Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Seminar for Applied Mathematics

Overcoming the curse of dimensionality in the numerical approximation of parabolic partial differential equations with gradient-dependent nonlinearities

M. Hutzenthaler and A. Jentzen and Th. Kruse

Research Report No. 2019-63

December 2019
Latest revision: January 2020

Seminar für Angewandte Mathematik
Eidgenössische Technische Hochschule
CH-8092 Zürich
Switzerland

Overcoming the curse of dimensionality in the numerical approximation of parabolic partial differential equations with gradient-dependent nonlinearities

Martin Hutzenthaler[1], Arnulf Jentzen[2,3], and Thomas Kruse[3]

[1] Faculty of Mathematics, University of Duisburg-Essen,
45117 Essen, Germany, e-mail: martin.hutzenthaler@uni-due.de

[2] SAM, Department of Mathematics, ETH Zürich,
8092 Zürich, Switzerland, e-mail: arnulf.jentzen@sam.math.ethz.ch

[3] Faculty of Mathematics and Computer Science, University of Münster,
48149 Münster, Germany, e-mail: ajentzen@uni-muenster.de

[4] Institute of Mathematics, University of Gießen,
35392 Gießen, Germany, e-mail: thomas.kruse@math.uni-giessen.de

December 10, 2019

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 ε. 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 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.

Contents

AMS 2010 subject classification: 65M75
Keywords and phrases: curse of dimensionality, partial differential equation, PDE, backward stochastic differential equation, BSDE, multilevel Picard, multilevel Monte Carlo, gradient-dependent nonlinearity

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 emerge in biology in models that aim to better understand biodiversity in ecosystems, 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 models that aim to better understand biodiversity the dimension of the PDE corresponds to the number of traits of the considered species in the considered ecosystem, 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 ε. 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 [10, 25, 47, 26, 11, 23] for BSDE based approximation methods for PDEs in which nested conditional expectations are discretized through suitable regression methods, we refer, e.g., to [33, 36, 35, 9] for branching diffusion approximation methods for PDEs, we refer, e.g., to [57, 51, 46, 44, 15, 3, 18, 22, 31, 34, 21, 54, 56, 5, 2, 50, 48, 32, 7, 52, 49, 27, 1, 42, 37, 12, 6, 14, 13, 30] for deep learning based approximation methods for PDEs, and we refer to [17, 41, 40, 4, 24] for numerical simulations, approximation results, and extensions of the in [16, 39] 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 are actually not working anymore in the case where the time horizon $T \in (0, \infty)$ exceeds a certain threshold (cf., e.g., [35, Theorem 3.12]). For MLP approximation methods it has been recently shown in [39, 40, 4] 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 [57, 51, 46, 44, 15, 3, 18, 22, 31, 34, 21, 54, 56, 5, 2, 50, 48, 32, 7, 52, 49, 27, 1, 42, 37, 12, 6, 14, 13, 30]) but so far there is only a 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. [38] and, e.g., [32, 56, 8, 20, 28, 43, 45, 55, 29]). To sum up, to the best of our knowledge until today the MLP approximation method (see [39]) 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 [39, 40, 4, 24] 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 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 Section 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 ε. 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 $d \in \mathbb{N}$, $t \in [0,T)$, $x = (x_1,x_2,\ldots,x_d)$, $\mathfrak{x} = (\mathfrak{x}_1,\mathfrak{x}_2,\ldots,\mathfrak{x}_d)$, $z = (z_1,z_2,\ldots,z_d)$, $\mathfrak{z} = (\mathfrak{z}_1,\mathfrak{z}_2,\ldots,\mathfrak{z}_d) \in \mathbb{R}^d$, $y,\mathfrak{y} \in \mathbb{R}$ that*

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

$$\big(\tfrac{\partial}{\partial t}u_d\big)(t,x) = (\Delta_x u_d)(t,x) + f_d\big(u(t,x),(\nabla_x u_d)(t,x)\big), \qquad u_d(0,x) = g_d(x), \tag{2}$$

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 $Z^{d,\theta}\colon \Omega \to \mathbb{R}^d$, $d \in \mathbb{N}$, $\theta \in \Theta$, be i.i.d. standard normal random variables, let $\mathfrak{r}^\theta\colon \Omega \to (0,1)$, $\theta \in \Theta$, 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 $\mathbf{U}^{d,\theta}_{n,M} = (\mathbf{U}^{d,\theta,0}_{n,M}, \mathbf{U}^{d,\theta,1}_{n,M},\ldots,\mathbf{U}^{d,\theta,d}_{n,M})\colon (0,T]\times\mathbb{R}^d\times\Omega \to \mathbb{R}^{1+d}$, $n,M,d \in \mathbb{Z}$, $\theta\in\Theta$, satisfy for all $n,M,d\in\mathbb{N}$, $\theta\in\Theta$, $t\in(0,T]$, $x \in \mathbb{R}^d$ that $\mathbf{U}^{d,\theta}_{-1,M}(t,x) = \mathbf{U}^{d,\theta}_{0,M}(t,x) = 0$ and

$$\begin{aligned} \mathbf{U}^{d,\theta}_{n,M}(t,x) = (g_d(x),0) &+ \sum_{i=1}^{M^n} \tfrac{1}{M^n}\big(g_d(x+[2t]^{1/2}Z^{(\theta,0,-i)}) - g_d(x)\big)\big(1,[2t]^{-1/2}Z^{d,(\theta,0,-i)}\big) \\ &+ \sum_{l=0}^{n-1}\sum_{i=1}^{M^{n-l}} \tfrac{2t[\mathfrak{r}^{(\theta,l,i)}]^{1/2}}{M^{n-l}}\big[f_d\big(\mathbf{U}^{d,(\theta,l,i)}_{l,M}(t(1-\mathfrak{r}^{(\theta,l,i)}),x+[2t\mathfrak{r}^{(\theta,l,i)}]^{1/2}Z^{d,(\theta,l,i)})\big) \\ &- \mathbb{1}_{\mathbb{N}}(l)f_d\big(\mathbf{U}^{d,(\theta,-l,i)}_{l-1,M}(t(1-\mathfrak{r}^{(\theta,l,i)}),x+[2t\mathfrak{r}^{(\theta,l,i)}]^{1/2}Z^{d,(\theta,l,i)})\big)\big]\big(1,[2t\mathfrak{r}^{(\theta,l,i)}]^{-1/2}Z^{d,(\theta,l,i)}\big), \end{aligned} \tag{3}$$

and for every $d,M,n\in\mathbb{N}$ let $\mathrm{RV}_{d,n,M}\in\mathbb{N}$ be the number of realizations of scalar random variables which are used to compute one realization of $\mathbf{U}^{d,0}_{n,M}(T,0)\colon\Omega\to\mathbb{R}$ (cf. (160) *for a precise definition). Then there exist $c\in\mathbb{R}$ and $N = (N_{d,\varepsilon})_{(d,\varepsilon)\in\mathbb{N}\times(0,1]}\colon \mathbb{N}\times(0,1]\to\mathbb{N}$ such that for all $d\in\mathbb{N}$, $\varepsilon\in(0,1]$ it holds that $\sum_{n=1}^{N_{d,\varepsilon}}\mathrm{RV}_{d,n,\lfloor n^{1/4}\rfloor} \leq cd^c\varepsilon^{-(2+\delta)}$ and*

$$\sup_{n\in\mathbb{N}\cap[N_{d,\varepsilon},\infty)}\Big[\mathbb{E}\big[|\mathbf{U}^{d,0,0}_{n,\lfloor n^{1/4}\rfloor}(T,0) - u_d(T,0)|^2\big] + \max_{i\in\{1,2,\ldots,d\}}\mathbb{E}\big[|\mathbf{U}^{d,0,i}_{n,\lfloor n^{1/4}\rfloor}(T,0) - (\tfrac{\partial}{\partial x_i}u_d)(T,0)|^2\big]\Big]^{1/2} \leq \varepsilon. \tag{4}$$

Theorem 1.1 is an immediate consequence of Corollary 5.4 in Section 5 below. Corollary 5.4, in turn, follows from Theorem 5.2 in Section 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 proves 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 $u_d\colon[0,T]\times\mathbb{R}^d\to\mathbb{R}$, $d\in\mathbb{N}$, 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 $f_d\colon\mathbb{R}\times\mathbb{R}^d\to\mathbb{R}$, $d\in\mathbb{N}$, describe the nonlinearities of the PDEs in (2) and the functions $g_d\colon\mathbb{R}^d\to\mathbb{R}$, $d\in\mathbb{N}$, 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 $\lfloor n^{1/4}\rfloor = \max([0,n^{1/4}]\cap\mathbb{N})$.

Theorem 1.1 above in this introductory section is a special case of the more general approximation results in Section 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, Theorem 5.2, and Corollary 5.4 in Section 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 $\mathfrak{r}^\theta\colon\Omega\to(0,1)$, $\theta\in(0,1)$, of i.i.d. random variables satisfies for all $b\in(0,1)$ that $\mathbb{P}(\mathfrak{r}^0\leq b)=\sqrt{b}$ 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 $\mathbb{P}(\mathfrak{r}^0\leq b) = b^\alpha$ (see, e.g., (146) in Theorem 5.2). Furthermore, the more general approximation result in Corollary 5.1 in Section 5 also provides an explicit upper bound for the constant $c \in \mathbb{R}$ in Theorem 1.1 above (see (133) in Corollary 5.1).

The remainder of this article is organized as follows. In Section 2 we establish a few identities and upper bounds for certain iterated deterministic integrals. The results of Section 2 are then used in Section 3 in which we introduce and analyze the considered MLP approximation methods. In Section 4 we establish suitable a priori bounds for exact solutions of PDEs of the form (2). In Section 5 we combine the findings of Sections 3 and 4 to etablish 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 Section 3 below. Our proof of Corollary 2.5 employs a few elementary identities and estimates for certain deterministic iterated integrals which are provided in Lemma 2.1, Lemma 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 $\rho\colon (0,1) \to (0,\infty)$ be $\mathcal{B}((0,1))/\mathcal{B}((0,\infty))$-measurable, and let $\varrho\colon [0,T]^2 \to (0,\infty)$ 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*

$$\int_{s_0}^{T} \frac{1}{(s_1-s_0)^\beta[\varrho(s_0,s_1)]^\gamma} \int_{s_1}^{T} \frac{1}{(s_2-s_1)^\beta[\varrho(s_1,s_2)]^\gamma} \cdots \int_{s_{j-1}}^{T} \frac{1}{(s_j-s_{j-1})^\beta[\varrho(s_{j-1},s_j)]^\gamma}\, ds_j \ldots ds_2\, ds_1 = (T-s_0)^{j(1+\gamma-\beta)} \left[\prod_{i=0}^{j-1} \int_0^1 \frac{(1-s)^{i(1+\gamma-\beta)}}{s^\beta[\rho(s)]^\gamma}\, ds\right]. \tag{5}$$

Proof of Lemma 2.1. We prove (5) 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

$$\int_{s_0}^{T} \frac{1}{(s_1-s_0)^\beta[\varrho(s_0,s_1)]^\gamma}\, ds_1 = \int_0^1 \frac{(T-s_0)^{1-\beta}}{z[\varrho(s_0,s_0+z(T-s_0))]^\gamma}\, dz = (T-s_0)^{1+\gamma-\beta}\int_0^1 \frac{1}{z^\beta[\rho(z)]^\gamma}\, dz. \tag{6}$$

This proves (5) 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

$$\begin{aligned}
&\int_{s_0}^{T} \frac{1}{(s_1-s_0)^\beta[\varrho(s_0,s_1)]^\gamma} \left[\int_{s_1}^{T} \frac{1}{(s_2-s_1)^\beta[\varrho(s_1,s_2)]^\gamma} \cdots \int_{s_j}^{T} \frac{1}{(s_{j+1}-s_j)^\beta[\varrho(s_j,s_{j+1})]^\gamma}\, ds_{j+1} \ldots ds_2\right] ds_1 \\
&= \int_{s_0}^{T} \frac{(T-s_1)^{j(1+\gamma-\beta)}}{(s_1-s_0)^\beta[\varrho(s_0,s_1)]^\gamma} \left[\prod_{i=0}^{j-1}\int_0^1 \frac{(1-s)^{i(1+\gamma-\beta)}}{s^\beta[\rho(s)]^\gamma}\, ds\right] ds_1 \\
&= \left[\prod_{i=0}^{j-1}\int_0^1 \frac{(1-s)^{i(1+\gamma-\beta)}}{s^\beta[\rho(s)]^\gamma}\, ds\right] \int_0^1 \frac{(T-s_0)^{j(1+\gamma-\beta)}(1-z)^{j(1+\gamma-\beta)}}{(T-s_0)^\beta z^\beta[\varrho(s_0,s_0+z(T-s_0))]^\gamma}(T-s_0)dz \\
&= (T-s_0)^{(j+1)(1+\gamma-\beta)} \left[\prod_{i=0}^{j}\int_0^1 \frac{(1-s)^{i(1+\gamma-\beta)}}{s^\beta[\rho(s)]^\gamma}\, ds\right].
\end{aligned} \tag{7}$$

Induction thus proves (5). This completes the proof of Lemma 2.1. □

Lemma 2.2. *Let $\alpha \in (0,1)$, $T, \gamma \in (0,\infty)$, $\beta \in (0, \alpha\gamma+1)$ and let $\rho\colon (0,1) \to (0,\infty)$ and $\varrho\colon [0,T]^2 \to (0,\infty)$ 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*

$$\int_{s_0}^{T} \frac{1}{(s_1-s_0)^\beta[\varrho(s_0,s_1)]^\gamma} \int_{s_1}^{T} \frac{1}{(s_2-s_1)^\beta[\varrho(s_1,s_2)]^\gamma} \cdots \int_{s_{j-1}}^{T} \frac{1}{(s_j-s_{j-1})^\beta[\varrho(s_{j-1},s_j)]^\gamma}\, ds_j \ldots ds_2\, ds_1 = \left[\frac{(T-s_0)^{(1+\gamma-\beta)}\Gamma(\alpha\gamma-\beta+1)}{(1-\alpha)^\gamma}\right]^j \left[\prod_{i=0}^{j-1} \frac{\Gamma(i(1+\gamma-\beta)+1)}{\Gamma(\alpha\gamma-\beta+i(1+\gamma-\beta)+2)}\right]. \tag{8}$$

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

$$B(x,y) = \int_0^1 s^{x-1}(1-s)^{y-1}\, ds. \tag{9}$$

Note that (9) 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

$$\begin{aligned}\int_0^1 \frac{(1-s)^{i(1+\gamma-\beta)}}{s^\beta [\rho(s)]^\gamma}\, ds &= \frac{1}{(1-\alpha)^\gamma} \int_0^1 s^{\alpha\gamma-\beta}(1-s)^{i(1+\gamma-\beta)}\, ds = \frac{B(\alpha\gamma-\beta+1, i(1+\gamma-\beta)+1)}{(1-\alpha)^\gamma} \\ &= \frac{\Gamma(\alpha\gamma-\beta+1)\Gamma(i(1+\gamma-\beta)+1)}{(1-\alpha)^\gamma \Gamma(\alpha\gamma-\beta+i(1+\gamma-\beta)+2)}.\end{aligned} \tag{10}$$

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

$$\begin{aligned}&\int_{s_0}^T \frac{1}{(s_1-s_0)^\beta [\varrho(s_0,s_1)]^\gamma} \int_{s_1}^T \frac{1}{(s_2-s_1)^\beta [\varrho(s_1,s_2)]^\gamma} \cdots \int_{s_{j-1}}^T \frac{1}{(s_j-s_{j-1})^\beta [\varrho(s_{j-1},s_j)]^\gamma}\, ds_j \ldots ds_2\, ds_1 \\ &= (T-s_0)^{j(1+\gamma-\beta)} \left[\prod_{i=0}^{j-1} \int_0^1 \frac{(1-s)^{i(1+\gamma-\beta)}}{s^\beta [\rho(s)]^\gamma}\, ds\right] \\ &= (T-s_0)^{j(1+\gamma-\beta)} \left[\prod_{i=0}^{j-1} \frac{\Gamma(\alpha\gamma-\beta+1)\Gamma(i(1+\gamma-\beta)+1)}{(1-\alpha)^\gamma \Gamma(\alpha\gamma-\beta+i(1+\gamma-\beta)+2)}\right] \\ &= \left[\frac{(T-s_0)^{(1+\gamma-\beta)}\Gamma(\alpha\gamma-\beta+1)}{(1-\alpha)^\gamma}\right]^j \left[\prod_{i=0}^{j-1} \frac{\Gamma(i(1+\gamma-\beta)+1)}{\Gamma(\alpha\gamma-\beta+i(1+\gamma-\beta)+2)}\right].\end{aligned} \tag{11}$$

This establishes (8). The proof of Lemma 2.2 is thus completed. □

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 $\rho\colon (0,1) \to (0,\infty)$ and $\varrho\colon [0,T]^2 \to (0,\infty)$ 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*

$$\begin{aligned}&\int_{s_0}^T \frac{1}{(s_1-s_0)^\beta [\varrho(s_0,s_1)]^\gamma} \int_{s_1}^T \frac{1}{(s_2-s_1)^\beta [\varrho(s_1,s_2)]^\gamma} \cdots \int_{s_{j-1}}^T \frac{1}{(s_j-s_{j-1})^\beta [\varrho(s_{j-1},s_j)]^\gamma}\, ds_j \ldots ds_2\, ds_1 \\ &\leq \left[\frac{(T-s_0)^{(1+\gamma-\beta)}\Gamma(\alpha\gamma-\beta+1)}{(1-\alpha)^\gamma (1+\gamma-\beta)^{\alpha\gamma-\beta+1}}\right]^j \left[e^{1+\gamma-\beta}((1+\gamma-\beta)(j-1)+1)\right]^{\frac{(\beta-\alpha\gamma)(\alpha\gamma-\beta+1)}{1+\gamma-\beta}} \left[\frac{\Gamma(\frac{1}{1+\gamma-\beta})}{\Gamma(j+\frac{1}{1+\gamma-\beta})}\right]^{\alpha\gamma-\beta+1}.\end{aligned} \tag{12}$$

Proof of Corollary 2.3. First, observe that Wendel's inequality for the gamma function (see, e.g., Wendel [58] and Qi [53, Section 2.1]) ensures that for all $x \in (0,\infty)$, $s \in [0,1]$ it holds that

$$\frac{\Gamma(x)}{\Gamma(x+s)} \leq \frac{1}{x^s}\left[\frac{x+s}{x}\right]^{1-s}. \tag{13}$$

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

$$\sum_{i=0}^{j-1} \frac{1}{i+\lambda} = \frac{1}{\lambda} + \sum_{i=1}^{j-1} \frac{1}{i+\lambda} \leq \frac{1}{\lambda} + \sum_{i=1}^{j-1} \int_i^{i+1} \frac{1}{s-1+\lambda}\, ds = \frac{1}{\lambda} + \int_1^j \frac{1}{s-1+\lambda}\, ds = \frac{1}{\lambda} + \ln\!\left(\frac{j-1}{\lambda}+1\right). \tag{14}$$

Combining this, (13), 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

$$
\begin{aligned}
&\left[\prod_{i=0}^{j-1} \frac{\Gamma(i(1+\gamma-\beta)+1)}{\Gamma(\alpha\gamma-\beta+i(1+\gamma-\beta)+2)}\right] \\
&\leq \left[\prod_{i=0}^{j-1} \left(\frac{\alpha\gamma-\beta+i(1+\gamma-\beta)+2}{i(1+\gamma-\beta)+1}\right)^{1-(\alpha\gamma-\beta+1)} \frac{1}{(i(1+\gamma-\beta)+1)^{\alpha\gamma-\beta+1}}\right] \\
&= \left[\prod_{i=0}^{j-1} \left(1+\frac{\alpha\gamma-\beta+1}{i(1+\gamma-\beta)+1}\right)^{\beta-\alpha\gamma} \frac{1}{(i(1+\gamma-\beta)+1)^{\alpha\gamma-\beta+1}}\right] \\
&\leq \left[\prod_{i=0}^{j-1} e^{\frac{(\beta-\alpha\gamma)(\alpha\gamma-\beta+1)}{i(1+\gamma-\beta)+1}} \frac{1}{(i(1+\gamma-\beta)+1)^{\alpha\gamma-\beta+1}}\right] \\
&= e^{\frac{(\beta-\alpha\gamma)(\alpha\gamma-\beta+1)}{1+\gamma-\beta} \sum_{i=0}^{j-1} \frac{1}{i+\frac{1}{1+\gamma-\beta}}} \left[\prod_{i=0}^{j-1} \frac{1}{(i(1+\gamma-\beta)+1)^{\alpha\gamma-\beta+1}}\right] \\
&\leq e^{\frac{(\beta-\alpha\gamma)(\alpha\gamma-\beta+1)}{1+\gamma-\beta}(1+\gamma-\beta+\ln((1+\gamma-\beta)(j-1)+1))} \left[\prod_{i=0}^{j-1} \frac{1}{(i(1+\gamma-\beta)+1)^{\alpha\gamma-\beta+1}}\right] \\
&= \left[e^{1+\gamma-\beta}((1+\gamma-\beta)(j-1)+1)\right]^{\frac{(\beta-\alpha\gamma)(\alpha\gamma-\beta+1)}{1+\gamma-\beta}} \left[\frac{\Gamma\left(\frac{1}{1+\gamma-\beta}\right)}{(1+\gamma-\beta)^j \Gamma\left(j+\frac{1}{1+\gamma-\beta}\right)}\right]^{\alpha\gamma-\beta+1}.
\end{aligned}
\tag{15}
$$

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

$$
\begin{aligned}
&\int_{s_0}^{T} \frac{1}{(s_1-s_0)^\beta [\varrho(s_0,s_1)]^\gamma} \int_{s_1}^{T} \frac{1}{(s_2-s_1)^\beta [\varrho(s_1,s_2)]^\gamma} \cdots \int_{s_{j-1}}^{T} \frac{1}{(s_j-s_{j-1})^\beta [\varrho(s_{j-1},s_j)]^\gamma} \, ds_j \ldots ds_2 \, ds_1 \\
&\leq \left[\frac{(T-s_0)^{(1+\gamma-\beta)}\Gamma(\alpha\gamma-\beta+1)}{(1-\alpha)^\gamma(1+\gamma-\beta)^{\alpha\gamma-\beta+1}}\right]^j \left[e^{1+\gamma-\beta}((1+\gamma-\beta)(j-1)+1)\right]^{\frac{(\beta-\alpha\gamma)(\alpha\gamma-\beta+1)}{1+\gamma-\beta}} \left[\frac{\Gamma\left(\frac{1}{1+\gamma-\beta}\right)}{\Gamma\left(j+\frac{1}{1+\gamma-\beta}\right)}\right]^{\alpha\gamma-\beta+1}.
\end{aligned}
\tag{16}
$$

This establishes (12). The proof of Corollary 2.3 is thus completed. □

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 $\rho\colon \Omega \to (0,1)$ and $\tau\colon \Omega \to (0,T)$ be random variables, let $W\colon [0,T] \times \Omega \to \mathbb{R}^d$ be a standard Brownian motion with continuous sample paths, let $f\colon [0,T) \to [0,\infty]$ 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 ρ, τ, W, and $\mathcal{G}$ are independent. Then it holds $\mathbb{P}$-a.s. that*

$$
\mathbb{E}[F(\rho, \tau, W_{\tau+(T-\tau)\rho} - W_\tau) | \mathcal{H}] = f(\tau). \tag{17}
$$

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

$$
\mathbb{E}[F(\rho, \tau, W_{\tau+(T-\tau)\rho} - W_\tau) | \mathcal{H}] = \mathbb{E}[F(\rho, \tau, W_{\tau+(T-\tau)\rho} - W_\tau) | \sigma(\tau, (W_{\min\{s,\tau\}})_{s\in[0,T]})]. \tag{18}
$$

Next note that Hutzenthaler et al. [39, Lemma 2.2] (applied with $\mathcal{G} = \sigma(\rho, (W_s)_{s\in[0,T]})$, $S = (0,T)$, $\mathcal{S} = \mathcal{B}((0,T))$, $U(t,\omega) = \mathbf{1}_A(t)\mathbf{1}_B((W_{\min\{s,t\}}(\omega))_{s\in[0,T]})F(\rho(\omega), t, W_{t+(T-t)\rho(\omega)}(\omega) - W_t(\omega))$, and $Y = \tau$ for $t \in (0,T)$, $\omega \in \Omega$ in the notation of [39, 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

$$
\begin{aligned}
&\mathbb{E}[\mathbf{1}_A(\tau)\mathbf{1}_B((W_{\min\{s,\tau\}})_{s\in[0,T]})F(\rho, \tau, W_{\tau+(T-\tau)\rho} - W_\tau)] \\
&= \int_{(0,T)} \mathbb{E}[\mathbf{1}_A(t)\mathbf{1}_B((W_{\min\{s,t\}})_{s\in[0,T]})F(\rho, t, W_{t+(T-t)\rho} - W_t)](\tau(\mathbb{P}))(dt).
\end{aligned}
\tag{19}
$$

Independence of Brownian increments hence proves that for all $A \in \mathcal{B}((0,T))$, $B \in \mathcal{B}(C([0,T], \mathbb{R}^d))$ it holds that

$$
\begin{aligned}
&\mathbb{E}[\mathbf{1}_A(\tau)\mathbf{1}_B((W_{\min\{s,\tau\}})_{s\in[0,T]})F(\rho, \tau, W_{\tau+(T-\tau)\rho} - W_\tau)] \\
&= \int_{(0,T)} \mathbb{E}[\mathbf{1}_A(t)\mathbf{1}_B((W_{\min\{s,t\}})_{s\in[0,T]})]\mathbb{E}[F(\rho, t, W_{t+(T-t)\rho} - W_t)](\tau(\mathbb{P}))(dt) \\
&= \int_{(0,T)} \mathbb{E}[\mathbf{1}_A(t)\mathbf{1}_B((W_{\min\{s,t\}})_{s\in[0,T]})]f(t)(\tau(\mathbb{P}))(dt).
\end{aligned}
\tag{20}
$$

Hence, Hutzenthaler et al. [39, Lemma 2.2] (applied with $\mathcal{G} = \sigma(\rho, (W_s)_{s\in[0,T]})$, $S = (0,T)$, $\mathcal{S} = \mathcal{B}((0,T))$, $U(t,\omega) = \mathbf{1}_A(t)\mathbf{1}_B((W_{\min\{s,t\}}(\omega))_{s\in[0,T]})g(t)$, and $Y = \tau$ for $t \in (0,T)$, $\omega \in \Omega$ in the notation of [39, 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

$$\begin{aligned}\mathbb{E}[\mathbf{1}_A(\tau)\mathbf{1}_B((W_{\min\{s,\tau\}})_{s\in[0,T]})F(\rho,\tau,W_{\tau+(T-\tau)\rho}-W_\tau)] &= \int_{(0,T)} \mathbb{E}[\mathbf{1}_A(t)\mathbf{1}_B((W_{\min\{s,t\}})_{s\in[0,T]})f(t)](\tau(\mathbb{P}))(dt) \\ &= \mathbb{E}[\mathbf{1}_A(\tau)\mathbf{1}_B((W_{\min\{s,\tau\}})_{s\in[0,T]})f(\tau)].\end{aligned} \tag{21}$$

This together with (18) proves (17). The proof of Lemma 2.4 is thus completed. □

Corollary 2.5. *Let $T \in (0,\infty)$, $d \in \mathbb{N}$, $j \in \mathbb{N}_0$, $\mathbf{e}_1 = (1,0,0\ldots,0)$, $\mathbf{e}_2 = (0,1,0,\ldots,0)$, $\ldots$, $\mathbf{e}_{d+1} = (0,0,\ldots,0,1) \in \mathbb{R}^{d+1}$, $\nu_0,\nu_1,\ldots,\nu_j \in \{1,2,\ldots,d+1\}$, $\alpha \in (0,1)$, $p \in (1,\infty)$ satisfy $\alpha(p-1) \le \frac{p}{2} \le \alpha(p-1)+1$, let $\langle\cdot,\cdot\rangle\colon \mathbb{R}^{d+1}\times\mathbb{R}^{d+1} \to \mathbb{R}$ be the standard scalar product on $\mathbb{R}^{d+1}$, let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space, let $W = (W^1,W^2,\ldots,W^d)\colon [0,T]\times\Omega \to \mathbb{R}^d$ be a standard Brownian motion with continuous sample paths, let $\mathfrak{r}^{(n)}\colon \Omega \to (0,1)$, $n \in \mathbb{N}_0$, be i.i.d. random variables, assume that W and $(\mathfrak{r}^{(n)})_{n\in\mathbb{N}_0}$ are independent, let $\rho\colon (0,1) \to (0,\infty)$ and $\varrho\colon [0,T]^2 \to (0,\infty)$ satisfy for all $b \in (0,1)$, $t \in [0,T)$, $s \in (t,T]$ that $\rho(b) = \frac{1-\alpha}{b^\alpha}$, $\mathbb{P}(\mathfrak{r}^{(0)} \le b) = \int_0^b \rho(u)\,du$, and $\varrho(t,s) = \frac{1}{T-t}\rho(\frac{s-t}{T-t})$, let $S\colon \mathbb{N}_0\times[0,T)\times\Omega \to [0,T)$ satisfy for all $n \in \mathbb{N}_0$, $t \in [0,T)$ that $S(0,t) = t$ and $S(n+1,t) = S(n,t) + (T-S(n,t))\mathfrak{r}^{(n)}$, and let $t \in [0,T)$. Then*

$$\begin{aligned}&\mathbb{E}\left[\left|\prod_{i=0}^{j} \tfrac{1}{\varrho(S(i,t),S(i+1,t))}\left\langle \mathbf{e}_{\nu_i}, \left(1, \tfrac{W_{S(i+1,t)}-W_{S(i,t)}}{S(i+1,t)-S(i,t)}\right)\right\rangle\right|^p\right] \\ &\le \left[\max\left\{(T-t)^{\frac{p}{2}}, \tfrac{2^{\frac{p}{2}}\Gamma(\frac{p+1}{2})}{\sqrt{\pi}}\right\} \tfrac{(T-t)^{\frac{p}{2}}\Gamma(\frac{p}{2})}{(1-\alpha)^{p-1}(\frac{p}{2})^{\alpha(p-1)-\frac{p}{2}+1}}\right]^{j+1} \left[e^{\frac{p}{2}}\left(\tfrac{pj}{2}+1\right)\right]^{\frac{1}{2p}} \left[\tfrac{\Gamma(\frac{2}{p})}{\Gamma(1+j+\frac{2}{p})}\right]^{\alpha(p-1)-\frac{p}{2}+1}.\end{aligned} \tag{22}$$

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 $v = (v_1,v_2,\ldots,v_d) \in \mathbb{R}^d$ satisfy $v_1 = v_2 = ... = v_d = 1$. Note that for all $r \in [0,T)$, $s \in [r,T]$, $i \in \{1,2,\ldots,d\}$, $n \in \mathbb{N}$ it holds that $S(n,r) > S(n-1,r)$ and

$$\mathbb{E}\big[|W_s^i - W_r^i|^p\big] = \frac{(2(s-r))^{\frac{p}{2}}\Gamma\big(\frac{p+1}{2}\big)}{\sqrt{\pi}}. \tag{23}$$

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

$$\begin{aligned}&\mathbb{E}\left[\prod_{i=k}^{j+1}\left|\tfrac{1}{\varrho(S(i-1,t),S(i,t))}\left\langle \mathbf{e}_{\nu_{i-1}}, \left(1, \tfrac{W_{S(i,t)}-W_{S(i-1,t)}}{S(i,t)-S(i-1,t)}\right)\right\rangle\right|^p \middle| \mathbb{F}_{k-1}\right] = \int_{S(k-1,t)}^T \frac{\left\langle \mathbf{e}_{\nu_{k-1}}, \left(1, \left(\frac{2}{s_k-S(k-1,t)}\right)^{\frac{p}{2}} \frac{\Gamma(\frac{p+1}{2})}{\sqrt{\pi}} v\right)\right\rangle}{[\varrho(S(k-1,t),s_k)]^{p-1}} \\ &\cdot \int_{s_k}^T \frac{\left\langle \mathbf{e}_{\nu_k}, \left(1, \left(\frac{2}{s_{k+1}-s_k}\right)^{\frac{p}{2}} \frac{\Gamma(\frac{p+1}{2})}{\sqrt{\pi}} v\right)\right\rangle}{[\varrho(s_k,s_{k+1})]^{p-1}} \cdots \int_{s_j}^T \frac{\left\langle \mathbf{e}_{\nu_j}, \left(1, \left(\frac{2}{s_{j+1}-s_j}\right)^{\frac{p}{2}} \frac{\Gamma(\frac{p+1}{2})}{\sqrt{\pi}} v\right)\right\rangle}{[\varrho(s_j,s_{j+1})]^{p-1}}\, ds_{j+1}\ldots ds_{k+1}\, ds_k.\end{aligned} \tag{24}$$

To prove (24) we proceed by backward induction on $k \in \{1,2,\ldots,j+1\}$. For the base case $k = j+1$ note that the fact that $S(j+1,t) = S(j,t) + (T-S(j,t))\mathfrak{r}^{(j)}$, Lemma 2.4 (applied with $F(r,s,x) = \big|\frac{1}{\varrho(s,s+(T-s)r)}\langle \mathbf{e}_{\nu_j}, (1, \frac{x}{(T-s)r})\rangle\big|^p$, $\rho = \mathfrak{r}^{(j)}$, $\tau = S(j,t)$, $\mathcal{G} = \mathbb{F}_{j-1}$ for $r \in (0,1)$, $s \in [0,T)$, $x \in \mathbb{R}^d$ in the notation of Lemma 2.4), Hutzenthaler et al. [39, Lemma 2.3], the hypothesis that W and $\mathfrak{r}^{(j)}$ are independent, (23), 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

$$\begin{aligned}&\mathbb{E}\left[\left|\tfrac{1}{\varrho(S(j,t),S(j+1,t))}\left\langle \mathbf{e}_{\nu_j}, \left(1, \tfrac{W_{S(j+1,t)}-W_{S(j,t)}}{S(j+1,t)-S(j,t)}\right)\right\rangle\right|^p \middle| \mathbb{F}_j\right] \\ &= \mathbb{E}\left[\left|\tfrac{1}{\varrho(S(j,t),S(j,t)+(T-S(j,t))\mathfrak{r}^{(j)})}\left\langle \mathbf{e}_{\nu_j}, \left(1, \tfrac{W_{S(j,t)+(T-S(j,t))\mathfrak{r}^{(j)}}-W_{S(j,t)}}{(T-S(j,t))\mathfrak{r}^{(j)}}\right)\right\rangle\right|^p \middle| \mathbb{F}_j\right] \\ &= \mathbb{E}\left[\left|\tfrac{1}{\varrho(s,s+(T-s)\mathfrak{r}^{(j)})}\left\langle \mathbf{e}_{\nu_j}, \left(1, \tfrac{W_{s+(T-s)\mathfrak{r}^{(j)}}-W_s}{(T-s)\mathfrak{r}^{(j)}}\right)\right\rangle\right|^p\right]\Bigg|_{s=S(j,t)} \\ &= \int_0^1 \frac{\mathbb{E}\left[\left|\left\langle \mathbf{e}_{\nu_j}, \left(1, \frac{W_{s+(T-s)r}-W_s}{(T-s)r}\right)\right\rangle\right|^p\right]}{[\varrho(s,s+(T-s)r)]^p}\rho(r)dr\Bigg|_{s=S(j,t)} \\ &= \int_s^T \frac{\mathbb{E}\left[\left|\left\langle \mathbf{e}_{\nu_j}, \left(1, \frac{W_{s_{j+1}}-W_s}{s_{j+1}-s}\right)\right\rangle\right|^p\right]}{[\varrho(s,s_{j+1})]^p} \frac{1}{T-s}\rho\left(\frac{s_{j+1}-s}{T-s}\right)\, ds_{j+1}\Bigg|_{s=S(j,t)} \\ &= \int_{S(j,t)}^T \frac{\left\langle \mathbf{e}_{\nu_j}, \left(1, \left(\frac{2}{s_{j+1}-S(j,t)}\right)^{\frac{p}{2}} \frac{\Gamma(\frac{p+1}{2})}{\sqrt{\pi}} v\right)\right\rangle}{[\varrho(S(j,t),s_{j+1})]^{p-1}}\, ds_{j+1}.\end{aligned} \tag{25}$$

This establishes (24) in the base case $k = j+1$. For the induction step $\{2, 3 \ldots, j+1\} \ni k+1 \rightsquigarrow k \in \{1, 2, \ldots, j\}$ assume that there exists $k \in \{1, 2, \ldots, j\}$ which satisfies that

$$\begin{aligned}
&\mathbb{E}\Bigg[\prod_{i=k+1}^{j+1}\bigg|\tfrac{1}{\varrho(S(i-1,t),S(i,t))}\big\langle \mathbf{e}_{\nu_{i-1}},\big(1,\tfrac{W_{S(i,t)}-W_{S(i-1,t)}}{S(i,t)-S(i-1,t)}\big)\big\rangle\bigg|^p\,\bigg|\,\mathbb{F}_k\Bigg] = \int_{S(k,t)}^{T}\frac{\big\langle \mathbf{e}_{\nu_k},\big(1,\big(\frac{2}{s_{k+1}-S(k,t)}\big)^{\frac{p}{2}}\frac{\Gamma(\frac{p+1}{2})}{\sqrt{\pi}}v\big)\big\rangle}{[\varrho(S(k,t),s_{k+1})]^{p-1}}\\
&\quad\cdot\int_{s_{k+1}}^{T}\frac{\big\langle \mathbf{e}_{\nu_{k+1}},\big(1,\big(\frac{2}{s_{k+2}-s_{k+1}}\big)^{\frac{p}{2}}\frac{\Gamma(\frac{p+1}{2})}{\sqrt{\pi}}v\big)\big\rangle}{[\varrho(s_{k+1},s_{k+2})]^{p-1}}\cdots\int_{s_j}^{T}\frac{\big\langle \mathbf{e}_{\nu_j},\big(1,\big(\frac{2}{s_{j+1}-s_j}\big)^{\frac{p}{2}}\frac{\Gamma(\frac{p+1}{2})}{\sqrt{\pi}}v\big)\big\rangle}{[\varrho(s_j,s_{j+1})]^{p-1}}\,ds_{j+1}\ldots ds_{k+2}\,ds_{k+1}.
\end{aligned}\tag{26}$$

Observe that (26), the tower property, the fact that the random variable $\frac{1}{\varrho(S(k-1,t),S(k,t))}\Big(1, \frac{W_{S(k,t)}-W_{S(k-1,t)}}{S(k,t)-S(k-1,t)}\Big)$ is $\mathbb{F}_k$-measurable, the induction hypothesis, the fact that $S(k,t) = S(k-1,t)+(T-S(k-1,t))\mathfrak{r}^{(k-1)}$, and the fact that conditioned on $S(k-1,t)$ the σ-algebras $\sigma(\mathfrak{r}^{(k-1)}, (W_t - W_{S(k-1,t)})_{t\in[S(k-1,t),T]})$ and $\mathbb{F}_{k-1}$ are independent ensure that it holds $\mathbb{P}$-a.s. that

$$\begin{aligned}
&\mathbb{E}\Bigg[\prod_{i=k}^{j+1}\bigg|\tfrac{1}{\varrho(S(i-1,t),S(i,t))}\big\langle \mathbf{e}_{\nu_{i-1}},\big(1,\tfrac{W_{S(i,t)}-W_{S(i-1,t)}}{S(i,t)-S(i-1,t)}\big)\big\rangle\bigg|^p\,\bigg|\,\mathbb{F}_{k-1}\Bigg]\\
&=\mathbb{E}\Bigg[\bigg|\tfrac{1}{\varrho(S(k-1,t),S(k,t))}\big\langle \mathbf{e}_{\nu_{k-1}},\big(1,\tfrac{W_{S(k,t)}-W_{S(k-1,t)}}{S(k,t)-S(k-1,t)}\big)\big\rangle\bigg|^p\\
&\qquad\cdot\mathbb{E}\Bigg[\prod_{i=k+1}^{j+1}\bigg|\tfrac{1}{\varrho(S(i-1,t),S(i,t))}\big\langle \mathbf{e}_{\nu_{i-1}},\big(1,\tfrac{W_{S(i,t)}-W_{S(i-1,t)}}{S(i,t)-S(i-1,t)}\big)\big\rangle\bigg|^p\,\bigg|\,\mathbb{F}_k\Bigg]\,\Bigg|\,\mathbb{F}_{k-1}\Bigg]\\
&=\mathbb{E}\Bigg[\bigg|\tfrac{1}{\varrho(S(k-1,t),S(k,t))}\big\langle \mathbf{e}_{\nu_{k-1}},\big(1,\tfrac{W_{S(k,t)}-W_{S(k-1,t)}}{S(k,t)-S(k-1,t)}\big)\big\rangle\bigg|^p\int_{S(k,t)}^{T}\frac{\big\langle \mathbf{e}_{\nu_k},\big(1,\big(\frac{2}{s_{k+1}-S(k,t)}\big)^{\frac{p}{2}}\frac{\Gamma(\frac{p+1}{2})}{\sqrt{\pi}}v\big)\big\rangle}{[\varrho(S(k,t),s_{k+1})]^{p-1}}\\
&\qquad\cdot\int_{s_{k+1}}^{T}\frac{\big\langle \mathbf{e}_{\nu_{k+1}},\big(1,\big(\frac{2}{s_{k+2}-s_{k+1}}\big)^{\frac{p}{2}}\frac{\Gamma(\frac{p+1}{2})}{\sqrt{\pi}}v\big)\big\rangle}{[\varrho(s_{k+1},s_{k+2})]^{p-1}}\cdots\int_{s_j}^{T}\frac{\big\langle \mathbf{e}_{\nu_j},\big(1,\big(\frac{2}{s_{j+1}-s_j}\big)^{\frac{p}{2}}\frac{\Gamma(\frac{p+1}{2})}{\sqrt{\pi}}v\big)\big\rangle}{[\varrho(s_j,s_{j+1})]^{p-1}}\,ds_{j+1}\ldots ds_{k+2}\,ds_{k+1}\,\bigg|\,\mathbb{F}_{k-1}\Bigg]\\
&=\mathbb{E}\Bigg[\Bigg|\frac{\big\langle \mathbf{e}_{\nu_{k-1}},\big(1,\frac{W_{s+(T-s)\mathfrak{r}^{(k-1)}}-W_s}{(T-s)\mathfrak{r}^{(k-1)}}\big)\big\rangle}{\varrho(s,s+(T-s)\mathfrak{r}^{(k-1)})}\Bigg|^p\int_{s+(T-s)\mathfrak{r}^{(k-1)}}^{T}\frac{\big\langle \mathbf{e}_{\nu_k},\big(1,\big(\frac{2}{s_{k+1}-(s+(T-s)\mathfrak{r}^{(k-1)})}\big)^{\frac{p}{2}}\frac{\Gamma(\frac{p+1}{2})}{\sqrt{\pi}}v\big)\big\rangle}{[\varrho(s+(T-s)\mathfrak{r}^{(k-1)},s_{k+1})]^{p-1}}\\
&\qquad\cdot\int_{s_{k+1}}^{T}\frac{\big\langle \mathbf{e}_{\nu_{k+1}},\big(1,\big(\frac{2}{s_{k+2}-s_{k+1}}\big)^{\frac{p}{2}}\frac{\Gamma(\frac{p+1}{2})}{\sqrt{\pi}}v\big)\big\rangle}{[\varrho(s_{k+1},s_{k+2})]^{p-1}}\cdots\int_{s_j}^{T}\frac{\big\langle \mathbf{e}_{\nu_j},\big(1,\big(\frac{2}{s_{j+1}-s_j}\big)^{\frac{p}{2}}\frac{\Gamma(\frac{p+1}{2})}{\sqrt{\pi}}v\big)\big\rangle}{[\varrho(s_j,s_{j+1})]^{p-1}}\,ds_{j+1}\ldots ds_{k+2}\,ds_{k+1}\Bigg]\Bigg|_{s=S(k-1,t)}.
\end{aligned}\tag{27}$$

This, Hutzenthaler et al. [39, Lemma 2.3], the hypothesis that W and $\mathfrak{r}^{(k-1)}$ are independent, (23), 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

$$\begin{aligned}
&\mathbb{E}\Bigg[\prod_{i=k}^{j+1}\bigg|\tfrac{1}{\varrho(S(i-1,t),S(i,t))}\big\langle \mathbf{e}_{\nu_{i-1}},\big(1,\tfrac{W_{S(i,t)}-W_{S(i-1,t)}}{S(i,t)-S(i-1,t)}\big)\big\rangle\bigg|^p\,\bigg|\,\mathbb{F}_{k-1}\Bigg]\\
&=\int_0^1\frac{\big\langle \mathbf{e}_{\nu_{k-1}},\big(1,\frac{(2((T-s)r))^{\frac{p}{2}}\Gamma(\frac{p+1}{2})}{\sqrt{\pi}|(T-s)r|^p}v\big)\big\rangle\rho(r)}{[\varrho(s,s+(T-s)r)]^p}\int_{s+(T-s)r}^{T}\frac{\big\langle \mathbf{e}_{\nu_k},\big(1,\big(\frac{2}{s_{k+1}-(s+(T-s)r)}\big)^{\frac{p}{2}}\frac{\Gamma(\frac{p+1}{2})}{\sqrt{\pi}}v\big)\big\rangle}{[\varrho(s+(T-s)r,s_{k+1})]^{p-1}}\\
&\qquad\cdot\int_{s_{k+1}}^{T}\frac{\big\langle \mathbf{e}_{\nu_{k+1}},\big(1,\big(\frac{2}{s_{k+2}-s_{k+1}}\big)^{\frac{p}{2}}\frac{\Gamma(\frac{p+1}{2})}{\sqrt{\pi}}v\big)\big\rangle}{[\varrho(s_{k+1},s_{k+2})]^{p-1}}\cdots\int_{s_j}^{T}\frac{\big\langle \mathbf{e}_{\nu_j},\big(1,\big(\frac{2}{s_{j+1}-s_j}\big)^{\frac{p}{2}}\frac{\Gamma(\frac{p+1}{2})}{\sqrt{\pi}}v\big)\big\rangle}{[\varrho(s_j,s_{j+1})]^{p-1}}\,ds_{j+1}\ldots ds_{k+2}\,ds_{k+1}\,dr\,\Bigg|_{s=S(k-1,t)}\\
&=\int_{S(k-1,t)}^{T}\frac{\big\langle \mathbf{e}_{\nu_{k-1}},\big(1,\big(\frac{2}{s_k-S(k-1,t)}\big)^{\frac{p}{2}}\frac{\Gamma(\frac{p+1}{2})}{\sqrt{\pi}}v\big)\big\rangle}{[\varrho(S(k-1,t),s_k)]^{p-1}}\int_{s_k}^{T}\frac{\big\langle \mathbf{e}_{\nu_k},\big(1,\big(\frac{2}{s_{k+1}-s_k}\big)^{\frac{p}{2}}\frac{\Gamma(\frac{p+1}{2})}{\sqrt{\pi}}v\big)\big\rangle}{[\varrho(s_k,s_{k+1})]^{p-1}}\\
&\qquad\cdot\int_{s_{k+1}}^{T}\frac{\big\langle \mathbf{e}_{\nu_{k+1}},\big(1,\big(\frac{2}{s_{k+2}-s_{k+1}}\big)^{\frac{p}{2}}\frac{\Gamma(\frac{p+1}{2})}{\sqrt{\pi}}v\big)\big\rangle}{[\varrho(s_{k+1},s_{k+2})]^{p-1}}\cdots\int_{s_j}^{T}\frac{\big\langle \mathbf{e}_{\nu_j},\big(1,\big(\frac{2}{s_{j+1}-s_j}\big)^{\frac{p}{2}}\frac{\Gamma(\frac{p+1}{2})}{\sqrt{\pi}}v\big)\big\rangle}{[\varrho(s_j,s_{j+1})]^{p-1}}\,ds_{j+1}\ldots ds_{k+2}\,ds_{k+1}\,ds_k.
\end{aligned}\tag{28}$$

This completes the induction step. Induction hence proves (24).

Next (24) implies that

$$\mathbb{E}\left[\prod_{i=1}^{j+1}\left|\tfrac{1}{\varrho(S(i-1,t),S(i,t))}\langle\mathbf{e}_{\nu_{i-1}},\big(1,\tfrac{W_{S(i,t)}-W_{S(i-1,t)}}{S(i,t)-S(i-1,t)}\big)\rangle\right|^p\right]\leq\left[\max\left\{(T-t)^{\frac{p}{2}},\frac{2^{\frac{p}{2}}\Gamma\big(\frac{p+1}{2}\big)}{\sqrt{\pi}}\right\}\right]^{j+1}$$
$$\cdot\int_t^T\frac{1}{(s_1-t)^{\frac{p}{2}}[\varrho(t,s_1)]^{p-1}}\int_{s_1}^T\frac{1}{(s_2-s_1)^{\frac{p}{2}}[\varrho(s_1,s_2)]^{p-1}}\cdots\int_{s_j}^T\frac{1}{(s_{j+1}-s_j)^{\frac{p}{2}}[\varrho(s_j,s_{j+1})]^{p-1}}\,ds_{j+1}\ldots ds_2\,ds_1. \tag{29}$$

Inequality (29), Corollary 2.3 (applied with $\beta=\frac{p}{2}$ and $\gamma=p-1$ in the notation of Corollary 2.3) together with $\frac{p}{2}\in[\alpha(p-1),\alpha(p-1)+1]$, and the fact that $(\frac{p}{2}-\alpha(p-1))(1-(\frac{p}{2}-\alpha(p-1)))\leq\frac{1}{4}$ show that

$$\left\|\prod_{i=1}^{j+1}\tfrac{1}{\varrho(S(i-1,t),S(i,t))}\langle\mathbf{e}_{\nu_{i-1}},\big(1,\tfrac{W_{S(i,t)}-W_{S(i-1,t)}}{S(i,t)-S(i-1,t)}\big)\rangle\right\|_{L^p(\mathbb{P};\mathbb{R})}^p=\mathbb{E}\left[\prod_{i=1}^{j+1}\left|\tfrac{1}{\varrho(S(i-1,t),S(i,t))}\langle\mathbf{e}_{\nu_{i-1}},\big(1,\tfrac{W_{S(i,t)}-W_{S(i-1,t)}}{S(i,t)-S(i-1,t)}\big)\rangle\right|^p\right]$$
$$\leq\left[\max\left\{(T-t)^{\frac{p}{2}},\frac{2^{\frac{p}{2}}\Gamma\big(\frac{p+1}{2}\big)}{\sqrt{\pi}}\right\}\right]^{j+1}\left[\tfrac{(T-t)^{\frac{p}{2}}\Gamma(\alpha(p-1)-\frac{p}{2}+1)}{(1-\alpha)^{p-1}(\frac{p}{2})^{\alpha(p-1)-\frac{p}{2}+1}}\right]^{j+1}\left[e^{\frac{p}{2}}\left(\tfrac{pj}{2}+1\right)\right]^{\frac{2(\frac{p}{2}-\alpha(p-1))(\alpha(p-1)-\frac{p}{2}+1)}{p}}$$
$$\cdot\left[\tfrac{\Gamma(\frac{2}{p})}{\Gamma(1+j+\frac{2}{p})}\right]^{\alpha(p-1)-\frac{p}{2}+1}$$
$$\leq\left[\max\left\{(T-t)^{\frac{p}{2}},\frac{2^{\frac{p}{2}}\Gamma\big(\frac{p+1}{2}\big)}{\sqrt{\pi}}\right\}\tfrac{(T-t)^{\frac{p}{2}}\Gamma(\alpha(p-1)-\frac{p}{2}+1)}{(1-\alpha)^{p-1}(\frac{p}{2})^{\alpha(p-1)-\frac{p}{2}+1}}\right]^{j+1}\left[e^{\frac{p}{2}}\left(\tfrac{pj}{2}+1\right)\right]^{\frac{1}{2p}}\left[\tfrac{\Gamma(\frac{2}{p})}{\Gamma(1+j+\frac{2}{p})}\right]^{\alpha(p-1)-\frac{p}{2}+1}. \tag{30}$$

This together with $\alpha(p-1)-\frac{p}{2}+1\leq p-1-\frac{p}{2}+1=\frac{p}{2}$ implies (22). The proof of Corollary 2.5 is thus completed. □

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 Subsection 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 Section 5 below (cf. Corollary 5.1, Theorem 5.2, and Corollary 5.4 in Section 5 below).

3.1 Description of MLP approximations

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

$$\max\{|f(t,x,u)-f(t,x,\mathfrak{u})|,|g(x)-g(\mathfrak{x})|\}\leq\left[\sum_{\nu=1}^{d+1}L_\nu\,|u_\nu-\mathfrak{u}_\nu|\right]+\left[\sum_{\nu=1}^{d}K_\nu|x_\nu-\mathfrak{x}_\nu|\right], \tag{31}$$

$$\mathbb{E}\left[\left\|g(x+W_T^0-W_t^0)\big(1,\tfrac{W_T^0-W_t^0}{T-t}\big)\right\|_1+\int_t^T\left\|[(F(\mathbf{u}))(t,x+W_s^0-W_t^0)]\big(1,\tfrac{W_s^0-W_t^0}{s-t}\big)\right\|_1ds\right]<\infty, \tag{32}$$

$$\mathbf{u}(t,x)=\mathbb{E}\left[g(x+W_T^0-W_t^0)\big(1,\tfrac{W_T^0-W_t^0}{T-t}\big)+\int_t^T[(F(\mathbf{u}))(t,x+W_s^0-W_t^0)]\big(1,\tfrac{W_s^0-W_t^0}{s-t}\big)\,ds\right], \tag{33}$$

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

$$\begin{aligned}\mathbf{U}^\theta_{n,M}(t,x) &= \big(g(x),0\big) + \sum_{i=1}^{M^n} \frac{\big(g(x+W^{(\theta,0,-i)}_T - W^{(\theta,0,-i)}_t) - g(x)\big)}{M^n}\Big(1, \tfrac{W^{(\theta,0,-i)}_T - W^{(\theta,0,-i)}_t}{T-t}\Big) \\ &+ \sum_{l=0}^{n-1}\sum_{i=1}^{M^{n-l}} \frac{\big(F(\mathbf{U}^{(\theta,l,i)}_{l,M}) - \mathbb{1}_{\mathbb{N}}(l)F(\mathbf{U}^{(\theta,-l,i)}_{l-1,M})\big)(\mathcal{R}^{(\theta,l,i)}_t, x + W^{(\theta,l,i)}_{\mathcal{R}^{(\theta,l,i)}_t} - W^{(\theta,l,i)}_t)}{M^{n-l}\varrho(t,\mathcal{R}^{(\theta,l,i)}_t)}\left(1, \frac{W^{(\theta,l,i)}_{\mathcal{R}^{(\theta,l,i)}_t} - W^{(\theta,l,i)}_t}{\mathcal{R}^{(\theta,l,i)}_t - t}\right).\end{aligned} \tag{34}$$

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* $\mathbf{U}^\theta_{n,M}\colon[0,T)\times\mathbb{R}^d\times\Omega\to\mathbb{R}$ *is a continuous random field,*

(ii) for all $n\in\mathbb{N}_0$, $\theta\in\Theta$ *it holds that* $\sigma(\mathbf{U}^\theta_{n,M}) \subseteq \sigma((\mathfrak{r}^{(\theta,\vartheta)})_{\vartheta\in\Theta}, (W^{(\theta,\vartheta)})_{\vartheta\in\Theta})$,

(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* $\mathbf{U}^{(\theta,i,j)}_{n,M}$ *and* $\mathbf{U}^{(\theta,k,l)}_{m,M}$ *are independent,*

(iv) for all $n\in\mathbb{N}_0$, $\theta\in\Theta$ *it holds hat* $\mathbf{U}^\theta_{n,M}$, W^θ, *and* $\mathfrak{r}^\theta$ *are independent,*

(v) for all $n\in\mathbb{N}_0$ *it holds that* $\mathbf{U}^\theta_{n,M}$, $\theta\in\Theta$, *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*

$$\frac{F(\mathbf{U}^{(\theta,-l,i)}_{l-1,M})(\mathcal{R}^{(\theta,l,i)}_t, x+W^{(\theta,l,i)}_{\mathcal{R}^{(\theta,l,i)}_t} - W^{(\theta,l,i)}_t)}{\varrho(t,\mathcal{R}^{(\theta,l,i)}_t)}\left(1, \frac{W^{(\theta,l,i)}_{\mathcal{R}^{(\theta,l,i)}_t} - W^{(\theta,l,i)}_t}{\mathcal{R}^{(\theta,l,i)}_t - t}\right) \tag{35}$$

and

$$\frac{F(\mathbf{U}^{(\theta,l,i)}_{l-1,M})(\mathcal{R}^{(\theta,l,i)}_t, x+W^{(\theta,l,i)}_{\mathcal{R}^{(\theta,l,i)}_t} - W^{(\theta,l,i)}_t)}{\varrho(t,\mathcal{R}^{(\theta,l,i)}_t)}\left(1, \frac{W^{(\theta,l,i)}_{\mathcal{R}^{(\theta,l,i)}_t} - W^{(\theta,l,i)}_t}{\mathcal{R}^{(\theta,l,i)}_t - t}\right) \tag{36}$$

are identically distributed.

Proof of Lemma 3.2. First, observe that (34), the hypothesis that for all $M\in\mathbb{N}$, $\theta\in\Theta$ it holds that $\mathbf{U}^\theta_{0,M}=0$, the fact that for all $\theta\in\Theta$ it holds that W^θ 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 $\varrho|_{\{(s,t)\in[0,T)^2\colon s<t\}} \in C(\{(s,t)\in[0,T)^2\colon s<t\},\mathbb{R})$, 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 $F(\mathbf{U}^\theta_{n,M})$ is $(\mathcal{B}((0,T)\times\mathbb{R}^d)\otimes\sigma(\mathbf{U}^\theta_{n,M}))/\mathcal{B}(\mathbb{R})$-measurable. The hypothesis that for all $M\in\mathbb{N}$, $\theta\in\Theta$ it holds that $\mathbf{U}^\theta_{0,M}=0$, (34), the fact that for all $\theta\in\Theta$ it holds that W^θ 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 $(\mathcal{B}([0,T))\otimes\sigma(\mathfrak{r}^\theta))/\mathcal{B}([0,T))$-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 $((\mathfrak{r}^{(\theta,i,j,\vartheta)}, W^{(\theta,i,j,\vartheta)}))_{\vartheta\in\Theta}$ and $((\mathfrak{r}^{(\theta,k,l,\vartheta)}, W^{(\theta,k,l,\vartheta)}))_{\vartheta\in\Theta}$ are independent prove Item (iii). Furthermore, observe that Item (ii) and the fact that for all $\theta\in\Theta$ it holds that $(\mathfrak{r}^{(\theta,\vartheta)})_{\vartheta\in\Theta}$, $(W^{(\theta,\vartheta)})_{\vartheta\in\Theta}$, W^θ, and $\mathfrak{r}^\theta$ are independent establish Item (iv). Next observe that the hypothesis that for all $\theta\in\Theta$ it holds that $\mathbf{U}^\theta_{0,M}=0$, 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. [39, Corollary 2.5], and induction on $\mathbb{N}_0$ establish Item (v). Furthermore, observe that Item (ii) and the fact that for all $\theta\in\Theta$, $l\in\mathbb{N}$, $i\in\mathbb{N}$ it holds that $(\mathfrak{r}^{(\theta,-l,i,\vartheta)})_{\vartheta\in\Theta}$, $(W^{(\theta,-l,i,\vartheta)})_{\vartheta\in\Theta}$, $W^{\theta,l,i}$, and $\mathfrak{r}^{\theta,l,i}$ are independent, and, e.g., Hutzenthaler et al. [39, 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 $\psi\colon\mathbb{R}^{d+1}\to\mathbb{R}$ it holds that

$$\begin{aligned}&\mathbb{E}\left[\psi\left(\frac{F(\mathbf{U}^{(\theta,-l,i)}_{l-1,M})(\mathcal{R}^{(\theta,l,i)}_t, x+W^{(\theta,l,i)}_{\mathcal{R}^{(\theta,l,i)}_t} - W^{(\theta,l,i)}_t)}{\varrho(t,\mathcal{R}^{(\theta,l,i)}_t)}\left(1, \frac{W^{(\theta,l,i)}_{\mathcal{R}^{(\theta,l,i)}_t} - W^{(\theta,l,i)}_t}{\mathcal{R}^{(\theta,l,i)}_t - t}\right)\right)\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[\psi\left(\frac{F(\mathbf{U}^{(\theta,-l,i)}_{l-1,M})(r,x+z)}{\varrho(t,r)}\left(1,\frac{z}{r-t}\right)\right)\right]\Bigg|_{r=\mathcal{R}^{(\theta,l,i)}_t,\, z=W^{(\theta,l,i)}_{\mathcal{R}^{(\theta,l,i)}_t} - W^{(\theta,l,i)}_t}\right].\end{aligned} \tag{37}$$

This, Item (v), Item (ii), and the fact that for all $\theta\in\Theta$, $l,i\in\mathbb{N}$ it holds that $(\mathfrak{r}^{(\theta,l,i,\vartheta)})_{\vartheta\in\Theta}$, $(W^{(\theta,l,i,\vartheta)})_{\vartheta\in\Theta})$, $W^{\theta,l,i}$, and $\mathfrak{r}^{\theta,l,i}$ are independent, and, e.g., Hutzenthaler et al. [39, 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 $\psi\colon\mathbb{R}^{d+1}\to\mathbb{R}$ it holds that

$$\begin{aligned}
&\mathbb{E}\!\left[\psi\!\left(\tfrac{F(\mathbf{U}^{(\theta,-l,i)}_{l-1,M})(\mathcal{R}^{(\theta,l,i)}_t,x+W^{(\theta,l,i)}_{\mathcal{R}^{(\theta,l,i)}_t}-W^{(\theta,l,i)}_t)}{\varrho(t,\mathcal{R}^{(\theta,l,i)}_t)}\left(1,\tfrac{W^{(\theta,l,i)}_{\mathcal{R}^{(\theta,l,i)}_t}-W^{(\theta,l,i)}_t}{\mathcal{R}^{(\theta,l,i)}_t-t}\right)\right)\right]\\
&=\mathbb{E}\!\left[\mathbb{E}\!\left[\psi\!\left(\tfrac{F(\mathbf{U}^{(\theta,l,i)}_{l-1,M})(r,x+z)}{\varrho(t,r)}\left(1,\tfrac{z}{r-t}\right)\right)\right]\Big|_{r=\mathcal{R}^{(\theta,l,i)}_t,\,z=W^{(\theta,l,i)}_{\mathcal{R}^{(\theta,l,i)}_t}-W^{(\theta,l,i)}_t}\right]\\
&=\mathbb{E}\!\left[\psi\!\left(\tfrac{F(\mathbf{U}^{(\theta,l,i)}_{l-1,M})(\mathcal{R}^{(\theta,l,i)}_t,x+W^{(\theta,l,i)}_{\mathcal{R}^{(\theta,l,i)}_t}-W^{(\theta,l,i)}_t)}{\varrho(t,\mathcal{R}^{(\theta,l,i)}_t)}\left(1,\tfrac{W^{(\theta,l,i)}_{\mathcal{R}^{(\theta,l,i)}_t}-W^{(\theta,l,i)}_t}{\mathcal{R}^{(\theta,l,i)}_t-t}\right)\right)\right].
\end{aligned}\tag{38}$$

This establishes Item (vi). The proof of Lemma 3.2 is thus completed. ☐

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*

$$\int_0^1\frac{1}{s^{\frac{q}{2}}\,[\rho(s)]^{q-1}}\,ds+\sup_{s\in[t,T)}\mathbb{E}\Big[\big|\big(F(0)\big)(s,x+W^0_s-W^0_t)\big|^q\Big]<\infty. \tag{39}$$

Then

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

$$\sup_{u\in(0,T]}\sup_{t\in[0,u)}\sup_{y\in\mathbb{R}^d}\mathbb{E}\Big[\Big|\big(g(y+W^\theta_u-W^\theta_t)-g(y)\big)\left\langle\mathbf{e}_\nu,\big(1,\tfrac{W^\theta_u-W^\theta_t}{u-t}\big)\right\rangle\Big|^q\Big]<\infty, \tag{40}$$

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

$$\sup_{s\in[t,T)}\left\{\mathbb{E}\Big[\big|(\mathbf{U}^\theta_{n,M}(s,x+W^\theta_s-W^\theta_t))_\nu\big|^q\Big]+\mathbb{E}\left[\frac{\big|\big(F(\mathbf{U}^\theta_{n,M})\big)(\mathcal{R}^\theta_s,x+W^\theta_{\mathcal{R}^\theta_s}-W^\theta_t)\big|^q}{[\varrho(s,\mathcal{R}^\theta_s)]^q}\left|\left\langle\mathbf{e}_\nu,\big(1,\tfrac{W^\theta_{\mathcal{R}^\theta_s}-W^\theta_s}{\mathcal{R}^\theta_s-s}\big)\right\rangle\right|^q\right]\right\}<\infty, \tag{41}$$

and

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

$$\begin{aligned}
\mathbb{E}\big[\mathbf{U}^\theta_{n,M}(t,x)\big]&=\mathbb{E}\Big[g(x+W^\theta_T-W^\theta_t)\Big(1,\tfrac{W^\theta_T-W^\theta_t}{T-t}\Big)\Big]\\
&\quad+\mathbb{E}\left[\big(F(\mathbf{U}^\theta_{n-1,M,Q})\big)(\mathcal{R}^\theta_t,x+W^\theta_{\mathcal{R}^\theta_t}-W^\theta_t)\Big(1,\tfrac{W^\theta_{\mathcal{R}^\theta_t}-W^\theta_t}{\mathcal{R}^\theta_t-t}\Big)\right].
\end{aligned}\tag{42}$$

Proof of Lemma 3.3. The Cauchy-Schwarz inequality, the Lipschitz property (31) of g, Jensen's inequality, and the scaling property of Brownian motion yield for all $\theta\in\Theta$, $q\in[1,\infty)$, $\nu\in\{1,2,\ldots,d+1\}$ that

$$\begin{aligned}
&\sup_{u\in(0,T]}\sup_{t\in[0,u)}\sup_{y\in\mathbb{R}^d}\left(\mathbb{E}\Big[\Big|\big(g(y+W^\theta_u-W^\theta_t)-g(y)\big)\left\langle\mathbf{e}_\nu,\big(1,\tfrac{W^\theta_u-W^\theta_t}{u-t}\big)\right\rangle\Big|^q\Big]\right)^2\\
&\le\sup_{u\in(0,T]}\sup_{t\in[0,u)}\sup_{y\in\mathbb{R}^d}\left(\mathbb{E}\Big[\big|g(y+W^\theta_u-W^\theta_t)-g(y)\big|^{2q}\Big]\,\mathbb{E}\Big[\Big|\left\langle\mathbf{e}_\nu,\big(1,\tfrac{W^\theta_u-W^\theta_t}{u-t}\big)\right\rangle\Big|^{2q}\Big]\right)\\
&\le\sup_{u\in(0,T]}\sup_{t\in[0,u)}\left(\mathbb{E}\left[\left|\sum_{i=1}^d K_i|W^{\theta,i}_u-W^{\theta,i}_t|\right|^{2q}\right]\mathbb{E}\left[1+\left|\tfrac{W^{\theta,1}_u-W^{\theta,1}_t}{u-t}\right|^{2q}\right]\right)\\
&\le\sup_{u\in(0,T]}\sup_{t\in[0,u)}\left(d^{2q-1}\left(\textstyle\sum_{i=1}^d(K_i)^{2q}\right)\tfrac{(u-t)^q}{T^q}\mathbb{E}[|W^{0,1}_T|^{2q}]\left(1+\tfrac{(u-t)^q}{(u-t)^{2q}T^q}\mathbb{E}[|W^{0,1}_T|^{2q}]\right)\right)\\
&=d^{2q-1}\left(\textstyle\sum_{i=1}^d(K_i)^{2q}\right)\mathbb{E}[|W^{0,1}_T|^{2q}]\left(1+\tfrac{1}{T^{2q}}\mathbb{E}[|W^{0,1}_T|^{2q}]\right)<\infty.
\end{aligned}\tag{43}$$

This proves Item (i). Next observe that the fact that W^θ, $\mathcal{R}^\theta$, and U^θ are independent and continuous random fields (see Lemma 3.2), Hutzenthaler et al. [39, Lemma 2.3]), Hölder's inequality (applied with the conjugate numbers $\frac{p+q}{2q},\frac{p+q}{p-q}\in(1,\infty)$), and the scaling property of Brownian motion demonstrate that for all $t\in[0,T)$,

$s \in [t,T)$, $n \in \mathbb{N}_0$, $\theta \in \Theta$, $q \in [1,p)$, $\nu \in \{1,2,\ldots,d+1\}$ it holds that

$$\mathbb{E}\left[\frac{\left|\left(F(\mathbf{U}^\theta_{n,M})\right)(\mathcal{R}^\theta_s, x+W^\theta_{\mathcal{R}^\theta_s}-W^\theta_t)\right|^q}{[\varrho(s,\mathcal{R}^\theta_s)]^q}\left|\left\langle \mathbf{e}_\nu, \left(1, \frac{W^\theta_{\mathcal{R}^\theta_s}-W^\theta_s}{\mathcal{R}^\theta_s - s}\right)\right\rangle\right|^q\right]$$
$$= \int_0^1 \mathbb{E}\left[\frac{\left|\left(F(\mathbf{U}^\theta_{n,M})\right)(s+u(T-s), x+W^\theta_{s+u(T-s)}-W^\theta_t)\right|^q}{[\varrho(s,s+u(T-s))]^q}\left|\left\langle \mathbf{e}_\nu, \left(1, \frac{W^\theta_{s+u(T-s)}-W^\theta_s}{s+u(T-s)-s}\right)\right\rangle\right|^q\right]\rho(u)\,du$$
$$\leq \int_0^1 \frac{\sup_{z\in[t,T)}\left(\mathbb{E}\left[\left|\left(F(\mathbf{U}^\theta_{n,M})\right)(z,x+W^\theta_z-W^\theta_t)\right|^{\frac{p+q}{2}}\right]\right)^{\frac{2q}{p+q}}(T-s)^q}{[\rho(u)]^{q-1}}\left(1+\left(\mathbb{E}\left[\left|\frac{W^{\theta,1}_{s+u(T-s)}-W^{\theta,1}_s}{u(T-s)}\right|^{\frac{q(p+q)}{p-q}}\right]\right)^{\frac{p-q}{p+q}}\right)du$$
$$= \left(\sup_{z\in[t,T)}\mathbb{E}\left[\left|\left(F(\mathbf{U}^\theta_{n,M})\right)(z,x+W^\theta_z-W^\theta_t)\right|^{\frac{p+q}{2}}\right]\right)^{\frac{2q}{p+q}}\int_0^1 \frac{(T-s)^q}{[\rho(u)]^{q-1}}\left(1+\frac{(u(T-s))^{\frac{q}{2}}}{(u(T-s))^q T^{\frac{q}{2}}}\left(\mathbb{E}\left[|W^{0,1}_T|^{\frac{q(p+q)}{p-q}}\right]\right)^{\frac{p-q}{p+q}}\right)du$$
$$\leq \left(\sup_{z\in[t,T)}\mathbb{E}\left[\left|\left(F(\mathbf{U}^\theta_{n,M})\right)(z,x+W^\theta_z-W^\theta_t)\right|^{\frac{p+q}{2}}\right]\right)^{\frac{2q}{p+q}}\left(T^q+\left(\mathbb{E}\left[|W^{0,1}_T|^{\frac{q(p+q)}{p-q}}\right]\right)^{\frac{p-q}{p+q}}\right)\int_0^1 \frac{u^{-\frac{q}{2}}}{[\rho(u)]^{q-1}}\,du. \tag{44}$$

Now we prove (41) by induction on $n \in \mathbb{N}_0$. In the base case $n = 0$, the fact that for all $\theta \in \Theta$ it holds that $U^\theta_{0,M} = 0$, (44), and (39) ensure that for all $\theta \in \Theta$, $q \in [1,p)$, $t \in [0,T)$, $\nu \in \{1,2,\ldots,d+1\}$ it holds that

$$\begin{aligned}&\sup_{s\in[t,T)}\left\{\mathbb{E}\left[\left|(\mathbf{U}^\theta_{0,M}(s,x+W^\theta_s-W^\theta_t))_\nu\right|^q\right]+\mathbb{E}\left[\frac{\left|\left(F(\mathbf{U}^\theta_{0,M})\right)(\mathcal{R}^\theta_s,x+W^\theta_{\mathcal{R}^\theta_s}-W^\theta_t)\right|^q}{[\varrho(s,\mathcal{R}^\theta_s)]^q}\left|\left\langle\mathbf{e}_\nu,\left(1,\frac{W^\theta_{\mathcal{R}^\theta_s}-W^\theta_s}{\mathcal{R}^\theta_s-s}\right)\right\rangle\right|^q\right]\right\}\\
&=\sup_{s\in[t,T)}\left\{\mathbb{E}\left[\frac{\left|\left(F(0)\right)(\mathcal{R}^\theta_s,x+W^\theta_{\mathcal{R}^\theta_s}-W^\theta_t)\right|^q}{[\varrho(s,\mathcal{R}^\theta_s)]^q}\left|\left\langle\mathbf{e}_\nu,\left(1,\frac{W^\theta_{\mathcal{R}^\theta_s}-W^\theta_s}{\mathcal{R}^\theta_s-s}\right)\right\rangle\right|^q\right]\right\}\\
&\leq\left(\sup_{z\in[t,T)}\mathbb{E}\left[\left|\left(F(0)\right)(z,x+W^0_z-W^0_t)\right|^{\frac{p+q}{2}}\right]\right)^{\frac{2q}{p+q}}\left(T^q+\left(\mathbb{E}\left[|W^{0,1}_T|^{\frac{q(p+q)}{p-q}}\right]\right)^{\frac{p-q}{p+q}}\right)\int_0^1\frac{u^{-\frac{q}{2}}}{[\rho(u)]^{q-1}}\,du<\infty.\end{aligned} \tag{45}$$

This establishes (41) in the base case $n = 0$. For the induction step $\mathbb{N}_0 \ni n-1 \rightsquigarrow n \in \mathbb{N}$ let $n \in \mathbb{N}$ and assume that for all $k \in \mathbb{N}_0 \cap [0,n)$, $\theta \in \Theta$, $q \in [1,p)$, $t \in [0,T)$, $\nu \in \{1,2,\ldots,d+1\}$ it holds that

$$\sup_{s\in[t,T)}\left\{\mathbb{E}\left[\left|(\mathbf{U}^\theta_{k,M}(s,x+W^\theta_s-W^\theta_t))_\nu\right|^q\right]+\mathbb{E}\left[\frac{\left|\left(F(\mathbf{U}^\theta_{k,M})\right)(\mathcal{R}^\theta_s,x+W^\theta_{\mathcal{R}^\theta_s}-W^\theta_t)\right|^q}{[\varrho(s,\mathcal{R}^\theta_s)]^q}\left|\left\langle\mathbf{e}_\nu,\left(1,\frac{W^\theta_{\mathcal{R}^\theta_s}-W^\theta_s}{\mathcal{R}^\theta_s-s}\right)\right\rangle\right|^q\right]\right\}<\infty. \tag{46}$$

Observe that (34) and Jensen's inequality ensure that for all $\theta \in \Theta$, $q \in [1,p)$, $t \in [0,T)$, $s \in [t,T)$, $\nu \in \{1,2,\ldots,d+1\}$ it holds that

$$\begin{aligned}&\mathbb{E}\left[\left|(\mathbf{U}^\theta_{n,M}(s,x+W^\theta_s-W^\theta_t))_\nu\right|^q\right]\leq(3+2n)^{q-1}|g(x)|^q+(3+2n)^{q-1}\mathbb{E}\left[|g(x+W^\theta_s-W^\theta_t)-g(x)|^q\right]\\
&+\frac{(3+2n)^{q-1}}{M^n}\sum_{i=1}^{M^n}\mathbb{E}\left[\left|g(x+W^\theta_s-W^\theta_t+W^{(\theta,0,-i)}_T-W^{(\theta,0,-i)}_s)-g(x+W^\theta_s-W^\theta_t))\right|^q\left|\left\langle\mathbf{e}_\nu,\left(1,\frac{W^{(\theta,0,-i)}_T-W^{(\theta,0,-i)}_s}{T-s}\right)\right\rangle\right|^q\right]\\
&+\sum_{l=0}^{n-1}\frac{(3+2n)^{q-1}}{M^{n-l}}\sum_{i=1}^{M^{n-l}}\mathbb{E}\left[\frac{\left|\left(F(\mathbf{U}^{(\theta,l,i)}_{l,M})\right)(\mathcal{R}^{(\theta,l,i)}_s,x+W^\theta_s-W^\theta_t+W^{(\theta,l,i)}_{\mathcal{R}^{(\theta,l,i)}_s}-W^{(\theta,l,i)}_s)\right|^q}{\left[\varrho(s,\mathcal{R}^{(\theta,l,i)}_s)\right]^q}\left|\left\langle\mathbf{e}_\nu,\left(1,\frac{W^{(\theta,l,i)}_{\mathcal{R}^{(\theta,l,i)}_s}-W^{(\theta,l,i)}_s}{\mathcal{R}^{(\theta,l,i)}_s-s}\right)\right\rangle\right|^q\right]\\
&+\sum_{l=1}^{n-1}\frac{(3+2n)^{q-1}}{M^{n-l}}\sum_{i=1}^{M^{n-l}}\mathbb{E}\left[\frac{\left|\left(F(\mathbf{U}^{(\theta,-l,i)}_{l-1,M})\right)(\mathcal{R}^{(\theta,l,i)}_s,x+W^\theta_s-W^\theta_t+W^{(\theta,l,i)}_{\mathcal{R}^{(\theta,l,i)}_s}-W^{(\theta,l,i)}_s)\right|^q}{\left[\varrho(s,\mathcal{R}^{(\theta,l,i)}_s)\right]^q}\left|\left\langle\mathbf{e}_\nu,\left(1,\frac{W^{(\theta,l,i)}_{\mathcal{R}^{(\theta,l,i)}_s}-W^{(\theta,l,i)}_s}{\mathcal{R}^{(\theta,l,i)}_s-s}\right)\right\rangle\right|^q\right].\end{aligned} \tag{47}$$

This, Hutzenthaler et al. [39, Corollary 2.5] together with Lemma 3.2, Item (i), and the induction hypothesis (46) yield for all $\theta \in \Theta$, $q \in [1,p)$, $t \in [0,T)$, $\nu \in \{1,2,\ldots,d+1\}$ that

$$\begin{aligned}&\sup_{s\in[t,T)} \mathbb{E}\Big[\big|(\mathbf{U}^\theta_{n,M}(s,x+W^\theta_s-W^\theta_t))_\nu\big|^q\Big] \le (3+2n)^{q-1}|g(x)|^q\\ &\quad + (6+2n)^q \sup_{u\in(0,T]}\sup_{s\in[0,u)}\sup_{y\in\mathbb{R}^d}\sup_{j\in\{1,2,\ldots,d+1\}} \mathbb{E}\Big[\big|g(y+W^0_u-W^0_s)-g(y)\big|^q \Big|\big\langle \mathbf{e}_j, \big(1, \tfrac{W^0_u-W^0_s}{u-s}\big)\big\rangle\Big|^q\Big]\\ &\quad + \sum_{l=0}^{n-1}(6+2n)^q \sup_{s\in[t,T)} \mathbb{E}\left[\frac{\big|\big(F(\mathbf{U}^0_{l,M})\big)(\mathcal{R}^0_s, x+W^0_{\mathcal{R}^0_s}-W^0_t)\big|^q}{[\varrho(s,\mathcal{R}^0_s)]^q}\Big|\big\langle \mathbf{e}_\nu, \big(1, \tfrac{W^0_{\mathcal{R}^0_s}-W^0_s}{\mathcal{R}^0_s-s}\big)\big\rangle\Big|^q\right] < \infty.\end{aligned} \tag{48}$$

Furthermore, Jensen's inequality, the Lipschitz property (31) of f, (48), and assumption (39) yield that for all $\theta \in \Theta$, $q \in [1,p)$, $t \in [0,T)$ it holds that

$$\begin{aligned}&\sup_{s\in[t,T)}\Big\{\mathbb{E}\Big[\big|\big(F(\mathbf{U}^\theta_{n,M})\big)(s,x+W^\theta_s-W^\theta_t)\big|^q\Big]\Big\}\\ &\le 2^{q-1}\sup_{s\in[t,T)}\Big(\mathbb{E}\Big[\big|\big(F(\mathbf{U}^\theta_{n,M})-F(0)\big)(s,x+W^\theta_s-W^\theta_t)\big|^q\Big]+\mathbb{E}\Big[\big|\big(F(0)\big)(s,x+W^\theta_s-W^\theta_t)\big|^q\Big]\Big)\\ &\le 2^{q-1}\sup_{s\in[t,T)}\left(\mathbb{E}\left[\left|\sum_{\nu=1}^{d+1}L_\nu \mathbf{U}^{\theta,\nu}_{n,M}(s,x+W^\theta_s-W^\theta_t)\right|^q\right]+\mathbb{E}\Big[\big|\big(F(0)\big)(s,x+W^\theta_s-W^\theta_t)\big|^q\Big]\right)\\ &\le (4d)^q\sum_{\nu=1}^{d+1}(L_\nu)^q\sup_{s\in[t,T)}\mathbb{E}\Big[\Big|\mathbf{U}^{\theta,\nu}_{n,M}(s,x+W^\theta_s-W^\theta_t)\Big|^q\Big]+2^q\sup_{s\in[t,T)}\mathbb{E}\Big[\big|\big(F(0)\big)(s,x+W^\theta_s-W^\theta_t)\big|^q\Big]\\ &<\infty.\end{aligned} \tag{49}$$

This, (44), and assumption (39) implies that for all $\theta \in \Theta$, $q \in [1,p)$, $t \in [0,T)$, $\nu \in \{1,2,\ldots,d+1\}$ it holds that

$$\begin{aligned}&\sup_{s\in[t,T)}\mathbb{E}\left[\frac{\big|\big(F(\mathbf{U}^\theta_{n,M})\big)(\mathcal{R}^\theta_s,x+W^\theta_{\mathcal{R}^\theta_s}-W^\theta_t)\big|^q}{[\varrho(s,\mathcal{R}^\theta_s)]^q}\Big|\big\langle\mathbf{e}_\nu,\big(1,\tfrac{W^\theta_{\mathcal{R}^\theta_s}-W^\theta_s}{\mathcal{R}^\theta_s-s}\big)\big\rangle\Big|^q\right]\\ &\le\left(\sup_{s\in[t,T)}\mathbb{E}\Big[\big|\big(F(\mathbf{U}^\theta_{n,M})\big)(s,x+W^\theta_s-W^\theta_t)\big|^{\frac{p+q}{2}}\Big]\right)^{\frac{2q}{p+q}}\left(T^q+\left(\mathbb{E}\Big[\Big|W^{0,1}_T\Big|^{\frac{q(p+q)}{p-q}}\Big]\right)^{\frac{p-q}{p+q}}\right)\int_0^1\frac{u^{-\frac{q}{2}}}{[\rho(u)]^{q-1}}\,du\\ &<\infty.\end{aligned} \tag{50}$$

This and (48) finish the induction step. Induction hence proves Item (ii).

Finally, we prove Item (iii). Note that (34), Item (ii), Corollary 2.5 in [39] together with the fact that for all $n \in \mathbb{N}_0$ it holds that $\mathbf{U}^\theta_{n,M}$, $\theta \in \Theta$, are identically distributed (see Item (i) in Lemma 3.2) and together with the fact that $(\mathcal{R}^\theta, W^\theta)$, $\theta \in \Theta$, are identically distributed, a telescoping sum, Hutzenthaler et al. [39, Lemma 2.3]), and the fact that for every $t \in [0,T)$, $\theta \in \Theta$ the distribution of R^θ_t has density $\varrho(t,\cdot)$ with respect to Lebesgue measure on $[t,T]$ yield that for all $n \in \mathbb{N}$, $\theta \in \Theta$, $t \in [0,T)$ it holds that

$$\begin{aligned}&\mathbb{E}\left[\mathbf{U}^\theta_{n,M}(t,x)\right]-\mathbb{E}\Big[g(x+W^0_T-W^0_t)\Big(1,\tfrac{W^0_T-W^0_t}{T-t}\Big)\Big]\\ &=\sum_{l=0}^{n-1}\sum_{i=1}^{M^{n-l}}\tfrac{1}{M^{n-l}}\mathbb{E}\left[\frac{\big(F(\mathbf{U}^{(\theta,l,i)}_{l,M})\big)(\mathcal{R}^{(\theta,l,i)}_t,x+W^{(\theta,l,i)}_{\mathcal{R}^{(\theta,l,i)}_t}-W^{(\theta,l,i)}_t)}{\varrho(t,\mathcal{R}^{(\theta,l,i)}_t)}\left(1,\frac{W^{(\theta,l,i)}_{\mathcal{R}^{(\theta,l,i)}_t}-W^{(\theta,l,i)}_t}{\mathcal{R}^{(\theta,l,i)}_t-t}\right)\right]\\ &\quad-\sum_{l=1}^{n-1}\sum_{i=1}^{M^{n-l}}\tfrac{1}{M^{n-l}}\mathbb{E}\left[\frac{\big(F(\mathbf{U}^{(\theta,-l,i)}_{l-1,M})\big)(\mathcal{R}^{(\theta,l,i)}_t,x+W^{(\theta,l,i)}_{\mathcal{R}^{(\theta,l,i)}_t}-W^{(\theta,l,i)}_t)}{\varrho(t,\mathcal{R}^{(\theta,l,i)}_t)}\left(1,\frac{W^{(\theta,l,i)}_{\mathcal{R}^{(\theta,l,i)}_t}-W^{(\theta,l,i)}_t}{\mathcal{R}^{(\theta,l,i)}_t-t}\right)\right]\\ &=\sum_{l=0}^{n-1}\left(\mathbb{E}\left[\frac{\big(F(\mathbf{U}^\theta_{l,M})\big)(\mathcal{R}^\theta_t,x+W^\theta_{\mathcal{R}^\theta_t}-W^\theta_t)}{\varrho(t,\mathcal{R}^\theta_t)}\left(1,\frac{W^\theta_{\mathcal{R}^\theta_t}-W^\theta_t}{\mathcal{R}^\theta_t-t}\right)\right]-\mathbb{1}_{\mathbb{N}}(l)\mathbb{E}\left[\frac{\big(F(\mathbf{U}^\theta_{l-1,M})\big)(\mathcal{R}^\theta_t,x+W^\theta_{\mathcal{R}^\theta_t}-W^\theta_t)}{\varrho(t,\mathcal{R}^\theta_t)}\left(1,\frac{W^\theta_{\mathcal{R}^\theta_t}-W^\theta_t}{\mathcal{R}^\theta_t-t}\right)\right]\right)\\ &=\mathbb{E}\left[\frac{\big(F(\mathbf{U}^\theta_{n-1,M})\big)(\mathcal{R}^\theta_t,x+W^\theta_{\mathcal{R}^\theta_t}-W^\theta_t)}{\varrho(t,\mathcal{R}^\theta_t)}\left(1,\frac{W^\theta_{\mathcal{R}^\theta_t}-W^\theta_t}{\mathcal{R}^\theta_t-t}\right)\right]\\ &=\int_t^T\mathbb{E}\left[\frac{\big(F(\mathbf{U}^\theta_{n-1,M})\big)(s,x+W^\theta_s-W^\theta_t)}{\varrho(t,s)}\left(1,\frac{W^\theta_s-W^\theta_t}{s-t}\right)\right]\mathbb{P}(\mathcal{R}^\theta_t\in ds)\\ &=\int_t^T\mathbb{E}\Big[\big(F(\mathbf{U}^\theta_{n-1,M,Q})\big)(s,x+W^\theta_s-W^\theta_t)\Big(1,\tfrac{W^\theta_s-W^\theta_t}{s-t}\Big)\Big]\,ds.\end{aligned} \tag{51}$$

This establishes Item (iii). The proof of Lemma 3.3 is thus completed. □

3.3 Error analysis for MLP approximations

Lemma 3.4 (Recursive bound for global error)**.** *Assume Setting 3.1, let $p \in (1,\infty)$, $M \in \mathbb{N}$, and assume for all $q \in [1,p)$, $t \in [0,T)$, $x \in \mathbb{R}^d$ that*

$$\int_0^1 \frac{1}{s^{\frac{q}{2}}\,[\rho(s)]^{q-1}}\,ds + \sup_{s\in[t,T)} \mathbb{E}\Big[\big|\big(F(0)\big)(s,x+W_s^0-W_t^0)\big|^q\Big] < \infty. \tag{52}$$

Then it holds for all $n \in \mathbb{N}$, $t \in [0,T)$, $x \in \mathbb{R}^d$, $\nu_0 \in \{1,2,\ldots,d+1\}$ that

$$\begin{aligned}
&\Big\|\mathbf{U}^{0,\nu_0}_{n,M}(t,x) - \mathbf{u}_{\nu_0}(t,x)\Big\|_{L^2(\mathbb{P};\mathbb{R})} \\
&\le \sum_{j=0}^{n-1} \sum_{\substack{\nu_1,\nu_2,\ldots,\nu_{j+1}\in\\ \{1,2,\ldots,d+1\}}} \tbinom{n-1}{j} \tfrac{\mathbb{1}_{\{1\}}(\nu_{j+1})2^j\left[\prod_{i=1}^j L_{\nu_i}\right]}{\sqrt{M^{n-j}}} \Big\| \Big(g(x+W_T^0-W_t^0) - g(x+W^0_{S(j,t)}-W_t^0)\Big) \\
&\qquad \cdot \big\langle \mathbf{e}_{\nu_j}, \big(1, \tfrac{W_T^0 - W^0_{S(j,t)}}{T-S(j,t)}\big)\big\rangle \prod_{i=1}^{j} \tfrac{1}{\varrho(S(i-1,t),S(i,t))} \big\langle \mathbf{e}_{\nu_{i-1}}, \big(1, \tfrac{W^0_{S(i,t)}-W^0_{S(i-1,t)}}{S(i,t)-S(i-1,t)}\big)\big\rangle \Big\|_{L^2(\mathbb{P};\mathbb{R})} \\
&+ \sum_{j=0}^{n-1} \sum_{\substack{\nu_1,\nu_2,\ldots,\nu_{j+1}\in\\ \{1,2,\ldots,d+1\}}} \tbinom{n-1}{j} \tfrac{\mathbb{1}_{\{1\}}(\nu_{j+1})2^j\left[\prod_{i=1}^j L_{\nu_i}\right]}{\sqrt{M^{n-j}}} \Big\| (F(0))(S(j+1,t), x+W^0_{S(j+1,t)} - W_t^0) \\
&\qquad \cdot \prod_{i=1}^{j+1} \tfrac{1}{\varrho(S(i-1,t),S(i,t))} \big\langle \mathbf{e}_{\nu_{i-1}}, \big(1, \tfrac{W^0_{S(i,t)}-W^0_{S(i-1,t)}}{S(i,t)-S(i-1,t)}\big)\big\rangle \Big\|_{L^2(\mathbb{P};\mathbb{R})} \\
&+ \sum_{j=0}^{n-1} \sum_{\substack{\nu_1,\nu_2,\ldots,\nu_{j+1}\in\\ \{1,2,\ldots,d+1\}}} \tbinom{n-1}{j} \tfrac{2^j\left[\prod_{i=1}^{j+1} L_{\nu_i}\right]}{\sqrt{M^{n-j-1}}} \Big\| \mathbf{u}_{\nu_{j+1}}(S(j+1,t), x+W^0_{S(j+1,t)} - W_t^0) \\
&\qquad \cdot \prod_{i=1}^{j+1} \tfrac{1}{\varrho(S(i-1,t),S(i,t))} \big\langle \mathbf{e}_{\nu_{i-1}}, \big(1, \tfrac{W^0_{S(i,t)}-W^0_{S(i-1,t)}}{S(i,t)-S(i-1,t)}\big)\big\rangle \Big\|_{L^2(\mathbb{P};\mathbb{R})}.
\end{aligned} \tag{53}$$

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 imply that all summands on the right-hand side of (34) are integrable. Lemma 3.2 yields that the summands in (34) are pairwise independent. Then (34) and the fact that the summation formula for variances of pairwise independent, integrable random variables imply that for all $m \in \mathbb{N}$, $x \in \mathbb{R}^d$, $t \in [0,T)$, $\nu \in \{1,2,\ldots,d+1\}$ it holds that

$$\begin{aligned}
&\mathrm{Var}\Big(\mathbf{U}^{0,\nu}_{m,M}(t,x)\Big) = \tfrac{1}{M^m}\mathrm{Var}\Big(\big(g(x+W_T^0-W_t^0)-g(x)\big)\,\big\langle \mathbf{e}_\nu, \big(1, \tfrac{W_T^0-W_t^0}{T-t}\big)\big\rangle\Big) \\
&+ \sum_{l=0}^{m-1} \tfrac{1}{M^{m-l}} \mathrm{Var}\left(\frac{\big(F(\mathbf{U}^{(0,l,1)}_{l,M}) - \mathbb{1}_{\mathbb{N}}(l)F(\mathbf{U}^{(0,-l,1)}_{l-1,M})\big)(\mathcal{R}_t^{(0,l,1)}, x+W^{(0,l,1)}_{\mathcal{R}_t^{(0,l,1)}} - W_t^{(0,l,1)})}{\varrho(t,\mathcal{R}_t^{(0,l,1)})} \Big\langle \mathbf{e}_\nu, \Big(1, \frac{W^{(0,l,1)}_{\mathcal{R}_t^{(0,l,1)}} - W_t^{(0,l,1)}}{\mathcal{R}_t^{(0,l,1)}-t}\Big)\Big\rangle\right) \\
&\le \tfrac{1}{M^m}\mathbb{E}\Big[\Big|\big(g(x+W_T^0-W_s^0)-g(x)\big)\,\big\langle \mathbf{e}_\nu, \big(1, \tfrac{W_T^0-W_s^0}{T-s}\big)\big\rangle\Big|^2\Big] \\
&+ \sum_{l=0}^{m-1} \tfrac{1}{M^{m-l}} \mathbb{E}\left[\left|\frac{\big(F(\mathbf{U}^{(0,l,1)}_{l,M}) - \mathbb{1}_{\mathbb{N}}(l)F(\mathbf{U}^{(0,-l,1)}_{l-1,M})\big)(\mathcal{R}_t^{(0,l,1)}, x+W^{(0,l,1)}_{\mathcal{R}_t^{(0,l,1)}} - W_t^{(0,l,1)})}{\varrho(t,\mathcal{R}_t^{(0,l,1)})} \Big\langle \mathbf{e}_\nu, \Big(1, \frac{W^{(0,l,1)}_{\mathcal{R}_t^{(0,l,1)}} - W_t^{(0,l,1)}}{\mathcal{R}_t^{(0,l,1)}-t}\Big)\Big\rangle\right|^2\right]
\end{aligned} \tag{54}$$

Combining this, the triangle inequality, and (31) yields that for all $m \in \mathbb{N}$, $x \in \mathbb{R}^d$, $t \in [0,T)$, $\nu \in \{1,2,\ldots,d+1\}$

it holds that

$$
\begin{aligned}
&\Big\|\mathbf{U}^{0,\nu}_{m,M}(t,x)-\mathbb{E}\Big[\mathbf{U}^{0,\nu}_{m,M}(t,x)\Big]\Big\|_{L^2(\mathbb{P};\mathbb{R})}=\Big(\mathrm{Var}\Big(\mathbf{U}^{0,\nu}_{m,M}(t,x)\Big)\Big)^{1/2}\\
&\le \tfrac{1}{\sqrt{M^m}}\Big\|\big(g(x+W^0_T-W^0_t)-g(x)\big)\left\langle \mathbf{e}_\nu,\big(1,\tfrac{W^0_T-W^0_t}{T-t}\big)\right\rangle\Big\|_{L^2(\mathbb{P};\mathbb{R})}\\
&+\sum_{l=0}^{m-1}\tfrac{1}{\sqrt{M^{m-l}}}\left\|\tfrac{\big(F(\mathbf{U}^{(0,l,1)}_{l,M})-\mathbb{1}_{\mathbb{N}}(l)F(\mathbf{U}^{(0,-l,1)}_{l-1,M})\big)(\mathcal{R}^{(0,l,1)}_t,x+W^{(0,l,1)}_{\mathcal{R}^{(0,l,1)}_t}-W^{(0,l,1)}_t)}{\varrho(t,\mathcal{R}^{(0,l,1)}_t)}\Big\langle \mathbf{e}_\nu,\Big(1,\tfrac{W^{(0,l,1)}_{\mathcal{R}^{(0,l,1)}_t}-W^{(0,l,1)}_t}{\mathcal{R}^{(0,l,1)}_t-t}\Big)\Big\rangle\right\|_{L^2(\mathbb{P};\mathbb{R})}\\
&\le \tfrac{1}{\sqrt{M^m}}\left(\Big\|\big(g(x+W^0_T-W^0_t)-g(x)\big)\left\langle \mathbf{e}_\nu,\big(1,\tfrac{W^0_T-W^0_t}{T-t}\big)\right\rangle\Big\|_{L^2(\mathbb{P};\mathbb{R})}+\left\|\tfrac{(F(0))(\mathcal{R}^0_t,x+W^0_{\mathcal{R}^0_t}-W^0_t)}{\varrho(t,\mathcal{R}^0_t)}\Big\langle \mathbf{e}_\nu,\Big(1,\tfrac{W^0_{\mathcal{R}^0_t}-W^0_t}{\mathcal{R}^0_t-t}\Big)\Big\rangle\right\|_{L^2(\mathbb{P};\mathbb{R})}\right)\\
&+\sum_{l=1}^{m-1}\tfrac{1}{\sqrt{M^{m-l}}}\left\|\sum_{\nu_1=1}^{d+1}L_{\nu_1}\tfrac{\left|\big(\mathbf{U}^{(0,l,1),\nu_1}_{l,M}-\mathbf{U}^{(0,-l,1),\nu_1}_{l-1,M}\big)(\mathcal{R}^{(0,l,1)}_t,x+W^{(0,l,1)}_{\mathcal{R}^{(0,l,1)}_t}-W^{(0,l,1)}_t)\right|}{\varrho(t,\mathcal{R}^{(0,l,1)}_t)}\left|\Big\langle \mathbf{e}_\nu,\Big(1,\tfrac{W^{(0,l,1)}_{\mathcal{R}^{(0,l,1)}_t}-W^{(0,l,1)}_t}{\mathcal{R}^{(0,l,1)}_t-t}\Big)\Big\rangle\right|\right\|_{L^2(\mathbb{P};\mathbb{R})}.
\end{aligned}
\tag{55}
$$

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

$$
\begin{aligned}
&\Big\|\mathbf{U}^{0,\nu}_{m,M}(t,x)-\mathbb{E}\Big[\mathbf{U}^{0,\nu}_{m,M}(t,x)\Big]\Big\|_{L^2(\mathbb{P};\mathbb{R})}\\
&\le \tfrac{1}{\sqrt{M^m}}\left(\Big\|\big(g(x+W^0_T-W^0_t)-g(x)\big)\left\langle \mathbf{e}_\nu,\big(1,\tfrac{W^0_T-W^0_t}{T-t}\big)\right\rangle\Big\|_{L^2(\mathbb{P};\mathbb{R})}+\left\|\tfrac{(F(0))(\mathcal{R}^0_t,x+W^0_{\mathcal{R}^0_t}-W^0_t)}{\varrho(t,\mathcal{R}^0_t)}\Big\langle \mathbf{e}_\nu,\Big(1,\tfrac{W^0_{\mathcal{R}^0_t}-W^0_t}{\mathcal{R}^0_t-t}\Big)\Big\rangle\right\|_{L^2(\mathbb{P};\mathbb{R})}\right)\\
&+\sum_{l=1}^{m-1}\sum_{\nu_1=1}^{d+1}\tfrac{L_{\nu_1}}{\sqrt{M^{m-l}}}\left\|\tfrac{(\mathbf{U}^{0,\nu_1}_{l,M}-\mathbf{u})(\mathcal{R}^0_t,x+W^0_{\mathcal{R}^0_t}-W^0_t)}{\varrho(t,\mathcal{R}^0_t)}\Big\langle \mathbf{e}_\nu,\Big(1,\tfrac{W^0_{\mathcal{R}^0_t}-W^0_t}{\mathcal{R}^0_t-t}\Big)\Big\rangle\right\|_{L^2(\mathbb{P};\mathbb{R})}\\
&+\sum_{l=1}^{m-1}\sum_{\nu_1=1}^{d+1}\tfrac{L_{\nu_1}}{\sqrt{M^{m-l}}}\left\|\tfrac{(\mathbf{U}^{0,\nu_1}_{l-1,M}-\mathbf{u})(\mathcal{R}^0_t,x+W^0_{\mathcal{R}^0_t}-W^0_t)}{\varrho(t,\mathcal{R}^0_t)}\Big\langle \mathbf{e}_\nu,\Big(1,\tfrac{W^0_{\mathcal{R}^0_t}-W^0_t}{\mathcal{R}^0_t-t}\Big)\Big\rangle\right\|_{L^2(\mathbb{P};\mathbb{R})}\\
&= \tfrac{1}{\sqrt{M^m}}\left(\Big\|\big(g(x+W^0_T-W^0_t)-g(x)\big)\left\langle \mathbf{e}_\nu,\big(1,\tfrac{W^0_T-W^0_t}{T-t}\big)\right\rangle\Big\|_{L^2(\mathbb{P};\mathbb{R})}+\left\|\tfrac{(F(0))(\mathcal{R}^0_t,x+W^0_{\mathcal{R}^0_t}-W^0_t)}{\varrho(t,\mathcal{R}^0_t)}\Big\langle \mathbf{e}_\nu,\Big(1,\tfrac{W^0_{\mathcal{R}^0_t}-W^0_t}{\mathcal{R}^0_t-t}\Big)\Big\rangle\right\|_{L^2(\mathbb{P};\mathbb{R})}\right)\\
&+\sum_{l=0}^{m-1}\sum_{\nu_1=1}^{d+1}\tfrac{L_{\nu_1}}{\sqrt{M^{m-l-1}}}\left(\tfrac{\mathbb{1}_{(0,m)}(l)}{\sqrt{M}}+\mathbb{1}_{(-1,m-1)}(l)\right)\left\|\tfrac{(\mathbf{U}^{0,\nu_1}_{l,M}-\mathbf{u})(\mathcal{R}^0_t,x+W^0_{\mathcal{R}^0_t}-W^0_t)}{\varrho(t,\mathcal{R}^0_t)}\Big\langle \mathbf{e}_\nu,\Big(1,\tfrac{W^0_{\mathcal{R}^0_t}-W^0_t}{\mathcal{R}^0_t-t}\Big)\Big\rangle\right\|_{L^2(\mathbb{P};\mathbb{R})}.
\end{aligned}
\tag{56}
$$

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 holds that

$$
\mathbb{E}\Big[\mathbf{U}^0_{m,M}(t,x)-g(x+W^0_T-W^0_t)\Big(1,\tfrac{W^0_T-W^0_t}{T-t}\Big)\Big]=\mathbb{E}\left[\int_t^T\big(F(\mathbf{U}^0_{m-1,M})\big)(s,x+W^0_s-W^0_t)\Big(1,\tfrac{W^0_s-W^0_t}{s-t}\Big)\,ds\right].
\tag{57}
$$

This, (33), linearity together with (32) and with Item (i) in Lemma 3.3, and Jensen's inequality show for all $m\in\mathbb{N}$, $t\in[0,T)$, $x\in\mathbb{R}^d$, $\nu\in\{1,2,\dots,d+1\}$ that

$$
\begin{aligned}
&\Big|\mathbb{E}\Big[\mathbf{U}^{0,\nu}_{m,M}(t,x)\Big]-\mathbf{u}_\nu(t,x)\Big|\\
&=\left|\mathbb{E}\left[\int_t^T\big(F(\mathbf{U}^0_{m-1,M})-F(\mathbf{u})\big)(s,x+W^0_s-W^0_t)\big\langle \mathbf{e}_\nu,\big(1,\tfrac{W^0_s-W^0_t}{s-t}\big)\big\rangle\,ds\right]\right|\\
&\le\sum_{\nu_1=1}^{d+1}L_{\nu_1}\mathbb{E}\left[\int_t^T\left|\Big(\mathbf{U}^{0,\nu_1}_{m-1,M}-\mathbf{u}_{\nu_1}\Big)(s,x+W^0_s-W^0_t)\right|\left|\big\langle \mathbf{e}_\nu,\big(1,\tfrac{W^0_s-W^0_t}{s-t}\big)\big\rangle\right|ds\right]\\
&=\sum_{\nu_1=1}^{d+1}L_{\nu_1}\mathbb{E}\left[\int_t^T\tfrac{\left|\big(\mathbf{U}^{0,\nu_1}_{m-1,M}-\mathbf{u}_{\nu_1}\big)(s,x+W^0_s-W^0_t)\right|}{\varrho(t,s)}\left|\big\langle \mathbf{e}_\nu,\big(1,\tfrac{W^0_s-W^0_t}{s-t}\big)\big\rangle\right|\varrho(t,s)\,ds\right]\\
&=\sum_{\nu_1=1}^{d+1}L_{\nu_1}\mathbb{E}\left[\tfrac{\left|\big(\mathbf{U}^{0,\nu_1}_{m-1,M}-\mathbf{u}_{\nu_1}\big)(\mathcal{R}^0_t,x+W^0_{\mathcal{R}^0_t}-W^0_t)\right|}{\varrho(t,\mathcal{R}^0_t)}\left|\Big\langle \mathbf{e}_\nu,\Big(1,\tfrac{W^0_{\mathcal{R}^0_t}-W^0_t}{\mathcal{R}^0_t-t}\Big)\Big\rangle\right|\right]\\
&\le\sum_{\nu_1=1}^{d+1}L_{\nu_1}\left\|\tfrac{\big(\mathbf{U}^{0,\nu_1}_{m-1,M}-\mathbf{u}_{\nu_1}\big)(\mathcal{R}^0_t,x+W^0_{\mathcal{R}^0_t}-W^0_t)}{\varrho(t,\mathcal{R}^0_t)}\Big\langle \mathbf{e}_\nu,\Big(1,\tfrac{W^0_{\mathcal{R}^0_t}-W^0_t}{\mathcal{R}^0_t-t}\Big)\Big\rangle\right\|_{L^2(\mathbb{P};\mathbb{R})}.
\end{aligned}
\tag{58}
$$

In the next step we combine the established bounds for the Monte Carlo error and for the time discretization error to obtain a bound for the *global error*. More formally, observe that (56) and (58) ensure that for all $m \in \mathbb{N}$, $t \in [0,T)$, $x \in \mathbb{R}^d$, $\nu \in \{1,2,\dots,d+1\}$ it holds that

$$
\begin{aligned}
&\Big\|\mathbf{U}^{0,\nu}_{m,M}(t,x)-\mathbf{u}_\nu(t,x)\Big\|_{L^2(\mathbb{P};\mathbb{R})}\\
&\le \Big\|\mathbf{U}^{0,\nu}_{m,M}(t,x)-\mathbb{E}\Big[\mathbf{U}^{0,\nu}_{m,M}(t,x)\Big]\Big\|_{L^2(\mathbb{P};\mathbb{R})}+\Big|\mathbb{E}\Big[\mathbf{U}^{0,\nu}_{m,M}(t,x)\Big]-\mathbf{u}_\nu(t,x)\Big|\\
&\le \tfrac{1}{\sqrt{M^m}}\Big\|\big(g(x+W^0_T-W^0_t)-g(x)\big)\big\langle \mathbf{e}_\nu,\big(1,\tfrac{W^0_T-W^0_t}{T-t}\big)\big\rangle\Big\|_{L^2(\mathbb{P};\mathbb{R})}\\
&\quad+\tfrac{1}{\sqrt{M^m}}\Big\|\tfrac{(F(0))(\mathcal{R}^0_t,x+W^0_{\mathcal{R}^0_t}-W^0_t)}{\varrho(t,\mathcal{R}^0_t)}\Big\langle \mathbf{e}_\nu,\Big(1,\tfrac{W^0_{\mathcal{R}^0_t}-W^0_t}{\mathcal{R}^0_t-t}\Big)\Big\rangle\Big\|_{L^2(\mathbb{P};\mathbb{R})}\\
&\quad+\sum_{l=0}^{m-1}\sum_{\nu_1=1}^{d+1}\tfrac{L_{\nu_1}}{\sqrt{M^{m-l-1}}}\Big(\tfrac{\mathbb{1}_{(0,m)}(l)}{\sqrt{M}}+\mathbb{1}_{(-1,m-1)}(l)\Big)\Big\|\tfrac{(\mathbf{U}^{0,\nu_1}_{l,M}-\mathbf{u}_{\nu_1})(\mathcal{R}^0_t,x+W^0_{\mathcal{R}^0_t}-W^0_t)}{\varrho(t,\mathcal{R}^0_t)}\Big\langle \mathbf{e}_\nu,\Big(1,\tfrac{W^0_{\mathcal{R}^0_t}-W^0_t}{\mathcal{R}^0_t-t}\Big)\Big\rangle\Big\|_{L^2(\mathbb{P};\mathbb{R})}\\
&\quad+\sum_{\nu_1=1}^{d+1}L_{\nu_1}\Big\|\tfrac{(\mathbf{U}^{0,\nu_1}_{m-1,M}-\mathbf{u}_{\nu_1})(\mathcal{R}^0_t,x+W^0_{\mathcal{R}^0_t}-W^0_t)}{\varrho(t,\mathcal{R}^0_t)}\Big\langle \mathbf{e}_\nu,\Big(1,\tfrac{W^0_{\mathcal{R}^0_t}-W^0_t}{\mathcal{R}^0_t-t}\Big)\Big\rangle\Big\|_{L^2(\mathbb{P};\mathbb{R})}\\
&\le \tfrac{1}{\sqrt{M^m}}\Big\|\big(g(x+W^0_T-W^0_t)-g(x)\big)\big\langle \mathbf{e}_\nu,\big(1,\tfrac{W^0_T-W^0_t}{T-t}\big)\big\rangle\Big\|_{L^2(\mathbb{P};\mathbb{R})}\\
&\quad+\tfrac{1}{\sqrt{M^m}}\Big\|\tfrac{(F(0))(\mathcal{R}^0_t,x+W^0_{\mathcal{R}^0_t}-W^0_t)}{\varrho(t,\mathcal{R}^0_t)}\Big\langle \mathbf{e}_\nu,\Big(1,\tfrac{W^0_{\mathcal{R}^0_t}-W^0_t}{\mathcal{R}^0_t-t}\Big)\Big\rangle\Big\|_{L^2(\mathbb{P};\mathbb{R})}\\
&\quad+\sum_{\nu_1=1}^{d+1}\tfrac{L_{\nu_1}}{\sqrt{M^{m-1}}}\Big\|\tfrac{\mathbf{u}_{\nu_1}(\mathcal{R}^0_t,x+W^0_{\mathcal{R}^0_t}-W^0_t)}{\varrho(t,\mathcal{R}^0_t)}\Big\langle \mathbf{e}_\nu,\Big(1,\tfrac{W^0_{\mathcal{R}^0_t}-W^0_t}{\mathcal{R}^0_t-t}\Big)\Big\rangle\Big\|_{L^2(\mathbb{P};\mathbb{R})}\\
&\quad+\sum_{l=1}^{m-1}\sum_{\nu_1=1}^{d+1}\tfrac{2L_{\nu_1}}{\sqrt{M^{m-l-1}}}\Big\|\tfrac{(\mathbf{U}^{0,\nu_1}_{l,M}-\mathbf{u}_{\nu_1})(\mathcal{R}^0_t,x+W^0_{\mathcal{R}^0_t}-W^0_t)}{\varrho(t,\mathcal{R}^0_t)}\Big\langle \mathbf{e}_\nu,\Big(1,\tfrac{W^0_{\mathcal{R}^0_t}-W^0_t}{\mathcal{R}^0_t-t}\Big)\Big\rangle\Big\|_{L^2(\mathbb{P};\mathbb{R})}.
\end{aligned}
\tag{59}
$$

We next iterate this inequality. More precisely, we show that it holds for all $n,k \in \mathbb{N}$, $t \in [0,T)$, $x \in \mathbb{R}^d$, $\nu_0 \in \{1,2,\dots,d+1\}$ that

$$
\begin{aligned}
&\Big\|\mathbf{U}^{0,\nu_0}_{n,M}(t,x)-\mathbf{u}_{\nu_0}(t,x)\Big\|_{L^2(\mathbb{P};\mathbb{R})}\\
&\le\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{\nu_1,\nu_2,\dots,\nu_{j+1}\in\\ \{1,2,\dots,d+1\}}}\tfrac{\mathbb{1}_{\{1\}}(\nu_{j+1})2^j\big[\prod_{i=1}^j L_{\nu_i}\big]}{\sqrt{M^{n-j}}}\Big\|\Big(g(x+W^0_T-W^0_t)-g(x+W^0_{S(j,t)}-W^0_t)\Big)\\
&\qquad\cdot\big\langle \mathbf{e}_{\nu_j},\big(1,\tfrac{W^0_T-W^0_{S(j,t)}}{T-S(j,t)}\big)\big\rangle\prod_{i=1}^{j}\tfrac{1}{\varrho(S(i-1,t),S(i,t))}\big\langle \mathbf{e}_{\nu_{i-1}},\big(1,\tfrac{W^0_{S(i,t)}-W^0_{S(i-1,t)}}{S(i,t)-S(i-1,t)}\big)\big\rangle\Big\|_{L^2(\mathbb{P};\mathbb{R})}\\
&\quad+\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{\nu_1,\nu_2,\dots,\nu_{j+1}\in\\ \{1,2,\dots,d+1\}}}\tfrac{\mathbb{1}_{\{1\}}(\nu_{j+1})2^j\big[\prod_{i=1}^j L_{\nu_i}\big]}{\sqrt{M^{n-j}}}\Big\|(F(0))(S(j+1,t),x+W^0_{S(j+1,t)}-W^0_t)\\
&\qquad\cdot\prod_{i=1}^{j+1}\tfrac{1}{\varrho(S(i-1,t),S(i,t))}\big\langle \mathbf{e}_{\nu_{i-1}},\big(1,\tfrac{W^0_{S(i,t)}-W^0_{S(i-1,t)}}{S(i,t)-S(i-1,t)}\big)\big\rangle\Big\|_{L^2(\mathbb{P};\mathbb{R})}\\
&\quad+\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{\nu_1,\nu_2,\dots,\nu_{j+1}\in\\ \{1,2,\dots,d+1\}}}\tfrac{2^j\big[\prod_{i=1}^{j+1} L_{\nu_i}\big]}{\sqrt{M^{n-j-1}}}\Big\|\mathbf{u}_{\nu_{j+1}}(S(j+1,t),x+W^0_{S(j+1,t)}-W^0_t)\\
&\qquad\cdot\prod_{i=1}^{j+1}\tfrac{1}{\varrho(S(i-1,t),S(i,t))}\big\langle \mathbf{e}_{\nu_{i-1}},\big(1,\tfrac{W^0_{S(i,t)}-W^0_{S(i-1,t)}}{S(i,t)-S(i-1,t)}\big)\big\rangle\Big\|_{L^2(\mathbb{P};\mathbb{R})}\\
&\quad+\sum_{\substack{l_1,l_2,\dots,l_k\in\mathbb{N},\\ l_1<l_2<\dots<l_k<n}}\sum_{\substack{\nu_1,\nu_2,\dots,\nu_k\in\\ \{1,2,\dots,d+1\}}}\tfrac{2^k\big[\prod_{i=1}^k L_{\nu_i}\big]}{\sqrt{M^{n-k-l_1}}}\Big\|\Big(\mathbf{U}^{0,\nu_k}_{l_1,M}-\mathbf{u}_{\nu_k}\Big)(S(k,t),x+W^0_{S(k,t)}-W^0_t)\\
&\qquad\cdot\prod_{i=1}^{k}\tfrac{1}{\varrho(S(i-1,t),S(i,t))}\big\langle \mathbf{e}_{\nu_{i-1}},\big(1,\tfrac{W^0_{S(i,t)}-W^0_{S(i-1,t)}}{S(i,t)-S(i-1,t)}\big)\big\rangle\Big\|_{L^2(\mathbb{P};\mathbb{R})}.
\end{aligned}
\tag{60}
$$

We prove (60) by induction on $k \in \mathbb{N}$. The base case $k=1$ follows immediately from (59). For the induction step $\mathbb{N} \ni k \rightsquigarrow k+1 \in \mathbb{N}$ let $k \in \mathbb{N}$ and assume that (60) holds for k. Inequality (59) and independence of $(W^0, \mathfrak{r}^{(m)}, \mathbf{U}^0_{m,M})_{m\in\mathbb{N}_0}$ (see Item (iv) in Lemma 3.2) yield that for all $l_1 \in \mathbb{N}$, $x \in \mathbb{R}^d$, $t \in [0,T)$, $\nu_0, \nu_1, \ldots, \nu_k \in \{1,2,\ldots,d+1\}$ it holds that

$$
\begin{aligned}
&\left\| \left(\mathbf{U}^{0,\nu_k}_{l_1,M} - \mathbf{u}_{\nu_k}\right)(S(k,t), x + W^0_{S(k,t)} - W^0_t) \prod_{i=1}^{k} \tfrac{1}{\varrho(S(i-1,t),S(i,t))} \left\langle \mathbf{e}_{\nu_{i-1}}, \left(1, \tfrac{W^0_{S(i,t)} - W^0_{S(i-1,t)}}{S(i,t)-S(i-1,t)}\right)\right\rangle \right\|_{L^2(\mathbb{P};\mathbb{R})} \\
&= \left\| \left\| \left(\mathbf{U}^{0,\nu_k}_{l_1,M} - \mathbf{u}_{\nu_k}\right)(s,y) \right\|_{L^2(\mathbb{P};\mathbb{R})} \Big|_{(s,y)=(S(k,t),x+W^0_{S(k,t)}-W^0_t)} \prod_{i=1}^{k} \tfrac{1}{\varrho(S(i-1,t),S(i,t))} \left\langle \mathbf{e}_{\nu_{i-1}}, \left(1, \tfrac{W^0_{S(i,t)} - W^0_{S(i-1,t)}}{S(i,t)-S(i-1,t)}\right)\right\rangle \right\|_{L^2(\mathbb{P};\mathbb{R})} \\
&\le \tfrac{1}{\sqrt{M^{l_1}}} \left\| \left(g(x+W^0_T - W^0_t) - g(x + W^0_{S(k,t)} - W^0_t)\right) \left\langle \mathbf{e}_{\nu_k}, \left(1, \tfrac{W^0_T - W^0_{S(k,t)}}{T-S(k,t)}\right)\right\rangle \prod_{i=1}^{k} \frac{\left\langle \mathbf{e}_{\nu_{i-1}}, \left(1, \frac{W^0_{S(i,t)} - W^0_{S(i-1,t)}}{S(i,t)-S(i-1,t)}\right)\right\rangle}{\varrho(S(i-1,t),S(i,t))} \right\|_{L^2(\mathbb{P};\mathbb{R})} \\
&+ \tfrac{1}{\sqrt{M^{l_1}}} \left\| (F(0))\,(S(k+1,t), x + W_{S(k+1,t)} - W^0_t) \prod_{i=1}^{k+1} \frac{\left\langle \mathbf{e}_{\nu_{i-1}}, \left(1, \frac{W^0_{S(i,t)} - W^0_{S(i-1,t)}}{S(i,t)-S(i-1,t)}\right)\right\rangle}{\varrho(S(i-1,t),S(i,t))} \right\|_{L^2(\mathbb{P};\mathbb{R})} \\
&+ \sum_{\nu_{k+1}=1}^{d+1} \tfrac{L_{\nu_{k+1}}}{\sqrt{M^{l_1-1}}} \left\| \mathbf{u}_{\nu_{k+1}}(S(k+1,t), x + W_{S(k+1,t)} - W^0_t) \prod_{i=1}^{k+1} \frac{\left\langle \mathbf{e}_{\nu_{i-1}}, \left(1, \frac{W^0_{S(i,t)} - W^0_{S(i-1,t)}}{S(i,t)-S(i-1,t)}\right)\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_{\nu_{k+1}=1}^{d+1} \tfrac{2L_{\nu_{k+1}}}{\sqrt{M^{l_1-l_0-1}}} \left\| (\mathbf{U}^{0,\nu_{k+1}}_{l_0,M} - \mathbf{u}_{\nu_{k+1}})(S(k+1,t), x + W_{S(k+1,t)} - W^0_t) \prod_{i=1}^{k+1} \frac{\left\langle \mathbf{e}_{\nu_{i-1}}, \left(1, \frac{W^0_{S(i,t)} - W^0_{S(i-1,t)}}{S(i,t)-S(i-1,t)}\right)\right\rangle}{\varrho(S(i-1,t),S(i,t))} \right\|_{L^2(\mathbb{P};\mathbb{R})}.
\end{aligned}
\tag{61}
$$

This and the induction hypothesis complete the induction step $\mathbb{N} \ni k \rightsquigarrow k+1 \in \mathbb{N}$. Induction hence establishes (60). Applying (60) with $k = n$ yields for all $n \in \mathbb{N}$, $t \in [0,T)$, $x \in \mathbb{R}^d$, $\nu_0 \in \{1,2,\ldots,d+1\}$ that

$$
\begin{aligned}
&\left\| \mathbf{U}^{0,\nu_0}_{n,M}(t,x) - \mathbf{u}_{\nu_0}(t,x) \right\|_{L^2(\mathbb{P};\mathbb{R})} \\
&\le \sum_{j=0}^{n-1} \sum_{\substack{l_1,l_2,\ldots,l_{j+1}\in\mathbb{N}, \\ l_1<l_2<\ldots<l_j<l_{j+1}=n}} \sum_{\substack{\nu_1,\nu_2,\ldots,\nu_{j+1}\in \\ \{1,2,\ldots,d+1\}}} \tfrac{\mathbb{1}_{\{1\}}(\nu_{j+1}) 2^j \left[\prod_{i=1}^{j} L_{\nu_i}\right]}{\sqrt{M^{n-j}}} \Big\| \left(g(x+W^0_T - W^0_t) - g(x + W^0_{S(j,t)} - W^0_t)\right) \\
&\qquad \cdot \left\langle \mathbf{e}_{\nu_j}, \left(1, \tfrac{W^0_T - W^0_{S(j,t)}}{T-S(j,t)}\right)\right\rangle \prod_{i=1}^{j} \tfrac{1}{\varrho(S(i-1,t),S(i,t))} \left\langle \mathbf{e}_{\nu_{i-1}}, \left(1, \tfrac{W^0_{S(i,t)} - W^0_{S(i-1,t)}}{S(i,t)-S(i-1,t)}\right)\right\rangle \Big\|_{L^2(\mathbb{P};\mathbb{R})} \\
&+ \sum_{j=0}^{n-1} \sum_{\substack{l_1,l_2,\ldots,l_{j+1}\in\mathbb{N}, \\ l_1<l_2<\ldots<l_j<l_{j+1}=n}} \sum_{\substack{\nu_1,\nu_2,\ldots,\nu_{j+1}\in \\ \{1,2,\ldots,d+1\}}} \tfrac{\mathbb{1}_{\{1\}}(\nu_{j+1}) 2^j \left[\prod_{i=1}^{j} L_{\nu_i}\right]}{\sqrt{M^{n-j}}} \Big\| (F(0))(S(j+1,t), x + W^0_{S(j+1,t)} - W^0_t) \\
&\qquad \cdot \prod_{i=1}^{j+1} \tfrac{1}{\varrho(S(i-1,t),S(i,t))} \left\langle \mathbf{e}_{\nu_{i-1}}, \left(1, \tfrac{W^0_{S(i,t)} - W^0_{S(i-1,t)}}{S(i,t)-S(i-1,t)}\right)\right\rangle \Big\|_{L^2(\mathbb{P};\mathbb{R})} \\
&+ \sum_{j=0}^{n-1} \sum_{\substack{l_1,l_2,\ldots,l_{j+1}\in\mathbb{N}, \\ l_1<l_2<\ldots<l_j<l_{j+1}=n}} \sum_{\substack{\nu_1,\nu_2,\ldots,\nu_{j+1}\in \\ \{1,2,\ldots,d+1\}}} \tfrac{2^j \left[\prod_{i=1}^{j+1} L_{\nu_i}\right]}{\sqrt{M^{n-j-1}}} \Big\| \mathbf{u}_{\nu_{j+1}}(S(j+1,t), x + W^0_{S(j+1,t)} - W^0_t) \\
&\qquad \cdot \prod_{i=1}^{j+1} \tfrac{1}{\varrho(S(i-1,t),S(i,t))} \left\langle \mathbf{e}_{\nu_{i-1}}, \left(1, \tfrac{W^0_{S(i,t)} - W^0_{S(i-1,t)}}{S(i,t)-S(i-1,t)}\right)\right\rangle \Big\|_{L^2(\mathbb{P};\mathbb{R})}.
\end{aligned}
\tag{62}
$$

This and the fact that for all $n \in \mathbb{N}$ and $j \in \{0,1,\ldots,n-1\}$ it holds that

$$
\sum_{\substack{l_1,l_2,\ldots,l_{j+1}\in\mathbb{N}, \\ l_1<l_2<\ldots<l_j<l_{j+1}=n}} 1 = \binom{n-1}{j}. \tag{63}
$$

proves (53). This finishes the proof of Lemma 3.4. ☐

Proposition 3.5 (Global approximation error)**.** *Assume Setting 3.1, let* $t \in [0,T)$, $x \in \mathbb{R}^d$, $\nu_0 \in \{1,2,\ldots,d+1\}$, $M, n \in \mathbb{N}$, $p \in [2,\infty)$, $\alpha \in (\frac{p-2}{2(p-1)}, \frac{p}{2(p-1)})$, $\beta = \frac{\alpha}{2} - \frac{(1-\alpha)(p-2)}{2p}$, $C \in \mathbb{R}$ *satisfy that*

$$
C = \max\left\{1, 2(T-t)^{\frac{1}{2}} |\Gamma(\tfrac{p}{2})|^{\frac{1}{p}} (1-\alpha)^{\frac{1}{p}-1} \max\{1, \|L\|_1\} \max\left\{(T-t)^{\frac{1}{2}}, 2^{\frac{1}{2}} |\Gamma(\tfrac{p+1}{2})|^{\frac{1}{p}} \pi^{-\frac{1}{2p}}\right\}\right\}, \tag{64}
$$

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

$$\left\|\mathbf{U}^{0,\nu_0}_{n,M}(t,x) - \mathbf{u}_{\nu_0}(t,x)\right\|_{L^2(\mathbb{P};\mathbb{R})} \leq \tfrac{1}{4}\left[1+\tfrac{pn}{2}\right]^{\frac{1}{8}} M^{-\frac{n}{2}}(2C)^n \exp\left(\tfrac{1}{8}+\beta M^{\frac{1}{2\beta}}\right)\Big[2C^{-1}\sqrt{\max\{T-t,3\}}\|K\|_1$$
$$+ \sup_{s\in[t,T)} \left\|(F(0))(s,x+W^0_s-W^0_t)\right\|_{L^{\frac{2p}{p-2}}(\mathbb{P};\mathbb{R})} + \sqrt{M}\sup_{s\in[t,T)}\max_{i\in\{1,2,\dots,d+1\}}\left\|\mathbf{u}_i(s,x+W^0_s-W^0_t)\right\|_{L^{\frac{2p}{p-2}}(\mathbb{P};\mathbb{R})}\Big]. \tag{65}$$

Proof of Proposition 3.5. Throughout this proof let $C_1 \in [0,\infty)$ satisfy

$$C_1 = \max\left\{T-t, 2|\Gamma(\tfrac{p+1}{2})|^{\frac{2}{p}}\pi^{-\frac{1}{p}}\right\}(T-t)|\Gamma(\tfrac{p}{2})|^{\frac{2}{p}}(1-\alpha)^{\frac{2}{p}-2}. \tag{66}$$

Without loss of generality we assume that $\sup_{s\in[t,T)}\left\||(F(0))(s,x+W^0_s-W^0_t)|\right\|_{L^{\frac{2p}{p-2}}(\mathbb{P};\mathbb{R})} < \infty$ (otherwise the assertion is trivial). It follows from (31) and the triangle inequality that for all $\nu \in \{1,2,\dots,d+1\}$, $s \in [0,T)$ it holds that

$$\begin{aligned}
&\sum_{\alpha=1}^d K_\alpha \left\|(W^{0,\alpha}_T - W^{0,\alpha}_s)\langle \mathbf{e}_\nu, (1, \tfrac{W^0_T-W^0_s}{T-s})\rangle\right\|_{L^2(\mathbb{P};\mathbb{R})}\\
&= \sum_{\alpha=1}^d K_\alpha\left(\sqrt{T-s}\mathbb{1}_{\{1\}}(\nu) + \tfrac{\mathbb{1}_{[2,\infty)}(\nu)}{T-s}\|(W^{0,\alpha}_T-W^{0,\alpha}_s)(W^{0,\nu-1}_T-W^{0,\nu-1}_s)\|_{L^2(\mathbb{P};\mathbb{R})}\right)\\
&= \sqrt{T-s}\|K\|_1\mathbb{1}_{\{1\}}(\nu) + \tfrac{\mathbb{1}_{[2,\infty)}(\nu)}{T-s}\left(K_{\nu-1}\|(W^{0,\nu-1}_T-W^{0,\nu-1}_s)^2\|_{L^2(\mathbb{P};\mathbb{R})} + \sum_{\alpha\in\{1,2,\dots,d\}\setminus\{\nu-1\}} K_\alpha\|W^{0,\alpha}_T-W^{0,\alpha}_s\|^2_{L^2(\mathbb{P};\mathbb{R})}\right)\\
&= \sqrt{T-s}\|K\|_1\mathbb{1}_{\{1\}}(\nu) + \mathbb{1}_{[2,\infty)}(\nu)\left(\sqrt{3}K_{\nu-1} + \sum_{\alpha\in\{1,2,\dots,d\}\setminus\{\nu-1\}} K_\alpha\right)\\
&\leq \max\{\sqrt{T-s},\sqrt{3}\}\|K\|_1.
\end{aligned} \tag{67}$$

Note that the fact that $p \in [2,\infty)$ and the fact that $\alpha \in \left(\frac{p-2}{2(p-1)}, \frac{p}{2(p-1)}\right)$ imply that $\frac{p}{2} \in [\alpha(p-1), \alpha(p-1)+1]$ and that $\alpha \in (0,1)$. This, (67), (31), and Corollary 2.5 (with $p=2$ in the notation of Corollary 2.5) show that for all $j \in \{0,1,\dots,n-1\}$, $\nu_1,\nu_2,\dots,\nu_j \in \{1,2,\dots,d+1\}$ it holds that

$$\begin{aligned}
&\left\|\left(g(x+W^0_T-W^0_t) - g(x+W^0_{S(j,t)}-W^0_t)\right)\langle \mathbf{e}_{\nu_j}, (1, \tfrac{W^0_T-W^0_{S(j,t)}}{T-S(j,t)})\rangle \prod_{i=1}^j \tfrac{1}{\varrho(S(i-1,t),S(i,t))}\langle \mathbf{e}_{\nu_{i-1}}, (1, \tfrac{W^0_{S(i,t)}-W^0_{S(i-1,t)}}{S(i,t)-S(i-1,t)})\rangle\right\|_{L^2(\mathbb{P};\mathbb{R})}\\
&\leq \left\|\left(\sum_{\alpha=1}^d K_\alpha\left|W^{0,\alpha}_T - W^{0,\alpha}_{S(j,t)}\right|\right)\langle \mathbf{e}_{\nu_j}, (1, \tfrac{W^0_T-W^0_{S(j,t)}}{T-S(j,t)})\rangle \prod_{i=1}^j \tfrac{1}{\varrho(S(i-1,t),S(i,t))}\langle \mathbf{e}_{\nu_{i-1}}, (1, \tfrac{W^0_{S(i,t)}-W^0_{S(i-1,t)}}{S(i,t)-S(i-1,t)})\rangle\right\|_{L^2(\mathbb{P};\mathbb{R})}\\
&= \left\|\left\|\left(\sum_{\alpha=1}^d K_\alpha\left|W^{0,\alpha}_T - W^{0,\alpha}_s\right|\right)\langle \mathbf{e}_{\nu_j}, (1, \tfrac{W^0_T-W^0_s}{T-s})\rangle\right\|_{L^2(\mathbb{P};\mathbb{R})}\Big|_{s=S(j,t)} \prod_{i=1}^j \tfrac{1}{\varrho(S(i-1,t),S(i,t))}\langle \mathbf{e}_{\nu_{i-1}}, (1, \tfrac{W^0_{S(i,t)}-W^0_{S(i-1,t)}}{S(i,t)-S(i-1,t)})\rangle\right\|_{L^2(\mathbb{P};\mathbb{R})}\\
&\leq \left\|\left(\sum_{\alpha=1}^d K_\alpha\left\|\left(W^{0,\alpha}_T - W^{0,\alpha}_s\right)\langle \mathbf{e}_{\nu_j}, (1, \tfrac{W^0_T-W^0_s}{T-s})\rangle\right\|_{L^2(\mathbb{P};\mathbb{R})}\Big|_{s=S(j,t)}\right) \prod_{i=1}^j \tfrac{1}{\varrho(S(i-1,t),S(i,t))}\langle \mathbf{e}_{\nu_{i-1}}, (1, \tfrac{W^0_{S(i,t)}-W^0_{S(i-1,t)}}{S(i,t)-S(i-1,t)})\rangle\right\|_{L^2(\mathbb{P};\mathbb{R})}\\
&\leq \sqrt{\max\{T-t,3\}}\|K\|_1\left\|\prod_{i=1}^j \tfrac{1}{\varrho(S(i-1,t),S(i,t))}\langle \mathbf{e}_{\nu_{i-1}}, (1, \tfrac{W^0_{S(i,t)}-W^0_{S(i-1,t)}}{S(i,t)-S(i-1,t)})\rangle\right\|_{L^2(\mathbb{P};\mathbb{R})}\\
&\leq \sqrt{\max\{T-t,3\}}\|K\|_1\left[\frac{(T-t)\max\left\{T-t, 2\frac{\Gamma(\frac{3}{2})}{\sqrt{\pi}}\right\}}{(1-\alpha)}\right]^{\frac{j}{2}}(ej)^{\frac{1}{8}}\left[\tfrac{1}{\Gamma(j+1)}\right]^{\frac{\alpha}{2}}.
\end{aligned} \tag{68}$$

The facts that $\Gamma\left(\frac{p+1}{2}\right) \geq \Gamma\left(\frac{3}{2}\right) = \frac{\sqrt{\pi}}{2}$, that $\frac{2(p-1)}{p} \geq 1$, and that $\alpha \leq 1$ prove that

$$\frac{(T-t)\max\left\{T-t, 2\frac{\Gamma(\frac{3}{2})}{\sqrt{\pi}}\right\}}{(1-\alpha)} \leq C_1. \tag{69}$$

Moreover, the fact that $p \geq 2$ ensures that for all $j \in \mathbb{N}_0$ it holds that $(ej)^{\frac{1}{8}} \leq \left[e\left(\frac{pj}{2}+1\right)\right]^{\frac{1}{8}}$ and that $\Gamma(j+1)^{\frac{\alpha}{2}} \geq \Gamma(j+1)^{\beta}$. This together with (68) and (69) proves that for all $j \in \{0,1,\ldots,n-1\}$, $\nu_1,\nu_2,\ldots,\nu_j \in \{1,2,\ldots,d+1\}$ it holds that

$$\begin{aligned}
&\left\|\Big(g(x+W^0_T-W^0_t)-g(x+W^0_{S(j,t)}-W^0_t)\Big)\left\langle \mathbf{e}_{\nu_j},\big(1,\tfrac{W^0_T-W^0_{S(j,t)}}{T-S(j,t)}\big)\right\rangle \prod_{i=1}^{j}\tfrac{1}{\varrho(S(i-1,t),S(i,t))}\left\langle \mathbf{e}_{\nu_{i-1}},\big(1,\tfrac{W^0_{S(i,t)}-W^0_{S(i-1,t)}}{S(i,t)-S(i-1,t)}\big)\right\rangle\right\|_{L^2(\mathbb{P};\mathbb{R})}\\
&\leq \sqrt{\max\{T-t,3\}}\|K\|_1 C_1^{\frac{j}{2}}\left[e\left(\tfrac{pj}{2}+1\right)\right]^{\frac{1}{8}}\left[\tfrac{1}{\Gamma(j+1)}\right]^{\beta}.
\end{aligned}\tag{70}$$

Corollary 2.5 and the facts that $p \geq 2$ and $\alpha > \frac{p-2}{2(p-1)}$ prove that

$$\begin{aligned}
&\left\|\prod_{i=1}^{j+1}\tfrac{1}{\varrho(S(i-1,t),S(i,t))}\left\langle \mathbf{e}_{\nu_{i-1}},\big(1,\tfrac{W^0_{S(i,t)}-W^0_{S(i-1,t)}}{S(i,t)-S(i-1,t)}\big)\right\rangle\right\|_{L^p(\mathbb{P};\mathbb{R})}\\
&\leq \left[\max\left\{T-t,2\tfrac{\Gamma(\frac{p+1}{2})^{\frac{2}{p}}}{\pi^{\frac{1}{p}}}\right\}\tfrac{(T-t)\Gamma(\frac{p}{2})^{\frac{2}{p}}}{(1-\alpha)^{\frac{2(p-1)}{p}}(\frac{p}{2})^{\frac{2(\alpha(p-1)-\frac{p}{2}+1)}{p}}}\right]^{\frac{j+1}{2}}\left[e^{\frac{p}{2}}\left(\tfrac{pj}{2}+1\right)\right]^{\frac{1}{2p^2}}\left[\tfrac{\Gamma(\frac{2}{p})}{\Gamma(1+j+\frac{2}{p})}\right]^{\frac{\alpha}{2}-\frac{(1-\alpha)(p-2)}{2p}}\\
&\leq \left[\max\left\{T-t,2\tfrac{\Gamma(\frac{p+1}{2})^{\frac{2}{p}}}{\pi^{\frac{1}{p}}}\right\}\tfrac{(T-t)\Gamma(\frac{p}{2})^{\frac{2}{p}}}{(1-\alpha)^{\frac{2(p-1)}{p}}}\right]^{\frac{j+1}{2}}\left[e\left(\tfrac{pj}{2}+1\right)\right]^{\frac{1}{8}}\left[\tfrac{1}{\Gamma(1+j+\frac{2}{p})}\right]^{\frac{\alpha}{2}-\frac{(1-\alpha)(p-2)}{2p}}\\
&= C_1^{\frac{j+1}{2}}\left[e\left(\tfrac{pj}{2}+1\right)\right]^{\frac{1}{8}}\left[\tfrac{1}{\Gamma(1+j+\frac{2}{p})}\right]^{\beta}
\end{aligned}\tag{71}$$

This together with Hölder's inequality and independence of $(\mathfrak{r}^{(n)})_{n\in\mathbb{N}}$ and W^0 proves that for all $j \in \{0,1,\ldots,n-1\}$, $\nu_1,\nu_2,\ldots,\nu_{j+1} \in \{1,2,\ldots,d+1\}$ it holds that

$$\begin{aligned}
&\left\|(F(0))(S(j+1,t),x+W^0_{S(j+1,t)}-W^0_t)\prod_{i=1}^{j+1}\tfrac{1}{\varrho(S(i-1,t),S(i,t))}\left\langle \mathbf{e}_{\nu_{i-1}},\big(1,\tfrac{W^0_{S(i,t)}-W^0_{S(i-1,t)}}{S(i,t)-S(i-1,t)}\big)\right\rangle\right\|_{L^2(\mathbb{P};\mathbb{R})}\\
&\leq \left\|(F(0))(S(j+1,t),x+W^0_{S(j+1,t)}-W^0_t)\right\|_{L^{\frac{2p}{p-2}}(\mathbb{P};\mathbb{R})}\left\|\prod_{i=1}^{j+1}\tfrac{1}{\varrho(S(i-1,t),S(i,t))}\left\langle \mathbf{e}_{\nu_{i-1}},\Big(1,\tfrac{W^0_{S(i,t)}-W^0_{S(i-1,t)}}{S(i,t)-S(i-1,t)}\Big)\right\rangle\right\|_{L^p(\mathbb{P};\mathbb{R})}\\
&=\left(\int_t^T\mathbb{E}\Big[\big\|(F(0))(s,x+W^0_s-W^0_t)\big\|^{\frac{2p}{p-2}}\Big]\mathbb{P}\Big(S(j+1,t)\in ds\Big)\right)^{\frac{p-2}{2p}}\\
&\quad\cdot\left\|\prod_{i=1}^{j+1}\tfrac{1}{\varrho(S(i-1,t),S(i,t))}\left\langle \mathbf{e}_{\nu_{i-1}},\big(1,\tfrac{W^0_{S(i,t)}-W^0_{S(i-1,t)}}{S(i,t)-S(i-1,t)}\big)\right\rangle\right\|_{L^p(\mathbb{P};\mathbb{R})}\\
&\leq \sup_{s\in[t,T)}\big\|(F(0))(s,x+W^0_s-W^0_t)\big\|_{L^{\frac{2p}{p-2}}(\mathbb{P};\mathbb{R})}\left[e\left(\tfrac{pj}{2}+1\right)\right]^{\frac{1}{8}}C_1^{\frac{j+1}{2}}\left[\tfrac{1}{\Gamma(1+j+\frac{2}{p})}\right]^{\beta}
\end{aligned}\tag{72}$$

and, analogously,

$$\begin{aligned}
&\left\|\mathbf{u}_{\nu_{j+1}}(S(j+1,t),x+W^0_{S(j+1,t)}-W^0_t)\prod_{i=1}^{j+1}\tfrac{1}{\varrho(S(i-1,t),S(i,t))}\left\langle \mathbf{e}_{\nu_{i-1}},\big(1,\tfrac{W^0_{S(i,t)}-W^0_{S(i-1,t)}}{S(i,t)-S(i-1,t)}\big)\right\rangle\right\|_{L^2(\mathbb{P};\mathbb{R})}\\
&\leq \sup_{s\in[t,T)}\big\|\mathbf{u}_{\nu_{j+1}}(s,x+W^0_s-W^0_t)\big\|_{L^{\frac{2p}{p-2}}(\mathbb{P};\mathbb{R})}\left[e\left(\tfrac{pj}{2}+1\right)\right]^{\frac{1}{8}}C_1^{\frac{j+1}{2}}\left[\tfrac{1}{\Gamma(1+j+\frac{2}{p})}\right]^{\beta}.
\end{aligned}\tag{73}$$

Next we apply Lemma 3.4. To this end note that $\sup_{s\in[t,T)}\big\|\big|(F(0))(s,x+W^0_s-W^0_t)\big|\big\|_{L^{\frac{2p}{p-2}}(\mathbb{P};\mathbb{R})} < \infty$ and $\frac{2p}{p-2} \geq 2$ ensure that for all $r \in [1,2)$ it holds that

$$\sup_{s\in[t,T)}\mathbb{E}\Big[\big|\big(F(0)\big)(s,x+W^0_s-W^0_t)\big|^r\Big] < \infty. \tag{74}$$

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

$$\int_0^1 \frac{1}{s^{\frac{r}{2}}\left[\rho(s)\right]^{r-1}}\,ds = \frac{1}{(1-\alpha)^{r-1}}\int_0^1 s^{-(r(\frac{1}{2}-\alpha)+\alpha)}\,ds < \infty. \tag{75}$$

Combing Lemma 3.4, (70), (72), and (73) proves that

$$\begin{aligned}
&\left\|\mathbf{U}^{0,\nu_0}_{n,M}(t,x)-\mathbf{u}_{\nu_0}(t,x)\right\|_{L^2(\mathbb{P};\mathbb{R})}\\
&\leq \sqrt{\max\{T-t,3\}}\|K\|_1\sum_{j=0}^{n-1}\sum_{\substack{\nu_1,\nu_2,\ldots,\nu_{j+1}\in\\ \{1,2,\ldots,d+1\}}}\tbinom{n-1}{j}\frac{\mathbb{1}_{\{1\}}(\nu_{j+1})\left[e\left(\frac{pj}{2}+1\right)\right]^{\frac{1}{8}}2^jC_1^{\frac{j}{2}}\left[\prod_{i=1}^{j}L_{\nu_i}\right]}{\sqrt{M^{n-j}}}\left[\frac{1}{\Gamma(j+1)}\right]^{\beta}\\
&+\sup_{s\in[t,T)}\left\|(F(0))(s,x+W^0_s-W^0_t)\right\|_{L^{\frac{2p}{p-2}}(\mathbb{P};\mathbb{R})}\sum_{j=0}^{n-1}\sum_{\substack{\nu_1,\nu_2,\ldots,\nu_{j+1}\in\\ \{1,2,\ldots,d+1\}}}\tbinom{n-1}{j}\frac{\mathbb{1}_{\{1\}}(\nu_{j+1})\left[e\left(\frac{pj}{2}+1\right)\right]^{\frac{1}{8}}2^jC_1^{\frac{j+1}{2}}\left[\prod_{i=1}^{j}L_{\nu_i}\right]}{\sqrt{M^{n-j}}\left(\Gamma\left(1+j+\frac{2}{p}\right)\right)^{\beta}}\\
&+\sum_{j=0}^{n-1}\sum_{\substack{\nu_1,\nu_2,\ldots,\nu_{j+1}\in\\ \{1,2,\ldots,d+1\}}}\tbinom{n-1}{j}\sup_{s\in[t,T)}\left\|\mathbf{u}_{\nu_{j+1}}(s,x+W^0_s-W^0_t)\right\|_{L^{\frac{2p}{p-2}}(\mathbb{P};\mathbb{R})}\frac{\left[e\left(\frac{pj}{2}+1\right)\right]^{\frac{1}{8}}2^jC_1^{\frac{j+1}{2}}\left[\prod_{i=1}^{j+1}L_{\nu_i}\right]}{\sqrt{M^{n-j-1}}\left(\Gamma\left(1+j+\frac{2}{p}\right)\right)^{\beta}}.
\end{aligned}\tag{76}$$

Observe that for all $j\in\mathbb{N}_0$ it holds that

$$\left[\sum_{\substack{\nu_1,\nu_2,\ldots,\nu_{j+1}\in\\ \{1,2,\ldots,d+1\}}}\mathbb{1}_{\{1\}}(\nu_{j+1})\prod_{i=1}^{j}L_{\nu_i}\right]=\|L\|_1^j.\tag{77}$$

Combining this with (76) proves that

$$\begin{aligned}
&\left\|\mathbf{U}^{0,\nu_0}_{n,M}(t,x)-\mathbf{u}_{\nu_0}(t,x)\right\|_{L^2(\mathbb{P};\mathbb{R})}\\
&\leq \sqrt{\max\{T-t,3\}}\|K\|_1\sum_{j=0}^{n-1}\frac{\left[e\left(\frac{pj}{2}+1\right)\right]^{\frac{1}{8}}2^jC_1^{\frac{j}{2}}\|L\|_1^j}{\sqrt{M^{n-j}}}\tbinom{n-1}{j}\left[\frac{1}{\Gamma(j+1)}\right]^{\beta}\\
&\quad+\sup_{s\in[t,T)}\left\|(F(0))(s,x+W^0_s-W^0_t)\right\|_{L^{\frac{2p}{p-2}}(\mathbb{P};\mathbb{R})}\sum_{j=0}^{n-1}\frac{\left[e\left(\frac{pj}{2}+1\right)\right]^{\frac{1}{8}}2^j\|L\|_1^jC_1^{\frac{j+1}{2}}}{\sqrt{M^{n-j}}}\tbinom{n-1}{j}\left[\frac{1}{\Gamma\left(1+j+\frac{2}{p}\right)}\right]^{\beta}\\
&\quad+\sup_{s\in[t,T),i\in\{1,2,\ldots,d+1\}}\left\|\mathbf{u}_i(s,x+W^0_s-W^0_t)\right\|_{L^{\frac{2p}{p-2}}(\mathbb{P};\mathbb{R})}\sum_{j=0}^{n-1}\frac{\left[e\left(\frac{pj}{2}+1\right)\right]^{\frac{1}{8}}2^j\|L\|_1^{j+1}C_1^{\frac{j+1}{2}}}{\sqrt{M^{n-j-1}}}\tbinom{n-1}{j}\left[\frac{1}{\Gamma\left(1+j+\frac{2}{p}\right)}\right]^{\beta}
\end{aligned}\tag{78}$$

This and the facts that $2\sqrt{C_1}\|L\|_1\leq 2\sqrt{C_1}\max\{1,\|L\|_1\}\leq C$, $p\geq 2$, and $C\geq 1$ imply that

$$\begin{aligned}
&\left\|\mathbf{U}^{0,\nu_0}_{n,M}(t,x)-\mathbf{u}_{\nu_0}(t,x)\right\|_{L^2(\mathbb{P};\mathbb{R})}\\
&\leq \sqrt{\max\{T-t,3\}}\|K\|_1\sum_{j=0}^{n-1}\frac{\left[e\left(\frac{pj}{2}+1\right)\right]^{\frac{1}{8}}C^j}{\sqrt{M^{n-j}}}\tbinom{n-1}{j}\left[\frac{1}{\Gamma(j+1)}\right]^{\beta}\\
&\quad+\frac{1}{2}\sup_{s\in[t,T)}\left\|(F(0))(s,x+W^0_s-W^0_t)\right\|_{L^{\frac{2p}{p-2}}(\mathbb{P};\mathbb{R})}\sum_{j=0}^{n-1}\frac{\left[e\left(\frac{pj}{2}+1\right)\right]^{\frac{1}{8}}C^{j+1}}{\sqrt{M^{n-j}}}\tbinom{n-1}{j}\left[\frac{1}{\Gamma(j+1)}\right]^{\beta}\\
&\quad+\frac{1}{2}\sup_{s\in[t,T),i\in\{1,2,\ldots,d+1\}}\left\|\mathbf{u}_i(s,x+W^0_s-W^0_t)\right\|_{L^{\frac{2p}{p-2}}(\mathbb{P};\mathbb{R})}\sum_{j=0}^{n-1}\frac{\left[e\left(\frac{pj}{2}+1\right)\right]^{\frac{1}{8}}C^{j+1}}{\sqrt{M^{n-j-1}}}\tbinom{n-1}{j}\left[\frac{1}{\Gamma(j+1)}\right]^{\beta}\\
&\leq\frac{\left[e\left(\frac{pn}{2}+1\right)\right]^{\frac{1}{8}}C^{n-1}}{\sqrt{M^{n-1}}}\left[\sum_{j=0}^{n-1}\tbinom{n-1}{j}\frac{(\sqrt{M})^j}{\Gamma(j+1)^{\beta}}\right]\left[\frac{\sqrt{\max\{T-t,3\}}\|K\|_1}{\sqrt{M}}+\frac{C\sup_{s\in[t,T)}\left\|(F(0))(s,x+W^0_s-W^0_t)\right\|_{L^{\frac{2p}{p-2}}(\mathbb{P};\mathbb{R})}}{2\sqrt{M}}\right.\\
&\qquad\left.+\frac{C\sup_{s\in[t,T),i\in\{1,2,\ldots,d+1\}}\left\|\mathbf{u}_i(s,x+W^0_s-W^0_t)\right\|_{L^{\frac{2p}{p-2}}(\mathbb{P};\mathbb{R})}}{2}\right].
\end{aligned}\tag{79}$$

Next note that for all $j\in\mathbb{N}_0$ it holds that

$$\frac{(\sqrt{M})^j}{\Gamma(j+1)^{\beta}}=\left(\frac{(M^{\frac{1}{2\beta}})^j}{\Gamma(j+1)}\right)^{\beta}\leq\left(\sum_{k=0}^{\infty}\frac{(M^{\frac{1}{2\beta}})^k}{\Gamma(k+1)}\right)^{\beta}=\exp\left(\beta M^{\frac{1}{2\beta}}\right)\tag{80}$$

and that

$$\sum_{j=0}^{n-1} \binom{n-1}{j} = 2^{n-1}. \tag{81}$$

Combining (78), (80), and (81) shows that

$$\begin{aligned} &\left\| \mathbf{U}_{n,M}^{0,\nu_0}(t,x) - \mathbf{u}_{\nu_0}(t,x) \right\|_{L^2(\mathbb{P};\mathbb{R})} \leq \frac{\left[e\left(\frac{pn}{2}+1\right)\right]^{\frac{1}{8}}(2C)^{n-1}\exp\left(\beta M^{\frac{1}{2\beta}}\right)}{\sqrt{M^{n-1}}} \cdot \left[\frac{\sqrt{\max\{T-t,3\}}\|K\|_1}{\sqrt{M}}\right. \\ &+ \frac{C\sup_{s\in[t,T)}\left\|(F(0))(s,x+W_s^0-W_t^0)\right\|_{L^{\frac{2p}{p-2}}(\mathbb{P};\mathbb{R})}}{2\sqrt{M}} + \left.\frac{C\sup_{s\in[t,T),i\in\{1,2,\ldots,d+1\}}\left\|\mathbf{u}_i(s,x+W_s^0-W_t^0)\right\|_{L^{\frac{2p}{p-2}}(\mathbb{P};\mathbb{R})}}{2}\right]. \end{aligned} \tag{82}$$

This completes the proof of Proposition 3.5. ☐

4 Regularity analysis for solutions of certain differential equations

The error analysis in Subsection 3.3 above provides upper bounds for the approximation errors of the MLP approximations in (34). 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 (65) in Proposition 3.5 in Subsection 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 4.2 below (see, e.g., El Karoui et al. [19, Proposition 2.1 and Equation (2.12)] for results related to Lemma 4.2 below).

4.1 Regularity analysis for solutions of backward stochastic differential equations (BSDEs)

Lemma 4.1. *Let $T \in (0,\infty)$, $t \in [0,T]$, $d \in \mathbb{N}$, $L_1, L_2, \ldots, L_{d+1} \in [0,\infty)$, let $\|\cdot\|_2 \colon \mathbb{R}^d \to [0,\infty)$ be the d-dimensional Euclidean norm, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space with a normal filtration $(\mathbb{F}_t)_{t\in[t,T]}$, let $f, \tilde{f} \colon [t,T] \times \mathbb{R} \times \mathbb{R}^d \times \Omega \to \mathbb{R}$ be functions satisfying that for all $s \in [t,T]$ the function $[t,s] \times \mathbb{R} \times \mathbb{R}^d \times \Omega \ni (u,y,z,\omega) \mapsto (f(u,y,z,\omega), \tilde{f}(u,y,z,\omega)) \in \mathbb{R}^2$ is $(\mathcal{B}([t,s]) \otimes \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R}^d) \otimes \mathbb{F}_s)/\mathcal{B}(\mathbb{R}^2)$ is measurable, assume that for all $s \in [t,T]$, $y, \tilde{y} \in \mathbb{R}$, $z, \tilde{z} \in \mathbb{R}^d$ it holds $\mathbb{P}$-a.s. that*

$$|f(s,y,z) - f(s,\tilde{y},\tilde{z})| \leq L_1|y-\tilde{y}| + \sum_{j=1}^{d} L_{j+1}|z_j - \tilde{z}_j|, \tag{83}$$

let $Y, \tilde{Y} \colon [t,T] \times \Omega \to \mathbb{R}$, $W \colon [t,T] \times \Omega \to \mathbb{R}^d$ be $(\mathbb{F}_s)_{s\in[t,T]}$-adapted processes with continuous sample paths, assume that $(W_{s+t} - W_t)_{s\in[0,T-t]}$ is a standard Brownian motion, let $Z, \tilde{Z} \colon [t,T] \times \Omega \to \mathbb{R}^d$ be $(\mathbb{F}_s)_{s\in[t,T]}$-adapted $(\mathcal{B}([t,T]) \otimes \mathcal{F})/\mathcal{B}(\mathbb{R}^d)$-measurable processes, assume that it holds $\mathbb{P}$-a.s. that $\int_t^T |f(s,Y_s,Z_s)|\,ds < \infty$, $\int_t^T |\tilde{f}(s,\tilde{Y}_s,\tilde{Z}_s)|\,ds < \infty$, $\int_t^T \|Z_s\|_2^2\,ds < \infty$, and $\int_t^T \|\tilde{Z}_s\|_2^2\,ds < \infty$, assume that for all $s \in [t,T]$ it holds $\mathbb{P}$-a.s. that

$$\begin{aligned} Y_s &= Y_T + \int_s^T f(u,Y_u,Z_u)\,du - \int_s^T \left(Z_u\right)^T dW_u, \\ \tilde{Y}_s &= \tilde{Y}_T + \int_s^T \tilde{f}(u,\tilde{Y}_u,\tilde{Z}_u)\,du - \int_s^T \left(\tilde{Z}_u\right)^T dW_u, \end{aligned} \tag{84}$$

and that $\mathbb{E}\Big[\sup_{s\in[t,T]} Y_s^2\Big] < \infty$ and $\mathbb{E}\Big[\sup_{s\in[t,T]} \tilde{Y}_s^2\Big] < \infty$. Then it holds $\mathbb{P}$-a.s. that

$$\left|Y_t - \tilde{Y}_t\right| \leq e^{L_1(T-t)}\left(\|Y_T - \tilde{Y}_T\|_{L^\infty(\mathbb{P};\mathbb{R})} + (T-t)\sup_{s\in[t,T],y\in\mathbb{R},z\in\mathbb{R}^d}\|f(s,y,z) - \tilde{f}(s,y,z)\|_{L^\infty(\mathbb{P};\mathbb{R})}\right). \tag{85}$$

Proof of Lemma 4.1. Throughout the proof let $\langle\cdot,\cdot\rangle \colon \mathbb{R}^d \times \mathbb{R}^d \to [0,\infty)$ the function that satisfies for all $v = (v_1,\ldots,v_d)$, $w = (w_1,\ldots,w_d) \in \mathbb{R}^d$ that $\langle v,w\rangle = \sum_{i=1}^d v_i w_i$. Without loss of generality we suppose that $\sup_{s\in[t,T],y\in\mathbb{R},z\in\mathbb{R}^d}\|f(s,y,z) - \tilde{f}(s,y,z)\|_{L^\infty(\mathbb{P};\mathbb{R})} < \infty$. Throughout this proof let $\|\cdot\| \colon \mathbb{R}^d \to [0,\infty)$, be the Euclidean norm, let $\langle\cdot,\cdot\rangle \colon \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be the Euclidean scalar product, let $A \colon [t,T]\times\Omega \to \mathbb{R}$, $B \colon [t,T]\times\Omega \to \mathbb{R}^d$ be the functions which satisfy for all $s \in [t,T]$, $j \in \{1,2,\ldots,d\}$ that

$$A_s = \begin{cases} \frac{f(s,Y_s,\tilde{Z}_s)-f(s,\tilde{Y}_s,\tilde{Z}_s)}{Y_s-\tilde{Y}_s} & \text{if } Y_s \neq \tilde{Y}_s \\ 0 & \text{else} \end{cases}, \tag{86}$$

$$B_s(j) = \begin{cases} \frac{f(s,Y_s,Z_s(1),\ldots,Z_s(j),\tilde{Z}_s(j+1)\ldots,\tilde{Z}_s(d))-f(s,Y_s,Z_s(1),\ldots,Z_s(j-1),\tilde{Z}_s(j)\ldots,\tilde{Z}_s(d))}{Z_s(j)-\tilde{Z}_s(j)} & \text{if } Z_s(j) \neq \tilde{Z}_s(j) \\ 0 & \text{else} \end{cases}, \tag{87}$$

and let $\Gamma\colon [t,T]\times\Omega\to\mathbb{R}$ be the function that satisfies for all $s\in[t,T]$ that $\Gamma_s = e^{\int_t^s A_r - \frac{\|B_r\|^2}{2}\,dr + \int_t^s \langle B_r, dW_r\rangle}$. Itô's formula implies that for all $s\in[t,T]$ it holds $\mathbb{P}$-a.s. that

$$\Gamma_s = 1 + \int_t^s \Gamma_r A_r\,dr + \int_t^s \Gamma_r \langle B_r, dW_r\rangle. \tag{88}$$

Then Itô's formula, (84), and (88) yield that for all $u\in[t,T)$ it holds $\mathbb{P}$-a.s. that

$$\Gamma_u Y_u - Y_t = \int_t^u \Gamma_s\Big(-f(s,Y_s,Z_s) + A_sY_s + \langle B_s, Z_s\rangle\Big)\,ds + \int_t^u \Gamma_s Y_s\langle B_s, dW_s\rangle + \int_t^u \Gamma_s\langle Z_s, dW_s\rangle \tag{89}$$

and that

$$\Gamma_u \tilde{Y}_u - \tilde{Y}_t = \int_t^u \Gamma_s\Big(-\tilde{f}(s,\tilde{Y}_s,\tilde{Z}_s) + A_s\tilde{Y}_s + \langle B_s, \tilde{Z}_s\rangle\Big)\,ds + \int_t^u \Gamma_s \tilde{Y}_s\langle B_s, dW_s\rangle + \int_t^u \Gamma_s\langle \tilde{Z}_s, dW_s\rangle. \tag{90}$$

Next (86), (87), and a telescoping sum imply for all $s\in[t,T]$ that

$$\begin{aligned}
&A_s(Y_s-\tilde{Y}_s) + \langle B_s, Z_s-\tilde{Z}_s\rangle = A_s(Y_s-\tilde{Y}_s) + \sum_{j=1}^d B_s(j)\big(Z_s(j)-\tilde{Z}_s(j)\big)\\
&= f(s,Y_s,\tilde{Z}_s) - f(s,\tilde{Y}_s,\tilde{Z}_s) + \sum_{j=1}^d \Big(f(s,Y_s,Z_s(1),\ldots,Z_s(j),\tilde{Z}_s(j+1),\ldots,\tilde{Z}_s(d))\\
&\qquad - f(s,Y_s,Z_s(1),\ldots,Z_s(j-1),\tilde{Z}_s(j),\ldots,\tilde{Z}_s(d))\Big)\\
&= f(s,Y_s,\tilde{Z}_s) - f(s,\tilde{Y}_s,\tilde{Z}_s) + f(s,Y_s,Z_s) - f(s,Y_s,\tilde{Z}_s)\\
&= f(s,Y_s,Z_s) - f(s,\tilde{Y}_s,\tilde{Z}_s).
\end{aligned} \tag{91}$$

This, (89), and (90) imply that for all $u\in[t,T]$ it holds $\mathbb{P}$-a.s. that

$$\begin{aligned}
&Y_t-\tilde{Y}_t - \Gamma_u(Y_u-\tilde{Y}_u) + \int_t^u \Gamma_s(Y_s-\tilde{Y}_s)\langle B_s, dW_s\rangle + \int_t^u \Gamma_s\langle Z_s-\tilde{Z}_s, dW_s\rangle\\
&= \int_t^u \Gamma_s\Big(f(s,Y_s,Z_s) - \tilde{f}(s,\tilde{Y}_s,\tilde{Z}_s) - \Big(A_s(Y_s-\tilde{Y}_s) + \langle B_s, Z_s-\tilde{Z}_s\rangle\Big)\Big)\,ds\\
&= \int_t^u \Gamma_s\Big(f(s,\tilde{Y}_s,\tilde{Z}_s) - \tilde{f}(s,\tilde{Y}_s,\tilde{Z}_s)\Big)\,ds.
\end{aligned} \tag{92}$$

For every $n\in\mathbb{N}$ let $\tau_n\colon\Omega\to[t,T)$ be the stopping time that satisfies

$$\tau_n = \inf\left(\left\{r\in[t,T]\colon \left|\int_t^r \Gamma_s(Y_s-\tilde{Y}_s)\langle B_s, dW_s\rangle\right| + \left|\int_t^r \Gamma_s\langle Z_s-\tilde{Z}_s, dW_s\rangle\right| \geq n\right\}\cup\{T-\tfrac{T}{n}\}\right). \tag{93}$$

Taking conditional expectations in (92) implies for all $n\in\mathbb{N}$ that $\mathbb{P}$-a.s. it holds that

$$Y_t-\tilde{Y}_t = \mathbb{E}\left[\Gamma_{\tau_n}(Y_{\tau_n}-\tilde{Y}_{\tau_n}) + \int_t^{\tau_n}\Gamma_s\Big(f(s,\tilde{Y}_s,\tilde{Z}_s) - \tilde{f}(s,\tilde{Y}_s,\tilde{Z}_s)\Big)\,ds\Big|\mathbb{F}_t\right]. \tag{94}$$

Next note that assumption (83) ensures that $\sup_{s\in[t,T]}\|B_s\|^2 \leq \sum_{j=2}^{d+1}(L_j)^2$. This and an exponential martingale argument imply that

$$\begin{aligned}
\mathbb{E}\left[\left(e^{-\int_t^T \frac{\|B_s\|^2}{2}\,ds + \int_t^T\langle B_s, dW_s\rangle}\right)^2\right] &= \mathbb{E}\left[e^{\int_t^T \|B_s\|^2\,ds} e^{-\int_t^T \frac{\|2B_s\|^2}{2}\,ds + \int_t^T\langle 2B_s, dW_s\rangle}\right]\\
&\leq e^{(T-t)\sum_{j=2}^{d+1}(L_j)^2}\,\mathbb{E}\left[e^{-\int_t^T \frac{\|2B_s\|^2}{2}\,ds + \int_t^T\langle 2B_s, dW_s\rangle}\right]\\
&= e^{(T-t)\sum_{j=2}^{d+1}(L_j)^2}.
\end{aligned} \tag{95}$$

Note that by assumption (83) it holds $\mathbb{P}$-a.s. that $\sup_{s\in[t,T]}|A_s|\leq L_1$. This, Doob's martingale inequality, and (95), show that

$$\begin{aligned}
\mathbb{E}\left[\sup_{s\in[t,T]}\Gamma_s^2\right] &= \mathbb{E}\left[\sup_{s\in[t,T]} e^{\int_t^s 2A_r\,dr}\left(e^{-\int_t^s \frac{\|B_r\|^2}{2}\,dr + \int_t^s\langle B_r, dW_r\rangle}\right)^2\right]\\
&\leq e^{2L_1(T-t)}\,\mathbb{E}\left[\sup_{s\in[t,T]}\left(e^{-\int_t^s \frac{\|B_r\|^2}{2}\,dr + \int_t^s\langle B_r, dW_r\rangle}\right)^2\right]\\
&\leq 4e^{2L_1(T-t)}\,\mathbb{E}\left[\left(e^{-\int_t^T \frac{\|B_r\|^2}{2}\,dr + \int_t^T\langle B_r, dW_r\rangle}\right)^2\right] \leq 4e^{(T-t)\left(2L_1+\sum_{j=2}^{d+1}(L_j)^2\right)}.
\end{aligned} \tag{96}$$

This together with the Cauchy-Schwarz inequality, and the assumptions that $\mathbb{E}\Big[\sup_{s\in[t,T]} Y_s^2\Big] < \infty$ and $\mathbb{E}\Big[\sup_{s\in[t,T]} \tilde{Y}_s^2\Big] < \infty$ prove that

$$\mathbb{E}\left[\sup_{s\in[t,T]} |\Gamma_s(Y_s - \tilde{Y}_s)|\right] \leq \sqrt{\mathbb{E}\left[\sup_{s\in[t,T]} \Gamma_s^2\right] \mathbb{E}\left[\sup_{s\in[t,T]} (Y_s - \tilde{Y}_s)^2\right]} < \infty. \tag{97}$$

Moreover, it holds that

$$\mathbb{E}\left[\int_t^T \Gamma_s \Big| f(s, \tilde{Y}_s, \tilde{Z}_s) - \tilde{f}(s, \tilde{Y}_s, \tilde{Z}_s)\Big| \, ds\right] \leq \mathbb{E}\left[\int_t^T \Gamma_s \, ds\right] \left[\sup_{s\in[t,T], y\in\mathbb{R}, z\in\mathbb{R}^d} \|f(s,y,z) - \tilde{f}(s,y,z)\|_{L^\infty(\mathbb{P};\mathbb{R})}\right] < \infty. \tag{98}$$

This, (97), Lebesgue's dominated convergence theorem, (94), continuity of Y, $\tilde{Y}$, Γ, and the fact that it holds $\mathbb{P}$-a.s. that $\lim_{n\to\infty} \tau_n = T$ ensure that $\mathbb{P}$-a.s. it holds that

$$\begin{aligned} Y_t - \tilde{Y}_t &= \lim_{n\to\infty} \mathbb{E}\left[\Gamma_{\tau_n}(Y_{\tau_n} - \tilde{Y}_{\tau_n}) + \int_t^{\tau_n} \Gamma_s \Big(f(s, \tilde{Y}_s, \tilde{Z}_s) - \tilde{f}(s, \tilde{Y}_s, \tilde{Z}_s)\Big) \, ds \Big| \mathbb{F}_t\right] \\ &= \mathbb{E}\left[\Gamma_T(Y_T - \tilde{Y}_T) + \int_t^T \Gamma_s \Big(f(s, \tilde{Y}_s, \tilde{Z}_s) - \tilde{f}(s, \tilde{Y}_s, \tilde{Z}_s)\Big) \, ds \Big| \mathbb{F}_t\right] \end{aligned} \tag{99}$$

This, the triangle inequality, the fact that the process $[t,T] \times \Omega \ni (s,\omega) \mapsto e^{-\int_t^s \frac{\|B_r\|^2}{2} dr + \int_t^s \langle B_r, dW_r\rangle} \in (0,\infty)$ is a martingale, and the fact $\sup_{s\in[t,T]} |A_s| \leq L_1$ yield that $\mathbb{P}$-a.s. it holds that

$$\begin{aligned} \big|Y_t - \tilde{Y}_t\big| &\leq \mathbb{E}\left[\Gamma_T |Y_T - \tilde{Y}_T| + \int_t^T \Gamma_s \Big| f(s, \tilde{Y}_s, \tilde{Z}_s) - \tilde{f}(s, \tilde{Y}_s, \tilde{Z}_s)\Big| \, ds \Big| \mathbb{F}_t\right] \\ &\leq \mathbb{E}[\Gamma_T | \mathbb{F}_t] \, \|Y_T - \tilde{Y}_T\|_{L^\infty(\mathbb{P};\mathbb{R})} + \mathbb{E}\left[\int_t^T \Gamma_s \, ds \Big| \mathbb{F}_t\right] \left[\sup_{s\in[t,T], y\in\mathbb{R}, z\in\mathbb{R}^d} \|f(s,y,z) - \tilde{f}(s,y,z)\|_{L^\infty(\mathbb{P};\mathbb{R})}\right] \\ &\leq e^{L_1(T-t)} \left(\|Y_T - \tilde{Y}_T\|_{L^\infty(\mathbb{P};\mathbb{R})} + (T-t) \sup_{s\in[t,T], y\in\mathbb{R}, z\in\mathbb{R}^d} \|f(s,y,z) - \tilde{f}(s,y,z)\|_{L^\infty(\mathbb{P};\mathbb{R})} \right). \end{aligned} \tag{100}$$

This proves (85). The proof of Lemma 4.1 is thus completed. □

4.2 Regularity analysis for solutions of partial differential equations (PDEs)

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

$$|f(t,x,y,z) - f(t,\mathfrak{x},\mathfrak{y},\mathfrak{z})| \leq L_0|y - \mathfrak{y}| + \textstyle\sum_{j=1}^d (L_j|z_j - \mathfrak{z}_j| + \mathfrak{L}_j|x_j - \mathfrak{x}_j|), \tag{101}$$

$$|g(x) - g(\mathfrak{x})| \leq \textstyle\sum_{i=1}^d K_i|x_i - \mathfrak{x}_i|, \qquad |u(t,x)| \leq \eta\big[1 + \sum_{i=1}^d |x_i|\big]^\eta, \qquad u(T,x) = g(x), \tag{102}$$

$$\text{and} \qquad \big(\tfrac{\partial}{\partial t} u\big)(t,x) + \tfrac{1}{2}(\Delta_x u)(t,x) + f\big(t,x,u(t,x),(\nabla_x u)(t,x)\big) = 0, \tag{103}$$

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

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

$$\begin{aligned} &\mathbb{E}\Big[\big\|g(x + W_{T-s})\big(1, \tfrac{W_{T-s}}{T-s}\big)\big\|\Big] \\ &\quad + \mathbb{E}\left[\int_s^T \big\|\big[f\big(t, x + W_{t-s}, u(t, x + W_{t-s}), (\nabla_x u)(t, x + W_{t-s})\big)\big]\big(1, \tfrac{W_{t-s}}{t-s}\big)\big\| \, dt\right] < \infty, \end{aligned} \tag{104}$$

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

$$\begin{aligned} (u(s,x), (\nabla_x u)(s,x)) &= \mathbb{E}\Big[g(x + W_{T-s})\big(1, \tfrac{W_{T-s}}{T-s}\big)\Big] \\ &\quad + \mathbb{E}\left[\int_s^T \big[f\big(t, x + W_{t-s}, u(t, x + W_{t-s}), (\nabla_x u)(t, x + W_{t-s})\big)\big] \big(1, \tfrac{W_{t-s}}{t-s}\big) \, dt\right], \end{aligned} \tag{105}$$

(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*

$$|u(t,x) - u(t,\mathfrak{x})| \le e^{L_0(T-t)}\Big(\textstyle\sum_{j=1}^d (K_j + (T-t)\mathfrak{L}_j)|x_j - \mathfrak{x}_j|\Big), \tag{106}$$

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

$$\big|(\tfrac{\partial}{\partial x_i}u)(t,x)\big| \le e^{L_0(T-t)}(K_i + (T-t)\mathfrak{L}_i), \tag{107}$$

and

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

$$\begin{aligned}&\sup_{s\in[0,T]}\sup_{t\in[s,T]} \left(\mathbb{E}[|u(t,x+W_t-W_s)|^p]\right)^{1/p}\\ &\le e^{L_0T}\left[\sup_{s\in[0,T]}(\mathbb{E}[|g(x+W_s)|^p])^{1/p} + T\sup_{s,t\in[0,T]}(\mathbb{E}[|f(t,x+W_s,0,0)|^p])^{1/p} + Te^{L_0T}\textstyle\sum_{j=1}^d L_j(K_j + T\mathfrak{L}_j)\right].\end{aligned} \tag{108}$$

Proof of Lemma 4.2. Throughout this proof let $\|\cdot\|_2 : \mathbb{R}^d \to [0,\infty)$ be the d-dimensional Euclidean norm. Item (i) and item (ii) follow from Lemma 4.2 in [41]. Next we prove item (iii). Throughout the proof of item (iii) 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 $Y, \mathfrak{Y} \colon [t,T]\times\Omega \to \mathbb{R}$, let $Z, \mathfrak{Z} \colon [t,T]\times\Omega\to\mathbb{R}^d$ be functions which satisfy for all $s \in [t,T]$ that

$$Y_s = u(s, x + W_s - W_t), \tag{109}$$
$$\mathfrak{Y}_s = u(s, \mathfrak{x} + W_s - W_t), \tag{110}$$
$$Z_s = (\nabla_x u)(s, x + W_s - W_t), \tag{111}$$
$$\mathfrak{Z}_s = (\nabla_x u)(s, \mathfrak{x} + W_s - W_t). \tag{112}$$

Then Itô's lemma yields that for all $s \in [t,T]$ it holds $\mathbb{P}$-a.s. that

$$Y_T = Y_s - \int_s^T f(r, x + W_r - W_t, Y_r, Z_r)\,dr + \int_s^T (Z_r)^T\,dW_r \tag{113}$$

and

$$\mathfrak{Y}_T = \mathfrak{Y}_s - \int_s^T f(r, \mathfrak{x} + W_r - W_t, \mathfrak{Y}_r, \mathfrak{Z}_r)\,dr + \int_s^T (\mathfrak{Z}_r)^T\,dW_r. \tag{114}$$

Next note that (102) implies that there exists $\lambda \in (\frac{1}{2},\infty)$ such that $\sup_{s\in[0,T],\xi\in\mathbb{R}^d} \frac{|u(s,\xi)|}{1+\|\xi\|_2^\lambda} < \infty$. For such a $\lambda \in (\frac{1}{2},\infty)$ Doob's inequality implies that

$$\begin{aligned}\left\|\sup_{s\in[t,T]}|Y_s|\right\|_{L^2(\mathbb{P};\mathbb{R})} &= \left\|\sup_{s\in[t,T]}|u(s,x+W_s-W_t)|\right\|_{L^2(\mathbb{P};\mathbb{R})}\\ &\le \left[\sup_{s\in[0,T],\xi\in\mathbb{R}^d}\tfrac{|u(s,\xi)|}{1+\|\xi\|_2^\lambda}\right]\left(1 + \left\|\sup_{s\in[t,T]}\|x+W_s-W_t\|_2^\lambda\right\|_{L^2(\mathbb{P};\mathbb{R})}\right)\\ &\le \left[\sup_{s\in[0,T],\xi\in\mathbb{R}^d}\tfrac{|u(s,\xi)|}{1+\|\xi\|_2^\lambda}\right]\left(1 + \left(\frac{2\lambda}{2\lambda-1}\right)^{1/\lambda}\big\|\|x+W_T-W_t\|_2^\lambda\big\|_{L^2(\mathbb{P};\mathbb{R})}\right) < \infty\end{aligned} \tag{115}$$

and likewise

$$\left\|\sup_{s\in[t,T]}|\mathfrak{Y}_s|\right\|_{L^2(\mathbb{P};\mathbb{R})} < \infty. \tag{116}$$

Moreover, (101) 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

$$|f(s, x + W_t - W_s, y, z) - f(s, x + W_t - W_s, \mathfrak{y}, \mathfrak{z})| \le L_0|y - \mathfrak{y}| + \sum_{j=1}^d L_j|z_j - \mathfrak{z}_j|. \tag{117}$$

and

$$|f(s, x + W_s - W_t, y, z) - f(s, \mathfrak{x} + W_s - W_t, y, z)| \le \sum_{j=1}^d \mathfrak{L}_j|x_j - \tilde{x}_j|. \tag{118}$$

This, (113), (114), (115), (116), and Lemma 4.1 prove that

$$\begin{aligned}
&|u(t,x)-u(t,\mathfrak{x})| = \big|Y_t - \mathfrak{Y}_t\big| \\
&\leq e^{L_0(T-t)}\Big(\|Y_T-\mathfrak{Y}_T\|_{L^\infty(\mathbb{P};\mathbb{R})} + (T-t)\sup_{s\in[t,T],y\in\mathbb{R},z\in\mathbb{R}^d}\|f(s,x+W_s-W_t,y,z)-f(s,\mathfrak{x}+W_s-W_t,y,z)\|_{L^\infty(\mathbb{P};\mathbb{R})}\Big) \\
&\leq e^{L_0(T-t)}\Big(\|g(x+W_T-W_t)-g(\mathfrak{x}+W_T-W_t)\|_{L^\infty(\mathbb{P};\mathbb{R})} + (T-t)\sum_{j=1}^d \mathfrak{L}_j|x_j-\mathfrak{x}_j|\Big) \\
&\leq e^{L_0(T-t)}\Big(\sum_{j=1}^d (K_j+(T-t)\mathfrak{L}_j)|x_j-\mathfrak{x}_j|\Big).
\end{aligned} \tag{119}$$

This proves item (iii). From item (iii) it then follows for all $t\in[0,T]$, $x\in\mathbb{R}^d$, $i\in\{1,2,\ldots,d\}$ that

$$\left|\big(\tfrac{\partial}{\partial x_i}u\big)(t,x)\right| = \left|\lim_{\mathbb{R}\setminus\{0\}\ni h\to 0}\tfrac{u(t,x+he_i)-u(t,x)}{h}\right| \leq e^{L_0(T-t)}(K_i+(T-t)\mathfrak{L}_i). \tag{120}$$

and this proves item (iv). Next we prove item (v). For the rest of the proof let $F\colon C([0,T)\times\mathbb{R}^d,\mathbb{R}^{1+d})\to C([0,T)\times\mathbb{R}^d,\mathbb{R})$, $\mathbf{u}\colon[0,T)\times\mathbb{R}^d\to\mathbb{R}^{d+1}$ be the functions which satisfies 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 $(F(\mathbf{v}))(t,x)=f(t,x,\mathbf{v}(t,x))$ and $\mathbf{u}(t,x)=(u(t,x),(\nabla_x u)(t,x))$. Item (ii), Tonelli's theorem, and the triangle inequality prove for all $s\in[0,T]$, $t\in[s,T]$, $x\in\mathbb{R}^d$, $p\in[1,\infty)$ that

$$\begin{aligned}
&\|u(t,x+W_t-W_s)\|_{L^p(\mathbb{P};\mathbb{R})} \\
&= \Big\|\int g(x+W_t-W_s+y)\mathbb{P}_{W_T-W_s}(dy) + \int_t^T\int (F(\mathbf{u}))(v,x+W_t-W_s+y)\,\mathbb{P}_{W_v-W_t}(dy)\,dv\Big\|_{L^p(\mathbb{P};\mathbb{R})} \\
&\leq \Big\|\int g(x+W_t-W_s+y)\mathbb{P}_{W_T-W_s}(dy)\Big\|_{L^p(\mathbb{P};\mathbb{R})} + \int_t^T\Big\|\int (F(\mathbf{u}))(v,x+W_t-W_s+y)\,\mathbb{P}_{W_v-W_t}(dy)\Big\|_{L^p(\mathbb{P};\mathbb{R})} dv
\end{aligned} \tag{121}$$

This, Jensen's inequality independence of Brownian increments yield for all $s\in[0,T]$, $t\in[s,T]$, $x\in\mathbb{R}^d$, $p\in[1,\infty)$ that

$$\begin{aligned}
\|u(t,x+W_t-W_s)\|_{L^p(\mathbb{P};\mathbb{R})} &\leq \Big(\int\int\Big|g(x+z+y)\Big|^p\mathbb{P}_{W_T-W_s}(dy)\,\mathbb{P}_{W_t-W_s}(dz)\Big)^{\frac{1}{p}} \\
&\quad+\int_t^T\Big(\int\int\Big|(F(\mathbf{u}))(v,x+z+y)\Big|^p\,\mathbb{P}_{W_T-W_s}(dy)\,\mathbb{P}_{W_t-W_s}(dz)\Big)^{\frac{1}{p}}\,dv. \\
&= \Big(\int\int\Big|g(x+z+y)\Big|^p\mathbb{P}_{(W_v-W_t,W_t-W_s)}(d(y,z))\Big)^{\frac{1}{p}} \\
&\quad+\int_t^T\Big(\int\int\Big|(F(\mathbf{u}))(v,x+z+y)\Big|^p\,\mathbb{P}_{(W_v-W_t,W_t-W_s)}(d(y,z))\Big)^{\frac{1}{p}}\,dv \\
&= \|g(x+W_T-W_s)\|_{L^p(\mathbb{P};\mathbb{R})} + \int_t^T\Big\|(F(\mathbf{u}))(v,x+W_v-W_s)\Big\|_{L^p(\mathbb{P};\mathbb{R})}dv
\end{aligned} \tag{122}$$

This, the triangle inequality, the global Lipschitz assumption 101 of f, and item (iv) show for all $s\in[0,T]$, $x\in\mathbb{R}^d$, $p\in[1,\infty)$ that

$$\begin{aligned}
&\sup_{t\in[s,T]}\|u(t,x+W_t-W_s)\|_{L^p(\mathbb{P};\mathbb{R})} \leq \|g(x+W_T-W_s)\|_{L^p(\mathbb{P};\mathbb{R})} + \int_s^T\Big\|(F(0))(v,x+W_v-W_s)\Big\|_{L^p(\mathbb{P};\mathbb{R})}dv \\
&\quad+L_0\int_s^T\Big\|u(v,x+W_v-W_s)\Big\|_{L^p(\mathbb{P};\mathbb{R})}dv + \sum_{j=1}^d L_j\int_s^T\Big\|\big(\tfrac{\partial}{\partial x_j}u\big)(v,x+W_v-W_s)\Big\|_{L^p(\mathbb{P};\mathbb{R})}dv \\
&\leq \|g(x+W_T-W_s)\|_{L^p(\mathbb{P};\mathbb{R})} + \int_s^T\Big\|(F(0))(v,x+W_v-W_s)\Big\|_{L^p(\mathbb{P};\mathbb{R})}dv \\
&\quad+L_0\int_s^T\sup_{t\in[v,T]}\Big\|u(t,x+W_t-W_s)\Big\|_{L^p(\mathbb{P};\mathbb{R})}dv + \sum_{j=1}^d e^{L_0T}TL_j(K_j+T\mathfrak{L}_\mathrm{j})
\end{aligned} \tag{123}$$

Note that there exists $\lambda\in(0,\infty)$ such that $\sup_{s\in[0,T],\xi\in\mathbb{R}^d}\frac{|u(s,\xi)|}{1+\|\xi\|_2^\lambda}<\infty$. For such a $\lambda\in(0,\infty)$ and for all

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

$$\sup_{s\in[0,T],t\in[s,T]} \|u(t,x+W_t-W_s)\|_{L^p(\mathbb{P};\mathbb{R})} \leq \left[\sup_{s\in[0,T],\xi\in\mathbb{R}^d} \tfrac{|u(s,\xi)|}{1+\|\xi\|_2^\lambda}\right]\left(1+\sup_{s\in[0,T],t\in[s,T]} \big\|\|x+W_t-W_s\|_2^\lambda\big\|_{L^p(\mathbb{P};\mathbb{R})}\right)<\infty. \tag{124}$$

This, (123), and Gronwall's inequality finally yield for all $x \in \mathbb{R}^d$, $p \in [1,\infty)$ that

$$\begin{aligned}&\sup_{s\in[0,T]}\sup_{t\in[s,T]} \|u(t,x+W_t-W_s)\|_{L^p(\mathbb{P};\mathbb{R})}\\ &\leq e^{L_0T}\sup_{s\in[0,T]}\|g(x+W_s)\|_{L^p(\mathbb{P};\mathbb{R})}+e^{L_0T}T\left(\sup_{s,t\in[0,T]}\big\|(F(0))(t,x+W_s)\big\|_{L^p(\mathbb{P};\mathbb{R})}+e^{L_0T}\sum_{j=1}^d L_j(K_j+T\mathfrak{L}_{\mathfrak{j}})\right).\end{aligned} \tag{125}$$

The proof of Lemma 4.2 is thus completed. □

5 Overall complexity analysis for MLP approximation methods

In this section we combine the findings of Sections 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 $\mathfrak{r}^\theta\colon \Omega \to (0,1)$, $\theta \in \Theta$, appearing in the MLP approximation methods in Corollary 5.1 (see (129) in Corollary 5.1), Theorem 5.2 (see (146) in Theorem 5.2), and Corollary 5.4 (see (159) 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 $\mathfrak{r}^\theta\colon \Omega \to (0,1)$, $\theta \in \Theta$, 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 $\mathbb{P}(\mathfrak{r}^\theta \leq b) = b^\alpha$ for some $\alpha \in (0,1)$. In particular, in contrast to previous MLP approximation methods studied in the scientific literature (see, e.g., [39, 40, 4]) it is crucial in this article to exclude the case where the random variables $\mathfrak{r}^\theta\colon \Omega \to (0,1)$, $\theta \in \Theta$, 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 $\mathbf{U}\colon \Omega \to \mathbb{R}^{d+1}$ in Lemma 5.3 coincides with a special case of the random fields in (146) in Theorem 5.2 (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 $\|\cdot\|_1 \colon (\cup_{n\in\mathbb{N}}\mathbb{R}^n) \to \mathbb{R}$ and $\|\cdot\|_\infty \colon (\cup_{n\in\mathbb{N}}\mathbb{R}^n) \to \mathbb{R}$ satisfy for all $n \in \mathbb{N}$, $x = (x_1,x_2,\ldots,x_n) \in \mathbb{R}^n$ that $\|x\|_1 = \sum_{i=1}^n |x_i|$ and $\|x\|_\infty = \max_{i\in\{1,2,\ldots,n\}} |x_i|$, let $T,\delta \in (0,\infty)$, $\varepsilon \in (0,1]$, $d \in \mathbb{N}$, $L = (L_0,L_1,\ldots,L_d) \in \mathbb{R}^{d+1}$, $K = (K_1,K_2,\ldots,K_d)$, $\mathfrak{L} = (\mathfrak{L}_1,\mathfrak{L}_2\ldots,\mathfrak{L}_d)$, $\xi \in \mathbb{R}^d$, $p \in (2,\infty)$, $\alpha \in (\frac{p-2}{2(p-1)},\frac{p}{2(p-1)})$, $\beta = \frac{\alpha}{2} - \frac{(1-\alpha)(p-2)}{2p} \in (0,\frac{\alpha}{2})$, $f \in C([0,T]\times\mathbb{R}^d\times\mathbb{R}\times\mathbb{R}^d,\mathbb{R})$, $g \in C(\mathbb{R}^d,\mathbb{R})$, 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 $t \in (0,T)$, $x = (x_1,x_2,\ldots,x_d)$, $\mathfrak{x} = (\mathfrak{x}_1,\mathfrak{x}_2,\ldots,\mathfrak{x}_d)$, $z = (z_1,z_2,\ldots,z_d)$, $\mathfrak{z} = (\mathfrak{z}_1,\mathfrak{z}_2,\ldots,\mathfrak{z}_d) \in \mathbb{R}^d$, $y,\mathfrak{y} \in \mathbb{R}$ that*

$$|f(t,x,y,z) - f(t,\mathfrak{x},\mathfrak{y},\mathfrak{z})| \leq L_0|y-\mathfrak{y}| + \textstyle\sum_{j=1}^d \big(L_j|z_j-\mathfrak{z}_j| + \mathfrak{L}_j|x_j-\mathfrak{x}_j|\big), \tag{126}$$

$$|g(x)-g(\mathfrak{x})| \leq \textstyle\sum_{i=1}^d K_i|x_i-\mathfrak{x}_i|, \qquad u(T,x) = g(x), \tag{127}$$

$$\text{and} \qquad \big(\tfrac{\partial}{\partial t}u\big)(t,x) + \tfrac{1}{2}(\Delta_x u)(t,x) + f\big(t,x,u(t,x),(\nabla_x u)(t,x)\big) = 0, \tag{128}$$

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

$$\begin{aligned}\mathbf{U}^\theta_{n,M}(t,x) &= (g(x),0) + \tfrac{1}{M^n}\sum_{i=1}^{M^n}\big(g(x+[T-t]^{1/2}Z^{(\theta,0,-i)})-g(x)\big)\big(1,[T-t]^{-1/2}Z^{(\theta,0,-i)}\big)\\ &+\sum_{l=0}^{n-1}\sum_{i=1}^{M^{n-l}}\tfrac{(T-t)(\mathfrak{r}^{(\theta,l,i)})^\alpha}{(1-\alpha)M^{n-l}}\big(1,[(T-t)\mathfrak{r}^{(\theta,l,i)}]^{-1/2}Z^{(\theta,l,i)}\big)\\ &\cdot\big[\big(F(\mathbf{U}^{(\theta,l,i)}_{l,M})-\mathbb{1}_{\mathbb{N}}(l)F(\mathbf{U}^{(\theta,-l,i)}_{l-1,M})\big)\big(t+(T-t)\mathfrak{r}^{(\theta,l,i)},x+[(T-t)\mathfrak{r}^{(\theta,l,i)}]^{1/2}Z^{(\theta,l,i)}\big)\big],\end{aligned} \tag{129}$$

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

$$\mathrm{RV}_{n,M}\leq dM^n+\sum_{l=0}^{n-1}\left[M^{(n-l)}(d+1+\mathrm{RV}_{l,M}+\mathbb{1}_{\mathbb{N}}(l)\,\mathrm{RV}_{l-1,M})\right], \tag{130}$$

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

$$C=\max\left\{\tfrac{1}{2},|\Gamma(\tfrac{p}{2})|^{\frac{1}{p}}(1-\alpha)^{\frac{1}{p}-1}\max\left\{T,\Gamma(\tfrac{p+1}{2})^{\frac{1}{p}}\pi^{-\frac{1}{2p}}\sqrt{2T}\right\}\max\{1,\|L\|_1\}\right\}. \tag{131}$$

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

$$\sup_{n\in\mathbb{N}\cap[N,\infty)}\left[\left\|\mathbf{U}^{0,0}_{n,\lfloor n^{2\beta}\rfloor}(0,\xi)-u(0,\xi)\right\|_{L^2(\mathbb{P};\mathbb{R})}+\max_{i\in\{1,2,\ldots,d\}}\left\|\mathbf{U}^{0,i}_{n,\lfloor n^{2\beta}\rfloor}(0,\xi)-(\tfrac{\partial}{\partial x_i}u)(0,\xi)\right\|_{L^2(\mathbb{P};\mathbb{R})}\right]\leq\varepsilon \tag{132}$$

and

$$\begin{aligned}\sum_{n=1}^{N}\mathrm{RV}_{n,\lfloor n^{2\beta}\rfloor}&\leq d\varepsilon^{-(2+\delta)}2^{3+\delta}\Bigg[1+\frac{\sqrt{\max\{T,3\}}\|K\|_1}{\sqrt{\lfloor(N-1)^{2\beta}\rfloor}}+Ce^{L_0T}\sup_{s\in[0,T]}\|g(\xi+\sqrt{s}Z^{(0)})\|_{L^{\frac{2p}{p-2}}(\mathbb{P};\mathbb{R})}\\&+Ce^{L_0T}(\|K\|_\infty+T\|\mathfrak{L}\|_\infty)+C\sup_{s,t\in[0,T)}\left\|(F(0))(t,\xi+\sqrt{s}Z^{(0)})\right\|_{L^{\frac{2p}{p-2}}(\mathbb{P};\mathbb{R})}\left(\frac{1}{\sqrt{\lfloor(N-1)^{2\beta}\rfloor}}+Te^{L_0T}\right)\\&+TCe^{2L_0T}\sum_{j=1}^{d}L_j(K_j+T\mathfrak{L}_j)\Bigg]^{2+\delta}\cdot\left[\sup_{n\in\mathbb{N}\cap[2,\infty)}\left(\frac{5^n\left[(n-1)^{2\beta}\left[e\left(\frac{p(n-1)}{2}+1\right)\right]^{\frac{1}{8}}(4Ce^{\beta})^{n-1}\right]^{2+\delta}}{(\lfloor(n-1)^{2\beta}\rfloor)^{\frac{\delta n}{2}}}\right)\right]<\infty.\end{aligned} \tag{133}$$

Proof of Corollary 5.1. Throughout the proof let $\varepsilon\in(0,1]$ and let $(\eta_{n,M})_{(n,M)\in\mathbb{N}^2}\subseteq\mathbb{R}$ satisfy for all $n,M\in\mathbb{N}$ that

$$\eta_{n,M}=\left\|\mathbf{U}^{0,0}_{n,M}(0,\xi)-u(0,\xi)\right\|_{L^2(\mathbb{P};\mathbb{R})}+\max_{i\in\{1,2,\ldots,d\}}\left\|\mathbf{U}^{0,i}_{n,M}(0,\xi)-(\tfrac{\partial}{\partial x_i}u)(0,\xi)\right\|_{L^2(\mathbb{P};\mathbb{R})}. \tag{134}$$

First note that it follows from Proposition 3.5, item (i) of Lemma 4.2, and the fact that the increments of a Brownian motion are normally distributed that for all $M,n\in\mathbb{N}$ it holds that

$$\begin{aligned}\eta_{n,M}&\leq 2\max\left\{\left\|\mathbf{U}^{0,0}_{n,M}(0,\xi)-u(0,\xi)\right\|_{L^2(\mathbb{P};\mathbb{R})},\max_{i\in\{1,2,\ldots,d\}}\left\|\mathbf{U}^{0,i}_{n,M}(0,\xi)-(\tfrac{\partial}{\partial x_i}u)(0,\xi)\right\|_{L^2(\mathbb{P};\mathbb{R})}\right\}\\&\leq\frac{2\left[e\left(\frac{pn}{2}+1\right)\right]^{\frac{1}{8}}(4C)^{n-1}\exp\left(\beta M^{\frac{1}{2\beta}}\right)}{\sqrt{M^{n-1}}}\Bigg[\frac{\sqrt{\max\{T,3\}}\|K\|_1}{\sqrt{M}}+\frac{C\sup_{s,t\in[0,T)}\left\|(F(0))(t,\xi+\sqrt{s}Z^{(0)})\right\|_{L^{\frac{2p}{p-2}}(\mathbb{P};\mathbb{R})}}{\sqrt{M}}\\&\quad+C\sup_{s,t\in[0,T)}\left[\max\left\{\left\|u(t,\xi+\sqrt{s}Z^{(0)})\right\|_{L^{\frac{2p}{p-2}}(\mathbb{P};\mathbb{R})},\max_{i\in\{1,2,\ldots,d\}}\left\|(\tfrac{\partial}{\partial x_i}u)(t,\xi+\sqrt{s}Z^{(0)})\right\|_{L^{\frac{2p}{p-2}}(\mathbb{P};\mathbb{R})}\right\}\right]\Bigg].\end{aligned} \tag{135}$$

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

$$\begin{aligned}\eta_{n,M}\leq&\frac{2\left[e\left(\frac{pn}{2}+1\right)\right]^{\frac{1}{8}}(4C)^{n}\exp\left(\beta M^{\frac{1}{2\beta}}\right)}{\sqrt{M^{n-1}}}\Bigg[\frac{\sqrt{\max\{T,3\}}\|K\|_1}{\sqrt{M}}+C\left(\frac{1}{\sqrt{M}}+Te^{L_0T}\right)\sup_{s,t\in[0,T)}\left\|(F(0))(t,\xi+\sqrt{s}Z^{(0)})\right\|_{L^{\frac{2p}{p-2}}(\mathbb{P};\mathbb{R})}\\&+Ce^{L_0T}(\|K\|_\infty+T\|\mathfrak{L}\|_\infty)+Ce^{L_0T}\sup_{s\in[0,T]}\|g(\xi+\sqrt{s}Z^{(0)})\|_{L^{\frac{2p}{p-2}}(\mathbb{P};\mathbb{R})}+TCe^{2L_0T}\sum_{j=1}^{d}L_j(K_j+T\mathfrak{L}_j)\Bigg].\end{aligned} \tag{136}$$

It follows from (126) and (127) that $\sup_{s,t\in[0,T)}\left\|(F(0))(t,\xi+\sqrt{s}Z^{(0)})\right\|_{L^{\frac{2p}{p-2}}(\mathbb{P};\mathbb{R})}<\infty$ and $\sup_{s\in[0,T]}\|g(\xi+\sqrt{s}Z^{(0)})\|_{L^{\frac{2p}{p-2}}(\mathbb{P};\mathbb{R})}<\infty$. This together with (136) proves that $\limsup_{n\to\infty}\eta_{n,\lfloor n^{2\beta}\rfloor}=0$. Let $N\in\mathbb{N}$ be the natural number given by

$$N=\min\left\{n\in\mathbb{N}\cap[2,\infty)\colon\sup_{m\in\mathbb{N}\cap[n,\infty)}\eta_{m,\lfloor m^{2\beta}\rfloor}\leq\varepsilon\right\} \tag{137}$$

and let $\mathfrak{C}\in[0,\infty)$ be the real number given by

$$\begin{aligned}\mathfrak{C}=2\Bigg[&1+\frac{\sqrt{\max\{T,3\}}\|K\|_1}{\sqrt{\lfloor(N-1)^{2\beta}\rfloor}}+C\sup_{s,t\in[0,T)}\left\|(F(0))(t,\xi+\sqrt{s}Z^{(0)})\right\|_{L^{\frac{2p}{p-2}}(\mathbb{P};\mathbb{R})}\left(\frac{1}{\sqrt{\lfloor(N-1)^{2\beta}\rfloor}}+Te^{L_0T}\right)\\&+Ce^{L_0T}(\|K\|_\infty+T\|\mathfrak{L}\|_\infty)+Ce^{L_0T}\sup_{s\in[0,T]}\|g(\xi+\sqrt{s}Z^{(0)})\|_{L^{\frac{2p}{p-2}}(\mathbb{P};\mathbb{R})}+TCe^{2L_0T}\sum_{j=1}^{d}L_j(K_j+T\mathfrak{L}_j)\Bigg].\end{aligned} \tag{138}$$

If $N=2$ then it holds that

$$\varepsilon \leq 1 \leq 4C\mathfrak{C}\left[e\left(\tfrac{p}{2}+1\right)\right]^{\frac{1}{8}} e^{\beta} = \frac{\mathfrak{C}\left[e\left(\frac{p(N-1)}{2}+1\right)\right]^{\frac{1}{8}}(4C)^{N-1}\exp\left(\beta(\lfloor (N-1)^{2\beta}\rfloor)^{\frac{1}{2\beta}}\right)}{\sqrt{(\lfloor (N-1)^{2\beta}\rfloor)^{N-2}}}. \tag{139}$$

If $N>2$ it follows from (137), (136) and (138) that

$$\varepsilon < \eta_{N-1,\lfloor (N-1)^{2\beta}\rfloor} \leq \frac{\mathfrak{C}\left[e\left(\frac{p(N-1)}{2}+1\right)\right]^{\frac{1}{8}}(4C)^{N-1}\exp\left(\beta(\lfloor (N-1)^{2\beta}\rfloor)^{\frac{1}{2\beta}}\right)}{\sqrt{(\lfloor (N-1)^{2\beta}\rfloor)^{N-2}}}. \tag{140}$$

Moreover, [39, Lemma 3.6] implies that for all $n \in \mathbb{N}$ it holds that $\mathrm{RV}_{n,\lfloor n^{2\beta}\rfloor} \leq d(5\lfloor n^{2\beta}\rfloor)^n$. This implies that

$$\sum_{n=1}^{N} \mathrm{RV}_{n,\lfloor n^{2\beta}\rfloor} \leq d\sum_{n=1}^{N}(5\lfloor n^{2\beta}\rfloor)^n \leq d\sum_{n=1}^{N}(5\lfloor N^{2\beta}\rfloor)^n = \frac{d(5\lfloor N^{2\beta})\rfloor)((5\lfloor N^{2\beta}\rfloor)^N-1)}{5\lfloor N^{2\beta}\rfloor-1} \leq 2d(5\lfloor N^{2\beta}\rfloor)^N. \tag{141}$$

Combining this with (139) and (140) proves that

$$\begin{aligned}
\sum_{n=1}^{N} \mathrm{RV}_{n,\lfloor n^{2\beta}\rfloor} &\leq 2d(5\lfloor N^{2\beta}\rfloor)^N = d(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{5^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{5^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{5^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{5^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] \\
&= 2d\varepsilon^{-(2+\delta)}\mathfrak{C}^{2+\delta}\left[\sup_{n\in\mathbb{N}\cap[2,\infty)}\left(\frac{5\left[(n-1)^{\frac{2\beta}{n}}\left[e\left(\frac{p(n-1)}{2}+1\right)\right]^{\frac{1}{8n}}(4Ce^{\beta})^{\frac{n-1}{n}}\right]^{2+\delta}}{(\lfloor (n-1)^{2\beta}\rfloor)^{\frac{\delta}{2}}}\right)^{n}\right] < \infty.
\end{aligned} \tag{142}$$

This establishes (133). The proof of Corollary 5.1 is thus completed. □

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 $\xi_d = (\xi_{d,1},\xi_{d,2},\ldots,\xi_{d,d}) \in \mathbb{R}^d$, $d \in \mathbb{N}$, 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 $F_d\colon C([0,T)\times\mathbb{R}^d,\mathbb{R}^{1+d}) \to C([0,T)\times\mathbb{R}^d,\mathbb{R})$, $d \in \mathbb{N}$, be functions, assume for all $d \in \mathbb{N}$, $t \in [0,T)$, $x = (x_1,x_2,\ldots,x_d)$, $\mathfrak{x} = (\mathfrak{x}_1,\mathfrak{x}_2,\ldots,\mathfrak{x}_d)$, $z = (z_1,z_2,\ldots,z_d)$, $\mathfrak{z} = (\mathfrak{z}_1,\mathfrak{z}_2,\ldots,\mathfrak{z}_d) \in \mathbb{R}^d$, $y,\mathfrak{y} \in \mathbb{R}$, $\mathbf{v} \in C([0,T)\times\mathbb{R}^d,\mathbb{R}^{1+d})$ that*

$$\max\{|f_d(t,x,y,z)-f_d(t,\mathfrak{x},\mathfrak{y},\mathfrak{z})|,|g_d(x)-g_d(\mathfrak{x})|\} \leq \textstyle\sum_{j=1}^{d} L_{d,j}\big(d^{\lambda}|x_j-\mathfrak{x}_j|+|y-\mathfrak{y}|+|z_j-\mathfrak{z}_j|\big), \tag{143}$$

$$\big(\tfrac{\partial}{\partial t}u_d\big)(t,x)+\tfrac{1}{2}(\Delta_x u_d)(t,x)+f_d\big(t,x,u(t,x),(\nabla_x u_d)(t,x)\big)=0, \qquad u_d(T,x)=g_d(x), \tag{144}$$

$$d^{-\lambda}(|g_d(0)|+|f_d(t,0,0,0)|+\max_{i\in\{1,2,\ldots,d\}}|\xi_{d,i}|)+\textstyle\sum_{i=1}^{d} L_{d,i} \leq \lambda, \quad \textit{and} \quad (F_d(\mathbf{v}))(t,x)=f_d(t,x,\mathbf{v}(t,x)), \tag{145}$$

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

$$
\begin{aligned}
\mathbf{U}^{d,\theta}_{n,M}(t,x) &= (g_d(x),0) + \sum_{i=1}^{M^n} \tfrac{1}{M^n}\big(g_d(x+[T-t]^{1/2}Z^{(\theta,0,-i)}) - g_d(x)\big)\big(1,[T-t]^{-1/2}Z^{d,(\theta,0,-i)}\big) \\
&\quad + \sum_{l=0}^{n-1}\sum_{i=1}^{M^{n-l}} \tfrac{(T-t)(\mathfrak{r}^{(\theta,l,i)})^{1-\alpha}}{\alpha M^{n-l}}\big(1,[(T-t)\mathfrak{r}^{(\theta,l,i)}]^{-1/2}Z^{d,(\theta,l,i)}\big) \\
&\quad \cdot \big[\big(F_d(\mathbf{U}^{d,(\theta,l,i)}_{l,M}) - \mathbb{1}_{\mathbb{N}}(l)F_d(\mathbf{U}^{d,(\theta,-l,i)}_{l-1,M})\big)\big(t+(T-t)\mathfrak{r}^{(\theta,l,i)}, x+[(T-t)\mathfrak{r}^{(\theta,l,i)}]^{1/2}Z^{d,(\theta,l,i)}\big)\big],
\end{aligned}
\tag{146}
$$

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

$$
\mathrm{RV}_{d,n,M} \le dM^n + \sum_{l=0}^{n-1}\Big[M^{(n-l)}(d+1+\mathrm{RV}_{d,l,M} + \mathbb{1}_{\mathbb{N}}(l)\,\mathrm{RV}_{d,l-1,M})\Big]. \tag{147}
$$

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

$$
\sup_{n\in\mathbb{N}\cap[N_{d,\varepsilon},\infty)} \Big[\mathbb{E}\big[|\mathbf{U}^{d,0,0}_{n,\lfloor n^\beta\rfloor}(0,\xi_d) - u_d(0,\xi_d)|^2\big] + \max_{i\in\{1,2,\ldots,d\}} \mathbb{E}\big[|\mathbf{U}^{d,0,i}_{n,\lfloor n^\beta\rfloor}(0,\xi_d) - (\tfrac{\partial}{\partial x_i}u_d)(0,\xi_d)|^2\big]\Big]^{1/2} \le \varepsilon. \tag{148}
$$

Proof of Theorem 5.2. Throughout this proof let $p = \frac{2\alpha}{\beta+2\alpha-1}$. Note that the fact that $0 < \beta < 1-\alpha$ ensures 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

$$
\frac{\beta}{2} = \frac{1-\alpha}{2} - \frac{(1-(1-\alpha))(p-2)}{2p}. \tag{149}
$$

If $\alpha > \frac{1}{2}$, then it holds that $1-\alpha < \frac{p}{2(p-1)}$. Moreover, if $\alpha > \frac{1}{2}$ the fact that $\beta > 0$ implies that $p < \frac{2\alpha}{2\alpha-1}$ and hence $1-\alpha > \frac{p-2}{2(p-1)}$. If $\alpha < \frac{1}{2}$, then it holds that $1-\alpha > \frac{p-2}{2(p-1)}$. Moreover, if $\alpha < \frac{1}{2}$ the fact that $\beta > \frac{1-2\alpha}{1-\alpha}$ implies that $p < \frac{2\alpha}{\frac{1-2\alpha}{1-\alpha}+2\alpha-1} = \frac{1-\alpha}{\frac{1}{2}-\alpha}$ and hence $1-\alpha < \frac{p}{2(p-1)}$. Furthermore, it holds that $\frac{p-2}{2(p-1)} < \frac{1}{2} < \frac{p}{2(p-1)}$. To summarize, it holds that

$$
p \in (2,\infty), \quad 1-\alpha \in \left(\frac{p-2}{2(p-1)}, \frac{p}{2(p-1)}\right), \quad \text{and} \quad \frac{\beta}{2} = \frac{1-\alpha}{2} - \frac{(1-(1-\alpha))(p-2)}{2p}. \tag{150}
$$

Next note that (143) ensures for all $d \in \mathbb{N}$ that

$$
\begin{aligned}
&\sup_{s,t\in[0,T)} \Big\|(F_d(0))(t,\xi_d+\sqrt{s}Z^{d,(0)})\Big\|_{L^{\frac{2p}{p-2}}(\mathbb{P};\mathbb{R})} \\
&\le \sup_{s,t\in[0,T)} \Big\|(F_d(0))(t,\xi_d+\sqrt{s}Z^{d,(0)}) - (F_d(0))(t,0)\Big\|_{L^{\frac{2p}{p-2}}(\mathbb{P};\mathbb{R})} + \sup_{t\in[0,T)} |(F_d(0))(t,0)| \\
&= \sup_{s,t\in[0,T)} \Big\|f_d(t,\xi_d+\sqrt{s}Z^{d,(0)},0,0) - f_d(t,0,0,0)\Big\|_{L^{\frac{2p}{p-2}}(\mathbb{P};\mathbb{R})} + \sup_{t\in[0,T)} |f_d(t,0,0,0)| \\
&\le \bigg(d^\lambda \sum_{j=1}^d L_{d,j}\big(|\xi_{d,j}| + \sqrt{T}\Big\|Z^{1,(0)}\Big\|_{L^{\frac{2p}{p-2}}(\mathbb{P};\mathbb{R})}\big)\bigg) + \sup_{t\in[0,T)} |f_d(t,0,0,0)| \\
&\le d^\lambda\bigg(\big[\max_{j\in\{1,2,\ldots,d\}} |\xi_{d,j}|\big] + \sqrt{T}\Big\|Z^{1,(0)}\Big\|_{L^{\frac{2p}{p-2}}(\mathbb{P};\mathbb{R})}\bigg)\bigg(\sum_{j=1}^d L_{d,j}\bigg) + \sup_{t\in[0,T)} |f_d(t,0,0,0)|.
\end{aligned}
\tag{151}
$$

Moreover, (143) proves for all $d \in \mathbb{N}$ that

$$
\begin{aligned}
&\sup_{s\in[0,T]} \|g_d(\xi_d+\sqrt{s}Z^{d,(0)})\|_{L^{\frac{2p}{p-2}}(\mathbb{P};\mathbb{R})} \le \sup_{s\in[0,T]} \|g_d(\xi_d+\sqrt{s}Z^{d,(0)}) - g_d(0)\|_{L^{\frac{2p}{p-2}}(\mathbb{P};\mathbb{R})} + |g_d(0)| \\
&\le \bigg(d^\lambda \sum_{j=1}^d L_{d,j}\big(|\xi_{d,j}| + \sqrt{T}\|Z^{1,(0)}\|_{L^{\frac{2p}{p-2}}(\mathbb{P};\mathbb{R})}\big)\bigg) + |g_d(0)| \\
&\le d^\lambda\bigg(\big[\max_{j\in\{1,2,\ldots,d\}} |\xi_{d,j}|\big] + \sqrt{T}\Big\|Z^{1,(0)}\Big\|_{L^{\frac{2p}{p-2}}(\mathbb{P};\mathbb{R})}\bigg)\bigg(\sum_{j=1}^d L_{d,j}\bigg) + |g_d(0)|.
\end{aligned}
\tag{152}
$$

Furthermore, it holds for all $d \in \mathbb{N}$ that

$$\max_{j\in\{1,2,\ldots,d\}} d^\lambda L_{j,d} \le d^\lambda \sum_{j=1}^d L_{j,d}, \quad \text{and} \quad \sum_{j=1}^d L_{d,j}(d^\lambda L_{d,j} + T d^\lambda L_{d,j}) \le d^\lambda (T+1)\Big(\sum_{j=1}^d L_{d,j}\Big)^2. \tag{153}$$

Combining (151), (152), and (153) with Corollary 5.1 (applied with $\alpha = 1-\alpha$, $\beta = \frac{\beta}{2}$, $p = \frac{1-\alpha}{\frac{\beta}{2}-\alpha+\frac{1}{2}}$, $L_0 = \sum_{j=1}^d L_{d,j}$, $L_j = L_{d,j}$, $K_j = d^\lambda L_{d,j}$, $\mathfrak{L}_j = d^\lambda L_{d,j}$ for $j \in \{1,\ldots,d\}$, $d \in \mathbb{N}$ in the notation of Corollary 5.1) and the fact that

$$\sup_{d\in\mathbb{N}} \left[\frac{1}{d^\lambda}\left(\max_{i\in\{1,2,\ldots,d\}} |\xi_{d,i}| + \sup_{t\in[0,T]} |f_d(t,0,0,0)| + |g_d(0)|\right) + \sum_{i=1}^d L_{d,i}\right] < \infty \tag{154}$$

proves (148). The proof of Theorem 5.2 is thus completed. □

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 $Z = (Z^1, Z^2, \ldots, Z^d)\colon \Omega \to \mathbb{R}^d$ be a standard normal random variable, let $\mathfrak{r}\colon \Omega \to (0,1)$ satisfy for all $b \in (0,1)$ that $\mathbb{P}(\mathfrak{r} \le b) = b^\alpha$, assume that Z and $\mathfrak{r}$ are independent, and let $\mathbf{U} = (\mathbf{U}^0, \mathbf{U}^1, \ldots, \mathbf{U}^d)\colon \Omega \to \mathbb{R}^{d+1}$ satisfy that $\mathbf{U} = T\alpha^{-1}\mathfrak{r}^{1-\alpha}\big(1, [T\mathfrak{r}]^{-1/2} Z\big)$. Then*

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

$$\mathbb{E}\big[|\mathbf{U}^i|^2\big] = \frac{T}{\alpha}\int_0^1 s^{-\alpha}\,ds = \begin{cases} \frac{T}{\alpha(1-\alpha)} & : \alpha < 1 \\ \infty & : \alpha = 1 \end{cases} \tag{155}$$

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 $i \in \{1,2,\ldots,d\}$ it holds that

$$\begin{aligned} \mathbb{E}\big[|\mathbf{U}^i|^2\big] &= \frac{T}{\alpha^2}\mathbb{E}\big[\mathfrak{r}^{2(1-\alpha)-1}|Z^i|^2\big] = \frac{T}{\alpha}\int_0^1 \mathbb{E}\big[s^{2(1-\alpha)-1}|Z^i|^2\big]s^{\alpha-1}\,ds = \frac{T}{\alpha}\int_0^1 s^{2(1-\alpha)-1+\alpha-1}\,ds \\ &= \frac{T}{\alpha}\int_0^1 s^{-\alpha}\,ds = \begin{cases} \frac{T}{\alpha(1-\alpha)} & : \alpha < 1 \\ \infty & : \alpha = 1 \end{cases}. \end{aligned} \tag{156}$$

This proves item (i). Moreover, observe that item (i) establishes item (ii). The proof of Lemma 5.3 is thus completed. □

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 $\xi_d = (\xi_{d,1}, \xi_{d,2}, \ldots, \xi_{d,d}) \in \mathbb{R}^d$, $d \in \mathbb{N}$, 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 $d \in \mathbb{N}$, $t \in [0,T)$, $x = (x_1, x_2, \ldots, x_d)$, $\mathfrak{x} = (\mathfrak{x}_1, \mathfrak{x}_2, \ldots, \mathfrak{x}_d)$, $z = (z_1, z_2, \ldots, z_d)$, $\mathfrak{z} = (\mathfrak{z}_1, \mathfrak{z}_2, \ldots, \mathfrak{z}_d) \in \mathbb{R}^d$, $y, \mathfrak{y} \in \mathbb{R}$ that*

$$\max\{|f_d(y,z) - f_d(\mathfrak{y},\mathfrak{z})|, |g_d(x) - g_d(\mathfrak{x})|\} \le \textstyle\sum_{j=1}^d L_{d,j}\big(d^\lambda|x_j - \mathfrak{x}_j| + |y-\mathfrak{y}| + |z_j - \mathfrak{z}_j|\big), \tag{157}$$

$$\big(\tfrac{\partial}{\partial t}u_d\big)(t,x) = (\Delta_x u_d)(t,x) + f_d\big(u(t,x), (\nabla_x u_d)(t,x)\big), \qquad u_d(0,x) = g_d(x), \tag{158}$$

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

$$\begin{aligned} \mathbf{U}^{d,\theta}_{n,M}(t,x) &= (g_d(x), 0) + \sum_{i=1}^{M^n} \tfrac{1}{M^n}\big(g_d(x + [2t]^{1/2} Z^{(\theta,0,-i)}) - g_d(x)\big)\big(1, [2t]^{-1/2} Z^{d,(\theta,0,-i)}\big) \\ &\quad + \sum_{l=0}^{n-1}\sum_{i=1}^{M^{n-l}} \tfrac{2t[\mathfrak{r}^{(\theta,l,i)}]^{1/2}}{M^{n-l}}\big[f_d\big(\mathbf{U}^{d,(\theta,l,i)}_{l,M}(t(1-\mathfrak{r}^{(\theta,l,i)}), x + [2t\mathfrak{r}^{(\theta,l,i)}]^{1/2} Z^{d,(\theta,l,i)})\big) \\ &\quad - \mathbb{1}_{\mathbb{N}}(l) f_d\big(\mathbf{U}^{d,(\theta,-l,i)}_{l-1,M}(t(1-\mathfrak{r}^{(\theta,l,i)}), x + [2t\mathfrak{r}^{(\theta,l,i)}]^{1/2} Z^{d,(\theta,l,i)})\big)\big]\big(1, [2t\mathfrak{r}^{(\theta,l,i)}]^{-1/2} Z^{d,(\theta,l,i)}\big), \end{aligned} \tag{159}$$

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

$$\mathrm{RV}_{d,n,M} \le dM^n + \sum_{l=0}^{n-1}\Big[M^{(n-l)}(d + 1 + \mathrm{RV}_{d,l,M} + \mathbb{1}_{\mathbb{N}}(l)\,\mathrm{RV}_{d,l-1,M})\Big]. \tag{160}$$

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

$$\sup_{n\in\mathbb{N}\cap[N_{d,\varepsilon},\infty)} \Big[\mathbb{E}\big[|\mathbf{U}^{d,0,0}_{n,\lfloor n^{1/4}\rfloor}(T,\xi_d) - u_d(T,\xi_d)|^2\big] + \max_{i\in\{1,2,\ldots,d\}} \mathbb{E}\big[|\mathbf{U}^{d,0,i}_{n,\lfloor n^{1/4}\rfloor}(T,\xi_d) - (\tfrac{\partial}{\partial x_i} u_d)(T,\xi_d)|^2\big]\Big]^{1/2} \leq \varepsilon. \quad (161)$$

Proof of Corollary 5.4. Corollary 5.4 is a direct consequence of 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). □

Acknowledgement

This project has been partially supported by the Deutsche Forschungsgesellschaft (DFG) via RTG 2131 *High-dimensional Phenomena in Probability – Fluctuations and Discontinuity* and via research grant HU 1889/6-1.

References

[1] Beck, C., Becker, S., Cheridito, P., Jentzen, A., and Neufeld, A. Deep splitting method for parabolic PDEs. *arXiv:1907.03452* (2019).

[2] Beck, C., Becker, S., Grohs, P., Jaafari, N., and Jentzen, A. Solving stochastic differential equations and Kolmogorov equations by means of deep learning. *arXiv:1806.00421* (2018).

[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. *arXiv:1907.06729* (2019).

[5] Becker, S., Cheridito, P., and Jentzen, A. Deep optimal stopping. *Journal of Machine Learning Research 20*, 74 (2019), 1–25.

[6] Becker, S., Cheridito, P., Jentzen, A., and Welti, T. Solving high-dimensional optimal stopping problems using deep learning. *arXiv:1908.01602* (2019), 42 pages.

[7] Berg, J., and Nyström, K. A unified deep artificial neural network approach to partial differential equations in complex geometries. *Neurocomputing 317* (2018), 28–41.

[8] 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. *arXiv:1809.03062, Revision requested from SIAM J. Math. Data Sci.* (2018).

[9] 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.

[10] 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.

[11] Briand, P., and Labart, C. Simulation of BSDEs by Wiener chaos expansion. *The Annals of Applied Probability 24*, 3 (2014), 1129–1171.

[12] Chan-Wai-Nam, Q., Mikael, J., and Warin, X. Machine learning for semi linear PDEs. *J. Sci. Comput. 79*, 3 (2019), 1667–1712.

[13] 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* (2019).

[14] Dockhorn, T. A discussion on solving partial differential equations using neural networks. *arXiv:1904.07200* (2019).

[15] 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.

[16] E, W., Hutzenthaler, M., Jentzen, A., and Kruse, T. Multilevel Picard iterations for solving smooth semilinear parabolic heat equations. *arXiv:1607.03295* (2016).

[17] 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.

[18] 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.

[19] El Karoui, N., Peng, S., and Quenez, M. C. Backward stochastic differential equations in finance. *Mathematical finance 7*, 1 (1997), 1–71.

[20] Elbrächter, D., Grohs, P., Jentzen, A., and Schwab, C. DNN Expression Rate Analysis of High-dimensional PDEs: Application to Option Pricing. *arXiv:1809.07669* (2018).

[21] Farahmand, A.-m., Nabi, S., and Nikovski, D. Deep reinforcement learning for partial differential equation control. *2017 American Control Conference (ACC)* (2017), 3120–3127.

[22] Fujii, M., Takahashi, A., and Takahashi, M. Asymptotic Expansion as Prior Knowledge in Deep Learning Method for high dimensional BSDEs. *arXiv:1710.07030* (2017).

[23] Geiss, C., and Labart, C. Simulation of BSDEs with jumps by Wiener chaos expansion. *Stochastic processes and their applications 126*, 7 (2016), 2123–2162.

[24] Giles, M. B., Jentzen, A., and Welti, T. Generalised multilevel Picard approximations. *arXiv:1911.03188* (2019).

[25] 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.

[26] 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.

[27] Goudenège, L., Molent, A., and Zanette, A. Machine Learning for Pricing American Options in High Dimension. *arXiv:1903.11275* (2019), 11 pages.

[28] 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. *to appear in Mem. Amer. Math. Soc.* (2019).

[29] Grohs, P., Hornung, F., Jentzen, A., and Zimmermann, P. Space-time error estimates for deep neural network approximations for differential equations. *arXiv:1908.03833* (2019).

[30] Grohs, P., Jentzen, A., and Salimova, D. Deep neural network approximations for monte carlo algorithms. *arXiv:1908.10828* (2019).

[31] 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.

[32] Han, J., and Long, J. Convergence of the Deep BSDE Method for Coupled FBSDEs. *arXiv:1811.01165* (2018).

[33] Henry-Labordère, P. Counterparty risk valuation: a marked branching diffusion approach. *arXiv:1203.2369* (2012).

[34] Henry-Labordère, P. Deep Primal-Dual Algorithm for BSDEs: Applications of Machine Learning to CVA and IM. *Available at SSRN: http://dx.doi.org/10.2139/ssrn.3071506* (2017).

[35] 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.

[36] 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.

[37] Huré, C., Pham, H., and Warin, X. Some machine learning schemes for high-dimensional nonlinear PDEs. *arXiv:1902.01599* (2019).

[38] 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. *arXiv:1901.10854, Revision requested from SN Partial Differential Equations and Applications* (2019).

[39] 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. *arXiv:1807.01212* (2018).

[40] Hutzenthaler, M., Jentzen, A., and von Wurstemberger, P. Overcoming the curse of dimensionality in the approximative pricing of financial derivatives with default risks. *arXiv:1903.05985* (2019).

[41] Hutzenthaler, M., and Kruse, T. Multi-level Picard approximations of high-dimensional semilinear parabolic differential equations with gradient-dependent nonlinearities. *to appear in SIAM J. Numer. Anal.* (2019).

[42] Jacquier, A., and Oumgari, M. Deep PPDEs for rough local stochastic volatility. *arXiv:1906.02551* (2019).

[43] 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* (2018).

[44] 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.

[45] Kutyniok, G., Petersen, P., Raslan, M., and Schneider, R. A theoretical analysis of deep neural networks and parametric PDEs. *arXiv:1904.00377* (2019).

[46] 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.

[47] 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.

[48] 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.

[49] Lye, K. O., Mishra, S., and Ray, D. Deep learning observables in computational fluid dynamics. *arXiv:1903.03040* (2019).

[50] 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.

[51] 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.

[52] Pham, H., and Warin, X. Neural networks-based backward scheme for fully nonlinear PDEs. *arXiv:1908.00412* (2019).

[53] Qi, F. Bounds for the ratio of two gamma functions. *Journal of Inequalities and Applications 2010*, 1 (2010), 493058.

[54] Raissi, M. Forward-Backward Stochastic Neural Networks: Deep Learning of High-dimensional Partial Differential Equations. *arXiv:1804.07010* (2018).

[55] 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* (2019).

[56] Sirignano, J., and Spiliopoulos, K. DGM: A deep learning algorithm for solving partial differential equations. *arXiv:1708.07469* (2017).

[57] 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.

[58] Wendel, J. Note on the gamma function. *The American Mathematical Monthly 55*, 9 (1948), 563–564.