



# Overcoming the Curse of Dimensionality in the Numerical Approximation of Parabolic Partial Differential Equations with Gradient-Dependent Nonlinearities

Martin Hutzenthaler<sup>1</sup> · Arnulf Jentzen<sup>2,3</sup> · Thomas Kruse<sup>4</sup>

Received: 2 March 2020 / Revised: 18 January 2021 / Accepted: 19 January 2021 / Published online: 13 July 2021  
© The Author(s) 2021

## Abstract

Partial differential equations (PDEs) are a fundamental tool in the modeling of many real-world phenomena. In a number of such real-world phenomena the PDEs under consideration contain gradient-dependent nonlinearities and are high-dimensional. Such high-dimensional nonlinear PDEs can in nearly all cases not be solved explicitly, and it is one of the most challenging tasks in applied mathematics to solve high-dimensional nonlinear PDEs approximately. It is especially very challenging to design approximation algorithms for nonlinear PDEs for which one can rigorously prove that they do overcome the so-called curse of dimensionality in the sense that the number of computational operations of the approximation algorithm needed to achieve an approximation precision of size  $\varepsilon > 0$  grows at most polynomially in both the PDE dimension  $d \in \mathbb{N}$  and the reciprocal of the prescribed approximation accuracy  $\varepsilon$ . In particular, to the best of our knowledge there exists no approximation algorithm in the scientific literature which has been proven to overcome the curse of dimensionality in the case of a class of nonlinear PDEs with general time horizons and gradient-dependent nonlinearities. It is the key contribution of this article to overcome this difficulty. More

---

Communicated by Endre Süli.

---

✉ Martin Hutzenthaler  
martin.hutzenthaler@uni-due.de

Arnulf Jentzen  
ajentzen@uni-muenster.de

Thomas Kruse  
thomas.kruse@math.uni-giessen.de

<sup>1</sup> Faculty of Mathematics, University of Duisburg-Essen, 45117 Essen, Germany

<sup>2</sup> Applied Mathematics: Institute for Analysis and Numerics, Faculty of Mathematics and Computer Science, University of Münster, 48149 Münster, Germany

<sup>3</sup> Seminar for Applied Mathematics, Department of Mathematics, ETH Zürich, 8092 Zürich, Switzerland

<sup>4</sup> Institute of Mathematics, University of Gießen, 35392 Gießen, Germany

specifically, it is the key contribution of this article (i) to propose a new full-history recursive multilevel Picard approximation algorithm for high-dimensional nonlinear heat equations with general time horizons and gradient-dependent nonlinearities and (ii) to rigorously prove that this full-history recursive multilevel Picard approximation algorithm does indeed overcome the curse of dimensionality in the case of such nonlinear heat equations with gradient-dependent nonlinearities.

**Keywords** Curse of dimensionality · Partial differential equation · PDE · Backward stochastic differential equation · BSDE · Multilevel Picard · Multilevel Monte Carlo · Gradient-dependent nonlinearity

**Mathematics Subject Classification** 65M75

## 1 Introduction

Partial differential equations (PDEs) play a prominent role in the modeling of many real-world phenomena. For instance, PDEs appear in financial engineering in models for the pricing of financial derivatives, PDEs such as the Schrödinger equation appear in quantum physics to describe the wave function of a quantum-mechanical system, PDEs are used in operations research to characterize the value function of control problems, PDEs provide solutions for backward stochastic differential equations (BSDEs) which itself appear in several models from applications, and stochastic PDEs such as the Zakai equation or the Kushner equation appear in nonlinear filtering problems to describe the density of the state of a physical system with only partial information available.

The PDEs in the above named models contain often nonlinearities and are typically high-dimensional, where, e.g., in the models from financial engineering the dimension of the PDE usually corresponds to the number of financial assets in the associated hedging or trading portfolio, where, e.g., in quantum physics the dimension of the PDE is, loosely speaking, three times the number of electrons in the considered physical system, where, e.g., in optimal control problems the dimension of the PDE is determined by the dimension of the state space of the control problem, and where, e.g., in nonlinear filtering problems the dimension of the PDE corresponds to the degrees of freedom in the considered physical system.

Such high-dimensional nonlinear PDEs can in nearly all cases not be solved explicitly and it is one of the most challenging tasks in applied mathematics to solve high-dimensional nonlinear PDEs approximately. In particular, it is very challenging to design approximation methods for nonlinear PDEs for which one can rigorously prove that they do overcome the so-called curse of dimensionality in the sense that the number of computational operations of the approximation method needed to achieve an approximation precision of size  $\varepsilon > 0$  grows at most polynomially in both the PDE dimension  $d \in \mathbb{N}$  and the reciprocal of the prescribed approximation accuracy  $\varepsilon$ .

Recently, several new stochastic approximation methods for certain classes of high-dimensional nonlinear PDEs have been proposed and studied in the scientific literature. In particular, we refer, e.g., to [11, 12, 26, 29, 30, 53] for BSDE-based approximation methods for PDEs in which nested conditional expectations are discretized through suitable regression methods, we refer, e.g., to [10, 39, 41, 42] for

branching diffusion approximation methods for PDEs, we refer, e.g., to [1–3, 6–8, 13, 14, 16, 17, 21, 24, 25, 31, 34–36, 40, 43, 48, 50, 52, 54–58, 60, 62, 63] for deep learning based approximation methods for PDEs, and we refer to [4, 5, 20, 28, 46, 47] for numerical simulations, approximation results, and extensions of the in [19, 45] recently introduced full-history recursive multilevel Picard approximation methods for PDEs. In the following we abbreviate *full-history recursive multilevel Picard* as *MLP*.

Branching diffusion approximation methods are also in the case of certain nonlinear PDEs as efficient as plain vanilla Monte Carlo approximations in the case of linear PDEs, but the error analysis only applies in the case where the time horizon  $T \in (0, \infty)$  and the initial condition, respectively, are sufficiently small and branching diffusion approximation methods fail to converge in the case where the time horizon  $T \in (0, \infty)$  exceeds a certain threshold (cf., e.g., [41, Theorem 3.12]). For MLP approximation methods it has been recently shown in [4, 45, 46] that such algorithms do indeed overcome the curse of dimensionality for certain classes of gradient-independent PDEs. Numerical simulations for deep learning based approximation methods for nonlinear PDEs in high dimensions are very encouraging (see, e.g., the above named references [1–3, 6–8, 13, 14, 16, 17, 21, 24, 25, 31, 34–36, 40, 43, 48, 50, 52, 54–58, 60, 62, 63]) but so far there is only partial error analysis available for such algorithms (which, in turn, is strongly based on the above-mentioned error analysis for the MLP approximation method; cf. [44] and, e.g., [9, 23, 32, 33, 36, 49, 51, 61, 62]). To sum up, to the best of our knowledge until today the MLP approximation method (see [45]) is the only approximation method in the scientific literature for which it has been shown that it does overcome the curse of dimensionality in the numerical approximation of semilinear PDEs with general time horizons.

The above-mentioned articles [4, 28, 45, 46] prove, however, only in the case of gradient-independent nonlinearities that MLP approximation methods overcome the curse of dimensionality and it remains an open problem to overcome the curse of dimensionality in the case of PDEs with general time horizons and gradient-dependent nonlinearities. This is precisely the subject of this article. More specifically, in this article we propose a new MLP approximation method for nonlinear heat equations with gradient-dependent nonlinearities and the main result of this article, Theorem 5.2 in Sect. 5 below, proves that the number of realizations of scalar random variables required by this MLP approximation method to achieve a precision of size  $\varepsilon > 0$  grows at most polynomially in both the PDE dimension  $d \in \mathbb{N}$  and the reciprocal of the prescribed approximation accuracy  $\varepsilon$ . To illustrate the findings of the main result of this article in more detail, we now present in the following theorem a special case of Theorem 5.2.

**Theorem 1.1** *Let  $T, \delta, \lambda \in (0, \infty)$ , let  $u_d = (u_d(t, x))_{(t,x) \in [0,T] \times \mathbb{R}^d} \in C^{1,2}([0, T] \times \mathbb{R}^d, \mathbb{R})$ ,  $d \in \mathbb{N}$ , be at most polynomially growing functions, let  $f_d \in C(\mathbb{R} \times \mathbb{R}^d, \mathbb{R})$ ,  $d \in \mathbb{N}$ , let  $g_d \in C(\mathbb{R}^d, \mathbb{R})$ ,  $d \in \mathbb{N}$ , let  $L_{d,i} \in \mathbb{R}$ ,  $d, i \in \mathbb{N}$ , assume for all  $d \in \mathbb{N}$ ,  $t \in [0, T]$ ,  $x = (x_1, x_2, \dots, x_d)$ ,  $\mathfrak{x} = (\mathfrak{x}_1, \mathfrak{x}_2, \dots, \mathfrak{x}_d)$ ,  $z = (z_1, z_2, \dots, z_d)$ ,  $\mathfrak{z} = (\mathfrak{z}_1, \mathfrak{z}_2, \dots, \mathfrak{z}_d) \in \mathbb{R}^d$ ,  $y, \mathfrak{y} \in \mathbb{R}$  that*

$$\begin{aligned} & \max\{|f_d(y, z) - f_d(\mathfrak{y}, \mathfrak{z})|, |g_d(x) - g_d(\mathfrak{x})|\} \\ & \leq \sum_{j=1}^d L_{d,j} (d^\lambda |x_j - \mathfrak{x}_j| + |y - \mathfrak{y}| + |z_j - \mathfrak{z}_j|), \end{aligned} \tag{1}$$

$$\begin{aligned} \left(\frac{\partial}{\partial t} u_d\right)(t, x) &= (\Delta_x u_d)(t, x) \\ &+ f_d(u_d(t, x), (\nabla_x u_d)(t, x)), \quad u_d(0, x) = g_d(x), \end{aligned} \quad (2)$$

and  $d^{-\lambda}(|g_d(0)| + |f_d(0, 0)|) + \sum_{i=1}^d L_{d,i} \leq \lambda$ , let  $(\Omega, \mathcal{F}, \mathbb{P})$  be a probability space, let  $\Theta = \cup_{n \in \mathbb{N}} \mathbb{Z}^n$ , let  $Z^{d,\theta}: \Omega \rightarrow \mathbb{R}^d$ ,  $d \in \mathbb{N}$ ,  $\theta \in \Theta$ , be i.i.d. standard normal random variables, let  $\tau^\theta: \Omega \rightarrow (0, 1)$ ,  $\theta \in \Theta$ , be i.i.d. random variables, assume for all  $b \in (0, 1)$  that  $\mathbb{P}(\tau^\theta \leq b) = \sqrt{b}$ , assume that  $(Z^{d,\theta})_{(d,\theta) \in \mathbb{N} \times \Theta}$  and  $(\tau^\theta)_{\theta \in \Theta}$  are independent, let  $\mathbf{U}_{n,M}^{d,\theta} = (\mathbf{U}_{n,M}^{d,\theta,0}, \mathbf{U}_{n,M}^{d,\theta,1}, \dots, \mathbf{U}_{n,M}^{d,\theta,d}): (0, T] \times \mathbb{R}^d \times \Omega \rightarrow \mathbb{R}^{1+d}$ ,  $n \in \mathbb{Z}$ ,  $M, d \in \mathbb{N}$ ,  $\theta \in \Theta$ , satisfy for all  $n, M, d \in \mathbb{N}$ ,  $\theta \in \Theta$ ,  $t \in (0, T]$ ,  $x \in \mathbb{R}^d$  that  $\mathbf{U}_{-1,M}^{d,\theta}(t, x) = \mathbf{U}_{0,M}^{d,\theta}(t, x) = 0$  and

$$\begin{aligned} &\mathbf{U}_{n,M}^{d,\theta}(t, x) - (g_d(x), 0) \\ &= \sum_{i=1}^{M^n} \frac{1}{M^n} (g_d(x + [2t]^{1/2} Z^{d,(\theta,0,-i)}) - g_d(x)) (1, [2t]^{-1/2} Z^{d,(\theta,0,-i)}) \\ &+ \sum_{l=0}^{n-1} \sum_{i=1}^{M^{n-l}} \frac{2t[\tau^{(\theta,l,i)}]^{1/2}}{M^{n-l}} \left[ f_d \left( \mathbf{U}_{l,M}^{d,(\theta,l,i)}(t(1 - \tau^{(\theta,l,i)}), x + [2t\tau^{(\theta,l,i)}]^{1/2} Z^{d,(\theta,l,i)}) \right) \right. \\ &- \mathbb{1}_{\mathbb{N}}(l) f_d \left( \mathbf{U}_{l-1,M}^{d,(\theta,-l,i)}(t(1 - \tau^{(\theta,l,i)}), x + [2t\tau^{(\theta,l,i)}]^{1/2} Z^{d,(\theta,l,i)}) \right) \\ &\left. \cdot (1, [2t\tau^{(\theta,l,i)}]^{-1/2} Z^{d,(\theta,l,i)}) \right], \end{aligned} \quad (3)$$

and for every  $d, M, n \in \mathbb{N}$  let  $\text{RV}_{d,n,M} \in \mathbb{N}$  be the number of realizations of scalar random variables which are used to compute one realization of  $\mathbf{U}_{n,M}^{d,0}(T, 0): \Omega \rightarrow \mathbb{R}$  (cf. (176) for a precise definition). Then there exist  $c \in \mathbb{R}$  and  $N = (N_{d,\varepsilon})_{(d,\varepsilon) \in \mathbb{N} \times (0,1]}: \mathbb{N} \times (0, 1] \rightarrow \mathbb{N}$  such that for all  $d \in \mathbb{N}$ ,  $\varepsilon \in (0, 1]$  it holds that  $\sum_{n=1}^{N_{d,\varepsilon}} \text{RV}_{d,n,\lfloor n^{1/4} \rfloor} \leq cd^c \varepsilon^{-(2+\delta)}$  and

$$\begin{aligned} &\sup_{n \in \mathbb{N} \cap [N_{d,\varepsilon}, \infty)} \left[ \mathbb{E}[|\mathbf{U}_{n,\lfloor n^{1/4} \rfloor}^{d,0,0}(T, 0) - u_d(T, 0)|^2] \right]^{1/2} \\ &+ \max_{i \in \{1, 2, \dots, d\}} \mathbb{E}[|\mathbf{U}_{n,\lfloor n^{1/4} \rfloor}^{d,0,i}(T, 0) - (\frac{\partial}{\partial x_i} u_d)(T, 0)|^2]^{1/2} \leq \varepsilon. \end{aligned} \quad (4)$$

Theorem 1.1 is an immediate consequence of Corollary 5.4 in Sect. 5 below. Corollary 5.4, in turn, follows from Theorem 5.2 in Sect. 5, which is the main result of this article. In the following we add a few comments regarding some of the mathematical objects appearing in Theorem 1.1 above.

The real number  $T \in (0, \infty)$  in Theorem 1.1 above describes the time horizon of the PDE under consideration (see (2) in Theorem 1.1 above). Theorem 1.1 reveals under suitable Lipschitz assumptions that the MLP approximation method in (3) above overcomes the curse of dimensionality in the numerical approximation of the gradient-dependent semilinear PDEs in (2) above. Theorem 1.1 even proves that the computational effort of the MLP approximation method in (3) required to obtain a precision of size  $\varepsilon \in (0, 1]$  is bounded by  $cd^c \varepsilon^{-(2+\delta)}$  where  $c \in \mathbb{R}$  is a constant which is completely independent of the PDE dimension  $d \in \mathbb{N}$  and where  $\delta \in (0, \infty)$  is an

arbitrarily small positive real number which describes the convergence order which we lose when compared to standard Monte Carlo approximations of linear heat equations. The real number  $\lambda \in (0, \infty)$  in Theorem 1.1 above is an arbitrary large constant which we employ to formulate the Lipschitz and growth assumptions in Theorem 1.1 (see (1) and below (2) in Theorem 1.1 above).

The functions  $u_d: [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$ ,  $d \in \mathbb{N}$ , in Theorem 1.1 above are the solutions of the PDEs under consideration; see (2) in Theorem 1.1 above. Note that for every  $d \in \mathbb{N}$  we have that (2) is a PDE where the time variable  $t \in [0, T]$  takes values in the interval  $[0, T]$  and where the space variable  $x \in \mathbb{R}^d$  takes values in the  $d$ -dimensional Euclidean space  $\mathbb{R}^d$ . The functions  $f_d: \mathbb{R} \times \mathbb{R}^d \rightarrow \mathbb{R}$ ,  $d \in \mathbb{N}$ , describe the nonlinearities of the PDEs in (2) and the functions  $g_d: \mathbb{R}^d \rightarrow \mathbb{R}$ ,  $d \in \mathbb{N}$ , describe the initial conditions of the PDEs in (2). The quantities  $\lfloor n^{1/4} \rfloor$ ,  $n \in \mathbb{N}$ , in (4) in Theorem 1.1 above describe evaluations of the standard floor function in the sense that for all  $n \in \mathbb{N}$  it holds that  $\lfloor n^{1/4} \rfloor = \max([0, n^{1/4}] \cap \mathbb{N})$ . Note that in this work for every  $d \in \mathbb{N}$ ,  $a \in \mathbb{R}$ ,  $b = (b_1, b_2, \dots, b_d) \in \mathbb{R}^d$  we sometimes write  $(a, b) \in \mathbb{R}^{d+1}$  as an abbreviation for the vector  $(a, b_1, b_2, \dots, b_d) \in \mathbb{R}^{d+1}$ . In particular, observe that for all  $d \in \mathbb{N}$ ,  $x = (x_1, x_2, \dots, x_d)$ ,  $z = (z_1, z_2, \dots, z_d) \in \mathbb{R}^d$ ,  $t \in (0, T]$  it holds that  $(g_d(x + [2t]^{1/2}z) - g_d(x))(1, [2t]^{-1/2}z) = (g_d(x + z\sqrt{2t}) - g_d(x))(1, z_1\sqrt{(2t)^{-1}}, z_2\sqrt{(2t)^{-1}}, \dots, z_d\sqrt{(2t)^{-1}})$  (cf. the first line in (3) above).

Theorem 1.1 proves under suitable assumptions (cf. (1) in Theorem 1.1 above for details) that the MLP approximation method in (3) above overcomes the curse of dimensionality in the numerical approximation of the gradient-dependent semilinear heat equations in (2) above. In order to give the reader a better understanding why the approximation scheme in (3) above is capable of overcoming the curse of dimensionality in the numerical approximation of gradient-dependent semilinear PDEs, we now outline a brief derivation of the approximation scheme in (3) in the special case where  $f \in C(\mathbb{R} \times \mathbb{R}^d, \mathbb{R})$ ,  $d \in \mathbb{N}$ , and  $g_d \in C(\mathbb{R}^d, \mathbb{R})$ ,  $d \in \mathbb{N}$ , are globally bounded functions. For this observe that the Feynman–Kac formula and the assumptions of Theorem 1.1 ensure that the solutions  $u_d: [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$ ,  $d \in \mathbb{N}$ , of the PDEs in (2) satisfy that for all  $d \in \mathbb{N}$ ,  $t \in [0, T]$ ,  $x \in \mathbb{R}^d$  it holds that

$$u_d(t, x) = \mathbb{E} \left[ g_d(x + [2t]^{1/2}Z^{d,0}) + \int_0^t f_d(u_d(s, x + [2(t-s)]^{1/2}Z^{d,0}), (\nabla_x u_d)(s, x + [2(t-s)]^{1/2}Z^{d,0})) ds \right]. \quad (5)$$

Note that (5) are stochastic fixed-point equations with the solutions  $u_d: [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$ ,  $d \in \mathbb{N}$ , of the PDEs in (2) being the fixed points of the stochastic fixed-point equations. MLP approximation methods are powerful approximation techniques for stochastic fixed-point equations but the MLP approach cannot be directly applied to (5) as the gradients  $(\nabla_x u_d)(t, x)$ ,  $t \in [0, T]$ ,  $x \in \mathbb{R}^d$ , of the solutions  $u_d: [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$ ,  $d \in \mathbb{N}$ , of the PDEs in (2) appear on the right-hand side of (5) but not on the left-hand side of (5).

To get over this obstacle, we reformulate (5) by means of the Bismut–Elworthy–Li formula (cf., e.g., Da Prato and Zabczyk [15, Theorem 2.1]) to obtain other stochastic fixed-point equations with no derivatives appearing on the right-hand side of the stochastic fixed point equations so that the MLP machinery can be brought into play. More formally, note that the Bismut–Elworthy–Li formula ensures that the solutions  $u_d : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$ ,  $d \in \mathbb{N}$ , of the PDEs in (2) satisfy that for all  $d \in \mathbb{N}$ ,  $t \in (0, T]$ ,  $x \in \mathbb{R}^d$  it holds that

$$\begin{aligned} (\nabla_x u_d)(t, x) &= \mathbb{E} \left[ g_d(x + [2t]^{1/2} Z^{d,0}) [2t]^{-1/2} Z^{d,0} \right. \\ &\quad \left. + \int_0^t f_d(u_d(s, x + [2(t-s)]^{1/2} Z^{d,0})) [2(t-s)]^{-1/2} Z^{d,0} ds \right]. \end{aligned} \quad (6)$$

Next let  $\mathbf{u}_d = (\mathbf{u}_{d,1}, \mathbf{u}_{d,2}, \dots, \mathbf{u}_{d,d+1}) : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^{d+1}$ ,  $d \in \mathbb{N}$ , satisfy for all  $d \in \mathbb{N}$ ,  $k \in \{2, 3, \dots, d+1\}$ ,  $x = (x_1, x_2, \dots, x_d) \in \mathbb{R}^d$ ,  $t \in [0, T]$  that  $\mathbf{u}_{d,1}(t, x) = u(t, x)$  and  $\mathbf{u}_{d,k}(t, x) = (\frac{\partial}{\partial x_{k-1}})u(t, x)$  and observe that (5) and (6) reveal that for all  $d \in \mathbb{N}$ ,  $t \in (0, T]$ ,  $x \in \mathbb{R}^d$  it holds that

$$\begin{aligned} \mathbf{u}_d(t, x) &= \mathbb{E} \left[ g_d(x + [2t]^{1/2} Z^{d,0}) (1, [2t]^{-1/2} Z^{d,0}) \right. \\ &\quad \left. + \int_0^t f_d(\mathbf{u}_d(s, x + [2(t-s)]^{1/2} Z^{d,0})) (1, [2(t-s)]^{-1/2} Z^{d,0}) ds \right]. \end{aligned} \quad (7)$$

Observe that (7) are stochastic fixed-point equations with no derivatives appearing on the right-hand side of (7) so that we are now in the position to apply the MLP approach to (7). Next we briefly sketch how MLP approximations for (7) can be obtained and, thereby, we briefly outline the derivation of the MLP approximation schemes in (3) above (cf., e.g., E et al. [18, Section 1.2]). For this observe that (7) can also be written as the fixed point equation  $u = \Phi(u)$  where  $\Phi$  is the self-mapping on the set of all bounded functions in  $C((0, T] \times \mathbb{R}^d, \mathbb{R}^{d+1})$  which is described through the right-hand side of (7). We now define Picard iterates  $u_n$ ,  $n \in \mathbb{N}_0$ , by means of the recursion that for all  $n \in \mathbb{N}$  it holds that  $u_0 = 0$  and  $u_n = \Phi(u_{n-1})$ . Next we observe that a telescoping sum argument and the fact that for all  $k \in \mathbb{N}$  it holds that  $u_k = \Phi(u_{k-1})$  demonstrate that for all  $n \in \mathbb{N}$  it holds that

$$u_n = u_1 + \sum_{k=1}^{n-1} [u_{k+1} - u_k] = \Phi(0) + \sum_{l=1}^{n-1} [\Phi(u_l) - \Phi(u_{l-1})]. \quad (8)$$

Roughly speaking, the MLP approximations in (3) can then be derived by approximating the expectations and temporal Lebesgue integrals in (8) within the fixed point function  $\Phi$  through appropriate Monte Carlo approximations with different numbers of Monte Carlo samples (cf., e.g., Heinrich [37], Heinrich and Sindambiwe [38], and Giles [27] for related multilevel Monte Carlo approximations).

Theorem 1.1 above in this introductory section is a special case of the more general approximation results in Sect. 5 in this article, and these more general approximations results treat more general PDEs than (2) as well as more general MLP approximation methods than (3). More specifically, in (2) above we have for every  $d \in \mathbb{N}$  that the nonlinearity  $f_d$  depends only on the PDE solution  $u_d$  and the spatial gradient  $\nabla_x u_d$  of the PDE solution but not on  $t \in [0, T]$  and  $x \in \mathbb{R}^d$ , while in Corollary 5.1 and Theorem 5.2 in Sect. 5 the nonlinearities of the PDEs may also depend on  $t \in [0, T]$  and  $x \in \mathbb{R}^d$ . Corollary 5.1 and Theorem 5.2 also provide error analyses for a more general class of MLP approximation methods. In particular, in Theorem 1.1 above the family  $\tau^\theta : \Omega \rightarrow (0, 1)$ ,  $\theta \in (0, 1)$ , of i.i.d. random variables satisfies for all  $b \in (0, 1)$  that  $\mathbb{P}(\tau^0 \leq b) = \sqrt{b}$  and Corollary 5.1 and Theorem 5.2 are proved under the more general hypothesis that there exists  $\alpha \in (0, 1)$  such that for all  $b \in (0, 1)$  it holds that  $\mathbb{P}(\tau^0 \leq b) = b^\alpha$  (see, e.g., (160) in Theorem 5.2). It is a crucial observation of this article that, roughly speaking, in the case of PDEs with gradient-dependent nonlinearities time integrals in the semigroup formulations of the PDEs under consideration cannot be approximated with continuous uniformly distributed random variables (corresponding to the case  $\alpha = 1$ ) as it was done, e.g., in [4, 45, 46]; cf. Sect. 5 for a more detailed discussion of this observation. Furthermore, the more general approximation result in Corollary 5.1 in Sect. 5 also provides an explicit upper bound for the constant  $c \in \mathbb{R}$  in Theorem 1.1 above (see (144) in Corollary 5.1).

We also refer to the references [5] and [20] for numerical simulations for MLP approximation methods for semilinear high-dimensional PDEs. We would like to point out that there are two variants of MLP approximation methods in the scientific literature. First, MLP approximation methods with the temporal Lebesgue integrals being approximated by deterministic quadrature rules have been proposed in [19, 20] in 2016. In [20] this first variant of MLP approximation methods has also been tested numerically, also in the case of gradient-dependent nonlinearities (see [20, Section 3.2]), but the existing convergence results for this first variant of MLP approximation methods are only applicable under very strong regularity assumptions and even exclude many semilinear heat equations with globally Lipschitz continuous gradient-independent nonlinearities (cf. [19, Corollary 3.19] and [47, Corollary 4.8]). Thereafter, MLP approximation methods with the temporal Lebesgue integrals being approximated by Monte Carlo approximations have been proposed in [45] in 2018 and further studied and extended in [4, 28, 46] as well as in the present article. This second variant of MLP approximation methods has been shown to overcome the curse of dimensionality under only mild (local) Lipschitz continuity assumptions on the nonlinearity of the PDEs under consideration (see, e.g., [45, Theorem 1.1], [4, Theorem 1.1], [46, Theorem 1.1], and [28, Corollary 1.2]) and this second variant of MLP approximation methods has also been tested numerically in [5] but only in the case of gradient-independent nonlinearities and it remains a topic of future research to perform numerical simulations for the MLP approximation methods proposed in this article (belonging somehow to the ‘second variant’ of MLP approximation methods) in the case of PDEs with gradient-dependent nonlinearities.

The remainder of this article is organized as follows. In Sect. 2 we establish a few identities and upper bounds for certain iterated deterministic integrals. The results of Sect. 2 are then used in Sect. 3 in which we introduce and analyze the considered

MLP approximation methods. In Sect. 4 we establish suitable a priori bounds for exact solutions of PDEs of the form (2). In Sect. 5 we combine the findings of Sects. 3 and 4 to establish in Theorem 5.2 below the main approximation result of this article.

## 2 Analysis of Certain Deterministic Iterated Integrals

In this section we establish in Corollary 2.5 below an upper bound for products of certain independent random variables. Corollary 2.5 below is a central ingredient in our error analysis for MLP approximations in Sect. 3 below. Our proof of Corollary 2.5 employs the essentially well-known factorization lemma for certain conditional expectations in Lemma 2.4 below as well as a few elementary identities and estimates for certain deterministic iterated integrals which are provided in Lemmas 2.1, 2.2, and Corollary 2.3 below.

### 2.1 Identities for Certain Deterministic Iterated Integrals

**Lemma 2.1** *Let  $T, \beta, \gamma \in (0, \infty)$ , let  $\rho: (0, 1) \rightarrow (0, \infty)$  be  $\mathcal{B}((0, 1))/\mathcal{B}((0, \infty))$ -measurable, and let  $\varrho: [0, T]^2 \rightarrow (0, \infty)$  satisfy for all  $t \in [0, T)$ ,  $s \in (t, T]$  that  $\varrho(t, s) = \frac{1}{T-t}\rho(\frac{s-t}{T-t})$ . Then it holds for all  $j \in \mathbb{N}$ ,  $s_0 \in [0, T)$  that*

$$\begin{aligned} & \int_{s_0}^T \frac{1}{(s_1 - s_0)^\beta [\varrho(s_0, s_1)]^\gamma} \int_{s_1}^T \frac{1}{(s_2 - s_1)^\beta [\varrho(s_1, s_2)]^\gamma} \\ & \quad \cdots \int_{s_{j-1}}^T \frac{1}{(s_j - s_{j-1})^\beta [\varrho(s_{j-1}, s_j)]^\gamma} ds_j \dots ds_2 ds_1 \\ &= (T - s_0)^{j(1+\gamma-\beta)} \left[ \prod_{i=0}^{j-1} \int_0^1 \frac{(1-s)^{i(1+\gamma-\beta)}}{s^\beta [\rho(s)]^\gamma} ds \right]. \end{aligned} \quad (9)$$

**Proof of Lemma 2.1** We prove (9) by induction on  $j \in \mathbb{N}$ . For the base case  $j = 1$  note that integration by substitution yields that for all  $s_0 \in [0, T)$  it holds that

$$\begin{aligned} \int_{s_0}^T \frac{1}{(s_1 - s_0)^\beta [\varrho(s_0, s_1)]^\gamma} ds_1 &= \int_0^1 \frac{(T - s_0)^{1-\beta}}{z[\varrho(s_0, s_0 + z(T - s_0))]^\gamma} dz \\ &= (T - s_0)^{1+\gamma-\beta} \int_0^1 \frac{1}{z^\beta [\rho(z)]^\gamma} dz. \end{aligned} \quad (10)$$

This proves (9) in the base case  $j = 1$ . For the induction step  $\mathbb{N} \ni j \rightsquigarrow j+1 \in \mathbb{N}$  note that the induction hypothesis and integration by substitution imply for all  $s_0 \in [0, T)$

that

$$\begin{aligned}
 & \int_{s_0}^T \frac{1}{(s_1 - s_0)^\beta [\varrho(s_0, s_1)]^\gamma} \left[ \int_{s_1}^T \frac{1}{(s_2 - s_1)^\beta [\varrho(s_1, s_2)]^\gamma} \right. \\
 & \quad \left. \cdots \int_{s_j}^T \frac{1}{(s_{j+1} - s_j)^\beta [\varrho(s_j, s_{j+1})]^\gamma} ds_{j+1} \dots ds_2 \right] ds_1 \\
 &= \int_{s_0}^T \frac{(T - s_1)^{j(1+\gamma-\beta)}}{(s_1 - s_0)^\beta [\varrho(s_0, s_1)]^\gamma} \left[ \prod_{i=0}^{j-1} \int_0^1 \frac{(1-s)^{i(1+\gamma-\beta)}}{s^\beta [\varrho(s)]^\gamma} ds \right] ds_1 \\
 &= \left[ \prod_{i=0}^{j-1} \int_0^1 \frac{(1-s)^{i(1+\gamma-\beta)}}{s^\beta [\varrho(s)]^\gamma} ds \right] \int_0^1 \frac{(T - s_0)^{j(1+\gamma-\beta)} (1-z)^{j(1+\gamma-\beta)}}{(T - s_0)^\beta z^\beta [\varrho(s_0, s_0 + z(T - s_0))]^\gamma} (T - s_0) dz \\
 &= (T - s_0)^{(j+1)(1+\gamma-\beta)} \left[ \prod_{i=0}^j \int_0^1 \frac{(1-s)^{i(1+\gamma-\beta)}}{s^\beta [\varrho(s)]^\gamma} ds \right].
 \end{aligned} \tag{11}$$

Induction thus proves (9). This completes the proof of Lemma 2.1.  $\square$

**Lemma 2.2** Let  $\alpha \in (0, 1)$ ,  $T, \gamma \in (0, \infty)$ ,  $\beta \in (0, \alpha\gamma + 1)$  and let  $\rho: (0, 1) \rightarrow (0, \infty)$  and  $\varrho: [0, T]^2 \rightarrow (0, \infty)$  satisfy for all  $r \in (0, 1)$ ,  $t \in [0, T)$ ,  $s \in (t, T]$  that  $\rho(r) = \frac{1-\alpha}{r^\alpha}$  and  $\varrho(t, s) = \frac{1}{T-t} \rho(\frac{s-t}{T-t})$ . Then it holds for all  $j \in \mathbb{N}$ ,  $s_0 \in [0, T)$  that

$$\begin{aligned}
 & \int_{s_0}^T \frac{1}{(s_1 - s_0)^\beta [\varrho(s_0, s_1)]^\gamma} \int_{s_1}^T \frac{1}{(s_2 - s_1)^\beta [\varrho(s_1, s_2)]^\gamma} \\
 & \quad \cdots \int_{s_{j-1}}^T \frac{1}{(s_j - s_{j-1})^\beta [\varrho(s_{j-1}, s_j)]^\gamma} ds_j \dots ds_2 ds_1 \\
 &= \left[ \frac{(T - s_0)^{(1+\gamma-\beta)} \Gamma(\alpha\gamma - \beta + 1)}{(1 - \alpha)^\gamma} \right]^j \\
 & \quad \cdot \left[ \prod_{i=0}^{j-1} \frac{\Gamma(i(1 + \gamma - \beta) + 1)}{\Gamma(\alpha\gamma - \beta + i(1 + \gamma - \beta) + 2)} \right]. \tag{12}
 \end{aligned}$$

**Proof of Lemma 2.2** Throughout this proof let  $B: (0, \infty) \times (0, \infty) \rightarrow \mathbb{R}$  satisfy for all  $x, y \in (0, \infty)$  that

$$B(x, y) = \int_0^1 s^{x-1} (1-s)^{y-1} ds. \tag{13}$$

Note that (13) and the fact that for all  $x, y \in (0, \infty)$  it holds that  $B(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$  ensure that for all  $i \in \mathbb{N}_0$  it holds that

$$\begin{aligned} \int_0^1 \frac{(1-s)^{i(1+\gamma-\beta)}}{s^\beta [\rho(s)]^\gamma} ds &= \frac{1}{(1-\alpha)^\gamma} \int_0^1 s^{\alpha\gamma-\beta} (1-s)^{i(1+\gamma-\beta)} ds \\ &= \frac{B(\alpha\gamma-\beta+1, i(1+\gamma-\beta)+1)}{(1-\alpha)^\gamma} \\ &= \frac{\Gamma(\alpha\gamma-\beta+1)\Gamma(i(1+\gamma-\beta)+1)}{(1-\alpha)^\gamma \Gamma(\alpha\gamma-\beta+i(1+\gamma-\beta)+2)}. \end{aligned} \quad (14)$$

Lemma 2.1 hence implies that for all  $j \in \mathbb{N}$ ,  $s_0 \in [0, T]$  it holds that

$$\begin{aligned} &\int_{s_0}^T \frac{1}{(s_1 - s_0)^\beta [\varrho(s_0, s_1)]^\gamma} \int_{s_1}^T \frac{1}{(s_2 - s_1)^\beta [\varrho(s_1, s_2)]^\gamma} \\ &\quad \cdots \int_{s_{j-1}}^T \frac{1}{(s_j - s_{j-1})^\beta [\varrho(s_{j-1}, s_j)]^\gamma} ds_j \dots ds_2 ds_1 \\ &= (T - s_0)^{j(1+\gamma-\beta)} \left[ \prod_{i=0}^{j-1} \int_0^1 \frac{(1-s)^{i(1+\gamma-\beta)}}{s^\beta [\rho(s)]^\gamma} ds \right] \\ &= (T - s_0)^{j(1+\gamma-\beta)} \left[ \prod_{i=0}^{j-1} \frac{\Gamma(\alpha\gamma-\beta+1)\Gamma(i(1+\gamma-\beta)+1)}{(1-\alpha)^\gamma \Gamma(\alpha\gamma-\beta+i(1+\gamma-\beta)+2)} \right] \\ &= \left[ \frac{(T - s_0)^{(1+\gamma-\beta)} \Gamma(\alpha\gamma-\beta+1)}{(1-\alpha)^\gamma} \right]^j \left[ \prod_{i=0}^{j-1} \frac{\Gamma(i(1+\gamma-\beta)+1)}{\Gamma(\alpha\gamma-\beta+i(1+\gamma-\beta)+2)} \right]. \end{aligned} \quad (15)$$

This establishes (12). The proof of Lemma 2.2 is thus completed.  $\square$

## 2.2 Estimates for Certain Deterministic Iterated Integrals

**Corollary 2.3** Let  $\alpha \in (0, 1)$ ,  $T, \gamma \in (0, \infty)$ ,  $\beta \in [\alpha\gamma, \alpha\gamma + 1]$  and let  $\rho: (0, 1) \rightarrow (0, \infty)$  and  $\varrho: [0, T]^2 \rightarrow (0, \infty)$  satisfy for all  $r \in (0, 1)$ ,  $t \in [0, T)$ ,  $s \in (t, T]$  that  $\rho(r) = \frac{1-\alpha}{r^\alpha}$  and  $\varrho(t, s) = \frac{1}{T-t} \rho(\frac{s-t}{T-t})$ . Then it holds for all  $j \in \mathbb{N}$ ,  $s_0 \in [0, T)$  that

$$\begin{aligned} &\int_{s_0}^T \frac{1}{(s_1 - s_0)^\beta [\varrho(s_0, s_1)]^\gamma} \int_{s_1}^T \frac{1}{(s_2 - s_1)^\beta [\varrho(s_1, s_2)]^\gamma} \\ &\quad \cdots \int_{s_{j-1}}^T \frac{1}{(s_j - s_{j-1})^\beta [\varrho(s_{j-1}, s_j)]^\gamma} ds_j \dots ds_2 ds_1 \\ &\leq \left[ \frac{(T - s_0)^{(1+\gamma-\beta)} \Gamma(\alpha\gamma-\beta+1)}{(1-\alpha)^\gamma (1+\gamma-\beta)^{\alpha\gamma-\beta+1}} \right]^j \end{aligned}$$

$$\cdot \left[ e^{1+\gamma-\beta} ((1+\gamma-\beta)(j-1)+1) \right]^{\frac{(\beta-\alpha\gamma)(\alpha\gamma-\beta+1)}{1+\gamma-\beta}} \left[ \frac{\Gamma\left(\frac{1}{1+\gamma-\beta}\right)}{\Gamma\left(j+\frac{1}{1+\gamma-\beta}\right)} \right]^{\alpha\gamma-\beta+1}. \quad (16)$$

**Proof of Corollary 2.3** First, observe that Wendel's inequality for the gamma function (see, e.g., Wendel [64] and Qi [59, Section 2.1]) ensures that for all  $x \in (0, \infty)$ ,  $s \in [0, 1]$  it holds that

$$\frac{\Gamma(x)}{\Gamma(x+s)} \leq \frac{1}{x^s} \left[ \frac{x+s}{x} \right]^{1-s}. \quad (17)$$

Moreover, note that the fact that for all  $x \in (0, \infty)$  it holds that  $\ln'(x) = x^{-1}$  demonstrates that for all  $j \in \mathbb{N}$ ,  $\lambda \in (0, \infty)$  it holds that

$$\begin{aligned} \sum_{i=0}^{j-1} \frac{1}{i+\lambda} &= \frac{1}{\lambda} + \sum_{i=1}^{j-1} \frac{1}{i+\lambda} \leq \frac{1}{\lambda} + \sum_{i=1}^{j-1} \int_i^{i+1} \frac{1}{s-1+\lambda} ds \\ &= \frac{1}{\lambda} + \int_1^j \frac{1}{s-1+\lambda} ds = \frac{1}{\lambda} + \ln\left(\frac{j-1}{\lambda} + 1\right). \end{aligned} \quad (18)$$

Combining this, (17), and the fact that  $\alpha\gamma - \beta + 1 \in [0, 1]$  with the fact that for all  $x \in [0, \infty)$  it holds that  $1+x \leq e^x$  proves that for all  $j \in \mathbb{N}$  it holds that

$$\begin{aligned} &\left[ \prod_{i=0}^{j-1} \frac{\Gamma(i(1+\gamma-\beta)+1)}{\Gamma(\alpha\gamma-\beta+i(1+\gamma-\beta)+2)} \right] \\ &\leq \left[ \prod_{i=0}^{j-1} \left( \frac{\alpha\gamma-\beta+i(1+\gamma-\beta)+2}{i(1+\gamma-\beta)+1} \right)^{1-(\alpha\gamma-\beta+1)} \frac{1}{(i(1+\gamma-\beta)+1)^{\alpha\gamma-\beta+1}} \right] \\ &= \left[ \prod_{i=0}^{j-1} \left( 1 + \frac{\alpha\gamma-\beta+1}{i(1+\gamma-\beta)+1} \right)^{\beta-\alpha\gamma} \frac{1}{(i(1+\gamma-\beta)+1)^{\alpha\gamma-\beta+1}} \right] \\ &\leq \left[ \prod_{i=0}^{j-1} e^{\frac{(\beta-\alpha\gamma)(\alpha\gamma-\beta+1)}{i(1+\gamma-\beta)+1}} \frac{1}{(i(1+\gamma-\beta)+1)^{\alpha\gamma-\beta+1}} \right] \\ &= e^{\frac{(\beta-\alpha\gamma)(\alpha\gamma-\beta+1)}{1+\gamma-\beta} \sum_{i=0}^{j-1} \frac{1}{i+\frac{1}{1+\gamma-\beta}}} \left[ \prod_{i=0}^{j-1} \frac{1}{(i(1+\gamma-\beta)+1)^{\alpha\gamma-\beta+1}} \right] \\ &\leq e^{\frac{(\beta-\alpha\gamma)(\alpha\gamma-\beta+1)}{1+\gamma-\beta} (1+\gamma-\beta+\ln((1+\gamma-\beta)(j-1)+1))} \left[ \prod_{i=0}^{j-1} \frac{1}{(i(1+\gamma-\beta)+1)^{\alpha\gamma-\beta+1}} \right] \\ &= \left[ e^{1+\gamma-\beta} ((1+\gamma-\beta)(j-1)+1) \right]^{\frac{(\beta-\alpha\gamma)(\alpha\gamma-\beta+1)}{1+\gamma-\beta}} \left[ \frac{\Gamma\left(\frac{1}{1+\gamma-\beta}\right)}{(1+\gamma-\beta)^j \Gamma\left(j+\frac{1}{1+\gamma-\beta}\right)} \right]^{\alpha\gamma-\beta+1}. \end{aligned} \quad (19)$$

Lemma 2.2 hence implies that for all  $j \in \mathbb{N}$ ,  $s_0 \in [0, T]$  it holds that

$$\begin{aligned} & \int_{s_0}^T \frac{1}{(s_1 - s_0)^\beta [\varrho(s_0, s_1)]^\gamma} \int_{s_1}^T \frac{1}{(s_2 - s_1)^\beta [\varrho(s_1, s_2)]^\gamma} \\ & \quad \cdots \int_{s_{j-1}}^T \frac{1}{(s_j - s_{j-1})^\beta [\varrho(s_{j-1}, s_j)]^\gamma} ds_j \dots ds_2 ds_1 \\ & \leq \left[ \frac{(T-s_0)^{(1+\gamma-\beta)} \Gamma(\alpha\gamma-\beta+1)}{(1-\alpha)^\gamma (1+\gamma-\beta)^{\alpha\gamma-\beta+1}} \right]^j \left[ e^{1+\gamma-\beta} ((1+\gamma-\beta)(j-1)+1) \right]^{\frac{(\beta-\alpha\gamma)(\alpha\gamma-\beta+1)}{1+\gamma-\beta}} \\ & \quad \cdot \left[ \frac{\Gamma\left(\frac{1}{1+\gamma-\beta}\right)}{\Gamma\left(j+\frac{1}{1+\gamma-\beta}\right)} \right]^{\alpha\gamma-\beta+1}. \end{aligned} \quad (20)$$

This establishes (16). The proof of Corollary 2.3 is thus completed.  $\square$

## 2.3 Estimates for Products of Certain Independent Random Variables

**Lemma 2.4** Let  $T \in (0, \infty)$ ,  $d \in \mathbb{N}$ ,  $F \in C((0, 1) \times [0, T) \times \mathbb{R}^d, [0, \infty))$ , let  $(\Omega, \mathcal{F}, \mathbb{P})$  be a probability space, let  $\rho: \Omega \rightarrow (0, 1)$  and  $\tau: \Omega \rightarrow (0, T)$  be random variables, let  $W: [0, T] \times \Omega \rightarrow \mathbb{R}^d$  be a standard Brownian motion with continuous sample paths, let  $f: [0, T) \rightarrow [0, \infty]$  satisfy for all  $t \in [0, T)$  that  $f(t) = \mathbb{E}[F(\rho, t, W_{t+(T-t)\rho} - W_t)]$ , let  $\mathcal{G} \subseteq \mathcal{F}$  be a sigma-algebra, let  $\mathcal{H} = \sigma(\mathcal{G} \cup \sigma(\tau, (W_{\min\{s, \tau\}})_{s \in [0, T]}))$ , and assume that  $\rho, \tau, W$ , and  $\mathcal{G}$  are independent. Then it holds  $\mathbb{P}$ -a.s. that

$$\mathbb{E}[F(\rho, \tau, W_{\tau+(T-\tau)\rho} - W_\tau) | \mathcal{H}] = f(\tau). \quad (21)$$

**Proof of Lemma 2.4** First, note that independence of  $\rho, \tau, W$ , and  $\mathcal{G}$  ensures that it holds  $\mathbb{P}$ -a.s. that

$$\mathbb{E}[F(\rho, \tau, W_{\tau+(T-\tau)\rho} - W_\tau) | \mathcal{H}] = \mathbb{E}[F(\rho, \tau, W_{\tau+(T-\tau)\rho} - W_\tau) | \sigma(\tau, (W_{\min\{s, \tau\}})_{s \in [0, T]})]. \quad (22)$$

Next observe that independence of  $\rho, \tau$ , and  $W$  and the tower property of conditional expectations (cf., e.g., Hutzenthaler et al. [45, Lemma 2.2] (applied with  $\mathcal{G} = \sigma(\rho, (W_s)_{s \in [0, T]})$ ,  $S = (0, T)$ ,  $\mathcal{S} = \mathcal{B}((0, T))$ ,  $U(t, \omega) = \mathbf{1}_A(t)\mathbf{1}_B((W_{\min\{s, t\}}(\omega))_{s \in [0, T]})F(\rho(\omega), t, W_{t+(T-t)\rho(\omega)}(\omega) - W_t(\omega))$ ,  $Y(\omega) = \tau(\omega)$  for  $t \in (0, T), \omega \in \Omega$  in the notation of [45, Lemma 2.2])) prove that for all  $A \in \mathcal{B}((0, T))$ ,  $B \in \mathcal{B}([0, T], \mathbb{R}^d)$  it holds that

$$\begin{aligned} & \mathbb{E}[\mathbf{1}_A(\tau)\mathbf{1}_B((W_{\min\{s, \tau\}})_{s \in [0, T]})F(\rho, \tau, W_{\tau+(T-\tau)\rho} - W_\tau)] \\ & = \int_{(0, T)} \mathbb{E}[\mathbf{1}_A(t)\mathbf{1}_B((W_{\min\{s, t\}})_{s \in [0, T]})F(\rho, t, W_{t+(T-t)\rho} - W_t)](\tau(\mathbb{P}))(dt). \end{aligned} \quad (23)$$

The fact that for all  $t \in [0, T], r \in [t, T]$  it holds that  $(W_{\min\{s, t\}})_{s \in [0, T]}$  and  $W_r - W_t$  are independent hence proves that for all  $A \in \mathcal{B}((0, T))$ ,  $B \in \mathcal{B}([0, T], \mathbb{R}^d)$  it

holds that

$$\begin{aligned} & \mathbb{E}[\mathbf{1}_A(\tau)\mathbf{1}_B((W_{\min\{s,\tau\}})_{s \in [0,T]})F(\rho, \tau, W_{\tau+(T-\tau)\rho} - W_\tau)] \\ &= \int_{(0,T)} \mathbb{E}[\mathbf{1}_A(t)\mathbf{1}_B((W_{\min\{s,t\}})_{s \in [0,T]})]\mathbb{E}[F(\rho, t, W_{t+(T-t)\rho} - W_t)](\tau(\mathbb{P}))(dt) \\ &= \int_{(0,T)} \mathbb{E}[\mathbf{1}_A(t)\mathbf{1}_B((W_{\min\{s,t\}})_{s \in [0,T]})]f(t)(\tau(\mathbb{P}))(dt). \end{aligned} \quad (24)$$

This and, e.g., Hutzenthaler et al. [45, Lemma 2.2] (applied with  $\mathcal{G} = \sigma(\rho, (W_s)_{s \in [0,T]})$ ,  $S = (0, T)$ ,  $\mathcal{S} = \mathcal{B}((0, T))$ ,  $U(t, \omega) = \mathbf{1}_A(t)\mathbf{1}_B((W_{\min\{s,t\}}(\omega))_{s \in [0,T]})g(t)$ ,  $Y(\omega) = \tau(\omega)$  for  $t \in (0, T)$ ,  $\omega \in \Omega$  in the notation of [45, Lemma 2.2]) proves that for all  $A \in \mathcal{B}((0, T))$ ,  $B \in \mathcal{B}(C([0, T], \mathbb{R}^d))$  it holds that

$$\begin{aligned} & \mathbb{E}[\mathbf{1}_A(\tau)\mathbf{1}_B((W_{\min\{s,\tau\}})_{s \in [0,T]})F(\rho, \tau, W_{\tau+(T-\tau)\rho} - W_\tau)] \\ &= \int_{(0,T)} \mathbb{E}[\mathbf{1}_A(t)\mathbf{1}_B((W_{\min\{s,t\}})_{s \in [0,T]})f(t)](\tau(\mathbb{P}))(dt) \\ &= \mathbb{E}[\mathbf{1}_A(\tau)\mathbf{1}_B((W_{\min\{s,\tau\}})_{s \in [0,T]})f(\tau)]. \end{aligned} \quad (25)$$

Combining this with (22) establishes (21). The proof of Lemma 2.4 is thus completed.  $\square$

**Corollary 2.5** Let  $T \in (0, \infty)$ ,  $d \in \mathbb{N}$ ,  $j \in \mathbb{N}_0$ ,  $\mathbf{e}_1 = (1, 0, 0, \dots, 0)$ ,  $\mathbf{e}_2 = (0, 1, 0, \dots, 0)$ , ...,  $\mathbf{e}_{d+1} = (0, 0, \dots, 0, 1) \in \mathbb{R}^{d+1}$ ,  $v_0, v_1, \dots, v_j \in \{1, 2, \dots, d+1\}$ ,  $\alpha \in (0, 1)$ ,  $p \in (1, \infty)$  satisfy  $\alpha(p-1) \leq \frac{p}{2} \leq \alpha(p-1)+1$ , let  $\langle \cdot, \cdot \rangle: \mathbb{R}^{d+1} \times \mathbb{R}^{d+1} \rightarrow \mathbb{R}$  be the standard scalar product on  $\mathbb{R}^{d+1}$ , let  $(\Omega, \mathcal{F}, \mathbb{P})$  be a probability space, let  $W = (W^1, W^2, \dots, W^d): [0, T] \times \Omega \rightarrow \mathbb{R}^d$  be a standard Brownian motion with continuous sample paths, let  $\mathbf{r}^{(n)}: \Omega \rightarrow (0, 1)$ ,  $n \in \mathbb{N}_0$ , be i.i.d. random variables, assume that  $W$  and  $(\mathbf{r}^{(n)})_{n \in \mathbb{N}_0}$  are independent, let  $\rho: (0, 1) \rightarrow (0, \infty)$  and  $\varrho: [0, T]^2 \rightarrow (0, \infty)$  satisfy for all  $b \in (0, 1)$ ,  $t \in [0, T]$ ,  $s \in (t, T]$  that  $\rho(b) = \frac{1-\alpha}{b^\alpha}$ ,  $\mathbb{P}(\mathbf{r}^{(0)} \leq b) = \int_0^b \rho(u) du$ , and  $\varrho(t, s) = \frac{1}{T-t} \rho(\frac{s-t}{T-t})$ , let  $S: \mathbb{N}_0 \times [0, T] \times \Omega \rightarrow [0, T]$  satisfy for all  $n \in \mathbb{N}_0$ ,  $t \in [0, T)$  that  $S(0, t) = t$  and  $S(n+1, t) = S(n, t) + (T - S(n, t))\mathbf{r}^{(n)}$ , and let  $t \in [0, T)$ . Then

$$\begin{aligned} & \mathbb{E}\left[\left|\prod_{i=0}^j \frac{1}{\varrho(S(i,t), S(i+1,t))} \langle \mathbf{e}_{v_i}, \left(1, \frac{W_{S(i+1,t)} - W_{S(i,t)}}{S(i+1,t) - S(i,t)}\right) \rangle\right|^p\right] \\ & \leq \left[ \max\left\{(T-t)^{\frac{p}{2}}, \frac{2^{\frac{p}{2}} \Gamma(\frac{p+1}{2})}{\sqrt{\pi}}\right\} \frac{(T-t)^{\frac{p}{2}} \Gamma(\frac{p}{2})}{(1-\alpha)^{p-1} (\frac{p}{2})^{\alpha(p-1)-\frac{p}{2}+1}} \right]^{j+1} \left[ e^{\frac{p}{2}} \left(\frac{pj}{2} + 1\right) \right]^{\frac{1}{2p}} \\ & \quad \cdot \left[ \frac{\Gamma(\frac{2}{p})}{\Gamma(1+j+\frac{2}{p})} \right]^{\alpha(p-1)-\frac{p}{2}+1}. \end{aligned} \quad (26)$$

**Proof of Corollary 2.5** Throughout this proof let  $\mathbb{F}_n \subseteq \mathcal{F}$ ,  $n \in \mathbb{N}_0$ , satisfy for all  $n \in \mathbb{N}$  that  $\mathbb{F}_0 = \{\emptyset, \Omega\}$  and that  $\mathbb{F}_n = \sigma(\mathbb{F}_{n-1} \cup \sigma(S(n, t), (W_{\min\{s, S(n, t)\}})_{s \in [0, T]}))$  and let  $v = (v_1, v_2, \dots, v_d) \in \mathbb{R}^d$  satisfy  $v_1 = v_2 = \dots = v_d = 1$ . Note that for all

$r \in [0, T]$ ,  $s \in [r, T]$ ,  $i \in \{1, 2, \dots, d\}$ ,  $n \in \mathbb{N}$  it holds that  $S(n, r) > S(n - 1, r)$  and

$$\mathbb{E}[|W_s^i - W_r^i|^p] = \pi^{-\frac{1}{2}} |2(s - r)|^{\frac{p}{2}} \Gamma(\frac{p+1}{2}). \quad (27)$$

Next we claim that for all  $k \in \{1, 2, \dots, j + 1\}$  it holds  $\mathbb{P}$ -a.s. that

$$\begin{aligned} & \mathbb{E}\left[\prod_{i=k}^{j+1} \left|\frac{1}{\varrho(S(i-1,t),S(i,t))} \langle \mathbf{e}_{v_{i-1}}, \left(1, \frac{W_{S(i,t)} - W_{S(i-1,t)}}{S(i,t) - S(i-1,t)}\right) \rangle\right|^p \middle| \mathbb{F}_{k-1}\right] \\ &= \int_{S(k-1,t)}^T \frac{\langle \mathbf{e}_{v_{k-1}}, \left(1, [\frac{2}{s_k - S(k-1,t)}]^{p/2} \pi^{-1/2} \Gamma(\frac{p+1}{2}) v\right) \rangle}{[\varrho(S(k-1,t), s_k)]^{p-1}} \int_{s_k}^T \frac{\langle \mathbf{e}_{v_k}, \left(1, [\frac{2}{s_{k+1} - s_k}]^{p/2} \pi^{-1/2} \Gamma(\frac{p+1}{2}) v\right) \rangle}{[\varrho(s_k, s_{k+1})]^{p-1}} \\ &\quad \cdots \int_{s_j}^T \frac{\langle \mathbf{e}_{v_j}, \left(1, [\frac{2}{s_{j+1} - s_j}]^{p/2} \pi^{-1/2} \Gamma(\frac{p+1}{2}) v\right) \rangle}{[\varrho(s_j, s_{j+1})]^{p-1}} ds_{j+1} \dots ds_{k+1} ds_k. \end{aligned} \quad (28)$$

We prove (28) by backward induction on  $k \in \{1, 2, \dots, j + 1\}$ . For the base case  $k = j + 1$  note that the fact that  $S(j + 1, t) = S(j, t) + (T - S(j, t))\tau^{(j)}$ , Lemma 2.4 (applied with  $F(r, s, x) = |\frac{1}{\varrho(s, s + (T-s)r)} \langle \mathbf{e}_v, \left(1, \frac{x}{(T-s)r}\right) \rangle|^p$ ,  $\rho = \tau^{(j)}$ ,  $\tau = S(j, t)$ ,  $\mathcal{G} = \mathbb{F}_{j-1}$  for  $r \in (0, 1)$ ,  $s \in [0, T)$ ,  $x \in \mathbb{R}^d$  in the notation of Lemma 2.4), e.g., Hutzenthaler et al. [45, Lemma 2.3], the hypothesis that  $W$  and  $\tau^{(j)}$  are independent, (27), and the fact that for all  $r \in [0, T)$ ,  $s \in (r, T]$  it holds that  $\varrho(r, s) = \frac{1}{T-r} \rho(\frac{s-r}{T-r})$  ensure that it holds  $\mathbb{P}$ -a.s. that

$$\begin{aligned} & \mathbb{E}\left[\left|\frac{1}{\varrho(S(j,t),S(j+1,t))} \langle \mathbf{e}_{v_j}, \left(1, \frac{W_{S(j+1,t)} - W_{S(j,t)}}{S(j+1,t) - S(j,t)}\right) \rangle\right|^p \middle| \mathbb{F}_j\right] \\ &= \mathbb{E}\left[\left|\frac{1}{\varrho(S(j,t),S(j,t)+(T-S(j,t))\tau^{(j)})} \langle \mathbf{e}_{v_j}, \left(1, \frac{W_{S(j,t)+(T-S(j,t))\tau^{(j)}} - W_{S(j,t)}}{(T-S(j,t))\tau^{(j)}}\right) \rangle\right|^p \middle| \mathbb{F}_j\right] \\ &= \mathbb{E}\left[\left|\frac{1}{\varrho(s,s+(T-s)\tau^{(j)})} \langle \mathbf{e}_{v_j}, \left(1, \frac{W_{s+(T-s)\tau^{(j)}} - W_s}{(T-s)\tau^{(j)}}\right) \rangle\right|^p\right]_{s=S(j,t)} \\ &= \int_0^1 \frac{\mathbb{E}\left[|\langle \mathbf{e}_{v_j}, \left(1, \frac{W_{s+(T-s)r} - W_s}{(T-s)r}\right) \rangle|^p\right] \rho(r)}{[\varrho(s, s + (T-s)r)]^p} dr \Big|_{s=S(j,t)} \\ &= \int_s^T \frac{\mathbb{E}\left[|\langle \mathbf{e}_{v_j}, \left(1, \frac{W_{s_{j+1}} - W_s}{s_{j+1} - s}\right) \rangle|^p\right] \rho\left(\frac{s_{j+1} - s}{T-s}\right)}{(T-s)[\varrho(s, s_{j+1})]^p} ds_{j+1} \Big|_{s=S(j,t)} \\ &= \int_{S(j,t)}^T \frac{\langle \mathbf{e}_{v_j}, \left(1, [\frac{2}{s_{j+1} - S(j,t)}]^{p/2} \pi^{-1/2} \Gamma(\frac{p+1}{2}) v\right) \rangle}{[\varrho(S(j,t), s_{j+1})]^{p-1}} ds_{j+1}. \end{aligned} \quad (29)$$

This establishes (28) in the base case  $k = j + 1$ . For the induction step  $\{2, 3, \dots, j+1\} \ni k + 1 \rightsquigarrow k \in \{1, 2, \dots, j\}$  assume that there exists  $k \in \{1, 2, \dots, j\}$  which satisfies

that

$$\begin{aligned}
& \mathbb{E} \left[ \prod_{i=k+1}^{j+1} \left| \frac{1}{\varrho(S(i-1,t), S(i,t))} \langle \mathbf{e}_{v_{i-1}}, \left( 1, \frac{W_{S(i,t)} - W_{S(i-1,t)}}{S(i,t) - S(i-1,t)} \right) \rangle \right|^p \middle| \mathbb{F}_k \right] \\
&= \int_{S(k,t)}^T \frac{\langle \mathbf{e}_{v_k}, \left( 1, [\frac{2}{s_{k+1}-S(k,t)}]^{p/2} \pi^{-1/2} \Gamma(\frac{p+1}{2}) v \right) \rangle}{[\varrho(S(k,t), s_{k+1})]^{p-1}} \\
&\quad \cdot \int_{s_{k+1}}^T \frac{\langle \mathbf{e}_{v_{k+1}}, \left( 1, [\frac{2}{s_{k+2}-s_{k+1}}]^{p/2} \pi^{-1/2} \Gamma(\frac{p+1}{2}) v \right) \rangle}{[\varrho(s_{k+1}, s_{k+2})]^{p-1}} \\
&\quad \cdots \int_{s_j}^T \frac{\langle \mathbf{e}_{v_j}, \left( 1, [\frac{2}{s_{j+1}-s_j}]^{p/2} \pi^{-1/2} \Gamma(\frac{p+1}{2}) v \right) \rangle}{[\varrho(s_j, s_{j+1})]^{p-1}} ds_{j+1} \dots ds_{k+2} ds_{k+1}. \tag{30}
\end{aligned}$$

Observe that the tower property, the fact that the random variable  $\frac{1}{\varrho(S(k-1,t), S(k,t))} \left( 1, \frac{W_{S(k,t)} - W_{S(k-1,t)}}{S(k,t) - S(k-1,t)} \right)$  is  $\mathbb{F}_k / \mathcal{B}(\mathbb{R})$ -measurable, and (30) ensure that it holds  $\mathbb{P}$ -a.s. that

$$\begin{aligned}
& \mathbb{E} \left[ \prod_{i=k}^{j+1} \left| \frac{1}{\varrho(S(i-1,t), S(i,t))} \langle \mathbf{e}_{v_{i-1}}, \left( 1, \frac{W_{S(i,t)} - W_{S(i-1,t)}}{S(i,t) - S(i-1,t)} \right) \rangle \right|^p \middle| \mathbb{F}_{k-1} \right] \\
&= \mathbb{E} \left[ \left| \frac{1}{\varrho(S(k-1,t), S(k,t))} \langle \mathbf{e}_{v_{k-1}}, \left( 1, \frac{W_{S(k,t)} - W_{S(k-1,t)}}{S(k,t) - S(k-1,t)} \right) \rangle \right|^p \right. \\
&\quad \cdot \mathbb{E} \left. \left[ \prod_{i=k+1}^{j+1} \left| \frac{1}{\varrho(S(i-1,t), S(i,t))} \langle \mathbf{e}_{v_{i-1}}, \left( 1, \frac{W_{S(i,t)} - W_{S(i-1,t)}}{S(i,t) - S(i-1,t)} \right) \rangle \right|^p \middle| \mathbb{F}_k \right] \middle| \mathbb{F}_{k-1} \right] \tag{31} \\
&= \mathbb{E} \left[ \left| \frac{1}{\varrho(S(k-1,t), S(k,t))} \langle \mathbf{e}_{v_{k-1}}, \left( 1, \frac{W_{S(k,t)} - W_{S(k-1,t)}}{S(k,t) - S(k-1,t)} \right) \rangle \right|^p \right. \\
&\quad \cdot \int_{S(k,t)}^T \frac{\langle \mathbf{e}_{v_k}, \left( 1, [\frac{2}{s_{k+1}-S(k,t)}]^{p/2} \pi^{-1/2} \Gamma(\frac{p+1}{2}) v \right) \rangle}{[\varrho(S(k,t), s_{k+1})]^{p-1}} \cdot \int_{s_{k+1}}^T \frac{\langle \mathbf{e}_{v_{k+1}}, \left( 1, [\frac{2}{s_{k+2}-s_{k+1}}]^{p/2} \pi^{-1/2} \Gamma(\frac{p+1}{2}) v \right) \rangle}{[\varrho(s_{k+1}, s_{k+2})]^{p-1}} \\
&\quad \cdots \int_{s_j}^T \frac{\langle \mathbf{e}_{v_j}, \left( 1, [\frac{2}{s_{j+1}-s_j}]^{p/2} \pi^{-1/2} \Gamma(\frac{p+1}{2}) v \right) \rangle}{[\varrho(s_j, s_{j+1})]^{p-1}} ds_{j+1} \dots ds_{k+2} ds_{k+1} \left. \middle| \mathbb{F}_{k-1} \right].
\end{aligned}$$

The fact that  $S(k, t) = S(k-1, t) + (T - S(k-1, t))\mathbf{r}^{(k-1)}$  and Lemma 2.4 (applied with  $\rho = \mathbf{r}^{(k-1)}$ ,  $\tau = S(k-1, t)$ ,  $\mathcal{G} = \mathbb{F}_{k-2}$  in the notation of Lemma 2.4) hence

prove that it holds  $\mathbb{P}$ -a.s. that

$$\begin{aligned} & \mathbb{E} \left[ \prod_{i=k}^{j+1} \left| \frac{1}{\varrho(S(i-1,t), S(i,t))} \langle \mathbf{e}_{v_{i-1}}, (1, \frac{W_{S(i,t)} - W_{S(i-1,t)}}{S(i,t) - S(i-1,t)}) \rangle \right|^p \middle| \mathbb{F}_{k-1} \right] \\ &= \mathbb{E} \left[ \left| \frac{\langle \mathbf{e}_{v_{k-1}}, (1, [\frac{W_{s+(T-s)\tau^{(k-1)} - W_s}{(T-s)\tau^{(k-1)}}]) \rangle}{\varrho(s, s + (T-s)\tau^{(k-1)})} \right|^p \right. \\ & \quad \cdot \int_{s+(T-s)\tau^{(k-1)}}^T \frac{\langle \mathbf{e}_{v_k}, (1, [\frac{2}{s_{k+1} - (s+(T-s)\tau^{(k-1)})}]^{p/2} \pi^{-1/2} \Gamma(\frac{p+1}{2}) v) \rangle}{[\varrho(s+(T-s)\tau^{(k-1)}, s_{k+1})]^{p-1}} \\ & \quad \cdot \int_{s_{k+1}}^T \frac{\langle \mathbf{e}_{v_{k+1}}, (1, [\frac{2}{s_{k+2} - s_{k+1}}]^{p/2} \pi^{-1/2} \Gamma(\frac{p+1}{2}) v) \rangle}{[\varrho(s_{k+1}, s_{k+2})]^{p-1}} \dots \\ & \quad \left. \cdot \int_{s_j}^T \frac{\langle \mathbf{e}_{v_j}, (1, [\frac{2}{s_{j+1} - s_j}]^{p/2} \pi^{-1/2} \Gamma(\frac{p+1}{2}) v) \rangle}{[\varrho(s_j, s_{j+1})]^{p-1}} ds_{j+1} \dots ds_{k+2} ds_{k+1} \right] \Big|_{s=S(k-1,t)}. \end{aligned} \quad (32)$$

This, e.g., Hutzenthaler et al. [45, Lemma 2.3], the hypothesis that  $W$  and  $\tau^{(k-1)}$  are independent, (27), and the fact that for all  $r \in [0, T]$ ,  $s \in (r, T]$  it holds that  $\varrho(r, s) = \frac{1}{T-r} \rho(\frac{s-r}{T-r})$  assure that it holds  $\mathbb{P}$ -a.s. that

$$\begin{aligned} & \mathbb{E} \left[ \prod_{i=k}^{j+1} \left| \frac{1}{\varrho(S(i-1,t), S(i,t))} \langle \mathbf{e}_{v_{i-1}}, (1, \frac{W_{S(i,t)} - W_{S(i-1,t)}}{S(i,t) - S(i-1,t)}) \rangle \right|^p \middle| \mathbb{F}_{k-1} \right] \\ &= \int_0^1 \frac{\langle \mathbf{e}_{v_{k-1}}, (1, [\frac{(2((T-s)r))^{p/2} \Gamma(\frac{p+1}{2})}{\sqrt{\pi}((T-s)r)]^p} v) \rangle \rho(r)}{[\varrho(s, s + (T-s)r)]^p} \int_{s+(T-s)r}^T \frac{\langle \mathbf{e}_{v_k}, (1, [\frac{2}{s_{k+1} - (s+(T-s)r)}]^{p/2} \pi^{-1/2} \Gamma(\frac{p+1}{2}) v) \rangle}{[\varrho(s+(T-s)r, s_{k+1})]^{p-1}} \\ & \quad \cdot \int_{s_{k+1}}^T \frac{\langle \mathbf{e}_{v_{k+1}}, (1, [\frac{2}{s_{k+2} - s_{k+1}}]^{p/2} \pi^{-1/2} \Gamma(\frac{p+1}{2}) v) \rangle}{[\varrho(s_{k+1}, s_{k+2})]^{p-1}} \dots \\ & \quad \cdot \int_{s_j}^T \frac{\langle \mathbf{e}_{v_j}, (1, [\frac{2}{s_{j+1} - s_j}]^{p/2} \pi^{-1/2} \Gamma(\frac{p+1}{2}) v) \rangle}{[\varrho(s_j, s_{j+1})]^{p-1}} ds_{j+1} \dots ds_{k+2} ds_{k+1} dr \Big|_{s=S(k-1,t)} \\ &= \int_{S(k-1,t)}^T \frac{\langle \mathbf{e}_{v_{k-1}}, (1, [\frac{2}{s_{k+1} - s_{k-1,t}}]^{p/2} \pi^{-1/2} \Gamma(\frac{p+1}{2}) v) \rangle}{[\varrho(s(k-1,t), s_k)]^{p-1}} \int_{s_k}^T \frac{\langle \mathbf{e}_{v_k}, (1, [\frac{2}{s_{k+1} - s_k}]^{p/2} \pi^{-1/2} \Gamma(\frac{p+1}{2}) v) \rangle}{[\varrho(s_k, s_{k+1})]^{p-1}} \\ & \quad \cdot \int_{s_{k+1}}^T \frac{\langle \mathbf{e}_{v_{k+1}}, (1, [\frac{2}{s_{k+2} - s_{k+1}}]^{p/2} \pi^{-1/2} \Gamma(\frac{p+1}{2}) v) \rangle}{[\varrho(s_{k+1}, s_{k+2})]^{p-1}} \dots \int_{s_j}^T \frac{\langle \mathbf{e}_{v_j}, (1, [\frac{2}{s_{j+1} - s_j}]^{p/2} \pi^{-1/2} \Gamma(\frac{p+1}{2}) v) \rangle}{[\varrho(s_j, s_{j+1})]^{p-1}} ds_{j+1} \\ & \quad \dots ds_{k+2} ds_{k+1} ds_k. \end{aligned} \quad (33)$$

Induction thus proves (28). Next observe that (28) implies that

$$\begin{aligned} & \mathbb{E} \left[ \prod_{i=1}^{j+1} \left| \frac{1}{\varrho(S(i-1,t), S(i,t))} \langle \mathbf{e}_{v_{i-1}}, (1, \frac{W_{S(i,t)} - W_{S(i-1,t)}}{S(i,t) - S(i-1,t)}) \rangle \right|^p \right] \\ & \leq \left[ \max \left\{ (T-t)^{\frac{p}{2}}, \frac{2^{\frac{p}{2}} \Gamma(\frac{p+1}{2})}{\sqrt{\pi}} \right\} \right]^{j+1} \end{aligned}$$

$$\begin{aligned} & \cdot \int_t^T \frac{1}{(s_1 - t)^{\frac{p}{2}} [\varrho(t, s_1)]^{p-1}} \int_{s_1}^T \frac{1}{(s_2 - s_1)^{\frac{p}{2}} [\varrho(s_1, s_2)]^{p-1}} \\ & \cdots \int_{s_j}^T \frac{1}{(s_{j+1} - s_j)^{\frac{p}{2}} [\varrho(s_j, s_{j+1})]^{p-1}} ds_{j+1} \dots ds_2 ds_1. \end{aligned} \quad (34)$$

Corollary 2.3 (applied with  $\beta = \frac{p}{2}$  and  $\gamma = p - 1$  in the notation of Corollary 2.3), the fact that  $\frac{p}{2} \in [\alpha(p-1), \alpha(p-1)+1]$ , and the fact that  $(\frac{p}{2} - \alpha(p-1))(1 - (\frac{p}{2} - \alpha(p-1))) \leq \frac{1}{4}$  therefore show that

$$\begin{aligned} & \left\| \prod_{i=1}^{j+1} \frac{1}{\varrho(S(i-1,t), S(i,t))} \langle \mathbf{e}_{v_{i-1}}, \left(1, \frac{W_{S(i,t)} - W_{S(i-1,t)}}{S(i,t) - S(i-1,t)}\right) \rangle \right\|_{L^p(\mathbb{P}; \mathbb{R})}^p \\ &= \mathbb{E} \left[ \prod_{i=1}^{j+1} \left| \frac{1}{\varrho(S(i-1,t), S(i,t))} \langle \mathbf{e}_{v_{i-1}}, \left(1, \frac{W_{S(i,t)} - W_{S(i-1,t)}}{S(i,t) - S(i-1,t)}\right) \rangle \right|^p \right] \\ &\leq \left[ \max \left\{ (T-t)^{\frac{p}{2}}, \frac{2^{\frac{p}{2}} \Gamma(\frac{p+1}{2})}{\sqrt{\pi}} \right\} \right]^{j+1} \left[ \frac{(T-t)^{\frac{p}{2}} \Gamma(\alpha(p-1) - \frac{p}{2} + 1)}{(1-\alpha)^{p-1} (\frac{p}{2})^{\alpha(p-1) - \frac{p}{2} + 1}} \right]^{j+1} \\ &\quad \cdot \left[ e^{\frac{p}{2}} \left( \frac{pj}{2} + 1 \right) \right]^{\frac{2(p - \alpha(p-1)(\alpha(p-1) - \frac{p}{2} + 1))}{p}} \cdot \left[ \frac{\Gamma(\frac{2}{p})}{\Gamma(1+j+\frac{2}{p})} \right]^{\alpha(p-1) - \frac{p}{2} + 1} \quad (35) \\ &\leq \left[ \max \left\{ (T-t)^{\frac{p}{2}}, \frac{2^{\frac{p}{2}} \Gamma(\frac{p+1}{2})}{\sqrt{\pi}} \right\} \frac{(T-t)^{\frac{p}{2}} \Gamma(\alpha(p-1) - \frac{p}{2} + 1)}{(1-\alpha)^{p-1} (\frac{p}{2})^{\alpha(p-1) - \frac{p}{2} + 1}} \right]^{j+1} \\ &\quad \cdot \left[ e^{\frac{p}{2}} \left( \frac{pj}{2} + 1 \right) \right]^{\frac{1}{2p}} \left[ \frac{\Gamma(\frac{2}{p})}{\Gamma(1+j+\frac{2}{p})} \right]^{\alpha(p-1) - \frac{p}{2} + 1}. \end{aligned}$$

Combining this with the fact that  $\alpha(p-1) - \frac{p}{2} + 1 \leq p - 1 - \frac{p}{2} + 1 = \frac{p}{2}$  establishes (26). The proof of Corollary 2.5 is thus completed.  $\square$

### 3 Full-History Recursive Multilevel Picard Approximation Methods

In this section we introduce and analyze a class of new MLP approximation methods for nonlinear heat equations with gradient-dependent nonlinearities. In the main result of this section, Proposition 3.5 in Sect. 3.3 below, we provide a detailed error analysis for these new MLP approximation methods. We will employ Proposition 3.5 in our proofs of the approximation results in Sect. 5 below (cf. Corollary 5.1, Theorem 5.2, Corollary 5.4, and Corollary 5.5 in Sect. 5 below). Our proof of Proposition 3.5 employs suitable recursive error bounds for the proposed MLP approximations, which we establish in Lemma 3.4 below. Our proof of Lemma 3.4, in turn, uses appropriate measurability and distribution properties of the proposed MLP approximations, which we establish in the elementary auxiliary result in Lemma 3.2 below, and appropriate

explicit representations for expectations of the proposed MLP approximations, which we establish in the auxiliary result in Lemma 3.3 below. In Setting 3.1 we formulate the mathematical framework which we employ to analyze the proposed MLP approximations. In particular, in (39) in Setting 3.1 we introduce the proposed MLP approximations.

### 3.1 Description of MLP Approximations

**Setting 3.1** Let  $\|\cdot\|_1 : (\cup_{n \in \mathbb{N}} \mathbb{R}^n) \rightarrow \mathbb{R}$  satisfy for all  $n \in \mathbb{N}$ ,  $x = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$  that  $\|x\|_1 = \sum_{i=1}^n |x_i|$ , let  $T \in (0, \infty)$ ,  $d \in \mathbb{N}$ ,  $\Theta = \cup_{n \in \mathbb{N}} \mathbb{Z}^n$ ,  $L = (L_1, L_2, \dots, L_{d+1}) \in [0, \infty)^{d+1}$ ,  $K = (K_1, K_2, \dots, K_d) \in [0, \infty]^d$ ,  $\mathbf{e}_1 = (1, 0, \dots, 0)$ ,  $\mathbf{e}_2 = (0, 1, 0, \dots, 0)$ , ...,  $\mathbf{e}_{d+1} = (0, 0, \dots, 1) \in \mathbb{R}^{d+1}$ ,  $\rho \in C((0, 1), (0, \infty))$ ,  $\mathbf{u} = (\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_{d+1}) \in C([0, T) \times \mathbb{R}^d, \mathbb{R}^{1+d})$ , let  $\langle \cdot, \cdot \rangle : \mathbb{R}^{d+1} \times \mathbb{R}^{d+1} \rightarrow \mathbb{R}$  satisfy for all  $v = (v_1, v_2, \dots, v_{d+1})$ ,  $w = (w_1, w_2, \dots, w_{d+1}) \in \mathbb{R}^{d+1}$  that  $\langle v, w \rangle = \sum_{i=1}^{d+1} v_i w_i$ , let  $\varrho : [0, T]^2 \rightarrow \mathbb{R}$  satisfy for all  $t \in [0, T)$ ,  $s \in (t, T)$  that  $\varrho(t, s) = \frac{1}{T-t} \rho(\frac{s-t}{T-t})$ , let  $(\Omega, \mathcal{F}, \mathbb{P})$  be a probability space, let  $W^\theta = (W^{\theta,1}, W^{\theta,2}, \dots, W^{\theta,d}) : [0, T] \times \Omega \rightarrow \mathbb{R}^d$ ,  $\theta \in \Theta$ , be i.i.d. standard Brownian motions with continuous sample paths, let  $\mathfrak{r}^\theta : \Omega \rightarrow (0, 1)$ ,  $\theta \in \Theta$ , be i.i.d. random variables, assume for all  $b \in (0, 1)$  that  $\mathbb{P}(\mathfrak{r}^0 \leq b) = \int_0^b \rho(s) ds$ , assume that  $(W^\theta)_{\theta \in \Theta}$  and  $(\mathfrak{r}^\theta)_{\theta \in \Theta}$  are independent, let  $\mathcal{R}^\theta : [0, T) \times \Omega \rightarrow [0, T)$ ,  $n \in \mathbb{N}_0$ , satisfy for all  $\theta \in \Theta$ ,  $t \in [0, T)$  that  $\mathcal{R}_t^\theta = t + (T-t)\mathfrak{r}^\theta$ , let  $f \in C([0, T] \times \mathbb{R}^d \times \mathbb{R}^{1+d}, \mathbb{R})$ ,  $g \in C(\mathbb{R}^d, \mathbb{R})$ ,  $F : C([0, T) \times \mathbb{R}^d, \mathbb{R}^{1+d}) \rightarrow C([0, T) \times \mathbb{R}^d, \mathbb{R})$  satisfy for all  $t \in [0, T)$ ,  $x = (x_1, x_2, \dots, x_d)$ ,  $\mathfrak{x} = (\mathfrak{x}_1, \mathfrak{x}_2, \dots, \mathfrak{x}_d) \in \mathbb{R}^d$ ,  $u = (u_1, u_2, \dots, u_{d+1})$ ,  $\mathbf{u} = (\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_{d+1}) \in \mathbb{R}^{1+d}$ ,  $\mathbf{v} \in C([0, T) \times \mathbb{R}^d, \mathbb{R}^{1+d})$  that

$$\begin{aligned} & \max\{|f(t, x, u) - f(t, x, \mathbf{u})|, |g(x) - g(\mathfrak{x})|\} \\ & \leq \left[ \sum_{v=1}^{d+1} L_v |u_v - \mathbf{u}_v| \right] + \left[ \sum_{v=1}^d K_v |x_v - \mathfrak{x}_v| \right], \end{aligned} \quad (36)$$

$$\begin{aligned} & \mathbb{E} \left[ \left\| g(x + W_T^0 - W_t^0) \left( 1, \frac{W_T^0 - W_t^0}{T-t} \right) \right\|_1 \right. \\ & \quad \left. + \int_t^T \left\| [(F(\mathbf{u}))(t, x + W_s^0 - W_t^0)] \left( 1, \frac{W_s^0 - W_t^0}{s-t} \right) \right\|_1 ds \right] < \infty, \end{aligned} \quad (37)$$

$$\begin{aligned} \mathbf{u}(t, x) = & \mathbb{E} \left[ g(x + W_T^0 - W_t^0) \left( 1, \frac{W_T^0 - W_t^0}{T-t} \right) \right. \\ & \quad \left. + \int_t^T [(F(\mathbf{u}))(t, x + W_s^0 - W_t^0)] \left( 1, \frac{W_s^0 - W_t^0}{s-t} \right) ds \right], \end{aligned} \quad (38)$$

and  $(F(\mathbf{v}))(t, x) = f(t, x, \mathbf{v}(t, x))$ , and let  $\mathbf{U}_{n,M}^\theta = (\mathbf{U}_{n,M}^{\theta,1}, \mathbf{U}_{n,M}^{\theta,2}, \dots, \mathbf{U}_{n,M}^{\theta,d+1}) : [0, T) \times \mathbb{R}^d \times \Omega \rightarrow \mathbb{R}^{1+d}$ ,  $n, M \in \mathbb{Z}$ ,  $\theta \in \Theta$ , satisfy for all  $n, M \in \mathbb{N}$ ,  $\theta \in \Theta$ ,

$t \in [0, T)$ ,  $x \in \mathbb{R}^d$  that  $\mathbf{U}_{-1,M}^\theta(t, x) = \mathbf{U}_{0,M}^\theta(t, x) = 0$  and

$$\begin{aligned} \mathbf{U}_{n,M}^\theta(t, x) &= (g(x), 0) + \sum_{i=1}^{M^n} \frac{(g(x + W_T^{(\theta, 0, -i)} - W_t^{(\theta, 0, -i)}) - g(x))}{M^n} \left( 1, \frac{W_T^{(\theta, 0, -i)} - W_t^{(\theta, 0, -i)}}{T-t} \right) \\ &+ \sum_{l=0}^{n-1} \sum_{i=1}^{M^{n-l}} \frac{(F(\mathbf{U}_{l,M}^{(\theta, l, i)}) - \mathbb{1}_N(l) F(\mathbf{U}_{l-1,M}^{(\theta, l, i)})) (\mathcal{R}_t^{(\theta, l, i)}, x + W_{\mathcal{R}_t^{(\theta, l, i)}}^{(\theta, l, i)} - W_t^{(\theta, l, i)})}{M^{n-l} \varrho(t, \mathcal{R}_t^{(\theta, l, i)})} \\ &\cdot \left( 1, \frac{W_{\mathcal{R}_t^{(\theta, l, i)}}^{(\theta, l, i)} - W_t^{(\theta, l, i)}}{\mathcal{R}_t^{(\theta, l, i)} - t} \right). \end{aligned} \quad (39)$$

### 3.2 Properties of MLP Approximations

**Lemma 3.2** (Measurability properties) Assume Setting 3.1 and let  $M \in \mathbb{N}$ . Then

- (i) for all  $n \in \mathbb{N}_0$ ,  $\theta \in \Theta$  it holds that  $\mathbf{U}_{n,M}^\theta : [0, T) \times \mathbb{R}^d \times \Omega \rightarrow \mathbb{R}^{d+1}$  is a continuous random field,
- (ii) for all  $n \in \mathbb{N}_0$ ,  $\theta \in \Theta$  it holds that  $\sigma(\mathbf{U}_{n,M}^\theta) \subseteq \sigma((\mathbf{r}^{(\theta, \vartheta)})_{\vartheta \in \Theta}, (W^{(\theta, \vartheta)})_{\vartheta \in \Theta})$ ,
- (iii) for all  $n, m \in \mathbb{N}_0$ ,  $i, j, k, l \in \mathbb{Z}$ ,  $\theta \in \Theta$  with  $(i, j) \neq (k, l)$  it holds that  $\mathbf{U}_{n,M}^{(\theta, i, j)}$  and  $\mathbf{U}_{m,M}^{(\theta, k, l)}$  are independent,
- (iv) for all  $n \in \mathbb{N}_0$ ,  $\theta \in \Theta$  it holds that  $\mathbf{U}_{n,M}^\theta$ ,  $W^\theta$ , and  $\mathbf{r}^\theta$  are independent,
- (v) for all  $n \in \mathbb{N}_0$  it holds that  $\mathbf{U}_{n,M}^\theta$ ,  $\theta \in \Theta$ , are identically distributed, and
- (vi) for all  $\theta \in \Theta$ ,  $l \in \mathbb{N}$ ,  $i \in \mathbb{N}$ ,  $t \in [0, T)$ ,  $x \in \mathbb{R}^d$  it holds that

$$\frac{F(\mathbf{U}_{l-1,M}^{(\theta, l, i)}) (\mathcal{R}_t^{(\theta, l, i)}, x + W_{\mathcal{R}_t^{(\theta, l, i)}}^{(\theta, l, i)} - W_t^{(\theta, l, i)})}{\varrho(t, \mathcal{R}_t^{(\theta, l, i)})} \left( 1, \frac{W_{\mathcal{R}_t^{(\theta, l, i)}}^{(\theta, l, i)} - W_t^{(\theta, l, i)}}{\mathcal{R}_t^{(\theta, l, i)} - t} \right) \quad (40)$$

and

$$\frac{F(\mathbf{U}_{l-1,M}^{(\theta, l, i)}) (\mathcal{R}_t^{(\theta, l, i)}, x + W_{\mathcal{R}_t^{(\theta, l, i)}}^{(\theta, l, i)} - W_t^{(\theta, l, i)})}{\varrho(t, \mathcal{R}_t^{(\theta, l, i)})} \left( 1, \frac{W_{\mathcal{R}_t^{(\theta, l, i)}}^{(\theta, l, i)} - W_t^{(\theta, l, i)}}{\mathcal{R}_t^{(\theta, l, i)} - t} \right) \quad (41)$$

are identically distributed.

**Proof of Lemma 3.2** First, observe that (39), the hypothesis that for all  $M \in \mathbb{N}, \theta \in \Theta$  it holds that  $\mathbf{U}_{0,M}^\theta = 0$ , the fact that for all  $\theta \in \Theta$  it holds that  $W^\theta$  and  $\mathcal{R}^\theta$  are continuous random fields, the hypothesis that  $f \in C([0, T] \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d, \mathbb{R})$ , the hypothesis that  $g \in C(\mathbb{R}^d, \mathbb{R})$ , the fact that  $\varrho|_{\{(s, t) \in [0, T]^2 : s < t\}} \in C(\{(s, t) \in [0, T]^2 : s < t\}, \mathbb{R})$ , and induction on  $\mathbb{N}_0$  establish Item (i). Next note that Item (i), the hypothesis that  $f \in C([0, T] \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d, \mathbb{R})$ , and, e.g., Beck et al. [2, Lemma 2.4] assure that for all  $n \in \mathbb{N}_0, \theta \in \Theta$  it holds that  $F(\mathbf{U}_{n,M}^\theta)$  is  $(\mathcal{B}([0, T]) \otimes \sigma(W^\theta)) / \mathcal{B}(\mathbb{R})$ -measurable. The hypothesis that for all  $M \in \mathbb{N}, \theta \in \Theta$  it holds that  $\mathbf{U}_{0,M}^\theta = 0$ , (39), the fact that for all  $\theta \in \Theta$  it holds that  $W^\theta$  is  $(\mathcal{B}([0, T]) \otimes \sigma(W^\theta)) / \mathcal{B}(\mathbb{R})$ -measurable, the fact that for

all  $\theta \in \Theta$  it holds that  $\mathcal{R}^\theta$  is  $(\mathcal{B}([0, T]) \otimes \sigma(\tau^\theta)) / \mathcal{B}([0, T])$ -measurable, and induction on  $\mathbb{N}_0$  hence prove Item (ii). In addition, note that Item (ii) and the fact that for all  $i, j, k, l \in \mathbb{Z}$ ,  $\theta \in \Theta$  with  $(i, j) \neq (k, l)$  it holds that  $((\tau^{(\theta, i, j, \vartheta)}, W^{(\theta, i, j, \vartheta)}))_{\vartheta \in \Theta}$  and  $((\tau^{(\theta, k, l, \vartheta)}, W^{(\theta, k, l, \vartheta)}))_{\vartheta \in \Theta}$  are independent prove Item (iii). Furthermore, observe that Item (ii) and the fact that for all  $\theta \in \Theta$  it holds that  $(\tau^{(\theta, \vartheta)})_{\vartheta \in \Theta}$ ,  $(W^{(\theta, \vartheta)})_{\vartheta \in \Theta}$ ,  $W^\theta$ , and  $\tau^\theta$  are independent establish Item (iv). Next observe that the hypothesis that for all  $\theta \in \Theta$  it holds that  $\mathbf{U}_{0, M}^\theta = 0$ , the hypothesis that  $(W^\theta)_{\theta \in \Theta}$  are i.i.d., the hypothesis that  $(\mathcal{R}^\theta)_{\theta \in \Theta}$  are i.i.d., Items (i)–(iii), Hutzenthaler et al. [45, Corollary 2.5], and induction on  $\mathbb{N}_0$  establish Item (v). Furthermore, observe that Item (ii), the fact that for all  $\theta \in \Theta$ ,  $l \in \mathbb{N}$ ,  $i \in \mathbb{N}$  it holds that  $(\tau^{(\theta, -l, i, \vartheta)})_{\vartheta \in \Theta}$ ,  $(W^{(\theta, -l, i, \vartheta)})_{\vartheta \in \Theta}$ ,  $W^{\theta, l, i}$ , and  $\tau^{\theta, l, i}$  are independent, and, e.g., Hutzenthaler et al. [45, Lemma 2.3] imply that for every  $\theta \in \Theta$ ,  $l, i \in \mathbb{N}$ ,  $t \in [0, T)$ ,  $x \in \mathbb{R}^d$  and every bounded  $\mathcal{B}(\mathbb{R}^{d+1}) / \mathcal{B}(\mathbb{R})$ -measurable  $\psi: \mathbb{R}^{d+1} \rightarrow \mathbb{R}$  it holds that

$$\begin{aligned} & \mathbb{E} \left[ \psi \left( \frac{F(\mathbf{U}_{l-1, M}^{(\theta, -l, i)})(\mathcal{R}_t^{(\theta, l, i)}, x + W_{\mathcal{R}_t^{(\theta, l, i)}}^{(\theta, l, i)} - W_t^{(\theta, l, i)})}{\varrho(t, \mathcal{R}_t^{(\theta, l, i)})} \left( 1, \frac{W_{\mathcal{R}_t^{(\theta, l, i)}}^{(\theta, l, i)} - W_t^{(\theta, l, i)}}{\mathcal{R}_t^{(\theta, l, i)} - t} \right) \right) \right] \\ &= \mathbb{E} \left[ \mathbb{E} \left[ \psi \left( \frac{F(\mathbf{U}_{l-1, M}^{(\theta, -l, i)})(r, x + z)}{\varrho(t, r)} \left( 1, \frac{z}{r-t} \right) \right) \right] \Big|_{(r, z) = (\mathcal{R}_t^{(\theta, l, i)}, W_{\mathcal{R}_t^{(\theta, l, i)}}^{(\theta, l, i)} - W_t^{(\theta, l, i)})} \right]. \end{aligned} \quad (42)$$

This, Item (v), Item (ii), the fact that for all  $\theta \in \Theta$ ,  $l, i \in \mathbb{N}$  it holds that  $(\tau^{(\theta, l, i, \vartheta)})_{\vartheta \in \Theta}$ ,  $(W^{(\theta, l, i, \vartheta)})_{\vartheta \in \Theta}$ ,  $W^{\theta, l, i}$ , and  $\tau^{\theta, l, i}$  are independent, and, e.g., Hutzenthaler et al. [45, Lemma 2.3] imply that for every  $\theta \in \Theta$ ,  $l, i \in \mathbb{N}$ ,  $t \in [0, T)$ ,  $x \in \mathbb{R}^d$  and every bounded  $\mathcal{B}(\mathbb{R}^{d+1}) / \mathcal{B}(\mathbb{R})$ -measurable  $\psi: \mathbb{R}^{d+1} \rightarrow \mathbb{R}$  it holds that

$$\begin{aligned} & \mathbb{E} \left[ \psi \left( \frac{F(\mathbf{U}_{l-1, M}^{(\theta, -l, i)})(\mathcal{R}_t^{(\theta, l, i)}, x + W_{\mathcal{R}_t^{(\theta, l, i)}}^{(\theta, l, i)} - W_t^{(\theta, l, i)})}{\varrho(t, \mathcal{R}_t^{(\theta, l, i)})} \left( 1, \frac{W_{\mathcal{R}_t^{(\theta, l, i)}}^{(\theta, l, i)} - W_t^{(\theta, l, i)}}{\mathcal{R}_t^{(\theta, l, i)} - t} \right) \right) \right] \\ &= \mathbb{E} \left[ \mathbb{E} \left[ \psi \left( \frac{F(\mathbf{U}_{l-1, M}^{(\theta, -l, i)})(r, x + z)}{\varrho(t, r)} \left( 1, \frac{z}{r-t} \right) \right) \right] \Big|_{(r, z) = (\mathcal{R}_t^{(\theta, l, i)}, W_{\mathcal{R}_t^{(\theta, l, i)}}^{(\theta, l, i)} - W_t^{(\theta, l, i)})} \right] \\ &= \mathbb{E} \left[ \psi \left( \frac{F(\mathbf{U}_{l-1, M}^{(\theta, l, i)})(\mathcal{R}_t^{(\theta, l, i)}, x + W_{\mathcal{R}_t^{(\theta, l, i)}}^{(\theta, l, i)} - W_t^{(\theta, l, i)})}{\varrho(t, \mathcal{R}_t^{(\theta, l, i)})} \left( 1, \frac{W_{\mathcal{R}_t^{(\theta, l, i)}}^{(\theta, l, i)} - W_t^{(\theta, l, i)}}{\mathcal{R}_t^{(\theta, l, i)} - t} \right) \right) \right]. \end{aligned} \quad (43)$$

This establishes Item (vi). The proof of Lemma 3.2 is thus completed.  $\square$

**Lemma 3.3** (Approximations are integrable) *Assume Setting 3.1, let  $p \in (1, \infty)$ ,  $M \in \mathbb{N}$ ,  $x \in \mathbb{R}^d$ , and assume for all  $q \in [1, p)$ ,  $t \in [0, T)$  that*

$$\int_0^1 \frac{1}{s^{\frac{q}{2}} [\rho(s)]^{q-1}} ds + \sup_{s \in [t, T)} \mathbb{E} \left[ \left| f(s, x + W_s^0 - W_t^0, 0, 0) \right|^q \right] < \infty. \quad (44)$$

Then

(i) it holds for all  $\theta \in \Theta$ ,  $q \in [1, \infty)$ ,  $v \in \{1, 2, \dots, d+1\}$  that

$$\sup_{u \in (0, T]} \sup_{t \in [0, u)} \sup_{y \in \mathbb{R}^d} \mathbb{E} \left[ |(g(y + W_u^\theta - W_t^\theta) - g(y)) \langle \mathbf{e}_v, (1, \frac{W_u^\theta - W_t^\theta}{u-t}) \rangle|^q \right] < \infty, \quad (45)$$

(ii) it holds for all  $n \in \mathbb{N}_0$ ,  $\theta \in \Theta$ ,  $q \in [1, p)$ ,  $t \in [0, T)$ ,  $v \in \{1, 2, \dots, d+1\}$  that

$$\begin{aligned} & \sup_{s \in [t, T)} \left| \mathbb{E} \left[ |(\mathbf{U}_{n,M}^\theta(s, x + W_s^\theta - W_t^\theta))_v|^q \right] \right| \\ & + \mathbb{E} \left[ \left| \frac{(F(\mathbf{U}_{n,M}^\theta))(\mathcal{R}_s^\theta, x + W_s^\theta - W_t^\theta)^q}{[\varrho(s, \mathcal{R}_s^\theta)]^q} \left| \langle \mathbf{e}_v, (1, \frac{W_s^\theta - W_t^\theta}{\mathcal{R}_s^\theta - s}) \rangle \right|^q \right| \right] < \infty, \end{aligned} \quad (46)$$

and

(iii) it holds for all  $n \in \mathbb{N}$ ,  $\theta \in \Theta$ ,  $t \in [0, T)$  that

$$\begin{aligned} \mathbb{E}[\mathbf{U}_{n,M}^\theta(t, x)] &= \mathbb{E} \left[ g(x + W_T^\theta - W_t^\theta) \left( 1, \frac{W_T^\theta - W_t^\theta}{T-t} \right) \right] \\ &+ \mathbb{E} \left[ \left( F(\mathbf{U}_{n-1,M,Q}^\theta) \right) (\mathcal{R}_t^\theta, x + W_{\mathcal{R}_t^\theta}^\theta - W_t^\theta) \left( 1, \frac{W_{\mathcal{R}_t^\theta}^\theta - W_t^\theta}{\mathcal{R}_t^\theta - t} \right) \right]. \end{aligned} \quad (47)$$

**Proof of Lemma 3.3** Observe that the Cauchy–Schwarz inequality, (36), Jensen’s inequality, and the fact that for all  $u \in (0, T]$ ,  $t \in [0, u)$ ,  $i \in \{1, 2, \dots, d\}$ ,  $\theta \in \Theta$  it holds that  $W_u^{\theta,i} - W_t^{\theta,i}$  and  $(u-t)^{1/2} T^{-1/2} W_T^{0,1}$  are identically distributed yield that for all  $\theta \in \Theta$ ,  $q \in [1, \infty)$ ,  $v \in \{1, 2, \dots, d+1\}$  it holds that

$$\begin{aligned} & \sup_{u \in (0, T]} \sup_{t \in [0, u)} \sup_{y \in \mathbb{R}^d} \left( \mathbb{E} \left[ |(g(y + W_u^\theta - W_t^\theta) - g(y)) \langle \mathbf{e}_v, (1, \frac{W_u^\theta - W_t^\theta}{u-t}) \rangle|^q \right] \right)^2 \\ & \leq \sup_{u \in (0, T]} \sup_{t \in [0, u)} \sup_{y \in \mathbb{R}^d} \left( \mathbb{E} \left[ |g(y + W_u^\theta - W_t^\theta) - g(y)|^{2q} \right] \mathbb{E} \left[ |\langle \mathbf{e}_v, (1, \frac{W_u^\theta - W_t^\theta}{u-t}) \rangle|^{2q} \right] \right) \\ & \leq \sup_{u \in (0, T]} \sup_{t \in [0, u)} \left( \mathbb{E} \left[ \left| \sum_{i=1}^d K_i |W_u^{\theta,i} - W_t^{\theta,i}| \right|^{2q} \right] \mathbb{E} \left[ 1 + \left| \frac{W_u^{\theta,1} - W_t^{\theta,1}}{u-t} \right|^{2q} \right] \right) \\ & \leq \sup_{u \in (0, T]} \sup_{t \in [0, u)} \left( d^{2q-1} \left( \sum_{i=1}^d (K_i)^{2q} \right) \frac{(u-t)^q}{T^q} \mathbb{E}[|W_T^{0,1}|^{2q}] \left( 1 + \frac{(u-t)^q}{(u-t)^{2q} T^q} \mathbb{E}[|W_T^{0,1}|^{2q}] \right) \right) \\ & = d^{2q-1} \left( \sum_{i=1}^d (K_i)^{2q} \right) \mathbb{E}[|W_T^{0,1}|^{2q}] \left( 1 + \frac{1}{T^{2q}} \mathbb{E}[|W_T^{0,1}|^{2q}] \right) < \infty. \end{aligned} \quad (48)$$

This proves Item (i). Next observe that Lemma 3.2 ensures that for all  $n \in \mathbb{N}_0$ ,  $\theta \in \Theta$  it holds that  $W^\theta$ ,  $\mathcal{R}^\theta$ , and  $\mathbf{U}_{n,M}^\theta$  are independent continuous random fields. Combining this and, e.g., Hutzenthaler et al. [45, Lemma 2.3] with Hölder’s inequality (applied with  $p = \frac{p+q}{2q}$ ,  $q = \frac{p+q}{p-q}$  in the notation of Hölder’s inequality) and the fact that for all  $s \in [0, T]$ ,  $h \in (0, \infty)$ ,  $\theta \in \Theta$  it holds that  $W_{s+h}^{\theta,1} - W_s^{\theta,1}$  and  $h^{1/2} T^{-1/2} W_T^{0,1}$  are

identically distributed demonstrates that for all  $t \in [0, T)$ ,  $s \in [t, T)$ ,  $n \in \mathbb{N}_0$ ,  $\theta \in \Theta$ ,  $q \in [1, p)$ ,  $v \in \{1, 2, \dots, d+1\}$  it holds that

$$\begin{aligned}
& \mathbb{E} \left[ \frac{|(F(\mathbf{U}_{n,M}^\theta))(\mathcal{R}_s^\theta, x + W_s^\theta - W_t^\theta)|^q}{[\varrho(s, \mathcal{R}_s^\theta)]^q} \left| \langle \mathbf{e}_v, (1, \frac{W_\theta^\theta - W_s^\theta}{\mathcal{R}_s^\theta - s}) \rangle \right|^q \right] \\
&= \int_0^1 \mathbb{E} \left[ \frac{|(F(\mathbf{U}_{n,M}^\theta))(s+u(T-s), x + W_{s+u(T-s)}^\theta - W_t^\theta)|^q}{[\varrho(s+s+u(T-s))]^q} \left| \langle \mathbf{e}_v, (1, \frac{W_{s+u(T-s)}^\theta - W_s^\theta}{s+u(T-s)-s}) \rangle \right|^q \right] \rho(u) du \\
&\leq \int_0^1 \frac{\sup_{z \in [t, T)} \left( \mathbb{E} \left[ |(F(\mathbf{U}_{n,M}^\theta))(z, x + W_z^\theta - W_t^\theta)|^{\frac{p+q}{2}} \right] \right)^{\frac{2q}{p+q}} (T-s)^q}{[\rho(u)]^{q-1}} \\
&\quad \cdot \left( 1 + \left( \mathbb{E} \left[ \left| \frac{W_{s+u(T-s)}^{\theta,1} - W_s^{\theta,1}}{u(T-s)} \right|^{\frac{q(p+q)}{p-q}} \right] \right)^{\frac{p-q}{p+q}} \right) du \\
&= \left( \sup_{z \in [t, T)} \mathbb{E} \left[ |(F(\mathbf{U}_{n,M}^\theta))(z, x + W_z^\theta - W_t^\theta)|^{\frac{p+q}{2}} \right] \right)^{\frac{2q}{p+q}} \\
&\quad \cdot \int_0^1 \frac{(T-s)^q}{[\rho(u)]^{q-1}} \left[ 1 + \frac{(u(T-s))^{\frac{q}{2}}}{(u(T-s))^q T^{\frac{q}{2}}} \left( \mathbb{E} [|W_T^{0,1}|^{\frac{q(p+q)}{p-q}}] \right)^{\frac{p-q}{p+q}} \right] du \\
&\leq \left( \sup_{z \in [t, T)} \mathbb{E} \left[ |(F(\mathbf{U}_{n,M}^\theta))(z, x + W_z^\theta - W_t^\theta)|^{\frac{p+q}{2}} \right] \right)^{\frac{2q}{p+q}} \\
&\quad \cdot \left( T^q + \left( \mathbb{E} [|W_T^{0,1}|^{\frac{q(p+q)}{p-q}}] \right)^{\frac{p-q}{p+q}} \right) \int_0^1 \frac{u^{-\frac{q}{2}}}{[\rho(u)]^{q-1}} du. \tag{49}
\end{aligned}$$

In the next step we prove (46) by induction on  $n \in \mathbb{N}_0$ . For the base case  $n = 0$  observe that (49), (44), and the fact that for all  $\theta \in \Theta$  it holds that  $U_{0,M}^\theta = 0$  ensure that for all  $\theta \in \Theta$ ,  $q \in [1, p)$ ,  $t \in [0, T)$ ,  $v \in \{1, 2, \dots, d+1\}$  it holds that

$$\begin{aligned}
& \sup_{s \in [t, T)} \left| \mathbb{E} \left[ |(\mathbf{U}_{0,M}^\theta)(s, x + W_s^\theta - W_t^\theta)_v|^q \right] \right. \\
& \quad \left. + \mathbb{E} \left[ \frac{|(F(\mathbf{U}_{0,M}^\theta))(\mathcal{R}_s^\theta, x + W_s^\theta - W_t^\theta)|^q}{[\varrho(s, \mathcal{R}_s^\theta)]^q} \left| \langle \mathbf{e}_v, (1, \frac{W_\theta^\theta - W_s^\theta}{\mathcal{R}_s^\theta - s}) \rangle \right|^q \right] \right| \\
&= \sup_{s \in [t, T)} \left| \mathbb{E} \left[ \frac{|(F(0))(\mathcal{R}_s^\theta, x + W_s^\theta - W_t^\theta)|^q}{[\varrho(s, \mathcal{R}_s^\theta)]^q} \left| \langle \mathbf{e}_v, (1, \frac{W_\theta^\theta - W_s^\theta}{\mathcal{R}_s^\theta - s}) \rangle \right|^q \right] \right| \\
&\leq \left( \sup_{z \in [t, T)} \mathbb{E} \left[ |(F(0))(z, x + W_z^0 - W_t^0)|^{\frac{p+q}{2}} \right] \right)^{\frac{2q}{p+q}} \\
&\quad \cdot \left( T^q + \left( \mathbb{E} [|W_T^{0,1}|^{\frac{q(p+q)}{p-q}}] \right)^{\frac{p-q}{p+q}} \right) \int_0^1 \frac{u^{-\frac{q}{2}}}{[\rho(u)]^{q-1}} du < \infty. \tag{50}
\end{aligned}$$

This establishes (46) in the base case  $n = 0$ . For the induction step  $\mathbb{N}_0 \ni n-1 \rightsquigarrow n \in \mathbb{N}$  assume that there exists  $n \in \mathbb{N}$  which satisfies for all  $k \in \mathbb{N}_0 \cap [0, n)$ ,  $\theta \in \Theta$ ,

$q \in [1, p)$ ,  $t \in [0, T)$ ,  $v \in \{1, 2, \dots, d+1\}$  that

$$\begin{aligned} & \sup_{s \in [t, T)} \left| \mathbb{E} \left[ |(\mathbf{U}_{k,M}^\theta(s, x + W_s^\theta - W_t^\theta))_v|^q \right] \right| \\ & + \mathbb{E} \left[ \left| \frac{(F(\mathbf{U}_{k,M}^\theta))(\mathcal{R}_s^\theta, x + W_s^\theta - W_t^\theta)}{[\varrho(s, \mathcal{R}_s^\theta)]^q} \right|^q \left| \langle \mathbf{e}_v, (1, \frac{W_{\mathcal{R}_s^\theta}^\theta - W_s^\theta}{\mathcal{R}_s^\theta - s}) \rangle \right|^q \right] < \infty. \end{aligned} \quad (51)$$

Observe that (39) and Jensen's inequality ensure that for all  $\theta \in \Theta$ ,  $q \in [1, p)$ ,  $t \in [0, T)$ ,  $s \in [t, T)$ ,  $v \in \{1, 2, \dots, d+1\}$  it holds that

$$\begin{aligned} & \mathbb{E} \left[ |(\mathbf{U}_{n,M}^\theta(s, x + W_s^\theta - W_t^\theta))_v|^q \right] \\ & \leq (3+2n)^{q-1} |g(x)|^q + (3+2n)^{q-1} \mathbb{E} \left[ |g(x + W_s^\theta - W_t^\theta) - g(x)|^q \right] + \frac{(3+2n)^{q-1}}{M^n} \\ & \cdot \sum_{i=1}^{M^n} \mathbb{E} \left[ \left| g(x + W_s^\theta - W_t^\theta + W_T^{(\theta,0,-i)} - W_s^{(\theta,0,-i)}) - g(x + W_s^\theta - W_t^\theta) \right|^q \right. \\ & \cdot \left. \left| \langle \mathbf{e}_v, (1, \frac{W_T^{(\theta,0,-i)} - W_s^{(\theta,0,-i)}}{T-s}) \rangle \right|^q \right] \\ & + \sum_{l=0}^{n-1} \frac{(3+2n)^{q-1}}{M^{n-l}} \\ & \cdot \sum_{i=1}^{M^{n-l}} \mathbb{E} \left[ \left| \frac{(F(\mathbf{U}_{l,M}^{(\theta,l,i)}))(\mathcal{R}_s^{(\theta,l,i)}, x + W_s^\theta - W_t^\theta + W_{\mathcal{R}_s^{(\theta,l,i)}}^{(\theta,l,i)} - W_s^{(\theta,l,i)})}{[\varrho(s, \mathcal{R}_s^{(\theta,l,i)})]^q} \right|^q \left| \langle \mathbf{e}_v, (1, \frac{W_{\mathcal{R}_s^{(\theta,l,i)}}^{(\theta,l,i)} - W_s^{(\theta,l,i)}}{\mathcal{R}_s^{(\theta,l,i)} - s}) \rangle \right|^q \right] \\ & + \sum_{l=1}^{n-1} \frac{(3+2n)^{q-1}}{M^{n-l}} \\ & \cdot \sum_{i=1}^{M^{n-l}} \mathbb{E} \left[ \left| \frac{(F(\mathbf{U}_{l-1,M}^{(\theta,-l,i)}))(\mathcal{R}_s^{(\theta,l,i)}, x + W_s^\theta - W_t^\theta + W_{\mathcal{R}_s^{(\theta,l,i)}}^{(\theta,l,i)} - W_s^{(\theta,l,i)})}{[\varrho(s, \mathcal{R}_s^{(\theta,l,i)})]^q} \right|^q \left| \langle \mathbf{e}_v, (1, \frac{W_{\mathcal{R}_s^{(\theta,l,i)}}^{(\theta,l,i)} - W_s^{(\theta,l,i)}}{\mathcal{R}_s^{(\theta,l,i)} - s}) \rangle \right|^q \right]. \end{aligned} \quad (52)$$

This, e.g., Hutzenthaler et al. [45, Corollary 2.5], Lemma 3.2, Item (i), and (51) yield that for all  $\theta \in \Theta$ ,  $q \in [1, p)$ ,  $t \in [0, T)$ ,  $v \in \{1, 2, \dots, d+1\}$  it holds that

$$\begin{aligned} & \sup_{s \in [t, T)} \mathbb{E} \left[ |(\mathbf{U}_{n,M}^\theta(s, x + W_s^\theta - W_t^\theta))_v|^q \right] \leq (3+2n)^{q-1} |g(x)|^q \\ & + (6+2n)^q \sup_{u \in (0, T]} \sup_{s \in [0, u]} \sup_{y \in \mathbb{R}^d} \sup_{j \in \{1, 2, \dots, d+1\}} \\ & \mathbb{E} \left[ \left| g(y + W_u^0 - W_s^0) - g(y) \right|^q \left| \langle \mathbf{e}_j, (1, \frac{W_u^0 - W_s^0}{u-s}) \rangle \right|^q \right] \\ & + \sum_{l=0}^{n-1} (6+2n)^q \sup_{s \in [t, T)} \mathbb{E} \left[ \left| \frac{(F(\mathbf{U}_{l,M}^0))(\mathcal{R}_s^0, x + W_s^0 - W_t^0)}{[\varrho(s, \mathcal{R}_s^0)]^q} \right|^q \left| \langle \mathbf{e}_v, (1, \frac{W_{\mathcal{R}_s^0}^0 - W_s^0}{\mathcal{R}_s^0 - s}) \rangle \right|^q \right] < \infty. \end{aligned} \quad (53)$$

Jensen's inequality and (44) hence yield that for all  $\theta \in \Theta$ ,  $q \in [1, p)$ ,  $t \in [0, T)$  it holds that

$$\begin{aligned}
& \sup_{s \in [t, T)} \left| \mathbb{E} \left[ |(F(\mathbf{U}_{n,M}^\theta))(s, x + W_s^\theta - W_t^\theta)|^q \right] \right| \\
& \leq 2^{q-1} \sup_{s \in [t, T)} \left( \mathbb{E} \left[ |(F(\mathbf{U}_{n,M}^\theta) - F(0))(s, x + W_s^\theta - W_t^\theta)|^q \right] \right. \\
& \quad \left. + \mathbb{E} \left[ |(F(0))(s, x + W_s^\theta - W_t^\theta)|^q \right] \right) \\
& \leq 2^{q-1} \sup_{s \in [t, T)} \left( \mathbb{E} \left[ \left| \sum_{v=1}^{d+1} L_v \mathbf{U}_{n,M}^{\theta,v}(s, x + W_s^\theta - W_t^\theta) \right|^q \right] \right. \\
& \quad \left. + \mathbb{E} \left[ |(F(0))(s, x + W_s^\theta - W_t^\theta)|^q \right] \right) \\
& \leq (4d)^q \sum_{v=1}^{d+1} (L_v)^q \sup_{s \in [t, T)} \mathbb{E} \left[ |\mathbf{U}_{n,M}^{\theta,v}(s, x + W_s^\theta - W_t^\theta)|^q \right] + 2^q \sup_{s \in [t, T)} \\
& \quad \cdot \mathbb{E} \left[ |(F(0))(s, x + W_s^\theta - W_t^\theta)|^q \right] \\
& < \infty.
\end{aligned} \tag{54}$$

This, (49), and (44) imply that for all  $\theta \in \Theta$ ,  $q \in [1, p)$ ,  $t \in [0, T)$ ,  $v \in \{1, 2, \dots, d+1\}$  it holds that

$$\begin{aligned}
& \sup_{s \in [t, T)} \mathbb{E} \left[ \frac{\left| (F(\mathbf{U}_{n,M}^\theta))(\mathcal{R}_s^\theta, x + W_s^\theta - W_t^\theta) \right|^q}{[\varrho(s, \mathcal{R}_s^\theta)]^q} \left| \langle \mathbf{e}_v, (1, \frac{W_s^\theta - W_t^\theta}{\mathcal{R}_s^\theta - s}) \rangle \right|^q \right] \\
& \leq \left( \sup_{s \in [t, T)} \mathbb{E} \left[ |(F(\mathbf{U}_{n,M}^\theta))(s, x + W_s^\theta - W_t^\theta)|^{\frac{p+q}{2}} \right] \right)^{\frac{2q}{p+q}} \\
& \quad \cdot \left( T^q + \left( \mathbb{E} \left[ \left| W_T^{0,1} \right|^{\frac{q(p+q)}{p-q}} \right] \right)^{\frac{p-q}{p+q}} \right) \int_0^1 \frac{u^{-\frac{q}{2}}}{[\varrho(u)]^{q-1}} du \\
& < \infty.
\end{aligned} \tag{55}$$

Combining this with (53) demonstrates that for all  $\theta \in \Theta$ ,  $q \in [1, p)$ ,  $t \in [0, T)$ ,  $v \in \{1, 2, \dots, d+1\}$  it holds that

$$\begin{aligned}
& \sup_{s \in [t, T)} \mathbb{E} \left[ \left| (\mathbf{U}_{n,M}^\theta(s, x + W_s^\theta - W_t^\theta))_v \right|^q \right. \\
& \quad \left. + \frac{\left| (F(\mathbf{U}_{n,M}^\theta))(\mathcal{R}_s^\theta, x + W_s^\theta - W_t^\theta) \right|^q}{[\varrho(s, \mathcal{R}_s^\theta)]^q} \left| \langle \mathbf{e}_v, (1, \frac{W_s^\theta - W_t^\theta}{\mathcal{R}_s^\theta - s}) \rangle \right|^q \right] < \infty. \tag{56}
\end{aligned}$$

Induction thus proves Item (ii). Next we prove Item (iii). Note that Item (iii), e.g., Hutzenthaler et al. [45, Corollary 2.5], Item (v) of Lemma 3.2, the fact that  $(\mathcal{R}^\theta, W^\theta)$ ,  $\theta \in \Theta$ , are identically distributed, e.g., Hutzenthaler et al. [45, Lemma 2.3], and the fact that for every  $t \in [0, T)$ ,  $s \in [t, T]$ ,  $\theta \in \Theta$  it holds that  $\mathbb{P}(\mathcal{R}_t^\theta \leq s) = \int_t^s \varrho(t, r) dr$  yield that for all  $n \in \mathbb{N}$ ,  $\theta \in \Theta$ ,  $t \in [0, T)$  it holds that

$$\begin{aligned}
& \mathbb{E}[\mathbf{U}_{n,M}^\theta(t, x)] - \mathbb{E}\left[g(x + W_T^0 - W_t^0)\left(1, \frac{W_T^0 - W_t^0}{T-t}\right)\right] \\
&= \sum_{l=0}^{n-1} \sum_{i=1}^{M^{n-l}} \frac{1}{M^{n-l}} \mathbb{E}\left[\frac{\left(F(\mathbf{U}_{l,M}^{(\theta,l,i)})\right)(\mathcal{R}_t^{(\theta,l,i)}, x + W_{\mathcal{R}_t^{(\theta,l,i)}}^{(\theta,l,i)} - W_t^{(\theta,l,i)})}{\varrho(t, \mathcal{R}_t^{(\theta,l,i)})}\left(1, \frac{W_{\mathcal{R}_t^{(\theta,l,i)}}^{(\theta,l,i)} - W_t^{(\theta,l,i)}}{\mathcal{R}_t^{(\theta,l,i)} - t}\right)\right] \\
&\quad - \sum_{l=1}^{n-1} \sum_{i=1}^{M^{n-l}} \frac{1}{M^{n-l}} \mathbb{E}\left[\frac{\left(F(\mathbf{U}_{l-1,M}^{(\theta,l,i)})\right)(\mathcal{R}_t^{(\theta,l,i)}, x + W_{\mathcal{R}_t^{(\theta,l,i)}}^{(\theta,l,i)} - W_t^{(\theta,l,i)})}{\varrho(t, \mathcal{R}_t^{(\theta,l,i)})}\left(1, \frac{W_{\mathcal{R}_t^{(\theta,l,i)}}^{(\theta,l,i)} - W_t^{(\theta,l,i)}}{\mathcal{R}_t^{(\theta,l,i)} - t}\right)\right] \\
&= \sum_{l=0}^{n-1} \left( \mathbb{E}\left[\frac{\left(F(\mathbf{U}_{l,M}^\theta)\right)(\mathcal{R}_t^\theta, x + W_{\mathcal{R}_t^\theta}^\theta - W_t^\theta)}{\varrho(t, \mathcal{R}_t^\theta)}\left(1, \frac{W_{\mathcal{R}_t^\theta}^\theta - W_t^\theta}{\mathcal{R}_t^\theta - t}\right)\right] \right. \\
&\quad \left. - \mathbb{1}_{\mathbb{N}}(l) \mathbb{E}\left[\frac{\left(F(\mathbf{U}_{l-1,M}^\theta)\right)(\mathcal{R}_t^\theta, x + W_{\mathcal{R}_t^\theta}^\theta - W_t^\theta)}{\varrho(t, \mathcal{R}_t^\theta)}\left(1, \frac{W_{\mathcal{R}_t^\theta}^\theta - W_t^\theta}{\mathcal{R}_t^\theta - t}\right)\right]\right) \\
&= \mathbb{E}\left[\frac{\left(F(\mathbf{U}_{n-1,M}^\theta)\right)(\mathcal{R}_t^\theta, x + W_{\mathcal{R}_t^\theta}^\theta - W_t^\theta)}{\varrho(t, \mathcal{R}_t^\theta)}\left(1, \frac{W_{\mathcal{R}_t^\theta}^\theta - W_t^\theta}{\mathcal{R}_t^\theta - t}\right)\right] \\
&= \int_t^T \mathbb{E}\left[\frac{\left(F(\mathbf{U}_{n-1,M}^\theta)\right)(s, x + W_s^\theta - W_t^\theta)}{\varrho(t, s)}\left(1, \frac{W_s^\theta - W_t^\theta}{s-t}\right)\right] \mathbb{P}(\mathcal{R}_t^\theta \in ds) \\
&= \int_t^T \mathbb{E}\left[\left(F(\mathbf{U}_{n-1,M,Q}^\theta)\right)(s, x + W_s^\theta - W_t^\theta)\left(1, \frac{W_s^\theta - W_t^\theta}{s-t}\right)\right] ds.
\end{aligned} \tag{57}$$

This establishes Item (iii). The proof of Lemma 3.3 is thus completed.  $\square$

### 3.3 Error Analysis for MLP Approximations

**Lemma 3.4** (Recursive bound for global error) *Assume Setting 3.1, let  $p \in (1, \infty)$ ,  $M \in \mathbb{N}$ , let  $S: \mathbb{N}_0 \times [0, T) \times \Omega \rightarrow [0, T)$  satisfy for all  $\theta \in \Theta$ ,  $n \in \mathbb{N}_0$ ,  $t \in [0, T)$  that  $S(0, t) = t$  and  $S(n+1, t) = \mathcal{R}_{S(n,t)}^{(n)}$ , and assume for all  $q \in [1, p)$ ,  $t \in [0, T)$ ,  $x \in \mathbb{R}^d$  that*

$$\int_0^1 \frac{1}{s^{\frac{q}{2}} [\rho(s)]^{q-1}} ds + \sup_{s \in [t, T)} \mathbb{E}\left[\left|f(s, x + W_s^0 - W_t^0, 0, 0)\right|^q\right] < \infty. \tag{58}$$

Then it holds for all  $n \in \mathbb{N}$ ,  $t \in [0, T)$ ,  $x \in \mathbb{R}^d$ ,  $v_0 \in \{1, 2, \dots, d+1\}$  that

$$\begin{aligned} & \| \mathbf{U}_{n,M}^{0,v_0}(t, x) - \mathbf{u}_{v_0}(t, x) \|_{L^2(\mathbb{P}; \mathbb{R})} \\ & \leq \sum_{j=0}^{n-1} \sum_{\substack{v_1, v_2, \dots, v_{j+1} \in \\ \{1, 2, \dots, d+1\}}} \binom{n-1}{j} \frac{1_{\{1\}}(v_{j+1}) 2^j [\prod_{i=1}^j L_{v_i}]}{\sqrt{M^{n-j}}} \left\| (g(x + W_T^0 - W_t^0) - g(x + W_{S(j,t)}^0 - W_t^0)) \right. \\ & \quad \cdot \langle \mathbf{e}_{v_j}, \left( 1, \frac{W_T^0 - W_{S(j,t)}^0}{T - S(j,t)} \right) \rangle \left[ \prod_{i=1}^j \frac{1}{\varrho(S(i-1,t), S(i,t))} \langle \mathbf{e}_{v_{i-1}}, \left( 1, \frac{W_{S(i,t)}^0 - W_{S(i-1,t)}^0}{S(i,t) - S(i-1,t)} \right) \rangle \right] \Big\|_{L^2(\mathbb{P}; \mathbb{R})} \\ & \quad + \sum_{j=0}^{n-1} \sum_{\substack{v_1, v_2, \dots, v_{j+1} \in \\ \{1, 2, \dots, d+1\}}} \binom{n-1}{j} \frac{1_{\{1\}}(v_{j+1}) 2^j [\prod_{i=1}^j L_{v_i}]}{\sqrt{M^{n-j}}} \left\| (F(0))(S(j+1, t), x + W_{S(j+1,t)}^0 - W_t^0) \right. \\ & \quad \cdot \left[ \prod_{i=1}^{j+1} \frac{1}{\varrho(S(i-1,t), S(i,t))} \langle \mathbf{e}_{v_{i-1}}, \left( 1, \frac{W_{S(i,t)}^0 - W_{S(i-1,t)}^0}{S(i,t) - S(i-1,t)} \right) \rangle \right] \Big\|_{L^2(\mathbb{P}; \mathbb{R})} \\ & \quad + \sum_{j=0}^{n-1} \sum_{\substack{v_1, v_2, \dots, v_{j+1} \in \\ \{1, 2, \dots, d+1\}}} \binom{n-1}{j} \frac{2^j [\prod_{i=1}^{j+1} L_{v_i}]}{\sqrt{M^{n-j-1}}} \left\| \mathbf{u}_{v_{j+1}}(S(j+1, t), x + W_{S(j+1,t)}^0 - W_t^0) \right. \\ & \quad \cdot \left[ \prod_{i=1}^{j+1} \frac{1}{\varrho(S(i-1,t), S(i,t))} \langle \mathbf{e}_{v_{i-1}}, \left( 1, \frac{W_{S(i,t)}^0 - W_{S(i-1,t)}^0}{S(i,t) - S(i-1,t)} \right) \rangle \right] \Big\|_{L^2(\mathbb{P}; \mathbb{R})}. \end{aligned} \quad (59)$$

**Proof of Lemma 3.4** First, we analyze the *Monte Carlo error*. Item (i) of Lemma 3.3, Item (ii) of Lemma 3.3, and Item (vi) of Lemma 3.2 ensure that for all  $l \in \mathbb{N}$ ,  $x \in \mathbb{R}^d$ ,  $t \in [0, T)$ ,  $v \in \{1, 2, \dots, d+1\}$  it holds that

$$\begin{aligned} & \mathbb{E} \left[ \left| (g(x + W_T^0 - W_t^0) - g(x)) \langle \mathbf{e}_v, \left( 1, \frac{W_T^0 - W_t^0}{T - t} \right) \rangle \right| \right. \\ & \quad \left. + \left| \frac{(F(\mathbf{U}_{l,M}^{(0,l,1)}) - \mathbb{1}_{\mathbb{N}(l)} F(\mathbf{U}_{l-1,M}^{(0,-l,1)})) (\mathcal{R}_t^{(0,l,1)}, x + W_{\mathcal{R}_t^{(0,l,1)}}^0 - W_t^{(0,l,1)})}{\varrho(t, \mathcal{R}_t^{(0,l,1)})} \langle \mathbf{e}_v, \left( 1, \frac{W_{\mathcal{R}_t^{(0,l,1)}}^0 - W_t^{(0,l,1)}}{\mathcal{R}_t^{(0,l,1)} - t} \right) \rangle \right| \right] < \infty. \end{aligned} \quad (60)$$

Moreover, observe that Lemma 3.2 yields that

(a) it holds for all  $x \in \mathbb{R}^d$ ,  $t \in [0, T)$ ,  $v \in \{1, 2, \dots, d+1\}$  that

$$(g(x + W_T^{(0,0,-i)} - W_t^{(0,0,-i)}) - g(x)) \langle \mathbf{e}_v, \left( 1, \frac{W_T^{(0,0,-i)} - W_t^{(0,0,-i)}}{T - t} \right) \rangle, \quad i \in \mathbb{N}_0, \quad (61)$$

are independent random variables,

(b) it holds for all  $x \in \mathbb{R}^d$ ,  $t \in [0, T)$ ,  $v \in \{1, 2, \dots, d+1\}$  that

$$\frac{(F(\mathbf{U}_{l,M}^{(0,l,i)}) - \mathbb{1}_{\mathbb{N}(l)} F(\mathbf{U}_{l-1,M}^{(0,-l,i)})) (\mathcal{R}_t^{(0,l,i)}, x + W_{\mathcal{R}_t^{(0,l,i)}}^0 - W_t^{(0,l,i)})}{\varrho(t, \mathcal{R}_t^{(0,l,i)})} \langle \mathbf{e}_v, \left( 1, \frac{W_{\mathcal{R}_t^{(0,l,i)}}^0 - W_t^{(0,l,i)}}{\mathcal{R}_t^{(0,l,i)} - t} \right) \rangle, \quad l, i \in \mathbb{N}_0, \quad (62)$$

are independent random variables, and

(c) it holds for all  $x \in \mathbb{R}^d$ ,  $t \in [0, T)$ ,  $v \in \{1, 2, \dots, d+1\}$  that

$$(g(x + W_T^{(0,0,-i)} - W_t^{(0,0,-i)}) - g(x)) \langle \mathbf{e}_v, \left( 1, \frac{W_T^{(0,0,-i)} - W_t^{(0,0,-i)}}{T - t} \right) \rangle, \quad i \in \mathbb{N}_0, \quad (63)$$

and

$$\frac{\left( F(\mathbf{U}_{l,M}^{(0,l,i)}) - \mathbb{1}_{\mathbb{N}}(l) F(\mathbf{U}_{l-1,M}^{(0,-l,i)}) \right) (\mathcal{R}_t^{(0,l,i)}, x + W_{\mathcal{R}_t^{(0,l,i)}}^{(0,l,i)} - W_t^{(0,l,i)})}{\varrho(t, \mathcal{R}_t^{(0,l,i)})} \left\langle \mathbf{e}_v, \left( 1, \frac{W_{\mathcal{R}_t^{(0,l,i)}}^{(0,l,1)} - W_t^{(0,l,i)}}{\mathcal{R}_t^{(0,l,i)} - t} \right) \right\rangle, \quad l, i \in \mathbb{N}_0, \quad (64)$$

are independent.

Combining this, (60), and (39) with the fact that for all  $N \in \mathbb{N}$ ,  $X_1, X_2, \dots, X_N \in L^1(\mathbb{P}; \mathbb{R})$  with  $\forall i \in \{1, 2, \dots, N\}$ ,  $j \in \{1, 2, \dots, N\} \setminus \{i\}$ ,  $A, B \in \mathcal{B}(\mathbb{R})$ :  $\mathbb{P}(X_i \in A) \cap \{X_j \in B\} = \mathbb{P}(X_i \in A)\mathbb{P}(X_j \in B)$  it holds that  $\text{Var}(\sum_{i=1}^N X_i) = \sum_{i=1}^N \text{Var}(X_i)$  implies that for all  $m \in \mathbb{N}$ ,  $x \in \mathbb{R}^d$ ,  $t \in [0, T]$ ,  $v \in \{1, 2, \dots, d+1\}$  it holds that

$$\begin{aligned} \text{Var}\left(\mathbf{U}_{m,M}^{0,v}(t, x)\right) &= \frac{1}{M^m} \text{Var}\left(\left(g(x + W_T^0 - W_t^0) - g(x)\right) \langle \mathbf{e}_v, \left(1, \frac{W_T^0 - W_t^0}{T-t}\right) \rangle\right) \\ &\quad + \sum_{l=0}^{m-1} \frac{1}{M^{m-l}} \\ &\quad \cdot \text{Var}\left(\frac{\left(F(\mathbf{U}_{l,M}^{(0,l,1)}) - \mathbb{1}_{\mathbb{N}}(l) F(\mathbf{U}_{l-1,M}^{(0,-l,1)})\right) (\mathcal{R}_t^{(0,l,1)}, x + W_{\mathcal{R}_t^{(0,l,1)}}^{(0,l,1)} - W_t^{(0,l,1)})}{\varrho(t, \mathcal{R}_t^{(0,l,1)})} \left\langle \mathbf{e}_v, \left(1, \frac{W_{\mathcal{R}_t^{(0,l,1)}}^{(0,l,1)} - W_t^{(0,l,1)}}{\mathcal{R}_t^{(0,l,1)} - t}\right) \right\rangle\right) \\ &\leq \frac{1}{M^m} \mathbb{E}\left[\left|\left(g(x + W_T^0 - W_s^0) - g(x)\right) \langle \mathbf{e}_v, \left(1, \frac{W_T^0 - W_s^0}{T-s}\right) \rangle\right|^2\right] \quad (65) \\ &\quad + \sum_{l=0}^{m-1} \frac{1}{M^{m-l}} \\ &\quad \cdot \mathbb{E}\left[\left|\frac{\left(F(\mathbf{U}_{l,M}^{(0,l,1)}) - \mathbb{1}_{\mathbb{N}}(l) F(\mathbf{U}_{l-1,M}^{(0,-l,1)})\right) (\mathcal{R}_t^{(0,l,1)}, x + W_{\mathcal{R}_t^{(0,l,1)}}^{(0,l,1)} - W_t^{(0,l,1)})}{\varrho(t, \mathcal{R}_t^{(0,l,1)})} \left\langle \mathbf{e}_v, \left(1, \frac{W_{\mathcal{R}_t^{(0,l,1)}}^{(0,l,1)} - W_t^{(0,l,1)}}{\mathcal{R}_t^{(0,l,1)} - t}\right) \right\rangle\right|^2\right]. \end{aligned}$$

The triangle inequality and (36) hence yield that for all  $m \in \mathbb{N}$ ,  $x \in \mathbb{R}^d$ ,  $t \in [0, T]$ ,  $v \in \{1, 2, \dots, d+1\}$  it holds that

$$\begin{aligned} \left\| \mathbf{U}_{m,M}^{0,v}(t, x) - \mathbb{E}\left[\mathbf{U}_{m,M}^{0,v}(t, x)\right] \right\|_{L^2(\mathbb{P}; \mathbb{R})} &= \left[ \text{Var}\left(\mathbf{U}_{m,M}^{0,v}(t, x)\right) \right]^{1/2} \\ &\leq \frac{1}{\sqrt{M^m}} \left\| \left(g(x + W_T^0 - W_t^0) - g(x)\right) \langle \mathbf{e}_v, \left(1, \frac{W_T^0 - W_t^0}{T-t}\right) \rangle \right\|_{L^2(\mathbb{P}; \mathbb{R})} \\ &\quad + \sum_{l=0}^{m-1} \frac{1}{\sqrt{M^{m-l}}} \left\| \frac{\left(F(\mathbf{U}_{l,M}^{(0,l,1)}) - \mathbb{1}_{\mathbb{N}}(l) F(\mathbf{U}_{l-1,M}^{(0,-l,1)})\right) (\mathcal{R}_t^{(0,l,1)}, x + W_{\mathcal{R}_t^{(0,l,1)}}^{(0,l,1)} - W_t^{(0,l,1)})}{\varrho(t, \mathcal{R}_t^{(0,l,1)})} \right. \\ &\quad \left. \cdot \left\langle \mathbf{e}_v, \left(1, \frac{W_{\mathcal{R}_t^{(0,l,1)}}^{(0,l,1)} - W_t^{(0,l,1)}}{\mathcal{R}_t^{(0,l,1)} - t}\right) \right\rangle \right\|_{L^2(\mathbb{P}; \mathbb{R})} \\ &\leq \frac{1}{\sqrt{M^m}} \left\| \left(g(x + W_T^0 - W_t^0) - g(x)\right) \langle \mathbf{e}_v, \left(1, \frac{W_T^0 - W_t^0}{T-t}\right) \rangle \right\|_{L^2(\mathbb{P}; \mathbb{R})} \end{aligned}$$

$$\begin{aligned}
& + \frac{1}{\sqrt{M^m}} \left\| \frac{(F(0))(\mathcal{R}_t^0, x + W_{\mathcal{R}_t^0}^0 - W_t^0)}{\varrho(t, \mathcal{R}_t^0)} \langle \mathbf{e}_v, \left(1, \frac{W_{\mathcal{R}_t^0}^0 - W_t^0}{\mathcal{R}_t^0 - t}\right) \rangle \right\|_{L^2(\mathbb{P}; \mathbb{R})} \\
& + \sum_{l=1}^{m-1} \frac{1}{\sqrt{M^{m-l}}} \left\| \sum_{v_1=1}^{d+1} L_{v_1} \frac{|(\mathbf{U}_{l,M}^{(0,l,1),v_1} - \mathbf{U}_{l-1,M}^{(0,-l,1),v_1})(\mathcal{R}_t^{(0,l,1)}, x + W_{\mathcal{R}_t^{(0,l,1)}}^0 - W_t^{(0,l,1)})|}{\varrho(t, \mathcal{R}_t^{(0,l,1)})} \right. \\
& \cdot \left. \langle \mathbf{e}_v, \left(1, \frac{W_{\mathcal{R}_t^{(0,l,1)}}^0 - W_t^{(0,l,1)}}{\mathcal{R}_t^{(0,l,1)} - t}\right) \rangle \right\|_{L^2(\mathbb{P}; \mathbb{R})}. \tag{66}
\end{aligned}$$

This and the triangle inequality ensure that for all  $m \in \mathbb{N}$ ,  $x \in \mathbb{R}^d$ ,  $t \in [0, T)$ ,  $v \in \{1, 2, \dots, d+1\}$  it holds that

$$\begin{aligned}
& \left\| \mathbf{U}_{m,M}^{0,v}(t, x) - \mathbb{E}[\mathbf{U}_{m,M}^{0,v}(t, x)] \right\|_{L^2(\mathbb{P}; \mathbb{R})} \\
& \leq \frac{1}{\sqrt{M^m}} \left\| (g(x + W_T^0 - W_t^0) - g(x)) \langle \mathbf{e}_v, \left(1, \frac{W_T^0 - W_t^0}{T-t}\right) \rangle \right\|_{L^2(\mathbb{P}; \mathbb{R})} \\
& + \frac{1}{\sqrt{M^m}} \left\| \frac{(F(0))(\mathcal{R}_t^0, x + W_{\mathcal{R}_t^0}^0 - W_t^0)}{\varrho(t, \mathcal{R}_t^0)} \langle \mathbf{e}_v, \left(1, \frac{W_{\mathcal{R}_t^0}^0 - W_t^0}{\mathcal{R}_t^0 - t}\right) \rangle \right\|_{L^2(\mathbb{P}; \mathbb{R})} \\
& + \sum_{l=1}^{m-1} \sum_{v_1=1}^{d+1} \frac{L_{v_1}}{\sqrt{M^{m-l}}} \left\| \frac{(\mathbf{U}_{l,M}^{0,v_1} - \mathbf{u})(\mathcal{R}_t^0, x + W_{\mathcal{R}_t^0}^0 - W_t^0)}{\varrho(t, \mathcal{R}_t^0)} \langle \mathbf{e}_v, \left(1, \frac{W_{\mathcal{R}_t^0}^0 - W_t^0}{\mathcal{R}_t^0 - t}\right) \rangle \right\|_{L^2(\mathbb{P}; \mathbb{R})} \\
& + \sum_{l=1}^{m-1} \sum_{v_1=1}^{d+1} \frac{L_{v_1}}{\sqrt{M^{m-l}}} \left\| \frac{(\mathbf{U}_{l-1,M}^{0,v_1} - \mathbf{u})(\mathcal{R}_t^0, x + W_{\mathcal{R}_t^0}^0 - W_t^0)}{\varrho(t, \mathcal{R}_t^0)} \langle \mathbf{e}_v, \left(1, \frac{W_{\mathcal{R}_t^0}^0 - W_t^0}{\mathcal{R}_t^0 - t}\right) \rangle \right\|_{L^2(\mathbb{P}; \mathbb{R})} \\
& = \frac{1}{\sqrt{M^m}} \left\| (g(x + W_T^0 - W_t^0) - g(x)) \langle \mathbf{e}_v, \left(1, \frac{W_T^0 - W_t^0}{T-t}\right) \rangle \right\|_{L^2(\mathbb{P}; \mathbb{R})} \\
& + \frac{1}{\sqrt{M^m}} \left\| \frac{(F(0))(\mathcal{R}_t^0, x + W_{\mathcal{R}_t^0}^0 - W_t^0)}{\varrho(t, \mathcal{R}_t^0)} \langle \mathbf{e}_v, \left(1, \frac{W_{\mathcal{R}_t^0}^0 - W_t^0}{\mathcal{R}_t^0 - t}\right) \rangle \right\|_{L^2(\mathbb{P}; \mathbb{R})} \\
& + \sum_{l=0}^{m-1} \sum_{v_1=1}^{d+1} \frac{L_{v_1}}{\sqrt{M^{m-l-1}}} \left( \frac{\mathbb{I}_{(0,m)}(l)}{\sqrt{M}} + \mathbb{I}_{(-1,m-1)}(l) \right) \\
& \cdot \left\| \frac{(\mathbf{U}_{l,M}^{0,v_1} - \mathbf{u})(\mathcal{R}_t^0, x + W_{\mathcal{R}_t^0}^0 - W_t^0)}{\varrho(t, \mathcal{R}_t^0)} \langle \mathbf{e}_v, \left(1, \frac{W_{\mathcal{R}_t^0}^0 - W_t^0}{\mathcal{R}_t^0 - t}\right) \rangle \right\|_{L^2(\mathbb{P}; \mathbb{R})}. \tag{67}
\end{aligned}$$

Next we analyze the *time discretization error*. Item (ii) of Lemma 3.3 ensures that for all  $m \in \mathbb{N}$ ,  $t \in [0, T)$ ,  $x \in \mathbb{R}^d$  it holdsthat

$$\begin{aligned} & \mathbb{E} \left[ \mathbf{U}_{m,M}^0(t, x) - g(x + W_T^0 - W_t^0) \left( 1, \frac{W_T^0 - W_t^0}{T-t} \right) \right] \\ &= \mathbb{E} \left[ \int_t^T \left( F(\mathbf{U}_{m-1,M}^0) \right) (s, x + W_s^0 - W_t^0) \left( 1, \frac{W_s^0 - W_t^0}{s-t} \right) ds \right]. \end{aligned} \quad (68)$$

Combining this, (38), (37), Item (i) of Lemma 3.3, and Jensen's inequality shows that for all  $m \in \mathbb{N}$ ,  $t \in [0, T)$ ,  $x \in \mathbb{R}^d$ ,  $v \in \{1, 2, \dots, d+1\}$  it holds that

$$\begin{aligned} & \left| \mathbb{E} \left[ \mathbf{U}_{m,M}^{0,v}(t, x) \right] - \mathbf{u}_v(t, x) \right| \\ &= \left| \mathbb{E} \left[ \int_t^T \left( F(\mathbf{U}_{m-1,M}^0) - F(\mathbf{u}) \right) (s, x + W_s^0 - W_t^0) \langle \mathbf{e}_v, \left( 1, \frac{W_s^0 - W_t^0}{s-t} \right) \rangle ds \right] \right| \\ &\leq \sum_{v_1=1}^{d+1} L_{v_1} \mathbb{E} \left[ \int_t^T \left| \left( \mathbf{U}_{m-1,M}^{0,v_1} - \mathbf{u}_{v_1} \right) (s, x + W_s^0 - W_t^0) \right| \left| \langle \mathbf{e}_v, \left( 1, \frac{W_s^0 - W_t^0}{s-t} \right) \rangle \right| ds \right] \\ &= \sum_{v_1=1}^{d+1} L_{v_1} \mathbb{E} \left[ \int_t^T \frac{\left| \left( \mathbf{U}_{m-1,M}^{0,v_1} - \mathbf{u}_{v_1} \right) (s, x + W_s^0 - W_t^0) \right|}{\varrho(t,s)} \left| \langle \mathbf{e}_v, \left( 1, \frac{W_s^0 - W_t^0}{s-t} \right) \rangle \right| \varrho(t,s) ds \right] \quad (69) \\ &= \sum_{v_1=1}^{d+1} L_{v_1} \mathbb{E} \left[ \frac{\left| \left( \mathbf{U}_{m-1,M}^{0,v_1} - \mathbf{u}_{v_1} \right) (\mathcal{R}_t^0, x + W_{\mathcal{R}_t^0}^0 - W_t^0) \right|}{\varrho(t,\mathcal{R}_t^0)} \left| \langle \mathbf{e}_v, \left( 1, \frac{W_{\mathcal{R}_t^0}^0 - W_t^0}{\mathcal{R}_t^0 - t} \right) \rangle \right| \right] \\ &\leq \sum_{v_1=1}^{d+1} L_{v_1} \left\| \frac{\left( \mathbf{U}_{m-1,M}^{0,v_1} - \mathbf{u}_{v_1} \right) (\mathcal{R}_t^0, x + W_{\mathcal{R}_t^0}^0 - W_t^0)}{\varrho(t,\mathcal{R}_t^0)} \langle \mathbf{e}_v, \left( 1, \frac{W_{\mathcal{R}_t^0}^0 - W_t^0}{\mathcal{R}_t^0 - t} \right) \rangle \right\|_{L^2(\mathbb{P};\mathbb{R})}. \end{aligned}$$

In the next step we combine the established bounds for the Monte Carlo error (see (67) above) and the time discretization error (see (69) above) to obtain a suitable bound for the overall approximation error. More formally, observe that (67) and (69) ensure that for all  $m \in \mathbb{N}$ ,  $t \in [0, T)$ ,  $x \in \mathbb{R}^d$ ,  $v \in \{1, 2, \dots, d+1\}$  it holds that

$$\begin{aligned} & \left\| \mathbf{U}_{m,M}^{0,v}(t, x) - \mathbf{u}_v(t, x) \right\|_{L^2(\mathbb{P};\mathbb{R})} \\ &\leq \left\| \mathbf{U}_{m,M}^{0,v}(t, x) - \mathbb{E} \left[ \mathbf{U}_{m,M}^{0,v}(t, x) \right] \right\|_{L^2(\mathbb{P};\mathbb{R})} + \left| \mathbb{E} \left[ \mathbf{U}_{m,M}^{0,v}(t, x) \right] - \mathbf{u}_v(t, x) \right| \\ &\leq \frac{1}{\sqrt{M^m}} \left\| \left( g(x + W_T^0 - W_t^0) - g(x) \right) \langle \mathbf{e}_v, \left( 1, \frac{W_T^0 - W_t^0}{T-t} \right) \rangle \right\|_{L^2(\mathbb{P};\mathbb{R})} \\ &\quad + \frac{1}{\sqrt{M^m}} \left\| \frac{(F(0)) \mathcal{R}_t^0, x + W_{\mathcal{R}_t^0}^0 - W_t^0}{\varrho(t,\mathcal{R}_t^0)} \langle \mathbf{e}_v, \left( 1, \frac{W_{\mathcal{R}_t^0}^0 - W_t^0}{\mathcal{R}_t^0 - t} \right) \rangle \right\|_{L^2(\mathbb{P};\mathbb{R})} \\ &\quad + \sum_{l=0}^{m-1} \sum_{v_1=1}^{d+1} \frac{L_{v_1}}{\sqrt{M^{m-l-1}}} \left( \frac{\mathbb{I}_{(0,m)}(l)}{\sqrt{M}} + \mathbb{I}_{(-1,m-1)}(l) \right) \\ &\quad \cdot \left\| \frac{\left( \mathbf{U}_{l,M}^{0,v_1} - \mathbf{u}_{v_1} \right) (\mathcal{R}_t^0, x + W_{\mathcal{R}_t^0}^0 - W_t^0)}{\varrho(t,\mathcal{R}_t^0)} \langle \mathbf{e}_v, \left( 1, \frac{W_{\mathcal{R}_t^0}^0 - W_t^0}{\mathcal{R}_t^0 - t} \right) \rangle \right\|_{L^2(\mathbb{P};\mathbb{R})} \end{aligned}$$

$$+ \sum_{v_1=1}^{d+1} L_{v_1} \left\| \frac{(\mathbf{U}_{m-1,M}^{0,v_1} - \mathbf{u}_{v_1})(\mathcal{R}_t^0, x + W_{\mathcal{R}_t^0}^0 - W_t^0)}{\varrho(t, \mathcal{R}_t^0)} \langle \mathbf{e}_{v_1}, \left(1, \frac{W_{\mathcal{R}_t^0}^0 - W_t^0}{\mathcal{R}_t^0 - t}\right) \rangle \right\|_{L^2(\mathbb{P}; \mathbb{R})}. \quad (70)$$

This shows that for all  $m \in \mathbb{N}$ ,  $t \in [0, T)$ ,  $x \in \mathbb{R}^d$ ,  $v \in \{1, 2, \dots, d+1\}$  it holds that

$$\begin{aligned} & \left\| \mathbf{U}_{m,M}^{0,v}(t, x) - \mathbf{u}_v(t, x) \right\|_{L^2(\mathbb{P}; \mathbb{R})} \leq \\ & \frac{1}{\sqrt{M^m}} \left\| \left( g(x + W_T^0 - W_t^0) - g(x) \right) \langle \mathbf{e}_v, \left(1, \frac{W_T^0 - W_t^0}{T-t}\right) \rangle \right\|_{L^2(\mathbb{P}; \mathbb{R})} \\ & + \frac{1}{\sqrt{M^m}} \left\| \frac{(F(0))(\mathcal{R}_t^0, x + W_{\mathcal{R}_t^0}^0 - W_t^0)}{\varrho(t, \mathcal{R}_t^0)} \langle \mathbf{e}_v, \left(1, \frac{W_{\mathcal{R}_t^0}^0 - W_t^0}{\mathcal{R}_t^0 - t}\right) \rangle \right\|_{L^2(\mathbb{P}; \mathbb{R})} \\ & + \sum_{v_1=1}^{d+1} \frac{L_{v_1}}{\sqrt{M^{m-1}}} \left\| \frac{\mathbf{u}_{v_1}(\mathcal{R}_t^0, x + W_{\mathcal{R}_t^0}^0 - W_t^0)}{\varrho(t, \mathcal{R}_t^0)} \langle \mathbf{e}_v, \left(1, \frac{W_{\mathcal{R}_t^0}^0 - W_t^0}{\mathcal{R}_t^0 - t}\right) \rangle \right\|_{L^2(\mathbb{P}; \mathbb{R})} \\ & + \sum_{l=1}^{m-1} \sum_{v_1=1}^{d+1} \frac{2L_{v_1}}{\sqrt{M^{m-l-1}}} \left\| \frac{(\mathbf{U}_{l,M}^{0,v_1} - \mathbf{u}_{v_1})(\mathcal{R}_t^0, x + W_{\mathcal{R}_t^0}^0 - W_t^0)}{\varrho(t, \mathcal{R}_t^0)} \langle \mathbf{e}_v, \left(1, \frac{W_{\mathcal{R}_t^0}^0 - W_t^0}{\mathcal{R}_t^0 - t}\right) \rangle \right\|_{L^2(\mathbb{P}; \mathbb{R})}. \end{aligned} \quad (71)$$

In the next step we iterate (71). More formally, we claim that for all  $n, k \in \mathbb{N}$ ,  $t \in [0, T)$ ,  $x \in \mathbb{R}^d$ ,  $v_0 \in \{1, 2, \dots, d+1\}$  it holds that

$$\begin{aligned} & \left\| \mathbf{U}_{n,M}^{0,v_0}(t, x) - \mathbf{u}_{v_0}(t, x) \right\|_{L^2(\mathbb{P}; \mathbb{R})} \\ & \leq \sum_{j=0}^{k-1} \sum_{\substack{l_1, l_2, \dots, l_{j+1} \in \mathbb{N}, \\ l_1 < l_2 < \dots < l_j < l_{j+1} = n}} \sum_{\substack{v_1, v_2, \dots, v_{j+1} \in \\ \{1, 2, \dots, d+1\}}} \frac{1_{\{1\}}(v_{j+1}) 2^j [\prod_{i=1}^j L_{v_i}]}{\sqrt{M^{n-j}}} \\ & \cdot \left\| (g(x + W_T^0 - W_t^0) - g(x + W_{S(j,t)}^0 - W_t^0)) \right. \\ & \cdot \left. \langle \mathbf{e}_{v_j}, \left(1, \frac{W_T^0 - W_{S(j,t)}^0}{T - S(j,t)}\right) \right] \left[ \prod_{i=1}^j \frac{1}{\varrho(S(i-1,t), S(i,t))} \langle \mathbf{e}_{v_{i-1}}, \left(1, \frac{W_{S(i,t)}^0 - W_{S(i-1,t)}^0}{S(i,t) - S(i-1,t)}\right) \rangle \right] \right\|_{L^2(\mathbb{P}; \mathbb{R})} \\ & + \sum_{j=0}^{k-1} \sum_{\substack{l_1, l_2, \dots, l_{j+1} \in \mathbb{N}, \\ l_1 < l_2 < \dots < l_j < l_{j+1} = n}} \sum_{\substack{v_1, v_2, \dots, v_{j+1} \in \\ \{1, 2, \dots, d+1\}}} \frac{1_{\{1\}}(v_{j+1}) 2^j [\prod_{i=1}^j L_{v_i}]}{\sqrt{M^{n-j}}} \\ & \cdot \left\| (F(0))(S(j+1,t), x + W_{S(j+1,t)}^0 - W_t^0) \right. \\ & \cdot \left. \left[ \prod_{i=1}^{j+1} \frac{1}{\varrho(S(i-1,t), S(i,t))} \langle \mathbf{e}_{v_{i-1}}, \left(1, \frac{W_{S(i,t)}^0 - W_{S(i-1,t)}^0}{S(i,t) - S(i-1,t)}\right) \rangle \right] \right\|_{L^2(\mathbb{P}; \mathbb{R})} \\ & + \sum_{j=0}^{k-1} \sum_{\substack{l_1, l_2, \dots, l_{j+1} \in \mathbb{N}, \\ l_1 < l_2 < \dots < l_j < l_{j+1} = n}} \sum_{\substack{v_1, v_2, \dots, v_{j+1} \in \\ \{1, 2, \dots, d+1\}}} \frac{2^j [\prod_{i=1}^{j+1} L_{v_i}]}{\sqrt{M^{n-j-1}}} \left\| \mathbf{u}_{v_{j+1}}(S(j+1,t), x + W_{S(j+1,t)}^0 - W_t^0) \right. \\ & \cdot \left. \left[ \prod_{i=1}^{j+1} \frac{1}{\varrho(S(i-1,t), S(i,t))} \langle \mathbf{e}_{v_{i-1}}, \left(1, \frac{W_{S(i,t)}^0 - W_{S(i-1,t)}^0}{S(i,t) - S(i-1,t)}\right) \rangle \right] \right\|_{L^2(\mathbb{P}; \mathbb{R})} \end{aligned}$$

$$\begin{aligned}
& + \sum_{l_1, l_2, \dots, l_k \in \mathbb{N}} \sum_{v_1, v_2, \dots, v_k \in \{1, 2, \dots, d+1\}} \frac{2^k [\prod_{i=1}^k L_{v_i}]}{\sqrt{M^{n-k-l_1}}} \left\| \left( \mathbf{U}_{l_1, M}^{0, v_k} - \mathbf{u}_{v_k} \right) (S(k, t), x + W_{S(k, t)}^0 - W_t^0) \right. \\
& \cdot \left. \left[ \prod_{i=1}^k \frac{1}{\varrho(S(i-1, t), S(i, t))} \langle \mathbf{e}_{v_{i-1}}, \left( 1, \frac{W_{S(i, t)}^0 - W_{S(i-1, t)}^0}{S(i, t) - S(i-1, t)} \right) \rangle \right] \right\|_{L^2(\mathbb{P}; \mathbb{R})}. \tag{72}
\end{aligned}$$

We now prove (72) by induction on  $k \in \mathbb{N}$ . Observe that (71) establishes (72) in the base case  $k = 1$ . For the induction step  $\mathbb{N} \ni k \rightsquigarrow k+1 \in \mathbb{N}$  assume that there exists  $k \in \mathbb{N}$  which satisfies that (72) holds for  $k$ . Observe that Item (iv) of Lemma 3.2 shows that for all  $m \in \mathbb{N}$  it holds that  $\mathbf{U}_{m, M}^0$ ,  $W^0$ , and  $(\mathbf{r}^{(k)})_{k \in \mathbb{N}_0}$  are independent. Combining this and (71) yields that for all  $l_1 \in \mathbb{N}$ ,  $x \in \mathbb{R}^d$ ,  $t \in [0, T]$ ,  $v_0, v_1, \dots, v_k \in \{1, 2, \dots, d+1\}$  it holds that

$$\begin{aligned}
& \left\| \left( \mathbf{U}_{l_1, M}^{0, v_k} - \mathbf{u}_{v_k} \right) (S(k, t), x + W_{S(k, t)}^0 - W_t^0) \right. \\
& \cdot \left. \prod_{i=1}^k \frac{1}{\varrho(S(i-1, t), S(i, t))} \langle \mathbf{e}_{v_{i-1}}, \left( 1, \frac{W_{S(i, t)}^0 - W_{S(i-1, t)}^0}{S(i, t) - S(i-1, t)} \right) \rangle \right\|_{L^2(\mathbb{P}; \mathbb{R})} \\
& = \left\| \left\| \left( \mathbf{U}_{l_1, M}^{0, v_k} - \mathbf{u}_{v_k} \right) (s, y) \right\|_{L^2(\mathbb{P}; \mathbb{R})} \right. \Big|_{(s, y) = (S(k, t), x + W_{S(k, t)}^0 - W_t^0)} \\
& \cdot \left. \prod_{i=1}^k \frac{1}{\varrho(S(i-1, t), S(i, t))} \langle \mathbf{e}_{v_{i-1}}, \left( 1, \frac{W_{S(i, t)}^0 - W_{S(i-1, t)}^0}{S(i, t) - S(i-1, t)} \right) \rangle \right\|_{L^2(\mathbb{P}; \mathbb{R})} \\
& \leq \frac{1}{\sqrt{M^{l_1}}} \left\| \left( g(x + W_T^0 - W_t^0) - g(x + W_{S(k, t)}^0 - W_t^0) \right) \langle \mathbf{e}_{v_k}, \left( 1, \frac{W_T^0 - W_{S(k, t)}^0}{T - S(k, t)} \right) \rangle \right. \\
& \cdot \left. \prod_{i=1}^k \frac{\langle \mathbf{e}_{v_{i-1}}, \left( 1, \frac{W_{S(i, t)}^0 - W_{S(i-1, t)}^0}{S(i, t) - S(i-1, t)} \right) \rangle}{\varrho(S(i-1, t), S(i, t))} \right\|_{L^2(\mathbb{P}; \mathbb{R})} \\
& + \frac{1}{\sqrt{M^{l_1}}} \left\| (F(0)) (S(k+1, t), x + W_{S(k+1, t)}^0 - W_t^0) \right. \\
& \cdot \left. \prod_{i=1}^{k+1} \frac{\langle \mathbf{e}_{v_{i-1}}, \left( 1, \frac{W_{S(i, t)}^0 - W_{S(i-1, t)}^0}{S(i, t) - S(i-1, t)} \right) \rangle}{\varrho(S(i-1, t), S(i, t))} \right\|_{L^2(\mathbb{P}; \mathbb{R})} \\
& + \sum_{v_{k+1}=1}^{d+1} \frac{L_{v_{k+1}}}{\sqrt{M^{l_1-1}}} \left\| \mathbf{u}_{v_{k+1}} (S(k+1, t), x + W_{S(k+1, t)}^0 - W_t^0) \right. \\
& \cdot \left. \prod_{i=1}^{k+1} \frac{\langle \mathbf{e}_{v_{i-1}}, \left( 1, \frac{W_{S(i, t)}^0 - W_{S(i-1, t)}^0}{S(i, t) - S(i-1, t)} \right) \rangle}{\varrho(S(i-1, t), S(i, t))} \right\|_{L^2(\mathbb{P}; \mathbb{R})} \\
& + \sum_{l_0=1}^{l_1-1} \sum_{v_{k+1}=1}^{d+1} \frac{2L_{v_{k+1}}}{\sqrt{M^{l_1-l_0-1}}} \left\| (\mathbf{U}_{l_0, M}^{0, v_{k+1}} - \mathbf{u}_{v_{k+1}}) (S(k+1, t), x + W_{S(k+1, t)}^0 - W_t^0) \right. \\
& \cdot \left. \prod_{i=1}^{l_1-1} \frac{\langle \mathbf{e}_{v_{i-1}}, \left( 1, \frac{W_{S(i, t)}^0 - W_{S(i-1, t)}^0}{S(i, t) - S(i-1, t)} \right) \rangle}{\varrho(S(i-1, t), S(i, t))} \right\|_{L^2(\mathbb{P}; \mathbb{R})}.
\end{aligned}$$

$$\cdot \prod_{i=1}^{k+1} \frac{\langle \mathbf{e}_{v_{i-1}}, \left(1, \frac{W_{S(i,t)}^0 - W_{S(i-1,t)}^0}{\varrho(S(i-1,t), S(i,t))}\right) \rangle}{\varrho(S(i-1,t), S(i,t))} \Bigg\|_{L^2(\mathbb{P}; \mathbb{R})}. \quad (73)$$

This and the induction hypothesis establish (72) in the case  $k + 1$ . Induction thus proves (72). Next note that (72) (applied with  $k = n$  in the notation of (72)) demonstrates that for all  $n \in \mathbb{N}$ ,  $t \in [0, T]$ ,  $x \in \mathbb{R}^d$ ,  $v_0 \in \{1, 2, \dots, d+1\}$  it holds that

$$\begin{aligned} & \left\| \mathbf{U}_{n,M}^{0,v_0}(t, x) - \mathbf{u}_{v_0}(t, x) \right\|_{L^2(\mathbb{P}; \mathbb{R})} \\ & \leq \sum_{j=0}^{n-1} \sum_{\substack{l_1, l_2, \dots, l_{j+1} \in \mathbb{N}, \\ l_1 < l_2 < \dots < l_j < l_{j+1} = n}} \sum_{\substack{v_1, v_2, \dots, v_{j+1} \in \\ \{1, 2, \dots, d+1\}}} \frac{1_{\{1\}(v_{j+1})} 2^j [\prod_{i=1}^j L_{v_i}]}{\sqrt{M^{n-j}}} \\ & \quad \cdot \left\| \left( g(x + W_T^0 - W_t^0) - g(x + W_{S(j,t)}^0 - W_t^0) \right) \right. \\ & \quad \cdot \left. \langle \mathbf{e}_{v_j}, \left(1, \frac{W_T^0 - W_{S(j,t)}^0}{T - S(j,t)}\right) \rangle \left[ \prod_{i=1}^j \frac{1}{\varrho(S(i-1,t), S(i,t))} \langle \mathbf{e}_{v_{i-1}}, \left(1, \frac{W_{S(i,t)}^0 - W_{S(i-1,t)}^0}{S(i,t) - S(i-1,t)}\right) \rangle \right] \right\|_{L^2(\mathbb{P}; \mathbb{R})} \\ & \quad + \sum_{j=0}^{n-1} \sum_{\substack{l_1, l_2, \dots, l_{j+1} \in \mathbb{N}, \\ l_1 < l_2 < \dots < l_j < l_{j+1} = n}} \sum_{\substack{v_1, v_2, \dots, v_{j+1} \in \\ \{1, 2, \dots, d+1\}}} \frac{1_{\{1\}(v_{j+1})} 2^j [\prod_{i=1}^j L_{v_i}]}{\sqrt{M^{n-j}}} \\ & \quad \cdot \left\| (F(0))(S(j+1, t), x + W_{S(j+1,t)}^0 - W_t^0) \right. \\ & \quad \cdot \left. \left[ \prod_{i=1}^{j+1} \frac{1}{\varrho(S(i-1,t), S(i,t))} \langle \mathbf{e}_{v_{i-1}}, \left(1, \frac{W_{S(i,t)}^0 - W_{S(i-1,t)}^0}{S(i,t) - S(i-1,t)}\right) \rangle \right] \right\|_{L^2(\mathbb{P}; \mathbb{R})} \\ & \quad + \sum_{j=0}^{n-1} \sum_{\substack{l_1, l_2, \dots, l_{j+1} \in \mathbb{N}, \\ l_1 < l_2 < \dots < l_j < l_{j+1} = n}} \sum_{\substack{v_1, v_2, \dots, v_{j+1} \in \\ \{1, 2, \dots, d+1\}}} \frac{2^j [\prod_{i=1}^{j+1} L_{v_i}]}{\sqrt{M^{n-j-1}}} \\ & \quad \cdot \left\| \mathbf{u}_{v_{j+1}}(S(j+1, t), x + W_{S(j+1,t)}^0 - W_t^0) \right. \\ & \quad \cdot \left. \left[ \prod_{i=1}^{j+1} \frac{1}{\varrho(S(i-1,t), S(i,t))} \langle \mathbf{e}_{v_{i-1}}, \left(1, \frac{W_{S(i,t)}^0 - W_{S(i-1,t)}^0}{S(i,t) - S(i-1,t)}\right) \rangle \right] \right\|_{L^2(\mathbb{P}; \mathbb{R})}. \end{aligned} \quad (74)$$

Combining this with the fact that for all  $n \in \mathbb{N}$ ,  $j \in \{0, 1, \dots, n-1\}$  it holds that

$$\sum_{\substack{l_1, l_2, \dots, l_{j+1} \in \mathbb{N}, \\ l_1 < l_2 < \dots < l_j < l_{j+1} = n}} 1 = \binom{n-1}{j}. \quad (75)$$

establishes (59). The proof of Lemma 3.4 is thus completed.  $\square$

**Proposition 3.5** (Global approximation error) Assume Setting 3.1, let  $t \in [0, T)$ ,  $x \in \mathbb{R}^d$ ,  $v_0 \in \{1, 2, \dots, d+1\}$ ,  $M, n \in \mathbb{N}$ ,  $p \in [2, \infty)$ ,  $\alpha \in (\frac{p-2}{2(p-1)}, \frac{p}{2(p-1)})$ ,  $\beta = \frac{\alpha}{2} - \frac{(1-\alpha)(p-2)}{2p}$ ,  $C \in \mathbb{R}$  satisfy that

$$C = \max \left\{ 1, 2(T-t)^{\frac{1}{2}} |\Gamma(\frac{p}{2})|^{\frac{1}{p}} (1-\alpha)^{\frac{1}{p}-1} \max\{1, \|L\|_1\} \cdot \max \left\{ (T-t)^{\frac{1}{2}}, 2^{\frac{1}{2}} |\Gamma(\frac{p+1}{2})|^{\frac{1}{p}} \pi^{-\frac{1}{2p}} \right\} \right\}, \quad (76)$$

and assume for all  $s \in (0, 1)$  that  $\rho(s) = \frac{1-\alpha}{s^\alpha}$ . Then

$$\begin{aligned} \|\mathbf{U}_{n,M}^{0,v_0}(t, x) - \mathbf{u}_{v_0}(t, x)\|_{L^2(\mathbb{P}; \mathbb{R})} &\leq \frac{1}{4} \left[ 1 + \frac{pn}{2} \right]^{\frac{1}{8}} M^{-\frac{n}{2}} (2C)^n \\ &\cdot \exp \left( \frac{1}{8} + \beta M^{\frac{1}{2\beta}} \right) \left[ 2C^{-1} \sqrt{\max\{T-t, 3\}} \|K\|_1 \right. \\ &+ \sup_{s \in [t, T)} \left\| f(s, x + W_s^0 - W_t^0, 0, 0) \right\|_{L^{\frac{2p}{p-2}}(\mathbb{P}; \mathbb{R})} \\ &+ \sqrt{M} \sup_{s \in [t, T)} \max_{i \in \{1, 2, \dots, d+1\}} \left\| \mathbf{u}_i(s, x + W_s^0 - W_t^0) \right\|_{L^{\frac{2p}{p-2}}(\mathbb{P}; \mathbb{R})} \left. \right]. \end{aligned} \quad (77)$$

**Proof of Proposition 3.5** Throughout this proof assume without loss of generality that

$$\sup_{s \in [t, T)} \left\| (F(0))(s, x + W_s^0 - W_t^0) \right\|_{L^{\frac{2p}{p-2}}(\mathbb{P}; \mathbb{R})} < \infty \quad (78)$$

(otherwise (77) is clear), let  $S: \mathbb{N}_0 \times [0, T) \times \Omega \rightarrow [0, T)$  satisfy for all  $n \in \mathbb{N}_0$ ,  $t \in [0, T)$  that  $S(0, t) = t$  and  $S(n+1, t) = \mathcal{R}_{S(n,t)}^{(n)}$ , and let  $\mathfrak{C} \in \mathbb{R}$  satisfy

$$\mathfrak{C} = \max \left\{ T-t, 2|\Gamma(\frac{p+1}{2})|^{\frac{2}{p}} \pi^{-\frac{1}{p}} \right\} (T-t) |\Gamma(\frac{p}{2})|^{\frac{2}{p}} (1-\alpha)^{\frac{2}{p}-2}. \quad (79)$$

Observe that the triangle inequality ensures that for all  $v \in \{1, 2, \dots, d+1\}$ ,  $s \in [0, T)$  it holds that

$$\begin{aligned}
& \sum_{\alpha=1}^d K_\alpha \left\| (W_T^{0,\alpha} - W_s^{0,\alpha}) \langle \mathbf{e}_v, (1, \frac{W_T^0 - W_s^0}{T-s}) \rangle \right\|_{L^2(\mathbb{P}; \mathbb{R})} \\
&= \sum_{\alpha=1}^d K_\alpha \left( \sqrt{T-s} \mathbb{1}_{\{1\}}(v) + \frac{\mathbb{1}_{[2,\infty)}(v)}{T-s} \| (W_T^{0,\alpha} - W_s^{0,\alpha})(W_T^{0,v-1} - W_s^{0,v-1}) \|_{L^2(\mathbb{P}; \mathbb{R})} \right) \\
&= \sqrt{T-s} \|K\|_1 \mathbb{1}_{\{1\}}(v) + \frac{\mathbb{1}_{[2,\infty)}(v)}{T-s} \left( K_{v-1} \| (W_T^{0,v-1} - W_s^{0,v-1})^2 \|_{L^2(\mathbb{P}; \mathbb{R})} \right) \\
&\quad + \frac{\mathbb{1}_{[2,\infty)}(v)}{T-s} \left( \sum_{\alpha \in \{1, 2, \dots, d\} \setminus \{v-1\}} K_\alpha \|W_T^{0,\alpha} - W_s^{0,\alpha}\|_{L^2(\mathbb{P}; \mathbb{R})}^2 \right) \\
&= \sqrt{T-s} \|K\|_1 \mathbb{1}_{\{1\}}(v) + \mathbb{1}_{[2,\infty)}(v) \left( \sqrt{3} K_{v-1} + \sum_{\alpha \in \{1, 2, \dots, d\} \setminus \{v-1\}} K_\alpha \right) \\
&\leq \max\{\sqrt{T-s}, \sqrt{3}\} \|K\|_1.
\end{aligned} \tag{80}$$

Moreover, observe that (36) assures that for all  $j \in \{0, 1, \dots, n-1\}$ ,  $v_1, v_2, \dots, v_j \in \{1, 2, \dots, d+1\}$  it holds that

$$\begin{aligned}
& \left\| \left( g(x + W_T^0 - W_t^0) - g(x + W_{S(j,t)}^0 - W_t^0) \right) \langle \mathbf{e}_{v_j}, (1, \frac{W_T^0 - W_{S(j,t)}^0}{T-S(j,t)}) \rangle \right. \\
& \quad \cdot \left. \left[ \prod_{i=1}^j \frac{1}{\varrho(S(i-1,t), S(i,t))} \langle \mathbf{e}_{v_{i-1}}, (1, \frac{W_{S(i,t)}^0 - W_{S(i-1,t)}^0}{S(i,t) - S(i-1,t)}) \rangle \right] \right\|_{L^2(\mathbb{P}; \mathbb{R})} \\
& \leq \left\| \left( \sum_{\alpha=1}^d K_\alpha |W_T^{0,\alpha} - W_{S(j,t)}^{0,\alpha}| \right) \langle \mathbf{e}_{v_j}, (1, \frac{W_T^0 - W_{S(j,t)}^0}{T-S(j,t)}) \rangle \right. \\
& \quad \cdot \left. \left[ \prod_{i=1}^j \frac{1}{\varrho(S(i-1,t), S(i,t))} \langle \mathbf{e}_{v_{i-1}}, (1, \frac{W_{S(i,t)}^0 - W_{S(i-1,t)}^0}{S(i,t) - S(i-1,t)}) \rangle \right] \right\|_{L^2(\mathbb{P}; \mathbb{R})} \\
& = \left\| \left( \sum_{\alpha=1}^d K_\alpha |W_T^{0,\alpha} - W_s^{0,\alpha}| \right) \langle \mathbf{e}_{v_j}, (1, \frac{W_T^0 - W_s^0}{T-s}) \rangle \right\|_{L^2(\mathbb{P}; \mathbb{R})} \Big|_{s=S(j,t)} \\
& \quad \cdot \left[ \prod_{i=1}^j \frac{1}{\varrho(S(i-1,t), S(i,t))} \langle \mathbf{e}_{v_{i-1}}, (1, \frac{W_{S(i,t)}^0 - W_{S(i-1,t)}^0}{S(i,t) - S(i-1,t)}) \rangle \right] \Big\|_{L^2(\mathbb{P}; \mathbb{R})}.
\end{aligned} \tag{81}$$

Next note that the hypothesis that  $p \in [2, \infty)$  and the fact that  $\alpha \in (\frac{p-2}{2(p-1)}, \frac{p}{2(p-1)})$  imply that  $\frac{p}{2} \in [\alpha(p-1), \alpha(p-1)+1]$  and  $\alpha \in (0, 1)$ . This, (81), (80), and Corollary 2.5 (applied with  $p = 2$  in the notation of Corollary 2.5) show that for all

$j \in \{0, 1, \dots, n-1\}$ ,  $v_1, v_2, \dots, v_j \in \{1, 2, \dots, d+1\}$  it holds that

$$\begin{aligned}
& \left\| \left( g(x + W_T^0 - W_t^0) - g(x + W_{S(j,t)}^0 - W_t^0) \right) \langle \mathbf{e}_{v_j}, (1, \frac{W_T^0 - W_{S(j,t)}^0}{T - S(j,t)}) \right\| \\
& \quad \cdot \left[ \prod_{i=1}^j \frac{1}{\varrho(S(i-1,t), S(i,t))} \langle \mathbf{e}_{v_{i-1}}, (1, \frac{W_{S(i,t)}^0 - W_{S(i-1,t)}^0}{S(i,t) - S(i-1,t)}) \rangle \right] \Big\|_{L^2(\mathbb{P}; \mathbb{R})} \\
& \leq \left\| \left( \sum_{\alpha=1}^d K_\alpha \left\| \left( W_T^{0,\alpha} - W_s^{0,\alpha} \right) \langle \mathbf{e}_{v_j}, (1, \frac{W_T^0 - W_s^0}{T - s}) \rangle \right\|_{L^2(\mathbb{P}; \mathbb{R})} \Big|_{s=S(j,t)} \right) \right. \\
& \quad \cdot \left. \prod_{i=1}^j \left[ \frac{1}{\varrho(S(i-1,t), S(i,t))} \langle \mathbf{e}_{v_{i-1}}, (1, \frac{W_{S(i,t)}^0 - W_{S(i-1,t)}^0}{S(i,t) - S(i-1,t)}) \rangle \right] \right\|_{L^2(\mathbb{P}; \mathbb{R})} \quad (82) \\
& \leq \sqrt{\max\{T-t, 3\}} \|K\|_1 \left\| \prod_{i=1}^j \frac{1}{\varrho(S(i-1,t), S(i,t))} \langle \mathbf{e}_{v_{i-1}}, (1, \frac{W_{S(i,t)}^0 - W_{S(i-1,t)}^0}{S(i,t) - S(i-1,t)}) \rangle \right\|_{L^2(\mathbb{P}; \mathbb{R})} \\
& \leq \sqrt{\max\{T-t, 3\}} \|K\|_1 \left[ \frac{(T-t) \max\{T-t, 2\pi^{-1/2}\Gamma(\frac{3}{2})\}}{(1-\alpha)} \right]^{\frac{j}{2}} (ej)^{\frac{1}{8}} \left[ \frac{1}{\Gamma(j+1)} \right]^{\frac{\alpha}{2}}.
\end{aligned}$$

The fact that  $\Gamma(\frac{p+1}{2}) \geq \Gamma(\frac{3}{2}) = \frac{\sqrt{\pi}}{2}$ , the fact that  $\frac{2(p-1)}{p} \geq 1$ , and the fact that  $\alpha < 1$  prove that

$$(T-t)(1-\alpha)^{-1} \max\{T-t, 2\pi^{-1/2}\Gamma(\frac{3}{2})\} \leq \mathfrak{C}. \quad (83)$$

Moreover, note that the fact that  $p \geq 2$  ensures that for all  $j \in \mathbb{N}_0$  it holds that  $(ej)^{1/8} \leq \left[ e \left( \frac{pj}{2} + 1 \right) \right]^{1/8}$  and  $\Gamma(j+1)^{\alpha/2} \geq \Gamma(j+1)^\beta$ . Combining this with (82) and (83) proves that for all  $j \in \{0, 1, \dots, n-1\}$ ,  $v_1, v_2, \dots, v_j \in \{1, 2, \dots, d+1\}$  it holds that

$$\begin{aligned}
& \left\| \left( g(x + W_T^0 - W_t^0) - g(x + W_{S(j,t)}^0 - W_t^0) \right) \langle \mathbf{e}_{v_j}, (1, \frac{W_T^0 - W_{S(j,t)}^0}{T - S(j,t)}) \right\| \\
& \quad \cdot \left[ \prod_{i=1}^j \frac{1}{\varrho(S(i-1,t), S(i,t))} \langle \mathbf{e}_{v_{i-1}}, (1, \frac{W_{S(i,t)}^0 - W_{S(i-1,t)}^0}{S(i,t) - S(i-1,t)}) \rangle \right] \Big\|_{L^2(\mathbb{P}; \mathbb{R})} \quad (84) \\
& \leq \sqrt{\max\{T-t, 3\}} \|K\|_1 \mathfrak{C}^{\frac{j}{2}} \left[ e \left( \frac{pj}{2} + 1 \right) \right]^{\frac{1}{8}} \left[ \frac{1}{\Gamma(j+1)} \right]^\beta.
\end{aligned}$$

Corollary 2.5, the fact that  $p \geq 2$ , the fact that  $\alpha > \frac{p-2}{2(p-1)}$ , and the fact that  $\frac{2}{p} \Gamma(\frac{2}{p}) \leq 1$  prove that for all  $j \in \{0, 1, \dots, n-1\}$ ,  $v_1, v_2, \dots, v_j \in \{1, 2, \dots, d+1\}$  it holds

that

$$\begin{aligned}
& \left\| \prod_{i=1}^{j+1} \frac{1}{\varrho(S(i-1,t), S(i,t))} \langle \mathbf{e}_{v_{i-1}}, \left( 1, \frac{W_{S(i,t)}^0 - W_{S(i-1,t)}^0}{S(i,t) - S(i-1,t)} \right) \rangle \right\|_{L^p(\mathbb{P}; \mathbb{R})} \\
& \leq \left[ \max \left\{ T - t, 2 \frac{\Gamma(\frac{p+1}{2})^{\frac{2}{p}}}{\pi^{\frac{1}{p}}} \right\} \frac{(T-t)\Gamma(\frac{p}{2})^{\frac{2}{p}}}{(1-\alpha)\frac{2(p-1)}{p}(\frac{p}{2})^{\frac{2(\alpha(p-1)-\frac{p}{2}+1)}{p}}} \right]^{\frac{j+1}{2}} \\
& \quad \cdot \left[ e^{\frac{p}{2}} \left( \frac{pj}{2} + 1 \right) \right]^{\frac{1}{2p^2}} \left[ \frac{\Gamma(\frac{p}{2})}{\Gamma(1+j+\frac{2}{p})} \right]^{\frac{\alpha}{2} - \frac{(1-\alpha)(p-2)}{2p}} \\
& \leq \left[ \max \left\{ T - t, 2 \frac{\Gamma(\frac{p+1}{2})^{\frac{2}{p}}}{\pi^{\frac{1}{p}}} \right\} \frac{(T-t)\Gamma(\frac{p}{2})^{\frac{2}{p}}}{(1-\alpha)\frac{2(p-1)}{p}} \right]^{\frac{j+1}{2}} \left[ e \left( \frac{pj}{2} + 1 \right) \right]^{\frac{1}{8}} \left[ \frac{1}{\Gamma(1+j+\frac{2}{p})} \right]^{\frac{\alpha}{2} - \frac{(1-\alpha)(p-2)}{2p}} \\
& = \mathfrak{C}^{\frac{j+1}{2}} \left[ e \left( \frac{pj}{2} + 1 \right) \right]^{\frac{1}{8}} \left[ \frac{1}{\Gamma(1+j+\frac{2}{p})} \right]^{\beta}.
\end{aligned} \tag{85}$$

Combining this with Hölder's inequality and the fact that the random variables  $W^0$  and  $\tau^{(n)}$ ,  $n \in \mathbb{N}$ , are independent proves that for all  $j \in \{0, 1, \dots, n-1\}$ ,  $v_1, v_2, \dots, v_{j+1} \in \{1, 2, \dots, d+1\}$  it holds that

$$\begin{aligned}
& \left\| (F(0))(S(j+1,t), x + W_{S(j+1,t)}^0 - W_t^0) \prod_{i=1}^{j+1} \frac{1}{\varrho(S(i-1,t), S(i,t))} \langle \mathbf{e}_{v_{i-1}}, \left( 1, \frac{W_{S(i,t)}^0 - W_{S(i-1,t)}^0}{S(i,t) - S(i-1,t)} \right) \rangle \right\|_{L^2(\mathbb{P}; \mathbb{R})} \\
& \leq \left\| (F(0))(S(j+1,t), x + W_{S(j+1,t)}^0 - W_t^0) \right\|_{L^{\frac{2p}{p-2}}(\mathbb{P}; \mathbb{R})} \\
& \quad \cdot \left\| \prod_{i=1}^{j+1} \frac{1}{\varrho(S(i-1,t), S(i,t))} \langle \mathbf{e}_{v_{i-1}}, \left( 1, \frac{W_{S(i,t)}^0 - W_{S(i-1,t)}^0}{S(i,t) - S(i-1,t)} \right) \rangle \right\|_{L^p(\mathbb{P}; \mathbb{R})} \\
& = \left( \int_t^T \mathbb{E} \left[ \left\| (F(0))(s, x + W_s^0 - W_t^0) \right\|_{L^{\frac{2p}{p-2}}(\mathbb{P}; \mathbb{R})}^{\frac{2p}{p-2}} \right] \mathbb{P}(S(j+1,t) \in ds) \right)^{\frac{p-2}{2p}} \\
& \quad \cdot \left\| \prod_{i=1}^{j+1} \frac{1}{\varrho(S(i-1,t), S(i,t))} \langle \mathbf{e}_{v_{i-1}}, \left( 1, \frac{W_{S(i,t)}^0 - W_{S(i-1,t)}^0}{S(i,t) - S(i-1,t)} \right) \rangle \right\|_{L^p(\mathbb{P}; \mathbb{R})} \\
& \leq \sup_{s \in [t, T)} \left\| (F(0))(s, x + W_s^0 - W_t^0) \right\|_{L^{\frac{2p}{p-2}}(\mathbb{P}; \mathbb{R})} \left[ e \left( \frac{pj}{2} + 1 \right) \right]^{\frac{1}{8}} \mathfrak{C}^{\frac{j+1}{2}} \left[ \frac{1}{\Gamma(1+j+\frac{2}{p})} \right]^{\beta}.
\end{aligned} \tag{86}$$

Moreover, observe that for all  $j \in \{0, 1, \dots, n-1\}$ ,  $v_1, v_2, \dots, v_{j+1} \in \{1, 2, \dots, d+1\}$  it holds that

$$\begin{aligned}
& \left\| \mathbf{u}_{v_{j+1}}(S(j+1,t), x + W_{S(j+1,t)}^0 - W_t^0) \prod_{i=1}^{j+1} \frac{1}{\varrho(S(i-1,t), S(i,t))} \langle \mathbf{e}_{v_{i-1}}, \left( 1, \frac{W_{S(i,t)}^0 - W_{S(i-1,t)}^0}{S(i,t) - S(i-1,t)} \right) \rangle \right\|_{L^2(\mathbb{P}; \mathbb{R})} \\
& \leq \sup_{s \in [t, T)} \left\| \mathbf{u}_{v_{j+1}}(s, x + W_s^0 - W_t^0) \right\|_{L^{\frac{2p}{p-2}}(\mathbb{P}; \mathbb{R})} \left[ e \left( \frac{pj}{2} + 1 \right) \right]^{\frac{1}{8}} \mathfrak{C}^{\frac{j+1}{2}} \left[ \frac{1}{\Gamma(1+j+\frac{2}{p})} \right]^{\beta}.
\end{aligned} \tag{87}$$

In the next step we intend to apply Lemma 3.4. For this we now verify the hypotheses of Lemma 3.4. More formally, observe that the fact that  $\sup_{s \in [t, T]} \|(F(0))(s, x + W_s^0 - W_t^0)\|_{L^{2p/(p-2)}(\mathbb{P}; \mathbb{R})} < \infty$  and the fact that  $\frac{2p}{p-2} \geq 2$  ensure that for all  $r \in [1, 2)$  it holds that

$$\sup_{s \in [t, T]} \mathbb{E} \left[ \left| (F(0))(s, x + W_s^0 - W_t^0) \right|^r \right] < \infty. \quad (88)$$

Moreover, observe that for all  $r \in [1, 2)$  it holds that

$$\int_0^1 \frac{1}{s^{\frac{r}{2}} [\rho(s)]^{r-1}} ds = \frac{1}{(1-\alpha)^{r-1}} \int_0^1 s^{-(r(\frac{1}{2}-\alpha)+\alpha)} ds < \infty. \quad (89)$$

Combining this, (88), and Lemma 3.4 with (84), (86), and (87) proves that

$$\begin{aligned} & \| \mathbf{U}_{n,M}^{0,v_0}(t, x) - \mathbf{u}_{v_0}(t, x) \|_{L^2(\mathbb{P}; \mathbb{R})} \\ & \leq \sqrt{\max\{T-t, 3\}} \| K \|_1 \\ & \cdot \sum_{j=0}^{n-1} \sum_{\substack{v_1, v_2, \dots, v_{j+1} \in \\ \{1, 2, \dots, d+1\}}} \binom{n-1}{j} \frac{1_{\{1\}}(v_{j+1}) \left[ e \left( \frac{pj}{2} + 1 \right) \right]^{\frac{1}{8}} 2^j \mathfrak{C}^{\frac{j}{2}} \left[ \prod_{i=1}^j L_{v_i} \right]}{\sqrt{M^{n-j}} \left[ \Gamma(j+1) \right]^{\beta}} \\ & + \sup_{s \in [t, T]} \left\| (F(0))(s, x + W_s^0 - W_t^0) \right\|_{L^{\frac{2p}{p-2}}(\mathbb{P}; \mathbb{R})} \\ & \cdot \sum_{j=0}^{n-1} \sum_{\substack{v_1, v_2, \dots, v_{j+1} \in \\ \{1, 2, \dots, d+1\}}} \binom{n-1}{j} \frac{1_{\{1\}}(v_{j+1}) \left[ e \left( \frac{pj}{2} + 1 \right) \right]^{\frac{1}{8}} 2^j \mathfrak{C}^{\frac{j+1}{2}} \left[ \prod_{i=1}^j L_{v_i} \right]}{\sqrt{M^{n-j}} \left( \Gamma \left( 1+j+\frac{2}{p} \right) \right)^{\beta}} \\ & + \sum_{j=0}^{n-1} \sum_{\substack{v_1, v_2, \dots, v_{j+1} \in \\ \{1, 2, \dots, d+1\}}} \binom{n-1}{j} \sup_{s \in [t, T]} \left\| \mathbf{u}_{v_{j+1}}(s, x + W_s^0 - W_t^0) \right\|_{L^{\frac{2p}{p-2}}(\mathbb{P}; \mathbb{R})} \\ & \cdot \frac{\left[ e \left( \frac{pj}{2} + 1 \right) \right]^{\frac{1}{8}} 2^j \mathfrak{C}^{\frac{j+1}{2}} \left[ \prod_{i=1}^{j+1} L_{v_i} \right]}{\sqrt{M^{n-j-1}} \left( \Gamma \left( 1+j+\frac{2}{p} \right) \right)^{\beta}}. \end{aligned} \quad (90)$$

Moreover, observe that for all  $j \in \mathbb{N}_0$  it holds that

$$\left[ \sum_{\substack{v_1, v_2, \dots, v_{j+1} \in \\ \{1, 2, \dots, d+1\}}} 1_{\{1\}}(v_{j+1}) \prod_{i=1}^j L_{v_i} \right] = \| L \|_1^j. \quad (91)$$

Combining this with (90) proves that

$$\begin{aligned}
& \left\| \mathbf{U}_{n,M}^{0,v_0}(t, x) - \mathbf{u}_{v_0}(t, x) \right\|_{L^2(\mathbb{P}; \mathbb{R})} \\
& \leq \sqrt{\max\{T-t, 3\}} \|K\|_1 \sum_{j=0}^{n-1} \frac{\left[ e\left(\frac{pj}{2}+1\right) \right]^{\frac{1}{8}} 2^j \mathfrak{C}^{\frac{j}{2}} \|L\|_1^j \binom{n-1}{j} \left[ \frac{1}{\Gamma(j+1)} \right]^\beta}{\sqrt{M^{n-j}}} \\
& + \sup_{s \in [t, T)} \left\| (F(0))(s, x + W_s^0 - W_t^0) \right\|_{L^{\frac{2p}{p-2}}(\mathbb{P}; \mathbb{R})} \\
& \cdot \sum_{j=0}^{n-1} \frac{\left[ e\left(\frac{pj}{2}+1\right) \right]^{\frac{1}{8}} 2^j \|L\|_1^j \mathfrak{C}^{\frac{j+1}{2}} \binom{n-1}{j} \left[ \frac{1}{\Gamma(1+j+\frac{2}{p})} \right]^\beta}{\sqrt{M^{n-j}}} \\
& + \sup_{s \in [t, T), i \in \{1, 2, \dots, d+1\}} \left\| \mathbf{u}_i(s, x + W_s^0 - W_t^0) \right\|_{L^{\frac{2p}{p-2}}(\mathbb{P}; \mathbb{R})} \\
& \cdot \sum_{j=0}^{n-1} \frac{\left[ e\left(\frac{pj}{2}+1\right) \right]^{\frac{1}{8}} 2^j \|L\|_1^{j+1} \mathfrak{C}^{\frac{j+1}{2}} \binom{n-1}{j} \left[ \frac{1}{\Gamma(1+j+\frac{2}{p})} \right]^\beta}{\sqrt{M^{n-j-1}}}.
\end{aligned} \tag{92}$$

This, the fact that  $2\sqrt{\mathfrak{C}}\|L\|_1 \leq 2\sqrt{\mathfrak{C}} \max\{1, \|L\|_1\} \leq C$ , the fact that  $p \geq 2$ , and the fact that  $C \geq 1$  imply that

$$\begin{aligned}
& \left\| \mathbf{U}_{n,M}^{0,v_0}(t, x) - \mathbf{u}_{v_0}(t, x) \right\|_{L^2(\mathbb{P}; \mathbb{R})} \\
& \leq \sqrt{\max\{T-t, 3\}} \|K\|_1 \sum_{j=0}^{n-1} \frac{\left[ e\left(\frac{pj}{2}+1\right) \right]^{\frac{1}{8}} C^j \binom{n-1}{j} \left[ \frac{1}{\Gamma(j+1)} \right]^\beta}{\sqrt{M^{n-j}}} \\
& + \frac{1}{2} \sup_{s \in [t, T)} \left\| (F(0))(s, x + W_s^0 - W_t^0) \right\|_{L^{\frac{2p}{p-2}}(\mathbb{P}; \mathbb{R})} \\
& \cdot \sum_{j=0}^{n-1} \frac{\left[ e\left(\frac{pj}{2}+1\right) \right]^{\frac{1}{8}} C^{j+1} \binom{n-1}{j} \left[ \frac{1}{\Gamma(j+1)} \right]^\beta}{\sqrt{M^{n-j}}} \\
& + \frac{1}{2} \sup_{s \in [t, T), i \in \{1, 2, \dots, d+1\}} \left\| \mathbf{u}_i(s, x + W_s^0 - W_t^0) \right\|_{L^{\frac{2p}{p-2}}(\mathbb{P}; \mathbb{R})} \\
& \cdot \sum_{j=0}^{n-1} \frac{\left[ e\left(\frac{pj}{2}+1\right) \right]^{\frac{1}{8}} C^{j+1} \binom{n-1}{j} \left[ \frac{1}{\Gamma(j+1)} \right]^\beta}{\sqrt{M^{n-j-1}}} \leq \frac{\left[ e\left(\frac{pn}{2}+1\right) \right]^{\frac{1}{8}} C^{n-1}}{\sqrt{M^{n-1}}} \\
& \cdot \left[ \sum_{j=0}^{n-1} \binom{n-1}{j} \frac{(\sqrt{M})^j}{\Gamma(j+1)^\beta} \right] \left[ \frac{\sqrt{\max\{T-t, 3\}} \|K\|_1}{\sqrt{M}} + \frac{C \sup_{s \in [t, T)} \|(F(0))(s, x + W_s^0 - W_t^0)\|_{L^{\frac{2p}{p-2}}(\mathbb{P}; \mathbb{R})}}{2\sqrt{M}} \right]
\end{aligned}$$

$$+ \frac{C \sup_{s \in [t, T], i \in \{1, 2, \dots, d+1\}} \|u_i(s, x + W_s^0 - W_t^0)\|_{L^{\frac{2p}{p-2}}(\mathbb{P}; \mathbb{R})}}{2} \Big]. \quad (93)$$

Next note that for all  $j \in \mathbb{N}_0$  it holds that  $\sum_{i=0}^{n-1} \binom{n-1}{i} = 2^{n-1}$  and

$$\frac{(\sqrt{M})^j}{\Gamma(j+1)^\beta} = \left( \frac{(M^{\frac{1}{2\beta}})^j}{\Gamma(j+1)} \right)^\beta \leq \left( \sum_{k=0}^{\infty} \frac{(M^{\frac{1}{2\beta}})^k}{\Gamma(k+1)} \right)^\beta = \exp(\beta M^{\frac{1}{2\beta}}). \quad (94)$$

Combining this with (92) assures that

$$\begin{aligned} \left\| \mathbf{U}_{n,M}^{0,v_0}(t, x) - \mathbf{u}_{v_0}(t, x) \right\|_{L^2(\mathbb{P}; \mathbb{R})} &\leq \frac{\left[ e \left( \frac{pn}{2} + 1 \right) \right]^{\frac{1}{8}} (2C)^{n-1} \exp(\beta M^{\frac{1}{2\beta}})}{\sqrt{M^{n-1}}} \cdot \left[ \frac{\sqrt{\max\{T-t, 3\}} \|K\|_1}{\sqrt{M}} \right. \\ &+ \frac{C \sup_{s \in [t, T]} \|(F(0))(s, x + W_s^0 - W_t^0)\|_{L^{\frac{2p}{p-2}}(\mathbb{P}; \mathbb{R})}}{2\sqrt{M}} + \left. \frac{C \sup_{s \in [t, T], i \in \{1, 2, \dots, d+1\}} \|u_i(s, x + W_s^0 - W_t^0)\|_{L^{\frac{2p}{p-2}}(\mathbb{P}; \mathbb{R})}}{2} \right]. \end{aligned} \quad (95)$$

This establishes (77). The proof of Proposition 3.5 is thus completed.  $\square$

## 4 Regularity Analysis for Solutions of Certain Differential Equations

The error analysis in Sect. 3.3 above provides upper bounds for the approximation errors of the MLP approximations in (39). The established upper bounds contain certain norms of the unknown exact solutions of the PDEs which we intend to approximate; see, e.g., the right-hand side of (77) in Proposition 3.5 in Sect. 3.3 above for details. In Lemma 4.2 below we establish suitable upper bounds for these norms of the unknown exact solutions of the PDEs which we intend to approximate. In our proof of Lemma 4.2 we employ certain a priori estimates for solutions of BSDEs which we establish in the essentially well-known result in Lemma 4.1 below (see, e.g., El Karoui et al. [22, Proposition 2.1 and Equation (2.12)] for results related to Lemma 4.1).

### 4.1 Regularity Analysis for Solutions of Backward Stochastic Differential Equations (BSDEs)

**Lemma 4.1** *Let  $T \in (0, \infty)$ ,  $t \in [0, T]$ ,  $d \in \mathbb{N}$ ,  $L, \mathfrak{L} \in [0, \infty)$ , let  $\|\cdot\| : \mathbb{R}^d \rightarrow [0, \infty)$  be the standard norm on  $\mathbb{R}^d$ , let  $\langle \cdot, \cdot \rangle : \mathbb{R}^d \times \mathbb{R}^d \rightarrow [0, \infty)$  be the standard scalar product on  $\mathbb{R}^d$ , let  $(\Omega, \mathcal{F}, \mathbb{P})$  be a probability space with a normal filtration  $(\mathbb{F}_s)_{s \in [t, T]}$ , let  $f_1, f_2 : [t, T] \times \mathbb{R} \times \mathbb{R}^d \times \Omega \rightarrow \mathbb{R}$  be functions, assume for all  $s \in [t, T]$ ,  $i \in \{1, 2\}$  that the function  $[t, s] \times \mathbb{R} \times \mathbb{R}^d \times \Omega \ni (u, y, z, \omega) \mapsto f_i(u, y, z, \omega) \in \mathbb{R}$  is  $(\mathcal{B}([t, s]) \otimes \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R}^d) \otimes \mathbb{F}_s)/\mathcal{B}(\mathbb{R})$ -measurable, assume that for all  $s \in [t, T]$ ,  $y, \mathfrak{y} \in \mathbb{R}$ ,  $z, \mathfrak{z} \in \mathbb{R}^d$  it holds  $\mathbb{P}$ -a.s. that*

$$|f_1(s, y, z) - f_1(s, \mathfrak{y}, \mathfrak{z})| \leq L|y - \mathfrak{y}| + \mathfrak{L}\|z - \mathfrak{z}\|, \quad (96)$$

let  $Y^1, Y^2: [t, T] \times \Omega \rightarrow \mathbb{R}$  and  $W: [t, T] \times \Omega \rightarrow \mathbb{R}^d$  be  $(\mathbb{F}_s)_{s \in [t, T]}$ -adapted stochastic processes with continuous sample paths, assume that  $(W_{s+t} - W_t)_{s \in [0, T-t]}$  is a standard Brownian motion, let  $Z^k = (Z^{k,1}, Z^{k,2}, \dots, Z^{k,d}): [t, T] \times \Omega \rightarrow \mathbb{R}^d$ ,  $k \in \{1, 2\}$ , be  $(\mathbb{F}_s)_{s \in [t, T]}$ -adapted  $(\mathcal{B}([t, T]) \otimes \mathcal{F})/\mathcal{B}(\mathbb{R}^d)$ -measurable stochastic processes, assume that  $\sum_{i=1}^2 \mathbb{E}[\sup_{s \in [t, T]} |Y_s^i|^2] < \infty$  and  $\mathbb{P}(\sum_{i=1}^2 \int_t^T |f_i(s, Y_s^i, Z_s^i)| + \|Z_s^i\|^2 ds < \infty) = 1$ , and assume that for all  $s \in [t, T]$ ,  $i \in \{1, 2\}$  it holds  $\mathbb{P}$ -a.s. that

$$Y_s^i = Y_T^i + \int_s^T f_i(u, Y_u^i, Z_u^i) du - \int_s^T \langle Z_u^i, dW_u \rangle. \quad (97)$$

Then it holds  $\mathbb{P}$ -a.s. that

$$\begin{aligned} |Y_t^1 - Y_t^2| &\leq e^{L(T-t)} \left( \|Y_T^1 - Y_T^2\|_{L^\infty(\mathbb{P}; \mathbb{R})} \right. \\ &\quad \left. + (T-t) \sup_{s \in [t, T]} \sup_{y \in \mathbb{R}} \sup_{z \in \mathbb{R}^d} \|f_1(s, y, z) - f_2(s, y, z)\|_{L^\infty(\mathbb{P}; \mathbb{R})} \right). \end{aligned} \quad (98)$$

**Proof of Lemma 4.1** Throughout this proof assume without loss of generality that  $\sup_{s \in [t, T], y \in \mathbb{R}, z \in \mathbb{R}^d} \|f_1(s, y, z) - f_2(s, y, z)\|_{L^\infty(\mathbb{P}; \mathbb{R})} < \infty$ , let  $A: [t, T] \times \Omega \rightarrow \mathbb{R}$  and  $B = (B^1, B^2, \dots, B^d): [t, T] \times \Omega \rightarrow \mathbb{R}^d$  satisfy for all  $s \in [t, T]$ ,  $j \in \{1, 2, \dots, d\}$  that

$$A_s = \begin{cases} \frac{f_1(s, Y_s^1, Z_s^2) - f_1(s, Y_s^2, Z_s^2)}{Y_s^1 - Y_s^2} & : Y_s^1 \neq Y_s^2 \\ 0 & : Y_s^1 = Y_s^2 \end{cases} \quad (99)$$

and

$$B_s^j = \begin{cases} \frac{f_1(s, Y_s^1, Z_s^{1,1}, \dots, Z_s^{1,j}, Z_s^{2,j+1}, \dots, Z_s^{2,d}) - f_1(s, Y_s^1, Z_s^{1,1}, \dots, Z_s^{1,j-1}, Z_s^{2,j}, \dots, Z_s^{2,d})}{Z_s^{1,j} - Z_s^{2,j}} & : Z_s^{1,j} \neq Z_s^{2,j} \\ 0 & : Z_s^{1,j} = Z_s^{2,j} \end{cases}, \quad (100)$$

and let  $\Gamma: [t, T] \times \Omega \rightarrow \mathbb{R}$  be a stochastic process with continuous sample paths which satisfies that for all  $s \in [t, T]$  it holds  $\mathbb{P}$ -a.s. that  $\Gamma_s = e^{\int_t^s A_r - \frac{\|B_r\|^2}{2} dr + \int_t^s \langle B_r, dW_r \rangle}$ . Note that Itô's formula implies that for all  $s \in [t, T]$  it holds  $\mathbb{P}$ -a.s. that

$$\Gamma_s = 1 + \int_t^s \Gamma_r A_r dr + \int_t^s \Gamma_r \langle B_r, dW_r \rangle. \quad (101)$$

Next observe that (97) ensures that for all  $u \in [t, T]$ ,  $i \in \{1, 2\}$  it holds  $\mathbb{P}$ -a.s. that  $Y_u^i = Y_t^i - \int_t^u f_i(r, Y_r^i, Z_r^i) dr + \int_t^u \langle Z_r^i, dW_r \rangle$ . Combining this and (101) with Itô's

formula yields that for all  $u \in [t, T)$  it holds  $\mathbb{P}$ -a.s. that

$$\begin{aligned} \Gamma_u Y_u^1 - Y_t^1 &= \int_t^u \Gamma_s \left( -f_1(s, Y_s^1, Z_s^1) + A_s Y_s^1 + \langle B_s, Z_s^1 \rangle \right) ds \\ &\quad + \int_t^u \Gamma_s Y_s^1 \langle B_s, dW_s \rangle + \int_t^u \Gamma_s \langle Z_s^1, dW_s \rangle \end{aligned} \quad (102)$$

and

$$\begin{aligned} \Gamma_u Y_u^2 - Y_t^2 &= \int_t^u \Gamma_s \left( -f_2(s, Y_s^2, Z_s^2) + A_s Y_s^2 + \langle B_s, Z_s^2 \rangle \right) ds \\ &\quad + \int_t^u \Gamma_s Y_s^2 \langle B_s, dW_s \rangle + \int_t^u \Gamma_s \langle Z_s^2, dW_s \rangle. \end{aligned} \quad (103)$$

Moreover, note that (99) and (100) imply that for all  $s \in [t, T]$  it holds that

$$\begin{aligned} A_s(Y_s^1 - Y_s^2) + \langle B_s, Z_s^1 - Z_s^2 \rangle &= A_s(Y_s^1 - Y_s^2) + \sum_{j=1}^d B_s^j (Z_s^{1,j} - Z_s^{2,j}) \\ &= f_1(s, Y_s^1, Z_s^2) - f_1(s, Y_s^2, Z_s^2) \\ &\quad + \sum_{j=1}^d \left( f_1(s, Y_s^1, Z_s^{1,1}, \dots, Z_s^{1,j}, Z_s^{2,j+1}, \dots, Z_s^{2,d}) \right. \\ &\quad \left. - f_1(s, Y_s^1, Z_s^{1,1}, \dots, Z_s^{1,j-1}, Z_s^{2,j}, \dots, Z_s^{2,d}) \right) \\ &= f_1(s, Y_s^1, Z_s^2) - f_1(s, Y_s^2, Z_s^2) + f_1(s, Y_s^1, Z_s^1) - f_1(s, Y_s^1, Z_s^2) \\ &= f_1(s, Y_s^1, Z_s^1) - f_1(s, Y_s^2, Z_s^2) \end{aligned} \quad (104)$$

This, (102), and (103) assure that for all  $u \in [t, T]$  it holds  $\mathbb{P}$ -a.s. that

$$\begin{aligned} Y_t^1 - Y_t^2 - \Gamma_u(Y_u^1 - Y_u^2) + \int_t^u \Gamma_s(Y_s^1 - Y_s^2) \langle B_s, dW_s \rangle + \int_t^u \Gamma_s \langle Z_s^1 - Z_s^2, dW_s \rangle \\ = \int_t^u \Gamma_s \left( f_1(s, Y_s^1, Z_s^1) - f_2(s, Y_s^2, Z_s^2) - \left( A_s(Y_s^1 - Y_s^2) + \langle B_s, Z_s^1 - Z_s^2 \rangle \right) \right) ds \\ = \int_t^u \Gamma_s \left( f_1(s, Y_s^2, Z_s^2) - f_2(s, Y_s^2, Z_s^2) \right) ds. \end{aligned} \quad (105)$$

Next let  $\tau_n: \Omega \rightarrow [t, T)$ ,  $n \in \mathbb{N}$ , be the  $(\mathbb{F}_s)_{s \in [t, T]}$ -stopping times which satisfy for all  $n \in \mathbb{N}$  that

$$\begin{aligned} \tau_n = \inf \left( \left\{ r \in [t, T] : \left| \int_t^r \Gamma_s(Y_s^1 - Y_s^2) \langle B_s, dW_s \rangle \right| \right. \right. \\ \left. \left. + \left| \int_t^r \Gamma_s \langle Z_s^1 - Z_s^2, dW_s \rangle \right| \geq n \right\} \cup \{T - \frac{T}{n}\} \right). \end{aligned} \quad (106)$$

Note that (105) and (106) assure that for all  $n \in \mathbb{N}$  it holds  $\mathbb{P}$ -a.s. that

$$Y_t^1 - Y_t^2 = \mathbb{E} \left[ \Gamma_{\tau_n} (Y_{\tau_n}^1 - Y_{\tau_n}^2) + \int_t^{\tau_n} \Gamma_s \left( f_1(s, Y_s^2, Z_s^2) - f_2(s, Y_s^2, Z_s^2) \right) ds \middle| \mathbb{F}_t \right]. \quad (107)$$

Moreover, observe that (96) ensures that  $\sup_{s \in [t, T]} \|B_s\|^2 \leq \mathcal{L}^2 d$ . The fact that  $\mathbb{E}[e^{-\int_t^T \frac{\|2B_s\|^2}{2} ds + \int_t^T \langle 2B_s, dW_s \rangle}] = 1$  hence implies that

$$\begin{aligned} \mathbb{E} \left[ \left| e^{-\int_t^T \frac{\|B_s\|^2}{2} ds + \int_t^T \langle B_s, dW_s \rangle} \right|^2 \right] &= \mathbb{E} \left[ e^{\int_t^T \|B_s\|^2 ds} e^{-\int_t^T \frac{\|2B_s\|^2}{2} ds + \int_t^T \langle 2B_s, dW_s \rangle} \right] \\ &\leq e^{(T-t)\mathcal{L}^2 d} \mathbb{E} \left[ e^{-\int_t^T \frac{\|2B_s\|^2}{2} ds + \int_t^T \langle 2B_s, dW_s \rangle} \right] \quad (108) \\ &= e^{(T-t)\mathcal{L}^2 d}. \end{aligned}$$

In addition, observe that (96) demonstrates that it holds  $\mathbb{P}$ -a.s. that  $\sup_{s \in [t, T]} |A_s| \leq L$ . This, Doob's martingale inequality, and (108) show that

$$\begin{aligned} \mathbb{E} \left[ \sup_{s \in [t, T]} |\Gamma_s|^2 \right] &= \mathbb{E} \left[ \sup_{s \in [t, T]} e^{\int_t^s 2A_r dr} \left| e^{-\int_t^s \frac{\|B_r\|^2}{2} dr + \int_t^s \langle B_r, dW_r \rangle} \right|^2 \right] \\ &\leq e^{2L(T-t)} \mathbb{E} \left[ \sup_{s \in [t, T]} \left| e^{-\int_t^s \frac{\|B_r\|^2}{2} dr + \int_t^s \langle B_r, dW_r \rangle} \right|^2 \right] \quad (109) \\ &\leq 4e^{2L(T-t)} \mathbb{E} \left[ \left| e^{-\int_t^T \frac{\|B_r\|^2}{2} dr + \int_t^T \langle B_r, dW_r \rangle} \right|^2 \right] \leq 4e^{(T-t)(2L+\mathcal{L}^2 d)}. \end{aligned}$$

Combining this with the Cauchy–Schwarz inequality and the assumption that  $\sum_{i=1}^2 \mathbb{E}[\sup_{s \in [t, T]} |Y_s^i|^2] < \infty$  proves that

$$\mathbb{E} \left[ \sup_{s \in [t, T]} |\Gamma_s(Y_s^1 - Y_s^2)| \right] \leq \sqrt{\mathbb{E} \left[ \sup_{s \in [t, T]} |\Gamma_s|^2 \right] \mathbb{E} \left[ \sup_{s \in [t, T]} |Y_s^1 - Y_s^2|^2 \right]} < \infty. \quad (110)$$

Next observe that the fact that for all  $s \in [t, T]$  it holds that  $\Gamma_s \geq 0$  demonstrates that

$$\begin{aligned} \mathbb{E} \left[ \int_t^T \Gamma_s |f_1(s, Y_s^2, Z_s^2) - f_2(s, Y_s^2, Z_s^2)| ds \right] &\leq \mathbb{E} \left[ \int_t^T \Gamma_s ds \right] \left[ \sup_{s \in [t, T]} \sup_{y \in \mathbb{R}} \sup_{z \in \mathbb{R}^d} \|f_1(s, y, z) - f_2(s, y, z)\|_{L^\infty(\mathbb{P}; \mathbb{R})} \right] < \infty. \quad (111) \end{aligned}$$

This, (110), Lebesgue's dominated convergence theorem, (107), the fact that  $Y^1, Y^2$ , and  $\Gamma$  are stochastic processes with continuous sample paths, and the fact that it holds  $\mathbb{P}$ -a.s. that  $\lim_{n \rightarrow \infty} \tau_n = T$  ensure that it holds  $\mathbb{P}$ -a.s. that

$$\begin{aligned} Y_t^1 - Y_t^2 &= \lim_{n \rightarrow \infty} \mathbb{E} \left[ \Gamma_{\tau_n} (Y_{\tau_n}^1 - Y_{\tau_n}^2) + \int_t^{\tau_n} \Gamma_s (f_1(s, Y_s^2, Z_s^2) - f_2(s, Y_s^2, Z_s^2)) ds \middle| \mathbb{F}_t \right] \\ &= \mathbb{E} \left[ \Gamma_T (Y_T^1 - Y_T^2) + \int_t^T \Gamma_s (f_1(s, Y_s^2, Z_s^2) - f_2(s, Y_s^2, Z_s^2)) ds \middle| \mathbb{F}_t \right]. \end{aligned} \quad (112)$$

Next observe that the fact that  $\sup_{s \in [t, T]} |A_s| \leq L$  demonstrates that for all  $s \in [t, T]$  it holds that

$$\mathbb{E}[\Gamma_s | \mathbb{F}_t] \leq e^{L(s-t)} \mathbb{E} \left[ e^{- \int_t^s \frac{\|B_r\|^2}{2} dr + \int_t^s \langle B_r, dW_r \rangle} \middle| \mathbb{F}_t \right] = e^{L(s-t)}. \quad (113)$$

Combining this, (112), and the triangle inequality shows that it holds  $\mathbb{P}$ -a.s. that

$$\begin{aligned} |Y_t^1 - Y_t^2| &\leq \mathbb{E} \left[ \Gamma_T |Y_T^1 - Y_T^2| + \int_t^T \Gamma_s |f_1(s, Y_s^2, Z_s^2) - f_2(s, Y_s^2, Z_s^2)| ds \middle| \mathbb{F}_t \right] \\ &\leq \mathbb{E}[\Gamma_T | \mathbb{F}_t] \|Y_T^1 - Y_T^2\|_{L^\infty(\mathbb{P}; \mathbb{R})} + \mathbb{E} \left[ \int_t^T \Gamma_s ds \middle| \mathbb{F}_t \right] \\ &\quad \left[ \sup_{s \in [t, T]} \sup_{y \in \mathbb{R}} \sup_{z \in \mathbb{R}^d} \|f_1(s, y, z) - f_2(s, y, z)\|_{L^\infty(\mathbb{P}; \mathbb{R})} \right] \\ &\leq e^{L(T-t)} \left( \|Y_T^1 - Y_T^2\|_{L^\infty(\mathbb{P}; \mathbb{R})} + (T-t) \right. \\ &\quad \left. \sup_{s \in [t, T]} \sup_{y \in \mathbb{R}} \sup_{z \in \mathbb{R}^d} \|f_1(s, y, z) - f_2(s, y, z)\|_{L^\infty(\mathbb{P}; \mathbb{R})} \right). \end{aligned} \quad (114)$$

This proves (98). The proof of Lemma 4.1 is thus completed.  $\square$

## 4.2 Regularity Analysis for Solutions of Partial Differential Equations (PDEs)

**Lemma 4.2** Let  $T \in (0, \infty)$ ,  $d \in \mathbb{N}$ ,  $\eta, L_0, L_1, \dots, L_d, \mathfrak{L}_1, \mathfrak{L}_2, \dots, \mathfrak{L}_d, K_1, K_2, \dots, K_d \in \mathbb{R}$ ,  $f \in C([0, T] \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d, \mathbb{R})$ ,  $g \in C(\mathbb{R}^d, \mathbb{R})$ ,  $u = (u(t, x))_{(t,x) \in [0,T] \times \mathbb{R}^d} \in C^{1,2}([0, T] \times \mathbb{R}^d, \mathbb{R})$ , let  $\|\cdot\|: \mathbb{R}^{d+1} \rightarrow [0, \infty)$  be a norm, assume for all  $t \in [0, T]$ ,  $x = (x_1, x_2, \dots, x_d)$ ,  $\mathfrak{x} = (\mathfrak{x}_1, \mathfrak{x}_2, \dots, \mathfrak{x}_d)$ ,  $z = (z_1, z_2, \dots, z_d)$ ,  $\mathfrak{z} = (\mathfrak{z}_1, \mathfrak{z}_2, \dots, \mathfrak{z}_d) \in \mathbb{R}^d$ ,  $y, \mathfrak{y} \in \mathbb{R}$  that

$$|f(t, x, y, z) - f(t, \mathfrak{x}, \mathfrak{y}, \mathfrak{z})| \leq L_0 |y - \mathfrak{y}| + \sum_{j=1}^d (L_j |z_j - \mathfrak{z}_j| + \mathfrak{L}_j |x_j - \mathfrak{x}_j|), \quad (115)$$

$$|g(x) - g(\mathfrak{x})| \leq \sum_{i=1}^d K_i |x_i - \mathfrak{x}_i|, \quad |u(t, x)| \leq \eta \left[ 1 + \sum_{i=1}^d |x_i| \right]^\eta, \quad u(T, x) = g(x), \quad (116)$$

$$\text{and } \left( \frac{\partial}{\partial t} u \right)(t, x) + \frac{1}{2} (\Delta_x u)(t, x) + f(t, x, u(t, x), (\nabla_x u)(t, x)) = 0, \quad (117)$$

let  $(\Omega, \mathcal{F}, \mathbb{P})$  be a probability space, and let  $W: [0, T] \times \Omega \rightarrow \mathbb{R}^d$  be a standard Brownian motion. Then

(i) it holds for all  $s \in [0, T]$ ,  $x \in \mathbb{R}^d$  that

$$\begin{aligned} & \mathbb{E}\left[\|g(x + W_{T-s})(1, \frac{W_{T-s}}{T-s})\|\right] \\ & + \mathbb{E}\left[\int_s^T \|f(t, x + W_{t-s}, u(t, x + W_{t-s}), (\nabla_x u)(t, x + W_{t-s}))\|(1, \frac{W_{t-s}}{t-s}) dt\right] < \infty, \end{aligned} \quad (118)$$

(ii) it holds for all  $s \in [0, T]$ ,  $x \in \mathbb{R}^d$  that

$$\begin{aligned} (u(s, x), (\nabla_x u)(s, x)) = & \mathbb{E}\left[g(x + W_{T-s})(1, \frac{W_{T-s}}{T-s})\right] \\ & + \mathbb{E}\left[\int_s^T [f(t, x + W_{t-s}, u(t, x + W_{t-s}), (\nabla_x u)(t, x + W_{t-s}))](1, \frac{W_{t-s}}{t-s}) dt\right], \end{aligned} \quad (119)$$

(iii) it holds for all  $t \in [0, T]$ ,  $x = (x_1, x_2, \dots, x_d)$ ,  $\mathfrak{x} = (\mathfrak{x}_1, \mathfrak{x}_2, \dots, \mathfrak{x}_d) \in \mathbb{R}^d$  that

$$|u(t, x) - u(t, \mathfrak{x})| \leq e^{L_0(T-t)} \left( \sum_{j=1}^d (K_j + (T-t)\mathfrak{L}_j) |x_j - \mathfrak{x}_j| \right), \quad (120)$$

(iv) it holds for all  $t \in [0, T]$ ,  $x \in \mathbb{R}^d$ ,  $i \in \{1, 2, \dots, d\}$  that

$$\left|(\frac{\partial}{\partial x_i} u)(t, x)\right| \leq e^{L_0(T-t)} (K_i + (T-t)\mathfrak{L}_i), \quad (121)$$

and

(v) it holds for all  $x \in \mathbb{R}^d$ ,  $p \in [1, \infty)$  that

$$\begin{aligned} & \sup_{s \in [0, T]} \sup_{t \in [s, T]} \left( \mathbb{E}[|u(t, x + W_t - W_s)|^p] \right)^{1/p} \\ & \leq e^{L_0 T} \left[ \sup_{s \in [0, T]} (\mathbb{E}[|g(x + W_s)|^p])^{1/p} + T \sup_{s, t \in [0, T]} (\mathbb{E}[|f(t, x + W_s, 0, 0)|^p])^{1/p} \right. \\ & \quad \left. + T e^{L_0 T} \sum_{j=1}^d L_j (K_j + T \mathfrak{L}_j) \right]. \end{aligned} \quad (122)$$

**Proof of Lemma 4.2** Throughout this proof let  $t \in [0, T]$ ,  $x = (x_1, x_2, \dots, x_d)$ ,  $\mathfrak{x} = (\mathfrak{x}_1, \mathfrak{x}_2, \dots, \mathfrak{x}_d) \in \mathbb{R}^d$ , let  $\|\cdot\|: \mathbb{R}^d \rightarrow [0, \infty)$  be the standard norm on  $\mathbb{R}^d$ , let  $\langle \cdot, \cdot \rangle: \mathbb{R}^d \times \mathbb{R}^d \rightarrow [0, \infty)$  be the standard scalar product on  $\mathbb{R}^d$ , let  $F: C([0, T] \times \mathbb{R}^d, \mathbb{R}^{1+d}) \rightarrow C([0, T] \times \mathbb{R}^d, \mathbb{R})$  and  $\mathbf{u}: [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^{d+1}$  satisfy for all  $s \in [0, T]$ ,  $y \in \mathbb{R}^d$ ,  $\mathbf{v} \in C([0, T] \times \mathbb{R}^d, \mathbb{R}^{1+d})$  that  $(F(\mathbf{v}))(s, y) = f(s, y, \mathbf{v}(s, y))$  and  $\mathbf{u}(s, y) = (u(s, y), (\nabla_y u)(s, y))$ , and let  $Y, \mathfrak{Y}: [t, T] \times \Omega \rightarrow \mathbb{R}$  and  $Z, \mathfrak{Z}: [t, T] \times \Omega \rightarrow \mathbb{R}^d$  be stochastic processes which satisfy for all  $s \in [t, T]$  that

$$Y_s = u(s, x + W_s - W_t), \quad \mathfrak{Y}_s = u(s, \mathfrak{x} + W_s - W_t), \quad Z_s = (\nabla_x u)(s, x + W_s - W_t), \quad (123)$$

and  $\mathfrak{Z}_s = (\nabla_x u)(s, \mathfrak{x} + W_s - W_t)$ . Observe that [47, Lemma 4.2] establishes Item (i) and Item (ii). Next we prove Item (iii). Note that Itô's lemma yields that for all  $s \in [t, T]$  it holds  $\mathbb{P}$ -a.s. that

$$Y_T = Y_s - \int_s^T f(r, x + W_r - W_t, Y_r, Z_r) dr + \int_s^T \langle Z_r, dW_r \rangle \quad (124)$$

and

$$\mathfrak{Y}_T = \mathfrak{Y}_s - \int_s^T f(r, \mathfrak{x} + W_r - W_t, \mathfrak{Y}_r, \mathfrak{Z}_r) dr + \int_s^T \langle \mathfrak{Z}_r, dW_r \rangle. \quad (125)$$

Next note that (116) implies that there exists  $\lambda \in (\frac{1}{2}, \infty)$  which satisfies that  $\sup_{s \in [0, T], \xi \in \mathbb{R}^d} \frac{|u(s, \xi)|}{1 + \|\xi\|^\lambda} < \infty$ . Observe that Doob's inequality implies that

$$\begin{aligned} \left\| \sup_{s \in [t, T]} |Y_s| \right\|_{L^2(\mathbb{P}; \mathbb{R})} &= \left\| \sup_{s \in [t, T]} |u(s, x + W_s - W_t)| \right\|_{L^2(\mathbb{P}; \mathbb{R})} \\ &\leq \left[ \sup_{s \in [0, T], \xi \in \mathbb{R}^d} \frac{|u(s, \xi)|}{1 + \|\xi\|^\lambda} \right] \left( 1 + \left\| \sup_{s \in [t, T]} \|x + W_s - W_t\|^\lambda \right\|_{L^2(\mathbb{P}; \mathbb{R})} \right) \\ &\leq \left[ \sup_{s \in [0, T], \xi \in \mathbb{R}^d} \frac{|u(s, \xi)|}{1 + \|\xi\|^\lambda} \right] \left( 1 + \left( \frac{2\lambda}{2\lambda - 1} \right)^{1/\lambda} \left\| \|x + W_T - W_t\|^\lambda \right\|_{L^2(\mathbb{P}; \mathbb{R})} \right) < \infty \end{aligned} \quad (126)$$

and

$$\left\| \sup_{s \in [t, T]} |\mathfrak{Y}_s| \right\|_{L^2(\mathbb{P}; \mathbb{R})} < \infty. \quad (127)$$

Moreover, note that (115) implies that for all  $s \in [t, T], y, \mathfrak{y} \in \mathbb{R}, z = (z_1, z_2, \dots, z_d), \mathfrak{z} = (\mathfrak{z}_1, \mathfrak{z}_2, \dots, \mathfrak{z}_d) \in \mathbb{R}^d$  it holds  $\mathbb{P}$ -a.s. that

$$|f(s, x + W_t - W_s, y, z) - f(s, x + W_t - W_s, \mathfrak{y}, \mathfrak{z})| \leq L_0 |y - \mathfrak{y}| + \sum_{j=1}^d L_j |z_j - \mathfrak{z}_j|. \quad (128)$$

and

$$|f(s, x + W_s - W_t, y, z) - f(s, \mathfrak{x} + W_s - W_t, y, z)| \leq \sum_{j=1}^d \mathfrak{L}_j |x_j - \tilde{x}_j|. \quad (129)$$

Combining this with (124), (125), (126), (127), and Lemma 4.1 proves that

$$|u(t, x) - u(t, \mathfrak{x})| = |Y_t - \mathfrak{Y}_t|$$

$$\begin{aligned}
&\leq e^{L_0(T-t)} \left( \|Y_T - \mathfrak{Y}_T\|_{L^\infty(\mathbb{P}; \mathbb{R})} \right. \\
&+ (T-t) \sup_{s \in [t, T]} \sup_{y \in \mathbb{R}} \sup_{z \in \mathbb{R}^d} \|f(s, x + W_s - W_t, y, z) \right. \\
&\quad \left. - f(s, \mathfrak{x} + W_s - W_t, y, z)\|_{L^\infty(\mathbb{P}; \mathbb{R})} \right) \\
&\leq e^{L_0(T-t)} \left( \|g(x + W_T - W_t) - g(\mathfrak{x} + W_T - W_t)\|_{L^\infty(\mathbb{P}; \mathbb{R})} \right. \\
&+ (T-t) \sum_{j=1}^d \mathfrak{L}_j |x_j - \mathfrak{x}_j| \Big) \\
&\leq e^{L_0(T-t)} \left( \sum_{j=1}^d (K_j + (T-t)\mathfrak{L}_j) |x_j - \mathfrak{x}_j| \right). \tag{130}
\end{aligned}$$

This establishes Item (iii). Next note that Item (iii) demonstrates that for all  $i \in \{1, 2, \dots, d\}$  it holds that

$$\left| \left( \frac{\partial}{\partial x_i} u \right)(t, x) \right| = \left| \lim_{\mathbb{R} \setminus \{0\} \ni h \rightarrow 0} \frac{u(t, x + he_i) - u(t, x)}{h} \right| \leq e^{L_0(T-t)} (K_i + (T-t)\mathfrak{L}_i). \tag{131}$$

This establishes Item (iv). In the next step we prove Item (v). Observe that Item (ii), Tonelli's theorem, and the triangle inequality prove that for all  $r \in [0, T]$ ,  $s \in [r, T]$ ,  $p \in [1, \infty)$  it holds that

$$\begin{aligned}
&\|u(s, x + W_s - W_r)\|_{L^p(\mathbb{P}; \mathbb{R})} \\
&= \left\| \int g(x + W_s - W_r + y) \mathbb{P}_{W_T - W_r}(dy) \right. \\
&\quad \left. + \int_s^T \int (F(\mathbf{u}))(v, x + W_s - W_r + y) \mathbb{P}_{W_v - W_s}(dy) dv \right\|_{L^p(\mathbb{P}; \mathbb{R})} \\
&\leq \left\| \int g(x + W_s - W_r + y) \mathbb{P}_{W_T - W_r}(dy) \right\|_{L^p(\mathbb{P}; \mathbb{R})} \\
&\quad + \int_s^T \left\| \int (F(\mathbf{u}))(v, x + W_s - W_r + y) \mathbb{P}_{W_v - W_s}(dy) \right\|_{L^p(\mathbb{P}; \mathbb{R})} dv. \tag{132}
\end{aligned}$$

This and Jensen's inequality show that for all  $r \in [0, T]$ ,  $s \in [r, T]$ ,  $p \in [1, \infty)$  it holds that

$$\begin{aligned}
&\|u(r, x + W_s - W_r)\|_{L^p(\mathbb{P}; \mathbb{R})} \leq \left( \int \int |g(x + z + y)|^p \mathbb{P}_{W_T - W_r}(dy) \mathbb{P}_{W_s - W_r}(dz) \right)^{\frac{1}{p}} \\
&\quad + \int_s^T \left( \int \int |(F(\mathbf{u}))(v, x + z + y)|^p \mathbb{P}_{W_T - W_r}(dy) \mathbb{P}_{W_s - W_r}(dz) \right)^{\frac{1}{p}} dv. \\
&= \left( \int \int |g(x + z + y)|^p \mathbb{P}_{(W_v - W_s, W_s - W_r)}(d(y, z)) \right)^{\frac{1}{p}}
\end{aligned}$$

$$\begin{aligned}
& + \int_s^T \left( \int \int |(F(\mathbf{u}))(v, x+z+y)|^p \mathbb{P}_{(W_v-W_s, W_s-W_r)}(d(y, z)) \right)^{\frac{1}{p}} dv \\
& = \|g(x+W_T-W_r)\|_{L^p(\mathbb{P}; \mathbb{R})} + \int_s^T \|(F(\mathbf{u}))(v, x+W_v-W_r)\|_{L^p(\mathbb{P}; \mathbb{R})} dv. \quad (133)
\end{aligned}$$

This, the triangle inequality, (115), and Item (iv) show that for all  $r \in [0, T]$ ,  $p \in [1, \infty)$  it holds that

$$\begin{aligned}
& \sup_{s \in [r, T]} \|u(r, x+W_s-W_r)\|_{L^p(\mathbb{P}; \mathbb{R})} \leq \|g(x+W_T-W_r)\|_{L^p(\mathbb{P}; \mathbb{R})} \\
& + \int_r^T \|(F(0))(v, x+W_v-W_r)\|_{L^p(\mathbb{P}; \mathbb{R})} dv \\
& + L_0 \int_r^T \|u(v, x+W_v-W_r)\|_{L^p(\mathbb{P}; \mathbb{R})} dv \\
& + \sum_{j=1}^d L_j \int_r^T \|(\frac{\partial}{\partial x_j} u)(v, x+W_v-W_r)\|_{L^p(\mathbb{P}; \mathbb{R})} dv \\
& \leq \|g(x+W_T-W_r)\|_{L^p(\mathbb{P}; \mathbb{R})} + \int_r^T \|(F(0))(v, x+W_v-W_r)\|_{L^p(\mathbb{P}; \mathbb{R})} dv \\
& + L_0 \int_r^T \sup_{s \in [v, T]} \|u(s, x+W_s-W_t)\|_{L^p(\mathbb{P}; \mathbb{R})} dv + \sum_{j=1}^d e^{L_0 T} T L_j (K_j + T \mathfrak{L}_j). \quad (134)
\end{aligned}$$

In addition, note that the fact that  $\sup_{s \in [0, T], \xi \in \mathbb{R}^d} \frac{|u(s, \xi)|}{1 + \|\xi\|^\lambda} < \infty$  ensures that for  $p \in [1, \infty)$  it holds that

$$\begin{aligned}
& \sup_{r \in [0, T]} \sup_{s \in [r, T]} \|u(s, x+W_s-W_r)\|_{L^p(\mathbb{P}; \mathbb{R})} \\
& \leq \left[ \sup_{r \in [0, T], \xi \in \mathbb{R}^d} \frac{|u(r, \xi)|}{1 + \|\xi\|^\lambda} \right] \left( 1 + \sup_{r \in [0, T]} \sup_{s \in [r, T]} \|\|x+W_s-W_r\|^\lambda\|_{L^p(\mathbb{P}; \mathbb{R})} \right) < \infty. \quad (135)
\end{aligned}$$

This, (134), and Gronwall's inequality yield that for all  $p \in [1, \infty)$  it holds that

$$\begin{aligned}
& \sup_{r \in [0, T]} \sup_{s \in [r, T]} \|u(s, x+W_s-W_r)\|_{L^p(\mathbb{P}; \mathbb{R})} \leq e^{L_0 T} \sup_{r \in [0, T]} \|g(x+W_r)\|_{L^p(\mathbb{P}; \mathbb{R})} \\
& + e^{L_0 T} T \left( \sup_{s, r \in [0, T]} \|(F(0))(s, x+W_r)\|_{L^p(\mathbb{P}; \mathbb{R})} + e^{L_0 T} \sum_{j=1}^d L_j (K_j + T \mathfrak{L}_j) \right). \quad (136)
\end{aligned}$$

The proof of Lemma 4.2 is thus completed.  $\square$

## 5 Overall Complexity Analysis for MLP Approximation Methods

In this section we combine the findings of Sects. 3 and 4 to establish in Theorem 5.2 below the main approximation result of this article; see also Corollary 5.1 and Corollary 5.4 below. The i.i.d. random variables  $r^\theta : \Omega \rightarrow (0, 1)$ ,  $\theta \in \Theta$ , appearing in the MLP approximation methods in Corollary 5.1 (see (140) in Corollary 5.1), Theorem 5.2 (see (160) in Theorem 5.2), and Corollary 5.4 (see (175) in Corollary 5.4) are employed to approximate the time integrals in the semigroup formulations of the PDEs under consideration. One of the key ingredients of the MLP approximation methods, which we propose and analyze in this article, is the fact that the density of these i.i.d. random variables  $r^\theta : \Omega \rightarrow (0, 1)$ ,  $\theta \in \Theta$ , is equal to the function  $(0, 1) \ni s \mapsto \alpha s^{\alpha-1} \in \mathbb{R}$  for some  $\alpha \in (0, 1)$ , or equivalently, that these i.i.d. random variables satisfy for all  $\theta \in \Theta$ ,  $b \in (0, 1)$  that  $\mathbb{P}(r^\theta \leq b) = b^\alpha$  for some  $\alpha \in (0, 1)$ . In particular, in contrast to previous MLP approximation methods studied in the scientific literature (see, e.g., [4, 45, 46]) it is crucial in this article to exclude the case where the random variables  $r^\theta : \Omega \rightarrow (0, 1)$ ,  $\theta \in \Theta$ , are continuous uniformly distributed on  $(0, 1)$  (corresponding to the case  $\alpha = 1$ ). To make this aspect more clear to the reader, we provide in Lemma 5.3 below an explanation why it is essential to exclude the continuous uniform distribution case  $\alpha = 1$ . Note that the random variable  $U : \Omega \rightarrow \mathbb{R}^{d+1}$  in Lemma 5.3 coincides with a special case of the random fields in (160) in Theorem 5.2 (applied with  $g_d(x) = 0$ ,  $f_d(s, x, y, z) = 1$ ,  $M = 1$ ,  $t = 0$  for  $s \in [0, T]$ ,  $x, z \in \mathbb{R}^d$ ,  $y \in \mathbb{R}$ ,  $d \in \mathbb{N}$  in the notation of Theorem 5.2).

### 5.1 Quantitative Complexity Analysis for MLP Approximation Methods

**Corollary 5.1** Let  $\|\cdot\|_1 : (\cup_{n \in \mathbb{N}} \mathbb{R}^n) \rightarrow \mathbb{R}$  and  $\|\cdot\|_\infty : (\cup_{n \in \mathbb{N}} \mathbb{R}^n) \rightarrow \mathbb{R}$  satisfy for all  $n \in \mathbb{N}$ ,  $x = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$  that  $\|x\|_1 = \sum_{i=1}^n |x_i|$  and  $\|x\|_\infty = \max_{i \in \{1, 2, \dots, n\}} |x_i|$ , let  $T, \delta \in (0, \infty)$ ,  $\varepsilon \in (0, 1]$ ,  $d \in \mathbb{N}$ ,  $L = (L_0, L_1, \dots, L_d) \in \mathbb{R}^{d+1}$ ,  $K = (K_1, K_2, \dots, K_d)$ ,  $\mathfrak{L} = (\mathfrak{L}_1, \mathfrak{L}_2, \dots, \mathfrak{L}_d)$ ,  $\xi \in \mathbb{R}^d$ ,  $p \in (2, \infty)$ ,  $\alpha \in (\frac{p-2}{2(p-1)}, \frac{p}{2(p-1)})$ ,  $\beta = \frac{\alpha}{2} - \frac{(1-\alpha)(p-2)}{2p}$ ,  $f \in C([0, T] \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d, \mathbb{R})$ ,  $g \in C(\mathbb{R}^d, \mathbb{R})$ , let  $u = (u(t, x))_{(t,x) \in [0,T] \times \mathbb{R}^d} \in C^{1,2}([0, T] \times \mathbb{R}^d, \mathbb{R})$  be an at most polynomially growing function, assume for all  $t \in (0, T)$ ,  $x = (x_1, x_2, \dots, x_d)$ ,  $\mathfrak{x} = (\mathfrak{x}_1, \mathfrak{x}_2, \dots, \mathfrak{x}_d)$ ,  $z = (z_1, z_2, \dots, z_d)$ ,  $\mathfrak{z} = (\mathfrak{z}_1, \mathfrak{z}_2, \dots, \mathfrak{z}_d) \in \mathbb{R}^d$ ,  $y, \mathfrak{y} \in \mathbb{R}$  that

$$|f(t, x, y, z) - f(t, \mathfrak{x}, \mathfrak{y}, \mathfrak{z})| \leq L_0 |y - \mathfrak{y}| + \sum_{j=1}^d (L_j |z_j - \mathfrak{z}_j| + \mathfrak{L}_j |x_j - \mathfrak{x}_j|), \quad (137)$$

$$|g(x) - g(\mathfrak{x})| \leq \sum_{i=1}^d K_i |x_i - \mathfrak{x}_i|, \quad u(T, x) = g(x), \quad (138)$$

$$\text{and } \left( \frac{\partial}{\partial t} u \right)(t, x) + \frac{1}{2} (\Delta_x u)(t, x) + f(t, x, u(t, x), (\nabla_x u)(t, x)) = 0, \quad (139)$$

let  $F : C([0, T] \times \mathbb{R}^d, \mathbb{R}^{1+d}) \rightarrow C([0, T] \times \mathbb{R}^d, \mathbb{R})$  satisfy for all  $t \in [0, T)$ ,  $x \in \mathbb{R}^d$ ,  $\mathbf{v} \in C([0, T] \times \mathbb{R}^d, \mathbb{R}^{1+d})$  that  $(F(\mathbf{v}))(t, x) = f(t, x, \mathbf{v}(t, x))$ , let  $(\Omega, \mathcal{F}, \mathbb{P})$  be a probability space, let  $\Theta = \cup_{n \in \mathbb{N}} \mathbb{Z}^n$ , let  $Z^\theta : \Omega \rightarrow \mathbb{R}^d$ ,  $\theta \in \Theta$ , be i.i.d. standard normal random variables, let  $r^\theta : \Omega \rightarrow (0, 1)$ ,  $\theta \in \Theta$ , be i.i.d. random variables, assume for all  $b \in (0, 1)$  that  $\mathbb{P}(r^\theta \leq b) = b^{1-\alpha}$ , assume that  $(Z^\theta)_{\theta \in \Theta}$  and  $(r^\theta)_{\theta \in \Theta}$

are independent, let  $\mathbf{U}_{n,M}^\theta = (\mathbf{U}_{n,M}^{\theta,0}, \mathbf{U}_{n,M}^{\theta,1}, \dots, \mathbf{U}_{n,M}^{\theta,d}) : [0, T) \times \mathbb{R}^d \times \Omega \rightarrow \mathbb{R}^{1+d}$ ,  $n, M \in \mathbb{Z}$ ,  $\theta \in \Theta$ , satisfy for all  $n, M \in \mathbb{N}$ ,  $\theta \in \Theta$ ,  $t \in [0, T)$ ,  $x \in \mathbb{R}^d$  that  $\mathbf{U}_{-1,M}^\theta(t, x) = \mathbf{U}_{0,M}^\theta(t, x) = 0$  and

$$\begin{aligned} \mathbf{U}_{n,M}^\theta(t, x) &= (g(x), 0) + \frac{1}{M^n} \sum_{l=1}^{M^n} (g(x + [T-t]^{1/2} Z^{(\theta,0,-i)}) \\ &\quad - g(x)) (1, [T-t]^{-1/2} Z^{(\theta,0,-i)}) \\ &\quad + \sum_{l=0}^{n-1} \sum_{i=1}^{M^{n-l}} \frac{(T-t)(\mathbf{r}^{(\theta,l,i)})^\alpha}{(1-\alpha)M^{n-l}} (1, [(T-t)\mathbf{r}^{(\theta,l,i)}]^{-1/2} Z^{(\theta,l,i)}) \\ &\quad \cdot [(F(\mathbf{U}_{l,M}^{(\theta,l,i)}) - \mathbb{1}_N(l) F(\mathbf{U}_{l-1,M}^{(\theta,-l,i)}))(t + (T-t)\mathbf{r}^{(\theta,l,i)}, x) \\ &\quad + [(T-t)\mathbf{r}^{(\theta,l,i)}]^{1/2} Z^{(\theta,l,i)})], \end{aligned} \quad (140)$$

let  $(\text{RV}_{n,M})_{(n,M) \in \mathbb{Z}^2} \subseteq \mathbb{Z}$  satisfy for all  $n, M \in \mathbb{N}$  that  $\text{RV}_{0,M} = 0$  and

$$\text{RV}_{n,M} \leq d M^n + \sum_{l=0}^{n-1} \left[ M^{(n-l)} (d + 1 + \text{RV}_{l,M} + \mathbb{1}_N(l) \text{RV}_{l-1,M}) \right], \quad (141)$$

and let  $C \in (0, \infty)$  satisfy that

$$C = \max \left\{ \frac{1}{2}, |\Gamma(\frac{p}{2})|^{\frac{1}{p}} (1-\alpha)^{\frac{1}{p}-1} \max \left\{ T, |\Gamma(\frac{p+1}{2})|^{\frac{1}{p}} \pi^{-\frac{1}{2p}} \sqrt{2T} \right\} \max\{1, \|L\|_1\} \right\}. \quad (142)$$

Then there exists  $N \in \mathbb{N} \cap [2, \infty)$  such that

$$\begin{aligned} \sup_{n \in \mathbb{N} \cap [N, \infty)} &\left[ \|\mathbf{U}_{n,\lfloor n^{2\beta} \rfloor}^{0,0}(0, \xi) - u(0, \xi)\|_{L^2(\mathbb{P}; \mathbb{R})} \right. \\ &\left. + \max_{i \in \{1, 2, \dots, d\}} \|\mathbf{U}_{n,\lfloor n^{2\beta} \rfloor}^{0,i}(0, \xi) - (\frac{\partial}{\partial x_i} u)(0, \xi)\|_{L^2(\mathbb{P}; \mathbb{R})} \right] \leq \varepsilon \end{aligned} \quad (143)$$

and

$$\begin{aligned} \sum_{n=1}^N \text{RV}_{n,\lfloor n^{2\beta} \rfloor} &\leq d \varepsilon^{-(2+\delta)} 2^{3+\delta} \left[ 1 + \frac{\sqrt{\max\{T, 3\}} \|K\|_1}{\sqrt{\lfloor (N-1)^{2\beta} \rfloor}} \right. \\ &\quad + C e^{L_0 T} \sup_{s \in [0, T]} \|g(\xi + \sqrt{s} Z^{(0)})\|_{L^{\frac{2p}{p-2}}(\mathbb{P}; \mathbb{R})} \\ &\quad + C e^{L_0 T} (\|K\|_\infty + T \|\mathfrak{L}\|_\infty) \\ &\quad + C \sup_{s,t \in [0, T]} \left\| (F(0))(t, \xi + \sqrt{s} Z^{(0)}) \right\|_{L^{\frac{2p}{p-2}}(\mathbb{P}; \mathbb{R})} \left( \frac{1}{\sqrt{\lfloor (N-1)^{2\beta} \rfloor}} + T e^{L_0 T} \right) \\ &\quad \left. + T C e^{2L_0 T} \sum_{j=1}^d L_j (K_j + T \mathfrak{L}_j) \right]^{2+\delta} \end{aligned}$$

$$\cdot \left[ \sup_{n \in \mathbb{N} \cap [2, \infty)} \left( \frac{10^n \left[ (n-1)^{2\beta} \left[ e \left( \frac{p(n-1)}{2} + 1 \right) \right]^{\frac{1}{8}} (4Ce^\beta)^{n-1} \right]^{2+\delta}}{(\lfloor (n-1)^{2\beta} \rfloor)^{\frac{\delta n}{2}}} \right) \right] < \infty. \quad (144)$$

**Proof of Corollary 5.1** Throughout the proof let  $(\eta_{n,M})_{(n,M) \in \mathbb{N}^2} \subseteq \mathbb{R}$  satisfy for all  $n, M \in \mathbb{N}$  that

$$\eta_{n,M} = \left\| \mathbf{U}_{n,M}^{0,0}(0, \xi) - u(0, \xi) \right\|_{L^2(\mathbb{P}; \mathbb{R})} + \max_{i \in \{1, 2, \dots, d\}} \left\| \mathbf{U}_{n,M}^{0,i}(0, \xi) - (\frac{\partial}{\partial x_i} u)(0, \xi) \right\|_{L^2(\mathbb{P}; \mathbb{R})}. \quad (145)$$

Note that Proposition 3.5 and Item (i) of Lemma 4.2 ensure that for all  $M, n \in \mathbb{N}$  it holds that

$$\begin{aligned} \eta_{n,M} &\leq 2 \max \left\{ \left\| \mathbf{U}_{n,M}^{0,0}(0, \xi) - u(0, \xi) \right\|_{L^2(\mathbb{P}; \mathbb{R})}, \max_{i \in \{1, 2, \dots, d\}} \left\| \mathbf{U}_{n,M}^{0,i}(0, \xi) - (\frac{\partial}{\partial x_i} u)(0, \xi) \right\|_{L^2(\mathbb{P}; \mathbb{R})} \right\} \\ &\leq \frac{2 \left[ e \left( \frac{pn}{2} + 1 \right) \right]^{\frac{1}{8}} (4C)^{n-1} \exp \left( \beta M^{\frac{1}{2p}} \right)}{\sqrt{M^{n-1}}} \left[ \frac{\sqrt{\max\{T, 3\}} \|K\|_1}{\sqrt{M}} + \frac{C \sup_{s, t \in [0, T]} \|(F(0))(t, \xi + \sqrt{s} Z^{(0)})\|_{L^{\frac{2p}{p-2}}(\mathbb{P}; \mathbb{R})}}{\sqrt{M}} \right. \\ &\quad \left. + C \sup_{s \in [0, T]} \left[ \max \left\{ \left\| u(s, \xi + \sqrt{s} Z^{(0)}) \right\|_{L^{\frac{2p}{p-2}}(\mathbb{P}; \mathbb{R})}, \max_{i \in \{1, 2, \dots, d\}} \left\| (\frac{\partial}{\partial x_i} u)(s, \xi + \sqrt{s} Z^{(0)}) \right\|_{L^{\frac{2p}{p-2}}(\mathbb{P}; \mathbb{R})} \right\} \right]. \end{aligned} \quad (146)$$

Combining this with Lemma 4.2 implies for all  $M, n \in \mathbb{N}$  that

$$\begin{aligned} \eta_{n,M} &\leq \frac{2 \left[ e \left( \frac{pn}{2} + 1 \right) \right]^{\frac{1}{8}} (4C)^n \exp \left( \beta M^{\frac{1}{2p}} \right)}{\sqrt{M^{n-1}}} \left[ \frac{\sqrt{\max\{T, 3\}} \|K\|_1}{\sqrt{M}} \right. \\ &\quad + C \left( \frac{1}{\sqrt{M}} + Te^{L_0 T} \right) \sup_{s, t \in [0, T]} \left\| (F(0))(t, \xi + \sqrt{s} Z^{(0)}) \right\|_{L^{\frac{2p}{p-2}}(\mathbb{P}; \mathbb{R})} \\ &\quad + Ce^{L_0 T} (\|K\|_\infty + T \|\mathcal{L}\|_\infty) \\ &\quad \left. + Ce^{L_0 T} \sup_{s \in [0, T]} \|g(\xi + \sqrt{s} Z^{(0)})\|_{L^{\frac{2p}{p-2}}(\mathbb{P}; \mathbb{R})} + T Ce^{2L_0 T} \sum_{j=1}^d L_j (K_j + T \mathcal{L}_j) \right]. \end{aligned} \quad (147)$$

Next observe that (137) and (138) demonstrate that

$$\begin{aligned} \sup_{s, t \in [0, T]} \left\| (F(0))(t, \xi + \sqrt{s} Z^{(0)}) \right\|_{L^{\frac{2p}{p-2}}(\mathbb{P}; \mathbb{R})} &< \infty \quad \text{and} \\ \sup_{s \in [0, T]} \|g(\xi + \sqrt{s} Z^{(0)})\|_{L^{\frac{2p}{p-2}}(\mathbb{P}; \mathbb{R})} &< \infty. \end{aligned} \quad (148)$$

Combining this with (147) proves that  $\limsup_{n \rightarrow \infty} \eta_{n, \lfloor n^{2\beta} \rfloor} = 0$ . Next let  $N \in \mathbb{N}$  satisfy that

$$N = \min \left\{ n \in \mathbb{N} \cap [2, \infty) : \sup_{m \in \mathbb{N} \cap [n, \infty)} \eta_{m, \lfloor m^{2\beta} \rfloor} \leq \varepsilon \right\} \quad (149)$$

and let  $\mathfrak{C} \in \mathbb{R}$  satisfy that

$$\begin{aligned} \mathfrak{C} = & 2 \left[ 1 + \frac{\sqrt{\max\{T, 3\}} \|K\|_1}{\sqrt{\lfloor (N-1)^{2\beta} \rfloor}} \right. \\ & + C \sup_{s, t \in [0, T]} \left\| (F(0))(t, \xi + \sqrt{s} Z^{(0)}) \right\|_{L^{\frac{2p}{p-2}}(\mathbb{P}; \mathbb{R})} \left( \frac{1}{\sqrt{\lfloor (N-1)^{2\beta} \rfloor}} + T e^{L_0 T} \right) \\ & + C e^{L_0 T} (\|K\|_\infty + T \|\mathcal{L}\|_\infty) \\ & \left. + C e^{L_0 T} \sup_{s \in [0, T]} \|g(\xi + \sqrt{s} Z^{(0)})\|_{L^{\frac{2p}{p-2}}(\mathbb{P}; \mathbb{R})} + T C e^{2L_0 T} \sum_{j=1}^d L_j (K_j + T \mathcal{L}_j) \right]. \end{aligned} \quad (150)$$

Note that the fact that  $C \geq \frac{1}{2}$  and (150) ensure that

$$1 \leq 4C\mathfrak{C} \left[ e \left( \frac{p}{2} + 1 \right) \right]^{\frac{1}{8}} e^\beta = \frac{\mathfrak{C} \left[ e \left( \frac{p(2-1)}{2} + 1 \right) \right]^{\frac{1}{8}} (4C)^{2-1} \exp \left( \beta (\lfloor (2-1)^{2\beta} \rfloor)^{\frac{1}{2\beta}} \right)}{\sqrt{(\lfloor (2-1)^{2\beta} \rfloor)^{2-2}}}. \quad (151)$$

Furthermore, observe that (147) and (150) demonstrate that

$$\eta_{N-1, \lfloor (N-1)^{2\beta} \rfloor} \leq \frac{\mathfrak{C} \left[ e \left( \frac{p(N-1)}{2} + 1 \right) \right]^{\frac{1}{8}} (4C)^{N-1} \exp \left( \beta (\lfloor (N-1)^{2\beta} \rfloor)^{\frac{1}{2\beta}} \right)}{\sqrt{(\lfloor (N-1)^{2\beta} \rfloor)^{N-2}}}. \quad (152)$$

Combining this with (149) and (151) proves that

$$\begin{aligned} \varepsilon & \leq \mathbb{1}_{\{2\}}(N) + \eta_{N-1, \lfloor (N-1)^{2\beta} \rfloor} \mathbb{1}_{(2, \infty)}(N) \\ & \leq \frac{\mathfrak{C} \left[ e \left( \frac{p(N-1)}{2} + 1 \right) \right]^{\frac{1}{8}} (4C)^{N-1} \exp \left( \beta (\lfloor (N-1)^{2\beta} \rfloor)^{\frac{1}{2\beta}} \right)}{\sqrt{(\lfloor (N-1)^{2\beta} \rfloor)^{N-2}}}. \end{aligned} \quad (153)$$

Moreover, observe that [45, Lemma 3.6] assures that for all  $n \in \mathbb{N}$  it holds that  $\text{RV}_{n, \lfloor n^{2\beta} \rfloor} \leq d(5 \lfloor n^{2\beta} \rfloor)^n$ . Hence, we obtain that

$$\begin{aligned} \sum_{n=1}^N \text{RV}_{n, \lfloor n^{2\beta} \rfloor} & \leq d \sum_{n=1}^N (5 \lfloor n^{2\beta} \rfloor)^n \leq d \sum_{n=1}^N (5 \lfloor N^{2\beta} \rfloor)^n \\ & = \frac{d(5 \lfloor N^{2\beta} \rfloor)((5 \lfloor N^{2\beta} \rfloor)^N - 1)}{5 \lfloor N^{2\beta} \rfloor - 1} \leq 2d(5 \lfloor N^{2\beta} \rfloor)^N. \end{aligned} \quad (154)$$

Note that the fact that  $2\beta \in (0, 1)$  ensures that

$$\lfloor N^{2\beta} \rfloor \leq \lfloor (N-1)^{2\beta} + 1 \rfloor = \lfloor (N-1)^{2\beta} \rfloor + 1 \leq 2 \lfloor (N-1)^{2\beta} \rfloor. \quad (155)$$

Combining this and (153) with (154) proves that

$$\begin{aligned} \sum_{n=1}^N \text{RV}_{n, \lfloor n^{2\beta} \rfloor} &\leq 2d(5 \lfloor N^{2\beta} \rfloor)^N = 2d(5 \lfloor N^{2\beta} \rfloor)^N \varepsilon^{2+\delta} \varepsilon^{-(2+\delta)} \\ &\leq 2d\varepsilon^{-(2+\delta)} (5 \lfloor N^{2\beta} \rfloor)^N \left( \frac{\mathfrak{C} \left[ e \left( \frac{p(N-1)}{2} + 1 \right) \right]^{\frac{1}{8}} (4C)^{N-1} \exp \left( \beta (\lfloor (N-1)^{2\beta} \rfloor)^{\frac{1}{2\beta}} \right)}{\sqrt{(\lfloor (N-1)^{2\beta} \rfloor)^{N-2}}} \right)^{2+\delta} \\ &= 2d\varepsilon^{-(2+\delta)} \mathfrak{C}^{2+\delta} \frac{5^N (\lfloor N^{2\beta} \rfloor)^N \left[ \left[ e \left( \frac{p(N-1)}{2} + 1 \right) \right]^{\frac{1}{8}} (4C)^{N-1} \exp \left( \beta (\lfloor (N-1)^{2\beta} \rfloor)^{\frac{1}{2\beta}} \right) \right]^{2+\delta}}{(\lfloor (N-1)^{2\beta} \rfloor)^{\frac{(N-2)(2+\delta)}{2}}} \\ &\leq 2d\varepsilon^{-(2+\delta)} \mathfrak{C}^{2+\delta} \frac{10^N (\lfloor (N-1)^{2\beta} \rfloor)^N \left[ \left[ e \left( \frac{p(N-1)}{2} + 1 \right) \right]^{\frac{1}{8}} (4C)^{N-1} \exp \left( \beta (\lfloor (N-1)^{2\beta} \rfloor)^{\frac{1}{2\beta}} \right) \right]^{2+\delta}}{(\lfloor (N-1)^{2\beta} \rfloor)^{\frac{(N-2)(2+\delta)}{2}}} \\ &= 2d\varepsilon^{-(2+\delta)} \mathfrak{C}^{2+\delta} \frac{10^N \left[ \lfloor (N-1)^{2\beta} \rfloor \left[ e \left( \frac{p(N-1)}{2} + 1 \right) \right]^{\frac{1}{8}} (4C)^{N-1} \exp \left( \beta (\lfloor (N-1)^{2\beta} \rfloor)^{\frac{1}{2\beta}} \right) \right]^{2+\delta}}{(\lfloor (N-1)^{2\beta} \rfloor)^{\frac{\delta N}{2}}} \\ &\leq 2d\varepsilon^{-(2+\delta)} \mathfrak{C}^{2+\delta} \frac{10^N \left[ (N-1)^{2\beta} \left[ e \left( \frac{p(N-1)}{2} + 1 \right) \right]^{\frac{1}{8}} (4Ce^\beta)^{N-1} \right]^{2+\delta}}{(\lfloor (N-1)^{2\beta} \rfloor)^{\frac{\delta N}{2}}} \\ &\leq 2d\varepsilon^{-(2+\delta)} \mathfrak{C}^{2+\delta} \left[ \sup_{n \in \mathbb{N} \cap [2, \infty)} \left( \frac{10^n \left[ (n-1)^{2\beta} \left[ e \left( \frac{p(n-1)}{2} + 1 \right) \right]^{\frac{1}{8}} (4Ce^\beta)^{n-1} \right]^{2+\delta}}{(\lfloor (n-1)^{2\beta} \rfloor)^{\frac{\delta n}{2}}} \right) \right] < \infty. \end{aligned} \quad (156)$$

This establishes (144). The proof of Corollary 5.1 is thus completed.  $\square$

## 5.2 Qualitative Complexity Analysis for MLP Approximation Methods

**Theorem 5.2** Let  $T, \delta, \lambda \in (0, \infty)$ ,  $\alpha \in (0, 1)$ ,  $\beta \in (\max\{\frac{1-2\alpha}{1-\alpha}, 0\}, 1 - \alpha)$ , let  $f_d \in C([0, T] \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d, \mathbb{R})$ ,  $d \in \mathbb{N}$ , let  $g_d \in C(\mathbb{R}^d, \mathbb{R})$ ,  $d \in \mathbb{N}$ , let  $\xi_d = (\xi_{d,1}, \xi_{d,2}, \dots, \xi_{d,d}) \in \mathbb{R}^d$ ,  $d \in \mathbb{N}$ , let  $L_{d,i} \in \mathbb{R}$ ,  $d, i \in \mathbb{N}$ , let  $u_d = (u_d(t, x))_{(t,x) \in [0,T] \times \mathbb{R}^d} \in C^{1,2}([0, T] \times \mathbb{R}^d, \mathbb{R})$ ,  $d \in \mathbb{N}$ , be at most polynomially growing functions, let  $F_d: C([0, T] \times \mathbb{R}^d, \mathbb{R}^{1+d}) \rightarrow C([0, T] \times \mathbb{R}^d, \mathbb{R})$ ,  $d \in \mathbb{N}$ , be functions, assume for all  $d \in \mathbb{N}$ ,  $t \in [0, T]$ ,  $x = (x_1, x_2, \dots, x_d)$ ,

$\mathfrak{x} = (\mathfrak{x}_1, \mathfrak{x}_2, \dots, \mathfrak{x}_d)$ ,  $z = (z_1, z_2, \dots, z_d)$ ,  $\mathfrak{z} = (\mathfrak{z}_1, \mathfrak{z}_2, \dots, \mathfrak{z}_d) \in \mathbb{R}^d$ ,  $y, \mathfrak{y} \in \mathbb{R}$ ,  $\mathbf{v} \in C([0, T) \times \mathbb{R}^d, \mathbb{R}^{1+d})$  that

$$\begin{aligned} & \max\{|f_d(t, x, y, z) - f_d(t, \mathfrak{x}, \mathfrak{y}, \mathfrak{z})|, |g_d(x) - g_d(\mathfrak{x})|\} \\ & \leq \sum_{j=1}^d L_{d,j} (d^\lambda |x_j - \mathfrak{x}_j| + |y - \mathfrak{y}| + |z_j - \mathfrak{z}_j|), \end{aligned} \quad (157)$$

$$\begin{aligned} & \left( \frac{\partial}{\partial t} u_d \right)(t, x) + \frac{1}{2} (\Delta_x u_d)(t, x) + f_d(t, x, u(t, x), (\nabla_x u_d)(t, x)) = 0, \\ & u_d(T, x) = g_d(x), \end{aligned} \quad (158)$$

$$\begin{aligned} & d^{-\lambda} (|g_d(0)| + |f_d(t, 0, 0, 0)| + \max_{i \in \{1, 2, \dots, d\}} |\xi_{d,i}|) + \sum_{i=1}^d L_{d,i} \leq \lambda, \\ & \text{and } (F_d(\mathbf{v}))(t, x) = f_d(t, x, \mathbf{v}(t, x)), \end{aligned} \quad (159)$$

let  $(\Omega, \mathcal{F}, \mathbb{P})$  be a probability space, let  $\Theta = \cup_{n \in \mathbb{N}} \mathbb{Z}^n$ , let  $Z^{d,\theta}: \Omega \rightarrow \mathbb{R}^d$ ,  $d \in \mathbb{N}$ ,  $\theta \in \Theta$ , be i.i.d. standard normal random variables, let  $\mathfrak{r}^\theta: \Omega \rightarrow (0, 1)$ ,  $\theta \in \Theta$ , be i.i.d. random variables, assume for all  $b \in (0, 1)$  that  $\mathbb{P}(\mathfrak{r}^0 \leq b) = b^\alpha$ , assume that  $(Z^{d,\theta})_{(d,\theta) \in \mathbb{N} \times \Theta}$  and  $(\mathfrak{r}^\theta)_{\theta \in \Theta}$  are independent, let  $\mathbf{U}_{n,M}^{d,\theta} = (\mathbf{U}_{n,M}^{d,\theta,0}, \mathbf{U}_{n,M}^{d,\theta,1}, \dots, \mathbf{U}_{n,M}^{d,\theta,d}): [0, T) \times \mathbb{R}^d \times \Omega \rightarrow \mathbb{R}^{1+d}$ ,  $n, M, d \in \mathbb{Z}$ ,  $\theta \in \Theta$ , satisfy for all  $n, M, d \in \mathbb{N}$ ,  $\theta \in \Theta$ ,  $t \in [0, T)$ ,  $x \in \mathbb{R}^d$  that  $\mathbf{U}_{-1,M}^{d,\theta}(t, x) = \mathbf{U}_{0,M}^{d,\theta}(t, x) = 0$  and

$$\begin{aligned} \mathbf{U}_{n,M}^{d,\theta}(t, x) &= (g_d(x), 0) + \sum_{i=1}^{M^n} \frac{1}{M^n} \\ & (g_d(x + [T-t]^{1/2} Z^{(\theta, 0, -i)}) - g_d(x)) (1, [T-t]^{-1/2} Z^{d,(\theta, 0, -i)}) \\ & + \sum_{l=0}^{n-1} \sum_{i=1}^{M^{n-l}} \frac{(T-t)(\mathfrak{r}^{(\theta, l, i)})^{1-\alpha}}{\alpha M^{n-l}} (1, [(T-t)\mathfrak{r}^{(\theta, l, i)}]^{-1/2} Z^{d,(\theta, l, i)}) \\ & \cdot [(F_d(\mathbf{U}_{l,M}^{d,(\theta, l, i)}) - \mathbb{1}_{\mathbb{N}}(l) F_d(\mathbf{U}_{l-1,M}^{d,(\theta, -l, i)})) \\ & \cdot (t + (T-t)\mathfrak{r}^{(\theta, l, i)}, x + [(T-t)\mathfrak{r}^{(\theta, l, i)}]^{1/2} Z^{d,(\theta, l, i)})], \end{aligned} \quad (160)$$

and let  $\text{RV}_{d,n,M} \in \mathbb{Z}$ ,  $d, n, M \in \mathbb{Z}$ , satisfy for all  $d, n, M \in \mathbb{N}$  that  $\text{RV}_{d,0,M} = 0$  and

$$\text{RV}_{d,n,M} \leq d M^n + \sum_{l=0}^{n-1} \left[ M^{(n-l)} (d+1 + \text{RV}_{d,l,M} + \mathbb{1}_{\mathbb{N}}(l) \text{RV}_{d,l-1,M}) \right]. \quad (161)$$

Then there exist  $c \in \mathbb{R}$  and  $N = (N_{d,\varepsilon})_{(d,\varepsilon) \in \mathbb{N} \times (0,1]}: \mathbb{N} \times (0, 1] \rightarrow \mathbb{N}$  such that for all  $d \in \mathbb{N}$ ,  $\varepsilon \in (0, 1]$  it holds that  $\sum_{n=1}^{N_{d,\varepsilon}} \text{RV}_{d,n,\lfloor n^\beta \rfloor} \leq c d^c \varepsilon^{-(2+\delta)}$  and

$$\begin{aligned} & \sup_{n \in \mathbb{N} \cap [N_{d,\varepsilon}, \infty)} \left[ \mathbb{E}[|\mathbf{U}_{n,\lfloor n^\beta \rfloor}^{d,0,0}(0, \xi_d) - u_d(0, \xi_d)|^2] \right. \\ & \left. + \max_{i \in \{1, 2, \dots, d\}} \mathbb{E}[|\mathbf{U}_{n,\lfloor n^\beta \rfloor}^{d,0,i}(0, \xi_d) - (\frac{\partial}{\partial x_i} u_d)(0, \xi_d)|^2] \right]^{1/2} \leq \varepsilon. \end{aligned} \quad (162)$$

**Proof of Theorem 5.2** Throughout this proof let  $p = \frac{2\alpha}{\beta+2\alpha-1}$ . Note that the hypothesis that  $\alpha \in (0, 1)$  and the hypothesis that  $\beta \in (\max\{\frac{1-2\alpha}{1-\alpha}, 0\}, 1 - \alpha)$  ensure that  $p > \frac{2\alpha}{1-\alpha+2\alpha-1} = 2$ . Moreover, observe that the fact that  $p = \frac{2\alpha}{\beta+2\alpha-1}$  demonstrates that

$$\frac{\beta}{2} = \frac{1 - \alpha}{2} - \frac{(1 - (1 - \alpha))(p - 2)}{2p}. \quad (163)$$

Next note that the hypothesis that  $\alpha \in (0, 1)$  and the hypothesis that  $\beta \in (\max\{\frac{1-2\alpha}{1-\alpha}, 0\}, 1 - \alpha)$  ensure that

$$\frac{p}{2(p-1)} = \frac{\alpha}{1-\beta} > \frac{\alpha}{1-\frac{1-2\alpha}{1-\alpha}} = 1 - \alpha. \quad (164)$$

Moreover, observe that the hypothesis that  $\alpha \in (0, 1)$  and the hypothesis that  $\beta \in (\max\{\frac{1-2\alpha}{1-\alpha}, 0\}, 1 - \alpha)$  prove that

$$\frac{p-2}{2(p-1)} = 1 - \frac{\alpha}{1-\beta} < 1 - \alpha. \quad (165)$$

Combining this with (164) ensures that

$$1 - \alpha \in \left( \frac{p-2}{2(p-1)}, \frac{p}{2(p-1)} \right). \quad (166)$$

Next note that (157) ensures for all  $d \in \mathbb{N}$  that

$$\begin{aligned} & \sup_{s,t \in [0,T]} \left\| (F_d(0))(t, \xi_d + \sqrt{s}Z^{d,(0)}) \right\|_{L^{\frac{2p}{p-2}}(\mathbb{P};\mathbb{R})} \\ & \leq \sup_{s,t \in [0,T]} \left\| (F_d(0))(t, \xi_d + \sqrt{s}Z^{d,(0)}) - (F_d(0))(t, 0) \right\|_{L^{\frac{2p}{p-2}}(\mathbb{P};\mathbb{R})} \\ & \quad + \sup_{t \in [0,T]} |(F_d(0))(t, 0)| \\ & = \sup_{s,t \in [0,T]} \left\| f_d(t, \xi_d + \sqrt{s}Z^{d,(0)}, 0, 0) - f_d(t, 0, 0, 0) \right\|_{L^{\frac{2p}{p-2}}(\mathbb{P};\mathbb{R})} \\ & \quad + \sup_{t \in [0,T]} |f_d(t, 0, 0, 0)| \\ & \leq \left( d^\lambda \sum_{j=1}^d L_{d,j} (|\xi_{d,j}| + \sqrt{T} \|Z^{1,(0)}\|_{L^{\frac{2p}{p-2}}(\mathbb{P};\mathbb{R})}) \right) \\ & \quad + \sup_{t \in [0,T]} |f_d(t, 0, 0, 0)| \\ & \leq d^\lambda \left( \left[ \max_{j \in \{1,2,\dots,d\}} |\xi_{d,j}| \right] + \sqrt{T} \|Z^{1,(0)}\|_{L^{\frac{2p}{p-2}}(\mathbb{P};\mathbb{R})} \right) \left( \sum_{j=1}^d L_{d,j} \right) \end{aligned}$$

$$+ \sup_{t \in [0, T]} |f_d(t, 0, 0, 0)|. \quad (167)$$

In addition, observe that (157) proves that for all  $d \in \mathbb{N}$  it holds that

$$\begin{aligned} & \sup_{s \in [0, T]} \|g_d(\xi_d + \sqrt{s} Z^{d, (0)})\|_{L^{\frac{2p}{p-2}}(\mathbb{P}; \mathbb{R})} \\ & \leq \sup_{s \in [0, T]} \|g_d(\xi_d + \sqrt{s} Z^{d, (0)}) - g_d(0)\|_{L^{\frac{2p}{p-2}}(\mathbb{P}; \mathbb{R})} + |g_d(0)| \\ & \leq \left( d^\lambda \sum_{j=1}^d L_{d,j} \left( |\xi_{d,j}| + \sqrt{T} \|Z^{1,(0)}\|_{L^{\frac{2p}{p-2}}(\mathbb{P}; \mathbb{R})} \right) \right) + |g_d(0)| \quad (168) \\ & \leq d^\lambda \left( \left[ \max_{j \in \{1, 2, \dots, d\}} |\xi_{d,j}| \right] + \sqrt{T} \|Z^{1,(0)}\|_{L^{\frac{2p}{p-2}}(\mathbb{P}; \mathbb{R})} \right) \left( \sum_{j=1}^d L_{d,j} \right) + |g_d(0)|. \end{aligned}$$

Furthermore, note that the fact that for all  $d \in \mathbb{N}$  it holds that  $(\sum_{j=1}^d |L_{d,j}|^2)^{1/2} \leq \sum_{j=1}^d L_{d,j}$  ensures that for all  $d \in \mathbb{N}$  it holds that

$$\begin{aligned} \max_{j \in \{1, 2, \dots, d\}} d^\lambda L_{d,j} & \leq d^\lambda \sum_{j=1}^d L_{d,j} \quad \text{and} \\ \sum_{j=1}^d L_{d,j} (d^\lambda L_{d,j} + T d^\lambda L_{d,j}) & \leq d^\lambda (T+1) \left[ \sum_{j=1}^d L_{d,j} \right]^2. \quad (169) \end{aligned}$$

This, (163), (166), the fact that  $p > 2$ , (167), (168), Corollary 5.1 (applied with  $\alpha = 1 - \alpha$ ,  $\beta = \frac{\beta}{2}$ ,  $p = p$ ,  $L_0 = \sum_{j=1}^d L_{d,j}$ ,  $L_j = L_{d,j}$ ,  $K_j = d^\lambda L_{d,j}$ ,  $\mathfrak{L}_j = d^\lambda L_{d,j}$  for  $j \in \{1, \dots, d\}$ ,  $d \in \mathbb{N}$  in the notation of Corollary 5.1), and the fact that

$$\sup_{d \in \mathbb{N}} \left[ \frac{1}{d^\lambda} \left( \max_{i \in \{1, 2, \dots, d\}} |\xi_{d,i}| + \sup_{t \in [0, T]} |f_d(t, 0, 0, 0)| + |g_d(0)| \right) + \sum_{i=1}^d L_{d,i} \right] < \infty \quad (170)$$

establish (162). The proof of Theorem 5.2 is thus completed.  $\square$

**Lemma 5.3** Let  $T \in (0, \infty)$ ,  $d \in \mathbb{N}$ ,  $\alpha \in (0, 1]$ , let  $(\Omega, \mathcal{F}, \mathbb{P})$  be a probability space, let  $Z = (Z^1, Z^2, \dots, Z^d): \Omega \rightarrow \mathbb{R}^d$  be a standard normal random variable, let  $\mathfrak{r}: \Omega \rightarrow (0, 1)$  satisfy for all  $b \in (0, 1)$  that  $\mathbb{P}(\mathfrak{r} \leq b) = b^\alpha$ , assume that  $Z$  and  $\mathfrak{r}$  are independent, and let  $\mathbf{U} = (\mathbf{U}^0, \mathbf{U}^1, \dots, \mathbf{U}^d): \Omega \rightarrow \mathbb{R}^{d+1}$  satisfy that  $\mathbf{U} = T\alpha^{-1}\mathfrak{r}^{1-\alpha}(1, [T\mathfrak{r}]^{-1/2}Z)$ . Then

(i) it holds for all  $i \in \{1, 2, \dots, d\}$  that

$$\mathbb{E}[|\mathbf{U}^i|^2] = \frac{T}{\alpha} \int_0^1 s^{-\alpha} ds = \begin{cases} \frac{T}{\alpha(1-\alpha)} & : \alpha < 1 \\ \infty & : \alpha = 1 \end{cases} \quad (171)$$

and

(ii) it holds that  $\mathbf{U} \in L^2(\mathbb{P}; \mathbb{R}^{d+1})$  if and only if  $\alpha \in (0, 1)$ .

**Proof of Lemma 5.3** Note that for all  $i \in \{1, 2, \dots, d\}$  it holds that

$$\begin{aligned}\mathbb{E}[|\mathbf{U}^i|^2] &= \frac{T}{\alpha^2} \mathbb{E}[\mathfrak{r}^{2(1-\alpha)-1} |Z^i|^2] \\ &= \frac{T}{\alpha} \int_0^1 \mathbb{E}[s^{2(1-\alpha)-1} |Z^i|^2] s^{\alpha-1} ds = \frac{T}{\alpha} \int_0^1 s^{2(1-\alpha)-1+\alpha-1} ds \quad (172) \\ &= \frac{T}{\alpha} \int_0^1 s^{-\alpha} ds = \begin{cases} \frac{T}{\alpha(1-\alpha)} & : \alpha < 1 \\ \infty & : \alpha = 1 \end{cases}.\end{aligned}$$

This proves Item (i). Moreover, observe that Item (i) establishes Item (ii). The proof of Lemma 5.3 is thus completed.  $\square$

**Corollary 5.4** Let  $T, \delta, \lambda \in (0, \infty)$ , let  $f_d \in C(\mathbb{R} \times \mathbb{R}^d, \mathbb{R})$ ,  $d \in \mathbb{N}$ , let  $g_d \in C(\mathbb{R}^d, \mathbb{R})$ ,  $d \in \mathbb{N}$ , let  $\xi_d = (\xi_{d,1}, \xi_{d,2}, \dots, \xi_{d,d}) \in \mathbb{R}^d$ ,  $d \in \mathbb{N}$ , let  $L_{d,i} \in \mathbb{R}$ ,  $d, i \in \mathbb{N}$ , let  $u_d = (u_d(t, x))_{(t,x) \in [0,T] \times \mathbb{R}^d} \in C^{1,2}([0, T] \times \mathbb{R}^d, \mathbb{R})$ ,  $d \in \mathbb{N}$ , be at most polynomially growing functions, assume for all  $d \in \mathbb{N}$ ,  $t \in [0, T]$ ,  $x = (x_1, x_2, \dots, x_d)$ ,  $\mathfrak{x} = (\mathfrak{x}_1, \mathfrak{x}_2, \dots, \mathfrak{x}_d)$ ,  $z = (z_1, z_2, \dots, z_d)$ ,  $\mathfrak{z} = (\mathfrak{z}_1, \mathfrak{z}_2, \dots, \mathfrak{z}_d) \in \mathbb{R}^d$ ,  $y, \mathfrak{y} \in \mathbb{R}$  that

$$\max\{|f_d(y, z) - f_d(\mathfrak{y}, \mathfrak{z})|, |g_d(x) - g_d(\mathfrak{x})|\} \leq \sum_{j=1}^d L_{d,j} (d^\lambda |x_j - \mathfrak{x}_j| + |y - \mathfrak{y}| + |z_j - \mathfrak{z}_j|), \quad (173)$$

$$(\frac{\partial}{\partial t} u_d)(t, x) = (\Delta_x u_d)(t, x) + f_d(u(t, x), (\nabla_x u_d)(t, x)), \quad u_d(0, x) = g_d(x), \quad (174)$$

and  $d^{-\lambda}(|g_d(0)| + |f_d(0, 0)| + \max_{i \in \{1, 2, \dots, d\}} |\xi_{d,i}|) + \sum_{i=1}^d L_{d,i} \leq \lambda$ , let  $(\Omega, \mathcal{F}, \mathbb{P})$  be a probability space, let  $\Theta = \cup_{n \in \mathbb{N}} \mathbb{Z}^n$ , let  $Z^{d,\theta} : \Omega \rightarrow \mathbb{R}^d$ ,  $d \in \mathbb{N}$ ,  $\theta \in \Theta$ , be i.i.d. standard normal random variables, let  $\mathfrak{r}^\theta : \Omega \rightarrow (0, 1)$ ,  $\theta \in \Theta$ , be i.i.d. random variables, assume for all  $b \in (0, 1)$  that  $\mathbb{P}(\mathfrak{r}^\theta \leq b) = \sqrt{b}$ , assume that  $(Z^{d,\theta})_{(d,\theta) \in \mathbb{N} \times \Theta}$  and  $(\mathfrak{r}^\theta)_{\theta \in \Theta}$  are independent, let  $\mathbf{U}_{n,M}^{d,\theta} = (\mathbf{U}_{n,M}^{d,\theta,0}, \mathbf{U}_{n,M}^{d,\theta,1}, \dots, \mathbf{U}_{n,M}^{d,\theta,d}) : (0, T] \times \mathbb{R}^d \times \Omega \rightarrow \mathbb{R}^{1+d}$ ,  $n \in \mathbb{Z}$ ,  $M, d \in \mathbb{N}$ ,  $\theta \in \Theta$ , satisfy for all  $n, M, d \in \mathbb{N}$ ,  $\theta \in \Theta$ ,  $t \in (0, T]$ ,  $x \in \mathbb{R}^d$  that  $\mathbf{U}_{-1,M}^{d,\theta}(t, x) = \mathbf{U}_{0,M}^{d,\theta}(t, x) = 0$  and

$$\begin{aligned}\mathbf{U}_{n,M}^{d,\theta}(t, x) &= (g_d(x), 0) + \sum_{i=1}^{M^n} \frac{1}{M^n} (g_d(x + [2t]^{1/2} Z^{(\theta,0,-i)}) \\ &\quad - g_d(x)) (1, [2t]^{-1/2} Z^{d,(\theta,0,-i)}) \\ &\quad + \sum_{l=0}^{n-1} \sum_{i=1}^{M^{n-l}} \frac{2t[\mathfrak{r}^{(\theta,l,i)}]^{1/2}}{M^{n-l}} \left[ f_d(\mathbf{U}_{l,M}^{d,(\theta,l,i)}(t(1 - \mathfrak{r}^{(\theta,l,i)}), x + [2t\mathfrak{r}^{(\theta,l,i)}]^{1/2} Z^{d,(\theta,l,i)})) \right. \\ &\quad \left. - \mathbb{1}_{\mathbb{N}}(l) f_d(\mathbf{U}_{l-1,M}^{d,(\theta,-l,i)}(t(1 - \mathfrak{r}^{(\theta,l,i)}), x + [2t\mathfrak{r}^{(\theta,l,i)}]^{1/2} Z^{d,(\theta,l,i)})) \right] \\ &\quad \cdot (1, [2t\mathfrak{r}^{(\theta,l,i)}]^{-1/2} Z^{d,(\theta,l,i)}), \quad (175)\end{aligned}$$

and let  $\text{RV}_{d,n,M} \in \mathbb{Z}$ ,  $d, n, M \in \mathbb{Z}$ , satisfy for all  $d, n, M \in \mathbb{N}$  that  $\text{RV}_{d,0,M} = 0$  and

$$\text{RV}_{d,n,M} \leq dM^n + \sum_{l=0}^{n-1} \left[ M^{(n-l)}(d+1 + \text{RV}_{d,l,M} + \mathbb{1}_N(l) \text{RV}_{d,l-1,M}) \right]. \quad (176)$$

Then there exist  $c \in \mathbb{R}$  and  $N = (N_{d,\varepsilon})_{(d,\varepsilon) \in \mathbb{N} \times (0,1]} : \mathbb{N} \times (0,1] \rightarrow \mathbb{N}$  such that for all  $d \in \mathbb{N}$ ,  $\varepsilon \in (0,1]$  it holds that  $\sum_{n=1}^{N_{d,\varepsilon}} \text{RV}_{d,n,\lfloor n^{1/4} \rfloor} \leq cd^c \varepsilon^{-(2+\delta)}$  and

$$\begin{aligned} & \sup_{n \in \mathbb{N} \cap [N_{d,\varepsilon}, \infty)} \left[ \mathbb{E}[|\mathbf{U}_{n,\lfloor n^{1/4} \rfloor}^{d,0,0}(T, \xi_d) - u_d(T, \xi_d)|^2] \right. \\ & \quad \left. + \max_{i \in \{1, 2, \dots, d\}} \mathbb{E}[|\mathbf{U}_{n,\lfloor n^{1/4} \rfloor}^{d,0,i}(T, \xi_d) - (\frac{\partial}{\partial x_i} u_d)(T, \xi_d)|^2] \right]^{1/2} \leq \varepsilon. \end{aligned} \quad (177)$$

**Proof of Corollary 5.4** Observe that Theorem 5.2 (applied with  $\alpha = \frac{1}{2}$ ,  $\beta = \frac{1}{4}$ ,  $T = 2T$ ,  $u_d(t, x) = u_d(T - \frac{t}{2}, x)$ ,  $f_d(t, x, y, z) = f_d(y, z)/2$  for  $t \in [0, 2T]$ ,  $x, z \in \mathbb{R}^d$ ,  $y \in \mathbb{R}$ ,  $d \in \mathbb{N}$  in the notation of Theorem 5.2) establishes (177). This completes the proof of Corollary 5.4.  $\square$

**Corollary 5.5** Let  $T, \delta, \lambda \in (0, \infty)$ , let  $g_d \in C(\mathbb{R}^d, \mathbb{R})$ ,  $d \in \mathbb{N}$ , let  $\xi_d = (\xi_{d,1}, \xi_{d,2}, \dots, \xi_{d,d}) \in \mathbb{R}^d$ ,  $d \in \mathbb{N}$ , let  $u_d = (u_d(t, x))_{(t,x) \in [0,T] \times \mathbb{R}^d} \in C^{1,2}([0, T] \times \mathbb{R}^d, \mathbb{R})$ ,  $d \in \mathbb{N}$ , be at most polynomially growing functions, assume for all  $d \in \mathbb{N}$ ,  $t \in [0, T]$ ,  $x = (x_1, x_2, \dots, x_d)$ ,  $\mathfrak{x} = (\mathfrak{x}_1, \mathfrak{x}_2, \dots, \mathfrak{x}_d)$  that  $|g_d(0)| + \max_{i \in \{1, 2, \dots, d\}} |\xi_{d,i}| \leq \lambda d^\lambda$ ,  $|g_d(x) - g_d(\mathfrak{x})| \leq \lambda d^\lambda (\sum_{j=1}^d |x_j - \mathfrak{x}_j|)$ ,  $u_d(0, x) = g_d(x)$ , and

$$(\frac{\partial}{\partial t} u_d)(t, x) = (\Delta_x u_d)(t, x) + \sin((\frac{\partial}{\partial x_1} u)(t, x)), \quad u_d(0, x) = g_d(x), \quad (178)$$

let  $(\Omega, \mathcal{F}, \mathbb{P})$  be a probability space, let  $\Theta = \cup_{n \in \mathbb{N}} \mathbb{Z}^n$ , let  $Z^{d,\theta} : \Omega \rightarrow \mathbb{R}^d$ ,  $d \in \mathbb{N}$ ,  $\theta \in \Theta$ , be i.i.d. standard normal random variables, let  $\mathfrak{r}^\theta : \Omega \rightarrow (0, 1)$ ,  $\theta \in \Theta$ , be i.i.d. random variables, assume for all  $b \in (0, 1)$  that  $\mathbb{P}(\mathfrak{r}^0 \leq b) = \sqrt{b}$ , assume that  $(Z^{d,\theta})_{(d,\theta) \in \mathbb{N} \times \Theta}$  and  $(\mathfrak{r}^\theta)_{\theta \in \Theta}$  are independent, let  $\mathbf{U}_{n,M}^{d,\theta} = (\mathbf{U}_{n,M}^{d,\theta,0}, \mathbf{U}_{n,M}^{d,\theta,1}, \dots, \mathbf{U}_{n,M}^{d,\theta,d}) : (0, T] \times \mathbb{R}^d \times \Omega \rightarrow \mathbb{R}^{1+d}$ ,  $n \in \mathbb{Z}$ ,  $M, d \in \mathbb{N}$ ,  $\theta \in \Theta$ , satisfy for all  $n, M, d \in \mathbb{N}$ ,  $\theta \in \Theta$ ,  $t \in (0, T]$ ,  $x \in \mathbb{R}^d$  that  $\mathbf{U}_{-1,M}^{d,\theta}(t, x) = \mathbf{U}_{0,M}^{d,\theta}(t, x) = 0$  and

$$\begin{aligned} \mathbf{U}_{n,M}^{d,\theta}(t, x) &= (g_d(x), 0) + \sum_{i=1}^{M^n} \frac{1}{M^n} (g_d(x + [2t]^{1/2} Z^{(\theta,0,-i)}) \\ &\quad - g_d(x)) (1, [2t]^{-1/2} Z^{d,(\theta,0,-i)}) \\ &+ \sum_{l=0}^{n-1} \sum_{i=1}^{M^{n-l}} \frac{2t[\mathfrak{r}^{(\theta,l,i)}]^{1/2}}{M^{n-l}} \left[ \sin(\mathbf{U}_{l,M}^{d,(\theta,l,i),1}(t(1 - \mathfrak{r}^{(\theta,l,i)}), x + [2t\mathfrak{r}^{(\theta,l,i)}]^{1/2} Z^{d,(\theta,l,i)})) \right. \\ &\quad \left. - \mathbb{1}_N(l) \sin(\mathbf{U}_{l-1,M}^{d,(\theta,-l,i),1}(t(1 - \mathfrak{r}^{(\theta,l,i)}), x + [2t\mathfrak{r}^{(\theta,l,i)}]^{1/2} Z^{d,(\theta,l,i)})) \right] \\ &\quad \cdot (1, [2t\mathfrak{r}^{(\theta,l,i)}]^{-1/2} Z^{d,(\theta,l,i)}), \end{aligned} \quad (179)$$

and let  $\text{RV}_{d,n,M} \in \mathbb{Z}$ ,  $d, n, M \in \mathbb{Z}$ , satisfy for all  $d, n, M \in \mathbb{N}$  that  $\text{RV}_{d,0,M} = 0$  and

$$\text{RV}_{d,n,M} \leq dM^n + \sum_{l=0}^{n-1} \left[ M^{(n-l)}(d+1 + \text{RV}_{d,l,M} + \mathbb{1}_N(l) \text{RV}_{d,l-1,M}) \right]. \quad (180)$$

Then there exist  $c \in \mathbb{R}$  and  $N = (N_{d,\varepsilon})_{(d,\varepsilon) \in \mathbb{N} \times (0,1]} : \mathbb{N} \times (0,1] \rightarrow \mathbb{N}$  such that for all  $d \in \mathbb{N}$ ,  $\varepsilon \in (0,1]$  it holds that  $\sum_{n=1}^{N_{d,\varepsilon}} \text{RV}_{d,n,\lfloor n^{1/4} \rfloor} \leq cd^c \varepsilon^{-(2+\delta)}$  and

$$\begin{aligned} & \sup_{n \in \mathbb{N} \cap [N_{d,\varepsilon}, \infty)} \left[ \mathbb{E}[|\mathbf{U}_{n,\lfloor n^{1/4} \rfloor}^{d,0,0}(T, \xi_d) - u_d(T, \xi_d)|^2] \right. \\ & \quad \left. + \max_{i \in \{1,2,\dots,d\}} \mathbb{E}[|\mathbf{U}_{n,\lfloor n^{1/4} \rfloor}^{d,0,i}(T, \xi_d) - (\frac{\partial}{\partial x_i} u_d)(T, \xi_d)|^2] \right]^{1/2} \leq \varepsilon. \end{aligned} \quad (181)$$

**Proof of Corollary 5.5** Throughout this proof let  $\kappa = 2\lambda + \sum_{i=1}^{\infty} \frac{1}{j^2}$ . Observe that the fact that for all  $z, \mathfrak{z} \in \mathbb{R}$  it holds that  $|\sin(z) - \sin(\mathfrak{z})| \leq |z - \mathfrak{z}|$  and the assumption that for all  $d \in \mathbb{N}$ ,  $x = (x_1, x_2, \dots, x_d)$ ,  $\mathfrak{x} = (\mathfrak{x}_1, \mathfrak{x}_2, \dots, \mathfrak{x}_d) \in \mathbb{R}^d$  it holds that  $|g_d(x) - g_d(\mathfrak{x})| \leq \lambda d^\lambda (\sum_{j=1}^d |x_j - \mathfrak{x}_j|)$  imply that for all  $d \in \mathbb{N}$ ,  $x = (x_1, x_2, \dots, x_d)$ ,  $\mathfrak{x} = (\mathfrak{x}_1, \mathfrak{x}_2, \dots, \mathfrak{x}_d)$ ,  $z = (z_1, z_2, \dots, z_d)$ ,  $\mathfrak{z} = (\mathfrak{z}_1, \mathfrak{z}_2, \dots, \mathfrak{z}_d) \in \mathbb{R}^d$  it holds that

$$\begin{aligned} \max\{|\sin(z_1) - \sin(\mathfrak{z}_1)|, |g_d(x) - g_d(\mathfrak{x})|\} & \leq \max\{|z_1 - \mathfrak{z}_1|, \lambda d^\lambda \sum_{j=1}^d |x_j - \mathfrak{x}_j|\} \\ & \leq |z_1 - \mathfrak{z}_1| + \lambda d^\lambda (\sum_{j=1}^d |x_j - \mathfrak{x}_j|) \\ & \leq \sum_{j=1}^d \left[ \left( \lambda d^{\lambda-\kappa} + \frac{1}{j^2} \right) (d^\kappa |x_j - \mathfrak{x}_j| + |z_j - \mathfrak{z}_j|) \right]. \end{aligned} \quad (182)$$

Moreover, the assumption that for all  $d \in \mathbb{N}$  it holds that  $|g_d(0)| + \max_{i \in \{1,2,\dots,d\}} |\xi_{d,i}| \leq \lambda d^\lambda$  implies that

$$\begin{aligned} & d^{-\kappa} \left( |g_d(0)| + |\sin(0)| + \max_{i \in \{1,2,\dots,d\}} |\xi_{d,i}| \right) + \left[ \sum_{j=1}^d \left( \lambda d^{\lambda-\kappa} + \frac{1}{j^2} \right) \right] \\ & \leq \lambda d^{\lambda-\kappa} + \lambda d^{\lambda-\kappa+1} + \left[ \sum_{i=1}^{\infty} \frac{1}{j^2} \right] \leq 2\lambda + \left[ \sum_{i=1}^{\infty} \frac{1}{j^2} \right] = \kappa. \end{aligned} \quad (183)$$

This, (182), and Corollary 5.4 establish (181). This completes the proof of Corollary 5.5.  $\square$

**Acknowledgements** We thank two anonymous reviewers and the handling editor for very helpful comments and suggestions. The first author acknowledges funding by the Deutsche Forschungsgesellschaft (DFG) via RTG 2131 *High-dimensional Phenomena in Probability – Fluctuations and Discontinuity* and via research grant HU 1889/6-1. The second author acknowledges funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany’s Excellence Strategy EXC 2044-390685587, Mathematics Muenster: Dynamics-Geometry-Structure.

**Funding** Open Access funding enabled and organized by Projekt DEAL.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

1. BECK, C., BECKER, S., CHERIDITO, P., JENTZEN, A., AND NEUFELD, A. Deep splitting method for parabolic PDEs. [arXiv:1907.03452](https://arxiv.org/abs/1907.03452) (2019). Revision requested from SIAM Journal of Scientific Computing.
2. BECK, C., BECKER, S., GROHS, P., JAAFARI, N., AND JENTZEN, A. Solving the Kolmogorov PDE by means of deep learning. [arXiv:1806.00421](https://arxiv.org/abs/1806.00421) (2018). Accepted in Journal of Scientific Computing.
3. BECK, C., E, W., AND JENTZEN, A. Machine learning approximation algorithms for high-dimensional fully nonlinear partial differential equations and second-order backward stochastic differential equations. *Journal of Nonlinear Science* 29, 4 (2019), 1563–1619.
4. BECK, C., HORNUNG, F., HUTZENTHALER, M., JENTZEN, A., AND KRUSE, T. Overcoming the curse of dimensionality in the numerical approximation of Allen-Cahn partial differential equations via truncated full-history recursive multilevel Picard approximations. *Journal of Numerical Mathematics* 28, 4 (2020), 197–222.
5. BECKER, S., BRAUNWARTH, R., HUTZENTHALER, M., JENTZEN, A., AND VON WURSTEMBERGER, P. Numerical simulations for full history recursive multilevel Picard approximations for systems of high-dimensional partial differential equations. *Commun. Comput. Phys.* 28, 5 (2020), 2109–2138.
6. BECKER, S., CHERIDITO, P., AND JENTZEN, A. Deep optimal stopping. *Journal of Machine Learning Research* 20, 74 (2019), 1–25.
7. BECKER, S., CHERIDITO, P., JENTZEN, A., AND WELTI, T. Solving high-dimensional optimal stopping problems using deep learning. Cambridge University Press (2021). <https://www.cambridge.org/core/journals/european-journal-of-applied-mathematics/article/solving-highdimensional-optimal-stopping-problems-using-deep-learning/A632772461C859353E6F8A7DAB8A1769>
8. BERG, J., AND NYSTRÖM, K. A unified deep artificial neural network approach to partial differential equations in complex geometries. *Neurocomputing* 317 (2018), 28–41.
9. BERNER, J., GROHS, P., AND JENTZEN, A. Analysis of the generalization error: Empirical risk minimization over deep artificial neural networks overcomes the curse of dimensionality in the numerical approximation of Black-Scholes partial differential equations. *SIAM Journal on Mathematics of Data Science* 2, 3 (2020), 631–657.
10. BOUCHARD, B., TAN, X., WARIN, X., AND ZOU, Y. Numerical approximation of BSDEs using local polynomial drivers and branching processes. *Monte Carlo Methods and Applications* 23, 4 (2017), 241–263.
11. BOUCHARD, B., AND TOUZI, N. Discrete-time approximation and Monte-Carlo simulation of backward stochastic differential equations. *Stochastic Processes and their applications* 111, 2 (2004), 175–206.
12. BRIAND, P., AND LABART, C. Simulation of BSDEs by Wiener chaos expansion. *The Annals of Applied Probability* 24, 3 (2014), 1129–1171.
13. CHAN-WAI-NAM, Q., MIKAEL, J., AND WARIN, X. Machine learning for semi linear PDEs. *J. Sci. Comput.* 79, 3 (2019), 1667–1712.
14. CHEN, Y., AND WAN, J. W. Deep neural network framework based on backward stochastic differential equations for pricing and hedging american options in high dimensions. [arXiv:1909.11532](https://arxiv.org/abs/1909.11532) (2019).
15. DA PRATO, G., AND ZABCZYK, J. Differentiability of the Feynman-Kac semigroup and a control application. *Atti Accad. Naz. Lincei Cl. Sci. Fis. Mat. Natur. Rend. Lincei (9) Mat. Appl.* 8, 3 (1997), 183–188.

16. DOCKHORN, T. A discussion on solving partial differential equations using neural networks. [arXiv:1904.07200](https://arxiv.org/abs/1904.07200) (2019).
17. E, W., HAN, J., AND JENTZEN, A. Deep learning-based numerical methods for high-dimensional parabolic partial differential equations and backward stochastic differential equations. *Commun. Math. Stat.* 5, 4 (2017), 349–380.
18. E, W., HAN, J., AND JENTZEN, A. Algorithms for Solving High Dimensional PDEs: From Nonlinear Monte Carlo to Machine Learning. [arXiv:2008.13333](https://arxiv.org/abs/2008.13333) (2020).
19. E, W., HUTZENTHALER, M., JENTZEN, A., AND KRUSE, T. Multilevel Picard iterations for solving smooth semilinear parabolic heat equations. [arXiv:1607.03295](https://arxiv.org/abs/1607.03295) (2016). Springer Nature Partial Differential Equations and Applications (in press).
20. E, W., HUTZENTHALER, M., JENTZEN, A., AND KRUSE, T. On multilevel Picard numerical approximations for high-dimensional nonlinear parabolic partial differential equations and high-dimensional nonlinear backward stochastic differential equations. *Journal of Scientific Computing* 79, 3 (2019), 1534–1571.
21. E, W., AND YU, B. The Deep Ritz method: A deep learning-based numerical algorithm for solving variational problems. *Communications in Mathematics and Statistics* 6, 1 (2018), 1–12.
22. EL KAROUI, N., PENG, S., AND QUENEZ, M. C. Backward stochastic differential equations in finance. *Mathematical finance* 7, 1 (1997), 1–71.
23. ELBRÄCHTER, D., GROHS, P., JENTZEN, A., AND SCHWAB, C. DNN Expression Rate Analysis of High-Dimensional PDEs: Application to Option Pricing. *Constr. Approx.* (2021). <https://doi.org/10.1007/s00365-021-09541-6>
24. FARAHMAND, A.-M., NABI, S., AND NIKOVSKI, D. Deep reinforcement learning for partial differential equation control. *2017 American Control Conference (ACC)* (2017), 3120–3127.
25. FUJII, M., TAKAHASHI, A., AND TAKAHASHI, M. Asymptotic Expansion as Prior Knowledge in Deep Learning Method for high dimensional BSDEs. [arXiv:1710.07030](https://arxiv.org/abs/1710.07030) (2017).
26. GEISS, C., AND LABART, C. Simulation of BSDEs with jumps by Wiener chaos expansion. *Stochastic processes and their applications* 126, 7 (2016), 2123–2162.
27. GILES, M. B. Multilevel Monte Carlo path simulation. *Oper. Res.* 56, 3 (2008), 607–617.
28. GILES, M. B., JENTZEN, A., AND WELTI, T. Generalised multilevel Picard approximations. [arXiv:1911.03188](https://arxiv.org/abs/1911.03188) (2019). Revision requested from IMA J. Num. Anal.
29. GOBET, E., LEMOR, J.-P., AND WARIN, X. A regression-based Monte Carlo method to solve backward stochastic differential equations. *The Annals of Applied Probability* 15, 3 (2005), 2172–2202.
30. GOBET, E., TURKEDJIEV, P., ET AL. Approximation of backward stochastic differential equations using malliavin weights and least-squares regression. *Bernoulli* 22, 1 (2016), 530–562.
31. GOUDENÈGE, L., MOLENT, A., AND ZANETTE, A. Machine Learning for Pricing American Options in High Dimension. [arXiv:1903.11275](https://arxiv.org/abs/1903.11275) (2019), 11 pages.
32. GROHS, P., HORNUNG, F., JENTZEN, A., AND VON WURSTEMBERGER, P. A proof that artificial neural networks overcome the curse of dimensionality in the numerical approximation of Black-Scholes partial differential equations. [arXiv:1809.02362](https://arxiv.org/abs/1809.02362) (2019). Accepted in Mem. Amer. Math. Soc.
33. GROHS, P., HORNUNG, F., JENTZEN, A., AND ZIMMERMANN, P. Space-time error estimates for deep neural network approximations for differential equations. [arXiv:1908.03833](https://arxiv.org/abs/1908.03833) (2019).
34. GROHS, P., JENTZEN, A., AND SALIMOVA, D. Deep neural network approximations for Monte Carlo algorithms. [arXiv:1908.10828](https://arxiv.org/abs/1908.10828) (2019). Accepted in Springer Nature Partial Differential Equations and Applications.
35. HAN, J., JENTZEN, A., AND E, W. Solving high-dimensional partial differential equations using deep learning. *Proceedings of the National Academy of Sciences* 115, 34 (2018), 8505–8510.
36. HAN, J., AND LONG, J. Convergence of the Deep BSDE Method for Coupled FBSDEs. [arXiv:1811.01165](https://arxiv.org/abs/1811.01165) (2018).
37. HEINRICH, S. Monte Carlo complexity of global solution of integral equations. *J. Complexity* 14, 2 (1998), 151–175.
38. HEINRICH, S., AND SINDAMBIWE, E. Monte Carlo complexity of parametric integration. *J. Complexity* 15, 3 (1999), 317–341. Dagstuhl Seminar on Algorithms and Complexity for Continuous Problems (1998).
39. HENRY-LABORDÈRE, P. Counterparty risk valuation: a marked branching diffusion approach. [arXiv:1203.2369](https://arxiv.org/abs/1203.2369) (2012).

40. HENRY-LABORDÈRE, P. Deep Primal-Dual Algorithm for BSDEs: Applications of Machine Learning to CVA and IM. Available at SSRN: <https://doi.org/10.2139/ssrn.3071506> (2017).
41. HENRY-LABORDÈRE, P., OUDJANE, N., TAN, X., TOUZI, N., AND WARIN, X. Branching diffusion representation of semilinear PDEs and Monte Carlo approximation. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques* 55, 1 (2019), 184–210.
42. HENRY-LABORDÈRE, P., TAN, X., AND TOUZI, N. A numerical algorithm for a class of BSDEs via the branching process. *Stochastic Process. Appl.* 124, 2 (2014), 1112–1140.
43. HURÉ, C., PHAM, H., AND WARIN, X. Some machine learning schemes for high-dimensional nonlinear PDEs. [arXiv:1902.01599](https://arxiv.org/abs/1902.01599) (2019).
44. HUTZENTHALER, M., JENTZEN, A., KRUSE, T., AND NGUYEN, T. A. A proof that rectified deep neural networks overcome the curse of dimensionality in the numerical approximation of semilinear heat equations. *SN Partial Differential Equations and Applications* 1 (2020), 1–34.
45. HUTZENTHALER, M., JENTZEN, A., KRUSE, T., NGUYEN, T. A., AND VON WURSTEMBERGER, P. Overcoming the curse of dimensionality in the numerical approximation of semilinear parabolic partial differential equations. *Proceedings of the Royal Society A* 476, 2244 (2020), 20190630.
46. HUTZENTHALER, M., JENTZEN, A., AND VON WURSTEMBERGER, P. Overcoming the curse of dimensionality in the approximative pricing of financial derivatives with default risks. *Electronic Journal of Probability* 25 (2020).
47. HUTZENTHALER, M., AND KRUSE, T. Multilevel Picard approximations of high-dimensional semilinear parabolic differential equations with gradient-dependent nonlinearities. *SIAM Journal on Numerical Analysis* 58, 2 (2020), 929–961.
48. JACQUIER, A., AND OUMGARI, M. Deep PPDEs for rough local stochastic volatility. [arXiv:1906.02551](https://arxiv.org/abs/1906.02551) (2019).
49. JENTZEN, A., SALIMOVA, D., AND WELTI, T. A proof that deep artificial neural networks overcome the curse of dimensionality in the numerical approximation of Kolmogorov partial differential equations with constant diffusion and nonlinear drift coefficients. [arXiv:1809.07321](https://arxiv.org/abs/1809.07321) (2018). Accepted in Communications in Mathematical Sciences.
50. JIANYU, L., SIWEI, L., YINGJIAN, Q., AND YAPING, H. Numerical solution of elliptic partial differential equation using radial basis function neural networks. *Neural Networks* 16, 5 (2003), 729–734.
51. KUTYNIOK, G., PETERSEN, P., RASLAN, M., AND SCHNEIDER, R. A theoretical analysis of deep neural networks and parametric PDEs. [arXiv:1904.00377](https://arxiv.org/abs/1904.00377) (2019).
52. LAGARIS, I. E., LIKAS, A., AND FOTIADIS, D. I. Artificial neural networks for solving ordinary and partial differential equations. *IEEE transactions on neural networks* 9 (5) (1998), 987–1000.
53. LEMOR, J.-P., GOBET, E., AND WARIN, X. Rate of convergence of an empirical regression method for solving generalized backward stochastic differential equations. *Bernoulli* 12, 5 (2006), 889–916.
54. LONG, Z., LU, Y., MA, X., AND DONG, B. PDE-Net: Learning PDEs from Data. In *Proceedings of the 35th International Conference on Machine Learning* (2018), pp. 3208–3216.
55. LYÉ, K. O., MISHRA, S., AND RAY, D. Deep learning observables in computational fluid dynamics. [arXiv:1903.03040](https://arxiv.org/abs/1903.03040) (2019).
56. MAGILL, M., QURESHI, F., AND DE HAAN, H. W. Neural networks trained to solve differential equations learn general representations. In *Advances in Neural Information Processing Systems* (2018), pp. 4071–4081.
57. MEADE, JR., A. J., AND FERNÁNDEZ, A. A. The numerical solution of linear ordinary differential equations by feedforward neural networks. *Math. Comput. Modelling* 19, 12 (1994), 1–25.
58. PHAM, H., AND WARIN, X. Neural networks-based backward scheme for fully nonlinear PDEs. [arXiv:1908.00412](https://arxiv.org/abs/1908.00412) (2019).
59. QI, F. Bounds for the ratio of two gamma functions. *Journal of Inequalities and Applications* 2010, 1 (2010), 493058.
60. RAISI, M. Forward-Backward Stochastic Neural Networks: Deep Learning of High-dimensional Partial Differential Equations. [arXiv:1804.07010](https://arxiv.org/abs/1804.07010) (2018).
61. REISINGER, C., AND ZHANG, Y. Rectified deep neural networks overcome the curse of dimensionality for nonsmooth value functions in zero-sum games of nonlinear stiff systems. [arXiv:1903.06652](https://arxiv.org/abs/1903.06652) (2019).
62. SIRIGNANO, J., AND SPILIOPOULOS, K. DGM: A deep learning algorithm for solving partial differential equations. [arXiv:1708.07469](https://arxiv.org/abs/1708.07469) (2017).
63. UCHIYAMA, T., AND SONEHARA, N. Solving inverse problems in nonlinear PDEs by recurrent neural networks. In *IEEE International Conference on Neural Networks* (1993), IEEE, pp. 99–102.

64. WENDEL, J. Note on the gamma function. *The American Mathematical Monthly* 55, 9 (1948), 563–564.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.