{t_{i-1}}^t \Big[\Delta(r)^{1/2} + |\chi(r,x) - \overline{\chi}(r,x)| + \Delta(r)|b(t_{i-1},x)|\Big]dr \\ &\le C(1+|x|h_i^{1/2})h_i^{3/2} + K\int_{t_{i-1}}^t |\chi(r,x) - \overline{\chi}(r,x)|dr\end{aligned}$$

and hence by Gronwall inequality, $\sup_{t\in I_i}|\chi(t,x)-\overline{\chi}(t,x)|\le e^{Kh_i}C(1+|x|\sqrt{h_i})h_i^{3/2}$. Thus

$$\mathbb{E}\Big[\sup_{t\in I_i}|\chi(t,X_{t_{i-1}})-\overline{\chi}(t,X_{t_{i-1}})|^p\Big]\le C_p h_i^{3p/2}\left(1+h_i^{p/2}\mathbb{E}\Big[|X_{t_{i-1}}|^p\Big]\right)\le C_p h_i^{3p/2} \tag{C.7}$$

with some $(i,n)$-independent positive constant $C_p$. Since $|\partial_x\widehat{u}^{i+1}|\le K'$ by Assumption 3.1 (i),

$$|y(t,x)-\overline{y}(t,x)| \;=\; |\widehat{u}^{i+1}(\chi(t_i,x))-\widehat{u}^{i+1}(\overline{\chi}(t_i,x))|\le K'|\chi(t_i,x)-\overline{\chi}(t_i,x)|\ .$$

Thus from (C.7),

$$\mathbb{E}\Big[\sup_{t\in I_i}|y(t,X_{t_{i-1}})-\overline{y}(t,X_{t_{i-1}})|^p\Big]\le C_p h_i^{3p/2} \tag{C.8}$$

with some $(i,n)$-independent positive constant $C_p$.

Let us now consider

$$\overline{\mathbf{y}}^{[1]}(t,x)=\partial_x\widehat{u}^{i+1}(\overline{\chi}(t_i,x))+\delta(t)\big[\partial_x b(t_i,\overline{\chi}(t_i,x))\big]\partial_x\widehat{u}^{i+1}(\overline{\chi}(t_i,x))\ .$$

Since both $|\partial_x\widehat{u}^{i+1}|$ and $|\partial_x b|$ are bounded, it is easy to see

$$\sup_{(t,x)\in I_i\times\mathbb{R}^d}|\overline{\mathbf{y}}^{[1]}(t,x)|\le C \tag{C.9}$$

with some positive constant $C$. For $t\in I_i$ with a given $x\in\mathbb{R}^d$, we have

$$\begin{aligned}\mathbf{y}^{[1]}(t,x)-\overline{\mathbf{y}}^{[1]}(t,x)&=\partial_x\widehat{u}^{i+1}(\chi(t_i,x))-\partial_x\widehat{u}^{i+1}(\overline{\chi}(t_i,x))\\&\quad+\int_t^{t_i}\Big(\partial_x b(r,\chi(r,x))\mathbf{y}^{[1]}(r,x)-\partial_x b(t_i,\overline{\chi}(t_i,x))\overline{\mathbf{y}}^{[1]}(t_i,x)\Big)dr\ .\end{aligned}$$

From (C.9), 1/2-Hölder continuity and global Lipschitz property of $\partial_x b$, we obtain

$$\begin{aligned}&|\mathbf{y}^{[1]}(t,x)-\overline{\mathbf{y}}^{[1]}(t,x)|\\&\le|\partial_x\widehat{u}^{i+1}(\chi(t_i,x))-\partial_x\widehat{u}^{i+1}(\overline{\chi}(t_i,x))|+\int_t^{t_i}\Big\{|\partial_x b(r,\chi(r,x))||\mathbf{y}^{[1]}(r,x)-\overline{\mathbf{y}}^{[1]}(r,x)|\\&\qquad+|\partial_x b(r,\chi(r,x))||\overline{\mathbf{y}}^{[1]}(r,x)-\overline{\mathbf{y}}^{[1]}(t_i,x)|+|\partial_x b(r,\chi(r,x))-\partial_x b(t_i,\overline{\chi}(t_i,x))||\overline{\mathbf{y}}^{[1]}(t_i,x)|\Big\}dr\\&\le K'|\chi(t_i,x)-\overline{\chi}(t_i,x)|+K\int_t^{t_i}|\mathbf{y}^{[1]}(r,x)-\overline{\mathbf{y}}^{[1]}(r,x)|dr\\&\qquad+Ch_i^2+C\int_t^{t_i}\Big(\delta(r)^{1/2}+|\chi(r,x)-\overline{\chi}(r,x)|+|\overline{\chi}(r,x)-\overline{\chi}(t_i,x)|\Big)dr\\&\le K\int_t^{t_i}|\mathbf{y}^{[1]}(r,x)-\overline{\mathbf{y}}^{[1]}(r,x)|dr+Ch_i^{3/2}\big(1+|x|\sqrt{h_i}\big)\ .\end{aligned}$$

Thus the backward Gronwall inequality (see, for example, Corollary 6.62 in [43]) gives

$$\sup_{t\in I_i}|\mathbf{y}^{[1]}(t,x)-\overline{\mathbf{y}}^{[1]}(t,x)|\le Ch_i^{3/2}(1+|x|\sqrt{h_i})e^{Kh_i}\ ,$$

and hence

$$\mathbb{E}\Big[\sup_{t\in I_i}|\mathbf{y}^{[1]}(t,X_{t_{i-1}})-\overline{\mathbf{y}}^{[1]}(t,X_{t_{i-1}})|^p\Big]\le C_p h_i^{3p/2}\ , \tag{C.10}$$

with some $(i,n)$-independent constant $C_p$ as desired.

By the boundedness of $|\partial_x^m \widehat{u}^{i+1}(x)|$ and $|\partial_x^m b|$ with $m \in \{1,2\}$, it is easy to see that $|\overline{G}^{[2]}|$ is also bounded

$$\sup_{(t,x)\in I_i\times\mathbb{R}^d} |\overline{G}^{[2]}(t,x)| \le C \tag{C.11}$$

with some positive constant $C$. Similar analysis done for $\overline{\mathbf{y}}^{[1]}$ using (C.11), 1/2-Hölder and Lipschitz continuities for $\partial_x b, \partial_x^2 b$, the backward Gronwall inequality yields

$$\sup_{t\in I_i} |G^{[2]}(t,x) - \overline{G}^{[2]}(t,x)| \le C h_i^{3/2}(1+|x|\sqrt{h_i}) \ ,$$

and hence

$$\mathbb{E}\Big[\sup_{t\in I_i} |G^{[2]}(t,X_{t_{i-1}}) - \overline{G}^{[2]}(t,X_{t_{i-1}})|^p\Big] \le C_p h_i^{3p/2} \tag{C.12}$$

with some $(i,n)$-independent positive constant $C_p$ as desired.

Finally, we consider

$$\overline{y}_0^{[2]}(t,x) = \delta(t)\mathrm{Tr}\Big(\overline{G}^{[2]}(t_i,x)[\sigma\sigma^\top]\big(t_i,\overline{\chi}(t_i,x)\big)\Big) \ .$$

From (C.11) and the linear-growth property of $\sigma$,

$$|\overline{y}_0^{[2]}(t,x)| \le C\delta(t)\big(1+|x|^2\big) \ , \tag{C.13}$$

is satisfied for every $(t,x)\in I_i\times\mathbb{R}^d$ with some positive constant $C$. We have

$$y_0^{[2]}(t,x) - \overline{y}_0^{[2]}(t,x) = \int_t^{t_i} \mathrm{Tr}\Big(G^{[2]}(r,x)[\sigma\sigma^\top](r,\chi(r,x)) - \overline{G}^{[2]}(t_i,x)[\sigma\sigma^\top](t_i,\overline{\chi}(t_i,x))\Big)dr$$

and thus

$$\begin{aligned}
|y_0^{[2]}(t,x) - \overline{y}_0^{[2]}(t,x)| \le& \int_t^{t_i} \mathrm{Tr}\Big\{\Big(|G^{[2]}(r,x) - \overline{G}^{[2]}(r,x)| + |\overline{G}^{[2]}(r,x) - \overline{G}^{[2]}(t_i,x)|\Big)\big|[\sigma\sigma^\top](r,\chi(r,x))\big| \\
&+\big|\overline{G}^{[2]}(t_i,x)\big|\Big|[\sigma\sigma^\top](r,\chi(r,x)) - [\sigma\sigma^\top](t_i,\overline{\chi}(t_i,x))\Big|\Big\}dr \\
\le& C h_i^{3/2}(1+|x|) + C h_i^2|x|^2(1+h_i|x|)
\end{aligned}$$

with some $(i,n)$-independent constant $C$. Thus we obtain, for any $p\ge 2$,

$$\mathbb{E}\Big[\sup_{t\in I_i} |y_0^{[2]}(t,X_{t_{i-1}}) - \overline{y}_0^{[2]}(t,X_{t_{i-1}})|^p\Big] \le C_p h_i^{3p/2} \tag{C.14}$$

as desired. From (C.7), (C.8), (C.10), (C.12), (C.14) and $\mathbf{y}^{[2]} = \mathbf{y}^{[1]}$, the claim is proved. □

We now introduce the processes $(\widehat{Y}_t^{i,[0]}, \widehat{Z}_t^{i,[0]})$ for each period $t\in I_i$. They are defined by $(\widetilde{\overline{Y}}_t^{i,[0]}, \widetilde{\overline{Z}}_t^{i,[0]})$ of (C.3) with the coefficient functions in Definition C.1 replaced by the approximations in Definition C.2, i.e.;

$$\begin{aligned}
\widehat{Y}_t^{i,[0]} \ :=& \ \overline{y}(t,X_{t_{i-1}}) + (X_t^{[1]})^\top \overline{\mathbf{y}}^{[1]}(t,X_{t_{i-1}}) \\
+& \ \frac{1}{2}\Big((X_t^{[2]})^\top \overline{\mathbf{y}}^{[2]}(t,X_{t_{i-1}}) + (X_t^{[1]})^\top \overline{G}^{[2]}(t,X_{t_{i-1}})X_t^{[1]} + \overline{y}_0^{[2]}(t,X_{t_{i-1}})\Big), \qquad \text{(C.15)} \\
\widehat{Z}_t^{i,[0]} \ :=& \ \overline{\mathbf{y}}^{[1]\top}(t,X_{t_{i-1}})\sigma(t,\overline{\chi}(t,X_{t_{i-i}})) \\
&+\Big((X_t^{[1]})^\top \partial_x\sigma(t,\overline{\chi}(t,X_{t_{i-1}}))\overline{\mathbf{y}}^{[2]}(t,X_{t_{i-1}}) + (X_t^{[1]})^\top \overline{G}^{[2]}(t,X_{t_{i-1}})\sigma(t,\overline{\chi}(t,X_{t_{i-1}}))\Big), \qquad \text{(C.16)}
\end{aligned}$$

where we have used Matrix notation for simplicity. The details of indexing can be checked from those given in Lemma C.4.

**Lemma C.6.** *There exits some $(i,n)$-independent positive constant $C_p$ such that the inequality*

$$\mathbb{E}\Big[||\widetilde{\overline{Y}}^{i,[0]} - \widehat{Y}^{i,[0]}||^p_{[t_{i-1},t_i]}\Big] + \mathbb{E}\Big[||\widetilde{\overline{Z}}^{i,[0]} - \widehat{Z}^{i,[0]}||^p_{[t_{i-1},t_i]}\Big] \leq C_p h_i^{3p/2}$$

*holds for every interval $I_i, i \in \{1, \cdots, n\}$ for any $p \geq 2$.*

*Proof.* It can be shown easily from Lemmas C.2 and C.5. □

**Corollary C.1.** *There exits some $(i,n)$-independent positive constant $C_p$ such that*

$$\mathbb{E}\Big[||\overline{Y}^{i,[0]} - \widehat{Y}^{i,[0]}||^p_{[t_{i-1},t_i]} + \Big(\int_{t_{i-1}}^{t_i} |\overline{Z}^{i,[0]}_r - \widehat{Z}^{i,[0]}_r|^2 dr\Big)^{p/2}\Big] \leq C_p h_i^{3p/2}$$

*holds for every interval $I_i, i \in \{1, \cdots, n\}$ with any $p \geq 2$.*

*Proof.* It follows directly from Lemmas C.3 and C.6. □

Since $X^{[1]}_{t_{i-1}} = X^{[2]}_{t_{i-1}} = 0$, we have a very simple expression at the starting time $t_{i-1}$ of each period $I_i = [t_{i-1}, t_i]$:

$$\begin{aligned}
\widehat{Y}^{i,[0]}_{t_{i-1}} &= \overline{y}(t_{i-1}, X_{t_{i-1}}) + \frac{1}{2}\overline{y}^{[2]}_0(t_{i-1}, X_{t_{i-1}}), \\
\widehat{Z}^{i,[0]}_{t_{i-1}} &= \overline{\mathbf{y}}^{[1]\top}(t_{i-1}, X_{t_{i-1}})\sigma(t_{i-1}, X_{t_{i-1}}).
\end{aligned}$$

We have the following continuity property of the approximated solution $(\widehat{Y}^{i,[0]}, \widehat{Z}^{i,[0]})$:

**Lemma C.7.** *There exists some $(i,n)$-independent positive constant $C_p$ such that the inequality*

$$\mathbb{E}\Big[\sup_{t\in I_i}\Big|\widehat{Y}^{i,[0]}_t - \widehat{Y}^{i,[0]}_{t_{i-1}}\Big|^p\Big] + \mathbb{E}\Big[\sup_{t\in I_i}\Big|\widehat{Z}^{i,[0]}_t - \widehat{Z}^{i,[0]}_{t_{i-1}}\Big|^p\Big] \leq C_p h_i^{p/2} \ ,$$

*holds for every interval $I_i, i \in \{1, \cdots, n\}$ with any $p \geq 2$.*

*Proof.* Since $\overline{y}(t,x) = \overline{y}(t_{i-1}, x)$ for $(t,x) \in I_i \times \mathbb{R}^d$, we have

$$\begin{aligned}
\widehat{Y}^{i,[0]}_t - \widehat{Y}^{i,[0]}_{t_{i-1}} &= X^{[1]\top}_t \overline{\mathbf{y}}^{[1]}(t, X_{t_{i-1}}) + \frac{1}{2}\Big(X^{[2]\top}_t \overline{\mathbf{y}}^{[2]}(t, X_{t_{i-1}}) + X^{[1]\top}_t \overline{G}^{[2]}(t, X_{t_{i-1}}) X^{[1]}_t\Big) \\
&+ \frac{1}{2}\Big(\overline{y}^{[2]}_0(t, X_{t_{i-1}}) - \overline{y}^{[2]}_0(t_{i-1}, X_{t_{i-1}})\Big) \ .
\end{aligned}$$

The bounds in (C.9) (remember that $\overline{\mathbf{y}}^{[1]} = \overline{\mathbf{y}}^{[2]}$), (C.11) and (C.13) as well as the estimates in Lemma C.2 imply

$$\mathbb{E}\Big[\sup_{t\in I_i}\Big|\widehat{Y}^{i,[0]}_t - \widehat{Y}^{i,[0]}_{t_{i-1}}\Big|^p\Big] \leq C_p \mathbb{E}\Big[||X^{[1]}||^p_{I_i} + ||X^{[2]}||^p_{I_i} + ||X^{[1]}||^{2p}_{I_i} + h_i^p\Big(1 + |X_{t_{i-1}}|^{2p}\Big)\Big] \leq C_p h_i^{p/2}$$

as desired. Similarly we have

$$\begin{aligned}
|\widehat{Z}^{i,[0]}_t - \widehat{Z}^{i,[0]}_{t_{i-1}}| &\leq |\overline{\mathbf{y}}^{[1]}(t, X_{t_{i-1}}) - \overline{\mathbf{y}}^{[1]}(t_{i-1}, X_{t_{i-1}})||\sigma(t, \overline{\chi}(t, X_{t_{i-1}}))| \\
&\quad + |\overline{\mathbf{y}}^{[1]}(t_{i-1}, X_{t_{i-1}})||\sigma(t, \overline{\chi}(t, X_{t_{i-1}})) - \sigma(t_{i-1}, X_{t_{i-1}})| \\
&\quad + |X^{[1]}_t|\Big|\partial_x\sigma(t, \overline{\chi}(t, X_{t_{i-1}}))\overline{\mathbf{y}}^{[2]}(t, X_{t_{i-1}}) + \overline{G}^{[2]}(t, X_{t_{i-1}})\sigma\big(t, \overline{\chi}(t, X_{t_{i-1}})\big)\Big| \\
&\leq C\Delta(t)\Big(1 + |X_{t_{i-1}}|\Big) + C\Big(\Delta(t)^{1/2} + \Delta(t)(1 + |X_{t_{i-1}}|)\Big) + C|X^{[1]}_t|\Big(1 + |X_{t_{i-1}}|\Big) \ ,
\end{aligned}$$

with some positive constant $C$. Thus, we obtain $\mathbb{E}\left[\sup_{t\in I_i}\left|\widehat{Z}^{i,[0]}_t - \widehat{Z}^{i,[0]}_{t_{i-1}}\right|^p\right] \le C_p h_i^{p/2}$ as desired. $\square$

## C.2 Approximation for $(\overline{Y}^{i,[1]}, \overline{Z}^{i,[1]})$

We now want to approximate the remaining BSDE (B.2) appeared in the decomposition of $(\overline{Y}^i, \overline{Z}^i)$. We shall see below that this can be achieved in a very simple fashion. We define the process $(\widehat{Y}^{i,[1]}, \widehat{Z}^{i,[1]})$ by

$$\widehat{Y}^{i,[1]}_t := \delta(t) f\Big(t_{i-1}, X_{t_{i-1}}, \widehat{Y}^{i,[0]}_{t_{i-1}}, \widehat{Z}^{i,[0]}_{t_{i-1}}\Big) , \tag{C.17}$$

$$\widehat{Z}^{i,[1]}_t := 0 , \tag{C.18}$$

for each period $t \in I_i, i \in \{1, \cdots, n\}$. Here, $\delta(t) = t_i - t$ as before.

**Lemma C.8.** *There exists some $(i, n)$-independent positive constant $C_p$ such that the inequality*

$$\mathbb{E}\Big[\big|\big|\overline{Y}^{i,[1]} - \widehat{Y}^{i,[1]}\big|\big|^p_{[t_{i-1},t_i]} + \Big(\int_{t_{i-1}}^{t_i} |\overline{Z}^{i,[1]}_r|^2 dr\Big)^{p/2}\Big] \le C_p h_i^{3p/2}$$

*holds for every interval $I_i, i \in \{1, \cdots, n\}$ with any $p \ge 2$.*

*Proof.* Let us put $\delta Y^{i,[1]}_t := \overline{Y}^{i,[1]}_t - \widehat{Y}^{i,[1]}_t$, $\delta Z^{i,[1]}_t := \overline{Z}^{i,[1]}_t$ for $t \in I_i$. Then, $(\delta Y^{i,[1]}, \delta Z^{i,[1]})$ is the solution of the following Lipschitz BSDE:

$$\delta Y^{i,[1]}_t = \int_t^{t_i} \delta f(r) dr - \int_t^{t_i} \delta Z^{i,[1]}_r dW_r ,$$

where $\delta f(r) := f(r, X_r, \overline{Y}^{i,[0]}_r, \overline{Z}^{i,[0]}_r) - f(t_{i-1}, X_{t_{i-1}}, \widehat{Y}^{i,[0]}_{t_{i-1}}, \widehat{Z}^{i,[0]}_{t_{i-1}})$. From Assumption 2.2 (ii), it satisfies with the positive constant $K$ that

$$\begin{aligned}
|\delta f(r)| \le\ & |f(r, X_r, \overline{Y}^{i,[0]}_r, \overline{Z}^{i,[0]}_r) - f(t_{i-1}, X_{t_{i-1}}, \overline{Y}^{i,[0]}_r, \overline{Z}^{i,[0]}_r)| \\
& + |f(t_{i-1}, X_{t_{i-1}}, \overline{Y}^{i,[0]}_r, \overline{Z}^{i,[0]}_r) - f(t_{i-1}, X_{t_{i-1}}, \widehat{Y}^{i,[0]}_r, \widehat{Z}^{i,[0]}_r)| \\
& + |f(t_{i-1}, X_{t_{i-1}}, \widehat{Y}^{i,[0]}_r, \widehat{Z}^{i,[0]}_r) - f(t_{i-1}, X_{t_{i-1}}, \widehat{Y}^{i,[0]}_{t_{i-1}}, \widehat{Z}^{i,[0]}_{t_{i-1}})| \\
\le\ & K\Big(\Delta(r)^{1/2} + (1 + |\overline{Y}^{i,[0]}_r| + |\overline{Z}^{i,[0]}_r|^2)|X_r - X_{t_{i-1}}|\Big) \\
& + K|\overline{Y}^{i,[0]}_r - \widehat{Y}^{i,[0]}_r| + K(1 + |\overline{Z}^{i,[0]}_r| + |\widehat{Z}^{i,[0]}_r|)|\overline{Z}^{i,[0]}_r - \widehat{Z}^{i,[0]}_r| \\
& + K|\widehat{Y}^{i,[0]}_r - \widehat{Y}^{i,[0]}_{t_{i-1}}| + K(1 + |\widehat{Z}^{i,[0]}_r| + |\widehat{Z}^{i,[0]}_{t_{i-1}}|)|\widehat{Z}^{i,[0]}_r - \widehat{Z}^{i,[0]}_{t_{i-1}}| .
\end{aligned}$$

From Lemma B.1, we know that $||\overline{Y}^{i,[0]}||_{\mathcal{S}^\infty[t_{i-1},t_i]} + ||\overline{Z}^{i,[0]}||_{\mathcal{S}^p[t_{i-1},t_i]} \le C_p$ for any $p \ge 2$. From (C.15), (C.16), Lemma C.2, and the boundedness properties of $(\overline{y}, \overline{\mathbf{y}}^{[i]}, \overline{G}^{[2]})$ shown in the proof for Lemma C.5, a similar inequality $||\widehat{Y}^{i,[0]}||_{\mathcal{S}^p[t_{i-1},t_i]} + ||\widehat{Z}^{i,[0]}||_{\mathcal{S}^p[t_{i-1},t_i]} \le C_p$ holds. The continuity property of the Lipschitz SDE $\mathbb{E}\Big[\sup_{t\in I_i} |X_t - X_{t_{i-1}}|^p\Big] \le C_p h_i^{p/2}$ is also well-known to hold for any $p \ge 2$. Then,

$$\begin{aligned}
\mathbb{E}\Big[||\delta Y^{i,[1]}||^p_{I_i} + \Big(\int_{t_{i-1}}^{t_i} |\delta Z^{i,[1]}_r|^2 dr\Big)^{p/2}\Big] &\le C_p \mathbb{E}\Big[\Big(\int_{t_{i-1}}^{t_i} |\delta f(r)| dr\Big)^p\Big] \\
&\le C_p \Bigg\{ h_i^{3p/2} + h_i^p \mathbb{E}\Big[1 + ||\overline{Y}^{i,[0]}||^{2p}_{I_i} + ||\overline{Z}^{i,[0]}||^{4p}_{I_i}\Big]^{\frac{1}{2}} \mathbb{E}\Big[\sup_{t\in I_i} |X_t - X_{t_{i-1}}|^{2p}\Big]^{\frac{1}{2}} \\
&\quad + h_i^p \Big(\mathbb{E}\Big[||\overline{Y}^{i,[0]} - \widehat{Y}^{i,[0]}||^p_{I_i}\Big] + \mathbb{E}\Big[\sup_{t\in I_i} |\widehat{Y}^{i,[0]}_t - \widehat{Y}^{i,[0]}_{t_{i-1}}|^p\Big]\Big)
\end{aligned}$$

$$+\mathbb{E}\Big[1+||\overline{Z}^{i,[0]}||_{I_i}^{2p}+||\widehat{Z}^{i,[0]}||_{I_i}^{2p}\Big]^{\frac{1}{2}}\mathbb{E}\Big[\Big(h_i\int_{t_{i-1}}^{t_i}|\overline{Z}_r^{i,[0]}-\widehat{Z}_r^{i,[0]}|^2dr\Big)^p\Big]^{\frac{1}{2}}$$

$$+h_i^p\mathbb{E}\Big[1+||\widehat{Z}^{i,[0]}||_{I_i}^{2p}+|\widehat{Z}_{t_{i-1}}^{i,[0]}|^{2p}\Big]^{\frac{1}{2}}\mathbb{E}\Big[\sup_{t\in I_i}|\widehat{Z}_t^{i,[0]}-\widehat{Z}_{t_{i-1}}^{i,[0]}|^{2p}\Big]^{\frac{1}{2}}\Bigg\}\leq C_ph_i^{3p/2}$$

follows from Corollary C.1 and Lemma C.7. □

The main result regarding the short-term approximation can be summarized as the next theorem.

**Theorem C.1.** *Under Assumptions 2.1, 2.2 and Assumption 3.1(i), the process* $(\widehat{Y}^i,\widehat{Z}^i)$ *defined by* $(\widehat{Y}_t^i:=\widehat{Y}_t^{i,[0]}+\widehat{Y}_t^{i,[1]},\quad \widehat{Z}_t^i:=\widehat{Z}_t^{i,[0]},t\in I_i)$ *is the short-term approximation of the solution* $(\overline{Y}^i,\overline{Z}^i)$ *of the qg-BSDE (3.1) and satisfies, with some* $(i,n)$*-independent positive constant* $C_p$*, that*

$$\mathbb{E}\Big[||\overline{Y}^i-\widehat{Y}^i||_{[t_{i-1},t_i]}^p+\Big(\int_{t_{i-1}}^{t_i}|\overline{Z}_r^i-\widehat{Z}_r^i|^2dr\Big)^{p/2}\Big]\leq C_ph_i^{3p/2}\ ,$$

*for every period* $I_i, i\in\{1,\cdots,n\}$ *and* $\forall p\geq 2$.

*Proof.* It follows directly from Proposition B.1, Corollary C.1 and Lemma C.8. □


### Acknowledgement

The research is partially supported by Center for Advanced Research in Finance (CARF).


# References


[1] Ankirchner, S., Imkeller, P. and Dos Reis, G., 2007 *Classical and Variational Differentiability of BSDEs with Quadratic Growth*, Electronic Journal of Probability, Vol. 12, 1418-1453.

[2] Barrieu, P. and El Karoui, N., 2013, *Monotone stability of quadratic semimartingales with applications to unbounded general quadratic BSDEs*, The Annals of Probability, Vol. 41, No. 3B, 1831-1863.

[3] Barthelmann, V., Novak, E. and Ritter, K., 2000, *High dimensional polynomial interpolation on sparse grids*, Advances in Computational Mathematics 12, 273-288.

[4] Bender, C. and Denk, R., 2007, *A forward scheme for backward SDEs*, Stochastic Processes and their Applications, 117, 1793-1812.

[5] Bender, C. and Steiner, J., 2012, *Least-squares Monte Carlo for Backward SDEs*, Numerical Methods in Finance (edited by Carmona et al.), 257-289, Springer, Berlin.

[6] Bally, V., and Pagès, G., 2003, *A quantization algorithm for solving discrete time multidimensional optimal stopping problems*, Bernoulli, 6, 1003-1049.

[7] Bianchetti, M and M Morini (eds.) (2013), Interest Rate Modeling after the Financial Crisis, *Risk Books, London.*

[8] Bismut, J.M., 1973, *Conjugate convex functions in optimal stochastic control*, J. Math. Anal. Apl. 44, 384-404.

[9] Bouchard, B. and Chassagneux, J-F., 2008, *Discrete-time approximation for continuously and discretely reflected BSDEs*, Stochastic Processes and their Applications, 118, 2269-2293.

[10] Bouchard, B. and Elie, R., 2008, *Discrete-time approximation of decoupled Forward-Backward SDE with jumps*, Stochastic Processes and their Applications, 118, 53-75.

[11] Bouchard, B. and Touzi, N., 2004, *Discrete-time approximation and Monte-Carlo simulation of backward stochastic differential equations*, Stochastic Processes and their Applications, 111, 175-206.

[12] Briand, P., Delyon, B., Hu, Y., Pardoux, E. and Stica, L., 2003, *$L^p$ solutions of backward stochastic differential equations*, Stochastic Processes and their Applications, 108, 109-129.

[13] Briand, P. and Confortola, F., 2008, *BSDEs with stochastic Lipschitz condition and quadratic PDEs in Hilbert spaces*, Stochastic Processes and their Applications, 118, pp. 818-838.

[14] Brigo, D, M Morini and A Pallavicini (2013), Counterparty Credit Risk, Collateral and Funding, *Wiley, West Sussex.*

[15] Chassagneux, J.F. and Richou, A., 2016, *Numerical Simulation of Quadratic BSDEs*, Annals of Applied Probabilities, Vol. 26, No. 1, 262-304.

[16] Chassagneux, J.F. and Richou, A., 2016, *Rate of convergence for the discrete-time approximation of reflected BSDEs arising in switching problems*, working paper, arXiv:1602:00015.

[17] Crépey, S, T Bielecki with an introductory dialogue by D Brigo (2014), Counterparty Risk and Funding, *CRC press, NY.*

[18] Crépey, S. and Song, S., 2015, *Counterparty Risk and Funding: Immersion and Beyond*, forthcoming in Finance and Stochastics.

[19] Crisan, D. and Monolarakis, K., 2014, *Second order discretization of backward SDEs and simulation with the cubature method*, Annals of Applied Probability, Vol. 24, No. 2, 652-678.

[20] Cvitanić, J. and Zhang, J., 2013, *Contract theory in continuous-time methods*, Springer, Berlin.

[21] Delong, L., 2013, *Backward Stochastic Differential Equations with Jumps and Their Actuarial and Financial Applications*, Springer-Verlag, LN.

[22] Delarue, F. and Menozzi, S., 2006, *A forward-backward stochastic algorithm for quasi-linear PDEs*, Annals of Applied Probability, Vol. 16, No. 1, 140-184.

[23] Delarue, F. and Menozzi, S., 2008, *An interpolated stochastic algorithm for quasi-linear PDEs*, Mathematics of Computation, Vol. 77, 261, pp 125-158.

[24] El Karoui, N. and Mazliak, L. (eds.), 1997, *Backward stochastic differential equations*, Addison Wesley Longman Limited, U.S..

[25] El Karoui, N., Peng, S. and Quenez, M.C., 1997, *Backward stochastic differential equations in Finance*, Mathematical Finance, Vol. 7, No. 1, 1-71.

[26] Fujii, M. and Takahashi, A., 2012, *Analytical approximation for non-linear FBSDEs with perturbation scheme*, International Journal of Theoretical and Applied Finance, 15, 5, 1250034 (24).

[27] Fujii, M., 2014, *Momentum-space approach to asymptotic expansion for stochastic filtering*, Annals of the Institute of Statistical Mathematics, Vol. 66, 93-120.

[28] Fujii, M. and Takahashi, A., 2015, *Perturbative expansion technique for non-linear FBSDEs with interacting particle method*, Asia-Pacific Financial Markets, Vol. 22, 3, 283-304.

[29] Fujii, M. and Takahashi, A., 2015, *Quadratic-exponential growth BSDEs with Jumps and their Malliavin's Differentiability*, Working paper, CARF-F-376. Available in arXiv.

[30] Fujii, M. and Takahashi, A., 2015, *Asymptotic Expansion for Forward-Backward SDEs with Jumps*, Working paper, CARF-F-372. Available in arXiv.

[31] Gobet, E., Lemor, J-P. and Warin, X., 2005, *A regression-based Monte Carlo method to solve backward stochastic differential equations*, Annals of Applied Probability, Vol. 15, No. 3, 2172-2202.

[32] Imkeller, P. and Dos Reis, G., 2010, *Path regularity and explicit convergence rate for BSDEs with truncated quadratic growth*, Stochastic Processes and their Applications, 120, 348-379. *Corrigendum for Theorem 5.5, 2010, 120, 2286-2288.*

[33] Kazamaki, N., 1994, *Continuous exponential martingales and BMO*, Lecture Notes in Mathematics, vol. 1579, Springer-Verlag, Berlin.

[34] Kloeden, P. and Platen, E., 1992, *Numerical solution of stochastic differential equations*, Applications of Mathematics (New York) 23. Springer, Berlin.

[35] Kobylanski, M., 2000, *Backward stochastic differential equations and partial differential equations with quadratic growth*, The Annals of Probability, Vol. 28, No. 2, 558-602.

[36] Kunitomo, N. and Takahashi, A., 2003, *On Validity of the Asymptotic Expansion Approach in Contingent Claim Analysis*, Annals of Applied Probability, 13, No.3, 914-952.

[37] Ma, J., Protter, P. and Yong, J., 1994, *Solving forward-backward stochastic differential equations explicitly-a four step scheme*, Probability Theory and Related Fields, 98, 339-359.

[38] Ma, J. and Yong, J., 2000, *Forward-backward stochastic differential equations and their applications*,Springer, Berlin.

[39] Ma, J. and Zhang, J., 2002, *Representation theorems for backward stochastic differential equations*, Annals of Applied probability, 12, 4, 1390-1418.

[40] Ma, X. and Zabaras, N., 2009, *An adaptive hierarchical sparse grid collocation algorithm for the solution of stochastic differential equations*, Journal of Computational Physics, 228, 3084-3113.

[41] Pagès, G. and Sagna, A., 2015, *Improved error bounds for quantization based numerical schemes for BSDE and nonlinear filtering*, arXiv:1510:01048.

[42] Pardoux, E. and Peng, S., 1990, *Adapted solution of a backward stochastic differential equations*, Systems Control Lett., 14, 55-61.

[43] Pardoux, E. and Rascanu, A., 2014, *Stochastic Differential Equations, Backward SDEs, Partial Differential Equations*, Springer International Publishing, Switzerland.

[44] Sauer, T., 1995, *Polynomial interpolation of minimal degree*, Numerische Mathematik, 78, 59-85.

[45] Yong, J. and Zhou, X.Y., 1999, *Stochastic Controls: Hamiltonian systems and HJB equations*, Springer, NY.

[46] Takahashi, A., 1999, *An Asymptotic Expansion Approach to Pricing Contingent Claims,* Asia-Pacific Financial Markets, 6, 115-151.

[47] Takahashi, A. and Yamada, T., 2015, *An asymptotic expansion of forward-backward SDEs with a perturbed driver*, Forthcoming in International Journal of Financial Engineering.

[48] Takahashi, A. and Yamada, T., 2016, *A weak approximation with asymptotic expansion and multidimensional Malliavin weights*, Annals of Applied Probability, Vol. 26, No. 2, 818-856.

[49] Zhang, G., Gunzburger, M. and Zhao, W., 2013, *A sparse-grid method for multi-dimensional backward stochastic differential equations*, Journal of Computational Mathematics, 31, pp. 221-248.

[50] Zhang, J., 2004, *A numerical scheme for BSDEs*, Annals of Applied Probability, Vol 14, No. 1, 459-488.

[51] Zhang, J., 2001, *Some fine properties of backward stochastic differential equations*, Ph.D Thesis, Purdue University.