

# MEAN SQUARE RATE OF CONVERGENCE FOR RANDOM WALK APPROXIMATION OF FORWARD-BACKWARD SDES

CHRISTEL GEISS,\* *University of Jyväskylä*

CÉLINE LABART,\*\* *Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, LAMA*

ANTTI LUOTO,\* *University of Jyväskylä*

## Abstract

Let  $(Y, Z)$  denote the solution to a forward-backward stochastic differential equation (FBSDE). If one constructs a random walk  $B^n$  from the underlying Brownian motion  $B$  by Skorokhod embedding, one can show  $L_2$ -convergence of the corresponding solutions  $(Y^n, Z^n)$  to  $(Y, Z)$ . We estimate the rate of convergence based on smoothness properties, especially for a terminal condition function in  $C^{2,\alpha}$ . The proof relies on an approximative representation of  $Z^n$  and uses the concept of discretized Malliavin calculus. Moreover, we use growth and smoothness properties of the partial differential equation associated to the FBSDE, as well as of the finite difference equations associated to the approximating stochastic equations. We derive these properties by probabilistic methods.

*Keywords:* Backward stochastic differential equations; approximation scheme; finite difference equation; convergence rate; random walk approximation

2010 Mathematics Subject Classification: Primary 60H10

Secondary 60H35; 60G50; 60H30

## 1. Introduction

Let  $(\Omega, \mathcal{F}, \mathbb{P})$  be a complete probability space carrying the standard Brownian motion  $B = (B_t)_{t \geq 0}$  and assume that  $(\mathcal{F}_t)_{t \geq 0}$  is the augmented natural filtration. Let  $(Y, Z)$  be the solution of the forward-backward stochastic differential equation (FBSDE)

$$\begin{aligned} X_s &= x + \int_0^s b(r, X_r) dr + \int_0^s \sigma(r, X_r) dB_r, \\ Y_s &= g(X_T) + \int_s^T f(r, X_r, Y_r, Z_r) dr - \int_s^T Z_r dB_r, \quad 0 \leq s \leq T. \end{aligned} \quad (1)$$

Let  $(Y^n, Z^n)$  be the solution of the FBSDE if the Brownian motion  $B$  is replaced by a scaled random walk  $B^n$  given by

$$B_t^n = \sqrt{h} \sum_{i=1}^{\lfloor t/h \rfloor} \varepsilon_i, \quad 0 \leq t \leq T, \quad (2)$$

Received 9 May 2019; revision received 7 February 2020; accepted 11 February 2020.

\* Postal address: Department of Mathematics and Statistics, University of Jyväskylä, Finland, P.O. Box 35 (MaD) FI-40014.

\*\* Postal address: Université Grenoble Alpes, Université Savoie Mont Blanc, CNRS, LAMA, 73000 Chambéry, France.

where  $h = \frac{T}{n}$  and  $(\varepsilon_i)_{i=1,2,\dots}$  is a sequence of independent and identically distributed (i.i.d.) Rademacher random variables. Then  $(Y^n, Z^n)$  solves the discretized FBSDE

$$\begin{aligned} X_s^n &= x + \int_{(0,s]} b(r, X_{r-}^n) d[B^n]_r + \int_{(0,s]} \sigma(r, X_{r-}^n) dB_r^n, \\ Y_s^n &= g(X_T^n) + \int_{(s,T]} f(r, X_{r-}^n, Y_{r-}^n, Z_{r-}^n) d[B^n]_r - \int_{(s,T]} Z_{r-}^n dB_r^n, \quad 0 \leq s \leq T. \end{aligned} \tag{3}$$

Many authors have investigated the approximation of backward stochastic differential equations (BSDEs) using random walks, analytically as well as numerically (see, for example, [7], [26, [29], [31], [32], [33], [16]). In 2001, Briand et al. [7] showed weak convergence of  $(Y^n, Z^n)$  to  $(Y, Z)$  for a Lipschitz continuous generator  $f$  and a terminal condition in  $L_2$ . The rate of convergence of this method remained an open problem. Bouchard and Touzi in [6] and Zhang in [42] proposed instead of random walks an approach based on the dynamic programming equation, for which they established a rate of convergence. But this approach involves conditional expectations. Various methods to approximate these conditional expectations have been developed ([23], [17], [12]). Also, forward methods have been introduced to approximate (1): a branching diffusion method ([24]), a multilevel Picard approximation ([39]), and Wiener chaos expansion ([8]). Many extensions of (1) have been considered, among them schemes for reflected BSDEs ([3], [14]), high order schemes ([9], [10]), fully-coupled BSDEs ([18], [5]), quadratic BSDEs ([13]), BSDEs with jumps ([21]), and McKean–Vlasov BSDEs ([1], [15], [11]).

The aim of this paper is to study the rate of the  $L_2$ -approximation of  $(Y_t^n, Z_t^n)$  to  $(Y_t, Z_t)$  when  $X$  satisfies (1). For this, we generate the random walk  $B^n$  by Skorokhod embedding from the Brownian motion  $B$ . In this case the  $L_p$ -convergence of  $B^n$  to  $B$  is of order  $h^{\frac{1}{4}}$  for any  $p > 0$ . The special case  $X = B$  has already been studied in [22], assuming a locally  $\alpha$ -Hölder continuous terminal function  $g$  and a Lipschitz continuous generator. An estimate for the rate of convergence was obtained which is of order  $h^{\frac{\alpha}{4}}$  for the  $L_2$ -norm of  $Y_t^n - Y_t$ , and of order  $\frac{h^{\frac{\alpha}{4}}}{\sqrt{T-t}}$  for the  $L_2$ -norm of  $Z_t^n - Z_t$ .

In the present paper, where we assume that  $X$  is a solution of the stochastic differential equation (SDE) in (1), rather strong conditions on the smoothness and boundedness of  $f$  and  $g$  and also of  $b$  and  $\sigma$  are needed. In Theorem 3.1, the main result of the paper, we show that the convergence rate for  $(Y_t^n, Z_t^n)$  to  $(Y_t, Z_t)$  in  $L_2$  is of order  $h^{\frac{1}{4} \wedge \frac{\alpha}{2}}$  provided that  $g''$  is locally  $\alpha$ -Hölder continuous. To the best of our knowledge, these are the first cases in which a convergence rate for the approximation of FBSDEs using random walks has been obtained.

**Remark 1.1.** For the diffusion setting—in contrast to the case  $X = B$ —we can derive the convergence rate for  $(Y_t^n, Z_t^n)$  to  $(Y_t, Z_t)$  in  $L_2$  only under strong smoothness conditions on the coefficients, which include also that  $g''$  is locally  $\alpha$ -Hölder continuous (see Assumption 2.3 below). These requirements appear to be necessary. This becomes visible in Subsection 2.2.2 where we introduce a discretized Malliavin weight to obtain a representation  $\hat{Z}^n$  for  $Z^n$ . While it holds that  $\hat{Z}^n = Z^n$  when  $X = B$ , in our case  $\hat{Z}^n$  does not coincide with  $Z^n$ . However, one can show that the difference  $\hat{Z}_t^n - Z_t^n$  converges to 0 in  $L_2$  as  $n \rightarrow \infty$  using a Hölder continuity property (see (63) in Remark 4.1) for the space derivative of the generator in (3). For this Hölder continuity property to hold one needs enough smoothness in space from the solution  $u^n$  to the finite difference equation associated to the discretized FBSDE (3). Provided that Assumption 2.3 holds, we show the smoothness properties for  $u^n$  in Proposition 4.2, applying methods known for Lévy driven BSDEs.

The paper is organized as follows. Section 2 contains the setting, the main assumptions, and the approximative representation  $\hat{Z}^n$  of  $Z^n$ . Our main results about the approximation rate for the case of no generator (i.e.  $f = 0$ ) and for the general case are in Section 3. One can see that in contrast to what is known for time discretization schemes, for random walk schemes the Lipschitz generator seems to cause more difficulties than the terminal condition: while in the case  $f = 0$  we need that  $g'$  is locally  $\alpha$ -Hölder continuous, in the case  $f \neq 0$  this property is required for  $g''$ . In Section 4 we recall some needed facts about Malliavin weights, the regularity of solutions to BSDEs, and properties of the associated partial differential equations (PDEs). Finally, we sketch how to prove growth and smoothness properties of solutions to the finite difference equation associated to the discretized FBSDE. Section 5 contains technical results which mainly arise from the fact that the construction of the random walk by Skorokhod embedding forces us to compare our processes on different ‘timelines’, one coming from the stopping times of the Skorokhod embedding, the other from the equidistant deterministic times due to the quadratic variation process  $[B^n]$ .

## 2. Preliminaries

### 2.1. The SDE and its approximation scheme

We introduce

$$X_t = x + \int_0^t b(s, X_s)ds + \int_0^t \sigma(s, X_s)dB_s, \quad 0 \leq t \leq T,$$

and its discretized counterpart

$$X_{t_k}^n = x + h \sum_{j=1}^k b(t_j, X_{t_{j-1}}^n) + \sqrt{h} \sum_{j=1}^k \sigma(t_j, X_{t_{j-1}}^n)\varepsilon_j, \quad t_j := j\frac{T}{n}, \quad j = 0, \dots, n, \quad (4)$$

where  $(\varepsilon_i)_{i=1,2,\dots}$  is a sequence of i.i.d. Rademacher random variables. Letting

$$\mathcal{G}_k := \sigma(\varepsilon_i: 1 \leq i \leq k) \quad \text{with} \quad \mathcal{G}_0 := \{\emptyset, \Omega\}, \quad (5)$$

it follows that the associated discrete-time random walk  $(B_{t_k}^n)_{k=0}^n$  is  $(\mathcal{G}_k)_{k=0}^n$ -adapted. Recall (2) and  $h = \frac{T}{n}$ . If we extend the sequence  $(X_{t_k}^n)_{k \geq 0}$  to a process in continuous time by defining  $X_t^n := X_{t_k}^n$  for  $t \in [t_k, t_{k+1})$ , it is the solution of the forward SDE (3).

We formulate our first assumptions. Assumption 2.1(ii) will not be used explicitly for our estimates, but it is required for Theorem 4.1 below.

#### Assumption 2.1.

- (i)  $b, \sigma \in C_b^{0,2}([0, T] \times \mathbb{R})$ , in the sense that the derivatives of order  $k = 0, 1, 2$  with respect to the space variable are continuous and bounded on  $[0, T] \times \mathbb{R}$ .
- (ii) The first and second derivatives of  $b$  and  $\sigma$  with respect to the space variable are assumed to be  $\gamma$ -Hölder continuous (for some  $\gamma \in (0, 1]$ , with respect to the parabolic metric  $d((t, x), (\bar{t}, \bar{x})) = (|t - \bar{t}| + |x - \bar{x}|^2)^{\frac{1}{2}}$ ) on all compact subsets of  $[0, T] \times \mathbb{R}$ .
- (iii)  $b, \sigma$  are  $\frac{1}{2}$ -Hölder continuous in time, uniformly in space.
- (iv)  $\sigma(t, x) \geq \delta > 0$  for all  $(t, x)$ .

**Assumption 2.2.**

(i)  $g$  is locally Hölder continuous with order  $\alpha \in (0, 1]$  and polynomially bounded in the following sense: there exist  $p_0 \geq 0, C_g > 0$  such that

$$\forall(x, \bar{x}) \in \mathbb{R}^2, \quad |g(x) - g(\bar{x})| \leq C_g(1 + |x|^{p_0} + |\bar{x}|^{p_0})|x - \bar{x}|^\alpha. \tag{6}$$

(ii) The function  $(t, x, y, z) \mapsto f(t, x, y, z)$  on  $[0, T] \times \mathbb{R}^3$  satisfies

$$|f(t, x, y, z) - f(\bar{t}, \bar{x}, \bar{y}, \bar{z})| \leq L_f \left( \sqrt{t - \bar{t}} + |x - \bar{x}| + |y - \bar{y}| + |z - \bar{z}| \right). \tag{7}$$

Notice that (6) implies

$$|g(x)| \leq K(1 + |x|^{p_0+1}) =: \Psi(x), \quad x \in \mathbb{R}, \tag{8}$$

for some  $K > 0$ . From the continuity of  $f$  we conclude that

$$K_f := \sup_{0 \leq t \leq T} |f(t, 0, 0, 0)| < \infty.$$

**Notation:**

- $\|\cdot\|_p := \|\cdot\|_{L_p(\mathbb{P})}$  for  $p \geq 1$ . For  $p = 2$  we write simply  $\|\cdot\|$ .
- If  $a$  is a function,  $C(a)$  represents a generic constant which depends on  $a$  and possibly also on its derivatives.
- $\mathbb{E}_{0,x} := \mathbb{E}(\cdot | X_0 = x)$ .
- Let  $\phi$  be a  $C^{0,1}([0, T] \times \mathbb{R})$  function. Then  $\phi_x$  denotes  $\partial_x \phi$ , the partial derivative of  $\phi$  with respect to  $x$ .

**2.2. The FBSDE and its approximation scheme**

Recall the FBSDE (1) and its approximation (3). The backward equation in (3) can equivalently be written in the form

$$Y_{t_k}^n = g(X_T^n) + h \sum_{m=k}^{n-1} f(t_{m+1}, X_{t_m}^n, Y_{t_m}^n, Z_{t_m}^n) - \sqrt{h} \sum_{m=k}^{n-1} Z_{t_m}^n \varepsilon_{m+1}, \quad 0 \leq k \leq n, \tag{9}$$

if one puts  $X_r^n := X_{t_m}^n, Y_r^n := Y_{t_m}^n$ , and  $Z_r^n := Z_{t_m}^n$  for  $r \in [t_m, t_{m+1})$ .

**Remark 2.1.** Equations (3) and (9) do not contain any martingale orthogonal to the random walk  $B^n$ , since we are in a special case where the orthogonal martingale is zero (see [7, p. 3] or [34, Proposition 1.7.5]). Indeed, for the symmetric simple random walk  $B^n$  the predictable representation property holds; i.e. for any  $\mathcal{G}_n$ -measurable (see (5)) random variable  $\xi = F(\varepsilon_1, \dots, \varepsilon_n)$  there exists a representation

$$F(\varepsilon_1, \dots, \varepsilon_n) = c + \sum_{m=1}^n h_m \varepsilon_m,$$

where  $c \in \mathbb{R}$  and  $h_m$  is  $\mathcal{G}_{m-1}$ -measurable for  $m = 1, \dots, n$ . To see this, consider

$$F(\varepsilon_1, \dots, \varepsilon_n) = \mathbb{E}[F(\varepsilon_1, \dots, \varepsilon_n)] + \sum_{m=1}^n \left( \mathbb{E}[F(\varepsilon_1, \dots, \varepsilon_n) | \mathcal{G}_m] - \mathbb{E}[F(\varepsilon_1, \dots, \varepsilon_n) | \mathcal{G}_{m-1}] \right).$$

Put  $c = \mathbb{E}[F(\varepsilon_1, \dots, \varepsilon_n)]$ . Our aim is to determine a  $\mathcal{G}_{m-1}$ -measurable  $h_m$  such that

$$\mathbb{E}[F(\varepsilon_1, \dots, \varepsilon_n) | \mathcal{G}_m] - \mathbb{E}[F(\varepsilon_1, \dots, \varepsilon_n) | \mathcal{G}_{m-1}] = h_m \varepsilon_m.$$

We define

$$F_m(\varepsilon_1, \dots, \varepsilon_m) := \mathbb{E}[F(\varepsilon_1, \dots, \varepsilon_n) | \mathcal{G}_m].$$

By the tower property it holds that

$$\begin{aligned} & F_m(\varepsilon_1, \dots, \varepsilon_m) - F_{m-1}(\varepsilon_1, \dots, \varepsilon_{m-1}) \\ &= F_m(\varepsilon_1, \dots, \varepsilon_m) - \mathbb{E}[F_m(\varepsilon_1, \dots, \varepsilon_m) | \mathcal{G}_{m-1}] \\ &= F_m(\varepsilon_1, \dots, \varepsilon_m) - \frac{F_m(\varepsilon_1, \dots, \varepsilon_{m-1}, 1) + F_m(\varepsilon_1, \dots, \varepsilon_{m-1}, -1)}{2} \\ &= \frac{F_m(\varepsilon_1, \dots, \varepsilon_{m-1}, 1) - F_m(\varepsilon_1, \dots, \varepsilon_{m-1}, -1)}{2} \varepsilon_m; \end{aligned}$$

hence

$$h_m = \frac{F_m(\varepsilon_1, \dots, \varepsilon_{m-1}, 1) - F_m(\varepsilon_1, \dots, \varepsilon_{m-1}, -1)}{2}.$$

One can derive an equation for  $Z^n = (Z_{t_k}^n)_{k=0}^{n-1}$  if one multiplies (9) by  $\varepsilon_{k+1}$  and takes the conditional expectation with respect to  $\mathcal{G}_k$ , so that

$$Z_{t_k}^n = \frac{\mathbb{E}_k^{\mathcal{G}}(g(X_T^n) \varepsilon_{k+1})}{\sqrt{h}} + \mathbb{E}_k^{\mathcal{G}}\left(\sqrt{h} \sum_{m=k+1}^{n-1} f(t_{m+1}, X_{t_m}^n, Y_{t_m}^n, Z_{t_m}^n) \varepsilon_{k+1}\right), \quad 0 \leq k \leq n-1, \tag{10}$$

where  $\mathbb{E}_k^{\mathcal{G}} := \mathbb{E}(\cdot | \mathcal{G}_k)$ .

**Remark 2.2.** For  $n$  large enough, the BSDE (3) has a unique solution  $(Y^n, Z^n)$  (see [36, Proposition 1.2]), and  $(Y_{t_k}^n, Z_{t_k}^n)_{k=0}^{n-1}$  is adapted to the filtration  $(\mathcal{G}_k)_{k=0}^{n-1}$ .

2.2.1. *Representation for Z.* We will use the following representation for  $Z$ , due to Ma and Zhang (see [30, Theorem 4.2]):

$$Z_t = \mathbb{E}_t\left(g(X_T) N_T^t + \int_t^T f(s, X_s, Y_s, Z_s) N_s^t ds\right) \sigma(t, X_t), \quad 0 \leq t \leq T, \tag{11}$$

where  $\mathbb{E}_t := \mathbb{E}(\cdot | \mathcal{F}_t)$ , and for all  $s \in (t, T]$ , we have (cf. Lemma 4.1)

$$N_s^t = \frac{1}{s-t} \int_t^s \frac{\nabla X_r}{\sigma(r, X_r) \nabla X_t} dB_r, \tag{12}$$

where  $\nabla X = (\nabla X_s)_{s \in [0, T]}$  is the variational process; i.e., it solves

$$\nabla X_s = 1 + \int_0^s b_x(r, X_r) \nabla X_r dr + \int_0^s \sigma_x(r, X_r) \nabla X_r dB_r, \tag{13}$$

with  $(X_s)_{s \in [0, T]}$  given in (1).

**Remark 2.3.** In the following we will assume that  $g''$  exists. In such a case we have the following representation for  $Z$ :

$$Z_t = \mathbb{E}_t\left(g'(X_T) \nabla X_T + \int_t^T f(s, X_s, Y_s, Z_s) N_s^t ds\right) \sigma(t, X_t), \quad 0 \leq t \leq T. \tag{14}$$

2.2.2. *Approximation for  $Z^n$ .* In this section we state the discrete counterpart to (11), which, in the general case of a forward process  $X$ , does not coincide with  $Z^n$  (given by (10)). In contrast to the continuous-time case, where the variational process and the Malliavin derivative are connected by  $\frac{\nabla X_t}{\nabla X_s} = \frac{D_s X_t}{\sigma(s, X_s)}$  ( $s \leq t$ ), we cannot expect equality for the corresponding expressions if we use the discretized versions of the processes  $(\nabla X_t)_t$  and  $(D_s X_t)_{s \leq t}$  introduced in (16). This counterpart  $\hat{Z}^n$  to  $Z$  is a key tool in the proof of the convergence of  $Z^n$  to  $Z$ . As we will see in the proof of Theorem 3.1, the study of  $\|Z^n - Z_{t_k}\|$  goes through the study of  $\|Z^n_{t_k} - \hat{Z}^n_{t_k}\|$  and  $\|\hat{Z}^n_{t_k} - Z_{t_k}\|$ .

Before defining the discretized versions of  $(\nabla X_t)_t$  and  $(D_s X_t)_{s \leq t}$ , we briefly introduce the discretized Malliavin derivative. We refer the reader to [4] for more information on this topic.

**Definition 2.1.** (Definition of  $T_{m,+}$ ,  $T_{m,-}$  and  $\mathcal{D}_m^n$ ) For any function  $F : \{-1, 1\}^n \rightarrow \mathbb{R}$ , the mappings  $T_{m,+}$  and  $T_{m,-}$  are defined by

$$T_{m,\pm} F(\varepsilon_1, \dots, \varepsilon_n) := F(\varepsilon_1, \dots, \varepsilon_{m-1}, \pm 1, \varepsilon_{m+1}, \dots, \varepsilon_n), \quad 1 \leq m \leq n.$$

For any  $\xi = F(\varepsilon_1, \dots, \varepsilon_n)$ , the discretized Malliavin derivative is defined by

$$\mathcal{D}_m^n \xi := \frac{\mathbb{E}[\xi \varepsilon_m | \sigma((\varepsilon_l)_{l \in \{1, \dots, n\} \setminus \{m\}})]}{\sqrt{h}} = \frac{T_{m,+} \xi - T_{m,-} \xi}{2\sqrt{h}}, \quad 1 \leq m \leq n. \tag{15}$$

**Definition 2.2.** (Definition of  $\phi_x^{(k,l)}$ ) Let  $\phi$  be a  $C^{0,1}([0, T] \times \mathbb{R})$  function. We define

$$\phi_x^{(k,l)} := \frac{\mathcal{D}_k^n \phi(t_l, X_{t_{l-1}}^n)}{\mathcal{D}_k^n X_{t_{l-1}}^n} := \int_0^1 \phi_x(t_l, \vartheta T_{k,+} X_{t_{l-1}}^n + (1 - \vartheta) T_{k,-} X_{t_{l-1}}^n) d\vartheta.$$

If  $\mathcal{D}_k^n X_{t_{l-1}}^n \neq 0$ , the second ‘:=’ holds as an identity.

We are now able to define the discretized versions of  $(\nabla X_t)_t$  and  $(D_s X_t)_{s \leq t}$ .

**Definition 2.3.** (Discretized processes  $(\nabla X_m^{n,t_k,x})_{m \in \{k, \dots, n\}}$  and  $(\mathcal{D}_k^n X_m^n)_{m \in \{k, \dots, n\}}$ ) For all  $m$  in  $\{k, \dots, n\}$  we define

$$\begin{aligned} \nabla X_{t_m}^{n,t_k,x} &= 1 + h \sum_{l=k+1}^m b_x(t_l, X_{t_{l-1}}^{n,t_k,x}) \nabla X_{t_{l-1}}^{n,t_k,x} \\ &\quad + \sqrt{h} \sum_{l=k+1}^m \sigma_x(t_l, X_{t_{l-1}}^{n,t_k,x}) \nabla X_{t_{l-1}}^{n,t_k,x} \varepsilon_l, \quad 0 \leq k \leq n, \\ \mathcal{D}_k^n X_m^n &= \sigma(t_k, X_{t_{k-1}}^n) + h \sum_{l=k+1}^m b_x^{(k,l)} \mathcal{D}_k^n X_{t_{l-1}}^n \\ &\quad + \sqrt{h} \sum_{l=k+1}^m \sigma_x^{(k,l)} (\mathcal{D}_k^n X_{t_{l-1}}^n) \varepsilon_l, \quad 0 < k \leq n. \end{aligned} \tag{16}$$

**Remark 2.4.**

- (i) Although  $\nabla X_{t_m}^{n,t_k,X_{t_k}^n}$  is not equal to

$$\frac{\mathcal{D}_{k+1}^n X_m^n}{\sigma(t_{k+1}, X_{t_k}^n)},$$

we can show that the difference of these terms converges in  $L_p$  (see Lemma 5.4).

(ii) With the notation introduced above, (10) can be rewritten as

$$Z_k^n = \mathbb{E}_k^{\mathcal{G}}(\mathcal{D}_{k+1}^n g(X_T^n)) + \mathbb{E}_k^{\mathcal{G}}\left(h \sum_{m=k+1}^{n-1} \mathcal{D}_{k+1}^n f(t_{m+1}, X_{t_m}^n, Y_{t_m}^n, Z_{t_m}^n)\right). \tag{17}$$

In order to define the discrete counterpart to (11), we first define the discrete counterpart to  $(N_s^t)_{s \in [t, T]}$  given in (12):

$$N_{t_\ell}^{n, t_k} := \sqrt{h} \sum_{m=k+1}^{\ell} \frac{\nabla X_{t_{m-1}}^{n, t_k, X_{t_k}^n}}{\sigma(t_m, X_{t_{m-1}}^n)} \frac{\varepsilon_m}{t_\ell - t_k}, \quad k < \ell \leq n. \tag{18}$$

Notice that there is some constant  $\widehat{\kappa}_2 > 0$  depending on  $b, \sigma, T, \delta$  such that

$$\left(\mathbb{E}_k^{\mathcal{G}} |N_{t_\ell}^{n, t_k}|^2\right)^{\frac{1}{2}} \leq \frac{\widehat{\kappa}_2}{(t_\ell - t_k)^{\frac{1}{2}}}, \quad 0 \leq k < \ell \leq n. \tag{19}$$

**Definition 2.4.** (Discrete counterpart to (14).) Let the process  $\widehat{Z}^n = (\widehat{Z}_{t_k}^n)_{k=0}^{n-1}$  be defined by

$$\widehat{Z}_{t_k}^n := \mathbb{E}_k^{\mathcal{G}}(\mathcal{D}_{k+1}^n g(X_T^n)) + \mathbb{E}_k^{\mathcal{G}}\left(h \sum_{m=k+1}^{n-1} f(t_{m+1}, X_{t_m}^n, Y_{t_m}^n, Z_{t_m}^n) N_{t_m}^{n, t_k}\right) \sigma(t_{k+1}, X_{t_k}^n). \tag{20}$$

**Remark 2.5.** In (20) the approximate expression  $\mathbb{E}_k^{\mathcal{G}}(g(X_T^n) N_{t_n}^{n, t_k} \sigma(t_{k+1}, X_{t_k}^n))$  also could have been used, but since we will assume that  $g''$  exists, we work with the correct term.

The study of the convergence  $\mathbb{E}_{0,x}^{\mathcal{G}} |Z_{t_k}^n - \widehat{Z}_{t_k}^n|^2$  requires stronger assumptions on the coefficients  $b, \sigma, f$ , and  $g$ .

**Assumption 2.3.** Assumptions 2.1 and 2.2 hold. Additionally, we assume that all first and second derivatives with respect to the variables  $x, y, z$  of  $b(t, x), \sigma(t, x)$ , and  $f(t, x, y, z)$  exist and are bounded Lipschitz functions with respect to these variables, uniformly in time. Moreover,  $g''$  satisfies (6).

**Proposition 2.1.** If Assumption 2.3 holds, then

$$\mathbb{E}_{0,x}^{\mathcal{G}} |Z_{t_k}^n - \widehat{Z}_{t_k}^n|^2 \leq C_{2.1} \widehat{\Psi}^2(x) h^\alpha,$$

where  $\mathbb{E}_{0,x}^{\mathcal{G}} := \mathbb{E}^{\mathcal{G}}(\cdot | X_0 = x)$ , the function  $\widehat{\Psi}$  is defined in (62) below, and  $C_{2.1}$  depends on  $b, \sigma, f, g, T, p_0$ , and  $\delta$ .

*Proof.* According to [7, Proposition 5.1] one has the representations

$$Y_{t_m}^n = u^n(t_m, X_{t_m}^n) \quad \text{and} \quad Z_{t_m}^n = \mathcal{D}_{m+1}^n u^n(t_{m+1}, X_{t_{m+1}}^n), \tag{21}$$

where  $u^n$  is the solution of the finite difference equation (44) with terminal condition  $u^n(t_n, x) = g(x)$ . Notice that by the definition of  $\mathcal{D}_{m+1}^n$  in (15) the expression  $\mathcal{D}_{m+1}^n u^n(t_{m+1}, X_{t_{m+1}}^n)$  depends in fact on  $X_{t_m}^n$ . Hence we can put

$$\begin{aligned} f(t_{m+1}, X_{t_m}^n, Y_{t_m}^n, Z_{t_m}^n) &= f(t_{m+1}, X_{t_m}^n, u^n(t_m, X_{t_m}^n), \mathcal{D}_{m+1}^n u^n(t_{m+1}, X_{t_{m+1}}^n)) \\ &=: F^n(t_{m+1}, X_{t_m}^n). \end{aligned}$$

From (20) and (17) we conclude the following (we use  $\mathbb{E} := \mathbb{E}_{0,x}^{\mathcal{G}}$  for  $\|\cdot\|$ ):

$$\begin{aligned} & \|Z_{t_k}^n - \hat{Z}_{t_k}^n\| \\ &= \left\| \mathbb{E}_k^{\mathcal{G}} \left( h \sum_{m=k+1}^{n-1} \mathcal{D}_{k+1}^n f(t_{m+1}, X_{t_m}^n, Y_{t_m}^n, Z_{t_m}^n) \right) \right. \\ & \quad \left. - \mathbb{E}_k^{\mathcal{G}} \left( h \sum_{m=k+1}^{n-1} f(t_{m+1}, X_{t_m}^n, Y_{t_m}^n, Z_{t_m}^n) N_{t_m}^{n,t_k} \sigma(t_{k+1}, X_{t_k}^n) \right) \right\| \\ & \leq \sum_{m=k+1}^{n-1} \frac{h}{m-k} \sum_{\ell=k+1}^m \left\| \mathbb{E}_k^{\mathcal{G}} \left[ \mathcal{D}_{k+1}^n F^n(t_{m+1}, X_{t_m}^n) - \mathcal{D}_{\ell}^n F^n(t_{m+1}, X_{t_m}^n) \frac{\sigma(t_{k+1}, X_{t_k}^n) \nabla X_{t_{\ell-1}}^{n,t_k, X_{t_k}^n}}{\sigma(t_{\ell}, X_{t_{\ell-1}}^n)} \right] \right\|. \end{aligned}$$

With the notation introduced in Definition 2.2 applied to  $F^n$ ,

$$\begin{aligned} & \left\| \mathcal{D}_{k+1}^n F^n(t_{m+1}, X_{t_m}^n) - \mathcal{D}_{\ell}^n F^n(t_{m+1}, X_{t_m}^n) \frac{\sigma(t_{k+1}, X_{t_k}^n) \nabla X_{t_{\ell-1}}^{n,t_k, X_{t_k}^n}}{\sigma(t_{\ell}, X_{t_{\ell-1}}^n)} \right\| \\ & \leq \|(\mathcal{D}_{k+1}^n X_{t_m}^n)(F_x^{n,(k+1,m+1)} - F_x^{n,(\ell,m+1)})\| \\ & \quad + \left\| F_x^{n,(\ell,m+1)} \left( (\mathcal{D}_{k+1}^n X_{t_m}^n) - (\mathcal{D}_{\ell}^n X_{t_m}^n) \frac{\sigma(t_{k+1}, X_{t_k}^n) \nabla X_{t_{\ell-1}}^{n,t_k, X_{t_k}^n}}{\sigma(t_{\ell}, X_{t_{\ell-1}}^n)} \right) \right\| \\ & =: A_1 + A_2. \end{aligned}$$

For  $A_1$  we use Definition 2.2 again and exploit the fact that

$$x \mapsto F_x^n(t_{m+1}, x) := \partial_x f(t_{m+1}, x, u^n(t_m, x), \mathcal{D}_{m+1}^n u^n(t_{m+1}, X_{t_{m+1}}^{n,t_m,x}))$$

is locally  $\alpha$ -Hölder continuous according to (63). By Hölder’s inequality and Lemma 5.4 Parts (i) and (iii),

$$\begin{aligned} A_1 & \leq \|\mathcal{D}_{k+1}^n X_{t_m}^n\|_4 \int_0^1 \|F_x^n(t_{m+1}, \vartheta T_{k+1,+} X_{t_m}^n + (1-\vartheta)T_{k+1,-} X_{t_m}^n) \\ & \quad - F_x^n(t_{m+1}, \vartheta T_{\ell,+} X_{t_m}^n + (1-\vartheta)T_{\ell,-} X_{t_m}^n)\|_4 d\vartheta \\ & \leq C(b, \sigma, f, g, T, p_0) \hat{\Psi}(x) h^{\frac{\alpha}{2}}. \end{aligned}$$

For the estimate of  $A_2$  we notice that by our assumptions the  $L_4$ -norm of  $F_x^{n,(\ell,m+1)}$  is bounded by  $C\Psi^2(x)$ , so that it suffices to estimate

$$\begin{aligned} & \left\| (\mathcal{D}_{k+1}^n X_{t_m}^n) - (\mathcal{D}_{\ell}^n X_{t_m}^n) \frac{\sigma(t_{k+1}, X_{t_k}^n) \nabla X_{t_{\ell-1}}^{n,t_k, X_{t_k}^n}}{\sigma(t_{\ell}, X_{t_{\ell-1}}^n)} \right\|_4 \\ & \leq \left\| (\mathcal{D}_{k+1}^n X_{t_m}^n) - \frac{\sigma(t_{k+1}, X_{t_k}^n) \mathcal{D}_{\ell}^n X_{t_m}^n}{\sigma(t_{\ell}, X_{t_{\ell-1}}^n)} \frac{\mathcal{D}_{k+1}^n X_{t_{\ell-1}}^n}{\sigma(t_{k+1}, X_{t_k}^n)} \right\|_4 \\ & \quad + \left\| \frac{\sigma(t_{k+1}, X_{t_k}^n) \mathcal{D}_{\ell}^n X_{t_m}^n}{\sigma(t_{\ell}, X_{t_{\ell-1}}^n)} \left( \nabla X_{t_{\ell-1}}^{n,t_k, X_{t_k}^n} - \frac{\mathcal{D}_{k+1}^n X_{t_{\ell-1}}^n}{\sigma(t_{k+1}, X_{t_k}^n)} \right) \right\|_4. \end{aligned} \tag{22}$$



The second expression on the right-hand side of (22) is bounded by  $C(b, \sigma, T, \delta)h^{\frac{1}{2}}$  as a consequence of Lemma 5.4 Parts (ii) and (iii). To show that the first expression is also bounded by  $C(b, \sigma, T, \delta)h^{\frac{1}{2}}$ , we rewrite it using (16) and get

$$\begin{aligned} & \left| \frac{\mathcal{D}_\ell^n X_{t_m}^n}{\sigma(t_\ell, X_{t_{\ell-1}}^n)} \mathcal{D}_{k+1}^n X_{t_{\ell-1}}^n - \mathcal{D}_{k+1}^n X_{t_m}^n \right| \\ &= \left| \left( 1 + \sum_{l=\ell+1}^m \frac{\mathcal{D}_l^n X_{t_{l-1}}^n}{\sigma(t_l, X_{t_{l-1}}^n)} (b_x^{(\ell,l)} h + \sigma_x^{(\ell,l)} \sqrt{h} \varepsilon_l) \right) \right. \\ & \quad \times \left( \sigma(t_{k+1}, X_{t_k}^n) + \sum_{l=k+2}^{\ell-1} \mathcal{D}_{k+1}^n X_{t_{l-1}}^n (b_x^{(k+1,l)} h + \sigma_x^{(k+1,l)} \sqrt{h} \varepsilon_l) \right) \\ & \quad \left. - \left( \sigma(t_{k+1}, X_{t_k}^n) + \left( \sum_{l=k+2}^{\ell-1} + \sum_{l=\ell}^m \right) \mathcal{D}_{k+1}^n X_{t_{l-1}}^n (b_x^{(k+1,l)} h + \sigma_x^{(k+1,l)} \sqrt{h} \varepsilon_l) \right) \right| \\ &\leq \left| \mathcal{D}_{k+1}^n X_{t_{\ell-1}}^n (b_x^{(k+1,\ell)} h + \sigma_x^{(k+1,\ell)} \sqrt{h} \varepsilon_\ell) \right| \\ & \quad + \left| \sum_{l=\ell+1}^m \left[ \frac{\mathcal{D}_l^n X_{t_{l-1}}^n}{\sigma(t_l, X_{t_{l-1}}^n)} \mathcal{D}_{k+1}^n X_{t_{l-1}}^n - \mathcal{D}_{k+1}^n X_{t_{l-1}}^n \right] (b_x^{(\ell,l)} h + \sigma_x^{(\ell,l)} \sqrt{h} \varepsilon_l) \right| \\ & \quad + \left| \sum_{l=\ell+1}^m \mathcal{D}_{k+1}^n X_{t_{l-1}}^n \left[ b_x^{(\ell,l)} h + \sigma_x^{(\ell,l)} \sqrt{h} \varepsilon_l - (b_x^{(k+1,l)} h + \sigma_x^{(k+1,l)} \sqrt{h} \varepsilon_l) \right] \right|. \tag{23} \end{aligned}$$

We take the  $L_4$ -norm of (23) and apply the Burkholder–Davis–Gundy (BDG) inequality and Hölder’s inequality. The second term on the right-hand side of (23) will be used for Gronwall’s lemma, while the first and last terms can be bounded by  $C(b, \sigma, T)h^{\frac{1}{2}}$ , using Lemma 5.4(iii). For the last term we also use the Lipschitz continuity of  $b_x$  and  $\sigma_x$  in space and Lemma 5.4(i). □

### 3. Main results

In order to compute the mean square distance between the solution to (1) and the solution to (3), we construct the random walk  $B^n$  from the Brownian motion  $B$  by Skorokhod embedding. Let

$$\tau_0 := 0 \quad \text{and} \quad \tau_k := \inf\{t > \tau_{k-1} : |B_t - B_{\tau_{k-1}}| = \sqrt{h}\}, \quad k \geq 1. \tag{24}$$

Then  $(B_{\tau_k} - B_{\tau_{k-1}})_{k=1}^\infty$  is a sequence of i.i.d. random variables with

$$\mathbb{P}(B_{\tau_k} - B_{\tau_{k-1}} = \pm\sqrt{h}) = \frac{1}{2},$$

which means that  $\sqrt{h}\varepsilon_k \stackrel{d}{=} B_{\tau_k} - B_{\tau_{k-1}}$ . We will denote by  $\mathbb{E}_{\tau_k}$  the conditional expectation with respect to  $\mathcal{F}_{\tau_k} := \mathcal{G}_k$ . In this case we also use the notation  $\mathcal{X}_{\tau_k} := X_{t_k}^n$  for all  $k=0, \dots, n$ , so that (4) turns into

$$\mathcal{X}_{\tau_k} = x + \sum_{j=1}^k b(t_j, \mathcal{X}_{\tau_{j-1}})h + \sum_{j=1}^k \sigma(t_j, \mathcal{X}_{\tau_{j-1}})(B_{\tau_j} - B_{\tau_{j-1}}), \quad 0 \leq k \leq n.$$

**Assumption 3.1.** We assume that the random walk  $B^n$  in (3) is given by

$$B_t^n = \sum_{k=1}^{\lfloor t/h \rfloor} (B_{\tau_k} - B_{\tau_{k-1}}), \quad 0 \leq t \leq T,$$

where the  $\tau_k, k = 1, \dots, n$ , are taken from (24).

**Remark 3.1.** Note that for  $p > 0$  there exists a  $C(p) > 0$  such that for all  $k = 1, \dots, n$  it holds that

$$\frac{1}{C(p)}(t_k h)^{\frac{1}{4}} \leq (\mathbb{E}|B_{\tau_k} - B_{\tau_{k-1}}|^p)^{\frac{1}{p}} \leq C(p)(t_k h)^{\frac{1}{4}}.$$

The upper estimate is given in Lemma 5.1. For  $p \in [4, \infty)$  the lower estimate follows from [2, Proposition 5.3]. We get the lower estimate for  $p \in (0, 4)$  by choosing  $0 < \theta < 1$  and  $0 < p < p_1$  such that  $\frac{1}{4} = \frac{1-\theta}{p} + \frac{\theta}{p_1}$ . Then it holds by the log-convexity of  $L_p$  norms (see for example [35, Lemma 1.11.5]) that

$$\|B_{\tau_k} - B_{\tau_{k-1}}\|_p^{1-\theta} \geq \frac{\|B_{\tau_k} - B_{\tau_{k-1}}\|_4}{\|B_{\tau_k} - B_{\tau_{k-1}}\|_{p_1}^\theta} \geq \frac{C(4)^{-1}(t_k h)^{\frac{1}{4}}}{(C(p_1)(t_k h)^{\frac{1}{4}})^\theta} \geq (C(p)(t_k h)^{\frac{1}{4}})^{1-\theta}.$$

Since for  $t \in [t_k, t_{k+1})$  it holds that  $B_t^n = B_{\tau_k}$  and  $\|B_t - B_{t_k}\|_p \leq C(p)h^{\frac{1}{2}}$ , we have for any  $p > 0$  that

$$\sup_{0 \leq t \leq T} \|B_t^n - B_t\|_p = O(h^{\frac{1}{4}}). \tag{25}$$

Proposition 3.1 states the convergence rate of  $(Y_v^n, Z_v^n)$  to  $(Y_v, Z_v)$  in  $L_2$  when  $f = 0$ , and Theorem 3.1 generalizes this result to any  $f$  which satisfies Assumption 2.3.

**Proposition 3.1.** Let Assumptions 2.1 and 3.1 hold. If  $f = 0$  and  $g \in C^1$  is such that  $g'$  is a locally  $\alpha$ -Hölder continuous function in the sense of (6), then for all  $0 \leq v < T$ , we have (for sufficiently large  $n$ ) that

$$\mathbb{E}_{0,x}|Y_v - Y_v^n|^2 \leq C_{3.1}^y \Psi(x)^2 h^{\frac{1}{2}} \quad \text{and} \quad \mathbb{E}_{0,x}|Z_v - Z_v^n|^2 \leq C_{3.1}^z \Psi(x)^2 h^{\frac{\alpha}{2}},$$

where  $C_{3.1}^y = C(C_g, b, \sigma, T, p_0, \delta)$  and  $C_{3.1}^z = C(C_{g'}, b, \sigma, T, p_0, \delta)$ .

**Theorem 3.1.** Let Assumptions 2.3 and 3.1 be satisfied. Then for all  $v \in [0, T)$  and large enough  $n$ , we have

$$\mathbb{E}_{0,x}|Y_v - Y_v^n|^2 + \mathbb{E}_{0,x}|Z_v - Z_v^n|^2 \leq C_{3.1} \hat{\Psi}(x)^2 h^{\frac{1}{2} \wedge \alpha},$$

where  $C_{3.1} = C(b, \sigma, f, g, T, p_0, \delta)$  and  $\hat{\Psi}$  is given in (62).

**Remark 3.2.** As observed above, the filtration  $\mathcal{G}_k$  coincides with  $\mathcal{F}_{\tau_k}$ , for all  $k = 0, \dots, n$ . The expectation  $\mathbb{E}_{0,x}$  appearing in Proposition 3.1 and in Theorem 3.1 is defined on the probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ .

**Remark 3.3.** In order to avoid too much notation for the dependencies of the constants, if for example only  $g$  is mentioned and not  $C_g$ , this means that the estimate might depend also on the bounds of the derivatives of  $g$ .

From (25) one can see that the convergence rates stated in Proposition 3.1 and Theorem 3.1 are the natural ones for this approach. The results are proved in the next two sections. In both proofs, we will use the following remark.

**Remark 3.4.** Since the process  $(X_t)_{t \geq 0}$  is strong Markov, we can express conditional expectations with the help of an independent copy of  $B$  denoted by  $\tilde{B}$ . For example,  $\mathbb{E}_{\tau_k} g(X_T^n) = \tilde{\mathbb{E}}g(\tilde{\mathcal{X}}_{\tau_n}^{\tau_k, \mathcal{X}_{\tau_k}})$  for  $0 \leq k \leq n$ , where

$$\tilde{\mathcal{X}}_{\tau_n}^{\tau_k, \mathcal{X}_{\tau_k}} = \mathcal{X}_{\tau_k} + \sum_{j=k+1}^n b(t_j, \tilde{\mathcal{X}}_{\tau_{j-1}}^{\tau_k, \mathcal{X}_{\tau_k}})h + \sum_{j=k+1}^n \sigma(t_j, \tilde{\mathcal{X}}_{\tau_{j-1}}^{\tau_k, \mathcal{X}_{\tau_k}})(\tilde{B}_{\tau_{j-k}} - \tilde{B}_{\tau_{j-k-1}}) \tag{26}$$

(we define  $\tilde{\tau}_k := 0$ ,  $\tilde{\tau}_j := \inf\{t > \tilde{\tau}_{j-1} : |\tilde{B}_t - \tilde{B}_{\tilde{\tau}_{j-1}}| = \sqrt{h}\}$  for  $j \geq 1$ , and  $\tau_n := \tau_k + \tilde{\tau}_{n-k}$  for  $n \geq k$ ). In fact, to represent the conditional expectations  $\mathbb{E}_{t_k}$  and  $\mathbb{E}_{\tau_k}$ , we work here with  $\tilde{\mathbb{E}}$  and the Brownian motions  $B'$  and  $B''$ , respectively, given by

$$B'_t = B_{t \wedge t_k} + \tilde{B}_{(t-t_k)^+} \quad \text{and} \quad B''_t = B_{t \wedge \tau_k} + \tilde{B}_{(t-\tau_k)^+}, \quad t \geq 0. \tag{27}$$

**3.1. Proof of Proposition 3.1: the approximation rates for the zero generator case**

To shorten the notation, we use  $\mathbb{E} := \mathbb{E}_{0,x}$ . Let us first deal with the error of  $Y$ . If  $v$  belongs to  $[t_k, t_{k+1})$  we have  $Y_v^n = Y_{t_k}^n$ . Then

$$\mathbb{E}|Y_v - Y_v^n|^2 \leq 2(\mathbb{E}|Y_v - Y_{t_k}|^2 + \mathbb{E}|Y_{t_k} - Y_{t_k}^n|^2).$$

Using Theorem 4.1 we bound  $\|Y_v - Y_{t_k}\|$  by

$$C_{4.1}^y \Psi(x)(v - t_k)^{\frac{1}{2}} = C(C_g, b, \sigma, T, p_0, \delta) \Psi(x)(v - t_k)^{\frac{1}{2}}$$

(since  $\alpha = 1$  can be chosen when  $g$  is locally Lipschitz continuous). It remains to bound

$$\mathbb{E}|Y_{t_k} - Y_{t_k}^n|^2 = \mathbb{E}|\mathbb{E}_{t_k} g(X_T) - \mathbb{E}_{\tau_k} g(X_T^n)|^2 = \mathbb{E}|\tilde{\mathbb{E}}g(\tilde{X}_{t_n}^{t_k, X_{t_k}}) - \tilde{\mathbb{E}}g(\tilde{\mathcal{X}}_{\tau_n}^{\tau_k, \mathcal{X}_{\tau_k}})|^2.$$

By (6) and the Cauchy–Schwarz inequality (with  $\Psi_1 := C_g(1 + |\tilde{X}_{t_n}^{t_k, X_{t_k}}|^{p_0} + |\tilde{\mathcal{X}}_{\tau_n}^{\tau_k, \mathcal{X}_{\tau_k}}|^{p_0})$ ),

$$\begin{aligned} |\tilde{\mathbb{E}}g(\tilde{X}_{t_n}^{t_k, X_{t_k}}) - \tilde{\mathbb{E}}g(\tilde{\mathcal{X}}_{\tau_n}^{\tau_k, \mathcal{X}_{\tau_k}})|^2 &\leq (\tilde{\mathbb{E}}(\Psi_1 |\tilde{X}_{t_n}^{t_k, X_{t_k}} - \tilde{\mathcal{X}}_{\tau_n}^{\tau_k, \mathcal{X}_{\tau_k}}|))^2 \\ &\leq \tilde{\mathbb{E}}(\Psi_1^2) \tilde{\mathbb{E}}|\tilde{X}_{t_n}^{t_k, X_{t_k}} - \tilde{\mathcal{X}}_{\tau_n}^{\tau_k, \mathcal{X}_{\tau_k}}|^2. \end{aligned}$$

Finally, we get by Lemma 5.2(v) that

$$\mathbb{E}|Y_{t_k} - Y_{t_k}^n|^2 \leq (\mathbb{E}(\tilde{\mathbb{E}}(\Psi_1^4)))^{\frac{1}{2}} \left( \mathbb{E}(\tilde{\mathbb{E}}|\tilde{X}_{t_n}^{t_k, X_{t_k}} - \tilde{\mathcal{X}}_{\tau_n}^{\tau_k, \mathcal{X}_{\tau_k}}|^4) \right)^{\frac{1}{2}} \leq C(C_g, b, \sigma, T, p_0) \Psi(x)^2 h^{\frac{1}{2}}.$$

Let us now deal with the error of  $Z$ . We use  $\|Z_v - Z_v^n\| \leq \|Z_v - Z_{t_k}\| + \|Z_{t_k} - Z_{t_k}^n\|$  and the representation

$$Z_t = \sigma(t, X_t) \tilde{\mathbb{E}}(g'(\tilde{X}_T^{t, X_t}) \nabla \tilde{X}_T^{t, X_t})$$

(see Theorem 4.2), where

$$\begin{aligned} \tilde{X}_s^{t,x} &= x + \int_t^s b(r, \tilde{X}_r^{t,x})dr + \int_t^s \sigma(r, \tilde{X}_r^{t,x})d\tilde{B}_{r-t}, \\ \nabla \tilde{X}_s^{t,x} &= 1 + \int_t^s b_x(r, \tilde{X}_r^{t,x}) \nabla \tilde{X}_r^{t,x} dr + \int_t^s \sigma_x(r, \tilde{X}_r^{t,x}) \nabla \tilde{X}_r^{t,x} d\tilde{B}_{r-t}, \quad 0 \leq t \leq s \leq T. \end{aligned} \tag{28}$$

For the first term we get by the assumption on  $g$  and Lemma 5.2 Parts (i) and (iii) that

$$\begin{aligned} \|Z_v - Z_{t_k}\| &= \|\sigma(v, X_v)\tilde{\mathbb{E}}(g'(\tilde{X}_T^{v, X_v})\nabla\tilde{X}_T^{v, X_v}) - \sigma(t_k, X_{t_k})\tilde{\mathbb{E}}(g'(\tilde{X}_T^{t_k, X_{t_k}})\nabla\tilde{X}_T^{t_k, X_{t_k}})\| \\ &\leq \|\sigma(v, X_v) - \sigma(t_k, X_{t_k})\|_4 \|\tilde{\mathbb{E}}(g'(\tilde{X}_T^{v, X_v})\nabla\tilde{X}_T^{v, X_v})\|_4 \\ &\quad + \|\sigma\|_\infty \|\tilde{\mathbb{E}}(g'(\tilde{X}_T^{v, X_v})\nabla\tilde{X}_T^{v, X_v}) - \tilde{\mathbb{E}}(g'(\tilde{X}_T^{t_k, X_{t_k}})\nabla\tilde{X}_T^{v, X_v})\| \\ &\quad + \|\sigma\|_\infty \|\tilde{\mathbb{E}}(g'(\tilde{X}_T^{t_k, X_{t_k}})\nabla\tilde{X}_T^{v, X_v}) - \tilde{\mathbb{E}}(g'(\tilde{X}_T^{t_k, X_{t_k}})\nabla\tilde{X}_T^{t_k, X_{t_k}})\| \\ &\leq C(C_{g'}, b, \sigma, T, p_0)\Psi(x)\left[h^{\frac{1}{2}} + \|X_v - X_{t_k}\|_4 + \left(\mathbb{E}\tilde{\mathbb{E}}|\tilde{X}_T^{v, X_v} - \tilde{X}_T^{t_k, X_{t_k}}|^{4\alpha}\right)^{\frac{1}{4}}\right. \\ &\quad \left. + \left(\mathbb{E}\tilde{\mathbb{E}}|\nabla\tilde{X}_T^{v, X_v} - \nabla\tilde{X}_T^{t_k, X_{t_k}}|^4\right)^{\frac{1}{4}}\right] \\ &\leq C(C_{g'}, b, \sigma, T, p_0)\Psi(x)h^{\frac{\alpha}{2}}. \end{aligned}$$

We compute the second term using  $Z_{t_k}^n$  as given in (17). Hence, with the notation from Definition 2.2,

$$\begin{aligned} \|Z_{t_k} - Z_{t_k}^n\|^2 &= \mathbb{E}\left|\sigma(t_k, X_{t_k})\tilde{\mathbb{E}}(g'(\tilde{X}_{t_n}^{t_k, X_{t_k}})\nabla\tilde{X}_{t_n}^{t_k, X_{t_k}}) - \tilde{\mathbb{E}}\mathcal{D}_{k+1}^n g(\tilde{\mathcal{X}}_{\tau_n}^{\tau_k, \mathcal{X}_{\tau_k}})\right|^2 \\ &\leq \|\sigma\|_\infty^2 \mathbb{E}\left|\tilde{\mathbb{E}}(g'(\tilde{X}_{t_n}^{t_k, X_{t_k}})\nabla\tilde{X}_{t_n}^{t_k, X_{t_k}}) - \frac{\tilde{\mathbb{E}}\mathcal{D}_{k+1}^n g(\tilde{\mathcal{X}}_{\tau_n}^{\tau_k, \mathcal{X}_{\tau_k}})}{\sigma(t_k, X_{t_k})}\right|^2 \\ &= \|\sigma\|_\infty^2 \mathbb{E}\left|\tilde{\mathbb{E}}(g'(\tilde{X}_{t_n}^{t_k, X_{t_k}})\nabla\tilde{X}_{t_n}^{t_k, X_{t_k}}) - \tilde{\mathbb{E}}\left(g_x^{(k+1, n+1)}\frac{\mathcal{D}_{k+1}^n \tilde{\mathcal{X}}_{\tau_n}^{\tau_k, \mathcal{X}_{\tau_k}}}{\sigma(t_k, X_{t_k})}\right)\right|^2. \end{aligned}$$

We insert  $\pm\tilde{\mathbb{E}}(g_x^{(k+1, n+1)}\nabla\tilde{X}_{t_n}^{t_k, X_{t_k}})$  and get by the Cauchy–Schwarz inequality that

$$\begin{aligned} &\left|\tilde{\mathbb{E}}(g'(\tilde{X}_{t_n}^{t_k, X_{t_k}})\nabla\tilde{X}_{t_n}^{t_k, X_{t_k}}) - \tilde{\mathbb{E}}\left(g_x^{(k+1, n+1)}\frac{\mathcal{D}_{k+1}^n \tilde{\mathcal{X}}_{\tau_n}^{\tau_k, \mathcal{X}_{\tau_k}}}{\sigma(t_k, X_{t_k})}\right)\right|^2 \\ &\leq 2\tilde{\mathbb{E}}|g'(\tilde{X}_{t_n}^{t_k, X_{t_k}}) - g_x^{(k+1, n+1)}|^2 \tilde{\mathbb{E}}|\nabla\tilde{X}_{t_n}^{t_k, X_{t_k}}|^2 + 2\tilde{\mathbb{E}}|g_x^{(k+1, n+1)}|^2 \tilde{\mathbb{E}}\left|\nabla\tilde{X}_{t_n}^{t_k, X_{t_k}} - \frac{\mathcal{D}_{k+1}^n \tilde{\mathcal{X}}_{\tau_n}^{\tau_k, \mathcal{X}_{\tau_k}}}{\sigma(t_k, X_{t_k})}\right|^2. \end{aligned} \tag{29}$$

For the estimate of  $\tilde{\mathbb{E}}|\nabla\tilde{X}_{t_n}^{t_k, X_{t_k}}|^2$  we use Lemma 5.2. Since  $g'$  satisfies (6) we proceed with

$$\begin{aligned} &\tilde{\mathbb{E}}|g'(\tilde{X}_{t_n}^{t_k, X_{t_k}}) - g_x^{(k+1, n+1)}|^2 \\ &\leq \int_0^1 \tilde{\mathbb{E}}\left|g'(\tilde{X}_{t_n}^{t_k, X_{t_k}}) - g'(\vartheta T_{k+1,+} \tilde{\mathcal{X}}_{\tau_n}^{\tau_k, \mathcal{X}_{\tau_k}} + (1-\vartheta)T_{k+1,-} \tilde{\mathcal{X}}_{\tau_n}^{\tau_k, \mathcal{X}_{\tau_k}})\right|^2 d\vartheta \\ &\leq \int_0^1 (\tilde{\mathbb{E}}\Psi_1^4)^{\frac{1}{2}} \left[\tilde{\mathbb{E}}\left|\tilde{X}_{t_n}^{t_k, X_{t_k}} - \vartheta T_{k+1,+} \tilde{\mathcal{X}}_{\tau_n}^{\tau_k, \mathcal{X}_{\tau_k}} - (1-\vartheta)T_{k+1,-} \tilde{\mathcal{X}}_{\tau_n}^{\tau_k, \mathcal{X}_{\tau_k}}\right|^{4\alpha}\right]^{\frac{1}{2}} d\vartheta, \end{aligned}$$

where  $\Psi_1 := C_g'(1 + |\tilde{X}_{t_n}^{I_k, X_{I_k}}|^{p_0} + |\vartheta T_{k+1,+} \tilde{\mathcal{X}}_{\tau_n}^{\tau_k, \mathcal{X}_{\tau_k}} + (1 - \vartheta)T_{k+1,-} \tilde{\mathcal{X}}_{\tau_n}^{\tau_k, \mathcal{X}_{\tau_k}}|^{p_0})$ . For  $\tilde{\mathbb{E}}\Psi_1^4$  and

$$\begin{aligned} & \tilde{\mathbb{E}} \left| \tilde{X}_{t_n}^{I_k, X_{I_k}} - (\vartheta T_{k+1,+} \tilde{\mathcal{X}}_{\tau_n}^{\tau_k, \mathcal{X}_{\tau_k}} + (1 - \vartheta)T_{k+1,-} \tilde{\mathcal{X}}_{\tau_n}^{\tau_k, \mathcal{X}_{\tau_k}}) \right|^{4\alpha} \\ & \leq 8 \left( \vartheta^{2\alpha} \tilde{\mathbb{E}} \left| \tilde{X}_{t_n}^{I_k, X_{I_k}} - T_{k+1,+} \tilde{\mathcal{X}}_{\tau_n}^{\tau_k, \mathcal{X}_{\tau_k}} \right|^{4\alpha} + (1 - \vartheta)^{2\alpha} \tilde{\mathbb{E}} \left| \tilde{X}_{t_n}^{I_k, X_{I_k}} - T_{k+1,-} \tilde{\mathcal{X}}_{\tau_n}^{\tau_k, \mathcal{X}_{\tau_k}} \right|^{4\alpha} \right) \\ & \leq C(b, \sigma, T)h^{2\alpha} + C(b, \sigma, T)(|X_{I_k} - \mathcal{X}_{\tau_k}|^{4\alpha} + h^\alpha), \end{aligned}$$

we use Lemma 5.4 and Lemma 5.2(v). For the last term in (29) we notice that

$$\tilde{\mathbb{E}}|g_x^{(k+1, n+1)}|^4 \leq C(C_g', b, \sigma, T, p_0)\Psi^4(x).$$

By Lemma 5.2 we have  $\tilde{\mathbb{E}}|\nabla \tilde{X}_{t_n}^{I_k, X_{I_k}} - \nabla \tilde{\mathcal{X}}_{\tau_n}^{\tau_k, \mathcal{X}_{\tau_k}}|^p \leq C(b, \sigma, T, p)h^{\frac{p}{4}}$ , and by Lemma 5.4,

$$\begin{aligned} & \tilde{\mathbb{E}} \left| \nabla \tilde{\mathcal{X}}_{\tau_n}^{\tau_k, \mathcal{X}_{\tau_k}} - \frac{\mathcal{D}_{k+1}^n \tilde{\mathcal{X}}_{\tau_n}^{\tau_k, \mathcal{X}_{\tau_k}}}{\sigma(t_k, X_{I_k})} \right|^p \\ & \leq C(p)\mathbb{E} \left| \nabla X_{t_n}^{n, t_k, X_{I_k}^n} - \frac{\mathcal{D}_{k+1}^n X_{t_n}^n}{\sigma(t_{k+1}, X_{I_k}^n)} \right|^p + C(p)\mathbb{E} \left| \frac{\mathcal{D}_{k+1}^n X_{t_n}^n}{\sigma(t_{k+1}, X_{I_k}^n)} - \frac{\mathcal{D}_{k+1}^n X_{t_n}^n}{\sigma(t_k, X_{I_k})} \right|^p \\ & \leq C(b, \sigma, T, p, \delta)h^{\frac{p}{4}}. \end{aligned}$$

Consequently,  $\|Z_{t_k} - Z_{t_k}^n\|^2 \leq C(C_g', b, \sigma, T, p_0, \delta)\Psi^2(x)h^{\frac{\alpha}{2}}$ .

**3.2. Proof of Theorem 3.1: the approximation rates for the general case**

Let  $u : [0, T) \times \mathbb{R} \rightarrow \mathbb{R}$  be the solution of the PDE (38) associated to (1). We use the representations  $Y_s = u(s, X_s)$  and  $Z_s = \sigma(s, X_s)u_x(s, X_s)$  stated in Theorem 4.2 and define

$$F(s, x) := f(s, x, u(s, x), \sigma(s, x)u_x(s, x)). \tag{30}$$

From (1) and (3) we conclude

$$\begin{aligned} \|Y_{t_k} - Y_{t_k}^n\| & \leq \|\mathbb{E}_{t_k}g(X_T) - \mathbb{E}_{\tau_k}g(X_T^n)\| \\ & \quad + \left\| \mathbb{E}_{t_k} \int_{t_k}^T f(s, X_s, Y_s, Z_s)ds - h\mathbb{E}_{\tau_k} \sum_{m=k}^{n-1} f(t_{m+1}, X_{t_m}^n, Y_{t_m}^n, Z_{t_m}^n) \right\|, \end{aligned}$$

where Proposition 3.1 provides the estimate for the terminal condition. We decompose the generator term as follows:

$$\begin{aligned} & \mathbb{E}_{t_k}f(s, X_s, Y_s, Z_s) - \mathbb{E}_{\tau_k}f(t_{m+1}, X_{t_m}^n, Y_{t_m}^n, Z_{t_m}^n) \\ & = [\mathbb{E}_{t_k}f(s, X_s, Y_s, Z_s) - \mathbb{E}_{t_k}f(t_m, X_{t_m}, Y_{t_m}, Z_{t_m})] + [\mathbb{E}_{t_k}F(t_m, X_{t_m}) - \mathbb{E}_{\tau_k}F(t_m, X_{t_m}^n)] \\ & \quad + [\mathbb{E}_{\tau_k}F(t_m, X_{t_m}^n) - \mathbb{E}_{\tau_k}F(t_m, X_{t_m})] + [\mathbb{E}_{\tau_k}f(t_m, X_{t_m}, Y_{t_m}, Z_{t_m}) - \mathbb{E}_{\tau_k}f(t_{m+1}, X_{t_m}^n, Y_{t_m}^n, Z_{t_m}^n)] \\ & =: d_1(s, m) + d_2(m) + d_3(m) + d_4(m). \end{aligned}$$

We use

$$\begin{aligned} & \left\| \mathbb{E}_{\tau_k} \int_{t_k}^T f(s, X_s, Y_s, Z_s) ds - h \mathbb{E}_{\tau_k} \sum_{m=k}^{n-1} f(t_{m+1}, X_{t_m}^n, Y_{t_m}^n, Z_{t_m}^n) \right\| \\ & \leq \sum_{m=k}^{n-1} \left( \left\| \int_{t_m}^{t_{m+1}} d_1(s, m) ds \right\| + h \sum_{i=2}^4 \|d_i(m)\| \right) \end{aligned}$$

and estimate the expressions on the right-hand side. For the function  $F$  defined in (30) we use Assumption 2.3 (which implies that (6) holds for  $\alpha = 1$ ) to derive by Theorem 4.2 and the mean value theorem that for  $x_1, x_2 \in \mathbb{R}$  there exists  $\xi \in [\min\{x_1, x_2\}, \max\{x_1, x_2\}]$  such that

$$\begin{aligned} |F(t, x_1) - F(t, x_2)| &= |f(t, x_1, u(t, x_1), \sigma(t, x_1)u_x(t, x_1)) - f(t, x_2, u(t, x_2), \sigma(t, x_2)u_x(t, x_2))| \\ &\leq C(L_f, \sigma) \left( 1 + c_{4.2}^2 \Psi(\xi) + \frac{c_{4.2}^3 \Psi(\xi)}{(T-t)^{\frac{1}{2}}} \right) |x_1 - x_2| \\ &\leq C(L_f, c_{4.2}^{2,3}, \sigma, T) (1 + |x_1|^{p_0+1} + |x_2|^{p_0+1}) \frac{|x_1 - x_2|}{(T-t)^{\frac{1}{2}}}. \end{aligned} \tag{31}$$

By (7), standard estimates on  $(X_s)$ , Theorem 4.1(i), and Proposition 4.1 for  $p = 2$ , we immediately get

$$\begin{aligned} \|d_1(s, m)\| &\leq C(L_f, C_{4.1}^y, C_{4.1}, b, \sigma, T) \Psi(x) h^{\frac{1}{2}} \\ &= C(b, \sigma, f, g, T, p_0, \delta) \Psi(x) h^{\frac{1}{2}}. \end{aligned}$$

For the estimate of  $d_2$  one exploits

$$\mathbb{E}_{t_k} F(t_m, X_{t_m}) - \mathbb{E}_{\tau_k} F(t_m, X_{t_m}^n) = \tilde{\mathbb{E}} F(t_m, \tilde{X}_{t_m}^{t_k, X_{t_k}}) - \tilde{\mathbb{E}} F(t_m, \tilde{X}_{t_m}^{n, t_k, X_{t_k}^n})$$

and then uses (31) and Lemma 5.2(v). This gives

$$\|d_2(m)\| \leq C(L_f, c_{4.2}^{2,3}, b, \sigma, T, p_0) \Psi(x) \frac{1}{(T-t_m)^{\frac{1}{2}}} h^{\frac{1}{4}}.$$

For  $d_3$  we start with Jensen’s inequality and then continue similarly as above to get

$$\|d_3(m)\| \leq \|F(t_m, X_{t_m}^n) - F(t_m, X_{t_m})\| \leq C(L_f, c_{4.2}^{2,3}, b, \sigma, T, p_0) \Psi(x) \frac{1}{(T-t_m)^{\frac{1}{2}}} h^{\frac{1}{4}},$$

and for the last term we get

$$\|d_4(m)\| \leq L_f (h^{\frac{1}{2}} + \|X_{t_m} - X_{t_m}^n\| + \|Y_{t_m} - Y_{t_m}^n\| + \|Z_{t_m} - Z_{t_m}^n\|).$$

This implies

$$\|Y_{t_k} - Y_{t_k}^n\| \leq C \Psi(x) h^{\frac{1}{4}} + h L_f \sum_{m=k}^{n-1} (\|Y_{t_m} - Y_{t_m}^n\| + \|Z_{t_m} - Z_{t_m}^n\|), \tag{32}$$

where  $C = C(L_f, C_{3.1}^y, C_{4.1}^y, C_{4.2}, c_{4.2}^{2,3}, b, \sigma, T, p_0) = C(b, \sigma, f, g, T, p_0, \delta)$ .

For  $\|Z_{t_k} - Z_{t_k}^n\|$  we use the representations (14) and (17), the approximation (20), and Proposition 2.1. Instead of  $N_{t_n}^{n,t_k}$  we will use here the notation  $N_{\tau_n}^{n,\tau_k}$  to indicate its measurability with respect to the filtration  $(\mathcal{F}_t)$ . It holds that

$$\begin{aligned} \|Z_{t_k}^n - Z_{t_k}\| &\leq \|Z_{t_k}^n - \hat{Z}_{t_k}^n\| + \|Z_{t_k} - \hat{Z}_{t_k}^n\| \\ &\leq C_{2.1} \hat{\Psi}(x) h^{\frac{\alpha}{2}} + \|\sigma(t_k, X_{t_k}) \tilde{\mathbb{E}} g'(\tilde{X}_{t_n}^{t_k, X_{t_k}}) \nabla \tilde{X}_{t_n}^{t_k, X_{t_k}} - \tilde{\mathbb{E}} \mathcal{D}_{k+1}^n g(\tilde{X}_{t_n}^{n, t_k, X_{t_k}^n})\| \\ &\quad + \left\| \mathbb{E}_{t_k} \int_{t_{k+1}}^T f(s, X_s, Y_s, Z_s) N_s^{t_k} ds \sigma(t_k, X_{t_k}) \right. \\ &\quad \left. - \mathbb{E}_{\tau_k} h \sum_{m=k+1}^{n-1} f(t_{m+1}, X_{t_m}^n, Y_{t_m}^n, Z_{t_m}^n) N_{\tau_m}^{n, \tau_k} \sigma(t_{k+1}, X_{t_k}^n) \right\| \\ &\quad + \left\| \mathbb{E}_{t_k} \int_{t_k}^{t_{k+1}} f(s, X_s, Y_s, Z_s) N_s^{t_k} ds \sigma(t_k, X_{t_k}) \right\|. \end{aligned} \tag{33}$$

For the terminal condition, Proposition 3.1 provides

$$\left\| \sigma(t_k, X_{t_k}) \tilde{\mathbb{E}} g'(\tilde{X}_{t_n}^{t_k, X_{t_k}}) \nabla \tilde{X}_{t_n}^{t_k, X_{t_k}} - \tilde{\mathbb{E}} \mathcal{D}_{k+1}^n g(\tilde{X}_{t_n}^{n, t_k, X_{t_k}^n}) \right\| \leq (C_{3.1}^z)^{\frac{1}{2}} \Psi(x) h^{\frac{1}{4}}. \tag{34}$$

We continue with the generator terms and use  $F$  defined in (30) to decompose the difference

$$\begin{aligned} &\mathbb{E}_{t_k} f(s, X_s, Y_s, Z_s) N_s^{t_k} \sigma(t_k, X_{t_k}) - \mathbb{E}_{\tau_k} f(t_{m+1}, X_{t_m}^n, Y_{t_m}^n, Z_{t_m}^n) N_{\tau_m}^{n, \tau_k} \sigma(t_{k+1}, X_{t_k}^n) \\ &= \mathbb{E}_{t_k} f(s, X_s, Y_s, Z_s) N_s^{t_k} \sigma(t_k, X_{t_k}) - \mathbb{E}_{t_k} f(t_m, X_{t_m}, Y_{t_m}, Z_{t_m}) N_{t_m}^{t_k} \sigma(t_k, X_{t_k}) \\ &\quad + \mathbb{E}_{t_k} F(t_m, X_{t_m}) N_{t_m}^{t_k} \sigma(t_k, X_{t_k}) - \mathbb{E}_{\tau_k} F(t_m, X_{t_m}^n) N_{\tau_m}^{n, \tau_k} \sigma(t_{k+1}, X_{t_k}^n) \\ &\quad + \mathbb{E}_{\tau_k} [F(t_m, X_{t_m}^n) - F(t_m, X_{t_m})] N_{\tau_m}^{n, \tau_k} \sigma(t_{k+1}, X_{t_k}^n) \\ &\quad + \mathbb{E}_{\tau_k} [f(t_m, X_{t_m}, Y_{t_m}, Z_{t_m}) - f(t_{m+1}, X_{t_m}^n, Y_{t_m}^n, Z_{t_m}^n)] N_{\tau_m}^{n, \tau_k} \sigma(t_{k+1}, X_{t_k}^n) \\ &=: \tau_1(s, m) + \tau_2(m) + \tau_3(m) + \tau_4(m), \end{aligned}$$

where  $s \in [t_m, t_{m+1})$ . For  $\tau_1$  we use that  $\mathbb{E}_{t_k} f(t_m, X_{t_k}, Y_{t_k}, Z_{t_k}) (N_s^{t_k} - N_{t_m}^{t_k}) = 0$ , so that

$$\begin{aligned} \|\tau_1(s, m)\| &\leq \|\mathbb{E}_{t_k} f(s, X_s, Y_s, Z_s) N_s^{t_k} \sigma(t_k, X_{t_k}) - \mathbb{E}_{t_k} f(t_m, X_{t_m}, Y_{t_m}, Z_{t_m}) N_s^{t_k} \sigma(t_k, X_{t_k})\| \\ &\quad + \|\mathbb{E}_{t_k} (f(t_m, X_{t_m}, Y_{t_m}, Z_{t_m}) - f(t_m, X_{t_k}, Y_{t_k}, Z_{t_k})) (N_s^{t_k} - N_{t_m}^{t_k}) \sigma(t_k, X_{t_k})\|. \end{aligned}$$

As before, we rewrite the conditional expectations with the help of the independent copy  $\tilde{B}$ . Then

$$\begin{aligned} &\mathbb{E}_{t_k} f(s, X_s, Y_s, Z_s) N_s^{t_k} - \mathbb{E}_{t_k} f(t_m, X_{t_m}, Y_{t_m}, Z_{t_m}) N_s^{t_k} \\ &= \tilde{\mathbb{E}}[(f(s, \tilde{X}_s^{t_k, X_{t_k}}, \tilde{Y}_s^{t_k, X_{t_k}}, \tilde{Z}_s^{t_k, X_{t_k}}) - f(t_m, \tilde{X}_{t_m}^{t_k, X_{t_k}}, \tilde{Y}_{t_m}^{t_k, X_{t_k}}, \tilde{Z}_{t_m}^{t_k, X_{t_k}})) \tilde{N}_s^{t_k}] \end{aligned}$$

and

$$\begin{aligned} &\mathbb{E}_{t_k} (f(t_m, X_{t_m}, Y_{t_m}, Z_{t_m}) - f(t_m, X_{t_k}, Y_{t_k}, Z_{t_k})) (N_s^{t_k} - N_{t_m}^{t_k}) \\ &= \tilde{\mathbb{E}}[(f(t_m, \tilde{X}_{t_m}^{t_k, X_{t_k}}, \tilde{Y}_{t_m}^{t_k, X_{t_k}}, \tilde{Z}_{t_m}^{t_k, X_{t_k}}) - f(t_m, X_{t_k}, Y_{t_k}, Z_{t_k})) (\tilde{N}_s^{t_k} - \tilde{N}_{t_m}^{t_k})]. \end{aligned}$$

We apply the conditional Hölder inequality, and from the estimates (37) and

$$\tilde{\mathbb{E}}|\tilde{N}_s^{t_k} - \tilde{N}_m^{t_k}|^2 \leq C(b, \sigma, T, \delta) \frac{h}{(s - t_k)^2}$$

we get

$$\begin{aligned} \|\tau_1(s, m)\| &\leq \frac{\kappa_2 \|\sigma\|_\infty}{(s - t_k)^{\frac{1}{2}}} \|f(s, X_s, Y_s, Z_s) - f(t_m, X_{t_m}, Y_{t_m}, Z_{t_m})\| \\ &\quad + C(b, \sigma, T, \delta) \frac{h^{\frac{1}{2}}}{s - t_k} \|f(t_m, X_{t_m}, Y_{t_m}, Z_{t_m}) - f(t_k, X_{t_k}, Y_{t_k}, Z_{t_k})\| \\ &\leq C(L_f, C_{4.1}^y, C_{4.1}, \kappa_2, b, \sigma, T, p_0, \delta) \Psi(x) \frac{h^{\frac{1}{2}}}{(s - t_k)^{\frac{1}{2}}}, \end{aligned}$$

since for  $0 \leq t < s \leq T$  we have by Theorem 4.1 and Proposition 4.1 that

$$\|f(s, X_s, Y_s, Z_s) - f(t, X_t, Y_t, Z_t)\| \leq C(L_f, C_{4.1}^y, C_{4.1}, b, \sigma, T, p_0) \Psi(x) (s - t)^{\frac{1}{2}}. \tag{35}$$

For the estimate of  $\tau_2$ , Lemma 5.2, Lemma 5.3, (31), and (37) yield

$$\begin{aligned} \|\tau_2(m)\| &= \|\tilde{\mathbb{E}}F(t_m, \tilde{X}_{t_m}^{t_k, X_{t_k}}) \tilde{N}_{t_m}^{t_k} \sigma(t_k, X_{t_k}) - \tilde{\mathbb{E}}F(t_m, \tilde{X}_{t_m}^{\tau_k, \mathcal{X}_{\tau_k}}) \tilde{N}_{t_m}^{n, \tau_k} \sigma(t_{k+1}, \mathcal{X}_{\tau_k})\| \\ &\leq \frac{C(\kappa_2, \sigma)}{(t_m - t_k)^{\frac{1}{2}}} \left( \mathbb{E} \tilde{\mathbb{E}}(F(t_m, \tilde{X}_{t_m}^{t_k, X_{t_k}}) - F(t_m, \tilde{X}_{t_m}^{\tau_k, \mathcal{X}_{\tau_k}}))^2 \right)^{\frac{1}{2}} \\ &\quad + (\mathbb{E} \tilde{\mathbb{E}}|F(t_m, \tilde{X}_{t_m}^{\tau_k, \mathcal{X}_{\tau_k}}) - F(t_m, \mathcal{X}_{\tau_k})|^2 \tilde{\mathbb{E}}|\tilde{N}_{t_m}^{t_k} \sigma(t_k, X_{t_k}) - \tilde{N}_{t_m}^{n, \tau_k} \sigma(t_{k+1}, \mathcal{X}_{\tau_k})|^2)^{\frac{1}{2}} \\ &\leq C(L_f, c_{4.2}^{2,3}, \kappa_2, b, \sigma, T, p_0, \delta) \frac{\Psi(x)}{(T - t_m)^{\frac{1}{2}} (t_m - t_k)^{\frac{1}{2}}} \frac{h^{\frac{1}{4}}}{1}. \end{aligned}$$

For  $\tau_3$  we use the conditional Hölder inequality, (31), (19), and Lemma 5.2:

$$\begin{aligned} \|\tau_3(m)\| &= \|\mathbb{E}_{\tau_k} [ [F(t_m, X_{t_m}^n) - F(t_m, X_{t_m})] N_{\tau_k}^{n, \tau_k} \sigma(t_{k+1}, \mathcal{X}_{\tau_k}) ] \| \\ &\leq \frac{C(\hat{\kappa}_2, \sigma)}{(t_m - t_k)^{\frac{1}{2}}} \|F(t_m, X_{t_m}^n) - F(t_m, X_{t_m})\| \\ &\leq C(L_f, c_{4.2}^{2,3}, b, \sigma, T, p_0, \delta) \frac{\Psi(x)}{(T - t_m)^{\frac{1}{2}} (t_m - t_k)^{\frac{1}{2}}} \frac{h^{\frac{1}{4}}}{1}. \end{aligned}$$

The term  $\tau_4$  can be estimated as follows:

$$\begin{aligned} \|\tau_4(m)\| &= \|\mathbb{E}_{\tau_k} [ [f(t_m, X_{t_m}, Y_{t_m}, Z_{t_m}) - f(t_{m+1}, X_{t_m}^n, Y_{t_m}^n, Z_{t_m}^n)] N_{\tau_k}^{n, \tau_k} \sigma(t_{k+1}, \mathcal{X}_{\tau_k}) ] \| \\ &\leq \frac{C(L_f, b, \sigma, T, \delta)}{(t_m - t_k)^{\frac{1}{2}}} (h^{\frac{1}{2}} + \|X_{t_m} - X_{t_m}^n\| + \|Y_{t_m} - Y_{t_m}^n\| + \|Z_{t_m} - Z_{t_m}^n\|). \end{aligned}$$

Finally, for the remaining term of the estimate of  $\|Z_{t_k} - Z_{t_k}^n\|$ , we use (35) and (37) to get

$$\begin{aligned} \|\mathbb{E}_{t_k} f(s, X_s, Y_s, Z_s) N_s^{t_k} \sigma(t_k, X_{t_k})\| &= \|\mathbb{E}_{t_k} [ (f(s, X_s, Y_s, Z_s) - f(s, X_{t_k}, Y_{t_k}, Z_{t_k})) N_s^{t_k} ] \sigma(t_k, X_{t_k}) \| \\ &\leq C(L_f, C_{4.1}^y, C_{4.1}, \kappa_2, b, \sigma, T, p_0) \Psi(x). \end{aligned}$$



Consequently, from (33), (34), and the estimates for the remaining term and for  $\tau_1, \dots, \tau_4$ , it follows that

$$\begin{aligned} \|Z_{t_k} - Z_{t_k}^n\| &\leq C_{2.1} \hat{\Psi}(x) h^{\frac{\alpha}{2}} + (C_{3.1}^z)^{\frac{1}{2}} \Psi(x) h^{\frac{1}{4}} + C(L_f, C_{4.1}^y, C_{4.1}, b, \sigma, T, p_0, \kappa_2) \Psi(x) h \\ &\quad + C(L_f, C_{4.1}^y, C_{4.1}, \kappa_2, b, \sigma, T, p_0, \delta) \Psi(x) h^{\frac{1}{2}} \int_{t_k}^T \frac{ds}{(s - t_k)^{\frac{1}{2}}} \\ &\quad + C(L_f, c_{4.2}^{2,3}, \kappa_2, b, \sigma, T, p_0, \delta) h \sum_{m=k+1}^{n-1} \frac{\Psi(x)}{(T - t_m)^{\frac{1}{2}}} \frac{h^{\frac{1}{4}}}{(t_m - t_k)^{\frac{1}{2}}} \\ &\quad + C(L_f, b, \sigma, T, \delta) h \sum_{m=k+1}^{n-1} (\|Y_{t_m} - Y_{t_m}^n\| + \|Z_{t_m} - Z_{t_m}^n\|) \frac{1}{(t_m - t_k)^{\frac{1}{2}}} \\ &\leq C(C_{2.1}, C_{3.1}^z) \hat{\Psi}(x) h^{\frac{\alpha}{2} \wedge \frac{1}{4}} + C(L_f, c_{4.2}^{2,3}, C_{4.1}^y, C_{4.1}, \kappa_2, b, \sigma, T, p_0, \delta) \Psi(x) h^{\frac{1}{4}} \\ &\quad + C(L_f, b, \sigma, T, \delta) \sum_{m=k+1}^{n-1} (\|Y_{t_m} - Y_{t_m}^n\| + \|Z_{t_m} - Z_{t_m}^n\|) \frac{1}{(t_m - t_k)^{\frac{1}{2}}} h. \end{aligned}$$

Then we use (32) and the above estimate to get

$$\begin{aligned} &\|Y_{t_k} - Y_{t_k}^n\| + \|Z_{t_k} - Z_{t_k}^n\| \\ &\leq C(C_{2.1}, C_{3.1}^z) \hat{\Psi}(x) h^{\frac{\alpha}{2} \wedge \frac{1}{4}} + C(L_f, C_{3.1}^y, C_{4.1}^y, C_{4.1}, c_{4.2}^{2,3}, \kappa_2, b, \sigma, T, p_0, \delta) \Psi(x) h^{\frac{1}{4}} \\ &\quad + C(L_f, b, \sigma, T, \delta) \sum_{m=k+1}^{n-1} (\|Y_{t_m} - Y_{t_m}^n\| + \|Z_{t_m} - Z_{t_m}^n\|) \frac{1}{(t_m - t_k)^{\frac{1}{2}}} h. \end{aligned}$$

Consequently, summarizing the dependencies, there is a  $C = C(b, \sigma, f, g, T, p_0, \delta)$  such that

$$\|Y_{t_k} - Y_{t_k}^n\| + \|Z_{t_k} - Z_{t_k}^n\| \leq C \hat{\Psi}(x) h^{\frac{\alpha}{2} \wedge \frac{1}{4}}.$$

By Theorem 4.1 (note that by Assumption 2.3 on  $g$  we have  $\alpha = 1$ ) it follows that

$$\|Y_v - Y_v^n\| \leq \|Y_v - Y_{t_k}\| + \|Y_{t_k} - Y_{t_k}^n\| \leq C_{4.1}^y \Psi(x) h^{\frac{1}{2}} + \hat{\Psi}(x) h^{\frac{\alpha}{2} \wedge \frac{1}{4}},$$

while Proposition 4.1 implies that

$$\|Z_v - Z_{t_k}\| \leq C_{4.1} \Psi(x) h^{\frac{1}{2}},$$

and hence we have

$$\mathbb{E}_{0,x} |Y_v - Y_v^n|^2 + \mathbb{E}_{0,x} |Z_v - Z_v^n|^2 \leq C_{3.1} \hat{\Psi}(x)^2 h^{\frac{1}{2} \wedge \alpha}$$

with  $C_{3.1} = C_{3.1}(b, \sigma, f, g, T, p_0, \delta)$ .

### 4. Some properties of solutions to BSDEs and their associated PDEs

#### 4.1. Malliavin weights

We use the SDE from (1) started in  $(t, x)$ ,

$$X_s^{t,x} = x + \int_t^s b(r, X_r^{t,x}) dr + \int_t^s \sigma(r, X_r^{t,x}) dB_r, \quad 0 \leq t \leq s \leq T, \tag{36}$$

and recall the Malliavin weight and its properties from [20, Subsection 1.1 and Remark 3].

**Lemma 4.1.** *Let  $H : \mathbb{R} \rightarrow \mathbb{R}$  be a polynomially bounded Borel function. If Assumption 2.1 holds and  $X^{t,x}$  is given by (36), then setting*

$$G(t, x) := \mathbb{E}H(X_T^{t,x})$$

*implies that  $G \in C^{1,2}([0, T] \times \mathbb{R})$ . Specifically, it holds for  $0 \leq t \leq r < T$  that*

$$\partial_x G(r, X_r^{t,x}) = \mathbb{E}[H(X_T^{t,x})N_T^{r,(t,x)} | \mathcal{F}_r^t],$$

*where  $(\mathcal{F}_r^t)_{r \in [t, T]}$  is the augmented natural filtration of  $(B_r^{t,0})_{r \in [t, T]}$ ,*

$$N_T^{r,(t,x)} = \frac{1}{T-r} \int_r^T \frac{\nabla X_s^{t,x}}{\sigma(s, X_s^{t,x}) \nabla X_r^{t,x}} dB_s,$$

*and  $\nabla X_s^{t,x}$  is given in (13). Moreover, for  $q \in (0, \infty)$  there exists a  $\kappa_q > 0$  such that*

$$\left(\mathbb{E}[|N_T^{r,(t,x)}|^q | \mathcal{F}_r^t]\right)^{\frac{1}{q}} \leq \frac{\kappa_q}{(T-r)^{\frac{1}{2}}} \quad \text{and} \quad \mathbb{E}[N_T^{r,(t,x)} | \mathcal{F}_r^t] = 0 \quad \text{almost surely,} \quad (37)$$

*and we have*

$$\|\partial_x G(r, X_r^{t,x})\|_{L_p(\mathbb{P})} \leq \kappa_q \frac{\|H(X_T^{t,x}) - \mathbb{E}[H(X_T^{t,x}) | \mathcal{F}_r^t]\|_p}{\sqrt{T-r}}$$

*for  $1 < q, p < \infty$  with  $\frac{1}{p} + \frac{1}{q} = 1$ .*

**4.2. Regularity of solutions to BSDEs**

The following result originates from [20, Theorem 1], where path-dependent cases were also included. We formulate it only for our Markovian setting but use  $\mathbb{P}_{t,x}$  since we are interested in an estimate for all  $(t, x) \in [0, T] \times \mathbb{R}$ . A sketch of a proof of this formulation can be found in [22].

**Theorem 4.1.** *Let Assumptions 2.1 and 2.2 hold. Then for any  $p \in [2, \infty)$  the following assertions are true.*

(i) *There exists a constant  $C_{4.1}^y > 0$  such that for  $0 \leq t < s \leq T$  and  $x \in \mathbb{R}$ ,*

$$\|Y_s - Y_t\|_{L_p(\mathbb{P}_{t,x})} \leq C_{4.1}^y \Psi(x) \left( \int_t^s (T-r)^{\alpha-1} dr \right)^{\frac{1}{2}}.$$

(ii) *There exists a constant  $C_{4.1}^z > 0$  such that for  $0 \leq t < s < T$  and  $x \in \mathbb{R}$ ,*

$$\|Z_s - Z_t\|_{L_p(\mathbb{P}_{t,x})} \leq C_{4.1}^z \Psi(x) \left( \int_t^s (T-r)^{\alpha-2} dr \right)^{\frac{1}{2}}.$$

The constants  $C_{4.1}^y$  and  $C_{4.1}^z$  depend on  $(L_f, K_f, C_g, c_{4.2}^{1,2}, \kappa_q, b, \sigma, T, p_0, p)$ , and  $\Psi(x)$  is defined in (8).

### 4.3. Properties of the associated PDE

The theorem below collects properties of the solution to the PDE associated to the FBSDE (1). For a proof see [43, Theorem 3.2], [41], and [22, Theorem 5.4].

**Theorem 4.2.** *Consider the FBSDE (1) and let Assumptions 2.1 and 2.2 hold. Then for the solution  $u$  of the associated PDE*

$$\begin{cases} u_t(t, x) + \frac{\sigma^2(t, x)}{2} u_{xx}(t, x) + b(t, x)u_x(t, x) + f(t, x, u(t, x), \sigma(t, x)u_x(t, x)) = 0, \\ t \in [0, T), \quad x \in \mathbb{R}, \\ u(T, x) = g(x), \quad x \in \mathbb{R}, \end{cases} \tag{38}$$

we have the following:

- (i)  $Y_t = u(t, X_t)$  almost surely, where  $u(t, x) = \mathbb{E}_{t,x} \left( g(X_T) + \int_t^T f(r, X_r, Y_r, Z_r) dr \right)$ , and  $|u(t, x)| \leq c_{4.2}^1 \Psi(x)$  for  $\Psi$  given in (8), where  $c_{4.2}^1$  depends on  $L_f, K_f, C_g, T$ , and  $p_0$ , as well as on the bounds and Lipschitz constants of  $b$  and  $\sigma$ .
- (ii) (a)  $\partial_x u$  exists and is continuous in  $[0, T) \times \mathbb{R}$ .  
 (b)  $Z_s^{t,x} = u_x(s, X_s^{t,x})\sigma(s, X_s^{t,x})$  almost surely.  
 (c)

$$|u_x(t, x)| \leq \frac{c_{4.2}^2 \Psi(x)}{(T-t)^{\frac{1-\alpha}{2}}},$$

where  $c_{4.2}^2$  depends on  $L_f, K_f, C_g, T, p_0$ , and  $\kappa_2 = \kappa_2(b, \sigma, T, \delta)$ , as well as on the bounds and Lipschitz constants of  $b$  and  $\sigma$ , and hence  $c_{4.2}^2 = c_{4.2}^2(L_f, K_f, C_g, b, \sigma, T, p_0, \delta)$ .

- (iii) (a)  $\partial_x^2 u$  exists and is continuous in  $[0, T) \times \mathbb{R}$ .  
 (b)

$$|\partial_x^2 u(t, x)| \leq \frac{c_{4.2}^3 \Psi(x)}{(T-t)^{1-\frac{\alpha}{2}}},$$

where  $c_{4.2}^3$  depends on  $L_f, C_g, T, p_0, \kappa_2 = \kappa_2(b, \sigma, T, \delta), C_{4.1}^y$ , and  $C_{4.1}^z$ , as well as on the bounds and Lipschitz constants of  $b$  and  $\sigma$ , and hence  $c_{4.2}^3 = c_{4.2}^3(L_f, K_f, C_g, b, \sigma, T, p_0, \delta)$ .

Using Assumption 2.3, we are now in a position to improve the bound on  $\|Z_s - Z_t\|_{L^p(\mathbb{P}_{t,x})}$  given in Theorem 4.1.

**Proposition 4.1.** *If Assumption 2.3 holds, then there exists a constant  $C_{4.1} > 0$  such that for  $0 \leq t < s \leq T$  and  $x \in \mathbb{R}$ ,*

$$\|Z_s - Z_t\|_{L^p(\mathbb{P}_{t,x})} \leq C_{4.1} \Psi(x)(s-t)^{\frac{1}{2}},$$

where  $C_{4.1}$  depends on  $c_{4.2}^{2,3}, b, \sigma, f, g, T, p_0$ , and  $p$ , and hence  $C_{4.1} = C_{4.1}(b, \sigma, f, g, T, p_0, p, \delta)$ .

*Proof.* From  $Z_s^{t,x} = u_x(s, X_s^{t,x})\sigma(s, X_s^{t,x})$  and

$$\nabla Y_s^{t,x} = \partial_x u(s, X_s^{t,x}) = u_x(s, X_s^{t,x})\nabla X_s^{t,x},$$

we conclude that

$$Z_s^{t,x} = \frac{\nabla Y_s^{t,x}}{\nabla X_s^{t,x}} \sigma(s, X_s^{t,x}), \quad 0 \leq t \leq s \leq T. \tag{39}$$

It is well-known (see e.g. [19]) that the solution  $\nabla Y$  of the linear BSDE

$$\begin{aligned} \nabla Y_s = & g'(X_T)\nabla X_T + \int_s^T (f_x(\Theta_r)\nabla X_r + f_y(\Theta_r)\nabla Y_r + f_z(\Theta_r)\nabla Z_r)dr \\ & - \int_s^T \nabla Z_r dB_r, \quad 0 \leq s \leq T, \end{aligned} \tag{40}$$

can be represented as

$$\begin{aligned} \frac{\nabla Y_s}{\nabla X_s} = & \mathbb{E}_s \left[ g'(X_T)\nabla X_T \Gamma_T^s + \int_s^T f_x(\Theta_r)\nabla X_r \Gamma_r^s dr \right] \frac{1}{\nabla X_s} \\ = & \tilde{\mathbb{E}} \left[ g'(\tilde{X}_T^{s,X_s})\nabla \tilde{X}_T^{s,X_s} \tilde{\Gamma}_T^{s,X_s} + \int_s^T f_x(\tilde{\Theta}_r^{s,X_s})\nabla \tilde{X}_r^{s,X_s} \tilde{\Gamma}_r^{s,X_s} dr \right], \quad 0 \leq t \leq s \leq T, \end{aligned} \tag{41}$$

where  $\Theta_r := (r, X_r, Y_r, Z_r)$ ,  $\Gamma^s$  denotes the adjoint process given by

$$\Gamma_r^s = 1 + \int_s^r f_y(\Theta_u)\Gamma_u^s du + \int_s^r f_z(\Theta_u)\Gamma_u^s dB_u, \quad s \leq r \leq T,$$

and

$$\tilde{\Gamma}_s^{t,x} = 1 + \int_t^s f_y(\tilde{\Theta}_r^{t,x})\tilde{\Gamma}_r^{t,x} dr + \int_t^s f_z(\tilde{\Theta}_r^{t,x})\tilde{\Gamma}_r^{t,x} d\tilde{B}_r, \quad t \leq s \leq T, \quad x \in \mathbb{R},$$

where  $\tilde{B}$  denotes an independent copy of  $B$ . Notice that  $\nabla X_t^{t,x} = 1$ , so that

$$\frac{\nabla Y_t^{t,x}}{\nabla X_t^{t,x}} = \nabla Y_t^{t,x} = \tilde{\mathbb{E}} \left[ g'(\tilde{X}_T^{t,x})\nabla \tilde{X}_T^{t,x} \tilde{\Gamma}_T^{t,x} + \int_t^T f_x(\tilde{\Theta}_r^{t,x})\nabla \tilde{X}_r^{t,x} \tilde{\Gamma}_r^{t,x} dr \right].$$

Then, by (39),

$$\begin{aligned} \|Z_s - Z_t\|_{L_p(\mathbb{P}_{t,x})} \leq & C(\sigma) \left[ \left\| \frac{\nabla Y_s}{\nabla X_s} - \frac{\nabla Y_t}{\nabla X_t} \right\|_{L_p(\mathbb{P}_{t,x})} \right. \\ & \left. + \|\nabla Y_t\|_{L_{2p}(\mathbb{P}_{t,x})} [(s-t)^{\frac{1}{2}} + \|X_s^{t,x} - x\|_{L_{2p}(\mathbb{P}_{t,x})}] \right]. \end{aligned}$$

Since  $(\nabla Y_s, \nabla Z_s)$  is the solution to the linear BSDE (40) with bounded  $f_x, f_y, f_z$ , we have that  $\|\nabla Y_t\|_{L_{2p}(\mathbb{P}_{t,x})} \leq C(b, \sigma, f, g, T, p)$ . Obviously,  $\|X_s^{t,x} - x\|_{L_{2p}(\mathbb{P}_{t,x})} \leq C(b, \sigma, T, p)(s-t)^{\frac{1}{2}}$ . So it remains to show that

$$\left\| \frac{\nabla Y_s}{\nabla X_s} - \frac{\nabla Y_t}{\nabla X_t} \right\|_{L_p(\mathbb{P}_{t,x})} \leq C\Psi(x)(s-t)^{\frac{1}{2}}.$$

We intend to use (41) in the following. There is a certain degree of freedom in how to connect  $B$  and  $\tilde{B}$  in order to compute conditional expectations. Here, unlike in (27), we define the processes

$$B'_u = B_{u \wedge s} + \tilde{B}_{u \vee s} - \tilde{B}_s \quad \text{and} \quad B''_u = B_{u \wedge t} + \tilde{B}_{u \vee t} - \tilde{B}_t, \quad u \geq 0,$$

as driving Brownian motions for  $\frac{\nabla Y_s}{\nabla X_s}$  and  $\frac{\nabla Y_t}{\nabla X_t}$ , respectively. This will especially simplify the estimate for  $\tilde{\mathbb{E}}|\tilde{\Gamma}_T^{s,X_s} - \tilde{\Gamma}_T^{t,x}|^q$  below. From the above relations we get the following (with  $X_s := X_s^{t,x}$ ):

$$\begin{aligned} \left\| \frac{\nabla Y_s}{\nabla X_s} - \frac{\nabla Y_t}{\nabla X_t} \right\|_{L_p(\mathbb{P}_{t,x})} &\leq \left\| \tilde{\mathbb{E}} \left[ g'(\tilde{X}_T^{s,X_s}) \nabla \tilde{X}_T^{s,X_s} \tilde{\Gamma}_T^{s,X_s} - g'(\tilde{X}_T^{t,x}) \nabla \tilde{X}_T^{t,x} \tilde{\Gamma}_T^{t,x} \right] \right\|_p \\ &\quad + \int_t^s \left\| \tilde{\mathbb{E}} \left[ f_x(\tilde{\Theta}_r^{t,x}) \nabla \tilde{X}_r^{t,x} \tilde{\Gamma}_r^{t,x} \right] \right\|_p dr \\ &\quad + \left\| \int_s^T \tilde{\mathbb{E}} \left[ f_x(\tilde{\Theta}_r^{s,X_s}) \nabla \tilde{X}_r^{s,X_s} \tilde{\Gamma}_r^{s,X_s} - f_x(\tilde{\Theta}_r^{t,x}) \nabla \tilde{X}_r^{t,x} \tilde{\Gamma}_r^{t,x} \right] dr \right\|_p \\ &=: J_1 + J_2 + J_3. \end{aligned}$$

Since  $g'$  is Lipschitz continuous and of polynomial growth, we have

$$J_1 \leq C(b, \sigma, g, T, p) \Psi(x)(s - t)^{\frac{1}{2}}$$

by Hölder’s inequality and the  $L_q$ -boundedness for any  $q > 0$  of all the factors, as well as from the estimates for  $\tilde{X}_T^{s,X_s} - \tilde{X}_T^{t,x}$  and  $\nabla \tilde{X}_T^{s,X_s} - \nabla \tilde{X}_T^{t,x}$  as in Lemma 5.2. For the  $\Gamma$  differences we first apply the inequalities of Hölder and BDG:

$$\begin{aligned} \tilde{\mathbb{E}}|\tilde{\Gamma}_T^{s,X_s} - \tilde{\Gamma}_T^{t,x}|^q &\leq C(T, q) \left[ (s - t)^{q-1} \tilde{\mathbb{E}} \int_t^s |f_y(\tilde{\Theta}_r^{s,X_s}) \tilde{\Gamma}_r^{s,X_s}|^q dr + \tilde{\mathbb{E}} \left( \int_t^s |f_z(\tilde{\Theta}_r^{s,X_s}) \tilde{\Gamma}_r^{s,X_s}|^2 dr \right)^{\frac{q}{2}} \right. \\ &\quad \left. + \tilde{\mathbb{E}} \int_s^T |f_y(\tilde{\Theta}_r^{s,X_s}) \tilde{\Gamma}_r^{s,X_s} - f_y(\tilde{\Theta}_r^{t,x}) \tilde{\Gamma}_r^{t,x}|^q dr \right. \\ &\quad \left. + \tilde{\mathbb{E}} \left( \int_s^T |f_z(\tilde{\Theta}_r^{s,X_s}) \tilde{\Gamma}_r^{s,X_s} - f_z(\tilde{\Theta}_r^{t,x}) \tilde{\Gamma}_r^{t,x}|^2 dr \right)^{\frac{q}{2}} \right]. \end{aligned}$$

Since  $f_y$  and  $f_z$  are bounded we have  $\tilde{\mathbb{E}}|\tilde{\Gamma}_r^{s,X_s}|^q + \tilde{\mathbb{E}}|\tilde{\Gamma}_r^{t,x}|^q \leq C(f, T, q)$ . Similarly to (31), since  $f_x, f_y, f_z$  are Lipschitz continuous with respect to the space variables,

$$\begin{aligned} |f_x(\tilde{\Theta}_r^{s,X_s}) - f_x(\tilde{\Theta}_r^{t,x})| &= |f_x(r, \tilde{X}_r^{s,X_s}, u(r, \tilde{X}_r^{s,X_s}), \sigma(r, \tilde{X}_r^{s,X_s})u_x(r, \tilde{X}_r^{s,X_s})) \\ &\quad - f_x(r, \tilde{X}_r^{t,x}, u(r, \tilde{X}_r^{t,x}), \sigma(r, \tilde{X}_r^{t,x})u_x(r, \tilde{X}_r^{t,x}))| \\ &\leq C(c_{4.2}^{2,3}, \sigma, f, T)(1 + |\tilde{X}_r^{s,X_s}|^{p_0+1} + |\tilde{X}_r^{t,x}|^{p_0+1}) \frac{|\tilde{X}_r^{s,X_s} - \tilde{X}_r^{t,x}|}{(T - r)^{\frac{1}{2}}}, \end{aligned}$$

so that Lemma 5.2 yields

$$\tilde{\mathbb{E}}|f_x(\tilde{\Theta}_r^{s,X_s}) - f_x(\tilde{\Theta}_r^{t,x})|^q \leq C(c_{4.2}^{2,3}, b, \sigma, f, T, p_0, q)(1 + |X_s|^{p_0+1} + |x|^{p_0+1})^q \frac{|X_s - x|^q + |s - t|^{\frac{q}{2}}}{(T - r)^{\frac{1}{2}}}.$$

The same holds for  $|f_y(\tilde{\Theta}_r^{s,X_s}) - f_y(\tilde{\Theta}_r^{t,x})|$  and  $|f_z(\tilde{\Theta}_r^{s,X_s}) - f_z(\tilde{\Theta}_r^{t,x})|$ . Applying these inequalities and Gronwall’s lemma, we arrive at

$$\|\tilde{\mathbb{E}}[\tilde{\Gamma}_T^{s,X_s} - \tilde{\Gamma}_T^{t,x}]\|_p \leq C(c_{4.2}^{2,3}, b, \sigma, f, g, T, p_0, p) \Psi(x) |s - t|^{\frac{1}{2}}$$

for  $p > 0$ .

For  $J_2 \leq C(t - s)$  it is enough to realise that the integrand is bounded. The estimate for  $J_3$  follows similarly to that of  $J_1$ . □

**4.4. Properties of the solution to the finite difference equation**

Recall the definition of  $\mathcal{D}_m^n$  given in (15). By (4),

$$X_{t_{m+1}}^{n,t_m,x} = x + hb(t_{m+1}, x) + \sqrt{h}\sigma(t_{m+1}, x)\varepsilon_{m+1}, \tag{42}$$

so that

$$T_{m+1,\pm} u^n(t_{m+1}, X_{t_{m+1}}^{n,t_m,x}) = u^n(t_{m+1}, x + hb(t_{m+1}, x) \pm \sqrt{h}\sigma(t_{m+1}, x)). \tag{43}$$

While for the solution to the PDE (38) one can observe in Theorem 4.2 the well-known smoothing property which implies that  $u$  is differentiable on  $[0, T] \times \mathbb{R}$  even though  $g$  is only Hölder continuous, in the following proposition, for the solution  $u^n$  to the finite difference equation we have to require from  $g$  the same regularity as we want for  $u^n$ .

**Proposition 4.2.** *Let Assumption 2.3 hold and assume that  $u^n$  is a solution of*

$$\begin{aligned} &u^n(t_m, x) - hf(t_{m+1}, x, u^n(t_m, x), \mathcal{D}_{m+1}^n u^n(t_{m+1}, X_{t_{m+1}}^{n,t_m,x})) \\ &= \frac{1}{2} [T_{m+1,+} u^n(t_{m+1}, X_{t_{m+1}}^{n,t_m,x}) + T_{m+1,-} u^n(t_{m+1}, X_{t_{m+1}}^{n,t_m,x})], \quad m = 0, \dots, n-1, \end{aligned} \tag{44}$$

with terminal condition  $u^n(t_n, x) = g(x)$ . Then, for sufficiently small  $h$ , the map  $x \mapsto u^n(t_m, x)$  is  $C^2$ , and it holds that

$$|u^n(t_m, x)| + |u_x^n(t_m, x)| \leq C_{u^n,1} \Psi(x), \quad |u_{xx}^n(t_m, x)| \leq C_{u^n,2} \Psi^2(x),$$

and

$$|u_{xx}^n(t_m, x) - u_{xx}^n(t_m, \bar{x})| \leq C_{u^n,3} (1 + |x|^{6p_0+7} + |\bar{x}|^{6p_0+7}) |x - \bar{x}|^\alpha, \tag{45}$$

uniformly in  $m = 0, \dots, n-1$ . The constants  $C_{u^n,1}$ ,  $C_{u^n,2}$ , and  $C_{u^n,3}$  depend on the bounds of  $f, g, b, \sigma$ , and their derivatives, and on  $T$  and  $p_0$ .

*Proof. Step 1.* From (44), since  $g$  is  $C^2$  and  $f_y$  is bounded, for sufficiently small  $h$  we conclude by induction (backwards in time) that  $u_x^n(t_m, x)$  exists for  $m = 0, \dots, n-1$ , and that

$$\begin{aligned} u_x^n(t_m, x) &= hf_x(t_{m+1}, x, u^n(t_m, x), \mathcal{D}_{m+1}^n u^n(t_{m+1}, X_{t_{m+1}}^{n,t_m,x})) \\ &\quad + hf_y(t_{m+1}, x, u^n(t_m, x), \mathcal{D}_{m+1}^n u^n(t_{m+1}, X_{t_{m+1}}^{n,t_m,x})) u_x^n(t_m, x) \\ &\quad + hf_z(t_{m+1}, x, u^n(t_m, x), \mathcal{D}_{m+1}^n u^n(t_{m+1}, X_{t_{m+1}}^{n,t_m,x})) \partial_x \mathcal{D}_{m+1}^n u^n(t_{m+1}, X_{t_{m+1}}^{n,t_m,x}) \\ &\quad + \frac{1}{2} \left( \partial_x T_{m+1,+} u^n(t_{m+1}, X_{t_{m+1}}^{n,t_m,x}) + \partial_x T_{m+1,-} u^n(t_{m+1}, X_{t_{m+1}}^{n,t_m,x}) \right). \end{aligned}$$

Similarly one can show that  $u_{xx}^n(t_m, x)$  exists and solves the derivative of the previous equation.

**Step 2.** As stated in the proof of Proposition 2.1, the finite difference equation (44) is the associated equation to (9) in the sense that we have the representations (21). We will use that  $u^n(t_m, x) = Y_{t_m}^{n,t_m,x}$  and exploit the BSDE

$$\begin{aligned} Y_{t_m}^{n,t_m,x} &= g(X_T^{n,t_m,x}) + \int_{(t_m, T]} f(s, X_{s-}^{n,t_m,x}, Y_{s-}^{n,t_m,x}, Z_{s-}^{n,t_m,x}) d[B^n]_s \\ &\quad - \int_{(t_m, T]} Z_{s-}^{n,t_m,x} dB_s^n, \end{aligned} \tag{46}$$

in which we will drop the superscript  $t_m, x$  from now on. For  $u_x^n(t_m, x)$  we will consider

$$\begin{aligned} \nabla Y_{t_m}^n &:= \partial_x Y_{t_m}^n = g'(X_T^n) \partial_x X_T^n + \int_{(t_m, T]} f_x \partial_x X_{s-}^n + f_y \partial_x Y_{s-}^n + f_z \partial_x Z_{s-}^n d[B^n]_s \\ &\quad - \int_{(t_m, T]} \partial_x Z_{s-}^n dB_s^n. \end{aligned} \tag{47}$$

Similarly as in the proof of [30, Theorem 3.1], the BSDE (47) can be derived from (46) as a limit of difference quotients with respect to  $x$ . Notice that the generator of (47) is random but has the same Lipschitz constant and linear growth bound as  $f$ . Assumption 2.3 allows us to find a  $p_0 \geq 0$  and a  $K > 0$  such that

$$|g(x)| + |g'(x)| + |g''(x)| \leq K(1 + |x|^{p_0+1}) = \Psi(x).$$

In order to get estimates simultaneously for (46) and (47) we prove the following lemma.

**Lemma 4.2.** *We fix  $n$  and assume a BSDE*

$$Y_{t_k} = \xi^n + \int_{(t_k, T]} f(s, X_{s-}, Y_{s-}, Z_{s-}) d[B^n]_s - \int_{(t_k, T]} Z_{s-} dB_s^n, \quad m \leq k \leq n, \tag{48}$$

with  $\xi^n = g(X_T^{n, t_m, x})$  or  $\xi^n = g'(X_T^{n, t_m, x}) \partial_x X_T^{n, t_m, x}$ , and  $X_s := X_s^{n, t_m, x}$  or  $X_s := \partial_x X_s^{n, t_m, x}$ , such that  $f: \Omega \times [0, T] \times \mathbb{R}^3 \rightarrow \mathbb{R}$  is measurable and satisfies

$$\begin{aligned} |f(\omega, t, x, y, z) - f(\omega, t, x', y', z')| &\leq L_f(|x - x'| + |y - y'| + |z - z'|), \\ |f(\omega, t, x, y, z)| &\leq (K_f + L_f)(1 + |x| + |y| + |z|). \end{aligned} \tag{49}$$

Then for any  $p \geq 2$ ,

(i)

$$\mathbb{E}|Y_{t_k}|^p + \frac{\gamma_p}{4} \mathbb{E} \int_{(t_k, T]} |Y_{s-}|^{p-2} |Z_{s-}|^2 d[B^n]_s \leq C\Psi^p(x)$$

for  $k = m, \dots, n$  and some  $\gamma_p > 0$ ,

(ii)  $\mathbb{E} \sup_{t_m < s \leq T} |Y_{s-}|^p \leq C\Psi^p(x)$ , and

(iii)  $\mathbb{E} \left( \int_{(t_m, T]} |Z_{s-}|^2 d[B^n]_s \right)^{\frac{p}{2}} \leq C\Psi^p(x)$ ,

for some constant  $C = C(b, \sigma, f, g, T, p, p_0)$ .

*Proof.*

(i) By Itô's formula (see [25, Theorem 4.57]) we get for  $p \geq 2$  that

$$\begin{aligned} |Y_{t_k}|^p &= |\xi^n|^p - p \int_{(t_k, T]} Y_{s-} |Y_{s-}|^{p-2} Z_{s-} dB_s^n \\ &\quad + p \int_{(t_k, T]} Y_{s-} |Y_{s-}|^{p-2} f(s, X_{s-}, Y_{s-}, Z_{s-}) d[B^n]_s \\ &\quad - \sum_{s \in (t_k, T]} [ |Y_s|^p - |Y_{s-}|^p - p Y_{s-} |Y_{s-}|^{p-2} (Y_s - Y_{s-}) ]. \end{aligned} \tag{50}$$

Following the proof of [27, Proposition 2] (which is carried out there in the Lévy process setting but can be done also for martingales with jumps, like  $B^n$ ) we can use the estimate

$$- \sum_{s \in (t_k, T]} [ |Y_s|^p - |Y_{s-}|^p - pY_{s-}|Y_{s-}|^{p-2}(Y_s - Y_{s-}) ] \leq -\gamma_p \sum_{s \in (t_k, T]} |Y_{s-}|^{p-2}(Y_s - Y_{s-})^2,$$

where  $\gamma_p > 0$  is computed in [40, Lemma A4]. Since

$$Y_{t_{\ell+1}} - Y_{t_{\ell+1}-} = f(t_{\ell+1}, X_{t_{\ell}}, Y_{t_{\ell}}, Z_{t_{\ell}})h - Z_{t_{\ell}}\sqrt{h}\varepsilon_{\ell+1}$$

we have

$$\begin{aligned} & - \sum_{s \in (t_k, T]} [ |Y_s|^p - |Y_{s-}|^p - pY_{s-}|Y_{s-}|^{p-2}(Y_s - Y_{s-}) ] \\ & \leq -\gamma_p \sum_{\ell=k}^{n-1} |Y_{t_{\ell}}|^{p-2} \left( f(t_{\ell+1}, X_{t_{\ell}}, Y_{t_{\ell}}, Z_{t_{\ell}})h - Z_{t_{\ell}}\sqrt{h}\varepsilon_{\ell+1} \right)^2 \\ & = -\gamma_p h \int_{(t_k, T]} |Y_{s-}|^{p-2} f^2(s, X_{s-}, Y_{s-}, Z_{s-})d[B^n]_s - \gamma_p \int_{(t_k, T]} |Y_{s-}|^{p-2} |Z_{s-}|^2 d[B^n]_s \\ & \quad + 2\gamma_p \int_{(t_k, T]} |Y_{s-}|^{p-2} f(s, X_{s-}, Y_{s-}, Z_{s-})Z_{s-}(B_s^n - B_{s-}^n)d[B^n]_s. \end{aligned}$$

Hence we get from (50) that

$$\begin{aligned} |Y_{t_k}|^p & \leq |\xi^n|^p - p \int_{(t_k, T]} Y_{s-}|Y_{s-}|^{p-2} Z_{s-}dB_s^n \\ & \quad + p \int_{(t_k, T]} Y_{s-}|Y_{s-}|^{p-2} f(s, X_{s-}, Y_{s-}, Z_{s-})d[B^n]_s \\ & \quad - \gamma_p \int_{(t_k, T]} |Y_{s-}|^{p-2} |Z_{s-}|^2 d[B^n]_s \\ & \quad + 2\gamma_p \int_{(t_k, T]} |Y_{s-}|^{p-2} f(s, X_{s-}, Y_{s-}, Z_{s-})Z_{s-}(B_s^n - B_{s-}^n)d[B^n]_s. \end{aligned}$$

From Young’s inequality and (49) we conclude that there is a  $c' = c'(p, K_f, L_f, \gamma_p) > 0$  such that

$$p|Y_{s-}|^{p-1} |f(s, X_{s-}, Y_{s-}, Z_{s-})| \leq \frac{\gamma_p}{4}|Y_{s-}|^{p-2} |Z_{s-}|^2 + c'(1 + |X_{s-}|^p + |Y_{s-}|^p),$$

and for  $\sqrt{h} < \frac{1}{8(L_f + K_f)}$  we find a  $c'' = c''(p, L_f, K_f, \gamma_p) > 0$  such that

$$\begin{aligned} 2\gamma_p \sqrt{h}|Y_{s-}|^{p-2} |f(s, X_{s-}, Y_{s-}, Z_{s-})||Z_{s-}| & \leq \frac{\gamma_p}{4}|Y_{s-}|^{p-2} |Z_{s-}|^2 \\ & \quad + c''(1 + |X_{s-}|^p + |Y_{s-}|^p). \end{aligned}$$

Then for  $c = c' + c''$  we have

$$\begin{aligned} |Y_{t_k}|^p & \leq |\xi^n|^p - p \int_{(t_k, T]} Y_{s-}|Y_{s-}|^{p-2} Z_{s-}dB_s^n + c \int_{(t_k, T]} 1 + |X_{s-}|^p + |Y_{s-}|^p d[B^n]_s \\ & \quad - \frac{\gamma_p}{2} \int_{(t_k, T]} |Y_{s-}|^{p-2} |Z_{s-}|^2 d[B^n]_s. \end{aligned} \tag{51}$$



By standard methods, approximating the terminal condition and the generator by bounded functions, it follows that for any  $a > 0$ ,

$$\mathbb{E} \sup_{t_k \leq s \leq T} |Y_s|^a < \infty \quad \text{and} \quad \mathbb{E} \left( \int_{(t_k, T]} |Z_{s-}|^2 d[B^n]_s \right)^{\frac{a}{2}} < \infty.$$

Hence  $\int_{(t_k, T]} Y_{s-} |Y_{s-}|^{p-2} Z_{s-} dB_s^n$  has expectation zero. Taking the expectation in (51) yields

$$\begin{aligned} \mathbb{E}|Y_{t_k}|^p + \frac{yp}{2} \mathbb{E} \int_{(t_k, T]} |Y_{s-}|^{p-2} |Z_{s-}|^2 d[B^n]_s &\leq \mathbb{E}|\xi^n|^p \\ &+ c \mathbb{E} \int_{(t_k, T]} 1 + |X_{s-}|^p + |Y_{s-}|^p d[B^n]_s. \end{aligned} \tag{52}$$

Since  $\mathbb{E}|\xi^n|^p$  and  $\mathbb{E} \int_{(t_k, T]} 1 + |X_{s-}|^p d[B^n]_s$  are polynomially bounded in  $x$ , Gronwall’s lemma gives

$$\|Y_{t_k}\|_p \leq C(b, \sigma, f, g, T, p, p_0)(1 + |x|^{p_0+1}), \quad k = m, \dots, n,$$

and inserting this into (52) yields

$$\begin{aligned} \left( \mathbb{E} \int_{(t_k, T]} |Y_{s-}|^{p-2} |Z_{s-}|^2 d[B^n]_s \right)^{\frac{1}{p}} &\leq C(b, \sigma, f, g, T, p, p_0)(1 + |x|^{p_0+1}), \\ k &= m, \dots, n - 1. \end{aligned}$$

(ii) From (51) we derive by the inequality of BDG and Young’s inequality that for  $t_m \leq t_k \leq T$ ,

$$\begin{aligned} &\mathbb{E} \sup_{t_k < s \leq T} |Y_{s-}|^p \\ &\leq \mathbb{E}|\xi^n|^p + C(p) \mathbb{E} \left( \int_{(t_k, T]} |Y_{s-}|^{2p-2} |Z_{s-}|^2 d[B^n]_s \right)^{\frac{1}{2}} \\ &\quad + c \mathbb{E} \int_{(t_k, T]} 1 + |X_{s-}|^p + |Y_{s-}|^p d[B^n]_s \\ &\leq \mathbb{E}|\xi^n|^p + c \mathbb{E} \int_{(t_k, T]} 1 + |X_{s-}|^p d[B^n]_s \\ &\quad + C(p) \mathbb{E} \left[ \sup_{t_k < s \leq T} |Y_{s-}|^{\frac{p}{2}} \left( \int_{(t_k, T]} |Y_{s-}|^{p-2} |Z_{s-}|^2 d[B^n]_s \right)^{\frac{1}{2}} \right] \\ &\quad + c \mathbb{E} \int_{(t_k, T]} |Y_{s-}|^p d[B^n]_s \\ &\leq \mathbb{E}|\xi^n|^p + c \mathbb{E} \int_{(t_k, T]} 1 + |X_{s-}|^p d[B^n]_s + C(p) \mathbb{E} \int_{(t_k, T]} |Y_{s-}|^{p-2} |Z_{s-}|^2 d[B^n]_s \\ &\quad + \mathbb{E} \sup_{t_k < s \leq T} |Y_{s-}|^p \left( \frac{1}{4} + c(T - t_k) \right). \end{aligned}$$

We assume that  $h$  is sufficiently small so that we find a  $t_k$  with  $c(T - t_k) < \frac{1}{4}$ . We rearrange the inequality to have  $\mathbb{E} \sup_{t_k < s \leq T} |Y_{s-}|^p$  on the left-hand side, and from (i) we conclude that

$$\begin{aligned} \mathbb{E} \sup_{t_k < s \leq T} |Y_{s-}|^p &\leq 2\mathbb{E}|\xi^n|^p + 2c\mathbb{E} \int_{(t_k, T]} 1 + |X_{s-}|^p d[B^n]_s \\ &\quad + 2C(p)\mathbb{E} \int_{(t_k, T]} |Y_{s-}|^{p-2} |Z_{s-}|^2 d[B^n]_s \\ &\leq C(b, \sigma, f, g, T, p, p_0)(1 + |x|^{(p_0+1)p}). \end{aligned}$$

Now we may repeat the above step for  $\mathbb{E} \sup_{t_\ell < s \leq t_k} |Y_{s-}|^p$  with  $c(t_k - t_\ell) < \frac{1}{4}$  and  $\xi^n = Y_T$  replaced by  $Y_{t_k}$ , and continue doing so until we eventually get the assertion (ii).

(iii) We proceed from (48):

$$\begin{aligned} \sup_{k \leq \ell \leq n} \left| \int_{(t_\ell, T]} Z_{s-} dB_s^n \right|^p \\ \leq C(p) \left( |\xi^n|^p + \sup_{k \leq \ell \leq n} |Y_{t_\ell}|^p + \left| \int_{(t_k, T]} |f(s, X_{s-}, Y_{s-}, Z_{s-})| d[B^n]_s \right|^p \right), \end{aligned}$$

so that by (49) and the inequalities of BDG and Hölder we have that

$$\begin{aligned} \mathbb{E} \left( \int_{(t_k, T]} |Z_{s-}|^2 d[B^n]_s \right)^{\frac{p}{2}} \\ \leq C(p) \left( \mathbb{E}|\xi^n|^p + \mathbb{E} \sup_{k \leq \ell \leq n} |Y_{t_\ell}|^p \right) + C(p, L_f, K_f) \mathbb{E} \left( \int_{(t_k, T]} 1 + |X_{s-}| + |Y_{s-}| d[B^n]_s \right)^p \\ + C(p, L_f, K_f)(T - t_k)^{\frac{p}{2}} \mathbb{E} \left( \int_{(t_k, T]} |Z_{s-}|^2 d[B^n]_s \right)^{\frac{p}{2}}. \end{aligned}$$

Hence for  $C(p, L_f, K_f)(T - t_k)^{\frac{p}{2}} < \frac{1}{2}$  we derive from the assertion (ii) and from the growth properties of the other terms that

$$\mathbb{E} \left( \int_{(t_k, T]} |Z_{s-}|^2 d[B^n]_s \right)^{\frac{p}{2}} \leq C(b, \sigma, f, g, T, p, p_0)(1 + |x|^{(p_0+1)p}). \tag{53}$$

Repeating this procedure eventually yields (iii). □

**Step 3.** Applying Lemma 4.2 to (46) and (47) we see that for all  $m = 0, \dots, n$  we have

$$|u^n(t_m, x)| = |Y_{t_m}^{n, t_m, x}| = (\mathbb{E}(Y_{t_m}^{n, t_m, x})^2)^{\frac{1}{2}} \leq C(b, \sigma, f, g, T, p_0)(1 + |x|^{p_0+1})$$

and

$$|u_x^n(t_m, x)| = (\mathbb{E}(\partial_x Y_{t_m}^{n, t_m, x})^2)^{\frac{1}{2}} \leq C(b, \sigma, f, g, T, p_0)(1 + |x|^{p_0+1}). \tag{54}$$

Our next aim is to show that  $u_{xx}^n(t_m, x)$  is locally Lipschitz in  $x$ . We first show that  $u_{xx}^n(t_m, x)$  has polynomial growth. We introduce the BSDE which describes  $u_{xx}^n(t_m, x)$ , for simplicity writing

$$f(t, x_1, x_2, x_3) := f(t, x, y, z) \quad \text{and} \quad D^a := \partial_{x_1}^{i_1} \partial_{x_2}^{i_2} \partial_{x_3}^{i_3} \quad \text{with} \quad a := (i_1, i_2, i_3),$$

and consider

$$\begin{aligned} \partial_x^2 Y_{t_m}^n &= g''(X_T^n)(\partial_x X_T^n)^2 + g'(X_T^n)\partial_x^2 X_T^n \\ &+ \int_{(t_m, T]} \sum_{\substack{a \in \{0, 1, 2\}^3 \\ i_1 + i_2 + i_3 = 2}} (D^a f)(s, X_{s-}^n, Y_{s-}^n, Z_{s-}^n)(\partial_x X_{s-}^n)^{i_1} (\partial_x Y_{s-}^n)^{i_2} (\partial_x Z_{s-}^n)^{i_3} d[B^n]_s \\ &+ \int_{(t_m, T]} \sum_{\substack{a \in \{0, 1\}^3 \\ i_1 + i_2 + i_3 = 1}} (D^a f)(s, X_{s-}^n, Y_{s-}^n, Z_{s-}^n)(\partial_x^2 X_{s-}^n)^{i_1} (\partial_x^2 Y_{s-}^n)^{i_2} (\partial_x^2 Z_{s-}^n)^{i_3} d[B^n]_s \\ &- \int_{(t_m, T]} \partial_x^2 Z_{s-}^n dB_s^n. \end{aligned} \tag{55}$$

We denote the generator of this BSDE by  $\hat{f}$  and notice that it is of the structure

$$\hat{f}(\omega, t, x, y, z) = f_0(\omega, t) + f_1(\omega, t)x + f_2(\omega, t)y + f_3(\omega, t)z.$$

Here  $f_0(\omega, t)$  denotes the integrand of the first integral on the right-hand side of (55), and from the previous results one concludes that  $\mathbb{E}(\int_{(t_m, T]} |f_0(s-)| d[B^n]_s)^p < \infty$ . The functions  $f_1(t) = (D^{(1,0,0)}f)(t, \cdot) = (\partial_x f)(t, \cdot)$ ,  $f_2(t) = (\partial_y f)(t, \cdot)$ , and  $f_3(t) = (\partial_z f)(t, \cdot)$  are bounded by our assumptions. We put

$$\hat{\xi}^n := g''(X_T^n)(\partial_x X_T^n)^2 + g'(X_T^n)\partial_x^2 X_T^n.$$

Denoting the solution by  $(\hat{Y}, \hat{Z})$ , we get for  $C(f_3)(T - t_m) \leq \frac{1}{2}$  that

$$\begin{aligned} &\mathbb{E}|\hat{Y}_{t_m}|^2 + \frac{1}{2}\mathbb{E} \int_{(t_m, T]} |\hat{Z}_{s-}|^2 d[B^n]_s \\ &\leq C \left[ \mathbb{E}|\hat{\xi}^n|^2 + \mathbb{E} \left( \int_{(t_m, T]} |f_0(s-)| d[B^n]_s \right)^2 + \mathbb{E} \int_{(t_m, T]} |\hat{X}_{s-}|^2 + |\hat{Y}_{s-}|^2 d[B^n]_s \right]. \end{aligned} \tag{56}$$

Now we derive the polynomial growth  $\mathbb{E}|\hat{\xi}^n|^2 \leq C\Psi^2(x)$  from the properties of  $g'$  and  $g''$  and from the fact that  $\mathbb{E} \sup_{t_m < s \leq T} |\partial_x^j X_s^n|^p$  is bounded for  $j = 1, 2$  under our assumptions. Then the estimate

$$\mathbb{E} \left( \int_{(t_m, T]} |f_0(s-)| d[B^n]_s \right)^2 \leq C\Psi^4(x)$$

can be derived from Lemma 4.2 Parts (ii) and (iii), so that Gronwall's lemma implies

$$|\hat{Y}_{t_m}^{t_m, x}| = |u_{xx}(t_m, x)| \leq C\Psi^2(x). \tag{57}$$

Finally, to show (45), one uses (55) and derives an inequality as in (56), but now for the difference  $\partial_x^2 Y_{t_m}^{n, t_m, x} - \partial_x^2 Y_{t_m}^{n, t_m, \bar{x}}$ .

Before proving (45), let us state the following lemma.

**Lemma 4.3** *Let Assumption 2.3 hold. We have*

$$\left( \mathbb{E} \sup_s |Z_{s-}^{n, t_m, x} - Z_{s-}^{n, t_m, \bar{x}}|^p \right)^{1/p} \leq C(\Psi(x)^2 + \Psi(\bar{x})^2)|x - \bar{x}|, \quad p \geq 2, \tag{58}$$

$$\mathbb{E} \left( \int_{(t_m, T]} |\partial_x Z_{s-}^{n, t_m, x} - \partial_x Z_{s-}^{n, t_m, \bar{x}}|^2 d[B^n]_s \right)^{\frac{p}{2}} \leq C(\Psi^{4p}(x) + \Psi^{4p}(\bar{x}))|x - \bar{x}|^p, \quad p \geq 2, \quad (59)$$

$$\mathbb{E} \left( \int_{(t_m, T]} |\partial_x^2 Z_{s-}^{n, t_m, x}|^2 d[B^n]_s \right)^{\frac{p}{2}} \leq C\Psi^{4p}(x), \quad p \geq 2, \quad (60)$$

for some constant  $C = C(b, \sigma, f, g, T, p, p_0)$ .

*Proof of Lemma 4.3.* Proof of (58): Introduce  $G(t_{k+1}, x) := \mathcal{D}_{k+1}^n u^n(t_{k+1}, X_{t_{k+1}}^{n, t_k, x})$ . Using the relations (42)–(43) and the bounds (54) and (57) for  $u_x^n$  and  $u_{xx}^n$ , respectively, one obtains

$$|G(t_{k+1}, x) - G(t_{k+1}, \bar{x})| \leq C(1 + |x|^{2(p_0+1)} + |\bar{x}|^{2(p_0+1)})|x - \bar{x}|, \quad x, \bar{x} \in \mathbb{R},$$

uniformly in  $t_{k+1}$ . Since  $Z_{t_k}^{n, t_m, x} = \mathcal{D}_{k+1}^n u^n(t_{k+1}, X_{t_{k+1}}^{n, t_k, \eta}) = G(t_{k+1}, \eta)$ , where  $\eta = X_{t_k}^{n, t_m, x}$ , the previous bound yields

$$|Z_{t_k}^{n, t_m, x} - Z_{t_k}^{n, t_m, \bar{x}}| \leq C(1 + |X_{t_k}^{n, t_m, x}|^{2(p_0+1)} + |X_{t_k}^{n, t_m, \bar{x}}|^{2(p_0+1)})|X_{t_k}^{n, t_m, x} - X_{t_k}^{n, t_m, \bar{x}}|$$

uniformly for each  $t_m \leq t_k < T$ . The inequality (58) then follows by applying the Cauchy–Schwarz inequality and standard  $L_p$ -estimates for the process  $X^n$ .

Proof of (59): This can be shown similarly to Lemma 4.2(iii), by considering the BSDE for the difference  $\partial_x Y_{t_m}^{n, t_m, x} - \partial_x Y_{t_m}^{n, t_m, \bar{x}}$  instead of (47) itself.

Proof of (60): This can again be shown by repeating the proof of Lemma 4.2(iii), but now for the BSDE (55). □

We return to the main proof. By our assumptions we have

$$\mathbb{E} |\hat{\xi}^{n, t_m, x} - \hat{\xi}^{n, t_m, \bar{x}}|^2 \leq C(\Psi^2(x) + \Psi^2(\bar{x}))(1 + |x|^2 + |\bar{x}|^2)|x - \bar{x}|^{2\alpha},$$

where we use  $|x - \bar{x}|^2 \leq C(1 + |x|^2 + |\bar{x}|^2)|x - \bar{x}|^{2\alpha}$ . (The term  $|x - \bar{x}|^2$  appears, for example, in the estimate of  $(\partial_x X_T^{n, t_m, x})^2 - (\partial_x X_T^{n, t_m, \bar{x}})^2$ .) To see that

$$\mathbb{E} \left( \int_{(t_m, T]} |f_0^{t_m, x}(s-) - f_0^{t_m, \bar{x}}(s-)| d[B^n]_s \right)^2 \leq C(\Psi^{10}(x) + \Psi^{10}(\bar{x}))(1 + |x|^2 + |\bar{x}|^2)|x - \bar{x}|^{2\alpha},$$

we check the terms with the highest polynomial growth. We have to deal with terms like

$$\mathbb{E} \left( \int_{(t_m, T]} |Z_{s-}^{n, t_m, x} - Z_{s-}^{n, t_m, \bar{x}}| |\partial_x Z_{s-}^{n, t_m, x}|^2 d[B^n]_s \right)^2$$

and

$$\mathbb{E} \left( \int_{(t_m, T]} |\partial_x Z_{s-}^{n, t_m, x}|^2 - |\partial_x Z_{s-}^{n, t_m, \bar{x}}|^2 d[B^n]_s \right)^2,$$

for example. We bound the first term by using (53) and (58):

$$\begin{aligned} & \mathbb{E} \left( \int_{(t_m, T]} |Z_{s-}^{n, t_m, x} - Z_{s-}^{n, t_m, \bar{x}}| |\partial_x Z_{s-}^{n, t_m, x}|^2 d[B^n]_s \right)^2 \\ & \leq \left( \mathbb{E} \sup_s |Z_{s-}^{n, t_m, x} - Z_{s-}^{n, t_m, \bar{x}}|^4 \right)^{\frac{1}{2}} \left( \mathbb{E} \left( \int_{(t_m, T]} |\partial_x Z_{s-}^{n, t_m, x}|^2 d[B^n]_s \right)^4 \right)^{\frac{1}{2}} \\ & \leq C(\Psi^4(x) + \Psi^4(\bar{x}))|x - \bar{x}|^2 \Psi^4(x). \end{aligned}$$

We bound the second term by using (53) and (59):

$$\begin{aligned} & \mathbb{E} \left( \int_{(t_m, T]} |\partial_x Z_{s-}^{n, t_m, x}|^2 - |\partial_x Z_{s-}^{n, t_m, \bar{x}}|^2 d[B^n]_s \right)^2 \\ & \leq C \mathbb{E} \int_{(t_m, T]} |\partial_x Z_{s-}^{n, t_m, x}|^2 + |\partial_x Z_{s-}^{n, t_m, \bar{x}}|^2 d[B^n]_s \int_{(t_m, T]} |\partial_x Z_{s-}^{n, t_m, x} - \partial_x Z_{s-}^{n, t_m, \bar{x}}|^2 d[B^n]_s \\ & \leq C(\Psi^2(x) + \Psi^2(\bar{x}))(\Psi^8(x) + \Psi^8(\bar{x}))|x - \bar{x}|^2 \\ & \leq C(\Psi^{10}(x) + \Psi^{10}(\bar{x}))(|x|^{2-2\alpha} + |\bar{x}|^{2-2\alpha})|x - \bar{x}|^{2\alpha} \\ & \leq C(\Psi^{10}(x) + \Psi^{10}(\bar{x}))(1 + |x|^2 + |\bar{x}|^2)|x - \bar{x}|^{2\alpha}. \end{aligned}$$

While all the other terms can be easily estimated using the results we have obtained already, for

$$\begin{aligned} & \mathbb{E} \left( \int_{(t_m, T]} |f_3^{t_m, x}(s-) - f_3^{t_m, \bar{x}}(s-)| \partial_x^2 Z_{s-}^{n, t_m, x} |d[B^n]_s \right)^2 \\ & \leq C(\Psi^{12}(x) + \Psi^{12}(\bar{x}))(1 + |x|^2 + |\bar{x}|^2)|x - \bar{x}|^{2\alpha} \end{aligned}$$

we need the bound (60).

The result then follows from Gronwall’s lemma. □

**Remark 4.1.** Under Assumption 2.3 we conclude that by Proposition 4.2 there exists a constant  $C = C(b, \sigma, f, g, T, p, p_0) > 0$  such that

$$\begin{aligned} |u^n(t_m, x) - u^n(t_m, \bar{x})| & \leq C(1 + \Psi(x) + \Psi(\bar{x}))|x - \bar{x}|, \\ |\mathcal{D}_{m+1}^n u^n(t_{m+1}, X_{t_{m+1}}^{n, t_m, x}) - \mathcal{D}_{m+1}^n u^n(t_{m+1}, X_{t_{m+1}}^{n, t_m, \bar{x}})| & \leq C(1 + \Psi^2(x) + \Psi^2(\bar{x}))|x - \bar{x}|, \\ |u_x^n(t_m, x) - u_x^n(t_m, \bar{x})| & \leq C(1 + \Psi^2(x) + \Psi^2(\bar{x}))|x - \bar{x}|, \\ |\partial_x \mathcal{D}_{m+1}^n u^n(t_{m+1}, X_{t_{m+1}}^{n, t_m, x}) - \partial_x \mathcal{D}_{m+1}^n u^n(t_{m+1}, X_{t_{m+1}}^{n, t_m, \bar{x}})| & \leq C(1 + \hat{\Psi}(x) + \hat{\Psi}(\bar{x}))|x - \bar{x}|^\alpha, \\ |\partial_x \mathcal{D}_{m+1}^n u^n(t_{m+1}, X_{t_{m+1}}^{n, t_m, x})| & \leq C(1 + \Psi^2(x)), \end{aligned} \tag{61}$$

uniformly in  $m = 0, 1, \dots, n - 1$ , where

$$\hat{\Psi}(x) := 1 + |x|^{6p_0+8}. \tag{62}$$

In addition, for

$$\partial_x F^n(t_{m+1}, x) := \partial_x f(t_{m+1}, x, u^n(t_m, x), \mathcal{D}_{m+1}^n u^n(t_{m+1}, X_{t_{m+1}}^{n, t_m, x})),$$

we have

$$|\partial_x F^n(t_{m+1}, x) - \partial_x F^n(t_{m+1}, \bar{x})| \leq C(1 + \hat{\Psi}(x) + \hat{\Psi}(\bar{x}))|x - \bar{x}|^\alpha \tag{63}$$

uniformly in  $m = 0, 1, \dots, n - 1$ . The latter inequality follows from the assumption that the partial derivatives of  $f$  are bounded and Lipschitz continuous with respect to the spatial variables, from estimates proved in Proposition 4.2, and from those stated in (61) above.

From the calculations it can be seen that in general Assumption 2.3 cannot be weakened if one needs  $\partial_x F^n(t_{m+1}, x)$  to be locally  $\alpha$ -Hölder continuous.

### 5. Technical results and estimates

In this section we collect some facts which are needed for the proofs of our results. We start with properties of the stopping times used to construct a random walk.

**Lemma 5.1** (Proposition 11.1 in [38], Lemma A.1 in [22]) *For all  $0 \leq k \leq m \leq n$  and  $p > 0$ , it holds for  $h = \frac{T}{n}$  and  $\tau_k$  as defined in (24) that*

- (i)  $\mathbb{E}\tau_k = kh$ ;
- (ii)  $\mathbb{E}|\tau_1|^p \leq C(p)h^p$ ;
- (iii)  $\mathbb{E}|B_{\tau_k} - B_{t_k}|^{2p} \leq C(p)\mathbb{E}|\tau_k - t_k|^p \leq C(p)(tkh)^{\frac{p}{2}}$ .

The next lemma lists some estimates concerning the diffusion  $X$  defined by (28) and its discretization (26), where we assume that  $B$  and  $\tilde{B}$  are connected as in (27).

**Lemma 5.2** *Under Assumption 2.1 on  $b$  and  $\sigma$ , for  $p \geq 2$  there exists a constant  $C = C(b, \sigma, T, p) > 0$  such that the following hold:*

- (i)  $\mathbb{E}|X_T^{s,y} - X_T^{t,x}|^p \leq C(|y - x|^p + |s - t|^{\frac{p}{2}})$ ,  $x, y \in \mathbb{R}, s, t \in [0, T]$ .
- (ii)  $\tilde{\mathbb{E}} \sup_{\tilde{\tau}_l \wedge t_m \leq r \leq \tilde{\tau}_{l+1} \wedge t_m} |\tilde{X}_{t_k+r}^{t_k,x} - \tilde{X}_{t_k+\tilde{\tau}_l \wedge t_m}^{t_k,x}|^p \leq Ch^{\frac{p}{4}}$ ,  $0 \leq k \leq n, 0 \leq l \leq n - k - 1, 0 \leq m \leq n - k$ .
- (iii)  $\mathbb{E}|\nabla X_T^{s,y} - \nabla X_T^{t,x}|^p \leq C(|y - x|^p + |s - t|^{\frac{p}{2}})$ ,  $x, y \in \mathbb{R}, s, t \in [0, T]$ .
- (iv)  $\mathbb{E} \sup_{0 \leq l \leq m} |\nabla X_{t_k+t_l}^{n,t_k,x}|^p \leq C$ ,  $0 \leq k \leq n, 0 \leq m \leq n - k$ .
- (v)  $\tilde{\mathbb{E}} |\tilde{X}_{t_k+t_m}^{t_k,x} - \tilde{\mathcal{X}}_{\tau_k+\tilde{\tau}_m}^{\tau_k,y}|^p \leq C(|x - y|^p + h^{\frac{p}{4}})$ ,  $0 \leq k \leq n, 0 \leq m \leq n - k$ .
- (vi)  $\tilde{\mathbb{E}} |\nabla \tilde{X}_{t_k+t_m}^{t_k,x} - \nabla \tilde{\mathcal{X}}_{\tau_k+\tilde{\tau}_m}^{\tau_k,y}|^p \leq C(|x - y|^p + h^{\frac{p}{4}})$ ,  $0 \leq k \leq n, 0 \leq m \leq n - k$ .

*Proof.*

- (i) This estimate is well-known.
- (ii) For the stochastic integral we use the inequality of BDG and then, since  $b$  and  $\sigma$  are bounded, we get by Lemma 5.1(ii) that

$$\begin{aligned} & \tilde{\mathbb{E}} \sup_{\tilde{\tau}_l \wedge t_m \leq r \leq \tilde{\tau}_{l+1} \wedge t_m} |\tilde{X}_{t_k+r}^{t_k,x} - \tilde{X}_{t_k+\tilde{\tau}_l \wedge t_m}^{t_k,x}|^p \\ & \leq C(p)(\|b\|_\infty^p \tilde{\mathbb{E}}|\tilde{\tau}_{l+1} - \tilde{\tau}_l|^p + \|\sigma\|_\infty^p \mathbb{E}|\tilde{\tau}_{l+1} - \tilde{\tau}_l|^{\frac{p}{2}}) \leq C(b, \sigma, T, p) h^{\frac{p}{2}}. \end{aligned}$$

- (iii) This can be easily seen because the process  $(\nabla X_r^{s,y})_{r \in [s,T]}$  solves the linear SDE (13) with bounded coefficients.
- (iv) The process solves (65). The estimate follows from the inequality of BDG and Gronwall’s lemma.

(v) Recall that from (4) and (26) we have

$$\tilde{\mathcal{X}}_{\tau_k+\tilde{\tau}_m}^{t_k,y} = \tilde{X}_{t_k+t_m}^{n,t_k,y} = y + \int_{(0,t_m]} b(t_k+r, \tilde{X}_{t_k+r-}^{n,t_k,y})d[\tilde{B}^n, \tilde{B}^n]_r + \int_{(0,t_m]} \sigma(t_k+r, \tilde{X}_{t_k+r-}^{n,t_k,y})d\tilde{B}_r^n,$$

and  $\tilde{X}_{t_k+t_m}^{t_k,x}$  is given by

$$\tilde{X}_{t_k+t_m}^{t_k,x} = x + \int_0^{t_m} b(t_k+r, \tilde{X}_{t_k+r}^{t_k,x})dr + \int_0^{t_m} \sigma(t_k+r, \tilde{X}_{t_k+r}^{t_k,x})d\tilde{B}_r.$$

To compare the stochastic integrals of the previous two equations we use the relation

$$\int_{(0,t_m]} \sigma(t_k+r, \tilde{X}_{t_k+r-}^{n,t_k,y})d\tilde{B}_r^n = \int_0^\infty \sum_{l=0}^{m-1} \sigma(t_{k+l+1}, \tilde{X}_{t_{k+l}}^{n,t_k,y})\mathbf{1}_{(\tilde{\tau}_l, \tilde{\tau}_{l+1}]}(r)d\tilde{B}_r.$$

We define an ‘increasing’ map,  $i(r) := t_{l+1}$  for  $r \in (t_l, t_{l+1}]$ , and a ‘decreasing’ map,  $d(r) := t_l$  for  $r \in (t_l, t_{l+1}]$ , and split the differences as follows (using Assumption 2.1(iii) for the coefficient  $b$ ):

$$\begin{aligned} & \mathbb{E}|\tilde{X}_{t_k+t_m}^{t_k,x} - \tilde{X}_{t_k+t_m}^{n,t_k,y}|^p \\ & \leq C(b, p) \left( |x - y|^p + \mathbb{E} \int_0^{t_m} |r - i(r)|^{\frac{p}{2}} + |\tilde{X}_{t_k+r}^{t_k,x} - \tilde{X}_{t_k+d(r)}^{t_k,x}|^p + |\tilde{X}_{t_k+d(r)}^{t_k,x} - \tilde{X}_{t_k+d(r)}^{n,t_k,y}|^p dr \right) \\ & \quad + C(p)\mathbb{E} \left| \int_{t_m \wedge \tilde{\tau}_m}^{t_m} \sigma(t_k+r, \tilde{X}_{t_k+r}^{t_k,x})d\tilde{B}_r \right|^p \\ & \quad + C(p)\mathbb{E} \left| \int_{t_m \wedge \tilde{\tau}_m}^{\tilde{\tau}_m} \sum_{l=0}^{m-1} \sigma(t_{k+l+1}, \tilde{X}_{t_{k+l}}^{n,t_k,y})\mathbf{1}_{(\tilde{\tau}_l, \tilde{\tau}_{l+1}]}(r)d\tilde{B}_r \right|^p \\ & \quad + C(p)\mathbb{E} \left| \int_0^{t_m \wedge \tilde{\tau}_m} \sigma(t_k+r, \tilde{X}_{t_k+r}^{t_k,x}) - \sum_{l=0}^{m-1} \sigma(t_{k+l+1}, \tilde{X}_{t_{k+l}}^{n,t_k,y})\mathbf{1}_{(\tilde{\tau}_l, \tilde{\tau}_{l+1}]}(r)d\tilde{B}_r \right|^p. \end{aligned} \tag{64}$$

We estimate the terms on the right-hand side as follows: by standard estimates for SDEs with bounded coefficients one has that

$$\mathbb{E} \int_0^{t_m} |r - i(r)|^{\frac{p}{2}} + |\tilde{X}_{t_k+r}^{t_k,x} - \tilde{X}_{t_k+d(r)}^{t_k,x}|^p dr \leq C(b, \sigma, T, p)h^{\frac{p}{2}}.$$

By the BDG inequality, the fact that  $\sigma$  is bounded, and Lemma 5.1, we conclude that

$$\begin{aligned} & \mathbb{E} \left| \int_{t_m \wedge \tilde{\tau}_m}^{t_m} \sigma(t_k+r, \tilde{X}_{t_k+r}^{t_k,x})d\tilde{B}_r \right|^p + \mathbb{E} \left| \int_{t_m \wedge \tilde{\tau}_m}^{\tilde{\tau}_m} \sum_{l=0}^{m-1} \sigma(t_{k+l+1}, \tilde{X}_{t_{k+l}}^{n,t_k,y})\mathbf{1}_{(\tilde{\tau}_l, \tilde{\tau}_{l+1}]}(r)d\tilde{B}_r \right|^p \\ & \leq C(\sigma, p)\|\sigma\|_\infty^p \mathbb{E}|\tilde{\tau}_m - t_m|^{\frac{p}{2}} \leq C(\sigma, p)(t_m h)^{\frac{p}{4}}. \end{aligned}$$

Finally, by the BDG inequality,

$$\begin{aligned} & \tilde{\mathbb{E}} \left| \int_0^{t_m \wedge \tilde{t}_m} \sigma(t_k + r, \tilde{X}_{t_k+r}^{t_k,x}) - \sum_{l=0}^{m-1} \sigma(t_{k+l+1}, \tilde{X}_{t_{k+l}}^{n,t_k,y}) \mathbf{1}_{(\tilde{t}_l, \tilde{t}_{l+1})}(r) d\tilde{B}_r \right|^p \\ & \leq C(p) \tilde{\mathbb{E}} \left( \int_0^{t_m} \sum_{l=0}^{m-1} |\sigma(t_k + r, \tilde{X}_{t_k+r}^{t_k,x}) - \sigma(t_{k+l+1}, \tilde{X}_{t_{k+l}}^{n,t_k,y})|^2 \mathbf{1}_{(\tilde{t}_l, \tilde{t}_{l+1})}(r) dr \right)^{\frac{p}{2}} \\ & \leq C(\sigma, p) \tilde{\mathbb{E}} \left( \sum_{l=0}^{m-1} \int_{\tilde{t}_l \wedge t_m}^{\tilde{t}_{l+1} \wedge t_m} |\tilde{t}_{l+1} - t_{l+1}|^{\frac{p}{2}} + |\tilde{t}_l - t_{l+1}|^{\frac{p}{2}} + |\tilde{X}_{t_k+r}^{t_k,x} - \tilde{X}_{t_k+\tilde{t}_l \wedge t_m}^{t_k,x}|^p \right. \\ & \quad \left. + |\tilde{X}_{t_k+\tilde{t}_l \wedge t_m}^{t_k,x} - \tilde{X}_{t_{k+l}}^{n,t_k,y}|^p dr \right) \\ & \leq C(\sigma, T, p) \left( h^{\frac{p}{2}} + \max_{1 \leq l < m} (\tilde{\mathbb{E}} |\tilde{t}_l - t_l|^p)^{\frac{1}{2}} \right. \\ & \quad \left. + \max_{0 \leq l < m} \left( \tilde{\mathbb{E}} \sup_{\tilde{t}_l \wedge t_m \leq r \leq \tilde{t}_{l+1} \wedge t_m} |\tilde{X}_{t_k+r}^{t_k,x} - \tilde{X}_{t_k+\tilde{t}_l \wedge t_m}^{t_k,x}|^{2p} \right)^{\frac{1}{2}} \right. \\ & \quad \left. + \tilde{\mathbb{E}} \sum_{l=0}^{m-1} |\tilde{X}_{t_k+\tilde{t}_l \wedge t_m}^{t_k,x} - \tilde{X}_{t_{k+l}}^{n,t_k,y}|^p (\tilde{t}_{l+1} - \tilde{t}_l) \right). \end{aligned}$$

Moreover, since  $\tilde{t}_{l+1} - \tilde{t}_l$  is independent of  $|\tilde{X}_{t_k+\tilde{t}_l \wedge t_m}^{t_k,x} - \tilde{X}_{t_k+t_l}^{n,t_k,y}|^p$ , by Lemma 5.1(i) we get

$$\begin{aligned} & \tilde{\mathbb{E}} \sum_{l=0}^{m-1} |\tilde{X}_{t_k+\tilde{t}_l \wedge t_m}^{t_k,x} - \tilde{X}_{t_{k+l}}^{n,t_k,y}|^p (\tilde{t}_{l+1} - \tilde{t}_l) \\ & = \tilde{\mathbb{E}} \sum_{l=0}^{m-1} |\tilde{X}_{t_k+\tilde{t}_l \wedge t_m}^{t_k,x} - \tilde{X}_{t_{k+l}}^{n,t_k,y}|^p (t_{l+1} - t_l) \\ & \leq C(T, p) \left( \tilde{\mathbb{E}} \int_0^{t_m} |\tilde{X}_{t_k+d(r)}^{t_k,x} - \tilde{X}_{t_k+d(r)}^{n,t_k,y}|^p dr + \max_{0 \leq l < m} \tilde{\mathbb{E}} |\tilde{X}_{t_k+\tilde{t}_l \wedge t_m}^{t_k,x} - \tilde{X}_{t_k+t_l}^{t_k,x}|^p \right). \end{aligned}$$

Using Lemma 5.1(iii), one concludes similarly as in the proof of (ii) that

$$\tilde{\mathbb{E}} |\tilde{X}_{t_k+\tilde{t}_l \wedge t_m}^{t_k,x} - \tilde{X}_{t_k+t_l}^{t_k,x}|^p \leq C(b, \sigma, T, p) h^{\frac{p}{4}}.$$

Then (64) combined with the above estimates implies that

$$\tilde{\mathbb{E}} |\tilde{X}_{t_k+t_m}^{t_k,x} - \tilde{X}_{t_k+t_m}^{n,t_k,y}|^p \leq C(b, \sigma, T, p) \left( |x - y|^p + h^{\frac{p}{4}} + \tilde{\mathbb{E}} \int_0^{t_m} |\tilde{X}_{t_k+d(r)}^{t_k,x} - \tilde{X}_{t_k+d(r)}^{n,t_k,y}|^p dr \right).$$

Gronwall's lemma yields

$$\tilde{\mathbb{E}} |\tilde{X}_{t_k+t_m}^{t_k,x} - \tilde{X}_{t_k+t_m}^{n,t_k,y}|^p \leq C(b, \sigma, T, p) (|x - y|^p + h^{\frac{p}{4}}).$$

(vi) We have

$$\begin{aligned} \nabla \tilde{X}_{t_k+t_m}^{n,t_k,y} &= 1 + \int_{(0,t_m]} b_x(t_k + r, X_{t_k+r-}^{n,t_k,y}) \nabla \tilde{X}_{t_k+r-}^{n,t_k,y} d[\tilde{B}^n, \tilde{B}^n]_r \\ & \quad + \int_{(0,t_m]} \sigma_x(t_k + r, \tilde{X}_{t_k+r-}^{n,t_k,y}) \nabla \tilde{X}_{t_k+r-}^{n,t_k,y} d\tilde{B}_r^n \end{aligned} \tag{65}$$



and

$$\nabla \tilde{X}_{t_k+t_m}^{t_k,x} = 1 + \int_0^{t_m} b_x(t_k+r, \tilde{X}_{t_k+r}^{t_k,x}) \nabla \tilde{X}_{t_k+r}^{t_k,x} dr + \int_0^{t_m} \sigma_x(t_k+r, \tilde{X}_{t_k+r}^{t_k,x}) \nabla \tilde{X}_{t_k+r}^{t_k,x} d\tilde{B}_r. \tag{66}$$

We may proceed similarly as in (v), except that this time the coefficients are not bounded but have linear growth. Here one uses that the integrands are bounded in any  $L_p(\mathbb{P})$ .  $\square$

Finally, we estimate the difference between the continuous-time Malliavin weight and its discrete-time counterpart.

**Lemma 5.3** *Let  $B$  and  $\tilde{B}$  be connected via (27). Under Assumption 2.1 it holds that*

$$\tilde{\mathbb{E}} |\tilde{N}_{t_m}^{t_k} \sigma(t_k, X_{t_k}) - \tilde{N}_{\tilde{t}_m}^{n, \tau_k} \sigma(t_{k+1}, \mathcal{X}_{\tau_k})|^2 \leq C(b, \sigma, T, \delta) \frac{|X_{t_k} - \mathcal{X}_{\tau_k}|^2 + h^{\frac{1}{2}}}{(t_m - t_k)^{\frac{3}{2}}}, \quad m = k + 1, \dots, n.$$

*Proof.* For  $N_{\tilde{t}_m}^{n, \tau_k}$  and  $N_{t_m}^{t_k}$  given by (12) and (18), respectively, we introduce the notation

$$\tilde{N}_{t_m}^{t_k} \sigma(t_k, X_{t_k}) =: \frac{1}{t_m - k} \int_0^{t_m - k} a_{t_k+s} d\tilde{B}_s, \quad \tilde{N}_{\tilde{t}_m}^{n, \tau_k} \sigma(t_{k+1}, \mathcal{X}_{\tau_k}) =: \frac{1}{t_m - k} \int_0^{\tilde{t}_m - k} a_{\tau_k+s}^n d\tilde{B}_s,$$

with

$$a_{t_k+s} := \nabla \tilde{X}_{t_k+s}^{t_k, X_{t_k}} \frac{\sigma(t_k, X_{t_k})}{\sigma(t_k+s, \tilde{X}_{t_k+s}^{t_k, X_{t_k}})}, \quad a_{\tau_k+s}^n := \sum_{\ell=1}^{m-k} \nabla \tilde{\mathcal{X}}_{\tau_k+\tilde{\tau}_{\ell-1}}^{\tau_k, \mathcal{X}_{\tau_k}} \frac{\sigma(t_{k+1}, \mathcal{X}_{\tau_k})}{\sigma(t_{k+\ell}, \tilde{\mathcal{X}}_{\tau_k+\tilde{\tau}_{\ell-1}})} \mathbf{1}_{s \in (\tilde{\tau}_{\ell-1}, \tilde{\tau}_{\ell}]}$$

By the inequality of BDG,

$$\begin{aligned} & (t_m - t_k)^2 \tilde{\mathbb{E}} |\tilde{N}_{t_m}^{t_k} \sigma(t_k, X_{t_k}) - \tilde{N}_{\tilde{t}_m}^{n, \tau_k} \sigma(t_{k+1}, \mathcal{X}_{\tau_k})|^2 \\ &= \tilde{\mathbb{E}} \left| \int_0^{t_m - k} a_{t_k+s} d\tilde{B}_s - \int_0^{\tilde{t}_m - k} a_{\tau_k+s}^n d\tilde{B}_s \right|^2 \\ &= \tilde{\mathbb{E}} \int_0^{t_m - k \wedge \tilde{t}_m - k} (a_{t_k+s} - a_{\tau_k+s}^n)^2 ds + \tilde{\mathbb{E}} \int_0^{\infty} a_{t_k+s}^2 \mathbf{1}_{(\tilde{t}_m - k, t_m - k)}(s) ds \\ &\quad + \tilde{\mathbb{E}} \int_0^{\infty} (a_{\tau_k+s}^n)^2 \mathbf{1}_{(t_m - k, \tilde{t}_m - k)}(s) ds \\ &\leq \sum_{\ell=1}^{m-k} \left( \tilde{\mathbb{E}} \sup_{s \in [0, t_m - k] \cap (\tilde{\tau}_{\ell-1}, \tilde{\tau}_{\ell}]} |a_{t_k+s} - a_{\tau_k+\tilde{\tau}_{\ell}}^n|^4 \right)^{\frac{1}{2}} (\tilde{\mathbb{E}} |\tilde{\tau}_{\ell} - \tilde{\tau}_{\ell-1}|^2)^{\frac{1}{2}} \\ &\quad + \left( \tilde{\mathbb{E}} \sup_{s \in [0, t_m - k]} |a_{t_k+s}|^4 + \tilde{\mathbb{E}} \max_{1 \leq \ell \leq m-k} |a_{\tau_k+\tilde{\tau}_{\ell}}^n|^4 \right)^{\frac{1}{2}} (\tilde{\mathbb{E}} |t_m - k - \tilde{t}_m - k|^2)^{\frac{1}{2}}. \end{aligned}$$

The assertion then follows from Lemma 5.1 and from the estimates

$$\tilde{\mathbb{E}} \sup_{s \in [0, t_m - k] \cap (\tilde{\tau}_{\ell-1}, \tilde{\tau}_{\ell}]} |a_{t_k+s} - a_{\tau_k+\tilde{\tau}_{\ell}}^n|^4 \leq C(b, \sigma, T, \delta) (|X_{t_k} - \mathcal{X}_{\tau_k}|^4 + h), \tag{67}$$

$$\tilde{\mathbb{E}} \sup_{s \in [0, t_m - k]} |a_{t_k+s}|^4 + \tilde{\mathbb{E}} \max_{1 \leq \ell \leq m-k} |a_{\tau_k+\tilde{\tau}_{\ell}}^n|^4 \leq 2 \|\sigma\|_{\infty}^4 \delta^{-4}. \tag{68}$$

So it remains to prove these inequalities. We put

$$\tilde{K}_{t_k+s}^{t_k} := \frac{\sigma(t_k, X_{t_k})}{\sigma(t_k + s, \tilde{X}_{t_k+s}^{t_k, X_{t_k}})} \quad \text{and} \quad \tilde{K}_{\tau_k+\tilde{\tau}_{\ell-1}}^{n, \tau_k} := \frac{\sigma(t_{k+1}, \mathcal{X}_{\tau_k})}{\sigma(t_{k+\ell}, \tilde{\mathcal{X}}_{\tau_k+\tilde{\tau}_{\ell-1}}^{\tau_k, \mathcal{X}_{\tau_k}})}$$

and notice that by Assumption 2.1 both expressions are bounded by  $\|\sigma\|_\infty \delta^{-1}$ . To show (67) let us split  $a_{t_k+s} - a_{\tau_k+\tilde{\tau}_\ell}^n$  in the following way:

$$\begin{aligned} a_{t_k+s} - a_{\tau_k+\tilde{\tau}_\ell}^n &= \tilde{K}_{t_k+s}^{t_k} (\nabla \tilde{X}_{t_k+s}^{t_k, X_{t_k}} - \nabla \tilde{X}_{t_k+t_{\ell-1}}^{t_k, X_{t_k}}) + \nabla \tilde{X}_{t_k+t_{\ell-1}}^{t_k, X_{t_k}} (\tilde{K}_{t_k+s}^{t_k} - \tilde{K}_{t_k+t_{\ell-1}}^{t_k}) \\ &\quad + \tilde{K}_{t_k+t_{\ell-1}}^{t_k} (\nabla \tilde{X}_{t_k+t_{\ell-1}}^{t_k, X_{t_k}} - \nabla \tilde{\mathcal{X}}_{\tau_k+\tilde{\tau}_{\ell-1}}^{\tau_k, \mathcal{X}_{\tau_k}}) + \nabla \tilde{\mathcal{X}}_{\tau_k+\tilde{\tau}_{\ell-1}}^{\tau_k, \mathcal{X}_{\tau_k}} (\tilde{K}_{t_k+t_{\ell-1}}^{t_k} - \tilde{K}_{\tau_k+\tilde{\tau}_{\ell-1}}^{n, \tau_k}). \end{aligned}$$

Then

$$\begin{aligned} &\tilde{\mathbb{E}} \sup_{s \in [\tilde{\tau}_{\ell-1} \wedge t_{m-k}, \tilde{\tau}_\ell \wedge t_{m-k}]} |\tilde{K}_{t_k+s}^{t_k} (\nabla \tilde{X}_{t_k+s}^{t_k, X_{t_k}} - \nabla \tilde{X}_{t_k+t_{\ell-1}}^{t_k, X_{t_k}})|^4 \\ &\leq \|\sigma\|_\infty^4 \delta^{-4} \tilde{\mathbb{E}} \sup_{s \in [\tilde{\tau}_{\ell-1} \wedge t_{m-k}, \tilde{\tau}_\ell \wedge t_{m-k}]} |\nabla \tilde{X}_{t_k+s}^{t_k, X_{t_k}} - \nabla \tilde{X}_{t_k+t_{\ell-1}}^{t_k, X_{t_k}}|^4 \leq C(b, \sigma, T, \delta)h, \end{aligned}$$

since one can show similarly to Lemma 5.2(ii) that

$$\tilde{\mathbb{E}} \sup_{s \in [\tilde{\tau}_{\ell-1} \wedge t_{m-k}, \tilde{\tau}_\ell \wedge t_{m-k}]} |\nabla \tilde{X}_{t_k+s}^{t_k, X_{t_k}} - \nabla \tilde{X}_{t_k+t_{\ell-1}}^{t_k, X_{t_k}}|^4 \leq C(b, \sigma, T, \delta)h.$$

Notice that  $\nabla \tilde{X}_t^{t_k, X_{t_k}}$  and  $\nabla \tilde{\mathcal{X}}_{\tau_m}^{\tau_k, \mathcal{X}_{\tau_k}}$  solve the linear SDEs (66) and (65), respectively. Therefore,

$$\tilde{\mathbb{E}} \sup_{s \in [0, t_{m-k}]} |\nabla \tilde{X}_{t_k+s}^{t_k, X_{t_k}}|^p \leq C(b, \sigma, T, p) \quad \text{and} \quad \tilde{\mathbb{E}} \max_{0 \leq \ell \leq m-k} |\nabla \tilde{\mathcal{X}}_{\tilde{\tau}_\ell+\tau_k}^{\tau_k, \mathcal{X}_{\tau_k}}|^p \leq C(b, \sigma, T, p). \quad (69)$$

For the second term we get

$$\begin{aligned} &\tilde{\mathbb{E}} \sup_{s \in [\tilde{\tau}_{\ell-1} \wedge t_{m-k}, \tilde{\tau}_\ell \wedge t_{m-k}]} |\nabla \tilde{X}_{t_k+t_{\ell-1}}^{t_k, X_{t_k}} (\tilde{K}_{t_k+s}^{t_k} - \tilde{K}_{t_k+t_{\ell-1}}^{t_k})|^4 \\ &\leq C(\sigma, \delta) (\tilde{\mathbb{E}} |\nabla \tilde{X}_{t_k+t_{\ell-1}}^{t_k, X_{t_k}}|^8)^{\frac{1}{2}} (\tilde{\mathbb{E}} \sup_{s \in [\tilde{\tau}_{\ell-1} \wedge t_{m-k}, \tilde{\tau}_\ell \wedge t_{m-k}]} (|t_\ell - s|^4 + |\tilde{X}_{t_k+s}^{t_k, X_{t_k}} - \tilde{X}_{t_k+t_\ell}^{t_k, X_{t_k}}|^8)^{\frac{1}{2}}) \\ &\leq C(b, \sigma, T, \delta)h. \end{aligned}$$

For the third term, Lemma 5.2(vi) implies that

$$\tilde{\mathbb{E}} |\tilde{K}_{t_k+t_{\ell-1}}^{t_k} (\nabla \tilde{X}_{t_k+t_{\ell-1}}^{t_k, X_{t_k}} - \nabla \tilde{\mathcal{X}}_{\tau_k+\tilde{\tau}_{\ell-1}}^{\tau_k, \mathcal{X}_{\tau_k}})|^4 \leq C(b, \sigma, T) \|\sigma\|_\infty^4 \delta^{-4} (|X_{t_k} - \mathcal{X}_{\tau_k}|^4 + h).$$

The last term we estimate similarly to the second one:

$$\begin{aligned} &\tilde{\mathbb{E}} |\nabla \tilde{\mathcal{X}}_{\tau_k+\tilde{\tau}_{\ell-1}}^{\tau_k, \mathcal{X}_{\tau_k}} (\tilde{K}_{t_k+t_{\ell-1}}^{t_k} - \tilde{K}_{\tau_k+\tilde{\tau}_{\ell-1}}^{n, \tau_k})|^4 \\ &\leq C(\sigma, \delta) (\tilde{\mathbb{E}} |\nabla \tilde{\mathcal{X}}_{\tau_k+\tilde{\tau}_{\ell-1}}^{\tau_k, \mathcal{X}_{\tau_k}}|^8)^{\frac{1}{2}} (|X_{t_k} - \mathcal{X}_{\tau_k}|^8 + \tilde{\mathbb{E}} |\mathcal{X}_{\tau_k+\tilde{\tau}_{\ell-1}}^{\tau_k, \mathcal{X}_{\tau_k}} - \tilde{X}_{t_k+t_{\ell-1}}^{t_k, X_{t_k}}|^8)^{\frac{1}{2}} \\ &\leq C(b, \sigma, T, \delta) (|X_{t_k} - \mathcal{X}_{\tau_k}|^4 + h). \end{aligned}$$

To see (68), use the estimates (69). □

We close this section with estimates concerning the effect of  $T_{m,\pm}$  and the discretized Malliavin derivative  $\mathcal{D}_k^n$  (see Definition 2.1) on  $X^n$ .

**Lemma 5.4** *Under Assumption 2.1, and for  $p \geq 2$ , we have the following:*

- (i)  $\mathbb{E}|X_{t_l}^n - T_{m,\pm}X_{t_l}^n|^p \leq C(b, \sigma, T, p)h^{\frac{p}{2}}, \quad 1 \leq l, m \leq n.$
- (ii)  $\mathbb{E}\left|\nabla X_{t_m}^{n,t_k,X_{t_k}^n} - \frac{\mathcal{D}_{k+1}^n X_{t_m}^n}{\sigma(t_{k+1}, X_{t_k}^n)}\right|^p \leq C(b, \sigma, T, p)h^{\frac{p}{2}}, \quad 0 \leq k < m \leq n.$
- (iii)  $\mathbb{E}|\mathcal{D}_k^n X_{t_m}^n|^p \leq C(b, \sigma, T, p), \quad 0 \leq k \leq m \leq n.$

*Proof.*

(i) By definition,  $T_{m,\pm}X_{t_l}^n = X_{t_l}^n$  for  $l \leq m - 1$ , and for  $l \geq m$  we have

$$\begin{aligned} T_{m,\pm}X_{t_l}^n &= X_{t_{m-1}}^n + b(t_m, X_{t_{m-1}}^n)h \pm \sigma(t_m, X_{t_{m-1}}^n)\sqrt{h} \\ &\quad + h \sum_{j=m+1}^l b(t_j, T_{m,\pm}X_{t_{j-1}}^n) + \sqrt{h} \sum_{j=m+1}^l \sigma(t_j, T_{m,\pm}X_{t_{j-1}}^n)\varepsilon_j. \end{aligned}$$

By the properties of  $b$  and  $\sigma$ , and thanks to the inequality of BDG and Hölder’s inequality, we see that

$$\begin{aligned} &\mathbb{E}|X_{t_l}^n - T_{m,\pm}X_{t_l}^n|^p \\ &\leq C(p) \left( \mathbb{E}|\sigma(t_m, X_{t_{m-1}}^n)\sqrt{h}(1 \pm \varepsilon_m)|^p + h^p \mathbb{E}\left| \sum_{j=m+1}^l (b(t_j, X_{t_{j-1}}^n) - b(t_j, T_{m,\pm}X_{t_{j-1}}^n)) \right|^p \right. \\ &\quad \left. + h^{\frac{p}{2}} \mathbb{E}\left| \sum_{j=m+1}^l (\sigma(t_j, X_{t_{j-1}}^n) - \sigma(t_j, T_{m,\pm}X_{t_{j-1}}^n))^2 \right|^{\frac{p}{2}} \right) \\ &\leq C(p) \left( \|\sigma\|_\infty^p h^{\frac{p}{2}} + h(\|b_x\|_\infty^p t_{l-m}^{p-1} + \|\sigma_x\|_\infty^p t_{l-m}^{\frac{p}{2}-1}) \sum_{j=m+1}^l \mathbb{E}|X_{t_{j-1}}^n - T_{m,\pm}X_{t_{j-1}}^n|^p \right). \end{aligned}$$

It remains to apply Gronwall’s lemma.

(ii) By the inequality of BDG and Hölder’s inequality,

$$\begin{aligned} &\mathbb{E}\left|\nabla X_{t_m}^{n,t_k,X_{t_k}^n} - \frac{\mathcal{D}_{k+1}^n X_{t_m}^n}{\sigma(t_{k+1}, X_{t_k}^n)}\right|^p \\ &\leq C(p, T) \left( |b_x(t_{k+1}, X_{t_k}^n)h + \sigma_x(t_{k+1}, X_{t_k}^n)\sqrt{h}\varepsilon_{k+1}|^p \right. \\ &\quad \left. + h^p \sum_{l=k+2}^m \mathbb{E}\left| b_x(t_l, X_{t_{l-1}}^n)\nabla X_{t_{l-1}}^{n,t_k,X_{t_k}^n} - b_x^{(k+1,l)} \frac{\mathcal{D}_{k+1}^n X_{t_{l-1}}^n}{\sigma(t_{k+1}, X_{t_k}^n)} \right|^p \right. \\ &\quad \left. + h^{\frac{p}{2}} \sum_{l=k+2}^m \mathbb{E}\left| \sigma_x(t_l, X_{t_{l-1}}^n)\nabla X_{t_{l-1}}^{n,t_k,X_{t_k}^n} - \sigma_x^{(k+1,l)} \frac{\mathcal{D}_{k+1}^n X_{t_{l-1}}^n}{\sigma(t_{k+1}, X_{t_k}^n)} \right|^p \right). \end{aligned}$$

Since by Lemma 5.4(i) we conclude that

$$\mathbb{E}|b_x^{(k+1,l)} - b_x(t_l, X_{t_{l-1}}^n)|^{2p} + \mathbb{E}|\sigma_x^{(k+1,l)} - \sigma_x(t_l, X_{t_{l-1}}^n)|^{2p} \leq C(b, \sigma, T, p)h^p,$$

and Lemma 5.2 implies that

$$\mathbb{E} \sup_{k+1 \leq l \leq m} \left| \nabla X_{t_{l-1}}^{n,t_k, X_{t_k}^n} \right|^{2p} \leq C(b, \sigma, T, p),$$

the assertion follows by Gronwall's lemma.

(iii) This is an immediate consequence of (i). □

### Acknowledgement

Christel Geiss would like to thank the Erwin Schrödinger Institute, Vienna, where a part of this work was written, for its hospitality and support.

### References

- [1] ALANKO, S. (2015). Regression-based Monte Carlo methods for solving nonlinear PDEs. Doctoral Thesis, New York University.
- [2] ANKIRCHNER, S., KRUSE, T. AND URUSOV, M. (2019). Wasserstein convergence rates for coin tossing approximations of continuous Markov processes. Preprint. Available at <https://arxiv.org/abs/1903.07880>.
- [3] BALLY, V. AND PAGÈS, G. (2003). A quantization algorithm for solving multidimensional discrete-time optimal stopping problems. *Bernoulli* **9**, 1003–1049.
- [4] BENDER, C. AND PARCZEWSKI, P. (2018). Discretizing Malliavin calculus. *Stoch. Proc. Appl.* **128**, 2489–2537.
- [5] BENDER, C. AND ZHANG, J. (2008). Time discretization and Markovian iteration for coupled FBSDEs. *Ann. Appl. Prob.* **18**, 143–177.
- [6] BOUCHARD, B. AND TOUZI, N. (2004). Discrete-time approximation and Monte Carlo simulation of backward stochastic differential equations. *Stoch. Proc. Appl.* **111**, 175–206.
- [7] BRIAND, P., DELYON, B. AND MÉMIN, J. (2001). Donsker-type theorem for BSDEs. *Electron. Commun. Probab.* **6**, 1–14.
- [8] BRIAND, P. AND LABART, C. (2014). Simulation of BSDEs by Wiener chaos expansion. *Ann. Appl. Prob.* **24**, 1129–1171.
- [9] CHASSAGNEUX, J.-F. (2014). Linear multistep schemes for BSDEs. *SIAM J. Numer. Anal.* **52**, 2815–2836.
- [10] CHASSAGNEUX, J.-F. AND CRISAN, D. (2014). Runge–Kutta schemes for backward stochastic differential equations. *Ann. Appl. Prob.* **24**, 679–720.
- [11] CHASSAGNEUX, J.-F., CRISAN, D. AND DELARUE, F. (2019). Numerical method for FBSDEs of McKean–Vlasov type. *Ann. Appl. Prob.* **29**, 1640–1684.
- [12] CHASSAGNEUX, J.-F. AND GARCIA TRILLOS, C. A. (2017). Cubature methods to solve BSDEs: error expansion and complexity control. Preprint. Available at <https://arxiv.org/abs/1702.00999>.
- [13] CHASSAGNEUX, J.-F. AND RICHOU, A. (2015). Numerical stability analysis of the Euler scheme for BSDEs. *SIAM J. Numer. Anal.* **53**, 1172–1193.
- [14] CHASSAGNEUX, J.-F. AND RICHOU, A. (2019). Rate of convergence for discrete-time approximation of reflected BSDEs arising in switching problems. *Stoch. Proc. Appl.* **129**, 4597–4637.
- [15] CHAUDRU DE RAYNAL, P. E. AND GARCIA TRILLOS, C. A. (2015). A cubature based algorithm to solve decoupled McKean–Vlasov forward-backward stochastic differential equations. *Stoch. Proc. Appl.* **125**, 2206–2255.
- [16] CHERIDITO, P. AND STADJE, M. (2013). BSΔEs and BSDEs with non-Lipschitz drivers: comparison, convergence and robustness. *Bernoulli* **19**, 1047–1085.
- [17] CRISAN, D., MANOLARAKIS, K. AND TOUZI, N. (2010). On the Monte Carlo simulation of BSDEs: an improvement on the Malliavin weights. *Stoch. Proc. Appl.* **120**, 1133–1158.
- [18] DELARUE, F. AND MENOZZI, S. (2006). A forward-backward stochastic algorithm for quasi-linear PDEs. *Ann. Appl. Prob.* **16**, 140–184.
- [19] EL KAROUI, N., PENG, S. AND QUENEZ, M. C. (1997). Backward stochastic differential equations in finance. *Math. Finance* **7**, 1–71.

- [20] GEISS, C., GEISS, S. AND GOBET, E. (2012). Generalized fractional smoothness and  $L_p$ -variation of BSDEs with non-Lipschitz terminal conditions. *Stoch. Proc. Appl.* **122**, 2078–2116.
- [21] GEISS, C. AND LABART, C. (2016). Simulation of BSDEs with jumps by Wiener chaos expansion. *Stoch. Proc. Appl.* **126**, 2123–2162.
- [22] GEISS, C., LABART, C. AND LUOTO, A. (2020). Random walk approximation of BSDEs with Hölder continuous terminal condition. *Bernoulli* **26**, 159–190.
- [23] GOBET, E., LEMOR, J.-P. AND WARIN, X. (2005). A regression-based Monte Carlo method to solve backward stochastic differential equations. *Ann. Appl. Prob.* **15**, 2172–2202.
- [24] HENRY-LABORDÈRE, P., TAN, X. AND TOUZI, N. (2014). A numerical algorithm for a class of BSDEs via the branching process. *Stoch. Proc. Appl.* **124**, 1112–1140.
- [25] JACOD, J. AND SHIRYAEV, A. N. (2003). *Limit Theorems for Stochastic Processes*. Springer, Berlin, Heidelberg.
- [26] JAŃCZAK-BORKOWSKA, K. (2012). Discrete approximations of generalized RBSDE with random terminal time. *Discuss. Math. Prob. Statist.* **32**, 69–85.
- [27] KRUSE, T AND POPIER, A. (2016). BSDEs with monotone generator driven by Brownian and Poisson noises in a general filtration. *Stochastics* **88**, 491–539.
- [28] KRUSE, T. AND POPIER, A. (2017).  $L_p$ -solution for BSDEs with jumps in the case  $p < 2$ . *Stochastics* **89**, 1201–1227.
- [29] MA, J., PROTTER, P., SAN MARTÍN, J. AND TORRES, S. (2007). Numerical method for backward stochastic differential equations. *Ann. Appl. Prob.* **12**, 302–316.
- [30] MA, J. AND ZHANG, J. (2002). Representation theorems for backward stochastic differential equations. *Ann. Appl. Prob.* **12**, 1390–1418.
- [31] MARTÍNEZ, M., SAN MARTÍN, J. AND TORRES, S. (2011). Numerical method for reflected backward stochastic differential equations. *Stoch. Anal. Appl.* **29**, 1008–1032.
- [32] MÉMIN, J., PENG, S. AND XU, M. (2008). Convergence of solutions of discrete reflected backward SDE's and simulations. *Acta Math. Appl. Sin. Engl. Ser.* **24**, 1–18.
- [33] PENG, S. AND XU, M. (2011). Numerical algorithms for backward stochastic differential equations with 1-d Brownian motion: convergence and simulations. *Math. Modelling Numer. Anal.* **45**, 335–360.
- [34] PRIVAULT, N. (2009). *Stochastic Analysis in Discrete and Continuous Settings—With Normal Martingales*. Springer, Berlin, Heidelberg.
- [35] TAO, T. (2010). *An Epsilon of Room, I: Real Analysis*. American Mathematical Society, Providence, RI.
- [36] TOLDO, S. (2006). Stability of solutions of BSDEs with random terminal time. *ESAIM Prob. Statist.* **10**, 141–163.
- [37] TOLDO, S. (2007). Corrigendum to ‘Stability of solutions of BSDEs with random terminal time’. *ESAIM Prob. Statist.* **11**, 381–384.
- [38] WALSH, J. B. (2003). The rate of convergence of the binomial tree scheme. *Finance Stoch.* **7**, 337–361.
- [39] WEINAN, E., HUTZENTHALER, M., JENTZEN, A. AND KRUSE, T. (2019). On multilevel Picard numerical approximations for high-dimensional nonlinear parabolic partial differential equations and high-dimensional nonlinear backward stochastic differential equations. *J. Sci. Comput.* **79**, 1534–1571.
- [40] YAO, S. (2017).  $L_p$  solutions of backward stochastic differential equations with jumps. *Stoch. Proc. Appl.* **127**, 3465–3511.
- [41] ZHANG, J. (2001). Some fine properties of backward stochastic differential equations, with applications. Doctoral Thesis, Purdue University.
- [42] ZHANG, J. (2004). A numerical scheme for BSDEs. *Ann. Appl. Prob.* **14**, 459–488.
- [43] ZHANG, J. (2005). Representation of solutions to BSDEs associated with a degenerate FSDE. *Ann. Appl. Prob.* **15**, 1798–1831.